



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ  
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ  
ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ  
ΕΡΓΑΣΤΗΡΙΟ ΕΠΕΞΕΡΓΑΣΙΑΣ ΕΙΚΟΝΑΣ ΒΙΝΤΕΟ ΚΑΙ ΠΟΛΥΜΕΣΩΝ

**Μέθοδοι Όρασης Υπολογιστών  
για Επαυξημένη Πραγματικότητα**

**ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ**

του

**Χρήστου Σ. Αρβανίτη**

**Επιβλέπων:** Στέφανος Κόλλιας  
Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούλιος 2012





ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ  
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ  
ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ  
ΕΡΓΑΣΤΗΡΙΟ ΕΠΕΞΕΡΓΑΣΙΑΣ ΕΙΚΟΝΑΣ ΒΙΝΤΕΟ ΚΑΙ ΠΟΛΥΜΕΣΩΝ

## Μέθοδοι Όρασης Υπολογιστών για Επαυξημένη Πραγματικότητα

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

**Χρήστου Σ. Αρβανίτη**

**Επιβλέπων:** Στέφανος Κόλλιας  
Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 20<sup>η</sup> Ιουλίου 2012.

(Υπογραφή)

.....  
Στέφανος Κόλλιας  
Καθηγητής Ε.Μ.Π.

(Υπογραφή)

.....  
Σταφυλοπάτης Ανδρέας-Γεώργιος  
Καθηγητής Ε.Μ.Π.

(Υπογραφή)

.....  
Στάμου Γεώργιος  
Λέκτορας Ε.Μ.Π.

Αθήνα, Ιούλιος 2012

.....  
**Χρήστος Αρβανίτης**

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

Copyright © Αρβανίτης Χρήστος (2012) Εθνικό Μετσόβιο Πολυτεχνείο.

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα. Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

# Ευχαριστίες

Η παρούσα διπλωματική εργασία εκπονήθηκε κατά το ακαδημαϊκό έτος 2011-2012 στο Εργαστήριο Ψηφιακής Επεξεργασίας Εικόνας, Βίντεο και Πολυμέσων του Εθνικού Μετσόβιου Πολυτεχνείου. Θα ήθελα να ευχαριστήσω τον επιβλέποντα Καθηγητή κ. Στέφανο Κόλλια για την εμπιστοσύνη που μου έδειξε αναθέτοντάς μου την εργασία αυτή και για τη δυνατότητα που μου έδωσε να ασχοληθώ με το συγκεκριμένο ενδιαφέρον θέμα. Επίσης, ευχαριστώ ιδιαίτερα τους Δρ. Ιωάννη Αβρίθη και Δρ. Κώστα Ραπαντζίκο για την καθοδήγησή τους. Τέλος, θα ήθελα να ευχαριστήσω τους γονείς, τα αδέρφια μου, φίλους και φίλες για την ηθική συμπαράσταση που μου προσέφεραν όλα αυτά τα χρόνια.



# Περίληψη

Η Επαυξημένη Πραγματικότητα αποτελεί ένα αναπτυσσόμενο πεδίο τα τελευταία χρόνια στον τομέα της Εικονικής Πραγματικότητας. Ένα σύστημα Επαυξημένης Πραγματικότητας προσθέτει ένα γραφικό μοντέλο στον πραγματικό χώρο δημιουργώντας στο χρήστη την ψευδαίσθηση της συνύπαρξης πραγματικού και εικονικού κόσμου. Η διαδικασία αυτή απαιτεί ακριβή εκτίμηση θέσης της κάμερας με μεθόδους όρασης υπολογιστών.

Σκοπός της παρούσας διπλωματικής είναι η μελέτη των μεθόδων εκτίμησης της θέσης της κάμερας, καθώς και η υλοποίηση μιας εφαρμογής Επαυξημένης Πραγματικότητας με βάση την αντίχνευση τοπικών χαρακτηριστικών σε διαδοχικά καρέ, χωρίς πρότερη γνώση του περιβάλλοντος. Για την εκτίμηση θέσης της κάμερας γίνεται χρήση των επιπολικών περιορισμών που ορίζει η Επιπολική γεωμετρία.

## Λέξεις Κλειδιά

Όραση υπολογιστών, επαυξημένη πραγματικότητα, τοπικά χαρακτηριστικά, επιπολική γεωμετρία, εκτίμηση θέσης της κάμερας





# **Abstract**

Augmented Reality is a growing area in the recent years, in the field of Virtual Reality. A system of Augmented Reality supplements the real world with virtual graphical models, creating the illusion of coexistence between real and virtual world. This process requires accurate camera pose estimation based on computer vision methods.

In the framework of this thesis, we aim to study the camera pose estimation methods based on the detection of local features in successive frames, without prior knowledge of the environment. Furthermore we present an implementation of an Augmented Reality application.

Epipolar constraints, defined by the Epipolar Geometry, are used for the camera pose estimation.

## **Keywords**

Computer vision, augmented reality, local features, epipolar geometry, camera pose estimation



# Περιεχόμενα

<b>1</b>	<b>Εισαγωγή</b>	<b>19</b>
1.1	Επαυξημένη Πραγματικότητα και Εφαρμογές . . . . .	19
1.1.1	Χαρακτηριστικά εφαρμογών Επαυξημένης Πραγματικότητας . . . . .	20
1.1.2	Το πρόβλημα ευθυγράμμισης (registration) . . . . .	21
1.1.3	Εφαρμογές της επαυξημένης πραγματικότητας . . . . .	21
1.2	Συνεισφορά εργασίας . . . . .	23
1.3	Δομή Διπλωματικής . . . . .	24
<b>2</b>	<b>Τοπικά χαρακτηριστικά</b>	<b>25</b>
2.1	Εισαγωγή . . . . .	25
2.2	Ιδιότητες Τοπικών Χαρακτηριστικών . . . . .	27
2.3	Scale-Invariant Feature Transform (SIFT) . . . . .	27
2.4	Speeded-Up Robust Features (SURF) . . . . .	28
2.5	Features from Accelerated Segment Test (FAST) . . . . .	29
2.6	BRISK (Binary Robust Invariant Scalable Keypoints) . . . . .	29
<b>3</b>	<b>Προβολική Γεωμετρία και Όραση Πολλαπλών Όψεων</b>	<b>31</b>
3.1	Εισαγωγή . . . . .	31
3.2	Προβολική Γεωμετρία . . . . .	31
3.3	Μοντέλο Κάμερας Μικροσκοπικής Οπής . . . . .	33
3.4	Επιπολική Γεωμετρία . . . . .	36

3.4.1	Ο Θεμελιώδης Πίνακας (Fundamental Matrix) $F$ . . . . .	37
3.4.2	Ο Βασικός Πίνακας (Essential matrix) $E$ . . . . .	38
3.4.3	Υπολογισμός Σχετικής Κίνησης της Κάμερας από το Βασικό πίνακα $E$ . . . . .	39
<b>4</b>	<b>Αλγόριθμοι εκτίμησης θέσης της κάμερας</b>	<b>41</b>
4.1	Εισαγωγή . . . . .	41
4.2	Μέθοδος εύρωστης εκτίμησης RANSAC . . . . .	41
4.3	Σφάλμα Sampson Θεμελιώδους πίνακα $F$ . . . . .	42
4.4	Κανονικοποιημένος Αλγόριθμος 8-σημείων . . . . .	43
4.5	Αλγόριθμος εκτίμησης πίνακα Εξωτερικών παραμέτρων της κάμερας . . . . .	45
<b>5</b>	<b>Προσέγγιση με Τοπικά Χαρακτηριστικά</b>	<b>47</b>
5.1	Εισαγωγή . . . . .	47
5.2	Τοπικά Χαρακτηριστικά BRISK . . . . .	47
5.3	Πλατφόρμα Οπτικής Οδομετρίας . . . . .	49
5.3.1	Τοπικά Χαρακτηριστικά πλατφόρμας . . . . .	49
5.3.2	Εκτίμηση Θέσης της Κάμερας . . . . .	50
5.4	Η εφαρμογή . . . . .	51
5.4.1	Εκτίμηση θέσης της κάμερας . . . . .	52
5.4.2	Παρουσίαση Επαύξησης . . . . .	54
5.5	Παρουσίαση Αποτελεσμάτων . . . . .	55
5.5.1	Εκτίμηση Θέσης από σύνολο δεδομένων . . . . .	56
5.5.2	Ταύτιση τοπικών χαρακτηριστικών . . . . .	58
5.5.3	Επαύξηση της πραγματικότητας-Ευθυγράμμιση . . . . .	63
<b>6</b>	<b>Συμπεράσματα-Μελλοντικές επεκτάσεις</b>	<b>69</b>
6.1	Συμπεράσματα . . . . .	69

6.2	Μελλοντικές επεκτάσεις . . . . .	70
-----	----------------------------------	----



# Κατάλογος σχημάτων

1.1	Συνεχές πραγματικότητας-εικονικότητας . . . . .	19
1.2	Εφαρμογές Επαυξημένης Πραγματικότητας . . . . .	22
2.1	Περιγραφείας Χαρακτηριστικών SIFT . . . . .	28
2.2	Προσέγγιση SURF τετραγωνικών φίλτρων . . . . .	29
3.1	Προβολική Γεωμετρία . . . . .	33
3.2	Μοντέλο Κάμερας Μικροσκοπικής Οπής . . . . .	33
3.3	Γεωμετρία Κάμερας Μικροσκοπικής Οπής . . . . .	34
3.4	Επιπολική Γεωμετρία . . . . .	36
3.5	Μονοπαραμετρική οικογένεια επιπολικών επιπέδων . . . . .	37
3.6	Οι 4 δυνατές λύσεις του πίνακα προβολής $P'$ . . . . .	40
5.1	Τοπικά Χαρακτηριστικά BRISK . . . . .	49
5.2	Διάγραμμα ροής Εκτίμησης Θέσης . . . . .	53
5.3	Διάγραμμα Ροής Δευτερεύοντος Νήματος . . . . .	55
5.4	Γραφικό μοντέλο OpenGL . . . . .	55
5.5	Διάγραμμα εκτίμησης θέσης . . . . .	56
5.6	Συγκριτικό διάγραμμα εκτίμησης θέσης . . . . .	57
5.7	Τοπικά χαρακτηριστικά LibViso2 σε διαδοχικά καρέ . . . . .	59
5.8	Τοπικά χαρακτηριστικά BRISK σε διαδοχικά καρέ . . . . .	60
5.9	Τοπικά χαρακτηριστικά SURF σε διαδοχικά καρέ . . . . .	61

5.10 Τοπικά χαρακτηριστικά SIFT σε διαδοχικά καρέ . . . . .	62
5.11 Επαύξηση της σκηνής με χρήση Σποραδικών χαρακτηριστικών . . . . .	64
5.12 Επαύξηση της σκηνής με χρήση τοπικών χαρακτηριστικών BRISK . . . . .	65
5.13 Επαύξηση της σκηνής με χρήση τοπικών χαρακτηριστικών SURF . . . . .	66
5.14 Επαύξηση της σκηνής με χρήση τοπικών χαρακτηριστικών SIFT . . . . .	67



## Κατάλογος πινάκων

5.1	Πίνακας ταυτισμένων τοπικών χαρακτηριστικών . . . . .	58
5.2	Πίνακας σύγκρισης χρόνων . . . . .	63

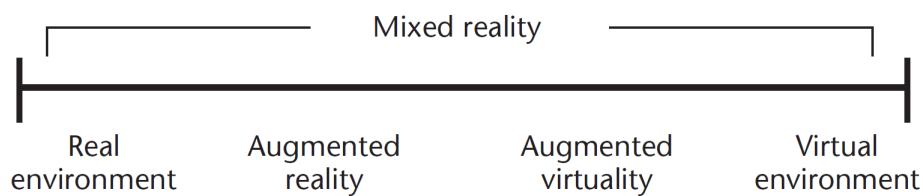


# Κεφάλαιο 1

## Εισαγωγή

### 1.1 Επαυξημένη Πραγματικότητα και Εφαρμογές

Ένα ταχέως αναπτυσσόμενο ερευνητικό πεδίο τα τελευταία χρόνια στο χώρο της εικονικής πραγματικότητας (virtual reality-VR) είναι το πεδίο της *Επαυξημένης Πραγματικότητας* (*augmented reality-AR*). Η Επαυξημένη Πραγματικότητα αποτελεί ένα μέρος του γενικότερου πεδίου της Μικτής Πραγματικότητας (Mixed Reality) (Σχήμα 1.1) [1].



Σχήμα 1.1: Συνεχές πραγματικότητας-εικονικότητας

Ένα σύστημα επαυξημένης πραγματικότητας συμπληρώνει τον πραγματικό κόσμο με εικονικά αντικείμενα ή πληροφορίες τα οποία συνυπάρχουν στον ίδιο χώρο με τον πραγματικό κόσμο. Τα συστήματα Επαυξημένης πραγματικότητας έχουν τις εξής ιδιότητες:

- Συνδυάζουν πραγματικά και εικονικά αντικείμενα σε ένα πραγματικό περιβάλλον.
- Αλληλεπιδρούν με το περιβάλλον σε πραγματικό χρόνο.
- Ευθυγραμμίζουν πραγματικά και εικονικά αντικείμενα ανάλογα με το οπτικό πεδίο του χρήστη.

- Ένα σύστημα επαυξημένης πραγματικότητας μπορεί να επιδράσει σε όλες τις αισθήσεις του χρήστη.

### 1.1.1 Χαρακτηριστικά εφαρμογών Επαυξημένης Πραγματικότητας

Μια εφαρμογή επαυξημένης πραγματικότητας, όσον αφορά την οπτική επαύξηση μια σκηνής, δεν περιορίζεται μόνο στην προσθήκη εικονικών αντικειμένων ή πληροφοριών σε ένα πραγματικό τοπίο αλλά μπορεί να περιλαμβάνει την απόκρυψη αντικειμένων του πραγματικού κόσμου, επικαλύπτοντας τα με ένα εικονικό αντικείμενο. Αυτή η διαδικασία πρέπει να γίνεται σε πραγματικό χρόνο δίνοντας στον χρήστη τη δυνατότητα αλληλεπίδρασης με το εικονικό περιβάλλον. Τα βήματα τα οποία ακολουθούνται σε εφαρμογή επαυξημένης πραγματικότητας είναι τα εξής [1] [2]:

- Παρουσίαση στη συσκευή απεικόνισης του πραγματικού περιβάλλοντος
- Δημιουργία των εικονικών αντικειμένων τα οποία θα προβληθούν είτε για να προσθέσουν πληροφορία στο πραγματικό περιβάλλον είτε για να επικαλύψουν το επιθυμητό μέρος του πραγματικού περιβάλλοντος
- Ευθυγράμμιση (registration) του πραγματικού κόσμου με τα εικονικά μοντέλα
- Απεικόνιση της εικονικής πληροφορίας στον πραγματικό κόσμο
- Εκτίμηση της κίνησης και θέσης του χρήστη ως προς τον πραγματικό κόσμο, ούτως ώστε να υπάρξει ανάλογη μεταβολή της απεικόνισης των εικονικών πληροφοριών

Μπορεί κανείς να αντιληφθεί ότι οι συσκευές απεικόνισης αποτελούν ένα σημαντικό κομμάτι σε κάθε εφαρμογή επαυξημένης πραγματικότητας. Για λόγους πληρότητας πιο κάτω αναφέρονται συνοπτικά οι κατηγορίες στις οποίες διακρίνουμε τις συσκευές αυτές:

- Φορετές συσκευές απεικόνισης (Head mounted Displays-HMD). Πρόκειται για ειδικές συσκευές οι οποίες τοποθετούνται στο κεφάλι του χρήστη και προβάλλουν τη μίξη της εικόνας στα μάτια του χρήστη
- Φορητές συσκευές. Αποτελούνται από μια κάμερα και μία επίπεδη οθόνη. Έτσι ο χρήστης αντιλαμβάνεται τη μίξη εικονικού και πραγματικού περιβάλλοντος μέσω της οθόνης του συστήματος
- Προβολικές συσκευές. Μέσω μιας προβολικής συσκευής (projector) η εικονική πληροφορία προβάλλεται απευθείας πάνω στον πραγματικό κόσμο.

### 1.1.2 Το πρόβλημα ευθυγράμμισης (registration)

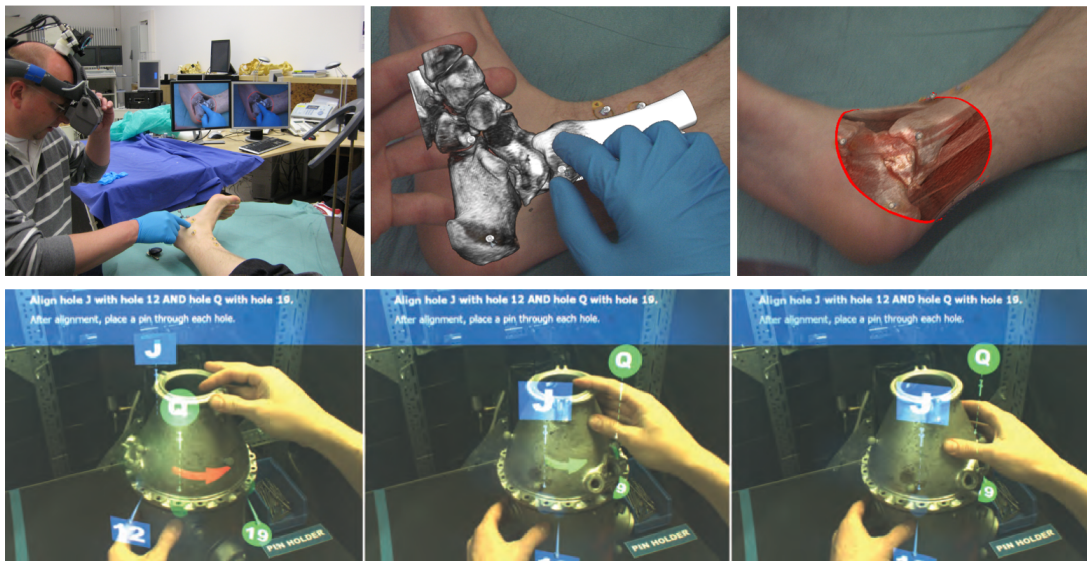
Το πιο σημαντικό κομμάτι σε μια εφαρμογή Επαυξημένης Πραγματικότητας είναι η επιτυχία ευθυγράμμισης του πραγματικού περιβάλλοντος με την εικονική πληροφορία. Το συγκεκριμένο πρόβλημα αποτελεί τον πιο βασικό περιορισμό στις εφαρμογές επαυξημένης πραγματικότητας, καθώς η ψευδαίσθηση συνύπαρξης πραγματικού και εικονικού κόσμου στηρίζεται στην απόλυτη ευθυγράμμιση των δύο κόσμων.

Η επίλυση του προβλήματος αυτού γίνεται δυσκολότερη λόγω της ευαισθησίας του ανθρώπινου οπτικού συστήματος καθώς και της αλληλεπίδρασης του χρήστη με το εικονικό αντικείμενο. Προφανώς ένα σύστημα επαυξημένης πραγματικότητας δεν μπορεί να ελέγχει την κίνηση του χρήστη αλλά πρέπει να ανταποκρίνεται με σωστή απεικόνιση σε χρονικό διάστημα μη αντιληπτό απ' αυτόν. Λόγω της απαίτησης για υψηλή ακρίβεια και των πολλαπλών πηγών σφαλμάτων τα σφάλματα ευθυγράμμισης δεν μπορούν να ελεγχθούν επαρκώς. Οι πηγές σφαλμάτων μπορούν να διαχωριστούν σε δύο κατηγορίες [3]:

- Στατικά σφάλματα, τα οποία προκαλούν απόκλιση της ευθυγράμμισης των περιβάλλοντων ακόμα και όταν η θέση του χρήστη ή το πραγματικό περιβάλλον δεν μεταβάλλονται καθόλου. Οφείλονται σε οπτική στρέβλωση των εικόνων λόγω κατασκευαστικών ατελειών της κάμερας, ή οποία αντιμετωπίζεται με τεχνικές βαθμονόμησης (calibration). Επίσης μπορεί να οφείλονται σε σφάλματα του συστήματος εκτίμησης της θέσης του χρήστη, τα οποία αποτελούν και τα πιο σημαντικά στατικά σφάλματα.
- Δυναμικά σφάλματα, τα οποία γίνονται αντιληπτά μόνο κατά τη μεταβολή του πραγματικού χώρου ή της θέσης του χρήστη και τα οποία οφείλονται σε καθυστερήσεις του συστήματος να προβάλλουν τη σωστή θέση και προσανατολισμό του εικονικού αντικειμένου τη σωστή χρονική στιγμή σε σχέση με το πραγματικό περιβάλλον. Για να αντιμετωπιστούν τέτοια σφάλματα υιοθετούνται τεχνικές μείωσης της καθυστέρησης του συστήματος, τεχνικές ταύτισης των καναλιών πραγματικής-εικονικής σκηνής πριν παρουσιαστούν στο χρήστη καθώς και τεχνικές εκτίμησης μελλοντικής θέσης του χρήστη για καλύτερη απόκριση του συστήματος.

### 1.1.3 Εφαρμογές της επαυξημένης πραγματικότητας

Με την ανάπτυξη της τεχνολογίας και κυρίως της υπολογιστικής ισχύος των σύγχρονων ηλεκτρονικών υπολογιστών οι ερευνητές και προγραμματιστές ανακαλύπτουν συνεχώς όλο και περισσότερους τομείς οι οποίοι θα μπορούσαν να επωφεληθούν από την επαύξηση της πραγματικότητας. Αν και αρχικά τα πρώτα συστήματα Επαυξημένης Πραγματικότητας στόχευαν σε στρατιωτικές, βιομηχανικές και ιατρικές εφαρμογές τα τελευταία χρόνια η επαυξημένη πραγματικότητα



Σχήμα 1.2: **Πάνω:** Εφαρμογή στην ιατρική. [4] **Κάτω:** Εφαρμογή στη βιομηχανία. [5]

εφαρμόζεται για εμπορικούς σκοπούς καθώς και για σκοπούς ψυχαγωγίας. Πιο κάτω αναφέρονται αναλυτικότερα οι τομείς στους οποίους εφαρμόζεται ή μπορεί να εφαρμοστεί η επαυξημένη πραγματικότητα [2]:

- Σε συστήματα προσωπικής ενημέρωσης: Με την τεχνολογία στους προσωπικούς ηλεκτρονικούς υπολογιστές να αναπτύσσεται όλο και περισσότερο πολλοί πιστεύουν ότι τα συστήματα Επαυξημένης Πραγματικότητας μπορούν να αποτελέσουν ένα προηγμένο, άμεσο και πιο φυσικό περιβάλλον διεπαφής χρήστη (user interface) για την καθημερινή προσωπική χρήση των υπολογιστών.
- Στην πλοήγηση σε εσωτερικούς ή εξωτερικούς χώρους: Η πλοήγηση σε ένα μουσείο ή συνεδριακό κέντρο θα ήταν πολύ πιο απλή με τη προβολή πληροφοριών στο χρήστη μέσω της οθόνης του κινητού του. Παράλληλα γίνονται προσπάθειες για ανάπτυξη εφαρμογών πλοήγησης αυτοκινήτων.
- Στον τουρισμό: Αρχαιολογικοί χώροι στους οποίους ο επισκέπτης θα έχει την δυνατότητα να θαυμάσει τα μνημεία πριν τα αγγίξει το πέρασμα του χρόνου, μουσεία τα οποία θα προβάλλουν ενημερωτικά κείμενα ή αναπαραστάσεις δίπλα από τα εκθέματα.
- Στην εκπαίδευση και στη βιομηχανία: Διαδραστικές εφαρμογές για τους μαθητές, και εφαρμογές εκπαίδευσης για μηχανικούς, ενώ ένας συντηρητής θα μπορούσε να βλέπει το ελαττωματικό εξάρτημα και τις τεχνικές προδιαγραφές του με έντονο χρώμα.
- Στην αρχιτεκτονική: Ένα εικονικό μοντέλο του κτιρίου θα μπορούσε να προβάλλεται στον

πραγματικό κόσμο για να βοηθήσει τον αρχιτέκτονα ή τους μελλοντικούς αγοραστές του κτιρίου να πάρουν κάποιες σημαντικές αποφάσεις.

- Στην ψυχαγωγία/διαφήμιση: Η επαύξηση της πραγματικότητας ήδη εφαρμόζεται σε πολλούς τομείς της ψυχαγωγίας και της διαφήμισης. Τρισδιάστατα εικονικά μοντέλα παρουσιάζονται να αλληλεπιδρούν με τους ηθοποιούς σχεδόν σε όλες τις κινηματογραφικές ταινίες ενώ στα δελτία ειδήσεων και στην αναμετάδοση αθλητικών γεγονότων εικονικές πληροφορίες παρουσιάζονται στον τηλεθεατή για να έχει μια καλύτερη ενημέρωση. Στον τομέα της διαφήμισης εικονικά μοντέλα παρουσιάζονται στο κέντρο του γηπέδου ή στις διαφημιστικές πινακίδες τα οποία γίνονται αντιληπτά μόνο από τον τηλεθεατή.
- Στην ιατρική: Ίσως από τους σημαντικότερους τομείς στους οποίους η επαύξηση της πραγματικότητας βρίσκει εφαρμογή. Οι περισσότερες προσπάθειες των ερευνητών σε αυτό τον τομέα στρέφονται στην υποβοήθηση εγχειρήσεων. Εικόνες από ακτινογραφίες ή αξονικές /μαγνητικές τομογραφίες μπορούν να προβάλλονται πάνω στο σώμα του ασθενούς όταν ο γιατρός τις χρειάζεται. Παρόλα αυτά, η ακρίβεια ευθυγράμμισης των δύο κόσμων σε ιατρικές εφαρμογές αποτελεί το μεγαλύτερο πρόβλημα για τους ερευνητές.

## 1.2 Συνεισφορά εργασίας

Η πλειοψηφία των εφαρμογών Επαυξημένης πραγματικότητας λειτουργεί με την προϋπόθεση ότι έχουν κάποια πρότερη γνώση του περιβάλλοντος. Αυτή η γνώση μπορεί να περιορίζεται σε κάποιες 3D συντεταγμένες στο χώρο, είτε στην ύπαρξη κάποιου προκαθορισμένου σημειωτή στη σκηνή. Η ευθυγράμμιση των δύο κόσμων, πραγματικού και εικονικού, μπορεί να προκύψει απευθείας από τη γνώση αυτή. Αυτή η διαδικασία αποτελεί την πιο συνηθισμένη προσέγγιση στις εφαρμογές Επαυξημένης πραγματικότητας. Παρόλα αυτά, πρότερη γνώση του περιβάλλοντος δεν είναι πάντα διαθέσιμη στο χρήστη.

Στα πλαίσια της παρούσας διπλωματικής θα μελετηθεί το πρόβλημα εκτίμησης θέσης της κάμερας σε άγνωστο περιβάλλον, χωρίς καμία πρότερη γνώση του συστήματος. Η προσέγγιση αυτή χρησιμοποιεί εκτίμηση θέσης της κάμερας από τα τοπικά χαρακτηριστικά, με βάση την Επιπολική γεωμετρία. Η μέθοδος αυτή έχει μελετηθεί στον τομέα της Ρομποτικής για αυτοεντοπισμό και χαρτογράφηση του χώρου από κινούμενα ρομπότ (Simultaneous Localization and Mapping). Παρόλα αυτά, η εφαρμογή της στην Επαυξημένη πραγματικότητα κρίνεται πολύ πιο δύσκολη.

Η κύρια δυσκολία βρίσκεται στη σωστή εκτίμηση της κάμερας, καθώς ο χρήστης έχει την ελευθερία να κινηθεί με γρήγορες κινήσεις στο χώρο ενώ ένα ρομπότ συνήθως έχει σταθερή ταχύτητα και μικρές κινήσεις. Ακόμη μία δυσκολία στις εφαρμογές Επαυξημένης πραγματικότητας είναι η δυνατότητα της εφαρμογής να ανταποκρίνεται σε πραγματικό χρόνο.

Μελετώντας το πρόβλημα αυτό, στα πλαίσια της παρούσας διπλωματικής, υλοποιήθηκε μια εφαρμογή Επαυξημένης Πραγματικότητας με χρήση τοπικών χαρακτηριστικών. Η εφαρμογή δεν έχει πρότερη γνώση του περιβάλλοντος και παρουσιάζει την επαυξημένη σκηνή σε πραγματικό χρόνο.

### 1.3 Δομή Διπλωματικής

Η δομή της παρούσας διπλωματικής στηρίζεται στη σταδιακή μελέτη των μεθόδων εκτίμησης θέσης της κάμερας με στόχο την πλήρη κατανόηση της εφαρμογής που υλοποιήθηκε. Το **Κεφάλαιο 2** περιέχει βιβλιογραφική επισκόπηση του πεδίου των τοπικών χαρακτηριστικών. Ακολουθώντας, στο **Κεφάλαιο 3** εισάγεται η Προβολική γεωμετρία, ως επέκταση της Ευκλείδειας, και μέσω αυτής παρουσιάζεται το μοντέλο κάμερας μικροσκοπικής οπής. Χρησιμοποιώντας το μοντέλο αυτό αναλύονται η Επιπολική γεωμετρία και οι επιπολικοί περιορισμοί μεταξύ δύο εικόνων. Στο **Κεφάλαιο 4** παρουσιάζονται οι αλγόριθμοι και τα μαθηματικά εργαλεία που χρησιμοποιούνται για την εκτίμηση θέσης της κάμερας μέσω των επιπολικών περιορισμών. Στη συνέχεια στο **Κεφάλαιο 5** παρουσιάζεται η εφαρμογή που υλοποιήθηκε στα πλαίσια της παρούσας διπλωματικής καθώς και τα βασικά τμήματα που την αποτελούν μέσα από διαγράμματα ροής και επιλεγμένα στιγμιότυπα. Τέλος συμπεράσματα και μελλοντικές επεκτάσεις της εφαρμογής παρουσιάζονται στο **Κεφάλαιο 6**.



## Κεφάλαιο 2

# Τοπικά χαρακτηριστικά

### 2.1 Εισαγωγή

Στο κεφάλαιο αυτό παρουσιάζονται μέσα από την υπάρχουσα βιβλιογραφία τα τοπικά χαρακτηριστικά εικόνων. Ιδιαίτερη αναφορά γίνεται στους ανιχνευτές τοπικών χαρακτηριστικών οι οποίοι χρησιμοποιούνται ευρέως σήμερα στις εφαρμογές Όρασης Υπολογιστών, κυρίως λόγω της υπολογιστικής τους αποδοτικότητας.

Τα οπτικά τοπικά χαρακτηριστικά (local visual features) αποτελούν περιοχές σε μια εικόνα οι οποίες διαφέρουν από την άμεση γειτονιά τους. Συνήθως σχετίζονται με τη μεταβολή σε μια περιοχή μίας ή περισσότερων ιδιοτήτων ταυτόχρονα, όπως ένταση, χρώμα ή υφή. Τα οπτικά τοπικά χαρακτηριστικά έχουν εφαρμοστεί με επιτυχία σε ένα ευρύ φάσμα συστημάτων και εφαρμογών, και η επιλογή τους γίνεται ανάλογα με την εκάστοτε εφαρμογή. Σε εφαρμογές όπως η 3D ανακατασκευή σκηνών, βαθμονόμηση καμερών και φυσικά εκτίμηση θέσης, η οποία όπως αναφέρθηκε είναι απαραίτητο σε εφαρμογές επαυξημένης πραγματικότητας, επιλέγονται τοπικά χαρακτηριστικά τα οποία είναι σταθερά και η θέση τους στην εικόνα μπορεί να καθοριστεί με ακρίβεια.

Η χρήση των τοπικών χαρακτηριστικών στην όραση υπολογιστών ολοένα και αυξάνεται καθώς η αύξηση της υπολογιστικής δυνατότητας των υπολογιστών δίνει περισσότερες δυνατότητες στην επεξεργασία εικόνων. Αν και η υπολογιστική δύναμη περιορίζει τις δυνατές εφαρμογές, τα τοπικά χαρακτηριστικά έχουν κάνει την εμφάνισή τους εδώ και πολλές δεκαετίες στον τομέα αυτό. Η θεωρία με την οποία υλοποιήθηκαν οι σημερινές μέθοδοι εξαγωγής τοπικών χαρακτηριστικών πρωτοεμφανίστηκε τη δεκαετία του 1970 και χωρίζεται σε δύο πεδία:

- Στην ανίχνευση τοπικών χαρακτηριστικών με χρήση ενός ανιχνευτή χαρακτηριστικών (key-point detector). Το εργαλείο αυτό έχει σκοπό να εντοπίζει τα χαρακτηριστικά σε μία εικόνα.

- Στην περιγραφή τοπικών χαρακτηριστικών με τη χρήση ενός περιγραφέα χαρακτηριστικών (keypoint descriptor). Αντίστοιχα ο περιγραφέας χαρακτηριστικών περιέχει όλες τις σημαντικές και μοναδικές πληροφορίες για κάθε τοπικό χαρακτηριστικό, ούτως ώστε να αναγνωρίζεται επιτυχώς σε κάθε εικόνα και ενδεχομένως να ταυτίζεται σε μία άλλη.

Όπως αναφέρουν οι Tutelaars et al. η θεωρία της ανίχνευσης τοπικών χαρακτηριστικών χρονολογείται από το 1954 [6]. Ερευνητές παρατήρησαν ότι πληροφορίες για το σχήμα κάποιου αντικειμένου συγκεντρώνονται σε σημεία όπου η αλλαγή του περιγράμματος του αντικειμένου μεγιστοποιείται [7]. Αρχικά η ανίχνευση τοπικών χαρακτηριστικών επικεντρώθηκε στην ανίχνευση απλών αντικειμένων σε εικόνες εργαλείων σχεδίασης παρά σε φυσικές εικόνες. Αργότερα, το 1978, προτάθηκε ο Hessian ανιχνευτής από τον Beaudet [8], βασισμένος στην Hessian μήτρα. Ο ανιχνευτής αυτός μπορεί να ανιχνεύσει περιοχές (blobs) και κορυφογραμμές (ridges) σε μία εικόνα. Η ανίχνευση τοπικών χαρακτηριστικών σε φυσικές εικόνες αναπτύχθηκε από τον Moravec στο τέλος της δεκαετίας του 1970. [9] Ο ανιχνευτής που πρότεινε στηρίχθηκε στην ιδέα της μεταβολής της έντασης σε μία τοπική γειτονιά ενός εικονοστοιχείου. Αργότερα οι Harris και Stephens βελτίωσαν αυτή την ιδέα προτείνοντας τον ανιχνευτή γωνιών Harris [10]. Οι παραπάνω μέθοδοι είχαν το μειονέκτημα ότι δεν ήταν ανεξάρτητες της κλίμακας της εικόνας (scale-invariant), δηλαδή δεν μπορούσαν να εξάγουν τα ίδια χαρακτηριστικά αν η κλίμακα της εικόνας μεταβαλλόταν. Σημαντική δουλειά στον τομέα των scale-invariant ανιχνευτών έγινε από τον Lindeberg τη δεκαετία 1990, [11] [12] συνδράμοντας στην ανάπτυξη:

- Των ανιχνευτών Harris-Affine και Hessian-Affine από τους Mikolajczyk και Schmid [13].
- Του Maximally stable extremal regions (MSER) από Matas et al. [14].
- Του Difference-of-Gaussian (DoG) από Lowe [15].
- Του Laplacian-of-Gaussian(LoG) Lindeberg [11].

Στο πεδίο της περιγραφής των τοπικών χαρακτηριστικών οι πιο διαδεδομένες μέθοδοι μέχρι σήμερα είναι:

- Scale-Invariant Feature Transform (SIFT) από τον Lowe το 1999 [15] και διάφορες παραλλαγές όπως ο PCA-SIFT [16] και GLOH [17].
- Ο Speeded-Up Robust Features (SURF) από Bay et al. [18], ο οποίος αποτελείται τόσο από ανιχνευτή όσο και από περιγραφέα τοπικών χαρακτηριστικών .

Στις ενότητες 2.3, 2.4 και 2.5 θα παρουσιαστούν λεπτομερώς τα τοπικά χαρακτηριστικά SIFT, SURF και FAST,.

## 2.2 Ιδιότητες Τοπικών Χαρακτηριστικών

Η επιλογή των κατάλληλων τοπικών χαρακτηριστικών για μία εφαρμογή ενδέχεται να μην είναι προφανής. Πιο κάτω αναφέρονται αναλυτικά οι ιδιότητες τις οποίες πρέπει να έχουν τα ιδανικά τοπικά χαρακτηριστικά [6]:

- **Επαναληπτικότητα:** Δεδομένου ότι δύο εικόνες περιέχουν την ίδια σκηνή υπό διαφορετικές λήψεις, ένα υψηλό ποσοστό των ίδιων τοπικών χαρακτηριστικών πρέπει να ανιχνεύεται και στις δύο εικόνες.
- **Διακριτότητα:** Οι περιοχές που ανιχνεύονται ως τοπικά χαρακτηριστικά πρέπει να διαφέρουν μεταξύ τους ούτως ώστε να είναι εφικτή η ταύτιση τους μεταξύ δύο εικόνων.
- **Πλήθος:** το πλήθος των τοπικών χαρακτηριστικών σε μια εικόνα πρέπει να είναι επαρκές για να τη χαρακτηρίζει πλήρως και ταυτόχρονα να ικανοποιεί τις απαιτήσεις της εφαρμογής.
- **Ακρίβεια:** Η θέση τους καθώς και η κλίμακά τους πρέπει να καθορίζονται με ακρίβεια.
- **Αποδοτικότητα:** Ιδιαίτερα για εφαρμογές οι οποίες τρέχουν σε πραγματικό χρόνο, ο χρόνος υπολογισμού των τοπικών χαρακτηριστικών πρέπει να είναι όσο το δυνατόν καλύτερος.

## 2.3 Scale-Invariant Feature Transform (SIFT)

Τα χαρακτηριστικά SIFT είναι ένας συνδυασμός του ανιχνευτή τοπικών χαρακτηριστικών DoG και ενός περιγραφέα χαρακτηριστικών. Ο περιγραφέας χαρακτηριστικών SIFT είναι ένας εύρωστος (robust) περιγραφέας, ο οποίος έχει τη δυνατότητα να παρέχει διακριτή περιγραφή του κάθε χαρακτηριστικού που ανιχνεύεται.

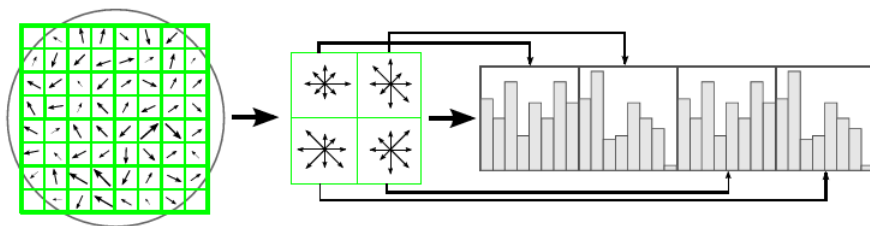
Όταν τα χαρακτηριστικά ανιχνευτούν αναπαριστώνται στο χώρο της κλίμακας (scale-space) με σκοπό τον εντοπισμό τοπικών μεγίστων. Αυτά τα τοπικά μέγιστα αποτελούν τη χαρακτηριστική κλίμακα του κάθε χαρακτηριστικού. Με αυτόν το τρόπο τα τοπικά χαρακτηριστικά με συντεταγμένες  $(x, y)$  τοποθετούνται σε ένα επίπεδο  $\sigma$  της πυραμίδας στο χώρο της κλίμακας  $L(x, y, \sigma)$ .

Για κάθε τοπικό χαρακτηριστικό  $L(x, y, \sigma)$  υπολογίζεται το μέγεθος της κλίσης (gradient magnitude)  $m(x, y)$  και ο προσανατολισμός  $\vartheta(x, y)$  ως προς την γειτονιά του, χρησιμοποιώντας διαφορές εικονοστοιχείων (pixels):

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2} \quad (2.1)$$

$$\theta(x, y) = \tan^{-1} \left( \frac{L(x, y+1) - L(x, y-1)}{L(x+1, y) - L(x-1, y)} \right) \quad (2.2)$$

Για κάθε χαρακτηριστικό δημιουργείται ένα ιστόγραμμα προσανατολισμού το οποίο καλύπτει όλη τη γειτονιά γύρω από αυτό. Για να καλυφθεί όλη η γειτονιά χρησιμοποιούνται 36 στοιχεία διαμέρισης, καλύπτοντας το καθένα γωνία  $10^\circ$ , τα οποία περιέχουν μια μέτρηση για την αλλαγή της έντασης προς μια κατεύθυνση (Σχήμα 2.1). Η μέγιστη μέτρηση αυτών των εντάσεων θεωρείται ο κυρίαρχος προσανατολισμός του χαρακτηριστικού, ενώ αν υπάρχουν στοιχεία διαμέρισης με ένταση  $\geq 80\%$  του κυρίαρχου προσανατολισμού τα στοιχεία αυτά θεωρούνται δευτερεύουσα περιγραφή για το συγκεκριμένο χαρακτηριστικό. Η μέγιστη ένταση των στοιχείων διαμέρισης της δευτερεύουσας περιγραφής θεωρείται και ο κυρίαρχος προσανατολισμός της.



Σχήμα 2.1: Περιγραφέας Χαρακτηριστικών SIFT

## 2.4 Speeded-Up Robust Features (SURF)

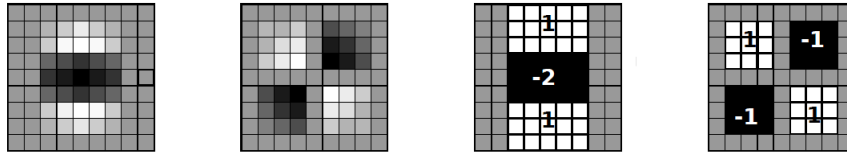
Τα τοπικά χαρακτηριστικά *Speeded-Up Robust Features (SURF)* είναι υπολογιστικά πολύ γρήγορα χωρίς να θυσιάζουν αποδοτικότητα στην ανίχνευση και περιγραφή των χαρακτηριστικών. Η χαμηλή απαίτηση τους για υπολογιστική δύναμη επιτυγχάνεται με τη χρησιμοποίηση “ολοκληρωτικών εικόνων” (integral images). Η τεχνική αυτή προσφέρει έναν τρόπο υπολογισμού των αποκρίσεων τετραγωνικών (box-type) φίλτρων σε πραγματικό χρόνο.

Η ανίχνευση τοπικών χαρακτηριστικών SURF στηρίζεται στη Hessian μήτρα, η οποία εκτιμάται χρησιμοποιώντας απλά τετραγωνικά φίλτρα αυξανόμενου μεγέθους. Προσεγγίζοντας τα τετραγωνικά φίλτρα όπως φαίνεται στο Σχήμα 2.2 η απόκριση των φίλτρων υπολογίζεται από την ορίζουσα της Hessian μήτρας, ως εξής:

$$\det(H_{approx}) = D_{xx}D_{yy} - (wD_{xy})^2 \quad (2.3)$$

Το βάρος  $w$  στα χαρακτηριστικά SURF είναι σταθερό  $w = 0.9$  σε όλες τις κλίμακες. Για την ανίχνευση χαρακτηριστικών διαφόρων μεγεθών είναι απαραίτητη η αναπαράσταση στο χώρο της κλίμακας (scale-space). Μια οκτάβα αποτελείται από τη φιλτραρισμένη εικόνα με συγκεκριμένο μέγεθος φίλτρου (π.χ.  $9 \times 9$ ,  $15 \times 15$ ,  $21 \times 21$ ,  $27 \times 27$ ). Σε κάθε οκτάβα το μέγεθος του φίλτρου διπλασιάζεται. Η μικρή υπολογιστική απαίτηση, χρησιμοποιώντας τις ολοκληρωτικές εικόνες, επιτρέπει στα χαρακτηριστικά SURF να αυξάνουν το μέγεθος των φίλτρων αντί να υποδειγματολη-

πτούν την εικόνα. Τα τοπικά χαρακτηριστικά επιλέγονται ως τα τοπικά μέγιστα των οκτάβων στο χώρο της κλίμακας.



Σχήμα 2.2: **Αριστερά προς δεξιά:** Διακριτή Gaussian παράγωγος 2ης τάξης ( $D_{yy}$  και  $D_{xy}$ ) και η SURF προσέγγιση της.

## 2.5 Features from Accelerated Segment Test (FAST)

Ο ανιχνευτής τοπικών χαρακτηριστικών FAST αποτελεί έναν ανιχνευτή γωνιών [19]. Η ανίχνευση γίνεται συγκρίνοντας τη φωτεινότητα των εικονοστοιχείων σε μια γειτονιά ακτίνας 16 εικονοστοιχείων γύρω από την υποψήφια γωνία. Τα εικονοστοιχεία κατατάσσονται σε σκοτεινότερα, παρόμοια ή φωτεινότερα υποσύνολα σε σχέση με το κεντρικό εικονοστοιχείο. Στη συνέχεια με τη βοήθεια ενός αλγορίθμου δημιουργείται ένα δέντρο απόφασης από τα υποσύνολα, το οποίο χρησιμοποιείται σαν ανιχνευτής γωνιών. Ανιχνεύονται ως τοπικά χαρακτηριστικά τα εικονοστοιχεία εκείνα που η γειτονιά τους περιέχει  $n$  φωτεινότερα ή σκοτεινότερα εικονοστοιχεία. Το κριτήριο αυτό αποτελεί το κριτήριο FAST. Σαν τελευταίο βήμα υπολογίζονται τα τοπικά μέγιστα του αθροίσματος της απόλυτης διαφοράς μεταξύ των εικονοστοιχείων της γειτονιάς και του κεντρικού εικονοστοιχείου. Το αποτέλεσμα της υλοποίησης αυτής είναι ένας πολύ αποδοτικός ανιχνευτής χαρακτηριστικών, με το μειονέκτημα όμως ότι δεν είναι αναλλοίωτος στις μεταβολές κλίμακας (scale-invariant) των εικόνων.

## 2.6 BRISK (Binary Robust Invariant Scalable Keypoints)

Η μέθοδος ανίχνευσης και περιγραφής τοπικών χαρακτηριστικών BRISK (Binary Robust Invariant Scalable Keypoints) χρησιμοποιεί το κριτήριο FAST για να ανιχνεύσει τις περιοχές τοπικών χαρακτηριστικών στο χώρο της κλίμακας, ο οποίος αποτελείται από  $n$  οκτάβες και  $n$  εσωτερικές οκτάβες. Η περιγραφή των χαρακτηριστικών γίνεται με τη χρήση μια δυαδικής συμβολοσειράς, η οποία περιέχει τις τιμές των συγκριτικών δοκιμών φωτεινότητας. Τα τοπικά χαρακτηριστικά BRISK θα παρουσιαστούν αναλυτικότερα στην Ενότητα 5.2.



## Κεφάλαιο 3

# Προβολική Γεωμετρία και Όραση Πολλαπλών Όψεων

### 3.1 Εισαγωγή

Το κεφάλαιο αυτό παρέχει το θεωρητικό υπόβαθρο για τη μοντελοποίηση πολλαπλών όψεων μιας σκηνής στα πλαίσια της προβολικής γεωμετρίας, το μοντέλο της κάμερας μικροσκοπικής οπής (pinhole model) και τα συστήματα συντεταγμένων που ορίζουν μια γραμμική προβολική κάμερα. Η δομή που ακολουθείται στηρίχθηκε στο βιβλίο *Image Processing, Analysis and Machine Vision* [20]. Τέλος αναλύεται η Επιπολική γεωμετρία η οποία περιγράφει την εσωτερική προβολική γεωμετρία δύο εικόνων, ακολουθώντας τη δομή του βιβλίου *Multiple View Geometry* [21]. Οι συγκεκριμένες ιδέες που παρουσιάζονται στο κεφάλαιο αυτό χρησιμοποιούνται κατά κόρον σε εφαρμογές Δομής από Κίνηση (Structure from Motion) καθώς και σε οπτικά συστήματα αυτοεπιτοπισμού και χαρτογράφησης (Simultaneous localization and mapping) [22].

### 3.2 Προβολική Γεωμετρία

Ο τρισδιάστατος κόσμος περιγράφεται με μεγάλη ακρίβεια από την Ευκλείδεια γεωμετρία, σύμφωνα με την οποία το σχήμα, το μήκος των ευθυγράμμων τμημάτων, γωνίες, αποστάσεις και λόγοι μηκών ευθυγράμμων τμημάτων διατηρούνται όταν εφαρμόζονται μετασχηματισμοί (μετάτωση και περιστροφή). Παρόλα αυτά, καθώς τα γνωστά μήκη, τα σχήματα και οι γωνίες δεν διατηρούνται στο επίπεδο της εικόνας η χρήση της Ευκλείδειας γεωμετρίας δεν κρίνεται βολική στην Όραση Υπολογιστών.

Η προβολική γεωμετρία, σαν επέκταση της Ευκλείδειας, έρχεται να καλύψει την αναποτελεσμα-

τικότητα της στο επίπεδο της εικόνας. Κατά την Ευκλείδεια γεωμετρία δύο παράλληλες ευθείες τέμνονται στο άπειρο, το οποίο όμως ως σημείο στην προβολική γεωμετρία είναι καθορισμένο και καλείται ιδανικό σημείο. Η προσθήκη λοιπόν ιδανικών σημείων μετασχηματίζει τον Ευκλείδειο χώρο στον προβολικό. Η κύρια ιδιότητα ενός προβολικού μετασχηματισμού είναι ότι διατηρεί την ευθύτητα των σημείων. Δηλαδή, συνευθειακά σημεία θα μετασχηματιστούν σε συνευθειακά. Ταυτόχρονα όμως δεν διατηρούνται το σχήμα, το μήκος των ευθυγράμμων τμημάτων, οι γωνίες, οι αποστάσεις και οι λόγοι μηκών των ευθυγράμμων τμημάτων.

Σχεδόν σε όλες τις εφαρμογές της Όρασης Υπολογιστών η προοπτική προβολή ενός 3Δ σημείου στο χώρο στο 2Δ επίπεδο της εικόνας πραγματοποιείται χρησιμοποιώντας την προβολική γεωμετρία και το μοντέλο κάμερας μικροσκοπικής οπής. Το επίπεδο της εικόνας αναπαριστάται από το προβολικό επίπεδο  $\mathcal{P}^2$ , ορισμένο ως το πραγματικό Ευκλείδειο επίπεδο  $\mathcal{R}^2$  επαυξημένο με μία γραμμή στο άπειρο. Ο 3Δ πραγματικός χώρος αναπαριστάται από το 3Δ προβολικό χώρο  $\mathcal{P}^3$ , ορισμένο ως τον πραγματικό Ευκλείδειο χώρο  $\mathcal{R}^3$  επαυξημένο με ένα επίπεδο στο άπειρο.

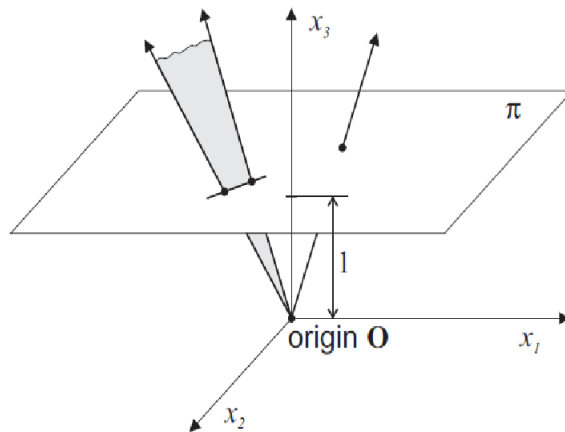
Τα σημεία στον προβολικό χώρο παρουσιάζονται στη μορφή των ομογενών συντεταγμένων, δηλαδή είναι ορισμένα ως προς ένα μη μηδενικό παράγοντα κλίμακας. Το σημείο  $\mathbf{m}_{\mathbb{P}^2} \simeq (x \ y \ 1)^T$ , όπου το  $\simeq$  συμβολίζει μέγεθος ως προς ένα παράγοντα κλίμακας, συνδέεται με όλα τα βαθμωτά πολλαπλάσια του ευκλείδειου διανύσματος  $\mathbf{m}_{\mathbb{R}^3} = (x \ y \ z)^T$  τα οποία τέμνουν το προβολικό επίπεδο σε ένα μοναδικό σημείο. Τα σημεία και οι γραμμές στο προβολικό επίπεδο αντιστοιχούν σε ακτίνες και επίπεδα στον Ευκλείδειο χώρο τα οποία τέμνουν την αρχή των αξόνων (Σχήμα 3.1). Το ίδιο εφαρμόζεται και στα σημεία του προβολικού χώρου  $\mathcal{P}^3$ : ένα σημείο του ευκλείδειου χώρου γράφεται σε ομογενείς συντεταγμένες ως  $\mathbf{M}_{\mathbb{P}^3} \simeq (X \ Y \ Z \ 1)^T$ .

Επομένως ένα 2Δ ή 3Δ σημείο σε Ευκλείδειες συντεταγμένες παρουσιάζεται στην προβολική γεωμετρία ως εξής:

$$\mathbf{m}_{\mathbb{R}^2} = \begin{bmatrix} x \\ y \end{bmatrix} \rightarrow \mathbf{m}_{\mathbb{P}^2} \simeq \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad \mathbf{M}_{\mathbb{R}^3} = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \rightarrow \mathbf{M}_{\mathbb{P}^3} \simeq \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

Οποιοδήποτε σημείο έχει την τελευταία του συντεταγμένη μηδενική είναι σημείο στο άπειρο στον προβολικό χώρο και, όπως αναφέρθηκε πιο πάνω, καλείται ιδανικό σημείο. Η αρχή των αξόνων στην προβολική γεωμετρία είναι  $(0 \ 0 \ 1)^T \in \mathcal{R}^2$  και  $(0 \ 0 \ 0 \ 1)^T \in \mathcal{P}^3$ .

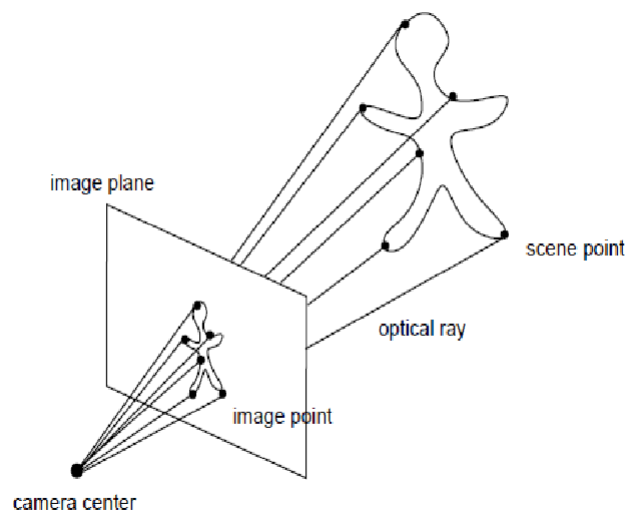




Σχήμα 3.1: Το προβολικό επίπεδο ενσωματωμένο στον Ευκλείδειο χώρο

### 3.3 Μοντέλο Κάμερας Μικροσκοπικής Οπής

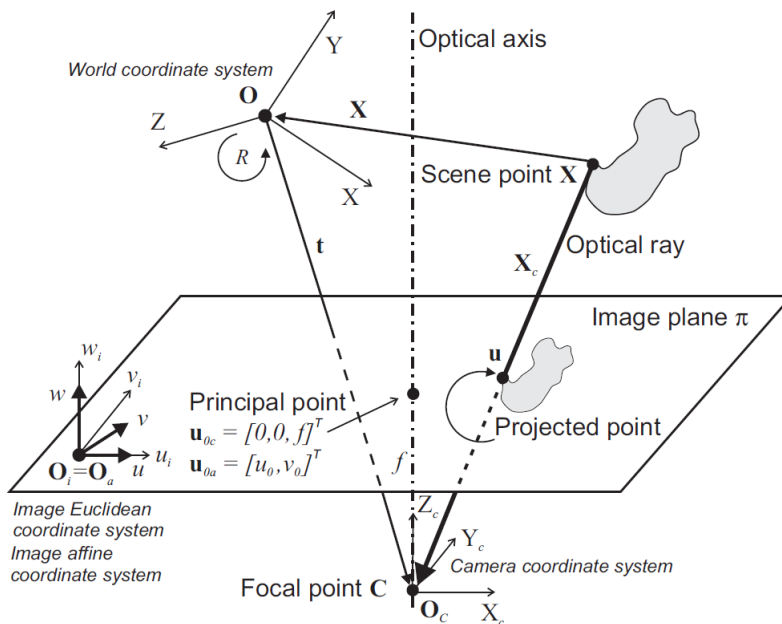
Το μοντέλο κάμερας μικροσκοπικής οπής χρησιμοποιείται ευρέως σε εφαρμογές όρασης υπολογιστών. Θεωρεί μια οπτική ακτίνα η οποία πηγάζει από το οπτικό κέντρο της κάμερας και διαπερνά το επίπεδο της εικόνας τέμνοντας ένα 3Δ σημείο του χώρου. Το σημείο τομής της οπτικής ακτίνας και του επιπέδου της εικόνας αποτελεί την 2Δ προβολή του 3Δ σημείου στο χώρο (Σχήμα 3.2).



Σχήμα 3.2: Μοντέλο Κάμερας Μικροσκοπικής Οπής-Το σημείο τομής της οπτικής ακτίνας από το κέντρο της κάμερας αποτελεί τη 2Δ προβολή του 3Δ σημείου στο επίπεδο της εικόνας

Όπως αναφέρθηκε, ο πραγματικός κόσμος προβάλλεται πάνω στο επίπεδο της εικόνας  $\pi$ . Το σημείο  $C$  αποτελεί το κέντρο προβολής και η ευθεία που περνάει από το  $C$  και τέμνει κάθετα το επίπεδο της εικόνας αποτελεί τον οπτικό άξονα. Οι οπτικοί φακοί της κάμερας αποτελούν τους άξονες του οπτικού κέντρου  $C$  και η απόσταση μέχρι το επίπεδο της εικόνας αποτελεί την παράμετρο μήκους εστίασης  $f$ . Με βάση την προβολική γεωμετρία στην ουσία η κάμερα πραγματοποιεί ένα γραμμικό μετασχηματισμό από τον 3D προβολικό χώρο  $\mathcal{P}^3$  στο 2D προβολικό επίπεδο  $\mathcal{P}^2$  με διαδοχικούς μετασχηματισμούς μεταξύ των 4 συστημάτων συντεταγμένων που διακρίνονται στο Σχήμα 3.3 και τα οποία είναι:

1. Το Ευκλείδειο σύστημα συντεταγμένων του κόσμου με αρχή των αξόνων το  $O$ .
2. Το Ευκλείδειο σύστημα συντεταγμένων της κάμερας με αρχή των αξόνων το οπτικό κέντρο προβολής  $C \equiv O_c$ .
3. Το Ευκλείδειο σύστημα συντεταγμένων της εικόνας.
4. Το αφινικό σύστημα συντεταγμένων της εικόνας το οποίο ταυτίζεται ως προς την αρχή των αξόνων με το Ευκλείδειο σύστημα συντεταγμένων της κάμερας αλλά ενδέχεται να διαφέρει ως προς τους άξονές του λόγω κατασκευαστικών ατελειών της κάμερας.



Σχήμα 3.3: Γεωμετρία Κάμερας Μικροσκοπικής Οπής

Για να λάβουμε τον τελικό προβολικό μετασχηματισμό έχουμε τους εξής 3 διαδοχικούς μετασχηματισμούς μεταξύ των παραπάνω συστημάτων:

- Μετασχηματισμός από το σύστημα συντεταγμένων του κόσμου  $O$  στο σύστημα συντεταγμένων της κάμερας  $O_c$  ( $1 \rightarrow 2$ ). Αποτελείται από μετατόπιση κατά  $\mathbf{t}$  και ακολούθως περιστροφή των αξόνων κατά  $\mathbf{R}$ . Επομένως, ένα σημείο στον πραγματικό κόσμο μετασχηματίζεται ως εξής (σε ομογενείς συντεταγμένες):

$$X_c = \begin{bmatrix} R & -Rt \\ 0^T & 1 \end{bmatrix} X \quad (3.1)$$

Οι παράμετροι  $\mathbf{t}$  και  $\mathbf{R}$  καλούνται *εξωτερικές παράμετροι της κάμερας (extrinsic parameters)*.

- Μετασχηματισμός προβολής του 3D σημείου  $X_c$  στο Ευκλείδειο σύστημα συντεταγμένων της εικόνας  $O_i$  ( $2 \rightarrow 3$ ). Για λόγους ευκολίας στο συγκεκριμένο μετασχηματισμό θέτουμε το μήκος εστίασης  $f = 1$  και επομένως σε ομογενείς συντεταγμένες:

$$u_i = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} X_c \simeq \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} X_c \quad (3.2)$$

- Μετασχηματισμός από το Ευκλείδειο σύστημα της εικόνας  $O_i$  στο αφινικό σύστημα της εικόνας ( $3 \rightarrow 4$ ) το οποίο περιέχει όλες τις εσωτερικές παραμέτρους της κάμερας:

$$u \simeq K u_i = \begin{bmatrix} f & s & -u_0 \\ 0 & g & -v_0 \\ 0 & 0 & 1 \end{bmatrix} u_i \quad (3.3)$$

Καλούμε τον πίνακα  $K$  ο οποίος περιέχει τις εσωτερικές παραμέτρους αυτές, πίνακα *εσωτερικών παραμέτρων κάμερας (intrinsic camera matrix)*.

Ο τελικός προβολικός μετασχηματισμός  $\mathcal{P}^3 \rightarrow \mathcal{P}^2$  της κάμερας μικροσκοπικής οπής προκύπτει από τις εξισώσεις 3.1, 3.2, 3.3 και είναι:

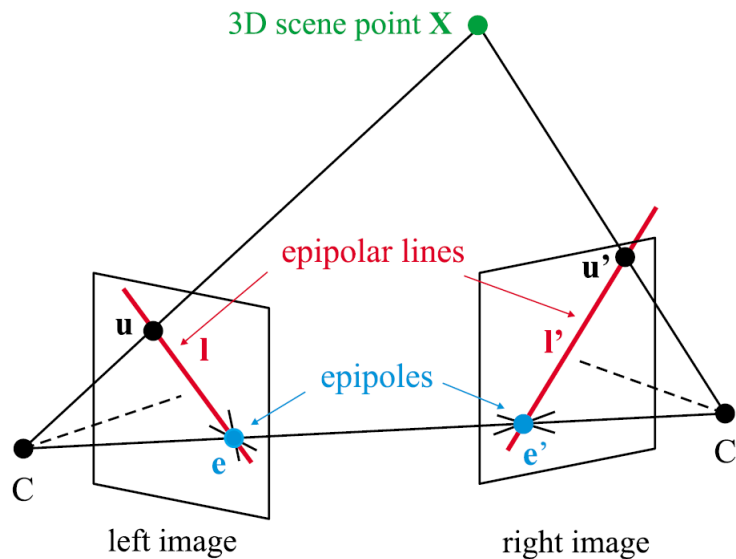
$$u \simeq K \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & -Rt \\ 0^T & 1 \end{bmatrix} X = K [R \mid -Rt] X = PX \quad (3.4)$$

Καλούμε τον πίνακα  $P$  πίνακα *προβολής (projection matrix)*. Ο πίνακας αυτός περιέχει όλες τις εσωτερικές και εξωτερικές παραμέτρους της κάμερας.

### 3.4 Επιπολική Γεωμετρία

Η επιπολική γεωμετρία περιγράφει την εσωτερική προβολική γεωμετρία μεταξύ δύο εικόνων. Εξαρτάται μόνο από τις εσωτερικές παραμέτρους της κάμερας και τη σχετική θέση των καμερών ενώ ταυτόχρονα είναι ανεξάρτητη από το σύστημα συντεταγμένων του κόσμου. Με βάση της ιδέες της επιπολικής γεωμετρίας μπορούμε να εξάγουμε πληροφορίες για τη θέση της κάμερας ή την 3Δ γεωμετρία του χώρου από ένα σύνολο επικαλυπτόμενων εικόνων μίας σκηνής.

Παρατηρώντας ένα 3Δ σημείο  $X$  στο χώρο, από δύο διαφορετικές όψεις, η προβολή του απεικονίζεται στο  $u$  στο επίπεδο εικόνας της 1<sup>ης</sup> κάμερας και στο σημείο  $u'$  στο επίπεδο εικόνας της 2<sup>ης</sup> κάμερας, όπως παρουσιάζεται στο Σχήμα 3.4. Τα σημεία των εικόνων  $u, u'$ , τα κέντρα των καμερών  $C, C'$  καθώς και το σημείο στο χώρο ανήκουν στο ίδιο επίπεδο  $\pi$ . Έστω ότι είναι γνωστό μόνο το σημείο  $u$ , το επίπεδο  $\pi$  ορίζεται από μια ευθεία που ενώνει τα κέντρα των καμερών (βασική ευθεία) και την οπτική ακτίνα που διέρχεται από το σημείο  $u$ . Επομένως, μπορούμε να πούμε ότι η ευθεία που αντιστοιχεί στο άγνωστο σημείο  $u'$  ανήκει στο επίπεδο  $\pi$ . Άρα το σημείο  $u'$  ανήκει στην ευθεία  $l'$ , η οποία είναι η τομή του επιπέδου  $\pi$  με το επίπεδο της δεύτερης εικόνας. Με βάση την επιπολική γεωμετρία λοιπόν μπορούμε να πούμε ότι θα αναζητήσουμε το άγνωστο σημείο  $u'$  στην ευθεία  $l'$ .



Σχήμα 3.4: Η επιπολική γεωμετρία και η γεωμετρική αναπαράσταση του επιπολικού περιορισμού.

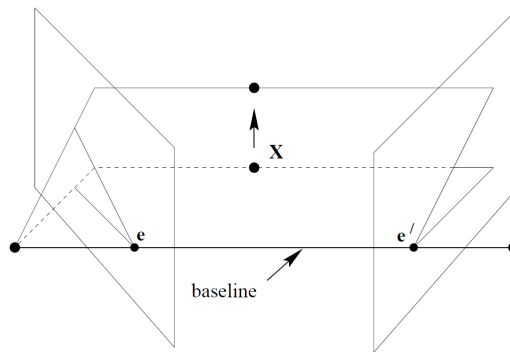
Έχοντας υπόψιν τα παραπάνω ισχύουν οι εξής ορισμοί:

**Ορισμός 3.1.** *Επιπολικό σημείο* καλείται το σημείο τομής της ευθείας που ενώνει τα κέντρα των καμερών (βασική ευθεία) με το επίπεδο της εικόνας. Ισοδύναμα, το επιπολικό σημείο είναι η απει-

κόνιση σε μια εικόνα του κέντρου της κάμερας της άλλης εικόνας.

**Ορισμός 3.2.** *Επιπολικό επίπεδο* είναι το επίπεδο που περιέχει τη βασική ευθεία. Υπάρχει μία μονοπαραμετρική οικογένεια από επιπολικά επίπεδα (“*epipolar pencil*”). Όπως φαίνεται στο Σχήμα 3.5 καθώς η θέση του 3D σημείου  $X$  μεταβάλλεται, τα επιπολικά επίπεδα περιστρέφονται γύρω από την βασική ευθεία δημιουργώντας τη μονοπαραμετρική οικογένεια επιπολικών επιπέδων.

**Ορισμός 3.3.** *Η επιπολική ευθεία* είναι η τομή του επιπολικού επιπέδου με το επίπεδο της εικόνας. Όλες οι επιπολικές ευθείες τέμνονται στο επιπολικό σημείο. Ένα επιπολικό επίπεδο τέμνει το αριστερό και δεξί επίπεδο της εικόνας στις επιπολικές ευθείες, και ορίζει τις αντιστοιχίες μεταξύ των ευθειών.



Σχήμα 3.5: Καθώς η θέση του 3D σημείου  $X$  μεταβάλλεται, τα επιπολικά επίπεδα περιστρέφονται γύρω από την βασική ευθεία δημιουργώντας τη μονοπαραμετρική οικογένεια επιπολικών επιπέδων (“*epipolar pencil*”).

### 3.4.1 Ο Θεμελιώδης Πίνακας (Fundamental Matrix) $F$

Ο Θεμελιώδης πίνακας  $F$  ενσωματώνει την εσωτερική γεωμετρία των εικόνων και αποτελεί την αλγεβρική αναπαράσταση της επιπολικής γεωμετρίας.

Με βάση τους ορισμούς 3.1, 3.2, 3.3 για κάθε σημείο  $u$  στο επίπεδο εικόνας της  $1^{ης}$  κάμερας αντιστοιχεί μία επιπολική ευθεία  $l'$  στο επίπεδο εικόνας της  $2^{ης}$  κάμερας (Σχήμα 3.4). Επίσης το σημείο  $u'$  το οποίο αντιστοιχεί στο  $u$  πρέπει να ανήκει σε αυτή την επιπολική ευθεία  $l'$ . Αντίστοιχα το ίδιο πρέπει να ισχύει για το σημείο  $u$  και την επιπολική ευθεία  $l$ , καθώς έχουμε συμμετρική σχέση των δύο εικόνων. Το γεγονός ότι η θέση των σημείων  $u$  και του σημείου  $u'$  που αντιστοιχεί σε αυτό στην δεύτερη εικόνα δεν είναι τυχαία καλείται *επιπολικός περιορισμός* (*epipolar constraint*). Ισχύουν λοιπόν:

$$l'^T u' = 0 \qquad l^T u = 0 \qquad (3.5)$$

$$e = PC' \quad e' = P'C \quad (3.6)$$

Από την Ενότητα 3.3 είναι γνωστό ότι η προβολή στο επίπεδο της εικόνας της  $1^{ηs}$  κάμερας ενός σημείου  $X$  στο 3Δ χώρο δίνεται από την σχέση  $u = PX$  (Εξ. 3.4). Αντίστροφα ισχύει  $X = P^\dagger u$  όπου ο πίνακας  $P^\dagger = P^T(PP^\dagger)^{-1}$  αποτελεί τον ψευδο-αντίστροφο πίνακα προβολής  $P$  της κάμερας. Η επιπολική ευθεία  $l'$  περνάει από το επιπολικό σημείο  $e'$  και το σημείο  $u' = P'X$  και επομένως ορίζεται ως εξής:

$$l' = e' \times u' \quad (3.7)$$

Αντικαθιστώντας τα σημεία  $e'$  και  $u'$  έχουμε:

$$l' = e' \times (P'P^\dagger u) \quad (3.8)$$

Επομένως η επιπολική ευθεία  $l'$  αποτελεί μια γραμμική αντιστοίχιση του σημείου  $u$  και ο πίνακας που αντιπροσωπεύει αυτή τη γραμμική αντιστοίχιση αποτελεί τον *Θεμελιώδη πίνακα*  $F$ :

$$F = e' \times P'P^\dagger \leftrightarrow l' = Fu \quad (3.9)$$

Αντίστοιχα η γραμμική αντιστοίχιση σημείων και στις δύο εικόνες εξάγεται από τη σχέση  $l'^T u' = 0$  και την Εξ.3.9 και ορίζει το Θεμελιώδη πίνακα ως εξής:

**Ορισμός 3.4.** Έστω δύο εικόνες από δύο κάμερες με διαφορετικά κέντρα, τότε ο Θεμελιώδης πίνακας  $F$  είναι ο μοναδικός  $3 \times 3$  ομογενής πίνακας βαθμού 2 που ικανοποιεί τη σχέση:

$$u'^T F u = 0 \quad (3.10)$$

για όλα τα ζεύγη αντίστοιχων σημείων  $u \leftrightarrow u'$

Η παραπάνω σχέση αποτελεί τον επιπολικό περιορισμό του Θεμελιώδους πίνακα σε αλγεβρική μορφή.

### 3.4.2 Ο Βασικός Πίνακας (Essential matrix) $E$

Μια ειδική περίπτωση του Θεμελιώδους πίνακα  $F$  αποτελεί ο Βασικός πίνακας  $E$ . Ο πίνακας αυτός ορίζεται ως προς κανονικοποιημένα εικονοστοιχεία (pixels) στην εικόνα. Έστω ένα ζεύγος καμερών για τις οποίες θεωρούμε ότι το σύστημα συντεταγμένων του κόσμου ταυτίζεται με το σύστημα συντεταγμένων της  $1^{ηs}$  κάμερας, και η  $2^{η}$  κάμερα έχει κινηθεί σε σχέση με την  $1^{η}$ . Τότε οι πίνακες προβολής των δύο καμερών θα είναι της μορφής:

$$P = K \begin{bmatrix} I & 0 \end{bmatrix} \quad P' = K' \begin{bmatrix} R & t \end{bmatrix} \quad (3.11)$$

Από την εξίσωση 3.9 και γνωρίζοντας ότι η θέση του επιπολικού σημείου είναι  $e' = P'C$  τότε:

$$P^\dagger = K \begin{pmatrix} K^{-1} \\ 0^T \end{pmatrix} \quad C = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad (3.12)$$

$$\begin{aligned}
(3.9) \Rightarrow F &= [P'C]_{\times} P' P^{\dagger} \\
&= [K'T]_{\times} K' R K^{-1} = K'^{-T} [t]_{\times} R K^{-1} \\
&= K'^{-T} R [R^T t]_{\times} K^{-1} = K'^{-T} R K^T [K R^T t]_{\times}
\end{aligned} \tag{3.13}$$

Αν ο πίνακας βαθμονόμησης  $K$  της κάμερας είναι γνωστός τότε προκύπτουν τα κανονικοποιημένα σημεία της εικόνας  $\hat{u} = K^{-1}u$  και ο κανονικοποιημένος πίνακας προβολής  $K^{-1}P = [R \mid t]$ . Ο Θεμελιώδης πίνακας που αντιστοιχεί στο ζεύγος των κανονικοποιημένων καμερών καλείται *Βασικός πίνακας* και προκύπτει από την Εξ. 3.13 ως εξής:

$$E = [t]_{\times} R = R [R^T t]_{\times} \tag{3.14}$$

**Ορισμός 3.5.** Ο Βασικός πίνακας  $E$  ως προς τα κανονικοποιημένα ζεύγη σημείων  $\hat{u} \leftrightarrow \hat{u}'$  δίνεται από την εξίσωση:

$$\hat{u}'^T E \hat{u} = 0 \tag{3.15}$$

Αντικαθιστώντας τα κανονικοποιημένα σημεία στην Εξ. 3.15 και συγκρίνοντας τη σχέση αυτή με την Εξ. 3.10 του  $F$ , προκύπτει η σχέση που συνδέει τους δύο πίνακες:

$$E = K'^T F K \tag{3.16}$$

Ο  $E$  είναι ένας ομογενής πίνακας, με 5 βαθμούς ελευθερίας. Αποδεικνύεται ότι ένας πίνακας  $3 \times 3$  είναι Βασικός πίνακας αν και μόνο αν 2 ιδιόμορφες τιμές του είναι ίσες και η τρίτη είναι μοναδική.

### 3.4.3 Υπολογισμός Σχετικής Κίνησης της Κάμερας από το Βασικό πίνακα $E$

Μπορούμε να υπολογίσουμε το Βασικό πίνακα  $E$  είτε από την Εξ. 3.15 χρησιμοποιώντας κανονικοποιημένα σημεία, είτε από την Εξ. 3.16 αφού έχει υπολογιστεί ο Θεμελιώδης πίνακας. Με γνωστό τον πίνακα  $E$  μπορούμε να εξάγουμε τους πίνακες προβολής των δύο καμερών.

Υποθέτοντας ότι ο πίνακας προβολής της 1<sup>ης</sup> κάμερας είναι  $P = [I \mid 0]$ , για να υπολογίσουμε τον πίνακα προβολής της 2<sup>ης</sup>  $P'$  είναι απαραίτητο να εκφράσουμε το Βασικό πίνακα  $E$  σαν παράγοντα ενός αντισυμμετρικού πίνακα  $S$  (skew-symmetric) και ενός πίνακα περιστροφής  $R$ . Επομένως  $E = SR$ . Αποσυνθέτωντας σε ιδιόμορφες τιμές (Singular Value Decomposition) τον πίνακα  $E$  γνωρίζουμε ότι θα είναι της μορφής  $E = U \text{diag}(1, 1, 0) V^T$ . Ορίζουμε τους πίνακες  $W$  και ως εξής:

$$W = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad Z = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{3.17}$$

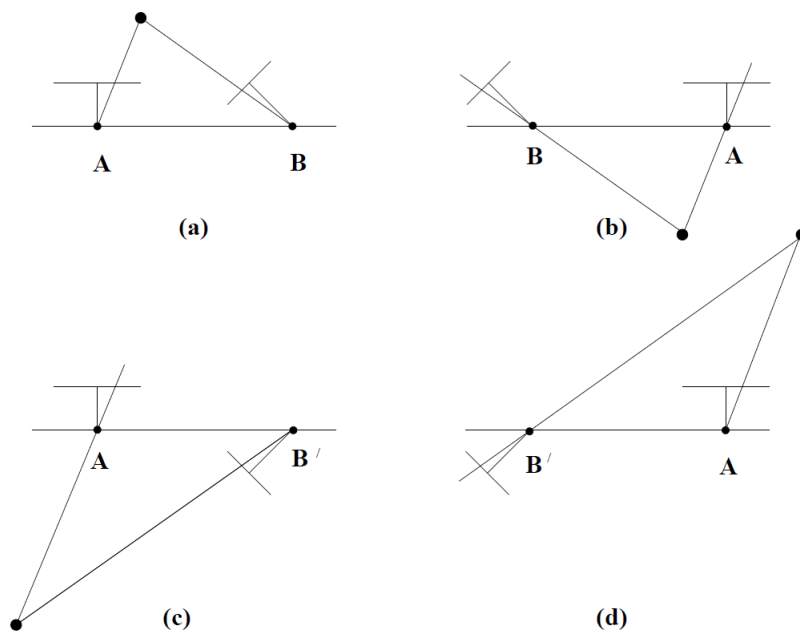
Μπορούμε να γράψουμε τον αντισυμμετρικό πίνακα  $S = UZU^T$  και το πίνακα περιστροφής  $R = UXV^T$  όπου  $X$  κάποιος πίνακας περιστροφής. Επομένως:

$$U \text{diag}(1, 1, 0)V^T = E = SR = (UZU^T)(UXV^T) \quad (3.18)$$

Από τις Εξ. 3.17, 3.18 έχουμε ότι  $ZX = \text{diag}(1, 1, 0)$  και επομένως καθορίζουμε τον πίνακα περιστροφής  $X = W$  ή  $X = W^T$ . Για τον αντισυμμετρικό πίνακα  $S$  έχουμε την εξής σχέση:  $S = [t]_{\times} = UZU^T$ . Επίσης ισχύει ότι  $St = 0$  και επομένως  $t = U(0, 0, 1)^T = u_3$ , δηλαδή το διάνυσμα μετατόπισης αντιστοιχεί στην τελευταία στήλη του πίνακα  $U$ . Καθώς το πρόσημο του Βασικού πίνακα  $E$  δεν μπορεί να καθορισθεί, δεν μπορούμε να καθορίσουμε ούτε το πρόσημο του διανύσματος μετατόπισης. Επομένως υποθέτοντας ότι ο πίνακας προβολής της 1<sup>ης</sup> κάμερας είναι  $P = [I \mid 0]$ , και με γνωστό τον πίνακα  $E$  βρίσκουμε 4 πιθανές λύσεις για τον πίνακα προβολής  $P' = [R \mid t]$  της 2<sup>ης</sup> κάμερας:

$$P' = [R \mid t] = [UWV^T \mid +u_3], [UWV^T \mid -u_3], [UW^TV^T \mid +u_3], [UW^TV^T \mid -u_3] \quad (3.19)$$

Γεωμετρικά οι 4 αυτές λύσεις παρουσιάζονται στο Σχήμα 3.6. Σαν σωστή λύση επιλέγεται εκείνη η οποία δίνει το μεγαλύτερο αριθμό ανακατασκευασμένων σημείων στον 3Δ χώρο μπροστά από τις κάμερες.



Σχήμα 3.6: Από τις 4 δυνατές λύσεις για τον πίνακα προβολής  $P'$  μόνο η λύση (a) μπορεί να επιλεγεί, καθώς δίνει το ανακατασκευασμένο σημείο του χώρου μπροστά από τις κάμερες.



## Κεφάλαιο 4

# Αλγόριθμοι εκτίμησης θέσης της κάμερας

### 4.1 Εισαγωγή

Στο κεφάλαιο αυτό παρουσιάζονται τα μαθηματικά εργαλεία και αλγόριθμοι που χρησιμοποιούνται για εκτίμηση θέσης της κάμερας. Αρχικά αναφέρεται η ευρέως διαδεδομένη μέθοδος εύρωστης εκτίμησης RANSAC, η οποία δίνει τη δυνατότητα εκτίμησης ενός μοντέλου από θορυβώδεις μετρήσεις. Στη συνέχεια παρουσιάζεται το σφάλμα Sampson υπολογισμένο για το Θεμελιώδη πίνακα, το οποίο γεωμετρικά είναι ένα μέτρο απόστασης του σημείου  $u'$  από την επιπολική ευθεία  $l' = Fu'$ . Τέλος παρουσιάζονται δύο αλγόριθμοι για τον υπολογισμό του Θεμελιώδους πίνακα  $F$  και την εξαγωγή του πίνακα εξωτερικών παραμέτρων από το Βασικό πίνακα  $E$ .

### 4.2 Μέθοδος εύρωστης εκτίμησης RANSAC

Αν και η μέθοδος RANSAC (*RANdom SAmple Consensus*) είναι μια γενική μέθοδος για εύρωστη εκτίμηση [23], έχει αναπτυχθεί κυρίως μέσα στα πλαίσια του προβλήματος εκτίμησης θέσης της κάμερας. Η μέθοδος αυτή είναι πολύ εύκολα υλοποιήσιμη και έχει το πλεονέκτημα ότι δεν χρειάζεται αρχική εκτίμηση του ζητούμενου μοντέλου. Από ένα σύνολο δεδομένων η μέθοδος RANSAC επιλέγει τυχαία το μικρότερο δυνατό υποσύνολο που είναι απαραίτητο για να υπολογιστεί το ζητούμενο μοντέλο. Με αυτό τον τρόπο μεγιστοποιείται η πιθανότητα ότι ένα τουλάχιστον υποσύνολο δεν θα περιέχει σφάλματα και επομένως θα δώσει σωστή εκτίμηση (Αλγόριθμος 1). Συγκεκριμένα, στις εφαρμογές Επαυξημένης πραγματικότητας ζητούμενο είναι η εκτίμηση των παραμέτρων  $p$  του πίνακα προβολής  $P$  από ένα σύνολο δεδομένων  $S$ , από τα οποία κάποια περιέ-

χουν σφάλματα. Υποθέτοντας ότι για να εκτιμηθεί η θέση της κάμερας χρειάζονται τουλάχιστον  $n$  δεδομένα, τότε επιλέγονται τυχαία  $N$  δείγματα  $n$  στοιχείων. Με κάθε δείγμα εκτιμώνται οι παράμετροι  $p_i$  του πίνακα προβολής και υπολογίζονται τα υποσύνολα  $S_i \subseteq S$  για τα οποία ισχύει η εκτίμηση αυτή. Το μεγαλύτερο υποσύνολο  $S_i$  αποτελεί την καλύτερη εκτίμηση των παραμέτρων. Ακολουθώντας η λύση εκτιμάται ξανά χρησιμοποιώντας όλα τα δεδομένα του υποσυνόλου  $S_i$ , και το σύστημα λύνεται με την μέθοδο ελαχίστων τετραγώνων. Τα δεδομένα του υποσυνόλου  $S_i$  αποτελούν τις επιλεγμένες τιμές της εκτίμησης (*inliers*), ενώ όλα τα υπόλοιπα τις ακραίες τιμές (*outliers*).

Η παράμετρος  $N$  της μεθόδου, αποτελεί βασικά τον αριθμό προσπαθειών να βρεθεί η καλύτερη εκτίμηση και επιλέγεται ούτως ώστε να είναι:

$$N \geq \frac{\log(1-p)}{1-w^n} \quad (4.1)$$

όπου  $p$  είναι η πιθανότητα τουλάχιστον ένα από τα δείγματα να μην περιέχει σφάλματα και  $w$  είναι η αναλογία των *inliers* ως προς τα *outliers*.

---

#### Algorithm 1 Αλγόριθμος RANSAC

---

- 1: Διάλεξε τυχαία ένα δείγμα  $n$  πλήθους από το σύνολο  $S$  των δεδομένων και υπολόγισε το μοντέλο από τα δείγματα αυτά.
- 2: Καθόρισε το σύνολο δεδομένων  $S_i$  για τα οποία το μοντέλο που υπολογίστηκε απέχει απόσταση  $\leq t$ . Το σύνολο  $S_i$  αποτελεί το σύνολο των *inliers* για το μοντέλο αυτό.
- 3: Αν το πλήθος των *inliers* είναι  $\geq T$  υπολόγισε το μοντέλο χρησιμοποιώντας όλα τα σημεία του συνόλου  $S_i$  και τερμάτισε.
- 4: Αν το πλήθος των *inliers* είναι  $< T$  διάλεξε ένα νέο υποσύνολο και επανέλαβε.
- 5: Μετά από  $N$  επαναλήψεις επιλέγεται το μεγαλύτερο σε πλήθος  $S_i$  και το μοντέλο επανυπολογίζεται με χρήση όλων των δεδομένων του  $S_i$ .

όπου τα κατώφλια  $t$  και  $T$  επιλέγονται από τον χρήστη.

---

### 4.3 Σφάλμα Sampson Θεμελιώδους πίνακα $F$

Έχοντας σαν δεδομένο τα ταυτισμένα τοπικά χαρακτηριστικά  $u \leftrightarrow u'$  δύο διαδοχικών εικόνων στο προβολικό επίπεδο (ομογενείς συντεταγμένες), το σφάλμα Sampson [24] γεωμετρικά είναι ένα μέτρο απόστασης του σημείου  $u'$  από την επιπολική ευθεία  $l' = Fu$ .

Έστω  $\mathcal{V}_{\mathcal{F}}$  ο χώρος του ζεύγους  $(u, u')$  έτσι ώστε να ισχύει ο επιπολικός περιορισμός του Θεμελιώδους πίνακα  $u'^T Fu = 0$  και ο τελεστής  $C_F(X) = u'^T Fu$  όπου  $X = (u, u') = (x, y, x', y')$ . Για μη θορυβώδεις μετρήσεις των εικονοστοιχείων  $u \leftrightarrow u'$  τότε  $(u, u') \in \mathcal{V}_{\mathcal{F}}$ . Αντίστοιχα για θορυβώδεις μετρήσεις  $(u, u') \notin \mathcal{V}_{\mathcal{F}}$  και το σφάλμα Sampson δίνει την απόσταση από το χώρο  $\mathcal{V}_{\mathcal{F}}$ .

Χρησιμοποιώντας προσέγγιση πρώτων όρων:

$$C_F(\hat{X}) = C_F(X + \delta X) = C_F + \frac{\partial C_F(X)}{\partial X} \delta X \quad (4.2)$$

Αν  $\hat{X} \in V_F \Rightarrow C_F(\hat{X}) = 0$  και επομένως:

$$C_F + \frac{\partial C_F(X)}{\partial X} \delta X = 0 \quad (4.3)$$

Η πιο πάνω εξίσωση μπορεί να γραφτεί, ορίζοντας τον πίνακα μερικών παραγόντων  $J = \frac{\partial C_F(X)}{\partial X}$  και τη συνάρτηση κόστους  $e = C_F(X)$ , ως εξής:

$$J\delta X = -e \quad (4.4)$$

Ζητούμενο είναι η ελαχιστοποίηση του  $\| \delta X \|$  υπό τον περιορισμό της Εξ. 4.4. Ορίζοντας τη συνάρτηση με Laplace πολλαπλασιαστές έχουμε:

$$\Psi = \delta X^T \delta X - 2\lambda(J\delta X + e) \quad (4.5)$$

Για την εύρεση του ελαχίστου της παραπάνω συνάρτησης θέτουμε  $\frac{\partial \Psi}{\partial \delta X} = 0$  και  $\frac{\partial \Psi}{\partial \lambda} = 0$ . Επομένως:

$$\lambda = -(JJ^T)^{-1}e \Rightarrow \delta X = -J^T(JJ^T)^{-1}e \quad (4.6)$$

Ορίζεται το σφάλμα Sampson ως το μέτρο της μετατόπισης του  $\delta X$ :

$$SE() = \| \delta X \|^2 = \delta X^T \delta X = e^T (JJ^T)^{-1} e \quad (4.7)$$

όπου  $J = \begin{bmatrix} \frac{\delta C_F}{\partial x} & \frac{\delta C_F}{\partial y} & \frac{\delta C_F}{\partial x'} & \frac{\delta C_F}{\partial y'} \end{bmatrix}$  η μήτρα μερικών παραγώνων της συνάρτησης  $C_F(X)$ . Επομένως το σφάλμα Sampson ως προς τον επιπολικό περιορισμό του Θεμελιώδους πίνακα  $F$  είναι:

$$SE() = \frac{(u_i'^T F x_i)^2}{(u'^T F)_x^2 + (u'^T F)_y^2 + (Fu)_{x'}^2 + (Fu)_{y'}^2} \quad (4.8)$$

## 4.4 Κανονικοποιημένος Αλγόριθμος 8-σημείων

Όπως αποδείχθηκε στην Ενότητα 3.4.1 ο Θεμελιώδης πίνακας  $F$  ικανοποιεί την εξίσωση  $u'Fu = 0$  για κάθε ζεύγος ταυτισμένων τοπικών χαρακτηριστικών  $u' = (x', y', 1) \leftrightarrow u = (x, y, 1)$  σε δύο εικόνες. Η σχέση αυτή μπορεί να γραφτεί για  $n$  ταυτισμένα σημεία ως ένα ομογενές

σύστημα:

$$A \cdot f = \begin{bmatrix} x'_1 x_1 & x'_1 y_1 & y'_1 x_1 & y'_1 y_1 & y'_1 & x_1 & y_1 & 1 \\ x'_2 x_2 & x'_2 y_2 & y'_2 x_2 & y'_2 y_2 & y'_2 & x_2 & y_2 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x'_n x_n & x'_n y_n & y'_n x_n & y'_n y_n & y'_n & x_n & y_n & 1 \end{bmatrix} \cdot \begin{pmatrix} f_{11} \\ f_{12} \\ f_{13} \\ f_{21} \\ f_{22} \\ f_{23} \\ f_{31} \\ f_{32} \\ f_{33} \end{pmatrix} = 0 \quad (4.9)$$

Το παραπάνω σύστημα είναι ομογενές η λύση  $f$  μπορεί να βρεθεί με ακρίβεια μίας πολλαπλασιαστικής σταθεράς. Για να υπάρχει μοναδική λύση ο πίνακας  $A$  πρέπει να έχει βαθμό ίσο με 8, ωστόσο αν τα σημεία είναι  $n > 8$  τότε η λύση δίνεται με τη μέθοδο Ελαχίστων Τετραγώνων (Least-Squares).

Καθώς οι μετρήσεις των σημείων συνήθως είναι θορυβώδεις ένα σημαντικό βήμα στον υπολογισμό του πίνακα  $F$  είναι ο μετασχηματισμός των σημείων πριν το σχηματισμό των εξισώσεων ούτως ώστε να μειωθεί η επίδραση του θορύβου στον υπολογισμό του. Αυτό επιτυγχάνεται μετατοπίζοντας τα σημεία ώστε να έχουν κέντρο το σημείο  $(0, 0)$  και η RMS απόσταση τους από το κέντρο να είναι ίση με  $\sqrt{2}$ .

Για να εξασφαλιστεί η ιδιότητα του Θεμελιώδους πίνακα για βαθμό ίσο με 2, μετά τον υπολογισμό του επιβάλλεται ο περιορισμός αυτός με τη χρήση της αποσύνθεσης σε ιδιόμορφες τιμές (Singular Value Decomposition) και λαμβάνοντας το διάνυσμα που αντιστοιχεί στην ελάχιστη ιδιόμορφη τιμή του πίνακα  $A$ . Ο κανονικοποιημένος αλγόριθμος 8-σημείων για τον υπολογισμό του Θεμελιώδους πίνακα  $F$  παρουσιάζεται πιο κάτω [25]:

---

**Algorithm 2** Κανονικοποιημένος Αλγόριθμος 8-σημείων

---

- 1: **Κανονικοποίηση:** Μετασχηματισμός των ταυτισμένων σημείων ως  $\hat{u}_i = T u_i$  και  $\hat{u}'_i = T' u'_i$  όπου  $T$ ,  $T'$  οι πίνακες μετασχηματισμού των σημείων.
- 2: **Γραμμική Επίλυση:** Υπολογισμός του Θεμελιώδους πίνακα  $\hat{F}$  από τα ζεύγη  $\hat{u}_i \leftrightarrow u'_i$ .  
Ανάλυση SVD του πίνακα  $A$  τ.ω.  $A = U D V^T$ ,  
 $\hat{F} \leftarrow v_3$
- 3: **Επιβολή περιορισμού**  $rank(\hat{F}) = 2$ : Ανάλυση SVD του πίνακα  $\hat{F}$  τ.ω.  $\hat{F} = U D V^T$   
όπου  $D = \text{diag}(r, s, t)$   
 $\hat{F}' \leftarrow U \text{diag}(r, s, 0) V^T$
- 4: **Αποκανονικοποίηση:** Θεμελιώδης πίνακας  $F \leftarrow T' \hat{F}' T$

όπου  $v_3$  είναι η τρίτη στήλη του πίνακα  $V$  που προκύπτει από την SVD ανάλυση.

---

## 4.5 Αλγόριθμος εκτίμησης πίνακα Εξωτερικών παραμέτρων της κάμερας

Όπως αναφέρθηκε στην Ενότητα 3.4.2 γνωρίζοντας το Θεμελιώδη πίνακα  $F$  δύο διαδοχικών εικόνων με ζεύγη ταυτισμένων σημείων  $u' \leftrightarrow u$  είναι δυνατόν να εξαχθεί ο πίνακας εξωτερικών παραμέτρων της κάμερας, δηλαδή η θέση της ως προς το σύστημα συντεταγμένων του χώρου. Αυτό γίνεται χρησιμοποιώντας το Βασικό πίνακα  $E = K^T F K$  (Εξ. 3.16).

Έστω ότι για δύο διαδοχικές εικόνες οι μη βαθμονομημένοι πίνακες προβολής είναι:

$$P_{1,c} = [I \mid 0] \quad P_{2,c} = [R \mid t] \quad (4.10)$$

Παρατηρώντας τον προβολικό πίνακα  $P_{1,c}$  είναι προφανές ότι το σύστημα συντεταγμένων του χώρου ταυτίζεται με το τοπικό σύστημα συντεταγμένων της 1<sup>ης</sup> κάμερας, και επομένως ο προβολικός πίνακας της 2<sup>ης</sup> περιγράφει τη σχετική κίνηση της κάμερας ως προς την αρχική της θέση. Με βάση την Εξ. 3.19 μπορούμε να γράψουμε τους τέσσερις πιθανούς μη βαθμονομημένους πίνακες προβολής της 2<sup>ης</sup> κάμερας ως εξής:

$$\begin{aligned} P_{2,c}^{(a)} &= [UWV^T \mid \hat{t}] \\ P_{2,c}^{(b)} &= [UWV^T \mid -\hat{t}] \\ P_{2,c}^{(c)} &= [UW^T V^T \mid \hat{t}] \\ P_{2,c}^{(d)} &= [UW^T V^T \mid -\hat{t}] \end{aligned} \quad (4.11)$$

Για να επιλεγεί ο σωστός πίνακας  $P_2$  εξετάζονται και οι τέσσερις περιπτώσεις βρίσκοντας το πλήθος των 3Δ σημείων που βρίσκεται μπροστά από τις δύο κάμερες. Έστω  $X$  το σημείο στο 3Δ χώρο και η προβολή του στα επίπεδα εικόνας  $u \leftrightarrow u'$  τα οποία με γνωστό τον πίνακα εσωτερικών παραμέτρων είναι:

$$\begin{aligned} u_c &= K^{-1}u \\ u'_c &= K^{-1}u' \end{aligned} \quad (4.12)$$

Τότε ισχύει:

$$\begin{aligned} u_c &= P_{1,c}X \\ u'_c &= P_{2,c}^j X \end{aligned} \quad (4.13)$$

Ορίζεται επομένως το γραμμικό σύστημα  $A_c X = 0$  όπου ο πίνακας  $A_c$  δίνεται από την σχέση:

$$A_c = \begin{bmatrix} x_c p_{1,c}^{3T} - p_{1,c}^{1T} \\ y_c p_{1,c}^{3T} - p_{1,c}^{2T} \\ x'_c p_{2,c}^{3T} - p_{2,c}^{1T} \\ y'_c p_{2,c}^{3T} - p_{2,c}^{2T} \end{bmatrix} \quad (4.14)$$

Υπολογιστικά, για να αποφευχθούν λάθη λόγω προσεγγίσεων, κάθε γραμμή του πίνακα  $A$  κανονικοποιείται ως προς τη νόρμα της.

$$\hat{A}_c = \left[ \begin{array}{cccc} \frac{A_1}{\|A_1\|} & \frac{A_2}{\|A_2\|} & \frac{A_3}{\|A_3\|} & \frac{A_4}{\|A_4\|} \end{array} \right]^T \quad (4.15)$$

Η επίλυση του ομογενούς συστήματος γίνεται με τη μέθοδο ελαχίστων τετραγώνων και από αυτή εξάγεται η 3Δ θέση του σημείου  $X$  στο χώρο. Η διαδικασία αυτή εφαρμόζεται για όλο το πλήθος  $N$  ζευγών ταυτισμένων σημείων και τελικά επιλέγεται ο κατάλληλος συνδυασμός προβολικών πινάκων  $P_{1,c}, P_{2,c}$ . Οι τελικοί προβολικοί πίνακες εξάγονται μετά από βαθμονόμηση τους και είναι:

$$\begin{aligned} P_1 &= KP_{1,c} \\ P_2 &= KP_{2,c} \end{aligned} \tag{4.16}$$

## Κεφάλαιο 5

# Προσέγγιση με Τοπικά Χαρακτηριστικά

### 5.1 Εισαγωγή

Στο κεφάλαιο αυτό παρουσιάζεται η εφαρμογή η οποία υλοποιήθηκε στα πλαίσια της παρούσας διπλωματικής εργασίας. Αποτελεί μια εφαρμογή Επαυξημένης Πραγματικότητας με τη χρήση τοπικών χαρακτηριστικών για εκτίμηση της σχετικής κίνησης μονής κάμερας σε πραγματικό χρόνο. Αρχικά παρουσιάζονται δύο από τα βασικά τμήματα της υλοποίησης, τα τοπικά χαρακτηριστικά BRISK, καθώς και η πλατφόρμα εκτίμησης θέσης της κάμερας LibViso2. Στη συνέχεια παρουσιάζεται η υλοποίηση της εφαρμογής με διαγράμματα ροής καθώς και με στιγμιότυπα τα οποία παρουσιάζουν την ευθυγράμμιση του εικονικού με τον πραγματικό χώρο σε άγνωστο περιβάλλον.

### 5.2 Τοπικά Χαρακτηριστικά BRISK

Τα τοπικά χαρακτηριστικά Binary Robust Invariant Scalable Keypoints (BRISK) προτάθηκαν από τους Leutenegger et al. [26]. Αποτελούνται από ανιχνευτή τοπικών χαρακτηριστικών και περιγραφέα. Έχουν τη δυνατότητα να παραμένουν αναλλοίωτα στις μεταβολές κλίμακας της εικόνας (scale invariant), ενώ η ανίχνευση και περιγραφή τους είναι υπολογιστικά αποδοτική. Η μέθοδος BRISK επιτυγχάνει ανεξαρτησία σε κλίμακα ανιχνεύοντας τοπικά μέγιστα, τα οποία αποτελούν περιοχές τοπικών χαρακτηριστικών. Τα τοπικά μέγιστα ανιχνεύονται, όχι μόνο στο επίπεδο της εικόνας αλλά και στο χώρο της κλίμακας (scale-space), χρησιμοποιώντας το κριτήριο FAST [19]. Ο ανιχνευτής χαρακτηριστικών της μεθόδου εκτιμά την πραγματική κλίμακα του κάθε τοπικού χαρακτηριστικού σε ένα συνεχή χώρο κλίμακας, σε αντίθεση με τις άλλες μεθόδους ανίχνευσης χαρακτηριστικών. Αυτό επιτυγχάνεται παρεμβάλλοντας μια τετραγωνική συνάρτηση στις διακρι-

τές τιμές της κλίμακας.

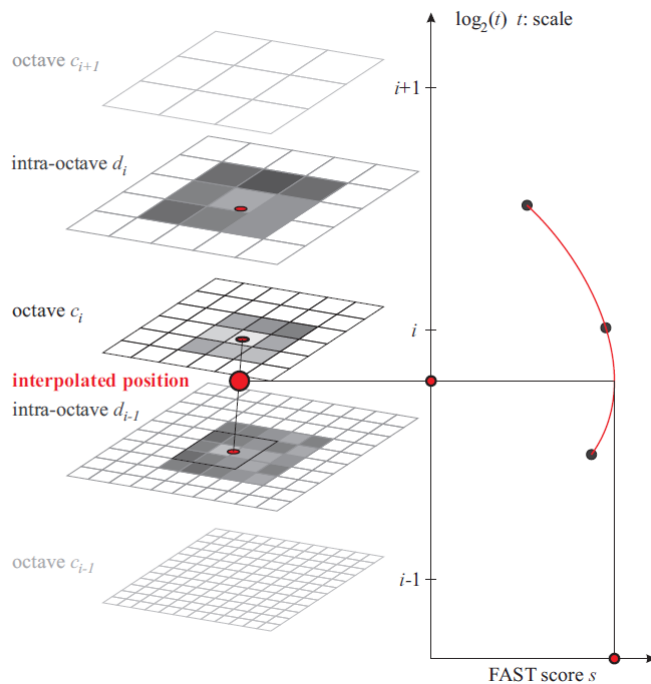
Η πυραμίδα που ορίζει το χώρο της κλίμακας αποτελείται από  $n$  οκτάβες  $c_i$  και  $n$  εσωτερικές οκτάβες  $d_i$ , όπου  $i = \{0, 1, \dots, n - 1\}$  και συνήθως αριθμό οκτάβων  $n = 4$ . Οι οκτάβες αυτές δημιουργούνται σταδιακά υποδειματοληπτώντας στο μισό την εικόνα που προκύπτει από την προηγούμενη οκτάβα, ξεκινώντας από την αρχική εικόνα ( $c_0$ ). Κάθε εσωτερική οκτάβα  $d_i$  τοποθετείται μεταξύ των στρωμάτων  $c_i$  και  $c_{i+1}$  της πυραμίδας. Η αρχική εσωτερική οκτάβα ( $d_0$ ) δημιουργείται υποδειματοληπτώντας την αρχική εικόνα ( $c_0$ ) κατά 1.5, ενώ όλες οι υπόλοιπες εσωτερικές οκτάβες δημιουργούνται υποδειματοληπτώντας στο μισό. Επομένως αν  $t$  είναι η κλίμακα τότε έχουμε για κάθε οκτάβα  $t(c_i) = 2^i$  και για κάθε εσωτερική οκτάβα  $t(d_i) = 2^i \cdot 1.5$ . Για να ισχύσει το κριτήριο FAST η μέθοδος BRISK χρησιμοποιεί μια μάσκα 9-16 εικονοστοιχείων (pixels), δηλαδή απαιτεί 9 διαδοχικά εικονοστοιχεία σε μία γειτονιά 16 εικονοστοιχείων να είναι αρκετά φωτεινότερα ή σκοτεινότερα από το κεντρικό εικονοστοιχείο. Αρχικά ο ανιχνευτής εφαρμόζει το κριτήριο FAST με μάσκα 9-16 εικονοστοιχείων σε κάθε οκτάβα και εσωτερική οκτάβα χρησιμοποιώντας το ίδιο κατώφλι  $T$ , ούτως ώστε να αναγνωρίσει τα πιθανά σημεία τοπικών χαρακτηριστικών. Στη συνέχεια τα σημεία αυτά για να ανιχνευτούν ως τοπικά χαρακτηριστικά πρέπει:

- Το κριτήριο FAST του σημείου που εξετάζεται να είναι το τοπικό μέγιστο σε μία γειτονιά 8 εικονοστοιχείων στο ίδιο επίπεδο της πυραμίδας.
- Τα κριτήρια FAST του επόμενου και προηγούμενου επιπέδου της πυραμίδας να είναι μικρότερα από το εξεταζόμενο σημείο.

Για την περίπτωση της αρχικής οκτάβας ( $c_0$ ) εφαρμόζεται το κριτήριο FAST με μάσκα 5-8 εικονοστοιχείων καθώς δεν υπάρχει προηγούμενο επίπεδο. Τέλος η θέση και η κλίμακα του κάθε τοπικού μεγίστου, δηλαδή τοπικού χαρακτηριστικού, αποκτάται με παρεμβολή τετραγωνικών συναρτήσεων (Σχήμα 5.1).

Η μέθοδος BRISK συνθέτει τον περιγραφέα κάθε τοπικού χαρακτηριστικού σαν μία δυαδική συμβολοσειρά μήκους 512 bits η οποία περιέχει τα αποτελέσματα συγκριτικών δοκιμών φωτεινότητας. Οι δοκιμές φωτεινότητας γίνονται χρησιμοποιώντας ένα πρότυπο δειγματοληψίας  $N$  δειγμάτων γύρω από κάθε τοπικό χαρακτηριστικό.





Σχήμα 5.1: Ανίχνευση τοπικών χαρακτηριστικών BRISK ως το μέγιστο μέγεθος FAST σε τρία επίπεδα της πυραμίδας χώρου κλίμακας.

### 5.3 Πλατφόρμα Οπτικής Οδομετρίας

Η πλατφόρμα LibViso2 (Library for Visual Odometry) [27] αποτελεί μια πλατφόρμα υπολογισμού της θέσης κινούμενης κάμερας με τη χρήση τοπικών χαρακτηριστικών (Ενότητα 2.1) και όρασης πολλαπλών όψεων (Κεφ. 3) για σκοπούς Δομής από Κίνηση (Structure from Motion) κινούμενων οχημάτων σε περιβάλλον πόλης. Υπάρχει δυνατότητα χρησιμοποίησης εικόνων από στερεοσκοπική ή μονή κάμερα από τις οποίες μπορεί να εξάγει και να ταυτίσει τοπικά χαρακτηριστικά, εκτίμησης θέσης της κάμερας ως προς το σύστημα συντεταγμένων του κόσμου, δημιουργίας 3Δ χάρτη τοπικών χαρακτηριστικών και τέλος 3Δ ανακατασκευής του χώρου. Στα πλαίσια της παρούσας διπλωματικής από την πλατφόρμα LibViso2 χρησιμοποιήθηκε η εξαγωγή τοπικών χαρακτηριστικών και ο υπολογισμός της θέσης της κάμερας με εικόνες από μονή κάμερα (monocular).

#### 5.3.1 Τοπικά Χαρακτηριστικά πλατφόρμας

Η είσοδος στον αλγόριθμο Εκτίμησης Θέσης της πλατφόρμας είναι τοπικά χαρακτηριστικά τα οποία ταυτίζονται μεταξύ της προηγούμενης και της τρέχουσας εικόνας (διαδοχικά καρέ). Τα τοπικά χαρακτηριστικά που χρησιμοποιεί το LibViso2 εξάγονται φιλτράροντας τις διαδοχικές ει-

κόνες με μάσκες blobs και γωνιών μεγέθους  $5 \times 5$ . Βρίσκοντας τα τοπικά μέγιστα και ελάχιστα των φιλτραρισμένων εικόνων για τις δύο μάσκες επιλέγονται τα υποψήφια τοπικά χαρακτηριστικά. Ακολούθως χρησιμοποιείται η τεχνική Bucketing[28][29] η οποία, συνοπτικά, χωρίζει την εικόνα σε πολλά επικαλυπτόμενα τετράγωνα (buckets) όπου στο καθένα επιλέγεται ένας μέγιστος αριθμός ταυτισμένων τοπικών χαρακτηριστικών. Αυτή η τεχνική αρχικά μειώνει το υπολογιστικό κόστος. Επίσης εγγυάται ότι τα τοπικά χαρακτηριστικά είναι ομοιόμορφα κατανομημένα κατά τον άξονα  $z$  της κάμερας δίνοντας καλή εκτίμηση της γραμμικής και γωνιακής ταχύτητας, με αποτέλεσμα καλή συνολική εκτίμηση θέσης της κάμερας. Τέλος τα τοπικά χαρακτηριστικά κατανέμονται σε ολόκληρη την εικόνα προσφέροντας καλή εκτίμηση σε δυναμικά μεταβαλλόμενες σκηνές. Η ταύτιση των συγκεκριμένων τοπικών χαρακτηριστικών μεταξύ δύο εικόνων γίνεται συγκρίνοντας την απόσταση της απόκρισης ενός οριζόντιου και κάθετου Sobel φίλτρου μεγέθους  $11 \times 11$ . Για λόγους ευκολίας τα τοπικά χαρακτηριστικά αυτά θα αναφέρονται ως τοπικά χαρακτηριστικά LibViso2.

### 5.3.2 Εκτίμηση Θέσης της Κάμερας

Ο πίνακας προβολής  $P$ , δηλαδή η κίνηση της κάμερας μεταξύ δύο διαδοχικών καρέ υπολογίζεται, όπως περιγράφετε στην Ενότητα 3.4, χρησιμοποιώντας τη Επιπολική Γεωμετρία. Απαραίτητη προϋπόθεση είναι να είναι γνωστός εκ των προτέρων ο πίνακας εσωτερικών παραμέτρων της κάμερας  $K$ . Ο Θεμελιώδης πίνακας  $F$  υπολογίζεται από τα ταυτισμένα τοπικά χαρακτηριστικά με τον απλοποιημένο αλγόριθμο RANSAC (RANdom SAMple Consensus) 8-σημείων. Επιχειρείται δηλαδή να γίνει μια εύρωστη εκτίμηση με τη μέθοδο RANSAC του Θεμελιώδους πίνακα  $F$  από τα θορυβώδη τοπικά χαρακτηριστικά με βάση τον αλγόριθμο 8-σημείων, έχοντας σαν κριτήριο επιλογής των *inliers* το σφάλμα Sampson(Εξ. 4.8). Ακολούθως υπολογίζεται ο Βασικός πίνακας  $E$  και αποσυνθέτοντάς τον επιλέγεται η σωστή λύση του πίνακα προβολής  $P$  της κάμερας για τα δύο διαδοχικά καρέ (Ενότητα 4.5). Η συνολική κίνηση της κάμερας για την  $k$  χρονική στιγμή σε σχέση με το σύστημα συντεταγμένων του κόσμου  $P_0 = [I \mid 0]$  υπολογίζεται ως εξής:

$$P_0^k = P_0^{k-1} \cdot P_{k-1}^k \quad (5.1)$$

---

**Algorithm 3** Αλγόριθμος RANSAC 8-σημείων για υπολογισμό του  $F$ 

---

- 1: **Κανονικοποίηση:** Μετασχηματισμός των ταυτισμένων σημείων ως  $\hat{u}_i = Tu_i$  και  $\hat{u}'_i = T'u'_i$  όπου  $T, T'$  οι πίνακες μετασχηματισμού των σημείων.
- 2: **for**  $j=1:N$  **do**
- 3:   Επέλεξε τυχαία 8 σημεία  $\hat{u} \leftrightarrow \hat{u}'$
- 4:   **Γραμμική Επίλυση:** Υπολογισμός του Θεμελιώδους πίνακα  $\hat{F}$  από τα ζεύγη  $\hat{u}_i \leftrightarrow \hat{u}'_i$ .  
    Ανάλυση SVD του πίνακα  $A$  τ.ω.  $A = UDV^T$   
     $\hat{F} \leftarrow v_3$
- 5:   **if**  $SE() < t$  **then**
- 6:      $inliers \leftarrow \hat{u}, \hat{u}'$
- 7:   **end if**
- 8:   **if**  $|inliers| > |inliersF|$  **then**
- 9:      $inliersF \leftarrow inliers$
- 10:   **end if**
- 11: **end for**
- 12: **Γραμμική Επίλυση** μόνο με τα σημεία  $inliersF$ :  
    Ανάλυση SVD του πίνακα  $A$  τ.ω.  $A = UDV^T$   
     $\hat{F} \leftarrow v_3$
- 13: **Επιβολή περιορισμού**  $rank(\hat{F}) = 2$ : Ανάλυση SVD του πίνακα  $\hat{F}$  τ.ω.  $\hat{F} = UDV^T$   
    όπου  $D = \text{diag}(r, s, t)$   
     $\hat{F}' \leftarrow U \text{diag}(r, s, 0) V^T$
- 14: **Αποκανονικοποίηση:** Θεμελιώδης πίνακας  $F \leftarrow T' \hat{F}' T$

όπου το κατώφλι επιλογής των  $inliers$  ορίζεται ως  $t = 0.0001$  και  $N = 2000$ ,  $v_3$  η τρίτη στήλη του πίνακα  $V$  που προκύπτει από την SVD ανάλυση.

---

## 5.4 Η εφαρμογή

Στα πλαίσια της παρούσας διπλωματικής εργασίας αναπτύχθηκε μια εφαρμογή Επαυξημένης Πραγματικότητας με χρήση τοπικών χαρακτηριστικών. Σε αντίθεση με άλλες μεθόδους που χρησιμοποιούνται στην Επαυξημένη Πραγματικότητα με τη χρήση τοπικών χαρακτηριστικών δεν είναι απαραίτητη η εκ των προτέρων γνώση 3D σημείων στο χώρο, ή η αναγνώριση συγκεκριμένων σημειωτών (markers) στην εικόνα. Αν και η μεθοδολογία αυτή χρησιμοποιείται κατά κόρον σε εφαρμογές ρομποτικής αυτοεντοπισμού και χαρτογράφησης (Simultaneous Localization and Mapping) για την εξαγωγή της θέσης κάμερας πρόσφατα έχει γίνει δυνατή η εφαρμογή της και στον τομέα της Επαυξημένης Πραγματικότητας. Η αύξηση της υπολογιστικής ισχύος και η ύπαρξη αλγορίθμων ανίχνευσης και ταύτισης αξιόπιστων τοπικών χαρακτηριστικών σε πραγματικό χρόνο έχουν συντελέσει καταλυτικά σε αυτό.

Η εφαρμογή χρησιμοποιεί τη δυνατότητα πολυνημάτωσης (multithreading) για να διαχειριστεί σε δύο διαφορετικά νήματα (threads) τη διαδικασία εξαγωγής της θέσης της κάμερας και παράλληλα

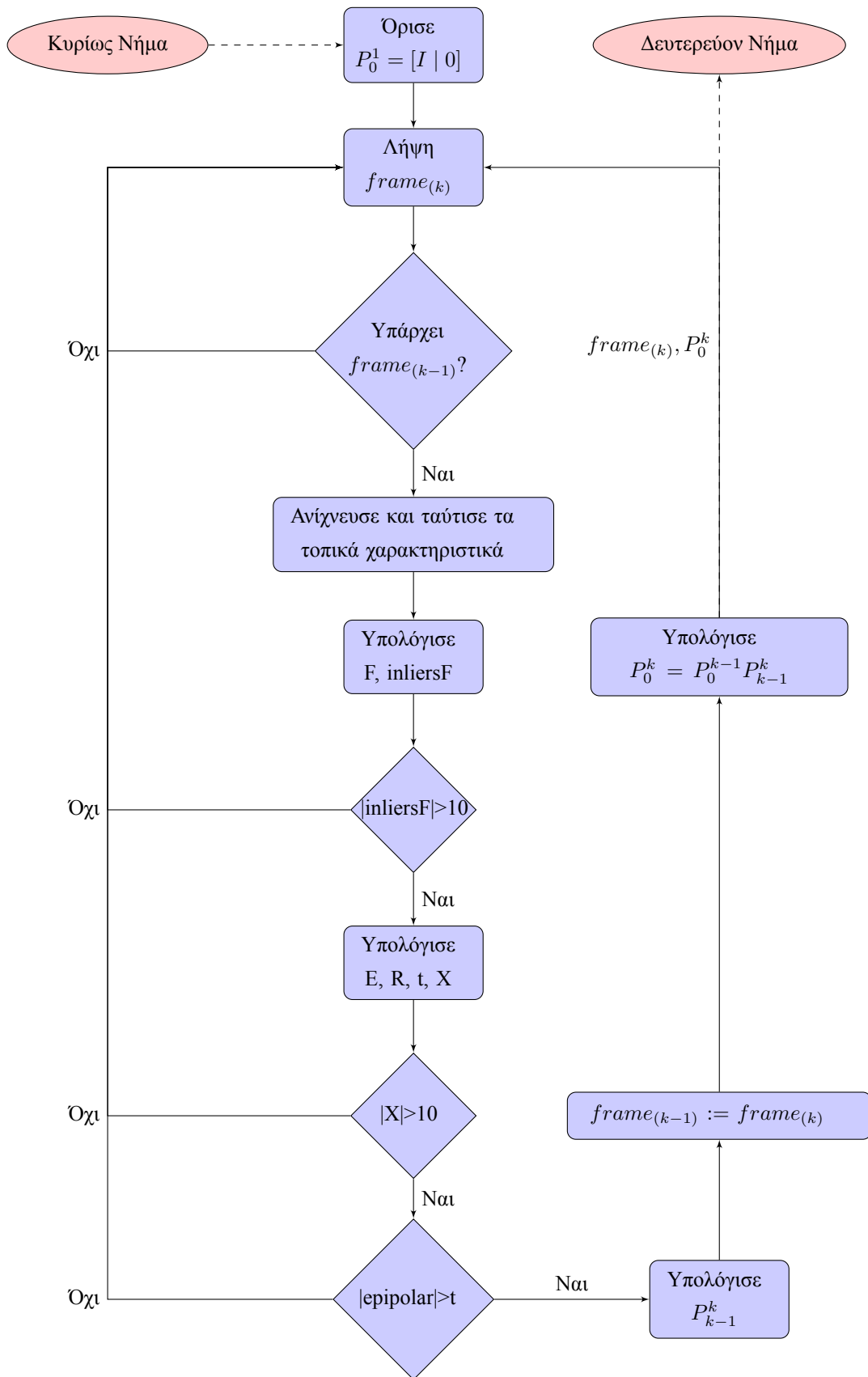
την παραγωγή του 3D εικονικού μοντέλου και την προβολή του στο επίπεδο της εικόνας, επανξάνοντας έτσι την πραγματικότητα. Το κυρίως νήμα (thread) αναλαμβάνει τη διαδικασία καταγραφής, επεξεργασίας και εξαγωγής της θέσης της κάμερας ενώ το δευτερεύον νήμα, που τρέχει παράλληλα με το κυρίως, αναλαμβάνει μόνο τη διαδικασία παρουσίασης της οπτικής πληροφορίας στο χρήστη με βάση την ανανεωμένη θέση της κάμερας σε κάθε επανάληψη.

#### 5.4.1 Εκτίμηση θέσης της κάμερας

Στο κυρίως νήμα τρέχει η εφαρμογή LibViso2 με σκοπό να ληφθεί η σχετική θέση της κάμερας ως το σύστημα συντεταγμένων του κόσμου, το οποίο ορίζεται ως το πρώτο καρέ που λαμβάνεται από το σύστημα. Ο πηγαίος κώδικας (source code) της πλατφόρμας, στα πλαίσια της παρούσας διπλωματικής, υπέστη κάποιες αλλαγές-προσθήκες ούτως ώστε να είναι δυνατή η λήψη των διαδοχικών εικόνων μέσω μιας κάμερας σε πραγματικό χρόνο. Επίσης προστέθηκε μια κλάση με την οποία υπάρχει η δυνατότητα επιλογής των τοπικών χαρακτηριστικών που εξάγονται από τα διαδοχικά καρέ. Ο χρήστης πέρα από τα τοπικά χαρακτηριστικά LibViso2 που χρησιμοποιεί η πλατφόρμα μπορεί να επιλέξει μεταξύ όλων των υλοποιήσεων ανιχνευτών και περιγραφέων τοπικών χαρακτηριστικών που υπάρχουν στη βιβλιοθήκη ελεύθερου λογισμικού OpenCV (Open Source Computer Vision), ή ακόμα και των τοπικών χαρακτηριστικών BRISK. Ένα συνοπτικό διάγραμμα ροής της διαδικασίας που ακολουθείται στο κυρίως νήμα για την εκτίμηση θέσης της κάμερας παρουσιάζεται στο Σχήμα 5.2.

Η διαδικασία που ακολουθείται για τον υπολογισμό της σχετικής κίνησης της κάμερας μεταξύ των δύο διαδοχικών καρέ περιλαμβάνει, όπως έχει αναφερθεί, χρήση ενός απλοποιημένου αλγορίθμου RANSAC 8-σημείων (Αλγόριθμος 3), υπολογισμό του Βασικού πίνακα  $E$  και τέλος εξαγωγή του σωστού πίνακα εξωτερικών παραμέτρων  $[R | t]$  (Ενότητα 4.5). Παρόλα αυτά, είναι προφανές ότι πρέπει να υπάρχουν συνεχείς έλεγχοι σε κάθε βήμα ούτως ώστε να υπάρξει μια καλή εκτίμηση της σχετικής κίνησης της κάμερας. Αν οποιοσδήποτε από τους παρακάτω ελέγχους αποτύχει τότε η συνολική κίνηση της κάμερας δεν ανανεώνεται για το συγκεκριμένο ζεύγος διαδοχικών καρέ.

- Αρχικά για να έχουμε μια καλή εκτίμηση του Θεμελιώδους πίνακα  $F$  ο αριθμός των ταυτισμένων σημείων που χαρακτηρίζονται ως *inliers* πρέπει να είναι:  $|inliersF| > 10$ .
- Επίσης πρέπει η κίνηση της κάμερας να είναι μεγαλύτερη από ένα κατώφλι ούτως ώστε ο Θεμελιώδης πίνακας  $F$  να μη θεωρείται θορυβώδης. Αυτό επιτυγχάνεται ελέγχοντας το μήκος της επιπολικής γραμμής στο επίπεδο της εικόνας του τρέχοντος καρέ, που πρέπει να είναι μεγαλύτερη από ένα συγκεκριμένο κατώφλι.
- Τέλος αφού έχει επιλεγεί ο σωστός πίνακας  $P$  από την αποσύνθεση του Σημαντικού πίνακα πρέπει το πλήθος των 3D σημείων  $X$  μπροστά από την κάμερα να είναι:  $|X| > 10$ .



Σχήμα 5.2: Διάγραμμα ροής Εκτίμησης Θέσης

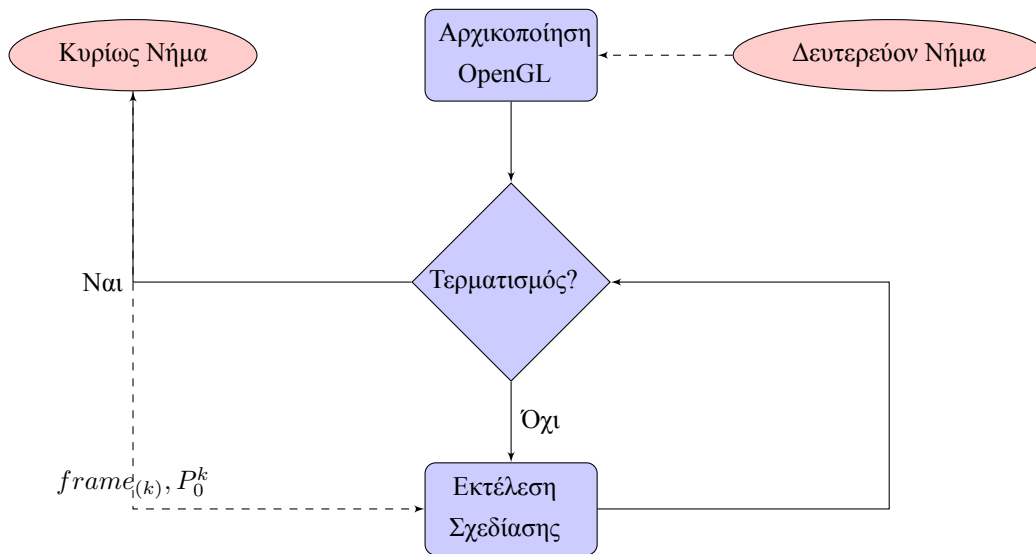
## 5.4.2 Παρουσίαση Επαύξεσης

Το δευτερεύον νήμα είναι υπεύθυνο για την παρουσίαση της επαύξεσης στο χρήστη, δηλαδή την προβολή του τρέχοντος καρέ προσθέτοντας το γραφικό μοντέλο στην επιθυμητή θέση με βάση τη θέση της κάμερας τη δεδομένη χρονική στιγμή. Στα πλαίσια της παρούσας διπλωματικής, λοιπόν, υλοποιήθηκε μια εφαρμογή σε OpenGL (Open Graphics Library) η οποία αναλαμβάνει αυτήν ακριβώς τη διαδικασία.

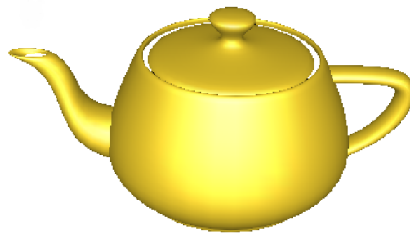
Η OpenGL είναι ένα χαμηλού επιπέδου, διαδραστικό API, που απαιτεί από τον προγραμματιστή να υπαγορεύει τα ακριβή βήματα που απαιτούνται για να σχεδιαστεί μια σκηνή. Επομένως, για τη δημιουργία του γραφικού μοντέλου για μία ρεαλιστική επαύξηση της πραγματικότητας απαιτείται να οριστούν οι κατάλληλοι μετασχηματισμοί της σκηνής και του γραφικού αντικειμένου. Ταυτόχρονα για να είναι πιο ρεαλιστική η προβολή του γραφικού μοντέλου έχει οριστεί φωτισμός στην σκηνή (Σχήμα 5.4). Η ιδιαιτερότητα της OpenGL βρίσκεται στο ότι, προκειμένου να μετασχηματίσουμε την κάμερα με βάση την ανανεωμένη της θέση, χρειάζεται να μετακινήσουμε όλη τη σκηνή με τον αντίστροφο μετασχηματισμό, μετακινώντας δηλαδή τη σχετική θέση της σκηνής ως προς την κάμερα. Η δομή υλοποίησης της OpenGL εφαρμογής στο δευτερεύον νήμα μπορεί να χωριστεί σε 4 διαφορετικές διαδικασίες:

- Αρχικοποίηση: Εδώ αρχικοποιούνται όλες οι συναρτήσεις της OpenGL που αφορούν τα συστήματα συντεταγμένων, το φωτισμό καθώς και το γραφικό μοντέλο.
- Συναρτήσεις εισόδου: Ανταποκρίνονται κατάλληλα στις εντολές του χρήστη.
- Συνάρτηση σχεδιασμού: Εκτελούνται όλες οι εντολές που ορίστηκαν στη διαδικασία αρχικοποίησης και παρουσιάζεται η επαυξημένη σκηνή στο χρήστη.
- Κυρίως κύκλος: Είναι ένας ατέρμονος βρόχος μέσα στον οποίο εκτελούνται όλες οι παραπάνω διαδικασίες μέχρι ο χρήστης να επιλέξει τον τερματισμό της εφαρμογής.

Το δευτερεύον νήμα επικοινωνεί με το κυρίως μόνο όταν υπάρχει διαθέσιμο από το κυρίως ανανεωμένο τρέχων καρέ ή ανανεωμένη σχετική κίνηση της κάμερας, όπως φαίνεται στο παρακάτω διάγραμμα ροής. Αντίστοιχα το δευτερεύον νήμα στέλνει στο κυρίως μόνο την εντολή του χρήστη για τερματισμό της εφαρμογής.



Σχήμα 5.3: Διάγραμμα Ροής Δευτερεύοντος Νήματος



Σχήμα 5.4: Γραφικό μοντέλο OpenGL

## 5.5 Παρουσίαση Αποτελεσμάτων

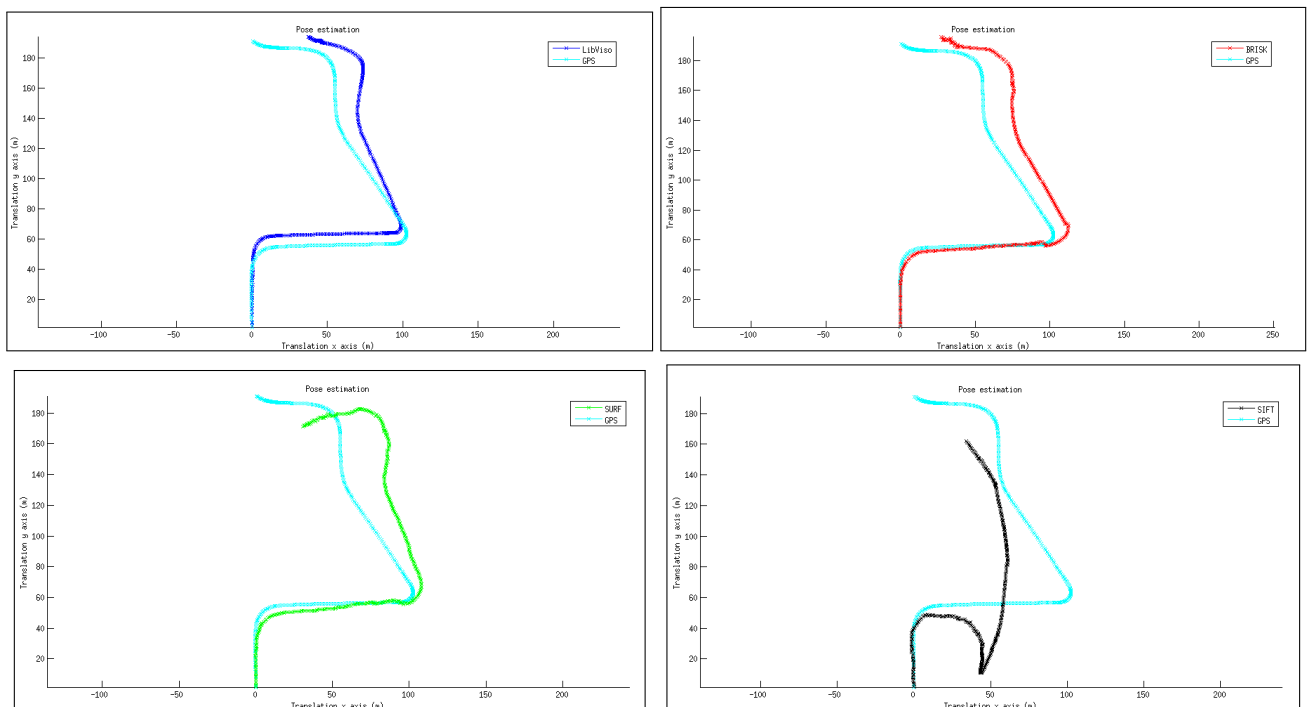
Η σωστή λειτουργία της εφαρμογής απαιτεί την ανίχνευση τοπικών χαρακτηριστικών μεταξύ δύο διαδοχικών καρέ και ακολούθως τη σωστή ταύτιση ικανοποιητικού αριθμού από αυτά. Η εκτίμηση του Θεμελιώδους πίνακα περιέχει μικρότερο σφάλμα όταν υπολογίζεται με μεγαλύτερο δείγμα ταυτισμένων τοπικών χαρακτηριστικών. Κατ' επέκταση και η εκτίμηση θέσης της κάμερας θα περιέχει μικρότερο σφάλμα. Ταυτόχρονα όμως δεν πρέπει να παραβλεφθεί η ανάγκη για ανίχνευση και ταύτιση των τοπικών χαρακτηριστικών σε πραγματικό χρόνο.

Στην ενότητα αυτή αρχικά, θα παρουσιαστούν τα αποτελέσματα εκτίμησης θέσης χρησιμοποιώντας τα χαρακτηριστικά LibViso2, BRISK, SURF και SIFT, σε ένα σύνολο διαδοχικών καρέ που παρέχεται από την πλατφόρμα οδομετρίας. Το σύνολο δεδομένων αυτό, λήφθηκε από κινούμενο όχημα με σταθερή ταχύτητα σε περιβάλλον πόλης και αποτελείται από 373 διαδοχικά καρέ ανάλυσης  $1344 \times 391 \text{ pixels}$ .

Στην συνέχεια, θα παρουσιαστούν τα αποτελέσματα της εφαρμογής ως προς την ανίχνευση και ταύτιση τοπικών χαρακτηριστικών σε επιλεγμένα στιγμιότυπα από ένα σύνολο 60 διαδοχικών καρτέ, τα οποία λήφθηκαν από “ζωντανή” κάμερα. Ακολούθως, παρουσιάζεται η επαύξηση της πραγματικότητας για τα στιγμιότυπα αυτά.

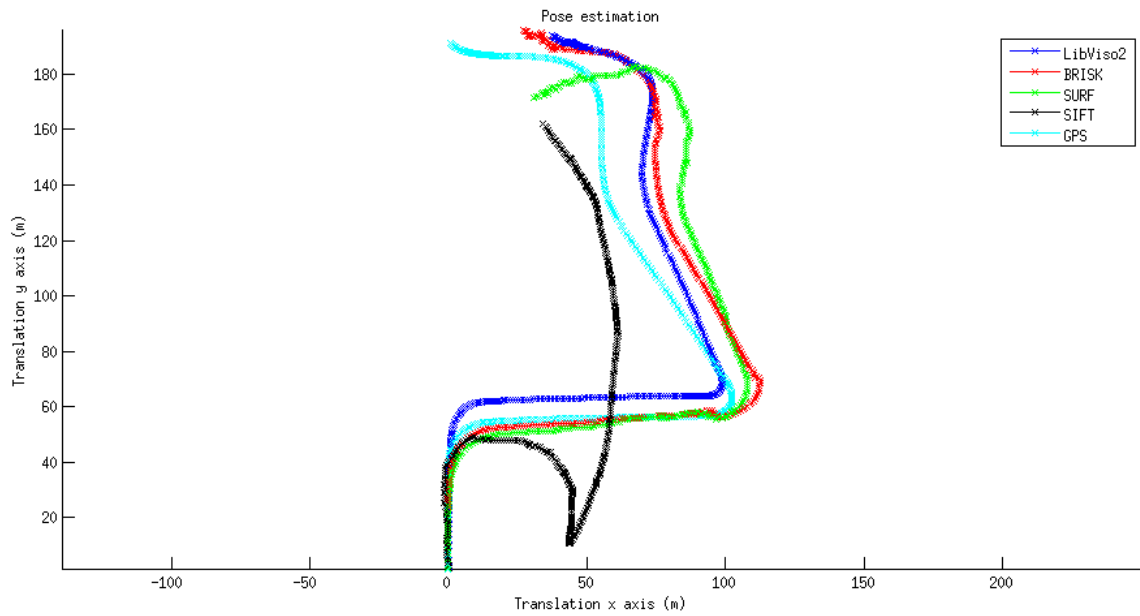
### 5.5.1 Εκτίμηση Θέσης από σύνολο δεδομένων

Όπως αναφέρθηκε η σκοπός της πλατφόρμας LibViso2 είναι η εκτίμηση θέσης κινούμενου οχήματος σε περιβάλλον πόλης. Στα πλαίσια της παρούσας διπλωματικής, χρησιμοποιήσαμε το σύνολο δεδομένων που παρέχει η πλατφόρμα για να αξιολογήσουμε τη εκτίμηση θέσης της κάμερας για τα διάφορα τοπικά χαρακτηριστικά. Το γεγονός ότι η κάμερα κινείται με σταθερή ταχύτητα προσφέρει μια πιο “ασφαλή” εκτίμηση θέσης της κάμερας.



Σχήμα 5.5: **Αριστερά προς δεξιά:** Εκτίμηση κίνησης της κάμερας χρησιμοποιώντας τα τοπικά χαρακτηριστικά LibViso2, BRISK, SURF και SIFT, από σύνολο δεδομένων της πλατφόρμας οπτικής οδομετρίας.





Σχήμα 5.6: Συγκριτικό διάγραμμα εκτίμησης θέσης

Παρατηρώντας το Σχήμα 5.5, βλέπει κανείς πόσο επηρεάζεται η εκτίμηση θέσης της κάμερας ανάλογα με τα τοπικά χαρακτηριστικά που χρησιμοποιούνται. Αν και χρησιμοποιώντας τα τοπικά χαρακτηριστικά LibViso2, BRISK και SURF εκτιμάται μία παρόμοια τροχιά από το σύνολο δεδομένων, η εκτίμηση που δίνεται με τη χρήση των χαρακτηριστικών SIFT διαφέρει αρκετά (Σχήμα 5.6)

### 5.5.2 Ταύτιση τοπικών χαρακτηριστικών

Το σύνολο των διαδοχικών καρτέ λήφθηκε από μονή κάμερα Microsoft LifeCam HD-3000 με ανάλυση εικόνας  $640 \times 480 \text{ pixels}$  σε άγνωστο περιβάλλον. Ο πίνακας εσωτερικών παραμέτρων υπολογίστηκε με τεχνική βαθμονόμησης χρησιμοποιώντας τις συναρτήσεις της βιβλιοθήκης OpenCV. Για κάθε καρτέ, το οποίο ανανεώνει τη θέση της κάμερας, έχει υπολογιστεί το πλήθος των ταυτισμένων χαρακτηριστικών καθώς και το ποσοστό των *inliers*. Στα Σχήματα 5.7, 5.8, 5.9, 5.10 παρουσιάζονται επιλεγμένα καρτέ, από τα οποία μπορεί να γίνει αντιληπτή η κίνηση της κάμερας. Τα ταυτισμένα τοπικά χαρακτηριστικά LibViso, BRISK, SURF και SIFT που έχουν επιλεγεί ως *inliers* παρουσιάζονται με κόκκινο χρώμα. Τα αποτελέσματα ως προς το πλήθος των *inliers* συνοψίζονται στον πιο κάτω πίνακα.

Καρτέ	Τοπικά Χαρακτηριστικά LibViso2			BRISK		
	Matches	Ποσοστό <i>inliers</i>	Πλήθος <i>inliers</i>	Matches	Ποσοστό <i>inliers</i>	Πλήθος <i>inliers</i>
1	185	58.4%	108	64	42.3%	27
2	223	68.6%	153	102	46.5%	47
3	215	53.5%	115	158	41.2%	65
4	214	56.5%	121	83	48.5%	40
5	197	67%	132	60	55%	33
6	207	64.3%	133	144	57%	82

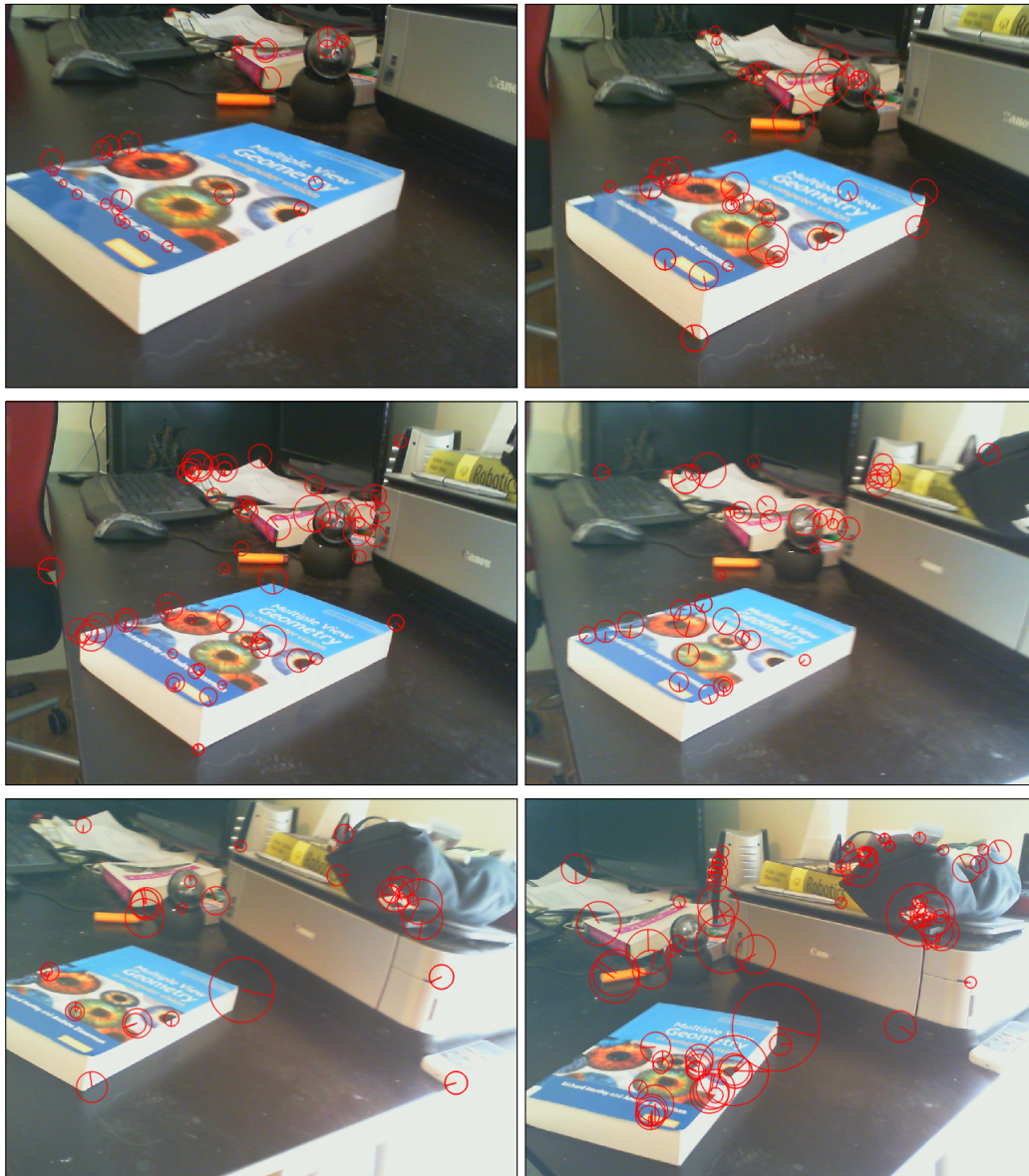
Καρτέ	SURF			SIFT		
	Matches	Ποσοστό <i>inliers</i>	Πλήθος <i>inliers</i>	Matches	Ποσοστό <i>inliers</i>	Πλήθος <i>inliers</i>
1	163	39.9%	65	188	19.2%	36
2	185	49.2%	91	265	26%	69
3	221	51.1%	113	354	25.4%	90
4	192	47.9%	92	147	32%	47
5	140	60%	84	108	24%	26
6	182	54.4%	99	264	32.5%	86

Πίνακας 5.1: Πίνακας αποτελεσμάτων ταυτισμένων τοπικών χαρακτηριστικών σε διαδοχικά καρτέ



Σχήμα 5.7: Αριστερά προς δεξιά: Τοπικά χαρακτηριστικά LibViso2 σε επιλεγμένα διαδοχικά καρέ. Με κόκκινο χρώμα παρουσιάζονται τα *inliers* του κάθε καρέ.

Είναι εύκολο να παρατηρήσει κανείς την ομοιόμορφη κατανομή των τοπικών χαρακτηριστικών στο επίπεδο εικόνας του κάθε καρέ. Αυτή η ομοιόμορφη κατανομή οφείλεται, όπως έχει αναφερθεί, στην τεχνική Bucketing.



Σχήμα 5.8: **Αριστερά προς δεξιά:** Τοπικά χαρακτηριστικά BRISK σε επιλεγμένα διαδοχικά καρέ. Με κόκκινο χρώμα παρουσιάζονται τα *inliers* του κάθε καρέ.

Τα τοπικά χαρακτηριστικά BRISK, SURF και SIFT ανιχνεύοντας τα χαρακτηριστικά τόσο στο επίπεδο της εικόνας, όσο και στο χώρο της κλίμακας έχουν τη δυνατότητα να εκτιμήσουν τη θέση, τον προσανατολισμό και την κλίμακα του κάθε τοπικού χαρακτηριστικού, όπως φαίνεται και στα στιγμιότυπα που παρουσιάζονται.



Σχήμα 5.9: **Αριστερά προς δεξιά:** Τοπικά χαρακτηριστικά SURF σε επιλεγμένα διαδοχικά καρέ. Με κόκκινο χρώμα παρουσιάζονται τα *inliers* του κάθε καρέ.



Σχήμα 5.10: **Αριστερά προς δεξιά:** Τοπικά χαρακτηριστικά SIFT σε επιλεγμένα διαδοχικά καρέ. Με κόκκινο χρώμα παρουσιάζονται τα *inliers* του κάθε καρέ.

Τα τοπικά χαρακτηριστικά BRISK, SURF και SIFT που χαρακτηρίζονται ως *inliers* και θα χρησιμοποιηθούν για την εκτίμηση θέσης της κάμερας είναι πολύ λιγότερα από τα χαρακτηριστικά LibViso2. Παρόλα αυτά, είναι δυνατή η εκτίμηση θέσης της κάμερας, καθώς το πλήθος τους επιτρέπει τον υπολογισμό του πίνακα  $F$  (Πίνακας 5.1).

### 5.5.3 Επαύξηση της πραγματικότητας-Ευθυγράμμιση

Συνεχίζοντας, η εφαρμογή εκτιμά τη θέση της κάμερας ως προς το αρχικό καρέ με τους αλγόριθμους που έχουν αναφερθεί στις προηγούμενες ενότητες. Η εκτίμηση θέσης της κάμερας, χρησιμοποιώντας τα τοπικά χαρακτηριστικά LibViso2, είναι σαφώς καλύτερη καθώς ο Θεμελιώδης πίνακας εκτιμάται με μεγαλύτερο πλήθος *inliers*, μειώνοντας έτσι το σφάλμα εκτίμησης του (Σχήμα 5.11, 5.12). Για τα στιγμιότυπα που επιλέχθηκαν παρουσιάζεται η επαύξηση της σκηνής χρησιμοποιώντας το γραφικό μοντέλο του Σχήματος 5.4.

Ο μέσος χρόνος που χρειάζεται η εφαρμογή για να εκτιμήσει τη θέση της κάμερας και να προβάλει την επαυξημένη σκηνή, για τα επιλεγμένα στιγμιότυπα που παρουσιάστηκαν στην προηγούμενη ενότητα, είναι για τα χαρακτηριστικά LibViso2  $151.15ms$  και για τα τοπικά χαρακτηριστικά BRISK  $145.86ms$  (Πίνακας 5.2). Οι χρόνοι αυτοί επιβεβαιώνουν τη δυνατότητα της εφαρμογής να τρέχει σε πραγματικό χρόνο για τα συγκεκριμένα χαρακτηριστικά.

Αντίθετα για τα τοπικά χαρακτηριστικά SURF και SIFT ο μέσος χρόνος που χρειάζεται είναι  $437.2ms$  και  $1003.8ms$  αντίστοιχα. Οι χρόνοι αυτοί καθιστούν τη χρήση των συγκεκριμένων τοπικών χαρακτηριστικών αδύνατη στη συγκεκριμένη εφαρμογή. Παρόλα αυτά, για λόγους πληρότητας η επαυξημένη σκηνή χρησιμοποιώντας στα χαρακτηριστικά SURF και SIFT παρουσιάζεται στα Σχήματα 5.13, 5.14.

Καρέ	Χρόνος Επανάληψης ( <i>ms</i> )			
	Τοπικά Χαρακτηριστικά LibViso2	BRISK	SURF	SIFT
1	154.5	143.2	492.0	918.1
2	153.3	143.7	514.2	1057.0
3	153.9	153.1	533.1	1012.0
4	156.0	138.7	508.1	928.4
5	140.0	150.1	477.7	1089.0
6	149.2	146.4	493.5	1018.8
Μέσος Χρόνος	151.15	145.86	437.2	1003.8

Πίνακας 5.2: Πίνακας σύγκρισης χρόνων για μια επανάληψη της εφαρμογής σε επιλεγμένα στιγμιότυπα.



Σχήμα 5.11: **Αριστερά προς δεξιά:** Η προβολή του γραφικού μοντέλου στη σκηνή. Η εκτίμηση θέσης της κάμερας έγινε χρησιμοποιώντας Σποραδικά χαρακτηριστικά.





Σχήμα 5.12: Αριστερά προς δεξιά: Η προβολή του γραφικού μοντέλου στη σκηνή. Η εκτίμηση θέσης της κάμερας έγινε χρησιμοποιώντας τοπικά χαρακτηριστικά BRISK.



Σχήμα 5.13: **Αριστερά προς δεξιά:** Η προβολή του γραφικού μοντέλου στη σκηνή. Η εκτίμηση θέσης της κάμερας έγινε χρησιμοποιώντας τοπικά χαρακτηριστικά SURF.



Σχήμα 5.14: Αριστερά προς δεξιά: Η προβολή του γραφικού μοντέλου στη σκηνή. Η εκτίμηση θέσης της κάμερας έγινε χρησιμοποιώντας τοπικά χαρακτηριστικά SIFT.



## Κεφάλαιο 6

# Συμπεράσματα-Μελλοντικές επεκτάσεις

### 6.1 Συμπεράσματα

Στόχος της παρούσας διπλωματικής αποτέλεσε η υλοποίηση μια εφαρμογής Επαυξημένης πραγματικότητας, εκτιμώντας τη θέση της κάμερας χωρίς καμία πρότερη γνώση του περιβάλλοντος σε πραγματικό χρόνο.

Αρχικά ο στόχος της εφαρμογής να τρέχει σε πραγματικό χρόνο έχει επιτευχθεί, όπως παρουσιάζεται και στον Πίνακα 5.2, μόνο για τα χαρακτηριστικά BRISK και LibViso2. Παρόλα αυτά, αν και με τη χρήση των τοπικών χαρακτηριστικών BRISK η εφαρμογή είναι αποδοτικότερη υπολογιστικά, η εκτίμηση θέσης της κάμερας είναι συνήθως χειρότερη από ότι με τα χαρακτηριστικά LibViso2. Το γεγονός αυτό οφείλεται κυρίως στο μικρό αριθμό *inliers* που ανιχνεύεται μεταξύ των διαδοχικών καρέ, όπως επίσης και στη μη ομοιόμορφη κατανομή των τοπικών χαρακτηριστικών στο επίπεδο της εικόνας. Επομένως, η χρήση των χαρακτηριστικών της πλατφόρμας LibViso2 δίνει καλύτερη εκτίμηση θέσης της κάμερας σε χρόνο που μπορεί να χαρακτηριστεί “πραγματικός”. Αξιολογώντας την εφαρμογή ως προς τα σφάλματα ευθυγράμμισης, τα δυναμικά σφάλματα έχουν εξαιρεθεί καθώς η σκηνή και το γραφικό μοντέλο παρουσιάζονται στο χρήστη από το δευτερεύον νήμα ταυτόχρονα. Επίσης, ο έλεγχος που πραγματοποιείται για το αν υπάρχει αρκετή μετατόπιση μεταξύ των δύο διαδοχικών καρέ δεν επιτρέπει στο γραφικό αντικείμενο να τρεμοπαίζει στην οθόνη του χρήστη (flickering).

Τα στατικά σφάλματα, λόγω λανθασμένης εκτίμησης θέσης, παρόλα αυτά αποτελούν πρόβλημα για την εφαρμογή. Δυστυχώς, σε περίπτωση αποτυχίας να ανανεωθεί η θέση της κάμερας για διαδοχικά ζεύγη καρέ η ευθυγράμμιση των δύο κόσμων δεν επανέρχεται. Η αποτυχία εκτίμησης της καινούργιας θέσης της κάμερας μπορεί να οφείλεται στους παρακάτω λόγους:

- Αποτυχία ανίχνευσης του ελάχιστου πλήθους *inliers*. Η αποτυχία αυτή μπορεί να οφείλεται σε γρήγορη μετατόπιση της κάμερας ή λήψη θολής εικόνας από την κάμερα.
- Περιστροφή της κάμερας σε ακραίες τιμές γωνιών.

Επομένως, η ανάπτυξη μιας εφαρμογής Επαυξημένης πραγματικότητας με τις μεθόδους που μελετήθηκαν είναι εφικτή. Οι μέθοδοι που χρησιμοποιήθηκαν αποτελούν τη βάση αλλά χρειάζεται περαιτέρω μελέτη των μεθόδων για εκτίμηση θέσης της κάμερας, ούτως ώστε να επιτευχθεί η ακρίβεια που απαιτεί μια τέτοια εφαρμογή.

## 6.2 Μελλοντικές επεκτάσεις

Όπως αναφέρθηκε, η μεθοδολογία πάνω στην οποία στηρίχθηκε η εφαρμογή για επαύξηση της πραγματικότητας περιέχει πολλά σφάλματα, όχι τόσο δυναμικά, αλλά κυρίως στατικά. Τα σφάλματα αυτά θέτουν υπό αμφισβήτηση την ταύτιση του εικονικού και πραγματικού κόσμου και επομένως καταρρίπτουν την ψευδαίσθηση της επαύξησης της πραγματικότητας.

Αρχικά κρίνεται απαραίτητη η βελτίωση της συμπεριφοράς του συστήματος στα στατικά σφάλματα. Αυτό θα μπορούσε να βελτιωθεί με τη χρήση κάποιων από τις παρακάτω μεθόδους ή ακόμα και με υβριδικό μοντέλο αυτών.

- Εισαγωγή κυρίαρχου επιπέδου: Μέτα από μία αρχική εκτίμηση θέσης της κάμερας και κατ'επέκταση των 3Δ συντεταγμένων των ταυτισμένων σημείων στο χώρο θα μπορούσε να δημιουργηθεί ένα εικονικό επίπεδο πάνω στο οποίο τα εικονικά αντικείμενα θα προβάλλονται. Χρησιμοποιώντας το επίπεδο αυτό ως το σύστημα συντεταγμένων του κόσμου, η σχετική θέση της κάμερας θα βρίσκεται πάντα ως προς το εικονικό αντικείμενο και όχι ως προς το πρώτο καρέ μειώνοντας έτσι τα στατικά σφάλματα [30].
- Επίσης η χρησιμοποίηση της εκτίμησης του Θεμελιώδους πίνακα  $F$  με τη μέθοδο RANSAC ως αρχική εκτίμηση και ο επανυπολογισμός του με τη μέθοδο εύρωστης εκτίμησης Μ-εκτιμητών (M-estimators) θα μείωνε σημαντικά τα σφάλματα υπολογισμού υπό θορυβώδεις μετρήσεις. Ακόμη θα μπορούσε να χρησιμοποιηθεί το κριτήριο ελαχιστοποίησης της επαναπροβολής των σημείων (reprojection error), αντικαθιστώντας το γεωμετρικό σφάλμα Sampson [31] [32].
- Όπως παρατηρήθηκε, τα στατικά σφάλματα αυξάνονται σημαντικά όταν το σύστημα δεν καταφέρει να εκτιμήσει την κίνηση της κάμερας για δυο διαδοχικά ή μία σειρά από διαδοχικά καρέ. Μια λύση στο πρόβλημα αυτό θα ήταν η χρησιμοποίηση των εκτεταμένων φίλτρων Kalman (Extended Kalman Filters) για να γίνεται μία πρόβλεψη θέσης της κάμερας όταν δεν είναι δυνατή η εκτίμηση της [33].

Τέλος, μία πολύ σημαντική προοπτική κάθε συστήματος Επαυξημένης Πραγματικότητας είναι η δυνατότητα των εικονικών αντικειμένων να αλληλεπιδρούν σε πραγματικό χρόνο με τα πραγματικά αντικείμενα όταν αυτά βρεθούν μπροστά από τα εικονικά ή αλλάξουν δυναμικά τη θέση τους.





# Βιβλιογραφία

- [1] R. Azuma, Y. Baillet, R. Behringer, S. Feiner, S. Julier, and B. MacIntyre, “Recent advances in augmented reality,” *IEEE Computer Graphics and Applications*, vol. 21(6), pp. 34–47, 2001.
- [2] R. Krevelen and R. Poelman, “A survey of augmented reality technologies, applications and limitations,” *The International Journal of Virtual Reality*, vol. 9, pp. 1–20, 2010.
- [3] R. Azuma, “A survey of augmented reality,” *Teleoperators and Virtual Environments*, vol. 6, pp. 355–385, 1997.
- [4] O. Kutter, A. Aichert, C. Bichlmeier, J. Traub, S. Heining, and B. Ockert, “Real-time volume rendering for high quality visualization in augmented reality,” in *International Workshop on Augmented environments for Medical Imaging including Augmented Reality in Computer-aided Surgery*, 2008.
- [5] S. J. Henserson and S. K. Feiner, “Augmented reality in the psychomotor phase of a procedural task,” *Mixed and Augmented Reality, IEEE / ACM International Symposium on*, vol. 0, pp. 191–200, 2011.
- [6] T. Tuytelaars and K. Mikolajczyk, “Local invariant feature detectors: A survey,” *Foundations and Trends in Computer Graphics and Vision*, vol. 3, pp. 177–280, 2007.
- [7] F. Attneave, “Some informational aspects of visual perception,” *Psychological review*, vol. 61, pp. 183–193, 1954.
- [8] P. R. Beaudet, “Rotationally invariant image operators,” in *Proceedings of the International Joint Conference on Pattern Recognition*, pp. 579–583, 1978.
- [9] H. Moravec, *Obstacle Avoidance and Navigation in the Real World by a Seeing Robot Rover*. PhD thesis, Stanford University, 1980.
- [10] C. Harris and M. Stephens, “A combined corner and edge detector,” in *4th Alvey Vision Conference*, pp. 147–151, 1988.

- [11] T. Lindeberg, "Feature detection and ridge detection with automatic scale selection," *International Journal of Computer Vision*, vol. 30(2), pp. 79–116, 1998.
- [12] T. Lindeberg, "Scale-space for discrete signals," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12(1), pp. 234–254, 1990.
- [13] K. Mikolajczyk and C. Schmid, "Scale and affine invariant interest point detectors," *International Journal of Computer Vision*, vol. 1(60), pp. 63–86, 2004.
- [14] J. Matas, O. Chum, M. Urban, and T. Pajdla., "Robust wide-baseline stereo from maximally stable extremal regions," in *British Machine Vision Conference*, pp. 384–393, 2002.
- [15] D. G. Lowe, "Object recognition from local scale-invariant features," in *International Conference on Computer Vision*, pp. 1150–1157, 1999.
- [16] Y. Ke and R. Sukthankar, "Pca-sift: A more distinctive representation for local image descriptors," in *IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 230–235, 1998.
- [17] L. G. Brown, "A survey of image registration techniques," *ACM Computing Surveys*, vol. 24(4), pp. 325–376, 1992.
- [18] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool, "Speeded-up robust features (surf)," *International Journal of Computer Vision and Image Understanding*, vol. 110(3), pp. 346–359, 2008.
- [19] E. Rosten and T. Drummond, "Machine learning for highspeed corner detection," in *In Proceedings of the European Conference on Computer Vision (ECCV)*, 2006.
- [20] M. Sonka, V. Hlavac, and R. Boyle, *Image Processing, Analysis and Machine Vision*. CL-Engineering, 2007.
- [21] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.
- [22] Z. Zhang, "Determining the epipolar geometry and its uncertainty: A review," *International Journal of Computer Vision*, vol. 27(2), pp. 161–198, 1998.
- [23] M. Fischler and R. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Communications ACM*, vol. 24(6), pp. 381–395, 1981.
- [24] P. Sampson, "Fitting conic sections to 'very scattered' data: An iterative refinement of the bookstein algorithm," *Computer Vision, Graphics and Image Processing*, vol. 18, pp. 97–108, 1982.

- [25] H. L. Higgins, "A computer algorithm for reconstructing a scene from two projections," *Nature*, vol. 293, pp. 133–135, 1981.
- [26] S. Leutenegger, M. Chli, and R. Siegwart, "Brisk: Binary robust invariant scalable keypoints," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2011.
- [27] A. Geiger, J. Ziegler, and C. Stiller, "Stereoscan: Dense 3d reconstruction in real-time," in *IEEE Intelligent Vehicles Symposium*, 2011.
- [28] B. Kitt, A. Geiger, and H. Lategahn, "Visual odometry based on stereo image sequences with ransac-based outlier rejection scheme," in *IEEE Intelligent Vehicles Symposium*, 2010.
- [29] Z. Zhang, R. Deriche, O. Faugeras, and Q.-T. Luong, "A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry," *Artificial Intelligence*, vol. 78(1-2), pp. 87–119, 1995.
- [30] G. Klein and D. Murray, "Parallel tracking and mapping for small AR workspaces," in *Proc. Sixth IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR '07)*, 2007.
- [31] P. Huber, *Robust Statistics*. New York, 1987.
- [32] L. Vacchetti, V. Lepetit, and P. Fua, "Stable real-time 3d tracking using online and offline information," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, pp. 1385–1391, 2004.
- [33] D. Koller, K. Daniilidis, and H.-H. Nagel, "Model-based object tracking in monocular image sequences of road traffic scenes," *International Journal of Computer Vision*, vol. 10, pp. 257–281, 1993.