



**Εθνικό Μετσόβιο Πολυτεχνείο**  
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών  
Τομέας Σημάτων, Ελέγχου και Ρομποτικής  
Όραση Υπολογιστών, Επικοινωνία Λόγου και Επεξεργασία Σημάτων

**Τρισδιάστατη Ανακατασκευή Αντικειμένων και Κτηρίων  
με Γεωμετρία Πολλαπλών Όψεων**

**Διπλωματική Εργασία**

της

**Γεωργίας Γκιοζάρη**

**Επιβλέπων:** Πέτρος Μαραγκός  
Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούλιος 2010





Εθνικό Μετσόβιο Πολυτεχνείο  
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών  
Τομέας Σημάτων, Ελέγχου και Ρομποτικής  
Όραση Υπολογιστών, Επικοινωνία Λόγου και Επεξεργασία Σημάτων

## Τρισδιάστατη Ανακατασκευή Αντικειμένων και Κτηρίων με Γεωμετρία Πολλαπλών Όψεων

### Διπλωματική Εργασία

της

**Γεωργίας Γκιοζάρη**

**Επιβλέπων:** Πέτρος Μαραγκός  
Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 20<sup>η</sup> Ιουλίου, 2010.

*(Υπογραφή)*

*(Υπογραφή)*

*(Υπογραφή)*

.....  
Πέτρος Μαραγκός  
Καθηγητής Ε.Μ.Π.

.....  
Νικόλαος Μήτρου  
Καθηγητής Ε.Μ.Π.

.....  
Γεράσιμος Ποταμιάνος  
Διευθυντής Ερευνών  
ΕΚΕΦΕ “Δημόκριτος”

Αθήνα, Ιούλιος 2010

*(Υπογραφή)*

.....

**Γεωργία Γκιοζάρη**

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

© 2010– All rights reserved



Εθνικό Μετσόβιο Πολυτεχνείο  
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών  
Τομέας Σημάτων, Ελέγχου και Ρομποτικής  
Όραση Υπολογιστών, Επικοινωνία Λόγου και Επεξεργασία Σημάτων

Copyright © – All rights reserved Γεωργία Γκιοζάρη, 2010.  
Με επιφύλαξη παντός δικαιώματος.

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.



# Ευχαριστίες

Θα ήθελα να ευχαριστήσω τον καθηγητή κ. Πέτρο Μαραγκό για την επίβλεψη αυτής της διπλωματικής εργασίας και για την ευκαιρία που μου έδωσε να την εκπονήσω στο εργαστήριο Συστημάτων Όρασης Υπολογιστών, Επικοινωνίας Λόγου και Επεξεργασίας Σημάτων. Ο χρόνος του και η καθοδήγησή του στα πλαίσια της παρούσας εργασίας αλλά και κατά τη διδασκαλία των μαθημάτων της Όρασης Υπολογιστών και της Αναγνώρισης Προτύπων μου έδωσαν την ευκαιρία να γνωρίσω αυτούς τους τομείς της επιστήμης. Θα ήθελα, επιπρόσθετα, να ευχαριστήσω θερμά τον Γιώργο Παπανδρέου και τον Ιάσονα Κόκκινο για την πολύτιμη βοήθειά τους και τις ιδέες που μοιράστηκαν μαζί μου κατά τη διάρκεια της εκπόνησης της εργασίας.

Εκτός του επιβλέποντα μου, θα ήθελα επίσης να ευχαριστήσω τους καθηγητές κ. Γεώργιο Παπαβασιλόπουλο και κ. Κώστα Τζαφέστα, με τους οποίους είχα την ευκαιρία να συνεργαστώ κατά τη διάρκεια της φοίτησής μου. Θα ήθελα να τους ευχαριστήσω θερμά για την εξαιρετική διδασκαλία τους, που αποτέλεσε αιτία να ασχοληθώ με το αντικείμενο του αυτομάτου ελέγχου και της ρομποτικής, καθώς και για τις συστάσεις τους στην προσπάθειά μου για επιδίωξη μεταπτυχιακών σπουδών στο εξωτερικό.

Θα ήθελα να ευχαριστήσω τον καθηγητή κ. Νικόλαο Μήτρου καθώς και τον διευθυντή ερευνών του Ε.Κ.Ε.Φ.Ε. "Δημόκριτος" κ. Γεράσιμο Ποταμίανο για το χρόνο τους και τη συμμετοχή τους στην εξεταστική επιτροπή της παρούσας διπλωματικής εργασίας.

Τέλος, θα ήθελα να ευχαριστήσω τους γονείς μου για την καθοδήγηση και την ηθική συμπαράσταση που μου προσέφεραν όλα αυτά τα χρόνια.





# Περίληψη

Η τρισδιάστατη ανακατασκευή από διαθέσιμες δισδιάστατες εικόνες αποτελεί έντονο πεδίο έρευνας στον τομέα της ΟΥ. Στόχος είναι η ακριβής αλλά και η οπτικά ικανοποιητική 3D αναπαράσταση του υπό εξέταση αντικειμένου, όπως αυτό απεικονίζεται από πολλαπλές όψεις. Τα στάδια για την επίτευξη του σκοπού αυτού περιλαμβάνουν την ανίχνευση χαρακτηριστικών από τις εικόνες, τη μελέτη της γεωμετρίας που συνδέει το δισδιάστατο με τον τρισδιάστατο χώρο καθώς και τη γνώση των μαθηματικών μοντέλων των καμερών.

Στην παρούσα εργασία, στόχος είναι η μελέτη και η εφαρμογή μεθόδων 3D ανακατασκευής από περιορισμένα σε πλήθος δεδομένα, συγκριτικά με τις απαιτήσεις των πιο γνωστών ως σήμερα αλγορίθμων. Για το σκοπό αυτό, αναπτύσσεται το θεωρητικό υπόβαθρο της προβολικής γεωμετρίας, της επιπολικής γεωμετρίας καθώς και της δομής από κίνηση (structure from motion), το οποίο αποτελεί βασικό στήριγμα για την πραγματοποίηση 3D ανακατασκευής. Εξετάζονται, υλοποιούνται και αξιολογούνται τρεις μέθοδοι για την εξαγωγή χαρακτηριστικών από εικόνες. Εν συνεχεία, υλοποιούνται και εφαρμόζονται αλγόριθμοι ανακατασκευής ενός αντικειμένου από πολλαπλές όψεις του με στόχο την όσο το δυνατό καλύτερη εκτίμηση του σχήματος του από ένα διασκορπισμένο νέφος σημείων. Από τις εικόνες, εξάγονται οι ακμές του αντικειμένου και υλοποιείται ένας αλγόριθμος για την αναπαράστασή τους στον 3D χώρο. Οι αλγόριθμοι και τα αποτελέσματά τους παρουσιάζονται στην εργασία.

Η κύρια συνεισφορά της παρούσας εργασίας σχετίζεται με την εύρεση, υλοποίηση και εφαρμογή ενός γρήγορου αλγορίθμου για την 3D ανακατασκευή και οπτική αναπαράσταση κτηρίων από δύο όψεις με απομακρυσμένα οπτικά κέντρα και με περιορισμένη γνώση συνδετικών χαρακτηριστικών μεταξύ των δύο εικόνων. Ο αλγόριθμος και τα αποτελέσματά του παρουσιάζονται στην εργασία.

## Λέξεις Κλειδιά

όραση υπολογιστών, τρισδιάστατη ανακατασκευή, οπτική αναπαράσταση, προβολική γεωμετρία, επιπολική γεωμετρία, δομή από κίνηση, μοντέλα καμερών, σημεία ενδιαφέροντος, νέφος σημείων, τρισδιάστατη ανακατασκευή ακμών, υπολογιστική γεωμετρία, ανίχνευση επιπέδων, ομογραφικός μετασχηματισμός.



# Abstract

Three-dimensional reconstruction from available two-dimensional images is an intense field of study in Computer Vision. The goal is the exact and the satisfying visual representation of the 3D object generated from multiple views. The accomplishment of this purpose includes study in many fields, such as feature extraction from images, the study of the geometry that connects the 3D and 2D space and the knowledge of the mathematical models of the cameras.

In this diploma thesis, the aim is to study and implement methods of 3D reconstruction from limited data, compared to the size of the data base required by the most popular algorithms up to date. For this reason, the theoretical foundation of projective geometry, epipolar geometry and structure from motion is presented. Three different methods for feature extraction from images are studied, implemented and evaluated. Next, algorithms for the 3D reconstruction of an object from multiple images are implemented and applied, aiming at the best possible shape estimation from a scattered point cloud. The edges of the object are extracted from the images and an algorithm is implemented for the representation of the edges in the 3D space. The algorithms and their results are presented in this study

The main contribution of this study is associated with finding, implementing and applying a fast algorithm for the 3D reconstruction and visualization of buildings from two views with remote optical centers and with limited knowledge of matching features between the images. The algorithm and its results are presented in this study.

## Keywords

computer vision, three-dimensional reconstruction, visual representation, projective geometry, epipolar geometry, structure from motion, camera models, keypoints, point cloud, 3D edge reconstruction, computational geometry, plane detection, homography



# Περιεχόμενα

Ευχαριστίες	1
Περίληψη	3
Abstract	5
<b>Περιεχόμενα</b>	<b>9</b>
<b>Κατάλογος σχημάτων</b>	<b>12</b>
<b>Κατάλογος πινάκων</b>	<b>13</b>
<b>1 Εισαγωγή</b>	<b>15</b>
1.1 Όραση Υπολογιστών . . . . .	15
1.2 Το περιεχόμενο της παρούσας εργασίας . . . . .	16
<b>2 Προβολική Γεωμετρία</b>	<b>19</b>
2.0.1 Προβολικός Χώρος . . . . .	19
2.0.2 Συντεταγμένες . . . . .	19
2.1 Προβολική Γεωμετρία και Μετασχηματισμοί στο 2Δ Χώρο . . . . .	20
2.1.1 Το 2Δ Προβολικό Επίπεδο . . . . .	20
2.1.1.1 Σημεία και Ευθείες . . . . .	20
2.1.1.2 Ιδανικά Σημεία και η Ευθεία στο Άπειρο . . . . .	21
2.1.1.3 Κωνικές Τομές . . . . .	22
2.1.2 Προβολικοί Μετασχηματισμοί . . . . .	23
2.1.3 Ιεραρχία Μετασχηματισμών . . . . .	24
2.2 Προβολική Γεωμετρία και Μετασχηματισμοί στο 3Δ Χώρο . . . . .	25
2.2.1 Σημεία και Προβολικοί Μετασχηματισμοί . . . . .	25
2.2.2 Επίπεδα, Ευθείες και Προβολικοί Μετασχηματισμοί . . . . .	25
2.2.2.1 Επίπεδα . . . . .	25
2.2.2.2 Ευθείες . . . . .	26
2.2.3 Ιεραρχία Μετασχηματισμών . . . . .	26
2.3 Μοντέλα Καμερών . . . . .	26
2.3.1 Πεπερασμένη Κάμερα . . . . .	27
2.3.2 Ιδιότητες Προβολικής Κάμερας . . . . .	30
2.3.2.1 Ανατομία Κάμερας . . . . .	30
2.3.2.2 Δράση της Προβολικής Κάμερας στα Σημεία . . . . .	32
2.3.2.3 Βάθος Σημείων . . . . .	32
2.3.3 Η Κάμερα στο Άπειρο . . . . .	33

<b>3</b>	<b>Επιπολική Γεωμετρία και Δομή από Κίνηση</b>	<b>35</b>
3.1	Επιπολική Γεωμετρία . . . . .	35
3.1.1	Ο Θεμελιώδης Πίνακας . . . . .	37
3.1.1.1	Αλγεβρικός Υπολογισμός του Θεμελιώδους Πίνακα . . . . .	38
3.1.1.2	Ιδιότητες του Θεμελιώδους Πίνακα . . . . .	38
3.1.2	Ο Σημαντικός Πίνακας . . . . .	39
3.2	Δομή από Κίνηση . . . . .	40
3.2.1	Triangulation . . . . .	40
3.2.2	Δομή από Κίνηση δύο Όψεων . . . . .	42
<b>4</b>	<b>Ημιαυτόματη Μέθοδος Ανίχνευσης Σημείων Φυγής</b>	<b>47</b>
4.1	Ανίχνευση Ακμών και Ευθειών . . . . .	47
4.1.1	Ανίχνευση Ακμών . . . . .	47
4.1.2	Ανίχνευση ευθειών . . . . .	48
4.2	Ε-σημαντικά Ευθύγραμμα Τμήματα . . . . .	52
4.3	Ανίχνευση Σημείων Φυγής . . . . .	56
4.4	Ε-σημαντικά Σημεία Φυγής . . . . .	56
4.5	Τελικά Αποτελέσματα . . . . .	61
<b>5</b>	<b>Τρισδιάστατη Ανακατασκευή Αντικειμένων</b>	<b>65</b>
5.1	Εύρεση Σημείων Ταιριάσματος . . . . .	66
5.1.1	SIFT Σημεία Ενδιαφέροντος . . . . .	66
5.1.1.1	Ανίχνευση Σημείων Ενδιαφέροντος . . . . .	66
5.1.1.2	Τοπικοί Περιγραφείς . . . . .	69
5.1.1.3	Ανίχνευση Σημείων Ταιριάσματος . . . . .	70
5.1.2	MSER Περιοχές Ενδιαφέροντος . . . . .	70
5.1.3	Μορφολογικές Περιοχές Ενδιαφέροντος . . . . .	71
5.1.4	Σύγκριση των Μεθόδων Ανίχνευσης Χαρακτηριστικών . . . . .	74
5.2	Υπολογισμός του Θεμελιώδους Πίνακα . . . . .	78
5.2.1	Βασικές Σχέσεις . . . . .	78
5.2.2	Ο Αλγόριθμος 8-σημείων . . . . .	79
5.2.3	Σφάλμα Sampson . . . . .	80
5.2.4	Ο Τελικός Αλγόριθμος . . . . .	81
5.3	Ανακατασκευή Τρισδιάστατου Νέφους Σημείων . . . . .	82
5.3.1	Προσδιορισμός των Παραμέτρων της Κάμερας . . . . .	82
5.3.2	Προσδιορισμός της 3Δ Θέσης των Σημείων . . . . .	82
5.4	Εκτίμηση Συνεχούς Τρισδιάστατης Επιφάνειας . . . . .	83
5.5	Τρισδιάστατη Ανακατασκευή Ακμών . . . . .	85
5.6	Τελικά Αποτελέσματα . . . . .	86
<b>6</b>	<b>Τρισδιάστατη Ανακατασκευή Κτηρίων</b>	<b>93</b>
6.1	Τρισδιάστατη Ανακατασκευή Κτηρίων με τη Μέθοδο της Ανίχνευσης Επιπέδων . . . . .	93
6.1.1	Υπολογισμός Θεμελιώδους Πίνακα . . . . .	93
6.1.2	Εκτίμηση των Εξωτερικών Παραμέτρων της Κάμερας . . . . .	95
6.1.3	Ανακατασκευή Τρισδιάστατου Νέφους Σημείων . . . . .	96
6.1.4	Ανίχνευση Επιπέδων . . . . .	96
6.1.5	Απόδοση Υφής . . . . .	98
6.1.6	Τελικά Αποτελέσματα . . . . .	100

---

6.2	Εφαρμογή Τρισδιάστατης Ανακατασκευής με Δυνατότητα Περιήγησης στα Δεδομένα ενός Μνημείου . . . . .	103
<b>7</b>	<b>Επίλογος</b>	<b>107</b>
7.1	Συμβολή της διπλωματικής εργασίας . . . . .	107
7.2	Μελλοντικές κατευθύνσεις . . . . .	107
<b>A'</b>	<b>Πράξεις πινάκων</b>	<b>109</b>
A'.1	Ιδιότητες αντισυμμετρικών πινάκων . . . . .	109
A'.2	Διαφορισμοί πινάκων . . . . .	110
A'.3	SVD ανάλυση . . . . .	110
<b>B'</b>	<b>Ο RANSAC αλγόριθμος</b>	<b>111</b>
<b>Γ'</b>	<b>Delaunay Triangulation</b>	<b>113</b>
	<b>Βιβλιογραφία</b>	<b>114</b>





# Κατάλογος σχημάτων

2.1	Το μοντέλο του προβολικού επίπεδου . . . . .	22
2.2	Το μοντέλο της pinhole κάμερας . . . . .	27
2.3	Τα συστήματα συντεταγμένων της εικόνας και της κάμερας . . . . .	28
2.4	Τα σημεία των στηλών του πίνακα της κάμερας $\mathbf{p}_i, i = 1, \dots, 3$ . . . . .	31
2.5	Δύο εκ των τριών επιπέδων που ορίζονται από τις γραμμές του πίνακα $P$ της κάμερας	31
3.1	Η διάταξη των δύο καμερών και τα επίπεδα των εικόνων . . . . .	36
3.2	Επιπολικές ευθείες . . . . .	36
3.3	Η οικογένεια επιπολικών επιπέδων . . . . .	37
3.4	Η μέθοδος του τριγωνισμού . . . . .	42
3.5	Η επιπολική γεωμετρία . . . . .	43
3.6	Οι τέσσερις δυνατοί συνδυασμοί της βαθμονομημένης ανακατασκευής από τον πίνακα $E$ . . . . .	46
4.1	Εικόνα εισόδου και οι ακμές της σύμφωνα με τον αλγόριθμο του Canny . . . . .	48
4.2	Ο αλγόριθμος διαχωρισμού ακμών . . . . .	50
4.3	PCA σε δισδιάστατη γκαουσιανή κατανομή δεδομένων . . . . .	51
4.4	Το αποτέλεσμα της μεθόδου ανίχνευσης ευθειών με τη βοήθεια του PCA . . . . .	53
4.5	Μία περιοχή φυγής $V_j$ και ένα σύνολο ευθειών που διέρχονται από αυτήν . . . . .	56
4.6	Περιοχές φυγής . . . . .	57
4.7	Κώνος αβεβαιότητας ενός ευθύγραμμου τμήματος . . . . .	58
4.8	Τα τμήματα του επιπέδου της εικόνας με ίση πιθανότητα να διακρίνει τις περιοχές φυγής για διάφορες τιμές της γωνιακής ακρίβειας $\theta$ . Μόνο οι εξωτερικές περιοχές φαίνονται. Οι εσωτερικές περιοχές σχηματίζουν ένα τετράγωνο πλέγμα. Οι άξονες αντιπροσωπεύουν αποστάσεις σχετικές με την ακτίνα της κυκλικής περιοχής της εικόνας $\Omega$ . Το σχήμα αυτό προέρχεται από το [3]. . . . .	59
4.9	Κατασκευή των εξωτερικών περιοχών φυγής . . . . .	59
4.10	Στο διάγραμμα απεικονίζεται ο ελάχιστος απαιτούμενος αριθμός συντρεχουσών ευθειών ώστε οι περιοχές φυγής να είναι 1-σημαντικές ως συνάρτηση του πλήθους των συνολικών ευθειών $N$ , για διάφορες τιμές της παραμέτρου $\theta$ . Το σχήμα αυτό προέρχεται από το [3]. . . . .	61
4.11	Η αρχική εικόνα εισόδου του αλγορίθμου ανίχνευσης σημείων φυγής (πηγή: google images). . . . .	62
4.12	Η εικόνα ακμών της εικόνας εισόδου. . . . .	62
4.13	Οι ευθείες που ανιχνεύονται με τη μέθοδο του PCA στην εικόνα ακμών . . . . .	63
4.14	Οι 1-σημαντικές ευθείες. . . . .	63
4.15	Τα 1-σημαντικά σημεία φυγής . . . . .	64
5.1	Από το σύνολο δεδομένων για την 3Δ ανακατασκευή . . . . .	66

5.2	Τρόπος κατασκευής των συναρτήσεων διαφοράς γκαουσιανών . . . . .	68
5.3	Τα τοπικά ακρότατα των συναρτήσεων $D(x, y, \sigma)$ . . . . .	68
5.4	Ο τοπικός SIFT περιγραφέας των σημείων ενδιαφέροντος . . . . .	70
5.5	Οι MSER περιοχές της εικόνας εισόδου . . . . .	72
5.6	Πολλαπλή κατωφλιοποίηση της συνάρτησης κορυφών . . . . .	73
5.7	Το αποτέλεσμα ανίχνευσης περιοχών ενδιαφέροντος με τον top hat μετασχηματισμό . . . . .	74
5.8	Οι περιοχές ενδιαφέροντος όπως προκύπτουν από τον γενικευμένο top hat τελεστή με χρήση reconstruction opening . . . . .	75
5.9	Σύγκριση των τριών αλγορίθμων ανίχνευσης χαρακτηριστικών για εφαρμογές στερέωσης (1) . . . . .	76
5.10	Σύγκριση των τριών αλγορίθμων ανίχνευσης χαρακτηριστικών για εφαρμογές στερέωσης (2) . . . . .	76
5.11	Σύγκριση των τριών αλγορίθμων ανίχνευσης χαρακτηριστικών για εφαρμογές στερέωσης (3) . . . . .	77
5.12	Σύγκριση των τριών αλγορίθμων ανίχνευσης χαρακτηριστικών για εφαρμογές στερέωσης (4) . . . . .	78
5.13	Η διαδικασία moving least squares . . . . .	84
5.14	SIFT ζεύγη σημείων ταιριάσματος . . . . .	86
5.15	Τα ζεύγη σημείων που αποτελούν inliers . . . . .	87
5.16	Το 3Δ νέφος σημείων . . . . .	88
5.17	Το αποτέλεσμα της μεθόδου MLS . . . . .	89
5.18	Ανίχνευση και αντιστοίχιση ακμών . . . . .	90
5.19	Το αποτέλεσμα της 3Δ ανακατασκευής ακμών με είσοδο 6 εικόνες της βάσης δεδομένων. Οι ακμές στις 2Δ εικόνες προέρχονται από τα Primal Sketch Tokens (βλ. [23]). . . . .	91
6.1	Τα δεδομένα για την 3Δ ανακατασκευή κτηρίων . . . . .	94
6.2	Ομογραφικός μετασχηματισμός του επιπέδου της εικόνας ως προς την επιφάνεια του 3Δ κόσμου . . . . .	98
6.3	Τα matchpoints στις δύο εικόνες του κτηρίου προς ανακατασκευή . . . . .	100
6.4	3Δ αναπαράσταση των σημείων ταξινομημένα σε επίπεδα . . . . .	100
6.5	3Δ αναπαράσταση των κτηρίων αποτελούμενα από τριγωνικά επίπεδα . . . . .	101
6.6	Το διορθωμένο επίπεδο ως αναπαράσταση υφής για την αντίστοιχη επίπεδη επιφάνεια του 3Δ χώρου . . . . .	101
6.7	Η 3Δ ανακατασκευή των κτηρίων . . . . .	102
6.8	Πλήρης 3Δ ανακατασκευή των κτηρίων . . . . .	102
6.9	Αδυναμία του αλγορίθμου ανακατασκευής κτηρίων . . . . .	102
6.10	Δεδομένα για την γρήγορη 3Δ ανακατασκευή μνημείου . . . . .	103
6.11	Τα ζεύγη SIFT σημείων μεταξύ των δύο εικόνων του ναού . . . . .	104
6.12	Το 3Δ νέφος των σημείων που αναπαριστούν τον ναό . . . . .	105
6.13	Το αποτέλεσμα της πρόχειρης 3Δ ανακατασκευής και οι θέσεις των καμερών . . . . .	105
6.14	Η εφαρμογή περιήγησης στις εικόνες της βάσης δεδομένων με βάση τη θέση των αντίστοιχων οπτικών κέντρων . . . . .	106

# Κατάλογος πινάκων

2.1	Γεωμετρικές ιδιότητες των ειδικών κατηγοριών 2Δ μετασχηματισμών . . . . .	24
2.2	Γεωμετρικές ιδιότητες των ειδικών κατηγοριών 3Δ μετασχηματισμών . . . . .	27



# Κεφάλαιο 1

## Εισαγωγή

### 1.1 Όραση Υπολογιστών

Η Όραση Υπολογιστών ασχολείται με την εξαγωγή χρήσιμων πληροφοριών από εικόνες ή ακολουθίες εικόνων. Ανώτερος στόχος αυτών που εργάζονται στην περιοχή είναι η κατασκευή συστημάτων τα οποία θα αντιλαμβάνονται τον ορατό κόσμο με ανθρώπινη ευφυΐα. Αν και η ερευνα στην Όραση Υπολογιστών έχει κάνει τα τελευταία χρόνια σημαντική πρόοδο, απέχουμε ακόμα πάρα πολύ από την εκπλήρωση αυτού του οράματος.

Η Όραση Υπολογιστών (βλ. [28] και [17]) έκανε τα πρώτα της βήματα τη δεκαετία του '60, ως ένα παρακλάδι της Τεχνητής Νοημοσύνης. Προκειμένου η ερευνα στον τομέα της ΟΥ να δώσει τα επιθυμητά αποτελέσματα, οι ερευνητές γρήγορα χώρισαν το συνολικό πρόβλημα σε υποπροβλήματα υψηλού επιπέδου. Αυτά με τη σειρά τους αποσυντέθηκαν σε προβλήματα μεσαίου επιπέδου, τα οποία τελικά διασπάστηκαν σε προβλήματα χαμηλού επιπέδου. Παρακάτω ακολουθεί ένας σύνομος κατάλογος επιμέρους προβλημάτων της ΟΥ, από τα πιο υψηλού στα πιο χαμηλού επιπέδου

- Μοντελοποίηση φυσικών και τεχνητών σχημάτων
- Αναγνώριση/κατηγοριοποίηση αντικειμένων
- Κατάτμηση εικόνων σε ομογενείς περιοχές
- Εκτίμηση 3D σχήματος από την κίνηση ή/και την σκίαση
- Εντοπισμός και εκτίμηση ταχύτητας κινούμενου αντικειμένου
- Στερέωση
- Ανάλυση και σύνθεση υφής
- Ανίχνευση ακμών
- Ομαλοποίηση εικόνων και ανάλυση σε πολλαπλές κλίμακες
- Μοντελοποίηση και ανάλυση χρώματος
- Σχηματισμός εικόνων από τον άνθρωπο και τη μηχανή

Εξαιτίας της πολυπλοκότητας του αντικειμένου και του γεγονότος ότι η ΟΥ είναι ένας έντονα αναπτυσσόμενος τομέας έρευνας, η ΟΥ χρησιμοποιεί μια πληθώρα μεθόδων και εργαλείων, δανεισμένων από πολλές και διαφορετικές μεταξύ τους επιστημονικές περιοχές. Η ερευνα στην περιοχή της ΟΥ έχει δεχτεί επιρροές από κλάδους όπως

- *Στατιστική* Στατιστικές μέθοδοι χρησιμοποιούνται για την ανάλυση εικόνων και για τη σύνθεση τους. Τα τελευταία χρόνια, η στατιστική χρησιμοποιείται σε μεγάλο βαθμό για σκοπούς αναγνώρισης και κατηγοριοποίησης αντικειμένων σε εικόνες και εκμάθηση συστημάτων για βέλτιστα αποτελέσματα.
- *Μαθηματική Μορφολογία* Η μαθηματική μορφολογία είναι μια μη-γραμμική γεωμετρική μεθοδολογία που χρησιμοποιείται στην ανάλυση εικόνων και στην όραση υπολογιστών. Βασίζεται στη θεωρία συνόλων, στην άλγεβρα πλεγμάτων (lattice algebra), στην κυρτή ανάλυση, στη στοχαστική και ολοκληρωτική γεωμετρία.
- *Θεωρία Γραφημάτων* Μέθοδοι των διακριτών μαθηματικών και της θεωρίας γράφων έχουν βρει τα τελευταία χρόνια εφαρμογή σε προβλήματα όρασης υπολογιστών, όπως π.χ. η κατάτμηση εικόνων σε ομογενείς περιοχές και η τρισδιάστατη ανακατασκευή.
- *Γεωμετρία* Διάφορα είδη γεωμετρίας έχουν διαδραματίσει σημαντικό ρόλο στην εξέλιξη της έρευνας στον τομέα της ΟΥ. Η προβολική γεωμετρία (βλ. [22] και [15]) άνοιξε το δρόμο για την τρισδιάστατη ανακατασκευή αντικειμένων από πολλαπλές εικόνες, ενώ η διαφορική και η υπολογιστική γεωμετρία έφεραν σε επαφή την όραση υπολογιστών με την υπολογιστική γραφική.

## 1.2 Το περιεχόμενο της παρούσας εργασίας

Σκοπός της συγκεκριμένης διπλωματικής εργασίας είναι η τρισδιάστατη ανακατασκευή αντικειμένων και κτηρίων που απεικονίζονται σε πολλαπλές εικόνες μιας βάσης δεδομένων. Ζητείται η εξαγωγή σχήματος από τις εικόνες καθώς και η αναπαράσταση της υψής του αντικειμένου για την ικανοποιητική οπτική αναπαράστασή του. Για την επίτευξη του σκοπού αυτού χρησιμοποιούνται μεθοδολογίες προερχόμενες από τα πεδία της προβολικής γεωμετρίας, της διαφορικής γεωμετρίας καθώς και αποτελέσματα γραμμικής άλγεβρας και πιθανοτήτων. Ζητούμενο δεν είναι μόνο η οπτική απεικόνιση των 3D αντικειμένων αλλά η ανάπτυξη αλγορίθμων για την πραγματοποίηση του σκοπού αυτού με περιορισμένο πλήθος εικόνων και έλλειψη γνώσης των παραμέτρων του προβλήματος. Μεγάλο μέρος της εργασίας κατέλαβε η πρακτική υλοποίηση των ιδεών και αλγορίθμων στον υπολογιστή.

Οι κατευθύνσεις και τα πεδία μελέτης για την πραγματοποίηση της παρούσας εργασίας αναφέρονται περιληπτικά παρακάτω

- Ημιαυτόματη ανίχνευση σημείων φυγής σε εικόνες. Για το σκοπό αυτό, χρησιμοποιήθηκε η έννοια της στατιστικής σημαντικότητας.
- Μελέτη και αξιολόγηση μεθόδων εξαγωγής χαρακτηριστικών για εφαρμογές στερέωσης. Έγινε σύγκριση μεταξύ των SIFT σημείων ενδιαφέροντος, των MSER περιοχών και των μορφολογικών περιοχών ενδιαφέροντος .
- Τρισδιάστατη ανακατασκευή νέφους σημείων από πολλαπλές όψεις ενός αντικειμένου και εκτίμηση συνεχούς επιφάνειας από το σύνολο των διασκορπισμένων σημείων.
- Τρισδιάστατη αναπαράσταση ακμών ενός αντικειμένου από ακμές στις δισδιάστατες εικόνες
- Τρισδιάστατη ανακατασκευή και οπτική αναπαράσταση κτηρίων.

Το κυρίως κείμενο είναι οργανωμένο σε 5 κεφάλαια. Στα πρώτα δύο κεφάλαια (Κεφάλαιο 2 και 3) εισάγεται ο αναγνώστης στον τομέα της προβολικής και επιπολικής γεωμετρίας αντίστοιχα. Στα

επόμενα τρία κεφάλαια (Κεφάλαια 4, 5 και 6) παρουσιάζονται οι εφαρμογές και οι υλοποιήσεις των αλγορίθμων που αποτέλεσαν το κύριο μέρος της παρούσας εργασίας.

Στο Κεφάλαιο 2 γίνεται συνοπτική και περιεκτική αναφορά στις ιδέες από το χώρο της προβολικής γεωμετρίας που χρησιμοποιούνται για την υλοποίηση της τρισδιάστατης ανακατασκευής. Συγκεκριμένα, παρουσιάζονται οι κυριότερες ιδέες από τη 2Δ και 3Δ προβολική γεωμετρία και αναπτύσσονται τα μοντέλα των καμερών που συνδέουν το 3Δ με το 2Δ χώρο.

Στο Κεφάλαιο 3 παρουσιάζονται οι βασικότερες έννοιες προερχόμενες από το πεδίο της επιπολικής γεωμετρίας. Επιπλέον, παρουσιάζεται ο τρόπος με τον οποίο οι ιδέες αυτές συνδέονται για την εξαγωγή πληροφοριών για τις παραμέτρους της κάμερας από πολλαπλές εικόνες.

Στο Κεφάλαιο 4 παρουσιάζεται ένας αλγόριθμος ημιαντόματης ανίχνευσης σημείων φυγής. Ο αλγόριθμος αυτός είναι ταχέος και το αποτέλεσμα του συμβάλει στη βαθμονόμηση των καμερών, όταν αυτή δεν είναι γνωστή.

Στο Κεφάλαιο 5 υλοποιείται ένας αλγόριθμος 3Δ ανακατασκευής ενός αντικειμένου από πολλαπλές εικόνες. Τα στάδια της ανακατασκευής περιλαμβάνουν τη δημιουργία ενός νέφους σημείων και την εκτίμηση συνεχούς επιφάνειας από το σύνολο των σημείων. Επιπλέον, παρουσιάζεται ένας τρόπος 3Δ ανακατασκευής των ακμών του αντικειμένου.

Στο Κεφάλαιο 6 παρουσιάζεται ένας εξαιρετικά ταχύς αλγόριθμος για την οπτική ανακατασκευή ενός κτηρίου από δύο μόνο εικόνες του κτηρίου αυτού. Ο αλγόριθμος περιλαμβάνει όχι μόνο την 3Δ αναπαράσταση του σχήματος αλλά και την απόδοση υψής στις όψεις του κτιρίου.





## Κεφάλαιο 2

# Προβολική Γεωμετρία

Στο κεφάλαιο αυτό εισάγονται οι κυριότερες ιδέες από το χώρο της προβολικής γεωμετρίας και σκοπός είναι να γίνει η σύνδεση τους με την όραση υπολογιστών. Συγκεκριμένα, δίνεται έμφαση στους μετασχηματισμούς του 2Δ και 3Δ κόσμου καθώς και στην γεωμετρία πολλαπλών όψεων. Οι ιδέες αυτές παρουσιάζονται αναλυτικά στα [22] και [15].

Οι προβολικοί μετασχηματισμοί είναι συχνοί στην καθημερινότητα. Σε μια φωτογραφία μπορεί κανείς να δει τετραγωνικά ή κυκλικά σχήματα που δεν ανταποκρίνονται στην πραγματικότητα. Ο μετασχηματισμός που αντιστοιχεί επίπεδα σχήματα του κόσμου στην εικόνα είναι ένα παράδειγμα προβολικού μετασχηματισμού.

Ο προβολικός μετασχηματισμός έχει τις παρακάτω ιδιότητες. Δεν διατηρεί το σχήμα, το μήκος των ευθυγράμμων τμημάτων, γωνίες, αποστάσεις και λόγους μηκών ευθυγράμμων τμημάτων. Δηλαδή, ένας κύκλος ενδέχεται να μετασχηματιστεί σε μία έλλειψη, ένα ορθογώνιο σε ρόμβο. Ωστόσο, η ιδιότητα που διατηρείται είναι αυτή της *ευθύτητας*. Δηλαδή, συνευθειακά σημεία θα μετασχηματιστούν σε συνευθειακά σημεία. Η τελευταία αυτή ιδιότητα της διατήρησης των ευθειών είναι αυτή που ουσιαστικά ορίζει τους προβολικούς μετασχηματισμούς.

### 2.0.1 Προβολικός Χώρος

Για τον ορισμό του *προβολικού χώρου* θα γίνει χρήση του Ευκλείδειου χώρου, για λόγους ευκολίας κατανόησης. Η Ευκλείδεια γεωμετρία περιγράφει γωνίες και σχήματα αντικειμένων. Ωστόσο, αποτυγχάνει να περιγράψει το σημείο τομής παράλληλων ευθειών. Συγκεκριμένα, δύο παράλληλες ευθείες τέμνονται στο *άπειρο*. Το *άπειρο*, όμως, δεν υπάρχει και αποτελεί απλά πλάσμα μιας βολικής φαντασίας. Είναι δυνατό να παρακαμφθεί το πρόβλημα αυτό επεκτείνοντας τον Ευκλείδειο χώρο προσθέτοντας τα σημεία στο άπειρο, όπου τέμνονται παράλληλες ευθείες. Τα σημεία αυτά καλούνται *ιδανικά σημεία*.

Η προσθήκη των *ιδανικών σημείων* μετασχηματίζει τον Ευκλείδειο χώρο στον *προβολικό χώρο*. Με άλλα λόγια, ο *προβολικός χώρος* είναι μια απλή επέκταση του Ευκλείδειου χώρου.

### 2.0.2 Συντεταγμένες

Ένα σημείο στον δισδιάστατο Ευκλείδειο χώρο περιγράφεται από το διατεταγμένο ζεύγος πραγματικών αριθμών  $(x, y)$ . Η προσθήκη μιας επιπλέον συνιστώσας, δηλαδή  $(x, y, 1)$ , δηλώνεται να αντιπροσωπεύει το ίδιο σημείο στο χώρο. Μια τέτοια αλλαγή δεν οδηγεί σε σφάλματα, καθώς είναι εύκολο να οδηγηθεί κανείς από τη μία περιγραφή στην άλλη. Στη συνέχεια, η παραπάνω αλλαγή διευρύνεται ορίζοντας ως εναλλακτική περιγραφή του σημείου οποιαδήποτε τριάδα της μορφής  $(kx, ky, k)$ , όπου  $k$  οποιοσδήποτε μη μηδενικός πραγματικός αριθμός. Και πάλι, είναι δυνατή η μετάβαση από την μία περιγραφή στην άλλη καθώς ο λόγος των δύο πρώτων συνιστω-

σών προς την τρίτη συνιστώσα της εναλλακτικής περιγραφής οδηγεί στο αρχικό σημείο  $(x, y)$ . Ουσιαστικά λοιπόν, σημεία περιγράφονται από *ισοδύναμες τάξεις* τριάδων, όπου δύο τριάδες είναι *ισοδύναμες* αν διαφέρουν κατά μία πολλαπλασιαστική σταθερά. Οι τριάδες που αποτελούν εναλλακτικούς περιγραφείς των σημείων ονομάζονται *ομογενείς συντεταγμένες*.

Επομένως σύμφωνα με τα παραπάνω, το σημείο  $(x, y, 1)$  αντιπροσωπεύει το σημείο  $(x, y)$ . Ποιο σημείο όμως αντιπροσωπεύει η τριάδα  $(x, y, 0)$ ; Το σημείο  $(\frac{x}{0}, \frac{y}{0})$  δεν είναι πεπερασμένο. Αυτά τα σημεία αποτελούν τα *ιδανικά σημεία*, δηλαδή τα σημεία στο *άπειρο*. Δηλαδή, τα *ιδανικά σημεία* περιγράφονται από ομογενείς συντεταγμένες με μηδενική την τελευταία συνιστώσα.

Οι ίδιες ιδέες που μετασχηματίζουν το δισδιάστατο Ευκλείδειο χώρο σε προβολικό χώρο με σημεία που περιγράφονται από ομογενείς συντεταγμένες, μπορούν να επεκταθούν σε οποιαδήποτε διάσταση. Ο Ευκλείδειος χώρος  $\mathbb{R}^n$  μπορεί να επεκταθεί στον προβολικό χώρο  $\mathbb{P}^n$  με περιγραφείς σημείων μέσω ομογενών διανυσμάτων. Για  $n = 2$  τα σημεία στο άπειρο σχηματίζουν μια ευθεία που καλείται *ευθεία στο άπειρο*, ενώ για  $n = 3$  σχηματίζουν επίπεδο που καλείται *επίπεδο στο άπειρο*.

## 2.1 Προβολική Γεωμετρία και Μετασχηματισμοί στο 2Δ Χώρο

Στην ενότητα αυτή μελετώνται οι ιδιότητες της προβολικής γεωμετρίας στον 2Δ χώρο. Συγκεκριμένα, μελετώνται οι περιγραφείς των σημείων, ευθειών, καμπυλών σε ομογενείς συντεταγμένες, και πως αυτές μετασχηματίζονται υπό προβολικούς μετασχηματισμούς. Εισάγονται οι έννοιες των *ιδανικών σημείων* και της *ευθείας στο άπειρο*. Στην ενότητα αυτή, έντονοι μικροί χαρακτήρες, όπως  $\mathbf{x}$ , συμβολίζουν ένα διάνυσμα στήλη, και το ανάστροφό του,  $\mathbf{x}^T$ , συμβολίζει ένα διάνυσμα γραμμή.

### 2.1.1 Το 2Δ Προβολικό Επίπεδο

Τα σημεία σε ένα επίπεδο περιγράφονται από το ζεύγος συντεταγμένων  $(x, y)$  στο  $\mathbb{R}^2$ . Οι ομογενείς συντεταγμένες του σημείου είναι  $(kx, ky, k)$  στο  $\mathbb{P}^2$ .

#### 2.1.1.1 Σημεία και Ευθείες

**Ομογενείς περιγραφείς ευθειών.** Η ευθεία σε ένα επίπεδο περιγράφεται από την εξίσωση  $ax + by + c = 0$ . Η επιλογή διαφορετικών τιμών για τις σταθερές  $a, b$  και  $c$  δίνει διαφορετικές ευθείες στο επίπεδο. Επομένως, μια ευθεία μπορεί να παρασταθεί ως ένα διάνυσμα  $(a, b, c)^T$ . Ωστόσο, το διάνυσμα αυτό δεν αποτελεί μοναδικό τρόπο περιγραφής της ευθείας αυτής, διότι η εξίσωση  $(ka)x + (kb)y + (kc) = 0$  και άρα το διάνυσμα  $(ka, kb, kc)$  περιγράφουν την ίδια ευθεία για κάθε μη μηδενικό  $k$ . Τα διανύσματα αυτά αν και διαφορετικά είναι *ισοδύναμα* καθώς παριστάνουν την ίδια ευθεία. Μια *ισοδύναμη τάξη* διανυσμάτων είναι γνωστή ως *ομογενές διάνυσμα*. Οποιοδήποτε διάνυσμα  $(a, b, c)^T$  είναι αντιπροσωπευτικό της *ισοδύναμης τάξης*.

**Ομογενείς περιγραφείς σημείων.** Ένα σημείο  $\mathbf{x} = (x, y)^T$  ανήκει στην ευθεία  $\mathbf{l} = (a, b, c)^T$  αν και μόνο αν  $ax + by + c = 0$ . Αυτό μπορεί να γραφεί και ως  $(x, y, 1)(a, b, c)^T = (x, y, 1)\mathbf{l} = 0$ . Να σημειωθεί ότι η για κάθε μη μηδενική σταθερά  $k$  και ευθεία  $\mathbf{l}$  η εξίσωση  $(kx, ky, k)\mathbf{l} = 0$  ισχύει αν και μόνο αν  $(x, y, 1)\mathbf{l} = 0$ . Επομένως, το σύνολο διανυσμάτων  $(kx, ky, k)^T$  αντιστοιχεί στο σημείο  $(x, y)$  στο  $\mathbb{R}^2$ . Για το λόγο αυτό, και τα σημεία και οι ευθείες μπορούν να παρασταθούν με ομογενείς συντεταγμένες. Ένα αυθαίρετο ομογενές διάνυσμα της μορφής  $\mathbf{x} = (x_1, x_2, x_3)^T$  στο  $\mathbb{P}^2$  αντιπροσωπεύει το σημείο  $(\frac{x_1}{x_3}, \frac{x_2}{x_3})$  στο  $\mathbb{R}^2$ .

Μια απλή εξίσωση καθορίζει αν ένα σημείο ανήκει στην ευθεία, συγκεκριμένα

**Αποτέλεσμα 2.1.** Ένα σημείο  $\mathbf{x}$  ανήκει στην ευθεία  $\mathbf{l}$  εάν και μόνο εάν  $\mathbf{x}^T \mathbf{l} = 0$

**Τομή ευθειών.** Για δύο ευθείες  $\mathbf{l} = (a, b, c)^T$  και  $\mathbf{l}' = (a', b', c')^T$  ορίζεται το διάνυσμα  $\mathbf{x} = \mathbf{l} \times \mathbf{l}'$ , όπου  $\times$  αντιπροσωπεύει το εξωτερικό γινόμενο. Από την ιδιότητα του εξωτερικού γινομένου  $\mathbf{l}^T(\mathbf{l} \times \mathbf{l}') = \mathbf{l}'^T(\mathbf{l} \times \mathbf{l}') = 0$  ισχύει ότι  $\mathbf{l}^T \mathbf{x} = \mathbf{l}'^T \mathbf{x} = 0$ . Επομένως, αν θεωρηθεί ότι  $\mathbf{x}$  αντιπροσωπεύει ένα σημείο, τότε το  $\mathbf{x}$  ανήκει και στις δύο ευθείες  $\mathbf{l}$  και  $\mathbf{l}'$ , και άρα αποτελεί το σημείο τομής των ευθειών.

**Αποτέλεσμα 2.2.** Το σημείο τομής δύο ευθειών  $\mathbf{l}, \mathbf{l}'$  είναι το σημείο  $\mathbf{x} = \mathbf{l} \times \mathbf{l}'$ .

**Ευθεία ορισμένη από σημεία.** Έστω δύο σημεία  $\mathbf{x}$  και  $\mathbf{x}'$ . Ορίζοντας το διάνυσμα  $\mathbf{l} = \mathbf{x} \times \mathbf{x}'$  και θεωρώντας ότι παριστάνει ευθεία, εύκολα προκύπτει ότι αυτή διέρχεται και από τα δύο σημεία.

**Αποτέλεσμα 2.3** Η ευθεία που διέρχεται από τα σημεία  $\mathbf{x}$  και  $\mathbf{x}'$  είναι η  $\mathbf{l} = \mathbf{x} \times \mathbf{x}'$

### 2.1.1.2 Ιδανικά Σημεία και η Ευθεία στο Άπειρο

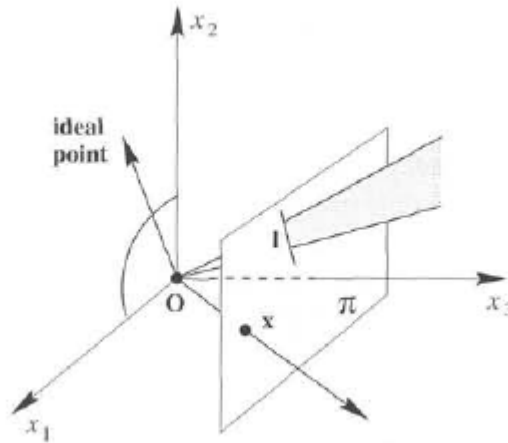
**Σημείο τομής παράλληλων ευθειών.** Έστω δύο παράλληλες ευθείες  $ax + by + c = 0$  και  $ax + by + c' = 0$ . Αυτές περιγράφονται από τα διανύσματα  $\mathbf{l} = (a, b, c)^T$  και  $\mathbf{l}' = (a, b, c')^T$ . Σύμφωνα με το Αποτέλεσμα 2.2 το σημείο τομής είναι  $\mathbf{l} \times \mathbf{l}' = (c' - c)(b, -a, 0)^T$ . Αγνοώντας την πολλαπλασιαστική σταθερά, το σημείο τομής των δύο ευθειών είναι το  $(b, -a, 0)^T$ . Το σημείο αυτό είναι σημείου απείρου και δεν αντιστοιχεί σε πεπερασμένο σημείο στο  $\mathbb{R}^2$ .

**Ευθεία στο άπειρο.** Από τα παραπάνω προκύπτει ότι ένα ομογενές διάνυσμα  $\mathbf{x} = (x_1, x_2, x_3)^T$  με  $x_3 \neq 0$  αντιστοιχεί σε πεπερασμένο σημείο στο  $\mathbb{R}^2$ . Τα σημεία με  $x_3 = 0$  θεωρούνται σημεία στο άπειρο. Το σύνολο όλων των ιδανικών σημείων (δηλαδή σημείων στο άπειρο) μπορεί να γραφεί ως  $(x_1, x_2, 0)^T$ , όπου ένα συγκεκριμένο σημείο καθορίζεται από το λόγο  $x_1 : x_2$ . Είναι φανερό ότι αυτό το σύνολο σημείων ανήκει σε μια ευθεία, την *ευθεία στο άπειρο*, που περιγράφεται από το διάνυσμα  $\mathbf{l}_\infty = (0, 0, 1)^T$ . Πράγματι, εύκολα επαληθεύει κανείς ότι  $(0, 0, 1)(x_1, x_2, 0)^T = 0$ .

Από το Αποτέλεσμα 2.2, το σημείο τομής της ευθείας  $\mathbf{l} = (a, b, c)^T$  και της ευθείας στο άπειρο είναι το ιδανικό σημείο  $(b, -a, 0)^T$ . Μια ευθεία  $\mathbf{l}' = (a, b, c')^T$  παράλληλη στην  $\mathbf{l}$ , έχει το ίδιο σημείο τομής με την ευθεία στο άπειρο  $\mathbf{l}_\infty$ . Σε μη ομογενείς συντεταγμένες, το διάνυσμα  $(b, -a)^T$  είναι ένα διάνυσμα εφαπτόμενο στην ευθεία και ορθογώνιο στο κάθετο διάνυσμα της ευθείας  $(a, b)$  και έτσι αντιπροσωπεύει τη διεύθυνση της ευθείας. Για το σκοπό αυτό, η ευθεία στο άπειρο μπορεί να θεωρηθεί ως το σύνολο των διαφορετικών διευθύνσεων των ευθειών στο 2Δ χώρο.

**Μοντελοποίηση του προβολικού επιπέδου.** Για ευκολότερη κατανόηση του  $\mathbb{P}^2$ , μπορεί κανείς να το σκεφτεί ως ένα σύνολο ακτίνων στο  $\mathbb{R}^3$ . Το σύνολο όλων των διανυσμάτων της μορφής  $k(x_1, x_2, x_3)^T$  με  $k$  μεταβαλλόμενο σχηματίζει μια ακτίνα που διέρχεται από την αρχή ( $k = 0$ ). Μια τέτοια ακτίνα αντιπροσωπεύει ένα σημείο στο  $\mathbb{P}^2$ . Με το μοντέλο αυτό, μια ευθεία στο  $\mathbb{P}^2$  είναι ένα επίπεδο που διέρχεται από την αρχή. Τα σημεία και οι ευθείες στο  $\mathbb{P}^2$  μπορούν να προκύψουν από την τομή των ακτίνων και των επιπέδων που διέρχονται από την αρχή με το επίπεδο  $x_3 = 1$ . Οι ακτίνες που αντιστοιχούν στα ιδανικά σημεία και το επίπεδο που αντιστοιχεί στην  $\mathbf{l}_\infty$  είναι παράλληλα στο επίπεδο  $x_3 = 1$ . Τα παραπάνω φαίνονται στο Σχήμα 2.1.

**Διαδικότητα.** Είναι εμφανές ότι ο ρόλος των ευθειών και των σημείων εναλλάσσεται στις σχέσεις που αφορούν τις ιδιότητές τους. Τα Αποτελέσματα 2.2 και 2.4 είναι τα ίδια, με τους ρόλους των σημείων και ευθειών να εναλλάσσονται.



**Σχήμα 2.1:** Το μοντέλο του προβολικού επιπέδου. Τα σημεία και οι ευθείες του  $\mathbb{P}^2$  αντιστοιχούν σε ακτίνες και επίπεδα, αντίστοιχα, που διέρχονται από την αρχή στο  $\mathbb{R}^3$ . Ευθείες που ανήκουν στο  $x_1x_2$ -επίπεδο αντιστοιχούν σε ιδανικά σημεία ενώ το  $x_1x_2$ -επίπεδο αντιπροσωπεύει την  $I_\infty$ . Το σχήμα προέρχεται από το [22].

### 2.1.1.3 Κωνικές Τομές

Μια κωνική τομή είναι μια καμπύλη που περιγράφεται από μια εξίσωση δευτέρου βαθμού στο επίπεδο. Στην Ευκλείδεια γεωμετρία διακρίνονται τρία είδη κωνικών τομών, οι υπερβολοειδείς, οι ελλειψοειδείς και οι παραβολικές.

Η εξίσωση μια κωνικής τομής σε μη ομογενείς συντεταγμένες είναι

$$ax_2 + bxy + cy^2 + dx + ey + f = 0 \quad (2.1)$$

Ομογενοποιώντας τις συντεταγμένες με αντικατάσταση της συνιστώσας  $x$  με  $\frac{x_1}{x_3}$  και της  $y$  με  $\frac{x_2}{x_3}$  προκύπτει η εξίσωση

$$ax_1^2 + bx_1x_2 + cx_2^2 + dx_1x_3 + ex_2x_3 + fx_3^2 = 0 \quad (2.2)$$

ή σε μορφή πινάκων

$$\mathbf{x}^T C \mathbf{x} = 0 \quad (2.3)$$

όπου ο πίνακας  $C$  δίνεται από

$$C = \begin{bmatrix} a & b/2 & d/2 \\ b/2 & c & e/2 \\ d/2 & e/2 & f \end{bmatrix}$$

Ο πίνακας  $C$  είναι συμμετρικός. Ο λόγος των στοιχείων του πίνακα είναι σημαντικός καθώς ο πολλαπλασιασμός του πίνακα με μια μη μηδενική σταθερά δεν επηρεάζει την εξίσωση. Επομένως, μια κωνική τομή έχει πέντε βαθμούς ελευθερίας (τα έξι στοιχεία ενός συμμετρικού πίνακα  $3 \times 3$  πλην μια πολλαπλασιαστική σταθερά).

Ο προσδιορισμός των παραμέτρων του πίνακα  $C$  και επομένως της κωνικής τομής απαιτεί τον προσδιορισμό των πέντε στοιχείων και άρα απαιτεί την γνώση πέντε σημείων που ανήκουν στην καμπύλη. Με άλλα λόγια, πέντε σημεία καθορίζουν μια κωνική τομή.

**Εφαπτόμενες στις κωνικές τομές.** Η ευθεία  $\mathbf{l}$  εφαπτόμενη σε μια κωνική τομή σε σημείο  $\mathbf{x}$  δίνεται από το παρακάτω αποτέλεσμα

**Αποτέλεσμα 2.4** Η ευθεία  $\mathbf{l}$  εφαπτόμενη στην  $C$  στο σημείο  $\mathbf{x}$  δίνεται από τη σχέση  $\mathbf{l} = C\mathbf{x}$ .

Πράγματι, η ευθεία αυτή διέρχεται από το σημείο  $\mathbf{x}$ , διότι  $\mathbf{l}^T \mathbf{x} = \mathbf{x}^T C \mathbf{x} = 0$ . Αν η ευθεία  $\mathbf{l}$  έχει ένα σημείο επαφής με την κωνική τομή, τότε είναι εφαπτόμενη. Διαφορετικά, αν η ευθεία έχει και δεύτερο σημείο τομής, έστω  $\mathbf{y}$ , τότε  $\mathbf{y}^T C \mathbf{y} = \mathbf{l}^T \mathbf{y} = 0$ . Τότε όμως ισχύει ότι  $(\mathbf{x} + \alpha \mathbf{y})^T C (\mathbf{x} + \alpha \mathbf{y}) = 0$  για κάθε  $\alpha$ . Επομένως, όλη η ευθεία  $\mathbf{l} = C\mathbf{x}$  που ενώνει τα σημεία  $\mathbf{x}$ ,  $\mathbf{y}$  ανήκει στην κωνική τομή  $C$ . Για μη εκφυλισμένες κωνικές τομές (βλ. παρακάτω), ένα τέτοιο συμπέρασμα δεν ισχύει.

**Δυαδικές κωνικές τομές.** Ο ορισμός της κωνικής τομής παραπάνω βασίζεται στον καθορισμό της καμπύλης από σημεία. Από την ιδιότητα της *δυαδικότητας* μπορεί να οριστεί μια κωνική τομή που προκύπτει από εξισώσεις σε ευθείες. Αυτή η κωνική τομή περιγράφεται επίσης από ένα  $3 \times 3$  πίνακα  $C^*$ . Μια ευθεία  $\mathbf{l}$  εφαπτόμενη στην κωνική τομή  $C$  ικανοποιεί τη σχέση  $\mathbf{l}^T C^* \mathbf{l} = 0$ . Ο συμβολισμός  $C^*$  δείχνει ότι ο πίνακας  $C^*$  είναι ο συζυγής πίνακας του  $C$ . Αν ο πίνακας συμμετρικός και αντιστρέψιμος, τότε ισχύει  $C^* = C^{-1}$ .

Στην περίπτωση που ο πίνακας  $C$  είναι πλήρης, από το Αποτέλεσμα 2.4 η εφαπτόμενη από ένα σημείο  $\mathbf{x}$  που ανήκει στην κωνική τομή  $C$  ικανοποιεί τη σχέση  $\mathbf{l} = C\mathbf{x}$ . Αντιστρέφοντας προκύπτει ότι  $\mathbf{x} = C^{-1}\mathbf{l}$ . Το σημείο  $\mathbf{x}$  ανήκει στην κωνική τομή οπότε ισχύει ότι  $\mathbf{x}^T C \mathbf{x} = 0$ . Επομένως, προκύπτει ότι  $(C^{-1}\mathbf{l})^T C (C^{-1}\mathbf{l}) = \mathbf{l}^T C^{-1} \mathbf{l} = 0$

**Εκφυλισμένες κωνικές τομές.** Αν ο πίνακας  $C$  δεν είναι πλήρους βαθμού, τότε η κωνική τομή λέγεται εκφυλισμένη.

## 2.1.2 Προβολικοί Μετασχηματισμοί

**Ορισμός 2.1.** Προβολικότητα είναι ένας αντιστρέψιμος μετασχηματισμός  $h$  από το  $\mathbb{P}^2$  στον εαυτό του τέτοιος ώστε τρία σημεία  $\mathbf{x}_1$ ,  $\mathbf{x}_2$  και  $\mathbf{x}_3$  ανήκουν στην ίδια ευθεία αν και μόνο αν  $h(\mathbf{x}_1)$ ,  $h(\mathbf{x}_2)$  και  $h(\mathbf{x}_3)$  είναι συνευθειακά.




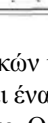
Η προβολικότητα ονομάζεται αλλιώς και προβολικός μετασχηματισμός ή ομογραφία.

**Θεώρημα 2.1.** Μια αντιστοιχηση  $h : \mathbb{P}^2 \rightarrow \mathbb{P}^2$  είναι μια προβολικότητα αν και μόνο αν υπάρχει ένας αντιστρέψιμος  $3 \times 3$  πίνακας  $H$  τέτοιος ώστε για κάθε σημείο στο  $\mathbb{P}^2$  που περιγράφεται από ένα διάνυσμα  $\mathbf{x}$  να ισχύει ότι  $h(\mathbf{x}) = H\mathbf{x}$ .

Πράγματι, έστω τρία συνευθειακά σημεία  $\mathbf{x}_1$ ,  $\mathbf{x}_2$  και  $\mathbf{x}_3$  ανήκουν στην ευθεία  $\mathbf{l}$ . Τότε  $\mathbf{l}^T \mathbf{x}_i = 0$  για  $i = 1, \dots, 3$ . Έστω ότι ο  $H$  είναι ένας μη αντιστρέψιμος  $3 \times 3$  πίνακας. Ισχύει ότι  $\mathbf{l}^T H^{-1} H \mathbf{x}_i = 0$ . Επομένως τα σημεία  $H \mathbf{x}_i$  ανήκουν στην ευθεία  $H^{-T} \mathbf{l}$  και άρα είναι συνευθειακά. Τελικά, διατηρείται η ιδιότητα της συγγραμικότητας και άρα πρόκειται για *προβολικότητα*.

Από τα παραπάνω, είναι δυνατό να δοθεί ένας εναλλακτικός ορισμός του προβολικού μετασχηματισμού.

**Ορισμός 2.2.** Ένας επίπεδος προβολικός μετασχηματισμός είναι ένας γραμμικός μετασχηματισμός σε ομογενείς διανύσματα που περιγράφεται από ένα αντιστρέψιμο  $3 \times 3$  πίνακα  $H$ , δηλαδή  $\mathbf{x}' = H\mathbf{x}$ ,

Group	Matrix	Distortion	Invariant properties
Projective 8 dof	$\begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix}$		Concurrency, collinearity, <b>order of contact</b> : intersection (1 pt contact); tangency (2 pt contact); inflections (3 pt contact with line); tangent discontinuities and cusps. cross ratio (ratio of ratio of lengths).
Affine 6 dof	$\begin{bmatrix} a_{11} & a_{12} & t_x \\ a_{21} & a_{22} & t_y \\ 0 & 0 & 1 \end{bmatrix}$		Parallelism, ratio of areas, ratio of lengths on collinear or parallel lines (e.g. midpoints), linear combinations of vectors (e.g. centroids). The line at infinity, $l_\infty$ .
Similarity 4 dof	$\begin{bmatrix} sr_{11} & sr_{12} & t_x \\ sr_{21} & sr_{22} & t_y \\ 0 & 0 & 1 \end{bmatrix}$		Ratio of lengths, angle. The circular points, <b>I, J</b>
Euclidean 3 dof	$\begin{bmatrix} r_{11} & r_{12} & t_x \\ r_{21} & r_{22} & t_y \\ 0 & 0 & 1 \end{bmatrix}$		Length, area

**Πίνακας 2.1:** Γεωμετρικές ιδιότητες των ειδικών κατηγοριών 2Δ μετασχηματισμών. Ο πίνακας  $A = [a_{ij}]$  είναι αντιστρέψιμος, ο πίνακας  $R = [r_{ij}]$  είναι ένας πίνακας περιστροφής. Η τρίτη στήλη δείχνει το αποτέλεσμα των μετασχηματισμών σε ένα τετράγωνο. Οι μετασχηματισμοί που βρίσκονται πιο ψηλά στον πίνακα περιλαμβάνουν τους πιο χαμηλούς. Ο πίνακας προέρχεται από το [22].

όπου ο πίνακας  $H$  δίνεται από

$$H = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix}$$

**Μετασχηματισμός ευθειών.** Παραπάνω έγινε φανερό πως όταν ένα σημείο  $\mathbf{x}$  ανήκει στην ευθεία  $\mathbf{l}$  τότε το νέο σημείο  $\mathbf{x}' = H\mathbf{x}$  ανήκει στην ευθεία  $\mathbf{l}' = H^{-T}\mathbf{l}$ . Συμπερασματικά, υπό το μετασχηματισμό  $\mathbf{x}' = H\mathbf{x}$ , μια ευθεία μετασχηματίζεται σε

$$\mathbf{l}' = H^{-T}\mathbf{l} \quad (2.4)$$

### 2.1.3 Ιεραρχία Μετασχηματισμών

Στην ενότητα αυτή περιγράφονται οι ειδικές περιπτώσεις του προβολικού μετασχηματισμού. Στον παρακάτω πίνακα φαίνονται οι περιπτώσεις αυτές.

Για κάθε ένα από τους μετασχηματισμούς στον Πίνακα 2.1 δίνεται παρακάτω μια σύντομη περιγραφή των χαρακτηριστικών τους.

- **Προβολικός Μετασχηματισμός** Όπως παρατηρήθηκε και παραπάνω, ο μετασχηματισμός αυτός αφήνει αναλοίωτη την ιδιότητα της συγγραμμικότητας. Διαθέτει οκτώ βαθμούς ελευθερίας, καθώς είναι ομογενής μετασχηματισμός (πίνακας εννέα στοιχείων πλην μιας πολλαπλασιαστικής σταθεράς).
- **Αφινικός Μετασχηματισμός** Ο αφινικός μετασχηματισμός διατηρεί τις παράλληλες ευθείες, τον λόγο εμβαδών, το λόγο μηκών σε παράλληλες ευθείες και την ευθεία στο άπειρο. Διαθέτει έξι βαθμούς ελευθερίας (τέσσερα στοιχεία του πίνακα  $A$  και δύο στοιχεία του διανύσματος  $t$ ).
- **Μετασχηματισμός Ομοιότητας** Ο μετασχηματισμός ομοιότητας διατηρεί το λόγο μηκών, τις γωνίες καθώς και τα κυκλικά σημεία **I, J**. Διαθέτει τέσσερις βαθμούς ελευθερίας (ένα βαθμό για την περιστροφή, ένα για την κλίμακα και δύο για την μετακίνηση).

- **Ευκλείδειος Μετασχηματισμός** Ο ευκλείδειος μετασχηματισμός διατηρεί τα μήκη και τα εμβαδά. Διαθέτει τρεις βαθμούς ελευθερίας (ένα για την περιστροφή και δύο για την μετακίνηση).

## 2.2 Προβολική Γεωμετρία και Μετασχηματισμοί στο 3Δ Χώρο

Στην ενότητα αυτή μελετώνται οι ιδιότητες της προβολικής γεωμετρίας στον 3Δ κόσμο. Οι έννοιες που εισάγονται είναι επεκτάσεις των εννοιών που μελετήθηκαν στην προηγούμενη ενότητα. Στην ενότητα αυτή, με κεφαλαία έντονα γράμματα θα συμβολίζουμε ένα διάνυσμα διάστασης  $4 \times 1$ .

### 2.2.1 Σημεία και Προβολικοί Μετασχηματισμοί

Ένα σημείο  $\mathbf{X}$  στο τρισδιάστατο χώρο περιγράφεται σε ομογενείς συντεταγμένες ως ένα διάνυσμα  $4 \times 1$ . Συγκεκριμένα, το ομογενές διάνυσμα  $\mathbf{X} = (X_1, X_2, X_3, X_4)^T$  με  $X_4 \neq 0$  αντιστοιχεί στο σημείο  $(X, Y, Z)$  του  $\mathbb{R}^3$  όπου

$$X = \frac{X_1}{X_4} \quad Y = \frac{X_2}{X_4} \quad Z = \frac{X_3}{X_4} \quad (2.5)$$

Ομογενή σημεία με  $X_4 = 0$  αντιστοιχούν σε σημεία στο άπειρο.

Ένας προβολικός μετασχηματισμός στον  $\mathbb{P}^3$  είναι ένας γραμμικός μετασχηματισμός που δρα σε ομογενή διανύσματα  $4 \times 1$ . Πρόκειται για έναν αντιστρέψιμο πίνακα  $H$  διάστασης  $4 \times 4$ :  $\mathbf{X}' = H\mathbf{X}$ . Ο πίνακας  $H$  που αντιπροσωπεύει τον μετασχηματισμό σε ομογενείς συντεταγμένες έχει 15 βαθμούς ελευθερίας (16 στοιχεία του πίνακα πλην μιας πολλαπλασιαστικής σταθεράς). Ο προβολικός μετασχηματισμός διατηρεί τις ευθείες, δηλαδή ευθείες μετασχηματίζονται σε ευθείες, καθώς και τα σημεία τομής ευθειών με επίπεδα.

### 2.2.2 Επίπεδα, Ευθείες και Προβολικοί Μετασχηματισμοί

Στον  $\mathbb{P}^3$  τα σημεία και τα επίπεδα είναι δυαδικά και η περιγραφή τους είναι ανάλογη με την δυαδικότητα σημείου-ευθείας στον  $\mathbb{P}^2$ .

#### 2.2.2.1 Επίπεδα

Στον τρισδιάστατο ευκλείδειο χώρο ένα επίπεδο περιγράφεται από την εξίσωση

$$\pi_1 X + \pi_2 Y + \pi_3 Z + \pi_4 = 0$$

Ομογενοποιώντας τις συντεταγμένες ενός σημείου που ανήκει στο επίπεδο, δηλαδή  $X \rightarrow \frac{X_1}{X_4}$ ,  $Y \rightarrow \frac{X_2}{X_4}$ ,  $Z \rightarrow \frac{X_3}{X_4}$  προκύπτει η εξίσωση

$$\pi_1 X_1 + \pi_2 X_2 + \pi_3 X_3 + \pi_4 X_4 = 0 \quad (2.6)$$

ή σε διανυσματική μορφή όπου  $\pi = (\pi_1, \pi_2, \pi_3, \pi_4)^T$ .

$$\pi^T \mathbf{X} = 0 \quad (2.7)$$

που εκφράζει ότι το σημείο  $\mathbf{X}$  ανήκει στο επίπεδο  $\pi$ .

Οι πρώτες τρεις συνιστώσες του  $\pi$  αντιστοιχούν στο κάθετο διάνυσμα του επιπέδου στην ευκλείδεια γεωμετρία.

**Τρία σημεία ορίζουν ένα επίπεδο.** Έστω τρία σημεία  $\mathbf{X}_i$  ανήκουν στο επίπεδο  $\pi$ . Τότε για κάθε σημείο ισχύει  $\pi^T \mathbf{X}_i = 0$ ,  $i = 1, \dots, 3$ . Σε μορφή πινάκων οι εξισώσεις γίνονται

$$\begin{bmatrix} \mathbf{X}_1^T \\ \mathbf{X}_2^T \\ \mathbf{X}_3^T \end{bmatrix} \pi = 0 \quad (2.8)$$

Αν τα τρία σημεία είναι μη συγγραμικά, ο πίνακας που σχηματίζεται από τα σημεία θα έχει βαθμό 3. Επομένως, το επίπεδο προκύπτει με ακρίβεια μιας πολλαπλασιαστικής σταθεράς από το δεξί null-space. Αν ο πίνακας έχει βαθμό 2, τότε και το null-space έχει διάσταση 2 και άρα τα σημεία συγγραμικά. Η λύση της εξίσωσης δίνει ένα σύνολο από επίπεδα που διέρχονται από τα σημεία αυτά.

**Το επίπεδο στο άπειρο.** Το επίπεδο στο άπειρο  $\pi_\infty = (0, 0, 0, 1)^T$  περιλαμβάνει όλες τις διευθύνσεις  $\mathbf{D} = (X_1, X_2, X_3, 0)^T$ . Παρακάτω, περιγράφονται δύο ιδιότητες του επιπέδου αυτού

- Δύο επίπεδα είναι παράλληλα αν και μόνο αν η ευθεία τομής τους ανήκει στο  $\pi_\infty$
- Μια ευθεία είναι παράλληλη σε μία άλλη ευθεία, ή σε ένα επίπεδο, αν το σημείο τομής ανήκει στο  $\pi_\infty$

Το επίπεδο  $\pi_\infty$  είναι μια γεωμετρική αναπαράσταση για τους 3 βαθμούς ελευθερίας που χρειάζονται για να καθοριστούν οι αφινικές ιδιότητες. Το επίπεδο στο άπειρο είναι αναλοίωτο από αφινικούς μετασχηματισμούς.

### 2.2.2.2 Ευθείες

Μια ευθεία ορίζεται από την ένωση δύο σημείων ή την τομή δύο επιπέδων. Ο τρόπος περιγραφής τους είναι μέσω αυτών των δύο δυνατοτήτων ορισμού μιας ευθείας. Συγκεκριμένα, μπορούν να παρασταθούν μέσω *null-space and span representation*, *pluecker matrices* και *pluecker line coordinates*. Οι τρόποι αυτοί αναλύονται στον [22].

### 2.2.3 Ιεραρχία Μετασχηματισμών

Υπάρχει ένα πλήθος περιπτώσεων του προβολικού μετασχηματισμού στον 3Δ χώρο. Κάθε κατηγορία μετασχηματισμού έχει τις δικές της ιδιότητες. Στον παρακάτω πίνακα φαίνεται η ιεραρχία των μετασχηματισμών, όπου σε κάθε στήλη αναγράφονται μόνο οι επιπρόσθετες ιδιότητες καθώς ανέρχεται κανείς τις στήλες.





## 2.3 Μοντέλα Καμερών

Η κάμερα είναι μια απεικόνιση μεταξύ του τρισδιάστατου κόσμου (χώρος αντικειμένων) και της δισδιάστατης εικόνας. Στην ενότητα αυτή μελετώνται οι κάμερες που μοντελοποιούνται μέσω πινάκων με διάφορες ιδιότητες. Ενδιαφέρον θα δοθεί κυρίως στις κάμερες *κεντρικής προβολής*. Όλες οι κάμερες κεντρικής προβολής είναι ειδικές περιπτώσεις της *γενικής προβολικής κάμερας*. Για την ανάλυση της γενικότερης αυτής κάμερας χρησιμοποιούνται εργαλεία της προβολικής γεωμετρίας.

Οι ειδικές περιπτώσεις της *γενικής προβολικής κάμερας* κληρονομούν τις ιδιότητες της. Τα ειδικά αυτά μοντέλα κατηγοριοποιούνται σε δύο μεγάλες κλάσεις, αυτές που μοντελοποιούν την κάμερα με πεπερασμένο κέντρο και αυτές που μοντελοποιούν την κάμερα με κέντρο στο άπειρο.

Στην ενότητα αυτή, με μικρά έντονα γράμματα θα περιγράφονται σημεία του  $\mathbb{P}^2$  ενώ με κεφαλαία έντονα γράμματα σημεία του  $\mathbb{R}^3$ .



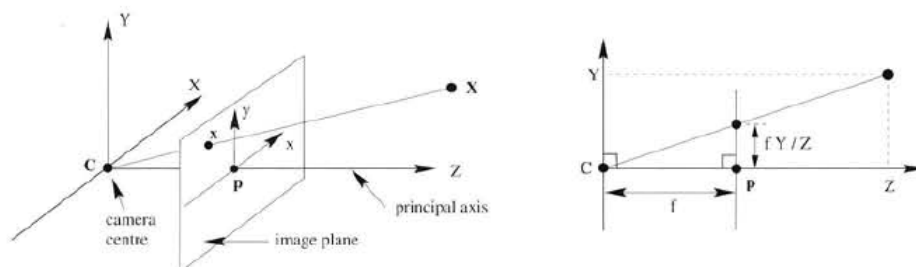
Group	Matrix	Distortion	Invariant properties
Projective 15 dof	$\begin{bmatrix} A & t \\ v^T & v \end{bmatrix}$		Intersection and tangency of surfaces in contact. Sign of Gaussian curvature.
Affine 12 dof	$\begin{bmatrix} A & t \\ 0^T & 1 \end{bmatrix}$		Parallelism of planes, volume ratios, centroids. The plane at infinity, $\pi_\infty$ , (see section 3.5).
Similarity 7 dof	$\begin{bmatrix} sR & t \\ 0^T & 1 \end{bmatrix}$		The absolute conic, $\Omega_\infty$ , (see section 3.6).
Euclidean 6 dof	$\begin{bmatrix} R & t \\ 0^T & 1 \end{bmatrix}$		Volume.

**Πίνακας 2.2:** Γεωμετρικές ιδιότητες των ειδικών κατηγοριών 3Δ μετασχηματισμών. Ο πίνακας  $A = [a_{ij}]$  είναι αντιστρέψιμος, ο πίνακας  $R = [r_{ij}]$  είναι ένας πίνακας περιστροφής. Η τρίτη στήλη δείχνει το αποτέλεσμα των μετασχηματισμών σε ένα κύβο. Οι μετασχηματισμοί που βρίσκονται πιο ψηλά στον πίνακα περιλαμβάνουν τους πιο χαμηλούς. Ο πίνακας προέρχεται από το [22].

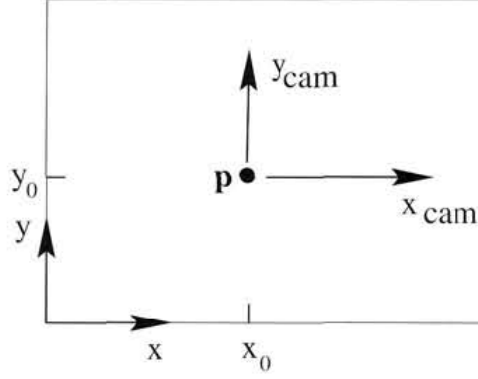
### 2.3.1 Πεπερασμένη Κάμερα

Στην υποενότητα αυτή μελετώνται τα πιο ειδικά και απλά μοντέλα όπως η pinhole κάμερα και περιγράφονται τα αναγκαία μαθηματικά μοντέλα.

**Pinhole Κάμερα.** Στο μοντέλο αυτό, τα σημεία του κόσμου προβάλλονται στο επίπεδο της εικόνας. Έστω ότι το κέντρο προβολής είναι η αρχή ενός ευκλείδειου συστήματος συντεταγμένων και έστω ότι το επίπεδο  $Z = f$  είναι το επίπεδο της κάμερας. Σύμφωνα με το μοντέλο της pinhole κάμερας, ένα σημείο  $\mathbf{X} = (X, Y, Z)^T$  στον 3Δ χώρο απεικονίζεται σε ένα σημείο στο επίπεδο της εικόνας. Το σημείο αυτό είναι το σημείο τομής της ευθείας που ενώνει το κέντρο προβολής με το σημείο του κόσμου  $\mathbf{X}$  με το επίπεδο της εικόνας. Τα παραπάνω φαίνονται στο Σχήμα 2.2.



**Σχήμα 2.2:** Το μοντέλο της pinhole κάμερας.  $C$  είναι το κέντρο της κάμερας και  $p$  είναι το *κόριο σημείο*. Να σημειωθεί ότι το επίπεδο της κάμερας έχει τοποθετηθεί μπροστά από το κέντρο της κάμερας. Το σχήμα προέρχεται από το [22].



**Σχήμα 2.3:** Τα συστήματα συντεταγμένων της εικόνας  $(x, y)$  και της κάμερας  $(x_{cam}, y_{cam})$ . Το σχήμα προέρχεται από το [22].

Εύκολα αποδεικνύεται ότι το σημείο  $(X, Y, Z)^T$  απεικονίζεται στο σημείο  $(f\frac{X}{Z}, f\frac{Y}{Z}, f)^T$  στο επίπεδο της εικόνας. Με άλλα λόγια, η αντιστοιχία δίνεται από την παρακάτω σχέση

$$(X, Y, Z)^T \rightarrow (f\frac{X}{Z}, f\frac{Y}{Z})^T \quad (2.9)$$

όπου αντιστοιχίζεται σημείο του κόσμου (χώρος  $\mathbb{R}^3$ ) σε σημείο της εικόνας (χώρος  $\mathbb{R}^2$ ).

Η παραπάνω σχέση μπορεί να παρασταθεί με τη μορφή πινάκων. Θεωρώντας τις ομογενείς συντεταγμένες των σημείων προκύπτει ότι

$$\begin{bmatrix} fX \\ fY \\ Z \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (2.10)$$

Αν  $\mathbf{x} = (fX, fY, Z)^T$  και  $\mathbf{X} = (X, Y, Z, 1)^T$  η παραπάνω σχέση γράφεται

$$\mathbf{x} = P\mathbf{X} \quad (2.11)$$

όπου  $P = \text{diag}(f, f, 1)[I|\mathbf{0}]$  ο πίνακας για το μοντέλο της pinhole κάμερας.

**Αντιστάθμιση του κύριου σημείου.** Γενικεύοντας το μοντέλο της pinhole κάμερας, είναι δυνατό να οριστεί ένα νέο μοντέλο αντισταθμίζοντας τη θέση του *κύριου σημείου*. Το *κύριο σημείο*  $\mathbf{p}$  ορίζεται ως το σημείο τομής του άξονα  $Z$  του τοπικού συστήματος συντεταγμένων της κάμερας με το επίπεδο της εικόνας. Στο μοντέλο της pinhole κάμερας, όπως φαίνεται και από το Σχήμα 2.2 το κύριο σημείο είναι και η αρχή του συστήματος συντεταγμένων της εικόνας για τον  $\mathbb{P}^2$ . Είναι δυνατό ωστόσο να γίνει μετατόπιση της αρχής του συστήματος του επιπέδου της εικόνας με αποτέλεσμα να πρέπει να αντισταθμιστεί η απόσταση αυτή στον πίνακα της κάμερας. Στο Σχήμα 2.3 φαίνεται το σύστημα συντεταγμένων της εικόνας καθώς και το κύριο σημείο  $\mathbf{p}$ .

Τα σημεία του 3Δ κόσμου αντιστοιχίζονται στα σημεία της εικόνας σύμφωνα με τον κανόνα

$$(X, Y, Z)^T \rightarrow (f\frac{X}{Z} + p_x, f\frac{Y}{Z} + p_y)^T \quad (2.12)$$

όπου  $(p_x, p_y)^T$  είναι οι συντεταγμένες του κύριου σημείου. Σε μορφή πινάκων η σχέση γίνεται

$$\mathbf{x} = K[I|\mathbf{0}]\mathbf{X}_{cam} \quad (2.13)$$

όπου ο πίνακας  $K$  ονομάζεται *πίνακας βαθμονόμησης* δίνεται από τη σχέση

$$K = \begin{bmatrix} f & 0 & p_x \\ 0 & f & p_y \\ 0 & 0 & 1 \end{bmatrix} \quad (2.14)$$

Να σημειωθεί ότι τα σημεία του 3Δ κόσμου με ομογενείς συντεταγμένες  $\mathbf{X}_{cam} = (X, Y, Z, 1)^T$  αναφέρονται ως προς σύστημα συντεταγμένων προσαρτημένο στην κάμερα (η αρχή του συστήματος ταυτίζεται με το κέντρο της κάμερας) και με τον κύριο άξονα της κάμερας να έχει την διεύθυνση του άξονα  $Z$ . Το σύστημα αυτό καλείται *σύστημα συντεταγμένων της κάμερας*.

**Περιστροφή και μετακίνηση της κάμερας.** Στην προηγούμενη παράγραφο θεωρήθηκε ότι το παγκόσμιο σύστημα συντεταγμένων ταυτίζεται με το σύστημα συντεταγμένων της κάμερας. Στη γενική περίπτωση, ωστόσο, τα δύο συστήματα διαφέρουν κατά μια περιστροφή και μία μετατόπιση. Αν  $\tilde{\mathbf{X}}$  είναι ένα μη ομογενές διάνυσμα διάστασης  $3 \times 1$  που περιγράφει ένα σημείο στον 3Δ κόσμο, ενώ  $\tilde{\mathbf{X}}_{cam}$  περιγράφει το ίδιο σημείο στο σύστημα συντεταγμένων της κάμερας, τότε μπορεί να γραφεί ότι  $\tilde{\mathbf{X}}_{cam} = R(\tilde{\mathbf{X}} - \tilde{\mathbf{C}})$ , όπου  $\tilde{\mathbf{C}}$  περιγράφει τις συντεταγμένες του κέντρου της κάμερας στο παγκόσμιο σύστημα συντεταγμένων και  $R$  είναι ένας  $3 \times 3$  πίνακας περιστροφής που περιγράφει τον προσανατολισμό του συστήματος συντεταγμένων της κάμερας.

Από τη σχέση 2.13 προκύπτει ότι

$$\mathbf{x} = KR[I | -\tilde{\mathbf{C}}]\mathbf{X} \quad (2.15)$$

όπου  $\mathbf{X}$  είναι σημεία στο παγκόσμιο σύστημα συντεταγμένων.

Είναι φανερό από την τελευταία σχέση ότι μια γενική pinhole κάμερα,  $P = KR[I | -\tilde{\mathbf{C}}]$  έχει 9 βαθμούς ελευθερίας, δηλαδή 3 για τον  $K$ , 3 για τον  $R$  και 3 για τον  $\tilde{\mathbf{C}}$ . Οι παράμετροι του πίνακα βαθμονόμησης  $K$  λέγονται *εσωτερικές παράμετροι* της κάμερας, ενώ οι παράμετροι των  $R$  και  $\tilde{\mathbf{C}}$  λέγονται *εξωτερικές παράμετροι*.

Ο πίνακας της κάμερας πολλές φορές γράφεται στην παρακάτω μορφή

$$P = K[R|\mathbf{t}] \quad (2.16)$$

όπου  $\mathbf{t} = -R\tilde{\mathbf{C}}$ .

**Πεπερασμένη προβολική κάμερα.** Στο προηγούμενο μοντέλο θεωρήθηκε ότι οι συντεταγμένες της εικόνας είναι ευκλείδειες με ίση κλίμακα και στις δύο διευθύνσεις των αξόνων. Γενικεύοντας το μοντέλο, υπάρχει η πιθανότητα τα pixels να μην είναι τετράγωνα. Αν, λοιπόν, οι συντεταγμένες μετρώνται σε pixels τότε είναι αναγκαίο να εισαχθούν άνισοι παράγοντες κλίμακας σε κάθε κατεύθυνση. Συγκεκριμένα, αν ο αριθμός των pixels ανά μονάδα απόστασης σε συντεταγμένες εικόνας είναι  $m_x$  και  $m_y$  στον  $x$ - και  $y$ -άξονα, τότε η μετατροπή από παγκόσμιες συντεταγμένες σε συντεταγμένες pixel προκύπτει πολλαπλασιάζοντας τον πίνακα βαθμονόμησης (σχέση) με τον πίνακα  $diag(m_x, m_y, 1)$ . Επομένως, ο πίνακας βαθμονόμησης γίνεται

$$K = \begin{bmatrix} \alpha_x & 0 & x_0 \\ 0 & \alpha_y & y_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (2.17)$$

όπου  $\alpha_x = fm_x$  και  $\alpha_y = fm_y$  αντιπροσωπεύουν το εστιακό μήκος της κάμερας σε διαστάσεις pixel στην  $x$ - και  $y$ -διεύθυνση. Το διάνυσμα  $\tilde{\mathbf{x}}_0 = (x_0, y_0)$  περιγράφει το κύριο σημείο σε pixel συντεταγμένες, δηλαδή  $x_0 = m_x p_x$  και  $y_0 = m_y p_y$ .

Γενικεύοντας ακόμα περισσότερο, ο πίνακας βαθμονόμησης μπορεί να έχει τη μορφή

$$K = \begin{bmatrix} \alpha_x & s & x_0 \\ 0 & \alpha_y & y_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (2.18)$$

Η επιπλέον παράμετρος  $s$  αναφέρεται ως *παράμετρος λοξότητας*. Για τις περισσότερες κάμερες λαμβάνει μηδενική τιμή.

Η κάμερα με

$$P = KR[I | -\tilde{\mathbf{C}}] \quad (2.19)$$

όπου ο πίνακας βαθμονόμησης δίνεται από τη σχέση 2.18 ονομάζεται *πεπερασμένη προβολική κάμερα*. Μια τέτοια κάμερα έχει 11 βαθμούς ελευθερίας.

## 2.3.2 Ιδιότητες Προβολικής Κάμερας

### 2.3.2.1 Ανατομία Κάμερας

Μια γενικευμένη προβολική κάμερα  $P = KR[I | -\tilde{\mathbf{C}}]$  μπορεί να αναλυθεί σε μπλοκ σύμφωνα με  $P = [M | \mathbf{p}_4]$  όπου ο πίνακας  $M$  είναι  $3 \times 3$ .

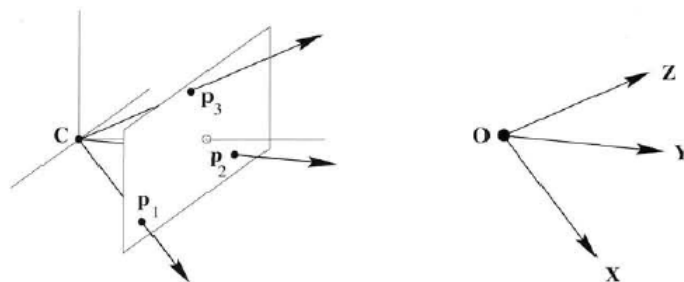
**Κέντρο κάμερας.** Ο πίνακας μια προβολικής κάμερας  $P$ , διάστασης  $3 \times 4$ , έχει βαθμό 3. Για το λόγο αυτό ο δεξιός μηδενικός υπόχωρός του έχει διάσταση 1. Έστω ότι ο μηδενικός υπόχωρος προκύπτει από το διάνυσμα  $\mathbf{C}$  έτσι ώστε  $PC = 0$ . Είναι εύκολο να δειχθεί (βλ. [22] p.158) ότι το διάνυσμα αυτό περιγράφει το κέντρο της κάμερας σε ομογενείς συντεταγμένες.

Στην περίπτωση των πεπερασμένων καμερών το αποτέλεσμα μπορεί να επαληθευτεί άμεσα, διότι  $\mathbf{C} = (\tilde{\mathbf{C}}^T, 1)^T$  για το οποίο ισχύει  $PC = 0$ . Το ίδιο ισχύει και για την περίπτωση που ο πίνακας  $M$  είναι μη αντιστρέψιμος. Στην περίπτωση αυτή, ένα διάνυσμα που γεννά τον μηδενικό υπόχωρο είναι το  $\mathbf{C} = (\mathbf{d}^T, 0)^T$  με  $M\mathbf{d} = 0$ . Το κέντρο της κάμερας είναι ένα σημείο στο άπειρο (οι κάμερες αυτές ανήκουν στην κατηγορία των *άπειρων καμερών*).

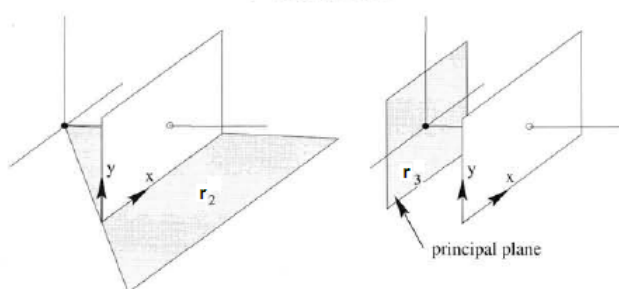
**Διανύσματα στήλες.** Οι στήλες της προβολικής κάμερας είναι τρισδιάστατα διανύσματα με γεωμετρική ερμηνεία. Συμβολίζοντας με  $\mathbf{p}_i, i = 1, \dots, 4$  τις στήλες του πίνακα  $P$  εύκολα προκύπτει ότι τα  $\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3$  είναι η απεικόνιση των τριών διευθύνσεων του παγκόσμιου συστήματος συντεταγμένων. Για παράδειγμα, ο  $X$ -άξονας έχει διεύθυνση  $\mathbf{D} = (1, 0, 0, 0)^T$ , που απεικονίζεται στο σημείο  $\mathbf{p}_1 = P\mathbf{D}$ . Τα παραπάνω φαίνονται στο Σχήμα 2.4 Το διάνυσμα  $\mathbf{p}_4$  είναι η απεικόνιση της αρχής των αξόνων.

**Διανύσματα γραμμές.** Τα διανύσματα γραμμές του πίνακα  $P$  είναι διάστασης  $4 \times 1$  και μπορούν να ερμηνευθούν ως επίπεδα τα οποία θα μελετηθούν παρακάτω. Οι γραμμές του πίνακα συμβολίζονται με  $\mathbf{r}_i^T, i = 1, \dots, 3$ .

- **Κύριο επίπεδο** Το κύριο επίπεδο είναι το επίπεδο που διέρχεται από το κέντρο της κάμερας και είναι παράλληλο στο επίπεδο της εικόνας. Αποτελείται από το σύνολο σημείων που προβάλλονται στην ευθεία στο άπειρο της εικόνας. Συγκεκριμένα,  $P\mathbf{X} = (x, y, 0)^T$ . Δηλαδή, ένα σημείο ανήκει στο κύριο επίπεδο της κάμερας αν και μόνο αν  $\mathbf{r}_3^T \mathbf{X} = 0$ . Με άλλα λόγια,  $\mathbf{r}_3$  είναι το διάνυσμα που περιγράφει το κύριο επίπεδο της κάμερας.
- **Επίπεδα αξόνων** Έστω το σύνολο σημείων  $\mathbf{X}$  του επιπέδου  $\mathbf{r}_1$ . Για τα σημεία αυτά ισχύει  $\mathbf{r}_1^T \mathbf{X} = 0$ . Επομένως, τα σημεία αυτά απεικονίζονται στα  $P\mathbf{X} = (0, y, z)^T$ , που είναι σημεία στον  $y$ -άξονα της εικόνας. Επιπλέον, αφού ισχύει  $PC = 0$  όπου  $\mathbf{C}$  το κέντρο της κάμερας, προκύπτει ότι και  $\mathbf{r}_1^T \mathbf{C} = 0$ , δηλαδή το κέντρο της κάμερας ανήκει στο επίπεδο



**Σχήμα 2.4:** Τα σημεία της εικόνας όπως ορίζονται από τις στήλες του πίνακα της κάμερας  $\mathbf{p}_i, i = 1, \dots, 3$ . Τα σημεία αυτά αποτελούν τις προβολές των διευθύνσεων των αξόνων του παγκόσμιου συστήματος συντεταγμένων. Το σχήμα προέρχεται από το [22].



**Σχήμα 2.5:** Δύο εκ των τριών επιπέδων που ορίζονται από τις γραμμές του πίνακα  $P$  της κάμερας. Το σχήμα προέρχεται από το [22].

$\mathbf{r}_1$ . Τελικά, το επίπεδο  $\mathbf{r}_1$  ορίζεται από το κέντρο της κάμερας και την ευθεία  $x = 0$  στην εικόνα. Όμοια, το επίπεδο  $\mathbf{r}_2$  ορίζεται από το κέντρο της κάμερας και την ευθεία  $y = 0$  της εικόνας.

Τα παραπάνω φαίνονται στο Σχήμα 2.5.

**Το κύριο σημείο.** Ο *κύριος άξονας* είναι η ευθεία που διέρχεται από το κέντρο της κάμερας  $C$  με διεύθυνση κάθετη στο κύριο επίπεδο  $\mathbf{r}_3$ . Ο άξονας τέμνει το επίπεδο της εικόνας στο *κύριο σημείο*. Ο προσδιορισμός του σημείου αυτού γίνεται ως εξής: Το κάθετο διάνυσμα ενός επιπέδου  $\pi = (\pi_1, \pi_2, \pi_3, \pi_4)^T$  δίνεται από το διάνυσμα  $(\pi_1, \pi_2, \pi_3)^T$ . Το διάνυσμα αυτό μπορεί να περιγραφεί και από το σημείο  $(\pi_1, \pi_2, \pi_3, 0)^T$  του επιπέδου στο άπειρο. Στην περίπτωση του κύριου επιπέδου της κάμερας το σημείο αυτό είναι το  $\mathbf{X}_3 = (p_{31}, p_{32}, p_{33}, 0)^T$ . Προβάλλοντας το σημείο στην εικόνα προκύπτει το κύριο σημείο, δηλαδή  $\mathbf{p} = P\mathbf{X}_3$ .

**Διάνυσμα του κύριου άξονα.** Αν και κάθε σημείο  $\mathbf{X}$  που δεν ανήκει στο κύριο επίπεδο μπορεί να αντιστοιχηθεί σε σημείο της εικόνας σύμφωνα με  $\mathbf{x} = P\mathbf{X}$ , στην πραγματικότητα μόνο τα σημεία στον χώρο που βρίσκονται μπροστά από την κάμερα μπορούν να απεικονιστούν. Έστω ο προβολικός πίνακας  $P = [M|\mathbf{p}_4]$ . Παραπάνω, έγινε φανερό ότι το διάνυσμα  $\mathbf{m}_3$  (τελευταία γραμμή του  $M$ ) δείχνει προς την διεύθυνση του κύριου άξονα. Ωστόσο, δεν είναι γνωστό αν το διάνυσμα αυτό δείχνει προς την θετική κατεύθυνση (μπροστά από την κάμερα). Συγκεκριμένα, ο πίνακας  $P$  έχει οριστεί με αβεβαιότητα προσήμου. Για το λόγο αυτό δεν μπορεί να καθοριστεί αν το  $\mathbf{m}_3$  ή το  $-\mathbf{m}_3$  δείχνει προς τα θετικά.

Από τη σχέση 2.13, ισχύει  $\mathbf{x} = P_{cam}\mathbf{X}_{cam} = K[I|\mathbf{0}]\mathbf{X}_{cam}$ , όπου  $\mathbf{X}_{cam}$  περιγράφει τα σημεία του 3Δ κόσμου στο σύστημα συντεταγμένων της κάμερας. Στην περίπτωση αυτή το διάνυσμα  $\mathbf{v} = \det(M)\mathbf{m}_3 = (0, 0, 1)^T$  δείχνει προς τα θετικά (μπροστά από την κάμερα) στην διεύθυνση του κύριου άξονα, ανεξαρτήτως πολλαπλασιαστικής σταθεράς του πίνακα  $P_{cam}$ . Για παράδειγμα, αν  $P_{cam} \rightarrow kP_{cam}$  τότε  $\mathbf{v} \rightarrow k^4\mathbf{v}$  που έχει την ίδια διεύθυνση. Αν τα σημεία του χώρου εκφράζονται στο παγκόσμιο σύστημα συντεταγμένων, τότε  $P = kK[R|\tilde{R}\tilde{C}]$ . Επειδή  $\det(R) > 0$  το διάνυσμα  $\mathbf{v}$  δεν αλλάζει διεύθυνση για οποιαδήποτε τιμή του  $k$ .

Συμπερασματικά, το διάνυσμα  $\mathbf{v} = \det(M)\mathbf{m}_3$  είναι ένα διάνυσμα στην διεύθυνση του κύριου άξονα που δείχνει προς το μπροστινό μέρος της κάμερας.

### 2.3.2.2 Δράση της Προβολικής Κάμερας στα Σημεία

**Εμπρόσθια προβολή.** Όπως έχει αναφερθεί παραπάνω, μια γενική προβολική κάμερα προβάλλει ένα σημείο του χώρου  $\mathbf{X}$  σε ένα σημείο στην εικόνα σύμφωνα με τον τύπο  $\mathbf{x} = P\mathbf{X}$ . Τα σημεία στο άπειρο της μορφής  $\mathbf{D} = (\mathbf{d}, 0)^T$  προβάλλονται στα σημεία

$$\mathbf{x} = P\mathbf{D} = [M|\mathbf{p}_4]\mathbf{D} = M\mathbf{d} \quad (2.20)$$

**Οπίσθια προβολή σημείων σε ακτίνες.** Δεδομένου ενός σημείου  $\mathbf{x}$  στην εικόνα, είναι επιθυμητό να προσδιοριστεί ένα σύνολο σημείων στον χώρο που προβάλλονται στο σημείο αυτό. Τα σημεία αυτά αποτελούν μια ακτίνα στο χώρο που διέρχεται από το κέντρο της κάμερας.

Για να οριστεί επαρκώς η ακτίνα χρειάζονται δύο σημεία. Ένα σημείο είναι το κέντρο της κάμερας  $\mathbf{C}$  ( $P\mathbf{C} = \mathbf{0}$ ). Ένα δεύτερο σημείο είναι το σημείο  $P^+\mathbf{x}$  όπου  $P^+$  είναι ο ψευδο-αντίστροφος του  $P$ . Το σημείο αυτό ανήκει στην ακτίνα καθώς η προβολή του στην εικόνα είναι το σημείο  $\mathbf{x}$ . Η ακτίνα δίνεται λοιπόν από την εξίσωση

$$\mathbf{X}(\lambda) = P^+\mathbf{x} + \lambda\mathbf{C} \quad (2.21)$$

### 2.3.2.3 Βάθος Σημείων

Στη συνέχεια, μελετάται η απόσταση των σημείων μπροστά ή πίσω από το κύριο επίπεδο της κάμερας. Έστω ο πίνακας της κάμερας  $P = [M|\mathbf{p}_4]$ , που προβάλλει το σημείο  $\mathbf{X} = (X, Y, Z, 1)^T = (\tilde{\mathbf{X}}, 1)^T$  του κόσμου στο σημείο της εικόνας  $\mathbf{x} = w(x, y, 1)^T = P\mathbf{X}$ . Όπως αναφέρθηκε παραπάνω, ο οπτικός άξονας της κάμερας έχει διεύθυνση  $\mathbf{m}_3$  (το διάνυσμα της τρίτης γραμμής του πίνακα  $M$ ). Αν  $\mathbf{C} = (\tilde{\mathbf{C}}, 1)^T$  είναι το κέντρο της κάμερας, τότε ισχύει  $w = \mathbf{r}_3^T\mathbf{X} = \mathbf{r}_3^T(\mathbf{X} - \mathbf{C})$  αφού  $P\mathbf{C} = \mathbf{0}$ . Ωστόσο,  $\mathbf{r}_3^T(\mathbf{X} - \mathbf{C}) = \mathbf{m}_3^T(\tilde{\mathbf{X}} - \tilde{\mathbf{C}})$ . Επομένως, το  $w = \mathbf{m}_3^T(\tilde{\mathbf{X}} - \tilde{\mathbf{C}})$  μπορεί να ερμηνευθεί ως η προβολή του τμήματος που ενώνει το κέντρο της κάμερας και το σημείο στο 3Δ χώρο πάνω στον οπτικό άξονα της κάμερας. Αν ο πίνακας  $M$  είναι κανονικοποιημένος έτσι ώστε  $\det(M) > 0$  και  $\|\mathbf{m}_3\| = 1$ , τότε το διάνυσμα  $\mathbf{m}_3$  είναι μοναδιαίο και δείχνει προς τα θετικά του οπτικού άξονα. Στην περίπτωση αυτή, η ποσότητα  $w$  είναι το βάθος του σημείου  $\mathbf{X}$  από το κέντρο της κάμερας  $\mathbf{C}$  στη θετική κατεύθυνση του.

Αν ο πίνακας  $M$  δεν έχει κανονικοποιηθεί τότε το βάθος προκύπτει από το παρακάτω αποτέλεσμα

**Αποτέλεσμα 2.5.** Έστω  $\mathbf{X} = (X, Y, Z, T)^T$  είναι ένα 3Δ σημείο και  $P = [M|\mathbf{p}_4]$  είναι ο προβολικός πίνακας μια κάμερας. Αν ισχύει  $P(X, Y, Z, T)^T = w(x, y, 1)$  τότε

$$\text{depth}(\mathbf{X}; P) = \frac{\text{sign}(\det(M))w}{T\|\mathbf{m}_3\|} \quad (2.22)$$

είναι το βάθος του σημείου  $\mathbf{X}$  μπροστά από το κύριο επίπεδο.

### 2.3.3 Η Κάμερα στο Άπειρο

Στην κατηγορία αυτή, το κέντρο των καμερών βρίσκεται στο επίπεδο του απείρου. Αυτό σημαίνει ότι το αριστερό  $3 \times 3$  μέρος του πίνακα  $P$  είναι μη αντιστρέψιμο. Το κέντρο της κάμερας μπορεί να βρεθεί από τη σχέση  $PC = \mathbf{0}$  όπως και στις πεπερασμένες κάμερες.

Οι κάμερες στο άπειρο μπορούν να διαχωριστούν σε δύο κατηγορίες, τις *αφινικές* και τις *μη αφινικές*. Συγκεκριμένα, οι αφινικές κάμερες είναι αυτές των οποίων η τελευταία σειρά του πίνακα  $P$  είναι της μορφής  $(0, 0, 1)^T$ . Με άλλα λόγια, οι αφινικές κάμερες αντιστοιχίζουν σημεία στο άπειρο σε σημεία στο άπειρο. Περισσότερες λεπτομέρειες που αφορούν στην κατηγορία αυτή υπάρχουν στο [22].





## Κεφάλαιο 3

# Επιπολική Γεωμετρία και Δομή από Κίνηση

Η επιπολική γεωμετρία είναι η εσωτερική προβολική γεωμετρία μεταξύ δύο εικόνων. Είναι ανεξάρτητη από τον 3Δ κόσμο και εξαρτάται μόνο από τις εσωτερικές παραμέτρους της κάμερας και τη σχετική θέση των καμερών.

Ο θεμελιώδης πίνακας  $F$  ενσωματώνει την εσωτερική γεωμετρία. Είναι ένας  $3 \times 3$  πίνακας βαθμού 2. Αν ένα σημείο στον 3Δ χώρο  $\mathbf{X}$  απεικονίζεται ως  $\mathbf{x}$  στην πρώτη εικόνα και ως  $\mathbf{x}'$  στην δεύτερη, τότε τα σημεία αυτά ικανοποιούν τη σχέση  $\mathbf{x}'^T F \mathbf{x} = 0$ .

Με βάση τις ιδέες της επιπολικής γεωμετρίας, σκοπός είναι η εκτίμηση της 3Δ δομής σημείων στο χώρο από ένα σύνολο επικαλυπτόμενων εικόνων. Η διαδικασία αυτή συνήθως απαιτεί την εκτίμηση της 3Δ γεωμετρίας (δομή) καθώς και τη θέση της κάμερας (κίνηση). Η διαδικασία αυτή είναι γνωστή ως *δομή από κίνηση* (βλ. [46]).

Στις παρακάτω ενότητες θα αναφερθούν οι βασικότερες έννοιες της επιπολικής γεωμετρίας (βλ. [22] και [15]) όσο και της δομής από κίνηση (βλ. [45]).

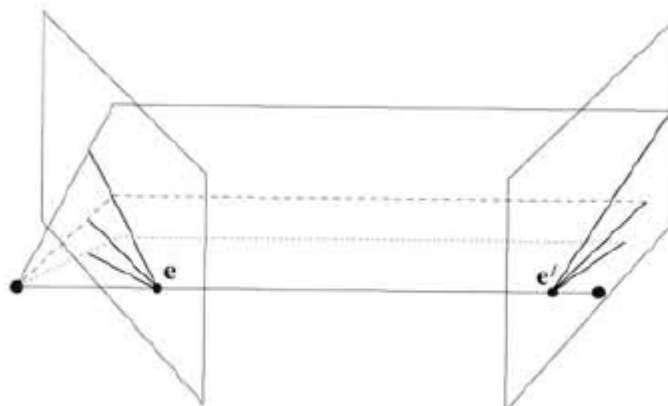
### 3.1 Επιπολική Γεωμετρία

Έστω ένα σημείο  $\mathbf{X}$  στον 3Δ χώρο που απεικονίζεται σε δύο εικόνες, στο  $\mathbf{x}$  στη πρώτη και  $\mathbf{x}'$  στη δεύτερη. Όπως φαίνεται στο Σχήμα 3.1 τα σημεία των εικόνων  $\mathbf{x}$ ,  $\mathbf{x}'$ , το σημείο του χώρου  $\mathbf{X}$  και τα κέντρα των καμερών ανήκουν στο ίδιο επίπεδο. Το επίπεδο αυτό σημειώνεται ως  $\pi$ . Προφανώς, οι ακτίνες που ενώνουν τα σημεία των εικόνων με τα αντίστοιχα κέντρα των καμερών τέμνονται στο  $\mathbf{X}$  και ανήκουν στο επίπεδο  $\pi$ .

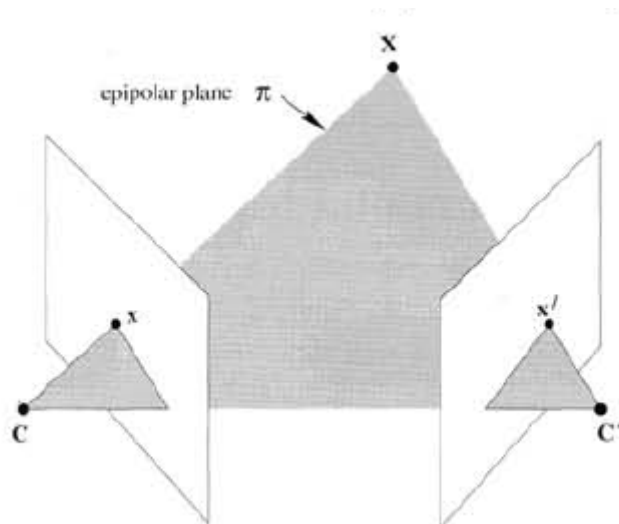
Έστω τώρα ότι γνωστό είναι μόνο το  $\mathbf{x}$ . Το επίπεδο  $\pi$  ορίζεται από τη βασική ευθεία (ευθεία που ενώνει τα κέντρα των καμερών) και την ακτίνα που αντιστοιχεί στο σημείο  $\mathbf{x}$ . Από τα παραπάνω, είναι γνωστό ότι η ευθεία που αντιστοιχεί στο (άγνωστο) σημείο  $\mathbf{x}'$  ανήκει στο  $\pi$ , και άρα το σημείο  $\mathbf{x}'$  ανήκει στην ευθεία  $l'$  που είναι η τομή του επιπέδου  $\pi$  με το επίπεδο της δεύτερης εικόνας (βλ. Σχήμα 3.2). Το αποτέλεσμα είναι ότι για την εύρεση του σημείου  $\mathbf{x}'$  απαιτείται αναζήτηση στην ευθεία  $l'$  και όχι σε ολόκληρο το επίπεδο  $\pi$ .

Παρακάτω δίνονται κάποιοι ορισμοί που σχετίζονται με την επιπολική γεωμετρία και στους οποίους θα γίνεται αναφορά σε όλη την εργασία.

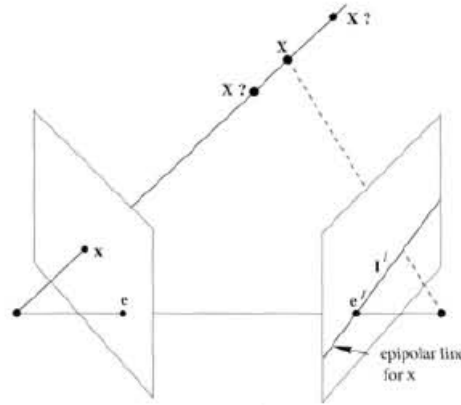
**Ορισμός 3.1. (το επιπολικό σημείο)** Το επιπολικό σημείο είναι το σημείο τομής της ευθείας που ενώνει τα κέντρα των καμερών (βασική ευθεία) με το επίπεδο της εικόνας. Ισοδύναμα, το επιπολικό σημείο είναι η απεικόνιση σε μια εικόνα του κέντρου της κάμερας της άλλης εικόνας.



**Σχήμα 3.1:** Οι δύο κάμερες φαίνονται από τα κέντρα τους  $C$  και  $C'$  και τα αντίστοιχα επίπεδα των εικόνων. Τα κέντρα των καμερών, το σημείο του χώρου  $X$  και η εικόνα του  $x, x'$  στις δύο εικόνες ανήκουν στο επίπεδο  $\pi$ . Το σχήμα αυτό προέρχεται από το [22].



**Σχήμα 3.2:** Η ακτίνα που ενώνει ένα σημείο της εικόνας  $x$  με το αντίστοιχο κέντρο της κάμερας ανήκει στον 3D χώρο. Η ακτίνα αυτή απεικονίζεται ως μια ευθεία  $l'$  στη δεύτερη εικόνα. Το σημείο του χώρου  $X$  που προβάλλεται στο  $x$  πρέπει να ανήκει στην ακτίνα αυτή και επομένως η εικόνα του  $X$  στη δεύτερη εικόνα πρέπει να ανήκει στην  $l'$ . Το σχήμα αυτό προέρχεται από το [22].



**Σχήμα 3.3:** Η οικογένεια επιπολικών επιπέδων. Τα σημεία  $e, e'$  είναι τα επιπολικά σημεία στις δύο εικόνες αντίστοιχα. Το σχήμα αυτό προέρχεται από το [22].

**Ορισμός 3.2. (το επιπολικό επίπεδο)** Το επιπολικό επίπεδο είναι ένα επίπεδο που περιέχει την βασική ευθεία. Υπάρχει μια μονοπαραμετρική οικογένεια από επιπολικά επίπεδα. (Σχήμα 3.3.)

**Ορισμός 3.3. (η επιπολική ευθεία)** 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.

### 3.1.1 Ο Θεμελιώδης Πίνακας

Ο θεμελιώδης πίνακας είναι η αλγεβρική αναπαράσταση της επιπολικής γεωμετρίας. Στην ενότητα, θα υπολογιστεί ο θεμελιώδης πίνακας μεταξύ ενός σημείου και της επιπολικής ευθείας, και θα αναφερθούν οι ιδιότητες του πίνακα.

Δεδομένου δύο εικόνες, όπως αναφέρθηκε παραπάνω, σε κάθε σημείο  $x$  σε μια εικόνα αντιστοιχεί μια επιπολική ευθεία  $l'$  στην άλλη. Όποιο σημείο  $x'$  στην δεύτερη εικόνα αποτελεί τμήρι του σημείου  $x$  πρέπει να ανήκει στην επιπολική ευθεία  $l'$ . Η επιπολική ευθεία είναι η προβολή στη δεύτερη εικόνα της ακτίνας που ενώνει το σημείο  $x$  με το κέντρο της πρώτης κάμερας  $C$ . Δηλαδή, υπάρχει μια αντιστοίχιση

$$x \rightarrow l' \quad (3.1)$$

από ένα σημείο στη μία εικόνα στην αντίστοιχη επιπολική ευθεία στην άλλη εικόνα. Θα αποδειχθεί παρακάτω ότι η αντιστοίχιση αυτή είναι ένας προβολικός μετασχηματισμός από σημεία σε ευθείες που αναπαρίσταται με το θεμελιώδη πίνακα  $F$ .

Έστω το σημείο  $x'$ , η επιπολική ευθεία  $l'$  που διέρχεται από το  $x'$  και το επιπολικό σημείο  $e'$  μπορεί να γραφεί ως  $l' = e' \times x' = [e']_{\times} x'$ . Το σημείο  $x'$  μπορεί να γραφεί ως  $x' = H_{\pi} x$ , προκύπτει

$$l' = [e']_{\times} H_{\pi} x = Fx \quad (3.2)$$

όπου  $F = [e']_{\times} H_{\pi}$  είναι ο θεμελιώδης πίνακας.

**Αποτέλεσμα 3.1.** Ο θεμελιώδης πίνακας  $F$  μπορεί να γραφεί ως  $F = [e']_{\times} H_{\pi}$ , όπου  $H_{\pi}$  είναι ο μετασχηματισμός από μία εικόνα σε μία άλλη μέσω οποιουδήποτε επιπέδου  $\pi$ . Επιπλέον, επειδή  $[e']_{\times}$  έχει βαθμό 2 και  $H_{\times}$  έχει βαθμό 3, ο  $F$  έχει βαθμό 2.

### 3.1.1.1 Αλγεβρικός Υπολογισμός του Θεμελιώδους Πίνακα

Υπάρχει αλγεβρικός τρόπος υπολογισμού του θεμελιώδους πίνακα από τους προβολικούς πίνακες των καμερών  $P, P'$ . Η ακτίνα που ενώνει το σημείο  $\mathbf{x}$  με το κέντρο της αντίστοιχης κάμερας προκύπτει από τη λύση της εξίσωσης  $P\mathbf{X} = \mathbf{x}$ . Η μονοπαραμετρική οικογένεια λύσεων είναι της μορφής

$$\mathbf{X}(\lambda) = P^+\mathbf{x} + \lambda\mathbf{C} \quad (3.3)$$

όπου  $P^+$  είναι ο ψευδοαντίστροφος του  $P$ , δηλαδή  $PP^+ = I$ , και  $\mathbf{C}$  είναι το κέντρο της κάμερας που ικανοποιεί τη σχέση  $P\mathbf{C} = \mathbf{0}$ . Η ακτίνα παραμετροποιείται από το βαθμωτό μέγεθος  $\lambda$ . Συγκεκριμένα, δύο σημεία στην ακτίνα είναι τα  $P^+\mathbf{x}$  (για  $\lambda = 0$ ), και το κέντρο της κάμερας  $\mathbf{C}$  (για  $\lambda = \infty$ ). Τα δύο αυτά σημεία έχουν εικόνες στη δεύτερη εικόνα στα  $P'P^+\mathbf{x}$  και  $P'\mathbf{C}$ , αντίστοιχα. Η επιπολική ευθεία είναι η ευθεία που ενώνει τα δύο αυτά προβαλλόμενα σημεία, δηλαδή  $\mathbf{I}' = (P'\mathbf{C}) \times (P'P^+\mathbf{x})$ . Το σημείο  $P'\mathbf{C}$  είναι το επιπολικό σημείο στη δεύτερη εικόνα, δηλαδή η προβολή του κέντρου της πρώτης κάμερας, και συμβολίζεται ως  $\mathbf{e}'$ . Επομένως,  $\mathbf{I}' = [\mathbf{e}'] \times (P'P^+)\mathbf{x} = F\mathbf{x}$  όπου ο πίνακας  $F$  δίνεται από τη σχέση

$$F = [\mathbf{e}'] \times P'P^+ \quad (3.4)$$

Η σχέση αυτή είναι ίδια με τη σχέση 3.2 όπου η ομογραφία  $H_\pi$  έχει την μορφή  $H_\pi = P'P^+$ . Τα παραπάνω δεν ισχύουν αν οι δύο κάμερες έχουν το ίδιο κέντρο, δηλαδή  $\mathbf{C} = \mathbf{C}'$ .

**Παράδειγμα.** Σε πολλές εφαρμογές ορίζεται το παγκόσμιο σύστημα συντεταγμένων να ταυτίζεται με το σύστημα συντεταγμένων μιας εκ των δύο καμερών. Δηλαδή, ισχύει για τους προβολικούς τους πίνακες

$$P = K[I|\mathbf{0}] \quad P' = K'[R|\mathbf{t}] \quad (3.5)$$

Τότε

$$P^+ = \begin{pmatrix} K^{-1} \\ \mathbf{0}^T \end{pmatrix}, \quad \mathbf{C} = \begin{bmatrix} \mathbf{0} \\ 1 \end{bmatrix}$$

και άρα

$$\begin{aligned} F &= [P'\mathbf{C}] \times P'P^+ \\ &= [K'\mathbf{t}] \times K'RK^{-1} = K'^{-T}[\mathbf{t}] \times RK^{-1} \\ &= K'^{-T}R[R^T\mathbf{t}] \times K^{-1} = K'^{-T}RK^T[KR^T\mathbf{t}] \times \end{aligned} \quad (3.6)$$

Τα επιπολικά σημεία που ορίζονται ως οι εικόνα του κέντρου της άλλης κάμερας είναι

$$\mathbf{e} = P \begin{bmatrix} -R^T\mathbf{t} \\ 1 \end{bmatrix} = KR^T\mathbf{t}, \quad \mathbf{e}' = P' \begin{bmatrix} \mathbf{0} \\ 1 \end{bmatrix} = K'\mathbf{t}$$

Επομένως, μπορούμε να γράψουμε τη σχέση 3.4

$$F = [\mathbf{e}'] \times K'RK^{-1} = K'^{-T}[\mathbf{t}] \times RK^{-1} = K'^{-T}R[R^T\mathbf{t}] \times K^{-1} = K'^{-T}RK^T[\mathbf{e}] \times \quad (3.7)$$

### 3.1.1.2 Ιδιότητες του Θεμελιώδους Πίνακα

**Συνθήκη αντιστοίχισης.** Ως τώρα, ο θεμελιώδης πίνακας όριζε την αντιστοίχιση  $\mathbf{x} \rightarrow \mathbf{I}'$ . Τώρα εξετάζεται η αντιστοίχιση σημείων ταιριάσματος  $\mathbf{x} \leftrightarrow \mathbf{x}'$ .

**Αποτέλεσμα 3.2.** Ο θεμελιώδης πίνακας ικανοποιεί τη συνθήκη ότι για κάθε ζεύγος αντίστοιχων σημείων  $\mathbf{x} \rightarrow \mathbf{x}'$  στις δύο εικόνες

$$\mathbf{x}'^T F \mathbf{x} = 0$$

Αυτο ισχύει γιατί αν το σημείο  $\mathbf{x}$  αντιστοιχεί στο σημείο  $\mathbf{x}'$ , τότε το  $\mathbf{x}'$  ανήκει στην επιπολική ευθεία  $\mathbf{l}' = F\mathbf{x}$  που αντιστοιχεί στο σημείο  $\mathbf{x}$ . Με άλλα λόγια,  $0 = \mathbf{x}'^T \mathbf{l}' = \mathbf{x}'^T F\mathbf{x}$ . Αντίστροφα, αν δύο σημεία ικανοποιούν τη σχέση  $\mathbf{x}'^T F\mathbf{x} = 0$ , τότε το  $\mathbf{x}$  ανήκει στην επιπολική ευθεία  $\mathbf{l}'$ . Επομένως, τα σημεία των εικόνων ανήκουν στο επιπολικό επίπεδο και άρα είναι αντίστοιχα. Το παραπάνω αποτέλεσμα επιτρέπει τον υπολογισμό του  $F$  μόνο από ζεύγη αντίστοιχων σημείων (βλ. Κεφάλαιο 5) και χωρίς τη γνώση των πινάκων των καμερών, όπως είδαμε παραπάνω. Από το παραπάνω αποτέλεσμα, δίνεται εναλλακτικός ορισμός για τον θεμελιώδη πίνακα.

**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

$$\mathbf{x}'^T F\mathbf{x} = 0$$

για όλα τα ζεύγη αντίστοιχων σημείων  $\mathbf{x} \leftrightarrow \mathbf{x}'$

**Μεταφορική ιδιότητα.** Αν  $F$  ο θεμελιώδης πίνακας για το ζευγάρι καμερών  $(P, P')$ , τότε  $F^T$  είναι ο θεμελιώδης πίνακας για το αντίστροφο ζευγάρι  $(P', P)$ .

**Επιπολικές ευθείες.** Για κάθε σημείο  $\mathbf{x}$  της πρώτης εικόνας, η αντίστοιχη επιπολική ευθεία είναι  $\mathbf{l}' = F\mathbf{x}$ . Όμοια, η  $\mathbf{l} = F^T \mathbf{x}'$  είναι η επιπολική ευθεία που αντιστοιχεί στο σημείο  $\mathbf{x}'$  της δεύτερης εικόνας.

**Το επιπολικό σημείο.** Για κάθε σημείο  $\mathbf{x}$  (εκτός του  $\mathbf{e}$ ) η επιπολική ευθεία  $\mathbf{l}' = F\mathbf{x}$  περιέχει το επιπολικό σημείο  $\mathbf{e}'$ . Επομένως,  $\mathbf{e}'$  ικανοποιεί τη σχέση  $\mathbf{e}'^T (F\mathbf{x}) = (\mathbf{e}'^T F)\mathbf{x} = 0$  για κάθε  $\mathbf{x}$ . Άρα  $\mathbf{e}'^T F = 0$ , δηλαδή το  $\mathbf{e}'$  γεννά το αριστερό μηδενικό υπόχωρο του πίνακα  $F$ . Όμοια,  $F\mathbf{e} = 0$ .

**Βαθμοί ελευθερίας.** Ο πίνακας  $F$  έχει 7 βαθμούς ελευθερίας: ένας  $3 \times 3$  ομογενής πίνακας έχει 8 ανεξάρτητα διανύσματα (9 στοιχεία πλην πολλαπλασιαστικής σταθεράς). Ωστόσο, ο πίνακας  $F$  ικανοποιεί τη σχέση  $\det(F) = 0$  που αφαιρεί ένα βαθμό ελευθερίας.

### 3.1.2 Ο Σημαντικός Πίνακας

Ο σημαντικός πίνακας είναι μια ειδική περίπτωση του θεμελιώδους πίνακα, στην περίπτωση που οι συντεταγμένες της εικόνας είναι κανονικοποιημένες. Δηλαδή, ο σημαντικός πίνακας ορίζεται με την υπόθεση μη ύπαρξης βαθμονόμησης. Παρακάτω δίνονται οι ιδιότητες του πίνακα αυτού.

**Κανονικοποιημένες συντεταγμένες.** Έστω μια κάμερα με πίνακα  $P = K[R | \mathbf{t}]$  και έστω  $\mathbf{x} = P\mathbf{X}$  ένα σημείο της εικόνας. Αν ο πίνακας βαθμονόμησης  $K$  είναι γνωστός, τότε εφαρμόζεται ο αντίστροφος του στο σημείο  $\mathbf{x}$  ώστε να προκύψει το σημείο  $\hat{\mathbf{x}} = K^{-1}\mathbf{x}$ . Τότε  $\hat{\mathbf{x}} = [R | \mathbf{t}]\mathbf{X}$ , όπου  $\hat{\mathbf{x}}$  είναι οι κανονικοποιημένες συντεταγμένες του σημείου της εικόνας. Μπορεί να θεωρηθεί ως η εικόνα του σημείου  $\mathbf{X}$  από κάμερα με πίνακα  $[R | \mathbf{t}]$  μοναδιαίο πίνακα βαθμονόμησης  $K = I$ . Ο πίνακας της κάμερας  $K^{-1}P = [R | \mathbf{t}]$  λέγεται *κανονικοποιημένος πίνακας κάμερας*.

Έστω ένα ζεύγος κανονικοποιημένων καμερών  $P = [I | \mathbf{0}]$  και  $P' = [R | \mathbf{t}]$ . Ο θεμελιώδης πίνακας που αντιστοιχεί στο ζεύγος των κανονικοποιημένων καμερών λέγεται *σημαντικός πίνακας* και προκύπτει από τη σχέση 3.6 ότι

$$E = [\mathbf{t}]_{\times} R = R[R^T \mathbf{t}]_{\times} \quad (3.8)$$

**Ορισμός 3.5.** Η καθοριστική εξίσωση για το σημαντικό πίνακα είναι

$$\hat{\mathbf{x}}'^T E \hat{\mathbf{x}} = 0 \quad (3.9)$$

ως προς τις κανονικοποιημένες συντεταγμένες για ζεύγη σημείων  $\mathbf{x} \leftrightarrow \mathbf{x}'$ .

Αντικαθιστώντας τα  $\hat{\mathbf{x}}$  και  $\hat{\mathbf{x}'}$  προκύπτει  $\mathbf{x}'^T K'^{-T} E K^{-1} \mathbf{x} = 0$ . Συγκρίνοντας τη σχέση αυτή με την εξίσωση  $\mathbf{x}'^T F \mathbf{x} = 0$ , προκύπτει ότι η σχέση που συνδέει τον θεμελιώδη με τον σημαντικό πίνακα είναι

$$E = K'^T F K \quad (3.10)$$

Ο σημαντικός πίνακας,  $E = [\mathbf{t}]_{\times} R$ , έχει μόνο 5 βαθμούς ελευθερίας καθώς ο πίνακας περιστροφής  $R$  και το διάνυσμα μετατόπισης  $\mathbf{t}$  έχουν 3 β.ελ. ο καθένας πλην μια πολλαπλασιαστική σταθερά - ο σημαντικός πίνακας είναι ομογενής.

**Αποτέλεσμα 3.3.** Ένας  $3 \times 3$  πίνακας είναι ένας σημαντικός πίνακας αν και μόνο αν οι δύο από τις ιδιάζουσες ιδιοτιμές του (singular values) είναι ίσες και η τρίτη είναι μηδέν.

*Απόδειξη*

Από τη σχέση 3.8  $E = [\mathbf{t}]_{\times} R = SR$  όπου  $S$  ένας αντισυμμετρικός πίνακας  $3 \times 3$  (περιττής διάστασης). Επομένως, σύμφωνα με το (βλ. Παράρτημα Α') ο πίνακας  $S$  θα έχει ιδιοτιμές  $(ik, -ik, 0)$  και μπορεί να αναλυθεί σύμφωνα με τον τύπο  $S = kUZU^T$ , όπου  $U$  ένας ορθογώνιος πίνακας και  $Z = \text{diag}(1, 1, 0)W$  με

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

Επόμενος, με ακρίβεια μιας πολλαπλασιαστικής σταθεράς ισχύει  $S = U \text{diag}(1, 1, 0) W U^T$  και  $E = SR = U \text{diag}(1, 1, 0) (W U^T R)$ . Η τελευταία εξίσωση παραπέμπει σε ανάλυση SVD του πίνακα  $E$  με δύο ίσες ιδιάζουσες ιδιοτιμές. Αντίστροφα, ένας πίνακας με δύο ίσες ιδιάζουσες ιδιοτιμές μπορεί να παραγοντοποιηθεί με τον  $SR$  τρόπο. Ο πίνακας  $V = R^T U W^T$  είναι ορθογώνιος ως γινόμενο ορθογωνίων πινάκων. Αφού  $E = U \text{diag}(1, 1, 0) V^T$ , είναι εμφανές ότι ο σημαντικός πίνακας έχει 6 βαθμούς ελευθερίας, καθώς οι  $U$  και  $V$  έχουν 3 β.ελ. ο καθένας. Επιπλέον, επειδή οι δύο ιδιάζουσες ιδιοτιμές είναι ίσες, η ανάλυση SVD δεν είναι μοναδική. Υπάρχει μια οικογένεια από SVD του. Για παράδειγμα, μια εναλλακτική ανάλυση SVD είναι  $E = (U \text{diag}(R_2, 1)) \text{diag}(1, 1, 0) (\text{diag}(R_2^T, 1) V^T)$

## 3.2 Δομή από Κίνηση

Αρχικά γίνεται αναφορά στη μέθοδο του τριγωνισμού (triangulation), δηλαδή στην εκτίμηση της 3Δ θέσης ενός σημείου όταν αυτό είναι εμφανές από πολλές όψεις. Στη δεύτερη υποενοότητα, γίνεται λόγος για τη δομή από κίνηση δύο όψεων, που αναφέρεται στον προσδιορισμό της επιπολικής γεωμετρίας μεταξύ δύο καμερών, και μπορεί επίσης να χρησιμοποιηθεί για τον προσδιορισμό των παραμέτρων τους.

### 3.2.1 Triangulation

Το πρόβλημα της εύρεσης της 3Δ θέσης ενός σημείου από ένα σύνολο αντίστοιχων σημείων στις εικόνες με γνωστές παραμέτρους των καμερών λέγεται *μέθοδος τριγωνισμού*.

Ο πιο απλός τρόπος είναι να θεωρηθεί το 3Δ σημείο  $\mathbf{P}$  (σε μη ομογενείς συντεταγμένες) το οποίο ανήκει πιο κοντά στις 3Δ ακτίνες που αντιστοιχούν στα 2Δ σημεία ταιριάσματος  $\{\mathbf{x}_j\}$  που

απεικονίζονται από τις κάμερες  $P_j = K_j[R_j \mid \mathbf{t}_j]$ , με  $\mathbf{t}_j = -R_j\tilde{\mathbf{C}}_j$  όπου  $\tilde{\mathbf{C}}_j$  είναι το κέντρο της  $j$ -στης κάμερας. Όπως φαίνεται στο Σχήμα 3.4, οι ακτίνες αυτές έχουν την αρχή τους στα κέντρα  $\tilde{\mathbf{C}}_j$ . Η  $j$ -στη ακτίνα έχει διεύθυνση  $\hat{\mathbf{v}}_j = \mathcal{N}(R_j^{-1}K_j^{-1}\mathbf{x}_j)$  διότι η ακτίνα σχηματίζεται από την ένωση των σημείων  $P_j^+\mathbf{x}_j$  και το  $(\tilde{\mathbf{C}}_j^T, 1)^T$ , δηλαδή τα 3Δ σημεία που ανήκουν στην ευθεία ικανοποιούν τη σχέση

$$\mathbf{X}_j(\lambda) = P_j^+\mathbf{x}_j + \lambda\mathbf{C}_j \quad (3.11)$$

Το σημείο του επιπέδου στο άπειρο  $\mathbf{D}_j = (R_j^{-1}K_j^{-1}\mathbf{x}_j)$  προβάλλεται στο σημείο  $P_j\mathbf{D}_j = K_j[R_j \mid \mathbf{t}_j]\mathbf{D}_j = \mathbf{x}_j$  και επομένως ανήκει στην  $j$ -στη ακτίνα. Τελικά προκύπτει ότι η διεύθυνση της ακτίνας είναι  $\hat{\mathbf{v}}_j = \mathcal{N}(R_j^{-1}K_j^{-1}\mathbf{x}_j)$ .

Το κοντινότερο σημείο του  $\tilde{\mathbf{P}}$  πάνω στην ακτίνα, που σημειώνεται με  $\tilde{\mathbf{Q}}_j$ , ελαχιστοποιεί την απόσταση

$$\|\tilde{\mathbf{C}}_j + d_j\hat{\mathbf{v}}_j - \tilde{\mathbf{P}}\|^2 \quad (3.12)$$

η οποία παρουσιάζει ελάχιστο στο  $d_j = \hat{\mathbf{v}}_j \cdot (\tilde{\mathbf{P}} - \tilde{\mathbf{C}}_j)$ . Επομένως,

$$\tilde{\mathbf{Q}}_j = \tilde{\mathbf{C}}_j + (\hat{\mathbf{v}}_j\hat{\mathbf{v}}_j^T)(\tilde{\mathbf{P}} - \tilde{\mathbf{C}}_j) = \tilde{\mathbf{C}}_j + (\tilde{\mathbf{P}} - \tilde{\mathbf{C}}_j)_{\parallel} \quad (3.13)$$

όπου το σύμβολο  $\mathbf{u}_{\parallel}$  δηλώνει την προβολή του διανύσματος  $\mathbf{u}$  στην 3Δ ακτίνα. Επιπλέον, εύκολα αποδεικνύεται ότι  $\hat{\mathbf{v}}(\hat{\mathbf{v}} \cdot \mathbf{u}) = (\hat{\mathbf{v}}\hat{\mathbf{v}}^T)\mathbf{u} = \mathbf{u}_{\parallel}$ .

Η απόσταση του σημείου  $\tilde{\mathbf{P}}$  από το  $\tilde{\mathbf{Q}}_j$  δίνεται από τη σχέση

$$r_j^2 = \|\tilde{\mathbf{P}} - \tilde{\mathbf{Q}}_j\|^2 = \|(I - \hat{\mathbf{v}}_j\hat{\mathbf{v}}_j^T)(\tilde{\mathbf{P}} - \tilde{\mathbf{C}}_j)\|^2 = \|(\tilde{\mathbf{P}} - \tilde{\mathbf{C}}_j)_{\perp}\|^2 \quad (3.14)$$

Η βέλτιστη τιμή του  $\tilde{\mathbf{P}}$ , που ανήκει πιο κοντά στις ακτίνες, μπορεί αν υπολογιστεί από τη λύση ενός προβλήματος ελαχίστων τετραγώνων, ελαχιστοποιώντας το συναρτησιακό  $\sum_j r_j^2$ . Σύμφωνα με το Παράρτημα Α' η λύση δίνεται από τη σχέση

$$\tilde{\mathbf{P}} = \left[ \sum_j I - \hat{\mathbf{v}}_j\hat{\mathbf{v}}_j^T \right]^{-1} \sum_j (I - \hat{\mathbf{v}}_j\hat{\mathbf{v}}_j^T)\tilde{\mathbf{C}}_j \quad (3.15)$$

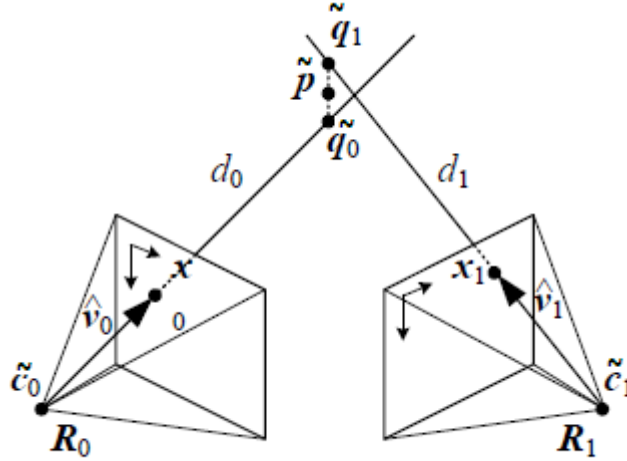
Εναλλακτική και πιο ευσταθής λύση είναι η εύρεση της λύσης  $\mathbf{P} = (X, Y, Z, W)^T$  (σε ομογενείς συντεταγμένες) που ελαχιστοποιεί τα παρακάτω υπόλοιπα

$$x_j = \frac{p_{00}^{(j)}X + p_{01}^{(j)}Y + p_{02}^{(j)}Z + p_{03}^{(j)}W}{p_{20}^{(j)}X + p_{21}^{(j)}Y + p_{22}^{(j)}Z + p_{23}^{(j)}W} \quad (3.16)$$

$$y_j = \frac{p_{10}^{(j)}X + p_{11}^{(j)}Y + p_{12}^{(j)}Z + p_{13}^{(j)}W}{p_{20}^{(j)}X + p_{21}^{(j)}Y + p_{22}^{(j)}Z + p_{23}^{(j)}W} \quad (3.17)$$

όπου  $(x_j, y_j)$  είναι οι συντεταγμένες του αντίστοιχου σημείου στην  $j$ -στη εικόνα και  $\{p_{00}^{(j)}, \dots, p_{23}^{(j)}\}$  είναι τα γνωστά στοιχεία του πίνακα  $P_j$ .

Το σύνολο των παραπάνω μη γραμμικών εξισώσεων μπορεί να μετατραπεί σε γραμμικό, αν οι εξισώσεις πολλαπλασιαστούν με τους αντίστοιχους παρονομαστές. Αν το σημείο του 3Δ χώρου εκφραστεί σε ομογενείς συντεταγμένες, δηλαδή  $\mathbf{P} = (X, Y, Z, W)^T$ , τότε το προκύπτον σύστημα είναι ομογενές και μπορεί να λυθεί με τη μέθοδο της ανάλυσης σε SVD (βλ. Παράρτημα Α'). Αν τεθεί  $W = 1$  τότε το σύστημα δεν είναι πλέον ομογενές και μπορεί να λυθεί με τη μέθοδο των ελαχίστων τετραγώνων. Η τελευταία μέθοδος θα αποτύχει σε ιδιάζουσες καταστάσεις, π.χ. αν οι ακτίνες είναι παράλληλες, δηλαδή αν το σημείο είναι μακριά από την κάμερα.



**Σχήμα 3.4:** Η μέθοδος του τριγωνισμού για την ανακατασκευή της 3Δ θέσης, βρίσκοντας το σημείο  $\tilde{\mathbf{P}}$  που βρίσκεται πιο κοντά σε όλες τις ακτίνες. Το σχήμα αυτό προέρχεται από το [45].

Τέλος, κατά τη διαδικασία του τριγωνισμού πρέπει να διασφαλιστεί ότι τα ανακατασκευασμένα σημεία του 3Δ χώρου να βρίσκονται μπροστά από όλες τις κάμερες. Επειδή κάτι τέτοιο δεν είναι εγγυημένο ότι δεν μπορεί να συμβεί, είναι σύνηθες να μεταφέρονται τα σημεία αυτά στο επίπεδο του απείρου, δηλαδή για τα σημεία που δεν βρίσκονται μπροστά από τις κάμερες να τίθεται  $W = 0$ .

### 3.2.2 Δομή από Κίνηση δύο Όψεων

Στην προηγούμενη υποενότητα, θεωρήσαμε ότι οι εξωτερικοί παράμετροι των καμερών είναι γνωστές. Στην υποενότητα αυτή, μελετάται η δομή από κίνηση, δηλαδή η ταυτόχρονη εκτίμηση της 3Δ δομής και της 3Δ θέσης των καμερών από αντιστοιχίες εικόνων.

Στο Σχήμα 3.5 φαίνεται ένα 3Δ σημείο  $\tilde{\mathbf{P}}$  όπως αυτό εικονίζεται από δύο κάμερες με σχετικό προσανατολισμό  $R$  και μετατόπιση  $\mathbf{t}$ . Εφόσον δεν είναι γνωστές οι θέσεις των καμερών, χωρίς πρόβλημα γενικότητας, είναι δυνατό να οριστεί η πρώτη κάμερα ως αρχή αναφοράς, δηλαδή  $\tilde{\mathbf{C}}_0 = \mathbf{0}$  και  $R_0 = I_{3 \times 3}$ . Με άλλα λόγια, το παγκόσμιο σύστημα συντεταγμένων ταυτίζεται με το σύστημα συντεταγμένων της πρώτης κάμερας.

Επομένως, ισχύει  $P_0 = K_0[I_{3 \times 3} | \mathbf{0}]$  και  $P_1 = K_1[R | \mathbf{t}]$ . Αν  $\tilde{\mathbf{P}}_0$  είναι οι συντεταγμένες ενός σημείου  $\tilde{\mathbf{P}}$  ως προς το σύστημα συντεταγμένων της πρώτης κάμερας, όπου το σημείο  $\tilde{\mathbf{P}}$  είναι εκφρασμένο ως προς το παγκόσμιο σύστημα συντεταγμένων, τότε ισχύει  $\tilde{\mathbf{P}}_1 = R\tilde{\mathbf{P}}_0 + \mathbf{t}$  και  $\tilde{\mathbf{P}}_0 = \tilde{\mathbf{P}}$ . Δηλαδή,

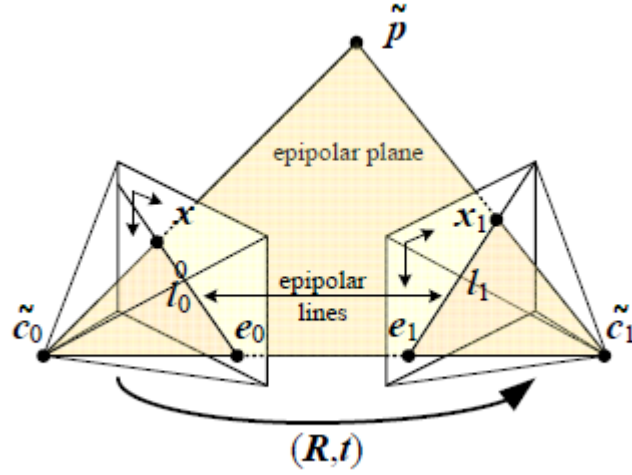
$$\tilde{\mathbf{P}}_1 = R\tilde{\mathbf{P}} + \mathbf{t} \quad (3.18)$$

Όπως αναφέρθηκε παραπάνω, στο παγκόσμιο σύστημα συντεταγμένων η ακτίνα που συνδέει το κέντρο της  $j$ -στής κάμερας με το 3Δ σημείο  $\tilde{\mathbf{P}}$  δίνεται από τη σχέση  $\hat{\mathbf{v}}_j = R_j^{-1}K_j^{-1}\mathbf{x}_j$ , όπου  $\mathbf{x}_j$  είναι η απεικόνιση του 3Δ σημείου στην  $j$ -στή εικόνα. Στο τοπικό σύστημα συντεταγμένων της  $j$ -στής κάμερας, η κατεύθυνση αυτή δίνεται από τη σχέση  $\hat{\mathbf{v}}_{j,cam} = R_j\hat{\mathbf{v}}_j = K_j^{-1}\mathbf{x}_j$ .

Λαμβάνοντας υπόψη την παραπάνω ανάλυση, ισχύει  $\tilde{\mathbf{P}} = \tilde{\mathbf{P}}_0 = d_0\hat{\mathbf{v}}_{0,cam}$  και  $\tilde{\mathbf{P}}_1 = d_1\hat{\mathbf{v}}_{1,cam}$ . Από τη σχέση 3.18

$$d_1\hat{\mathbf{v}}_{1,cam} = \tilde{\mathbf{P}}_1 = R\tilde{\mathbf{P}}_0 + \mathbf{t} = R(d_0\hat{\mathbf{v}}_{0,cam}) + \mathbf{t} \quad (3.19)$$





**Σχήμα 3.5:** Η επιπολική γεωμετρία. Η γεωμετρική αναπαράσταση του επιπολικού περιορισμού. Το σχήμα αυτό προέρχεται από το [45].

Πολλαπλασιάζοντας εξωτερικά από τα αριστερά την εξίσωση 3.19 με  $\mathbf{t}$  προκύπτει

$$d_1[\mathbf{t}]_{\times} \hat{\mathbf{v}}_{1,cam} = d_0[\mathbf{t}]_{\times} R \hat{\mathbf{v}}_{0,cam} \quad (3.20)$$

Πολλαπλασιάζοντας εσωτερικά την εξίσωση 3.20 με  $\hat{\mathbf{v}}_{1,cam}$  προκύπτει

$$d_0 \hat{\mathbf{v}}_{1,cam}^T ([\mathbf{t}]_{\times} R) \hat{\mathbf{v}}_{0,cam} = d_1 \hat{\mathbf{v}}_{1,cam}^T [\mathbf{t}]_{\times} \hat{\mathbf{v}}_{1,cam} = 0 \quad (3.21)$$

διότι ο πίνακας  $[\mathbf{t}]_{\times}$  είναι αντισυμμετρικός και όταν πολλαπλασιάζεται από τα δεξιά και τα αριστερά από το ίδιο διάνυσμα δίνει αποτέλεσμα 0. Έτσι καταλήγει κανείς στον *επιπολικό περιορισμό*

$$\hat{\mathbf{x}}_1^T ([\mathbf{t}]_{\times} R) \hat{\mathbf{x}}_0 = 0 \quad (3.22)$$

όπου  $\hat{\mathbf{v}}_{j,cam} = \hat{\mathbf{x}}_j = K_j^{-1} \mathbf{x}_j$  για  $j = 0, 1$  οι κανονικοποιημένες συντεταγμένες των αντίστοιχων σημείων της εικόνας και  $E = [\mathbf{t}]_{\times} R$  ο *σημαντικός πίνακας*.

Ένας άλλος τρόπος απόδειξης του επιπολικού περιορισμού είναι η παρατήρηση ότι το διάνυσμα που ενώνει τα δύο κέντρα  $\tilde{\mathbf{C}}_1 - \tilde{\mathbf{C}}_0 = -R^{-1} \mathbf{t}$  και οι κατευθύνσεις των ακτινών  $R_0^{-1} \hat{\mathbf{x}}_0$  και  $R_1^{-1} \hat{\mathbf{x}}_1$  (εκφρασμένα στο παγκόσμιο σύστημα συντεταγμένων) πρέπει να είναι συνεπίπεδα. Δηλαδή πρέπει το τριπλό γινόμενο

$$(\hat{\mathbf{x}}_0, R_1^{-1} \hat{\mathbf{x}}_1, -R^{-1} \mathbf{t}) = (R \hat{\mathbf{x}}_0, \hat{\mathbf{x}}_1, -\mathbf{t}) = \hat{\mathbf{x}}_1 \cdot (\mathbf{t} \times R \hat{\mathbf{x}}_0) = \hat{\mathbf{x}}_1^T ([\mathbf{t}]_{\times} R) \hat{\mathbf{x}}_0 = 0 \quad (3.23)$$

που ισχύει καθώς  $R$  είναι ορθογώνιος πίνακας.

Δεδομένου του επιπολικού περιορισμού, εξετάζεται πως μπορεί να ανακτηθεί η πληροφορία για την κίνηση της κάμερας. Έστω ότι υπάρχουν  $N$  το πλήθος αντιστοιχίσεις  $\mathbf{x}_{i0} \leftrightarrow \mathbf{x}_{i1}$  με  $i = 1, \dots, N$ , τότε σχηματίζονται  $N$  ομογενείς εξισώσεις από τα 9 στοιχεία του πίνακα  $E = \{e_{00}, \dots, e_{22}\}$ ,

$$\begin{aligned} & x_{i0} x_{i1} e_{00} + y_{i0} x_{i1} e_{01} + x_{i1} e_{02} \\ & + x_{i0} y_{i1} e_{00} + y_{i0} y_{i1} e_{11} + y_{i1} e_{12} \\ & + x_{i0} e_{20} + y_{i0} e_{21} + e_{22} = 0 \end{aligned} \quad (3.24)$$

όπου  $\mathbf{x}_{ij} = (x_{ij}, y_{ij}, 1)$ . Η παραπάνω εξίσωση μπορεί να γραφεί ως

$$[\mathbf{x}_{i1} \mathbf{x}_{i0}^T] \oplus E = Z_i \oplus E = \mathbf{z}_i \cdot \mathbf{f} = 0 \quad (3.25)$$

όπου  $\oplus$  συμβολίζει την πράξη του πολλαπλασιασμού στοιχείο προς στοιχείο και στη συνέχεια την άθροιση των τελικών στοιχείων. Τα διανύσματα  $\mathbf{z}_i$  και  $\mathbf{f}$  έχουν ως στοιχεία τα στοιχεία των πινάκων  $Z_i = \mathbf{x}_{i1} \mathbf{x}_{i0}^T$  και  $E$ , αντίστοιχα. Αν δίνονται  $N \geq 8$  τέτοιες εξισώσεις, είναι δυνατό να υπολογιστεί μια εκτίμηση - με ακρίβεια πολλαπλασιαστικής σταθεράς - του πίνακα  $E$ , με τη μέθοδο SVD.

Από τη μορφή της εξίσωσης 3.25, παρατηρεί κανείς ότι οι μετρήσεις των σημείων πολλαπλασιάζονται μεταξύ τους. Αν οι μετρήσεις αυτές περιέχουν θόρυβο, τότε ο θόρυβος αυτός αμβλύνεται εξαιτίας των πράξεων μεταξύ των μετρήσεων. Για να ελαττωθεί η επίδραση του θορύβου στη λύση της εξίσωσης, ο (Hartley, 1997) προτείνει την προεπεξεργασία των δεδομένων, έτσι ώστε το κέντρο των σημείων να μεταφερθεί στην αρχή και η μέση απόσταση από την αρχή να τεθεί σε μια συγκεκριμένη τιμή, δηλαδή

$$\tilde{x}_i = s(\hat{x}_i - \mu_{\hat{x}}) \quad (3.26)$$

$$\tilde{y}_i = s(\hat{y}_i - \mu_{\hat{y}}) \quad (3.27)$$

έτσι ώστε  $\sum_i \tilde{x}_i = \sum_i \tilde{y}_i = 0$  και  $\sum_i \sqrt{\tilde{x}_i^2 + \tilde{y}_i^2} = \sqrt{2}N$ , όπου  $N$  το πλήθος των σημείων. Η επεξεργασία αυτή μπορεί να γραφεί με τη μορφή πινάκων ως

$$\tilde{\mathbf{x}}_i = T_i \hat{\mathbf{x}}_i \quad (3.28)$$

όπου ο πίνακας μετασχηματισμού  $T_i$  είναι της μορφής

$$T_i = \begin{bmatrix} s & 0 & -\mu_{\hat{x}} \\ 0 & s & -\mu_{\hat{y}} \\ 0 & 0 & 1 \end{bmatrix}$$

με  $\mu_{\hat{x}} = \frac{\sum_i \hat{x}_i}{N}$ , όμοια για  $\mu_{\hat{y}}$  και  $s = \frac{\sum_i \sqrt{(\hat{x}_i - \mu_{\hat{x}})^2 + (\hat{y}_i - \mu_{\hat{y}})^2}}{N}$ .

Αφού υπολογιστεί ο πίνακας  $\tilde{E}$  από τα μετασχηματισμένα σημεία  $\{\tilde{\mathbf{x}}_{i0}, \tilde{\mathbf{x}}_{i1}\}$ , ο πίνακας  $E$  που αντιστοιχεί στα αρχικά (πριν το μετασχηματισμό) σημεία δίνεται από τη σχέση

$$E = T_1^{-1} \tilde{E} T_0^{-1} \quad (3.29)$$

Αφού υπολογιστεί ο σημαντικός πίνακας  $E$ , η διεύθυνση της μετατόπισης  $\mathbf{t}$  μπορεί να εκτιμηθεί. Η απόλυτη απόσταση δεν μπορεί να υπολογιστεί από μετρήσεις σημείων της εικόνας, ανεξάρτητα από το πόσες εικόνες ή πόσα σημεία διατίθενται. Γνώση διαστάσεων ή αρχικών θέσεων είναι απαραίτητη για να υπολογιστούν οι απόλυτες αποστάσεις.

Μια εκτίμηση αυτής της διεύθυνσης  $\hat{\mathbf{t}}$  προκύπτει από την παρατήρηση, σε συνθήκες έλλειψης θορύβου, ότι ο σημαντικός πίνακας είναι μη πλήρους βαθμού, δηλαδή  $\hat{\mathbf{t}}^T E = 0$ . Επομένως, θα ισχύει

$$E = [\hat{\mathbf{t}}]_{\times} R = U \Sigma V^T = \begin{bmatrix} \mathbf{u}_0 & \mathbf{u}_1 & \hat{\mathbf{t}} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \\ \mathbf{v}_3^T \end{bmatrix} \quad (3.30)$$

Ωστόσο, τις περισσότερες φορές ο σημαντικός πίνακας  $E$  υπολογίζεται από θορυβώδης μετρήσεις. Στην περίπτωση αυτή, το αριστερό ιδιάζον διάνυσμα (singular vector) που σχετίζεται με την μικρότερη ιδιάζουσα ιδιοτιμή είναι η διεύθυνση  $\hat{\mathbf{t}}$ . Οι άλλες δύο ιδιάζουσες ιδιοτιμές πρέπει να είναι όμοιες, όχι αναγκαστικά ίσες με 1 καθώς ο σημαντικός πίνακας είναι ομογενής πίνακας και άρα υπολογίζεται με ακρίβεια πολλαπλασιαστικής σταθεράς.

Μετά τον υπολογισμό της διεύθυνσης της μετατόπισης  $\hat{\mathbf{t}}$ , σκοπός είναι η εκτίμηση του προσανατολισμού  $R$  της δεύτερης κάμερας. Έστω το δεξιόστροφο ορθοκανονικό σύστημα βάσεων  $S = \{\hat{\mathbf{t}}, \hat{\mathbf{s}}_0, \hat{\mathbf{s}}_1\}$  του 3Δ χώρου  $\Omega$ . Αν  $v \in \Omega$  τότε  $v = \alpha\hat{\mathbf{t}} + \beta\hat{\mathbf{s}}_0 + \gamma\hat{\mathbf{s}}_1$ . Ο τελεστής  $[\hat{\mathbf{t}}]_{\times}$  επιδρά στο διάνυσμα ως εξής

$$\hat{\mathbf{t}} \times v = \beta\hat{\mathbf{t}} \times \hat{\mathbf{s}}_0 + \gamma\hat{\mathbf{t}} \times \hat{\mathbf{s}}_1 \quad (3.31)$$

$$= \beta\hat{\mathbf{s}}_1 - \gamma\hat{\mathbf{s}}_0 \quad (3.32)$$

Δηλαδή ο τελεστής  $[\hat{\mathbf{t}}]_{\times}$  μηδενίζει την  $\hat{\mathbf{t}}$  συνιστώσα του διανύσματος και περιστρέφει το αποτέλεσμα κατά  $+90^\circ$  (αν το σύστημα βάσεων  $S$  είναι αριστερόστροφο τότε περιστρέφεται κατά  $-90^\circ$ ). Σε μορφή πινάκων, η παραπάνω διαδικασία μπορεί να γραφεί

$$[\hat{\mathbf{t}}]_{\times} = SZR_{90^\circ}S^T = \begin{bmatrix} \hat{\mathbf{s}}_0 & \hat{\mathbf{s}}_1 & \hat{\mathbf{t}} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \hat{\mathbf{s}}_0^T \\ \hat{\mathbf{s}}_1^T \\ \hat{\mathbf{t}} \end{bmatrix} \quad (3.33)$$

με  $\hat{\mathbf{t}} = \hat{\mathbf{s}}_0 \times \hat{\mathbf{s}}_1$ . Από τις σχέσεις 3.30 και 3.33

$$E = [\hat{\mathbf{t}}]_{\times}R = SZR_{90^\circ}S^TR = U\Sigma V^T \quad (3.34)$$

Από την παραπάνω σχέση συμπεραίνουμε ότι  $U = S$ . Όπως είδαμε και παραπάνω, όταν ο σημαντικός πίνακας είναι απαλλαγμένος από θόρυβο τότε  $\Sigma = Z$ . Επομένως,

$$R_{90^\circ}U^TR = V^T \quad (3.35)$$

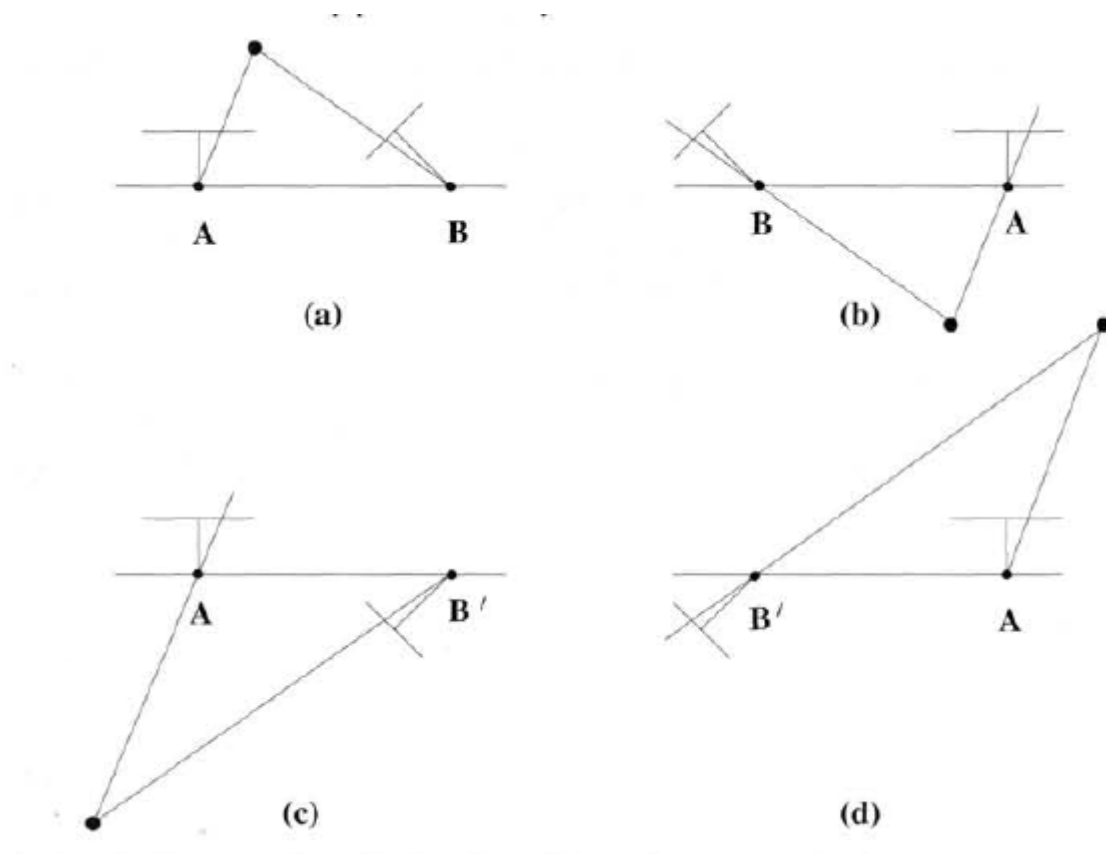
και

$$R = UR_{90^\circ}^TV^T \quad (3.36)$$

Δυστυχώς, οι  $E$  και  $\hat{\mathbf{t}}$  έχουν υπολογιστεί με ακρίβεια προσήμου. Επιπλέον, οι πίνακες  $U$  και  $V$  δεν είναι δεδομένο ότι είναι περιστροφικοί πίνακες (αν αντιστραφεί το πρόσημο τους δίνουν το ίδιο SVD αποτέλεσμα). Για το λόγο αυτό, είναι αναγκαίο να παραχθούν και οι τέσσερις δυνατοί πίνακες περιστροφής

$$R = \pm UR_{\pm 90^\circ}^TV^T \quad (3.37)$$

και παραμένουν αυτοί οι δύο με  $\det(R) = 1$ . Οι δύο εναπομείναντες πίνακες ζευγαρώνονται με τον τις διευθύνσεις μετατόπισης  $\pm\hat{\mathbf{t}}$  και τελικά επιλέγεται ο συνδυασμός με τον οποίο ο μεγαλύτερος αριθμός σημείων να φαίνεται από το μπροστινό μέρος των δύο καμερών. Συγκεκριμένα, όπως φαίνεται στο Σχήμα 3.6 επιλέγεται ο συνδυασμός της εικόνας (a), καθώς το σημείο φαίνεται από το μπροστινό μέρος των καμερών.



**Σχήμα 3.6:** Οι τέσσερις δυνατοί συνδυασμοί της βαθμονομημένης ανακατασκευής από τον πίνακα  $E$ . Η εικόνα (a) δίνει το μόνο συνδυασμό στον οποίο το σημείο φαίνεται από το μπροστινό μέρος των καμερών. Το σχήμα αυτό προέρχεται από το [22].

## Κεφάλαιο 4

# Ημιαυτόματη Μέθοδος Ανίχνευσης Σημείων Φυγής

Στο κεφάλαιο αυτό επιχειρείται η εύρεση μεθόδου ανίχνευσης σημείων φυγής σε μια εικόνα. Για το σκοπό αυτό, μελετάται και υλοποιείται ένας αλγόριθμος ανίχνευσης τέτοιων σημείων. Η βασική ιδέα πίσω από τον αλγόριθμο είναι αυτή της *στατιστικής σημαντικότητας*. Δηλαδή με λίγα λόγια, ένα γεγονός που συμβαίνει είναι σημαντικό αν η πιθανότητα να συμβεί είναι μικρή. Όσο πιο μικρή είναι αυτή η πιθανότητα, τόσο πιο σημαντικό είναι το γεγονός.

Ένα σύνολο παράλληλων ευθειών στον 3Δ χώρο προβάλλεται στην 2Δ εικόνα σε ένα σύνολο συντρεχουσών ευθειών. Το σημείο τομής αυτών των ευθειών στο επίπεδο της εικόνας καλείται *σημείο φυγής*. Αν οι ευθείες είναι παράλληλες στο επίπεδο της εικόνας, τότε το σημείο αυτό ανήκει στην ευθεία στο άπειρο της εικόνας. Παρά το γεγονός ότι η σύγκλιση ευθειών στο επίπεδο της εικόνας δεν σημαίνει απαραίτητα παραλληλία στον 3Δ χώρο, εξαιρέσεις δεν παρατηρούνται συχνά σε φυσικές εικόνες. Επομένως, η εύρεση συνόλων παράλληλων ευθειών στον 3Δ χώρο ανάγεται στο πρόβλημα εύρεσης συνόλου ευθειών που συγκλίνουν στο ίδιο σημείο στο επίπεδο της εικόνας.

Η ανίχνευση των σημείων φυγής μπορεί να χρησιμοποιηθεί για τον προσδιορισμό των παραμέτρων βαθμονόμησης (εσωτερικοί παράμετροι κάμερας) ή για άλλες εφαρμογές, όπως μέτρηση αποστάσεων από μια εικόνα (single-view metrology).

Τα στάδια του αλγορίθμου περιλαμβάνουν ανίχνευση ακμών και ευθειών στην εικόνα, διατήρηση των στατιστικά σημαντικών ευθειών, ανίχνευση πιθανών σημείων φυγής και διατήρηση στατιστικά σημαντικών σημείων φυγής. Κάθε ένα από τα παραπάνω στάδια αναλύονται στις παρακάτω ενότητες.

### 4.1 Ανίχνευση Ακμών και Ευθειών

Για την ανίχνευση των ευθειών στην εικόνα υλοποιείται ένας αλγόριθμος εύρεσης ευθυγράμμων τμημάτων από την εικόνα ακμών με χρήση ιδεών της ανάλυσης σε πρωτεύουσες συνιστώσες ή γνωστή στη διεθνή βιβλιογραφία ως PCA (βλ. [25]).

#### 4.1.1 Ανίχνευση Ακμών

Η ανίχνευση ακμών είναι μια γνωστή διαδικασία στον τομέα της επεξεργασίας εικόνων και της όρασης υπολογιστών και αναφέρεται στους αλγορίθμους ανίχνευσης σημείων όπου η φωτεινότητα της εικόνας αλλάζει απότομα ή παρουσιάζει ασυνέχειες.

Πολλές μέθοδοι έχουν βρεθεί και υλοποιηθεί για το σκοπό αυτό. Ο πιο γνωστός αλγόριθμος είναι αυτός του Canny (βλ. [11]).



**Σχήμα 4.1:** Εφαρμογή του αλγορίθμου ανίχνευσης ακμών Canny στην αριστερή εικόνα. Το αποτέλεσμα φαίνεται στη δεξιά εικόνα.

Ένα pixel ακμής σε ένα 2Δ σήμα (όπως μια εικόνα) ορίζεται όχι μόνο από τη θέση του αλλά και από τον προσανατολισμό της καμπύλης της αντίστοιχης ακμής. Έτσι, προσανατολισμένες 2Δ ακμές μπορούν να βρεθούν αναζητώντας μέγιστα της παραγώγου της εικόνας  $f$  σε διάφορες διευθύνσεις, ιδανικά στην διεύθυνση κάθετη στο όριο της ακμής. Η κάθετη αυτή διεύθυνση μπορεί να υπολογιστεί από το διάνυσμα  $\vec{n} = \frac{\nabla f}{\|\nabla f\|}$ .

Επομένως, ο Canny ανιχνευτής ακμών για εικόνες αποτελείται από τα παρακάτω βήματα: (1) Ομαλοποίηση της εικόνας  $f(x, y)$  με ένα ισοτροπικό 2Δ γκαουσιανό φίλτρο  $G_\sigma$ , (2) εύρεση των μηδενισμών της δεύτερης κατευθυντικής παραγώγου  $D_{\eta\eta}f$  της εικόνας στην κατεύθυνση  $\vec{n} = \frac{\nabla f}{\|\nabla f\|}$  (αυτό είναι ισοδύναμο με την εύρεση μεγίστων της κλίσης στην κατεύθυνση της κλίσης), (3) Απομόνωση μόνο εκείνων των σημείων που ανήκουν σε συνδεδεμένα τόξα, των οποίων τα σημεία έχουν ενέργεια ακμής που ξεπερνά ένα συγκεκριμένο διπλό κατώφλι υστέρησης. Στο Σχήμα 4.1 φαίνονται τα αποτελέσματα του αλγορίθμου

#### 4.1.2 Ανίχνευση ευθειών

Αφού ανιχνευθούν οι ακμές, εφαρμόζεται ο αλγόριθμος ανίχνευσης ευθειών. Παρακάτω αναλύονται τα βήματα του αλγορίθμου.

##### (a) Διαχωρισμός ακμών

Από την εικόνα ακμών διαχωρίζονται οι ακμές γραμμής και οι ακμές στήλης. Αρχικά γίνεται διαχωρισμός των pixels που αποτελούν τις ακμές σε "pixels γραμμής" και "pixels στήλης". Συγκεκριμένα, αν το εξεταζόμενο pixel ανήκει σε ακμή γραμμής τότε αποκτά την ιδιότητα του "pixel γραμμής". Αντίστοιχα για τα "pixels στήλης". Ο αλγόριθμος 1 περιγράφει τον αλγόριθμο διαχωρισμού των pixels.

Κάποια pixels ακμών μπορεί να ανήκουν και σε ακμές γραμμής και στήλης, αν είναι κεντρικά pixels σε δύο ακμές. Για το λόγο αυτό, μετά την παραπάνω διεργασία θα υπάρχουν τεσσάρων ειδών ιδιότητες για κάθε pixel ακμής, δηλαδή "pixel γραμμής", "pixel στήλης", "pixel γραμμής και στήλης" ή αλλιώς "κεντρικό pixel" και "ιδιαίτερο pixel". Η τελευταία ιδιότητα αναφέρεται στα pixels τα οποία δεν ανήκουν στις παραπάνω κατηγορίες.

Στη συνέχεια, επειδή τα ευθύγραμμα τμήματα έχουν σταθερή κατεύθυνση, διαχωρίζονται ακμές γραμμής και στήλης με κανόνα 8-συνεκτικότητας. Συγκεκριμένα, προκύπτουν τα σύνολα  $I_1 = \{(i,j): \text{αν } (i,j) \text{ δεν είναι "pixel στήλης"}\}$  και  $I_2 = \{(i,j): \text{αν } (i,j) \text{ δεν είναι "pixel γραμμής"}\}$ . Από τα σύνολα αυτά, προκύπτουν με κανόνα 8-συνεκτικότητας ακμές γραμμής και στήλης αντίστοιχα. Αν προκύψουν σε κάποιο σύνολο ακμές που αποτελούνται από "ιδιαίτερα pixels" και δεν αποτελούν τμήμα ακμής στο άλλο σύνολο τότε αυτή καταχωρείται στο σύνολο  $I_2$  (μεταφέρονται τα pixels της στο σύνολο αυτό). Ενώ αν η ακμή από "ιδιαίτερα pixels" στο αρχικό σύνολο είναι

---

**Algorithm 1** Διαχωρισμός σε pixels γραμμής και στήλης από την εικόνα ακμών  $p(i, j)$ , όπου  $(i, j)$  η θέση του pixel

---

**ΓΙΑ** (όλα τα  $i$  και  $j$ )

**ΑΝ** ( $p(i, j)$  και  $p(i + 1, j)$  είναι pixels ακμής )

δήλωσε τα  $p(i, j)$  και  $p(i + 1, j)$  ως "pixels γραμμή"

**ΑΝ** ( $p(i, j)$  και  $p(i - 1, j)$  είναι pixels ακμής )

πρόσθεσε στα  $p(i, j)$  και  $p(i - 1, j)$  την ιδιότητα της "pixels στήλης"

**ΑΛΛΙΩΣ ΑΝ** ( $p(i, j)$  είναι pixel ακμής αλλά δεν έχει καμία ιδιότητα )

δήλωσε το ως "ιδιαίτερο pixel"

---

τμήμα ακμής στο άλλο τότε η ακμή διαγράφεται από το αρχικό. Ο αλγόριθμος 2 περιγράφει τον παραπάνω αλγόριθμο.

---

**Algorithm 2** Διαχωρισμός ακμών γραμμής και στήλης από την ταξινόμηση των pixels

---

**ΓΙΑ** (όλα τα  $i$  και  $j$ )

**ΑΝ** ( $(i, j)$  δεν είναι "pixel στήλης" )

καταχώρηση του pixel στο σύνολο  $I_1$

**ΑΛΛΙΩΣ ΑΝ** ( $(i, j)$  δεν είναι "pixel γραμμής" )

καταχώρηση του pixel στο σύνολο  $I_2$

Εξαγωγή από τους πίνακες  $I_1$  και  $I_2$  των ευθυγράμμων τμημάτων με κανόνα 8-συνεκτικότητας και δήλωση των τμημάτων ως "ακμές γραμμής" και "ακμές στήλης" αντίστοιχα. Διαγραφή των ακμών από "ιδιαίτερα pixels" που ανήκουν σε ακμή του άλλου συνόλου. Μεταφορά των ακμών από "ιδιαίτερα pixels" που δεν ανήκουν σε κάποια άλλη ακμή.

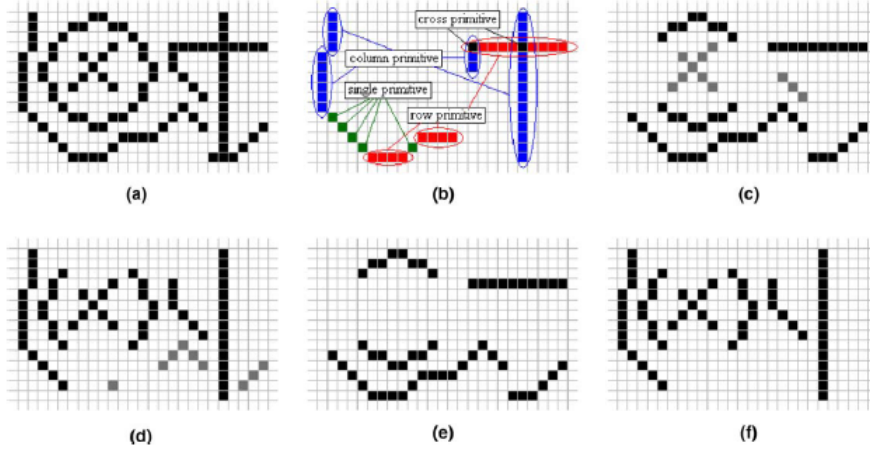
---

Στο Σχήμα 4.2 φαίνεται το αποτέλεσμα του διαχωρισμού των ακμών από μια αρχική εικόνα ακμών (a). Στην εικόνα (b) φαίνεται η ταξινόμηση των pixels στις διάφορες κατηγορίες. Στις εικόνες (c) και (d) παρατηρούνται οι ακμές γραμμής και στήλης, όπως προκύπτουν από την συνεκτικότητα των pixels που αποτελούν τα σύνολα  $I_1$ ,  $I_2$  αντίστοιχα. Με ανοιχτό γκρι χρώμα φαίνονται τα pixels τα πρέπει να διαγραφούν από τα δύο σύνολα. Στις εικόνες (e) και (f) παρατηρούνται οι τελικές ακμές γραμμής και στήλης αντίστοιχα μετά τις απαραίτητες αλλαγές που απαιτεί ο αλγόριθμος.

(b) Ανίχνευση ευθύγραμμων τμημάτων

Αφού γίνει ο διαχωρισμός των ακμών, επιχειρείται η ανίχνευση ευθυγράμμων τμημάτων με την χρήση της μεθόδου PCA (βλ. [25]).

Η μέθοδος του PCA (βλ. [14]) μπορεί να οριστεί ως ορθογώνια προβολή των δεδομένων σε ένα μικρότερης διάστασης διανυσματικό χώρο, έτσι ώστε η μεταβλητότητα των προβαλλόμενων δεδομένων να είναι μέγιστη. Έστω  $\{\mathbf{x}_n\}$  με  $n = 1, \dots, N$   $D$ -διάστατα δεδομένα. Σκοπός είναι η προβολή τους σε χώρο διάστασης  $M < D$ , μειοδοποιώντας τη μεταβλητότητα. Με άλλα λόγια, σκοπός είναι η εύρεση ορθογώνιων μοναδιαίων διανυσμάτων  $\mathbf{u}_i$  με  $i = 1, \dots, M$  έτσι ώστε ο χώρος διάστασης  $M$  να γεννάται από αυτά τα διανύσματα. Ένα παράδειγμα της μεθόδου φαίνεται στο Σχήμα 4.3. Με γκρι χρώμα φαίνονται τα δεδομένα. Το μεγάλο διάνυσμα αποτελεί την κατεύθυνση κατά την οποία τα δεδομένα παρουσιάζουν τη μεγαλύτερη μεταβλητότητα και αποτελεί την



**Σχήμα 4.2:** Το αποτέλεσμα του αλγορίθμου διαχωρισμού ακμών στην αρχική εικόνα (a). Στην εικόνα (b) γίνεται ο διαχωρισμός των pixels στις 4 ιδιότητες. Στα σχήματα (c) και (d) φαίνονται τα συνεκτικά pixels των συνόλων  $I_1, I_2$  αντίστοιχα. Στα σχήματα (e) και (f) φαίνονται οι τελικές ακμές γραμμής και στήλης αντίστοιχα. Το σχήμα αυτό προέρχεται από το [25].

πρωτεύουσα συνιστώσα.

Όπως αναφέρθηκε και παραπάνω, σκοπός της μεθόδου PCA είναι η εύρεση ορθογώνιων μοναδιαίων διανυσμάτων  $\mathbf{u}_i$  με  $i = 1, \dots, M$  τα οποία να γεννούν τον χώρο διάστασης  $M$ . Η προβολή των δεδομένων στο χώρο αυτό δίνει τη μεγαλύτερη μεταβλητότητα ή ισοδύναμα το μικρότερο σφάλμα.

Εύκολα προκύπτει ότι για  $M = 1$  η πρωτεύουσα συνιστώσα είναι το ιδιοδιάνυσμα του πίνακα συμμεταβλητότητας των δεδομένων που αντιστοιχεί στην μεγαλύτερη ιδιοτιμή. Συγκεκριμένα, για μονοδιάστατο χώρο είναι επιθυμητό να βρεθεί μοναδιαίο διάνυσμα  $\mathbf{u}_1$ . Αν  $\{\mathbf{x}_n\}, n = 1, \dots, N$  τα δεδομένα και

$$\tilde{\mathbf{x}} = \frac{1}{N} \sum_{n=0}^N \mathbf{x}_n \quad (4.1)$$

είναι η μέση τιμή τους τότε η μεταβλητότητα των προβαλλόμενων δεδομένων δίνεται από τη σχέση

$$\frac{1}{N} \sum_{n=0}^N (\mathbf{u}_1^T \mathbf{x}_n - \mathbf{u}_1^T \tilde{\mathbf{x}})^2 = \mathbf{u}_1^T S \mathbf{u}_1 \quad (4.2)$$

όπου  $S$  είναι ο πίνακας συμμεταβλητότητας που ορίζεται από

$$S = \frac{1}{N} \sum_{n=0}^N (\mathbf{x}_n - \tilde{\mathbf{x}})(\mathbf{x}_n - \tilde{\mathbf{x}})^T \quad (4.3)$$

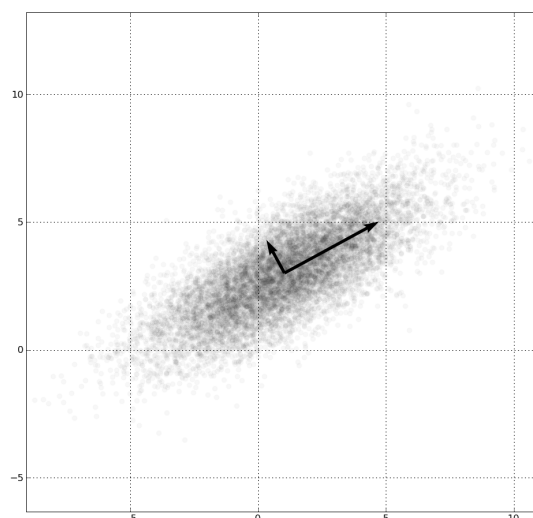
Σκοπός είναι η μεγιστοποίηση της μεταβλητότητας  $\mathbf{u}_1^T S \mathbf{u}_1$  ως προς  $\mathbf{u}_1$  με περιορισμό το μοναδιαίο μέτρο του διανύσματος. Με άλλα λόγια, επιθυμείτε να μεγιστοποιηθεί το συναρτησιακό

$$J = \mathbf{u}_1^T S \mathbf{u}_1 + \lambda(\mathbf{u}_1^T \mathbf{u}_1 - 1) \quad (4.4)$$

Μηδενίζοντας την παράγωγο του  $J$  ως προς  $\mathbf{u}_1$  προκύπτει η σχέση

$$S \mathbf{u}_1 = \lambda \mathbf{u}_1 \quad (4.5)$$





**Σχήμα 4.3:** PCA σε διδιάστατη γκαουσιανή κατανομή δεδομένων. Τα διανύσματα αντιστοιχούν στις πρωτεύουσες συνιστώσες των δεδομένων

η οποία δηλώνει ότι το διάνυσμα  $\mathbf{u}_1$  πρέπει να είναι ιδιοδιάνυσμα του πίνακα  $S$ . Αν πολλαπλασιάσουμε από αριστερά τη σχέση 4.5 με  $\mathbf{u}_1^T$  και λαμβάνοντας υπόψη τον περιορισμό του μοναδιαίου μέτρου προκύπτει ότι

$$\mathbf{u}_1^T S \mathbf{u}_1 = \lambda \quad (4.6)$$

Επομένως, η μεταβλητότητα θα είναι μέγιστη όταν επιλεγθεί εκείνο το ιδιοδιάνυσμα που αντιστοιχεί στην μεγαλύτερη ιδιοτιμή. Το ιδιοδιάνυσμα αυτό είναι γνωστό ως η *πρωτεύουσα συνιστώσα*.

Η ιδέα του PCA μπορεί να χρησιμοποιηθεί για την ανίχνευση δεδομένων που ανήκουν σε ευθεία γραμμή στο 2D χώρο. Έστω  $\{\mathbf{x}_n\}$ ,  $n = 1, \dots, N$  με  $\mathbf{x}_n = (x_n, y_n)$  συνευθειακά σημεία στο 2D χώρο. Όπως αποδείχθηκε παραπάνω, τα ιδιοδιανύσματα του πίνακα συμμεταβλητότητας των δεδομένων εκφράζουν τις κύριες κατευθύνσεις κατανομής των δεδομένων ενώ οι αντίστοιχες ιδιοτιμές το μήκος της κατανομής αυτής. Στην ιδανική περίπτωση που τα σημεία είναι συνευθειακά, η μικρότερη εκ των δύο ιδιοτιμών  $\lambda_2$  πρέπει να είναι μηδενική. Αν στα σημεία έχει προστεθεί θόρυβος με αποτέλεσμα να μην σχηματίζουν τέλεια ευθεία (π.χ. pixels σε ψηφιακή εικόνα) τότε η ιδιοτιμή δεν μπορεί να λάβει τη μηδενική τιμή. Η ανοχή του θορύβου θα εξαρτηθεί από το μέγεθος της θετικής ιδιοτιμής  $\lambda_2$  (ο πίνακας  $S$  είναι θετικά ημιορισμένος). Συγκεκριμένα, ο πίνακας συμμεταβλητότητας είναι

$$S = \begin{bmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{bmatrix}$$

όπου τα στοιχεία δίνονται από τις σχέσεις

$$s_{11} = \frac{1}{N} \sum_{n=1}^N (x_n - \tilde{x})^2 \quad (4.7)$$

$$s_{12} = s_{21} = \frac{1}{N} \sum_{n=1}^N (x_n - \tilde{x})(y_n - \tilde{y}) \quad (4.8)$$

$$s_{22} = \frac{1}{N} \sum_{n=1}^N (y_n - \tilde{y})^2 \quad (4.9)$$

όπου  $\tilde{x} = \frac{1}{N} \sum_{n=1}^N x_n$ ,  $\tilde{y} = \frac{1}{N} \sum_{n=1}^N y_n$  οι μέσες τιμές των δύο συντεταγμένων των σημείων.

Από τον πίνακα  $S$  εύκολα υπολογίζονται οι ιδιοτιμές  $\lambda_1, \lambda_2$  ( $\lambda_1 > \lambda_2$ ) και τα ιδιοδιανύσματα του  $\mathbf{u}_1, \mathbf{u}_2$ . Η τιμή της μικρότερης ιδιοτιμής  $\lambda_2$  θα αποτελέσει κριτήριο για την ύπαρξη ευθείας. Πρέπει, λοιπόν, να βρεθεί μια τιμή κατωφλίου  $t$  έτσι ώστε αν  $\lambda_2 > t$  τα σημεία να μη θεωρούνται συνευθειακά. Στο [25] προτείνεται ο παρακάτω εμπειρικός τύπος  $t = \frac{l^2}{30^2} t_a$ , όπου  $l$  είναι το μήκος του ευθυγράμμου τμήματος (ακμή γραμμής ή ακμή στήλης) και  $t_a$  λαμβάνει τιμές μεταξύ 0.25 και 0.3. Το μήκος της ακμής μετράται ως το πλήθος των pixels που την αποτελούν.

(c) Αποτελέσματα

Στο Σχήμα 4.4 φαίνονται τα αποτελέσματα στην αρχική εικόνα (a). Στην εικόνα (b) φαίνονται οι ακμές της αρχικής εικόνας όπως προκύπτουν από τον ανιχνευτή ακμών του Canny. Στις εικόνες (c) και (d) φαίνονται οι ακμές γραμμής και στήλης αντίστοιχα. Στην εικόνα (e) φαίνονται τα ευθύγραμμα τμήματα που αναγνωρίστηκαν από τις ακμές με  $t_a = 0.25$ .

## 4.2 Ε-σημαντικά Ευθύγραμμα Τμήματα

Στο στάδιο αυτό σκοπός είναι ο διαχωρισμός των στατιστικά σημαντικών ευθύγραμμων τμημάτων, δηλαδή των ευθύγραμμων ακμών της εικόνας που με κάποια πιθανότητα αντιστοιχούν σε ευθύγραμμα τμήματα στον 3Δ κόσμο. Συγκεκριμένα, ευθύγραμμες ακμές σε μια ψηφιακή εικόνα με θόρυβο δεν αντιστοιχούν απαραίτητα σε ευθείς στον 3Δ χώρο. Οι ακμές αυτές μπορεί να έχουν προέλθει με τυχαία ευθυγράμμιση κάποιων pixels. Στο [12] εισάγεται η έννοια της *στατιστικής σημαντικότητας* για την εύρεση σημαντικών ευθύγραμμων τμημάτων. Η μέθοδος που προτείνεται στο [12] προσαρμόζεται στη μέθοδο ανίχνευσης ευθειών που περιγράφηκε παραπάνω.

Έστω  $A$  ένα ευθύγραμμο τμήμα με μήκος  $l$  και  $x_1, x_2, \dots, x_l$  είναι  $l$  ανεξάρτητα σημεία του  $A$ . Αν  $\mathbf{X}_i$  είναι μια τυχαία μεταβλητή που παίρνει την τιμή 1 όταν η διεύθυνση στο pixel  $x_i$  είναι ευθυγραμμισμένη με αυτή του τμήματος  $A$  με ακρίβεια  $p$  και την τιμή 0 αλλιώς. Επομένως, η τυχαία μεταβλητή  $\mathbf{X}_i$  ακολουθεί την παρακάτω κατανομή bernoulli

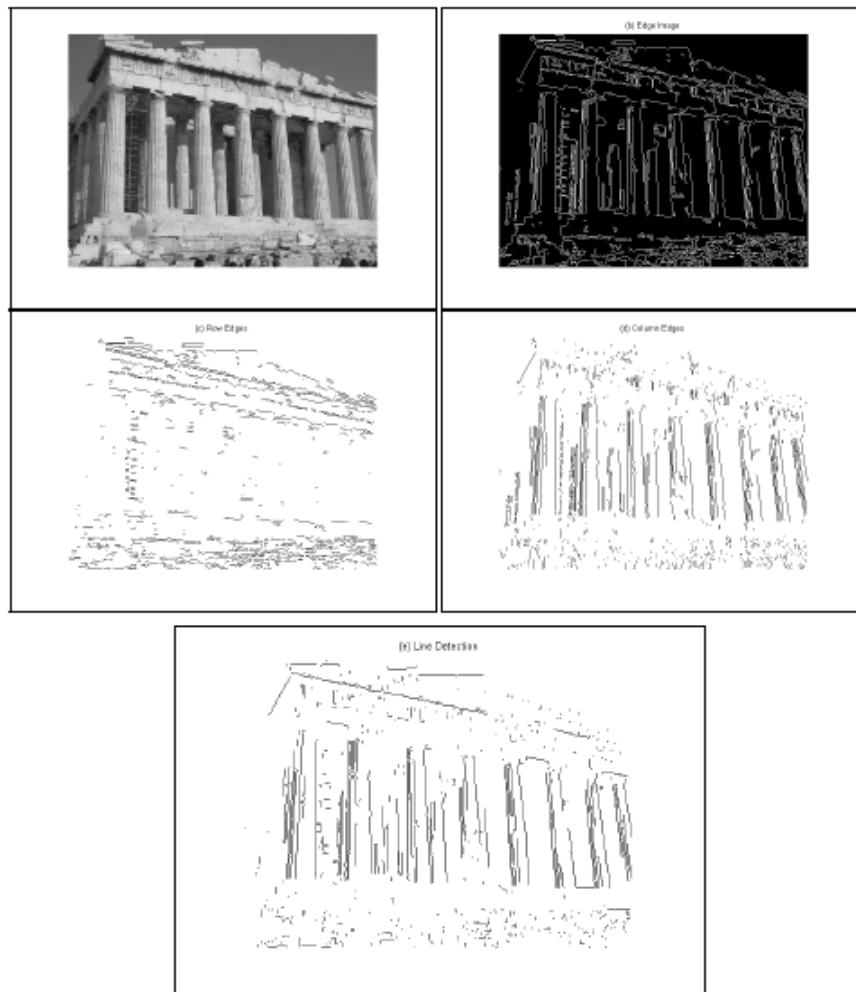
$$\mathbb{P}[\mathbf{X}_i = 1] = p, \mathbb{P}[\mathbf{X}_i = 0] = 1 - p \quad (4.10)$$

Η τυχαία μεταβλητή που παριστάνει τον αριθμό των  $x_i$  που έχουν τη σωστή διεύθυνση είναι

$$\mathbf{S}_l = \mathbf{X}_1 + \mathbf{X}_2 + \dots + \mathbf{X}_l \quad (4.11)$$

Εξαιτίας της ανεξαρτησίας των  $\mathbf{X}_i$ , η τ.μ.  $\mathbf{S}_l$  δίνεται από την διωνυμική κατανομή

$$\mathbb{P}[\mathbf{S}_l = k] = \binom{l}{k} p^k (1 - p)^{l-k} \quad (4.12)$$



**Σχήμα 4.4:** Το αποτέλεσμα της μεθόδου ανίχνευσης ευθειών με τη βοήθεια του PCA στην εικόνα (a) (πηγή: google images). Οι ακμές της εικόνας ως αποτέλεσμα του Canny αλγορίθμου ανίχνευσης ακμών φαίνονται στην εικόνα (b). Οι ακμές γραμμής και ακμές στήλης στις εικόνες (c) και (d) αντίστοιχα. Τα ευθύγραμμα τμήματα που ανιχνεύθηκαν φαίνονται στην εικόνα (e).

Δεδομένου ενός ευθύγραμμου τμήματος μήκους  $l$  ζητείται να βρεθεί αν το τμήμα είναι  $\epsilon$ -σημαντικό ή όχι μεταξύ όλων των τμημάτων της εικόνας και όχι μόνο μεταξύ αυτών με το ίδιο μήκος. Έστω  $m(l)$  ο αριθμός των προσανατολισμένων τμημάτων μήκους  $l$  στην  $N \times N$  θορυβώδη εικόνα. Ως αριθμός προσανατολισμένων τμημάτων στην εικόνα ορίζεται ο αριθμός των ζευγών  $(x, y)$  σημείων (κάθε ευθύγραμμο τμήμα ορίζεται από ένα αρχικό και ένα τελικό σημείο) και άρα προκύπτει

$$\sum_{l=1}^{l_{max}} m(l) = N^2(N^2 - 1) \simeq N^4 \quad (4.13)$$

καθώς μια  $N \times N$  εικόνα αποτελείται από  $N^2$  σημεία και επομένως διακρίνονται  $N^2(N^2 - 1)$  προσανατολισμένα ζεύγη. Για χάρη απλότητας διατηρείται η τιμή  $N^4$ . Όλα τα προσανατολισμένα τμήματα στην εικόνα αριθμούνται με δείκτες από  $i = 1$  ως  $i = N^4$ .

**Ορισμός 2.1. (κατώφλια ανίχνευσης)** Ως κατώφλια ανίχνευσης καλείται ένα σύνολο από θετικές τιμές  $w(l, \epsilon, N)$ ,  $1 \leq l \leq l_{max}$  τέτοιο ώστε

$$\sum_{l=1}^{l_{max}} w(l, \epsilon, N)m(l) \leq \epsilon$$

**Ορισμός 2.2. ( $\epsilon$ -σημαντικά ευθύγραμμα τμήματα)** Ένα ευθύγραμμο τμήμα μήκους  $l$  σε μια  $N \times N$  εικόνα λέγεται  $\epsilon$ -σημαντικό αν περιέχει τουλάχιστον  $k(l)$  σημεία με διεύθυνση ευθυγραμμισμένη με αυτή του τμήματος, όπου  $k(l)$  δίνεται από

$$k(l) = \min\{k \in \mathbb{N}, \mathbb{P}[\mathbf{S}_l \geq k] \leq w(l, \epsilon, N)\}$$

Ακολουθεί η εξήγηση του ορισμού. Για  $1 \leq i \leq N^4$ , έστω  $e_i$  το γεγονός "το  $i$ -στο τμήμα είναι  $\epsilon$ -σημαντικό" και  $\chi_{e_i}$  είναι η χαρακτηριστική συνάρτηση του γεγονότος  $e_i$ . Ισχύει

$$\mathbb{P}[\chi_{e_i} = 1] = \mathbb{P}[\mathbf{S}_{l_i} \geq k(l_i)] \quad (4.14)$$

όπου  $l_i$  είναι το μήκος του  $i$ -στού τμήματος. Να σημειωθεί ότι αν  $l_i$  είναι μικρό τότε έχουμε  $\mathbb{P}[\mathbf{S}_{l_i} \geq k(l_i)] = 0$ . Έστω  $R$  μια τυχαία μεταβλητή που αντιπροσωπεύει τον ακριβή αριθμό των  $e_i$  που συμβαίνουν ταυτόχρονα. Τότε  $R = \chi_{e_1} + \chi_{e_2} + \dots + \chi_{e_{N^4}}$ , τότε η αναμενόμενη του  $R$  είναι

$$\mathbb{E}[R] = \mathbb{E}[\chi_{e_1}] + \mathbb{E}[\chi_{e_2}] + \dots + \mathbb{E}[\chi_{e_{N^4}}] = \sum_{l=1}^{l_{max}} m(l)\mathbb{P}[\mathbf{S}_l \geq k(l)] \quad (4.15)$$

Ωστόσο εξ'ορισμού ισχύει

$$\mathbb{P}[\mathbf{S}_l \geq k(l)] \leq w(l, \epsilon, N) \quad (4.16)$$

έτσι ώστε

$$\mathbb{E}[R] \leq \sum_{l=1}^{l_{max}} w(l, \epsilon, N)m(l) \leq \epsilon \quad (4.17)$$

Αυτό σημαίνει ότι η αναμενόμενη τιμή του αριθμού των  $\epsilon$ -σημαντικών τμημάτων σε μια εικόνα είναι μικρότερη από  $\epsilon$ . Μένει, λοιπόν, να οριστούν τα κατώφλια ανίχνευσης. Επειδή δεν υπάρχει λόγος να υπάρχουν ξεχωριστά κατώφλια ανίχνευσης ανάλογα με το μήκος των τμημάτων, από τις σχέσεις 4.13 και 4.17 προκύπτει

$$\forall l \geq 1, w(l, \epsilon, N) = \frac{\epsilon}{N^4} \quad (4.18)$$

**Ορισμός 2.3.** (ε-σημαντικά τμήματα - τελικός ορισμός) Ένα ευθύγραμμο τμήμα μήκους  $l$  είναι ε-σημαντικό σε μια  $N \times N$  εικόνα αν περιέχει τουλάχιστον  $k(l)$  σημεία με διεύθυνση ευθυγραμμισμένη με αυτή του τμήματος, όπου  $k(l)$  δίνεται από τη σχέση

$$k(l) = \min\{k \in \mathbb{N}, \mathbb{P}[\mathbf{S}_l \geq k] \leq \frac{\epsilon}{N^4}\}$$

και ισχύει

$$\mathcal{B}(l, k, p) = \mathbb{P}[\mathbf{S}_l \geq k]$$

όπου  $\mathcal{B}(l, k, p) = \sum_{i=k}^l \binom{l}{i} p^i (1-p)^{l-i}$ .

*Ο Αλγόριθμος*

Αφού έχουν ανιχνευθεί τα ευθύγραμμα τμήματα στην εικόνα, επιθυμητός είναι ο χαρακτηρισμός τους ως στατιστικά σημαντικά ή όχι, με βάση τους παραπάνω ορισμούς. Τα ευθύγραμμα τμήματα χαρακτηρίζονται είτε από τα σημεία αρχής και τέλους τους είτε από τις παραμέτρους  $(\alpha_i, \beta_i, \gamma_i)$  τέτοιοι ώστε να ισχύει για το  $i$ -στο ευθύγραμμο τμήμα  $\alpha_i x + \beta_i y + \gamma_i = 0$ . Αρχικά, για κάθε σημείο της εικόνας υπολογίζεται η διεύθυνση τους. Συγκεκριμένα, η διεύθυνση ενός σημείου  $(n, m)$  της εικόνας σχετίζεται με την τιμή στα σημεία της  $2 \times 2$  γειτονιάς και ορίζεται από τη σχέση

$$dir(n, m) = \frac{1}{\|Du(n, m)\|} Du(n, m) \perp \quad (4.19)$$

όπου

$$Du(n, m) = \begin{pmatrix} u_x \\ u_y \end{pmatrix} = \frac{1}{2} \begin{pmatrix} u(n, m) + u(n+1, m) - u(n, m+1) - u(n+1, m+1) \\ u(n+1, m) + u(n+1, m+1) - u(n, m) - u(n, m+1) \end{pmatrix} \quad (4.20)$$

Δύο σημεία  $\mathbf{X}, \mathbf{Y}$  έχουν την ίδια διεύθυνση με ακρίβεια  $p$  αν

$$|\text{Angle}(dir(\mathbf{X}), dir(\mathbf{Y}))| \leq p\pi \quad (4.21)$$

Στη συνέχεια, θέτοντας  $\epsilon = 1$  υπολογίζονται τα μεγέθη  $k(l)$  σύμφωνα με τον Ορισμό 2.3 για  $1 \leq l \leq l_{max}$ , όπου  $l_{max}$  είναι το μήκος του μεγαλύτερου ευθυγράμμου τμήματος που έχει ανιχνευθεί.

Για κάθε ευθεία  $i$  που έχει ανιχνευθεί, αποφασίζεται αν είναι ε-σημαντική. Σκοπός είναι να υπολογιστεί το πλήθος των pixels από τα οποία διέρχεται η ευθεία σε όλη την έκταση της εικόνας και τα οποία είναι ευθυγραμμισμένα με αυτήν. Για το λόγο αυτό, με κέντρο τα pixel τα οποία ανήκουν στην ευθεία εξετάζεται μια  $3 \times 3$  γειτονιά για την εύρεση ευθυγραμμισμένων pixels σύμφωνα με τον τύπο

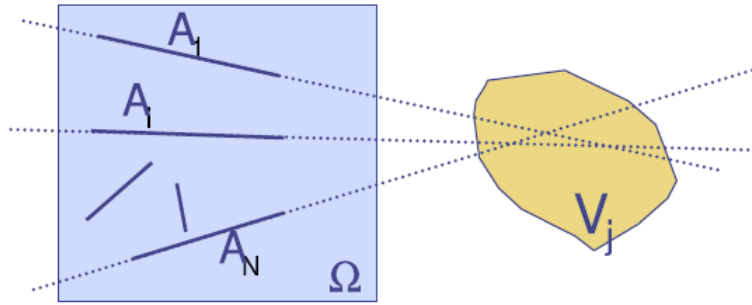
$$|\text{Angle}(dir(\mathbf{X}), direction_i)| \leq p\pi \quad (4.22)$$

όπου  $dir(\mathbf{X})$  είναι η διεύθυνση του εξεταζόμενου pixel  $\mathbf{X}$  με βάση τη σχέση 4.19 και  $direction_i$  είναι η κλίση της ευθείας στην αρχική εικόνα.

Έστω  $N_i$  το πλήθος των ευθυγραμμισμένων pixel της εικόνας με την  $i$ -οστή ευθεία, όπως υπολογίζεται με τον παραπάνω τρόπο. Στη συνέχεια, υπολογίζεται το μήκος ενός υποτιθέμενου ε-σημαντικού τμήματος με  $N_i$  πλήθος ευθυγραμμισμένων σημείων σύμφωνα με τον τύπο

$$l_{i, \epsilon-mean} = \max\{l : k(l) \leq N_i\} \quad (4.23)$$

Η απόφαση για την ε-σημαντικότητα του  $i$ -οστού τμήματος εξαρτάται από την τιμή  $l_{i, mean}$ . Συγκεκριμένα, το τμήμα θεωρείται ε-σημαντικό αν  $l_{i, mean} \geq l_{thres}$ , όπου  $l_{thres}$  μια προκαθορισμένη τιμή.



**Σχήμα 4.5:** Μία περιοχή φυγής  $V_j$  και ένα σύνολο ευθειών που διέρχονται από αυτήν. Το σχήμα αυτό προέρχεται από το [3].

### 4.3 Ανίχνευση Σημείων Φυγής

Στην ενότητα αυτή επιχειρείται η ανίχνευση πιθανών σημείων φυγής στην εικόνα. Συγκεκριμένα, επιχειρείται η εύρεση συνόλου ευθύγραμμων τμημάτων της εικόνας οι οποίες τέμνονται σε ένα σημείο  $v_\infty$ . Εξαιτίας όμως της αδυναμίας για ακριβές μετρήσεις, δεν θα υπάρξει ποτέ ένας μεγάλος αριθμός ευθειών που θα τέμνονται σε ένα σημείο. Ωστόσο, θα υπάρξει ένα σύνολο ευθειών που θα διέρχονται από ένα μικρό υποσύνολο  $\mathbf{V}$  του επιπέδου της εικόνας, που θα καλείται *περιοχή φυγής*.

Για την κάλυψη όλων των δυνατών περιπτώσεων, θα πρέπει να οριστεί ένα πεπερασμένο σύνολο τέτοιων περιοχών  $\{V_j\}_{j=1}^M$ ,  $V_i \cap V_j, \forall i \neq j$ , που θα καλύπτει το (άπειρο) επίπεδο της εικόνας, δηλαδή

$$\bigcup_{j=1}^M V_j = \mathbb{R}^2 \quad (4.24)$$

Έστω  $N$  οι ευθείες που έχουν χαρακτηριστεί ως στατιστικά σημαντικές και σημειώνονται ως  $l_1, l_2, \dots, l_N$ . Η ιδιότητα που μελετάται είναι το κατά πόσο μια ομάδα από  $k$  τέτοιες ευθείες έχουν σημείο τομής σε μία από τις περιοχές φυγής  $V_j$ . Η περιγραφή αυτή φαίνεται στο Σχήμα 4.5

### 4.4 Ε-σημαντικά Σημεία Φυγής

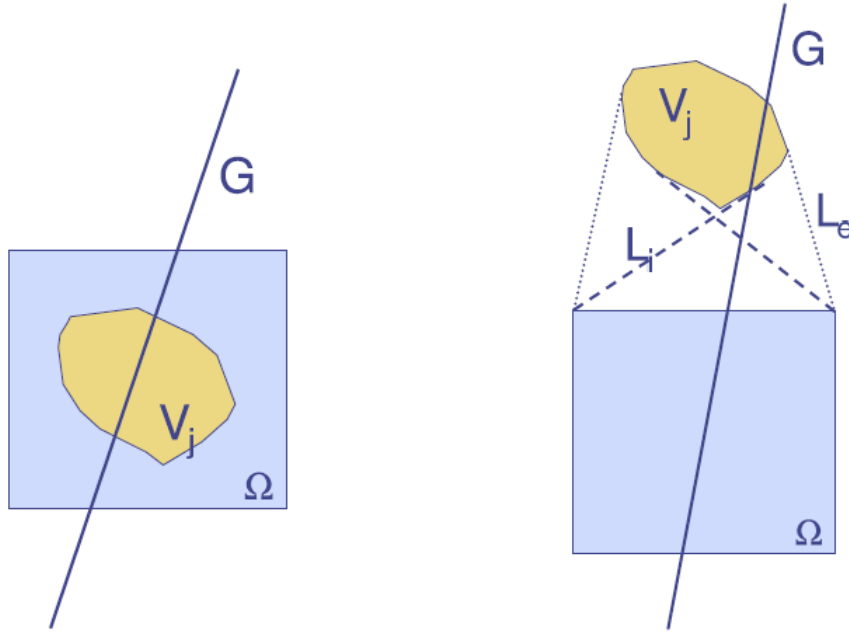
Υπό την υπόθεση ότι όλες οι γραμμές είναι ανεξάρτητες με την ίδια κατανομή, η πιθανότητα  $k$  ευθείες εκ των  $N$  να διέρχονται από την περιοχή  $V_j$  είναι  $\mathcal{B}(N, k, p_j)$  (Bernoulli κατανομή), όπου  $p_j$  είναι η πιθανότητα μια ευθεία να διέρχεται από την περιοχή  $V_j$ . Με βάση τα παραπάνω, ορίζεται η  $\epsilon$ -σημαντικότητα των περιοχών φυγής

**Ορισμός 2.4. ( $\epsilon$ -σημαντική περιοχή φυγής)** Η περιοχή φυγής  $V_j$  μέσα από την οποία διέρχονται  $k$  εκ των  $N$  ευθείες είναι  $\epsilon$ -σημαντική αν

$$\mathcal{B}(N, k, p_j) \leq \frac{\epsilon}{M} \quad (4.25)$$

όπου  $p_j$  η πιθανότητα μια ευθεία να διέρχεται μέσα από την περιοχή φυγής  $V_j$ .

Είναι εμφανές ότι η ανάλυση για την στατιστική σημαντικότητα των περιοχών φυγής είναι παρόμοια με αυτή των ευθύγραμμων τμημάτων. Η μόνη διαφορά είναι η διαφορετική ερμηνεία των



**Σχήμα 4.6:** (a) Εσωτερική περιοχή φυγής, (b) Εξωτερική περιοχή φυγής. Το σχήμα αυτό προέρχεται από το [3].

παραμέτρων. Ο συνολικός αριθμός των τμημάτων  $N$  παίζει τον ρόλο του μήκους του τμήματος, ο αριθμός των περιοχών φυγής  $M$  παίζει το ρόλο των συνολικών τμημάτων.

Η πιθανότητα  $p_j$  επηρεάζεται από τη γεωμετρία της περιοχής φυγής  $V_j$ . Η εξάρτηση αυτή μπορεί να υπολογιστεί και να οδηγήσει σε κλειστό τύπο. Παρακάτω, παρατίθενται τα αποτελέσματα. Για την απόδειξη των σχέσεων ο αναγνώστης παραπέμπεται στο [37] και στο [3].

Στην περίπτωση που η περιοχή φυγής  $V_j$  ανήκει στην περιοχή της εικόνας  $\Omega$ , δηλαδή  $V_j \subseteq \Omega$ , η πιθανότητα δίνεται από τον τύπο

$$p_j = \frac{Per(V_j)}{Per(\Omega)} \quad (4.26)$$

όπου  $Per(V_j)$  είναι η περίμετρος της περιοχής φυγής και  $Per(\Omega)$  είναι η περίμετρος της περιοχής της εικόνας.

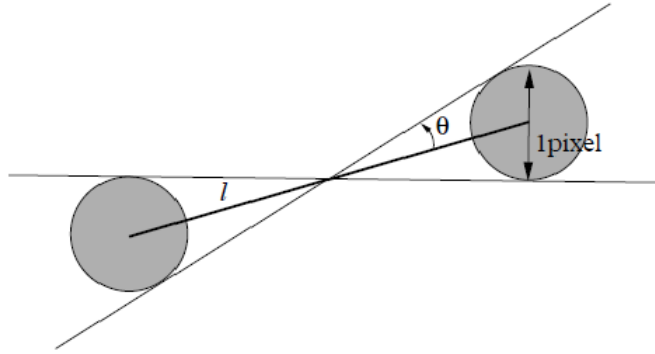
Στην περίπτωση που η περιοχή φυγής είναι εξωτερικά της περιοχής της εικόνας, δηλαδή  $V_j \cap \Omega = \emptyset$ , η πιθανότητα δίνεται από τον τύπο

$$p_j = \frac{L_i - L_e}{Per(\Omega)} \quad (4.27)$$

όπου  $L_e$  είναι η περίμετρος του κυρτού περιβλήματος των  $\Omega$  και  $V_j$  και  $L_i$  είναι το μήκος της περιοχής που αποτελείται από τις δύο περιοχές  $V_j$  και  $\Omega$  και τις εφαπτόμενες ευθείες σε αυτές.

Το Σχήμα 4.6 απεικονίζει τις δύο αυτές περιπτώσεις.

Επομένως, για να προσδιοριστεί η τιμή της πιθανότητας  $p_j$  για την περιοχή φυγής, πρέπει να προσδιοριστεί η γεωμετρία των περιοχών φυγής  $V_j$ ,  $j = 1, \dots, M$ . Ο χωρισμός σε περιοχές πρέπει να είναι τέτοιος ώστε η πιθανότητα  $p_j = \mathbb{P}[G \cap V_j \neq \emptyset]$  ένα τυχαίο ευθύγραμμο τμήμα  $G$  της εικόνας να διέρχεται από την περιοχή φυγής  $V_j$  να είναι σταθερή για όλες τις περιοχές. Χωρίς αυτή την ιδιότητα, συγκεκριμένες περιοχές φυγής θα απαιτούσαν περισσότερες συγκλίνουσες ευθείες για να γίνουν στατιστικά σημαντικές από άλλες, δηλαδή δεν θα ήταν εξίσου ανιχνεύσιμες. Από



**Σχήμα 4.7:** Κώνος αβεβαιότητας ενός ευθύγραμμου τμήματος. Η θέση των ακριανών σημείων του τμήματος μήκους  $l$  μπορεί να υπολογιστεί με σφάλμα μισού pixel. Για το λόγο αυτό, η ευθεία που αντιστοιχεί στο ευθύγραμμο τμήμα βρίσκεται σε ένα κώνο αβεβαιότητας με κέντρο το βαρύκεντρο του τμήματος και γωνία  $\theta = \arcsin(\frac{1}{l})$ . Το σχήμα αυτό προέρχεται από το [3].

την ισοπίθανη συνθήκη που επιβάλλεται, συμπεραίνει κανείς ότι το μέγεθος των περιοχών  $V_j$  αυξάνεται δραματικά με την απόσταση από την εικόνα και άρα το σφάλμα θέσης ενός σημείου φυγής αυξάνεται και αυτό με την απόσταση.

Το μέγεθος και το σχήμα των περιοχών φυγής πρέπει να είναι σε συμφωνία με την γωνιακή ακρίβεια των ευθύγραμμων τμημάτων που έχουν ανιχνευθεί. Εξαιτίας του διακριτού χαρακτήρα της ψηφιακής εικόνας, η ευθεία μπορεί να θεωρηθεί ως ένας κώνος με γωνία  $\theta = \arcsin(\frac{1}{l})$ , όπως φαίνεται στο Σχήμα 4.7. Οι περιοχές φυγής πρέπει να έχουν μέγεθος συγκριτικό με το πλάτος των αντίστοιχων κώνων φυγής.

Για τη γεωμετρική κατασκευή, ο χωρισμός σε περιοχές διακρίνεται σε δύο κατηγορίες, σε  $\{V_j^{(i)}\}$  και  $\{V_j^{(e)}\}$ . Η πρώτη αποτελείται από τις εσωτερικές περιοχές που περιλαμβάνονται ολοκληρωτικά στην εικόνα  $\Omega$ . Η δεύτερη αποτελείται από εξωτερικές περιοχές που βρίσκονται εκτός της περιοχής της εικόνας. Για ευκολία, η περιοχή  $\Omega$  προσεγγίζεται από την εγγεγραμμένο κύκλο στην (τετράγωνη) περιοχή της εικόνας με ακτίνα  $R$ . Για να ικανοποιηθεί η απαίτηση της γωνιακής ακρίβειας όλες οι εξωτερικές περιοχές  $V$  θα πρέπει να είναι τμήματα των τομέων με γωνία  $2\theta$  που βρίσκονται σε απόσταση  $d$  και  $d'$  από το κέντρο της εικόνας  $O$ . Η τμηματοποίηση στις εξωτερικές περιοχές φαίνεται στο Σχήμα 4.8. Στην περίπτωση εξωτερικών περιοχών με γωνιακή ακρίβεια  $2\theta$  σε αποστάσεις  $d$  και  $d'$ , η πιθανότητα γίνεται

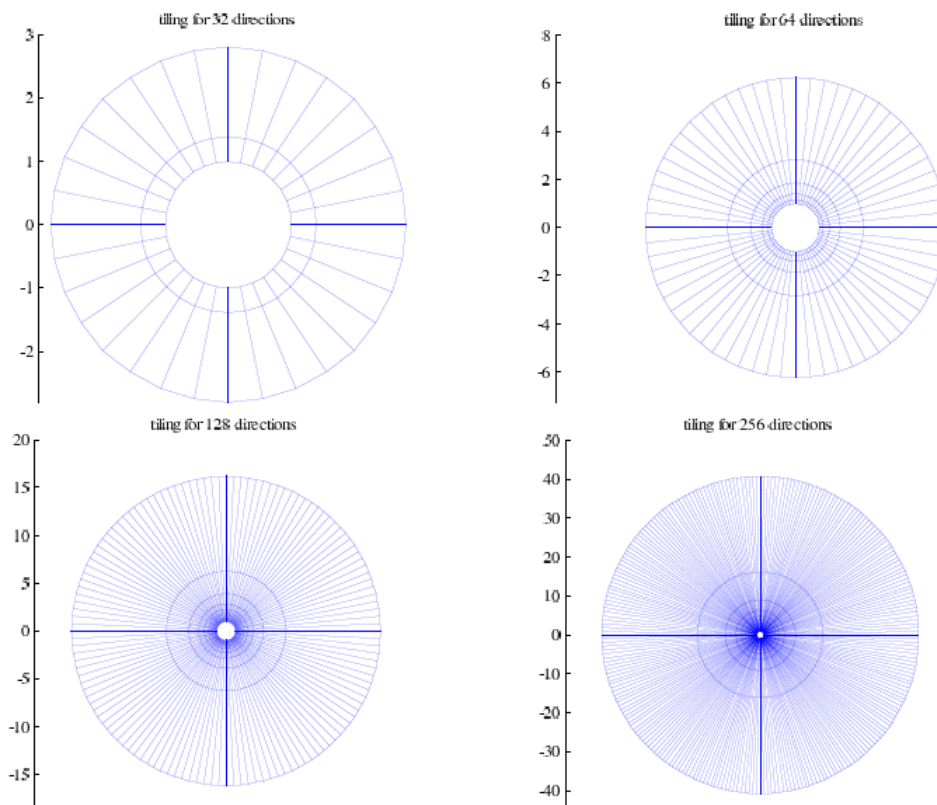
$$p_e(d, d') = \frac{L_i - L_e}{Per(\Omega)} = \frac{1}{\pi} \left( 2\theta + \left[ \beta + \frac{1}{\cos(\beta)} - \tan(\beta) \right]_{\beta=\arccos(R \cos(\beta)/d)}^{\beta=\arccos(R \cos(\beta)/d')} \right) \quad (4.28)$$

Όσον αφορά τις εσωτερικές περιοχές φυγής, ο δίσκος  $\Omega$  χωρίζεται σε τετράγωνα τμήματα. Η πλευρά του κάθε τετραγώνου επιλέγεται να είναι ίση με την πλευρά του εξωτερικού τμήματος που είναι κοντινότερο στην περιοχή της εικόνας, δηλαδή  $2R \sin(\theta)$  (προκύπτει από νόμο συνημιτόνων στο Σχήμα 4.9). Η περίμετρος των εσωτερικών περιοχών ισούται με  $8R \sin(\theta)$ , και η πιθανότητα μια ευθεία να διέρχεται από μια εσωτερική περιοχή φυγής είναι

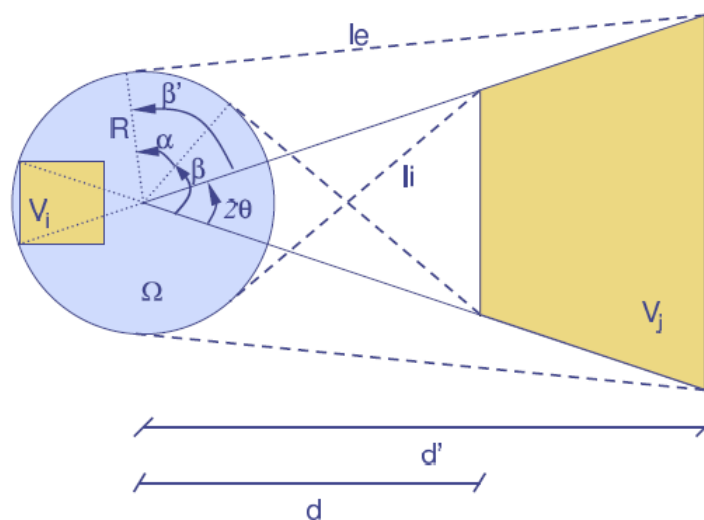
$$p_i = \frac{Per(V)}{Per(\Omega)} = \frac{4 \sin(\theta)}{\pi} \quad (4.29)$$

Με τον τρόπο αυτό, διασφαλίζεται ότι όλες οι εσωτερικές περιοχές θα έχουν την ίδια πιθανότητα και το μέγεθος τους είναι σε συμφωνία με την γωνιακή ακρίβεια  $\theta$  του ευθύγραμμου τμήματος.





**Σχήμα 4.8:** Τα τμήματα του επιπέδου της εικόνας με ίση πιθανότητα να διακρίνει τις περιοχές φυγής για διάφορες τιμές της γωνιακής ακρίβειας  $\theta$ . Μόνο οι εξωτερικές περιοχές φαίνονται. Οι εσωτερικές περιοχές σχηματίζουν ένα τετράγωνο πλέγμα. Οι άξονες αντιπροσωπεύουν αποστάσεις σχετικές με την ακτίνα της κυκλικής περιοχής της εικόνας  $\Omega$ . Το σχήμα αυτό προέρχεται από το [3].



**Σχήμα 4.9:** Κατασκευή των εξωτερικών περιοχών φυγής  $V_j$  έτσι ώστε να διακρίνονται από την ίδια πιθανότητα μια ευθεία να διέρχεται από αυτές σε σύγκριση με τις εσωτερικές περιοχές. Το σχήμα αυτό προέρχεται από το [3].

Σύμφωνα με τα παραπάνω, για τις εξωτερικές περιοχές τα μεγέθη  $d$  και  $d'$  επιλέγονται έτσι ώστε να διασφαλίσουν ότι όλες οι εξωτερικές περιοχές έχουν την ίδια πιθανότητα  $p_i = p_e$ . Ένας τρόπος για να γίνει αυτό, είναι να ξεκινήσει κανείς από την πρώτη σειρά εξωτερικών περιοχών θέτοντας  $d_1 = R$  και μετά να επιλέξει την  $d'_1$  λύνοντας την εξίσωση  $p_e(d_1, d'_1) = p_i$  ως προς  $d'_1$ . Στη συνέχεια, για την δεύτερη γραμμή από εξωτερικές περιοχές λύνεται η εξίσωση  $p_e(d_2, d'_2) = p_i$  ως προς  $d'_2$  όπου  $d_2 = d'_1$ . Η διαδικασία αυτή συνεχίζεται μέχρι  $d' \geq d_\infty$ , όπου  $d_\infty$  ικανοποιεί τη σχέση

$$\lim_{d' \rightarrow \infty} p_e(d_\infty, d') = p_i \quad (4.30)$$

Για τον υπολογισμό αυτού του ορίου, πρέπει να παρατηρηθεί ότι για  $d' \rightarrow \infty$  έχουμε  $\beta' \rightarrow \frac{\pi}{2}$  και άρα  $(\frac{1}{\cos(\beta')} - \tan(\beta')) \rightarrow 0$ , οπότε

$$\lim_{d' \rightarrow \infty} p_e(d, d') = \frac{1}{\pi} \left( 2\theta + \frac{\pi}{2} - \beta - \frac{1}{\cos(\beta)} + \tan(\beta) \right) \quad (4.31)$$

όπου  $\beta = \arccos\left(\frac{R \cos(\theta)}{d}\right)$ . Από την παραπάνω εξίσωση μπορεί εύκολα να αποδειχθεί ότι η απόσταση  $d_\infty$  είναι συνεχής και ικανοποιεί την εξίσωση

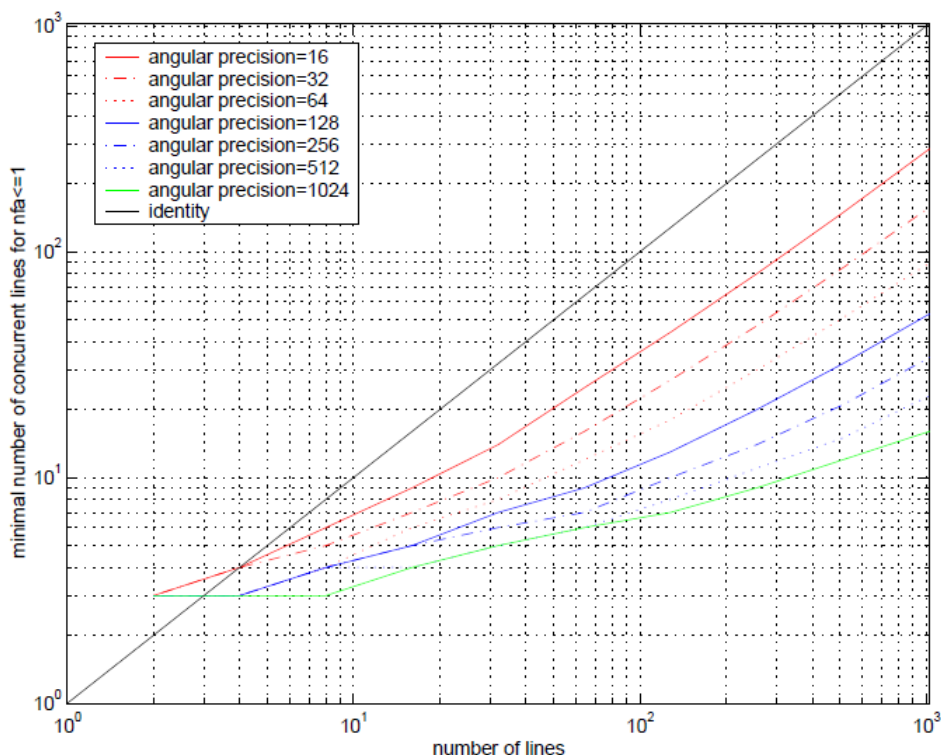
$$4 \sin(\theta) = 2\theta + \frac{\pi}{2} - \beta_\infty - \frac{1}{\cos(\beta_\infty)} + \tan(\beta_\infty) \quad (4.32)$$

όπου  $\beta_\infty = \arccos\left(\frac{R \cos(\theta)}{d_\infty}\right)$

Οι περιοχές στην τελευταία σειρά θα είναι απεριόριστες, με πιθανότητα λιγότερη από  $p_i$ . Στο Σχήμα 4.8 φαίνονται ορισμένα παραδείγματα χωρισμού του επιπέδου της εικόνας για διάφορες τιμές της γωνιακής ακρίβειας  $\theta$ .

### Ο Αλγόριθμος

1. **Καθορισμός παραμέτρων του αλγορίθμου.** Πριν από την υλοποίηση και εκτέλεση του αλγορίθμου, πρέπει να προσδιοριστούν οι παράμετροι. Συγκεκριμένα, πρέπει να οριστεί κατάλληλη τιμή για τη γωνιακή ακρίβεια  $\theta$ . Σκοπός είναι να επιτευχθεί η μεγαλύτερη δυνατή ακρίβεια και άρα το μικρότερο σφάλμα θέσης των σημείων φυγής. Από την άλλη, μεγάλη ακρίβεια είναι δυνατό να οδηγήσει σε μη ανιχνεύσιμη περιοχή φυγής. Το Σχήμα 4.10 δείχνει το ελάχιστο πλήθος των συντρεχουσών ευθειών που απαιτούνται ώστε μια περιοχή φυγής να είναι 1-σημαντική, ως συνάρτηση των συνολικών ευθειών  $N$  στην εικόνα για διάφορες τιμές της γωνιακής ακρίβειας  $\theta$ . Με βάση το Σχήμα 4.10 και θεωρώντας ότι το μέγιστο πλήθος ευθειών σε μια εικόνα είναι  $N \leq 100$  επιλέγεται τιμή της γωνιακής ακρίβειας  $\theta = \frac{\pi}{128}$ .
2. **Προσδιορισμός περιοχών φυγής.** Ο χώρος χωρίζεται με τον τρόπο που αναφέρθηκε παραπάνω. Η κυκλική περιοχή της εικόνας  $\Omega$  χωρίζεται σε τετράγωνα με πλευρά  $2R \sin(\theta)$  όπου  $R = \sqrt{\left(\frac{N}{2}\right)^2 + \left(\frac{M}{2}\right)^2}$  με  $M, N$  είναι οι πλευρές της εικόνας. Η περιοχή στο εξωτερικό της εικόνας χωρίζεται σε κύκλους ακτίνας  $d$  και  $d'$  διαδοχικά για  $d' \leq d_\infty$  με βάση τη σχέση 4.28 και εν συνεχεία σε τόξα με άνοιγμα  $2\theta$ . Με τον τρόπο αυτό δημιουργούνται οι περιοχές  $\{V_j^{(i)}\}_{j=1}^{N_i}$  και  $\{V_j^{(e)}\}_{j=1}^{N_e}$ .
3. **Προσδιορισμός ευθειών που διατρέχουν τις περιοχές φυγής** Έχοντας προσδιορίσει τις περιοχές, υπολογίζεται το πλήθος των ευθειών που διατρέχουν τις περιοχές σύμφωνα με τον αλγόριθμο 3



**Σχήμα 4.10:** Στο διάγραμμα απεικονίζεται ο ελάχιστος απαιτούμενος αριθμός συντρεχουσών ευθειών ώστε οι περιοχές φυγής να είναι 1-σημαντικές ως συνάρτηση του πλήθους των συνολικών ευθειών  $N$ , για διάφορες τιμές της παραμέτρου  $\theta$ . Το σχήμα αυτό προέρχεται από το [3].

4. **Ε-σημαντικότητα περιοχών φυγής.** Προσδιορίζεται η  $\epsilon$ -σημαντικότητα των περιοχών φυγής. Για τον λόγο αυτό, πρέπει να βρεθεί το κατώφλι του πλήθους ευθειών που πρέπει να αντιστοιχίζονται σε κάθε περιοχή ώστε η περιοχή αυτή να είναι στατιστικά σημαντικά. Για  $\epsilon = 1$  και  $p = \frac{4 \sin(\theta)}{\pi}$  υπολογίζεται το κατώφλι  $N_o$ . Επομένως, μια περιοχή  $V_j$  είναι  $\epsilon$ -σημαντική αν  $N_j \geq N_o$ , όπου  $N_j$  το πλήθος των ευθειών που διέρχονται από αυτήν. Πριν το τελικό στάδιο, πρέπει να διασφαλιστεί ότι δεν θα προκύψουν εσφαλμένα σημεία φυγής. Για το λόγο αυτό, κάθε ευθύγραμμο τμήμα πρέπει να ανήκει σε μια περιοχή φυγής, η οποία καλύτερα εξηγεί τον προσανατολισμό της. Συγκεκριμένα, ένα τμήμα αντιστοιχίζεται στην περιοχή  $V_j$  από όπου διέρχονται οι περισσότερες ευθείες σε σχέση με τις υπόλοιπες περιοχές από όπου διέρχεται αυτή. Με τον τρόπο αυτό, μια ευθεία δεν μπορεί να συμμετέχει στην ανίχνευση ενός δευτέρου λανθασμένου σημείου φυγής. Μετά τις βελτιστοποιήσεις αυτές, ο επαναυπολογισμός των  $\epsilon$ -σημαντικών περιοχών φυγής οδηγεί στο τελικό αποτέλεσμα της ανίχνευσης σημείων φυγής.

## 4.5 Τελικά Αποτελέσματα

Παρακάτω παρουσιάζονται τα αποτελέσματα του αλγορίθμου ανίχνευσης σημείων φυγής. Στο Σχήμα 4.11 φαίνεται η αρχική εικόνα της οποίας θέλουμε να ανιχνεύσουμε τα σημεία φυγής.

Στο Σχήμα 4.12 φαίνεται η εικόνα των ακμών μετά την εκτέλεση του αλγορίθμου ανίχνευσης ακμών Canny στην αρχική εικόνα.

Στο Σχήμα 4.13 φαίνονται οι ευθείες όπως έχουν ανιχνευθεί με την μέθοδο ανίχνευσης που

---

**Algorithm 3** Υπολογισμός πλήθους ευθειών που διατρέχουν τις περιοχές φυγής

---

**ΓΙΑ** (όλες τις ευθείες  $l = 1$  ως  $N_l$ )

**ΓΙΑ** (όλα τα pixels  $(i_l, j_l)$  που ανήκουν στην ευθεία  $l$ )

**ΑΝ**  $((i_l, j_l)$  ανήκει στην περιοχή  $\Omega$ )

$N_k^{(i)} = N_k^{(i)} + 1$  αν το  $(i_l, j_l)$  ανήκει στην περιοχή  $V_k^{(i)}$

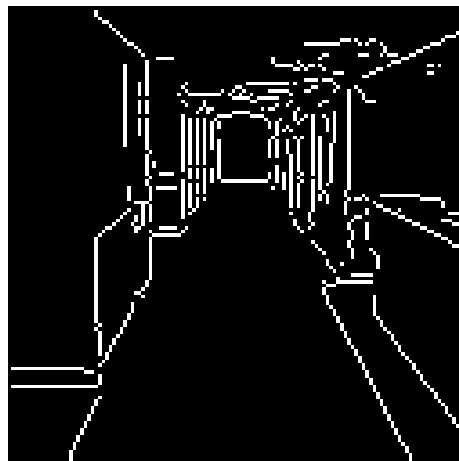
**ΑΛΛΙΩΣ**

$N_k^{(e)} = N_k^{(e)} + 1$  αν το  $(i_l, j_l)$  ανήκει στην περιοχή  $V_k^{(e)}$

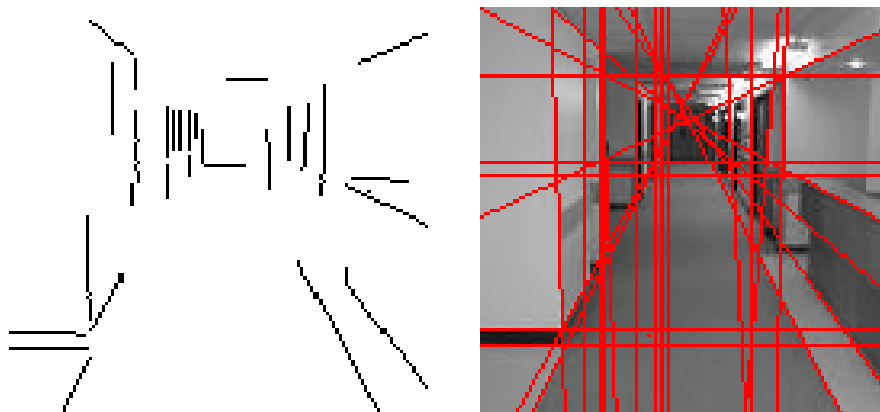
---



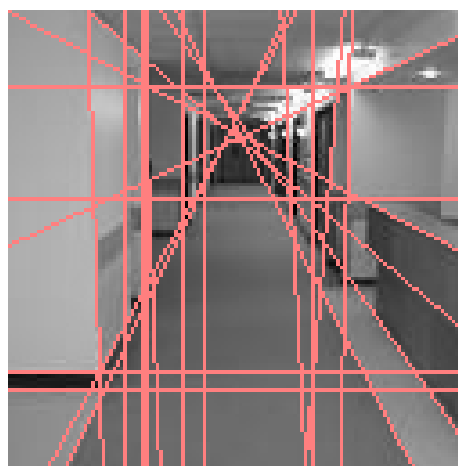
**Σχήμα 4.11:** Η αρχική εικόνα εισόδου του αλγορίθμου ανίχνευσης σημείων φυγής (πηγή: google images).



**Σχήμα 4.12:** Η εικόνα ακμών της εικόνας εισόδου.



**Σχήμα 4.13:** (a) Τα ευθύγραμμα τμήματα όπως ανιχνεύθηκαν με τη μέθοδο PCA, (b) οι ευθείες που αντιστοιχούν στα ευθύγραμμα τμήματα της αρχικής εικόνας.

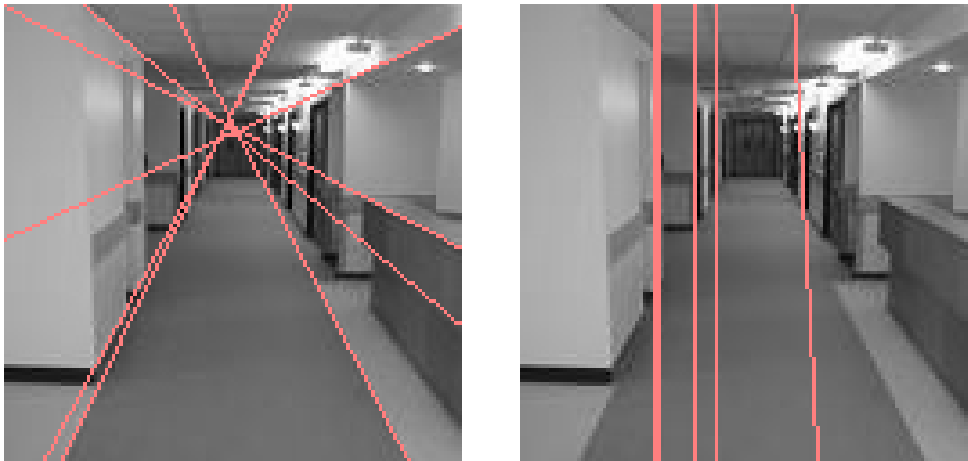


**Σχήμα 4.14:** Οι 1-σημαντικές ευθείες.

περιγράφηκε παραπάνω. Στην εικόνα (a) φαίνονται τα ευθύγραμμα τμήματα των ακμών ενώ στην εικόνα (b) φαίνονται οι αντίστοιχες ευθείες πάνω στην αρχική εικόνα

Στο Σχήμα 4.14 φαίνονται οι 1-σημαντικές ευθείες, δηλαδή οι ευθείες που είναι στατιστικά σημαντικές σύμφωνα με τα όσα αναπτύχθηκαν παραπάνω

Τέλος, στο Σχήμα 4.15 φαίνονται τα 1-σημαντικά σημεία φυγής. Παρατηρούνται δύο μόνο σημεία φυγής που αντιστοιχούν στα αντίστοιχα σύνολα παράλληλων ευθειών στον 3Δ κόσμο.



**Σχήμα 4.15:** Τα 1- σημαντικά σημεία φυγής

## Κεφάλαιο 5

# Τρισδιάστατη Ανακατασκευή Αντικειμένων

Στο κεφάλαιο αυτό επιχειρείται η 3Δ ανακατασκευή αντικειμένων. Σκοπός είναι η 3Δ αναπαράσταση συγκεκριμένων αντικειμένων από φωτογραφίες που απεικονίζουν το αντικείμενο από όσο το δυνατό περισσότερες όψεις, ώστε να επιτευχθεί μια πιο ακριβής ανακατασκευή.

Την τελευταία δεκαετία, η έρευνα στον τομέα της 3Δ αναπαράστασης από 2Δ εικόνες έχει σημειώσει σημαντική πρόοδο. Έχουν αναπτυχθεί πολλές μεθοδολογίες για την επίτευξη ενός οπτικά ικανοποιητικού αποτελέσματος. Κάποιες από τις τεχνολογίες που έχουν αναπτυχθεί περιλαμβάνουν γραφοτομές (βλ. [48]), ή οπτικά κελύφη (βλ. [36], [18]). Στο [39] γίνεται μια σύγκριση μεταξύ των πιο πρόσφατων μεθόδων ανακατασκευής. Τα αποτελέσματα των μεθοδολογιών αυτών είναι πολύ εντυπωσιακά ωστόσο απαιτούν μια σχετικά μεγάλη βάση δεδομένων. Στην παρούσα εφαρμογή στόχος είναι να γίνει μια 3Δ ανακατασκευή από λιγότερες εικόνες από αυτές που απαιτούν οι παραπάνω μέθοδοι.

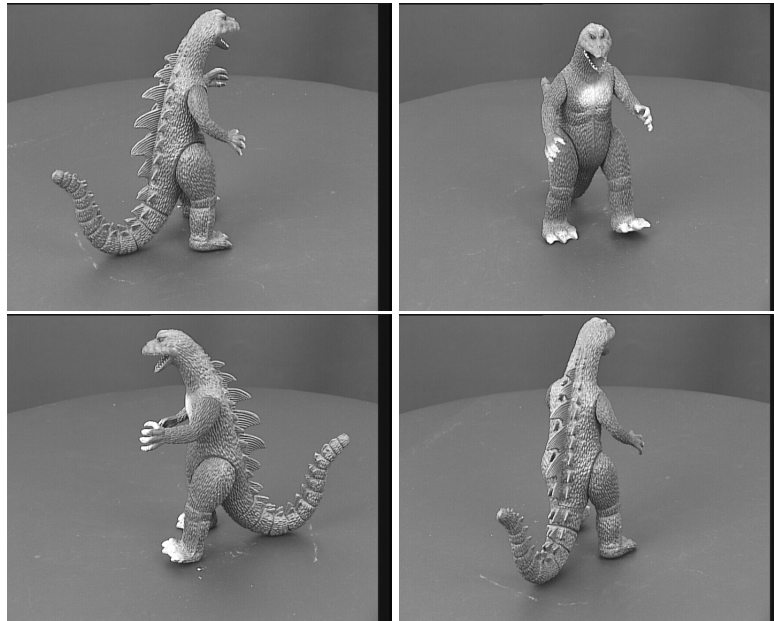
Τα δεδομένα που απαιτούνται για την 3Δ αναπαράσταση ενός αντικειμένου είναι εικόνες, οι οποίες απεικονίζουν το αντικείμενο από διάφορες θέσεις και συγκεκριμένο προσανατολισμό. Στη συγκεκριμένη εφαρμογή, η λήψη των εικόνων έγινε με γνωστές τις θέσεις και τον προσανατολισμό των καμερών. Με άλλα λόγια, ήταν γνωστή η θέση του κέντρου του συστήματος συντεταγμένων της  $i$ -στης κάμερας  $\tilde{C}_i$  και ο προσανατολισμός του εν λόγω συστήματος  $R_i$  ως προς το παγκόσμιο σύστημα συντεταγμένων, όπως αυτό ορίστηκε από τους ερευνητές. Επιπλέον, από το EXIF tag των φωτογραφιών είναι δυνατό να οριστούν οι εσωτερικές παράμετροι των καμερών, δηλαδή είναι γνωστός και ο πίνακας βαθμονόμησης  $K_i$ . Επομένως, για την  $i$ -οστή κάμερα εύκολα μπορεί να υπολογιστεί ο πίνακας της κάμερας

$$P_i = K_i R_i \left[ I \mid -\tilde{C}_i \right] \quad (5.1)$$

Επιπλέον, για να είναι δυνατή η ανακατασκευή από τις εικόνες πρέπει μεταξύ διαδοχικών φωτογραφιών να παρουσιάζεται επικάλυψη. Δηλαδή, πρέπει να παρατηρείται πλήθος κοινών σημείων μεταξύ διαδοχικών εικόνων του αντικειμένου από το σύνολο των δεδομένων.

Στη συγκεκριμένη εφαρμογή, η βάση αποτελείται από 36 εικόνες οι οποίες απεικονίζουν το αντικείμενο από διαφορετικές θέσεις. Για την παρούσα εφαρμογή χρησιμοποιήθηκαν οι μισές από τις εικόνες αυτές. Στο Σχήμα 5.1 φαίνονται κάποιες από τις εικόνες που απαρτίζουν το σύνολο των δεδομένων.

Στις παρακάτω ενότητες περιγράφεται ο αλγόριθμος για την 3Δ ανακατασκευή και παρουσιάζονται τα τελικά αποτελέσματα αυτής.



**Σχήμα 5.1:** Από το σύνολο δεδομένων για την 3Δ ανακατασκευή (πηγή: [www.robots.ox.ac.uk/vgg/data/data-mview.html](http://www.robots.ox.ac.uk/vgg/data/data-mview.html)).

## 5.1 Εύρεση Σημείων Ταιριάσματος

### 5.1.1 SIFT Σημεία Ενδιαφέροντος

Στο πρώτο στάδιο του αλγορίθμου ανακατασκευής, σκοπός είναι να βρεθούν ζεύγη κοινών σημείων μεταξύ διαδοχικών εικόνων από το σύνολο των δεδομένων. Ο αλγόριθμος για την ανίχνευση των σημείων αυτών μεταξύ δύο 2Δ εικόνων αποτελείται από τρία στάδια, την ανίχνευση σημείων ενδιαφέροντος ανεξάρτητα από αφινικούς μετασχηματισμούς σε κάθε εικόνα ξεχωριστά, την περιγραφή αυτών των σημείων με αποτελεσματικό τρόπο ώστε να μην εξαρτάται από παράγοντες κλιμάκωσης ή φωτεινότητας γύρω από το εκάστοτε σημείο και την ανίχνευση σημείων ταιριάσματος μεταξύ των σημείων ενδιαφέροντος στις δύο εικόνες. Η μέθοδος που ακολουθείται για την ανίχνευση σημείων ενδιαφέροντος αμετάβλητων από αφινικούς μετασχηματισμούς περιγράφεται αναλυτικά στο [27]. Τα σημεία που δίνει είναι γνωστά στη βιβλιογραφία και ως SIFT σημεία.

Έχουν προταθεί και άλλοι αλγόριθμοι ανίχνευσης σημείων ενδιαφέροντος. Στο [33] συγκρίνονται και αξιολογούνται οι διάφορες μέθοδοι.

#### 5.1.1.1 Ανίχνευση Σημείων Ενδιαφέροντος

Ένας απλός τρόπος ανίχνευσης σημείων ενδιαφέροντος προτάθηκε αρχικά από τον C. Harris (βλ. [21]). Τα σημεία που εντοπίζει ο αλγόριθμος αυτός αποτελούν συνδυασμό σημείων ακμής και γωνιών. Η μέθοδος του D. Lowe ακολουθεί την παρακάτω διαδικασία.

Έστω  $I(x, y)$  μια δισδιάστατη εικόνα. Για την ανίχνευση σημείων, η εικόνα θα διέλθει από μια σειρά φίλτρων αποτελεσματικών για την εύρεση πιθανών θέσεων που στη συνέχεια μελετώνται περαιτέρω. Για να ανιχνευθούν θέσεις που είναι αμετάβλητες σε αλλαγές κλίμακας στην εικόνα μπορεί να επιτευχθεί με την αναζήτηση ευσταθών χαρακτηριστικών σε όλες τις δυνατές κλίμακες.

Για την σάρωση στο χώρο κλίμακας, η εικόνα διέρχεται από ένα γκαουσιανό φίλτρο με αποτέλεσμα την εικόνα  $L(x, y, \sigma)$  που προέρχεται από τη συνέλιξη του φίλτρου  $G(x, y, \sigma)$  με την



εικόνας εισόδου  $I(x, y)$ :

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \quad (5.2)$$

όπου  $*$  είναι ο 2Δ τελεστής της συνέλιξης και

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \quad (5.3)$$

Για να ανιχνευθούν αποτελεσματικά ευσταθείς χαρακτηριστικές θέσεις στο χώρο κλίμακας, ο D. Lowe προτείνει την χρησιμοποίηση της συνάρτησης διαφοράς γκαουσιανών  $D(x, y, \sigma)$  η οποία προκύπτει από την παρακάτω σχέση

$$\begin{aligned} D(x, y, \sigma) &= \left( G(x, y, k\sigma) - G(x, y, \sigma) \right) * I(x, y) \\ &= L(x, y, k\sigma) - L(x, y, \sigma) \end{aligned} \quad (5.4)$$

Η συνάρτηση της διαφοράς γκαουσιανών ισούται προσεγγιστικά με την κανονικοποιημένη ως προς την κλίμακα Laplacian of Gaussian,  $\sigma^2 \nabla^2 G$ . Σύμφωνα με τον [32], τα ακρότατα της  $\sigma^2 \nabla^2 G$  δίνουν τα πιο ευσταθή χαρακτηριστικά της εικόνας σε σχέση με άλλες συναρτήσεις.

Η σχέση μεταξύ των  $D$  και  $\sigma^2 \nabla^2 G$  μπορεί να γίνει κατανοητή από την εξίσωση διάχυσης

$$\frac{\partial G}{\partial \sigma} = \sigma \nabla^2 G$$

Προσεγγίζοντας τη μερική παράγωγο με κανονικοποιημένη διαφορά για κοντινές κλίμακες  $k\sigma$  και  $\sigma$

$$\sigma \nabla^2 G = \frac{\partial G}{\partial \sigma} \approx \frac{G(x, y, k\sigma) - G(x, y, \sigma)}{k\sigma - \sigma}$$

και επομένως

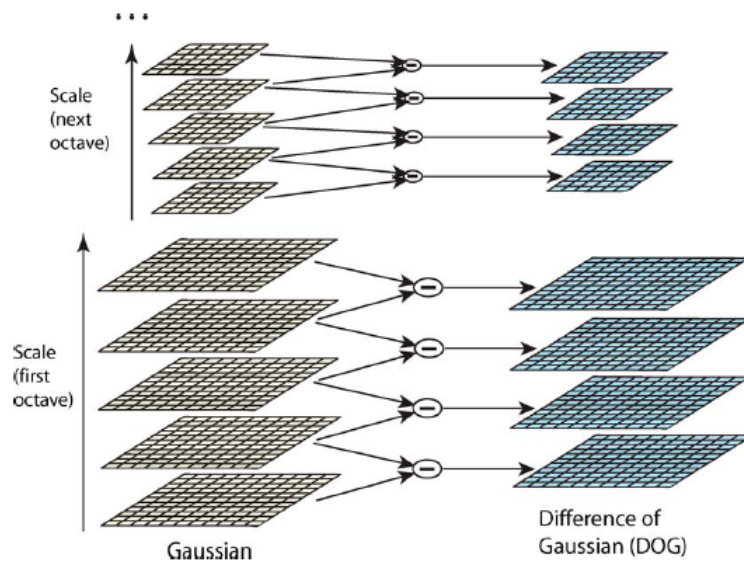
$$G(x, y, k\sigma) - G(x, y, \sigma) \approx (k - 1)\sigma^2 \nabla^2 G$$

Η σχέση αυτή δείχνει ότι όταν η διαφορά γκαουσιανών έχει κλίμακες που διαφέρουν κατά έναν σταθερό παράγοντα περιλαμβάνει την κανονικοποίηση στην κλίμακα  $\sigma^2$  που απαιτείται για την λαπλασιανή. Ο παράγοντας  $(k - 1)$  είναι σταθερός σε όλες τις κλίμακες και έτσι δεν επηρεάζει την εύρεση ακροτάτων. Το λάθος προσέγγισης τείνει στο μηδέν όταν το  $k$  τείνει στο 1 άλλα στην πραγματικότητα ο D.Lowe έχει βρει ότι η προσέγγιση δεν έχει καμία επίδραση στην ευστάθεια των ακροτάτων ακόμα και για σχετικά μεγάλα  $k$ .

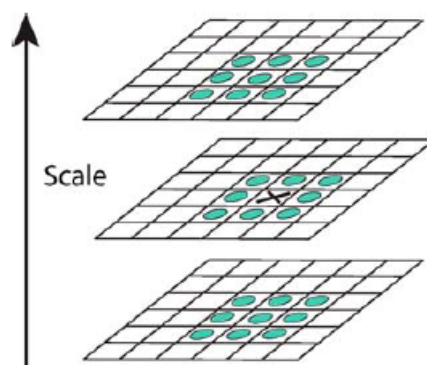
Στο Σχήμα 5.2 φαίνεται ένας τρόπος κατασκευής των  $D(x, y, \sigma)$ . Η αρχική εικόνα συνελίσσεται διαδοχικά με γκαουσιανές για να παράγει εικόνες οι οποίες διαχωρίζονται από ένα σταθερό παράγοντα  $k$  στο χώρο κλίμακας. Κάθε οκτάδα στο χώρο κλίμακας επιλέγεται να διαχωριστεί κατά ένα ακέραιο πλήθος διαστημάτων  $s$ , δηλαδή  $k = 2^{1/s}$  ώστε τελικά να καταλήξει σε διπλάσια κλίμακα από την αρχική. Για κάθε οκτάδα, παράγονται  $s + 3$  εικόνες. Διαδοχικές εικόνες αφαιρούνται μεταξύ τους για να παράγουν τις διαφορές των γκαουσιανών. Αφού έχει επεξεργαστεί μια οκτάδα, επαναδειγματοληπτούμε τις γκαουσιανές εικόνες που έχουν την διπλάσια τιμή κλίμακας από την αρχική αφαιρώντας ένα pixel κάθε δύο pixels σε κάθε γραμμή και στήλη. Η εικόνα αυτή αποτελεί αρχική εικόνα για την επόμενη οκτάδα, ώστε να μειωθεί η υπολογιστική ανάγκη.

Για να ανιχνευθούν τα τοπικά ακρότατα των  $D(x, y, \sigma)$ , κάθε σημείο της εικόνας συγκρίνεται με τα 8 γειτονικά σημεία της ίδιας εικόνας καθώς και τα 9 γειτονικά σημεία στην πιο πάνω και πιο κάτω κλίμακα από την τρέχουσα  $\sigma$  (βλ. Σχήμα 5.3). Επιλέγεται ως ακρότατο μόνο αν έχει τιμή μεγαλύτερη ή μικρότερη από όλους τους γείτονες.

Μετά την ανίχνευση ενός υποψήφιου σημείου ενδιαφέροντος, το οποίο έχει προκύψει από τις συγκρίσεις με τα γειτονικά pixels, το επόμενο βήμα είναι να γίνει μια λεπτομερής εξέταση στα διπλανά δεδομένα για τη θέση, την κλίμακα και τον λόγο καμπυλότητας. Αυτή η πληροφορία



**Σχήμα 5.2:** Τρόπος κατασκευής των συναρτήσεων διαφοράς γκαουσιανών. Οι εικόνες συνελίσσονται με γκαουσιανά φίλτρα για να παράγουν ένα σύνολο από εικόνες στο χώρο κλίμακας (αριστερά). Οι διαδοχικές γκαουσιανές εικόνες αφαιρούνται για να παράγουν τις διαφορές γκαουσιανών  $D(x, y, \sigma)$  (δεξιά). Μετά από κάθε οκτάδα, οι γκαουσιανές εικόνες στην κορυφή της υποδειγματοληπτούνται κατά ένα παράγοντα 2, και η διαδικασία επαναλαμβάνεται. Το σχήμα προέρχεται από το [27].



**Σχήμα 5.3:** Τα τοπικά ακρότατα των συναρτήσεων  $D(x, y, \sigma)$  υπολογίζονται συγκρίνοντας την τιμή κάθε σημείου (pixel  $X$ ) με τα γειτονικά σημεία της  $4 \times 3$  γειτονιάς στην τρέχουσα καθώς και στην αμέσως μικρότερη και μεγαλύτερη κλίμακα. Το σχήμα προέρχεται από το [27].

επιτρέπει την απόρριψη σημείων που δεν διαφέρουν σημαντικά από τα γειτονικά τους και άρα είναι ευαίσθητα στο θόρυβο. Η μεθοδολογία αυτή βασίζεται στο [27]. Η απομάκρυνση των σημείων αυτών δεν είναι αρκετή ώστε να εξασφαλιστεί ευστάθεια για τα υπόλοιπα σημεία. Η συνάρτηση της διαφοράς γκαουσιανών θα λαμβάνει μεγάλη τιμή κατά μήκος ακμών, ακόμα και αν η θέση του σημείου δεν είναι με ακρίβεια πάνω στην ακμή και άρα είναι ασταθής σε σχετικά μικρή επίδραση του θορύβου. Για το λόγο αυτό, πρέπει να εξαιρεθούν τα σημεία που δεν είναι ορθά ορισμένα πάνω στις ακμές.

Για να προκύψουν σημεία ενδιαφέροντος αμετάβλητα όχι μόνο ως προς την κλίμακα αλλά και ως προς την περιστροφή, σε κάθε σημείο αποδίδεται μια κατεύθυνση, η οποία είναι και η κυρίαρχη κατεύθυνση όπως προκύπτει με μετρήσεις σε μια γειτονιά του σημείου ενδιαφέροντος. Αν δεν υπάρχει εμφανής κυρίαρχη κατεύθυνση, τότε είναι δυνατό σε ένα σημείο ενδιαφέροντος να αποδοθούν περισσότερες από μια κατευθύνσεις.

### 5.1.1.2 Τοπικοί Περιγραφείς

Στο προηγούμενο στάδιο, σε κάθε σημείο ενδιαφέροντος αποδόθηκε μία θέση στην εικόνα και μία κλίμακα. Το επόμενο βήμα είναι η απόδοση στο σημείο προσανατολισμό και η εύρεση ενός τρόπου περιγραφής για την τοπική περιοχή της εικόνας γύρω από τα σημεία ενδιαφέροντος ώστε να είναι εύκολα διακριτός και αμετάβλητος σε μεταβολές, όπως η περιστροφή της εικόνας και η φωτεινότητα.

Το Σχήμα 5.4 δείχνει τον τρόπο υπολογισμού των περιγραφών των σημείων ενδιαφέροντος. Αρχικά, υπολογίζονται το μέτρο και η γωνία της κλίσης σύμφωνα με τους τύπους 5.5 σε μια περιοχή γύρω από τη θέση του σημείου ενδιαφέροντος, θολώνοντας την εικόνα με ένα γκαουσιανό φίλτρο κλίμακας ίση με την κλίμακα του σημείου ενδιαφέροντος.

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2} \quad (5.5)$$

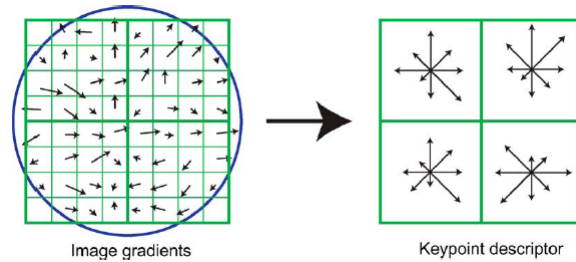
$$\theta(x, y) = \arctan\left(\frac{L(x, y+1) - L(x, y-1)}{L(x+1, y) - L(x-1, y)}\right) \quad (5.6)$$

Μια γκαουσιανή συνάρτηση με  $\sigma$  ίσο με το μισό του πλάτους του παραθύρου του περιγραφέα και κέντρο το σημείο ενδιαφέροντος χρησιμοποιείται για την ανάθεση βαρών στο μέτρο του κάθε σημείου της περιοχής. Η διαδικασία αυτή φαίνεται με το κυκλικό παράθυρο του αριστερού σχήματος στο Σχήμα 5.4 Η χρήση του παραθύρου εξασφαλίζει ομαλές μεταβάσεις στον περιγραφέα με μικρές μεταβάσεις στην θέση του παραθύρου και δίνει μικρότερη έμφαση σε απομακρυσμένα από το κέντρο σημεία.

Ο περιγραφές φαίνεται στο δεξί μέρος του Σχήματος 5.4. Με την δημιουργία ιστογραμμάτων κατεύθυνσης σε γειτονίες (εδώ  $4 \times 4$ ) επιτρέπεται η σημαντική ολίσθηση στις θέσεις της κλίσης. Το σχήμα δείχνει 8 διευθύνσεις (δηλαδή πρόκειται για 8 bin ιστογράμματα), με το μήκος του κάθε διανύσματος να αντιπροσωπεύει το μέγεθος της αντίστοιχης θέσης στο ιστογράμματα.

Ο περιγραφέας είναι ένα διάνυσμα που περιέχει τις τιμές όλων των θέσεων του ιστογράμματος κατεύθυνσης, που αντιστοιχούν στα μήκη των διανυσμάτων στα δεξιά του Σχήματος 5.4 Το Σχήμα 5.4 δείχνει ένα  $2 \times 2$  πίνακα από ιστογράμματα κατεύθυνσης. Στην εφαρμογή, χρησιμοποιήθηκε ένας  $4 \times 4$  πίνακας από ιστογράμματα 8 θέσεων το καθένα. Επομένως, το διάνυσμα του περιγραφέα είναι διάστασης  $4 \times 4 \times 8 = 128$  για κάθε σημείο ενδιαφέροντος.

Για να εξασφαλιστεί η ανεξαρτησία από τις αλλαγές φωτεινότητας, το διάνυσμα κανονικοποιείται ώστε να έχει μοναδιαίο μέτρο. Με τον τρόπο αυτό, ο περιγραφέας παραμένει αμετάβλητος σε αφινικές αλλαγές της φωτεινότητας.



Σχήμα 5.4: Ο τοπικός SIFT περιγραφέας των σημείων ενδιαφέροντος. Το σχήμα προέρχεται από το [27].

### 5.1.1.3 Ανίχνευση Σημείων Ταιριάσματος

Ο πιο εύκολος τρόπος αντιστοίχισης σε ένα σημείο ενδιαφέροντος στην εικόνα  $I$  με ένα άλλο σημείο ταιριάσματος στην εικόνα  $J$  είναι η εύρεση του κοντινότερου γείτονα στη βάση δεδομένων των σημείων ενδιαφέροντος στη δεύτερη εικόνα. Ο κοντινότερος γείτονας ορίζεται ως το σημείο ενδιαφέροντος με την ελάχιστη Ευκλείδεια απόσταση από τον περιγραφέα του αρχικού σημείου.

Ωστόσο, πολλά χαρακτηριστικά σε μια εικόνα μπορεί να μην έχουν σωστό ταίρι σε κάποια άλλη εικόνα, επειδή δεν έχουν βρεθεί σε καμία άλλη εικόνα. Για το λόγο αυτό, θα ήταν χρήσιμο να βρεθεί ένας τρόπος να εξαλειφθούν τα χαρακτηριστικά που δεν έχουν καλά ταίρια στη βάση δεδομένων. Ένας ολικός τελεστής κατωφλίου στην απόσταση στο κοντινότερο χαρακτηριστικό δεν θα λειτουργεί αποτελεσματικά, καθώς μερικοί περιγραφείς είναι πιο ευδιάκριτοι από άλλους. Ένας πιο αποτελεσματικός τρόπος, είναι να συγκριθεί η απόσταση με τον κοντινότερο γείτονα σε σχέση με αυτήν με τον δεύτερο κοντινότερο γείτονα. Το μέτρο αυτό δουλεύει αποτελεσματικά, διότι σωστά ταιριάσματα πρέπει να είναι σημαντικά πιο κοντά από το κοντινότερο λανθασμένο ταίριασμα. Υπάρχει περίπτωση να υπάρχουν εσφαλμένα ταιριάσματα εξαιτίας της μεγάλης διάστασης του διανύσματος του περιγραφέα. Στην εφαρμογή επιλέγεται λόγος αποστάσεων ίσος με 0.6.

Στο Σχήμα 5.14 φαίνονται τα σημεία ταιριάσματος για δύο διαδοχικές εικόνες  $I_{11}, I_{12}$  της βάσης δεδομένων. Οι γραμμές αντιστοιχίζουν ζεύξη κοινών σημείων στις δύο εικόνες με βάση τα παραπάνω. Τα σημεία αυτά βρίσκονται στην άκρη των ευθύγραμμων τμημάτων αντιστοίχισης.

### 5.1.2 MSER Περιοχές Ενδιαφέροντος

Ένας άλλος πολύ διαδεδομένος τρόπος εξαγωγής χαρακτηριστικών από εικόνες είναι ο αλγόριθμος υπολογισμού ακραίων περιοχών μέγιστης ευστάθειας MSER (βλ. [29]). Η μέθοδος αυτή χρησιμοποιείται στην ΟΥ για ανίχνευση κηλίδων (blobs).

Ο αλγόριθμος MSER ανιχνεύει τις περιοχές μιας εικόνας που μπορούν να ανιχνευθούν επαναληπτικά καθώς κατέχουν διακριτικές και σταθερές ιδιότητες. Το σύνολο των περιοχών που ανιχνεύονται είναι κλειστό ως προς συνεχείς μετασχηματισμούς στις συντεταγμένες των σημείων της εικόνας και είναι κλειστό ως προς μονοτονικούς μετασχηματισμούς της έντασης της εικόνας. Οι MSER περιοχές διακρίνονται μονάχα από την ιδιότητα της συνάρτησης της έντασης και από το εξωτερικό τους όριο.

Ο αλγόριθμος μπορεί να περιγραφεί άτυπα παρακάτω. Έστω όλες οι δυνατές κατωφλιοποιήσεις μιας γκριζας εικόνας  $I$  και έστω τα pixels που βρίσκονται υπό του κατωφλίου να λέγονται μαύρα και αυτά άνω του κατωφλίου άσπρα. Μια συνεχής επίδειξη των κατωφλιοποιημένων εικόνων  $I_t$  όπου  $t$  το κατώφλι, τότε ένας παρατηρητής θα έβλεπε αρχικά μια άσπρη εικόνα. Στη συνέχεια, μαύρες περιοχές κάνουν την εμφάνισή τους οι οποίες ολοένα και αυξάνονται σε έκταση μέχρι να ενωθούν και να προκύψει τελικά μια μαύρη εικόνα. Το σύνολο όλων των συνεκτικών

συνιστώσων από όλες τις κατωφλιοποιημένες εικόνες αποτελεί το σύνολο των *μεγίστων περιοχών*. Οι *ελάχιστες περιοχές* προκύπτουν με την ίδια διαδικασία αφού πρώτα αντιστραφεί η ένταση της εικόνας. Το σύνολο των ακραίων περιοχών αποτελείται από το σύνολο των μεγίστων και των ελαχίστων περιοχών.

Αναλυτικά, τα pixels ταξινομούνται με βάση την έντασή τους (σε αύξουσα ή φθίνουσα σειρά). Αφού τα ταξινομημένα pixels τοποθετηθούν στην εικόνα, υπολογίζονται και αποθηκεύονται οι συνεκτικές συνιστώσες. Η συγχώνευση δύο συνεκτικών συνιστώσων σηματοδοτεί τον τερματισμό ύπαρξης της μικρότερης σε μέγεθος συνιστώσας και τα pixels της μικρότερης συνιστώσας συμπληρώνουν αυτά της μεγαλύτερης. Οι MSER περιοχές περιγράφονται από τη θέση του τοπικού ελαχίστου (ή μεγίστου) της έντασης της περιοχής και ένα κατώφλι που αποτελεί το τοπικό ελάχιστο του ρυθμού μεταβολής της συνάρτησης της περιοχής. Με τον τρόπο αυτό οι MSER περιοχές είναι ευσταθείς για ένα μεγάλο πεδίο τιμών κατωφλίου. Παρακάτω φαίνονται οι ορισμοί που εισάγονται στο [29] για τις MSER περιοχές

- **Η εικόνα**  $I$  είναι μια αντιστοίχιση  $I : D \subset \mathbb{Z}^2 \rightarrow S$ . Οι ακραίες περιοχές είναι καλά ορισμένες αν
  1. Το σύνολο  $S$  είναι ταξινομημένο.
  2. Έχει οριστεί μια σχέση γειτνίασης  $A \subset D \times D$ . Για 4-συνεκτικότητα ισχύει για  $p, q \in D$  ότι  $(pAq)$  αν και μόνο αν  $\sum_{i=1}^d |p_i - q_i| \leq 1$ .
- **Η περιοχή**  $Q$  είναι ένα συνεχές υποσύνολο του  $D$ , δηλαδή αν  $p, q \in Q$  τότε υπάρχει μια ακολουθία  $p, a_1, a_2, \dots, a_n, q$  έτσι ώστε  $(pAa_1), (a_1Aa_2), \dots, (a_nAq)$ .
- **Εξωτερικό όριο περιοχής**  $\partial Q = \{q \in D \setminus Q : \exists p \in Q : qAp\}$ , δηλαδή το όριο  $\partial Q$  της περιοχής  $Q$  είναι ένα σύνολο από pixels τα οποία γειτνιάζουν με ένα τουλάχιστον pixel της περιοχής  $Q$  αλλά δεν ανήκουν σε αυτήν.
- **Ακραία περιοχή**  $Q \subset D$  είναι μια περιοχή τέτοια ώστε για κάθε  $p \in Q, q \in \partial Q: I(p) > I(q)$  (περιοχή μέγιστης έντασης) ή  $I(p) < I(q)$  (περιοχή ελάχιστης έντασης).
- **Ακραία περιοχή μέγιστης ευστάθειας MSER** Αν  $Q_1, \dots, Q_{i-1}, Q_i, \dots$  είναι μια ακολουθία από ένθετες ακραίες περιοχές, δηλαδή  $Q_i \subset Q_{i+1}$ , τότε η ακραία περιοχή  $Q_{i^*}$  είναι μέγιστης ευστάθειας αν και μόνο αν  $q(i) = |Q_{i+\Delta} \setminus Q_{i-\Delta}| / |Q_i|$  έχει τοπικό ελάχιστο στο  $i^*$  ( $|\cdot|$  εκφράζει τον πληθικό αριθμό και  $\Delta \in S$  είναι μια παράμετρος της μεθόδου).

Στο Σχήμα 5.5 φαίνεται το αποτέλεσμα του αλγορίθμου σε μια εικόνα. Οι MSER περιοχές αντικαθίστανται από ελλείψεις με σχηματικές ροπές ίδιες με αυτές του σχήματος των περιοχών μέχρι και δευτέρας τάξης.

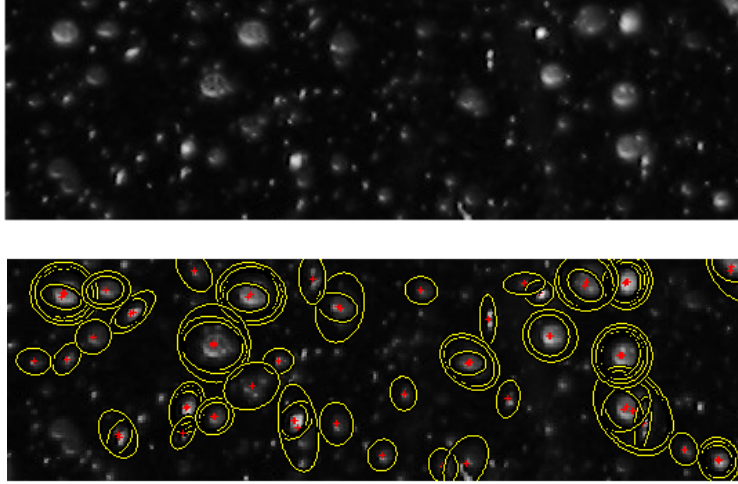
### 5.1.3 Μορφολογικές Περιοχές Ενδιαφέροντος

Ένας άλλος τρόπος ανίχνευσης χαρακτηριστικών περιοχών σε μια εικόνα είναι με τη χρήση μορφολογικών φίλτρων. Οι μορφολογικοί τελεστές στις εικόνες ορίζονται στο [28].

Ο μορφολογικός τελεστής του opening στην εικόνα  $f$  με δομικό στοιχείο  $B$  δίνεται από τον τύπο

$$f \circ B \equiv (f \ominus B) \oplus B \quad (5.7)$$

όπου οι τελεστές  $\ominus, \oplus$  αναφέρονται στους μετασχηματισμούς *erosion* και *dilation* αντίστοιχα. Με άλλα λόγια, η εικόνα υπόκειται σε *erosion* και εν συνεχεία σε *dilation*.



**Σχήμα 5.5:** Οι MSER περιοχές της εικόνας εισόδου που απεικονίζει νυφάδες χιονιού. Οι ελλείψεις έχουν αντικαταστήσει τις περιοχές των συνεκτικών συνιστώσων, παρουσιάζοντας ίδιες σχηματικές ροπές ως και δευτέρας τάξης.

Αφαιρώντας από την εικόνα  $f$  το αποτέλεσμα της 5.7 για ένα συμπαγές και κυρτό σύνολο  $B$  δίνει ως αποτέλεσμα την εικόνα κορυφών. Η υλοποίηση αυτή αποτελεί τον *Top hat μετασχηματισμό* και δίνεται από στη σχέση 5.8.

$$peak(f) = f - (f \circ B) \quad (5.8)$$

Η έξοδος  $peak(f)$  παίρνει μη αρνητικές τιμές και διασφαλίζει ότι περιέχει μόνο κορυφές. Η τιμή του  $peak(f)$  σε κάποια θέση του pixel ορίζει την ένταση της κορυφής στη θέση αυτή. Το σύνολο  $B$  είναι ένας σημαντικός παράγοντας του top hat τελεστή, καθώς το σχήμα και το μέγεθος των κορυφών εξαρτάται από το μέγεθος και το σχήμα του  $B$ .

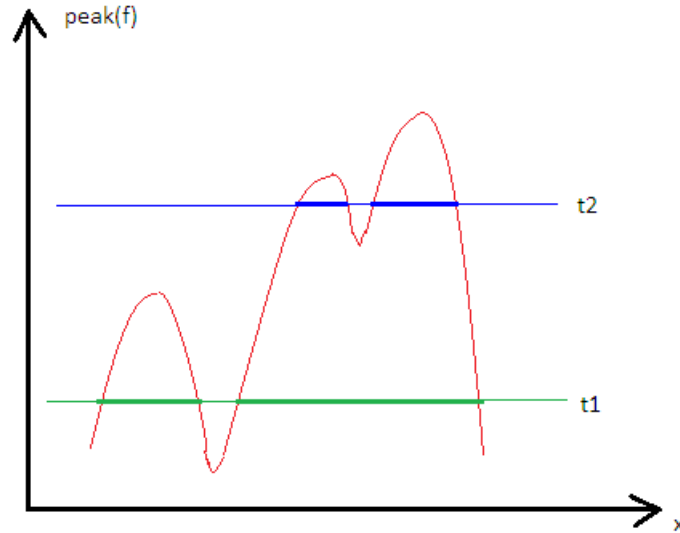
Η παραγωγή δυαδικών περιοχών κορυφών γίνεται με την κατωφλιοποίηση της εικόνας  $peak(f)$ . Επειδή είναι δύσκολη η εύρεση της ιδανικής τιμής κατωφλίου, η εικόνα  $peak(f)$  κατωφλιοποιείται για διάφορες τιμές κατωφλίου  $t$ , παράγοντας την εικόνα  $peak(f)_t$  έτσι ώστε

$$peak(f)_t(x, y) = \begin{cases} 1, & \text{αν } peak(f)(x, y) \geq t \\ 0, & \text{αν αλλιώς} \end{cases} \quad (5.9)$$

Ο λόγος για τον οποίο πραγματοποιείται πολλαπλή κατωφλιοποίηση της εικόνας των κορυφών φαίνεται στο Σχήμα 5.6. Στο Σχήμα 5.6 φαίνεται μια μονοδιάστατη συνάρτηση κορυφών  $peak(f)$  της συνάρτησης  $f(x)$ . Η κατωφλιοποίηση της συνάρτησης των κορυφών με την τιμή  $t_1$  επιστρέφει δύο συνεκτικές περιοχές, ωστόσο δεν μπορεί να διακρίνει κάποιες κορυφές της συνάρτησης. Η κατωφλιοποίηση με την τιμή  $t_2$  διακρίνει τις κορυφές που απέτυχε να διακρίνει η  $peak(f)_{t_1}$  αλλά δεν διακρίνει την κορυφή που αυτή κατάφερε να διακρίνει.

Η τιμή του κατωφλίου  $t$  ανήκει στο πεδίο τιμών της εικόνας  $f$ . Στη συνέχεια, από την ακολουθία των δυαδικών εικόνων  $peak(f)_t$  που προκύπτουν εξάγονται οι συνεκτικές συνιστώσες (χρησιμοποιείται γειτνίαση 8-συνεκτικότητας). Συγκεκριμένα, για κάθε τιμή του κατωφλίου  $t$  εξάγονται οι συνεκτικές συνιστώσες  $\{X_{t(i)}\}_{i=1}^{N_t}$ . Οι συνιστώσες αυτές αποτελούν επιπεδοσύνολα με ακαθόριστο σχήμα. Σε κάθε τέτοιο επιπεδοσύνολο αντιστοιχίζεται μια έλλειψη, με τις ίδιες σχηματικές ροπές με το σχήμα των περιοχών των συνόλων ως και δευτέρας τάξης. Με άλλα λόγια, σε κάθε συνιστώσα  $X_{t(i)}$  αντιστοιχίζεται μια πεντάδα  $u_{t(i)}, v_{t(i)}, a_{t(i)}, b_{t(i)}, c_{t(i)}$  έτσι ώστε η έλλειψη που δίνεται από τη εξίσωση

$$\eta_{t(i)} : a_{t(i)}(x - u_{t(i)})^2 + 2b_{t(i)}(x - u_{t(i)})(y - v_{t(i)}) + c_{t(i)}(y - v_{t(i)})^2 = 1$$



**Σχήμα 5.6:** Η πολλαπλή κατωφλιοποίηση της συνάρτησης κορυφών μιας μονοδιάστατης συνάρτησης  $f(x)$ . Είναι εμφανές ότι το αποτέλεσμα της κατωφλιοποίησης δεν μπορεί να δώσει όλες τις κορυφές που διακρίνονται. Με την πολλαπλή κατωφλιοποίηση μπορούν να διακριθούν όλες οι κορυφές.

να περιγράψει βέλτιστα την περιοχή αυτή.

Αφού εξαχθούν οι συνεκτικές συνιστώσες για όλες τις τιμές κατωφλίου, επιδιώκεται η διαγραφή εκείνων των περιοχών που δεν προσφέρουν επιπλέον πληροφορία. Συγκεκριμένα, έστω  $X_{t_1(i)}$  και  $X_{t_2(j)}$  με  $t_1 \neq t_2$  δύο συνεκτικές συνιστώσες που έχουν προέλθει από τις εικόνες  $peak(f)_{t_1}$  και  $peak(f)_{t_2}$  αντίστοιχα. Τότε, η συνιστώσα  $X_{t_1(i)}$  διαγράφεται αν  $X_{t_1(i)} \subset X_{t_2(j)}$  και επίσης τα κέντρα των ελλείψεων  $(u_{t_1(i)}, v_{t_1(i)})$  και  $(u_{t_2(j)}, v_{t_2(j)})$  ταυτίζονται.

Αν  $\{Y_{t(i)}\}_{i=1}^{M_t}$  οι  $M_t$  πλήθος συνεκτικές συνιστώσες μετά τη διαγραφή των περιττών συνιστωσών της εικόνας  $peak(f)_t$ , οι top hat περιοχές ενδιαφέροντος προκύπτουν από την ένωση όλων των τελικών συνιστωσών (ή ελλείψεων που τις χαρακτηρίζουν) για κάθε τιμή του κατωφλίου  $t$ . Δηλαδή,

$$TopHat(f) = Y_{t_1(1)} \cup \dots \cup Y_{t_1(M_{t_1})} \cup Y_{t_2(1)} \cup \dots \cup Y_{t_2(M_{t_2})} \cup \dots \cup Y_{t_k(1)} \cup \dots \cup Y_{t_k(M_{t_k})} \quad (5.10)$$

όπου  $t_1, \dots, t_k$  είναι οι διάφορες τιμές των κατωφλίων που επιλέγονται.

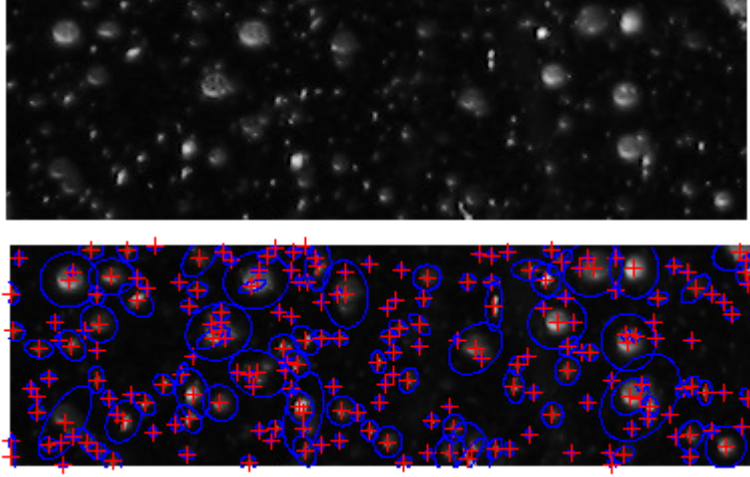
Η ίδια διαδικασία επαναλαμβάνεται για την εικόνα  $f' = 1 - f$  όπου  $f$  η αρχική εικόνα που λαμβάνει τιμές στο διάστημα  $[0, 1]$ . Με τον τρόπο αυτό, υπολογίζονται οι περιοχές ενδιαφέροντος της εικόνας που αντιστοιχούν σε κοιλάδες.

Στο Σχήμα 5.7 φαίνεται το αποτέλεσμα της ανίχνευση των περιοχών ενδιαφέροντος με το μετασχηματισμό top hat.

Η εξαγωγή κορυφών με τον top hat μετασχηματισμό δεν περιορίζει το ύψος των ανιχνευθέντων κορυφών. Σε ορισμένες εφαρμογές ωστόσο είναι επιθυμητή η ανίχνευση των κορυφών με μέγιστο ύψος  $h > 0$ . Τέτοιες κορυφές λέγονται *τρούλοι* (domes) και υπολογίζονται ως εξής. Αφαιρώντας τη σταθερά  $h$  από την εικόνα  $f$  προκύπτει η εικόνα

$$g = f - h \quad (5.11)$$

Διαστέλλοντας την εικόνα  $g$  με ένα συμμετρικό, συμπαγές και κυρτό σύνολο  $B$  με ολοένα αυξανόμενη διάμετρο και περιορίζοντας τις τιμές του αποτελέσματος, ώστε να μην προκύψουν τιμές μεγαλύτερες από τις αντίστοιχες τιμές της εικόνας  $f$  παράγει στο όριο μια εικόνα από κοιλάδες και επίπεδα οροπέδια. Ο τελεστής αυτός λέγεται *reconstruction opening* της  $g$  στην  $f$  και συμβολίζεται  $\rho^-(g|f)$ , όπου  $f$  είναι η εικόνα αναφοράς και  $g$  ο σημαδευτής (marker). Αφαιρώντας το



**Σχήμα 5.7:** Το αποτέλεσμα ανίχνευσης περιοχών ενδιαφέροντος με τον top hat μετασχηματισμό. Το σύνολο  $B$  είναι μια εικόνα δίσκου διάστασης  $9 \times 9$ . Οι τιμές κατωφλίου ανήκουν στο διάστημα  $[0.1, 0.9]$  με σταθερό βήμα ίσο με  $0.1$ .

reconstruction opening από την αρχική εικόνα προκύπτει το αποτέλεσμα

$$domes(f) = f - \rho^-(g|f) \quad (5.12)$$

Ο τελεστής αυτός εξάγει κορυφές που το ύψος τους δεν μπορεί να υπερβεί το ύψος  $h$  αλλά έχουν μη περιορισμένο υποστηρικτικό φάρδος. Σε αντίθεση, ο τελεστής  $peak(f)$  μπορεί να δώσει κορυφές απεριόριστου ύψους αλλά το υποστηρικτικό φάρδος εξαρτάται από το σύνολο  $B$ .

Η προκύπτουσα εικόνα των κορυφών  $domes(f)$  κατωφλιοποιείται με τιμές κατωφλίου  $t$  στο πεδίο τιμών  $(0, h)$ . Στη συνέχεια, εξάγονται οι συνεκτικές συνιστώσες για κάθε εικόνα και επεξεργάζονται με την ίδια μεθοδολογία που περιγράφηκε παραπάνω για τον απλό top hat μετασχηματισμό.

Η διαδικασία επαναλαμβάνεται για την εικόνα  $f' = 1 - f$  όπου  $f$  η αρχική εικόνα που λαμβάνει τιμές στο διάστημα  $[0, 1]$ . Με τον τρόπο αυτό, υπολογίζονται οι περιοχές ενδιαφέροντος της εικόνας που αντιστοιχούν σε κοιλάδες.

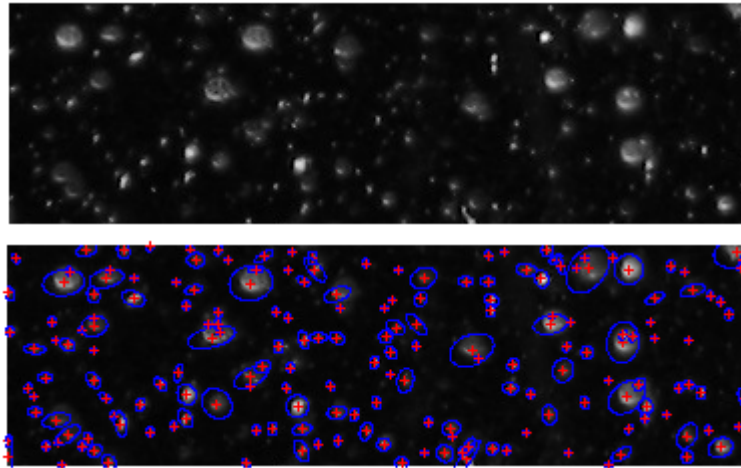
Στο Σχήμα 5.8 φαίνεται το αποτέλεσμα του γενικευμένου top hat τελεστή, όπου οι ελλείψεις περιγράφουν τις περιοχές ενδιαφέροντος.

Είναι εμφανές ότι ο γενικευμένος top hat μετασχηματισμός ανιχνεύει πιο "στενές" περιοχές ενδιαφέροντος. Αυτό οφείλεται στο γεγονός ότι επιλέχθηκαν οι πιο έντονες κορυφές ως περιοχές ενδιαφέροντος. Στο Σχήμα 5.8, ανιχνεύονται όλα τα χαρακτηριστικά της εικόνας με αρκετή ακρίβεια. Σε αντίθεση, από το Σχήμα 5.7 είναι εμφανές ότι ανιχνεύονται τα χαρακτηριστικά αλλά όχι με ικανοποιητική ακρίβεια.

#### 5.1.4 Σύγκριση των Μεθόδων Ανίχνευσης Χαρακτηριστικών

Έχοντας μελετήσει τους διάφορους τρόπους ανίχνευσης και εξαγωγής χαρακτηριστικών, στόχος είναι η αξιολόγηση τους. Η αξιολόγηση των SIFT, MSER και του γενικευμένου Top hat αλγορίθμου θα γίνει αναφορικά με την ικανότητα τους να χρησιμοποιηθούν για εφαρμογές στερέωσης. Συγκεκριμένα, αυτό που ενδιαφέρει είναι οι αλγόριθμοι να ανιχνεύουν ζεύγη σημείων σε δύο διαφορετικές εικόνες του ίδιου αντικειμένου. Πρωτίστως, ενδιαφέρει η απόδοση των μεθόδων στην ανίχνευση έγκυρων ζευγών, δηλαδή ζευγών σημείων που ικανοποιούν τον επιπολικό περιορισμό με συγκεκριμένη ακρίβεια. Για τις δύο υπό εξέταση εικόνες, ανιχνεύονται τα χαρακτηριστικά όπως





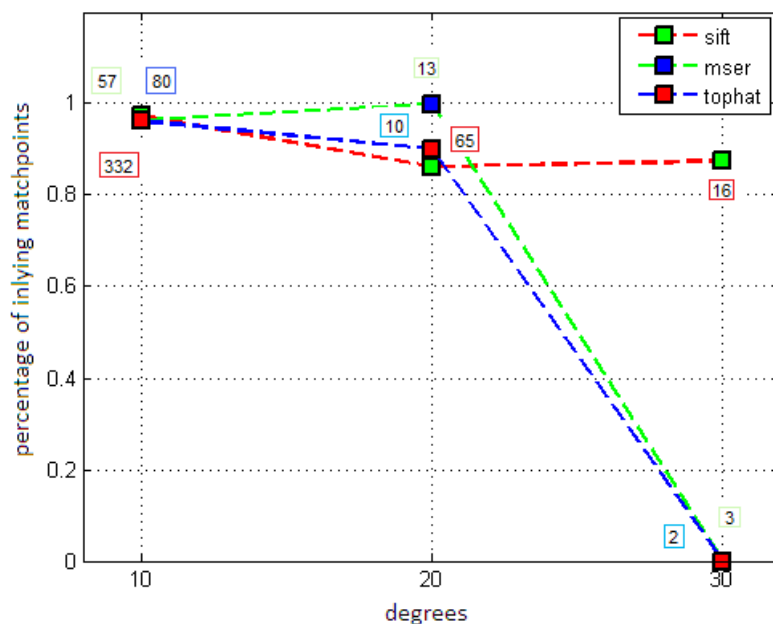
**Σχήμα 5.8:** Οι περιοχές ενδιαφέροντος όπως προκύπτουν από τον γενικευμένο top hat τελεστή με χρήση reconstruction opening. Η τιμή του ύψους  $h = 0.4$  για την εικόνα  $f$  με τιμές στο πεδίο τιμών  $[0,1]$ . Η τιμή κατωφλίου ανήκει στο διάστημα  $(0,h)$  με σταθερό βήμα ίσο με  $0.1$ .

προκύπτουν από τους τρεις αλγόριθμους. Τα χαρακτηριστικά αυτά περιγράφονται με τον SIFT τοπικό περιγραφέα που αναλύθηκε στην ενότητα 5.1.1.2. Οι περιγραφείς των σημείων στις εκάστοτε εικόνες συγκρίνονται μεταξύ τους για την εύρεση σημείων ταιριάσματος (βλ. ενότητα 5.1.1.3). Η διαδικασία του ταιριάσματος εξαρτάται από την παράμετρο  $distRatio$ . Το κατώφλι αυτό καθορίζει το μέγιστο λόγο απόστασης του διανύσματος περιγραφής ενός χαρακτηριστικού στη μια εικόνα μεταξύ των διανυσμάτων περιγραφής των δύο κοντινότερων προς αυτό χαρακτηριστικών στη δεύτερη εικόνα ώστε να προκύψει ζεύγος σημείων. Η εγκυρότητα των ζευγών που υπολογίζονται εξαρτάται από τον υπολογισμό του θεμελιώδους πίνακα με τον αλγόριθμο RANSAC (βλ. ενότητα 5.2), με τον οποίο κρίνεται αν ένας ζεύγος είναι έγκυρο. Η παράμετρος  $ransacRatio$  είναι το σχετικό κατώφλι του RANSAC αλγόριθμου, δηλαδή ένα ζεύγος σημείων  $\mathbf{x} \leftrightarrow \mathbf{x}'$  όπως προέκυψε από τη διαδικασία ταιριάσματος τ.ω. οι αποστάσεις  $d_{\perp}, d'_{\perp}$  (βλ. σχέση 5.27) υπερβαίνουν το απόλυτο κατώφλι απόστασης δεν αποτελεί έγκυρο ζεύγος (inlier). Το απόλυτο κατώφλι του RANSAC προκύπτει από το γινόμενο του σχετικού κατωφλίου με τη μέγιστη διάσταση της εικόνας.

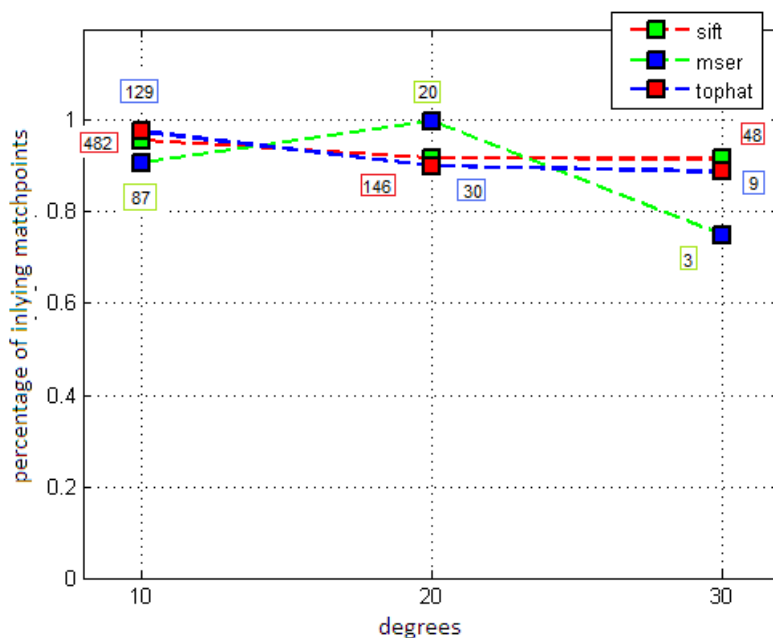
Στην παρούσα αξιολόγηση, γίνονται τα εξής πειράματα. Αρχικά, εφαρμόζονται οι τρεις αλγόριθμοι σε μια βάση δεδομένων αποτελούμενοι από εικόνες με κέντρα καμερών που παρουσιάζουν γωνιακή διαφορά  $10^\circ$ . Η εικόνα αναφοράς (θέση  $0^\circ$ ) μαζί με τις εικόνες (θέση  $\phi, \phi = 10^\circ, 20^\circ, 30^\circ$ ) της βάσης αποτελούν εισόδου στους αλγόριθμους με σκοπό τον υπολογισμό των έγκυρων σημείων ταιριάσματος (matchpoints). Οι τιμές των σχετικών κατωφλίων τέθηκαν ίσες με  $distRatio = 0.55$  και  $ransacRatio = 0.005$ . Στο Σχήμα 5.9 φαίνονται τα αποτελέσματα. Στον x-άξονα φαίνεται η γωνιακή διαφορά των εικόνων που συγκρίνονται και στον y-άξονα φαίνεται το ποσοστό των έγκυρων ζευγών από το συνολικό πλήθος ζευγών που ανιχνεύθηκαν. Πάνω από τα σημεία του γραφήματος παρουσιάζονται σε κουτιά το συνολικό πλήθος ζευγών που ανιχνεύθηκαν με την εκάστοτε μέθοδο. Με κόκκινο περίγραμμα παρουσιάζεται το πλήθος των συνολικών ζευγών όπως προέκυψε με τον SIFT αλγόριθμο, με πράσινο περίγραμμα τα MSER ζεύγη και με μπλε τα ζεύγη όπως προέκυψαν με τον γενικευμένο top hat τελεστή.

Στο δεύτερο πείραμα, εφαρμόζεται η ίδια διαδικασία για τιμή κατωφλίου  $distRatio = 0.65$ . Στο Σχήμα 5.10 φαίνεται το αποτέλεσμα

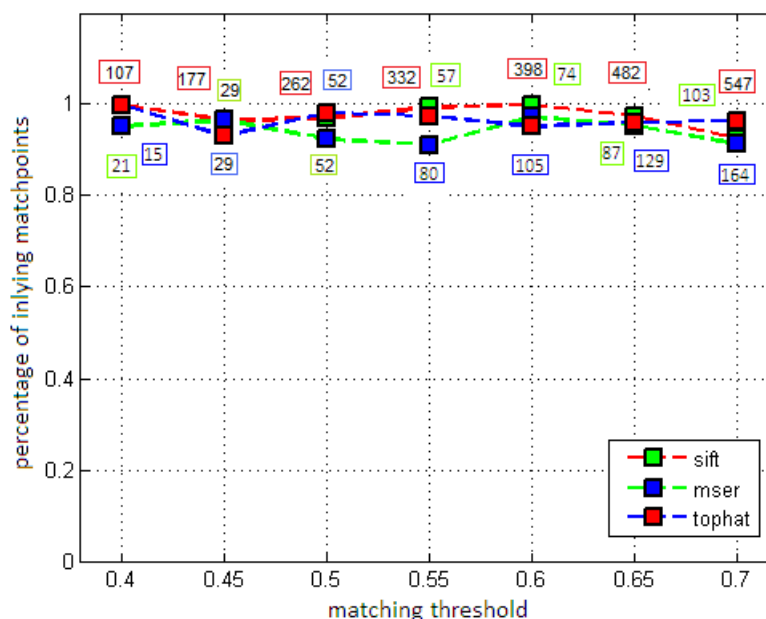
Στο τρίτο πείραμα, εξετάζονται δύο εικόνες με γωνιακή διαφορά των κεντρών των καμερών τους ίση με  $10^\circ$ . Η τιμή του σχετικού κατωφλίου είναι  $ransacRatio = 0.005$  και μεταβάλλεται η



**Σχήμα 5.9:** Σύγκριση των τριών αλγορίθμων ανίχνευσης χαρακτηριστικών για εφαρμογές στερέωσης. Στο διάγραμμα παρουσιάζεται το ποσοστό των έγκυρων ζευγών σημείων στις δύο εικόνες (από τη βάση δεδομένων του Σχήματος 5.1) ως προς τη γωνιακή διαφορά των κέντρων των καμερών των εικόνων για τους αλγορίθμους ξεχωριστά. Επιπλέον, παρουσιάζονται για κάθε ζεύγος εικόνων το συνολικό πλήθος ζευγών σημείων που ανιχνεύθηκαν. Τιμές κατωφλίου  $distRatio = 0.55$  και  $ransacRatio = 0.005$ .



**Σχήμα 5.10:** Σύγκριση των τριών αλγορίθμων ανίχνευσης χαρακτηριστικών για εφαρμογές στερέωσης. Στο διάγραμμα παρουσιάζεται το ποσοστό των έγκυρων ζευγών σημείων στις δύο εικόνες ως προς τη γωνιακή διαφορά των κέντρων των καμερών των εικόνων για τους αλγορίθμους ξεχωριστά. Επιπλέον, παρουσιάζονται για κάθε ζεύγος εικόνων το συνολικό πλήθος ζευγών σημείων που ανιχνεύθηκαν. Τιμές κατωφλίου  $distRatio = 0.65$  και  $ransacRatio = 0.005$ .



**Σχήμα 5.11:** Σύγκριση των τριών αλγορίθμων ανίχνευσης χαρακτηριστικών για εφαρμογές στερέωσης. Στο διάγραμμα παρουσιάζεται το ποσοστό των έγκυρων ζευγών σημείων στις δύο εικόνες ως προς την τιμή του κατωφλίου  $distRatio$  για τους αλγορίθμους ξεχωριστά. Επιπλέον, παρουσιάζονται για κάθε ζεύγος εικόνων το συνολικό πλήθος ζευγών σημείων που ανιχνεύθηκαν. Τιμές κατωφλίου  $ransacRatio = 0.005$  και γωνιακή διαφορά εικόνων ίση με  $10^\circ$ .

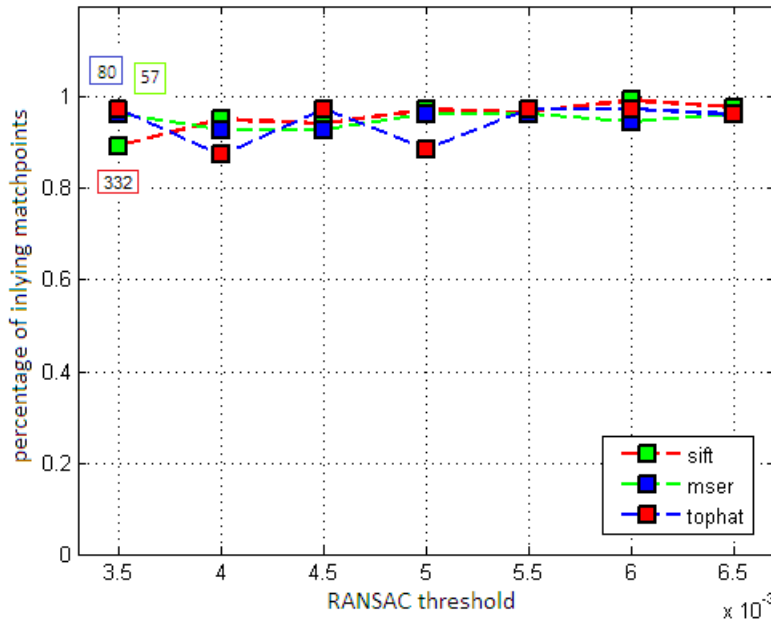
τιμή του κατωφλίου  $distRatio$  στο πεδίο τιμών  $[0.4, 0.7]$ . Στο Σχήμα 5.11 φαίνεται το αποτέλεσμα

Τέλος, στο τέταρτο πείραμα, και πάλι εξετάζονται δύο εικόνες με γωνιακή διαφορά κέντρων καμερών ίση με  $10^\circ$ . Η τιμή του κατωφλίου είναι  $distRatio = 0.55$  ενώ μεταβάλλεται η τιμή του σχετικού κατωφλίου  $ransacRatio$  στο διάστημα  $[0.0035, 0.0065]$ . Στο Σχήμα 5.12 φαίνεται το αποτέλεσμα. Να σημειωθεί πως στο πείραμα αυτό αλλάζει μόνο το ποσοστό των έγκυρων ζευγών και όχι το πλήθος των ζευγών που ανιχνεύονται συνολικά καθώς η παράμετρος ταιριάσματος και οι υπο εξέταση εικόνες είναι αμετάβλητες.

Από τα Σχήματα 5.9 και 5.10 είναι εμφανές ότι η απόδοση των αλγορίθμων μειώνεται με αύξηση της γωνιακής διαφοράς των εικόνων. Συγκεκριμένα, τόσο το ποσοστό των έγκυρων ζευγών σημείων όσο και το συνολικό τους πλήθος μειώνεται όσο μεγαλύτερη η απόσταση του οπτικού σημείου από τη θέση αναφοράς. Μάλιστα, για  $distRatio = 0.55$  ο MSER αλγόριθμος και ο μορφολογικός αλγόριθμος δεν δίνουν έγκυρα ζεύγη για γωνιακή διαφορά ίση με  $30^\circ$ . Από τα δύο Σχήματα παρατηρούμε ότι για γωνιακή διαφορά ίση με  $10^\circ$  οι τρεις μέθοδοι έχουν την ίδια περίπου απόδοση με τον SIFT αλγόριθμο να υπερέχει ως προς το πλήθος των ζευγών που ανιχνεύονται. Για γωνιακή διαφορά  $20^\circ$  υπερέχει ο MSER αλγόριθμος ως προς το ποσοστό αλλά όχι ως προς το πλήθος έγκυρων ζευγών. Για μεγαλύτερη γωνιακή διαφορά, η υπεροχή του SIFT αλγορίθμου είναι εμφανής.

Από το Σχήμα 5.11 παρατηρείται ότι η μέθοδος MSER έχει την χειρότερη απόδοση ως προς το ποσοστό έγκυρων ζευγών. Η μέθοδος του γενικευμένου top hat μετασχηματισμού είναι καλύτερη συγκριτικά με τον MSER αλγόριθμο ως προς το ποσοστό και ως προς το πλήθος έγκυρων ταιριασμάτων. Ο SIFT αλγόριθμος φαίνεται να υπερέχει και ως προς το ποσοστό και ως προς το πλήθος.

Στο Σχήμα 5.12 φαίνεται για άλλη μια φορά η υπεροχή του SIFT αλγορίθμου. Οι άλλες δύο



**Σχήμα 5.12:** Σύγκριση των τριών αλγορίθμων ανίχνευσης χαρακτηριστικών για εφαρμογές στερέωσης. Στο διάγραμμα παρουσιάζεται το ποσοστό των έγκυρων ζευγών σημείων στις δύο εικόνες ως προς την τιμή του σχετικού κατωφλίου *ransacRatio* για τους αλγορίθμους ξεχωριστά. Επιπλέον, παρουσιάζονται για κάθε ζεύγος εικόνων το συνολικό πλήθος ζευγών σημείων που ανιχνεύθηκαν. Τιμές κατωφλίου *distRatio* = 0.55 και γωνιακή διαφορά εικόνων ίση με  $10^\circ$ .

μέθοδου ενώ παρουσιάζουν πολύ καλά ποσοστά έγκυρων σημείων, ανιχνεύουν μικρό συγκριτικά πλήθος ζευγών.

Από τα παραπάνω γίνεται εμφανές ότι ο SIFT αλγόριθμος υπερέχει. Αυτήν η υπεροχή οφείλεται στο γεγονός ότι ανιχνεύει μεγάλο πλήθος σημείων ενδιαφέροντος και κατά επέκταση ζεύγη σημείων (matchpoints) στις δύο εικόνες. Ωστόσο κανένας αλγόριθμος δεν έχει ικανοποιητικά αποτελέσματα για γωνιακή διαφορά εικόνων μεγαλύτερη των  $40^\circ$ . Αυτό είναι αναμενόμενο καθώς οι εικόνες με μεγάλη απόσταση οπτικών σημείων δεν έχουν κοινά ευδιάκριτα σημεία. Η υπεροχή των SIFT σημείων οδηγεί στη χρήση τους στην παρούσα εργασία για την 3Δ ανακατασκευή.

## 5.2 Υπολογισμός του Θεμελιώδους Πίνακα

Αφού ανιχνευθούν τα ζεύγη σημείων ταιριάσματος μεταξύ δύο διαδοχικών εικόνων  $I$  και  $J$  της βάσης δεδομένων, σκοπός είναι ο υπολογισμός του θεμελιώδους πίνακα που αντιστοιχεί στις δύο αυτές εικόνες. Ο υπολογισμός αυτός γίνεται με βάση τον 8-point αλγόριθμο (βλ. [22]).

Παρακάτω δίνονται οι βασικές σχέσεις που σχετίζονται με τον υπολογισμό του πίνακα  $F$  περιγράφεται αναλυτικά ο αλγόριθμος.

### 5.2.1 Βασικές Σχέσεις

Ο θεμελιώδης πίνακας ικανοποιεί τη σχέση

$$\mathbf{x}'F\mathbf{x} = 0 \quad (5.13)$$

για κάθε ζεύγος σημείων  $\mathbf{x} \leftrightarrow \mathbf{x}'$  στις δύο εικόνες. Η παραπάνω εξίσωση μπορεί να γραφεί στη μορφή

$$\begin{aligned} x'x f_{11} + x'y f_{12} + x'f_{13} + y'x f_{21} + y'y f_{22} + y'f_{23} + x f_{31} + y f_{32} + f_{33} = 0 \\ (x'x, x'y, x', y'x, y'y, y', x, y, 1) \cdot \mathbf{f} = 0 \end{aligned} \quad (5.14)$$

όπου  $\mathbf{x} = (x, y, 1)^T$  και  $\mathbf{x}' = (x', y', 1)$ . Αν υπάρχουν  $N$  ζεύγη σημείων, προκύπτει ομογενές γραμμικό σύστημα εξισώσεων της μορφής

$$\mathbf{A}\mathbf{f} = 0 \quad (5.15)$$

όπου ο πίνακας  $A$  περιέχει στην  $i$ -στη γραμμή  $A_{i-throw} = (x'_i x_i, x'_i y_i, x'_i, y'_i x_i, y'_i y_i, y'_i, x_i, y_i, 1)$ .

Επειδή το σύστημα εξισώσεων είναι ομογενές η λύση  $\mathbf{f}$  μπορεί να προσδιοριστεί με ακρίβεια πολλαπλασιαστικής σταθεράς. Για να υπάρχει λύση, ο πίνακας  $A$  πρέπει να έχει βαθμό το πολύ 8, και αν έχει βαθμό ακριβώς 8 τότε η λύση είναι μοναδική (ως προς σταθερά). Ωστόσο, καθώς οι μετρήσεις των σημείων είναι θορυβώδεις, ο βαθμός του πίνακα ενδέχεται να είναι 9 (αν  $N \geq 9$ ). Στην περίπτωση αυτή, η λύση είναι το δεξί ιδιάζον διάνυσμα που αντιστοιχεί στην μικρότερη ιδιάζουσα ιδιοτιμή, όπως προκύπτει από την SVD ανάλυση του πίνακα  $A$ .

Μια σημαντική ιδιότητα του θεμελιώδους πίνακα είναι ότι έχει βαθμό 2. Ωστόσο, η λύση που προκύπτει από την επίλυση του παραπάνω συστήματος μέσω SVD ανάλυσης δεν δίνει συνήθως πίνακα βαθμού 2. Ο πιο βολικός τρόπος να εξασφαλιστεί η ιδιότητα αυτή είναι να αντικατασταθεί ο πίνακας  $F$  όπως υπολογίζεται από την SVD ανάλυση. Ο πίνακας  $F$  αντικαθίσταται από τον πίνακα  $F'$  που ελαχιστοποιεί την Frobenius νόρμα  $\|F - F'\|$  υπό τον περιορισμό  $\det(F') = 0$ . Για να γίνει αυτό, χρησιμοποιείται και πάλι η μέθοδος της SVD ανάλυσης. Συγκεκριμένα, έστω  $F = UDV^T$  η SVD ανάλυση του  $F$  με  $D = \text{diag}(r, s, t)$  όπου  $r \geq s \geq t$ . Τότε, ο πίνακας  $F' = U \text{diag}(r, s, 0) V^T$  αποτελεί μια λύση στο πρόβλημα ελαχιστοποίησης της Frobenius νόρμας.

### 5.2.2 Ο Αλγόριθμος 8-σημείων

Ο αλγόριθμος αυτός αποτελεί τον πιο εύκολο τρόπο υπολογισμού του θεμελιώδους πίνακα. Το σημαντικότερο στάδιο του αλγορίθμου είναι ο κατάλληλος μετασχηματισμός των δεδομένων. Ο μετασχηματισμός αυτός γίνεται πριν από τον σχηματισμό των εξισώσεων και αποσκοπεί στην μείωση της επίδρασης του θορύβου των μετρήσεων στον υπολογισμό του πίνακα  $F$ .

Συγκεκριμένα, τα σημεία μετατοπίζονται ώστε το κέντρο του να μεταφερθεί στην αρχή του συστήματος συντεταγμένων και κλιμακώνονται ώστε η RMS απόσταση των σημείων από το κέντρο να είναι ίση με  $\sqrt{2}$ .

Ο Αλγόριθμος 4 περιγράφει τον αλγόριθμο 8-σημείων

---

**Algorithm 4** Αλγόριθμος 8-σημείων: Υπολογισμός του θεμελιώδους πίνακα  $F$  από  $\geq 8$  ζεύγη σημείων  $\{\mathbf{x}_i \leftrightarrow \mathbf{x}'_i\}$

---

- (i). **Κανονικοποίηση:** Μετασχηματισμός των δεδομένων  $\hat{\mathbf{x}}_i = T\mathbf{x}_i$  και  $\hat{\mathbf{x}}'_i = T'\mathbf{x}'_i$ , όπου  $T$  και  $T'$  οι πίνακες μετασχηματισμού
  - (ii). Υπολογισμός του θεμελιώδους πίνακα  $\hat{F}$  που αντιστοιχεί στα ζεύγη κανονικοποιημένων σημείων  $\hat{\mathbf{x}}_i \leftrightarrow \hat{\mathbf{x}}'_i$ 
    - (a). **Γραμμική επίλυση:** Προσδιορισμός του  $\hat{F}$  από την SVD ανάλυση του  $\hat{A}$
    - (b). **Επιβολή περιορισμού:** Αντικατάσταση του  $\hat{F}$  με τον  $\hat{F}'$  ώστε  $\det(\hat{F}') = 0$
  - (iii). **Αποκανονικοποίηση:** Ο θεμελιώδης πίνακας δίνεται από τη σχέση  $F = T'\hat{F}'T$
-

### 5.2.3 Σφάλμα Sampson

Έστω  $\mathbf{x} \leftrightarrow \mathbf{x}'$  αντίστοιχα σημεία των δύο εικόνων, όπου  $\mathbf{x} = (x, y, 1)^T$  και  $\mathbf{x}' = (x', y', 1)^T$ . Επίσης, έστω  $\mathcal{V}_F$  ο χώρος των ζευγών  $(\mathbf{x}, \mathbf{x}')$  έτσι ώστε  $\mathbf{x}'^T F \mathbf{x} = 0$  και ο τελεστής  $C_F(X) = \mathbf{x}'^T F \mathbf{x}$  όπου  $X = (x, y, x', y')$ .

Αν οι μετρήσεις των ζευγών των σημείων  $\mathbf{x} \leftrightarrow \mathbf{x}'$  είναι ιδανικές, τότε για  $(\mathbf{x}, \mathbf{x}') \in \mathcal{V}_F$ . Ωστόσο, η ύπαρξη θορύβου στις μετρήσεις δίνει ζεύγη σημείων που δεν ανήκουν στο χώρο  $\mathcal{V}_F$ . Σκοπός είναι, λοιπόν, ο προσδιορισμός της ελάχιστης απόστασης των μη ιδανικών μετρήσεων  $(\mathbf{x}, \mathbf{x}')$  από το χώρο αυτό, ώστε να εκτιμηθεί το αντίστοιχο "κοντινότερο" ιδανικό ζεύγος σημείων που ικανοποιεί τον επιπολικό περιορισμό. Αυτό θα γίνει με την προσέγγιση πρώτων όρων.

$$C_F(\hat{X}) = C_F(X + \delta X) = C_F(X) + \frac{\partial C_F(X)}{\partial X} \delta X \quad (5.16)$$

Αν  $\hat{X} \in \mathcal{V}_F$  τότε  $C_F(\hat{X}) = 0$ . Επομένως,

$$C_F(X) + \frac{\partial C_F(X)}{\partial X} \delta X = 0 \quad (5.17)$$

Ορίζοντας το κόστος  $e = C_F(X)$  και τον πίνακα μερικών παραγώγων  $J = \frac{\partial C_F(X)}{\partial X}$ . Επομένως,

$$J \delta X = -e \quad (5.18)$$

Για το πρόβλημα ελαχιστοποίησης, ζητούμενο είναι να ελαχιστοποιηθεί το  $\|\delta X\|$  υπό τον περιορισμό  $J \delta X = -e$ . Για το σκοπό αυτό, ορίζεται το συναρτησιακό

$$\Psi = \delta X^T \delta X - 2\lambda^T (J \delta X + e) \quad (5.19)$$

Για την εύρεση του ελαχίστου, η πρώτη παράγωγος του  $\Psi$  ως προς  $\delta X$  τίθεται ίση με 0

$$\frac{\partial \Psi}{\partial \delta X} = 2\delta X^T - 2\lambda^T J = 0 \quad (5.20)$$

Επομένως,  $\delta X = J^T \lambda$ . Επιπλέον, η πρώτη παράγωγος του  $\Psi$  ως προς  $\lambda$  τίθεται ίση με 0.

$$\frac{\partial \Psi}{\partial \lambda} = J \delta X + e = 0 \quad (5.21)$$

δηλαδή προκύπτει ο περιορισμός. Από τις σχέσεις 5.20 και 5.21, προκύπτει

$$\lambda = -(J J^T)^{-1} e \quad (5.22)$$

Τελικά, από την παραπάνω σχέση προκύπτει

$$\delta X = -J^T (J J^T)^{-1} e \quad (5.23)$$

Ως σφάλμα Sampson ορίζεται το μέτρο της μετατόπισης  $\delta X$ , δηλαδή

$$SE = \|\delta X\|^2 = \delta X^T \delta X = e^T (J J^T)^{-1} e \quad (5.24)$$

όπου  $J = \left[ \frac{\partial C_F}{\partial x} \frac{\partial C_F}{\partial y} \frac{\partial C_F}{\partial x'} \frac{\partial C_F}{\partial y'} \right]$ . Οι μερικές παράγωγοι υπολογίζονται εύκολα. Συγκεκριμένα, ισχύει

$$\frac{\partial \mathbf{x}'^T F \mathbf{x}}{\partial x} = \frac{\partial \mathbf{x}'^T F \mathbf{x}}{\partial \mathbf{x}} \frac{\partial \mathbf{x}}{\partial x} = \mathbf{x}'^T F (1, 0, 0)^T = (\mathbf{x}'^T F)_1 \quad (5.25)$$

Όμοια,  $\frac{\partial \mathbf{x}'^T F \mathbf{x}}{\partial y} = (\mathbf{x}'^T F)_2$ ,  $\frac{\partial \mathbf{x}'^T F \mathbf{x}}{\partial x'} = (F \mathbf{x})_1$  και  $\frac{\partial \mathbf{x}'^T F \mathbf{x}}{\partial y'} = (F \mathbf{x})_2$ . Εύκολα, λοιπόν, προκύπτει ότι το σφάλμα Sampson δίνεται από τη σχέση

$$SE() = \frac{(\mathbf{x}'^T F \mathbf{x}_i)^2}{(\mathbf{x}'^T F)_1^2 + (\mathbf{x}'^T F)_2^2 + (F \mathbf{x})_1^2 + (F \mathbf{x})_2^2} \quad (5.26)$$

Γεωμετρικά, το σφάλμα Sampson είναι ένα μέτρο της απόστασης του σημείου  $\mathbf{x}'$  από την επιπολική ευθεία  $\mathcal{I}' = F \mathbf{x}$ . Αν  $\mathbf{x}' \in \mathcal{I}'$  τότε  $\mathbf{x}'^T F \mathbf{x} = 0$  και άρα  $SE = 0$ .

### 5.2.4 Ο Τελικός Αλγόριθμος

Ο τελικός αλγόριθμος υπολογισμού του θεμελιώδους πίνακα χρησιμοποιεί τον αλγόριθμο 8-σημείων και τον RANSAC αλγόριθμο (βλ. Παράρτημα Β').

Πριν την εφαρμογή του αλγορίθμου, αρχικοποιούνται οι παράμετροι του RANSAC αλγορίθμου. Συγκεκριμένα, αρχικοποιούνται το κατώφλι απόστασης (distance threshold), η πιθανότητα  $p$  τουλάχιστον ένα επιλεγθέν σύνολο που αποτελείται από  $s = 8$  ζεύγη σημείων να μην διαθέτει καμία ακραία τιμή (outlier) καθώς και η πιθανότητα  $\epsilon$  ένα ζεύγος σημείων να είναι ακραία τιμή. Οι δύο τελευταίες τιμές καθορίζουν τις επαναλήψεις  $N$ , δηλαδή το πλήθος των επιλεγμένων συνόλων  $s$  δεδομένων, για τις οποίες θα εφαρμοστεί ο αλγόριθμος ώστε να καταλήξουμε στο θεμελιώδη πίνακα  $F$ .

Σε κάθε επανάληψη  $j = 1, \dots, N$  του αλγορίθμου υπολογισμού, επιλέγονται τυχαία 8 ζεύγη σημείων από τη βάση δεδομένων, τα οποία χρησιμοποιούνται για την εφαρμογή του αλγορίθμου 8-σημείων. Από τον πίνακα  $F_j$  που προκύπτει, υπολογίζεται το πλήθος των ακραίων σημείων με βάση το κατώφλι απόστασης. Συγκεκριμένα, για όλα τα ζεύγη σημείων  $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$  της βάσης υπολογίζονται οι αποστάσεις

$$d_{\perp}^{(i)} = \frac{(\mathbf{x}'_i^T F \mathbf{x}_i)^2}{(F \mathbf{x}_i)_1^2 + (F \mathbf{x}_i)_2^2} d'_{\perp}^{(i)} = \frac{(\mathbf{x}'_i^T F \mathbf{x}_i)^2}{(F^T \mathbf{x}'_i)_1^2 + (F^T \mathbf{x}'_i)_2^2} \quad (5.27)$$

όπου  $(F \mathbf{x}_i)_j^2$  παριστάνει το τετράγωνο της  $j$ -στης συνιστώσας του διανύσματος  $F \mathbf{x}_i$ . Εύκολα μπορεί κανείς να δει ότι το μέγεθος  $d_{\perp}^{(i)}$  παριστάνει την απόσταση του σημείου  $\mathbf{x}'_i$  από την επιπολική ευθεία  $\mathbf{l}' = F \mathbf{x}$ . Όμοια για το  $d'_{\perp}^{(i)}$ . Αν ισχύει  $d_{\perp}^{(i)} > \text{distancethreshold}$  ή  $d'_{\perp}^{(i)} > \text{distancethreshold}$ , τότε το ζεύγος σημείων  $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$  είναι outlier.

Μετά από  $N$  επαναλήψεις, επιλέγεται αυτός ο θεμελιώδης πίνακας που δίνει το μεγαλύτερο πλήθος από inliers ή (σε περίπτωση ισοβαθμίας) το μικρότερο σφάλμα υπολοίπου

$$\text{resErr} = \frac{(d_{\perp}^{(i)})^2 + (d'_{\perp}^{(i)})^2}{\text{InlierCount}} \quad (5.28)$$

όπου οι αποστάσεις  $d_{\perp}$  έχουν υπολογιστεί για εκείνα τα ζεύγη  $\mathbf{x} \leftrightarrow \mathbf{x}'$  που είναι inliers για την τρέχουσα επιλογή του θεμελιώδους πίνακα και  $\text{InlierCount}$  είναι το πλήθος τους.

Έστω  $F^{(b)}$  ο θεμελιώδης πίνακας όπως προκύπτει από τον RANSAC αλγόριθμο και  $\mathbf{x}^{(b)} \leftrightarrow \mathbf{x}'^{(b)}$  τα ζεύγη σημείων που αποτελούν inliers για το θεμελιώδη συγκεκριμένο πίνακα. Με βάση το συνολικό σφάλμα Sampson  $SE_{total} = \sum_i (SE(X_i^{(b)}))^2$ , είναι δυνατό να υπολογιστεί μια διόρθωση  $\delta F$  τέτοια ώστε ο θεμελιώδης πίνακας  $F^{(b)} + \delta F$  να ελαχιστοποιεί το συνολικό σφάλμα  $SE_{total}$ . Αυτό μπορεί να πραγματοποιηθεί με την λύση μη γραμμικών ελαχίστων τετραγώνων της μορφής

$$\min_F \|\mathbf{SE}(F)\|_2^2 = \min_F (SE_1(F)^2 + \dots + SE_n(F)^2) \quad (5.29)$$

όπου στην παραπάνω εξίσωση συμβολίζεται  $SE_j = SE(X_j)$ . Επειδή η παράμετρος του προβλήματος ελαχιστοποίησης (εδώ ο πίνακας  $F$ ) εμπλέκεται στις εξισώσεις  $S_j$  μη γραμμικά, δεν υπάρχει κλειστός τύπος που να δίνει τη λύση του. Οι μέθοδοι για την εύρεση λύσης είναι αριθμητικές, όπως η μέθοδος Levenberg-Marquardt, και απαιτούν αρχική τιμή. Η επίλυση του παραπάνω προβλήματος επιστρέφει την τιμή του θεμελιώδους πίνακα  $F_0$  που ελαχιστοποιεί συνολικό σφάλμα Sampson, με αρχική τιμή τον πίνακα  $F^{(b)}$ . Τα ζεύγη σημείων που θα χρησιμοποιηθούν για την 3Δ ανακατασκευή είναι τα  $\mathbf{x}^{(b)} \leftrightarrow \mathbf{x}'^{(b)}$  όπως προκύπτουν από τον RANSAC αλγόριθμο ενώ ο θεμελιώδης πίνακας των δύο εικόνων είναι ο  $F_0$ .

### 5.3 Ανακατασκευή Τρισδιάστατου Νέφους Σημείων

Αφού προσδιορίστηκε ο θεμελιώδης πίνακας των δύο υπό εξέταση εικόνων, σκοπός είναι να προσδιοριστεί η 3Δ θέση των σημείων που αντιστοιχούν στα ζεύγη σημείων. Προτού γίνει αυτό, πρέπει να προσδιοριστούν οι εσωτερικές και οι εξωτερικές παράμετροι των δύο καμερών. Στη συνέχεια, με τις μεθόδους δομής από κίνηση προσδιορίζεται η θέση των σημείων στο χώρο.

#### 5.3.1 Προσδιορισμός των Παραμέτρων της Κάμερας

**Εσωτερικές παράμετροι.** Οι εσωτερικές παράμετροι, δηλαδή τα στοιχεία του πίνακα βαθμονόμησης  $K_i$ , θεωρούνται γνωστές για την  $i$ -στη κάμερα. Οι παράμετροι αυτοί βρίσκονται συνήθως στο EXIF tag που συνοδεύει τις εικόνες και αποθηκεύτηκαν εκεί από την κάμερα κατά τη στιγμή λήψης της εικόνας. Ωστόσο, αν δεν είναι γνωστές μπορούν να εκτιμηθούν.

Ένας τρόπος εκτίμησης των εσωτερικών παραμέτρων σχετίζεται με τον προσδιορισμό σημείων φυγής της εικόνας. Για τον προσδιορισμό των σημείων φυγής σε μια εικόνα μπορεί να χρησιμοποιηθεί η μέθοδος του κεφαλαίου 4

Η γνώση του πίνακα βαθμονόμησης δίνει τη δυνατότητα μετρικής ανακατασκευής, δηλαδή διατηρούνται τα σχήματα, οι γωνίες και οι σχετικές αποστάσεις. Η πραγματική διάσταση ωστόσο απαιτεί πρότερη γνώση της διάστασης ενός τμήματος του αντικειμένου. Για το σκοπό, ο προσδιορισμός των παραμέτρων αυτών είναι απαραίτητος.

**Εξωτερικές παράμετροι.** Για τον προσδιορισμό των εξωτερικών παραμέτρων, δηλαδή θέση  $\mathbf{t}_i$  και προσανατολισμό  $R_i$  της κάμερας, μπορεί να χρησιμοποιηθεί η ιδέα του κεφαλαίου 3, που αναφέρεται στην δομή από κίνηση δύο όψεων. Στην περίπτωση αυτή, ο προβολικός πίνακας της μίας κάμερας τίθεται αυθαίρετα σε μία τιμή και προσδιορίζεται ο προβολικός πίνακας της δεύτερης.

Στη συγκεκριμένη εφαρμογή, οι προβολικοί πίνακες των καμερών δίνονται, καθώς μετρήθηκαν κατά τη διαδικασία λήψης των εικόνων.

#### 5.3.2 Προσδιορισμός της 3Δ Θέσης των Σημείων

Έχοντας γνωστούς του προβολικούς πίνακες  $P_1$  και  $P_2$  δύο διαδοχικών εικόνων, γίνεται ο προσδιορισμός της θέσης των 3Δ σημείων. Παρακάτω περιγράφεται ο αλγόριθμος που βασίζεται στις ιδέα του τριγωνισμού του κεφαλαίου 3

Έστω τα ζεύγη σημείων στις δύο εικόνες  $\mathbf{x} \leftrightarrow \mathbf{x}'$  όπου  $\mathbf{x} \in I_1$  και  $\mathbf{x}' \in I_2$  με  $I_1, 2$  οι δυο 2Δ εικόνες αντίστοιχα. Τότε ισχύει

$$\begin{aligned}\mathbf{x} &= P_1 \mathbf{X} \\ \mathbf{x}' &= P_2 \mathbf{X}\end{aligned}\tag{5.30}$$

Οι εξισώσεις 5.30 μπορούν να συνδυαστούν σε μία της μορφής  $A\mathbf{X} = 0$ , που είναι γραμμική ως προς  $\mathbf{X}$ . Ο πίνακας  $A$  μπορεί να προκύψει από την παρατήρηση ότι από τις σχέσεις 5.30 και για την πρώτη εικόνα ισχύει  $\mathbf{x} \times (P\mathbf{X}) = 0$ . Επομένως, αν  $\mathbf{x} = w(x, y, 1)^T$  τότε η προηγούμενη σχέση γράφεται

$$\begin{aligned}x(\mathbf{p}_1^{3T} \mathbf{X}) - (\mathbf{p}_1^{1T} \mathbf{X}) &= 0 \\ y(\mathbf{p}_1^{3T} \mathbf{X}) - (\mathbf{p}_1^{2T} \mathbf{X}) &= 0 \\ x(\mathbf{p}_1^{2T} \mathbf{X}) - y(\mathbf{p}_1^{1T} \mathbf{X}) &= 0\end{aligned}\tag{5.31}$$

όπου  $\mathbf{p}_1^{iT}$  είναι η  $i$ -στη γραμμή του πίνακα  $P_1$ . Η τελευταία από τις παραπάνω εξισώσεις είναι γραμμικός συνδυασμός των άλλων δύο. Επιπλέον, είναι φανερό ότι οι εξισώσεις είναι γραμμικές



ως προς  $\mathbf{X}$ . Όμοια ισχύει και για τη δεύτερη εικόνα. Επομένως, ο πίνακας  $A$  προκύπτει

$$A = \begin{bmatrix} x\mathbf{p}_1^{3T} - \mathbf{p}_1^{1T} \\ y\mathbf{p}_1^{3T} - \mathbf{p}_1^{2T} \\ x'\mathbf{p}_2^{3T} - \mathbf{p}_2^{1T} \\ y'\mathbf{p}_2^{3T} - \mathbf{p}_2^{2T} \end{bmatrix}$$

όπου  $\mathbf{x} = w(x, y, 1)^T$  και  $\mathbf{x}' = w'(x', y', 1)^T$ .

Υπολογιστικά, για να αποφευχθούν λάθη λόγω προσεγγίσεων κάθε γραμμή του πίνακα  $A$  κανονικοποιείται, δηλαδή

$$A = \begin{bmatrix} A_1/\|A_1\| \\ A_2/\|A_2\| \\ A_3/\|A_3\| \\ A_4/\|A_4\| \end{bmatrix}$$

Η επίλυση της εξίσωσης  $A\mathbf{X} = 0$  γίνεται με τη μέθοδο της SVD ανάλυσης. Με άλλα λόγια, η λύση  $\mathbf{X}$  είναι το δεξί ιδιάζον διάνυσμα του  $A$  που αντιστοιχεί στη μικρότερη ιδιάζουσα ιδιοτιμή. Αν η λύση είναι  $\mathbf{X} = (X, Y, Z, W)^T$  τότε το 3Δ σημείο έχει συντεταγμένες  $\tilde{\mathbf{X}} = (\frac{X}{W}, \frac{Y}{W}, \frac{Z}{W})^T$ .

Σύμφωνα με τη μέθοδο που περιγράφηκε παραπάνω, από δύο εικόνες  $I_1, I_2$  με ζεύγη σημείων  $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i, i = 1, \dots, N$  όπου  $N$  το πλήθος των ζευγών, προκύπτουν τα αντίστοιχα 3Δ σημεία  $\tilde{\mathbf{X}}_i$  του χώρου. Αν η βάση δεδομένων αποτελείται από  $M$  το πλήθος εικόνες, τότε ο αλγόριθμος εφαρμόζεται κάθε φορά σε δύο διαδοχικές. Η εφαρμογή του αλγορίθμου στις εικόνες  $I_i$  και  $I_{i+1}, i = 1, \dots, M-1$  τοποθετεί στον χώρο τα 3Δ σημεία  $\mathbf{X}_j^i, j = 1, \dots, N_j$  όπου  $N_j$  το πλήθος από ζεύγη σημείων στις εικόνες. Στο Σχήμα 5.16 φαίνεται το σύνολο των 3Δ σημείων (νέφος σημείων) όπως προέκυψαν από την εφαρμογή του αλγορίθμου σε μια βάση δεδομένων 36 εικόνων. Ο αλγόριθμος εύρεσης σημείων ταιριάσματος δέχεται ως είσοδο τις δύο εικόνες  $I_i$  και  $I_{i+1}$  για κάθε  $i = 1, \dots, 35$ . Στη συνέχεια, τα ζεύγη σημείων που προκύπτουν αποτελούν είσοδο για τον αλγόριθμο υπολογισμού του θεμελιώδους πίνακα. Τα ζεύγη που είναι inliers για την τελική επιλογή του πίνακα  $F$  σύμφωνα με τον RANSAC αλγόριθμο (βλ. Σχήμα 5.15) αποτελούν με τη σειρά τους είσοδο για την 3Δ ανακατασκευή του νέφους σημείων.

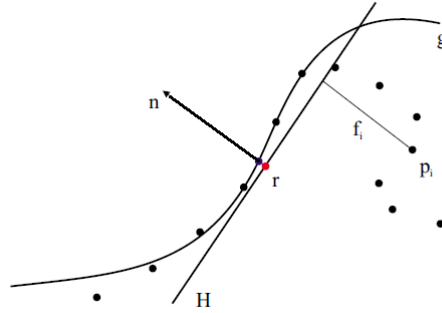
## 5.4 Εκτίμηση Συνεχούς Τρισδιάστατης Επιφάνειας

Η 3Δ ανακατασκευή, όπως παρουσιάστηκε παραπάνω, έχει ως αποτέλεσμα τον υπολογισμό ενός πλήθους σημείων στον 3Δ χώρο με ακανόνιστη κατανομή. Το "σύννεφο σημείων" που αναπαριστά το αντικείμενο δεν μπορεί να δώσει μια ικανοποιητική οπτικά ανακατασκευή. Για το λόγο αυτό, είναι αναγκαία η εκτίμηση μιας συνεχούς επιφάνειας από το σύνολο των σημείων.

Πολλές μέθοδοι έχουν βρεθεί για την επίλυση του προβλήματος αυτού, χρησιμοποιώντας ιδέες από υπολογιστική και διαφορική γεωμετρία (βλ. [34], [35], [20], [4]). Η πιο γνωστή από αυτές είναι η μέθοδος του ball pivoting (βλ. [9]). Οι περισσότερες από τις μεθόδους αυτές απαιτούν συγκεκριμένη προϋπάρχουσα κατανομή σημείων στο χώρο για την εκτίμηση της επιφάνειας και για το λόγο αυτό η εφαρμογή τους γίνεται σε 3Δ δεδομένα που έχουν προκύψει από 3D scanners.

Καταλληλότερη μέθοδος για την αναπαράσταση μιας ομαλής επιφάνεια από 3Δ σημεία στο χώρο με ακανόνιστη κατανομή είναι αυτή των κινούμενων ελαχίστων τετραγώνων MLS (βλ. [1] και [2]). Η μέθοδος αυτή βασίζεται στην ιδέα ότι ένα σημείο από το "σύννεφο σημείων" ορίζει μια επιφάνεια.

Συγκεκριμένα, έστω  $S_p = \{\mathbf{p}_i\}_{i=1}^{M_p}$  το σύνολο των 3Δ σημείων που αποτελούν το νέφος σημείων. Για κάθε σημείο  $\mathbf{r}$  του συνόλου  $S_p$  και για μια περιοχή γύρω από το σημείο αυτό, επιδιώκεται η εκτίμηση του κατάλληλου πολυωνύμου  $N$ -οστού βαθμού που καλύτερα περιγράφει την εν



**Σχήμα 5.13:** Η διαδικασία moving least squares. Αν  $\mathbf{r}$  είναι το σημείο αναφοράς, προσδιορίζεται ένα επίπεδο  $H$  που διέρχεται από το σημείο αυτό με τέτοια κατεύθυνση ώστε να ελαχιστοποιείται η απόσταση των γειτονικών σημείων από αυτό. Στη συνέχεια, εκτιμάται το πολυώνυμο  $N$ -βαθμού που καλύτερα περιγράφει την κατανομή των γειτονικών σημείων του σημείου αναφοράς.

λόγω κατανομή. Στην εργασία χρησιμοποιήθηκε βαθμός  $N = 4$ . Στο Σχήμα 5.13 φαίνεται η διαδικασία

Για την υλοποίηση της μεθόδου, για κάθε σημείο  $\mathbf{r}$  υπολογίζεται ένα επίπεδο αναφοράς  $H$  το οποίο διέρχεται από το σημείο αυτό και είναι τέτοιο ώστε να ελαχιστοποιείται η ευκλείδεια νόρμα των αποστάσεων των γειτονικών σημείων από το επίπεδο (βλ. Σχήμα 5.13). Αν  $\mathbf{n}$  το (άγνωστο) κάθετο διάνυσμα στο επίπεδο, τότε ζητούμενη είναι η ελαχιστοποίηση του συναρτησιακού

$$\sum_{i=1}^N \langle \mathbf{n}, \mathbf{p}_i - \mathbf{r} \rangle^2 \cdot \theta(\|\mathbf{p}_i - \mathbf{r}\|) \quad (5.32)$$

όπου  $\mathbf{p}_i$  είναι τα γειτονικά σημεία ως προς το σημείο αναφοράς  $\mathbf{r}$  και η συνάρτηση  $\theta(\cdot)$  είναι μια ακτινική φθίνουσα συνάρτηση (συνήθως επιλέγεται η γκαουσιανή).

Για την ελαχιστοποίηση της παραπάνω σχέσης αρκεί να μηδενίσουμε την παράγωγο της ως προς  $\mathbf{n}$ , και προκύπτει γραμμική εξίσωση ως προς  $\mathbf{n}$ , δηλαδή

$$\sum_{i=1}^N 2\theta_i \cdot \langle \mathbf{n}, \mathbf{p}_i - \mathbf{r} \rangle \cdot (\mathbf{p}_i - \mathbf{r}) = 0 \quad (5.33)$$

όπου  $\theta_i = \theta(\|\mathbf{p}_i - \mathbf{r}\|)$ .

Στη συνέχεια, για να εκτιμηθούν οι παράμετροι του πολυωνύμου, ελαχιστοποιείται ως προς αυτές το συναρτησιακό

$$\sum_{i=1}^N (g(x_i, y_i) - f_i)^2 \cdot \theta_i \quad (5.34)$$

όπου  $(x_i, y_i)$  είναι οι συντεταγμένες των σημείων προβολής των  $\mathbf{p}_i$  στο επίπεδο  $H$  ως προς ένα τοπικό σύστημα αναφοράς ( με  $z$ -άξονα παράλληλο με το διάνυσμα  $\mathbf{n}$  και αρχή το σημείο  $\mathbf{r}$ ) και  $f_i = \langle \mathbf{n}, \mathbf{p}_i - \mathbf{r} \rangle$ . Η παράγωγος της παραπάνω σχέσης ως προς τις παραμέτρους του πολυωνύμου  $g(x, y)$  καταλήγει σε γραμμικό σύστημα εξισώσεων, η επίλυση του οποίου μας δίνει μια εκτίμηση για το πολυώνυμο.

Έχοντας βρει το πολυώνυμο που καλύτερα απεικονίζει την κατανομή των σημείων στη γειτονιά του εκάστοτε σημείου αναφοράς, είναι δυνατό να εκτιμήσουμε περαιτέρω σημεία δειγματοληπτώντας την συνεχή επιφάνεια που ορίζεται από το πολυώνυμο. Με τον τρόπο αυτό πυκνώνει η

κατανομή και το αντικείμενο προς ανακατασκευή απεικονίζεται με μεγαλύτερο πλήθος σημείων, τα οποία τοπικά προσεγγίζουν μια συνεχή επιφάνεια.

---

**Algorithm 5** Αλγόριθμος MLS: Εκτίμηση συνεχούς επιφάνειας από νέφος σημείων

---

Για κάθε σημείο  $\mathbf{r} \in S_p$

- (i). **Επίπεδο αναφοράς:** Υπολογισμός επιπέδου αναφοράς  $H$  που διέρχεται από το σημείο  $\mathbf{r}$  και με κατεύθυνση  $\mathbf{n}$  που δίνεται από τη σχέση 5.33
  - (ii). Υπολογισμός των συντελεστών του πολυωνύμου  $g(x, y)$  που ικανοποιεί τη σχέση 5.34 για τα σημεία που απέχουν απόσταση λιγότερη από  $r$  από το σημείο αναφοράς  $\mathbf{r}$ .
  - (iii). **Δειγματοληψία:** Στο τοπικό σύστημα συντεταγμένων με κέντρο το σημείο  $\mathbf{r}$  και z-άξονα παράλληλο του διανύσματος  $\mathbf{n}$  δειγματοληπτείται η συνάρτηση του πολυωνύμου στην περιοχή γύρω από το σημείο αναφοράς και σε απόσταση  $r$  από αυτό.
- 

## 5.5 Τρισδιάστατη Ανακατασκευή Ακμών

Μια επιπλέον ενδιαφέρουσα εφαρμογή είναι η 3Δ ανακατασκευή ακμών. Συγκεκριμένα, σκοπός είναι να απεικονιστούν οι ακμές στον 3Δ χώρο ώστε να δώσουν μια πιο ακριβής πληροφόρηση αναφορικά με το σχήμα και τις λεπτομέρειες του αντικειμένου. Η ανακατασκευή αυτή γίνεται χωρίς καμία a priori πληροφόρηση για τη γεωμετρία του αντικειμένου και ως δεδομένα εισόδου χρησιμοποιούνται μόνο οι εικόνες της βάσης δεδομένων.

Η ανίχνευση των ακμών του αντικειμένου γίνεται με χρήση των Primal Sketch Tokens, Kokkinos et al. (βλ. [23]). Αυτός ο αλγόριθμος ανιχνεύει σημεία και χαρακτηριστικά ενδιαφέροντος σε 2Δ εικόνες. Στο [23] αναφέρονται λεπτομέρειες για τον αλγόριθμο αυτό.

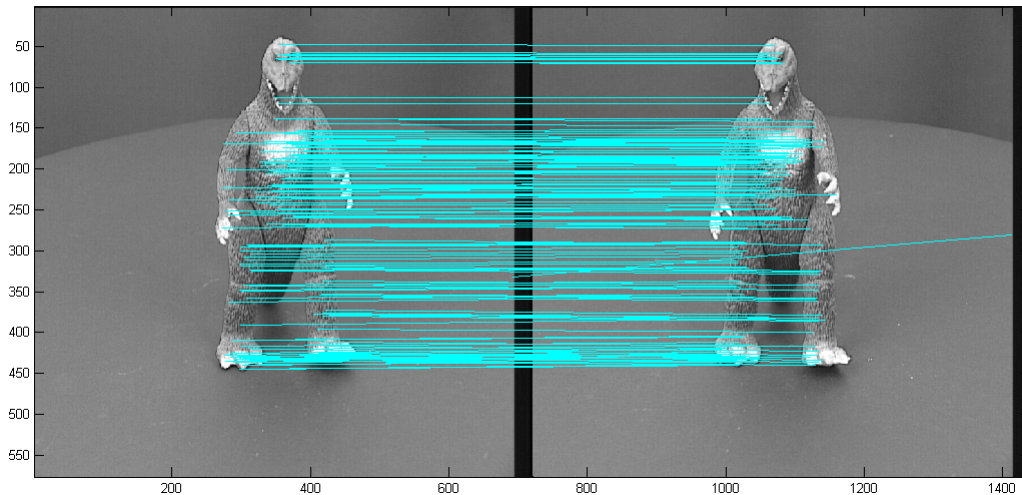
Αφού ανιχνευθούν οι ακμές στην εικόνα  $I_i$  της βάσης δεδομένων, για τα σημεία που αποτελούν τις ακμές επιχειρείται η εύρεση των πιθανών αντίστοιχων σημείων τους στην εικόνα  $I_{i+1}$  της βάσης. Συγκεκριμένα, αν  $\mathbf{x} \in Edge^{(j)}$  όπου  $Edge^{(j)}$  η  $j$ -στη ακμή της εικόνας αναφοράς, σκοπός είναι να βρεθεί το αντίστοιχο σημείο  $\mathbf{x}'$  της δεύτερης εικόνας έτσι ώστε  $\mathbf{x} \leftrightarrow \mathbf{x}'$ .

Σύμφωνα με το κεφάλαιο 3, το σημείο  $\mathbf{x}' \in \mathcal{V}$ , όπου  $\mathcal{V} = F\mathbf{x}$  η επιπολική ευθεία στη δεύτερη εικόνα. Για τον προσδιορισμό του σημείου, εξετάζεται κάθε σημείο της ευθείας  $\mathcal{V}$ . Συγκεκριμένα, για κάθε σημείο  $\mathbf{x}'_k$ ,  $k = 1, \dots, N_{\mathcal{V}}$  υπολογίζεται η κανονικοποιημένη αλληλό-συσχέτιση μεταξύ των δύο περιοχών των εικόνων με κέντρα τα σημεία  $\mathbf{x}$  και  $\mathbf{x}'_k$  σύμφωνα με τον τύπο

$$NCC[\mathbf{x}, \mathbf{x}'_k](u, v) = \frac{\sum_{(x,y)} (T_i(x, y) - \bar{T}_i)(T_{i+1}(x - u, y - v) - \bar{T}_{i+1})}{\sqrt{\sum_{(x,y)} (T_i(x, y) - \bar{T}_i)^2 \sum_{(x,y)} (T_{i+1}(x - u, y - v) - \bar{T}_{i+1})^2}} \quad (5.35)$$

όπου  $T_i, T_{i+1}$  είναι οι περιοχές των εικόνων  $I_i, I_{i+1}$  γύρω από τα σημεία  $\mathbf{x}, \mathbf{x}'_k$  αντίστοιχα, και  $\bar{T}_i, \bar{T}_{i+1}$  η μέση τιμή τους. Η διάσταση των περιοχών καθορίζεται από το μέγεθος του παραθύρου  $W$ . Το διάστημα τιμών των  $(u, v)$  καθορίζεται από το μέγεθος του παραθύρου αυτού.

Το σημείο  $\mathbf{x}'_k$  της δεύτερης εικόνας θεωρείται matchpoint του  $\mathbf{x}$  αν  $\max(u, v)NCC[\mathbf{x}, \mathbf{x}'_k](u, v) > NCC_{thres}$ . Αν δεν βρεθεί σημείο πάνω στην επιπολική ευθεία που να ικανοποιεί τη σχέση αυτή, τότε θεωρείται ότι το σημείο της ακμής στην πρώτη εικόνα δεν έχει αντίστοιχο σημείο στη δεύτερη. Αν περισσότερα από τα μισά σημεία από τα αποτελείται η ακμή δεν μπορεί να αντιστοιχηθεί με κάποιο σημείο της αντίστοιχης επιπολικής ευθείας της δεύτερης εικόνα τότε θεωρείται ότι η ακμή δεν αποτελεί ακμή και στη δεύτερη εικόνα. Σε αντίθετη περίπτωση, από τα ζεύγη σημείων



**Σχήμα 5.14:** Οι γραμμές αντιστοιχίζουν ζεύγη SIFT σημείων ταιριάσματος μεταξύ των εικόνων (a)  $I_{11}$  και (b)  $I_{12}$  της βάσης δεδομένων.

$\mathbf{x} \leftrightarrow \mathbf{x}'$  γίνεται εκτίμηση της 3Δ θέσης του σημείου στο χώρο και επόμενος της θέσης της 3Δ ακμής.

## 5.6 Τελικά Αποτελέσματα

Στην ενότητα αυτή, παρουσιάζονται τα αποτελέσματα των αλγορίθμων που παρουσιάστηκαν στο κεφάλαιο αυτό. Η βάση δεδομένων αποτελείται από 36 εικόνες. Ένα δείγμα από τις εικόνες αυτές φαίνεται στο Σχήμα 3.1.

Στο Σχήμα 5.14 φαίνεται το αποτέλεσμα του αλγορίθμου SIFT για τον προσδιορισμό σημείων ταιριάσματος. Τα ζεύγη σημείων των δύο εικόνων ενώνονται με τις γραμμές που φαίνονται στο σχήμα.

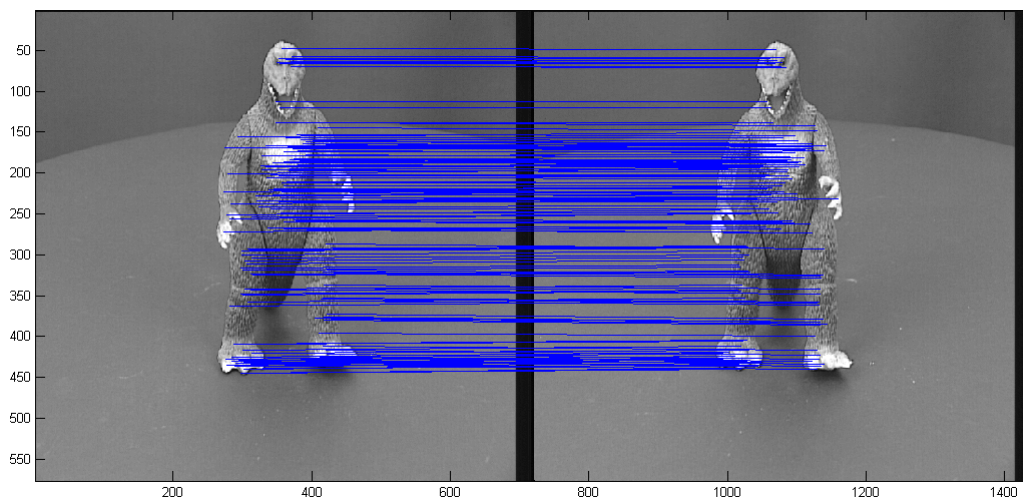
Στο Σχήμα 5.15 φαίνονται τα ζεύγη σημείων που αποτελούν inliers του τελικού θεμελιώδους πίνακα όπως περιγράφηκε στην ενότητα 5.2. Τα ζεύγη σημείων ενώνονται με τις γραμμές που φαίνονται στο σχήμα.

Συγκρίνοντας τα Σχήματα 5.14 και 5.15, είναι εμφανές ότι κάποια αρχικά ζεύγη σημείων όπως τα υπολόγισε ο αλγόριθμος SIFT δεν ικανοποιούν τον επιπολικό περιορισμό με κάποια δεδομένη ακρίβεια. Τα ζεύγη αυτά έχουν εξαλειφθεί από το Σχήμα 5.15. Αυτός είναι ένας τρόπος να αντισταθμιστεί η μη ικανοποιητική απόδοση του αλγορίθμου ανίχνευσης σημείων ενδιαφέροντος. Με άλλα λόγια, ακόμα και αν αρχικά εντοπιστούν σημεία τα οποία δεν είναι σημεία ενδιαφέροντος, στη συνέχεια καθώς είναι πολύ απίθανο τέτοια σημεία να βρουν το ταίρι τους σε άλλες εικόνες, εξαλείφονται και δεν χρησιμοποιούνται. Δηλαδή, την έλλειψη προσεκτικής επιλογής σημείων ενδιαφέροντος την αντισταθμίζει ο αλγόριθμος RANSAC.

Στο Σχήμα 5.16 φαίνεται το 3Δ νέφος σημείων με τον τρόπο που περιγράφηκε στην ενότητα 5.4. Το σύνολο των σημείων αυτών φαίνεται από έξι διαφορετικές θέσεις στο χώρο για την καλύτερη παρατήρηση.

Από το 3Δ νέφος σημείων ζητείται η εκτίμηση μια ομαλής επιφάνειας που καλύτερα περιγράφει τη συγκεκριμένη κατανομή. Αυτό γίνεται με τη μέθοδο των κινούμενων ελαχίστων τετραγώνων που περιγράφηκε στην ενότητα ... . Στο Σχήμα 5.17 φαίνεται το αποτέλεσμα της μεθόδου από πολλές οπτικές γωνίες για καλύτερη παρατήρηση.

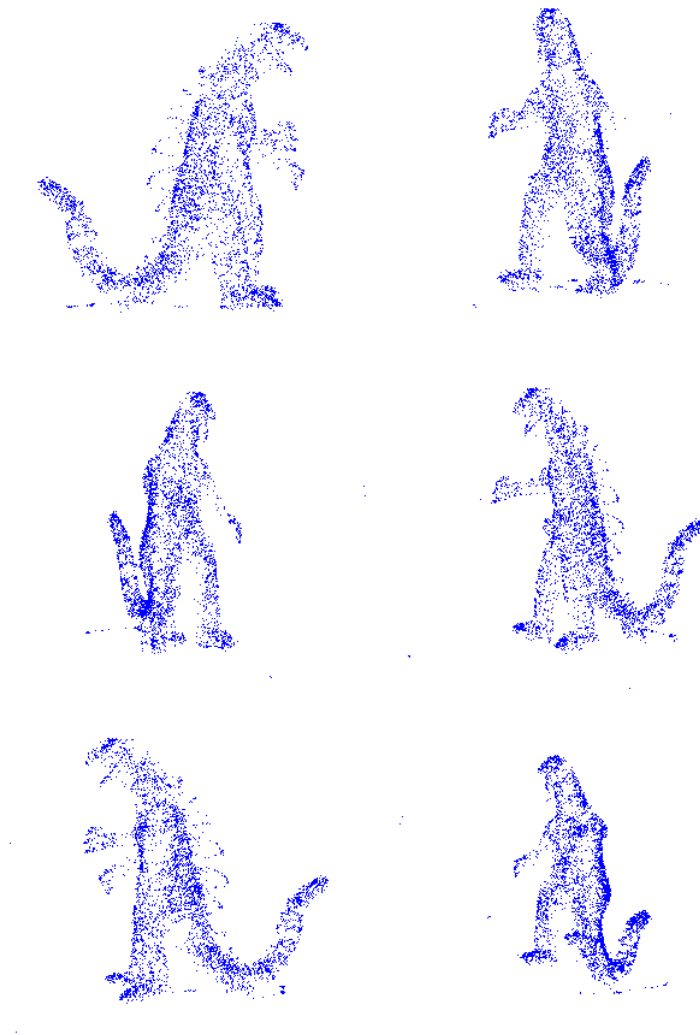
Για την 3Δ ανακατασκευή των ακμών χρησιμοποιούνται 6 εικόνες από τη βάση δεδομένων.



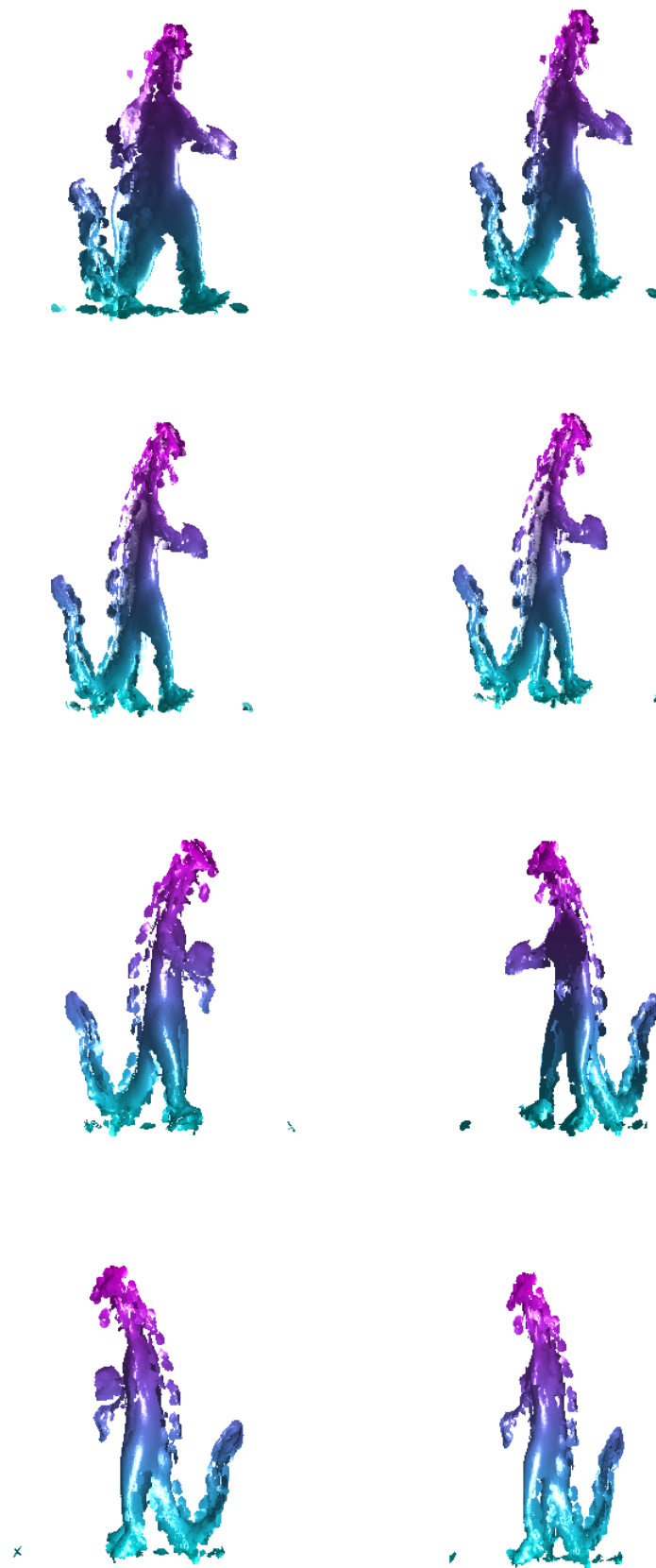
**Σχήμα 5.15:** Οι γραμμές αντιστοιχίζουν ζεύγη σημείων των εικόνων  $I_{11}$  και  $I_{12}$  που είναι inliers σύμφωνα με τον αλγόριθμο RANSAC. Τα ζεύγη αυτά αποτελούν υποσύνολο των ζευγών που υπολόγισε ο SIFT αλγόριθμος.

Για καλύτερα αποτελέσματα θα μπορούσαν να χρησιμοποιηθούν όλες οι διαθέσιμες εικόνες της βάσης. Ωστόσο κάτι τέτοιο δεν είναι απαραίτητο καθώς το αποτέλεσμα θα βελτιωνόταν ελάχιστα αφού οι ακμές επαναλαμβάνονται στις διαδοχικές εικόνες της βάσης. Για την υλοποίηση, επιλέχθηκε το εμβαδόν της υπό εξέταση περιοχής να αποτελεί το 0.4% της συνολικής εικόνας και  $NCC_{thres} = 0.6$ . Στο Σχήμα 5.18(a) φαίνεται το αποτέλεσμα της εφαρμογής των Primal Sketch Tokens στην εικόνα ενώ στο 5.18(b) η αντιστοίχιση των ακμών με τον αλγόριθμο που περιγράφηκε στην αντίστοιχη ενότητα του κεφαλαίου.

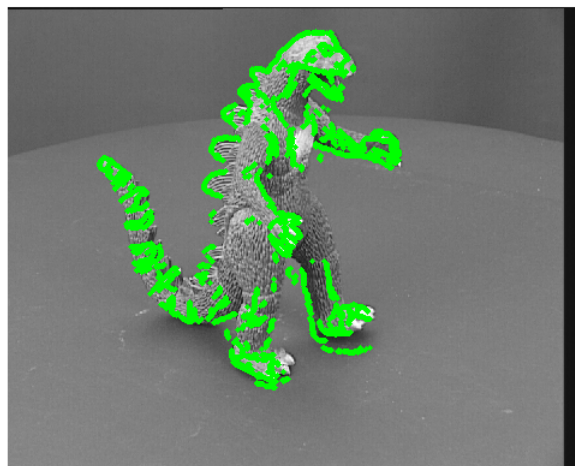
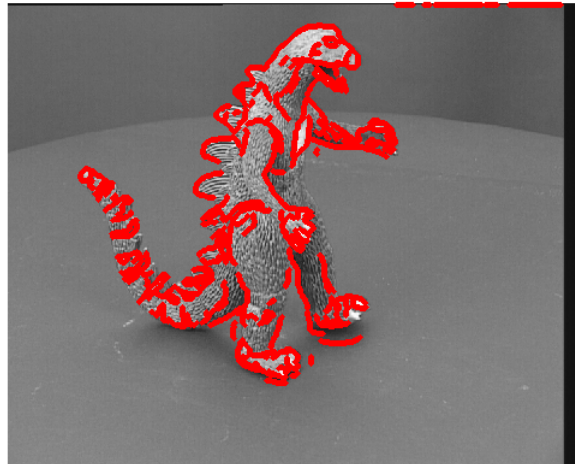
Στο Σχήμα 5.19 φαίνεται το αποτέλεσμα της 3Δ ανακατασκευής των ακμών από τις 6 εικόνες της βάσης δεδομένων. Το αποτέλεσμα παρουσιάζεται από πολλές όψεις για την καλύτερη οπτική παρατήρηση του.



**Σχήμα 5.16:** Το 3D νέφος σημείων. Τα 3D σημεία όπως υπολογίστηκαν από τη βάση δεδομένων των 36 εικόνων με τον SIFT αλγόριθμο ανίχνευσης κοινών σημείων έχουν τοποθετηθεί στο χώρο. Στις εικόνες του σχήματος φαίνεται η κατανομή των σημείων από διάφορες θέσεις του χώρου.

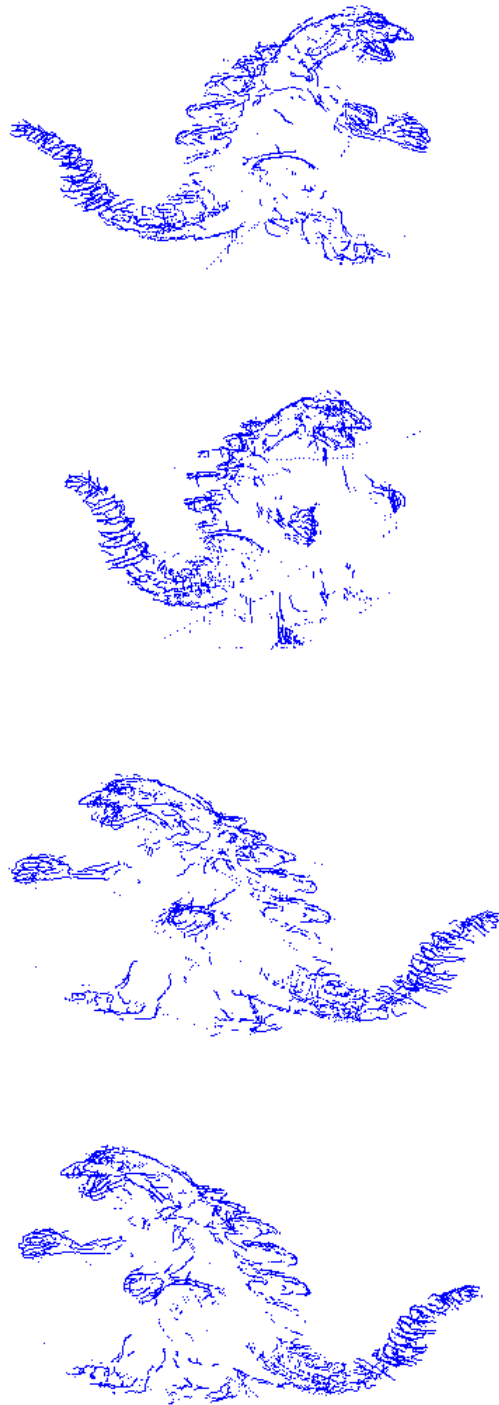


**Σχήμα 5.17:** Το αποτέλεσμα της μεθόδου MLS (moving least squares). Το νέφος των διάσπαρτων σημείων έχει δώσει τη θέση του σε μια συνεχή επιφάνεια που εκτιμά την αρχική κατανομή των σημείων. Το πολυώνυμο που χρησιμοποιήθηκε τοπικά για τον υπολογισμό της επιφάνειας είναι 4ου βαθμού.



**Σχήμα 5.18:** (a) Το αποτέλεσμα της ανίχνευσης ακμών με τον αλγόριθμο των Primal Sketch Tokens, (b) Η αντιστοίχιση των ακμών που βρέθηκαν από την πρώτη εικόνα στη δεύτερη εικόνα





**Σχήμα 5.19:** Το αποτέλεσμα της 3Δ ανακατασκευής ακμών με είσοδο 6 εικόνες της βάσης δεδομένων. Οι ακμές στις 2Δ εικόνες προέρχονται από τα Primal Sketch Tokens (βλ. [23]).



## Κεφάλαιο 6

# Τρισδιάστατη Ανακατασκευή Κτηρίων

Στο κεφάλαιο αυτό, μελετάται η 3Δ ανακατασκευή κτηρίων. Σκοπός είναι η 3Δ αναπαράσταση από φωτογραφίες που απεικονίζουν το κτήριο από δύο μόνο όψεις. Για οπτικά καλύτερη αναπαράσταση, οι δύο εικόνες χαρακτηρίζονται από απομακρυσμένα κέντρα κάμερας. Επιπλέον, επιθυμητό είναι να δίνεται η δυνατότητα στο χρήστη να περιηγηθεί στα δεδομένα με ταυτόχρονη την απεικόνιση της τρισδιάστατης ανακατασκευής.

Πολλές μέθοδοι έχουν προταθεί για την 3Δ ανακατασκευή κτηρίων και αρχιτεκτονικών κτισμάτων (βλ. [?], [36], [50], [49], [7] και [6]). Η πιο πρόσφατη εφαρμογή με στόχο την 3Δ ανακατασκευή εξωτερικών σκηνών από εικόνες και την περιήγηση στις διάφορες εικόνες είναι αυτή του PhotoTourism (βλ. [43]).

### 6.1 Τρισδιάστατη Ανακατασκευή Κτηρίων με τη Μέθοδο της Ανίχνευσης Επιπέδων

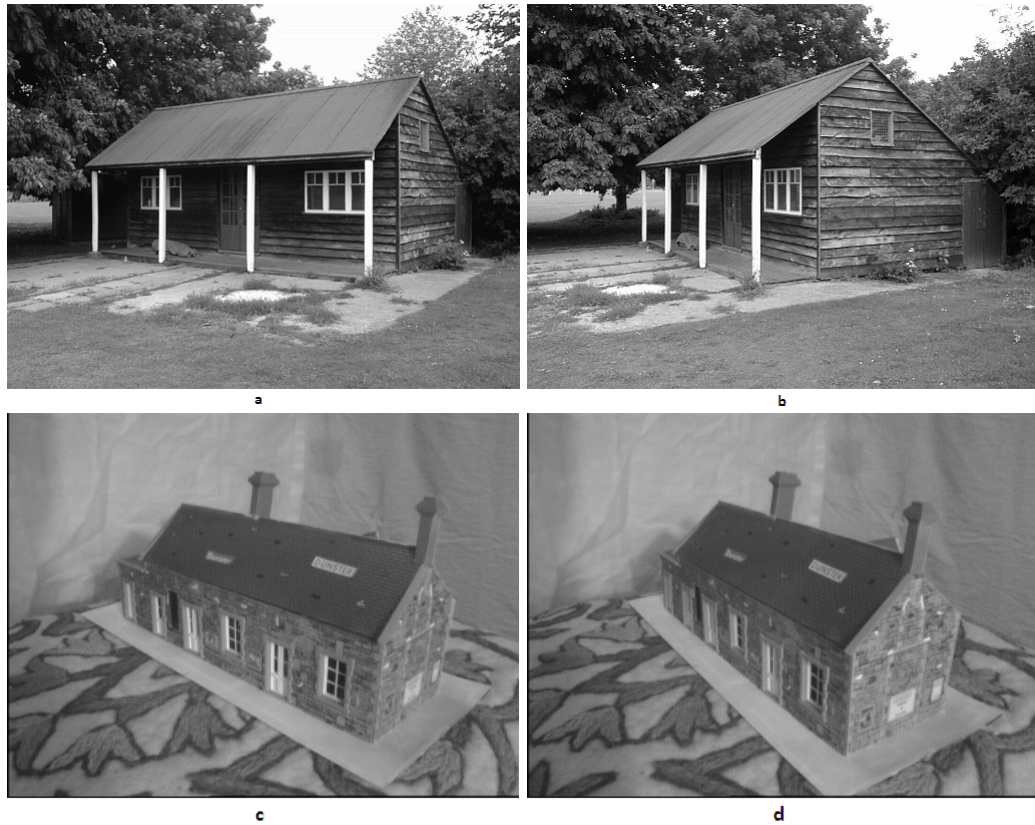
Τα δεδομένα που απαιτούνται για την 3Δ αναπαράσταση του κτηρίου είναι δύο εικόνες, που έχουν ληφθεί από διαφορετικές θέσεις και οπτικές γωνίες του κτηρίου. Για τις εικόνες, είναι γνωστές οι εσωτερικές παράμετροι (παράμετροι βαθμονόμησης) των καμερών. Με άλλα λόγια είναι γνωστοί οι πίνακες  $K_i$  με  $i = 1, 2$  των δύο καμερών. Η θέση  $\mathbf{t}_i$  και ο προσανατολισμός  $R_i$  τους θα εκτιμηθούν από τον σημαντικό πίνακα  $E$  που αντιστοιχεί στις δύο εικόνες.

Στο Σχήμα 6.1 φαίνονται τα δεδομένα που απαιτούνται για την 3Δ ανακατασκευή των κτηρίων. Στο σχήματα (a) και (b) φαίνονται οι δύο όψεις ενός κτηρίου ενώ στις (c) και (d) οι δύο όψεις του δεύτερου κτηρίου. Με βάση της εικόνες αυτές θα γίνει η 3Δ αναπαράσταση των κτηρίων αυτών.

Ο αλγόριθμος, όπως περιγράφεται στις παρακάτω ενότητες, παρουσιάζει πολλές ομοιότητες με τον αλγόριθμο του κεφαλαίου 5. Ωστόσο, η 3Δ ανακατασκευή στηρίζεται στη γεωμετρία των κτηρίων. Συγκεκριμένα, η ύπαρξη επιπέδων στο χώρο παίζει καθοριστικό ρόλο στην 3Δ αναπαράσταση κτηρίων. Τα επίπεδα αυτά αποτελούν τις προσόψεις του κτηρίου και οδηγούν σε μια οπτικά ικανοποιητική ανακατασκευή, καθώς γίνεται δυνατή η απόδοση υφής στην ανακατασκευή.

#### 6.1.1 Υπολογισμός Θεμελιώδους Πίνακα

Με τρόπο όμοιο με αυτόν που περιγράφηκε στην ενότητα 5.1.1. ανιχνεύονται ζεύγη σημείων στις δύο εικόνες. Όταν οι δύο κάμερες έχουν μεγάλη σε μήκος βασική ευθεία (baseline) τότε οι αλγόριθμοι εξαγωγής χαρακτηριστικών, π.χ. SIFT, δεν μπορούν να δώσουν ικανό πλήθος ζευγών σημείων για 3Δ ανακατασκευή. Στις περιπτώσεις αυτές, συνηθίζεται να δίνονται χειροκίνητα τα ζεύγη αυτά.



**Σχήμα 6.1:** Τα δεδομένα για την 3Δ ανακατασκευή κτηρίων(πηγή: [www.robots.ox.ac.uk/vgg/data/data-view.html](http://www.robots.ox.ac.uk/vgg/data/data-view.html)). Στις εικόνες (a) και (b) καθώς και στις (c) και (d) φαίνονται οι εικόνες που θα χρησιμοποιηθούν για την αναπαράσταση των δύο κτηρίων. Επιπλέον, είναι γνωστοί οι πίνακες βαθμονόμησης των αντίστοιχων καμερών.

Με δεδομένα τα ζεύγη σημείων  $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$  με  $i = 1, \dots, N$  μεταξύ των δύο εικόνων, υπολογίζεται ο θεμελιώδης πίνακας  $F$ , σύμφωνα με τον αλγόριθμο που παρουσιάστηκε αναλυτικά στο κεφάλαιο 5. Για να δώσει ο αλγόριθμος μια καλή εκτίμηση του θεμελιώδους πίνακα, πρέπει ο αριθμός από ζεύγη σημείων  $N$  να είναι όσο το δυνατό μεγαλύτερος και σε καμία περίπτωση μικρότερος από 8.

### 6.1.2 Εκτίμηση των Εξωτερικών Παραμέτρων της Κάμερας

Από το θεμελιώδη πίνακα  $F$  και τους πίνακες βαθμονόμησης των καμερών, ο σημαντικός πίνακας δίνεται από τη σχέση

$$E = K_2^T F K_1$$

Σκοπός είναι να εκτιμηθούν οι προβολικοί πίνακες για τις κάμερες. Για το ζεύγος εικόνων, θεωρείται ότι οι *κανονικοποιημένοι* (δηλαδή μη βαθμονομημένοι) προβολικοί πίνακες δίνονται από τις σχέσεις

$$P_{1,cam} = [I|\mathbf{0}] \quad P_{2,cam} = [R|\mathbf{t}] \quad (6.1)$$

Με άλλα λόγια, θεωρείται ότι το παγκόσμιο σύστημα συντεταγμένων ταυτίζεται με το τοπικό σύστημα συντεταγμένων της πρώτης κάμερας. Από τα παραπάνω και από τη θεωρία του κεφαλαίου 3, ο σημαντικός πίνακας δίνεται από τη σχέση

$$E = [\mathbf{t}]_{\times} R \quad (6.2)$$

Όπως αποδείχθηκε στο κεφάλαιο 3, η διεύθυνση της μετατόπισης  $\hat{\mathbf{t}}$  μπορεί να εκτιμηθεί, συγκεκριμένα μια εκτίμηση του προκύπτει από την SVD ανάλυση του πίνακα  $E$  όπου στο διάνυσμα  $\hat{\mathbf{t}}$  αποδίδεται το αριστερό ιδιάζον διάνυσμα του πίνακα  $E$  που αντιστοιχεί στην μικρότερη ιδιάζουσα ιδιοτιμή.

Από την SVD ανάλυση του πίνακα  $E$  μπορεί να εκτιμηθεί και ο πίνακας προσανατολισμού  $R$  της δεύτερης κάμερας. Συγκεκριμένα, αν η SVD ανάλυση του  $E = U\Sigma V^T$ , τότε προκύπτουν, σύμφωνα με τη θεωρία του κεφαλαίου 3, υποψήφιοι πίνακες προσανατολισμού που δίνονται από τη σχέση

$$R = \pm U R_{\pm 90^\circ}^T V^T \quad (6.3)$$

Εύκολα προκύπτει ότι όλοι οι παραπάνω πίνακες είναι ορθογώνιοι, αφού οι πίνακες  $U, V$  είναι ορθογώνιοι πίνακες.

Από αυτούς, επειδή ο πίνακας  $R$  της κάμερας πρέπει να είναι πίνακας περιστροφής (rotation matrix), υπολογίζονται οι πίνακες που ικανοποιούν επιπλέον τη σχέση  $\det(R) = 1$ . Έστω  $R_1$  και  $R_2$  οι πίνακες αυτοί (μόνο δύο από τους τέσσερις υποψήφιους πίνακες της σχέσης 6.3 είναι δυνατό να είναι πίνακες περιστροφής).

Συνδυάζοντας τους πίνακες αυτούς με τις δυνατές διευθύνσεις της μετατόπισης προκύπτουν τέσσερις διαφορετικές τιμές του *κανονικοποιημένου* προβολικού πίνακα της δεύτερης κάμερας. Συγκεκριμένα, προκύπτουν οι υποψήφιοι πίνακες

$$\begin{aligned} P_{2,cam}^{(a)} &= [R_1|\hat{\mathbf{t}}] & P_{2,cam}^{(b)} &= [R_1|-\hat{\mathbf{t}}] \\ P_{2,cam}^{(c)} &= [R_2|\hat{\mathbf{t}}] & P_{2,cam}^{(d)} &= [R_2|-\hat{\mathbf{t}}] \end{aligned} \quad (6.4)$$

Σκοπός είναι να βρεθεί ο προβολικός πίνακας της δεύτερης κάμερας, έτσι ώστε τα περισσότερα 3D σημεία να ανήκουν στο μπροστινό μέρος των δύο καμερών. Για να επιτευχθεί αυτό εξετάζονται όλα τα δυνατά ζεύγη. Συγκεκριμένα, έστω οι προβολικοί πίνακες για τις κάμερες  $P_{1,cam}, P_{2,cam}^{(j)}$

με  $j = a, b, c, d$  και έστω  $\mathbf{x}_1 \leftrightarrow \mathbf{x}_2$  είναι ένα ζεύγος σημείων εκ των συνολικών  $N$  στο πλήθος στις δύο εικόνες. Τα σημεία αυτά κανονικοποιούνται σύμφωνα με τις παρακάτω σχέσεις

$$\begin{aligned} \mathbf{x}_{1,cam} &= K_1^{-1} \mathbf{x}_1 \\ \mathbf{x}_{2,cam} &= K_2^{-1} \mathbf{x}_2 \end{aligned} \quad (6.5)$$

Αν  $\mathbf{X}$  είναι το σημείο του 3Δ χώρου το οποίο απεικονίζεται στα σημεία  $\mathbf{x}_1$  και  $\mathbf{x}_2$  στις δύο εικόνες αντίστοιχα, τότε ισχύει

$$\begin{aligned} \mathbf{x}_{1,cam} &= P_{1,cam} \mathbf{X} \\ \mathbf{x}_{2,cam} &= P_{2,cam}^{(j)} \mathbf{X} \end{aligned} \quad (6.6)$$

Με παρόμοια διαδικασία με αυτή που αναφέρθηκε στο κεφάλαιο 5, η εύρεση του σημείου  $\mathbf{X}$  προκύπτει από την επίλυση του συστήματος  $A_{cam} \mathbf{X} = 0$  όπου ο πίνακας  $A_{cam}$  δίνεται από τη σχέση

$$A_{cam} = \begin{bmatrix} x_{1,cam} \mathbf{P}_{1,cam}^{3T} - \mathbf{P}_{1,cam}^{1T} \\ y_{1,cam} \mathbf{P}_{1,cam}^{3T} - \mathbf{P}_{1,cam}^{2T} \\ x_{2,cam} \mathbf{P}_{2,cam}^{3T} - \mathbf{P}_{2,cam}^{1T} \\ y_{2,cam} \mathbf{P}_{2,cam}^{3T} - \mathbf{P}_{2,cam}^{2T} \end{bmatrix}$$

όπου  $\mathbf{x}_{1,cam} = w_1(x_{1,cam}, y_{1,cam}, 1)^T$  και  $\mathbf{x}_{2,cam} = w_2(x_{2,cam}, y_{2,cam}, 1)^T$ .

Μετά τον υπολογισμό του του 3Δ σημείου  $\mathbf{X}$ , εξετάζεται αν ανήκει στο μπροστινό μέρος της κάθε κάμερας. Αυτό γίνεται με τον υπολογισμό του βάθους του σημείου από την κάμερα σύμφωνα με το Κεφαλαίο 2. Συγκεκριμένα, υπολογίζονται τα βάθη από τις δύο κάμερες

$$\begin{aligned} depth_1(\mathbf{X}, P_{1,cam}) &= \frac{w_1}{T} \\ depth_2(\mathbf{X}, P_{2,cam}) &= \frac{\det(R)w_2}{T \|r_3\|} \end{aligned} \quad (6.7)$$

Το σημείο  $\mathbf{X}$  ανήκει στο μπροστινό μέρος των δύο καμερών αν  $depth_1 \geq 0$  και  $depth_2 \geq 0$ .

Η διαδικασία αυτή εφαρμόζεται για όλα τα  $N$  ζεύγη σημείων και τελικά επιλέγεται εκείνος ο συνδυασμός κανονικοποιημένων προβολικών πινάκων που δίνουν το μεγαλύτερο πλήθος εμπρόσθιων σημείων. Με τον τρόπο αυτό προκύπτουν οι τελικοί προβολικοί πίνακες, ως  $P_1 = K_1 P_{1,cam}$  και  $P_2 = K_2 P_{2,cam}$ , όπου  $P_{1,cam}, P_{2,cam}$  είναι οι πίνακες που επιλέχθησαν παραπάνω.

### 6.1.3 Ανακατασκευή Τρισδιάστατου Νέφους Σημείων

Έχοντας εκτιμήσει τους πίνακες  $P_1$  και  $P_2$  των αντίστοιχων καμερών και γνωρίζοντας τα ζεύγη σημείων  $\mathbf{x}_1 \leftrightarrow \mathbf{x}_2$  των δύο εικόνων αντίστοιχα, είναι δυνατή η εύρεση των 3Δ σημείων που αντιστοιχούν στα ζεύγη σημείων.

Η μέθοδος για την εύρεση των σημείων αυτών είναι ίδια με αυτή της ενότητας 5.3, όπου εκτιμάται το 3Δ σημείο  $\mathbf{X}$  με γνώση των προβολικών πινάκων δύο καμερών και τις συντεταγμένες των σημείων όπου απεικονίζεται το 3Δ σημείο στις δύο εικόνες.

### 6.1.4 Ανίχνευση Επιπέδων

Από το 3Δ νέφος σημείων επιδιώκεται η ανίχνευση διάφορων επιπέδων, όπως αυτά ορίζονται από τα σημεία στο 3Δ χώρο. Κάποιοι τρόποι που έχουν προταθεί για το σκοπό αυτό περιγράφονται στα [47], [26], [51]. Για την ανίχνευση των επιπέδων στην παρούσα εργασία εφαρμόζεται ο 3Δ Hough μετασχηματισμός (επέκταση του 2Δ μετασχηματισμού [13]).

Ένα επίπεδο  $\Pi$  στον 3Δ χώρο χαρακτηρίζεται από την τριπλέτα  $(s, \theta, \phi)$  όπου  $s \geq 0$  δηλώνει την απόσταση του επιπέδου από την αρχή του παγκόσμιου συστήματος συντεταγμένων  $OXYZ$ ,  $\theta \in [0, 2\pi)$  και  $\phi \in [-\frac{\pi}{2}, \frac{\pi}{2})$  δηλώνουν τις δύο γωνίες που σχετίζονται με την αναπαράσταση σε σφαιρικές συντεταγμένες του μοναδιαίου κάθετου διανύσματος  $\mathbf{n} = (n_x, n_y, n_z)$  στο επίπεδο  $\Pi$ . Οι τρεις συνιστώσες του  $\mathbf{n}$  γράφονται

$$n_x = \cos(\phi) \cdot \cos(\theta) \quad n_y = \cos(\phi) \cdot \sin(\theta) \quad n_z = \sin(\phi) \quad (6.8)$$

Δειγματοληπτώντας ομοιόμορφα τις παραμέτρους  $(s, \theta, \phi)$  έχουμε αντίστοιχα  $N_s, N_\theta, N_\phi$  στοιχεία ως εξής

- $\Xi = \{s_k = k \cdot \Delta_s\}$  όπου  $\Delta_s = S_{max}/N_s, k \in \{0, \dots, N_s - 1\}$  και  $S_{max}$  είναι η μέγιστη απόσταση (επιλέγεται επαρκώς μεγάλη ώστε να περιλαμβάνει όλα τα δυνατά επίπεδα)
- $\Theta = \{\theta_k = k \cdot \Delta_\theta\}$  όπου  $\Delta_\theta = 2\pi/N_\theta$  και  $k \in \{0, \dots, N_\theta - 1\}$
- $\Psi = \{\phi = (k - N_\phi/2) \cdot \Delta_\phi\}$  όπου  $\Delta_\phi = \pi/N_\phi$  και  $k \in \{0, \dots, N_\phi - 1\}$

Ο 3Δ μετασχηματισμός Hough είναι μια διαδικασία από το χώρο παραμέτρων στους πραγματικούς αριθμούς, δηλαδή  $h : (\Xi \times \Theta \times \Psi) \rightarrow R$  και περιγράφεται παρακάτω.

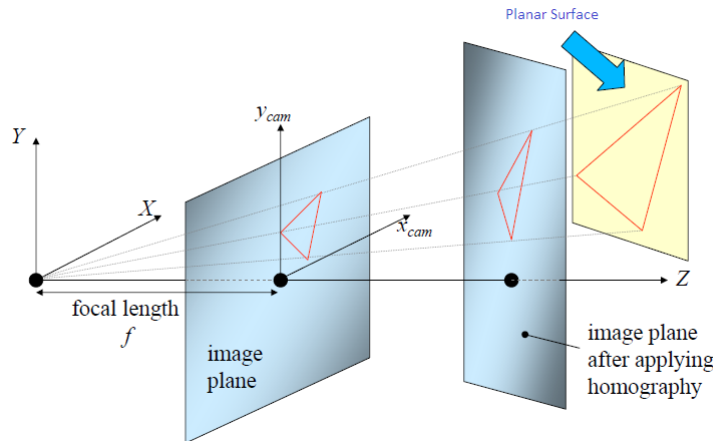
Αρχικά, η  $h$  αρχικοποιείται στο μηδέν. Για κάθε σημείο  $\mathbf{p}$  στο σύνολο των 3Δ σημείων, τα επίπεδα που διέρχονται από το  $\mathbf{p}$  με προσανατολισμό  $(\theta_j, \phi_k)$  κατασκευάζονται. Η απόσταση του επιπέδου από την αρχή των αξόνων είναι

$$s_{jk}^p = n_x(\theta_j, \phi_k) \cdot x_p + n_y(\theta_j, \phi_k) \cdot y_p + n_z(\theta_j, \phi_k) \cdot z_p$$

. Η απόλυτη τιμή της απόστασης αυτής αποδίδεται στην κοντινότερη τιμή από το σύνολο  $\Xi$  και στη συνέχεια η αντίστοιχη τιμή του αθροιστή  $h(\bar{s}_{jk}^p, \theta_j, \phi_k)$  αυξάνεται κατά ένα. Επίπεδα στον 3Δ χώρο ορίζονται εκείνα που δίνουν τις μεγαλύτερες τιμές του αθροιστή.

Από το αποτέλεσμα του Hough μετασχηματισμού, πρέπει να προσδιοριστούν εκείνα τα επίπεδα που αντιστοιχούν σε πραγματικά επίπεδα στον 3Δ χώρο. Το μειονέκτημα της μεθόδου έγκειται στον προσδιορισμό της τιμής του κατωφλίου, η οποία θα οδηγήσει σε ορθή τρισδιάστατη ανακατασκευή. Συγκεκριμένα, όταν ο αριθμός των ζευγών των σημείων είναι μικρός, όπως στη πρώτη εφαρμογή εκ των δύο, τότε είναι πολύ πιθανό να ανιχνευθούν λανθασμένα επίπεδα. Στην περίπτωση αυτή, είναι αναγκαίο ο αλγόριθμος να γνωρίζει το μέγιστο αριθμό επιπέδων  $M$  που καλείται να ανιχνεύσει. Στη συγκεκριμένη εφαρμογή, θεωρήθηκε γνωστό ότι  $M = 3$  και το κατώφλι ορίστηκε ίσο με  $t_1 = 0.3N$ , όπου  $N$  το πλήθος των σημείων στον 3Δ κόσμο. Στην περίπτωση που είναι δυνατό να προσδιοριστούν περισσότερα ζεύγη σημείων μεταξύ των δύο εικόνων, δεν είναι απαραίτητη η ακριβής γνώση του αριθμού των επιπέδων. Στην τελευταία περίπτωση, το κατώφλι μπορεί να οριστεί ως ένα μικρό ποσοστό του συνολικού αριθμού των 3Δ σημείων. Για τον αποδοτικό προσδιορισμό της τιμής του κατωφλίου, είναι δυνατό να γίνει η υπόθεση ότι από δύο όψεις ενός κτηρίου δεν είναι πιθανό να απεικονίζονται περισσότερα από 8-10 επίπεδα. Με τον τρόπο αυτό, μπορεί να θεωρηθεί ότι η τιμή του κατωφλίου θα ορίζεται ως  $t = 0.12N$ . Στην δεύτερη εφαρμογή, όπου το πλήθος των σημείων ήταν μεγαλύτερο, ορίστηκε η τιμή του κατωφλίου να είναι ίση με  $t_2 = 0.15N$ . Επομένως, αν ο Hough μετασχηματισμός ανιχνεύσει πιθανά επίπεδα αποτελούμενα από σημεία περισσότερα από το προκαθορισμένο κατώφλι, τότε τα επίπεδα αυτά θεωρούνται πραγματικά.

Στη συνέχεια, και αφού αναγνωριστούν τα επίπεδα (όψεις) που σχηματίζονται από το σύνολο των 3Δ σημείων, αυτά χωρίζονται σε μικρότερα τριγωνικά επίπεδα με ιδέες Delaunay Triangulation (βλ. Παράρτημα Γ'). Συγκεκριμένα, σε κάθε επίπεδο  $\Pi$  αποδίδονται τα σημεία που το ορίζουν (όπως προκύπτει από τον 3Δ Hough μετασχηματισμό). Το επίπεδο χωρίζεται σε μικρότερα τριγωνικά επίπεδα με κορυφές τα σημεία του επιπέδου μέσω του delaunay αλγορίθμου, ο οποίος χωρίζει το επίπεδο με τον βέλτιστο τρόπο.



**Σχήμα 6.2:** Ομογραφικός μετασχηματισμός του επιπέδου της εικόνας ως προς την επιφάνεια του 3Δ κόσμου. Το αποτέλεσμα του μετασχηματισμού δίνει μια νέα εικόνα front-to-parallel προοπτικής ως προς την επιφάνεια αναφοράς.

### 6.1.5 Απόδοση Υφής

Για την οπτική αναπαράσταση του ανακατασκευασμένου κτηρίου επιθυμητή είναι η απόδοση της υφής στα επίπεδα που ανιχνεύθηκαν με τον παραπάνω τρόπο. Η υφή ενός επιπέδου εξάγεται από τις εικόνες της βάσης δεδομένων. Συγκεκριμένα, η καλύτερη αναπαράσταση της υφής προκύπτει από την παρατήρηση της επιφάνειας από μέτωπο παράλληλης προοπτικής (front-to-parallel perspective). Η διαδικασία αυτή περιλαμβάνει τη "διόρθωση" της αρχικής εικόνας. Με άλλα λόγια, η αρχική εικόνα μετασχηματίζεται ώστε η νέα εικόνα να είναι παράλληλη με την αναπαράσταση της επιφάνειας του 3Δ κόσμου.

Δεδομένων των συντεταγμένων τριών σημείων στην εικόνα και των αντίστοιχων στον 3Δ κόσμο, όπου τα σημεία ορίζουν ένα επίπεδο, το επίπεδο μπορεί να μετασχηματιστεί σε μέτωπο παράλληλης προοπτικής. Η διαδικασία αυτή απαιτεί ομογραφικό μετασχηματισμό.

Ως ομογραφία ορίζεται ο μετασχηματισμός μιας εικόνας, που αντιστοιχεί σε μετασχηματισμό στερεού σώματος για το επίπεδο προβολής της κάμερας αναφορικά με την επιφάνεια που απεικονίζεται στην εικόνα. Στο Σχήμα 6.2 φαίνεται ένα παράδειγμα ομογραφικού μετασχηματισμού. Το επίπεδο της εικόνας αριστερά μετασχηματίζεται έτσι ώστε το νέο επίπεδο να είναι παράλληλο με την επίπεδη επιφάνεια του 3Δ κόσμου. Δηλαδή, η νέα εικόνα απεικονίζει το επίπεδο με front-to-parallel προοπτική.

Σκοπός είναι να βρεθεί ο μετασχηματισμός  $H$  έτσι ώστε να προκύπτει η "διορθωμένη" εικόνα

$$\mathbf{x}_{rect} = H\mathbf{x}$$

με τις ιδιότητες, το κάθετο διάνυσμα της νέας εικόνας να είναι παράλληλο με το κάθετο διάνυσμα της επιφάνειας αναφοράς του 3Δ χώρου (front-to-parallel) και ο τελικός μετασχηματισμός να είναι ομογραφία.

Έστω ότι για την αρχική εικόνα ισχύει

$$\mathbf{x} = K[R|\mathbf{t}]\mathbf{X} \quad (6.9)$$

όπου η τελευταία σειρά του πίνακα  $R$  δίνει τη διεύθυνση του οπτικού άξονα. Για τη "διορθωμένη" εικόνα

$$\mathbf{x}_{rect} = K[R_{rect}|\mathbf{t}_{rect}]\mathbf{X} \quad (6.10)$$



όπου η τελευταία σειρά του πίνακα  $R_{rect}$  είναι η διεύθυνση του νέου οπτικού άξονα, ο οποίος πρέπει να είναι παράλληλος με το κάθετο διάνυσμα  $\mathbf{n}$  της επιφάνειας αναφοράς του 3Δ χώρου.

Η ομογραφία που αντιστοιχίζει ένα σημείο  $\mathbf{x}$  της αρχικής εικόνας σε ένα σημείο  $\mathbf{x}_{rect}$  της διορθωμένης εικόνας, δίνεται από τη σχέση

$$H = KSK^{-1} \quad (6.11)$$

όπου  $S$  είναι ένας πίνακας περιστροφής (rotation matrix) με τελευταία γραμμή το διάνυσμα  $\mathbf{n}^T R^{-1}$ .

*Απόδειξη.* Ο μετασχηματισμός  $H$  στην αρχική εικόνα  $\mathbf{x}$  δίνει

$$\begin{aligned} \mathbf{x}_{rect} &= H\mathbf{x} = HK[R|\mathbf{t}]\mathbf{X} \\ &= KSK^{-1}K[R|\mathbf{t}]\mathbf{X} \\ &= K[SR|S\mathbf{t}]\mathbf{X} \\ &= K[R_{rect}|\mathbf{t}_{rect}]\mathbf{X} \end{aligned}$$

Ο πίνακας  $R_{rect} = SR$  είναι πίνακας περιστροφής και η τελευταία σειρά δίνεται από τη σχέση  $R_{rect}^{(3)} = S^{(3)}R = \mathbf{n}^T R^{-1}R = \mathbf{n}^T$ .  $\square$

Από τρία σημεία  $\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3$  του επίπεδου του 3Δ χώρου (τρία σημεία ορίζουν ένα επίπεδο) το κάθετο διάνυσμα στο επίπεδο δίνεται από τη σχέση  $\mathbf{n} = (\mathbf{p}_1 - \mathbf{p}_2) \times (\mathbf{p}_1 - \mathbf{p}_3)$ . Ο πίνακας  $S$  είναι ένας πίνακας περιστροφής με τελευταία γραμμή το διάνυσμα  $\mathbf{n}^T R^{-1}$ . Ο πίνακας  $S$  μπορεί να οριστεί ως πίνακας Householder ο οποίος δίνεται από τη σχέση

$$S = I - 2 \cdot \frac{v \cdot v^T}{v^T \cdot v} \quad (6.12)$$

όπου  $v = S_3^T + \|S_3\| \cdot \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$  και  $S_3 = \frac{\mathbf{n}^T}{\|\mathbf{n}\|} \cdot R^{-1}$ .

*Απόδειξη.* Πράγματι, αν  $A = I - 2 \cdot \frac{v \cdot v^T}{v^T \cdot v}$  όπου  $v = u^T + \|u\| \cdot \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$  με  $u = \frac{\mathbf{n}^T}{\|\mathbf{n}\|} \cdot R^{-1}$  τότε η τελευταία γραμμή του πίνακα  $A$  προκύπτει

$$A_3 = [0 \ 0 \ 1]A = [0 \ 0 \ 1] \frac{v^T v I - 2vv^T}{v^T v} \quad (6.13)$$

Με απλές πράξεις γραμμικής άλγεβρας

$$\begin{aligned} vv^T &= u^T u + \|u\|u^t[001] + \|u\| [0 \ 0 \ 1]^T u + \|u\|^2 [0 \ 0 \ 1]^T [0 \ 0 \ 1] \\ v^T v &= 2\|u\|^2 + 2\|u\|u^{(3)} \end{aligned} \quad (6.14)$$

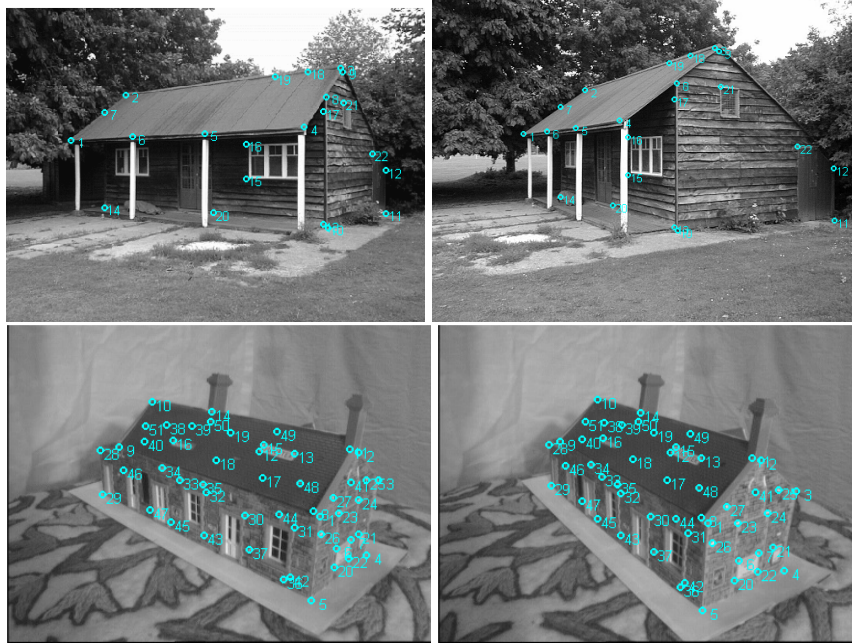
Επομένως, η σχέση 6.13 από τη σχέση 6.14 γίνεται

$$A_3 = \frac{[0 \ 0 \ 1](2\|u\|^2 + 2\|u\|u^{(3)}) - 2(u^{(3)}u + u^{(3)}[0 \ 0 \ 1] + \|u\|u + \|u\|^2[0 \ 0 \ 1])}{2\|u\|(\|u\| + u^{(3)})} \quad (6.15)$$

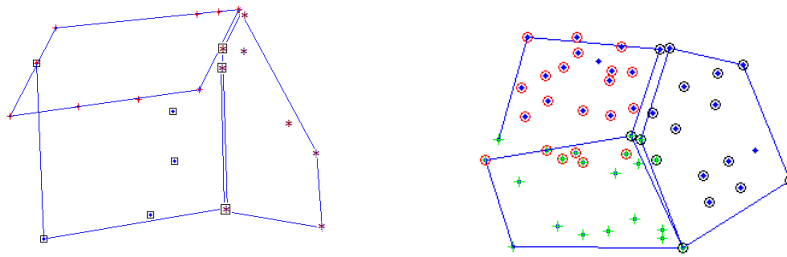
$$A_3 = -\frac{u}{\|u\|} \quad (6.16)$$

Είναι εμφανές ότι το διάνυσμα της τελευταίας γραμμής του πίνακα  $A$  είναι παράλληλο με το διάνυσμα  $\mathbf{n}^T R^{-1}$ , όπως απαιτείται.  $\square$

Για κάθε επίπεδη επιφάνεια στον 3Δ χώρο μπορεί να οριστεί η ομογραφία που μετασχηματίζει την εικόνα σε front-to-parallel προοπτική αναφορικά με το επίπεδο.



Σχήμα 6.3: Τα matchpoints στις δύο εικόνες του κτηρίου προς ανακατασκευή.



Σχήμα 6.4: 3Δ αναπαράσταση των σημείων ταξινομημένα σε επίπεδα (a) για την πρώτη εφαρμογή, (b) για την δεύτερη εφαρμογή. Οι ευθείες έχουν προστεθεί για την καλύτερη παρατήρηση του νέφους των σημείων. Δεν αποτελούν έξοδο του αλγορίθμου.

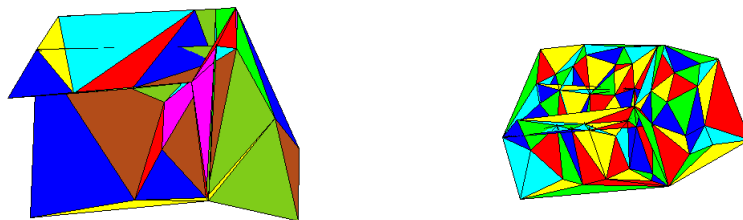
### 6.1.6 Τελικά Αποτελέσματα

Η βάση δεδομένων για την εφαρμογή του αλγορίθμου ανακατασκευής κτηρίου φαίνονται στο Σχήμα 6.1. Στο Σχήμα 6.3 φαίνονται τα ζεύγη σημείων στις δύο εικόνες.

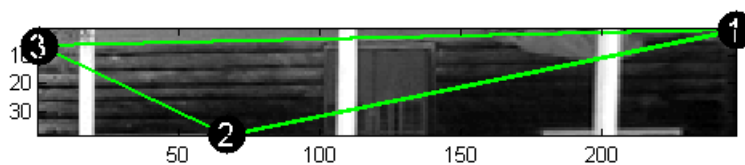
Από το 3Δ νέφος σημείων επιχειρείται η ανίχνευση διάφορων επιπέδων με τη μέθοδο του 3D Hough μετασχηματισμού. Στο Σχήμα 6.4 φαίνονται τα επίπεδα όπως ορίζονται από τα σημεία που ανήκουν σε αυτά με διαφορετικό σχήμα και χρώμα για κάθε επίπεδο. Οι γραμμές έχουν συμπληρωθεί στο σχήμα για την εύκολη παρατήρηση των επιπέδων.

Μετά από την ανίχνευση των επιπέδων στο χώρο, κάθε επίπεδο (όψη) που αναγνωρίστηκε χωρίζεται βέλτιστα σε μικρότερα τριγωνικά επίπεδα με τον αλγόριθμο Delaunay Triangulation. Δηλαδή, κάθε επίπεδο αποτελείται από πλήθος τριγωνικών επιπέδων με κορυφές τα 3Δ σημεία που ανήκουν σε αυτό. Στο Σχήμα 6.5 φαίνεται το αποτέλεσμα για τις δύο εφαρμογές

Αφού ορίστηκαν τα τριγωνικά επίπεδα, για κάθε τέτοιο επίπεδο εκτελείται ομογραφικός μετασχηματισμός στην εικόνα που καλύτερα "βλέπει" το επίπεδο αυτό. Συγκεκριμένα, επιλέγεται ως



**Σχήμα 6.5:** 3Δ αναπαράσταση των κτηρίων αποτελούμενα από τριγωνικά επίπεδα για τις δύο εφαρμογές. Εφαρμόστηκε ο αλγόριθμος του Delaunay Triangulation.



**Σχήμα 6.6:** Το διορθωμένο επίπεδο ως αναπαράσταση υφής για την αντίστοιχη επίπεδη επιφάνεια του 3Δ χώρου

αρχική η εικόνα με τη μικρότερη γωνία μεταξύ του οπτικού άξονα της και του κάθετου διανύσματος  $\mathbf{n}$  του 3Δ επιπέδου. Στη συνέχεια, εκτελείται ο ομογραφικός μετασχηματισμός στην εικόνα αυτή ώστε να προκύψει μια νέα εικόνα παράλληλη στο τριγωνικό 3Δ επίπεδο.

Στο Σχήμα 6.6 φαίνεται μια "διορθωμένη" εικόνα. Συγκεκριμένα, στην εικόνα απεικονίζεται η όψη του κτηρίου από front-to-parallel προοπτική. Η όψη αυτή προέκυψε από το μετασχηματισμό μιας εκ των δύο εικόνων του κτηρίου.

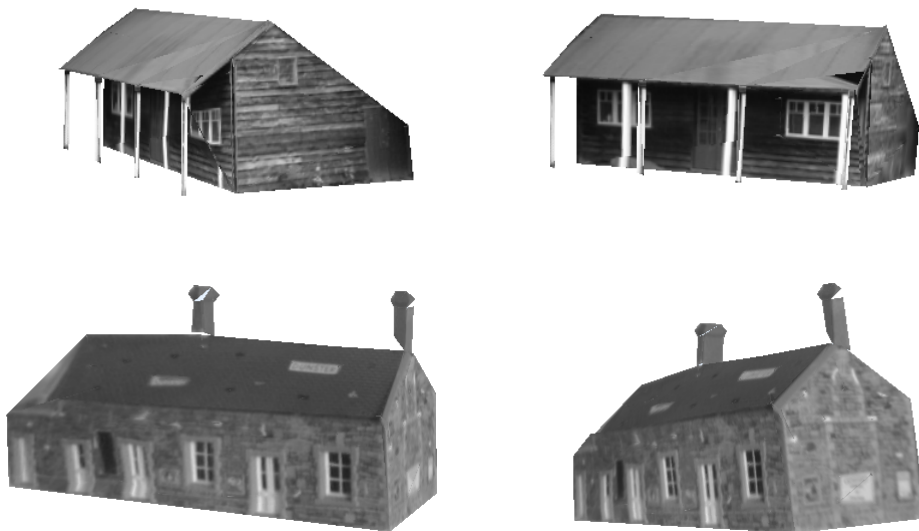
Έχοντας υπολογίσει τους ομογραφικούς μετασχηματισμούς για όλα τα τριγωνικά επίπεδα, οι νέες εικόνες που προκύπτουν μπορούν να αποδοθούν ως υφή στην 3Δ αναπαράσταση του κτηρίου. Στο Σχήμα 6.7 φαίνονται οι 3Δ ανακατασκευές με την απόδοση υφής. Τα 3Δ μοντέλα υλοποιήθηκαν σε VRML.

Ωστόσο, οι λεπτομέρειες των κτηρίων που δεν μπορούν να αναγνωριστούν από τον αλγόριθμο, όπως κολώνες οι καμινάδες (μη επίπεδες όψεις), δεν ανακατασκευάζονται με τον τρόπο αυτό καθώς δεν αποτελούν διακριτά επίπεδα στο χώρο. Για να αναπαρασταθούν, ο χρήστης πρέπει να δώσει μόνος του τις λεπτομέρειες αυτές. Στο Σχήμα 6.8 φαίνεται η πλήρης 3Δ ανακατασκευή των κτηρίων.

Ένα άλλο μειονέκτημα της μεθόδου αποτελεί το γεγονός ότι οι μη κυρτές εσοχές, όπως αυτή που ορίζεται από τη στέγη και την μια πρόσοψη του κτηρίου (βλ. Σχήμα 6.9). Για να ξεπεραστεί το πρόβλημα αυτό, είναι αναγκαία η πρότερη γνώση ή η καθοδήγηση από τον χρήστη ώστε να απορρίψει ενδεχομένως σημεία που ανήκουν μεν σε κάποια σωστά ορισμένο επίπεδο ωστόσο οδηγούν σε λανθασμένη ανακατασκευή.



**Σχήμα 6.7:** Η 3Δ ανακατασκευή των κτηρίων όπως προέκυψε από την υλοποίηση του αλγορίθμου ανακατασκευής



**Σχήμα 6.8:** Πλήρης 3Δ ανακατασκευή των κτηρίων



**Σχήμα 6.9:** Το μειονέκτημα του αλγορίθμου ανακατασκευής κτηρίων βρίσκεται στην αδυναμία του να αναγνωρίσει μη έγκυρα σημεία μιας όψης του κτηρίου, τα οποία ωστόσο βρίσκονται πάνω στο επίπεδο της όψης αυτής.



**Σχήμα 6.10:** Δύο από τις εικόνες της βάσης δεδομένων που απεικονίζουν τον ναό στη Σικελία (πηγή: <http://vision.middlebury.edu/mview/data/>).

## 6.2 Εφαρμογή Τρισδιάστατης Ανακατασκευής με Δυνατότητα Περιήγησης στα Δεδομένα ενός Μνημείου

Η ιδέα πίσω από την εφαρμογή αυτή ήταν γρήγορη η 3Δ ανακατασκευή από ένα μικρό σύνολο εικόνων ενός μνημείου που το απεικονίζουν περιμετρικά και η δυνατότητα περιήγησης στις εικόνες αυτές ανάλογα με την θέση της αντίστοιχης κάμερας. Την εφαρμογή αυτή θα μπορούσε να τρέξει ένας επισκέπτης σε πραγματικό χρόνο με είσοδο τα δικά του δεδομένα από το εκάστοτε μνημείο και τις παραμέτρους βαθμονόμησης που η ίδια η κάμερα παρέχει σε κάθε εικόνα (EXIF tag).

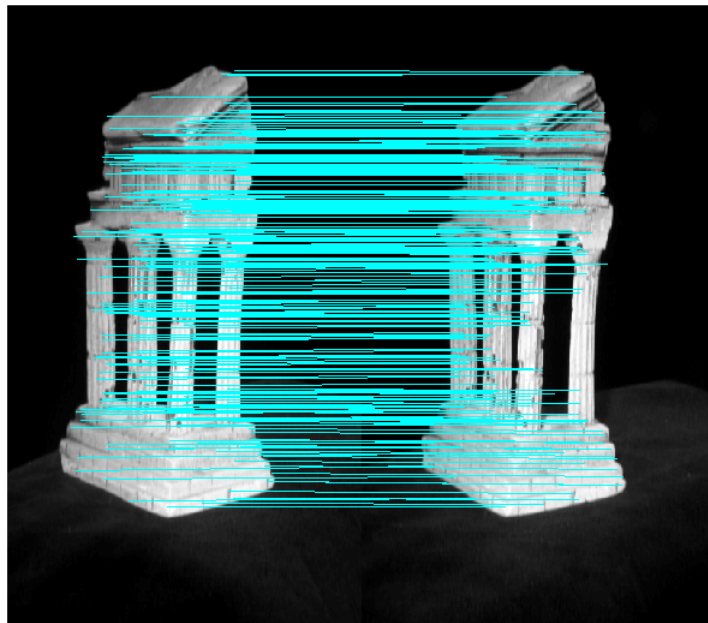
Η βάση δεδομένων αποτελείται από 41 εικόνες που απεικονίζουν τον Ναό των Διόσκουρων στη Σικελία. Δύο από τις εικόνες αυτές φαίνονται στο Σχήμα 6.10.

Στη συνέχεια, υπολογίζονται τα ζεύγη σημείων μεταξύ διαδοχικών εικόνων της βάσης με χρήση του SIFT σημείων ενδιαφέροντος. Στο Σχήμα 6.11 φαίνονται τα ζεύγη σημείων που ανιχνεύθηκαν μεταξύ των εικόνων του Σχήματος 6.10.

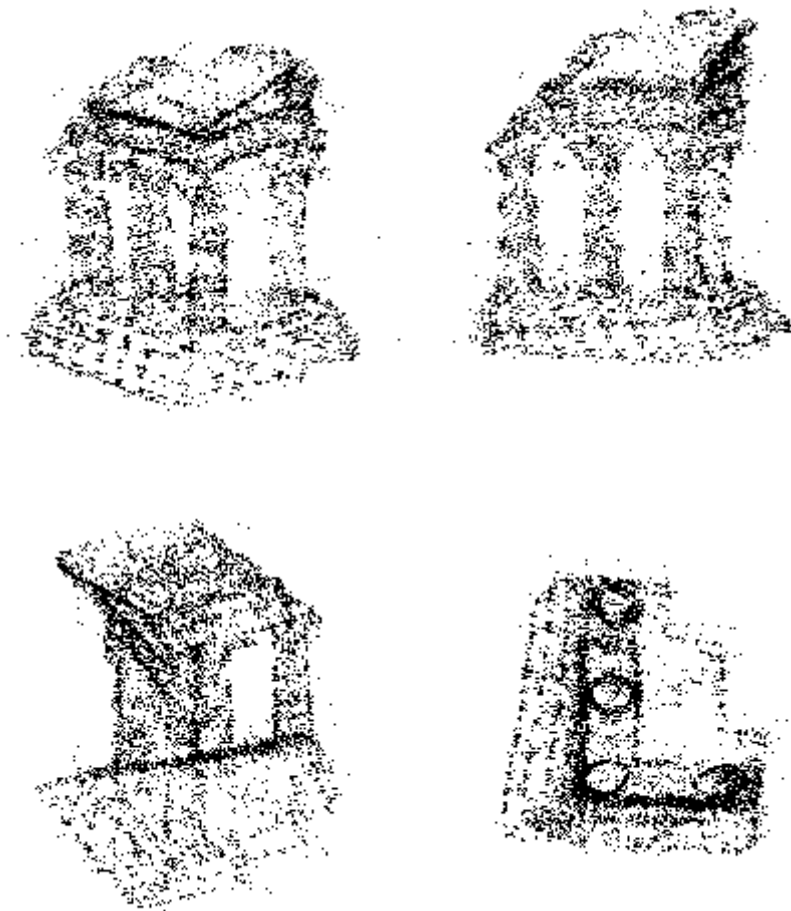
Με γνώση των παραμέτρων των καμερών (αρκούν οι παράμετροι βαθμονόμησης της κάμερας για κάθε εικόνα) πραγματοποιείται ανακατασκευή των σημείων, υπολογίζοντας τη 3Δ θέση τους για κάθε ζεύγος σημείων των εικόνων. Ο αλγόριθμος ανακατασκευής στην πιο απλή μορφή του παρέχει αποτελέσματα εντός ολίγων λεπτών. Στο Σχήμα 6.12 φαίνεται το νέφος σημείων.

Με γνώση των παραμέτρων των καμερών είτε με εκτίμηση τους με τις μεθόδους που αναφέρθηκαν στην εργασία, είναι εύκολο να υπολογιστεί η θέση των οπτικών κέντρων των καμερών. Συγκεκριμένα, από το Κεφάλαιο 2 είναι γνωστό ότι  $P_i C_i = 0$ , όπου  $P_i$  είναι ο προβολικός πίνακας της  $i$ -οστής κάμερας και  $C_i$  το κέντρο της σε ομογενείς συντεταγμένες. Στο Σχήμα 6.13 φαίνεται το αποτέλεσμα του 3Δ νέφους σημείων καθώς και τα οπτικά κέντρα των καμερών (κόκκινα σημεία).

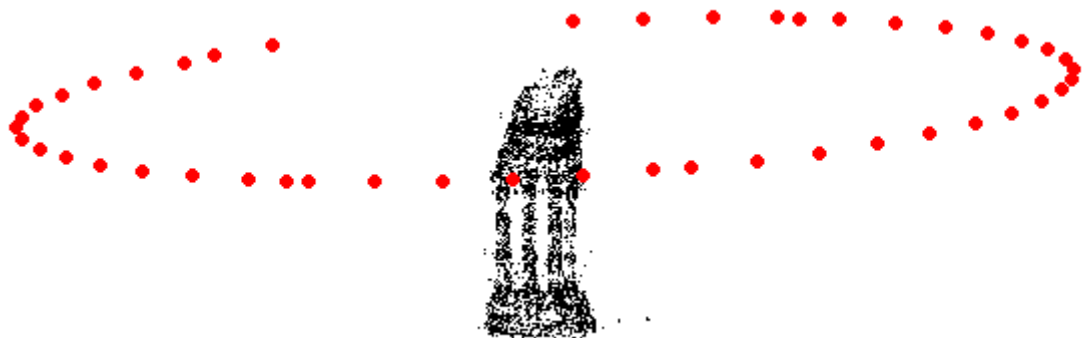
Εν συνεχεία, ο χρήστης μπορεί να επιλέξει μία από τις διαθέσιμες θέσεις των οπτικών κέντρων των καμερών και άμεσα η εικόνα που απεικονίζει την αντίστοιχη όψη του ναού προβάλλεται. Η εφαρμογή φαίνεται σε στιγμιότυπο στο Σχήμα 6.14.



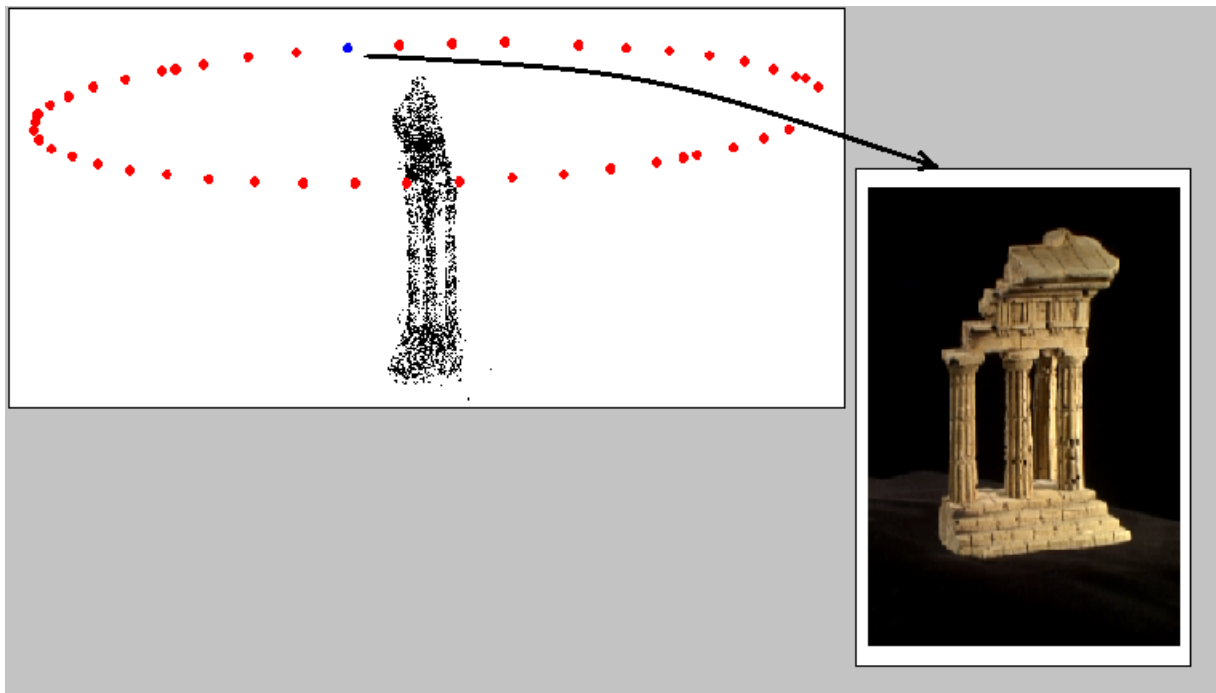
**Σχήμα 6.11:** Τα ζεύγη SIFT σημείων μεταξύ των δύο εικόνων του Ναού, αφού διαγράφηκαν τα ζεύγη που αποτελούσαν outliers σύμφωνα με τον RANSAC αλγόριθμο.



**Σχήμα 6.12:** Το 3D νέφος σημείων όπως προέκυψε από τις μεθόδους ανακατασκευής. Ο αλγόριθμος που χρησιμοποιείται είναι απλός αλλά γρήγορος και επομένως επιτρέπει τη χρήση του για εφαρμογές σε πραγματικό χρόνο.



**Σχήμα 6.13:** Το αποτέλεσμα της πρόχειρης 3D ανακατασκευής και οι θέσεις των καμερών με κόκκινα σημεία.



**Σχήμα 6.14:** Η εφαρμογή περιήγησης στις εικόνες της βάσης δεδομένων με βάση τη θέση των αντίστοιχων οπτικών κέντρων. Με μπλέ κουκίδα σημειώνεται η κάμερα που επιλέχθηκε από τον χρήστη. Η εικόνα στα δεξιά αντιστοιχεί στην όψη από το συγκεκριμένο οπτικό κέντρο.



## Κεφάλαιο 7

# Επίλογος

### 7.1 Συμβολή της διπλωματικής εργασίας

Στην παρούσα διπλωματική εργασία μελετήθηκε το πρόβλημα της 3Δ ανακατασκευής από εικόνες. Οι κατευθύνσεις με τις οποίες ασχοληθήκαμε ήταν η τρισδιάστατη ανακατασκευή αντικειμένων από πολλαπλές όψεις (περισσότερες από 5) και η ανακατασκευή κτηρίων από δύο μόνο όψεις.

Οι δυσκολίες που εμφανίστηκαν κατά τη διάρκεια της παρούσας εργασίας σχετίζονται τόσο με την μελέτη και εξερεύνηση μεθόδων για γρήγορη και ικανοποιητική τρισδιάστατη ανακατασκευή όσο και με την πρακτική υλοποίηση των αλγορίθμων υπολογιστικά.

Η συμβολή της παρούσας εργασίας περιγράφεται περιληπτικά παρακάτω

- Υλοποίηση ενός αλγορίθμου ημιαυτόματης ανίχνευσης σημείων φυγής με βάση τις ιδέες της στατιστικής σημαντικότητας.
- Μελέτη, υλοποίηση και αξιολόγηση μεθόδων εξαγωγής χαρακτηριστικών και σημείων ενδιαφέροντος από εικόνες.
- Υλοποίηση και εφαρμογή μεθόδου ανακατασκευής αντικειμένων από πολλαπλές όψεις με σκοπό τη βέλτιστη εκτίμηση του σχήματος από διασκορπισμένο νέφος σημείων. Ο αλγόριθμος περιλαμβάνει ιδέες από προβολική και διαφορική γεωμετρία.
- Υλοποίηση και εφαρμογή αλγορίθμου για την 3Δ ανακατασκευή των ακμών ενός αντικειμένου, που βασίζεται στην υλοποίηση των Primal Sketch Tokens (βλ. [23]).
- Υλοποίηση και εφαρμογή ενός γρήγορου αλγορίθμου για την τρισδιάστατη αναπαράσταση κτηρίων από δύο μόνο όψεις με απομακρυσμένα οπτικά κέντρα και κατ' επέκταση περιορισμένη γνώση κοινών χαρακτηριστικών μεταξύ των εικόνων. Ο αλγόριθμος περιλαμβάνει ιδέες από υπολογιστική γεωμετρία.

### 7.2 Μελλοντικές κατευθύνσεις

Μολονότι το πεδίο της 3Δ ανακατασκευής έχει μελετηθεί διεξοδικά την τελευταία δεκαετία, οι επιστημονικές μελέτες έχουν μειονεκτήματα τα οποία δεν επιτρέπουν την εύκολη χρήση των μεθόδων τους. Για παράδειγμα, η πλειοψηφία των αλγορίθμων 3Δ ανακατασκευής αντικειμένων απαιτούν μεγάλο πλήθος εικόνων και λήψη των δεδομένων αυτών με συγκεκριμένους τρόπους για ικανοποιητική απόδοση, κάτι που δεν είναι εφικτό σε πολλές εφαρμογές και έτσι χάνει την πρακτική αξία του.

Ο τομέας της 3Δ ανακατασκευής περιλαμβάνει διάφορα πεδία της ΟΥ που επιτρέπουν ακόμα διεξοδική μελέτη. Στην παρούσα εργασία έγινε αναφορά σε μεθόδους εξαγωγής χαρακτηριστικών και στη διαδικασία ταιριάσματος τους σε πολλαπλές όψεις. Συγκεκριμένα, η χρήση μορφολογικών τελεστών για την ανίχνευση περιοχών ενδιαφέροντος δίνει υποσχόμενα αποτελέσματα και περαιτέρω μελέτη της μπορεί να δώσει ακόμα καλύτερα. Επίσης, σημαντικό ρόλο παίζει η διαδικασία ταιριάσματος και άλλων χαρακτηριστικών μεταξύ των εικόνων, όπως ακμές, υφή κτλ.

Επιπλέον, σημαντική κρίνεται η εύρεση μεθόδων για την ακριβέστερη εκτίμηση της συνεχούς επιφάνειας που καλύτερα περιγράφει ένα διασκορπισμένο νέφος σημείων, χωρίς συγκεκριμένη κατανομή σημείων στο χώρο.

Τέλος, στον αλγόριθμο ανακατασκευής κτηρίων που προτείνεται στην παρούσα εργασία κρίνεται απαραίτητο να βρεθούν τρόποι για την αναπαράσταση λεπτομερειών που δεν είναι εύκολο να ανακατασκευαστούν. Αυτό περιλαμβάνει υλοποίηση αντίστοιχων μεθόδων και για άλλες γεωμετρικές μορφές εκτός των επιπέδων, ώστε να διατηρηθεί η ταχύτητα του αλγορίθμου. Επιπλέον, το μειονέκτημα της μεθόδου λόγω της ύπαρξης τιμής κατωφλίου για τον αλγόριθμο ανίχνευσης επιπέδων, θα μπορούσε ενδεχομένως να αντιμετωπιστεί με τη χρήση και άλλων χαρακτηριστικών στις εικόνες όπως με το ταίριασμα ευθειών και τη χρήση τους για τον προσδιορισμό των επιπέδων.

## Παράρτημα Α΄

# Πράξεις πινάκων

### Α΄.1 Ιδιότητες αντισυμμετρικών πινάκων

Έστω  $A$  ένας τετραγωνικός, πραγματικός και αντισυμμετρικός πίνακας, δηλαδή  $A^T = -A$ . Τότε ο πίνακας  $iA$  είναι ερμιτιανός καθώς  $(iA)^H = -iA^H = iA$ . Αυτό σημαίνει ότι ο πίνακας  $iA$  μπορεί να διαγωνοποιηθεί (βλ. [44]) ως εξής

$$iA = UDU^H \quad (\text{A1.1})$$

όπου  $UU^H = I$  και  $D$  είναι ένας ο διαγώνιος πίνακας των πραγματικών ιδιοτιμών. Άμεσα προκύπτει ότι

$$A = U(-iD)U^H \quad (\text{A1.2})$$

Αν ο πίνακας  $A$  είναι περιττής διάστασης τότε ισχύει  $\det(A) = 0$

Απόδειξη.

$$\det(A^T) = \det(A) = \det(-A^T)$$

Αν  $n$  είναι η διάσταση του πίνακα τότε

$$\begin{aligned} \det(A) &= (-1)^n \det(A^T) \\ \det(A) &= (-1)^n \det(A) \end{aligned} \quad (\text{A1.3})$$

Αφού  $n$  περιττός αριθμός προκύπτει ότι  $\det(A) = 0$ . Επομένως ο πίνακας  $A$  πρέπει να έχει τουλάχιστον μια μηδενική ιδιοτιμή.  $\square$

Επιπλέον, αν  $A$  ένας αντισυμμετρικός πίνακας και  $\det(A - \lambda I) = 0$  τότε ισχύει

$$\begin{aligned} \det(A + \lambda I) &= \det(-(-A - \lambda I)) = (-1)^n \det(-A - \lambda I) \\ &= (-1)^n \det(A^T - \lambda I) = \det(A - \lambda I) = 0 \end{aligned} \quad (\text{A1.4})$$

Επομένως, οι ιδιοτιμές ενός αντισυμμετρικού πίνακα είναι της μορφής  $\pm \lambda$ . Από τη σχέση A1.2 προκύπτει

$$A = U\Lambda U^H \quad \Lambda = \begin{bmatrix} i\lambda_1 & & & \\ & -i\lambda_1 & & \\ & & \ddots & \\ & & & \ddots \end{bmatrix} \quad (\text{A1.5})$$

Από τη θεωρία της γραμμικής άλγεβρας υπάρχει ένας ορθογώνιο μετασχηματισμός τέτοιος ώστε  $\Lambda = KZK^H$  όπου

$$Z = \begin{bmatrix} 0 & \lambda_1 \\ -\lambda_1 & 0 \end{bmatrix}$$

Από τη σχέση A1.5 προκύπτει ότι

$$A = OZO^H \quad (\text{A'.1})$$

όπου  $O = UK$  και  $OO^H = I$ .

## A'.2 Διαφορισμοί πινάκων

Στην ενότητα αυτή μελετάται το πρόβλημα ελαχιστοποίησης

$$\min_{\mathbf{x}} \sum_j r(\mathbf{x})_j^2 \quad (\text{A2.1})$$

όπου  $r(\mathbf{x})_j^2 = \|S_j(\mathbf{x} - \mathbf{a}_j)\|^2 = (\mathbf{x} - \mathbf{a}_j)^T A_j^T A_j (\mathbf{x} - \mathbf{a}_j)$ .

Μηδενίζοντας την παράγωγο του συναρτησιακού  $J(\mathbf{x}) = \sum_j r(\mathbf{x})_j^2$  προκύπτει

$$\begin{aligned} \frac{\partial}{\partial \mathbf{x}} \sum_j r(\mathbf{x})_j^2 &= \sum_j \frac{\partial}{\partial \mathbf{x}} r(\mathbf{x})_j^2 \\ &= \sum_j 2(\mathbf{x} - \mathbf{a}_j)^T A_j^T A_j = 0 \end{aligned} \quad (\text{A2.2})$$

Η σχέση A2.2 οδηγεί στη λύση του προβλήματος

$$\mathbf{x} = \left( \sum_j A_j^T A_j \right)^{-1} \sum_j A_j^T A_j \mathbf{a}_j \quad (\text{A2.3})$$

Αν  $A_j = I - \mathbf{v}\mathbf{v}_j^T$  με  $\|\mathbf{v}_j\| = 1$  εύκολα αποδεικνύεται ότι  $A_j^T A_j = I - \mathbf{v}\mathbf{v}_j^T$ .

## A'.3 SVD ανάλυση

Η SVD ανάλυση ενός πίνακα  $S$  διάστασης  $N \times M$  αποτελεί ένα τρόπο παραγοντοποίησης του πίνακα αυτού. Συγκεκριμένα, οποιοσδήποτε πραγματικός ή φανταστικός πίνακας μπορεί να γραφεί στην παρακάτω μορφή

$$S = UDV^H \quad (\text{A3.1})$$

όπου  $U$  είναι ένας  $M \times M$  μοναδιαίος πίνακας,  $V$  είναι ένας  $N \times N$  μοναδιαίος πίνακας και  $D$  είναι ένας  $M \times N$  διαγώνιος πίνακας με μη αρνητικά πραγματικά στοιχεία στη διαγώνιο του. Τα στοιχεία αυτά είναι οι θετικές τετραγωνικές ρίζες των μη μηδενικών ιδιοτιμών του πίνακα  $S^H S$ .

Η παραγοντοποίηση ενός πίνακα στην παραπάνω μορφή επιτρέπει τη λύση των εξισώσεων

$$S\mathbf{x} = 0 \quad (\text{A3.2})$$

Η λύση της εξίσωσης αυτής αποδεικνύεται ότι είναι το δεξί ιδιάζον διάνυσμα (στήλη του πίνακα  $V$ ) που αντιστοιχεί στη μηδενική ιδιάζουσα ιδιοτιμή (στοιχείο της διαγωνίου του πίνακα  $D$ ).

Ωστόσο, η εξίσωση A3.2 δεν έχει πάντοτε ακριβείς λύσεις. Για το λόγο αυτό, η επίλυση της εξίσωσης μετατρέπεται στο πρόβλημα ελαχιστοποίησης της ποσότητας  $S\mathbf{x}$  υπό τον περιορισμό  $\|\mathbf{x}\| = 1$ . Εύκολα αποδεικνύεται ότι η λύση αντιστοιχεί στο δεξί ιδιάζον διάνυσμα του πίνακα  $S$  που αντιστοιχεί στη μικρότερη ιδιάζουσα ιδιοτιμή.

## Παράρτημα Β΄

# Ο RANSAC αλγόριθμος

Στο παράρτημα αυτό μελετάται ο RANSAC αλγόριθμος (βλ. [16]). Ο αλγόριθμος αυτός έχει ως σκοπό την εύρεση του μοντέλου που καλύτερα περιγράφει το σύνολο δεδομένων  $S$  που περιέχει ακραία στοιχεία (outliers). Τα βήματα του αλγορίθμου περιγράφονται παρακάτω

---

**Algorithm 6** Αλγόριθμος RANSAC

---

- (i). Διάλεξε τυχαία ένα δείγμα αποτελούμενο από  $s$  το πλήθος δεδομένα εκ του συνόλου  $S$  και βρες το μοντέλο από τα δείγματα αυτά
- (ii). Καθόρισε το σύνολο δεδομένων  $S_i$  που απέχει απόσταση μικρότερη από  $t$  από το μοντέλο του βήματος (i). Το σύνολο αυτό αποτελεί το σύνολο των inliers για το μοντέλο αυτό.
- (iii). Αν το πλήθος των inliers είναι μεγαλύτερο από κάποιο κατώφλι  $T$ , υπολόγισε το μοντέλο χρησιμοποιώντας όλα τα σημεία του συνόλου  $S_i$  και τερμάτισε.
- (iv). Αν το πλήθος των inliers είναι μικρότερο από  $T$ , διάλεξε ένα νέο υποσύνολο και επανέλαβε.
- (v). Μετά από  $N$  επαναλήψεις, το πιο βέλτιστο σύνολο  $S_i$  επιλέγεται και το μοντέλο επαναυπολογίζεται με χρήση όλων των σημείων του  $S_i$ .

---

Επειδή είναι αδύνατο και περιττό ο αλγόριθμος να επαναλαμβάνεται για όλα τα δυνατά δείγματα, ο αριθμός των επαναλήψεων  $N$  επιλέγεται έτσι ώστε να διασφαλιστεί με πιθανότητα  $p$  ότι τουλάχιστον ένα από τα τυχαία δείγματα των  $s$  σημείων δεν περιέχει outliers. Συνήθως, η τιμή του  $p$  επιλέγεται να είναι ίση με 0.99. Αν υποθεθεί ότι  $w$  είναι η πιθανότητα κάποιο σημείο να είναι inlier τότε  $\epsilon = 1 - w$  είναι η πιθανότητα ότι είναι outlier. Τότε τουλάχιστον  $N$  επιλογές  $s$  σημείων πρέπει να γίνουν έτσι ώστε  $(1 - w^s)^N = 1 - p$ . Άρα

$$N = \frac{\log(1 - p)}{\log(1 - (1 - \epsilon)^s)}$$

Τα κατώφλια  $t$  και  $T$  επιλέγονται ανάλογα με το εκάστοτε πρόβλημα από το χρήστη.



## Παράρτημα Γ'

# Delaunay Triangulation

Στο παράρτημα αυτό, γίνεται αναφορά στον Delaunay τριγωνισμό (βλ. [8]).

Ο Delaunay τριγωνισμός ενός συνόλου  $P$  σημείων του επιπέδου είναι ένας τριγωνισμός  $DT(P)$  έτσι ώστε κανένα σημείο του  $P$  να μην ανήκει στον περιγεγραμμένο κύκλο οποιουδήποτε τριγώνου του  $DT(P)$ . Η μέθοδος αυτή μεγιστοποιεί την ελάχιστη γωνία όλων των γωνιών των τριγώνων του τριγωνισμού, αποφεύγοντας την ύπαρξη λεπτών τριγώνων.

Σύμφωνα με τον ορισμό των  $DT$ , ένα τριγωνικό πλέγμα είναι Delaunay τριγωνισμός αν όλοι οι περιγεγραμμένοι κύκλοι όλων των τριγώνων είναι άδειοι (δηλαδή δεν υπάρχει κάποιο σημείο του  $P$  στο εσωτερικό τους). Αυτός είναι ο ορισμός του Delaunay τριγωνισμού στον 2Δ χώρο. Στον 3Δ χώρο, ο ορισμός παραμένει ο ίδιος, με εναλλαγή του κύκλου με σφαίρα.

Το αποτέλεσμα του Delaunay τριγωνισμού σχετίζεται με το διάγραμμα Voronoi (βλ. [5]). Συγκεκριμένα, αν ενωθούν τα κέντρα των κύκλων των περιγεγραμμένων τριγώνων  $DT(P)$  για ένα διακριτό σύνολο  $P$  παράγει το διάγραμμα Voronoi.

Πολλοί αλγόριθμοι έχουν προταθεί για τον υπολογισμό του Delaunay τριγωνικού πλέγματος από σύνολο σημείων στο χώρο. Σε αυτούς περιλαμβάνονται flip αλγόριθμοι, incremental αλγόριθμοι και αλγόριθμοι διαίρει και βασίλευε.





# Βιβλιογραφία

- [1] Marc Alexa, Johannes Behr, Daniel Cohen-Or, Shachar Fleishman, David Levin, and Claudio T. Silva. Point set surfaces. In *VIS '01: Proceedings of the conference on Visualization '01*, pages 21–28, Washington, DC, USA, 2001. IEEE Computer Society.
- [2] Marc Alexa, Johannes Behr, Daniel Cohen-Or, Shachar Fleishman, David Levin, and Claudio T. Silva. Computing and rendering point set surfaces. *IEEE Transactions on Visualization and Computer Graphics*, 9(1):3–15, 2003.
- [3] Andrés Almansa, Agnès Desolneux, and Sébastien Vamech. Vanishing point detection without any a priori information. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(4):502–507, 2003.
- [4] Nina Amenta and Marshall Bern. Surface reconstruction by voronoi filtering. In *SCG '98: Proceedings of the fourteenth annual symposium on Computational geometry*, pages 39–48, New York, NY, USA, 1998. ACM.
- [5] Franz Aurenhammer. Voronoi diagrams—a survey of a fundamental geometric data structure. *ACM Comput. Surv.*, 23(3):345–405, 1991.
- [6] C. Baillard, C. Schmid, A. Zisserman, and A. Fitzgibbon. Automatic line matching and 3d reconstruction of buildings from multiple views. In *In ISPRS Conference on Automatic Extraction of GIS Objects from Digital Imagery, IAPRS Vol.32, Part 3-2W5*, pages 69–80, 1999.
- [7] C. Baillard and A. Zisserman. A plane-sweep strategy for the 3d reconstruction of buildings from multiple images. In *ISPRS Journal of Photogrammetry and Remote Sensing*, pages 56–62, 2000.
- [8] Mark de Berg, Otfried Cheong, Marc van Kreveld, and Mark Overmars. *Computational Geometry: Algorithms and Applications*. Springer-Verlag TELOS, Santa Clara, CA, USA, 2008.
- [9] Fausto Bernardini, Joshua Mittleman, Holly Rushmeier, Cláudio Silva, and Gabriel Taubin. The ball-pivoting algorithm for surface reconstruction. *IEEE Transactions on Visualization and Computer Graphics*, 5(4):349–359, 1999.
- [10] Adam Bowen, Andrew Mullins, and Nasir Rajpoot. Bayesian surface estimation from multiple cameras using a prior based on the visual hull and its application to image based rendering.
- [11] J Canny. A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 8(6):679–698, 1986.

- [12] Agnès Desolneux, Lionel Moisan, and Jean-Michel Morel. Meaningful alignments. *Int. J. Comput. Vision*, 40(1):7–23, 2000.
- [13] Richard O. Duda and Peter E. Hart. Use of the hough transformation to detect lines and curves in pictures. *Commun. ACM*, 15(1):11–15, 1972.
- [14] Richard O. Duda, Peter E. Hart, and David G. Stork. *Pattern Classification (2nd Edition)*. Wiley-Interscience, 2000.
- [15] Olivier Faugeras, Quang-Tuan Luong, and T. Papadopoulou. *The Geometry of Multiple Images: The Laws That Govern The Formation of Images of A Scene and Some of Their Applications*. MIT Press, Cambridge, MA, USA, 2001.
- [16] Martin A. Fischler and Robert C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6):381–395, 1981.
- [17] David A. Forsyth and Jean Ponce. *Computer Vision: A Modern Approach*. Prentice Hall Professional Technical Reference, 2002.
- [18] Yasutaka Furukawa and Jean Ponce. *High-Fidelity Image-Based Modeling*. Technical Report (UIUC), 2006.
- [19] Seitz S. Curless B. Goesele, M. Multi-view stereo revisited. In *In: Proceedings of CVPR 2006, New York, New York, USA, June 17-22, 2006.*, 2006.
- [20] Gaël Guennebaud and Markus Gross. Algebraic point set surfaces. In *SIGGRAPH '07: ACM SIGGRAPH 2007 papers*, page 23, New York, NY, USA, 2007. ACM.
- [21] C. Harris and M. Stephens. A combined corner and edge detector. In *Proceedings of the 4th Alvey Vision Conference*, pages 147–151, 1988.
- [22] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, second edition, 2004.
- [23] Iasonas Kokkinos, Petros Maragos, and Alan Yuille. Bottom-Up and Top-Down Object Detection Using Primal Sketch Features and Graphical Models. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2006.
- [24] A. Laurentini. The visual hull concept for silhouette-based image understanding. *IEEE Trans. Pattern Anal. Mach. Intell.*, 16(2):150–162, 1994.
- [25] Yun-Seok Lee, Han-Suh Koo, and Chang-Sung Jeong. A straight line detection using principal component analysis. *Pattern Recogn. Lett.*, 27(14):1744–1754, 2006.
- [26] Manolis I. A. Lourakis and Antonis A. Argyros. Chaining planar homographies: Fast and reliable 3d plane tracking. In *Machine Vision Conference*.
- [27] David G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60(2):91–110, 2004.
- [28] Petros Maragos. *Image Analysis and Computer Vision*. NTUA Press, 2005.
- [29] J. Matas, O. Chum, U. Martin, and T. Pajdla. Robust wide baseline stereo from maximally stable extremal regions. In *Proceedings of the British Machine Vision Conference, volume 1, pages 384-393*, 2002.

- [30] Wojciech Matusik, Chris Buehler, Ramesh Raskar, Steven J. Gortler, and Leonard McMillan. Image-based visual hulls. In *IN PROCEEDINGS OF ACM SIGGRAPH 2000*, pages 369–374, 2000.
- [31] Wojciech Matusik, Chris Buehler, Ramesh Raskar, Steven J. Gortler, and Leonard McMillan. Image-based visual hulls. In *SIGGRAPH '00: Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, pages 369–374, New York, NY, USA, 2000. ACM Press/Addison-Wesley Publishing Co.
- [32] K. Mikolajczyk. Detection of local features invariant to affine transformations, application to matching and recognition. *PhD Thesis, Institut National de Polytechniques de Grenoble, France*, 2002.
- [33] Krystian Mikolajczyk and Cordelia Schmid. A performance evaluation of local descriptors. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(10):1615–1630, 2005.
- [34] Mark Pauly, Markus Gross, and Leif P. Kobbelt. Efficient simplification of point-sampled surfaces. In *VIS '02: Proceedings of the conference on Visualization '02*, pages 163–170, Washington, DC, USA, 2002. IEEE Computer Society.
- [35] Mark Pauly, Richard Keiser, Leif P. Kobbelt, and Markus Gross. Shape modeling with point-sampled geometry. *ACM Trans. Graph.*, 22(3):641–650, 2003.
- [36] Marc Pollefeys, Luc Van Gool, Maarten Vergauwen, Kurt Cornelis, Frank Verbiest, and Jan Tops. Image-based 3d acquisition of archaeological heritage and applications. In *VAST '01: Proceedings of the 2001 conference on Virtual reality, archeology, and cultural heritage*, pages 255–262, New York, NY, USA, 2001. ACM.
- [37] L. Santalo. *Integral geometry and geometric probability, Encyclopedia of Mathematics and its Applications vol.1*. Addison-Wesley, 1976.
- [38] C. Schmid and A. Zisserman. Automatic line matching across views. In *CVPR '97: Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition (CVPR '97)*, page 666, Washington, DC, USA, 1997. IEEE Computer Society.
- [39] Curless B. Diebel J. Scharstein D. Szeliski R. Seitz, S. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *In IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'2006), volume 1*, pages 519–526. IEEE Computer Society, 2006.
- [40] Steven M. Seitz and Charles R. Dyer. Physically-valid view synthesis by image interpolation. In *In Proc. IEEE Workshop on Representations of Visual Scenes*, pages 18–25, 1995.
- [41] Steven M. Seitz and Charles R. Dyer. View morphing. In *SIGGRAPH '96: Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, pages 21–30, New York, NY, USA, 1996. ACM.
- [42] J. Serra. *Image Analysis and Mathematical Morphology*. Academic Press, 1982.
- [43] Noah Snavely, Steven M. Seitz, and Richard Szeliski. Modeling the world from internet photo collections. *Int. J. Comput. Vision*, 80(2):189–210, 2008.
- [44] G. Strang. *Introduction to Linear Algebra*. Wellesley-Cambridge Press, 2003.
- [45] R. Szeliski. *Computer Vision: Algorithms and Applications*. Online, 2010.

- 
- [46] Shimon Ullman. Computational studies in the interpretation of structure and motion: Summary and extension. In *Human and Machine Vision*. Academic Press, 1983.
- [47] Narunas Vaskevicius, Andreas Birk, and Kaustubh Pathak. Fast plane detection and polygonalization in noisy 3d range images.
- [48] George Vogiatzis, Carlos Hernández Esteban, Philip H. S. Torr, and Roberto Cipolla. Multiview stereo via volumetric graph-cuts and occlusion robust photo-consistency. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29(12):2241–2246, 2007.
- [49] George Vosselman and Er Dijkman. 3d building model reconstruction from point clouds and ground plans. *Int. Arch. of Photogrammetry and Remote Sensing*, pages 37–43, 2001.
- [50] Tomás Werner and Andrew Zisserman. New techniques for automated architectural reconstruction from photographs. In *ECCV '02: Proceedings of the 7th European Conference on Computer Vision-Part II*, pages 541–555, London, UK, 2002. Springer-Verlag.
- [51] Foerstner W. Yang, M. Y. Plane Detection in Point Cloud Data. Technical Report TR-IGG-P-2010-01, Department of Photogrammetry, University of Bonn, 2010.

