NATIONAL TECHNICAL
UNIVERSITY OF ATHENS
SCHOOL OF RURAL AND
SURVEYING ENGINEERING

# Georeferencing Procedures for Oblique Aerial Images

## Doctoral Dissertation

Styliani Verykokou


National Technical University of Athens
School of Rural and Surveying Engineering

# Georeferencing Procedures for Oblique Aerial Images

Doctoral Dissertation

**Styliani Verykokou**
Dipl. Rural and Surveying Engineer, NTUA

Supervisor:
Charalabos Ioannidis, Professor, NTUA

Athens, February 2020

Advisory Committee:

**Charalabos Ioannidis**, Professor, NTUA, Supervisor
**George Karras**, Professor Emeritus, NTUA
**Nikolaos Doulamis**, Associate Professor, NTUA

Examination Committee:

**Charalabos Ioannidis**, Professor, NTUA
**Andreas Georgopoulos**, Professor, NTUA
**Nikolaos Doulamis**, Associate Professor, NTUA
**Demetre Argialas**, Professor, NTUA
**Konstantinos Karantzalos**, Associate Professor, NTUA
**Olga Georgoula**, Professor, AUTh
**Vassilios Tsioukas**, Professor, AUTh

Co-financed by Greece and the European Union

This research has been co-financed by:

- the Eugenides Foundation, Greece, and
- the European Union and Greek national funds through the Operational Program "Competitiveness, Entrepreneurship and Innovation" of the Partnership and Cooperation Agreement for the period 2014-2020, under the Call "Research – Create – Innovate" (project code: T1EΔK-02589).


Εθνικό Μετσόβιο Πολυτεχνείο

Σχολή Αγρονόμων και Τοπογράφων Μηχανικών

# Διερεύνηση Μεθόδων Γεωαναφοράς για Πλάγιες Εναέριες Εικόνες

Διδακτορική Διατριβή

**Στυλιανή Βερυκόκου**

Διπλωματούχος Αγρονόμος και Τοπογράφος Μηχανικός ΕΜΠ

Επιβλέπων Καθηγητής:

Χαράλαμπος Ιωαννίδης, Καθηγητής ΕΜΠ

Αθήνα, Φεβρουάριος 2020

Τριμελής Συμβουλευτική Επιτροπή:

**Χαράλαμπος Ιωαννίδης**, Καθηγητής ΕΜΠ, Επιβλέπων
**Γεώργιος Καρράς**, Ομότιμος Καθηγητής ΕΜΠ
**Νικόλαος Δουλάμης**, Αναπληρωτής Καθηγητής ΕΜΠ

Επταμελής Εξεταστική Επιτροπή:

**Χαράλαμπος Ιωαννίδης**, Καθηγητής ΕΜΠ
**Ανδρέας Γεωργόπουλος**, Καθηγητής ΕΜΠ
**Νικόλαος Δουλάμης**, Αναπληρωτής Καθηγητής ΕΜΠ
**Δημήτριος Αργιαλάς**, Καθηγητής ΕΜΠ
**Κωνσταντίνος Καράντζαλος**, Αναπληρωτής Καθηγητής ΕΜΠ
**Όλγα Γεωργούλα**, Καθηγήτρια ΑΠΘ
**Βασίλειος Τσιούκας**, Καθηγητής ΑΠΘ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Η εκπόνηση της διδακτορικής διατριβής συγχρηματοδοτήθηκε από:

- το Ίδρυμα Ευγενίδου και
- την Ευρωπαϊκή Ένωση και εθνικούς πόρους μέσω του Επιχειρησιακού Προγράμματος «Ανταγωνιστικότητα, Επιχειρηματικότητα και Καινοτομία» του Συμφώνου Εταιρικής Σχέσης για την περίοδο 2014-2020, υπό την Ενιαία Δράση «Ερευνώ – Δημιουργώ – Καινοτομώ» (κωδικός προγράμματος: ΤΙΕΔΚ-02589).

***To my parents, Andreas and Tasoula***

*All that I am, or hope to be, I owe to them*

# Acknowledgements

The past five and a half years of my doctoral studies have been a challenging and intense, yet fruitful and enjoyable period. At the end of this unforgettable journey, I would like to extend my heartful thanks to those who stood next to me and supported me during this period.

First and foremost, my greatest gratitude, highest respect and most sincere appreciation are expressed to my supervisor, ***Prof. Charalabos Ioannidis***, who gave me the opportunity to pursue doctoral studies, opened my eyes in photogrammetry and made me love the research. His mentorship has played a monumental role during this doctoral dissertation. From the very beginning, he has been a great supporter and mentor; the trust he provided me throughout my doctoral studies as well as his valuable scientific advice and encouragement have been a particularly precious resource in my efforts. He has guided my research in the right direction, while simultaneously according me a great degree of freedom. I could not be happier to have had him as my supervisor. *Dear Prof. Ioannidis, I would like to thank you for your guidance, for your advice when the paths of scientific research seemed to me insurmountable, as well as for your complete moral support in some difficult moments during my doctoral studies. You have always been there, whenever I needed you. I thank you for our endless scientific discussions; even in weekends and on holiday periods; even during your travels abroad. I thank you for the great opportunities you gave me to collaborate with you and your team in two research projects, namely, "5DMuPLIS" and "METEORA". It has always been an honor to work with you. Moreover, I gratefully acknowledge the opportunities you gave me to travel abroad both for presenting our research in scientific conferences and for working on research projects. Dear Prof. Ioannidis, thanks to the impetus you provided me, I managed to travel to Limassol and Paphos, Cyprus; Sofia, Bulgaria; Tel Aviv and Jerusalim, Israel; Delft, the Netherlands; and Prague, Czech Republic, and gain unforgettable memories and experience. I thank you for everything.*

At the same time, I would like to thank the two other members of the advisory committee, ***Prof. George Karras*** and ***Prof. Nikolaos Doulamis***, for their support and important advice, as well as for their very constructive comments on my doctoral dissertation. Also, I would like to thank all the other members of the examination committee, ***Prof. Andreas Georgopoulos***, ***Prof. Demetre Argialas***, ***Prof. Konstantinos Karantzalos***, ***Prof. Olga Georgoula*** and ***Prof. Vassilios Tsioukas***, for their willingness to participate in the examination of this doctoral dissertation.

I will never forget that ***Prof. George Karras*** was the man who urged me to pursue doctoral studies in photogrammetry, when I entered his office and I asked for a recommendation letter for a master program. *Dear Prof. Karras, I will always be grateful for this recommendation and trust. You inspired me during your undergraduate course in photogrammetry and you were always willing to*

*watch the progress of my doctoral studies and provide me with particularly valuable comments that helped to greatly improve my research.*

***Prof. Nikolaos Doulamis*** was the man who encouraged me to engage in C++ programming and was always there to advise me. *Dear Prof. Doulamis, I would like to thank you for giving me the opportunity to cooperate with you under several research projects, for our daily conversations in the Lab. of Photogrammetry and for your interest towards my work.*

Also, I would like to thank ***Prof. Chryssy Potsiou***, for granting me the opportunity to closely cooperate with her under the research project "5DMuPLIS" and for always being there to advise me and inspire me to work hard and achieve my goals. *Dear Prof. Potsiou, dear Mrs. Chryssy, I have learned a lot from our cooperation. I thank you for being truly there for me and believing in me.*

Furthermore, I would like to thank ***Prof. Anastasios Doulamis***, for granting me the great opportunity to closely work with him under two research projects, namely, "INACHUS" and "4D-CH-World". *Dear Prof. Doulamis, I would like to thank you for trusting me and believing in me. I have gained very precious experience and knowledge from these research projects. I thank you for giving me such opportunities and always motivating me to work hard.*

Also, I would like to thank ***Prof. Andreas Georgopoulos***, the director of the Lab. of Photogrammetry, for giving me the great opportunity to work in the research project "State-of-the-art mapping technologies for Public Work Studies and Environmental Impact Assessment Studies". Although I was not lucky to have him as a professor in any of my undergraduate photogrammetry lessons, he was always interested in my work and asked me for the progress of my doctoral studies. *Dear Prof. Georgopoulos, I thank you for that. The Lab. of Photogrammetry has been my home for the past five and a half years.*

In addition, I would like to express my sincere thanks to the ***Eugenides Foundation***, for the three years of financial support, that greatly contributed in making my doctoral research come to life. The doctoral scholarship that I received by the Eugenides Foundation was especially significant for me and let me focus on my doctoral studies. I gratefully acknowledge this support.

Furthermore, I would like to acknowledge the provision of datasets used throughout this doctoral dissertation by ***ISPRS*** and ***EuroSDR***, acquired in the context of the ISPRS scientific initiatives "ISPRS/EuroSDR benchmark on high density image matching for DSM computation" and "ISPRS benchmark for multi-platform photogrammetry".

Moreover, I would like to thank all my colleagues and friends in the Lab. of Photogrammetry for sharing this amazing journey with me. It would be an omission if I did not particularly thank ***Ms. Sofia Soile***, Msc Eng., member of the Lab. of Photogrammetry, for our very close and enjoyable cooperation all these years. *Dear Ms. Sofia, I have learned a lot working with you; but most importantly, I have gained a friend. Thank you for our great moments in the Lab. of Photogrammetry. Thank you for our unforgettable moments during our travels abroad – especially, our travel in Delft; I will always have the best memories of it.*

On a personal level, the love and continuous interest from my beloved ones is very much appreciated. I would like to sincerely thank all my friends and relatives who supported me throughout my endeavor and showed understanding whenever the volume of work was prohibitive for other activities. Special thanks go to my lifelong friends, for always being there for me, and my

wonderful cousins, for verifying every day that even though we are 500 kilometers apart, distance cannot make our hearts apart.

I have been blessed with a very loving and supportive family. Nobody has been more important for me in the completion of my doctoral dissertation than my family members. Undoubtedly the biggest thanks goes to my parents, ***Andreas Verykokos*** and ***Tasoula Verykokou***. Without them by my side, the completion of my doctoral dissertation would never be possible. They have always been my precious, constant and unwavering support. They believed in me from an early age and offered me all the skills to complete myself as a human and as a scientist. I can never express my gratitude to them for their unconditional love, patience, trust and support. *Mom, Dad, you were always there for me to provide your deepest love and countless hugs; to encourage me and cherish me; to advise me and help me in any means; Thank you for believing in me. Thank you for supporting me. Thank you for your patience. Thank you for taking care of my son – and me – during this period. Thank you for being so wonderful parents and even more wonderful grandparents. Mom, thank you for sacrificing your personal time to offer to me your unconditional help all these years and especially since Christos's arrival in our lives. Your help was very meaningful for me and gave me the time to dedicate myself to my child and my doctoral studies. I have no idea how I would have made it without you. I can never express my gratitude for that. Dad, without you pushing me and encouraging me, always being informed for the status of my experiments and being truly interested in my studies, reading my papers – every single one – and being proud of me, admiring my achievements and always having faith in me, I could not have managed to finish my doctoral dissertation. You have been the person that pushed me to start this life-changing journey of my doctoral studies and morally supported me during this. You were the person to whom I was complaining about any failure of my experiments and with whom I was celebrating any success. Mom, Dad, this doctoral dissertation is also your accomplishment, hence I dedicate it to you.*

As for my husband, ***Konstantinos Pierroutsakos***, I find it difficult to express my love and appreciation, because it is boundless. He is my best friend; my partner in life; an amazing husband and father; the person who has seen me through the ups and downs of the entire doctoral process. *Konstantinos, without you I would not have been able to balance my research with everything else. Thank you for joining me in this scholarly adventure; thank you for understanding me when I worked during weekends and whenever I dedicated our personal time to my doctoral studies; thank you for your patience when I was stressed; thank you for always being by my side; and especially thank you for making with me a wonderful family.*

***Christos***, my baby boy, has been the light of my life for the last two and a half years. He came in our lives in the middle of this doctoral journey and filled it with joy. Considering that he is a toddler, it is, indeed, hard to imagine how he could contribute to my doctoral dissertation, but, in fact, it is his bright little smile that was filling me with hope and was giving me a reason to keep going. During difficult periods of my doctoral studies, he was always there to put a smile on my face and he was always giving me the extra strength and motivation to accomplish my goals. *Christos, thank you for making me a mom. You are the most precious gift in our lives. We love you so much, more than words could ever say.*

I would also like to say a big thanks to my aunt ***Voula Kontou***, who has always been there for me – as far back as I can remember myself – to love me, support me, take care of me, be proud of me, laugh with me and talk with me. Now she does the same with my son. *My dearest Voula, thank you for your so special presence in our lives; thank you for being the most wonderful and incredible*

*aunt I have ever met; thank you for loving my son with all your heart; thank you for taking care of him; and especially thank you for the fact that Christos loves you so much.*

Finally, my grandparents, ***Dimitris Verykokos*** and ***Stella Verykokou***, have been a source of constant and unconditional love for me for as long as I can remember. Although they have passed away, they deserve my most sincere love and appreciation. My grandfather was my most enthusiastic supporter. He was always showing how proud he is of me and how much he was loving me. My grandmother was always there to take care of me and advise me. I am grateful to the two of them. *I love you with all my heart.*

*Styliani Verykokou*

# Abstract

Although vertical aerial images have played the leading role in photogrammetric applications for more than a century, in recent years oblique aerial images have gained popularity, mainly because of certain fundamental advantages they provide compared to nadir views, in combination with the progress made in photogrammetric and computer vision algorithms that enable their automatic processing. The main subject of this dissertation is the metric exploitation of datasets containing oblique aerial images, with emphasis on the establishment of automatic georeferencing algorithms, both in terms of exterior orientation estimation and in terms of image-to-ground 2D transformation, i.e., polynomial/projective transformation that approximates the relationship between each image and the ground reference system. Furthermore, it investigates how the automatic extraction of elements describing the geometry of the scene depicted in oblique aerial images contributes in their metric exploitation, focusing on georeferencing procedures. All methods developed throughout this dissertation, both for single images and for multi-image datasets, do not require any knowledge of approximate positioning and orientation data (e.g., from onboard sensors), so that they can also be applied in cases where such data are not available (e.g., amateur images from unmanned aerial vehicles, old datasets). Moreover, they require minimum user interaction, being in principle easily adoptable by operators without expertise or even basic knowledge of photogrammetry. Several experiments with different datasets of oblique aerial images are conducted through developed software solutions that implement the proposed algorithms. In addition, error analysis, investigation of the impact of different variables on the results of each method and their comparison with those obtained by well-known existing software packages are performed.

Specifically, an algorithm for automatic rough georeferencing of large datasets of multi-perspective oblique and vertical aerial images of the same unknown interior orientation and flying height is established, in terms of estimating the 2D transformation from each image to the ground reference system, along with image rectification as well as extraction of the ground footprints of the images. The method requires the measurement of a minimum number of points of known horizontal coordinates in one image. Also, the geometry of oblique aerial images of man-made environments favors the automatic extraction of points lying on planes as well as sets of parallel lines, mainly horizontal and vertical ones, and thereby the automatic detection of vanishing points. The research conducted throughout this dissertation proves that knowledge of these quantities, i.e., coplanar points as well as horizontal and vertical lines and corresponding vanishing points, may be used for rough georeferencing purposes and metric exploitation of single oblique images. In this context, automatic methods for detecting coplanar points and vanishing points in oblique aerial images are introduced; exterior orientation techniques that adopt these methods are established; a method for automatic transfer of coplanar GCPs in multiple images based on their measurements in a single image, for exterior orientation estimation purposes, is proposed; a 2D georeferencing framework for images of a piecewise planar scene is introduced; and a method for measuring vertical and

horizontal distances from a single unoriented oblique image is presented. Furthermore, robust photogrammetry-based incremental and global structure from motion (SfM) workflows that can be applied in challenging datasets of oblique aerial images are established. The developed SfM algorithms eliminate all erroneous feature points through combination of multiple geometric constraints and a robust iterative bundle adjustment framework, which improves the accuracy of the exterior orientation results. Finally, a scale-based weighting strategy for feature point observations in bundle adjustment is introduced, targeted to highly overlapping oblique imagery.

# Εκτεταμένη Περίληψη στα Ελληνικά

Η ραγδαία πρόοδος που σημειώθηκε τα τελευταία χρόνια στην ψηφιακή τεχνολογία δημιούργησε τις προϋποθέσεις για την εντυπωσιακή εξέλιξη των επιστημονικών περιοχών της φωτογραμμετρίας και της όρασης υπολογιστών. Το γνωστικό τους πεδίο διευρύνεται συνεχώς και οι μεθοδολογίες τους αναμορφώνονται και εξελίσσονται ώστε να ικανοποιούν τις απαιτήσεις των σύγχρονων εφαρμογών. Έτσι, ενώ στο παρελθόν η διαδικασία εξαγωγής μετρητικής πληροφορίας από αεροφωτογραφίες στηριζόταν σχεδόν αποκλειστικά σε κατακόρυφες εναέριες λήψεις, τα τελευταία χρόνια οι πλάγιες εναέριες εικόνες ήλθαν στο προσκήνιο και αποτελούν όχι μόνο ένα συμπληρωματικό σύνολο δεδομένων για τις παραδοσιακές κατακόρυφες εικόνες, αλλά και τη βασική πηγή πληροφοριών για διάφορες εφαρμογές. Προς αυτή την κατεύθυνση, έχει διεξαχθεί σημαντική έρευνα για εφαρμογές οι οποίες αξιοποιούν σύνολα πλάγιων εναέριων εικόνων κυρίως από μετρητικές φωτομηχανές ή συστήματα πολλαπλών φωτομηχανών, υπό την προϋπόθεση ύπαρξης δεδομένων γεωαναφοράς από αισθητήρες καταγραφής θέσης και αδρανειακούς αισθητήρες (GNSS/INS) και πολυεικονικών μετρήσεων σημείων γνωστών συντεταγμένων. Ωστόσο, η έρευνα που έχει διεξαχθεί σχετικά με τη μετρητική αξιοποίηση πλάγιων εναέριων εικόνων που δεν συνοδεύονται από δεδομένα θέσης και προσανατολισμού και διαθέτουν ελάχιστο αριθμό σημείων γνωστών συντεταγμένων χωρίς καλή κατανομή είναι πολύ περιορισμένη. Ταυτόχρονα, η γεωαναφορά τους αποτελεί μια πολύ σημαντική διαδικασία που είναι προαπαιτούμενη για την πλειονότητα των φωτογραμμετρικών εφαρμογών.

Στο πλαίσιο αυτό και υπό το πρίσμα των ανωτέρω εξελίξεων, στόχος της διδακτορικής διατριβής είναι η ανάπτυξη αλγορίθμων φωτογραμμετρίας και όρασης υπολογιστών για τη μετρητική αξιοποίηση πλάγιων εναέριων εικόνων, με έμφαση σε διαδικασίες γεωαναφοράς πλάγιων εικόνων και συνδυασμένων συνόλων κατακόρυφων και πλάγιων εικόνων πολλαπλών προοπτικών, που στηρίζονται στον ελάχιστο δυνατό αριθμό μετρήσεων και δεν απαιτούν τη χρήση προσεγγιστικών δεδομένων θέσης και προσανατολισμού (π.χ. από αισθητήρες GNSS/INS), ώστε να είναι εφαρμόσιμοι και στην περίπτωση κατά την οποία τέτοιου είδους δεδομένα δεν είναι διαθέσιμα (π.χ. εικόνες από ερασιτεχνικά μη επανδρωμένα αεροσκάφη, παλιά σύνολα δεδομένων). Η γεωαναφορά επιτυγχάνεται είτε μέσω υπολογισμού του μετασχηματισμού κάθε εικόνας στο επίγειο σύστημα αναφοράς είτε μέσω υπολογισμού των παραμέτρων εξωτερικού προσανατολισμού κάθε εικόνας. Επίσης, η γεωμετρία των πλάγιων εναέριων εικόνων ενός δομημένου περιβάλλοντος ευνοεί τον αυτόματο εντοπισμό κατακόρυφων και οριζόντιων γραμμών και αντίστοιχων σημείων φυγής, όπως επίσης και σημείων που κείνται σε επίπεδα. Στο πλαίσιο αυτό, επιμέρους στόχος της διδακτορικής διατριβής είναι η διερεύνηση της συμβολής της αυτόματης εξαγωγής γεωμετρικών στοιχείων της σκηνής που απεικονίζεται σε πλάγιες εναέριες εικόνες στη γεωαναφορά αυτών και στη μέτρηση οριζόντιων και κατακόρυφων αποστάσεων σε μεμονωμένες πλάγιες εναέριες εικόνες. Οι ανωτέρω στόχοι εκπληρώθηκαν με την ανάπτυξη αλγορίθμων υψηλού βαθμού αυτοματοποίησης, οι οποίοι επιτρέπουν τη χρήση τους και σε χρήστες χωρίς εμπειρία ή βασικές γνώσεις φωτογραμμετρίας και

συμβάλλουν στην ένταξη των πλάγιων εναέριων εικόνων στις κλασικές φωτογραμμετρικές διαδικασίες.

Πριν από την παρουσίαση των αλγορίθμων που αναπτύχθηκαν στο πλαίσιο της διδακτορικής διατριβής καθώς και των αποτελεσμάτων που επιτεύχθηκαν μέσω αυτών, καταγράφονται βασικοί ορισμοί και γεωμετρικές σχέσεις για πλάγιες εναέριες εικόνες και γίνεται εκτενής έρευνα στη διεθνή βιβλιογραφία κυρίως σε θέματα γεωαναφοράς και προσανατολισμού συνόλων εικόνων που περιέχουν πλάγιες εναέριες εικόνες. Συγκεκριμένα, περιγράφεται η κατηγοριοποίησή τους σε εικόνες μικρής και μεγάλης κλίσης, καταγράφονται οι κύριες διατάξεις λήψης πλάγιων εναέριων εικόνων και παρουσιάζονται οι βασικοί ορισμοί, σχέσεις και ιδιότητες που σχετίζονται με αυτού του είδους τις εικόνες. Τα ανωτέρω είναι απαραίτητα για την κατανόηση της γεωμετρίας τους και των μεθόδων που αναπτύχθηκαν και περιγράφονται στο πλαίσιο της διατριβής. Όσον αφορά στη βιβλιογραφική ανασκόπηση, αρχικά γίνεται κατηγοριοποίηση της έρευνας ανάλογα με το είδος εφαρμογών στις οποίες χρησιμοποιούνται οι πλάγιες εναέριες εικόνες (π.χ. τρισδιάστατη μοντελοποίηση, απόδοση υφής, εξαγωγή κτηρίων, τρισδιάστατη ανακατασκευή κτηρίων, δημιουργία ψηφιακού μοντέλου εδάφους, αξιολόγηση καταστροφών, ταξινόμηση σκηνής, ανίχνευση προσόψεων κτηρίων). Καταγράφονται με λεπτομέρεια οι εξής προσεγγίσεις προσανατολισμού και γεωαναφοράς πλάγιων εναέριων εικόνων καθώς επίσης και συνόλων πλάγιων και κατακόρυφων εναέριων εικόνων: υπολογισμός εξωτερικού προσανατολισμού μέσω υπαρχόντων λογισμικών· υπολογισμός εξωτερικού προσανατολισμού μέσω ανάπτυξης λύσεων που εφαρμόζουν τη μέθοδο υπολογισμού δομής από κίνηση (structure from motion – SfM) και συνόρθωση δέσμης· άμεση γεωαναφορά με χρήση δεδομένων από αισθητήρες αεροσκαφών· γεωαναφορά ως μονοεικονική διαδικασία· και εναλλακτικές μέθοδοι γεωαναφοράς με χρήση ευθειών από εικόνες και μεθόδων συνταύτισης. Παράλληλα, καταγράφονται αναλυτικά τα πιο σημαντικά αποτελέσματα της έρευνας που έχει διεξαχθεί μέχρι σήμερα για τη γεωαναφορά συνόλων πλάγιων εναέριων εικόνων σχετικά με την εξαγωγή χαρακτηριστικών από εικόνες και την ψηφιακή συνταύτιση, τη χρήση γράφων, τη χρήση των σχετικών θέσεων μεταξύ των φωτομηχανών ενός συστήματος πολλαπλών καμερών, την επιβολή γεωμετρικών περιορισμών κατά τη διαδικασία υπολογισμού εξωτερικού προσανατολισμού, τη στρατηγική απόδοσης βαρών κατά τη συνόρθωση δέσμης, ζητήματα ακρίβειας σε σύνολα κατακόρυφων και πλάγιων εικόνων πολλαπλών προοπτικών, την επίδραση διαφορετικών παραμέτρων στην ακρίβεια μετρήσεων μετά από συνόρθωση δέσμης, τεχνικές ανίχνευσης σημείων φυγής για υπολογισμό εξωτερικού προσανατολισμού, μεθόδους συνταύτισης μεταξύ εικόνων και προϋπαρχόντων τρισδιάστατων μοντέλων ή δισδιάστατων χαρτών και ζητήματα προσεγγιστικής γεωαναφοράς.

Κατόπιν, παρουσιάζονται οι επιμέρους μέθοδοι που αναπτύχθηκαν και τα αποτελέσματά τους σε πραγματικά σύνολα δεδομένων. Συγκεκριμένα, αναπτύχθηκε και παρουσιάζεται αλγόριθμος αυτόματης γεωαναφοράς μεγάλων συνόλων κατακόρυφων και πλάγιων εναέριων εικόνων πολλαπλών προοπτικών που δεν συνοδεύονται από δεδομένα θέσης και προσανατολισμού, άγνωστου εσωτερικού προσανατολισμού και ύψους πτήσης. Ο αλγόριθμος υπολογίζει τις παραμέτρους μετασχηματισμού κάθε εικόνας στο επίγειο σύστημα αναφοράς, υιοθετώντας ένα μοντέλο δισδιάστατου μετασχηματισμού. Βασίζεται στη μέτρηση σημείων γνωστών οριζοντιογραφικών συντεταγμένων σε μια εικόνα σε συνδυασμό με συνταύτιση και υπολογισμό του δισδιάστατου μετασχηματισμού μεταξύ επικαλυπτόμενων εικόνων για τη μεταφορά της πληροφορίας γεωαναφοράς μεταξύ των εικόνων. Επιπρόσθετα, πραγματοποιεί επανασύσταση των εικόνων, εξάγει τα ίχνη τους, κατηγοριοποιεί τις εικόνες σε ομάδες βάσει των ιχνών τους και εντοπίζει τις εικόνες που καλύπτουν μια συγκεκριμένη περιοχή. Στο πλαίσιο της διδακτορικής διατριβής, εξετάζονται διάφοροι δισδιάστατοι μετασχηματισμοί συντεταγμένων που προσεγγίζουν τη γεωμετρική σχέση μεταξύ ομόλογων σημείων σε επικαλυπτόμενες εικόνες και προκύπτουν

συμπεράσματα σχετικά με την καταλληλότητά τους ανάλογα με το είδος της περιοχής που απεικονίζουν οι εικόνες, την κλίμακά τους και τη διαφορά προοπτικής τους. Επίσης, εξετάζεται η επίδραση της επιλογής της εικόνας στην οποία γίνεται η μέτρηση των σημείων γνωστών συντεταγμένων στην ακρίβεια γεωαναφοράς, εντοπίζονται τα χαρακτηριστικά της που ελαχιστοποιούν το σφάλμα γεωαναφοράς και πραγματοποιείται ανάλυση των σφαλμάτων σε σχέση με τον αριθμό μετασχηματισμών που μεσολαβούν μεταξύ μιας εικόνας και του επίγειου συστήματος αναφοράς. Τέλος, αξιολογείται η επίδοση του αλγορίθμου SURF στη συνταύτιση κατακόρυφων και πλάγιων εναέριων εικόνων πολλαπλών προοπτικών.

Τα πειράματα πραγματοποιήθηκαν με χρήση τριών συνόλων εικόνων διαφορετικών χαρακτηριστικών. Το πρώτο μπλοκ εικόνων αποτελείται από 150 κατακόρυφες και πλάγιες εναέριες εικόνες πολλαπλών προοπτικών, που λήφθηκαν από σύστημα τύπου Leica RCD30 Oblique Penta σε διαμόρφωση Μαλτέζικου σταυρού, με χρήση μιας ναδιρικής φωτομηχανής και τεσσάρων φωτομηχανών κεκλιμένων κατά 35˚ σε σχέση με την κατακόρυφο. Οι εικόνες, διαστάσεων 60 MP, αντιστοιχούν σε εστιακή απόσταση 53 mm και σε μέγεθος εδαφοψηφίδας 6 – 13 cm. Καλύπτουν περιοχή περίπου 3 $km^2$ με σχετικά χαμηλό ανάγλυφο στο μεγαλύτερο μέρος της, ενώ υψομετρική διαφορά περίπου 100 m συγκεντρώνεται σε μικρό τμήμα της. Απεικονίζουν κυρίως πολυκατοικίες και βιομηχανικά κτήρια, γέφυρες και ποτάμι. Το δεύτερο σύνολο εικόνων απαρτίζεται από 81 κατακόρυφες και πλάγιες εναέριες εικόνες πολλαπλών προοπτικών που λήφθηκαν σε διαμόρφωση Μαλτέζικου σταυρού από σύστημα τύπου PentaCam IGI. Οι εικόνες, διαστάσεων 50 MP, αντιστοιχούν σε εστιακή απόσταση 80 mm και σε μέγεθος εδαφοψηφίδας 8 – 12 cm. Καλύπτουν περιοχή περίπου 3 $km^2$ που περιλαμβάνει αρκετές μονοκατοικίες και ιστορικά κτήρια καθώς και γεωργικές περιοχές. Η υψομετρική διαφορά εντός της εικονιζόμενης περιοχής είναι περίπου 40 m. Το τρίτο σύνολο εικόνων αποτελείται από ένα ακανόνιστο μπλοκ 42 πλάγιων εναέριων εικόνων που λήφθηκαν από φωτομηχανή τύπου Canon EOS-1Ds Mark III με τηλεφακό εστιακής απόστασης 1200 mm, με γωνία κλίσης περίπου 60˚. Οι εικόνες, διαστάσεων 21 MP, αντιστοιχούν σε μέγεθος εδαφοψηφίδας 5 – 25 cm. Καλύπτουν έκταση περίπου 4 $km^2$, η οποία αποτελείται κυρίως από πολυκατοικίες, λιμένα, θάλασσα και ακαλλιέργητη γη. Η υψομετρική διαφορά εντός της εικονιζόμενης περιοχής είναι περίπου 40 m. Τα αποτελέσματα του προτεινόμενου αλγορίθμου γεωαναφοράς εξαρτώνται κυρίως από τον υπολογισμό του κατάλληλου δισδιάστατου μετασχηματισμού μεταξύ κάθε εικόνας και του επίγειου συστήματος αναφοράς. Η εικόνα στην οποία πρέπει να μετρηθεί ένας ελάχιστος αριθμός σημείων γνωστών οριζοντιογραφικών συντεταγμένων – ο οποίος εξαρτάται από τον χρησιμοποιούμενο μετασχηματισμό – θα πρέπει να έχει επικάλυψη με μεγάλο αριθμό εικόνων, σχετικά κεντρική θέση στο μπλοκ των εικόνων και ήπιο ανάγλυφο. Η εν λόγω μέθοδος γεωαναφοράς μπορεί να εφαρμοστεί σε σύνολα πλάγιων και κατακόρυφων εικόνων που χαρακτηρίζονται από μικρές υψομετρικές διαφορές. Η χρήση μη συνεπίπεδων σημείων για τον υπολογισμό δισδιάστατου μετασχηματισμού μεταξύ εικόνων είναι η κύρια πηγή σφαλμάτων. Η ακρίβεια γεωαναφοράς μειώνεται με την αύξηση του αριθμού μετασχηματισμών που μεσολαβούν μεταξύ μιας εικόνας και του επίγειου συστήματος αναφοράς. Τέλος, η εφαρμογή του αλγορίθμου SURF αποδείχτηκε ότι έχει ικανοποιητικά αποτελέσματα στη συνταύτιση εικόνων πολλαπλών προοπτικών για σκοπούς γεωαναφοράς τους.

Επιπλέον, αναπτύχθηκε και παρουσιάζεται αλγόριθμος αυτόματου εντοπισμού σημείων που κείνται σε συγκεκριμένο επίπεδο ενδιαφέροντος σε σύνολα πλάγιων και κατακόρυφων εναέριων εικόνων. Ο αλγόριθμος βασίζεται σε μια τεχνική συνταύτισης προτύπων (template matching) σε συνδυασμό με τον αλγόριθμο RANSAC για την εκτίμηση των παραμέτρων του προβολικού μετασχηματισμού (ομογραφίας) μεταξύ των απεικονίσεων του επιπέδου ενδιαφέροντος σε ζεύγη εικόνων και τη μεταφορά της πληροφορίας του επιπέδου σε όλες τις εικόνες. Αναπτύχθηκαν και

περιγράφονται οι εξής μέθοδοι που εφαρμόζουν τον εν λόγω αλγόριθμο σε εικόνες που περιλαμβάνουν επίπεδες επιφάνειες: μέθοδος πολυεικονικής γεωαναφοράς μέσω προβολικού μετασχηματισμού· μέθοδος υπολογισμού προσεγγιστικών στοιχείων εξωτερικού προσανατολισμού σε σύνολα εικόνων· και μέθοδος μεταφοράς συνεπίπεδων εικονοσημείων γνωστών συντεταγμένων σε επικαλυπτόμενες εικόνες βάσει των εικονοσυντεταγμένων τους σε μία εικόνα. Το προτεινόμενο σχήμα γεωαναφοράς που χρησιμοποιεί συνεπίπεδα σημεία ενδιαφέροντος προϋποθέτει τη μέτρηση συνεπίπεδων φωτοσταθερών σε μια εικόνα και μπορεί να εφαρμοστεί σε σύνολα εικόνων ανεξάρτητα από τη διαφορά υψομέτρου της σκηνής, αρκεί οι εικόνες να απεικονίζουν επίπεδο ίδιου υψομέτρου με το υψόμετρο των συνεπίπεδων φωτοσταθερών.

Τα πειράματα πραγματοποιήθηκαν με δύο σύνολα πλάγιων εικόνων από μη επανδρωμένο αεροσκάφος, που περιλαμβάνουν 23 και 35 εικόνες αντίστοιχα που εικονίζουν ένα κτήριο. Οι εικόνες, διαστάσεων 24 MP, λήφθηκαν από φωτομηχανή τύπου Sony Nex-7 και αντιστοιχούν σε εστιακή απόσταση 16 mm και σε μέγεθος εδαφοψηφίδας 1 – 3 cm. Τα πειράματα έδειξαν ότι η προτεινόμενη μέθοδος εντοπισμού συνεπίπεδων σημείων εντοπίζει σωστά σημεία ενδιαφέροντος που κείνται σε συγκεκριμένο επίπεδο, οριζόμενο από τον χρήστη. Επίσης, τα προσεγγιστικά στοιχεία εξωτερικού προσανατολισμού εικόνων που έχουν γεωαναφερθεί μέσω συνεπίπεδων σημείων ενδιαφέροντος μπορούν να χρησιμοποιηθούν ως αρχικές τιμές εξωτερικού προσανατολισμού σε διαδικασία συνόρθωσης δέσμης.

Επιπλέον, αναπτύχθηκε και παρουσιάζεται αλγόριθμος αυτόματης εξαγωγής του ναδιρικού σημείου μεμονωμένων πλάγιων εναέριων εικόνων. Ο αλγόριθμος στηρίζεται στην ανίχνευση ακμών και ευθύγραμμων τμημάτων, στον εντοπισμό πολλαπλών σημείων φυγής, στον εντοπισμό της γραμμής του αληθούς ορίζοντα, σημείων φυγής οριζόντιων διευθύνσεων και αντίστοιχων οριζόντιων ευθύγραμμων τμημάτων, στον υπολογισμό αρχικής εκτίμησης του ναδιρικού σημείου και στην τελική εκτίμηση του ναδιρικού σημείου και κατακόρυφων ευθύγραμμων τμημάτων. Αναπτύχθηκαν και περιγράφονται δύο μέθοδοι που εφαρμόζουν τον εν λόγω αλγόριθμο: μέθοδος υπολογισμού προσεγγιστικών στοιχείων εξωτερικού προσανατολισμού συνόλων πλάγιων εναέριων εικόνων και μέθοδος υπολογισμού οριζόντιων και κατακόρυφων αποστάσεων σε μεμονωμένες πλάγιες εικόνες.

Τα πειράματα πραγματοποιήθηκαν ομοίως με τα ανωτέρω αναφερόμενα σύνολα 23 και 35 πλάγιων εικόνων που λήφθηκαν με φωτομηχανή τύπου Sony Nex-7 από μη επανδρωμένο αεροσκάφος. Τα πειράματα έδειξαν ότι ο προτεινόμενος αλγόριθμος εντοπίζει σωστά οριζόντια και κατακόρυφα ευθύγραμμα τμήματα. Το ναδιρικό σημείο που υπολογίζεται μέσω του αλγορίθμου μπορεί να χρησιμοποιηθεί για την εξαγωγή οριζόντιων και κατακόρυφων αποστάσεων, σε περίπτωση που δεν απαιτούνται μετρήσεις ακριβείας. Οι προσεγγιστικές παράμετροι εξωτερικού προσανατολισμού που υπολογίζονται με βάση το αυτόματα εντοπισμένο ναδιρικό σημείο των εικόνων αντιστοιχούν σε μειωμένη ακρίβεια σε σχέση με την ακρίβεια των προσεγγιστικών παραμέτρων εξωτερικού προσανατολισμού που υπολογίζονται με χρήση των αυτόματα εντοπισμένων συνεπίπεδων σημείων ενδιαφέροντος των εικόνων. Και οι δύο μέθοδοι υπολογισμού αρχικών τιμών εξωτερικού προσανατολισμού μέσω γεωμετρικών στοιχείων των εικόνων (συνεπίπεδων σημείων, ναδιρικού σημείου) μπορούν να εφαρμοστούν σε σύνολα γεωαναφερμένων εικόνων στα οποία ο αριθμός των μετασχηματισμών ομογραφίας που μεσολαβούν μεταξύ κάθε εικόνας και του επίγειου συστήματος αναφοράς είναι μικρός (π.χ. μέχρι 20 μετασχηματισμοί ομογραφίας). Σε διαφορετική περίπτωση, πρέπει να μετρηθούν τέσσερα συνεπίπεδα φωτοσταθερά σε μια επιπρόσθετη εικόνα, έτσι ώστε ο μέγιστος αριθμός μετασχηματισμών μεταξύ κάθε εικόνας και του επίγειου συστήματος αναφοράς να διατηρείται μικρός.

Ακόμη, αναπτύχθηκαν και παρουσιάζονται αλγόριθμοι που υπολογίζουν τα στοιχεία του εξωτερικού προσανατολισμού συνόλων εικόνων μέσω της βαθμωτής μεθόδου SfM (incremental SfM) και μέσω της καθολικής μεθόδου SfM (global SfM). Ο προτεινόμενος καθολικός αλγόριθμος SfM χρησιμοποιεί διαφορετικές τεχνικές για τον υπολογισμό αρχικών τιμών εξωτερικού προσανατολισμού, συμπεριλαμβανομένων των μεθόδων που αναπτύχθηκαν στο πλαίσιο της διατριβής, οι οποίες κάνουν χρήση στοιχείων της γεωμετρίας των σκηνών που απεικονίζονται στις εικόνες (συνεπίπεδων σημείων καθώς και οριζόντιων και κατακόρυφων ευθύγραμμων τμημάτων). Οι προτεινόμενοι αλγόριθμοι SfM εφαρμόζουν εύρωστες μεθόδους επαναληπτικής συνόρθωσης δέσμης για την απομάκρυνση των λανθασμένων ομολογιών που παραμένουν μετά τη διαδικασία της ψηφιακής συνταύτισης. Ιδιαίτερα για σύνολα πλάγιων εναέριων εικόνων διαφορετικών προοπτικών, στα οποία το φαινόμενο των λανθασμένων ομολογιών είναι εντονότερο σε σχέση με εκείνες σε παραδοσιακά σύνολα κατακόρυφων εικόνων, η απαλοιφή τους είναι ιδιαίτερα σημαντική. Επίσης, αναπτύχθηκε και περιγράφεται μια στρατηγική απόδοσης βαρών στις παρατηρήσεις εικονοσημείων κατά τη διαδικασία συνόρθωσης δέσμης, η οποία προορίζεται για πλάγιες εναέριες εικόνες μεγάλου βαθμού επικάλυψης. Τέλος, διερευνάται η επίδραση διαφόρων παραμέτρων στην ακρίβεια των στοιχείων εξωτερικού προσανατολισμού, όπως: της μεθόδου SfM (βαθμωτή ή καθολική) και της μεθόδου υπολογισμού αρχικών τιμών εξωτερικού προσανατολισμού, για την περίπτωση της καθολικής μεθόδου· της απομάκρυνσης όλων των λανθασμένων ομολογιών μέσω της προτεινόμενης μεθόδου επαναληπτικής συνόρθωσης δέσμης· του αλγορίθμου εξαγωγής σημείων ενδιαφέροντος (SURF και ASIFT)· της τεχνικής απόδοσης βαρών κατά τη συνόρθωση δέσμης· και του αριθμού των χρησιμοποιούμενων σημείων γνωστών συντεταγμένων.

Τα πειράματα πραγματοποιήθηκαν σε οκτώ σύνολα εικόνων διαφορετικών χαρακτηριστικών. Τα πέντε σύνολα εικόνων που χρησιμοποιήθηκαν προέρχονται από φωτομηχανή τύπου Sony Nex-7 από μη επανδρωμένο αεροσκάφος και περιλαμβάνουν 23, 35, 60, 80 και 100 πλάγιες εικόνες, αντίστοιχα. Τα υπόλοιπα τρία σύνολα εικόνων λήφθηκαν από σύστημα τύπου Leica RCD30 Oblique Penta από επανδρωμένο αεροσκάφος. Το πρώτο σύνολο εικόνων περιλαμβάνει 10 πλάγιες εναέριες εικόνες της ίδιας προοπτικής που λήφθηκαν από την ίδια φωτομηχανή του συστήματος Leica RCD30 Oblique Penta. Το δεύτερο σύνολο εικόνων περιλαμβάνει 40 πλάγιες εναέριες εικόνες διαφορετικών προοπτικών, που λήφθηκαν από τις τέσσερις φωτομηχανές του συστήματος Leica RCD30 Oblique Penta. Τέλος, το τρίτο σύνολο εικόνων περιλαμβάνει 50 κατακόρυφες και πλάγιες εναέριες εικόνες που λήφθηκαν σε διαμόρφωση Μαλτέζικου σταυρού από τις πέντε φωτομηχανές του εν λόγω συστήματος. Η προτεινόμενη βαθμωτή μέθοδος SfM που χρησιμοποιεί επαναληπτική συνόρθωση δέσμης για την απομάκρυνση των λανθασμένων ομολογιών και η αντίστοιχη προτεινόμενη καθολική μέθοδος SfM που χρησιμοποιεί ως αρχικές τιμές εξωτερικού προσανατολισμού αυτές που προκύπτουν από μια βαθμωτή μέθοδο SfM προτείνονται στο πλαίσιο της διατριβής ως εκείνες με τα καλύτερα αποτελέσματα. Η απομάκρυνση όλων των λανθασμένων ομολογιών μέσω των προτεινόμενων μεθόδων επαναληπτικής συνόρθωσης δέσμης οδηγεί σε αποτελέσματα εξωτερικού προσανατολισμού υψηλότερης ακρίβειας σε σχέση με εκείνα που θα επιτυγχάνονταν χωρίς την απομάκρυνσή τους. Ο αλγόριθμος SURF αποδείχθηκε ότι μπορεί να χρησιμοποιείται για την εξαγωγή σημείων ενδιαφέροντος σε σύνολα πλάγιων εναέριων εικόνων με ικανοποιητικά αποτελέσματα. Η στρατηγική απόδοσης βαρών κατά τη συνόρθωση δέσμης πλάγιων εναέριων εικόνων βελτιώνει τα αποτελέσματα εξωτερικού προσανατολισμού εικόνων που αντιστοιχούν σε μεγάλο ποσοστό επικάλυψης και στις οποίες το ίδιο σημείο ενδιαφέροντος εντοπίζεται σε μεγάλο αριθμό εικόνων (τουλάχιστον πέντε). Επίσης, τα αποτελέσματα εξωτερικού προσανατολισμού βελτιώνονται με τη μέτρηση περισσότερων φωτοσταθερών με καλύτερη κατανομή. Ακόμη, ο συνδυασμένος αεροτριγωνισμός κατακόρυφων και πλάγιων εικόνων

διαφορετικών προοπτικών που έχουν ληφθεί από το ίδιο σύστημα πολλαπλών φωτομηχανών έχει καλύτερα αποτελέσματα σε σχέση με τον αεροτριγωνισμό μόνο των πλάγιων εικόνων. Τα αποτελέσματα εξωτερικού προσανατολισμού που προκύπτουν μέσω των καλύτερων συνδυασμών των ανωτέρω αναφερόμενων μεθόδων και παραμέτρων σε σύνολα πλάγιων εναέριων εικόνων και σε συνδυασμένα σύνολα πλάγιων και κατακόρυφων εικόνων είναι βελτιωμένης ακρίβειας σε σχέση με τα αντίστοιχα αποτελέσματα εμπορικών λογισμικών. Ταυτόχρονα, οι προτεινόμενες μέθοδοι αντιστοιχούν σε μεγαλύτερο βαθμό αυτοματοποίησης σε σχέση με υφιστάμενα λογισμικά, στην περίπτωση ύπαρξης συνεπίπεδων σημείων γνωστών συντεταγμένων.

# Table of Contents

# List of Figures

# List of Tables

# List of Abbreviations

| | |
|---|---|
| 2D | Two-dimensional |
| 3D | Three-dimensional |
| ASIFT | Affine-SIFT |
| BA | Bundle adjustment |
| DEM | Digital elevation model |
| DSM | Digital surface model |
| DTM | Digital terrain model |
| EO | Exterior orientation |
| EuroSDR | European Spatial Data Research |
| GCP | Ground control point |
| GIS | Geographic information system |
| GNSS | Global navigation satellite system |
| GSD | Ground sample distance |
| GPS | Global positioning system |
| IMU | Inertial measurement unit |
| INS | Inertial navigation system |
| ISPRS | International Society for Photogrammetry and Remote Sensing |
| KML | Keyhole markup language |
| LiDAR | Light detection and ranging |
| LoD | Level of detail |
| MVS | Multi-view stereo |
| PnP | Perspective-n-Point |
| RANSAC | Random sample consensus |
| SAR | Synthetic aperture radar |
| SfM | Structure from motion |
| SIFT | Scale-invariant feature transform |
| SURF | Speeded-up robust features |
| UAV | Unmanned aerial vehicle |
| UTM | Universal Transverse Mercator |
| WGS 84 | World Geodetic System 1984 |

# Chapter 1
# Introduction

## 1.1 Background

Oblique aerial images have been a significant part of the history of aerial photography since its earliest days. Although vertical airborne images played the leading role in photogrammetric applications for more than a century, in recent years oblique images gain ground as a result of the progress made in photogrammetric and computer vision algorithms, the increasing market availability of such images and certain fundamental advantages they provide compared to vertical images. During the last twenty years, several companies have entered the geospatial market by offering complete coverage of many cities with this kind of images and by providing software for their metric exploitation. In addition, multi-camera systems have become a well-established technology for both metric and visualization applications and give users the opportunity to obtain large datasets of multi-perspective oblique and vertical aerial images. As far as the target group of users of oblique aerial images is concerned, it includes not only photogrammetrists but also engineers performing a work study, as well as other scientists, decision makers and professionals, who need aerial imagery in order to accomplish a project. Furthermore, the wide spread of affordable unmanned aerial vehicles (UAVs) has made aerial images – vertical and mainly oblique ones – accessible to a broader public, even for artistic and entertainment purposes. In this context, taking into account the extensive use of oblique images in recent years, two benchmarks have been launched by the European Spatial Data Research organization (EuroSDR) and the International Society for Photogrammetry and Remote Sensing (ISPRS) in order to evaluate the progress of the available tools in processing oblique aerial imagery (Cavegn et al., 2014; Nex et al., 2015).

The main reason why oblique aerial images are being used more and more these days is the fact that they reveal vertical structures (mainly building facades) which are either not or just marginally depicted in traditional vertical aerial images, in combination with the representation of landscapes in a way similar to the human perception of scenes from ground view. Thus, they can be used for visualization purposes by non-specialists in the interpretation of aerial images. Moreover, they are a valuable tool for security operations, by providing visualization of critical locations, for

identification of surrounding areas and planning of access and exit routes, as well as for urban and infrastructural planning. As far as photogrammetric applications are concerned, they can be used for a wide range of applications, like 3D modelling, 3D building reconstruction, texture mapping, building extraction, damage assessment, facade detection and mapping applications. For the aforementioned cases, oblique aerial images may be used either independently, as the basic source of information, or in combination with vertical views and/or terrestrial images.

Image georeferencing is the process of assigning spatial information to an image to define its location with respect to a ground coordinate system. It is usually accomplished in terms of exterior orientation estimation and optionally image orthorectification, while in some cases a simple georeferencing by means of mapping of the center point of the images and/or their footprint and/or image rectification via a 2D transformation suffices. Georeferencing is generally the first step of most photogrammetric applications, being in most cases a prerequisite for their metric exploitation. Thus, much research has been conducted towards this direction for oblique aerial images as well. This kind of images poses several challenges in automated photogrammetric processes due to complex image geometry, as a result of occlusions and significant scale variations inside each oblique image, as well as due to the strongly varying appearance at images that depict the same scene from different viewing angles.

Although satisfying results are achieved through existing software solutions, especially via those solving the structure from motion (SfM) problem, being originally targeted to terrestrial image sequences, some significant research issues regarding georeferencing of oblique aerial images still remain open. Such issues mainly concern (i) the automatic georeferencing of oblique aerial images without availability of GNSS/INS (global navigation satellite system/inertial navigation system) information; (ii) the investigation of the image matching results achievable from oblique aerial images using a feature detector different from SIFT, which is mainly used by the majority of researchers; (iii) the establishment of robust image matching and outlier rejection techniques for oblique aerial images; (iv) the automatic extraction of elements describing the geometry of the scene depicted in oblique aerial images (e.g., vertical and horizontal lines, coplanar points) for exterior orientation estimation; (v) the establishment of a weighting strategy for image tie points of different scale for bundle adjustment, without requiring initial orientation values from onboard sensors; and (vi) the investigation of the exterior orientation results that can be achieved in non-ideal bundle adjustment scenarios characterized by the lack of well-distributed ground control points (GCPs) and minimum image measurements without any data from onboard sensors.

## 1.2 Research Objectives

While in the past the process of extracting metric information from aerial images was based almost exclusively on vertical views, oblique aerial images have come to the foreground, being not only a complementary dataset to traditional vertical images but also often the basic source of information for various kinds of applications. Indeed, significant research has been carried out in this direction. This research focuses on oblique aerial images mainly captured by metric cameras or multi-camera systems, provided that georeferencing data from positioning and inertial sensors (GNSS/INS) as well as multi-image measurements of GCPs are available. However, the research carried out on the metric exploitation of oblique aerial images that are not accompanied by rough positioning and orientation data and correspond to a minimum number of GCPs without a good distribution is very limited. At the same time, image georeferencing is a very important process, being a prerequisite for the majority of photogrammetric applications.

In this context and in the light of the aforementioned research background, the overall aim of this doctoral dissertation is the development of algorithms of photogrammetry and computer vision for the metric exploitation of oblique aerial images, with emphasis on georeferencing procedures for both oblique images and combined datasets of vertical and multi-perspective oblique images, based on a minimum number of measurements and without requiring approximate positioning and orientation data (e.g., from GNSS/INS sensors), so that they are also applicable in cases where such information is not available (e.g., amateur UAV images, old datasets). Georeferencing may be achieved either by calculating the transformation of each image to the ground reference system or by calculating the exterior orientation parameters of each image. Also, the geometry of oblique aerial images of man-made environments favors the automatic detection of vertical and horizontal lines and corresponding vanishing points, as well as coplanar points. In this context and in line with the overall aim, this dissertation aims to investigate the contribution of the automatic extraction of geometric elements of the scene depicted in oblique aerial images for their metric exploitation, focusing on georeferencing procedures.

Hence, the following research objectives that stem from the overall aim of this dissertation have been defined:

- the establishment of an automatic image co-registration scheme based on 2D transformations for rough georeferencing of datasets of multi-perspective oblique and vertical aerial imagery of unknown interior orientation and flying height;
- the establishment of an automatic workflow for detecting coplanar feature points that lie on a user-defined plane among oblique and vertical large-scale aerial images depicting a piecewise planar scene;
- the establishment of an automatic workflow for detecting horizontal and vertical line segments, horizontal vanishing points, the horizon line and the nadir point of an oblique aerial image of a man-made environment;
- the investigation of the contribution of the automatic extraction of geometric elements of the scene depicted in oblique aerial images (horizontal and vertical lines and vanishing points as well as coplanar points) for calculation of approximate exterior orientation parameters to be refined in a bundle adjustment procedure as well as measurement of horizontal and vertical distances in single oblique aerial images;
- the establishment of incremental and global SfM workflows that adopt robust methods for outlier removal in challenging datasets containing oblique and vertical aerial images for exterior orientation estimation; and
- the investigation of the impact that each of the following has on the exterior orientation results of datasets containing oblique aerial images: (i) the SfM workflow (global or incremental) and the method used for obtaining initial exterior orientation parameters, in the case of a global workflow; (ii) the removal of erroneous matches that are not detected during the image matching process; (iii) the feature detection algorithm; (iv) a weighting strategy for automatic tie point measurements during bundle adjustment; and (v) the number of GCPs.

## 1.3 Contribution and Originality

The overall contribution of this dissertation lies in the establishment of automated methods that can be applied for the metric exploitation of datasets containing oblique aerial images without availability of GNSS/INS data or multi-image measurement of well-distributed ground control data,

mainly for georeferencing purposes. The main contributions, which are analyzed in the next paragraphs, are the following:

- the establishment of a RANSAC-based method for detecting coplanar feature points lying on a user-defined plane and its implementation (i) within a developed 2D georeferencing framework, so that the latter is applicable not only in small-scale aerial image datasets but also in large-scale image datasets of a piecewise planar scene, without any constraint on its elevation difference, and (ii) for transferring coplanar GCPs based on their single-image measurements;
- the establishment of a RANSAC-based method for detecting the nadir point of an oblique aerial image of a man-made scene, along with its horizon line as well as horizontal and vertical line segments;
- the establishment of a robust iterative bundle adjustment framework for eliminating all erroneous tie points that remain after tests and constraints imposed during image matching procedures, within incremental and global SfM workflows; and
- the investigation of different methods for estimating approximate camera exterior orientation parameters and the investigation of different aspects that affect the accuracy of the exterior orientation results estimated by SfM methods, including the SfM workflow, the removal of erroneous correspondences not detected during image matching, the feature extraction algorithm and a weighting strategy for feature points during bundle adjustment.

The developed 2D georeferencing framework for datasets of oblique and vertical aerial images differs from the other state-of-the-art georeferencing methods discussed in literature in that it is applied to large datasets of images without availability of GNSS/INS data from onboard sensors, nor camera characteristics or interior orientation parameters, using a minimum number of points of known horizontal coordinates measured in a single image. The related work on existing rough georeferencing approaches applied in oblique aerial imagery is limited to a few research approaches that rely on metadata from onboard sensors and on information for the camera, contrary to the developed approach. What is more, these approaches are applied in single images, without making use of any image registration framework, like the one proposed in this dissertation. Whereas more research has been conducted towards 2D georeferencing of vertical aerial images, only a few methods are applied in datasets of images, rather than on single images, and perform a global 2D transformation from every image to the reference coordinate system. None of these methods is simultaneously applied in large datasets of images with GCP measurements in a single image without GNSS/INS data. Also, the whole developed rough georeferencing software suite consists of a combination of useful tools for image rectification, extraction of footprints, subdivision of images into groups based on the area they cover and determination of those that cover a specific area; this combination of tools automates several applications for georeferencing, archiving and searching among large datasets and is not found in existing software packages.

Several RANSAC-based approaches have been proposed in literature that either search for a dominant homography that verifies the relationship between the largest number of corresponding points in two views or group the detected feature points into coplanar sets that share a common homography for the detection of multiple planes. However, the coplanar point detection algorithm introduced in this dissertation goes one step further by choosing, among the coplanar point sets, the one that belongs to a specific plane of interest. In this context, its main contribution compared to the existing RANSAC-based approaches is the combination of a template matching method with the RANSAC algorithm in an iterative framework for the detection of feature points that lie on a specific user-defined plane instead of the dominant one or each dominant one that can be iteratively

detected by RANSAC. Moreover, the implementation of the coplanar point detection method within the developed 2D georeferencing framework for constraining the detected correspondences to lie in the plane where the GCPs lie, provided that the latter are coplanar, gives satisfactory georeferencing results in large-scale image datasets of a piecewise planar scene. In this way, the developed 2D georeferencing framework can be applied in image datasets regardless of the elevation difference of the scene, provided that all images depict a planar area with the same elevation as the elevation of the coplanar GCPs. Such a 2D georeferencing framework, that may be applied in images of a piecewise planar scene without constraint on the elevation difference of the scene, is novel and has neither been implemented for vertical aerial images nor for oblique ones. Moreover, the estimation of approximate exterior orientation parameters of georeferenced image datasets of a piecewise planar scene using the developed coplanar point detection scheme is an alternative method for computing rough exterior orientation parameters to be used as initial values in a bundle adjustment procedure. In addition, the transfer of coplanar GCPs among images based on their single-image measurements is another field of application of the developed coplanar point detection method.

The developed method that detects horizontal and vertical line segments, the true horizon line, horizontal vanishing points and the vertical vanishing point of an oblique aerial image is more robust for being applied in oblique aerial images than the ones that assume a Manhattan world. Especially, the assumption of the availability of three orthogonal vanishing points in the scene may be non-realistic for oblique aerial images of man-made scenes that depict several structures with horizontal edges of quite differing orientations, not necessarily orthogonal to each other. The developed method only assumes the availability of a vertical vanishing direction and several horizontal vanishing directions, which is a realistic assumption in oblique imagery of urban areas. Also, it deals with the challenging case of detecting the nadir point of an oblique aerial image depicting vertical line segments that are either small compared to the horizontal segments and/or do not correspond to a significant number, compared to the rest of the visible line segments in the image, without any a priori knowledge of its approximate position. This is the case for most oblique aerial images (mainly low oblique ones), as the imaged horizontal structures usually stand out more than the vertical ones. In this way, the developed method is targeted to the geometry of oblique aerial images of man-made environments. Whereas methods for detecting the vertical vanishing point have also been applied in oblique aerial imagery by a few researchers, they do not deal with the aforementioned challenging cases; they either make assumptions on the imaged scene, or they apply a standard vanishing point detection scheme without exploiting further the oblique image geometry. Moreover, its contribution compared to the other state-of-the-art RANSAC-based methods for vanishing point detection lies in the introduction of variations of the RANSAC algorithm, according to which the random choice of initial data is controlled through various criteria for estimation of the appropriate model; these constraints enable the developed variations of RANSAC to be applied even in datasets with a very high rate of outliers (e.g., more than 80%), in which the application of a classical RANSAC-based approach may fail. Also, the combination of the developed vanishing point detection technique with mathematical relationships for the determination of vertical and horizontal distances, provides a complete automated framework for measuring distances from a single oblique image without any knowledge for the camera exterior orientation parameters, nor GCP measurements. In addition, the estimation of approximate exterior orientation parameters of georeferenced image datasets of a piecewise planar scene using the developed vanishing point detection scheme is an alternative method for computing rough exterior orientation parameters as initial values in a bundle adjustment process.

One of the challenges of SfM algorithms is obtaining correct feature point correspondences among multiple views. The developed SfM algorithms face this challenge by eliminating all erroneous tie points through the combination of multiple checks and geometric constraints imposed during the image matching procedure and a robust iterative bundle adjustment framework, thus leading to improved accuracy of the exterior orientation results in comparison with well-established SfM techniques. Also, the developed scale-based weighting strategy for feature point observations in bundle adjustment of highly overlapping oblique images differs from the one that has already been presented in literature for oblique images both in its weighting model and in the fact that it does not require information from onboard sensors. In addition, the assessment of the suitability of a scale-based weighting strategy using different configurations of oblique images, captured by manned and unmanned aerial platforms, including both solely oblique ones acquired by a single camera and by a multi-camera system as well as oblique and vertical ones obtained by a multi-camera system, corresponding to different overlap scenarios, is only presented in this dissertation. Furthermore, the fact that an extensive analysis on different aspects that affect the accuracy of the exterior orientation results is made, is another contribution. This analysis includes the investigation of the impact of the SfM workflow, the removal of erroneous correspondences not detected during the image matching process, the feature detection algorithm, the weighting strategy for image feature points during bundle adjustment, and the number of GCPs. Finally, the investigation of the accuracy that can be achieved in the exterior orientation parameters of datasets that include oblique aerial images using a minimum number of points of known horizontal ground coordinates of a bad distribution, which are measured in a single image, is another point of contribution of this dissertation.

## 1.4 Structure of the Dissertation

*Chapter 1* initially presents the current state in the usage of oblique aerial images and briefly describes the research background along with some open research issues. Then, the overall aim along with the research objectives of this doctoral dissertation are outlined and its contribution and originality are discussed. Finally, an overview of the content and structure of this dissertation is presented.

In *Chapter 2*, basic definitions and geometric relationships for oblique aerial images are presented. Specifically, the well-established classification of oblique aerial images into low and high oblique ones is discussed, the camera configurations for capturing datasets containing oblique aerial images are presented and basic terms, geometric relationships and properties associated with oblique aerial images are outlined. These ones are necessary for understanding the underlying geometry of oblique aerial images and the methods introduced in this dissertation. Furthermore, mathematical relationships for computing vertical and horizontal distances from a single oblique aerial image are presented and a related error analysis is made.

*Chapter 3* provides a state-of-the-art on recent research concerning georeferencing of datasets comprising oblique aerial images and outlines the most significant research outcomes. It starts with a presentation of the brief history of oblique aerial imagery followed by a discussion on current research trends and applications using such images. A detailed description of georeferencing approaches is then made for aerial images acquired either by multi-camera systems or by single airborne cameras. Methodologies accomplished using either existing software or self-developed research solutions are presented, regardless of whether the researchers focused entirely on the georeferencing part or presented results from other photogrammetric processes as well. Moreover,

the significant points, results and conclusions derived from the literature review are highlighted and categorized, depending on the topic they cover and their contribution in the georeferencing process.

The next four chapters present the methods introduced in this dissertation along with the results achieved through testing by developed software. Firstly, in *Chapter 4*, a methodology for automatic rough absolute georeferencing of large datasets of multi-perspective oblique and vertical aerial images with unknown interior orientation and flying height, without the need for GNSS/INS information or ground elevation data is presented. The developed approach is based on the measurement of a minimum number of points of known horizontal ground coordinates in only one image and the calculation of the polynomial or projective 2D transformation from each image to the ground reference system, based on image matching and the transfer of georeferencing information among images. Except for providing rough georeferencing information, this methodology automates the process of selecting, among a large number of oblique and vertical aerial images, the ones that cover a specific region. The results of the developed methodology, which was tested using two datasets of multi-perspective oblique and vertical aerial images and one dataset of oblique aerial images, are presented in terms of accuracy and the performance of the SURF feature detector in multi-view oblique and vertical aerial images is evaluated.

In *Chapter 5*, an automatic method for detecting coplanar feature points that lie on a user-defined plane among oblique and vertical large-scale aerial images depicting a piecewise planar scene is introduced, and two georeferencing frameworks that implement the developed coplanar point detection scheme are discussed, i.e., (i) the georeferencing of image datasets of a piecewise planar scene via computation of their 2D homography transformation parameters to the reference coordinate system; and (ii) the estimation of approximate exterior orientation parameters of georeferenced image datasets of a piecewise planar scene. The developed coplanar point detection technique is based on a template matching method in combination with a RANSAC scheme to iteratively estimate the homography transformation between the plane of interest as depicted in the images of each pair and works incrementally by increasing the number of the detected feature points in each image and transferring the plane information among all images. The results of experiments, conducted using two datasets of UAV oblique images, are presented in terms of automatically detecting points that lie on a user-defined plane, 2D georeferencing and approximate exterior orientation estimation.

In *Chapter 6*, a method for automatic estimation of horizontal and vertical line segments, horizontal vanishing points, the horizon line and the nadir point of an oblique aerial image of a man-made environment is introduced and two applications based on the knowledge of these elements are discussed, i.e., (i) the measurement of vertical and horizontal distances from a single oblique aerial image; and (ii) the estimation of approximate exterior orientation parameters of oblique aerial images of georeferenced image datasets, that is, datasets of images with known transformation parameters from the reference coordinate system. The developed vanishing point detection technique uses edge detection and line extraction algorithms, combined with robust model fitting through RANSAC-based methods that take into account the underlying geometry of oblique imagery, as well as least-squares techniques. It works on image space for extracting multiple horizontal vanishing points, along with the true horizon line, with the ultimate goal of detecting the vanishing point of the vertical direction. The results of experiments, conducted using two datasets of UAV oblique images, are presented in terms of automatically detecting vertical and horizontal line segments, determining vertical and horizontal distances and estimating approximate exterior orientation parameters.

In *Chapter 7*, photogrammetry-based incremental and global SfM workflows that adopt robust methods for outlier removal in challenging datasets of oblique and vertical images are presented. The developed SfM techniques require the measurement of four coplanar GCPs in one image for georeferencing the exterior orientation results and the derived point cloud. Also, a scale-based weighting strategy for feature point observations in bundle adjustment is introduced, targeted to highly overlapping oblique images for improving the accuracy of the exterior orientation parameters. The results of experiments, conducted using five datasets of UAV oblique images and three datasets of aerial images acquired by a multi-view camera system, are presented. The experimental results investigate how the automatic extraction of elements describing the geometry of the scene depicted in oblique aerial images of man-made environments can contribute in obtaining initial values for their exterior orientation parameters required by a bundle adjustment process. Furthermore, this chapter investigates the impact that each of the following has on the exterior orientation results: (i) the SfM workflow (global or incremental SfM) and the method used for obtaining initial exterior orientation parameters; (ii) the removal of erroneous matches that are not detected during the image matching process; (iii) the feature detection algorithm; (iv) the developed weighting strategy for automatic tie point measurements during bundle adjustment of oblique imagery; and (v) the number of GCPs.

Finally, in *Chapter 8*, a summary and the contribution of this dissertation are presented, the conclusions derived by the research conducted during the course of this dissertation are discussed and ideas for future research are outlined.

# Chapter 2
# Definitions and Geometric Background

## 2.1 Introduction

In this chapter, basic definitions and geometric relationships for oblique aerial images, necessary for understanding their underlying geometry and the methods introduced in the following chapters of this dissertation, are presented. The chapter starts with the definition of oblique aerial images and their well-established classification into low and high oblique ones along with their advantages and disadvantages. Then, the camera configurations for capturing only oblique aerial images or combination of oblique and vertical aerial images are outlined. Subsequently, terms associated with the geometry of oblique aerial images are defined, the angular orientation in terms of azimuth, tilt and swing is presented, the displacement due to tilt is defined and the combined effect of displacements due to relief and tilt is presented. Also, relationships for estimation of scales (*x*-scale, *y*-scale) in an oblique aerial image are presented and some basic geometrical relationships among characteristic image points and angles of oblique aerial images are defined. This chapter ends with the presentation of mathematical relationships that can be used for determining both vertical and horizontal distances from a single oblique image along with an error analysis.

## 2.2 Types of Oblique Aerial Images

Oblique aerial images are captured with the camera optical axis intentionally inclined with respect to the vertical (Moffitt and Mikhail, 1980). They differ from "tilted" or "near vertical" aerial images, which refer to those taken with the camera axis unintentionally inclined at a small angle from the vertical. According to Troney (1952), this angle is in the order of 3 or 4 degrees, while Moffitt and Mikhail (1980) define this angle as usually being less than 2 degrees. According to Höhle (2008), the tilt angle of oblique aerial images, that is, the angle between the camera axis and the vertical,

may vary from 5 grad up to 55 grad, whereas Shufelt (1999) defines oblique aerial images as the ones corresponding to a tilt angle of ≥ 5 degrees.

Oblique aerial images are characterized as either low oblique or high oblique, depending on whether they depict the apparent horizon line, that is, the actual line where the surface of the earth appears to meet the sky (section 2.4.1). Low oblique aerial images are considered as those not including the apparent horizon line, whereas the aerial images in which the apparent horizon line is visible are defined as high oblique images (Troney, 1952; Moffitt, 1962; Moffitt and Mikhail, 1980; Wolf, 1983). "Non-horizon oblique" and "horizon oblique" are other designations for low and high oblique aerial images respectively (Troney, 1952).

High oblique aerial images cover a bigger ground area than low oblique ones taken by the same camera from the same flying height, thus being richer in content. However, they correspond to a greater change of scale and stronger illumination differences throughout the image area, which makes their processing more complicated than the processing of low oblique ones. Moreover, high oblique aerial images correspond to more occlusions in comparison with low oblique aerial images; this fact raises the need for acquisition of a greater number of high oblique aerial images for the scope of complete image coverage and 3D reconstruction. On the other hand, high oblique aerial images are easier for interpretation by non-experts, who are accustomed to seeing ground features in a quite similar perspective.

## 2.3 Camera Configurations

Apart from the acquisition of single oblique aerial images from both manned and unmanned aerial platforms, multi-camera and multi-lens systems that capture a combination of only oblique or both oblique and vertical aerial imagery are available in the market, adopting different number and types of cameras as well as different acquisition geometries. Maltese-cross configurations, fan configurations and block configurations are widely used for the acquisition of such kind of imagery. A state-of-the-art review of multi-camera systems and camera configurations is given by Petrie (2009), Lemmens (2014a; 2014b), Rupnik et al. (2014) and Remondino and Gerke (2015). In the following, the most widely used camera configurations are presented.

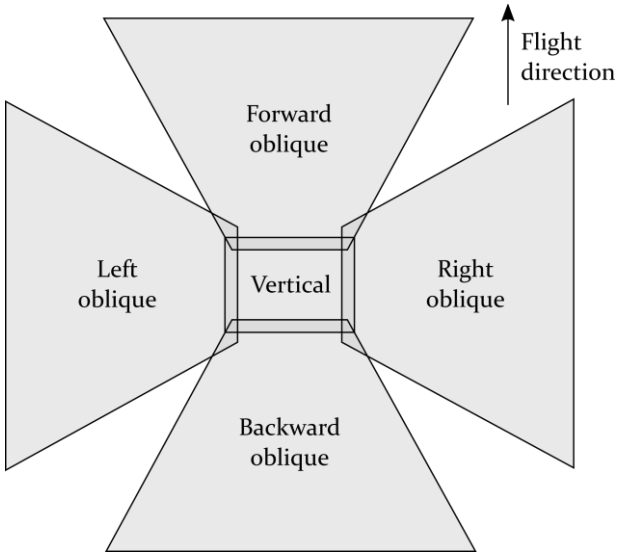


Figure 2-1. Diagram showing the Maltese-cross ground coverage

The Maltese-cross configuration (Figure 2-1) is the most commonly used one. It usually consists of a single nadir camera pointing vertical and four oblique cameras tilted towards the cardinal directions. The Maltese-cross coverage may also be accomplished by three cameras, that is, one vertical and two oblique ones pointing in opposite directions, which capture simultaneously a vertical image and two oblique ones in the across track direction and are rapidly rotated by 90 degrees to capture the second pair of oblique images pointing in the along track direction.

The fan configuration (Figure 2-2) refers to the combination of two or more cameras viewing at a different angle with their optical axes lying in the same vertical plane. Specifically, it may refer to one vertical and one oblique camera or two oblique cameras oriented either across or along track. The fan configuration may also be adopted by three cameras, among which the middle one takes vertical images and the other two take oblique images. Another approach of providing wide across track coverage of the ground area through fan configuration is the usage of a multi-lens camera comprising four, six or eight lenses that point obliquely to both sides of the flight line. Furthermore, stepping frame cameras capture a sequence of oblique images across track at a very high speed to provide a large field of view across the flight direction using a fan configuration; a single camera or twin cameras placed on a rotating mount may be used. The overlap between the individual images acquired through a fan configuration may allow for rectification and stitching.

Finally, according to the block configuration, a group of oblique cameras, usually four or six ones, are arranged in a block taking oblique aerial images. The overlap between the captured images allows them to be rectified and stitched together into a larger format composite image.

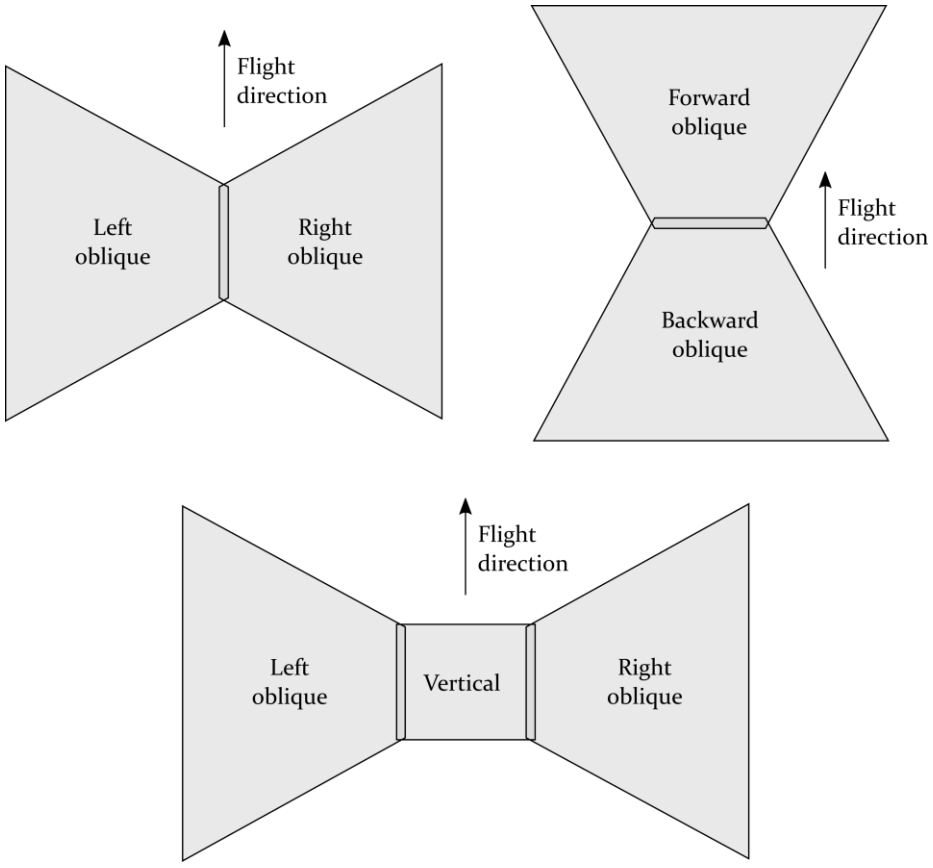


Figure 2-2. Diagrams showing the ground coverage of several fan configuration scenarios; top: fan configuration adopted by two oblique cameras oriented across track (left) and along track (right); bottom: fan configuration adopted by three cameras

## 2.4 Geometry of Oblique Imagery

In this section, terms associated with oblique aerial images along with their geometric properties are defined; the angular orientation in terms of azimuth, tilt and swing is presented; the displacement due to tilt is defined; the combined effect of relief and tilt is presented; the relationships for computing the *x*-scale and *y*-scale in an oblique aerial image are presented; and some basic geometrical relationships among characteristic points and angles in an oblique aerial image are established.

### 2.4.1 Terms and Geometric Properties

In this section, terms associated with the geometry of oblique aerial images are defined and their basic geometric properties are presented (Hydrographic Office Navy Department and Photographic Intelligence Center, 1945; American Society of Photogrammetry, 1952; Troney, 1952; Moffitt and Mikhail, 1980; Paine and Kiser, 2012).

The *principal plane* of an oblique aerial image is the vertical plane that includes the optical axis of the camera and the vertical line from the projection center; it intersects the oblique image plane at the *principal line*. The latter passes through the image nadir point and the principal point. Both the principal plane and the principal line make sense only in oblique imagery. The principal line is the line of biggest inclination in an oblique aerial image. All image points that lie on lines perpendicular to the principal line correspond to the same scale, assuming horizontal ground.

The *isocenter* (*I΄*) is the point where the bisector of the angle between the vertical line from the projection center and the optical axis of the camera intersects the oblique image plane (Figure 2-3, Figure 2-4). It is the intersection point of the principal line and the line that results from the intersection of the oblique image plane and the plane of an assumed truly vertical image with the

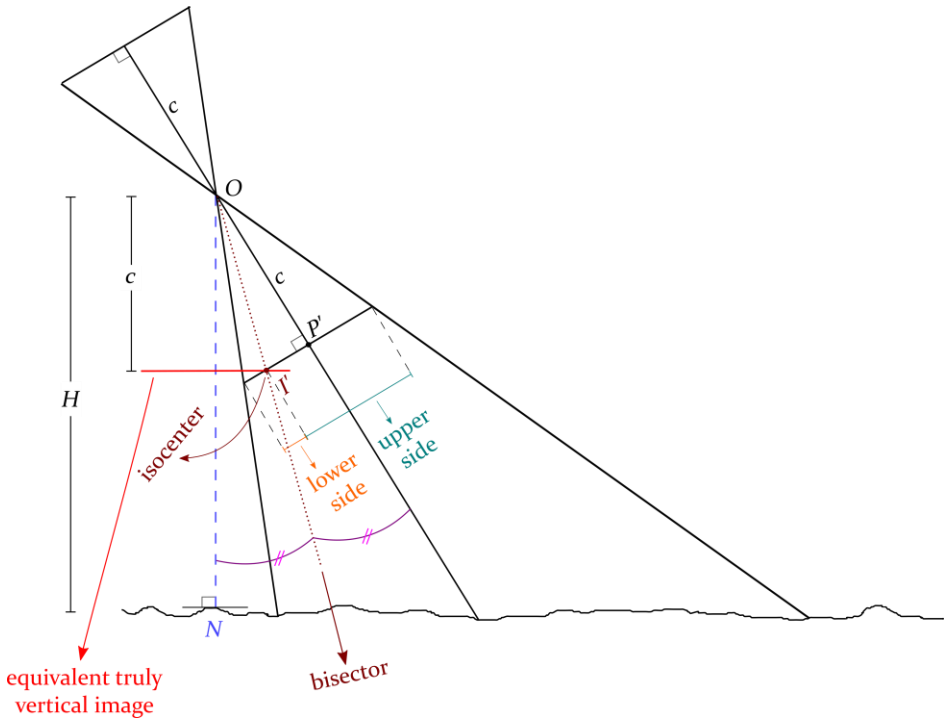


Figure 2-3. Section at the principal plane of an oblique aerial image showing the isocenter and the upper and lower side of the oblique image

same projection center taken by the same camera with the same focal length (equivalent truly vertical image). The isocenter lies on the principal plane, the oblique image plane and the plane of the equivalent truly vertical image. The displacement due to tilt in an oblique image is radial with respect to the isocenter. Similarly, this term makes sense only in oblique imagery. Its main characteristic is the fact that angles measured from the isocenter of an oblique aerial image are true, i.e., they are equal to the real angles measured from the ground isocenter.

The *axis of tilt*, or equivalently the *isometric parallel*, is the line where the oblique image plane meets the plane of an equivalent truly vertical image (Figure 2-4). It is perpendicular to the principal line and passes through the isocenter. The axis of tilt, like every image line perpendicular to the principal line, is a horizontal line and corresponds to zero displacement due to tilt relative to an equivalent truly vertical image. Along the axis of tilt, the image scale is constant and equals to $c/H$ (where $c$ is the camera constant and $H$ is the flying height), assuming horizontal ground. The axis of tilt separates the oblique image into two sides, i.e., the upper side and the lower side (Figure 2-3). The local scale is larger than $c/H$ below the axis of tilt ("lower" side of image) and smaller than $c/H$ above the axis of tilt ("upper" side of the image).

The *true horizon line* is the intersection of the horizontal plane including the perspective center and the oblique image plane (Figure 2-5). It is perpendicular to the principal line and parallel to the axis of tilt. The *apparent horizon line* is the actual image line where the surface of the earth appears to meet the sky. It is a real line depicted in a high oblique aerial image and appears below the fantastic true horizon line. The apparent horizon line is depicted in a small-scale high oblique aerial image as a slightly curved line in case of flat ground or as a broken line in which a slight curvature may be discernible, in case of strong relief. The *horizon point* ($K'$) is the intersection of the principal line with the horizontal plane through the perspective center, or equivalently, the intersection of the principal line with the true horizon line (Figure 2-5).

The *dip angle* ($\delta$) is the angle measured on the principal plane of an oblique aerial image between the true horizon line and the apparent horizon line. The *apparent depression angle* ($\gamma$) is the angle measured on the principal plane between the camera optical axis and the apparent horizon. The *depression angle* ($\theta$) is the angle measured on the principal plane between the camera axis and the true horizon. Due to earth curvature and flying height, the apparent horizon appears below the true

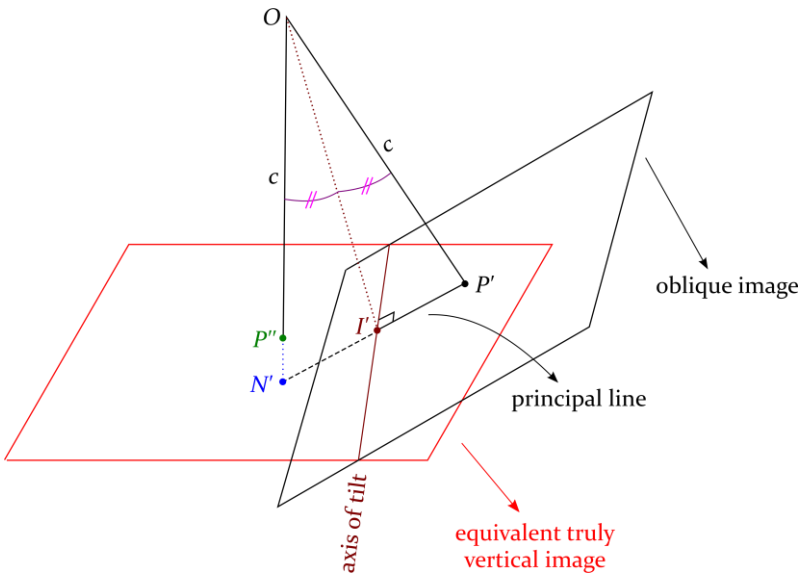


Figure 2-4. Oblique image geometry: axis of tilt, principal line, isocenter

horizon; thus the apparent depression angle incremented by the dip angle gives the depression angle (Figure 2-6).

### 2.4.2 Angular Orientation in Azimuth-Tilt-Swing

The *azimuth* angle (*a*) is the clockwise horizontal angle measured about the ground nadir point from a line parallel to the ground *Y*-axis to the principal plane of the image. The *tilt* angle (*t*) is the angle between the vertical line from the projection center and the camera axis. The *swing* angle (*s*) is the clockwise angle measured at the oblique image plane from the positive *y*-axis to the principal line, at the side of the image nadir point (Figure 2-5) (Moffitt and Mikhail, 1980; Punmia et al., 2005). The rotation matrix, $\boldsymbol{R}$, expressed in angles azimuth, tilt and swing is given by equation (2.1) (Dewitt, 1996).

$$\boldsymbol{R} = \begin{bmatrix} -\cos s \cdot \cos a - \sin s \cdot \cos t \cdot \sin a & \cos s \cdot \sin a - \sin s \cdot \cos t \cdot \cos a & -\sin s \cdot \sin t \\ \sin s \cdot \cos a - \cos s \cdot \cos t \cdot \sin a & -\sin s \cdot \sin a - \cos s \cdot \cos t \cdot \cos a & -\cos s \cdot \sin t \\ -\sin t \cdot \sin a & -\sin t \cdot \cos a & \cos t \end{bmatrix} \quad (2.1)$$

The angles azimuth, tilt and swing are computed from the elements $r_{ij}$ of the rotation matrix, according to the set of equations (2.2).

$$a = \tan^{-1}\left(-r_{31} / -r_{32}\right); \quad t = \cos^{-1}\left(r_{33}\right); \quad s = \tan^{-1}\left(-r_{13} / -r_{23}\right) \quad (2.2)$$

### 2.4.3 Tilt Displacement

The displacement due to tilt (tilt displacement) for a point is defined as the distance from the isocenter to the image of the point in an equivalent truly vertical photograph minus the distance from the isocenter to the image of the point in the oblique photograph (Wolf, 1983). Image points located on the lower side of the oblique image are radially displaced away from the isocenter, thus corresponding to larger scale than the scale that they would have in an equivalent truly vertical image. Points imaged on the upper side of the image are radially displaced towards the isocenter, thus corresponding to smaller scale than that in an equivalent truly vertical image.

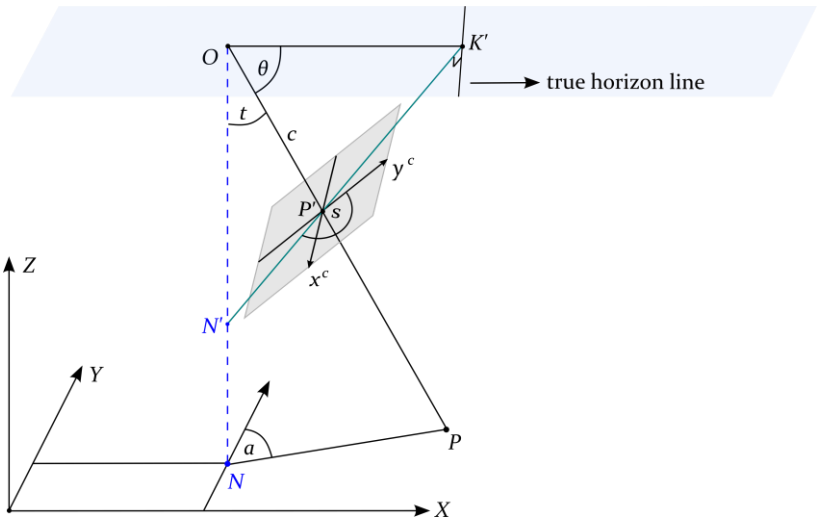


Figure 2-5. Orientation angles azimuth (*a*), tilt (*t*) and swing (*s*), depression angle (θ), true horizon line and horizon point (*K′*)

The algebraic value of the displacement due to tilt ($\Delta r_{tilt}$) for a point of an oblique aerial image is given by equation (2.3) (Misulia, 1946; Wolf, 1983), where $r_I$ is the radial distance from the isocenter to the point, $c$ is the camera constant, $t$ is the tilt angle and $\xi$ is the angle measured clockwise from the positive end of the principal line to the radial line from the isocenter to the point.

$$\Delta r_{tilt} = \frac{r_I^{\,2} \cdot \cos^2 \xi \cdot \sin t}{c - r_I \cdot \cos \xi \cdot \sin t} \tag{2.3}$$

If the point is located on the lower side of the oblique image, its corrected radial distance from the isocenter, $r_I^c$, is computed as: $r_I^c = r_I - \Delta r_{tilt}$. If the point is located on the upper side of the oblique image, its corrected radial distance from the isocenter is computed as: $r_I^c = r_I + \Delta r_{tilt}$

On the contrary, the displacement of a point due to its elevation is radial from the nadir point and always positive, that is, points elevated with respect to the datum are radially displaced outwards, away from the nadir point. The relief displacement ($\Delta r_{relief}$) for a point is given by equation (2.4), where $r_N$ is the radial distance from the isocenter to the point, $h$ is its elevation and $H$ is the flying height above datum.

$$\Delta r_{relief} = \frac{r_N \cdot h}{H} \tag{2.4}$$

The combined effect of relief displacement and tilt displacement is illustrated in Figure 2-7. Each point with subscript the number 1 $(A'_1, B'_1, C'_1, D'_1, E'_1)$ corresponds to the position that the point would have without the effects of relief and tilt displacement (corrected position). Each point with subscript the number 2 $(A'_2, B'_2, C'_2, D'_2, E'_2)$ represents the position of the point after it has undergone relief displacement, that is, after it has been radially displaced outwards from the nadir point. Each point with subscript the number 3 $(A'_3, B'_3, C'_3, D'_3, E'_3)$ represents the position of the point after it has undergone tilt displacement (observed position of the image point). Specifically, point $A'_2$ lies on the axis of tilt; thus, it is not displaced due to tilt and point $A'_3$ coincides with point $A'_2$. Point $B'_2$ is radially displaced away from the isocenter to be represented in position $B'_3$, as it is located in the lower side of the image. Point $C'_2$ is radially displaced towards the isocenter by the amount $(C'_2 C'_3)$, as it lies in the upper side of the image. Points $D'_2$ and $E'_2$ are

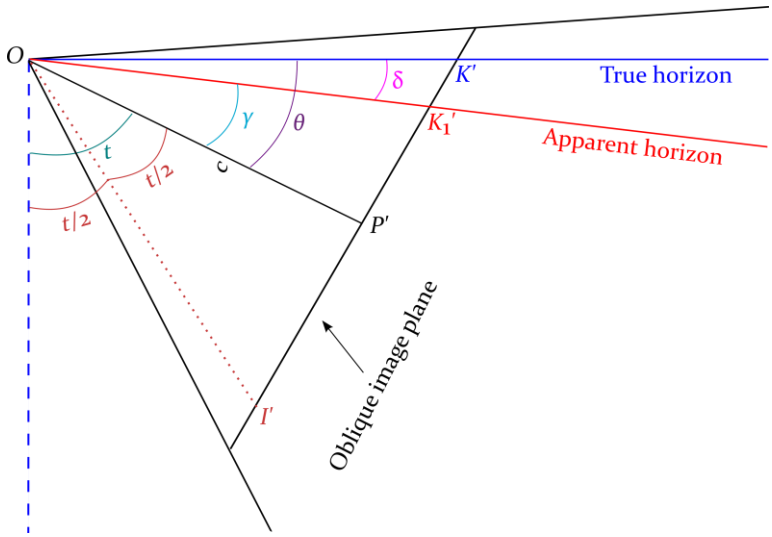


Figure 2-6. Section at the principal plane of an oblique aerial image showing the dip angle ($\delta$), the apparent depression angle ($\gamma$), the depression angle ($\theta$) and the tilt angle ($t$)

displaced along the principal line radially inwards and outwards with respect to the isocenter, respectively; for these points, the relief and tilt displacement are cumulative.

### 2.4.4 Scale

The scale at a point on an oblique aerial image in not the same in all directions. The scale of lines perpendicular to the principal line (*x*-scale) and the scale of lines parallel to the principal line (*y*-scale) are not the same. The scale of an oblique aerial image varies along the direction of the principal line. The scale of line segments perpendicular to the principal line is constant, whereas the scale of all the other line segments of an oblique aerial image varies. In this section, for the sake of completeness, the derivations of the well-known relationships for the computation of the scale of lines perpendicular to the principal line and infinitesimal line segments parallel to the principal line are shown. The reader may also refer to Lane (1950), Moffitt (1967), Moffitt and Mikhail (1980) and Wolf (1984).

#### 2.4.4.1 Scale of Lines Perpendicular to the Principal Line

The scale of an oblique aerial image is constant along each line that is perpendicular to the principal line. The scale of a line perpendicular to the principal line is called *x*-scale. The *x*-scale increases along the principal line from the horizon point to the nadir point. Let $T'$ be the image of a real world point $T$ on the oblique photograph and let $T_1'$ be the projection of $T'$ on the principal line (Figure 2-8); thus the line segment $T'T_1'$ is perpendicular to the principal line. The point $T_1$ is the projection of $T$ on the principal plane. The line segments $T'T_1'$ and $TT_1$ are horizontal.

The *x*-scale along the line that passes through $T'$ and $T_1'$ is given by equation (2.5), taking into account the similar triangles $OT'T_1'$ and $OTT_1$.

$$s_x = \frac{T'T_1'}{TT_1} = \frac{OT'}{OT} = \frac{OT_1'}{OT_1} \tag{2.5}$$

By trigonometric relations on the right triangles $OP'T_1'$ and $ON_1T_1$, equations (2.6) and (2.7) are

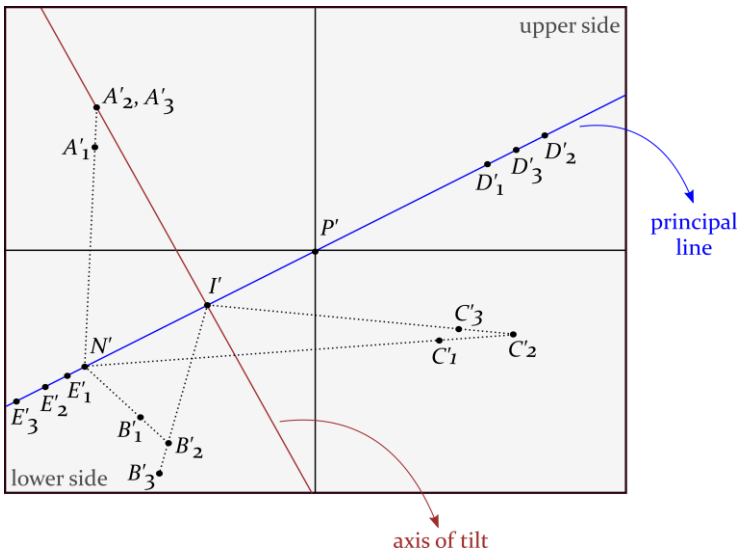


Figure 2-7. Combined effect of relief and tilt in an oblique aerial image

derived, where $\zeta$ is the angle formed by the camera axis and the ray to the point $T_1'$ and $\Delta H$ is the difference in elevation between the point $T$ and the ground nadir point, being positive if $T$ is above the ground nadir point $N$ and negative if $T$ is below the ground nadir point.

$$OT_1' = \frac{c}{\cos\zeta} \tag{2.6}$$

$$\begin{aligned} OT_1 &= \frac{H-\Delta H}{\cos(t+\zeta)}, \text{ if } T_1' \text{ lies in the ray } \overrightarrow{P'K'} \\ OT_1 &= \frac{H-\Delta H}{\cos(t-\zeta)}, \text{ if } T_1' \text{ lies in the ray } \overrightarrow{P'N'} \end{aligned} \tag{2.7}$$

Using equations (2.6) and (2.7), equation (2.5) takes the following form (equations (2.8)), for computation of the $x$-scale of a point $T'$ or $T_1'$.

$$\begin{aligned} s_x &= \frac{c\cos(t+\zeta)}{(H-\Delta H)\cos\zeta}, \text{ if } T_1' \text{ lies in the ray } \overrightarrow{P'K'} \\ s_x &= \frac{c\cos(t-\zeta)}{(H-\Delta H)\cos\zeta}, \text{ if } T_1' \text{ lies in the ray } \overrightarrow{P'N'} \end{aligned} \tag{2.8}$$

Alternatively, the $x$-scale along a line $l$ perpendicular to the principal line may be expressed in terms of the distance $d$ between the true horizon line and the line $l$ (measured parallel to the principal line); $d$ is calculated by the set of equations (2.9), which is derived by applying trigonometric relations in the right triangles $OP'K'$ and $OP'T_1'$.

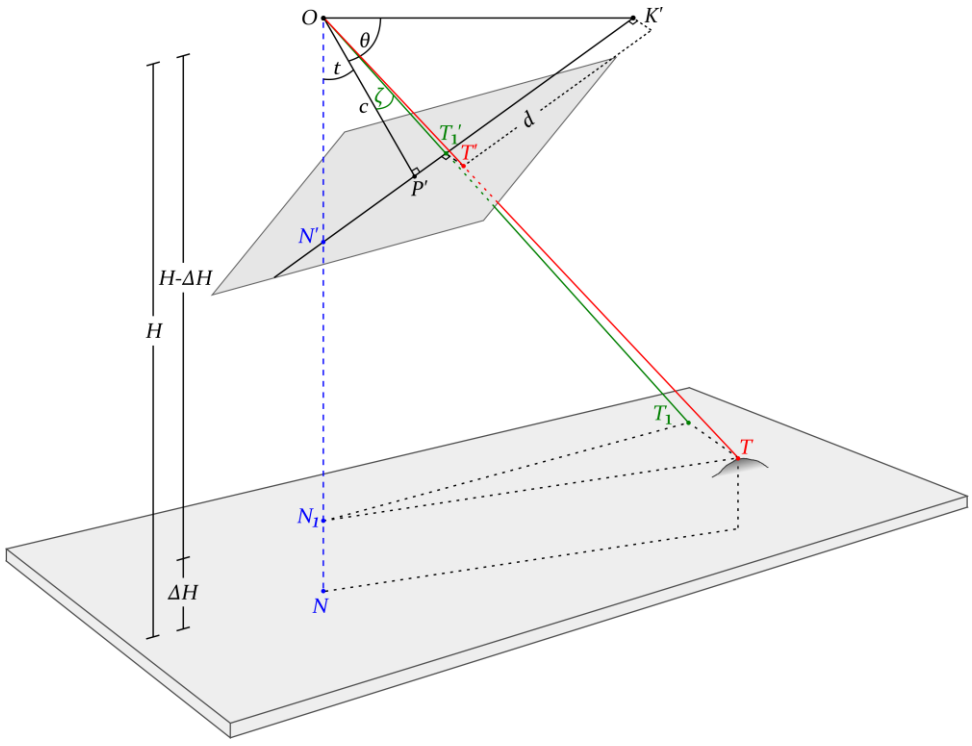


Figure 2-8. Oblique aerial image geometry: calculation of the $x$-scale of a line perpendicular to the principal line

$$\begin{aligned} d &= K'P' - P'T_1' = c\left(\tan\theta - \tan\zeta\right), \text{ if } T_1' \text{ lies in the ray } \overrightarrow{P'K'} \\ d &= K'P' + P'T_1' = c\left(\tan\theta + \tan\zeta\right), \text{ if } T_1' \text{ lies in the ray } \overrightarrow{P'N'} \end{aligned} \tag{2.9}$$

By trigonometric identity, the set of equations (2.9) takes the following form (equations (2.10)).

$$\begin{aligned} &\text{if } T_1' \text{ lies in the ray } \overrightarrow{P'K'}: \\ d &= c\frac{\sin\left(\theta-\zeta\right)}{\cos\theta\cos\zeta} = c\frac{\sin\left(90°-t-\zeta\right)}{\cos\theta\cos\zeta} = c\frac{\sin\left(90°-\left(t+\zeta\right)\right)}{\cos\theta\cos\zeta} = c\frac{\cos\left(t+\zeta\right)}{\cos\theta\cos\zeta} \\ &\text{if } T_1' \text{ lies in the ray } \overrightarrow{P'N'}: \\ d &= c\frac{\sin\left(\theta+\zeta\right)}{\cos\theta\cos\zeta} = c\frac{\sin\left(90°-t+\zeta\right)}{\cos\theta\cos\zeta} = c\frac{\sin\left(90°-\left(t-\zeta\right)\right)}{\cos\theta\cos\zeta} = c\frac{\cos\left(t-\zeta\right)}{\cos\theta\cos\zeta} \end{aligned} \tag{2.10}$$

Substituting equations (2.10) in equations (2.8), the *x*-scale of a point or line perpendicular to the principal line is given by equation (2.11) (Lane, 1950; Moffitt, 1967; Moffitt and Mikhail, 1980; Wolf, 1984).

$$s_x = \frac{d\cos\theta}{H - \Delta H} \tag{2.11}$$

### 2.4.4.2 Scale of Lines Parallel to the Principal Line

The scale of a line parallel to the principal line, i.e., perpendicular to the lines of constant scale, varies along this line. The scale in a direction parallel to the principal line is called *y*-scale. Since the scale in the direction parallel to the principal line varies from point to point, the *y*-scale may be considered to be constant only for an infinitesimal distance. Let $T'$ and $T_2'$ be the images of the real world points $T$ and $T_2$ on the oblique photograph that lie on the principal plane, such that the line segment $T'T_2'$ lies on the principal line and the line segments $T'T_2'$ and $TT_2$ are of infinitesimal length (Figure 2-9). The *y*-scale of the infinitesimal line segment $T'T_2'$ is computed by equation (2.12).

$$s_y = \frac{T'T_2'}{TT_2} \tag{2.12}$$

Equation (2.13) is derived by the law of sines in the triangle $T_2TW$ (Figure 2-9).

$$\frac{TT_2}{\sin\left(T_2WT\right)} = \frac{TW}{\sin\left(TT_2W\right)} \quad \rightarrow \quad TT_2 = TW\frac{\sin\left(T_2WT\right)}{\sin\left(TT_2W\right)} \tag{2.13}$$

The line segments $OT$ and $OT_2$ may be considered to be parallel, as the line segment $TT_2$ is infinitesimal. Thus, the angle $PT_2O$, may be considered to be equal to the angle $PTO$. Hence, the angle $TT_2W$, which is equal to the angle $PT_2O$, is computed as follows (equations (2.14)).

$$\begin{aligned} \text{angle } TT_2W &= \theta - \zeta, \text{ if } T' \text{ and } T_2' \text{ lie in the ray } \overrightarrow{P'K'} \\ \text{angle } TT_2W &= \theta + \zeta, \text{ if } T' \text{ and } T_2' \text{ lie in the ray } \overrightarrow{P'N'} \end{aligned} \tag{2.14}$$

In this way, the angle $T_2WT$ is computed as follows (equations (2.15)).

$$\begin{aligned} &\text{angle } T_2WT = 90° + \zeta, \text{ if } T' \text{ and } T_2' \text{ lie in the ray } \overrightarrow{P'K'} \\ &\text{angle } T_2WT = 90° - \zeta, \text{ if } T' \text{ and } T_2' \text{ lie in the ray } \overrightarrow{P'N'} \end{aligned} \tag{2.15}$$

Hence, equation (2.13) takes the following form (equations (2.16)).

$$\begin{aligned} TT_2 &= TW\frac{\sin(90° + \zeta)}{\sin(\theta - \zeta)}, \text{ if } T' \text{ and } T_2' \text{ lie in the ray } \overrightarrow{P'K'} \\ TT_2 &= TW\frac{\sin(90° - \zeta)}{\sin(\theta + \zeta)}, \text{ if } T' \text{ and } T_2' \text{ lie in the ray } \overrightarrow{P'N'} \end{aligned} \tag{2.16}$$

Substituting the set of equations (2.16) in equation (2.12), equations (2.17) are derived.

$$\begin{aligned} s_y &= \frac{T'T_2'\sin(\theta - \zeta)}{TW\sin(90° + \zeta)}, \text{ if } T' \text{ and } T_2' \text{ lie in the ray } \overrightarrow{P'K'} \\ s_y &= \frac{T'T_2'\sin(\theta + \zeta)}{TW\sin(90° - \zeta)}, \text{ if } T' \text{ and } T_2' \text{ lie in the ray } \overrightarrow{P'N'} \end{aligned} \tag{2.17}$$

From the similar triangles $OT'T_2'$ and $OTW$ and equation (2.5), equation (2.18) is derived.

$$\frac{T'T_2'}{TW} = \frac{OT'}{OT} = s_x \tag{2.18}$$

Substituting the set of equations (2.8) in equation (2.18) and substituting the result in the set of equations (2.17), the set of equations (2.19) is derived.

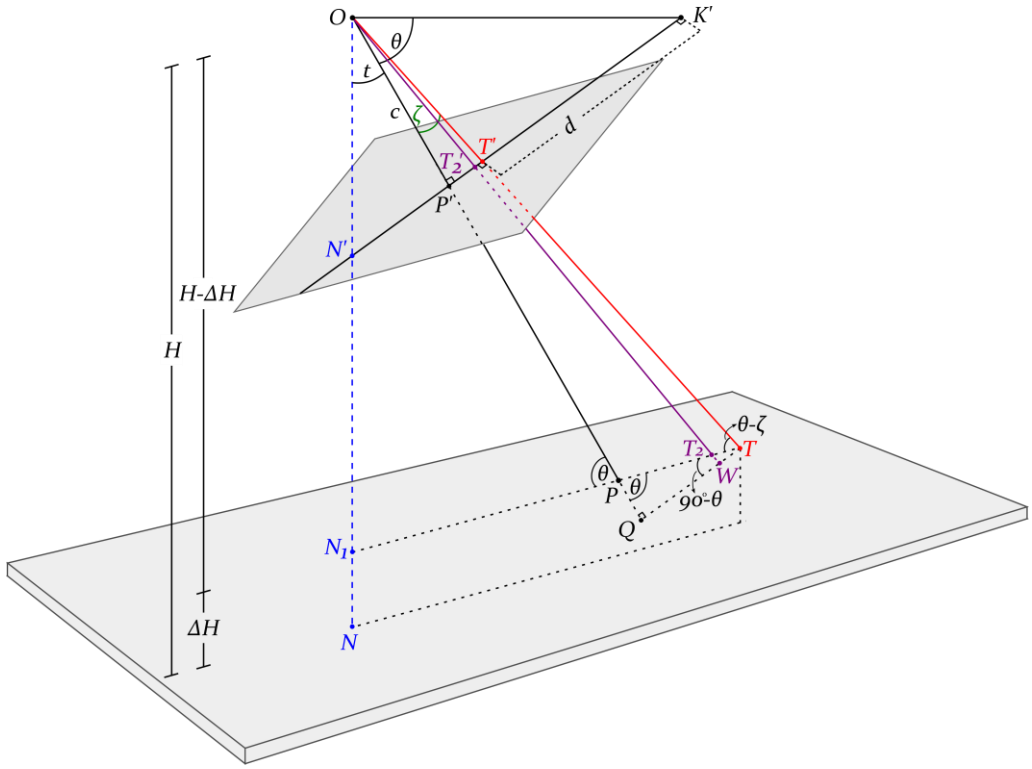


Figure 2-9. Oblique aerial image geometry: calculation of the *y*-scale of an infinitesimal line segment perpendicular to the true horizon line

$$\text{if } T' \text{ and } T_2' \text{ lie in the ray } \overrightarrow{P'K'}:$$

$$s_y = \frac{c\cos(t+\zeta)\sin(\theta-\zeta)}{(H-\Delta H)\cos\zeta\sin(90°+\zeta)} = \frac{c\sin(\theta-\zeta)\sin(\theta-\zeta)}{(H-\Delta H)\cos\zeta\cos\zeta} = \frac{c}{(H-\Delta H)}\left(\frac{\sin(\theta-\zeta)}{\cos\zeta}\right)^2$$

$$\text{if } T' \text{ and } T_2' \text{ lie in the ray } \overrightarrow{P'N'}: \tag{2.19}$$

$$s_y = \frac{c\cos(t-\zeta)\sin(\theta+\zeta)}{(H-\Delta H)\cos\zeta\sin(90°-\zeta)} = \frac{c\sin(\theta+\zeta)\sin(\theta+\zeta)}{(H-\Delta H)\cos\zeta\cos\zeta} = \frac{c}{(H-\Delta H)}\left(\frac{\sin(\theta+\zeta)}{\cos\zeta}\right)^2$$

From equations (2.10), equation (2.20) are derived.

$$\frac{\sin(\theta-\zeta)}{\cos\zeta} = \frac{d\cos\theta}{c}, \text{ if } T' \text{ lies in the ray } \overrightarrow{P'K'}$$

$$\frac{\sin(\theta+\zeta)}{\cos\zeta} = \frac{d\cos\theta}{c}, \text{ if } T' \text{ lies in the ray } \overrightarrow{P'N'} \tag{2.20}$$

Substituting equations (2.20) in equations (2.19), the *y*-scale of the infinitesimal line segment $T'T_2'$ (or a point $T'$) is computed by equation (2.21).

$$s_y = \frac{d^2 \cdot \cos^2\theta}{c(H-\Delta H)} \tag{2.21}$$

Whereas the *x*-scale represents the scale of a point in a direction perpendicular to the principal line, the *y*-scale represents its scale in the direction parallel to the principal line. In the case of line segments, the *x*-scale represents the scale of a line segment which is perpendicular to the principal line, whereas the *y*-scale represents the scale of a line segmemt that is parallel to the principal line, provided that the latter is of infinitesimal length, as the *y*-scale can be considered constant only for an infinitesimal distance. These scales can be used in order to compute ground distances. Since the *x*-scale and the *y*-scale differ by the factor $d\cos\theta/c$ (i.e., $s_y/s_x = d\cos\theta/c$) and the angle $\theta$ is positive and smaller than 90º, $s_y < s_x$ if $d < c$, whereas $s_y > s_x$ if $d > c$. Thus, for a specific image, the smaller the distance *d* between a point and the true horizon line is, the smaller the *y*-scale is compared to the *x*-scale, assuming that $d < c$. In the case that $d > c$ for a specific image, the bigger *d* is, the bigger the *y*-scale is compared to the *x*-scale. The length of a line segment that is neither perpendicular nor parallel to the principal line can be roughly computed by obtaining the lengths of its components, perpendicular and parallel to the principal line, using the *x*-scale and the *y*-scale, respectively. However, it should be noted that since the scale is constant only along lines perpendicular to the principal line and the *y*-scale assumes an infinitesimal distance, ground distances of line segments that are not perpendicular to the principal line may be only approximately computed using the scales given by equations (2.11) and (2.21).

### 2.4.5 Basic Geometrical Relationships

In this section, basic mathematical relationships among the principal point, the horizon point, the nadir point, the isocenter, the tilt angle, the depression angle, the apparent depression angle and the swing angle as well as between the flying height and the dip angle are presented and their derivations are shown in this section for the sake of completeness. These mathematical relationships are particularly useful for determining the geometry of a single oblique aerial image. For instance,

if the nadir point is automatically estimated from an oblique aerial image of known interior orientation (e.g., via the method presented in section 6.3 of this dissertation), the tilt angle, the depression angle, the swing angle, the horizon point and the isocenter may also be calculated using the relationships presented in this section.

### 2.4.5.1 Tilt and Depression Angle, Nadir Point and Horizon Point

The depression angle is the complement of the tilt angle (Figure 2-9); thus, it is calculated by equation (2.22).

$$\theta = 90° - t \tag{2.22}$$

The depression angle may be calculated as a function of the distance between the principal point and the horizon point (Figure 2-9), according to equation (2.23).

$$K'P' = c\tan\theta \quad \rightarrow \quad \theta = \arctan\left(\frac{K'P'}{c}\right) \tag{2.23}$$

The tilt angle may be calculated using the distance between the principal point and the image of the nadir point (Figure 2-9), according to equation (2.24).

$$P'N' = c\tan t \quad \rightarrow \quad t = \arctan\left(\frac{P'N'}{c}\right) \tag{2.24}$$

Taking into account the fact that the tilt angle is the complement of the depression angle, the distance between the principal point and the image of the nadir point may, alternatively, be expressed as a function of the depression angle, according to equation (2.25).

$$P'N' = c\tan\left(90° - \theta\right) \quad \rightarrow \quad P'N' = c\cot\theta \tag{2.25}$$

The distance between the horizon point and the nadir point is calculated by equation (2.26), as a function of the depression angle, or by equation (2.27), as a function of the tilt angle (Figure 2-9).

$$K'N' = K'P' + P'N' \quad \rightarrow \quad K'N' = c\tan\theta + c\cot\theta \quad \rightarrow \quad K'N' = c\left(\tan\theta + \cot\theta\right) \tag{2.26}$$

$$K'N' = c\left(\cot t + \tan t\right) \tag{2.27}$$

### 2.4.5.2 Isocenter

The distance between the isocenter and the principal point is calculated using equation (2.28), as a function of the tilt angle (Figure 2-6), or by equation (2.29), as a function of the depression angle.

$$P'I' = c\tan\left(\frac{t}{2}\right) \tag{2.28}$$

$$P'I' = c\tan\left(\frac{90° - \theta}{2}\right) \quad \rightarrow \quad P'I' = c\tan\left(45° - \frac{\theta}{2}\right) \tag{2.29}$$

The distance between the horizon point and the isocenter is calculated by equation (2.30), as a function of the depression angle (Figure 2-6), using equations (2.23) and (2.29).

$$K'I' = K'P' + P'I' \quad \rightarrow \quad K'I' = c\left(\tan\theta + \tan\left(45° - \frac{\theta}{2}\right)\right) \tag{2.30}$$

Equation (2.31) is derived by mathematical operations and trigonometric identities.

$$\tan\left(45° - \frac{\theta}{2}\right) = \frac{1-\sin\theta}{\cos\theta} \tag{2.31}$$

By substituting equation (2.31) in equation (2.30), equation (2.32) is finally derived for the calculation of the distance between the horizon point and the isocenter as a function of the depression angle.

$$K'I' = c\left(\frac{\sin\theta}{\cos\theta} + \frac{1-\sin\theta}{\cos\theta}\right) \quad \rightarrow \quad K'I' = \frac{c}{\cos\theta} \tag{2.32}$$

The distance between the horizon point and the isocenter may also be calculated by equation (2.33) as a function of the tilt angle, using equation (2.32) and taking into account the fact that the tilt angle is the complement of the depression angle.

$$K'I' = \frac{c}{\sin t} \tag{2.33}$$

### 2.4.5.3 Swing Angle

The swing angle may be calculated according to the set of equations (2.34), using the coordinates of the nadir point in the image coordinate system centered at the principal point ($x^c$, $y^c$) (Figure 2-10).

$$s = \begin{Bmatrix} 90° + \arctan\left|\dfrac{x^c_{N'}}{y^c_{N'}}\right| & \text{if } x^c_{N'} \le 0 \\ 90° - \arctan\left|\dfrac{x^c_{N'}}{y^c_{N'}}\right| & \text{if } x^c_{N'} > 0 \end{Bmatrix} \tag{2.34}$$

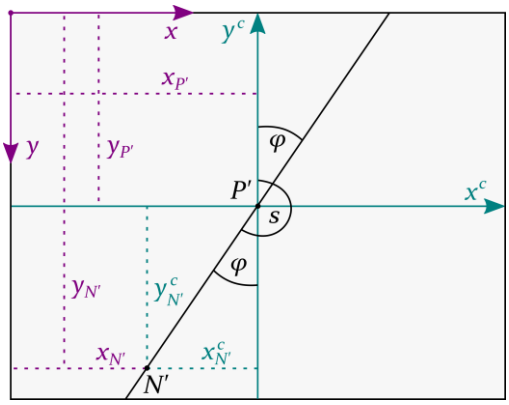


Figure 2-10. Calculation of the swing angle of an oblique aerial image

The coordinates of the nadir point in the image coordinate system centered at the principal point ($x^c$, $y^c$) are calculated using its corresponding pixel coordinates and the pixel coordinates of the principal point. Assuming that the image coordinate system in pixels ($x$, $y$) has its origin at the top left pixel, the swing angle is computed by the set of equations (2.35), as a function of the pixel coordinates of the principal point and the nadir point (Figure 2-10).

$$s = \begin{cases} 90° + \arctan\left|\dfrac{x_{N'} - x_{P'}}{y_{P'} - y_{N'}}\right| & \text{if } x_{N'} \le x_{P'} \\ 90° - \arctan\left|\dfrac{x_{N'} - x_{P'}}{y_{P'} - y_{N'}}\right| & \text{if } x_{N'} > x_{P'} \end{cases} \tag{2.35}$$

### 2.4.5.4 Dip Angle

The dip angle is geometrically calculated using the right triangle $OCK_1$ (Figure 2-11) according to equation (2.36), where $OK_1$ is the trace of the apparent horizon in the principal plane and $R_{earth}$ is the radius of the earth, assuming that the earth is approximated by a sphere.

$$\tan\delta = \frac{OK_1}{R_{earth}} \quad \rightarrow \quad \delta = \arctan\left(\frac{OK_1}{R_{earth}}\right) \tag{2.36}$$

In the right triangle $OCK_1$, equation (2.37) is derived.

$$OK_1 = \sqrt{\left(R_{earth} + H\right)^2 - {R_{earth}}^2} \quad \rightarrow \quad OK_1 = \sqrt{2R_{earth}H + H^2} \tag{2.37}$$

Substituting equation (2.37) in equation (2.36), equation (2.38) is derived for the calculation of the dip angle.

$$\delta = \arctan\left(\frac{\sqrt{2R_{earth}H + H^2}}{R_{earth}}\right) \tag{2.38}$$

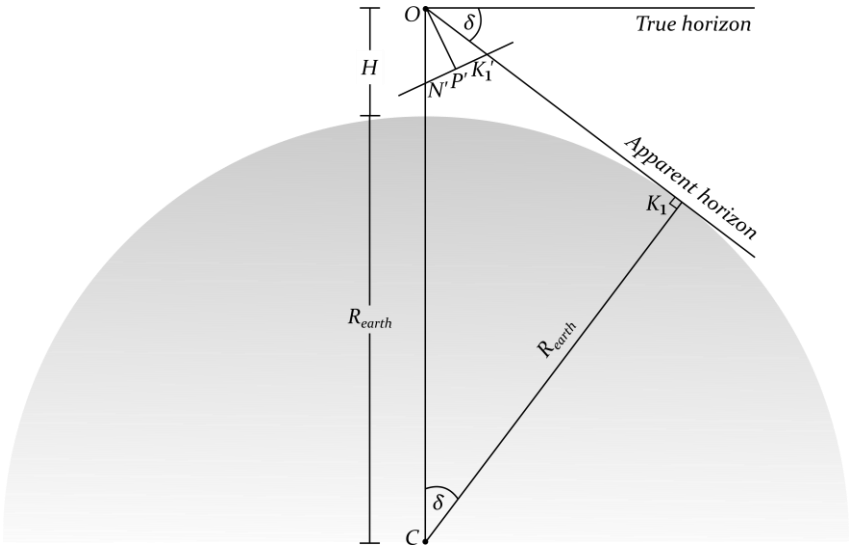


Figure 2-11. Section at the principal plane of an oblique aerial image illustrating the calculation of the dip angle

Considering that the radius of earth is constant, the dip angle is a function of the flying height and increases with it. By ignoring the term $H^2$ of equation (2.38), as it is very small compared to the quantity $2R_{earth}H$, and substituting the average value of the radius of the earth, the dip angle is approximated by equation (2.39) (Moffitt and Mikhail, 1980).

$$\delta = 106.5\sqrt{H},\ \delta \text{ in seconds, } H \text{ in meters} \tag{2.39}$$

#### 2.4.5.5 Apparent Depression Angle

The apparent depression angle is calculated by equation (2.40), where $K_1'$ is the point of intersection of the apparent horizon line and the principal line (Figure 2-6).

$$\tan\gamma = \frac{P'K_1'}{c} \quad \rightarrow \quad \gamma = \arctan\left(\frac{P'K_1'}{c}\right) \tag{2.40}$$

Equation (2.41) expresses the relationship between the apparent depression angle, the depression angle and the dip angle (Figure 2-6).

$$\theta = \gamma + \delta \tag{2.41}$$

## 2.5 Determination of Distances

In this section, the mathematical relationships that can be used for the determination of both vertical and horizontal distances from a single oblique image are presented. The prerequisite for the monocular derivation of such metric information from an oblique image is the knowledge of the following: (i) the nadir point of the image; (ii) the camera interior orientation; and (iii) the flying height measured from the bottom point of the vertical object, in the case of measuring a vertical distance, or the flying height measured from the two points of same elevation, in the case of measuring a horizontal distance. The nadir point can be automatically determined via the method presented in section 6.3 of this dissertation.

### 2.5.1 Vertical Distances

Hartley and Zisserman (2003) compute the height of a vertical object from a single undistorted image using knowledge of the vanishing line of the horizontal direction, as well as the nadir point and image measurements of the top and bottom point of the vertical object, along with one object of known height for which the top and base are imaged. This is a well-established relationship that is mainly adopted for the case of terrestrial images, for which knowledge of the height of a reference small object depicted in such images is generally available. Molinari et al. (2014) make measurements of vertical distances from a single oblique image, using knowledge of all camera settings, flight parameters as well as detailed digital terrain models (DTMs) and digital surface models (DSMs). In this section, a mathematical model for measuring vertical distances from a single oblique image is proved anew. The main motivation for proving anew such a relationship is the computation of the height of objects from a single oblique aerial image for which neither the exterior orientation parametes are available nor the height of a reference object is known. The external reference information required by the relationship presented in this section is the flying

height measured from the bottom point of the vertical object, which, in the case of an oblique aerial image, is generally more easily available than the height of a reference object depicted in the image.

Specifically, the relationship presented in this section differs from the one proved by Molinari et al. (2014) on the grounds that it does not require knowledge of the camera exterior orientation parameters and availability of detailed DTMs/DSMs. Hence, it may also be applied in cases where such data are not available. Similarly to the relationship presented in this section, the one derived by Hartley and Zisserman (2003) does not require any camera exterior orientation information. The relationship presented in this section differs from the one derived by Hartley and Zisserman (2003) in the initial data. Hartley and Zisserman (2003) assume knowledge of the height of a vertical object, for which the top and base are imaged, whereas the relationship derived in this section assumes knowledge of the flying height measured from the bottom point of the vertical object. In addition to the image nadir point (which is also a prerequisite of the relationship presented in this section), the relationship of Hartley and Zisserman (2003) assumes knowledge of the vanishing line of the horizontal direction. However, this is not a major difference, as knowledge of the image nadir point permits the determination of the true horizon line assuming known camera interior orientation parameters. This is achieved through estimation of the tilt angle, computation of the distance between the horizon point and the nadir point, computation of the coordinates of the horizon point and estimation of the equation of the horizon line. Besides, the method presented in section 6.3 of this dissertation for detecting the image nadir point also computes the true horizon line.

The derivation of the formula for the calculation of the height of a vertical object from a single oblique aerial image is presented in the following. In Figure 2-12, $B'$ and $T'$ are the projections of the bottom point $B$ and the top point $T$, respectively, of an edge of a vertical object in an oblique aerial image taken at a flying height $H$ above a specific datum. $\Delta H$ is the elevation difference between the datum plane and the bottom point of the vertical object (point $B$), being positive if $B$ is above the datum plane and negative if $B$ is below the datum plane. In Figure 2-12 it is assumed that the flying height is measured from the surface directly below the aerial platform, i.e., the datum plane is the horizontal plane that contains the ground nadir point $N$. $H - \Delta H$ is the flying height measured from the bottom point of the vertical object, i.e., the vertical distance between the camera projection center and the horizontal plane containing the bottom point of the vertical object to be measured. This is the quantity that is required by the formula which calculates the height of a vertical object. In the case that $H$ is the flying height above a typical datum and practically refers to the $Z$ coordinate of the camera projection center (e.g., orthometric height in the the World Geodetic System 1984 – WGS 84 datum), $\Delta H$ refers to the Z coordinate of the bottom point of the vertical object to be measured, which may be obtained by a free program for earth observation, such as Google Earth, and then $H - \Delta H$ may be calculated. In the case that $H$ has been measured from the suface directly below the aerial platform and, as a result, $\Delta H$ refers to the difference in elevation between the ground nadir point and the bottom point of the vertical object to be measured, the difference $H - \Delta H$ may also be obtained by a free program like Google Earth, provided that the region vertically under the aerial platform can be located in the map.

Using the similar triangles $ON_1T_1$ and $TBT_1$ (Figure 2-12), the height ($h$) of the vertical object is calculated according to equation (2.42), where $N_1$ is the intersection of the horizontal plane containing the bottom point $B$ with the vertical line from the perspective center and $T_1$ is the intersection of the horizontal plane containing the bottom point $B$ with the optical ray to the top point $T$.

$$h = (H - \Delta H)\frac{BT_1}{N_1T_1} \tag{2.42}$$

The distance $BT_1$ is calculated by equation (2.43), where $\beta_B$ is the angle between the vertical line from the perspective center and the optical ray to the bottom point $B$ of the vertical object and $\beta_T$ is the angle between the vertical line from the perspective center and the optical ray to the top point $T$ of the vertical object.

$$BT_1 = N_1T_1 - N_1B \quad \rightarrow \quad BT_1 = (H - \Delta H)(\tan\beta_T - \tan\beta_B) \tag{2.43}$$

By substituting equation (2.43) into equation (2.42), equation (2.44) is derived, that computes the height of the vertical object.

$$h = (H - \Delta H)\left(1 - \frac{\tan\beta_B}{\tan\beta_T}\right) \tag{2.44}$$

The cosine of the angle $\beta_B$ is calculated according to equation (2.45), by applying the law of cosines in the triangle $OB'N'$.

$$\cos\beta_B = \frac{(OB')^2 + (ON')^2 - (B'N')^2}{2(OB')(ON')} \tag{2.45}$$

The tangent of this angle is computed by equation (2.46)

$$\tan\beta_B = \frac{\sqrt{1-\cos^2\beta_B}}{\cos\beta_B} \quad \rightarrow \quad \tan\beta_B = \frac{\sqrt{4(OB')^2(ON')^2 - \left[(OB')^2 + (ON')^2 - (B'N')^2\right]^2}}{(OB')^2 + (ON')^2 - (B'N')^2} \tag{2.46}$$

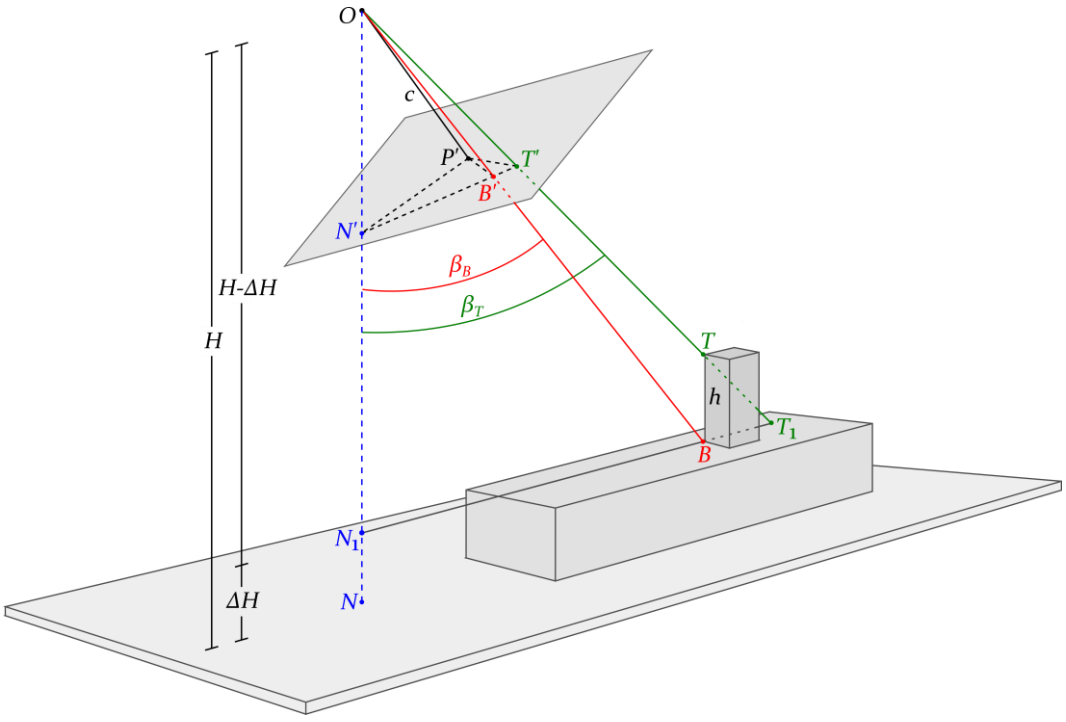


Figure 2-12. Oblique aerial image geometry: calculation of the height of a vertical object

The tangent of the angle $\beta_T$ is calculated in a similar way, according to equation (2.47)

$$\tan\beta_T = \frac{\sqrt{4(OT')^2(ON')^2 - \left[(OT')^2 + (ON')^2 - (T'N')^2\right]^2}}{(OT')^2 + (ON')^2 - (T'N')^2} \tag{2.47}$$

Equations (2.48), (2.49) and (2.50) are derived from the right triangles *OP'B'*, *OP'T'* and *OP'N*, respectively.

$$ON' = \sqrt{c^2 + (P'N')^2} \tag{2.48}$$

$$OB' = \sqrt{c^2 + (P'B')^2} \tag{2.49}$$

$$OT' = \sqrt{c^2 + (P'T')^2} \tag{2.50}$$

By substituting equations (2.46), (2.47), (2.48), (2.49) and (2.50) in equation (2.44), the height of a vertical object is calculated according to equation (2.51). The image coordinates used in equation (2.51) are those corrected from systematic errors (e.g., radial and tangential distortion).

$$h = (H - \Delta H)\left(1 - \frac{\dfrac{\sqrt{4\left(c^2 + (P'B')^2\right)\left(c^2 + (P'N')^2\right) - \left[2c^2 + (P'B')^2 + (P'N')^2 - (B'N')^2\right]^2}}{2c^2 + (P'B')^2 + (P'N')^2 - (B'N')^2}}{\dfrac{\sqrt{4\left(c^2 + (P'T')^2\right)\left(c^2 + (P'N')^2\right) - \left[2c^2 + (P'T')^2 + (P'N')^2 - (T'N')^2\right]^2}}{2c^2 + (P'T')^2 + (P'N')^2 - (T'N')^2}}\right)$$

$$\begin{aligned}
(P'B')^2 &= (x_{B'} - x_{P'})^2 + (y_{B'} - y_{P'})^2 \\
(P'N')^2 &= (x_{N'} - x_{P'})^2 + (y_{N'} - y_{P'})^2 \\
(B'N')^2 &= (x_{N'} - x_{B'})^2 + (y_{N'} - y_{B'})^2 \\
(P'T')^2 &= (x_{T'} - x_{P'})^2 + (y_{T'} - y_{P'})^2 \\
(T'N')^2 &= (x_{N'} - x_{T'})^2 + (x_{N'} - y_{T'})^2
\end{aligned}$$

where

- $x_{B'}$, $y_{B'}$, $x_{T'}$, $y_{T'}$: the image coordinates of the bottom and the top point of the vertical object
- $x_{N'}$, $y_{N'}$: the image coordinates of the nadir point
- $x_{P'}$, $y_{P'}$: the image coordinates of the principal point

(2.51)

$\Delta H > 0$ if the datum plane is lower in elevation than the bottom of the vertical object
$\Delta H < 0$ if the datum plane is higher in elevation than the bottom of the vertical object

### 2.5.2 Horizontal Distances

The calculation of horizontal distances from a single oblique aerial image is based on the computation of ground coordinates in a local coordinate system. The computation of ground coordinates in such a local ground coordinate system (hereinafter referred to as auxiliary ground coordinate system) has been presented and proved by the photogrammetric community (Moffitt and Mikhail, 1980; Wolf, 1983). In this section, for the scope of completeness, the derivation of the

formulas that compute horizontal coordinates in the local auxiliary ground coordinates system is also presented.

The calculation of horizontal coordinates in the auxiliary ground coordinate system requires the use of an auxiliary image coordinate system. For the definition of the auxiliary image coordinate system, it is assumed that the image has been corrected from systematic errors (e.g., lens distortion) or that systematic errors are insignificant and can be ignored. The origin of the auxiliary image coordinate system is located at the nadir point; $y$ axis coincides with the principal line and $x$ axis is perpendicular to it, thus being a horizontal line (Figure 2-13). It has been proved that the coordinates of an image point $A'$ in this system are calculated according to equations (2.52) (Wolf, 1983), where the angle $\varphi$ is a function of the swing angle $s$: $\varphi = s - 180^{o}$.

$$\begin{aligned} x'_{A'} &= x^c_{A'} \cos\varphi - y^c_{A'} \sin\varphi \\ y'_{A'} &= x^c_{A'} \sin\varphi + y^c_{A'} \cos\varphi + c\tan t \end{aligned} \tag{2.52}$$

The coordinates of $A'$ in the image coordinate system centered at the principal point $(x^c, y^c)$ are calculated using its corresponding pixel coordinates and the pixel coordinates of the principal point. Assuming that the image coordinate system in pixels $(x, y)$ has its origin at the top left pixel, the auxiliary image coordinates of a ground point $A$ are calculated by the set of equations (2.53).

$$\begin{aligned} x'_{A'} &= (x_{A'} - x_{P'})\cos\varphi - (y_{P'} - y_{A'})\sin\varphi \\ y'_{A'} &= (x_{A'} - x_{P'})\sin\varphi + (y_{P'} - y_{A'})\cos\varphi + c\tan t \end{aligned} \tag{2.53}$$

The auxiliary ground coordinate system has its origin at the ground nadir point, while $X'$ and $Y'$ axes lie on the same vertical planes with the axes of the auxiliary image coordinate system, being positive in the same directions as them (Moffitt and Mikhail, 1980; Wolf, 1983). Figure 2-14 illustrates the calculation of the auxiliary ground coordinates of a point $A$. $A_1$ is the projection of $A$ on the vertical plane that contains $Y'$ axis; $A'_1$ is the projection of $A'$ on the auxiliary $x'$ axis, thus being the image of $A_1$, and the line $A'_1D'$ is perpendicular to the vertical line from the perspective center. Equation (2.54) is derived using the similar triangles $OA'A'_1 - OAA_1$ and $OD'A' - ON_1A$.

$$\frac{A'A'_1}{AA_1} = \frac{OD'}{ON_1} \rightarrow \frac{|x'_{A'}|}{|X'_A|} = \frac{ON' - D'N'}{H - \Delta H} \tag{2.54}$$

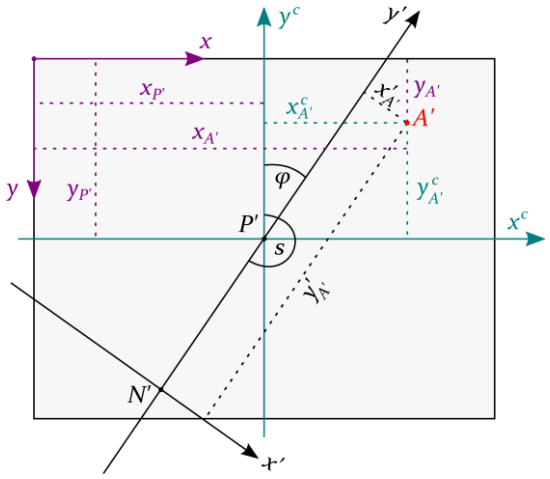


Figure 2-13. Auxiliary coordinate system of an oblique aerial image

The absolute value of the ground $X'$ coordinate of $A$ is calculated according to equation (2.55), which is derived using equation (2.54) and applying trigonometric relations in the right triangles $OP'N'$ and $D'A'_1N'$.

$$\left|X'_A\right| = \frac{(H-\Delta H)\left|x'_{A'}\right|}{\frac{c}{\cos t} - y'_{A'}\sin t} \rightarrow \left|X'_A\right| = \frac{(H-\Delta H)\left|x'_{A'}\right|\cos t}{c - y'_{A'}\sin t\cos t} \tag{2.55}$$

The absolute value of the ground $Y'$ coordinate of $A$ is calculated by equation (2.56) in a similar way, using the similar triangles $OD'A'_1 - ON_1A_1$ and the right triangle $D'A'_1N'$.

$$\frac{\left|y'_{A'}\right|\cos t}{\left|Y'_A\right|} = \frac{\frac{c}{\cos t} - \left|y'_{A'}\right|\sin t}{H-\Delta H} \rightarrow Y'_A = \frac{(H-\Delta H)\left|y'_{A'}\right|\cos^2 t}{c - \left|y'_{A'}\right|\sin t\cos t} \tag{2.56}$$

Thus, the $X'$ and $Y'$ coordinates of a point $A$ in the auxiliary ground coordinate system are calculated through the set of equations (2.57), having the same sign as its auxiliary image coordinates, $x'$ and $y'$, respectively.

$$X'_A = \begin{cases} \left|\frac{(H-\Delta H)x'_{A'}\cos t}{c - y'_{A'}\sin t\cos t}\right| & \text{if } x'_{A'} \geq 0 \\ -\left|\frac{(H-\Delta H)x'_{A'}\cos t}{c - y'_{A'}\sin t\cos t}\right| & \text{if } x'_{A'} < 0 \end{cases} \qquad Y'_A = \begin{cases} \left|\frac{(H-\Delta H)y'_{A'}\cos^2 t}{c - y'_{A'}\sin t\cos t}\right| & \text{if } y'_{A'} \geq 0 \\ -\left|\frac{(H-\Delta H)y'_{A'}\cos^2 t}{c - y'_{A'}\sin t\cos t}\right| & \text{if } y'_{A'} < 0 \end{cases} \tag{2.57}$$

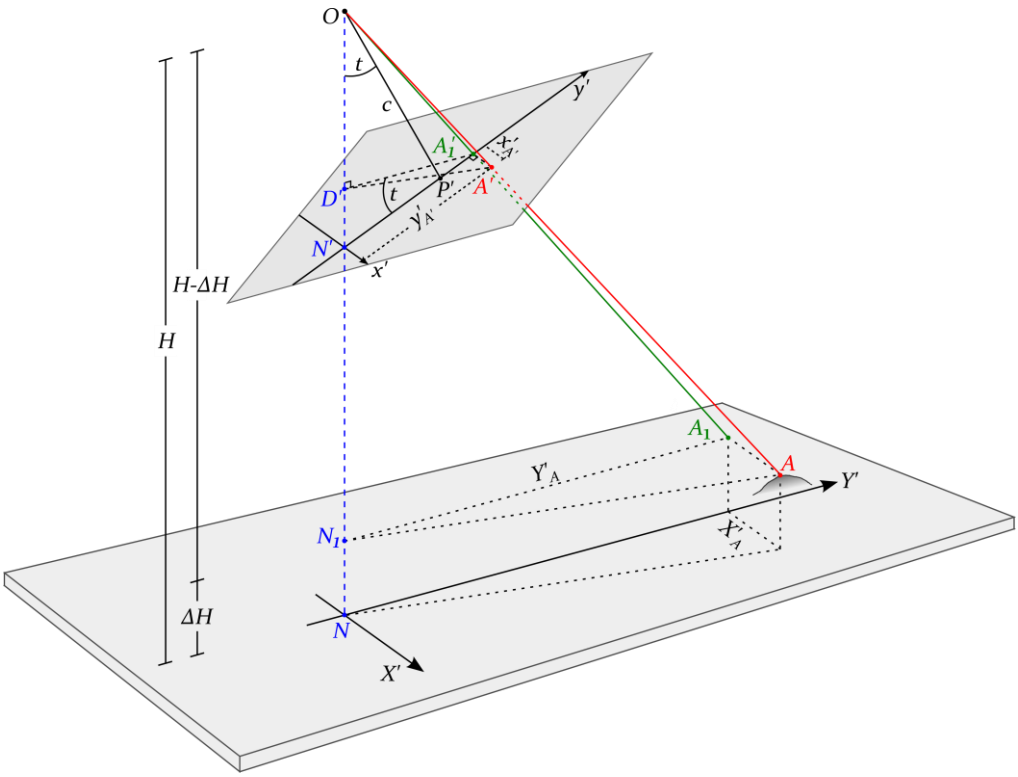


Figure 2-14. Oblique aerial image geometry: calculation of the coordinates of a point in the auxiliary ground coordinate system

If the camera tilt angle *t*, used in the set of equations (2.57) is not known a priori, it may be calculated through equation (2.24), provided that the principal point and camera constant are known and the image nadir point has been determined (e.g., through the method presented in section 6.3 of this dissertation).

The distance between two points *A* and *C* is calculated by equation (2.58), after the determination of the auxiliary ground coordinates of *A* ($X'_A$, $Y'_A$) and *C* ($X'_C$, $Y'_C$).

$$d = \sqrt{\left(X'_C - X'_A\right)^2 + \left(Y'_C - Y'_A\right)^2} \tag{2.58}$$

### 2.5.3 Assessment of Errors

The variables that may affect the accuracy of measurement of a vertical or horizontal distance, due to their values and errors, are the following:

- the flying height;
- the elevation difference $\Delta H$ between the datum plane and the bottom point of the vertical object, for the case of measuring a vertical distance, or the points of the same elevation, for the case of measuring a horizontal distance;
- the image positions of the two points being measured, which determine the length of their distance;
- the camera interior orientation parameters; and
- the position of the nadir point in the image, which depends on the camera tilt angle and the camera constant.

Assuming that:

- the camera interior orientation is known with adequate accuracy (thus setting the errors of the principal point, the camera constant and the distortion coefficients to zero);
- the measurements of the image points are equally accurate (assuming an error of 0.5 pixel in the measured coordinates of each point); and
- the nadir point is accurately determined (assuming an error of 10 pixels in both the *x* and *y* coordinates of the nadir point);

and considering a camera resolution of 60 MP and a pixel size of 6 μm, the standard errors of the measured vertical and horizontal distances were calculated by error propagation for the following four scenarios.

- *Scenario 1*: $\sigma_H$=0.5 m and $\sigma_{\Delta H}$=0.15 m; the flying height *H* is accurately known and the elevation difference $\Delta H$ is either known with high accuracy (if precise elevation data are available for the region under the aerial platform and the region of interest) or set to zero, provided that the ground is almost flat.
- *Scenario 2*: $\sigma_H$=0.5 m and $\sigma_{\Delta H}$=10 m; *H* is accurately known and $\Delta H$ is either roughly estimated (e.g., by using a free program like Google Earth) or set to zero, assuming small variations in ground elevations.
- *Scenario 3*: $\sigma_H$=15 m and $\sigma_{\Delta H}$=0.15 m; *H* is approximately estimated and $\Delta H$ is known with high accuracy or set to zero, provided that the ground is almost flat.
- *Scenario 4*: $\sigma_H$=15 m and $\sigma_{\Delta H}$=10 m. *H* and $\Delta H$ are approximately estimated.

Table 2-1. Standard errors of vertical distances ranging from 3 m to 40 m, measured from an oblique aerial image

| **Standard errors (m) of vertical distances 3 – 40 m** | | | | | |
|---|---|---|---|---|---|
| ***Scenario 1*** **($\sigma_H$=0.5 m, $\sigma_{\Delta H}$=0.15 m)** | ***t*, *c*** / ***H*** | ***t*=15°, *c*=50 mm** | ***t*=15°, *c*=150 mm** | ***t*=35°, *c*=50 mm** | ***t*=35°, *c*=150 mm** |
| | ***H*=200 m** | 0.47-1.10 | 0.03-0.12 | 0.05-0.12 | 0.01-0.11 |
| | ***H*=500 m** | 1.16-1.58 | 0.08-0.11 | 0.11-0.13 | 0.03-0.05 |
| | ***H*=1000 m** | 2.31-2.52 | 0.15-0.17 | 0.23-0.23 | 0.06-0.07 |
| ***Scenario 2*** **($\sigma_H$=0.5 m, $\sigma_{\Delta H}$=10 m)** | ***t*, *c*** / ***H*** | ***t*=15°, *c*=50 mm** | ***t*=15°, *c*=150 mm** | ***t*=35°, *c*=50 mm** | ***t*=35°, *c*=150 mm** |
| | ***H*=200 m** | 0.49-2.28 | 0.15-2.00 | 0.16-2.00 | 0.15-2.00 |
| | ***H*=500 m** | 1.16-1.77 | 0.10-0.81 | 0.13-0.81 | 0.07-0.80 |
| | ***H*=1000 m** | 2.32-2.55 | 0.16-0.43 | 0.23-0.46 | 0.07-0.41 |
| ***Scenario 3*** **($\sigma_H$=15 m, $\sigma_{\Delta H}$=0.15 m)** | ***t*, *c*** / ***H*** | ***t*=15°, *c*=50 mm** | ***t*=15°, *c*=150 mm** | ***t*=35°, *c*=50 mm** | ***t*=35°, *c*=150 mm** |
| | ***H*=200 m** | 0.52-3.19 | 0.23-3.00 | 0.23-3.00 | 0.23-3.00 |
| | ***H*=500 m** | 1.17-1.98 | 0.12-1.20 | 0.15-1.21 | 0.10-1.20 |
| | ***H*=1000 m** | 2.32-2.59 | 0.16-0.62 | 0.24-0.64 | 0.08-0.60 |
| ***Scenario 4*** **($\sigma_H$=15 m, $\sigma_{\Delta H}$=10 m)** | ***t*, *c*** / ***H*** | ***t*=15°, *c*=50 mm** | ***t*=15°, *c*=150 mm** | ***t*=35°, *c*=50 mm** | ***t*=35°, *c*=150 mm** |
| | ***H*=200 m** | 0.54-3.77 | 0.27-3.61 | 0.27-3.61 | 0.27-3.61 |
| | ***H*=500 m** | 1.17-2.14 | 0.13-1.45 | 0.16-1.45 | 0.11-1.44 |
| | ***H*=1000 m** | 2.33-2.62 | 0.16-0.74 | 0.24-0.76 | 0.08-0.72 |

For all the aforementioned scenarios, different combinations of the camera constant (50 mm and 150 mm), the tilt angle (15º and 35º) and the flying height (200 m, 500 m and 1000 m) were applied. The resulting standard errors are presented in Table 2-1 and Table 2-2, for vertical distances ranging from 3 m to 40 m and horizontal distances ranging from 10 m to 90 m, respectively.

The conclusions outlined in the following may be derived using the error propagation results presented in Table 2-1 and Table 2-2.

- For images of similar mean scale, an increase in the camera constant and in the tilt of the camera axis leads to higher accuracy in height measurements.
- The standard error of a horizontal distance is almost independent from the camera constant and the tilt angle.
- For images of similar mean scale, an increase in the error of the flying height and in the error of the elevation difference $\Delta H$ generally increases the error of both vertical and horizontal distances. An increase in the flying height reduces the effect of these errors on the standard error of the distance being measured.
- For images of similar mean scale, the standard error of a greater vertical or horizontal distance is generally worse than the error of a smaller one. An increase in the flying height reduces the effect of the length of the line segment on its standard error. Let the standard error of a vertical or horizontal distance be expressed as a percentage of this distance. For the case of vertical distances, this percentage decreases with an increase in the length of the vertical line segment. For the case of horizontal distances, this percentage remains constant, regardless of the length of the horizontal line segment.

Table 2-2. Standard errors of horizontal distances ranging from 10 m to 90 m, measured from an oblique aerial image

| **Standard errors (m) of horizontal distances 10 – 90 m** | | | | | |
|---|---|---|---|---|---|
| ***Scenario 1*** **($\sigma_H$=0.5 m, $\sigma_{\Delta H}$=0.15 m)** | ***t*, *c*** / ***H*** | ***t*=15°, *c*=50 mm** | ***t*=15°, *c*=150 mm** | ***t*=35°, *c*=50 mm** | ***t*=35°, *c*=150 mm** |
| | ***H*=200 m** | 0.03-0.24 | 0.03-0.24 | 0.03-0.24 | 0.03-0.24 |
| | ***H*=500 m** | 0.04-0.13 | 0.02-0.10 | 0.04-0.11 | 0.02-0.10 |
| | ***H*=1000 m** | 0.07-0.12 | 0.03-0.06 | 0.07-0.10 | 0.03-0.06 |
| ***Scenario 2*** **($\sigma_H$=0.5 m, $\sigma_{\Delta H}$=10 m)** | ***t*, *c*** / ***H*** | ***t*=15°, *c*=50 mm** | ***t*=15°, *c*=150 mm** | ***t*=35°, *c*=50 mm** | ***t*=35°, *c*=150 mm** |
| | ***H*=200 m** | 0.50-4.51 | 0.50-4.51 | 0.50-4.51 | 0.50-4.51 |
| | ***H*=500 m** | 0.20-1.80 | 0.20-1.80 | 0.20-1.80 | 0.20-1.80 |
| | ***H*=1000 m** | 0.12-0.91 | 0.10-0.90 | 0.12-0.91 | 0.11-0.90 |
| ***Scenario 3*** **($\sigma_H$=15 m, $\sigma_{\Delta H}$=0.15 m)** | ***t*, *c*** / ***H*** | ***t*=15°, *c*=50 mm** | ***t*=15°, *c*=150 mm** | ***t*=35°, *c*=50 mm** | ***t*=35°, *c*=150 mm** |
| | ***H*=200 m** | 0.75-6.75 | 0.75-6.75 | 0.75-6.75 | 0.75-6.75 |
| | ***H*=500 m** | 0.30-2.70 | 0.30-2.70 | 0.30-2.70 | 0.30-2.70 |
| | ***H*=1000 m** | 0.17-1.36 | 0.15-1.35 | 0.17-1.35 | 0.15-1.35 |
| ***Scenario 4*** **($\sigma_H$=15 m, $\sigma_{\Delta H}$=10 m)** | ***t*, *c*** / ***H*** | ***t*=15°, *c*=50 mm** | ***t*=15°, *c*=150 mm** | ***t*=35°, *c*=50 mm** | ***t*=35°, *c*=150 mm** |
| | ***H*=200 m** | 0.90-8.11 | 0.90-8.11 | 0.90-8.11 | 0.90-8.11 |
| | ***H*=500 m** | 0.36-3.25 | 0.36-3.25 | 0.36-3.25 | 0.36-3.25 |
| | ***H*=1000 m** | 0.19-1.63 | 0.18-1.62 | 0.19-1.63 | 0.18-0.62 |

## 2.6 Conclusions

In this chapter, basic definitions and geometric relationships for oblique aerial images have been presented. Specifically, the well-established classification of oblique aerial images into low and high oblique ones has been discussed, the camera configurations for capturing datasets containing oblique aerial images have been presented and basic terms associated with oblique aerial images along with their geometric properties and relationships have been outlined. These terms and relationships are necessary for understanding the underlying geometry of oblique aerial images and the methods introduced in the following chapters of this dissertation. Furthermore, mathematical relationships for computing vertical and horizontal distances from a single oblique aerial image have been established and a related error analysis has been presented. While the measurement of horizontal distances by a method similar to the one presented in this chapter has already been outlined by the photogrammetric community, the mathematical model presented in this chapter for the determination of vertical distances, has been proved anew in this dissertation and may be used instead of other well-established relationships for measuring the height of a vertical object. The mathematical models for determining vertical and horizontal distances may be used in combination with the nadir point detection technique, which is presented in chapter 6 of this dissertation, for the establishment of an automated framework for measuring vertical and horizontal distances from a single oblique image with unknown exterior orientation parameters and unavailable GCP measurements.

# Chapter 3
# State-of-the-Art

## 3.1 Introduction

The review of the earlier as well as the recent and ongoing research is a powerful asset for the development of photogrammetry regarding the usage of oblique airborne imagery. The georeferencing, usually in terms of exterior orientation estimation and optionally image orthorectification, or by means of image center or footprint calculation or image planar rectification, in cases that high accuracy is not needed, is generally the first stage of the majority of photogrammetric processes, required for the extraction of metric information from oblique imagery. Hence, the scientific community has focused on georeferencing methodologies, thus providing useful information for further research. In this context, this chapter provides a state-of-the-art mainly on recent research, conducted the last twenty years, on georeferencing of datasets comprising oblique aerial imagery and outlines the most significant research outcomes. It starts with a brief historical background of oblique aerial imagery followed by current research trends and applications using such images. A detailed presentation of state-of-the-art georeferencing approaches is then made, and a critical discussion of the most significant research results finally follows. This extensive literature review provides the general framework, as well as the foundation, for the basic contributions of the present dissertation.

## 3.2 Oblique Aerial Images: From the Nineteenth Century to the Present

In this section, a brief history of oblique aerial images from the nineteenth century to the beginning of the twenty-first century is outlined and a categorization of the literature regarding state-of-the-art georeferencing procedures as well as other research trends and applications using oblique aerial images is presented.

### 3.2.1 Historical Background

Oblique aerial images are not a new data source. Their history dates back to the nineteenth century, when Jacques Mandé Daguerre took the first oblique image (actually one of the first images in general) from the top of a tall building in Paris in 1839 (Mattison, 2007). The same year he introduced the process called "daguerreotype", the first commercially successful photographic process in the history of photography (Aber et al., 2019). Once a photographic technique was established, an aerial platform was needed for taking aerial images. Such platforms available at that time were mainly balloons and kites. Whereas it is known that Gaspar Felix Tournachon took the first aerial photograph of a village in France from a balloon in 1858, the oldest surviving aerial photograph is the oblique acquired from a balloon in Boston by James Wallace Black in 1860 (Colwell, 1997). In the same period, Colonel Aimé Laussedat, known as the "father of photogrammetry", who had been investigating the usefulness of photography in topographic mapping since 1849, experimented with aerial photography using balloons and kites (Georgopoulos, 1982; Mattison, 2007).

The usage of oblique aerial images for the scope of damage assessment has been introduced in early twentieth century by George R. Lawrence, who took such images over San Francisco after the earthquake of 1906 using a kite-mounted camera (Kerle, 2016). Furthermore, multi-lens camera systems providing the possibility for oblique views were introduced this period. For instance, Theodor Scheimpflug created a camera viewing obliquely in eight directions in 1900 and Claudius Aschenbrenner developed a nine-lens camera with one nadir and eight oblique-pointing lenses in 1920 (Jacobsen, 2008), while in the 1930s the Fairchild T-3A five-lens camera systems were built, with one vertical and four oblique-pointing cameras, and were used by the U.S. Geological Survey and the U.S. Army for mapping applications (Talley, 1938).

During World War II, oblique aerial images were being exploited for surveillance and reconnaissance purposes (Nocerino et al., 2012) and much interest was shown towards this kind of images for military applications. Also, high oblique aerial images were being used for the scope of rapid small-scale mapping. Two methods with much historical interest that have been widely used are the Canadian grid method (Canadian Topographical Survey, 1928; Church, 1944) and the trimetrogon method (Fitzgerald, 1944). The first one, was being used for mapping large flat areas of Canada. It involves the graphical construction of a perspective grid on an oblique aerial image depicting an area of low relief and the transfer of outlines and points of interest from this grid to a rectangular grid, the squares of which represent the quadrilaterals distorted due to tilt, superimposed over the image. The trimetrogon method involves use of three cameras equipped with wide angle "metrogon" lenses in a single assembly, among which the middle one points vertically downwards and the other two view obliquely at either side of the flight path corresponding to a tilt angle of 60 degrees. The technique of mapping from this type of imagery required the use of instruments, like the rectoblique plotter or photoangulator, radial intersectors, the stereoblique plotter, the topoangulator and the photoalidade (Lewis, 1945; Thompson, 1952). Other techniques for mapping using oblique aerial imagery through graphical methods without special equipment were also proposed in this period (Desjardins, 1945; Rich, 1947). What is more, the advantages of convergent low oblique aerial images were soon realized and this type of imagery was also used for mapping purposes, with the help of an instrument designed especially for convergent low oblique images, namely the Twinplex (Thompson, 1950; Radlinski, 1952).

Use of oblique images has been reported for lunar mapping during the Apollo missions in the late 1960s and early 1970s (Doyle, 1970; Edmundson et al., 2016). Furthermore, infrared oblique aerial

images have been used for the purpose of detecting active faults (Babcock, 1971). In the 1970s, the use of computers has been introduced for map production out of oblique aerial imagery (Scollar, 1975; Palmer, 1977; Scollar et al., 1977) and applications involving archaeological mapping took place in this decade and in the subsequent years (Palmer, 1976; Smith, 1989; Doneus, 2001). Also, methods for mapping of snow cover patterns (Kirnbauer et al., 1991), image registration (Zheng and Chellappa, 1994) as well as detection and creation of 3D shape descriptions of buildings (Lin and Nevatia, 1995) using oblique aerial images were proposed. However, despite the significant research conducted for oblique aerial image exploitation during the last century and the enormous progress made in the field of photogrammetry, oblique aerial images were mainly used for military applications until the beginning of the twenty-first century, when Pictometry International introduced its digital five-camera system which captures oblique and nadir imagery along with its software, opening the way for civil applications using oblique aerial images.

### 3.2.2 Current Research Trends and Applications

In the following paragraphs, a presentation and classification of current research on georeferencing of oblique aerial imagery is made, followed by a presentation of state-of-the-art applications, apart from georeferencing, that make use of oblique aerial imagery.

#### 3.2.2.1 Georeferencing

Over the last twenty years, significant research has been conducted on the georeferencing of oblique and vertical aerial images, taken mostly from manned aircrafts but also from unmanned aerial vehicles (UAVs). The research focuses on georeferencing of:

- datasets of multi-view oblique and vertical aerial images taken by multi-camera systems (Jacobsen, 2008; Smith et al., 2008; Gerke and Nyaruhuma, 2009; Fritsch et al., 2012; Wiedemann and Moré, 2012; Fritsch and Rothermel, 2013; Karel et al., 2013; Rupnik et al., 2013; Zhao et al., 2014; Jhan et al., 2015; Liu and Shao, 2015; Rau et al., 2015; Rupnik et al., 2015; Gerke et al., 2016; Jacobsen and Gerke, 2016; Moe et al., 2016; Nakada et al., 2016; Ostrowski and Bakuła, 2016; Passini et al., 2016; Sun et al., 2016; Xie et al., 2016; Jiang and Jiang, 2017a; Sun et al., 2017; Toschi et al., 2017; Jiang and Jiang, 2018);
- datasets of multi-view oblique aerial imagery taken by multi-camera systems (Gerke, 2011; Habbecke and Kobbelt, 2012; Tommaselli et al., 2013; Geniviva et al., 2014; Saeedi and Mao, 2014; Ostrowski, 2016; Ostrowski and Bakuła, 2016; Jiang and Jiang, 2017a; Jiang and Jiang, 2018);
- datasets of oblique and vertical aerial imagery taken by a single camera (Stilla et al., 2009; Markelin et al., 2014; Zhou and Liu, 2015);
- datasets of oblique aerial imagery taken by a single camera (Frueh et al., 2004; Ding et al., 2008; Le Besnerais et al., 2008; Wang et al., 2008; Wang and Neumann, 2009; Verhoeven et al., 2012; Karel et al., 2013; Sørensen et al., 2015; Doneus et al., 2016; Ostrowski and Bakuła, 2016; Jiang and Jiang, 2017a; Mertes et al., 2017; Midgley and Tonkin, 2017; Jiang and Jiang, 2018); and
- monocular oblique aerial imagery (Karjalainen et al., 2006; Xiong et al., 2014).

The majority of researchers have dealt with the estimation of the exterior orientation parameters of airborne imagery, whereas a few of them presented rough geotagging (Geniviva et al., 2014; Doneus et al., 2016), image planar rectification (Doneus et al., 2016) and orthorectification

(Verhoeven et al., 2012; Karel et al., 2013; Zhou and Liu, 2015; Mertes et al., 2017; Midgley and Tonkin, 2017) approaches. Most methods implemented until today require the availability of initial georeferencing information from GNSS/INS sensors. However, the methodologies proposed by some researchers either for orientation of oblique aerial image datasets (Karel et al., 2013) or for monocular image orientation (Xiong et al., 2014) do not require data from onboard sensors. Also, the workflows followed by Verhoeven et al. (2012), Tommaselli et al. (2013), Rau et al. (2015), Sørensen et al. (2015), Jacobsen and Gerke (2016), Passini et al. (2016), Midgley and Tonkin (2017) and Mertes et al. (2017) using existing software solutions did not make use of GNSS/INS information. Except for the works performed by Stilla et al. (2009), Zhao et al. (2014), Geniviva et al. (2014), Liu and Shao (2015), Zhou and Liu (2015) and Doneus et al. (2016), who implemented direct georeferencing approaches, the research conducted so far mostly focused on indirect georeferencing procedures.

Some researchers (Karjalainen et al., 2006; Smith et al., 2008; Gerke and Nyaruhuma, 2009; Gerke, 2011; Habbecke and Kobbelt, 2012; Verhoeven et al., 2012; Wiedemann and Moré, 2012; Karel et al., 2013; Rupnik et al., 2013; Geniviva et al., 2014; Xiong et al., 2014; Zhao et al., 2014; Liu and Shao, 2015; Zhou and Liu, 2015; Doneus et al., 2016; Ostrowski and Bakuła, 2016; Sun et al., 2016; Xie et al., 2016; Jiang and Jiang, 2017a; Jiang and Jiang, 2018) focused entirely on the georeferencing part. Other researchers (Frueh et al., 2004; Ding et al., 2008; Jacobsen, 2008; Le Besnerais et al., 2008; Wang et al., 2008; Stilla et al., 2009; Wang and Neumann, 2009; Fritsch et al., 2012; Fritsch and Rothermel, 2013; Tommaselli et al., 2013; Markelin et al., 2014; Saeedi and Mao, 2014; Jhan et al., 2015; Rau et al., 2015; Rupnik et al., 2015; Sørensen et al., 2015; Gerke et al., 2016; Jacobsen and Gerke, 2016; Moe et al., 2016; Nakada et al., 2016; Ostrowski, 2016; Passini et al., 2016; Mertes et al., 2017; Midgley and Tonkin, 2017; Sun et al., 2017; Toschi et al., 2017) presented results and conclusions concerning other parts of photogrammetric procedures as well. They either used existing commercial or free software for their research (Jacobsen, 2008; Smith et al., 2008; Fritsch et al., 2012; Verhoeven et al., 2012; Fritsch and Rothermel, 2013; Tommaselli et al., 2013; Markelin et al., 2014; Saeedi and Mao, 2014; Jhan et al., 2015; Rau et al., 2015; Sørensen et al., 2015; Gerke et al., 2016; Jacobsen and Gerke, 2016; Nakada et al., 2016; Ostrowski, 2016; Ostrowski and Bakuła, 2016; Passini et al., 2016; Mertes et al., 2017; Midgley and Tonkin, 2017) or developed their own solutions for the whole georeferencing stage (Frueh et al., 2004; Karjalainen et al., 2006; Ding et al., 2008; Le Besnerais et al., 2008; Wang et al., 2008; Gerke and Nyaruhuma, 2009; Stilla et al., 2009; Wang and Neumann, 2009; Gerke, 2011; Habbecke and Kobbelt, 2012; Karel et al., 2013; Geniviva et al., 2014; Saeedi and Mao, 2014; Xiong et al., 2014; Zhao et al., 2014; Zhou and Liu, 2015; Doneus et al., 2016; Gerke et al., 2016; Sun et al., 2016; Xie et al., 2016; Jiang and Jiang, 2017a; Sun et al., 2017; Jiang and Jiang, 2018) or for part of it (Wiedemann and Moré, 2012; Rupnik et al., 2013; Liu and Shao, 2015; Rupnik et al., 2015; Moe et al., 2016; Toschi et al., 2017), in combination with existing software. Most researchers who used oblique aerial images for their experiments regarding indirect orientation performed aerial triangulation techniques (Jacobsen, 2008; Le Besnerais et al., 2008; Smith et al., 2008; Gerke and Nyaruhuma, 2009; Gerke, 2011; Fritsch et al., 2012; Wiedemann and Moré, 2012; Fritsch and Rothermel, 2013; Tommaselli et al., 2013; Jhan et al., 2015; Liu and Shao, 2015; Sørensen et al., 2015; Jacobsen and Gerke, 2016; Nakada et al., 2016; Passini et al., 2016; Sun et al., 2016; Xie et al., 2016). Other researchers implemented structure from motion (SfM) approaches (Verhoeven et al., 2012; Karel et al., 2013; Rupnik et al., 2013; Rau et al., 2015; Rupnik et al., 2015; Ostrowski, 2016; Ostrowski and Bakuła, 2016; Jiang and Jiang, 2017a; Mertes et al., 2017; Midgley and Tonkin, 2017; Sun et al., 2017; Toschi et al., 2017; Jiang and Jiang, 2018), while Markelin et al. (2014), Gerke et al. (2016) and Moe et al. (2016) performed both SfM and aerial triangulation procedures. Finally,

techniques of matching image features with features extracted from existing 3D models (Frueh et al., 2004; Ding et al., 2008; Wang et al., 2008; Wang and Neumann, 2009), vector maps (Karjalainen et al., 2006; Habbecke and Kobbelt, 2012) and orthoimages (Geniviva et al., 2014), image registration (Saeedi and Mao, 2014) and vanishing point detection techniques (Ding et al., 2008; Habbecke and Kobbelt, 2012) for georeferencing purposes have also been proposed.

Several multi-camera systems have been used by researchers who implemented georeferencing procedures during the last fifteen years, as summarized in the following:

- multi-camera systems consisting of five cameras adopting the Maltese-cross configuration, like the one adopted by the Pictometry system (Smith et al., 2008; Gerke, 2011; Habbecke and Kobbelt, 2012; Geniviva et al., 2014; Saeedi and Mao, 2014), the MIDAS 5 multi-camera system (Jacobsen, 2008; Rupnik et al., 2013; Rupnik et al., 2015; Passini et al., 2016), the SWDC-5 multi-camera system (Xiong et al., 2014; Xie et al., 2016), the Leica RCD30 Penta oblique camera system (Zhao et al., 2014; Sun et al., 2016; Xie et al., 2016), the UltraCam Osprey I system (Rupnik et al., 2015), the IGI Penta-DigiCam system (Jacobsen, 2008; Gerke et al., 2016; Nakada et al., 2016; Ostrowski, 2016; Ostrowski and Bakuła, 2016; Xie et al., 2016), the UltraCam Osprey Prime camera system (Moe et al., 2016; Toschi et al., 2017), a system composed of five SONY A850 DSLR cameras (Rau et al., 2015), a system composed of five Sony NEX-7 cameras (Jiang and Jiang, 2017a; Jiang and Jiang, 2018), the TOPDC-5 penta-view system (Sun et al., 2017) and a self-developed system comprising commercial off-the-shelf cameras (Liu and Shao, 2015); the Maltese-cross configuration was the most widely used in the experiments;
- a three-camera system with one nadir camera and two oblique ones in a rotating sensor head able to rapidly rotate by 90 degrees in order to achieve a Maltese-cross ground coverage too, i.e., the Aerial Oblique System (Wiedemann and Moré, 2012);
- a two-camera system with one camera pointing vertically and one obliquely, i.e., the FLI-MAP-400 system (Gerke and Nyaruhuma, 2009);
- a dual head camera system comprising two obliquely-pointing cameras (Tommaselli et al., 2013; Jiang and Jiang, 2017a; Jiang and Jiang, 2018);
- a system of four oblique cameras in a block configuration pointing forward, backwards, left and right without overlap, i.e., IGI Quattro-DigiCAM Oblique, together with a block configuration of four near-nadir cameras stitched together into a large-format image, i.e., IGI Quattro-DigiCAM (Fritsch et al., 2012; Fritsch and Rothermel, 2013); and
- a self-developed system comprising five consumer grade cameras composed as a 1×5 array (Jhan et al., 2015).

#### 3.2.2.2 Other Applications

Apart from the research on georeferencing of oblique aerial imagery, which will be discussed in detail in section 3.3, several researchers have focused on multi-view oblique and nadir airborne images or solely oblique aerial images, by proposing methodologies for various applications, as outlined in the following:

- texture mapping (Frueh et al., 2004; Børgesen, 2008; Wang et al., 2008; Stilla et al., 2009; Rau and Chu, 2010; Frommholz et al., 2015; Frommholz et al., 2016; Iwaszczuk and Stilla, 2016; Kang et al., 2016);
- 3D building reconstruction (Panday and Gerke, 2011; Frommholz et al., 2015; Frommholz et al., 2016) and urban modelling (Haala et al., 2015);

- dense image matching (Le Besnerais et al., 2008; Gerke, 2009; Fritsch et al., 2012; Zhang, 2013; Haala et al., 2015; Zhang et al., 2015b; Sun et al., 2017) and quasi-dense image matching (Yao et al., 2014);
- DSM generation (Le Besnerais et al., 2008);
- building extraction (Xiao et al., 2010; Xiao, 2012; Xiao et al., 2012; Nex et al., 2013; Xiao, 2013; Gerke et al., 2014; Frommholz et al., 2015; Vetrivel et al., 2015; Frommholz et al., 2016);
- verification of buildings (Nyaruhuma et al., 2012; Nyaruhuma, 2013);
- damage assessment (Gerke and Kerle, 2011; Vetrivel et al., 2015; Vetrivel et al., 2016; Tu et al., 2017; Vetrivel et al., 2017);
- image matching (Weith-Glushko and Salvaggio, 2005; Huachao et al., 2012; Mao, 2013; Xiao et al., 2013; Saeedi and Mao, 2014; Hu et al., 2015; Hu et al., 2016; Jiang and Jiang, 2017b; Onyango et al., 2017);
- matching of point clouds (Zhang, 2014);
- multi-head camera calibration (Tommaselli et al., 2013);
- generation of virtual images through rectification, registration and mosaicking of images simultaneously acquired by multi-camera systems (Tommaselli et al., 2013; Jhan et al., 2015);
- monoplotting and/or single image metric exploitation (Höhle, 2008; Molinari et al., 2014; Murtiyoso et al., 2014; Xiong et al., 2014);
- 3D scene classification (Gerke and Xiao, 2013; Rau et al., 2015; Sun et al., 2017);
- facade detection, partitioning and reconstruction (Zhang et al., 2015a);
- validation of vector data (Mishra et al., 2008);
- data fusion, e.g., fusion with LiDAR data (Prandi et al., 2008), 3D models (Frueh et al., 2004; Ding et al., 2008; Wang et al., 2008; Wang and Neumann, 2009), cadastral maps (Habbecke and Kobbelt, 2012), point clouds for 4D reconstruction of lighting poles (Lin and West, 2014) and SAR data (Schack et al., 2016; Fagir et al., 2017); and
- detection of trees (Lin et al., 2015).

Also, existing free or commercial software packages have been evaluated using multi-perspective oblique and vertical aerial imagery or solely oblique aerial imagery for many applications, as summarized in the following:

- 3D modelling (Hamruni, 2010; Aicardi et al., 2016; Brocks and Bareth, 2016; Casella and Franzini, 2016; Madani et al., 2016; Moe et al., 2016);
- 3D building reconstruction (Yalcin and Selcuk, 2015; Moe et al., 2016; Remondino et al., 2016; Toschi et al., 2017);
- dense point cloud and/or DSM creation (Wackrow and Chandler, 2011; Harwin and Lucieer, 2012; Fritsch and Rothermel, 2013; Deuber et al., 2014; Markelin et al., 2014; Sørensen et al., 2015; Mertes et al., 2017; Midgley and Tonkin, 2017);
- detection of facades (Hsu et al., 2012), roofs, roads, trees and grass (Rau et al., 2015) as well as extraction of windows on building facades (Frommholz et al., 2015; Frommholz et al., 2016) through object-based image analysis;
- calibration of airborne multi-camera systems (Jacobsen, 2008; Liu and Shao, 2015; Jacobsen and Gerke, 2016; Kemper et al., 2016; Passini et al., 2016);
- analysis of performance of commercial oblique cameras (Soler et al., 2016); and
- glacier monitoring (Gleitsmann and Kappas, 2004; Gleitsmann and Kappas, 2006; Mertes et al., 2017; Midgley and Tonkin, 2017).

Furthermore, experiences gained from the use of oblique aerial imagery have been reported in literature for the verification of the potential of their use by national mapping agencies (Remondino et al., 2016) as well as for mapping purposes by public administrations or private customers (Rupnik et al., 2015; Moe et al., 2016), for land administration (Lemmens et al., 2007), cultural heritage documentation (Höhle, 2013), public surveys (Nakada et al., 2016) as well as measurements in urban environments (Ostrowski, 2016).

## 3.3 Georeferencing Approaches

In this section, the recent research concerning the georeferencing of oblique as well as oblique and vertical airborne imagery is presented and categorized depending on the kind of imagery involved, as illustrated in Figure 3-1.

### 3.3.1 Exterior Orientation Estimation via Existing Software

In the following paragraphs, the state-of-the-art approaches for the estimation of the exterior orientation of datasets including oblique airborne imagery are outlined. Firstly, the approaches that use existing aerial triangulation software for bundle block adjustment are outlined, followed by those making use of SfM software. The works conducted using both existing software and research solutions, either independently or in combination, follow. The software used for the task of exterior orientation calculation of datasets including oblique aerial imagery by the works outlined in this chapter, along with their frequency of use, are illustrated in Figure 3-2. These software packages were used either exclusively for the scope of image orientation, like the works presented in this section, or for accuracy assessment of the research methodologies that will be presented in the subsequent sections.

Smith et al. (2008) showed that the simultaneous orientation of oblique and vertical aerial images may yield satisfactory results, by presenting the conclusions derived from different photogrammetric aerial triangulation scenarios performed using commercial software. Initial georeferencing information and numerous GCPs were used as input data for the experiments performed, which include the orientation of (i) solely vertical images; (ii) solely oblique images; and (iii) combined vertical and oblique images taken either by a multi-view camera system or by different camera systems. The results indicated that the quality was improved by the introduction of oblique images into the block, as they provide strong intersection angles at the measured points.

Jacobsen (2008) performed aerial triangulation scenarios of vertical and oblique aerial images taken by a multi-camera system through bundle adjustment, both with and without using GPS measurements and internal system calibration parameters via aerial triangulation software. One major problem that arose during the simultaneous adjustment of vertical and oblique imagery was the failure of automatic aerial triangulation due to failure of the image matching process and the resulting need for manual measurement of tie points. Nevertheless, classical aerial triangulation software packages, like the one used by the author, may successfully perform a non-automatic bundle block adjustment of datasets including oblique aerial images, even with a small number of GCPs, as each tie point is visible in a large number of images. The author concluded that both GPS/INS information and internal system calibration are essential for successful orientation of oblique and vertical aerial imagery taken by a multi-camera system with less manual work.

Fritsch et al. (2012) and Fritsch and Rothermel (2013) computed the exterior orientation parameters of datasets of oblique and stitched large format vertical aerial imagery through existing software via an integrated sensor orientation solution using GNSS and IMU data. The oblique images were collected by a multi-view system comprising four oblique pointing cameras without overlap; each nadir image was generated by geometric and radiometric stitching from four near vertical images with small overlap collected by a multi-camera system at a different flight. The authors concluded that the simultaneous adjustment of oblique and stitched vertical aerial images handled as images

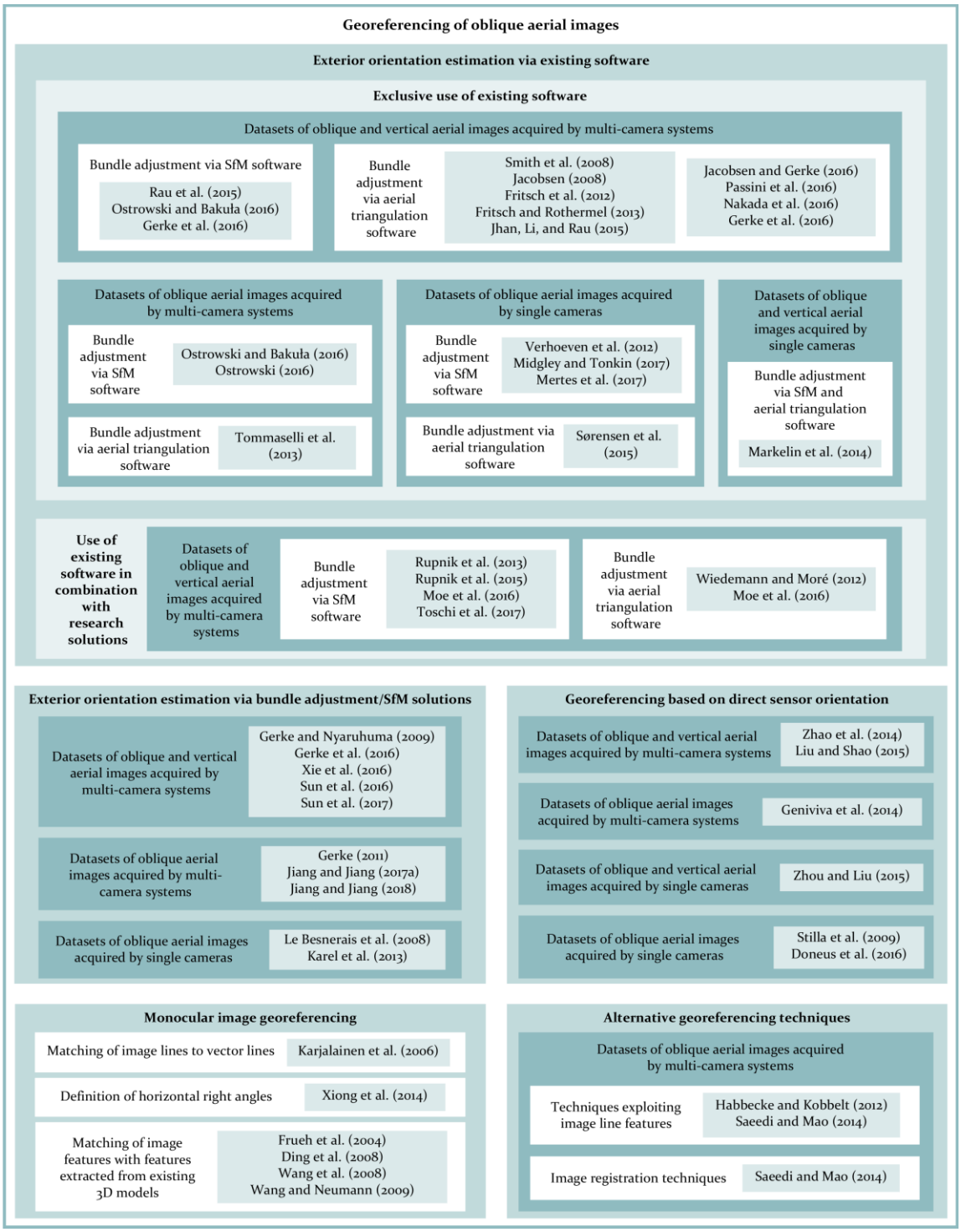


Figure 3-1. Categorization of recent research on georeferencing of oblique as well as oblique and vertical aerial images

taken by a single camera yields satisfactory exterior orientation results for the purposes of dense image matching.

Tommaselli et al. (2013) used a commercial software package to perform self-calibrating bundle adjustment scenarios of virtual images generated from oblique aerial imagery taken by a multi-camera system after system calibration, image rectification, image registration as well as fusion and brightness adjustment. The obtained results proved to be similar to those from a conventional frame camera with a single sensor. Similarly to them, Jhan et al. (2015) used commercial software for performing aerial triangulation of stitched large-format images, generated after registration and mosaicking of oblique and vertical images acquired by a multi-camera system. Contrary to the mosaicking approach of Tommaselli et al. (2013), which was based on corresponding points in the overlapping area of the oblique imagery detected through area-based matching and refined by least-squares matching, the mosaicking by Jhan et al. (2015) was performed using a modified projective transformation model estimated using the interior orientation parameters of each camera and the relative orientation parameters between the oblique cameras and the nadir one, calculated through camera calibration in an indoor calibration field. Comparisons between aerial triangulation scenarios of separate images and mosaicked images revealed similar accuracy, showing that the mosaicked images may be used for topographic mapping. However, the standard deviation of unit weight of the block adjustment is slightly increased when mosaicked images are used, due to error propagation.

Sørensen et al. (2015) performed bundle block adjustment of oblique aerial images acquired by a single calibrated camera using commercial software. The relative orientation of the imagery was performed through automatic tie point extraction and the final bundle adjustment was performed after manual measurement of GCPs in the imagery. The methodology followed yielded satisfactory exterior orientation results that can be used for generation of terrain models of geological outcrops.

Jacobsen and Gerke (2016) performed self-calibrating bundle block adjustment with additional parameters for datasets of oblique and vertical aerial imagery taken by a multi-camera system for the scope of sub-camera calibration, without fixing the sub-cameras to the same projection center nor using GNSS/INS information. It was concluded that the bundle adjustment with self-calibration

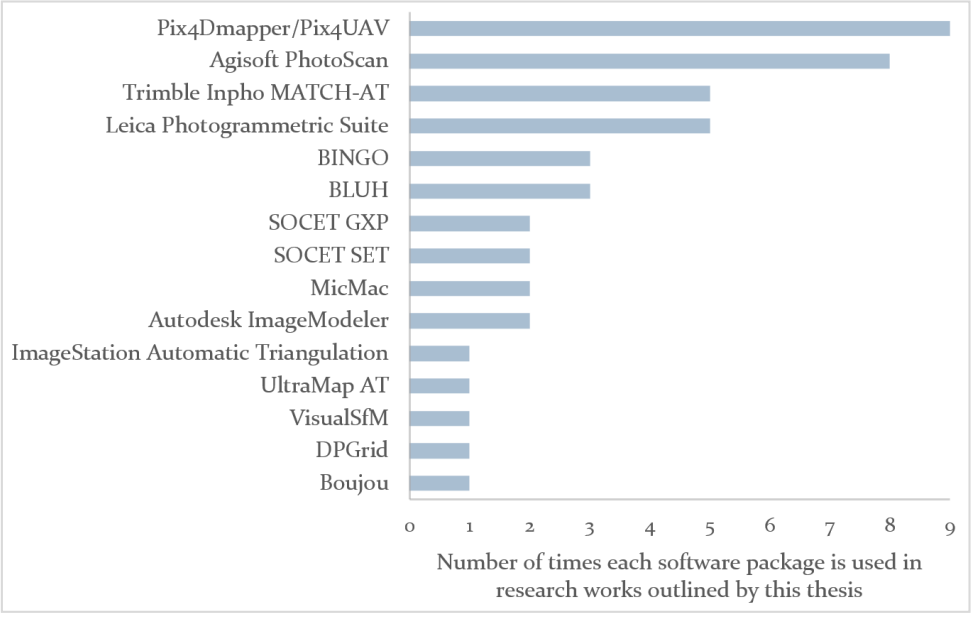


Figure 3-2. Use of existing software packages for exterior orientation estimation of datasets including oblique aerial imagery

and additional parameters, checked by remaining systematic errors, yields satisfactory results for such blocks of imagery. However, the assumption for a different projection center of each sub-camera weakens the block adjustment. The same conclusions were reported by Passini et al. (2016), who adopted a similar procedure using the same software packages both for image matching and self-calibrating bundle adjustment for the orientation of different image datasets, collected by a different multi-camera system.

Nakada et al. (2016) used commercial software for the calculation of the exterior orientation of imagery acquired by a multi-camera system accompanied by GNSS/IMU sensors via different block configuration scenarios using GCPs. These scenarios include the adjustment of (i) vertical images only; (ii) forward view images only; (iii) left view images only; (iv) oblique images only (forward, backward, left and right); (v) oblique images after the separate orientation of the vertical images; and (vi) multi-view oblique and vertical images (simultaneous orientation). Best results were derived in blocks (i), (v) and (vi), showing that the integration of vertical imagery improved the accuracy of the adjustment.

Whereas the aforementioned works were performed using existing software created especially for aerial triangulation purposes and only the tie points were automatically generated by some authors (Jhan et al., 2015; Jacobsen and Gerke, 2016; Passini et al., 2016) via SfM software, Verhoeven et al. (2012), Rau et al. (2015), Ostrowski and Bakuła (2016), Ostrowski (2016), Midgley and Tonkin (2017) and Mertes et al. (2017) used solely SfM software for the task of orientation estimation. Verhoeven et al. (2012) focused on the production of orthomosaics using uncalibrated oblique aerial imagery taken by single cameras for different case studies. The relative camera poses were initially estimated using a classical SfM workflow and, along with the subsequently computed dense point cloud, were then transformed into the earth coordinate frame using GCPs via a seven-parameter similarity transformation, for the scope of orthomosaic generation. The entire process was conducted using existing software with satisfactorily accurate results for archaeological large-scale mapping purposes. However, not optimal image alignments were reported, whereas the processing was proved to be computer resource intensive without achieving full automation, as manual selection and measurement of GCPs as well as visualization of images and parameter tuning took place.

Rau et al. (2015) used a commercial SfM software package for the orientation of multi-view oblique and vertical aerial images acquired by a multi-camera system using GCPs and reported satisfactory interior and exterior orientation results for the subsequent process of dense point cloud generation for classification purposes.

Ostrowski and Bakuła (2016) showed that the orientation of oblique and vertical aerial images as well as only oblique aerial images taken by a multi-camera system, accompanied with GNSS/INS data, can be carried out effectively in commercial SfM software. Contrary to the observation by Nakada et al. (2016) that the addition of vertical images improves the accuracy of bundle block adjustment in aerial triangulation software, Ostrowski and Bakuła (2016) concluded that the introduction of vertical images does not substantially affect the precision and effectiveness of the orientation via SfM software. A similar workflow for the exterior orientation of multi-view oblique aerial images acquired by a multi-camera system using direct measurements from onboard sensors (GNSS/INS) and GCPs via commercial SfM software was followed by Ostrowski (2016), with the scope of verifying the accuracy of measurements in oblique aerial imagery. The bundle adjustment yielded satisfactory sub-pixel accuracies, which were almost equal in all three dimensions, while

the accuracy of measurements of both planimetric coordinates and distances lied on the level of two ground pixels.

Midgley and Tonkin (2017) and Mertes et al. (2017) used the same commercial SfM software for the task of orienting oblique aerial images depicting glaciers with the ultimate goal of producing point clouds, digital elevation models (DEMs) and orthorectified imagery using GCPs. The unknown interior orientation parameters were calculated through self-calibration. The orientation results revealed the potential of using historical oblique aerial imagery for creating reliable DEMs and orthoimages of glaciated regions. The horizontal and vertical errors that may occur are usually due to poor image quality, image resolution, and a limited spread of GCPs, according to Midgley and Tonkin (2017).

Markelin et al. (2014) used both SfM and conventional photogrammetric triangulation software for the exterior orientation estimation of blocks of hyperspectral UAV images acquired by a single spectral camera. Their proposed workflow is based on the determination of approximate exterior orientation values via free SfM software and the optional calculation of refined image orientation through bundle adjustment with self-calibration via typical photogrammetric software. This workflow was implemented for the combined orientation of nadir and oblique image blocks depicting different areas and yielded satisfactory results. Some target areas allowed the measurement of GCPs, while on other areas, like forests, the authors relied on direct georeferencing results. Whereas the most accurate solution would be to solve the exterior orientation of each band of each data cube separately, as each band in the data cube has a slightly different position and orientation because they are exposed sequentially, the authors calculated the orientation of three different bands in a unique solution for the scope of efficiency; then, the orientations of the other bands were interpolated based on the flight trajectory of the reference bands using the known time difference between the bands.

Whereas the aforementioned experiments were conducted solely via existing software solutions, Gerke et al. (2016) reported the use of both existing SfM and aerial triangulation software as well as research solutions independently for the task of orientation estimation. They presented the results of experiments performed using the ISPRS/EuroSDR benchmark for multi-platform photogrammetry (Nex et al., 2015) in terms of tie point matching, influence of oblique imagery onto the bundle adjustment, role of image overlap and GCP distribution. The experiments involved tests via two software packages using the entire datasets of images and GCPs as well as tests using the subset of images and GCPs released to the participants of the benchmark, for comparing the performance of different bundle block adjustment approaches. Among the participants' tests, both research and commercial systems were used, proving that the commercial ones perform very well compared to research solutions. As far as the latter are concerned, some points presented by the research community are worth mentioning: (i) feature points can be extracted in virtual images generated from the original ones by affine transformation; (ii) the relative orientation of the cameras of a multi-camera system should not be used in the bundle block adjustment, in case of unknown synchronization accuracy of the cameras; (iii) the oblique and vertical airborne images can be geometrically rectified using initial exterior orientation parameters and a rough DEM, in order to alleviate the affine deformation and improve the efficiency of image matching; (iv) initial exterior orientation parameters for the oblique images can be obtained from the calibrated relative orientation parameters (rotation and translation) between vertical and oblique images and the exterior orientation parameters of the vertical images; and (v) a weighting strategy for image observations in bundle adjustment can be adopted, depending on the size and shape of the pixel projected on the ground.

Wiedemann and Moré (2012), Rupnik et al. (2013), Rupnik et al. (2015), Moe et al. (2016) and Toschi et al. (2017) used existing software in combination with self-developed solutions for the estimation of the exterior orientation parameters of datasets of oblique and vertical aerial imagery acquired by multi-camera systems. Wiedemann and Moré (2012) discussed five orientation strategies for oblique and vertical airborne images acquired by a calibrated multi-camera system mechanically connected with a GPS/IMU system. The first approach was the aerial triangulation of stitched image triples, which was discarded due to the small overlapping zone between the images and the resulting accuracy loss. The second approach included the aerial triangulation of all the vertical and oblique images via the commercial classical photogrammetric aerial triangulation software also used by Smith et al. (2008) and Jacobsen (2008) for orientation of vertical and oblique aerial images acquired by a multi-camera system; this approach resulted in a lot of erroneous matching results, as also confirmed by the aforementioned authors, who performed manual tie point extraction. The third approach consisted of two steps, that is, (i) the aerial triangulation of the vertical images via another aerial triangulation software package, also used by other researchers (Fritsch et al., 2012; Fritsch and Rothermel, 2013; Moe et al., 2016; Nakada et al., 2016) and (ii) a mathematical transformation of the oblique ones using the eccentricity of the oblique cameras, via a developed software solution. However, minor residual errors in the orientation of the vertical images may cause serious errors in the orientation of the oblique images. The fourth approach, which significantly reduces the number of unknowns, was based on an extended functional model for bundle block adjustment of all images using the relative camera orientations, considering the mechanical connection between vertical and oblique images. This approach assumes that the multi-camera system remains in a horizontal position. The fifth approach was the orientation of the images using a-priori knowledge provided by DTMs, DSMs and orthoimages for the improvement of the image matching process, by concentrating the tie points to well-suited areas of the images. In conclusion, the third approach seems to yield satisfactory results for most applications, in case of a stable mechanical connection between the cameras of the multi-view system, yet it still requires manual work. Approaches using a-priori knowledge from external sources (e.g., DTMs/DSMs, orthoimages) for improving image matching strategies as well as an extended mathematical model for bundle adjustment have to be established.

Rupnik et al. (2013) presented an approach relying on a developed tool that analyses image connectivity and existing open-source SfM software for the simultaneous orientation of large datasets of oblique and vertical airborne images provided with GNSS/IMU data, thus eliminating the need for ground control. The methodology followed by the developed tool relied on three main steps, that is, (i) the construction of an image connectivity graph containing links between images based on their overlapping area and the camera viewing direction, using the direct georeferencing information from onboard sensors; (ii) SIFT feature extraction (Lowe, 2004) and image matching, performed via open-source software using the image pairs from the connectivity graph; and (iii) the creation of an image concatenation queue for the detection of triplets of images within the connectivity graph for computing approximate orientations and assigning to all triplets a position in the concatenation order. The relative orientation was performed via existing open-source software and the absolute orientation exploited the GNSS information, without taking into account the IMU data nor using GCPs. The main contributions of this approach are the image connectivity graph, which, according to the authors, reduces the time required for image matching by 15-20 times, as well as the concatenation queue, which succeeded in linking all images of the tested block. Hence, the authors showed that open-source software can be combined with self-developed solutions for more effective orientation of multi-view image datasets.

Rupnik et al. (2015) investigated the precision and accuracy of 3D points triangulated from diverse multi-camera systems by conducting simulated experiments and performing real case studies. The simulated scenarios investigated the precision of the triangulated points, the productivity and completeness of the reconstructing scene depending on sensor size, forward and side overlap, camera focal length and field of view. The tests were performed using real multi-view image datasets via existing open-source software, exploiting the processing pipeline presented by Rupnik et al. (2013). The performed tests investigated how the accuracy of the triangulated 3D points is affected by the introduction of oblique imagery, the self-calibration, the spatial distribution of GCPs and the GNSS support. Significant conclusions have arisen from their work and are presented in section 3.4 of this chapter.

Moe et al. (2016) estimated the exterior orientation parameters of multi-perspective oblique and vertical images taken by a multi-camera system using a self-developed forward ray intersection tool to analyze the quality of GPS/IMU observations and calibration parameters as well as commercial SfM and aerial triangulation software packages for the process of aerial triangulation. The multi-camera system used was considered as a rigid assembly of cameras, rather than a system of five independent cameras through the usage of the relative orientation parameters (offsets, misalignments) from the camera calibration certificate of oblique views. In all tests, the initial interior orientation parameters as well as additional parameters of all sensors were computed by self-calibration and all software packages delivered similar aerial triangulation results. The strategy presented by Moe et al. (2016) for exterior orientation estimation using an existing SfM software package and a self-developed tool to extend the aerial triangulation results of the nadir images to include the oblique views by means of the relative orientation parameters was also followed by Toschi et al. (2017).

Taking into consideration the aforementioned works that were conducted using existing software solutions, it appears that the selection of the software is crucial for the orientation task when oblique aerial images are involved. As far as image datasets acquired by multi-camera systems are concerned, the approach followed by Smith et al. (2008) and Jacobsen (2008) is not advised, as the aerial triangulation software used failed in the automatic image matching process and manual work was needed, which makes the whole process both less productive and time consuming. On the other hand, the image matching algorithms implemented in the aerial triangulation software used by Fritsch et al. (2012), Fritsch and Rothermel (2013), Nakada et al. (2016) and Moe et al. (2016) automatically yield successful results for multi-view oblique and vertical aerial images. Similarly, the tie point extraction algorithm implemented by the software used by Sørensen et al. (2015) provided satisfactory results for oblique aerial images. The use of SfM software either only for the matching process followed by bundle block adjustment through aerial triangulation software (Jhan et al., 2015; Gerke et al., 2016; Jacobsen and Gerke, 2016; Passini et al., 2016) or for the whole orientation estimation task (Verhoeven et al., 2012; Rupnik et al., 2013; Rau et al., 2015; Rupnik et al., 2015; Sørensen et al., 2015; Gerke et al., 2016; Moe et al., 2016; Ostrowski, 2016; Ostrowski and Bakuła, 2016; Mertes et al., 2017; Midgley and Tonkin, 2017; Toschi et al., 2017) or for computation of initial exterior orientation parameters refined by the solution of a conventional photogrammetric workstation (Markelin et al., 2014) yields automatic image matching results in all cases. Thus, it was proved that existing SfM software solutions, which are optimized for terrestrial and UAV imagery, can effectively compute the exterior orientation parameters of datasets including oblique aerial images as well. Furthermore, it was shown that an aerial triangulation of virtual images generated from imagery acquired by a multi-camera system is possible through existing software solutions (Tommaselli et al., 2013; Jhan et al., 2015). In all cases, except for the approaches by Verhoeven et al. (2012), Tommaselli et al. (2013), Rau et al. (2015), Sørensen et al.

(2015), Passini et al. (2016), Midgley and Tonkin (2017) and Mertes et al. (2017) initial georeferencing information acquired by onboard sensors was used for conducting either an integrated sensor orientation with additional ground control (Jacobsen, 2008; Smith et al., 2008; Jhan et al., 2015; Jacobsen and Gerke, 2016; Moe et al., 2016; Nakada et al., 2016; Toschi et al., 2017) or without GCPs (Fritsch et al., 2012; Wiedemann and Moré, 2012; Fritsch and Rothermel, 2013), or a GNSS-supported sensor orientation without GCPs (Rupnik et al., 2013; Rupnik et al., 2015) or with the use of GCPs (Markelin et al., 2014; Gerke et al., 2016; Ostrowski, 2016; Ostrowski and Bakuła, 2016). On the contrary, Verhoeven et al. (2012), Tommaselli et al. (2013), Sørensen et al. (2015), Passini et al. (2016), Midgley and Tonkin (2017) and Mertes et al. (2017) presented indirect sensor orientation solutions using GCPs. Also, it was shown that the use of existing software may be combined with self-developed solutions for more efficient and fast orientation estimation (Rupnik et al., 2013; Rupnik et al., 2015).

### 3.3.2 Exterior Orientation Estimation via Bundle Adjustment and SfM Solutions

Whereas the aforementioned works were conducted entirely or partially using existing software solutions for exterior orientation estimation of datasets including oblique airborne imagery, the research works presented in this section were performed without using existing software for the task of exterior orientation calculation. Le Besnerais et al. (2008) implemented a workflow for the calculation of the exterior orientation parameters of small-baseline oblique aerial images accompanied with GNSS/INS data without GCPs. Their workflow is based on interest point tracking using the Kanade-Lucas-Tomasi tracker (Shi and Tomasi, 1994) and joint refinement of exterior orientation parameters along with the 3D coordinates of some scene points through bundle adjustment, using a sparse Levenberg Marquardt optimization (Lourakis and Argyros, 2004). In order to deal with potential outliers, the optimization was conducted twice, that is, firstly with all tracked features and then after a step for removing outliers. As it was desired that the subsequently generated height map is defined in the geometry of one of the images taken as the reference view, the authors used a diagonal covariance matrix with near-zero values for the reference image and larger standard deviations for the perspective center coordinates and camera orientation angles of the other images. Satisfactory standard deviations for camera exterior orientation parameters after bundle adjustment were reported for being used for height map computation.

Gerke and Nyaruhuma (2009) introduced the manual incorporation of geometric scene constraints into the triangulation of oblique and vertical aerial imagery, using approximations for the exterior orientation parameters and GCPs. The proposed constraints, i.e., horizontal lines, vertical lines and right angles, which are defined at building facades depicted in oblique aerial imagery, have to be manually defined in at least two images. The constraints along with the GCPs are integrated into the adjustment as soft constraints, by formulating pseudo-observations with high weights in the stochastic model, thus forcing the unknowns (i.e., the points defining the constraints) to satisfy the corresponding geometric conditions. The proposed approach was tested using a small block of oblique and vertical imagery acquired by a multi-camera system carrying two cameras via different configurations and the results were compared to commercial software packages, revealing a slightly better overall performance of the proposed method, which can also be applied in a free net adjustment.

The above algorithm concerning the introduction of scene constraints into the bundle adjustment is discussed in more detail by Gerke (2011), who made a thorough evaluation for multi-view oblique-

only imagery acquired by a multi-camera system and UAV vertical imagery. As far as the experiments with multi-view oblique imagery are concerned, various tests with and without scene constraints took place, differing in the number of GCPs, the estimation of lens distortion and the performance of self-calibration. As far as the proposed constrained aerial triangulation with the minimum number of GCPs is concerned, best results were derived by the self-calibrating bundle block adjustment with lens distortion estimation. However, the self-calibration is very sensitive in case of classical aerial triangulation without constraints using the minimum required number of GCPs. An advantage of the proposed method is the reduction of the number of GCPs through the use of scene constraints. However, the requirement for manual definition of constraints is a major limitation of this approach.

Karel et al. (2013) developed a software which georeferences oblique aerial images in two steps that include co-registration of the images and georeferencing of the resulting block as a whole, using image meta-information (footprints, flying height, interior orientation) as well as pre-existing orthophoto maps and DSMs/DTMs for the extraction of GCPs and the computation of their heights. The automatic relative orientation of the images is accomplished using SIFT features and an incremental SfM approach, using two graph structures, i.e., one graph for keypoint tracking, where the graph vertices correspond to the keypoints of all images, and one weighted image connectivity graph, where the graph vertices correspond to the images and the edges are weighted by the number of matched feature points combined with a measure of perspective distortion. Also, automatic orthoimage generation is accomplished by their software, by establishing correspondences between the imagery and the pre-existing orthoimage to extract horizontal GCPs along with DSMs/DTMs for obtaining their third dimension and optionally generating a dense DSM, if the one available is not sufficient, through semi-global matching. Although the software was tested by them using oblique aerial images, it can handle both oblique and vertical aerial images acquired by a multi-camera system, as discussed by Gerke et al. (2016) in the context of the participation of the authors in the benchmark tests. For the experiments performed using the multi-view image dataset released for the benchmark (Nex et al., 2015), SIFT features were extracted in synthetic images generated from the original ones after the implementation of affine transformations. Their approach yielded larger residuals at check points than the research solution proposed by the Southwest Jiaotong University and standard commercial SfM software packages that were tested using the same benchmark datasets, as outlined by Gerke et al. (2016). As far as the Southwest Jiaotong University approach, proposed by Hu, is concerned, it involves geometric rectification of the multi-view images using the rough exterior orientation parameters and a coarse DEM, SIFT-based image matching followed by least-squares matching for achieving subpixel accuracies, organization of the correspondences into tracks, selection of those that will be used in the bundle adjustment and exterior orientation estimation, preferably by keeping the vertical images fixed, if they have already been adjusted. Their methodology was tested using the benchmark dataset, by implementing independent bundle adjustment of the oblique aerial images, and yielded results comparable to the ones achieved by commercial SfM software (Gerke et al., 2016).

Xie et al. (2016) presented an asymmetric re-weighting method for the combined bundle adjustment of oblique and vertical aerial imagery captured by multi-camera systems, according to which each correspondence is given different weights in the $x$-direction and in the $y$-direction, which depend on image scale and the shape of the projection of each pixel on the ground. They relied on the concept that the error distribution for all observations in oblique imagery does not conform to the same Gaussian distribution, in contrast to nadir imagery, thus introducing a priori weights, namely, a normalized scale factor and a relative anisotropic ground sample distance (GSD). The assignment of different weights to observations is a fundamental issue for oblique images, as the scale varies

significantly throughout such images. The performance of the proposed method was proved to be superior to that of the standard bundle adjustment method that assumes Gaussian errors in terms of accuracy. One of its major limitations is the fact that relatively accurate initial information of the imaged scene is required, thus impeding its adoption by oblique imagery acquired by an aerial platform without onboard GNSS/IMU sensors.

Sun et al. (2016) proposed a bundle adjustment model for oblique and vertical airborne imagery acquired by multi-camera systems to reduce the number of unknown parameters and the dimensions of a non-linear optimization problem. According to this model, referred to as reduced bundle adjustment, the oblique poses are parameterized using the vertical poses and the relative poses between the vertical and oblique cameras, which are assumed to be constant at all exposure stations, thus reducing the dimensions of the normal equations and the total number of unknown parameters, which consist of the vertical poses, the relative poses and the feature points. A disadvantage of this method is its influence by the noise of relative pose parameters of the oblique cameras. Unless the noise is large, the root mean square error of the reduced bundle adjustment is only slightly bigger than that of the traditional bundle adjustment method. Nevertheless, the reduced bundle adjustment method requires a lower computational complexity as well as memory cost, which is more obvious when large datasets of imagery are involved.

Sun et al. (2017) proposed four orientation estimation approaches depending on the availability of geographic data, i.e., (i) an incremental SfM method followed by a constrained bundle adjustment using GCPs, if available; (ii) a global SfM exploiting GNSS information or an incremental SfM followed by a 3D similarity transformation from the calculated camera positions to the system of the GNSS data, if such information is available; (iii) computation of an approximate 3D similarity transformation using points extracted from 3D virtual earth platforms, if only known entities such as landmarks are available; and (iv) usage of the vertical vanishing point or objects with standard size to retrieve the vertical image orientation and scale, if explicit geographic information is unavailable. Also, they implemented a SfM workflow to compute the exterior orientation of calibrated multi-view oblique and vertical aerial images acquired by a multi-camera system using GPS/IMU information. Whereas the authors do not reveal much about the used SfM strategy and the achieved exterior orientation results, as they focus more on the subsequent semantic labelling of the computed dense point cloud, they report an even distribution of the sparse 3D points generated by the SfM process.

Jiang and Jiang (2017a; 2018) presented a SfM solution that mainly addresses three issues concerning the orientation of oblique UAV images, that is, selection of image pairs, image matching and geometric verification of matches. The determination of image pairs is conducted based on the overlap of their footprints, calculated using rough georeferencing information from onboard sensors along with information about the mean altitude of the ground or auxiliary elevation data. Then, a graph, weighted by the overlapping area along with the intersection angle of image pairs, is created and a maximum spanning tree expansion algorithm (Graham and Hell, 1985) is implemented, followed by local structural analysis, for removing redundant image pairs. As far as the image matching process is concerned, it consists of three main steps, implemented for each image pair: (i) a homography is fitted to the extracted correspondences after downsampling the images; (ii) the overlapping area of both images is determined and a tiling strategy is implemented in the overlapping area of the first image; and (iii) the computed homography is used to calculate the homologous tile in the second image and restrict feature matching. Finally, the correspondences are divided into true and false matches using an algorithm that filters obvious outliers prior to the use of RANSAC by Fischler and Bolles (1981), which is implemented for geometric verification, to

refine the final matches. The algorithm presented by Jiang and Jiang (2018) for outlier determination before RANSAC is based on (i) the projection of the corresponding feature points onto a specified elevation plane in order to transform their complex geometric relationship in image space to a simple 2D translation in object space and (ii) the rejection of wrong feature points using the estimated values of their motion direction and their motion direction change, taking into account the fact that directions of motions and direction change of motions for true matches vary in a limited range, while these values are random for false matches. The SfM workflow follows an incremental scheme, similar to the one proposed by Snavely et al. (2006). The results prove that the proposed tiling strategy improves the number and distribution of matches. However, it further increases the computational time of image orientation due to the increased number of tie points. Their proposed algorithm for outlier removal outperforms other state-of-the-art algorithms based on the RANSAC scheme for oblique imagery. The comparison of the overall SfM workflow with commercial and open-source software packages demonstrates higher efficiency in feature matching in terms of computational time, comparable reconstruction completeness in terms of the numbers of oriented images and 3D points, as well as similar relative and absolute bundle adjustment accuracy with the tested commercial software.

In conclusion, except for the methodologies by Le Besnerais et al. (2008) and Sun et al. (2017), that do not add new scientific information on the exterior orientation estimation of datasets including particularly oblique aerial imagery, each of the other three proposed bundle adjustment methods, i.e., the aerial triangulation with constraints by Gerke and Nyaruhuma (2009) and Gerke (2011), the asymmetric re-weighting method by Xie et al. (2016) and the reduced bundle adjustment approach by Sun et al. (2016), along with the SfM pipeline presented by Jiang and Jiang (2017a; 2018), contribute to the orientation estimation of oblique aerial imagery in a different way. The approach by Gerke and Nyaruhuma (2009) and Gerke (2011) reduces the number of required GCPs and simultaneously delivers similar results with the traditional bundle adjustment method with several well-distributed GCPs. The asymmetric re-weighting method by Xie et al. (2016) yields better results than the traditional bundle adjustment method, whereas the reduced bundle adjustment approach by Sun et al. (2016) yields slightly poorer results for the sake of reduced computational time. Thus, the method by Gerke and Nyaruhuma (2009) and Gerke (2011) is advised when a reduced number of GCPs is available, provided that the time factor is not crucial due to the requirement for manual work; the method by Sun et al. (2016) may be used when fast orientation results are required; whereas the asymmetric re-weighting method by Xie et al. (2016) may be applied in highly overlapping oblique aerial image datasets, if relatively accurate GNSS/IMU information is available. Finally, the matching pair selection method proposed by Jiang and Jiang (2017a; 2018) may be applied in large datasets accompanied by georeferencing data with many redundant images, while their feature-based matching tiling strategy and outlier removal method are reported to achieve reliable correspondences of a significantly increased number, which may however be a disadvantage in cases that the efficiency of image orientation is the key criterion for selecting the matching method.

### 3.3.3 Georeferencing Based on Direct Sensor Orientation

Whereas the aforementioned approaches concern indirect sensor orientations methodologies, in this section direct georeferencing approaches are outlined. Stilla et al. (2009) focused on the accuracy of direct georeferencing of sequences of infrared oblique aerial images using GNSS/INS data for texture mapping purposes. They concluded that GNSS/INS data cannot be used without corrections for performing texturing of a 3D city model, due to the shift and misalignment of the camera

coordinate system relative to the IMU coordinate system as well as the time shift effect. They proved this conclusion by calculating the differences between the directly measured and the calculated exterior orientation parameters by self-calibrating bundle block adjustment using GCPs. Larger differences were observed for the pitch angle; the differences in roll and yaw angles, despite being significantly smaller, could affect the texturing results. The mean values of the calculated differences were used as corrections to the measured angular exterior orientation parameters of images, yielding satisfying texture mapping results for a LoD2 (level of detail 2) city model.

Zhao et al. (2014) implemented direct georeferencing approaches for oblique and vertical aerial imagery taken by multi-camera systems in different coordinate systems (earth centered earth fixed frame, local tangent frame and map projection frame) using GNSS/INS data. A significant contribution of their work is the correction for systematic errors in the projection frame, namely, the earth curvature correction and the correction for projection deformation unconformity in the vertical and horizontal directions. A new approach for the determination of the first correction was proposed in their work; the respective correction formulas were defined by generalizing oblique images from nadir ones and applying the well-established formalism of earth-curvature correction for vertical imagery. Whereas the projection frame is convenient for map production and the direct georeferencing process is straightforward, for achieving high accuracy it is advised to adopt the earth centered earth fixed frame or the local tangent frame.

Whereas the methods proposed by Zhao et al. (2014) constitute self-developed research solutions, Liu and Shao (2015) used existing software for direct georeferencing of multi-view image datasets. They performed both classical aerial triangulation and direct georeferencing based on GPS/IMU data to obtain the exterior orientation parameters of oblique and vertical aerial imagery captured by a low-cost multi-lens system comprising off-the-shelf cameras. After comparison of the exterior orientation results derived by both methods, the authors concluded that the direct georeferencing approach can yield satisfactory results for rapid photogrammetry.

The aforementioned approaches by Zhao et al. (2014) and Liu and Shao (2015) serve for calculation of exterior orientation parameters of multi-view image datasets through direct sensor orientation without rectifying the imagery or computing their center point or footprint, like the methodologies proposed by Geniviva et al. (2014), Zhou and Liu (2015) and Doneus et al. (2016). Specifically, Geniviva et al. (2014) georeferenced each image by means of locating their center rather than refining the exterior orientation parameters obtained by GPS/IMU sensors. According to their proposed methodology, the GPS and IMU information for an oblique image is used for calculating its bounding box and selecting the corresponding rectangular area from existing georeferenced orthorectified imagery. Feature extraction takes place in both oblique and orthorectified imagery and image matching follows using the ASIFT (Affine-SIFT) algorithm (Morel and Yu, 2009). The outputs of the proposed methodology, which is fully automated, are the coordinates of the center of each oblique image. The proposed methodology could be improved by constraining the corresponding points to lie on the ground plane for more accurate registration results.

Zhou and Liu (2015) proposed a method for the automatic orthorectification and mosaicking of oblique and vertical aerial images acquired by a single zoom lens camera using GNSS/INS measurements, without GCPs. The a-priori estimation of the relationships between the camera focal length, the distortion coefficients and the principal point in the laboratory, the computation of the camera exterior orientation parameters by direct georeferencing and the determination of the lens distortion parameters based on the focal length of the camera using the calibration results are adopted by their proposed method, in order to geo-orthorectify the aerial images. The rectified

images are then stitched together to produce a georeferenced orthomosaic. Similarly to Zhao et al. (2014), Zhou and Liu (2015) concluded that the achievable accuracy is sufficient for rapid response situations.

Doneus et al. (2016) proposed an approach for georeferencing oblique aerial images based on GNSS/IMU data for automated GIS-based image archiving. The georeferencing involves computation of the location of both the center point of the images and their footprint as well as the optional implementation of a projective transformation for each image. The computed ground-projected image center is not the true ground principal point, but the ground image center. The ground location of each image footprint is calculated using the focal length, the image size and elevation data (DTM/DSM). A simple planar rectification of each image through projective transformation took place, instead of an orthorectification as implemented by Zhou and Liu (2015). This transformation was calculated by the set of image and ground points defining the image footprint and yields a rectified image suitable for monoplotting. Unlike Zhou and Liu (2015), Doneus et al. (2016) did not take into account lens distortion. Validation experiments of the proposed approach using real datasets took place. The exterior orientation values provided by the onboard sensors were compared with the parameters computed by commercial SfM software using GCPs, yielding satisfactory results for image archiving purposes. Furthermore, points plotted on a reference orthoimage were overlaid on the transformed images, proving good results near the image center and worse accuracy near the image borders. Apart from these points, the locations of the footprint corners of the rectified images were visually compared with their expected locations derived from the orthoimage, yielding an accuracy lying within a range of 10% of the respective image dimension. For instance, for a side of of 4000 m of the trapezoidal footprint of a high oblique aerial image, positional errors of object points up to 400 m may arise. Consequently, whereas the projective transformation of the images is sufficient for archiving purposes, it is not accurate enough for direct interpretative mapping.

### 3.3.4 Monocular Image Georeferencing

As far as georeferencing of single oblique aerial images is concerned, Frueh et al. (2004), Karjalainen et al. (2006), Ding et al. (2008), Wang et al. (2008), Wang and Neumann (2009) and Xiong et al. (2014) have presented research solutions. Karjalainen et al. (2006) presented a methodology for the calculation of the exterior orientation of a single aerial image (vertical or oblique) without the need for GCPs, using digital vector maps including 3D lines or 2D lines along with a DEM or constant input elevation for the endpoints of lines as well as good initial exterior and interior orientation parameter values. The lines are projected onto the image, and the final exterior orientation is determined using a local line scanning technique for finding the lines of the vector map from the image using the 1D Canny edge detector (Canny, 1986), followed by a line photogrammetry method using the coplanarity equation. An advantage of the proposed approach is the fact that edge detection is not necessary prior to feature matching, as edges are detected in real time in the neighborhood of the approximate location of the ground control line in the image, provided by the initial orientation values. A disadvantage of this approach is that manual work may be needed for the adjustment of the initial exterior orientation values, in the case that they are not accurate enough, as this method requires relatively accurate initial values. For this reason, the authors have built a software in which vectors from a GIS file are projected onto an image using the initial interior and exterior orientation parameters; the user may manually change the vector locations onto the image through slide switches, so that they are correctly projected onto it, to refine exterior orientation. However, according to this technique, it would also be possible to find the final

exterior orientation manually, without the need for the proposed automatic procedure. Another limitation of this approach is the required computational time, which may lie between a few minutes to hours, for the orientation of a single image, depending on the number of lines in the vector map as well as the computing capabilities. In conclusion, this exterior orientation estimation approach is similar to classical methods, differing in the way the ground control is transferred to the image, as instead of GCPs, the authors use lines.

An alternative way of getting ground control is the usage of angles between lines, as proposed by Xiong et al. (2014). The authors presented a method for the determination of the camera pose from a monocular oblique aerial image of an urban scene by introducing horizontal right angles as geometric constraints. Their concept is based on the fact that if the exterior orientation of an oblique image is recovered, the projection of an imaged right-angled building corner to an arbitrary object space horizontal plane will be a right angle. More than two horizontal right angles, manually measured, are needed for obtaining a reliable exterior orientation solution via least-squares adjustment. Unless ground measurements are available, the computed exterior orientation parameters are relative values with respect to a specific horizontal plane. Although the proposed approach was tested using calibrated cameras, the algorithm can be extended to handle non-calibrated cameras as well. However, this method needs manual work for definition of right angles. Whereas manual work is also needed by the method proposed by Karjalainen et al. (2006), compared to them the approach by Xiong et al. (2014) requires negligible computational time for image orientation once the right angles have been measured.

The methodologies proposed by Frueh et al. (2004), Ding et al. (2008), Wang et al. (2008) and Wang and Neumann (2009) are implemented in multiple images for the scope of texture mapping, yet the exterior orientation of each image is computed independently. Frueh et al. (2004) relied on the availability of a rough estimation of the camera parameters, which are refined via matching 2D lines on oblique aerial imagery with projections of 3D lines extracted from a pre-existing untextured 3D city model for texture mapping. The process of rating a camera pose, which consists of the camera exterior and interior orientation parameters without taking into account lens distortion, is accomplished by comparing the back-projected 3D model lines with the image 2D lines based on their slope and proximity of their endpoints. The pose with the highest rating is computed through optimization. The proposed methodology was tested in oblique aerial images taken from a helicopter and yielded satisfactory image orientation results that led to visually acceptable texturing of an existing 3D model.

While the methodology of Frueh et al. (2004) uses approximations for all six exterior orientation parameters, the automatic methodology of Ding et al. (2008) for the calculation of the exterior orientation parameters of calibrated oblique aerial imagery uses the rough camera position acquired by a GPS system, approximation for the heading angle and an existing 3D LiDAR model. In a first step, the pitch and roll angles are roughly computed by estimating the vanishing point of vertical lines for each image, so that coarse estimates of all camera parameters are available. Also, in this step vanishing points corresponding to non-vertical lines are detected in order to be subsequently used for 2D image corner extraction. In a second step, 2D corners that correspond to orthogonal building corners are extracted both from the pre-existing DSM and from each image and are matched to provide refined camera parameters using the algorithm of Lowe (1987). Their approach was tested using a large number of oblique aerial images of regions with different characteristics, achieving 61% correct camera pose recovery rate across all regions. The highest score of 91% was observed in a region with buildings densely distributed among few trees, whereas this approach yielded poor results in open fields with sparsely distributed buildings, due to low building density

and small number of matches between corners extracted from the imagery and the DSM. A computational time of 191 seconds per image was achieved through the used datasets.

Contrary to the approach of Ding et al. (2008), Wang et al. (2008) suggested the use of approximations for all six camera exterior orientation parameters for their refinement through image registration with an existing 3D model. According to their methodology, image lines are extracted and matched with the corresponding lines from the 3D model, after estimating the image positions of the 3D lines using the rough exterior orientation parameters. For each pair of corresponding lines, two equations may be formed based on the coplanarity condition and accurate exterior orientation parameters may be calculated using a least-squares solution if more than three pairs of conjugate lines have been detected. Their study focuses more on the theoretical part of exterior orientation refinement for automatic texture acquisition and only a few experimental results are given.

The methodology proposed by Wang and Neumann (2009) for the registration of oblique airborne imagery with 3D LiDAR data, using approximations for the six exterior orientation parameters similarly to the approaches by Frueh et al. (2004) and Wang et al. (2008), relies on the establishment of corresponding features between each image and the LiDAR data. Each feature consists of three line segments that are connected one after another into a chain. The detection of matches between image features and features extracted from the LiDAR data and projected onto each image is followed by a two-level RANSAC algorithm for removal of outliers. A homography is considered to express the transformation between the projected features of the LiDAR data and the corresponding image features for being used in the aforementioned RANSAC procedure, as it is assumed that this transformation is caused by camera rotation, considering that the error of the perspective center given by the GPS sensor is small compared to the flying height. The camera exterior orientation parameters are estimated and refined using the intersection points in the obtained corresponding features. Their approach was tested using two image datasets, yielding a correct pose recovery rate of 98.5%, which is higher than the corresponding rate reported by Ding et al. (2008). What is more, the feature proposed by Wang and Neumann (2009) seems to be more distinctive than that of Ding et al. (2008), as the average percentage of inliers in the putative feature matches is 19%, compared to the corresponding percentage of 4% in Ding et al. (2008). The proposed two-level RANSAC scheme is robust in the case that the outliers are much more than the inliers.

In conclusion, the fact that the entire methodologies proposed by Frueh et al. (2004), Ding et al. (2008), Wang et al. (2008) and Wang and Neumann (2009) are automated constitutes a significant advantage compared to the methodologies proposed by Karjalainen et al. (2006) and Xiong et al. (2014). The solution presented by Karjalainen et al. (2006) is efficient for estimating the exterior orientation parameters of multiple images only if accurate information from onboard GNSS/INS sensors is available, thus avoiding the manual work, whereas the one proposed by Xiong et al. (2014) is practically not applicable to multiple images, as it requires a significant amount of manual work.

### 3.3.5 Alternative Georeferencing Techniques

The aforementioned approaches concerning georeferencing of datasets including oblique aerial imagery either make use of existing software solutions or implement methodologies for the conventional tasks of bundle adjustment, SfM and direct sensor orientation. In this section, alternative georeferencing techniques are outlined. Both Habbecke and Kobbelt (2012) and Saeedi

and Mao (2014) presented methodologies for the estimation of the exterior orientation of datasets of oblique aerial imagery acquired by multi-camera systems. A research solution for automatic registration of oblique aerial imagery with digital vector maps containing footprints of buildings using initial exterior orientation information from onboard sensors was the subject of the research of Habbecke and Kobbelt (2012). They estimated the exterior orientation of oblique airborne imagery by (i) detecting the vertical vanishing point of each oblique image; (ii) finding vertex-to-line correspondences between map corner vertices projected onto the images and image line segments located in the neighborhood of the projected map vertices through RANSAC for obtaining initial exterior orientation parameters for each individual image as well as interior orientation parameters if they are unknown; and (iii) detecting and matching across images horizontal edges on building facades and performing a global optimization over all camera parameters to compensate for the unknown height offset between pairs of images which varies over the images according to an unknown linear height function. Their proposed methodology is automatic and yields satisfactory orientation results provided that a cadastral map is available and vertical and horizontal lines are visible in the imagery accompanied by GNSS/INS measurements.

Similarly, Saeedi and Mao (2014) used multi-view image datasets for exterior orientation estimation exploiting image line features. Their methodology is based on the availability of a rough estimation of the camera exterior orientation, interior orientation and flight information of oblique aerial images with large perspective differences. The exterior orientation parameters are computed through matching image features across multi-view oblique aerial images for image registration or 3D reconstruction purposes. The proposed features use two straight lines and their intersection, which may correspond to two connected edges and a vertex of the roof of a building. For every such feature extracted from an image, its possible correspondences are searched in another image within a narrow belt along the corresponding epipolar line. Putative matches are established via comparison with the computed descriptors and the outliers are removed using a projective matrix optimization method. Based on the results of the latter method, a second round of matching takes place within a more confined search space leading to more accurate determination of correspondences, fundamental matrix and camera parameters. The proposed features were proved to outperform SIFT and SURF (Bay et al., 2008) features, whereas ASIFT features outperform the proposed ones in case of images with rich texture. The advantage of this method compared to the approach presented by Habbecke and Kobbelt (2012) is the fact that a cadastral map is not required. However, emphasis was given on the feature matching process rather than the exterior orientation refinement.

## 3.4 Significant Research Outcomes

In this section, significant points, results and conclusions derived from the recent research concerning georeferencing of datasets including oblique aerial imagery are outlined.

### 3.4.1 Feature Extraction and Image Matching

Some traditional photogrammetric aerial triangulation software fail in the process of automatic generation of tie points when oblique aerial images are added to the block (Jacobsen, 2008; Smith et al., 2008; Wiedemann and Moré, 2012). Whereas matching between overlapping oblique aerial imagery pointing to the same cardinal direction and between oblique and vertical aerial imagery is successful in general, the matching of multi-view oblique images enclosing a large direction difference does not work well (Gerke, 2011; Gerke et al., 2016; Moe et al., 2016).

As far as the implemented methods of feature extraction and image matching are concerned, the SIFT algorithm was used by most researchers via existing SfM software (Verhoeven et al., 2012; Rupnik et al., 2013; Markelin et al., 2014; Jhan et al., 2015; Rau et al., 2015; Rupnik et al., 2015; Gerke et al., 2016; Jacobsen and Gerke, 2016; Moe et al., 2016; Ostrowski, 2016; Ostrowski and Bakuła, 2016; Passini et al., 2016; Mertes et al., 2017; Midgley and Tonkin, 2017) as well as in self-developed solutions (Gerke and Nyaruhuma, 2009; Gerke, 2011; Karel et al., 2013; Gerke et al., 2016; Jiang and Jiang, 2017a; Jiang and Jiang, 2018). Also, the ASIFT algorithm was used for image matching between unreferenced oblique images and georeferenced nadir imagery (Geniviva et al., 2014), for the georeferencing of the oblique ones. Moreover, good features to track (Shi and Tomasi, 1994) were used for the scope of image matching between oblique aerial imagery (Le Besnerais et al., 2008) via the Kanade-Lucas-Tomasi feature tracker. All these algorithms produced satisfactory results. However, the problem of matching between oblique and vertical imagery as well as between multi-view oblique imagery leads in the concentration of the majority of the correspondences on the ground plane (Geniviva et al., 2014). Another kind of features proposed by the research community, comprising two line segments and their intersection (Saeedi and Mao, 2014), was claimed to yield satisfactory results for multi-perspective oblique aerial imagery.

Furthermore, the extraction of SIFT features in stitched views generated from the original images after applying affine transformations onto them was proposed by researchers participating in the ISPRS/EuroSDR benchmark for multi-platform photogrammetry (Nex et al., 2015), in order to reduce the significant perspective distortions between multi-view imagery (Gerke et al., 2016). In addition, the geometric rectification of multi-view oblique and vertical aerial images using initial exterior orientation parameters and elevation data may reduce the affine deformation, thus improving the efficiency of image matching (Gerke et al., 2016). For the scope of extracting an increased number of feature points of good distribution throughout oblique images, the tiling strategy proposed by Jiang and Jiang (2018) may be implemented. Finally, obvious false matches can be filtered prior to the application of a RANSAC-based algorithm for geometric verification of the corresponding feature points, via the strategy proposed by Jiang and Jiang (2018).

Among the robust feature detectors that are scale and rotation invariant, the use of SURF for matching of datasets including oblique aerial images has not been investigated yet. SURF feature points are scale and rotation invariant; skew, anisotropic scaling and perspective effects are also covered to some degree and the matching of such features is accomplished in less time than the matching of other features with longer descriptor vectors, such as SIFT or ASIFT. Thus, in this dissertation, the performance of SURF in multi-view oblique and vertical aerial images is evaluated (sections 4.4.3.2 and 7.10.6) and a comparison with the ASIFT detector is also made (section 7.10.6).

### 3.4.2 Usage of Graph Structures

The exploitation of the georeferencing information provided by GNSS and IMU systems in constraining feature point extraction to relevant images only using a graph structure significantly speeds up the image matching process (Rupnik et al., 2013; Rupnik et al., 2015; Jiang and Jiang, 2017a; Jiang and Jiang, 2018) and reduces the number of possible outliers (Rupnik et al., 2015). Such graph structures usually represent images as nodes; edges connect nodes that correspond to overlapping images. Besides relying on available georeferencing metadata, the creation of a graph may also be accomplished through an image-based process for determining overlapping images. The graph edges can be given weights indicating the importance of image pairs. Such weights may

be estimated in different ways, e.g., using the overlapping area of the images along with their intersection angle (Jiang and Jiang, 2017a; Jiang and Jiang, 2018), or using the product of the number of correspondences and a measure of perspective distortion (Karel et al., 2013). Such graphs can guide the addition of images in incremental SfM procedures and reduce the risk of divergence of the bundle adjustment, as the similarity of images within triplets is maximized, thus increasing the ratio of good to bad matches and ensuring the cohesion of the block (Rupnik et al., 2013; Rupnik et al., 2015). In this dissertation, both an image-based method (section 4.3.1) and a GNSS/INS-based method (section 7.3) for the automatic creation of an undirected weighted graph that determines the image pairs are presented. Graphs can also be used for removing unnecessary image pairs in cases of large datasets, as proposed by Jiang and Jiang (2017a; 2018) using a maximum spanning tree algorithm followed by local structural analysis. Moreover, graphs can be used for feature tracking, by introducing the extracted feature points of all images as vertices and inserting an edge for each correspondence.

### 3.4.3 Usage of the Relative Poses Between the Cameras of a Multi-Camera System

According to Jacobsen (2008), the orientation of a block of oblique and vertical aerial imagery captured by a multi-camera system requires an internal system calibration, which may be accomplished via bundle adjustment of a sub-block of imagery using GPS/IMU measurements and GCPs. The exterior orientation parameters of each oblique aerial image can be expressed as a combination of the orientation parameters of the corresponding vertical image and the relative poses between the oblique and vertical cameras. This introduction of the mechanical connection between oblique and vertical aerial images taken by a multi-camera system, as adopted by Wiedemann and Moré (2012), Moe et al. (2016) and Sun et al. (2016), can significantly reduce the number of unknown parameters of the bundle adjustment (Wiedemann and Moré, 2012; Sun et al., 2016) and the dimensions of the normal equations (Sun et al., 2016), thus leading to a minimization of the computational complexity and memory cost (Sun et al., 2016). Whereas the improvement in terms of efficiency of aerial triangulation is significant, especially in the case of a large-scale scenario encompassing a large number of images, the accuracy of this "reduced bundle adjustment" is slightly poorer than the accuracy of the traditional bundle adjustment, according to which the oblique and vertical images are assumed to have independent exterior orientations, because the constraint of rigid relative poses between the cameras cannot be fully satisfied in practical cases (Sun et al., 2016).

Researchers participating in the ISPRS/EuroSDR benchmark for multi-platform photogrammetry (Nex et al., 2015), did not take the relative orientation of the images from the multi-camera system into account in the bundle block adjustment, due to the unknown synchronization accuracy of the cameras on the platform (Gerke et al., 2016). According to another benchmark participant, initial exterior orientation parameters for oblique views can be computed using the calibrated relative orientation parameters of the cameras of the platform along with the exterior orientation parameters of the nadir images (Gerke et al., 2016).

### 3.4.4 Geometric Scene Constraints for Exterior Orientation Estimation

Horizontal and vertical lines as well as right angles in object space widely exist in oblique aerial images of urban scenes (Xiong et al., 2014) and can be used as geometric constraints either in the process of bundle adjustment of multi-view imagery (Gerke and Nyaruhuma, 2009; Gerke, 2011)

or for the calculation of the exterior orientation of a single oblique image (Xiong et al., 2014). The introduction of scene constraints into the bundle adjustment of oblique imagery enhances the overall scene geometry as well as the results of bundle block adjustment and self-calibration, leading to further stabilization of the block (Gerke and Nyaruhuma, 2009). Furthermore, it permits a significant reduction in the number of necessary GCPs (Gerke and Nyaruhuma, 2009; Gerke, 2011). The incorporation of such constraints in the adjustment as fictive observations (i.e., soft constraints) permits both a flexible weighting of the constraints through the stochastic model as well as an a posteriori statistical check of the significance of the constraints (Gerke and Nyaruhuma, 2009; Gerke, 2011).

The method proposed by Gerke and Nyaruhuma (2009) and Gerke (2011) assumes that the scene structures are visible in at least two images, being unsuitable for monocular image orientation. As far as the latter is concerned, it has been shown that in the case of single oblique imagery, the horizontal-angle-based method proposed by Xiong et al. (2014) is more accurate than the vertical-line-based method proposed by Zhang et al. (2005) for use in aerial imagery of urban scenes. However, the method that utilizes horizontal right angles, as proposed by Xiong et al. (2014), fails if only gabled-roof buildings are depicted in the imagery.

The research conducted so far has dealt with the manual definition of scene constraints for oblique aerial imagery. According to Xiong et al. (2014), the fact that oblique airborne images depict many buildings of different orientations, in combination with the fact that the images of buildings are small in large-scale scenarios, make the automatic extraction of vanishing points difficult or even unsuitable for exterior orientation estimation (Xiong et al., 2014). However, the workflows proposed in this dissertation for the automatic determination of horizontal and vertical line segments and corresponding vanishing points (section 6.3) can be used in the research towards the automatic imposition of scene constraints in bundle adjustment of oblique aerial imagery.

### 3.4.5 Vanishing Point Detection Techniques

Methodologies implementing edge detection and line extraction techniques for vanishing point detection have been proposed and applied (Ding et al., 2008; Habbecke and Kobbelt, 2012) towards the calculation of exterior orientation parameters of oblique aerial images. The detection of the vertical vanishing point (image nadir point) reduces the degrees of freedom of the camera exterior orientation from six to four (Habbecke and Kobbelt, 2012). Using this point, the direction of the camera axis, i.e., the pitch and roll angles (Ding et al., 2008) or the tilt and swing angles (sections 2.4.5.1, 2.4.5.3 and 6.3 of this dissertation) at the moment of capture can be calculated. The nadir point may be successfully detected in oblique aerial imagery where vertical line segments are visible. However, slight inaccuracies in automatically detected vanishing points lead to inaccuracies in the estimation of the exterior orientation parameters (Habbecke and Kobbelt, 2012). Thus, a global optimization using all images may improve the orientation results (Habbecke and Kobbelt, 2012) through bundle adjustment. Furthermore, in addition to the extraction of vertical segments, horizontal line segments may also contribute not only to nadir point detection of each image (section 6.3 of this dissertation), but also to the exterior orientation estimation of oblique aerial imagery (Ding et al., 2008; Habbecke and Kobbelt, 2012) (section 6.5). In conclusion, methods implementing vanishing point detection techniques in single images without a subsequent optimization step using all images may serve for the calculation of initial exterior orientation parameters. The latters can be refined through a conventional bundle adjustment procedure. This is also shown in chapter 7 of this dissertation.

### 3.4.6 Weighting Strategy

Xie et al. (2016) reported that the estimation of an a priori weight for each image observation based on its scale and the amount of anisotropy in its GSDs, thus discarding the assumption of Gaussian distribution of errors used in the case of nadir imagery, leads to more precise bundle adjustment of oblique aerial imagery. A weighting strategy for image observations based on the size and shape of the ground pixel is also proposed for bundle adjustment procedures of multi-view imagery by Gerke et al. (2016). In order to investigate the impact of a scale-based weighting strategy on the exterior orientation results, a weighting strategy based on the scale of each image observation is also discussed in this dissertation (section 7.6) and its suitability is assessed using different configurations of oblique imagery (section 7.10.7).

### 3.4.7 Accuracy Issues in Multi-View Oblique and Vertical Image Blocks

Whereas the introduction of oblique imagery into the bundle block adjustment leads to only a slight improvement of the planimetric precision (Rupnik et al., 2015), the obtainable height accuracy is significantly improved (Rupnik et al., 2015; Gerke et al., 2016; Moe et al., 2016), mainly thanks to the improved ray intersection geometry (Moe et al., 2016). Specifically, according to Rupnik et al. (2015), the standard deviation in the *Z* direction is 2-5 times better when all images, both vertical and oblique, are used in the block. Also, the introduction of oblique images reduces the magnitude of noise in 3D object space (Rupnik et al., 2015) and stabilizes the block (Rupnik et al., 2015; Gerke et al., 2016), especially in cases of small overlap and lack of GNSS or well-distributed GCPs (Rupnik et al., 2015). The introduction of oblique imagery into blocks leads to more resistant geometry (Gerke et al., 2016; Ostrowski and Bakuła, 2016), thanks to more tie points and higher overlaps, and permits a significant reduction in the number of GCPs, without affecting orientation results (Ostrowski and Bakuła, 2016).

As far as the introduction of vertical images into oblique image blocks is concerned, the conclusions of the research conducted so far are contradictory. According to Ostrowski and Bakuła (2016), the introduction of nadir imagery into oblique image blocks does not substantially affect the precision and effectiveness of the orientation. However, according to Nakada et al. (2016), the accuracy of the bundle block adjustment is improved by the introduction of vertical imagery into oblique image blocks.

### 3.4.8 Influence of Different Aspects on the Accuracy of Measurements after Bundle Block Adjustment

The bundle adjustment of oblique imagery may result in high quality photogrammetric measurements (Ostrowski, 2016). It has been shown that an increased tilt angle leads to larger point redundancy, improved precision, and homogenous standard deviation in all three dimensions (Rupnik et al., 2015). The imposition of the viewing direction constraint, whereby only observations in nadir cameras or oblique cameras of the same viewing direction are considered, decreases the precision along the Z axis via a reduction of both the number of observations and the intersection angles and causes inhomogeneous planimetric precision (Rupnik et al., 2015).

An increase in overlap between images improves the precision (Rupnik et al., 2015; Gerke et al., 2016) and helps compensate for deficits of image calibration (Gerke et al., 2016). Rupnik et al. (2015) suggested that an overlap of 80% along-track and 60% across-track would be a good

compromise, taking into account both the precision and productivity aspects. According to Gerke et al. (2016) in the case of the 80%/80% overlap, the random errors seem to be 2-3 times smaller compared to the 60%/60% overlap. A cross overlap of more than 60% improves the precision of triangulated points after bundle adjustment (Fritsch and Rothermel, 2013). The change in precision among the 80%/60%, 80%/80% and 90%/90% scenarios is less dramatic (Rupnik et al., 2015). In the case of 60%/60% overlap in multi-view imagery blocks, remaining errors from self-calibration cannot be sufficiently compensated by GCPs (Gerke et al., 2016).

Although the influence of self-calibration on bundle adjustment depends on several factors, Rupnik et al. (2015) suggest that it is generally better to perform self-calibration, as the bundle adjustment with self-calibration reduces the noise and deformations and improves height accuracy. According to Jacobsen and Gerke (2016) and Passini et al. (2016), self-calibration improves the results of bundle adjustment. Also, the large number of correspondences between images increases the reliability of the self-calibration parameters and improves the accuracy of the bundle adjustment results (Passini et al., 2016).

A poor distribution of GCPs mostly reduces the accuracy along the Z axis (Rupnik et al., 2015). However, in the case of strong overlaps (80%/80%), the GCP distribution does not have a significant influence (Gerke et al., 2016). A good GCP distribution is a prerequisite for the acquisition of superior results in bundle adjustment when additional parameters for self-calibration are used and for avoiding systematic errors in object space (Gerke et al., 2016).

The division of large blocks of multi-view imagery into sub-blocks leads to more efficient and productive orientation of images (Ostrowski and Bakuła, 2016). Also, according to Rupnik et al. (2015), GNSS observations stabilize the bundle adjustment solution.

### 3.4.9 Matching between Images and Existing 3D or 2D Data

The exterior orientation of oblique aerial imagery can be computed through matching image features with features extracted from a pre-existing 3D model of the scene, provided that a rough estimation of the camera parameters is available. Such methodologies are based on matching of:

- 2D image lines with projections of 3D lines extracted from a 3D model (Frueh et al., 2004; Wang et al., 2008);
- 2D orthogonal corners extracted from images and from a DSM of the scene (Ding et al., 2008); and
- three connected line segments extracted from images and 3D LiDAR data (Wang and Neumann, 2009).

The research conducted so far aimed at using the refined orientation values for texture mapping of the 3D models with the oriented oblique aerial imagery. However, some problems may arise using these methods, like the fact that the 3D lines may not match perfectly with the extracted image lines due to possible inaccuracies in the existing 3D model, such as erroneous building dimensions due to overhanging roofs (Frueh et al., 2004). Generally, in the case of airborne oblique imagery accompanied by dual frequency post processed GPS data, it can be assumed that the error of the camera location provided by the GPS sensor is very small, compared to the flying height; thus, according to Wang and Neumann (2009), the transformation between the projected 3D features (referred to above) of a 3D model to the images and the corresponding image features may be assumed to be a homography, as it is mainly caused by camera rotation. This transformation may be used in a RANSAC procedure for computation of reliable matches between projected features

from a 3D model and image features. Finally, the feature consisting of three connected line segments proposed by Wang and Neumann (2009) is claimed to be more distinctive than the feature comprising 2D orthogonal corners proposed by Ding et al. (2008), thereby increasing the percentage of inliers in the matching process (Wang and Neumann, 2009).

Except for matching oblique imagery with 3D models, techniques of their matching with existing 2D map vector data with (Karjalainen et al., 2006) or without (Habbecke and Kobbelt, 2012) elevation information have also been proposed for exterior orientation estimation, either by projecting the vector lines on the imagery and performing a local line scanning technique that attempts to locate the linear features from the imagery or by establishing vertex-to-line correspondences between map corner vertices and extracted image line segments (Habbecke and Kobbelt, 2012). All the aforementioned techniques yield satisfactory results provided that a sufficient number of lines are visible in the oblique imagery, which is almost always the case for images depicting urban environments.

### 3.4.10 Rough Georeferencing Issues

Instead of georeferencing images by means of camera exterior orientation parameters, some researchers georeference images coarsely by means of locating their center point (Geniviva et al., 2014; Doneus et al., 2016) and their footprint (Doneus et al., 2016) and rectifying the images using a 2D transformation (Doneus et al., 2016). The rough georeferencing of large datasets of images may enable rapid and almost automatic GIS-based archiving of the imagery (Doneus et al., 2016) for easy detection of those that cover a specific region. Automatic rough georeferencing may be accomplished either using GNSS/INS data (Geniviva et al., 2014; Doneus et al., 2016) or without using this kind of information, as it is shown in sections 4.3 and 5.4 of this dissertation.

The parameters that are crucial for the georeferencing accuracy based on GNSS/INS data, in terms of locating the center and footprint of each image as well as 2D image rectification are, according to Doneus et al. (2016), the angular orientation from the INS, the positional accuracy from the GNSS sensor and the ray intersection angles with the terrain. An error in the INS-based rotation angles has stronger influence on the mapped result in cases of high oblique images. Also, the accuracy of points near the image center is better than the accuracy of points near the image borders (Doneus et al., 2016). Other factors that influence georeferencing are the camera lens distortion if neglected, the true camera constant if the nominal value of the focal length is used instead, as well as the accuracy and spatial resolution of the available DTM/DSM (Doneus et al., 2016).

## 3.5 Conclusions

This chapter has provided a review of recent methods for georeferencing of datasets that include oblique airborne images. In addition to providing information concerning the types of oblique aerial images and common camera configurations as well as their brief history and current research trends and applications, this chapter has included a detailed description of georeferencing workflows for aerial images acquired by multi-camera systems (combinations of vertical and oblique imagery or solely oblique imagery) as well as images acquired by single airborne cameras, including either datasets comprising oblique and nadir imagery or only oblique imagery, or monocular oblique images. Processing pipelines accomplished using existing software and methodologies conducted via self-developed research solutions have been presented, regardless of whether the researchers focused entirely on the georeferencing part or presented results from other photogrammetric

processes as well. Moreover, the significant points, results and conclusions derived from the georeferencing methodologies implemented by the researchers throughout the last years have been highlighted and categorized, depending on the topic they cover and their contribution in the georeferencing process.

Based on the most significant research outcomes, derived by the literature review, some open research issues regarding georeferencing of oblique aerial images have arisen. Such issues, which are addressed in the next chapters of this dissertation, concern (i) the automatic rough georeferencing of oblique aerial images without availability of GNSS/INS information; (ii) the establishment of robust image matching and outlier removal techniques for oblique aerial images; (iii) the automatic extraction of elements describing the geometry of the scene depicted in oblique aerial images (e.g., vertical/horizontal lines and coplanar points) for exterior orientation estimation; and (iv) the establishment of a weighting strategy for image observations of different scale in bundle adjustment processes, without requiring initial orientation values from onboard sensors. Furthermore, other significant issues treated in the following chapters of this dissertation, which have not been addressed by the researchers who have dealt with oblique aerial images, are (i) the investigation of the exterior orientation results that can be achieved in non-ideal bundle adjustment scenarios characterized by the lack of well-distributed GCPs and minimum image measurements without data from onboard sensors; and (ii) the increase of the automation of the exterior orientation process, under the assumption of coplanar GCPs. Furthermore, the interest of the research community towards georeferencing of oblique aerial images has focused more on multi-view datasets, mainly captured in a Maltese-cross configuration, which are accompanied by both good initial approximations for the exterior orientation parameters by GNSS/INS sensors and GCPs. Thus, in the following chapters of this dissertation, emphasis is given in UAV oblique image datasets and aerial image datasets not accompanied with data from onboard sensors.

# Chapter 4
# Georeferencing via 2D Transformations

## 4.1 Introduction

For the large majority of users of airborne oblique and vertical imagery, rough georeferencing information is essential, either for amateur photographic shots and visualization purposes, in the form of geotagging, or in the context of a photogrammetric or interpretation application. Over the last fifteen years, much research has been conducted concerning georeferencing of both oblique aerial images and datasets of vertical and oblique airborne images, relying on the availability of approximate georeferencing information (Frueh et al., 2004; Karjalainen et al., 2006; Ding et al., 2008; Jacobsen, 2008; Le Besnerais et al., 2008; Smith et al., 2008; Wang et al., 2008; Gerke and Nyaruhuma, 2009; Stilla et al., 2009; Wang and Neumann, 2009; Gerke, 2011; Fritsch et al., 2012; Habbecke and Kobbelt, 2012; Wiedemann and Moré, 2012; Fritsch and Rothermel, 2013; Rupnik et al., 2013; Geniviva et al., 2014; Markelin et al., 2014; Saeedi and Mao, 2014; Zhao et al., 2014; Jhan et al., 2015; Liu and Shao, 2015; Rupnik et al., 2015; Zhou and Liu, 2015; Doneus et al., 2016; Gerke et al., 2016; Moe et al., 2016; Nakada et al., 2016; Ostrowski, 2016; Ostrowski and Bakuła, 2016; Sun et al., 2016; Xie et al., 2016; Jiang and Jiang, 2017a; Sun et al., 2017; Toschi et al., 2017; Jiang and Jiang, 2018). As far as photogrammetric applications are concerned, rough georeferencing information is generally provided by onboard sensors. However, in cases of non-metric and interpretation applications, as well as in cases of old datasets, information from GNSS/INS sensors may not be available. In this chapter, a methodology is proposed for automatic rough absolute georeferencing of large datasets of multi-perspective oblique and vertical aerial images of urban regions with unknown interior orientation and flying height, without the need for GNSS/INS sensors or the collection of GCPs in the field. The developed approach is based on the measurement of a minimum number of points in only one image and the input of their ground coordinates, extracted from a free program for earth observation, such as Google Earth. Except for providing rough georeferencing information for photogrammetric as well as interpretation and

visualization applications, this methodology automates the process of selecting, among a large number of oblique and vertical aerial images, the ones that cover a specific region in order to be used in other applications, such as 3D modelling or texturing. This can be done without any metadata or flight information. Thus, it eliminates the need for manual searches across all images for the determination of the appropriate ones, which can be a very time-consuming process. Another instance is the reuse of old datasets of oblique and vertical aerial images, which are stored in archives without being associated with georeferencing information (Redweik et al., 2016). In general, the developed methodology may be applied for the georeferencing of multi-perspective aerial images without any metadata.

## 4.2 Related Work

An interesting review and classification of rectification methods for geometric correction of images is made by Novak (1992). Such methods are classified into polynomial, projective and differential ones. The first two classes of methods use analytical relationships for the transformation between an image and a reference system, which are approximate relationships that do not require knowledge of the geometry and orientation of the camera. The third method, i.e., an orthorectification, models the geometric reality of the imaging process by means of the collinearity equations and corrects for relief displacements in the image. The transformations investigated in this dissertation for georeferencing of datasets containing oblique aerial images belong to the first two categories (i.e., polynomial and projective transformations), and the results obtained using such kind of transformations are assessed. Thus, the developed framework is referred to as "rough" georeferencing, on the grounds that only 2D transformations that approximate the relation between the images and the ground reference system are used.

The related work on existing rough georeferencing approaches applied in oblique aerial imagery is presented in chapter 3 of this dissertation. Among them, Geniviva et al. (2014) and Doneus et al. (2016) roughly georeferenced datasets of oblique aerial images without implementing any indirect solution for estimation of the camera exterior orientation parameters that requires automatic or manual multi-image measurements (i.e., SfM/aerial triangulation). Geniviva et al. (2014) presented a rough georeferencing technique for single oblique and vertical aerial images accompanied by GNSS/INS data by means of locating their center point through image matching with a georeferenced vertical image. Doneus et al. (2016) georeferenced oblique aerial images acquired by a single camera by means of locating their center point and their footprint through direct georeferencing data from GNSS/INS sensors along with elevation data, and roughly rectifying them using a 2D projective transformation. What is more, the georeferencing of oblique aerial images in terms of orthoimagery generation, rather than rectification via a 2D transformation, has been studied by several researchers (Verhoeven et al., 2012; Karel et al., 2013; Zhou and Liu, 2015; Mertes et al., 2017; Midgley and Tonkin, 2017). Among them, only Zhou and Liu (2015) implemented a direct georeferencing process for computation of the camera exterior orientation needed by the orthorectification process, whereas the other researchers estimated the camera exterior orientation parameters through a SfM procedure. However, orthorectification procedures for oblique aerial imagery are out of the scope of this chapter, which deals with image georeferencing through a 2D transformation.

The aforementioned rough georeferencing approaches for oblique aerial images require the availability of GNSS/INS data from onboard sensors and camera characteristics (focal length) or interior orientation parameters. However, the methodology introduced in this chapter does not rely,

as mentioned, on any metadata from onboard sensors nor on information for the camera. Furthermore, both rough georeferencing techniques proposed by Geniviva et al. (2014) and Doneus et al. (2016) are applied to single images, without making use of any image registration framework, like the one proposed in this chapter, which can be applied in large datasets of oblique aerial imagery. Finally, an investigation into the applicability of different types of 2D transformations for georeferencing oblique aerial imagery is addressed solely in this chapter. Specifically, the developed rough georeferencing framework calculates the 2D transformation parameters from each image to the ground reference system. With this information as input, image rectification is performed, the ground footprints of the images are determined and additional information for subdivision of images into groups and selection of images is also generated. Furthermore, error analysis and accuracy assessment for the investigation of the influence of different parameters on the georeferencing results is performed.

Whereas the research concerning georeferencing of oblique aerial imagery through estimation of a 2D transformation from each image to the ground reference system is limited to a few publications, more research has of course been conducted towards georeferencing of vertical aerial images based on a 2D transformation. Specifically, Jaramillo (2000) georeferenced single vertical aerial images through rectification via 2D conformal, affine and projective transformation using GCP measurements, with the latter transformation achieving the best results. Rocchini and Rita (2005) georeferenced single vertical aerial imagery via rectification through polynomial transformations of first (affine) and second order using numerous GCPs. Lin et al. (2010) presented an image georeferencing method based on registration with a reference orthoimage through corner detection and matching using a 2D affine transformation. Cléry et al. (2014) proposed a georeferencing framework for a single old vertical aerial image via estimation of the 2D homography transformation between the image and the reference coordinate system through registration with a topographic database using line feature matching. However, all above georeferencing frameworks are applied to single oblique imagery, in contrast to the developed georeferencing framework, presented in this chapter, that computes georeferencing information among large datasets of imagery. Similarly to the aforementioned research works that do not involve the implementation of any image matching stage, Li and Briggs (2012) presented an image-to-vector georeferencing method applied to a single image, according to which an image of a relatively small geographic area is automatically located relatively to a substantially larger vector map, assuming either a similarity or an affine transformation. The steps of their proposed method include automatic identification of corresponding road intersections between the image and the vector map, transformation verification, globalization and optimization.

On the other hand, Chen and Tseng (2016) presented a georeferencing method for aerial image datasets using an image registration scheme based on SIFT for historical vertical aerial imagery without GNSS/INS data, using a 2D affine or projective transformation and manual GCP measurements. Similarly to the developed georeferencing method presented in this chapter, they rectified the imagery based on the computed transformation parameters. However, they tested their approach using a small number of images (12 images) and a big number of GCPs (17 GCPs), as opposed to the purpose of the experiments performed in this chapter, that assess the 2D transformation-based georeferencing of large datasets containing oblique imagery with the minimum required number of GCPs measured in a single image. Also, Gotovac et al. (2017) georeferenced UAV vertical images by means of calculating their affine transformation parameters for the transformation of each image to the reference coordinate system. Their proposed georeferencing method involves SIFT feature extraction, image matching, estimation of the affine transformation between image pairs, warping of images to a planar compositing surface via the

affine transformation parameters and image blending, estimation of the affine transformation parameters of the generated mosaic of aerial images and automatic affine rectification of the mosaic through an existing GIS software package. For the scope of georeferencing, the authors used the rough camera positions available from the GPS sensor of the UAV platform, without using any GCPs. Also, Kim (2018) presented a georeferencing procedure for a strip of historic aerial photographs that involves the manual measurement of GCPs in the first, middle and last image of the strip, their georeferencing using a 2D affine transformation, matching of SURF features between the overlapping images and georeferencing of each new image using the ground coordinates of the conjugate points estimated using the affine parameters of the already georeferenced image of each pair. A final step that includes the warping of images using the estimated 2D affine transformation parameters through available GIS tools is implemented. Similarly, to Chen and Tseng (2016), Kim (2018) tested the proposed approach using a small number of images (12 images).

Instead of using a global 2D transformation throughout the entire image, Liew et al. (2014) presented a piecewise linear transformation scheme that divides the image into triangular regions and separately employs different linear transformations to each region for the purpose of image rectification, using manual GCP measurements. The comparison of their proposed method to an affine rectification shows higher error rates in the case of the affine transformation. However, their proposed framework was only tested using flat areas and results using aerial images of hilly/mountainous areas have not been investigated. Their previous work (Liew et al., 2012) on rectification of single vertical aerial imagery via a commercial software package using polynomial transformations of first (affine), second and third order, as well as the "adjust transformation" via GCP measurements in single images, served as a basis for the comparison of the aforementioned transformations. The implementation of the "adjust transformation" by the used software requires at least three GCPs and optimizes both global and local accuracy, by performing a polynomial transformation using two sets of control points and adjusting the control points locally to better match the target control points using a triangulated irregular network technique.

Also, several accurate georeferencing procedures for aerial images have been reported in literature, such as a method for accurate geo-registration of aerial images using already georeferenced images and height maps or DSMs (Zhuo et al., 2017), as well as a method for georeferencing of an orthomosaic of aerial images using polynomial transformations of first and second order (Nevins, 2017). However, these are out of the scope of this dissertation, that investigates image 2D georeferencing without any elevation data for the imaged scene nor the implementation of an orthorectification process. Despite the obvious superiority of an orthorectification over a 2D image transformation (Rocchini et al., 2012), the purpose in this dissertation is to investigate the applicability of an image georeferencing framework based on 2D transformation and propose a complete rough georeferencing framework, rather than another orthorectification technique. Finally, several methods for georeferencing satellite imagery using 2D transformations have been discussed in literature, such as the methods by Hild et al. (2000), Sadeghian et al. (2001), Willneff and Poon (2006), Liu et al. (2011), Panigrahi et al. (2011) and Shen et al. (2017), which are out of the subject of this dissertation.

In conclusion, among the aforementioned methods applied in vertical aerial imagery, only the methods by Chen and Tseng (2016), Gotovac et al. (2017) and Kim (2018) are applied in datasets of images, rather than single images, and perform a global 2D transformation from every image to the reference coordinate system. Among them, only the method proposed by Gotovac et al. (2017) is applied in a large dataset of images, like the method presented in this chapter, which investigates automatic rough georeferencing procedures for large datasets of imagery. Whereas the methods

presented by Chen and Tseng (2016) and Kim (2018) require several GCP measurements in more than one image, the developed method investigates the accuracy that can be achieved using a bad distribution of a minimum number of points of horizontal ground coordinates, which are measured in a single image. On the other hand, the georeferencing framework proposed by Gotovac et al. (2017) performs georeferencing through GNSS/INS data; such kind of data is not used in the developed 2D georeferencing approach.

Finally, methods for estimating the relative orientation parameters of image pairs in unordered image datasets and computing the exterior orientation parameters of each image in the same – arbitrary or not – coordinate system, through image registration schemes like the one used by the georeferencing methodology presented in this dissertation, have been proposed in literature. For instance, Kakogiannou (2017) calculated rough exterior orientation parameters of images in an arbitrary coordinate system for being refined in a bundle adjustment solution, through extraction of feature points; image matching and estimation of the fundamental matrix of each image pair; rough estimation of the camera constant for each image pair; computation of the essential matrix using the mean value of the estimations of the camera constant; calculation of the relative orientation parameters of each image pair through decomposition of the essential matrix and subsequently using the coplanarity condition; estimation of model coordinates of points; and expression of the orientation parameters and the 3D coordinates of points in a common coordinate system. Also, several other researchers have implemented the aforementioned methodology or variations of it, e.g., Läbe and Förstner (2006) and Lerma et al. (2010). Whereas the methodology proposed in this dissertation is also based on image matching, estimation of the fundamental matrix of each image pair and computation of the relative positions of the images, it differs from the aforementioned ones in that it computes for every image its 2D transformation parameters to the ground reference system instead of its exterior orientation parameters. Its results, i.e., the 2D transformation parameters, are especially intended for being used for rectifying the images and inserting them in a GIS environment, extracting their footprints, subdividing them into groups based on the region they cover and detecting thoses depicting a specific region. Also, the developed methodology does not perform any bundle adjustment stage, thus being faster than the aforementioned ones. On the other hand, it depends on the number of transformations involved between each image and the ground reference system and it does not use any strict mathematical model, as the 2D transformations used are approximate. Thus, the error arising from the developed methodology is bigger than the aforementioned methods and increases depending on the number of transformations.

## 4.3 Developed Georeferencing Method

In this section, the developed methodology for rough georeferencing of large datasets of multi-view oblique and vertical aerial imagery is presented. It relies on the combination of existing algorithms for transferring geospatial information among images and consists of three main steps:

- determination of overlapping images;
- detection of valid correspondences among images; and
- storage of the transformation parameters from each image to the reference coordinate system.

The parameter values used in this three-step process are presented in Table 4-1.

Table 4-1. Parameter values used by the developed georeferencing algorithm

| Symbol | Step | Value |
|---|---|---|
| *resizingFactor1* | Determination of overlapping images | 4; if maxDimension < 2000 pixels<br>5; if maxDimension < 6000 pixels<br>7; if maxDimension < 10000 pixels<br>8; if maxDimension ≥ 10000 pixels |
| *resizingFactor2* | Image matching | 2; if maxDimension < 2000 pixels<br>3; if maxDimension < 6000 pixels<br>5; if maxDimension < 10000 pixels<br>6; if maxDimension ≥ 10000 pixels |
| *hessianThres1* | Determination of overlapping images | 500 |
| *hessianThres2* | Image matching | 400 |
| *descriptorsDistThres* | Determination of overlapping images; Image matching | 0.2 |
| *fundamentalThres* | Determination of overlapping images; Image matching | 3 |
| *minMatches* | Determination of overlapping images | 20 |
| *transformThres* | Image matching | 80 |
| *matchesThresForGeoreferencing* | Georeferencing | 60 |

### 4.3.1 Image-Based Determination of Overlapping Images

In order to minimize the computational burden and the processing time of the main procedure of image matching, which is a fundamental step of the developed methodology, the overlapping images are determined in a first step, so that the subsequent search of correspondences takes place solely in such images. The initial step is the resampling of images to a sufficiently low resolution, in order to speed up the process. The resizing factor (*resizingFactor1*), that is, the integer factor by which the dimensions of each image are divided, depends on the initial size of the images. Feature points are then extracted in each image after it has been converted to greyscale, using the feature detector SURF, which is robust in changes of scale and rotation while being computationally less intensive than other feature detectors with longer descriptor vectors, e.g., SIFT. The choice of this algorithm has taken into consideration the nature of oblique aerial images (different perspective, scale and orientation) and their possibly large number, which implies large computational time. SURF feature points are scale and rotation invariant; skew, anisotropic scaling and perspective effects are also covered to some degree and the matching of such features is accomplished faster than other features with longer descriptors. The first step of SURF is the detection of feature points located in blob-like structures of an image, based on the determinant of an approximation of the Hessian matrix, which is computed for every pixel of the image at all scale levels of each octave into which the scale space is divided. The second step is the computation of a 64-dimensional descriptor vector for each interest point, which indicates the underlying intensity structure of a square region around it, oriented along its dominant orientation, with the size of the region depending on the scale of the point.

At the stage of finding correspondences, the feature points extracted from a given image are compared to the feature points extracted from all the other images, using the criterion of the minimum Euclidean distance between their descriptor vectors. The correspondences are rejected if the distance between the descriptors of the matched interest points is above a maximum accepted threshold (*descriptorsDistThres*). Additionally they are geometrically verified via the RANSAC algorithm, through the computation of the fundamental matrix using the eight-point algorithm (Hartley, 1997). The threshold (*fundamentalThres*) used by the RANSAC algorithm is the maximum allowable distance from a point to an epipolar line, beyond which the point is considered to be an outlier. The remaining correspondences may still contain some outliers lying on the epipolar lines, which however do not affect the reliability of the resulting information, that is, if the images need to be matched in the next step. The main output of this procedure is the number of matches between the images of every image pair, which determines whether the images are overlapping. The developed workflow is illustrated in Figure 4-1.

An undirected weighted graph that connects the overlapping images is created, using as input the number of matches between the images of each pair. Its nodes correspond to the images; an edge connects two nodes that correspond to overlapping images. Two images with a minimum number of matches (*minMatches*) are considered to be overlapping. Each edge of the graph is given a weight, which corresponds to the number of correspondences between the images that it connects.

### 4.3.2 Image Matching

SURF feature points are extracted anew in all images, after reduction of their dimensions. The resizing factor used here (*resizingFactor2*) is smaller than the resizing factor (*resizingFactor1*) used in the previous step (section 4.3.1). The Hessian threshold (*hessianThres2*) used for point extraction is lower than the one used in the previous step (*hessianThres1*), resulting in a larger number of interest points in every image. The Hessian threshold is data-specific and should be determined after experiments with real datasets. Indicative values used in the experiments performed in this dissertation are presented in Table 4-1.

During the image matching procedure, which takes place only for overlapping images if at least one of them has not been georeferenced, a cross-check test is implemented, which checks if the matches found in image *j* for feature points in image *i* are the same as the ones found in image *i* for feature points in image *j*; the correspondences that are found in both directions are retained and the other ones are removed. Hence, the similarity of the descriptors is verified through reverse

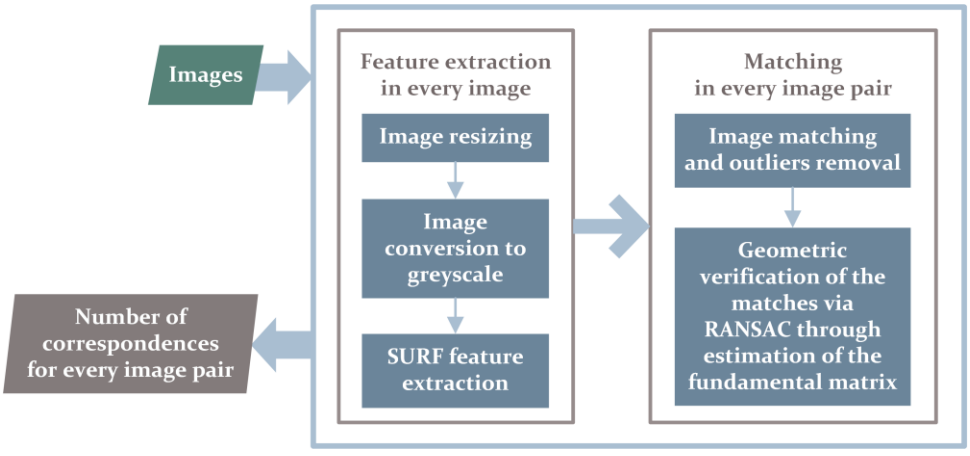


Figure 4-1. The developed workflow for image-based determination of overlapping images

comparison, using the criterion of the minimum Euclidean distance, which is constrained to be below a predefined value (*descriptorsDistThres*). The RANSAC algorithm is applied for the removal of outliers via the computation of the fundamental matrix, using a distance threshold (*fundamentalThres*). In spite of this geometric constraint, some incorrect matches remain; feature points erroneously considered to match a feature point in another image and happenning to lie on the epipolar line of the homologous feature point under consideration are not detected via RANSAC. Thus, a point-to-point constraint needs to be imposed. This point-to-point constraint is imposed through a 2D transformation, which is also used for the purpose of georeferencing, as it is discussed in section 4.3.3.

Specifically, a 2D affine transformation, a 2D projective transformation (homography) and 2D polynomial transformations of second, third, fourth and fifth order were fitted to the correspondences via RANSAC, using a distance threshold (*transformThres*) for the determination of the erroneous matches. This threshold is the maximum allowable distance from an image point to its estimated image position using the computed transformation parameters, beyond which the point and its homologous one found from the aforementioned matching process, are considered to be outliers. The threshold *transformThres* has to be small enough to reject the outliers yet sufficiently large to cope with all of these transformations being approximate and not strictly representing the actual relation between two central projections (with the sole exception of the homography in the case of planar scenes or images taken from the same point). Its value, as outlined in Table 4-1, has been found through several tests with the image datasets described in section 4.4.1. The results indicated that the geometric constraints imposed by all of these transformations successfully reject the wrong correspondences, but end up in rejecting correct matches as well (false negatives, causing omission errors). The affine transformation, the homography and the second order polynomial transformation, which are expressed by equations (4.1), (4.2) and (4.3) respectively, result in the smallest number of erroneously rejected matches. Thus, these relations are applied in the next step, for the rough georeferencing of the images.

$$\text{Affine transformation: } x_2 = a_1 + a_2 x_1 + a_3 y_1 \quad y_2 = a_4 + a_5 x_1 + a_6 y_1 \tag{4.1}$$

$$\text{Projective transformation: } x_1 = \frac{h_1 x_1 + h_2 y_1 + h_3}{h_7 x_1 + h_8 y_1 + 1} \quad y_1 = \frac{h_4 x_1 + h_5 y_1 + h_6}{h_7 x_1 + h_8 y_1 + 1} \tag{4.2}$$

$$\text{Second order polynomial transformation: } \begin{cases} x_2 = b_1 + b_2 x_1 + b_3 y_1 + b_4 x_1^2 + b_5 x_1 y_1 + b_6 y_1^2 \\ y_2 = b_7 + b_8 x_1 + b_9 y_1 + b_{10} x_1^2 + b_{11} x_1 y_1 + b_{12} y_1^2 \end{cases} \tag{4.3}$$

### 4.3.3 Georeferencing

The georeferencing of the images is based on the measurement of a minimum number of points of known ground coordinates in only one image and the computation of the scaled relative positions of the images. The required number of points of known coordinates depends on the transformation that approximates the relation between the image on which the points of known coordinates are measured and the reference coordinate system:

- three points are required in the case of the 6-parameter affine transformation;
- four points are required in the case of the 8-parameter projective transformation; and
- six points are required in the case of the 12-parameter second order polynomial transformation.

The results concerning the suitability of these transformations are presented in section 4.4.2.2.

The georeferencing stage for an image of a pair is executed after the image matching stage for this pair, if at least one of the images of the pair has not been georeferenced. The first step of the georeferencing procedure is the least-squares estimation of the transformation parameters between the images of the pair being matched, using their initial estimation obtained by RANSAC and all the inliers returned by the same algorithm. If one of the images being matched is georeferenced and the number of valid correspondences is bigger than a threshold (*matchesThresForGeoreferencing*), the transformation parameters between the non-georeferenced image and the reference coordinate system are computed. This threshold is defined in order to ensure that the georeferencing information is transferred between images with a sufficient number of matches and generally significant overlap. Additionally, measures concerning the geometric stability of the transformation, regarding the distribution of the corresponding feature points, could also be examined in this step, as well as in all steps examining the number of correspondences, e.g., a check that the feature points are located in at least three quadrants of the images, or a measure to evaluate the area that the overlapping region covers. If none of the images are georeferenced, only their transformation parameters are stored. Thus, some images may pass through this step without having been georeferenced because either their overlapping images are not georeferenced or the number of correspondences detected between these images and their overlapping ones is not adequate. Consequently, a further step follows, whereby the transformation parameters from each remaining non-georeferenced image to the ground coordinate reference system are computed. This is done using the estimated transformation parameters between the non-georeferenced image and its corresponding georeferenced one with the maximum number of matches.

If an affine transformation approximates the relation between two images and the relation between an image and the reference coordinate system, the transformation from every other image to the ground coordinate system is described by an affine transformation as well, because it is a first order polynomial transformation. Consequently, in this case, the number of the transformation parameters from an image to the reference coordinate system remains the same, regardless of the number of equivalent linear transformations involved between this image and the reference system. The same applies to the case of the projective transformation. In contrast, if the second order polynomial transformation expresses the relation between two images, as well as between one image and the reference coordinate system, the order of the polynomial transformation from every other image to the ground reference system changes with the number of second order transformations involved between the image and the reference ground system. For instance, an image which has sufficient overlap with the one that is georeferenced using the points of known coordinates requires thirty parameters for its georeference (fourth order polynomial); an image which overlaps with the aforementioned one is georeferenced using fifty-six parameters (sixth order polynomial) and so on. Thus, the transformation from only one image to the reference coordinate system is described by a second order polynomial transformation; every other image demands a polynomial of higher order. In this case, all the parameters which define subsequent second order polynomial transformations are stored in the georeferencing file of an image; in this way, the calculation of the world coordinates for a pixel of the image is easily carried out by applying the subsequent second order polynomial transformations.

The main output of the georeferencing procedure is considered to be the parameters which define the transformation between each image and the world coordinate system. With this information as input, the images can be resampled and inserted in a GIS environment at approximately the correct position, orientation and scale, the image footprints can be exported in vector format, the image

datasets may be subdivided into groups of images based on the region they cover and the images that depict a specific region can be automatically detected. These tools are discussed in section 4.4.2. The developed workflow for the matching and georeferencing of the images is illustrated in Figure 4-2.

## 4.4 Experimental Results

In this section, the georeferencing software suite that has been developed implementing the methodology outlined in section 4.3 is presented, along with the three test datasets and the results that have been achieved.

### 4.4.1 Test Datasets

The first dataset, hereinafter referred to as *RCD30 Oblique Penta Dataset*, consists of 150 multi-perspective oblique and vertical aerial images taken by a Leica RCD30 Oblique Penta camera system with a 60-megapixel sensor from a flying height of about 520 m. The images were captured

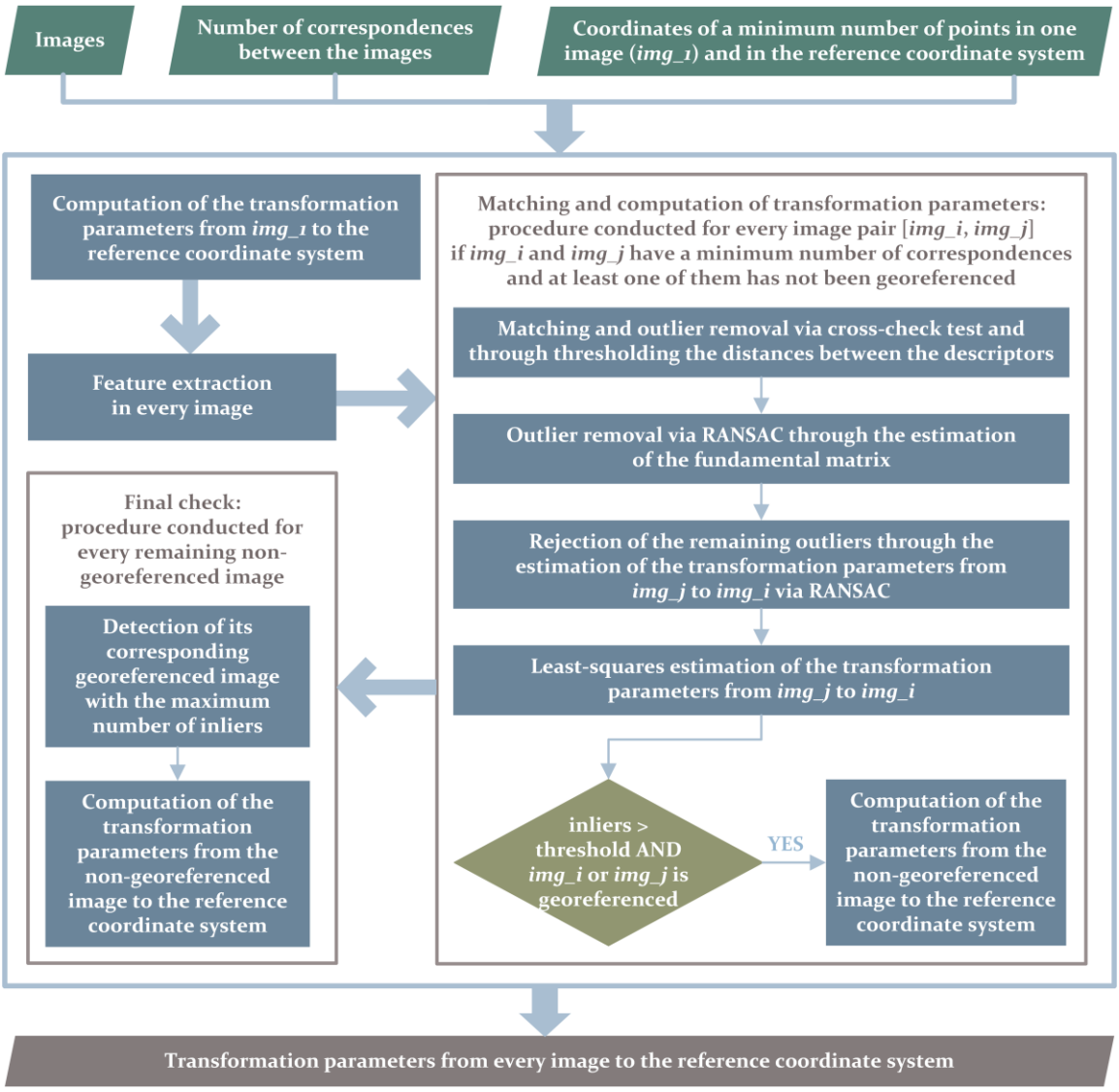


Figure 4-2. The developed workflow for automatic georeferencing of oblique and vertical aerial image datasets

in a Maltese-cross configuration with four oblique cameras, the axes of which were tilted at 35˚ with respect to the vertical, and one nadir camera. The photogrammetric block consists of three strips. The 9000 × 6732 images are free of distortion and correspond to a pixel size of 6 μm. Their calibrated focal length is 53 mm and their GSD varies between 6 and 13 cm. The approximate image overlap in nadir view is 70%. The images cover a total area of about 3 $km^2$ with a relatively low relief throughout most of the area; the elevation difference of about 100 m is concentrated in a small region. The images depict mainly apartment and industrial buildings (with a few high buildings), bridges and a river.

The second dataset, hereinafter referred to as *PentaCam IGI Dataset*, consists of 81 multi-perspective oblique and vertical aerial images taken in a Maltese-cross configuration by a PentaCam IGI camera system with a 50-megapixel sensor from a flying height of about 850 m. The images, with 8176 × 6132 pixels each, have a pixel size of 6 μm. The vertical images correspond to a GSD of 10 cm and a focal length of 50 mm, while the oblique images have a GSD ranging from 8 to 12 cm, a focal length of 80 mm and correspond to a tilt angle of 45˚ from the vertical. The images cover a total area of about 3 $km^2$, which consists of several detached houses and historic buildings, as well as agricultural areas. The elevation difference of this area is about 40 m. The first and second datasets were acquired for the ISPRS scientific initiative "ISPRS/EuroSDR benchmark on high density image matching" (Cavegn et al., 2014; Nex et al., 2015).

The third dataset, hereinafter referred to as *Canon EOS-1Ds Dataset*, consists of an irregular block of 42 oblique airborne images acquired by a Canon EOS-1Ds Mark III camera, which features a 21-megapixel sensor (5616 × 3744 pixels), using a telephoto lens with a focal length of 1200 mm. The images, which have a pixel size of 6.4 μm, correspond to a GSD ranging from 5 to 25 cm. They were taken from a flying height of about 4000 m with a tilt angle of about 60˚ and depict an area of about 4 $km^2$, which consists mostly of apartment buildings, a port, sea and uncultivated land. The terrain of the area changes over its whole extent, having an elevation difference of about 40 m.

### 4.4.2 Developed Software and Experiments

A software suite consisting of five software tools has been developed. The software aims at achieving the automatic rough georeferencing of oblique and vertical aerial images, based on the presented methodology, and relies on this to export additional information in a suitable format. Specifically, the five tools undertake the following operations:

- determining which images overlap;
- rough georeferencing and rectification of images;
- extraction of image footprints;
- subdivision of images into groups; and
- detecting images that cover a specific region.

These tools are intended for computers running Microsoft Windows; they have been developed in the C++ programming language and make use of some functionalities offered by the OpenCV library (OpenCV Team, 2019) for image manipulations, the GeographicLib library (Karney, 2017) for conversion between geographic (latitude, longitude) and Universal Transverse Mercator (UTM) coordinates and the Boost libraries (Boost Team, 2019) for graph creation. A graphical user interface for the software suite has been developed using the Microsoft .Net Framework 4 in the C# programming language, and the C# .Net code calls the C++ applications. In this section, the

applications are presented and the results obtained using the three different datasets of imagery are discussed.

#### 4.4.2.1 Determination of Overlapping Images

A software tool that determines which images overlap one another has been developed based on the methodology discussed in section 4.3.1. The input data consist of the images, without any metadata. The output of the tool consists of the following:

- a file containing the undirected weighted graph that connects the overlapping images;
- a file containing the number of correspondences between each image pair;
- a file containing the image pairs which have a minimum number of matches;
- a file containing a categorization of the image pairs depending on the number of correspondences between the images; and
- a file containing the number of overlapping images for each image.

The results of the tool, as it was tested using the three different datasets presented in section 4.4.1, showed that the overlapping images can be readily detected, even those with very little overlap (of the order of 10%), except for the images that were taken from very different angles, as it is discussed in section 4.4.3.2. SURF feature points proved to be robust for the detection of correspondences between aerial images with significant perspective differences, even when they are resampled to a low resolution (less than 2.5 MP). However, the visual inspection of the detected feature point correspondences, which is also possible through the tool, in combination with the examination of their number, showed that in every image pair which passed through the matching process and consisted of non-corresponding images, 9 to 13 wrong correspondences were found. This was observed regardless of the number of feature points extracted in the images, which ranged from 1000 to 2000, depending on the image content and resolution. Thus, a threshold of a minimum of 20 correspondences (*minMatches*, Table 4-1) has been posed in order to distinguish overlapping images from non-corresponding ones. The matching of two images via this tool took 1 to 2 seconds, depending on the initial size of the images and the number of the detected feature points. The total time spent for the determination of the overlapping images ranged from 17 minutes for the 42 images of the third dataset, to 48 minutes for the 81 images of the second dataset, and 375 minutes for the 150 images of the first dataset. All experiments were performed on a 64-bit Intel Core i7 2.4 GHz laptop computer with 16 GB of RAM.

#### 4.4.2.2 Generation of World Files and Rectification of Images

The aim of this tool is the rough georeferencing of images, so that they can be visualized in a GIS environment in the correct position, orientation and scale. This is achieved through the generation of a world file for each image. The rectification of each image is also required in the case that a transformation of order higher than the affine is used. (A world file is a six-line text file used by a GIS; it stores the six parameters of an affine transformation required for the conversion from an image to the reference coordinate system.) An example of the main window of the tool is illustrated in Figure 4-3, showing its graphical user interface. The input to this application involves: (i) the image dataset; (ii) the starting image, which is the image where the points with known ground coordinates will be measured; (iii) a file containing the ground coordinates of the necessary points in the reference coordinate system; and (iv) a file containing the pixel coordinates of the points of

known ground coordinates (which can be replaced by the direct measurement of these points via this tool).

As a world file stores affine transformation parameters, if an affine transformation is deemed to represent the relationship between two images, as well as that between the images and the ground coordinate system, image rectification is not required via this tool (as it is automatically conducted in a GIS environment). However, if any other transformation defined by more than six parameters is applied, such as a homography or a second order polynomial transformation, image rectification is required. In such cases, the world file stores the parameters of a translation and scaling transformation which verify the set of equations (4.4).

$$\begin{aligned} X &= X_{min} + x_{rec}\, pixelSize \\ Y &= Y_{max} - y_{rec}\, pixelSize \end{aligned} \tag{4.4}$$

In the set of equations (4.4), $X$ and $Y$ are the coordinates of a point in the reference coordinate system, $x_{rec}$ and $y_{rec}$ denote the pixel coordinates of the point ($X$, $Y$) in the rectified image coordinate system, while $X_{min}$ and $Y_{max}$ indicate the minimum $X$ and the maximum $Y$ coordinates among the points of the rectified image in the reference coordinate system. The $y$-scale is defined by the negative value of the pixel size of the rectified image. This is because the origin of the image coordinates is located in the upper left corner but the origin of a map coordinate system is located in the lower left corner.

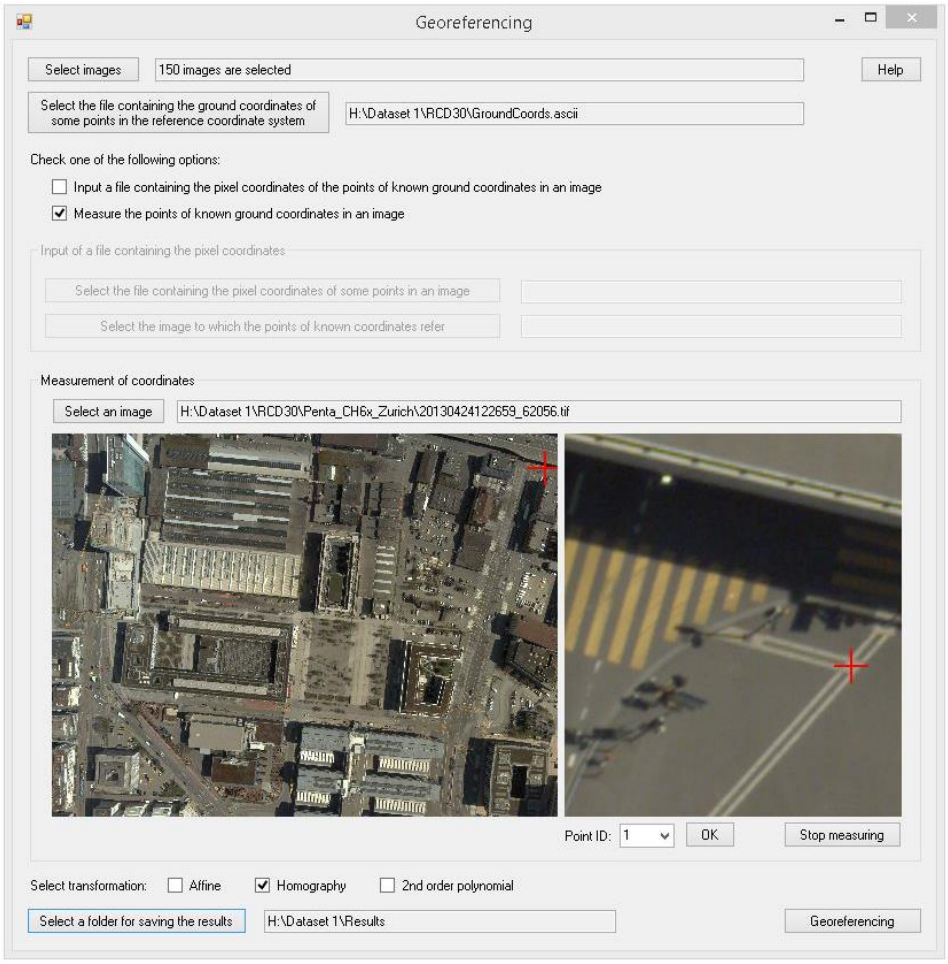


Figure 4-3. Main window of the georeferencing tool

The parameters of the translation ($X_{min}$, $Y_{max}$) and scaling (*pixelSize*, *–pixelSize*) are estimated by implementing the following steps: (i) computation of the *X* and *Y* coordinates of the four corners of the image in the reference coordinate system, using the calculated transformation parameters from the image to the reference coordinate system and the pixel coordinates of the four corners; (ii) estimation of $X_{min}$ and $Y_{max}$ as the minimum and maximum, respectively, between the *X* and *Y* coordinates of the corners of the image in the reference coordinate system; (iii) computation of the pixel size in the *x*-direction of the rectified image by dividing the maximum difference between the *X* coordinates of the corners of the image in the reference coordinate system with the number of columns of the original image; (iv) computation of the pixel size in the *y*-direction of the rectified image by dividing the maximum difference between the *Y* coordinates of the corners of the image in the reference coordinate system with the number of rows of the original image; and (v) computation of the pixel size (*pixelSize*) of the new rectified image as the minimum between the computed pixel size in the *x*-direction and in the *y*-direction.

The image rectification process implements the following steps: (i) computation of the width of the rectangle area that includes the ground footprint of the rectified image as the difference between the maximum and minimum *X* coordinates of the corners of the image in the reference coordinate system; (ii) computation of the height of the rectangle area that includes the ground footprint of the rectified image as the difference between the maximum and minimum *Y* coordinates of the corners of the image in the reference coordinate system; (iii) computation of the number of rows and columns of the rectified image based on the height and width, respectively of the rectangle area that includes the ground footprint of the rectified image divided by its pixel size; (iv) computation of the inverse parameters that express the transformation from the reference coordinate system to the image coordinate system; and (v) for every pixel of the rectified image, computation of the corresponding coordinates in the reference coordinate system, estimation of the pixel coordinates of the original image and image resampling.

The results of this tool were evaluated using the three datasets, by applying the affine transformation, the homography and the second order polynomial transformation for the transfer of the georeferencing information. Only the affine transformation and the homography provided acceptable results, as it is discussed in the following.

As far as the *RCD30 Oblique Penta Dataset* of oblique and vertical aerial imagery is concerned, it requires a transformation which does not preserve the property of parallelism among lines, because the shape of the footprints of the oblique images is trapezoidal and cannot be approximated by a parallelogram. In addition, the elevation difference between the regions depicted in the images is relatively small in comparison with the flying height. For these reasons, the homography yielded satisfactory results for the needs of rough georeferencing. The georeferenced rectified images that were exported by the application were inserted into a GIS environment for visualization purposes (Figure 4-4, left). Figure 4-5 shows details of the overlapping images with their footprints superimposed on them.

Furthermore, the second order polynomial transformation was tested with the *RCD30 Oblique Penta Dataset*. Although at first glance the rectified images seemed to smoothly overlap one another (Figure 4-6, left), they are geometrically distorted and the ground dimensions of the footprints of the images change significantly, as illustrated in Figure 4-6, right. Consequently, this polynomial transformation is unsuitable for the georeferencing of datasets of oblique aerial imagery. The affine transformation did not yield good results because it preserves parallelism, thus being of course inappropriate because of the trapezoidal footprint of the oblique airborne images.

For similar reasons, the homography yielded satisfactory results when applied to the *PentaCam IGI Dataset*, because of the trapezoidal footprint of the oblique images (Figure 4-4, right). The maximum error is observed in agricultural areas, where the SURF algorithm fails to detect a significant number of corresponding points due to the lack of rich texture for image matching, as illustrated in Figure 4-7. The fact that the matching of feature points is not successful in areas like the one depicted in Figure 4-7, limits the applicability of the developed methodology to images that depict mainly urban areas. The georeferencing and rectification of the *PentaCam IGI Dataset* were successful because every image covers not only agricultural areas but also urban ones, which allow the extraction of a sufficient number of feature points and the detection of correspondences.

In the *Canon EOS-1Ds Dataset* of oblique aerial imagery, the affine transformation yielded satisfactory results (Figure 4-8, left) because these images were taken from a very long distance using a high oblique camera and this combination results in almost rectangular footprints. The projective transformation (homography) yielded almost equivalent results with the affine one. The polynomial transformation proved to be inappropriate, as with the first dataset.

The matching of two overlapping images via this first tool took 7 to 20 seconds (on a 2.4 GHz laptop computer with 16 GB of RAM) and the rectification of an image using the homography transformation took 3.5 to 4.5 seconds, depending on the image size. The total time spent on this software tool for the georeferencing of the images ranged from 7 minutes in the case of the *Canon EOS-1Ds Dataset* (42 images), to 35 minutes for the *PentaCam IGI Dataset* (81 images) and 94

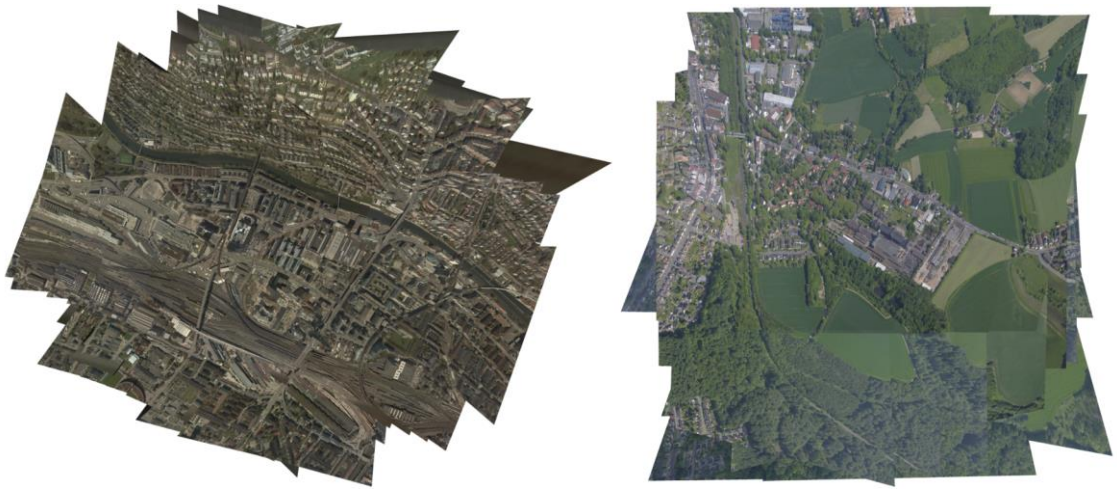

Figure 4-4. Insertion of the rectified georeferenced oblique and vertical aerial imagery of the *RCD30 Oblique Penta Dataset* (left) and the *PentaCam IGI Dataset* (right) in a GIS environment (homography transformation used)

Figure 4-5. Details of the rectified georeferenced images of the *RCD30 Oblique Penta Dataset* with the borders of their footprints superimposed on them in green (homography transformation used)

minutes for the *RCD30 Oblique Penta Dataset* (150 images). These first two software tools (i.e., the tool that determines the overlapping images and the one that generates world files and rectifies images) are the most time consuming ones. The processes executed by the other three tools (extracting image footprints; subdivision of images into groups; and detecting the images that cover a specific region) took no more than 3 seconds.

#### 4.4.2.3 Extraction of Image Footprints

After computation of the transformation parameters from each image coordinate system to the reference coordinate system, the ground coordinates of the corners of each image can be calculated. Thus, another software tool that aims at the extraction of the footprints of images has been developed. This application exports the footprints of the images in Keyhole Markup Language (KML) format, so that they can be superimposed on an earth browser such as Google Earth (Figure 4-8, right). The points of known coordinates are defined in a Cartesian coordinate system, usually in the UTM system, whereas KML files use geographic coordinates in WGS 84 in order to define the ground positions of points. Thus the conversion from the Cartesian coordinates of the corners of image footprints into longitude and latitude is required. The application supports the conversion from UTM to geographic coordinates through the use of the GeographicLib library (Karney, 2017), and requires that the points of known coordinates, which are the input data of the aforementioned software tool, have UTM coordinates, which can be easily obtained from Google Earth for the context of rough georeferencing.

#### 4.4.2.4 Subdivision of Images into Groups Based on their Footprints

Another developed tool aims at the subdivision of the image dataset into groups of images based on the region they cover. The user of this tool may draw a polygon, which approximates a rectangle,

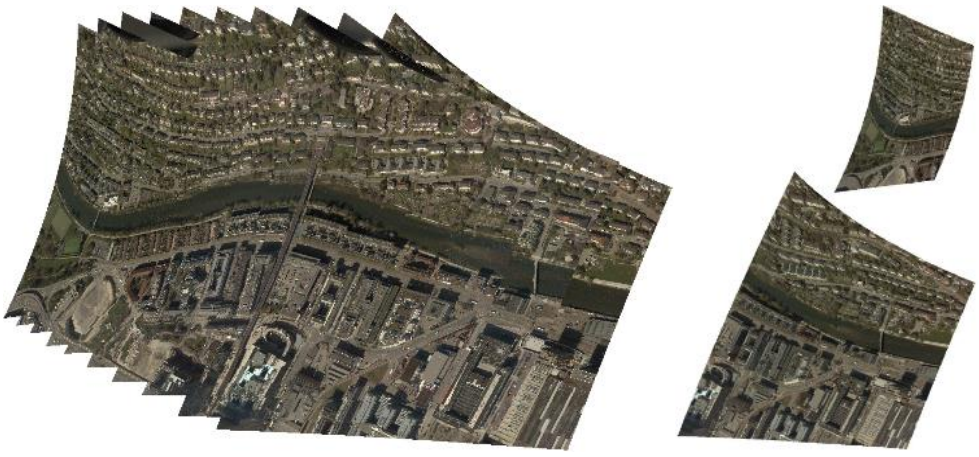

Figure 4-6. Left: a strip of rectified georeferenced oblique aerial imagery from the *RCD30 Oblique Penta Dataset* after the application of a second order polynomial transformation; right: the first (bottom) and last (top) rectified images of this strip, indicating changes in ground footprint

Figure 4-7. Detected correspondences in parts of overlapping images of the *PentaCam IGI Dataset*, depicting an agricultural region

in Google Earth and save it in a KML file. The inputs of the application are this file, in combination with the number of rows and columns into which the polygon will be divided, and either the calculated transformation parameters from each image to the reference coordinate system or the image footprints, if they have been exported using the related software tool that has been developed. This tool divides the polygon into cells and determines in which cell the footprint of each image belongs. Furthermore, it creates a folder for each cell and stores each image footprint, as well as the polygon of the cell, in KML format, in the appropriate folder. In this way, the images that cover a specific ground area can easily be determined (Figure 4-9) for further use.

#### 4.4.2.5 Determination of Images that Cover a Specific Region

In order to automate the process of selecting images that cover a specific region, a tool has been developed. The input data consists of the following:

- a file containing the coordinates of some points that define a polygon, or a KML file representing a polygon drawn in Google Earth, both covering the region of interest; and
- the transformation parameters from each image to the reference coordinate system or the image footprints.

The application exports a file with the names of the images that depict the specific area, which are determined based on the location of the polygon vertices relative to the border of the footprint of each image.

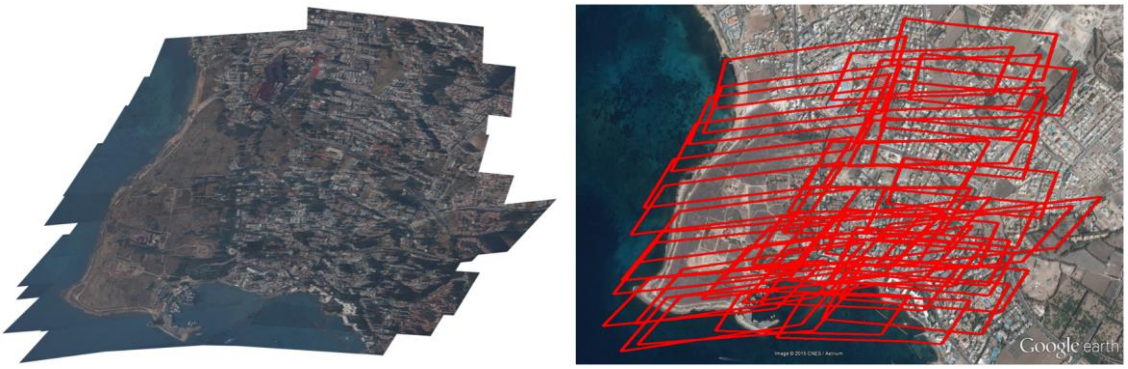


Figure 4-8. Left: insertion of the rectified georeferenced oblique aerial imagery of the *Canon EOS-1Ds Dataset* in a GIS environment (affine transformation used); right: superimposition of the footprints of the *Canon EOS-1Ds Dataset* on Google Earth imagery

Figure 4-9. Subdivision of the *Canon EOS-1Ds Dataset* into groups of images based on their footprints: the user draws the outer green polygon in Google Earth environment; this polygon is automatically divided into cells and each image footprint (in red) is placed in the appropriate cell, using the developed tool

### 4.4.3 Assessment of Results

In this section, the results of the developed methodology are presented in terms of accuracy for every test dataset. The effect of the starting image on the total accuracy is addressed and the characteristics of this image that minimize the georeferencing error are reported. Furthermore, an error analysis depending on the number of transformations applied between an image and the ground reference system is presented. Finally, the performance of the SURF algorithm in multi-view oblique and vertical aerial images is evaluated.

#### 4.4.3.1 Accuracy

In order to estimate the accuracy of the developed methodology, the results were compared with reference data. Specifically, the whole block of 150 images from the *RCD30 Oblique Penta Dataset* was accompanied with reference exterior orientation parameters computed through a highly accurate triangulation (Cavegn et al., 2014). The Agisoft PhotoScan software (Agisoft LLC, 2019) was used to orient the 81 images of the *PentaCam IGI Dataset* via a self-calibrating bundle block adjustment using the coordinates of the camera positions and 12 GCPs, measured in all corresponding images, provided by ISPRS and EuroSDR (Nex et al., 2015). An orthoimage of the region covered by the *Canon EOS-1Ds Dataset* with a GSD of 0.5 m was available.

Several tests were conducted for the georeferencing of the images of every dataset, leading to different results depending on the selection of the starting image, where the points of known coordinates are measured. The tests showed that the developed methodology leads to better results when the points of known coordinates are measured in the image with the maximum number of overlapping images, provided that the imaged terrain relief is relatively low (e.g., a maximum elevation difference of about 25 m for a flying height of 850 m). Generally, such an image lies approximately at the center of the block. However, in the case of a large block where the total block area is much bigger than the area covered by each individual image (such as the *Canon EOS-1Ds Dataset*), the image with the maximum number of overlapping images may not be a central one. In this case, the image with the best combination of a large number of corresponding images, a small elevation difference compared to the flying height and a position close to the center of the block is the starting one. The number of overlapping images for every individual image is an output of the software tool presented in section 4.4.2.1. The points of known coordinates must not be concentrated in a small part of the starting image and may be either ground-level points or points on the top of man-made structures. The ground coordinates of the measured points in each starting image were retrieved from Google Earth; three points were used in the case of the affine transformation (for the *Canon EOS-1Ds Dataset*) and four points in the case of the homography (for the *RCD30 Oblique Penta Dataset* and the *PentaCam IGI Dataset*).

The results of the georeferencing and image rectification procedure achieved using the best starting image with the aforementioned characteristics for every dataset are presented in Table 4-2. Several check points, which were well-distributed over the three blocks of aerial imagery, were measured in a GIS environment in each rectified georeferenced image that depicts them. Their coordinates were compared with the reference coordinates of the same points. The reference coordinates of the check points of the *RCD30 Oblique Penta Dataset* and the *PentaCam IGI Dataset* were computed through photogrammetric space intersection via measurements of these points in the corresponding oriented images. The reference coordinates of the check points of the *Canon EOS-1Ds Dataset* were retrieved from the available orthoimage. The average absolute difference between the $X$ and $Y$ ground coordinates of every point measured in the rectified imagery and the reference $X$ and $Y$

Table 4-2. Accuracy assessment for the test datasets; $N_{meas}$ indicates the number of measurements of each point

| Point ID | $N_{meas}$ | $Avg_{\|\Delta X\|}$ (m) | $Avg_{\|\Delta Y\|}$ (m) | $Max_{\|\Delta X\|}$ (m) | $Max_{\|\Delta Y\|}$ (m) | $Min_{\|\Delta X\|}$ (m) | $Min_{\|\Delta Y\|}$ (m) | $Avg_{Avg\|\Delta X\|}$ (m) | $Avg_{Avg\|\Delta Y\|}$ (m) |
|---|---|---|---|---|---|---|---|---|---|
| ***RCD30 Oblique Penta Dataset*** | | | | | | | | | |
| A1 | 34 | 14.2 | 16.0 | 47.4 | 69.7 | 0.6 | 0.9 | 12.1 | 12.0 |
| A2 | 30 | 7.0 | 5.1 | 17.3 | 19.2 | 0.5 | 0.1 | | |
| A3 | 27 | 4.8 | 6.7 | 27.1 | 53.4 | 0.2 | 0.1 | | |
| A4 | 26 | 8.1 | 4.9 | 29.2 | 12.9 | 0.0 | 0.0 | | |
| A5 | 22 | 6.8 | 5.4 | 41.3 | 30.6 | 0.2 | 0.0 | | |
| A6 | 9 | 37.5 | 35.2 | 81.1 | 66.5 | 8.5 | 21.0 | | |
| A7 | 6 | 49.3 | 65.6 | 216.1 | 130.1 | 11.8 | 49.9 | | |
| A8 | 3 | 30.0 | 15.0 | 49.1 | 32.0 | 16.0 | 3.0 | | |
| ***PentaCam IGI Dataset*** | | | | | | | | | |
| B1 | 51 | 4.0 | 5.6 | 14.7 | 22.8 | 0.0 | 0.0 | 5.0 | 5.5 |
| B2 | 41 | 2.1 | 2.8 | 6.5 | 7.9 | 0.1 | 0.0 | | |
| B3 | 20 | 6.7 | 9.8 | 38.5 | 40.3 | 0.1 | 1.0 | | |
| B4 | 19 | 8.5 | 5.7 | 18.0 | 21.2 | 0.0 | 0.6 | | |
| B5 | 11 | 10.2 | 6.3 | 20.0 | 11.9 | 1.2 | 0.5 | | |
| B6 | 5 | 6.4 | 8.1 | 13.1 | 23.4 | 0.7 | 0.9 | | |
| ***Canon EOS-1Ds Dataset*** | | | | | | | | | |
| C1 | 8 | 8.1 | 7.8 | 8.7 | 14.8 | 6.5 | 3.0 | 26.5 | 13.1 |
| C2 | 6 | 15.2 | 21.0 | 18.0 | 32.1 | 11.4 | 14.3 | | |
| C3 | 6 | 59.5 | 11.7 | 60.5 | 13.5 | 57.7 | 9.2 | | |
| C4 | 5 | 5.9 | 10.0 | 13.4 | 23.4 | 1.0 | 3.4 | | |
| C5 | 5 | 42.0 | 4.6 | 44.2 | 13.2 | 40.1 | 0.8 | | |
| C6 | 4 | 47.9 | 17.1 | 57.2 | 24.2 | 42.3 | 11.3 | | |
| C7 | 3 | 9.0 | 3.3 | 11.4 | 6.1 | 5.4 | 0.1 | | |
| C8 | 3 | 22.1 | 4.4 | 23.3 | 9.2 | 20.5 | 2.0 | | |
| C9 | 3 | 19.2 | 21.5 | 21.5 | 29.6 | 18.0 | 13.1 | | |
| C10 | 2 | 48.2 | 50.8 | 92.4 | 68.7 | 4.0 | 33.0 | | |

coordinates of the same point ($Avg_{|\Delta X|}$ and $Avg_{|\Delta Y|}$), the maximum and minimum differences for every point ($Max_{|\Delta X|}$, $Max_{|\Delta Y|}$, $Min_{|\Delta X|}$ and $Min_{|\Delta Y|}$) and the weighted average absolute differences for all the measured points ($Avg_{Avg|\Delta X|}$ and $Avg_{Avg|\Delta Y|}$) were computed (Table 4-2). $Avg_{Avg|\Delta X|}$ and $Avg_{Avg|\Delta Y|}$ were calculated according to the set of equations (4.5), where *n* is the number of check points and $N_{meas_i}$ is the number of measurements used for the calculation of each $Avg_{|\Delta Xi|}$ and $Avg_{|\Delta Yi|}$.

$$Avg_{Avg|\Delta X|} = \frac{\sum_{i=1}^{n}\left(N_{meas_i} Avg_{|\Delta X_i|}\right)}{\sum_{i=1}^{n} N_{meas_i}} \qquad Avg_{Avg|\Delta Y|} = \frac{\sum_{i=1}^{n}\left(N_{meas_i} Avg_{|\Delta Y_i|}\right)}{\sum_{i=1}^{n} N_{meas_i}} \tag{4.5}$$

The starting images of all the blocks that led to the results presented in Table 4-2 have the maximum number of overlapping images (34, 30 and 12 for the *RCD30 Oblique Penta Dataset*, the *PentaCam*

Table 4-3. Accuracy assessment depending on the choice of starting image; see also Figure 4-10

| Starting image ID | Starting image characteristics (including elevation range *DH*) | $Avg_{Avg\|\Delta X\|}$ (m) | $Avg_{Avg\|\Delta Y\|}$ (m) | Number of measurements |
|---|---|---|---|---|
| | ***RCD30 Oblique Penta Dataset*** | | | |
| 1 | 34 overlapping images, *DH* = 4 m | 12.1 | 12.0 | 157 |
| 2 | 18 overlapping images, *DH* = 4 m | 27.8 | 32.1 | 157 |
| 3 | 29 overlapping images, *DH* = 25 m | 62.5 | 100.5 | 157 |
| 4 | 18 overlapping images, *DH* = 49 m | 91.8 | 158.2 | 157 |
| 5 | 18 overlapping images, *DH* = 74 m | 213.6 | 213.2 | 157 |
| | ***PentaCam IGI Dataset*** | | | |
| 1 | 30 overlapping images, *DH* = 20 m | 5.0 | 5.5 | 147 |
| 2 | 12 overlapping images, *DH* = 13 m | 9.3 | 9.8 | 147 |
| 3 | 19 overlapping images, *DH* = 20 m | 22.7 | 17.0 | 147 |
| 4 | 7 overlapping images, *DH* = 30 m | 93.6 | 38.2 | 147 |
| 5 | 15 overlapping images, *DH* = 32 m | 120.7 | 205.8 | 147 |
| | ***Canon EOS-1Ds Dataset*** | | | |
| 1 | 12 overlapping images, *DH* = 24 m | 26.5 | 13.1 | 45 |
| 2 | 7 overlapping images, *DH* = 18 m | 72.4 | 23.3 | 45 |
| 3 | 8 overlapping images, *DH* = 20 m | 130.3 | 26.4 | 45 |
| 4 | 3 overlapping images, *DH* = 13 m | 110.·1 | 62.1 | 45 |

Table 4-4. Accuracy assessment depending on the number of transformations between the image and the ground coordinate system

| Number of transformations | $Avg_{\|\Delta X\|}$ (m) | $Avg_{\|\Delta Y\|}$ (m) | Number of measurements |
|---|---|---|---|
| 1 | 1.7 | 3.5 | 8 |
| 2 | 5.2 | 5.0 | 197 |
| 3 | 9.6 | 8.4 | 81 |
| 4 | 18.2 | 10.1 | 118 |
| 5 | 25.6 | 18.8 | 121 |
| 6 | 26.3 | 26.3 | 98 |
| 7 | 41.7 | 38.8 | 47 |
| 8 | 53.5 | 56.6 | 25 |
| 9 | 69.6 | 72.0 | 12 |

*IGI Dataset* and the *Canon EOS-1Ds Dataset*, respectively) and a relatively small elevation difference *DH* (4 m, 20 m and 24 m, respectively).

For comparison reasons, the same check points were measured in the same georeferenced images of every dataset, which were generated according to the developed methodology using different starting images. The results are presented in Table 4-3, in the form of $Avg_{Avg|\Delta X|}$ and $Avg_{Avg|\Delta Y|}$. The results indicate that the developed methodology for georeferencing and image rectification may be applied provided that the starting image has a large number of corresponding images, a small elevation difference and a central (rather than a more peripheral) position in the block. The footprints of the starting images that were tested for every dataset are illustrated in Figure 4-10. In general, the georeferencing error increases significantly when the area covered by the starting image has a large elevation difference (e.g. see Table 4-3, stating image 5 for the *RCD30 Oblique Penta Dataset* and starting image 5 for the *PentaCam IGI Dataset*). This is an expected outcome, since

the transformation of the starting image largely determines the georeferencing of the other images of the dataset. Whereas a starting image depicting a flat area and corresponding to a large number of overlapping images and a central position in the block would be ideal, in some cases this is not possible, as the areas of low relief may be concentrated, for instance, near the borders of the block. In such cases, the most significant criterion for the choice of the starting image is the low relief, since in the case that the starting image transformation does not represent its image-to-ground transformation with adequate accuracy, the georeferencing error of all the other images that depend on this first transformation would be obviously large.

Furthermore, the error of a rectified image increases with the number of transformations applied between this image and the ground reference system. This assumption is verified by conducting several measurements of check points in the georeferenced images of the test datasets. The results are shown in Table 4-4, in the form of average absolute differences between the measured coordinates of points in the rectified images and their reference coordinates for each group of images that correspond to the same number of transformations from the ground coordinate system. Generally, the smaller the ratio of the total area of the block to the area of one image, the higher the accuracy to be expected because fewer transformations are applied. This is the reason why the georeferenced rectified images of the second dataset (Figure 4-10, top-right) correspond to the smallest georeferencing error (Table 4-2). The fact that the error of a rectified image increases with the number of transformations also verifies the conclusion derived from Table 4-3, that a large number of overlapping images is generally required for the starting image, as this large number of overlapping images implies a smaller total number of transformations.

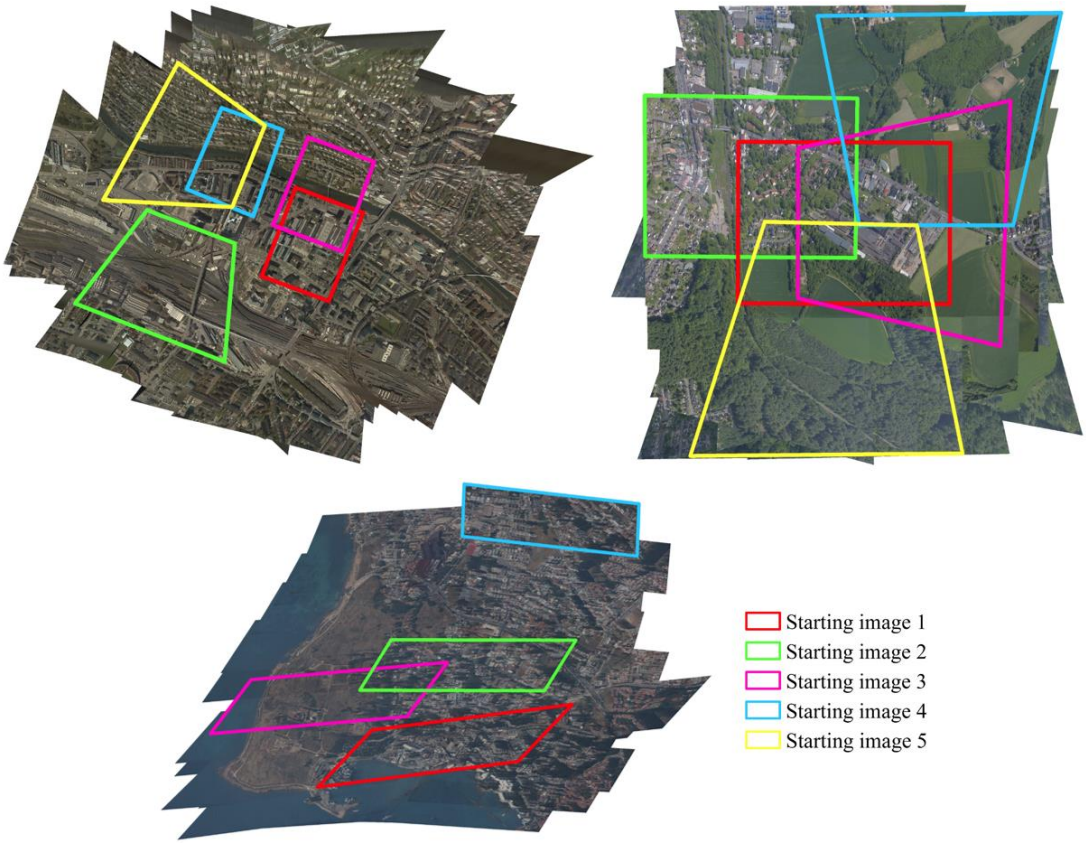


Figure 4-10. Footprints of the starting images that led to the results presented in Table 4-3, superimposed on the georeferenced rectified blocks of images for the *RCD30 Oblique Penta Dataset* (top-left), the *PentaCam IGI Dataset* (top-right) and the *Canon EOS-1Ds Dataset* (bottom)

Generally, an average absolute difference of the order of 50 m between the $X$ and $Y$ ground coordinates of points in the rectified imagery and their reference coordinates is acceptable for applications that require a rough georeferencing of small-scale aerial image datasets, such as archiving of images without any metadata (e.g., old datasets) and automated selection of images based on the area they cover. The developed georeferencing framework covers the accuracy requirements of such kinds of applications (e.g., see Table 4-2).

#### 4.4.3.2 Performance of SURF in Multi-View Aerial Images

The performance of SURF under scale changes, image rotation, image blur, addition of noise, illumination changes and viewpoint changes has been evaluated by several researchers (Bauer et al., 2007; Juan and Gwun, 2009; Khan et al., 2011; Wu et al., 2013; Srivastava, 2015). Viewpoint changes, which can be approximated either by affine transformations or, more generally, by perspective transformations, are the most difficult type of transformations. Observing the scene

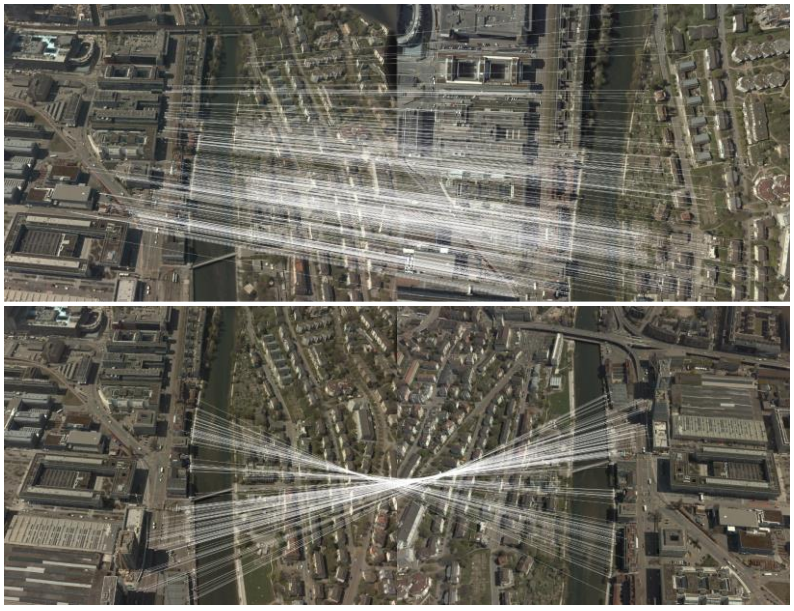

Figure 4-11. Results of the matching procedure using the SURF feature point detector and descriptor for the *RCD30 Oblique Penta Dataset* under viewpoint changes of 35˚ (top) and 70˚ (bottom)

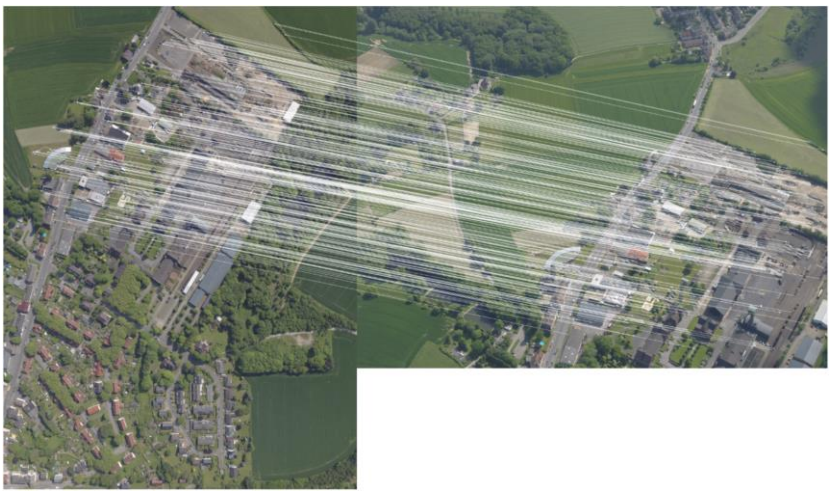

Figure 4-12. Results of the matching procedure using the SURF feature-point detector and descriptor in the *PentaCam IGI Dataset* under a viewpoint change of 45˚

from a different angle has a large impact on the repeatability of the SURF algorithm and thus on the matching procedure. Whereas SURF is rotation invariant, its robustness decreases significantly under large viewpoint changes for non-planar surfaces in natural scenes. However, SURF has a good repeatability when the viewpoint change is small. For instance, according to the literature relating to terrestrial images of natural scenes, the total number of matches is greatly reduced even under viewpoint changes of 10˚, while the percentage of good matches drops to zero for viewpoint changes of more than 40˚ (Bauer et al., 2007). The performance of this algorithm in matching vertical and multi-view oblique aerial images has not been sufficiently investigated yet, except for some simple applications in limited numbers of image pairs (e.g., Floridi, 2016).

The SURF algorithm yielded satisfactory results in matching oblique and vertical aerial images of the test datasets. Valid correspondences were detected in the overlapping oblique and vertical aerial images of both the *RCD30 Oblique Penta Dataset* that correspond to a viewpoint change of 35˚ (Figure 4-11, top) and the *PentaCam IGI Dataset* that correspond to a viewpoint change of 45˚ (Figure 4-12). Furthermore, the matching of multi-perspective oblique aerial images was conducted satisfactorily, provided that the viewpoint change was only up to 70˚ (Figure 4-11, bottom). The matching of multi-view oblique aerial images of the *PentaCam IGI Dataset* that correspond to a viewpoint change greater than 70˚ was not accomplished because the developed software tool that determines the overlapping images (section 4.4.2.1) failed to detect valid correspondences in the low-resolution copies of such images. However, the georeferencing procedure was accomplished without any problem because SURF feature points were matched between oblique and vertical aerial images as well as between oblique images that correspond to the same perspective.

An important characteristic of the SURF feature points that were detected as correspondences between multi-view aerial images is the fact that they all lie in planar surfaces. More than 95% of them consists of ground points, while only very few points lie on the top of planar constructions. Thus, the robustness of SURF was practically eliminated under viewpoint changes equal to or greater than 35˚ for non-planar objects depicted in the aerial images of the test datasets. Nevertheless, the matching procedure was accomplished because the images contain some planar surfaces not occluded by obstacles, where the repeatability of SURF is not affected.

## 4.5 Conclusions

The georeferencing of oblique and vertical aerial images is an issue of great importance for a wide range of applications. In this chapter, an automatic methodology for the rough georeferencing of datasets of multi-perspective oblique and vertical aerial imagery, without the need for GNSS/INS data or ground elevation data, has been presented; a related software suite that has been developed and consists of five interdependent tools has been outlined; and the results using three different datasets have been discussed. The georeferencing methodology uses a 2D transformation to express the relation between every image and the ground reference system. It is based on image matching and the transfer of the georeferencing information among images and requires the input of horizontal ground coordinates of some points measured in a central image of the block with a large number of overlapping images and a small elevation difference throughout the area it covers. It can be applied in images taken by frame cameras from an aerial platform depicting mainly urban regions, which contain rich textural information. Images depicting large agricultural areas with repetitive patterns, areas with ice and snow cover or regions covered by water surfaces or sand may cause the image matching process to fail, thus being inappropriate for the developed methodology.

The SURF feature detector proved to yield satisfactory image matching results in multi-perspective image datasets with rich texture.

The results of the developed methodology depend greatly on the definition of the proper transformation between the starting image and the ground reference system, as well as between the other images, for the transfer of the georeferencing information. A homography yields satisfactory results in cases of small elevation differences in the regions depicted in the images compared with flying height. An affine transformation may be applied in the simpler cases of image footprints that can be approximated by a rectangle. However, these transformations are not applicable if the images cover regions with large elevation differences. Obviously, the georeferencing error increases with the number of transformations applied between an image and the ground reference system.

The main source of errors of the georeferencing methodology presented in this chapter is the adoption of a 2D transformation model between images using feature points that may lie on different planes. As a result, whereas this method is applicable in image datasets that are characterized by low elevation differences, it cannot be implemented in large-scale datasets of man-made environments that depict multiple planes in different elevations. Thus, an extension of the developed georeferencing methodology applicable in large-scale images of a piecewise planar scene, that constrains feature point correspondences among all images to lie on the plane where the GCPs lie, provided that the latters are coplanar, is discussed in section 5.4 of this dissertation. The purpose of such an additional investigation for large-scale images of man-made structures is to reduce the error of the computed image-to-ground 2D homography transformation caused by using feature points that lie on planes different from the one where the GCPs lie.

One innovative point of the developed image-based georeferencing framework outlined in this chapter is its application in datasets containing multiple oblique aerial images without availability of GNSS/INS data from onboard sensors and camera characteristics (focal length) or interior orientation parameters, as the available rough georeferencing techniques for such kinds of images are applied in single imagery accompanied with such kinds of data. Furthermore, the investigation of the performance of the SURF feature extractor in image matching of oblique aerial images is another aspect of this chapter, as this algorithm has generally not been used by the research community for matching of oblique images. As far as the existing 2D georeferencing methods applied for vertical aerial images are concerned, the non-requirement for GNSS/INS data along with the necessity for a minimum number of single-image horizontal GCP measurements constitute advantages of the developed framework. Finally, the whole developed software suite consists of a combination of useful tools for image rectification, extraction of footprints, subdivision of images into groups based on the area they cover and determination of those ones that cover a specific area; this combination of tools automates several applications for georeferencing, archiving and searching among large datasets and is not found in existing software packages.

The developed framework can also be applied for georeferencing of large collections of old/historic images. Such images are usually not accompanied with georeferencing information from onboard sensors and camera calibration reports as well as GCPs are hard to be found for such images. Hence, real world 2D ground coordinates can be allocated in the imagery based on the developed georeferencing technique with far less time and labor in comparison with a traditional manual georeferencing technique used by a GIS platform.

# Chapter 5
# Detection of Coplanar Points

## 5.1 Introduction

Planar structures are very common in man-made environments. Thus, the majority of both oblique and vertical large-scale images of urban scenes, usually acquired by UAVs, depict more than one planes (e.g., planar roofs of buildings, building facades, planar ground) with sufficient resolution so as to be visually identifiable. In this chapter, a technique for detecting coplanar feature points that lie on a user-defined plane among oblique and vertical large-scale aerial imagery depicting a piecewise planar scene is introduced. The technique is based on a template matching method in combination with a RANSAC scheme to iteratively estimate the homography transformation between the plane of interest as depicted in the overlapping images of image pairs and works incrementally by increasing the number of the detected feature points in each image and transferring the plane information among all images. Two georeferencing frameworks that implement the developed coplanar point detection scheme are discussed, namely:

- the georeferencing of image datasets of a piecewise planar scene via computation of their 2D homography transformation parameters from the reference coordinate system; and
- the estimation of approximate exterior orientation parameters of georeferenced image datasets of a piecewise planar scene.

## 5.2 Related Work

Several researchers have focused on the detection of planar structures in image sequences, because of their simplicity and attractive geometric properties combined with their abundance in urban areas and generally in man-made environments. Their simplicity relies on the fact that the transformation between a real world plane and its corresponding image plane as well as the transformation between perspective views of a plane in different images are linear. Coplanar points and planar regions impose strong geometric constraints for several photogrammetric tasks. They have been successfully employed in diverse image-based applications, like SfM and 3D reconstruction

(Szeliski and Torr, 1998; Bartoli et al., 2001; Kaucic et al., 2001; Bartoli and Sturm, 2003; Fraundorfer et al., 2006; Zhou et al., 2012), autonomous robot/vehicle navigation (Ohnishi and Imiya, 2006; Conrad and DeSouza, 2010; Nishida et al., 2011), obstacle detection (Lourakis and Orphanoudakis, 1998), object recognition and tracking (Mondragón et al., 2010) and camera self-calibration (Malis and Cipolla, 2002; Zhou et al., 2012).

Several methods for detecting coplanar points and planes have been proposed so far using robust homography estimation techniques, such as RANSAC-based model fitting methods (Vincent and Laganiere, 2001; Schindler, 2003; Gorges et al., 2004; Fraundorfer et al., 2006; Kähler and Denzler, 2007; Fouhey et al., 2010; Mondragón et al., 2010; Nishida et al., 2011; Zhou et al., 2012), maximum likelihood estimators (Bartoli et al., 2001), expectation maximization methods (Conrad and DeSouza, 2010), the J-linkage method (Fouhey et al., 2010) introduced by Toldo and Fusiello (2008), and least median squares (LMedS) techniques (Lourakis and Orphanoudakis, 1998; Lourakis et al., 2002; Odone et al., 2002). The RANSAC algorithm provides an efficient framework for the estimation of the projective transformation between points lying on image planes in cases of data containing many outliers and is also used by the developed approach. Among the RANSAC-based approaches, some of them work purely on sparse sets of corresponding feature points for homography computation (e.g., Vincent and Laganiere, 2001; Gorges et al., 2004; Kähler and Denzler, 2007; Zhou et al., 2012), while other methods are based on optical flow detection (e.g., Mondragón et al., 2010; Nishida et al., 2011). Schindler (2003) and Fraundorfer et al. (2006) apply a computationally more intensive region growing scheme to sparse sets of correspondences in order to delineate, expand and refine the detected planar regions. The method introduced in this chapter relies solely on a sparse set of correspondences to detect coplanar feature points and the homography induced by the plane of interest among images.

Some RANSAC-based approaches search for a dominant homography that verifies the relationship between the largest number of corresponding points in two views (e.g., Mondragón et al., 2010; Nishida et al., 2011; Zhou et al., 2012). Other methods group the detected feature points into coplanar sets that share a common homography for the detection of multiple planes (e.g., Vincent and Laganiere, 2001; Schindler, 2003; Gorges et al., 2004; Fraundorfer et al., 2006; Kähler and Denzler, 2007). The method introduced in this chapter works by iteratively grouping corresponding coplanar feature points into sets based on the plane where they lie, with the ultimate goal of detecting feature points that lie on a user-defined plane. In this way, whereas it also detects points that belong in several planes, the developed algorithm goes one step further by choosing among the coplanar point sets the one that belongs to a specific plane of interest. In this context, its main difference compared to the existing RANSAC-based approaches is the combination of a template matching method with the RANSAC algorithm in an iterative framework for the detection of feature points that lie on a specific user-defined plane and not on the dominant one, or each dominant one that can be iteratively detected by RANSAC.

## 5.3 Developed Method for Detecting Coplanar Points

In this section, the developed algorithm for the automatic detection of feature points that lie on a specific image plane is described. The prerequisites for the implementation of the developed algorithm in an image dataset of a piecewise planar scene are (i) the image coordinates of four coplanar points in a single image that lie on the plane of interest; (ii) feature point correspondences between the overlapping images of the dataset; and (iii) the fundamental matrices between the overlapping images. It is not necessary for all images to depict the same part of the planar surface

that is defined through the use of the four image points and is visible in the image where these points have been measured. However, all images have to depict the plane of interest where the coplanar points will be detected. In addition, at least two images of the dataset should depict the same part of the planar surface that is defined by the four coplanar points, i.e., these coplanar points should be visible in at least two images. Also, for the successful detection of coplanar points in an image pair, the images should correspond to a similar perspective, i.e., an image-to-image translation transformation should roughly express their relationship. Thus, the developed methodology may be implemented in aerial image datasets whereby the angular orientation of the camera remains approximately the same or changes gradually from image to image. Generally, it may be adopted by datasets containing successive overlapping aerial images taken by the same camera. However, it cannot be implemented in multi-perspective image datasets taken by a multi-camera system whereby the oblique and vertical images or the oblique ones are taken by different cameras with different viewing angles. In such cases, the developed methodology may be applied in each set of images taken by the same camera of the multi-camera system. The flowchart of the developed method for detecting homologous coplanar points in an image pair, as described in the following paragraphs, is illustrated in Figure 5-1. The image of the pair where at least four coplanar points have been identified is hereinafter symbolized as *img_i* and the other image of the pair is hereinafter symbolized as *img_j*. Table 5-1 presents indicative values for the different parameters and thresholds that can be used by the developed methodology.

### 5.3.1 Template Matching

The first step of the developed method for detecting coplanar points in an image pair is a template matching process, which may be successfully implemented provided that the images of the pair correspond to a similar perspective. In order to keep the computational time of the whole process of detecting homologous coplanar points at a reasonable level, a maximum number of points (*numPoints*) that will be involved in the template matching process is initially defined. This number equals to a maximum threshold (*maxPointsThres*), if the already identified coplanar points of *img_i* are more than this threshold; otherwise, if less than *maxPointsThres* coplanar points have been identified in *img_i*, *numPoints* equals to the number of coplanar points of *img_i*.

A template window, hereinafter symbolized as *template*, is defined around the *numPoints* coplanar points of *img_i* that lie on the plane of interest. In the case that more than *maxPointsThres* coplanar points have been identified in *img_i*, *numPoints* coplanar points of *img_i* are automatically selected so as to be evenly distributed throughout the whole area of the coplanar points of *img_i*. The homologous window of each one of the *numPoints* points of *img_i* is estimated in *img_j* through template matching. According to this method, *template* slides pixel to pixel in *img_j*, left to right and up to down. At each location of this sliding window, a metric is calculated that represents the similarity between *template* and *img_j* and is stored in a result matrix $\boldsymbol{M}$. The metric used by the developed algorithm is computed using equation (5.1), where *i* and *j* represent the row and column number of *img_j*, while *i΄* and *j΄* represent the row and column number of *template*.

$$\boldsymbol{M}(i,j)=\frac{\sum_{i',j'}\left[template(i',j')\cdot img\_j(i+i',j+j')\right]}{\sqrt{\sum_{i',j'}template(i',j')^2\cdot\sum_{i',j'}img\_j(i+i',j+j')^2}} \tag{5.1}$$

After normalization of the matrix $\boldsymbol{M}$ in the range [0, 1], the best match is found as a global maximum. The top left corner of the homologous window of *template* in *img_j* is the location of

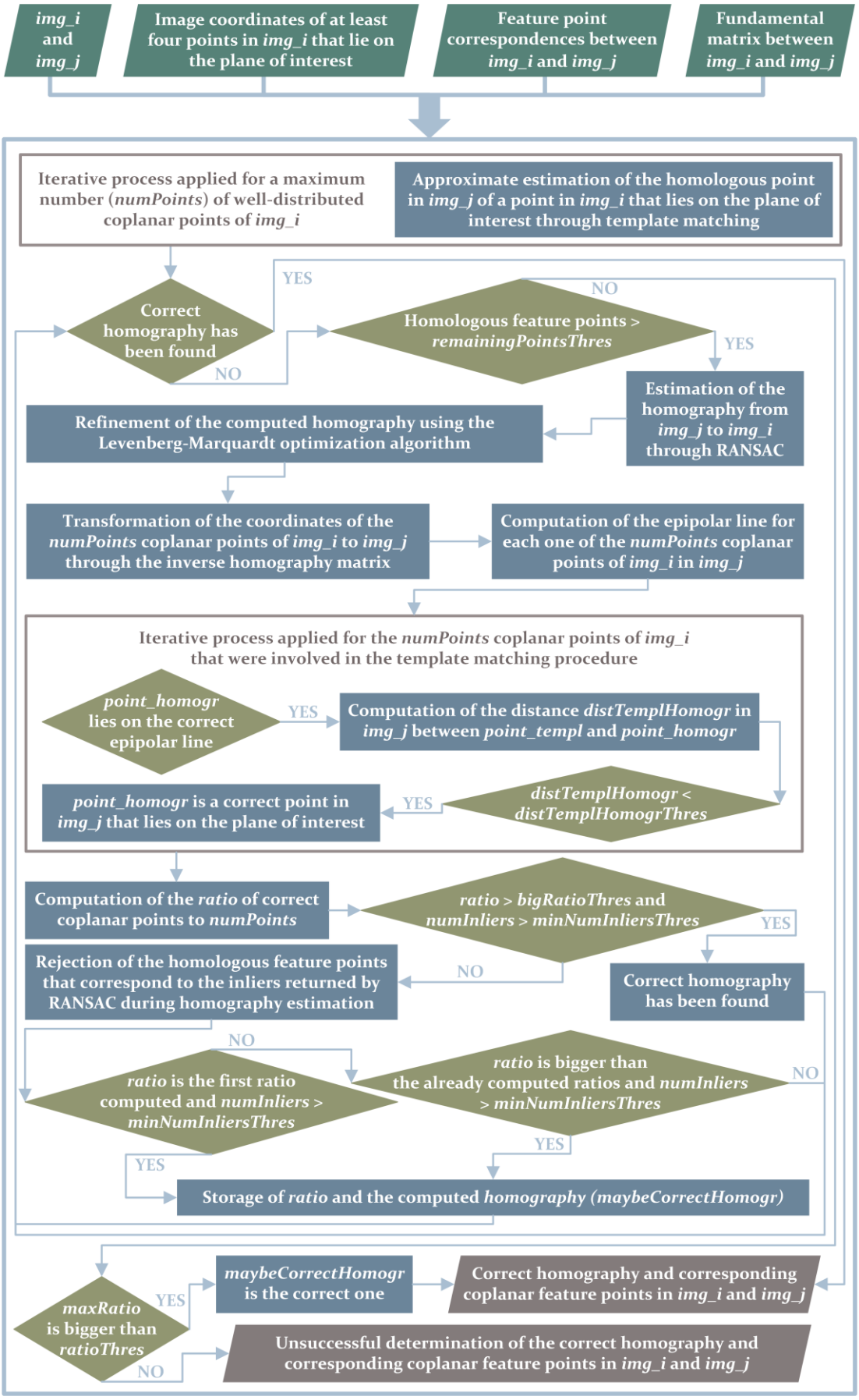


Figure 5-1. Flowchart illustrating the developed coplanar point detection methodology

Table 5-1. Variables and parameter values used in the developed coplanar point detection algorithm

| Symbol | Value |
|---|---|
| *maxPointsThres* | 100 |
| *numPoints* | min[*maxPointsThres*, number of coplanar points of *img_i*] |
| *homographyThres* | 1 pixel |
| *distFromEpipolarLineThres* | 5 pixels |
| *distTemplHomogrThres* | 15 pixels |
| *minNumInliersThres* | 70 |
| *bigRatioThres* | 0.9 |
| *ratioThres* | 0.6 |
| *remainingPointsThres* | 70 |

the global maximum and its width equals to the width of *template*. This template matching procedure applies for the *numPoints* coplanar points of *img_i*. The center point of each rectangle detected in *img_j* as the homologous window of each *template* is assumed to be the rough homologous point of each one of the *numPoints* coplanar points of *img_i* in *img_j*, found through template matching.

### 5.3.2 Detection of Coplanar Feature Points

An iterative procedure (*iter_process_1*) is implemented for the estimation of the homography induced by the plane of interest between *img_i* and *img_j*. Specifically, a homography transformation from *img_j* to *img_i* is fitted to the corresponding feature points via RANSAC using a small distance threshold (*homographyThres*) between the extracted feature points in *img_i* and the projected ones via homography, so that the inliers that verify the computed transformation belong to a plane. The homography computed via RANSAC using the best combination of four corresponding feature points that yields the largest number of inliers is then refined through the Levenberg-Marquardt (Moré, 1978) optimization algorithm using all inliers detected by RANSAC. Subsequently, the pixel coordinates of the *numPoints* coplanar points that were involved in the template matching process are transformed from *img_i* to *img_j* using the inverse homography matrix. Then, the epipolar line $\boldsymbol{l}_j$ in *img_j* that corresponds to each one of the *numPoints* coplanar points $\boldsymbol{p}_i$ of *img_i* is determined using the fundamental matrix $\boldsymbol{F}$, according to equation (5.2). The epipolar line is expressed by three coefficients ($a$, $b$, $c$) that satisfy the equation $ax + by + c = 0$; the line coefficients are defined up to a scale, being normalized so that $a^2 + b^2 = 1$. Each point is expressed in pixels in homogeneous coordinates.

$$\boldsymbol{l}_j = \boldsymbol{F} \cdot \boldsymbol{p}_i \quad \text{where } \boldsymbol{l}_j = \begin{bmatrix} a & b & c \end{bmatrix}^T \text{ and } \boldsymbol{p}_i = \begin{bmatrix} x & y & 1 \end{bmatrix}^T \tag{5.2}$$

An iterative process (*iter_process_2*) is implemented after the computation of the epipolar lines, as a loop inside *iter_process_1*, for checking whether the computed homography transformation is the correct one, i.e., whether it expresses the relation between the perspective views of the plane of interest in *img_i* and *img_j*. The number of iterations of *iter_process_2* equals to the number of coplanar points involved in the template matching process (*numPoints*), as one point is used in the procedure implemented in each iteration. The approximate estimation in *img_j* of the point $\boldsymbol{p}_i$ of *img_i* obtained through template matching, which is used in an iteration, is hereinafter symbolized

as *point_templ*. The estimation of the corresponding point in *img_j* obtained using the calculated inverse homography matrix is hereinafter symbolized as *point_homogr*.

The first step of *iter_process_2* is the computation of the distance (*distFromEpipolarLine*) between *point_homogr* and the corresponding epipolar line of $\boldsymbol{p_i}$ in *img_j*. If this distance is lower than a threshold (*distFromEpipolarLineThres*), *point_homogr* is considered to lie on the correct epipolar line, estimated according to equation (5.2), and the distance (*distTemplHomogr*) in *img_j* between *point_homogr* and *point_templ* is calculated. If *distTemplHomogr* is smaller than a threshold (*distTemplHomogrThres*) the homography-based estimation of $\boldsymbol{p_i}$ in *img_j* is considered to be the homologous point of $\boldsymbol{p_i}$ in *img_j* and *point_homogr* is considered to be a point that lies on the plane of interest in *img_j*. Otherwise, if *distFromEpipolarLine* is equal to or bigger than *distFromEpipolarLineThres,* or if *distTemplHomogr* is equal to or bigger than *distTemplHomogrThres*, *point_homogr* is not considered to be a correct match with $\boldsymbol{p_i}$.

After the termination of *iter_process_2*, when all the *numPoints* coplanar points of *img_i* have been involved in *iter_process_2*, the ratio (*ratio*) of correct matches that lie on the plane of interest to *numPoints* is computed. If *ratio* is bigger than a big threshold (*bigRatioThres*), the computed homography is considered to be the correct one that expresses the relation between the perspective views of the plane of interest in *img_i* and *img_j*, provided that the number of inliers (*numInliers*) identified by RANSAC that verify the computed homography is bigger than a threshold (*minNumInliersThres*). In this case, *iter_process_1* is terminated. If *numInliers* is equal to or less than *minNumInliersThres*, the homography has been computed using an insignificant number of correspondences, thus being prone to inaccurate estimation. If ratio is smaller than *bigRatioThres*, the computed homography possibly represents the relation between points that belong to another plane visible in *img_i* and *img_j*. In this case, points that are returned as inliers by RANSAC are removed from the set of matches so that in the next iteration of *iter_process_1* they will not be involved in the homography estimation procedure. If *ratio* is bigger than the previously computed ratios or if it is the first computed ratio during the first iteration of *iter_process_1*, it is stored by the algorithm along with the computed homography transformation (*maybeCorrectHomography*) that corresponds to it, provided that the number of inliers returned by RANSAC that verify the computed homography is bigger than *minNumInliersThres*. The iterative process *iter_process_1* is terminated if the correct homography has been found or if the number of remaining correspondences between *img_i* and *img_j* is equal to or smaller than a threshold (*remainingPointsThres*). In the case that the correct homography has not been detected through *iter_process_1*, if *ratio* with the maximum value that corresponds to *maybeCorrectHomography* is bigger than a threshold (*ratioThres)*, which is smaller than *bigRatioThres, maybeCorrectHomography* is considered to be the correct one. The reason why two thresholds have been defined for the ratio of correct matches that lie on the plane of interest to *numPoints,* is to avoid unnecessary iterations in *iter_process_1* for the detection of the homography transformation that yields the maximum number of inliers when the correct homography, i.e., the one that corresponds to *bigRatioThres*, has been found.

The outputs of the whole process, if the correct homography has been detected, is the homography induced by the plane of interest in *img_i* and *img_j* and the corresponding feature points in *img_i* and *img_j* that verify the correct homography and lie on the plane of interest. Otherwise, if *maybeCorrectHomography* is smaller than or equal to *ratioThres*, neither the correct homography nor the feature points that lie on the plane of interest can be determined. In this case, the whole procedure may be repeated using more relaxed thresholds (e.g., using a double value of *homographyThres*, a value for *distTemplHomogrThres* increased by 30% and a value for

*minNumInliersThres* reduced by 20%) until the correct homography and the coplanar points of interest that verify the correct homograhy are found.

## 5.4 Georeferencing of Image Datasets of a Piecewise Planar Scene

In this section, the developed methodology for rough georeferencing of datasets of multi-view oblique and vertical aerial imagery of a piecewise planar scene, in the case that GNSS/INS information is not available, is presented. It is based on the method presented in section 4.3 of this dissertation, with some variations in the required input data and the methodology that is implemented. These variations are outlined in the following.

- Four coplanar points of known horizontal ground coordinates have to be measured in a starting image.
- The images are not downscaled for the image matching process, so that a much bigger number of feature points is extracted in each image, which is required for the successful detection of coplanar points by the developed method.
- Different hessian thresholds are used by the SURF feature detector in the step of image-based determination of overlapping images as well as in the image matching step for overlapping images. Specifically, *hessianThres1*, which is used in the process of determining overlapping images, is larger than the one used in section 4.3, with the purpose of extracting fewer correspondences, as the coplanar point detection process detects coplanar points in highly overlapping images; thus, poorly overlapping images are not used by the developed method and the matching of an image with the maximum number of overlapping images is not the goal of this method. On the contrary, because of the absence of such a constraint imposed by coplanar points by the georeferencing method presented in section 4.3, the extraction of correspondences between images that are not highly overlapping not only can be implemented but also it is required for obtaining more accurate results, as the error of a georeferenced image increases with the number of transformations applied between this image and the ground reference system (Table 4-4). As far as *hessianThres2* is concerned, i.e., the SURF threshold used during the image matching step, it is bigger than the value proposed in section 4.3, as the images are not downscaled and a bigger threshold would result in too many feature points.
- The corresponding feature points that lie on the plane defined by the GCPs (*planeOfGCPs*) in the starting image are used for transferring the geospatial information among images during image registration.
- The homography induced by the *planeOfGCPs* between the overlapping images of each pair is computed.
- The homography that expresses the relationship between the *planeOfGCPs* in each image and the coordinate reference system is computed.
- A smaller value for *matchesThresForGeoreferencing* is used. This is the threshold of correspondences between a pair of images, which ensures that georeferencing information is transferred between images with a sufficient number of matches. If the number of valid correspondences of an image pair is bigger than *matchesThresForGeoreferencing*, the homography parameters between the non-georeferenced image and the reference coordinate system are computed. This threshold is smaller compared to the one defined in section 4.3, because the detected correspondences lie in one specific plane (*planeOfGCPs*) and may be less than the ones detected by the methodology presented in section 4.3, which are not restricted to lie on a specific plane.

Table 5-2. Parameter values used by the developed georeferencing algorithm for images of a piecewise planar scene that differ from the ones outlined in Table 4-1

| Symbol | Step | Value |
|---|---|---|
| *resizingFactor2* | Image matching | 1 (no resizing) |
| *hessianThres1* | Determination of overlapping images | 700 |
| *hessianThres2* | Image matching | 500 |
| *matchesThresForGeoreferencing* | Georeferencing | 30 |

The main reason for the implementation of the developed georeferencing method in large-scale datasets of a piecewise planar scene is the reduction of the error that would arise using the methodology presented in section 4.3 due to the estimation of the homography between image pairs using feature points that lie on different planes, thus causing error accumulation. Whereas in datasets of smaller scale without large elevation difference compared to the flying height, like the ones (section 4.4.1) used for testing the method presented in section 4.3, satisfying results are achieved for rough georeferencing purposes, the method presented in section 4.3 would result into large georeferencing errors for datasets that are characterized by large elevation difference compared to the flying height and multiple different planar surfaces. Thus, the constraint imposed by the extraction of coplanar feature points that lie on the plane of GCPs in all images is necessary for such kinds of datasets.

The parameter values used by the methodology presented in this section for rough georeferencing of datasets of oblique and vertical aerial imagery of a piecewise planar scene, which are different from the ones used by the georeferencing framework presented in section 4.3 (Table 4-1), are outlined in Table 5-2.

### 5.4.1 Image-Based Determination of Overlapping Images

In order to minimize the processing time of the main procedure of image matching, the overlapping images are determined in a first step through image matching using the SURF algorithm, according to the methodology presented in section 4.3.1 of this dissertation.

### 5.4.2 Image Matching

SURF feature points are extracted in all images of original size. Feature-based matching is implemented in each pair of overlapping images if at least one image of the pair has not been georeferenced, applying a cross-check test and a thresholding of the Euclidean distances between the descriptors of the points considered to be homologous. The RANSAC algorithm is applied for outlier removal (i) through estimation of the fundamental matrix using the eight-point algorithm and (ii) through homography estimation to reject the remaining erroneous matches that happen to lie on the epipolar line of the homologous feature point under consideration.

### 5.4.3 Georeferencing

The georeferencing of the images is based on the measurement of at least four points of known ground coordinates in only one image and the transfer of the geospatial information among images through computation of the relative position of the images using only correspondences that lie on the plane of the GCPs.

After the end of the image matching process for a pair of overlapping images, the coplanar point detection method described in section 5.3 is implemented for computation of the homography transformation between the images of this pair induced by the plane where the GCPs lie, using only corresponding feature points that lie on the plane of the GCPs. If one of the images being matched is georeferenced, and if the number of valid correspondences that lie on the plane of the GCPs is bigger than a threshold (*matchesThresForGeoreferencing*), the homography parameters between the non-georeferenced image and the reference coordinate system are calculated. If none of the images are georeferenced, only the homography transformation parameters between the images are stored.

The aforementioned image matching and homography computation technique is applied for every pair of overlapping images of similar perspective. By the end of this process, the homography transformation from each remaining non-georeferenced image to the ground reference system is computed, using the estimated homography parameters between the non-georeferenced image and its corresponding georeferenced one with the maximum number of correspondences. The developed workflow for the matching and georeferencing of images of a piecewise planar scene is illustrated in Figure 5-2.

The main output of the georeferencing procedure is considered to be the homography parameters which define the transformation between each image and the coordinate reference system. With this information as input:

- the images can be rectified and visualized in a GIS environment, with the plane of GCPs at each image being at the correct position, orientation and scale, while the position, orientation and scale of the rest of the scene depicted in each image being approximate;
- the image footprints can be exported in vector format;
- the images can be subdivided into groups based on the region they cover; and
- the images that depict a specific region can be automatically detected.

## 5.5 Approximate Exterior Orientation Estimation of Georeferenced Image Datasets of a Piecewise Planar Scene

In this section, a method for the estimation of approximate exterior orientation parameters of the images of a piecewise planar scene is proposed. It is hereinafter referred to as *CP-based EO Estimation* method. It can be applied in image datasets after the implementation of the georeferencing methodology presented in section 5.4, which computes the homography transformation parameters between each image and the reference coordinate system. The initial data that are required consist of:

- the image coordinates of the coplanar points, detected according to the methodology presented in section 5.3;
- the homography parameters, calculated according to the georeferencing methodology described in section 5.4, that define the transformation from the plane of interest (i.e., the plane of GCPs) of each image to the reference coordinate system;
- the $Z$ coordinate of the GCPs that define the plane of interest; and
- the camera interior orientation parameters.

If the camera interior orientation parameters are not available, the camera constant may be computed using the homography matrix of an image – or multiple images – of the dataset, calculated

according to the georeferencing methodology described in section 5.4, assuming a principal point located at the center of the image (Sturm and Maybank, 1999; Grammatikopoulos, 2007). Besides, the developed methodology computes an approximate solution for camera exterior orientation parameters and thus, the assumption of a principal point located at the center of the image is reasonable.

The developed exterior orientation method is applied in each image separately and implements two main steps, that is, (i) estimation of ground coordinates for the detected coplanar points of each image; and (ii) monocular exterior orientation estimation for each image, as illustrated in Figure 5-3. In the following sections, the developed orientation procedure as applied in each image separately is described. It can be applied for estimation of the approximate exterior orientation parameters of (i) the entire dataset of images for input to a global SfM process; and mainly (ii) a starting image pair for initialization of an incremental SfM process, as it is discussed in chapter 7 of this dissertation, in the case that the images depict a piecewise planar scene.

### 5.5.1 Estimation of Ground Coordinates for the Coplanar Feature Points

The horizontal ground coordinates ($X$, $Y$) of the coplanar feature points of an image (*img_i)*, detected according to the method presented in section 5.3, are calculated using their pixel coordinates ($x_{pix}$, $y_{pix}$) in *img_i* in combination with the estimated homography transformation parameters from *img_i* to the real world plane (equation (4.2)), computed according to the

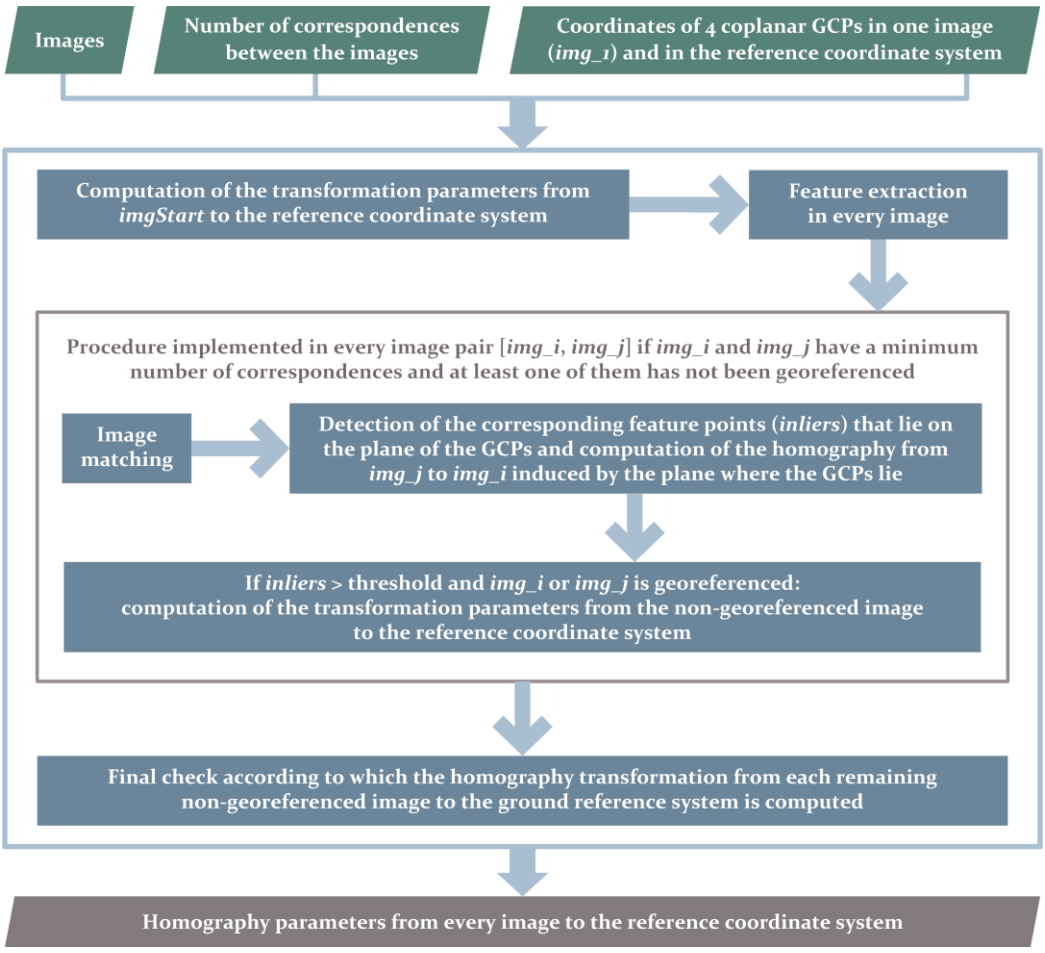


Figure 5-2. The developed workflow for automatic georeferencing of images of a piecewise planar scene

georeferencing method introduced in section 5.4. The third ($Z$) coordinate of the coplanar feature points is equal to the $Z$ coordinate of the coplanar GCPs, used for the georeferencing process.

### 5.5.2 Exterior Orientation Estimation

Otto von Gruber was the first researcher to adopt a parameterization that allows estimating directly the exterior orientation parameters along with the principal point coordinates (assuming a known camera constant) from homographies (von Gruber, 1932). In this section, the method used by the developed *CP-based EO Estimation* for calculating the exterior orientation parameters of *img_i* through space resection, after computation of initial approximations for the camera exterior orientation parameters through an iterative solution that solves the Perspective-n-Point (PnP) problem (Lepetit et al., 2009) for coplanar image points, followed by the conversion of the estimated extrinsic parameters in a form required by the space resection step, is presented.

#### 5.5.2.1 Perspective-n-Point Solution for Coplanar Points

The relationship between the homography matrix $\boldsymbol{H}$ that transforms the coplanar points of an image to the reference coordinate system and the exterior and interior orientation parameters is expressed by equation (5.3).

$$\boldsymbol{H} = \frac{1}{\lambda}\boldsymbol{K}\left[\boldsymbol{R}'|\boldsymbol{t}'\right] \tag{5.3}$$

In equation (5.3), $\lambda$ is a scale factor, $\boldsymbol{K}$ is the camera matrix that contains the intrinsic parameters in pixels, that is, the camera constant ($c_{pix}$) and the principal point coordinates ($x_{0\,pix}$, $y_{0\,pix}$), as expressed by equation (5.4), and $[\boldsymbol{R}'|\boldsymbol{t}']$ is the 3×4 joint rotation-translation matrix, as expressed by equation (5.5), that contains the elements $r'_{ij}$ of the rotation matrix and the elements $t_i$ of the translation vector that transform the world coordinates to image coordinates.

$$\boldsymbol{K} = \begin{bmatrix} c_{pix} & 0 & x_{0\,pix} \\ 0 & c_{pix} & y_{0\,pix} \\ 0 & 0 & 1 \end{bmatrix} \tag{5.4}$$

$$\left[\boldsymbol{R}'|\boldsymbol{t}'\right] = \begin{bmatrix} r'_{11} & r'_{12} & r'_{13} & t_1 \\ r'_{21} & r'_{22} & r'_{23} & t_2 \\ r'_{31} & r'_{32} & r'_{33} & t_3 \end{bmatrix} \tag{5.5}$$

In the case that the interior orientation parameters are not available, a principal point located at the center of each image and a camera constant equal to the camera focal length, extracted from the image metadata, may be assumed for construction of the camera matrix $\boldsymbol{K}$. Otherwise, the camera constant may be computed using an estimated homography matrix, as stated above (section 5.5), also assuming a principal point located at the center of each image.

Let $\boldsymbol{h_1}$, $\boldsymbol{h_2}$ and $\boldsymbol{h_3}$ be the vectors that contain the three elements of each column of the homography matrix $\boldsymbol{H}$ and let $\boldsymbol{r_1}$, $\boldsymbol{r_2}$ and $\boldsymbol{t}'$ express the first, second and fourth column of the joint rotation-translation matrix $[\boldsymbol{R}'|\boldsymbol{t}']$. Then, equation (5.3) takes the following form:

$$\begin{bmatrix} \boldsymbol{h}_1 & \boldsymbol{h}_2 & \boldsymbol{h}_3 \end{bmatrix} = \frac{1}{\lambda} \boldsymbol{K} \begin{bmatrix} \boldsymbol{r}_1 & \boldsymbol{r}_2 & \boldsymbol{t}' \end{bmatrix} \tag{5.6}$$

Thus, the vectors $\boldsymbol{r}_1$, $\boldsymbol{r}_2$ and $\boldsymbol{t}'$ are computed by equations (5.8), (5.9) and (5.10), after computation of the scale factor $\lambda$, according to equation (5.7).

$$\lambda = \frac{1}{\left\| \boldsymbol{K}^{-1} \cdot \boldsymbol{h}_1 \right\|} \tag{5.7}$$

$$\boldsymbol{h}_1 = \frac{1}{\lambda} \cdot \boldsymbol{K} \cdot \boldsymbol{r}_1 \quad \rightarrow \quad \boldsymbol{r}_1 = \lambda \cdot \boldsymbol{K}^{-1} \cdot \boldsymbol{h}_1 \tag{5.8}$$

$$\boldsymbol{h}_2 = \frac{1}{\lambda} \cdot \boldsymbol{K} \cdot \boldsymbol{r}_2 \quad \rightarrow \quad \boldsymbol{r}_2 = \lambda \cdot \boldsymbol{K}^{-1} \cdot \boldsymbol{h}_2 \tag{5.9}$$

$$\boldsymbol{h}_3 = \frac{1}{\lambda} \cdot \boldsymbol{K} \cdot \boldsymbol{t}' \quad \rightarrow \quad \boldsymbol{t}' = \lambda \cdot \boldsymbol{K}^{-1} \cdot \boldsymbol{h}_3 \tag{5.10}$$

The three columns $\boldsymbol{r}_1$, $\boldsymbol{r}_2$ and $\boldsymbol{r}_3$ of the rotation matrix $\boldsymbol{R}'$, which is an orthogonal matrix, are orthonormal vectors. Thus, the vector $\boldsymbol{r}_3$ is perpendicular to the plane where $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ lie. Therefore, $\boldsymbol{r}_3$ is calculated via the external product of $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ according to equation (5.11).

$$\boldsymbol{r}_3 = \boldsymbol{r}_1 \times \boldsymbol{r}_2 \tag{5.11}$$

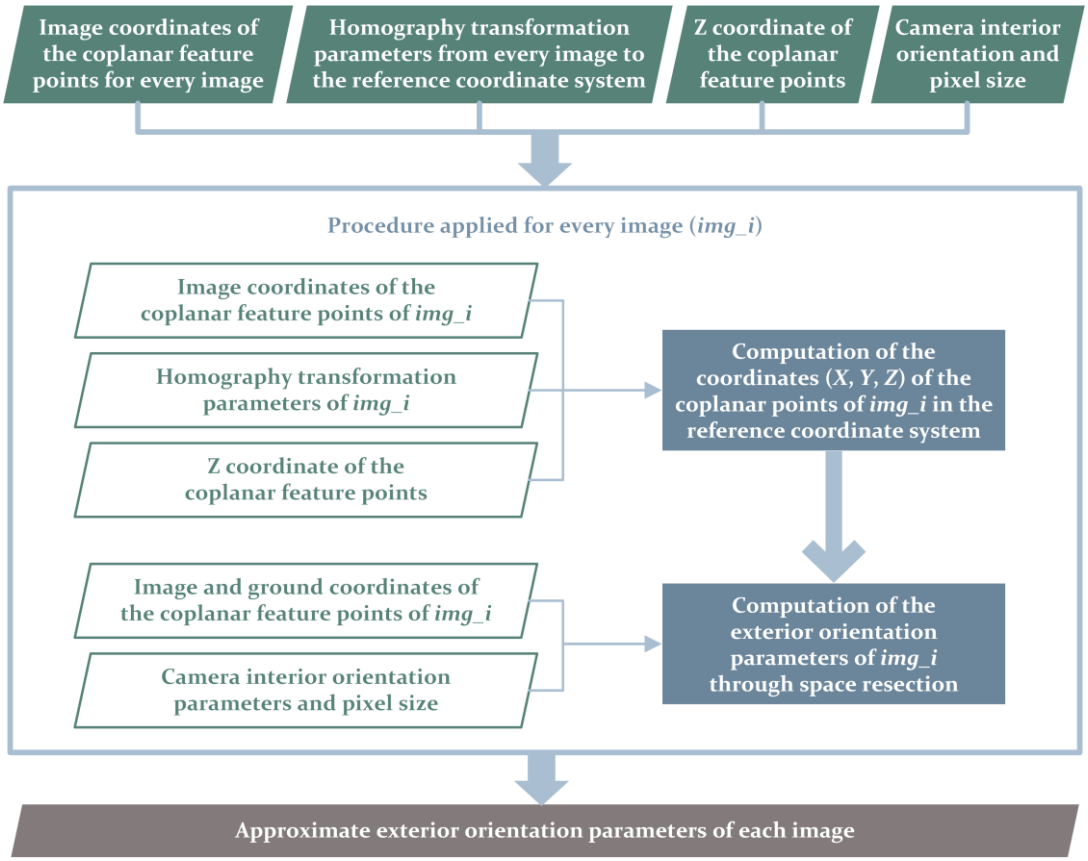


Figure 5-3. The developed workflow for computation of approximate exterior orientation parameters of an image dataset of a piecewise planar scene through the *CP-based EO Estimation* method

However, because of noise in data, the computed rotation matrix $\boldsymbol{R}' = [\boldsymbol{r}_1 \ \boldsymbol{r}_2 \ \boldsymbol{r}_3]$ does not generally satisfy the properties of a rotation matrix. Therefore, the best rotation matrix from the computed 3×3 matrix is calculated according to the method discussed by Zhang (2000). Specifically, the computed matrix $\boldsymbol{R}'$ is initially factored into a singular value decomposition (SVD), according to equation (5.12).

$$\boldsymbol{R}' = \boldsymbol{U} \cdot \boldsymbol{W} \cdot \boldsymbol{V}^{\mathrm{T}} \tag{5.12}$$

In equation (5.12), $\boldsymbol{U}$ and $\boldsymbol{V}$ are 3×3 orthogonal matrices and $\boldsymbol{W}$ is a 3×3 diagonal matrix, with its diagonal entries being the singular values of $\boldsymbol{R}'$. By setting $\boldsymbol{W} = \boldsymbol{I}$, the best rotation matrix $\boldsymbol{R}'$ to approximate the computed one is estimated by equation (5.13).

$$\boldsymbol{R}' = \boldsymbol{U} \cdot \boldsymbol{V}^{\mathrm{T}} \tag{5.13}$$

Then, the matrix $\boldsymbol{R}'$ is converted into a three-dimensional vector $\boldsymbol{r}$ according to Rodrigues' rotation formula, which is a convenient and more compact representation for optimization procedures. This vector $\boldsymbol{r}$ is parallel to the rotation axis and its magnitude equals to the angle of rotation $\theta$. The elements $r_x$, $r_y$ and $r_z$ of the three-dimensional unit vector $\boldsymbol{r}_{unit}$ that describes the axis of rotation along with the angle $\theta$ are computed based on equation (5.14).

$$\sin\theta \cdot \begin{bmatrix} 0 & -r_z & r_y \\ r_z & 0 & -r_x \\ -r_y & r_x & 0 \end{bmatrix} = \frac{\boldsymbol{R}' - (\boldsymbol{R}')^{\mathrm{T}}}{2} \tag{5.14}$$

The rotation vector $\boldsymbol{r}$ is estimated according to equation (5.15).

$$\boldsymbol{r} = \theta \cdot \begin{bmatrix} r_x & r_y & r_z \end{bmatrix}^{\mathrm{T}} \tag{5.15}$$

Finally, the exterior orientation parameters of each image separately are estimated through a Levenberg-Marquardt optimization method which minimizes the reprojection error, that is, the sum of squared distances between the observed image points and the projected ones in the image, using the estimated extrinsic parameters (in the form of translation vector $\boldsymbol{t}'$ and rotation vector $\boldsymbol{r}$) and the camera intrinsic parameters (matrix $\boldsymbol{K}$ along with the camera distortion coefficients, in the case that they are known, or assuming zero distortion).

### 5.5.2.2 Initial Estimation of Exterior Orientation Parameters

After the estimation of the rotation and translation vector through the PnP solution for coplanar points, the rotation vector is converted into a rotation matrix. Initially, the elements $r_x$, $r_y$ and $r_z$ of the three-dimensional unit vector $\boldsymbol{r}_{unit}$ that describes the axis of rotation are computed according to equation (5.16) and the rotation matrix $\boldsymbol{R}'$ is estimated through equation (5.17).

$$\boldsymbol{r}_{unit} = \frac{1}{\theta} \cdot \boldsymbol{r} \quad \rightarrow \quad \begin{bmatrix} r_x \\ r_y \\ r_z \end{bmatrix} = \frac{1}{\theta} \cdot \begin{bmatrix} r_1 \\ r_2 \\ r_3 \end{bmatrix} \tag{5.16}$$

$$\boldsymbol{R}' = \cos\theta \cdot \boldsymbol{I} + \left(1 - \cos\theta\right) \cdot \boldsymbol{r}_{unit} \cdot \boldsymbol{r}_{unit}^{\mathrm{T}} + \sin\theta \cdot \begin{bmatrix} 0 & -r_z & r_y \\ r_z & 0 & -r_x \\ -r_y & r_x & 0 \end{bmatrix} \tag{5.17}$$

The rotation matrix $\boldsymbol{R}'$ which is computed according to equation (5.17) using the rotation vector estimated by the PnP solution, along with the translation vector $\boldsymbol{t}'$, computed through the PnP solution, express the transformation from the world coordinate system to the camera coordinate system; thus, the matrix $[\boldsymbol{R}'|\boldsymbol{t}']$ describes how the world is transformed relative to the camera. However, the rotation matrix $\boldsymbol{R}$ used by the collinearity equations expresses the rotation from the world coordinate system to the camera coordinate system, thus describing the orientation of the camera with respect to the world coordinate system. In addition, the coordinates of the camera perspective center ($X_0$, $Y_0$, $Z_0$) in the world coordinate reference system are used by the collinearity equations.

Let $\boldsymbol{t} = [X_0 \;\; Y_0 \;\; Z_0]^{\mathrm{T}}$ be the vector with the coordinates of the camera perspective center in the world coordinate reference system; let $\boldsymbol{R}$ be the rotation matrix that describes the orientation of the camera with respect to the world coordinate axes; and let $[\boldsymbol{R}|\boldsymbol{t}]$ be the 3×4 exterior orientation matrix that contains the elements of the rotation matrix along with the perspective center coordinates. For reasons of convenience, in order to prove the relationship between the exterior orientation elements (matrix $\boldsymbol{R}$ and vector $\boldsymbol{t}$) and the camera pose parameters (matrix $\boldsymbol{R}'$ and vector $\boldsymbol{t}'$), an additional row of [0 0 0 1] is added at the bottom of each matrix, $[\boldsymbol{R}|\boldsymbol{t}]$ and $[\boldsymbol{R}'|\boldsymbol{t}']$, so that they are transformed into square matrices in order to be invertible. Then, the exterior orientation matrix is obtained by inverting the camera pose matrix obtained through the PnP solution, as shown in equation (5.18).

$$\begin{bmatrix} r_{11} & r_{12} & r_{13} & X_0 \\ r_{21} & r_{22} & r_{23} & Y_0 \\ r_{31} & r_{32} & r_{33} & Z_0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} r'_{11} & r'_{12} & r'_{13} & t_1 \\ r'_{21} & r'_{22} & r'_{23} & t_2 \\ r'_{31} & r'_{32} & r'_{33} & t_3 \\ 0 & 0 & 0 & 1 \end{bmatrix}^{-1} \quad \rightarrow \quad \left[\begin{array}{c|c} \boldsymbol{R} & \boldsymbol{t} \\ \hline \boldsymbol{0}^{\mathrm{T}} & 1 \end{array}\right] = \left[\begin{array}{c|c} \boldsymbol{R}' & \boldsymbol{t}' \\ \hline \boldsymbol{0}^{\mathrm{T}} & 1 \end{array}\right]^{-1} \tag{5.18}$$

By applying matrix decomposition and multiplication and taking into account the properties of inverse matrices along with the fact that the inverse of a rotation matrix equals to its transpose, equation (5.18) takes the following form (equation (5.19)).

$$\left[\begin{array}{c|c} \boldsymbol{R} & \boldsymbol{t} \\ \hline \boldsymbol{0}^{\mathrm{T}} & 1 \end{array}\right] = \left[\begin{array}{c|c} \boldsymbol{R}' & \boldsymbol{t}' \\ \hline \boldsymbol{0}^{\mathrm{T}} & 1 \end{array}\right]^{-1} = \begin{bmatrix} r'_{11} & r'_{12} & r'_{13} & t_1 \\ r'_{21} & r'_{22} & r'_{23} & t_2 \\ r'_{31} & r'_{32} & r'_{33} & t_3 \\ 0 & 0 & 0 & 1 \end{bmatrix}^{-1} = \left[ \begin{bmatrix} 1 & 0 & 0 & X_0 \\ 0 & 1 & 0 & Y_0 \\ 0 & 0 & 1 & Z_0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} r'_{11} & r'_{12} & r'_{13} & 0 \\ r'_{21} & r'_{22} & r'_{23} & 0 \\ r'_{31} & r'_{32} & r'_{33} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \right]^{-1}$$

$$= \begin{bmatrix} r'_{11} & r'_{12} & r'_{13} & 0 \\ r'_{21} & r'_{22} & r'_{23} & 0 \\ r'_{31} & r'_{32} & r'_{33} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}^{-1} \cdot \begin{bmatrix} 1 & 0 & 0 & t_1 \\ 0 & 1 & 0 & t_2 \\ 0 & 0 & 1 & t_3 \\ 0 & 0 & 0 & 1 \end{bmatrix}^{-1} = \left[\begin{array}{c|c} \boldsymbol{R}' & \boldsymbol{0} \\ \hline \boldsymbol{0}^{\mathrm{T}} & 1 \end{array}\right]^{-1} \cdot \left[\begin{array}{c|c} \boldsymbol{I} & \mathbf{t}' \\ \hline \boldsymbol{0}^{\mathrm{T}} & 1 \end{array}\right]^{-1}$$

$$= \left[\begin{array}{c|c} \left(\boldsymbol{R}'\right)^{-1} & \boldsymbol{0} \\ \hline \boldsymbol{0}^{\mathrm{T}} & 1 \end{array}\right] \cdot \left[\begin{array}{c|c} \boldsymbol{I} & -\mathbf{t}' \\ \hline \boldsymbol{0}^{\mathrm{T}} & 1 \end{array}\right] = \left[\begin{array}{c|c} \left(\boldsymbol{R}'\right)^{\mathrm{T}} & \boldsymbol{0} \\ \hline \boldsymbol{0}^{\mathrm{T}} & 1 \end{array}\right] \cdot \left[\begin{array}{c|c} \boldsymbol{I} & -\mathbf{t}' \\ \hline \boldsymbol{0}^{\mathrm{T}} & 1 \end{array}\right]$$

$$= \begin{bmatrix} r'_{11} & r'_{21} & r'_{31} & 0 \\ r'_{12} & r'_{22} & r'_{32} & 0 \\ r'_{13} & r'_{23} & r'_{33} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 & -t_1 \\ 0 & 1 & 0 & -t_2 \\ 0 & 0 & 1 & -t_3 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} r'_{11} & r'_{21} & r'_{31} & -\left(r'_{11}t_1 + r'_{21}t_2 + r'_{31}t_3\right) \\ r'_{12} & r'_{22} & r'_{32} & -\left(r'_{12}t_1 + r'_{22}t_2 + r'_{32}t_3\right) \\ r'_{13} & r'_{23} & r'_{33} & -\left(r'_{13}t_1 + r'_{23}t_2 + r'_{33}t_3\right) \\ 0 & 0 & 0 & 1 \end{bmatrix} = \left[\begin{array}{c|c} \left(\boldsymbol{R}'\right)^{\mathrm{T}} & -\left(\boldsymbol{R}'\right)^{\mathrm{T}} \cdot \boldsymbol{t}' \\ \hline \boldsymbol{0}^{\mathrm{T}} & 1 \end{array}\right] \quad \rightarrow$$

$$\left[\begin{array}{c|c} \boldsymbol{R} & \boldsymbol{t} \\ \hline \boldsymbol{0} & 1 \end{array}\right] = \left[\begin{array}{c|c} \left(\boldsymbol{R}'\right)^{\mathrm{T}} & -\left(\boldsymbol{R}'\right)^{\mathrm{T}} \cdot \boldsymbol{t}' \\ \hline \boldsymbol{0} & 1 \end{array}\right] \tag{5.19}$$

Thus, the rotation matrix $\boldsymbol{R}$ is computed via equation (5.20) and the linear exterior orientation parameters ($X_0$, $Y_0$, $Z_0$) are computed via equation (5.21).

$$\boldsymbol{R} = \left(\boldsymbol{R}'\right)^{\mathrm{T}} \tag{5.20}$$

$$\boldsymbol{t} = -\left(\boldsymbol{R}'\right)^{\mathrm{T}} \cdot \boldsymbol{t}' \quad \rightarrow \quad \begin{bmatrix} X_0 \\ Y_0 \\ Z_0 \end{bmatrix} = -\left(\boldsymbol{R}'\right)^{\mathrm{T}} \cdot \boldsymbol{t}' \tag{5.21}$$

The rotation angles $\omega$, $\varphi$ and $\kappa$ are computed from the estimated rotation matrix, according to the set of equations (5.22).

$$\omega = \tan^{-1}\left(\frac{-r_{32}}{r_{33}}\right) \qquad \varphi = \sin^{-1}\left(r_{31}\right) \qquad \kappa = \tan^{-1}\left(\frac{-r_{21}}{r_{11}}\right) \tag{5.22}$$

### 5.5.2.3 Photogrammetric Space Resection

The exterior orientation parameters ($X_0$, $Y_0$, $Z_0$, $\omega$, $\varphi$, $\kappa$) of *img_i* are estimated through photogrammetric space resection using the collinearity equations as the mathematical model, assuming zero distortion, through a least-squares adjustment of indirect observations. As stated in section 5.5.2.1, if calibration information is not available, a principal point ($x_0$, $y_0$) located at the center of each image and a camera constant ($c$) equal to the camera focal length, extracted from the image metadata, may be assumed, or the camera constant may be estimated using an image-to-ground homography matrix.

Initially, the pixel coordinates of the coplanar points in *img_i* ($x_{pix}$, $y_{pix}$) are converted in the photogrammetric system ($x$, $y$). The collinearity equations are not linear with respect to the unknowns, i.e., the exterior orientation parameters. Thus, they are linearized through the Taylor series expansion around approximate values of the exterior orientation parameters, neglecting terms of order higher than 1. In this way, the unknowns of the least-squares solution are the differences between the approximate values of the camera exterior orientation parameters and the ones that result from the least-squares process.

The number of equations that can be formed to solve the space resection problem based on the collinearity conditions is: $n = 2N$, where $N$ is the number of points of known ground coordinates that are involved in the least-squares solution. The system of linear equations is represented in a matrix form according to equation (5.23), assuming equal weights for the observations, and the solution is given by equation (5.24).

$$\boldsymbol{A}\boldsymbol{dx} = \boldsymbol{dl} + \boldsymbol{\upsilon} \tag{5.23}$$

$$\boldsymbol{dx} = \left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\right)^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{dl} \tag{5.24}$$

In equations (5.23) and (5.24), $\boldsymbol{dx}$ is the six-dimensional vector with the differences between the approximate values of the unknown parameters and those ones computed by the least-squares solution (equation (5.25)), $\boldsymbol{dl}$ is the $n$-dimensional vector with the differences between the measured coordinates of the GCPs in the reference coordinate system and their approximate values, computed using the approximate values of the unknown parameters, $\boldsymbol{\upsilon}$ is the $n$-dimensional residual vector and $\boldsymbol{A}$ is the $n\times 6$ coefficient matrix with the partial derivatives with respect to the unknowns.

$$\boldsymbol{dx} = \begin{bmatrix} dX_0 & dY_0 & dZ_0 & d\omega & d\varphi & d\kappa \end{bmatrix}^{\mathrm{T}} \tag{5.25}$$

Then, the best values of the exterior orientation parameters are computed by the addition of the vector $\boldsymbol{dx}$ and the vector that contains the approximate exterior orientation parameters. If the values of the vector $\boldsymbol{dx}$ are bigger than some thresholds (*linearEOThres* for $dX_0$, $dY_0$ and $dZ_0$ and *angularEOThres* for $d\omega$, $\delta\varphi$ and $d\kappa$ – see Table 5-3), the best values of the exterior orientation parameters are considered to be the approximate ones and the least-squares procedure is repeated until these thresholds are reached. The space resection algorithm is illustrated in the form of a flowchart in Figure 5-4.

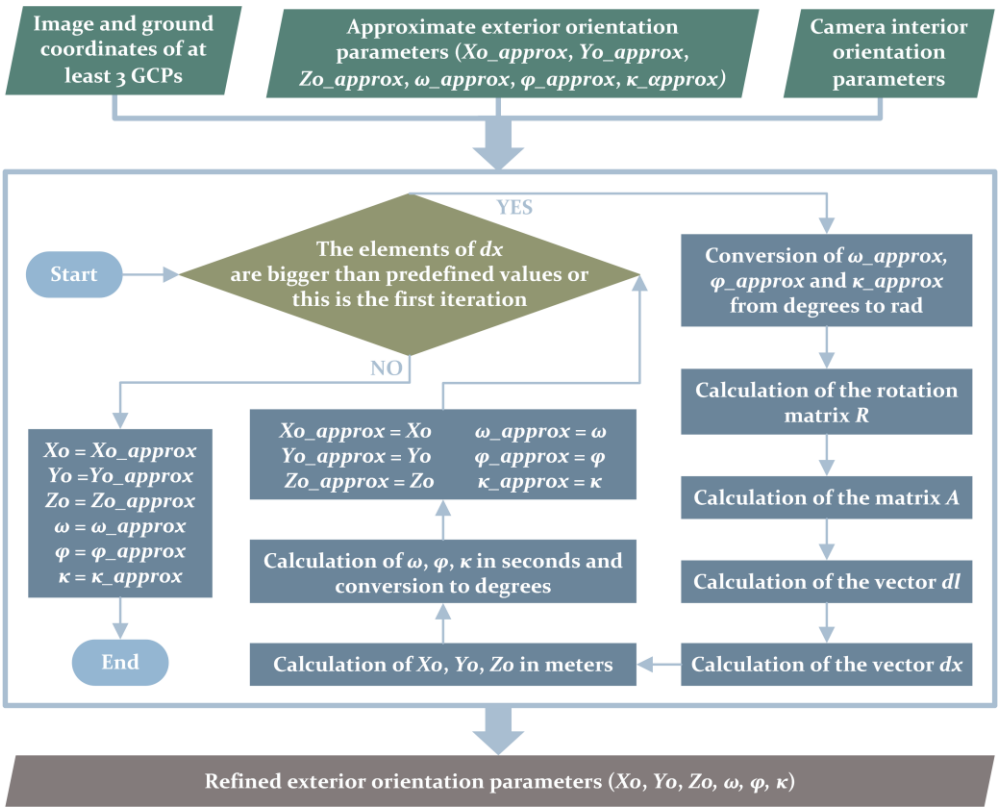


Figure 5-4. Workflow illustrating the space resection algorithm

Table 5-3. Parameter values used by the space resection algorithm

| Symbol | Value |
|---|---|
| *linearEOThres* | 0.001 m |
| *angularEOThres* | 2′′ |

## 5.6 Experimental Results

In this section, the developed software suite that implements the methodology presented in sections 5.3, 5.4 and 5.5 is presented, along with the test datasets and the obtained results.

### 5.6.1 Developed Software

A software suite that implements the coplanar point detection methodology presented in this chapter for the scope of image georeferencing and approximate exterior orientation estimation has been developed in the C++ programming language, making use of some functionalities offered by the OpenCV library (OpenCV Team, 2019) for image manipulations, the Boost libraries (Boost Team, 2019) for graph creation, the Eigen Library (Eigen Team, 2019) for matrix and vector operations and the GDI+ library (Microsoft, 2019) for extracting image metadata. The following developed software tools are used for the experiments:

- a software tool for determining overlapping images (section 4.4.2.1);
- a software tool for detecting coplanar feature points in image datasets, lying on the plane of the coplanar GCPs;
- a software tool for georeferencing images based on the detected coplanar feature points; and
- a software tool for computing approximate exterior orientation parameters of georeferenced images.

### 5.6.2 Test Datasets

The developed software that implements the proposed algorithm was tested using two datasets of UAV oblique aerial images. Specifically, *UAV Dataset 1* incorporates 23 images and *UAV Dataset 2* incorporates 35 images, as illustrated in Figure 5-5. The images, which correspond to an overlap bigger than 90% across track, are part of a dataset acquired for the ISPRS scientific initiative "ISPRS benchmark for multi-platform photogrammetry", run in collaboration with EuroSDR (Nex et al., 2015). They were taken by a Sony Nex-7 camera over the city hall in Dortmund, Germany, with a GSD varying from 1 to 3 cm. Each image has a size of 6000×4000 pixels and a focal length of 16 mm. The dataset provided by the benchmark also contains the ground coordinates of some targets on the facades of the city hall that exist only in the terrestrial images.

### 5.6.3 Reference Data

In order to obtain (i) the ground coordinates of four coplanar points for being measured in *img_1*; (ii) reference exterior orientation parameters for comparison with the ones computed via the developed solution; and (iii) the ground coordinates of 15 coplanar check points lying on the plane of GCPs, a block of 645 images (446 oblique and vertical aerial images and 199 terrestrial images)

provided by the ISPRS/EuroSDR benchmark for multi-platform photogrammetry was oriented using the Agisoft PhotoScan software (Agisoft LLC, 2019) through SfM with self-calibrating bundle adjustment using the "high alignment" setting, according to which the images are not downscaled for the tie point extraction process. Specifically, 34 targets of known coordinates on the facades of the city hall were measured in the corresponding terrestrial images (239 measurements in total) and 19 tie points were manually measured in multiple terrestrial and oblique UAV imagery (260 measurements in total). The average reprojection error of the manually measured GCPs and tie points is 0.15 pixels and 0.33 pixels, respectively, and the average absolute difference between the computed and the input coordinates of the GCPs is 2.9 cm. The computed ($X$, $Y$, $Z$) ground coordinates of four coplanar points, lying on the roof of the city hall, which belong to the manually measured tie points of the reference dataset, triangulated via PhotoScan, served as the ground control of the starting image (*img_1*) of *UAV Dataset 1* and *UAV Dataset 2* (Figure 5-6).

### 5.6.4 Results

The results of the coplanar point detection methodology and the georeferencing as well as the exterior orientation estimation processes applied in *UAV Dataset 1* and *UAV Dataset 2* through the developed software tools are outlined in the following sections. As far as the interior orientation parameters used in the experiments are concerned, approximate data were used; specifically, the

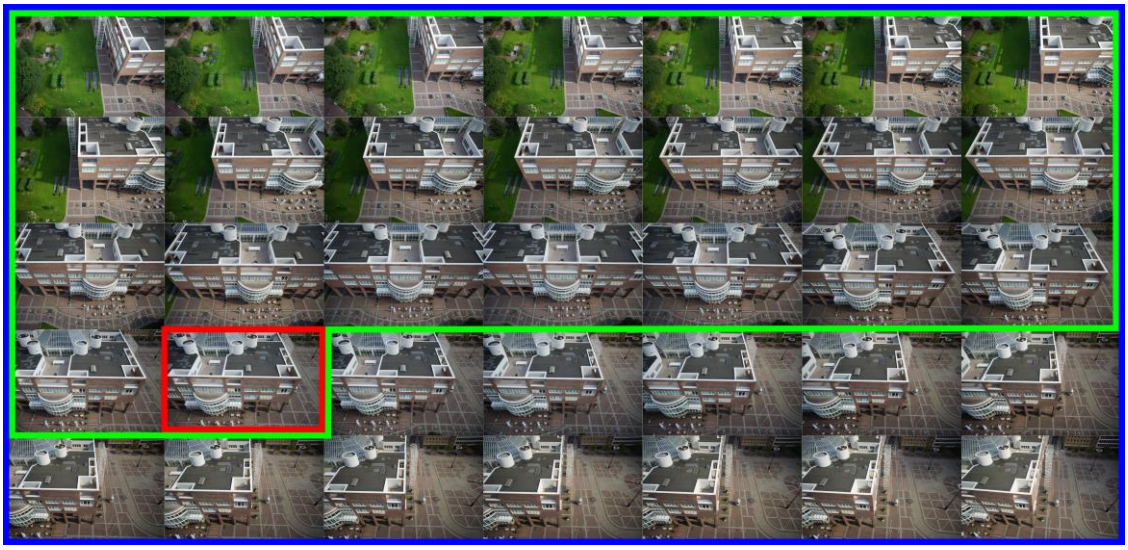

Figure 5-5. *UAV Dataset 1* (outlined in green), *UAV Dataset 2* (outlined in blue) and *img_1* (outlined in red) used for manual measurements of 4 coplanar GCPs

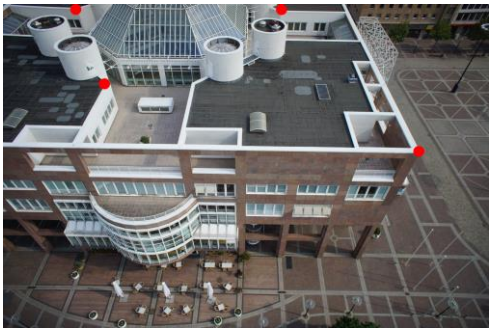

Figure 5-6. Manually measured coplanar GCPs (marked in red) in *img_1* of *UAV Dataset 1* and *UAV Dataset 2*

nominal name of the focal length, extracted from the image metadata, was used as the camera constant of the images, while the principal point was assumed to coincide with the center of the images.

#### 5.6.4.1 Detection of Coplanar Points

Coplanar feature points that lie on the plane where the GCPs lie are identified correctly in each image of *UAV Dataset 1* and *UAV Dataset 2*. An average number of 768 coplanar feature points that lie on the plane of interest is extracted per image, using the developed methodology. Figure 5-7 illustrates the iterative process of transferring the coplanar GCPs from *img_1* to another image *(img_k)* and detecting feature points that lie one the plane of GCPs in both images. The left images of this figure illustrate the automatically detected corresponding coplanar feature points between *img_1* and *img_k* in three iterations of the coplanar point detection process, among which the bottom image corresponds to the last iteration in which the coplanar points are those ones lying on the plane of GCPs. The right images show the estimated epipolar lines of the GCPs, which are similar for all iterations, as they do not depend on the computed homography, along with the homography-based estimations of GCPs using the detected coplanar feature points shown to the left. In the top-right and middle-right images, at least three out of the four homography-based estimations of the GCPs are not close to the template-matching-based estimation of the GCPs; thus the iterations are continued until the homography-based estimation of GCPs are close enough to the estimations obtained through template matching, using the defined threshold. The bottom-right image satisfies this constraint and the iterations are terminated. Moreover, the right images of Figure 5-7 prove the fact that only the epipolar line constraint, according to which the homography-based estimation of the coplanar points of an image lie on the epipolar line of these points in another image (*img_k*), does not suffice for rejecting any wrong homography-based estimations of the coplanar points in *img_k*. This is illustrated by the fact that the estimations of GCPs in *img_k* using all the computed homographies (among which the two out of the three homographies are induced by another plane different from the one where the GCPs lie) lie on their corresponding epipolar line (Figure 5-7, right). Zoom-in views of the estimations of the GCPs in an *img_k* obtained through template matching and through the correctly identified homography transformation are illustrated in Figure 5-8 for the same pair of *img_1* and *img_k*; the homography-based estimations of the GCPs are the correct ones. Figure 5-8 also shows the homologous template matching window of each GCP in *img_2*. The height of the template matching window of the first GCP (Figure 5-8, top-left) is different from the height of the template matching window of the other ones, as this GCP is located close to the top border of *img_1* and its size is automatically adjusted so that the template matching window lies inside the image borders. The aforementioned images (Figure 5-7 and Figure 5-8) illustrate the results of some steps of the coplanar point detection process for the first image pair of *UAV Dataset 2;* analogous results are also derived for the rest of the images.

Figure 5-9 shows the automatically detected coplanar feature points in the images of *UAV Dataset 2* (which includes the images of *UAV Dataset 1*). It is observed that the great majority (more than 99%) of the identified feature point are coplanar and lie on the plane of the GCPs. It is also observed that more points are identified in certain images of the dataset. Images that participate in the matching process being part of more image pairs are those ones with a bigger number of detected coplanar feature points. For instance, if corresponding coplanar feature points are identified between *img_1*-*img_2*, *img_1*-*img_3*, *img_1*-*img_4*, *img_3*-*img_5* and *img_3*-*img_6*, then *img_1* and *img_3* will correspond to a bigger number of identified coplanar points. Furthermore, the number of coplanar points that are detected also depends on the parameter values used by the

coplanar point detection algorithm, as in the case that corresponding coplanar feature points have not been identified between the images of a pair, some thresholds are redefined and the coplanar point detection process is reapeted, until coplanar points are identified.

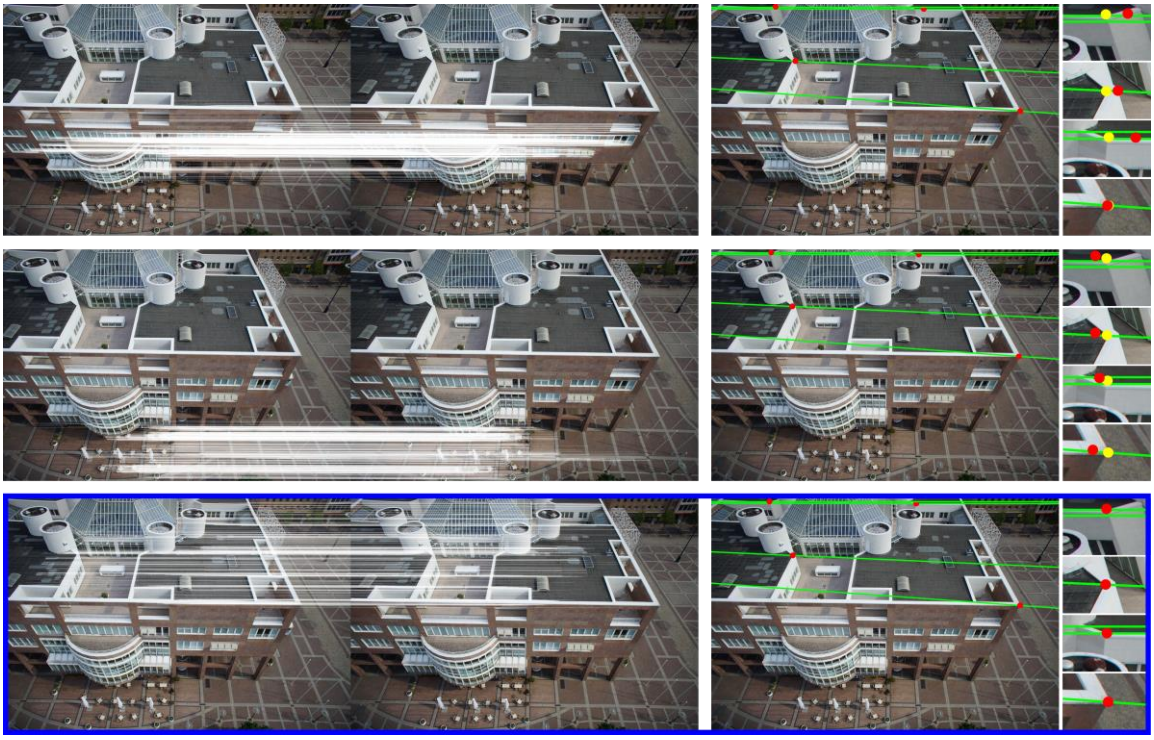

Figure 5-7. Visualization of the iterative process of transferring the coplanar GCPs from *img_1* to an *img_k* of *UAV Dataset 2*; left: detected coplanar corresponding feature points between *img_1* and *img_k*; right: *img_k* showing the epipolar lines of the GCPs (in green) and the homography-based estimations of GCPs (in red) using the detected coplanar feature points shown to the left, along with zoom-in views of the homography-based GCP estimations (in red) and the actual GCPs (in yellow); the blue outline indicates the identified correspondences that lie on the plane of GCPs (left) along with the correct homography-based estimations of the GCPs in *img_k* (right)

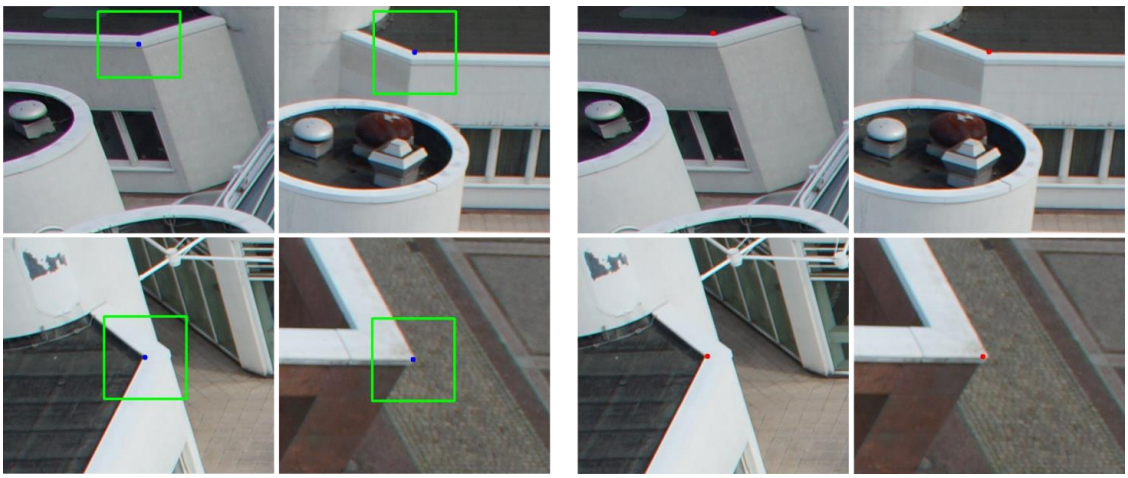

Figure 5-8. Zoom-in views of coplanar GCP estimations in an *img_k* of *UAV Dataset 2*; left: template-matching-based estimations of the four GCPs (in blue) along with the homologous template matching window (outlined in green); right: final (correct) homography-based estimations of the GCPs (in red)

#### 5.6.4.2 Georeferencing

The images of *UAV Dataset 1* and *UAV Dataset 2* are rectified using the homography transformation parameters estimated by the georeferencing procedure outlined in section 5.4. After the rectification of each image, the structures that lie on the plane of GCPs are at the correct position, orientation and scale with an expected error, which is bigger in images that are georeferenced using a big number of homography transformations from the ground reference system and smaller in those ones that are georeferenced using fewer homography transformations. The rectified structures that lie on a plane parallel to the one where the GCPs lie differ from the real world structures by a translation and scaling transformation; they correspond to uniform scale as well as correct shape and orientation, up to an expected error. The position, orientation and scale of the rest of the structures, which lie on a different plane from the plane of GCPs, non-parallel to it, are not correct.

Figure 5-10, top shows two rectified images of *UAV Dataset 2*, with the outlines of structures that lie on the plane of GCPs, which are manually designed using rectified *img_1* as a background, being superimposed on them. Figure 5-10, bottom shows zoom-in views of three rectified images of *UAV Dataset 2* that are georeferenced using 4, 8 and 13 transformations (from left to right, respectively)

Figure 5-9. Automatically detected coplanar feature points in the images of *UAV Dataset 2*; *img_1* is marked by a red outline

from the ground reference system. These images correspond to the calculation of 3, 7 and 12 homography transformations from *img_1* coordinate system, respectively. The visual inspection of the outlines superimposed on these images verifies that the distance between the outlines of planar structures designed using the rectified *img_1* as a background and the outlines of rectified structures of an image increases as the number of calculated transformations from this image to *img_1* coordinate system increases. In other words, the error of the structures that lie on the plane of GCPs of a rectified image increases with the number of transformations applied between this image and the ground reference system. This conclusion is also verified by conducting several measurements of check points in the georeferenced images of *UAV Dataset 2* (which includes the images of *UAV Dataset 1*). The results are shown in Table 5-4, in the form of average absolute differences ($Avg_{|\Delta X|}$, $Avg_{|\Delta Y|}$) between the measured coordinates of points in the rectified images and their reference coordinates for each group of images that correspond to the same range of transformations from the ground coordinate system. The reference coordinates of the check points, which lie on the plane of GCPs, have been obtained through photogrammetric intersection using manual measurements of these points in the corresponding images using their reference interior and exterior orientation parameters, via the Agisoft PhotoScan software (section 5.6.3). The number of images that correspond to the same number of homography transformations from the ground reference system is outlined in Table 5-5. Whereas a smaller number of homography transformations would seem to be enough for the test datasets for the purpose of georeferencing, the coplanar point detection method, used by the georeferencing process, has been successfully implemented in successive or highly overlapping images. Specifically, each image of the test datasets has been used up to three times for transferring coplanar points to another overlapping image. This is because of the used thresholds of the developed coplanar point detection algorithm, which ensure that coplanar points lying on the plane of interest will be detected.

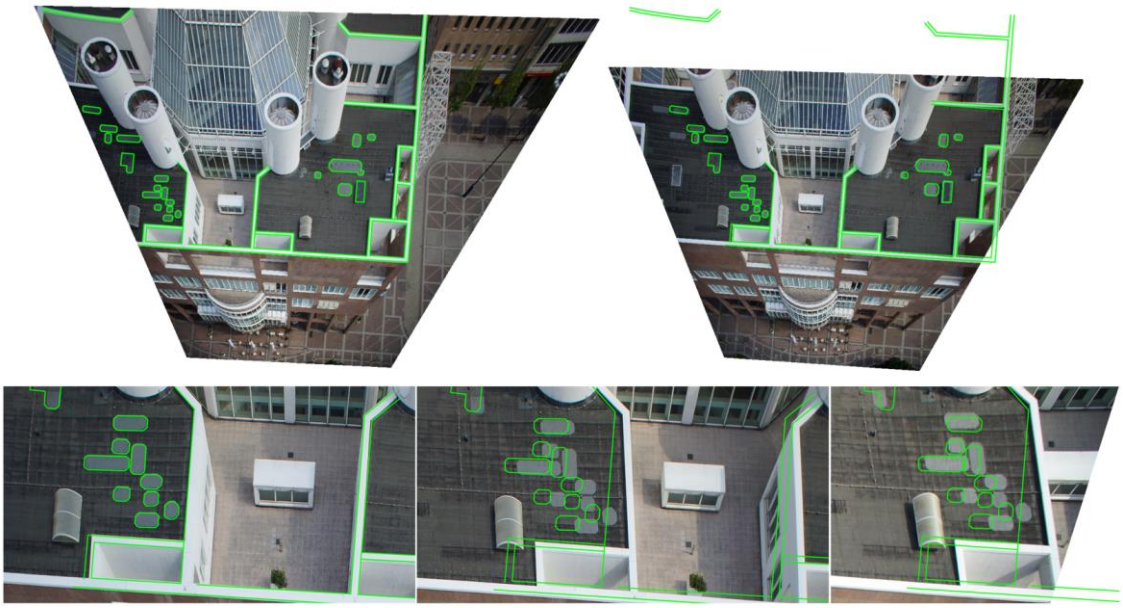

Figure 5-10. Superimposition of manually digitized outlines of planar structures of the georeferenced rectified *img_1* of *UAV Dataset 2* lying on the plane of GCPs, on rectified images of the same dataset; top: *img_1* (left) georeferenced using 1 homography transformation and image georeferenced using 4 homography transformations (right); bottom: zoom-in views of images georeferenced using 4 (left), 8 (middle) and 13 (right) homography transformations

Table 5-4. Accuracy assessment depending on the number of homography transformations between the image and the ground coordinate system for *UAV Dataset 2*

| Number of transformations | Avg$_{\|\Delta X\|}$ (m) | Avg$_{\|\Delta Y\|}$ (m) | Number of measurements |
|---|---|---|---|
| 1 | 0.09 | 0.19 | 10 |
| 2 | 0.43 | 0.22 | 30 |
| 3-4 | 0.67 | 0.15 | 36 |
| 5-6 | 1.46 | 0.39 | 60 |
| 7-9 | 1.51 | 0.81 | 69 |
| 10-12 | 1.96 | 0.47 | 35 |
| 13-15 | 2.06 | 0.38 | 21 |
| 16-18 | 2.50 | 0.40 | 21 |

Table 5-5. Number of images that correspond to the same number of homography transformations from the ground reference system

| Number of transformations | Number of images for *UAV Dataset 1* | Number of images for *UAV Dataset 2* |
|---|---|---|
| 1 | 1 | 1 |
| 2 | 2 | 3 |
| 3 | 1 | 2 |
| 4 | 1 | 2 |
| 5 | 2 | 4 |
| 6 | 1 | 3 |
| 7 | 2 | 3 |
| 8 | 1 | 3 |
| 9 | 1 | 3 |
| 10 | 1 | 1 |
| 11 | 2 | 2 |
| 12 | 1 | 1 |
| 13 | 1 | 1 |
| 14 | 1 | 1 |
| 15 | 1 | 1 |
| 16 | 1 | 1 |
| 17 | 2 | 2 |
| 18 | 1 | 1 |

### 5.6.4.3 Approximate Exterior Orientation Estimation

The method outlined in section 5.5 for approximate estimation of the exterior orientation parameters of georeferenced image datasets of a piecewise planar scene (*CP-based EO Estimation*), has been applied in *UAV Dataset 1* and *UAV Dataset 2*, which have been georeferenced through the method presented in section 5.4 yielding the results presented in section 5.6.4.2. In the following, the differences ($\Delta X_0$, $\Delta Y_0$, $\Delta Z_0$, $\Delta\omega$, $\Delta\varphi$, $\Delta\kappa$) between the exterior orientation parameters computed through one of the developed solutions and the reference exterior orientation parameters are presented in the form of tables (Table 5-6 for *UAV Dataset 1* and Table 5-7 for *UAV Dataset 2*). Specifically, the average absolute differences (Avg-Abs), the maximum (Max-Abs) and minimum (Min-Abs) absolute differences, the average differences (Avg) and the standard deviation of

Table 5-6. Differences between the approximate exterior orientation parameters computed via the developed coplanar point detection method and the reference ones for *UAV Dataset 1*

| ***UAV Dataset 1*: 23 oblique UAV images** | | | | | | |
|---|---|---|---|---|---|---|
| **Method** | **Metric** | **Avg-Abs** | **Max-Abs** | **Min-Abs** | **Avg** | **Stdev** |
| ***CP-based EO Estimation* (SURF; 4 GCPs in one image)** | $\Delta X_o$ (m) | 2.252 | 5.865 | 0.170 | 2.084 | 1.652 |
| | $\Delta Y_o$ (m) | 1.267 | 3.651 | 0.204 | -0.127 | 1.612 |
| | $\Delta Z_o$ (m) | 1.236 | 3.640 | 0.373 | -1.196 | 0.804 |
| | $\Delta\omega$ (deg.) | 0.565 | 1.177 | 0.011 | 0.544 | 0.367 |
| | $\Delta\varphi$ (deg.) | 3.110 | 7.023 | 0.033 | 3.110 | 2.363 |
| | $\Delta\kappa$ (deg.) | 1.148 | 2.895 | 0.224 | 0.488 | 1.216 |
| | $\Delta LinearEO$ (m) | 1.585 | 5.865 | 0.170 | 0.254 | 1.356 |
| | $\Delta AngularEO$ (deg.) | 1.608 | 7.023 | 0.011 | 1.381 | 1.315 |

Table 5-7. Differences between the approximate exterior orientation parameters computed via the developed coplanar point detection method and the reference ones for *UAV Dataset 2*

| ***UAV Dataset 2*: 35 oblique UAV images** | | | | | | |
|---|---|---|---|---|---|---|
| **Method** | **Metric** | **Avg-Abs** | **Max-Abs** | **Min-Abs** | **Avg** | **Stdev** |
| ***CP-based EO Estimation* (SURF; 4 GCPs in one image)** | $\Delta X_o$ (m) | 2.714 | 5.865 | 0.012 | 0.136 | 3.165 |
| | $\Delta Y_o$ (m) | 1.181 | 3.651 | 0.049 | 0.208 | 1.507 |
| | $\Delta Z_o$ (m) | 1.215 | 3.640 | 0.373 | -1.189 | 0.681 |
| | $\Delta\omega$ (deg.) | 0.557 | 1.177 | 0.011 | 0.453 | 0.453 |
| | $\Delta\varphi$ (deg.) | 2.906 | 7.023 | 0.033 | 1.181 | 3.314 |
| | $\Delta\kappa$ (deg.) | 1.733 | 3.718 | 0.224 | -0.659 | 1.891 |
| | $\Delta LinearEO$ (m) | 1.704 | 5.865 | 0.012 | -0.282 | 1.784 |
| | $\Delta AngularEO$ (deg.) | 1.732 | 7.023 | 0.011 | 0.325 | 1.886 |

differences (Stdev) between the exterior orientation parameters computed through one of the developed solutions and the reference ones, along with Avg-Abs, Max-Abs, Min-Abs, Avg and Stdev for the linear ($\Delta LinearEO$) and angular ($\Delta AngularEO$) exterior orientation parameters, computed through the set of equations (5.26), are outlined.

$$\begin{Bmatrix} f_1(\Delta LinearEO) = \left( f_1(\Delta X_0) + f_1(\Delta Y_0) + f_1(\Delta Z_0) \right)/3 \\ f_1(\Delta AngularEO) = \left( f_1(\Delta\omega) + f_1(\Delta\varphi) + f_1(\Delta\kappa) \right)/3 \end{Bmatrix}$$
$$\begin{Bmatrix} f_2(\Delta LinearEO) = f_2\left( f_2(\Delta X_0), f_2(\Delta Y_0), f_2(\Delta Z_0) \right) \\ f_2(\Delta AngularEO) = f_2\left( f_2(\Delta\omega), f_2(\Delta\varphi), f_2(\Delta\kappa) \right) \end{Bmatrix} \quad (5.26)$$
$$\text{where } f_1 = \text{Avg-Abs, Avg, Stdev; } f_2 = \text{Max-Abs, Min-Abs}$$

The diagrams illustrated in Figure 5-11 and Figure 5-12 show the distribution of absolute $\Delta LinearEO$ and $\Delta AngularEO$, respectively, for the images of *UAV Dataset 2* (which includes the images of *UAV Dataset 1*), revealing a greater absolute value of $\Delta LinearEO$ and $\Delta AngularEO$ for images that are far away from the image with GCP measurements (*img_1*). The values of absolute $\Delta LinearEO$ and $\Delta AngularEO$ generally increase with the number of homography transformations between an image and the ground reference system. This is also illustrated in the form of bar charts in Figure 5-13, which illustrate the average absolute $\Delta LinearEO$ and $\Delta AngularEO$ for the developed *CP-based EO Estimation* method for each group of images of *UAV Dataset 2* that corresponds to the same number of transformations from the ground reference system (see also Table 5-5 for the

number of images that correspond to the same number of homography transformations from the ground reference system). The suitability of the derived approximate exterior orientation results to be used as initial values in a bundle adjustment procedure is investigated in section 7.10.4 of this dissertation.

## 5.7 Conclusions

In this chapter, a method for detecting coplanar feature points that lie on a user-defined plane among oblique and vertical large-scale aerial imagery depicting a piecewise planar scene has been presented and two applications that implement the developed coplanar point detection scheme have been discussed, i.e., the 2D homography-based georeferencing of image datasets of a piecewise planar scene and the estimation of approximate exterior orientation parameters of georeferenced image datasets of a piecewise planar scene. The experiments, which were conducted using two datasets of highly overlapping UAV oblique images of similar perspective, reveal that the coplanar point detection method yields successful results in terms of automatically detecting points that lie on a user-defined plane. Moreover, the implementation of the coplanar point detection method

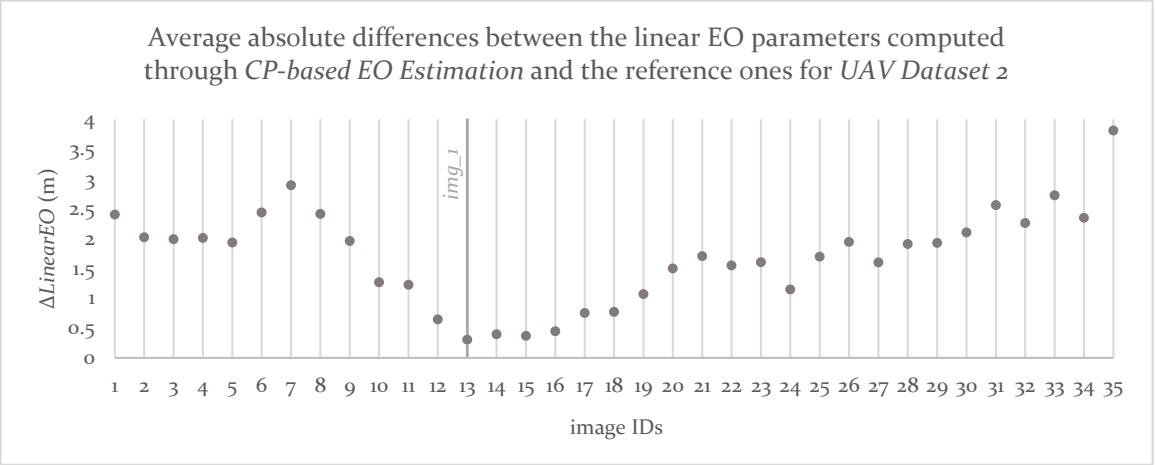


Figure 5-11. Average absolute differences between the linear EO parameters computed through *CP-based EO Estimation* (using the SURF feature extractor and four GCP measurements in *img_1*) and the reference ones for the images of *UAV Dataset 2*

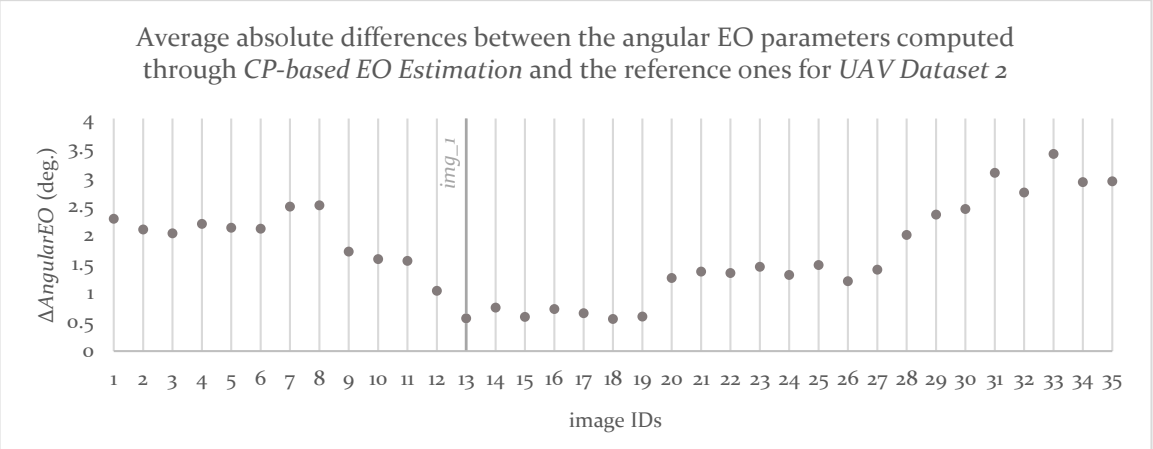


Figure 5-12. Average absolute differences between the angular EO parameters computed through *CP-based EO Estimation* (using the SURF feature extractor and four GCP measurements in *img_1*) and the reference ones for the images of *UAV Dataset 2*

within the developed georeferencing framework presented in section 4.3 of this dissertation for constraining the detected correspondences to lie in a single plane, i.e., the one where the points of known ground coordinates are measured, along with the appropriate changes discussed in this chapter, gives satisfactory georeferencing results in large-scale image datasets of a piecewise planar scene. In this way, the fields of application of the developed 2D georeferencing framework are not restricted to image datasets depicting regions with small elevation differences compared with the flying height or datasets with image footprints that can be approximated by a rectangle. On the contrary, thanks to the proposed constraint and corresponding changes, it can be applied in image datasets regardless of the elevation difference of the scene, provided however that all images depict a planar area with the same elevation as the elevation of the coplanar GCPs.

Also, the computation of approximate exterior orientation parameters of images of a piecewise planar scene, for which the 2D homography transformation parameters to the ground coordinate system have been computed, has been addressed. The results show that the approximate exterior orientation parameters, estimated using the method presented in this chapter, may be used as initial values into a bundle adjustment procedure, as will be shown in chapter 7 of this dissertation. However, one of the drawbacks of the approximate exterior orientation method presented in this chapter is the decrease of the accuracy of the estimated exterior orientation parameters with an increase of the number of homography transformations involved between each image and the reference coordinate system. As a result, the developed method for estimating approximate exterior orientation parameters of images of a piecewise planar scene cannot be implemented in large image datasets, for which the images that are far away from the one where the GCPs have been measured are georeferenced using a big number of homography transformations (e.g., more than 20

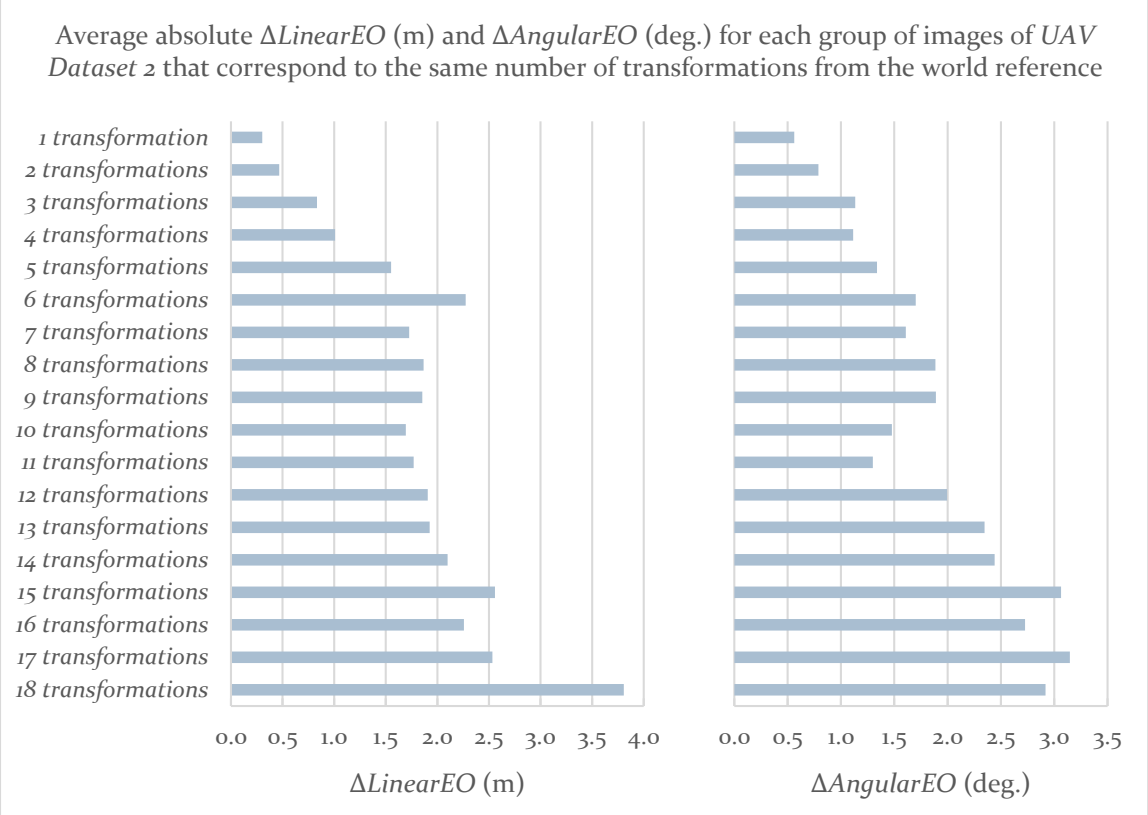


Figure 5-13. Average absolute Δ*LinearEO* (m) and Δ*AngularEO* (deg.) for the developed *CP-based EO Estimation* method for each group of images of *UAV Dataset 2* that corresponds to the same number of transformations from the world reference system

transformations); in that case, the corresponding error would be large and thus prevent these values from being used as initial exterior orientation parameters in a bundle adjustment procedure. In such large datasets, the approximate exterior orientation scheme presented in this chapter would yield satisfying results if GCPs were measured in an additional image, thus enforcing the maximum total number of homography transformations from an image to the reference coordinate system to be less than a defined threshold (e.g., 20 transformations).

# Chapter 6
# Detection of Vertical and Horizontal Vanishing Points

## 6.1 Introduction

Under the assumption of a pinhole camera model, projections of parallel space lines in an image intersect in one point, the so-called vanishing point of the direction of the lines. A vanishing point of a direction is defined as the intersection of the image plane with a projection ray, parallel to that direction. Vanishing points may be finite points, located either inside or outside the image limits, or points at infinity, if image lines are parallel to the real world lines. The latter case results from an image plane parallel to the direction of the parallel real world lines.

Oblique aerial images depict both vertical structures and horizontal structures of urban environments, which are characterized by the existence of many parallel lines, usually generated by edges of buildings and roads. Thus, the geometry of oblique aerial images of man-made environments favors the automatic extraction of sets of parallel lines, mainly horizontal and vertical ones, and thereby the automatic detection of their vanishing points. In addition, knowledge of these quantities is the foundation for many higher level tasks, including oblique imagery metric exploitation. In this chapter, an automatic method for extracting horizontal vanishing points and the vertical vanishing point (i.e., the nadir point) of oblique aerial images, along with their corresponding horizontal and vertical line segments, is introduced and two applications based on the knowledge of these elements are discussed, namely:

- the measurement of vertical and horizontal distances from a single oblique aerial image; and
- the estimation of approximate exterior orientation parameters of oblique aerial images.

## 6.2 Related Work

The automatic extraction of vanishing points is one of the most fundamental tasks in computer vision. Knowledge of vanishing points provides useful information for several kind of applications, like single-image measurements (Criminisi et al., 2000), autonomous robot navigation (Shi and Samarabandu, 2006; Chang et al., 2012), road detection (Kong et al., 2009; Moghadam et al., 2012), camera calibration (Cipolla et al., 1999b; Košecká and Zhang, 2002; Grammatikopoulos et al., 2006; Grammatikopoulos, 2007; Grammatikopoulos et al., 2007; Antunes et al., 2017), facade detection (Liu and Liu, 2014), camera orientation (Košecká and Zhang, 2002; Košecká and Zhang, 2005; Ding et al., 2008; Schindler et al., 2008; Baatz et al., 2010), multi-image rotational registration (Antone and Teller, 2000; Teller et al., 2003), image rectification (Baatz et al., 2010) and extraction of dominant rectangular structures (Košecká and Zhang, 2005).

One class of vanishing point detection methods works in image space, either using the image plane itself for representing vanishing points or using the projective plane for treating finite and infinite vanishing points equally (e.g., Minagawa et al., 2000; Rother, 2002; Wildenauer and Vincze, 2007; Grammatikopoulos, 2007; Grammatikopoulos et al., 2007; Tardif, 2009; Nieto and Salgado, 2011; Zhang and Koch, 2013). Another class of methods considers a projection sphere, rather than a plane, for representation of vanishing points (e.g., Antone and Teller, 2000; Ribeiro and Hancock; Košecká and Zhang, 2002; Mirzaei and Roumeliotis, 2011; Kroeger et al., 2015). Most of these approaches operate on the so-called Gaussian sphere, originally proposed by Barnard (1983), i.e., a unit sphere centered at the projection center of the camera that represents all directions in the 3D space. Its main characteristic is the fact that it may represent all vanishing points within a finite space, even those ones that do not lie within the image domain, being at infinity. Whereas the Gaussian sphere representation assumes a calibrated camera, approaches that operate on a projection sphere without interior orientation information have also been proposed (e.g., Košecká and Zhang, 2002). Less common is the use of a dual space in which points become lines and converging lines become aligned points, as adopted by the vanishing point detection algorithms proposed by Zhao et al. (2013) and Lezama et al. (2014).

In order to simplify the vanishing point detection problem, a large number of methods is based on the Manhattan world assumption (e.g., Košecká and Zhang, 2002; Rother, 2002; Grammatikopoulos, 2007; Grammatikopoulos et al., 2007; Tardif, 2009; Mirzaei and Roumeliotis, 2011; Zhang and Koch, 2013), according to which, lines are orthogonal or parallel to each other. The Manhattan world scenario implies the existence of three dominant orthogonal vanishing points, assuming the following possible configurations for a line triplet: (i) lines are orthogonal to each other; (ii) two lines are parallel and the third one is orthogonal to them; (iii) all lines are parallel to each other. A few approaches (e.g., Schindler and Dellaert, 2004; Antunes and Barreto, 2013) are based on an extension of the Manhattan world, namely the Atlanta world, introduced by Schindler and Dellaert (2004), which includes multiple Manhattan world configurations (i.e., multiple groups of orthogonal vanishing directions) that share one vanishing direction and is applicable in cases where not all vanishing directions are aligned along the same Manhattan grid. Also, methods that do not make any Manhattan world assumption have been proposed (e.g., Almansa et al., 2003; Zhai et al., 2016; Zhou et al., 2017; Zhang et al., 2018).

As far as the methods used for grouping edges or lines and corresponding vanishing directions are concerned, several researchers have implemented expectation-maximization-based methods (e.g., Košecká and Zhang, 2002; Wildenauer and Vincze, 2007; Schindler and Dellaert, 2004; Nieto and Salgado, 2011; Minagawa et al., 2000; Teller et al., 2003). RANSAC-based algorithms have also

been applied for clustering and detecting vanishing points (e.g., Wildenauer and Vincze, 2007; Mirzaei and Roumeliotis, 2011; Zhang and Koch, 2013; Zhao et al., 2013). Other approaches rely on J-linkage (e.g., Tardif, 2009) and direct clustering (e.g., Lezama et al., 2014).

The developed approach works on image space for extracting multiple horizontal vanishing points, along with the true horizon line of each input oblique aerial image, with the ultimate goal of detecting the vanishing point of the vertical direction, that is, the nadir point of each image. Whereas the workspace of a Gaussian sphere is an attractive alternative for handling infinite vanishing points equally to finite ones, the case of any infinite horizontal or vertical vanishing point is not applicable for oblique aerial images, as their image plane is neither horizontal nor vertical; thus, the horizontal and vertical image lines are not parallel to the real world lines.

Whereas the developed method assumes the existence of both horizontal and vertical lines, it does not rely on the assumption of the Manhattan world, as it only assumes the existence of a vertical vanishing direction and several horizontal vanishing directions, which do not have to be orthogonal to each other. Thus, it is more robust for being applied in oblique aerial images, compared to the ones assuming a Manhattan world, as it does not imply the availability of three orthogonal vanishing points in the scene. Especially, the latter assumption may be non-realistic for oblique aerial images of man-made scenes that depict several structures (e.g., buildings) with horizontal edges of quite differing orientations, not necessarily orthogonal to each other. Besides, the availability of a vertical vanishing direction and several horizontal vanishing directions in oblique imagery of urban areas is a realistic assumption, which simplifies the final nadir point detection goal of the developed method.

The developed method also deals with the challenging case of detecting the nadir point of oblique aerial images depicting vertical line segments that are either small compared to the horizontal segments and/or do not correspond to a significant number, compared to the rest of the visible line segments in the image, without any a priori knowledge of the vertical vanishing point position. This is the case for most oblique aerial images (mainly low oblique ones), as the imaged horizontal structures usually stand out more than the vertical ones. In this way, the developed method is targeted to the geometry of oblique aerial images of man-made environments. Whereas methods for detecting the vertical vanishing point have also been applied in oblique aerial imagery by a few researchers, they do not deal with the aforementioned challenging cases (e.g., short vertical line segments and/or vertical lines of a small number); they either make assumptions on the imaged scene (Xiao et al., 2012), or they apply a standard vanishing point detection scheme without exploiting further the oblique image geometry (Ding et al., 2008; Habbecke and Kobbelt, 2012). Specifically, Ding et al. (2008) estimated the nadir point of an oblique image using a Gaussian sphere representation, as proposed by Barnard (1983); Habbecke and Kobbelt (2012) used a RANSAC-based approach for nadir point computation; and Xiao et al. (2012), made the assumption that the vertical lines in object space also appear as vertical lines in image space (assuming an aircraft flying horizontally during image acquisition and oblique images significantly tilted in only one direction).

Finally, the developed method is based on a RANSAC framework for estimating horizontal vanishing points, the true horizon line and the nadir point. Its contribution compared to the other state-of-the-art RANSAC-based methods for vanishing point detection lies in the introduction of variations of the RANSAC algorithm, according to which the random choice of initial data is controlled through various criteria for estimation of the appropriate model; these constraints enable the developed variations of RANSAC to be applied even in datasets with a very high rate of outliers

(e.g., more than 80%), in which the application of a classical RANSAC-based approach may fail (Debnath et al., 2015; Hasheminasaba et al., 2015).

## 6.3 Developed Vanishing Point Detection Method

In the following, the developed method for automatic estimation of horizontal and vertical line segments, horizontal vanishing points, the horizon line and the nadir point of an oblique aerial image of a man-made environment is presented. It requires an estimation of the camera interior orientation parameters, i.e., the camera constant and the principal point, along with the lens distortion parameters, if they are available. Mainly, in the case of images taken by a camera with a low quality lens system, the lens distortion effect has to be corrected before the step of extracting edges and line segments in the image, because straight lines in space would be mapped to curves in the images, thus leading to an inaccurate estimation of the nadir point. In this case, the principal point, the camera constant and the lens distortion coefficients have to be a priori known, according to the method presented in this section. However, methods for estimating the camera interior orientation parameters and vanishing points simultanesouly have also been proposed in literature. For instance, Grammatikopoulos (2007) and Grammatikopoulos et al. (2007) presented an approach for automatic estimation of interior orientation from images with three vanishing points of orthogonal directions that involves extraction of image line segments and their clustering into groups corresponding to three dominant vanishing points; the camera interior orientation parameters (camera constant, location of principal point, and two coefficients of radial lens distortion) and the vanishing points are estimated in a one-step adjustment of all participating image points. Also, Förstner (2010) presented a vanishing point scheme that can easily be extended to determine parameters for lens distortion, as proposed by Grammatikopoulos (2007) and Grammatikopoulos et al. (2007). In the case that the lens distortion effect can be neglected, rough estimations for the camera constant and the principal point suffice, according to the developed methodology. If such information is not available, it can be roughly computed via the method discussed in section 6.3.1. The developed workflow, assuming knowledge of the camera interior orientation parameters, is illustrated in Figure 6-1. Table 6-1 presents indicative values for the thresholds that can be used by the developed methodology.

### 6.3.1 Rough Camera Interior Orientation Estimation

In the case that the lens distortion effect in the input oblique aerial image cannot be neglected, the principal point, the camera constant and the lens distortion coefficients have to be a priori known and the image has to be undistorted before any further processing, according to the developed methodology. However, in the case that the lens distortion effect can be neglected and camera interior orientation parameters are unavailable, the process described in this section can be implemented.

Specifically, if an undistorted oblique aerial image is used or the lens distortion effect can be neglected, the camera constant and the location of the principal point are used (i) for controlling the random choice of initial candidate horizontal line segments during the implementation of the RANSAC algorithm for true horizon detection; (ii) for computing an initial estimation of the nadir point using the detected true horizon line; and (iii) for controlling the random choice of initial candidate vertical line segments during the implementation of the RANSAC algorithm for nadir point detection. Thus, rough estimates suffice, as they are only used in the process of obtaining

Table 6-1. Parameter values used by the developed nadir point detection algorithm

| Symbol | Value |
|---|---|
| *distLineSegmentVPThres* | 50 pixels |
| *maxVanishingPointsThres* | 20 |
| *minRemainingSegmentsThres* | 10 |
| *distVPsHorizonLineThres* | 250 pixels |
| *distNadirVerticalLineThres* | min[2000 pixels, maxDimension/3] |
| *distFromInitialNadirThres* | min[800 pixels, maxDimension/8] |
| *angularThres* | 10° |
| *distLineSegmentNadirThres* | 200 pixels |

approximate estimations of the nadir point; its final estimation is obtained without any requirement for camera interior orientation parameters.

If the camera interior orientation is unknown, these parameters can be roughly estimated. The principal point may be considered to coincide with the image center. The camera constant may be determined using the estimation of the principal point if two vanishing points and the angle between the two sets of parallel 3D lines that converge to these vanishing points are known (Förstner et al., 2004). Hence, it may be calculated using the vanishing points of two perpendicular directions, which may be easily indicated manually in an image of a man-made scene without particular knowledge of the scene geometry. This is achieved by measuring two parallel horizontal line segments and two vertical ones. The intersection point of each set of parallel segments is the vanishing point of the corresponding direction. The camera constant ($c$) is then calculated according to equation (6.1), where $x_P$, $y_P$ are the image coordinates of the principal point, $x_N$, $y_N$ are the image coordinates of the nadir point and $x_H$, $y_H$ are the image coordinates of the vanishing point of a horizontal direction (Karras et al., 1993).

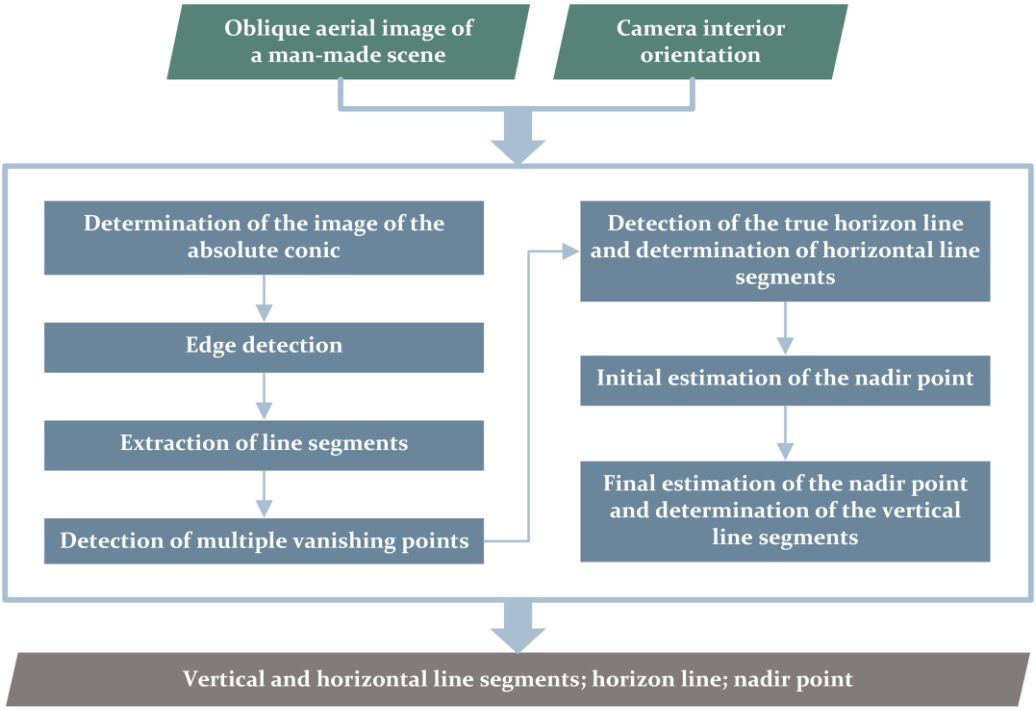


Figure 6-1. The developed workflow for determining vertical and horizontal line segments, the horizon line and the nadir point of a calibrated oblique aerial image of a man-made scene

$$c = \sqrt{\left| -\left(x_H - x_P\right)\left(x_N - x_P\right) - \left(y_H - y_P\right)\left(y_N - y_P\right) \right|} \tag{6.1}$$

Except for the aforementioned method for obtaining an estimation of the camera constant using two vanishing points based on assumption on the principal point, which has been implemented in the developed software and used in the experiments, other methods for estimating the camera interior orientation parameters from a single image of a man-made environment have also been reported in literature and can also be used for estimation of the camera calibration parameters. Grammatikopoulos et al. (2003) presented mathematical models for single image calibration, relying on measurements of straight lines in three orthogonal directions, including camera calibration based on (i) the direct use of the three basic image vanishing points; (ii) the use of image line parameters; and (iii) the direct use of image point observations. Several researchers have dealt with camera calibration of single images of man-made scenes using knowledge of three vanishing points, like Caprile and Torre (1990), Cipolla et al. (1999a), Deutscher et al. (2002) and Rother (2002). Also, the computation of the camera constant of a single image using two vanishing points, assuming knowledge of the principal point, has been studied, e.g., by Guillou et al. (2000). Methods for multi-image camera calibration using three vanishing points of orthogonal directions have also been studied in literature (Xie et al., 2004; Grammatikopoulos et al., 2006; Grammatikopoulos et al., 2007; Xie and Xue, 2016). Moreover, the identification of two vanishing points in multiple images from the same camera permits the estimation of its camera interior orientation parameters through a multi-image approach, as it has been shown by Grammatikopoulos et al. (2004), or a two-image approach, as it has been implemented by Lugang et al. (2014). Finally, unlike approaches based on vanishing points, line algorithms have also been implemented for the scope of camera calibration of single images, assuming the existence of three mutually orthogonal directions, like the method presented by Karras and Petsa (1999) relying on the line algorithm developed by Petsa and Patias (1994).

### 6.3.2 Determination of the Image of the Absolute Conic

Once the camera is calibrated, the image of the absolute conic is computed. The absolute conic is a second degree curve lying on the plane at infinity, consisting of complex points (Hartley and Zisserman, 2003; Kalisperakis, 2010). The image of this curve is an imaginary point conic that depends on the camera interior orientation, being independent on the camera position or orientation. Thus, for all images taken by the same camera, the image of the absolute conic, $\boldsymbol{\omega}$, is the same. It is represented by a 3×3 matrix and is calculated via equation (6.2), using the calibration matrix $\boldsymbol{K}$ and considering zero skew and unit aspect ratio, to wit, square pixels.

$$\boldsymbol{\omega} = \left(\boldsymbol{K}\boldsymbol{K}^T\right)^{-1}; \qquad \boldsymbol{K} = \begin{bmatrix} c & 0 & x_P \\ 0 & c & y_P \\ 0 & 0 & 1 \end{bmatrix} \tag{6.2}$$

### 6.3.3 Extraction of Line Segments

The initial step is the extraction of edges using the Canny operator (Canny, 1986) in the image, after it has been converted to greyscale. In the case of high accuracy requirements or if the image is taken by a camera with a low quality lens system, the lens distortion effect has to be corrected over the image before the extraction of Canny edges. The “high” threshold, which is used by the

Canny operator in the hysteresis thresholding step, is automatically calculated according to the Otsu algorithm (Otsu, 1979) as proposed by Fang et al. (2009), while the "low" threshold is set to be the half of the "high" one. As far as the Canny thresholds are concerned, if the gradient of a pixel is higher than the high threshold, it is accepted as an edge; if it is below the low threshold, it is rejected; if it is between these two thresholds, it is accepted in the case that it is connected to a pixel with a gradient above the high threshold. Subsequently, the progressive probabilistic Hough transform (Matas et al., 1998) is applied in the binary edge map obtained using the Canny operator for the extraction of line segments.

### 6.3.4 Detection of Multiple Vanishing Points

The detection of multiple vanishing points is the next step. A variation of the RANSAC algorithm is applied for the estimation of the intersection point of most line segments. According to this variation, the random iterative choice of two line segments is controlled so that the initial estimation of their intersection point lies outside the image limits, having a smaller *y* coordinate than the height of the image (assuming that the origin of the image coordinate system is located at the top left corner, *x*-axis points to the right and *y*-axis points downwards). This constraint is imposed in order to ensure that mostly vanishing points of horizontal directions will be detected, taking into consideration the fact that the nadir point is located below the bottom limit of an oblique aerial image and assuming that the true horizon line lies outside the image, towards its upper side. The inliers, i.e., the line segments that converge to the vanishing point estimated by RANSAC, are those that correspond to a distance from the estimated vanishing point smaller than a threshold (*distLineSegmentVPThres*). Subsequently, the estimation of the detected vanishing point is refined through least-squares adjustment, using its initial estimation and the inliers that converge to this point obtained by RANSAC. Then, the line segments that converge to the detected vanishing point are removed from the set of line segments. This procedure is repeated for the detection of multiple vanishing points, either until a maximum number of vanishing points (*maxVanishingPointsThres*) has been detected or until a minimum number of line segments remains (*minRemainingSegmentsThres*).

### 6.3.5 Detection of the True Horizon Line

The true horizon line is the intersection of the horizontal plane that contains the projection center and the oblique image plane (section 2.4.1). All horizontal line segments converge to points on this line, which consists of the vanishing points of horizontal directions. The true horizon line is estimated via a variation of the RANSAC algorithm as the straight line that connects the largest number of the vanishing points detected through the method presented in section 6.3.4. Thus, the iterative random choice of two vanishing points is controlled in order to ensure that the initial estimation of the nadir point, which is obtained based on the estimation of the equation of the true horizon line using the two random vanishing points, as described in the following paragraph, has a greater *y* coordinate than the image height.

The first step for obtaining an initial estimation of the nadir point is the determination of the equation of the principal line, which is perpendicular to the true horizon line and passes through the principal point. Then, the position of the horizon point $K'$ (section 2.4.1) is determined as the intersection of the true horizon line and the principal line. The distance between the horizon point and the principal point is then computed. The depression angle, $\theta$, is subsequently calculated using equation (2.23) (section 2.4.5.1) and the distance $P'N'$ between the principal point and the nadir

point is then estimated, according to equation (2.25) (section 2.4.5.1). The nadir point coordinates are determined using the distance $P'N'$ and the coordinates of the principal point, taking into consideration the fact that the nadir point lies on the principal line. In the resulting system of two equations, the solution according to which the nadir point is located below the principal point is the correct one.

The inliers, i.e., the vanishing points that converge to points on the horizon line estimated by RANSAC, are those that correspond to a distance from the estimated horizon line smaller than a threshold (*distVPsHorizonLineThres*). The estimation of the true horizon line obtained by RANSAC is refined through least-squares adjustment using the vanishing points that are characterized as inliers by RANSAC. In this way, the final estimation of the true horizon line is obtained.

### 6.3.6 Estimation of the Nadir Point

After the final estimation of the true horizon line through least-squares adjustment, the equation of the principal line is determined, the horizon point is computed, the distance between the horizon point and the principal point is estimated, the depression angle and the distance between the principal point and the nadir point are calculated and an estimation of the nadir point coordinates is obtained (see also section 6.3.5). However, this is not the final solution, because the errors arising from the calculation of multiple horizontal vanishing points and the estimation of the equation of the true horizon line are involved.

The horizontal line segments that constitute the inliers are removed from the set of line segments. If the distance between the initial estimation of the nadir point and a line that passes through the endpoints of a remaining segment is greater than a predefined value (*distNadirVerticalLineThres*), this segment is considered to be non-vertical and is rejected. A variation of the RANSAC algorithm is used for the final estimation of the nadir point, using the non-rejected line segments. According to this variation, the iterative estimation of the nadir point using two line segments is considered to be correct if (i) the *y* coordinate of the nadir point exceeds the image height; (ii) the distance between the nadir point and its initial estimation, obtained using the true horizon line, is smaller than a threshold (*distFromInitialNadirThres*); and (iii) the angle between the line segments that converge to the nadir point and the line segments that converge to a horizontal vanishing point (e.g., the horizon point) differs from the right angle by an angular threshold (*angularThres*). Let $\boldsymbol{v}_N$ be the nadir point estimation and $\boldsymbol{v}_{Hi}$ be the vanishing point of a horizontal direction, both in homogeneous coordinates; the cosine of the angle $\mu$ between the lines that converge to these vanishing points is computed via equation (6.3), using the image of the absolute conic $\boldsymbol{\omega}$ (Hartley and Zisserman, 2003).

$$\cos\mu = \frac{\boldsymbol{v}_N^{\mathrm{T}}\boldsymbol{\omega}\boldsymbol{v}_{H_i}}{\sqrt{\left(\boldsymbol{v}_N^{\mathrm{T}}\boldsymbol{\omega}\boldsymbol{v}_N\right)\left(\boldsymbol{v}_{H_i}^{\mathrm{T}}\boldsymbol{\omega}\boldsymbol{v}_{H_i}\right)}} \tag{6.3}$$

The inliers, i.e., the vertical line segments that converge to the nadir point estimated by RANSAC, are those that correspond to a distance from the estimated vanishing point smaller than a threshold (*distLineSegmentNadirThres*). The final estimation of the nadir point is obtained through least-squares adjustment using all line segments that converge to the RANSAC-based estimation of the nadir point, returned by RANSAC.

## 6.4 Metric Exploitation of a Single Oblique Image

The estimation of distances in single aerial images for their metric exploitation is an issue of great importance. Significant work has been conducted using old datasets of images, mainly after World War II (Williams, 1969; Williamson and Brill, 1990). The mathematical relationships that may be used for calculation of vertical and horizontal distances in a single oblique aerial image of known interior orientation and flying height, based on the knowledge of the nadir point, have already been presented in section 2.5 of this dissertation. Thus, the methodology presented in section 6.3 can be applied for the detection of the vanishing point of the vertical direction, before the calculation of vertical and horizontal distances.

## 6.5 Approximate Exterior Orientation Estimation of Georeferenced Image Datasets of a Man-Made Scene

In this section, a method for the estimation of approximate exterior orientation parameters of the images of a man-made scene based on the knowledge of the nadir point of each image, estimated by the vanishing point detection method introduced in section 6.3, is presented. It is hereinafter referred to as *VP-based EO Estimation* method. It can be applied in georeferenced image datasets, i.e., images with known transformation parameters to the ground reference system, provided that their nadir point has been determined. Taking into account the fact that both horizontal and vertical line segments have to be clearly visible for the implementation of the nadir point detection technique described in section 6.3, the images have to depict at least one horizontal plane for the detection of the nadir point based on the developed method. Hence, the georeferencing method presented in section 5.4 for images of a piecewise planar scene can be applied for the scope of calculating the transformation parameters from each image to the reference coordinate system. The initial data required consist of the following:

- the homography parameters, calculated according to the georeferencing methodology described in section 5.4, that define the transformation from the plane of the coplanar GCPs (that are required for the implementation of the georeferencing methodology) of each image to the reference coordinate system;
- the nadir point of each image, computed according to the method presented in section 6.3;
- the $Z$ coordinate of the GCPs that define the plane of interest; and
- the camera interior orientation parameters; the latter ones may also be roughly estimated through the process described in section 6.3.1.

The *VP-based EO Estimation* method is applied in each georeferenced image separately. It can be applied for the estimation of approximate exterior orientation parameters of an image dataset for being adjusted inside a global SfM process, as it is discussed in chapter 7 of this dissertation, in the case that the images depict a piecewise planar scene where horizontal and vertical line segments are clearly visible. The developed workflow is illustrated in Figure 6-2. In the following paragraphs, the developed orientation procedure, as applied in each image separately, is described.

The image coordinates of the nadir point and the principal point of every image are transformed to the reference coordinate system using their image coordinates and the computed homography transformation parameters. The coordinates of the nadir point in the reference coordinate system are the horizontal coordinates of the projection center ($X_0$, $Y_0$).

The tilt angle ($t$) is computed according to equation (2.24) (section 2.4.5.1) using the image coordinates of the principal point and the nadir point as well as the camera constant. The swing angle ($s$) is calculated according to equation (2.35) (section 2.4.5.3), using the image coordinates of the principal point and the nadir point. The azimuth angle ($a$) is computed via equation (6.4).

$$a' = \tan^{-1}\left|\frac{X_P - X_N}{Y_P - Y_N}\right| \begin{cases} a = a' \ \ if \ X_P > X_N \ and \ Y_P > Y_N \\ a = \pi - a' \ \ if \ X_P > X_N \ and \ Y_P < Y_N \\ a = \pi + a' \ \ if \ X_P < X_N \ and \ Y_P < Y_N \\ a = 2\pi - a' \ \ if \ X_P < X_N \ and \ Y_P > Y_N \\ a = 0 \ \ if \ X_P = X_N \ and \ Y_P \geq Y_N \\ a = \pi \ \ if \ X_P = X_N \ and \ Y_P < Y_N \\ a = \pi/2 \ \ if \ X_P > X_N \ and \ Y_P = Y_N \\ a = 3\pi/2 \ \ if \ X_P < X_N \ and \ Y_P = Y_N \end{cases} \tag{6.4}$$

As far as the third coordinate of the projection center ($Z_0$) is concerned, it requires the estimation of the flying height, measured from the plane of GCPs, which is computed through equation (6.5), using the tilt angle and the horizontal distance $PN$ between the transformed principal point ($X_P$, $Y_P$) and nadir point ($X_N$, $Y_N$) in the reference coordinate system (Figure 6-3). Then, $Z_0$ is calculated via equation (6.6), using the computed flying height ($H$) and the elevation of the coplanar GCPs ($Z_{GCPs}$).

$$H = \frac{PN}{\tan t} \tag{6.5}$$

$$Z_0 = H + Z_{GCPs} \tag{6.6}$$

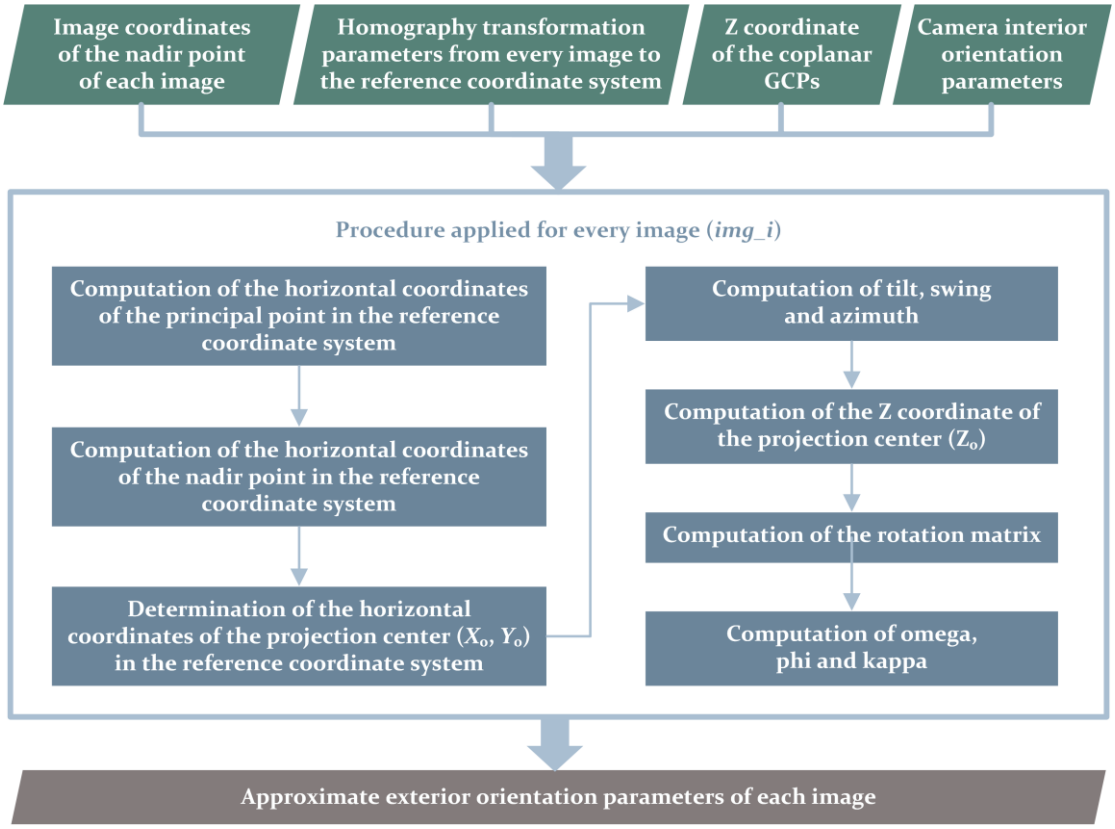


Figure 6-2. The developed workflow for computation of approximate exterior orientation parameters of an image dataset of a man-made scene through the *VP-based EO Estimation* method

Using the angles tilt, swing and azimuth, the 3×3 rotation matrix, $\boldsymbol{R}$, is computed according to equation (2.1) (section 2.4.2). Then, the orientation of the camera can be computed in terms of omega ($\omega$), phi ($\varphi$) and kappa ($\kappa$) according to equation (5.22) (section 5.5.2.2), using the elements, $r_{ij}$, of the rotation matrix $\boldsymbol{R}$.

## 6.6 Experimental Results

In this section, the developed software suite that implements the methodology presented in sections 6.3, 6.4 and 6.5 is presented, along with the test datasets and the results achieved.

### 6.6.1 Developed Software

A software suite that implements the proposed vanishing point detection methodology that extracts vertical and horizontal line segments and locates the true horizon line and the nadir point of images with the ultimate goals of (i) extracting vertical and horizontal distances from single images; and (ii) estimating approximate exterior orientation parameters for image datasets, has been developed. The vanishing point detection tool along with the tool for approximate exterior orientation estimation have been developed in the C# programming language, using the Microsoft .NET Framework and make use of some functionalities offered by the OpenCV library (OpenCV Team, 2019) for image manipulations, using its .Net wrapper, Emgu CV (Emgu CV Team, 2019). The tool for determining vertical and horizontal distances has been developed in the C++ programming language and makes use of some functionalities offered by the OpenCV library for image manipulations and the GDI+ library (Microsoft, 2019) for extracting image metadata. The following developed software tools were used for the experiments:

- a software tool for determining overlapping images (section 4.4.2.1);
- a software tool for detecting coplanar feature points in image datasets, lying on the plane of the coplanar GCPs (section 5.6.1);
- a software tool for georeferencing images, by computing their 2D homography transformation parameters to the coordinate reference system, based on the detected coplanar feature points (section 5.6.1);

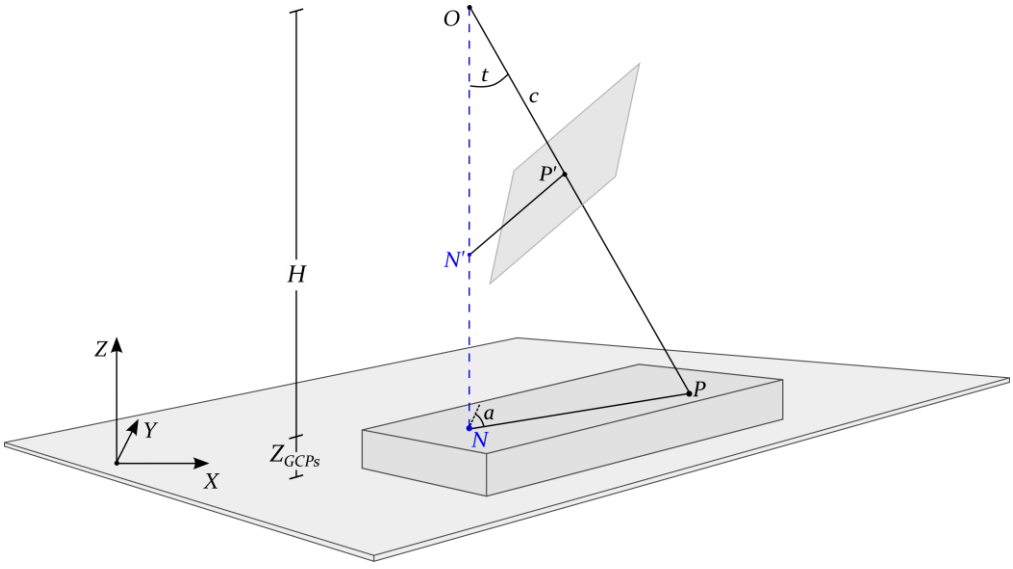


Figure 6-3. Oblique aerial image geometry: calculation of the flying height *H* from the plane of GCPs

- a software tool for extracting vertical and horizontal line segments, along with the true horizon line and the nadir point of single images; this tool requires the manual measurement of two horizontal and two vertical line segments in a single image in the case that the camera interior orientation is unknown;
- a software tool for extracting vertical and horizontal distances of single images based on the knowledge of their nadir point and, optionally, the camera interior orientation parameters; this tool requires the manual measurement of the vertical or horizontal distance; and
- a software tool for computing approximate exterior orientation parameters of georeferenced images based on the knowledge of the nadir point of each image.

### 6.6.2 Test Datasets and Reference Data

The developed software that implements the proposed algorithm was tested using two image datasets, i.e., *UAV Dataset 1*, which incorporates 23 images, and *UAV Dataset 2*, which incorporates 35 images, as described in section 5.6.2 (Figure 5-5). The reference data described in section 5.6.3, were used (both concerning the reference exterior orientation parameters and the ground coordinates of the coplanar GCPs measured in *img_1*). Furthermore, in order to obtain reference vertical and horizontal distances to be compared with the ones derived through the developed software, 15 horizontal and 15 vertical distances visible in different images of *UAV Dataset 2* were calculated through multi-image measurements in the oriented block of 645 images (see section 5.6.3) provided by the ISPRS/EuroSDR benchmark for multi-platform photogrammetry using the Agisoft PhotoScan software. Specifically, the coordinates of the end points of the aforementioned line segments were computed through triangulation by Agisoft PhotoScan and the corresponding distances were estimated, to serve as reference distances (section 6.6.3.2).

### 6.6.3 Results

The coplanar point detection and georeferencing results obtained for *UAV Dataset 1* and *UAV Dataset 2* have been presented in sections 5.6.4.1 and 5.6.4.2 of this dissertation. The coplanar point detection and georeferencing methodology are required for computing approximate exterior orientation parameters for image datasets through the *VP-based EO Estimation* method. In this section, the results of (i) the developed vanishing point detection methodology; (ii) the method for computing distances from single images and (iii) the *VP-based EO Estimation* method applied in *UAV Dataset 1* and *UAV Dataset 2* through the developed software tools are outlined. As far as the interior orientation parameters used in the experiments are concerned, approximate data were used, similarly to the experiments presented in chapter 5 (section 5.6.4), so that the exterior orientation results are comparable. Specifically, the nominal name of the focal length, extracted from the image metadata, was used as the camera constant of the images, while the principal point was assumed to coincide with the center of the images. However, the developed method for obtaining the approximate value of the camera constant, in case that knowledge of the focal length is not available (section 6.3.1), was also tested through the developed software suite. The camera constant computed through the method presented in section 6.3.1, using different combinations of images as well as vertical and horizontal line segments, differs approximately by 1.70% from the reference value, calculated via the self-calibrating bundle block adjustment (section 5.6.3). Thus, it could very well be used instead of the nominal value of the focal length in the experiments, which differs by

1.65% from the reference one. Finally, in addition to using the aforementioned nominal values of the focal length and principal point in the experiments conducted for the determination of distances from a single oblique image, tests using another calibration scenario were also performed.

#### 6.6.3.1 Detection of Horizontal and Vertical Line Segments

Figure 6-4 illustrates the basic steps of the automatic vanishing point extraction process for an image of *UAV Dataset 2*. Specifically, it shows: (i) the detected line segments (top-left) using the progressive probabilistic Hough transform (section 6.3.3); (ii) the detected line segments that converge to multiple vanishing points, in different color based on their direction (top-right), i.e., those returned by the developed variation of the RANSAC algorithm that estimates vanishing points that are possibly horizontal (section 6.3.4); (iii) the detected horizontal line segments (bottom-left), i.e., those ones returned by the developed variation of the RANSAC algorithm that estimates the true horizon line (section 6.3.5); (iv) the detected vertical line segments (bottom-right), i.e., those returned by the developed variation of the RANSAC algorithm that estimates the nadir point (section 6.3.6). From Figure 6-4 (top-right) it is observed that some sets of parallel space lines have been erroneously identified by the developed vanishing point extraction algorithm as belonging to different directions, because the corresponding imaged lines have been found to converge to different – yet close to one another – vanishing points. An increase of the RANSAC threshold used for the detection of multiple vanishing points (*distLineSegmentVPThres*, section 6.3.4) could fix this, but it would also involve the risk of identification of non-parallel space lines as parallel. Besides, these line segments and the corresponding vanishing points only affect the detection of the horizon line, which is used for obtaining an approximate solution for the nadir point and not the

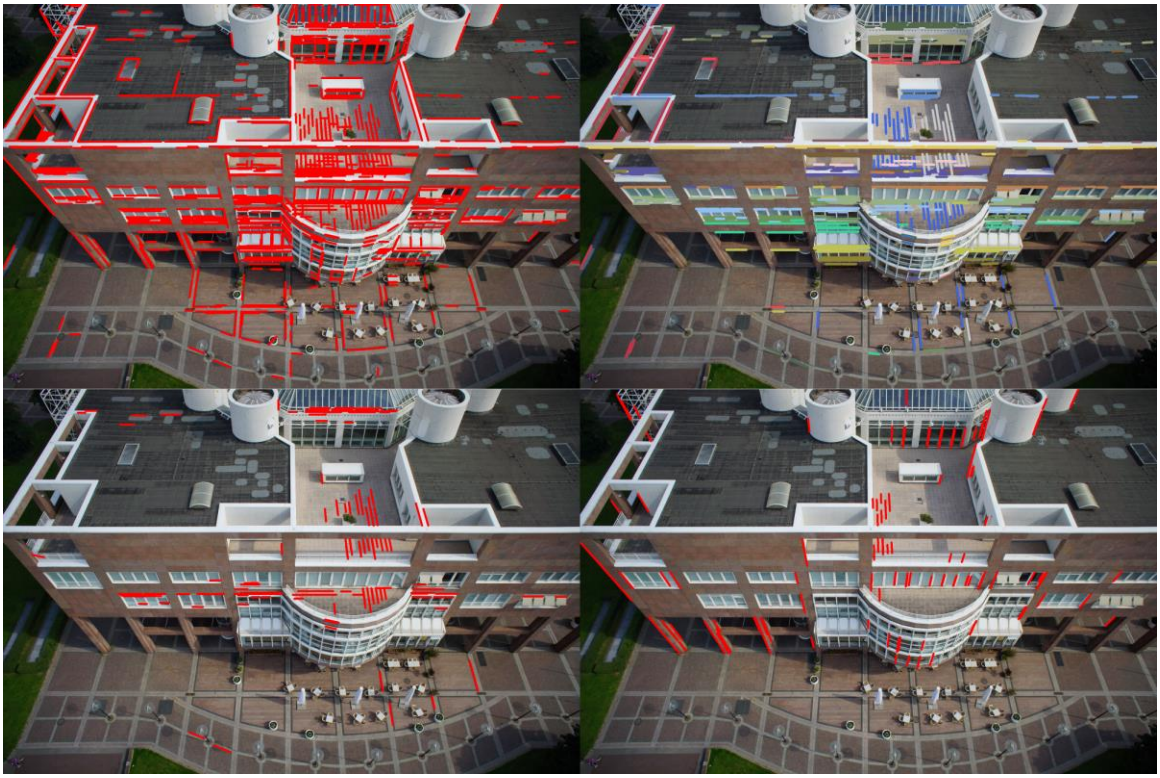

Figure 6-4. Automatic vanishing point extraction process for an image of *UAV Dataset 2*: detected line segments (top-left); detected horizontal line segments, in different color based on their direction (top-right); detected horizontal line segments (bottom-left); detected vertical line segments (bottom-right)

final one. Hence, the fact that some parallel space lines have not been identified as parallel by the developed algorithm neither affects the final result, i.e., the nadir point, nor the detected horizontal (Figure 6-4, bottom-left) and vertical (Figure 6-4, bottom-right) line segments. Figure 6-4, bottom-left depicts those line segments from Figure 6-4, top-right that have been categorized as horizontal, i.e., their vanishing points are used for the detection of the horizon line.

An average number of 1100 line segments is detected in each image of *UAV Dataset 1* and *UAV Dataset 2*, out of which an average number of 100 line segments is identified as horizontal and an average number of 100 line segments is identified as vertical. Hence the percentage of horizontal line segments identified by the developed method is about 10% of the number of extracted line segments; the same applies to the vertical line segments. Whereas the vast majority of horizontal line segments that are detected through the method described in section 6.3.5 is correct (Figure 6-5), some non-horizontal line segments are wrongly identified as being horizontal (Figure 6-6), because they happen to converge close to a correctly detected horizontal line segment. However, they do

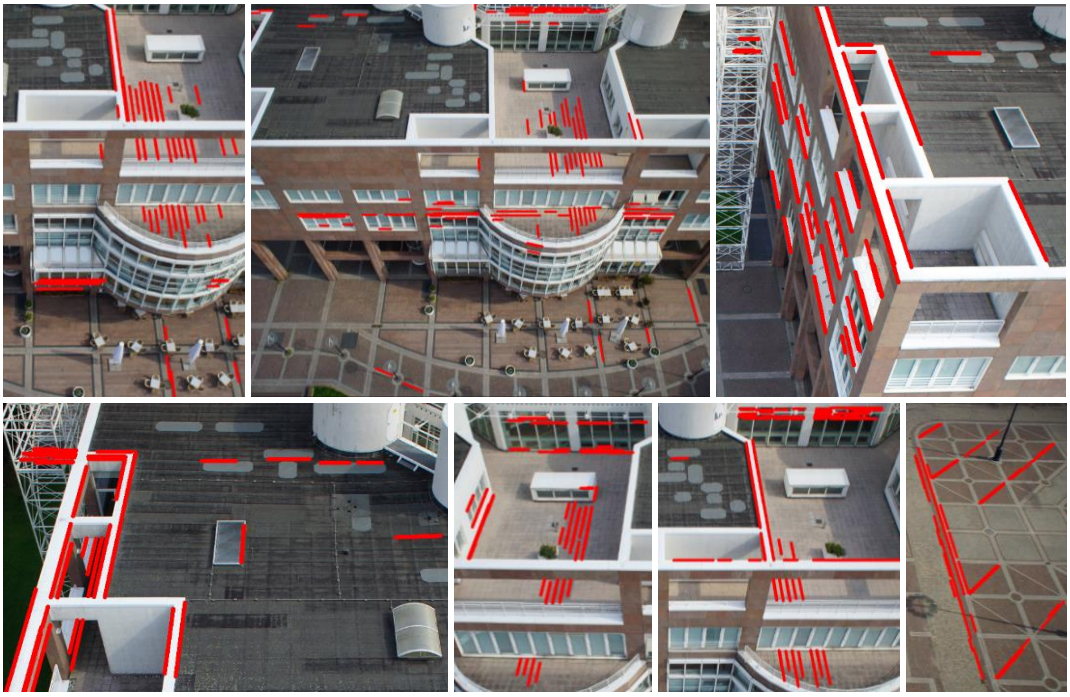

Figure 6-5. Automatically detected correct horizontal line segments superimposed in magnified parts of images of *UAV Dataset 2*

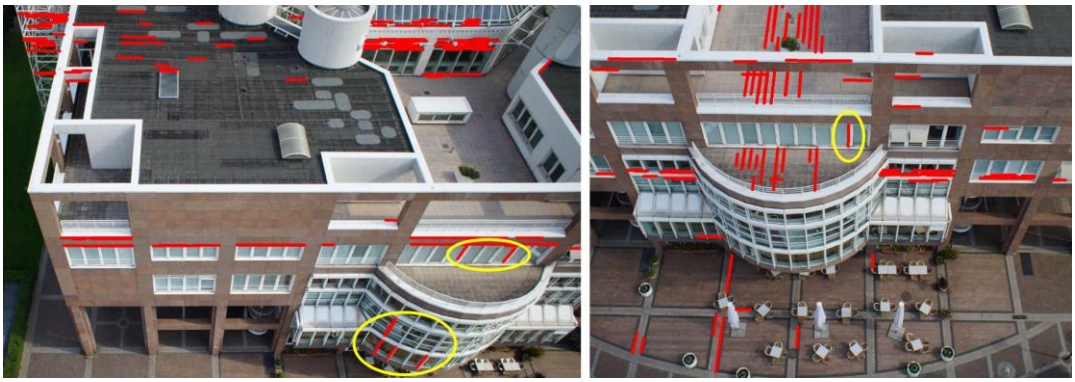

Figure 6-6. Automatically detected horizontal line segments superimposed in magnified parts of images of *UAV Dataset 2*, among which the wrongly identified horizontal line segments are marked by yellow ellipses

not affect the correctly detected horizontal vanishing point. Similarly, the great majority of line segments that are identified through the method presented in section 6.3.6 as being vertical is correct (Figure 6-7); some non-vertical line segments that converge close to the correctly detected nadir point are wrongly identified as vertical ones (Figure 6-8). However, these also do not affect the final goal of the whole methodology, i.e., the detection of the vertical vanishing point. The visual inspection of the results proves that the average percentage of correctly identified vertical and

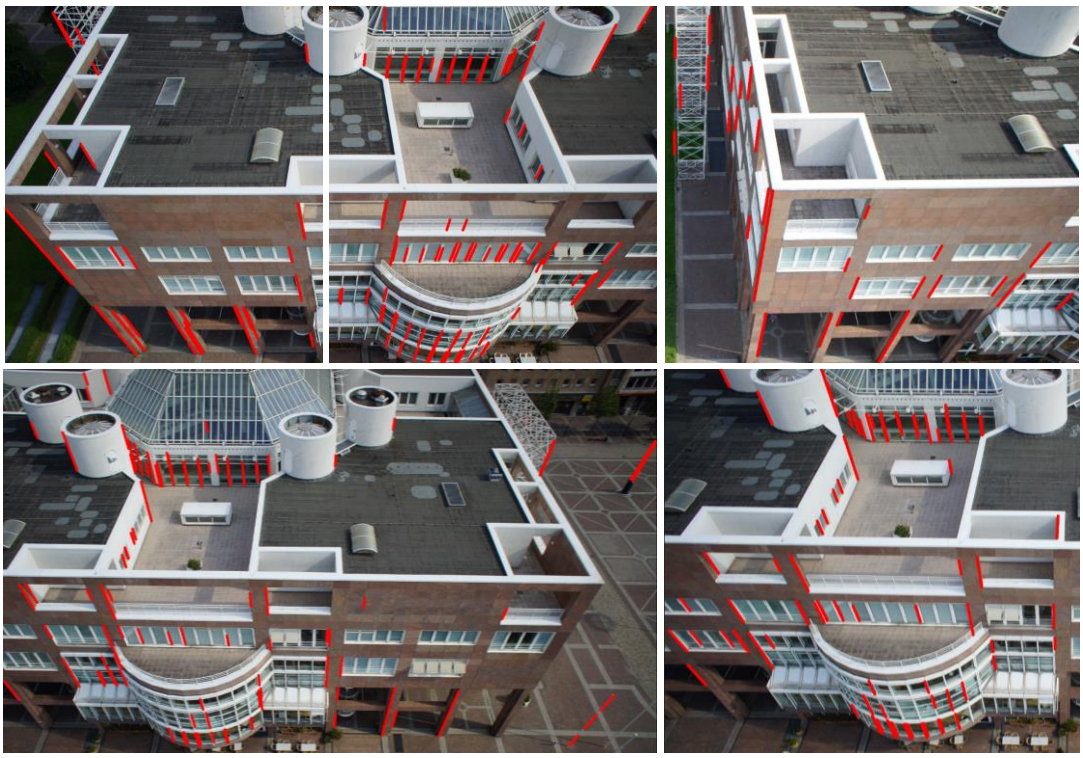

Figure 6-7. Automatically detected correct vertical line segments superimposed in magnified parts of images of *UAV Dataset 2*

Figure 6-8. Automatically detected vertical line segments superimposed in magnified parts of images of *UAV Dataset 2*, among which the wrongly identified vertical line segments are marked by yellow ellipses

horizontal line segments exceeds 95%. Figure 6-9 illustrates the vertical vanishing points detected in all test images via the developed framework, i.e. those converging to the detected nadir point.

Figure 6-9. Automatically detected vertical line segments in the images of *UAV Dataset 2*

Table 6-2. Differences between the vertical distances computed through the developed solution and the reference ones for *UAV Dataset 2*

| Scenario | Avg-Abs | Max-Abs | Min-Abs | Avg | Stdev | Avg-R | Max-R | Min-R |
|---|---|---|---|---|---|---|---|---|
| ***NominalVal*** | 0.34 m | 0.66 m | 0.02 m | 0.32 m | 0.22 m | 3.79% | 5.58% | 1.00% |
| ***CalibratedVal*** | 0.30 m | 0.58 m | 0.02 m | 0.27 m | 0.21 m | 3.51% | 5.34% | 0.81% |

Table 6-3. Differences between the horizontal distances computed through the developed solution and the reference ones for *UAV Dataset 2*

| Scenario | Avg-Abs | Max-Abs | Min-Abs | Avg | Stdev | Avg-R | Max-R | Min-R |
|---|---|---|---|---|---|---|---|---|
| ***NominalVal*** | 0.60 m | 1.38 m | 0.07 m | 0.60 m | 0.46 m | 3.07% | 4.32% | 1.13% |
| ***CalibratedVal*** | 0.25 m | 0.93 m | 0.02 m | 0.20 m | 0.27 m | 1.65% | 4.82% | 0.17% |

#### 6.6.3.2 Determination of Vertical and Horizontal Distances from a Single Image

As far as the experiments conducted for the extraction of distances from a single image are concerned, horizontal distances lying on the ground plane and vertical distances from the ground plane were measured. An average flying height was used as input in the developed software for all images of the experiments. Also, the nadir point calculated through the method presented in section 6.3 for each image was used. 15 vertical line segments, ranging from 1.8 m to 21.4 m and 15 horizontal line segments of differing orientations, ranging from 1.9 m to 52.3 m, were measured across different images of *UAV Dataset 2* (that includes the images of *UAV Dataset 1*) through the developed software that implements the proposed framework (section 6.4) under two different calibration scenarios. According to the first scenario (*NominalVal*), the nominal value of the focal length, extracted from the image metadata, approximates the camera constant and the principal point is considered to coincide with the image center. According to the second scenario (*CalibratedVal*), the camera constant and the principal point coordinates, as calculated by the self-calibrating bundle adjustment in the block of 645 UAV images (section 5.6.3) are used, assuming zero distortion. The measured vertical and horizontal distances were computed through the developed software for the aforementioned two camera calibration scenarios, in order to investigate the impact of the camera calibration information on the results and assess the results that can be achieved without precise calibration information. The results of the measured vertical and horizontal distances under the two calibration scenarios were compared with the reference values of the same distances, which were estimated using the corresponding triangulated points through Agisoft PhotoScan (section 6.6.2) and are shown in Table 6-2 and Table 6-3. Bar charts illustrating the reference values of the vertical and horizontal distances measured by the developed software for the aforementioned two calibration scenarios, along with the absolute differences between the computed distances and the reference ones are shown in Figure 6-10 and Figure 6-11, respectively.

Specifically, the average absolute difference (Avg-Abs) between the calculated distances through the developed method and the reference ones, along with the maximum absolute difference (Max-Abs), the minimum absolute difference (Min-Abs), the average difference (Avg), the standard deviation of differences (Stdev), the average difference-to-distance ratio (Avg-R), the maximum difference-to-distance ratio (Max-R) and the minimum difference-to-distance ratio (Min-R) are presented in Table 6-2 (for the case of vertical distances) and Table 6-3 (for the case of horizontal distances). The difference between the vertical distances calculated through the developed algorithm and the reference ones range from 0.010$h$ to 0.056$h$ in the case of the *NominalVal* scenario and from 0.008$h$ to 0.053$h$ in the case of the *CalibratedVal* scenario, where $h$ is the

reference value of the corresponding vertical distance. The difference between the horizontal distances calculated through the developed algorithm and the reference ones range from 0.011*d* to 0.043*d* in the case of the *NominalVal* scenario and from 0.002*d* to 0.048*d* in the case of the *CalibratedVal* scenario, where *d* is the reference value of the corresponding horizontal distance. From the results outlined in Table 6-2 and Table 6-3 along with the bar charts illustrated in Figure 6-10 and Figure 6-11, it is deduced that the usage of calibrated interior orientation parameters improves the accuracy of measurements of both horizontal and vertical distances. However, the results obtained using the nominal camera parameters or, generally, approximate camera interior orientation parameters are also acceptable for purposes of extracting monocular measurements from an uncalibrated image of unknown exterior orientation parameters.

### 6.6.3.3 Approximate Exterior Orientation Estimation

The method outlined in section 6.5 for approximate estimation of the exterior orientation parameters of georeferenced image datasets of a piecewise planar scene (*VP-based EO Estimation*), has been applied in *UAV Dataset 1* and *UAV Dataset 2*, which have been georeferenced through the method proposed in section 5.4, yielding the results presented in section 5.6.4.2; additionally, they are accompanied with the knowledge of the nadir point position for each image, obtained through the method presented in section 6.3, yielding the results outlined in section 6.6.3.1. In the following, the differences ($\Delta X_0$, $\Delta Y_0$, $\Delta Z_0$, $\Delta\omega$, $\Delta\varphi$, $\Delta\kappa$) between the exterior orientation parameters computed through one of the developed solutions and the reference exterior orientation parameters are presented in Table 6-4 for *UAV Dataset 1* and Table 6-5 for *UAV Dataset 2*. Specifically, the

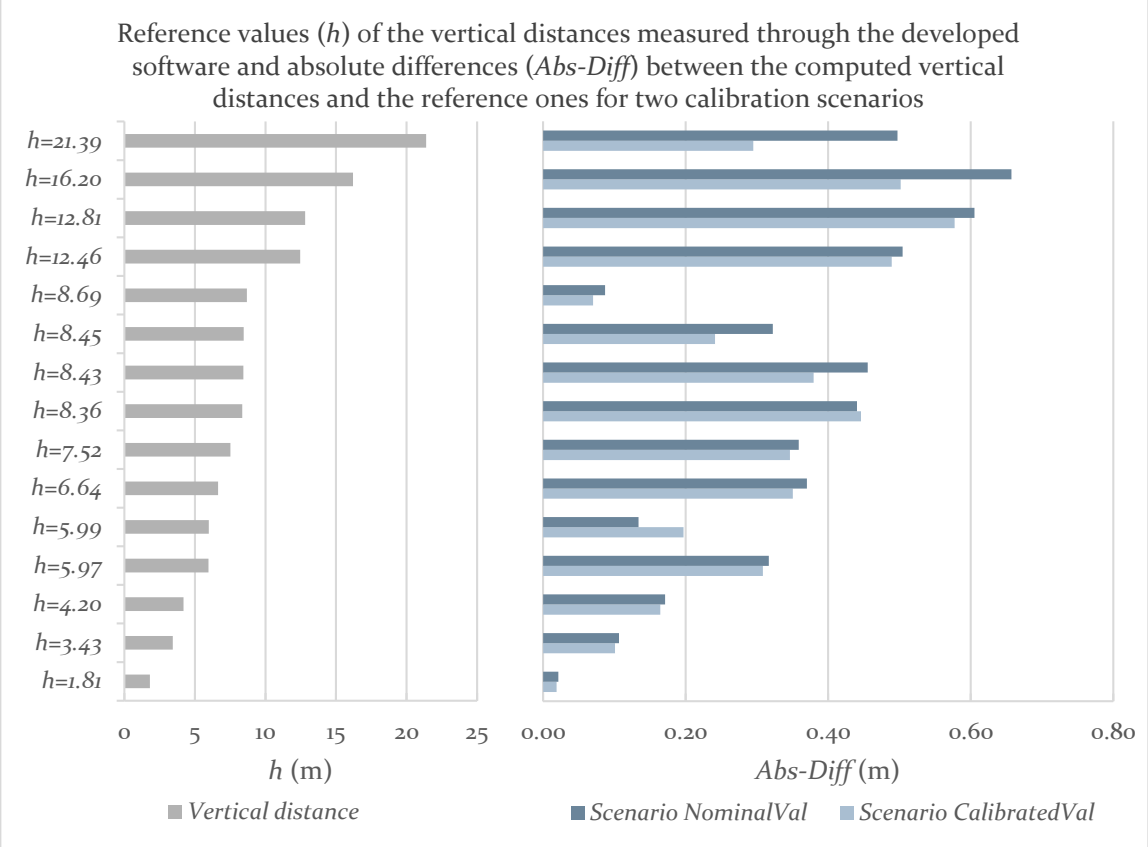


Figure 6-10. Reference values (*h*) of the vertical distances measured through the developed software and absolute differences (*Abs-Diff*) between the computed vertical distances and the reference ones for two calibration scenarios

average absolute differences (Avg-Abs), the maximum (Max-Abs) and minimum (Min-Abs) absolute differences, the average differences (Avg) and the standard deviation of differences (Stdev) between the exterior orientation parameters computed through one of the developed solutions and the reference ones, along with Avg-Abs, Max-Abs, Min-Abs, Avg and Stdev for the linear (Δ*LinearEO*) and angular (Δ*AngularEO*) exterior orientation parameters, (computed through the set of equations (5.26), as described in section 5.6.4.3) are outlined.

The diagrams in Figure 6-12 and Figure 6-13 show the distribution of absolute Δ*LinearEO* and Δ*AngularEO*, respectively, for the images of *UAV Dataset 2*, revealing a greater absolute value of Δ*LinearEO* and Δ*AngularEO* for images that are far away from the image with GCP measurements (*img_1*). The values of absolute Δ*LinearEO* and Δ*AngularEO* generally increase with the number of homography transformations between an image and the ground reference system, similarly to the case of the approximate exterior orientation parameters computed through *CP-based EO Estimation* (section 5.6.4.3). However, the comparison of the approximte exterior orientation results of *CP-based EO Estimation* and *VP-based EO Estimation* proves that the first method yields better results in terms of accuracy; this is due to the fact that the errors arising from the coplanar point detection technique are included in the *VP-based EO Estimation* method, which applies the former for the scope of georeferencing. The suitability of the derived approximate exterior orientation results through the *VP-based EO Estimation* method for being used as initial values in a bundle adjustment procedure is investigated in section 7.10.4 of this dissertation.

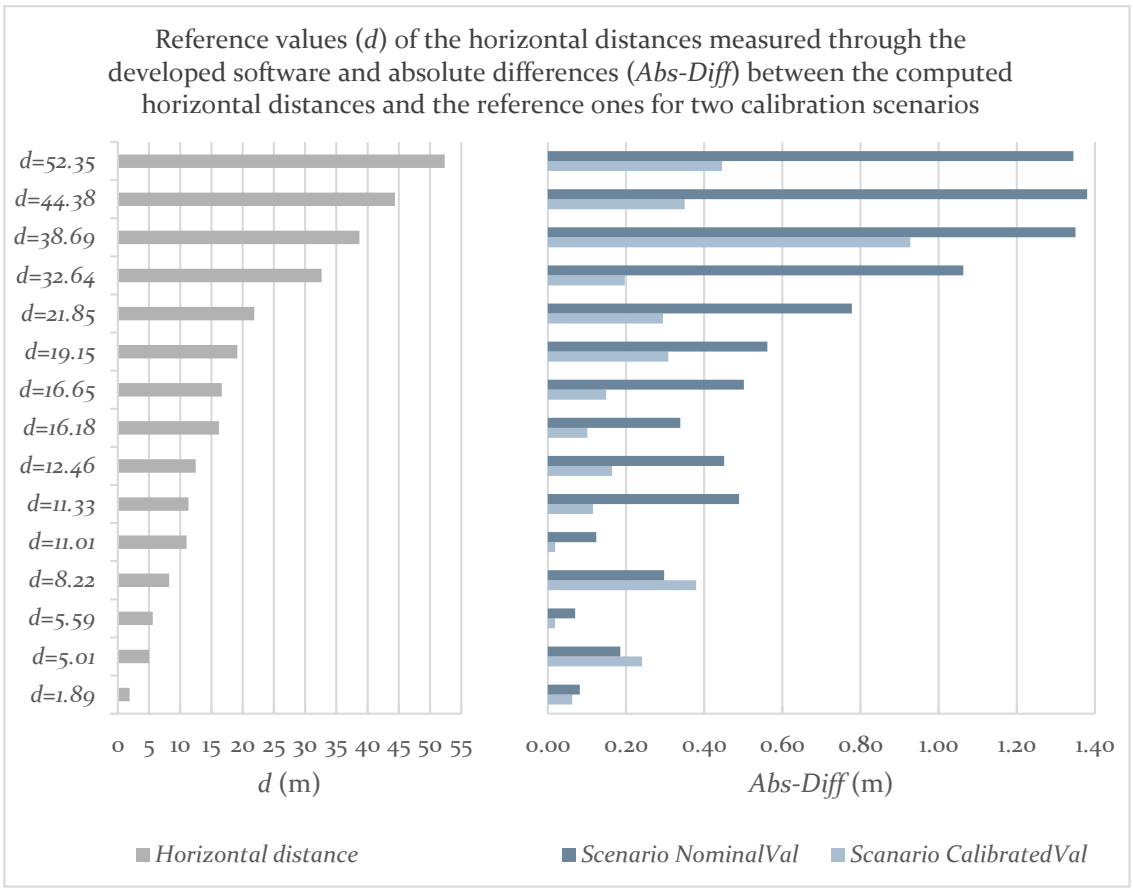


Figure 6-11. Reference values (*d*) of the horizontal distances measured through the developed software and absolute differences (*Abs-Diff*) between the computed horizontal distances and the reference ones for two calibration scenarios

Table 6-4. Differences between the approximate exterior orientation parameters computed through the developed vanishing point detection method and the reference ones for *UAV Dataset 1*

| ***UAV Dataset 1*: 23 oblique UAV images** | | | | | | |
|---|---|---|---|---|---|---|
| **Method** | **Metric** | **Avg-Abs** | **Max-Abs** | **Min-Abs** | **Avg** | **Stdev** |
| ***VP-based EO Estimation* (SURF; 4 GCPs in one image)** | $\Delta X_o$ (m) | 5.549 | 13.289 | 0.101 | 5.490 | 4.331 |
| | $\Delta Y_o$ (m) | 0.549 | 1.424 | 0.002 | 0.354 | 0.526 |
| | $\Delta Z_o$ (m) | 0.848 | 2.176 | 0.010 | -0.814 | 0.577 |
| | $\Delta\omega$ (deg.) | 1.260 | 4.348 | 0.173 | 0.544 | 1.538 |
| | $\Delta\varphi$ (deg.) | 4.262 | 9.429 | 0.242 | 4.079 | 3.202 |
| | $\Delta\kappa$ (deg.) | 5.683 | 12.623 | 0.128 | 5.582 | 4.632 |
| | $\Delta$*LinearEO* (m) | 2.315 | 13.289 | 0.002 | 1.677 | 1.812 |
| | $\Delta$*AngularEO* (deg.) | 3.735 | 12.623 | 0.128 | 3.402 | 3.124 |

Table 6-5. Differences between the approximate exterior orientation parameters computed through the developed vanishing point detection method and the reference ones for *UAV Dataset 2*

| ***UAV Dataset 2*: 35 oblique UAV images** | | | | | | |
|---|---|---|---|---|---|---|
| **Method** | **Metric** | **Avg-Abs** | **Max-Abs** | **Min-Abs** | **Avg** | **Stdev** |
| ***VP-based EO Estimation* (SURF; 4 GCPs in one image)** | $\Delta X_o$ (m) | 5.590 | 13.289 | 0.101 | 1.664 | 6.454 |
| | $\Delta Y_o$ (m) | 0.546 | 1.696 | 0.002 | 0.226 | 0.622 |
| | $\Delta Z_o$ (m) | 0.740 | 2.201 | 0.002 | -0.632 | 0.683 |
| | $\Delta\omega$ (deg.) | 1.226 | 4.839 | 0.173 | 0.502 | 1.566 |
| | $\Delta\varphi$ (deg.) | 4.448 | 9.429 | 0.242 | 1.034 | 4.983 |
| | $\Delta\kappa$ (deg.) | 5.375 | 12.623 | 0.128 | 2.028 | 6.230 |
| | $\Delta$*LinearEO* (m) | 2.292 | 13.289 | 0.002 | 0.419 | 2.586 |
| | $\Delta$*AngularEO* (deg.) | 3.683 | 12.623 | 0.128 | 1.188 | 4.260 |

## 6.7 Conclusions

In this chapter, an algorithm for detecting horizontal vanishing points and the vertical vanishing point of single oblique aerial images has been presented and two applications that implement the developed algorithm have been discussed, i.e., the extraction of horizontal and vertical distances from a single oblique aerial image and the estimation of approximate exterior orientation parameters of image datasets with known 2D transformation parameters to the ground reference system. The experiments, conducted using two datasets of UAV oblique images, reveal that the developed vanishing point detection method yields successful results in terms of automatically detecting horizontal and vertical lines. Also, the nadir point that is automatically estimated through the developed framework can be used for obtaining vertical and horizontal distances from a single oblique image, in the case that high accuracy is not the goal; in such case, nominal values for the camera interior orientation parameters (focal length and principal point at the image center) can also be used or, alternatively, the camera constant can be roughly estimated by the method discussed in this chapter, assuming a principal point at the center of the image. However, in the case of an application that requires highly accurate measurements of horizontal and vertical distances, the presented vanishing point scheme is not suitable, as it is responsible for the main source of inaccuracies in the final results. A future study on the improvement of the accuracy of the automatically extracted nadir point by elaborating the developed methodology combined with detailed investigation of the different parameters and thresholds that could be changed with the purpose of increasing the accuracy of the extracted nadir point would be interesting. Finally, as

expected, the usage of calibrated interior orientation parameters improves the accuracy of measurements of both horizontal and vertical distances.

Regarding the approximate exterior orientation parameters computed through *VP-based EO Estimation*, these correspond to larger error compared to the ones estimated through *CP-based EO Estimation*, due to the fact that the errors arising from the developed coplanar point detection technique are included in the *VP-based EO Estimation* method, which applies the former for the purpose of georeferencing. Similarly to the case of *CP-based EO Estimation*, the error of the estimated approximate exterior orientation parameters through *VP-based EO Estimation* increases with an increase of the number of transformations involved between each image and the reference coordinate system, during the georeferencing procedure. As a result, the *VP-based EO Estimation* method cannot be implemented in large image datasets, for which images far away from that of the GCPs are georeferenced using more transformations (e.g., above 20); in that case, the corresponding error in the exterior orientation parameters would be large and thus prevent these values from being used as initial exterior orientation parameters in a bundle adjustment process. As already discussed in chapter 5, in such large datasets the *VP-based EO Estimation* scheme would yield satisfying

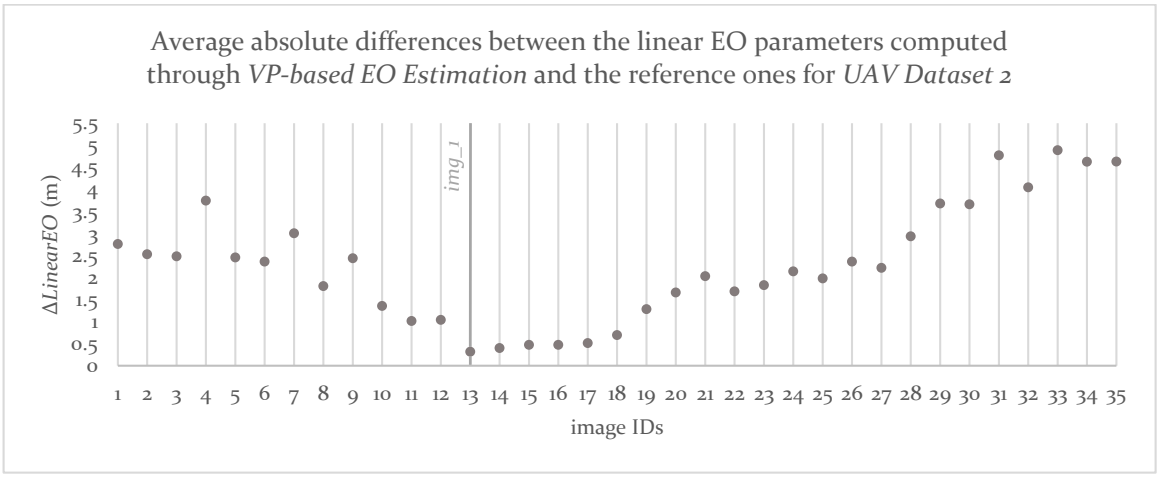


Figure 6-12. Average absolute differences between the linear EO parameters computed through V*P-based EO Estimation* (using the SURF feature extractor and four GCP measurements in *img_1*) and the reference ones for the images *of UAV Dataset 2*

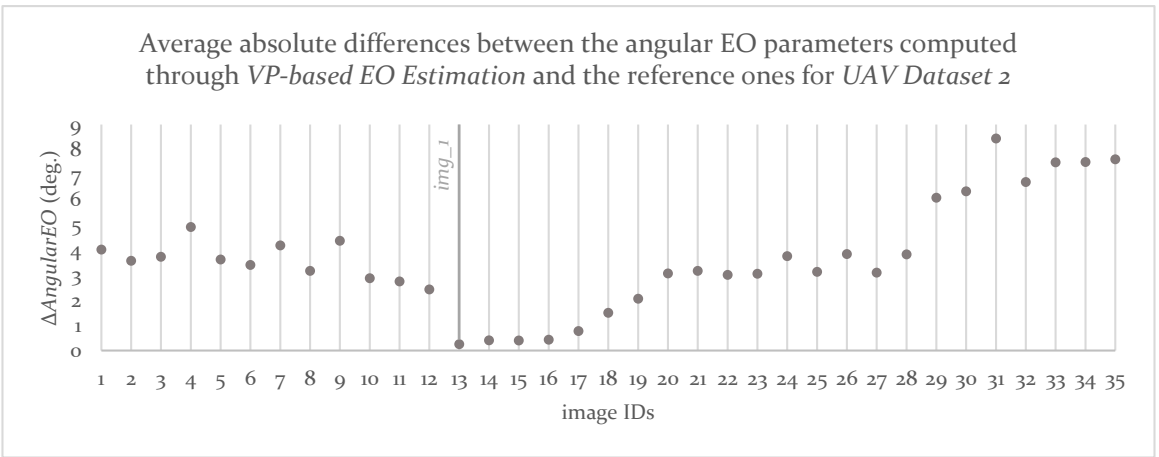


Figure 6-13. Average absolute differences between the angular EO parameters computed through *VP-based EO Estimation* (using the SURF feature extractor and four GCP measurements in *img_1*) and the reference ones for the images *of UAV Dataset 2*

results if GCPs were measured in an additional image, thus keeping the maximum total number of transformations of an image below a defined threshold (e.g., 20 transformations).

# Chapter 7
# Robust Structure from Motion

## 7.1 Introduction

Advances in photogrammetry and computer vision have led to the development of the structure from motion (SfM) approach, which has seen tremendous evolution over the years. SfM refers to the process of estimating the camera poses that correspond to a 2D image sequence and reconstructing a sparse 3D scene geometry. The combination of SfM and multi-view stereo (MVS) methods offers an automated workflow for the generation of high-accuracy dense 3D point clouds.

In this chapter, photogrammetry-based incremental and global SfM workflows that adopt robust methods for outlier removal in challenging datasets of oblique and vertical images are presented. The developed SfM techniques require the measurement of four coplanar GCPs in only one image. One of the challenges that they are intended to address is matching correctly feature points among multiple views; the developed algorithms solve this by eliminating all erroneous tie points through the combination of multiple checks and geometric constraints imposed during the image matching procedure and robust iterative bundle adjustment methods. Finally, a scale-based weighting strategy for feature point observations in bundle adjustment is introduced, targeted to highly overlapping oblique imagery with the purpose of improving the accuracy of the exterior orientation parameters.

The global SfM workflow presented in this chapter uses different techniques for calculating initial exterior orientation parameters, i.e., (i) an incremental SfM workflow; (ii) the *CP-based EO Estimation* method presented in chapter 5; (iii) the *VP-based EO Estimation* method presented in chapter 6; and (iv) georeferencing metadata from onboard sensors. The experimental results of this chapter investigate how the automatic extraction of elements describing the geometry of the scene depicted in oblique aerial images particularly of man-made environments can contribute in (i) obtaining initial values for their exterior orientation parameters required by a bundle adjustment process; and (ii) reducing the manual work required for image-based measurements of coplanar GCPs to single image measurements. Furthermore, this chapter investigates the impact that each of

the following has on the exterior orientation results: (i) the SfM workflow (global or incremental SfM) and the method used for obtaining initial exterior orientation parameters; (ii) the removal of erroneous matches that are not detected during the image matching process; (iii) the feature detection algorithm; (iv) a weighting strategy for automatic tie point measurements during bundle adjustment of oblique imagery; and (v) the number of GCPs. With the exception of the *VP-based EO Estimation* method for rough exterior orientation estimation, which can be implemented only in oblique aerial images, and the scale-based weighting strategy for bundle adjustment of oblique imagery, which is presented in this chapter, the developed SfM methods may also be applied in vertical aerial images as well. Figure 7-1 shows the incorporation of the coplanar point detection and vanishing point detection techniques, presented in this dissertation, into the SfM algorithms proposed in this chapter.

The main practical contribution of the developed exterior orientation estimation frameworks described in this chapter is the improvement of the accuracy of the exterior orientation results thanks to the developed robust iterative bundle adjustment methods for detection and removal of wrong tie points that remain after the checks and constraints implemented in the image matching process. What is more, in the special case that four coplanar GCPs are available, the developed exterior orientation frameworks require their manual measurement in only one image. A difference of the developed algorithms compared to state-of-the-art SfM methods is the estimation of the orientation of each new image based on a photogrammetric workflow, contrary to existing SfM algorithms that use computer vision methods (e.g., usage of photogrammetric space resection instead of decomposition of the essential matrix for estimation of the camera exterior orientation for the images of a pair). The fact that the algorithms are fully automatic in combination with their requirement for a minimum number of measurements in solely a single image makes them easily adoptable by operators without expertise or even basic knowledge on photogrammetry.

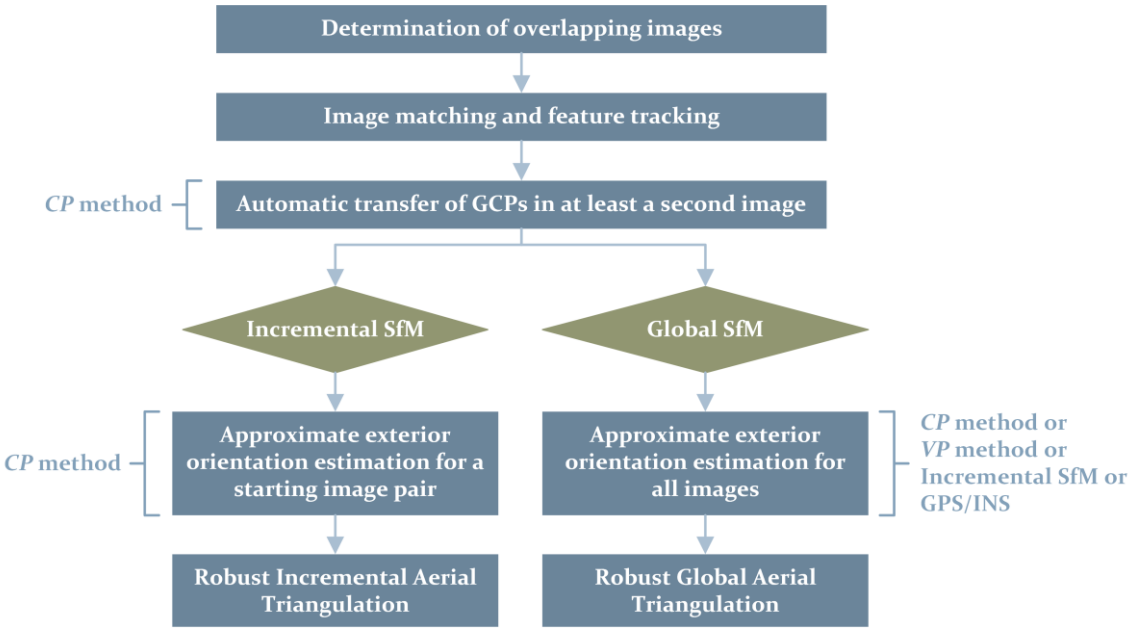


Figure 7-1. Incorporation of the coplanar point detection method (*CP*) and the vanishing point extraction method (*VP*) introduced in this dissertation into the developed incremental and global SfM workflow

## 7.2 Related Work

In this section, the related work on SfM methods is presented, including incremental, hierarchical and global techniques. Furthermore, the related work on exterior orientation estimation of oblique aerial images, which is presented in more detail in chapter 3 of this dissertation, is discussed, focusing on the differences of the SfM algorithms presented in this chapter with the state-of-the-art methods.

### 7.2.1 SfM Methods

A variety of SfM methods have been proposed so far and can be basically categorized as incremental (e.g., Snavely et al., 2006; Frahm et al., 2010; Agarwal et al., 2011; Wu, 2013; Shah et al., 2014), hierarchical (e.g., Farenzena et al., 2009; Gherardi et al., 2010; Ni and Dellaert, 2012) and global methods (e.g., Arie-Nachimson et al., 2012; Jiang et al., 2013; Moulon et al., 2013; Wilson and Snavely, 2014; Cui and Tan, 2015). Among them, incremental methods register one camera at each iteration, hierarchical methods gradually merge short sequences or partial reconstructions and global methods register all cameras simultaneously.

Incremental SfM methods are the most widely used ones. Such methods refer to the gradual incorporation of successive views in the sparse 3D reconstruction, as one image is added at a time. According to the general workflow followed by incremental SfM methods, two or three starting images are automatically selected and the corresponding camera poses are retrieved either from the fundamental matrix of the starting pair or the trifocal tensor of the starting triplet respectively, up to a projective transformation of the 3D space if precise calibration information is not available. 3D coordinates of the corresponding points in the starting images are retrieved through triangulation. Several methods may be used for the registration of successive views, such as epipolar constraints to relate each image to its predecessor, resection or merging of partial reconstructions (Robertson and Cipolla, 2009). A metric reconstruction is obtained either via an auto-calibration procedure or using known calibration data. The final stage of the SfM pipeline is usually a bundle adjustment, that is, a nonlinear optimization in order to refine the camera poses and the 3D coordinates of points. Except for the final bundle adjustment, intermediate bundle adjustment processes are also necessary in order to ensure successful camera pose estimation and sparse 3D point cloud extraction. The georeferencing of the derived SfM results is usually accomplished by estimating the 3D similarity transformation between the arbitrary SfM coordinate system and the world reference system using GCPs and/or GNSS data.

In the context of hierarchical SfM methods, the images are organized into a hierarchical cluster structure that guides the camera pose and sparse reconstruction process. A general workflow of an hierarchical SfM algorithm adopts methods for partial sparse 3D reconstructions and camera pose estimations for small sets of images, additions of images in these partial reconstructions and merging of clusters. Taking into account the fact that the partial reconstructions are in different coordinate systems, a 3D similarity transformation is usually implemented to register the cameras into a common coordinate system using the common 3D points between the partial 3D reconstructions, which are recomputed in order to refer to the same common coordinate system. Similarly to incremental SfM approaches, except for a final bundle adjustment, intermediate bundle adjustment processes are necessary. The georeferencing of the SfM results is usually accomplished as a last step by estimating the 3D similarity transformation between the SfM coordinate system and the world reference system using GCPs and/or GNSS data.

Frequent intermediate bundle adjustment procedures required by both incremental and hierarchical methods increase their computational time. Global methods do not face this challenge. Initial estimates for the camera exterior orientation parameters and the sparse 3D feature points required by the bundle block adjustment stage may be computed in the context of a global SfM framework either through a factorization-based method or through a motion averaging method. According to factorization methods, the camera pose and the scene geometry are simultaneously computed using all image measurements. Factorization methods were originally proposed by Tomasi and Kanade (1992) for orthographic cameras but have also been extended to incorporate the paraperspective camera model (Poelman and Kanade, 1997) as well as perspective cameras (Sturm and Triggs, 1996). One significant limitation of the factorization methods is that every 3D point must be visible in every image, as in case of missing data there is not any closed-form solution to the factorization. However, several methods have been proposed so far that handle this limitation, either using advanced matrix factorization techniques or low-rank approximations (e.g., Chen and Suter, 2004; Buchanan and Fitzgibbon, 2005; Ke and Kanade, 2005; Okatani and Deguchi, 2007). As far as the motion averaging methods are concerned, they generally consist of two steps, that is, rotation averaging and translation averaging. The process of rotation averaging computes a global rotation matrix for each camera based on the relative rotations between cameras, while in the context of translation averaging, a global translation for each camera is estimated based on a set of relative translations between cameras along with the computed global rotations. Alternatively, initial values for the exterior orientation parameters required by the bundle adjustment process of a global SfM method may be available from other sources, e.g., from positioning and orientation information by onboard GNSS/INS sensors. In this case, initial values for the 3D feature points may be computed using the available exterior orientation parameters through a triangulation procedure (e.g., photogrammetric intersection) and a factorization or a motion averaging method is not needed.

### 7.2.2 Exterior Orientation Estimation of Oblique Aerial Images

In recent years, much research has been conducted on the calculation of exterior orientation parameters of datasets of multi-view oblique and vertical aerial images acquired by multi-camera systems (Smith et al., 2008; Gerke and Nyaruhuma, 2009; Wiedemann and Moré, 2012; Rupnik et al., 2013; Rupnik et al., 2015; Gerke et al., 2016; Moe et al., 2016; Nakada et al., 2016; Ostrowski and Bakuła, 2016; Sun et al., 2016; Xie et al., 2016; Jiang and Jiang, 2017a; Jiang and Jiang, 2018), datasets of multi-view oblique aerial imagery acquired by multi-camera systems (Gerke, 2011; Ostrowski and Bakuła, 2016; Jiang and Jiang, 2017a; Jiang and Jiang, 2018), datasets of oblique aerial imagery acquired by a single camera (Ding et al., 2008; Ostrowski and Bakuła, 2016; Jiang and Jiang, 2017a; Mertes et al., 2017; Jiang and Jiang, 2018) and monocular oblique aerial imagery (Karjalainen et al., 2006; Xiong et al., 2014). Extended discussion of this research has been made in chapter 3 of this dissertation.

The only self-developed SfM solution targeted to oblique aerial images without making use of any existing software packages is the one proposed by Jiang and Jiang (2017a; 2018). Among the other research works that develop their own solutions regarding the exterior orientation estimation of datasets containing oblique imagery, only Rupnik et al. (2013) and Rupnik et al. (2015) presented a complete SfM framework, which is based on the combination of their own solution and existing open-source software, whereas Gerke and Nyaruhuma (2009), Gerke (2011), Sun et al. (2016) and Xie et al. (2016) proposed solely aerial triangulation methods through bundle block adjustment without focusing on the step of finding correspondences. Among the self-developed approaches,

Karjalainen et al. (2006) and Xiong et al. (2014) presented exterior orientation methods for single images, which are out of the scope of this chapter.

Specifically, Jiang and Jiang (2017a; 2018) presented a SfM solution that mainly addresses the issues of selection of image pairs and removal of redundant ones, image matching using a tiling strategy and geometric verification of matches using a method that rejects obvious outliers prior to the implementation of a RANSAC-based approach. Whereas in this chapter the issue of filtering unnecessary image pairs is not addessed, as the orientation of very large datasets of oblique aerial images is out of the scope of this dissertation, the determination of image pairs based on a graph is also implemented in the SfM algorithm developed in the context of this dissertation, using a different weighting strategy for the graph generation, which is not solely based on available georeferencing data, like the approach of Jiang and Jiang (2017a; 2018). Furthermore, whereas their tiling strategy is proved to generate an increased number of corresponding feature points of good distribution, the large number of correspondences significantly increases the image orientation computational time. Thus, taking into account that a sufficient number of feature point matches can be derived in downscaled images along with the fact that the computational time of bundle adjustment dramatically increases in cases of large numbers of tie points, such a strategy is not implemented in this dissertation. Finally, a different outlier removal strategy is implemented in this chapter, which rejects all outliers after the implementation of RANSAC, in the step of bundle adjustment.

Rupnik et al. (2013) and Rupnik et al. (2015) presented an approach which relied on a developed tool that analyses image connectivity and existing open-source SfM software for the simultaneous orientation of large datasets of oblique and vertical aerial images accompanied by georeferencing information. Their main contribution is the image connectivity graph, which contains links between images based on their overlapping area and the camera looking direction, and guides the concatenation of images during relative orientation. In this dissertation, two automatic ways for the construction of the image connectivity graph are presented (i.e., image-based method and GNSS/INS-based method), without relying only on georeferencing information, thus extending the approach by Rupnik et al. (2013) and Rupnik et al. (2015).

Gerke and Nyaruhuma (2009) and Gerke (2011) introduced the manual incorporation of geometric scene constraints into the triangulation of datasets containing oblique aerial imagery. The developed constraints, i.e., horizontal lines, vertical lines and right angles, are integrated into the bundle adjustment solution. Whereas the reduction of the number of GCPs through the use of scene constraints is an advantage of the approach proposed by Gerke and Nyaruhuma (2009) and Gerke (2011), the requirement for manual definition of constraints constitutes a major limitation. The automatic determination of vertical and horizontal line segments in oblique imagery by the methodology introduced in this dissertation (section 6.3) could assist towards the automatic definition of scene constraints. Sun et al. (2016) proposed a reduced bundle adjustment model for oblique and vertical airborne imagery acquired by multi-camera systems, which requires a lower computational complexity as well as memory cost compared to traditional bundle adjustment. According to this model, the oblique poses are parameterized using the vertical ones and the relative poses between the vertical and oblique cameras. A disadvantage of this method, due to which it is not implemented in this dissertation for the case of the images acquired by a multi-view system, is its sensitivity to the noise of relative pose parameters of the oblique cameras, due to which the reduced bundle adjustment results in bigger root mean square error, compared to the traditional bundle adjustment method.

Xie et al. (2016) presented an asymmetric re-weighting method for the combined bundle adjustment of oblique and vertical aerial imagery captured by multi-camera systems, according to which each correspondence is given different weights in the *x*-direction and *y*-direction, which depend on image scale and the shape of the projection of each pixel on the ground. The assignment of different weights to observations is a fundamental issue for oblique images, which is also dealt with in this chapter. Contrary to the approach presented by Xie et al. (2016), the weighting strategy that is presented in this chapter is based exclusively on the scale of each image observation, thus treating each image feature point as an entity with equal weights in its image coordinates (*x*, *y*). In this way, the different weights reflect the difference in scale between the image feature points. One of the major limitations of the method proposed by Xie et al. (2016) is the fact that relatively accurate initial information of the imaged scene is required, thus impeding its adoption by oblique imagery acquired by an aerial platform without onboard GNSS/IMU sensors. However, in the SfM framework introduced in this dissertation, the accurate input information required by the weighted bundle adjustment is provided by a conventional unweighted bundle adjustment process which does not require information from onboard sensors, as initial exterior orientation parameters are also proposed to be computed via different methods. Finally, whereas Xie et al. (2016) tested their strategy using oblique and vertical imagery acquired by multi-camera systems, in this chapter the suitability of the scale-based weighting strategy is assessed in different configurations of oblique imagery, both solely oblique ones acquired by a single camera and by a multi-camera system as well as oblique and vertical ones obtained by a multi-camera system, corresponding to different overlap scenarios, also including UAV imagery.

## 7.3 GNSS/INS-Based Determination of Overlapping Images

If georeferencing and orientation metadata from GNSS/INS sensors along with focal length and pixel size are available (e.g., encoded in the image headers) and if an average flying height or a mean ground elevation is specified, the GNSS/INS-based procedure presented in this section may be implemented for determining overlapping images. If the images are not accompanied with such kind of metadata, image-based determination of overlapping images takes place. The image-based method for determining overlapping images has been presented in section 4.3.1 of this dissertation.

The developed workflow for GNSS/INS-based determination of overlapping imagery is illustrated in Figure 7-2. The steps described in the following are implemented for each image separately. The first one is the extraction of the longitude, latitude, altitude, focal length and pixel size from the image metadata. The conversion of latitude and longitude to easting and northing is the next step. Then, the angular camera orientation, which is usually encoded in terms of yaw ($a_y$), pitch ($a_p$) and roll ($a_r$), is expressed in terms of omega ($\omega$), phi ($\varphi$) and kappa ($\kappa$). For this reason, the rotation matrix $\boldsymbol{R}$ is firstly calculated according to equation (7.1) and the angles $\omega$, $\varphi$ and $\kappa$ are computed using the elements $r_{ij}$ of the rotation matrix, according to the set of equations (5.22) (section 5.5.2.2).

$$\boldsymbol{R} = \begin{bmatrix} \cos a_y \cos a_p & \cos a_y \sin a_p \sin a_r - \sin a_y \cos a_r & \cos a_y \sin a_p \cos a_r + \sin a_y \sin a_r \\ \sin a_y \cos a_p & \sin a_y \sin a_p \sin a_r + \cos a_y \cos a_r & \sin a_y \sin a_p \cos a_r - \cos a_y \sin a_r \\ -\sin a_p & \cos a_p \sin a_r & \cos a_p \cos a_r \end{bmatrix} \quad (7.1)$$

In a next step, the horizontal ground coordinates of the four image corners are computed in the reference coordinate system through the collinearity equations, using (i) their image coordinates expressed in the photogrammetric system; (ii) the approximate camera exterior orientation

parameters (coordinates of the projection center from GNSS data along with the angles $\omega$, $\varphi$ and $\kappa$); (iii) the camera focal length; and (iv) the mean ground elevation. Finally, the area of the image footprint is calculated using the horizontal ground coordinates of the image corners. After the implementation of the aforementioned workflow for all images, the intersection polygon between the ground footprints of the images of each possible image pair, their overlapping area and their percentage of overlap are estimated. The percentage of overlap for all image pairs is the output of the whole process.

Finally, an undirected weighted graph that connects the overlapping images is created. Its nodes correspond to the images; an edge connects two nodes that correspond to overlapping images. Two images with a minimum percentage of overlap (e.g., bigger than 25%) are assumed to be overlapping. Each edge of the graph is given a weight, which is the percentage of overlap between the images that it connects.

## 7.4 Image Matching

The image matching procedure presented in section 4.3.2 of this dissertation is applied for the detection of corresponding feature points between image pairs. Except for the use of the SURF feature detector, briefly described in section 4.3.1, other detectors may also be used for the feature extraction process. Especially for the case of multi-perspective oblique aerial images, the feature point detector to be used has to be scale and rotation invariant and ideally cover partially or fully the parameters of an affine transform. In the context of this dissertation, the Affine-SIFT (ASIFT) detector, proposed by Morel and Yu (2009), is also tested in the experiments; thus, it is briefly outlined in the following. ASIFT is a fully affine invariant feature extraction algorithm. It simulates several sample affine views of the original images, obtained by varying the two camera axis orientation parameters, namely, the latitude and the longitude angles, and then applies the SIFT algorithm to the simulated images. SIFT searches over all scales and image locations by using a

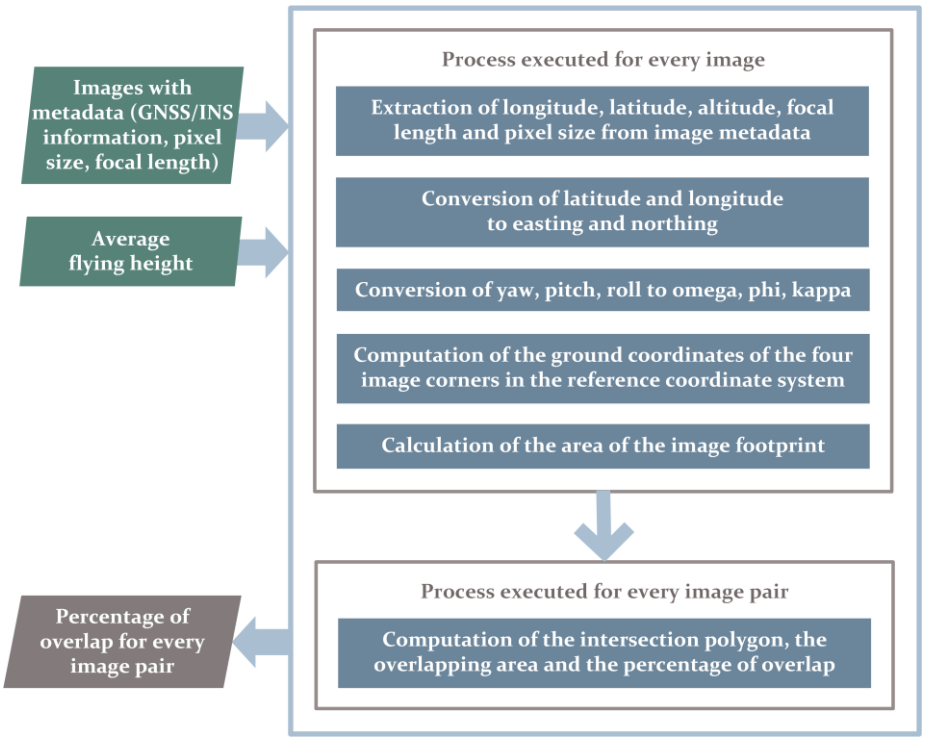


Figure 7-2. The developed workflow for GNSS/INS-based determination of overlapping images

difference-of-Gaussian function to identify potential feature points that are invariant to scale and orientation. At each candidate location, keypoint localization takes place, by determining their location and scale, and the final keypoints are selected based on measures of their stability. One or more orientations are assigned to each keypoint location based on local image gradient directions. The local image gradients are measured at the selected scale in a region around each keypoint and an 128-dimensional descriptor vector for each keypoint is generated, which allows for significant levels of local shape distortion and change in illumination. The SIFT algorithm is fully invariant to four out of the six parameters of an affine transform, while the ASIFT algorithm covers all its six parameters.

The corresponding feature points for each image pair, detected by the method presented in section 4.3.2, are stored in the same position of two vectors, for efficient implementation of the feature tracking process. Also, they are sorted by ascending order of the *x* pixel coordinates of the feature points of the first image of each pair, for faster search for points during the feature tracking process. For efficient implementation of the developed SfM algorithm, IDs corresponding to numbers ranging from 1 (that corresponds to the image where the GCPs are measured) to the total number of images are assigned to each image.

## 7.5 Feature Tracking

The correspondences are organized into tracks, each containing the coordinates of the feature points in different images that correspond to the same 3D point. Feature tracking takes place for each pair of overlapping images after the image matching procedure for this pair. The developed methodology for the implementation of feature tracking is based on the proper formation of one vector and two matrices, as explained in the following; an index of zero is assumed to correspond to the first row and column of these data structures.

- A vector (*processedImages*) that stores information about whether feature points have been stored for each image. Its size equals to the number of images. Its elements may take the values of 0 and 1; an element at position *r* that equals to 0 indicates that feature points have not been stored for the image with ID *r*+1; a value of 1 indicates that feature points have been stored for the corresponding image.
- A matrix (*imgPoints*) that stores the image coordinates of feature points, with the number of rows being equal to the number of images and the number of columns being equal to the number of tracks, that is, the number of 3D points. Each row stores information for one image and each column stores information for each track. For instance, the element at row *r* and column *c* corresponds to the pixel coordinates of the feature point with ID *c*+1 at image with ID *r*+1. If a point is not visible in an image, the coordinates of (-1, -1) are stored in the corresponding element of this matrix.
- A matrix (*visibility*) that stores information about whether a point is visible in each image. Its dimensions are the same with the ones of the *imgPoints* matrix. Its elements may take the values of 0 and 1, with 0 indicating non-visibility and 1 indicating visibility. For instance, an element at row *r* and column *c* that equals to 0 indicates that the feature point with ID *c*+1 is not visible in the image with ID *r*+1.

The *processedImages* vector and the *visibility* matrix are initialized with the value of 0, whereas the *imgPoints* matrix is initialized with the values of (-1, -1); their initialization takes place before the image matching process. Feature tracking is applied for each image pair by filling the aforementioned vector and matrices. The algorithm is presented in the form of pseudocode in Figure

7-3; *rowImg_k* and *colImg_k* represent the row and column number, respectively, of the vector or matrix, that corresponds to an image *k; maxColNum* represents the maximum column number of *imgPoints* that contains a feature point; *pointImg_kp* represents the feature point with ID *p* in image *k*, respectively.

Finally, among the tracks detected through the method presented in this section, only tracks that include feature point measurements in at least three images are used by the SfM processes described in the following sections, while the rest is discarded. In this way, more stable tracks are used in the bundle block adjustment procedures.

## 7.6 Weighting Strategy for Bundle Adjustment of Oblique Images

The scale variations within a nadir image depend only on the anaglyph of the imaged scene and they may be considered insignificant, especially in comparison with the scale variations inside the corresponding oblique image taken from the same point. Thus, the bundle adjustment procedures designed for nadir images usually assume that all point observations are of the same quality, by applying identical weights to all points. However, the difference of scale within an oblique aerial image may be significant, depending on its tilt angle. Hence, in order to investigate the results that can be achieved if different weights are assigned to each image point during the bundle adjustment

**Algorithm**: Feature tracking

1: **if** (*processedImages*(*rowImg_i*) = 0 **and** *processedImages*(*rowImg_j* ) = 0) **then**
2: **for each** pair *p* of corresponding feature points **do**
3: find *maxColNum*
4: **for each** *img k* ∈[ *i, j* ] **do**
5: *colImg_k* ← *maxColNum* + 1
6: *imgPoints(rowImg_k, colImg_k)* ← *pointImg_kp*
7: *visibility*(*rowImg_k, colImg_k*) ← 1
8: **for each** *img k* ∈[ *i, j* ] **do**
9: *processedImages*(*rowImg_k*) ← 1
10: **else if** ((*processedImages*(*rowImg_i*) = 1 **and** *processedImages*(*rowImg_j* ) = 0) **or** (*processedImages*(*rowImg_i*) = 0 **and** *processedImages*(*rowImg_j* ) = 1)) **then**
11: **for each** pair *p* of corresponding feature points **do**
12: *m* ← index for the processed *img*
13: *k* ← index for the non-processed *img*
14: **if** *pointImg_mp* has been saved **then**
15: find *colImg_m* where *pointImg_mp* has been saved
16: *colImg_k* ← *colImg_m*
17: **do** steps 6-7
18: **else**
19: **do** steps 3-7
20: *processedImages*(*rowImg_k* ) ← 1
21: **else**
22: **for each** pair *p* of corresponding feature points **do**
23: **if** *pointImg_ip* **or** *pointImg_jp* has been saved **then**
24: **if** its corresponding point has not been saved **then**
25: *m* ← index for *img* with the already saved point
26: *k* ← index for the other *img*
27: **do** steps 15-17
28: **else**
29: **do** steps 3-7

Figure 7-3. The developed feature tracking algorithm applied for an image pair

process, a weighting strategy is presented, designed for the aerial triangulation of datasets containing oblique aerial images. The main idea is the assignment of greater weights to image points that correspond to larger scale. Thus, the weight of a target point observation ($p_T$) is a function of its scale ($s_T$) and is given by equation (7.2).

$$p_T = c_{const} \cdot s_T \tag{7.2}$$

In equation (7.2), $c_{const}$, is a constant that can be multiplied with the scale of each point in order to normalize the weights. The value of $c_{const}$ is the same for every point and does not influence the results of the bundle adjustment process. Xie et al. (2016) estimated this constant as the inverse scale of a point of the equivalent nadir image taken by the mean flying height with the same camera. In this way, an average ground elevation has to be calculated using the approximate $Z$ coordinates of all tie points. In order to avoid this calculation and constrain the weights to have values close to unity, one random point observation may be given a unit weight by calculating $c_{const}$ as the inverse scale of a random point of the oblique image.

The scale of a point of an oblique aerial image (i.e., the $x$-scale that corresponds to the line that passes through the point and is perpendicular to the principal line) is calculated geometrically according to equation (7.3) (see also equation (2.5), section 2.4.4.1).

$$s_T = \frac{OT'}{OT} \tag{7.3}$$

In equation (7.3), $OT$ is the distance between the target $T$ and the projective center and $OT'$ is the distance between the projective center and the image of the target $T'$ (Figure 7-4). These distances are geometrically calculated using the Pythagorean Theorem. Thus, the scale of a point is finally estimated by equation (7.4), where $P'$ is the principal point, $N_1$ is the projection of the nadir point at the horizontal plane of the target point $T$, $H_T$ is the flying height that corresponds to the target

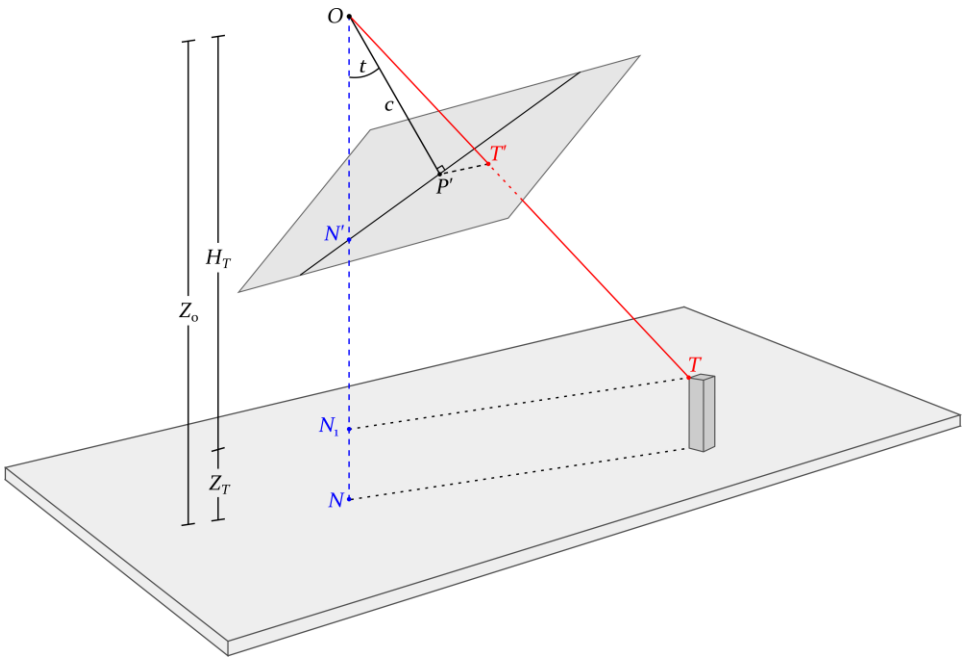


Figure 7-4. Geometry of an oblique aerial image illustrating the ray to a target point $T$ for calculation of its scale

point $T$, $Z_T$ is its elevation, $x_{T'}$, $y_{T'}$ are its image coordinates and $X_T$, $Y_T$, $Z_T$ are its coordinates in the ground reference system.

$$s_T = \frac{\sqrt{c^2 + \left(P'T'\right)^2}}{\sqrt{H_T{}^2 + \left(N_1 T\right)^2}} \quad \rightarrow \quad s_T = \frac{\sqrt{c^2 + \left(x_{T'} - x_0\right)^2 + \left(y_{T'} - y_0\right)^2}}{\sqrt{\left(Z_0 - Z_T\right)^2 + \left(X_0 - X_T\right)^2 + \left(Y_0 - Y_T\right)^2}} \tag{7.4}$$

Substituting equation (7.4) in equation (7.2) and calculating $c_{const}$ as the inverse scale of a random point $R$ ($X_R$, $Y_R$, $Z_R$), the weight of a target point observation is given by equation (7.5).

$$p_T = \frac{\sqrt{\left(Z_0 - Z_R\right)^2 + \left(X_0 - X_R\right)^2 + \left(Y_0 - Y_R\right)^2}}{\sqrt{c^2 + \left(x_{R'} - x_0\right)^2 + \left(y_{R'} - y_0\right)^2}} \cdot \frac{\sqrt{c^2 + \left(x_{T'} - x_0\right)^2 + \left(y_{T'} - y_0\right)^2}}{\sqrt{\left(Z_0 - Z_T\right)^2 + \left(X_0 - X_T\right)^2 + \left(Y_0 - Y_T\right)^2}} \tag{7.5}$$

The rejection of all outiers before a weighted bundle adjustment process is particularly significant, as any remaining erroneous tie points with their resulting wrong coordinates in the ground reference system involve the risk of being assigned large (wrongly calculated) weights, that will negatively affect the bundle adjustment results. Whereas this may be avoided by constraining weights to be between predefined upper and lower thresholds, this is not the most appropriate solution. Thus, the elimination of all outliers is particularly important before the implementation of a weighted bundle adjustment. In the following, a robust iterative framework for rejecting any remaining outliers of the SfM process is proposed.

## 7.7 Automatic Transfer of GCPs

The method described in this section for automatic transfer of coplanar GCPs among overlapping images significantly reduces the required manual work. It may be applied if at least four coplanar GCPs have been measured in a single image; it automatically determines their position in other overlapping images. The method may be implemented either if coplanar GCPs have been measured only in one image or if sets of coplanar GCPs (lying in different planes) have been measured in the same or different images. In each case, the developed method has to be applied for each set of coplanar GCPs. It consists of three main steps, that is, high-resolution image matching, template matching and transfer of GCPs.

The images to which the GCPs are going to be transferred (*img_k*) are determined as the ones with a large number of correspondences with the image that includes the GCP measurements (*img_1*). In order to ensure that the GCPs will have a similar appearance in all overlapping images to which the algorithm will be applied, so that the template matching method will derive successful results, a minimum number of correspondences extracted by the image matching process (section 7.4) is defined for selecting the images on which the GCPs will be automatically recognized. This number is algorithm-specific and data-specific, depending on the parameterization used by the feature extractor.

In the following, the steps applied to a pair of images (*img_1*, *img_k*) is described. Initially, the image matching procedure, discussed in section 7.4, is applied in these images, followed by the coplanar point detection process described in section 5.3 for computation of coplanar feature points that lie on the plane of GCPs as well as the homography transformation induced by the plane of GCPs from *img_k* to *img_1*. Then, the GCPs are transferred from *img_1* to *img_k* through the inverse homography matrix, in order to automatically obtain GCP measurements in *img_k*.

## 7.8 Robust Incremental SfM

Whereas, the elimination of all outliers is a prerequisite before the implementation of a weighted bundle adjustment process, it is also of high importance for the implementation of a bundle adjustment process that assumes equal weights for point observations, as it is shown in the experimental section of this chapter (section 7.10). In this section, the workflow proposed for the implementation of an incremental SfM method that incorporates robust iterative bundle adjustment procedures for the elimination of all outliers that have passed through the tests and constraints imposed by the image matching procedure, hereinafter referred to as *Robust Incremental SfM*, is presented. Although the SfM algorithm requires the input of the image coordinates of at least four coplanar GCPs in one image, it assumes that the GCPs are also visible in at least a second one, for automatic transfer of the GCPs. The other images involved in the SfM process are not required to depict the plane where the GCPs lie; they may depict a scene without any constraint on its anaglyph. The developed SfM method requires:

- a dataset of images without the need for GNSS/INS information;
- the pixel coordinates of four coplanar GCPs, measured in one image;
- the ground coordinates of the GCPs; and
- the interior orientation of the imagery along with their pixel size or at least nominal values for the focal length and pixel size of the imagery, which are almost always available in the image metadata.

Indicative parameter values used by the developed *Robust Incremental SfM* workflow are outlined in Table 7-1. In the following, the steps followed by the developed *Robust Incremental SfM* method, after the stages of determining overlapping images, image matching, feature tracking and automatic transfer of GCPs are described.

### 7.8.1 Orientation of a First Image Pair

In this section, a method for the computation of the exterior orientation parameters of a starting pair of images that depicts a piecewise planar scene using coplanar GCP measurements in only one image of the pair for initialization of an incremental SfM process is presented. Let *img_1* be the image that includes the manual GCP measurements and *img_2* be the one, among the images where the GCPs have been automatically transferred (section 7.7), with the maximum number of correspondences with *img_1*. In the case that different GCPs have been measured in more than one images, *img_1* and *img_2* are determined as the pair of images with the maximum number of correspondences among all possible pairs of *img_1* and *img_2*. Figure 7-5 illustrates the developed workflow for computation of the exterior orientation parameters of a starting image pair for SfM initialization. The steps of the developed workflow are described in the following.

The homography transformation from the plane of GCPs in *img_1* to the real world plane, in the reference coordinate system defined by the GCPs, is computed using the pixel coordinates of the GCPs and their ground coordinates in the reference system. The ($X$, $Y$) ground coordinates of the corresponding features of *img_1* and *img_2* that belong to the plane of the GCPs are calculated using their pixel coordinates ($x_{pix}$, $y_{pix}$) in *img_1* in combination with the estimated homography transformation from *img_1* to the real world plane. Their third ($Z$) coordinate is set equal to the $Z$ coordinate of the coplanar GCPs.

Table 7-1. Parameter values used by the developed *Robust Incremental SfM* algorithm

| Symbol | Value |
|---|---|
| *distThresToVerifyPoints* | 15 m |
| *minMatchesForImgi* | 5 |
| *maxBAIterations* | 3 |
| *maxdxDifferenceThres* | 3 m; if *images* ≤ 40 |
| | 4 m; if *images* ≤ 70 |
| | 5 m; if *images* ≤ 100 |
| | round(*images*/20) m; if *images* > 100 |
| *maxImagesWithoutGlobalBA* | 5 |

The exterior orientation parameters of *img_1* and *img_2* are estimated through conventional photogrammetric space resection (section 5.5.2.3) using the collinearity equations as the mathematical model. For this reason, the coordinates of the coplanar points in each image along with the ground coordinates of the coplanar points in the reference coordinate system and the interior orientation parameters of the camera that acquired the images are combined in a least-squares solution for the iterative computation of the exterior orientation parameters of each image separately. The initial exterior orientation parameters required for the least-squares solution are estimated through a Levenberg-Marquardt (Moré, 1978) optimization method that solves the PnP problem (Lepetit et al., 2009), which minimizes the reprojection error, that is, the sum of squared distances between the observed image points and the projected ones in the image using the estimated camera pose parameters.

### 7.8.2 Photogrammetric Space Intersection

The starting images of the iterative procedure of the SfM algorithm are *img_1* and *img_2*, i.e., the image where the GCPs have been measured and the image with the maximum number of correspondences with *img_1*, where the GCPs have been automatically transferred, respectively; the exterior orientation parameters of these images have already been estimated (section 7.8.1). The image that is incrementally added in each iteration of the SfM process is hereinafter symbolized as *img_j*. The already oriented images that overlap with *img_j*, which are used in combination with *img_j* in each SfM iteration for determining the 3D coordinates of their common feature points are hereinafter symbolized as *imgs_i*; each one is referred to as *img_i*. Hence, for the first iteration of SfM, *img_i* is *img_1* and *img_j* is *img_2*.

The first stage of each iteration of the SfM process is the computation of the 3D coordinates of the feature points between *img_j* and each *img_i* in the reference system. In order to obtain more stable tracks for the bundle adjustment process, only feature points that belong to a track of feature point measurements in at least three images are used, as discussed in section 7.5. A least-squares solution that solves the conventional problem of photogrammetric space intersection is adopted for every pair of corresponding points, based on the collinearity equations.

### 7.8.3 Removal of Wrong Points

If a point has already been triangulated through a different combination of overlapping images, the differences between its newly computed (*X*, *Y*, *Z*) ground coordinates and its ground coordinates as computed in a previous iteration of the SfM process are calculated. If these differences are above a

maximum accepted threshold (*distThresToVerifyPoints*), the entire track is considered to be erroneous and is removed from the initial data. If the computed differences are below this threshold, the already calculated 3D ground coordinates that correspond to the track are considered to be correct. Otherwise, if the point has not already been triangulated, its coordinates, which have been calculated in the current SfM iteration, are stored by the algorithm.

### 7.8.4 Photogrammetric Space Resection

The non-oriented image with the maximum number of already triangulated feature points is considered to be *img_j*; *imgs_i* are the oriented images that have a least number (*minMatchesForImgi*) of common feature points with *img_j*. The exterior orientation parameters of *img_j* are estimated through photogrammetric space resection as described in section 5.5.2.3, using the image coordinates of the already triangulated feature points expressed in the photogrammetric system, their 3D coordinates in the reference coordinate system and the interior orientation parameters of the camera.

### 7.8.5 Bundle Adjustment

The mathematical model of the bundle adjustment is expressed by the collinearity equations, which are not linear with respect to the unknowns, i.e., the exterior orientation parameters of the images, the 3D coordinates of the tie points in the reference coordinate system and the interior orientation parameters in case of an auto-calibration. The number of equations that can be formed to solve the bundle adjustment problem based on the collinearity conditions is $n = 2N_{tie} + 2N_{GCP}$, where $N_{tie}$ is the number of image feature points and $N_{GCP}$ is the number of image appearences of GCPs. The solution to the system of linearized equations is represented in matrix form by equation (7.6), where

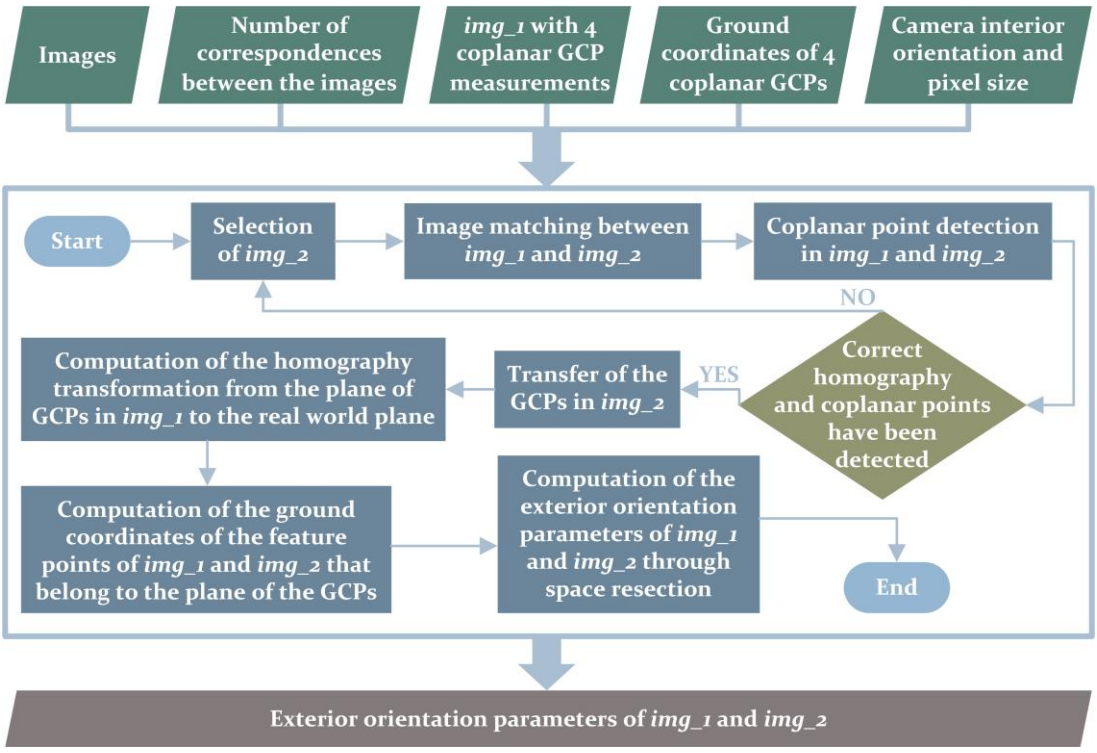


Figure 7-5. The developed workflow for computation of the exterior orientation parameters of a starting pair of images for SfM initialization using coplanar GCP measurements in a single image

***dx*** is the vector with the differences between the approximate values of the unknowns and those ones computed by the least-squares solution; ***A*** is the $n \times 6$ coefficient matrix with the partial derivatives with respect to the unknowns; ***P*** is the $n \times n$ diagonal weighting matrix for the observations; and ***dl*** is the $n$-dimensional vector with the differences between the image coordinates of the tie points and GCPs in the photogrammetric system and their estimations obtained using the approximate values of the unknown parameters. For reasons of consistency of the observation equations in terms of units, the elements of the columns of ***A*** that refer to the angular exterior orientation elements are multiplied by $\pi/1800000$, as the angular differences of the vector ***dx***, ($d\omega$, $d\varphi$ and $d\kappa$) are expressed in $10^{-4}$ degrees.

$$\boldsymbol{dx} = \left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{A}\right)^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{dl} \tag{7.6}$$

#### 7.8.5.1 Iterative Fixed Bundle Adjustment

Some erroneous tie points may still exist; these are the points that have passed through the checks applied in the image matching procedure (section 4.3.2) and have not already been triangulated by a different combination of overlapping images, because in that case they would have been rejected, as described in section 7.8.3. An iterative process implementing a least-squares bundle adjustment method is applied in order to remove possible erroneous tie points and ensure that there are not any outliers in the solution. This stage is implemented in every iteration of the SfM procedure after the space resection step, with the exception of the first SfM iteration; in this case, the process described in section 7.8.5.2 is adopted. Specifically, this iterative procedure starts with a *fixed* bundle adjustment solution, which refines the exterior orientation of *img_j* and the 3D ground coordinates of the feature points that have been triangulated until the current SfM iteration; the exterior orientation parameters of all the other oriented images, as well as the interior orientation parameters are considered to be fixed. The already calculated ground coordinates of the feature points either refined through a previous bundle adjustment or estimated only through space intersection (which applies to the newly triangulated feature points), as described in section 7.8.2, serve as initial values for the 3D ground coordinates of the points required for the bundle adjustment. The exterior orientation of *img_j* estimated through space resection (section 7.8.4), is used as initial exterior orientation for this image in the bundle adjustment. A maximum number of iterations (*maxBAIterations*) for this bundle adjustment method is defined and the maximum value of the differences between the coordinates of the feature points as estimated in the ultimate and penultimate iteration is detected. These differences are computed in the vector ***dx***, which contains the differences between the calculated values of the unknown elements as estimated in two consecutive iterations of the least-squares solution. If the maximum value of these differences is above a maximum accepted adaptive threshold (*maxdxDifferenceThres*), the value of which increases with the number of oriented images, the corresponding point is removed from the set of points and the bundle adjustment with the predefined number of iterations is repeated, until all outliers have been detected. At the end of this iterative procedure, the same bundle adjustment method is applied without constraint on the maximum number of iterations, using the correct feature points, for refining their 3D ground coordinates and the exterior orientation parameters of *img_j*. The iterations of this last bundle adjustment step are continued until all the values of the calculated vector ***dx*** are below predefined thresholds, depending on the accuracy requirements of each application.

### 7.8.5.2 Iterative Global Bundle Adjustment

The iterative *global* bundle adjustment is implemented (i) after the space resection step, described in section 7.8.4, in the first iteration of the SfM method; (ii) after the iterative *fixed* bundle adjustment procedure, described in section 7.8.5.1, if a maximum number of images (*maxImagesWithoutGlobalBA*) has been added in the incremental SfM procedure without a *global* bundle adjustment step; and (iii) after the iterative *fixed* bundle adjustment procedure (section 7.8.5.1), if *img_j* is the last image of the dataset being oriented. Similarly to the process described

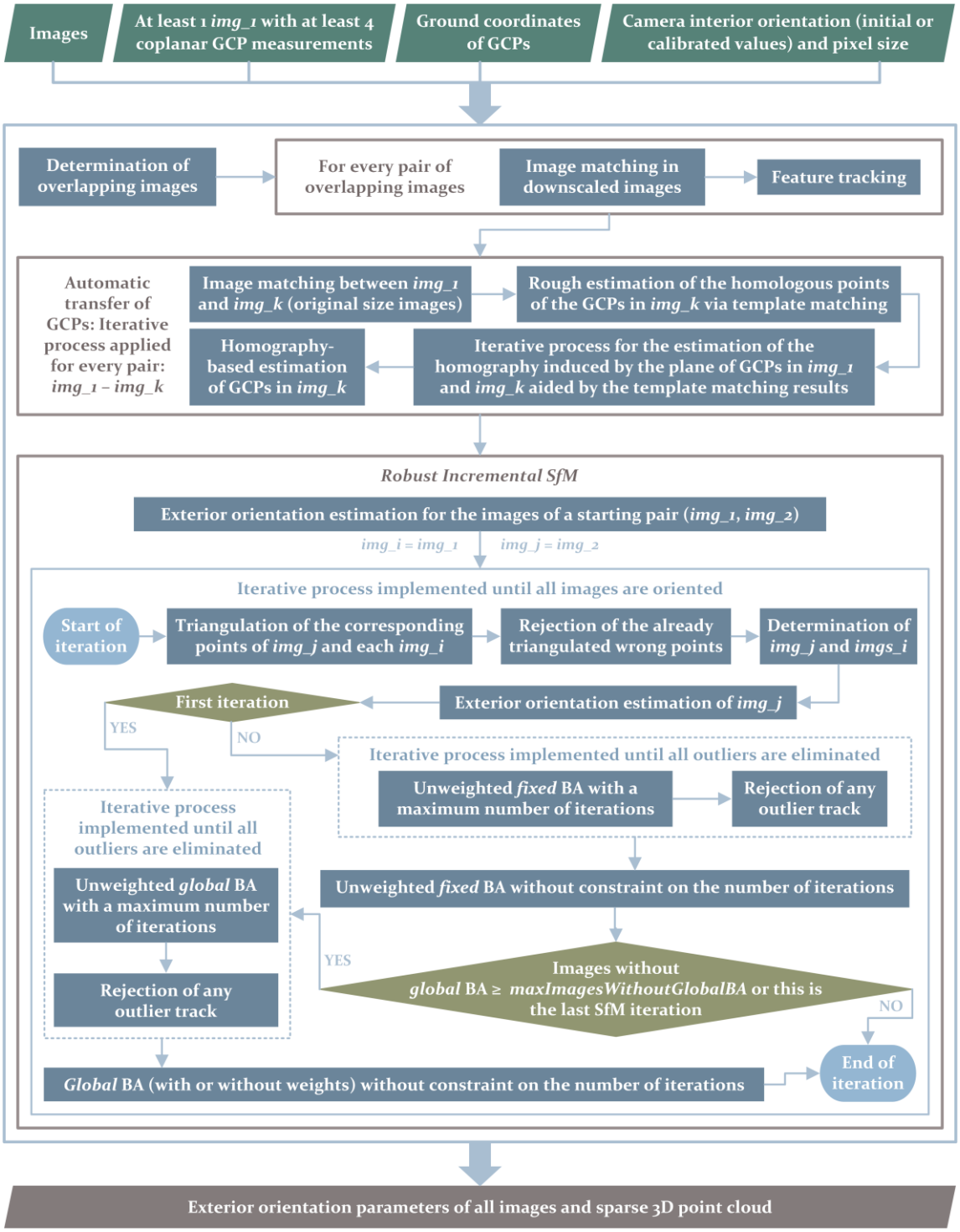


Figure 7-6. The developed workflow for *Robust Incremental SfM*

in section 7.8.5.1, an iterative procedure that implements a least-squares bundle adjustment is applied. However, contrary to the iterative *fixed* bundle adjustment solution, the iterative *global* bundle adjustment in this stage refines the exterior orientation parameters of all oriented images, in addition to the 3D coordinates of the feature points that have been triangulated until the current iteration of the SfM process and, optionally, the camera interior orientation parameters. Any possible outliers are detected and removed via estimating the differences of the 3D coordinates of the feature points as calculated in the penultimate and the ultimate iteration of the bundle adjustment, as described in section 7.8.5.1. At the end of this iterative procedure, the same *global* bundle adjustment method is applied without any constraint regarding the number of iterations, for refining the 3D coordinates of the remaining correct feature points, the exterior orientation of all already oriented images and optionally the camera interior orientation.

Figure 7-6 shows the workflow for the implementation of the developed *Robust Incremental SfM* method, including the steps of determining overlapping images, image matching, feature tracking and transfer of GCPs.

## 7.9 Robust Global SfM

In this section, a photogrammetry-based framework that solves the global SfM problem, given initial exterior orientation estimations and camera calibration information along with GCP measurements is presented. The developed workflow incorporates robust iterative bundle adjustment procedures for the elimination of all outliers that have passed through the tests and constraints imposed by the image matching procedure. It is hereinafter referred to as *Robust Global SfM*. Initial exterior orientation parameters, required by the iterative least-squares solution of the bundle block adjustment, may be calculated by different methods, e.g., (i) by an incremental SfM workflow without iterative procedures for outlier removal (*Incremental SfM*); (ii) by the *CP-based EO Estimation* method, presented in section 5.5, followed by a global SfM method without iterative procedures for outlier removal (*Global SfM*); (iii) by the *VP-based EO Estimation* method, presented in section 6.5, followed by *Global SfM*; and (iv) by georeferencing metadata from onboard sensors. Specifically, the developed SfM method requires:

- a dataset of images;
- the pixel coordinates of four coplanar GCPs, measured in one image;
- the ground coordinates of the GCPs;
- the interior orientation of the images along with their pixel size, or at least the focal length and pixel size of the imagery which are almost always available in the image metadata; and
- approximations for the camera exterior orientation parameters of the images

In this section, the steps followed by the developed *Robust Global SfM* method, after the stages of determining overlapping images, image matching, feature tracking and automatic transfer of GCPs are described.

### 7.9.1 Photogrammetric Space Intersection

Approximate values for the 3D coordinates of the corresponding feature points between each pair of overlapping images have to be computed. Similarly to the case of the *Robust Incremental SfM* method, in order to obtain more stable tracks for the bundle adjustment process, only tracks that include feature point measurements in at least three images are used. Specifically, a least-squares

solution that solves the conventional problem of photogrammetric space intersection is adopted for every pair of corresponding points that belong to a track of at least three feature points, based on the collinearity equations.

### 7.9.2 Removal of Erroneous Points

Every 3D feature point is visible in more than two images. Thus, the distance between their ground coordinates, which are computed through photogrammetric intersection using different combinations of overlapping images, are computed. If this distance is above a maximum accepted threshold (*distThresToVerifyPoints*, Table 7-1) for at least one combination of image pairs, the entire track is considered to be erroneous and is removed from the initial data. In this way, another check that removes any remaining outliers that have passed through the geometric constraints imposed by the image matching procedure is implemented.

### 7.9.3 Iterative Global Bundle Adjustment

The ground coordinates of the feature points estimated through space intersection (section 7.9.1), along with the approximate exterior orientation parameters of the images serve as initial values for

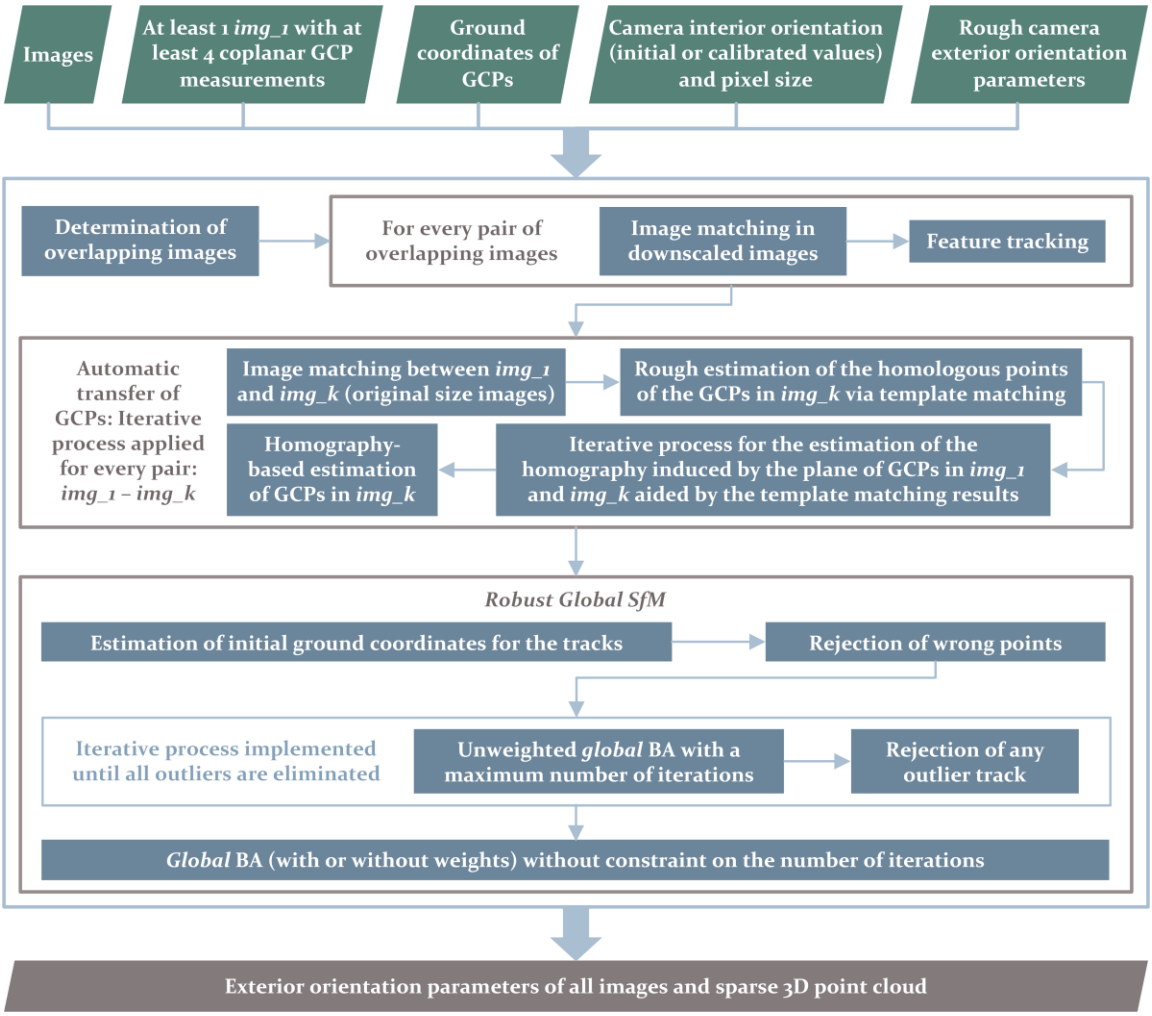


Figure 7-7. The developed workflow for *Robust Global SfM*

the bundle block adjustment. An approximation of the camera constant (e.g., knowledge of the focal length), if an auto-calibration process is implemented, is required. The iterative *global* bundle adjustment workflow presented in section 7.8.5.2 is used for the rejection of any remaining outliers that have passed through the tests and constraints imposed during the previous steps as well as the refinement of the approximate exterior orientation parameters that have been used and the computed 3D coordinates of the tracks.

Figure 7-7 shows the workflow for the implementation of the developed *Robust Global SfM* method, including the steps of determining overlapping images, image matching, feature tracking and transfer of GCPs.

## 7.10 Experimental Results

This section investigates the impact that each of the following has on the exterior orientation results: (i) the SfM workflow (global or incremental SfM) and the method used for obtaining initial exterior orientation parameters in the case that a global SfM workflow is applied; (ii) the removal of erroneous matches that are not detected during the image matching process; (iii) a scale-based weighting strategy for automatic tie point measurements during bundle adjustment; (iv) the feature detection algorithm; and (v) the number of GCPs.

### 7.10.1 Developed Software

A software suite that implements the proposed incremental and global SfM methods has been developed in the C++ programming language, making use of some functionalities offered by the OpenCV library (OpenCV Team, 2019) for image manipulations, the Eigen Library (Eigen Team, 2019) for matrix and vector operations, the GeographicLib library (Karney, 2017) for conversion between geographic (latitude, longitude) and UTM coordinates, the Boost libraries (Boost Team, 2019) for graph creation and the GDI+ library (Microsoft, 2019) for extraction of image metadata. The following developed software tools are used for the experiments:

- a software tool for determining overlapping images through the image-based method (section 4.4.2.1) or through the GNSS/INS-based method;
- a software tool for image matching and feature tracking;
- a software tool for detecting coplanar feature points in images lying on the plane of the coplanar GCPs (section 5.6.1) and transferring those coplanar GCPs among images;
- a software tool for *Incremental SfM*;
- a software tool for *Global SfM*;
- a software tool for *Robust Incremental SfM*; and
- a software tool for *Robust Global SfM*

### 7.10.2 Test Datasets and Reference Data

In order to assess the developed exterior orientation workflows, several datasets are used, as described in the following.

#### 7.10.2.1 UAV Datasets

The UAV datasets used in the experiments are part of the dataset acquired by the ISPRS/EuroSDR benchmark for multi-platform photogrammetry, run in collaboration with EuroSDR (Nex et al., 2015). The characteristics of this dataset are presented in section 5.6.2 of this dissertation. Five UAV datasets are used in the experiments, as illustrated in Figure 7-8:

- *UAV Dataset 1*: 23 oblique UAV images;
- *UAV Dataset 2:* 35 oblique UAV images;
- *UAV Dataset 3*: 60 oblique UAV images;
- *UAV Dataset 4*: 80 oblique UAV images; and
- *UAV Dataset 5*: 100 oblique UAV images.

For the first two UAV datasets (i.e., *UAV Dataset 1* and *UAV Dataset 2*), initial exterior orientation parameters have been computed using both the developed *CP-based EO Estimation* method as well as the *VP-based EO Estimation* technique. The starting image (*img_1*) of *UAV Dataset 2*, where the four coplanar GCPs have been measured, does not lie on the borders of the dataset; on the contrary, the starting image (*img_1*) of the other datasets (i.e., *UAV Dataset 1*, *UAV Dataset 3*, *UAV Dataset 4* and *UAV Dataset 5*) is the first image of the sequence. The same starting image for all UAV datasets is used, as illustrated in Figure 7-8. The coplanar GCPs, measured in *img_1* of all UAV datasets, is illustrated in Figure 5-6 (section 5.6.3). The reference data used in the experiments performed using all UAV datasets are presented in section 5.6.3.

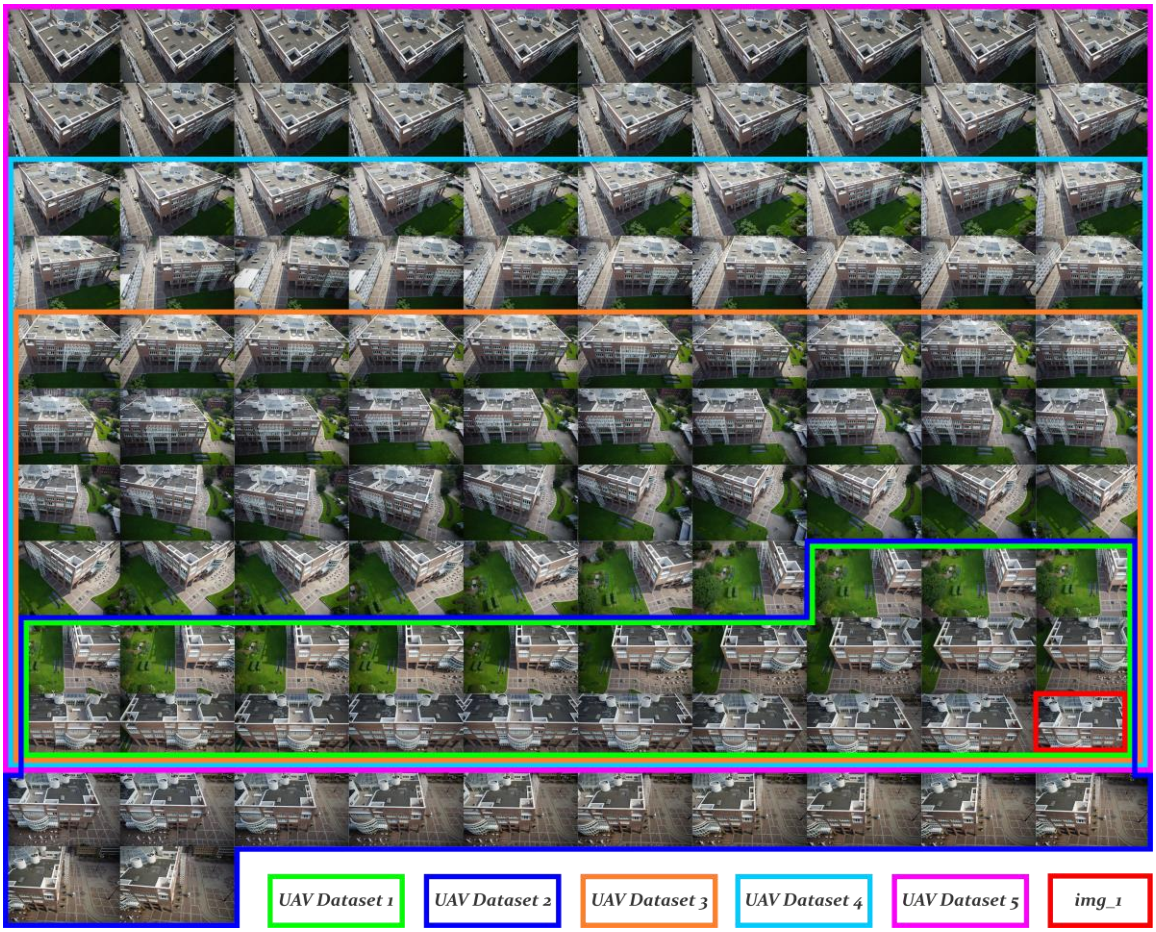


Figure 7-8. *UAV Datasets 1-5* and *img_1* used for manual measurement of 4 coplanar GCPs

### 7.10.2.2 Aerial Image Datasets

The aerial image datasets used in the experiments are part of a dataset that consists of multi-perspective oblique and vertical aerial images, which are provided by the ISPRS/EuroSDR benchmark on high density image matching for DSM computation (Cavegn et al., 2014). The characteristics of this dataset have been presented in section 4.4.1 (see *RCD30 Oblique Penta Dataset*). The photogrammetric block used in the experiments presented in this chapter consists of one strip. Figure 7-9 and Figure 7-10 illustrate the ground footprints of the captured imagery along with the overlap both among the simultaneously acquired multi-view imagery and among the successive images taken from a single camera. The images are accompanied with reference exterior

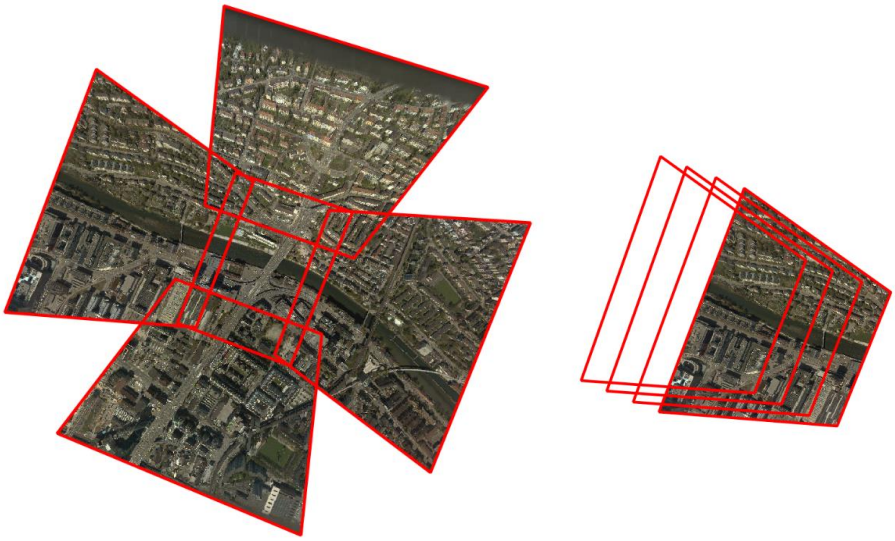

Figure 7-9. Ground footprints of the images acquired by the Leica RCD30 Oblique Penta camera system, generated using the georeferencing software suite presented in section 4.4.2; left: images acquired by the five cameras at one time instance; right: successive images taken by one oblique camera

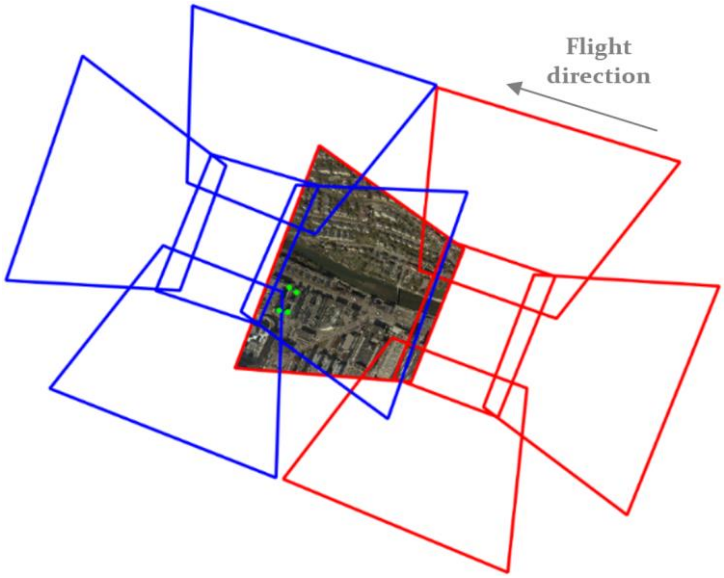


Figure 7-10. Ground footprints of the images that correspond to the first (shown in red) and the last (shown in blue) acquisition of the strip captured by the Leica RCD30 Oblique Penta camera system, along with the image used for manual measurements of four coplanar GCPs (shown in green) generated using the georeferencing software suite presented in section 4.4.2

orientation parameters computed through a highly accurate triangulation (Cavegn et al., 2014), which are used for assessing the results derived using the developed aerial triangulation framework. The coordinates of the coplanar GCPs that served as the ground control of the starting image (*img_1*) of all three aerial image datasets were computed through space intersection using the reference exterior orientation parameters.

Three datasets encompassing different configurations of oblique aerial images are used in the experiments, as described in the following (Figure 7-11):

- *RCD30 Dataset 1*: 10 oblique aerial images of the same perspective, captured by one camera of the multi-view system;
- *RCD30 Dataset 2*: 40 multi-perspective oblique aerial images, captured by the four cameras of the multi-view system; and
- *RCD30 Dataset 3*: 50 multi-perspective oblique and vertical aerial images, including 40 oblique and 10 vertical aerial images captured in a Maltese-cross configuration by the five cameras of the multi-view system.

### 7.10.3 Performance of Individual Steps

In this section, the performance of individual steps of the developed SfM methods, in terms of automatic transfer of GCPs, feature tracking and robust outlier removal are outlined.

#### 7.10.3.1 Feature Tracking

As far as the five UAV datasets are concerned, about 50% of the initial 3D feature points (tracks) are identified in more than three images and are finally used by the bundle adjustment processes of the developed SfM methods, both using the SURF and ASIFT feature extractors. For the three RCD30 aerial image datasets, this percentage is much smaller; as far as the 10 single-perspective oblique aerial images are concerned (*RCD30 Dataset 1*), about 33% of the initially identified feature points reach the bundle adjustment procedures, while this percentage is about 20% for the multi-perspective aerial image datasets (*RCD30 Dataset 2* and *RCD30 Dataset 3*) using both SURF and ASIFT. For instance, for the case of the SURF feature detector, for *UAV Dataset 5* (which includes 100 images), 3572 tracks out of the initially extracted 7483 tracks include tie point measurements in at least three images and are finally used by the bundle adjustment scheme; for *RCD30 Dataset*

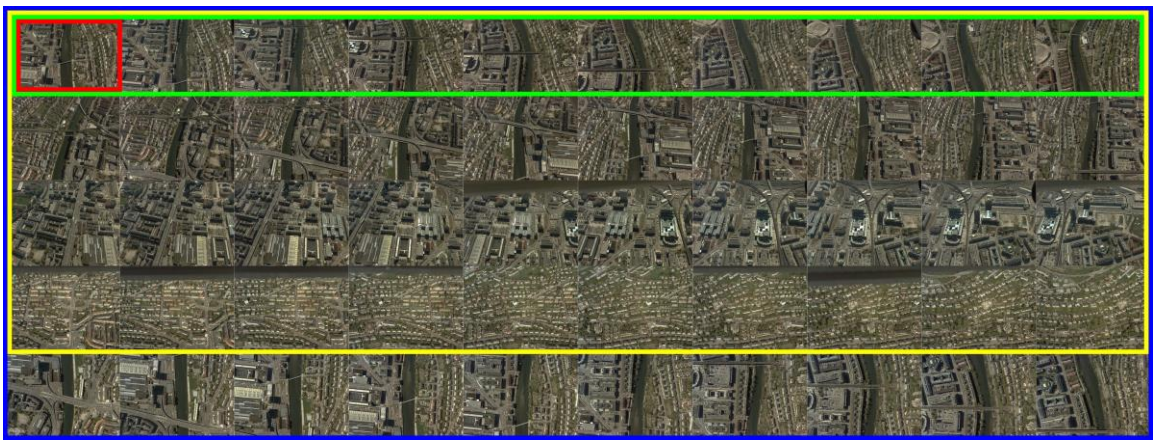

Figure 7-11. *RCD30 Dataset 1* (outlined in green), *RCD30 Dataset 2* (outlined in yellow), *RCD30 Dataset 3* (outlined in blue) and *img_1* used for manual measurement of 4 coplanar GCPs

*3* (which includes 50 images), 7844 out of the initially extracted 43878 tracks are used by the bundle adjustment scheme; for *RCD30 Dataset 1* (which includes 10 images), 2028 out of the initially extracted 6099 tracks are used by the bundle adjustment scheme. The mean number of images that depict the same 3D feature point is about 5.3 for the case of SURF features of all UAV datasets and 5.5 for the case of ASIFT features. As far as the RCD30 datasets are concerned, a mean number of about 3.5 images have been identified to depict the same 3D point, for the case of both SURF and ASIFT. As far as the time cost is concerned, for instance it is mentioned that in the case of the SURF detector, the image matching and feature tracking steps for *UAV Dataset 5* took 26 minutes while these steps for *RCD30 Dataset 3* were executed in 32 minutes; the experiments were performed on a 64-bit Intel Core i7 2.4 GHz laptop computer with 16 GB of RAM.

### 7.10.3.2 Automatic Transfer of GCPs

The coplanar point detection method described in section 5.3 has been applied in the images that overlap with *img_1* for all datasets, for the recognition of the four manually measured coplanar GCPs in at least another image of the datasets. For the case of *UAV Dataset 2* (for which *img_1* is not the first image of the sequence), the four GCPs have been automatically recognized in five images; for the case of the other UAV datasets, the GCPs have been automatically recognized in three images; for the case of the RCD30 aerial image datasets, the four GCPs have been automatically identified in one image. The results of the coplanar point detection method as applied in *UAV Dataset 1* and *UAV Dataset 2* have already been presented in section 5.6.4. The results that refer to the images that are georeferenced based on image matching with *img_1* are the same for the images of all UAV datasets to which the GCPs have been transferred, as the same starting image (*img_1*) along with the same GCP measurements have been used for all datasets. Figure 7-12 shows the automatically detected coplanar feature points in *img_1* of the RCD30 datasets (shown in red), based on the manually measured coplanar GCPs (shown in green). Specifically, 409 corresponding feature points in *img_1* and its overlapping image with the maximum number of correspondences (*img_2*) have been identified by RANSAC as being coplanar with the manually measured GCPs. Figure 7-13 illustrates the iterative process of transferring the coplanar GCPs from *img_1* to *img_2*

Figure 7-12. Automatically detected coplanar feature points in *img_1* of RCD30 datasets (shown in red) that lie on the plane of the four manually measured coplanar GCPs (shown in green)

of the RCD30 datasets and detecting feature points that lie on the plane of GCPs in both images. In the second iteration of the coplanar point detection process, the correct homography along with the coplanar feature points that lie on the plane of GCPs were identified. The process of automatic transfer of GCPs in an image was executed in less than 1 minute for all test datasets using a 64-bit Intel Core i7 2.4 GHz laptop computer with 16 GB of RAM.

#### 7.10.3.3 Robust Outlier Removal

Except for the case of *RCD30 Dataset 1*, for which all outliers have been eliminated by the image matching process, for all other datasets, the remaining erroneous correspondences are rejected through the robust iterative bundle adjustment framework. The percentage of wrong tracks to tracks that reach the first bundle adjustment process of the robust aerial triangulation method lies between 0.3% and 1.2% using both SURF and ASIFT. For instance, it is mentioned that in the case of the SURF feature detector, in *UAV Dataset 3*, 13 out of 2337 tracks of at least three correspondences that reach bundle adjustment are iteratively rejected; for the case of *UAV Dataset 5*, 37 out of 3572 tracks are removed; and for the case of *RCD30 Dataset 3*, 23 out of 7844 tracks are rejected. Similar numbers of outliers have also been derived in the case of ASIFT features (e.g., 15 out of 2416 ASIFT feature point correspondences of *UAV Dataset 3* are iteratively rejected by the robust outlier removal scheme). As far as time cost is concerned, for instance it is mentioned that for the case of *UAV Dataset 3* (that corresponds to 2337 tracks and 12551 automatic tie point measurements), about 3 minutes were required for the rejection of each outlier through the bundle adjustment, as tested using a 64-bit Intel Core i7 2.4 GHz laptop computer with 16 GB of RAM.

### 7.10.4 Comparison of SfM Methods

In all experiments outlined in sections 7.10.4, 7.10.5, 7.10.6 and 7.10.7, four GCPs are measured in one starting image (*img_1* for each dataset, as discussed in sections 7.10.2.1 and 7.10.2.2). In

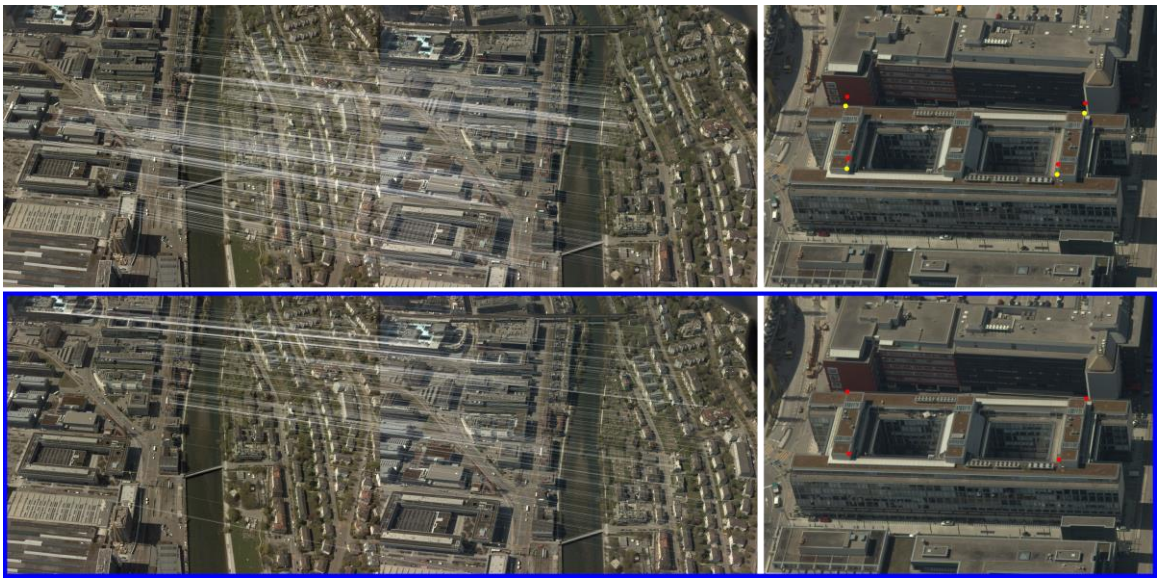

Figure 7-13. Visualization of the iterative process of transferring the coplanar GCPs from *img_1* to *img_2* for the RCD30 datasets; left: detected coplanar corresponding feature points between *img_1* and *img_2*; right: zoom-in views of the homography-based GCP estimations (in red) and the actual GCPs (in yellow); the blue outline indicates the identified correspondences that lie on the plane of GCPs (left) along with the correct homography-based estimations of the GCPs in *img_2* (right)

addition, the same fixed interior orientation parameters are specified for each experiment outlined in sections 7.10.4, 7.10.5, 7.10.6, 7.10.7 and 7.10.9, so that the derived results are comparable with each other as well as with the results derived by commercial software (section 7.10.8). As far as the five UAV datasets are concerned, their interior orientation parameters are the ones computed through the auto-calibrating bundle adjustment using the whole block of 645 images of the benchmark (section 5.6.3) while the interior orientation parameters of the three RCD30 aerial image datasets captured by multi-camera system are provided by the benchmark.

In all experiments mentioned in this section, the SURF algorithm is used for feature extraction; in addition, the bundle adjustment procedures assume equal weights for all point observations. Initially, in sections 7.10.4.1 and 7.10.4.2, the impact of the workflow (*Robust Incremental SfM* or *Robust Global SfM*) along with the method for obtaining approximate exterior orientation parameters in the case of *Robust Global SfM* (*Incremental SfM*; *CP-based EO Estimation* followed by *Global SfM*; and *VP-based EO Estimation* followed by *Global SfM*) are assessed using the images of *UAV Dataset 1* and *UAV Dataset 2*. Specifically, the following four workflows are assessed:

i. *Robust Incremental SfM;*
ii. *Incremental SfM* followed by *Robust Global SfM*, hereinafter referred to as *IncrSfM_RobustGlobalSfM;*
iii. *CP-based EO Estimation* followed by *Global SfM* and *Robust Global SfM*, hereinafter referred to as *CP_RobustGlobalSfM*; and
iv. *VP-based EO Estimation* followed by *Global SfM* and *Robust Global SfM*, hereinafter referred to as *VP_RobustGlobalSfM*.

For reasons of convenience, the workflows that implement a *Robust Global SfM* as a last step have been abbreviated, as presented above (workflows ii, iii and iv). In the case that a *Robust Global SfM* method uses initial values from an *Incremental SfM* or a *Global SfM* method (i.e., in cases of workflows ii, iii and iv), the *Robust Global SfM* method does not implement again the steps of image matching, feature tracking and automatic transfer of GCPs, as they have already been implemented by the SfM method that precedes in the workflow. Thus, in such cases, the *Robust Global SfM* method only consists of the aerial triangulation stage (computation of approximate ground coordinates for the tie points, outlier removal and iterative bundle adjustment procedures).

In the following, the differences ($\Delta X_0$, $\Delta Y_0$, $\Delta Z_0$, $\Delta\omega$, $\Delta\varphi$, $\Delta\kappa$) between the exterior orientation parameters computed through one of the developed solutions and the reference exterior orientation parameters are presented in the form of tables. Specifically, the average absolute differences (Avg-Abs), the maximum (Max-Abs) and minimum (Min-Abs) absolute differences, the average differences (Avg) and the standard deviation of differences (Stdev) between the exterior orientation parameters computed through one of the developed solutions and the reference ones, along with Avg-Abs, Max-Abs, Min-Abs, Avg and Stdev for the linear ($\Delta$*LinearEO*) and angular ($\Delta$*AngularEO*) exterior orientation parameters, computed through the set of equations (5.26) (section 5.6.4.3), are outlined for each experiment.

Except for the metric that estimates the average angular difference using the Euler angles $\omega$, $\varphi$, $\kappa$ ($\Delta$*AngularEO*), which is used in this dissertation, other metrics have also been reported in literature that may be used to estimate angular difference (Huynh, 2009; Hartley et al., 2013). For instance, the distance of quaternions, $D_{quaternions}$, is another metric. This metric requires the conversion from the angles $\omega$, $\varphi$, $\kappa$ to a quaternion-representation $\boldsymbol{q} = [q_1 \;\; q_2 \;\; q_3 \;\; q_4]^T$, or equivalently $\boldsymbol{q} = q_1 + \boldsymbol{i}q_2 + \boldsymbol{j}q_3 + \boldsymbol{k}q_4$, where $q_1$ is the real part of the quaternion, and $\boldsymbol{i}$, $\boldsymbol{j}$ and $\boldsymbol{k}$ are unit vectors pointing along

Table 7-2. Differences between the EO parameters computed through *Robust Incremental SfM* (using the SURF feature extractor, equal weights for observations in bundle adjustment (BA) and four GCP measurements in *img_1*) and the reference ones for *UAV Dataset 1*

| ***UAV Dataset 1*: 23 oblique UAV images** | | | | | | |
|---|---|---|---|---|---|---|
| **Method** | **Metric** | **Avg-Abs** | **Max-Abs** | **Min-Abs** | **Avg** | **Stdev** |
| ***Robust Incremental SfM* (SURF; unweighted BA; 4 GCPs in one image)** | $\Delta X_o$ (m) | 0.479 | 1.136 | 0.019 | 0.403 | 0.335 |
| | $\Delta Y_o$ (m) | 0.171 | 0.210 | 0.141 | 0.171 | 0.280 |
| | $\Delta Z_o$ (m) | 0.071 | 0.154 | 0.008 | -0.070 | 0.312 |
| | $\Delta\omega$ (deg.) | 0.614 | 0.736 | 0.437 | 0.614 | 0.372 |
| | $\Delta\varphi$ (deg.) | 0.157 | 0.412 | 0.009 | -0.111 | 0.307 |
| | $\Delta\kappa$ (deg.) | 0.093 | 0.203 | 0.001 | 0.047 | 0.266 |
| | $\Delta$*LinearEO* (m) | 0.240 | 1.136 | 0.008 | 0.168 | 0.309 |
| | $\Delta$*AngularEO* (deg.) | 0.288 | 0.736 | 0.001 | 0.183 | 0.315 |

Table 7-3. Differences between the EO parameters computed through *Robust Incremental SfM* (using the SURF feature extractor, equal weights for observations in BA and four GCP measurements in *img_1*) and the reference ones for *UAV Dataset 2*

| ***UAV Dataset 2*: 35 oblique UAV images** | | | | | | |
|---|---|---|---|---|---|---|
| **Method** | **Metric** | **Avg-Abs** | **Max-Abs** | **Min-Abs** | **Avg** | **Stdev** |
| ***Robust Incremental SfM* (SURF; unweighted BA; 4 GCPs in one image)** | $\Delta X_o$ (m) | 0.236 | 0.402 | 0.002 | -0.174 | 0.191 |
| | $\Delta Y_o$ (m) | 0.228 | 0.493 | 0.012 | 0.214 | 0.172 |
| | $\Delta Z_o$ (m) | 0.137 | 0.276 | 0.001 | -0.030 | 0.155 |
| | $\Delta\omega$ (deg.) | 0.565 | 0.683 | 0.337 | 0.565 | 0.096 |
| | $\Delta\varphi$ (deg.) | 0.475 | 0.822 | 0.136 | -0.475 | 0.151 |
| | $\Delta\kappa$ (deg.) | 0.093 | 0.172 | 0.011 | 0.046 | 0.094 |
| | $\Delta$*LinearEO* (m) | 0.200 | 0.493 | 0.001 | 0.003 | 0.173 |
| | $\Delta$*AngularEO* (deg.) | 0.378 | 0.822 | 0.011 | 0.045 | 0.114 |

the three spatial axes and forming the imaginary part of the quaternion. The formula for calculating a quaternion from the set of rotation angles $\omega$, $\varphi$, $\kappa$ applied sequentially (from $\omega$ to $\kappa$) is derived by calculating the product of the quaternions computed for each of the rotation angles, as expressed in equation (7.7).

$$\left.\begin{aligned} \boldsymbol{q}_{\omega} &= \cos(\omega/2) + \boldsymbol{i}\sin(\omega/2) \\ \boldsymbol{q}_{\varphi} &= \cos(\varphi/2) + \boldsymbol{j}\sin(\varphi/2) \\ \boldsymbol{q}_{\kappa} &= \cos(\kappa/2) + \boldsymbol{k}\sin(\kappa/2) \end{aligned}\right\} \; \boldsymbol{q} = \boldsymbol{q}_{\kappa}\boldsymbol{q}_{\varphi}\boldsymbol{q}_{\omega} \tag{7.7}$$

For the case of the experiments presented in this dissertation, $D_{quaternions}$ indicates the distance between the unit quaternion that corresponds to the Euler angles ($\omega$, $\varphi$, $\kappa$) computed by a tested SfM method or commercial software for an image ($\boldsymbol{q}$) and the unit quaternion that corresponds to the reference Euler angles for the same image ($\boldsymbol{q}_{ref}$); it is calculated through equation (7.8), where $\|\cdot\|$ symbolizes the Euclidean norm in $\mathbb{R}^4$. Taking into account the fact that $\boldsymbol{q}$ and $-\boldsymbol{q}$ represent the same rotation, the quaternion distance is calculated as the minimum between the Euclidean norm of the difference of quaternions and the Euclidean norm of the sum of quaternions. This distance is dimensionless and gives values in the range $[\,0, \sqrt{2}\,]$.

$$D_{quaternions} = \min\left\{\left\|\boldsymbol{q} - \boldsymbol{q}_{ref}\right\|, \left\|\boldsymbol{q} + \boldsymbol{q}_{ref}\right\|\right\} \tag{7.8}$$

In order to obtain a metric that gives an estimation of the angular difference between the calculated angles and the reference ones in degrees, rather than a dimensionless metric, $\Delta AngularEO$ is used throughout this dissertation to analyze the obtained results. However, similar results in terms of percentage of angular difference are also derived using the quaternion distance, as it is discussed in section 7.10.10 of this dissertation.

### 7.10.4.1 Robust Incremental SfM

Table 7-2 and Table 7-3 present the results achieved through the developed *Robust Incremental SfM* method for *UAV Dataset 1* and *UAV Dataset 2*. A relatively smaller average absolute

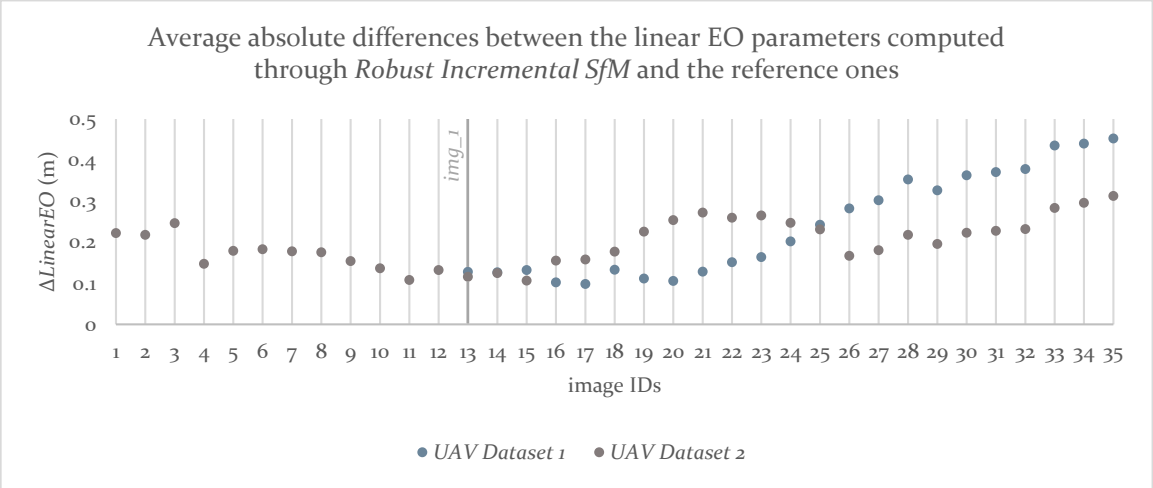


Figure 7-14. Average absolute differences between the linear EO parameters computed through *Robust Incremental SfM* (using the SURF feature extractor, equal weights for observations in BA and four GCP measurements in *img_1*) and the reference ones for the images *of UAV Dataset 1* and *UAV Dataset 2*

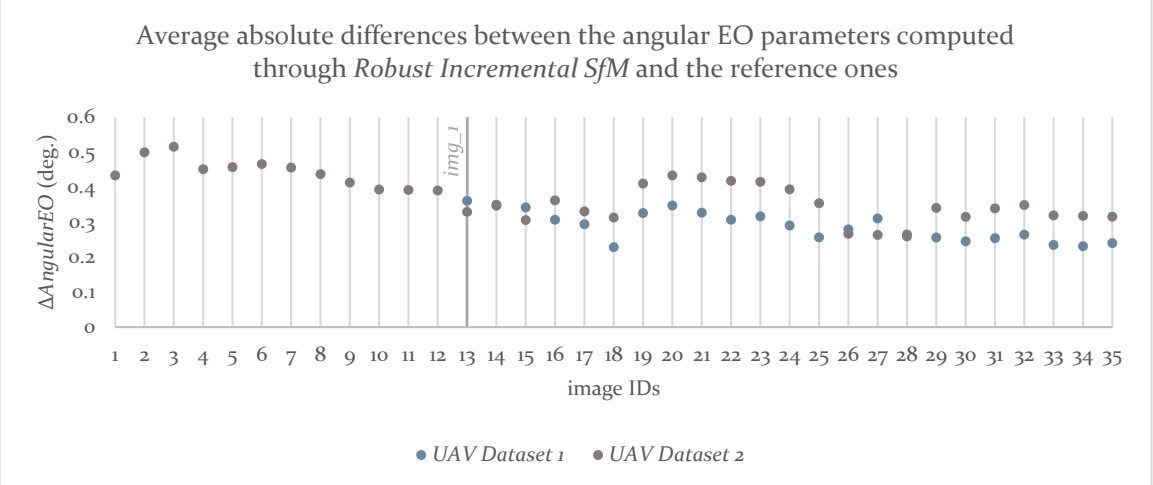


Figure 7-15. Average absolute differences between the angular EO parameters computed through *Robust Incremental SfM* (using the SURF feature extractor, equal weights for observations in BA and four GCP measurements in *img_1*) and the reference ones for the images *of UAV Dataset 1* and *UAV Dataset 2*

Table 7-4. Differences between the EO parameters computed (i) through *Incremental SfM* and (ii) through *Robust Global SfM* using as initial EO parameters the ones estimated by *Incremental SfM* (using the SURF feature extractor, equal weights for observations in BA and four GCP measurements in *img_1*) and the reference ones for *UAV Dataset 1*

| ***UAV Dataset 1*: 23 oblique UAV images** | | | | | | |
|---|---|---|---|---|---|---|
| **Method** | **Metric** | **Avg-Abs** | **Max-Abs** | **Min-Abs** | **Avg** | **Stdev** |
| ***Incremental SfM* (SURF; unweighted BA; 4 GCPs in one image)** | $\Delta X_o$ (m) | 0.434 | 1.034 | 0.001 | 0.357 | 0.325 |
| | $\Delta Y_o$ (m) | 0.175 | 0.216 | 0.139 | 0.175 | 0.284 |
| | $\Delta Z_o$ (m) | 0.077 | 0.154 | 0.006 | -0.076 | 0.316 |
| | $\Delta\omega$ (deg.) | 0.614 | 0.737 | 0.442 | 0.614 | 0.362 |
| | $\Delta\varphi$ (deg.) | 0.169 | 0.409 | 0.015 | -0.136 | 0.288 |
| | $\Delta\kappa$ (deg.) | 0.091 | 0.210 | 0.000 | 0.055 | 0.240 |
| | $\Delta$*LinearEO* (m) | 0.229 | 1.034 | 0.001 | 0.152 | 0.308 |
| | $\Delta$*AngularEO* (deg.) | 0.292 | 0.737 | 0.000 | 0.178 | 0.297 |
| ***Robust Global SfM:* initial EO from *Incremental SfM* (SURF; unweighted BA; 4 GCPs in one image)** | $\Delta X_o$ (m) | 0.303 | 0.622 | 0.015 | 0.179 | 0.272 |
| | $\Delta Y_o$ (m) | 0.215 | 0.264 | 0.139 | 0.215 | 0.264 |
| | $\Delta Z_o$ (m) | 0.050 | 0.124 | 0.002 | -0.031 | 0.291 |
| | $\Delta\omega$ (deg.) | 0.563 | 0.712 | 0.369 | 0.563 | 0.298 |
| | $\Delta\varphi$ (deg.) | 0.166 | 0.399 | 0.017 | -0.096 | 0.214 |
| | $\Delta\kappa$ (deg.) | 0.104 | 0.204 | 0.011 | 0.039 | 0.160 |
| | $\Delta$*LinearEO* (m) | 0.189 | 0.622 | 0.002 | 0.121 | 0.275 |
| | $\Delta$*AngularEO* (deg.) | 0.277 | 0.712 | 0.011 | 0.169 | 0.224 |

$\Delta$*LinearEO* value (reduction by 17%) accompanied by a relatively bigger average absolute *ΔAngularEO* value (increase by 31%) are observed for *UAV Dataset 2*, which is the dataset of 35 images for which the GCPs have been measured in an internal image of the block, compared to *UAV Dataset 1*, which includes GCP measurements in the first image of the sequence of 23 UAV images. The maximum absolute *ΔLinearEO* is significantly bigger in the case of *UAV Dataset 1* compared to *UAV Dataset 2*, because of the distribution of GCPs inside the datasets. The diagram in Figure 7-14 shows the distribution of absolute $\Delta$*LinearEO* for all images of *UAV Dataset 1* and *UAV Dataset 2*, revealing a greater absolute value of $\Delta$*LinearEO* for images that are far away from the image with GCP measurements (*img_1*) for both datasets. The increase of *ΔLinearEO* with the distance from *img_1* is greater in the case of *UAV Dataset 1*, which includes GCP measurements in the first image of the sequence. However, a more even distribution of absolute $\Delta$*AngularEO* is observed for the images of both *UAV Dataset 1* and *UAV Dataset 2*, as illustrated in the diagram of Figure 7-15, revealing that the differences between the angular exterior orientation parameters of an image computed through *Robust Incremental SfM* and the reference ones do not depend on its relative position with respect to the image that includes GCP measurements (*img_1*).

### 7.10.4.2 Robust Global SfM

The developed *Robust Global SfM* workflow is tested using *UAV Dataset 1* and *UAV Dataset 2* assuming three different scenarios for obtaining initial values of exterior orientation parameters.

#### *IncrSfM_RobustGlobalSfM*

The first method for obtaining approximate exterior orientation parameters is an *Incremental SfM* method, without iterative bundle adjustment procedures for outlier removal. Table 7-4 and Table

Table 7-5. Differences between the EO parameters computed (i) through *Incremental SfM* and (ii) through *Robust Global SfM* using as initial EO parameters the ones estimated by *Incremental SfM* (using the SURF feature extractor, equal weights for observations in BA and four GCP measurements in *img_1*) and the reference ones for *UAV Dataset 2*

| ***UAV Dataset 2*: 35 oblique UAV images** | | | | | | |
|---|---|---|---|---|---|---|
| **Method** | **Metric** | **Avg-Abs** | **Max-Abs** | **Min-Abs** | **Avg** | **Stdev** |
| ***Incremental SfM* (SURF; unweighted BA; 4 GCPs in one image)** | $\Delta X_o$ (m) | 0.548 | 3.497 | 0.048 | 0.020 | 0.878 |
| | $\Delta Y_o$ (m) | 0.179 | 0.620 | 0.004 | 0.164 | 0.147 |
| | $\Delta Z_o$ (m) | 0.153 | 0.655 | 0.004 | 0.077 | 0.208 |
| | $\Delta\omega$ (deg.) | 0.524 | 1.000 | 0.032 | 0.514 | 0.226 |
| | $\Delta\varphi$ (deg.) | 0.676 | 3.419 | 0.046 | -0.226 | 0.945 |
| | $\Delta\kappa$ (deg.) | 0.133 | 0.420 | 0.001 | 0.102 | 0.150 |
| | $\Delta$*LinearEO* (m) | 0.293 | 3.497 | 0.004 | 0.087 | 0.411 |
| | $\Delta$*AngularEO* (deg.) | 0.444 | 3.419 | 0.001 | 0.130 | 0.440 |
| ***Robust Global SfM:* initial EO from *Incremental SfM* (SURF; unweighted BA; 4 GCPs in one image)** | $\Delta X_o$ (m) | 0.302 | 0.641 | 0.022 | -0.080 | 0.325 |
| | $\Delta Y_o$ (m) | 0.202 | 0.390 | 0.022 | 0.187 | 0.143 |
| | $\Delta Z_o$ (m) | 0.097 | 0.264 | 0.005 | 0.030 | 0.121 |
| | $\Delta\omega$ (deg.) | 0.535 | 0.657 | 0.267 | 0.535 | 0.119 |
| | $\Delta\varphi$ (deg.) | 0.392 | 0.764 | 0.009 | -0.357 | 0.279 |
| | $\Delta\kappa$ (deg.) | 0.085 | 0.226 | 0.007 | 0.042 | 0.095 |
| | $\Delta$*LinearEO* (m) | 0.200 | 0.641 | 0.005 | 0.046 | 0.196 |
| | $\Delta$*AngularEO* (deg.) | 0.337 | 0.764 | 0.007 | 0.073 | 0.164 |

7-5 present the results achieved (i) through the *Incremental SfM* method; and (ii) through the *Robust Global SfM* method, using as initial exterior orientation parameters the ones calculated through *Incremental SfM*. For both UAV datasets, an improvement in the exterior orientation parameters computed through *Robust Global SfM* relative to the ones computed through *Incremental SfM* is observed. This is an expected outcome, as the *Incremental SfM* process, which does not incorporate any iterative bundle adjustment techniques, does not eliminate all erroneous point observations. Specifically, the average absolute $\Delta$*LinearEO* and $\Delta$*AngularEO* are reduced by 17% and 5% thanks to the robust iterative outlier removal techniques of *Robust Global SfM* for *UAV Dataset 1*. The corresponding decrease for *UAV Dataset 2* reaches the percentages of 32% for $\Delta$*LinearEO* and 24% for $\Delta$*AngularEO*. The improvement in the results of *UAV Dataset 2* is greater, as more outliers are removed thanks to the *Robust Global SfM* method. Besides, in *UAV Dataset 2*, a much greater maximum absolute value of both $\Delta$*LinearEO* and $\Delta$*AngularEO* is observed (Table 7-5) in the results of *Incremental SfM* (3.497 m – 3.419 deg.) compared to the results of *Robust Global SfM* (0.641 m – 0.764 deg.), revealing some gross errors that are due to the remaining outliers.

Furthermore, the conclusion derived from the results achieved through the *Robust Incremental SfM* method that a greater absolute value of $\Delta$*LinearEO* corresponds to images that are far away from *img_1* is also valid for the case of the *Robust Global SfM* method initialized by *Incremental SfM* for both datasets, as illustrated in the diagram of Figure 7-16. Similarly, the distribution of absolute $\Delta$*AngularEO* is more even for both datasets, like the corresponding distribution of $\Delta$*AngularEO* derived for *Robust Incremental SfM*, as illustrated in Figure 7-15. Similar conclusions are also derived from the results of the *Incremental SfM* method. Finally, better results are achieved through the *Robust Global SfM* method for *UAV Dataset 1*, compared to *UAV Dataset 2*, which corresponds to an increase of $\Delta$*LinearEO* by 6% and an increase of $\Delta$*AngularEO* by 22%.

Table 7-6. Differences between the EO parameters computed (i) through *Global SfM* using as initial EO parameters the ones estimated through *CP-based EO Estimation* (Table 5-6) and (ii) through *Robust Global SfM* using as initial EO parameters the ones estimated by *Global SfM* (using the SURF feature extractor, equal weights for observations in BA and four GCP measurements in *img_1*) and the reference ones for *UAV Dataset 1*

| ***UAV Dataset 1*: 23 oblique UAV images** | | | | | | |
|---|---|---|---|---|---|---|
| **Method** | **Metric** | **Avg-Abs** | **Max-Abs** | **Min-Abs** | **Avg** | **Stdev** |
| ***Global SfM:* initial EO from *CP-based EO Estimation* (SURF; unweighted BA; 4 GCPs in one image)** | $\Delta X_o$ (m) | 0.960 | 1.849 | 0.025 | 0.890 | 0.482 |
| | $\Delta Y_o$ (m) | 0.105 | 0.177 | 0.017 | -0.007 | 0.330 |
| | $\Delta Z_o$ (m) | 0.134 | 0.330 | 0.002 | 0.121 | 0.341 |
| | $\Delta\omega$ (deg.) | 0.665 | 0.773 | 0.562 | 0.665 | 0.515 |
| | $\Delta\varphi$ (deg.) | 0.475 | 1.155 | 0.004 | 0.349 | 0.524 |
| | $\Delta\kappa$ (deg.) | 0.079 | 0.180 | 0.006 | -0.017 | 0.494 |
| | $\Delta$*LinearEO* (m) | 0.400 | 1.849 | 0.002 | 0.335 | 0.385 |
| | $\Delta$*AngularEO* (deg.) | 0.407 | 1.155 | 0.004 | 0.332 | 0.511 |
| ***Robust Global SfM:* initial EO from *Global SfM* (SURF; unweighted BA; 4 GCPs in one image)** | $\Delta X_o$ (m) | 0.494 | 1.027 | 0.047 | 0.403 | 0.323 |
| | $\Delta Y_o$ (m) | 0.137 | 0.199 | 0.079 | 0.137 | 0.278 |
| | $\Delta Z_o$ (m) | 0.068 | 0.144 | 0.001 | 0.038 | 0.311 |
| | $\Delta\omega$ (deg.) | 0.626 | 0.869 | 0.494 | 0.626 | 0.364 |
| | $\Delta\varphi$ (deg.) | 0.227 | 0.487 | 0.001 | 0.036 | 0.301 |
| | $\Delta\kappa$ (deg.) | 0.097 | 0.339 | 0.015 | -0.048 | 0.268 |
| | $\Delta$*LinearEO* (m) | 0.233 | 1.027 | 0.001 | 0.193 | 0.304 |
| | $\Delta$*AngularEO* (deg.) | 0.317 | 0.869 | 0.001 | 0.205 | 0.311 |

### *CP_RobustGlobalSfM*

The second workflow for obtaining approximate exterior orientation parameters consists of the coplanar point detection method for exterior orientation estimation (*CP-based EO Estimation*), as presented in section 5.5, followed by a *Global SfM* method without iterative procedures for outlier removal. The results of *CP-based EO Estimation* have been outlined in Table 5-6 and Table 5-7 for *UAV Dataset 1* and *UAV Dataset 2*, respectively. Hence, Table 7-6 and Table 7-7 present the results

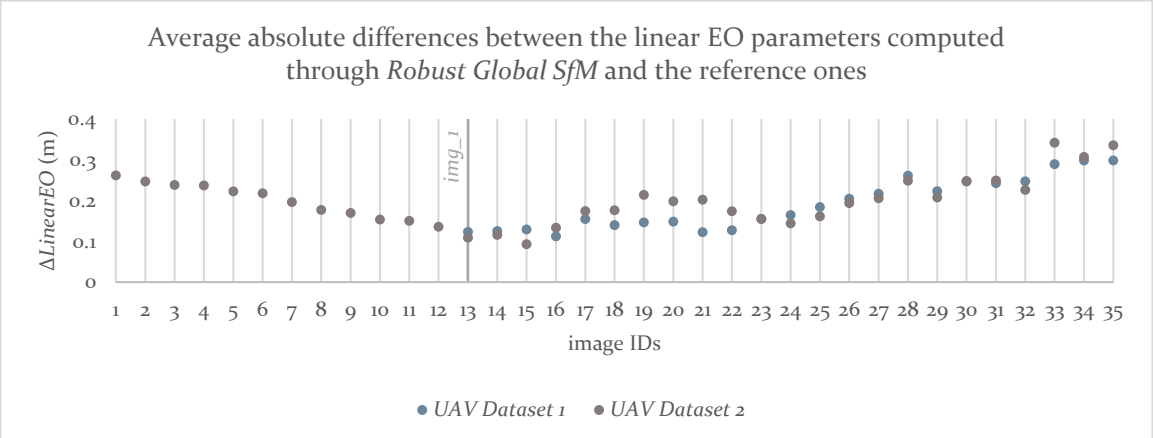


Figure 7-16. Average absolute differences between the linear EO parameters computed through *Robust Global SfM* initialized by *Incremental SfM* (using the SURF feature extractor, equal weights for observations in BA and four GCP measurements in *img_1*) and the reference ones for the images of *UAV Dataset 1* and *UAV Dataset 2*

Table 7-7. Differences between the EO parameters computed (i) through *Global SfM* using as initial EO parameters the ones estimated through *CP-based EO Estimation* (Table 5-7) and (ii) through *Robust Global SfM* using as initial EO parameters the ones estimated by *Global SfM* (using the SURF feature extractor, equal weights for observations in BA and four GCP measurements in *img_1*) and the reference ones for *UAV Dataset 2*

| ***UAV Dataset 2*: 35 oblique UAV images** | | | | | | |
|---|---|---|---|---|---|---|
| **Method** | **Metric** | **Avg-Abs** | **Max-Abs** | **Min-Abs** | **Avg** | **Stdev** |
| ***Global SfM:* initial EO from *CP-based EO Estimation* (SURF; unweighted BA; 4 GCPs in one image)** | $\Delta X_o$ (m) | 0.449 | 1.501 | 0.003 | 0.084 | 0.540 |
| | $\Delta Y_o$ (m) | 0.190 | 0.745 | 0.003 | -0.002 | 0.243 |
| | $\Delta Z_o$ (m) | 0.117 | 0.578 | 0.001 | 0.055 | 0.154 |
| | $\Delta\omega$ (deg.) | 0.736 | 1.450 | 0.300 | 0.736 | 0.206 |
| | $\Delta\varphi$ (deg.) | 0.627 | 2.429 | 0.069 | 0.001 | 0.770 |
| | $\Delta\kappa$ (deg.) | 0.128 | 0.566 | 0.002 | 0.069 | 0.162 |
| | $\Delta$*LinearEO* (m) | 0.252 | 1.501 | 0.001 | 0.046 | 0.312 |
| | $\Delta$*AngularEO* (deg.) | 0.497 | 2.429 | 0.002 | 0.269 | 0.379 |
| ***Robust Global SfM:* initial EO from *Global SfM* (SURF; unweighted BA; 4 GCPs in one image)** | $\Delta X_o$ (m) | 0.247 | 1.005 | 0.010 | -0.089 | 0.312 |
| | $\Delta Y_o$ (m) | 0.192 | 0.390 | 0.001 | 0.171 | 0.148 |
| | $\Delta Z_o$ (m) | 0.104 | 0.273 | 0.001 | 0.071 | 0.118 |
| | $\Delta\omega$ (deg.) | 0.565 | 0.729 | 0.363 | 0.565 | 0.083 |
| | $\Delta\varphi$ (deg.) | 0.436 | 1.544 | 0.012 | -0.183 | 0.515 |
| | $\Delta\kappa$ (deg.) | 0.101 | 0.540 | 0.002 | -0.048 | 0.150 |
| | $\Delta$*LinearEO* (m) | 0.181 | 1.005 | 0.001 | 0.051 | 0.193 |
| | $\Delta$*AngularEO* (deg.) | 0.367 | 1.544 | 0.002 | 0.112 | 0.249 |

achieved (i) through the *Global SfM* method, using the approximate exterior orientation parameters calculated by *CP-based EO Estimation*; and (ii) through the *Robust Global SfM* method, using as initial exterior orientation parameters the ones calculated via *Global SfM*. For both datasets, a significant improvement in the exterior orientation parameters computed through *Global SfM* relative to the ones computed through *CP-based EO Estimation* is observed. For instance, the average absolute $\Delta$*LinearEO* for the case of *Global SfM* (Table 7-6) is 4 times smaller than the one achieved through *CP-based EO Estimation* (Table 5-6) for *UAV Dataset 1*; for *UAV Dataset 2* this improvement of accuracy is greater, with the average absolute $\Delta$*LinearEO* being 6.8 times smaller for the case of *Global SfM* (Table 7-7) than the one achieved through *CP-based EO Estimation* (Table 5-7). Similar results are also observed for $\Delta$*AngularEO*; the latter one is 4 times smaller for *UAV Dataset 1* (Table 7-6) and 3.5 times smaller for *UAV Dataset 2* (Table 7-7) for the case of *Global SfM*, compared to the ones corresponding to *CP-based EO Estimation* (Table 5-6 and Table 5-7).

What is more, for both datasets an improvement in the exterior orientation parameters computed through *Robust Global SfM* relative to the ones computed through *Global SfM* is observed (Table 7-6 and Table 7-7). This is also expected, as the *Global SfM* process does not eliminate all outliers. Specifically, the average absolute $\Delta$*LinearEO* and $\Delta$*AngularEO* are reduced by 42% and 22% thanks to the robust iterative bundle adjustment techniques of *Robust Global SfM* for *UAV Dataset 1* (Table 7-6). The corresponding decrease for *UAV Dataset 2* reaches the percentages of 28% for $\Delta$*LinearEO* and 26% for $\Delta$*AngularEO* (Table 7-7). Finally, similarly to the aforementioned results, a higher absolute value of $\Delta$*LinearEO* corresponds to images that are far away from the image with GCP measurements (*img_1*).

Table 7-8. Differences between the EO parameters computed (i) through *Global SfM* using as initial EO parameters the ones estimated through *VP-based EO Estimation* (Table 6-4) and (ii) through *Robust Global SfM* using as initial EO parameters the ones estimated by *Global SfM* (using the SURF feature extractor, equal weights for observations in BA and four GCP measurements in *img_1*) and the reference ones for *UAV Dataset 1*

| ***UAV Dataset 1*: 23 oblique UAV images** | | | | | | |
|---|---|---|---|---|---|---|
| **Method** | **Metric** | **Avg-Abs** | **Max-Abs** | **Min-Abs** | **Avg** | **Stdev** |
| ***Global SfM:* initial EO from *VP-based EO Estimation* (SURF; unweighted BA; 4 GCPs in one image)** | $\Delta X_o$ (m) | 1.043 | 2.618 | 0.062 | 0.925 | 0.970 |
| | $\Delta Y_o$ (m) | 0.129 | 0.366 | 0.000 | -0.037 | 0.151 |
| | $\Delta Z_o$ (m) | 0.182 | 0.547 | 0.007 | 0.064 | 0.239 |
| | $\Delta\omega$ (deg.) | 0.659 | 1.223 | 0.060 | 0.659 | 0.281 |
| | $\Delta\varphi$ (deg.) | 0.729 | 2.228 | 0.043 | 0.466 | 0.747 |
| | $\Delta\kappa$ (deg.) | 0.168 | 0.432 | 0.013 | 0.144 | 0.159 |
| | $\Delta$*LinearEO* (m) | 0.451 | 2.618 | 0.000 | 0.317 | 0.453 |
| | $\Delta$*AngularEO* (deg.) | 0.519 | 2.228 | 0.013 | 0.423 | 0.396 |
| ***Robust Global SfM:* initial EO from *Global SfM* (SURF; unweighted BA; 4 GCPs in one image)** | $\Delta X_o$ (m) | 0.416 | 1.056 | 0.027 | 0.306 | 0.371 |
| | $\Delta Y_o$ (m) | 0.236 | 0.457 | 0.107 | 0.236 | 0.114 |
| | $\Delta Z_o$ (m) | 0.075 | 0.269 | 0.001 | -0.022 | 0.098 |
| | $\Delta\omega$ (deg.) | 0.496 | 0.761 | 0.123 | 0.496 | 0.206 |
| | $\Delta\varphi$ (deg.) | 0.556 | 0.905 | 0.022 | -0.115 | 0.586 |
| | $\Delta\kappa$ (deg.) | 0.168 | 0.367 | 0.018 | 0.062 | 0.202 |
| | $\Delta$*LinearEO* (m) | 0.242 | 1.056 | 0.001 | 0.173 | 0.195 |
| | $\Delta$*AngularEO* (deg.) | 0.407 | 0.905 | 0.018 | 0.148 | 0.331 |

### *VP_RobustGlobalSfM*

The third workflow for obtaining initial exterior orientation parameters for the *Robust Global SfM* process consists of the vanishing point detection method for exterior orientation estimation (*VP-based EO Estimation*), as presented in section 6.3, followed by a *Global SfM* method without iterative bundle adjustment procedures for rejection of the remaining erroneous tie point observations. The results of *VP-based EO Estimation* have been presented in Table 6-4 and Table 6-5 for *UAV Dataset 1* and *UAV Dataset 2*, respectively. Thus, Table 7-8 and Table 7-9 outline the results obtained (i) through the *Global SfM* method, using as initial exterior orientation parameters the ones calculated via *VP-based EO Estimation*; and (ii) through the *Robust Global SfM* method, using as initial exterior orientation parameters the ones calculated via *Global SfM*. For both datasets, a significant improvement in the exterior orientation parameters estimated through *Global SfM* relative to the ones estimated by *VP-based EO Estimation* is observed. The average absolute $\Delta$*LinearEO* and $\Delta$*AngularEO* for *Global SfM* are 5.1 times and 7.2 times smaller, respectively, than the ones achieved through *VP-based EO Estimation* for *UAV Dataset 1*; for *UAV Dataset 2*, the average absolute $\Delta$*LinearEO* and $\Delta$*AngularEO* are 1.9 times and 3.2 times smaller, respectively, than the ones achieved through *VP-based EO Estimation*.

Furthermore, as expected, the exterior orientation parameters computed through *Robust Global SfM* for both datasets are improved, relative to the ones computed through *Global SfM* initialized by *VP-based EO Estimation*. The average absolute $\Delta$*LinearEO* and $\Delta$*AngularEO* are reduced by 46% and 22% thanks to the robust iterative bundle adjustment techniques of *Robust Global SfM* for *UAV Dataset 1*. The corresponding decrease of $\Delta$*LinearEO* and $\Delta$*AngularEO* for *UAV Dataset 2* reaches

Table 7-9. Differences between the EO parameters computed (i) through *Global SfM* using as initial EO parameters the ones estimated through *VP-based EO Estimation* (Table 6-5) and (ii) through *Robust Global SfM* using as initial EO parameters the ones estimated by *Global SfM* (using the SURF feature extractor, equal weights for observations in BA and four GCP measurements in *img_1*) and the reference ones for *UAV Dataset 2*

| ***UAV Dataset 2*: 35 oblique UAV images** | | | | | | |
|---|---|---|---|---|---|---|
| **Method** | **Metric** | **Avg-Abs** | **Max-Abs** | **Min-Abs** | **Avg** | **Stdev** |
| ***Global SfM:* initial EO from *VP-based EO Estimation* (SURF; unweighted BA; 4 GCPs in one image)** | $\Delta X_o$ (m) | 2.461 | 4.956 | 0.128 | -0.144 | 2.926 |
| | $\Delta Y_o$ (m) | 0.562 | 1.695 | 0.025 | -0.514 | 0.532 |
| | $\Delta Z_o$ (m) | 0.524 | 1.707 | 0.011 | 0.474 | 0.560 |
| | $\Delta\omega$ (deg.) | 0.736 | 2.193 | 0.160 | 0.727 | 0.328 |
| | $\Delta\varphi$ (deg.) | 2.553 | 6.881 | 0.032 | -0.998 | 3.262 |
| | $\Delta\kappa$ (deg.) | 0.182 | 0.588 | 0.004 | 0.004 | 0.226 |
| | $\Delta$*LinearEO* (m) | 1.182 | 4.956 | 0.011 | -0.061 | 1.340 |
| | $\Delta$*AngularEO* (deg.) | 1.157 | 6.881 | 0.004 | -0.089 | 1.272 |
| ***Robust Global SfM:* initial EO from *Global SfM* (SURF; unweighted BA; 4 GCPs in one image)** | $\Delta X_o$ (m) | 0.321 | 0.522 | 0.001 | -0.141 | 0.321 |
| | $\Delta Y_o$ (m) | 0.266 | 0.537 | 0.008 | 0.202 | 0.234 |
| | $\Delta Z_o$ (m) | 0.147 | 0.394 | 0.000 | 0.013 | 0.184 |
| | $\Delta\omega$ (deg.) | 0.540 | 0.742 | 0.224 | 0.540 | 0.150 |
| | $\Delta\varphi$ (deg.) | 0.571 | 1.106 | 0.099 | -0.571 | 0.281 |
| | $\Delta\kappa$ (deg.) | 0.112 | 0.177 | 0.002 | 0.047 | 0.113 |
| | $\Delta$*LinearEO* (m) | 0.245 | 0.537 | 0.000 | 0.025 | 0.246 |
| | $\Delta$*AngularEO* (deg.) | 0.408 | 1.106 | 0.002 | 0.005 | 0.181 |

the percentages of 79% and 64%, respectively. Finally, similarly to the aforementioned results, a greater absolute value of $\Delta$*LinearEO* corresponds to images that are far away from *img_1*.

### 7.10.4.3 Comparison and Discussion

In sections 7.10.4.1 and 7.10.4.2, four exterior orientation estimation workflows have been tested using the oblique images of *UAV Dataset 1* and *UAV Dataset 2*. The three out of the four workflows yield similar results. Specifically, the *Robust Incremental SfM* method yields similar results with *IncrSfM_RobustGlobalSfM* and *CP_RobustGlobalSfM*. Among these workflows, the smallest average absolute $\Delta$*LinearEO* corresponds to *IncrSfM_RobustGlobalSfM* for *UAV Dataset 1* and *CP_RobustGlobalSfM* for *UAV Dataset 2*; the smallest average absolute $\Delta$*AngularEO* corresponds to *IncrSfM_RobustGlobalSfM* for both datasets. Between the *Robust Incremental SfM* method and the *Robust Global SfM* method that does not exploit the scene geometry for computation of initial exterior orientation parameters (*IncrSfM_RobustGlobalSfM),* the *Robust Global SfM* method yields slightly improved results. However, these differences are insignificant and the results achieved are almost equivalent.

The *VP_RobustGlobalSfM* method yields the results that correspond to the biggest differences from the reference ones. Thus, among the geometry-based methods for computing approximate exterior orientation parameters, the coplanar point detection method is more suitable than the vanishing point detection technique. The fact that the *VP-based EO Estimation* method yields worse approximate exterior orientation results than the *CP-based EO Estimation* method is due to the fact that it uses the coplanar point detection method for the scope of image georeferencing, prior to the detection of vanishing points. As a result, the errors arising from the coplanar point detection

Table 7-10. Differences between the EO parameters computed (i) through *Incremental SfM* and (ii) through *Robust Global SfM* using as initial EO parameters the ones estimated by *Incremental SfM* (using the SURF feature extractor, equal weights for observations in BA and four GCP measurements in *img_1*) and the reference ones for *UAV Dataset 3*

| ***UAV Dataset 3*: 60 oblique UAV images** | | | | | | |
|---|---|---|---|---|---|---|
| **Method** | **Metric** | **Avg-Abs** | **Max-Abs** | **Min-Abs** | **Avg** | **Stdev** |
| ***Incremental SfM* (SURF; unweighted BA; 4 GCPs in one image)** | $\Delta X_o$ (m) | 0.867 | 1.373 | 0.016 | 0.826 | 0.489 |
| | $\Delta Y_o$ (m) | 0.145 | 0.754 | 0.002 | -0.023 | 0.187 |
| | $\Delta Z_o$ (m) | 0.328 | 1.121 | 0.004 | 0.274 | 0.392 |
| | $\Delta\omega$ (deg.) | 0.690 | 1.134 | 0.343 | 0.071 | 0.711 |
| | $\Delta\varphi$ (deg.) | 0.331 | 0.699 | 0.006 | 0.098 | 0.353 |
| | $\Delta\kappa$ (deg.) | 0.229 | 1.279 | 0.005 | 0.214 | 0.224 |
| | $\Delta$*LinearEO* (m) | 0.447 | 1.373 | 0.002 | 0.359 | 0.356 |
| | $\Delta$*AngularEO* (deg.) | 0.417 | 1.279 | 0.005 | 0.128 | 0.429 |
| ***Robust Global SfM:* initial EO from *Incremental SfM* (SURF; unweighted BA; 4 GCPs in one image)** | $\Delta X_o$ (m) | 0.657 | 1.084 | 0.005 | 0.590 | 0.451 |
| | $\Delta Y_o$ (m) | 0.250 | 0.497 | 0.006 | 0.221 | 0.157 |
| | $\Delta Z_o$ (m) | 0.177 | 0.570 | 0.001 | 0.095 | 0.219 |
| | $\Delta\omega$ (deg.) | 0.518 | 0.950 | 0.013 | -0.009 | 0.566 |
| | $\Delta\varphi$ (deg.) | 0.214 | 0.532 | 0.003 | -0.136 | 0.227 |
| | $\Delta\kappa$ (deg.) | 0.189 | 0.597 | 0.003 | 0.154 | 0.191 |
| | $\Delta$*LinearEO* (m) | 0.361 | 1.084 | 0.001 | 0.302 | 0.276 |
| | $\Delta$*AngularEO* (deg.) | 0.307 | 0.950 | 0.003 | 0.003 | 0.328 |

Table 7-11. Differences between the EO parameters computed (i) through *Incremental SfM* and (ii) through *Robust Global SfM* using as initial EO parameters the ones estimated by *Incremental SfM* (using the SURF feature extractor, equal weights for observations in BA and four GCP measurements in *img_1*) and the reference ones for *UAV Dataset 4*

| ***UAV Dataset 4*: 80 oblique UAV images** | | | | | | |
|---|---|---|---|---|---|---|
| **Method** | **Metric** | **Avg-Abs** | **Max-Abs** | **Min-Abs** | **Avg** | **Stdev** |
| ***Incremental SfM* (SURF; unweighted BA; 4 GCPs in one image)** | $\Delta X_o$ (m) | 0.610 | 1.060 | 0.002 | 0.561 | 0.401 |
| | $\Delta Y_o$ (m) | 0.521 | 2.036 | 0.012 | -0.331 | 0.780 |
| | $\Delta Z_o$ (m) | 0.219 | 0.905 | 0.000 | 0.133 | 0.312 |
| | $\Delta\omega$ (deg.) | 0.543 | 1.295 | 0.028 | 0.041 | 0.604 |
| | $\Delta\varphi$ (deg.) | 0.268 | 0.714 | 0.008 | -0.033 | 0.302 |
| | $\Delta\kappa$ (deg.) | 0.446 | 1.890 | 0.002 | 0.429 | 0.460 |
| | $\Delta$*LinearEO* (m) | 0.450 | 2.036 | 0.000 | 0.121 | 0.498 |
| | $\Delta$*AngularEO* (deg.) | 0.419 | 1.890 | 0.002 | 0.146 | 0.455 |
| ***Robust Global SfM:* initial EO from *Incremental SfM* (SURF; unweighted BA; 4 GCPs in one image)** | $\Delta X_o$ (m) | 0.611 | 1.048 | 0.011 | 0.559 | 0.407 |
| | $\Delta Y_o$ (m) | 0.579 | 1.873 | 0.018 | -0.246 | 0.773 |
| | $\Delta Z_o$ (m) | 0.177 | 0.415 | 0.002 | -0.042 | 0.209 |
| | $\Delta\omega$ (deg.) | 0.376 | 0.849 | 0.006 | 0.039 | 0.441 |
| | $\Delta\varphi$ (deg.) | 0.238 | 0.524 | 0.015 | -0.236 | 0.159 |
| | $\Delta\kappa$ (deg.) | 0.398 | 1.154 | 0.008 | 0.371 | 0.395 |
| | $\Delta$*LinearEO* (m) | 0.456 | 1.873 | 0.002 | 0.091 | 0.463 |
| | $\Delta$*AngularEO* (deg.) | 0.338 | 1.154 | 0.006 | 0.058 | 0.332 |

Table 7-12. Differences between the EO parameters computed (i) through *Incremental SfM* and (ii) through *Robust Global SfM* using as initial EO parameters the ones estimated by *Incremental SfM* (using the SURF feature extractor, equal weights for observations in BA and four GCP measurements in *img_1*) and the reference ones for *UAV Dataset 5*

| ***UAV Dataset 5*: 100 oblique UAV images** | | | | | | |
|---|---|---|---|---|---|---|
| **Method** | **Metric** | **Avg-Abs** | **Max-Abs** | **Min-Abs** | **Avg** | **Stdev** |
| ***Incremental SfM* (SURF; unweighted BA; 4 GCPs in one image)** | $\Delta X_o$ (m) | 0.507 | 1.062 | 0.013 | 0.405 | 0.455 |
| | $\Delta Y_o$ (m) | 0.694 | 1.919 | 0.011 | -0.550 | 0.806 |
| | $\Delta Z_o$ (m) | 0.235 | 0.903 | 0.000 | 0.082 | 0.330 |
| | $\Delta\omega$ (deg.) | 0.441 | 1.268 | 0.002 | 0.035 | 0.526 |
| | $\Delta\varphi$ (deg.) | 0.229 | 0.714 | 0.006 | -0.032 | 0.269 |
| | $\Delta\kappa$ (deg.) | 0.624 | 1.855 | 0.002 | 0.611 | 0.526 |
| | $\Delta$*LinearEO* (m) | 0.479 | 1.919 | 0.000 | -0.021 | 0.530 |
| | $\Delta$*AngularEO* (deg.) | 0.431 | 1.855 | 0.002 | 0.204 | 0.440 |
| ***Robust Global SfM:* initial EO from *Incremental SfM* (SURF; unweighted BA; 4 GCPs in one image)** | $\Delta X_o$ (m) | 0.496 | 1.050 | 0.004 | 0.406 | 0.449 |
| | $\Delta Y_o$ (m) | 0.753 | 1.824 | 0.007 | -0.491 | 0.835 |
| | $\Delta Z_o$ (m) | 0.193 | 0.679 | 0.002 | -0.083 | 0.234 |
| | $\Delta\omega$ (deg.) | 0.313 | 0.813 | 0.007 | 0.052 | 0.387 |
| | $\Delta\varphi$ (deg.) | 0.239 | 0.524 | 0.012 | -0.237 | 0.143 |
| | $\Delta\kappa$ (deg.) | 0.567 | 1.275 | 0.008 | 0.545 | 0.474 |
| | $\Delta$*LinearEO* (m) | 0.481 | 1.824 | 0.002 | -0.056 | 0.506 |
| | $\Delta$*AngularEO* (deg.) | 0.373 | 1.275 | 0.007 | 0.120 | 0.335 |

technique are included in the vanishing point detection method. Another conclusion that is derived from the aforementioned results is the fact that the exterior orientation parameters computed through a *Robust Global SfM* method depend on the approximate exterior orientation parameters used as input. Better approximate exterior orientation parameters result in more accurate exterior orientation parameters (e.g., see the results of *CP_RobustGlobalSfM* and *VP_RobustGlobalSfM*).

Last but not least, the automatic elimination of all erroneous tie point observations through the developed SfM methods that implement robust iterative outlier removal techniques significantly improves the estimated exterior orientation parameters. This conclusion holds both for *Robust Incremental SfM* and *Robust Global SfM*.

Taking into account the facts that:

- similar exterior orientation results are achieved through the *Robust Incremental SfM* method and the *Robust Global SfM* method initialized via (i) *Incremental SfM* and (ii) *CP-based EO Estimation* followed by *Global SfM* (*CP_RobustGlobalSfM*);
- the *CP-based EO Estimation* is only applied in datasets for which every image depicts a planar surface that lies on the same plane with the one where the GCPs lie;
- the *CP-based EO Estimation* is more prone to errors, due to the possibility of wrong estimation of the homography induced by the plane of interest between two images; and
- slightly better accuracy is achieved by *IncrSfM_RobustGlobalSfM* compared to *Robust Incremental SfM*,

the experiments performed hereinafter follow the *Robust Global SfM* workflow, initialized either through an *Incremental SfM* or through georeferencing metadata provided by onboard sensors.

### 7.10.5 Analysis of Results Using Robust Global SfM

In order to further assess the developed exterior orientation framework, more results using the *Robust Global SfM* method are outlined in this section. In all experiments that are presented in this section, the SURF algorithm is used for feature extraction; in addition, the bundle adjustment procedures assume equal weights for all point observations.

The results from *Robust Global SfM* using as approximate exterior orientation parameters the ones calculated by an *Incremental SfM* are presented for *UAV Dataset 3*, *UAV Dataset 4* and *UAV Dataset 5* in Table 7-10, Table 7-11 and Table 7-12, respectively. (The results for *UAV Dataset 1* and *UAV Dataset 2* have already been presented in Table 7-4 and Table 7-5, respectively, of section 7.10.4.2). The conclusions derived for *UAV Dataset 1* and *UAV Dataset 2* are also valid for *UAV*

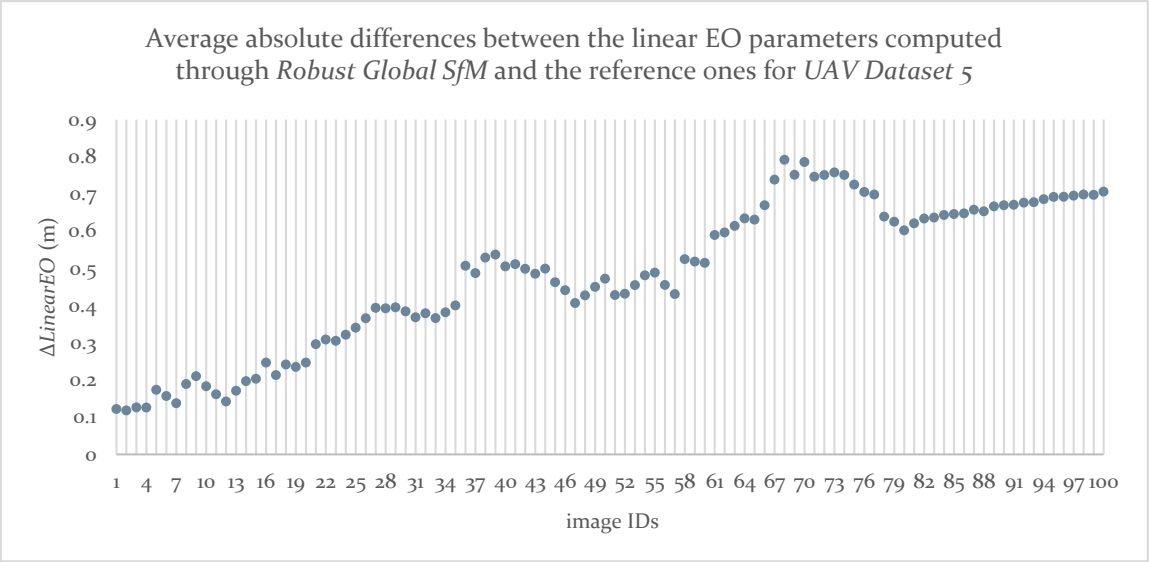


Figure 7-17. Average absolute differences between the linear EO parameters computed through *Robust Global SfM* initialized by *Incremental SfM* (using the SURF feature extractor, equal weights for observations in BA and four GCP measurements in *img_1*) and the reference ones for the images *of UAV Dataset 5*; *img_1* is the one with image ID = 1

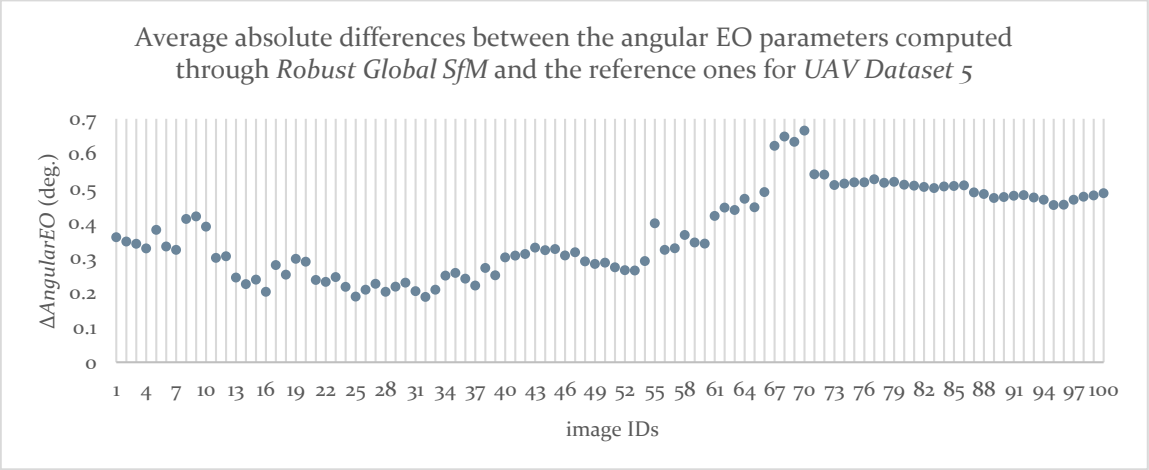


Figure 7-18. Average absolute differences between the angular EO parameters computed through *Robust Global SfM* initialized by *Incremental SfM* (using the SURF feature extractor, equal weights for observations in BA and four GCP measurements in *img_1*) and the reference ones for the images of *UAV Dataset 5*; *img_1* is the one with image ID = 1

Table 7-13. Differences between the EO parameters computed through *Robust Global SfM*, which are the same with the ones achieved through *Global SfM,* using as initial EO parameters the ones provided by georeferencing data (using the SURF feature extractor, equal weights for observations in BA and four GCP measurements in *img_1*) for *RCD30 Dataset 1*

| ***RCD30 Dataset 1*: 10 single-perspective oblique aerial images** | | | | | | |
|---|---|---|---|---|---|---|
| **Method** | **Metric** | **Avg-Abs** | **Max-Abs** | **Min-Abs** | **Avg** | **Stdev** |
| ***Global SfM* & *Robust Global SfM;* initial EO from geo-referencing data (SURF; unweighted BA; 4 GCPs in one image)** | $\Delta X_o$ (m) | 0.197 | 0.452 | 0.031 | 0.112 | 0.209 |
| | $\Delta Y_o$ (m) | 0.329 | 0.887 | 0.066 | -0.021 | 0.414 |
| | $\Delta Z_o$ (m) | 0.326 | 0.568 | 0.057 | -0.111 | 0.348 |
| | $\Delta\omega$ (deg.) | 0.017 | 0.044 | 0.000 | -0.005 | 0.022 |
| | $\Delta\varphi$ (deg.) | 0.016 | 0.038 | 0.000 | 0.015 | 0.014 |
| | $\Delta\kappa$ (deg.) | 0.009 | 0.022 | 0.001 | -0.003 | 0.010 |
| | $\Delta$*LinearEO* (m) | 0.284 | 0.887 | 0.031 | -0.007 | 0.324 |
| | $\Delta$*AngularEO* (deg.) | 0.014 | 0.044 | 0.000 | 0.002 | 0.015 |

*Dataset 3*, *UAV Dataset 4* and *UAV Dataset 5*. Specifically, as expected, an improvement in the exterior orientation parameters computed through *Robust Global SfM* relative to the ones computed through *Incremental SfM* is observed. This improvement is greater for the angular exterior orientation parameters, resulting in a reduction of the average absolute $\Delta$*AngularEO* by 26%, 19% and 13% for *UAV Dataset 3*, *UAV Dataset 4* and *UAV Dataset 5*, respectively. Whereas for *UAV Dataset 4* and *UAV Dataset 5* the average absolute $\Delta$*LinearEO* remains practically unchanged by the implementation of robust iterative methods for outlier removal through *Robust Global SfM*, the average absolute $\Delta$*LinearEO* for *UAV Dataset 3* is reduced by 19%. Furthermore, the value of absolute $\Delta$*LinearEO* for an image generally increases with the distance of the image from *img_1*, as it is observed for *UAV Dataset 1* and *UAV Dataset 2*, while the absolute $\Delta$*AngularEO* does not generally increase with the distance from *img_1*. Figure 7-17 and Figure 7-18 illustrate these conclusions for *UAV Dataset 5*.

The *Robust Global SfM* method has also been tested using *RCD30 Dataset 1*, *RCD30 Dataset 2* and *RCD30 Dataset 3*. The GNSS information that accompanies the imagery along with rough estimates of the camera orientation angles serve as initial values for the exterior orientation parameters of the imagery. In order to investigate the impact of the robust iterative outlier removal technique on the estimated exterior orientation parameters, a *Global SfM* without any iterative outlier removal techniques has also been applied to these three datasets, using the same approximate exterior orientation parameters as the ones used by *Robust Global SfM*. The results for both *Global SfM* and *Robust Global SfM* are illustrated in Table 7-13 for *RCD30 Dataset 1*, Table 7-14 for *RCD30 Dataset 2* and Table 7-15 for *RCD30 Dataset 3*. For the dataset of 10 single perspective oblique aerial images (*RCD30 Dataset 1*), all erroneous feature point correspondences have been eliminated through the tests and constraints imposed by the image matching procedure; thus, the iterative framework of bundle adjustment methods for outlier removal has not been implemented for *RCD30 Dataset 1*. For the other two datasets, the results have been improved thanks to the iterative framework for outlier removal. Specifically, the average absolute $\Delta$*LinearEO* has been reduced by 6% and 9% thanks to the use of *Robust Global SfM* for *RCD30 Dataset 2* and *RCD30 Dataset 3*, respectively. This reduction of average absolute $\Delta$*AngularEO* corresponds to a percentage of 19% for *RCD30 Dataset 2* and 11% for *RCD30 Dataset 3*.

The results of both *Global SfM* and *Robust Global SfM* demonstrate a better accuracy in the exterior orientation parameters for the dataset of 10 single-perspective oblique images (*RCD30 Dataset 1*).

Table 7-14. Differences between the EO parameters computed (i) through *Global SfM* and (ii) through *Robust Global SfM*, using as initial EO parameters the ones provided by georeferencing data (using the SURF feature extractor, equal weights for observations in BA and four GCP measurements in *img_1*) for *RCD30 Dataset 2*

| ***RCD30 Dataset 2*: 40 multi-perspective oblique aerial images** | | | | | | |
|---|---|---|---|---|---|---|
| **Method** | **Metric** | **Avg-Abs** | **Max-Abs** | **Min-Abs** | **Avg** | **Stdev** |
| ***Global SfM;* initial EO from georeferencing data (SURF; unweighted BA; 4 GCPs in one image)** | $\Delta X_o$ (m) | 0.322 | 0.665 | 0.029 | 0.049 | 0.361 |
| | $\Delta Y_o$ (m) | 0.320 | 1.106 | 0.014 | 0.047 | 0.414 |
| | $\Delta Z_o$ (m) | 0.484 | 1.531 | 0.008 | -0.106 | 0.637 |
| | $\Delta\omega$ (deg.) | 0.029 | 0.119 | 0.002 | -0.009 | 0.036 |
| | $\Delta\varphi$ (deg.) | 0.047 | 0.111 | 0.002 | 0.012 | 0.054 |
| | $\Delta\kappa$ (deg.) | 0.050 | 0.082 | 0.016 | 0.006 | 0.051 |
| | $\Delta$*LinearEO* (m) | 0.375 | 1.531 | 0.008 | -0.003 | 0.471 |
| | $\Delta$*AngularEO* (deg.) | 0.042 | 0.119 | 0.002 | 0.003 | 0.047 |
| ***Robust Global SfM:* initial EO from georeferencing data (SURF; unweighted BA; 4 GCPs in one image)** | $\Delta X_o$ (m) | 0.296 | 0.739 | 0.016 | 0.028 | 0.350 |
| | $\Delta Y_o$ (m) | 0.331 | 1.204 | 0.005 | 0.022 | 0.419 |
| | $\Delta Z_o$ (m) | 0.430 | 1.252 | 0.033 | -0.097 | 0.556 |
| | $\Delta\omega$ (deg.) | 0.023 | 0.091 | 0.001 | -0.006 | 0.029 |
| | $\Delta\varphi$ (deg.) | 0.036 | 0.109 | 0.004 | 0.009 | 0.045 |
| | $\Delta\kappa$ (deg.) | 0.042 | 0.076 | 0.012 | 0.006 | 0.044 |
| | $\Delta$*LinearEO* (m) | 0.352 | 1.252 | 0.005 | -0.016 | 0.442 |
| | $\Delta$*AngularEO* (deg.) | 0.034 | 0.109 | 0.001 | 0.003 | 0.039 |

Table 7-15. Differences between the EO parameters computed (i) through *Global SfM* and (ii) through *Robust Global SfM*, using as initial EO parameters the ones provided by georeferencing data (using the SURF feature extractor, equal weights for observations in BA and four GCP measurements in *img_1*) for *RCD30 Dataset 3*

| ***RCD30 Dataset 3*: 50 multi-perspective oblique and vertical aerial images** | | | | | | |
|---|---|---|---|---|---|---|
| **Method** | **Metric** | **Avg-Abs** | **Max-Abs** | **Min-Abs** | **Avg** | **Stdev** |
| ***Global SfM;* initial EO from georeferencing data (SURF; unweighted BA; 4 GCPs in one image)** | $\Delta X_o$ (m) | 0.335 | 1.132 | 0.013 | 0.112 | 0.410 |
| | $\Delta Y_o$ (m) | 0.243 | 0.828 | 0.000 | 0.068 | 0.290 |
| | $\Delta Z_o$ (m) | 0.340 | 1.327 | 0.001 | -0.121 | 0.457 |
| | $\Delta\omega$ (deg.) | 0.029 | 0.142 | 0.001 | -0.010 | 0.037 |
| | $\Delta\varphi$ (deg.) | 0.032 | 0.123 | 0.003 | 0.017 | 0.039 |
| | $\Delta\kappa$ (deg.) | 0.052 | 0.098 | 0.001 | 0.010 | 0.057 |
| | $\Delta$*LinearEO* (m) | 0.306 | 1.327 | 0.000 | 0.020 | 0.386 |
| | $\Delta$*AngularEO* (deg.) | 0.038 | 0.142 | 0.001 | 0.006 | 0.044 |
| ***Robust Global SfM:* initial EO from georeferencing data (SURF; unweighted BA; 4 GCPs in one image)** | $\Delta X_o$ (m) | 0.273 | 0.785 | 0.007 | 0.102 | 0.304 |
| | $\Delta Y_o$ (m) | 0.240 | 1.035 | 0.001 | 0.079 | 0.291 |
| | $\Delta Z_o$ (m) | 0.323 | 1.100 | 0.004 | -0.112 | 0.402 |
| | $\Delta\omega$ (deg.) | 0.025 | 0.110 | 0.001 | -0.010 | 0.031 |
| | $\Delta\varphi$ (deg.) | 0.027 | 0.095 | 0.001 | 0.016 | 0.028 |
| | $\Delta\kappa$ (deg.) | 0.051 | 0.105 | 0.001 | 0.009 | 0.056 |
| | $\Delta$*LinearEO* (m) | 0.279 | 1.100 | 0.001 | 0.023 | 0.333 |
| | $\Delta$*AngularEO* (deg.) | 0.034 | 0.110 | 0.001 | 0.005 | 0.038 |

The largest absolute differences from the reference data are observed for the dataset of 40 multi-perspective oblique aerial images (*RCD30 Dataset 2*), as there is very poor overlap between the

four oblique aerial images acquired by the multi-view camera system at a single time instance. The dataset of 50 multi-perspective oblique and vertical aerial images (*RCD30 Dataset 3*) corresponds to better accuracy in the exterior orientation parameters than the oblique-only multi-perspective dataset (*RCD30 Dataset 2*), as the five images that are acquired by the multi-camera system provide a sufficiently stronger geometry to tie the side oblique aerial images of the strip together. Furthermore, the average differences between the exterior orientation parameters computed through *Robust Global SfM* and the reference ones are close to zero (corresponding to an average value of less than 12 cm for the linear exterior orientation parameters and less than 0.017 deg. for the angular exterior orientation parameters), demonstrating that the aerial triangulation solution has insignificant systematic errors.

The diagram illustrated in Figure 7-19 shows the overall improvement of the exterior orientation parameters thanks to the developed robust iterative bundle adjustment framework that is incorporated in the *Robust Global SfM* method for all test datasets. The non-robust SfM methods that are used for comparison with *Robust Global SfM* are the *Incremental SfM* method for *UAV Dataset 1 – UAV Dataset 5* and the *Global SfM* method for *RCD30 Dataset 1 – RCD30 Dataset 3*. Except for *RCD30 Dataset 1*, for which all outliers have been eliminated through the checks and constraints imposed by the image matching procedure, for all the other datasets the combined comparison of Δ*LinearEO* and Δ*AngularEO* obtained through a robust and a non-robust SfM method proves significant improvement of the exterior orientation results thanks to the elimination of all erroneous tie points by a robust SfM method.

### 7.10.6 Impact of the Feature Extraction Algorithm

The experiments presented in sections 7.10.4 and 7.10.5 were conducted using the SURF algorithm for feature extraction. In order to assess the impact of the feature extraction algorithm on the exterior orientation results, experiments using the ASIFT feature detector for *UAV Dataset 3* as well as for *RCD30 Dataset 1 – RCD30 Dataset 3* have also been performed. Whereas the use of ASIFT is also tested for single-perspective oblique images (*UAV Dataset 3* and *RCD30 Dataset 1*), the goal of

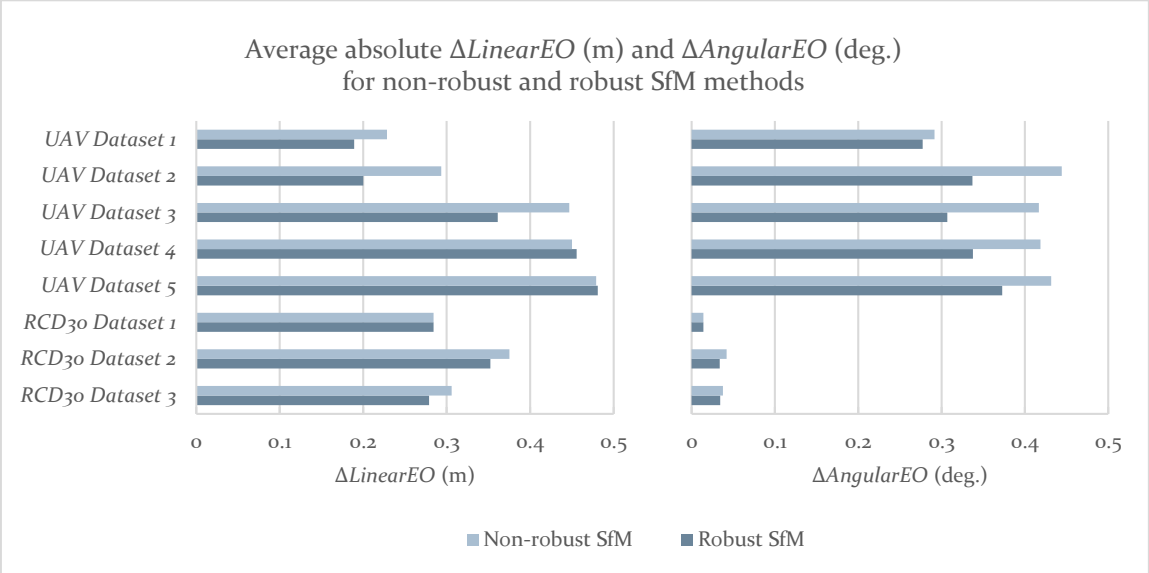


Figure 7-19. Average absolute Δ*LinearEO* (m) and Δ*AngularEO* (deg.) for non-robust SfM methods (i.e., *Incremental SfM* for *UAV Dataset 1 – UAV Dataset 5* and *Global SfM* for *RCD30 Dataset 1 – RCD30 Dataset 3*) and *Robust Global SfM* (using the SURF feature extractor, equal weights for observations in BA and four GCP measurements in *img_1*) for all test datasets

Table 7-16. Differences between the EO parameters computed (i) through *Incremental SfM* and (ii) through *Robust Global SfM* using as initial EO parameters the ones estimated by *Incremental SfM* (using the ASIFT feature extractor, equal weights for observations in BA and four GCP measurements in *img_1*) and the reference ones for *UAV Dataset 3*

| ***UAV Dataset 3*: 60 oblique UAV images** | | | | | | |
|---|---|---|---|---|---|---|
| **Method** | **Metric** | **Avg-Abs** | **Max-Abs** | **Min-Abs** | **Avg** | **Stdev** |
| ***Incremental SfM* (ASIFT; unweighted BA; 4 GCPs in one image)** | $\Delta X_o$ (m) | 0.410 | 1.161 | 0.020 | 0.096 | 0.485 |
| | $\Delta Y_o$ (m) | 0.313 | 1.112 | 0.002 | -0.107 | 0.432 |
| | $\Delta Z_o$ (m) | 0.577 | 1.973 | 0.002 | 0.473 | 0.721 |
| | $\Delta\omega$ (deg.) | 1.173 | 4.238 | 0.011 | 1.163 | 1.186 |
| | $\Delta\varphi$ (deg.) | 0.393 | 1.615 | 0.005 | 0.009 | 0.472 |
| | $\Delta\kappa$ (deg.) | 1.141 | 4.749 | 0.001 | 1.108 | 1.610 |
| | $\Delta$*LinearEO* (m) | 0.433 | 1.973 | 0.002 | 0.154 | 0.546 |
| | $\Delta$*AngularEO* (deg.) | 0.902 | 4.749 | 0.001 | 0.760 | 1.089 |
| ***Robust Global SfM:* initial EO from *Incremental SfM* (ASIFT; unweighted BA; 4 GCPs in one image)** | $\Delta X_o$ (m) | 0.501 | 0.914 | 0.003 | 0.501 | 0.355 |
| | $\Delta Y_o$ (m) | 0.223 | 0.721 | 0.000 | 0.223 | 0.229 |
| | $\Delta Z_o$ (m) | 0.366 | 0.831 | 0.003 | 0.366 | 0.414 |
| | $\Delta\omega$ (deg.) | 0.646 | 1.809 | 0.027 | 0.646 | 0.627 |
| | $\Delta\varphi$ (deg.) | 0.436 | 1.676 | 0.041 | 0.436 | 0.506 |
| | $\Delta\kappa$ (deg.) | 0.349 | 1.886 | 0.001 | 0.349 | 0.460 |
| | $\Delta$*LinearEO* (m) | 0.363 | 0.914 | 0.000 | 0.363 | 0.332 |
| | $\Delta$*AngularEO* (deg.) | 0.477 | 1.886 | 0.001 | 0.477 | 0.531 |

this research was mainly to test the ASIFT feature detector in oblique aerial images of different perspective, i.e., for *RCD30 D ataset 2* and *RCD30 Dataset 3*, and compare it with SURF, which is mainly used for the experiments in this dissertation, as ASIFT enables matching of images with a severe viewpoint change and has been proved to be robust to affine distortions (Morel and Yu, 2009).

Table 7-16 presents the results obtained for *UAV Dataset 3* via *Incremental SfM* and via *Robust Global SfM* using initial exterior orientation parameters from *Incremental SfM*, using the ASIFT feature detector. A small improvement (reduction by 3%) of average absolute $\Delta$*LinearEO* for *Incremental SfM* is observed for the results obtained using ASIFT, compared to those ones obtained using SURF (Table 7-10). The value of average absolute $\Delta$*LinearEO* obtained using SURF and ASIFT is approximately the same for the case of *Robust Global SfM* (increase by 0.5% for ASIFT). However, the angular exterior orientation parameters obtained both by *Incremental SfM* and *Robust Global SfM* are of higher accuracy for the case of SURF. Specifically, an approximately double value of average absolute $\Delta$*AngularEO* for *Incremental SfM* and an increase of average absolute $\Delta$*AngularEO* by 55% for *Robust Global SfM* correspond to the results of ASIFT compared to those ones obtained using SURF. Finally, as expected, the ASIFT results achieved through *Robust Global SfM* are improved compared to those ones derived by *Incremental SfM* (reduction of average absolute $\Delta$*LinearEO* and $\Delta$*AngularEO* by 16% and 47% respectively).

Whereas more correspondences are generally detected by the ASIFT algorithm in comparison with those extracted by SURF (also depending on the defined thresholds), both SURF and ASIFT feature detectors result in some erroneous correspondences, which are not detected by the tests and constraints imposed by the image matching procedure. For instance, Figure 7-20 and Figure 7-21 illustrate the correspondences detected using SURF and ASIFT, respectively, for an image pair of

*UAV Dataset 3*. The bottom images of these figures present magnifications of the matching results between the pair of images depicted at the top image; the tie points included in the blue and red ellipses are not correct. These erroneous matches mainly lie on glass surfaces of windows that incorporate some repetitive patterns, as illustrated in Figure 7-20 and Figure 7-21. Thus, the developed robust iterative bundle adjustment methods that reject these remaining erroneous matches are indispensable for both feature extractors.

Table 7-17, Table 7-18 and Table 7-19 present the results obtained via *Global SfM* and *Robust Global SfM* using approximate exterior orientation parameters from the georeferencing data, using the ASIFT feature detector for *RCD30 Dataset 1*, *RCD30 Dataset 2* and *RCD30 Dataset 3*, respectively. Similarly to the application of SURF to *RCD30 Dataset 1*, all erroneous ASIFT feature point correspondences have been eliminated through the tests and constraints imposed by the image matching procedure for *RCD30 Dataset 1*, which is the dataset of 10 single perspective oblique aerial images. For the other two datasets, the results have been improved thanks to the iterative outlier removal framework of *Robust Global SfM*. Specifically, the average absolute $\Delta LinearEO$ has been reduced by 5% and 15% thanks to the use of *Robust Global SfM* for *RCD30 Dataset 2* and *RCD30 Dataset 3*, respectively. This reduction of average absolute $\Delta AngularEO$ corresponds to a percentage of 13% for *RCD30 Dataset 2* and 17% for *RCD30 Dataset 3*.

Figure 7-20. Image matching results between a pair of UAV images using SURF; top: detected correspondences; bottom: details illustrating remaining outliers, marked with red and blue ellipses

Figure 7-21. Image matching results between a pair of UAV images using ASIFT; top: detected correspondences; bottom: details illustrating remaining outliers, marked with red and blue ellipses

Table 7-17. Differences between the EO parameters computed through *Robust Global SfM*, which are the same with the ones achieved through *Global SfM,* using as initial EO parameters the ones provided by georeferencing data (using the ASIFT feature extractor, equal weights for observations in BA and four GCP measurements in *img_1*) for *RCD30 Dataset 1*

| ***RCD30 Dataset 1*: 10 single-perspective oblique aerial images** | | | | | | |
|---|---|---|---|---|---|---|
| **Method** | **Metric** | **Avg-Abs** | **Max-Abs** | **Min-Abs** | **Avg** | **Stdev** |
| ***Global SfM* & *Robust Global SfM;* initial EO from geo-referencing data (ASIFT; unweighted BA; 4 GCPs in one image)** | $\Delta X_o$ (m) | 0.182 | 0.365 | 0.003 | 0.114 | 0.183 |
| | $\Delta Y_o$ (m) | 0.224 | 0.542 | 0.027 | 0.008 | 0.286 |
| | $\Delta Z_o$ (m) | 0.440 | 0.760 | 0.083 | -0.059 | 0.499 |
| | $\Delta\omega$ (deg.) | 0.021 | 0.049 | 0.005 | -0.005 | 0.025 |
| | $\Delta\varphi$ (deg.) | 0.018 | 0.034 | 0.003 | 0.015 | 0.014 |
| | $\Delta\kappa$ (deg.) | 0.011 | 0.030 | 0.001 | 0.001 | 0.014 |
| | $\Delta$*LinearEO* (m) | 0.282 | 0.760 | 0.003 | 0.021 | 0.323 |
| | $\Delta$*AngularEO* (deg.) | 0.016 | 0.049 | 0.001 | 0.004 | 0.018 |

Table 7-18. Differences between the EO parameters computed (i) through *Global SfM* and (ii) through *Robust Global SfM*, using as initial EO parameters the ones provided by georeferencing data (using the ASIFT feature extractor, equal weights for observations in BA and four GCP measurements in *img_1*) for *RCD30 Dataset 2*

| ***RCD30 Dataset 2*: 40 multi-perspective oblique aerial images** | | | | | | |
|---|---|---|---|---|---|---|
| **Method** | **Metric** | **Avg-Abs** | **Max-Abs** | **Min-Abs** | **Avg** | **Stdev** |
| ***Global SfM;* initial EO from georeferencing data (ASIFT; unweighted BA; 4 GCPs in one image)** | $\Delta X_o$ (m) | 0.356 | 1.247 | 0.006 | 0.122 | 0.478 |
| | $\Delta Y_o$ (m) | 0.432 | 2.227 | 0.019 | 0.106 | 0.591 |
| | $\Delta Z_o$ (m) | 0.444 | 1.960 | 0.002 | -0.071 | 0.615 |
| | $\Delta\omega$ (deg.) | 0.062 | 0.331 | 0.000 | -0.010 | 0.087 |
| | $\Delta\varphi$ (deg.) | 0.042 | 0.144 | 0.001 | 0.016 | 0.050 |
| | $\Delta\kappa$ (deg.) | 0.052 | 0.110 | 0.010 | 0.003 | 0.056 |
| | $\Delta$*LinearEO* (m) | 0.411 | 2.227 | 0.002 | 0.052 | 0.561 |
| | $\Delta$*AngularEO* (deg.) | 0.052 | 0.331 | 0.000 | 0.003 | 0.065 |
| ***Robust Global SfM:* initial EO from georeferencing data (ASIFT; unweighted BA; 4 GCPs in one image)** | $\Delta X_o$ (m) | 0.343 | 1.325 | 0.013 | 0.112 | 0.439 |
| | $\Delta Y_o$ (m) | 0.329 | 1.891 | 0.003 | 0.077 | 0.463 |
| | $\Delta Z_o$ (m) | 0.497 | 1.782 | 0.004 | -0.089 | 0.637 |
| | $\Delta\omega$ (deg.) | 0.046 | 0.283 | 0.002 | -0.007 | 0.066 |
| | $\Delta\varphi$ (deg.) | 0.038 | 0.131 | 0.001 | 0.017 | 0.045 |
| | $\Delta\kappa$ (deg.) | 0.051 | 0.107 | 0.011 | 0.004 | 0.055 |
| | $\Delta$*LinearEO* (m) | 0.390 | 1.891 | 0.003 | 0.033 | 0.513 |
| | $\Delta$*AngularEO* (deg.) | 0.045 | 0.283 | 0.001 | 0.005 | 0.055 |

The exterior orientation results of *RCD30 Dataset 2* computed through *Robust Global SfM* using the SURF (Table 7-13) and ASIFT feature detector (Table 7-17) are similar. However, the exterior orientation parameters of *RCD30 Dataset 2* and *RCD30 Dataset 3*, estimated by *Global SfM* and *Robust Global SfM* using the SURF feature detector (Table 7-14 and Table 7-15, respectively), are generally of higher accuracy than the ones obtained using ASIFT (Table 7-18 and Table 7-19, respectively). Specifically, an increase of average absolute $\Delta$*LinearEO* by 10% and 25% for *Global SfM* as well as 11% and 17% for *Robust Global SfM*, for *RCD30 Dataset 2* and *RCD30 Dataset 3*, respectively, are observed. The increase of average absolute $\Delta$*AngularEO* corresponds to a

Table 7-19. Differences between the EO parameters computed (i) through *Global SfM* and (ii) through *Robust Global SfM*, using as initial EO parameters the ones provided by georeferencing data (using the ASIFT feature extractor, equal weights for observations in BA and four GCP measurements in *img_1*) for *RCD30 Dataset 3*

| ***RCD30 Dataset 3*: 50 multi-perspective oblique and vertical aerial images** | | | | | | |
|---|---|---|---|---|---|---|
| **Method** | **Metric** | **Avg-Abs** | **Max-Abs** | **Min-Abs** | **Avg** | **Stdev** |
| ***Global SfM;* initial EO from georeferencing data (ASIFT; unweighted BA; 4 GCPs in one image)** | $\Delta X_o$ (m) | 0.418 | 1.387 | 0.016 | 0.123 | 0.524 |
| | $\Delta Y_o$ (m) | 0.253 | 1.192 | 0.025 | 0.114 | 0.324 |
| | $\Delta Z_o$ (m) | 0.482 | 1.386 | 0.004 | -0.118 | 0.590 |
| | $\Delta\omega$ (deg.) | 0.030 | 0.143 | 0.001 | -0.009 | 0.039 |
| | $\Delta\varphi$ (deg.) | 0.047 | 0.152 | 0.000 | 0.022 | 0.054 |
| | $\Delta\kappa$ (deg.) | 0.047 | 0.100 | 0.002 | 0.005 | 0.053 |
| | $\Delta$*LinearEO* (m) | 0.385 | 1.387 | 0.004 | 0.040 | 0.479 |
| | $\Delta$*AngularEO* (deg.) | 0.041 | 0.152 | 0.000 | 0.006 | 0.049 |
| ***Robust Global SfM:* initial EO from georeferencing data (ASIFT; unweighted BA; 4 GCPs in one image)** | $\Delta X_o$ (m) | 0.330 | 0.933 | 0.001 | 0.075 | 0.401 |
| | $\Delta Y_o$ (m) | 0.248 | 0.782 | 0.010 | 0.100 | 0.286 |
| | $\Delta Z_o$ (m) | 0.399 | 1.101 | 0.002 | -0.086 | 0.459 |
| | $\Delta\omega$ (deg.) | 0.026 | 0.110 | 0.000 | -0.009 | 0.032 |
| | $\Delta\varphi$ (deg.) | 0.036 | 0.098 | 0.000 | 0.017 | 0.041 |
| | $\Delta\kappa$ (deg.) | 0.040 | 0.087 | 0.000 | 0.005 | 0.045 |
| | $\Delta$*LinearEO* (m) | 0.326 | 1.101 | 0.001 | 0.030 | 0.382 |
| | $\Delta$*AngularEO* (deg.) | 0.034 | 0.110 | 0.000 | 0.004 | 0.039 |

percentage of 23% and 8% for *Global SfM* as well as 32% and 0% (similar results) for *Robust Global SfM*, for *RCD30 Dataset 2* and *RCD30 Dataset 3*, respectively.

Figure 7-22, Figure 7-23 and Figure 7-24 show the corresponding feature points extracted for pairs of images of *RCD30 Dataset 3*, using SURF and ASIFT. Specifically, Figure 7-22 depicts the image matching results between an oblique image taken by the front camera of the Leica RCD30 multi-view system and an oblique image taken by the right camera; Figure 7-23 illustrates the image matching results between an oblique and a nadir image; Figure 7-24 (top and middle) illustrates the image matching results between same perspective oblique images. Whereas significantly more correspondences are extracted by ASIFT, as it is shown in these figures, these correspondences lie on the same areas with the ones detected by SURF; thus, ASIFT does not extract any feature points in areas where SURF fails to find any corresponding feature points. Hence, the conclusion already derived in section 4.4.3.2 of this dissertation, that SURF is robust in finding correspondences in multi-view oblique imagery and oblique and vertical aerial imagery, is verified in this chapter through comparison with ASIFT, which is a well-established detector that has been proved to be robust to affine distortions. Besides, the thresholds used in the ASIFT detector for the extraction of feature points that cover the maximum possible area of the images (similar to the area covered by SURF feature points) generates ASIFT features that correspond to a significantly greater number than SURF ones and are very close to each other. As a result, a process of rejecting the redundant points, taking into account their distribution (e.g., by keeping one feature point per 20 pixels) is absolutely necessary for the case of ASIFT features, whereas it could be omitted in the case of SURF features. The usage of a different parameterization in ASIFT for extracting a less number of feature points results in feature points that do not cover the necessary overlapping areas of image pairs. This is illustrated in Figure 7-24, which shows that the extracted correspondences using

ASIFT, with a parameterization that yields approximately the same number of feature points as SURF, result in areas of images being without feature points.

In conclusion, taking into account the significant increase in computational time required by the ASIFT algorithm and the fact that the exterior orientation parameters estimated through SfM using SURF features are of equal or better accuracy compared to the ones estimated via SfM using ASIFT features for the test datasets, the SURF algorithm seems to be ideal for SfM processes even for datasets of oblique imagery of different perspective.

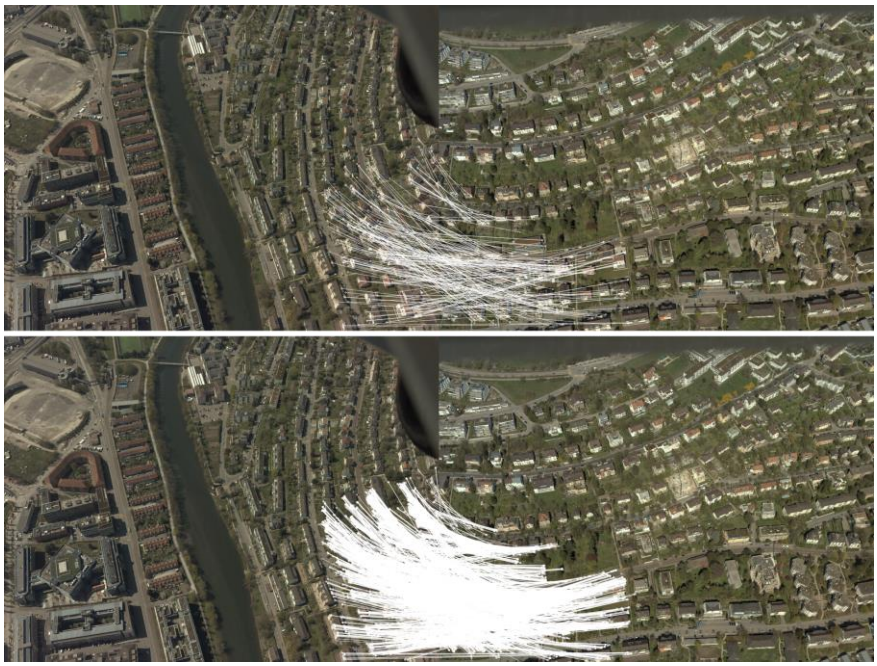

Figure 7-22. Image matching results between a pair of images of *RCD30 Dataset 3* acquired by the front and the right camera of the multi-view system using SURF (top) and ASIFT (bottom)

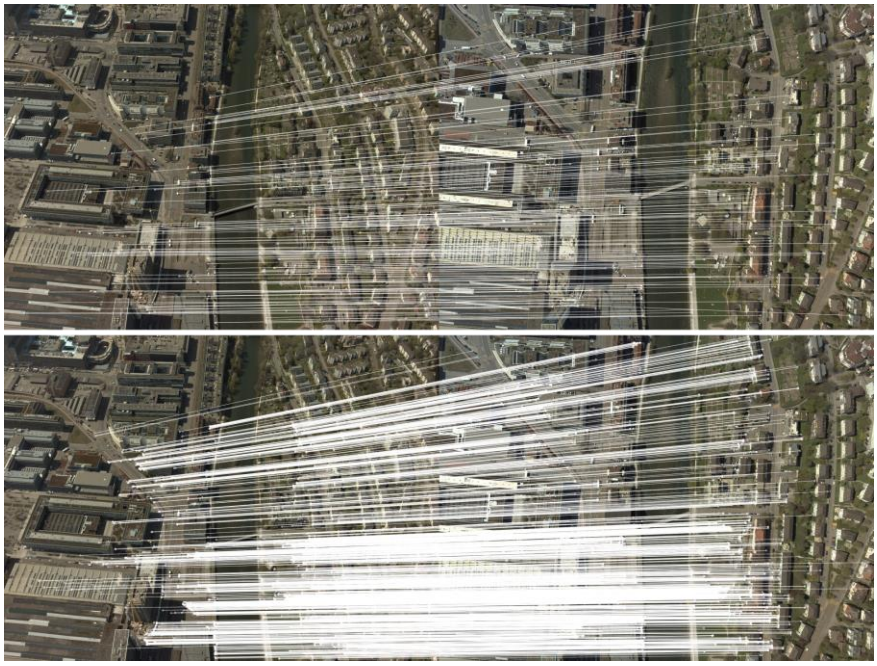

Figure 7-23. Image matching results between a pair of images of *RCD30 Dataset 3* acquired by the front and the nadir camera of the multi-view system using SURF (top) and ASIFT (bottom)

### 7.10.7 Impact of the Weighting Strategy

The experiments presented in sections 7.10.4, 7.10.5 and 7.10.6 were conducted assuming equal weights for feature point observations in the bundle adjustment process. In order to investigate the impact of a weighting strategy of bundle adjustment on the exterior orientation results, the weighting strategy presented in section 7.6 has been adopted and results are derived for all test datasets. Specifically, the results produced:

- through *Robust Global SfM*, using initial values from an *Incremental SfM*, for the five UAV datasets (*UAV Dataset 1 – UAV Dataset 5*), using the SURF feature extractor;
- through *Robust Incremental SfM* for *UAV Dataset 1* and *UAV Dataset 2*, using the SURF feature extractor;
- through *Robust Global SfM*, using initial values from an *Incremental SfM*, for *UAV Dataset 3*, using the ASIFT feature extractor;
- through *Robust Global SfM*, using initial values from georeferencing data for the three RCD30 datasets (*RCD30 Dataset 1 – RCD30 Dataset 3*), using the SURF feature extractor; and
- through *Robust Global SfM*, using initial values from georeferencing data for the three RCD30 datasets (*RCD30 Dataset 1 – RCD30 Dataset 3*), using the ASIFT feature extractor

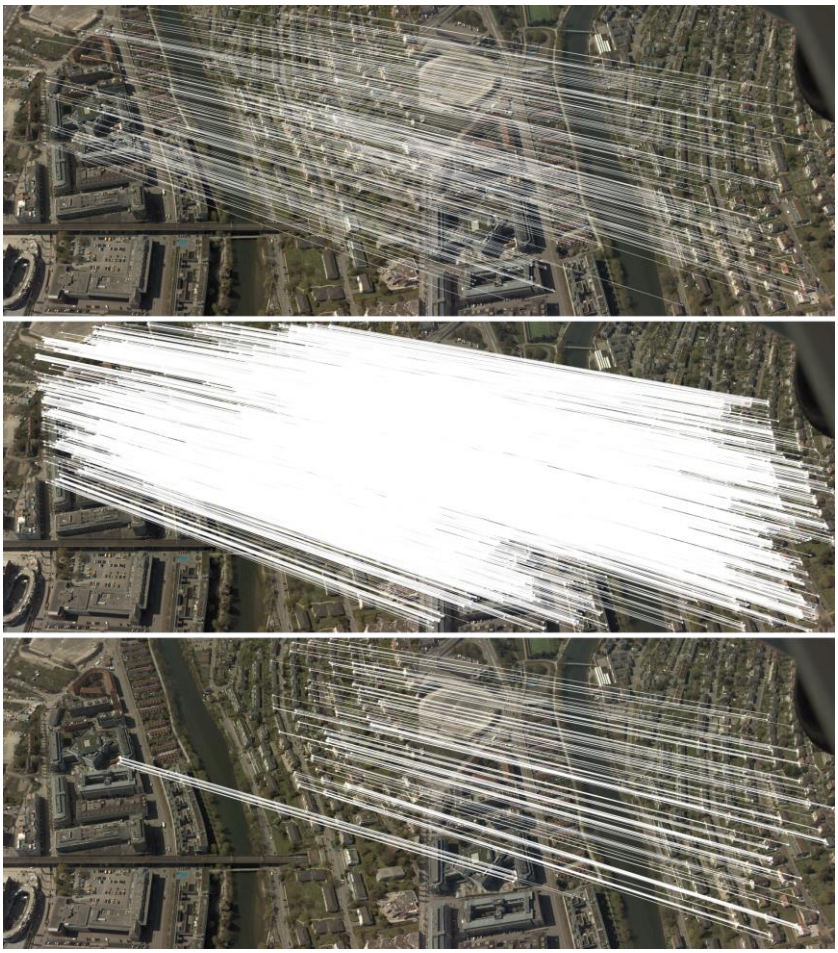

Figure 7-24. Image matching results between a pair of same-perspective images of *RCD30 Dataset 3* using SURF (top) and ASIFT (middle) with the parameterization followed in the experiments, as well as ASIFT with a different parameterization that yields approximately the same number of corresponding feature points as SURF (bottom)

Table 7-20. Differences between the EO parameters computed through *Robust Global SfM* using as initial EO parameters the ones estimated by *Incremental SfM* (using the SURF feature extractor, weighted BA and four GCP measurements in *img_1*) and the reference ones for the five UAV datasets

| ***Robust Global SfM:*** **initial EO from *Incremental SfM* (SURF; weighted BA; 4 GCPs in one image)** | | | | | | |
|---|---|---|---|---|---|---|
| **Dataset** | **Metric** | **Avg-Abs** | **Max-Abs** | **Min-Abs** | **Avg** | **Stdev** |
| ***UAV Dataset 1*: 23 oblique UAV images** | $\Delta X_o$ (m) | 0.320 | 0.656 | 0.021 | 0.212 | 0.270 |
| | $\Delta Y_o$ (m) | 0.191 | 0.234 | 0.132 | 0.191 | 0.257 |
| | $\Delta Z_o$ (m) | 0.047 | 0.104 | 0.003 | -0.019 | 0.286 |
| | $\Delta\omega$ (deg.) | 0.574 | 0.699 | 0.405 | 0.574 | 0.297 |
| | $\Delta\varphi$ (deg.) | 0.162 | 0.395 | 0.007 | -0.061 | 0.213 |
| | $\Delta\kappa$ (deg.) | 0.096 | 0.198 | 0.011 | 0.033 | 0.168 |
| | $\Delta$*LinearEO* (m) | 0.186 | 0.656 | 0.003 | 0.128 | 0.271 |
| | $\Delta$*AngularEO* (deg.) | 0.277 | 0.699 | 0.007 | 0.182 | 0.226 |
| ***UAV Dataset 2*: 35 oblique UAV images** | $\Delta X_o$ (m) | 0.283 | 0.606 | 0.000 | -0.068 | 0.306 |
| | $\Delta Y_o$ (m) | 0.179 | 0.338 | 0.011 | 0.163 | 0.127 |
| | $\Delta Z_o$ (m) | 0.093 | 0.270 | 0.001 | 0.039 | 0.117 |
| | $\Delta\omega$ (deg.) | 0.554 | 0.658 | 0.326 | 0.554 | 0.091 |
| | $\Delta\varphi$ (deg.) | 0.367 | 0.741 | 0.020 | -0.322 | 0.283 |
| | $\Delta\kappa$ (deg.) | 0.073 | 0.195 | 0.004 | 0.026 | 0.086 |
| | $\Delta$*LinearEO* (m) | 0.185 | 0.606 | 0.000 | 0.045 | 0.183 |
| | $\Delta$*AngularEO* (deg.) | 0.332 | 0.741 | 0.004 | 0.086 | 0.153 |
| ***UAV Dataset 3*: 60 oblique UAV images** | $\Delta X_o$ (m) | 0.676 | 1.126 | 0.040 | 0.619 | 0.451 |
| | $\Delta Y_o$ (m) | 0.228 | 0.453 | 0.000 | 0.178 | 0.172 |
| | $\Delta Z_o$ (m) | 0.187 | 0.584 | 0.004 | 0.121 | 0.225 |
| | $\Delta\omega$ (deg.) | 0.506 | 0.892 | 0.045 | 0.023 | 0.550 |
| | $\Delta\varphi$ (deg.) | 0.206 | 0.556 | 0.007 | -0.128 | 0.227 |
| | $\Delta\kappa$ (deg.) | 0.162 | 0.618 | 0.010 | 0.125 | 0.179 |
| | $\Delta$*LinearEO* (m) | 0.364 | 1.126 | 0.000 | 0.306 | 0.283 |
| | $\Delta$*AngularEO* (deg.) | 0.291 | 0.892 | 0.007 | 0.007 | 0.319 |
| ***UAV Dataset 4*: 80 oblique UAV images** | $\Delta X_o$ (m) | 0.648 | 1.103 | 0.034 | 0.605 | 0.409 |
| | $\Delta Y_o$ (m) | 0.568 | 1.876 | 0.003 | -0.277 | 0.765 |
| | $\Delta Z_o$ (m) | 0.157 | 0.431 | 0.004 | 0.018 | 0.197 |
| | $\Delta\omega$ (deg.) | 0.383 | 0.813 | 0.000 | 0.062 | 0.442 |
| | $\Delta\varphi$ (deg.) | 0.213 | 0.553 | 0.005 | -0.203 | 0.169 |
| | $\Delta\kappa$ (deg.) | 0.373 | 1.131 | 0.007 | 0.346 | 0.390 |
| | $\Delta$*LinearEO* (m) | 0.458 | 1.876 | 0.003 | 0.115 | 0.457 |
| | $\Delta$*AngularEO* (deg.) | 0.323 | 1.131 | 0.000 | 0.068 | 0.334 |
| ***UAV Dataset 5*: 100 oblique UAV images** | $\Delta X_o$ (m) | 0.511 | 1.105 | 0.004 | 0.452 | 0.447 |
| | $\Delta Y_o$ (m) | 0.762 | 1.865 | 0.025 | -0.533 | 0.841 |
| | $\Delta Z_o$ (m) | 0.156 | 0.541 | 0.000 | -0.008 | 0.214 |
| | $\Delta\omega$ (deg.) | 0.309 | 0.778 | 0.000 | 0.092 | 0.385 |
| | $\Delta\varphi$ (deg.) | 0.216 | 0.552 | 0.000 | -0.209 | 0.156 |
| | $\Delta\kappa$ (deg.) | 0.551 | 1.263 | 0.007 | 0.529 | 0.481 |
| | $\Delta$*LinearEO* (m) | 0.476 | 1.865 | 0.000 | -0.030 | 0.500 |
| | $\Delta$*AngularEO* (deg.) | 0.359 | 1.263 | 0.000 | 0.137 | 0.341 |

Table 7-21. Differences between the EO parameters computed through *Robust Incremental SfM* (using the SURF feature extractor, weighted BA and four GCP measurements in *img_1*) and the reference ones for *UAV Dataset 1* and *UAV Dataset 2*

| ***Robust Incremental SfM* (SURF; weighted BA; 4 GCPs in one image)** | | | | | | |
|---|---|---|---|---|---|---|
| **Dataset** | **Metric** | **Avg-Abs** | **Max-Abs** | **Min-Abs** | **Avg** | **Stdev** |
| ***UAV Dataset 1*: 23 oblique UAV images** | $\Delta X_o$ (m) | 0.498 | 1.147 | 0.023 | 0.427 | 0.333 |
| | $\Delta Y_o$ (m) | 0.150 | 0.176 | 0.107 | 0.150 | 0.274 |
| | $\Delta Z_o$ (m) | 0.056 | 0.133 | 0.001 | -0.054 | 0.306 |
| | $\Delta\omega$ (deg.) | 0.622 | 0.735 | 0.468 | 0.622 | 0.371 |
| | $\Delta\varphi$ (deg.) | 0.142 | 0.402 | 0.019 | -0.074 | 0.308 |
| | $\Delta\kappa$ (deg.) | 0.087 | 0.199 | 0.000 | 0.040 | 0.276 |
| | $\Delta$*LinearEO* (m) | 0.235 | 1.147 | 0.001 | 0.174 | 0.305 |
| | $\Delta$*AngularEO* (deg.) | 0.284 | 0.735 | 0.000 | 0.196 | 0.318 |
| ***UAV Dataset 2*: 35 oblique UAV images** | $\Delta X_o$ (m) | 0.180 | 0.455 | 0.022 | -0.072 | 0.193 |
| | $\Delta Y_o$ (m) | 0.157 | 0.405 | 0.007 | 0.027 | 0.184 |
| | $\Delta Z_o$ (m) | 0.140 | 0.338 | 0.001 | -0.025 | 0.166 |
| | $\Delta\omega$ (deg.) | 0.778 | 1.221 | 0.087 | 0.778 | 0.317 |
| | $\Delta\varphi$ (deg.) | 0.290 | 0.686 | 0.001 | -0.154 | 0.299 |
| | $\Delta\kappa$ (deg.) | 0.096 | 0.341 | 0.001 | 0.084 | 0.113 |
| | $\Delta$*LinearEO* (m) | 0.159 | 0.455 | 0.001 | -0.024 | 0.181 |
| | $\Delta$*AngularEO* (deg.) | 0.388 | 1.221 | 0.001 | 0.236 | 0.243 |

Table 7-22. Differences between the EO parameters computed through *Robust Global SfM* using as initial EO parameters the ones estimated by *Incremental SfM* (using the SURF feature extractor, weighted BA and four GCP measurements in *img_1*) and the reference ones for *UAV Dataset 3*

| ***Robust Global SfM:* initial EO from *Incremental SfM* (ASIFT; weighted BA; 4 GCPs in one image)** | | | | | | |
|---|---|---|---|---|---|---|
| **Dataset** | **Metric** | **Avg-Abs** | **Max-Abs** | **Min-Abs** | **Avg** | **Stdev** |
| ***UAV Dataset 3*: 60 oblique UAV images** | $\Delta X_o$ (m) | 0.465 | 0.836 | 0.027 | 0.411 | 0.324 |
| | $\Delta Y_o$ (m) | 0.213 | 0.747 | 0.002 | 0.163 | 0.236 |
| | $\Delta Z_o$ (m) | 0.392 | 0.973 | 0.001 | 0.315 | 0.441 |
| | $\Delta\omega$ (deg.) | 0.663 | 1.488 | 0.057 | 0.336 | 0.631 |
| | $\Delta\varphi$ (deg.) | 0.432 | 1.286 | 0.008 | 0.089 | 0.477 |
| | $\Delta\kappa$ (deg.) | 0.313 | 1.989 | 0.000 | 0.288 | 0.459 |
| | $\Delta$*LinearEO* (m) | 0.356 | 0.973 | 0.001 | 0.296 | 0.334 |
| | $\Delta$*AngularEO* (deg.) | 0.469 | 1.989 | 0.000 | 0.238 | 0.522 |

are presented in this section. Table 7-20 outlines the results achieved through the *Robust Global SfM* method using initial values from an *Incremental SfM*, for the five UAV datasets (*UAV Dataset 1 – UAV Dataset 5*) using the SURF feature extractor and applying the weighting strategy presented in section 7.6 in the final bundle adjustment of *Robust Global SfM*. The results of *Incremental SfM* used for initialization have already been presented in Table 7-4, Table 7-5, Table 7-10, Table 7-11 and Table 7-12 for *UAV Dataset 1*, *UAV Dataset 2*, *UAV Dataset 3*, *UAV Dataset 4* and *UAV Dataset 5*, respectively. In addition, these tables have presented the results of *Robust Global SfM* assuming equal weights for point observations, which are used for comparison reasons in this section. Specifically, the average absolute $\Delta$*AngularEO* for the case of weighted bundle adjustment is smaller than the one that corresponds to an unweighted bundle adjustment for all UAV datasets

Table 7-23. Differences between the EO parameters computed through *Robust Global SfM* using initial EO parameters by georeferencing data (using the SURF feature extractor, weighted BA and four GCP measurements in *img_1*) and the reference ones for the three RCD30 aerial image datasets

| ***Robust Global SfM:* initial EO from *Incremental SfM* (SURF; weighted BA; 4 GCPs in one image)** | | | | | | |
|---|---|---|---|---|---|---|
| **Dataset** | **Metric** | **Avg-Abs** | **Max-Abs** | **Min-Abs** | **Avg** | **Stdev** |
| ***RCD30 Dataset 1*: 10 single-perspective oblique aerial images** | $\Delta X_o$ (m) | 0.198 | 0.371 | 0.049 | 0.106 | 0.200 |
| | $\Delta Y_o$ (m) | 0.338 | 0.893 | 0.065 | -0.014 | 0.420 |
| | $\Delta Z_o$ (m) | 0.329 | 0.572 | 0.053 | -0.108 | 0.351 |
| | $\Delta\omega$ (deg.) | 0.018 | 0.043 | 0.002 | -0.006 | 0.021 |
| | $\Delta\varphi$ (deg.) | 0.017 | 0.039 | 0.001 | 0.014 | 0.015 |
| | $\Delta\kappa$ (deg.) | 0.008 | 0.020 | 0.000 | -0.002 | 0.010 |
| | $\Delta$*LinearEO* (m) | 0.288 | 0.893 | 0.049 | -0.005 | 0.324 |
| | $\Delta$*AngularEO* (deg.) | 0.014 | 0.043 | 0.000 | 0.002 | 0.016 |
| ***RCD30 Dataset 2*: 40 multi-perspective oblique aerial images** | $\Delta X_o$ (m) | 0.268 | 0.629 | 0.009 | 0.023 | 0.322 |
| | $\Delta Y_o$ (m) | 0.326 | 1.183 | 0.003 | 0.032 | 0.418 |
| | $\Delta Z_o$ (m) | 0.442 | 1.306 | 0.019 | -0.098 | 0.572 |
| | $\Delta\omega$ (deg.) | 0.023 | 0.093 | 0.002 | -0.007 | 0.029 |
| | $\Delta\varphi$ (deg.) | 0.036 | 0.103 | 0.000 | 0.009 | 0.044 |
| | $\Delta\kappa$ (deg.) | 0.042 | 0.075 | 0.020 | 0.006 | 0.043 |
| | $\Delta$*LinearEO* (m) | 0.346 | 1.306 | 0.003 | -0.014 | 0.437 |
| | $\Delta$*AngularEO* (deg.) | 0.034 | 0.103 | 0.000 | 0.003 | 0.039 |
| ***RCD30 Dataset 3*: 50 multi-perspective oblique and vertical aerial images** | $\Delta X_o$ (m) | 0.273 | 0.837 | 0.001 | 0.076 | 0.333 |
| | $\Delta Y_o$ (m) | 0.269 | 1.100 | 0.005 | 0.068 | 0.320 |
| | $\Delta Z_o$ (m) | 0.307 | 0.939 | 0.009 | -0.108 | 0.377 |
| | $\Delta\omega$ (deg.) | 0.022 | 0.107 | 0.000 | -0.009 | 0.028 |
| | $\Delta\varphi$ (deg.) | 0.027 | 0.089 | 0.000 | 0.015 | 0.032 |
| | $\Delta\kappa$ (deg.) | 0.047 | 0.102 | 0.001 | 0.009 | 0.052 |
| | $\Delta$*LinearEO* (m) | 0.283 | 1.100 | 0.001 | 0.012 | 0.343 |
| | $\Delta$*AngularEO* (deg.) | 0.032 | 0.107 | 0.000 | 0.005 | 0.037 |

(case: *Robust Global SfM*, SURF feature extraction); the percentages of improvement lie between 0.04% and 5%. The average absolute $\Delta$*LinearEO*, computed for the case of weighted bundle adjustment, is smaller than the one calculated for the case of unweighted bundle adjustment, for *UAV Dataset 1*, *UAV Dataset 2* and *UAV Dataset 5*, with percentages of reduction lying between 1% and 8%. Whereas for *UAV Dataset 3* and *UAV Dataset 4*, the average absolute $\Delta$*LinearEO* is worse for the case of weighted bundle adjustment, these differences are small, corresponding to percentages of less than 1%.

Similar results are also derived for the UAV datasets for the case of *Robust Incremental SfM* using the SURF feature detector (Table 7-21, in comparison with Table 7-2 and Table 7-3 for *UAV Dataset 1* and *UAV Dataset 2*, respectively) and *Robust Global SfM* using ASIFT (Table 7-22, in comparison with Table 7-16 for *UAV Dataset 3*). Specifically, both average absolute $\Delta$*LinearEO* and $\Delta$*AngularEO* have been improved for the case of weighted bundle adjustment for all three tests, except for the value of $\Delta$*LinearEO* that correponds to an increase by less than 3% for $\Delta$*AngularEO* for the case of *Robust Incremental SfM* using SURF. The results of comparison between unweighted and weighted bundle adustment for the UAV datasets are illustrated in the form of a bar chart in Figure 7-25.

Table 7-24. Differences between the EO parameters computed through *Robust Global SfM* using initial EO parameters by georeferencing data (using the ASIFT feature extractor, weighted BA and four GCP measurements in *img_1*) and the reference ones for the three RCD30 aerial image datasets

| ***Robust Global SfM:*** **initial EO from *Incremental SfM* (ASIFT; weighted BA; 4 GCPs in one image)** | | | | | | |
|---|---|---|---|---|---|---|
| **Dataset** | **Metric** | **Avg-Abs** | **Max-Abs** | **Min-Abs** | **Avg** | **Stdev** |
| ***RCD30 Dataset 1*: 10 single-perspective oblique aerial images** | $\Delta X_o$ (m) | 0.181 | 0.382 | 0.006 | 0.111 | 0.192 |
| | $\Delta Y_o$ (m) | 0.226 | 0.555 | 0.031 | 0.013 | 0.287 |
| | $\Delta Z_o$ (m) | 0.433 | 0.761 | 0.070 | -0.057 | 0.490 |
| | $\Delta\omega$ (deg.) | 0.020 | 0.048 | 0.005 | -0.006 | 0.024 |
| | $\Delta\varphi$ (deg.) | 0.016 | 0.039 | 0.002 | 0.014 | 0.014 |
| | $\Delta\kappa$ (deg.) | 0.011 | 0.030 | 0.001 | 0.001 | 0.014 |
| | $\Delta$*LinearEO* (m) | 0.280 | 0.761 | 0.006 | 0.022 | 0.323 |
| | $\Delta$*AngularEO* (deg.) | 0.016 | 0.048 | 0.001 | 0.003 | 0.017 |
| ***RCD30 Dataset 2: 40 multi-perspective oblique aerial images*** | $\Delta X_o$ (m) | 0.362 | 1.406 | 0.002 | 0.118 | 0.483 |
| | $\Delta Y_o$ (m) | 0.306 | 1.769 | 0.017 | 0.083 | 0.420 |
| | $\Delta Z_o$ (m) | 0.440 | 1.693 | 0.007 | -0.097 | 0.582 |
| | $\Delta\omega$ (deg.) | 0.045 | 0.270 | 0.000 | -0.007 | 0.064 |
| | $\Delta\varphi$ (deg.) | 0.040 | 0.133 | 0.001 | 0.017 | 0.048 |
| | $\Delta\kappa$ (deg.) | 0.051 | 0.104 | 0.011 | 0.004 | 0.055 |
| | $\Delta$*LinearEO* (m) | 0.369 | 1.769 | 0.002 | 0.035 | 0.495 |
| | $\Delta$*AngularEO* (deg.) | 0.046 | 0.270 | 0.000 | 0.005 | 0.056 |
| ***RCD30 Dataset 3: 50 multi-perspective oblique and vertical aerial images*** | $\Delta X_o$ (m) | 0.342 | 0.973 | 0.004 | 0.089 | 0.415 |
| | $\Delta Y_o$ (m) | 0.252 | 0.817 | 0.002 | 0.110 | 0.291 |
| | $\Delta Z_o$ (m) | 0.391 | 1.115 | 0.003 | -0.089 | 0.450 |
| | $\Delta\omega$ (deg.) | 0.026 | 0.114 | 0.002 | -0.010 | 0.032 |
| | $\Delta\varphi$ (deg.) | 0.037 | 0.104 | 0.001 | 0.018 | 0.041 |
| | $\Delta\kappa$ (deg.) | 0.040 | 0.090 | 0.002 | 0.005 | 0.046 |
| | $\Delta$*LinearEO* (m) | 0.329 | 1.115 | 0.002 | 0.037 | 0.385 |
| | $\Delta$*AngularEO* (deg.) | 0.034 | 0.114 | 0.001 | 0.004 | 0.040 |

Table 7-23 and Table 7-24 present the results achieved through *Robust Global SfM* for the three aerial image datasets (*RCD30 Dataset 1 – RCD3 Dataset 3*), using SURF and ASIFT, respectively and applying the developed weighting strategy in the final bundle adjustment. The corresponding results of an unweighted bundle adustment have been presented in Table 7-13, Table 7-14 and Table 7-15, for the case of SURF and Table 7-17, Table 7-18 and Table 7-19 for the case of ASIFT. The results of comparison between unweighted and weighted bundle adustment for the RCD30 aerial image datasets are illustrated in the form of a bar chart in Figure 7-26.

The comparison of these results shows that some of the exterior orientation parameters have been improved through the use of a weighted bundle adustment, whereas others have got worse; the improvement of the exterior orientation parameters as a result of the weighted bundle adjutstment is generally bigger than the corresponding worsening of the results. Specifically, among all tests, the percentages of increase of average absolute $\Delta$*LinearEO* and $\Delta$*AngularEO* in the results of a weighted bundle adjustment are smaller than 2% for all tests using the RCD30 aerial image datasets, while the reduction of $\Delta$*LinearEO* and $\Delta$*AngularEO* in the results of a weighted bundle adjustment reaches the percentage of 6%.

Although the weighting strategy in the bundle block adjustment has improved the results of the UAV datasets, the corresponding results of the aerial image datasets (*RCD30 Dataset 1 – RCD3 Dataset 3*) have not generally been improved. The difference of the behaviour of the weighting strategy lies on the different nature of the specific UAV and aerial image datasets; this difference lies on the smaller number of images of the aerial image datasets that depict a 3D feature point, due to the smaller image overlap along with the difference in perspective among the images of two out of the three aerial image datasets (*RCD30 Dataset 2* and *RCD30 Dataset 3*). As a result, for such kind of datasets, some feature points that are located on the upper side of an oblique image (in relation to the tilt axis) and correspond to small scale will be given small weights, despite being significant to tie the whole block of images due to possible lack of other feature point

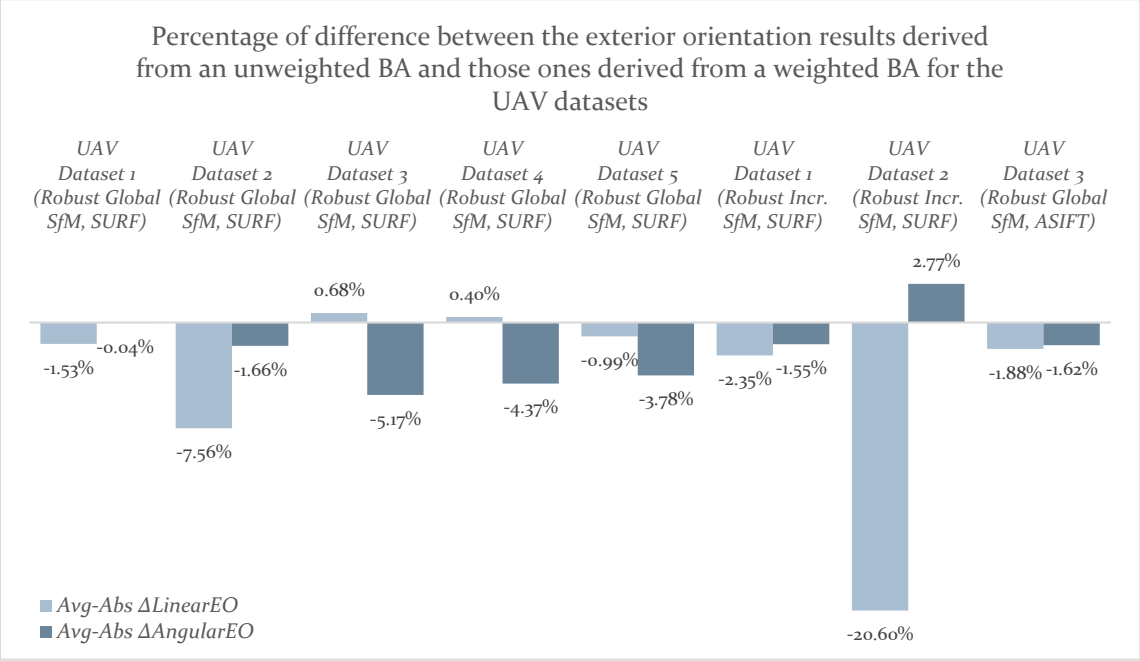


Figure 7-25. Comparison of Δ*LinearEO* and Δ*AngularEO* as calculated between an unweighted BA and a weighted BA through different methods for the UAV datasets; a negative percentage of difference implies a reduction of Δ*LinearEO* or Δ*AngularEO* for the weighted BA, that is, an improvement in the results of the weighted BA

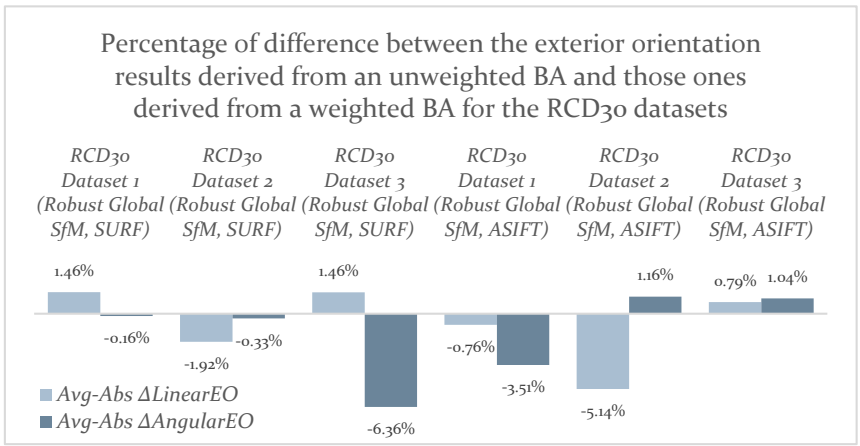


Figure 7-26. Comparison of Δ*LinearEO* and Δ*AngularEO* as calculated between an unweighted BA and a weighted BA through different methods for the RCD30 aerial image datasets; a negative percentage of difference implies an improvement in the results of the weighted BA

Table 7-25. Differences between the exterior orientation (EO) parameters computed through commercial software and the reference ones for the five UAV test datasets

| **Commercial SfM Software (4 GCPs in two images)** | | | | | | |
|---|---|---|---|---|---|---|
| **Dataset** | **Metric** | **Avg-Abs** | **Max-Abs** | **Min-Abs** | **Avg** | **Stdev** |
| ***UAV Dataset 1*: 23 oblique UAV images** | $\Delta X_o$ (m) | 0.248 | 0.412 | 0.003 | -0.248 | 0.128 |
| | $\Delta Y_o$ (m) | 0.233 | 0.472 | 0.048 | 0.233 | 0.142 |
| | $\Delta Z_o$ (m) | 0.153 | 0.297 | 0.014 | -0.097 | 0.152 |
| | $\Delta\omega$ (deg.) | 0.651 | 0.691 | 0.623 | 0.651 | 0.019 |
| | $\Delta\varphi$ (deg.) | 0.397 | 0.440 | 0.317 | -0.397 | 0.024 |
| | $\Delta\kappa$ (deg.) | 0.047 | 0.084 | 0.002 | -0.005 | 0.051 |
| | $\Delta$*LinearEO* (m) | 0.212 | 0.472 | 0.003 | -0.037 | 0.141 |
| | $\Delta$*AngularEO* (deg.) | 0.365 | 0.691 | 0.002 | 0.083 | 0.031 |
| ***UAV Dataset 2*: 35 oblique UAV images** | $\Delta X_o$ (m) | 0.268 | 0.472 | 0.003 | -0.268 | 0.147 |
| | $\Delta Y_o$ (m) | 0.231 | 0.474 | 0.048 | 0.231 | 0.141 |
| | $\Delta Z_o$ (m) | 0.155 | 0.302 | 0.014 | -0.099 | 0.154 |
| | $\Delta\omega$ (deg.) | 0.657 | 0.701 | 0.623 | 0.657 | 0.024 |
| | $\Delta\varphi$ (deg.) | 0.406 | 0.440 | 0.317 | -0.406 | 0.028 |
| | $\Delta\kappa$ (deg.) | 0.047 | 0.083 | 0.000 | -0.006 | 0.051 |
| | $\Delta$*LinearEO* (m) | 0.218 | 0.474 | 0.003 | -0.045 | 0.147 |
| | $\Delta$*AngularEO* (deg.) | 0.370 | 0.701 | 0.000 | 0.081 | 0.034 |
| ***UAV Dataset 3*: 60 oblique UAV images** | $\Delta X_o$ (m) | 0.514 | 0.789 | 0.236 | -0.514 | 0.147 |
| | $\Delta Y_o$ (m) | 0.482 | 0.969 | 0.153 | 0.482 | 0.214 |
| | $\Delta Z_o$ (m) | 0.204 | 0.391 | 0.011 | -0.204 | 0.115 |
| | $\Delta\omega$ (deg.) | 0.495 | 0.724 | 0.019 | 0.351 | 0.416 |
| | $\Delta\varphi$ (deg.) | 0.606 | 0.843 | 0.350 | -0.606 | 0.176 |
| | $\Delta\kappa$ (deg.) | 0.148 | 0.407 | 0.003 | -0.039 | 0.179 |
| | $\Delta$*LinearEO* (m) | 0.400 | 0.969 | 0.011 | -0.079 | 0.158 |
| | $\Delta$*AngularEO* (deg.) | 0.417 | 0.843 | 0.003 | -0.098 | 0.257 |
| ***UAV Dataset 4*: 80 oblique UAV images** | $\Delta X_o$ (m) | 0.571 | 0.822 | 0.236 | -0.571 | 0.148 |
| | $\Delta Y_o$ (m) | 0.628 | 1.140 | 0.153 | 0.628 | 0.307 |
| | $\Delta Z_o$ (m) | 0.178 | 0.392 | 0.010 | -0.178 | 0.109 |
| | $\Delta\omega$ (deg.) | 0.465 | 0.727 | 0.022 | 0.169 | 0.490 |
| | $\Delta\varphi$ (deg.) | 0.629 | 0.851 | 0.350 | -0.629 | 0.167 |
| | $\Delta\kappa$ (deg.) | 0.186 | 0.417 | 0.003 | -0.105 | 0.199 |
| | $\Delta$*LinearEO* (m) | 0.459 | 1.140 | 0.010 | -0.040 | 0.188 |
| | $\Delta$*AngularEO* (deg.) | 0.427 | 0.851 | 0.003 | -0.188 | 0.285 |
| ***UAV Dataset 5*: 100 oblique UAV images** | $\Delta X_o$ (m) | 0.505 | 0.765 | 0.235 | -0.505 | 0.129 |
| | $\Delta Y_o$ (m) | 0.677 | 1.079 | 0.153 | 0.677 | 0.308 |
| | $\Delta Z_o$ (m) | 0.146 | 0.393 | 0.001 | -0.132 | 0.133 |
| | $\Delta\omega$ (deg.) | 0.456 | 0.730 | 0.023 | 0.060 | 0.500 |
| | $\Delta\varphi$ (deg.) | 0.616 | 0.852 | 0.350 | -0.616 | 0.155 |
| | $\Delta\kappa$ (deg.) | 0.189 | 0.404 | 0.001 | -0.124 | 0.185 |
| | $\Delta$*LinearEO* (m) | 0.443 | 1.079 | 0.001 | 0.013 | 0.190 |
| | $\Delta$*AngularEO* (deg.) | 0.420 | 0.852 | 0.001 | -0.227 | 0.280 |

correspondences throughout the images. Thus, in such cases, the use of a weighting strategy depending on the scale of each point observation is risky and may deteriorate the exterior orientation

Table 7-26. Differences between the EO parameters computed through commercial software and the reference ones for the three aerial image datasets

| **Commercial SfM Software (4 GCPs in two images)** | | | | | | |
|---|---|---|---|---|---|---|
| **Dataset** | **Metric** | **Avg-Abs** | **Max-Abs** | **Min-Abs** | **Avg** | **Stdev** |
| ***RCD30 Dataset 1*: 10 single-perspective oblique aerial images** | $\Delta X_o$ (m) | 0.534 | 0.881 | 0.006 | 0.528 | 0.291 |
| | $\Delta Y_o$ (m) | 0.213 | 0.493 | 0.038 | 0.205 | 0.155 |
| | $\Delta Z_o$ (m) | 0.164 | 0.321 | 0.005 | 0.040 | 0.191 |
| | $\Delta\omega$ (deg.) | 0.047 | 0.066 | 0.013 | -0.047 | 0.014 |
| | $\Delta\varphi$ (deg.) | 0.048 | 0.082 | 0.015 | 0.048 | 0.021 |
| | $\Delta\kappa$ (deg.) | 0.015 | 0.028 | 0.002 | -0.010 | 0.014 |
| | $\Delta$*LinearEO* (m) | 0.303 | 0.881 | 0.005 | 0.258 | 0.212 |
| | $\Delta$*AngularEO* (deg.) | 0.036 | 0.082 | 0.002 | -0.003 | 0.016 |
| ***RCD30 Dataset 2*: 40 multi-perspective oblique aerial images** | $\Delta X_o$ (m) | 0.522 | 0.975 | 0.021 | 0.528 | 0.363 |
| | $\Delta Y_o$ (m) | 0.391 | 1.106 | 0.010 | 0.205 | 0.285 |
| | $\Delta Z_o$ (m) | 0.264 | 0.866 | 0.010 | 0.040 | 0.300 |
| | $\Delta\omega$ (deg.) | 0.054 | 0.171 | 0.001 | -0.047 | 0.037 |
| | $\Delta\varphi$ (deg.) | 0.034 | 0.082 | 0.006 | 0.048 | 0.032 |
| | $\Delta\kappa$ (deg.) | 0.034 | 0.086 | 0.000 | -0.010 | 0.035 |
| | $\Delta$*LinearEO* (m) | 0.392 | 1.106 | 0.010 | 0.258 | 0.316 |
| | $\Delta$*AngularEO* (deg.) | 0.041 | 0.171 | 0.000 | -0.003 | 0.035 |
| ***RCD30 Dataset 3*: 50 multi-perspective oblique and vertical aerial images** | $\Delta X_o$ (m) | 0.567 | 1.109 | 0.032 | 0.519 | 0.374 |
| | $\Delta Y_o$ (m) | 0.298 | 0.974 | 0.008 | 0.279 | 0.252 |
| | $\Delta Z_o$ (m) | 0.253 | 0.702 | 0.001 | -0.178 | 0.257 |
| | $\Delta\omega$ (deg.) | 0.048 | 0.153 | 0.001 | -0.046 | 0.036 |
| | $\Delta\varphi$ (deg.) | 0.036 | 0.086 | 0.000 | 0.030 | 0.030 |
| | $\Delta\kappa$ (deg.) | 0.032 | 0.088 | 0.002 | 0.027 | 0.030 |
| | $\Delta$*LinearEO* (m) | 0.373 | 1.109 | 0.001 | 0.206 | 0.294 |
| | $\Delta$*AngularEO* (deg.) | 0.039 | 0.153 | 0.000 | 0.004 | 0.032 |

results. In conclusion, the use of the developed weighting strategy in the bundle block adjustment generally improves the exterior orientation results of datasets that incorporate the same characteristics as the five UAV datasets, i.e., in cases of same-perspective images that correspond to a big percentage of overlap and include a significant number of feature point observations among every pair of overlapping images.

### 7.10.8 Comparison with Commercial Software Results

For comparison reasons, the exterior orientation parameters of all test datasets were computed through well-established commercial SfM software. The same metrics that were presented for the exterior orientation results that were computed through the developed solutions are also calculated for the results obtained through the commercial packages. The results are presented in Table 7-25 for the five UAV test datasets (*UAV Dataset 1 – UAV Dataset 5*) and in Table 7-26 for the three aerial image datasets *(RCD30 Dataset 1 – RCD30 Dataset 3*). The commercial software package used for obtaining the results presented in Table 7-25 for the five UAV datasets is Pix4DMapper (Pix4D, 2019); the results outlined in Table 7-26 are derived through the commercial software Agisoft PhotoScan (Agisoft LLC, 2019). Both software packages adopt an incremental SfM workflow.

In all tests conducted using the commercial software solutions, four coplanar GCPs are measured in *img_1* for each test dataset; *img_1* along with its GCP measurements are the same with the ones used by the developed software for testing the proposed approaches in the previous sections. However, both commercial software solutions require the manual measurement of GCPs in at least two images for georeferencing the SfM results; thus, the GCPs are also manually measured in another image (*img_2*) in both commercial software solutions, contrary to the experiments performed using the developed solutions, which require the measurement of coplanar GCPs in one image. Similarly to the experiments performed using the developed solutions, in every dataset for both commercial software packages the same fixed interior orientation parameters are specified so that the derived results are comparable. Finally, in the experiments performed using both commercial software packages, the feature extraction process took place in downscaled images with a resizing factor of 4, so that the results are comparable to the ones achieved through the developed software solutions, which apply a resizing factor of 5 for the images during feature extraction. Quite similar results were also derived using smaller resizing factors.

As far as the UAV oblique image datasets are concerned, bigger absolute differences are generally observed using the commercial software solution in the datasets that include a bigger number of images. Similar results are derived for the dataset of 23 images (*UAV Dataset 1*) and the dataset of 35 images (*UAV Dataset 2*); the latter one is the only dataset for which the GCP measurements have been conducted in internal images of the block. This shows that in spite of the increase of the number of images by about 50%, the accuracy of the exterior orientation parameters remains the same in the case that GCPs are measured in internal images of the block. Furthermore, an increase of the number of images by about 160% (*UAV Dataset 1 – UAV Dataset 3*) yields an increase of average absolute Δ*LinearEO* by 89% and average absolute Δ*AngularEO* by 14%, while an increase of the number of images by about 250% (*UAV Dataset 1 – UAV Dataset 4*) yields an increase of Δ*LinearEO* by 117% and Δ*AngularEO* by 17%. Quite similar results are achieved for the datasets

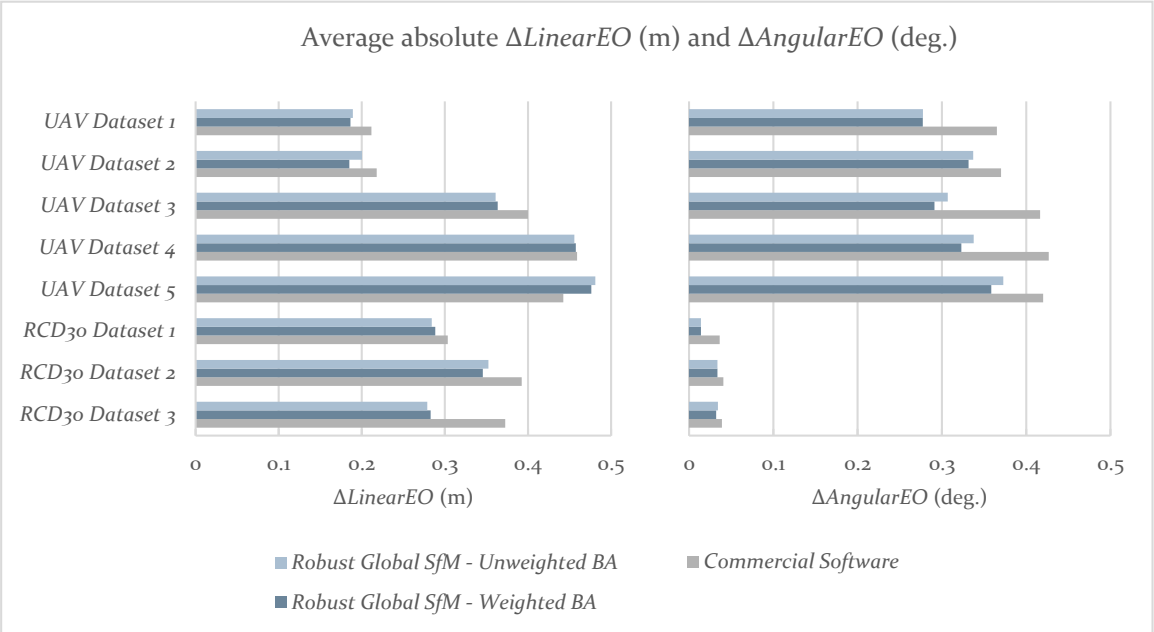


Figure 7-27. Average absolute Δ*LinearEO* (m) and Δ*AngularEO* (deg.) for the developed *Robust Global SfM* methods (i.e., *Robust Global SfM* initialized by *Incremental SfM* for *UAV Dataset 1 – UAV Dataset 5* and *Global SfM* initialized by georeferencing data for *RCD30 Dataset 1 – RCD30 Dataset 3*) (using the SURF feature extractor and four GCP measurements in *img_1*) and the commercial software solutions for all test datasets

of 80 and 100 UAV images, i.e., *UAV Dataset 4* and *UAV Dataset 5*, respectively, with *UAV Dataset 5* corresponding to relatively smaller average absolute Δ*LinearEO* and Δ*AngularEO* compared to *UAV Dataset 4* (decrease by 3% and 2% respectively).

As far as the aerial image datasets taken by the multi-camera system are concerned, a better accuracy in the exterior orientation parameters is observed for the dataset of 10 single-perspective oblique images (*RCD30 Dataset 1*). The largest absolute differences from the reference data are

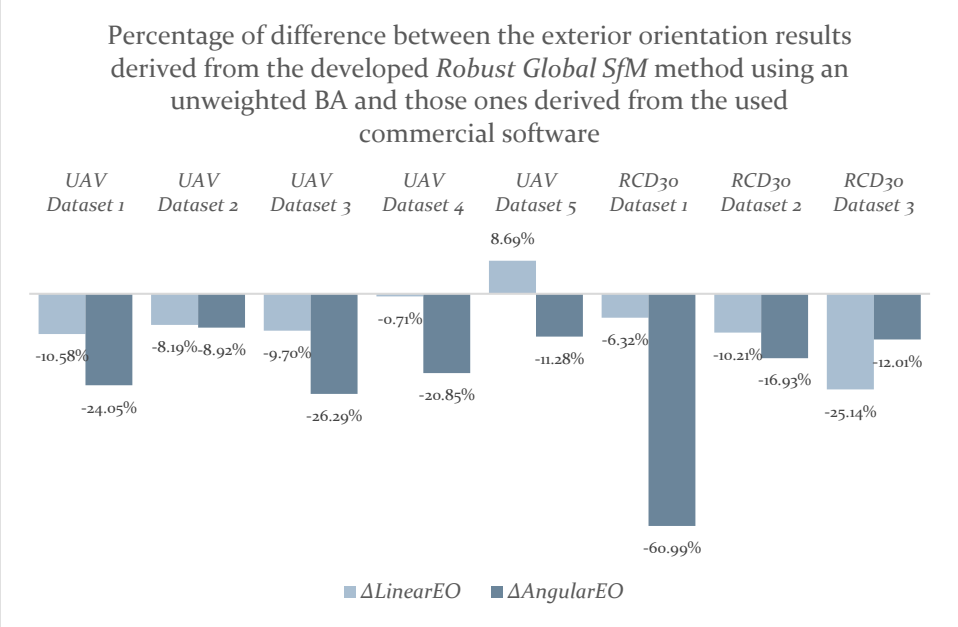


Figure 7-28. Comparison of Δ*LinearEO* and Δ*AngularEO* as calculated by an unweighted BA through *Robust Global SfM* and by commercial software for all test datasets; a negative percentage of difference implies a reduction of Δ*LinearEO* or Δ*AngularEO* for *Robust Global SfM*, that is, an improvement in the results of *Robust Global SfM*

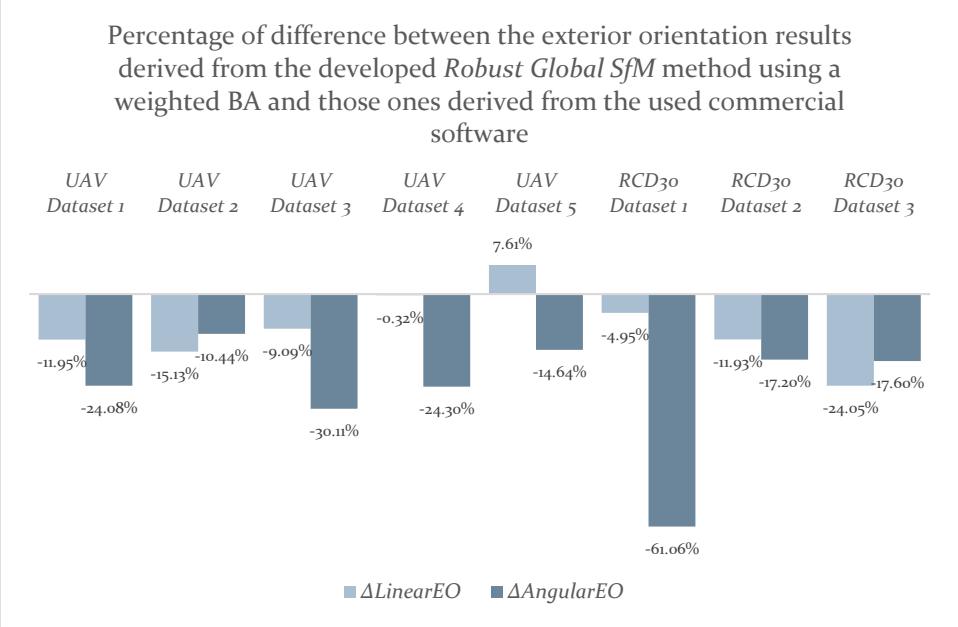


Figure 7-29. Comparison of Δ*LinearEO* and Δ*AngularEO* as calculated by a weighted BA through *Robust Global SfM* and by commercial software for all test datasets; a negative percentage of difference implies an improvement in the results of *Robust Global SfM*

observed for the dataset of 40 multi-perspective oblique aerial images (*RCD30 Dataset 2*), as there is very poor overlap between the four oblique aerial images acquired by the multi-view camera system at a single time instance. The dataset of 50 multi-perspective oblique and vertical aerial images (*RCD30 Dataset 3*) corresponds to better accuracy in the exterior orientation parameters than the oblique-only multi-perspective dataset (*RCD30 Dataset 2*), as the five images that are acquired by the multi-camera system provide a sufficiently stronger geometry to tie the side oblique aerial images of the strip together. The relative increase of the average absolute Δ*LinearEO* and Δ*AngularEO* for *RCD30 Dataset 2* compared to *RCD30 Dataset 1* are 29% and 14%, respectively; the corresponding values for *RCD30 Dataset 3* compared to *RCD30 Dataset 1* are 23% and 8% for average absolute Δ*LinearEO* and Δ*AngularEO*, respectively.

The diagrams depicted in Figure 7-27, Figure 7-28 and Figure 7-29 illustrate the differences of the exterior orientation results as derived by *Robust Global SfM*, using bundle adjustment both without and with weights, and the commercial software solutions. Specifically, Figure 7-27 shows the average absolute Δ*LinearEO* and Δ*AngularEO* for *Robust Global SfM* and the commercial software solutions in the form of bar charts, which illustrate the overall reduction of Δ*LinearEO* and Δ*AngularEO* for the developed *Robust Global SfM* solution, either using a weighted bundle adjustment or assuming equal weights for point observations in bundle adjustment. The improved results of the developed *Robust Global SfM* solution are also verified by the bar charts of Figure 7-28 and Figure 7-29, which illustrate the percentage of difference between the exterior orientation results derived by *Robust Global SfM* (with the Figure 7-28 corresponding to the case of unweighted bundle adjustment and Figure 7-29 corresponding to a weighted bundle adjustment) and those ones derived by commercial software. With the only exception of average absolute Δ*LinearEO* of *UAV Dataset 5*, which is smaller in the case of commercial software, the rest of average absolute

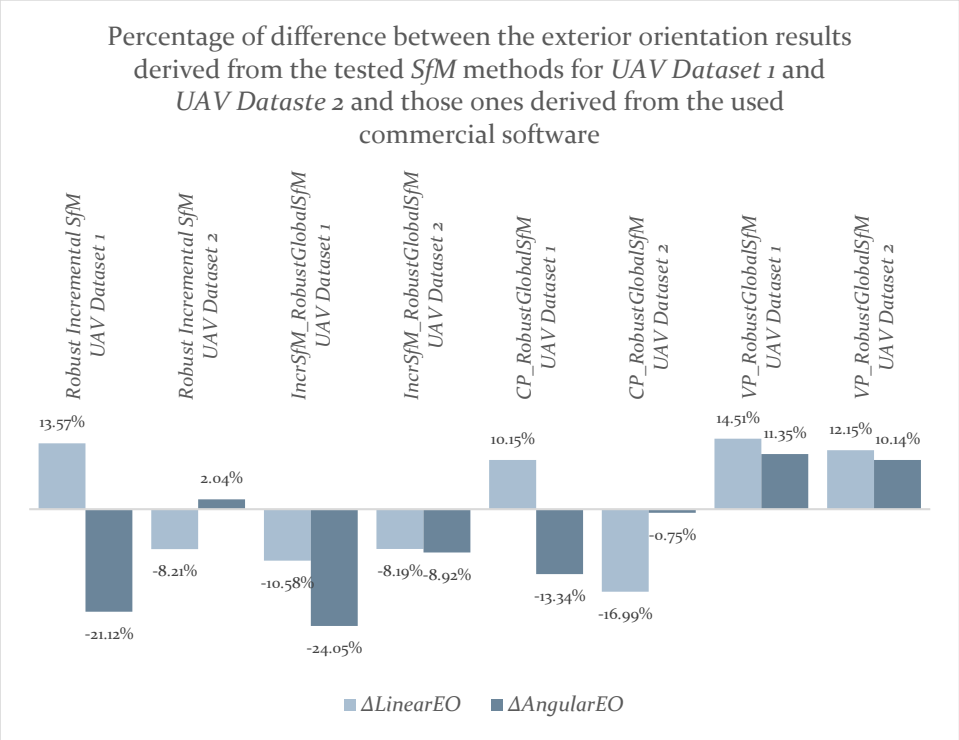


Figure 7-30. Comparison of Δ*LinearEO* and Δ*AngularEO* as calculated by the developed SfM methods (using the SURF feature extractor, unweighted BA and four GCP measurements in *img_1*) and by commercial software for *UAV Dataset 1* and *UAV Dataset 2*; a negative percentage of difference implies a reduction of Δ*LinearEO* or Δ*AngularEO* for the tested SfM methods, that is, an improvement in the results of the tested SfM methods

Δ*LinearEO* and Δ*AngularEO* are improved in the case of the developed *Robust Global SfM* method. The overall improvement of average absolute Δ*LinearEO* for *Robust Global SfM* reaches the percentage of 8%, in the case of unweighted bundle adjustment, and 9%, in the case of weighted bundle adjustment; the corresponding improvement of Δ*AngularEO* for *Robust Global SfM* reaches the percentage of 23% for unweighted bundle adjustment and 25% for weighted bundle adjustment. Thus, the developed *Robust Global SfM* framework is a robust alternative to existing commercial software solutions. Finally, bigger systematic error is derived by the commercial software for all datasets, which is proved by the bigger absolute values of the average differences between the computed and the reference exterior orientation parameters, compared to the results achieved through the developed *Robust Global SfM* solution.

Finally, the bar chart of Figure 7-30 illustrates the percentage of difference between the exterior orientation results derived by all exterior orientation methods presented in this dissertation (*Robust Incremental SfM*, *IncrSfm_RobustGlobalSfM*, *CP_RobustGlobalSfM*, *VP_RobustGlobalSfM*) and those ones derived by commercial software for *UAV Dataset 1* and *UAV Dataset 2*, applying an unweighted bundle adjustment and the SURF feature extractor, along with four GCP measurements in *img_1*. The results of comparison with commercial software verify the already derived conclusions. Specifically, *VP_RobustGlobalSfM* yields greater values of average absolute Δ*LinearEO* and Δ*AngularEO* for both datasets compared to the commercial software results, while *CP_RobustGlobalSfM* and *Robust Incremental SfM* yield equivalent results with the commercial software. The results of *IncrSfM_RobustGlobalSfM* correspond to smaller average absolute Δ*LinearEO* and Δ*AngularEO* compared to the commercial solution for both datasets, as already mentioned.

### 7.10.9 Further Discussion

In the previous sections, the impact of the SfM workflow and the method for obtaining initial values, the feature extraction algorithm (SURF and ASIFT) and the scale-based weighting strategy on the exterior orientation results have been assessed, assuming the worst scenario of availability of four GCP measurements in one image. Similar results are also derived using more GCP measurements. For instance, the impact of the iterative robust outlier removal framework and the weighting strategy in bundle adjustment are verified in this section by experiments with *UAV Dataset 3* using measurements of 4 GCPs in the starting image of the sequence and 4 different GCPs in the last image of the sequence. Furthermore, tests assuming a different scale-based weighting strategy have also been implemented and the results achieved for *UAV Dataset 4* are outlined.

#### 7.10.9.1 Impact of the Number of GCPs

The results achieved for *UAV Dataset 3* via the developed *Robust Global SfM* framework using the same four GCP measurements in the starting image of the sequence and measuring also another set of four coplanar GCPs in the last image of the sequence are presented in Table 7-27. As already mentioned in section 7.10.3.2, the four coplanar GCPs that were measured in *img_1* have been automatically recognized in three images using the developed coplanar point detection methodology; the four coplanar GCPs that were measured in the last image of the sequence have been automatically recognized in two images. Table 7-27 also outlines the results derived by commercial software (Pix4DMapper) using the aforementioned eight GCPs, manually measured in four images (two images for each set of four GCPs, which is the least possible number of images for which the GCPs have to be measured).

Table 7-27. Differences between the EO parameters computed through the developed SfM solutions using bundle adjustment with and without weights, the SURF feature extractor and eight GCP measurements and through commercial software and the reference ones for *UAV Dataset 3*

| ***UAV Dataset 3*: 60 oblique UAV images** | | | | | | |
|---|---|---|---|---|---|---|
| **Method** | **Metric** | **Avg-Abs** | **Max-Abs** | **Min-Abs** | **Avg** | **Stdev** |
| ***Incremental SfM* (SURF; unweighted BA; 8 GCPs in two images)** | $\Delta X_o$ (m) | 0.196 | 0.583 | 0.005 | 0.085 | 0.216 |
| | $\Delta Y_o$ (m) | 0.201 | 0.566 | 0.003 | 0.156 | 0.180 |
| | $\Delta Z_o$ (m) | 0.291 | 0.578 | 0.063 | -0.005 | 0.323 |
| | $\Delta\omega$ (deg.) | 0.788 | 1.449 | 0.111 | 0.270 | 0.767 |
| | $\Delta\varphi$ (deg.) | 0.435 | 1.037 | 0.013 | -0.417 | 0.316 |
| | $\Delta\kappa$ (deg.) | 0.263 | 1.146 | 0.002 | -0.043 | 0.335 |
| | $\Delta$*LinearEO* (m) | 0.229 | 0.583 | 0.003 | 0.079 | 0.240 |
| | $\Delta$*AngularEO* (deg.) | 0.495 | 1.449 | 0.002 | -0.063 | 0.472 |
| ***Robust Global SfM:* initial EO from *Incremental SfM* (SURF; unweighted BA; 8 GCPs in two images)** | $\Delta X_o$ (m) | 0.201 | 0.360 | 0.009 | 0.091 | 0.213 |
| | $\Delta Y_o$ (m) | 0.212 | 0.396 | 0.002 | 0.212 | 0.091 |
| | $\Delta Z_o$ (m) | 0.162 | 0.352 | 0.002 | 0.008 | 0.180 |
| | $\Delta\omega$ (deg.) | 0.587 | 0.790 | 0.102 | 0.169 | 0.586 |
| | $\Delta\varphi$ (deg.) | 0.456 | 0.861 | 0.012 | -0.456 | 0.188 |
| | $\Delta\kappa$ (deg.) | 0.167 | 0.507 | 0.004 | -0.087 | 0.193 |
| | $\Delta$*LinearEO* (m) | 0.192 | 0.396 | 0.002 | 0.104 | 0.161 |
| | $\Delta$*AngularEO* (deg.) | 0.403 | 0.861 | 0.004 | -0.124 | 0.322 |
| ***Robust Global SfM:* initial EO from *Incremental SfM* (SURF; weighted BA; 8 GCPs in two images)** | $\Delta X_o$ (m) | 0.219 | 0.409 | 0.006 | 0.124 | 0.217 |
| | $\Delta Y_o$ (m) | 0.194 | 0.377 | 0.020 | 0.191 | 0.096 |
| | $\Delta Z_o$ (m) | 0.181 | 0.410 | 0.004 | 0.046 | 0.203 |
| | $\Delta\omega$ (deg.) | 0.569 | 0.787 | 0.085 | 0.198 | 0.563 |
| | $\Delta\varphi$ (deg.) | 0.392 | 0.823 | 0.003 | -0.392 | 0.196 |
| | $\Delta\kappa$ (deg.) | 0.137 | 0.405 | 0.005 | -0.056 | 0.162 |
| | $\Delta$*LinearEO* (m) | 0.198 | 0.410 | 0.004 | 0.120 | 0.172 |
| | $\Delta$*AngularEO* (deg.) | 0.366 | 0.823 | 0.003 | -0.083 | 0.307 |
| **Commercial Software (8 GCPs in four images)** | $\Delta X_o$ (m) | 0.135 | 0.256 | 0.001 | 0.105 | 0.113 |
| | $\Delta Y_o$ (m) | 0.156 | 0.384 | 0.017 | 0.156 | 0.090 |
| | $\Delta Z_o$ (m) | 0.142 | 0.236 | 0.008 | -0.142 | 0.052 |
| | $\Delta\omega$ (deg.) | 0.551 | 0.982 | 0.057 | 0.208 | 0.534 |
| | $\Delta\varphi$ (deg.) | 0.552 | 0.918 | 0.106 | -0.552 | 0.308 |
| | $\Delta\kappa$ (deg.) | 0.185 | 0.681 | 0.000 | -0.037 | 0.254 |
| | $\Delta$*LinearEO* (m) | 0.144 | 0.384 | 0.001 | 0.040 | 0.085 |
| | $\Delta$*AngularEO* (deg.) | 0.429 | 0.982 | 0.000 | -0.127 | 0.365 |

The results presented in Table 7-27 verify the conclusion already derived by the results achieved using four coplanar GCPs, that the robust iterative outlier removal framework improves the exterior orientation results. Specifically, the average absolute $\Delta$*LinearEO* of *Robust Global SfM* has been improved by 16% in the case of a weighted bundle adjustment and 14% in the case of weighted bundle adjustment compared to the corresponding values of *Incremental SfM* without any iterative bundle adjustment technique for outlier removal; the corresponding improvement of average absolute $\Delta$*AngularEO* reaches the percentage of 19% for the case of unweighted bundle adjustment and 26% for the case of weighted bundle adjustment. Moreover, the results of weighted bundle adjustment are generally improved compared to those ones derived by an unweighted bundle

Table 7-28. Differences between the EO parameters computed through *Robust Global SfM*, using as initial EO parameters the ones calculated by *Incremental SfM* (using the SURF feature extractor, weighted BA using a different weighting strategy and four GCP measurements in *img_1*) for *UAV Dataset 4*

| ***UAV Dataset 4*: 80 oblique UAV images** | | | | | | |
|---|---|---|---|---|---|---|
| **Method** | **Metric** | **Avg-Abs** | **Max-Abs** | **Min-Abs** | **Avg** | **Stdev** |
| ***Robust Global SfM:* initial EO from *Incremental SfM* (SURF; weighted BA – different weights; 4 GCPs in one image)** | $\Delta X_o$ (m) | 0.330 | 0.933 | 0.001 | 0.075 | 0.401 |
| | $\Delta Y_o$ (m) | 0.248 | 0.782 | 0.010 | 0.100 | 0.286 |
| | $\Delta Z_o$ (m) | 0.399 | 1.101 | 0.002 | -0.086 | 0.459 |
| | $\Delta\omega$ (deg.) | 0.026 | 0.110 | 0.000 | -0.009 | 0.032 |
| | $\Delta\varphi$ (deg.) | 0.036 | 0.098 | 0.000 | 0.017 | 0.041 |
| | $\Delta\kappa$ (deg.) | 0.040 | 0.087 | 0.000 | 0.005 | 0.045 |
| | $\Delta LinearEO$ (m) | 0.326 | 1.101 | 0.001 | 0.030 | 0.382 |
| | $\Delta AngularEO$ (deg.) | 0.034 | 0.110 | 0.000 | 0.004 | 0.039 |

adjustment; whereas the average absolute $\Delta LinearEO$ is approximately the same for both weighted and unweighted bundle adjustment, the average absolute $\Delta AngularEO$ has been improved by 9% in the case of weighted bundle adjustment.

The results of *Robust Global SfM* using both unweighted and weighted bundle adjustment show an overall improvement thanks to the measurement of another set of four GCPs, in comparison with the corresponding results achieved through *Robust Global SfM* using measurements of four GCPs. This is an expected conclusion, taking into account the improved distribution of GCPs. For instance, the average absolute $\Delta LinearEO$ has been reduced by 46% for the case of *Robust Global SfM* with weighted bundle adjustment and eight GCP measurements; the average absolute $\Delta AngularEO$ shows an increase by a smaller percentage (26%) for the case of *Robust Global SfM* with weighted bundle adjustment and eight GCP measurements. Similar results are derived by the commercial software solution; its results correspond to significantly smaller $\Delta LinearEO$ for the case of 8 GCPs (reduction by 64%) and a slightly bigger $\Delta AngularEO$ (increase by 3%). Finally, the results achieved by the developed solution and the commercial software package are comparable, showing an improved $\Delta LinearEO$ in the case of the commercial software and an improved $\Delta AngularEO$ in the case of the developed *Robust Global SfM* method.

### 7.10.9.2 Investigation of the Results Achieved via Another Weighting Strategy

In section 7.6, the weight of each point observation in a bundle adjustment process has been defined as the product of a constant with the scale of this point; this constant may be calculated as the inverse scale of a random point, in order to normalize the weights. Tests have also been conducted using another weighting strategy, according to which the weight of each point is defined as the product of a constant with the square of the scale of this point. In this case, this constant may be defined as the inverse of the square of scale of a random point, for the scope of normalizing the weights to be close to the unit, and the weight is given by equation (7.9) (for the definition of symbols, see section 7.6).

$$p_T = \frac{\left(Z_0 - Z_R\right)^2 + \left(X_0 - X_R\right)^2 + \left(Y_0 - Y_R\right)^2}{c^2 + \left(x_{R'} - x_0\right)^2 + \left(y_{R'} - y_0\right)^2} \cdot \frac{c^2 + \left(x_{T'} - x_0\right)^2 + \left(y_{T'} - y_0\right)^2}{\left(Z_0 - Z_T\right)^2 + \left(X_0 - X_T\right)^2 + \left(Y_0 - Y_T\right)^2} \tag{7.9}$$

Table 7-29. Average absolute Δ*AngularEO* and average $D_{quaternions}$ for some experiments performed using all test datasets via *Robust Global SfM* initialized by georeferencing metadata (for the RCD30 datasets) or *Incremental SfM* (for the UAV datasets), along with percentage of difference of average absolute Δ*AngularEO* and average $D_{quaternions}$ between each tested method and commercial software; a negative percentage shows an improvement of the results obtained by the developed SfM methods

| **Dataset and method/ parameterization** | **Results of different metrics** | | **Percentage of difference of the metrics between each developed solution and the commercial software** | |
|---|---|---|---|---|
| | **Avg-Abs Δ*AngularEO*** | **Avg $D_{quaternions}$** | **Avg-Abs Δ*AngularEO*** | **Avg $D_{quaternions}$** |
| ***RCD30 Dataset 1* (SURF, unweighted BA)** | 0.014 deg. | $2.667 \cdot 10^{-4}$ | -61% | -56% |
| ***RCD30 Dataset 2* (SURF, unweighted BA)** | 0.034 deg. | $5.474 \cdot 10^{-4}$ | -17% | -20% |
| ***RCD30 Dataset 3* (SURF, unweighted BA)** | 0.034 deg. | $5.919 \cdot 10^{-4}$ | -12% | -11% |
| ***RCD30 Dataset 1* (SURF, weighted BA)** | 0.014 deg. | $2.620 \cdot 10^{-4}$ | -61% | -56% |
| ***RCD30 Dataset 2* (SURF, weighted BA)** | 0.034 deg. | $5.415 \cdot 10^{-4}$ | -17% | -21% |
| ***RCD30 Dataset 3* (SURF, weighted BA)** | 0.032 deg. | $5.608 \cdot 10^{-4}$ | -18% | -16% |
| ***RCD30 Dataset 1* (ASIFT, unweighted BA)** | 0.016 deg. | $2.906 \cdot 10^{-4}$ | -55% | -52% |
| ***RCD30 Dataset 2* (ASIFT, unweighted BA)** | 0.045 deg. | $7.419 \cdot 10^{-4}$ | 11% | 8% |
| ***RCD30 Dataset 3* (ASIFT, unweighted BA)** | 0.034 deg. | $5.739 \cdot 10^{-4}$ | -13% | -14% |
| ***RCD30 Dataset 1* (commercial software)** | 0.036 deg. | $6.009 \cdot 10^{-4}$ | | |
| ***RCD30 Dataset 2* (commercial software** | 0.041 deg. | $6.849 \cdot 10^{-4}$ | | |
| ***RCD30 Dataset 3* (commercial software)** | 0.039 deg. | $6.641 \cdot 10^{-4}$ | | |
| ***UAV Dataset 1* (SURF, unweighted BA)** | 0.277 deg. | $49.495 \cdot 10^{-4}$ | -24% | -21% |
| ***UAV Dataset 2* (SURF, weighted BA)** | 0.332 deg. | $57.196 \cdot 10^{-4}$ | -10% | -10% |
| ***UAV Dataset 3* (SURF, weighted BA)** | 0.291 deg. | $47.295 \cdot 10^{-4}$ | -30% | -36% |
| ***UAV Dataset 4* (SURF, weighted BA)** | 0.323 deg. | $58.116 \cdot 10^{-4}$ | -24% | -26% |
| ***UAV Dataset 5* (SURF, unweighted BA)** | 0.373 deg. | $69.260 \cdot 10^{-4}$ | -11% | -10% |
| ***UAV Dataset 1* (commercial software)** | 0.365 deg. | $62.271 \cdot 10^{-4}$ | | |
| ***UAV Dataset 2* (commercial software)** | 0.370 deg. | $63.850 \cdot 10^{-4}$ | | |
| ***UAV Dataset 3* (commercial software)** | 0.417 deg. | $74.471 \cdot 10^{-4}$ | | |
| ***UAV Dataset 4* (commercial software)** | 0.427 deg. | $78.945 \cdot 10^{-4}$ | | |
| ***UAV Dataset 5* (commercial software)** | 0.420 deg. | $77.143 \cdot 10^{-4}$ | | |

Experiments using the aforementioned weighting strategy have been conducted for all UAV datasets, for which the results in the case of the scale-based weighted bundle adjustment using the approach presented in section 7.6 have been improved, compared to those ones derived using an unweighted bundle adjustment. The aforementioned weighting strategy that differentiates the weights of point observations more than the strategy presented in section 7.6 yields worse results. For instance, Table 7-28 shows the results achieved through *Robust Global SfM* for *UAV Dataset 4* by calculating weights in the final bundle adjustment according to equation (7.9). The average absolute Δ*LinearEO* and Δ*AngularEO* are greater by 5% and 4%, respectively, compared to those ones that correspond to the developed weighting strategy, presented in section 7.6.

## 7.10.10 Quaternion Distance versus Difference of Euler Angles

The evaluation of methods that compute camera exterior orientatation parameters is a notoriously complex issue, because of the fact that the parameters are correlated with each other and possibly with other parameters as well. A typical evaluation of orientation methods is conducted through comparison of the 3D ground coordinates of points, obtained using the camera interior and exterior orientation parameters, with reference data. However, in the context of this dissertation, a more direct evaluation has been preferred, according to which, the computed exterior orientation parameters were compared with their reference values.

In the previous sections, the average absolute difference Δ*AngularEO* between the angular exterior orientation parameters computed through one of the tested SfM methods or commercial solution and the reference ones in terms of Euler angles were outlined and analyzed. In this section, the average distance $D_{quaternions}$ between the unit quaternions that correspond to the Euler angles ($\omega$, $\varphi$, $\kappa$) computed by a tested SfM method or commercial software and the unit quaternions that correspond to the reference Euler angles (section 7.10.4, equations (7.7) and (7.8)) is outlined for some of the experiments performed in the previous sections, along with average absolute

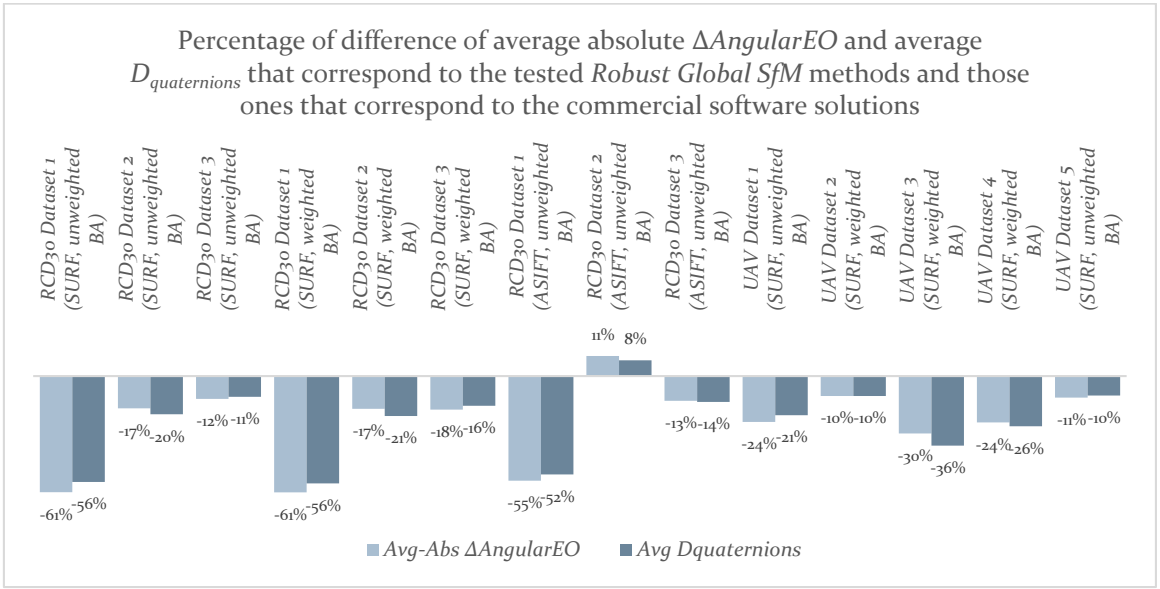


Figure 7-31. Comparison of average absolute Δ*AngularEO* and $D_{quaternions}$ as calculated by *Robust Global SfM* initialized by georeferencing metadata (for the RCD30 datasets) or *Incremental SfM* (for the UAV datasets) (using four GCP measurements in *img_1*) and by the commercial software solutions for all test datasets; a negative percentage of difference implies an improvement in the results of the *Robust Global SfM* methods

$\Delta AngularEO$, for comparison reasons. Table 7-29 shows these results, along with the percentage of difference between the average absolute $\Delta AngularEO$ and average $D_{quaternions}$ that correspond to the tested developed SfM methods and those ones that correspond to the commercial software, in order to compare the results of these metrics. These percentages reveal negligible relative difference between the average absolute $\Delta AngularEO$, computed from the Euler angles, and $D_{quaternions}$, computed from the quaternions that correspond to the Euler angles. Hence, the analysis of the angular exterior orientation results that has been made in the previous sections using absolute difference of Euler angles would be the same if a quaternion representation had been adopted. The diagram illustrated in Figure 7-31 verifies these conclusions.

## 7.11 Conclusions

In this chapter, photogrammetry-based incremental and global SfM workflows that adopt robust methods for outlier removal in challenging datasets of oblique and vertical images have been presented. The developed SfM techniques require the measurement of four coplanar GCPs in only one image. The experiments that have been performed using datasets of oblique images acquired both by manned and unmanned aerial platforms investigate the impact that each of the following has on the exterior orientation results: (i) the SfM workflow (global or incremental SfM) and the method used for obtaining initial exterior orientation parameters in the case of a global SfM workflow; (ii) the removal of erroneous matches that are not detected during the image matching process; (iii) a scale-based weighting strategy for automatic tie point measurements during bundle adjustment; (iv) the feature detection algorithm; and (v) the number of GCPs. The conclusions that may be derived by the experiments performed in this chapter are summarized in the following.

- Among the geometry-based methods for calculating approximate exterior orientation parameters, the coplanar point detection method yields more accurate results than the vanishing point detection method, mainly due to the fact that the errors arising from the coplanar point detection technique are included in the vanishing point detection method, which applies the former for georeferencing.
- Better approximate exterior orientation parameters used as input in a *Global SfM* or *Robust Global SfM* method result in more accurate exterior orientation estimation.
- Similar exterior orientation results are achieved through *Robust Incremental SfM* and through *Robust Global SfM* initialized via (i) *Incremental SfM* and (ii) *CP-based EO Estimation* followed by *Global SfM*.
- Whereas for small datasets (up to 35 images, as tested in the context of this dissertation), the *Incremental SfM* method and the *CP-based EO Estimation* method followed by *Global SfM* yield similar approximate exterior orientation parameters, the implementation of an *Incremental SfM* method is a "safer" technique, mainly for large datasets, because it does not impose any geometric constraint on the imaged scene. On the contrary, the *CP-based EO Estimation* method is only applied in datasets for which every image depicts a planar surface that lies on the same plane with the one where the GCPs lie. However, the coplanar point detection method is ideal for being applied in at least a pair of images that includes *img_1*, for the automatic transfer of GCPs, as well as for the starting pair of an *Incremental SfM* method for initialization of its exterior orientation parameters.
- Georeferencing metadata from onboard sensors may also be used for computation of approximate exterior orientation parameters for input in a *Global SfM* or *Robust Global SfM* method.

- A significant improvement of the exterior orientation results is observed thanks to the elimination of all erroneous tie points by the iterative bundle adjustment framework of *Robust Global SfM* or *Robust Incremental SfM*. In other words, the developed *Robust Global SfM* and *Robust Incremental SfM* methods yield more accurate exterior orientation parameters than the corresponding *Global SfM* and *Incremental SfM* methods, in addition to producing a more clear outlier-free sparse point cloud.
- The elimination of all erroneous matches through the developed robust iterative outlier removal framework is a prerequisite for the application of a scale-based weighting strategy in bundle block adjustment.
- The linear exterior orientation parameters estimated by a SfM method are more accurate for images that are close to the image that includes GCP measurements.
- The combined aerial triangulation of oblique and vertical images acquired by the same multi-camera system yields more accurate results than the aerial triangulation of the corresponding oblique-only dataset, as the total dataset of images that includes the nadir ones provides a sufficiently stronger geometry to tie the side oblique aerial images of the block together.
- The SURF feature extractor is preferable for being used in SfM processes for exterior orientation estimation of datasets containing oblique aerial images compared to ASIFT, which is significantly slower and yields results of equal or inferior accuracy compared to SURF.
- The developed scale-based weighting strategy in the bundle block adjustment of oblique aerial images improves the exterior orientation results of images that correspond to a big percentage of overlap and include a significant number of feature point observations among most pairs of overlapping images (e.g., an average of five images depicting the same feature point). The use of a scale-based weighting strategy is not suitable in the case of datasets for which a small number of images (e.g., an average of three images) depict the same 3D feature point, either due to small overlap or due to different perspective between images that makes the image matching process challenging. For such kind of datasets, some feature points that correspond to small scale will be given small weights, despite being significant to tie the whole block of images due to possible lack of other feature point correspondences throughout the images.
- The developed *Robust Incremental SfM* and *Robust Global SfM* frameworks are robust alternatives to existing commercial software solutions. The overall exterior orientation results derived by the developed *Robust SfM* methods are improved compared to the solutions derived using well-established SfM software, proving that the iterative outlier removal framework improves the exterior orientation results.
- The exterior orientation results are improved by the measurement of more GCPs with a better distribution throughout the whole block of images.

# Chapter 8
# Conclusions

## 8.1 Summary and Contribution of the Dissertation

The significant advances that took place in recent years in the field of digital technology have created the conditions for the impressive evolution of the scientific domains of photogrammetry and computer vision. Their cognitive field is constantly expanding and their methodologies are evolving to meet the demands of modern applications. Thus, while in the past the process of extracting metric information from aerial images was based almost exclusively on vertical views, oblique aerial images have also come to the foreground, being not only a complementary dataset to traditional vertical images but also the basic source of information for various applications.

Georeferencing, usually in terms of exterior orientation estimation or by means of estimation of a 2D transformation model from images to the reference coordinate system, in cases that high accuracy is not the goal, is the first stage of the majority of photogrammetric processes, required for the extraction of metric information from imagery. In this context, the subject of this dissertation was the metric exploitation of datasets containing oblique aerial images with emphasis on the establishment of automatic georeferencing procedures, both in terms of exterior orientation estimation and in terms of image-to-ground 2D transformation, along with image footprint extraction and image rectification. Furthermore, this dissertation investigated how the automatic extraction of elements describing the geometry of the scene depicted in oblique aerial images can contribute in obtaining rough exterior orientation parameters, georeferencing images of a piecewise planar scene adopting a 2D homography transformation model, performing measurements of distances from single oblique images and transferring coplanar GCPs. All methods developed in this dissertation, both for single images and for multi-image datasets, require minimum user interaction and are applicable in images without approximate georeferencing metadata from onboard GNSS/INS sensors.

Specifically, a methodology for automatic rough georeferencing of large datasets of multi-perspective oblique and vertical aerial images of unknown interior orientation and flying height was presented. It uses a 2D transformation to express the relation from every image to the ground

coordinate system. It is based on image matching and the transfer of the georeferencing information among images and requires the measurement of a minimum number of points in one image and the input of their horizontal ground coordinates, which can be extracted from a free program for earth observation for rough georeferencing purposes. The effect of the selection of this image on the total accuracy was addressed and its characteristics that minimize the georeferencing error were reported. Furthermore, an error analysis depending on the number of transformations applied between an image and the ground reference system was presented and the performance of the SURF algorithm in multi-view oblique and vertical aerial images was evaluated. Finally, an investigation into the applicability of different types of 2D transformations for georeferencing purposes was made. The main output of this georeferencing procedure is the parameters that define the transformation between each image and the ground reference system. With this information as input, the images can be resampled and inserted in a GIS environment at approximately the correct position, orientation and scale, the image footprints can be exported in vector format, the image datasets may be subdivided into groups of images based on the region they cover and the images that depict a specific region can be automatically detected. Except for providing rough georeferencing information to photogrammetric as well as interpretation and visualization applications, this methodology automates the process of selecting, among a large number of images, those ones that cover a specific region.

Oblique aerial images of man-made environments are also of particular interest. Planar structures are very common in man-made environments. Thus, the majority of both oblique and vertical images of urban scenes depict at least one plane with sufficient resolution so as to be visually identifiable. What is more, oblique images depict both vertical and horizontal structures of urban environments, which are characterized by the existence of many parallel lines, usually generated by edges of buildings, roads and signs. Thus, the geometry of oblique aerial images of man-made environments favors the automatic extraction of sets of parallel lines, mainly horizontal and vertical ones, and thereby the automatic detection of vanishing points. In this context, automatic methods for detecting coplanar points and vanishing points, that can be applied in oblique aerial imagery for monocular metric exploitation and georeferencing reasons, were introduced.

Specifically, a technique for detecting coplanar feature points that lie on a user-defined plane among oblique and vertical large-scale aerial imagery depicting a piecewise planar scene was introduced. It is based on a template matching method in combination with a RANSAC scheme to iteratively estimate the homography transformation between the plane of interest as depicted in the images of each pair and works incrementally, by increasing the number of the detected feature points in each image and transferring the plane information among all images. Two georeferencing frameworks that implement the developed coplanar point detection scheme were discussed, namely, the georeferencing of image datasets of a piecewise planar scene via computation of their 2D homography transformation parameters to the reference coordinate system and the estimation of approximate exterior orientation parameters of georeferenced image datasets of a piecewise planar scene. The implementation of the coplanar point detection method within the developed georeferencing framework for constraining the detected correspondences to lie in a single plane yields satisfactory georeferencing results in large-scale image datasets of a piecewise planar scene. Also, an automatic method for extracting the nadir point of oblique aerial images through estimation of horizontal and vertical line segments and corresponding horizontal vanishing points was introduced and two applications which are based on the knowledge of the nadir point of images were discussed, namely, the measurement of vertical and horizontal distances from a single oblique aerial image and the estimation of approximate exterior orientation parameters of oblique aerial

images for which the transformation parameters to the reference coordinate system have been computed through the workflow presented in this dissertation.

The aforementioned techniques generate rough georeferencing information. The final issue addressed in this dissertation was the establishment of photogrammetry-based incremental and global SfM workflows that adopt robust methods for outlier removal, for exterior orientation estimation of datasets including oblique aerial images. The developed algorithms eliminate all the erroneous tie points through the combination of multiple checks and geometric constraints imposed during the image matching procedure and robust iterative bundle adjustment methods. Also, they adopt a framework for automatic transfer of GCPs based on single-image measurements. Furthermore, a scale-based weighting strategy for feature point observations in bundle adjustment was introduced, targeted to highly overlapping oblique imagery. While rough exterior orientation parameters required by the bundle adjustment, are calculated by the developed incremental SfM workflow, the developed global SfM workflow may use different techniques for calculating initial exterior orientation parameters, i.e., (i) an incremental SfM workflow; (ii) the developed coplanar point detection method; (iii) the developed vanishing point extraction technique; and (iv) georeferencing metadata from onboard sensors. Finally, the impact that each of the following has on the exterior orientation results is investigated throughout this dissertation: (i) the SfM workflow (global or incremental SfM) and the method used for obtaining initial exterior orientation parameters, in the case of global SfM; (ii) the removal of erroneous matches not detected during the image matching process; (iii) the feature detection algorithm; (iv) the weighting strategy for automatic tie point measurements during bundle adjustment; and (v) the number of GCPs.

## 8.2 Conclusions

In this section, the basic conclusions derived from the research carried out in the context of this dissertation are presented. These conclusions concern the developed 2D georeferencing frameworks, the developed methods for extracting coplanar points as well as horizontal and vertical lines and corresponding vanishing points and their applications in georeferencing procedures, as well as the developed SfM approaches. Whereas some methods presented in this dissertation may also be applied to vertical images, some others are applicable only to oblique aerial images. These methods are the following: (i) the vanishing point detection method (which locates horizontal and vertical lines and corresponding vanishing points on a single image); (ii) the *VP-based EO Estimation* method, which calculates the exterior orientation parameters of image datasets based on the developed vanishing point detection method; (iii) the measurement of vertical and horizontal distances from a single image; (iv) the scale-based weighting strategy for image feature points during bundle adjustment; and (v) the *VP_RobustGlobalSfM* method, which consist of a *VP-based EO Estimation* method followed by *Global SfM* and *Robust Global SfM*. In addition, whereas the developed robust iterative outlier removal framework can also be applied to vertical images, it is particularly beneficiary for the orientation of challenging datasets of oblique aerial images that capture a scene under different perspectives or combination of vertical and oblique images. Finally, the results achieved using the specific feature detectors tested in this dissertation (SURF and ASIFT) and the corresponding conclusions concern datasets that include oblique aerial images. All developed algorithms were tested using datasets that contain oblique aerial images.

The results of the developed 2D rough georeferencing framework that does not use any coplanarity constraint on the extracted feature points, depend greatly on the definition of the proper transformation model between the starting image and the ground reference system, as well as

between the other images, for the transfer of georeferencing information. A homography yields satisfactory results in cases of small elevation differences in the regions depicted in the images compared with the flying height. An affine transformation may be applied if the image footprints can be approximated by a rectangle. However, these transformations are not applicable if the images cover regions with large elevation differences. For achieving the best possible results, the starting image, where the GCPs are measured, should have a large number of corresponding images, a small elevation difference and a central (rather than a more peripheral) position in the block. This georeferencing framework may be implemented for image datasets that do not depict any planar surfaces easily recognizable (e.g., for small-scale oblique and vertical images aerial images).

However, in cases of large-scale images of a piecewise planar scene (e.g., UAV oblique or vertical images depicting man-made environments) accompanied with measurements of at least four coplanar GCPs in one image, the developed homography-based georeferencing framework that uses only coplanar corresponding feature points, extracted through the method introduced in this dissertation, should be applied. The implementation of the coplanarity constraint for feature point correspondences throughout the entire dataset yields satisfactory georeferencing results in case of a piecewise planar scene. In this way, the fields of application of the developed 2D georeferencing framework are not restricted to image datasets that depict regions with small elevation differences compared with the flying height or datasets with image footprints that can be approximated by a rectangle. On the contrary, thanks to the proposed constraint, it can be applied in image datasets regardless of the elevation difference of the scene, provided that all images depict a planar area with the same elevation as the elevation of the coplanar GCPs. Furthermore, according to both georeferencing frameworks, the error of a georeferenced image increases with the number of transformations applied between this image and the ground reference system.

As far as the developed geometry-based methods for automatic extraction of coplanar points as well as horizontal and vertical lines and corresponding vanishing points are concerned, satisfying results were reported. The coplanar point detection method yields successful results in terms of automatically detecting points that lie on a user-defined plane among overlapping aerial images of similar perspective. Also, the vanishing point detection method yields successful results in terms of automatically detecting horizontal and vertical lines in single oblique aerial images of man-made environments. The visual inspection of the results in 35 oblique UAV images proved that more than 99% of the feature points identified as lying on a specific user-defined plane were correct and the average percentage of correctly identified vertical and horizontal line segments was above 95%. The nadir point that is automatically estimated through the scheme introduced in this dissertation can be used in the developed automatic framework for obtaining vertical and horizontal distances from a single oblique image, in the case that high accuracy is not the goal; in such case, nominal values for the camera interior orientation parameters (focal length and principal point) can also be used or, alternatively, the camera constant can be roughly estimated by the method discussed in this dissertation, assuming a principal point at the center of the image. However, the usage of calibrated interior orientation parameters improves the accuracy of measurements of both horizontal and vertical distances compared to the case where the nominal camera parameters are used.

As far as the contribution of both geometry-based methods in SfM procedures are concerned, they can be used for the estimation of approximate exterior orientation parameters for being refined in a global SfM method. The rough exterior orientation parameters that are computed through *VP-based EO Estimation* correspond to bigger error compared to the ones that can be estimated through *CP-based EO Estimation*, due to the fact that the errors arising from the developed coplanar point detection technique are included in the *VP-based EO Estimation* method. The error of the

approximate exterior orientation parameters estimated through both methods increases with the number of transformations involved between each image and the reference coordinate system, during the 2D georeferencing procedure. As a result, none of these methods should be implemented in large image datasets, for which the images that are far away from the one that includes GCP measurements are georeferenced using a big number of transformations (e.g., more than 20 transformations). In that case, the corresponding error in the exterior orientation parameters would be large and thus prevent these values from being used as initial exterior orientation parameters in a bundle adjustment process. In such large datasets, these rough exterior orientation schemes would yield satisfying results if GCPs were measured in an additional image, thus enforcing the maximum total number of transformations from an image to the reference coordinate system to be less than a certain threshold. The coplanar point detection method is ideal for being applied for the automatic transfer of GCPs, based on single-image GCP measurements, as well as for computation of the exterior orientation parameters of a starting pair of images during an incremental SfM method.

As far as the developed SfM workflows are concerned, similar exterior orientation results are achieved through *Robust Incremental SfM* and through *Robust Global SfM* initialized via (i) *Incremental SfM* and (ii) *CP-based EO Estimation* followed by *Global SfM*, in cases of small datasets (e.g., up to 35 images). However, the *Incremental SfM* method is a "safer" technique, mainly for large datasets, as it does not impose any geometric constraint on the dataset. A significant improvement of the exterior orientation results, in addition to the creation of a more clear outlier-free sparse point cloud, is observed thanks to the elimination of all erroneous correspondences by the developed iterative bundle adjustment framework of *Robust Global SfM* or *Robust Incremental SfM*. However, the developed robust SfM frameworks are not suitable for very large image datasets (e.g., of the order of 1000 images per dataset), because of the serious time consumption that the iterative bundle adjustment framework for outlier removal would cause.

The elimination of all erroneous matches is a prerequisite for the implementation of a scale-based weighting strategy in bundle adjustment of oblique images. The developed weighting strategy improves the exterior orientation results of oblique aerial images that correspond to a big percentage of overlap and include a significant number of feature point observations among most pairs of overlapping images (e.g., an average of five images depicting the same feature point). However, it is not suitable in the case of datasets for which a small number of images (e.g., an average of three images) depict the same 3D feature point, either due to small overlap or due to different perspective between images, which makes the image matching process challenging. Also, the experiments that were conducted proved that the combined aerial triangulation of oblique and vertical aerial images acquired by the same multi-camera system yields more accurate results than the aerial triangulation of the corresponding oblique-only dataset. Furthermore, the SURF feature detector appears to be ideal for being used in SfM processes for exterior orientation estimation of datasets containing oblique aerial images.

Some additional conclusions were derived from the research conducted throughout this dissertation concerning SfM methods. Initially, better approximate exterior orientation parameters that are used as input in a global SfM method result in more accurate exterior orientation estimation. Also, the linear exterior orientation parameters estimated by a SfM method are more accurate for images that are close to the one that includes GCP measurements. The exterior orientation results are improved by the measurement of more GCPs with a better distribution throughout the whole block of images. Finally, the developed SfM frameworks are robust alternatives to existing commercial software solutions.

## 8.3 Future Prospects

There is, of course, room for additional work in each of the topics covered in this dissertation. Hence, in this final section, some ideas for future research are discussed. An issue that falls under the broader research area of this dissertation is the incorporation of geometric scene constraints into the triangulation of datasets containing oblique aerial images. Whereas such methods have been studied in literature (Gerke and Nyaruhuma, 2009; Gerke, 2011), reporting improvement of the block geometry and reduction of the number of necessary GCPs, the research conducted so far has dealt with the manual definition of scene constraints, which constitutes a significant limitation. Thus, an interesting future prospect would be the exploitation of the workflows introduced in this dissertation for the automatic determination of horizontal and vertical line segments and automatic detection of coplanar points, towards the automatic imposition of scene constraints in bundle adjustment of oblique aerial images and the investigation of the results that may be achieved.

An extensive research with the goal of improving the accuracy of the automatically extracted nadir point in an oblique aerial image, through additions/changes in the developed methodology combined with detailed investigation of the different parameters and thresholds, would also be interesting. In addition, the automatic transfer of coplanar GCPs among images of different perspective should also be investigated, in order to overcome the limitation that arises from the use of the template matching process, which yields successful results in highly overlapping images of similar perspective. Moreover, the automatic determination of the images where GCPs should be measured in cases of large datasets (e.g., more than 200 images) for being used in the developed SfM methods, the investigation of the results that can be achieved and the influence of self-calibration on the results are interesting aspects for future research. Also, the issue of excluding unnecessary image pairs from being used in the process of exterior orientation estimation using the developed SfM methods in cases of large datasets containing highly overlapping oblique aerial images, which is not addressed in this dissertation, is another interesting prospect for future research. Finally, the rejection of outliers in cases of very large datasets of images (e.g., of the order of 1000 images), where the developed robust iterative bundle adjustment framework would cause serious time consumption thus being inappropriate, is another issue worth being investigated.

# References


Aber, J., Marzolff, I., Ries, J. & Aber, S. (2019). *Small-format aerial photography and UAS imagery,* Amsterdam, the Netherlands, Elsevier.

Agarwal, S., Furukawa, Y., Snavely, N., Simon, I., Curless, B., Seitz, S.M. & Szeliski, R. (2011). Building Rome in a day. *Communications of the ACM,* 54, 105-112.

Agisoft LLC. (2019). *Agisoft PhotoScan Software* [Online]. Available: http://www.agisoft.com/ [Accessed 10 April 2019].

Aicardi, I., Chiabrando, F., Grasso, N., Lingua, A.M., Noardo, F. & Spanò, A. (2016). UAV photogrammetry with oblique images: First analysis on data acquisition and processing. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* XLI-B1, 835-842.

Almansa, A., Desolneux, A. & Vamech, S. (2003). Vanishing point detection without any a priori information. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 25, 502-507. Doi: 10.1109/TPAMI.2003.1190575.

American Society of Photogrammetry (1952). *Manual of Photogrammetry,* Washington, D.C., American Society of Photogrammetry.

Antone, M.E. & Teller, S. (2000). Automatic recovery of relative camera rotations for urban scenes. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2000).* Hilton Head Island, SC, USA: IEEE, 1-8. Doi: 10.1109/CVPR.2000.854809.

Antunes, M. & Barreto, J.P. (2013). A global approach for the detection of vanishing points and mutually orthogonal vanishing direction. *EEE Conference on Computer Vision and Pattern Recognition.* Portland, OR, USA: IEEE, 1336 - 1343. Doi: 10.1109/CVPR.2013.176.

Antunes, M., Barreto, J.P., Aouada, D. & Ottersten, B. (2017). Unsupervised vanishing point detection and camera calibration from a single Manhattan image with radial distortion. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).* Honolulu, HI, USA: IEEE, 4288-4296. Doi: 10.1109/CVPR.2017.708.

Arie-Nachimson, M., Kovalsky, S.Z., Kemelmacher-Shlizerman, I., Singer, A. & Basri, R. (2012). Global motion estimation from point matches. *Second International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission.* Zurich, Switzerland: IEEE, 81-88. Doi: 10.1109/3DIMPVT.2012.46.

Baatz, G., Köser, K., Chen, D., Grzeszczuk, R. & Pollefeys, M. (2010). Handling urban location recognition as a 2D homothetic problem. *In:* Daniilidis, K., Maragos, P. & Paragios, N. (eds.) *Computer Vision – ECCV 2010. ECCV 2010. Lecture Notes in Computer Science, vol. 6316.* Springer, Berlin, Heidelberg.

Babcock, E.A. (1971). Detection of active faulting using oblique infrared aerial photography in the Imperial Valley, California. *Geological Society of America Bulletin,* 82, 3189-3196.

Barnard, S.T. (1983). Interpreting perspective images. *Artificial Intelligence,* 21, 435-462. Doi: 10.1016/S0004-3702(83)80021-6.

Bartoli, A. & Sturm, P. (2003). Constrained structure and motion from multiple uncalibrated views of a piecewise planar scene. *International Journal of Computer Vision,* 52, 45-64.

Bartoli, A., Sturm, P. & Horaud, R. (2001). Projective structure and motion from two views of a piecewise planar scene. *8th IEEE International Conference on Computer Vision.* IEEE, 593-598.

Bauer, J., Sunderhauf, N. & Protzel, P. (2007). Comparing several implementations of two recently published feature detectors. *International Conference on Intelligent and Autonomous Systems.* Toulouse, France.

Bay, H., Ess, A., Tuytelaars, T. & Van Gool, L. (2008). Speeded-up robust features (SURF). *Computer Vision and Image Understanding,* 110, 346-359.

Boost Team. (2019). *Boost C++ Libraries* [Online]. Available: https://www.boost.org/ [Accessed 10 April 2019].

Børgesen, C. (2008). *Projection of oblique photos onto 3D models of buildings.* Doctoral Thesis, Technical University of Denmark.

Brocks, S. & Bareth, G. (2016). Evaluating dense 3D reconstruction software packages for oblique monitoring of crop canopy surface. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* XLI-B5, 785-789.

Buchanan, A.M. & Fitzgibbon, A.W. (2005). Damped Newton algorithms for matrix factorization with missing data. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05).* San Diego, CA, USA: IEEE, 316-322. Doi: 10.1109/CVPR.2005.118.

Canadian Topographical Survey (1928). A graphical method of plotting oblique aerial photographs. *Bulletin of the Topographical Survey.* Department of the Interior, Dominion of Canada.

Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 8, 679-698.

Caprile, B. & Torre, V. (1990). Using vanishing points for camera calibration *International Journal of Computer Vision,* 4, 127-140.

Casella, V. & Franzini, M. (2016). Modelling steep surfaces by various configurations of nadir and oblique photogrammetry. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* III-1, 175-182.

Cavegn, S., Haala, N., Nebiker, S., Rothermel, M. & Tutzauer, P. (2014). Benchmarking high density image matching for oblique airborne imagery. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* XL-3, 45-52.

Chang, C.-K., Siagian, C. & Itti, L. (2012). Mobile robot monocular vision navigation based on road region and boundary estimation. *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems.* Vilamoura, Portugal: IEEE, 1043-1050. Doi: 10.1109/IROS.2012.6385703.

Chen, H.R. & Tseng, Y.H. (2016). Study of automatic image rectification and registration of scanned historical aerial photographs. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* XLI-B8, 1229-1236. Doi: 10.5194/isprs-archives-XLI-B8-1229-2016.

Chen, P. & Suter, D. (2004). Recovering the missing components in a large noisy low-rank matrix: application to SfM. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 25, 1051-1063. Doi: 10.1109/TPAMI.2004.52.

Church, E. (1944). Perspective grid method for constructing planimetric maps from high oblique photographs. *Elements of aerial photogrammetry.* Syracuse, New York: Syracuse University Press.

Cipolla, R., Drummond, T. & Robertson, D. (1999a). Camera calibration from vanishing points in images of architectural scenes. *10th British Machine Vision Conference (BMVC 1999).* Nottingham, UK: British Machine Vision Association. Doi: 10.5244/C.13.38.

Cipolla, R., Drummond, T. & Robertson, D.P. (Year) Published. Camera calibration from vanishing points in image of architectural scenes. British Machine Vision Conference 1999 (BMVC 1999), 1999b Nottingham, UK. 382-391. Doi: 10.5244/C.13.38.

Cléry, I., Pierrot-Deseilligny, M. & Vallet, B. (2014). Automatic georeferencing of a heritage of old analog aerial photographs. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* II-3, 33-40. Doi: 10.5194/isprsannals-II-3-33-2014.

Colwell, R.N. (1997). History and place of photographic interpretation. *In:* Philipson, W.R. (ed.) *Manual of Photographic Interpretation.* 2nd ed. Bethesda, Maryland, United States: American Society for Photogrammetry and Remote Sensing (ASPRS).

Conrad, D. & DeSouza, G.N. (2010). Homography-based ground plane detection for mobile robot navigation using a modified EM algorithm. *2010 IEEE International Conference on Robotics and Automation.* Anchorage, Alaska, USA: IEEE, 910-915. Doi: 10.1109/ROBOT.2010.5509457.

Criminisi, A., Reid, I. & Zisserman, A. (2000). Single view metrology. *International Journal of Computer Vision,* 40, 123-148. Doi: 10.1023/A:1026598000963.

Cui, Z. & Tan, P. (2015). Global structure-from-motion by similarity averaging. *2015 IEEE International Conference on Computer Vision (ICCV).* 864-872. Doi: 10.1109/ICCV.2015.105.

Debnath, S., Banerjee, A. & Namboodiri, V. (2015). Adapting RANSAC SVM to detect outliers for robust classification. *In:* Xie, X., Jones, M.W. & Tam, G.K.L. (eds.) *British Machine Vision Conference (BMVC).* Swansea, UK: BMVA Press, 168.1-168.11. Doi: 10.5244/C.29.168.

Desjardins, L. (1945). Useful Graphical Constructions on Aerial Photographs. *Photogrammetric Engineering,* XI, 194-229.

Deuber, M., Cavegn, S. & Nebiker, S. (2014). Dense image matching - Performance analysis on oblique imagery. *GIM International,* 28, 23-25.

Deutscher, J., Isard, M. & MacCormick, J. (2002). Automatic camera calibration from a single Manhattan image. *In:* Heyden, A., Sparr, G., Nielsen, M. & P., J. (eds.) *Computer Vision - ECCV 2002. ECCV 2002. Lecture Notes in Computer Science, vol. 2353.* Springer, Berlin, Heidelberg.

Dewitt, B.A. (1996). Initial approximations for the three-dimensional conformal coordinate transformation. *Photogrammetric Engineering & Remote Sensing,* 62, 79-83.

Ding, M., Lyngbaek, K. & Zakhor, A. (2008). Automatic registration of aerial imagery with untextured 3D LiDAR models. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2008).* Anchorage, Alaska, USA: IEEE.

Doneus, M. (2001). Precision mapping and interpretation of oblique aerial photographs. *Archaeological Prospection,* 8, 13-27.

Doneus, M., Wieser, M., Verhoeven, G., Karel, W., Fera, M. & Pfeifer, N. (2016). Automated archiving of archaeological aerial images. *Remote Sensing,* 8, 209:1-22.

Doyle, F.J. (1970). Photographic systems for Apollo. *Photogrammetric Engineering,* 36, 1039-1044.

Edmundson, K.L., Alexandrov, O., Archinal, B.A., Becker, K.J., Becker, T.L., Kirk, R.L., Moratto, Z.M., Nefian, A.V., Richie, J.O. & Robinson, M.S. (2016). Photogrammetric processing of Apollo 15 metric camera oblique images. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* XLI-B4, 375-381. Doi: 10.5194/isprsarchives-XLI-B4-375-2016.

Eigen Team. (2019). *Eigen Library* [Online]. Available: http://eigen.tuxfamily.org/ [Accessed 10 April 2019].

Emgu CV Team. (2019). *Emgu CV* [Online]. Available: http://www.emgu.com/ [Accessed 10 April 2019].

Fagir, J., Schubert, A., Frioud, M. & Henke, D. (2017). SAR and oblique aerial optical image fusion for urban area image segmentation. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* XLII-1/W1, 639-642. Doi: 10.5194/isprs-archives-XLII-1-W1-639-2017.

Fang, M., Yue, G. & Yu, Q. (2009). The study on an application of Otsu method in Canny Operator. *2009 International Symposium on Information Processing (ISIP'09).* Huangshan, China, 109-112.

Farenzena, M., Fusiello, A. & Gherardi, R. (2009). Structure-and-motion pipeline on a hierarchical cluster tree. *2009 IEEE 12th International Conference on Computer Vision Workshops (ICCV Workshops).* Kyoto, Japan: IEEE, 1489-1496. Doi: 10.1109/ICCVW.2009.5457435.

Fischler, M. & Bolles, R. (1981). Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM,* 24, 381-395.

Fitzgerald, G. (1944). Reconnaissance mapping with Trimetrogon Photography. *Manual of Photogrammetry.* Preliminary ed. Washington, D.C.: American Society of Photogrammetry.

Floridi, V. (2016). *Implementation and evaluation of point operators for automatic sparse stereo-matching.* Master Thesis, National Technical University of Athens.

Förstner, W. (2010). Optimal vanishing point detection and rotation estimation of single images from a Legoland scene. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences,* XXXVIII, Part 3A, 157-162.

Förstner, W., Wrobel, B., Paderes, F., Craig, R., Fräser, C. & Dollo, J. (2004). Analytical photogrammetric operations. *In:* McGlone, J.C., Mikhail, E.M. & Bethel, J.S. (eds.) *Manual of Photogrammetry.* Fifth Edition ed.: ASPRS.

Fouhey, D.F., Scharstein, D. & Briggs, A.J. (2010). Multiple plane detection in image pairs using J-linkage. *20th International Conference on Pattern Recognition.* Istanbul, Turkey: IEEE, 336-339. Doi: 10.1109/ICPR.2010.91.

Frahm, J.-M., Fite-Georgel, P., Gallup, D., Johnson, T., Raguram, R., Wu, C., Jen, Y.-H., Dunn, E., Clipp, B., Lazebnik, S. & Pollefeys, M. (2010). Building Rome on a cloudless day. *In:* Daniilidis, K., Maragos, P. & Paragios, N. (eds.) *Computer Vision – ECCV 2010. ECCV 2010. Lecture Notes in Computer Science, vol. 6314.* Springer, Berlin, Heidelberg.

Fraundorfer, F., Schindler, K. & Bischof, H. (2006). Piecewise planar scene reconstruction from sparse correspondences. *Image and Vision Computing,* 24, 395-406. Doi: 10.1016/j.imavis.2006.01.007.

Fritsch, D., Kremer, J. & Grimm, A. (2012). Towards all-in-one photogrammetry. *GIM International,* 26, 18-23.

Fritsch, D. & Rothermel, M. (2013). Oblique image data processing - Potential, experiences and recommendations. *In:* Fritsch, D. (ed.) *Photogrammetric Week '13.* Berlin: Wichmann.

Frommholz, D., Linkiewicz, M., Meissner, H., Dahlke, D. & Poznanska, A. (2015). Extracting semantically annotated 3D building models with textures from oblique aerial imagery. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* XL-3/W2, 53-58.

Frommholz, D., Linkiewicz, M. & Poznanska, A.M. (2016). Inlining 3D reconstruction, multi-source texture mapping and semantic analysis using oblique aerial imagery. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* XLI-B3, 605-612.

Frueh, C., Sammon, R. & Zakhor, A. (2004). Automated texture mapping of 3D city models with oblique aerial imagery. *2nd International Symposium on 3D Data Processing, Visualization and Transmission.* Thessaloniki, Greece: IEEE, 396-403.

Geniviva, A., Faulring, J. & Salvaggio, C. (2014). Automatic georeferencing of imagery from high-resolution, low-altitude, low-cost aerial platforms. *Proc. SPIE 9089, Geospatial InfoFusion and Video Analytics IV; and Motion Imagery for ISR and Situational Awareness II*, 90890D.

Georgopoulos, A. (1982). Balloon and kite photography: An historical review. *International Archives of Photogrammetry,* 24, V/1, 196-206.

Gerke, M. (2009). Dense matching in high resolution oblique airborne images. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* XXXVIII (3/W4), 77-82.

Gerke, M. (2011). Using horizontal and vertical building structure to constrain indirect sensor orientation. *ISPRS Journal of Photogrammetry and Remote Sensing,* 66, 307-316.

Gerke, M. & Kerle, N. (2011). Automatic structural seismic damage assessment with airborne oblique Pictometry imagery. *Photogrammetric Engineering & Remote Sensing,* 77, 885-898.

Gerke, M., Nex, F., Remondino, F., Jacobsen, K., Kremer, J., Karel, W., Hu, H. & Ostrowski, W. (2016). Orientation of oblique airborne image sets-experiences from the ISPRS/EuroSDR benchmark on multi-platform photogrammetry. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* XLI-B1, 185-191.

Gerke, M. & Nyaruhuma, A. (2009). Incorporating scene constraints into the triangulation of airborne oblique images. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* XXXVIII-1-4-7/W5.

Gerke, M. & Xiao, J. (2013). Supervised and unsupervised MRF based 3D scene classification in multiple view airborne oblique images. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* II-3/W3, 25-30.

Gerke, M., Xiao, J. & Nyaruhuma, A.P. (2014). Oblique Airborne Photogrammetry - Automatic building detection and verification. *GIM International,* 28, 23-25.

Gherardi, R., Farenzena, M. & Fusiello, A. (2010). Improving the efficiency of hierarchical structure-and-motion. *2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).* San Francisco, CA, USA: IEEE, 1594-1600. Doi: 10.1109/CVPR.2010.5539782.

Gleitsmann, L. & Kappas, M. (2004). Digital multi-image photogrammetry combined with oblique aerial photography enables glacier monitoring survey flights below clouds in Alaska. *2004 IEEE International Geoscience and Remote Sensing Symposium, IGARSS '04.* Anchorage, AK, USA: IEEE, 1148-1151. Doi: 10.1109/IGARSS.2004.1368617.

Gleitsmann, L. & Kappas, M. (2006). Glacier monitoring survey flights below clouds in Alaska: Oblique aerial photography utilising digital multiple-image photogrammetry to cope with adverse weather. *EARSeL eProceedings,* 5, 42-50.

Gorges, N., Hanheide, M., Christmas, W., Bauckhage, C., Sagerer, G. & Kittler, J. (2004). Mosaics from arbitrary stereo video sequences. *In:* Rasmussen, C.E., Bülthoff, H.H., Schölkopf, B. & Giese, M.A. (eds.) *Pattern Recognition. DAGM 2004. Lecture Notes in Computer Science, vol. 3175.* Springer, Berlin, Heidelberg.

Gotovac, D., Kružić, S., Gotovac, S. & Papić, V. (2017). A model for automatic geomapping of aerial images mosaic acquired by UAV. *2nd International Multidisciplinary Conference on Computer and Energy Science* Split, Croatia: IEEE, 1-6.

Graham, R.L. & Hell, P. (1985). On the history of the minimum spanning tree problem. *Annals of the History of Computing,* 7, 43-57. Doi: 10.1109/MAHC.1985.10011.

Grammatikopoulos, L. (2007). *Geometric information from single images in photogrammetry and computer vision.* Doctoral Thesis, National Technical University of Athens.

Grammatikopoulos, L., Karras, G. & Petsa, E. (2003). Camera calibration approaches using single images of man-made objects. *XIX CIPA International Symposium.* Antalya, Turkey, 216-219.

Grammatikopoulos, L., Karras, G. & Petsa, E. (2004). Camera calibration combining images with two vanishing points. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* 35, 99-104.

Grammatikopoulos, L., Karras, G. & Petsa, E. (2007). An automatic approach for camera calibration from vanishing points. *ISPRS Journal of Photogrammetry and Remote Sensing,* 62, 64-76. Doi: 10.1016/j.isprsjprs.2007.02.002.

Grammatikopoulos, L., Karras, G., Petsa, E. & Kalisperakis, I. (2006). A unified approach for automatic camera calibration from vanishing points. *The International Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* XXXVI, Part 5.

Guillou, E., Meneveaux, D., Maisel, E. & Bouatouch, K. (2000). Using vanishing points for camera calibration and coarse 3D reconstruction from a single image. *The Visual Computer,* 16, 396-410. Doi: 10.1007/PL00013394.

Haala, N., Rothermel, M. & Cavegn, S. (2015). Extracting 3D urban models from oblique aerial images. *2015 Joint Urban Remote Sensing Event (JURSE)* Lausanne, Switzerland: IEEE. Doi: 10.1109/JURSE.2015.7120479.

Habbecke, M. & Kobbelt, L. (2012). Automatic registration of oblique aerial images with cadastral maps. *In:* Kutulakos, K.N. (ed.) *Trends and Topics in Computer Vision. ECCV 2010. Lecture Notes in Computer Science, vol. 6554.* Springer, Berlin, Heidelberg.

Hamruni, A.M. (2010). *The use of oblique and vertical images for 3D urban modelling.* Doctoral Thesis, The University of Nottingham.

Hartley, R., Trumpf, J., Dai, Y. & Li, H. (2013). Rotation averaging. *International Journal of Computer Vision,* 101, 267-305. Doi: 10.1007/s11263-012-0601-0.

Hartley, R. & Zisserman, A. (2003). *Multiple View Geometry in Computer Vision,* Cambridge, UK, Cambridge University Press.

Hartley, R.I. (1997). In defense of the eight-point algorithm. *IEEE Transactions on Pattern Recognition and Machine Intelligence*, 19, 580-593.

Harwin, S. & Lucieer, A. (2012). Assessing the accuracy of georeferenced point clouds produced via multi-view stereopsis from unmanned aerial vehicle (UAV) imagery. *Remote Sensing,* 4, 1573-1599. Doi: 10.3390/rs4061573.

Hasheminasaba, M., Ebadi, H. & Sedaghat, A. (2015). An integrated RANSAC and graph based mismatch elimination approach for wide-baseline image matching. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* XL-1/W5, 297-300. Doi: 10.5194/isprsarchives-XL-1-W5-297-2015.

Hild, H., Haala, N. & Fritsch, D. (2000). A strategy for automatic image to map registration. *International Archives of Photogrammetry and Remote Sensing,* XXXIII, Part B2, 287-294.

Höhle, J. (2008). Photogrammetric measurements in oblique aerial images. *Photogrammetrie, Fernerkundung, Geoinformation* 1, 7-14.

Höhle, J. (2013). Oblique aerial images and their use in cultural heritage documentation. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* XL-5/W2, 349-354.

Hsu, Y.-C., Jhan, J.-P. & Rau, J.-Y. (2012). Facade detection in oblique aerial image using object based image analysis. *33rd Asian Conference on Remote Sensing.* Pattaya, Thailand.

Hu, H., Ding, Y., Zhu, Q., Wu, B., Xie, L. & Chen, M. (2016). Stable least-squares matching for oblique images using bound constrained optimization and a robust loss function. *ISPRS Journal of Photogrammetry and Remote Sensing,* 118, 53-67.

Hu, H., Zhu, Q., Du, Z., Zhang, Y. & Ding, Y. (2015). Reliable Spatial Relationship Constrained Feature Point Matching of Oblique Aerial Images. *Photogrammetric Engineering & Remote Sensing,* 81, 49-58.

Huachao, Y., Shubi, Z. & Yongbo, W. (2012). Robust and precise registration of oblique images based on scale-invariant feature transformation algorithm. *IEEE Geoscience and Remote Sensing Letters,* 9, 783-787.

Huynh, D.Q. (2009). Metrics for 3D Rotations: comparison and analysis. *Journal of Mathematical Imaging and Vision,* 35, 155-164. Doi: 10.1007/s10851-009-0161-2.

Hydrographic Office Navy Department & Photographic Intelligence Center - Division of Naval Intelligence - Navy Department (1945). *Introduction to Oblique Photogrammetry.*

Iwaszczuk, D. & Stilla, U. (2016). Quality assessment of building textures extracted from oblique airborne thermal imagery. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* III-1, 3-8.

Jacobsen, K. (2008). Geometry of vertical and oblique image combinations. *In:* Maktav, D. (ed.) *28th Symposium of the European Association of Remote Sensing Laboratories.* Istanbul, Turkey: IOS Press, 16-23.

Jacobsen, K. & Gerke, M. (2016). Sub-camera calibration of a penta-camera. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* XL-3/W4, 35-40.

Jaramillo, I. (2000). Mathematical rectification of aerial photographs for cadastral mapping in Bolivia. *International Archives of Photogrammetry and Remote Sensing,* XXXIII, Part B4, 394-399.

Jhan, J.-P., Li, Y.-T. & Rau, J.-Y. (2015). A modified projective transformation scheme for mosaicking multi-camera imaging system equipped on a large payload fixed-wing UAS. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* XL-3/W2, 87-93. Doi: 10.5194/isprsarchives-XL-3-W2-87-2015.

Jiang, N., Cui, Z. & Tan, P.A. (2013). A global linear method for camera pose registration. *2013 IEEE International Conference on Computer Vision (ICCV).* IEEE, 481-488 Doi: 10.1109/ICCV.2013.66.

Jiang, S. & Jiang, W. (2017a). Efficient structure from motion for oblique UAV images based on maximal spanning tree expansion. *ISPRS Journal of Photogrammetry and Remote Sensing,* 132, 140-161. Doi: 10.1016/j.isprsjprs.2017.09.004.

Jiang, S. & Jiang, W. (2017b). On-board GNSS/IMU assisted feature extraction and matching for oblique UAV images. *Remote Sensing,* 9. Doi: 10.3390/rs9080813.

Jiang, S. & Jiang, W. (2018). Efficient SfM for oblique UAV images: from match pair selection to geometrical verification. *Remote Sensing,* 10, 1246:1-22. Doi: 10.3390/rs10081246.

Juan, L. & Gwun, O. (2009). A comparison of SIFT, PCA-SIFT and SURF. *International Journal of Image Processing,* 3, 143-152.

Kähler, O. & Denzler, J. (2007). Detecting coplanar feature points in handheld image sequences. *International Conference on Computer Vision Theory and Applications (VISAPP 2007)* 447-452.

Kakogiannou, M. (2017). *Development and evaluation of automatic self-calibrating bundle block adjustment of unordered image sequences.* Diploma Thesis, National Technical University of Athens.

Kalisperakis, I. (2010). *Epipolar geometry in projective and euclidean space*. Doctoral Thesis, National Technical University of Athens.

Kang, J., Deng, F., Li, X. & Wan, F. (2016). Automatic texture reconstruction of 3D city model from oblique images. *The Internationcal Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* XLI-B1, 341-347.

Karel, W., Doneus, M., Verhoeven, G., Briese, C., Ressl, C. & Pfeifer, N. (2013). Oriental: Automatic geo-referencing and ortho-rectification of archaeological aerial photographs. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* II-5/W1, 175-180.

Karjalainen, M., Hyyppä, J. & Kuittinen, R. (2006). Determination of exterior orientation using linear features from vector maps. *The Photogrammetric Record,* 21, 329-341.

Karney, C. (2017). *GeographicLib* [Online]. Available: https://geographiclib.sourceforge.io/ [Accessed 25 September 2018].

Karras, G., Patias, P. & Petsa, E. (1993). Experiences with rectification of non-metric digital images when ground control is not available. *XV International CIPA Symposium.* Bucharest.

Karras, G. & Petsa, E. (1999). Metric information from uncalibrated single images. *CIPA International Symposium.* Olinda, Brasil.

Kaucic, R., Dano, N. & Hartley, R. (2001). Plane-based projective reconstruction. *8th IEEE International Conference on Computer Vision (ICCV 2001).* IEEE, 420-427. Doi: 10.1109/ICCV.2001.937548.

Ke, Q. & Kanade, T. (2005). Robust $L_1$ Norm Factorization in the Presence of Outliers and Missing Data by Alternative Convex Programming. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05).* San Diego, CA, USA: IEEE, 739-746. Doi: 10.1109/CVPR.2005.309.

Kemper, G., Melykuti, B. & Yu, C. (2016). Calibration procedures on oblique camera setups. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* XLI-B1, 205-209.

Kerle, N. (2016). Disasters: Risk assessment, management, and post-disaster studies using remote sensing. *In:* Thenkabail, P.S. (ed.) *Remote sensing of water resources, disasters and urban studies.* Boca Raton, Florida, United States: CRC Press.

Khan, N.Y., McCane, B. & Wyvill, G. (2011). SIFT and SURF performance evaluation against various image deformations on benchmark dataset. *IEEE International Conference on Digital Image Computing: Techniques and Applications.* Noosa, Queensland, Australia: IEEE, 501-506. Doi: 10.1109/DICTA.2011.90.

Kim, J.S. (2018). The application of near-automated georeferencing technique to a strip of historic aerial photographs in GIS. *Library Hi Tech,* 6, 43-56. Doi: 10.1108/LHT-10-2016-0115.

Kirnbauer, R., Blôschl, G., Waldhâusl, P. & Hochstöger, F. (1991). An analysis of snow cover patterns as derived from oblique aerial photographs *In:* Bergmann, H., Lang, H., Frey, W., Issler, D. & Salm, B. (eds.) *Snow, Hydrology and Forests In High Alpine Areas*

*(Proceedings of the Vienna Symposium, August 1991).* Great Britain: International Association of Hydrological Sciences.

Kong, H., Audibert, J.-Y. & Ponce, J. (2009). Vanishing point detection for road detection. *2009 IEEE Conference on Computer Vision and Pattern Recognition.* Miami, FL, USA: IEEE, 96-103. Doi: 10.1109/CVPR.2009.5206787.

Košecká, J. & Zhang, W. (2002). Video compass. *In:* Heyden A., Sparr G., Nielsen M. & P., J. (eds.) *Computer Vision — ECCV 2002. ECCV 2002. Lecture Notes in Computer Science, vol. 2353*: Springer, Berlin, Heidelberg.

Košecká, J. & Zhang, W. (2005). Extraction, matching and pose recovery based on dominant rectangular structures. *Computer Vision and Image Understanding,* 100, 274-293. Doi: 10.1016/j.cviu.2005.04.005.

Kroeger, T., Dai, D. & Van Gool, L. (2015). Joint vanishing point extraction and tracking. *IEEE Conference on Computer Vision and Pattern Recognition.* Boston, MA, USA: IEEE, 2449 - 2457. Doi: 10.1109/CVPR.2015.7298859.

Läbe, T. & Förstner, W. (2006). Automatic relative orientation of images. *5th Turkish-German Joint Geodetic Days.* Berlin.

Lane, B.B. (1950). Scales of oblique photographs. *Photogrammetric Engineering*, 409-414.

Le Besnerais, G., Sanfourche, M. & Champagnat, F. (2008). Dense height map estimation from oblique aerial image sequences. *Computer Vision and Image Understanding,* 109, 204-225. Doi: 10.1016/j.cviu.2007.07.003.

Lemmens, M. (2014a). Digital oblique aerial cameras (1): A survey of features and systems. *GIM International,* 28.

Lemmens, M. (2014b). Digital oblique aerial cameras (2): A survey of features and systems. *GIM International,* 28.

Lemmens, M., Lemmen, C. & Wubbe, M. (2007). Pictometry: Potentials for land administration. *6th FIG Regional Conference.* San José, Costa Rica.

Lepetit, V., Moreno-Noguer, F. & Fua, P. (2009). EPnP: An Accurate O(n) Solution to the PnP Problem. *International Journal of Computer Vision,* 81, 155-166.

Lerma, J.L., Navarro, S., Cabrelles, M. & Seguí, A.E. (2010). Camera calibration with baseline distance constraints. *The Photogrammetric Record,* 25, 140-158. Doi: 10.1111/j.1477-9730.2010.00579.x.

Lewis, J.G. (1945). Mechanical instruments for plotting from oblique photographs. *Photogrammetric Engineering*, 326-330.

Lezama, J., von Gioi, R.G., Randall, G. & Morel, J.-M. (2014). Finding Vanishing Points via Point Alignments in Image Primal and Dual Domains. *2014 IEEE Conference on Computer Vision and Pattern Recognition.* Columbus, OH, USA: IEEE, 509 - 515. Doi: 10.1109/CVPR.2014.72.

Li, Y. & Briggs, R. (2012). An automated system for image-to-vector georeferencing. *Cartography and Geographic Information Science,* 39, 199-217. Doi: 10.1559/152304063941199.

Liew, L.H., Lee, B.Y., Wang, Y.C. & Cheah, W.S. (2014). Rectification of aerial images using piecewise linear transformation. *IOP Conference Series: Earth and Environmental Science,* 18, 1-6. Doi: 10.1088/1755-1315/18/1/012009.

Liew, L.H., Wang, Y.C. & Cheah, W.S. (2012). Evaluation of control points' distribution on distortions and geometric transformations for aerial images rectification *Procedia Engineering,* 41, 1002-1008. Doi: 10.1016/j.proeng.2012.07.275.

Lin, C. & Nevatia, R. (1995). 3-D descriptions of buildings from an oblique view aerial image. *International Symposium on Computer Vision.* Coral Gables, FL, USA,: IEEE, 377-382. Doi: 10.1109/ISCV.1995.477031.

Lin, H., Du, P., Zhao, W., Zhang, L. & Sun, H. (Year) Published. Image registration based on corner detection and affine transformation. 3rd International Congress on Image and Signal Processing, 2010. IEEE, 2184-2188. Doi: 10.1109/CISP.2010.5647722.

Lin, Y., Jiang, M., Yao, Y., Zhang, L. & Lin, J. (2015). Use of UAV oblique imaging for the detection of individual trees in residential environments. *Urban Forestry & Urban Greening,* 14, 404-412.

Lin, Y. & West, G. (2014). Attempt of UAV oblique images and MLS point clouds for 4D modelling of roadside pole-like objects. *Proc. SPIE 9262, Lidar Remote Sensing for Environmental Monitoring XIV*, 92620Q.

Liu, J. & Liu, Y. (2014). Local Regularity-driven city-scale facade detection from aerial images. *2014 IEEE Conference on Computer Vision and Pattern Recognition.* Columbus, OH, USA, 3778-3785. Doi: 10.1109/CVPR.2014.489.

Liu, L. & Shao, Z. (2015). Calibration of digital camera integration accuracy for low-cost oblique aerial photogrammetry. *Geo-spatial Information Science,* 18, 90-96.

Liu, S., Fraser, C.S., Zhang, C., Ravanbakhsh, M. & Tong, X. (2011). Georeferencing performance of THEOS satellite imagery. *The Photogrammetric Record,* 26, 250-262. Doi: 10.1111/j.1477-9730.2011.00639.x.

Lourakis, M. & Argyros, A. (2004). The design and implementation of a generic sparse bundle adjustment software package based on the levenberg-marquardt algorithm. Heraklion, Crete, Greece: Institute of Computer Science - Foundation for Research and Technology Hellas (FORTH), 1-21.

Lourakis, M.A. & Orphanoudakis, S.C. (1998). Visual detection of obstacles assuming a locally planar ground *In:* Chin, R.. & Pong, T.C. (eds.) *Computer Vision - ACCV'98. Lecture Notes in Computer Science, vol. 1352.* Springer, Berlin, Heidelberg.

Lourakis, M.I.A., Argyros, A.A. & Orphanoudakis, S.C. (2002). Detecting planes in an uncalibrated image pair. *13th British Machine Vision Conference (BMVC 2002).* Cardiff, United Kingdom. Doi: 10.5244/C.16.57.

Lowe, D.G. (1987). Three-dimensional object recognition from single two-dimensional images. *Artificial Intelligence,* 31, 355-395.

Lowe, D.G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision,* 60, 91-110.

Lugang, Z., Chengke, W. & Jifeng, N. (2014). A camera calibration method based on two orthogonal vanishing points. *Concurrency and Computation: Practice and Experience,* 26, 1185-1199. Doi: 10.1002/cpe.3055.

Madani, A.T., Ahmad, A.Z., Christoph, L., Hammadi, Z. & Sabeal, M.A. (2016). Dubai 3D textured mesh using high quality resolution vertical/oblique aerial imagery. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* XLI-B3, 151-154.

Malis, E. & Cipolla, R. (2002). Camera self-calibration from unknown planar structures enforcing the multiview constraints between collineations. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 24, 1268-1272.

Mao, M. (2013). *A novel feature for registration of oblique aerial images under large viewpoint variations.* Master Thesis, Simon Fraser University.

Markelin, L., Honkavaara, E., Näsi, R., Nurmine, K. & Hakala, T. (2014). Geometric processing workflow for vertical and oblique hyperspectral frame images collected using UAV. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* XL-3, 205-210.

Matas, J., Galambos, C. & Kittler, J. (1998). Progressive Probabilistic Hough Transform. *British Machine Vision Conference '98 (BMVC 1998).* Southampton, UK, 256-265.

Mattison, D. (2007). Aerial photography. *In:* Hannavy, J. (ed.) *Encyclopedia of nineteenth-century photography.* USA: Routledge.

Mertes, J.R., Gulley, J.D., Benn, D.I., Thompson, S.S. & Nicholson, L.I. (2017). Using structure-from-motion to create glacier DEMs and orthoimagery from historical terrestrial and oblique aerial imagery. *Earth Surface Processes and Landforms,* 42, 2350-2364. Doi: 10.1002/esp.4188.

Microsoft. (2019). *GDI+* [Online]. Available: https://docs.microsoft.com/el-gr/windows/desktop/gdiplus/ [Accessed 10 April 2019].

Midgley, N.G. & Tonkin, T.N. (2017). Reconstruction of former glacier surface topography from archive oblique aerial images. *Geomorphology,* 282, 18-26. Doi: 10.1016/j.geomorph.2017.01.008.

Minagawa, A., Tagawa, N., Moriya, T. & Gotoh, T. (2000). Vanishing point and vanishing line estimation with line clustering. *IEICE Transctions on Information and Systems,* 83, 1574-1582.

Mirzaei, F.M. & Roumeliotis, S.I. (2011). Optimal estimation of vanishing points in a Manhattan world. *2011 IEEE International Conference on Computer Vision (ICCV).* Barcelona, Spain: IEEE, 2454-2461. Doi: ICCV.2011.6126530.

Mishra, P., Ofek, E. & Kimchi, G. (2008). Validation of vector data using oblique images. *16th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems.* Irvine, CA, USA.

Misulia, M.G. (1946). A derivation for the image displacement due to tilt. *Photogrammetric Engineering,* 461-463.

Moe, K., Toschi, I., Poli, D., Lago, F., Schreiner, C., Legat, K. & Remondino, F. (2016). Changing the production pipeline - Use of oblique aerial cameras for mapping purposes. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* XLI-B4, 631-637.

Moffitt, F.H. (1962). *Elements of Photogrammetry,* USA, International Textbook Company.

Moffitt, F.H. (1967). *Photogrammetry,* Scranton, Pennsylvania, International Textbook Company.

Moffitt, F.H. & Mikhail, E.M. (1980). *Photogrammetry,* New York, Harper & Row Inc.

Moghadam, P., Starzyk, J.A. & Wijesoma, W.S. (2012). Fast vanishing-point detection in unstructured environments. *IEEE Transactions on Image Processing,* 21, 425-430. Doi: 10.1109/TIP.2011.2162422.

Molinari, M., Medda, S. & Villani, S. (2014). Vertical measurements in oblique aerial imagery. *ISPRS International Journal of Geo-Information,* 3, 914-928.

Mondragón, I.F., Campoy, P., Martínez, C. & Olivares-Méndez, M.A. (2010). 3D pose estimation based on planar object tracking for UAVs control. *2010 IEEE International Conference on Robotics and Automation.* Anchorage, Alaska, USA: IEEE, 35-41. Doi: 10.1109/ROBOT.2010.5509287.

Moré, J.J. (1978). The Levenberg-Marquardt algorithm: implementation and theory. *In:* Watson, G.A. (ed.) *Numerical Analysis. Lecture Notes in Mathematics, vol. 630.* Springer, Berlin, Heidelberg.

Morel, J.-M. & Yu, G. (2009). ASIFT: A new framework for fully affine invariant image comparison. *SIAM Journal on Imaging Sciences,* 2, 438-469. Doi: 10.1137/080732730.

Moulon, P., Monasse, P. & Marlet, R. (2013). Global fusion of relative motions for robust, accurate and scalable structure from motion. *2013 IEEE International Conference on Computer Vision (ICCV).* Sydney, NSW, Australia, 3248-3255. Doi: 10.1109/ICCV.2013.403.

Murtiyoso, A., Remondino, F., Rupnik, E., Nex, F. & Grussenmeyer, P. (2014). Oblique aerial photography tool for building inspection and damage assessment. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* XL-1, 309-313.

Nakada, R., Takigawa, M., Ohga, T. & Fujii, N. (2016). Verification of potency of aerial digital oblique cameras for aerial photogrammetry in Japan. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* XLI-B1, 63-68.

Nevins, R.P. (2017). *Georeferencing unmanned aerial systems imagery via registration with geobrowser reference imagery.* Master Thesis, Ohio State University.

Nex, F., Gerke, M., Remondino, F., Przybilla, H.-J., Bäumker, M. & Zurhorst, A. (2015). ISPRS benchmark for multi-platform photogrammetry. *The International Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* II-3/W4, 135-142.

Nex, F., Rupnik, E. & Remondino, F. (2013). Building footprints extraction from oblique imagery. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* II-3/W3, 61-66.

Ni, K. & Dellaert, F. (2012). HyperSfM. *Second International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT).* Zurich, Switzerland: IEEE, 144-151.

Nieto, M. & Salgado, L. (2011). Simultaneous estimation of vanishing points and their converging lines using the EM algorithm. *Pattern Recognition Letters,* 32, 1691-1700. Doi: 10.1016/j.patrec.2011.07.018.

Nishida, K., Fujiki, J., Tsushiya, C., Tanaka, S. & Kurita, T. (2011). Road plane detection using differential homography estimated by pair feature matching of local regions. *In:* Hamza, M.H. & Zhang, J.J. (eds.) *Conference on Signal Processing, Pattern Recognition, and Applications.* Innsbruck, Austria, 16-18. Doi: 10.2316/P.2011.721-073.

Nocerino, E., Menna, F. & Remondino, F. (2012). Multi-temporal analysis of landscapes and urban areas. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* XXXIX-B4, 85-90. Doi: 10.5194/isprsarchives-XXXIX-B4-85-2012.

Novak, K. (1992). Rectification of Digital Imagery. *Photogrammetric Engineering & Remote Sensing,* 58 339-344.

Nyaruhuma, A.P. (2013). *Automatic verification of buildings using oblique airborne images.* Doctoral Thesis, University of Twente.

Nyaruhuma, A.P., Gerke, M. & Vosselman, G. (2012). Verification of 3D building models using mutual information in airborne oblique images. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* I-3, 275-280.

Odone, F., Fusiello, A. & Trucco, E. (2002). Layered representation of a video shot with mosaicing. *Pattern Analysis & Applications,* 5, 296–305. Doi: 10.1007/s100440200026.

Ohnishi, N. & Imiya, A. (2006). Dominant plane detection from optical flow for robot navigation. *Pattern Recognition Letters,* 27, 1009-1021. Doi: 10.1016/j.patrec.2005.11.012.

Okatani, T. & Deguchi, K. (2007). On the Wiberg algorithm for matrix factorization in the presence of missing components. *International Journal of Computer Vision,* 72, 329-337. Doi: 10.1007/s11263-006-9785-5.

Onyango, F.A., Nex, F., Peter, M.S. & Jende, P. (2017). Accurate estimation of orientation parameters of UAV images through image registration with aerial oblique imagery. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* XLII-1/W1, 599-605. Doi: 10.5194/isprs-archives-XLII-1-W1-599-2017.

OpenCV Team. (2019). *OpenCV Library* [Online]. Available: https://opencv.org/ [Accessed 10 April 2019].

Ostrowski, W. (2016). Accuracy of measurements in oblique aerial images for urban environment. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* XLII-2/W2, 79-85.

Ostrowski, W. & Bakuła, K. (2016). Towards efficiency of oblique images orientation. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* XL-3/W4, 91-96.

Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics,* 9, 62-66.

Paine, D.P. & Kiser, J.D. (2012). *Aerial photography and image interpretation,* Hoboken, New Jersey, John Wiley & Sons, Inc.

Palmer, R. (1976). A method of transcribing archaeological sites from oblique aerial photographs *Journal of Archaeological Science,* 3, 391-394.

Palmer, R. (1977). A computer method for transcribing information graphically from oblique aerial photographs to maps. *Journal of Archaeological Science,* 4, 283-290.

Panday, U.S. & Gerke, M. (2011). Fitting of parametric building models to oblique aerial images. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* XXXVIII-4/W19, 233-238. Doi: 10.5194/isprsarchives-XXXVIII-4-W19-233-2011.

Panigrahi, N., Mohan, B.K. & Athithan, G. (2011). Pre-processing algorithm for rectification of geometric distortions in satellite images. *Defence Science Journal,* 61, 174-179. Doi: 10.14429/dsj.61.421.

Passini, R., Jacobsen, K., Weaver, W. & Day, D. (2016). An study on oblique camera systems calibration and the stability of their parameters. *ASPRS 2016 Annual Conference: IGTF-2016 (Imaging and Geospatial Technology Forum).* Fort Worth, Texas, USA.

Petrie, G. (2009). Systematic oblique aerial photography using multiple digital cameras. *Photogrammetric Engineering & Remote Sensing,* 75, 102-107.

Petsa, E. & Patias, P. (1994). Formulation & assessment of straightline based algorithms for digital photogrammetry. *International Archives of Photogrammetry and Remote Sensing,* 30 (B5), 310-317.

Pix4D. (2019). *Pix4D - Professional drone mapping and photogrammetry software* [Online]. Available: https://pix4d.com/ [Accessed 10 April 2019].

Poelman, C.J. & Kanade, T. (1997). A paraperspective factorization method for shape and motion recovery. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 19, 206-218. Doi: 10.1109/34.584098.

Prandi, F., Achille, C., Brumana, R., Fassi, F. & Fregonese, L. (2008). LiDAR and Pictometry images integrated use for 3D model generation. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* XXXVII (Part B2), 763-770.

Punmia, B.C., Jain, A.K. & Jain, A.K. (2005). *Higher surveying,* New Delhi, Laxmi Publications (P) Ltd.

Radlinski, W.A. (1952). Convergent low oblique photography and its application to the twinplex. *Photogrammetric Engineering,* 18, 591-597.

Rau, J.-Y. & Chu, C.-Y. (2010). Photo-realistic 3D mapping from aerial oblique imagery. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* XXXVIII - Part 1.

Rau, J.-Y., Jhan, J.-P. & Hsu, Y.-C. (2015). Analysis of oblique aerial images for land cover and point cloud classification in an urban environment. *IEEE Transactions on Geoscience and Remote Sensing,* 53, 1304-1318. Doi: 10.1109/TGRS.2014.2337658.

Redweik, P., Garzón, V. & sa Pereira, T. (2016). Recovery of stereo aerial coverage from 1934 and 1938 into the digital era. *The Photogrammetric Record* 31, 9-28. Doi: doi.org/10.1111/phor.12137.

Remondino, F. & Gerke, M. (2015). Oblique aerial imagery - a review. *In:* Fritsch, D. (ed.) *Photogrammetric Week '15.* Belin: Wichmann/VDE Verlag.

Remondino, F., Toschi, I., Gerke, M., Nex, F., Holland, D., McGill, A., Talaya Lopez, J. & Magarinos, A. (2016). Oblique aerial imagery for NMA - Some best practices. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* XLI-B4, 639-645.

Ribeiro, E. & Hancock, E.R. (2000). Perspective pose from spectral voting. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2000).* Hilton Head Island, SC, USA: IEEE, 656 - 662. Doi: 10.1109/CVPR.2000.855882.

Rich, J.L. (1947). Reconnaissance mapping from oblique aerial photographs without ground control. *Photogrammetric Engineering,* 600-609.

Robertson, D.P. & Cipolla, R. (2009). Structure from motion. *In:* Varga, M. (ed.) *Practical image processing and computer vision.* New York: John Wiley and Sons Ltd.

Rocchini, D., Metz, M., Frigeri, A., Delucchi, L., Marcantonio, M. & Neteler, M. (2012). Robust rectification of aerial photographs in an open source environment. *Computers & Geosciences,* 39, 145-151. Doi: 10.1016/j.cageo.2011.06.002.

Rocchini, D. & Rita, A.D. (2005). Relief effects on aerial photos geometric correction. *Applied Geography,* 25, 159-168. Doi: 10.1016/j.apgeog.2005.03.002.

Rother, C. (2002). A new approach for vanishing point detection in architectural environments. *Image and Vision Computing,* 20, 647-655. Doi: 10.1016/S0262-8856(02)00054-9.

Rupnik, E., Nex, F. & Remondino, F. (2013). Automatic orientation of large blocks of oblique images. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences,* XL-1/W1, 299-304.

Rupnik, E., Nex, F. & Remondino, F. (2014). Oblique multi-camera systems - orientation and dense matching issues. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* XL-3/W1, 107-114. Doi: 10.5194/isprsarchives-XL-3-W1-107-2014.

Rupnik, E., Nex, F., Toschi, I. & Remondino, F. (2015). Aerial multi-camera systems: Accuracy and block triangulation issues. *ISPRS Journal of Photogrammetry and Remote Sensing,* 101, 233-246.

Sadeghian, S., Mohammad Javad Valadan Zoej, Delavar, M.R. & Abootalebi, A. (2001). Precision rectification of high resolution satellite imagery without ephemeris data. *International Journal of Applied Earth Observation and Geoinformation,* 3, 366-371. Doi: 10.1016/S0303-2434(01)85044-3.

Saeedi, P. & Mao, M. (2014). Two-edge-corner image features for registration of geospatial images with large view variations. *International Journal of Geosciences,* 5, 1324-1344.

Schack, L., Soergel, U. & Heipke, C. (2016). Graph matching for the registration of persistent scatterers to optical oblique imagery. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* III-7, 195-202.

Schindler, G. & Dellaert, F. (2004). Atlanta world: an expectation maximization framework for simultaneous low-level edge grouping and camera calibration in complex man-made environments. *2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2004).* Washington, D.C.: IEEE. Doi: CVPR.2004.1315033.

Schindler, G., Krishnamurthy, P., Lublinerman, R., Liu, Y. & Dellaert, F. (2008). Detecting and matching repeated patterns for automatic geo-tagging in urban environments. *2008 IEEE Conference on Computer Vision and Pattern Recognition.* Anchorage, AK, USA: IEEE, 1-7. Doi: CVPR.2008.4587461.

Schindler, K. (2003). Generalized use of homographies for piecewise planar reconstruction. *In:* Bigun, J. & Gustavsson, T. (eds.) *Image Analysis. SCIA 2003. Lecture Notes in Computer Science, vol. 2749.* Springer, Berlin, Heidelberg.

Scollar, I. (1975). Transformation of extreme oblique aerial photographs to maps or plans by conventional means or by computer. *Aerial Reconnaissance for Archaeology,* 52-59.

Scollar, I., Huang, T.S., Weidner, B. & Tang, G. (1977). An installation for interactive transfer of information from oblique aerial photos to maps. *Digitale Bildverarbeitung Digital Image Processing.* Berlin, Heidelberg: Springer.

Shah, R., Deshpande, A. & Narayanan, P.J. (2014). Multistage SFM: Revisiting incremental structure from motion. *2nd International Conference on 3D Vision (3DV).* Tokyo, Japan: IEEE, 417-424. Doi: 10.1109/3DV.2014.95.

Shen, X., Li, Q., Wu, G. & Zhu, J. (2017). Bias compensation for rational polynomial coefficients of high-resolution satellite imagery by local polynomial modeling. *Remote Sensing,* 9, 200:1-12. Doi: 10.3390/rs9030200.

Shi, J. & Tomasi, C. (1994). Good Features to Track. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR94).* Seattle, 593-600.

Shi, W. & Samarabandu, J. (2006). Corridor line detection for vision based indoor robot navigation. *2006 Canadian Conference on Electrical and Computer Engineering.* Ottawa, Ont., Canada: IEEE, 1988-1991. Doi: 10.1109/CCECE.2006.277812.

Shufelt, J.A. (1999). Performance evaluation and analysis of monocular building extraction from aerial imagery. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 21, 311 - 326. Doi: 10.1109/34.761262.

Smith, M.J. (1989). A photogrammetric system for archaeological mapping using oblique non-metric photography. *The Photogrammetric Record,* 13, 95-105.

Smith, M.J., Kokkas, N., Hamruni, A.M., Critchley, D. & Jamieson, A. (2008). Investigation into the orientation of oblique and vertical digital images. *EuroCOW 2008.* Castelldefels, Spain.

Snavely, N., Seitz, S.M. & Szeliski, R. (2006). Photo tourism: exploring photo collections in 3D. *ACM Transactions on Graphics,* 25, 835-846.

Soler, M.E., Kornus, W., Magariños, A. & Pla, M. (2016). Analyzing RCD30 oblique performance in a production environment. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* XLI-B3, 99-105.

Sørensen, E.V., Pedersen, A.K., García-Sellés, D. & Strunck, M.N. (2015). Point clouds from oblique stereo-imagery: Two outcrop case studies across scales and accessibility. *European Journal of Remote Sensing,* 48, 593-614. Doi: 10.5721/EuJRS20154833.

Srivastava, S. (2015). SIFT vs SURF: quantifying the variation in transformations. *arXiv:1504.06740 [cs.CV]*.

Stilla, U., Kolecki, J. & Hoegner, L. (2009). Texture mapping of 3D building models with oblique direct geo-referenced airborne IR image sequences. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* XXXVIII-1-4-7/W5.

Sturm, P.F. & Maybank, S.J. (1999). On plane-based camera calibration: A general algorithm, singularities, applications. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition.* Fort Collins, Colorado, USA: IEEE. Doi: 10.1109/CVPR.1999.786974.

Sturm, P.F. & Triggs, B. (1996). A factorization based algorithm for multi-image projective structure and motion. *In:* Buxton, B. & Cipolla, R. (eds.) *Computer Vision — ECCV '96. ECCV 1996. Lecture Notes in Computer Science, vol. 1065.*: Springer, Berlin, Heidelberg.

Sun, X., Shen, S., Cui, H., Hu, L. & Hu, Z. (2017). Geographic, geometrical and semantic reconstruction of urban scene from high resolution oblique aerial images. *IEEE/CAA Journal of Automatica Sinica,* PP, 1-13. Doi: 10.1109/JAS.2017.7510673.

Sun, Y., Sun, H., Yan, L., Fan, S. & Chen, R. (2016). RBA: Reduced Bundle Adjustment for oblique aerial photogrammetry. *ISPRS Journal of Photogrammetry and Remote Sensing,* 121, 128–142.

Szeliski, R. & Torr, P.H.S. (1998). Geometrically constrained structure from motion: Points on planes. *In:* Koch, R. & Van Gool, L. (eds.) *SMILE 1998: 3D Structure from Multiple Images of Large-Scale Environments. Lecture Notes in Computer Science, vol. 1506.* Springer, Berlin, Heidelberg.

Talley, B.B. (1938). Multiple lens aerial cameras. *Engineering applications of aerial and terrestrial photogrammetry.* New York: Pitman Publishing Corporation.

Tardif, J.-P. (2009). Non-iterative approach for fast and accurate vanishing point detection. *2009 IEEE 12th International Conference on Computer Vision.* Kyoto, Japan: IEEE, 1250 - 1257. Doi: ICCV.2009.5459328.

Teller, S., Antone, M., Bodnar, Z., Bosse, M., Coorg, S., Jethwa, M. & Master, N. (2003). Calibrated, registered images of an extended urban area. *International Journal of Computer Vision,* 53, 93-107.

Thompson, M.M. (1950). The Twinplex: A new stereoplotting instrument. *Geological Survey Circular.* Washington, D.C.: United States Department of the Interior, Geological Survey.

Thompson, M.M. (1952). Development of Photogrammetry in the U.S. Geological Survey. *Geological Survey Circular.* Washington, D.C.

Toldo, R. & Fusiello, A. (2008). Robust multiple structures estimation with J-linkage. *In:* Forsyth, D., Torr, P. & Zisserman, A. (eds.) *Computer Vision – ECCV 2008. ECCV 2008. Lecture Notes in Computer Science, vol. 5302.* . Springer, Berlin, Heidelberg.

Tomasi, C. & Kanade, T. (1992). Shape and motion from image streams under orthography: a factorization method. *International Journal of Computer Vision,* 9, 137-154. Doi: 10.1007/BF00129684.

Tommaselli, A.M.G., Galo, M., de Moraes, M.V.A., Marcato, J.J., Caldeira, C.R.T. & Lopes, R.F. (2013). Generating virtual images from oblique frames. *Remote Sensing,* 5, 1875-1893. Doi: 10.3390/rs5041875.

Toschi, I., Ramos, M.M., Nocerino, E., Menna, F., Remondino, F., Moe, K., Poli, D., Legat, K. & Fassi, F. (2017). Oblique photogrammetry supporting 3D urban reconstruction of complex scenarios. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* XLII-1/W1, 519-526. Doi: 10.5194/isprs-archives-XLII-1-W1-519-2017.

Troney, L.G. (1952). *Handbook of aerial mapping and photogrammetry,* USA, Cambridge University Press.

Tu, J., Sui, H., Feng, W., Sun, K., Xu, C. & Han, Q. (2017). Detecting building façade damage from oblique aerial images using local symmetry feature and the Gini Index. *Remote Sensing Letters,* 8, 676-685. Doi: 10.1080/2150704X.2017.1312027.

Verhoeven, G., Doneus, M., Briese, C. & Vermeulen, F. (2012). Mapping by matching: a computer vision-based approach to fast and accurate georeferencing of archaeological aerial

photographs. *Journal of Archaeological Science,* 39, 2060-2070. Doi: 10.1016/j.jas.2012.02.022.

Vetrivel, A., Duarte, D., Nex, F., Gerke, M., Kerle, N. & Vosselman, G. (2016). Potential of multi-temporal oblique airborne imagery for structural damage assessment. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* III-3, 355-362, 355-362.

Vetrivel, A., Gerke, M., Kerle, N., Nex, F. & Vosselman, G. (2017). Disaster damage detection through synergistic use of deep learning and 3D point cloud features derived from very high resolution oblique aerial images, and multiple-kernel-learning. *ISPRS Journal of Photogrammetry and Remote Sensing*. Doi: 10.1016/j.isprsjprs.2017.03.001.

Vetrivel, A., Gerke, M., Kerle, N. & Vosselman, G. (2015). Identification of damage in buildings based on gaps in 3D point clouds from very high resolution oblique airborne images. *ISPRS Journal of Photogrammetry and Remote Sensing,* 105, 61-78.

Vincent, E. & Laganiere, R. (2001). Detecting planar homographies in an image pair. *2nd International Symposium on Image and Signal Processing and Analysis. In conjunction with 23rd International Conference on Information Technology Interfaces.* Pula, Croatia, Croatia: IEEE, 182-187. Doi: 10.1109/ISPA.2001.938625.

von Gruber, O. (1932). *Photogrammetry. Collected Lectures and Essays* London, Chapman & Hall.

Wackrow, R. & Chandler, J.H. (2011). Minimising systematic error surfaces in digital elevation models using oblique convergent imagery. *The Photogrammetric Record,* 26, 16-32. Doi: 10.1111/j.1477-9730.2011.00623.x.

Wang, L. & Neumann, U. (2009). A robust approach for automatic registration of aerial images with untextured aerial LiDAR data. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2009).* Miami, Florida, USA: IEEE, 2623-2630.

Wang, M., Bai, H. & Hu, F. (2008). Automatic texture acquisition for 3D model using oblique aerial images. *First International Conference on Intelligent Networks and Intelligent Systems.* Wuhan, China: IEEE, 495-498. Doi: 10.1109/ICINIS.2008.122.

Weith-Glushko, S. & Salvaggio, C. (2005). Automatic tie-point generation for oblique aerial imagery: An algorithm. *ASPRS 2005 Annual Conference.* Baltimore, Maryland.

Wiedemann, A. & Moré, J. (2012). Orientation strategies for aerial oblique images. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* XXXIX-B1, 185-189.

Wildenauer, H. & Vincze, M. (2007). Vanishing point detection in complex man-made worlds. *14th International Conference on Image Analysis and Processing (ICIAP 2007).* Modena, Italy: IEEE, 615-622. Doi: 10.1109/ICIAP.2007.4362845.

Williams, J.C.C. (1969). *Simply Photogrammetry,* Academic Press Inc.

Williamson, J.R. & Brill, M.H. (1990). *Dimensional analysis through perspective: A reference manual,* USA, Kendal/Hunt Publishing Co.

Willneff, J. & Poon, J. (2006). Georeferencing from orthorectified and non-orthorectified high-resolution satellite imagery. *The 13th Australasian Remote Sensing and Photogrammetry Conference.* Canberra, Australia, 20-24.

Wilson, K. & Snavely, N. (2014). Robust Global Translation with 1DSfM. *In:* Fleet, D., Pajdla, T., Schiele, B. & Tuytelaars, T. (eds.) *Computer Vision - ECCV 2014. ECCV 2014. Lecture Notes in Computer Science, vol. 8691.* Springer, Cham.

Wolf, P.R. (1983). *Elements of photogrammetry with air photo interpretation and remote sensing,* New York, McGraw-Hill Inc.

Wolf, P.R. (1984). *Elements of Photogrammetry,* USA, McGraw-Hill Inc., US.

Wu, C. (2013). Towards linear-time incremental structure from motion. *2013 International Conference on 3D Vision.* Seattle, Washington, USA: IEEE, 127-134. Doi: 10.1109/3DV.2013.25.

Wu, J., Cui, Z., Sheng, V.S., Zhao, P., Su, D. & Gong, S. (2013). A comparative study of SIFT and its variants. *Measurement Science Review,* 13, 122-131. Doi: 10.2478/msr-2013-0021.

Xiao, J. (2012). Automatic building outlining from multi-view oblique images. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* I-3, 323-328.

Xiao, J. (2013). *Automatic building detection using oblique imagery.* Doctoral Thesis, University of Twente.

Xiao, J., Gerke, M. & Vosselman, G. (2010). Automatic detection of buildings with rectangular flat roofs from multi-view oblique imagery. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* XXXVIII (Part 3A), 251-256.

Xiao, J., Gerke, M. & Vosselman, G. (2012). Building extraction from oblique airborne imagery based on robust façade detection. *ISPRS Journal of Photogrammetry and Remote Sensing,* 68, 56-68.

Xiao, X., Guo, B., Shi, Y., Gong, W., Li, J. & Zhang, C. (2013). Robust and rapid matching of oblique UAV images of urban area. *Proc. SPIE 8919, MIPPR 2013: Pattern Recognition and Computer Vision*, 89190Y.

Xie, L., Hu, H., Wang, J., Zhu, Q. & Chen, M. (2016). An asymmetric re-weighting method for the precision combined bundle adjustment of aerial oblique images. *ISPRS Journal of Photogrammetry and Remote Sensing,* 117, 92-107.

Xie, W.-H. & Xue, Y.-C. (2016). Multi-image vanishing point based camera calibration. *2016 International Conference on Applied Mechanics, Mechanical and Materials Engineering (AMMME 2016).* Xiamen, China. Doi: 10.12783/dtmse/ammme2016/6883.

Xie, W., Zhang, Z. & Zhang, J. (2004). Multi-image based camera calibration without control points. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences,* XXXV, Part B5, 36-41.

Xiong, X., Zhang, Y., Zhu, J. & Zheng, M. (2014). Camera pose determination and 3-D measurement from monocular oblique images with horizontal right angle constraints. *IEEE Geoscience and Remote Sensing Letters,* 11, 1976-1980.

Yalcin, G. & Selcuk, O. (2015). 3D city modelling with oblique photogrammetry method. *Procedia Technology,* 19, 424-431. Doi: 10.1016/j.protcy.2015.02.060.

Yao, G.-B., Deng, K.-Z., Zhang, L. & Aic, H.-B. (2014). Robust perspective invariant quasidense matching across large oblique images. *Journal of Electronic Imaging,* 23, 043005. Doi: 10.1117/1.JEI.23.4.043005.

Zhai, M., Workman, S. & Jacobs, N. (2016). Detecting vanishing points using global image context in a non-Manhattan world. *IEEE Conference on Computer Vision and Pattern Recognition.* San Francisco, California: IEEE, 5657-5665.

Zhang, J. (2013). *Dense point cloud extraction from oblique imagery.* Master Thesis, Rochester Institute of Technology.

Zhang, J., Yao, H., Jiang, W. & Shen, X. (2015a). Hierarchical repetition extraction for building façade reconstruction from oblique aerial images. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* XL-4/W5, 183-187.

Zhang, L. & Koch, R. (2013). Vanishing points estimation and line classication in a Manhattan world. *In:* Lee, K.M., Matsushita, Y., Rehg, J.M. & Hu, Z. (eds.) *Computer Vision – ACCV 2012. ACCV 2012. Lecture Notes in Computer Science, vol. 7725.* Springer, Berlin, Heidelberg.

Zhang, M. (2014). *Towards 3D matching of point clouds derived from oblique and nadir airborne imagery.* Master Thesis, Rochester Institute of Technology.

Zhang, X., Gao, X., Lu, W., He, L. & Liu, Q., 311, 260–269. doi: (2018). Dominant vanishing point detection in the wild with application in composition analysis. *Neurocomputing,* 311, 260-269. Doi: 10.1016/j.neucom.2018.05.071.

Zhang, Y., Zhang, Z. & Zhang, J. (2005). The feasibility study on the aerial triangulation over urban area constrained by vertical lines. *Proc. SPIE 6043, MIPPR 2005: SAR and Multispectral Image Processing*, 60431S.

Zhang, Z. (2000). A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 22, 1330–1334. Doi: 10.1109/34.888718.

Zhang, Z., Dai, C.G., Ji, S. & Zhao, M.Y. (2015b). Adaptive hierarchical dense matching of multi-view airborne oblique imagery. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences,* XL-3/W2, 289-294.

Zhao, H., Zhang, B., Wu, C., Zuo, Z., Chen, Z. & Bi, J. (2014). Direct georeferencing of oblique and vertical imagery in different coordinate systems. *ISPRS Journal of Photogrammetry and Remote Sensing*, 95, 122-133. Doi: 10.1016/j.isprsjprs.2014.06.001.

Zhao, Y.-G., Wang, X., Feng, L.-B., Chen, G., Wu, T.-P. & Tang, C.-K. (2013). Calculating vanishing points in dual space. *In:* Yang, J., Fang, F. & Sun, C. (eds.) *Intelligent Science and Intelligent Data Engineering. IScIDE 2012. Lecture Notes in Computer Science, vol. 7751*. Springer, Berlin, Heidelberg.

Zheng, Q. & Chellappa, R. (1994). Automatic registration of oblique aerial images. *IEEE International Conference on Image Processing* Austin, TX, USA: IEEE, 218-222. Doi: 10.1109/ICIP.1994.413307.

Zhou, Q. & Liu, J. (2015). Automatic orthorectification and mosaicking of oblique images from a zoom lens aerial camera. *Optical Engineering*, 54, 013104:1-14.

Zhou, Z., Farhat, F. & Wang, J.Z. (2017). Detecting dominant vanishing points in natural scenes with application to composition-sensitive image retrieval. *IEEE Transactions on Multimedia*, 19, 2651-2665. Doi: 10.1109/TMM.2017.2703954.

Zhou, Z., Jin, H. & Ma, Y. (2012). Robust plane-based structure from motion. *2012 IEEE Conference on Computer Vision and Pattern Recognition.* IEEE, 1482-1489. Doi: 10.1109/CVPR.2012.6247837.

Zhuo, X., Koch, T., Kurz, F., Fraundorfer, F. & Reinartz, P. (2017). Automatic UAV image geo-registration by matching UAV images to georeferenced image data. *Remote Sensing*, 9, 376:1-25. Doi: 10.3390/rs9040376.

# List of Publications of Styliani Verykokou

## International Peer-Reviewed Journal Articles

**Verykokou, S.** & Ioannidis, C. (forthcoming 2020). Exterior orientation estimation of oblique aerial images using SfM-based robust bundle adjustment. *International Journal of Remote Sensing*.

**Verykokou, S.** & Ioannidis, C. (2018). Oblique aerial images: a review focusing on georeferencing procedures. *International Journal of Remote Sensing*, 39, 3452-3496. Doi: 10.1080/01431161.2018.1444294.

**Verykokou, S.**, Ioannidis, C., Athanasiou, G., Doulamis, N. & Amditis, A. (2018). 3D reconstruction of disaster scenes for urban search and rescue. *Multimedia Tools and Applications*, 77, 9691–9717. Doi: 10.1007/s11042-017-5450-y.

**Verykokou, S.** & Ioannidis, C. (2016). Automatic rough georeferencing of multiview oblique and vertical aerial image datasets of urban scenes. *The Photogrammetric Record*, 31, 281-303. Doi: 10.1111/phor.12156.

Soile, S., **Verykokou, S.** & Ioannidis, C. (2015). Stereo-orthoimage as a GIS tool for reliable restitution of Cultural Monuments. *International Journal of Heritage in the Digital Era*, 4, 275-294. Doi: 10.1260/2047-4970.4.3-4.275.

## International Peer-Reviewed Conference Papers

Ioannidis, C., Potsiou, C., Soile, S., **Verykokou, S.** & Boutsi, A-M. (2019). Geospatial big data management for a holistic documentation of complex sites – the case of Meteora, Greece. *FIG Commission 3 Annual Meeting and Workshop 2019: Advances in Geodata Analytics for Smart Cities and Regions.* Cluj-Napoca, Romania.

**Verykokou, S.** & Ioannidis, C. (2019). A global photogrammetry-based structure from motion framework: application in oblique aerial images. *FIG Working Week 2019: Geospatial information for a smarter life and environmental resilience.* Hanoi, Vietnam.

**Verykokou, S.** & Ioannidis, C. (2018). A photogrammetry-based structure from motion algorithm using robust iterative bundle adjustment techniques. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-4/W6, 73-80. Doi: 10.5194/isprs-annals-IV-4-W6-73-2018.

Potsiou, C., Ioannidis, C., Soile, S. & **Verykokou, S.** (2016). A technical and policy tool for urban upgrading and affordable housing planning. *FIG Commission 3 Annual Workshop and Annual Meeting: From Volume to Quality - Bridging the Gap for Spatial Data Infrastructure.* Iasi, Romania.

**Verykokou, S.**, Doulamis, A., Athanasiou, G., Ioannidis, C. & Amditis, A. (2016). Multi-scale 3D modelling of damaged cultural sites: use cases and image-based workflows. *Digital Heritage. Progress in Cultural Heritage: Documentation, Preservation, and Protection. EuroMed 2016. LNCS 10058*, 50-62. Doi: 10.1007/978-3-319-48496-9_5.

Ioannidis, C., Soile, S. & **Verykokou, S.** (2016). Low cost technique for accurate geometric documentation of complex monuments by non-experts. *Digital Heritage. Progress in Cultural Heritage: Documentation, Preservation and Protection. EuroMed 2016. LNCS 10058*, 63-75. Doi: 10.1007/978-3-319-48496-9_6.

Ioannides, M., Ioannidis, C., Archimadrite Enkleistriotis, N., Castrillo, D., Chatzigrigoriou, P., Papageorgiou, E., Leventis, G., Nikolakopoulou, V., Athanasiou, V., Bourexis, F., Soile, S., **Verykokou, S.**, de Castrillo, M.C. & Sovis, C. (2016). Towards monuments' holistic digital documentation: the Saint Neophytos Enkleistriotis case study. *Digital Heritage. Progress in Cultural Heritage: Documentation, Preservation and Protection. EuroMed 2016. LNCS 10058*, 442-473. Doi: 10.1007/978-3-319-48496-9_36.

**Verykokou, S.**, Doulamis, A., Athanasiou, G., Ioannidis, C. & Amditis, A. (2016). UAV-based 3D modelling of disaster scenes for urban search and rescue. *2016 IEEE International Conference on Imaging Systems and Techniques (IST).* Chania, Crete Island, Greece: IEEE, 106-111. Doi: 10.1109/IST.2016.7738206.

**Verykokou, S.** & Ioannidis, C. (2016). Exterior orientation estimation of oblique aerial imagery using vanishing points. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLI-B3, 123-130. Doi: 10.5194/isprs-archives-XLI-B3-123-2016.

Ioannidis, C., Potsiou, C., Soile, S., **Verykokou, S.**, Mourafetis, G. & Doulamis, N. (2016). Technical aspects for the creation of a multi-dimensional Land Information System. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLI-B2, 115-122. Doi: 10.5194/isprs-archives-XLI-B2-115-2016.

Ioannidis, C., **Verykokou, S.**, Soile, S. & Potsiou, C. (2015). 5D multi-purpose Land Information System. *3rd Eurographics Workshop on Urban Data Modelling and Visualisation (UDMV 2015).* Delft, the Netherlands.

**Verykokou, S.** & Ioannidis, C. (2015). Metric exploitation of a single low oblique aerial image. *FIG Working Week 2015.* Sofia, Bulgaria.

Ioannidis, C., Soile, S., Stamos, A., Vassilaki, D., Maltezos, E. & **Verykokou, S.** (2015). Exploitation of satellite optical and SAR data for public work studies. *Third International Conference on Remote Sensing and Geoinformation of Environment.* Paphos, Cyprus: SPIE. 95350I. Doi: 10.1117/12.2193483.

**Verykokou, S.**, Ioannidis, C. & Kontogianni, G. (2014). 3D visualization via augmented reality: the case of the Middle Stoa in the Ancient Agora of Athens. *Digital Heritage. Progress in Cultural Heritage: Documentation, Preservation, and Protection. EuroMed 2014. LNCS 8740*, 279-289. Doi: 10.1007/978-3-319-13695-0_27.

**Verykokou, S.** & Ioannidis, C. (2014). 3D Visualization through planar pattern based augmented reality. *FIG Congress 2014: Engaging the Challenges - Enhancing the Relevance.* Kuala Lumpur, Malaysia.