



National Technical University of Athens
School of Rural and Surveying Engineering

Image-based bathymetry mapping for shallow waters

Panagiotis AGRAFIOTIS

Ph.D. Dissertation

Supervisor : Prof. ANDREAS GEORGOPOULOS

Athens, July 2020



Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Αγρονόμων και Τοπογράφων Μηχανικών

Ανάπτυξη τεχνικών
βαθυμετρικής αποτύπωσης
σε ρηχά νερά με χρήση εικόνων

Παναγιώτης ΑΓΡΑΦΙΩΤΗΣ

Διδακτορική Διατριβή

Επιβλέπων : Καθ. ΑΝΔΡΕΑΣ ΓΕΩΡΓΟΠΟΥΛΟΣ

Αθήνα, Ιούλιος 2020

For Alexandra

Committee

Advisory Committee

Andreas Georgopoulos
Professor, National Technical University of Athens, Supervisor

Konstantinos Karantzas
Associate Professor, National Technical University of Athens

Dimitrios Skarlatos
Associate Professor, Cyprus University of Technology

Examination Committee

Andreas Georgopoulos
Professor, National Technical University of Athens, Supervisor

Konstantinos Karantzas
Associate Professor, National Technical University of Athens

Dimitrios Skarlatos
Associate Professor, Cyprus University of Technology

Charalabos Ioannidis
Professor, National Technical University of Athens

Anastasios Doulamis
Associate Professor, National Technical University of Athens

Costas Armenakis
Professor, York University

Giorgos Tolias
Assistant Professor, Czech Technical University in Prague



**NATIONAL TECHNICAL
UNIVERSITY OF ATHENS**
RESEARCH COMMITTEE

This Ph.D. Dissertation was supported by the Special Account for Research Grants of National Technical University of Athens for the period from February 2018 to July 2020.

Abstract

Accurate bathymetric mapping is a key element for offshore activities, hydrological studies such as coastal engineering applications, sedimentary processes, hydrographic surveying as well as archaeological mapping and biological research. Through Structure from Motion (SfM) and Multi View Stereo (MVS) techniques, images can provide a low-cost alternative comparing to expensive LiDAR and acoustic surveys, offering as well, important visual information. Despite their relative low cost, their major drawback is that optical properties and illumination conditions of water severely affect overwater and underwater imagery and data acquisition process. However, when it comes to shallow waters, refraction seems to be the main factor affecting the geometry and the radiometry of the primary data and consequently of the results of overwater and underwater image-based 3D reconstruction methods. To that direction, the research carried out is concerned with the study and development of new methods for improving the performance and the accuracy of image-based mapping and 3D reconstructing the bottom in shallow waters for small- and large-scale surveys.

The first method developed, addresses the systematic refraction errors on point clouds derived from SfM-MVS procedures in a generalized and accurate way. The developed method, based on Support Vector Machines, can accurately predict shallow bathymetric information from low altitude aerial image datasets over a calm water surface, supporting several coastal engineering applications in non-turbid waters and textured bottoms.

The second method developed, is an image correction method which is built upon the state-of-the-art and firstly exploits the machine learning procedures that recover depth on the derived image-based dense point clouds (also a novel contribution of the thesis) and then corrects the refraction effect on the original imaging dataset. This way the operational SfM-MVS processing pipelines are ultimately executed on a refraction-free set of aerial imagery datasets resulting into highly accurate bathymetric maps and image-based products. This method also achieves a reduction on the noise of the sparse point clouds, which resulted from the SfM process and improves the accuracy and the quality of the produced orthoimages and textures.

In the third method developed, an FCN classifier based on SegNet architecture is exploited in order to detect areas with caustics on the underwater imagery with very high accuracy, reliability and repeatability over different types of caustics, different types of seabed and luminosity conditions. The classifier is being trained using the first real world benchmark dataset on underwater caustics which is also a deliverable of the thesis. Having detected the rippling caustics on the initial imagery, a color transferring approach is performed, images are stereo-rectified and their respective disparity maps are generated. In the final step, the pixels classified as "caustics" are replaced by the corresponding pixels on the matched images that are classified as "non-caustics" using the disparity maps and the corrected stereo-rectified images are projected back onto the initial image model in order to facilitate further SfM and MVS processing and texturing. Results suggest an increment of 25% of

the matched key-points between the images and more consistent 3D reconstruction, without any missing information. For all the developed methods, experimental results and validation over synthetic and real-world data are demonstrating their high potential, both in terms of bathymetric accuracy as well as texture and orthophoto quality. Also, frameworks were designed in a way to achieve high generalization over different platforms, cameras, flight heights, flight patterns and overlapping.

Acknowledgments

I would like to express my gratitude to the people who have supported me during the elaboration of this PhD thesis. First and foremost to my three advisors; Andreas Georgopoulos, for his outstanding ethos and for believing in me and encouraging me since my early steps on research on 2012 with his unshakable optimism, inspiration and friendship. Andreas except of being an important pillar of this PhD and my research career so far, shared with me his principles and ethics in research, shaping my research and personal identity; Konstantinos Karantzalos for the valuable discussions on the problems and the solutions found during this dissertation, for introducing me to machine and deep learning for remote sensing data, for his consistent willingness to help and for trusting on my ideas; Dimitrios Skarlatos for introducing me to the deepest secrets of the underwater environment during our collaboration in the context of the H2020 iMARECULTURE RIA project, various shipwrecks' excavations and beyond. Dimitrios gave me innumerable research freedom and trust in the context of those projects, help, comments and suggestions and was actively involved in most of the parts of this thesis. The exemplary cooperation between all the members of the advisory committee served as a constant motivation for improvement, setting a remarkable foundation for my scientific career.

I also thank the rest of the members of the Photogrammetry Lab. of NTUA and the members of the Photogrammetric Vision Lab. of CUT for their constant willingness to help whenever it was necessary. I would also like to extend my gratitude to Dr. Ioannis Papadakis for the valuable discussions we had regarding the refraction effect. I would like also to thank all the other members of the examination committee; Prof. Charalabos Ioannidis, Assoc. Prof. Anastasios Doulamis, Prof. Costas Armenakis and Assist. Prof. Giorgos Tolia, for their willingness to participate in the examination of this doctoral dissertation. Assoc. Prof. Anastasios Doulamis is also acknowledged for giving me the chance to work in various local and EU projects, extending my interests and knowledge. Assist. Prof. Giorgos Tolia is also acknowledged for his time and his valuable guidelines and comments regarding the deep learning aspects of this thesis. I would like also to acknowledge the Dep. of Land and Surveys of Cyprus for providing the LiDAR data, the Cyprus Dep. of Antiquities for permitting the flight over the Amathounta and commissioning the flight over Agia Napa, the Rural and Surveying Engineer Dimitrios Kontos for conducting the flights of Cyclades datasets, and Honor Frost Foundation for partially funding the research in Cyprus' datasets.

My beloved parents, Giorgos and Stavroula deserve special gratitude for their support on my decisions. I wish also to extend my gratitude to all my friends for supporting me.

The last acknowledgement is reserved for the most important person in my life, Alexandra, who with her endless love, comprehension, encouragement and patience, has been my principal support during these years, supporting all my crazy ideas and goals and giving me the strength to achieve them. Thank you for always standing beside me.

Εκτεταμένη Περίληψη

Η ακριβής βαθυμετρία και χαρτογράφηση του πυθμένα θαλασσών, λιμνών και ποταμών είναι σημαντική για πλήθος δραστηριοτήτων όπως υδρολογικές μελέτες, μελέτες ακτομηχανικής, χαρτογράφηση βυθισμένων αρχαιοτήτων, εντοπισμός και χαρτογράφηση θαλάσσιας χλωρίδας και πανίδας, διαγράμματα ναυσιπλοΐας κ.α. Μέσω τεχνικών φωτογραμμετρίας, όρασης υπολογιστών και τηλεπισκόπησης, οι εικόνες μπορούν να παρέχουν μία χαμηλού κόστους εναλλακτική σε σχέση με τα συστήματα Sound Navigation And Ranging (SONAR) και βαθυμετρικών συστημάτων Light Detection And Ranging (LiDAR), παρέχοντας παράλληλα και την άκρως σημαντική για πολλές εφαρμογές, εικονιστική πληροφορία. Ωστόσο, παρά το χαμηλό τους κόστος και τις δυνατότητες να παρέχουν μεγάλη ακρίβεια μέτρησης, το μεγάλο τους μειονέκτημα είναι ότι επηρεάζονται άμεσα από τις οπτικές ιδιότητες του νερού και τις συνθήκες φωτισμού. Όσον αφορά στα ρηγά νερά μέχρι 10-15 μέτρα βάθος, η διάθλαση φαίνεται να είναι ο κύριος παράγοντας που επηρεάζει αρνητικά τη γεωμετρία και τη ραδιομετρία των εικόνων, είτε αυτές λαμβάνονται εκτός του νερού, είτε εντός αυτού, και συνεπώς των 3D ανακατασκευών που παράγονται από αυτές. Ωστόσο, αν αντιμετωπιστούν αυτά τα προβλήματα επαρκώς, οι τεχνικές αυτές είναι πιο αποτελεσματικές από τα εναέρια συστήματα βαθυμετρικού LiDAR τα οποία πάσχουν από προβλήματα πολλαπλών διαδρομών στα ρηγά νερά και από τα συστήματα SONAR, φερόμενα από σκάφη, τα οποία επηρεάζονται από την οπισθοσκέδαση του ήχου και περιορίζονται από τη δυσκολία προσέγγισης των σκαφών στις ρηχότερες περιοχές.

Για την αντιμετώπιση των γεωμετρικών προβλημάτων που προκαλούνται από τη διάθλαση στις εικόνες και στα προϊόντα επεξεργασίας τους (3D νέφη, 3D επιφάνειες κλπ.), στη βιβλιογραφία έχουν προταθεί διάφορες μέθοδοι, τόσο για υποβρύχιες εικόνες όσο και για εικόνες που λαμβάνονται εκτός του νερού. Όταν η κάμερα και το απεικονιζόμενο αντικείμενο βρίσκονται εντός του νερού, το πρόβλημα της διάθλασης αντιμετωπίζεται μέσω της βαθμονόμησης, καθώς η απόσταση που διανύουν οι ακτίνες φωτός εντός της κάμερας και της θήκης θεωρούνται αμελητέες. Ωστόσο, όταν η κάμερα βρίσκεται εκτός του νερού και το αντικείμενο μέσα σε αυτό, όπως στις περιπτώσεις αποτυπώσεων μεγαλύτερων εκτάσεων με επανδρωμένα ή μη εναέρια μέσα, το πρόβλημα γίνεται πιο πολύπλοκο. Σε αυτές τις περιπτώσεις μόνο η βαθμονόμηση δεν είναι αποτελεσματική και πιο εξειδικευμένες τεχνικές απαιτούνται. Μέχρι τώρα, οι τεχνικές αυτές αναπτύσσονται για τη διόρθωση των γεωμετρικών προβλημάτων που δημιουργούνται από τη διάθλαση υπό συγκεκριμένες συνθήκες, σε συγκεκριμένα ύψη πτήσης, με συγκεκριμένα συστήματα καμερών και σε περιορισμένα σύνολα δεδομένων (datasets) που απεικονίζουν μόνο μία περιοχή, καθιστώντας την ενσωμάτωσή τους στη σύγχρονη διαδικασία παραγωγής 3D από εικόνες μέσω τεχνικών Structure from Motion-Multi-View Stereo (SfM-MVS) μη εφικτή. Επιπλέον, μέχρι τώρα, η αποτελεσματικότητά τους δεν υπερβαίνει τα 5-7.5 μέτρα βάθος, περιορίζοντας αρκετά το εύρος εφαρμογής τους. Αντίθετα με τις γεωμετρικές διορθώσεις, οι ραδιομετρικές επιδράσεις της διάθλασης στις εικόνες και συγκεκριμένα το φαινόμενο των ζαστις, δηλαδή των ιριδισμών που σχηματίζονται στον πυθμένα λόγω της διάθλασης του φωτός στη κυματώδη επιφάνεια του νερού, αντιμετωπίζεται εύκολα στις εναέριες λήψεις. Αυτό επιτυγχάνεται

αυξάνοντας το ύψος πτήσης έτσι ώστε οι επιδράσεις αυτές να είναι μικρότερες της εδαφοψηφίδας ή λαμβάνοντας εικόνες με τον ήλιο χαμηλά στον ορίζοντα. Ωστόσο, στις περιπτώσεις που η κάμερα βρίσκεται μέσα στο νερό και συνεπώς η απόσταση από το αντικείμενο είναι περιορισμένη, οι μέθοδοι που υπάρχουν στη βιβλιογραφία περιορίζονται στη διόρθωση των επιδράσεων του φαινομένου μόνο όταν ο πυθμένας είναι επίπεδος ή όταν η κάμερα είναι ακίνητη, κάτι μη ρεαλιστικό για τις επιχειρησιακές εφαρμογές της υποβρύχιας 3D αποτύπωσης. Επιπλέον, μέθοδοι που βασίζονται σε νευρωνικά δίκτυα, αντικαθιστούν τις περιοχές της εικόνας με μη πραγματική πληροφορία.

Στόχος της διατριβής είναι η μελέτη και η ανάπτυξη νέων μεθόδων για τη βελτίωση της ακρίβειας και της δυνατότητας γενίκευσης των τεχνικών φωτογραμμετρίας, όρασης υπολογιστών και τηλεπισκόπησης που έχουν στόχο την 3D ανακατασκευή του πυθμένα ρηχών περιοχών, τόσο σε μικρής όσο και σε μεγάλης κλίμακας αποτυπώσεις. Στη διατριβή αυτή, για τις αποτυπώσεις μικρής κλίμακας, οι γεωμετρικές επιδράσεις της διάθλασης αντιμετωπίζονται σε εναέριες εικόνες χαμηλού υψομέτρου που λαμβάνονται από μη επανδρωμένα εναέρια μέσα (Unmanned Aerial Vehicles – UAV). Για τις αποτυπώσεις μεγάλης κλίμακας, οι ραδιομετρικές επιδράσεις της διάθλασης αντιμετωπίζονται σε υποβρύχιας εικόνες. Έτσι, αντιμετωπίζονται οι δύο μεγαλύτεροι παράγοντες σφαλμάτων στις μετρήσεις του πυθμένα ρηχών περιοχών μέσω τεχνικών φωτογραμμετρίας και όρασης υπολογιστών, με προϋπόθεση την καλή ορατότητα του πυθμένα και την ήρεμη επιφάνεια του υδάτινου όγκου.

Προς αυτή την κατεύθυνση, σχετικά με τη διόρθωση των γεωμετρικών επιδράσεων της διάθλασης, αναπτύχθηκε μία πρωτότυπη μέθοδος που εκμεταλλεύεται τεχνικές μηχανικής μάθησης ώστε να προβλέπει το σωστό βάθος των 3D νεφών σημείων που έχουν προκύψει από τεχνικές SfM-MVS, γνωρίζοντας μόνο το φαινόμενο βάθος. Η μέθοδος αυτή εκπαιδεύεται σε μερικά εκατομμύρια ζεύγη πραγματικού και φαινομένου βάθους και αποδεικνύεται εξαιρετικά ακριβής και ικανή να γενικεύει σε διαφορετικές περιοχές και συνθήκες λήψης δεδομένων (ύψος λήψης, κάμερα, επικάλυψη κ.α.). Επιπλέον αυτής της μεθόδου, μία δεύτερη μέθοδος αναπτύσσεται επεκτείνοντας μία state-of-the-art μέθοδο με στόχο τη διόρθωση των γεωμετρικών επιδράσεων της διάθλασης στις αρχικές εικόνες. Η επέκταση που πραγματοποιείται μέσω της εκμετάλλευσης της πρώτης μεθόδου βελτιώνει σε μεγάλο βαθμό την υφιστάμενη μέθοδο, πενταπλασιάζοντας τη βυθομετρική ακρίβεια και μετατρέποντας τη state-of-the-art μέθοδο από επαναληπτική σε άμεση. Παράλληλα επεκτείνει τα επιχειρησιακά της βάθη από τα 7.5 μέτρα στα 15 μέτρα τουλάχιστον. Η μέθοδος αυτή επιτρέπει στη συνέχεια τη χρήση των εικόνων αυτών σε οποιοδήποτε λογισμικό παραγωγής 3D νεφών σημείων και επιφανειών, παραδίδοντας σωστότερες ορθοεικόνες και πιο λεπτομερείς υφές. Αποτελέσματα από πειράματα σε πραγματικά και συνθετικά δεδομένα αποδεικνύουν ότι η ακρίβεια των μεθόδων αυτών είναι εντός των αυστηρών πλαισίων του Διεθνούς Οργανισμού Υδρογραφίας και πως υπερτερούν άλλων state-of-the-art μεθόδων με τις οποίες συγκρίνονται εκτενώς. Για την αντιμετώπιση των ραδιομετρικών επιδράσεων της διάθλασης και συγκεκριμένα του φαινομένου των caustics, μία νέα μέθοδος προτείνεται, η οποία σε αντίθεση με τις μεθόδους που παρουσιάζονται στη βιβλιογραφία, δεν περιορίζεται από τη μορφή του ανάγλυφου του πυθμένα. Η μέθοδος εκμεταλλεύεται τεχνικές βαθιάς μηχανικής μάθησης και τη γεωμετρία της 3D σκηνής μέσω των χαρτών βάθους (παράλλαξης)

και διορθώνει το φαινόμενο στην εικόνα, αντικαθιστώντας μόνο τα «προβληματικά» εικονοστοιχεία με τιμές από τα αντίστοιχα της συνταυτισμένης εικόνας. Η μέθοδος αυτή έχει ως αποτέλεσμα την διόρθωση των εικόνων με πραγματική πληροφορία, διατηρώντας όσο το δυνατόν περισσότερα εικονοστοιχεία της αρχικής εικόνας ακέραια. Έστερα από πειράματα σε πραγματικά δεδομένα, αποδεικνύεται ότι οδηγεί σε αύξηση των συνταυτίσεων μεταξύ των εικόνων και τελικώς σε ένα πιο πλήρες 3D νέφος σημείων, δίνοντας δυνατότητες για μείωση του απαιτούμενου αριθμού εικόνων και συνεπώς μείωση του χρόνου στο πεδίο και του χρόνου επεξεργασίας. Επιπλέον αυτών, παρουσιάζεται το πρώτο σύνολο εικόνων με ζαυστις το οποίο θα είναι διαθέσιμο στην επιστημονική κοινότητα μαζί με αληθή δεδομένα διόρθωσης εικόνων και ταξινόμησης.

Όλες οι μέθοδοι που αναπτύχθηκαν στη διατριβή, δεν περιορίζονται από τα συστήματα και τους αισθητήρες λήψης καθώς τα πειράματα που διεξήχθησαν απέδειξαν ότι λειτουργούν κάτω από πλήθος συνθηκών και δεδομένων. Η διατριβή αναπτύσσεται σε 8 κεφάλαια:

Στο Κεφάλαιο 1 γίνεται μία εισαγωγή στις τεχνικές αποτύπωσης πυθμένα και παρουσιάζονται οι λόγοι που καθιστούν την αποτύπωση αυτή αναγκαία. Παρουσιάζονται τα προβλήματα που θα αντιμετωπιστούν, οι γενικότεροι περιορισμοί των υφιστάμενων μεθόδων στη βιβλιογραφία και τα κίνητρα για την ενασχόληση με το συγκεκριμένο αντικείμενο της διατριβής. Στη συνέχεια, σημειώνονται οι γενικοί και ειδικοί στόχοι της διατριβής και παραθέτονται οι πρωτότυπες συνεισφορές. Τέλος, δίνεται η δομή της διατριβής.

Στο Κεφάλαιο 2 παρουσιάζεται το θεωρητικό υπόβαθρο της διατριβής. Περιγράφονται οι βασικές οπτικές και ακουστικές ιδιότητες του νερού και αναλύονται οι γεωμετρικές και ραδιομετρικές επιπτώσεις της διάθλασης στις εικόνες που λαμβάνονται σε ρηγά νερά, μέσα και έξω από αυτά, τόσο για τις μονοεικονικές περιπτώσεις, όσο και για τις πολυεικονικές. Συγκεκριμένα, για τις εικόνες που λαμβάνονται εκτός του νερού και απεικονίζουν τον πυθμένα, το πρόβλημα της φαινομενικής ανύψωσης του πυθμένα παρουσιάζεται. Για τις εικόνες που λαμβάνονται εντός του νερού, το ραδιομετρικό πρόβλημα των caustics περιγράφεται και αναλύεται.

Στο Κεφάλαιο 3 παρουσιάζεται η σχετική βιβλιογραφία. Συγκεκριμένα, ύστερα από εκτεταμένη έρευνα που πραγματοποιήθηκε, περιγράφονται οι πιο σημαντικές βιβλιογραφικές αναφορές σχετικά με την αντιμετώπιση των γεωμετρικών και ραδιομετρικών επιδράσεων της διάθλασης. Για κάθε μία από τις δύο αυτές επιδράσεις, σχολιάζονται οι περιορισμοί των υφιστάμενων μεθόδων, τα ανοιχτά προβλήματα που παρατηρήθηκαν καθώς και το πως αυτά αντιμετωπίζονται στη παρούσα διατριβή.

Στο Κεφάλαιο 4 παρουσιάζονται οι δύο προτεινόμενες μέθοδοι για τη διόρθωση των γεωμετρικών επιδράσεων της διάθλασης σε εναέριες εικόνες χαμηλού υψομέτρου. Κύριος στόχος αυτών είναι η αρχικά η διόρθωση των επιδράσεων της διάθλασης στα 3D νέφη σημείων που έχουν προκύψει από μεθόδους SfM-MVS και στη συνέχεια, εάν απαιτείται από την εφαρμογή, η διόρθωση των αρχικών εικόνων. Αξίζει να σημειωθεί εδώ ότι η πρώτη μέθοδος μπορεί να χρησιμοποιηθεί και ανεξάρτητα, όταν μόνο νέφη σημείων ή προϊόντα τους ζητώνται. Κύριος στόχος της πρώτης μεθόδου είναι η δημιουργία 3D νεφών σημείων με ακριβή βαθυμετρική πληροφορία χωρίς τη γνώση

άλλων δεδομένων παρά μόνο του φαινομένου βάθους των σημείων. Για να επιτευχθεί αυτό, προτείνεται η εκπαίδευση ενός μοντέλου μηχανικής μάθησης που βασίζεται σε Support Vector Machines το οποίο θα είναι ικανό να εκτιμήσει τα σωστά βάθη των σημείων, γνωρίζοντας μόνο τα φαινόμενα. Η ανεξαρτησία της μεθόδου αυτής από το σύστημα UAV, την κάμερα, το ύψος πτήσης και την επικάλυψη των εικόνων θεωρείται εξαιρετικά σημαντική καθώς αυτό το χαρακτηριστικό θα προσδώσει στην προτεινόμενη μέθοδο την άκρως επιθυμητή ικανότητα γενίκευσης, κάτι που δεν συναντάται στις προϋπάρχουσες μεθόδους. Ύστερα από εκτενή έλεγχο, ένα μοντέλο που βασίζεται σε linear Support Vector Regression (SVR) υιοθετείται. Όπως έδειξαν τα πειραματικά αποτελέσματα σε πραγματικά δεδομένα, η μέθοδος αυτή, αντίθετα με την απλή γραμμική παλινδρόμηση (simple linear regression), αποδείχθηκε ότι είναι πιο ισχυρή και ικανή να προβλέπει πολύ ακριβή και όμοια μοντέλα σε κάθε περιοχή ελέγχου, παρά τον έντονο θόρυβο στα δεδομένα εκπαίδευσης, δίνοντας την επιθυμητή ικανότητα γενίκευσης της μεθόδου. Μέχρι σήμερα, η χρήση της απλής γραμμικής παλινδρόμησης που βασίζεται στην ελαχιστοποίηση του τετραγωνικού σφάλματος, απόκρυβε από τους μελετητές τη συστηματικότητα των γεωμετρικών σφαλμάτων και περιόριζε τη χρήση των προβλεπόμενων μοντέλων μόνο στις περιοχές εκπαίδευσής τους. Ωστόσο, στην παρούσα διατριβή, αποδεικνύεται μέσω πλήθους πειραμάτων ότι η προτεινόμενη μέθοδος, δίνει τη δυνατότητα εκπαίδευσης μοντέλων που είναι κατάλληλα για τη διόρθωση 3D νεφών σημείων άλλων περιοχών, με άλλα χαρακτηριστικά λήψης, με πολύ μεγάλη ακρίβεια.

Ωστόσο, για κάποιες εφαρμογές, η γνώση του βυθού μόνο σε επίπεδο 3D νέφους σημείων δεν είναι αρκετή, καθώς απαιτείται και εικονιστική πληροφορία για το βυθό. Για να ικανοποιηθεί αυτή η απαίτηση, η δεύτερη μέθοδος διόρθωσης των γεωμετρικών επιδράσεων της διάθλασης που παρουσιάζεται στη διατριβή διορθώνει τις αρχικές εικόνες από το φαινόμενο. Για την επίτευξη της διόρθωσης αυτής, είναι απαραίτητη η γνώση των σωστών βαθών του πυθμένα στη μορφή ενός Ψηφιακού Μοντέλου Επιφανείας (ΨΜΕ) καθώς και των εσωτερικών και εξωτερικών προσανατολισμών των καμερών. Μέχρι τώρα, τα σωστά αυτά βάθη υπολογίζονταν ύστερα από τέσσερις επαναλήψεις, χωρίς ωστόσο μεγάλη ακρίβεια για βάθη πάνω από 7.5μ. Στη διατριβή αυτή, η χρονοβόρα επαναληπτική διαδικασία παρακάμπτεται από την εκμετάλλευση της πρώτης προτεινόμενης μεθόδου. Αφού διορθωθεί το 3D νέφος σημείων του πυθμένα και συγχωνευθεί με αυτό των ξηρών περιοχών για τη δημιουργία μίας εικόνας ΨΜΕ η παρακάτω διαδικασία ακολουθείται: αρχικά αφαιρείται η ακτινική και έκκεντρη διαστρόφη των φακών από τις εικόνες και για λόγους ταχύτητας, ένας κανάβος δημιουργείται στις εικόνες κάθε πέντε εικονοστοιχεία. Για κάθε οπτική ακτίνα που διέρχεται από τις κορυφές του κανάβου, η τομή της με το ΨΜΕ υπολογίζεται, μέσω της συνθήκης συγγραμμικότητας. Γνωρίζοντας λοιπόν το μήκος αυτής της οπτικής ακτίνας καθώς και το ποσοστό αυτού που βρίσκεται εντός του αέρα και εντός του νερού, μία νέα «μεικτή» εστιακή υπόσταση υπολογίζεται για κάθε εικονοστοιχείο που ανήκει στις κορυφές του κανάβου. Έχοντας υπολογίσει τη νέα αυτή «μεικτή» εστιακή υπόσταση, η νέα θέση του εικονοστοιχείου υπολογίζεται πάλι μέσω της συνθήκης συγγραμμικότητας και η επανασύσταση της τελικής εικόνας πραγματοποιείται μέσω δικυβικής παρεμβολής. Για τη νέα θέση των εικονοστοιχείων που δεν ανήκουν στις κορυφές του κανάβου, πραγματοποιείται τριγωνισμός Delaunay στις κορυφές του κανάβου και για

κάθε τρίγωνο πραγματοποιείται αφινικός μετασχηματισμός. Η διαδικασία αυτή περιγράφεται στη βιβλιογραφία ως piecewise linear transformation. Φυσικά, ο κανάβος αντί για βήμα πέντε εικονοστοιχεία μπορεί να έχει ένα εικονοστοιχείο ή και περισσότερα, αυξάνοντας ή μειώνοντας το χρόνο επεξεργασίας αντίστοιχα. Ωστόσο, ύστερα από πειράματα σε εικόνες 3000 x 4000 εικονοστοιχείων, παρατηρήθηκε ότι μεταξύ του βήματος κανάβου του ενός εικονοστοιχείου και των πέντε, δεν παρατηρείται καμία ουσιαστική διαφορά στην διορθωμένη εικόνα. Αντίθετα με τις υπάρχουσες τεχνικές στη βιβλιογραφία, η διόρθωση των εικόνων, επιτρέπει στη συνέχεια την πραγματοποίηση των διαδικασιών SfM-MVS με μεγαλύτερη ακρίβεια, παραδίδοντας πιο ακριβείς εξωτερικούς και εσωτερικούς προσανατολισμούς καμερών, άμεσα συσχετισμένους με το ήδη διαθέσιμο διορθωμένο ΨΜΕ, μειώνοντας παράλληλα τον θόρυβο των παραγόμενων 3D νεφών σημείων που προκαλείται από τη διάθλαση. Επιπλέον, επιτρέπει την παραγωγή ορθοεικόνων μεγαλύτερης ακρίβειας και αξιοπιστίας καθώς και πιο λεπτομερών υφών 3D μοντέλων του πυθμένα. Αντίθετα με τα αποτελέσματα των αναλυτικών μεθόδων διόρθωσης διάθλασης, η μέθοδος αυτή επιτρέπει τη χρήση των εικόνων σε οποιοδήποτε εμπορικό ή μη λογισμικό SfM-MVS επεξεργασίας.

Στο Κεφάλαιο 5 παρουσιάζονται τα πειράματα που πραγματοποιήθηκαν για την αξιολόγηση των δύο προαναφερθέντων μεθόδων για τη γεωμετρική διόρθωση των επιδράσεων της διάθλασης. Τα πειράματα συνοδεύονται από τα αντίστοιχα αποτελέσματα και την εκτεταμένη αξιολόγηση τους. Προς αυτή την κατεύθυνση και με στόχο την εκτενή διερεύνηση και αξιολόγηση των μεθόδων με όρους ακρίβειας, αποτελεσματικότητας και γενίκευσης σε διαφορετικές συνθήκες και περιοχές, πολυάριθμα πειράματα πραγματοποιήθηκαν κάνοντας χρήση συνθετικών και πραγματικών δεδομένων. Τα συνθετικά δεδομένα, δημιουργήθηκαν ώστε να αποδειχθεί η σχέση πραγματικού και φαινομένου βάθους που παρουσιάζεται στο Κεφάλαιο 4 αλλά και να μελετηθούν υπολειπόμενα σφάλματα που δεν οφείλονται στις προτεινόμενες μεθοδολογίες. Επιπλέον, μέσω αυτών αποφεύγεται η εισαγωγή σφαλμάτων που μπορεί να περιέχονται στα αληθή δεδομένα βάθους μετρημένα από LiDAR ή SONAR και εξαιρούνται από την εκπαίδευση σφάλματα που οφείλονται στην ορατότητα των υδάτων και τον κυματισμό. Για τη δημιουργία τους, η μεθοδολογία της δεύτερης μεθόδου αντιστράφηκε και ως εκ τούτου, μόνο η πρώτη μέθοδος αξιολογήθηκε σε αυτά. Τα ΨΜΕ μοντέλα του πυθμένα που χρησιμοποιήθηκαν δημιουργήθηκαν από περιοδικές συναρτήσεις ώστε το βάθος να είναι συνάρτηση της οριζοντιογραφικής θέσης κάθε σημείου. Τα συνθετικά δεδομένα, έδωσαν επίσης τη δυνατότητα να ελεγχθούν και να αξιολογηθούν απλούστερες μέθοδοι διόρθωσης των φαινομένων βαθών που υπάρχουν στη βιβλιογραφία, αποδεικνύοντας και πάλι την υπερίσχυση της προτεινόμενης μεθόδου. Τα πραγματικά δεδομένα αποτελούνται από μπλοκ εικόνων από πέντε διαφορετικές περιοχές στην Ελλάδα και την Κύπρο και χαρακτηρίζονται από διαφορετικά χαρακτηριστικά πυθμένα, πτήσης και δύο διαφορετικά συστήματα UAV. Μαζί με τα μοντέλα που εκπαιδεύθηκαν στα πραγματικά δεδομένα, και αυτά που εκπαιδεύθηκαν στα συνθετικά δεδομένα, εφαρμόστηκαν στα πραγματικά, επιβεβαιώνοντας τις δυνατότητες διόρθωσης και γενίκευσης των προτεινόμενων μεθόδων. Τα αποτελέσματα των προτεινόμενων μεθόδων ελέγχθηκαν μέσω της σύγκρισής τους με αληθή δεδομένα LiDAR και σημείων που συλλέχθηκαν με Ολοκληρωμένο Γεωδαιτικό Σταθμό. Τα υπολειπόμενα σφάλματα δείχνουν ότι και οι δύο

μέθοδοι είναι ικανές για τη διόρθωση του βάθους και των γεωμετρικών επιδράσεων της διάθλασης με μεγάλη ακρίβεια, ξεπερνώντας παράλληλα τις επιδόσεις άλλων state-of-the-art μεθόδων με τις οποίες συγκρίθηκαν.

Στο Κεφάλαιο 6 παρουσιάζεται η προτεινόμενη μέθοδος για τη διόρθωση των ραδιομετρικών επιδράσεων της διάθλασης στις υποβρύχιες εικόνες. Πιο συγκεκριμένα, η μέθοδος που παρουσιάζεται αρχικά εκμεταλλεύεται ένα Fully Convolutional Neural Network (FCN) ώστε να ταξινομηθεί η εικόνα σε επηρεασμένες και μη περιοχές και να παραχθούν οι αντίστοιχες δυαδικές εικόνες-μάσκες. Στη συνέχεια, η μέθοδος βασισμένη στη γεωμετρία της σκληνής αντικαθιστά κάθε ένα από τα εικονοστοιχεία που βρίσκονται εντός των επηρεασμένων περιοχών με τα αντιστοιχισμένα των συνταυτισμένων εικόνων. Για να επιτευχθεί αυτό, αρχικά, οι δυαδικές εικόνες-μάσκες χρησιμοποιούνται στο στάδιο της ανίχνευσης των χαρακτηριστικών σημείων των εικόνων ώστε αυτά να ανιχνευθούν μόνο στις μη επηρεασμένες περιοχές, καθώς διαφορετικά η διαδικασία της συνταύτισης θα αποτύχει. Στη συνέχεια, οι εικόνες συνταυτίζονται και η επιπλοκή γεωμετρία ανακτάται. Οι εικόνες αυτές υπόκεινται σε χρωματική επεξεργασία ώστε να επιτευχθεί η ομοιογένεια των τόνων στην νέα διορθωμένη εικόνα. Στη συνέχεια, οι εικόνες μετατρέπονται σε επιπολικές και οι χάρτες παράλλαξης τους υπολογίζονται μέσω της τεχνικής Consistent Semi-Global Matching (CSGM). Στη συνέχεια, μέσω των χαρτών παράλλαξης, κάθε εικονοστοιχείο σε μία εικόνα που ανήκει στην κλάση «caustics», αντικαθίσταται με το αντίστοιχο των συνταυτισμένων εικόνων, αν δεν ανήκει στην ίδια κλάση. Τέλος, οι διορθωμένη επιπολική εικόνα προβάλλεται στην αρχική της κεντρική προβολή. Η μεθοδολογία αυτή επιτρέπει τη διόρθωση εικόνων που απεικονίζουν πυθμένα οποιουδήποτε ανάγλυφου. Επιπλέον, δίνει τη δυνατότητα για περεταίρω εκμετάλλευση της μεθόδου σε εναέριες ή δορυφορικές εικόνες για την αντικατάσταση περιοχών που καλύπτονται από νέφη ή άλλη πηγή απόκρυψης. Η μέθοδος αυτή, σέβεται την παράλλαξη των αντικειμένων στις εικόνες και ως εκ τούτου παραδίδει εικόνες κατάλληλες για περεταίρω επεξεργασία με στόχο την παραγωγή 3Δ.

Στο Κεφάλαιο 7 παρουσιάζονται τα πειραματικά αποτελέσματα της παραπάνω προτεινόμενης μεθόδου καθώς και το πρώτο σύνολο εικόνων με ζαυστις το οποίο θα είναι διαθέσιμο στην επιστημονική κοινότητα μαζί με αληθή δεδομένα διόρθωσης εικόνων και ταξινόμησης. Παράλληλα με την αξιολόγηση του συνόλου της μεθόδου και των τελικών της αποτελεσμάτων, δικαιολογείται η χρήση μίας βαθιάς αρχιτεκτονικής όπως το FCN που χρησιμοποιήθηκε στην παρούσα διατριβή, έναντι άλλων ρηχότερων. Σε επίπεδο εικόνας, η προτεινόμενη μέθοδος καταφέρνει να διορθώσει τις επηρεασμένες περιοχές με μεγάλη ακρίβεια και αξιοπιστία. Σε επίπεδο βελτίωσης της 3Δ ανακατασκευής, τα αποτελέσματα είναι ακόμα πιο εντυπωσιακά καθώς μέσω της διόρθωσης, επιτυγχάνεται η δημιουργία πλήρους 3Δ νέφους σημείων, σε περιοχές που οι εικόνες με ζαυστις το απέτρεπαν.

Στο Κεφάλαιο 8 παρουσιάζονται η σύνοψη και η συμβολή της διδακτορικής διατριβής καθώς και τα συμπεράσματα που προέκυψαν, πλαισιωμένα από ιδέες για μελλοντική έρευνα στο πεδίο.



**ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ
ΠΟΛΥΤΕΧΝΕΙΟ**
ΕΙΔΙΚΟΣ ΛΟΓΑΡΙΑΣΜΟΣ
ΚΟΝΔΥΛΙΩΝ ΕΡΕΥΝΑΣ

Η διδακτορική διατριβή εκπονήθηκε με χρηματοδότηση από τον Ειδικό Λογαριασμό Έρευνας Ε.Μ.Π για το διάστημα από τον Φεβρουάριο 2018 μέχρι τον Ιούλιο 2020.

Contents

1	Introduction	1
1.1	Motivation and Current Challenges	3
1.1.1	Geometric effects	4
1.1.2	Radiometric effects	5
1.2	Goal of the Thesis	6
1.2.1	Objectives	7
1.2.2	Original Contributions of the Thesis	7
1.3	Outline of the Thesis	9
2	Theoretical background	11
2.1	Properties of water	11
2.1.1	Optical properties	12
2.1.2	Acoustic properties	12
2.2	Geometric Effects Caused by Specific Water Properties	13
2.2.1	Snell's law	13
2.2.2	Single view geometry in the through-water depth determination	14
2.2.3	Stereo view geometry in the through-water depth determination	15
2.2.4	Multiple-view geometry in the through-water depth determination	17
2.3	Radiometric Effects	18
2.3.1	Caustics effect	19
2.3.2	Rippling Caustics in key point detection and matching processes	20
3	Related Work	25
3.1	Compensation for Refraction's Geometric Effects	25
3.1.1	Analytical and Image-based Refraction Correction	26
3.1.2	Image-based Bathymetry Estimation using Machine Learning and simple Regression Models	29
3.1.3	Discussion	30
3.2	Radiometric Effects	31
3.2.1	Caustics Detection and Removal	31
3.2.2	Discussion	32
4	Correcting the Geometric Effects of Refraction	33
4.1	Correcting the Geometric Effects of Refraction in the 3D Space	34
4.1.1	Proposed method	34
4.1.2	The proposed refraction correction model	35
4.2	Correcting the Geometric Effects of Refraction in the 2D (Image) Space	41
4.2.1	Proposed Method	41
4.2.2	Proposed Refraction Correction Model	42

4.2.3	Summary	47
5	Experimental Results and Validation of the Developed Geometric Correction Methods	49
5.1	Synthetic Training and Test Datasets	50
5.1.1	Synthetic Data Generation	50
5.2	Experimental Results and Validation on the Synthetic Datasets	56
5.2.1	Correcting the Geometric Effects of Refraction in the 3D space	56
5.2.2	Is a ratio-based depth determination better?	69
5.2.3	Summary	73
5.3	Real World Test Datasets	74
5.3.1	Datasets Description	75
5.3.2	Factors affecting the low altitude aerial imagery over shallow waters	80
5.3.3	Ground Control Points measurements	81
5.3.4	Initial Dense Point Clouds Generation	81
5.3.5	The influence of the Base-to-Height ratio (B/H) of stereopairs on the apparent depths	83
5.3.6	True depth data	85
5.3.7	Seabed point cloud extraction	87
5.4	Experimental Results and Validation on Real World Data	89
5.4.1	Correcting the Geometric Effects of Refraction in the 3D space	89
5.4.2	Correcting the Geometric Effects of Refraction in the 2D (Image) Space	122
5.5	Comparative analysis of the proposed methods	139
5.6	Discussion	145
6	Correcting the Radiometric Effects of Refraction	149
6.1	Proposed method for pixelwise image radiometric correction	149
6.1.1	Image classification	151
6.1.2	Stereo-rectification and Disparity maps generation	152
6.1.3	RGB values replacement and back-projection of the images	155
6.2	Summary	155
7	Experimental Results and Validation of the Developed Pixelwise Correction Method	157
7.1	The first real world dataset on underwater caustics	157
7.1.1	Data Collection	159
7.2	Pixelwise image correction	161
7.2.1	Image classification	161
7.2.2	FCN selection and justification	168
7.2.3	Pixelwise image radiometric correction	171
7.2.4	Improvements on the 3D reconstruction	173
7.3	Discussion	177

8 Conclusion of the Thesis	181
8.1 Summary	181
8.2 Original Contributions of the Thesis	183
8.3 Future Work	185
8.4 Research Framework and Related Publications	186
8.4.1 Compensation for Refraction's Geometric Effects	187
8.4.2 Compensation for Refraction's Radiometric Effects	187
8.4.3 Close Range Underwater 3D reconstruction and Color Enhancement	188
Bibliography	191

List of Tables

5.1	The six simulated aerial UAV campaigns, the derived datasets and relative information.	55
5.2	Differences between the calculated and the real elevations of the point clouds generated using the non-refracted datasets. \bar{x} is the average distance of the point cloud from the true elevations and s its standard deviation.	56
5.3	The results of the comparisons between the predicted models for all tests performed.	68
5.4	The results of the comparisons between the true depth and the uncorrected depth, the depths derived by the implementation of the SVR based approach and the ratio based approach.	72
5.5	Indicative parts of the images containing (a) waves breaking on the shore, (b) turbid waters, (c) sun glint, (d) caustics, (e) seabed with texture and (f) seabed without texture (except of the lower left area of the image).	79
5.6	Results of this comparison between the initial dense point clouds and the point clouds generated with the double B/H ratio.	84
5.7	The differences between the initial point clouds generated using all the available imagery and the point clouds generated using imagery acquired in specific altitudes.	85
5.8	True data specifications. Numbers in parenthesis are for the sparse point clouds	86
5.9	The results of the comparisons between the predicted models for all tests performed. \bar{x} is the average distance of the point cloud from the true and s its standard deviation. Negative values suggest overestimation of the depth and positive suggest underestimation.	111
5.10	Quantitative Evaluation. \bar{x} is the average distance of the point cloud from the true values and s is its standard deviation. Negative values suggest overestimation of the depth and positive suggest underestimation.	129
5.11	Comparative quantitative Evaluation. \bar{x} is the average distance of the point cloud from the true values and s is its standard deviation. Negative values suggest overestimation of the depth and positive suggest underestimation. The best values achieved for each test site are highlighted in bold format.	144
7.1	The testing performance of the various models after 10 epochs of training.	165
7.2	The testing performance of the various methods. Number in bold are the best metrics achieved.	169

7.3 The evaluation metrics between the uncorrected (uc) and the corrected (c) images. D is the surface density and R the roughness. Numbers for these two metrics are relative. 177

List of Figures

2.1	Snell's law in standard form (left) and three-dimensional vector format (right).	13
2.2	Geometry of the single view in the though-water depth determination (reproduced after [Shan 1994])	14
2.3	The geometry of two media photogrammetry in stereo case [Fryer 1985].	16
2.4	The geometry of two-media photogrammetry for the multiple view case, retrieved from [Agrafiotis 2019a, Agrafiotis 2019b].	18
2.5	Due to waves on the water surface, the refraction of natural sky illumination is spatially varying. This creates 3D patterns of variable light flux and caustics underwater, and 2D illumination patterns on the illuminated objects. These patterns vary in time due to the dynamics of the surface waves.	19
2.6	Images with caustics of the same seabed area from the same camera position, captured with an interval of 5 seconds (from left to right) and the detected key points using SIFT [Lowe 1999]. In the first row the full images are demonstrated while in the second row zoomed areas are depicted.	21
2.7	Matching results between the image captured at t and $t+5$ seconds (a), between the image captured at $t+5$ and $t+10$ seconds (b) and between the image without caustics and itself (c).	22
4.1	The proposed method for correcting the refraction in the point clouds based on a Linear Support Vector Regression (SVR) model	35
4.2	The geometry of the apparent point cloud generation using SfM-MVS techniques.	36
4.3	Illustration of the employed SVM regression model, showing the regression line together with the insensitive "tube". Also shown are examples of the slack variables ξ_n and $\hat{\xi}_n$. Points above the ε "tube" have $\xi_n > 0$ and $\hat{\xi}_n = 0$, points below the -tube have $\hat{\xi}_n > 0$ and $\xi_n = 0$, and points inside the ε "tube" have $\xi_n = \hat{\xi}_n = 0$	37
4.4	The established correlations based on a simple linear regression and SVM linear regression models, trained on Amathounda and Agia Napa datasets.	40
4.5	The workflow of the proposed method for image correction.	42
4.6	The geometry of a single image formation and the repositioning of an image point based on the calculated focal length (cmixed).	43
4.7	The process of the creation of the corrected new image by resampling using the initial grid and the transformed grid and bicubic interpolation. It is assumed a nadir image.	46

5.1	The two different synthetically generated DTMs. X, Y, Z are in meters. DTM1 is illustrated in (a) and DTM2 in (b)	52
5.2	The real seabed texture used for both the DTMs	52
5.3	Example images of the generated synthetic imagery without refraction for the DTM1 at 200m flying (a) and for the DTM2 at the same flying height (b). The respective generated synthetic images with refraction are presented in (c) and (d), having exactly the same exterior orientation with the unrefracted ones.	54
5.4	An example 2D histogram of the differences between the real (resulted from the Z function) and the SfM-MVS calculated elevations using the non-refracted imagery for the 150m[DTM1][non-refracted] dataset, together with explanatory sections (1D histograms) along X and Y axis.	57
5.5	The 2D histograms of the differences between the real (resulted from the Z function) and the SfM-MVS calculated elevations using the imagery for the 150m[DTM1][non-refracted] dataset (a), the 200m[DTM1][non-refracted] dataset (b) and the 200m[DTM2][non-refracted] dataset (c) respectively, in relation to the true elevation.	58
5.6	The Depth (Z) - Z_0 2D histograms of the synthetic datasets and the respective point densities of the 100% of the correlated points: The histograms for 150m[DTM1] (a), for 200m[DTM1] (b) and for 200m[DTM2] (c) datasets are presented. Figures (d), (e) and (f) are presenting the depth (Z) - Z_0 distribution and the respective point densities of the 95.4% of the correlated points respectively.	59
5.7	The Z - Z_0 distribution of the employed synthetic datasets and their respective, predicted linear models.	62
5.8	The 2D histograms of the differences between the real and the corrected image-based dense point clouds derived from the SfM-MVS for the 150m[DTM1] dataset (a), the 200m[DTM1] dataset (b) and the 200m[DTM2] dataset (c) respectively, in relation to the real depth. Figures (d), (e) and (f) are presenting the differences and the respective point densities of the 95.4% of the correlated points respectively. The dashed lines in red represent the accuracy limits generally accepted for hydrography as introduced by the IHO [Guenther 2000].	63
5.9	An example 2D histogram of the differences between the true depths and the corrected depths produced for the predicted model, together with explanatory sections (1D histograms) along X and Y axis.	64
5.10	The 2D histograms of the differences between the true depth, resulting from the Z function and the corrected point clouds after the application of the proposed method, in relation to the real depth. The red dashed lines represent the accuracy limits generally accepted for hydrography, as introduced by the IHO [Guenther 2000].	67

5.11	The 2D histograms of the differences between the true depths and the corrected depths produced after the multiplication of the coefficient with the apparent depths in relation to the true depth. For the presented cases the ratio is calculated on the same test site on which testing is performed. The red dashed lines represent the accuracy limits generally accepted for hydrography, as introduced by the IHO [Guenther 2000].	69
5.12	The 2D histograms of the differences of a cross site approach; In the left column, the 2D histogram of the differences between the true depths and the corrected depths of the 200m[DTM2] synthetic site produced by the proposed method trained on the 200m[DTM1] synthetic site is presented. In middle right column, the respective histogram of differences of the derived depths after the multiplication with the average ratio calculated on the 200m[DTM1] synthetic site. In the right column, the respective histogram of differences of the derived depths after following the method presented in [Chirayath 2019] are presented The red dashed lines represent the accuracy limits generally accepted for hydrography, as introduced by the IHO [Guenther 2000].	70
5.13	The five different test areas used.	75
5.14	Indicative images from the Agia Napa test area. Depth decreases from the left to the right image.	76
5.15	Indicative images from the Amathounta test area. Depth decreases from the left to the right image.	76
5.16	Indicative images from the Dekelia test area. Depth decreases from the left image to the right image.	77
5.17	Indicative images from the Cyclades-1 test area. Images are depicting the same area of the site captured from 88m (left), 70m (center) and 35m (right).	77
5.18	Indicative images from the Cyclades-2 test area. Left and center images are depicting the same area of the site captured from 75m and 33m respectively, while the image on the right depicts the rocky formations of the seabed and the shore.	78
5.19	The five aerial UAV campaigns, the derived datasets and related information.	80
5.20	The initial dense point clouds for Amathounta (a), Agia Napa (b), Dekelia (c), Cyclades-1 (d) and Cyclades-2 (e) test sites. The yellow triangles represent the GCPs positions. Scale bars are in meters.	82
5.21	The spatial patterns of the remaining errors on the compared point clouds of Agia Napa (a) and Amathounta (b) datasets. M3C2 distances and scale bars are in meters (m).	84
5.22	The initial seabed point clouds for Amathounta (a), Agia Napa (b), Dekelia (f), Cyclades-1 (i) and Cyclades-2 (j) test sites. Scalebars and apparent depths are in meters.	89

5.23	The initial seabed point clouds for Amathounta (a), Agia Napa (b) and Dekelia (c) test sites and the respective areas of the point clouds used for training the presented method. Scale bars are in meters.	91
5.24	(a): Part I (left) and Part II (right) datasets from the Agia Napa test site in blue, the Amathounta test site in orange, and the Dekelia test site in magenta, all having the same scale (b): The distribution of the Z and Z_0 values for each dataset.	92
5.25	The depth to apparent depth ($Z-Z_0$) distribution of the datasets prepared for training: the Agia Napa Part I dataset over the full Agia Napa dataset (a). The Agia Napa Part II dataset over the full Agia Napa dataset and (b) the Amathounta dataset (c); the merged dataset over the Agia Napa and Amathounta datasets (d) and the Dekelia dataset; and (e) the Dekelia dataset over the full Agia Napa dataset (f).	93
5.26	The $Z-Z_0$ distribution of the employed datasets and their respective, predicted linear models.	97
5.27	The $Z-Z_0$ distribution of the used datasets and the respective point densities are presented.	98
5.28	The $Z-Z_0$ distribution of the employed datasets and their respective, predicted linear models.	100
5.29	Typical stereopairs and their matches in the deepest areas of the Agia Napa [Part I] (a) and the Agia Napa [Part II] (b) test sites. The valid matches are presented with blue color while the invalid matches with red. The orange dotted line represents the isodepth line of 10m.	102
5.30	The initial M3C2 distances (in m) between the (reference) LiDAR point cloud and the uncorrected image-based point clouds derived from the SfM-MVS. (a) M3C2 distances of Amathounta, (b) M3C2 distances of Agia Napa [Part II], (c) M3C2 distances for Agia Napa [Part I] test site and (d) M3C2 distances for Dekelia test site.	103
5.31	The 2D histograms of the M3C2 distances between the true and the uncorrected and corrected image-based dense point clouds derived from the SfM for Amathounta, Agia Napa, Dekelia, Agia Napa [Part I], Agia Napa [Part II], Dekelia [GPS], Cyclades-1 and Cyclades-2 test sites respectively, in relation to the real depth. The dashed lines in red represent the accuracy limits generally accepted for hydrography as introduced by the IHO [Guenther 2000].	105
5.32	The 2D histograms of the M3C2 distances between the LiDAR point cloud and the corrected point clouds after the application of the proposed method in relation to the real depth. The red dashed lines represent the accuracy limits generally accepted for hydrography, as introduced by the IHO [Guenther 2000].	108
5.33	The footprints of the representative cross sections. Scalebars are in meters (m).	113

5.34	The footprints of the representative cross sections. Scalebars are in meters (m).	114
5.35	The cloud to cloud (M3C2) distances between the LiDAR point cloud and the corrected point clouds after the implementation of the proposed method, demonstrating the spatial patterns of the remaining errors. M3C2 distances are in (m).	116
5.36	(a): the corrected image-based point cloud, (b): the non-colored LiDAR point cloud and (c): the colored LiDAR point cloud based on (a).	118
5.37	Three different sample areas of the Agia Napa seabed with depth between 9.95m to 11.75m. The first two columns depict two sample areas of the Agia Napa[I] seabed while the third depicts a sample area of the Agia Napa [Part II] seabed. The first row depicts the three areas on the image-based point cloud, the second row on the LiDAR point cloud colored with the corrected image-based point cloud and the third row depicts the three areas on the LiDAR point cloud colored using the uncorrected image-based one. Point size is multiplied by a factor of 4 in order to facilitate visual comparisons.	119
5.38	An example of the seamless hole filling area and indicative results.	120
5.39	The extracted seabed point clouds after the depth correction using the SVR model trained on the Dekelia dataset for Amathounta (a), Agia Napa (b), Cyclades-1 (d) and Cyclades-2 (e) test sites. Scalebars visualizing the depths are in meters (m).	123
5.40	The updated DSM after the depth correction using the SVR model trained on the Dekelia dataset for Amathounta (a), Agia Napa (b), Cyclades-1 (c) and Cyclades-2 (d) test sites.	124
5.41	Initial aerial data (first row), refraction-free images (second row) after the depth correction using the SVR model trained on the Dekelia dataset and their subtraction (third row).	125
5.42	The 2D histograms of the M3C2 distances between the true and the uncorrected and corrected image-based sparse point clouds derived from the SfM for Amathounta (first coloumn), Agia Napa (second coloumn), Cyclades-1 (third coloumn) and Cyclades-2 (fourth coloumn) test sites respectively. The dashed lines in red represent the accuracy limits generally accepted for hydrography as introduced by the IHO [Guenther 2000]. M3C2 distances are in (m).	127
5.43	Indicative textured 3D models from Amathounta (first row), Agia Napa (second row), Cyclades-1 (third row), and Cyclades-2 (fourth row) test sites with the initial uncorrected (left column) and the refraction-free (right column) datasets. For the generation of the presented refraction-free datasets the SVR model trained on the Dekelia dataset was exploited.	131

5.44	Parts of the orthoimages of Amathounta (first row), and Agia Napa (second row), Cyclades-1 (third row) and Cyclades-2 (fourth row) test sites. In the left column the orthoimages generated using the uncorrected imagery are presented while in the right column the results using the refraction corrected imagery are presented. For the generation of the presented refraction-free datasets the SVR model trained on the Dekelia dataset was exploited.	133
5.45	The computed radial distance differences and the c_{mixed} differences for all the test sites.	135
5.46	The radial distance differences for Amathounta and Agia Napa test sites are presented in (a) and (b) for depths of 5m and 15m respectively and the radial distance differences for Cyclades-1 and Cyclades-2 test sites are presented in (c) and (d) for depths of 5m and 15m respectively.	138
5.47	The histograms of the M3C2 distances between the true and the uncorrected and corrected image-based sparse point clouds derived from the SfM for Amathounta (first column), Agia Napa (second column), Cyclades-1 (third column) and Cyclades-2 (fourth column) test sites respectively. The dashed lines in red represent the accuracy limits generally accepted for hydrography as introduced by the IHO [Guenther 2000]. M3C2 distances are in (m).	142
6.1	The proposed method for the pixelwise correction of the rippling caustics on the underwater imagery.	150
6.2	An illustration of the SegNet architecture. A decoder upsamples its input using the transferred pool indices from its encoder to produce a sparse feature map. It then performs convolution with a trainable filter bank to densify the feature map. The final decoder output feature maps are fed to a soft-max classifier for pixel-wise classification. (Retrieved from [Badrinarayanan 2017])	152
7.1	Sample images from the constructed dataset. First column: RGB images; Second column: the reference images; Third column: the difference images between the images with caustics and the reference images; Fourth column: the thresholded difference image; Fifth column: the RGB images with the detected contours.	158
7.2	The reference image created using the artificial shadow in (a) and the reference image created using the pixels having the lowest values compared with the pixels of the same position in the rest of the imagery in (b). In the red rectangle the multiple fished appearing in the reference image are highlighted.	160
7.3	Training loss and cross-validation accuracy for all the models trained.	163

7.4	Training loss and validation accuracy for all the models trained on datasets 3, 4, 5, 6 and 7 for 0.01 base learning rate (a) and for 0.0001 base learning rate (b)	164
7.5	Typical examples of the RGB images (first column), the ground truth binary images (second column) and the predicted binary images (third column) using the models already described.	166
7.6	Various real-world RGB images (left column) and the respective predictions using the model trained on all the available datasets (right column).	167
7.7	The input RGB image together with the predictions of the tested architectures for an image from dataset 4.	169
7.8	The input RGB image together with the predictions of the tested architectures for an image of a real world application, not contained on the datasets.	170
7.9	Examples of corrected images. The first column depicts the original images with caustics, the second column depicts the binary images predicted by the trained FCN model and the third column demonstrates the corrected imagery.	172
7.10	A corrected image by the proposed methodology with the color transferring step (a) and the same image corrected by the proposed methodology without applying the color transferring step (b).	173
7.11	The 3D reconstructions using the reference imagery (left column), the uncorrected imagery (middle column) and the corrected imagery by the proposed methodology (right column). Each row refers to a different test case.	174
7.12	The resulting 3D point clouds for the rest three tests performed over real-world imagery. In the left column the 3D point cloud generated using the original uncorrected imagery is presented while in the right column, the respective 3D point cloud of the corrected imagery is presented.	175

Introduction

Contents

1.1 Motivation and Current Challenges	3
1.1.1 Geometric effects	4
1.1.2 Radiometric effects	5
1.2 Goal of the Thesis	6
1.2.1 Objectives	7
1.2.2 Original Contributions of the Thesis	7
1.3 Outline of the Thesis	9

Since the ancient years, mankind is depended on water bodies either in their natural forms, such as oceans, lakes, rivers, wetlands, or in their man-made counterparts, such as flumes, channels, basins, dams etc. Water has also been an valuable resource for irrigation and an important mean for transportation and trade, other than being the most effective way to explore new continents in ancient times. Evidence of human life from the very beginning hides under the water level, off the coasts, under shallow seas or deep oceans, but also inland water bodies all around the world. Since the dawn of civilization, humans have been drawn to the world's oceans and the unknown wonders that lay beneath their mysterious surfaces. After all, about 71% of the Earth's surface is water-covered, and the oceans hold about 96.5% of all Earth's water. This is about 362 million square kilometers of the total surface area [Eakins 2010]. Even though only a small fraction has been mapped by direct observation so far, the last few years have seen an increase in the recognition of the importance of seafloor mapping [Wölfel 2019]. This need for better knowledge of the bottom of water bodies (seabed, riverbed, lakebed), is driven by habitat destruction, marine pollution, cultural heritage at risk, recent tragedies such as sinking of ships and plane crashes as well as natural disasters, navigation and the increasing demand for offshore energy and marine resources.

Oceans and seas host an incredibly rich biodiversity, influence short- and long-term climate and have a high impact, not only on economy, but on the whole life and evolution of human society. Oceans are mitigating weather extremes and generate the oxygen necessary for life. They produce income through many activities and food for many species while they store the excess carbon dioxide generated by human activities. However, together with various geological phenomena, the effects of increasing greenhouse gas emissions threaten coastal and marine ecosystems

through changes in ocean temperature and melting of ice, which in turn affect ocean currents, weather patterns, and sea level. This degradation of coastal and marine ecosystems threatens the physical, economic and food security of coastal communities. To forecast or mitigate these effects, the sustainable management, conservation and restoration of coastal and marine ecosystems seems necessary. To perform these actions, accurate, reliable and inexpensive mapping of the coastal areas seems to be crucial.

At the same time, human activities generate considerable amounts of waste much of which escapes management schemes and ends up in the environment via rivers, stormwater, wind and sewage, or can be disposed of directly at beaches, lakes and sea [UNEP/MAP 2015, Consoli 2020]. This amount of marine litter, mainly composed of plastic materials, has become a global environmental issue in coastal environments. Traditional monitoring programs are based on in-situ visual census, which require human effort and are time-demanding [Gonçalves 2020]. Therefore, it is crucial to implement innovative and low-cost mapping strategies to improve the environmental monitoring of marine litter on the coast.

Submerged Cultural Heritage such as shipwrecks and underwater ruins, are coming also under increasing threat. Many archaeological sites under water are subject to heavy looting. Their exploitation and the sale of the objects found is reminiscent of events that took place a hundred years ago on many archaeological sites on land. However, submerged archaeological sites are not only threatened by human activities, but also by changes in the environment caused by earthquakes, storms, temperature variations, changing currents or coastal erosion, just like coastal and marine ecosystems. Usually, a site that lies buried under sediment reaches a stable anaerobic state with low levels of oxygen and light that reduces the process of material degradation caused by chemical, physical and biological factors [UNESCO]. To that direction, recording, documenting and, ultimately, protecting underwater cultural heritage is an obligation of mankind and dictated by international treaties like the Convention on the Protection of the Underwater Cultural Heritage that fosters and encourages the use of "non-destructive techniques and survey methods in preference over the recovery of objects" [Menna 2018]. Image-based and mapping techniques represent an invaluable set of effective tools for reconnaissance, documentation, monitoring, but also public dissemination and awareness of underwater cultural heritage assets [Menna 2018].

Seabed, riverbed and lakebed bathymetry is also essential for the safety of navigation. In the shallow waters and coastal areas, safety of vessel navigation is the most prioritized rationale for bathymetric mapping close to the coast, around shoals and along shipping routes. This mapping is the basis for the production of nautical charts [Hell 2012] and it is also essential for establishing the limits of the extended continental shelf (ECS) under the United Nations Convention on Law of the Sea (UNCLOS) [Jakobsson 2003]. Marine infrastructure development, such as cable laying, pipeline and platform installation, rig anchoring, or deployment of machines requires high-resolution bathymetric data too [Wölf 2019]. This need for high-resolution bathymetric data for monitoring of bottom activities will increase

in the future [Clark 2017, Ellis 2017] together with the identification and characterization of areas suitable for seafloor mining which also rely on precise bathymetric information [Hein 2009].

However, not until the twentieth century, did technology truly allowed coastal marine scientists to visualize and interpret the topography of benthic environments on a wide scale [Makowski 2016]. With the advent of modern remote sensing techniques (e.g., seismic reflection profiling, aerial imagery, satellites, LiDAR), researchers have now gained the ability to effectively interpret and map large portions of the dynamically changing coastal environments along continental shelf margins. Summing up the above, accurate bathymetric mapping seems to be a key element for offshore activities, hydrological studies such as coastal engineering applications, sedimentary processes, hydrographic surveying as well as archaeological mapping and biological research. Through Structure from Motion (SfM) and Multi View Stereo (MVS) techniques, images can provide a low-cost alternative compared to other methods such as LiDAR and acoustic surveys, offering as well, important visual information. To that direction, this thesis is concerned with the study and development of new methods for improving the performance and the accuracy of image-based mapping and 3D reconstructing the bottom in shallow waters for small- and large-scale surveys. Next, some background on the motivations, the current challenges and the applicability will be presented, as well as a description of the thesis' objectives and the outline of this document.

1.1 Motivation and Current Challenges

Image-based methods represent an effective tool for measuring, mapping, researching and monitoring the bottom of water bodies such as sea, rivers and lakes. Depending on the needs and on the environmental conditions such as depth, bottom characteristics, water surface state and water turbidity, sensors, techniques and methods may need to be used differently or may even may not be suitable at all [Menna 2018]. Despite their relative low cost compared to other methods and systems, their major drawback is that optical properties and illumination conditions of water severely affect overwater and underwater imagery and data acquisition process [Agrafiotis 2018a]. Colors are lost as the depth increases, resulting in a green-blue image effect due to light absorption, which mainly influences red wavelength. Therefore, red channel histogram has fewer values compared to green and blue. Water also absorbs light energy and scatters optical rays creating blurred images, reducing the exploitable visibility to a few meters [Mangeruga 2018, Agrafiotis 2017, Agrafiotis 2018b]. However, when it comes to shallow waters, refraction seems to be the main factor affecting the geometry and the radiometry of the primary data and consequently of the results of overwater and underwater image-based 3D reconstruction methods.

1.1.1 Geometric effects

Although in the literature, the **geometric effects** of refraction have been addressed thoroughly when the camera sensor is underwater, water refraction is still posing the most significant geometrical errors, affecting mostly the depth determination when the camera is over the water i.e. in small and large scale aerial mapping applications. In through water cases, i.e. when the camera is outside the water, the water surface variation due to waves [Skarlatos 2019, Agrafiotis 2018b], the solar reflections and especially the effects of refraction which differ in each image, lead to significant geometric errors and unstable solutions [Mangeruga 2018, Menna 2018] contrary to underwater scenarios, where according to the literature [Lavest 2000, Shortis 2019, Elnashef 2019], camera calibration is sufficient to correct the effects of refraction.

More specifically, the amount of the refraction of a light beam is affected by the amount of the water that covers the point of origin situated on the bottom and the angle of incidence of the beam in the water/air interface (see also Section 2.2). In underwater cases this interface is located at a constant distance and orientation from the camera sensor due to the fixed camera housings and thus it can be modelled. In overwater cases, images captured from different positions, different angles and different distances alter the amount of water that covers the imaged points, but also they significantly alter the angles of incidence of the light rays, thus affecting the amount of refraction and leading to unreliable results [Agrafiotis 2015]. Therefore, the common calibration procedures of the camera, fail to model and describe the effects of refraction in the images in such cases, as they fail to provide a reliable result. Hence, a much more sophisticated solution should be developed for correcting the refraction effects in these cases. This is why bathymetry mapping from aerial platforms in shallow waters is considered a much more time-consuming and costly process, compared to onshore (dry land) aerial mapping. However, it is still a more efficient operation than ship-borne echo-sounding methods or underwater close-range photogrammetric methods [Agrafiotis 2018a], especially when it comes to shallower clearwater areas. A very important additional feature of image-based bottom mapping is that a permanent record of other features is obtained in the coastal region, such as tidal levels, coastal dunes, benthic communities, marine litter, rock platforms and beach erosion [Karara 1979]. These benefits are especially evident in coastal zones, where most of the economic activities are concentrated, even though many alternatives for bathymetry have been reported recently [Kasvi 2019].

On the one hand, image-based mapping techniques generally fail due to wave breaking effects and water refraction, which till now, has been addressed through customized and application-oriented refraction correction algorithms or by modifying the collinearity equation. On the other hand, echo-sounding techniques fail due to multipath errors and artefacts, requiring special algorithms and post processing strategies [Skarlatos 2018, Menna 2013]. At the same time, bathymetric LiDAR systems with concurrent imaging data acquisition deliver valuable and accurate information, albeit at great cost, especially for small scale surveys. In addition, even

though in terrestrial environments, multi-view stereo reconstruction and texture mapping pipelines can be executed operationally, this is not the same for the marine environment and, especially, for the water bodies' bottom. In particular, despite the accurate and precise depth maps that can be derived from bathymetric LiDAR [Green 2000], the production of bottom orthoimages is still a challenging task due to refraction.

Hence, this dynamically changing zone is an area in need of further methodological development, since there are no robust and cost-effective solutions for seamless underwater and overwater mapping tasks.

1.1.2 Radiometric effects

The **radiometric effects** of water refraction include the chromatic aberration, which can be handled using high quality lens and the rippling caustics or sun flickering which are of really high importance [Agrafiotis 2018a, Menna 2018]. These complex physical phenomena are resulting from the light rays being refracted by a curved surface, such as the wavy interface between air and water. Caustics effects are apparent both in overwater and underwater imagery. In the first case, where large scale operations are mostly involved, they can be avoided by increasing the flying height, thus increasing the Ground Sampling Distance (GSD) (see Subsection 5.3.2) or by acquiring imagery with the sun angle less than 30 degrees over the horizon. However, in the second case, in most of which high detailed 3D reconstructions and textures are required, this is not possible, since the camera will exit the water, introducing also the severe geometric errors caused by the refraction on the water surface. Moreover, rippling caustics' generation precedes wavy water surface, a water state in which overwater mapping is not suggested since additional errors would be introduced due to the waves [Skarlatos 2019, Agrafiotis 2018b]. However, this is not the case for the underwater campaigns. Till now, to avoid these intense lighting artefacts on the bottom, image acquisition is performed under overcast conditions, or with the sun low on horizon [Agrafiotis 2018a, Menna 2018].

Hence, caustics seem to be the main factor degrading the underwater RGB image quality and affecting the image-based 3D reconstruction process in very shallow waters [Agrafiotis 2018a]. These effects are adversely affecting image matching algorithms by throwing off most of them, leading to less accurate matches [Agrafiotis 2018a] and causing issues in the Simultaneous Localization and Mapping (SLAM) based navigation of the Remotely Operated Vehicles (ROV) and Autonomous Underwater Vehicles (AUV) on shallow waters [Trabes 2017]. Also, they are the main cause for dissimilarities in the generated textures and orthoimages [Agrafiotis 2018a].

Thus far, rippling caustics effect did not attract enough attention from the experts. However, during the last years, where automated SfM and MVS techniques are widely used also from non-experts, especially for mapping submerged cultural heritage or benthic ecosystems in shallow waters, caustics are getting a lot of attention since in many cases they prevent the 3D reconstruction

[Agrafiotis 2018a, Forbes 2018]. At the same time, only few techniques have been proposed for the removal of caustics from images and video in the context of image enhancement. However, recent literature [Agrafiotis 2018a, Forbes 2018] indicates that the more successful the caustics removal is, the more valid matches are appearing in the stereo pairs and further processing with SfM-MVS techniques becomes possible for a number of applications.

1.2 Goal of the Thesis

As mentioned in the introduction, the general aim of this thesis is summarized in the following:

"Improve the efficiency, the accuracy and the applicability of the image-based methods in respect to overwater and underwater 3D mapping of the bottom in shallow waters by addressing geometric and radiometric issues caused by the refraction effect".

Regarding the overwater 3D reconstruction in shallow waters, the thesis will deal with the geometric distortions caused by water refraction on low altitude aerial imagery, posing the most significant challenges on depth determination in this field. Regarding the underwater 3D reconstruction in shallow waters, the thesis will deal with the radiometric effects caused by the water refraction, mainly the rippling caustics or sun flickering, which seem to be the main factor degrading image quality and affecting the 3D reconstruction process in very shallow waters.

Since the term "shallow waters" is quite general, regarding the aerial image-based bathymetry mapping, it should be understood as the depth for which the bottom is still visible in detail, presenting texture, random patterns, and adequate features to ensure salient point detection and matching. Depending on the water column characteristics, the water surface state, and the flight height this might vary. Experiments performed in this thesis for mapping the seabed in shallow waters, demonstrated that in the Eastern Mediterranean Sea and the Aegean Sea, the maximum working depth of the SfM-MVS pipelines varies from 10m to 15m. Nevertheless they produce very noisy point clouds in depths more than 10m, restricting the maximum depth for applying the developed methods for producing results of high accuracy to 10 m, a depth also reported in the literature [Legleiter 2019]. However, this is also related with the exploited SfM – MVS software for these experiments. Regarding the underwater image-based 3D reconstruction, the term "shallow" should be understood as the depth for which caustics are still visible on the seabed. Again here, this depends on the water column characteristics and the water surface state and might vary. Experiments performed demonstrated that in the Aegean Sea, the maximum depth on which this phenomenon is intense varies from 5m to 10m. This shallower zone of the coastal waters is of high importance since together with the growing economical and touristic activity, rich and threatened biodiversity and submerged

Cultural Heritage such as underwater ruins can be found there, stressing the need for precise 3D documentation and monitoring. Considering that Greece's coastline is estimated to 13.676 Km and that a roughly estimated average inclination of this coastal zone is of about 10%, it can be calculated that this shallower zone counts about 1367.6 sq. Km, equal to the total area covered by its internal waters i.e. lakes and rivers.

1.2.1 Objectives

The goal of this thesis can be divided into the following more specific objectives:

O.1. To correct the geometric errors caused by the refraction in overwater bathymetry mapping in clear shallow waters. Exploring and designing different methods to address the refraction effects that are apparent in the low-altitude aerial imagery in order to deliver more accurate and reliable bathymetric information. The interest of this topic is focused on exploring image-based refraction correction also exploiting the current advances in the machine learning techniques.

O.2. To improve the quality and the accuracy of the SfM results and of the rest of the derivatives of the image-based 3D reconstruction processes. Delivering methods that not only correct the geometric errors caused by the refraction in overwater bathymetry mapping in clear shallow waters, but they also improve the accuracy and the quality of the photogrammetric derivatives such as the point clouds, the orthoimages and the textures.

O.3. To democratize refraction correction methods: Deliver scalable and transferable methods. Main aim is to deliver scalable and transferable results, not dedicated to specific applications. This will make the developed methods useful to the scientific community and able to deal with different datasets, different areas and different acquisition systems and parameters.

O.4. To deal with the caustics effect in underwater shallow waters 3D mapping. Exploring and designing a comprehensive method to improve the underwater 3D reconstruction in shallow waters. Focus is given to the pixelwise correction of the underwater images affected by the radiometric effects of refraction with no restrictions on the seabed anaglyph.

O.5. To evaluate the developed methods on real world data. Evaluation of the proposed methods with real experiments using low altitude aerial imagery (UAV) and underwater imagery over different test areas in the Eastern Mediterranean Sea. The application of the developed methodologies is certainly a particularly important purpose of this thesis. For this reason, it was a priority to perform the experimentation with various real data as a premise to demonstrate the achieved research advances as well as their limitations.

1.2.2 Original Contributions of the Thesis

To address the objectives of the thesis, the main contributions to the research community are summarized in the following:

Development of two different methods for correcting the geometric effects of refraction in overwater bathymetry mapping in shallow waters:

The **first method** developed, is a novel method to address the systematic refraction errors on point clouds derived from SfM-MVS procedures in a massive and accurate way. The developed method, based on a linear Support Vector Regression model, can accurately recover shallow bathymetric information from low altitude aerial image datasets over a calm water surface, supporting several coastal engineering applications in non turbid waters and textured bottoms.

The **second method** developed, is an image correction method which is built upon the state-of-the-art and firstly exploits the machine learning procedures that recover depth on the derived image-based dense point clouds (also a contribution of this thesis) and then corrects the refraction effect on the original imaging dataset. This way the operational structure from motion (SfM) and Multi View Stereo (MVS) processing pipelines are ultimately executed on a refraction-free set of aerial datasets resulting into highly accurate bathymetric maps and image-based products, when calm water surface, non turbid waters and textured bottoms exist. This second method also achieves a reduction on the noise of the sparse point clouds which resulted from the SfM process and improves the accuracy and the quality of the produced orthoimages and textures. *These are related with O.1 and O.2.*

High bathymetric accuracy in the through-water imagery techniques.

For both developed methods for correcting the refraction effects, experimental results and validation over synthetic and real-world data demonstrated their high potential, both in terms of bathymetric accuracy as well as texture and orthophoto quality. *This is related with O.1. and O.5.*

Generalization of the developed refraction correction methods. For both developed methods for correcting the refraction effects, frameworks were designed in such a way to achieve high generalization over different low-altitude aerial platforms, cameras, flight heights, flight patterns and overlaps. Experimental results and validation over synthetic and real-world data, also applying a transfer learning approach, demonstrated the scalability and transferability of the methods that proved to be independent of specific applications. *This is related with O.3. and O.5.*

Development of a novel method for correcting the radiometric effects of refraction in close-range underwater 3D mapping in shallow waters:

A new method is proposed which contrarywise to the state-of-the-art can handle seabed of any anaglyph. This method exploits recent advantages in image classification to detect the areas of the image that are not affected by the phenomenon and based on the recovered geometry of the stereo pairs, corrects the rest of the pixels of the images. *This is related with O.4.*

Improve the 3D reconstruction when images with caustics are used.

Testing on real-world underwater imagery indicates that the proposed method for pixelwise image correction increases the completeness of the 3D point clouds, reaching the levels of the 3D point clouds being generated using images without caustics. *This is related with O.4.*

Creation of the first large-scale real-world underwater caustics benchmark dataset. The dataset contains 1465 underwater images. The dataset consists of 7 sub-datasets, including two-view, three-view and multiple view imagery depicting various types of seabed and caustics. Moreover, the corresponding reference results for all imagery are provided together with generated ground truth images for facilitating the training and testing of new machine learning and deep learning methods. The dataset counts 7345 images in total. The specific dataset provides an opportunity to evaluate, at least to some extent, the performance of different image classification approaches as well as the performance of image correction methods both in image and 3D reconstruction level. *This is related with O.4. and O.5.*

1.3 Outline of the Thesis

The contents of this thesis can be divided into three parts. The first part is an overview of the geometric and radiometric issues caused by the refraction effect in bathymetry mapping applications and describes the state-of-the-art for both cases (Chapter 2 and Chapter 3). The second part presents the two developed geometric correction methods and specifically the refraction correction methodologies for accurate bathymetry mapping using low altitude aerial imagery (Chapter 4). Experimental results over synthetic and real-world datasets endorse the different contributions (Chapter 5). The third part of the thesis presents the methodologies developed for improving the underwater 3D reconstruction by dealing with the radiometric issues caused by the refraction and, specifically, the caustics effect (Chapter 6). Experimental results over real-world datasets endorse the different contributions (Chapter 7). Finally, the last part of the thesis summarizes the conclusions and comments on eventual further work (Chapter 8). A brief description of each chapter is presented below.

Chapter 2: *Theoretical Background.* This chapter introduces the basic optical and acoustic properties of the water and analyzes the refraction's geometric effects in through water imaging, stereo imaging and multiple view imaging. Moreover, it reports on the caustics' generation, a radiometric effect of refraction.

Chapter 3: *Related Work.* This chapter is an overview of the most remarkable works already performed and which tackle the geometric and radiometric effects of refraction. A discussion on these works and on the open issues follows.

Chapter 4: *Correcting the Geometric Effects of Refraction.* This chapter proposes the two developed methods for refraction's geometric effects correction. For each one, the proposed method and the proposed refraction correction model is

described and justified.

Chapter 5: *Experimental Results and Validation of the Developed Geometric Corrections Methods.* Following Chapter 4, experimental results on synthetic and real world test datasets are presented and evaluated for both the presented methods as well as in comparison with other state-of-the-art methods. A summary and a discussion, including further work concludes the chapter.

Chapter 6: *Correcting the Radiometric Effects of Refraction.* This chapter, presents a new method proposed in this thesis regarding the pixelwise correction of the radiometric effects of refraction on the underwater imagery and specifically of the rippling caustics.

Chapter 7: *Experimental Results and Validation of the Developed Pixelwise Correction Method.* This chapter, initially presents the first real world benchmark dataset on underwater caustics. Subsequently, the experimental results of the proposed method for correcting the radiometric effects of the refraction on real world test datasets are presented and evaluated in terms of image classification, image correction and 3D reconstruction improvements.. A discussion concludes the chapter.

Chapter 8: *Conclusion.* This chapter concludes the thesis by summarizing the work and points out contributions and future work. It also comments on the research evolution and the publications accomplished during the compilation of the dissertation.

Theoretical background

Contents

2.1 Properties of water	11
2.1.1 Optical properties	12
2.1.2 Acoustic properties	12
2.2 Geometric Effects Caused by Specific Water Properties	13
2.2.1 Snell's law	13
2.2.2 Single view geometry in the through-water depth determination	14
2.2.3 Stereo view geometry in the through-water depth determination	15
2.2.4 Multiple-view geometry in the through-water depth determination	17
2.3 Radiometric Effects	18
2.3.1 Caustics effect	19
2.3.2 Rippling Caustics in key point detection and matching processes	20

2.1 Properties of water

Traditional surveying on dry land uses active or passive imaging techniques or devices almost entirely based on electromagnetic waves. Although on land these techniques have been successfully used for decades, the water's physical characteristics are such that 3D reconstruction and mapping of underwater environments becomes much more complex or even unfavorable in certain cases [Menna 2018]. Water is a medium inherently different from air and the first essential difference resides in the medium's density. Seawater is nearly 800 times denser than air and its density is not constant depending on depth, being a function of temperature, salinity and pressure. Water, especially from the sea, has a very high conductivity, with the well-known consequence of damaging electronic devices if they are not properly protected. Thus, the use of sensors underwater implies the use of special housings which must be waterproof and able to resist to high pressure to avoid failure. The high conductivity and permittivity of water has also the effect of attenuating electric waves to a great extent, making radio signals seldomly used underwater for communication and ranging measurements. Recording data in the water occurs nowadays mainly using optical or acoustic sensors [Menna 2018].

2.1.1 Optical properties

The refractive index of water is higher than air being about 1.33 times the one for freshwater and 1.34-1.35 the one for saltwater at 25°C. One of the most known optical phenomena concerning water is that any ray of light is deflected after hitting the water surface depending on the angle of incidence. This phenomenon is common to any surface separating media with different refractive indices and is known as Snell's law.

Light attenuation in water is ruled by scattering and absorption. Pure waters (either fresh or salty) are optically pure media totally exempt from any suspended particles; in pure water, light is absorbed only because of interaction of light with molecules and ions [Makowski 2016]. Long visible wavelengths, such as red light, are absorbed first, short visible wavelengths, like blue, last. For this reason, only 1% of the light entering the sea reaches 100m depth.

When suspended particles or sediments resulting from natural phenomena or human activities are present in water, a further light attenuation is present, and water is said to be turbid or cloudy. The more the particles, the higher the turbidity. Depending on the characteristics of these particles (such as size and color), light is absorbed rapidly and selectively (wavelength dependency), thus affecting the water color and visibility. Turbidity of water is generally quantified using the Secchi distance, an old and simple method introduced in 1865 [Cialdi 1865]; A white circular disk is immersed in water and the distance at which the disk is no more visible is defined as one Secchi distance.

2.1.2 Acoustic properties

While water absorbs electromagnetic waves very quickly, it transmits sound very well. Sound is a mechanical wave of pressure and displacement originating from continuous vibrations in a surrounding medium. Sound is used in SONAR (SOund NAvigation Ranging) systems, a technique well known in nature to animals like marine mammals for communication and navigation.

SONAR systems use sound propagation between a transmitter and a receiver and are used to remotely sense the interior of bodies of water, their floor and even the structures beneath the bottom. Indeed, when the sound wave encounters a surface interface between media with different physical characteristics, for example sea water and a sandy seabed, part of the energy is reflected, part is refracted and keeps traveling in the new media after a bending of the direction of propagation, according to Snell's law, and the remaining is scattered. The energy returned to the sonar device can provide information about the distance and the physical properties of the object.

Humans have known the principles of modern SONAR for hundreds of years. In 1490 Leonardo da Vinci, aware of the very good propagation of sound in water, had already suggested the use of an underwater tube to listen and discover approaching ships [Fahy 2003]. But it is only after the Titanic struck an iceberg that patents for

active systems able to detect objects and their distance were filed [Medwin 1997].

Sound is transmitted through gases, plasma, and liquids as longitudinal waves, also called compression waves. Sound waves are also reflected, like optical rays that are affected by Snell's law, refracted, scattered and attenuated by spreading and absorption in the medium through which they travel. Acoustic waves are spread as concentric spherical surfaces, whose intensity is inversely proportional to the square of the distance from the source.

2.2 Geometric Effects Caused by Specific Water Properties

As already described in the introduction, overwater imagery, also called through water imagery, is mostly affected by the geometric effects of refraction. However, the problem becomes even more complex when stereo-view or multi-view geometry is applied. In the following paragraphs, the Snell's Law (also known as Snell-Descartes Law or the Law of refraction) as well as the geometric effects of refraction in the through water depth determination for the single view, the stereo view and the multiple-view cases are discussed and analyzed.

2.2.1 Snell's law

The most well-known form of Snell's Law for a light ray traveling through different materials (Figure 2.1) is given by Equation 2.1. This formula is used to describe the relationship between the angles of incidence and refraction, when referring to light or other waves passing through a boundary between two different isotropic media, such as water, glass, or air. Snell's Law states that the ratio of the sines of the angles of incidence and refraction is equivalent to the ratio of phase velocities in the two media, or equivalent to the reciprocal of the ratio of the indices of refraction.

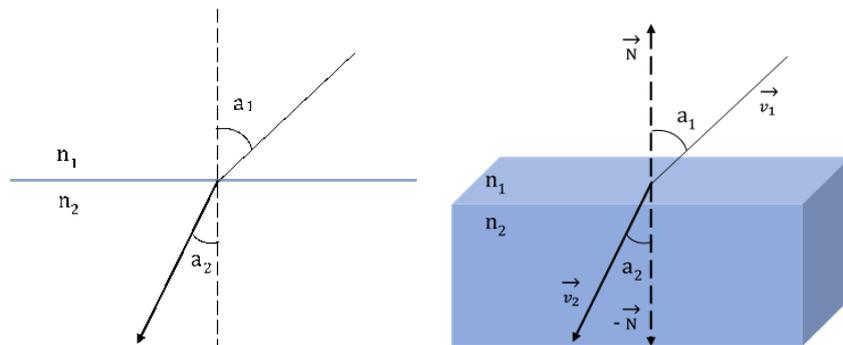


Figure 2.1: Snell's law in standard form (left) and three-dimensional vector format (right).

$$n_1 \sin(a_1) = n_2 \sin(a_2) \quad (2.1)$$

where, n_1 and n_2 are the refraction indices ($n_2 > n_1$), and a_1 and a_2 the incidence and refraction angles respectively (Figure 2.1 left).

In a generalized three-dimensional expression of the Snell's law (Figure 2.1 right), the border plane/surface is described by its normal \vec{N} , the incidence and refraction vectors are \vec{v}_1 and \vec{v}_2 respectively. In that case, the refraction vector can be calculated by Equation 2.2 [StarkEffects.com 2017].

$$\vec{v}_2 = \frac{n_1}{n_2} [\vec{N} \times (-\vec{N} \times \vec{v}_1)] - \vec{N} \sqrt{1 - \left(\frac{n_1}{n_2}\right)^2 (\vec{N} \times \vec{v}_1) \cdot (\vec{N} \times \vec{v}_1)} \quad (2.2)$$

The law follows from Fermat's principle of least time, which in turn follows from the propagation of light as waves.

2.2.2 Single view geometry in the through-water depth determination

The relationship of the object point, camera station and image point is shown in Figure 2.2, where the boundary of the two media is assumed to be a plane.

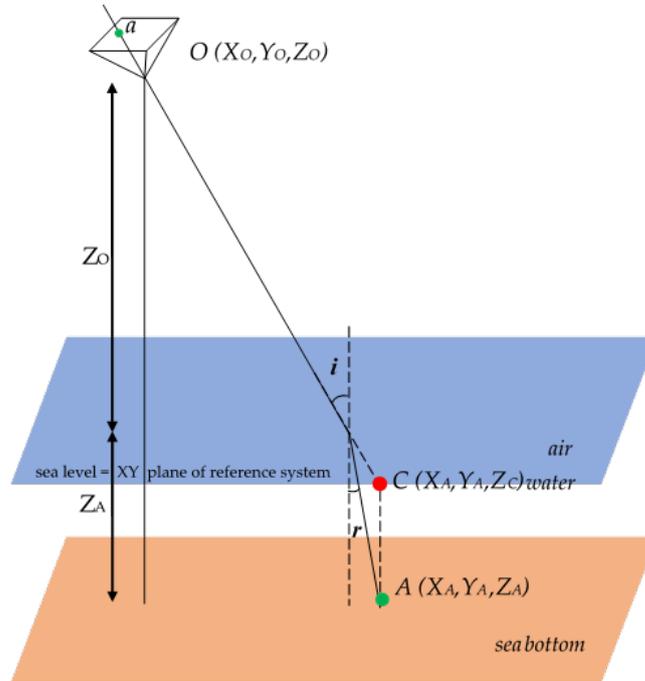


Figure 2.2: Geometry of the single view in the through-water depth determination (reproduced after [Shan 1994])

The XY plane of the photogrammetric world coordinate system X, Y, Z is defined as the boundary plane, hence $Z=\text{constant}$ for the water level and the Z axis is perpendicular to this plane with its positive direction upwards.

An object point $A(X_A, Y_A, Z_A)$ is imaged at point a on the image plane. The camera projection center is $O(X_O, Y_O, Z_O)$. The incident angle and refraction angle of the image line is i and r respectively [Shan 1994]. Because of refraction, the apparent position of the object point $A(X_A, Y_A, Z_A)$ is point $C(X_A, Y_A, Z_C)$ which has the same horizontal position with point $A(X_A, Y_A, Z_A)$ but seems to be at a shallower depth. As shown in Figure 2.2, points $C(X_A, Y_A, Z_C)$, $O(X_O, Y_O, Z_O)$ and point a on the image plane are collinear and satisfy the collinearity equation. Thus, without some form of correction, refraction acts to produce an image of the submerged object surface which appears to lie at a shallower depth than the real one, and it is worthy of attention that in each shot the collinearity condition for the correct position of the point A is violated [Fryer 1985].

2.2.3 Stereo view geometry in the through-water depth determination

The geometry of the two-media photogrammetry in the stereo case is presented in Figure 2.3 which is reproduced after [Fryer 1985]. Important is that for simplicity and as with other studies, the water surface is assumed planar. However, in reality the air-water surface will not be a plane, thus complicating the geometry of refraction through the surface [Fryer 1985].

To that direction, a scenario is considered, where the effective depth is being calculated by the standard collinearity. If a point has been matched successfully on the two images O_1 and O_2 , then the standard collinearity intersection would have returned the point $C(X_C, Y_C, Z_C)$, which is the apparent position of point $A(X_A, Y_A, Z_A)$. Starting from the apparent (erroneous) position $C(X_C, Y_C, Z_C)$ of the point $A(X_A, Y_A, Z_A)$ on the bottom, its image-coordinates a_1 and a_2 , can be backtracked in the images O_1 and O_1 , using the standard collinearity equation – the one used for forward calculation of the point’s coordinates, without taking refraction into consideration.

If the systematic error resulting from the refraction effect is ignored, the two apparent underwater rays A_1C and A_2C will not intersect exactly on the normal, passing from the underwater point $A(X_A, Y_A, Z_A)$, but approximately at $C(X_C, Y_C, Z_C)$, the apparent position of the point. In the special case where the point $A(X_A, Y_A, Z_A)$ is equidistant from the camera stations, the point $C(X_C, Y_C, Z_C)$ would be placed exactly on the normal [Fryer 1985], even if the refraction effect is ignored. However, even in this special case, Z_C would be the apparent (wrong) depth of the point.

According to the authors in [Wolff 2000], the simple solution for point $C(X_C, Y_C, Z_C)$ is not optimal in a statistical sense due to the presence of noise, however experiments demonstrate a very good approximation. Although in most of the cases the accuracy is sufficient, a maximum likelihood estimate could be performed based on this approximation to improve it.

Given that a_1 and a_2 are homologue points in the image plane, and they represent the same point in object space, the correct point position can be calculated explicitly, i.e. deterministically. Given that the water surface (the refraction surface) is the XY plane of the reference system, all points on it (such as A_1 and A_2) have $Z=0$ and their 3D space coordinates can be directly calculated using collinearity (Equation 2.3) which is linearly solved with only two unknowns [Skarlatos 2018].

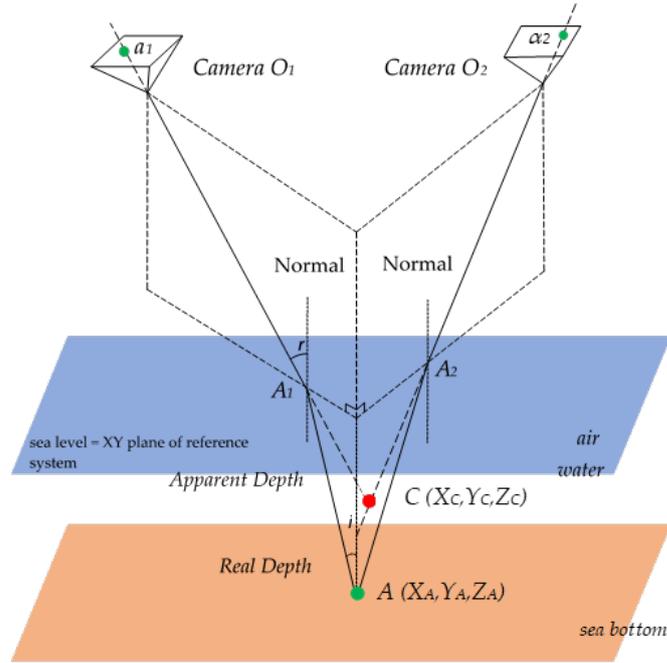


Figure 2.3: The geometry of two media photogrammetry in stereo case [Fryer 1985].

Having calculated the points A_1 and A_2 on the refraction surface:

$$\begin{bmatrix} x_1 \\ y_1 \\ -c \end{bmatrix} = \lambda_1 * R_{1\omega\varphi\kappa} * \begin{bmatrix} X_{O_1} - X_{A_1} \\ Y_{O_1} - Y_{A_1} \\ Z_{O_1} - Z_{A_1} \end{bmatrix}, \quad \text{with } Z_{A_1} = 0 \quad (2.3)$$

In this artificial example also described in [Skarlatos 2018], several assumptions have been posed, such as:

- The water surface is planar, i.e. without waves
- The water surface level is the reference ($Z=0$) plane of the coordinate system
- The image coordinates have already been corrected with respect to the principal point and lens distortion
- Cameras may have large rotation angles.

However, there is no assumption made of the camera axis being perpendicular to the planar water surface.

2.2.3.1 Vector intersection

Given that \vec{v}_1 (Figure 2.1) is the vector $\overrightarrow{O_1A_1}$ (Figure 2.3), and that the refraction index of air and water are known, one can calculate \vec{v}_2 . Point A_1 is already calculated in a previous step, hence point A in the bottom, is described by the equation $\vec{A} = \vec{A}_1 + t_1\vec{v}_2$, with t_1 unknown.

A similar equation can be formed from the second collinearity at point A_2 , $\vec{A} = \vec{A}_2 + t_2\vec{v}_2$, with \vec{v}_2 having been calculated from vector $\overrightarrow{O_2A_2}$ transformed by Snell's law. These two equations, in Cartesian format can be rewritten in Equation 2.4. The system (Equation 2.4) of six equations, has five unknowns (X_A, Y_A, Z_A, t_1, t_2) with one degree of freedom, similar to the intersection of two collinearity equations, and can be solved linearly using least squares. This two-step solution of the collinearity equation intersection with water refraction, is deterministic and direct, i.e. without iterations, and can be easily modified to address multi ray intersection, instead of a stereopair [Skarlatos 2018].

$$\begin{pmatrix} X_A \\ Y_A \\ Z_A \\ X_A \\ Y_A \\ Z_A \end{pmatrix} = \begin{pmatrix} X_{A_1} + t_1^*X_{v_2} \\ Y_{A_1} + t_1^*Y_{v_2} \\ 0 + t_1^*Z_{v_2} \\ X_{A_2} + t_2^*X_{v_2} \\ Y_{A_2} + t_2^*Y_{v_2} \\ 0 + t_2^*Z_{v_2} \end{pmatrix} \quad (2.4)$$

2.2.4 Multiple-view geometry in the through-water depth determination

In Figure 4, the multiple view geometry which applies to aerial imagery, is demonstrated: the erroneous, apparent position $C(X_C, Y_C, Z_C)$ is calculated by the collinearity equation [Agrafiotis 2019a]. Starting from this apparent position of a the point $A(X_A, Y_A, Z_A)$, its image-coordinates $a_1, a_2, a_3, \dots, a_n$, can be backtracked on images $O_1, O_2, O_3, \dots, O_n$ using the standard collinearity equation.

If a point has been matched successfully on the images $O_1, O_2, O_3, \dots, O_n$, then the intersection using standard collinearity would have returned point $C(X_C, Y_C, Z_C)$ which is the apparent and in a shallower position of point $A(X_A, Y_A, Z_A)$; and in the multiple view case is the adjusted position of all possible intersections for each stereopair, which are the red dots in Figure 2.4. When this adjustment fails to deliver reliable results i.e. in the case of adjusting points captured with very different incidence angles and thus affected differently by the refraction effect, the noise in the generated point clouds and consequently the estimated depths is increased (see also Subsection 5.3.5). Thus, also in the multiple view case, without some form of correction, refraction produces images and conse-

quently a point cloud of the submerged surface which appears to lie at a shallower depth than the real surface.

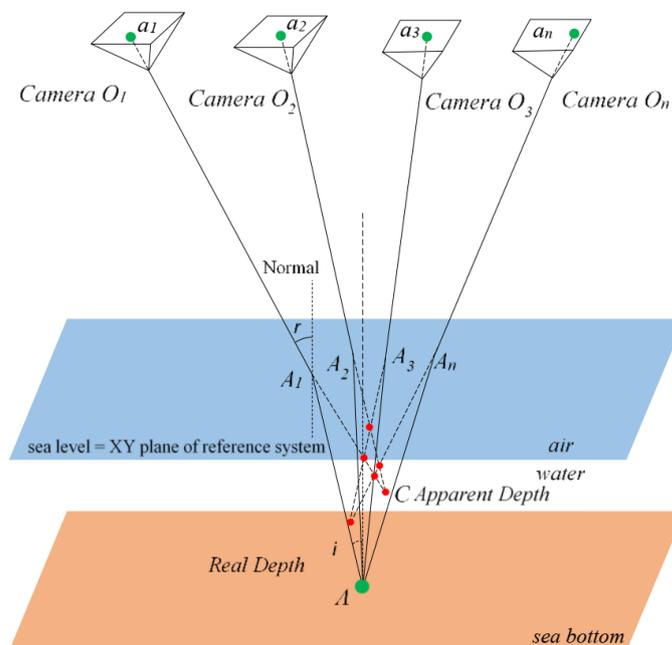


Figure 2.4: The geometry of two-media photogrammetry for the multiple view case, retrieved from [Agrafiotis 2019a, Agrafiotis 2019b].

2.3 Radiometric Effects

When underwater imagery is used for 3D mapping in shallow waters, and compared to overwater imagery, it is affected the most by the radiometric effects of refraction and especially by rippling caustics [Agrafiotis 2018a, Menna 2018], due to the smaller camera-to-object distance. In the following paragraphs, the nature of caustics and especially the rippling caustics effect is discussed and analyzed together with the related issues affecting the image-based underwater 3D reconstruction.

As already stated, underwater 3D reconstruction and mapping are based on various systems and methodologies but the most accurate of them are based on images as primary data providing also the important visual information [Menna 2018]. However, despite the relative low cost of such methods in relation to others, they suffer a major drawback; optical properties and illumination conditions of water severely affect image quality. Light is absorbed linearly to depth, resulting in a green-blue image due to strong absorption of the red wavelength. Therefore, the red channel histogram has less information in comparison to the green and blue ones. In addition, water absorbs light energy and scatters optical rays creating blurred images [Agrafiotis 2017, Agrafiotis 2018b, Mangeruga 2018]. Although the

above phenomena affect RGB imagery in every depth, when it comes to shallow waters, rippling caustics or sun flickering, seem to be the main factor degrading image quality [Agrafiotis 2018a].

More specifically, image-based surveys in shallow waters are affected by rippling caustics which generate bright fluctuations of a spatio-temporal scene radiance due to sun light refraction. Submerged objects on a strongly textured sea-bottom are illuminated by a natural random pattern, which is spatially and temporally varying, following the complex dynamics of the wavy water surface. These are referred as rippling caustics, sunlight flickers or caustic waves and characterized by fast moving patterns, which can significantly degrade the quality of underwater imagery on bottom mapping scenarios [Swirski 2009, Swirski 2010, Trabes 2015] (Figure 2.5).

2.3.1 Caustics effect

In optics, a caustic or caustic network is the envelope of light rays resulting by the projection of light rays being reflected or refracted by a curved surface [Lynch 2001], such as the wavy interface between air and water. The caustic is a curve or surface to which each of the light rays is tangent, defining a boundary of an envelope of rays as a curve of concentrated light [Weinstein 1969]. Therefore, the caustics can be the patches of light or their bright edges. These shapes often have cusp or spinode singularities [Wikipedia].

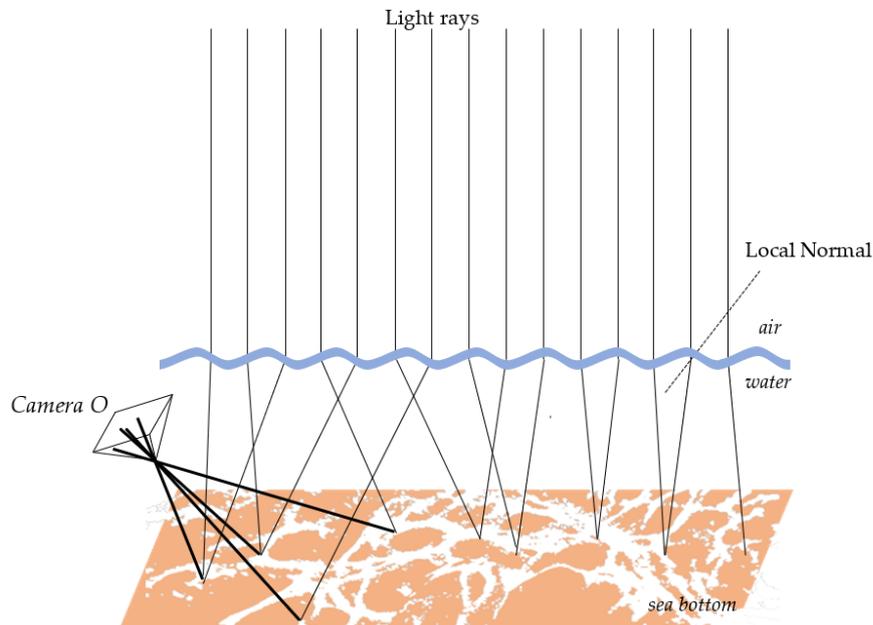


Figure 2.5: Due to waves on the water surface, the refraction of natural sky illumination is spatially varying. This creates 3D patterns of variable light flux and caustics underwater, and 2D illumination patterns on the illuminated objects. These patterns vary in time due to the dynamics of the surface waves.

Concentration of light, especially sunlight, can burn. The word caustic, in fact, comes from the Greek *καυστος*, burnt, via the Latin causticus, burning [Wikipedia]. A common situation where caustics are visible is when light shines on a drinking glass. The glass casts a shadow, but also produces a curved region of bright light [Wikipedia]. Rippling caustics are commonly formed when light shines through waves on a body of water [Wikipedia] (Figure 2.5).

Kinematically, wavy fringes can be clearly recognized from specks due to their relative smooth motion in contrast to the quick twinkling behavior of the scattering. So, sunlight rippling caustics have an optical flow field while specks do not. Also, morphologically, sunlight rippling caustics have characteristic strip and ring-like outlines (Figure 2.5) in contrast with the randomly dispersive location of the specks. On the contrary, morphological characteristics of the caustic waves, like the connectivity existing among many brilliant points (Figure 2.5 and Figure 2.5), is much more discriminating than the spectra of the brightness gradient. Connected bright pixels differentiate from the underlying background scene by the fact that the scene generally looks rather matte textured, but not dominantly brilliant. Furthermore, the brilliant points of the scattering are relatively small, well spread and generally not connected to each other [Trabes 2015].

2.3.2 Rippling Caustics in key point detection and matching processes

To make the effects of this phenomenon clear to the reader as well as to demonstrate the severe effects of rippling caustics in the underwater imagery in shallow waters, and consequently their effects in key point detection and matching processes, real world data are presented and processed below.

In Figure 2.6 three consecutive images with caustics of the same seabed area, captured from exactly the same camera position and orientation and with an interval of 5 seconds (from left to right) are presented (more details on the dataset used can be found in Chapter 5).

It is obvious that rippling caustics, being dynamic phenomena, cause a differentiation of the pixel values of the same area of the bottom in [Lowe 1999] key point detector are also mapped on these images. As can be seen in the zoomed areas of the images depicted in the second row of Figure 6, the detected key points on and around the areas affected by the phenomenon are totally different for the images acquired at time t , $t+5$ seconds and $t+10$ seconds.

Indeed, the boundaries of the rippling caustics on the seabed, appear to be a dominant area in the feature detection step, due to the dominant gradients between the bright and darker areas of the image. On the contrary, the interior area of rippling caustics, which is very bright or burnt, is inappropriate for feature detection since it is characterized by the absence of texture. Together with the variation of the phenomenon through time, these are exactly the reasons why these effects are adversely affecting image matching algorithms by throwing off most of them, leading to less accurate matches [Agrafiotis 2018a] and causing issues even in SLAM

navigation on shallow waters [Trabes 2017].

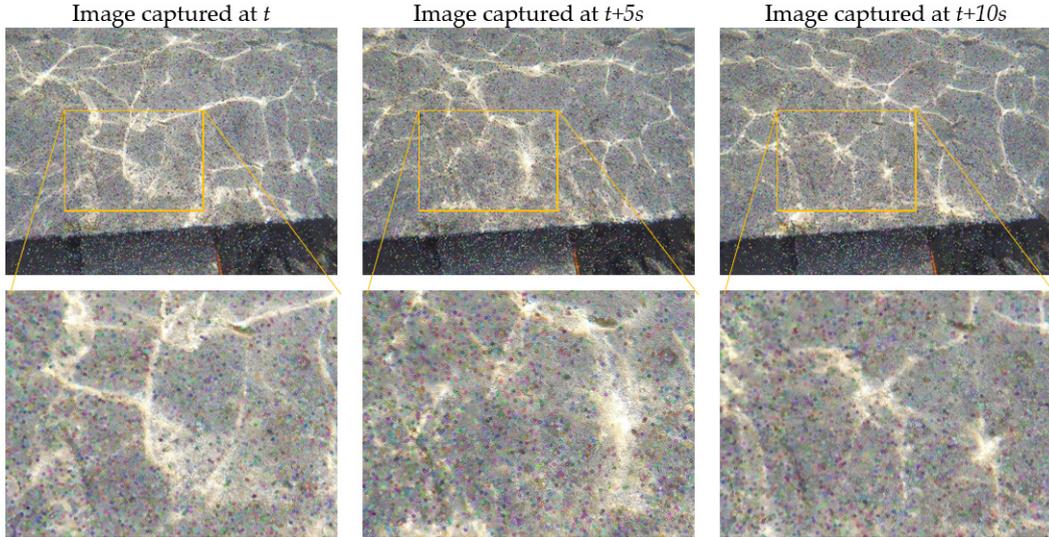
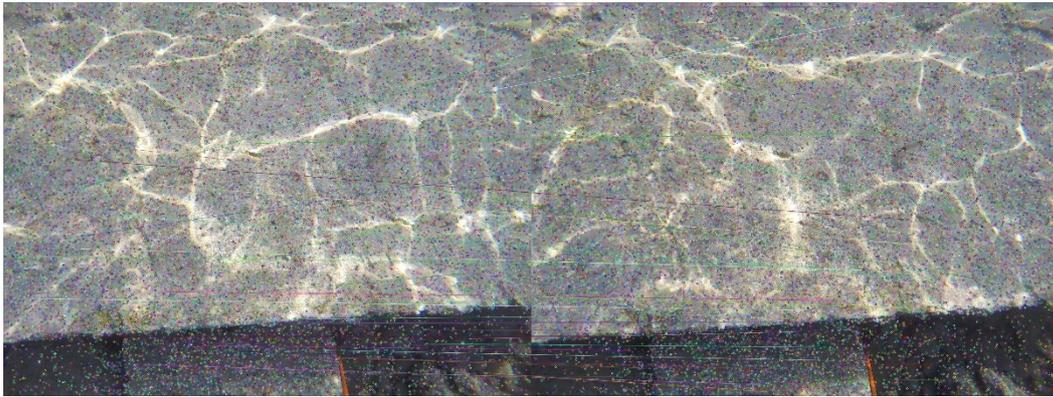


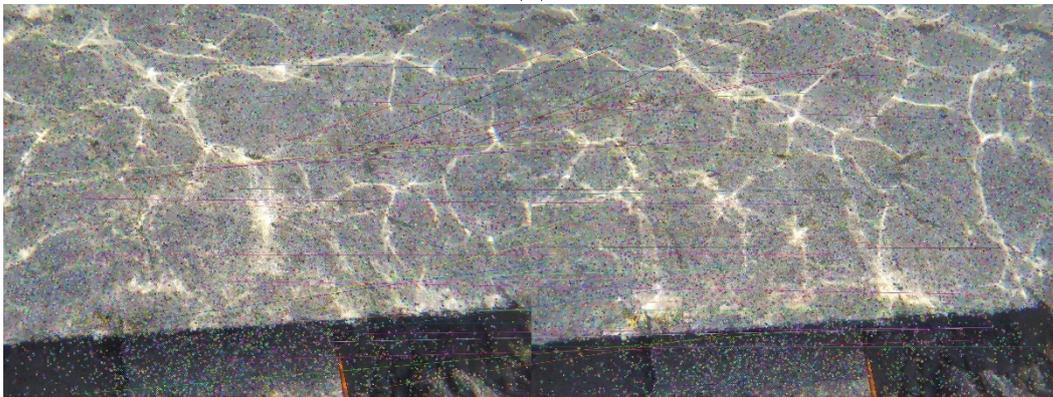
Figure 2.6: Images with caustics of the same seabed area from the same camera position, captured with an interval of 5 seconds (from left to right) and the detected key points using SIFT [Lowe 1999]. In the first row the full images are demonstrated while in the second row zoomed areas are depicted.

In order to prove the aforementioned negative effects of rippling caustics on the key point detection and matching processes, experiments were performed on the images presented in Figure 6 and their respective caustics free image, using SIFT [Lowe 1999] key point descriptor and Brute-Force matching. It is of really high importance to state here that the attempted matching process was performed on images having exactly the same camera orientation, thus depicting exactly the same area of the seabed in a different time.

The matching of the above three consecutive frames is attempted and compared with the matching results over the same caustics-free images, where the caustics-free image is matched with itself. In more detail, following the key point detection step using SIFT [Lowe 1999], the detected features were then matched to the corresponding features on the other images and the mapping of these features between these two images was stored in a vector. This matching was based on n-space Euclidean distance and performed both from left-to-right and right-to-left for redundancy. To filter these matches, the RANSAC algorithm was utilized to identify the inliers of the obtained point correspondences [Fischler 1981]. At the end of this step, a set of matched points is found in the given scenes, which are presented in Figure 2.7 with lines of different color. The matched points between the images captured at t and $t+5$ seconds are presented in Figure 2.7a while the matched points between the images captured at $t+5$ and $t+10$ seconds are presented in Figure 2.7b.



(a)



(b)



(c)

Figure 2.7: Matching results between the image captured at t and $t+5$ seconds (a), between the image captured at $t+5$ and $t+10$ seconds (b) and between the image without caustics and itself (c).

In Figure 2.7c the matches between the image itself, without rippling caustics, are presented. As also performed in [Agrafiotis 2018a], during the performed tests, it was decided not to evaluate the number of the total and valid matches only, but

also evaluate the geometry of the matches, since some valid results of the RANSAC filtering are still matching the wrong points.

Taking into account the above, one can observe in Figure 2.7a and Figure 2.7b that the image pairs have a lot of intersecting matches that are violating the epipolar geometry, a phenomenon that is eliminated in the image pair of respective caustics free image. Quantitative results suggested that in the first two matching examples, the matched points between the first two images were 47, between the second and the third were 43 while between the same caustic free image 1044.

The above results, suggest that the rippling caustics effect, indeed affects key point detection and the key point matching process in the underwater imagery in shallow waters, thus affecting the image-based 3D reconstruction in these areas.

Related Work

Contents

3.1 Compensation for Refraction's Geometric Effects	25
3.1.1 Analytical and Image-based Refraction Correction	26
3.1.2 Image-based Bathymetry Estimation using Machine Learning and simple Regression Models	29
3.1.3 Discussion	30
3.2 Radiometric Effects	31
3.2.1 Caustics Detection and Removal	31
3.2.2 Discussion	32

3.1 Compensation for Refraction's Geometric Effects

Refraction effect has driven scholars to suggest several correction models for two-media photogrammetry, most of which are dedicated to specific applications. Two-media photogrammetry is divided into through-water and in-water photogrammetry. The "through-water" term is used when the camera is above the water surface and the object is underwater, hence part of the ray is traveling through air and part of it through water. It is most commonly used in aerial photogrammetry [Skarlatos 2018, Mulsow 2010, Mandlbürger 2018, Mandlbürger 2019, Dietrich 2017], satellite photogrammetry [Cao 2020, Cao 2019] or in close range applications [Georgopoulos 2012, Butler 2002]. It was even argued that if the water depth to flight height ratio is considerably low, then compensation of water refraction is unnecessary [Karara 1979]. However, as shown by Skarlatos and Agrafiotis in [Skarlatos 2018], the water depth to flying height ratio is irrelevant in cases of aerial imaging, and water refraction correction is necessary. The "in-water" photogrammetry term is used when both camera and object are in the water.

Two-media photogrammetric techniques have been reported already, in the early 1960s [Tewinkel 1963, Shmutter 1967], and the basic optical principles proved that photogrammetric processing through water is possible. In the literature, two main approaches to correct refraction in through-water photogrammetry can be found: analytical or image-based. The first is based on the modification of the collinearity equation [Wang 1990, Butler 2002, Shan 1994, Fryer 1983], while the latter is based on re-projecting the original photo to correct the water refraction [Georgopoulos 2012, Skarlatos 2018, Agrafiotis 2015].

3.1.1 Analytical and Image-based Refraction Correction

According to the author in [Maas 2015], there are two ways to analytically correct the effects of refraction in two-media photogrammetry: the deterministic (explicit) and unsettled (ambiguous and iterative). The iterative method is employed to resolve refraction, either in object space or in image space. In the latter case, the whole multimedia problem is simplified to the radial shift computation module [Agrafiotis 2015], which is also one of the principles followed in this study (see Section 4.2).

The two-media analytically based methods have been used for underwater mapping using special platforms for hoisting the camera [Fryer 1985, Whittlesey 1975, Westaway 2001]. For mapping an underwater area, authors in [Elfick 1984] used a "floating pyramid" for lifting two cameras. The base of the pyramid was made of Plexiglas for avoiding wave effects and sun glint, which according to [Partama 2018], limits the performance of two-media approaches. A contemporary, but similar approach is adopted by [Butler 2002] for mapping the bottom of a river. There, the extracted digital surface model (DSM) points have been corrected from refraction effects, using a dedicated algorithm which corrects the final XYZ of the points. A Plexiglas surface was again used for hosting the control points. In [Ferreira 2006], for modelling a scale-model rubble-mound breakwater, refraction effects were described via a linearized Taylor series, which is valid for paraxial rays. That way, a virtual principal point was defined, where all rays were converging. It has been shown that this approach offers satisfactory compensation for the two media involved.

Author in [Mulrow 2010] developed a general but sophisticated analytical model, as part of the bundle adjustment solution. The model allows for simultaneous determination of camera orientation, 3D object point coordinates, an arbitrary number of media (refraction) surfaces, and the determination of the corresponding refractive index. The author supported the proposed model with experimental data and a corresponding solution. Another approach [Wolff 2000], introduced a new 3D feature-based surface reconstruction made by using the trifocal tensor. The method was extended to multimedia photogrammetry and applied to an experiment on the photogrammetric reconstruction of fluvial sediment surfaces. In [Ke 2008], authors were modelling the refraction effect to accurately measure deformations on submerged objects. They used an analytical formulation for refraction at an interface, out of which a non-linear solution approach was developed to perform stereo calibration. Aerial images have also been used for mapping the sea bottom with an accuracy of approximately 10% of the water depth in cases of clear waters and shallow depths, i.e., 5-10m [Byrne 1977, Harris 1972, Masry 1975].

Authors in [Woodget 2015] presented a relatively simple approach for refraction correction of bathymetric elevation models using a small angle approximation. There, Snell's Law was applied to the riverbed surface to increase the apparent water depths by a multiplier of 1.34 (i.e., the refractive index of clear water) to account for the effects of refraction. The refraction correction is calculated by assuming

that the incidence and refraction angles are less than 10° . Results present errors of 0.05–0.08 m, while the average water depth ranged between 0.14 m and 0.18 m, and maximum water depth between 0.50 m and 0.70 m. In the same context, authors in [Chirayath 2019] suggest that for a flat fluid sea surface and nadir camera viewpoint, the apparent depth is typically three-fourths the actual depth on Earth. An investigation on the drawbacks of this rational is performed in Subsection 5.2.2.

In [Dietrich 2017], a more sophisticated multi-angle refraction correction approach was proposed. This author used the 3D dense point clouds and applied an iterative refraction correction method based on the positions and view angles of all the cameras used within the photogrammetric reconstruction of the position of each point in the cloud. However, for correcting each submerged point in this approach, the erroneous camera positions calculated using the refracted imagery are exploited. The methods presented in [Woodget 2015] and [Dietrich 2017] are evaluated in comparison to the methods presented in this thesis in Section 5.5.

In [Maas 2015], the authors adopt a simplified model of in-water and through-water photogrammetry by making assumptions about the camera being perpendicular to a planar interface. In [Telem 2010], the authors avoided the full strict model of two-media photogrammetry if the effects could be considered as radial symmetric displacements on the camera plane—effects which could be absorbed from camera constant and radial lens distortion parameters. With the same concept, authors in [Lavest 2000] state that the effective focal length in underwater photogrammetry is approximately equal to the focal length in air, multiplied by the refractive index of water. This model was extended by [Agrafiotis 2015], by also considering the dependency on the percentages of air and water within the total camera-to-object distance.

However, these oversimplified models can only be valid for two-media photogrammetry, where the camera is strictly viewing perpendicularly onto a planar object. Obviously, this is not the general case, because the bottom and the water surfaces are seldomly flat and planar, while the images can be perfectly perpendicular only with sophisticated and expensive camera suspension mounts. When UAV (drone) photographs are being used to create a 3D model of the bottom and create the corresponding orthoimage object [Shan 1994], these assumptions are not valid. For these reasons, the method proposed by Agrafiotis and Georgopoulos in [Agrafiotis 2015] was extended by Skarlatos and Agrafiotis in [Skarlatos 2018], where the same principle was differentially applied over the entire area of the photograph. Authors in [Skarlatos 2018] proposed an iterative simplified algorithm for water refraction correction within the existing photogrammetric pipeline. In this proposed correction model, a provisional DSM was used in order to correct the refraction on the image iteratively. After computing a provisional but erroneous DSM, images were corrected from the refraction effect, and a new DSM was calculated using the temporarily corrected imagery. The solution converged after three to four iterations. Results over two test sites suggested that the mean distance between the LiDAR true depth data and the resulting dense point clouds was between 0.20m and 0.50m, depending on the depth. The approach presented in [Skarlatos 2018] followed after the method

presented by Georgopoulos and Agrafiotis in [Georgopoulos 2012]. Authors there corrected the refraction effect on the image by reprocessing the radial distance of each pixel, according to a depth map derived for each image, of which the grey values were assigned according to the real depth, which was measured using traditional surveying techniques. Then, a new image was gradually constructed using the determined radial distances.

In [Murase 2008] authors developed a refraction correction model for airborne imagery by examining the horizontal differences between the observed and true positions when objects were aligned along an airplane track or when the incident angles were identical. The procedure was applied in real world and the comparison of the corrected depths with measured depths at 658 points showed a mean error and standard deviation of 0.06m and 0.36m, respectively, for a measured depth range of 0.20m to 3.4m. The same approach was also applied in [Hodúl 2018] for refraction correction on bathymetry derived from a WorldView-2 stereopair. Results suggested a mean error of 0.03m and an *RMSE* of 1.18m. Authors in [Cao 2019] and [Cao 2020] present a two-media photogrammetry technique for WorldView-2 stereo imagery. Results over tests performed using high-resolution WV2 multispectral stereo-pairs illustrated that depth estimates can be derived with a vertical accuracy of about 1.14m to 1.91m in water depths from 5 to 20m.

In [Qian 2018], authors presented a first approach for simultaneously recovering the 3D shape of both the wavy water surface and the moving underwater scene and tested over both synthetic and real data. There, after acquiring the correspondences across different views, the unknown water surface and underwater scene were estimated through minimizing an objective function under a normal consistency constraint. In [Kasvi 2019], authors attempted to address the accurate mapping of shallow water bathymetry by comparing three approaches: one based on echo sounding, and two based on photogrammetry—bathymetric SfM and optical modelling, also referred to as the spectral depth approach or optical-empirical modelling. The bathymetric SfM exploited here is the one described in [Dietrich 2017]. The results from the photogrammetric methods suggest that the quality of the bathymetric SfM was highly sensitive to flow turbidity and color, and therefore, depth. However, it suffers less from substrate variability, turbulent flow, or large stones and cobbles on the riverbed than optical modelling. Color and depth did affect optical model performance, but clearly less than the bathymetric SfM. Finally, in a recent work presented in [Mandlbürger 2018] and extended in [Mandlbürger 2019] an interesting investigation on the potential of through-water DIM (dense image matching) for the high-resolution mapping of generally low textured shallow water areas is presented. For correcting the refraction effect, after extracting a water surface model from bathymetric LiDAR data, the approach presented in [Wimmer 2016] is adopted. In the case study presented there, the DIM-derived underwater surfaces of coastal and inland water bodies are compared to concurrently acquired bathymetric LiDAR data. Results present deviations in the decimeter-range over depths of more than 5m compared to the laser data as reference.

3.1.2 Image-based Bathymetry Estimation using Machine Learning and simple Regression Models

There is also a number of single image approaches for bathymetry retrieval using satellite imagery that are reported here for completeness reasons. Most of those methods are based on the relationship between the reflectance and the depth. Some of them exploit an SVM framework to predict the correct depth [Wang 2019, Misra 2018]; experiments therein showed that the localized model reduced the bathymetry estimation error by 60% from a Root Mean Square Error (RMSE) of 1.23m to 0.48m. In [Mohamed 2016] a methodology is introduced using an ensemble learning (EL) fitting algorithm of least squares boosting (LSB) for bathymetric map calculations in shallow lakes from high resolution satellite images and water depth measurement samples using an echo sounder. The bathymetric information retrieved from the three methods [Wang 2019, Misra 2018, Mohamed 2016] was evaluated using echo sounder data. The LSB fitting ensemble resulted in an *RMSE* of 0.15m where the Principal Component Analysis (PCA) and Generalized Linear Model (GLM) yielded *RMSEs* of 0.19m and 0.18m respectively, over shallow water depths less than 2 m. In [Traganos 2018], among other pre-processing steps, authors implemented and compared four different empirical SDB (satellite-derived bathymetry) approaches to derive bathymetry from pre-processed Google Earth Engine Sentinel-2 composites. Empirical SDB methods require certain bands in the visible wavelength—with blue and green being the most widely used—and a set of known in situ depths as the only inputs in simple or multiple linear regressions, which leads to bathymetry estimations in a given area. Accuracies of the calibrated model in the two validation sites reached an *RMSE* of 1.67 m. In [Niroumand-Jadidi 2019], a multiple regression bathymetry model was employed for substrate mapping in shallow fluvial systems having depth <1 m. To do so the authors analyzed spectroscopic measurements in a hydraulic laboratory setting, simulated water-leaving radiances under various optical scenarios and evaluated the potential to map bottom composition from a WorldView-3 image. In [Shintani 2017], authors compare the potential of through-water photogrammetry and spectral depth approaches to extract water depth for environmental applications. Imagery and cross sections were collected over a 140m reach of a river of a maximum depth of 1.2m approximately, using a UAV and real time kinematic (RTK)-GPS. There, for the site-specific refraction correction method, sparse calibration data resulting from two-thirds of the cross-section data were used to calibrate a simple linear regression equation between the predicted and measured water depths, and one-third of the data were used for validation and testing. This method achieved a mean error of 0.17–0.18m at a maximum depth of 1.2 m. In [Caballero 2019], authors examine the relatively high-resolution Multi-spectral Instrument (MSI) onboard Sentinel-2A and the Moderate-Resolution Ocean and Land Color Instrument onboard Sentinel-3A for generating bathymetric maps through a conventional ratio transform model in environments with some turbidity. The MSI retrieved bathymetry at 10m spatial resolution with errors of 0.58m, at depths ranging between 0 and 18m (limit of LiDAR survey) at the first test site, and

of 0.22m at depths ranging between 0 and 5m in the second test site, in conditions of low turbidity. In [Legleiter 2019], authors focused on defining the limits of spectrally based mapping in a large river. They used multibeam echosounder (MBES) surveys and hyperspectral images from a deep, clear-flowing channel to develop techniques for inferring the maximum detectable depth directly from an image and identifying optically deep areas that exceed the maximum depth. There, results suggest a limit of 9.5m depth. In [Kim 2019], authors evaluated the capability of a geographically weighted regression model to retrieve bathymetry of a shallow stream, of which water depth is less than about 1m from simple RGB imagery.

Moreover, authors in [Entwistle 2019] exploited simple regression models correlating theodolite and SfM-MVS points for site specific refraction correction for depths up to 2.40 m. Comparative analysis performed suggested that the submerged SfM data, captured bed levels generally to within ± 0.25 m with only a weak relationship recorded between error and flow depth. Authors in [Ma 2019] further explore the bathymetric extraction techniques using overlapping in shallow water areas through two-medium ray refraction and multispectral information inversion. This integrated approach combines the bathymetry obtained from the refraction method and the multispectral inversion technique to represent the integrated bathymetry of the shallow water area. Evaluation performed on sparse control data indicated *RMSEs* in the range of 0.35m to 0.90 m. Finally, in sandy texture-less seabed and riverbed areas, alternative methods based on the waves height and velocity or the river stream surface velocity are exploited [Thuan 2019, Collins 2020, Holman 2013, Legleiter 2020].

3.1.3 Discussion

This section has introduced the most representative works carried out so far regarding the correction of the geometric effects of refraction towards bathymetry estimation. Even bottom mapping has been a subject of great interest for years with plenty of published works in many different topics such as acoustic mapping (SONAR), active optical mapping (bathymetric LiDAR), and underwater close range photogrammetric mapping, refraction correction on image-based bathymetry mapping is still in its initial phase and a relatively reduced number of approaches have been reported so far. However, recent advances in photogrammetry and image-based 3D reconstruction, exploiting SfM and MVS techniques, brought to the forefront the problem again, motivating scholars to deal with refraction on a promising different basis.

Till now, most of the published methods are dedicated to specific applications and specific data, preventing their generalization over different cases. Moreover, most of the works found in the literature are dealing with only one test site, for which they develop a dedicated refraction correction approach. Even though some of the above methodologies and models produce quite accurate results, most of them are dependent on several parameters, such as the interior and the exterior orientation of the cameras and the water surface, carrying also alignment errors and

noise introduced by the refraction, in their final results. To that direction, most of the presented methods lack in cross validation with other test sites, using different image acquisition systems and parameters. A cause for this might be the lack of publicly available imagery data with ground truth, due to the great difficulty in their acquisition.

Regarding the approaches using satellite imagery, these are not based in SfM-MVS processing but in empirical SDB or model-based bathymetry, in order to predict the depth according to the spectral value. However, in most of the studies presented in these articles, with the exception of [Traganos 2018], test and evaluation steps are implemented on percentages of the same test site and at very shallow depths, leading to site-specific approaches. In most cases, the 80% of the available data is used for training the model while the remaining 20% is used for evaluation.

On the contrary, main aims of the proposed methods in this thesis are the creation of synthetic datasets and the use of various different real world test sites for testing and evaluating the proposed approaches and of course the proposal of generic and data independent solutions to the refraction problem that affects the aerial image-based bathymetry mapping.

3.2 Radiometric Effects

3.2.1 Caustics Detection and Removal

For many years, the computer graphics research community has focused on the generation of caustics and as a result many techniques have been proposed which generate photorealistic results. At the same time only a few techniques have been proposed for the removal of caustics from images and video in the context of image enhancement. A brief overview of the most relevant work to caustics removal are provided next.

Trabes et al., in [Trabes 2015] propose a technique which involves tuning a filter for sunlight-deflickering of dynamically changing underwater scenes. They employ a continuous parameter optimization inside a basic filter, which provides feedback for further improving the performance of the filter. Being an optimization, the filter's performance is highly sensitive to sub-optimal parameters and in particular, the segmentation parameter which is part of the objective function in the optimization. A different approach was proposed in [Gracias 2008] where a mathematical solution was presented involving the calculation of the temporal median between images within a sequence. A strong assumption of this work, is the fact that feature matching (Harris corner detection variant in [Gracias 2000]) is employed for the formation of the sequence which makes this approach very susceptible to the light variations in the images and in particular caustics effects.

The same authors later extend their work in [Shihavuddin 2012] and propose an online sunflicker removal method which treats caustics as a dynamic texture. As reported in the paper this only works if the seabed or bottom surface is flat. Similar approaches have also been proposed for general cases of dehazing and descattering

of images such as [Joshi 2010, He 2010, Fattal 2008].

In [Schechner 2004] the authors propose a method based on processing a number of consecutive frames. These frames are analyzed by a non-linear algorithm which preserves consistent image components while filtering out fluctuations. Their proposed method however does not take into account the camera motion which almost always leads to registration inaccuracies.

Finally, Forbes et al., in [Forbes 2018] proposed a solution based on two small and easily trainable CNNs (Convolutional Neural Networks). To detect caustics, a small CNN was trained over synthetic data in order to overcome the obstacle of not having ground truth data available for real world underwater caustics. Using sample underwater video images with caustics they created a set of synthetic data using 3D objects for underwater seabed, multiple lighting and global illumination for rendering caustic effects with the virtual camera located below the water surface as in real world shallow water imaging. In addition to the rendered caustics images, they rendered the masks containing confidence values of a caustic occurring at a pixel and also corresponding caustic-free images. This proposed solution was evaluated in terms of keypoint detection, image matching and 3D reconstruction performance in [Agrafiotis 2018a].

3.2.2 Discussion

This section has introduced the most representative works carried out so far towards the correction of rippling caustics effects. Despite the innovative and complex aforementioned techniques, addressing caustic removal with procedural methods requires that strong assumptions are made on the many varying parameters involved e.g. scene rigidity, camera motion, etc. Moreover, real ground truth for caustics is not easily available, preventing recent advances in machine learning to jump in. To deal with these issues, in this thesis a dataset containing 1465 underwater images is presented. In this dataset, the corresponding reference results for all the imagery are provided together with generated ground truth images for facilitating the training and testing of new machine learning and deep learning methods.

Correcting the Geometric Effects of Refraction

Contents

4.1 Correcting the Geometric Effects of Refraction in the 3D Space	34
4.1.1 Proposed method	34
4.1.2 The proposed refraction correction model	35
4.2 Correcting the Geometric Effects of Refraction in the 2D (Image) Space	41
4.2.1 Proposed Method	41
4.2.2 Proposed Refraction Correction Model	42
4.2.3 Summary	47

In order to deal with the refraction effect in the through-water imagery and address the aforementioned issues, two different but complementary correction methods are proposed in this thesis. Their main aim is to correct the geometric effects of refraction on the depth values of 3D point clouds produced by the SfM-MVS pipelines and then correct the original imagery. Worth noting though that the point cloud correction method can be also used as a standalone solution.

To address the systematic refraction errors on the depth values of 3D point clouds derived from SfM-MVS procedures in a generic, reliable and accurate way, a new method is developed. This method is published in [Agrafiotis 2019b] and [Agrafiotis 2019a] and is based on machine learning tools which can accurately recover shallow bathymetric information from aerial image datasets (UAV imagery in the cases studied in this thesis) over a calm water surface, supporting several coastal engineering and mapping applications. In particular, the main objective of this method is to deliver image-based 3D point clouds with accurate depth information by training suitable machine learning models. The rationale behind this objective is that these models would "learn" the systematic underestimation of the depths by correlating the apparent and the real depth of hundreds of thousands of points and would be then capable of estimating the correct depth, when fed with only the apparent one. To this end, a linear Support Vector Regression (SVR) model was employed and trained to predict the actual depth from the apparent depth of a point in a point cloud produced by image-based techniques.

Since, 3D point cloud data are not enough for many applications, a second method correcting the geometric effects of refraction on the image space was also developed and published in [Agrafiotis 2020]. Main objective of this method is to overcome the refraction effect directly in the primary data. The rationale behind this method is that having corrected the primary data, the execution of SfM-MVS processing pipelines on a refraction-free set of aerial images resulting into highly accurate bathymetric maps is facilitated. This method exploits SfM results and builds upon and exceeds the state-of-the-art approach described in [Skarlatos 2018] by integrating the machine learning model proposed in the first method. In the following Sections 4.1 and 4.2 the proposed correction models of the two methods are presented in detail.

4.1 Correcting the Geometric Effects of Refraction in the 3D Space

In this section, the developed method for addressing the systematic refraction errors in the depth values of the 3D point clouds derived from SfM-MVS procedures in a generic, reliable and accurate way is described. In the following text of the thesis, this method will be referred as Method 1. The developed technique is based on machine learning tools which can accurately recover shallow bathymetric information from aerial image datasets (UAV imagery in the later examined cases) over a calm water surface, clear waters and seabed with texture. The proposed method can also be used as a stand-alone solution for delivering refraction effect free image-based 3D point clouds in cases where imaging products are not necessary.

4.1.1 Proposed method

The main objective of the Method 1 presented in this section and published in [Agrafiotis 2019b] and [Agrafiotis 2019a] is to deliver image-based 3D point clouds with accurate depth information by training models which can estimate the correct depth of a set of points having knowledge of only their apparent (erroneous) depth. Additionally, the independence of this method from the UAV imaging system, the camera type, the flying height, the camera to water surface distance, and the base-to-height (B/H) ratio is also considered of great importance. This would attribute to the developed method a very important advantage regarding the current state-of-the-art methods. To this end, the employment of a machine learning model to be trained to correlate the apparent depth values of hundreds of thousands of points with their respective real depths is proposed. Having trained a model with great generalization potential over different areas and in shallow waters, the availability of a ground truth bathymetric data for a particular site will not be a prerequisite. The overall workflow of the proposed method is illustrated in Figure 4.1.

Following the aerial image data collection and Ground Control Point (GCPs) measurements, an initial SfM-MVS is executed ignoring the refraction effects on

4.1. Correcting the Geometric Effects of Refraction in the 3D Space 35

the images. Following the generation of the initial dense point cloud of the area, the bottom (seabed, lakebed or riverbed) points are extracted. To perform this extraction, points having an elevation equal or less than the water surface level (measured at the time of the flight with an RTK GPS) are considered as bottom points while the rest as dry land points. Next, having selected the correction model, the initial dense (or sparse) point cloud with systematic depth underestimation and the true depth data are correlated in order to train the model.

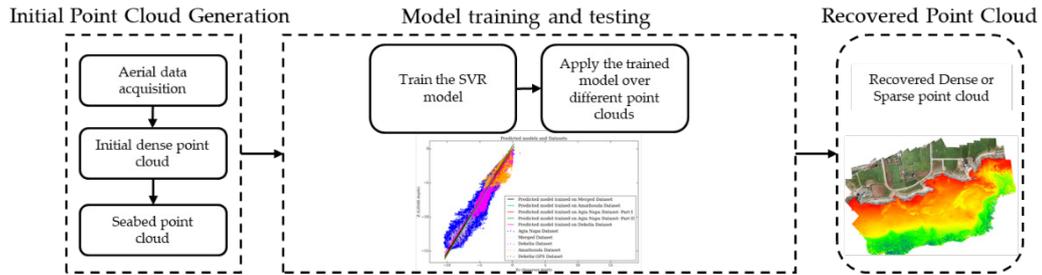


Figure 4.1: The proposed method for correcting the refraction in the point clouds based on a Linear Support Vector Regression (SVR) model

The steps of machine learning model selection as well as the training of this model are performed only during the initial phase, since after training, the model will be able to correct the depths in any image-based point cloud of the bottom regardless of the availability of true bathymetric data. Details for the important step of model selection and formation are elaborated next.

4.1.2 The proposed refraction correction model

In Figure 4.2, the XY plane of the world coordinate system X, Y, Z is defined as the boundary plane and the Z axis is perpendicular to this plane with its positive direction upwards. Without the presence of the water, point $A(X_A, Y_A, Z_A)$ on the bottom would be calculated by SfM-MVS in the same correct position in the 3D point cloud, being affected only by the interior and exterior orientation errors of the cameras. However, when water is present, due to the refraction effect, point $A(X_A, Y_A, Z_A)$ is calculated in the apparent position $A'(X_A, Y_A, Z'_A)$, lying always in a shallower depth than the real point $A(X_A, Y_A, Z_A)$. The same applies for all the points of a 3D point cloud which are situated under the surface of the water (See also Subsection 2.2.4).

Consequently, the main aim of the refraction correction method presented here is to compensate for the systematic underestimation of depths in the 3D point clouds of the bottom by learning the correlation of the apparent depth of a 3D point Z_0 with its real depth Z (Figure 4.2).

State-of-the-art approaches are using non-sophisticated regression methods with maximum some few hundreds of sparsely collected points, measured by traditional

surveying techniques and having maximum depths of no more than 2m to 3m. On the contrary, in this thesis, in order to achieve a transferable and generic solution, the correlation process would need hundreds of thousands of points. These points should include also deeper areas, in order to deliver a model that fits every low altitude aerial campaign.

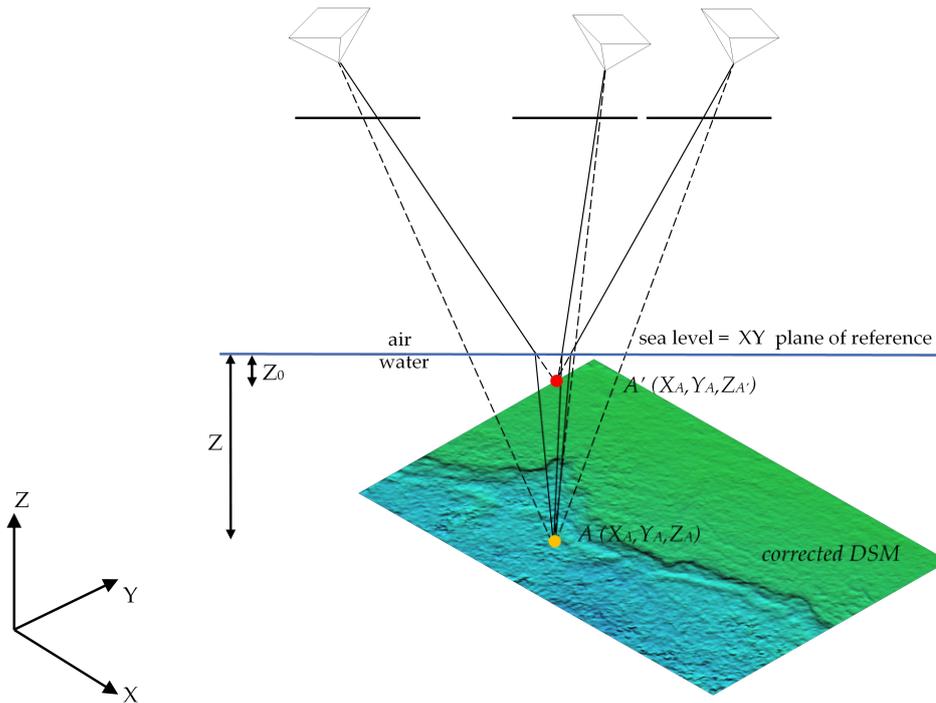


Figure 4.2: The geometry of the apparent point cloud generation using SfM-MVS techniques.

4.1.2.1 The linear SVR model

To achieve this correlation, a linear SVR approach was selected to be used for training and then predicting the correct depths. According to [Guenther 2000], one of the main advantages of SVR is that it has excellent generalization capability with high prediction accuracy. An additional reasoning behind the selection of the linear SVR, instead of a simple linear regression approach, is explained and supported by short experimental results in 4.1.2.3. Support vector machines (SVMs), are popular for solving problems in classification, regression and novelty detection. An important property of SVMs is that the determination of the model parameters corresponds to a convex optimization problem, and so any local solution is also a global optimum [Bishop 2006]. SVMs can also be applied to regression problems by the introduction of an alternative loss function [Smola 1996], preserving, however, the property of

4.1. Correcting the Geometric Effects of Refraction in the 3D Space 37

sparseness [Bishop 2006]. However, the loss function must be modified to include a distance measure. The regression problem is a generalization of the classification problem, in which the model returns a continuous-valued output, as opposed to an output from a finite set [Awad 2015]. As in classification, SVR is characterized by the use of kernels, sparse solution, and Vapnik - Chervonenkis [Vapnik 2015] control of the margin and the number of support vectors [Awad 2015].

As a supervised-learning approach, SVR training is performed using a symmetrical loss function with equal penalization of high and low misestimates. Using Vapnik's [Vapnik 2013] ε -insensitive approach, a flexible tube of minimal radius ε is formed symmetrically around the estimated function, such that the absolute values of errors less than a certain threshold ε (epsilon margin) are ignored both above and below the estimate. In this manner, points outside the tube are penalized, but those within the tube, either above or below the function, receive no penalty (Figure 4.3).

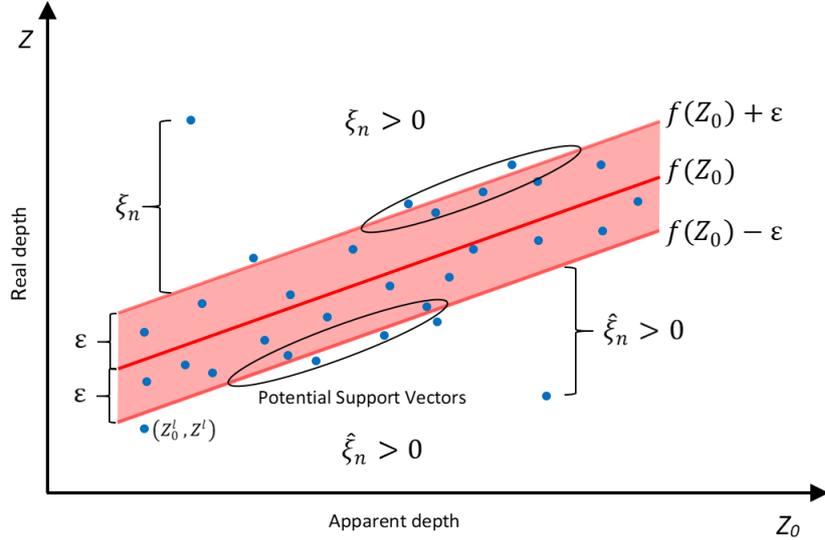


Figure 4.3: Illustration of the employed SVM regression model, showing the regression line together with the insensitive "tube". Also shown are examples of the slack variables ξ_n and $\hat{\xi}_n$. Points above the ε "tube" have $\xi_n > 0$ and $\hat{\xi}_n = 0$, points below the ε "tube" have $\hat{\xi}_n > 0$ and $\xi_n = 0$, and points inside the ε "tube" have $\xi_n = \hat{\xi}_n = 0$.

This will offer to the proposed method the advantage of ignoring the noise in the point clouds introduced by the refraction effect (See also Section 5.2 and 4.1.2.3). For the correction of the refraction effect on the 3D point clouds, the problem is formulated as follows:

Consider the problem of approximating the set of depths:

$$D = \left\{ (Z_0^1, Z^1), \dots, (Z_0^l, Z^l) \right\}, \quad Z_0 \in R^n, \quad Z \in R \quad (4.1)$$

with a linear function

$$f(Z_0) = \langle w, Z_0 \rangle + b \quad (4.2)$$

SVR formulates this function approximation problem as an optimization problem that attempts to find the narrowest tube centered around the surface, while minimizing the prediction error, that is, the distance between the predicted and the desired outputs [Awad 2015]. The optimal regression function formulated and first stated by [Vapnik 2013] is given by the minimum of the functional,

$$\Phi(w, Z_0) = \frac{1}{2} \|w\|^2 + C \sum_i (\xi_n + \hat{\xi}_n) \quad (4.3)$$

where C is a pre-specified positive numeric value that controls the penalty imposed on observations that lie outside the epsilon margin (ε) and helps prevent overfitting, i.e., regularization. This value determines the trade-off between the flatness of $f(Z_0)$ and the amount up to which deviations larger than ε are tolerated as illustrated in Figure 4.3. The value of ε determines the width of the tube; a smaller value indicates a lower tolerance for error and affects the number of support vectors and, consequently, the solution sparsity. $\xi_n, \hat{\xi}_n$ are slack variables representing upper and lower constraints of the outputs of the system; Z is the real depth; and Z_0 is the apparent depth of a point X, Y . By introducing Lagrange multipliers $a_n \geq 0, \hat{a}_n \geq 0, \mu_n \geq 0$, and $\hat{\mu}_n \geq 0$, and optimizing the Lagrangian [Bishop 2006], we now substitute for $f(Z_0)$ using (Equation 6) and then set the derivatives of the Lagrangian with respect to w, b, ξ_n , and $\hat{\xi}_n$ to zero, giving Equation 4.4.

$$\frac{\partial L}{\partial w} = 0 \Rightarrow w = \sum_{n=1}^N (a_n + \hat{a}_n) \Phi(Z_{0n}) \quad (4.4)$$

Substituting Equation 4.3 into Equation 4.2, the predictions for new Z_0 can be made using

$$f(Z_0) = \sum_{n=1}^N (a_n + \hat{a}_n) k(Z_0, Z_{0n}) + b \quad (4.5)$$

which is again expressed in terms of the kernel function k .

The support vectors are those data points that contribute to predictions given by (Equation 9) and the parameter b can be found by considering a data point for which $0 < a_n < C$ [Bishop 2006].

4.1.2.2 Model optimization

The SVR model expressed in Equation 4.5 needs to be solved using an optimization procedure. To perform this, a numerical optimization procedure to search for the coefficients of the hyperplane could be used, however, this is inefficient and is not the approach implemented in widely used SVR implementations. Instead, for the approach presented in this thesis, the most popular method for fitting an SVR model, which is the Sequential Minimal Optimization (SMO) [Platt 1999], is being

4.1. Correcting the Geometric Effects of Refraction in the 3D Space 39

used. This method is a specialized optimization procedure that re-formulates the optimization problem to be a Quadratic Programming problem.

SMO puts chunking to the extreme by iteratively selecting subsets only of size 2 and optimizing the target function with respect to them. It has been reported to have good convergence properties and it is easily implemented [Smola 2004]. The key point is that for a working set of 2 the optimization subproblem can be solved analytically without explicitly invoking a quadratic optimizer. While readily derived for pattern recognition by [Platt 1999], authors in [Smola 2004] had to mimic the original reasoning to obtain an extension to Regression Estimation.

The main difference in implementations of SMO for regression can be found in the way the constant offset b is determined [Keerthi 2001] and which criterion is used to select a new set of variables. Again, this strategy is presented extensively in [Smola 2004]. For a more detailed discussion the SMO see [Smola 2004, Platt 1999]. Unlike interior point algorithms SMO does not automatically provide a value for b . However, this can be chosen by having a close look at the Lagrange multipliers a_n obtained.

4.1.2.3 The selection of the linear SVR correction model

As already stated, an SVR method with linear kernel was adopted in order to address the systematic refraction errors to the depths of point clouds derived from SfM-MVS procedures.

Based on the above, the proposed model will be trained using the real (Z) and the apparent (Z_0) depths of a great number of 3D points to minimize Equation 4.3 and form a model following Equation 4.5 capable of predicting the real depth in the cases where only the apparent depths are available. However, in order to decide the most appropriate method to be used, the relation of the real (Z) and the apparent (Z_0) depths of real world test sites (which are described later in Section 5.3) was studied (Figure 4.4), prior to the selection of the specific regression approach. The use of synthetic data (5), was not considered appropriate for these experiments since the noise in these kind of datasets is eliminated, and thus the results would be unrealistic.

In the study performed, results suggested that the relationship of the real (Z) and the apparent depths (Z_0) of the available points follows a rather linear model (points in blue and orange color in Figure 4.4), and as such, a deeper learning architecture was not considered necessary. To that direction, the use of a Simple Linear Regression model was first examined due to its simplicity; fitting tests were performed in two of the real world test sites presented extensively in Section 5.3 and predicted values were compared to the real (Z) depth data (Figure 4.4). The same experiments were also performed with the Support Vector Regression (SVR) approach with linear kernel.

Regarding the Simple Linear Regression, this approach was rejected, since the predicted models were producing larger errors than the ones produced by the SVR linear regression, and they were highly dependent on the training dataset and its

density, being very sensitive to the noise of the point cloud. This is explained by the fact that the two regression methods differ in the loss function: SVM minimizes hinge loss while linear regression minimizes logistic loss, which diverges faster than hinge loss, being more sensitive to outliers.

This is apparent also in Figure 4.4, where the predicted models using a simple linear regression and an SVM linear regression trained on Amathouda and Agia Napa datasets (described in Section 5.3) are plotted.

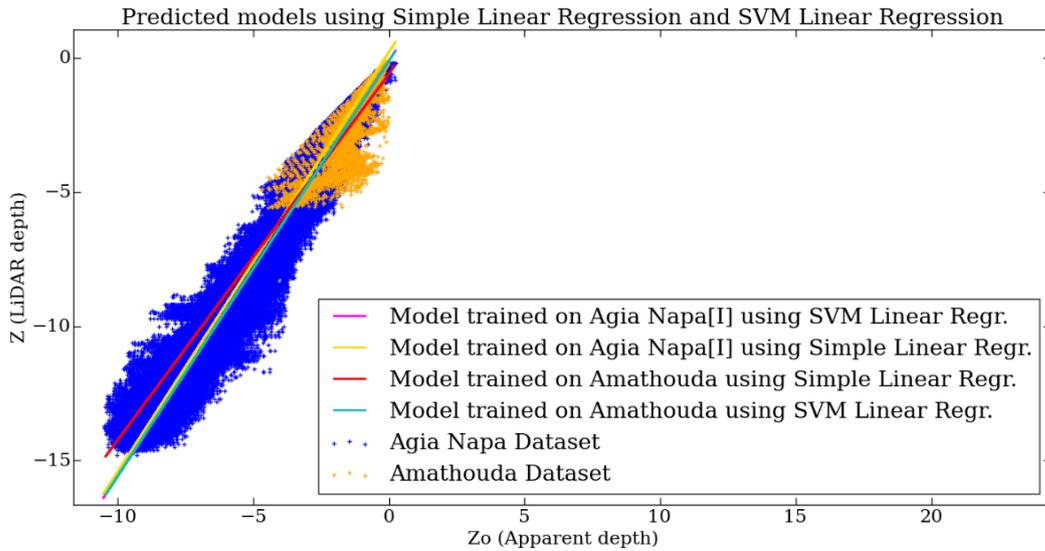


Figure 4.4: The established correlations based on a simple linear regression and SVM linear regression models, trained on Amathouda and Agia Napa datasets.

In the case of training on the Amathouda dataset, it was obvious that the two predicted models (lines in red and cyan color in Figure 4.4) differed considerably as the depth increased, leading to different depth predictions. However, in the case of the models trained in Agia Napa dataset, the two predicted models (lines in magenta and yellow color in Figure 4.4) overlapped, as did the predicted model of the SVM linear regression, trained on Amathouda. These results suggest that the SVR linear regression is less dependent on the density and the noise of the data and, ultimately, is the more robust method. Indeed, according to [Misra 2018], in order to obtain sparse solutions with the SVR, the quadratic error function is replaced by the ϵ -insensitive error function expressed in [Vapnik 2013]. This makes SVR less sensitive to noisy inputs (such as the point clouds that are generated from images with refraction) building a more robust and generalized model, predicting depths over different cases and test sites, outperforming Simple Linear Regression. This also explains the why in the literature till now, simple Linear Regression models trained on sparse data, have been used only for site-specific refraction correction and without much success. The predicted models in those approaches, based on

sparse ground truth data collected mainly using geodetic measurements and applied on shallow rivers, were heavily affected by the noise being apparent on each of the 3D point clouds used, thus not being able to conclude to an accurate generalized model for other bathymetric cases or even produce very accurate results in areas of the same sites used for training.

4.2 Correcting the Geometric Effects of Refraction in the 2D (Image) Space

For a variety of applications having the correct depth information solely in form of 3D point clouds is not enough. To address this requirement, in this second method presented in this thesis, the original imagery is resampled to eliminate the refraction phenomenon. In the following text of the thesis, this method will be referred as Method 2. This method which is published in [Agrafiotis 2020] builds upon and exceeds the state-of-the-art method presented in [Skarlatos 2018] by exploiting the generation of a DSM by the predicted correct depths by the Method 1 presented in Section 4.1.

This, facilitates a much more accurate and reliable SfM process, delivering a more accurate external and internal orientation of the images, directly related to the already available corrected DSM, reducing also the noise in the point clouds caused by the refraction. This also enables the faster and more accurate generation of orthoimagery and detailed textures of the 3D models of the bottom which are very important for several applications such as archeological or benthic community mapping, marine litter detection and image-based bottom classification. Contrary to the analytical based correction approaches, the outputs of the method presented in [Skarlatos 2018] and consequently of the proposed method, can be used by any commercial or built-in SfM-MVS software in order to produce accurate bathymetric and visual results while analytical correction approaches must be incorporated in the SfM-MVS process, something not allowed in most commercial software.

4.2.1 Proposed Method

As mentioned before, the straightforward way to apply water refraction correction into the existing pipeline of commercial SfM-MVS software is image transformation and resampling [Skarlatos 2018]. However, in order to apply refraction correction, the focal distance that varies per pixel dependent on the air-water ratio [Agrafiotis 2015] should be known a priori, which is not the case for any aerial data acquisition system. In the method presented here, to correct image refraction for aerial image-based bathymetry mapping, a new method is built upon a state-of-the-art method presented in [Skarlatos 2018] and expands it by exploiting the correction model presented in Section 4.1. This way, instead of using the erroneous provisional DSMs and iterate the whole process three to four times, as suggested in [Skarlatos 2018], which is a time-consuming and computationally expensive solu-

tion, a DSM created by the predicted corrected depths is used to correct the aerial imagery and facilitate orthoimages and textures generation. The overall workflow of the proposed method is illustrated in Figure 4.5.

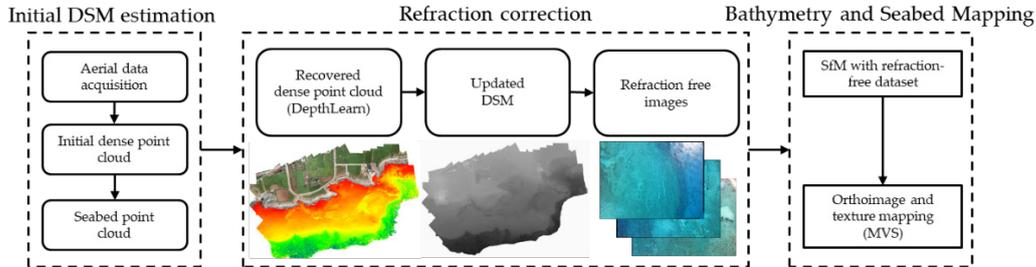


Figure 4.5: The workflow of the proposed method for image correction.

As in Method 1, following the aerial image data collection and Ground Control Points' (GCPs) measurements, an initial SfM-MVS procedure is executed. The resulting initial dense point cloud with systematic depth underestimation is then corrected by employing the correction model of Method 1 for the bottom points. The updated DSM with recovered bathymetry is then used to separately rectify the initial low-altitude aerial image dataset. A new SfM is finally executed based on the refraction-free dataset in order to update the interior and exterior orientation of the cameras, in preparation for orthoimage generation and texturing. During the last stages, texture and orthoimages can be generated based on the updated DSM generated by the corrected dense point clouds using the correction model of Method 1 and the initial dry-land points. In case users wish, they can update also the MVS step. Details for the refraction correction step are elaborated below.

4.2.2 Proposed Refraction Correction Model

The refraction correction method is divided into three modules; the SVR based refraction correction on the dense point clouds by exploiting Method 1, the updated DSM generation using the corrected dense point clouds and the final module of refraction correction in the image space. The three modules are presented in detail in the following sections.

4.2.2.1 SVR based refraction correction on the dense point clouds

The SVR method described in Section 4.1 (Method 1) is adopted in order predict the apparent depth of each point of the initial bottom dense point clouds and get the recovered dense point clouds using Equation 4.5. This way, the proposed method, goes beyond the already available iterative approaches for image refraction correction such as [Skarlatos 2018], since the bathymetric information is a-priori available facilitating a faster and more accurate image correction.

4.2.2.2 Updated DSM generation using the corrected dense point clouds

After correcting the dense point clouds from the refraction effect, a new updated DSM of the area is created. To achieve that, the corrected dense point clouds are firstly merged with the dry land dense points. This merged DSM is created directly in the geographic reference system where the camera positions are already calculated in order to facilitate later the calculation of ground intersections (X, Y, Z) using the projection centers of the images and the image coordinates of points.

4.2.2.3 Refraction correction in the image space

In order to implement the refraction correction procedure, certain principles were followed. The relation between an object point on the bottom, the camera position and the respective image point is shown in Figure 4.6, where the air-water interface is assumed to be planar.

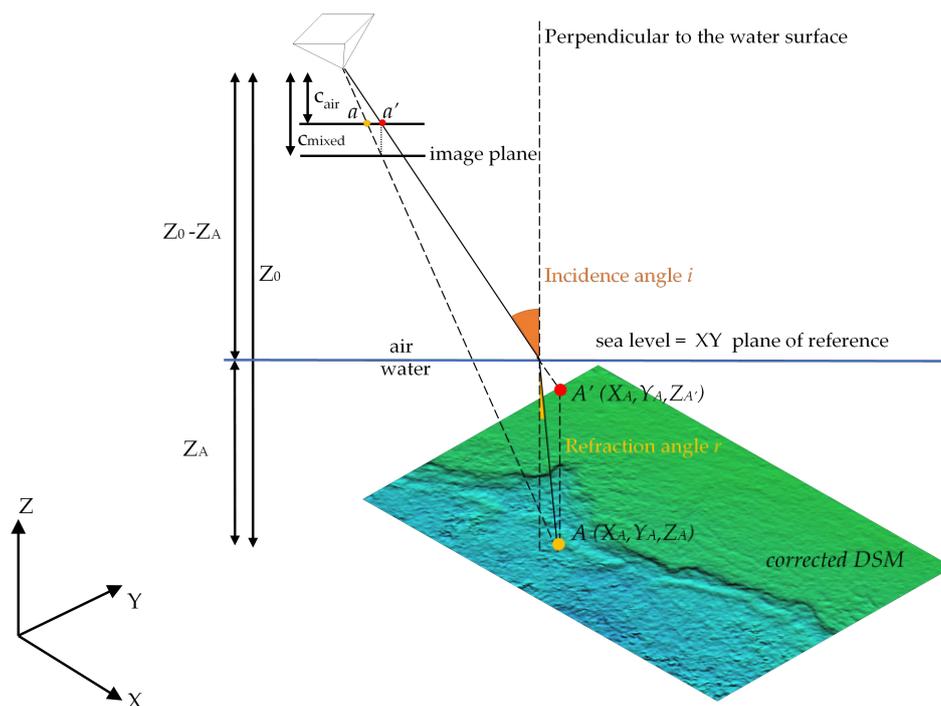


Figure 4.6: The geometry of a single image formation and the repositioning of an image point based on the calculated focal length (c_{mixed}).

The XY plane of the photogrammetric coordinate system X, Y, Z is defined as the boundary plane, hence $Z=0$ for the water level; the Z axis is perpendicular to this plane with its positive direction upwards. Without the presence of the water, a bottom point $A(X_A, Y_A, Z_A)$ is imaged at position a on the image plane,

fulfilling the collinearity equation. However, when water is present, the bottom point $A(X_A, Y_A, Z_A)$ is imaged at position a' in the image plane due to the refraction effect. The apparent position of point $A(X_A, Y_A, Z_A)$ is point $A'(X_A, Y_A, Z'_A)$, lying always in a shallower depth than the real point $A(X_A, Y_A, Z_A)$ and the image point a' is always closer to the projection of the principal point of the camera on the image plane than point a . Considering this, the main aim of the image refraction correction method presented here is to resample the original image containing image points like a' to a refraction-free image containing image points like a .

Starting from a pixel a on the image plane having image coordinates (x, y) , it is possible to estimate its bottom coordinates X_A, Y_A, Z_A using the interior and exterior orientation of the image, calculated already by the SfM process, and the refraction corrected DSM (merged DSM). The line-DSM intersection using a single collinearity function, is an iterative process itself, but the Z coordinate can be easily calculated. Also, exploiting the results of the SfM process such as the georeferenced images following the bundle adjustment, it is easy to identify if a point lies on the ground or on the bottom by its height in the merged DSM. Following, if point A has Z_A , above zero (usually zero is assigned to mean water level for seabed mapping applications), it can be omitted from further processing. If zero does not correspond to water level then a reference point on the shore, measured during GCPs acquisition, may indicate the elevation of the water surface. That way, lakes and calm rivers can also be accommodated by a reference point on the shore and setting the water level Z in the code.

In particular, the image refraction correction steps followed in this method and presented also in [Skarlatos 2018] are:

(a) Grid creation.

A grid is created based on the original image size and every 5 pixels. Experiments performed also using a grid step of 1 pixel suggested that a grid step of 5 pixels is the fastest solution that does not compromise the accuracy of the results. This is because 5 pixels on the ground is almost equal to the GSD size of the updated DSM used for each of the test sites. As such, the area of 3x3 pixels in the interior of the 5x5 cell, is smaller or equal to but not bigger than the GSD of the updated DSM. As a result, for each of the 3x3 pixels, in most cases, the same depth value is assigned. The created grid is then duplicated into the transformed grid. The nodes of this grid will serve later as control points in the final image warping.

(b) Interior orientation.

Contrarywise to [Skarlatos 2018], for every (x_α, y_α) point of the grid, the parameters of the interior orientation retrieved only from the imagery over the dry areas are applied and the new position of the node $(x_\alpha, y_\alpha)'$ is calculated. For the radial distortion the Brown's model [Brown 1972, Fryer 1986] formulated in Equation 4.6, is applied:

$$r'_a = r_a (1 + k_1 r_a^2 + k_2 r_a^4 + k_3 r_a^6) \quad (4.6)$$

where k_1, k_2, k_3 are the radial distortion coefficients and r is the distance from a point (x_α, y_α) to the principal point. For the tangential distortion the Equation 4.7, as formulated in [Mullen 2004] is used:

$$\begin{aligned}\Delta_x &= [P_1 (r_a^2 + 2x_\alpha^2) + 2P_2 x_\alpha y_\alpha] [1 + P_3 r_a^2 + P_4 r_a^4 \dots] \\ \Delta_y &= [2P_1 x_\alpha y_\alpha + P_2 (r_a^2 + 2y_\alpha^2)] [1 + P_3 r_a^2 + P_4 r_a^4 \dots]\end{aligned}\quad (4.7)$$

where P_1, P_2, P_3 and P_4 are the tangential distortion coefficients of the lens and r is the distance from a point (x_α, y_α) to the principal point.

(c) Collinearity equation intersection with the updated DSM. Then, the collinearity equation is applied and its intersection (X_A, Y_A, Z_A) with the ground determines the point imaged at (x_α, y_α) , as calculated using the available georeferenced DSM.

(d) calculation of c_{mixed} . In [Agrafiotis 2015] it is proved that in two media photogrammetry, the relation of the effective camera constant (c_{mixed}) to the one in air (c_{air}) is depending on the percentages of air and water within the total camera-to-object distance. Following, for every point on the updated DSM which has $Z < 0\text{m}$, this Z is used to calculate c_{mixed} for this specific point, according to the percentage of air and water distance travelled by the ray. The c_{mixed} is then used for the reprojection of that point back to the original image plane, which was taken with c_{air} . Adopting the Equation 4.8 of [Agrafiotis 2015] where the focal length (i.e. camera constant) of the camera is expressed by

$$c_{mixed} = (P_{air} n_{air} + P_{water} n_{water}) c_{air} \quad (4.8)$$

where P_{air} and P_{water} are the percentages of the ray travelling in air and water respectively, and n_{air} and n_{water} the refraction indices respectively, it is easy to calculate c_{mixed} by using camera focal length on air c_{air} , considering n_{air} equal to 1 and approximating n_{water} by Equation 4.9 [Quan 1995] or measuring it directly.

$$\begin{aligned}n(S, T, l) &= 1.447824 + 3.0110e^{-4}S - 1.8029e^{-5}T \\ &\quad - 1.6916e^{-6}T^2 - 4.89040e^{-1}l \\ &\quad + 7.28364e^{-1}l^2 - 3.83745e^{-1}l^3 \\ &\quad - S (7.9362e^{-7}T - 8.0597e^{-9}T^2 + 4.249e^{-4}l \\ &\quad - 5.847e^{-4}l^2 + 2.812e^{-4}l^3)\end{aligned}\quad (4.9)$$

where S, T, l the water salinity, water temperature and light wavelength respectively.

The most challenging part of Equation 4.8 is the calculation of the ray travelling percentages in water and air, since in most of the cases this is a priori unknown and changes for every point (pixel) of the image. Nevertheless, in this method, the refraction corrected updated DSM is available from Method 1 and thus it is possible to retrieve this information with high accuracy from the external orientation of the image and the real depth (Z_A) of point A (Figure 4.6). Having estimated the c_{mixed}

for the point A , the correct position a of the image point a' , can be calculated by applying the collinearity equation (to the opposite direction).

(e) Point reprojection. Having estimated the c_{mixed} for point A , the correct position a (x_α , y_α) of the image point a' , can be calculated by applying the collinearity equation (Equation 4.10) and solving for x_α , y_α :

$$\begin{bmatrix} x_\alpha \\ y_\alpha \\ -c_{mixed} \end{bmatrix} = \lambda_\alpha \times R_{\alpha\omega\varphi\kappa} \times \begin{bmatrix} X_O - X_A \\ Y_O - Y_A \\ Z_O - Z_A \end{bmatrix} \quad (4.10)$$

Where (X_O, Y_O, Z_O) is the position of the perspective center, (X_A, Y_A, Z_A) the position of point A on the seabed, $R_{\alpha\omega\varphi\kappa}$ the rotation matrix of the camera, and λ_α the image scale at point a . Then, the inverse camera calibration model is applied by solving Equation 4.6 for r in order to calculate (x_α, y_α) on the image plane and place (x_α, y_α) at the corresponding cell in the transformed grid.

(f) New image creation-Grid based image warping. Finally, a new image is created by resampling using the initial grid and the transformed grid and bicubic interpolation (Figure 4.7).

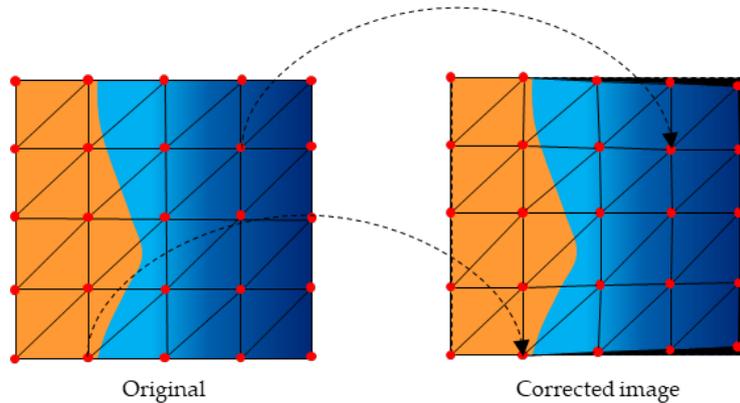


Figure 4.7: The process of the creation of the corrected new image by resampling using the initial grid and the transformed grid and bicubic interpolation. It is assumed a nadir image.

To that end, a Delaunay triangulation [Green 1978] of the fixed control points is created, and maps corresponding moving control points (the nodes of the transformed grid) to the fixed control points (the nodes of the grid). The process is presented in Figure 4.7 in which the red dots are the nodes of the grid created every 5 pixels, the black triangles are the ones generated after the Delaunay triangulation, the orange area represents a dry area on the image while the gradient blue area stands for the part of the image covered by water. The darkest the blue color,

the deeper the area. In this figure, it can be noticed that the nodes of the grid, the control points on the image as reported before, are not moved in the corrected image. On the contrary the control points over the water covered area are moved in proportion to the depth and the radial distance from the perspective center.

The mapping is linear (affine) for each triangle and continuous across the control points. However, each triangle has its own mapping (transformation) since in a piecewise linear transformation, linear (affine) transformations are applied separately to each triangular region of the image [Goshtasby 1986, Goshtasby 2005].

4.2.2.4 SfM with Refraction-Free Dataset – Generation of Orthoimages and Textured 3D Models

For updating the SfM solution, the refraction-free images are now employed. In this way, the correct interior and exterior orientations for the cameras will be calculated and will facilitate the generation of more accurate and reliable orthoimages and textures. Orthoimages and textured models can be generated using the corrected imagery and the 3D mesh produced using the refraction corrected point cloud; MVS could even be performed from scratch, should there be a specific need for it.

4.2.3 Summary

In this chapter, two different but also complementary methods for correcting the geometric effects of refraction were presented. Both methods are intended to be used over calm water surface, clear water bodies and textured seabed. The first one, relies on machine learning tools which can accurately recover shallow bathymetric information from low-altitude aerial image datasets. In particular, the main objective of this method is to train and employ a linear Support Vector Regression (SVR) model able to learn the systematic underestimation of the depths by correlating the apparent and the real depth of many points and facilitate the estimation of the correct depth, by feeding it only with the apparent one.

The second developed method for correcting the geometric effects of refraction in the image space, builds upon the image correction method presented in [Skarlatos 2018] by first exploiting the first method, and then correcting the original imaging dataset from refraction, exploiting image transformation and resampling techniques. Main objective of this method is to overcome the refraction effect directly in the raw data. Having corrected the raw data, the execution of the SfM-MVS processing pipelines on a refraction-free set of aerial images resulting into highly accurate bathymetric maps is facilitated.

In the next chapter, the accuracy limits of these methods together with their advantages and disadvantages are revealed by implementing them in synthetic and real-world data.

Experimental Results and Validation of the Developed Geometric Correction Methods

Contents

5.1	Synthetic Training and Test Datasets	50
5.1.1	Synthetic Data Generation	50
5.2	Experimental Results and Validation on the Synthetic Datasets	56
5.2.1	Correcting the Geometric Effects of Refraction on the point clouds	56
5.2.2	Is a ratio-based depth determination better?	69
5.2.3	Summary	73
5.3	Real World Test Datasets	74
5.3.1	Datasets Description	75
5.3.2	Factors affecting the low altitude aerial imagery over shallow waters	80
5.3.3	Ground Control Points measurements	81
5.3.4	Initial Dense Point Clouds Generation	81
5.3.5	The influence of the Base-to-Height ratio (B/H) of stereopairs on the apparent depths	83
5.3.6	True depth data	85
5.3.7	Seabed point cloud extraction	87
5.4	Experimental Results and Validation on Real World Data	89
5.4.1	Method 1: Correcting the Geometric Effects of Refraction to the point clouds	89
5.4.2	Method 2: Correcting the Geometric Effects of Refraction on the Image Space	122
5.5	Comparative analysis of the proposed methods	139
5.6	Discussion	145

This chapter aims to present the experiments performed together with the respective results and validation of the two developed methods of this thesis regarding the correction of the geometric effects of refraction. To that direction and in order to

check and evaluate thoroughly the proposed methods in terms of robustness, effectiveness, accuracy and generalization over different areas and mission characteristics, several tests were performed using synthetic and real-world datasets.

Section 5.1 presents the synthetic test datasets creation and interpretation, Section 5.2 discusses the experimental results and validation performed over the synthetic data while Section 5.3 presents the real-world datasets acquisition and interpretation. Section 5.4 presents and discusses the experimental results and validation performed on the real world dataset while in Section 5.5 a comparative analysis of the presented results is performed. Section 5.6 gives a summary, discusses the achieved results and suggests further work to be done.

5.1 Synthetic Training and Test Datasets

To initially check and evaluate Method 1 which is described in Chapter 3, for correcting the geometric effects of refraction on the point clouds in a controlled environment and confirm the assumption of the linearity of the relation of the true (Z) and apparent (Z_0) depths (Figure 4.3), synthetic data were generated and processed accordingly. These data will later facilitate the implementation of a transfer of the learning approach for correcting the real world data that are processed in Section 5.4. To that direction, two different Digital Terrain Models (DTMs) were generated using a periodic mathematical function. A suitable texture pattern was used to color these DTMs as explained below and six synthetic aerial datasets were generated; for three of them, refraction was added on the imagery while for the other three, images were generated without refraction for comparison reasons. Then, using these data, SfM-MVS processes were performed and the point clouds created using the refracted imagery were corrected using the proposed method.

The advantage of the synthetic data is the accuracy and reliability of the depth information and the exact knowledge of exterior and interior orientations of the cameras used. Moreover, when it comes to seabed imaging, errors and limitations in image matching caused by the visibility restrictions due to the depth, especially in depths more than 10m, and errors introduced by the wavy surface are excluded. Also, by using a mathematical function to generate and describe the DTMs, incompatibilities and errors that might be transferred to the solution by the true depth data are avoided, leading to independent and objective results, allowing also the cheap generation of high quality training data. In the following sections, the synthetic data generation process is described in detail while next, the tests performed, and the respective experimental results are presented and discussed.

5.1.1 Synthetic Data Generation

Main aim of the work described in this section is the synthetic data generation, given that the interior and exterior orientations of the cameras and the DTM accuracy are a priori set. A real texture for the seabed was also used. To that direction, images of the synthetic seabed were generated with the absence and the presence of

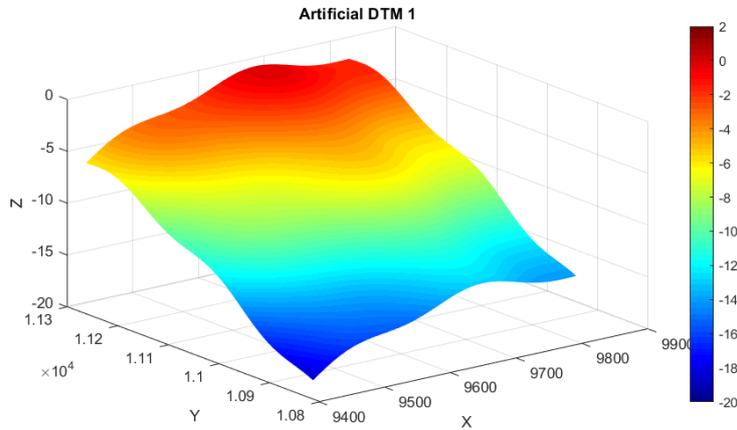
a synthetic water surface in order to train and test Method 1, which is based on the SVR model. The generation of data of different heights, different flight paths and different DTMs at this stage, was considered also of high importance.

5.1.1.1 Digital Terrain Model (DTM) Generation

In order to have knowledge of the exact elevation for each position X , Y of the DTM, a continuous function $Z = f(X, Y)$ was adopted to generate it. This way, when comparing the true and the apparent depths, the interpolation between discrete points is avoided, since it would introduce additional errors in the process, not related with the refraction effect. Towards that, two different surfaces were generated; the first one, named DTM1, describes in a realistic way the general seabed anaglyph in shallow water areas, while the second one, named DTM2, is characterized by changes in the elevation and more intense slopes. Although DTM2 is not a realistic seabed anaglyph, it was considered important to test the proposed method in this type of seabed in order to highlight its robustness or discover potential limitations. Equation 5.1 describes the function $Z = f(X, Y)$ that was adopted for both TMs generated:

$$Z(X, Y) = (Z_0 + (X - X_0) \times inc + (Y - Y_0) \times inc + ap \times \sin((X - X_0) \times ep) - ap \times \sin((Y + Y_0) \times ep) - as \times \sin((X - X_0) \times es) - as \times \sin((Y + Y_0) \times es)) \quad (5.1)$$

where for **DTM1**: $X_0 = 9312.94$, $Y_0 = 10729.49$, $Z_0 = -19$, the inclination $inc = 0.005$, and the parameters of the two sinusoidal functions used: $ap = 7$, $as = 0.5$, $ep = 0.00448785722$, $es = 0.0314150006$, while for **DTM2**: $X_0 = 9312.94$, $Y_0 = 10729.49$, $Z_0 = -19$, the inclination $inc = 0.005$, and the parameters of the two sinusoidal functions used: $ap = 6$, $as = 3$, $ep = 0.00448785722$, $es = 0.0314150006$. The generated DTMs are illustrated in Figure 5.1.



(a)

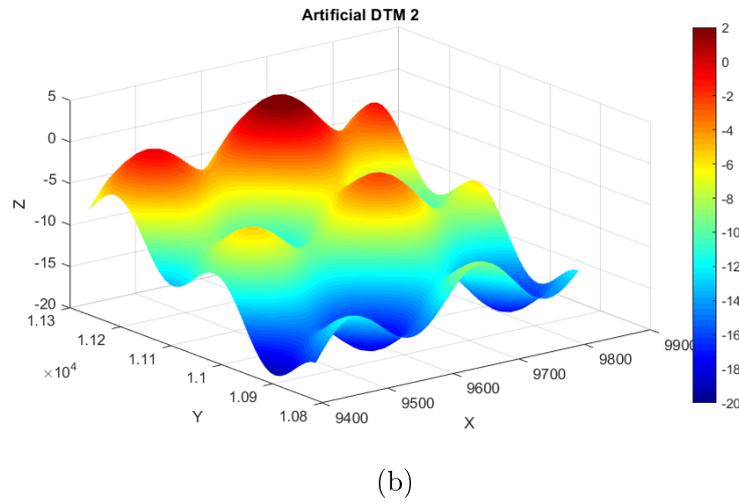


Figure 5.1: The two different synthetically generated DTMs. X, Y, Z are in meters. DTM1 is illustrated in (a) and DTM2 in (b)

The maximum elevation of DTM1 (Figure 5.1a) is -0.13m while the maximum elevation of DTM2 (Figure 5.1b) is 2.65m. Their minimum elevation (or maximum depth) is -18.57m and -19.05m respectively. For both DTMs, the synthetic water surface level was set to elevation of 0m. The two DTMs were then textured using an already available typical texture of a seabed area in the Mediterranean Sea, depicting rock formations, sandy areas and seagrass covered areas (Figure 5.2).



Figure 5.2: The real seabed texture used for both the DTMs

To texture both the DTMs, affine transformations were applied on this image in order to cover the whole area of the DTM. The GSD of the generated texture was set to 0.03m x 0.03m to facilitate the synthetic imagery generation in adequate resolution later.

5.1.1.2 Synthetic Images Generation

For the proof of concept of the first proposed method, it was necessary to generate synthetic images having exactly the same exterior orientations but to image the seabed areas with and without the presence of the water. As such, six synthetic datasets were created for two representative flying heights; for DTM1, two flights were simulated at a flying height of 150m and two flights were simulated at a flying height of 200m while for DTM2 two flights were simulated at a flying height of 200m.

For generating the images, the already mentioned texture was treated as an orthoimage. As such, the inverse process of the orthoimage generation was performed to create images from the specified camera positions, having also the specified interior orientation. For the generation of the non-refracted images, the straight-forward approach presented in [Skarlatos 2006] was implemented, expanding also to generate RGB imagery. For introducing the refraction effect on the imagery, the inverse approach of the correction method described in Section 4.2 was followed. Regarding the camera parameters, for DTM1, a focal length of 3.61mm with pixel size of $1.56\mu\text{m}$ and a typical image size of 4000 x 3000 was selected, while for DTM2, a focal length of 4.50mm was used, having the same pixel size and the same image size with the images of DTM1, representing a vast majority of the commercial light weight RGB sensors used for low altitude UAV image-based mapping (i.e. the DJI FC330, the GoPro Hero 4 etc.). In Figure 5.3, examples of the synthetic images are given.

In Figure 5.3a, an example of a generated synthetic image without refraction for DTM1 at 200m flying height is presented while in Figure 5.3b an example of a generated synthetic image without refraction for DTM2 at 200m flying height is presented. In Figure 5.3c, and Figure 5.3d the respective generated synthetic images with refraction are presented, having exactly the same exterior orientation with the non-refracted ones. On the left edges of all the images, the scale difference due to the characteristics of DTM1 and DTM2 respectively is obvious. In the same figure, it can be noticed that the images which are affected by the refraction phenomenon (Figure 5.3c and Figure 5.3d), are magnified, depicting a smaller area on the seabed. This is in line with Equation 4.8 and the results presented in [Agrafiotis 2015] where it is stated that when water is added in the light ray path, the effective camera constant is always larger than the camera constant in air.

Table 5.1 presents the details of the simulated flights and subsequent image-based processing for the six simulated aerial UAV campaigns. It may be noticed that they have different average flying heights and overlaps indicating that the suggested solution is not limited to specific flying heights and specific overlapping. Moreover, as an effect of refraction on the images, it can be noticed that in the cases of

the refracted datasets, the *RMSEs* are increased, compared with the ones resulted from the non-refracted datasets. This effect of refraction on the image matching and consequently the SfM accuracy is discussed in more detail in 5.4.1.2 and 5.4.1.3.

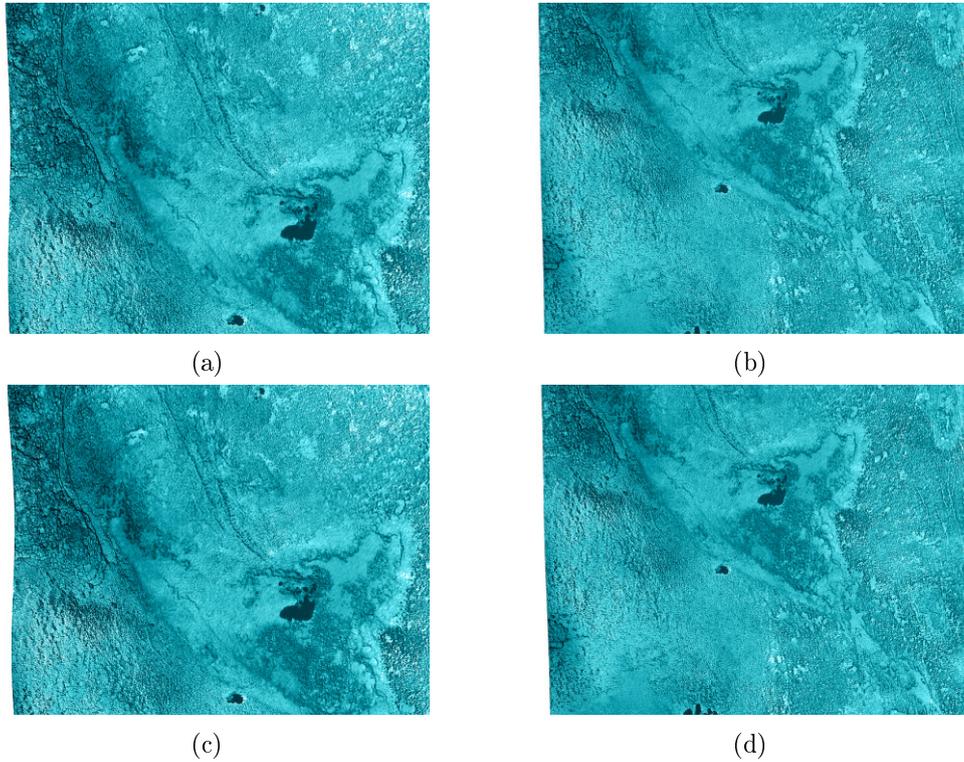


Figure 5.3: Example images of the generated synthetic imagery without refraction for the DTM1 at 200m flying (a) and for the DTM2 at the same flying height (b). The respective generated synthetic images with refraction are presented in (c) and (d), having exactly the same exterior orientation with the unrefracted ones.

Dataset	150m [DTM1] [non-refracted]	150m [DTM1]	200m [DTM1] [non-refracted]	200m [DTM1]	200m [DTM2] [non-refracted]	200m [DTM2]
# Images	24	24	12	12	16	16
Flying height [m]	150	150	200	200	200	200
Avg. Base-to-Height (B/H) ratio along strip (without refraction)	0.45	0.45	0.45	0.45	0.31	0.31
Avg. Base-to-Height (B/H) ratio across strip (without refraction)	0.52	0.52	0.52	0.52	0.69	0.69
Along strip overlap (without refraction)	65%	65%	65%	65%	70%	70%
Across strip overlap (without refraction)	70%	70%	70%	70%	50%	50%
Image footprint on the ground [m] (without refraction)	260 x 195	260 x 195	344 x 258	344 x 258	277 x 208	277 x 208
GSD [m] (without refraction)	0.065	0.065	0.086	0.086	0.069	0.069
$RMSE_X$ [m]	0.005	0.006	0.002	0.004	0.004	0.012
$RMSE_Y$ [m]	0.002	0.014	0.001	0.014	0.002	0.063
$RMSE_Z$ [m]	0.001	0.015	0.001	0.007	0.001	0.016
Pixel size [μ m]	1.56	1.56	1.56	1.56	1.56	1.56
Camera constant c [mm] (without refraction)	3.61	3.61	3.61	3.61	4.5	4.5

Table 5.1: The six simulated aerial UAV campaigns, the derived datasets and relative information.

5.2 Experimental Results and Validation on the Synthetic Datasets

In this section, the experiments and the respective results of the tests performed on the synthetically generated datasets are presented. The experimental results include testing and validation only of Method 1, which corrects the geometric effects of refraction on the image-based point clouds, since for the introduction of the geometric effects of the refraction to the synthetic imagery the inverted Method 2 was used. To that direction, the steps described in Chapter 3, generate the required data and correct the effects of refraction on the submerged point clouds.

5.2.1 Correcting the Geometric Effects of Refraction in the 3D space

5.2.1.1 Additional errors in the point cloud generation process, not related with the refraction effect

In order to estimate the amount and the type (random or systematic) of the errors in depth determination that are not related to the proposed refraction correction method but are introduced by the SfM-MVS processing and possibly by the synthetic data generation pipeline, the following comparison was performed: The elevations of the X , Y points generated from these processes using the non-refracted synthetic datasets, were compared with the respective true elevations calculated using Equation 5.1 for exactly the same X , Y points. Results are presented in Table 5.2 while the histograms of the calculated differences in relation to the true elevation are illustrated in Figure 5.4 and Figure 5.5.

Dataset [non-refracted]	\bar{x} [m]	s [m]	$RMSE_Z$ [m]	Expected $RMSE_Z$ according to [Smith 2015] [m]	Evaluated Points
150m [DTM1]	0.055	0.043	0.070	0.23	2.507.667
200m [DTM1]	0.069	0.030	0.078	0.31	5.789.766
200m [DTM2]	0.071	0.047	0.085	0.31	2.162.523

Table 5.2: Differences between the calculated and the real elevations of the point clouds generated using the non-refracted datasets. \bar{x} is the average distance of the point cloud from the true elevations and s its standard deviation.

As expected, in Table 5.2, Figure 5.4 and Figure 5.5, it can be noticed that the average elevation difference is increasing as the flying height and the complexity of the DTM increase. Indeed, according to [Nesbit 2019], nadir-only datasets generally generate higher point counts due to the easier and more accurate dense image matching, higher precision, and compared to true elevation data, mean errors closer to 0m. However, as also reported in [Nesbit 2019], the spatial distribution of the

5.2. Experimental Results and Validation on the Synthetic Datasets 57

errors of those point clouds reveals a systematic pattern. Moreover, in [Smith 2015], authors performed a survey of the published geoscience literature and reported that when compact cameras are used, which is the case in most UAV systems, the ratio of the root mean squared error of the elevations ($RMSE_Z$) to the flying height for more than 40 scenarios was found to be $\sim 1:640$. However, as can be seen in Table 2, in the examined cases, mainly due to the absence of lens distortions (radial or decentering), the mean differences \bar{x} and the $RMSE_Z$ between the calculated and the true elevations of the points are almost equal or less than the GSD size of the images of each datasets.

Figure 5.4 serves as an explanatory figure of the 2D histograms that are presented next in the thesis. There, the 2D histogram of the comparisons performed for the 150m[DTM1][non-refracted] synthetic dataset, which correlates the elevation and the differences between the true and the calculated elevations, is accompanied with two 1D histograms; the one on top is the histogram of the differences for the elevation of 6m (a section on the 2D histogram parallel to the X axis) while the one on the right is the histogram of the elevations for the difference equals with 0.055m (a section on the 2D histogram parallel to the Y axis), which is the average elevation difference for this dataset.

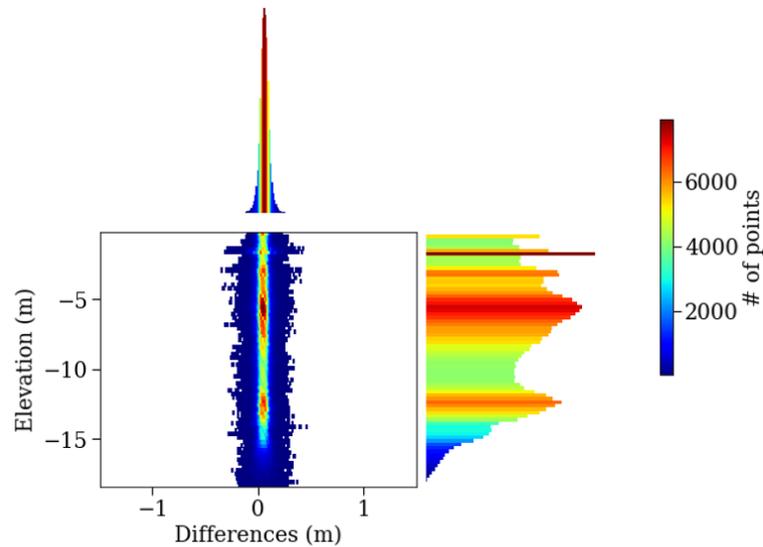


Figure 5.4: An example 2D histogram of the differences between the real (resulted from the Z function) and the SfM-MVS calculated elevations using the non-refracted imagery for the 150m[DTM1][non-refracted] dataset, together with explanatory sections (1D histograms) along X and Y axis.

Regarding the histogram of the elevations, it is directly related to the DTM used for the synthetic data generation, and thus, for DTM1 and DTM2 would be different. On the contrary, the histogram of the differences is described by a Gaussian

distribution for all the tests performed (Figure 5.5), as described in [Smith 2015].

In Figure 5.5, the histograms of the differences between the true (resulted from the Z function) and the SfM-MVS calculated elevations using the non-refracted imagery, for all the three non-refracted synthetic datasets are presented. There, it can be noticed that SfM-MVS processing, is underestimating the elevations, however, as will be proved in the next section, this systematic effect, also reported in [Smith 2015, Nesbit 2019], can be accommodated by the SVR model. These systematic errors, which are almost equal or less than the GSD size, and are probably caused by the absence of Ground Control Points (GCPs) in the dataset, are accompanied by larger differences, that are scattered almost equally on either side of the mean value, forming the Gaussian distribution already reported. These larger differences, that according to the literature [Smith 2015, Nesbit 2019] are present in most of the parallel-axis (nadir) image blocks, are justifying the calculated standard deviations of Table 5.2 and they are indicating the noise in the elevations of the point cloud, introduced mainly by the SfM-MVS process.

It is expected that this noise will be more intense when refraction will be added on the images, primarily due to the erroneous key point matching of points at the edges of the images which are severely affected by the effect (see also Subsection 2.2.4 and Subsection 5.3.5) and secondarily by the usage of RGB imagery, since the amount of refraction is different for each wavelength.

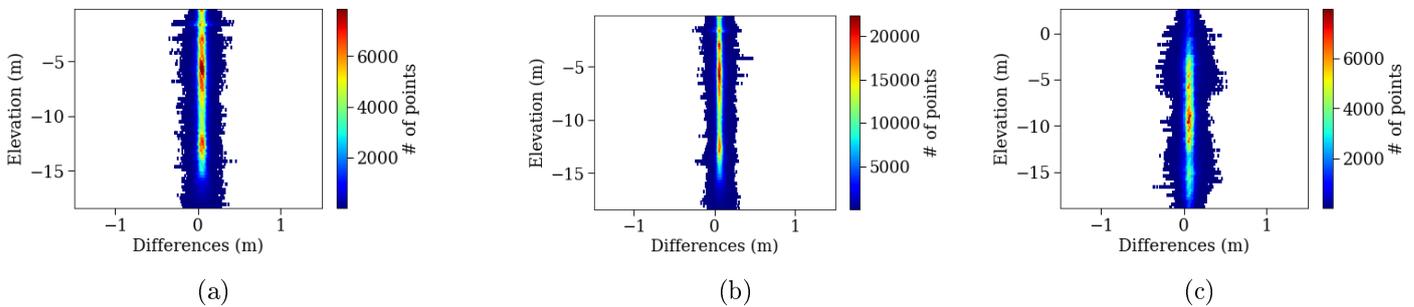


Figure 5.5: The 2D histograms of the differences between the real (resulted from the Z function) and the SfM-MVS calculated elevations using the imagery for the 150m[DTM1][non-refracted] dataset (a), the 200m[DTM1][non-refracted] dataset (b) and the 200m[DTM2][non-refracted] dataset (c) respectively, in relation to the true elevation.

5.2.1.2 Training, Validation and Testing

Having estimated the additional errors that are introduced in the point cloud generation process and are not related with the refraction effect, the linear SVR based method presented in Chapter 3 is implemented in the already described synthetic datasets with refraction. In order to demonstrate that the proposed method has significant potential for applications over different areas and shallow waters, and that a predicted model can be used in areas where true depth data are not avail-

5.2. Experimental Results and Validation on the Synthetic Datasets 59

able, training and cross-testing was performed also in different synthetic sites. As such, a model trained using the image-based depths (the apparent) and the ground true depths of the 3D point cloud of a synthetic test site, was tested by predicting the correct depths on the rest of the synthetic test sites, where the real depths were also available but used only for evaluation. Later, in Subsection 5.4.1, these trained models will be used to predict the correct depths over the real world datasets as well.

For the test performed here, three different training sets were formed (Figure 5.6) and then validated against 9 different testing sets. Training and testing processes were performed in different test sites each. In the first row of Figure 5.6, the depth (Z) – Z_0 2D histograms of the synthetic datasets and the respective point densities of the 100% of the points is presented. Since with a first glimpse, the distribution of the density of the correlated Z and Z_0 seems to be linear, but thick enough, in the second row of Figure 5.6, the depth (Z) – Z_0 2D histograms of the denser bins is presented, indicating that the amount of the noise surrounding the core of the 2D histograms, representing the majority of the points, is insignificant.

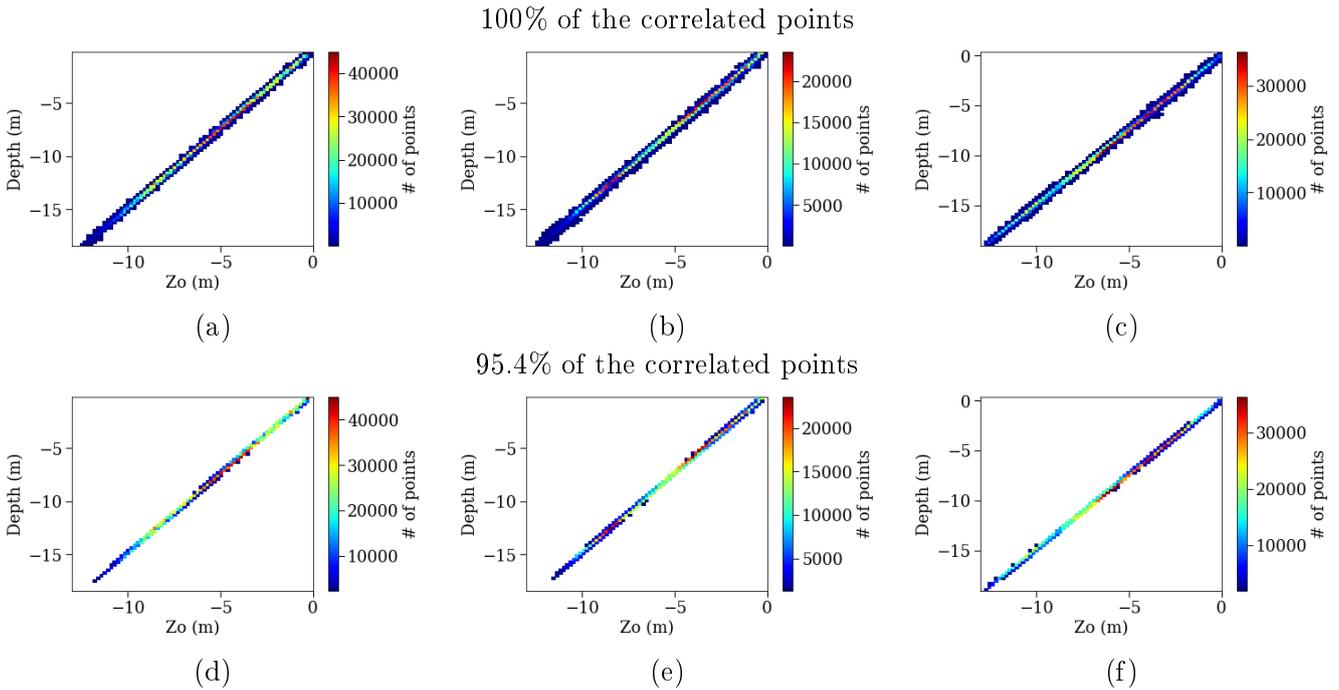


Figure 5.6: The Depth (Z) – Z_0 2D histograms of the synthetic datasets and the respective point densities of the 100% of the correlated points: The histograms for 150m[DTM1] (a), for 200m[DTM1] (b) and for 200m[DTM2] (c) datasets are presented. Figures (d), (e) and (f) are presenting the depth (Z) – Z_0 distribution and the respective point densities of the 95.4% of the correlated points respectively.

Below an example of the results presented in Figure 5.6 is given. Considering the first row that represents the 100% of the denser bins, a point having a true depth

of 10 m, can be present in the refracted point cloud with apparent depths ranging from 7.23m to 6.40m. However, when the second row of Figure 5.6 is examined, representing the 95.4% of the denser bins, this range is reduced to 0.42 m, ranging from 7.02m to 6.60m. The remaining differences of ± 0.21 m, are similar to the errors caused by the SfM-MVS process, reported before and are apparent also in the datasets without refraction. On the contrary, the gross errors that are represented only by the 4.6% of the denser bins of the 2D histograms, can be attributed to the refraction effect and the erroneous matches achieved in some cases (see also Subsection 2.2.4. Finally, it should be noticed that in the first row of Figure 5.6, a marginal trend is observed towards the underestimation of the depths, which is again explained by the erroneous matches achieved in the outer areas of the images which are affected the most by the refraction effect, due to the large incidence angles of the light rays captured there (see also Subsection 2.2.4 and Subsection 5.3.5).

Training

As already stated, SVR has a great potential to deliver robust models, even using sparse data for training (See 4.1.2.3). Due to the large size of the available data for correlation in all the three synthetic datasets, the use of percentages of the total data for training was investigated. To that direction, models were trained using the 80% of the data, the 30% of the data and the 5% of the data. Results suggested that all three training approaches for each test site delivered exactly the same results, without any improvement or compromise in the accuracy. However, training on the 80% of the data required much more computational resources compared to the one with 30% and the one with 5% of the data. Considering this, training was performed using the 5% of the available points. This way, the 5% of the points of these datasets were randomly picked to form the training and the validation set. This decision, was also aiming at excluding possible noise existing in the data and keep for training the data that are closer to the areas of high density presented in Figure 5.6.

Considering the above, the first training approach used the 5% of the 150m[DTM1], the second training approach used the 5% of the 200m[DTM1] while the third training approach used the 5% of the 200m[DTM2] synthetic datasets. The $Z-Z_0$ distribution of the points used for this training can be seen in Figure 5.6. The three training approaches are summarized below:

1. 5% of the 150m[DTM1][5%],
2. 5% of the 200m[DTM1][5%] and
3. 5% of the 200m[DTM2][5%]

Testing

Using the models trained on the 5% of the 150m[DTM1] dataset, the corrected depth over the 95% of the 150m[DTM1], the 95% of the 200m[DTM1] the 95% of the 200m[DTM2] datasets was predicted. Respectively, using the models trained

5.2. Experimental Results and Validation on the Synthetic Datasets 61

on the 5% of the 200m[DTM1] dataset, the corrected depth over the 95% of the 150m[DTM1], the 95% of the 200m[DTM1] the 95% of the 200m[DTM2] datasets was predicted. Finally, using the models trained on the 5% of the 200m[DTM2] dataset, the corrected depth over the 95% of the 150m[DTM1], the 95% of the 200m[DTM1] the 95% of the 200m[DTM2] datasets was predicted. The tests performed are summarized below:

1. For the model trained on the 5% of the 150m[DTM1] dataset, testing was performed on:
 - (a) the 95% of the 150m[DTM1] dataset,
 - (b) the 95% of the 200m[DTM1] dataset and
 - (c) the 95% of the 200m[DTM2] dataset
2. For the model trained on the 5% of the 200m[DTM1] dataset
 - (a) the 95% of the 150m[DTM1] dataset,
 - (b) the 95% of the 200m[DTM1] dataset and
 - (c) the 95% of the 200m[DTM2] dataset
3. For the model trained on the 5% of the 200m[DTM2] dataset
 - (a) the 95% of the 150m[DTM1] dataset,
 - (b) the 95% of the 200m[DTM1] dataset and
 - (c) the 95% of the 200m[DTM2] dataset

To facilitate the visual comparison between the predicted models, Figure 5.7 demonstrates the 3 models, predicting depths over the 95% of the 150m[DTM1] dataset. It is obvious that the models are overlapping, although they are trained in datasets having different flying heights, different, interior and exterior orientations, different B/H ratios (see also Subsection 5.3.5) and different seabed anaglyph. They succeed in following the Z - Z_0 distribution of the large percentage of the points (see also Figure 5.6). In the next section it will be shown that the percentage of the points that the models succeed in following the Z - Z_0 distribution within the accuracy levels of IHO [Guenther 2000] is more than 95.4%. This is based on the 2D histograms of the differences between the true depths and the corrected depths of the point clouds after the application of the proposed method.

In Figure 5.7, a number of outlier points appear to lie away from the predicted models. However, as can be seen in Figure 5.6, these points are less than 5% of the total correlated points and they are not enough to affect the final accuracy of the results. In the next paragraphs, the above models are evaluated in terms of accuracy and they highlight the high performance of this proposed method discussing also issues and differences observed between the predicted and real depths calculated by the Z function (Equation 5.1).

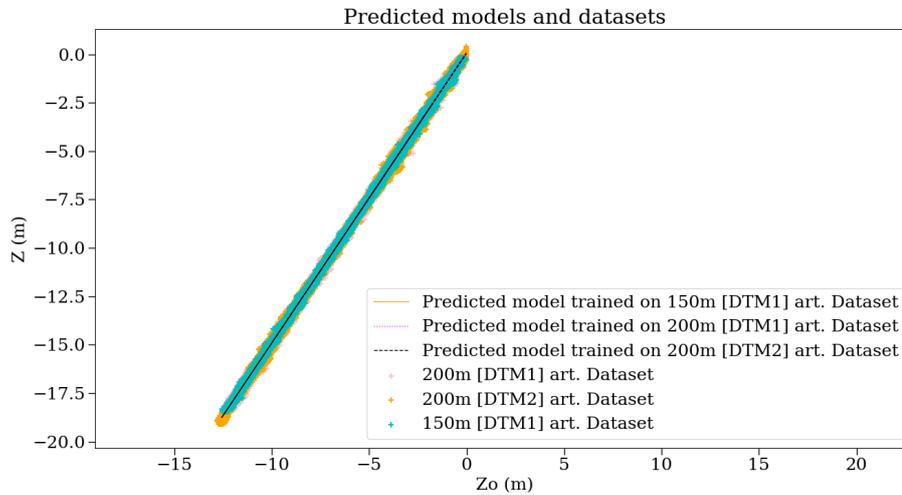


Figure 5.7: The Z - Z_0 distribution of the employed synthetic datasets and their respective, predicted linear models.

5.2.1.3 Evaluation of the predicted depths in the synthetic datasets

As a first step towards the evaluation of the trained models and their predicted depths, the amount of the errors that is introduced by the refraction effect is demonstrated by comparing the initial (uncorrected or apparent) depths (Z_0) of the synthetic datasets with the true depths (Z) calculated by the Z function (Equation 5.1). In Figure 5.8, it can be easily observed that the differences between the true depths $Z = f(X, Y)$ and the uncorrected depths are increasing proportionally to the depth, reaching a mean difference of 5.5m at the depth of 15m. This means that the mean difference is in the order of 36.7% of the true depth. These results clarify that the refraction effect cannot be ignored in bathymetric and seabed mapping applications. In all six cases demonstrated in Figure 5.8, the average value of the differences is significant, being between 2.48m to 2.96m with a standard deviation (1 sigma) of 1.47m and 1.54m and *RMSEs* of 2.87m and 3.34m respectively. As can be observed in the same figure, the majority of the average differences in all of the histograms are far from the red dashed lines representing the accuracy limits generally accepted for hydrography, as introduced by the International Hydrographic Organization (IHO) [Guenther 2000]. As before, the 100% and the 95.4% of the denser bins are demonstrated respectively in the first and the second row of Figure 5.8. As expected, the noise reported before (see 5.2.1.1), is transferred to the differences between the elevations of the uncorrected dense point clouds and the true depth data. Again here, the trend towards the underestimation of the depths is confirmed. However, as in the case of the non-refracted data, this noise is eliminated when the 95.4% of the points is studied.

5.2. Experimental Results and Validation on the Synthetic Datasets 63

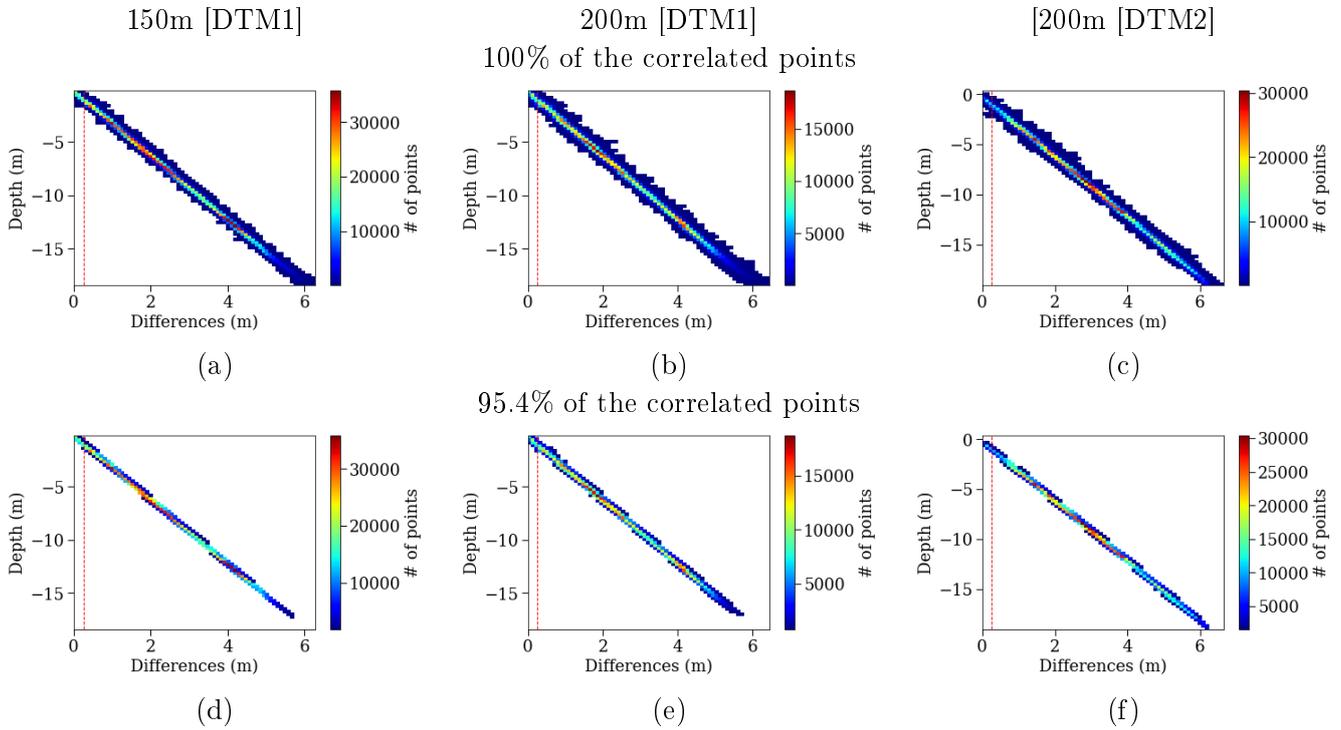


Figure 5.8: The 2D histograms of the differences between the real and the corrected image-based dense point clouds derived from the SfM-MVS for the 150m[DTM1] dataset (a), the 200m[DTM1] dataset (b) and the 200m[DTM2] dataset (c) respectively, in relation to the real depth. Figures (d), (e) and (f) are presenting the differences and the respective point densities of the 95.4% of the correlated points respectively. The dashed lines in red represent the accuracy limits generally accepted for hydrography as introduced by the IHO [Guenther 2000].

Similarly to Figure 5.4, Figure 5.9 serves as an explanatory figure of the 2D histograms that are presented next. There, the 2D histogram of the comparisons performed for the 150m[DTM1][refracted] dataset, which correlates the elevation and the differences between the true and the corrected depths resulted by Method 1, is accompanied with two 1D histograms; the one on top is the histogram of the differences for the elevation of 6m (a section on the 2D histogram parallel to the X axis) while the one on the right is the histogram of the elevations for the difference of 0m (a section on the 2D histogram parallel to the Y axis), which is the average elevation difference for this dataset.

Similarly to Figure 5.4, also in Figure 5.9, the histogram of the elevations is directly related to the DTM used for the synthetic data generation, and thus, it would be different for the DTM1 and DTM2. On the contrary, the histogram of the differences is described by a Gaussian distribution for all the tests performed (Figure 5.10). Figure 5.10 presents the 2D histograms of the differences between the true depths $Z = f(X, Y)$ and the corrected depths produced from the predicted model trained on each synthetic dataset in relation to the true depth. The dashed lines in

red represent the accuracy limits generally accepted for hydrography as introduced by the IHO [Guenther 2000]. In the first, third and fifth rows of Figure 5.10, the depth (Z) - Differences 2D histograms of the synthetic datasets and the respective point densities of the 100% of the points are presented while in the second, fourth and sixth rows of Figure 5.10, the depth (Z) - Differences 2D histograms of the 95.4% of the points with the higher density are presented. It is observed that the amount of the points that surround the core of the 2D histograms and exceeds the IHO limits, is insignificant, representing only the 4.6% of the points.

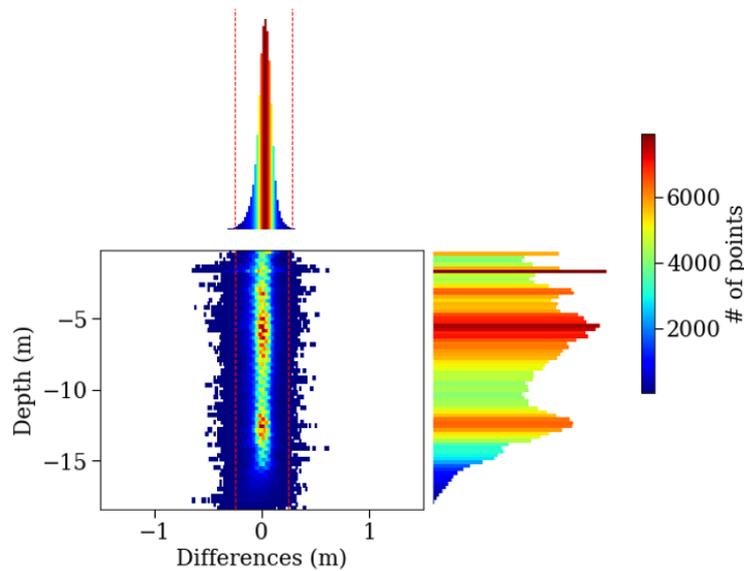


Figure 5.9: An example 2D histogram of the differences between the true depths and the corrected depths produced for the predicted model, together with explanatory sections (1D histograms) along X and Y axis.

Contrary to Figure 5.5, for all cases presented in Figure 5.10, the majority of the points are presenting mean differences close to 0m. These results, confirm that, as predicted in Section 5.2, the proposed method for correcting the geometric effects of refraction based on trained SVR models, can absorb successfully the systematic effects that are caused by the absence of Ground Control Points (GCPs) in the synthetic datasets. On the contrary, the noise in the elevations of the point clouds, introduced mainly by the SfM-MVS process is still present in Figure 5.10 and, as expected, is more intense due to the refraction effect (see also 2.2.4 and Subsection 5.3.5). However, this noise is limited to the 4.6% of the less dense bins and consequently the 4.6% of the compared points.

One more important observation on the results presented in Figure 5.10 is that in some of the 2D histograms, an inclination of the high density core is detected, deviating from the normal. This inclination seems to be directly related to the depth since as the depth increases, the high density core is moving away from the

5.2. Experimental Results and Validation on the Synthetic Datasets 65

vertical of 0m differences. This effect is more evident in the middle column, where the models trained over the 200m[DTM1][95%] are used and evaluated. However, even in these cases, in which the synthetic data are examined, this deviation is no more than 0.03m in the depth of 15m. When real datasets are under testing, this deviation is expected to be increased, especially in depths of more than 10 m, since errors and limitations caused by the visibility restrictions due to the depth and the water column characteristics would be introduced.

Table 5.3 presents the results of each one of the 9 tests performed with every detail. Both in Figure 5.10 and Table 5.3, an impressive improvement in depth determination accuracy is observed. More specifically, for all synthetic test sites used, the initial mean distance being in the range of 2.48m to 2.96m was reduced to the range of 0m to 0.027m, the initial standard deviation being in the range of 1.47m to 1.58m was reduced to the range of 0.07m to 0.11m while the initial *RMSEs* being in the range of 2.87m to 3.34m were reduced to the range of 0.07m to 0.11m. It is also important to note that the large distances between the clouds observed in Figure 5.8 disappeared and the vast majority of the points (more than 95.4%) that represent the relation between the depth (Z) and the calculated differences are within IHO limits. This improvement was observed in every test performed, proving that the proposed point cloud correction method achieves a great reduction of the geometric effects caused by the refraction to the bottom point clouds, eliminating the errors in depth determination.

To compare the achieved results with the differences found by the comparison between the uncorrected and the true depths (Figure 5.8), it is reminded that at the depth of 15m the difference reaches the 5.5m which is the 36.7% of the depth. On the contrary, when comparing the corrected depths resulted by the proposed method with the true depths, differences at the depth of 15m are within the range of -0.20m to 0.20m. Nevertheless, most of them lies within the range of -0.05m to 0.05m, i.e. the 1.33% and 0.33% respectively of the true depth for all the tests performed.

An additional measure to evaluate the predicted models used was the computation of the coefficient R^2 , which is the fitting score and is defined as:

$$R^2 = 1 - \frac{\sum (Z_{\text{true}} - Z_{\text{predicted}})^2}{\sum (Z_{\text{true}} - Z_{\text{true.mean}})^2} \quad (5.2)$$

The best possible score is 1.0 and it may also be negative [Pedregosa 2011]. Z_{true} is the true value of the depth of the points not used for training, while the $Z_{\text{predicted}}$ is the predicted depth for these points, using the model trained on another set of points. As can be seen in Table 5.3, the fitting score achieved for all the trained models indicates that they describe the data in an optimal way touching the best possible score.

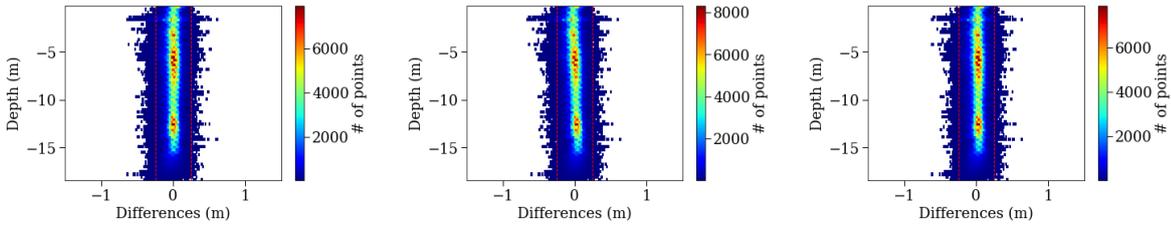
Training Site for Testing on 150m[DTM1][5%]

150m[DTM1][95%]

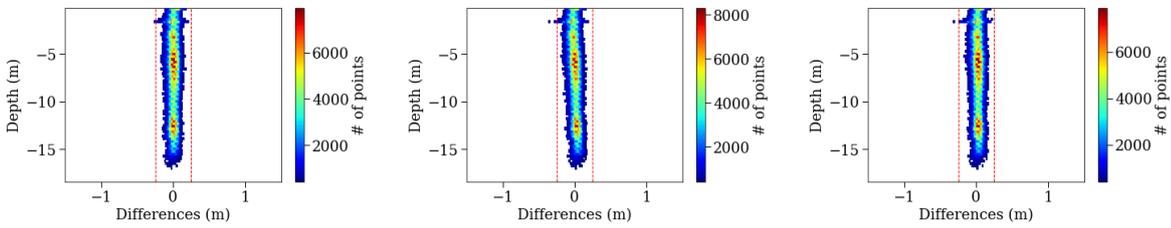
200m[DTM1][95%]

200m[DTM2][95%]

100% of the points

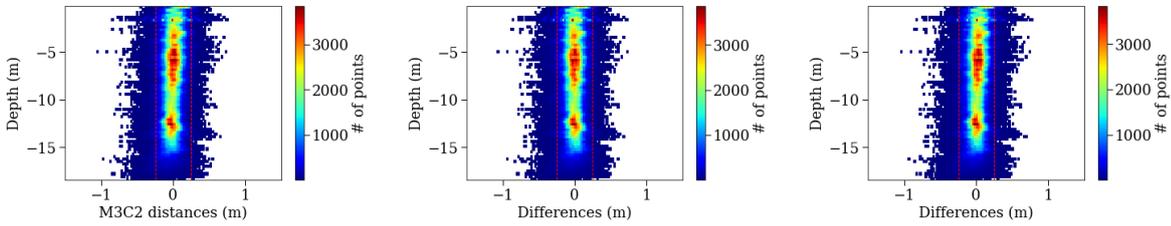


95.4% of the points

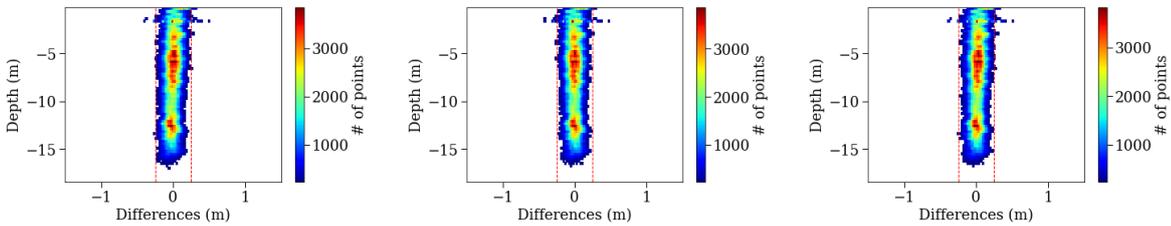


Training Site for Testing on 200m[DTM1][5%]

100% of the points

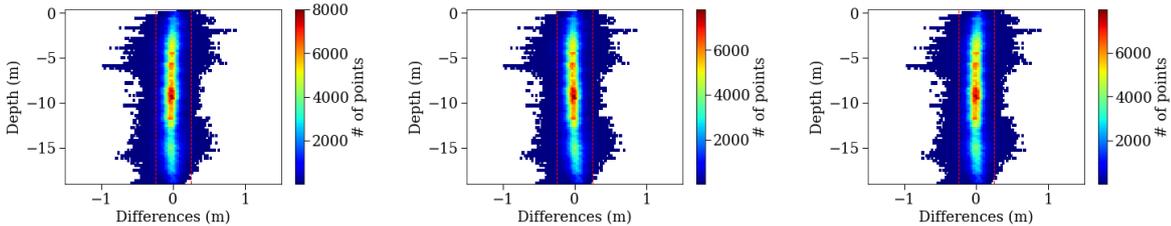


95.4% of the points



Training Site for Testing on 200m[DTM2][5%]

100% of the points



5.2. Experimental Results and Validation on the Synthetic Datasets 67

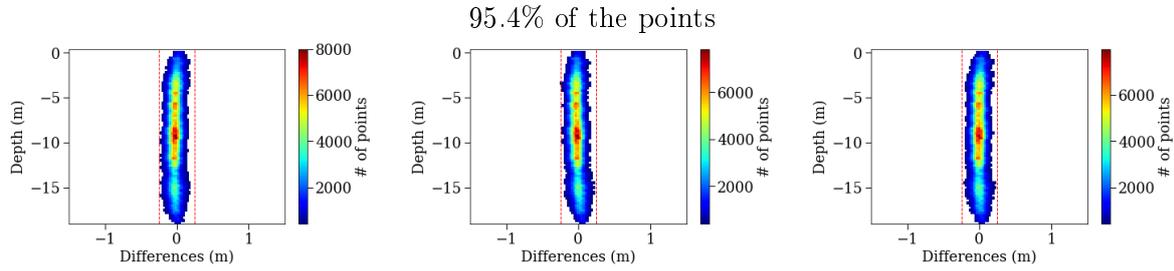


Figure 5.10: The 2D histograms of the differences between the true depth, resulting from the Z function and the corrected point clouds after the application of the proposed method, in relation to the real depth. The red dashed lines represent the accuracy limits generally accepted for hydrography, as introduced by the IHO [Guenther 2000].

Results in Table 5.3 highlight the robustness of the proposed depth correction framework. However, it should be noted that despite the great improvement on the mean distances (\bar{x}) and the $RSMEs$, the standard deviation remains almost equal to the pixel size for each of the synthetic test sites and it is doubled, compared with the experiments performed using the non-refracted images. Through the tests performed using the non-refracted datasets (Figure 5.5), it was found that noise causing these standard deviations are attributed mainly to the noise in the elevations of the point clouds, introduced by the SfM-MVS process. In the tests performed using the refracted synthetic datasets, this noise is magnified due to the refraction effect (see also Subsection 2.2.4 and Subsection 5.3.5). However, the 2D histograms presented in Figure 5.10 ensure that the majority of the differences are concentrated on a narrow area around the difference of 0m.

Dataset [training %]	Training points	Fitting Score	Evaluation Site [testing %]	Max /Min depth of test site	Evaluation points	Uncorrected data			Corrected data			
						\bar{x} [m]	s [m]	$RMSE_Z$ [m]	\bar{x} [m]	s [m]	$RMSE_Z$ [m]	
150m [DTM1] [5%]	129.552	0.9997	150m [DTM1] [95%]	18.7/0	2.461.488	2.48	1.47	2.874	0.006	0.070	0.073	
150m [DTM1] [5%]	129.552	0.9993	200m [DTM1] [95%]	18.7/0	1.409.673	2.49	1.49	2.883	-0.001	0.108	0.111	
150m [DTM1] [5%]	129.552	0.9996	200m [DTM2] [95%]	19.05/0	2.097.713	2.96	1.54	3.337	-0.022	0.090	0.093	
200m [DTM1] [5%]	74.193	0.9997	150m [DTM1] [95%]	18.7/0	2.461.488	2.48	1.47	2.874	0.007	0.075	0.076	
200m [DTM1] [5%]	74.193	0.9994	200m [DTM1] [95%]	18.7/0	1.409.673	2.49	1.49	2.883	0	0.109	0.109	
200m [DTM1] [5%]	74.193	0.9996	200m [DTM2] [95%]	19.05/0	2.097.713	2.96	1.54	3.337	-0.014	0.090	0.093	
200m [DTM2] [5%]	110.405	0.9996	200m [DTM2] [95%]	19.05/0	2.097.713	2.96	1.54	3.337	0.001	0.090	0.077	
200m [DTM2] [5%]	110.405	0.9997	150m [DTM1] [95%]	18.7/0	2.097.713	2.48	1.47	2.874	0.027	0.073	0.111	
200m [DTM2] [5%]	110.405	0.9994	200m [DTM1] [95%]	18.7/0	1.409.673	2.49	1.49	2.883	0.019	0.109	0.090	
Overall Average						2.643	1.500	3.031	0.003	0.090	0.093	
						s	0.224	0.029	0.216	0.014	0.015	0.014

Table 5.3: The results of the comparisons between the predicted models for all tests performed.

5.2.2 Is a ratio-based depth determination better?

The already generated synthetic data provided the opportunity to demonstrate in a nutshell that sophisticated solutions such as the ones presented in this thesis are necessary for addressing the depth determination problem in through-water cases, compared to simple solutions such as the multiplication of the apparent depths with a standard coefficient. In [Chirayath 2019], authors report that for flat fluid sea surface and nadir camera viewpoint the apparent depth equals to the 75.2% of the real depth or that the real depth equals to the 133% of the apparent depth, considering that the refractive index of the water is 1.33. A more sophisticated approach is followed in [Woodget 2015] on which the apparent depths are calculated by the refractive index of the water (1.34 in their approach) only for rays where the incidence and refraction angles are less than 10° . Since the correction using directly the refractive index of the water, proposed in both the methods in [Woodget 2015] and in [Chirayath 2019] apply only in the underwater cases (see [Lavest 2000, Agrafiotis 2015]), the coefficient used here, is calculated as the average ratio of the true depths to the respective apparent depths, being much closer to reality.

Figure 5.11 presents the 2D histograms of the differences between the true depths $Z = f(X, Y)$ and the corrected depths produced after the multiplication of the coefficient to the apparent depths in relation to the true depth. To highlight the ineffectiveness of this simple approach based just on the coefficient, the apparent depths of each test site are corrected with the coefficient computed as the average of the ratios of the true to the apparent depths of the same site, representing the best possible results achieved by this approach.

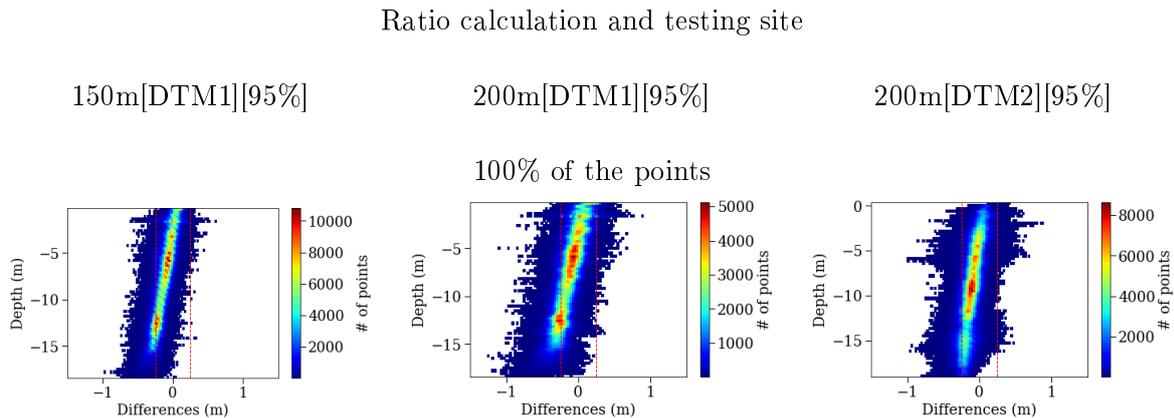


Figure 5.11: The 2D histograms of the differences between the true depths and the corrected depths produced after the multiplication of the coefficient with the apparent depths in relation to the true depth. For the presented cases the ratio is calculated on the same test site on which testing is performed. The red dashed lines represent the accuracy limits generally accepted for hydrography, as introduced by the IHO [Guenther 2000].

In Figure 5.12, differences between the true depths and the corrected depths produced after the implementation of the SVR based method presented in Section 4.1, the simple average coefficient method and the method presented in [Chirayath 2019] are presented. There, in the left column, the 2D histogram of the differences between the true depths $Z = f(X, Y)$ and the corrected depths of the 200m[DTM2] synthetic site resulted by the SVR model trained on the 200m[DTM1] synthetic site are presented. In the middle column, the 2D histogram of the differences between the true depths $Z = f(X, Y)$ and the depths resulted after the multiplication with the average ratio calculated on the 200m[DTM1] synthetic site are presented for the 200m[DTM2] synthetic site.

Both the above examples are intending to demonstrate the performance of each method when the training and the coefficient calculation respectively are not performed on the testing site but on a different one. Finally, in the right column, the 2D histogram of the differences between the ground truth depths $Z = f(X, Y)$ and the depths resulted after approach presented in [Chirayath 2019] are presented in order to highlight the magnitude of the remaining errors compared with the rest of the approaches. The dashed lines in red represent the accuracy limits generally accepted for hydrography as introduced by the IHO [Guenther 2000].

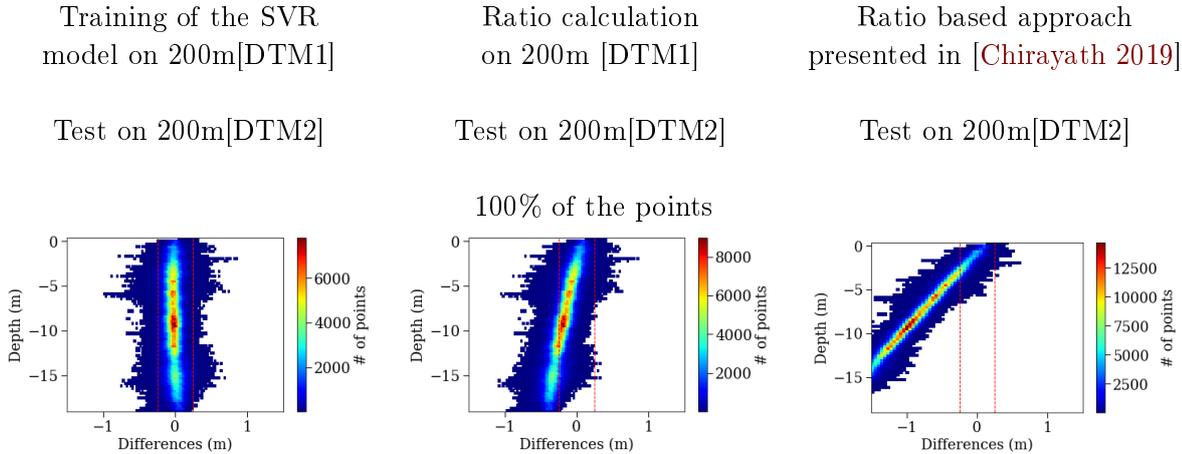


Figure 5.12: The 2D histograms of the differences of a cross site approach; In the left column, the 2D histogram of the differences between the true depths and the corrected depths of the 200m[DTM2] synthetic site produced by the proposed method trained on the 200m[DTM1] synthetic site is presented. In middle right column, the respective histogram of differences of the derived depths after the multiplication with the average ratio calculated on the 200m[DTM1] synthetic site. In the right column, the respective histogram of differences of the derived depths after following the method presented in [Chirayath 2019] are presented. The red dashed lines represent the accuracy limits generally accepted for hydrography, as introduced by the IHO [Guenther 2000].

In both Figure 5.11 and Figure 5.12, the results of the average coefficient based

5.2. Experimental Results and Validation on the Synthetic Datasets 71

method exceed the IHO limits as the depth increases and a large inclination of the histograms is observed, proving that the average ratio is not sufficient to deliver precise results in all depths. Even more essential for arguing against this simple method are the histograms presented in Figure 5.12. There, it is proved that when the average coefficient is calculated on one test site, its exploitation over a different test site produces way larger errors in depth determination, since the differences of the majority of the points are exceeding the IHO limits. However, this is not the case for the accurate results produced after the proposed approach based on the SVR. In Table 5.4 the respective metrics for the presented tests are presented. Regarding the results of the method presented in [Chirayath 2019], these are unacceptable for depths of more than 3m and no further comparisons and statistics will be given regarding this method.

In Table 5.4, it can be observed that when the coefficient is used to correct the data, the mean distance is increased by 12903% to 403704%, the standard deviation is increased by 129% to 184% while the $RMSE_Z$ is increased by 180% to 285%. Especially when observing the results after the cross-site testing, it is clear once more that this solution is not offering the same levels of accuracy as the proposed one based on the SVR. Most importantly, the coefficient-based solution, is not capable of generalizing over different test sites and thus it is proved inadequate for fulfilling the objectives of the approaches presented here.

Depending on the characteristics of the site, for correcting the refraction till the depth of 1m to 2m, a coefficient based on the ratios calculated for this specific test site could be used, however, even in this case, the expected errors in depth determination would be in the scale of 10% to 20% of the real depth, i.e. 0.1m to 0.2m in the depth of 1m.

Nevertheless, it should be noted that for the experiments performed in this section, the average ratios were calculated using 2.461.488, 1.409.673 and 2.097.713 ratios respectively, an amount of true data not available in most of the real world cases. Moreover, even as reported before, the coefficient could be used by non-experts to correct the depth till 1m to 2m, this is the area where the ratio values span the most leading to incorrect calculations when there is not available a huge amount of ratios. Thus, it is not considered realistic to calculate a representative accurate coefficient, especially when only some dozens or hundreds of points are measured with conventional surveying methods. A solution to the refraction correction using a coefficient is presented in [Woodget 2015] and results proving it's inadequateness for providing accurate results over real world test sites are presented in Section 5.5.

Dataset [training %]	Evaluation Site [testing %]	Max /Min depth of test site	Evaluation points	Uncorrected data			SVR corrected data			Ratio corrected data		
				\bar{x} [m]	s [m]	$RMSE_Z$ [m]	\bar{x} [m]	s [m]	$RMSE_Z$ [m]	\bar{x} [m]	s [m]	$RMSE_Z$ [m]
150m [DTM1]	150m [DTM1]	18.7/0	2.461.488	2.48	1.47	2.874	0.006	0.070	0.073	-0.088	0.121	0.149
200m [DTM1]	200m [DTM1]	18.7/0	1.409.673	2.49	1.49	2.883	0	0.109	0.109	-0.108	0.164	0.197
200m [DTM2]	200m [DTM2]	19.05/0	2.097.713	2.96	1.54	3.337	0.001	0.090	0.077	-0.079	0.116	0.140
200m [DTM1]	200m [DTM2]	19.05/0	2.097.713	2.96	1.54	3.337	-0.014	0.090	0.093	-0.181	0.166	0.247
Overall Average				2.643	1.500	3.031	-0.002	0.090	0.088	-0.114	0.142	0.183
s				0.224	0.029	0.216	0.009	0.016	0.016	0.046	0.027	0.049

Table 5.4: The results of the comparisons between the true depth and the uncorrected depth, the depths derived by the implementation of the SVR based approach and the ratio based approach.

5.2. Experimental Results and Validation on the Synthetic Datasets 73

5.2.3 Summary

Main aim of the work described in Section 5.1 and Section 5.2 was to test and evaluate the first of the two proposed methods in a controlled environment and confirm the assumption of the linearity of the relation of the real and the apparent depths (Figure 4.3). Moreover, an important additional aim was to train SVR models, to be used also for the correction of the depths over the real-world datasets.

To achieve this goal, initially, three synthetic datasets were generated and processed accordingly. Compared with the real world data which will be processed in the next section, the advantage of the synthetic data is the accuracy and the reliability of the depth information and the exact knowledge of exterior and interior orientations of the cameras used. More importantly, when it comes to seabed imaging, errors and limitations caused by the visibility restrictions due to the depth, especially in depths more than 10m, as well as errors introduced by the wavy surface are excluded, delivering unaffected results regarding the proposed method. Also, by using a mathematical function to generate and describe the DTMs, incompatibilities and errors that might be transferred to the solution by ground truth data generated using other systems such as LiDAR or echosounders are avoided, leading to independent and objective results.

Following the synthetic data generation, three SVR models -one for each synthetic dataset- were trained by correlating the known depths calculated by the Z function and the apparent depths of the point clouds derived from the conventional SfM-MVS procedures using the refracted synthetic aerial imagery.

Experimental results over the three test areas which were characterized by different flying heights, different depths, and different seabed anaglyph, along with the quantitative validation performed, indicated the high potential of the developed method (Method 1). Nine out of nine testing approaches (Table 5.3), i.e. 100% of the tests, proved that the proposed method for correcting the depths of the SfM-MVS derived point clouds meets and exceeds the accuracy standards generally accepted for hydrography, as introduced by the IHO [Guenther 2000]. It is reminded that these accuracy standards report that the vertical accuracy requirement for shallow water hydrography can be set as a total of $\pm 0.25\text{m}$ (one sigma) from all sources, including tides [Guenther 2000]. Moreover, considering IHO's white paper on the regulations for international charts and charts specifications, the proposed method is qualified to the maximum Zone of Confidence (ZOC) level A1 [IHO 2019].

In addition to the above, the performed experiments in the controlled environment of the synthetic datasets produced a number of important outcomes regarding the systematic errors and the noise in the elevations caused by the SfM-MVS processing in the case of aerial data acquisition campaigns and especially the noise that is added to the Z determination due to the refraction effect. In all the tests performed using the non-refracted datasets, it was noticed that the average elevation differences are increasing as the flying height and the complexity of the DTM increases, something reported also in the literature [Smith 2015, Nesbit 2019]. Moreover, the SfM-MVS processing of the same datasets resulted in underestimating the

elevations of the synthetic test sites. However, in the experiments performed using the refracted datasets, this systematic effect was eliminated, since it was proved that it can be accommodated by the SVR model. This systematic effect, which is almost equal to the GSD size, is accompanied by larger differences, which are not systematic and they are scattered almost equally on either side of the mean value, forming the Gaussian distribution already reported and shown in Figure 5.4 and Figure 5.9. These effects, reported also in [Smith 2015, Nesbit 2019], are present in most of the parallel-axis (nadir) image blocks and they are justifying the calculated standard deviations of Table 5.2 and Table 5.3 since they are indicating the noise in the elevations of the point cloud, introduced mainly by the SfM-MVS process. However, the use of oblique images for seabed mapping applications is not suggested since this would asymmetrically increase the incidence angles and consequently the refraction and the noise in the point clouds instead of reducing it.

This noise is not related to the proposed method and cannot be avoided by any of the proposed approaches presented in the literature, since it seems to be a fundamental feature of the SfM-MVS process and as expected, it is more intense when refraction is added to the images (Figure 5.10) (see also Subsection 2.2.4 and Subsection 5.3.5). However, this noise is limited to a mere 4.6% of the less dense bins and consequently the 4.6% of the compared points. Regarding the additional noise due of the refraction, it is expected to be reduced when the results of Method 2 will be used (see Section 5.5).

Additionally to the above, experiments performed, proved that simpler methods based on a coefficient equal to the refraction index or an average coefficient calculated by the ratio of the real and apparent depths are producing less accurate results and thus they should be avoided.

To conclude, although the proposed Method 1, together with the refraction, has to deal also with external errors due to the SfM-MVS processing, it managed to achieve impressive results in the synthetic datasets, by reducing the mean distances between the ground truth and the apparent depths almost to 0m. As a typical example it is reported that the initial mean difference of 5.5m, i.e. the 36.7% of the true depth, between the uncorrected and the true depths at the depth of 15m, was reduced to the range of -0.05m to 0.05m, i.e. the 0.33% of the true depth in the vast majority of the points. Even in the worst cases were the standard deviation is added to the mean value (1 sigma), the achieved accuracy meets and exceeds the accuracy standards generally accepted for hydrography, as introduced by the IHO [Guenther 2000] and [IHO 2019].

5.3 Real World Test Datasets

Following the training, testing and evaluation of the first of the two proposed methods using the synthetically generated data, both the proposed methods have been applied and tested with five different real world test areas with datasets acquired for operational shallow bathymetry mapping tasks but under optimal sea state and

water visibility conditions. The real world data used for this purpose consist of three shallow water areas in Cyprus and two in Greece, representing typical examples of seabed and water column characteristics in the Eastern Mediterranean Sea (Figure 5.13). The application of these methods to the real world data as well as the transfer of the learning approach between the synthetic and the real world data, confirmed their robustness and applicability and indicated the generalization potential of the two proposed methods.

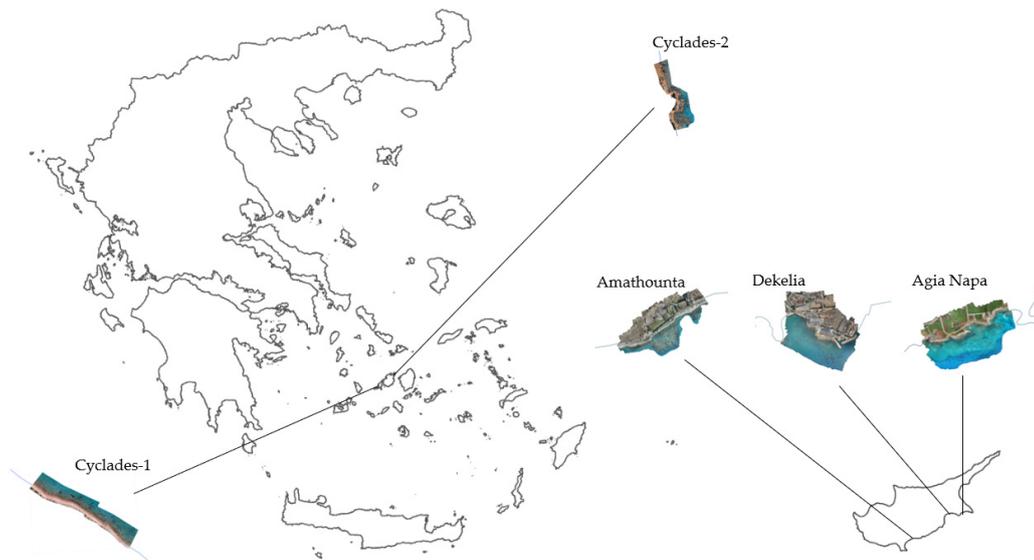


Figure 5.13: The five different test areas used.

5.3.1 Datasets Description

In the next paragraphs, the five different datasets are described together with the necessary technical details about data acquisition, SfM-MVS processing, georeferencing and achieved accuracies. Additionally, indicative images for each test area are given, demonstrating the different seabed characteristics, such as rock formations, sand and seagrass, the state of the water surface and the water visibility.

5.3.1.1 Agia Napa Test Area

The first area used is in Agia Napa, Cyprus, where the seabed reaches the depth of 14.8m. The flight was executed with a Swinglet CAM fixed-wing UAV with a Canon IXUS 220HS camera having 4.3mm focal length, $1.55\mu\text{m}$ pixel size and 4000×3000 pixels format. In total 383 images were acquired, from an average flying height of approximately 209m, resulting to a 6.3cm average GSD (Ground Sampling Distance). Figure 5.14 presents three indicative images from the test area, demonstrating the different seabed characteristics (rock formations, sand and seagrass) and the water

visibility. 40 control points located only on land, and measured with RTK accuracy ($\pm 0.03\text{m}$) were used in total. The georeferencing of the photogrammetric block was realized with an *RMSE* of 5.03, 4.74, 7.36 cm in *X, Y, Z* respectively and average reprojection error of 1.11 pixels after the adjustment.



Figure 5.14: Indicative images from the Agia Napa test area. Depth decreases from the left to the right image.

5.3.1.2 Amathounta Test Area

The second test area used is in Amathounta, Cyprus, where the seabed reaches a maximum depth of 5.57m. The flight here was executed with the same UAV as in the case of Agia Napa. A total of 182 images were acquired, from an average flying height of 103 m, resulting to 3.3cm average GSD. Figure 5.15 presents three indicative images from the test area, demonstrating the different seabed characteristics (rock formations, sand, and seagrass) and the water visibility. 29 control points measured with RTK accuracy were used to georeference the photogrammetric block, located only on land, resulting to an adjustment with *RMSEs* of 2.77, 3.33, 4.57 cm in *X, Y, Z* respectively and reprojection error of 0.645 pixel.



Figure 5.15: Indicative images from the Amathounta test area. Depth decreases from the left to the right image.

5.3.1.3 Dekelia Test Area

The third test area used is in Dekelia, Cyprus, where the seabed reaches a depth of 10.1m. The flight here was executed again with the same UAV as in the previous cases. In total 78 images were acquired, from an average flying height of 188 m,

resulting to 5.9 cm average GSD. Figure 5.16 presents three indicative images from the test area, demonstrating the different seabed characteristics (rock formations, sand, and seagrass) and the water visibility. For georeferencing the image block 17 control points were used, located only on land, resulting to *RMSEs* of 3.30, 3.70, and 3.90 cm in *X*, *Y*, and *Z*, respectively, and a reprojection error of 0.645 px in the bundle adjustment of the images.



Figure 5.16: Indicative images from the Dekelia test area. Depth decreases from the left image to the right image.

5.3.1.4 Cyclades-1 Test Area

The fourth area used is in Cyclades, Greece, where the seabed reaches a depth of 6.9m. The flight was executed with a Phantom 4 UAV with a FC330 camera having 3.61mm focal length, 1.56 μ m pixel size and 4000 x 3000 pixels format. In total 449 images were acquired, from three different average flying heights of approximately 88m, 70m and 35m with 0.038m, 0.030m and 0.015m respective GSDs. In total 449 images were acquired, from three different average flying heights of approximately 88m, 70m and 35m with 0.038m, 0.030m and 0.015m respective GSDs. Figure 5.17 presents three indicative images from the test area. Images are depicting the same area of the site from the three different flight heights. 14 control points in total, located only on land, were measured with RTK accuracy. The georeferencing of the photogrammetric block was realized with *RMSE* of 1.45, 1.03, 1.88 cm in *X*, *Y*, *Z* respectively and average reprojection error of 0.28 pixels after the adjustment.

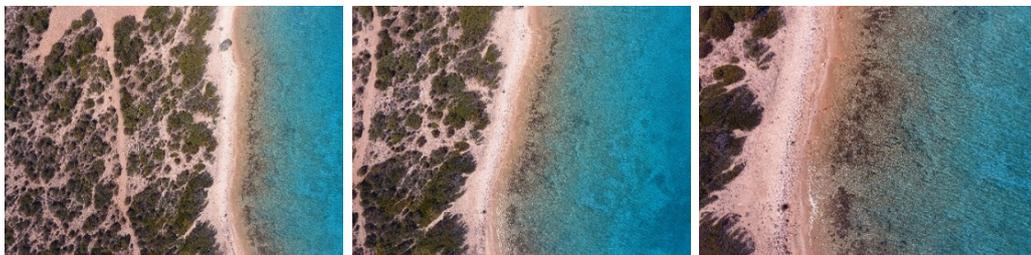


Figure 5.17: Indicative images from the Cyclades-1 test area. Images are depicting the same area of the site captured from 88m (left), 70m (center) and 35m (right).

5.3.1.5 Cyclades-2 Test Area

The fifth area used is also in Cyclades, Greece, where the seabed reaches a depth of 4.05m. The flight was executed with the same Phantom 4 UAV with a FC330 camera. In total 203 images were acquired, from two different average flying heights of approximately 75m and 33m with 0.032m and 0.014m respective GSDs. Figure 5.18 presents three indicative images from the test area. The left and the center images are depicting the same sandy area of the site from the two different flying heights, while the right image demonstrates the rocky formations of the seabed. 11 control points measured with RTK accuracy were used in total, located only on land. The georeferencing of the photogrammetric block was realized with *RMSE* (Root Mean Square Error) of 2.03, 1.44, 2.13 cm in *X, Y, Z* respectively and average reprojection error of 0.30 pixels after the adjustment.



Figure 5.18: Indicative images from the Cyclades-2 test area. Left and center images are depicting the same area of the site captured from 75m and 33m respectively, while the image on the right depicts the rocky formations of the seabed and the shore.

Table 5.5 presents the flights and the SfM-MVS processing details of the five different test areas. It may be noticed that the five areas have different average flying heights, indicating that the suggested solution is not limited to specific flying heights or indeed GSD's. The common system used for measuring the GCPs for the test sites located in Cyprus is the Cyprus Geodetic Reference System (CGRS) 1993 while for the test sites located in Greece, the Greek Geodetic Reference System (GGRS) 1987. The true data presented in Subsection 5.3.6 and used for training and testing the presented methods are in the same reference systems too. As such, projection errors and scale factors are not involved in these processes.

Test site	Amathounta	Agia Napa	Dekelia	Cyclades-1	Cyclades-2
# Images	182	383	78	449	203
Control points used	29	40	17	14	11
Average (Avg.) flying height [m]	103	209	188	88/70/35	75/33
Avg. Base-to-Height (B/H) ratio along strip	0.39	0.35	0.32	0.28/0.31/0.28	0.24/0.22
Avg. Base-to-Height (B/H) ratio across strip	0.66	0.62	0.38	0.5/0.47/0.46	0.42/0.46
Avg. forward overlap	65%	69%	70%	84%/82%/84%	86%/87%
Avg. side overlap	54%	57%	73%	62%/64%/64%	68%/64%
Image footprint on the ground [m]	149 x 111	301 x 226	271 x 203	152 x 114/ 121 x 91/ 61 x 45	130 x 97/ 57 x 43
GSD [m]	0.033	0.063	0.059	0.038/ 0.030/ 0.015	0.032/ 0.014
GCPs $RMSE_X$ [m]	0.028	0.050	0.033	1.45	2.03
GCPs $RMSE_Y$ [m]	0.033	0.047	0.037	1.03	1.44
GCPs $RMSE_Z$ [m]	0.046	0.074	0.039	1.88	2.13
Reprojection error on all points [pix]	0.645	1.106	0.717	1.12	0.86
Reprojection error on control points [pix]	1.48	0.76	0.77	0.28	0.30
Pixel size [μm]	1.55	1.55	1.55	1.56	1.56
Total # of tie points of the uncorrected images	28.5K	135K	71K	186K	72K
Nominal camera constant c [pix]	2774.19	2774.19	2774.19	2314.10	2314.10
Adjusted camera constant c of the blocks of the uncorrected images [pix]	2827.05	2852.34	2831.66	2352.23	2334.39

Table 5.5: Indicative parts of the images containing (a) waves breaking on the shore, (b) turbid waters, (c) sun glint, (d) caustics, (e) seabed with texture and (f) seabed without texture (except of the lower left area of the image).

5.3.2 Factors affecting the low altitude aerial imagery over shallow waters

As with any method for bottom mapping using low-altitude aerial optical (RGB) imagery, the implementation of the proposed method in real world cases is affected by factors related to the sea conditions, the visibility and the presence of texture in the seabed (Figure 5.19).

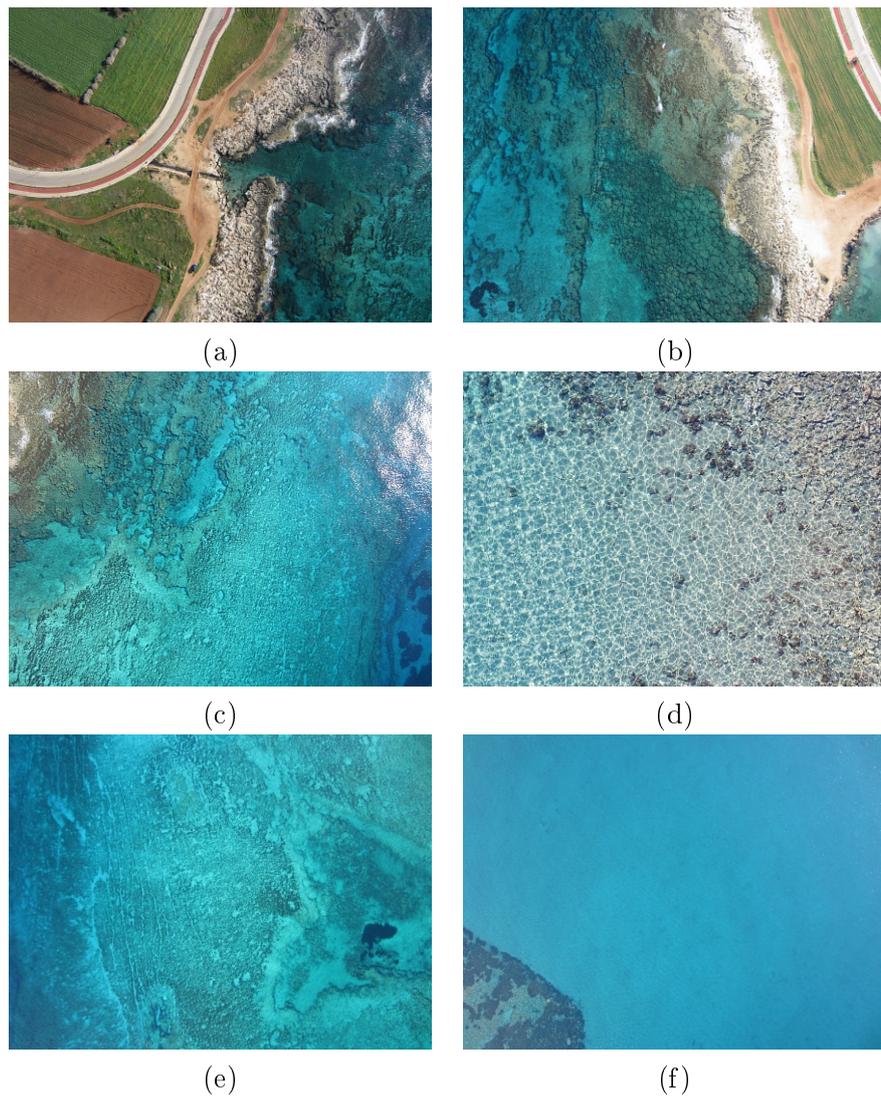


Figure 5.19: The five aerial UAV campaigns, the derived datasets and related information.

Calm water surface. To achieve accurate bathymetric results, water surface needs to be as flat as possible at the time of image acquisition, to ensure optimal bottom visibility and satisfy the geometric assumption of flat-water surface. In case of a wavy surface (5.19a), errors will be introduced [Okamoto 1982, Georgopoulos 2012,

Agrafiotis 2019a, Skarlatos 2018, Fryer 1985, Harwin 2015].

Turbidity. Obviously, water turbidity and water visibility are additional constraints (Figure 5.19b). According to [Collins 1979], Secchi depths near the test sites of Cyprus are between 20-30 meters. In the Cyclades test areas in Greece, Secchi depths are in the same range according to the Secchi depths archive of the Hellenic Navy Hydrographic Service (HNHS).

Sun glint. Sun glint, which depends on the sun angle, may cause direct reflections on the water surface and create annoying bright spots on images (Figure 5.19c), therefore the sun elevation should ideally be low during image acquisition (less than 30 degrees).

Caustics. Caustics (Figure 5.19d) may also appear in the aerial imagery, especially when the bottom of the shallower areas (less than 3m depth) is imaged having a sun angle more than 30 degrees over the horizon and small waves also exist. However, this phenomenon is eliminated when the GSD size is equal or more than the width of the caustics on the bottom and no further actions should be taken in these cases.

Texture. Just like in any automated photogrammetric project, the bottom must possess texture or some random pattern (Figure 5.19e) to ensure point detection and matching. This means that automated photogrammetric bathymetry might fail in sandy or extensively seagrass-covered bottoms (Figure 5.19f). Results would be acceptable even with a less dense point cloud due to reduced key-point detection and matching difficulties, because normally a sandy bottom does not present any abrupt height differences or detailed forms.

5.3.3 Ground Control Points measurements

Due to the geometry of the blocks and the shoreline geography, there is a deficiency in respect to GCPs distributions over the aerial blocks. While it is recommended to fully surround the area of interest with GCPs, this is cannot be the case in coastal mapping, as control points are usually measured only on land, and not on the bottom. In these cases, it is crucial for the GCPs to be placed also far from the shoreline, in order to ensure a proper self-calibration, having GCPs scattered all over the images and to avoid large depth errors at the farthest parts of the block. In general, a minimum number of 10-15 GCPs are required for a reliable self-calibration procedure [Harwin 2015]. The problem may be solved if control points on the bottom are acquired somehow, otherwise, appropriate flight design and additional cross-flight lines may reduce the effect. However, this is not possible for depths more than 3m – 4m. In Figure 5.20 the GCP locations for Amathounta, Agia Napa, Dekelia, Cyclades-1 and Cyclades-2 are presented respectively.

5.3.4 Initial Dense Point Clouds Generation

For the implementation of the proposed refraction correction methods towards accurate bathymetry mapping using aerial imagery, an initial SfM-MVS step should

be performed. To that direction, a standard SfM-MVS approach can be followed in order to obtain the required data for applying the proposed refraction correction methods, i.e. the interior and exterior orientation of the cameras and the initial dense point cloud (Figure 5.20).

For the SfM-MVS step, the implementation of a particular software does not affect the quality of the results and it should be noted that they could be produced in a similar way using any commercial or open source automated photogrammetric software. For the approaches presented here, Agisoft's Metashape, a commercial software applying SfM-MVS is used to generate the required data. Regarding the interior orientation of the cameras, a self-calibration approach is followed and thus accurate GCPs' measurements are of really high importance.

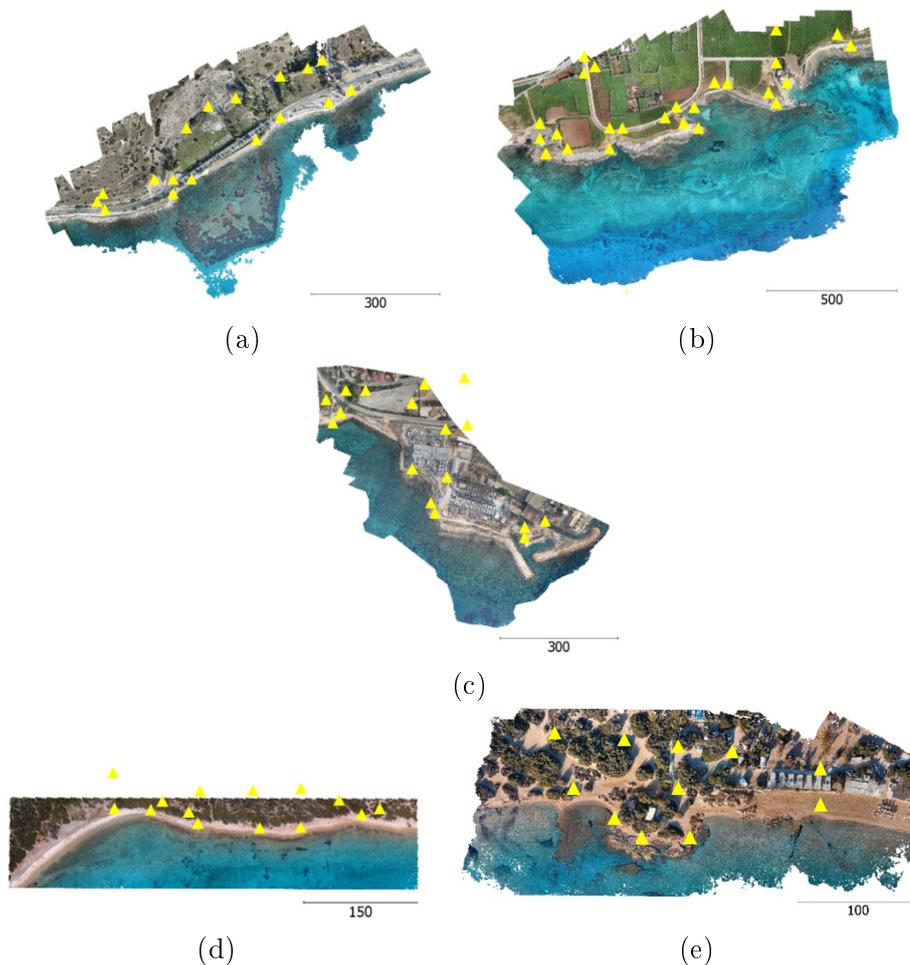


Figure 5.20: The initial dense point clouds for Amathounta (a), Agia Napa (b), Dekelia (c), Cyclades-1 (d) and Cyclades-2 (e) test sites. The yellow triangles represent the GCPs positions. Scale bars are in meters.

5.3.5 The influence of the Base-to-Height ratio (B/H) of stereopairs on the apparent depths

According to the literature [Hasegawa 2000] the accuracy of a digital surface model (DSM) increases rapidly in proportion to the B/H ratio at lower values. This happens because in the lower B/H values, even image matching is harder, rays intersect with larger angles, reducing the uncertainty of depth determination. In the eight different flights conducted for acquiring imagery data over the five test sites, the values of the B/H ratio along the strip were between 0.22 (1:4.6) and 0.39 (1:2.6) due to the forward overlap of 65% - 87%, while the values of the across-strip B/H ratio were between 0.38 (1:2.6) and 0.66 (1:1.5) due to the side overlap of 54% to 73% (Table 5.5).

Considering Snell's Law and the fact that the UAV flights over the five different test sites were planned with similar base-to-height (B/H) ratios, additional data were also created by doubling this ratio for all the test areas. In the initial datasets, image-based point clouds were generated as well, using a different (higher) base-to-height (B/H) ratios. To realize that, one every other image was used across the strip for generating the dense point cloud, thus changing entirely the B/H ratio of the dataset. In order not to compromise the accuracy of the SfM and especially the calibration and triangulation steps, the original B/H ratio was kept only for images containing GCPs on the land. The resultant, along-strip B/H ratio over the water was calculated to be 0.78 (1:1.3) for the Amathounta dataset, 0.70 (1:1.4) for the Agia Napa dataset, 0.64 for the Dekelia dataset, 0.57-0.63-0.57 (1:1.75 - 1:1.59 - 1:1.75) for the Cyclades-1 dataset and 0.47-0.45 (1:2.13 - 1:2.22) for the Cyclades-2 dataset (Table 5.6).

To evaluate the differences between the initial point clouds having the original B/H ratio and the new point clouds having the new B/H ratio, the point clouds were compared using the multiscale model to model cloud comparison (M3C2) [Lague 2013] in Cloud Compare freeware [Girardeau-Montaut 2015]. This also helped to demonstrate the changes and the differences that were applied by the altered B/H ratio.

The results of this comparison are presented in Table Table 5.6 and indicate a mean difference equal or less than the GSD for each of the compared sites, and as such, it was considered insignificant for the purposes of the work presented here. By carefully examining the spatial patterns of the larger remaining errors on the compared point clouds (Figure 5.21), it was established that these were mainly caused due to the noise of the point clouds, since less images are used in the dense image matching process for the underwater area. Considering the standard deviations, these are in the range of every test performed in this thesis, including the synthetic data (see Sections 5.2 and 5.4.1.3).

Taking into consideration the above, it was inferred that the change of the B/H ratio was not affecting the apparent depths of the generated point clouds to a measurable degree, although, according to Snell's Law, they should have been affected, and specifically, they should have presented apparent depths smaller than those of

the points generated using the original B/H ratio. It should be noticed here that the negative sign of the computed Gaussian means in both cases, might suggest this slight decrease on the computed apparent depths of the points. However, this difference is considered insignificant.

Test site	Flight height [m]	GSD [m]	Initial B/H ratio	Altered B/H ratio	\bar{x} [m]	s [m]	$RMSE_Z$ [m]
Amathounta	103	0.033	0.39	0.78	-0.042	0.073	0.084
Agia Napa	209	0.063	0.35	0.70	-0.018	0.053	0.055
Dekelia	188	0.059	0.32	0.64	-0.029	0.115	0.116
Cyclades-1	88	0.038	0.28	0.57			
Cyclades-1	70	0.030	0.31	0.63	-0.007	0.116	0.118
Cyclades-1	35	0.015	0.28	0.57			
Cyclades-2	75	0.032	0.24	0.47			
Cyclades-2	33	0.014	0.22	0.45	0.008	0.113	0.116

Table 5.6: Results of this comparison between the initial dense point clouds and the point clouds generated with the double B/H ratio.

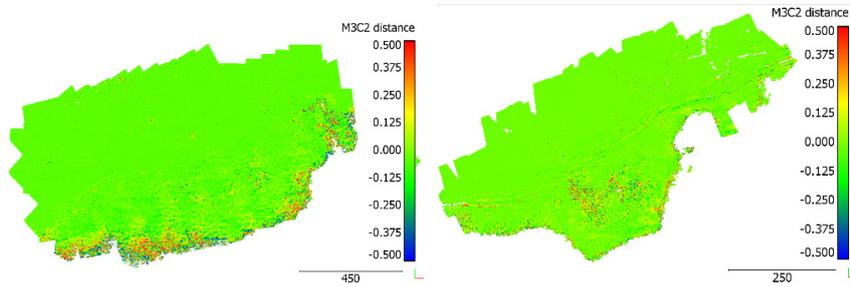


Figure 5.21: The spatial patterns of the remaining errors on the compared point clouds of Agia Napa (a) and Amathounta (b) datasets. M3C2 distances and scale bars are in meters (m).

Additionally to the above presented investigation, dense point clouds were generated using only the imagery of the same flight height in the test sites of which data were collected from more than one flights. To that direction, in the Cyclades-1 test area, one point cloud was generated using the imagery acquired in the average flight height of 88m and a different one was generated using the imagery acquired in average flight height of 70m. Similarly, in the Cyclades-2 test area, one point cloud was generated using the imagery acquired in the average flight height of 75m and a different one was generated using the imagery acquired in average flight height of 33m.

To evaluate the differences between the initial point clouds having the original B/H ratio and the new point clouds having the new B/H ratio, the point clouds

were compared using the multiscale model to model cloud comparison (M3C2) [Lague 2013] in Cloud Compare freeware [Girardeau-Montaut 2015]. In these tests, it is also considered that the estimated differences are not affecting the apparent depths of the generated point clouds to a measurable degree. These outcomes are also supported by the tests and evaluation performed using the synthetic datasets (see Section 5.2).

Test site	Flight height 1 [m]	GSD [m]	Flight height 2 [m]	\bar{x} [m]	s [m]	$RMSE_Z$ [m]
Cyclades-1	88	0.038	35	-0.008	0.130	0.130
Cyclades-1	88	0.038	88/70/35	-0.024	-0.144	0.140
Cyclades-1	70	0.030	88/70/35	-0.001	0.054	0.053
Cyclades-2	75	0.032	33	-0.007	0.116	0.118

Table 5.7: The differences between the initial point clouds generated using all the available imagery and the point clouds generated using imagery acquired in specific altitudes.

These small differences can be explained by the MVS procedure, especially the dense image matching step; a higher B/H ratio leads to larger bases, i.e., distances between the cameras, and this leads to increased percentage of light ray intersections using the image border areas, which present a larger intersection angle with the water surface. According to Snell’s Law, this obviously leads to a stronger refraction effect, meaning that light rays captured in the outer part of the image are highly affected by the refraction effect. Considering the above, dense image matching algorithms [Remondino 2013] aiming to obtain a corresponding pixel for almost every pixel in the image, are facing great difficulties in matching these pixels in the images that are created by capturing these light rays that are highly affected by the refraction effect. Even when they are matched, the generated depth values on the disparity images deviate from the depth values that result from points matched closer to the perpendicular to the water surface, and during the disparity map merging process, they are considered as outliers [Matthies 1989, Okutomi 1993, Ylimäki 2018].

This is also proven by the fact that most of the seabed points of the sparse and the dense point clouds of the test sites, are generated using no more than six images, having their projection centers close to that point, while for the points on the land, about 10 images are used, including images whose projection centers are more distant. Taking into account the above, the following processing and experiments refer only to the initial image-based point clouds, having the original B/H ratio, as reported in Table 5.5.

5.3.6 True depth data

In the real world datasets, LiDAR point clouds of the submerged areas, depths measured with dipping tape and georeferenced with RTK GPS measurements and

Total Station points were used for evaluating the developed methodologies. LiDAR bathymetry is currently considered the "golden standard" for shallow water depths [Skarlatos 2018, Menna 2018]. The point clouds were generated with the Leica HawkEye III (Leica Geosystems AG, Heerbrugg, Switzerland), a deep penetrating bathymetric airborne LiDAR system. This multisensory system includes an up to 500kHz infrared channel (1064nm) for topographic applications, a 10kHz bathymetric channel (532nm) for high accuracy and high data density for deep water and a 35kHz bathymetric channel optimized for shallow water and the transition zone towards the shore [Leica Geosystems 2015]. Even though the specific LiDAR system can, according to the manufacturers, produce point clouds with accuracy of 0.02m in topographic applications, when it comes to bathymetric applications the system's range error is in the order of ± 0.15 m for depths up to 1.5 Secchi depths [Leica Geosystems 2015], similar to other conventional topographic airborne scanners. Dipping tape depth measurements georeferenced with RTK GPS data were available only for the Dekelia test area. In total, 208 points were measured, having a maximum depth of 7.0m and a minimum depth of 0.30m. These points formed the Dekelia (GPS) dataset. Regarding the measurements for the Dekelia (GPS) dataset, a realistic nominal bathymetric accuracy can be estimated to 0.15m, since they might be affected by the boat swelling and errors related with the methodology implementation itself. Regarding the Total Station points, a realistic nominal bathymetric accuracy can be estimated to 0.05m, taking also into account the penetration of the weight of the tape or the pole into the seabed. Table 5.8 presents the details of the true depth data used.

Test site	# points for method 1	# points for method 2	Source	Point density [points/m ²]	Average pulse spacing [m]	Flying height [m]	Nominal Bathymetric Accuracy [m]
Amathounta	6K (1K)	1K	LiDAR	0.4	-	600	0.15
Agia Napa	1.3M (75K)	75K	LiDAR	1.1	1.65	600	0.15
Dekelia	500K	-	LiDAR	1.1	1.65	600	0.15
Dekelia	208	-	Dipping tape and GPS	-	-	-	0.15
Cyclades-1	23	23	Total Station	-	-	-	0.05
Cyclades-2	34	34	Total Station	-	-	-	0.05

Table 5.8: True data specifications. Numbers in parenthesis are for the sparse point clouds

According to the literature [Skinner 2011, Bailly 2010, Fernandez-Diaz 2013, Westfeld 2017], LiDAR bathymetry data can be affected by significant systematic

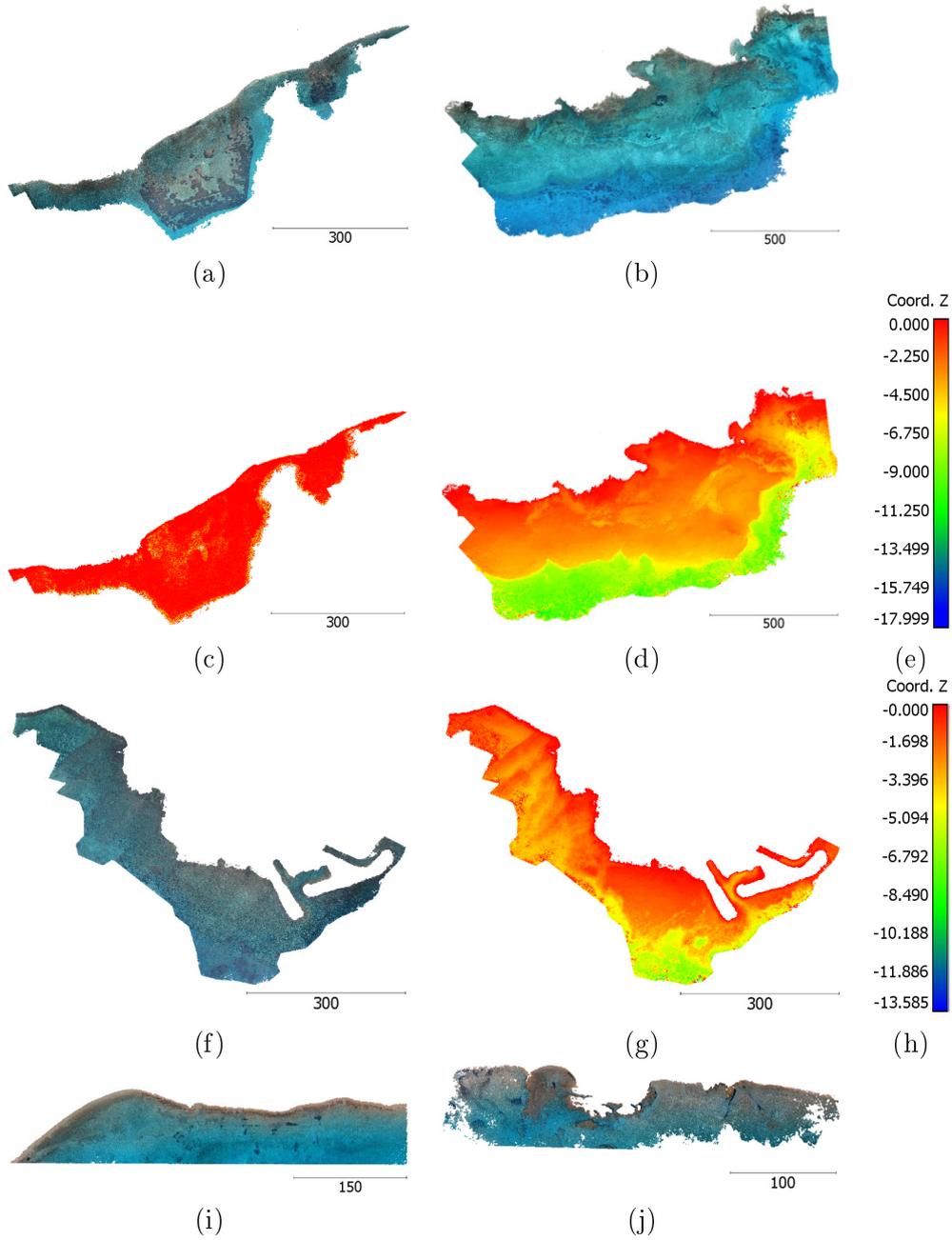
errors leading to much greater errors. In [Skinner 2011] the average error in elevations for the wetted river channel surface area was -0.5% and ranged from -12% to 13%. In [Bailly 2010] authors detected a random error of 0.19m-0.32m for the riverbed elevation from the Hawkeye II sensor. In [Fernandez-Diaz 2013] the standard deviation of the bathymetry elevation differences calculated reaches 0.79m, with 50% of the differences falling between 0.33m to 0.56m. However, according to the authors, it appears that most of these differences are due to sediment transport between observation epochs. In [Westfeld 2017] authors report that the *RMSE* of the lateral coordinate displacement is 2.5% of the water depth for the smooth, rippled sea swell. Assuming a mean water depth of 5m, these facts lead to an *RMSE* of 0.12m. If a light sea state with small wavelets is assumed, results with an *RMSE* of 3.8% which corresponds to 0.19m in 5m water are expected. It becomes obvious that wave patterns can cause significant systematic effects in coordinate locations in the bottom. Even for very calm sea states, the lateral displacement can be up to 0.30m at 5m water depth [Westfeld 2017].

Considering the above, it is highlighted here that in the real world implementation of the proposed methods, LiDAR point clouds can be used for training the suggested model, since this is the State-of-the-Art method used for shallow water bathymetry of large areas [Menna 2018], even though in some cases the absolute accuracy of the resulting point clouds is deteriorated. These issues do not affect the principle of the main goal of the proposed methods and this is proved in the next sections where a transfer learning approach is also presented.

5.3.7 Seabed point cloud extraction

The linear SVR model developed and exploited in Method 1 for depth correction should be only applied to the seabed points of the cloud. Therefore, these points were extracted from the initial cloud. To perform this extraction, points having an elevation equal or less than the sea surface level (measured at the time of the flight with an RTK GPS) were considered as seabed points while the rest as dry land points. A filtering step was then applied based on a Statistical Outlier Removal (SOR) [Rusu 2011] approach in order to remove noise from the point cloud and facilitate the creation of a more accurate DSM in the later steps. In particular, SOR is based on the distribution of point to k-nearest neighbors' distances. Based on the initial density of the point clouds, it was decided that a k=8 is adequate for the experiments performed in the context of this thesis since it is highly unlikely for some of those 8 neighbors to be in a significant distance. For each point, its mean distance to all its neighbors is computed. By assuming that the resulting distribution of the distances is Gaussian, all points whose mean distances are over a threshold can be considered as outliers and trimmed from the point cloud. This approach was considered necessary because image-based seabed point clouds produced using uncorrected imagery are prone to noise due to the effect of refraction on the Dense Image Matching process, especially in deeper areas or areas of poor texture. In Figure 5.22 the seabed point clouds for all test sites used

in this thesis are presented with natural colors (Figure 5.22a, b, f, i, j), while their corresponding bathymetry is illustrated in Figure 5.22c, d, g, k and l.



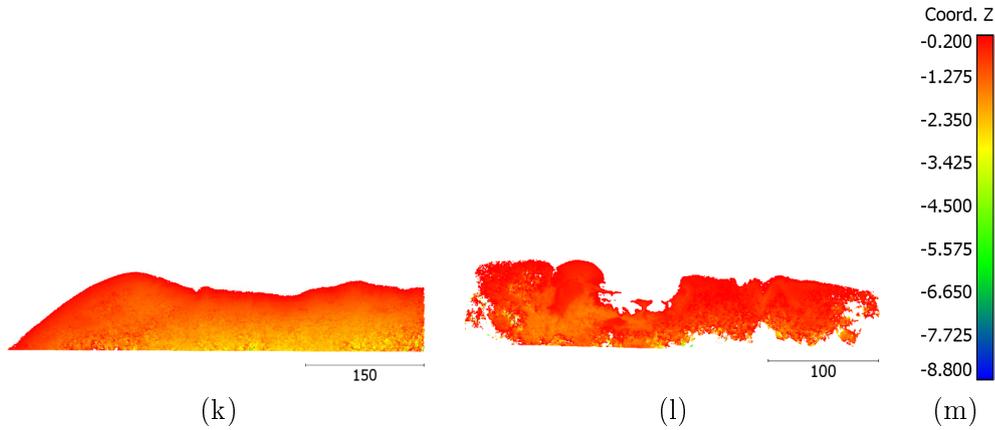


Figure 5.22: The initial seabed point clouds for Amathounta (a), Agia Napa (b), Dekelia (f), Cyclades-1 (i) and Cyclades-2 (j) test sites. Scalebars and apparent depths are in meters.

5.4 Experimental Results and Validation on Real World Data

In this section, the two complementary methods presented in Chapter 3 and Section 4.2 are tested and evaluated in real world operational conditions. Additionally, a transfer learning approach between the synthetic and the real world data is performed, where the already trained models on the synthetic data presented in Section 5.2 are exploited to predict the depths over the real world point clouds and correct the initial imagery over the real world data.

5.4.1 Correcting the Geometric Effects of Refraction in the 3D space

Method 1 is implemented in the already described real world datasets. In order to demonstrate that the linear SVR models proposed, have significant potential for applications over different areas and shallow waters, and that a predicted model can be used in areas where true data are not available, training and testing processes were always performed in different test sites, including testing of the models trained on the synthetic data over the real world datasets. Hence, a model trained on the image-based point cloud and the true depth point cloud of a test site, was tested on many other different test sites, where true data were also available but used only for evaluating the predicted depths. To apply the presented method in the real-world datasets in order to correct the depth of the dense point clouds that are affected by refraction, several steps should be performed that are analyzed in the subsequent sections.

5.4.1.1 Pre-training Phase

Unlike the synthetically generated data that are free of external errors and mismatches between the image-based and the true depth point clouds, the pre-training phase is a very crucial procedure for every supervised learning method when real world data are being used for training. Within this phase, the training data which will afterwards be fed to the regressor are created. The quality and quantity of the produced training data have a direct relation to the final precision of the regression and the created model. If the training data are not enough, not representative or not appropriate for the specific application, any further regression will fail to succeed, regardless of the applied regression method. For that reason, the obtained data must be adapted to the studied application, fulfilling its requirements in reliability, size, and diversity.

The main goal of the method to be tested in the real-world data is the correction of the geometric effects of refraction on the 3D point clouds of the seabed together with the attainment of a model able to generalize over different cases. So, the prerequisite for the training data is to describe in a representative way the different seabed mapping campaigns and different seabed characteristics. On the other hand, the phase of obtaining the necessary training data is a very time consuming and sometimes costly procedure in all learning-based applications since a vast amount of data must be processed. However, as most learning-based applications, the costs of the pretraining phase for this refraction correction method is not considered.

Pre-processing of training data is performed beforehand, without burdening the main refraction correction method with time consumption. Pre-training and regression are regarded as separate steps, allowing the pre-training phase to sacrifice its time efficiency for the sake of providing optimal results. This rule was also followed in this case, allowing the use of time-consuming approaches, to make sure that the best possible matches will be produced from the training points. Exactly this, i.e. the final precision of the refraction correction model, was the only criterion for performing this pre-training phase.

Considering the above, and in order to facilitate the training of the proposed bathymetry correction model using adequate and representative real data that will lead to a successful SVR model, the test sites of Amathounta, Agia Napa and Dekelia were considered as appropriate. Compared with Cyclades-1 and Cyclades-2 datasets, in the three selected areas dense true data are available, which have been collected using the LiDAR system described in Subsection 5.3.6. These true data extend over the whole area covered by the image-based dense point clouds, thus, facilitating a representative and robust correlation between the apparent (Z_0) and the real depth (Z) of the 3D point clouds. Cyclades-1 and Cyclades-2 datasets will be used only during the testing phase.

Towards that direction, the extracted seabed point clouds of the three test areas were evaluated in terms of surface density and roughness, following the point cloud quality estimation methodology described in [Mangeruga 2018] and only areas of low noise were used for further processing. An additional and very important criterion

was the GCPs distribution in the photogrammetric block. Taking into account the above considerations, the selected areas to be used for training the SVR model are illustrated with polygons in Figure 5.23. These areas are characterized by a wide range of depths, various seabed types including sandy or rocky bottom as well as bottom covered with seagrass. They are also characterized by a stronger geometry of the photogrammetric block, also considering the distribution of the GCPs.

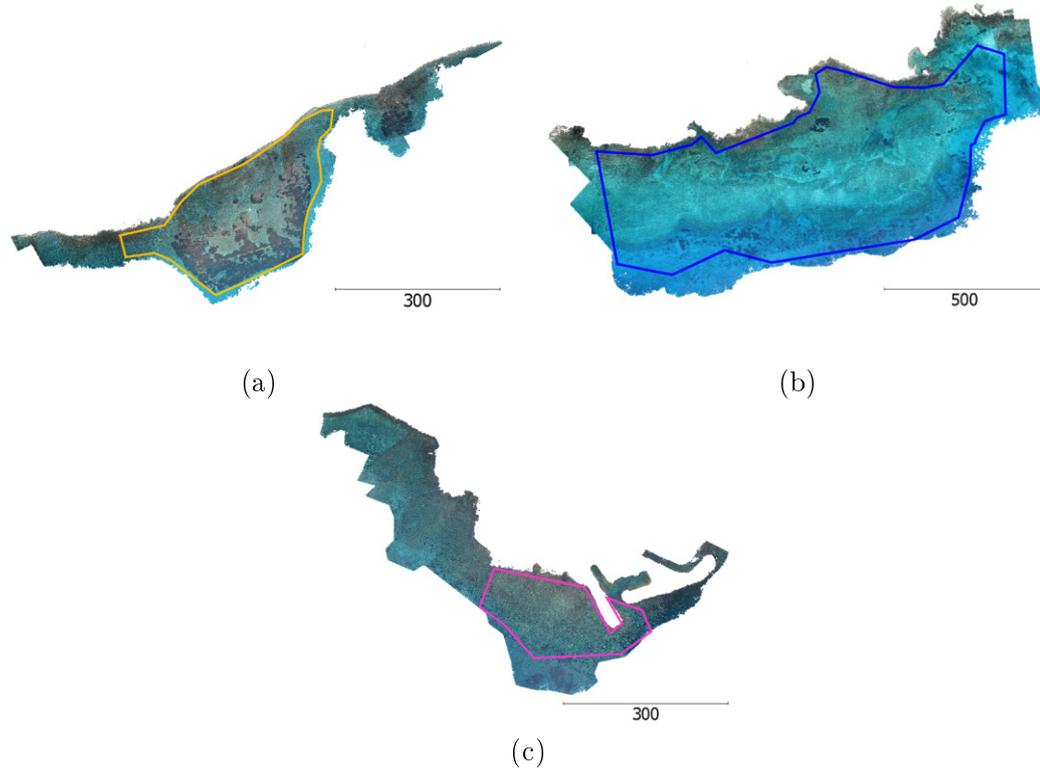


Figure 5.23: The initial seabed point clouds for Amathounta (a), Agia Napa (b) and Dekelia (c) test sites and the respective areas of the point clouds used for training the presented method. Scale bars are in meters.

This was considered necessary, although an SOR filtering was already applied on these 3D point clouds, in order to increase the final precision of the refraction correction model. Consequently, most of the outlier data stemming from a small number of mismatches in areas with strongly changing reflectance behavior (i.e. areas with sun glint) were removed from the dataset.

In addition, due to the differences in density between the LiDAR and the image-based point clouds, to facilitate the training, testing and validation of the model, the number of points of the image-based point clouds was reduced exactly to that of the LiDAR point clouds, for the three test sites. That way, two depth values would correspond to each X, Y-position of the bottom: the apparent depth Z_0 and the LiDAR depth Z . Since image-based and LiDAR points do not share the same horizontal coordinates, meshes were generated using the image-based point clouds,

since they were denser to facilitate interpolation. The respective apparent depth Z_0 was then interpolated from the LiDAR point cloud. Points having $Z_0 \geq Z$ were also considered as outliers, since this cannot be valid when the refraction phenomenon is present. Moreover, it is reminded that points having $Z_0 \geq 0\text{m}$, i.e., above the water surface, were already removed, thus avoiding errors in the training process.

Due to the large size of the reference dataset in Agia Napa test site, it was split in two parts to augment the available datasets; the two areas were selected to have similar number of points and describe various seabed morphologies: Part I having 627.522 points (Figure 5.24a in the red rectangle on the left and Figure 5.25a) and Part II having 661.208 points (Figure 5.24a in the red rectangle on the right and Figure 5.25b). The Amathounta dataset (Figure 5.24a in orange color and Figure 5.25c) and the Dekelia dataset (Figure 5.24a in magenta color and Figure 5.25e) were not split, since the available LiDAR points were much less, and in the case of Amathounta, also quite scattered.

Finally, a virtual dataset containing almost the same number of points from two of the discussed datasets was created. This dataset was generated using the total of the Amathounta dataset points and 1% of the Agia Napa Part II dataset. The $Z-Z_0$ distribution of this "Merged Dataset" is presented in Figure 5.25d in comparison with the Agia Napa dataset. In the same figure, the $Z-Z_0$ distribution of the $Z-Z_0$ distribution of the Agia Napa dataset are depicted.

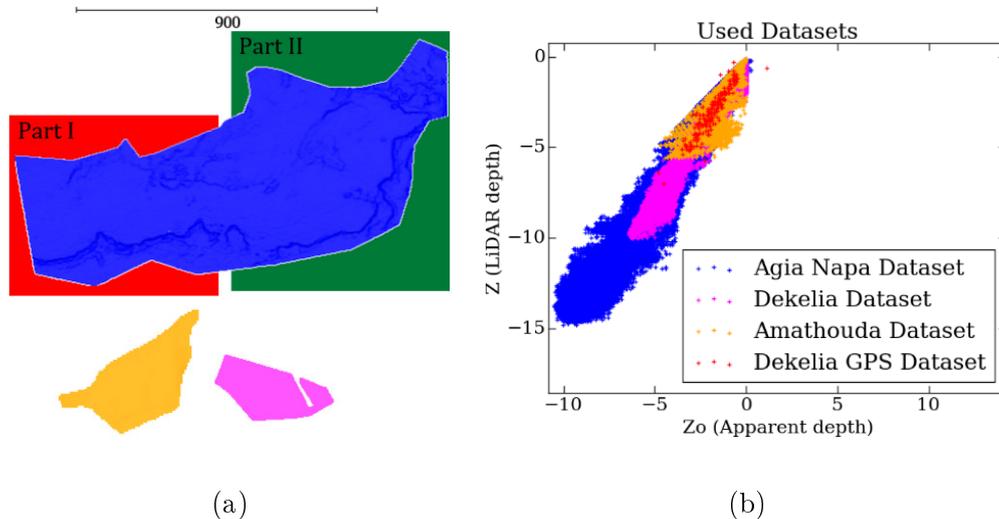


Figure 5.24: (a): Part I (left) and Part II (right) datasets from the Agia Napa test site in blue, the Amathounta test site in orange, and the Dekelia test site in magenta, all having the same scale (b): The distribution of the Z and Z_0 values for each dataset.

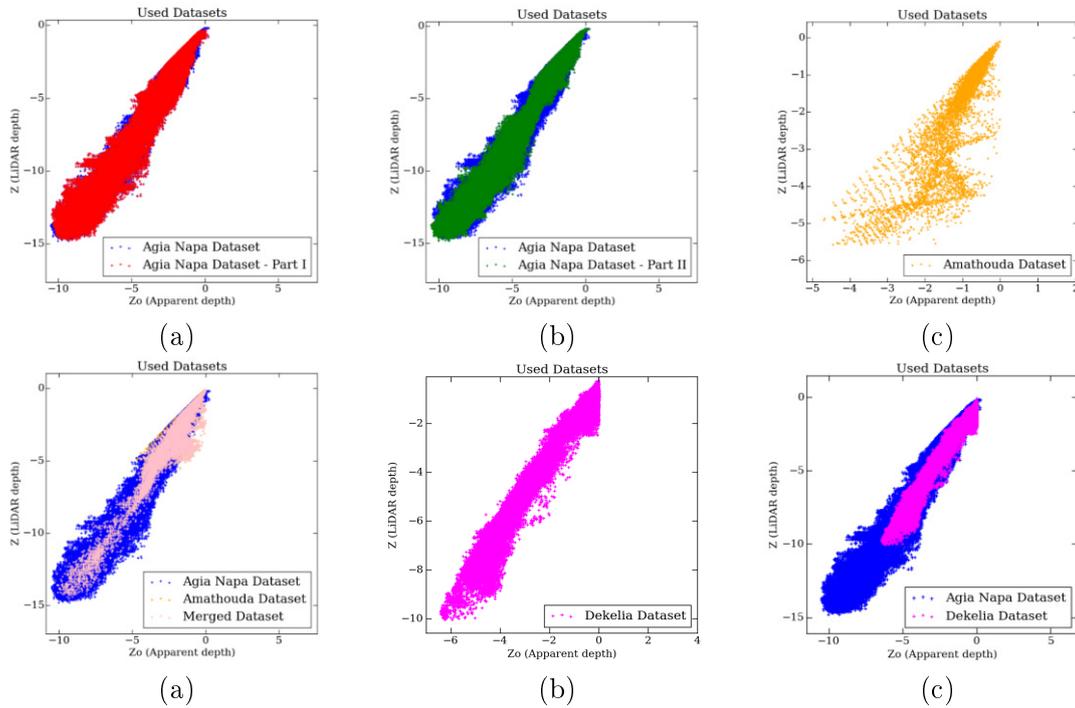


Figure 5.25: The depth to apparent depth ($Z-Z_0$) distribution of the datasets prepared for training: the Agia Napa Part I dataset over the full Agia Napa dataset (a). The Agia Napa Part II dataset over the full Agia Napa dataset and (b) the Amathouda dataset (c); the merged dataset over the Agia Napa and Amathouda datasets (d) and the Dekelia dataset; and (e) the Dekelia dataset over the full Agia Napa dataset (f).

5.4.1.2 Training, Validation and Testing

Seven different training sets were formed and then validated against 42 different testing sets. In addition to the above, the trained model over the synthetic dataset of 150m[DTM1] was used to be tested over seven more testing sets. Training and testing processes were performed in different test sites. As such, a model trained on a test site, was tested on many other different test sites where true data were also available but used only for evaluating the accuracy of the predicted depths. In this way, it was ensured that the trained bathymetry correction models would have significant potential for generalization over different areas and shallow waters, and that a predicted model from a training site could be used in areas where true data were not available.

Training

As already stated, SVR has great potential to deliver robust models, even using sparse data for training. Due to the large size of the available data for correlation in the Agia Napa Part I and the Agia Napa Part II, the use of percentages of the total

data for training was investigated. To that direction, models were trained using the 80% of the data, the 30% of the data and the 5% of the data. Results suggested that all the three training approaches for each test site delivered the same results, without any improvement or compromise in the accuracy. However, training on the 80% of the data required much more computational resources compared to the 30% and the 5% of the data. Most importantly, this decision, was a last attempt to filter the data, aiming to avoid possible artifacts existing in the data preparation process and keep for training only the data that are closer to the areas of high density presented in Figure 5.27. Considering this, in the Agia Napa Part I and the Agia Napa Part II datasets, training was performed using the 5% and the 30% of the available points. This way, the 30% and the 5% of the points of these datasets were randomly picked to form the training and the validation set.

The first and the second training approaches used 5% and 30% of the Agia Napa Part II dataset, respectively. The third and the fourth training approaches used 5% and 30% of the Agia Napa Part I dataset respectively, while the fifth training approach used 100% of the Amathounta dataset. The $Z-Z_0$ distribution of the points used for this training can be seen in Figure 5.25c. The sixth training approach used 100% of the Dekelia dataset and the $Z-Z_0$ distribution of the points used for this training can be seen in Figure 5.25e. Finally, a seventh training approach was performed using the Merged dataset. The seven training approaches are summarized below:

1. 5% of the Agia Napa Part I dataset
2. 30% of the Agia Napa Part I dataset
3. 5% of the Agia Napa Part II dataset
4. 30% of the Agia Napa Part II dataset
5. 100% of the Amathounta dataset
6. 100% of the Dekelia dataset
7. the Merged dataset

Testing

Using the models trained on the 5% and the 30% of the Agia Napa Part II dataset, the correct depth over the Agia Napa Part I, Amathounta, Dekelia, Dekelia (GPS), Cyclades-1 and Cyclades-2 datasets was predicted. Using the models trained on 5% and 30% of the Agia Napa Part I dataset respectively, the correct depth over the Agia Napa Part II, Amathounta, Dekelia, Dekelia (GPS), Cyclades-1 and Cyclades-2 datasets was predicted. Using the model resulted from the training approach performed on the 100% of the Amathounta dataset, the correct depth over the Agia Napa Part I, the Agia Napa Part II, the Dekelia, the Dekelia (GPS), the Cyclades-1 and the Cyclades-2 datasets was predicted. It is important to notice here that the

maximum depth of the training dataset was 5.57m, while the maximum depths of the testing datasets were 14.8m and 14.7m respectively. Using the model that was trained over the 100% of the Dekelia dataset the correct depth over the Agia Napa Part I, the Agia Napa Part II, the Amathounta, the Cyclades-1 and the Cyclades-2 datasets was predicted. Again, it is important to notice here that the maximum depth of the training dataset was 10.09m while the maximum depths of the testing datasets were 14.8m and 14.7m respectively. Using the model trained on the training approach using the Merged Dataset, the correct depth over the Agia Napa Part I, the Agia Napa Part II, the Dekelia, the Dekelia (GPS) datasets, the Amathounta, the Cyclades-1 and the Cyclades-2 datasets was predicted. The $Z-Z_0$ distribution of the points used for the above training approaches can be seen in Figure 5.25. Finally, it is of really high importance that the model trained on the 150m[DTM1] synthetic dataset was used for predicting the correct depth over the real world test sites and specifically: the Agia Napa Part I, the Agia Napa Part II, the Dekelia, the Dekelia (GPS) datasets, the Amathounta, the Cyclades-1 and the Cyclades-2. The tests performed are summarized below:

1. For the model trained on the 5% of the Agia Napa Part I dataset, testing was performed on:
 - (a) Agia Napa Part II
 - (b) Amathounta,
 - (c) Dekelia
 - (d) Dekelia (GPS)
 - (e) Cyclades-1 and
 - (f) Cyclades-2
2. For the model trained on the 30% of the Agia Napa Part I dataset, testing was performed on:
 - (a) Agia Napa Part II
 - (b) Amathounta,
 - (c) Dekelia
 - (d) Dekelia (GPS)
 - (e) Cyclades-1 and
 - (f) Cyclades-2
3. For the model trained on the 5% of the Agia Napa Part II dataset, testing was performed on:
 - (a) Agia Napa Part I
 - (b) Amathounta,
 - (c) Dekelia

- (d) Dekelia (GPS)
 - (e) Cyclades-1 and
 - (f) Cyclades-2
4. For the model trained on the 30% of the Agia Napa Part II dataset, testing was performed on:
 - (a) Agia Napa Part I
 - (b) Amathounta,
 - (c) Dekelia
 - (d) Dekelia (GPS)
 - (e) Cyclades-1 and
 - (f) Cyclades-2
 5. For the model trained on the 100% of the Amathounta dataset, testing was performed on:
 - (a) Agia Napa Part I
 - (b) Agia Napa Part II
 - (c) Dekelia
 - (d) Dekelia (GPS)
 - (e) Cyclades-1 and
 - (f) Cyclades-2
 6. For the model trained on the 100% of the Dekelia dataset, testing was performed on:
 - (a) Agia Napa Part I
 - (b) Agia Napa Part II
 - (c) Amathounta
 - (d) Cyclades-1 and
 - (e) Cyclades-2
 7. For the model trained on the Merged dataset, testing was performed on:
 - (a) Agia Napa Part I
 - (b) Agia Napa Part II
 - (c) Amathounta
 - (d) Dekelia
 - (e) Dekelia (GPS)
 - (f) Cyclades-1 and

(g) Cyclades-2

8. For the model trained the 5% of the 150m [DTM1] dataset testing was performed on:

- (a) Agia Napa Part I
- (b) Agia Napa Part II
- (c) Amathounta
- (d) Dekelia
- (e) Dekelia (GPS)
- (f) Cyclades-1 and
- (g) Cyclades-2

To facilitate the comparison between the predicted models, Figure 5.26 demonstrates six of them predicting depths over the Agia Napa Part I dataset: the green-colored dashed line stands for the predicted model trained on the Merged dataset, the red-colored line represents the predicted model trained on the Amathounta dataset, the green-colored line represents the predicted model trained on the Agia Napa Part I [30%] dataset, the orange-colored line stands for the predicted model trained on the Agia Napa Part II [30%] dataset, the magenta-colored line represents the predicted model trained on the Dekelia dataset and the black-colored dashed line represents the predicted model trained on the 150m [DTM1] synthetic dataset.

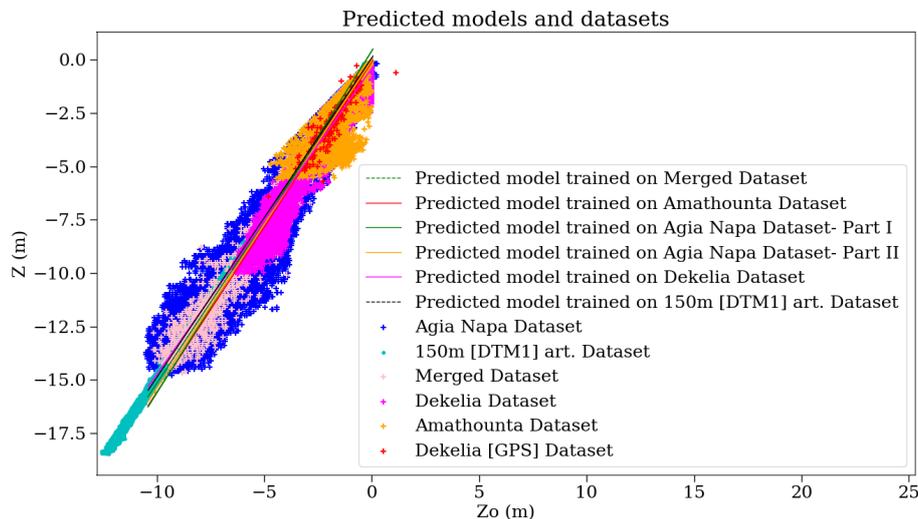


Figure 5.26: The $Z-Z_0$ distribution of the employed datasets and their respective, predicted linear models.

As can be seen in Figure 5.26, all the models succeed in following the $Z-Z_0$ distribution of the majority of the points. Although, in Figure 5.26, a number of

outlier points appear to lie away from the predicted models, these points are quite sparse and not enough to affect the final accuracy of the results. This is also shown in Figure 5.27, where the $Z-Z_0$ distribution and the respective point densities are presented for the 100% and the 95.4% of the points of the initial datasets. It must be noticed here that for the Cyclades-1 and the Cyclades-2 datasets, only the 100% of the points is presented, due to the small number of the available true data. There, it is obvious that most of the points is concentrated in the narrow distribution which can be described adequately by a linear model.

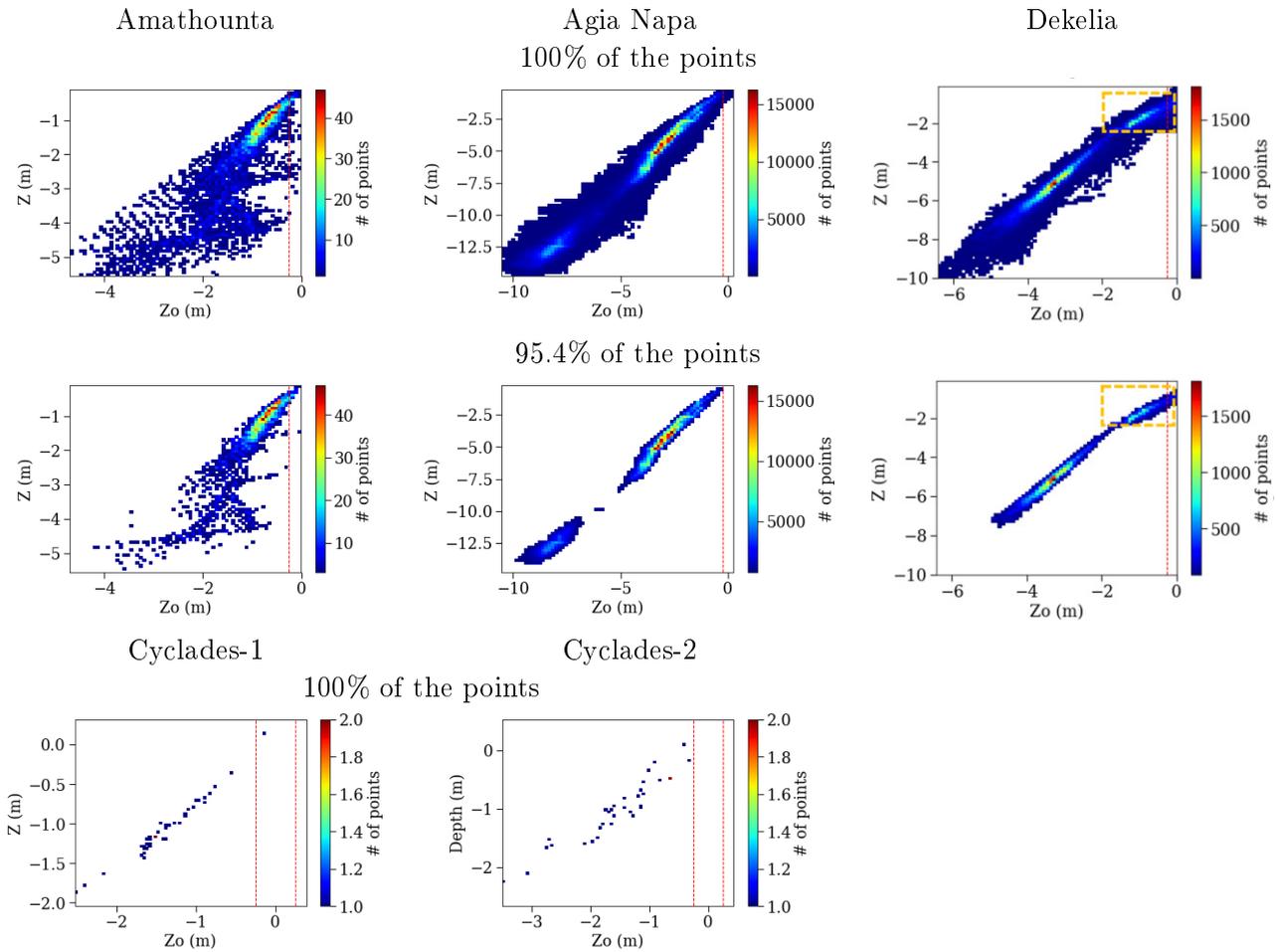


Figure 5.27: The $Z-Z_0$ distribution of the used datasets and the respective point densities are presented.

By examining the 2D histograms of Figure 5.27, a slight deviation from this linear distribution in the shallower part of the Dekelia dataset is observed and it is highlighted in the orange rectangle. This deviation, being limited in the shallower area till the depth of 2.5m, can be attributed only to the erroneous SfM and consequently MVS results in this shallower zone, generated by images containing dry land and water covered areas. Moreover, errors in the LiDAR data processing might

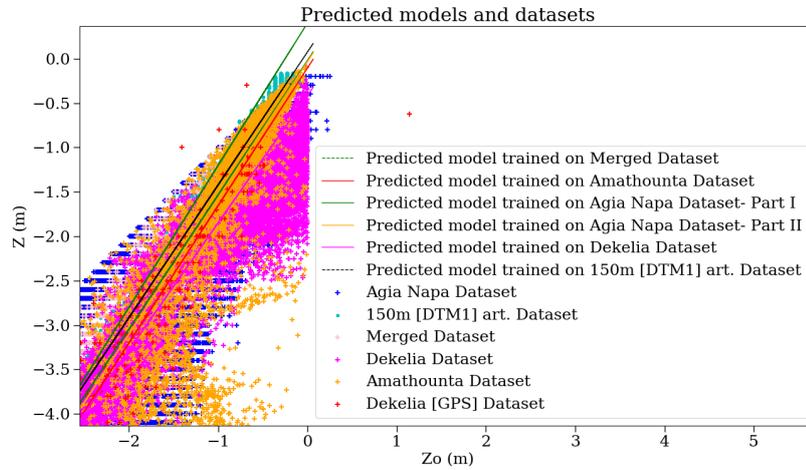
also be a cause.

In Figure 5.28, three zoomed areas of the $Z-Z_0$ distribution of the employed datasets and their respective, predicted linear models is presented in order to easier identify the differences between the datasets and the models. By observing Figure 5.26 and especially Figure 5.28, the slight differences between the trained models are evident. As can be seen, the models trained on the Agia Napa [Part I] and the Dekelia datasets, are deviating the most by the rest of the models in the top of the scatter diagram which represents the shallower areas of the test sites (Figure 5.28a). However, when calculating their differences from the model trained on the 150m [DTM1] synthetic dataset at the depth of 0.5m, they are 0.17m and 0.31m respectively. Although both of the differences are quite big for this depth, regarding the model trained on the Agia Napa [Part I] dataset, the deviation is within the IHO limits [Guenther 2000].

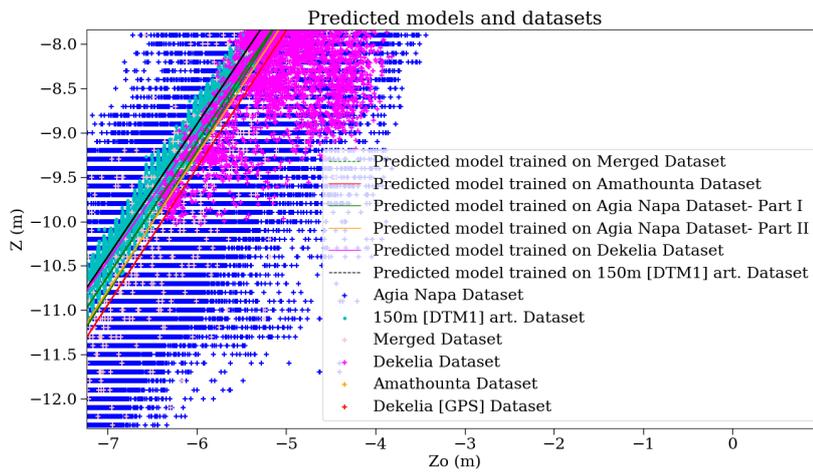
When the middle part of the scatter diagram (Figure 5.28b) is examined, the above models are not deviating more than 0.07m and 0.14m respectively at the depth of 9.5m, however, the deviation of the model trained on the Amathounta dataset reaches 0.33m. This larger deviation of the Amathounta dataset was expected since compared to the Agia Napa [Part I] and the Dekelia datasets, the Amathounta reaches the maximum depth of 5.57m only, not offering the necessary points for training at the depth of 9.5m. On the contrary, most of the points of Amathounta dataset have maximum depth of 2m while at this depth, the Agia Napa [Part I] and Dekelia datasets have quite low point densities.

At the lowest part of the scatter diagram (Figure 5.28c), it is obvious that the models trained on the Agia Napa [Part II], the Merged and the Amathounta datasets are deviating the most. Specifically, at the maximum depth of 14.8m where real word data are available the deviations are 0.31m, 0.43m and 0.43m respectively. It is also important to highlight here that the differences between the predicted model trained on the Amathounta dataset and the predicted models trained on Agia Napa datasets are not remarkable, even though the maximum depth of Amathounta dataset is 5.57m and the maximum depths of Agia Napa datasets are 14.8m and 14.7m respectively. The biggest difference observed among these three predicted models is between the predicted model trained on Amathounta dataset and the predicted model trained on the Dekelia dataset: 0.52m at 14.8m depth, or 3.51% of the real depth.

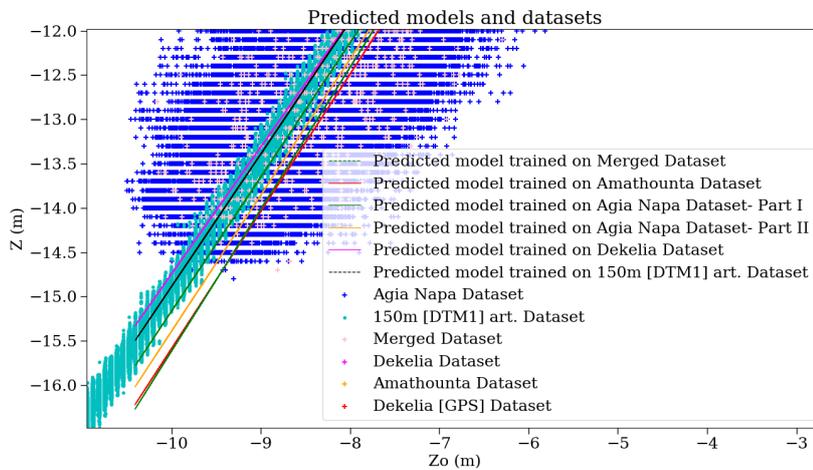
Another important observation is that at the larger depths, except of the model trained on the Dekelia dataset, all the rest of the trained models are predicting shallower depths than the depths predicted by the model trained on the synthetic dataset. This trend is very small, and it is not directly visible in the scatter diagrams of the $Z-Z_0$ distribution in Figure 5.23, Figure 5.24 and Figure 5.25. Only when the scatter diagram of Dekelia test site is observed, a slight bending of the points towards the shallower depths can be noticed after the depth of 8m-9m. However, as seen in Figure 5.26 and Figure 5.28, this is not affecting the training of the SVR model, probably due to the sparsity of the data.



(a)



(b)



(c)

Figure 5.28: The $Z-Z_0$ distribution of the employed datasets and their respective, predicted linear models.

The observed trend can be attributed to the unfavorable distribution of the GCPs in the aerial block and their $RMSE_Z$. Considering that for the Agia Napa test site the $RMSE_Z$ of the GCPs is 0.074m, and that the deeper parts of the test site lie at a distance of about 400m – 600m from the control points lying in the coast line, it is not unlikely for accumulated error up to 1m in Z to appear in these areas. However, the direction of this error is not systematic but random i.e. it may reduce or it may increase the apparent depths.

The generation of slightly shallower apparent depths in the real world datasets, can also be attributed to the incomplete calibration of the images. This can be true considering the GCPs distribution in the block and that most of the lightweight compact commercial cameras that are used for UAV mapping, due to their lenses, are introducing barrel radial distortion to the images. In these cases, radial distortion can increase the effect of the refraction phenomenon in the images by placing a pixel in a larger radial distance than the distance that it would be placed only by the effects of the refraction. This, leads the SfM-MVS processing to generate point clouds of slightly shallower apparent depths than the apparent depths that would be produced if only the refraction effect were introduced to the images, as it was the case with the synthetic datasets.

Moreover, in the scatter plots and the 2D histograms of the Agia Napa [Part I], the Agia Napa [Part II] and the Dekelia datasets, it is obvious that after the depth of 9m to 10m, the range of the Z_0 of the points along the X axis is increased enough, compared with the range of the Z_0 of the points in the shallower depths.

An important source of this increased range of the apparent depths and the trend reported before is the deterioration of the image quality by the combination of the intense refraction effect with the decreased visibility due to absorption and scattering, at depths of more than 9m to 10m. Since sea water's density is not constant through the depth, being a function of temperature, salinity and pressure, the larger the depth, the more the layers of different refraction index the rays are passing. Also, due to the refraction effect, matchable pixels are placed sometimes far from the epipolar line that passes through one of the points, which leads to them being filtered and not used, or generating wrong correspondences.

The limitation of the depth is also reported in [Legleiter 2019] where the maximum detectable depth in the tests performed is 9.5m. However, this might be also related with the specific SfM-MVS software used.

Indeed, by investigating the SfM results of the Agia Napa [Part I] and Agia Napa [Part II] datasets, it was found that after the 9m–10m depth, even a dense point cloud was generated, this, was not supported by the necessary matches in the SfM process, in order to ensure an accurate and reliable result. This can be seen in the typical stereopairs of the deepest areas of the Agia Napa [Part I] and Agia Napa [Part II] test sites presented in Figure 5.29. There, the iso-depth line of 10m is presented with the orange dotted line, the valid matches are presented with the blue colored line while the invalid matches with the red colored line.

Although it can be said that for the specific cases of Agia Napa [Part I] and Agia Napa [Part II] test sites, depths after 9m to 10m are characterized mainly by sandy

bottom with scattered areas of seagrass and rocks, and this might be the cause of the mismatches, it is observed that seagrass offer distinctive blobs for pyramid matching due to their large gradients resulting from the high contrast between the seagrass and the sand. However, in these depths, even if more detailed features such as the scattered rocky formations exist, which in the shallower areas serve as distinctive and unique points for detection and matching, they are not suitable for key points, obviously due to the blurriness and the general deterioration of the image quality because of the depth.

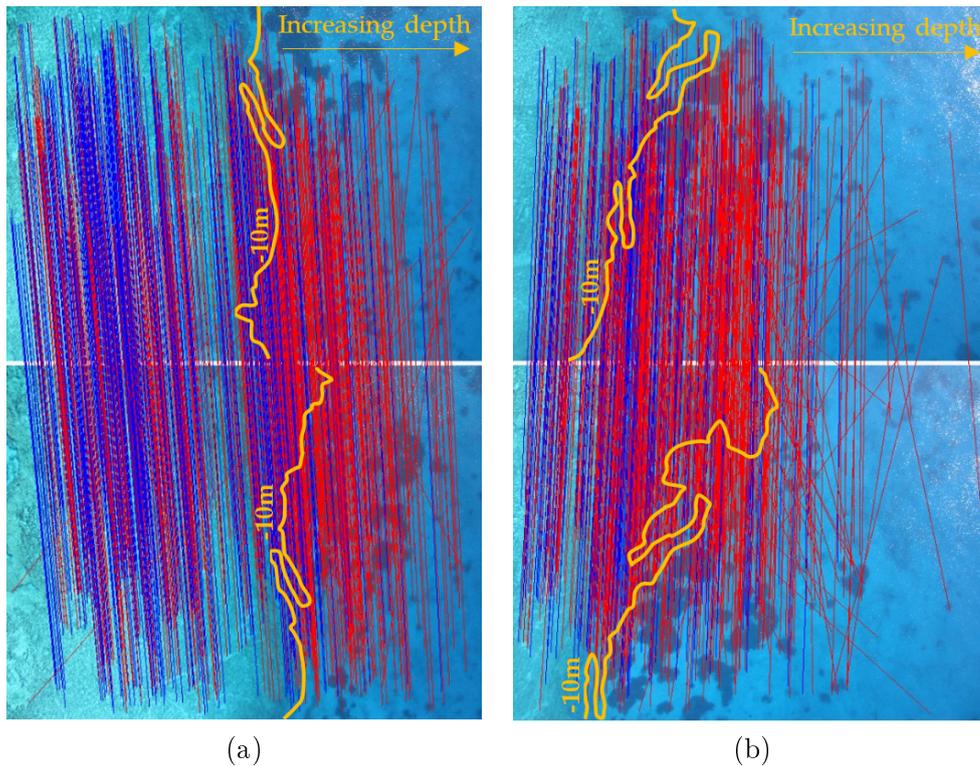


Figure 5.29: Typical stereopairs and their matches in the deepest areas of the Agia Napa [Part I] (a) and the Agia Napa [Part II] (b) test sites. The valid matches are presented with blue color while the invalid matches with red. The orange dotted line represents the isodepth line of 10m.

In the next paragraphs, the results of the proposed method are evaluated based on cloud to cloud distances. Considering the aforementioned reasons of degradation of the quality of the point clouds in depths of more than 9m to 10m in the Agia Napa [Part I] and Agia Napa [Part II] datasets and in order to distinguish between the errors caused by the SfM-MVS processing and the limitations of the proposed method separate results are presented for the points having the full range of depth and for the points having depth up to 10m. Additionally, cross sections of the seabed are presented to highlight the high performance of the proposed method, and the issues and differences observed between the tested and true depth point clouds.

5.4.1.3 Evaluation of the predicted depths

Point cloud comparison

To demonstrate the magnitude of the errors introduced by the refraction effect in the case of the real world datasets, the apparent depths of the initial (uncorrected) dense point clouds are compared with the depths of the true data presented in Table 5.8 using the multiscale model to model cloud comparison (M3C2) [Lague 2013] module in Cloud Compare freeware [Girardeau-Montaut 2015]. The initial point clouds used for this comparison are the point clouds already presented in Subsection 5.3.7 which are generated by performing SfM-MVS using the imagery affected by refraction.

According to the literature, the M3C2 algorithm offers accurate surface change measurement which is independent of point density [Lague 2013].

In Figure 5.30 and Figure 5.31, it can be observed that the distances between the depths of the reference data and those of the original image-based point clouds are increasing proportionally to the depth. These comparisons and the reported errors are in accordance with the comparisons performed with the synthetic data in Section 5.2 and confirm that the refraction effect cannot be ignored in seabed mapping and bathymetric applications based on SfM-MVS processing for shallow waters.

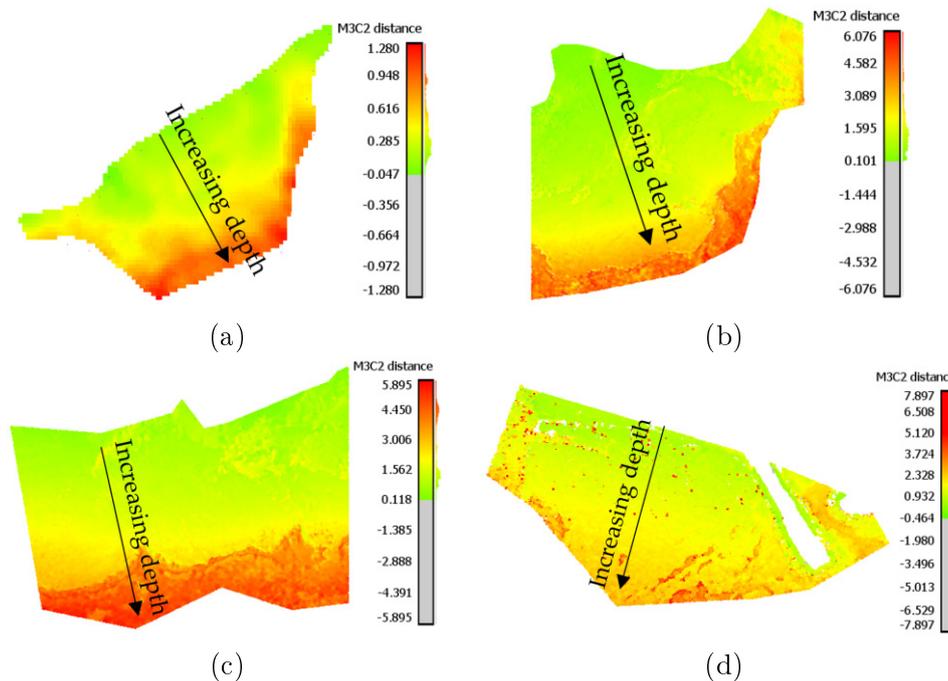


Figure 5.30: The initial M3C2 distances (in m) between the (reference) LiDAR point cloud and the uncorrected image-based point clouds derived from the SfM-MVS. (a) M3C2 distances of Amathounta, (b) M3C2 distances of Agia Napa [Part II], (c) M3C2 distances for Agia Napa [Part I] test site and (d) M3C2 distances for Dekelia test site.

In all eight test cases demonstrated in Figure 5.30 and Figure 5.31, the Gaussian mean of the differences was significant. It reached 0.44m (RMSE 0.51m) in the Amathounta test site (5.30a), 2.23m (RMSE 2.64 m) in the Agia Napa test site (full, [I] and [II]) (5.30b,c), 1.72m (RMSE 1.53m) in the Dekelia test site (5.30d) and 1.15m in the Dekelia [GPS] test site, 0.32m (RMSE 0.33m) in the Cyclades-1 test site and 0.54m (RMSE 0.62m) in the Cyclades-2 test site. Since these values might be considered "negligible" in some applications, it is important to stress that in the Amathounta test site, more than 30% of the compared image-based points presented a difference of 0.60- 1.00m from the LiDAR points; in Agia Napa, the same percentage presented differences of 3.00–6.07m; i.e., 20% - 41.1% percent of the real depth, in Dekelia, the same percentage of points presented differences of 1.76m – 4.94m, in Cyclades-1 the same percentage presented differences of 0.28m -0.52m while in Cyclades-2 the same percentage of point presented differences of 0.7m - 2.00m.

In Figure 5.30, some differences between the level of detail of the image-based and the LiDAR point clouds are also evident; the rocky formations on the seabed and the areas with abrupt changes in depth systematically present larger differences compared to the surrounding area. This is understood when observing all the subfigures of Figure 5.30 but especially Figure 5.30d, where many areas in red (indicating increased differences) are scattered on the seabed, regardless of depth.

As can be observed in Figure 5.31, the majority of the M3C2 distances in all histograms are far from the red dashed lines representing the accuracy limits generally accepted for hydrography [Guenther 2000]. Regarding the Dekelia [GPS], the Cyclades-1 and the Cyclades-2 test sites, despite their sparse and small numbered true depth points that lead to sparse histograms, a distribution similar to those of the Amathounta, Agia Napa, and Dekelia test sites can be observed.

To show again that the vast majority of the points of the 2D histograms, is concentrated in a narrow area of linear shape, in Figure 5.31, the 100% and the 95.4% of the points of the differences are presented for the datasets of which the number and the density of the points allow it.

As also noticed in Figure 5.27, a slight deviation from this linear distribution in the shallower part of the Dekelia dataset is observed and it is highlighted again with the orange rectangle. This deviation is the reason for the different behavior of the model trained on the Dekelia dataset reported before, when the results presented in Figure 5.28 were discussed; at the larger depths, except of the model trained on the Dekelia dataset, all the rest of the trained models are predicting shallower depths than the depths that are predicted by the model trained on the synthetic dataset. Specifically, these points are affecting the model by changing its inclination.

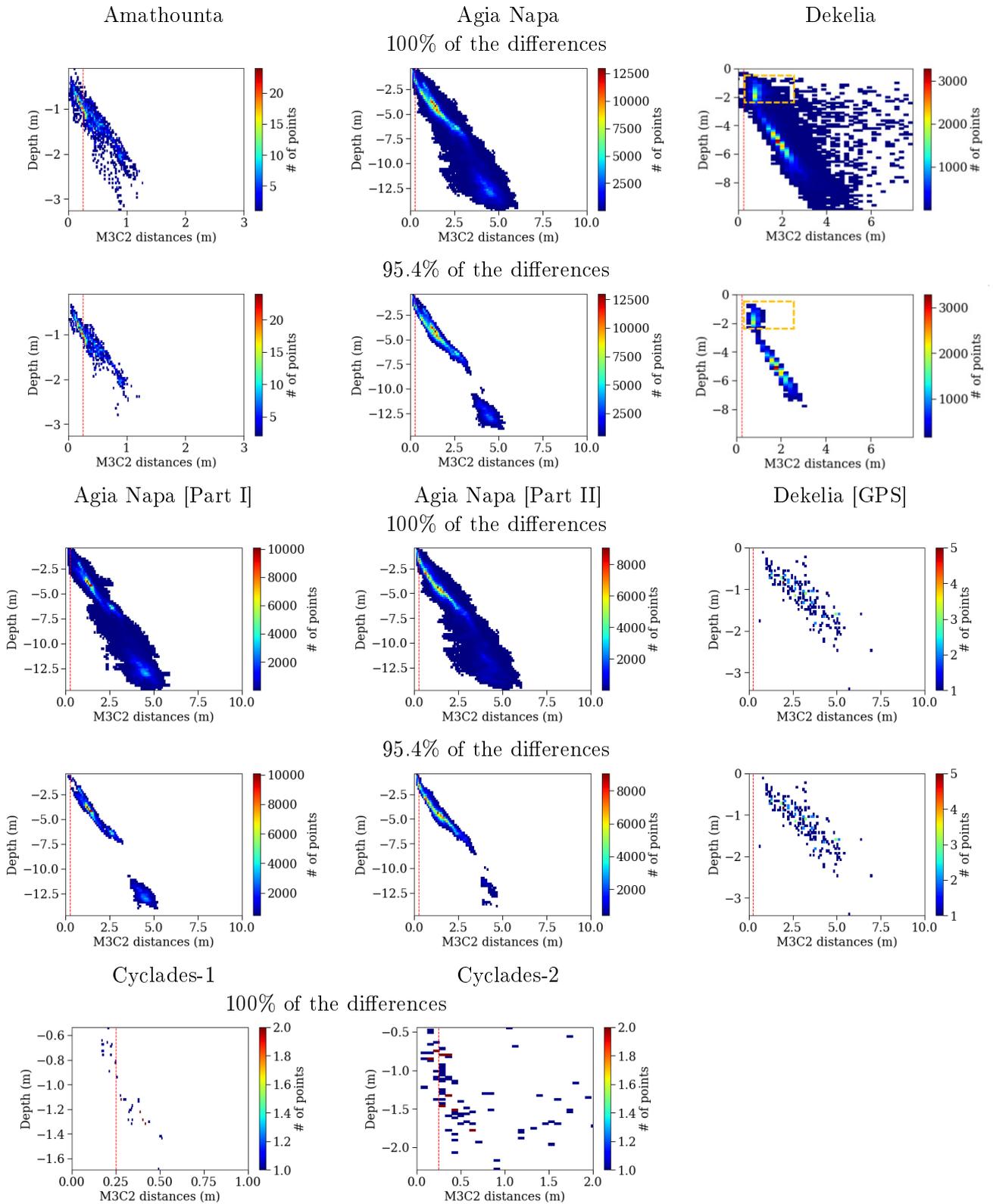


Figure 5.31: The 2D histograms of the M3C2 distances between the true and the uncorrected and corrected image-based dense point clouds derived from the SfM for Amathounta, Agia Napa, Dekelia, Agia Napa [Part I], Agia Napa [Part II], Dekelia [GPS], Cyclades-1 and Cyclades-2 test sites respectively, in relation to the real depth. The dashed lines in red represent the accuracy limits generally accepted for hydrography as introduced by the IHO [Guenther 2000].

Figure 5.32 presents the 2D histograms of the M3C2 distances between the depths of the true depth point clouds and the depths of the point clouds produced from the predicted model trained on each dataset in relation with the true depth. As in the previous 2D histograms, the dashed lines in red represent the accuracy limits generally accepted for hydrography as introduced by the IHO [Guenther 2000]. As it happened with the already presented 2D histograms, in Figure 5.32 the 2D histograms for the 100% and the 95.4% of the points of the differences are demonstrated. However, since it is already clear that the majority of the depths of the points is corrected with high accuracy, the 2D histograms for the 95.4% of the points and consequently the 95.4% of the compared elevations is given only for the models tested on the Agia Napa [Part I] test dataset. There, it is obvious that the vast majority of the compared elevations, present small differences, compared with the ground truth elevations.

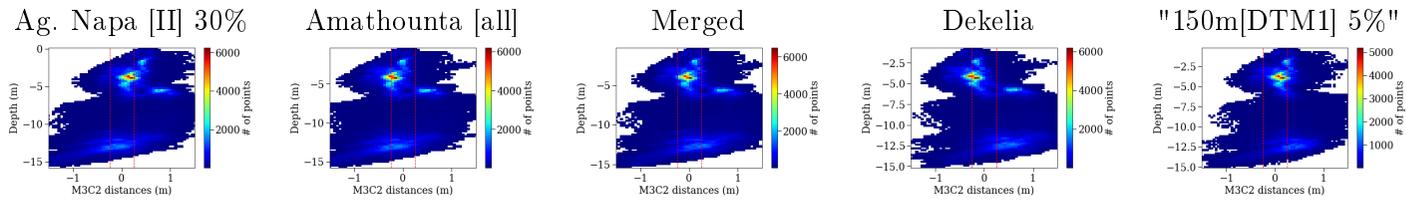
Table 5.9 presents the results of each one of the 51 tests performed with every detail. Additionally to the above, the results of the tests over the Agia Napa [Part I] and Agia Napa [Part II] test sites till the depth of 10m are given, raising the total number of the evaluation cases to 59.

Both in Figure 5.32 and Table 5.9, a great improvement is spotted. By observing the results, it is clear that the model trained on the synthetic dataset 150m[DTM1] 5% is delivering the best results compared with the models trained over the real world datasets. The increased accuracy of the predictions of this model were expected, since it was trained on external error free and fully compatible data. These results, confirm the proper modeling of the refraction effect in the synthetic data generation and consequently the reliability of the second method for correcting the geometric effects of the refraction effect, present in Section 4.2 and evaluated in Subsection 5.4.2. More specifically, the model trained on the synthetic data achieves the reduction of the mean error as follows: in Agia Napa [Part I] test site, from the initial 2.23m to 0.15 m; in Agia Napa [Part II] test site, from the initial 2.23m to 0.09m; in Amathounta, the initial mean distance of 0.44m was reduced to 0.04m; in Cyclades-1 test site, the initial mean distance of -0.32m was reduced to -0.05m; in Cyclades-2, the initial mean distance of -0.54m was reduced to -0.05m while in Dekelia test site, the initial mean distance of 1.72m was reduced to -0.16m, including outlier, points such as seagrass, which are not captured in the true depth point clouds for all the cases or are caused due to image-based point cloud noise again in areas with seagrass or poor texture. When the maximum evaluated depth is limited to 10m for the Agia Napa [Part I] and the Agia Napa [Part II] test sites, the initial mean distance of 1.52m is reduced to 0.11m and -0.11m respectively.

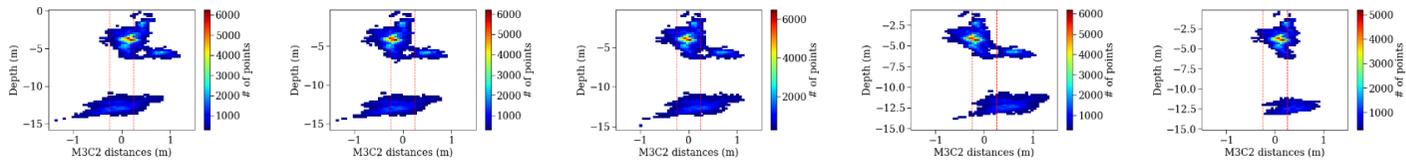
These results serve as a typical example of a successful transfer learning approach. Indeed, synthetic data are increasingly being used for machine learning applications; a model is trained on a synthetically generated dataset with the intention of transfer learning to real data. In the literature, efforts are reported for the construction of general-purpose synthetic data generators to enable data science experiments [Patki 2016].

Training Site for Testing on Agia Napa Part I

100% of the points

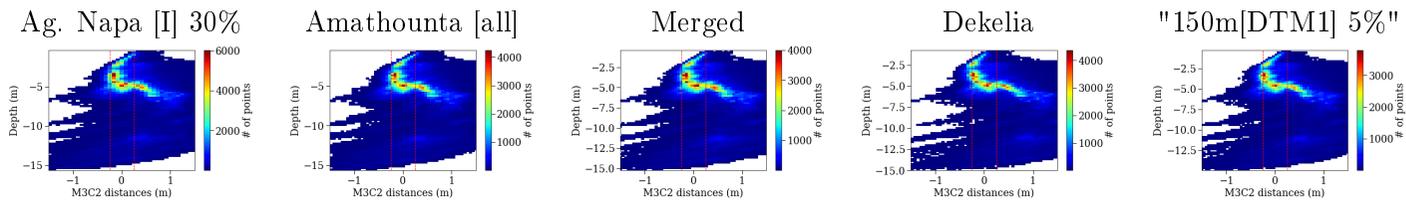


95.4% of the points



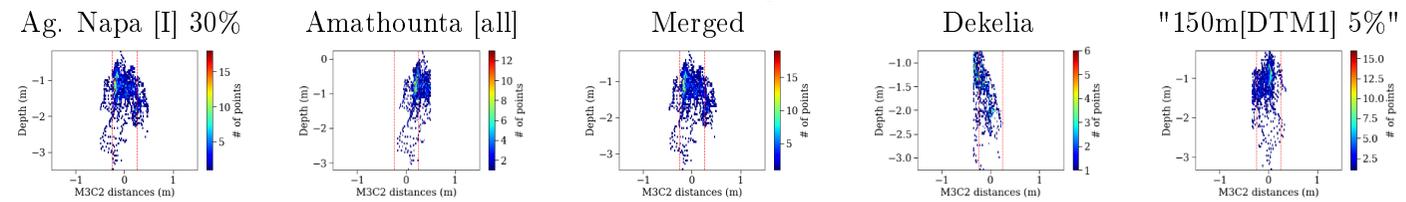
Training Site for Testing on Agia Napa Part II

100% of the points



Training Site for Testing on Amathounta

100% of the points



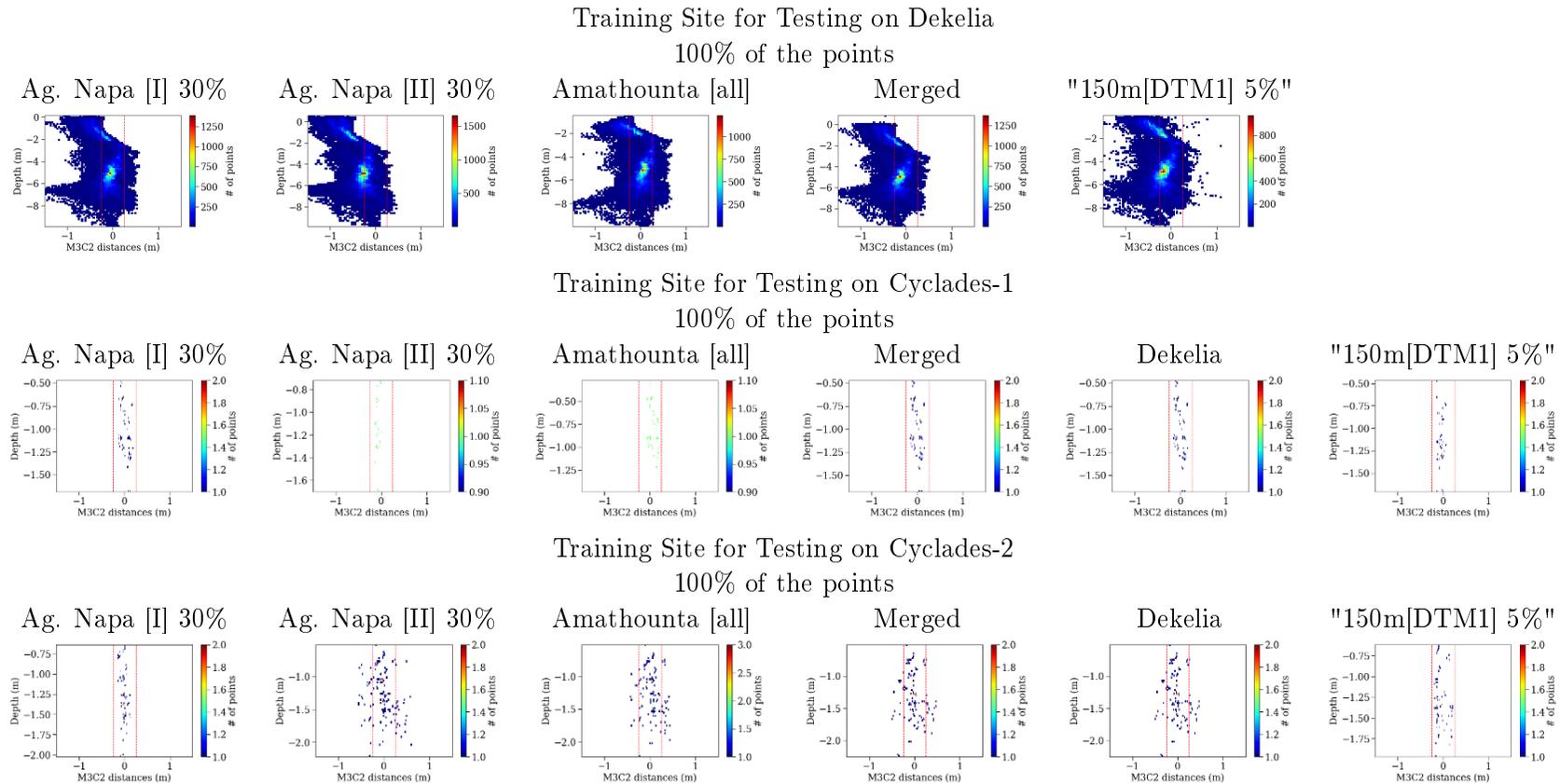


Figure 5.32: The 2D histograms of the M3C2 distances between the LiDAR point cloud and the corrected point clouds after the application of the proposed method in relation to the real depth. The red dashed lines represent the accuracy limits generally accepted for hydrography, as introduced by the IHO [Guenther 2000].

Training				Evaluation			Statistical Analysis [m]					
Site	# points	%	Fitting Score	Site	Max /Min depth	# points	Uncorrected data			Corrected data		
							\bar{x}	s	RMSE	\bar{x}	s	RMSE
Ag. Napa[I]	627.552	5	0.984	Ag. Napa[II]	14.7/0.30	661.208	2.23	1.42	2.64	-0.15	0.49	0.51
Ag. Napa[I]	627.552	30	0.984	Ag. Napa[II]	14.7/0.30	661.208	2.23	1.42	2.64	-0.14	0.50	0.51
Ag. Napa[I]	627.552	5	0.984	Ag. Napa[II]	10/0.30	576.803	1.52	0.80	1.72	-0.20	0.30	0.36
Ag. Napa[I]	627.552	5	0.984	Amathounta	5.57/0.10	5400	0.44	0.26	0.51	-0.03	0.19	0.19
Ag. Napa[I]	627.552	30	0.984	Amathounta	5.57/0.10	5400	0.44	0.26	0.51	-0.03	0.19	0.19
Ag. Napa[I]	627.552	5	0.984	Dekelia	10.1/0.09	101.887	1.72	0.76	1.89	-0.12	0.25	0.25
Ag. Napa[I]	627.552	30	0.984	Dekelia	10.1/0.09	101.887	1.72	0.76	1.89	-0.12	0.25	0.25
Ag. Napa[I]	627.552	5	0.984	Dekelia (GPS)	7.0/0.30	208	-1.15	0.55	1.27	-0.12	0.46	0.48
Ag. Napa[I]	627.552	30	0.984	Dekelia (GPS)	7.0/0.30	208	-1.15	0.55	1.27	-0.12	0.46	0.48
Ag. Napa[I]	627.552	5	0.984	Cyclades-1	6.9/0.0	23	-0.32	0.10	0.33	0.02	0.09	0.09
Ag. Napa[I]	627.552	30	0.984	Cyclades-1	6.9/0.0	23	-0.32	0.10	0.33	0.02	0.09	0.09
Ag. Napa[I]	627.552	5	0.984	Cyclades-2	4.05/0.0	34	-0.54	0.29	0.62	-0.01	0.06	0.06
Ag. Napa[I]	627.552	30	0.984	Cyclades-2	4.05/0.0	34	-0.54	0.29	0.62	-0.01	0.06	0.06
Ag. Napa[II]	661.208	5	0.967	Ag. Napa[I]	14.8/0.20	627.552	2.23	1.42	2.64	0.14	0.49	0.51
Ag. Napa[II]	661.208	30	0.967	Ag. Napa[I]	14.8/0.20	627.552	2.23	1.42	2.64	0.14	0.49	0.51
Ag. Napa[II]	661.208	5	0.967	Ag. Napa[I]	10/0.20	404.876	1.52	0.80	1.72	0.17	0.19	0.24
Ag. Napa[II]	661.208	5	0.967	Amathounta	5.57/0.10	5400	0.44	0.26	0.51	0.25	0.11	0.27
Ag. Napa[II]	661.208	30	0.967	Amathounta	5.57/0.10	5400	0.44	0.26	0.51	0.25	0.11	0.27
Ag. Napa[II]	661.208	5	0.967	Dekelia	10.1/0.09	101.887	1.72	0.76	1.89	-0.23	0.28	0.32
Ag. Napa[II]	661.208	30	0.967	Dekelia	10.1/0.09	101.887	1.72	0.76	1.89	-0.22	0.28	0.32
Ag. Napa[II]	661.208	5	0.967	Dekelia (GPS)	7.0/0.30	208	-1.15	0.55	1.27	-0.27	0.48	0.51
Ag. Napa[II]	661.208	30	0.967	Dekelia (GPS)	7.0/0.30	208	-1.15	0.55	1.27	-0.25	0.48	0.51

Ag. Napa[II]	661.208	5	0.967	Cyclades-1	6.9/0.0	23	-0.32	0.10	0.33	-0.08	0.03	0.08
Ag. Napa[II]	661.208	30	0.967	Cyclades-1	6.9/0.0	23	-0.32	0.10	0.33	-0.08	0.03	0.08
Ag. Napa[II]	661.208	5	0.967	Cyclades-2	4.05/0.0	34	-0.54	0.29	0.62	-0.02	0.25	0.25
Ag. Napa[II]	661.208	30	0.967	Cyclades-2	4.05/0.0	34	-0.54	0.29	0.62	-0.02	0.25	0.25
Amathounta	5400	100	-	Ag. Napa[I]	14.8/0.20	627.552	2.23	1.42	2.64	-0.10	0.45	0.47
Amathounta	5400	100	-	Ag. Napa[I]	10/0.20	404.876	1.52	0.80	1.72	-0.14	0.20	0.25
Amathounta	5400	100	-	Ag. Napa[II]	14.7/0.30	661.208	2.23	1.42	2.64	-0.26	0.49	0.56
Amathounta	5400	100	-	Ag. Napa[II]	10/0.30	576.803	1.52	0.80	1.72	-0.25	0.44	0.43
Amathounta	5400	100	-	Dekelia	10.1/0.09	101.887	1.72	0.76	1.89	0.02	0.26	0.25
Amathounta	5400	100	-	Dekelia (GPS)	7.0/0.30	208	-1.15	0.55	1.27	-0.02	0.46	0.47
Amathounta	5400	100	-	Cyclades-1	6.9/0.0	23	-0.32	0.10	0.33	0.07	0.08	0.10
Amathounta	5400	100	-	Cyclades-2	4.05/0.0	34	-0.54	0.29	0.62	0.05	0.19	0.20
Merged	11873	100	-	Ag. Napa[I]	14.8/0.20	627.552	2.23	1.42	2.64	0.13	0.45	0.47
Merged	11873	100	-	Ag. Napa[I]	10/0.20	404.876	1.52	0.80	1.72	0.01	0.20	0.20
Merged	11873	100	-	Ag. Napa[II]	14.7/0.30	661.208	2.23	1.42	2.64	-0.06	0.50	0.51
Merged	11873	100	-	Ag. Napa[II]	10/0.30	576.803	1.52	0.80	1.72	-0.14	0.31	0.34
Merged	11873	100	-	Amathounta	5.57/0.10	5400	0.44	0.26	0.51	0.00	0.18	0.18
Merged	11873	100	-	Dekelia	10.1/0.09	101.887	1.72	0.76	1.89	-0.19	0.24	0.31
Merged	11873	100	-	Dekelia(GPS)	7.0/0.30	208	-1.15	0.55	1.27	-0.17	0.46	0.50
Merged	11873	100	-	Cyclades-1	6.9/0.0	23	-0.32	0.10	0.33	-0.01	0.09	0.09
Merged	11873	100	-	Cyclades-2	4.05/0.0	34	-0.54	0.29	0.62	-0.02	0.19	0.21
Dekelia	101.887	100	-	Ag. Napa[I]	14.8/0.20	627.552	2.23	1.42	2.64	0.13	0.51	0.55
Dekelia	101.887	100	-	Ag. Napa[I]	10/0.20	404.876	1.52	0.80	1.72	0.17	0.19	0.25
Dekelia	101.887	100	-	Ag. Napa[II]	14.7/0.30	661.208	2.23	1.42	2.64	-0.15	0.56	0.58
Dekelia	101.887	100	-	Ag. Napa[II]	10/0.30	576.803	1.52	0.80	1.72	-0.23	0.50	0.55
Dekelia	101.887	100	-	Amathounta	5.57/0.10	5400	0.44	0.26	0.51	-0.21	0.18	0.28
Dekelia	101.887	100	-	Cyclades-1	6.9/0.0	23	-0.32	0.10	0.33	-0.01	0.09	0.09

Dekelia	101.887	100	-	Cyclades-2	4.05/0.0	34	-0.54	0.29	0.92	-0.01	0.21	0.21
				Ag. Napa[I]	14.8/0.20	627.552	2.23	1.42	2.64	0.15	0.33	0.36
				Ag. Napa[I]	10/0.20	404.876	1.52	0.80	1.72	0.11	0.21	0.24
				Ag. Napa[II]	14.7/0.30	661.208	2.23	1.42	2.64	0.09	0.49	0.50
				Ag. Napa[II]	10/0.30	576.803	1.52	0.80	1.72	-0.11	0.24	0.25
150m[DTM1]	129.552	5	-	Amathounta	5.57/0.10	5400	0.44	0.26	0.51	0.04	0.13	0.14
				Cyclades-1	6.9/0.0	23	-0.32	0.10	0.33	-0.05	0.06	0.07
				Cyclades-2	4.05/0.0	34	-0.54	0.29	0.62	-0.05	0.12	0.13
				Dekelia	10.1/0.09	101.887	1.72	0.76	1.89	-0.16	0.14	0.18
				Dekelia(GPS)	7.0/0.30	208	-1.15	0.55	1.27	-0.30	0.45	0.56
				Overall Average			0.71	0.66	1.40	-0.05	0.28	0.31
				s			1.22	0.45	0.83	0.13	0.16	0.16

Table 5.9: The results of the comparisons between the predicted models for all tests performed. \bar{x} is the average distance of the point cloud from the true and s its standard deviation. Negative values suggest overestimation of the depth and positive suggest underestimation.

In general, synthetic data present several advantages over the real data; once the synthetic environment is created, it is fast and cheap to produce as much data as needed for training a model. Additionally, synthetic data can have perfectly accurate labels, including labeling that may be very expensive or impossible to obtain by hand. This advantage is crucial in the cases of seabed mapping, where true data suffer from additional errors, introducing uncertainty to the model. If results are not satisfactory enough, the synthetic environment can be modified to improve the model and training. Finally, synthetic data can be used as a substitute for certain real data segments that contain, e.g., sensitive information. This usage of synthetic data has been adopted for several computer vision applications [Peng 2015, Sankaranarayanan 2018, Saleh 2018, Richardson 2016, Barbosa 2018], however their use in remote sensing applications is limited.

Following the model trained on the synthetic dataset, the models trained on the Agia Napa [Part I] and the Merged datasets are delivering the next best statistical indices. Results are also acceptable from the models trained on the rest of the datasets.

It should also be noticed that when the Dekelia (GPS) dataset is tested, it delivers mean distances very close to the mean distances calculated for the Dekelia test site, however, when the standard deviations and the *RMSEs* are compared, they are constantly larger by 0.20m to 0.25m indicating a biased dataset that introduces external errors in the process. However, this is not the case for the Dekelia test site, proving that the results are not related with the trained models.

It is also important to note that the large distances between the clouds observed in Figure 5.31 disappeared. This improvement was observed in every test performed, proving that the proposed method based on the SVR model achieves a great reduction of the errors caused by the refraction in the seabed point clouds.

Seabed Cross Sections

Several differences observed between the image-based point clouds and the true data, increasing the mean distance, the standard deviation and the *RMSEs*, are not due to the depth correction method which is under evaluation. To that direction, cross sections of the seabed were generated for the test sites where dense true data were available (i.e. LiDAR data), in order to prove the performance of the proposed method, highlight and exclude differences between the compared point clouds that are not related to the tested method. In Figure 5.33 the footprints of the representative cross sections in the Agia Napa [Part I], Amathounta and Dekelia test areas are demonstrated, together with three parts of the section (top, middle, and end).

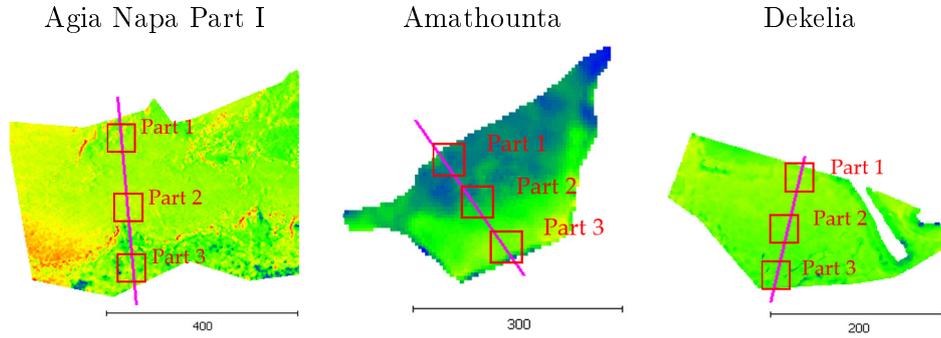


Figure 5.33: The footprints of the representative cross sections. Scalebars are in meters (m).

To demonstrate the performance of the proposed Method 1, for all three test sites used for generating the cross-sections, it was considered appropriate to use the point clouds having the elevations predicted from some of the worst cases presented in Table 5.9. As such, for the sections generated on the Agia Napa [Part I] test site, the elevations predicted using the model trained on the 30% of the Agia Napa Part [II] were used; for the sections generated on the Amathounta test site the elevations predicted using the model trained on Dekelia dataset were used; and for the sections generated on the Dekelia test site the elevations predicted using the model trained on Merged dataset were used.

In Figure 5.34, where the sections are presented, the blue line corresponds to the water surface, while the green one corresponds to the LiDAR data. The cyan line is the corrected depth after the application of the proposed method, while the red line corresponds to the depths derived from the initial uncorrected image-based point cloud. The first row of the cross sections represents Part 1, the second Part 2, and the third Part 3.

These parts highlight the high performance of the algorithm but also the differences between the point clouds, not related with the proposed method, as reported above. Specifically, in the first and the second part of the section presented, it can be noticed that even if the corrected image-based point cloud almost matches the LiDAR one on the left and the right side of the sections, in the middle parts, great differences are observed. These abrupt changes are caused by gross errors in the true data generation i.e. poor stitching among LiDAR flight lines, which however, are not related to the depth correction method. However, in the third part of the section, it is obvious that even when the depth reaches 14 m, the corrected image-based point cloud matches the LiDAR one, indicating a very high performance of the proposed Method 1. Excluding these differences, the corrected image-based point cloud presents deviations of less than 0.05m (0.36% remaining error at 14m depth) from the LiDAR point cloud.

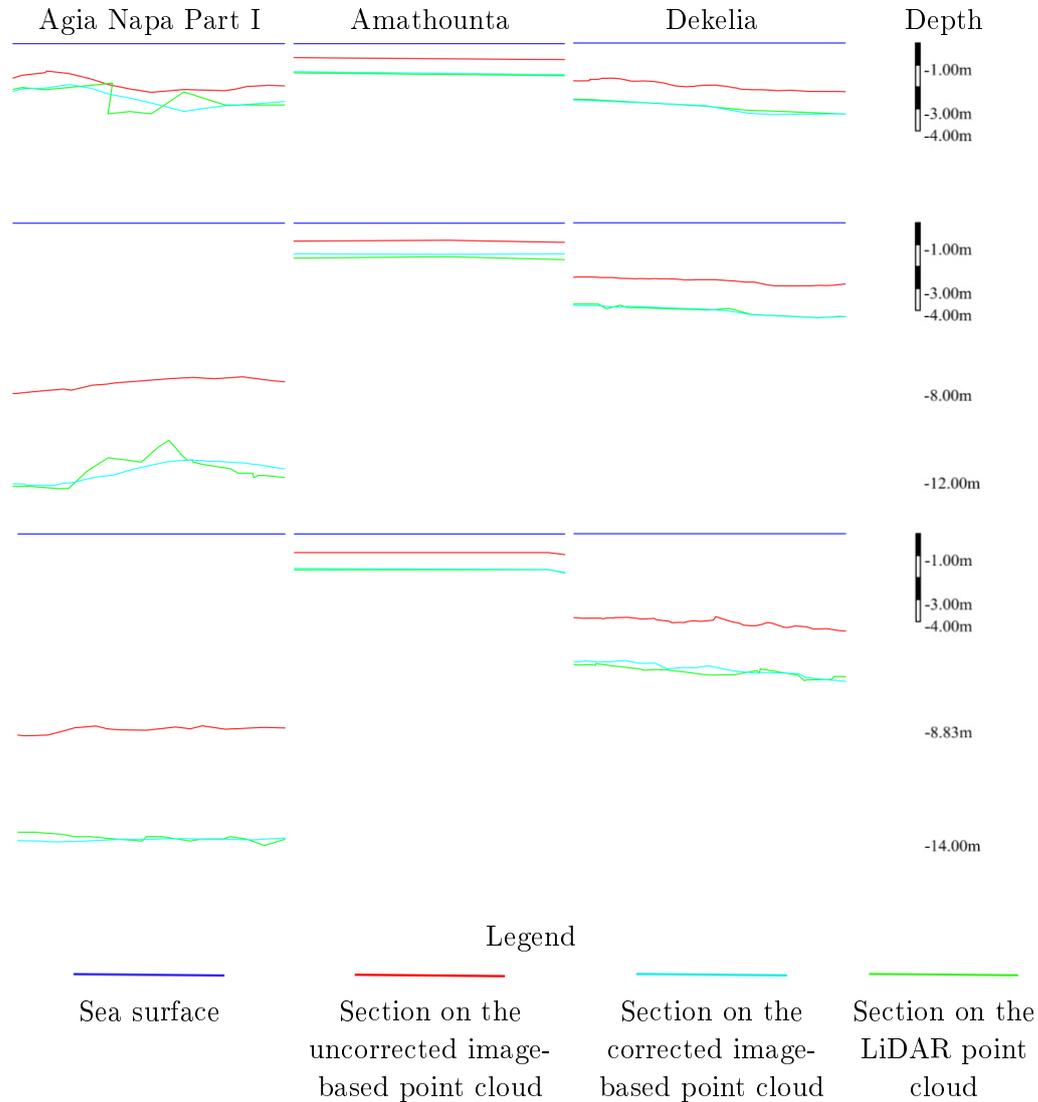


Figure 5.34: The footprints of the representative cross sections. Scalebars are in meters (m).

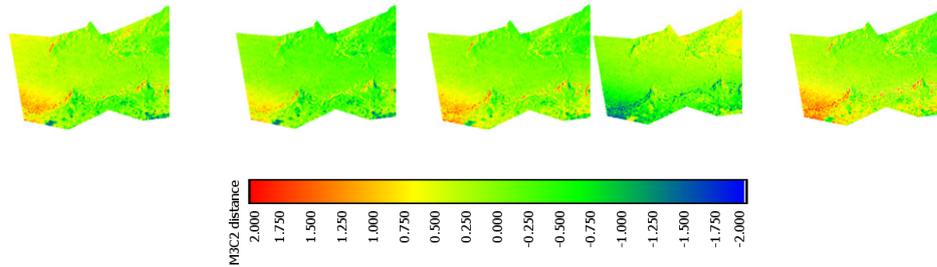
Distribution Patterns of Remaining Errors

In order to demonstrate the distribution patterns of the remaining errors, the multi-scale cloud to cloud distances (M3C2) between the true depth point clouds and the point clouds that are presented in Figure 5.31 and Table 5.9 are visualized for the test sites where dense true data were available (i.e. LiDAR data) in Figure 5.35.

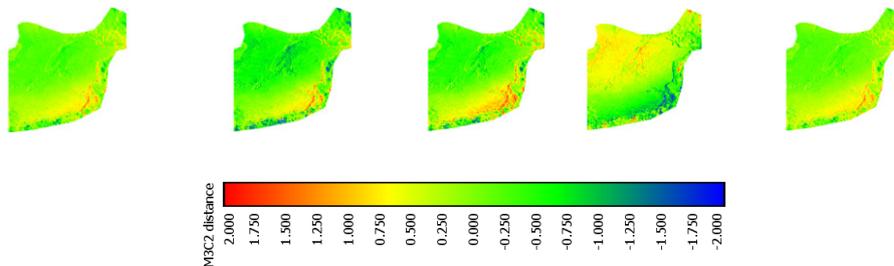
The larger differences between the predicted and the LiDAR depths are highlighted with reddish and blueish colors and are observed in certain specific areas, or areas with same characteristics. In more detail, the lower-left area of Agia Napa [Part I] test site and the lower-right area of Agia Napa [Part II] test site, present constantly larger errors than other areas of the same depth. This can be explained by their unfavorable position in the photogrammetric block, as these areas (i) are

far from the control points situated on the shore, and (ii) considering the entire Agia Napa test site, they lie in the periphery of the block. Additionally, they are imaging the deeper areas of the site which are affected the most by the refraction effect, where, as proved before and in combination with (ii), matches are very likely to be erroneous. However, it is noticeable that these two areas presented smaller deviations from the LiDAR point cloud when the model was trained in Amathounta test site, a totally characteristically different and shallower test site. Additionally, areas with small rock formations also presented differences. This is attributed to the different level of detail in these areas between the LiDAR point cloud and the image-based one, since LiDAR's average point spacing is about 1.1m. These small rock formations in many cases led M3C2 to detect larger distances in those parts of the site and were responsible for the increased standard deviation of the M3C2 distances presented in Table 5.9. Regarding the Dekelia test site, larger differences were observed in the area around the breakwater, something that was expected due to its construction characteristics (Figure 5.20c), and due to the fact that in this area the sea is quite turbid. Moreover, when the SVR model is trained on the Dekelia dataset, it is observed that the M3C2 distances are of the opposite sign, compared with the rest of the results. This is explained by the different behavior of the model, also reported in the annotation of Figure 5.28c.

Training Site for Testing on Agia Napa Part I
 100% of the points
 Ag. Napa [II] 30% Amathounta [all] Merged Dekelia "150m[DTM1] 5%"



Training Site for Testing on Agia Napa Part II
 100% of the points
 Ag. Napa [I] 30% Amathounta [all] Merged Dekelia "150m[DTM1] 5%"



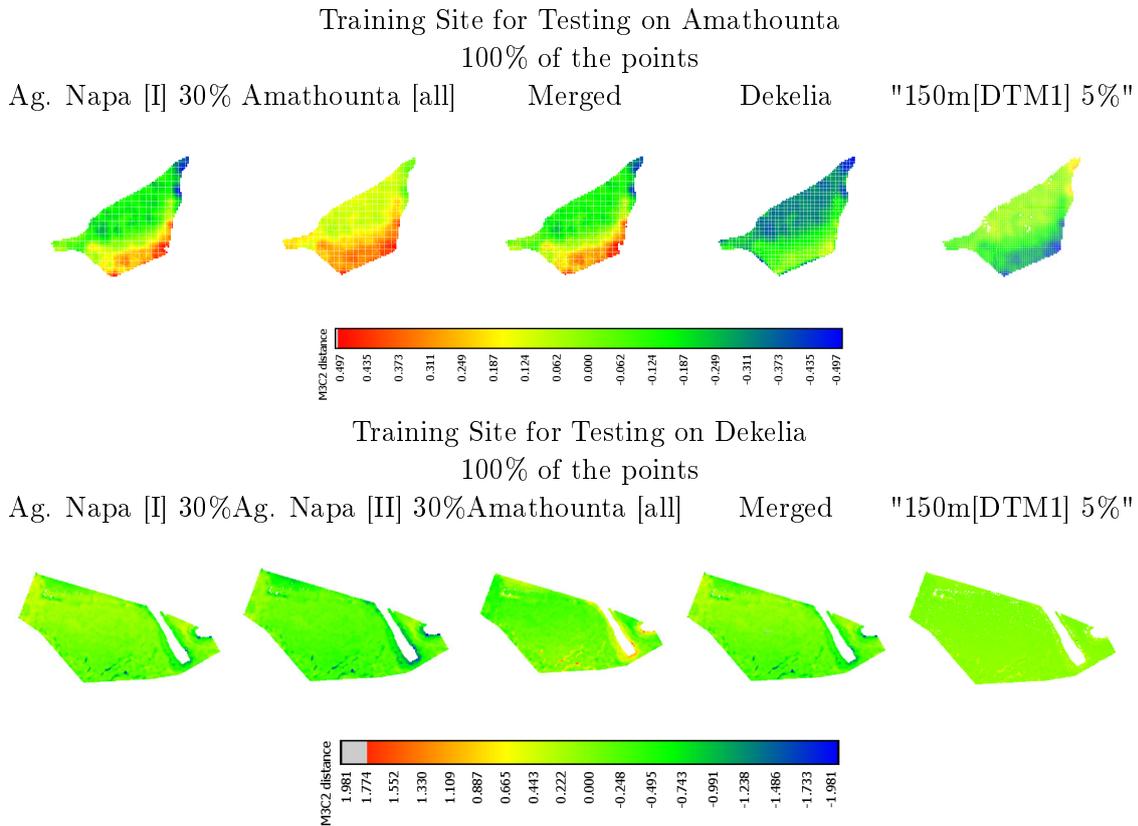


Figure 5.35: The cloud to cloud (M3C2) distances between the LiDAR point cloud and the corrected point clouds after the implementation of the proposed method, demonstrating the spatial patterns of the remaining errors. M3C2 distances are in (m).

5.4.1.4 Testing and validating the Method 1 by fusing the corrected image-based with LiDAR point clouds

To demonstrate the performance of the proposed Method 1 and highlight possible ways of exploitation of the results achieved, fusion of the LiDAR and the depth-corrected image-based point clouds was performed. The results are presented and evaluated with qualitative and quantitative methods. To that direction, examples of color transferring from the image-based point cloud to the LiDAR point cloud are presented, together with hole filling actions in both point clouds. For this approach, the Agia Napa test area was used, since due to its coverage and its depth, it presents many qualities for demonstrating various results.

Background and Related Work

Many issues arise regarding the fusion of the LiDAR and image-based point clouds and specifically those produced from aerial imagery, while facing the refraction effect. Bottom point cloud data resulting from bathymetric LiDAR systems deliver

geolocalized point clouds and correct depth information after refraction correction. However, in most of the cases, these data are quite sparse for underwater archaeological or biological research, missing required density and RGB color information.

Nowadays, this color information is very important for achieving reliable results from many machine learning and deep learning classification approaches. Moreover, LiDAR data in the very shallow nearshore zone ($< 2\text{m}$ depth) suffer from the difficulty of extraction of the surface and bottom positions, which are typically mixed in the green LiDAR signal [Allouis 2010], even though recent approaches [Schwarz 2019] are promising to meet this challenge. On the contrary, although UAV image-based point clouds generated through SfM-MVS processes are accompanied with color information, they suffer from the refraction effect described in Section 2.2. Moreover, as demonstrated in Subsection 5.3.2, the image-based methods for bottom point cloud generation, are not achieving great results in sandy bottom and non-static areas covered with seagrass since the texture of the images in these areas is quite poor and area based least squares matching and correlation algorithms fail, leading to the lack of correspondences in these areas. Additionally, sun glint on the water surface, may compromise the results. However, they are carrying much information important for exploiting them in recent classification and semantic segmentation approaches of bottom morphology, marine flora and fauna and coral reefs [van den Bergh 2017, Chirayath 2019].

Fusion of these two types of data will facilitate a better sensing and understanding of the bottom of the non-turbid shallow water bodies, exploiting the complementary attributes of both point cloud sources. However, till now, by not dealing with the refraction's geometric effects and consequently the erroneous depth information on the initial image-based point clouds with a robust and reliable way, fusion is impossible due to the incompatibility of the point clouds.

Even though aerial point cloud fusion tends to be a well-established methodology [Cabezas 2014] for inland and dry scenes, few works exist in the literature addressing bottom point cloud fusion using at least one overwater source of data such as airborne LiDAR or UAV imagery [Alvarez 2018, Legleiter 2012, Coleman 2011, Cheng 2014, Leon 2013]. Even though these works are dealing with fusion of bathymetric data, they are not dealing with the fusion of UAV image-based point clouds resulting from SfM-MVS processes with LiDAR point clouds. Most probable reason for this might be the strong geometric effects of refraction which affect the image-based bathymetry, leading to great errors and as a result great differences and depth inconsistencies comparing to the LiDAR data.

Color transfer to LiDAR data

Visual information is crucial for a number of applications. To that direction, a simple interpolation approach was followed to assign color values to the LiDAR point clouds. For every point of the LiDAR point cloud, the color value of the nearest neighbor in the corrected image-based point cloud is assigned. In Figure

5.36a the corrected image-based point cloud is presented while in Figure 5.36b the non-colored LiDAR point cloud is shown. In Figure 5.36c, the resulting colored LiDAR point cloud of the Agia Napa test area is presented.

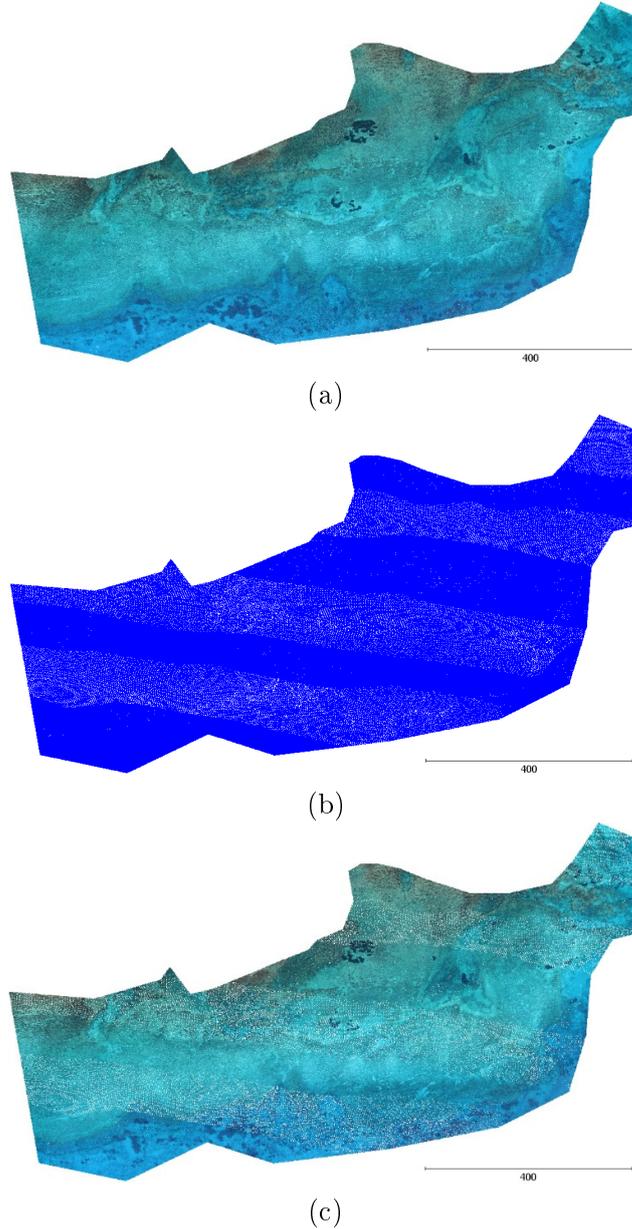


Figure 5.36: (a): the corrected image-based point cloud, (b): the non-colored LiDAR point cloud and (c): the colored LiDAR point cloud based on (a).

The first two columns of Figure 5.37 depict an area of Agia Napa[I] dataset, while the third one depicts an area of Agia Napa [Part II] dataset. The first row of images in Figure 5.37 depict the three sample areas on the image based point cloud, the second row of images demonstrate the same areas on the LiDAR point cloud colored

with the corrected image-based point cloud and the third row of images depict the three areas on the LiDAR point cloud colored with the uncorrected image based one.

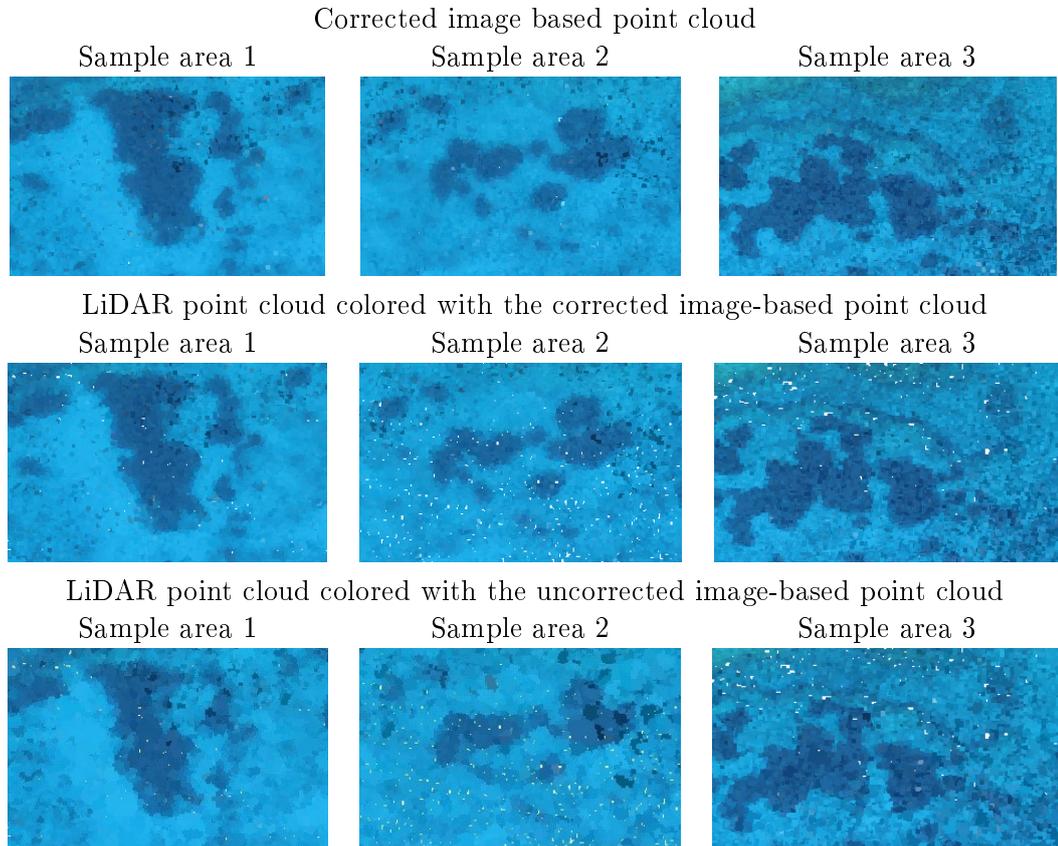


Figure 5.37: Three different sample areas of the Agia Napa seabed with depth between 9.95m to 11.75m. The first two columns depict two sample areas of the Agia Napa[I] seabed while the third depicts a sample area of the Agia Napa [Part II] seabed. The first row depicts the three areas on the image-based point cloud, the second row on the LiDAR point cloud colored with the corrected image-based point cloud and the third row depicts the three areas on the LiDAR point cloud colored using the uncorrected image-based one. Point size is multiplied by a factor of 4 in order to facilitate visual comparisons.

As can be easily noticed, the color of the seabed point clouds is almost identical in the first two rows, while certain areas in the third row present many differences, fluctuations, and noise. In more detail, by observing the total of the colored seabed point clouds, it was noticed that as the depth increases colors are assigned in a wrong way, something that was expected due to the large distance between the two point-clouds. This leads to a wrong representation of the seabed in depth of more than 2 to 3 meters. As can be noticed in Figure 5.37, not only the resulted assigned colors create a more rough and unclear seabed, but seagrass and rocky areas are

represented in a totally different shape and area affecting future exploitation and processing of the point cloud datasets such as seabed classification, orthoimage generation etc.

Seamless Hole filling

Seamless hole filling of seabed areas is performed and demonstrated in Agia Napa test area. To that direction, an area of the corrected and the uncorrected image-based point cloud, covered with seagrass and lying at 11.70m depth was deleted to create an artificial hole on the point cloud. Then this area was filled with the respective LiDAR point cloud. In Figure 5.38a, the seagrass area deleted to create the desirable hole is depicted with the yellow polyline. In Figure 5.38b, the image-based point cloud is demonstrated with red color while the LiDAR points used for filling are shown in blue color. The yellow rectangular describes the area of sections shown in Figure 5.38e and Figure 5.38f.

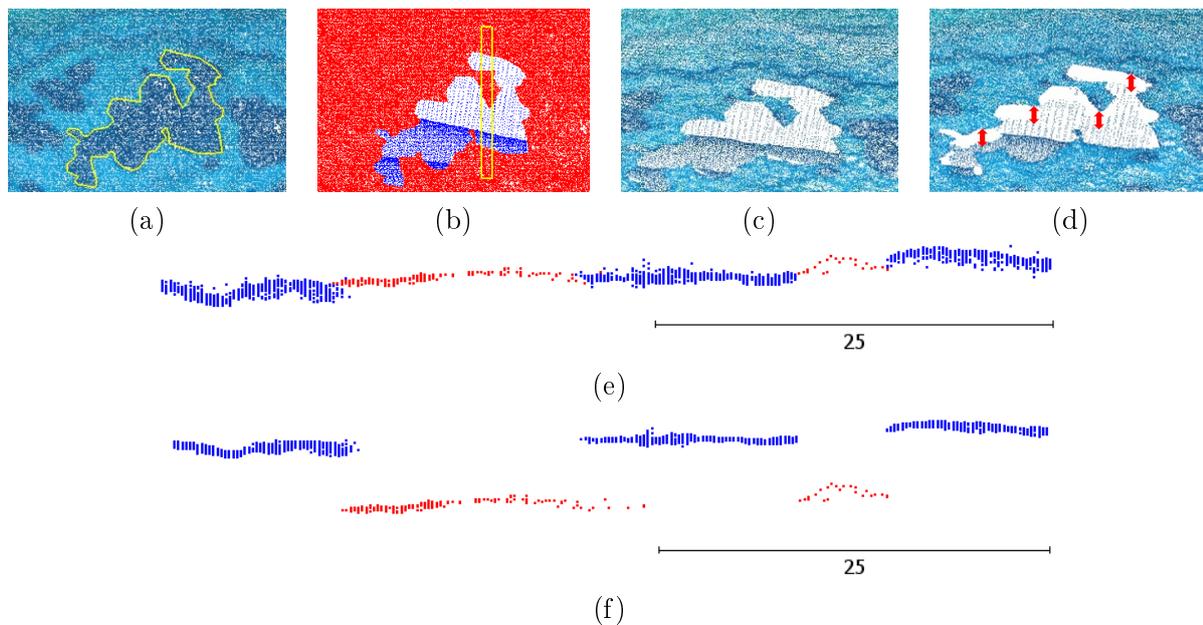


Figure 5.38: An example of the seamless hole filling area and indicative results.

In Figure 5.38c the hole on the corrected image-based point cloud is filled with the LiDAR data while in (d) the hole in the uncorrected image-based point clouds is filled with the LiDAR data. In the first case Figure 5.38c, it is obvious that the two point clouds are at the same depth and facilitate a seamless hole filling, even at the depth of 11.70m. However, in Figure 5.38d it is obvious that the depth difference of the two point clouds (highlighted with red arrows in Figure 5.38d) creates a huge obstacle in merging the two point clouds and having a seamless result. To further prove the above, cross sections were created of the area in the yellow rectangle in Figure 5.38b.

In the two generated cross sections of the point clouds demonstrated in Figure

5.38e for the corrected image based point cloud and in Figure 5.38f for the uncorrected image based point cloud, the LiDAR point cloud is in red color while the image based one in blue. It is clear that the proposed algorithm corrects the image-based point cloud to a degree facilitating the merging and the fusion of the point clouds of these two different sources.

5.4.1.5 Summary

In the above tested and validated Method 1 using real world data, several SVR models were trained based on known depth observations from bathymetric LiDAR, dipping tape and total station surveys, facilitating the estimation with high accuracy of the real depths of 3D point clouds derived from conventional SfM-MVS procedures using low altitude aerial imagery. State-of-the-art approaches are using non-sophisticated regression methods with maximum some few hundreds of sparsely collected points of maximum depths of no more than 2m to 3m. On the contrary this method, in order to achieve a transferable and generic solution, exploited the correlation of hundreds of thousands of points. These points included also deeper areas, in order to avoid external and gross errors (See also 5.4.1.3) and deliver models that fit every low altitude aerial campaign.

Several differences in the Z and Z_0 distribution encountered, affecting the SVR model training in specific test sites were discussed in detail and justified properly revealing that most of them are attributed to the quality of the true data. Moreover, other issues affecting the SfM-MVS performance i.e. the camera calibration and the restricted visibility due to scattering, absorption and refraction at larger depths were discussed in detail and their effects on the generated data were thoroughly explained.

Experimental results over five test areas in Cyprus and Greece, characterized by different flying heights, different depths, and different seabed characteristics, along with the quantitative validation performed, indicated the high potential of the developed method. Results suggested that the model trained on the synthetic dataset outperforms the models trained on the real-world datasets. This successful transfer learning approach was expected due to the increased quality of the synthetic training data, compared with the real-world data. However, the results of the models trained on the real-world data, closely follow in performance and achieved accuracies. As such, it was proved that the proposed method meets and exceeds the accuracy standards generally accepted for hydrography, as introduced by the IHO, where in its simplest form, the vertical accuracy requirement for shallow water hydrography can be set as a total of $\pm 0.25\text{m}$ (one sigma) from all sources, including tides [Guenther 2000]. Moreover, considering IHO's white paper on the regulations for international charts and charts' specifications, the proposed method is qualified to the maximum Zone of Confidence (ZOC) level A1 [IHO 2019]. Finally, the performed fusion of the LiDAR and the depth-corrected image-based point clouds, indicated the performance of the proposed method and highlighted possible ways of exploitation of the results achieved.

5.4.2 Correcting the Geometric Effects of Refraction in the 2D (Image) Space

The image space refraction correction method presented in Section 4.2 (Method 2) is implemented in the already described real world datasets, for real world shallow bathymetry mapping tasks, ensuring also optimal water surface state and water visibility conditions. To apply the presented method in the real-world datasets for correcting the original images which are affected by the refraction effect, several steps should be performed that are analyzed in the subsequent sections.

5.4.2.1 Data preparation

SVM based refraction correction on the dense point clouds

As already described in Section 4.2, where the proposed method is described, to correct the apparent depth of each point of the initial bottom dense point clouds for the test sites and produce the recovered dense point clouds, the Support Vector Regression (SVR) method for refraction correction of the point clouds described in Chapter 3 is exploited. As proved by the results presented in Subsection 5.4.1, the trained SVR models have significant potential for generalization over different areas and they can be used when no ground truth data are available. As such, for the experiments presented here, the already trained SVR models on Dekelia and on the synthetic "150m[DTM1] 5%" dataset, are used to predict the correct depths of the dense point clouds for the rest four test areas: Amathounta, Agia Napa, Cyclades-1 and Cyclades-2. Both models are selected since they are totally independent from the rest of the four test sites used in the experiments, proving also the significant potential of generalization of the model presented in Chapter 3. It is important also to stress that the data used for training the Dekelia model, were obtained with a different camera than the one used in Cyclades-1 and Cyclades-2 test sites and various flight patterns.

This process, goes beyond currently available iterative approaches for image refraction correction (i.e. [Skarlatos 2018]) because the bathymetric information is available a-priori, facilitating faster and more accurate image correction.

Updated DSM generation using the corrected dense point clouds

After correcting the dense point clouds for refraction effects with the two different SVR models trained on the Dekelia and on the synthetic dataset, two new DSMs of the area were created based on the results of each model. To achieve that, the recovered dense point clouds were first merged with the dry land dense points. The updated DSMs were created directly in the geographic reference system where the camera positions were already calculated in order to facilitate later the calculation of ground intersections (X, Y, Z) using the projection center of the camera and the image coordinates of points. In Figure 5.39 the merged point clouds are illustrated with the same color-scale, while in Figure 5.40 the generated updated DSMs of the four test areas are presented. Results presented in these two figures are produced after the depth correction using the SVR model trained on the Dekelia dataset. By

visually comparing the seabed areas between Figure 5.39 and Figure 5.22c, d, g, h, the difference between the initial and the recovered depths can be observed. These differences once again confirm that the refraction effect cannot be ignored.

Regarding the updated DSMs generation, for the Amathounta test site (Figure 5.40a), a GSD of 0.13m was achieved, for the Agia Napa test site (Figure 5.40b) a GSD of 0.28 m, for Cyclades-1 (Figure 5.40d) a GSD of 0.12 m, and for Cyclades-2 (Figure 5.40e) a GSD of 0.07 m, equal to the mean distance between the points of the dense point clouds of the four test sites, respectively.

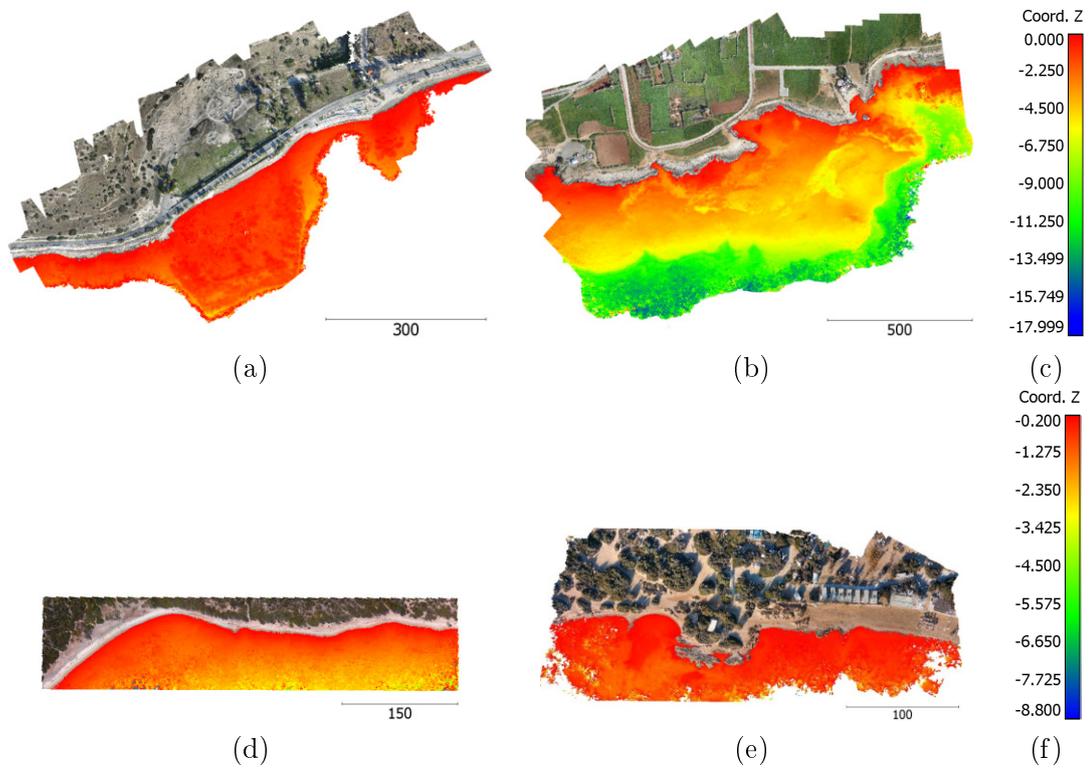


Figure 5.39: The extracted seabed point clouds after the depth correction using the SVR model trained on the Dekelia dataset for Amathounta (a), Agia Napa (b), Cyclades-1 (d) and Cyclades-2 (e) test sites. Scalebars visualizing the depths are in meters (m).

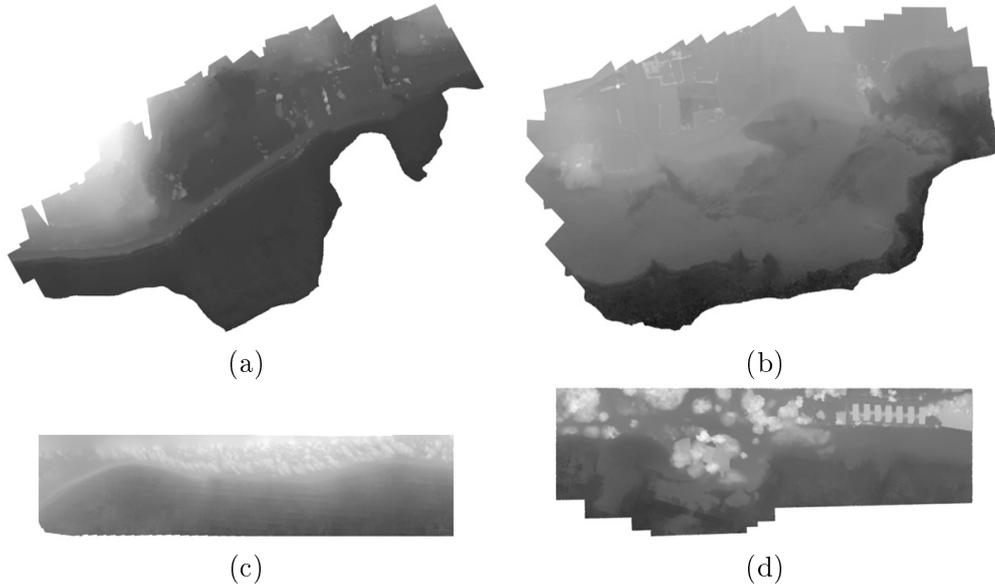


Figure 5.40: The updated DSM after the depth correction using the SVR model trained on the Dekelia dataset for Amathounta (a), Agia Napa (b), Cyclades-1 (c) and Cyclades-2 (d) test sites.

5.4.2.2 Refraction correction in the image space

Refraction-free images

Figure 5.41a and Figure 5.41b illustrate two sample images from the Agia Napa dataset along with the corresponding refraction-free ones in Figure 5.41d and in Figure 5.41e, after the depth correction using the SVR model trained on the Dekelia dataset. In a similar way, Figure 5.41c illustrates an original image from the Amathounta test site together with the corresponding refraction-free image in Figure 5.41f, after the depth correction using the SVR model trained on the Dekelia dataset. Since c_{mixed} is by default larger than c_{air} in all depth cases because $n_{water} > n_{air}$, an overall image shrinkage occurred, hence raw information was kept within the initial image frame, while areas with missing information due to shrinkage were kept black. This can be observed in Figure 5.41d, e and f zoomed areas. In Figure 5.41g, h and i, the pixel values' differences between the respective images of the first and the second row are illustrated.

By observing the corrected and the initial imagery, one could notice that differences are only apparent in underwater regions and that these differences are increasing with depth (from right to left, for all the images in Figure 5.41) as expected. Moreover, in all cases (Figure 5.41g, h and i), there is a small circular area (red dotted circle) in the nadiral point of each image, and, around it, the refraction effect is minimal. This area is bigger as the depth of the imaged seabed decreases (from the left image to the right image) since the effect of refraction is less in shallow areas than in deeper areas. This is also explained by the nature of the refraction phenomenon on the imagery [Georgopoulos 2012, Agrafiotis 2015].

Indeed, the refraction phenomenon on the imagery behaves like a radial shift, considering the central projection image formation model. The center of this area of the image, having almost no refraction, is the nadiral point of the image. This area of the image is created by light rays that are characterized by a very small incidence angle, which, according to Snell's law, leads to a very small refraction angle (Figure 4.6). Moreover, assuming a flat air-water interface and a perfectly perpendicular optical axis, the light ray that matches with the nadiral axis has a zero angle of incidence and thus is not affected by the refraction phenomenon. However, the latter is never achieved in a real-world dataset, such as the ones used here. On the contrary, the amount of correction and thus the differences between the images are increasing as the distance from the image center increases, since the amount of refraction is more intense for light rays captured far from the image center.

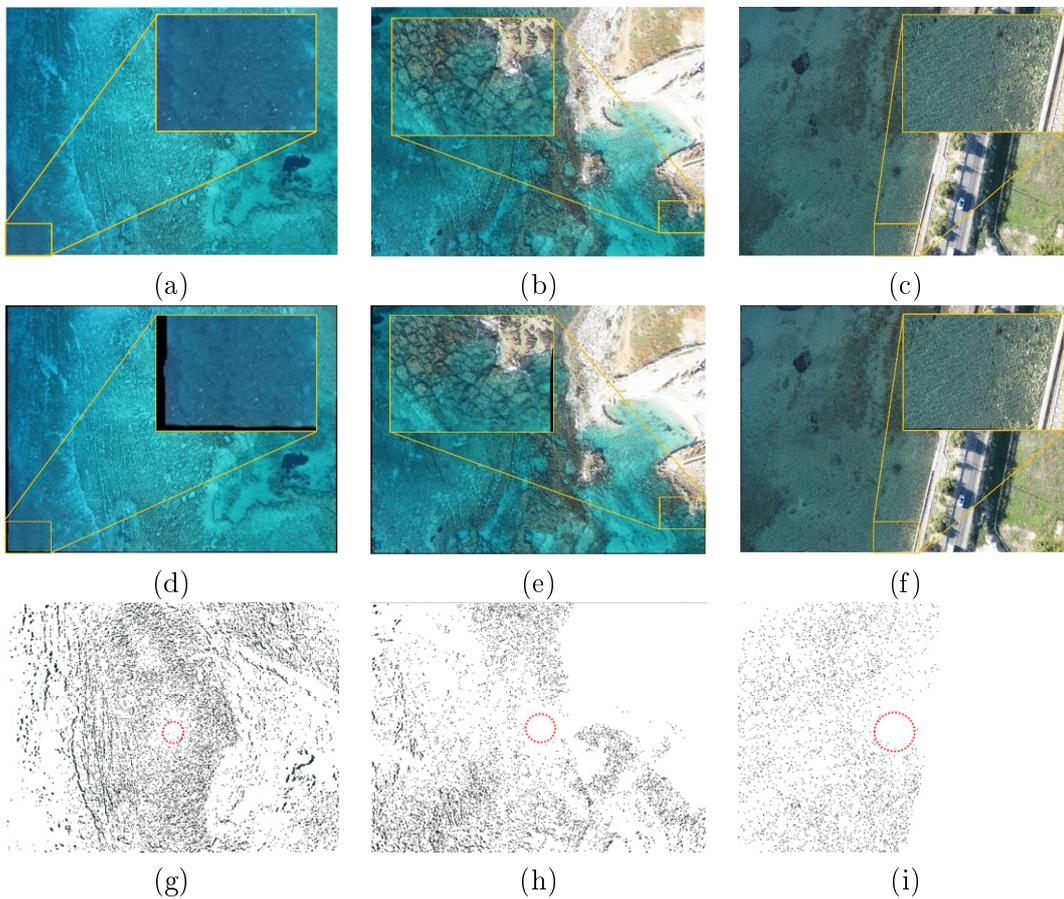


Figure 5.41: Initial aerial data (first row), refraction-free images (second row) after the depth correction using the SVR model trained on the Dekelia dataset and their subtraction (third row).

5.4.2.3 Quantitatively assessing the improvements on the SfM results

A well-established methodology for evaluating SfM performance is to compare the sparse point cloud to some ground truth with the same data representation [Bianco 2018]. Towards this end, the comparison of the sparse point clouds with the LiDAR data was performed for the Amathounta and Agia Napa test sites. Since for the image correction method (Method 2) a corrected dense point cloud was used, there was no need to evaluate the dense point cloud again. This way the positive effects of the image correction on the accuracy of the results and the performance of the image correction method was demonstrated. For the Cyclades-1 and Cyclades-2 test areas, since LiDAR data were not available, comparisons were performed using checkpoints measured with a geodetic total station on-site and acquired together with the image data. Together with the sparse point cloud produced by the corrected data, results from four state of the art refraction correction algorithms are presented and evaluated.

Quantitative assessment of the produced sparse point clouds

Since the dense point cloud regeneration after the correction of the imagery is not necessary, and there is no ground truth for the exterior orientations of the cameras, to evaluate the results of the proposed image correction method, the sparse point clouds of the updated SfM solutions were evaluated. To this end, the sparse point clouds generated using the initial imagery, and the sparse point clouds resulted from the corrected imagery using the proposed Method 2 (exploiting the SVR model trained on the Dekelia dataset and the SVR model trained on the synthetic dataset) were compared with the LiDAR point clouds or the Total Station points using the M3C2 [Lague 2013] module in Cloud Compare freeware [Girardeau-Montaut 2015] to demonstrate the changes and the differences produced by the proposed image correction method. The M3C2 algorithm offers accurate surface change measurement that is independent of point density [Lague 2013].

In Figure 5.42 (first row) it can be observed that the M3C2 distances between the reference data and the original image-based point clouds increase in proportion to the depth. In all the test cases demonstrated in the first row of Figure 5.42, the Gaussian mean of the differences (\bar{x}) is significant. It reaches 0.67m in the Amathounta test site, 1.71m in the Agia Napa test site, 0.32m in Cyclades-1 test site and 0.54m in Cyclades-2 test site (Table 5.10). These comparisons clarify that the refraction effect cannot be ignored in such applications. The dashed lines in red represent the accuracy limits generally accepted for hydrography as introduced by the IHO [Guenther 2000]. As can be observed in the first row of Figure 5.42, the majority of the points in both areas are far from the red dashed lines representing the accuracy limits generally accepted for hydrography as introduced by the IHO. The second row of Figure 5.42 presents the histograms of the M3C2 distances between the true values and the sparse point clouds generated from the corrected imagery of the image correction method presented in this thesis (Section 4.2) using the SVR model trained on the Dekelia dataset. The third row of Figure 5.41 presents the

histograms of the M3C2 distances between the true and the sparse point clouds produced from corrected imagery of the image correction method using the SVR model trained on the synthetic dataset.

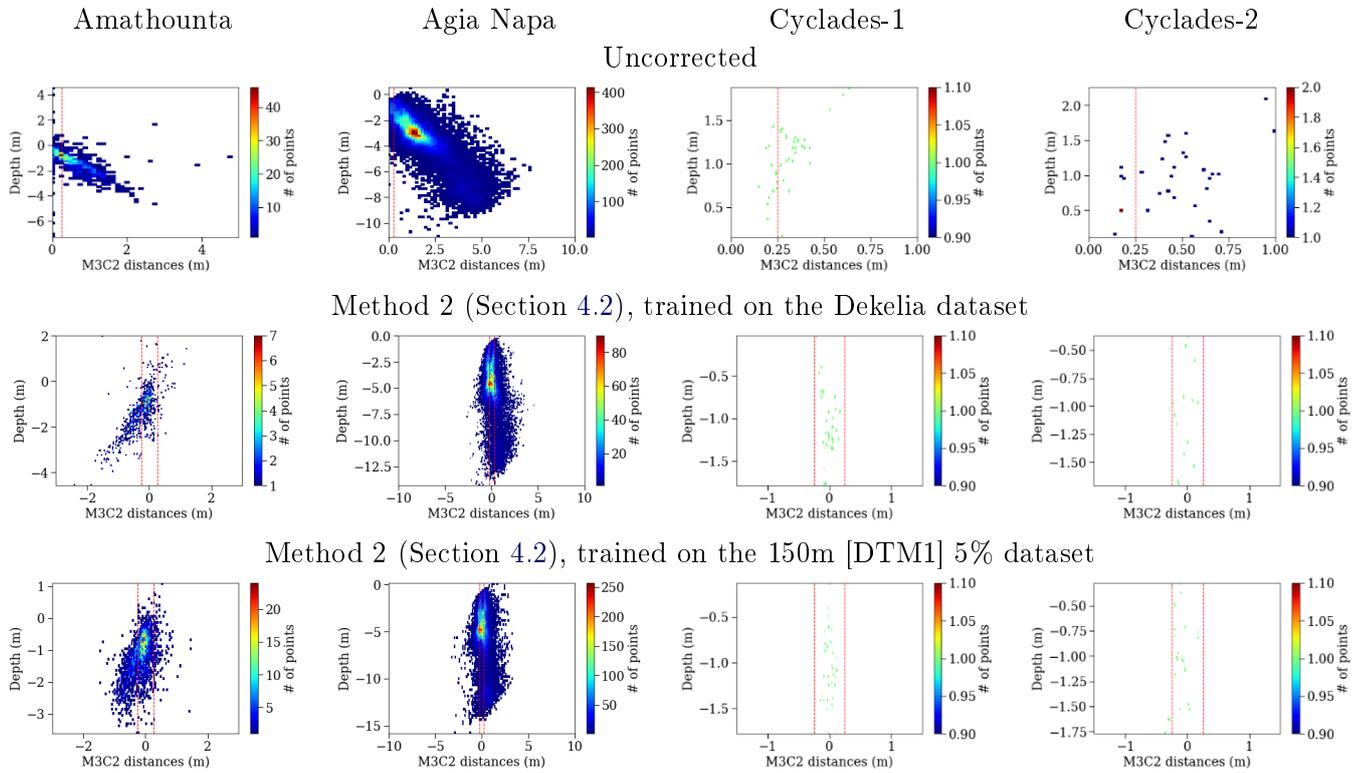


Figure 5.42: The 2D histograms of the M3C2 distances between the true and the uncorrected and corrected image-based sparse point clouds derived from the SfM for Amathounta (first coloumn), Agia Napa (second coloumn), Cyclades-1 (third coloumn) and Cyclades-2 (fourth coloumn) test sites respectively. The dashed lines in red represent the accuracy limits generally accepted for hydrography as introduced by the IHO [Guenther 2000]. M3C2 distances are in (m).

In Table 5.10 the results of all the tests performed are presented. Both in the second and third row of Figure 5.42 and Table 5.10, a great improvement in the depths of the sparse point clouds and thus the SfM accuracy is achieved by the image based refraction correction method (Method 2) presented in this thesis (see Section 4.2) compared with the results over the uncorrected ones, illustrated also in the first row of Figure 5.42. More specifically, when exploiting the SVR model trained on the Dekelia dataset, in Amathounta test site, the initial 0.67m mean distance is reduced to -0.19m, in Agia Napa test site the initial 1.71m mean distance is reduced to -0.04m, in Cyclades-1 the 0.32m initial mean distance is reduced -0.02m while in Cyclades-2 the 0.54m initial mean distance is reduced to -0.06m.

As expected, when exploiting the SVR model trained on the "150m [DTM1] 5%" dataset, the calculated statistical indices are better, confirming the high performance

of the model trained on the synthetic dataset and the successful transfer learning approach. In more detail, in Amathounta test site, the initial 0.67m mean distance is reduced to -0.04m, in Agia Napa test site the initial 1.71m mean distance is reduced to -0.04 m, in Cyclades-1 the 0.32m initial mean distance is reduced 0m while in Cyclades-2 the 0.54m initial mean distance is reduced to -0.07m. In this later case, even the mean distance is higher by 0.01 m, the standard deviation is reduced by 0.08m and the *RMSE* by 0.06m.

Again, these results include outlier points such as seagrass that are not captured in the true depth clouds for all the cases or are caused due to point cloud noise again in areas with seagrass or poor texture. For all the test sites, but especially for the results produced when the SVR model trained on the synthetic dataset is used, the resulting mean distances in the sparse point cloud meet and exceed the accuracy standards generally accepted for hydrography as introduced by the IHO, where in its simplest form, the vertical accuracy requirement for shallow water hydrography can be set as a total of $\pm 0.25\text{m}$ (one sigma) from all sources, including tides [Guenther 2000]. Moreover, considering IHO's white paper on the regulations for international charts and charts' specifications, the results of the proposed methods are qualified to the maximum Zone of Confidence (ZOC) level A1 [IHO 2019].

In Table 5.10, the results of the statistical analysis of the tests performed over the Agia Napa test site, till 10m depth are also presented. These results are better, compared with the results over the full test site that reaches the depth of 14.8m, however they do not differentiate enough, as it happened with the results of the Method 1 for the correction of the geometric effects of refraction presented in Table 5.8. The observed insignificant differences reveal a very important advantage of this second method for the correction of the geometric effects of refraction; by eliminating the geometric effects of refraction on the initial imagery, a more accurate SfM process is performed, supported by more matches in the depths of more than 10m.

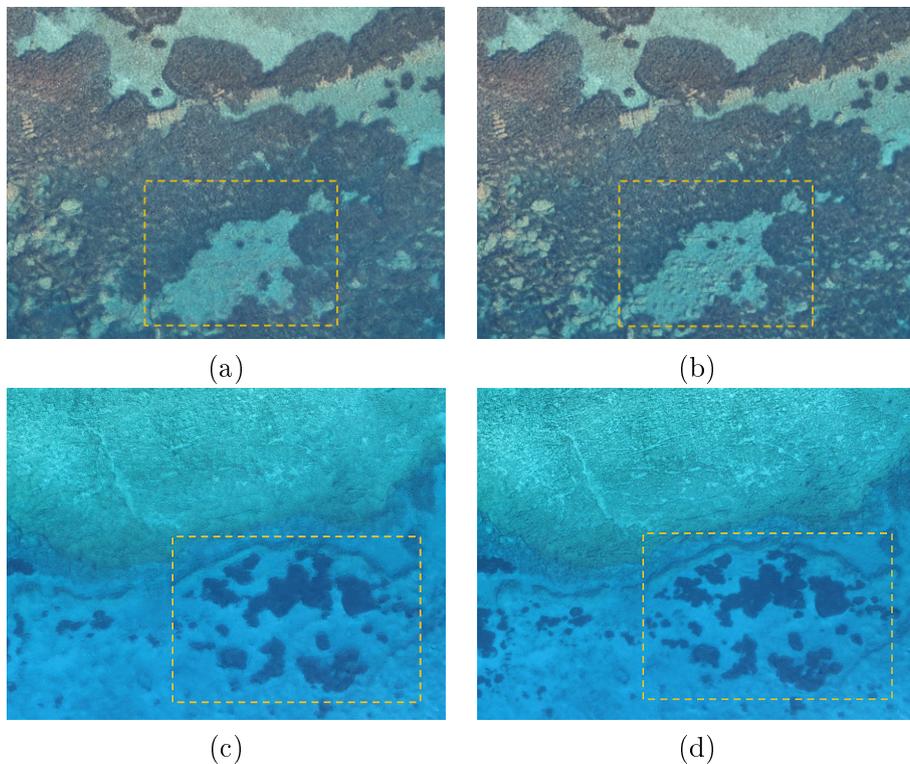
Test Site															
	Amathounta			Agia Napa			Agia Napa [10m]			Cyclades-1			Cyclades-2		
# of check points	1K			32K			21K			23			34		
Statistical Analysis [m]															
Derived point clouds from different methods	\bar{x} [m]	s [m]	RMSE [m]												
Uncorrected images	0.67	2.19	2.28	1.71	1.18	2.08	1.71	1.18	2.08	0.32	0.10	0.33	0.54	0.29	0.62
Method 2 (Dekelia)	-0.19	0.28	0.31	-0.04	0.37	0.38	0.01	0.37	0.37	-0.02	0.09	0.09	-0.06	0.14	0.15
Method 2 (150m [DTM1]5%)	-0.04	0.12	0.13	-0.04	0.24	0.24	-0.03	0.24	0.24	0.00	0.06	0.07	-0.07	0.06	0.09
IHO limit (1σ) [Guenther 2000]	± 0.25														

Table 5.10: Quantitative Evaluation. \bar{x} is the average distance of the point cloud from the true values and s is its standard deviation. Negative values suggest overestimation of the depth and positive suggest underestimation.

5.4.2.4 Qualitatively assessing the improvements on the textured seabed models

Figure 5.43 illustrates parts of the textured 3D models of Amathounta, Agia Napa, Cyclades-1, and Cyclades-2 test sites. There, the respective textured models created by a mesh, produced using the refraction corrected point cloud, and the original uncorrected imagery are presented in the left column in order to highlight that when textured models are needed, the correction of the seabed point cloud and the mesh is not enough and image correction is also necessary. In the right column of the same figure, the textured models created using the mesh, produced using the refraction corrected point cloud, and the corrected imagery implementing the proposed Method 2 are presented.

By comparing these results, it is clearly obvious, especially in the areas inside the orange dashed rectangle, that the details of the seabed on the textured models created using the uncorrected imagery are degraded, especially when depth increases (Figure 5.42c-d) or present detailed rock formations with abrupt changes in elevation (Figure 5.43b-d and Figure 5.43). This leads to a significant loss of information on the textured 3D model and to an unreliable representation of the seabed, highlighting the critical need to correct the original dataset from refraction and update the SfM solution. At the same time, as can be observed in Figure 5.43e-f, dry land areas are remaining unchanged, highlighting the effects of the refraction on the reliability of the created models of the seabed.



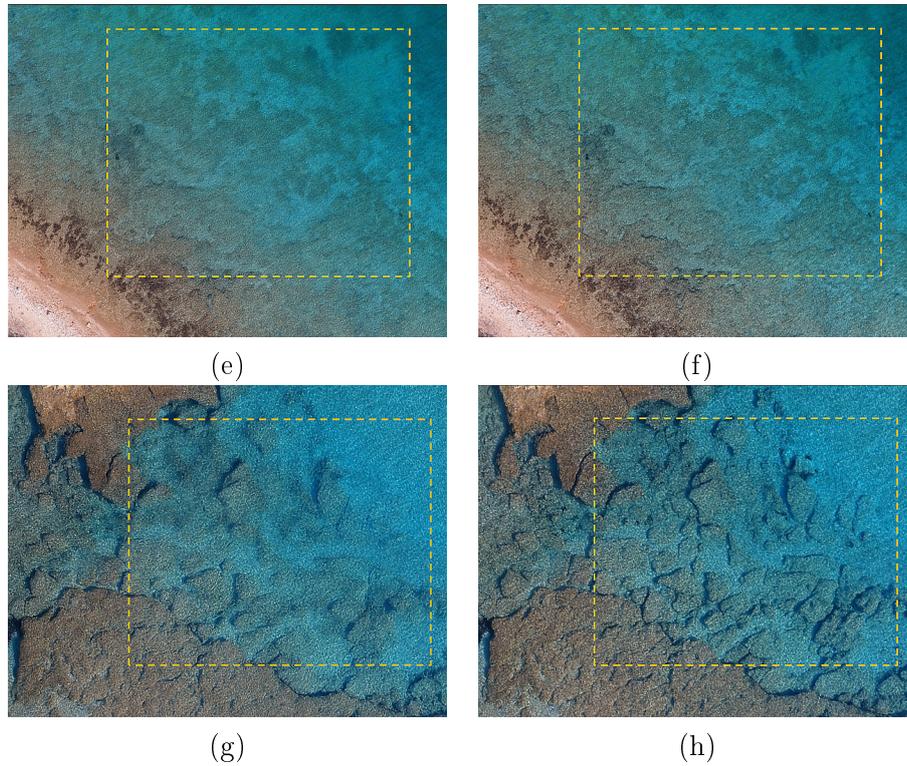


Figure 5.43: Indicative textured 3D models from Amathounta (first row), Agia Napa (second row), Cyclades-1 (third row), and Cyclades-2 (fourth row) test sites with the initial uncorrected (left column) and the refraction-free (right column) datasets. For the generation of the presented refraction-free datasets the SVR model trained on the Dekelia dataset was exploited.

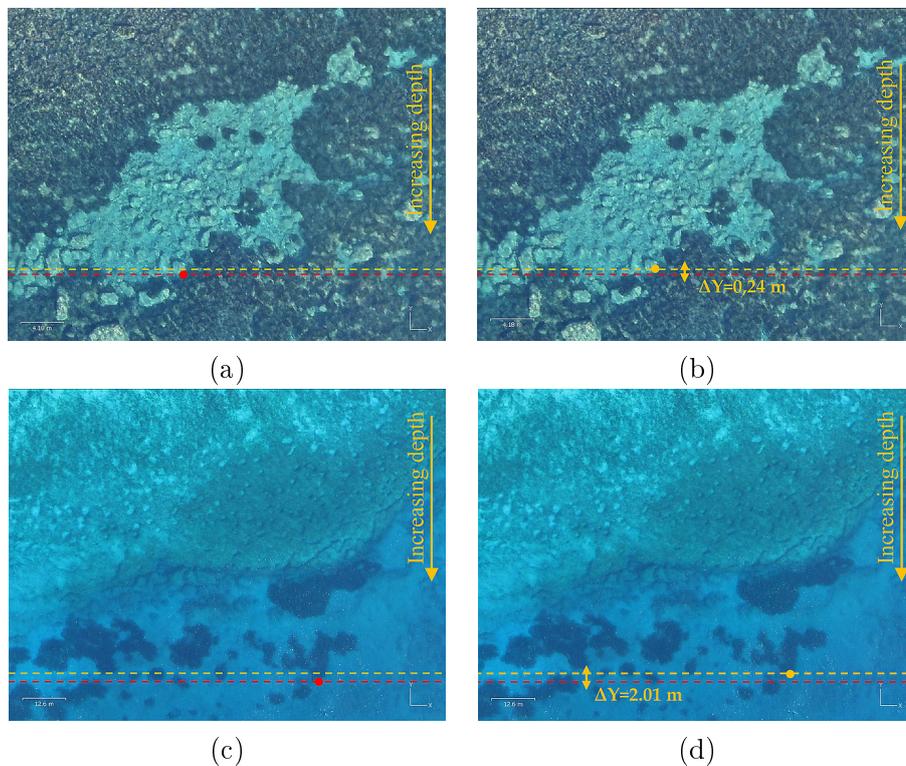
5.4.2.5 Quantitatively assessing the improvements on the orthoimages

Figure 5.44 illustrates parts of the orthoimages of Amathounta, Agia Napa, Cyclades-1, and Cyclades-2 test sites. Orthoimages created using the updated DSM and the original imagery. They are presented in the left column, while in the right column, orthoimages using the updated DSM and the corrected imagery are depicted. For the generation of the corrected imagery, the SVR model trained on the Dekelia dataset was exploited.

Dashed lines are used to facilitate the comparison between the left and the right column; red dashed lines pass from the red dots, which represent a specific point in the orthoimage of each test site produced using the uncorrected imagery. The orange dashed line passes from the orange dots, which represent the same specific point (being in red in the left column) in the orthoimage of each test site produced using the corrected imagery. The horizontal displacements of the marked points on the orthoimages are reported also in Figure 5.43. Contrarywise to the textured 3D models (Figure 5.42), in the orthoimages, no important qualitative differences can be spotted.

However, remarkable horizontal differences can be found in the direction that the depth increases. In more detail, in the Amathounta test site (Figure 5.43a,b), where the GSD of the orthoimage was 0.036 m/pix, differences of 5–8 pixels can be spotted in an average depth of 1.60m, leading to planimetric errors of 0.182m–0.291m. In the Agia Napa test site (Figure 5.43c,d), where the GSD of the orthoimage was 0.077m/pix, differences of 23–26 pixels can be spotted in a depth of 13.80m, leading to planimetric errors of 1.78m–2.07m. In the Cyclades-1 test site (Figure 5.43e,f), where the GSD of the orthoimage was 0.031m/pix, differences of 18–21 pixels can be spotted in a depth of 2.24m, leading to planimetric errors of 0.554m– 0.647m. Finally, in the Cyclades-2 test site (Figure 5.43g,h), where the GSD of the orthoimage was 0.021m/pix, differences of 6–17 pixels can be spotted in a depth of 1.98m, leading to planimetric errors of 0.122m–0.342m.

Results strongly suggest that when depth increases, planimetric errors on the orthoimages are increasing if no refraction corrections are applied to the images. In the shallower areas, these errors might not be obvious. However, in deeper areas, such as Agia Napa (Figure 5.43c,d), Cyclades-1 (Figure 5.43e,f), and Cyclades-2 (Figure 5.43g,h), the planimetric differences are evident and very important. The magnitude of these errors depends on the depth error and the position of the point on the image. These differences cannot be neglected for accurate seabed representation in an orthoimage. Consequently, the refraction correction of the imagery to be used for the orthoimage generation is necessary even if a correct DSM of the seabed is available.



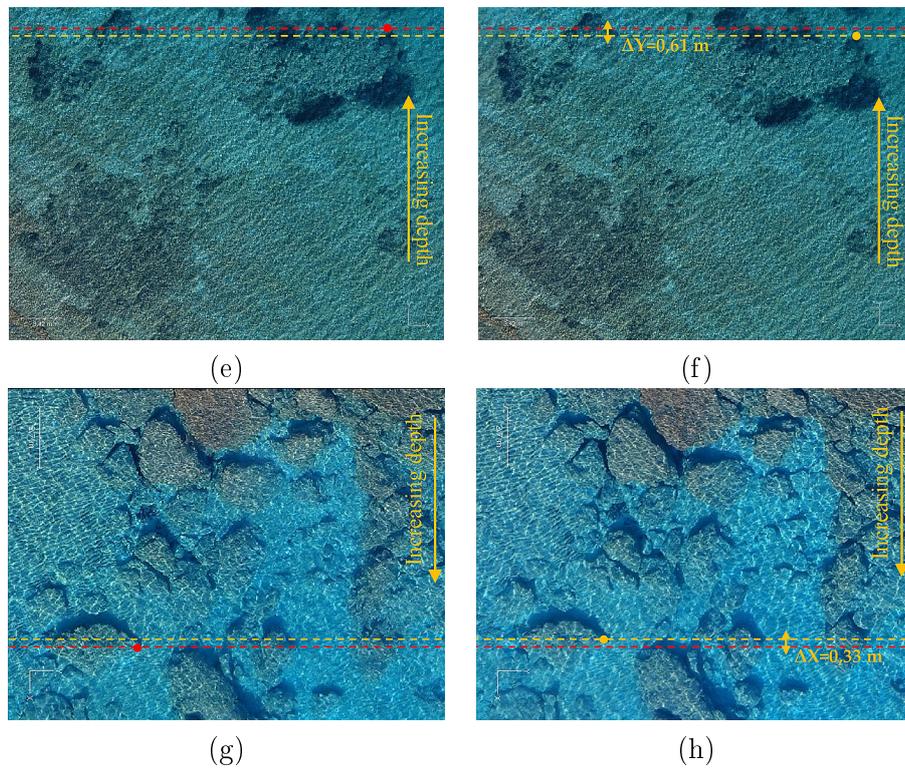


Figure 5.44: Parts of the orthoimages of Amathounta (first row), and Agia Napa (second row), Cyclades-1 (third row) and Cyclades-2 (fourth row) test sites. In the left column the orthoimages generated using the uncorrected imagery are presented while in the right column the results using the refraction corrected imagery are presented. For the generation of the presented refraction-free datasets the SVR model trained on the Dekelia dataset was exploited.

5.4.2.6 Error propagation within the sequential steps of the proposed Method 2

In this section, certain factors that affect the overall accuracy of the produced refraction-free images are discussed. In order to estimate the effect and the amount of error that propagates, the corrected dense point clouds were compared to their respective true depth point clouds, described in detail in Subsection 5.3.6. For the Amathounta test site, the mean M3C2 distance [Lague 2013] between the compared dense point clouds reached 0.03m with a standard deviation of 0.35 m; for the Agia Napa test site, the mean M3C2 distance was calculated to 0.085m with a standard deviation of 0.49 m; for the Cyclades-1 test site, the mean M3C2 distance was calculated to -0.01 m with a standard deviation of 0.09 m; while for the Cyclades-2 test site, the mean M3C2 distance was calculated to -0.01 m with a standard deviation of 0.31m. These values might not be exactly the same with the ones presented in Table 5.9 since for the tests performed here, the total number of the generated points was used, and not the specific training/testing areas.

To evaluate the effect of these errors in the image correction procedure, through their use in the updated DSM creation, two examples are provided with a seabed point A lying at 5m depth at the Amathounta and Agia Napa datasets. Considering that the average flying height for Amathounta dataset was 103m and for Agia Napa 209 m, the following can be calculated:

For the Amathounta test site, point A is having a real depth 5m and a calculated depth of 5.03m (real depth plus the mean distance computed by M3C2). Then, the c_{mixed} for the real depth and the c_{mixed}' for the calculated depth can be computed using Equation 4.8;

$$c_{\text{mixed}} = (P_{\text{air}} n_{\text{air}} + P_{\text{water}} n_{\text{water}}) c_{\text{air}} = (0.954 \cdot 1 + 0.046 \cdot 1.34) \cdot 2827.05 = 2871.550 \text{ pixels}$$

$$c_{\text{mixed}}' = (P_{\text{air}}' n_{\text{air}} + P_{\text{water}}' n_{\text{water}}) c_{\text{air}} = (0.953 \cdot 1 + 0.047 \cdot 1.34) \cdot 2827.05 = 2871.804 \text{ pixels}$$

The difference between c_{mixed} and c_{mixed}' is calculated $\Delta c_{\text{mixed}} = c_{\text{mixed}} - c_{\text{mixed}}' = 0.255 \text{ pixels}$

To compute the effect of this difference in the image coordinates of points a and a' that are the projections of the non-refracted and the refracted ray coming from seabed point A , respectively, two different radial distances were considered for testing: 100 pixels and 1500 pixels.

For a point a having radial distance r 100 pixels;
 $r_{c_{\text{mixed}}} = 101.574 \text{ pixels}$ and $r_{c_{\text{mixed}}'} = 101.583 \text{ pixels}$, $\Delta r_{c_{\text{mixed}}} = 0.009 \text{ pixels}$

For a point a having radial distance r 1500 pixels;
 $r_{c_{\text{mixed}}} = 1523.61 \text{ pixels}$ and $r_{c_{\text{mixed}}'} = 1523.75 \text{ pixels}$, $\Delta r_{c_{\text{mixed}}} = 0.160 \text{ pixels}$

For the same point in Agia Napa test site having a real depth 5m and a calculated depth of 5.085m (real depth plus the mean distance computed by M3C2), the c_{mixed} for the real depth and the c_{mixed}' for the calculated depth were calculated using Equation 4.8;

$$c_{\text{mixed}} = (P_{\text{air}} n_{\text{air}} + P_{\text{water}} n_{\text{water}}) c_{\text{air}} = (0.977 + 0.0236 \cdot 1.34) \cdot 2852.34 = 2874.999 \text{ pixels}$$

$$c_{\text{mixed}}' = (P_{\text{air}}' n_{\text{air}} + P_{\text{water}}' n_{\text{water}}) c_{\text{air}} = (0.976 \cdot 1 + 0.024 \cdot 1.34) \cdot 2852.34 = 2875.375 \text{ pixels}$$

The difference between c_{mixed} and c_{mixed}' is calculated $\Delta c_{\text{mixed}} = c_{\text{mixed}} - c_{\text{mixed}}' = 0.376 \text{ pixels}$

To compute the effect of this difference in the image coordinates of points a

and a' that are the projections of the non-refracted and the refracted ray coming from seabed point A, respectively, two different radial distances were considered for testing: 100 pixels and 1500 pixels.

For a point a having radial distance r 100 pixels;

$$r_{c_{\text{mixed}}} = 100.794 \text{ pixels and } r_{c_{\text{mixed}}'} = 100.808 \text{ pixels, } \Delta r_{c_{\text{mixed}}} = 0.014 \text{ pixels}$$

For a point a having radial distance r 1500 pixels;

$$r_{c_{\text{mixed}}} = 1511.92 \text{ pixels and } r_{c_{\text{mixed}}'} = 1512.11 \text{ pixels, } \Delta r_{c_{\text{mixed}}} = 0.19 \text{ pixels}$$

In the tests performed above, the differences observed between $c_{\text{mixed}} - c_{\text{mixed}}'$ and $r_{c_{\text{mixed}}} - r_{c_{\text{mixed}}}'$ when adding the mean distance as a constant value are considered negligible. In addition to the aforementioned analytical estimation, Figure 5.45 presents the differences Δc_{mixed} of the c_{mixed} calculated using the real depth and the c_{mixed}' calculated using the calculated depth of a point. Solid lines represent the Δc_{mixed} calculated for Amathounta (in green), Agia Napa (in blue), Cyclades-1 (in cyan for 35m flight height and in orange for 70 m), and Cyclades-2 (in light green for 35m flight height and in red for 70 m) by adding the mean distance calculated for each site as a constant value to the hypothetical real depth. Dashed lines represent the Δc_{mixed} calculated by adding the mean distance and the standard deviation, while dotted lines represent the Δc_{mixed} calculated by adding the mean distance minus the standard deviation. That way, the maximum Δc_{mixed} can be calculated for each depth.

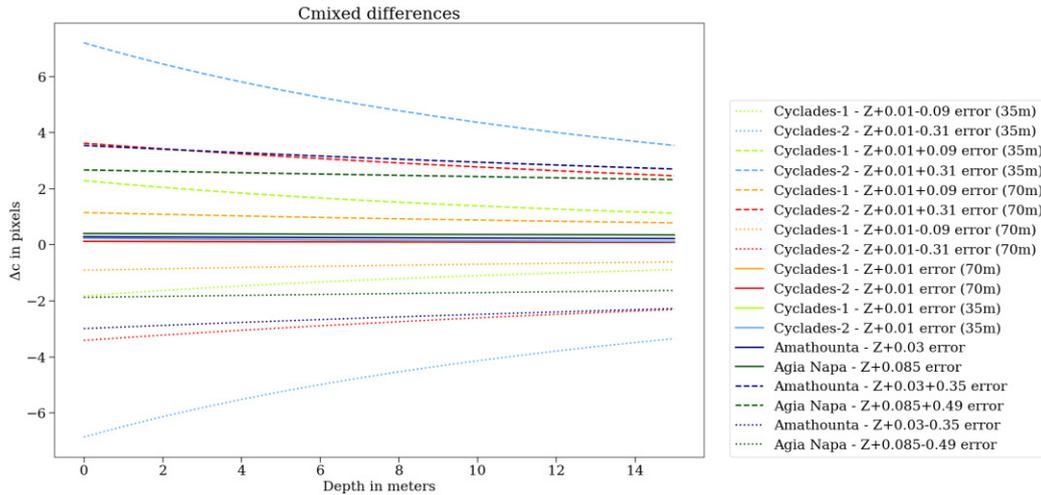
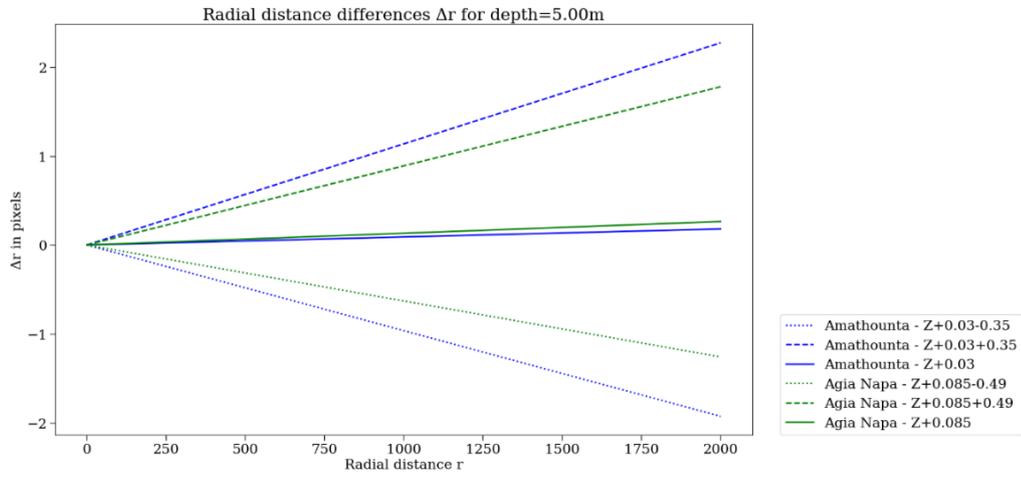


Figure 5.45: The computed radial distance differences and the c_{mixed} differences for all the test sites.

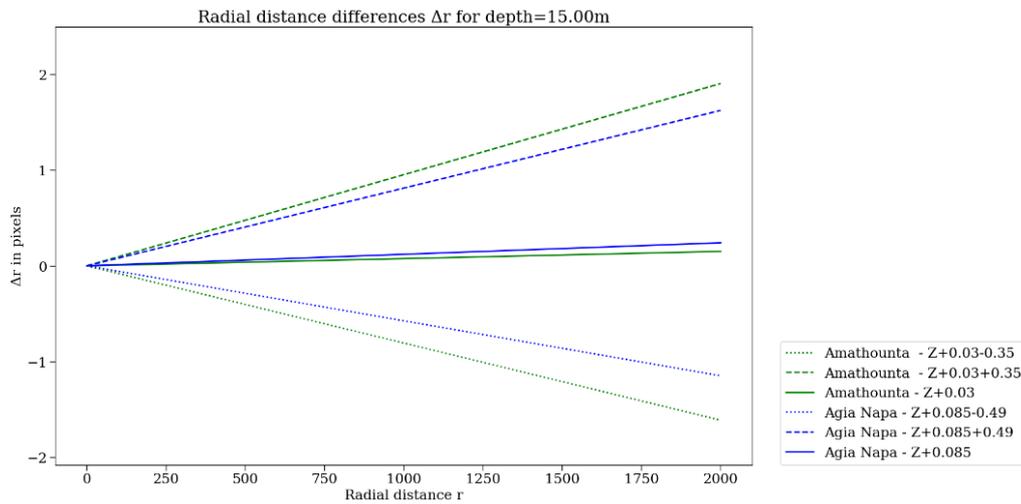
Results suggest that the larger the depth, the smaller the differences between c_{mixed} and c_{mixed}' . Moreover, when the standard deviation was added, the Δc_{mixed} calculated for the Cyclades-2 test site with a flight height of 35 m, the c_{mixed} was

affected the most by the errors present in the DSM, even though they were smaller than the errors at the Agia Napa test site. This result can be explained by the fact that the total added error in Cyclades-2 (35 m) had a greater effect on the percentages of air and water in the total light ray distance, being shallower and having a very small flight height. Excluding this test case, which was the worst, Δc_{mixed} was not more than 3.5 pixels in Amathounta and Cyclades-2 (70 m) test site and no more than 2.7 pixels in Agia Napa test site, while the Δc_{mixed} for the Cyclades-1 test site remained below 2 pixels for all the flight heights. In the typical case, Δc_{mixed} was not more than 0.4 pixels for all the tested cases.

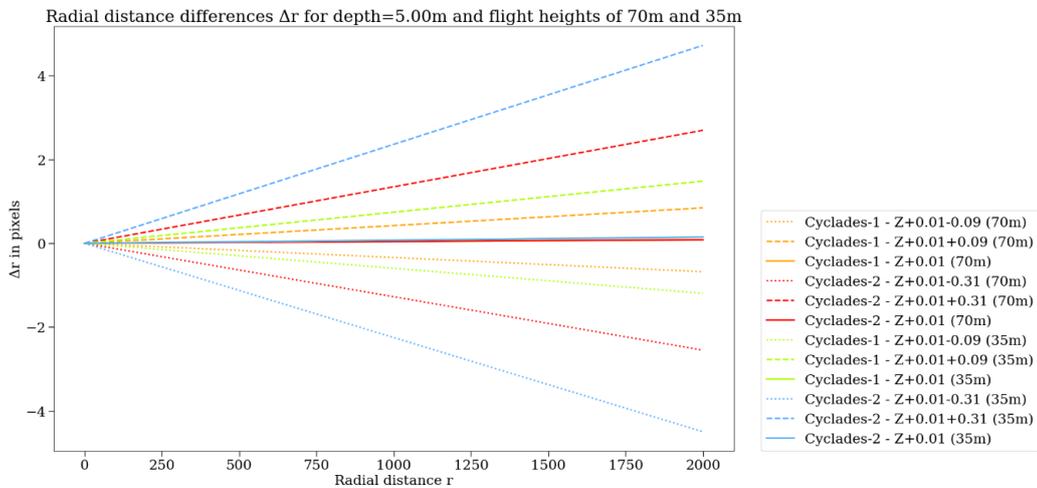
Figure 5.46 presents the radial distance differences $\Delta r_{c_{\text{mixed}}}$ in relation to the radial distance r for two different depths, 50 and 15 m, for all the test sites. Solid lines represent the Δc_{mixed} calculated for Amathounta (in green), Agia Napa (in blue), Cyclades-1 (in cyan for 35m flight height and in orange for 70 m), and Cyclades-2 (in light green for 35m flight height and in red for 70m) by adding the mean distance calculated for each site as a constant value to the hypothetical real depth. Dashed lines represent the $\Delta r_{c_{\text{mixed}}}$ calculated by adding the mean distance and the standard deviation, while dotted lines represent the $\Delta r_{c_{\text{mixed}}}$ calculated by adding the mean distance minus the standard deviation. That way, the maximum $\Delta r_{c_{\text{mixed}}}$ can be calculated for each of the two tested depths. The larger the value (error) added in the real depths, the larger the radial distance differences $\Delta r_{c_{\text{mixed}}}$. On the contrary, by observing the results among Figure 5.46a–d, the deeper the point is, the smaller the radial distance differences $\Delta r_{c_{\text{mixed}}}$ are. For example, in the Amathounta test site, presenting the larger differences between Figure 5.46a,b, the maximum $\Delta r_{c_{\text{mixed}}}$ calculated for the depth of 5m was 2.2 pixels on the edge of the image and 1.9 pixels for the depth of 15m. For the same depths, the maximum $\Delta r_{c_{\text{mixed}}}$ in the Agia Napa test site was calculated as 1.8 and 1.6 pixels, respectively. Coming to the typical case, $\Delta r_{c_{\text{mixed}}}$ reached 0.2 and 0.12 pixels in the Amathounta test site, while for the same depths in the Agia Napa test site, $\Delta r_{c_{\text{mixed}}}$ reached 0.25 and 0.23 pixels, respectively, for the radial distance of 2000 pixels, being the edge of the image. Coming now to Figure 5.46c,d that present the $\Delta r_{c_{\text{mixed}}}$ over the Cyclades-1 and Cyclades-2 datasets, it is obvious that the smaller the flight height, the larger the $\Delta r_{c_{\text{mixed}}}$ calculated, as observed also in Figure 5.45. For the Cyclades-1 dataset, the maximum $\Delta r_{c_{\text{mixed}}}$ calculated for the depth of 5m was 1.1 pixels on the edge of the image and 0.8 pixels for the depth of 15m and flight height of 35 m; hence for flight height of 70 m, the calculated $\Delta r_{c_{\text{mixed}}}$ was even smaller. For the same depths, the maximum $\Delta r_{c_{\text{mixed}}}$ in the Cyclades-2 test site was calculated as 4.5 and 2.8 pixels, respectively. Coming to the typical case, $\Delta r_{c_{\text{mixed}}}$ reached 0.18 and 0.11 pixels in the Cyclades-1 test site, while for the same depths in the Cyclades-2 test site, $\Delta r_{c_{\text{mixed}}}$ reached 0.22 and 0.18 pixels, respectively, for the radial distance of 2000 pixels, being the edge of the image. Again, results suggested that the larger the depth, the smaller the differences between $r_{c_{\text{mixed}}}$ and $r_{c_{\text{mixed}}}'$ (, since they clearly depend on the calculated c_{mixed} . Similarly, the Amathounta and Cyclades-2 test sites were more affected when the standard deviation was added as a constant error, presenting $\Delta r_{c_{\text{mixed}}}$.



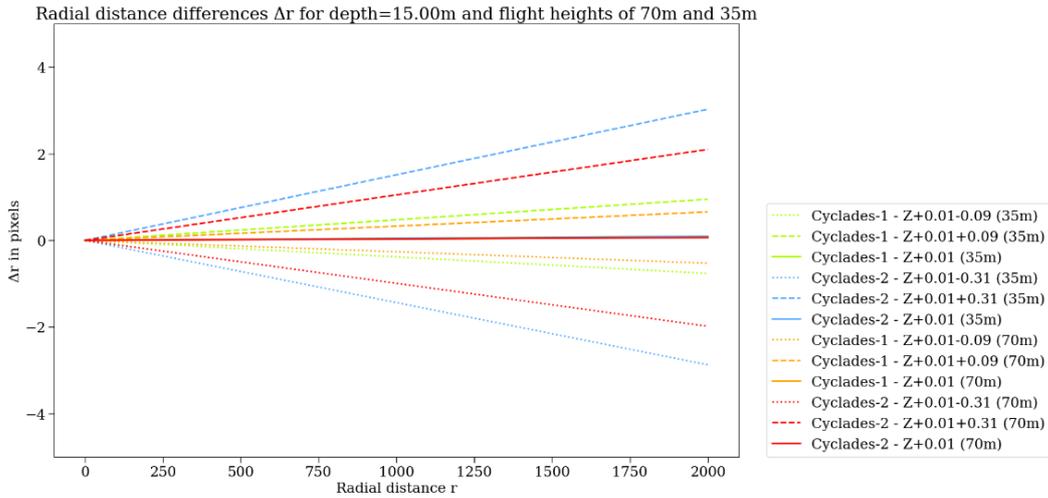
(a)



(b)



(c)



(d)

Figure 5.46: The radial distance differences for Amathounta and Agia Napa test sites are presented in (a) and (b) for depths of 5m and 15m respectively and the radial distance differences for Cyclades-1 and Cyclades-2 test sites are presented in (c) and (d) for depths of 5m and 15m respectively.

Considering the results in Figure 5.45 and Figure 5.46, in the majority of the cases, the magnitude of the errors introduced in the DSM by the refraction correction method applied on the dense point clouds (Method 1) is not capable of affecting the overall accuracy of the proposed method for correcting image refraction (Method 2), especially in the typical case. Subpixel errors were transferred in cases, where the mean errors applied were not affecting the image correction quality and accuracy. However, areas having larger mean M3C2 distances might exist due to the presence of seagrass and sandy bottom areas, leading to larger Δc_{mixed} and $\Delta r_{\text{c}_{\text{mixed}}}$, and thus affecting the image correction method (Method 2). This is the reason why the mean distances calculated for this approach (Table 5.10) are always bigger than the results of the method presented in Chapter 3, used to correct the dense point clouds. Taking this into account and in order to express the results of Figure 5.45 and Figure 5.46 about the transferable errors in the whole process into measurable distances, the differences between the mean distances of the sparse point clouds corrected by Method 1 and Method 2 were calculated; For Amathounta test site, it was 0.10 m; for Agia Napa test site, it was 0.01 m; for Cyclades-1 test site, it was 0.04 m, and for Cyclades-2 test site, it was 0.02m. Considering sea level alterations, ripples, and the anaglyph of the seabed, it is concluded that the magnitude of the errors propagated within Method 2 was not capable of affecting the overall accuracy of the proposed method, being in most of the cases less than the GSD.

5.4.2.7 Summary

The developed Method 2 was tested and validated over four different test sites in Greece and Cyprus having different flight plans, different UAV systems, different sensors, but similar optimal weather conditions and water visibility. The qualitative evaluation of the proposed method was performed through the visual assessment of the textured models created by the updated DSM and the original uncorrected imagery and the textured models created by the updated DSM and refraction-free imagery. Results, exploiting both an SVR model trained on a real world dataset and an SVR model trained on a synthetic dataset, suggested a great improvement in the detail and the quality of the textured 3D models when the results of the proposed Method 2 are used.

The quantitative evaluation of the proposed methods was performed through the comparison of the sparse point clouds produced after the corrected imagery with the respective true depth point clouds and through a metric comparison between orthoimages created by the updated DSM and the original uncorrected imagery and orthoimages created by the updated DSM and refraction-free imagery. By evaluating the results of the statistical analysis of the tests performed over the Agia Napa test site till the depth of 10m and the results over the full Agia Napa test site that reaches the depth of 14.8m, small insignificant differences were observed. These differences revealed a particularly important advantage of this second method for the correction of the geometric effects of refraction. This advantage is the facilitation of a more accurate and reliable SfM, supported by more correct matches in the depths of more than 10m, since the intense refraction in those depths which acts as a limiting factor (see 5.4.1.2 and Section 5.5) for point matching is eliminated.

Results from all the tests performed suggested that the proposed refraction correction method produces refraction-free images in a highly accurate and reliable way. Also, when these refraction-free images are used, more accurate and reliable orthoimages of the seabed can be generated. For all the test sites, this second proposed method met and exceeded the accuracy standards generally accepted for hydrography, as introduced by the IHO, where in its simplest form, the vertical accuracy requirement for shallow water hydrography can be set as a total of $\pm 0.25\text{m}$ from all sources, including tides [Guenther 2000]. Also, both the proposed methods are qualified to the maximum Zone of Confidence (ZOC) level A1, considering IHO's white paper on the regulations for international charts and charts specifications [IHO 2019].

5.5 Comparative analysis of the proposed methods

As already stated in 5.4.2.3, a well-established methodology for evaluating the SfM performance is to compare the sparse point cloud to some truth with the same data representation [Brown 1972]. In order to stress out the accuracy of the two presented methodologies compared with some state of the art refraction correction approaches, the true depth point clouds of the four test areas were compared with

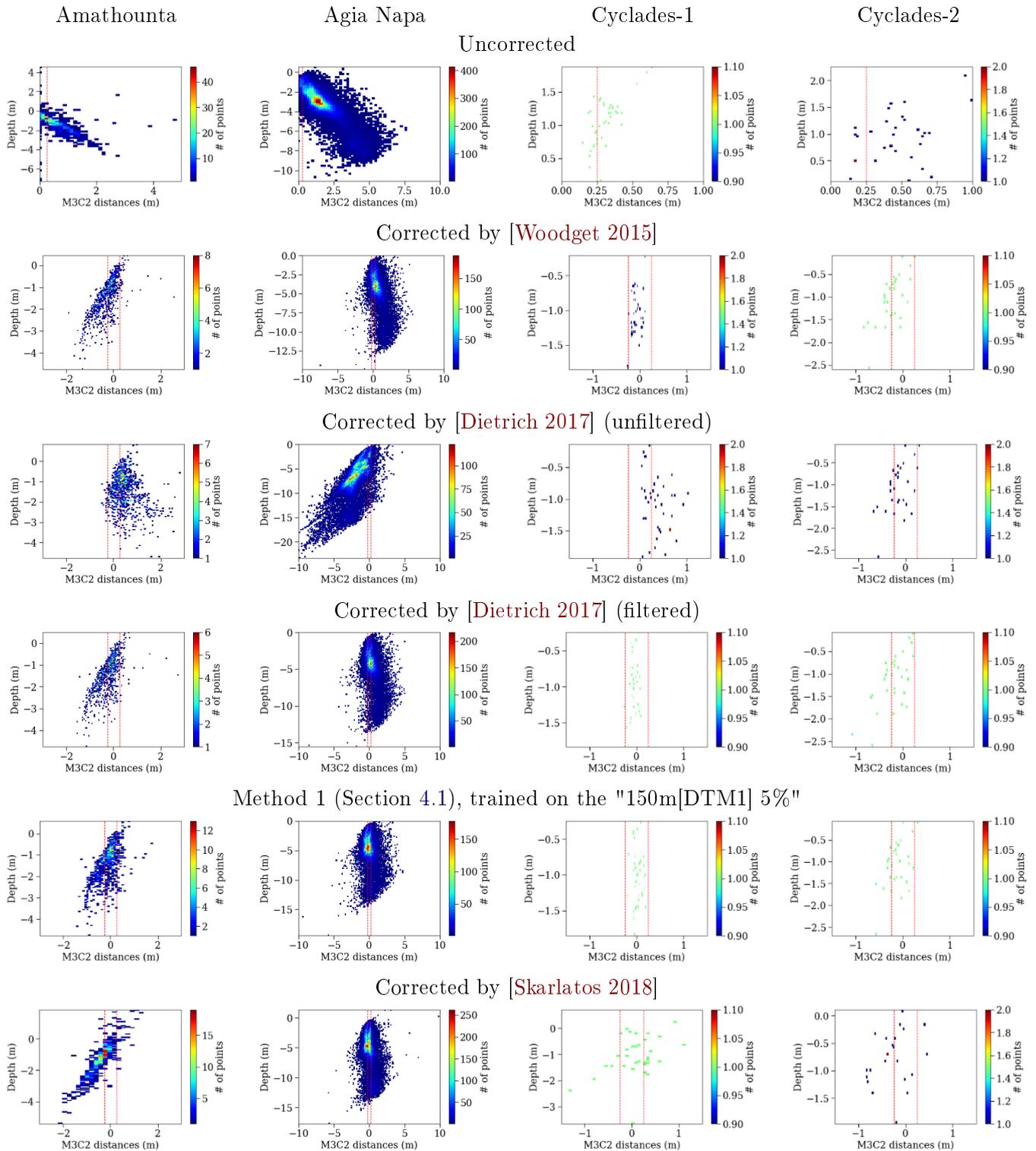
the sparse point clouds resulted from: the machine learning based method presented in Section 4.1 (Method 1), the image-based refraction correction method presented in Section 4.2 (Method 2), the image-based refraction correction method described in [Skarlatos 2018] (see Chapter 3) and the 3D space correction approaches presented in [Dietrich 2017] and in [Woodget 2015]. Since the two methodologies proposed in this thesis are already described in Chapter 3 and Section 4.2 and methods presented in [Skarlatos 2018, Dietrich 2017, Woodget 2015] have already and adequately been described in Chapter 3, in the following paragraph only some additional implementation details are given.

For applying the methods described in [Woodget 2015] and [Dietrich 2017], the Python script delivered by the authors was used [Dietrich 2019]. Regarding the parameters used for filtering the depth values produced by [Dietrich 2017], a maximum angle of 35° (as suggested by the author) and a maximum distance of 110m were selected for Amathounta test site, for Agia Napa test site, a maximum angle of 35° and a maximum distance of 220m were selected, for Cyclades-1 test site, a maximum angle of 35° and a maximum distance of 95m were selected, while for Cyclades-2 test site, a maximum angle of 35° and a maximum distance of 85m were selected. Regarding the presented results of the method presented in [Skarlatos 2018], they are generated after four iterations in order to achieve a more accurate depth correction. The results of the M3C2 distances among the true depth point clouds of the four test areas and the respective sparse point clouds resulted from the tested refraction correction approaches are presented in Figure 5.47 and Table 5.11.

In Figure 5.47 (first row), it can be observed that the distances between the reference data and the original (uncorrected) image-based point clouds increase in proportion to the depth. In all the test cases demonstrated in the first row of Figure 5.47, the Gaussian mean of the differences (\bar{x}) was significant. It reached 0.67m in the Amathounta test site, 1.71m in the Agia Napa test site, 0.32m in Cyclades-1 test site, and 0.54m in Cyclades-2 test site (Table

The second, third and fourth rows of Figure 61 illustrate the M3C2 distances in relation to the depth between the true and the sparse point clouds produced by [Woodget 2015] and [Dietrich 2017] (filtered in fourth and unfiltered in third). The fifth row illustrates the M3C2 distances between the true and the sparse point clouds produced by the Method 1 trained on the synthetic dataset. All the above compared approaches are correcting the geometric effects of refraction in the 3D space and specifically in the sparse or dense point cloud.

The sixth row of Figure 5.47 illustrates the M3C2 distances in relation to the depth between the true and the sparse point clouds produced by and [Skarlatos 2018] while the seventh and the eight row present the histograms of the M3C2 distances between the true and the sparse point clouds produced from corrected imagery of the Method 2, exploiting the SVR models trained on the Dekelia and the synthetic datasets respectively. The above three compared approaches are correcting the geometric effects of refraction in the image space.



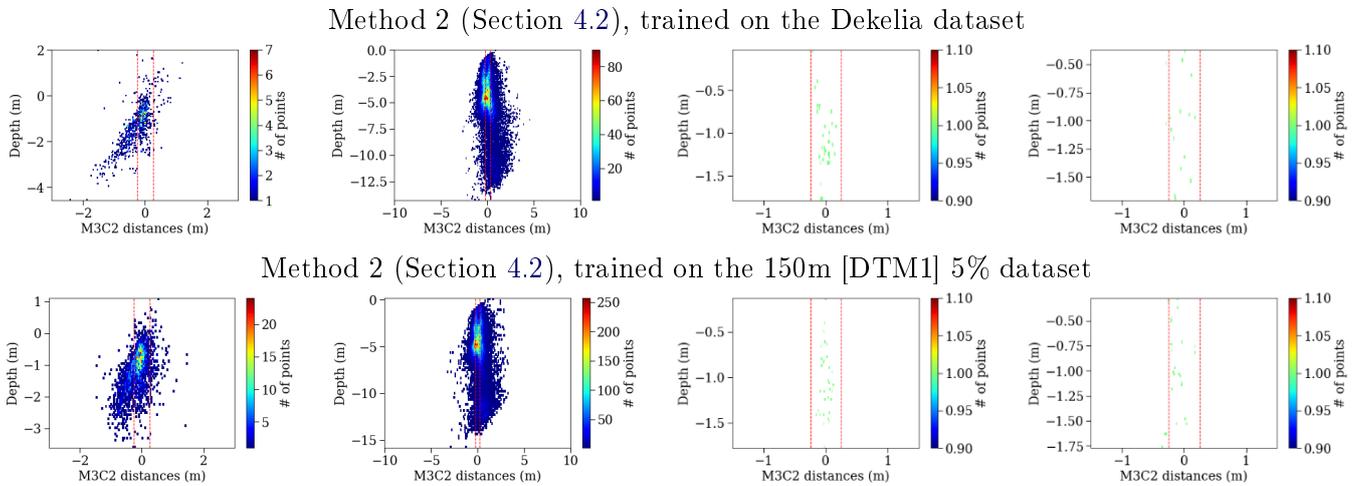


Figure 5.47: The histograms of the M3C2 distances between the true and the uncorrected and corrected image-based sparse point clouds derived from the SfM for Amathounta (first column), Agia Napa (second column), Cyclades-1 (third column) and Cyclades-2 (fourth column) test sites respectively. The dashed lines in red represent the accuracy limits generally accepted for hydrography as introduced by the IHO [Guenther 2000]. M3C2 distances are in (m).

By observing Figure 5.47, the improved performance of both of the proposed methods for correcting the geometric effects of refraction, compared with the rest of the approaches, is emerging. Also, a very important outcome of this figure is that when the initial imagery is corrected, the resulted sparse point cloud is characterized by less noise in the elevations. This is evident by comparing the 2D histograms of the image correction method presented in this thesis with the 2D histograms resulted for the methods that are applied in the 3D point clouds directly. There, it is obvious that the 2D histograms resulted from the image correction method are narrower in the X axis, indicating less scattered apparent elevations. This is a very important advantage of the proposed method, indicating that when correcting the initial imagery from refraction, SfM delivers more accurate and more reliable results with less noise in the sparse and consequently in the dense point clouds. This is explained by the fact that the geometric effects of refraction in the images are affecting feature matching in an important degree; matchable key points are not matched together or they are matched with the wrong key points, because due to the refraction effect, each image is distorted in a different way. This effect drives feature matching algorithms to match points that are fulfilling the epipolar geometry only because of the refraction effect.

Table 5.11 presents the results of all the tests performed. Both in Figure 5.47 and Table 5.11, a great improvement in the sparse point clouds, and thus the SfM accuracy, was achieved by the two methods presented in this thesis compared with the uncorrected ones. More specifically, regarding the Method 1, trained on the synthetic dataset, in Amathounta test site, the initial 0.67m mean distance was

reduced to -0.04m , in Agia Napa test site, the initial 1.71m mean distance was reduced to 0.06m , in Cyclades-1, the 0.32m initial mean distance was reduced to -0.05m , while in Cyclades-2, the 0.54m initial mean distance was reduced to -0.05m too. These results, are quite similar with the results of this method on the image-based dense point clouds presented in Table 9, proving the robustness and the consistency of the method.

Regarding the Method 2, in Amathounta test site, the initial 0.67m mean distance was reduced to -0.04m , in Agia Napa test site, the initial 1.71m mean distance was reduced to -0.04m , in Cyclades-1, the 0.32m initial mean distance was reduced to 0m , while in Cyclades-2, the 0.54m initial mean distance was reduced to -0.07m . In all the cases, results include outlier points, such as seagrass, that were not captured in the ground truth true depth clouds for all the cases or were caused due to point cloud noise again in areas with seagrass or poor texture.

The aforementioned reduction of noise in the point clouds generated using the refraction corrected imagery by the Method 2, is also proven in Table 5.11. There, it is obvious that even although the two proposed methods deliver quite similar results, the standard deviations are reduced in the results of the second method.

Compared with the results of the rest of the tested refraction correction approaches that are targeting the point cloud, for all the tested areas, the Method 1 outperformed all the rest. This result also justifies the selection of this method for the correction of the dense point clouds, before the DSM generation in the Method 2.

The Method 2, compared with on the accuracy levels reached by the Method 1, managed to achieve slightly better accuracy in all the test sites, meeting too the standards of IHO [Guenther 2000]. However, this second method, delivered better standard deviations and RSMEs, especially in the deeper site of Agia Napa compared with Method 1. This way it is proved that the Method 2 achieves the improvement not only of the accuracy but also of the quality of the generated point clouds by reducing the noise in them generated point clouds.

Regarding the rest of the compared refraction correction approaches, both [Dietrich 2017] and [Woodget 2015] are developed for correcting the effects of refraction in very shallow waters riverbeds and they seem to produce quite accurate and reliable results, especially when filtering is applied in [Dietrich 2017].

The improved accuracy of the Method 2 can be explained by the fact that approaches such as [Dietrich 2017] and [Woodget 2015] certainly affect SfM and MVS processes and degrades the results, as they are using the camera positions and the point clouds that which are produced using the refracted imagery, certainly affects SfM and MVS processes and degrades the results. On the contrary, the Method 2, by correcting the initial imagery from refraction, overcomes the transferred errors resulting from the initial orientations, since SfM results are being updated, delivering a more accurate geometry of the block and consequently more accurate SfM results. Compared with the similar method presented in [Skarlatos 2018], it seems that the iterative way is not effective enough, especially when the method is implemented in deeper sites.

Correction is applied on:		Test Site											
		Amathounta			Agia Napa			Cyclades-1			Cyclades-2		
		# of check points			75K			23			34		
Derived point clouds from different methods		Statistical Analysis [m]											
		\bar{x}	s	RMSE	\bar{x}	s	RMSE	\bar{x}	s	RMSE	\bar{x}	s	RMSE
		[m]	[m]	[m]	[m]	[m]	[m]	[m]	[m]	[m]	[m]	[m]	[m]
Uncorrected images		0.67	2.19	2.28	1.71	1.18	2.08	0.32	0.10	0.33	0.54	0.29	0.62
Point clouds	[Woodget 2015]	-0.27	0.40	0.49	0.63	1.02	0.98	-0.08	0.10	0.12	-0.23	0.26	0.34
	[Dietrich 2017]	0.49	0.54	0.73	-1.55	1.49	1.75	0.38	0.25	0.46	-0.15	0.24	0.28
	[Dietrich 2017] (filt.)	-0.22	0.40	0.45	0.43	0.72	0.84	-0.06	0.09	0.10	-0.20	-0.30	0.36
Method 1 (150m [DTM1] 5%)		-0.04	0.13	0.14	0.06	0.41	0.42	-0.05	0.06	0.07	-0.05	0.12	0.13
Images	[Skarlatos 2018]	-0.39	0.88	0.96	-0.05	0.74	0.74	0.15	0.42	0.46	-0.28	0.36	0.46
	Method 2 (Dekelia)	-0.19	0.28	0.31	-0.04	0.37	0.38	-0.02	0.09	0.09	-0.06	0.14	0.15
	Method 2 (150m [DTM1] 5%)	-0.04	0.12	0.13	-0.04	0.24	0.24	0.00	0.06	0.07	-0.07	0.06	0.09
IHO limit (1σ)		0.25			0.25			0.25			0.25		

Table 5.11: Comparative quantitative Evaluation. \bar{x} is the average distance of the point cloud from the true values and s is its standard deviation. Negative values suggest overestimation of the depth and positive suggest underestimation. The best values achieved for each test site are highlighted in bold format.

Regarding the Method 1, even though it is also correcting point clouds generated by the refracted imagery, it is not exploiting the interior and exterior orientations of the cameras, as done in [Dietrich 2017] and [Woodget 2015], thus being less affected by those factors. Moreover, the SVR model developed here was trained on point clouds generated using refracted imagery and, consequently, had "learned" to avoid errors introduced by the erroneous cameras' positions.

To conclude, the results of the two proposed methods outperformed the results of the other state of the arte methods for all the test sites, while Method 2 outperformed Method 1. The resulting accuracies in the sparse point cloud met and exceeded the accuracy standards generally accepted for hydrography, as introduced by the IHO, where in its simplest form, the vertical accuracy requirement for shallow water hydrography can be set as a total of $\pm 0.25\text{m}$ (one sigma) from all sources, including tides [Guenther 2000]. Both the proposed methods are also qualified to the maximum Zone of Confidence (ZOC) level A1, considering IHO's white paper on the regulations it's for international charts and charts specifications [IHO 2019].

5.6 Discussion

In this chapter, two different methods for correcting the geometric effects of refraction have been presented:

Method 1. Refraction correction on the point clouds exploiting a linear Support Vector Regression (SVR) model: The first one, delivers image-based 3D point clouds with accurate depth information by training models which are capable of estimating the correct depth of a set of points having knowledge of only their apparent (erroneous) depth. In this method, the training of an SVR model based on known depth observations from bathymetric LiDAR surveys is proposed, in order to be able to estimate with high accuracy, the real depths of point clouds derived from conventional SfM-MVS procedures using low-altitude aerial imagery. The developed method was tested and validated over a synthetic and five different real-world test sites in Greece and Cyprus having different flight plans, different UAV systems, different sensors, but similar optimal weather conditions and water visibility.

The experiments performed demonstrated that the model has significant potential for generalization over different areas and in shallow waters, and it can be used when no LiDAR data are available. This can be attributed to the large amount of data used for training and testing and the SVR model robustness, compared to the simple Linear Regression which is very sensitive to the point cloud noise. The method proposed is independent from the UAV imaging system, the camera type, the flying height, the camera to water surface distance, and the base-to-height (B/H) ratio. Additional data, e.g., camera orientations and camera intrinsic parameters, for predicting the correct depth of a point cloud, are not required. This is a very important advantage of the proposed method compared to the current state-of-the-art methods which have been designed to address refraction errors in bathymetry

mapping using aerial imagery.

The proposed method can also be used in areas where LiDAR data of low density are available, in order to create a denser bottom representation through data fusion techniques. In particular, the data fusion procedure implemented demonstrated its high potential towards the efficient coloring of LiDAR point clouds, gap and hole filling, and surface seabed model updating, thus minimizing costs and time for many applications. This achievement is of high importance when it comes to multitemporal and multi-source monitoring of seabed areas; It facilitates the update of specific parts of the seabed, as well as filling holes in point clouds generated by a source failing to produce reliable results in specific areas; i.e., areas with sun glint in aerial imagery, areas covered by ships or other floating objects in LiDAR, surveys, etc.

As in any supervised learning approach, the quality of the training data affects the accuracy of the predicted models. As such, regarding the training of Method 1 using real world data, slight limitations were imposed due to incompatibilities between the LiDAR point cloud and the image-based one. Among others, the different level of detail imposed additional errors in the point cloud comparison and might have compromised the absolute accuracy of the method. However, these issues were solved by the generation of the synthetic data. Additionally, the unfavorable position of the GCPs in the photogrammetric block situated on the shore may also have compromised the accuracy of the block and the point cloud, respectively, in the deeper areas of the bottom that are far from those. However, those issues do not affect the principle of the main goal of the presented method, which is to systematically solve the depth underestimation problem, by predicting the correct depth. To improve the accuracy of the predicted models, and thus the depth accuracy of the corrected point clouds, concurrently acquired images and true data (i.e., bathymetric LiDAR data), having the same level of detail, should be available in order to avoid differences in the point clouds.

Method 2: Correcting the Geometric Effects of Refraction on the Image Space: The second method, firstly exploits the first one, and then corrects the refraction effect in the original imaging dataset, exploiting image transformation and resampling techniques. That way, the operational SfM and MVS processing pipelines are executed on a refraction-free set of aerial datasets, resulting in highly accurate bathymetric maps. At the same time, this method allows for robust, cost-effective mapping of the coastal zone, both land and bottom, for image and dense point cloud deliverables. The developed method was tested and validated over different test sites in Greece and Cyprus having different flight plans, different UAV systems, different sensors, but the similar optimal weather conditions and water visibility. The performed qualitative and quantitative evaluations of this method suggested a great improvement in the detail and the quality of the textured 3D models when the results of the proposed method are used. Compared with the point clouds of the true values, the sparse point clouds which were produced based on the refraction-free dataset, managed to achieve accuracy within the IHO standards [Guenther 2000]. The method presented here, by correcting the initial imagery from

refraction, overcomes the transferred errors resulting from the initial orientations, since SfM results are being updated, delivering a more accurate geometry of the block and more accurate SfM results, eliminating the noise on the generated point clouds.

By correcting the imagery for refraction effects, a variety of applications becomes possible, offering more information in addition to the standard bottom surface/anaglyph. With the proposed method, the generation of accurate, reliable, and detailed orthoimages and textures for the bottom 3D models is achieved as an additional and very valuable deliverable for many applications. Orthoimages can be produced based on the updated DSM, the updated camera positions, and the refraction-free images. Having corrected the imagery, which is the primary source of information for all the image-based mapping applications, semantic information can be then used as a primary input to the SfM-MVS process in order to produce semantically labeled 3D point clouds. It also facilitates a wide range of exploitations of the available visual information, such as bottom classification, coral reefs monitoring, general benthic community mapping, marine litter mapping, and more.

As with any method for bathymetry mapping using low-altitude aerial RGB imagery, the implementation of both the proposed methods is restricted by various environmental factors, including the water surface state conditions, the visibility of the bottom, and the presence of texture on the bottom. To achieve accurate bathymetric results, the water surface must be as flat as possible at the time of image acquisition in order to have optimal bottom visibility and satisfy the geometric assumption of a flat-water surface. Obviously, water turbidity and water visibility are additional constraints. As in any photogrammetric method, the bottom must include texture, random patterns, and adequate features to ensure key point detection and matching. However, in poorly textured seabed such as a sandy seabed, results would be acceptable even with a less dense point cloud, because normally a sandy bottom does not present any abrupt height differences or detailed forms. Also, the small and sparse seagrass areas offer valuable blob matching possibilities. Sun glint, which depends on the sun angle, may cause direct reflections on the water surface and create bright spots on images. Therefore, the sun elevation should be low during image acquisition (less than 30 degrees), with large angles of incidence with respect to the water surface and the sensor. Finally, caustics [37] may also appear in the aerial imagery, especially when the bottom of the shallower areas (less than 3m depth) is imaged having a sun angle more than 30 degrees over the horizon, and small waves also exist. This phenomenon is eliminated when the GSD is equal to or greater than the width of the caustics on the bottom. Even with the above restrictions, both proposed methods for correcting the geometric effects of refraction, compared with other state-of-the-art refraction correction methods, for all the tested areas, are outperforming them, delivering accurate and reliable bathymetric maps.

Applicability in Greece and the Mediterranean Sea: Based on the tests performed in the real world datasets, for the successful implementation of the two proposed methods for shallow water seabed mapping using low altitude aerial im-

agery, except of the calm water surface, clear waters and textured seabed is a prerequisite.

According to the literature, the Mediterranean Sea is oligotrophic; it is rich in oxygen and poor in nutrients [Zenetos 2002], while this oligotrophy increases from west to east. This trophic status is translated into increased water clarity and light penetrance due to the low chlorophyll-a concentration. This is also confirmed by the large Secchi depths measured, especially in the Eastern Mediterranean (i.e. [Collins 1979]). The water column is also stratified; the top rarely mixes with the bottom due to differences in temperature, forming thermoclines. Water is only truly mixed with extreme weather conditions or in areas of upwelling or downwelling caused by a combination of geographical features and sea and atmospheric circulation. The result of all these factors is the increased clarity of the water column, allowing the clear bottom visibility for implementation of the proposed methods.

Additionally, as reported in [Furlani 2014], rocky coasts represent most of the coastline bordering the Mediterranean Sea, since they are intimately related to the geological history of the landscape that overlooks the basin. More specifically, in the northeastern sector of the Sea, on which the three-quarters of the total length of the Mediterranean coastline is confined [Stewart 2009], about the 65% of it is rocky. This coastline represents the northern part of the Eastern Mediterranean and includes the Ionian, the Adriatic and the western coasts of the Aegean Sea. Greek coasts are characterized by steep cliffs alternated with low-lying deltaic areas which can produce coastal features, such as tombolos [Furlani 2014]. Specifically, Greece has predominantly a rocky coast (70%) and only a minor part of the country's coastline is characterized by sandy beaches and dunes as well as wetlands and lagoons [European Commission 2009]. However, as reported in the tests performed, even in these minor cases where only sandy bottom exist in the scene, results would be acceptable even with a less dense point cloud due to reduced key-point detection and matching difficulties because normally a sandy bottom does not present any abrupt height differences or detailed forms.

The above conditions, together with the mild Mediterranean weather which gives ground for large time windows during the year, allow for the ideal implementation of the proposed methods, reducing the cost of bathymetric surveys and delivering additional optical products, facilitating further studies and processes.

Correcting the Radiometric Effects of Refraction

Contents

6.1 Proposed method for pixelwise image radiometric correction	149
6.1.1 Image classification	151
6.1.2 Stereo-rectification and Disparity maps generation	152
6.1.3 RGB values replacement and back-projection of the images	155
6.2 Summary	155

This chapter aims to present the research performed regarding the pixelwise correction of the radiometric effects of refraction on the shallow waters' underwater imagery and specifically of the rippling caustics. To that direction, a new method is proposed which contrarywise to the state-of-the-art can handle seabed of any anaglyph, correcting the images using real pixel information. This method exploits recent advantages in image classification to detect the areas of the image that are not affected by the phenomenon and based on the recovered geometry of the overlapping images, corrects the rest of the pixels of the images. Section 6.1 presents the proposed method while Section 6.2 summarizes the most important contributions of the approach.

6.1 Proposed method for pixelwise image radiometric correction

Till now, the radiometric correction of the images affected by rippling caustics on the underwater imagery is achieved either by replacing the affected parts of the imagery by using parts of co-registered images using homography [Gracias 2000, Gracias 2008] or by using CNN approaches to predict the missing information on the parts that are affected [Forbes 2018]. As for the first group of approaches, these rely on the feature detection and matching processes in order to co-register the images, which as shown in Subsection 2.3.2 are severely affected by the phenomenon. If no measure is taken, these errors in image registration can be consequently transferred to the whole correction process. Moreover, since they exploit the homography relation, they can be useful only in flat seabed cases. Regarding the prediction of missing RGB information on the affected parts of the

images using CNNs, this is very important for a variety of applications, however, when it comes to reliable and accurate underwater 3D reconstruction and mapping, especially of submerged cultural heritage or benthic community, artificial parts on the imagery should rather be avoided.

In order to deal with the feature detection and matching problem in images containing caustics, the very accurate detection of the not affected areas of the images is proposed in this thesis,, enabling feature detection only on these areas. For the approach presented here, it is of utmost importance to maintain as much of the original RGB information of the images as possible. Next, in order to correct the affected imagery, a pixelwise method based on the stereo or the multi-view geometry is proposed. This method works on the overlapping area of the imagery and takes advantage of the pixel correspondences in this area. The overall workflow of the proposed method is illustrated in Figure 6.1 while the different modules of the method are described in detail in the following sections.

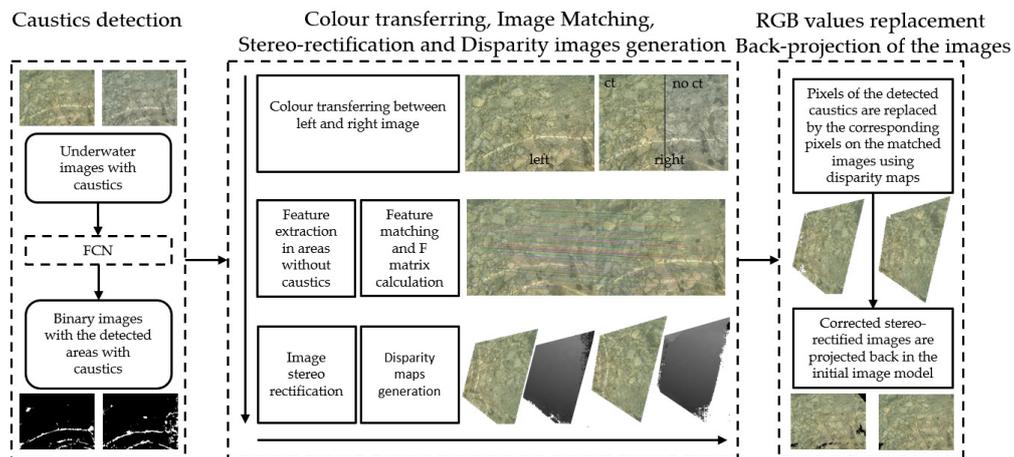


Figure 6.1: The proposed method for the pixelwise correction of the rippling caustics on the underwater imagery.

Initially, in order to detect the unaffected areas with very high accuracy, reliability and repeatability over different caustics patterns, different types of seabed, luminosity and visibility conditions, a Fully Convolutional Neural Network (FCN) classifier based on SegNet architecture [Badrinarayanan 2017] is exploited. The classifier is being trained using the real world dataset on underwater caustics which is presented in Section 7.1 and its selection over shallower learning architectures is justified in Section 7.2. Having detected the unaffected and consequently the affected areas on the initial imagery since it is a binary classification problem, a color transferring approach is performed between the initial imagery. Then the images that are matched using the key points detected on the not-affected areas of the images classified as "non-caustics", are stereo-rectified and their respective disparity maps are generated. In this step, the binary images resulted by the FCN are also re-projected

using the homography matrices computed for the initial imagery rectification. Having all the necessary data available, the pixels classified as "non-caustics" are not processed while the rest of the pixels are replaced by their corresponding pixels on the matched images using the disparity maps. Finally, the corrected stereo-rectified images are projected back onto the initial camera model in order to facilitate further SfM and MVS processing and texturing.

6.1.1 Image classification

Although the unaffected areas are appearing on the images as pixels having lower exposure than the caustics' pixels, for the method presented here, a sophisticated approach was chosen to be implemented. This was decided because of the very intense effects of chromatic aberration appearing on the boundaries of caustics, not being able to be detected by a simpler method such as thresholding. Moreover, it was also important to classify as not-affected areas white colored areas and reflective materials of the seabed such as e.g. marbles and GCP markers placed for photogrammetric campaigns in the shallower waters (see Section 7.2).

Since underwater imagery can be captured by very different distances from the seabed, different imaging sensors, different luminosity and visibility conditions and caustics size might vary depending on the water surface state, it was considered important for the deep fully convolutional network selected to be able to deal with multi-modal and multi-scale image data for semantic labeling. To that direction a similar approach to the one presented in [Audebert 2018] is adopted which is based on SegNet architecture [Badrinarayanan 2017]. For information on the selection of the FCN over shallower learning architectures see Section 7.2.

6.1.1.1 SegNet's architecture

SegNet [Badrinarayanan 2017] is based on an encoder-decoder architecture that produces an output with the same resolution as the input, as illustrated in Figure 63.

This property is highly desirable as, in order to achieve a pixelwise image correction, labeling of the data should be at original image resolution. SegNet allows such task to do as the decoder is able to upsample the feature maps using the unpooling operation [Audebert 2018].

The encoder network consists of 13 convolutional layers which correspond to the first 13 convolutional layers in the VGG16 [Simonyan 2014] designed for object classification. Each encoder layer has a corresponding decoder layer and hence the decoder network has 13 layers. The final decoder output is fed to a multi-class soft-max classifier to produce class probabilities for each pixel independently.

The encoder 5 convolution blocks, each containing 2 or 3 convolutional layers of kernel 3 x 3 with a padding of 1 followed by a rectified linear unit (ReLU) and a batch normalization (BN) [Ioffe 2015]. Each convolution block is followed by a max-pooling layer of size 2 x 2. The decoder performs both the upsampling and the classification. It learns how to restore the full spatial resolution while trans-

forming the encoded feature maps into the final labels. Its structure is symmetrical with respect to the encoder. Pooling layers are replaced by unpooling layers as described in [Zeiler 2014]. The unpooling relocates the activation from the smaller feature maps into a zero-padded upsampled map. The activations are relocated at the indices computed at the pooling stages, i.e. the argmax from the maxpooling. This unpooling allows to replace the highly abstracted features of the decoder to the saliency points of the low-level geometrical feature maps of the encoder. This is especially effective on small objects that would otherwise be misplaced or misclassified. After the unpooling, the convolution blocks densify the sparse feature maps. This process is repeated until the feature maps reach the input resolution.

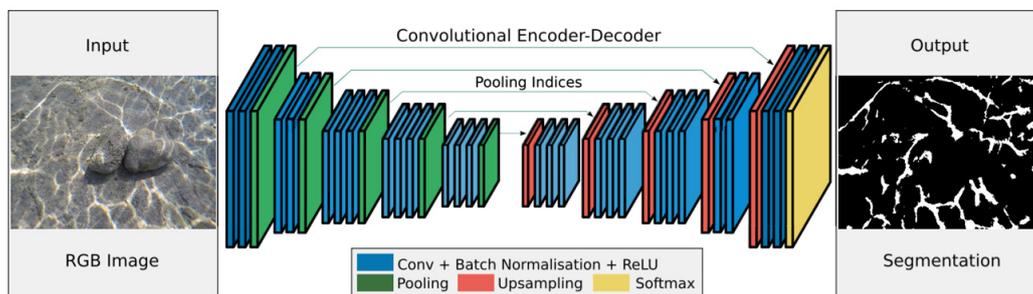


Figure 6.2: An illustration of the SegNet architecture. A decoder upsamples its input using the transferred pool indices from its encoder to produce a sparse feature map. It then performs convolution with a trainable filter bank to densify the feature map. The final decoder output feature maps are fed to a soft-max classifier for pixel-wise classification. (Retrieved from [Badrinarayanan 2017])

6.1.1.2 Multi-scale aspects and Class balancing

To tackle the challenge of the multiscale data on the underwater environment, the multiscale approach proposed in [Audebert 2018] is adopted. There, to generate output predictions at several resolutions, branching the deep network is suggested. In this approach, each output has its own loss which is backpropagated to earlier layers of the network. Therefore, considering the SegNet model exploited here, it is not only predicting one semantic map at full resolution, but also branches the model earlier in the decoder to predict maps of smaller resolutions. Those smaller maps are then interpolated to full resolution and averaged to obtain the final full resolution semantic map [Audebert 2018].

6.1.2 Stereo-rectification and Disparity maps generation

The most important steps of the proposed method are the accurate matching and stereo-rectification of the imagery and consequently the accurate disparity maps generation and filtering. However, prior to these steps, a color transferring approach is performed between the images of the stereopair or even the whole block.

6.1.2.1 Color transferring in the CIE- $l\alpha\beta$ color space

Color transferring is an important step of the proposed method. It will prevent the appearance of visible patches of pixels in the corrected imagery and will deliver seamless corrected imagery. Indeed, due to different exposure and white color balancing of the images, it is very likely that the replaced regions are clearly recognizable by color differences creating edges in the image which are not part of the object itself. The main goal of this step is to obtain a visual consistent and pleasing corrected image, while fine details are preserved.

To apply this color transferring approach, the well-established methodology presented in [Reinhard 2001] is exploited. It is based on the $l\alpha\beta$ color space developed by [Ruderman 1998] which minimizes the correlation between the three image channels. This facilitates different operations in different color channels with the confidence that undesirable cross-channel artifacts will not occur [Reinhard 2001]. Additionally, this color space is logarithmic, which means to a first approximation that uniform changes in channel intensity tend to be equally detectable [Reinhard 2001]. In the implemented method, firstly, the RGB images are converted to $l\alpha\beta$ color space. There, the l axis represents an achromatic channel, while the α and β channels are chromatic yellow–blue and red–green opponent channels. To achieve the color transferring, the mean and standard deviations along each of the three axes are computed for both the source and target images. These mean values are subtracted from the image data points [Reinhard 2001] i.e. the values of each channel. Consequently, the resulted data points are scaled comprising the new image by factors determined by the respective standard deviations. After this transformation, the new data points (of the new image) have standard deviations that conform to the image. Next, instead of adding the averages that were previously subtracted, the averages computed for the image are added. Finally, the result is converted back to RGB via log-LMS, LMS, and XYZ color spaces and the new image is ready for further processing.

6.1.2.2 Feature detection, matching and epipolar geometry retrieval

Next, feature detection is performed using BRISK [Leutenegger 2011] detector. BRISK is adaptive, offering high quality performance as in other state-of-the-art algorithms, albeit at a dramatically lower computational cost (an order of magnitude faster than SURF in some cases) [Leutenegger 2011]. To detect matchable features, the generated binary images from the previous step are exploited here. To that direction, for detecting the necessary features, BRISK [Leutenegger 2011] is working only in areas that are classified as "non-caustics" in the binary mask. Even for this specific approach BRISK detector was chosen due to its lower computational cost. SIFT, SURF, ORB or other state-of-the-art detectors can also be used to detect the necessary features.

The detected features on an image are then matched to the corresponding features on the overlapping images and the mapping of these features between these

images is stored in a vector. This matching is based on n -space Euclidean distance and it is performed both from left-to-right and right-to-left for redundancy. However, since in feature matching, several blunders might occur, the RANSAC [Fischler 1981] algorithm is utilized to identify the inliers of the obtained point correspondences. The algorithm takes all the matched points as input, formulates a mathematical model which incorporates the majority of the points, and filters out the remaining points which are considered as outliers. To accomplish that, the fundamental matrix is computed and the measure for thresholding inliers points is the distance from the epipolar line. At the end of this step, a set of matched points is found in the given scenes.

6.1.2.3 Stereo-rectification and disparity maps generation

Knowing the epipolar geometry of the overlapping images, the initial imagery is then projected to form the stereo-rectified imagery in pairs. By using exactly the same matrices which describe the epipolar geometry of the two images, the predicted binary images resulted by the FCN are also projected in their stereo-rectified form. These projections are necessary to reduce the complexity of the problem across horizontal epipolar lines and facilitate the pixelwise image radiometric correction. Since for the vast majority of the diver acquired underwater imagery, the optical axes of the cameras are not parallel to each other, in order to have horizontal epipolar lines parallel to the baseline, the reprojection of both image planes onto a common plane parallel to the baseline needs to be performed. The rectification approach followed was firstly proposed in [Loop 1999] and involves the decomposition of each rectifying homography into a projective and an affine component. Then the projective component which minimizes a well-defined projective distortion criterion is found. The affine component of each homography is further decomposed into a pair of simpler transforms; one designed to satisfy the constraints for rectification, the other is used to further reduce the distortion introduced by the projective component.

The stereo-rectified imagery is then used for the disparity maps generation. This is achieved by stereo-processing the imagery by Semi-Global Matching [Hirschmuller 2007, Hirschmuller 2005], evaluating in 8 line directions instead of 16, to reduce the processing time. The exact approach followed is described in [Hirschmuller 2007] where the Consistent Semi-Global Matching (CSGM) is firstly presented. Compared with SGM, CSGM uses the same steps but also the intensity consistent disparity selection. It uses a pixelwise, mutual information based matching cost for compensating radiometric differences of input images, a feature necessary when processing images with caustics. Pixelwise matching is supported by a smoothness constraint that is usually expressed as a global cost function. CSGM performs a fast approximation by pathwise optimizations from all the 8 directions. Additionally, postprocessing steps for removing outliers, recovering from specific problems caused due to the caustics' effects and the interpolation of gaps are also applied.

Occlusions and mismatches can be distinguished as part of the left/right con-

sistency check. Regarding the interpolation for the gaps caused by the mismatched pixel areas on the caustics, it is performed by propagating valid disparities through neighboring invalid disparity areas. To achieve the best possible interpolation, this is done similarly to SGM along paths from 8 directions. According to [Hirschmuller 2007], where the implementing postprocessing was first presented, this approach emphasizes the use of all information without a preference to foreground or background. Moreover, instead of the mean, the median is used for maintaining discontinuities in cases where the mismatched area is at an object border. The implemented interpolation method has the advantage that it is independent of the used stereo matching method. Finally, median filtering can be useful for removing remaining irregularities and additionally smooths the resulting disparity image.

6.1.3 RGB values replacement and back-projection of the images

Considering that for each stereo-rectified image, the binary image of "non-caustics" and "caustics" classes and the disparity maps are available in the same projection, the pixelwise image correction can be performed. To that direction, for each pixel of an image that is not classified as "non-caustics" in the binary image, the RGB values of the corresponding pixel in the overlapping image are found, using the disparity map and vice versa. The relation between the source and the target pixels is expressed in Equation 6.1 and Equation 6.2.

$$x_{target} = x_{source} - disparity_{source} \quad (6.1)$$

$$x_{source} = disparity_{target} + x_{target} \quad (6.2)$$

where x_{target} is the horizontal coordinate of the pixel in the n+1 image of the stereo pair and x_{source} the horizontal coordinate of the pixel in the n image of the block. This process is performed only if the target pixel is classified as "non-caustics".

That way, the areas that are not classified as "non-caustics", are pixel-wisely corrected by exploiting the rigorous geometry of the stereo-pair. By using this approach, the disparity (x-parallax) of each pixel is taken into account, facilitating a more accurate and reliable correction approach that is not adversely affecting the later SfM and MVS steps, but it improves them. This is achieved as this horizontal displacement between rectified feature points is related to the depth of the feature. This way the method can be used to recover the images over 3D structures without affecting the 3D position of the SfM and MVS processing calculated later.

6.2 Summary

In this chapter, a novel method for correcting the radiometric effects of caustics on the underwater imagery in shallow areas was presented. The method firstly relies on state-of-the-art deep learning tools which can accurately classify the pixels of the image as "non-caustics" or "caustics" and then exploits the 3D geometry of the scene

in order to achieve a pixel-wise correction, by transferring color values between the overlapping images. The method depends on the good matches among the images, since an accurate fundamental matrix calculation is a prerequisite. However, if no good matches can be achieved, even when the detected masks are exploited in the key point detection step, images are not even appropriate for image-based 3D reconstruction, which this method intends to improve.

In the literature, the already available methods are based either on image co-registration being functional only on flat seabed, or on CNNs that are correcting the image in a similar way to inpainting methods. Contrarywise to those methods, the proposed one, achieves to correct the underwater imagery irrelevant of seabed types and anaglyphs by transferring true information between the overlapping images exploiting the recovered geometry.

Most importantly, although this approach was developed intending to correct the radiometrically affected areas on the underwater imagery, it can also be exploited to correct overwater, aerial and satellite imagery for specularities, shadows, and occlusions caused by illumination conditions, objects or even clouds.

Experimental Results and Validation of the Developed Pixelwise Correction Method

Contents

7.1	The first real world dataset on underwater caustics	157
7.1.1	Data Collection	159
7.2	Pixelwise image correction	161
7.2.1	Image classification	161
7.2.2	FCN selection and justification	167
7.2.3	Pixelwise image radiometric correction	169
7.2.4	Improvements on the 3D reconstruction	172
7.3	Discussion	176

This chapter, aims to present the results achieved after the implementation of the method presented in Chapter 6 regarding the pixelwise correction of the radiometric effects of refraction on the underwater imagery and specifically of the rippling caustics. Moreover, the first real world benchmark dataset on underwater caustics is presented. This dataset is then also used for the evaluation of the proposed image correction approach as well as the improvements performed on the 3D reconstruction by the developed radiometric correction methods. Section 7.1 presents the first real world dataset on underwater caustics and the respective details on data generation while in Section 7.2 the experimental results of the proposed pixelwise image correction method are presented and evaluated. Finally, Section 7.3 concludes the chapter by summarizing the performed work and discussing the contributions achieved by the presented approach.

7.1 The first real world dataset on underwater caustics

In the literature, the datasets presented in [Swirski 2013] are the only available related with rippling caustics. These datasets contain raw material and results of the underwater experiments performed in a pool, in the Red Sea and in the Mediterranean Sea. However, the images are of low resolution and the dataset is not created with the purpose of being used with machine learning and deep learning

frameworks due to the lack of ground truth images. Also, the scenes of this dataset are not representing a realistic set up for underwater image-based 3D reconstruction of seabed.

In this thesis, in order to fill this gap, the first real-world underwater caustics benchmark dataset which contains 1465 underwater images is presented (Figure 7.1).

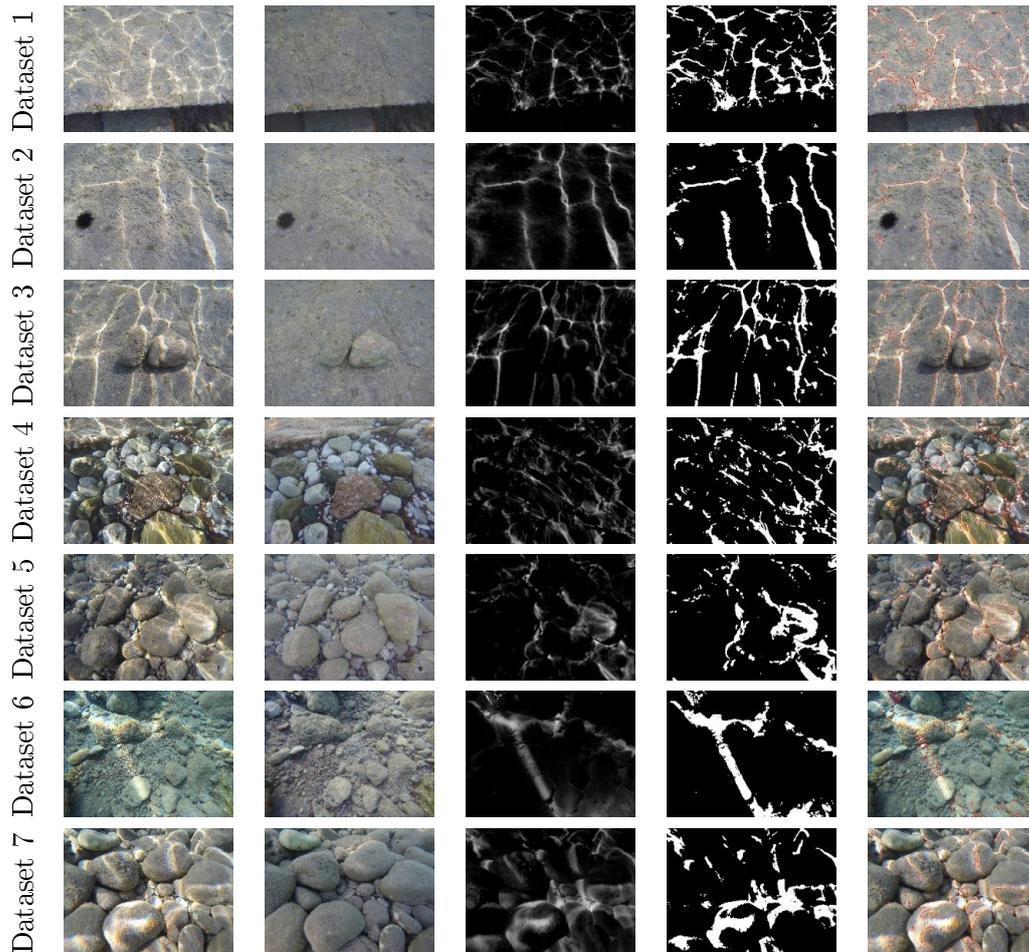


Figure 7.1: Sample images from the constructed dataset. First column: RGB images; Second column: the reference images; Third column: the difference images between the images with caustics and the reference images; Fourth column: the thresholded difference image; Fifth column: the RGB images with the detected contours.

Together with the RGB imagery, the corresponding generated ground truth images are delivered for facilitating the training and testing of machine learning and deep learning methods for image classification. The dataset contains 7345 images in total. The specific dataset also provides the necessary data to evaluate, at least to some extent, the performance of 3D reconstruction approaches.

7.1.1 Data Collection

Data were acquired using a GoPro Hero 4 Black action camera with image dimensions of 4000 x 3000 pixels, focal length of 2.77mm and pixel size of 1.55 μm and a tripod. Action cameras are widely used for underwater image acquisition [Menna 2018]. The dataset was captured in near-shore underwater sites at depths varying from 0.5 to 2m. No artificial light sources were used. Due to the wind, the turbulent surface of the water created dynamic rippling caustics on the seabed. In total 1465 images were collected, separated in 7 different datasets; five of them containing stereo images, one of them tri-stereo images and one consists of multi-stereo imagery acquired in 7 different camera poses (Figure 7.1).

As can be seen in Figure 7.1 where each row stands for a typical image from each dataset, the collected imagery presents a large variability in terms of scene complexity, color, caustics complexity, frequency and scale. Images were collected with five seconds interval to describe as much different instances of the caustics as possible. The 7 parts of the datasets capture flat and 3D seabed surfaces. It is also important to state that most of the parts of the dataset are appropriate for 3D reconstruction since they consist of at least two stereo images.

7.1.1.1 Reference and Ground Truth Image Generation

To capture the reference images of the dataset (Figure 7.1, second column), artificial shadow was introduced in the scene, above the water, in order to protect the water surface from the sun rays (Figure 7.2a). Since the images of the same branch of the dataset (i.e. left or right) are captured by the same camera position, an alternative approach was also implemented for the reference images generation. This was to create an image using the pixels having the lowest values compared with the pixels of the same position in the rest of the imagery. To perform that, images were transferred to the lab color space in order to choose the pixels with the lower luminosity. Although this approach is generating images with great quality, in some cases, in the created reference image objects appear that are apparent in one of the scenes, having lower values in the lab space compared with the rest of the pixels of this position. This led in many cases to the appearance of multiple fishes in the reference image, even in the branch of the dataset there is only one fish moving in the images (Figure 7.2b).

Having captured the reference images for each part of the dataset, the per-element differences $\Delta(x, y)$ between the reference and each one image of the dataset containing caustics is calculated and saved in a new image, the difference image (Figure 7.1, third column). To generate a more accurate difference image, containing only caustics and not other differences in the pixels' colors, a color transferring approach between the images containing caustics and the reference images is performed, as described in 6.1.2.1. Color transferring is performed from the images with caustics to the reference image. Also, since the difference image contains noise due to scattering and passing particles, the image is smoothed with a spatial Gaussian pre-filter with a kernel size of 3 x 3 to 7 x 7 pixels, depending on the scene.

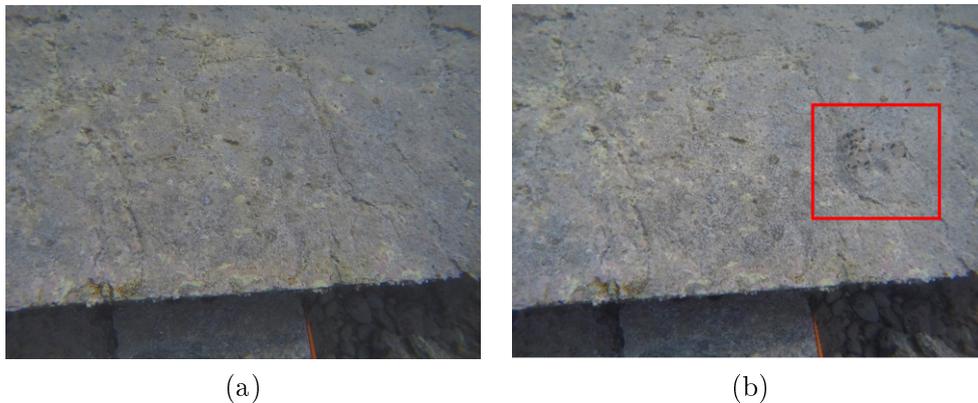


Figure 7.2: The reference image created using the artificial shadow in (a) and the reference image created using the pixels having the lowest values compared with the pixels of the same position in the rest of the imagery in (b). In the red rectangle the multiple fish appearing in the reference image are highlighted.

In order to compensate changes in illumination conditions during the acquisition phase and prepare the imagery for the ground truth image generation, the difference images' pixel values are scaled and shifted so their minimum value is 0 and the maximum 255. Finally, the thresholded difference images are created (Figure 7.1, fourth column). In order to ensure that the ground truth images do not include any false positives, a Canny edge detector [Canny 1987] is employed in order to facilitate the projection of the detected contours on the original images containing caustics for visual inspection (Figure 7.1, fifth column).

Although the above procedure is very consistent for generating the ground truth images, it must be noticed that caustics are very complex physical phenomena and by their nature they have not a very specific boundary. When the camera to object distance is quite small (i.e. 0.5m) and the incidence angle of the light rays to the surface of the water is also large, their boundary is characterized by very intense chromatic aberration effects. On the contrary, when the incidence angle is closer to the perpendicular, their boundary is of the same color values as the core of the caustics. The chromatic aberration intensity is also depending on the direction of the ripples compared with the lighting rays' direction. As the angle of the ripples with the lighting direction increases, the aberration increases too.

7.1.1.2 Underwater camera calibration

Since the collected dataset facilitates also the 3D reconstruction of the imaged areas of the seabed, underwater camera self-calibration [Zhang 2000] was performed in order to deliver the retrieved parameters together with the dataset. Self-calibration techniques do not use any calibration object. Just by moving a camera in a static scene, the rigidity of the scene provides in general two constraints [Maybank 1992] on the cameras' internal parameters from one camera displacement by using image information alone. Therefore, if images are taken by the same camera with fixed

internal parameters, correspondences between three images are sufficient to recover both the internal and external parameters [Hartley 1994]. Results are included in the dataset as a separate file. Below the camera calibration is given with unrefined principal point:

$$[f_x, f_y, c_x, c_y, k_1, k_2, p_1, p_2]=$$

$$[2416.65, 2424.26, 2000, 1500, -0.092734, 0.108091, 0.002016, -0.000773]$$

and with refined principal point:

$$[f_x, f_y, c_x, c_y, k_1, k_2, p_1, p_2]=$$

$$[2343.20, 2346.41, 1930.11, 1481.97, -0.101428, 0.098774, -0.000378, 0.000014]$$

The camera models used are of OpenCV [OpenCV]. For more information on the models, readers may refer to [OpenCV, Zhang 2000, Bouguet].

7.2 Pixelwise image correction

In this section, the steps of the proposed method for pixel-wise image correction are tested over real-world datasets, proving its robustness, accuracy and reliability.

7.2.1 Image classification

To form a reliable image classification model, the dataset presented in Section 7.1 was used to train, cross-validate and test the exploited FCN. To that direction and in order to prove and evaluate the potential of the FCN for generalization over different types of scenes, seven training-testing approaches were performed. To achieve that, the datasets were categorized based on the similarity of the scene; datasets 1, 2 and 3 are considered to have similar background, although caustics are of very different scale and complexity, datasets 5 and 7 are also considered similar. The rest of the datasets are not similar to each other. They represent totally different types of seabed and caustics.

7.2.1.1 Training and cross-validating the models

Considering the above, for the first case, the model is trained over datasets 4, 5, 6 and 7 and testing is performed over datasets 1, 2 and 3. For the second case, the model is trained over datasets 3, 4, 5, 6 and 7 and testing is performed over datasets 1 and 2. The third case involves training on 1, 2, 3, 4, and 6 and testing on datasets 5 and 7. In the fourth approach, the model is trained over datasets 1, 2, 3, 5, and 7 and tested on datasets 4 and 6. Finally, in the fifth, sixth and seventh approach, the model is trained over only datasets 1, 3 and 6 respectively and tested over the same datasets; 2, 4, 5 and 7. The number of the images used for training these three models are exactly the same with the previous training approaches. These

last three approaches aimed to prove the necessity of the variability of the scenes of the dataset and to reply to the question whether it is better to train the model using N images coming from one dataset or using N images coming from many datasets.

For training the FCN, a sliding window approach to extract 256 x 256 patches was used. The stride of the sliding window defines the size of the overlapping regions between two consecutive patches. At training time, a smaller stride allows us to extract more training samples and acts as data augmentation. At testing time, a smaller stride allows us to average predictions on the overlapping regions, which reduces border effects and improves the overall accuracy. During training, cross-validation and testing, a 32 pixels stride was used. Models are implemented using the Torch framework. Torch is a scientific computing framework with wide support for machine learning algorithms that puts GPUs first. All the models were trained using Stochastic Gradient Descent (SGD) with a base learning rate of 0.001, a momentum of 0.9, a weight decay of 0.0005 and a batch size of 5. The weights of the encoder in SegNet were initialized with those of VGG-16 trained on ImageNet [Deng 2009], while the decoder weights are randomly initialized using the policy from [He 2015]. The learning rate was divided by 10 after 7 epochs. In both cases, the learning rate of the pre-initialized weights was set as half the learning of the new weights as suggested in [Audebert 2018]. Regarding the number of the epochs of training, various applied approaches indicated that more than 10 epochs offer nothing more to the network's performance. To this direction, the network was trained on 10 epochs where for each epoch were used 10.000 samples acquired from 10 images of each dataset over 2000 iterations. For the training approaches on which only one dataset was used, the same number of samples was retrieved from 50 images.

Figure 7.3 presents the training loss and cross-validation accuracy for all the models trained. There, (a) presents the graphs for training on datasets 4, 5, 6 and 7, (b) presents the graphs for training on datasets 3, 4, 5, 6 and 7, (c) presents the graphs for training on datasets 1, 2, 3, 4 and 6, (d) presents the graphs for training on datasets 1, 2, 3, 5 and 7, (e) presents the graphs for training on dataset 1, (f) presents the graphs for training on dataset 3 and finally (g) presents the graphs for training on dataset 6. Figure 7.3 (h) is the legend. This legend applies also to Figure 7.4. Lines in green color represent the running-window cross-validation accuracy while lines in orange the cross-validation accuracy on all the validation data for each 100 iterations. Lines in red represent the training loss for the same running-window while lines in blue represent the mean loss for each iteration. In Figure 7.3, one can notice that the cross-validation accuracy (orange line) is over 92-94% and the mean training loss (blue line) is less than 0.15 even from the first 2500 iterations (1.25 epochs). However, large oscillation in the running window cross-validation accuracy and training loss is present in most of the training approaches. To investigate the source of this oscillation, initially the base learning rate is altered; In Figure 7.4 the training losses and cross-validation accuracies of the model trained on datasets 3, 4, 5, 6 and 7 using 0.01 (Figure 7.4a) and 0.0001 (Figure 7.4b) base learning rate are presented. These figures can be compared with Figure 7.3b.

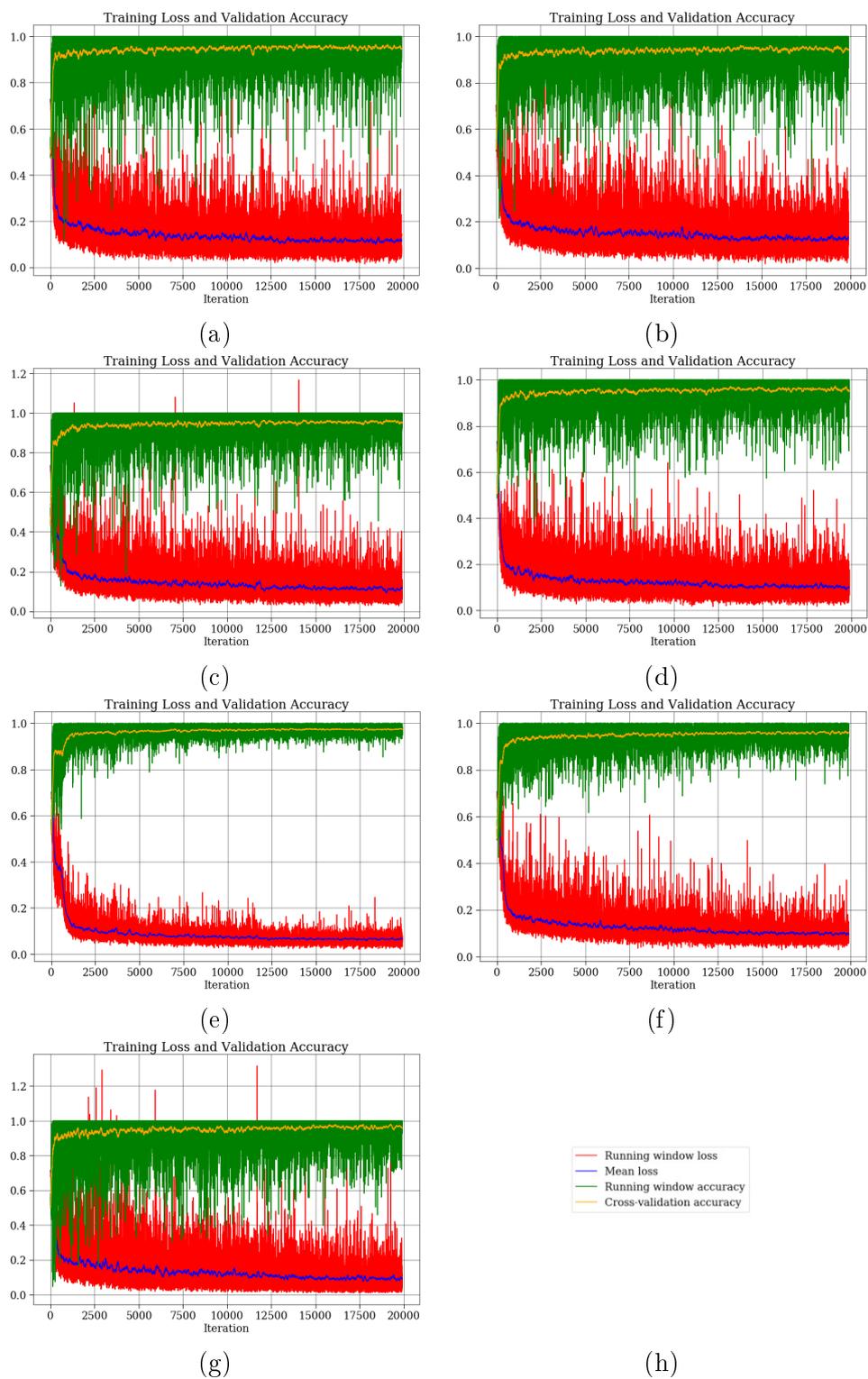


Figure 7.3: Training loss and cross-validation accuracy for all the models trained.

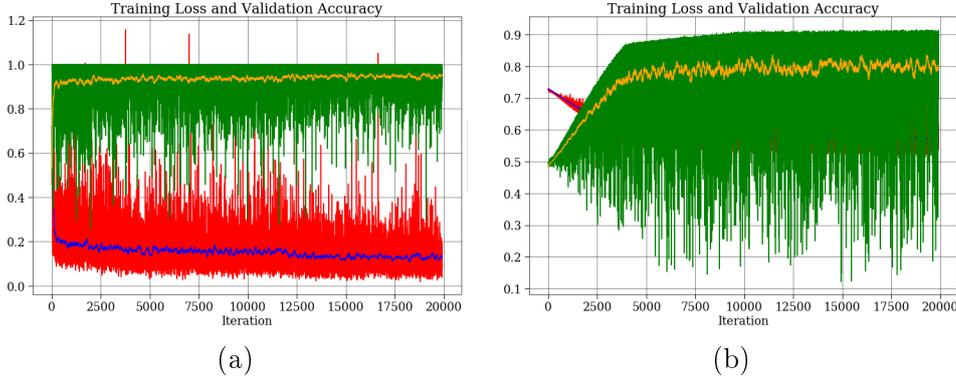


Figure 7.4: Training loss and validation accuracy for all the models trained on datasets 3, 4, 5, 6 and 7 for 0.01 base learning rate (a) and for 0.0001 base learning rate (b)

The oscillation observed in Figure 7.3 is directly related to the complexity and the clarity of the caustics on the datasets used for training; The model trained on dataset 1 presents the least oscillation from all the rest training approaches. The model trained on dataset 3 follows.

By observing Figure 7.4, it is clear that the reduction of the learning rate did not decrease the oscillation of the accuracy but on the contrary it increased it while the increase of the learning rate did not affect it. This implies that the oscillation of the running-window values is caused by the variance of the cross-validation data and especially the use of a random samples. Increasing the window size or reducing batch size will effect in reduction of the oscillation, however this is not the case for the performed experiments. Regarding the accuracy of all the cross-validation performed every 100 iterations for all the validation data and the mean loss, no remarkable oscillation is noticed, proving the reliability of the trained models.

7.2.1.2 Testing the models

To test the different models trained, several metrics are used; *precision* which gives the ability of a classification model to return only relevant instances, *recall* which gives the ability of a classification model to identify all relevant instances, *F1 score* which is a single metric that combines *recall* and *precision* using the harmonic mean and *accuracy* which is the ratio of the correctly labeled subjects to the whole pool of subjects. While *recall* expresses the ability to find all relevant instances in a dataset, *precision* expresses the proportion of the data points that the model considers as relevant and they are actually relevant.

$$precision = \frac{TP}{TP+FP} \quad recall = \frac{TP}{TP+FN} \quad accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

Where *TP* are the true positives: data points labeled as "non-caustics" that are actually "non-caustics", *FP* are the false positives: data points labeled as "non-caustics" that are actually "caustics", *TN* are the true negatives: data points

labeled as "caustics" that are actually "caustics" and FN are the false negatives: data points labeled as "caustics" that are actually "non-caustics". Table 7.1 presents the various metrics calculated after testing the several trained models on 10 epochs. It is obvious that for all the trained models, the achieved metrics are quite high, indicating the great generalization of the models. It is also evident that the models trained only on datasets 3 or 6 are achieving lower metrics compared to the rest, proving the necessity of the specific dataset as a whole.

The small variations observed in the metrics were expected due to the diverse characteristics of the datasets used for training and testing. It is also an indication of the performance of the trained models over different types of caustics. However, these variations are not affecting the overall accuracy of the model in a severe degree. Actually, these differences in the accuracy are mostly attributed to the indefinite boundary of the caustics effect and as such the differences between the predicted and the ground truth values, especially in Datasets 3 and 6.

Training data	Testing data	Precision (%)	Recall (%)	F1 (%)	Accuracy (%)
4,5,6,7	1,2,3	92.85	99.50	96.06	92.82
3,4,5,6,7	1,2	99.23	94.27	96.68	94.16
1,2,3,4,6	5,7	96.30	98.55	97.42	95.36
1,2,3,5,7	4,6	96.17	95.97	96.07	93.02
1	2,4,5,7	94.83	98.13	96.45	93.64
3	2,4,5,7	97.09	90.73	93.80	89.44
6	2,4,5,7	94.43	96.38	95.39	91.80

Table 7.1: The testing performance of the various models after 10 epochs of training.

In Figure 7.5 the input RGB images with the respective ground truth and predictions are illustrated. For the first and the second row of the figure which is showing the predictions over an image from 1 and 2 dataset respectively, the binary image is predicted by the model trained on 3, 4, 5, 6 and 7 datasets. Especially in the first row, it can be noticed that the pixels classified as "caustics" exceed those in the ground truth. These results come from the training of the model in more close-range imagery, giving it the ability to detect the intense chromatic aberration on the boundaries of the caustics. However, this cannot be achieved in the ground truth image. Following, the third and the fourth row present the predicted binary masks using the model trained over the 1, 2, 3, 4, and 6 datasets for images belonging to the fifth and the seventh datasets respectively. Finally, the fifth and the sixth rows are illustrating the predicted binary masks using the model trained on the 1, 2, 3, 5 and 7 datasets for images belonging to the fourth and the sixth datasets respectively.

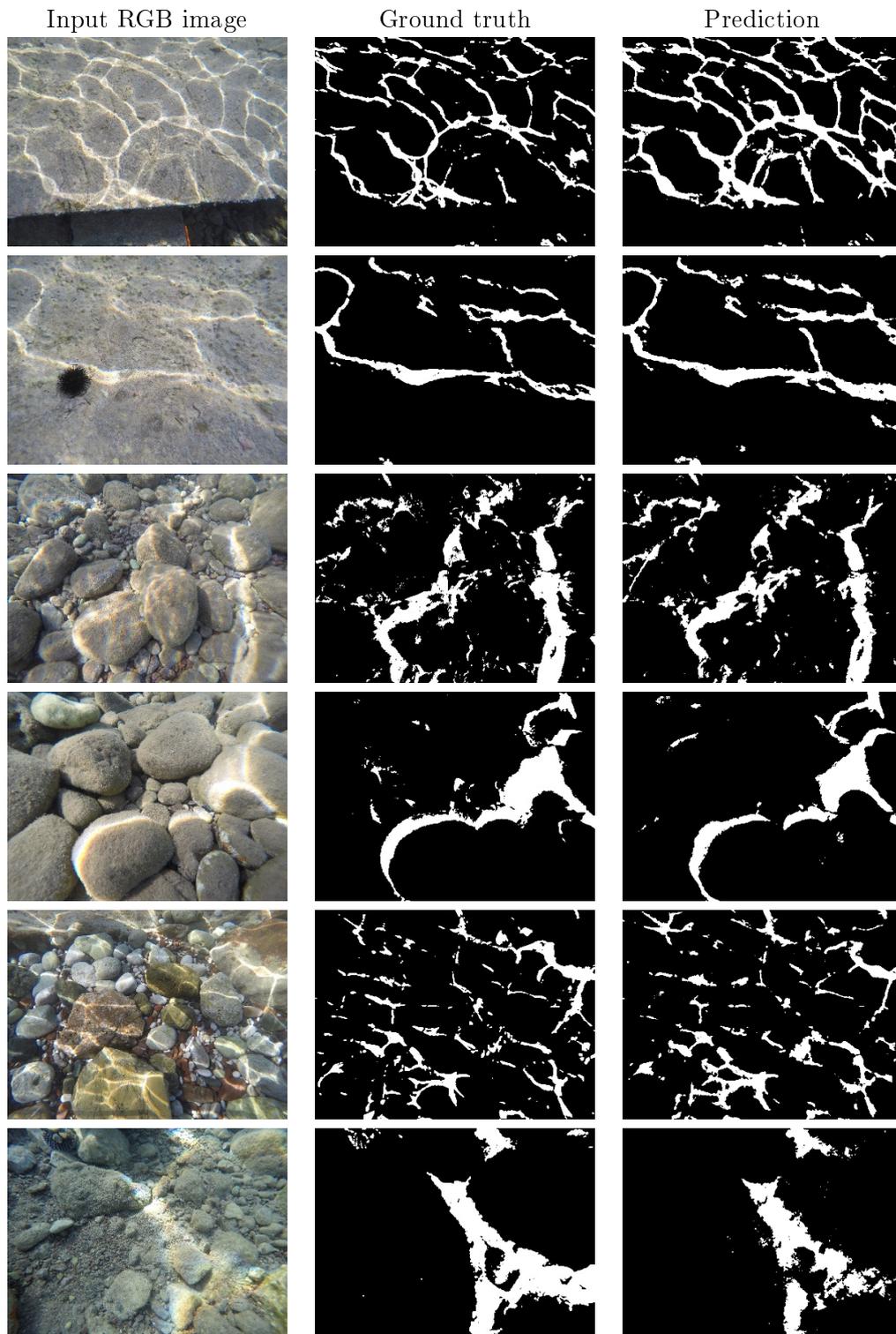


Figure 7.5: Typical examples of the RGB images (first column), the ground truth binary images (second column) and the predicted binary images (third column) using the models already described.

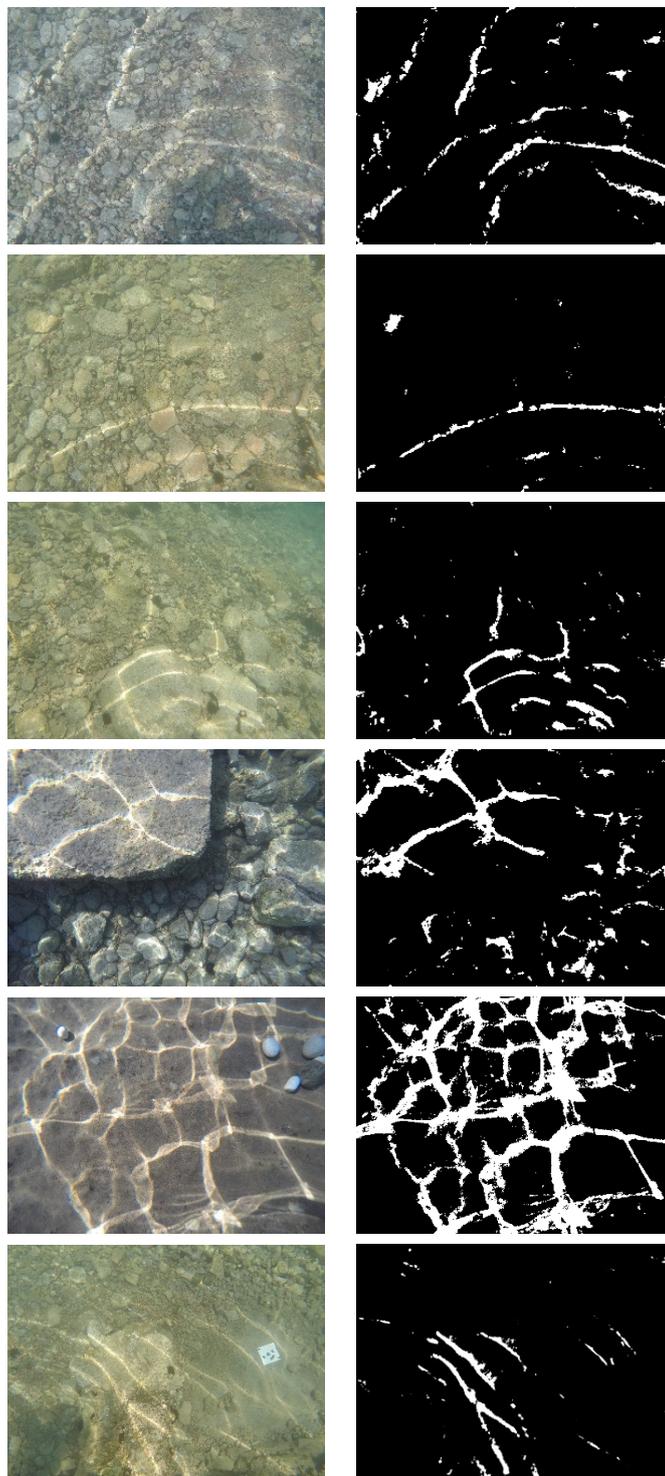


Figure 7.6: Various real-world RGB images (left column) and the respective predictions using the model trained on all the available datasets (right column).

Overall, the results delivered by all the trained models are very satisfying, clas-

sifying with very high accuracy and reliability the underwater images. Differences on the sensitivity of the models are expected, however this highlights the need to train a final model using all the available datasets.

To that direction, having proved the generalization capabilities of the exploited FCN, a model was trained exploiting all the available datasets, in order to be used for classifying unseen underwater imagery. For training this model, the same parameters as before are used. Results over images captured for real world underwater photogrammetric applications in shallow waters are presented in Figure 7.6. There, it is obvious that the trained model can generalize over different types of seabed and caustics with high reliability. A very important outcome is also that the model achieves to classify really bright areas of the image that are not caustics as "non-caustics" correctly. This can be seen on the last two images of Figure 7.6 where the white stones on the seabed and a white GCP are classified as "non-caustics". This example with the GCP is of really high importance for the underwater photogrammetric applications.

7.2.2 FCN selection and justification

Concerning the selection of a deep architecture for addressing the image classification problem, before the decision that a FCN would be more appropriate for the proposed solution, several tests were performed using thresholding and shallower architectures. Training and testing was performed for classifying the underwater imagery using AdaBoost [Freund 1995] and specifically AdaBoost-SAMME [Hastie 2009] having Decision Trees [Safavian 1991] as the base estimator, Decision Trees [Safavian 1991], K-Nearest Neighbors (KNN) [Altman 1992], Quadratic Discriminant Analysis (QDA), Random Forests [Breiman 2001] and linear Support Vector Machines [Schölkopf 2002]. Also, a simple thresholding approach was performed. Regarding the thresholding, its value was changed over the different images while for the rest of the methods, the default values as set in [Pedregosa 2011] were used. Resulting metrics from training over datasets 4, 5, 6 and 7 and testing over datasets 1, 2 and 3 are presented in Table 7.2. There, for reasons of direct comparison, the metrics achieved using the FCN, trained over only 1 epoch are presented too. However, these metrics are in accordance to the ones presented in Table 7.2.

As can be seen in Table 7.2, the metrics achieved by the FCN outperform the rest of the methods used for image classification. The Linear SVM follows with almost 10% less accuracy and after that, AdaBoost comes with almost 13% less accuracy. By comparing the precision score of the rest of the classifiers, they are achieving quite similar percentages, however this is not the case for the recall and the F1 scores where only the linear SVM and the AdaBoost classifiers are quite close. It is also evident that AdaBoost is outperforming the results of the Decision Trees, however this was expected.

For the majority of the performed tests, the shallower classifiers as well as the thresholding were not able to distinguish between caustics and other bright artifacts on the scene i.e. white rocks with high accuracy. Moreover, they did not succeed

in reliably detecting the boundary of the caustics when it is characterized by the intense chromatic aberration effect. This is also reflected in the achieved metrics in Table 7.2.

Two typical examples are presented in Figure 7.7 and Figure 7.8.

Training data	Precision (%)	Recall (%)	F1 (%)	Accuracy (%)
AdaBoost	88.07	90.00	89.02	80.48
Decision Trees	87.99	86.75	87.37	77.93
KNN	87.99	87.23	87.61	78.30
QDA	87.80	81.14	84.34	73.49
Random Forests	87.99	87.13	87.56	78.21
Linear SVM	88.09	91.19	89.61	81.40
FCN (1st epoch)	92.42	97.50	95.06	92.12

Table 7.2: The testing performance of the various methods. Number in bold are the best metrics achieved.

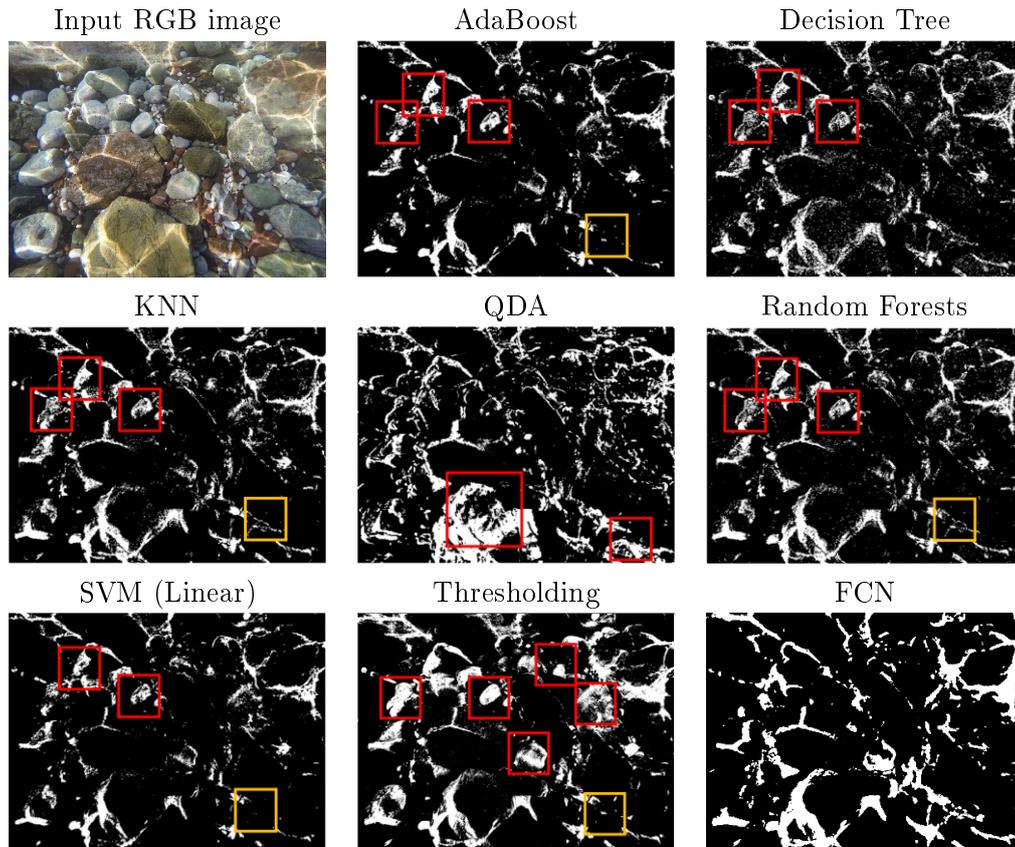


Figure 7.7: The input RGB image together with the predictions of the tested architectures for an image from dataset 4.

There, the input RGB image together with the predictions of the tested shallow architectures are given. In the red rectangles some typical cases of false negatives i.e. pixels that are wrongly classified as "caustics" are highlighted. On the other hand, orange rectangles are indicating areas of false positive, i.e. pixels that are wrongly classified as "non-caustics". This, together with the performed experiments over the different datasets, highlights the difficulty of those methods to generalize over the different types of seabed and caustics.

The metric and visual results of the above experiments, justify the use of a deeper architecture for classifying the underwater imagery affected by caustics, as also used in [Forbes 2018]. Even if the resulted metrics of most of the tests performed using the shallower architectures are lower than those achieved by the exploited FCN architecture, they do not prohibit their use. However, one has to consider the specific goal of the application. As such, for a more close to real time application, where only key point detection and matching is needed i.e. visual odometry etc., AdaBoost, Decision Trees, KNNs, QDA or even simple thresholding could be used due to their minimal prediction times, compared to the SVMs and the FCN architectures.

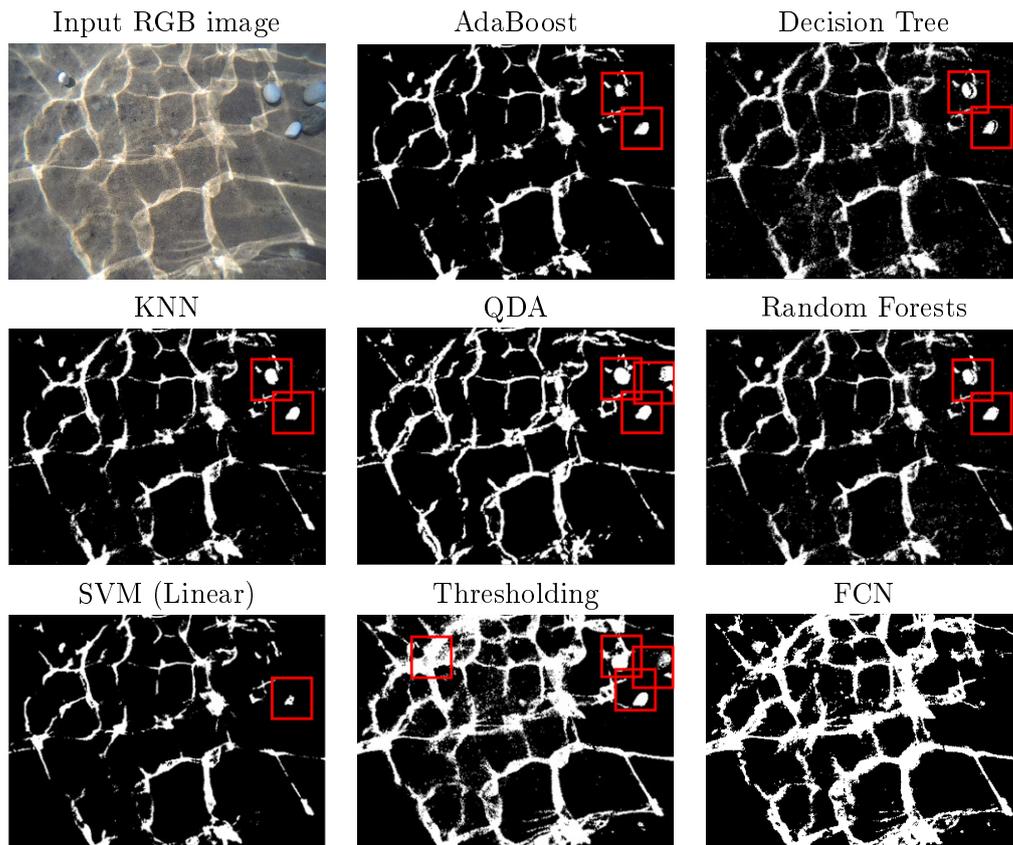


Figure 7.8: The input RGB image together with the predictions of the tested architectures for an image of a real world application, not contained on the datasets.

However, when it comes to the maximum improvements on the sparse and the

dense 3D reconstruction using the less necessary images, a deep architecture such as the selected FCN exploited in this thesis is the best solution. By classifying the pixels of the underwater imagery with the highest accuracy, only the pixels that it is really necessary to be replaced are undergoing the correction process, keeping as much as possible of the original imagery unprocessed. This way, there is more area on the image available for the keypoint detectors, ensuring a more robust SfM process. Additionally, as already proved, the FCN, achieves to detect in an accurate and reliable degree the intense chromatic aberration effects on the boundaries of the caustics, something really important for facilitating a more realistic corrected imagery and as will be shown in Subsection 7.2.4, enabling the generation of a more complete dense 3D point cloud.

7.2.3 Pixelwise image radiometric correction

In this section, results regarding the pixelwise image correction are presented and evaluated. Figure 7.9 presents typical examples of corrected images.

The first column depicts the original images with caustics, the second column depicts the binary images predicted by the trained FCN model and the third column demonstrates the corrected imagery. By observing the corrected imagery, it is clear that the proposed method achieves impressive visual results, since in the vast majority of the corrected images the rippling caustics are not apparent, and the replaced pixels are not clearly obvious. However, in some cases like the one presented in the first row of Figure 7.9, the corrected areas are obvious. This is a result of an inadequate color transferring approach, indicating great differences in luminosity between the overlapping images used for the correction. In fact, in some cases, these differences cannot be compensated by the color transferring only. However, as will be shown by the experiments presented in Subsection 7.2.4, this is not affecting the 3D reconstruction performance in a measurable degree. To solve this issue, Poisson blending [Pérez 2003] could be implemented on the boundary of the corrected with the uncorrected areas, however, this would affect the pixel values in an uncontrollable degree, affecting even more the SfM-MVS processes.

In order to demonstrate the importance of the color transferring step, proposed in this methodology, in cases where no extreme differences between the source and the target images are apparent, a typical example is given in Figure 7.10. There, a corrected image created by the proposed methodology with the color transferring step is presented in (a) while the same image corrected by the proposed methodology without applying the color transferring step is presented in (b). It is obvious that in the case where the color transferring is not applied, the replaced pixels are obviously enough, since they are characterized by lower RGB values.

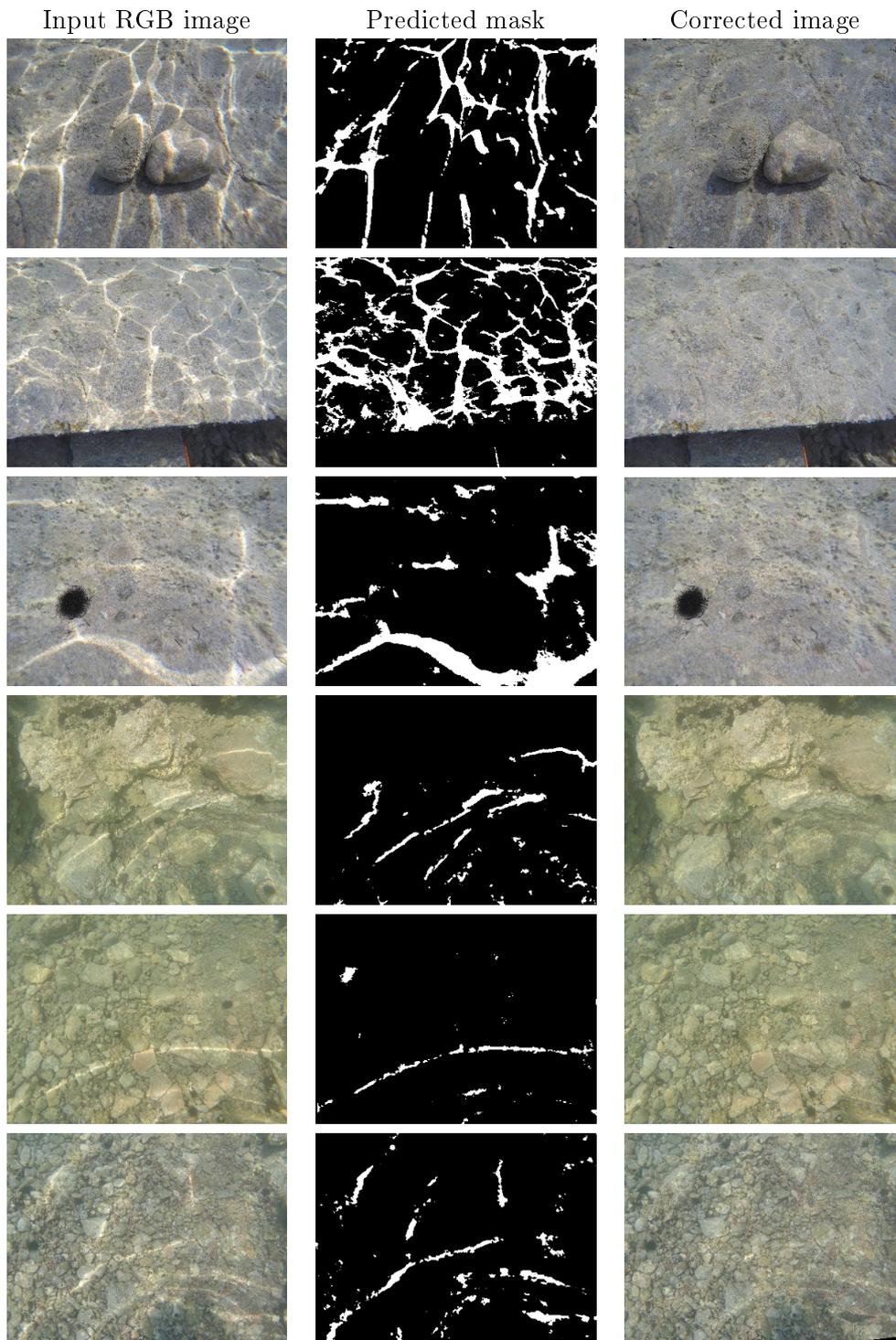


Figure 7.9: Examples of corrected images. The first column depicts the original images with caustics, the second column depicts the binary images predicted by the trained FCN model and the third column demonstrates the corrected imagery.



Figure 7.10: A corrected image by the proposed methodology with the color transferring step (a) and the same image corrected by the proposed methodology without applying the color transferring step (b).

Although it is not clearly obvious in the scale at which the images are presented in Figure 7.9, the proposed approach achieved really high accuracy in the pixel replacement process. When having a closer look to the corrected imagery, no offsets and other pixel displacements are obvious, especially in formations on the scene, being continuous between a corrected and a non-corrected area of the image.

7.2.4 Improvements on the 3D reconstruction

Following the pixelwise correction of the underwater imagery, the corrected data were processed with a commercial SfM-MVS software for evaluating the improvements on the 3D reconstruction, which was the initial objective of the effort. To that direction, six different test cases were processed. Typical images for these six test cases can be seen in Figure 7.9. The first three test cases are using images of the already presented benchmark dataset; the first test case is using tri-stereo imagery from Dataset 3, the second test case is using stereo imagery from Dataset 1 while for the third test case multiple-view stereo imagery (7 consecutive images) from Dataset 2 was used. These three test cases will facilitate the evaluation of the proposed correction methodology in terms of 3D reconstruction improvement, compared also to the generated 3D point clouds using the already available reference imagery. The 3D reconstruction results are presented in Figure 7.11. Results of the rest of the four datasets presented in Section 7.1 are not presented here since due to the complexity of the scene and the poor texture of the smooth and glossy rocks, the improvements are not easy to be highlighted, since the 3D point clouds of the reference imagery are incomplete too. As such, the datasets presented in Figure 7.11 are those that are generating the most complete 3D reconstructions, facilitating a detailed and direct comparison of the results.

By comparing the 3D point clouds generated using the reference imagery (left column of Figure 7.11) to the 3D point clouds generated using the uncorrected and the corrected imagery, significant differences can be noticed. It is clear that caustics are preventing a proper 3D reconstruction of the scene, since when images affected

by them are used, the resulting 3D point clouds are incomplete for all the tested cases. More specifically, DIM algorithms are failing to match the corresponding points of the affected areas and only the unaffected areas are appearing in the 3D point cloud. It is also important to highlight that as reported also in Table 14, when using the uncorrected imagery, not all of the images were aligned for the test case 1 and test case 3. In the latter, this resulted in a much less covered area by the 3D point cloud.

Coming to the third column, it is obvious that when the imagery corrected by the proposed methodology is used, the completeness of the 3D point cloud is improved to a great degree, delivering point clouds very similar to the ones generated by the reference imagery. Some insignificant differences can be observed on the perimeter of the reconstructed area. These differences are not attributed to some defect of the corrected imagery but on the extremely intense chromatic aberration that is apparent at the areas of the image having large radial distance when the luminosity of the scene is increased. This effect is not that intense on the reference images since they are characterized by lower illumination. In these cases, chromatic aberration is caused by the lens of the camera, as a result of the different refractive indices of the lens for each wavelength of light.

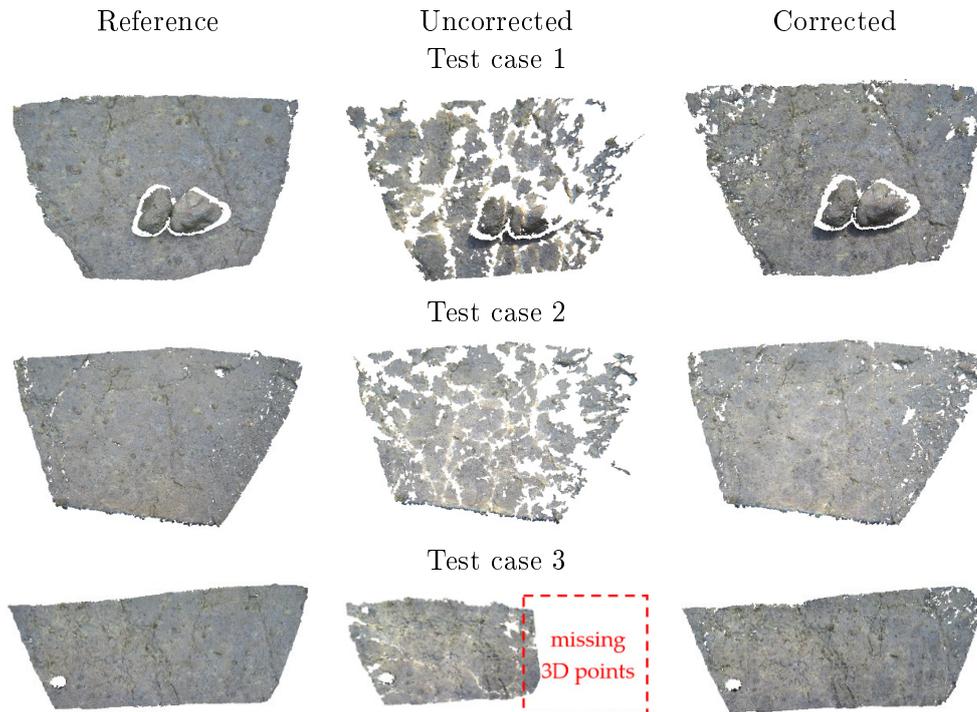


Figure 7.11: The 3D reconstructions using the reference imagery (left column), the uncorrected imagery (middle column) and the corrected imagery by the proposed methodology (right column). Each row refers to a different test case.

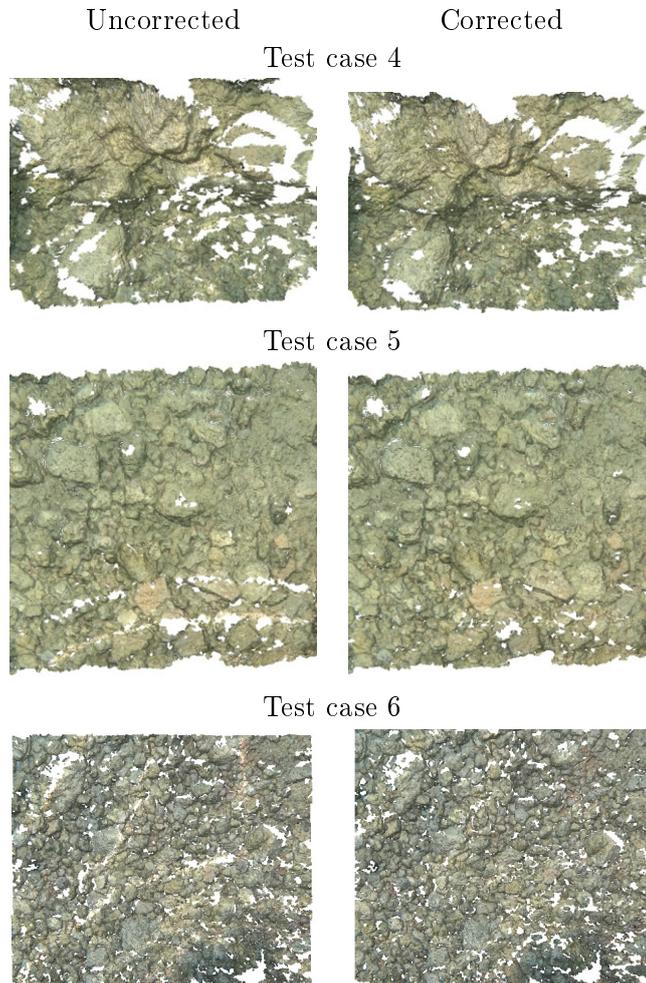


Figure 7.12: The resulting 3D point clouds for the rest three tests performed over real-world imagery. In the left column the 3D point cloud generated using the original uncorrected imagery is presented while in the right column, the respective 3D point cloud of the corrected imagery is presented.

In Figure 7.12, three different tests performed over imagery used for real underwater 3D reconstruction tasks are presented. Again, as in the first three test cases, when using real world imagery, the imagery resulted by the proposed methodology achieved to deliver a more complete 3D point cloud in the areas covered by caustics.

To facilitate a deeper evaluation of the improvements on the 3D reconstruction, all the 3D point clouds presented above were imported into Cloud Compare freeware [Girardeau-Montaut 2015] for further investigation. In particular, the following parameters and statistics that are widely used also in the literature for evaluating 3D point clouds [Agrafiotis 2017, Agrafiotis 2018b, Mangeruga 2018], were computed for each point cloud:

Total number of sparse and dense points. All the 3D points of the point cloud were considered for this metric, including any outliers and noise [Girardeau-Montaut 2015]. For the purposes of the work presented here, the total number of 3D points reveals the effect of the correction methodology on the matchable pixels among the images. The more corresponding pixels are found in the DIM step on the images, the more points are generated. Higher values of total number of points are considered better in these cases; however, this should be cross-checked with the surface density metric, since it might be an indication of noise on the point cloud.

Surface Density. The density is estimated by counting the number of neighbours N (inside a sphere of radius R) for each point [Girardeau-Montaut 2015]. The surface density used for this evaluation is defined as $\frac{N}{P_i \times R^2}$, i.e., the number of neighbours divided by the neighbourhood surface. Cloud Compare [Girardeau-Montaut 2015] estimates the surface density for all the points of the cloud and then it calculates the average value for an area of 1m² in a proportional way. Surface density is considered to be a positive metric, since it defines the number of the points on a potential generated surface, excluding the noise being present as points out of this surface. This is also the reason of using the surface density metric instead of the volume density metric.

Roughness. For each point, the "roughness" value is equal to the distance between this point and the best fitting plane computed on its nearest neighbour [Girardeau-Montaut 2015], which are the points within a sphere centred on the point. Roughness is considered to be a negative metric since it is an indication of noise on the point cloud, assuming an overall smooth surface. Table 7.3 presents the above metrics, together with the ratio of the aligned images by the total images, the average percentage of the pixels of the images containing caustics based on the predicted binary images, the number of the matched points (the number of the points of the sparse point cloud) and the respective percentage of the change achieved between the uncorrected and the corrected imagery, the number of the dense points and the respective percentage of change, the density D and the respective percentage of change and finally the roughness R .

By observing the metrics presented in Table 7.3, a first important outcome is that when the corrected imagery is used, more images are aligned in the image alignment step. Moreover, it can be noticed that for all the performed tests, when the corrected imagery is used, more points are matched, delivering a denser sparse point cloud and a more robust 3D geometry of the scene. These first two outcomes were expected, and the background is already reported in Subsection 2.3.2. The increase of the matched points, is proportional to the number of the images covered with caustics. However, there is not a strict relation between those two. This is not satisfied for test case 1 and test case 3, since not the same number of images are aligned. An increase in the number of the dense points is also observed. However, this is not of the same magnitude as that for the matched points. Again, here test case 1 and test case 3 are excluded for the same reasons as before. Coming now to the density of the point clouds and their roughness, no significant differences can be

observed. Regarding the density, this was expected since the unaffected areas of the images are remaining the same, so there is no reason for generating more 3D points there. However, the fact that the roughness of the point clouds is remaining the same is of great importance; In the literature [Mangeruga 2018] it is reported that most of the underwater image enhancement methods are increasing the roughness of the generated 3D point clouds. Nevertheless, this does not apply for the proposed methodology, highlighting the accuracy and the quality of the performed pixelwise image correction.

Metric		Images	Caustics	Matched	Matched	Dense	Dense	D	D	R
Test case		aligned/ total	in pixels [%]	points	points change [%]	points	points change [%]		change [%]	
1	uc	2/3	28.1	1327		37505		7764.99		0.003
	c	3/3		2210		+66.5		72261		
2	uc	2/2	15	606		1783027		106908.18		0.001
	c	2/2		820		+35.3		2118934		
3	uc	5/7	15.4	2562		556926		12364.76		0.001
	c	7/7		4173		+62.9		6137345		
4	uc	2/2	8	3920		188203		1590.53		0.007
	c	2/2		4916		+25.4		190442		
5	uc	2/2	2.7	4315		144435		3146.87		0.004
	c	2/2		5257		+21.8		147948		
6	uc	2/2	4.5	2433		253778		3996.95		0.003
	c	2/2		3075		+26.4		259839		

Table 7.3: The evaluation metrics between the uncorrected (uc) and the corrected (c) images. D is the surface density and R the roughness. Numbers for these two metrics are relative.

7.3 Discussion

In this chapter, the novel method proposed in chapter 6 for correcting the radiometric effects of caustics in the underwater imagery in shallow areas was tested and evaluated. Towards that, initially, the first real world benchmark dataset on underwater caustics detection has been presented. This dataset facilitated the training of the exploited FCN in order to achieve high classification accuracy. The dataset was also used for the evaluation of the proposed pixelwise image correction approach as well as the improvements performed on the 3D reconstruction by the developed radiometric correction methods.

Since the method firstly relies on an FCN based on SegNet architecture, various tests were performed to justify its use and to prove that the exploited FCN model

can generalize over different types of seabed and caustics. This is a very important feature of the proposed algorithm since it will facilitate a more accurate and reliable pixelwise image correction in all underwater conditions. Results suggest that the exploited framework achieves very high accuracies in this binary classification problem even from the first 5 epochs. This is attributed to the use of the pretrained weights of VGG-16 on ImageNet [Deng 2009], which require only some more epochs to update the weights over the new data and fine tune the layers of the FCN and of course to the exploited dataset for training which contains many instances of the phenomenon over different types of seabed, giving to the model the capability to generalize early. The different accuracies achieved over the different datasets were also expected due to the diverse characteristics of the datasets used for training and testing. However, these variations in the accuracy are not affecting to a severe degree the overall accuracy of the model. These differences in the accuracy are mostly attributed to the ill defined boundary of the caustics effect and as such they represent the differences between the predicted and the ground truth values. Differences on the sensitivity of the model were also expected, however this highlighted the need to train a final model using all the available datasets. A test performed by training the model only on one dataset and testing it over the rest proved that when more datasets are used for training, better results are achieved. Results over images captured for real world underwater photogrammetric applications in shallow waters suggested that the trained model can generalize over different types of seabed and caustics with high reliability.

Following the testing and the evaluation of the FCN model used for the detection of the unaffected areas, the pixelwise image correction was performed on overlapping images of the presented dataset as well as on overlapping images from real world underwater mapping applications. Experiments performed illustrated the robustness and the reliability of the method over different types of seabed, different types of caustics and different anaglyph of the scene. The need of the color transferring step that is proposed was also highlighted.

The final step of the evaluation of the method concerned the improvements on the 3D reconstruction of the scenes. There, the effectiveness of the proposed method was clearly obvious, since complete 3D point clouds were delivered, leaving no doubts about the achieved results. It was proven that when the corrected imagery is used for performing 3D reconstruction tasks, more images are aligned, and more points are matched. This delivers a more robust, complete and reliable 3D reconstruction. Moreover, it was also considered very important that the proposed method did not increase the roughness of the generated dense point clouds for all the testing cases. Most importantly, although this approach was developed intending to correct the radiometrically affected areas on the underwater imagery, it can also be exploited to correct overwater, aerial and satellite imagery for specularities, shadows, and occlusions caused by the illumination conditions, objects or even clouds.

Till now, when no action is taken for the radiometric effects of the refraction, a huge number of small-base consecutive images have to be captured in order to facilitate a possible successful matching between the images and a more complete

3D reconstruction. The proposed method will enable the users to capture less images and deliver more complete results, covering also larger areas. This will reduce the SfM-MVS processing time and the revisit of the underwater site for extra data will be avoided. To prove that, for the performed experiments the less possible images were used i.e. mainly two overlapping images, indicating that when the corrected images are used, there is no need for a huge amount of data to fill the gaps in the 3D model caused by the rippling caustics effect.

A reasonable question of the reader may be that since the disparity maps are used for the pixelwise image correction, based on the results presented regarding the 3D dense point clouds of the affected images, they should present holes in the areas classified as caustics. As also reported in Section 6.1.2, to overcome this issue, postprocessing steps for removing outliers, recovering from the problems caused due to the caustics effects and the interpolation of gaps are applied. To make the interpolation more efficient, the disparity maps are generated with the epipolar imagery being downscaled to 10%-25% of the original size, adjusting also the size of the kernel for the census transform and the size of the kernel for blurring the images and median filtering. Then the disparity maps are upscaled back in the original epipolar images size.

By using the proposed method, a more chromatically consistent and realistic representation of the seabed is achieved. From the experiments performed, some remaining artifacts noticed on the images, especially in the areas having large radial distance, are attributed to the really intense chromatic aberration effect. However this was not to be dealt with within the context of this thesis. It was found that this effect is negatively affecting mainly the SfM-MVS process, compared with the skipping of the color transferring step. To overcome these issues, a channel-based correction of the refraction effect on the water-lens-air interface has to be performed. However, it can be easily avoided by using dome ports instead of flat ports and by avoiding fish-eye lenses and generally very small focal lengths.

Finally, it should be noticed that other methods tested for correcting the underwater imagery i.e. a piecewise linear transformation between the overlapping areas of the images, exploiting the matched keypoints as control points for the transformation, fail to deliver an accurate and reliable 3D reconstruction, since the x-parallax of the 3D objects is not taken into account, delivering point clouds of different depths.

Conclusion of the Thesis

Contents

8.1	Summary	179
8.2	Contributions	181
8.3	Future Work	183
8.4	Research Framework and Related Publications	184
8.4.1	Compensation for Refraction's Geometric Effects	185
8.4.2	Compensation for Refraction's Radiometric Effects	185
8.4.3	Close Range Underwater 3D reconstruction and Color Enhancement	186

This chapter concludes the work presented throughout this thesis. Firstly, it summarizes the whole thesis by reviewing the contents presented in each chapter and then, it points out the major original research contributions that were proposed in the text. In addition, some interesting future research issues are mentioned in the respective section. Finally, the publications related to this work are listed, accompanied by the research framework of the thesis.

8.1 Summary

Accurate bathymetric mapping proved to be a key element for offshore activities, hydrological studies such as coastal engineering applications, sedimentary processes, hydrographic surveying as well as archaeological mapping and biological research. Structure from Motion (SfM) and Multi View Stereo (MVS) techniques represent a valuable tool for mapping the shallower areas of the clear-water bodies and images can provide a low-cost alternative compared to other methods such as LiDAR and acoustic surveys, offering important visual information as well. To that direction, this thesis was concerned with the development of new methods for improving the performance and the accuracy of image-based mapping and 3D reconstruction of the bottom in shallow waters for small- and large-scale surveys.

Chapter 2 presented the theoretical background and the elementary principles behind the work carried out in the thesis. The chapter reports on the properties of the water and then focuses on the geometric and the radiometric effects that are caused by the optical water properties and specifically the refraction effect. Following the Snell's Law description, the geometric effects of refraction on depth

determination are explained for the single, the stereo and the multiple view geometry cases. The chapter finally discusses the radiometric effects caused by the refraction and focuses on the rippling caustics creation and their effect on the underwater close-range image-based 3D reconstruction and mapping.

Chapter 3 reviewed the most important work related to this thesis worldwide. Following an extensive report on the most representative works carried out so far regarding the correction of the geometric effects of refraction towards bathymetry estimation, interesting observations resulted. Among them, it was found that refraction correction on image-based bathymetry mapping is still in its initial phase and a relatively reduced number of approaches have been reported so far. However, it can definitely benefit from the recent advances in photogrammetry and image-based 3D reconstruction, exploiting SfM and MVS techniques. Till now, most of the published methods are dedicated to specific applications and specific data, using limited models that are preventing their generalization over different cases. Moreover, most of the works found in the literature are dealing with only one test site and lack cross validation with other test sites, using different image acquisition systems and parameters. On the contrary, in this thesis both artificially generated synthetic datasets and various different real world test sites are exploited for testing and evaluating the proposed approaches. The performed cross-testing between the different synthetic and real-world test sites, proved the potential for the generalized applicability and the data independency of the proposed solutions to the refraction problem which affects the aerial image-based bathymetry mapping. Following, the most representative works carried out so far towards the correction of rippling caustics effects are introduced. Despite the innovative and complex techniques mentioned, addressing caustics removal with procedural methods requires that strong assumptions are made on the many varied parameters involved e.g. scene rigidity, camera motion, etc. On the contrary, the proposed method for pixelwise image correction stands independently of the seabed anaglyph and the camera motion being able to generalize and deliver accurate and reliable results over different underwater environments.

Chapter 4 presented the two different methods developed for correcting the geometric effects of refraction. There, the mathematical background of the two proposed methods is given and the specific selections of methods and models is justified. In contrast with the rest of the state-of-the-art methods for correcting the geometric effects of refraction in the overwater imagery, both methods are independent of the system to be used, thus they can generalize over different conditions and seabed mapping tasks.

The following Chapter 5 presented the experimental results and the validation performed for the two proposed methods for overcoming the geometric effects of refraction. To that direction, synthetic data were also used for the proof of concept as well as for applying a transfer learning approach, independently of the quality of the reference data to be used. Following, the experimental results of the two proposed methods over five real world test sites were presented, accompanied with extensive analysis of the results. This chapter concludes with the comparison of

these two methods with other three state-of-the-art methods. Results indicated that the proposed methods outperform the rest of the examined methods, while the proposed Method 2 outperforms the proposed Method 1.

Chapter 6 introduced the reader to the second part of this thesis; the correction of the radiometric effects of the refraction. There, a totally new approach for dealing with this issue is described, exploiting an FCN and the n-view geometry of a scene. The method is designed in such a way to deal with every type of caustics and every anaglyph of the seabed, correcting in a pixelwise method the underwater imagery, exploiting real pixel information from the overlapping images.

Chapter 7 presented the experimental results and the validation performed in order to prove the generalization and the applicability of the method as well as to highlight the improvements achieved on the 3D reconstruction. Results suggested that the SegNet based FCN model can achieve really high accuracy in classifying the not affected areas of the image, being trained in very few epochs. However, to achieve this, real world training data have to be used, like the ones included in the first real world dataset on underwater caustics presented in this chapter too. Various 3D reconstructions performed using the corrected and the uncorrected imagery proved the significance of the contributions of the methods, since a more complete 3D point cloud is delivered when the first images are used.

8.2 Original Contributions of the Thesis

The work described in the previous sections has accomplished the aim of this thesis for improving the efficiency, the accuracy and the applicability of the image-based methods in respect to overwater and underwater 3D mapping of the bottom in shallow waters by addressing geometric and radiometric issues caused by the refraction effect. In the development of this goal, various important research contributions were achieved. Although these contributions are listed in Subsection 1.2.2, they are listed again below in more detailed view, reporting also on the results achieved:

Development of two different methods for correcting the geometric effects of refraction in overwater bathymetry mapping in shallow waters: The first method developed, to address the systematic errors caused by the refraction effect on the point clouds derived from SfM-MVS processes in a generalized, global and accurate way, exploits Support Vector Machines. The developed method, is based on a linear SVR model, which can accurately recover shallow bathymetric information from low altitude aerial image datasets over a calm water surface, supporting several coastal engineering applications in non-turbid waters and textured bottoms. The SVR model is less sensitive to noisy inputs such as the point clouds that are generated from images with refraction, building a more robust and generalized model, predicting depths over different cases and test sites, outperforming Simple Linear Regression. The linearity of the relation between the

real and the apparent depths resulting from the SfM-MVS processing was proved using synthetic data, revealing also uncertainties related with the SfM-MVS process itself being apparent in every bathymetric application based on those.

The second method developed, is built upon and exceeds the state-of-the-art method presented in [Skarlatos 2018] and firstly exploits the machine learning procedures that recover depth on the derived image-based dense point clouds, which is an additional contribution of this thesis, and then corrects the refraction effect on the original imaging dataset. This way the operational SfM and MVS processing pipelines are executed on a refraction-free set of aerial datasets resulting into highly accurate bathymetric maps and image-based products, when calm water surface, non-turbid waters and textured bottoms exist. This method outperforms in accuracy the first one, however both of them are outperforming the rest of the compared state-of-the-art algorithms. It is also of most of importance that this second method achieves an important reduction on the noise of the sparse point clouds which resulted from the SfM process and improves the accuracy and the quality of the produced orthoimages and textures.

High bathymetric accuracy in the through-water imagery techniques and generalization of the developed refraction correction methods:

For both developed methods for correcting the refraction effects, experimental results and validation over synthetic and real-world data demonstrated their high potential, both in terms of bathymetric accuracy as well as texture and orthophoto quality. For both developed methods for correcting the refraction effects, frameworks were designed in such a way to achieve high generalization over different low-altitude aerial platforms, cameras, flight heights, flight patterns and overlaps. Experiments and validation performed extensively over synthetic and real-world data, also applying a transfer learning approach between the synthetic and the real world datasets, demonstrated the scalability and transferability of the methods that proved to be independent of specific applications.

Development of a novel method for correcting the radiometric effects of refraction in close-range underwater 3D mapping in shallow waters:

A new method was proposed which contrarywise to the state-of-the-art can confront seabed of any anaglyph. This method exploits a SegNet based FCN to classify the image pixels as affected and non-affected. Then based on the recovered geometry of the stereo pairs and specifically the epipolar geometry and the generated disparity maps, corrects the affected pixels of the images by exploiting the pixels' stereo correspondences. This way, only true information is used for correcting the imagery, avoiding artificial pixel values predicted by machine learning and deep learning techniques. This approach is not restricted only on this specific problem and it could be easily implemented for addressing problems caused by permanent or temporal occlusions on the images such as clouds on the aerial and satellite imagery.

Improvement on the 3D reconstruction when radiometrically affected images are used:

By detecting the radiometrically affected and non-affected areas of the images, key point detection and description can be only performed on the non-affected areas. This facilitates the matching of the imagery, bypassing the effects of caustics. However, by correcting the underwater imagery, a more complete MVS step can be performed, delivering more 3D points with no gaps in the areas covered by caustics. Testing on real-world underwater imagery indicated that the proposed method for pixelwise image correction increases the completeness of the 3D point clouds, reaching the levels of the 3D point clouds being generated using images without caustics.

Creation of the first large-scale real-world underwater caustics benchmark dataset:

The dataset contains 1465 underwater images. Moreover, the corresponding reference results for all imagery are provided together with generated ground truth images for facilitating the training and testing of new machine learning and deep learning methods. The dataset counts 7345 images in total. The specific dataset provides an opportunity to evaluate, at least to some extent, the performance of different caustics detection and image classification approaches. It also facilitates the evaluation of the 3D reconstructed scenes, either metrically or by direct visual comparison of the generated results.

8.3 Future Work

During the development of this research work, more problems and topics of interest for future research have arisen. The following points are the ones which have been found as the most logical directions to continue this research:

Improvements on the key point detection and matching processes regarding the overwater imagery. Further research on this will facilitate the generation of less noisy point clouds, increasing the accuracy of the proposed methods and of image-based bathymetry in general. Issues to be considered are also the exploitation of more sophisticated key point detectors, based on machine and deep learning, to avoid matches that are resulting in highly noisy sparse 3D point clouds. This also concerns the underwater imagery being affected by caustics where a solution could be a way to learn the detectors to avoid caustics' pixels similar to [Papadaki 2020]. However, this would not improve the MVS step.

Improvements on the bundle adjustment and self-calibration methods regarding the overwater imagery. It is a fact that the self-calibration methods applied on the overwater imagery are not delivering very consistent results.

Also, it would be of really high importance for the users to firstly perform the SfM step on the images containing only land and then match on them the rest of the imagery containing sea. This way, the block on the dry part of the model will be more reliable, delivering also the self-calibration parameters to be used for the rest of the block over the water.

Further exploitation of synthetic data, since the ground truth data collection are a very time consuming and costly process when it comes to large scale seabed mapping. This could expand the use of recent advances in learning to a great degree, since more training and testing data would be available. This could also improve color consistency, both on the overwater and the underwater imagery.

Fuse methods for SDB and low-altitude aerial imagery in order to overcome errors caused by the SfM-MVS processes by exploiting the color information. Work on this area will merge these approaches and would deliver more accurate and reliable bathymetric information. It would be also to bypass the errors caused by mismatches and wrong correspondences.

8.4 Research Framework and Related Publications

The results and conclusions presented in this thesis have been possible after the realization of countless tests and experiments, which were the fruit of numerous efforts done in the development of the different methods. All the work done during the evolution of this thesis is summarized here with references to the most relevant research publications done by the author. The complete list of publications can be consulted in the next Section.

Regarding the compensation for refraction's geometric effects, even the related publications to the thesis started on 2018 with [Skarlatos 2018], the author was holding deep knowledge on the field, already from 2012 [Georgopoulos 2012, Agrafiotis 2015]. Having already this background knowledge, the author initially contributed in [Skarlatos 2018], which is largely based in the outcomes of his previous research presented in [Agrafiotis 2015] and [Georgopoulos 2012]. Then, he led the research on the latest three publications [Agrafiotis 2020, Agrafiotis 2019b, Agrafiotis 2019a] where he presented his approaches towards the shallow water bathymetry mapping from UAV imagery based on machine learning.

Coming to the compensation for refraction's radiometric effects, the author has published [Agrafiotis 2018a]. There, the main challenges of underwater photogrammetry in very shallow waters are reported and analyzed while a caustics effect removal approach based on CNNs is applied and evaluated.

All the above research, was surrounded by several different triggers for research on the field published in the following Section 8.3.3 "Close Range Underwater 3D reconstruction and Color Enhancement". Most of those (except

[Agrafiotis 2018b, Menna 2018, Agrafiotis 2017] are related with the EU Research and Innovation Action (RIA) project "iMARECULTURE" on which the author served as a Researcher C and Project Manager. Through this project and other projects like the Nissia shipwreck (I and II) and Mazuts shipwreck excavation projects, the author had the opportunity to have a really wide overview of the state-of-the-art in underwater 3D reconstruction and mapping, in underwater color correction and enhancement as well as in underwater Augmented Reality, collaborating and sharing knowledge with researchers from many different disciplines and fields. Finally, during the same period, the author was actively involved in EU FP7 project INACHUS and various local Greek projects publishing also high quality research, which, however, is not included in this thesis.

8.4.1 Compensation for Refraction's Geometric Effects

Agrafiotis, P., Karantzas, K., Georgopoulos, A., Skarlatos, D. Correcting Image Refraction: Towards Accurate Aerial Image-Based Bathymetry Mapping in Shallow Waters. *Remote Sensing*, 12, 322, 2020

Agrafiotis, P., Skarlatos, D., Georgopoulos, A., Karantzas, K. DepthLearn: Learning to Correct the Refraction on Point Clouds Derived from Aerial Imagery for Accurate Dense Shallow Water Bathymetry Based on SVMs-Fusion with LiDAR Point Clouds. *Remote Sensing*, 11(19), 2225, 2019

Agrafiotis, P., Skarlatos, D., Georgopoulos, A., and Karantzas, K.: SHALLOW WATER BATHYMETRY MAPPING FROM UAV IMAGERY BASED ON MACHINE LEARNING, *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-2/W10, 9-16, <https://doi.org/10.5194/isprs-archives-XLII-2-W10-9-2019>, 2019.

Skarlatos, D., Agrafiotis, P. A Novel Iterative Water Refraction Correction Algorithm for Use in Structure from Motion Photogrammetric Pipeline. *Journal of Marine Science and Engineering*, 6(3), 77, 2018

8.4.2 Compensation for Refraction's Radiometric Effects

Agrafiotis, P., Skarlatos, D., Forbes, T., Poullis, C., Skamantzari, M., Georgopoulos, A. UNDERWATER PHOTOGRAMMETRY IN VERY SHALLOW WATERS: MAIN CHALLENGES AND CAUSTICS EFFECT REMOVAL. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42, 2, 2018

8.4.3 Close Range Underwater 3D reconstruction and Color Enhancement

Skarlatos D., Agrafiotis P. Image-Based Underwater 3D Reconstruction for Cultural Heritage: From Image Collection to 3D. Critical Steps and Considerations. In: Liarokapis F., Voulodimos A., Doulamis N., Doulamis A. (eds) Visual Computing for Cultural Heritage. Springer Series on Cultural Computing. Springer, Cham., 2020 (co-first/equal authorship)

Vlachos, M., Berger, L., Mathelier, R., Agrafiotis, P., Skarlatos, D. SOFTWARE COMPARISON FOR UNDERWATER ARCHAEOLOGICAL PHOTOGRAMMETRIC APPLICATIONS. International Archives of the Photogrammetry, Remote Sensing Spatial Information Sciences, 2019

Skarlatos, D., Menna, F., Nocerino, E., Agrafiotis, P. Precision Potential of Underwater Networks for Archaeological Excavation Through Trilateration and Photogrammetry. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 42(2/W10), 175-180, 2019

Agrafiotis, P., Drakonakis, G.I., Skarlatos, D. Underwater Image Enhancement before ThreeDimensional (3D) Reconstruction and Orthoimage Production Steps: Is It Worth? In Latest Developments in Reality-Based 3D Surveying and Modelling; Remondino, F., Georgopoulos, A., González-Aguilera, D., Agrafiotis, P., Eds.; MDPI: Basel, Switzerland, pp. 239–256, 2018

Mangeruga, M., Bruno, F., Cozza, M., Agrafiotis, P., Skarlatos, D. Guidelines for Underwater Image Enhancement Based on Benchmarking of Different Methods. Remote Sensing, 10, 1652, 2018

Menna, F., Agrafiotis, P., Georgopoulos, A. State-of-the-art and applications in archaeological underwater 3D recording and mapping. Journal of Cultural Heritage. 33, 231-248, 2018

Cejka, J., Žuži, M., Agrafiotis, P., Skarlatos, D., Bruno, F., and Liarokapis, F., Improving Marker-Based Tracking for Augmented Reality in Underwater Environments, EUROGRAPHICS Workshop on Graphics and Cultural Heritage, pp. 1–10, 2018

Skarlatos, D., Agrafiotis, P., Menna, F., Nocerino, E., Remondino, F. Ground control networks for underwater photogrammetry in archaeological excavations. In Proceedings of the 3rd IMEKO International Conference on Metrology for Archaeology and Cultural Heritage, MetroArcheo (pp. 23-25), 2017

Agrafiotis, P., Drakonakis, G. I., Georgopoulos, A., and Skarlatos, D.: THE

EFFECT OF UNDERWATER IMAGERY RADIOMETRY ON 3D RECONSTRUCTION AND ORTHOIMAGERY, *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-2/W3, 25-31, <https://doi.org/10.5194/isprs-archives-XLII-2-W3-25-2017>, 2017.

Liarokapis, F., Kouřil, P., Agrafiotis, P., Demesticha, S., Chmelík, J., Skarlatos, D. 3D Modelling and Mapping For Virtual Exploration of Underwater Archaeology Assets. *Proc. of the International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 425-431, 2017

Bruno, F., Lagudi, A., Ritacco, G., Agrafiotis, P., Skarlatos, D., Čejka, J., ... Mudur, S. Development and integration of digital technologies addressed to raise awareness and access to European underwater cultural heritage. An overview of the H2020 i-MARECULTURE project. In *OCEANS 2017-Aberdeen* (pp. 1-10). IEEE, 2017

Skarlatos, D., Agrafiotis, P., Balogh, T., Bruno, F., Castro, F., Petriaggi, B. D., ... Kikillos, F. Project iMARECULTURE: advanced VR, iMmersive serious games and augmented REality as tools to raise awareness and access to European underwater CULTURal heritagE. In *Euro-Mediterranean Conference* (pp. 805-813). Springer, Cham, 2016

Bibliography

- [Agrafiotis 2015] P Agrafiotis and A Georgopoulos. *Camera Constant in the Case of Two Media Photogrammetry*. ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, pages 1–6, 2015. 4, 25, 26, 27, 41, 45, 53, 69, 124, 186
- [Agrafiotis 2017] P Agrafiotis, GI Drakonakis, A Georgopoulos and D Skarlatos. *The effect of underwater imagery radiometry on 3d reconstruction and orthoimagery*. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 42, no. 2/W3, 2017. 3, 18, 175, 187
- [Agrafiotis 2018a] P. Agrafiotis, D. Skarlatos, T. Forbes, C. Poullis, M. Skamantzari and A. Georgopoulos. *Underwater photogrammetry in very shallow waters: main challenges and caustics effect removal*. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. XLII-2, pages 15–22, 2018. 3, 4, 5, 6, 18, 19, 20, 22, 32, 186
- [Agrafiotis 2018b] Panagiotis Agrafiotis, Georgios I Drakonakis, Dimitrios Skarlatos and Andreas Georgopoulos. *Underwater Image Enhancement before Three-Dimensional (3D) Reconstruction and Orthoimage Production Steps: Is It Worth*. Latest Developments in Reality-Based 3D Surveying and Modelling; Remondino, F., Georgopoulos, A., González-Aguilera, D., Agrafiotis, P., Eds, 2018. 3, 4, 5, 18, 175, 187
- [Agrafiotis 2019a] P. Agrafiotis, D. Skarlatos, A. Georgopoulos and K. Karantzalos. *Shallow water bathymetry mapping from uav imagery based on machine learning*. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. XLII-2/W10, pages 9–16, 2019. xix, 17, 18, 33, 34, 81, 186
- [Agrafiotis 2019b] Panagiotis Agrafiotis, Dimitrios Skarlatos, Andreas Georgopoulos and Konstantinos Karantzalos. *DepthLearn: Learning to Correct the Refraction on Point Clouds Derived from Aerial Imagery for Accurate Dense Shallow Water Bathymetry Based on SVMs-Fusion with LiDAR Point Clouds*. Remote Sensing, vol. 11, no. 19, page 2225, 2019. xix, 18, 33, 34, 186
- [Agrafiotis 2020] Panagiotis Agrafiotis, Konstantinos Karantzalos, Andreas Georgopoulos and Dimitrios Skarlatos. *Correcting image refraction: Towards accurate aerial image-based bathymetry mapping in shallow waters*. Remote Sensing, vol. 12, no. 2, page 322, 2020. 34, 41, 186
- [Allouis 2010] Tristan Allouis, Jean-Stéphane Bailly, Yves Pastol and Catherine Le Roux. *Comparison of LiDAR waveform processing methods for very shallow water bathymetry using Raman, near-infrared and green signals*. Earth

- Surface Processes and Landforms: The Journal of the British Geomorphological Research Group, vol. 35, no. 6, pages 640–650, 2010. 117
- [Altman 1992] Naomi S Altman. *An introduction to kernel and nearest-neighbor nonparametric regression*. The American Statistician, vol. 46, no. 3, pages 175–185, 1992. 168
- [Alvarez 2018] Laura V Alvarez, Hernan A Moreno, Antonio R Segales, Tri G Pham, Elizabeth A Pillar-Little and Phillip B Chilson. *Merging unmanned aerial systems (UAS) imagery and echo soundings with an adaptive sampling technique for bathymetric surveys*. Remote Sensing, vol. 10, no. 9, page 1362, 2018. 117
- [Audebert 2018] Nicolas Audebert, Bertrand Le Saux and Sébastien Lefèvre. *Beyond RGB: Very high resolution urban remote sensing with multimodal deep networks*. ISPRS Journal of Photogrammetry and Remote Sensing, vol. 140, pages 20–32, 2018. 151, 152, 162
- [Awad 2015] Mariette Awad and Rahul Khanna. *Support vector regression*. In Efficient Learning Machines, pages 67–80. Springer, 2015. 37, 38
- [Badrinarayanan 2017] Vijay Badrinarayanan, Alex Kendall and Roberto Cipolla. *Segnet: A deep convolutional encoder-decoder architecture for image segmentation*. IEEE transactions on pattern analysis and machine intelligence, vol. 39, no. 12, pages 2481–2495, 2017. xxiv, 150, 151, 152
- [Bailly 2010] Jean-Stéphane Bailly, Yann Le Coarer, Pascal Languille, Carl-Johann Stigermark and Tristan Allouis. *Geostatistical estimations of bathymetric LiDAR errors on rivers*. Earth Surface Processes and Landforms, vol. 35, no. 10, pages 1199–1210, 2010. 86, 87
- [Barbosa 2018] Igor Barros Barbosa, Marco Cristani, Barbara Caputo, Aleksander Rognhaugen and Theoharis Theoharis. *Looking beyond appearances: Synthetic training data for deep cnns in re-identification*. Computer Vision and Image Understanding, vol. 167, pages 50–62, 2018. 112
- [Bianco 2018] Simone Bianco, Gianluigi Ciocca and Davide Marelli. *Evaluating the performance of structure from motion pipelines*. Journal of Imaging, vol. 4, no. 8, page 98, 2018. 126
- [Bishop 2006] Christopher M Bishop. Pattern recognition and machine learning. springer, 2006. 36, 37, 38
- [Bouguet] Jean-Yves Bouguet. *Camera Calibration and 3D Reconstruction*. http://www.vision.caltech.edu/bouguetj/calib_doc/, Last accessed on 2020-05-30. 161
- [Breiman 2001] Leo Breiman. *Random forests*. Machine learning, vol. 45, no. 1, pages 5–32, 2001. 168

- [Brown 1972] Duane C Brown. *Calibration of close-range cameras*. In XII Congress of the Intern. Society of Photog., Ottawa, Canada, 1972, 1972. 44, 139
- [Butler 2002] Justin Butler, Stuart Lane, Jim Chandler and Ekaterini Porfiri. *Through-water close range digital photogrammetry in flume and field environments*. The Photogrammetric Record, vol. 17, no. 99, pages 419–439, 2002. 25, 26
- [Byrne 1977] PM Byrne and FR Honey. *Air survey and satellite imagery tools for shallow water bathymetry*. In Proceedings of the 20th Australian Survey Congress, Shrewsbury, UK, pages 103–119, 1977. 26
- [Caballero 2019] Isabel Caballero, Richard P Stumpf and Andrew Meredith. *Preliminary assessment of turbidity and chlorophyll impact on bathymetry derived from Sentinel-2A and Sentinel-3A satellites in South Florida*. Remote Sensing, vol. 11, no. 6, page 645, 2019. 29
- [Cabezas 2014] Randi Cabezas, Oren Freifeld, Guy Rosman and John W Fisher. *Aerial reconstructions via probabilistic data fusion*. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 4010–4017, 2014. 117
- [Canny 1987] JF Canny. *A computational approach to edge detection, Readings in computer vision: issues, problems, principles, and paradigms*. IEEE Trans. Pattern, 1987. 160
- [Cao 2019] Bincai Cao, Yong Fang, Zhengzhi Jiang, Li Gao and Haiyan Hu. *Shallow water bathymetry from WorldView-2 stereo imagery using two-media photogrammetry*. European Journal of Remote Sensing, vol. 52, no. 1, pages 506–521, 2019. 25, 28
- [Cao 2020] Bin Cao, Ruru Deng and Shulong Zhu. *Universal algorithm for water depth refraction correction in through-water stereo remote sensing*. International Journal of Applied Earth Observation and Geoinformation, vol. 91, page 102108, 2020. 25, 28
- [Cheng 2014] Liang Cheng, Lei Ma, Wenting Cai, Lihua Tong, Manchun Li and Peijun Du. *Integration of hyperspectral imagery and sparse sonar data for shallow water bathymetry mapping*. IEEE Transactions on Geoscience and Remote Sensing, vol. 53, no. 6, pages 3235–3249, 2014. 117
- [Chirayath 2019] Ved Chirayath and Alan Li. *Next-Generation Optical Sensing Technologies for Exploring Ocean Worlds-NASA FluidCam, MiDAR, and NeMO-Net*. Frontiers in Marine Science, vol. 6, page 521, 2019. xxi, 27, 69, 70, 71, 117

- [Cialdi 1865] Alessandro Cialdi and Pietro A Secchi. *Sur la transparence de la mer*. Comptes rendus hebdomadaire de séances de l'Academie des Sciences, vol. 61, pages 100–104, 1865. 12
- [Clark 2017] Malcolm R Clark, Helen L Rouse, Geoffroy Lamarche, Joanne Ellis and Christopher Wayne Hickey. Preparation of environmental impact assessments: General guidelines for offshore mining and drilling with particular reference to new zealand. National Institute of Water and Atmospheric Research, 2017. 3
- [Coleman 2011] John B Coleman, Xiaobai Yao, Thomas R Jordan and Marguertie Madden. *Holes in the ocean: Filling voids in bathymetric lidar data*. Computers & geosciences, vol. 37, no. 4, pages 474–484, 2011. 117
- [Collins 1979] MB Collins and FT Banner. *Secchi disc depths, suspensions and circulation, northeastern Mediterranean Sea*. Marine Geology, vol. 31, no. 1-2, pages M39–M46, 1979. 81, 148
- [Collins 2020] Adam Collins, Katherine L Brodie, Spicer Bak, Tyler Hesser, Matthew Farthing, Douglas W Gamble and Joseph W Long. *A 2D fully convolutional neural network for nearshore and surfzone bathymetry inversion from synthetic imagery of the surfzone using the wave model Celeris*. In Ocean Sciences Meeting 2020. AGU, 2020. 30
- [Consoli 2020] Pierpaolo Consoli, Gianfranco Scotti, Teresa Romeo, Maria Cristina Fossi, Valentina Esposito, Michela D'Alessandro, Pietro Battaglia, Francois Galgani, Fabio Figurella, Hannah Pragnell-Raaschet *al*. *Characterization of seafloor litter on Mediterranean shallow coastal waters: Evidence from Dive Against Debris®*, a citizen science monitoring approach. Marine Pollution Bulletin, vol. 150, page 110763, 2020. 2
- [Deng 2009] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li and Li Fei-Fei. *Imagenet: A large-scale hierarchical image database*. In 2009 IEEE conference on computer vision and pattern recognition, pages 248–255. Ieee, 2009. 162, 178
- [Dietrich 2017] James T Dietrich. *Bathymetric structure-from-motion: extracting shallow stream bathymetry from multi-view stereo photogrammetry*. Earth Surface Processes and Landforms, vol. 42, no. 2, pages 355–364, 2017. 25, 27, 28, 140, 141, 143, 144, 145
- [Dietrich 2019] James Thomas Dietrich. *pyBathySfM v4.0*, 2019. <https://github.com/geojames/pyBathySfM>, Last accessed on 2020-05-30. 140
- [Eakins 2010] BW Eakins and GF Sharman. *Volumes of the World's Oceans from ETOPO1*. NOAA National Geophysical Data Center, Boulder, CO, vol. 7, 2010. 1

- [Elfick 1984] Michael H Elfick and John G Fryer. *Mapping in shallow water*. International Archives of Photogrammetry and Remote Sensing, vol. 25, no. A5, pages 240–247, 1984. 26
- [Ellis 2017] JI Ellis, MR Clark, HL Rouse and Geoffroy Lamarche. *Environmental management frameworks for offshore mining: the New Zealand approach*. Marine Policy, vol. 84, pages 178–192, 2017. 3
- [Elnashef 2019] Bashar Elnashef and Sagi Filin. *Direct linear and refraction-invariant pose estimation and calibration model for underwater imaging*. ISPRS Journal of Photogrammetry and Remote Sensing, vol. 154, pages 259–271, 2019. 4
- [Entwistle 2019] Neil S Entwistle and George L Heritage. *Small unmanned aerial model accuracy for photogrammetrical fluvial bathymetric survey*. Journal of Applied Remote Sensing, vol. 13, no. 1, page 014523, 2019. 30
- [European Commission 2009] European Commission. *The economics of climate change adaptation in EU coastal areas*, 2009. https://ec.europa.eu/maritimeaffairs/sites/maritimeaffairs/files/docs/body/report_en.pdf, Last accessed on 2020-05-30. 148
- [Fahy 2003] Frank Fahy and John Walker. *Fundamentals of noise and vibration*. CRC Press, 2003. 12
- [Fattal 2008] Raanan Fattal. *Single image dehazing*. ACM transactions on graphics (TOG), vol. 27, no. 3, pages 1–9, 2008. 32
- [Fernandez-Diaz 2013] Juan Carlos Fernandez-Diaz, Craig L Glennie, William E Carter, Ramesh L Shrestha, Michael P Sartori, Abhinav Singhania, Carl J Legleiter and Brandon T Overstreet. *Early results of simultaneous terrain and shallow water bathymetry mapping using a single-wavelength airborne LiDAR sensor*. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 7, no. 2, pages 623–635, 2013. 86, 87
- [Ferreira 2006] Ricardo Ferreira, João P Costeira, Carlos Silvestre, I Sousa and JA Santos. *Using stereo image reconstruction to survey scale models of rubble-mound structures*. In Proceedings of the First International Conference on the Application of Physical Modelling to Port and Coastal Protection, Porto, Portugal, pages 8–10. Citeseer, 2006. 26
- [Fischler 1981] Martin A Fischler and Robert C Bolles. *Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography*. Communications of the ACM, vol. 24, no. 6, pages 381–395, 1981. 21, 154
- [Forbes 2018] Timothy Forbes, Mark Goldsmith, Sudhir Mudur and Charalambos Poullis. *DeepCaustics: classification and removal of caustics from underwater*

- imagery*. IEEE Journal of Oceanic Engineering, vol. 44, no. 3, pages 728–738, 2018. 6, 32, 149, 170
- [Freund 1995] Yoav Freund and Robert E Schapire. *A decision-theoretic generalization of on-line learning and an application to boosting*. In European conference on computational learning theory, pages 23–37. Springer, 1995. 168
- [Fryer 1983] John G Fryer. *Photogrammetry through shallow water*. Australian Journal of Geodesy, Photogrammetry and Surveying, vol. 38, pages 25–38, 1983. 25
- [Fryer 1985] John G Fryer and HT Kniest. *Errors in depth determination caused by waves in through-water photogrammetry*. The Photogrammetric Record, vol. 11, no. 66, pages 745–753, 1985. xix, 15, 16, 26, 81
- [Fryer 1986] John G Fryer and Duane C Brown. *Lens distortion for close-range photogrammetry*. Photogrammetric engineering and remote sensing, vol. 52, no. 1, pages 51–58, 1986. 44
- [Furlani 2014] Stefano Furlani, Marta Pappalardo, Ll Gómez-Pujol and A Chelli. *The rock coast of the Mediterranean and Black seas*. Geological Society, London, Memoirs, vol. 40, no. 1, pages 89–123, 2014. 148
- [Georgopoulos 2012] Andreas Georgopoulos and Panagiotis Agrafiotis. *Documentation of a submerged monument using improved two media techniques*. In 2012 18th International Conference on Virtual Systems and Multimedia, pages 173–180. IEEE, 2012. 25, 28, 81, 124, 186
- [Girardeau-Montaut 2015] Daniel Girardeau-Montaut. *CloudCompare (Version 2.11 Alpha) [GPL Software]*, 2015. <http://www.cloudcompare.org/>, Last accessed on 2020-05-30. 83, 85, 103, 126, 175, 176
- [Gonçalves 2020] Gil Gonçalves, Umberto Andriolo, Luís Pinto and Filipa Bessa. *Mapping marine litter using UAS on a beach-dune system: a multidisciplinary approach*. Science of The Total Environment, vol. 706, page 135742, 2020. 2
- [Goshtasby 1986] Ardeshir Goshtasby. *Piecewise linear mapping functions for image registration*. Pattern Recognition, vol. 19, no. 6, pages 459 – 466, 1986. 47
- [Goshtasby 2005] A. Ardeshir Goshtasby. *2-d and 3-d image registration: For medical, remote sensing, and industrial applications*. Wiley-Interscience, USA, 2005. 47
- [Gracias 2000] Nuno Gracias and José Santos-Victor. *Underwater video mosaics as visual navigation maps*. Computer Vision and Image Understanding, vol. 79, no. 1, pages 66–91, 2000. 31, 149

- [Gracias 2008] Nuno Gracias, Shahriar Negahdaripour, Laszlo Neumann, Ricard Prados and Rafael Garcia. *A motion compensated filtering approach to remove sunlight flicker in shallow water images*. In OCEANS 2008, pages 1–7. IEEE, 2008. 31, 149
- [Green 1978] P. J. Green and R. Sibson. *Computing Dirichlet Tessellations in the Plane*. The Computer Journal, vol. 21, no. 2, pages 168–173, 05 1978. 46
- [Green 2000] Edmund Green, Peter Mumby, Alasdair Edwards and Christopher Clark. *Remote sensing: Handbook for tropical coastal management*. United Nations Educational, Scientific and Cultural Organization (UNESCO), 2000. 5
- [Guenther 2000] Gary C Guenther, A Grant Cunningham, Paul E LaRocque and David J Reid. *Meeting the accuracy challenge in airborne bathymetry*. Report technique, NATIONAL OCEANIC ATMOSPHERIC ADMINISTRATION/NESDIS SILVER SPRING MD, 2000. xx, xx, xxi, xxi, xxii, xxii, xxiii, xxiv, 36, 61, 62, 63, 64, 67, 69, 70, 73, 74, 99, 104, 105, 106, 108, 121, 126, 127, 128, 129, 139, 142, 143, 145, 146
- [Harris 1972] William D Harris, Harris WD and Umbach MJ. *Underwater mapping*. Photogrammetric Engineering, 1972. 26
- [Hartley 1994] Richard I Hartley. *An algorithm for self calibration from several views*. In Cvpr, volume 94, pages 908–912. Citeseer, 1994. 161
- [Harwin 2015] Steve Harwin, Arko Lucieer and Jon Osborn. *The impact of the calibration method on the accuracy of point clouds derived using unmanned aerial vehicle multi-view stereopsis*. Remote Sensing, vol. 7, no. 9, pages 11933–11953, 2015. 81
- [Hasegawa 2000] Hiroyuki Hasegawa, Kaoru Matsuo, Mamoru Koarai, Nobuyuki Watanabe, Hiroshi Masaharu and Yoshikazu Fukushima. *DEM accuracy and the base to height (B/H) ratio of stereo images*. International Archives of Photogrammetry and Remote Sensing, vol. 33, no. B4/1; PART 4, pages 356–359, 2000. 83
- [Hastie 2009] Trevor Hastie, Saharon Rosset, Ji Zhu and Hui Zou. *Multi-class adaboost*. Statistics and its Interface, vol. 2, no. 3, pages 349–360, 2009. 168
- [He 2010] Kaiming He, Jian Sun and Xiaoou Tang. *Single image haze removal using dark channel prior*. IEEE transactions on pattern analysis and machine intelligence, vol. 33, no. 12, pages 2341–2353, 2010. 32
- [He 2015] Kaiming He, Xiangyu Zhang, Shaoqing Ren and Jian Sun. *Delving deep into rectifiers: Surpassing human-level performance on imagenet classification*. In Proceedings of the IEEE international conference on computer vision, pages 1026–1034, 2015. 162

- [Hein 2009] James R Hein, Tracey A Conrad and Rachel E Dunham. *Seamount characteristics and mine-site model applied to exploration-and mining-lease-block selection for cobalt-rich ferromanganese crusts*. Marine Georesources and Geotechnology, vol. 27, no. 2, pages 160–176, 2009. 3
- [Hell 2012] Benjamin Hell, Barry Broman, Lars Jakobsson, Martin Jakobsson, Åke Magnusson and Patrik Wiberg. *The use of bathymetric data in society and science: A review from the Baltic Sea*. Ambio, vol. 41, no. 2, pages 138–150, 2012. 2
- [Hirschmuller 2005] Heiko Hirschmuller. *Accurate and efficient stereo processing by semi-global matching and mutual information*. In 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05), volume 2, pages 807–814. IEEE, 2005. 154
- [Hirschmuller 2007] Heiko Hirschmuller. *Stereo processing by semiglobal matching and mutual information*. IEEE Transactions on pattern analysis and machine intelligence, vol. 30, no. 2, pages 328–341, 2007. 154, 155
- [Hodúl 2018] Matúš Hodúl, Stephen Bird, Anders Knudby and René Chénier. *Satellite derived photogrammetric bathymetry*. ISPRS Journal of Photogrammetry and Remote Sensing, vol. 142, pages 268–277, 2018. 28
- [Holman 2013] Rob Holman, Nathaniel Plant and Todd Holland. *cBathy: A robust algorithm for estimating nearshore bathymetry*. Journal of geophysical research: Oceans, vol. 118, no. 5, pages 2595–2609, 2013. 30
- [IHO 2019] IHO. *Zones of Confidence*, 2019. <https://www.admiralty.co.uk/>, Last accessed on 2020-05-30. 73, 74, 121, 128, 139, 145
- [Ioffe 2015] Sergey Ioffe and Christian Szegedy. *Batch normalization: Accelerating deep network training by reducing internal covariate shift*. arXiv preprint arXiv:1502.03167, 2015. 151
- [Jakobsson 2003] Martin Jakobsson, Larry Mayer and Andrew Armstrong. *Analysis of data relevant to establishing outer limits of a continental shelf under law of the sea article 76*. The International hydrographic review, vol. 4, no. 1, pages 36–52, 2003. 2
- [Joshi 2010] N Joshi, M Cohen and SM Rainier. *Lucky imaging for multi-image denoising, sharpening, and haze removal, Computational Photography (ICCP)*. In 2010 IEEE International Conference on, Cambridge, MA, USA, pages 1–8, 2010. 32
- [Karara 1979] Houssam M Karara. *Handbook of non-topographic photogrammetry*. Rapport technique, 1979. 4, 25

- [Kasvi 2019] E Kasvi, J Salmela, E Lotsari, T Kumpula and SN Lane. *Comparison of remote sensing based approaches for mapping bathymetry of shallow, clear water rivers*. *Geomorphology*, vol. 333, pages 180–197, 2019. 4, 28
- [Ke 2008] X Ke, MA Sutton, Susan M Lessner and M Yost. *Robust stereo vision and calibration methodology for accurate three-dimensional digital image correlation measurements on submerged objects*. *The Journal of Strain Analysis for Engineering Design*, vol. 43, no. 8, pages 689–704, 2008. 26
- [Keerthi 2001] S. Sathiya Keerthi, Shirish Krishnaj Shevade, Chiranjib Bhattacharyya and Karuturi Radha Krishna Murthy. *Improvements to Platt’s SMO algorithm for SVM classifier design*. *Neural computation*, vol. 13, no. 3, pages 637–649, 2001. 39
- [Kim 2019] Jun Song Kim, Donghae Baek, Il Won Seo and Jaehyun Shin. *Retrieving shallow stream bathymetry from UAV-assisted RGB imagery using a geospatial regression method*. *Geomorphology*, vol. 341, pages 102–114, 2019. 30
- [Lague 2013] Dimitri Lague, Nicolas Brodu and Jérôme Leroux. *Accurate 3D comparison of complex topography with terrestrial laser scanner: Application to the Rangitikei canyon (NZ)*. *ISPRS journal of photogrammetry and remote sensing*, vol. 82, pages 10–26, 2013. 83, 85, 103, 126, 133
- [Lavest 2000] Jean-Marc Lavest, Gérard Rives and Jean-Thierry Lapresté. *Underwater camera calibration*. In *European Conference on Computer Vision*, pages 654–668. Springer, 2000. 4, 27, 69
- [Legleiter 2012] Carl J Legleiter. *Remote measurement of river morphology via fusion of LiDAR topography and spectrally based bathymetry*. *Earth Surface Processes and Landforms*, vol. 37, no. 5, pages 499–518, 2012. 117
- [Legleiter 2019] Carl J Legleiter and Ryan L Fosness. *Defining the limits of spectrally based bathymetric mapping on a large river*. *Remote Sensing*, vol. 11, no. 6, page 665, 2019. 6, 30, 101
- [Legleiter 2020] Carl J Legleiter and Paul J Kinzel. *Inferring Surface Flow Velocities in Sediment-Laden Alaskan Rivers from Optical Image Sequences Acquired from a Helicopter*. *Remote Sensing*, vol. 12, no. 8, page 1282, 2020. 30
- [Leica Geosystems 2015] Leica Geosystems. *Leica HawkEye III*, 2015. https://leica-geosystems.com/-/media/files/leicageosystems/products/datasheets/leica_hawkeye_iii_ds.ashx?la=en, Last accessed on 2020-05-30. 86
- [Leon 2013] Javier X Leon, Stuart R Phinn, Sarah Hamylton and Megan I Saunders. *Filling the ‘white ribbon’—a multisource seamless digital elevation model for*

- Lizard Island, northern Great Barrier Reef*. International journal of remote sensing, vol. 34, no. 18, pages 6337–6354, 2013. 117
- [Leutenegger 2011] Stefan Leutenegger, Margarita Chli and Roland Y Siegwart. *BRISK: Binary robust invariant scalable keypoints*. In 2011 International conference on computer vision, pages 2548–2555. Ieee, 2011. 153
- [Loop 1999] Charles Loop and Zhengyou Zhang. *Computing rectifying homographies for stereo vision*. In Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149), volume 1, pages 125–131. IEEE, 1999. 154
- [Lowe 1999] David G Lowe. *Object recognition from local scale-invariant features*. In Proceedings of the seventh IEEE international conference on computer vision, volume 2, pages 1150–1157. Ieee, 1999. xix, 20, 21
- [Lynch 2001] David K Lynch, William Charles Livingston and William Livingston. *Color and light in nature*. Cambridge University Press, 2001. 19
- [Ma 2019] Zhenling Ma, Xiao Xu, Yannan Chen and Weijie Wang. *Bathymetric Extraction using Overlapping Orthoimages*. In IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium, pages 7948–7951. IEEE, 2019. 30
- [Maas 2015] Hans-Gerd Maas. *On the accuracy potential in underwater/multimedia photogrammetry*. Sensors, vol. 15, no. 8, pages 18140–18152, 2015. 26, 27
- [Makowski 2016] Christopher Makowski and Charles W Finkl. *History of modern seafloor mapping*. In Seafloor Mapping along Continental Shelves, pages 3–49. Springer, 2016. 3, 12
- [Mandlbürger 2018] G Mandlbürger. *A CASE STUDY ON THROUGH-WATER DENSE IMAGE MATCHING*. International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences, vol. 42, no. 2, 2018. 25, 28
- [Mandlbürger 2019] Gottfried Mandlbürger. *Through-Water Dense Image Matching for Shallow Water Bathymetry*. Photogrammetric Engineering & Remote Sensing, vol. 85, no. 6, pages 445–455, 2019. 25, 28
- [Mangeruga 2018] Marino Mangeruga, Fabio Bruno, Marco Cozza, Panagiotis Agrafiotis and Dimitrios Skarlatos. *Guidelines for underwater image enhancement based on benchmarking of different methods*. Remote Sensing, vol. 10, no. 10, page 1652, 2018. 3, 4, 18, 90, 175, 177
- [Masry 1975] SE Masry. *Measurement of water depth by the analytical plotter*. The International Hydrographic Review, 1975. 26

- [Matthies 1989] Larry Matthies, Takeo Kanade and Richard Szeliski. *Kalman filter-based algorithms for estimating depth from image sequences*. International Journal of Computer Vision, vol. 3, no. 3, pages 209–238, 1989. 85
- [Maybank 1992] Stephen J Maybank and Olivier D Faugeras. *A theory of self-calibration of a moving camera*. International journal of computer vision, vol. 8, no. 2, pages 123–151, 1992. 160
- [Medwin 1997] Herman Medwin and Clarence S Clay. Fundamentals of acoustical oceanography. Academic press, 1997. 13
- [Menna 2013] Fabio Menna, Erica Nocerino, Salvatore Troisi and Fabio Remondino. *A photogrammetric approach to survey floating and semi-submerged objects*. In Videometrics, Range Imaging, and Applications XII; and Automated Visual Inspection, volume 8791, page 87910H. International Society for Optics and Photonics, 2013. 4
- [Menna 2018] Fabio Menna, Panagiotis Agrafiotis and Andreas Georgopoulos. *State of the art and applications in archaeological underwater 3D recording and mapping*. Journal of Cultural Heritage, vol. 33, pages 231–248, 2018. 2, 3, 4, 5, 11, 18, 86, 87, 159, 187
- [Misra 2018] Ankita Misra, Zoran Vojinovic, Balaji Ramakrishnan, Arjen Lujendijk and Roshanka Ranasinghe. *Shallow water bathymetry mapping using Support Vector Machine (SVM) technique and multispectral imagery*. International journal of remote sensing, vol. 39, no. 13, pages 4431–4450, 2018. 29, 40
- [Mohamed 2016] Hassan Mohamed, Abdelazim Negm, Mohamed Zahran and Oliver C Saavedra. *Bathymetry determination from high resolution satellite imagery using ensemble learning algorithms in Shallow Lakes: Case study El-Burullus Lake*. International Journal of Environmental Science and Development, vol. 7, no. 4, page 295, 2016. 29
- [Mullen 2004] Roy Mullen. Manual of photogrammetry. Asprs Publications, 2004. 45
- [Mulsow 2010] Christian Mulsow. *A flexible multi-media bundle approach*. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci, vol. 38, pages 472–477, 2010. 25, 26
- [Murase 2008] Toshimi Murase, Miho Tanaka, Tomomi Tani, Yuko Miyashita, Naoto Ohkawa, Satoshi Ishiguro, Yasuhiro Suzuki, Hajime Kayanne and Hiroya Yamano. *A photogrammetric correction procedure for light refraction effects at a two-medium boundary*. Photogrammetric engineering & remote sensing, vol. 74, no. 9, pages 1129–1136, 2008. 28

- [Nesbit 2019] Paul Ryan Nesbit and Christopher H Hugenholtz. *Enhancing uav-sfm 3d model accuracy in high-relief landscapes by incorporating oblique images*. Remote Sensing, vol. 11, no. 3, page 239, 2019. 56, 58, 73, 74
- [Niroumand-Jadidi 2019] Milad Niroumand-Jadidi, Nima Pahlevan and Alfonso Vitti. *Mapping substrate types and compositions in shallow streams*. Remote Sensing, vol. 11, no. 3, page 262, 2019. 29
- [Okamoto 1982] Atsushi Okamoto *et al.* *Wave influences in two-media photogrammetry*. Photogrammetric Engineering and Remote Sensing, vol. 48, no. 9, pages 1487–1499, 1982. 81
- [Okutomi 1993] Masatoshi Okutomi and Takeo Kanade. *A multiple-baseline stereo*. IEEE Transactions on pattern analysis and machine intelligence, vol. 15, no. 4, pages 353–363, 1993. 85
- [OpenCV] OpenCV. *Camera Calibration and 3D Reconstruction*. https://docs.opencv.org/2.4/modules/calib3d/doc/camera_calibration_and_3d_reconstruction.html, Last accessed on 2020-05-30. 161
- [Papadaki 2020] Alexandra I Papadaki and Ronny Hansch. *Match or No Match: Keypoint Filtering Based on Matching Probability*. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pages 1014–1015, 2020. 185
- [Partama 2018] I GD Yudha Partama, Ariyo Kanno, Motoyasu Ueda, Yoshihisa Akamatsu, Ryutei Inui, Masahiko Sekine, Koichi Yamamoto, Tsuyoshi Imai and Takaya Higuchi. *Removal of water-surface reflection effects with a temporal minimum filter for UAV-based shallow-water photogrammetry*. Earth Surface Processes and Landforms, vol. 43, no. 12, pages 2673–2682, 2018. 26
- [Patki 2016] Neha Patki, Roy Wedge and Kalyan Veeramachaneni. *The synthetic data vault*. In 2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA), pages 399–410. IEEE, 2016. 106
- [Pedregosa 2011] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourget *et al.* *Scikit-learn: Machine learning in Python*. the Journal of machine Learning research, vol. 12, pages 2825–2830, 2011. 65, 168
- [Peng 2015] Xingchao Peng, Baochen Sun, Karim Ali and Kate Saenko. *Learning deep object detectors from 3d models*. In Proceedings of the IEEE International Conference on Computer Vision, pages 1278–1286, 2015. 112
- [Pérez 2003] Patrick Pérez, Michel Gangnet and Andrew Blake. *Poisson image editing*. In ACM SIGGRAPH 2003 Papers, pages 313–318. 2003. 171

- [Platt 1999] John C Platt. *Using analytic QP and sparseness to speed training of support vector machines*. In Advances in neural information processing systems, pages 557–563, 1999. 38, 39
- [Qian 2018] Yiming Qian, Yinqiang Zheng, Minglun Gong and Yee-Hong Yang. *Simultaneous 3d reconstruction for water surface and underwater scene*. In Proceedings of the European Conference on Computer Vision (ECCV), pages 754–770, 2018. 28
- [Quan 1995] Xiaohong Quan and Edward S. Fry. *Empirical equation for the index of refraction of seawater*. Appl. Opt., vol. 34, no. 18, pages 3477–3480, Jun 1995. 45
- [Reinhard 2001] Erik Reinhard, Michael Adhikhmin, Bruce Gooch and Peter Shirley. *Color transfer between images*. IEEE Computer graphics and applications, vol. 21, no. 5, pages 34–41, 2001. 153
- [Remondino 2013] Fabio Remondino, Maria Grazia Spera, Erica Nocerino, Fabio Menna, Francesco Nex and Sara Gonizzi-Barsanti. *Dense image matching: comparisons and analyses*. In 2013 Digital Heritage International Congress (DigitalHeritage), volume 1, pages 47–54. IEEE, 2013. 85
- [Richardson 2016] Elad Richardson, Matan Sela and Ron Kimmel. *3D face reconstruction by learning from synthetic data*. In 2016 Fourth International Conference on 3D Vision (3DV), pages 460–469. IEEE, 2016. 112
- [Ruderman 1998] Daniel L Ruderman, Thomas W Cronin and Chuan-Chin Chiao. *Statistics of cone responses to natural images: implications for visual coding*. JOSA A, vol. 15, no. 8, pages 2036–2045, 1998. 153
- [Rusu 2011] Radu Bogdan Rusu and Steve Cousins. *3d is here: Point cloud library (pcl)*. In 2011 IEEE international conference on robotics and automation, pages 1–4. IEEE, 2011. 87
- [Safavian 1991] S Rasoul Safavian and David Landgrebe. *A survey of decision tree classifier methodology*. IEEE transactions on systems, man, and cybernetics, vol. 21, no. 3, pages 660–674, 1991. 168
- [Saleh 2018] Fatemeh Sadat Saleh, Mohammad Sadegh Aliakbarian, Mathieu Salzmann, Lars Petersson and Jose M Alvarez. *Effective use of synthetic data for urban scene semantic segmentation*. In European Conference on Computer Vision, pages 86–103. Springer, 2018. 112
- [Sankaranarayanan 2018] Swami Sankaranarayanan, Yogesh Balaji, Arpit Jain, Ser Nam Lim and Rama Chellappa. *Learning from synthetic data: Addressing domain shift for semantic segmentation*. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 3752–3761, 2018. 112

- [Schechner 2004] Yoav Y Schechner and Nir Karpel. *Attenuating natural flicker patterns*. In Oceans' 04 MTS/IEEE Techno-Ocean'04 (IEEE Cat. No. 04CH37600), volume 3, pages 1262–1268. IEEE, 2004. 32
- [Schölkopf 2002] Bernhard Schölkopf, Alexander J Smola, Francis Bach *et al.* *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press, 2002. 168
- [Schwarz 2019] Roland Schwarz, Gottfried Mandlbürger, Martin Pfennigbauer and Norbert Pfeifer. *Design and evaluation of a full-wave surface and bottom-detection algorithm for LiDAR bathymetry of very shallow waters*. ISPRS journal of photogrammetry and remote sensing, vol. 150, pages 1–10, 2019. 117
- [Shan 1994] Jie Shan. *Relative orientation for two-media photogrammetry*. The Photogrammetric Record, vol. 14, no. 84, pages 993–999, 1994. xix, 14, 15, 25, 27
- [Shihavuddin 2012] ASM Shihavuddin, Nuno Gracias and Rafael Garcia. *Online Sunflicker Removal using Dynamic Texture Prediction*. In VISAPP (1), pages 161–167, 2012. 31
- [Shintani 2017] Christina Shintani and Mark A Fonstad. *Comparing remote-sensing techniques collecting bathymetric data from a gravel-bed river*. International journal of remote sensing, vol. 38, no. 8-10, pages 2883–2902, 2017. 29
- [Shmutter 1967] B Shmutter and L Bonfiglioli. *Orientation problem in two-medium photogrammetry*. Photogrammetric Engineering, vol. 33, no. 12, pages 1421–1428, 1967. 25
- [Shortis 2019] Mark Shortis. *Camera calibration techniques for accurate measurement underwater*. In 3D Recording and Interpretation for Maritime Archaeology, pages 11–27. Springer, 2019. 4
- [Simonyan 2014] Karen Simonyan and Andrew Zisserman. *Very deep convolutional networks for large-scale image recognition*. arXiv preprint arXiv:1409.1556, 2014. 151
- [Skarlatos 2006] Dimitrios Skarlatos and Andreas Georgopoulos. *The method of two overlapping orthoimages for checking the produced DTM*. Rapport technique, 2006. 53
- [Skarlatos 2018] Dimitrios Skarlatos and Panagiotis Agrafiotis. *A novel iterative water refraction correction algorithm for use in structure from motion photogrammetric pipeline*. Journal of Marine Science and Engineering, vol. 6, no. 3, page 77, 2018. 4, 16, 17, 25, 27, 34, 41, 42, 44, 47, 81, 86, 122, 140, 141, 143, 144, 184, 186

- [Skarlatos 2019] Dimitrios Skarlatos, Fabio Menna, Erica Nocerino and Panagiotis Agrafiotis. *Precision Potential of Underwater Networks for Archaeological Excavation Through Trilateration and Photogrammetry*. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 42, no. 2/W10, pages 175–180, 2019. 4, 5
- [Skinner 2011] Kenneth D Skinner. *Evaluation of LiDAR-acquired bathymetric and topographic data accuracy in various hydrogeomorphic settings in the Deadwood and South Fork Boise rivers, west-central Idaho, 2007*. Rapport technique, 2011. 86, 87
- [Smith 2015] Mark William Smith and Damià Vericat. *From experimental plots to experimental landscapes: topography, erosion and deposition in sub-humid badlands from structure-from-motion photogrammetry*. Earth Surface Processes and Landforms, vol. 40, no. 12, pages 1656–1671, 2015. 56, 57, 58, 73, 74
- [Smola 1996] Alex J Smola et al. *Regression estimation with support vector learning machines*. PhD thesis, Master’s thesis, Technische Universität München, 1996. 36
- [Smola 2004] Alex J Smola and Bernhard Schölkopf. *A tutorial on support vector regression*. Statistics and computing, vol. 14, no. 3, pages 199–222, 2004. 39
- [StarkEffects.com 2017] StarkEffects.com. *Snell’s Law in Vector Form*, 2017. <http://www.starkeffects.com/snells-law-vector.shtml>, Last accessed on 2020-05-30. 14
- [Stewart 2009] Iain Stewart and Christophe Morhange. Coastal geomorphology and sea-level change, pages pp. 385–414. 01 2009. 148
- [Swirski 2009] Yohay Swirski, Yoav Y Schechner, Ben Herzberg and Shahriar Negahdaripour. *Stereo from flickering caustics*. In 2009 IEEE 12th International Conference on Computer Vision, pages 205–212. IEEE, 2009. 19
- [Swirski 2010] Yohay Swirski, Yoav Y Schechner, Ben Herzberg and Shahriar Negahdaripour. *Underwater stereo using natural flickering illumination*. In OCEANS 2010 MTS/IEEE SEATTLE, pages 1–7. IEEE, 2010. 19
- [Swirski 2013] Yohay Swirski and Yoav Y Schechner. *3Deflicker from motion*. In IEEE International Conference on Computational Photography (ICCP), pages 1–9. IEEE, 2013. 157
- [Telem 2010] Gili Telem and Sagi Filin. *Photogrammetric modeling of underwater environments*. ISPRS journal of photogrammetry and remote sensing, vol. 65, no. 5, pages 433–444, 2010. 27
- [Tewinkel 1963] GC Tewinkel. *Water depths from aerial photographs*. Photogrammetric Engineering, vol. 29, no. 6, pages 1037–1042, 1963. 25

- [Thuan 2019] Duong Hai Thuan, Rafael Almar, Patrick Marchesiello and Nguyen Trung Viet. *Video sensing of nearshore bathymetry evolution with error estimate*. Journal of Marine Science and Engineering, vol. 7, no. 7, page 233, 2019. 30
- [Trabes 2015] Emanuel Trabes and Mario A Jordán. *Employing Feedback to Filter Caustic Waves in Underwater Scenes in Motion*. In Argentine Symposium on Artificial Intelligence (ASAI 2015)-JAIIO 44 (Rosario, 2015), 2015. 19, 20, 31
- [Trabes 2017] Emanuel Trabes and Mario Alberto Jordan. *A Node-Based Method for SLAM Navigation in Self-Similar Underwater Environments: A Case Study*. Robotics, vol. 6, no. 4, page 29, 2017. 5, 21
- [Traganos 2018] Dimosthenis Traganos, Dimitris Poursanidis, Bharat Aggarwal, Nektarios Chrysoulakis and Peter Reinartz. *Estimating satellite-derived bathymetry (SDB) with the google earth engine and sentinel-2*. Remote Sensing, vol. 10, no. 6, page 859, 2018. 29, 31
- [UNEP/MAP 2015] UNEP/MAP. *Assessment of the Status of Marine Litter, in the Mediterranean*. Rapport technique, 2015. 2
- [UNESCO] UNESCO. *Underwater Cultural Heritage*. <http://www.unesco.org/new/en/culture/themes/underwater-cultural-heritage/about-the-heritage/protection/threats/>, Last accessed on 2020-05-30. 2
- [van den Bergh 2017] Jarrett van den Bergh, Joey Schutz, Ved Chirayath and Alan Li. *A 3D Active Learning Application for NeMO-Net, the NASA Neural Multi-Modal Observation and Training Network for Global Coral Reef Assessment*. In AGU Fall Meeting Abstracts, 2017. 117
- [Vapnik 2013] Vladimir Vapnik. *The nature of statistical learning theory*. Springer science & business media, 2013. 37, 38, 40
- [Vapnik 2015] Vladimir N Vapnik and A Ya Chervonenkis. *On the uniform convergence of relative frequencies of events to their probabilities*. In Measures of complexity, pages 11–30. Springer, 2015. 37
- [Wang 1990] Zhizhuo Wang. *Principles of photogrammetry:(with remote sensing)*. Press of Wuhan Technical University of Surveying and Mapping, 1990. 25
- [Wang 2019] Lei Wang, Hongxing Liu, Haibin Su and Jun Wang. *Bathymetry retrieval from optical images with spatially distributed support vector machines*. GIScience & remote sensing, vol. 56, no. 3, pages 323–337, 2019. 29
- [Weinstein 1969] LA Weinstein. *“Open Resonators and Open Waveguides”*, Golem Press. 1969. 19

- [Westaway 2001] Richard M Westaway, Stuart N Lane and D Murray Hicks. *Remote sensing of clear-water, shallow, gravel-bed rivers using digital photogrammetry*. Photogrammetric Engineering and Remote Sensing, vol. 67, no. 11, pages 1271–1282, 2001. 26
- [Westfeld 2017] Patrick Westfeld, Hans-Gerd Maas, Katja Richter and Robert Weiß. *Analysis and correction of ocean wave pattern induced systematic coordinate errors in airborne LiDAR bathymetry*. ISPRS Journal of Photogrammetry and Remote Sensing, vol. 128, pages 314–325, 2017. 86, 87
- [Whittlesey 1975] Julian H Whittlesey. *Elevated and airborne photogrammetry and stereo photography*. Photography in archaeological research, pages 223–259, 1975. 26
- [Wikipedia] Wikipedia. *Caustic (optics)*. [https://en.wikipedia.org/wiki/Caustic_\(optics\)](https://en.wikipedia.org/wiki/Caustic_(optics)), Last accessed on 2020-05-30. 19, 20
- [Wimmer 2016] M Wimmer. *Comparison of active and passive optical methods for mapping river bathymetry*. TU Wien, 2016. 28
- [Wolff 2000] Kirsten Wolff and Wolfgang Förstner. *Exploiting the multi view geometry for automatic surfaces reconstruction using feature based matching in multimedia photogrammetry*. International Archives of Photogrammetry and Remote Sensing, vol. 33, no. B5/2; PART 5, pages 900–907, 2000. 15, 26
- [Wölf 2019] Anne-Cathrin Wölf, Helen Snaith, Sam Amirebrahimi, Colin W Devey, Boris Dorschel, Vicki Ferrini, Veerle AI Huvenne, Martin Jakobsson, Jennifer Jencks, Gordon Johnstone *et al.* *Seafloor Mapping—the challenge of a truly global ocean bathymetry*. Frontiers in Marine Science, vol. 6, page 283, 2019. 1, 2
- [Woodget 2015] AS Woodget, PE Carbonneau, Fleur Visser and Ian P Maddock. *Quantifying submerged fluvial topography using hyperspatial resolution UAS imagery and structure from motion photogrammetry*. Earth Surface Processes and Landforms, vol. 40, no. 1, pages 47–64, 2015. 26, 27, 69, 71, 140, 141, 143, 144, 145
- [Ylimäki 2018] Markus Ylimäki, Janne Heikkilä and Juho Kannala. *Accurate 3-D Reconstruction with RGB-D Cameras using Depth Map Fusion and Pose Refinement*. In 2018 24th International Conference on Pattern Recognition (ICPR), pages 1977–1982. IEEE, 2018. 85
- [Zeiler 2014] Matthew D Zeiler and Rob Fergus. *Visualizing and understanding convolutional networks*. In European conference on computer vision, pages 818–833. Springer, 2014. 152

- [Zenetos 2002] A Zenetos, I Siokou-Frangou, O Gotsis-Skretas and S Groom. *The Mediterranean Sea—blue oxygen-rich, nutrient-poor waters*. Europe's biodiversity, biogeographical regions and seas. Seas around Europe. EEA (European Environment Agency), National Centre for Marine Research (NCMR), Copenhagen, 2002. 148
- [Zhang 2000] Zhengyou Zhang. *A flexible new technique for camera calibration*. IEEE Transactions on pattern analysis and machine intelligence, vol. 22, no. 11, pages 1330–1334, 2000. 160, 161