



ΣΧΟΛΗ ΤΕΧΝΟΛΟΓΙΚΩΝ ΕΦΑΡΜΟΓΩΝ
ΤΜΗΜΑ ΤΟΠΟΓΡΑΦΙΑΣ
ΕΡΓΑΣΤΗΡΙΟ ΦΩΤΟΓΡΑΦΙΚΗΣ ΣΗΜΕΙΩΣΗΣ

**ΣΥΓΚΡΙΣΗ ΚΑΙ ΑΞΙΟΛΟΓΗΣΗ ΑΛΓΟΡΙΘΜΩΝ ΑΝΙΧΝΕΥΣΗΣ
ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ ΣΗΜΕΙΩΝ ΣΕ ΕΙΚΟΝΕΣ**

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

ΓΕΡΑΣΙΜΟΣ ΜΑΤΙΔΗΣ

ΜΑΪΟΣ 2015



ΣΧΟΛΗ ΤΕΧΝΟΛΟΓΙΚΩΝ ΕΦΑΡΜΟΓΩΝ
ΤΜΗΜΑ ΤΟΠΟΓΡΑΦΙΑΣ
ΕΡΓΑΣΤΗΡΙΟ ΦΩΤΟΓΡΑΦΙΑΣ

**ΣΥΓΚΡΙΣΗ ΚΑΙ ΑΞΙΟΛΟΓΗΣΗ ΑΛΓΟΡΙΘΜΩΝ ΑΝΙΧΝΕΥΣΗΣ
ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ ΣΗΜΕΙΩΝ ΣΕ ΕΙΚΟΝΕΣ**

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

ΓΕΡΑΣΙΜΟΣ ΜΑΤΙΔΗΣ

Τριμελής Εξεταστική Επιτροπή

Λάζαρος Γραμματικόπουλος(Επιβλέπων)	Καθηγητής εφαρμογών ΤΕΙ Αθήνας
Έλλη Πέτσα	Καθηγήτρια ΤΕΙ Αθήνας
Ηλίας Καλησπεράκης	Επιστημονικός συνεργάτης ΤΕΙ Αθήνας

ΜΑΪΟΣ 2015

Ευχαριστίες

Κατ’ αρχάς, θέλω να ευχαριστήσω τον επιβλέποντα καθηγητή μου, Λάζαρο Γραμματικόπουλο, για τη βοήθεια και τη συνεχή καθοδήγησή του στην εκπόνηση της παρούσας εργασίας και γενικότερα, για την εμπιστοσύνη και την εκτίμηση που μου δείχνει όλα τα χρόνια της παρουσίας μου στη σχολή. Ευχαριστώ την καθηγήτρια του εργαστηρίου Φωτογραμμετρίας, Έλλη Πέτσα, που όποτε της ζητήθηκε, μοιράστηκε μαζί μου τις γνώσεις της, καθώς επίσης και για το βιβλιογραφικό υλικό που μου προμήθευσε. Ακόμα ευχαριστώ τον, επίσης καθηγητή του εργαστηρίου, Ηλία Καλησπεράκη, για την παροχή προγραμματιστικού υλικού, που διευκόλυνε και επιτάχυνε σε μεγάλο βαθμό την ολοκλήρωση της εργασίας μου. Περισσότερο απ’ όλους, ευχαριστώ την οικογένεια και τους κοντινούς μου ανθρώπους για την ηθική -και όχι μόνο- στήριξή τους όλα αυτά τα χρόνια.

Αφιερώνω ολόκληρο τον κύκλο που κλείνει με την παρούσα εργασία, στη μνήμη του πατέρα μου, που υπήρξε πάντα ο πιο ένθερμος υποστηρικτής κάθε μου προσπάθειας.

Περίληψη

Η μετάβαση της Φωτογραμμετρίας στην ψηφιακή της περίοδο αποτέλεσε το πρώτο ουσιαστικό βήμα για την αυτοματοποίηση των φωτογραμμετρικών διαδικασιών. Έκτοτε η έρευνα στο πεδίο αυτό, αλλά και στο συγγενές πεδίο της όρασης υπολογιστών, προσανατολίζεται σε μεγάλο βαθμό σε θέματα επεξεργασίας εικόνας και ειδικά στη διαδικασία της ψηφιακής συνταύτισης εικόνων. Αυτό είναι εύλογο αν σκεφτεί κανείς ότι το ζήτημα της εύρεσης ομόλογων σημείων μεταξύ δύο ή περισσοτέρων εικόνων τίθεται σχεδόν σε κάθε φωτογραμμετρική εφαρμογή.

Ήδη λοιπόν από τις αρχές της δεκαετίας του '80, άρχισαν να αναπτύσσονται αλγόριθμοι αυτόματης ανίχνευσης χαρακτηριστικών σημείων. Με την πάροδο των χρόνων εμφανίζονταν συνεχώς καινούργιοι, προσπαθώντας να βελτιωθούν και να καταστούν ανθεκτικότεροι σε γεωμετρικές μεταβολές στις εικόνες, όπως ύπαρξη στροφών, διαφορά κλίμακας, καθώς και σε μεταβολές στην φωτεινότητα, ή ύπαρξη θορύβου. Ο σκοπός της παρούσας εργασίας είναι να μελετήσει κάποιους από αυτούς και να συγκρίνει, ειδικά τους αλγόριθμους SIFT και SURF, οι οποίοι χρησιμοποιούνται ευρέως σε εφαρμογές ψηφιακής συνταύτισης λόγω της αποτελεσματικότητας και της αξιοπιστίας τους. Στη βιβλιογραφία υπάρχει πληθώρα αναφορών στη σύγκριση αυτή, όμως ως επί το πλείστον, επικεντρώνεται στην απόδοση των αλγορίθμων όσον αφορά το πλήθος και την ακρίβεια των σημείων που συνταυτίζονται. Εδώ, επιχειρείται να συμπεριληφθεί στην σύγκριση και ένα ζήτημα που έχει μεγάλη σημασία, ιδίως για τη Φωτογραμμετρία, αυτό της αξιολόγησης της κατανομής των σημείων στις εικόνες.

Technological Educational Institute of Athens
School of Technological Applications
Department of Surveying

**Comparison and evaluation of feature points detection algorithms
on images**

Gerasimos Matidis

Diploma Thesis

May 2015

Abstract

The transition of Photogrammetry in the digital period, was the first essential step for the automation of the photogrammetric procedures. Since then, the research in this field, but also in the relative field of computer vision, is highly oriented towards issues of image processing and particularly the procedure of image matching. This is very reasonable if someone considers that the issue of establishment of image correspondences arises in almost every photogrammetric application between two or more images.

Already from the early '80s, algorithms for automatic feature points detection have started to appear. Over the years, improved algorithms have been developed, trying to be more resistant to geometric changes, like existence of rotation, scale difference, as well as changes in illumination, or existence of noise on the images. The purpose of this diploma thesis is to examine some of these and to compare especially the algorithms of SIFT and SURF, which are being used widely in applications of image matching, because of their efficiency and reliability. In the literature, there are numerous references about this comparison, but they are mostly focused on the performance of the algorithms regarding the number and the precision of matched points. In this contribution, an additional aspect of great photogrammetric importance is included in the comparison, this of evaluation of the distribution of points on the images.

Ευχαριστίες	i
Περίληψη	ii
Abstract	iii
Περιεχόμενα	iv
Εισαγωγή	vii

ΚΕΦΑΛΑΙΟ 1

Σχέση μεταξύ εικόνας και χώρου	1
1.1 Συστήματα αναφοράς στη Φωτογραμμετρία	1
1.1.1. Σύστημα του φωτογραφικού επιπέδου	1
1.1.2 Σύστημα της εικόνας	1
1.1.3 Σύστημα του χώρου	2
1.2 Σχέση μεταξύ του συστήματος της εικόνας και επίγειου συστήματος	2
1.2.1 Συνθήκη της Συγγραμμικότητας	3

ΚΕΦΑΛΑΙΟ 2

Επιπολική γεωμετρία	6
2.1 Σχετικός προσανατολισμός	7
2.1.1 Συνθήκη συνεπιπεδότητας	7
2.1.2 Δεσμευμένος επιπολικός πίνακας	8
2.2 Δισδιάστατη επιπολική γεωμετρία	9
2.2.1 Επιπολικός πίνακας	10
2.2.2 Υπολογισμός του επιπολικού πίνακα	11
2.2.2.1 Αλγόριθμος των 8 σημείων (<i>8-point algorithm</i>)	11
2.2.2.2 Κανονικοποιημένος αλγόριθμος των 8 σημείων	12
2.2.2.3 Αλγόριθμος των 7 σημείων (<i>7-point algorithm</i>)	13

<i>2.2.2.4 Μη γραμμική επίλυση</i>	13
<i>2.2.2.5 Αλγόριθμος RANSAC (RAndom SAmpling Consensus)</i>	14

ΚΕΦΑΛΑΙΟ 3

Ψηφιακή συνταύτιση	16
3.1 Επιφανειακή συνταύτιση	16
3.2 Συνταύτιση χαρακτηριστικών και τελεστές ανίχνευσης σημείων ενδιαφέροντος	18
3.2.1 Αλγόριθμος Moravec	19
3.2.2 Αλγόριθμος Harris	20
3.2.3 Αλγόριθμος SIFT	21
3.2.3.1 <i>Ανίχνευση χαρακτηριστικών σημείων</i>	22
3.2.3.2 <i>Ανάθεση προσανατολισμού</i>	24
3.2.3.3 <i>Δημιουργία των Descriptors</i>	25
3.2.3.4 <i>Συνταύτιση χαρακτηριστικών</i>	25
3.2.4 Αλγόριθμος SURF	26
3.2.4.1 <i>Ανίχνευση χαρακτηριστικών σημείων</i>	26
3.2.4.2 <i>Ανάθεση προσανατολισμού</i>	28
3.2.4.3 <i>Δημιουργία των Descriptors</i>	30
3.2.4.4 <i>Συνταύτιση χαρακτηριστικών</i>	30
3.2.5 Αλγόριθμος BRISK	31

ΚΕΦΑΛΑΙΟ 4

Εφαρμογές και αξιολόγηση	35
4.1 Αλγόριθμοι SURF και SIFT	41
4.1.1 Μικρό μήκος Βάσης του στερεοζεύγους	43
4.1.2 Διαφορά κλίμακας	48

4.1.3 Στροφή	53
4.1.4 Μεγάλο μήκος Βάσης του στερεοζεύγους	58
4.1.5 Διαφορά φωτεινότητας	63
4.1.6 Προσθήκη θορύβου	68
4.1.7 Κατανομή των σημείων στις εικόνες	73
<i>4.1.7.1 Μικρό μήκος Βάσης του στερεοζεύγους</i>	<i>74</i>
<i>4.1.7.2 Διαφορά κλίμακας</i>	<i>75</i>
<i>4.1.7.3 Στροφή</i>	<i>76</i>
<i>4.1.7.4 Μεγάλο μήκος Βάσης του στερεοζεύγους</i>	<i>77</i>
<i>4.1.7.5 Διαφορά φωτεινότητας</i>	<i>78</i>
<i>4.1.7.6 Προσθήκη θορύβου</i>	<i>79</i>

ΚΕΦΑΛΑΙΟ 5

Συμπεράσματα και προτάσεις	80
5.1 Συμπεράσματα	80
5.2 Προτάσεις για το μέλλον	81
Βιβλιογραφία	82

Εισαγωγή

Από τις αρχές της δεκαετίας του '80 που άρχισε να ανθεί ο τομέας της όρασης υπολογιστών (computer vision), εντάθηκε και η ανάγκη για αυτοματοποίηση ορισμένων φωτογραμμετρικών διαδικασιών, κυρίως εκείνων που πραγματεύονται τη γεωμετρία μεταξύ δύο ή περισσότερων επικαλυπτόμενων εικόνων, όπως για παράδειγμα, ο σχετικός προσανατολισμός, η επιπολική επανασύσταση και η μέθοδος της δέσμης. Απότερος σκοπός βέβαια, όσον αφορά στη Φωτογραμμετρία, είναι η εξαγωγή και η αξιοποίηση 3D πληροφορίας. Ταυτόχρονα, η συνύπαρξη των δύο αυτών επιστημονικών πεδίων, καθώς και λόγοι απλοποίησης και γενίκευσης του προβλήματος της μελέτης του χώρου μέσω εικόνων, ανέδειξαν την ανάγκη να ξεφύγουν οι εν λόγω διαδικασίες από τα στενά όρια των μετρικών και βαθμονομημένων μηχανών. Επανήλθε έτσι στο προσκήνιο η ιδέα της αποκατάστασης των ομόλογων δεσμών δύο επικαλυπτόμενων εικόνων, χωρίς απαραίτητα να είναι γνωστή η εσωτερική γεωμετρία της μηχανής, όπως εξ' ορισμού απαιτεί ο σχετικός προσανατολισμός. Μια ιδέα που προήλθε από το πεδίο της φωτογραμμετρίας σχεδόν έναν αιώνα πριν, εντούτοις δεν διερευνήθηκε επαρκώς, καθώς δεν υπήρχε το ουσιαστικό κίνητρο της πρακτικής χρησιμότητας.

Η θεωρητική τεκμηρίωση της ιδέας αυτής ήρθε από το πεδίο της όρασης υπολογιστών, με τη διατύπωση του επιπολικού πίνακα (fundamental matrix) από τους Faugeras (1992) και Hartley (1992) ο οποίος εκφράζει τη δισδιάστατη 'εσωτερική' γεωμετρία δύο επικαλυπτόμενων εικόνων άγνωστου εσωτερικού προσανατολισμού. Ο επιπολικός πίνακας αποτελεί τη γενικευμένη έκφραση του δεσμευμένου επιπολικού πίνακα¹ (essential matrix) που διατυπώθηκε από τους Longuet-Higgins (1981) και αποτελεί μια γραμμική προσέγγιση του σχετικού προσανατολισμού.

Η πρώτη όμως ουσιαστική εργασία (κατόπιν βέβαια της συλλογής κι επεξεργασίας των εικόνων) που πρέπει να γίνει, προκειμένου να συσχετιστούν/συνταυτιστούν δύο ή περισσότερες εικόνες (σαν μια πλήρως αυτοματοποιημένη διαδικασία), είναι να εντοπιστούν σε κάθε μια από αυτές οι γεωμετρικές εκείνες οντότητες (σημεία, γωνίες, ακμές, επιφάνειες) που θα χρησιμεύσουν για την επίτευξη των παραπάνω. Οι οντότητες που ενδιαφέρουν την παρούσα εργασία είναι τα σημεία, τα οποία εδώ αναφέρονται είτε ως 'χαρακτηριστικά σημεία', είτε ως 'σημεία ενδιαφέροντος'. Η διαδικασία αυτή, που είναι γνωστή σαν συνταύτιση χαρακτηριστικών (Feature Based Matching), αν και είναι εύκολη για τον

¹ Οι όροι 'επιπολικός πίνακας' και 'δεσμευμένος επιπολικός πίνακας' που χρησιμοποιούνται για την απόδοση των όρων 'fundamental matrix' και 'essential matrix' αντίστοιχα, εισήχθησαν για πρώτη φορά από τον Ηλία Καλησπεράκη στη διδακτορική του διατριβή (Καλησπεράκης, 2010).

άνθρωπο, είναι εξαιρετικά πολύπλοκη για τον Η/Υ, που αντιλαμβάνεται την εικόνα σαν έναν πίνακα διακριτών τιμών που αντιστοιχούν σε χρωματικούς τόνους.

Ο αλγόριθμος του Moravec, που προτάθηκε το 1980, είναι από τους πρώτους αλγόριθμους ανίχνευσης χαρακτηριστικών σημείων που εμφανίστηκαν, ενώ ακολούθησαν την ίδια δεκαετία οι αλγόριθμοι Fosler (1987) και Harris (1988). Οι αλγόριθμοι αυτοί αντιμετώπισαν σε ικανό βαθμό το πρόβλημα της ψηφιακής συνταύτισης, αλλά παρουσίαζαν αστάθεια σε γεωμετρικές κυρίως μεταβολές μεταξύ των εικόνων, όπως η διαφορά κλίμακας και οι στροφές. Ο πρώτος ανθεκτικός αλγόριθμος στις μεταβολές αυτές, είναι ο SIFT (2004), ο οποίος αποτέλεσε μια νέα προσέγγιση στο ζήτημα της ψηφιακής συνταύτισης, καθώς συνοδεύει τα σημεία που ανιχνεύει με ένα διάνυσμα περιγραφής (Descriptor) της τοπικής μεταβολής φωτεινότητας. Έκτοτε, δύο οι αλγόριθμοι που προτάθηκαν ακολουθούν αυτό το μοντέλο ‘ανίχνευσης-περιγραφής’ χαρακτηριστικών σημείων. Δύο από αυτούς είναι οι SURF (2006) και BRISK (2011) που εξετάζονται εδώ.

Αντικείμενο της εργασίας

Αντικείμενο της παρούσας πτυχιακής εργασίας είναι να μελετήσει τη συμπεριφορά των αλγορίθμων SIFT και SURF σε διάφορες γεωμετρικές και ραδιομετρικές μεταβολές μεταξύ δύο επικαλυπτόμενων εικόνων (στερεοζεύγη) και να συγκρίνει τις επιδόσεις τους, ως προς το πλήθος, την ακρίβεια και την κατανομή των ομόλογων σημείων που εξάγουν. Το μαθηματικό μοντέλο που επιλέγεται και οφείλουν να πληρούν τα ομόλογα σημεία είναι αυτό του επιπολικού πίνακα και ο έλεγχος της ακρίβειας (ή πιο σωστά, η επιλογή της ακρίβειας) υπολογισμού του, γίνεται με βάση τον αλγόριθμο RANSAC.

Διάρθρωση της εργασίας

Η διάρθρωση της εργασίας έχει ως εξής:

- Στο κεφάλαιο 1 γίνεται αναφορά στα συστήματα μέτρησης που χρησιμοποιούνται στη Φωτογραμμετρία. Επίσης, γίνεται μια αναλυτική παρουσίαση της συνθήκης συγγραμμικότητας.
- Το κεφάλαιο 2 ασχολείται με την επιπολική γεωμετρία του στερεοζεύγους. Αρχικά, αναφέρεται η περίπτωση όπου ο εσωτερικός προσανατολισμός των εικόνων είναι γνωστός. Παρουσιάζονται συνοπτικά δύο τρόποι επίλυσης του σχετικού προσανατολισμού, μέσω της συνθήκης συνεπιπεδότητας και μέσω του δεσμευμένου επιπολικού πίνακα. Στη συνέχεια, γίνεται αναφορά στον επιπολικό πίνακα και σε

ορισμένους αλγόριθμους, γραμμικούς και μη, που χρησιμοποιούνται για τον υπολογισμό του. Τέλος, παρουσιάζεται ο αλγόριθμος RANSAC, ο οποίος αποτελεί τη συνηθέστερη μέθοδο που χρησιμοποιείται για να απομακρύνθουν χονδροειδή σφάλματα και να αξιολογηθεί η ακρίβεια υπολογισμού του επιπολικού πίνακα.

- Το κεφάλαιο 3 είναι αφιερωμένο στην μελέτη της ψηφιακής συνταύτισης εικόνων. Μετά από κάποιες εισαγωγικές παρατηρήσεις, γίνεται μια συνοπτική αναφορά στην επιφανειακή συνταύτιση. Εν συνεχείᾳ, εξετάζεται η συνταύτιση χαρακτηριστικών και παρουσιάζονται, κατά χρονολογική σειρά εμφάνισης, οι αλγόριθμοι Moravec, Harris, SIFT, SURF και BRISK. Οι δύο πρώτοι παρουσιάζονται συνοπτικά, ενώ οι τρεις τελευταίοι αναλυτικά.
- Το κεφάλαιο 4 αποτελείται από τις εφαρμογές που έγιναν στο πλαίσιο της εργασίας. Αρχικά, παρουσιάζονται τα δεδομένα που χρησιμοποιήθηκαν (εικόνες, αλγόριθμοι), τα μέτρα αξιολόγησης και οι συνθήκες κάτω από τις οποίες εξετάστηκαν οι αλγόριθμοι (π.χ. στροφή, αλλαγή κλίμακας κλπ.). Αναφέρονται επίσης και οι ρυθμίσεις με τις οποίες ‘έτρεξαν’ οι αλγόριθμοι. Στη συνέχεια, ακολουθούν τα αποτελέσματα (εικόνες, πίνακες, διαγράμματα) και η αξιολόγηση των αλγορίθμων, ξεχωριστά για κάθε περίπτωση που εφαρμόστηκαν. Το κεφάλαιο κλείνει με τα αποτελέσματα της μελέτης της κατανομής των αλγορίθμων.
- Το κεφάλαιο 5 περιλαμβάνει τα συμπεράσματα που προέκυψαν από τις εφαρμογές, καθώς και προτάσεις για μελλοντική μελέτη.

ΚΕΦΑΛΑΙΟ 1

Σχέση μεταξύ εικόνας και χώρου

Στο κεφάλαιο αυτό παρουσιάζονται οι κυριότερες μαθηματικές εκφράσεις που περιγράφουν τη γενική σχέση εικόνας και χώρου. Επίσης, γίνεται μια συνοπτική παρουσίαση των συστημάτων αναφοράς που χρησιμοποιούνται στη Φωτογραμμετρία.

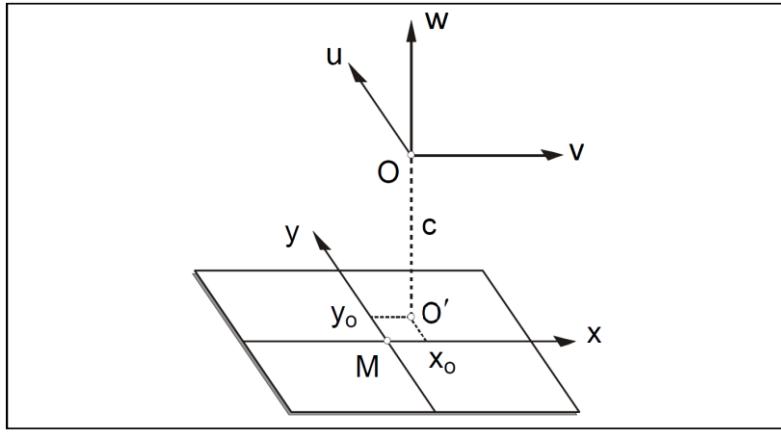
1.1 Συστήματα αναφοράς στη Φωτογραμμετρία

1.1.1. Σύστημα του φωτογραφικού επιπέδου

Συνηθέστερα το αποκαλούμε σύστημα των εικονοσυντεταγμένων. Είναι ένα δισδιάστατο δεξιόστροφο σύστημα χυ που ορίζουμε πάνω στο επίπεδο της εικόνας και έχει αρχή το κέντρο M της εικόνας (Πέτσα,2000). Οι μετρήσεις εκφράζονται πάντα σε mm, παρόλο που μπορούν αρχικώς να μετρηθούν σε διαφορετικό σύστημα μέτρησης και να μετασχηματιστούν στη συνέχεια, μέσω του αφινικού μετασχηματισμού στο σύστημα των εικονοσυντεταγμένων. Στην περίπτωση που έχουμε απευθείας ψηφιακές εικόνες, ο ζητούμενος μετασχηματισμός υποβαθμίζεται από αφινικό σε μια απλή μετάθεση.

1.1.2 Σύστημα της εικόνας

Πρόκειται για ένα τρισδιάστατο σύστημα uvw που έχει αρχή το κέντρο λήψης (ή κέντρο προβολής) O. Οι áξονες u και v είναι παράλληλοι με τους αντίστοιχους στο σύστημα των εικονοσυντεταγμένων, ενώ ο w είναι κάθετος στην εικόνα. Τα δυο συστήματα διαφέρουν κατά τη διεύθυνση w απόσταση ίση με c (σταθερά του φακού). Στην ιδανική περίπτωση η προβολή O' του κέντρου λήψης πάνω στην εικόνα θα συνέπιπτε με το κέντρο M της εικόνας, λόγω όμως της ατελούς κατασκευής των φακών, αυτή προκύπτει μετατεθειμένη κατά x₀ και y₀ στους áξονες x και y αντίστοιχα (σχήμα 1.1).



Σχήμα 1.1: Σχέση μεταξύ των συστημάτων των εικονοσυντεταγμένων και της εικόνας.

1.1.3 Σύστημα του χώρου

Πρόκειται για ένα τρισδιάστατο σύστημα συντεταγμένων XYZ που χρησιμεύει για τον προσδιορισμό των απεικονιζόμενων αντικειμένων στο χώρο και μπορεί να είναι οποιοδήποτε γεωδαιτικό σύστημα, γνωστό ή αυθαίρετο.

1.2 Σχέση μεταξύ του συστήματος της εικόνας και του επίγειου συστήματος

Προκειμένου να συνδεθούν δυο τρισορθογώνια καρτεσιανά συστήματα είναι απαραίτητο να υποστεί το ένα εξ' αυτών τρεις παράλληλες μεταθέσεις κατά τους άξονες και τρείς στροφές περί αυτών (ευκλείδειος μετασχηματισμός), έτσι ώστε να ταυτιστεί με το άλλο. Στην περίπτωση μας το ζητούμενο είναι να μετασχηματίσουμε το επίγειο σύστημα, ώστε να ταυτιστεί με αυτό της εικόνας.

Έστω τυχαίο σημείο P του εδάφους με συντεταγμένες X_P, Y_P, Z_P στο επίγειο σύστημα XYZ και X_O, Y_O, Z_O , τότε οι συντεταγμένες του P στο σύστημα UVW της εικόνας θα είναι $X_P - X_O, Y_P - Y_O$ και $Z_P - Z_O$ αντίστοιχα. Για να ολοκληρωθεί ο μετασχηματισμός χρειάζονται επιπλέον τρείς στροφές, μια περί τον άξονα X κατά γωνία ω , μια περί τον άξονα Y κατά γωνία φ και μια περί τον Z κατά γωνία κ . Οι στροφές αυτές εκφράζονται με τους πίνακες στροφής $R_1(\omega), R_2(\varphi), R_3(\kappa)$ αντίστοιχα, που ορίζονται ως εξής (Δερμάνης, 1985):

$$R_1(\omega) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\omega & \sin\omega \\ 0 & -\sin\omega & \cos\omega \end{bmatrix} \quad (1.1)$$

$$R_2(\varphi) = \begin{bmatrix} \cos\varphi & 0 & -\sin\varphi \\ 0 & 1 & 0 \\ \sin\varphi & 0 & \cos\varphi \end{bmatrix} \quad (1.2)$$

$$R3(\kappa) = \begin{bmatrix} \cos\kappa & \sin\kappa & 0 \\ -\sin\kappa & \cos\kappa & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (1.3)$$

Εντέλει, ο μετασχηματισμός σε μορφή πινάκων, εκφράζεται ως εξής:

$$\begin{bmatrix} U \\ V \\ W \end{bmatrix} = R3(\kappa) R2(\varphi) R1(\omega) \begin{bmatrix} X_p - X_0 \\ Y_p - Y_0 \\ Z_p - Z_0 \end{bmatrix} = R \begin{bmatrix} X_p - X_0 \\ Y_p - Y_0 \\ Z_p - Z_0 \end{bmatrix} \begin{bmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{23} \\ R_{31} & R_{32} & R_{33} \end{bmatrix} \quad (1.4)$$

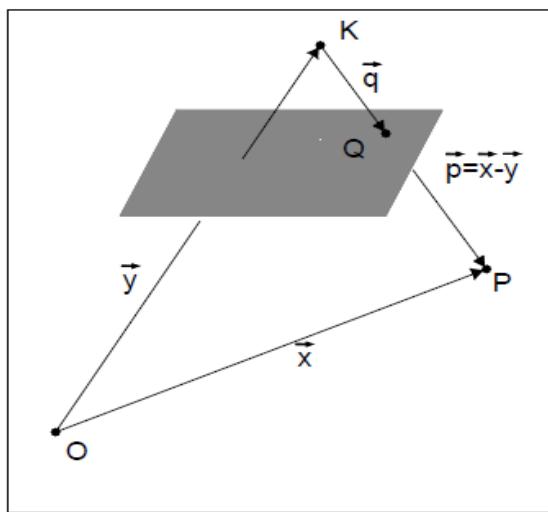
όπου R είναι ο γενικός πίνακας στροφής:

$$R = R3(\kappa) R2(\varphi) R1(\omega) = \begin{bmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{23} \\ R_{31} & R_{32} & R_{33} \end{bmatrix} =$$

$$= \begin{bmatrix} \cos\kappa \cos\varphi & \cos\kappa \sin\varphi \sin\omega + \sin\kappa \cos\omega & -\cos\kappa \sin\varphi \cos\omega + \sin\kappa \sin\omega \\ -\sin\kappa \cos\varphi & -\sin\kappa \sin\varphi \sin\omega + \cos\kappa \cos\omega & \sin\kappa \sin\varphi \cos\omega + \cos\kappa \sin\omega \\ \sin\varphi & -\cos\varphi \sin\omega & \cos\varphi \cos\omega \end{bmatrix} \quad (1.5)$$

1.2.1 Συνθήκη της Συγγραμμικότητας

Η συνθήκη της συγγραμμικότητας αποτελεί αναμφίβολα την πιο θεμελιώδη έννοια στη Φωτογραμμετρία, καθώς είναι η αναγκαία εκείνη συνθήκη που επιτρέπει την περιγραφή της σχέσης εικόνας – χώρου. Εκφράζει τη δέσμευση ότι οποιοδήποτε σημείο του αντικειμένου, το κέντρο προβολής και η εικόνα του σημείου πάνω στο αρνητικό (ή αισθητήρα, στην περίπτωση ψηφιακών μηχανών) βρίσκονται σε ευθεία κατά τη στιγμή της λήψης. Πιο αναλυτικά, όπως φαίνεται και στο σχ. 1.2 έχουμε:



Σχήμα 1.2: Γεωμετρία της φωτογραφικής λήψης (Δερμάνης, 1991)

$$\overrightarrow{OP} = \overrightarrow{OK} + \overrightarrow{KQ},$$

όπου P , τυχαίο σημείο του χώρου και Q η εικόνα του πάνω στη φωτογραφία,
ή διαφορετικά

$$\vec{x} = \vec{y} + \vec{p} \quad (1.6)$$

και

$$\vec{p} = \lambda \vec{q} \quad (1.7)$$

όπου λ είναι ο καθαρός αριθμός που ονομάζουμε συντελεστή κλίμακας και
έχει διαφορετική τιμή σε κάθε σημείο της εικόνας (σημειακή κλίμακα).

Η (1.7) αποτελεί τη μαθηματική έκφραση της συνθήκης συγγραμμικότητας
και με την βοήθεια της (1.6) γράφεται ως:

$$\lambda \vec{q} = \vec{x} - \vec{y} \quad (1.8)$$

όπου:

$$\vec{x} = \begin{bmatrix} X_P \\ Y_P \\ Z_P \end{bmatrix} \quad \vec{y} = \begin{bmatrix} X_0 \\ Y_0 \\ Z_0 \end{bmatrix} \quad (1.9)$$

οι συνιστώσες των διανυσμάτων \vec{x} και \vec{y} στο επίγειο σύστημα XYZ και

$$\vec{q} = \begin{bmatrix} X_P - X_0 \\ Y_P - Y_0 \\ -c \end{bmatrix} \quad (1.10)$$

οι συνιστώσες του διανύσματος \vec{q} στο σύστημα της εικόνας UVW.

Αν τώρα συνυπολογίσουμε και τη συνολική στροφή που πρέπει να υποστεί το
επίγειο σύστημα και με βάση τις (1.9) και (1.10), η (1.8) γράφεται ως:

$$\begin{aligned} \lambda \begin{bmatrix} X_P - X_0 \\ Y_P - Y_0 \\ -c \end{bmatrix} &= R \left(\begin{bmatrix} X_P \\ Y_P \\ Z_P \end{bmatrix} - \begin{bmatrix} X_0 \\ Y_0 \\ Z_0 \end{bmatrix} \right) \Leftrightarrow \\ \lambda \begin{bmatrix} X_P - X_0 \\ Y_P - Y_0 \\ -c \end{bmatrix} &= \begin{bmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{23} \\ R_{31} & R_{32} & R_{33} \end{bmatrix} \begin{bmatrix} X_P - X_0 \\ Y_P - Y_0 \\ Z_P - Z_0 \end{bmatrix} \end{aligned} \quad (1.11)$$

Αναλυτικά, προκύπτουν οι σχέσεις:

$$\lambda (X_P - X_0) = R_{11} (X_P - X_0) + R_{12} (Y_P - Y_0) + R_{13} (Z_P - Z_0) \quad (1.12)$$

$$\lambda (Y_P - Y_0) = R_{21} (X_P - X_0) + R_{22} (Y_P - Y_0) + R_{23} (Z_P - Z_0) \quad (1.13)$$

$$\lambda (-c) = R_{31} (X_P - X_0) + R_{32} (Y_P - Y_0) + R_{33} (Z_P - Z_0) \quad (1.14)$$

Διαιρώντας κατά μέλη τις (1.12) και (1.13) αντίστοιχα με την (1.14), απαλλασσόμαστε από τον συντελεστή λ και έχουμε:

$$\frac{x_p - x_0}{-c} = \frac{R_{11}(X_p - X_0) + R_{12}(Y_p - Y_0) + R_{13}(Z_p - Z_0)}{R_{31}(X_p - X_0) + R_{32}(Y_p - Y_0) + R_{33}(Z_p - Z_0)} \Leftrightarrow$$

$$x_p = x_0 - c \frac{R_{11}(X_p - X_0) + R_{12}(Y_p - Y_0) + R_{13}(Z_p - Z_0)}{R_{31}(X_p - X_0) + R_{32}(Y_p - Y_0) + R_{33}(Z_p - Z_0)}$$
(1.15)

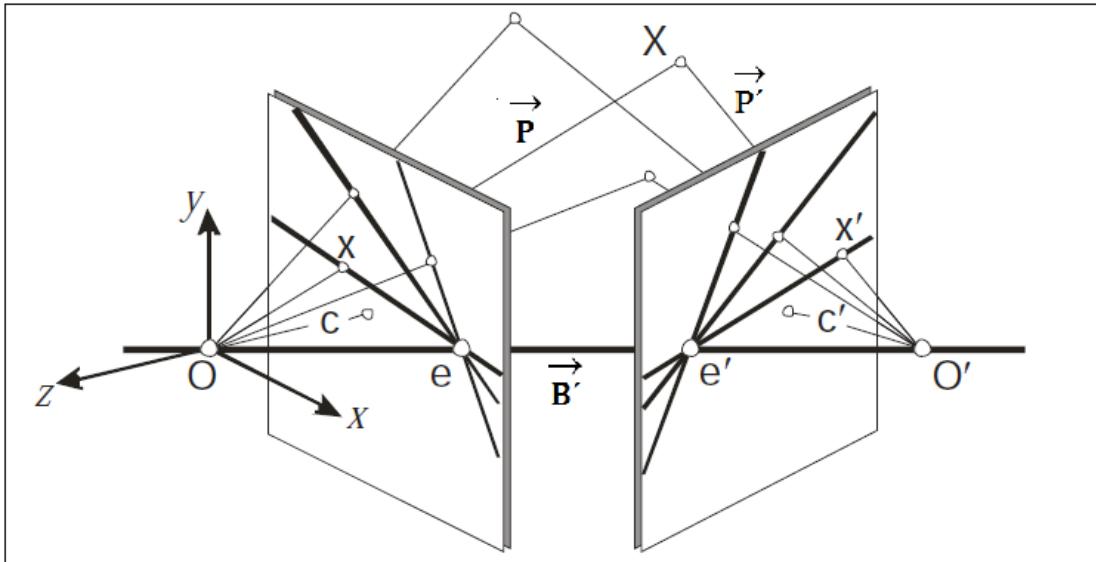
$$\frac{y_p - y_0}{-c} = \frac{R_{21}(X_p - X_0) + R_{22}(Y_p - Y_0) + R_{23}(Z_p - Z_0)}{R_{31}(X_p - X_0) + R_{32}(Y_p - Y_0) + R_{33}(Z_p - Z_0)} \Leftrightarrow$$

$$y_p = y_0 - c \frac{R_{21}(X_p - X_0) + R_{22}(Y_p - Y_0) + R_{23}(Z_p - Z_0)}{R_{31}(X_p - X_0) + R_{32}(Y_p - Y_0) + R_{33}(Z_p - Z_0)}$$
(1.16)

Οι εξισώσεις (1.15) και (1.16) αποτελούν τη συνήθη έκφραση της συνθήκης συγγραμμικότητας.

ΚΕΦΑΛΑΙΟ 2

Επιπολική γεωμετρία



Σχήμα 2.1: Γεωμετρία του στερεοζεύγους (Καλησπεράκης, 2010)

Με βάση το σχήμα 2.1 που απεικονίζει τη διάταξη δύο επικαλυπτόμενων εικόνων, ορίζουμε τα εξής:

- Βάση του στερεοζεύγους, είναι το διάνυσμα που ενώνει τα κέντρα προβολής των δύο εικόνων και συμβολίζεται με το γράμμα B .
- Επιπολικά σημεία ή πόλοι, είναι τα σημεία τομής της βάσης με τα επίπεδα των εικόνων και συμβολίζονται με e και e' στην αριστερή και δεξιά εικόνα αντίστοιχα. Είναι προφανές και από το σχήμα ότι η βάση είναι δυνατό να τέμνει μόνο τη μια εικόνα ή και καμία από τις δύο, οπότε ο ένας ή και οι δύο πόλοι να βρίσκονται έξω από τα επίπεδα των εικόνων (π.χ. στην περίπτωση αεροφωτογραφιών όπου τα επίπεδα των εικόνων είναι σχεδόν παράλληλα στη βάση).
- Επιπολικό επίπεδο, είναι κάθε επίπεδο που αποτελείται από το διάνυσμα της βάσης $\vec{B} = OO'$ και τα διανύσματα $\vec{P} = OX$ και $\vec{P}' = O'X$, όπου X τυχαίο σημείο του αντικειμένου.
- Επιπολικές γραμμές, είναι όλες εκείνες οι ευθείες που σχηματίζονται από την προβολή των επιπολικών επιπέδων πάνω στις εικόνες. Η ιδιαιτερότητα των ευθειών αυτών είναι ότι συγκλίνουν (τέμνονται) στο ίδιο σημείο και συγκεκριμένα στους πόλους, σχηματίζοντας δέσμες.

2.1 Σχετικός προσανατολισμός

Ο σχετικός προσανατολισμός του στερεοζεύγους ορίζεται ως το ζήτημα της αποκατάστασης της σχετικής θέσης και σχετικών στροφών μεταξύ δύο επικαλυπτόμενων εικόνων. Αντιμετωπίζεται κυρίως με προσδιορισμό ομόλογων σημείων και είναι απαραίτητο να έχει γίνει εκ των προτέρων αποκατάσταση της δέσμης για κάθε εικόνα (εσωτερικός προσανατολισμός). Η αποκατάστασή του επιτρέπει την δημιουργία τρισδιάστατου μοντέλου του αντικειμένου σε αυθαίρετο σύστημα και κλίμακα. Οι 5 παράμετροι του σχετικού είναι οι συνιστώσες κατά y και z της βάσης ανηγμένες ως προς την x, δηλαδή: ($b_y = \frac{B_y}{B_x}$ και $b_z = \frac{B_z}{B_x}$) (2.1)

και οι σχετικές στροφές ω, φ και κ της δεξιάς ως προς την αριστερή εικόνα. Λόγω του ότι ο σχετικός δεν απαιτεί καμία γεωμετρική πληροφορία για τον απεικονιζόμενο χώρο, επιτρέπει την πλήρη αξιοποίηση της 'εσωτερικής' ακρίβειας των εικόνων (Πέτσα, 2000, Καλησπεράκης, 2010).

2.1.1 Συνθήκη συνεπιπεδότητας

Κάθε ζεύγος ομόλογων σημείων ορίζει –κατόπιν αποκατάστασης του εσωτερικού προσανατολισμού των εικόνων– στον χώρο δύο ευθείες οι οποίες σχηματίζονται από τα κέντρα προβολής και το σημείο. Οι ευθείες αυτές οφείλουν να τέμνονται και για να συμβαίνει αυτό, όπως ορίζει η συνθήκη συνεπιπεδότητας, πρέπει τα διανύσματα \vec{P} και \vec{P}' και \vec{B} (διάνυσμα βάσης) να ανήκουν στο ίδιο επίπεδο (σχ. 2.1). Αλγεβρικά αυτό εκφράζεται με το μεικτό γινόμενο:

$$\vec{B} \cdot (\vec{P} \times \vec{P}') = 0 \quad (2.2)$$

ή ισοδύναμα

$$\vec{P} \cdot (\vec{B} \times \vec{P}') = 0 \quad (2.3)$$

όπου:

$$B = \begin{bmatrix} B_x \\ B_y \\ B_z \end{bmatrix}, P = \begin{bmatrix} x - x_o \\ y - y_o \\ -c \end{bmatrix} \text{ και } P' = R \begin{bmatrix} x' - x'_o \\ y' - y'_o \\ -c' \end{bmatrix} \quad (2.4)$$

όπου R ο πίνακας στροφής της δεξιάς εικόνας ως προς την αριστερή.

Το εξωτερικό γινόμενο $\vec{P} \times \vec{P}'$ ορίζει διάνυσμα \vec{D} το οποίο είναι κάθετο σε αυτά, άρα και στο συνεπίπεδο διάνυσμα \vec{B} της βάσης, οπότε ισχύει:

$$\vec{B} \cdot \vec{D} = 0 \quad (2.5)$$

Τελικά, το εσωτερικό γινόμενο υπολογίζεται από την ορίζουσα:

$$\begin{vmatrix} B_x & B_y & B_z \\ P_1 & P_2 & P_3 \\ P'_1 & P'_2 & P'_3 \end{vmatrix} = 0 \quad (2.6)$$

Όπου, με βάση τους ορισμούς (2.1) θα πάρει τη μορφή:

$$\begin{vmatrix} 1 & b_y & b_z \\ P_1 & P_2 & P_3 \\ P'_1 & P'_2 & P'_3 \end{vmatrix} = 0 \quad (2.7)$$

2.1.2 Δεσμευμένος επιπολικός πίνακας

Ο δεσμευμένος επιπολικός πίνακας (essential matrix), διατυπώθηκε από τους Longuet και Higgins (1981) στην προσπάθειά τους να δημιουργήσουν ένα αλγόριθμο που να αποκαθιστά τον σχετικό προσανατολισμό δύο εικόνων, χωρίς την απαίτηση αρχικών τιμών (Καλησπεράκης, 2010).

Ο δεσμευμένος επιπολικός πίνακας ορίζεται με την υπόθεση ότι η μηχανή ακολουθεί το μοντέλο της κάμερας μικροσκοπικής οπής (pinhole camera model) ($a=1$, $s=0$). Εφόσον οι πίνακες βαθμονόμησης K και K' είναι γνωστοί, μπορούν να οριστούν οι κανονικοποιημένες συντεταγμένες του ομόλογου ζεύγους ως εξής:

$$\tilde{x} = K^{-1}x = K^{-1} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \sim \begin{bmatrix} x - x_o \\ y - y_o \\ -c \end{bmatrix} \quad \text{και} \quad \tilde{x}' = K'^{-1}x' = K^{-1} \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} \sim \begin{bmatrix} x' - x_o \\ y' - y_o \\ -c \end{bmatrix} \quad (2.8)$$

Με βάση τα παραπάνω, η (2.3) μπορεί να γραφεί ως εξής:

$$\tilde{x} \cdot (B \times R \tilde{x}') = 0 \quad (2.9)$$

η οποία μπορεί να απλοποιηθεί επιπλέον με τη βοήθεια του αντισυμμετρικού πίνακα του εξωτερικού γινομένου, οπότε θα γίνει:

$$\tilde{x}^T [B]_{\times} R \tilde{x}' = 0 \quad (2.10)$$

Ο δεσμευμένος επιπολικός πίνακας ορίζεται ως:

$$E = [B]_{\times} R \quad (2.11)$$

οπότε η (2.10) παίρνει την τελική μορφή:

$$\tilde{x}^T E \tilde{x}' = 0 \quad (2.12)$$

Ο Ε είναι ένας πίνακας διαστάσεων 3x3 ο οποίος έχει 5 βαθμούς ελευθερίας και λόγω του αντισυμμετρικού του μέρους είναι μη αντιστρέψιμος, άρα η ορίζουσά του είναι μηδενική.

Με τη βοήθεια της μεθόδου ανάλυσης ιδιαζουσών τιμών (Singular Value Decomposition), ο Ε μπορεί να εκφραστεί με τη μορφή:

$$E = UDV^T, \text{ όπου } D = \text{diag}(\lambda_1, \lambda_2, \lambda_3), \text{ με } \lambda_1 \geq \lambda_2 \geq \lambda_3 \geq 0 \quad (2.13)$$

Οι $\lambda_1, \lambda_2, \lambda_3$ αποτελούν τις ιδιάζουσες τιμές του πίνακα. Εφόσον ο Ε είναι μη αντιστρέψιμος ισχύει ότι $\lambda_3 = 0$ και επιπλέον, οι Huang & Faugeras (1989) απέδειξαν ότι $\lambda_1 = \lambda_2$ (Καλησπεράκης, 2010).

Συνοψίζοντας λοιπόν, αποδεικνύεται ότι ένας πίνακας 3x3 καλείται δεσμευμένος επιπολικός, αν και μόνο αν δύο ιδιάζουσες τιμές του είναι ίσες και η τρίτη μηδενική.

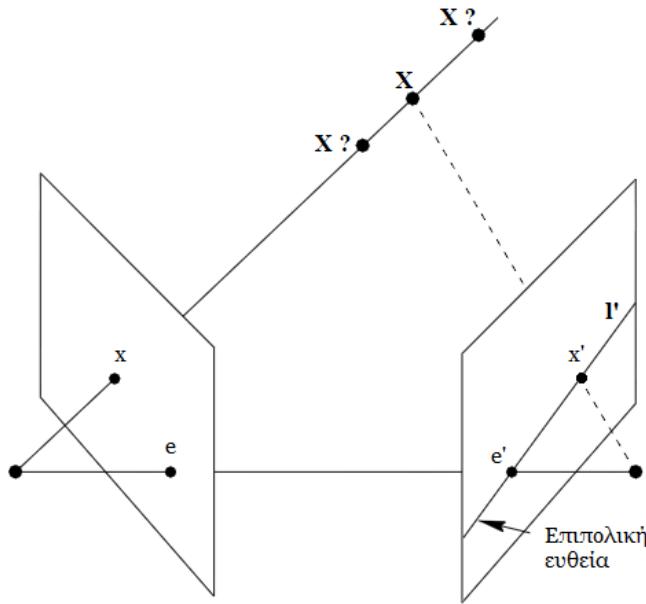
2.2 Δισδιάστατη επιπολική γεωμετρία

Με την εξέλιξη στο πεδίο της όρασης υπολογιστών (κυρίως στα τέλη της δεκαετίας του '80) και την ολοένα και αυξανόμενη ενασχόληση με ζητήματα που μέχρι πρότινος αφορούσαν αποκλειστικά την επιστήμη της φωτογραφίας, παρουσιάστηκε η ανάγκη εξαγωγής τρισδιάστατης πληροφορίας από εικόνες αγνώστου εσωτερικού προσανατολισμού.

Το πρόβλημα αυτό δεν μπορούσε να αντιμετωπιστεί απευθείας στις τρείς διαστάσεις, καθώς σύμφωνα με τις βασικές φωτογραφικές αρχές, για να είναι δυνατόν να οριστεί μια ακτίνα στο χώρο, πρέπει απαραίτητα να είναι γνωστή η εσωτερική γεωμετρία της μηχανής (αποκατάσταση εσωτερικού προσανατολισμού).

Η λύση δόθηκε κυρίως από την έρευνα στο πεδίο της όρασης υπολογιστών, με τη μελέτη της προβολικής σχέσης ανάμεσα στις δέσμες επιπολικών ευθειών απευθείας στο δισδιάστατο σύστημα των εικόνων (Καλησπεράκης, 2010), στηριζόμενη σε μια βασική δέσμευση της επιπολικής γεωμετρίας, αυτή του 'επιπολικού περιορισμού' (epipolar constraint).

Σύμφωνα με αυτή, για κάθε σημείο x της πρώτης εικόνας, το ομόλογο του x' στη δεύτερη εικόνα δεν μπορεί να βρίσκεται οπουδήποτε πάνω σε αυτή, αλλά 'κινείται' αυστηρά πάνω σε μια ευθεία και ειδικότερα, πάνω σε μια επιπολική γραμμή (σχ.2.2).



Σχήμα 2.2 Επιπολικός περιορισμός (Hartley & Zisserman, 2000)

2.2.1 Επιπολικός πίνακας

Η αποκατάσταση της σχέσης μεταξύ των δύο εικόνων, ορίζει επομένως μια μη αντιστρέψιμη ετερογραφία, καθώς σε ένα σημείο της πρώτης εικόνας αντιστοιχεί μια ευθεία της δεύτερης, αλλά σε μια ευθεία της δεύτερης εικόνας δεν αντιστοιχεί ένα σημείο της πρώτης, αλλά μια ευθεία επίσης. Από το τελευταίο διαφαίνεται και η ύπαρξη μιας ομογραφίας που συνδέει αμφιμονοσήμαντα τις ομόλογες επιπολικές ευθείες.

Αυτή η ‘εσωτερική γεωμετρία’ μεταξύ των εικόνων, εκφράζεται αλγεβρικά μέσω του επιπολικού πίνακα F (fundamental matrix) (Faugeras, 1992, Hartley, 1992).

Ο υπολογισμός του F ισοδυναμεί με την εύρεση της προαναφερθείσας ετερογραφίας. Η επιπολική ευθεία (στη δεύτερη εικόνα) ενός σημείου x (στην πρώτη) δίνεται από τη σχέση:

$$l' = Fx \quad (\text{και } l = Fx', \text{ στην αντίθετη περίπτωση}) \quad (2.14)$$

ενώ ο επιπολικός περιορισμός εκφράζεται από τη σχέση:

$$x'^T F x = 0 \quad (2.15)$$

Ο επιπολικός πίνακας έχει διαστάσεις 3×3 και ορίζεται από 8 παραμέτρους, αφού υπάρχει αδυναμία ορισμού κλίμακας (Άδαμ, 2011). Είναι μη αντιστρέψιμος, άρα οφείλει να έχει μηδενική ορίζουσα. Επομένως, οι βαθμοί ελευθερίας του είναι 7 και μπορεί να οριστεί πλήρως από 7 ζεύγη ομόλογων

σημείων. Επιπλέον, σε αντίθεση με τον δεσμευμένο επιπολικό, που χρησιμοποιεί κανονικοποιημένες συντεταγμένες, ο επιπολικός χρησιμοποιεί ομογενείς.

Ο επιπολικός πίνακας συνδέεται με τον δεσμευμένο επιπολικό, του οποίου αποτελεί γενίκευση, με τις παρακάτω σχέσεις:

$$F = K'^{-T} E K^{-1} \Leftrightarrow E = K'^T F K \quad (2.16)$$

Τέλος, τα διανύσματα θέσης των πόλων, προκύπτουν από τις σχέσεις:

$$Fe = 0 \text{ και } F^T e' = 0 \quad (2.17)$$

2.2.2 Υπολογισμός του επιπολικού πίνακα

2.2.2.1 Αλγόριθμος των 8 σημείων (8-point algorithm)

Όπως αναφέρθηκε στην προηγούμενη ενότητα, ο επιπολικός πίνακας οφείλει να ικανοποιεί τη σχέση (2.15) για κάθε ζεύγος ομόλογων σημείων, όπου $x = [x \ y \ 1]^T$ και $x' = [x' \ y' \ 1]^T$ οι ομογενείς συντεταγμένες τους. Η σχέση αυτή μπορεί να εκφραστεί για $n \geq 8$ ομολογίες σημείων, σαν ένα ομογενές σύστημα εξισώσεων, ως εξής:

$$A_{nx9}f = 0 \quad (2.18)$$

Όπου f είναι ο πίνακας F εκφρασμένος σε διανυσματική μορφή:

$$f = [f_{11} \ f_{12} \ f_{13} \ f_{21} \ f_{22} \ f_{23} \ f_{31} \ f_{32} \ f_{33}]^T \quad (2.19)$$

και

$$A = \begin{bmatrix} x_1 x'_1 & y_1 x'_1 & x'_1 & x_1 y'_1 & y_1 y'_1 & y'_1 & x_1 & y_1 & 1 \\ \vdots & \vdots \\ \vdots & \vdots \\ x_n x'_n & y_n x'_n & x'_n & x_n y'_n & y_n y'_n & y'_n & x_n & y_n & 1 \end{bmatrix} \quad (2.20)$$

Για 8 ζεύγη σημείων, το σύστημα έχει μοναδική λύση, ως προς μια πολλαπλασιαστική σταθερά. Για περισσότερα ζεύγη, απαιτείται συνόρθωση με τη μέθοδο των ελαχίστων τετραγώνων.

Για να αρθεί η απειρία λύσεων που συνοδεύει ένα ομογενές σύστημα, υπάρχουν δύο τρόποι (Καλησπεράκης, 2010, Αδάμ, 2011, Κυπαρίσση, 2011).

Ο πρώτος είναι να οριστεί αυθαίρετα ένα στοιχείο του f . Ο δεύτερος είναι να επιβληθεί στον f η δέσμευση της μοναδιαίας νόρμας ($\|f\| = 1$). Ο δεύτερος είναι προτιμότερος, γιατί έτσι εξαλείφει την πιθανότητα να δοθεί τιμή σε ένα στοιχείο του F , που πιθανώς είναι μηδενικό.

Εάν τώρα, γίνει ανάλυση ιδιαζουσών τιμών για τον πίνακα A ($A=UDV^T$), τότε αποδεικνύεται ότι η λύση της M.E.T. δίνεται από την τελευταία στήλη του πίνακα V , οι τιμές της οποίας δίνουν το διάνυσμα f για το οποίο ισχύει ότι:

$$\min \|Af\|^2 = 0 \quad (2.21)$$

2.2.2.2 Κανονικοποιημένος αλγόριθμος των 8 σημείων

Η γραμμική επίλυση μέσω του αλγόριθμου των 8 σημείων, δεν λαμβάνει υπ' όψιν της τη θεμελιώδη δέσμευση της μηδενικής ορίζουσας για τον επιπολικό πίνακα. Επιπλέον, η λύση του συστήματος της (2.18) παρουσιάζει διαφοροποιήσεις ανάλογα με το σύστημα συντεταγμένων που εκφράζονται τα ομόλογα ζεύγη (Κυπαρίσση, 2011) καθώς επίσης, είναι ιδιαιτέρως ευαίσθητη στο θόρυβο που ενυπάρχει στις μετρήσεις. Για το λόγο αυτόν ο Hartley (1997) πρότεινε τις εξής βελτιώσεις:

Καταρχήν, εφαρμόζεται στις συντεταγμένες των ομόλογων σημείων ένας μετασχηματισμός ομοιότητας, τέτοιος ώστε το κέντρο βάρους τους να συμπίπτει με το κέντρο της εικόνας και η μέση απόστασή τους από αυτό να ισούται με $\sqrt{2}$. Αν T και T' οι ζητούμενοι μετασχηματισμοί, τότε οι κανονικοποιημένες συντεταγμένες είναι:

$$x_t = Tx \text{ και } x'_t = T'x' \quad (2.22)$$

Ισχύει επίσης ότι:

$$F = T'F_t T \quad (2.23)$$

Όπου F_t είναι ο επιπολικός πίνακας που υπολογίζεται από τις κανονικοποιημένες συντεταγμένες.

Κατόπιν, προκειμένου να εξασφαλιστεί η δέσμευση της μηδενικής ορίζουσας, γίνεται ανάλυση ιδιαζουσών τιμών του επιπολικού πίνακα, επομένως:

$$F = U \text{diag}(\lambda_1, \lambda_2, \lambda_3) V^T \quad (2.24)$$

Ο F αντικαθίσταται από τον F' , ο οποίος προκύπτει με μηδενισμό της τελευταίας ιδιαζουσας τιμής ($\lambda_3 = 0$), οπότε ισχύει:

$$F' = U \text{diag}(\lambda_1, \lambda_2, 0) V^T \text{ με } \|F' - F\| = \min \quad (2.25)$$

2.2.2.3 Αλγόριθμος των 7 σημείων (7-point algorithm)

Όπως αναφέρθηκε και στην ενότητα 2.2.1, ο επιπολικός πίνακας έχει 7 βαθμούς ελευθερίας, οπότε αρκούν 7 ομόλογα ζεύγη για να υπολογιστεί. Στην περίπτωση αυτή, δημιουργείται ένα ομογενές σύστημα γραμμικών εξισώσεων ίδιο με αυτό της (2.18), όπου ο πίνακας A έχει διαστάσεις 7x9. Στη συνέχεια, μπορούμε να παραμετρικοποιήσουμε τον δεξί μηδενοχώρο του συστήματος ως εξής:

$$F = v_1 + \lambda v_2 = F_1 + \lambda F_2 \quad (2.26)$$

όπου: v_1, v_2 οι δύο τελευταίες στήλες του πίνακα V που προκύπτει από την ανάλυση ιδιαζουσών τιμών του πίνακα A (Hartley, 1994)

και F_1, F_2 οι αντίστοιχοι πίνακες που προκύπτουν από τις v_1, v_2 , με την αντίστροφη διαδικασία που δημιουργήθηκε το διάνυσμα f στην (2.18).

Η δέσμευση της μηδενικής ορίζουσας του επιπολικού πίνακα, οδηγεί σε ένα πολυώνυμο γ' βαθμού ως προς λ. Ισχύει:

$$\det(F) = \det(F_1 + \lambda F_2) = \alpha_3 \lambda^3 + \alpha_2 \lambda^2 + \alpha_1 \lambda + \alpha_0 = 0 \quad (2.27)$$

Από την επίλυση, προκύπτουν το πολύ τρείς πραγματικές λύσεις για τον F.

2.2.2.4 Μη γραμμική επίλυση

Προς αποφυγή των μειονεκτημάτων της γραμμικής επίλυσης, που επισημάνθηκαν στην αρχή της ενότητας 2.2.2.2, έχουν κατά καιρούς προταθεί διάφοροι μη γραμμικοί αλγόριθμοι υπολογισμού του επιπολικού πίνακα. Οι περισσότεροι εξ' αυτών διαφέρουν, πέραν του γεωμετρικού μοντέλου που χρησιμοποιούν, ως προς τη συνάρτηση κόστους που ελαχιστοποιούν (Αδάμ, 2011, Κυπαρίσση, 2011). Για τον ορισμό προσεγγιστικών τιμών των παραμέτρων, χρησιμοποιείται συνήθως ο αλγόριθμος των 8 σημείων.

Ο συνηθέστερος μη γραμμικός αλγόριθμος, χρησιμοποιεί σαν συνάρτηση κόστους το άθροισμα των τετραγώνων των αποστάσεων των σημείων από τις επιπολικές γραμμές, ελαχιστοποιούν δηλαδή την τιμή:

$$\sum_i d^2(x'_i, Fx_i) \quad (2.28)$$

Κατά την ελαχιστοποίηση του παραπάνω αθροίσματος, εισάγονται συστηματικά σφάλματα όσον αφορά τη θέση των πόλων, οι οποίοι τείνουν προς το κέντρο της εικόνας (Καλησπεράκης, 2010). Για το λόγο αυτό προτείνεται η αντικατάστασή του από την ακόλουθη συνάρτηση, που

λαμβάνει υπ' όψιν της, την απόσταση των σημείων από τις επιπολικές τους ευθείες ταυτόχρονα και για τις δύο εικόνες:

$$\sum_i (d^2(x'_i, Fx_i) + d^2(x_i, F'x'_i)) \quad (2.29)$$

2.2.2.5 Αλγόριθμος RANSAC (*R*andom *S*ampling *C*onsensus)

Ένα σημαντικό ζήτημα που προκύπτει και δεν μπορεί να αντιμετωπιστεί από τους παραπάνω αλγόριθμους, είναι η ύπαρξη λανθασμένων αντιστοιχιών που οφείλονται σε χονδροειδή σφάλματα, που αναπόφευκτα υπεισέρχονται στις μετρήσεις κατά τη διαδικασία της αυτόματης συνταύτισης εικόνων (Καλησπεράκης, 2010). Τα σφάλματα αυτά είναι ικανά να οδηγήσουν σε εσφαλμένη ή και σε αδύνατη λύση κατά τον υπολογισμό του επιπολικού πίνακα. Κατά καιρούς έχουν προταθεί διάφοροι αλγόριθμοι ανίχνευσης και απομάκρυνσης χονδροειδών σφαλμάτων. Ο πιο δημοφιλής όλων είναι ο αλγόριθμος Ransac των Fischler και Bolles (1981), που πρώτοι οι Torr και Murray (1993) χρησιμοποίησαν για να υπολογίσουν τον επιπολικό πίνακα.

Ο Ransac είναι ένας επαναληπτικός αλγόριθμος που στηρίζεται σε στατιστικές μεθόδους, οπότε είναι μη ντετερμινιστικά δομημένος, καθώς δίνει αποτέλεσμα με συγκεκριμένη πιθανότητα (Φάκα, 2010).

Πιο αναλυτικά, ο αλγόριθμος Ransac περιλαμβάνει τις ακόλουθες διαδικασίες:

- Από ένα σύνολο ομόλογων ζευγών επιλέγονται τυχαία 7 αντιστοιχίες, οι ελάχιστες δηλαδή που απαιτούνται για τον υπολογισμό του επιπολικού πίνακα. Οι αντιστοιχίες αυτές θεωρούνται καταρχήν σωστές (inliers). Να σημειωθεί εδώ ότι παρουσιάζεται ο αλγόριθμος για την περίπτωση που ακολουθείται το μαθηματικό μοντέλο του αλγορίθμου των 7 σημείων, επειδή αυτή η εκδοχή χρησιμοποιήθηκε στις εφαρμογές της εργασίας. Η ίδια διαδικασία ακολουθείται και για τον αλγόριθμο των 8 σημείων.
- Με βάση τις 7 ομολογίες υπολογίζεται ο επιπολικός πίνακας (ή τρείς το πολύ πίνακες, ανάλογα με τις λύσεις που θα προκύψουν).
- Όλα τα υπόλοιπα σημεία εξετάζονται μέσω της Sampson απόστασής τους από τις ομόλογες επιπολικές ευθείες, που προέκυψαν από τον επιπολικό πίνακα που υπολογίστηκε στο προηγούμενο βήμα. Αν η απόσταση για κάθε ζεύγος σημείων, δεν ξεπερνά ένα ορισμένο κατώφλι (που συνήθως εκφράζεται σε ποσοστό επί της διάστασης του pixel), τότε η αντιστοιχία θεωρείται inlier, ενώ σε αντίθετη περίπτωση θεωρείται outlier (λανθασμένη).

- Όλες οι παραπάνω διαδικασίες εκτελούνται για k επαναλήψεις.
- Επικρατέστερη λύση για τον F επιλέγεται εκείνη που επιστρέφει τις περισσότερες inliers αντιστοιχίες.
- Υπολογίζεται εκ νέου ο επιπολικός πίνακας από το σύνολο των inliers του προηγούμενου βήματος. Είναι προφανές ότι σε αυτή τη φάση λόγω της περίσσειας δεδομένων, δεν μπορεί να χρησιμοποιηθεί ο αλγόριθμος των 7 σημείων, οπότε γίνεται συνόρθωση με τη M.E.T. χρησιμοποιώντας είτε τους αλγόριθμους των 8 σημείων ή κάποια μη γραμμική μέθοδο.

Η πιθανότητα p να έχει υπολογιστεί ο επιπολικός μόνο από (πραγματικές) inliers αντιστοιχίες, ύστερα από k επαναλήψεις, δίνεται από την εξίσωση:

$$k = \frac{\log(1-p)}{\log(1-w^n)} \quad (2.30)$$

όπου n είναι ο ελάχιστος απαιτούμενος αριθμός ομολογιών για τον υπολογισμό του επιπολικού πίνακα και w η εκτίμηση του ποσοστού των inliers στα δεδομένα:

$$w = \frac{\text{inliers}}{\text{σύνολο ομολογιών}} \quad (2.31)$$

ΚΕΦΑΛΑΙΟ 3

Ψηφιακή συνταύτιση

Με τον όρο ‘ψηφιακή συνταύτιση’ αναφέρεται στη βιβλιογραφία το ζήτημα της αυτόματης αποκατάστασης ομόλογων σημείων μεταξύ δύο ή περισσοτέρων επικαλυπτόμενων εικόνων (Άδαμ, 2011, Κυπαρίσση, 2011). Η σημασία της είναι μεγάλη και στη φωτογραμμετρία, καθώς στην αποκατάσταση ομολογιών στηρίζονται αρκετές φωτογραμμετρικές εργασίες όπως ο σχετικός προσανατολισμός, η μέθοδος της δέσμης, ο υπολογισμός του επιπολικού πίνακα κ.α.

Το πρόβλημα της συνταύτισης είναι ‘έλλιπως ορισμένο’ (Heipke, 1992, Καρφής, 2009) καθώς:

- Είναι δυνατό να μην έχει λύση, εξαιτίας της ύπαρξης αποκρύψεων στις εικόνες.
- Είναι δυνατό να μην έχει ενιαία λύση, εξαιτίας επαναλαμβανόμενων προτύπων στις εικόνες.
- Είναι δυνατό να μην έχει σταθερή λύση, εξαιτίας χαμηλής υφής ή ύπαρξης θορύβου στην εικόνα.

Προκειμένου να συσχετιστούν δύο οντότητες, χρησιμοποιείται κάποιο μέτρο ομοιότητας (π.χ. άθροισμα των απολύτων διαφορών, ευκλείδεια απόσταση κ.α.). Όταν η συνταύτιση γίνεται με μέτρα που λαμβάνουν υπ’ όψιν τους τις τιμές φωτεινότητας σε δύο περιοχές των δύο εικόνων, τότε αποκαλείται επιφανειακή συνταύτιση (area-based matching), ενώ όταν συγκρίνονται χαρακτηριστικές δομές όπως για παράδειγμα σημεία ή ακμές, τότε αποκαλείται συνταύτιση χαρακτηριστικών(feature-based matching).

3.1 Επιφανειακή συνταύτιση

Η βασική αρχή λειτουργίας της μεθόδου επιφανειακής συνταύτισης είναι η εξής: Δημιουργείται στην πρώτη εικόνα ένα παράθυρο αναφοράς γύρω από το υπό αντιστοίχιση pixel (εκ των πραγμάτων, διαστάσεων $(2N+1) \times (2N+1)$), που είναι σταθερό και λειτουργεί ως πρότυπο. Στη δεύτερη εικόνα, δημιουργείται ένα παράθυρο αντιστοίχισης ίσων διαστάσεων με το πρότυπο, το οποίο κινείται μέσα σε μια καθορισμένη περιοχή και με τη χρήση ενός μέτρου ομοιότητας αναζητά την περιοχή που ταιριάζει περισσότερο με το πρότυπο. Η περιοχή αναζήτησης θεωρητικά, μπορεί να είναι ολόκληρη η

εικόνα, μπορεί όμως να περιοριστεί ανάλογα με τη γνώση που έχουμε για τη σχέση των εικόνων, μειώνοντας έτσι και το υπολογιστικό κόστος. Έτσι, αν είναι γνωστό, για παράδειγμα, το ποσοστό επικάλυψης των εικόνων η περιοχή καθορίζεται σύμφωνα με αυτό, ενώ, στην περίπτωση που έχουμε επιπολικές εικόνες, η αναζήτηση περιορίζεται σε μία 'λωρίδα' περί της ομόλογης επιπολικής γραμμής. Επίσης, οι διαστάσεις του προτύπου θα πρέπει να είναι τέτοιες που να εξασφαλίζουν την ύπαρξη υφής (Κυπαρίσση, 2011). Για όλα τα παραπάνω, υπάρχει πληθώρα αναφορών στη βιβλιογραφία, δεν είναι όμως μέλημα της παρούσας εργασίας να μελετηθούν εκτενέστερα.

Μερικά από τα μέτρα ομοιότητας που χρησιμοποιούνται στην επιφανειακή συνταύτιση είναι τα εξής (όπου $g_1(i, j)$ και $g_2(i, j)$ είναι οι τιμές φωτεινότητας στη θέση (i, j) των δύο εικόνων αντίστοιχα):

- Άθροισμα των Απολύτων Διαφορών (Sum of Absolute Differences)

$$\Delta = \sum_{i=1}^N \sum_{j=1}^N |g_1(i, j) - g_2(i, j)| \quad (3.1)$$

- Άθροισμα των Ανηγμένων Απολύτων Διαφορών (Zero-mean Sum of Absolute Differences)

$$\Delta = \sum_{i=1}^N \sum_{j=1}^N |(g_1(i, j) - \bar{g}_1) - (g_2(i, j) - \bar{g}_2)| \quad (3.2)$$

- Άθροισμα των Τετραγώνων Διαφορών (Sum of Squared Differences)

$$\Delta = \sum_{i=1}^N \sum_{j=1}^N (g_1(i, j) - g_2(i, j))^2 \quad (3.3)$$

- Άθροισμα των Ανηγμένων Τετραγώνων Διαφορών (Zero-mean Sum of Squared Differences)

$$\Delta = \sum_{i=1}^N \sum_{j=1}^N ((g_1(i, j) - \bar{g}_1) - (g_2(i, j) - \bar{g}_2))^2 \quad (3.4)$$

- Κανονικοποιημένος Συντελεστής Συσχέτισης (ανηγμένη έκφραση) (Zero-mean Normalized Cross Correlation)

$$\rho = \frac{\sigma_{12}}{\sigma_1 \sigma_2} = \frac{\sum_{i=1}^N \sum_{j=1}^N (g_1(i, j) - \bar{g}_1)(g_2(i, j) - \bar{g}_2)}{\sqrt{\sum_{i=1}^N \sum_{j=1}^N (g_1(i, j) - \bar{g}_1)^2 \sum_{i=1}^N \sum_{j=1}^N (g_2(i, j) - \bar{g}_2)^2}} \quad (3.5)$$

όπου σ_1 , σ_2 οι τυπικές αποκλίσεις των παραθύρων αναφοράς και αντιστοίχισης αντίστοιχα και σ_{12} η συμμεταβλητότητά τους.

Ο κανονικοποιημένος συντελεστής συσχέτισης είναι το συνηθέστερο μέτρο ομοιότητας που χρησιμοποιείται στην επιφανειακή συνταύτιση. Λαμβάνει υπ' όψιν του τη σχετική μεταβολή φωτεινότητας εντός των παραθύρων, πράγμα που σημαίνει ότι παραμένει αμετάβλητος σε ενιαίες ραδιομετρικές μεταβολές στις εικόνες. Το πεδίο ορισμού του είναι το $[-1,1]$ και όσο αυτός τείνει προς τις ακραίες τιμές του, τόσο ισχυρότερη είναι η συσχέτιση (θετική ή αρνητική).

Μια εξίσου διαδεδομένη μέθοδος είναι και η ελαχιστοτετραγωνική συνταύτιση (Least Squares Matching). Η βασική της αρχή είναι η ελαχιστοποίηση των διαφορών φωτεινότητας (άθροισμα των τετραγώνων διαφορών) μεταξύ του παραθύρου αναφοράς και του παραθύρου αντιστοίχισης. Επίσης, στην ελαχιστοτετραγωνική συνταύτιση το παράθυρο αντιστοίχισης υπόκειται σε δύο μετασχηματισμούς. Ένα ραδιομετρικό, για να αντισταθμιστούν οι διαφορές στη φωτεινότητα και την αντίθεση των εικόνων κι ένα γεωμετρικό, προκειμένου να προβλεφθεί η θέση, ο προσανατολισμός και το σχήμα του παραθύρου στη δεύτερη εικόνα. Προκειμένου να καλυφθεί η ανάγκη για προσεγγιστικές τιμές, χρησιμοποιείται συχνά ο συντελεστής συσχέτισης.

Επιπλέον, ένα πλεονέκτημα της μεθόδου αυτής έναντι του συντελεστή συσχέτισης (που πρωτογενώς δίνει εκτίμηση ενός pixel), είναι η μεγάλη ακρίβεια συνταύτισης που μπορεί να υπερβεί ακόμα και το 0.1 pixel.

3.2 Συνταύτιση χαρακτηριστικών και τελεστές ανίχνευσης σημείων ενδιαφέροντος

Η συνταύτιση χαρακτηριστικών προσεγγίζει το πρόβλημα μελετώντας, όχι απλά χρωματικούς τόνους, αλλά χαρακτηριστικές δομές μεταξύ των εικόνων, όπως σημεία, ευθείες, γραμμές ή περιοχές. Σε σχέση με την επιφανειακή συνταύτιση είναι πιο ανθεκτική σε γεωμετρικές και ραδιομετρικές μεταβολές. Επίσης, είναι λιγότερο απαιτητική υπολογιστικά, καθώς η αναζήτηση ομολογιών δεν γίνεται σε ολόκληρες περιοχές ή όλη την εικόνα, αλλά περιορίζεται μεταξύ δομών που συγκεντρώνουν ορισμένα χαρακτηριστικά (οι δομές που θα μας απασχολήσουν εδώ είναι τα σημεία).

Η συνταύτιση χαρακτηριστικών γίνεται σε δύο στάδια. Στο πρώτο, ανιχνεύονται σε κάθε εικόνα χωριστά οι χαρακτηριστικές δομές, που θα αντιστοιχηθούν. Συγκεκριμένα, για την εξαγωγή σημείων, χρησιμοποιούνται οι λεγόμενοι τελεστές ανίχνευσης σημείων ενδιαφέροντος (point operators). Κάποιοι από τους τελεστές αυτούς παρουσιάζονται στις επόμενες παραγράφους.

Στο δεύτερο στάδιο, γίνεται η αντιστοίχιση των σημείων. Αυτή συνήθως γίνεται με κριτήριο την ελαχιστοποίηση κάποιας συνάρτησης κόστους μεταξύ των χαρακτηριστικών των σημείων (ευκλείδειες αποστάσεις, τετράγωνα

αποστάσεων κ.α.), αν και μπορεί να γίνει και με μεθόδους επιφανειακής συνταύτισης, με τη δημιουργία ενός παραθύρου αναφοράς γύρω από τα σημεία που ανιχνεύθηκαν στο προηγούμενο στάδιο (Αδάμ, 2011).

3.2.1 Αλγόριθμος Moravec

Ο Moravec ήταν από τους πρώτους που ανέπτυξαν αλγόριθμο ανίχνευσης σημείων, εισάγοντας παράλληλα για πρώτη φορά τον όρο ‘σημεία ενδιαφέροντος’ (points of interest) (Καρφής, 2009). Ο τελεστής Moravec ορίζει σαν κρίσιμη οντότητα το pixel και για να το χαρακτηρίσει σαν σημείο ενδιαφέροντος εξετάζει τη μεταβολή της έντασης του τόνου του γκρίζου στις 8 κύριες διευθύνσεις γύρω από αυτό.

Αναλυτικότερα, δημιουργούνται δύο παράθυρα διαστάσεων 3x3 pixel, ένα που λειτουργεί σαν πρότυπο και περιβάλλει το υπό εξέταση pixel κι ένα παράθυρο αναζήτησης που κινείται γύρω από αυτό, έχοντας για κέντρο του κάθε ένα από τα 8 γειτονικά pixel. Για τις 8 αυτές διευθύνσεις υπολογίζει κάθε φορά το άθροισμα των τετραγώνων των διαφορών (SSD) του τόνου του γκρίζου των αντίστοιχων pixel στα δύο παράθυρα, ως εξής:

$$V(x, y)_{(u, v)} = \sum_{i=-1}^1 \sum_{j=-1}^1 (I(x+i, y+j) - I(x+u+i, y+v+j))^2 \quad (3.6)$$

όπου (u, v) είναι οι 8 μετατοπίσεις του παραθύρου αναζήτησης:

$$(0,1), (1,1), (1,0), (1,-1), (0,-1), (-1,-1), (-1,0), (-1,1)$$

Όσο μεγαλύτερη είναι η τιμή του V , τόσο μεγαλύτερη είναι και η μεταβολή του τόνου, άρα πιθανότερο το σημείο να είναι ‘γωνιακό’. Κριτήριο για αυτό λοιπόν, θα είναι η ελάχιστη τιμή του V , που αποκαλείται ‘μέτρο γωνιακότητας’ (Καρφής, 2009) και συμβολίζεται ως:

$$C(x, y) = \min (V(x, y)_{(u, v)}) \quad (3.7)$$

Αυτό που απομένει για να χαρακτηριστεί ένα pixel σημείο ενδιαφέροντος είναι να ορίσουμε ένα ‘κατώφλι’, δηλαδή μια τιμή την οποία θα πρέπει να υπερβαίνει η ποσότητα $\text{min}V$.

Είναι σαφές ότι θα εμφανιστούν ως άκρα pixel τα οποία είναι γειτονικά, πράγμα που στην πράξη σημαίνει ότι αναφέρονται στο ίδιο σημείο. Επειδή όμως ο αλγόριθμος πρέπει να ‘δώσει’ μόνο ένα pixel ως άκρο, εκτελείται η διαδικασία της ‘αφαίρεσης των μη μεγίστων’, που είναι και το τελικό στάδιο του αλγόριθμου. Κατασκευάζει πρώτα ένα πίνακα (χάρτη γωνιακότητας), υπολογίζοντας για κάθε pixel την τιμή του C κι εν συνεχείᾳ κρατάει όσες τιμές

υπερβαίνουν το κατώφλι και μηδενίζει τις υπόλοιπες. Έπειτα, στις περιοχές που συγκεντρώνονται μη μηδενικές τιμές εντοπίζει αυτή με τη μέγιστη τιμή και μηδενίζει όλες τις υπόλοιπες, δίνοντας έτσι το ζητούμενο pixel (άκρο).

Ο αλγόριθμος Moravec, λόγω του ότι υπολογίζει τις μεταβολές των εντάσεων μόνο στις 8 κύριες διευθύνσεις, επηρεάζεται σημαντικά από το σύστημα αναφοράς της εικόνας (Καρφής, 2009). Αυτό πρακτικά σημαίνει ότι εάν μια εικόνα στραφεί, ο αλγόριθμος θα εξάγει σε μεγάλο βαθμό, διαφορετικά σημεία απ' ότι εξήγαγε αρχικά (φαινόμενο ανισοτροπίας). Επίσης, είναι ιδιαίτερα ευαίσθητος στην ύπαρξη μεμονωμένου θορύβου, καθώς τα σημεία αυτά παρουσιάζουν έντονες μεταβολές σε όλες τις διευθύνσεις.

3.2.2 Αλγόριθμος Harris

Ο συγκεκριμένος αλγόριθμος ανίχνευσης γωνιών προτάθηκε από τους Harris και Stephens το 1988. Ο αλγόριθμος αυτός χρησιμοποιεί τη συνάρτηση αυτοσυσχέτισης, δηλαδή της μεταβολής των εντάσεων γύρω από ένα pixel και είναι η εξής:

$$V(x, y)_{(u, v)} = \sum \left(u \frac{\partial I}{\partial x} + v \frac{\partial I}{\partial y} \right)^2 \quad (3.8)$$

Από τη σχέση αυτή προκύπτει ο πίνακας αυτοσυσχέτισης:

$$M = \begin{bmatrix} \frac{\partial I}{\partial x} & \frac{\partial I}{\partial x} \frac{\partial I}{\partial y} \\ \frac{\partial I}{\partial x} \frac{\partial I}{\partial y} & \frac{\partial I}{\partial y} \end{bmatrix} \quad (3.9)$$

Επιπλέον, στο αλγόριθμο Harris εφαρμόζεται μέσω της διαδικασίας της συνέλιξης, ένα φίλτρο γκαουσιανής εξομάλυνσης της εικόνας, γεγονός που τον καθιστά ανθεκτικότερο στην ύπαρξη θορύβου σε σχέση με τον αλγόριθμο του Moravec.

Αποδεικνύεται, ότι για να χαρακτηριστεί ένα pixel σημείο ενδιαφέροντος, αρκεί να εξεταστούν οι ιδιοτιμές του πίνακα M ως προς τις μέγιστες κι ελάχιστες τιμές τους.

- Εάν οι δύο ιδιοτιμές έχουν μικρή τιμή, σημαίνει ότι η περιοχή είναι σχετικά ομοιογενής, άρα το pixel δεν αποτελεί χαρακτηριστικό σημείο.

- Εάν η μια ιδιοτιμή είναι μεγάλη και η άλλη μικρή, τότε το υπό εξέταση pixel βρίσκεται πάνω σε ακμή.
- Τέλος, εάν οι δύο ιδιοτιμές έχουν σχετικά μεγάλη τιμή, τότε το pixel θεωρείται σημείο ενδιαφέροντος, καθώς θα εμφανίζει έντονη μεταβολή σε όλες τις διευθύνσεις.

Επειδή ο υπολογισμός των ιδιοτιμών έχει μεγάλο υπολογιστικό κόστος (Κελασίδη, 2009) οι Harris και Stephens πρότειναν το εξής μέτρο:

$$R = \det(M) - k(\text{tr}(M))^2 = \alpha\beta - \kappa(\alpha + \beta) \quad (3.10)$$

όπου: α , β είναι οι ιδιοτιμές του πίνακα M και k είναι σταθερά, η οποία καθορίζεται εμπειρικά.

Για την αποφυγή του υπολογισμού του k , εναλλακτικά χρησιμοποιείται το μέτρο που πρότεινε η Alison Noble το 1989:

$$R = \frac{\det(M)}{\text{tr}(M)} \quad (3.11)$$

Παρ' όλο που ο αλγόριθμος Harris, αντιμετώπισε αποτελεσματικότερα ορισμένα προβλήματα του Moravec (π.χ. θόρυβο στις εικόνες), εντούτοις παραμένει κι αυτός περιστροφικά ασταθής, γιατί αν και μπορεί να περιγράψει τη μεταβολή της έντασης σε όλες τις διευθύνσεις, στηρίζεται γι' αυτό στον υπολογισμό της οριζόντιας και κατακόρυφης κλίσης της εικόνας.

3.2.3 Αλγόριθμος SIFT

Με την εξέλιξη της επιστήμης της όρασης των υπολογιστών και της χρήσης της σε όλο και περισσότερες εφαρμογές, εντάθηκε και η ανάγκη να αναπτυχθούν αλγόριθμοι που να είναι όσο το δυνατόν ανθεκτικότεροι στην ύπαρξη γεωμετρικών και ραδιομετρικών μεταβολών. Ο Lowe (2004) παρουσίασε τον αλγόριθμο SIFT (Scale Invariant Feature Transform), ο οποίος χρησιμοποιείται ευρύτατα σε απαιτητικές εφαρμογές ψηφιακής συνταύτισης εικόνων. Ο αλγόριθμος αυτός, όπως μαρτυρά κι η ονομασία του, είναι ιδιαίτερα ανθεκτικός στις μεταβολές κλίμακας στις εικόνες, καθώς επίσης και στις στροφές και μερικώς στις αλλαγές φωτεινότητας.

Σε αντίθεση με τους προαναφερθέντες αλγόριθμους που εξάγουν μόνο χαρακτηριστικά σημεία, ο SIFT συνοδεύει κάθε σημείο που εξάγει με ένα διάνυσμα περιγραφής (Descriptor) των μεταβολών της έντασης, γύρω από το χαρακτηριστικό σημείο. Το γεγονός αυτό, μεταξύ άλλων, περιορίζει κατά πολύ το υπολογιστικό κόστος κατά τη διαδικασία της συνταύτισης, καθώς η εύρεση

ομολογιών δεν γίνεται πλέον απ' ευθείας στις εικόνες, αλλά περιορίζεται σε υπολογισμούς μεταξύ 'πινάκων'.

3.2.3.1 Ανίχνευση χαρακτηριστικών σημείων

Η βασική ιδέα του αλγορίθμου είναι να εντοπίσει χαρακτηριστικά που μένουν αναλλοίωτα σε αλλαγές κλίμακας της εικόνας (Κυπαρίσση, 2011). Για το λόγο αυτό η αναζήτηση γίνεται σε πυραμίδες εικόνων (σχ. 3.1), που προκύπτουν μέσω συνέλιξης της εικόνας με Gaussian φίλτρο εξομάλυνσης. Ισχύει:

$$L(x, y, \sigma) = G(x, y, \sigma) \otimes I(x, y) \quad (3.12)$$

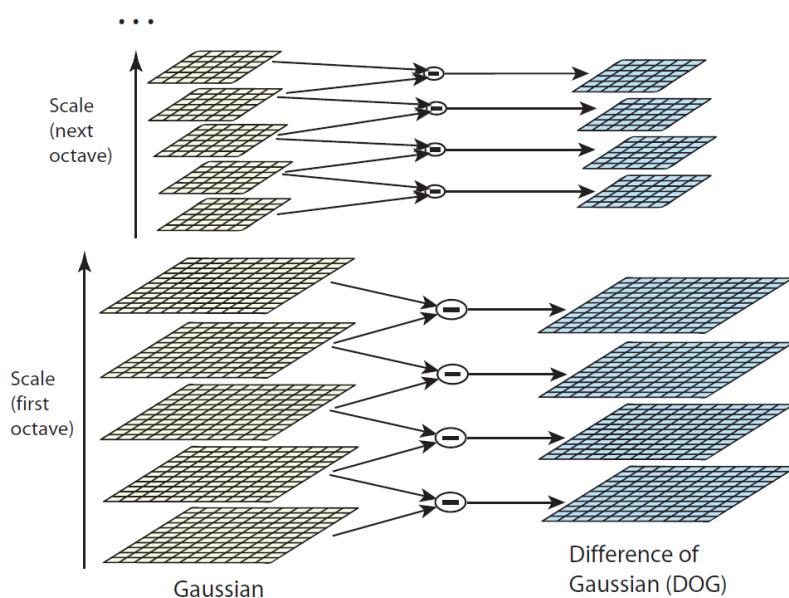
όπου:

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

Αναλυτικότερα, η αρχική εικόνα συνελλίσεται με φίλτρα Gauss διαφόρων τυπικών αποκλίσεων, παράγοντας έτσι ένα μπλοκ εικόνων (οκτάβα), οι οποίες διαδοχικά διαφέρουν μεταξύ τους κατά ένα σταθερό συντελεστή k . Στη συνέχεια, μέσω της επιλογής κάθε δεύτερου pixel, η εικόνα υποδιπλασιάζεται και ακολουθώντας την ίδια διαδικασία, δημιουργείται μια νέα οκτάβα. Στην πράξη συνήθως αρκεί να δημιουργηθεί μια πυραμίδα από 4 οκτάβες, που η κάθε μια να αποτελείται από 5 εικόνες.

Ακολουθεί η δημιουργία μιας νέας πυραμίδας (DoG), η οποία προκύπτει από την αφαίρεση διαδοχικών εικόνων κάθε οκτάβας.

$$D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma) \quad (3.13)$$

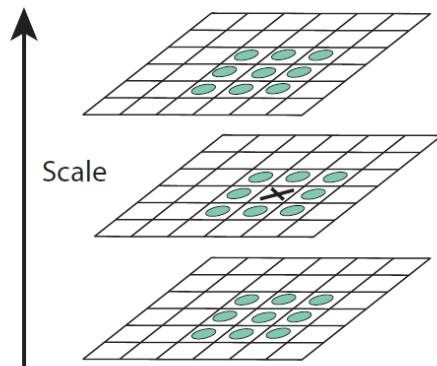


Σχήμα 3.1 Πυραμίδες Gaussian και DoG (Lowe, 2004)

Η διαδικασία ανίχνευσης σημείων γίνεται στην πυραμίδα DoG, καθώς στην πυραμίδα αυτή, εφόσον οι εικόνες διαφέρουν κατά σταθερό συντελεστή k , η εύρεσή τους είναι ανεξάρτητη της κλίμακας της εικόνας (Αδάμ, 2011).

$$G(x, y, k\sigma) - G(x, y, \sigma) \approx (k - 1)\sigma^2 \nabla^2 G^2 \quad (3.14)$$

Ένα pixel συγκρίνεται με τα 8 γειτονικά, καθώς και με τα 9 γειτονικά του στην προηγούμενη και την επόμενη εικόνα. Αν αποτελεί μέγιστο ή ελάχιστο σε σχέση με τα 26 γειτονικά pixel, θεωρείται χαρακτηριστικό σημείο (σχήμα 3.2).



Σχήμα 3.2 Διαδικασία ανίχνευσης χαρακτηριστικών σημείων

Μετά την εξαγωγή των χαρακτηριστικών σημείων γίνεται ο υπολογισμός της θέσης τους με υποψηφιδική ακρίβεια. Η παρεμβολή γίνεται χρησιμοποιώντας β' βαθμού πολυώνυμο taylor στη συνάρτηση $D(x, y, \sigma)$. Η πληροφορία αυτή επιτρέπει, εκτός των άλλων, να απορριφθούν σημεία με χαμηλή αντίθεση, καθώς είναι ιδιαίτερα ευαίσθητα στην ύπαρξη θορύβου.

$$D(x) = D + \frac{\partial D^T}{\partial x} x + \frac{1}{2} x^T \frac{\partial^2 D}{\partial x^2} x \quad (3.15)$$

Βρίσκοντας την παράγωγο της συνάρτησης $D(x)$ και θέτοντας τη ίση με μηδέν, μπορεί να υπολογιστεί η απόσταση του 'πραγματικού' σημείου από το υπό εξέταση pixel. Ισχύει:

$$\hat{x} = - \frac{\partial^2 D^{-1}}{\partial x^2} \frac{\partial D}{\partial x} \quad (3.16)$$

Αν το \hat{x} είναι μεγαλύτερο από 0.5 το σημείο βρίσκεται εγγύτερα σε άλλο pixel.

Για να απορρίψουμε τα σημεία με χαμηλή αντίθεση, χρησιμοποιούμε τη συνάρτηση που προκύπτει αν στην (3.15) αντικαταστήσουμε την (3.16), οπότε έχουμε:

$$D(\hat{x}) = D + \frac{\partial D^T}{\partial x} \hat{x} \quad (3.17)$$

Τα σημεία για τα οποία ισχύει $|D(\hat{x})| < 0.03$ απορρίπτονται.

Τέλος, για τον εντοπισμό και απομάκρυνση των σημείων που βρίσκονται πάνω σε ακμές χρησιμοποιείται ο Εσσιανός πίνακας (Hessian matrix):

$$H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{yx} & D_{yy} \end{bmatrix} \quad (3.18)$$

Όπως και στον αλγόριθμο Harris, δεν απαιτείται ο υπολογισμός των ιδιοτιμών του πίνακα, αλλά εξετάζεται η σχέση μεταξύ της ορίζουσας και του ίχνους του.

3.2.3.2 Ανάθεση προσανατολισμού

Με την ανάθεση κυρίαρχου προσανατολισμού στα χαρακτηριστικά σημεία με βάση τις τοπικές ιδιότητες της εικόνας, ο αλγόριθμος καθίσταται ιδιαίτερα ανθεκτικός στις στροφές μεταξύ των εικόνων, αφού με βάση αυτόν δημιουργείται ο descriptor (§ 3.2.3.3) κάθε σημείου.

Συγκεκριμένα, αφού καθοριστεί μια περιοχή γύρω από το σημείο, υπολογίζεται για κάθε ένα pixel αυτής το διανυσματικό του μέτρο m και ο προσανατολισμός του θ με βάση τις εξισώσεις:

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2} \quad (3.19)$$

$$\theta(x, y) = \tan^{-1} (L(x, y+1) - L(x, y-1)) / (L(x+1, y) - L(x-1, y)) \quad (3.20)$$

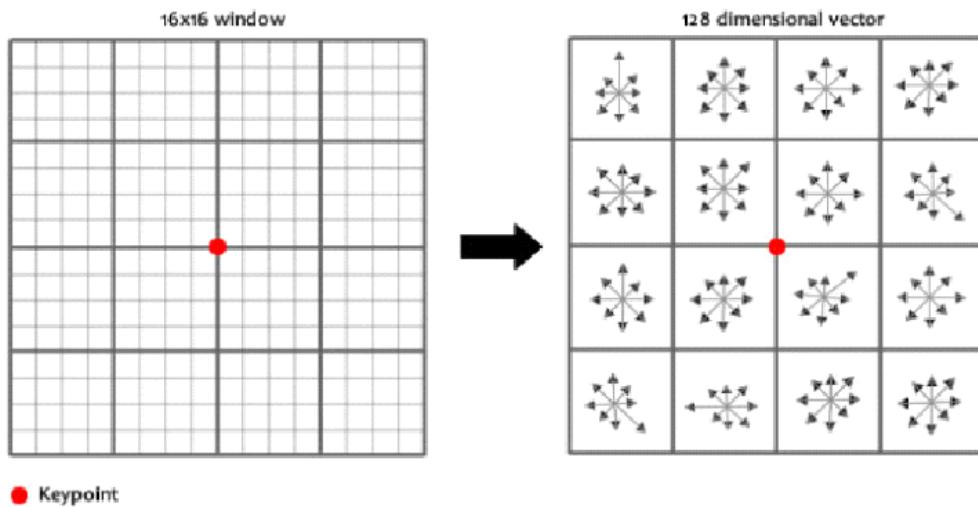
Στη συνέχεια, κατασκευάζεται ένα ιστόγραμμα προσανατολισμού, το οποίο χωρίζεται σε 36 περιοχές εύρους 10° η καθεμία και παριστάνει τους προσανατολισμούς των σημείων σταθμισμένους ανάλογα με το μέτρο τους. Η επικρατέστερη τιμή του ιστογράμματος ορίζει τον προσανατολισμό του σημείου. Σε περίπτωση που στο ιστόγραμμα εμφανίζονται μια ή περισσότερες τιμές που υπερβαίνουν το 80% της επικρατούσας τιμής, τότε ορίζονται περισσότερα χαρακτηριστικά σημεία, τα οποία έχουν ίδια θέση, αλλά διαφορετικό προσανατολισμό.

3.2.3.3 Δημιουργία των Descriptors

Για τη δημιουργία των descriptors ορίζεται μια περιοχή 4×4 περί το σημείο ενδιαφέροντος, στο οποίο υπολογίζεται για κάθε pixel το διανυσματικό του μέτρο, -με προσθήκη βάρους ανάλογα με την απόσταση του από το κέντρο του παραθύρου- στις 8 κύριες διευθύνσεις (σχήμα 3.3). Δημιουργείται έτσι ένα διάνυσμα (πίνακας) διαστάσεων 128×1 ($4 \times 4 \times 8 = 128$). Το διάνυσμα αυτό στρέφεται κατά τον προσανατολισμό που υπολογίστηκε στο προηγούμενο βήμα, ώστε να είναι ανεξάρτητο από τις στροφές των εικόνων. Τέλος, το διάνυσμα κανονικοποιείται, υπό τη δέσμευση της μοναδιαίας ευκλείδειας νόρμας, δηλαδή:

$$\|d\| = \sqrt{\sum_{i=1}^{128} |d_i|^2} = 1 \quad (3.21)$$

Η κανονικοποίηση αυτή καθιστά τον αλγόριθμο περισσότερο ανθεκτικό σε ομοιογενείς αλλαγές στη φωτεινότητα και την αντίθεση (Καλησπεράκης, 2010).



Σχήμα 3.3 Δημιουργία του Descriptor

3.2.3.4 Συνταύτιση χαρακτηριστικών

Για να πραγματοποιηθεί η αποκατάσταση των ομόλογων σημείων, που είναι και το τελευταίο στάδιο του αλγόριθμου, ο Lowe προτείνει την τεχνική του ‘πλησιέστερου γείτονα’. Οι Descriptors αντιστοιχίζονται με βάση την ευκλείδεια απόστασής τους. Εκείνοι που έχουν την ελάχιστη μεταξύ τους απόσταση κι εφόσον αυτή δεν ξεπερνά ένα ορισμένο ανώφλι, αντιστοιχούν σε

ομόλογα σημεία. Ένας επιπλέον έλεγχος για την αποτελεσματικότητα της συνταύτισης είναι η εύρεση του δεύτερου πλησιέστερου γείτονα κι η σύγκρισή του με τον πρώτο, καθώς όπως επισημαίνει ο Lowe, μια σωστή αντιστοιχία πρέπει να είναι σημαντικά πιο κοντά, από ότι η εγγύτερη λανθασμένη αντιστοιχία.

3.2.4 Αλγόριθμος SURF

Το 2006, οι Herbert Bay, Andreas Ess, Tinne Tuytelaars και Luc Van Gool παρουσίασαν τον αλγόριθμο SURF (Speeded-Up Robust Features), ως μια ταχύτερη μέθοδο ανίχνευσης και περιγραφής χαρακτηριστικών σημείων. Ο SURF στηρίζεται στη βασική φιλοσοφία του αλγόριθμου SIFT, όπως το ότι ανιχνεύει χαρακτηριστικά σημεία σε διαφορετικά επίπεδα κλιμάκωσης της εικόνας, για τα οποία δημιουργεί επίσης descriptors οι οποίοι περιστρέφονται, ώστε να μην επηρεάζονται από τις στροφές των εικόνων. Η μεθοδολογία του όμως, όπως θα φανεί παρακάτω, είναι αρκετά διαφορετική.

3.2.4.1 Ανίχνευση χαρακτηριστικών σημείων

Ο SURF για την ανίχνευση χαρακτηριστικών σημείων στην εικόνα, χρησιμοποιεί μια βασική προσέγγιση του Εσσιανού πίνακα. Για το λόγο αυτό χρησιμοποιούνται αθροιστικές εικόνες (Integral images)(Viola, Jones, 2001), οι οποίες επιτρέπουν τη δραστική μείωση του υπολογιστικού κόστους. Η αθροιστική εικόνα (Βερυκόκου, 2013) είναι μια εικόνα ίσων διαστάσεων με την αρχική και δημιουργείται αν στην τιμή της έντασης κάθε pixel της αρχικής εικόνας, προστεθούν οι τιμές της έντασης όλων των pixels, από την αρχή της εικόνας μέχρι τη θέση αυτού. Επομένως, η τιμή I_Σ στη θέση (x, y) προκύπτει από τη σχέση:

$$I_\Sigma(x, y) = \sum_{i=0}^{i \leq x} \sum_{j=0}^{j \leq y} I(x, y) \quad (3.22)$$

όπου $I(x,y)$, η τιμή της έντασης του pixel στην αρχική εικόνα, ενώ σε μια αθροιστική εικόνα, η ίδια τιμή προκύπτει ως εξής:

$$I_\Sigma(x, y) = I(x, y) + I_\Sigma(x - 1, y) + I_\Sigma(x, y - 1) - I_\Sigma(x - 1, y - 1) \quad (3.23)$$

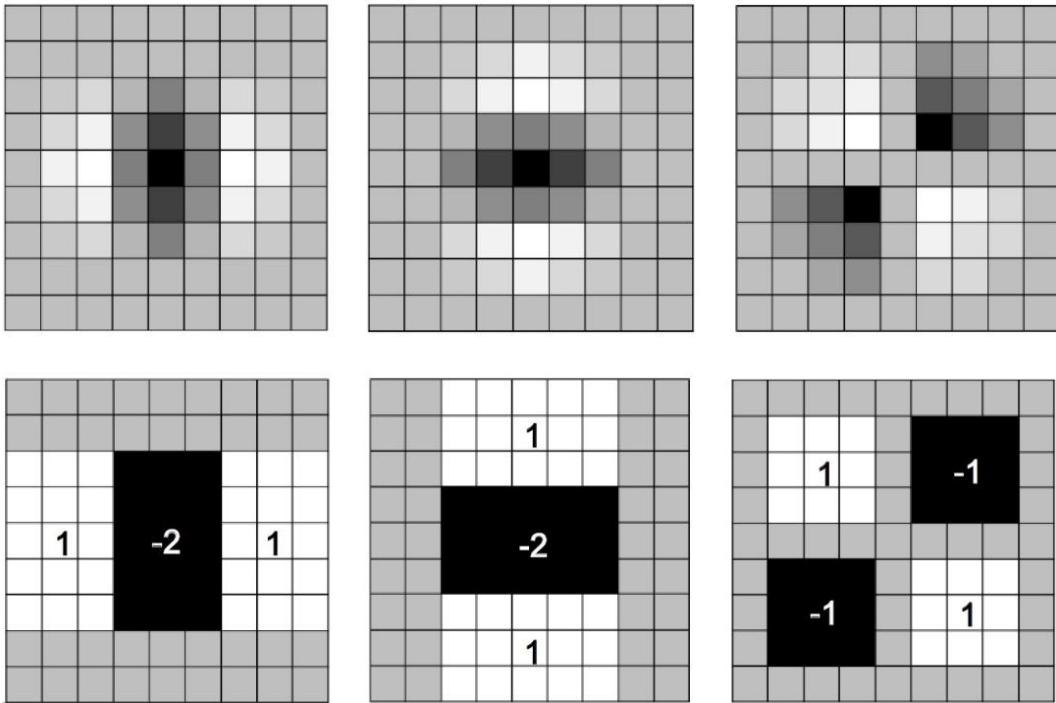
Ως γνωστόν, ο Εσσιανός πίνακας, είναι ένας τετραγωνικός πίνακας $N \times N$ διαστάσεων, που περιέχει τις, δευτέρας τάξης, μερικές παραγώγους μιας συνάρτησης N μεταβλητών, σε μια δεδομένη θέση. Στην περίπτωσή μας, λόγω

της δισδιάστατης φύσης των εικόνων, ισχύει $N=2$. Ο Εσσιανός πίνακας στη θέση (x, y) , για κλίμακα σ , είναι ο εξής:

$$H(x, y, \sigma) = \begin{bmatrix} L_{xx}(x, y, \sigma) & L_{xy}(x, y, \sigma) \\ L_{xy}(x, y, \sigma) & L_{yy}(x, y, \sigma) \end{bmatrix} \quad (3.24)$$

όπου L_{xx} , L_{xy} και L_{yy} είναι οι συνελίξεις της εικόνας με τα, δευτέρας παραγώγου Gauss φίλτρα $\frac{\partial}{\partial x^2} g(\sigma)$, $\frac{\partial}{\partial xy} g(\sigma)$ και $\frac{\partial}{\partial y^2} g(\sigma)$ αντίστοιχα (σχ. 3.4).

Επειδή όμως η παραπάνω διαδικασία είναι υπολογιστικά απαιτητική, ο αλγόριθμος πραγματοποιεί τις συνελίξεις, χρησιμοποιώντας τα λεγόμενα φίλτρα τύπου ‘κουτιού’ (box type filters) ή απλά, box φίλτρα (σχ. 3.4). Τα φίλτρα αυτά αποτελούν προσέγγιση των, δευτέρας παραγώγου Gauss φίλτρων και σε συνδυασμό με τη χρήση αθροιστικών εικόνων, μειώνουν κατά πολύ το χρόνο εκτέλεσης της διαδικασίας.



Σχήμα 3.4 Επάνω: δευτέρας παραγώγου Gauss φίλτρα. Κάτω: αντίστοιχα box φίλτρα προσέγγισης. Οι γκρι περιοχές έχουν μηδενική τιμή.

Εντέλει, προκύπτει ο προσεγγιστικός Εσσιανός πίνακας (Fast Hessian Matrix):

$$H_{approx}(x, y, \sigma) = \begin{bmatrix} D_{xx}(x, y, \sigma) & D_{xy}(x, y, \sigma) \\ D_{xy}(x, y, \sigma) & D_{yy}(x, y, \sigma) \end{bmatrix} \quad (3.25)$$

όπου D_{xx} , D_{xy} και D_{yy} είναι οι συνελίξεις της εικόνας με τα αντίστοιχα box φίλτρα στη θέση (x, y) για κλίμακα σ . Η ορίζουσά του είναι η εξής:

$$D_{xx}D_{yy} - \left(w(D_{xy})\right)^2 \quad (3.26)$$

όπου ο όρος w , χρησιμοποιείται ώστε να ελαχιστοποιηθεί το σφάλμα προσέγγισης του Εσσιανού πίνακα. Θεωρητικά, ο w μεταβάλλεται ανάλογα με την κλίμακα, στην πράξη ωστόσο, χρησιμοποιείται η προσεγγιστική τιμή $w=0.9$. Η ορίζουσα του προσεγγιστικού Εσσιανού πίνακα είναι το κριτήριο που χρησιμοποιεί ο αλγόριθμος, ώστε να ορίσει αν ένα σημείο στη θέση (x,y) και για κλίμακα σ , αποτελεί χαρακτηριστικό σημείο. Πιο αναλυτικά, θα πρέπει η ελάχιστη τιμή της ορίζουσας $\text{Det}(H_{\text{approx}})$ ενός σημείου να υπερβαίνει ένα προκαθορισμένο κατώφλι.

Όπως περιγράφηκε στην §3.2.3.1, ο αλγόριθμος SIFT, προκειμένου να καταστεί ανθεκτικός σε μεταβολές κλίμακας, δημιουργεί πυραμίδες εικόνας. Στην ίδια λογική στηρίζεται και ο SURF, μόνο που για τη δημιουργία κάθε νέας οκτάβας, αντί να υποδιπλασιάζει τις διαστάσεις της εικόνας, αυξάνει το μέγεθος του φίλτρου μειώνοντας με αυτόν τον τρόπο ακόμα περισσότερο το χρόνο εκτέλεσης του αλγόριθμου. Αναλυτικότερα, ακολουθείται η εξής διαδικασία:

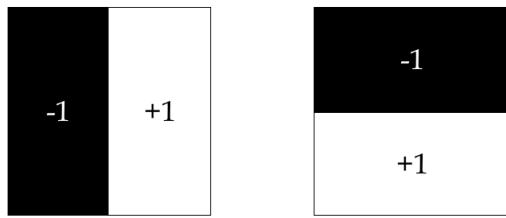
Η πρώτη οκτάβα δημιουργείται με τη συνέλιξη της εικόνας με το 9×9 box φίλτρο, που προσεγγίζει το Gauss φίλτρο τυπικής απόκλισης $\sigma=0.2$. Στη συνέχεια ακολουθούν άλλες τρεις συνελίξεις, αυξάνοντας κάθε φορά τη διάσταση του φίλτρου κατά 6 pixel, οπότε έχουμε φίλτρα διαστάσεων 15×15 , 21×21 και 27×27 . Η δεύτερη οκτάβα σχηματίζεται αυξάνοντας το αρχικό φίλτρο κατά 6 pixel και στη συνέχεια αυξάνεται άλλες τρεις φορές κατά 12 pixel, οπότε έχουμε φίλτρα διαστάσεων 15×15 , 27×27 , 39×39 και 51×51 . Τα αντίστοιχα φίλτρα για την τρίτη οκτάβα έχουν διαστάσεις 27×27 , 51×51 , 75×75 και 99×99 . Σε περιπτώσεις εικόνων μεγάλων διαστάσεων, δημιουργείται ακόμα μια οκτάβα συνελίσσοντας την εικόνα με φίλτρα 51×51 , 99×99 , 147×147 και 195×195 .

Στη συνέχεια, στις περιοχές που ανιχνεύθηκαν σαν χαρακτηριστικά σημεία διαδοχικά pixel, εκτελείται η διαδικασία της ‘αφαίρεσης των μη-μεγίστων, επιλέγοντας τελικά, το pixel με τη μεγαλύτερη τιμή της $\text{Det}(H_{\text{approx}})$.

Τέλος, για τον προσδιορισμό της θέσης των σημείων με υποψηφιδική ακρίβεια, πραγματοποιείται παρεμβολή, παρόμοια με τον SIFT, εκφράζοντας τη $\text{Det}(H_{\text{approx}}(x))$ με την βοήθεια β' βαθμού πολυώνυμο Taylor και θέτοντας την παράγωγό του ίση με μηδέν.

3.2.4.2 Ανάθεση προσανατολισμού

Προκειμένου να οριστεί ο κυρίαρχος προσανατολισμός κάθε χαρακτηριστικού σημείου, ο αλγόριθμος κάνει χρήση Haar Wavelet φίλτρων (σχ. 3.5), προκειμένου να υπολογίσει τις κλίσεις της εικόνας, σε μια περιοχή γύρω από το σημείο.

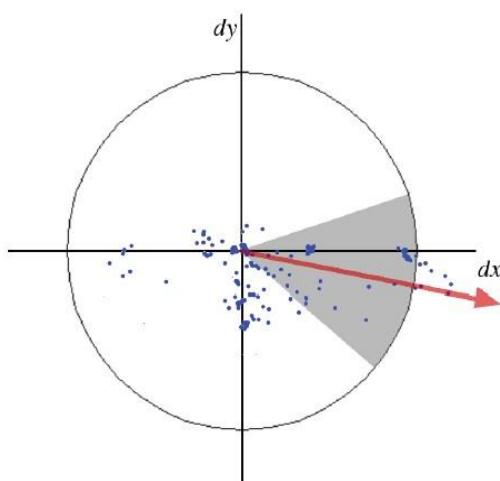


Σχήμα 3.5: Haar Wavelet φίλτρα, τα οποία χρησιμοποιούνται για τον υπολογισμό της κατά x (αριστερό) και της κατά y (δεξιά) κλίσης της εικόνας στο εκάστοτε σημείο (pixel) που εφαρμόζονται.

Συγκεκριμένα, ορίζεται μια περιοχή ακτίνας s γύρω από το σημείο, όπου s είναι η κλίμακα στην οποία ανιχνεύθηκε το σημείο. Στην περιοχή αυτή, επιλέγεται ένα δείγμα από pixel, στα οποία πραγματοποιούνται συνελίξεις με τα Haar Wavelets. Το βήμα δειγματοληψίας επιλέγεται να ισούται με s . Επίσης, η διάσταση της πλευράς των Haar Wavelets ορίζεται σε $4s$.

Τα αποτελέσματα των συνελίξεων σταθμίζονται με βάση μια συνάρτηση κανονικής κατανομής, με κέντρο το χαρακτηριστικό σημείο και τυπική απόκλιση $\sigma=2s$. Εν συνεχείᾳ, οι τιμές αυτές αναπαριστώνται σαν σημεία σε ένα δισδιάστατο καρτεσιανό σύστημα, με τετμημένη την κατά x και τεταγμένη την κατά y κλίση των pixel, όπως υπολογίστηκαν από τις συνελίξεις.

Τέλος, η ανάθεση του κυρίαρχου προσανατολισμού, γίνεται με τη χρήση ενός κυκλικού τομέα εύρους 60° (σχ. 3.6). Ο τομέας περιστρέφεται γύρω από την αρχή των αξόνων, αναζητώντας τη θέση στην οποία θα περιέχει το μέγιστο αριθμό των σημείων. Ο κυρίαρχος προσανατολισμός προκύπτει από τον υπολογισμό της γωνίας που σχηματίζεται στο κέντρο του κυκλικού τομέα.



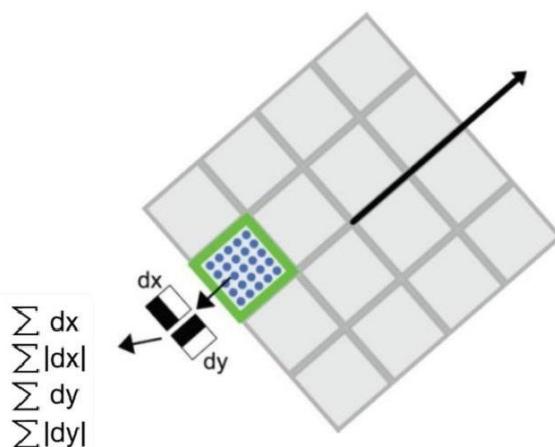
Σχήμα 3.6: Τελικό στάδιο της διαδικασίας ανάθεσης προσανατολισμού. Ο κυκλικός τομέας περιστρέφεται, αναζητώντας τη θέση στην οποία θα περιέχει το μέγιστο αριθμό των σημείων.

3.2.4.3 Δημιουργία των Descriptors

Για τη δημιουργία των Descriptors, ο αλγόριθμος χρησιμοποιεί, επίσης, τα Haar Wavelet φίλτρα.

Καταρχήν, ορίζεται μια τετράγωνη περιοχή διάστασης 20s, με κέντρο το χαρακτηριστικό σημείο (σχ. 3.7). Η περιοχή αυτή χωρίζεται σε $4 \times 4 = 16$ επί μέρους τετράγωνα. Σε καθεμία από τις 16 υποπεριοχές επιλέγονται $5 \times 5 = 25$ pixel, ομοιόμορφα κατανεμημένα, στα οποία πραγματοποιείται συνέλιξη με Haar Wavelet φίλτρα, διάστασης 2s, για τον υπολογισμό της οριζόντιας και κάθετης κλίσης, ακριβώς όπως στο προηγούμενο βήμα. Να σημειωθεί εδώ ότι, οι όροι ‘οριζόντια’ και ‘κάθετη’ κλίση, αναφέρονται με βάση τον κυρίαρχο προσανατολισμό. Επιπλέον, για την αύξηση της αξιοπιστίας του αλγόριθμου, ως προς τις γεωμετρικές παραμορφώσεις, τα αποτελέσματα των συνελίξεων σταθμίζονται με βάση μια συνάρτηση κανονικής κατανομής, με κέντρο το χαρακτηριστικό σημείο και τυπική απόκλιση $\sigma = 3.3s$.

Τέλος, για κάθε υποπεριοχή υπολογίζονται οι τιμές $\sum dx$, $\sum dy$, $\sum |dx|$ και $\sum |dy|$, όπου: $\sum dx$, $\sum dy$ το άθροισμα των κατά x και y κλίσεων αντίστοιχα, για τα 25 pixels και $\sum |dx|$, $\sum |dy|$ το άθροισμα των απόλυτων τιμών τους. Άρα, για την περιγραφή όλης της περιοχής, καταλήγουμε σε ένα διάνυσμα $4 \times 16 = 64$ στοιχείων. Το διάνυσμα αυτό κανονικοποιείται, όπως και στον SIFT, ώστε η ευκλείδεια νόρμα του να έχει μοναδιαίο μέτρο.



Σχήμα 3.7: Ο Descriptor του SURF (Βερυκόκου, 2013)

3.2.4.4 Συνταύτιση χαρακτηριστικών

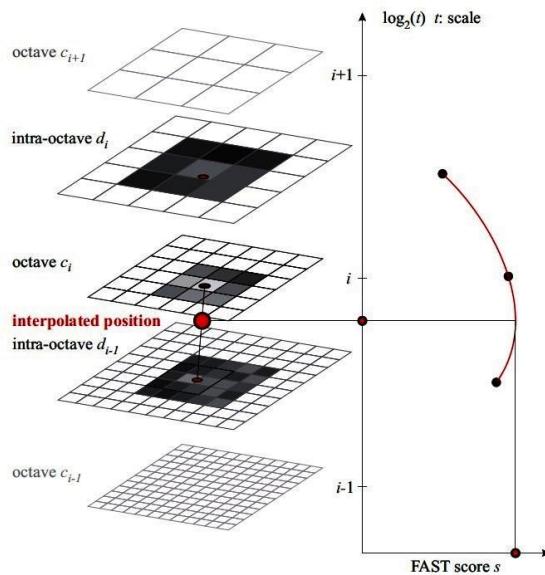
Η συνταύτιση χαρακτηριστικών μπορεί να γίνει κι εδώ με την τεχνική του ‘πλησιέστερου γείτονα’, με κριτήριο την ελάχιστη ευκλείδεια απόσταση μεταξύ των Descriptors και τον ορισμό μια ανώτατης ανεκτής τιμής της (ανώφλι).

Προκειμένου να επιταχυνθεί η διαδικασία της συνταύτισης, ο αλγόριθμος εξετάζει για κάθε χαρακτηριστικό σημείο το πρόσημο του ίχνους (Λαπλασιανή) του Εσσιανού πίνακα. Αν το πρόσημο είναι θετικό τότε περιγράφεται φωτεινό χαρακτηριστικό σε σκοτεινό φόντο, ενώ αν είναι αρνητικό περιγράφεται σκοτεινό χαρακτηριστικό σε φωτεινό φόντο. Έτσι, η συνταύτιση επιχειρείται μόνο σε σημεία με ίδιο πρόσημο της Λαπλασιανής.

3.2.5 Αλγόριθμος BRISK

Το 2011, οι Stefan Leutenegger, Margarita Chli και Roland Siegwart, παρουσίασαν το αλγόριθμο BRISK (Binary Robust Invariant Scalable Keypoints), ο οποίος είναι σημαντικά πιο γρήγορος από τους SIFT και SURF, χάρη στις απλούστερες μεθόδους που χρησιμοποιεί για την ανίχνευση και περιγραφή των χαρακτηριστικών σημείων.

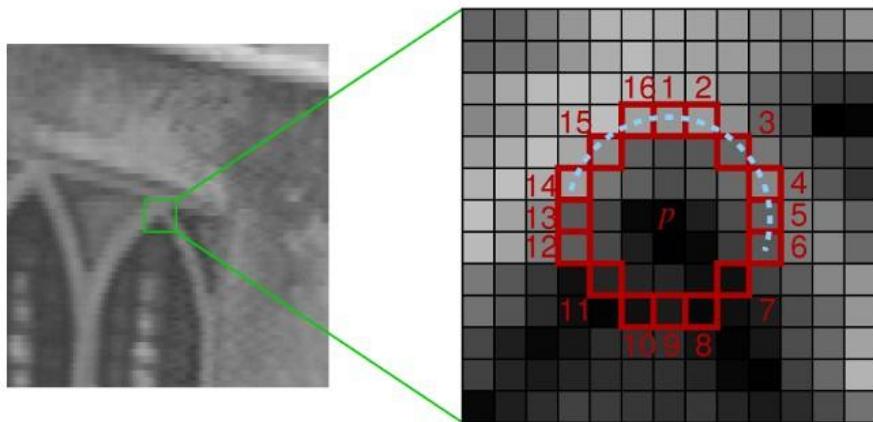
Καταρχήν, όπως και οι δύο προαναφερθέντες, έτσι και ο αλγόριθμος BRISK, ανιχνεύει χαρακτηριστικά σημεία σε πυραμίδα εικόνας που αναπαριστά το χώρο κλίμακας (Scale-Space Keypoint Detection). Η πυραμίδα (σχήμα 3.8) αποτελείται από n οκτάβες c_i και n εσωτερικές οκτάβες d_i (intra-octaves), όπου $i = \{0, 1, \dots, n - 1\}$ και συνήθως $n=4$. Οι οκτάβες δημιουργούνται με υποδιπλασιασμό της προηγούμενης, ξεκινώντας από την αρχική εικόνα, που συμβολίζεται με c_0 . Κάθε εσωτερική οκτάβα τοποθετείται ανάμεσα από τα επίπεδα c_i και c_{i+1} . Η αρχική εσωτερική οκτάβα προκύπτει με τη μείωση του μεγέθους των διαστάσεων της αρχική εικόνας κατά 1.5 φορά. Οι υπόλοιπες εσωτερικές οκτάβες προκύπτουν με προοδευτικό υποδιπλασιασμό της αρχικής.



Σχήμα 3.8: Αναπαράσταση του χώρου-κλίμακας

Για την ανίχνευση χαρακτηριστικών σημείων στα επίπεδα κλίμακας, ο αλγόριθμος βασίζεται στο κριτήριο FAST (Features from Accelerated Segment Test). Η διαδικασία που ακολουθείται είναι η εξής:

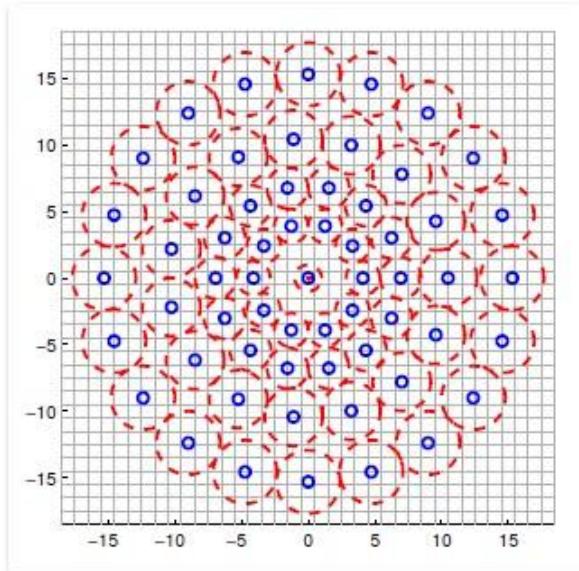
Γύρω από κάθε υποψήφιο χαρακτηριστικό σημείο p , εξετάζεται ένα κυκλικό τμήμα μήκους 16 pixel, όπως φαίνεται στο σχήμα 3.8. Αν I_p η τιμή της έντασης στο σημείο p και t ένα ορισμένο κατώφλι, τότε το σημείο p είναι χαρακτηριστικό, αν και μόνο αν, υπάρχουν n συνεχόμενα pixel του κυκλικού τμήματος, τα οποία να είναι φωτεινότερα από I_p+t ή σκοτεινότερα από I_p-t . Η τιμή του n συνήθως επιλέγεται να είναι ίση με 9 (μάσκα 9-16). Για τη μείωση του υπολογιστικού κόστους, ο αλγόριθμος εξετάζει αρχικά, μόνο τα pixel 1, 5, 9, 13. Αν τουλάχιστον τρία από αυτά δεν είναι φωτεινότερα από I_p+t ή σκοτεινότερα από I_p-t , τότε το σημείο δεν αποτελεί υποψήφιο χαρακτηριστικό και η διερεύνηση σταματάει εκεί.



Σχήμα 3.9: Κριτήριο FAST. Στη συγκεκριμένη περίπτωση εφαρμόζεται η μάσκα 9-16.

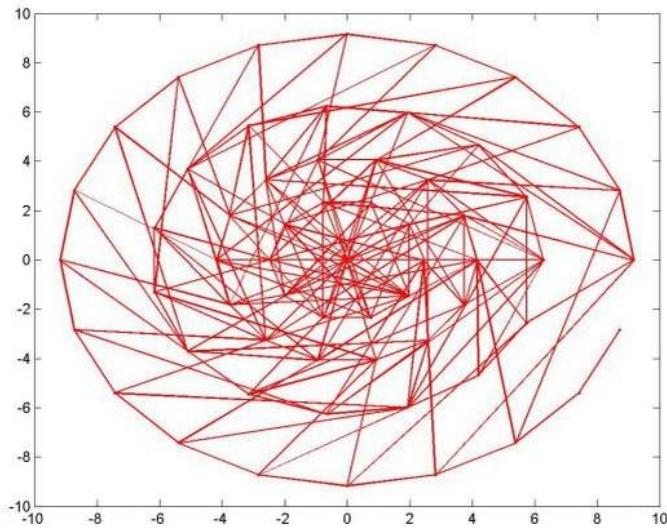
Η βασικότερη ίσως διαφορά του BRISK από τους αλγόριθμους SIFT και SURF, έγκειται στην περιγραφή των χαρακτηριστικών σημείων, καθώς ο πρώτος δημιουργεί Descriptors που αποτελούνται από μια δυαδική συμβολοσειρά, μήκους 512 bit.

Πιο αναλυτικά, ο αλγόριθμος χρησιμοποιεί ένα συγκεκριμένο πρότυπο δειγματοληψίας (Sampling Pattern), που αποτελείται από N pixel, κατανεμημένα σε διάταξη ομόκεντρων κύκλων, όπως φαίνεται στο σχήμα 3.10:



Σχήμα 3.10: Πρότυπο δειγματοληψίας του αλγόριθμου BRISK για $N=60$ σημεία. Οι μπλε κύκλοι παριστάνουν τα pixel, ενώ οι κόκκινοι κύκλοι γύρω τους, σχεδιάζονται με ακτίνα ανάλογη της τυπικής απόκλισης των Gauss φίλτρων που χρησιμοποιήθηκαν για την εξομάλυνσή των τιμών της έντασής τους.

Στη συνέχεια, στηριζόμενος πάλι σε συγκεκριμένο πρότυπο δειγματοληψίας (σχ. 3.11), δημιουργεί ζεύγη από pixel, τα οποία χρησιμοποιεί για τους υπολογισμούς του, προκειμένου να δομήσει τον Descriptor.



Σχήμα 3.11: Πρότυπο δειγματοληψίας με το οποίο ο αλγόριθμος δημιουργεί ζεύγη από pixel

Ο αλγόριθμος διαχωρίζει τα ζεύγη σε κοντινά και μακρινά, βάσει συγκεκριμένων τιμών κατωφλίωσης για κάθε περίπτωση. Τα μακρινά ζεύγη τα χρησιμοποιεί για να υπολογίσει τον κυρίαρχο προσανατολισμό του σημείου, ενώ τα κοντινά τα χρησιμοποιεί για να δομήσει τον Descriptor. Για $N=60$, δημιουργούνται 512 κοντινά και 870 μακρινά ζεύγη.

Για καθένα από τα μακρινά ζεύγη (p_i, p_j) υπολογίζεται η τοπική κλίση από τη σχέση:

$$g(p_i, p_j) = (p_i - p_j) \cdot \frac{I(p_j, \sigma_j) - I(p_i, \sigma_i)}{\|p_j - p_i\|^2} \quad (3.27)$$

όπου: $I(p_i, \sigma_i)$ και $I(p_j, \sigma_j)$ οι τιμές της έντασης των σημείων μετά από Gaussian εξομάλυνση τυπικής απόκλισης σ_i και σ_j αντίστοιχα.

Το διάνυσμα προσανατολισμού του σημείου, δίνεται από τη σχέση:

$$\mathbf{g} = \begin{pmatrix} g_x \\ g_y \end{pmatrix} = \frac{1}{L} \cdot \sum g(p_i, p_j) \quad (3.28)$$

όπου: L το πλήθος των ζευγών.

Για τη δημιουργία του Descriptor, καταρχήν, το πρότυπο δειγματοληψίας στρέφεται κατά γωνία $a = \text{arctan2}(g_x, g_y)$ γύρω από το σημείο. Στη συνέχεια, ο αλγόριθμος πραγματοποιεί συγκρίσεις φωτεινότητας μεταξύ των 512 κοντινών ζευγών και δημιουργεί μια δυαδική συμβολοσειρά μήκους 512 bit, όπου κάθε bit, παίρνει τις τιμές ο ή 1, σύμφωνα με το εξής κριτήριο:

$$b = \begin{cases} 1, & I(p_j^\alpha, \sigma_j) > I(p_i^\alpha, \sigma_i) \\ 0, & \text{σε κάθε άλλη περίπτωση} \end{cases} \quad (3.29)$$

Τέλος, οι Binary Descriptors συνταυτίζονται με βάση τη μεταξύ τους απόσταση Hamming. Τα αντίστοιχα bit των δύο Descriptors εξετάζονται ένα προς ένα, ως προς την ομοιότητά τους. Αν ο αριθμός των bits που διαφέρουν, δεν ξεπερνά ένα ορισμένο ανώφλι, που συνήθως ορίζεται ως ποσοστό, τότε τα υπό εξέταση χαρακτηριστικά σημεία θεωρούνται ομόλογα.

ΚΕΦΑΛΑΙΟ 4

Εφαρμογές και αξιολόγηση

Στο πλαίσιο της βαθύτερης κατανόησης όσων αναφέρθηκαν στα προηγούμενα κεφάλαια, προγραμματίστηκαν αλγόριθμοι που, σε συνδυασμό με ήδη υπάρχοντες, υλοποίησαν τις περισσότερες από τις διαδικασίες που περιγράφηκαν. Κύριος σκοπός βέβαια, είναι να αξιολογηθούν και να συγκριθούν οι αλγόριθμοι SIFT και SURF, ως προς την ποιότητα των αποτελεσμάτων τους.

Εύκολα γίνεται αντιληπτό, με βάση και την αναλυτική τους παρουσίαση στο προηγούμενο κεφάλαιο, ότι το κυριότερο κίνητρο στην εξέλιξη και διατύπωση νέων αλγορίθμων είναι να μειωθεί, όσο είναι δυνατόν, ο χρόνος εκτέλεσής τους (άρα μείωση του υπολογιστικού φόρτου), πρωτίστως λόγω της ανάγκης για εφαρμογές πραγματικού χρόνου. Ο SURF είναι σημαντικά πιο γρήγορος από τον SIFT, χάρη στη χρήση αθροιστικών εικόνων και της συνέλιξής τους με box φίλτρα, αλλά και της μικρότερης διάστασης των descriptors που εξάγει. Από την άλλη, ο BRISK είναι ακόμα γρηγορότερος, χάρη στο κριτήριο FAST για την ανίχνευση και τη χρήση δυαδικής συμβολοσειράς για την περιγραφή των σημείων. Ο πίνακας 4.1, είναι ενδεικτικός των διαφορών αυτών και προέκυψε από ένα τυχαίο ζεύγος εικόνων των δεδομένων.

	SIFT	SURF	BRISK
Χρόνος ανίχνευσης για κάθε σημείο (ms)	1326,9	55,8	31,9
Χρόνος συνταύτισης για κάθε ζεύγος (ms)	3178,2	486,2	64,2

Πίνακας 4.1: Ενδεικτικοί χρόνοι εκτέλεσης των αλγορίθμων.

Το παραπάνω, όπως είναι αναμενόμενο, δεν αφήνει ανεπηρέαστες την ποιότητα και την αξιοπιστία των αλγορίθμων. Ο αλγόριθμος BRISK, λόγω της ταχύτητας του (κυρίως όσον αφορά το χρόνο συνταύτισης), καθίσταται ιδιαίτερα ελκυστικός όταν πρόκειται για εφαρμογές πραγματικού χρόνου. Όμως τα αποτελέσματά του δεν είναι πάντα το ίδιο αξιόπιστα σε σχέση με τους άλλους δύο αλγόριθμους και σε ορισμένες περιπτώσεις παρουσιάζει σημαντική αστάθεια (τουλάχιστον όσον αφορά την έκδοση που χρησιμοποιείται εδώ). Επιπλέον, υπάρχουν περιπτώσεις που ο αριθμός των τελικών ομολογιών που εξάγει είναι σχετικά πολύ μικρός, κάτι που συνήθως μεταφράζεται σε ‘κακή’ κατανομή των σημείων στις εικόνες. Ως εκ τούτου

κρίνεται σκόπιμο να παραληφθεί από τη σύγκριση και να παρουσιαστούν μόνο τα αποτελέσματα των αλγορίθμων SIFT και SURF.

Για τον προγραμματισμό και την εκτέλεση των διαδικασιών που χρειάστηκαν, χρησιμοποιήθηκε η πλατφόρμα του Matlab R2014a (έκδοση 8.3).

Το προγραμματιστικό υλικό που χρησιμοποιήθηκε (με αναφορά στις πηγές) έχει ως εξής:

- Για τον αλγόριθμο SURF χρησιμοποιήθηκαν συναρτήσεις από τη βιβλιοθήκη του Matlab (<http://www.mathworks.com/help/vision/ref/detectsurffeatures.html>)
- Για τον αλγόριθμο SIFT χρησιμοποιήθηκε κώδικας που συνέταξε ο Ηλίας Καλησπεράκης, συνδυαστικά με τον κώδικα που διανέμει ελεύθερα ο David Lowe στην επίσημη ιστοσελίδα του (<http://www.cs.ubc.ca/~lowe/keypoints/>).
- Για τον Ransac χρησιμοποιήθηκε επίσης, κώδικας του Ηλία Καλησπεράκη και κώδικας από την ιστοσελίδα του Peter Kovesi (<http://www.csse.uwa.edu.au/~pk/research/matlabfns/>).
- Τέλος, για την αρχική συνταύτιση των σημείων, χρησιμοποιήθηκε κώδικας από τη σελίδα www.vlfeat.org (http://www.vlfeat.org/matlab/vl_ubcmatch.html).

Τα παραπάνω συνδυάστηκαν με κώδικα του γράφοντα, ώστε να δημιουργηθούν ξεχωριστά scripts για καθέναν από τους δύο αλγόριθμους, που αναλυτικά πραγματοποιούν τα εξής:

1. Φόρτωση του ζεύγους εικόνων και μείωση της ανάλυσής τους, ώστε να οριοθετηθεί ο αριθμός των εξαγόμενων σημείων.
2. Ανίχνευση χαρακτηριστικών σημείων.
3. Υπολογισμός των Descriptors.
4. Αρχική συνταύτιση των Descriptors, με κριτήριο την ελάχιστη τιμή του τετραγώνου της ευκλείδειας απόστασης.
5. Εφαρμογή του αλγόριθμου Ransac, χρησιμοποιώντας τη μέθοδο των 7 σημείων για τον υπολογισμό του επιπολικού πίνακα.
6. Εκ νέου υπολογισμός του επιπολικού πίνακα, με τη μέθοδο του κανονικοποιημένου αλγόριθμου των 8 σημείων, χρησιμοποιώντας ως δεδομένα εισόδου τα inliers που υπολογίστηκαν στο προηγούμενο βήμα.

Παρακάτω θα αναφερθούν αναλυτικά οι παράμετροι που χρησιμοποιήθηκαν σε κάθε βήμα, κατά περίπτωση.

Επιπλέον, δημιουργήθηκε script ώστε να μελετηθεί η κατανομή των σημείων που εξάγουν οι αλγόριθμοι, η λειτουργία του οποίου θα παρουσιαστεί αναλυτικά στην παράγραφο 4.1.7.

Πρωτογενή δεδομένα

Για την μελέτη των αλγορίθμων χρησιμοποιήθηκαν δύο σετ εικόνων (εικόνες 4.1 και 4.2), οι οποίες ελήφθησαν από τον γράφοντα με ψηφιακή φωτογραφική μηχανή. Ο εσωτερικός προσανατολισμός της μηχανής σκοπίμως δεν θα αναφερθεί εδώ, καθώς οι εφαρμογές που παρουσιάζονται στην παρούσα εργασία καταλήγουν στον υπολογισμό του επιπολικού πίνακα, τονίζεται μόνο ότι είναι κοινός για κάθε σετ εικόνων. Το πρώτο σετ απεικονίζει το κτήριο που στεγάζεται το τμήμα οινολογίας, στη δυτική πλευρά του Τ.Ε.Ι. Αθήνας, ενώ το δεύτερο απεικονίζει ένα τραπέζι, πάνω στο οποίο υπάρχουν διάφορα αντικείμενα. Οι εικόνες των δύο σετ διαφέρουν ως προς την απόσταση λήψης τους ($H \sim 25m$ στο πρώτο και $H \sim 1m$ στο δεύτερο) και ως προς την υφή τους, καθώς εκείνες του δευτέρου παρουσιάζουν εντονότερη υφή. Οι αρχικές διαστάσεις του πρώτου σετ είναι 1632×1224 , ενώ του δευτέρου 3072×2304 .



Εικόνα 4.1: Dataset_1



Εικόνα 4.2: Dataset_2

Συνθήκες στις οποίες εξετάζονται οι αλγόριθμοι

Οι συνθήκες στις οποίες εξετάζονται οι αλγόριθμοι, αφορούν στις διάφορες μεταβολές στη γεωμετρία και τη ραδιομετρία των εικόνων και είναι οι εξής:

Γεωμετρικές μεταβολές

1. Μικρό (σχετικά) μήκος Βάσης του στερεοζεύγους ($OO' < 0,50$ m για το Dataset_1 και $OO' < 0,20$ m για το Dataset_2).
2. Διαφορά κλίμακας μεταξύ των εικόνων.
3. Στροφή μεταξύ των εικόνων.
4. Μεγάλο μήκος Βάσης του στερεοζεύγους ($OO' > 10$ m για το Dataset_1 και $OO' > 2$ m για το Dataset_2).

Ραδιομετρικές μεταβολές

5. Διαφορά φωτεινότητας μεταξύ των εικόνων.
6. 'Υπαρξη θορύβου στις εικόνες, Συγκεκριμένα, ο θόρυβος ακολουθεί γκαουσιανή (κανονική) κατανομή με $\sigma=0.01$.

Εδώ αξίζει να σημειωθεί ότι, στη Φωτογραμμετρία, το μήκος της Βάσης αυτό καθαυτό, δεν έχει ιδιαίτερη σημασία. Αυτό που ενδιαφέρει είναι το μήκος της Βάσης ως προς την απόσταση λήψης ($\frac{B}{H}$). Για το λόγο αυτό, στις περιπτώσεις 1 και 4, οι όροι ‘μικρό’ και ‘μεγάλο’ μήκος Βάσης, αναφέρονται ως προς την απόσταση λήψης. Απλώς, επειδή στις δύο αυτές περιπτώσεις η απόσταση λήψης για τα αντίστοιχα στερεοζεύγη είναι σχεδόν η ίδια, καθώς και για λόγους συντομίας, η φράση ‘ως προς την απόσταση λήψης’ παραλείπεται.

Κριτήρια αξιολόγησης.

Προκειμένου να αξιολογηθεί η απόδοση των αλγορίθμων σε κάθε περίπτωση, δημιουργήθηκαν μέτρα που αποτελούνται από τους παρακάτω λόγους εκφρασμένους σε ποσοστά επί τοις εκατό:

$$1. \quad a = \frac{\text{Αριθμός αντιστοιχιών από την αρχική συνταύτιση των σημείων}}{\text{Μέσος όρος των σημείων που ανιχνεύθηκαν στις δύο εικόνες}}$$

$$2. \quad b = \frac{\text{Inliers}}{\text{Αριθμός αντιστοιχιών από την αρχική συνταύτιση των σημείων}}$$

Το πρώτο μέτρο είναι ενδεικτικό της ικανότητας του αλγόριθμου να εξάγει ομόλογα σημεία στις εικόνες και λόγω του παρονομαστή του, η σημασία του είναι μεγαλύτερη όσο μεγαλώνει η επικάλυψη των εικόνων. Στις εικόνες που χρησιμοποιούνται εδώ, η επικάλυψη είναι πάνω από 90-95% (με εξαίρεση βέβαια την περίπτωση αλλαγής κλίμακας).

Το δεύτερο μέτρο είναι ίσως σημαντικότερο, όσον αφορά τη συνολική απόδοση του αλγόριθμου. Το μέτρο αυτό συσχετίζει τις ομολογίες της αρχικής συνταύτισης, που προσδιορίζονται βάσει υπολογισμών μεταξύ τιμών των pixel, με τα inliers που προκύπτουν με καθαρά γεωμετρικό τρόπο. Εντέλει, όσο μεγαλύτερο είναι το ποσοστό του, τόσο πιο καλή είναι η περιγραφή (descriptors) των σημείων ενδιαφέροντος.

Παράδειγμα



Εικόνα 4.3: Αρχικά σημεία που εντοπίζονται σε κάθε εικόνα



Εικόνα 4.4: Αντιστοιχίες που προκύπτουν μετά την αρχική συνταύτιση



Εικόνα 4.5: Inliers που προκύπτουν μετά την εφαρμογή του Ransac ($\sigma_0=0.7$)

Στο παραπάνω παράδειγμα έχουμε:

$$\text{Μ.Ο. σημείων στις εικόνες} = (8315+8664)/2=8489.5$$

$$\text{Αριθμός αντιστοιχιών μετά την αρχική συνταύτιση} = 3071$$

$$\text{Inliers που προκύπτουν μετά από Ransac} = 2156$$

Οπότε, οι τιμές των δύο μέτρων είναι:

$$a = \frac{3071}{8489.5} \cdot 100\% = 36.17\%$$

και

$$b = \frac{2156}{3071} \cdot 100\% = 70.21\%$$

Άρα, από τα αρχικά σημεία των εικόνων, συνταυτίστηκε το 36.17 % και από τις αντιστοιχίες αυτές, μόνο το 70.21 % αποτελούν inliers.

Πέραν των δύο αυτών μέτρων, οι αλγόριθμοι αξιολογούνται επίσης, με βάση την κατανομή των σημείων που εξάγουν (§ 4.1.7).

4.1 Αλγόριθμοι SURF και SIFT

Στην ενότητα αυτή παρουσιάζονται και συγκρίνονται τα αποτελέσματα των αλγορίθμων SURF και SIFT.

Οι αλγόριθμοι υλοποιούνται με τις ακόλουθες παραμέτρους:

Διαστάσεις των εικόνων

Η ανάλυση των εικόνων μεταβάλλεται, ώστε να γίνει 1306 x 980 για τις εικόνες του Dataset_1 και 1229 x 922 για το Dataset_2.

SURF

Κατώφλι για την ορίζουσα του Fast-Hessian πίνακα: 200

Οκτάβες που δημιουργούνται: 4

Επίπεδα σε κάθε οκτάβα: 5

SIFT

Ο SIFT υλοποιήθηκε με τις προεπιλεγμένες ρυθμίσεις, οι οποίες ‘αποκρύπτονται’ από τον συντάκτη του, στο M-file αρχείο που διανέμει.

RANSAC

Ο αλγόριθμος RANSAC εφαρμόζεται σε κάθε Dataset δύο φορές, τη μια ορίζοντας σαν μέγιστη επιτρεπτή απόσταση σημείου από την αντίστοιχη επιπολική του γραμμή στην άλλη εικόνα, ίση με 0.7 pixel και την άλλη με μια ‘αυστηρή’ τιμή, ίση με 0.3 pixel. Ο αριθμός των επαναλήψεων είναι κοινός και στις δύο περιπτώσεις και είναι ίσος με 3000.

Να σημειωθεί εδώ, ότι σε όλες τις περιπτώσεις, ο αλγόριθμος εκτελείται αρχικά με μια πολύ μεγάλη τιμή κατωφλίου (της τάξης των 30 pixel) κι έπειτα, εκτελείται ξανά για τα inliers που προέκυψαν, με τιμές 0.7 και 0.3, όπως προαναφέρθηκε. Με τον τρόπο αυτό απομακρύνονται τα ιδιαιτέρως

χονδροειδή σφάλματα, που μπορούν να επηρεάσουν δυσμενώς την ακρίβεια υπολογισμού του επιπολικού πίνακα.

Αρχική συνταύτιση

Η αρχική συνταύτιση και για τους δύο αλγόριθμους γίνεται με βάση το τετράγωνο της ευκλείδειας απόστασης μεταξύ των descriptors. Η τιμή του κατωφλίου είναι ίση με 1,5 και έχει την εξής σημασία: προκειμένου να συνταυτιστούν δύο περιγραφείς D_1 και D_2 , θα πρέπει η απόσταση $d(D_1, D_2)$ πολλαπλασιαζόμενη επί την τιμή του κατωφλίου, να είναι μικρότερη από την απόσταση $d(D_1, D'_2)$, όπου D'_2 είναι ο δεύτερος κοντινότερος Descriptor στον D_1 .

Διάρθρωση των υποενοτήτων που ακολουθούν

Στην αρχή κάθε υποενότητας παρουσιάζονται τα στερεοζεύγη με τα αποτελέσματα των SURF και SIFT για κάθε Dataset ξεχωριστά. Η διάταξη που παρουσιάζονται, με βάση και την αρίθμηση τους, είναι η εξής:

1. Απεικόνιση των αρχικών σημείων που ανιχνεύθηκαν σε κάθε εικόνα.
2. Απεικόνιση των σημείων με βάση την αρχική συνταύτιση.
3. Απεικόνιση των σημείων που απομένουν μετά την εφαρμογή του RANSAC με $\sigma=0,7$.
4. Απεικόνιση των σημείων που απομένουν μετά την εφαρμογή του RANSAC με $\sigma=0,3$.

Στη συνέχεια παρουσιάζεται πίνακας που δείχνει αναλυτικά τον αριθμό των σημείων σε κάθε ένα από τα παραπάνω βήματα. Ακολουθεί η παρουσίαση δύο γραφημάτων (ένα για κάθε Dataset) που παρουσιάζουν σε αντιπαραβολή τα αποτελέσματα των δύο αλγορίθμων, με βάση τα κριτήρια αξιολόγησης που αναφέρθηκαν στην εισαγωγή του κεφαλαίου. Η κάθε υποενότητα κλείνει με τον σχολιασμό των αποτελεσμάτων.

Τέλος, η κατανομή των σημείων σε κάθε περίπτωση μελετάται ξεχωριστά στην §4.1.7.

4.1.1 Μικρό μήκος Βάσης του στερεοζεύγους

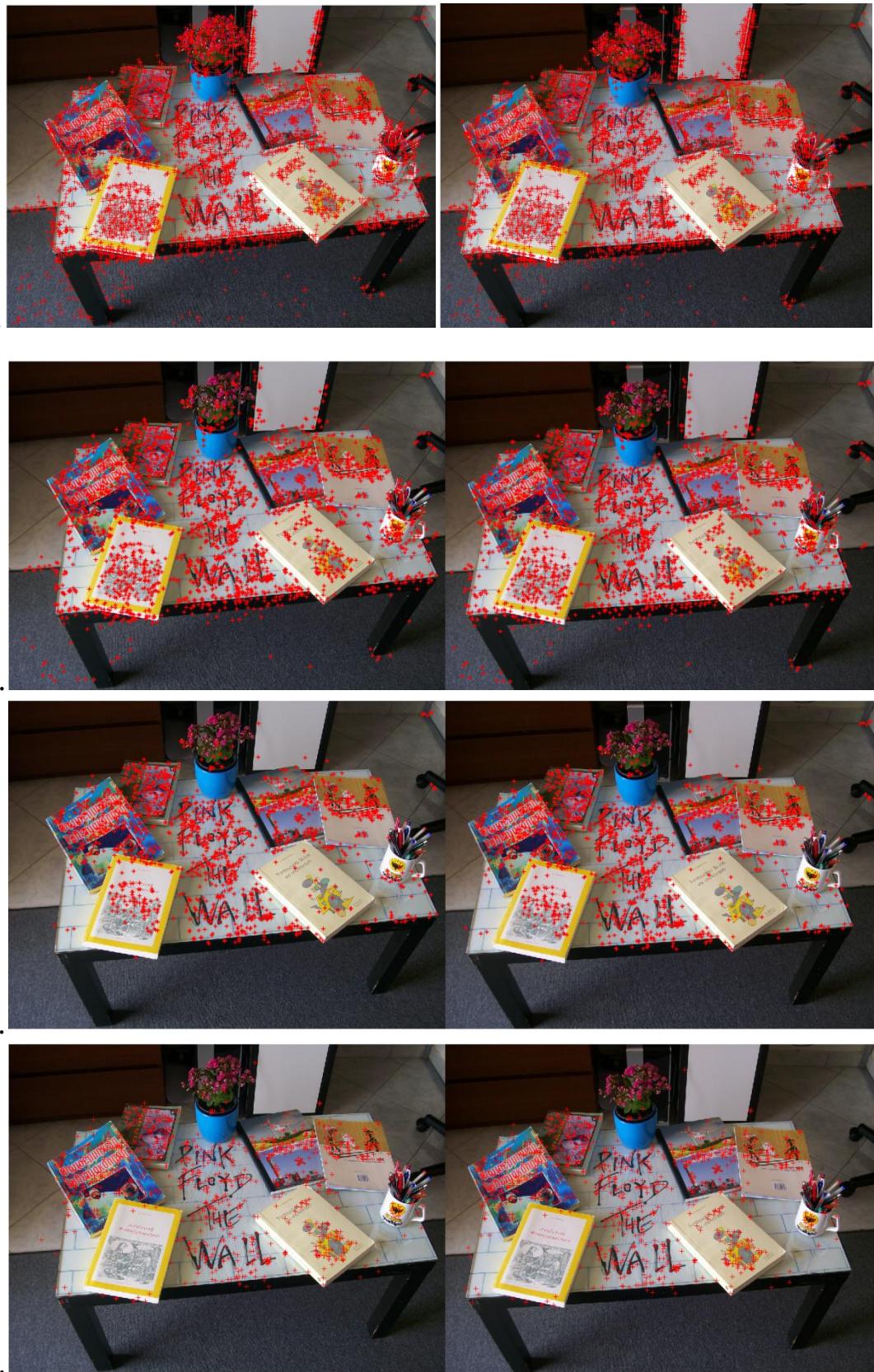
SURF (0023-0024)



SIFT (0023-0024)



SURF (0973-0974)

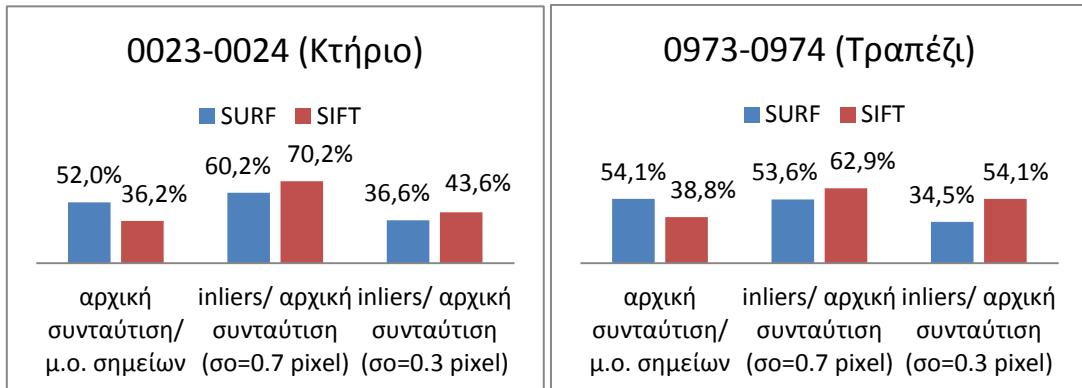


SIFT



Στερεοζεύγος	Αλγόριθμος	Σημεία στην αριστερή εικόνα	Σημεία στη δεξιά εικόνα	Αρχική συνταύτιση	Inliers (Ransac, $\sigma_o=0.7$)	Inliers (Ransac, $\sigma_o=0.3$)
0023-0024	SURF	6220	6128	3210	1934	1176
	SIFT	8315	8664	3071	2156	1338
0973-0974	SURF	5118	4828	2690	1442	929
	SIFT	8336	8447	3253	2046	1760

Πίνακας 4.2



Γραφήματα 4.1 και 4.2

Αυτό που παρατηρείται, με βάση τα γραφήματα 4.1 και 4.2, είναι τα σχετικά υψηλά ποσοστά και των δύο μέτρων αξιολόγησης. Αυτό οφείλεται, κυρίως, στη διάταξη των εικόνων, καθώς συνιστούν μια ‘εύκολη’ γεωμετρία στερεοζεύγους (μικρή Βάση, σχεδόν μηδενική στροφή, σχεδόν ίδια απόσταση λήψης). Ο SURF συσχετίζει σχεδόν τα μισά από τα σημεία που ανιχνεύει, ενώ ο SIFT το 1/3 και στα δύο ζεύγη. Όσον αφορά το ποσοστό των inliers, ο SIFT υπερτερεί του SURF σε όλες τις περιπτώσεις και των δύο όμως, τα ποσοστά είναι υψηλά. Ο SURF για RANSAC με $\sigma_o=0.7$, δίνει ελαφρώς υψηλότερα ποσοστά στο πρώτο ζεύγος, ενώ για $\sigma_o=0.3$ τα ποσοστά του είναι σχεδόν ίδια και στα δύο ζεύγη. Στον SIFT παρατηρείται το εξής: στο πρώτο ζεύγος, καθώς η τιμή του κατωφλίου μειώνεται από 0.7 σε 0.3, το ποσοστό των inliers μειώνεται σχεδόν κατά 30%, ενώ η αντίστοιχη μείωση στο δεύτερο είναι περίπου 9%, το οποίο σημαίνει ότι η ανίχνευση και περιγραφή των σημείων στο δεύτερο ζεύγος έγιναν εξαρχής με μεγαλύτερη ακρίβεια.

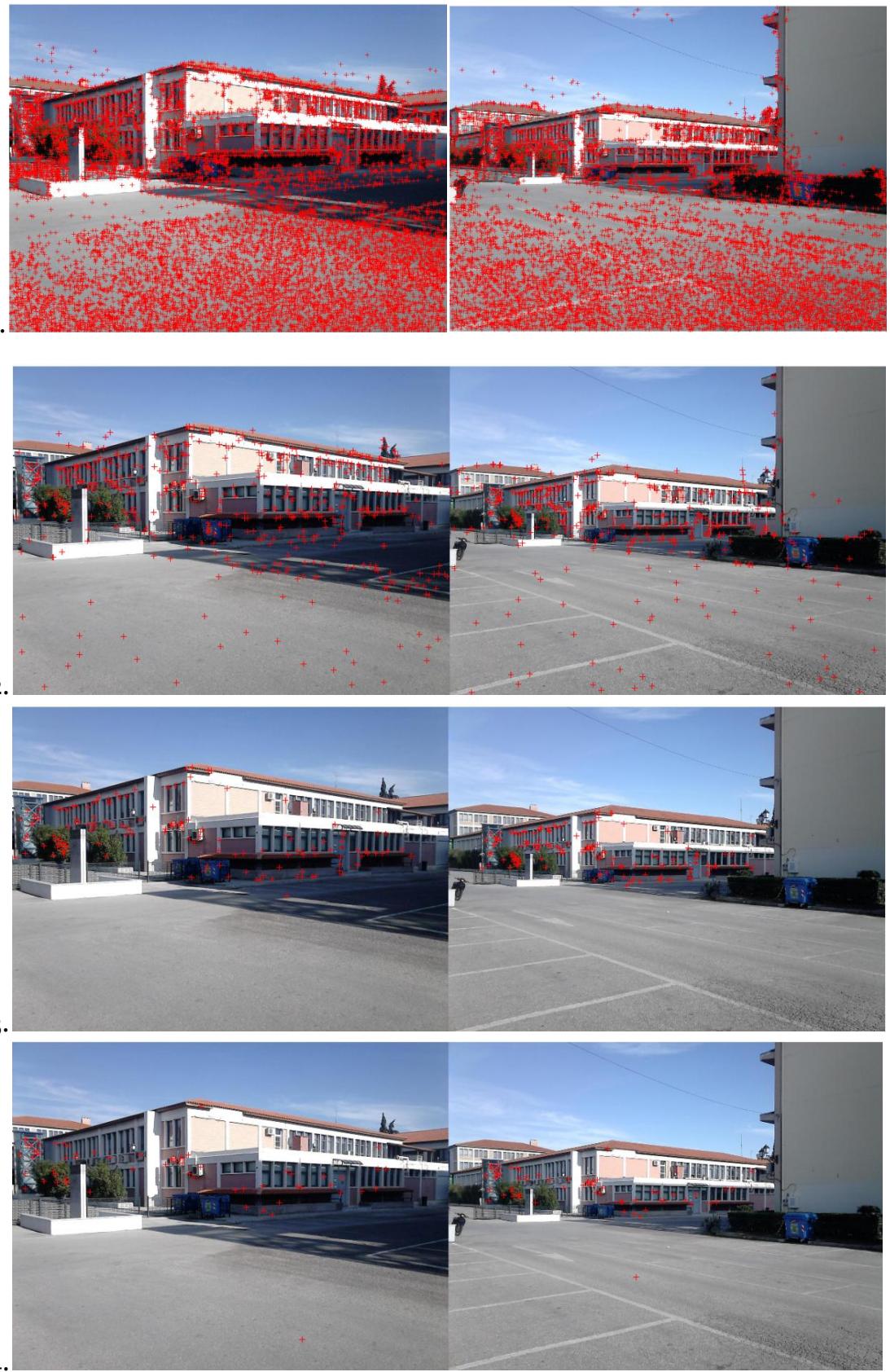
Όσον αφορά στην ανίχνευση σημείων, παρατηρείται ότι ο SIFT εξάγει μεγάλο αριθμό σημείων σε ομοιογενείς περιοχές (ουρανός, άσφαλτος, χαλί), ενώ τα σημεία του SURF επικεντρώνονται στις περιοχές με έντονη υφή. Αυτό εξηγεί και τα χαμηλότερα ποσοστά αρχικής συνταύτισης του SIFT σε σχέση με τον SURF, καθώς οι αποστάσεις των descriptors σε αυτές τις περιοχές είναι σχετικά κοντινές, οπότε δύσκολα θα βρεθεί descriptor που η απόσταση του από τον εγγύτερό του στην άλλη εικόνα, πολλαπλασιαζόμενη επί 1.5, θα είναι μικρότερη από την απόσταση που θα έχει με τον αμέσως εγγύτερο, όπως ορίζει το κριτήριο που γίνεται η αρχική συνταύτιση (βλ. ενότητα 4.1).

4.1.2 Διαφορά κλίμακας

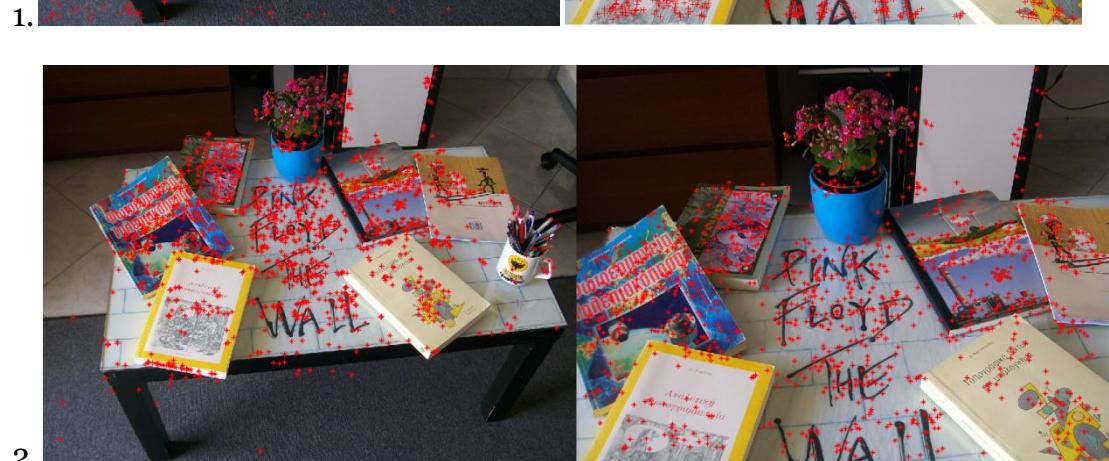
SURF (0037-0061)



SIFT (0037-0061)



SURF (0973-0977)

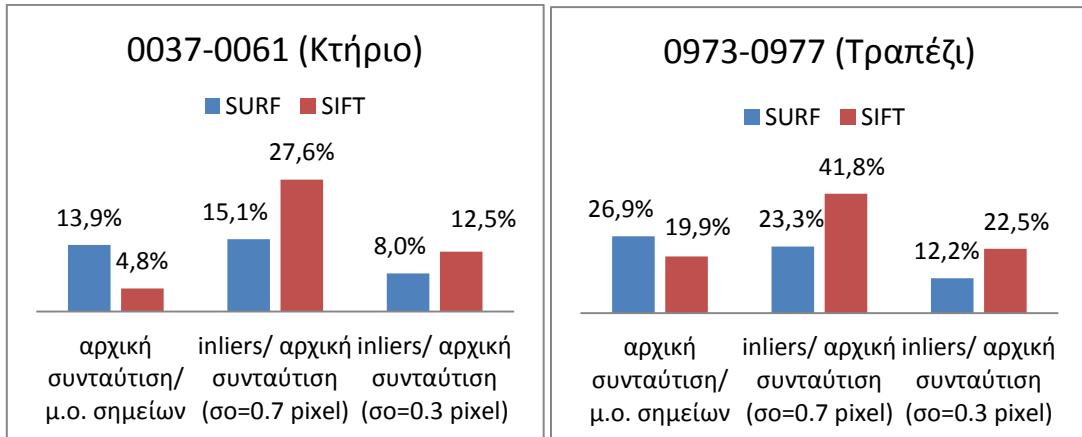


SIFT (0973-0977)



Στερεοζεύγος	Αλγόριθμος	Σημεία στην αριστερή εικόνα	Σημεία στη δεξιά εικόνα	Αρχική συνταύτιση	Inliers (Ransac, $\sigma_0=0.7$)	Inliers (Ransac, $\sigma_0=0.3$)
0037-0061	SURF	5587	4664	715	108	57
	SIFT	13547	10277	576	159	72
0973-0977	SURF	5118	5063	1368	319	167
	SIFT	8336	6831	1506	629	339

Πίνακας 4.3



Γραφήματα 4.3 και 4.4

Σε αυτή την περίπτωση, λόγω της μικρότερης επικάλυψης μεταξύ των εικόνων αναμένεται να είναι μικρότερα και τα ποσοστά της αρχικής συνταύτισης. Το ποσοστό για τον SURF είναι διπλάσιο στο δεύτερο ζεύγος σε σχέση με το πρώτο. Επίσης, υπερτερεί και στις δύο περιπτώσεις του SIFT. Όσον αφορά τον SIFT, ενώ στο δεύτερο ζεύγος εμφανίζει ένα ικανοποιητικό ποσοστό αρχικής συνταύτισης (19,9%), στο πρώτο ζεύγος το αντίστοιχο ποσοστό είναι εξαιρετικά χαμηλό (4,8%). Όπως τονίστηκε και στην §4.1.1, αυτό οφείλεται στο ότι ο SIFT εξάγει μεγάλο αριθμό σημείων σε ομοιογενείς περιοχές (άσφαλτος) και επιπλέον, στη συγκεκριμένη περίπτωση, ένα μεγάλο μέρος μιας τέτοιας περιοχής, βρίσκεται εκτός της επικαλυπτόμενης περιοχής (βλ. εικ.1 του ζεύγους 0037-0061).

Όσον αφορά τα ποσοστά των inliers παρατηρείται ότι:

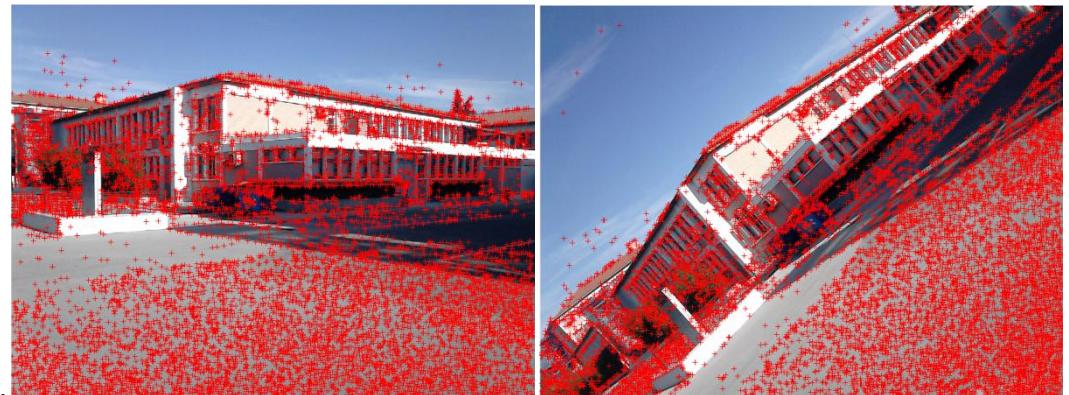
- Ο SIFT είναι σχεδόν διπλάσιος του SURF σε όλες τις περιπτώσεις, γεγονός που δείχνει ότι ο Descriptor του SIFT είναι ανθεκτικότερος σε μεταβολές της κλίμακας των εικόνων.
- Και για τους δύο αλγόριθμους και στα δύο ζεύγη εικόνων, το ποσοστό των inliers υποδιπλασιάζεται καθώς μειώνεται το κατώφλι του RANSAC από 0.7 σε 0.3.
- Και για τους δύο αλγόριθμους, όλα τα ποσοστά των inliers είναι αρκετά μεγαλύτερα στο δεύτερο ζεύγος, από τα αντίστοιχα τους στο πρώτο, γεγονός που κυρίως οφείλεται στο ότι, το δεύτερο ζεύγος παρουσιάζει αναλογικά, μικρότερη μεταβολή κλίμακας.

4.1.3 Στροφή

SURF (0037-0039)



SIFT (0037-0039)



SURF (0973-0975)

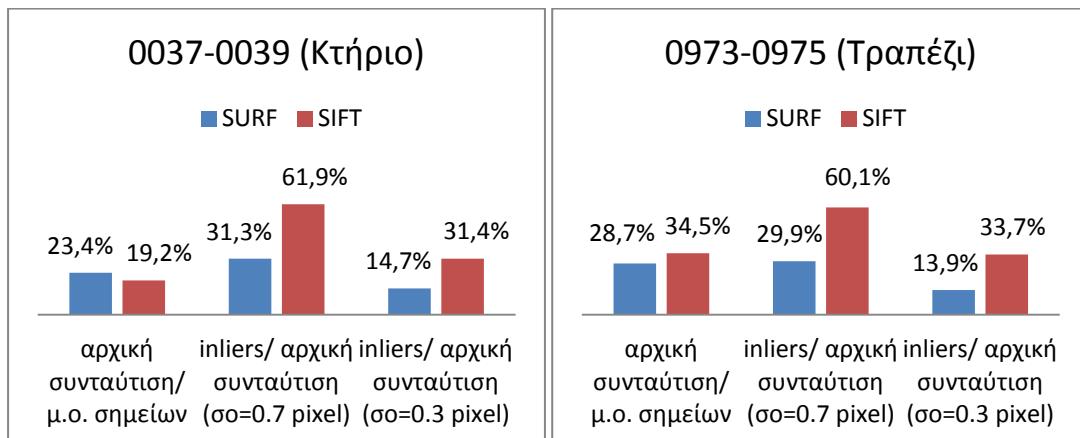


SIFT (0973-0975)



Στερεοζεύγος	Αλγόριθμος	Σημεία στην αριστερή εικόνα	Σημεία στη δεξιά εικόνα	Αρχική συνταύτιση	Inliers (Ransac, $\sigma_o=0.7$)	Inliers (Ransac, $\sigma_o=0.3$)
0037-0039	SURF	5587	3923	1113	348	164
	SIFT	13547	13819	2632	1628	826
0973-0975	SURF	5118	4071	1320	395	183
	SIFT	8336	7065	2656	1596	896

Πίνακας 4.4



Γραφήματα 4.5 και 4.6

Όπως φαίνεται και στα γραφήματα 4.5 και 4.6, στην περίπτωση της στροφής μεταξύ των εικόνων, ο SIFT υπερέχει κατά πολὺ του SURF. Μάλιστα, είναι η πρώτη φορά που παρουσιάζει μεγαλύτερο ποσοστό αρχικής συνταύτισης (δεύτερο ζεύγος).

Όσον αφορά τα ποσοστά των inliers παρατηρείται ότι:

- Ο SIFT παρουσιάζει υπερδιπλάσια ποσοστά έναντι του SURF σε όλες τις περιπτώσεις, γεγονός που κατά κύριο λόγο, δείχνει ότι η μέθοδος του πρώτου να αναθέτει προσανατολισμό στους Descriptors είναι πολὺ πιο αποτελεσματική.
- Και για τους δύο αλγόριθμους και στα δύο ζεύγη εικόνων, το ποσοστό των inliers υποδιπλασιάζεται καθώς μειώνεται το κατώφλι του RANSAC από 0.7 σε 0.3.
- Τα ποσοστά και των δύο αλγορίθμων στο πρώτο ζεύγος παρουσιάζουν αμελητέες διαφορές από τα αντίστοιχα τους στο δεύτερο ζεύγος.

4.1.4 Μεγάλο μήκος Βάσης του στερεοζεύγους

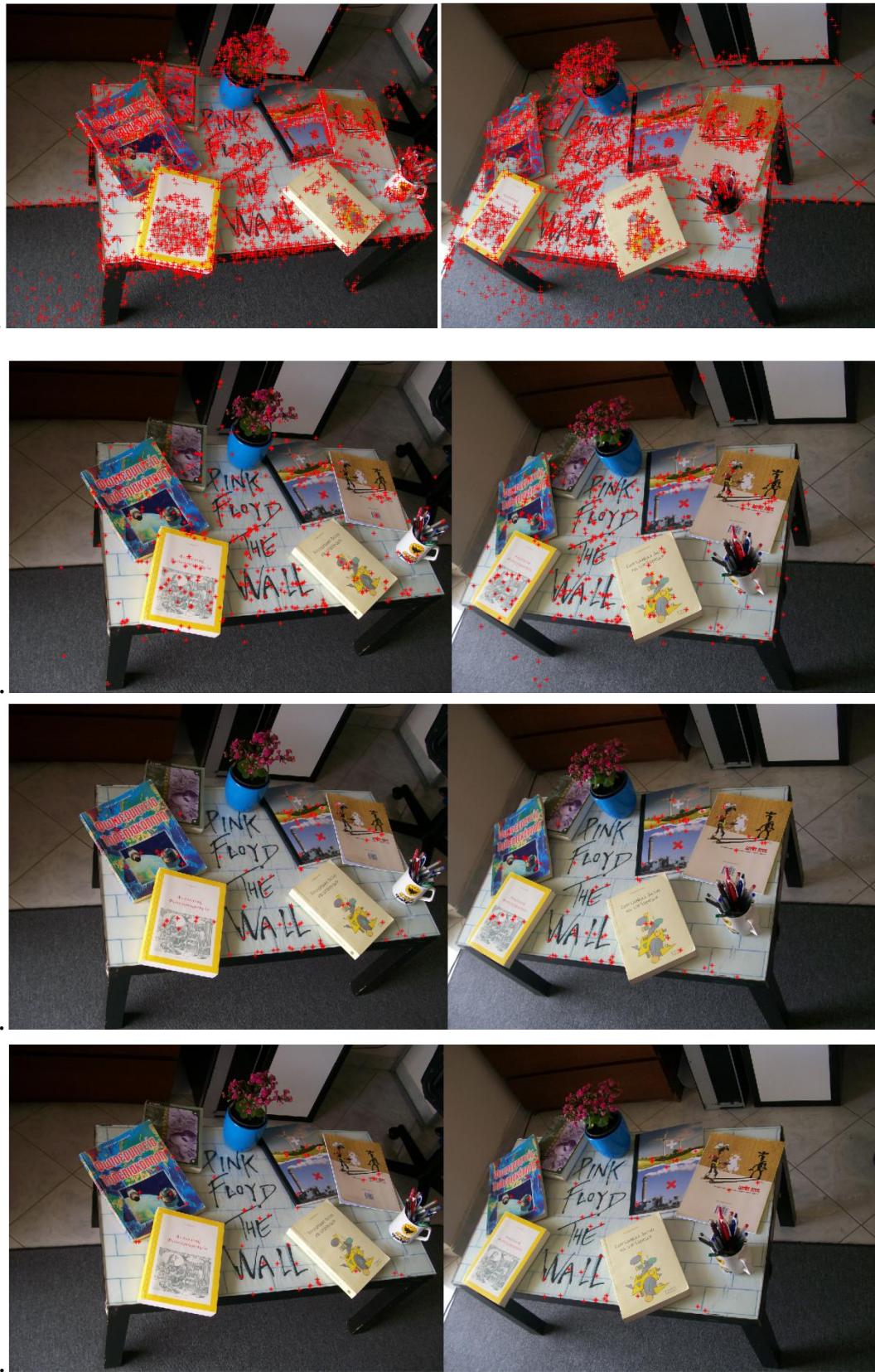
SURF (0023-0029)



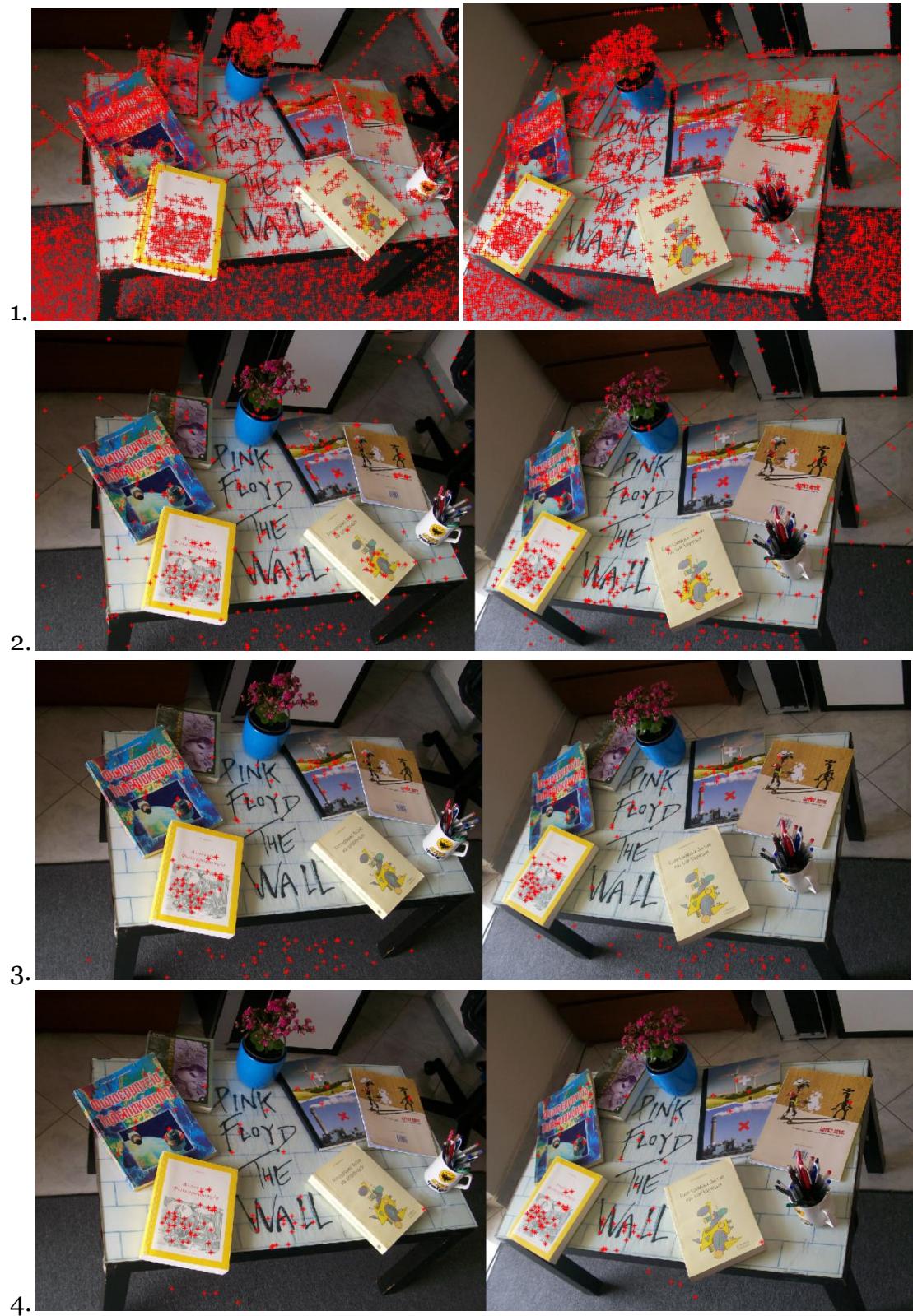
SIFT (0023-0029)



SURF (0981-0982)

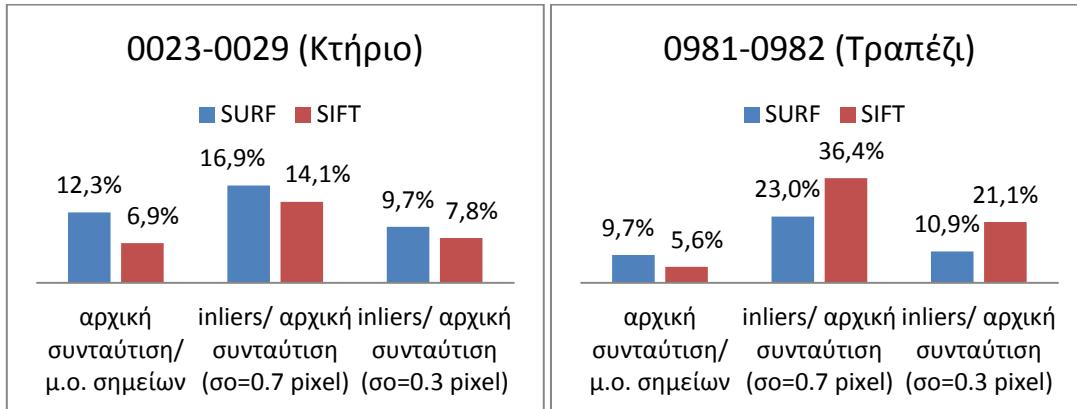


SIFT (0981-0982)



Στερεοζεύγος	Αλγόριθμος	Σημεία στην αριστερή εικόνα	Σημεία στη δεξιά εικόνα	Αρχική συνταύτιση	Inliers (Ransac, $\sigma_o=0.7$)	Inliers (Ransac, $\sigma_o=0.3$)
0023-0029	SURF	6220	5530	720	122	70
	SIFT	8315	5902	489	69	38
0981-0982	SURF	4446	4052	413	95	45
	SIFT	8387	7754	450	164	95

Πίνακας 4.5



Γραφήματα 4.6 και 4.7

Όπως φαίνεται στον πίνακα 4.5 και στα γραφήματα 4.6 και 4.7, οι τιμές των ποσοστών των μέτρων αξιολόγησης καθώς και των σημείων της αρχικής συνταύτισης και των inliers, είναι κατ' αντιστοιχία μικρότερες από όλες τις περιπτώσεις που μελετήθηκαν, γεγονός που σημαίνει ότι η συγκεκριμένη γεωμετρία στερεοζεύγους είναι η πλέον επισφαλής, όσον αφορά την απόδοση των αλγορίθμων. Στις προηγούμενες περιπτώσεις, οι άξονες λήψης των εικόνων ήταν πολύ κοντά και σχεδόν παράλληλοι μεταξύ τους, οπότε η επίδραση της προοπτικής παραμόρφωσης στο σχήμα των αντικειμένων ήταν παρόμοια. Εδώ όμως, οι δύο άξονες απέχουν αρκετά και συγκλίνουν μεταξύ τους. Αυτό έχει ως αποτέλεσμα, τα αντικείμενα να εμφανίζονται μεγάλες διαφορές ως προς την προοπτικότητά τους στις εικόνες του στερεοζεύγους, γεγονός που δυσχεραίνει τη διαδικασία της συνταύτισης. Βέβαια και στην περίπτωση της μεταβολής κλίμακας (§4.1.2), υπάρχει διαφορά στην προοπτικότητα των εικόνων, αλλά η κίνηση των λήψεων κατά τη διεύθυνση του άξονα της λήψης φαίνεται να μεταβάλλει σε μικρότερο βαθμό την προοπτικότητα τους, καθώς και την εκτροπή αναγλύφου τους, σε σχέση με την αντίστοιχη κίνηση κατά τη διεύθυνση της Βάσης.

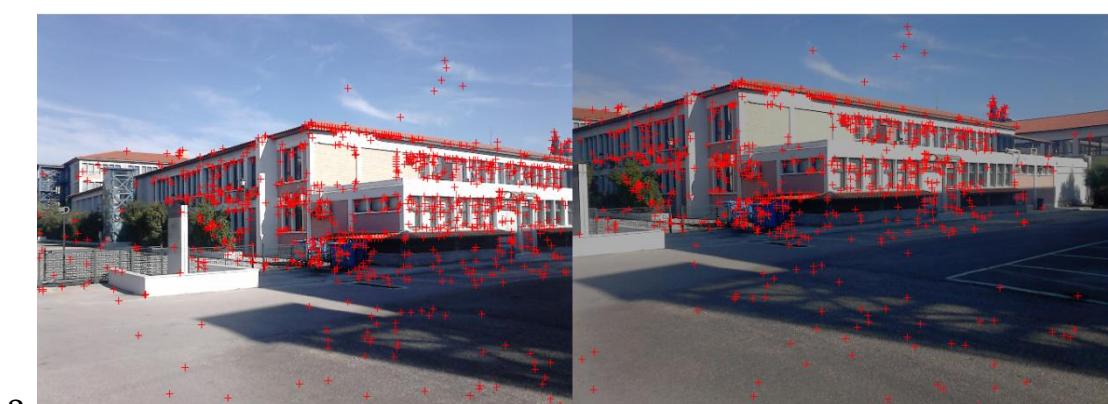
Ο SURF και σε αυτή την περίπτωση, εμφανίζει υψηλότερο ποσοστό αρχικής συνταύτισης. Επιπλέον, παρατηρείται ότι πρώτη φορά εμφανίζει μεγαλύτερο ποσοστό inliers, σε σχέση με τον SIFT (πρώτο ζεύγος), η διαφορά τους ωστόσο είναι μικρή. Από την άλλη, για το δεύτερο ζεύγος, το ποσοστό των inliers του SIFT είναι σημαντικά μεγαλύτερο (σχεδόν διπλάσιο) έναντι του SURF και στις δύο περιπτώσεις εφαρμογής του RANSAC.

4.1.5 Διαφορά φωτεινότητας

SURF (0023-0028)



SIFT (0023-0028)



SURF (0979-0980)

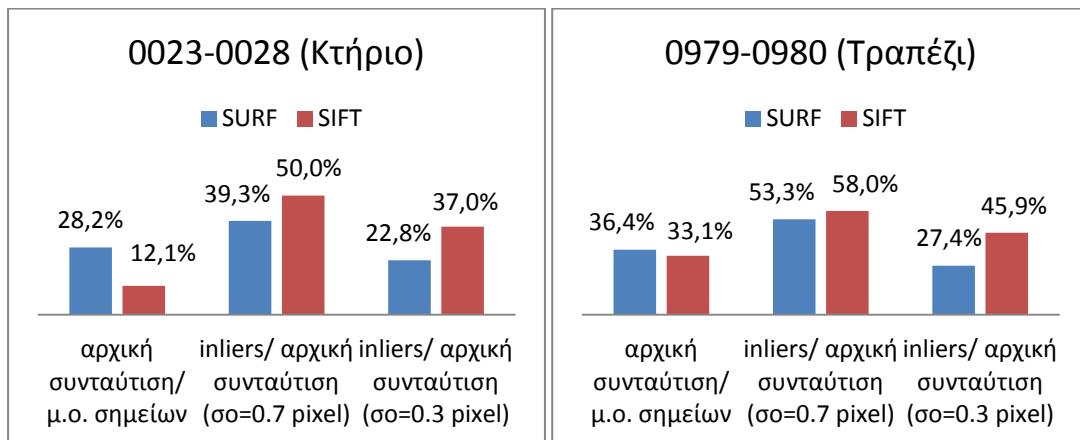


SIFT (0979-0980)



Στερεοζεύγος	Αλγόριθμος	Σημεία στην αριστερή εικόνα	Σημεία στη δεξιά εικόνα	Αρχική συνταύτιση	Inliers (Ransac, $\sigma_0=0.7$)	Inliers (Ransac, $\sigma_0=0.3$)
0023-0028	SURF	6220	3986	1439	566	328
	SIFT	8315	8758	1032	516	382
0979-0980	SURF	2920	5207	1481	790	406
	SIFT	6720	7463	2344	1360	1075

Πίνακας 4.6



Γραφήματα 4.8 και 4.9

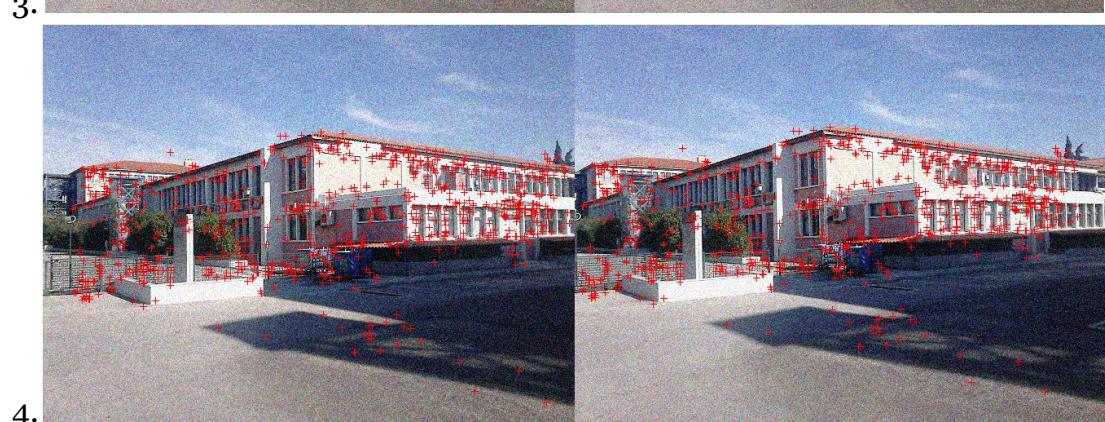
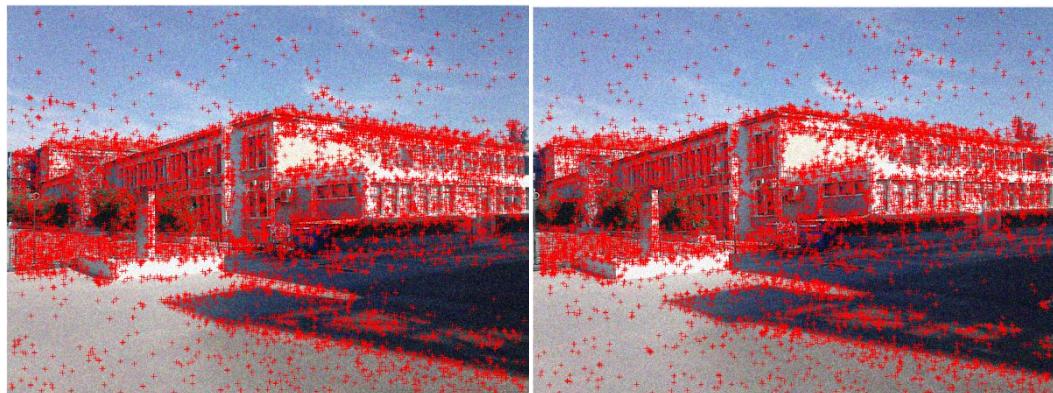
Όπως φαίνεται στον πίνακα 4.6, ο αριθμός των σημείων που εξάγει ο SURF μειώνεται σημαντικά, καθώς μειώνεται η φωτεινότητα στην εικόνα, ενώ ο SIFT δεν φαίνεται να επηρεάζεται από αυτό. Όσον αφορά την αρχική συνταύτιση, ο SURF συνταυτίζει περίπου το 1/3 των σημείων που εξάγει. Τα ποσοστά του μέτρου αυτού όμως, στην προκειμένη περίπτωση, δεν είναι ενδεικτικά της απόδοσης του αλγόριθμου, γιατί υπάρχει μεγάλη διαφορά μεταξύ του πλήθους των σημείων της μιας από την άλλη εικόνα κάθε ζεύγους. Ως εκ τούτου, θα λέγαμε ότι η ικανότητα του αλγόριθμου να ανιχνεύει ομόλογα σημεία στις εικόνες, υποεκτιμάται. Ο SIFT από την άλλη, συσχετίζει περίπου το 1/3 των σημείων στο δεύτερο ζεύγος, ενώ στο πρώτο, μόλις το 1/10. Αυτό, όπως έχει ήδη εξηγηθεί, οφείλεται στο μεγάλο αριθμό σημείων που εξάγει στις ομοιογενείς περιοχές της εικόνας, πράγμα που οδηγεί τον αλγόριθμο για την αρχική συνταύτιση που χρησιμοποιείται εδώ, να δίνει λιγότερες ομολογίες (βλ. §4.1).

Όσον αφορά τα ποσοστά των inliers, παρατηρείται ότι:

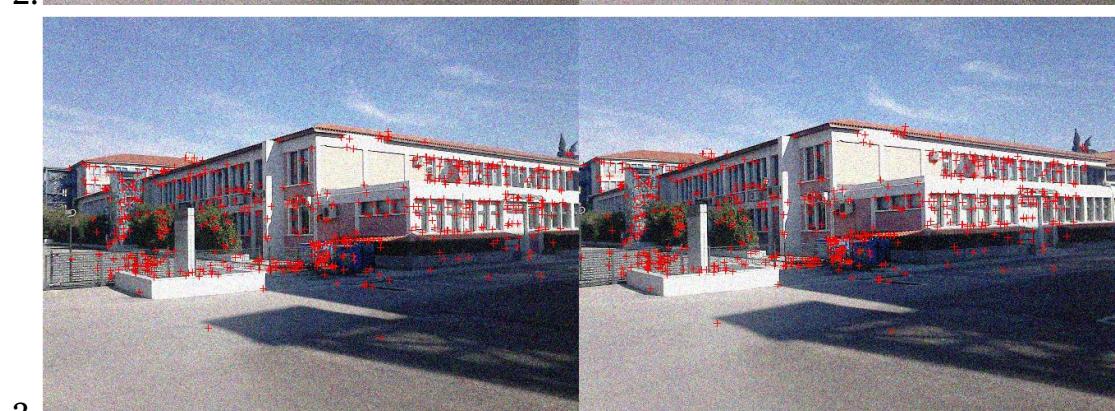
- Τα ποσοστά και των δύο αλγορίθμων είναι αρκετά υψηλά, με τον SIFT να υπερτερεί και πάλι του SURF.
- Καθώς, η τιμή του κατωφλίου του RANSAC μειώνεται από 0.7 σε 0.3, το ποσοστό των inliers για τον SURF μειώνεται περίπου στο μισό, ενώ η αντίστοιχη μείωση για τον SIFT είναι της τάξης του 25%, ενδεικτικό της υψηλής απόδοσης του Descriptor του.

4.1.6 Προθήκη θορύβου

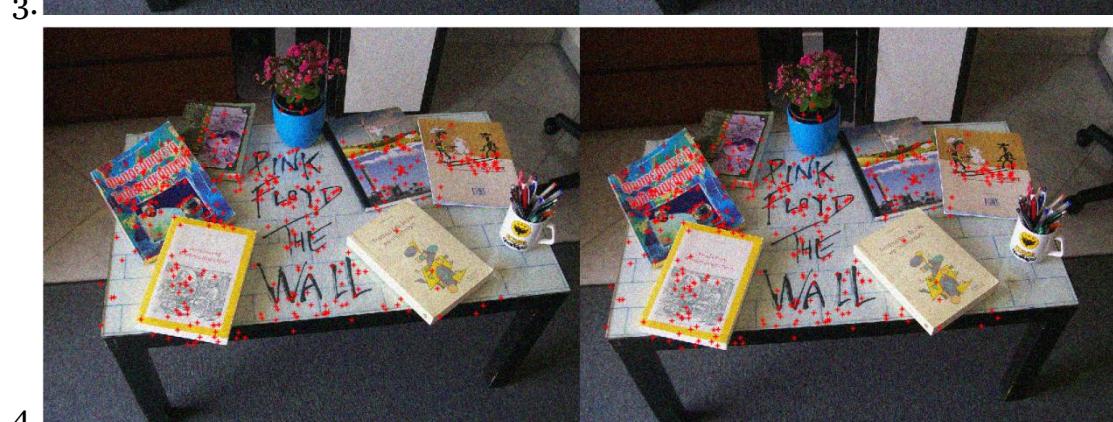
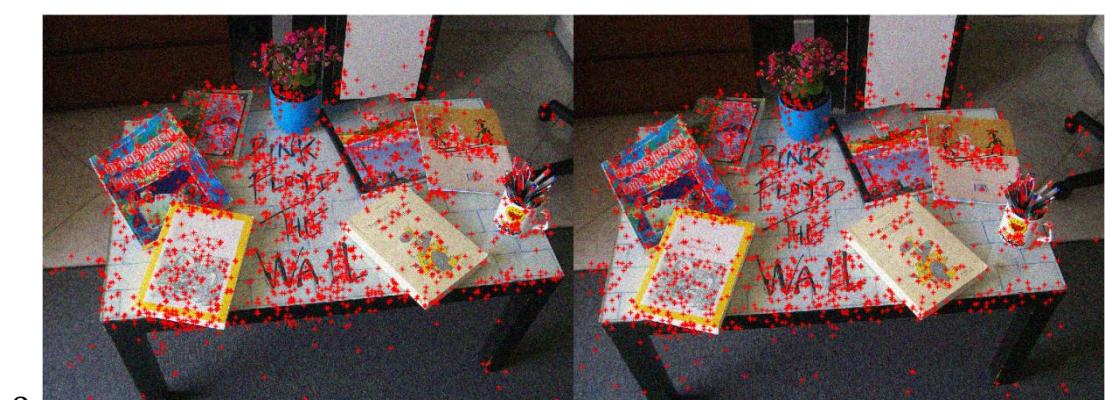
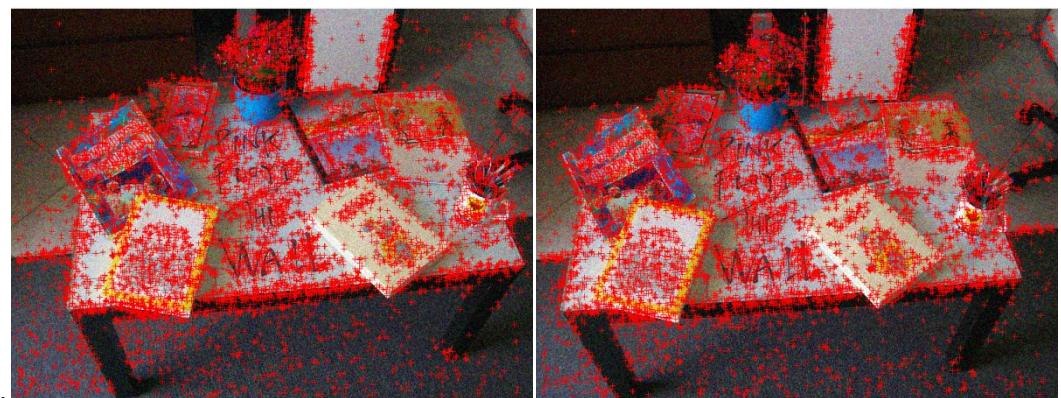
SURF (0023-0024)



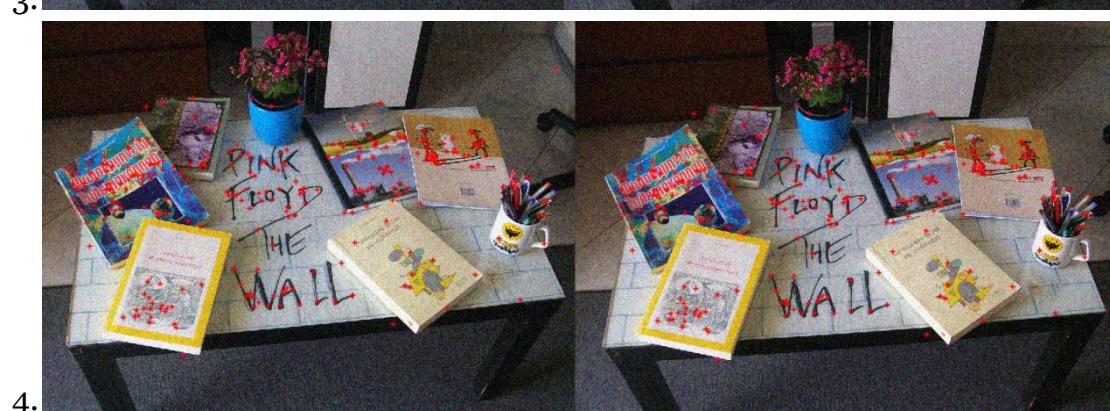
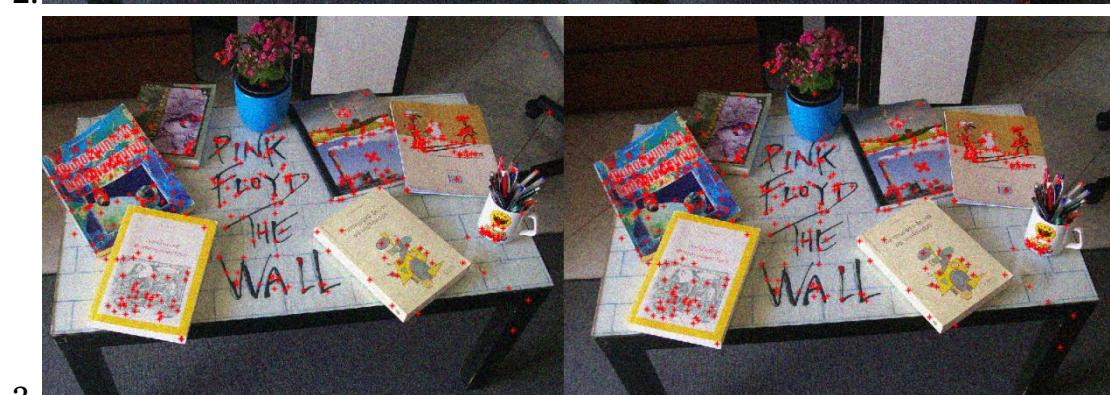
SIFT (0023-0024)



SURF (0973-0974)

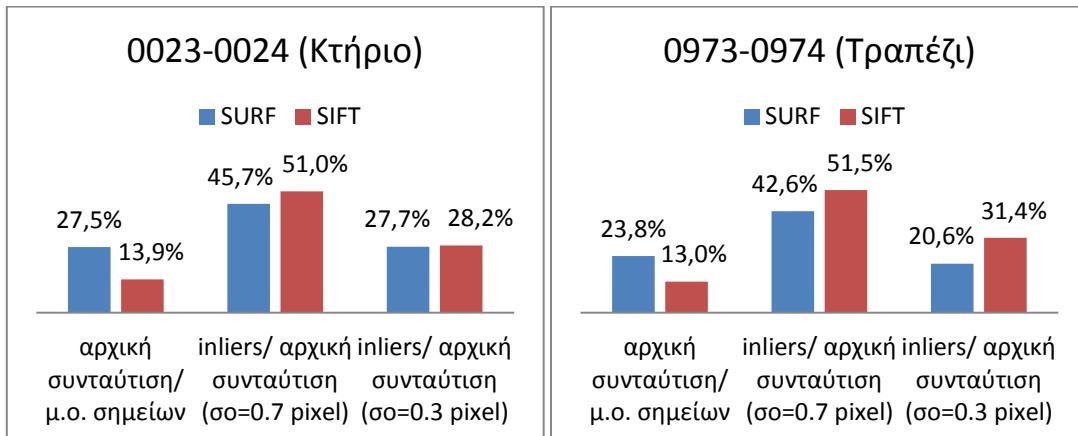


SIFT (0973-0974)



Στερεοζεύγος	Αλγόριθμος	Σημεία στην αριστερή εικόνα	Σημεία στη δεξιά εικόνα	Αρχική συνταύτιση	Inliers (Ransac, $\sigma_o=0.7$)	Inliers (Ransac, $\sigma_o=0.3$)
0023-0024	SURF	9423	9434	2592	1184	719
	SIFT	8810	8718	1221	623	344
0973-0974	SURF	8365	8029	1954	832	402
	SIFT	7878	7913	1024	527	322

Πίνακας 4.7



Γραφήματα 4.10 και 4.11

Όπως έχει ήδη αναφερθεί, ο θόρυβος που προστέθηκε στις εικόνες, ακολουθεί κανονική κατανομή, με $\sigma=0.01$.

Παρατηρείται ότι, ο αριθμός των σημείων που εξάγει ο SURF σε καθεμία εικόνα των δύο στερεοζεύγων, παρουσιάζει αύξηση άνω του 50%, σε σχέση με τον αντίστοιχο αριθμό των σημείων που ανιχνεύθηκαν στα ίδια στερεοζεύγη, χωρίς την παρουσία θορύβου (βλέπε πίνακα 4.2, §4.1.1). Αντίστοιχα, ο αριθμός των σημείων που εξάγει ο SIFT είναι περίπου ίδιος στο πρώτο ζεύγος, ενώ στο δεύτερο παρουσιάζει μικρή μείωση, της τάξης του 5-6%. Αυτό σημαίνει ότι ο ανιχνευτής του SIFT είναι πιο ανθεκτικός στην ύπαρξη θορύβου σε σχέση με αυτόν του SURF.

Ο SURF συνταυτίζει περίπου το 25% των σημείων που εξάγει, ποσοστό σχεδόν διπλάσιο από το αντίστοιχο του SIFT.

Όσον αφορά το ποσοστό των inliers, παρατηρείται ότι στο πρώτο ζεύγος, τα ποσοστά των inliers είναι σχεδόν τα ίδια και για τους δύο αλγόριθμους, ενώ στο δεύτερο ζεύγος τα ποσοστά του SIFT είναι υψηλότερα, περίπου κατά 10%.

4.1.7 Κατανομή των σημείων στις εικόνες

Στην παράγραφο αυτή θα μελετηθούν και θα συγκριθούν οι δύο αλγόριθμοι, ως προς την κατανομή των σημείων που εξάγονται. Σαν δεδομένα θα χρησιμοποιηθούν τα inliers που προέκυψαν για κάθε αλγόριθμο, έπειτα από την εφαρμογή του Ransac με τιμή κατωφλίου ίση με 0.7 pixel. Η κατανομή για τιμή κατωφλίου ίση με 0.3 pixel, δεν κρίνεται σκόπιμο να μελετηθεί, γιατί αυτή η περίπτωση υλοποιήθηκε περισσότερο, για να μελετηθεί η ακρίβεια των αποτελεσμάτων που δίνουν οι αλγόριθμοι.

Για τον σκοπό αυτό, συντάχθηκε κώδικας στο Matlab, που αναλυτικά, εκτελεί τις εξής εργασίες:

- Δημιουργία μιας ορθογώνιας περιοχής, τέτοιας ώστε να οριοθετηθεί η περιοχή της εικόνας στην οποία υπάρχει υφή, καθώς στις εικόνες των δεδομένων υπάρχει μεγάλο ποσοστό ομοιογενών περιοχών, που αν συμπεριληφθούν στη μελέτη θα επηρεάσουν αρνητικά την αξιοπιστία των αποτελεσμάτων.
- Η περιοχή αυτή χωρίζεται σε 12 ίσα ορθογώνια, 4 κατά τη διάσταση j και 3 κατά τη διάσταση i της εικόνας (x και y αντίστοιχα, στο σύστημα των εικονοσυντεταγμένων). Έπειτα, υπολογίζεται ο αριθμός των σημείων που περιέχονται σε κάθε τετράγωνο.
- Ο αριθμός των σημείων κάθε τετραγώνου, διαιρείται με τον συνολικό αριθμό των σημείων που περιέχονται στην περιοχή μελέτης. Ο λόγος αυτός εκφράζεται με τη μορφή ποσοστού επί τοις εκατό. Τα αποτελέσματα, καθώς και η περιοχή μελέτης προβάλλονται πάνω στην εικόνα.
- Υπολογίζεται ο μέσος όρος των ποσοστών και η τυπική τους απόκλιση. Ο μέσος όρος είναι ίδιος για όλες τις εικόνες και ισούται με:
$$\bar{x} = \frac{100\%}{12(\text{τετράγωνα})} = 8.33$$
, ενώ η τυπική τους απόκλιση δίνεται από τον τύπο:
$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (P_i - \bar{x})^2}$$
, όπου n=12 και P_i το ποσοστό του κάθε τετραγώνου.

Η τυπική απόκλιση των ποσοστών από το μέσο όρο, είναι το μέτρο που χρησιμοποιείται εδώ, ώστε να αξιολογηθεί η κατανομή των σημείων. Όσο μικρότερη είναι η τιμή της, τόσο εγγύτερα στο μέσο όρο είναι τα ποσοστά, άρα πιο ομοιόμορφη η κατανομή των σημείων εντός της περιοχής μελέτης.

Η εικόνα πάνω στην οποία εξετάζεται η κατανομή, σε κάθε περίπτωση, είναι η αριστερή (σταθερή) εικόνα κάθε στερεοζεύγους, ενώ ειδικά, στην περίπτωση μεταβολής της κλίμακας χρησιμοποιείται, όπως είναι λογικό, η εικόνα με την εγγύτερη λήψη. Παρακάτω παρουσιάζονται τα αποτελέσματα για κάθε μια από τις 6 συνθήκες που εξετάζονται οι εικόνες, ως εξής:

- Παρουσιάζονται συνολικά 4 εικόνες. Οι δύο πρώτες απεικονίζουν την αριστερή εικόνα του πρώτου στερεοζεύγους (κτηρίου), την πρώτη φορά περιέχοντας τα inliers του SURF και τη δεύτερη τα inliers του SIFT. Οι άλλες δύο απεικονίζουν την αριστερή εικόνα του δεύτερου στερεοζεύγους (τραπεζιού), στις αντίστοιχες περιπτώσεις.
- Πάνω στις εικόνες, προβάλλεται η ορθογώνια περιοχή μελέτης, χωρισμένη στα 12 ορθογώνια, όπως ήδη περιγράφηκε. Μέσα σε κάθε ορθογώνιο απεικονίζεται η τιμή του ποσοστού των σημείων του, επί των συνολικών σημείων της εικόνας.
- Κάτω από κάθε εικόνα, παρουσιάζεται η τιμή της τυπικής απόκλισης των ποσοστών της.
- Στη συνέχεια, ακολουθεί η σύγκριση των τιμών των τυπικών αποκλίσεων και ο γενικότερος σχολιασμός των αποτελεσμάτων, ώστε να κριθεί ποιος από τους δύο αλγόριθμους παρουσιάζει καλύτερη κατανομή, κατά περίπτωση.

4.1.7.1 Μικρό μήκος Βάσης του στερεοζεύγους



$\sigma=4.08$ (SURF)



$\sigma=2.68$ (SIFT)



$\sigma=5.86$ (SURF)

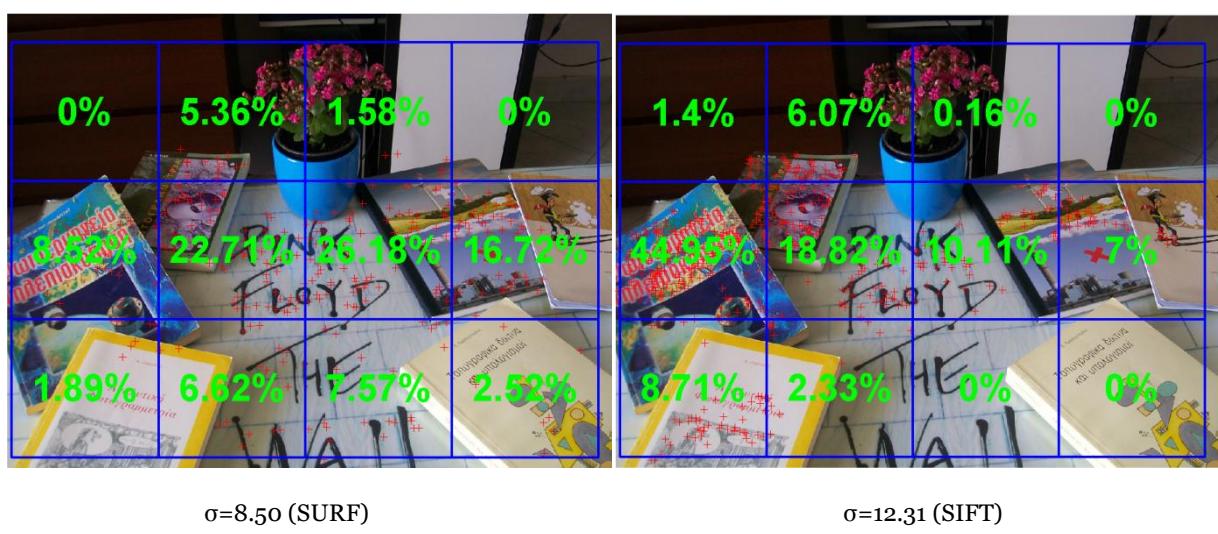
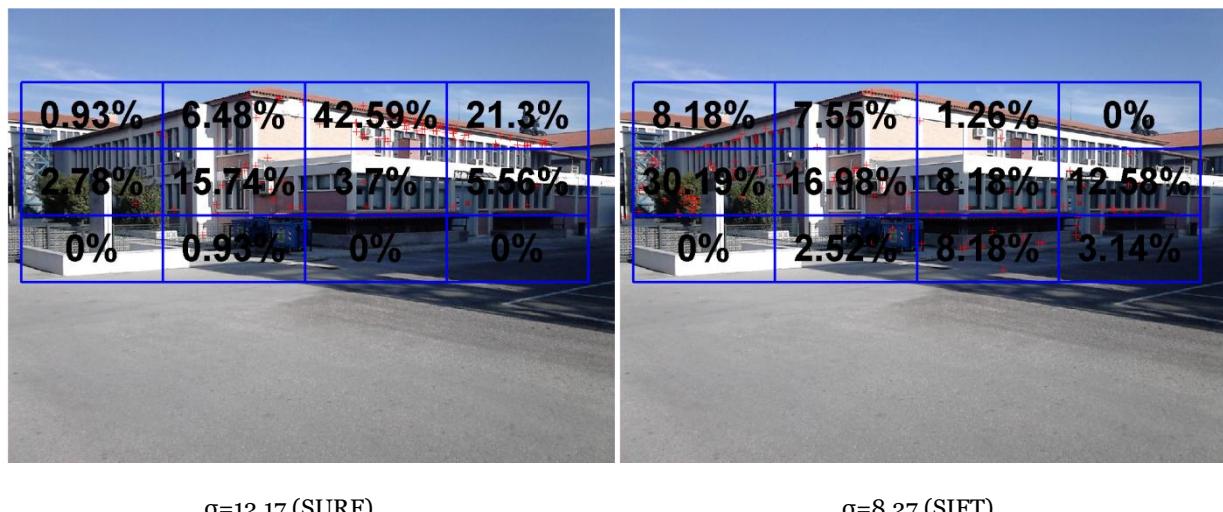


$\sigma=5.31$ (SIFT)

Σε αυτή την περίπτωση, παρατηρείται ότι η κατανομή είναι καλή σε όλες τις εικόνες. Αυτό φαίνεται και από τις τιμές των τυπικών αποκλίσεων, που είναι σχετικά χαμηλές. Παρατηρείται επίσης, ότι και στις 4 εικόνες υπάρχουν σημεία και στα 12 ορθογώνια των περιοχών μελέτης.

Ο SIFT στην πρώτη εικόνα παρουσιάζει αρκετά καλύτερη κατανομή από τον SURF, αφού η τυπική του απόκλιση είναι ίση με 2.68 έναντι 4.08 του SURF, ενώ και στη δεύτερη εικόνα υπερέχει του SURF, αλλά με μικρή διαφορά ($\sigma=5.31$ έναντι $\sigma=5.86$ αντίστοιχα).

4.1.7.2 Διαφορά κλίμακας



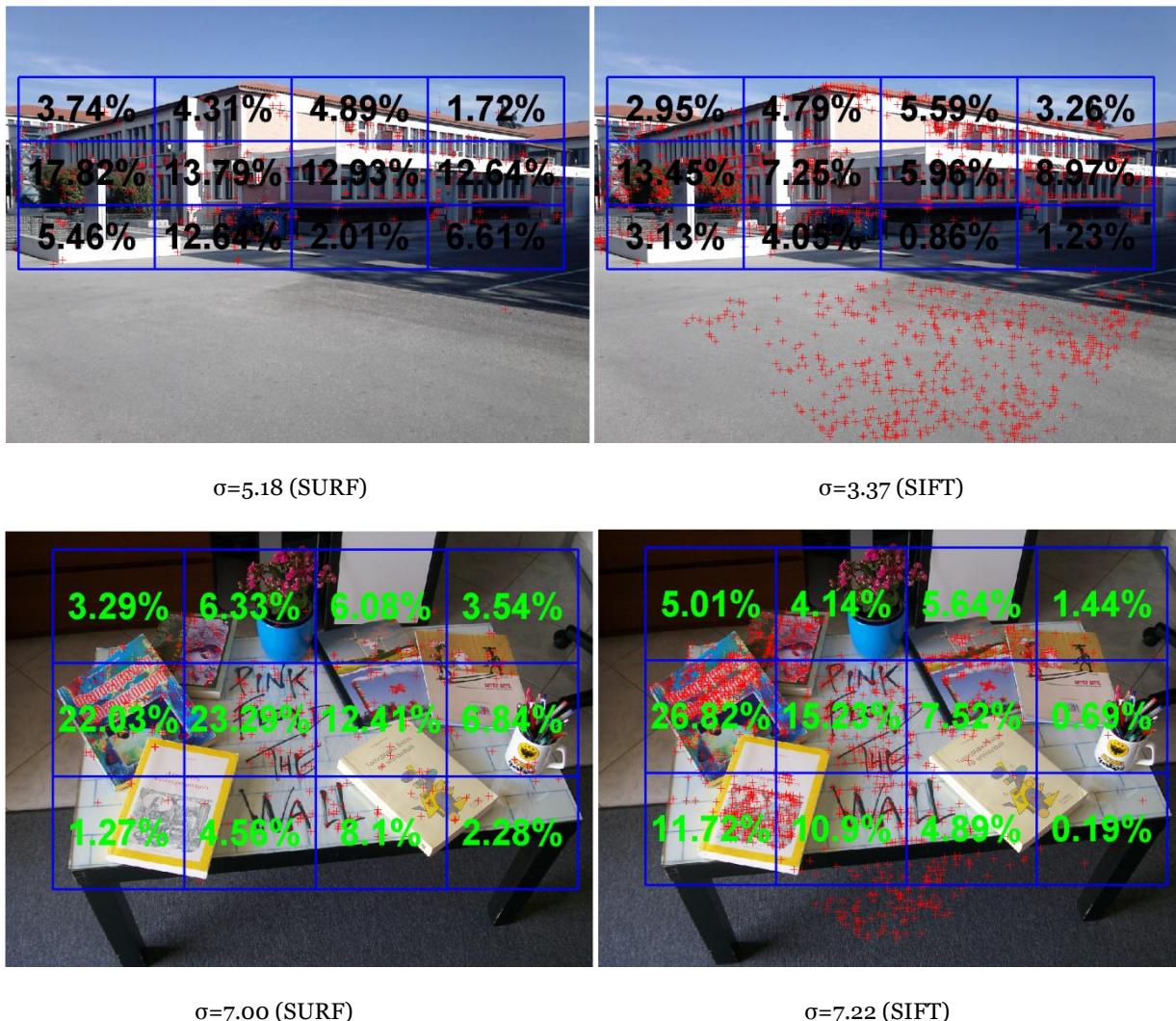
Όπως φαίνεται και στις εικόνες, η συμπεριφορά των δύο αλγορίθμων ως προς την κατανομή των σημείων, είναι αντίθετη στην πρώτη εικόνα σε σχέση με τη δεύτερη.

Στην πρώτη εικόνα ο SURF παρουσιάζει κακή κατανομή, έχοντας σχεδόν μηδαμινά σημεία στο κάτω, καθώς και στο αριστερό μέρος της ορθογώνιας

περιοχής. Από την άλλη ο SIFT παρουσιάζει μια μέτρια, αλλά αρκετά καλύτερη κατανομή από τον SURF, έχοντας ωστόσο 2 περιοχές με μηδενικό ποσοστό, κάτω αριστερά και πάνω δεξιά.

Στη δεύτερη εικόνα, τα πράγματα αντιστρέφονται. Ο SURF παρουσιάζει αρκετά καλύτερη κατανομή από τον SIFT. Επίσης, έχει 2 περιοχές με μηδενικό ποσοστό, αλλά αυτό είναι κάπως αναμενόμενο, καθώς οι περιοχές που απεικονίζουν τα δύο αυτά τετράγωνα (πάνω αριστερά και πάνω δεξιά), είναι σχετικά ομοιόμορφες. Όσον αφορά τον SIFT, η κατανομή του είναι κακή, αφού περισσότερο από το 55% των σημείων του βρίσκονται στα αριστερά της εικόνας, ενώ από τα 6 τετράγωνα δεξιά της περιοχής μελέτης τα 4 έχουν μηδενικό ποσοστό.

4.1.7.3 Στροφή



Στις εικόνες με στροφή, όπως και σε εκείνες με μικρό μήκος βάσης, η κατανομή είναι καλή σε όλες τις περιπτώσεις, όπως φαίνεται και από τις

μικρές τιμές των τυπικών αποκλίσεων. Παρατηρείται επίσης ότι υπάρχουν σημεία και στα 12 τετράγωνα της ορθογώνιας περιοχής, για όλες τις εικόνες.

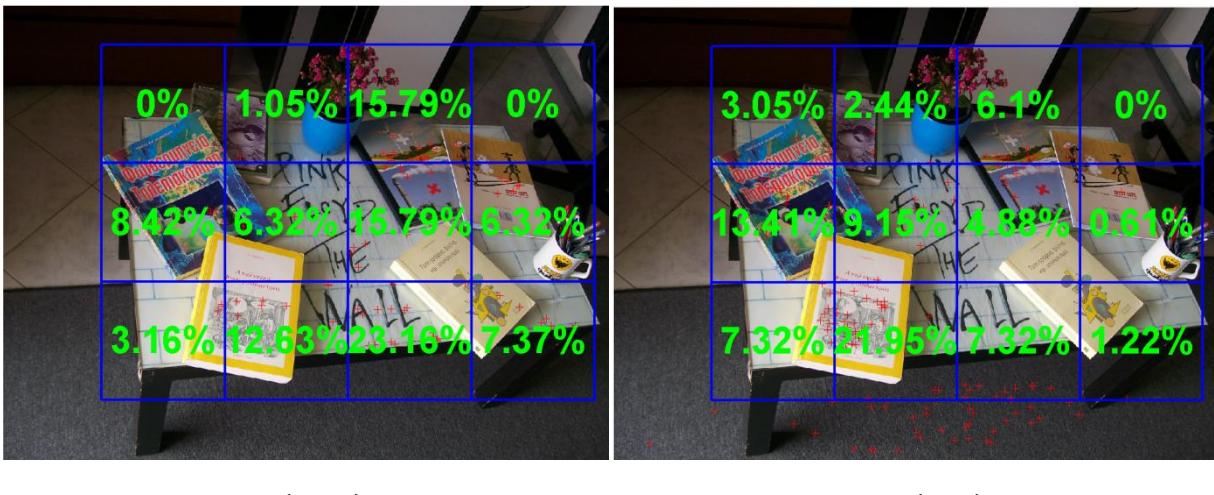
Στην πρώτη εικόνα, ο SIFT παρουσιάζει αρκετά καλύτερη κατανομή από τον SURF, αφού η τυπική του απόκλιση είναι ίση με 3.37 έναντι 5.18 του SURF, ενώ στη δεύτερη εικόνα ο SURF είναι οριακά καλύτερος, με τυπική απόκλιση ίση με 7.00 έναντι 7.22 του SIFT.

4.1.7.4 Μεγάλο μήκος Βάσης του στερεοζεύγους



$\sigma=16.97$ (SURF)

$\sigma=9.49$ (SIFT)



$\sigma=6.93$ (SURF)

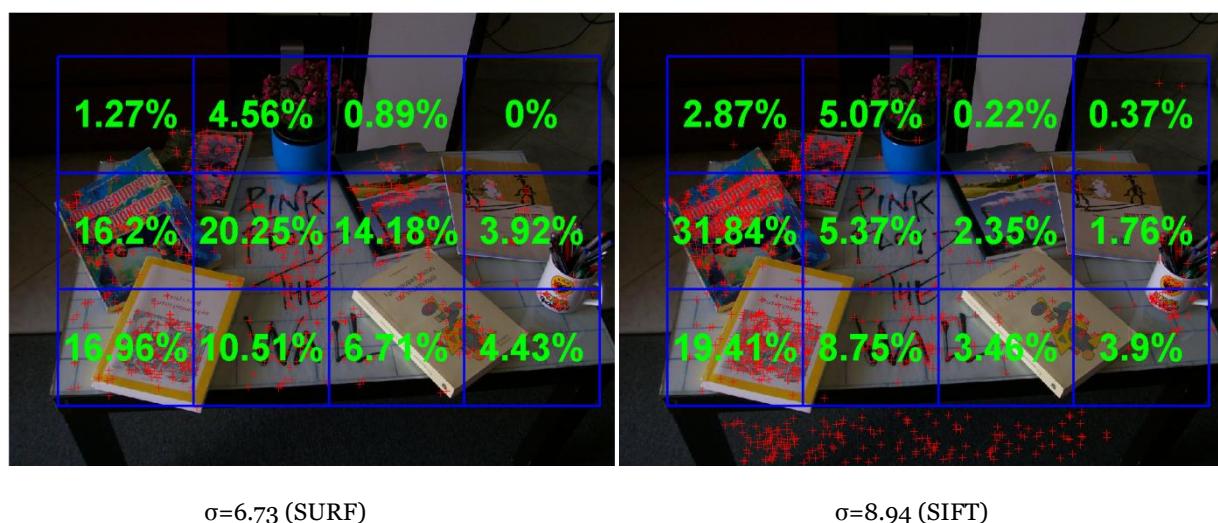
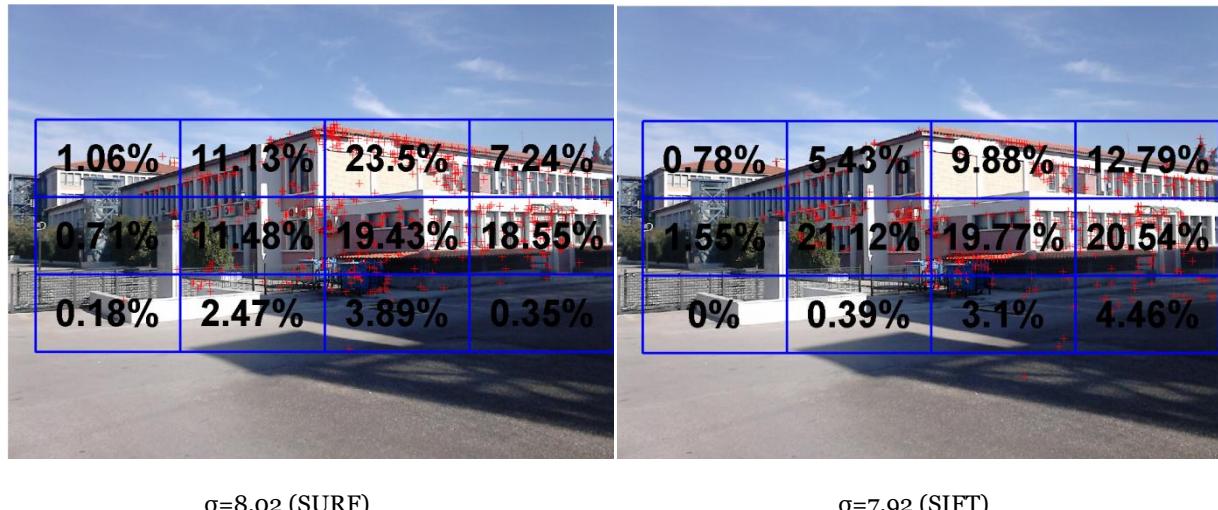
$\sigma=5.99$ (SIFT)

Όπως παρατηρείται και στις εικόνες, η περίπτωση μεγάλου μήκους Βάσης, εκτός του ότι είναι η δυσμενέστερη όσον αφορά τον αριθμό των σημείων που εξάγει (βλ. §4.1.4), παρουσιάζει και τη χειρότερη κατανομή.

Στην πρώτη εικόνα, ο SURF έχει μηδενικό ποσοστό σε 8 από τις 12 περιοχές, ενώ ο SIFT σε 5.

Στη δεύτερη εικόνα, παρατηρείται μια σχετικά καλή κατανομή και για τους δύο αλγόριθμους, με τον SIFT να έχει καλύτερη κατανομή από τον SURF και τυπικές αποκλίσεις 5.99 και 6.93 αντίστοιχα.

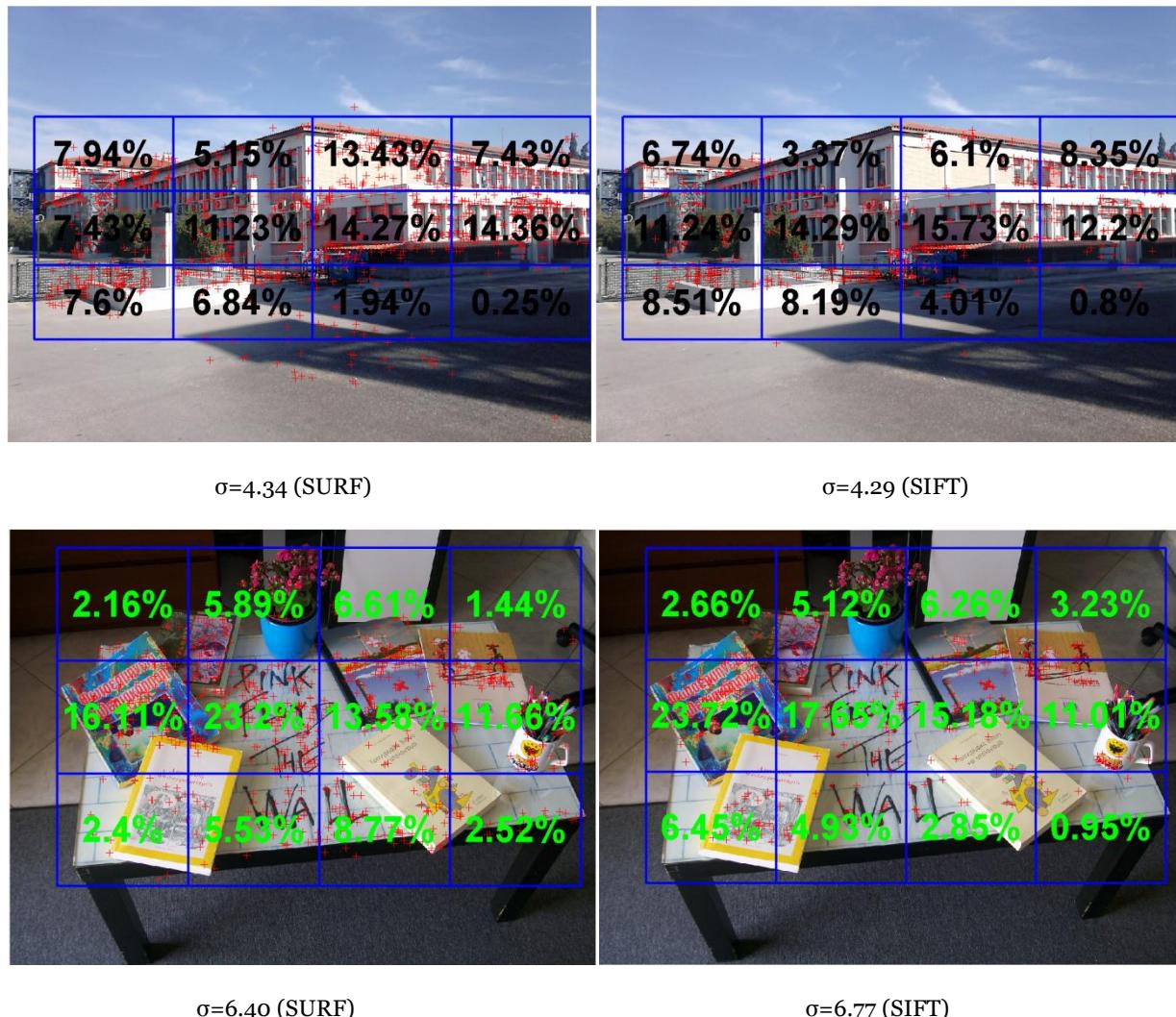
4.1.7.5 Διαφορά φωτεινότητας



Στην πρώτη εικόνα παρατηρείται μέτρια κατανομή και για τους δύο αλγόριθμους, καθώς αμφότεροι παρουσιάζουν μικρά ποσοστά στα αριστερά της περιοχής μελέτης. Όσον αφορά τις τυπικές τους αποκλίσεις, ο SIFT είναι οριακά καλύτερος του SURF, με τιμές 7.92 και 8.02 αντίστοιχα.

Στη δεύτερη εικόνα ο SURF παρουσιάζει πολύ καλύτερη κατανομή από τον SIFT. Οι τιμές των τυπικών αποκλίσεων 6.73 και 8.94 αντίστοιχα.

4.1.7.6 Προσθήκη θορύβου



Η κατανομή εδώ είναι καλή, σε όλες τις εικόνες. Αυτό είναι αναμενόμενο για δύο λόγους:

Πρώτον, τα στερεοζεύγη που χρησιμοποιήθηκαν είναι ίδια με την περίπτωση μικρού μήκους Βάσης, η οποία είναι η καλύτερη περίπτωση, βάσει όλων των κριτηρίων.

Δεύτερον, ο θόρυβος προστέθηκε στις εικόνες, με βάση συνάρτηση κανονικής (gauss) κατανομής.

ΚΕΦΑΛΑΙΟ 5

Συμπεράσματα και προτάσεις

Οι αλγόριθμοι SIFT και SURF είναι οι πλέον αναγνωρίσιμοι μεταξύ των αλγορίθμων εξαγωγής χαρακτηριστικών σημείων και σίγουρα, δύο από τους πλέον χρησιμοποιούμενους. Ένα γενικότερο συμπέρασμα που προκύπτει και από την παρούσα εργασία είναι ότι η δημοτικότητα τους αυτή είναι δικαιολογημένη.

Στο κεφάλαιο 4, μελετήθηκε επισταμένα η απόδοσή των δύο αλγορίθμων, σε 6 διαφορετικές συνθήκες μεταξύ των εικόνων. Αρχικά μελετήθηκαν ως προς το πλήθος και την ακρίβεια των ομόλογων σημείων που δίνουν και εν συνεχείᾳ, ως προς την κατανομή των σημείων αυτών. Στο κεφάλαιο αυτό, επιχειρείται να συνδυαστούν οι πληροφορίες αυτές και να απαντηθεί συμπερασματικά, ποιος από τους δύο αποδίδει καλύτερα κατά περίπτωση.

5.1 Συμπεράσματα

Η πιο ξεκάθαρη περίπτωση από όσες μελετήθηκαν είναι αυτή της ύπαρξης στροφής μεταξύ των εικόνων. Σε αυτήν, ο αλγόριθμος SIFT εμφανίζει διπλάσια και πλέον ποσοστά inliers σε σχέση με τον SURF, ενώ και η κατανομή του είναι καλύτερη. Φαίνεται ότι, η μέθοδος του SIFT (§ 3.2.3.2) να εξετάζει τις κλίσεις των pixel της περιοχής μελέτης ομαδοποιώντας τις ανά 10° , καταφέρνει να καθορίσει με μεγαλύτερη ακρίβεια τον προσανατολισμό του χαρακτηριστικού σημείου, απ' ότι ο SURF που χρησιμοποιεί το κυκλικό τμήμα εύρους 60° (§ 3.2.4.2). Σημαντικό ρόλο παίζει, επίσης, ότι ο SIFT μπορεί να δώσει περισσότερους από έναν, προσανατολισμούς στο ίδιο χαρακτηριστικό σημείο (θεωρώντας βέβαια, ξεχωριστά σημεία).

Στην περίπτωση διαφοράς κλίμακας μεταξύ των εικόνων, τα ποσοστά των inliers του SIFT είναι σχεδόν διπλάσια από αυτά του SURF. Η μέθοδος του SURF να δημιουργεί επίπεδα κλίμακας (οκτάβες) προσεγγίζοντάς τα με διαφορετικού μεγέθους box φίλτρα, αποδίδει αρκετά καλά, δεν παύει όμως να είναι προσέγγιση, σε αντίθεση με τον SIFT που κάνει το ίδιο με τον κλασικό τρόπο της υποδειγματοληψίας της εικόνας. Όσον αφορά την κατανομή, στο πρώτο στερεοζεύγος (κτήριο) υπερτερεί ο SIFT, ενώ στο δεύτερο (τραπέζι) ο SURF. Γενικότερα παρατηρείται μια τάση, ο SURF να παρουσιάζει καλύτερη κατανομή σε εικόνες με πιο έντονη υφή, πράγμα που χρήζει περεταίρω μελέτης.

Στις περιπτώσεις του μικρού μήκους Βάσης του στερεοζεύγους, της διαφοράς φωτεινότητας, καθώς και της ύπαρξης θορύβου στις εικόνες, ο SIFT εξακολουθεί να δίνει καλύτερα αποτελέσματα, ωστόσο οι διαφορές με τον SURF είναι μικρές. Με δεδομένο ότι ο SURF είναι ασυγκρίτως γρηγορότερος από τον SIFT, σε τέτοιες περιπτώσεις στερεοζευγών, αν ανάμεσα στα ζητούμενα της εκάστοτε εφαρμογής, είναι και ο χρόνος εκτέλεσής της, τότε η επιλογή του αλγορίθμου SURF είναι προτιμότερη.

Τέλος, η περίπτωση του μεγάλου μήκους Βάσης (σε σχέση πάντα με την απόσταση λήψης) του στερεοζεύγους είναι μια ‘δύσκολη’ διάταξη λήψης και για τους δύο αλγορίθμους, εξαιτίας των διαφορετικών παραμορφώσεων, που επιφέρει η διαφορετική γωνία λήψης στις εικόνες. Ο αριθμός των ομολογιών που εξάγουν και οι δύο αλγόριθμοι είναι σχετικά μικρός και η κατανομή τους είναι επίσης κακή, ειδικά του SURF. Καλό λοιπόν είναι να αποφεύγεται η συγκεκριμένη διάταξη λήψης, κατά τη συλλογή των δεδομένων. Αν δεν υπάρχει αυτή η δυνατότητα, τότε ο SIFT αποδίδει καλύτερα και σε αυτή την περίπτωση.

5.2 Προτάσεις για το μέλλον

Όπως ειπώθηκε και στην παράγραφο 5.1, η κατανομή των αλγορίθμων μπορεί να μελετηθεί διεξοδικότερα. Για παράδειγμα, θα μπορούσε να μελετηθεί η κατανομή τους σε ζεύγη εικόνων για τα οποία θα υπήρχε πιο σαφής εικόνα σχετικά με τις διαφορές στην υφή ή την αντίθεση τους. Επίσης, θα μπορούσαν να μελετηθούν οι αλγόριθμοι ως προς την κατανομή των σημείων που ανιχνεύουν αρχικά σε μια εικόνα, καθώς όπως παρατηρήθηκε στις εφαρμογές του κεφαλαίου 5, ο SURF ανιχνεύει κυρίως σημεία σε περιοχές της εικόνας με έντονη υφή, ενώ ο SIFT δίνει αρκετά σημεία και σε σχετικά ομοιογενείς περιοχές.

Στην εποχή μας, εμφανίζονται διαρκώς καινούργιοι αλγόριθμοι που προσπαθούν να συνδυάσουν αξιόπιστα αποτελέσματα με όσο δυνατόν μικρότερο υπολογιστικό φόρτο (καλύτερους χρόνους). Οι πιο πρόσφατοι από αυτούς περιγράφουν τα χαρακτηριστικά σημεία συγκρίνοντας, με δειγματοληπτική επιλογή, ζεύγη από pixel στην περιοχή γύρω από αυτά, καταλήγοντας τελικά σε μια δυαδική συμβολοσειρά (Binary Descriptors). Τέτοιοι είναι ο BRISK (§ 3.2.5), ο FREAK², ο BRIEF³ κ.α. Θα μπορούσε λοιπόν να γίνει μια αντίστοιχη σύγκριση ανάμεσα σε αυτούς.

² Fast Retina Keypoint

³ Binary Robust Independent Elementary Features

Βιβλιογραφία

Αδάμ Κ., 2011. *Γεωμετρία των στερεοζεύγους από βαθμονομημένες και από μη βαθμονομημένες μηχανές*. Διπλωματική Εργασία, ΣΑΤΜ ΕΜΠ.

Αρβανίτης Χ., *Μέθοδοι όρασης υπολογιστών για επαυξημένη πραγματικότητα*. Διπλωματική Εργασία, ΣΗΜΜΥ ΕΜΠ.

Βερυκόκου Σ., 2013. *Ανάπτυξη εφαρμογών επαυξημένης πραγματικότητας βάσει επιπέδου προτύπου*. Διπλωματική Εργασία, ΣΑΤΜ ΕΜΠ.

Γραμματικόπουλος Λ., 2007. *Γεωμετρική Πληροφορία από Μεμονωμένες Εικόνες στην Φωτογραμμετρία και την Όραση Υπολογιστών*. Διδακτορική Διατριβή, ΕΜΠ, σσ.140.

Δερμάνης Α., 1991. *Αναλυτική Φωτογραμμετρία*. Εκδόσεις Ζήτη, Θεσσαλονίκη

Καλησπεράκης, Η., 2010. *Η Επιπολική Γεωμετρία στον Προβολικό και τον Ευκλείδειο Χώρο*. Διδακτορική Διατριβή, ΣΑΤΜ ΕΜΠ.

Καρφής, Α., 2009. *Διερεύνηση Φωτογραμμετρικών Αυτοματισμών με Χρήση της Επιπολικής Γεωμετρίας*. Διπλωματική Εργασία, ΣΑΤΜ ΕΜΠ.

Κελασίδη Ε., 2009, *Ολοκληρωμένο σύστημα οδομετρίας για κινούμενα ρομπότ με χρήση μετρήσεων από πολλαπλούς αισθητήρες*. ΤΗΜΤΥ Πανεπιστήμιο Πατρών.

Κυπαρούση, Β., 2011. *Συνένωση Νεφών Σημείων από Ομολογίες Εικόνων*. Διπλωματική Εργασία, ΣΑΤΜ ΕΜΠ.

Πατιάς, Π., 1991. *Εισαγωγή στη Φωτογραμμετρία*. Ζήτη, Θεσσαλονίκη

Πέτσα, Ε., 2000. *Θεμελιώδεις Έννοιες και Θεμελιώδη Προβλήματα της Φωτογραμμετρίας*. Σπουδαστικές Σημειώσεις, Τμήμα Τοπογραφίας ΤΕΙ Αθήνας.

Φάκα Σ., 2010, *Τρισδιάστατη ανακατασκευή αντικειμένων από φωτογραφίες (με χρήση Matlab)*, Μεταπτυχιακή Διπλωματική Εργασία, Τμήμα Φυσικής Πανεπιστημίου Πατρών.

Bay H., Tuytelaars T. and van Gool L., 2006. SURF: *Speeded Up Robust Features*. Proceedings of the 9th European Conference on Computer Vision, Springer LNCS volume 3951, part 1. pp. 404–417.

Faugeras, O.D., 1992. 1992. *What can be seen in three dimensions with an uncalibrated stereo-rig?* European Conference on Computer Vision, Springer, pp. 563-578.

- Fischler, M., Bolles, R., 1981. *Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography*. Communications of the Association for Computing Machinery, 24(6), pp. 381-395.
- Harris, C., Stephens, M., 1988. *A combined corner and edge detector*. Alvey Vision Conference, pp. 147–151.
- Hartley, R., 1997a. *In defense of the eight-point algorithm*. PAMI, 19(6), pp. 580-593.
- Hartley R., Zisserman A., 2003. *Multiple View Geometry in Computer Vision*. 2nd edition, Cambridge University Press.
- Juan, L., Gwun, O. (2009). *A comparison of SIFT, PCA-SIFT and SURF*. International Journal of Image Processing, 3 (4), 143-152.
- Leutenegger, Stefan, Margarita Chli and Roland Y. Siegwart. “*BRISK: Binary robust invariant scalable keypoints*.” Computer Vision (ICCV), 2011 IEEE International Conference on. IEEE, 2011.
- Longuet-Higgins, H.C., 1981. *A computer algorithm for reconstructing a scene from two projections*. Nature, 293, pp. 133-135.
- Lowe, D., 2004. *Distinctive image features from scale invariant keypoints*. IJCV, 2(60), pp. 91-110.
- Moravec, H.P., 1979. *Visual mapping by robot rover*. Proc. 6th International Joint Conference on Artificial Intelligence, pp. 598-600.
- Torr, P.H.S., Murray, D., 1993. *Outlier detection and motion segmentation*. Sensor Fusion VI, Proc. SPIE, vol. 2059, pp. 432-443.