



Εθνικό Μετσόβιο Πολυτεχνείο
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ & ΥΠΟΛΟΓΙΣΤΩΝ

**Ανίχνευση τοπικών χαρακτηριστικών
για ανάκτηση οπτικής πληροφορίας**

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

του

ΧΡΗΣΤΟΥ Ι. ΒΑΡΥΤΙΜΙΔΗ

Διπλωματούχου Ηλεκτρολόγου Μηχανικού &
Μηχανικού Υπολογιστών Ε.Μ.Π. (2008)

Αθήνα, Μάρτιος 2016



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
& ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ & ΥΠΟΛΟΓΙΣΤΩΝ

Ανίχνευση τοπικών χαρακτηριστικών για ανάκτηση οπτικής πληροφορίας

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

TOU

ΧΡΗΣΤΟΥ Ι. ΒΑΡΥΤΙΜΙΔΗ

Διπλωματούχου Ηλεκτρολόγου Μηχανικού & Μηχανικού Υπολογιστών Ε.Μ.Π. (2008)

Συμβουλευτική Επιτροπή: Στέφανος Κόλλιας
Ανδρέας Σταφυλοπάτη
Γεώργιος Στάμου

Εγκρίθηκε από την επταμελή επιτροπή την 2^α Μαρτίου 2016.

Στέφανος Κόλλιας Ανδρέας Σταφυλοπάτης Γεώργιος Στάμου
Καθηγητής Ε.Μ.Π. Καθηγητής Ε.Μ.Π. Επ. Καθηγητής Ε.Μ.Π.

... Πέτρος Μαραγκός Ιωάννης Εμίρης Παναγιώτης Τσανάκας
Καθηγητής Ε.Μ.Π. Καθηγητής Ε.Κ.Π.Α. Καθηγητής Ε.Μ.Π.

...
Κωνσταντίνος Λουκάς
Επ. Καθηγητής Ε.Κ.Π.Α.
Αθήνα, Μάρτιος 2016

Copyright © Χρήστος Βαρυτιμίδης, 2016
Με επιφύλαξη παντός δικαιώματος. All rights reserved

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

Περιεχόμενα

1 Εισαγωγή	1
1.1 Τοπικά χαρακτηριστικά	3
1.2 α -σχήματα.....	5
1.3 Ανάκτηση ιατρικών βίντεο	7
1.4 Δομή της εργασίας.....	8
2 Σχετικές Εργασίες	11
2.1 Υπολογιστική γεωμετρία – α -σχήματα	11
2.2 Τοπικά χαρακτηριστικά	13
2.3 Βαθιά νευρωνικά δίκτυα	17
2.4 Ανάκτηση βίντεο	20
2.5 Περιοχές εικόνων που περικλείουν αντικείμενα	21
3 Ανίχνευση χαρακτηριστικών με α-σχήματα	25
3.1 Δειγματοληψία	26
3.1.1 Αναπαράσταση Εικόνων	26
3.1.2 Ομοιόμορφη δειγματοληψία	27
3.1.3 Ανομοιόμορφη δειγματοληψία	27
3.2 Τριγωνοποιήσεις	30
3.2.1 Κανονική τριγωνοποίηση	30
3.2.2 Τριγωνοποίηση Delaunay	31
3.2.3 Τριγωνοποίηση Delaunay με περιορισμούς	32
3.3 α -σχήματα.....	33
3.3.1 Σταθμισμένα α -σχήματα	33
3.3.2 Ανισοτροπικά σταθμισμένα α -σχήματα	37
3.4 Ανίχνευση χαρακτηριστικών	40
3.4.1 Συνεκτικές συνιστώσες και γειτνίαση	40
3.4.2 Επιλογή χαρακτηριστικών	43
3.4.3 Αλγόριθμος.....	45

3.5	Πειραματική αξιολόγηση	49
3.5.1	Μέτρα σημαντικότητας	53
3.5.2	Δειγματοληψία, τριγωνοποιήσεις και α -σχήματα	53
4	Δειγματοληψία εικόνων με διάχυση σφάλματος	63
4.1	Εισαγωγή	63
4.1.1	Κβαντισμός εικόνων με διάχυση σφάλματος	65
4.2	Δειγματοληψία με διάχυση σφάλματος	66
4.2.1	Διάχυση σφάλματος στην κλίση	67
4.2.2	Διάχυση σφάλματος στην Hessian	69
4.2.3	Παραδείγματα – Συζήτηση	69
4.3	Πειραματική αξιολόγηση	71
4.3.1	Βελτιστοποίηση παραμέτρων	72
4.3.2	Σύγκριση με τη βιβλιογραφία	75
4.4	Συμπεράσματα	77
5	Περιγραφή και ανάκτηση βίντεο	83
5.1	Ανάκτηση σε ιατρικά βίντεο	83
5.2	Σύνολο δεδομένων	84
5.3	Ανίχνευση πλάνων σε βίντεο	86
5.4	Περιγραφή βίντεο	87
5.4.1	Περιγραφή βίντεο με τοπικά χαρακτηριστικά	87
5.4.2	Περιγραφή βίντεο με νευρωνικά δίκτυα	89
5.5	Πειραματική αξιολόγηση	89
6	Συμπεράσματα	93
6.1	Συνεισφορά	94
6.2	Μελλοντικές επεκτάσεις	96
Αποδόσεις Ξένων Όρων	97	
Βιβλιογραφία	99	
Κατάλογος Δημοσιεύσεων	111	

Κατάλογος Εικόνων

1.1	Παραδείγματα ανίχνευσης τοπικών χαρακτηριστικών	4
1.2	Παράδειγμα της α -διήθησης	5
1.3	Περίληψη του προτεινόμενου αλγορίθμου ανίχνευσης τοπικών χαρακτηριστικών WαSH	6
2.1	Περιορισμοί των α -σχημάτων	12
2.2	Τοπικά χαρακτηριστικά που εξάγει ο προτεινόμενος αλγόριθμος WαSH ...	17
2.3	Πρόταση περιοχών του αλγορίθμου Selective Search	22
2.4	Πρόταση περιοχών του αλγορίθμου Objectness	23
3.1	Δείγματα με τη μέθοδο ανομοιόμορφης δειγματοληψίας	29
3.2	Τριγωνοποίηση Delaunay και Delaunay με περιορισμούς	33
3.3	Μέγεθος της ακμής ενός συνόλου 2 σημείων $\{p, q\}$	34
3.4	Μέγεθος του τριγώνου ενός συνόλου 3 σημείων $\{p, q, r\}$	35
3.5	Μέτρηση του μεγέθους ρ_T μιας ακμής	38
3.6	Παράδειγμα της σχέσης γειτνίασης μεταξύ άπλοκων	41
3.7	Παράδειγμα της άνω α -διήθησης	42
3.8	Ανίχνευση χαρακτηριστικών σε αντικείμενα χωρίς υφή	48
3.9	Ανίχνευση χαρακτηριστικών σε αντικείμενα χωρίς υφή, με προεπιλεγμένες παραμέτρους	49
3.10	Ανίχνευση χαρακτηριστικών σε εξωτερική σκηνή	50
3.11	Ανίχνευση χαρακτηριστικών σε εσωτερική σκηνή	51
3.12	Αποτίμηση των προτεινόμενων μέτρων σημαντικότητας	54
3.13	Μέση επαναληψιμότητα και βαθμός ταιριάσματος για διαφορετικές τριγωνοποιήσεις, α -σχήματα και δειγματοληψία	55
3.14	Επαναληψιμότητα και βαθμός ταιριάσματος: σύγκριση με τη βιβλιογραφία Α'	56
3.15	Επαναληψιμότητα και βαθμός ταιριάσματος: σύγκριση με τη βιβλιογραφία Β'	57
3.16	Ανιχνευμένα χαρακτηριστικά σε εικόνα του συνόλου Oxford 5K	58

4.1	Συντελεστές διάχυσης σφάλματος του αλγορίθμου Floyd–Steinberg	65
4.2	Δειγματοληψία με διάχυση σφάλματος.....	66
4.3	Παράδειγμα των διαφορετικών μεθόδων δειγματοληψίας	67
4.4	Ταίριασμα τοπικών χαρακτηριστικών στο σύνολο δεδομένων <i>graffiti</i>	70
4.5	Τοπικά χαρακτηριστικά <i>WaSH</i> με διαφορετικές δειγματοληψίες	71
4.6	Ρύθμιση των παραμέτρων των διαφορετικών μεθόδων δειγματοληψίας	79
4.7	Παραδείγματα δειγματοληψιών για διαφορετικές τιμές κατωφλίων.....	80
4.8	Επαναληψιμότητα και βαθμός ταιριάσματος για διαφορετικές δειγματοληψίες Α'	81
4.9	Επαναληψιμότητα και βαθμός ταιριάσματος για διαφορετικές δειγματοληψίες Β'	82
5.1	Καρέ από βίντεο λαπαροσκοπικής χολοκυστεκτομής.....	85
5.2	Τροχιές τοπικών χαρακτηριστικών	88
5.3	Τοπολογία δικτύου <i>Alexnet</i>	90
5.4	Απεικόνιση των παραμέτρων νευρωνικών δικτύων.....	91
5.5	Πειραματικά αποτελέσματα ανάκτησης βίντεο χρησιμοποιώντας περιγραφείς νευρωνικών δικτύων.....	92

Κατάλογος Πινάκων

3.1	Συνδυασμοί τριγωνοποιήσεων και α -σχημάτων	47
3.2	Αποτελέσματα του πειράματος αναζήτησης εικόνων για διαφορετικά α -σχήματα σε λεξικό 200K λέξεων	59
3.3	Αποτελέσματα του πειράματος αναζήτησης εικόνων για διαφορετικά α -σχήματα σε λεξικό 50K λέξεων	60
3.4	Αποτελέσματα του πειράματος αναζήτησης εικόνων για διαφορετικά α -σχήματα συγκρίνοντας με τον ανιχνευτή MFD	61
4.1	Αποτελέσματα του πειράματος αναζήτησης στο σύνολο εικόνων Oxford ...	75
4.2	Αποτελέσματα του πειράματος αναζήτησης στο σύνολο εικόνων Paris	75
5.1	Πειραματικά αποτελέσματα ανάκτησης βίντεο με χρήση τοπικών χαρακτηριστικών	90

I don't know anything, but I do
know that *everything is interesting*
if you go into it *deeply* enough.

Richard Feynman

ΠΡΟΛΟΓΟΣ

Η παρούσα διατριβή είναι το αποτέλεσμα της ενασχόλησής μου με το επιστημονικό πεδίο της ανάλυσης εικόνων και βίντεο. Η διατριβή εκπονήθηκε στο εργαστήριο Ψηφιακής Επεξεργασίας Εικόνας, Βίντεο και Πολυμέσων, υπό την επίβλεψη του καθηγητή κ. Στέφανου Κόλλια, στον οποίο οφείλω ιδιαίτερη ευγνωμοσύνη για την ανάθεση της εκπόνησης της διατριβής, καθώς και την εμπιστοσύνη που μου έδειξε σε όλα τα στάδια της, μέχρι και σήμερα.

Στη διάρκεια του διδακτορικού, είχα την τύχη να συνεργαστώ με τους ερευνητές Γιάννη Αβρίθη και Κώστα Ραπαντζίκο, τους οποίους θέλω να ευχαριστήσω θερμά. Ο τρόπος σκέψης του Γιάννη, το βάθος με το οποίο αναλύει κάθε τι με το οποίο ασχολείται, πιστεύω ότι με επηρέασαν αρκετά. Η διάθεσή του για διεξοδική συζήτηση σε ερευνητικά και άλλα θέματα, μου πρόσφερε πολύτιμες γνώσεις.

Καθοριστικό ρόλο στην πορεία και την εξέλιξη του διδακτορικού μου έπαιξε η συνεργασία μου με τον Κώστα Ραπαντζίκο. Η συνεργασία μας ξεκίνησε πριν ακόμα ξεκινήσω την εκπόνηση της διατριβής, από τη διπλωματική μου εργασία στη σχολή. Ο αφαιρετικός και σφαιρικός τρόπος με τον οποίο προσεγγίζει κάθε θέμα, καθώς και η διαισθητική ερμηνεία που αναζητάει, με βοήθησαν να βλέπω τα πράγματα από διαφορετικές οπτικές γωνίες.

Η εκπόνηση της διατριβής δε θα είχε ολοκληρωθεί αν δεν είχα την στήριξη των γονιών μου Γιάννη και Γιώτας, της αδερφής μου Νόπης και της συντρόφου μου Ευτυχίας. Η υπομονή και η αγάπη τους μου έδωσαν τη δύναμη να ολοκληρώσω το διδακτορικό μου. Τέλος, θέλω να ευχαριστήσω και τον γιο μου Στράτο, που τους τελευταίους 6 μήνες με το γέλιο του κάνει κάθε μέρα μου πιο ευχάριστη.

ΠΕΡΙΛΗΨΗ

Η ανάλυση εικόνων και βίντεο βασίζεται στην ανάπτυξη τεχνικών που εξάγουν οπτικά χαρακτηριστικά και δημιουργούν ενδιάμεσες αναπαραστάσεις, για την επίλυση της ανίχνευσης και αναγνώρισης αντικειμένων, της αναζήτησης και ανάκτησης εικόνων και άλλων σύνθετων προβλημάτων. Η εξαγωγή τοπικών χαρακτηριστικών για τον εντοπισμό περιοχών ενδιαφέροντος σε στατικές εικόνες και χωρο-χρονικών σημείων ενδιαφέροντος σε ακολουθίες εικόνων (βίντεο), καθώς και η περιγραφή με τοπικούς ή καθολικούς περιγραφείς, δημιουργούν συμπαγείς αναπαραστάσεις της οπτικής πληροφορίας και χρησιμοποιούνται σήμερα ευρύτατα στην περιοχή της όρασης υπολογιστών. Παρά το γεγονός ότι τα τελευταία χρόνια έχουν προταθεί πολλοί αλγόριθμοι εξαγωγής τοπικών χαρακτηριστικών, το πεδίο έρευνας είναι ανοιχτό στην ανάπτυξη και χρησιμοποίηση νέων μεθόδων που μπορούν να παράγουν αποτελεσματικές αναπαραστάσεις και καλύτερες αποδόσεις. Επίσης, η ραγδαία ανάπτυξη και οι πολύ καλές αποδόσεις βαθιών αρχιτεκτονικών νευρωνικών δικτύων τα τελευταία χρόνια οδήγησε σε ακόμη πιο αποδοτικές εφαρμογές των περιγραφέων σε εικόνες και βίντεο.

Στην παρούσα διατριβή περιγράφονται νέοι αλγόριθμοι εξαγωγής τοπικών χαρακτηριστικών για στατικές εικόνες, οι οποίοι βασίζονται σε γεωμετρικές ιδιότητες των εικόνων. Συγκεκριμένα, χρησιμοποιούνται τα α-σχήματα για την περιγραφή ενός συνόλου σημείων που εξάγονται από την εικόνα. Δεδομένου του συνόλου σημείων, τα α-σχήματα περιγράφουν αντικείμενα της εικόνας σε διάφορες κλίμακες και με διαφορετικό βαθμό λεπτομέρειας. Για την δειγματοληψία σημείων της εικόνας προτείνονται κατ' αρχήν δύο οικογένειες μεθόδων: δειγματοληψία επί ακμών και δειγματοληψία με διάχυση σφάλματος. Για τη δειγματοληψία επί ακμών, πέρα από μια βασική προσέγγιση ομοιόμορφης δειγματοληψίας που εξάγει δείγματα με σταθερή πυκνότητα, προτείνεται μια μέθοδος ανομοιόμορφης δειγματοληψίας που με εξαγωγή τοπικού αφινικού σχήματος μεταβάλει τοπικά την πυκνότητα δειγματοληψίας. Για τη δειγματοληψία με διάχυση σφάλματος προτείνονται δύο διαφορετικές προσεγγίσεις με χρήση συνεχών συναρτήσεων της φωτεινότητας της εικόνας. Η πρώτη βασίζεται στη χρήση πρώτων παραγώγων της φωτεινότητας (μέτρο κλίσης), ενώ η δεύτερη στη χρήση δεύτερων παραγώγων (απόκριση Hessian).

Στο πλαίσιο της ανάπτυξης του προτεινόμενου ανιχνευτή WαSH, εξετάζεται η χρήση

διαφορετικών τριγωνοποιήσεων και α-σχημάτων και προτείνονται ανισοτροπικά σταθμι- σμένα α-σχήματα που εκμεταλλεύονται το τοπικό σχήμα της κάθε περιοχής της τριγωνο- ποίησης. Για την επιλογή περιοχών ενδιαφέροντος, παρουσιάζονται νέα μέτρα αξιολόγη- σης των διαφορετικών συνεκτικών συνιστωσών των α-σχημάτων. Ο αλγόριθμος εξαγω- γής τοπικών χαρακτηριστικών αξιολογείται ποιοτικά και ποσοτικά, εξετάζοντας όλες τις προτεινόμενες παραλλαγές σε κάθε βήμα του αλγορίθμου. Ο προτεινόμενος ανιχνευτής εξάγει συγκριτικά μικρό αριθμό χαρακτηριστικών από περιοχές της εικόνας που αντιστοι- χούν σε τμήματα αντικειμένων που επαναλαμβάνονται συχνά. Η απόδοσή του συγκρίνε- ται με αυτή των μεθόδων της αιχμής της τεχνολογίας, ξεπερνώντας τις στις περισσότερες περιπτώσεις.

Στη συνέχεια της διατριβής, προτείνουμε έναν αλγόριθμο εξαγωγής εύρωστων ανα- παραστάσεων από ακολουθίες εικόνων (βίντεο), με χρήση βαθιών νευρωνικών δικτύων. Αρχικά γίνεται κατάτμηση των βίντεο σε πλάνα, με μια νέα μέθοδο που βασίζεται σε αλ- γορίθμους πρότασης περιοχών που έχουν αυξημένη πιθανότητα να περιλαμβάνουν αντι- κείμενα. Εξάγουμε ένα καθολικό μέτρο της πιθανότητας εμφάνισης αντικειμένων σε καρέ βίντεο, και σηματοδοτούμε την αλλαγή πλάνου στα καρέ που εμφανίζεται μεγάλη διακύ- μανση του εξεταζόμενου μέτρου. Ο αλγόριθμος χρησιμοποιεί τους χάρτες χαρακτηριστι- κών που εξάγονται από βαθιά νευρωνικά δίκτυα που έχουν χρησιμοποιηθεί για αναγνώ- ριση αντικειμένων σε εικόνες για την περιγραφή των καρέ. Έπειτα εξάγεται ένας καθολικός και συμπαγής περιγραφέας για όλο το πλάνο, συναθροίζοντας τις μέγιστες αποκρίσεις των περιγραφών. Η επίδοση του αλγορίθμου αποτιμάται σε ένα πείραμα ανάκτησης ιατρικών βίντεο, όπου επιτυγχάνεται βελτίωση της απόδοσης έναντι μεθόδων που χρησιμοποιούν τοπικά χαρακτηριστικά.

ABSTRACT

Low-level image analysis offers an intermediate image representation that is used by high-level computer vision algorithms (e.g. object detection and recognition, image and video retrieval, image matching). Local features extracted as regions of interest, or spatio-temporal interest points extracted from videos, combined with local descriptors, as well as global descriptors, offer a compact representation of visual information. Despite the fact that many local feature detectors have been proposed recently, this field of research is still open to new methods, as new and more complex application fields are introduced. Lately, the interest of the computer vision community has focused on deep neural networks, based on recent results in image classification tasks.

We propose a new local feature detector, based on geometric constructions. In particular, we propose using α -shapes to describe the shape of a set of points sampled on an image. Given the point set, α -shapes describe image objects in different scales and with different level of detail. For image sampling, we propose two different approaches: sampling on image edges and sampling using error diffusion. For sampling image edges, we propose a method that exploits the local affine shape in order to adapt sampling density, as well as a baseline method that uses fixed density sampling. We also propose sampling using error diffusion on two different functions of image intensity. The first one is based on first-order derivatives of image intensity (gradient strength), while the second one is based on second-order derivatives (Hessian response).

We use different triangulations of the samples and different α -shapes, and propose the *anisotropically weighted α -shapes* that exploit the local shape of each simplex of the triangulation. For selecting regions of interest, we propose different importance measures for the connected components of α -shapes. We qualitatively and quantitatively evaluate the proposed local feature extraction algorithm, under all proposed variations for each algorithm step. Our detector extracts a relatively small number of features from image regions that correspond to highly repeatable object parts. Its performance exceeds the state-of-the-art in most cases.

We also propose an efficient method for describing video clips, using deep neural networks. We segment videos in shots, using a novel method that exploits a global “objectness”

measure. For describing video frames, we exploit neural networks feature maps, and then aggregate the responses to create a single descriptor for the video shot. We evaluate the proposed method on a surgical video retrieval experiment, where other methods based on local features are outperformed.

Κεφάλαιο 1

Εισαγωγή

Οι εικόνες (στατικές και κινούμενες) αποτελούν μέσο αποθήκευσης πληροφορίας μεγάλης εκφραστικότητας. Η σημερινή τεχνολογία καθιστά τη σύλληψη, την αποθήκευση και το διαμοιρασμό εικόνων, εύκολη διαδικασία για κάθε χρήστη. Ως επακόλουθο, τα τελευταία χρόνια αυξάνεται ραγδαία το πλήθος των εικόνων και βίντεο που γίνονται διαθέσιμα στο διαδίκτυο. Σύμφωνα με τον ιστότοπο αποθήκευσης και διαμοιρασμού εικόνων flickr¹, οι χρήστες του ανεβάζουν περίπου 3.5 εκατομμύρια εικόνες κάθε μέρα, ενώ σύμφωνα με τον ιστότοπο διαμοιρασμού βίντεο youtube² οι χρήστες του ανεβάζουν κατά μέσο όρο 100 ώρες βίντεο κάθε λεπτό. Ακόμα, τα μέσα κοινωνικής δικτύωσης (facebook, instagram, google+ κ.α.) είτε απλά υποστηρίζουν, είτε βασίζονται στο διαμοιρασμό εικόνων και βίντεο, προσφέροντας παράλληλα άπλετο χώρο αποθήκευσης στους χρήστες.

Με την αύξηση του μεγέθους των προσωπικών συλλογών εικόνων και βίντεο κάθε χρήστη, δημιουργείται η ανάγκη οργάνωσής τους με βάση το οπτικό τους περιεχόμενο. Υπάρχουν διαθέσιμες εφαρμογές που οργανώνουν συλλογές εικόνων με βάση: α) τα γενικά “αντικείμενα” που περιέχονται σε αυτές (π.χ. ανθρώπους, ζώα, αυτοκίνητα), χρησιμοποιώντας μεθόδους ανίχνευσης αντικειμένων, β) συγκεκριμένα πρόσωπα που εμφανίζονται, χρησιμοποιώντας μεθόδους αναγνώρισης αντικειμένων, γ) μοναδικά αντικείμενα (π.χ. ορόσημα (landmarks), έργα τέχνης), με μεθόδους ανάκτησης όμοιων εικόνων, ή δ) το γενικό περιεχόμενο της σκηνής (π.χ. βουνό, θάλασσα, αστικό τοπίο), χρησιμοποιώντας τεχνολογίες κατάτμησης και αναγνώρισης σκηνών. Η αύξηση των διαθέσιμων εικόνων σε σελίδες του διαδικτύου, οδηγεί στην ανάπτυξη αλγορίθμων ανάκτησης εικόνων με βάση λεκτικά ερωτήματα, ακόμα και ανάκτησης εικόνων και άλλων πληροφοριών με χρήση ερωτήματος εικόνας. Ωστόσο, η ανάλυση περιεχομένου σε βίντεο δεν έχει χρησιμοποιηθεί σε εφαρμογές και ιστότοπους διαμοιρασμού, παρά τον τεράστιο όγκο των μη επισημασμένων βίντεο, λόγω του υπολογιστικού κόστους της διαδικασίας.

¹<http://www.flickr.com/>

²<http://www.youtube.com>

Για την αναπαράσταση της οπτικής πληροφορίας σε αλγορίθμους όπως η ανίχνευση και αναγνώριση αντικειμένων, η ανάκτηση εικόνων και βίντεο κ.α., απαιτείται ένα αρχικό στάδιο μεταφοράς από το 2Δ χώρο της εικόνας (3Δ σε περίπτωση έγχρωμων εικόνων), σε ένα άλλο χώρο πιο περιγραφικό, συνήθως μεγαλύτερης διάστασης. Η αναπαράσταση σε ένα τέτοιο χώρο γίνεται με τη χρήση τοπικών περιγραφέων όπως οι SIFT [70], SURF [8], HOG [20]. Οι περιγραφείς αυτοί έχουν συνήθως 64 έως 128 διαστάσεις, ενώ πιο πρόσφατοι περιγραφείς, όπως οι FAST [98, 99], BRIEF [11] και BRISK [62] προσπαθούν να μειώσουν τις διαστάσεις του χώρου, μειώνοντας σημαντικά τον απαιτούμενο χώρο για κάθε περιγραφέα. Οι περιγραφείς εξάγονται είτε από πυκνά (συνήθως επικαλυπτόμενα) μπλοκ της εικόνας [111], είτε από περιοχές επιλεγμένες από έναν ανιχνευτή τοπικών χαρακτηριστικών.

Πρόσφατα δημιουργήθηκαν μεγάλα σύνολα δεδομένων με καθημερινές φωτογραφίες χρηστών από το διαδίκτυο [23, 65], με στόχο την αναγνώριση αντικειμένων σε αυτά. Η εργασία των Krizhevsky et al. [54] εκμεταλλεύτηκε το μέγεθος αυτών των συνόλων για την εκπαίδευση ενός Συνελικτικού Νευρωνικού Δικτύου – ΣΝΔ (Convolutional Neural Network – CNN) μεγάλου βάθους για ταξινόμηση εικόνων, ξεκινώντας από τις τιμές χρώματος σε κάθε εικονοστοιχείο. Η χρήση βαθιών ΣΝΔ δε θα ήταν εφικτή αν δε συνέπιπτε με την ραγδαία βελτίωση της παραλληλίας σε απλούς υπολογισμούς με χρήση καρτών γραφικών. Η χρήση βαθιών ΣΝΔ επεκτάθηκε γρήγορα στην ανίχνευση αντικειμένων, ανάκτηση εικόνων και βίντεο, ακόμα και στην εξαγωγή καθολικών περιγραφέων για χρήση από αλγορίθμους αναζήτησης και ταξινόμησης.

Στην παρούσα διατριβή ασχολούμαστε με την ερευνητική περιοχή της ανίχνευσης τοπικών χαρακτηριστικών σε εικόνες, και της ανάκτησης βίντεο χρησιμοποιώντας περιγραφείς βαθιών ΣΝΔ. Στα πλαίσια της έρευνάς μας, αναπτύσσουμε τον ανιχνευτή τοπικών χαρακτηριστικών W α SH, ο οποίος εντοπίζει περιοχές της εικόνας που αντιστοιχούν σε ευδιάκριτα τμήματα αντικειμένων που είτε περικλείονται από ακμές της εικόνας, είτε τις αντιλαμβανόμαστε ως τέτοιες, παρά την έλλειψη πραγματικών ακμών (βλ. εικόνα 1.1), χρησιμοποιώντας γεωμετρικές κατασκευές. Ως αρχικό στάδιο του ανιχνευτή W α SH, εξετάζουμε τη δειγματοληψία εικόνων και προτείνουμε διαφορετικές προσεγγίσεις με βάση δυαδικές ακμές ή συνεχείς συναρτήσεις της φωτεινότητας. Προτείνουμε τα ανισοτροπικά σταθμισμένα α -σχήματα, μια γεωμετρική κατασκευή που αναπαριστά δομές σε σύνολα σημείων, σε διαφορετικές κλίμακες, με σκοπό τη χρησιμοποίησή τους στον προτεινόμενο αλγόριθμο ανίχνευσης τοπικών χαρακτηριστικών. Επίσης, αναπτύσσουμε έναν αλγόριθμο για αποδοτική ανάκτηση βίντεο με βάση το περιεχόμενο, εξάγοντας ολικούς περιγραφείς με βαθιά ΣΝΔ, και τον εφαρμόζουμε σε ένα σύνολο ιατρικών βίντεο.

1.1 Τοπικά χαρακτηριστικά

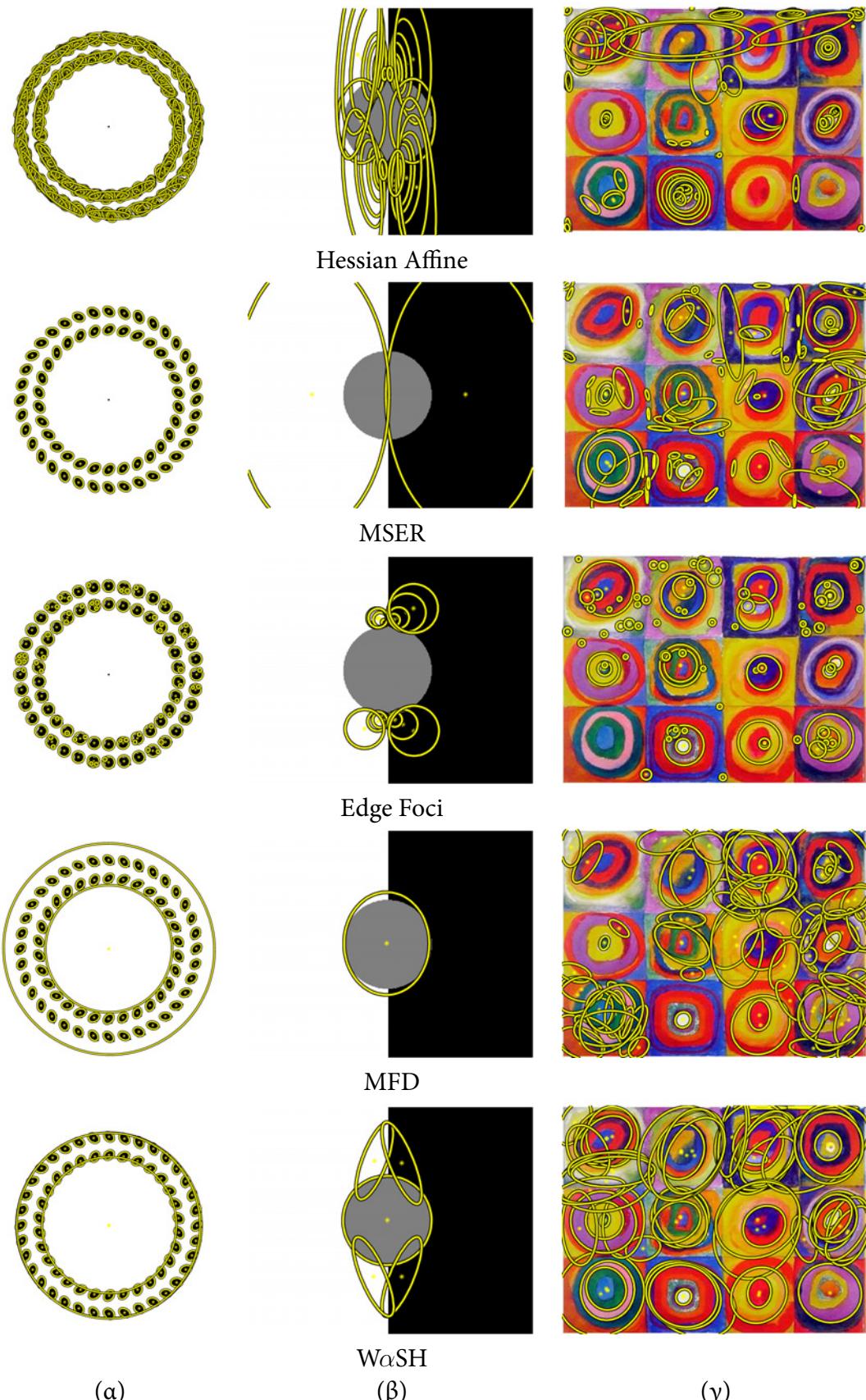
Τα τοπικά χαρακτηριστικά προσφέρουν μια ισορροπία ανάμεσα στις καθολικές αναπαραστάσεις (όπως οι περιγραφείς MPEG-7 και HOG [20]) και τα πυκνά χαρακτηριστικά που εξάγονται χρησιμοποιώντας ένα σταθερό πλέγμα με επικαλύψεις επί της εικόνας. Αγνοώντας τις μη-εξέχουσες περιοχές των εικόνων, και εστιάζοντας σε ευδιάκριτες περιοχές, τα τοπικά χαρακτηριστικά προσφέρουν επαναληψιμότητα, υψηλή διακριτική ικανότητα και υπολογιστική αποδοτικότητα, μειώνοντας τον απαραίτητο όγκο για την αναπαράσταση της οπτικής πληροφορίας. Οι ιδιότητές τους αυτές, βοηθούν στην αύξηση της απόδοσης πολλών εφαρμογών της όρασης υπολογιστών όπου χρησιμοποιούνται, όπως της αναζήτησης εικόνων μεγάλης κλίμακας, της ανίχνευσης αντικειμένων και της 3Δ ανακατασκευής εικόνων.

Από τους πολλούς ανιχνευτές τοπικών χαρακτηριστικών που έχουν προταθεί τα τελευταία χρόνια, οι Hessian-Affine [77], MSER [75] και SURF [8] χρησιμοποιούνται κατά κόρον στις περισσότερες εφαρμογές όρασης υπολογιστών. Οι συγκεκριμένες μέθοδοι έχουν δοκιμαστεί εξαντλητικά σε πολλές συγκριτικές δοκιμές και θεωρούνται αρκετά ώριμες ώστε να περιληφθούν σε δημόσια διαθέσιμες βιβλιοθήκες λογισμικού για όραση υπολογιστών, όπως η OpenCV³ και η VLFeat⁴, καθώς και σε εμπορικές εφαρμογές (συρραφής εικόνων, 3Δ αναπαραστάσεων κ.α.). Παρόλα αυτά, η ταχύτητα, η σταθερότητα των ανιχνευμένων χαρακτηριστικών, καθώς και η κατανομή τους στις εικόνες δεν είναι ιδανικές, όπως βλέπουμε και στην Εικόνα 1.1. Η ταχύτητα του ανιχνευτή Hessian-Affine είναι περιορισμένη, και τα τοπικά χαρακτηριστικά που ανιχνεύει συχνά δεν καλύπτουν μεγάλο μέρος της εικόνας, ενώ πολλαπλά χαρακτηριστικά ανιχνεύονται σε μια περιοχή σε γειτονικές θέσεις και πολλαπλές κλίμακες. Ο ανιχνευτής MSER είναι γρήγορος, αλλά συχνά ανιχνεύει αραιές περιοχές κανονικού σχήματος, των οποίων η διακριτικότητα είναι περιορισμένη, ενώ δεν ανιχνεύει περιοχές που δεν έχουν τοπικά ακραίες τιμές φωτεινότητας. Ο ανιχνευτής SURF εξάγει χαρακτηριστικά αρκετά γρήγορα, τα οποία όμως δεν είναι αρκετά σταθερά σε μετασχηματισμούς των εικόνων. Σε πρόσφατες δημοσιεύσεις, νέοι ανιχνευτές τοπικών χαρακτηριστικών συγκρίνονται με τους προαναφερθέντες ανιχνευτές της βιβλιογραφίας, χρησιμοποιώντας πέρα από τις συνήθεις μετρικές (π.χ. επαναληψιμότητα / βαθμός ταιριάσματος) και διάφορες εφαρμογές όρασης υπολογιστών, όπως ταξινόμηση εικόνων [108] και ανάκτηση [6, 126, 1].

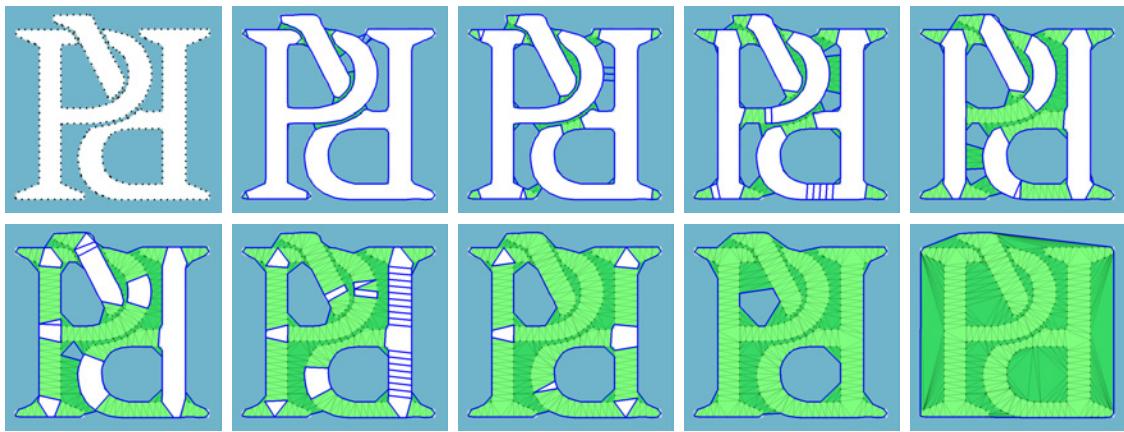
Στην παρούσα διατριβή περιγράφουμε τον νέο ανιχνευτή τοπικών χαρακτηριστικών *WaSH* (*Weighted α -SShapes detector*) που βασίζεται σε μια αναπαράσταση του τοπικού σχήματος, παρά σε πρώτες και δεύτερες παραγώγους της φωτεινότητας της εικόνας. Προς αυτή την κατεύθυνση χρησιμοποιούμε τα α -σχήματα, μια κατασκευή της υπολογιστικής

³<http://opencv.org/>

⁴<http://www.vlfeat.org/>



Εικόνα 1.1: Παραδείγματα ανίχνευσης τοπικών χαρακτηριστικών σε περιοχές που (α) δεν περικλείονται πλήρως από ακμές και (β,γ) δεν έχουν ακραίες τιμές φωτεινότητας.



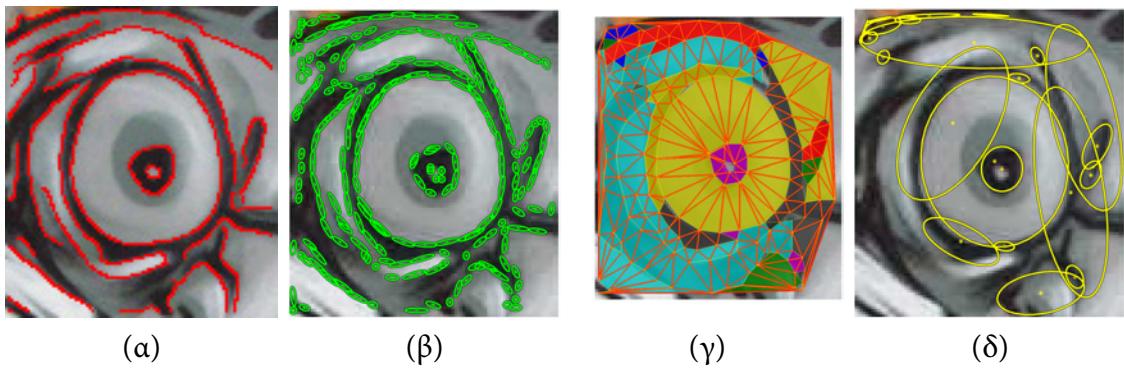
Εικόνα 1.2: Παράδειγμα της α -διήθησης. Από επάνω αριστερά προς τα δεξιά και κάτω, παρουσιάζονται διαφορετικά στιγμιότυπα της διήθησης για διαφορετικές τιμές του α . Ξεκινώντας από το σύνολο σημείων για $\alpha = 0$, προσθέτοντας ακμές και τρίγωνα καταλήγουμε στο κυρτό περίβλημα για $\alpha = \infty$. Παρατηρούμε ότι οι κοιλότητες των α -σχημάτων αντιστοιχούν σε κοιλότητες των αντικειμένων και ομοιόμορφες περιοχές.

γεωμετρίας που εισήγαγαν οι Edelsbrunner et al. [28]. Δεδομένου ενός συνόλου σημείων και μιας τριγωνοποίησης αυτών, τα α -σχήματα είναι υποσύνολα της τριγωνοποίησης τα οποία παραμετροποιούνται από τον βαθμωτό α . Στόχος της κατασκευής αυτής είναι η περιγραφή των δομών της εικόνας που μπορούν να σχηματιστούν από το δεδομένο σύνολο σημείων.

1.2 α -σχήματα

Τα α -σχήματα αποτελούν γενίκευση του κυρτού περιβλήματος ενός συνόλου σημείων, που παραμετροποιείται από τον βαθμωτό α , περιλαμβάνοντας διαφορετικά υποσύνολα της τριγωνοποίησης του συνόλου σημείων για κάθε α . Ξεκινώντας από το σύνολο των σημείων για $\alpha = 0$, καταλήγουν στο κυρτό περίβλημα του συνόλου στην ακραία περίπτωση όπου $\alpha = \infty$ (βλ. Εικόνα 1.2).

Το σύνολο όλων των α -σχημάτων (για όλες τις δυνατές τιμές του α) είναι μια διήθηση της τριγωνοποίησης του συνόλου σημείων, δηλ. μια μερική διάταξη των άπλοκων (σημεία, ακμές και τρίγωνα στις δύο διαστάσεις) [127]. Τα α -σχήματα βασίζονται στην τριγωνοποίηση Delaunay, η οποία χρησιμοποιεί μόνο τις συντεταγμένες των σημείων στο χώρο για την κατασκευή. Μια τέτοια αναπαράσταση μπορεί να είναι φτωχή για πολλές εφαρμογές της όρασης υπολογιστών, καθώς πολύ περισσότερη πληροφορία μπορεί να αντληθεί από το πεδίο της εικόνας. Τα σταθμισμένα α -σχήματα [27] βασίζονται σε μια κανονική τριγωνοποίηση και παρέχουν μια πιο ευέλικτη αναπαράσταση, καθώς συνδέουν με κάθε σημείο του συνόλου και έναν επιπλέον βαθμωτό αριθμό, πέρα από τις συντεταγμένες του. Επεκτείνοντας ακόμα περισσότερο τα α -σχήματα, οι Teichmann και Capps [107] εισήγα-



Εικόνα 1.3: Περίληψη του προτεινόμενου αλγορίθμου ανίχνευσης τοπικών χαρακτηριστικών WαSH. Από τις δυαδικές ακμές της εικόνας εισόδου (α), εξάγονμε δείγματα (β). Κατασκευάζουμε την α -διήθηση μιας τριγωνοποίησης των δειγμάτων (γ) και παρακολουθούμε την εξέλιξη των συνεκτικών συνιστωσών της. Τα χαρακτηριστικά της εικόνας εξάγονται με την επιλογή εύρωστων και εξεχουσών συνιστωσών (δ).

γαν τα ανισοτροπικά α -σχήματα, χρησιμοποιώντας μια ακόμα πιο πλούσια αναπαράσταση ανά σημείο, ενώ οι Labelle και Shewchuk [55] όρισαν ακόμα πιο αυστηρά την έννοια των ανισοτροπικών α -σχημάτων, ξεκινώντας από ένα καλά ορισμένο ανισοτροπικό διάγραμμα Voronoi.

Ο ανιχνευτής τοπικών χαρακτηριστικών WαSH που προτείνουμε στην παρούσα εργασία, βασίζεται στην αναπαράσταση των α -σχημάτων πάνω σε σημεία που προκύπτουν από δειγματοληψία εικόνων. Η δειγματοληψία γίνεται επί δυαδικών ακμών των εικόνων με ομοιόμορφο ή ανομοιόμορφο τρόπο, ή χρησιμοποιώντας τεχνικές διάχυσης σφάλματος. Για την κατασκευή των α -σχημάτων χρησιμοποιούμε τη θέση των δειγμάτων, το μέτρο της κλίσης της εικόνας, καθώς και το τοπικό σχήμα. Για να μπορέσουμε να συλλάβουμε το τοπικό σχήμα, χρησιμοποιούμε έναν αποδοτικό τρόπο για να ξεπεράσουμε τη βασική αδυναμία των α -σχημάτων, που είναι η επιλογή της κατάλληλης τιμής της παραμέτρου α που περιγράφει με τον καλύτερο δυνατό τρόπο το υποκείμενο σύνολο σημείων. Για το λόγο αυτό, χρησιμοποιούμε ένα δέντρο συνιστωσών με το οποίο μπορούμε να παρακολουθούμε την εξέλιξη των α -σχημάτων καθώς μεταβάλλεται η τιμή του α .

Τα βασικά βήματα του αλγορίθμου του ανιχνευτή WαSH φαίνονται στην Εικόνα 1.3 και είναι τα ακόλουθα:

- Εξαγωγή δειγμάτων από το πεδίο της εικόνας
- Τριγωνοποίηση των δειγμάτων και κατασκευή της α -διήθησης
- Επιλογή συνεκτικών συνιστωσών της α -διήθησης σε ένα δέντρο συνιστωσών, ως χαρακτηριστικά της εικόνας

Για κάθε βασικό βήμα του αλγορίθμου εξετάζουμε νέες μεθόδους, με στόχο τη βελτιστοποίηση της απόδοσης του αλγορίθμου σε εφαρμογές ταιριάσματος και ανάκτησης εικό-

νων, λαμβάνοντας υπ' όψιν και την υπολογιστική του πολυπλοκότητα.

Για την εξαγωγή δειγμάτων εξετάζουμε δύο διαφορετικές προσεγγίσεις. Η πρώτη χρησιμοποιεί ακμές τις εικόνας και εξαγωγή δειγμάτων επί των ακμών. Για την επιλογή των δειγμάτων κατά τη διάσχιση των ακμών χρησιμοποιείται είτε ένα σταθερό βήμα, είτε μια μέθοδος με εξαγωγή του τοπικού σχήματος για δειγματοληψία με μεταβλητή πυκνότητα. Η δεύτερη προσέγγιση εφαρμόζει διάχυση σφάλματος σε συνεχείς συναρτήσεις της φωτεινότητας για την εξαγωγή δειγμάτων.

Για την τριγωνοποίηση των δειγμάτων εξετάζουμε τη χρήση της τριγωνοποίησης Delaunay σαν βασική προσέγγιση, της Delaunay με περιορισμούς, και της κανονικής τριγωνοποίησης με προστιθέμενα βάρη. Σκοπός μας είναι η βελτίωση του εντοπισμού περιοχών με χαλαρά ορισμένα όρια. Για την κατασκευή της α -διήθησης χρησιμοποιούμε τα απλά και τα σταθμισμένα α -σχήματα, ενώ προτείνουμε τα ανισοτροπικά σταθμισμένα α -σχήματα που εκμεταλλεύονται το τοπικό σχήμα για τη διαμόρφωση της κατασκευής.

Τέλος, για την επιλογή χαρακτηριστικών από τα α -σχήματα και τις περιοχές που καλύπτουν, εξετάζουμε διαφορετικά κριτήρια σημαντικότητας, με στόχο την επιλογή χαρακτηριστικών με μεγάλη επαναληψιμότητα. Στο τελευταίο μέρος της διατριβής μεταβαίνουμε από την περιγραφή στατικών εικόνων στην περιγραφή βίντεο, και ασχολούμαστε με το πρόβλημα της ανάκτησης βίντεο.

1.3 Ανάκτηση ιατρικών βίντεο

Ενώ το πλήθος των διαθέσιμων βίντεο σε ιστοσελίδες διαχείρισης περιεχομένου και σε προσωπικές συλλογές χρηστών αυξάνεται με έντονο ρυθμό, στην πλειοψηφία τους παραμένουν ακατέργαστα, χωρίς συνοδευτικές πληροφορίες που να μπορούν να χρησιμοποιηθούν για την αποδοτική αναζήτηση και ανάκτησή τους. Μοναδικός τρόπος ανάκτησης σε μη επισημειωμένα από χρήστες βίντεο παραμένει η ανάκτηση με βάση το περιεχόμενο (οπτικό ή και ηχητικό σε ειδικές εφαρμογές). Τα βίντεο χειρουργικού περιεχομένου, δηλαδή βίντεο που συλλαμβάνονται κατά τη διάρκεια χειρουργικών επεμβάσεων, αποτελούν ιδιαίτερη πρόκληση για τις μεθόδους ανάκτησης με βάση το οπτικό περιεχόμενο, λόγω της φύσης του περιεχομένου (ιστοί με διαφορετική εμφάνιση, ποικιλία εσωτερικών οργάνων, απότομες κινήσεις της κάμερας, αλλαγές στη φωτεινότητα κ.α.).

Κατά τις δύο τελευταίες δεκαετίες, η ελάχιστα επεμβατική χειρουργική – EEX (*minimally invasive surgery* – MIS) έχει γίνει ιδιαίτερα δημοφιλής για μεγάλο αριθμό χειρουργικών επεμβάσεων. Σε σύγκριση με τις “ανοιχτές” χειρουργικές επεμβάσεις, οι EEX επιτυγχάνουν μικρότερο τραύμα, μείωση του χρόνου αποκατάστασης του τραύματος, λιγότερο πόνο, καθώς και μείωση του κινδύνου μολύνσεων. Ο χειρουργός χρησιμοποιεί μια ενδοσκοπική κάμερα και επιμήκη όργανα, η χρήση των οποίων απαιτεί ιδιαίτερες τεχνικές ικα-

Δομή της εργασίας

νότητες λόγω του περιορισμένου χώρου εργασίας, του φαινομένου του υπομοχλίου, καθώς και της 2Δ αναπαράστασης του χειρουργικού πεδίου [39]. Οι παραπάνω δυσκολίες, σε συνδυασμό με την εισαγωγή σύνθετων τεχνικών και οργάνων, καθιστούν την εκπαίδευση και ανασκόπηση της διαδικασίας μέσω των καταγεγραμμένων βίντεο ιδιαίτερα ωφέλιμη.

Η εκπαίδευση των ειδικών χειρουργών, περιλαμβάνει παρακολούθηση σχετικών βίντεο, τα οποία ανακτώνται από βάσεις δεδομένων χρησιμοποιώντας έναν περιορισμένο αριθμό λέξεων κλειδιά. Σε πολλές περιπτώσεις, όμως, ο εκπαιδευόμενος επιθυμεί να παρακολουθήσει συγκεκριμένα τμήματα της διαδικασίας και όχι ολόκληρο το βίντεο, γεγονός που οδηγεί στην αναζήτηση των σχετικών σκηνών από τον χρήστη. Τέτοια τμήματα βίντεο μπορεί να σχετίζονται με συγκεκριμένες διεργασίες (π.χ. εισαγωγή ενός εργαλείου), όργανα (π.χ. ψαλίδι), ή αποτελέσματα της διαδικασίας (π.χ. αιμορραγία, καυτηριασμός ιστού). Η επισημείωση σκηνών σε βίντεο με τέτοια χαρακτηριστικά είναι διαδικασία που γίνεται συνήθως από ανθρώπους, ενώ είναι χρονοβόρα και ανιαρή. Επίσης, στη διαδικασία υπεισέρχεται και η υποκειμενικότητα των επισημειωτών, καθώς δεν ενδιαφέρονται όλοι οι χρήστες για τις ίδιες διεργασίες. Κατά συνέπεια, η αυτόματη δεικτοδότηση και ανάκτηση σκηνών σε ιατρικά βίντεο χρησιμοποιώντας ένα σύνολο από τυποποιημένες επισημάνσεις είναι εξαιρετικά χρήσιμη.

1.4 Δομή της εργασίας

Το κείμενο της διατριβής οργανώνεται ως εξής:

Στο κεφάλαιο 2 παραθέτουμε τη σχετική με την εργασία βιβλιογραφία. Πρώτα παρουσιάζουμε τις εργασίες από το χώρο της υπολογιστικής γεωμετρίας που εξετάζουν τα α-σχήματα, και μετά τους πιο σημαντικούς ανίχνευτές τοπικών χαρακτηριστικών. Στην ενότητα 2.3 παρουσιάζουμε την εξέλιξη των βαθιών νευρωνικών δικτύων και τις εφαρμογές τους σε πληθώρα προβλημάτων ανάλυσης εικόνων και βίντεο. Τέλος, παρουσιάζουμε μεθόδους της βιβλιογραφίας για ανάκτηση βίντεο, και εύρεση υποπαραθύρων για ανίχνευση αντικειμένων. Οι μέθοδοι ανίχνευσης αντικειμένων συνδέονται με την ανάκτηση βίντεο στη μέθοδο που προτείνουμε.

Στο κεφάλαιο 3 περιγράφουμε τον ανίχνευτή χαρακτηριστικών WASH. Παρουσιάζουμε τις διαφορετικές μεθόδους δειγματοληψίας επί ακμών που προτείνουμε, και περιγράφουμε τις διαφορετικές τριγωνοποιήσεις που χρησιμοποιούμε ως βάση για την κατασκευή των α-σχημάτων, τα οποία ορίζουμε και περιγράφουμε στην ενότητα 3.3. Στην ίδια ενότητα παρουσιάζουμε και την προτεινόμενη μέθοδο των ανισοτροπικά σταθμισμένων α-σχημάτων. Έπειτα, περιγράφουμε τον εντοπισμό και την παρακολούθηση των συνεκτικών συνιστώσων της α-διήθησης μαζί με το σύστημα γειτνίασης μεταξύ άπλοκων που χρησιμοποιούμε. Στην ενότητα 3.4.2 περιγράφουμε τη διαδικασία επιλογής των τοπικών χαρακτη-

ριστικών από τις συνεκτικές συνιστώσες που παρακολουθούμε με τα δέντρα συνιστωσών, και παρουσιάζουμε τις διαφορετικές μετρικές που χρησιμοποιούμε. Παρουσιάζουμε τον αλγόριθμο ανίχνευσης χαρακτηριστικών που προτείνουμε στο σύνολό του, παραθέτουμε παραδείγματα ανιχνευμένων περιοχών και σχολιάζουμε τις ιδιότητές τους, συγκρίνοντάς τα με τα αποτελέσματα άλλων ανιχνευτών. Στην ενότητα 3.5 παρουσιάζουμε τα πειραματικά αποτελέσματα της αξιολόγησης της προτεινόμενης μεθόδου.

Στο κεφάλαιο 4 προτείνουμε τη δειγματοληψία εικόνων με χρήση διάχυσης σφάλματος σε συναρτήσεις της φωτεινότητας. Αρχικά περιγράφουμε τη διαδικασία του dithering στην ενότητα 4.1.1, και την προσαρμόζουμε για δειγματοληψία σε διαφορετικές συναρτήσεις της φωτεινότητας στην ενότητα 4.2. Συγκρίνουμε ποιοτικά και ποσοτικά τις δειγματοληψίες μεταξύ τους, χρησιμοποιώντας τες στον αλγόριθμο ανίχνευσης χαρακτηριστικών WASH. Τα αποτελέσματα της πειραματικής αποτίμησης παρουσιάζονται στην ενότητα 4.2.3.

Στο κεφάλαιο 5 προτείνουμε μια νέα μέθοδο για ανάκτηση βίντεο με χρήση περιγραφέων νευρωνικών δικτύων, την οποία εφαρμόζουμε σε ένα πείραμα ανάκτησης χειρουργικών βίντεο. Περιγράφουμε το πρόβλημα της ανάκτησης σε χειρουργικά βίντεο, καθώς και τη διαδικασία που ακολουθήσαμε για την δημιουργία του συνόλου δεδομένων. Στην ενότητα 5.3 περιγράφουμε τη διαδικασία κατάτμησης των βίντεο σε πλάνα, ενώ στην ενότητα 5.4 τη διαδικασία εξαγωγής περιγραφέων χρησιμοποιώντας τοπικά χαρακτηριστικά και νευρωνικά δίκτυα. Η απόδοση της μεθόδου αποτιμάται σε ένα πείραμα ανάκτησης πλάνων σε χειρουργικά βίντεο.

Στο κεφάλαιο 6 παρουσιάζουμε τα συμπεράσματα που εξάγουμε και κάνουμε μια σύνοψη της συνεισφοράς της διατριβής. Τέλος, προτείνουμε μελλοντικές κατευθύνσεις έρευνας για τις ερευνητικές περιοχές που εξετάσαμε.

Κεφάλαιο 2

Σχετικές Εργασίες

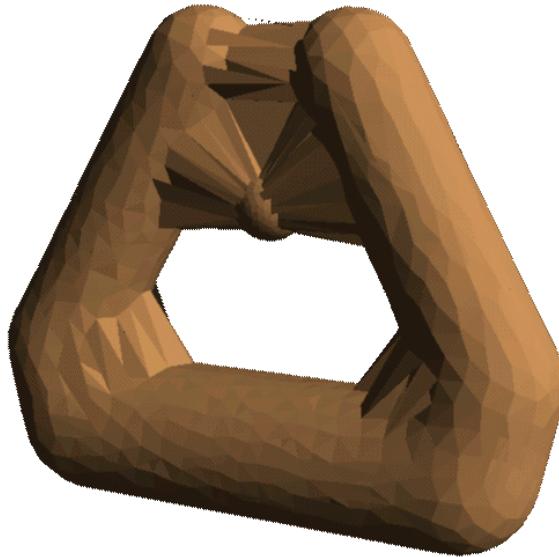
Σε αυτό το κεφάλαιο παρουσιάζουμε τη σχετική με το ερευνητικό μας πεδίο βιβλιογραφία. Αρχικά παρουσιάζουμε εργασίες από το χώρο της ανίχνευσης τοπικών χαρακτηριστικών, με περιγραφή των πιο συχνά χρησιμοποιούμενων ανιχνευτών καθώς και των διαθέσιμων προτύπων συγκριτικής αξιολόγησής τους. Στη συνέχεια παρουσιάζουμε εργασίες από την επιστημονική περιοχή της υπολογιστικής γεωμετρίας, που σχετίζονται με τα α -σχήματα και τις τριγωνοποιήσεις που εξετάζουμε. Τέλος, παρουσιάζουμε εργασίες που προτείνουν τη χρήση ΣΝΔ για ταξινόμηση και ανάκτηση εικόνων και βίντεο.

2.1 Υπολογιστική γεωμετρία – α -σχήματα

Σε αντίθεση με την όραση υπολογιστών, στην υπολογιστική γεωμετρία τα α -σχήματα χρησιμοποιούνται σε πληθώρα εφαρμογών. Μία από τις πρώτες εφαρμογές είναι η ανακατασκευή επιφάνειας στις τρεις διαστάσεις από ένα μη οργανωμένο σύνολο σημείων [28, 27]. Τα α -σχήματα ως γενίκευση του κυρτού περιβλήματος μπορούν να κατασκευάσουν διαφορετικές επιφάνειες για διαφορετικές τιμές του α , προσφέροντας μια ποικιλία ανακατασκευών διαφορετικής κλίμακας για ένα δεδομένο σύνολο σημείων.

Τα α -σχήματα έχουν χρησιμοποιηθεί επίσης στην μελέτη κοιλοτήτων (*pockets*). Ως κοιλότητα ορίζεται μια περιοχή του σχήματος με περιορισμένη πρόσβαση από το εξωτερικό του. Στις μελέτες αυτές μετράται η επιφάνεια και ο όγκος μακρομορίων, καθώς και η πρόσδεση πρωτεϊνών [27]. Οι Meine et al. [76] χρησιμοποιούν τα α -σχήματα για την ανακατασκευή ορίων από θορυβώδεις χάρτες ακμών. Η μέθοδος εξάγει σημεία με δειγματοληψία σε δυαδικές ακμές και κατασκευάζει την τριγωνοποίηση Delaunay των σημείων. Με οδηγό τις ακμές της τριγωνοποίησης συμπληρώνει τα όρια των αντικειμένων της εικόνας, όπου οι ακμές λόγω θορύβου δεν είναι συνεχείς. Για την επιλογή των ακμών που θα αξιοποιηθούν, εξετάζεται το μέγεθος των τριγώνων καθώς και το μέσο χρώμα.

Οι Teichmann και Capps [107] εισήγαγαν τα ανισοτροπικά α -σχήματα, σε μια προ-



Εικόνα 2.1: Περιορισμοί των α-σχημάτων. Οι επιφάνειες των αντικειμένων δεν αναπαριστώνται σωστά σε σημεία ρωγμών, αρθρώσεων και όταν τα αντικείμενα βρίσκονται κοντά μεταξύ τους (από [107]).

σπάθεια να ξεπεραστούν περιορισμοί των α-σχημάτων που οδηγούν σε λανθασμένη αναπαράσταση του πραγματικού σχήματος του συνόλου σημείων. Τα κύρια προβλήματα που αντιμετωπίζουν φαίνονται στο παράδειγμα της Εικόνας 2.1 και είναι:

ρωγμές: λανθασμένη ένωση σε ρωγμές του πραγματικού σχήματος. Για α αρκετά μεγάλο ώστε να σχηματίζεται η επιφάνεια του αντικειμένου, εμφανίζονται τμήματα της τριγωνοποίησης μέσα σε πραγματικές ασυνέχειες (ρωγμές) του σχήματος, όπως στο πάνω μέρος του σχήματος όπου ενώνονται τα δύο άκρα του πετάλου.

γειτνίαση: λανθασμένη ένωση ξένων αντικειμένων. Αντίστοιχα με την περίπτωση των ρωγμών, εμφανίζονται τμήματα της τριγωνοποίησης που οδηγούν σε ένωση ξένων μεταξύ τους αντικειμένων, όπως η ένωση της σφαίρας με το πέταλο.

αρθρώσεις: απώλεια λεπτομέρειας του σχήματος σε απότομες στροφές ή ενώσεις (αρθρώσεις). Στο κάτω μέρος του σχήματος, βλέπουμε ότι στις απότομες γωνίες του πετάλου εμφανίζονται τρίγωνα που αλλοιώνουν τη λεπτομέρεια του αντικειμένου.

Για να ξεπεράσουν του περιορισμούς αυτούς, οι συγγραφείς προτείνουν την εισαγωγή επιπλέον πληροφορίας σε κάθε σημείο της τριγωνοποίησης που προκύπτει από την κατεύθυνση της κλίσης στη δισδιάστατη περίπτωση και του κανονικοποιημένου κάθετου διανύσματος στην τρισδιάστατη περίπτωση. Για κάθε τρίγωνο (και ακμή) της τριγωνοποίησης υπολογίζεται ο μέσος όρος της κατεύθυνσης των τριών (δύο) σημείων που το ορίζουν. Η κατεύθυνση αυτή αντιστοιχεί σε έναν πίνακα τανυστή, ο οποίος χρησιμοποιείται για να μετακινήσει τα σημεία του τριγώνου (ακμής) ώστε να υπολογιστούν εκ νέου

οι κρίσιμες τιμές της παραμέτρου α για το τρίγωνο (ακμή) και να μεταβάλει τοπικά την αρχική τριγωνοποίηση Delaunay με τις απαραίτητες αναστροφές (*flips*).

Οι Labelle και Shewchuk [55] προτείνουν το θεωρητικό υπόβαθρο ώστε να κατασκευάσουν μια καλά ορισμένη ανισοτροπική τριγωνοποίηση. Ξεκινούν με την κατασκευή του ανισοτροπικού διαγράμματος Voronoi, όπου εφαρμόζουν τις απαραίτητες παραδοχές ώστε το διάγραμμα να είναι καλά ορισμένο, χωρίς ορφανές περιοχές (που δεν περιλαμβάνουν την πηγή τους). Έπειτα, εκμεταλλευόμενοι τη δυικότητα μεταξύ των διαγραμμάτων Voronoi και των τριγωνοποιήσεων, κατασκευάζουν μια ανισοτροπική τριγωνοποίηση, θέτοντας τη βάση για τα αντίστοιχα ανισοτροπικά α -σχήματα.

Οι Emrírges et al. [31] προτείνουν έναν αυξητικό αλγόριθμο για την κατασκευή Voronoi διαγραμμάτων όπου οι εστίες είναι ελλείψεις και όχι σημεία. Με αυτό τον τρόπο αξιοποιούν το σχήμα των εστιών για την κατασκευή μιας διαμέρισης του χώρου που ταιριάζει καλύτερα στα δεδομένα. Παρόλα αυτά, ο χώρος έξω από τις ελλείψεις παραμένει Ευκλεϊδειος, με αποτέλεσμα ελλείψεις μικρής κλίμακας να μην επηρεάζουν σημαντικά το τελικό αποτέλεσμα.

Σχετική είναι και η εργασία των Cazals et al. [14], που παρουσιάζουν τα σύμμορφα α -σχήματα (*conformal α -shapes*) και την αντίστοιχη διήθηση. Τα σύμμορφα α -σχήματα βασίζονται σε μία παράμετρο κλίμακας $\hat{\alpha}$ που αφορά όλο το σχήμα καθώς και δύο τοπικές παραμέτρους που προσαρμόζονται από τις ιδιότητες μιας γειτονιάς κάθε σημείου. Η μέθοδος χρησιμοποιείται για ανακατασκευή επιφανειών όπου η δειγματοληψία σημείων είναι ανομοιόμορφη. Στην ίδια κατεύθυνση, οι Zomorodian et al. [127] προβλέπουν τη δομή πρωτεΐνων χρησιμοποιώντας τα α -σχήματα για την ανίχνευση ατόμων που αλληλεπιδρούν σε ένα μόριο πρωτεΐνης.

2.2 Τοπικά χαρακτηριστικά

Από τους πρώτους ανιχνευτές χαρακτηριστικών της βιβλιογραφίας είναι οι ανιχνευτές γωνιών, με χαρακτηριστικά παραδείγματα αυτούς του Beaudet [9] που βασίζεται στον πίνακα δεύτερων παραγώγων (Hessian) και των Harris και Stevens [44] που χρησιμοποιεί τον πίνακα ροπών 2^{nd} τάξης. Αυτοί οι πρώτοι ανιχνευτές γωνιών οδήγησαν στην δημιουργία ανιχνευτών περιοχών ενδιαφέροντος, που δεν αντιστοιχούν μόνο σε γωνίες. Μεγάλη επιρροή στον χώρο των ανιχνευτών χαρακτηριστικών είχε η εργασία των Lindeberg et al. [66], οι οποίοι μελέτησαν τους χώρους κλίμακας προτείνοντας χαρακτηριστικά που εμφανίζουν ανεξαρτησία σε μετασχηματισμούς κλίμακας. Ως επέκταση της ανεξαρτησίας κλίμακας δημιουργήθηκαν οι θεωρητικές βάσεις για τη δημιουργία χαρακτηριστικών με ανεξαρτησία σε αφινικούς μετασχηματισμούς [67].

Βασιζόμενος σε αυτές τις ιδέες, ο Lowe [70] πρότεινε τον ανιχνευτή χαρακτηριστικών

SIFT (*Scale Invariant Feature Transform*), ο οποίος είναι ένας από τους πιο διαδεδομένους και χρησιμοποιημένους ανιχνευτές. Βασίζεται σε διαφορές εικόνων φιλτραρισμένων με Gaussian φίλτρα διαφορετικών τυπικών αποκλίσεων και εντοπίζει περιοχές ανεξάρτητες σε αλλαγές κλίμακας, καθώς και περιστροφές. Ο περιγραφέας που προτάθηκε στην ίδια εργασία χρησιμοποιείται στην πλειονότητα των εφαρμογών όρασης υπολογιστών. Αργότερα, οι Bay et al. [8] πρότειναν τον ανιχνευτή SURF (*Speeded Up Robust Features*), μία προσέγγιση του SIFT, που χρησιμοποιεί ολοκληρωτικές εικόνες (integral images) για να επιταχύνει τη διαδικασία εντοπισμού των χαρακτηριστικών, καθώς και του υπολογισμού των περιγραφέων. Ο SURF χρησιμοποιείται συχνά σε εφαρμογές όπου ο χρόνος που απαιτείται για την εξαγωγή των χαρακτηριστικών είναι κρίσιμος.

Βασιζόμενοι στον ανιχνευτή Harris, οι Mikolajczyk et al. [77] πρότειναν τους ανιχνευτές *Harris-Affine* και *Hessian-Affine*, που εμφανίζουν ανεξαρτησία σε αφινικούς μετασχηματισμούς, με χρήση ενός βήματος προσαρμογής στο αφινικό σχήμα της εντοπισμένης περιοχής. Οι Matas et al. [75] πρότειναν τον ανιχνευτή περιοχών MSER (*Maximally Stable Extremal Regions*), ο οποίος ανιχνεύει περιοχές που έχουν σταθερή φωτεινότητα, αποφεύγοντας τα συνήθη προβλήματα των ανιχνευτών που βασίζονται σε παραγώγους της φωτεινότητας, όπως ο θόρυβος και ο ακριβής εντοπισμός στο χώρο της εικόνας. Η ιδέα του αλγορίθμου είναι η δημιουργία κατατμήσεων με τρόπο αντίστοιχο του πλημμυρισμού (watershed), και η επιλογή περιοχών που διατηρούν σταθερό το εμβαδό τους για μεγάλο αριθμό κατωφλίων. Ο MSER χρησιμοποιείται ευρέως στο χώρο της όρασης υπολογιστών σε πολλές εφαρμογές, λόγω της υψηλής του απόδοσης (ειδικά σε εικόνες με επίπεδες δομές) και της ταχύτητάς του. Καθώς ο MSER ανιχνεύει περιοχές της εικόνας αυθαίρετου σχήματος που περικλείονται από καλά κλεισμένα όρια, και όχι απλά σημεία με συγκεκριμένη κλίμακα και περιστροφή, το αποτέλεσμα της ανίχνευσης μπορεί να χρησιμοποιηθεί ώστε να προκύψουν διαφορετικοί περιγραφείς (περιγραφείς σχήματος, φωτεινότητας, κατανομής ακμών κ.α.).

Πρόσφατα, οι Alcantarilla et al. [1] πρότειναν τον ανιχνευτή KAZE, ο οποίος βασίζεται στον ανιχνευτή Hessian-Affine, αλλά ανιχνεύει τα τοπικά μέγιστα του πίνακα Hessian σε ένα μη γραμμικό χώρο κλίμακας που κατασκευάζεται από φιλτράρισμα διάχυσης. Η επίδοση του KAZE στο πρότυπο συγκριτικής αξιολόγησης των Mikolajczyk et al. [78] είναι υψηλή, παρόλα αυτά η κατασκευή του μη γραμμικού χώρου κλίμακας έχει υψηλό υπολογιστικό κόστος. Επίσης, ο αριθμός των χαρακτηριστικών που εξάγονται από τον ανιχνευτή KAZE είναι αρκετά υψηλός, όπως και αυτός του Hessian-Affine, γεγονός που αποτρέπει από τη χρήση του σε εφαρμογές μεγάλης κλίμακας. Οι ίδιοι συγγραφείς βελτίωσαν την υπολογιστική πολυπλοκότητα του αλγορίθμου [2], διατηρώντας σταθερή την απόδοση του ανιχνευτή.

Τα τελευταία χρόνια έχει δοθεί μεγάλο βάρος στο υπολογιστικό κόστος των ανιχνευ-

τών χαρακτηριστικών, με στόχο τη χρησιμοποίησή τους σε εφαρμογές πραγματικού χρόνου (παρακολούθηση αντικειμένων, ανίχνευση αντικειμένων, κ.α.). Σε αυτή την κατεύθυνση έχει προταθεί ο ανιχνευτής FAST [98, 99], καθώς και παραλαγές του (FAST-9 [73], FAST-ER [100]). Η βασική ιδέα του αλγορίθμου είναι η χρήση συγκρίσεων σε τιμές φωτεινότητας μικρών περιοχών, ώστε να εντοπιστούν γωνίες. Το αποτέλεσμα είναι η γρήγορη ανίχνευση πολλών σημείων ενδιαφέροντος, στα οποία μπορεί να περιλαμβάνονται και θορυβώδη σημεία τα οποία καλείται να διαχωρίσει επόμενο βήμα της εφαρμογής. Οι ανιχνευτές BRISK [62] και ORB [101] βασίζονται στον FAST και παρέχουν ανίχνευση και περιγραφή των τοπικών χαρακτηριστικών σε πραγματικό χρόνο. Η απόδοσή τους στο ταίριασμα εικόνων είναι συγκρίσιμη με τους SIFT και SURF, ενώ ο χρόνος εντοπισμού είναι σημαντικά μικρότερος. Οι προαναφερθέντες αλγόριθμοι εξάγουν χαρακτηριστικά που είναι ανεξάρτητα κλίμακας και περιστροφής, αλλά όχι αφινικά ανεξάρτητα, γεγονός που περιορίζει τις διαθέσιμες εφαρμογές και την αναμενόμενη απόδοση.

Τοπικά χαρακτηριστικά από ακμές

Η χρήση των ακμών σε εικόνες για την εξαγωγή χαρακτηριστικών δεν είναι ιδιαίτερα διαδεδομένη, παρά το γεγονός ότι οι ακμές ορίζουν τα φυσικά όρια αντικειμένων και περιοχών. Το γεγονός αυτό οφείλεται κυρίως στο υπολογιστικό κόστος της ανίχνευσης ακμών που θα πρέπει να προστεθεί σε αυτό του τελικού αλγορίθμου εξαγωγής χαρακτηριστικών, καθώς και στην ευαισθησία των αλγορίθμων ανίχνευσης ακμών, όπου τα κατώφλια που χρησιμοποιούνται συχνά πρέπει να προσαρμόζονται για κάθε εικόνα.

Μία από τις πρώτες προσπάθειες να χρησιμοποιηθούν ακμές για την εξαγωγή χαρακτηριστικών έγινε από τους Tuytelaars και Van Gool [110] με τον ανιχνευτή EBR (*Edge-Based Region detector*). Ο EBR ξεκινά από γωνιακά σημεία και ακολουθώντας τις προσπίπτουσες στο σημείο ακμές υπολογίζει φωτομετρικές ποσότητες για να επιλέξει εξέχουσες περιοχές. Είναι κατάλληλος για εικόνες που περιέχουν ακμές που τέμνονται (π.χ. ανθρώπινες κατασκευές όπως κτήρια), αλλά όχι για ταίριασμα γενικού σκοπού, όπως έδειξαν οι Mikolajczyk et al. [78] στην πειραματική τους αξιολόγηση.

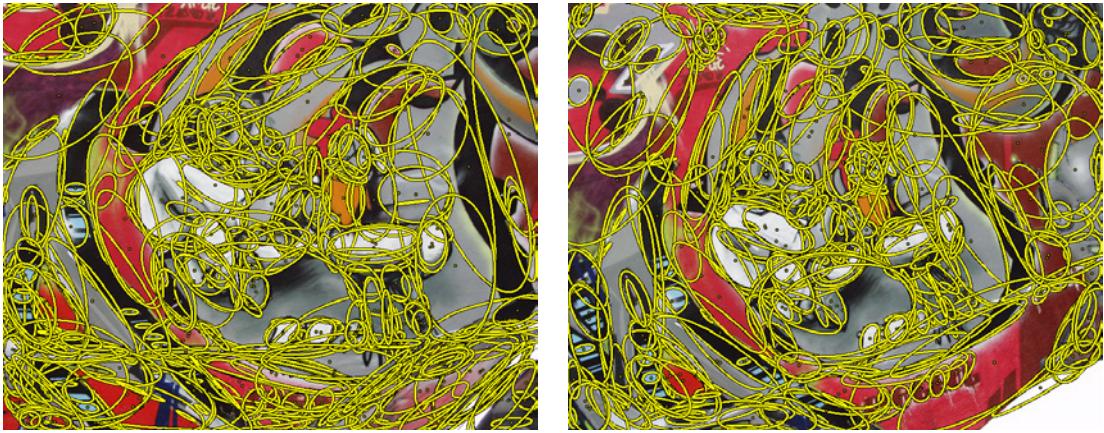
Οι Mikolajczyk et al. [79] πρότειναν έναν ανιχνευτή που βασίζεται σε ακμές, χρησιμοποιώντας τον αλγόριθμο του Canny [12] για την ανίχνευση ακμών σε συνδυασμό με αυτόματη επιλογή κλίμακας. Για να μειωθεί η υπολογιστική πολυπλοκότητα του αλγορίθμου χρησιμοποιούνται πυκνά δείγματα επί των ανιχνευμένων ακμών. Ο ανιχνευτής χρησιμοποιήθηκε σε εφαρμογή για αναγνώριση αντικειμένων, δίνοντας καλά αποτελέσματα. Οι Rapanantzikos et al. [94], ξεκινώντας πάλι από ακμές του αλγορίθμου Canny, υπολογίζουν τον δυαδικό μετασχηματισμό απόστασης. Εντοπίζουν περιοχές με μία διαδικασία ομαδοποίησης που ξεκινά από τοπικά μέγιστα του μετασχηματισμού απόστασης και αξιοποιεί το μέτρο της κλίσης των γειτονικών ακμών.

Εκτός των προαναφερθέντων μεθόδων που χρησιμοποιούν άμεσα δυαδικές ακμές της εικόνας, στη βιβλιογραφία υπάρχουν και άλλες πρόσφατες μέθοδοι ανίχνευσης χαρακτηριστικών που συνδέονται έμμεσα με τις ακμές, καθώς χρησιμοποιούν το μέτρο της κλίσης χωρίς κατωφλίωση. Οι Zitnick et al. [126] εφαρμόζουν μια τράπεζα κατευθυνόμενων φίλτρων στην εικόνα εισόδου και ανιχνεύουν τις εστίες ακμών *EF* (*Edge Foci*). Οι εστίες ακμών είναι σημεία που ισαπέχουν από πολλά μικρά τμήματα ακμών. Η διεύθυνση των μικρών τμημάτων ακμών είναι κάθετη προς την διεύθυνση που ορίζει το διάνυσμα από την ακμή προς την εστία. Οι εστίες ακμών αντιστοιχούν σε εσωτερικά σημεία γωνιών και κυκλικές δομές, αξιοποιώντας το σχήμα της τοπικής γειτονιάς. Ο υπολογισμός των χαρακτηριστικών *EF* είναι αρκετά χρονοβόρος.

Οι Αβρίθης και Ραπαντζίκος [6] υπολογίζουν τον σταθμισμένο μετασχηματισμό διάμεσον άξονα (*weighted medial axis transform*) από την κλίση της εικόνας και δημιουργούν μία διαμέριση της εικόνας με ένα τρόπο παρόμοιο με τον μετασχηματισμό τοπολογικού πλημμυρισμού. Χρησιμοποιώντας μια ιεραρχική δομή της διαμέρισης, ο ανιχνευτής *MFD* (*Medial Features Detector*) εντοπίζει χαρακτηριστικά λαμβάνοντας υπόψιν την αντίθεση και το τοπικό σχήμα. Παρά το γεγονός ότι οι προαναφερθείσες μέθοδοι χρησιμοποιούν την κλίση της εικόνας που είναι μια αναπαράσταση πιο πλούσια από τις δυαδικές ακμές, το μέτρο της κλίσης είναι ευαίσθητο σε διακυμάνσεις του φωτισμού και της κλίμακας.

Τα χαρακτηριστικά που εξάγουν οι ανιχνευτές που εξετάσαμε, περιγράφονται με διάφορες αναπαραστάσεις. Υπάρχουν ανιχνευτές που εξάγουν χαρακτηριστικά ανεξάρτητα ως προς μετασχηματισμούς ομοιότητας (όπως οι SIFT και SURF). Τα χαρακτηριστικά αυτά περιγράφονται από τέσσερις παραμέτρους (δύο για τη θέση στο επίπεδο, ένας για την κλίμακα και ένας για την περιστροφή). Άλλοι ανιχνευτές εξάγουν χαρακτηριστικά που είναι αφινικά ανεξάρτητα (όπως οι Hessian-Affine) και περιγράφονται από πέντε παραμέτρους (δύο για τη θέση στο επίπεδο, και τρεις που περιγράφουν τις δύο κλίμακες και την περιστροφή της έλλειψης που αντιστοιχεί στην αφινική περιοχή). Τέλος, υπάρχουν ανιχνευτές που εξάγουν το αυθαίρετο σχήμα της περιοχής ενδιαφέροντος (όπως ο MSER και κάποιοι που βασίζονται σε ακμές της εικόνας, π.χ. [94, 6]). Αυτά τα αυθαίρετα σχήματα μπορούν να περιγραφούν πιο συμπυκνωμένα με διάφορους τρόπος [82], εξάγοντας διαφορετικές γεωμετρικές ιδιότητές τους. Παρόλα αυτά, στην πλειοψηφία των περιπτώσεων χρησιμοποιούνται ελλειψεις που προσαρμόζονται στις περιοχές, όπως συμβαίνει και με τους ανιχνευτές που εξάγουν αφινικές περιοχές.

Ο νέος αλγόριθμος ανίχνευσης τοπικών χαρακτηριστικών *W_αSH* (*Weighted α-SHapes detector*) που προτείνουμε, ξεκινά από την εξαγωγή δειγμάτων πάνω στα όρια των αντικειμένων της εικόνας. Προτείνουμε δύο οικογένειες δειγματοληψίας, βασιζόμενες είτε σε ακμές όπως οι [79, 110], είτε σε διάχυση σφάλματος. Έπειτα κατασκευάζουμε μια 2Δ τριγωνοποίηση των σημείων και την α -διήθηση, όπως οι [28, 29, 107, 55], δοκιμάζοντας δια-



Εικόνα 2.2: Τοπικά χαρακτηριστικά που εξάγει ο προτεινόμενος αλγόριθμος *WaSH*, με χρήση ανιστροπικά σταθμισμένων α-σχημάτων και ανομοιόμορφη δειγματοληψία επί ακμών.

φορετικές τριγωνοποιήσεις και προτείνοντας μια νέα ανιστροπική α -διήθηση. Χρησιμοποιούμε δέντρα συνιστωσών, μια ιεραρχική αναπαράσταση φωλιασμένων συνόλων, για να παρακολουθήσουμε την εξέλιξη της κατασκευής, όπως οι [75] για την εξέλιξη των επιπεδοσυνόλων (level sets). Για την επιλογή των χαρακτηριστικών προτείνουμε διάφορα κριτήρια που σχετίζονται κυρίως με το σχήμα των συνιστωσών, όπως οι [75, 6].

Η μέθοδος μας ανιχνεύει περιοχές της εικόνας που μπορούν να συνορεύουν ταυτόχρονα με πιο σκούρες και πιο φωτεινές περιοχές, σε αντίθεση με τον MSER που ανιχνεύει μόνο περιοχές που έχουν ακραία τιμή φωτεινότητας, καθώς και περιοχές που δεν είναι πλήρως κλειστές (βλ. Εικόνα 1.1). Επίσης, ανιχνεύονται περιοχές που ορίζονται από εσοχές στο όριο αντικειμένων, όμοια με τις κοιλότητες [27] (βλ. Εικόνα 2.2).

2.3 Βαθιά νευρωνικά δίκτυα

Πρόσφατα επανεμφανίστηκαν στο προσκήνιο της ανάλυσης εικόνων και όρασης υπολογιστών τα *Συνελικτικά Νευρωνικά Δίκτυα* (ΣΝΔ), ερευνητικό πεδίο της μηχανικής μάθησης με μακρά ιστορία. Ένα από τα πρώτα συστήματα όρασης υπολογιστών ήταν το “Mark I Perceptron”, που προτάθηκε το 1957 από τον Rosenblatt [96], και έπαιξε σπουδαίο ρόλο στην εξέλιξη των τεχνιτών νευρωνικών δικτύων. Βασισμένο στις ανακαλύψεις για το βιολογικό μοντέλο όρασης των Hubel και Wiesl [46], το “Neocognitron” του Fukushima [38] είναι το πρώτο μοντέλο αναγνώρισης εικόνων που έχει ιεραρχική δομή και είναι ανεξάρτητο σε μετατοπίσεις. Οι LeCun et al. [59] έδειξαν την αποτελεσματικότητα της στοχαστικής επικλινούς καθόδου (stochastic gradient descent) με την οπισθοδιάδοση σφάλματος (backpropagation) [102] για την εκπαίδευση συνελικτικών νευρωνικών δικτύων. Τα ΣΝΔ βρήκαν αργότερα εφαρμογή στην οπτική αναγνώριση χαρακτήρων με μεγάλη επιτυχία [60].

Πρόσφατα, οι Krizhevsky et al. [54] εκπαίδευσαν το πρώτο πραγματικά βαθύ συνελικτικό νευρωνικό δίκτυο (*AlexNet*), αποτελούμενο από 7 κρυμμένα επίπεδα, 650,000 νευρώνες και 60 εκατομμύρια παραμέτρους. Η εκπαίδευση του δικτύου έγινε με μεγάλη παραλληλία σε κάρτες γραφικών, χρησιμοποιώντας το σύνολο εικόνων Imagenet [23] και η επίδοσή του στην ταξινόμηση εικόνων στο ίδιο σύνολο εικόνων ξεπέρασε κατά πολύ τις καλύτερες ανταγωνιστικές μεθόδους. Επίσης, οι Le et al. [57] εκπαίδευσαν ένα βαθύ ΣΝΔ πολύ μεγάλης κλίμακας για την εκμάθηση και εξαγωγή χαρακτηριστικών υψηλού επιπέδου από μη επισημασμένες εικόνες, με σκοπό την κατηγοριοποίηση εικόνων. Για την εκπαίδευση χρησιμοποίησαν 10 εκατομμύρια εικόνες από καρέ βίντεο, ενώ η εκμάθηση έγινε χωρίς τη χρήση καρτών γραφικών, και χρειάστηκε 3 ημέρες σε μια συστάδα 1000 υπολογιστών με 16000 πυρήνες. Τελευταία έχουν προταθεί αρχιτεκτονικές με ακόμα περισσότερα επίπεδα, βελτιώνοντας περαιτέρω τις αποδόσεις σε προβλήματα ταξινόμησης εικόνων [106, 104, 64].

Τα ΣΝΔ εξετάζονται τα τελευταία χρόνια σε πολλές ερευνητικές εργασίες, όπου προτείνονται νέες αρχιτεκτονικές, ή εφαρμογές τους σε άλλα προβλήματα πέρα από την ταξινόμηση εικόνων. Παρακάτω περιγράφουμε τις σημαντικότερες από τις πλέον πρόσφατες εργασίες που χρησιμοποιούν ΣΝΔ, για διαφορετικά προβλήματα της ανάλυσης εικόνων και βίντεο.

Εξαγωγή περιγραφέων

Οι Sermanet et al. [103] εκπαίδευσαν ένα δίκτυο αντίστοιχο του AlexNet, και το χρησιμοποίησαν για ανίχνευση αντικειμένων, πέρα από την ταξινόμηση εικόνων. Αντιμετωπίζοντας τα πλήρως συνδεδεμένα επίπεδα σαν συνελικτικά επίπεδα μεγέθους $1 \times 1 \times N$ (όπου N το μέγεθος του πλήρως συνδεδεμένου επιπέδου), πρότειναν την αποδοτική χρήση των ΣΝΔ για ανίχνευση αντικειμένων με χρήση της μεθοδολογίας ολισθαίνοντος παραθύρου (sliding window). Επίσης, πρότειναν τη χρήση των αποκρίσεων των ΣΝΔ σε μια εικόνα ως περιγραφείς, χρησιμοποιώντας την τελική έξοδο του δικτύου (τελευταίο πληρως συνδεδεμένο επίπεδο).

Οι Danahue et al. [24] χρησιμοποίησαν το δίκτυο AlexNet των Krizhevsky et al. που εκπαιδεύτηκε για ταξινόμηση εικόνων στις 1000 κατηγορίες του συνόλου Imagenet, στην εξαγωγή περιγραφέων και εφαρμογή σε αναγνώριση σκηνής και αναγνώριση υποκατηγοριών, σε διαφορετικά σύνολα εικόνων. Εξέτασαν τη χρήση του τελευταίου συνελικτικού επιπέδου, καθώς και των δύο πρώτων πλήρως συνδεδεμένων επιπέδων, για την εξαγωγή περιγραφέων, ενώ ο ταξινομητής που χρησιμοποίησαν είναι ένα γραμμικό SVM. Η μέθοδος ξεπέρασε τις ανταγωνιστικές μεθόδους της βιβλιογραφίας με χρήση τοπικών χαρακτηριστικών SURF, δείχνοντας ότι είναι δυνατή η μεταφορά της μάθησης ενός βαθιού ΣΝΔ σε διαφορετικά προβλήματα με διαφορετικά σύνολα εικόνων.

Οι Girshick et al. [41] χρησιμοποίησαν νευρωνικά δίκτυα για την εξαγωγή περιγραφέων από μια περιοχή εικόνας, με σκοπό την ανίχνευση αντικειμένων. Εξήγαγαν προτάσεις περιοχών που έχουν μεγάλη πιθανότητα να περικλείουν αντικείμενα (βλ. ενότητα 2.5) χρησιμοποιώντας τον αλγόριθμο επιλεκτικής αναζήτησης (selective search) [112]. Για την εξαγωγή περιγραφέων από τις προτεινόμενες περιοχές χρησιμοποίησαν το προτελευταίο πλήρως συνδεδεμένο επίπεδο της τοπολογίας δικτύου των Krizhevsky et al. [54], το οποίο πέτυχε καλύτερη ακρίβεια στο σύνολο Pascal VOC [32] από το πρώτο πλήρως συνδεδεμένο επίπεδο καθώς και το τελευταίο συνελικτικό. Η εκπαίδευση του δικτύου ξεκίνησε στο σύνολο Imagenet, ενώ συνέχισαν την εκπαίδευση στο σύνολο εικόνων Pascal VOC 2010, με σκοπό τη μικρορύθμιση των παραμέτρων του δικτύου. Για την ταξινόμηση των περιοχών χρησιμοποίησαν ένα γραμμικό SVM για κάθε κατηγορία, και τα αποτελέσματα της μεθόδου ξεπέρασαν αυτά της βιβλιογραφίας.

Ακολούθησαν πολλές μέθοδοι που χρησιμοποίησαν τα ΣΝΔ για την εξαγωγή περιγραφέων εικόνων, ελέγχοντας την απόδοση διαφορετικών επιπέδων των δικτύων σε διάφορες εφαρμογές [83, 95, 7, 121, 16].

Βίντεο

Οι Karapthy et al. [51] πρότειναν τη χρήση βαθιών ΣΝΔ για την ταξινόμηση βίντεο σε συλλογές μεγάλης κλίμακας. Εξέτασαν διαφορετικές προσεγγίσεις για τη μεταφορά των ΣΝΔ από την εφαρμογή σε στατικές εικόνες σε βίντεο και τη συγχώνευση πληροφορίας. Ωστόσο, η εξαγωγή περιγραφέων ανά καρέ, έδωσε αντίστοιχα αποτελέσματα με τις πιο σύνθετες προσεγγίσεις που εξετάστηκαν. Για την αποδοτική εφαρμογή των ΣΝΔ σε ένα καρέ του βίντεο πρότειναν τη χρήση δύο δικτύων με χαμηλή ανάλυση, ενώ τα βίντεο χωρίστηκαν σε πλάνα μήκους μόλις $0,5sec$, και για την ταξινόμηση ενός βίντεο χρησιμοποίησαν το μέσο όρο των αποτελεσμάτων της ταξινόμησης 20 τυχαίων πλάνων.

Οι Ng et al. [123] πρότειναν το συνδυασμό της χρωματικής πληροφορίας με την οπτική ροή των καρέ για την εξαγωγή ενός καθολικού περιγραφέα βίντεο. Πειραματίστηκαν με διάφορες τεχνικές για τη συγχώνευση πληροφορίας από τα καρέ, επιλέγοντας τη συγκέντρωση μεγίστου (max pooling). Εξέτασαν επίσης τη συγκέντρωση σε διαφορετικά επίπεδα του ΣΝΔ.

Χωρο-χρονικά χαρακτηριστικά

Οι Le et al. [58] εκμεταλλεύτηκαν ιδέες των ΣΝΔ και κατασκεύασαν ένα δίκτυο 2 επιπέδων για την εκμάθηση και εξαγωγή χωρο-χρονικών χαρακτηριστικών από μη επισημασμένα βίντεο για αναγνώριση δράσεων. Οι Ji et al. [48] χρησιμοποίησαν ΣΝΔ για αναγνώριση δράσεων, εξάγοντας χωρο-χρονικά χαρακτηριστικά από βίντεο. Πρότειναν τη

χρήση 3Δ φίλτρων στα συνελλικτικά επίπεδα, εξάγοντας 3Δ αποκρίσεις. Ωστόσο, χρησιμοποίησαν επιλεγμένα χαρακτηριστικά (όπως παραγώγους και οπτική ροή) για το πρώτο επίπεδο του δικτύου, και ανιχνευτή προσώπων για τον περιορισμό του βίντεο εισόδου σε τμήμα των καρέ. Οι Tran et al. [109] χρησιμοποίησαν ολόκληρα καρέ βίντεο για την αναγνώριση δράσεων και κατέληξαν στα καλύτερα αποτελέσματα χρησιμοποιώντας μικρά 3Δ φίλτρα μεγέθους $3 \times 3 \times 3$ (πλάτος, ύψος, χρόνος).

2.4 Ανάκτηση βίντεο

Στο κεφάλαιο 5 περιγράφουμε μια νέα μεθοδολογία για ανάκτηση βίντεο με βάση το περιεχόμενο. Στη μέθοδο που προτείνουμε, καθώς και σε πολλές άλλες της βιβλιογραφίας, πραγματοποιείται πρώτα μια κατάτμηση των βίντεο σε πλάνα (*shots*). Ένα πλάνο οριοθετείται από μια έντονη μετάβαση (κόψιμο), ή ομαλή μετάβαση με αλλαγή φωτεινότητας [63]. Στόχος της κατάτμησης σε πλάνα είναι να χωριστεί ένα βίντεο σε τμήματα που έχουν συνοχή ως προς το περιεχόμενό τους. Πολλοί αλγόριθμοι έχουν προταθεί για την ανίχνευση πλάνων σε βίντεο, χρησιμοποιώντας ιστογράμματα φωτεινότητας και χρώματος, εκτίμηση κίνησης, ή κατάτμηση γράφων [87, 63, 122].

Η εξαγωγή περιγραφέων από τα καρέ του βίντεο και η ενοποίησή τους για κάθε πλάνο, με σκοπό τη δεικτοδότηση των πλάνων για ταίριασμα με βάση το περιεχόμενο, οδηγεί σε μεγάλο χρονικό κόστος κατά την αναζήτηση ενός ερωτήματος [52]. Οι Douze et al. [25] προτείνουν μια μέθοδο για την κωδικοποίηση των περιγραφέων που εξάγουν από όλα τα καρέ του βίντεο σε μια χρονική αναπαράσταση, με στόχο την αποδοτική αναζήτηση πλάνων, καθώς και τη χρονική ευθυγράμμιση πλάνων από διαφορετικά βίντεο. Εκμεταλλεύονται τις ιδιότητες των κυκλοειδών πινάκων για να επιτύχουν αποδοτικές συγκρίσεις μεταξύ πλάνων στο πεδίο της συχνότητας.

Το πρόβλημα της συμπαγούς αναπαράστασης βίντεο αντιμετωπίζεται και από αλγόριθμους αναγνώρισης δράσεων. Οι Sun et al. [105] κατασκευάζουν χωρο-χρονικές τροχιές, ταιριάζοντας τοπικά χαρακτηριστικά SIFT [70] μεταξύ δύο διαδοχικών καρέ. Για κάθε τροχιά κατασκευάζουν περιγραφείς, συνδυάζοντας τους περιγραφείς SIFT όλων των τοπικών χαρακτηριστικών κάθε τροχιάς, παίρνοντας τον μέσο όρο σε κάθε διάσταση του περιγραφέα. Έπειτα, χρησιμοποιώντας το μοντέλο *Bag-of-Words (BoW)* επιτυγχάνουν το αποδοτικό ταίριασμα τμημάτων βίντεο. Η περιγραφή βίντεο με χρήση τοπικών χαρακτηριστικών που προτείνουμε στο κεφάλαιο 5 σχετίζεται με αυτή την εργασία, καθώς χρησιμοποιούμε χωρο-χρονικές τροχιές ταιριάζοντας τοπικά χαρακτηριστικά. Ωστόσο, επεκτείνουμε την περιγραφή ταιριάζοντας περισσότερα χαρακτηριστικά από καρέ σε καρέ με χρήση του επαναληπτικού αλγορίθμου RANSAC.

Οι Wang et al. [118] εξάγουν πυκνά σημεία σε κάθε καρέ και υπολογίζουν την οπτική

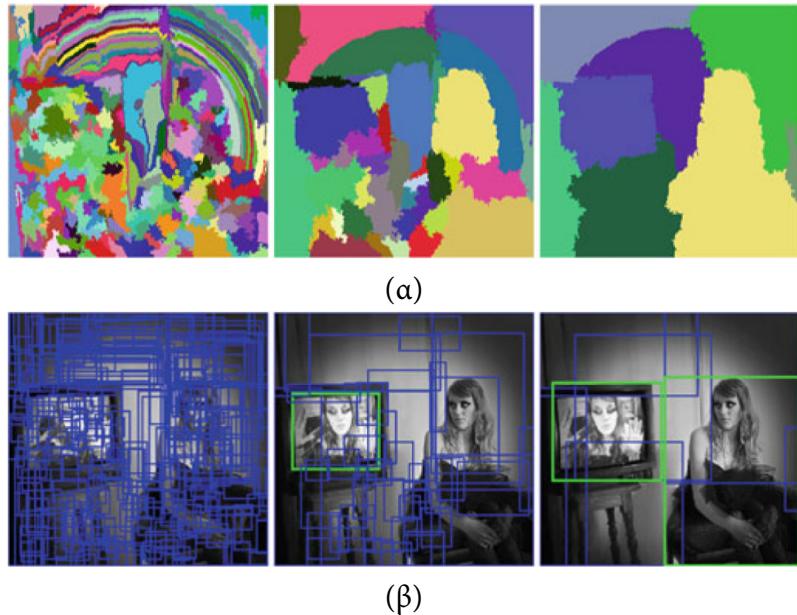
ροή μεταξύ των καρέ για την κατασκευή τροχιών χαρακτηριστικών. Εξάγουν χωρο-χρονικούς περιγραφείς HOG [20], HOF [56], και MBH [21] επί των χωρο-χρονικών τροχιών και χρησιμοποιούν το μοντέλο BoW για την αναγνώριση δράσεων. Ο περιγραφέας κατευθυνόμενων κλίσεων *HOG* (*Histograms of Oriented Gradients*) επικεντρώνεται στη στατική περιγραφή της εμφάνισης, συναθροίζοντας το μέτρο της κλίσης σε μια τοπική γειτονιά για ένα σύνολο διαφορετικών γωνιών της κλίσης. Ο περιγραφέας οπτικής ροής *HOF* (*Histograms of Optical Flow*) επικεντρώνεται στην τοπική πληροφορία κίνησης, εξάγοντας ιστογράμματα της οπτικής ροής, αντίστοιχα με αυτά του HOG. Ο περιγραφέας ορίων κίνησης *MBH* (*Motion Boundary Histograms*) κωδικοποιεί τη σχετική κίνηση μεταξύ εικονοστοιχείων, υπολογίζοντας ιστογράμματα των παραγώγων της οπτικής ροής στην οριζόντια και κατακόρυφη κατεύθυνση. Οι ίδιοι συγγραφείς [119], βελτιώνουν τις χωρο-χρονικές τροχιές, αφαιρώντας την κίνηση της κάμερας. Για τον υπολογισμό της κίνησης της κάμερας εξάγουν τοπικά χαρακτηριστικά SURF [8] τα οποία ταιριάζουν μεταξύ διαδοχικών καρέ με χρήση του αλγορίθμου RANSAC [35].

2.5 Περιοχές εικόνων που περικλείονται αντικείμενα

Για την κατάτμηση των βίντεο σε πλάνα στο κεφάλαιο 5, εξετάζουμε το μέγεθος “*objectness*” σε κάθε καρέ. Το μέγεθος “*objectness*” για περιοχές εικόνων έχει εισαχθεί πρόσφατα [3], και αντιστοιχεί σε ένα μέτρο που σηματοδοτεί την πιθανότητα ύπαρξης ενός αντικειμένου σε ένα υποπαράθυρο εικόνας, ανεξάρτητα από το συγκεκριμένο είδος του αντικειμένου. Το “*objectness*” σχετίζεται με τον εντοπισμό εξεχουσών περιοχών [15], και έχει επεκταθεί και σε χωρο-χρονικές περιοχές σε βίντεο [113]. Οι αλγόριθμοι που προτείνουν περιοχές υψηλού “*objectness*” χωρίζονται σε δύο κατηγορίες: α) αυτούς που ξεκινούν από υπερκατάτμηση της εικόνας και με συνενώσεις περιοχών καταλήγουν σε υποψήφιες περιοχές, και β) αυτούς που ξεκινούν από ξεχωριστά υποπαράθυρα της εικόνας και υπολογίζουν το “*objectness*” σε αυτά.

Στην πρώτη κατηγορία ανήκει ο αλγόριθμος επιλεκτικής αναζήτησης (*selective search*) [112], που συνενώνει περιοχές με βάση την ομοιότητά τους ως προς το χρώμα και την υφή. Για την βελτιστοποίηση των αποτελεσμάτων χρησιμοποιεί πολλαπλούς χρωματικούς χώρους και αλγορίθμους αρχικής κατάτμησης, με αποτέλεσμα την υψηλή πολυπλοκότητα του αλγορίθμου. Στην Εικόνα 2.3 βλέπουμε τις κατατμήσεις και τις τελικές περιοχές.

Οι Manen et al. [74] προτείνουν έναν αλγόριθμο που βασίζεται στην επιλεκτική αναζήτηση, αλλά ξεκινάει από κατάτμηση σε υπερεικονοστοιχεία, και εισάγοντας τυχαιότητα στον τρόπο συνένωσης των αρχικών τμημάτων, μειώνει τον απαιτούμενο χρόνο για την πρόταση περιοχών με υψηλό “*objectness*”. Οι Krähenbühl και Koltun [53] προτείνουν τη χρήση του γεωδαιτικού μετασχηματισμού απόστασης (GMA) για τη συνένωση τμημάτων

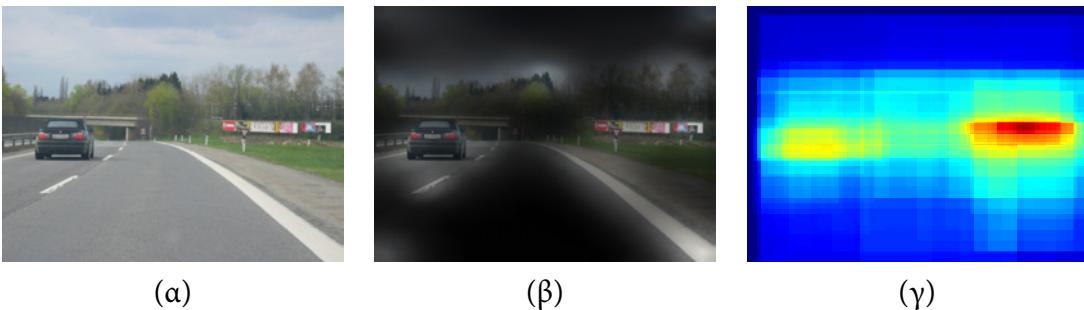


Εικόνα 2.3: Πρόταση περιοχών του αλγορίθμου *Selective Search*. (α) Κατατμήσεις της εικόνας σε διαφορετικές κλίμακες και (β) οι αντίστοιχες προτάσεις περιοχών που προκύπτουν. Με πράσινο χρώμα εμφανίζονται οι περιοχές που περικλείονται αντικείμενα της εικόνας. (από [112])

της εικόνας, ξεκινώντας από μια κατάτμηση της εικόνας σε υπερεικονοστοιχεία. Δημιουργείται ένας μικρός αριθμός σπόρων για τον υπολογισμό του ΓΜΑ, και επιλέγοντας ένα σύνολο από κρίσιμα επιπεδοσύνολα του μετασχηματισμού εξάγονται οι περιοχές με υψηλό “objectness”.

Στη δεύτερη κατηγορία αλγορίθμων, οι Alexe et al. [3] προτείνουν τον αλγόριθμο *Objectness*, όπου κατασκευάζουν όλα τα πιθανά υποπαράθυρα της εικόνας και επιλέγουν αυτά με τη μεγαλύτερη πιθανότητα να περιέχουν αντικείμενα. Για την επιλογή χρησιμοποιούνται διάφορα κριτήρια σχετικά με το χρώμα, τις ακμές, τον εντοπισμό εξεχουσών περιοχών, ενώ το σημαντικότερο κριτήριο αποτελεί το κατά πόσο ένα παράθυρο περικλείει κλειστά περιγράμματα από ενώσεις υπερεικονοστοιχείων (βλ. Εικόνα 2.4).

Οι Rahtu et al. [93] βελτιώνουν την απόδοση του [3] προτείνοντας τη χρήση διαφορετικών χαρακτηριστικών για τη βαθμολόγηση των υποπαραθύρων. Τέλος, ο αλγόριθμος BING των Cheng et al. [17] βαθμονομεί τα υποψήφια υποπαράθυρα μετρώντας την ομοιότητά τους με ένα πρότυπο ακμών που έχει μάθει από ένα σύνολο αντικειμένων διαφόρων κατηγοριών. Χρησιμοποιώντας συρρικνωμένες εικόνες σε μέγεθος 8×8 εικονοστοιχείων, επιτυγχάνει την εξαγωγή των υποψήφιων περιοχών με πολύ μικρό χρονικό κόστος, της τάξης των $msec$. Ωστόσο, οι Zhao et al. [124] υποστήριξαν πως η επιτυχία της μεθόδου BING οφείλεται στον τρόπο επιλογής των αρχικών υποψήφιων υποπαραθύρων, και όχι στην εκμάθηση ενός αντιπροσωπευτικού προτύπου, καθώς πέτυχαν αντίστοιχα αποτελέσματα χρησιμοποιώντας διαφορετικά πρότυπα, όπως ένα Gaussian παράθυρο και ένα



Εικόνα 2.4: Πρόταση περιοχών του αλγορίθμου Objectness. Για την αρχική εικόνα (α), υπολογίζεται ο χάρτης σημαντικότητας (β) και σε συνδυασμό με άλλα κριτήρια βαθμολογείται κάθε υποψήφιο υποπαράθυρο της εικόνας (γ).

φίλτρο απλού μέσου όρου.

Πρόσφατα έχουν χρησιμοποιηθεί για την αναγνώριση δράσεων σε βίντεο αλγόριθμοι πρότασης περιοχών σε εικόνες. Οι Jain et al. [47] επεκτείνουν τις προτάσεις του αλγορίθμου επιλεκτικής αναζήτησης από τις εικόνες στα βίντεο, με σκοπό την επιλογή υποψήφιων κύβων στον 3Δ χωρο-χρόνο, που περιλαμβάνουν δράσεις αντί για αντικείμενα. Οι Γκιο-ξάρη et al. [42] εξάγουν προτάσεις περιοχών με τον αλγόριθμο επιλεκτικής αναζήτησης σε καρέ βίντεο, και επιλέγουν μόνο τις προτάσεις που ταιριάζουν με την εκτίμηση κίνησης που υπολογίζουν. Έπειτα, εξάγουν περιγραφείς των περιοχών χρησιμοποιώντας δύο νευρωνικά δίκτυα, ένα που επικεντρώνεται στην εμφάνιση και ένα στην κίνηση.

Περιοχές εικόνων που περικλείονται αντικείμενα

Κεφάλαιο 3

Ανίχνευση χαρακτηριστικών με α-σχήματα

Σε αυτό το κεφάλαιο παρουσιάζουμε τον αλγόριθμο που προτείνουμε για ανίχνευση τοπικών χαρακτηριστικών με γεωμετρικές κατασκευές. Αρχικά θα παρουσιάσουμε την κατασκευή των α-σχημάτων σε εικόνες, η οποία περιλαμβάνει τρία βήματα: α) τη δειγματοληψία σημείων από την εικόνα, β) την τριγωνοποίηση των δειγμάτων, και γ) τη μέτρηση ποσοτήτων στις όψεις της τριγωνοποίησης για την κατασκευή της α-διήθησης και των διαφορετικών α-σχημάτων. Έπειτα θα περιγράψουμε πως χρησιμοποιούμε τα α-σχήματα για την ομαδοποίηση μικρών περιοχών της εικόνας και το σχηματισμό περιοχών που έχουν εξέχουσα οπτική σημασία. Θα παρουσιάσουμε διαφορετικές μεθόδους για τη μέτρηση της σημαντικότητας κάθε περιοχής που κατασκευάζεται και την τελική επιλογή των περιοχών που θα οδηγήσουν στα τοπικά χαρακτηριστικά. Τέλος, θα παρουσιάσουμε ολοκληρωμένο τον προτεινόμενο αλγόριθμο εξαγωγής χαρακτηριστικών WaSH, μαζί με ποιοτικά παραδείγματα, με τα οποία θα εξηγήσουμε τις ιδιότητες του ανιχνευτή.

Για το στάδιο της δειγματοληψίας εικόνων, προτείνουμε δύο εναλλακτικές μεθόδους, οι οποίες χρησιμοποιούν δυαδικές ακμές της εικόνας. Τα σημεία ανήκουν στις ακμές, ενώ η πυκνότητά τους είναι σταθερή στη μία περίπτωση και μεταβαλλόμενη στη δεύτερη, εξαρτώμενη από τοπικές ιδιότητες των ακμών.

Στο βήμα της τριγωνοποίησης των σημείων εξετάζουμε τρεις εναλλακτικές τριγωνοποιήσεις της βιβλιογραφίας: την κανονική με προστιθέμενα βάρη, τη Delaunay και τη Delaunay με περιορισμούς. Οι διαφορετικές τριγωνοποιήσεις συνδυάζονται με την κατασκευή διαφορετικών α-σχημάτων. Από τη Delaunay τριγωνοποίηση (είτε με, είτε χωρίς περιορισμούς), κατασκευάζουμε τα απλά α-σχήματα. Από την κανονική τριγωνοποίηση κατασκευάζουμε τα σταθμισμένα και τα ανισοτροπικά σταθμισμένα α-σχήματα. Τα ανισοτροπικά σταθμισμένα α-σχήματα συνδυάζονται με τη μέθοδο που χρησιμοποιούμε για την ανομοιόμορφη εξαγωγή των δειγμάτων επί ακμών, όπως θα αναλύσουμε παρακάτω.

3.1 Δειγματοληψία

Το πρώτο βήμα για την κατασκευή των α -σχημάτων είναι η δημιουργία του συνόλου σημείων. Τα σημεία θα προκύψουν από δειγματοληψία της εικόνας, ώστε τα α -σχήματα ως τελική κατασκευή να αναπαριστούν τα αντικείμενα της εικόνας. Προτείνουμε δύο μεθόδους δειγματοληψίας σε ακμές της εικόνας (μονοδιάστατα μονοπάτια)· την *ομοιόμορφη* και την *ανομοιόμορφη* δειγματοληψία. Στην ομοιόμορφη δειγματοληψία, τα σημεία επιλέγονται ομοιόμορφα κατά μήκος των ακμών, χρησιμοποιώντας ένα σταθερό βήμα δειγματοληψίας. Αντίθετα, στην ανομοιόμορφη δειγματοληψία η πυκνότητα των δειγμάτων καθορίζεται από μία μέτρηση της τοπικής ανισοτροπίας, η οποία μεταβάλει το βήμα δειγματοληψίας για κάθε δείγμα.

Και στις δύο μεθόδους δειγματοληψίας, τα δείγματα δεν αντιστοιχούν σε σημεία-κλειδιά, όπως μέγιστα της καμπυλότητας ή της κλίσης. Ο στόχος μας είναι να λάβουμε τον χαμηλότερο δυνατό αριθμό δειγμάτων που θα μπορεί να συλλάβει το σχήμα των ορίων των αντικειμένων σε μια εικόνα.

3.1.1 Αναπαράσταση Εικόνων

Θεωρούμε ότι η εικόνα εισόδου είναι μονόχρωμη και δίνεται ως συνάρτηση f του επέδου. Σε περίπτωση που η εικόνα εισόδου είναι έγχρωμη, απορρίπτουμε όλη τη χρωματική πληροφορία και κρατάμε μόνο τη φωτεινότητα της εικόνας. Υπολογίζουμε την κλίση της εικόνας ∇f , καθώς και το μέτρο της κλίσης $G = \|\nabla f\|$. Θέτουμε ως \hat{G} το κανονικοποιημένο στο διάστημα $[0, 1]$ μέτρο της κλίσης.

Εφαρμόζουμε έναν δυαδικό ανιχνευτή ακμών στο μέτρο της κλίσης \hat{G} . Έπειτα δειγματοληπτούμε επί του χάρτη ακμών E για να εξάγουμε ένα διακριτό σύνολο σημείων των ακμών $P \subseteq \mathbb{R}^2$. Το πεδίο ορισμού και το σύνολο τιμών των συναρτήσεων f, G είναι φραγμένα και διακριτά, και το σύνολο P είναι πεπερασμένο.

Ο δυαδικός χάρτης ακμών E κατασκευάζεται από τον ανιχνευτή ακμών *Canny* [12]. Μετά το στάδιο καταστολής των μη-μεγίστων του αλγορίθμου, ακολουθούμε όλα τα υποψήφια εικονοστοιχεία ακμής και εφαρμόζουμε τα κατώφλια υστέρησης. Ο χάρτης ακμών δεν απαιτείται να είναι καθαρός (γεγονός που θα απαιτούσε προσαρμογή των κατωφλίων ανά εικόνα), και κατ' επέκταση διατηρούμε τα κατώφλια υστέρησης του αλγορίθμου (χαμηλό και υψηλό) σταθερά.

Η διαδικασία της δειγματοληψίας ξεκινά από ένα τυχαίο εικονοστοιχείο που ανήκει στις ακμές τις εικόνας (από το προηγούμενο στάδιο του ανιχνευτή ακμών) το οποίο λαμβάνεται ως δείγμα. Κατόπιν λαμβάνουμε δείγματα ακολουθώντας τις ακμές, χρησιμοποιώντας μια γειτονιά συνεκτικότητας 8 (*8-connected neighborhood*) και μετρώντας βήματα.

3.1.2 Ομοιόμορφη δειγματοληψία

Στην ομοιόμορφη δειγματοληψία χρησιμοποιούμε ένα σταθερό διάστημα δειγματοληψίας s . Μετράμε τα βήματα που διανύουμε από το αρχικό δείγμα καθώς κινούμαστε επί της ακμής, ελέγχοντας την 8-συνδεδεμένη γειτονιά κάθε εικονοστοιχείου ακμής. Όταν ο αριθμός των βημάτων γίνει ίσος με s , λαμβάνουμε ένα καινούριο δείγμα στην τρέχουσα θέση και μηδενίζουμε τον μετρητή βημάτων. Αντί της παραπάνω μεθόδου, θα μπορούσαμε να υπολογίσουμε την Ευκλείδεια απόσταση μεταξύ των δειγμάτων σε κάθε θέση. Το υπολογιστικό κόστος, όμως, θα ανέβαινε και με τη μέθοδο αυτή θα οδηγούμασταν σε υποδειγματοληψία των ακμών σε περιοχές υψηλής καμπυλότητας.

Για κάθε σημείο $p \in P$ που έχει προκύψει από τη δειγματοληψία, ορίζουμε το βάρος του $w(p) \geq 0$ να είναι ανάλογο του κανονικοποιημένου μέτρου της κλίσης στο σημείο $\hat{G}(p)$,

$$w(p) = \hat{G}(p) \left(\frac{s}{2} \right)^2, \quad (3.1)$$

όπου $\hat{G}(p) \in [0, 1]$. Ο λόγος της επιλογής αυτής για το βάρος w θα γίνει σαφής στο τέλος της ενότητας 3.3.1.

3.1.3 Ανομοιόμορφη δειγματοληψία

Στην μέθοδο της ανομοιόμορφης δειγματοληψίας ακμών προσαρμόζουμε το διάστημα δειγματοληψίας s στο τοπικό σχήμα της ακμής για κάθε δείγμα. Υπολογίζουμε το τοπικό αφινικό σχήμα στο τρέχον σημείο και χρησιμοποιώντας τις ιδιότητες του σχήματος καθορίζουμε το διάστημα δειγματοληψίας ως το επόμενο δείγμα. Με αυτό τον τρόπο τα δείγματα θα περιγράφουν καλύτερα και πιο αποδοτικά το σχήμα των ακμών των αντικειμένων. Η περιγραφή βελτιώνεται καθώς σε περιοχές υψηλής καμπυλότητας και λεπτομερειών μικρής κλίμακας τα δείγματα θα είναι πυκνά, ώστε η αναπαράσταση μέσω των δειγμάτων να είναι ακριβής. Αντίθετα, σε περιοχές χαμηλής καμπυλότητας (ευθύγραμμα τμήματα) η δειγματοληψία θα είναι αραιή, μειώνοντας το υπολογιστικό κόστος της μεθόδου χωρίς να αλλοιώνουμε την περιγραφή του σχήματος.

Η μέθοδος που χρησιμοποιούμε για την προσαρμογή του τοπικού αφινικού σχήματος βασίζεται στην εργασία των Lindeberg et al. [67, 4]. Δεδομένου ενός δειγματοληπτημένου σημείου p , ακολουθούμε μια επαναληπτική διαδικασία όπου σε κάθε επανάληψη i υπολογίζουμε το τοπικό σχήμα $U^{(i)}$ της προσαρμοσμένης στο σχήμα γειτονιάς \mathbf{p} του p . Το τοπικό σχήμα $U^{(i)}$ μοντελοποιείται από έναν 2×2 πίνακα μετασχηματισμού. Αρχικοποιούμε το τοπικό σχήμα ως ισοτροπικό μοναδιαίας κλίμακας, δηλ. ο $U^{(0)}$ είναι ο μοναδιαίος πίνακας I . Στην επανάληψη i μετασχηματίζουμε τη γειτονιά \mathbf{p} σε $\hat{\mathbf{p}} = U^{(i)}\mathbf{p}$, επανυπολογίζουμε την φωτεινότητα της εικόνας στο $\hat{\mathbf{p}}$, και τέλος υπολογίζουμε το μέσο πίνακα $\rho_{\text{op}}^{(i)}$

τάξης

$$\mu = \begin{bmatrix} \mu_{11} & \mu_{12} \\ \mu_{12} & \mu_{22} \end{bmatrix} = \begin{bmatrix} L_x^2 & L_x L_y \\ L_x L_y & L_y^2 \end{bmatrix}. \quad (3.2)$$

επί της γειτονιάς $\hat{\mathbf{p}}$, όπου L_x, L_y είναι οι παράγωγοι 1^{ης} τάξης στις κατευθύνσεις x, y αντίστοιχα. Για τη μετασχηματισμένη γειτονιά $\hat{\mathbf{p}}$ θα λαμβάνουμε τιμές σε ένα πίνακα σταθερού μεγέθους, ίσο με την αρχική γειτονιά \mathbf{p} (βλ. ενότητα 3.5.2). Για να αποφύγουμε την απόκλιση της διαδικασίας σε περίπτωση ευθείας γραμμής σε μια μικρή γειτονιά του p , θέτουμε ένα άνω όριο k στη εκκεντρότητα της έλλειψης που αντιστοιχεί στον πίνακα μ . Ακολούθωντας τους Almansa και Lindeberg [4], μετράμε τις παρακάτω ποσότητες επί του πίνακα μ :

$$B = \mu_{11} + \mu_{22} \quad (3.3)$$

$$C = \mu_{11} - \mu_{22} \quad (3.4)$$

$$S = 2\mu_{12} \quad (3.5)$$

$$Q = \sqrt{C^2 + S^2} \quad (3.6)$$

Το Q μετράει το βαθμό ανισοτροπίας, ενώ το $\hat{Q} = Q/B$ είναι κανονικοποιημένο στο διάστημα $[0, 1]$. Η εκκεντρότητα της έλλειψης που αντιστοιχεί στον πίνακα μ δίνεται από τον τύπο $(1 + \hat{Q})/(1 - \hat{Q})$. Σύμφωνα με τους Almansa και Lindeberg [4], η έλλειψη που αντιστοιχεί στον κανονικοποιημένο πίνακα διάχυσης

$$\Sigma = (\mu + \epsilon I)^{-1} \quad (3.7)$$

όπου $\epsilon = \frac{Q}{k-1}$, έχει τον ίδιο προσανατολισμό όπως η έλλειψη του μ , χωρίς να είναι τόσο επιμήκης: η εκκεντρότητά της είναι άνω φραγμένη από το k .

Υπολογίζουμε μια τετραγωνική ρίζα του Σ και κανονικοποιούμε τις ιδιοτιμές του $\lambda_{\min}, \lambda_{\max}$, ώστε

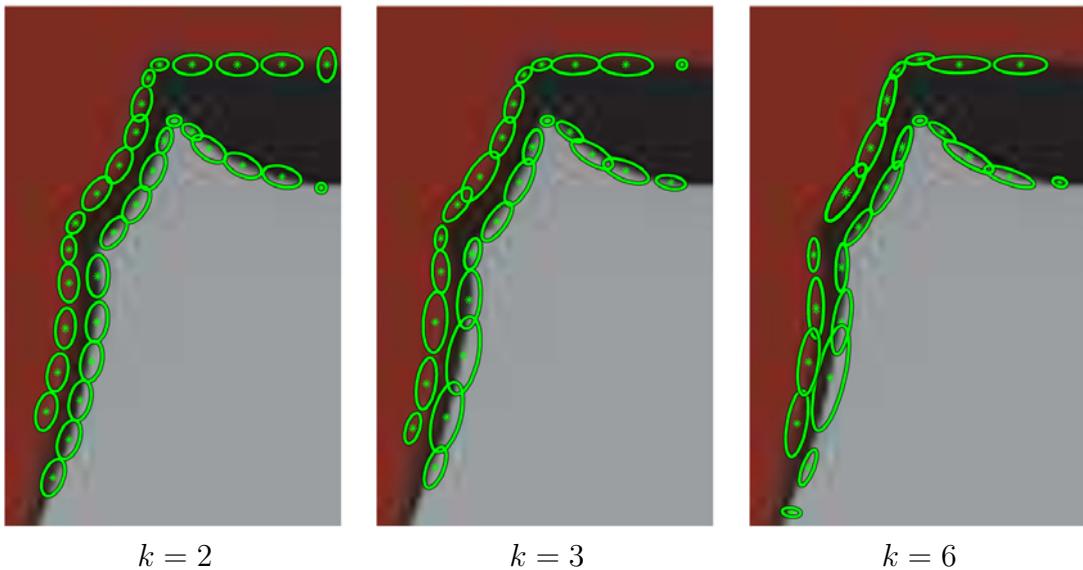
$$\lambda_{\min} \lambda_{\max} = 1. \quad (3.8)$$

Τέλος, εφαρμόζουμε επάλληλα τους μετασχηματισμούς που υπολογίσαμε πολλαπλασιάζοντας τον $\Sigma^{-\frac{1}{2}}$ με τον $U^{(i)}$ έτσι ώστε

$$U^{(i+1)} = \Sigma^{-\frac{1}{2}} U^{(i)}. \quad (3.9)$$

Για τη σύγκλιση της μεθόδου ελέγχουμε τον λόγο των ιδιοτιμών του Σ . Αν η τιμή του είναι κοντά στη μονάδα θεωρούμε ότι η μέθοδος έχει συγκλίνει, οπότε υπολογίζουμε τον τελικό πίνακα τοπικού σχήματος U . Τυπικά, η σύγκλιση επέρχεται σε 3 ως 5 επαναλήψεις.

Η μεγαλύτερη από τις εκάστοτε ιδιοτιμές χρησιμοποιείτε για τον υπολογισμό της κλίμακας της προσαρμοσμένης γειτονιάς \mathbf{p} . Δεδομένου ότι το σχήμα της \mathbf{p} περιγράφεται από



Εικόνα 3.1: Δείγματα με τη μέθοδο ανομοιόμορφης δειγματοληψίας. Η αύξηση του άνω ορίου της εκκεντρότητας k των περιοχών οδηγεί σε αραιότερη δειγματοληψία.

τον $U^{(i)}$, η κλίμακά της είναι το γινόμενο των μεγαλύτερων ιδιοτιμών επί όλων των επαναλήψεων μέχρι την επανάληψη i . Το διάστημα δειγματοληψίας s ελέγχεται από αυτή την κλίμακα, και πλέον εξαρτάται από το σημείο p :

$$s(p) = \prod_i \lambda_{\max}^{(i)}. \quad (3.10)$$

Με τη μέθοδο αυτή, έπειτα από το δείγμα στο σημείο p θα λάβουμε καινούριο δείγμα μετά από $s(p)$ βήματα επί της ακμής. Τέλος, αν U είναι ο τελικός πίνακας τοπικού σχήματος που μετρήθηκε στο p όταν η μέθοδος έφτασε σε σύγκλιση, ο μετρικός τανυστής

$$M_p = U^T U \quad (3.11)$$

Θα χρησιμοποιηθεί για να ορίσει τις αποστάσεις στη γειτονιά του p . Ο τανυστής αυτός θα χρησιμοποιηθεί στα ανισοτροπικά σταθμισμένα α -σχήματα στην ενότητα 3.3.2.

Στην Εικόνα 3.1 βλέπουμε πως η προτεινόμενη μέθοδος ανομοιόμορφης δειγματοληψίας συλλαμβάνει τις αλλαγές στο τοπικό σχήμα ενώ ταυτόχρονα ελαχιστοποιεί τον αριθμό των δειγμάτων. Σε περιοχές όπου οι ακμές γίνονται ευθύγραμμα τμήματα η δειγματοληψία είναι αραιή, αλλά παραμένει περιορισμένη από το άνω φράγμα k στην εκκεντρότητα των ελλείψεων. Σε περιοχές όπου η καμπυλότητα των ακμών αυξάνεται, η δειγματοληψία γίνεται πιο πυκνή ώστε να συλλάβει τις τοπικές λεπτομέρειες του σχήματος. Στην ίδια εικόνα μπορούμε να δούμε την επίδραση της παραμέτρου k . Μικρές τιμές για το k οδηγούν σε λεπτομερή αναπαράσταση, αυξάνοντας, όμως, την πολυπλοκότητα της μεθόδου στα επόμενα βήματα. Πολύ μεγάλες τιμές του k έχουν ως αποτέλεσμα μια αραιή αναπαράσταση, ενώ μπορούν να οδηγήσουν σε αφαίρεση κόμβων στην τριγωνοποίηση λόγω της επικάλυψης των ελλείψεων, όπως θα εξηγήσουμε στην ενότητα 3.2.1.

Συζήτηση

Χρησιμοποιώντας την μέθοδο ανομοιόμορφης δειγματοληψίας που προτείνουμε, το διάστημα δειγματοληψίας επηρεάζεται από την απόσταση μεταξύ γειτονικών ακμών, αν και δεν υπάρχει ακριβής θεωρητική σχέση μεταξύ των δύο. Μια ϵ -δειγματοληψία (ϵ -sampling) των ακμών, όπως περιγράφεται στο [14] παρέχει τις απαιτούμενες θεωρητικές εγγυήσεις. Εντούτοις, για την ϵ -δειγματοληψία ακμών απαιτείται η κατασκευή του διάμεσου άξονα (medial axis), η οποία έχει αυξημένο υπολογιστικό κόστος και συνήθως απαιτεί μια αρχική μορφή δειγματοληψίας.

Εναλλακτικά, η επιλογή της προσαρμογής του διαστήματος δειγματοληψίας σύμφωνα με την καμπυλότητα των ακμών είναι πολύ απλή υπολογιστικά καθώς καταλήγει σε μονοδιάστατους τελεστές, αλλά δε λαμβάνει υπόψιν τις γειτονικές ακμές. Το γεγονός αυτό μπορεί να οδηγήσει σε πολύ αραιές αναπαραστάσεις κοντινών ακμών που τείνουν προς ευθύγραμμα τμήματα, έχοντας αρνητικές επιπτώσεις στη σωστή αναπαράσταση μέσω της τριγωνοποίησης και τελικά στην εξαγωγή χαρακτηριστικών. Η μέθοδος που προτείνουμε, λαμβάνει υπόψιν την τοπική γειτονιά και έχει χαμηλό υπολογιστικό κόστος, γεγονός που την καθιστά μια συμβιβαστική λύση μεταξύ των δύο αυτών εναλλακτικών.

3.2 Τριγωνοποιήσεις

Σε αυτό το κεφάλαιο θα περιγράψουμε τις διαφορετικές τριγωνοποιήσεις που χρησιμοποιούμε στον ανιχνευτή χαρακτηριστικών. Οι τριγωνοποιήσεις αποτελούν τη βάση επί της οποίας υπολογίζονται τα α -σχήματα, όπως θα δούμε στην ενότητα 3.3.

3.2.1 Κανονική τριγωνοποίηση

Μια τριγωνοποίηση ενός συνόλου σημείων P στον χώρο \mathbb{R}^n λέγεται *κανονική* εάν μπορεί να κατασκευαστεί ως η προβολή του κάτω κυρτού περιβλήματος μιας ανύψωσης του P στον χώρο \mathbb{R}^{n+1} [22]. Οι κανονικές τριγωνοποιήσεις αποτελούν μια οικογένεια τριγωνοποιήσεων που διαφοροποιούνται μεταξύ τους μέσω της συνάρτησης ύψους που χρησιμοποιούμε για την ανύψωση στον χώρο \mathbb{R}^{n+1} .

Για την κατασκευή μιας κανονικής τριγωνοποίησης στο δισδιάστατο χώρο, ένα σύνολο σημείων ανυψώνεται στον χώρο τριών διαστάσεων εκχωρώντας μια τιμή ύψους σε κάθε σημείο p , χρησιμοποιώντας τη συνάρτηση ύψους $\omega : \mathbb{R}^2 \rightarrow \mathbb{R}$. Ένα ανυψωμένο σημείο ανήκει στον χώρο τριών διαστάσεων και βρίσκεται στη θέση $(p_x, p_y, \omega(p))$. Έπειτα, κατασκευάζεται το κάτω κυρτό περίβλημα του τρισδιάστατου συνόλου σημείων. Προβάλοντας το κάτω κυρτό περίβλημα στον χώρο δύο διαστάσεων δημιουργούμε μια διαμέριση του δισδιάστατου κυρτού περιβλήματος των αρχικών σημείων. Η διαμέριση αυτή είναι μια

κανονική τριγωνοποίηση.

Κάθε τριγωνοποίηση αποτελεί τη δυική αναπαράσταση ενός διαγράμματος τύπου Voronoi, όπου οι κορυφές της τριγωνοποίησης αποτελούν εστίες του διαγράμματος. Η δυικότητα εκφράζεται ως εξής: α) οι ακμές του διαγράμματος αντιστοιχούν σε ακμές της τριγωνοποίησης, ενώ β) τα σημεία όπου τέμνονται τρεις ακμές σε ένα διάγραμμα αντιστοιχούν σε τρίγωνα της τριγωνοποίησης. Έτσι, κάθε τριγωνοποίηση έχει ένα διάγραμμα Voronoi που της αντιστοιχεί, και το αντίθετο.

Χρησιμοποιούμε για την κατασκευή μιας κανονικής τριγωνοποίησης την συνάρτηση ύψους με προστιθέμενα βάρη

$$\omega(p) = \|p\|^2 - w(p), \quad (3.12)$$

όπου το $w(p)$ έχει οριστεί στην εξίσωση 3.1.

Κάθε σημείο $p \in P$ μαζί με το βάρος του $w(p)$ αποτελούν ένα ζευγάρι $(p, w(p))$ το οποίο ονομάζεται σταθμισμένο σημείο. Μπορούμε να δούμε ένα σταθμισμένο σημείο σαν ένα κύκλο με κέντρο το p και ακτίνα $\sqrt{w(p)}$. Η αναπαράσταση αυτή θα χρησιμοποιηθεί στο υπόλοιπο της εργασίας όταν αναφέρουμε τον κύκλο που αντιστοιχεί στο σταθμισμένο σημείο p . Επίσης θα χρησιμοποιούμε το ίδιο σύμβολο p είτε για ένα σταθμισμένο σημείο, είτε για τον κύκλο που αντιστοιχεί στο σημείο, διευκρινίζοντας σε κάθε περίπτωση.

Τα βάρη w θα πρέπει να επιλέγονται με προσοχή καθώς μπορεί να προκαλέσουν την εξαφάνιση σημείων στην τριγωνοποίηση. Ένα σημείο $p \in P$ μπορεί να διαγραφεί από την τριγωνοποίηση όταν ο κύκλος του p περικλείεται εξολοκλήρου στον κύκλο ενός άλλου σημείου q . Σε αυτή την περίπτωση, η απόσταση οποιουδήποτε σημείου στο χώρο από το p θα είναι μεγαλύτερη από την απόσταση του ίδιου σημείου από το q , γεγονός που θα μηδενίζει την επιφάνεια της περιοχής που αντιστοιχεί στο σημείο p στο διάγραμμα Voronoi. Αντίστοιχα, στον ανυψωμένο χώρο τριών διαστάσεων, το ανυψωμένο σημείο p δε θα ανήκει στο κάτω περίβλημα.

Η κανονική τριγωνοποίηση μπορεί να κατασκευαστεί με αυξητικό αλγόριθμο, προσθέτοντας τα σημεία ένα ένα. Οι Edelsbrunner και Shah [30] απέδειξαν ότι είναι δυνατή μια τέτοια κατασκευή με χρονική πολυπλοκότητα $O(n \log n + n)$ στην περίπτωση των δύο διαστάσεων, όπου $n = |P|$.

3.2.2 Τριγωνοποίηση Delaunay

Η τριγωνοποίηση Delaunay είναι η πιο διαδεδομένη τριγωνοποίηση που χρησιμοποιείται σε πληθώρα εφαρμογών (δισδιάστατα και τρισδιάστατα γραφικά, βιολογία κ.α.). Ανήκει στην οικογένεια των κανονικών τριγωνοποιήσεων, όπου η συνάρτηση ύψους είναι η $\omega(p) = \|p\|^2$. Η περίπτωση αυτή αντιστοιχεί στην κανονική τριγωνοποίηση με προστιθέμενα βάρη που περιγράψαμε στην ενότητα 3.2.1, όπου μηδενίζονται τα βάρη των σημείων,

δηλαδή $w(p) = 0$ για όλα τα σημεία $p \in P$ (βλ. εξίσωση 3.12). Καθώς η συνάρτηση ύψους υπολογίζεται χρησιμοποιώντας μόνο τις συντεταγμένες των σημείων, δε χρειάζεται επιπρόσθετη είσοδος στον αλγόριθμο τριγωνοποίησης.

Οι γεωμετρικές ιδιότητες μιας τριγωνοποίησης Delaunay είναι οι ακόλουθες:

1. Τρεις κορυφές ορίζουν τρίγωνο εάν και μόνο εάν ο περιγεγραμμένος ανοικτός κύκλος τους είναι κενός, δηλαδή δεν περιέχει άλλες κορυφές.
2. Δύο κορυφές ορίζουν ακμή εάν και μόνο εάν υπάρχει κενός κύκλος με τις κορυφές στην περιφέρειά του, δηλαδή υπάρχει κλειστός κύκλος με χορδή που ορίζεται από τις δύο κορυφές και δεν περιέχει καμία άλλη κορυφή.

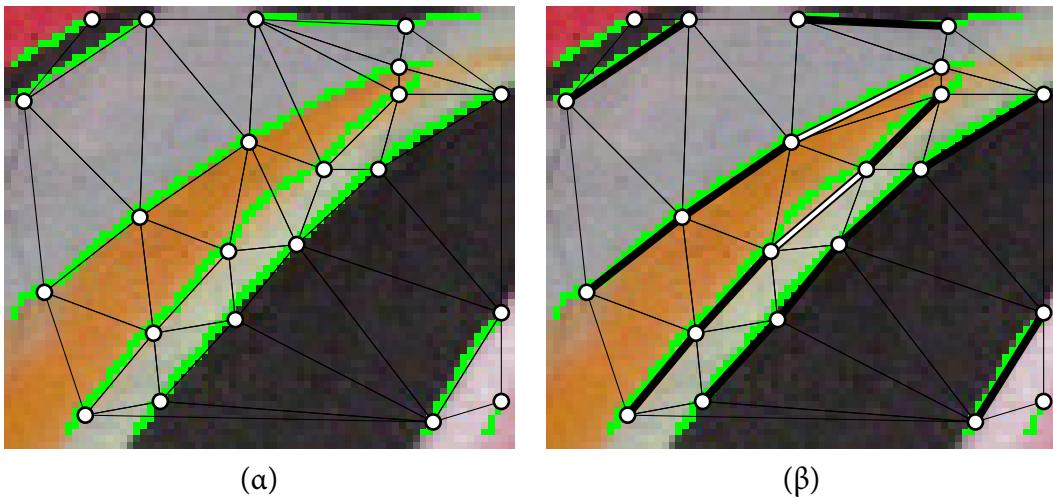
Για την κατασκευή της τριγωνοποίησης Delaunay, χρησιμοποιείται ένας αυξητικός αλγόριθμος με αναστροφή ακμών (*edge flipping*) χρονικής πολυπλοκότητας $O(n \log n)$. Ο αλγόριθμος ξεκινά από ένα σύνολο τριών σημείων που σχηματίζουν ένα τρίγωνο και συνεχίζει προσθέτοντας ένα σημείο τη φορά. Κατά την εισαγωγή ενός νέου σημείου, ελέγχεται αν τα νέα τρίγωνα που δημιουργούνται ικανοποιούν τις συνθήκες Delaunay. Σε περίπτωση μη συμμόρφωσης ενός ζεύγους τριγώνων με τις συνθήκες Delaunay, η κοινή ακμή τους αναστρέφεται.

3.2.3 Τριγωνοποίηση Delaunay με περιορισμούς

Η τριγωνοποίηση Delaunay με περιορισμούς (*constrained Delaunay*) είναι μια ειδική περίπτωση της τριγωνοποίησης Delaunay, που όμως ικανοποιεί μόνο μερικώς τις ιδιότητες Delaunay. Εκτός από τις συντεταγμένες των σημείων p , χρησιμοποιούμε ως είσοδο για την κατασκευή της τριγωνοποίησης και ένα σύνολο από περιορισμούς. Οι περιορισμοί υποδεικνύουν τις ακμές που θα πρέπει να υπάρχουν στην τριγωνοποίηση, ανεξάρτητα αν παραβιάζουν τις συνθήκες Delaunay. Κάθε περιορισμός c_{ij} αναπαριστάται από μια ακμή της μορφής $c_{ij} = (v_i, v_j)$, δηλώνοντας ότι οι κόμβοι v_i και v_j θα πρέπει να συνδέονται άμεσα με μια ακμή στην τριγωνοποίηση.

Στην περίπτωση που εξετάζουμε, οι περιορισμοί μπορούν να χρησιμοποιηθούν σε συνδυασμό με την δειγματοληψία επί ακμών (βλ. ενότητα 3.1). Μπορούμε, λοιπόν να ορίσουμε ως περιορισμούς τα συνεχόμενα τμήματα επί των ακμών. Συγκεκριμένα, καθώς ακολουθούμε μια ακμή εικόνας για να επιλέξουμε τα σημεία-δείγματα, προσθέτουμε ως περιορισμό κάθε ζευγάρι συνεχόμενων σημείων, ώστε τα σημεία αυτά να ενώνονται με μια ακμή στην τριγωνοποίηση. Με τον τρόπο αυτό εξασφαλίζουμε ότι οι ακμές της τριγωνοποίησης δε θα διασχίζουν τις ακμές τις εικόνας, και δημιουργούμε μια αναπαράσταση της εικόνας που σέβεται τα όρια των αντικειμένων.

Για την κατασκευή της τριγωνοποίησης Delaunay με περιορισμούς συνήθως χρησιμοποιείται μια παραλλαγή του αυξητικού αλγορίθμου που περιγράφηκε στην περίπτωση



Εικόνα 3.2: (α) Τριγωνοποίηση Delaunay. (β) Τριγωνοποίηση Delaunay με περιορισμούς. Οι περιορισμοί φαίνονται με παχιές μαύρες γραμμές. Οι περιορισμοί που προκαλούν αναστροφές ακμών φαίνονται με παχιές άσπρες γραμμές. Τα εικονοστοιχεία των ακμών φαίνονται με πράσινο χρώμα. Παρατηρούμε ότι οι περιορισμοί επιβάλλουν στις ακμές της τριγωνοποίησης να ακολουθούν τις ακμές τις εικόνας, με αποτέλεσμα η τριγωνοποίηση να αναπαριστά τις δομές της εικόνας με μεγαλύτερη ακρίβεια.

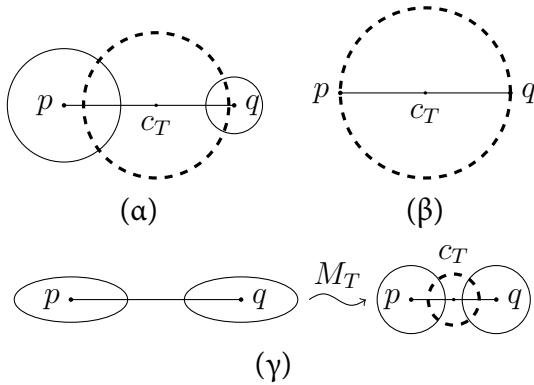
Delaunay (βλ. ενότητα 3.2.2). Εδώ, οι περιορισμοί προστίθεται πάντα ως ακμές. Εντούτοις, αν δύο ακμές από περιορισμούς τέμνονται, θα πρέπει πρώτα να χωριστούν στο σημείο τομής τους. Η περίπτωση αυτή μπορεί να συμβεί σε σημεία όπου διασταυρώνονται ακμές της εικόνας. Σε περίπτωση που ακμή από περιορισμό τέμνεται με μια απλή ακμή της τριγωνοποίησης, τότε η απλή ακμή θα πρέπει να αφαιρεθεί με μια σειρά από αναστροφές ακμών. Ένα παράδειγμα τριγωνοποίησης Delaunay με περιορισμούς φαίνεται στην Εικόνα 3.2.

3.3 α -σχήματα

Η κατασκευή των α -σχημάτων βασίζεται στην τριγωνοποίηση Delaunay. Τα σταθμισμένα α -σχήματα (*weighted α -shapes*) βασίζονται στην κανονική τριγωνοποίηση με συνάρτηση ύψους με προστιθέμενα βάρη (βλ. εξισωση 3.12). Τα ανισοτροπικά σταθμισμένα α -σχήματα (*anisotropically weighted α -shapes*) που προτείνουμε, βασίζονται στην ίδια κανονική τριγωνοποίηση και χρησιμοποιούν τον μετρικό τανυστή (βλ. εξισωση 3.11) που εξάγεται στην περίπτωση την ανομοιόμορφης δειγματοληψίας επί ακμών για τη μέτρηση των απαραίτητων αποστάσεων για την κατασκευή τους.

3.3.1 Σταθμισμένα α -σχήματα

Η περιγραφή των σταθμισμένων α -σχημάτων βασίζεται στην εργασία του Edelsbrunner [29], αλλά απλοποιείται στην περίπτωση των δύο διαστάσεων. Αρχικά δίνουμε έναν ορι-



Εικόνα 3.3: Μέγεθος της ακμής ενός συνόλου 2 σημείων $\{p, q\}$, ως το τετράγωνο της ακτίνας του κύκλου (σταθμισμένου σημείου) c_T που είναι ορθογώνιος ως προς τους κύκλους p, q , σε (α) μια κανονική και (β) μια Delaunay τριγωνοποίηση. (γ) Στα ανισοτροπικά σταθμισμένα α -σχήματα, ο χώρος μετασχηματίζεται σε ισοτροπικό προτού μετρηθεί το μέγεθος της ακμής.

σμό της κανονικής τριγωνοποίησης, εναλλακτικό αυτού που δόθηκε στην ενότητα 3.2.1. Ταυτόχρονα, ορίζουμε και άλλες ποσότητες που θα φανούν χρήσιμες στον ορισμό των σταθμισμένων α -σχημάτων. Υπενθυμίζουμε ότι το σύμβολο p αναφέρεται σε ένα σημείο σε δύο διαστάσεις, ένα σταθμισμένο σημείο $(p, w(p))$ ή τον κύκλο που αντιστοιχεί σε αυτό, αποσαφηνίζοντας σε κάθε περίπτωση. Κάθε τρίγωνο, ευθύγραμμο τμήμα ή σημείο μιας τριγωνοποίησης είναι ένα 2-, 1-, or 0-άπλοκο αντίστοιχα, στο οποίο θα αναφερόμαστε γενικά ως άπλοκο.

Δεδομένων δύο σταθμισμένων σημείων $p, q \in P$, ορίζουμε την ποσότητα

$$\pi(p, q) = \|p - q\|^2 - w(p) - w(q). \quad (3.13)$$

Οι κύκλοι των p, q τέμνονται σε ορθή γωνία εάν και μόνο εάν $\pi(p, q) = 0$, όπου λέμε ότι οι κύκλοι p, q είναι ορθογώνιοι. Δεδομένου ενός σημείου x με μηδενικό βάρος, ονομάζουμε την ποσότητα $\pi(p, x)$ δύναμη του σημείου x ως προς τον κύκλο p [19].

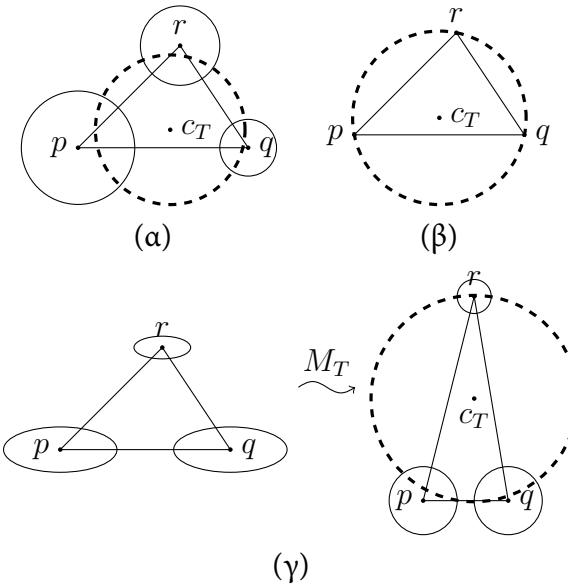
Δεδομένου ενός υποσυνόλου σημείων $T \subseteq P$ αποτελούμενο από τρία μη συνευθειακά σημεία, υπάρχει ένας μοναδικός κύκλος ορθογώνιος με όλους τους κύκλους του T . Το κέντρο του έχει ίσες δυνάμεις ως προς τους τρεις κύκλους του T και ονομάζεται ριζικό κέντρο τους (*radical center*) [19]. Το σταθμισμένο σημείο που αντιστοιχεί στο ριζικό κέντρο συμβολίζεται με c_T .

Θεωρούμε ένα 2-άπλοκο (τρίγωνο) $\sigma_T = \text{conv}(T)$, το κυρτό περίβλημα του T . Το άπλοκο αυτό λέγεται κανονικό εάν

$$\pi(p, c_T) = 0 \quad \text{για όλα τα } p \in T, \quad (3.14)$$

$$\pi(p, c_T) > 0 \quad \text{για όλα τα } p \in P \setminus T, \quad (3.15)$$

όπου η εξίσωση (3.14) ισοδυναμεί με το c_T να είναι ορθογώνιο με όλα τα σημεία του T . Η συλλογή \mathcal{R} όλων των κανονικών τριγώνων που δημιουργούνται από τα σημεία του P



Εικόνα 3.4: Μέγεθος του τριγώνου ενός συνόλου 3 σημείων $\{p, q, r\}$, ως το τετράγωνο της ακτίνας του κύκλου (σταθμισμένου σημείου) c_T που είναι ορθογώνιος ως προς τους κύκλους p, q, r , σε (α) μια κανονική και (β) μια Delaunay τριγωνοποίηση. (γ) Στα ανισοτροπικά σταθμισμένα α-σχήματα, ο χώρος μετασχηματίζεται σε ισοτροπικός προτού μετρηθεί το μέγεθος του τριγώνου.

λέγεται κανονική τριγωνοποίηση του P . Ο ορισμός αυτός είναι ισοδύναμος με αυτόν της ενότητας 3.2.1.

Αν μηδενίσουμε τα βάρη όλων των σημείων, θέσουμε, δηλαδή, $w(p) = 0$ για όλα τα $p \in T$, το σταθμισμένο σημείο c_T μετατρέπεται στον περιγεγραμμένο κύκλο του τριγώνου σ_T . Αντίστοιχα, η συνάρτηση υπολογισμού της δύναμης σημείου (3.13) μετατρέπεται στην τετραγωνισμένη Ευκλείδεια απόσταση. Σε αυτή τη περίπτωση όπου δεν υπάρχουν βάρη στα σημεία, η τριγωνοποίηση γίνεται *Delaunay*, με τις ιδιότητες που περιγράφηκαν στην ενότητα 3.2.2.

Η συλλογή \mathcal{K} όλων των 2-άπλοκων (τριγώνων) καθώς και των όψεών τους (ακμές και σημεία) στο \mathcal{R} είναι ένα απλοειδές σύμπλοκο. Αν ορίσουμε ένα μέγεθος $\rho_T \geq 0$ για κάθε άπλοκο $\sigma_T \in \mathcal{K}$, τότε το σταθμισμένο α -σύμπλοκο του P είναι το υποσύνολο του \mathcal{K} που περιλαμβάνει όλα τα άπλοκα μέχρι ένα δεδομένο μέγεθος $\alpha \geq 0$,

$$\mathcal{K}_\alpha = \{\sigma_T \in \mathcal{K} : \rho_T < \alpha\}. \quad (3.16)$$

Τέλος, το σταθμισμένο α -σχήμα του P [27] είναι η ένωση όλων αυτών των άπλοκων,

$$\mathcal{W}_\alpha = \bigcup_{\sigma \in \mathcal{K}_\alpha} \sigma. \quad (3.17)$$

Το \mathcal{W}_α είναι ένα πολύτοπο που δεν είναι ούτε κυρτό, ούτε και συνεκτικό στη γενική περίπτωση. Στην ειδική περίπτωση όπου $\alpha = +\infty$, το $\mathcal{W}_{+\infty}$ αντιστοιχεί στο κυρτό περίβλημα του P

$$\mathcal{W}_{+\infty} = \bigcup_{\sigma \in \mathcal{R}} \sigma = \text{conv}(P). \quad (3.18)$$

Η διαφορά μεταξύ του α -σχήματος και του α -σύμπλοκου είναι ανάλογη της διαφοράς μεταξύ του κυρτού περιβλήματος και της τριγωνοποίησης ενός συνόλου σημείων.

Για την κατασκευή των α -σχημάτων, θα πρέπει να ορίσουμε το μέγεθος ρ_T ενός τυχαίου σημείου, μιας ακμής και ενός τριγώνου σ_T , όταν το T περιέχει 1, 2 ή 3 σημεία του P , αντίστοιχα. Το μέγεθος ενός σημείου $p \in P$ είναι πάντα ίσο με μηδέν. Για τις ακμές και τα τρίγωνα, χρησιμοποιούμε τους ορθογώνιους κύκλους, όπως φαίνεται στις Εικόνες 3.3 και 3.4.

Δεδομένου ενός συνόλου δύο σημείων $T = \{p, q\} \subseteq P$, υπάρχει μια άπειρη συλλογή κύκλων (*pencil*) που είναι ορθογώνιοι στους κύκλους p, q . Τα κέντρα τους βρίσκονται επί του ριζικού άξονα (*radical axis*) [19], που είναι μια γραμμή κάθετη στην ακμή $\sigma_T = \sigma_{\{p, q\}}$ της οποίας τα σημεία έχουν ίσες δυνάμεις ως προς τους κύκλους p, q . Από αυτή την άπειρη συλλογή, επιλέγουμε τον κύκλο με την μικρότερη ακτίνα που ικανοποιεί την εξίσωση 3.15. Το μέγεθος ρ_T του ευθύγραμμου τμήματος σ_T ισούται με το τετράγωνο της ακτίνας του κύκλου c_T . Στην περίπτωση που δεν επηρεάζουν την επιλογή του c_T άλλα σημεία πέρα από τα p, q , το κέντρο του θα βρίσκεται επί του ευθύγραμμου τμήματος σ_T , όπως φαίνεται στην Εικόνα 3.3a. Στην περίπτωση που η τριγωνοποίηση είναι η Delaunay (μηδενικά βάρη), το σ_T είναι χορδή του c_T που περνά από το κέντρο του (διάμετρος), και το μέγεθος ρ_T ισούται με το τετράγωνο της απόστασης μεταξύ p, q (βλ. Εικόνα 3.3β).

Αντίστοιχα, στην περίπτωση του τριγώνου έχουμε τρία μη συνευθειακά σημεία $T = \{p, q, r\} \subseteq P$, για τα οποία υπάρχει ένας μοναδικός κύκλος c_T ορθογώνιος στα p, q, r . Το μέγεθος ρ_T του τριγώνου θα ισούται με το τετράγωνο της ακτίνας του κύκλου c_T , όπως φαίνεται και στην Εικόνα 3.4a. Στην περίπτωση Delaunay, ο κύκλος c_T είναι ο περιγεγραμμένος κύκλος του τριγώνου σ_T (βλ. Εικόνα 3.4β). Δεδομένου ότι ο κύκλος c_T είναι ορθογώνιος και ως προς τα τρία σημεία p, q, r , συνεπάγεται ότι θα είναι και ως προς τα τρία ζευγάρια σημείων που αντιστοιχούν στις ακμές του. Άρα, εξ ορισμού ένα τρίγωνο θα έχει μέγεθος μεγαλύτερο ή ίσο από τα μεγέθη των ακμών του, όπως παρατηρούμε συγκρινούντας τις Εικόνες 3.3a,β με τις Εικόνες 3.4a,β.

Από τον ορισμό της συνάρτησης βάρους (βλ. εξίσωση 3.1) συμπεραίνουμε ότι το βάρος $w(p)$ κάθε σημείου p παίρνει τιμές στο διάστημα $[0, (s/2)^2]$, όπου s είναι το διάστημα δειγματοληψίας. Κατ' επέκταση, η ακτίνα του κύκλου p παίρνει τιμές στο διάστημα $[0, s/2]$. Το γεγονός αυτό διασφαλίζει ότι για δύο συνεχόμενα σημεία $T = \{p, q\}$ που έχουν ληφθεί επί μιας ακμής με βήμα s , οι κύκλοι που τους αντιστοιχούν δεν επικαλύπτονται, όπως φαίνεται στην Εικόνα 3.3a. Έτσι, εξασφαλίζεται ότι δε μπορεί ένας κύκλος να περικλείεται πλήρως σε έναν άλλο, με αποτέλεσμα να διατηρούνται όλα τα σημεία της δειγματοληψίας ως κορυφές στην τριγωνοποίηση. Στην περίπτωση που οι δύο κύκλοι εφάπτονται, το μέγεθος της αντίστοιχης ακμής σ_T γίνεται ίσο με μηδέν.

Ο ορισμός των σταθμισμένων α -σχημάτων που δόθηκε παραπάνω μπορεί να εφαρμο-

στεί στην κανονική τριγωνοποίηση, τη Delaunay, καθώς και τη Delaunay με περιορισμούς. Στην περίπτωση Delaunay, απλά θέτουμε όλα τα βάρη ίσα με μηδέν. Στην περίπτωση Delaunay με περιορισμούς δε μας αρκεί να θέσουμε μόνο τα βάρη ίσα με μηδέν. Θέτουμε, επιπλέον, τα μεγέθη των ακμών από περιορισμούς ίσα με μηδέν. Με αυτό τον τρόπο διασφαλίζουμε αυστηρά ότι όλες οι ακμές της τριγωνοποίησης που αντιστοιχούν σε ακμές της εικόνας θα διατηρούν τα όρια μεταξύ γειτονικών περιοχών, όπως θα δούμε στην ενότητα 3.4.1.

3.3.2 Ανισοτροπικά σταθμισμένα α-σχήματα

Τα ανισοτροπικά σταθμισμένα α-σχήματα που προτείνουμε, βασίζονται στα σταθμισμένα α-σχήματα και την κανονική τριγωνοποίηση. Χρησιμοποιούν σταθμισμένα σημεία, σε συνδυασμό με τον μετρικό τανυστή M (βλ. εξίσωση 3.11) που εξάγεται κατά τη διαδικασία της ανομοιόμορφης δειγματοληψίας με προσαρμογή στο τοπικό σχήμα. Για την εφαρμογή στην περίπτωση ομοιόμορφης δειγματοληψίας απαιτείται η εκ των υστέρων εκτέλεση της διαδικασίας προσαρμογής σχήματος για τον υπολογισμό του M σε κάθε δείγμα.

Αρχικά κατασκευάζουμε μια κανονική τριγωνοποίηση χρησιμοποιώντας τα σταθμισμένα σημεία όπως περιγράφηκε στην ενότητα 3.3.1. Δεδομένου του μετρικού τανυστή M_p που αντιστοιχεί σε κάθε σημείο p , υπολογίζουμε έναν μετρικό τανυστή M_T για κάθε άπλοκο T της τριγωνοποίησης. Συγκεκριμένα, ορίζουμε τον τανυστή M_T ενός 1- ή 2-άπλοκου (ακμής ή τριγώνου) T ως τη μέση τιμή των μετρικών τανυστών των σημείων T που τα απαρτίζουν.

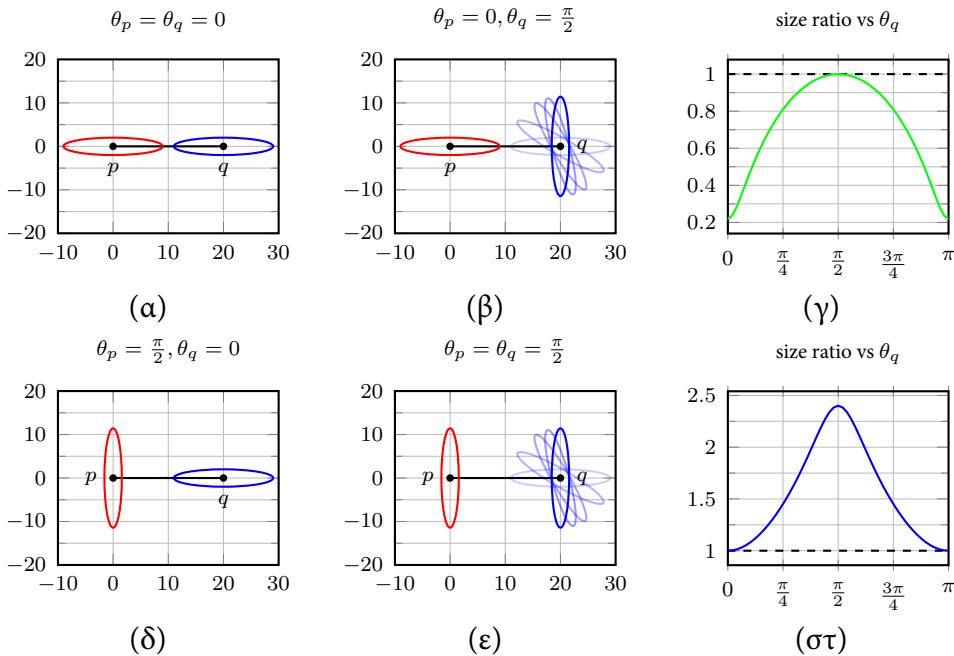
Όπως στην περίπτωση των απλών σημείων, έτσι και στις ακμές και τα τρίγωνα κανονικοποιούμε τον τανυστή ώστε να έχει μοναδιαία ορίζουνσα. Με αυτό τον τρόπο διασφαλίζουμε ότι το εμβαδό ενός σχήματος που παραμορφώνεται από την εφαρμογή του τανυστή θα παραμείνει αμετάβλητο. Έτσι, παρά το γεγονός ότι οι κύκλοι των σταθμισμάτων σημείων θα μετατραπούν σε ελλείψεις, ή οι ελλείψεις στην περίπτωση της ανομοιόμορφης δειγματοληψίας θα μετασχηματιστούν σε άλλες ελλείψεις, το εμβαδό τους (που αντιστοιχεί στη δύναμή τους) δε θα μεταβληθεί.

Δεδομένων τριών σημείων p, q, r της τριγωνοποίησης, με μετρικούς τανυστές M_p, M_q, M_r αντίστοιχα, ορίζουμε τον τανυστή

$$M_T = \frac{M_p + M_q}{\sqrt{\det(M_p + M_q)}} \quad (3.19)$$

για την ακμή που ορίζεται από τα $T = \{p, q\}$, και τον τανυστή

$$M_T = \frac{M_p + M_q + M_r}{\sqrt{\det(M_p + M_q + M_r)}} \quad (3.20)$$



Εικόνα 3.5: Μέτρηση του μεγέθους ρΤ μιας ακμής στην ανισοτροπική περίπτωση, για διαφορετικές γωνίες θ_p, θ_q των αντίστοιχων ελλείψεων, όπου η θ_p είναι σταθερή και μηδενική (οριζόντια) στην πρώτη γραμμή (α, β) ή ίση με $\pi/2$ (κατακόρυφη) στη δεύτερη γραμμή (δ, ε). Στη δεξιά στήλη (γ, σ) βλέπουμε τον λόγο του ρΤ ως προς την αντίστοιχη μέτρηση στην ισοτροπική περίπτωση, καθώς μεταβάλλεται η γωνία θ_q .

για το τρίγωνο που ορίζεται από τα $T = \{p, q, r\}$.

Στην Εικόνα 3.5 βλέπουμε ένα παράδειγμα εφαρμογής της εξίσωσης 3.19 στην ακμή που ορίζουν τα σημεία p και q για διαφορετικές γωνίες μεταξύ των αντίστοιχων ελλείψεων. Το μέγεθος που μετράμε στα γραφήματα (γ) και (σ) αντιστοιχεί στο λόγο του μεγέθους της ακμής που προκύπτει μετά την εφαρμογή του τανυστή, προς το αρχικό μέγεθος που αντιστοιχεί στην ισοτροπική περίπτωση. Όταν οι μεγάλοι άξονες των ελλείψεων είναι ευθυγραμμισμένοι, όπως στην Εικόνα 3.5α, το μέγεθος της ακμής μειώνεται (β . $\theta_q = 0$ και $\theta_q = \pi$ στην Εικόνα 3.5γ). Σε αυτή την περίπτωση ισοδυναμεί το να μετακινήσουμε τα σημεία p, q πιο κοντά μεταξύ τους. Στην αντίθετη περίπτωση, όπου οι μεγάλοι άξονες των ελλείψεων είναι παράλληλοι μεταξύ τους (αλλά όχι ευθυγραμμισμένοι), όπως στην Εικόνα 3.5ε, το μέγεθος της ακμής αυξάνεται, σαν να απομακρύνουμε τα σημεία μεταξύ τους (β . $\theta_q = \pi/2$ στο διάγραμμα 3.5στ). Τέλος, στην περίπτωση που οι μεγάλοι άξονες των ελλείψεων είναι κάθετοι μεταξύ τους, όπως στις Εικόνες 3.5β, δ οι τανυστές αλληλοαναιρούνται, με αποτέλεσμα ο τανυστής που προκύπτει να είναι ο μοναδιαίος πίνακας και το μέγεθος της ακμής να μην μεταβάλλεται (β . $\theta_q = \pi/2$ στην Εικόνα 3.5γ, και $\theta_q = 0, \theta_q = \pi$ στην Εικόνα 3.5στ).

Μπορούμε να δούμε την επίδραση του μετρικού τανυστή M_T ως την εφαρμογή ενός γραμμικού μετασχηματισμού στο χώρο του άπλοκου T , που δίνεται από έναν πίνακα τοπι-

κού σχήματος U , τέτοιο ώστε $M_T = U^T U$. Ο μετασχηματισμός αυτός μετατρέπει το χώρο σε “ισοτροπικό”, όπου οι ελλείψεις μετατρέπονται σε κύκλους, όπως στις Εικόνες 3.3γ και 3.4γ, πριν τον υπολογισμό του μεγέθους των άπλοκων. Ισοδύναμα, μπορούμε να θεωρήσουμε το εσωτερικό γινόμενο που ορίζεται τοπικά ως $\langle x, y \rangle_T = \langle Ux, Uy \rangle = x^T M_T y$ για $x, y \in \mathbb{R}^2$, και κατ’ επέκταση μια τοπική συνάρτηση απόστασης d_T που ορίζεται ως

$$d_T(x, y) = \sqrt{(x - y)^T M_T (x - y)} \quad (3.21)$$

για $x, y \in \mathbb{R}^2$. Χρησιμοποιύμε τη συνάρτηση απόστασης d_T για να υπολογίσουμε όλες τις αποστάσεις μεταξύ δύο ή τριών σημείων σε μια ακμή ή τρίγωνο T . Οι αποστάσεις αντιστοιχούν στις Ευκλείδειες αποστάσεις μετρημένες στο μετασχηματισμένο χώρο των Εικόνων 3.3γ, 3.4γ, αντίστοιχα. Δεδομένων των αποστάσεων και των βαρών των σημείων, μετράμε τα μεγέθη των άπλοκων με τον ίδιο τρόπο όπως και στην ισοτροπική περίπτωση.

Σύγκριση με βιβλιογραφία

Η προτεινόμενη μέθοδος για χρήση ανισοτροπικών μετρικών στα α-σχήματα, διαφέρει από τη μέθοδο των Teichmann και Capps [107], οι οποίοι χρησιμοποιούν μόνο την πληροφορία της περιστροφής σε κάθε σημείο χωρίς καμία κλίμακα. Ο λόγος των αξόνων της έλλειψης είναι σταθερός, ενώ η συνολική κλίμακα (που καθορίζει το εμβαδό της έλλειψης) είναι εξωτερική παράμετρος που χρειάζεται βελτιστοποίηση για να προσδιοριστεί. Αντίθετα, με την προτεινόμενη μέθοδο εκμεταλλευόμαστε όλη την πληροφορία του μετρικού τανυστή, όπως εξάγεται από τη διαδικασία τοπικής προσαρμογής σχήματος, η οποία αντιστοιχεί σε γενικότερους γραμμικούς μετασχηματισμούς.

Η τριγωνοποίηση που κατασκευάζουν οι Teichmann και Capps [107] ξεκινά ως Delaunay, χωρίς καμμία πληροφορία βάρους ή κλίμακας από τα σημεία. Εξετάζοντας κάθε τρίγωνο, υπολογίζουν το μέσο όρο των κατευθύνσεων των ελλείψεων των τριών κορυφών. Έπειτα μετακινούν τα σημεία σύμφωνα με την μέση κατεύθυνση όπως περιγράψαμε παραπάνω, αντί να παραμορφώσουν το χώρο. Η διαφορά με την προτεινόμενη μέθοδο έγκειται και πάλι στην επιπλέον πληροφορία που χρησιμοποιούμε από κάθε σημείο. Η τριγωνοποίηση που χρησιμοποιείται στο [107] μεταβάλλεται τοπικά με αναστροφές ακμών ώστε να ικανοποιούνται οι συνθήκες Delaunay των μετατοπισμένων σημείων. Για να αποφευχθούν οι ατέρμονοι κύκλοι ανατροπών, μια ακμή επιτρέπεται να ανατραπεί μόνο μια φορά. Το γεγονός ότι η σειρά επίσκεψης των τριγώνων της αρχικής Delaunay τριγωνοποίησης επηρεάζει το τελικό αποτέλεσμα, κάνει τον αλγόριθμο μη ντετερμινιστικό. Επίσης η διαδικασία αυξάνει το κόστος κατασκευής των α-σχημάτων. Στην προτεινόμενη μέθοδο κατασκευάζουμε μια κανονική τριγωνοποίηση την οποία δε μεταβάλουμε. Με αυτό τον τρόπο επιτυγχάνουμε αποτελέσματα συγκρίσιμα με των Teichmann και Capps [107], χωρίς να χρειάζεται να μεταβάλουμε την τριγωνοποίηση.

Οι Labelle και Shewchuk [55] προτείνουν την κατασκευή μιας ανισοτροπικής τριγωνοποίησης που έχει πιο αυστηρές εγγυήσεις για τις ιδιότητες των τριγώνων. Ξεκινούν από την κατασκευή ενός ανισοτροπικού διαγράμματος Voronoi, χρησιμοποιώντας ένα μετρικό τανυστή σε κάθε σημείο, που ορίζει μια τοπική μετρική απόστασης ανά κελί, όπως στην εξίσωση 3.21. Για να μπορέσει το διάγραμμα Voronoi να αντιστοιχηθεί με μια τριγωνοποίηση, θα πρέπει να χαλαρώσει ο ορισμός του ώστε να μην υπάρχουν ορφανές περιοχές (περιοχές που δεν περικλείουν την εστία στην οποία ανήκουν). Η χρονική πολυπλοκότητα της κατασκευής ενός ανισοτροπικού διαγράμματος Voronoi είναι $O(n^{2+\epsilon})$, όπου το ϵ είναι μια θετική σταθερά, γεγονός που την καθιστά χρονικά απαγορευτική.

3.4 Ανίχνευση χαρακτηριστικών

Στις προηγούμενες ενότητες παρουσιάσαμε τις μεθόδους τριγωνοποίησης ενός δεδομένου συνόλου σημείων, καθώς και την κατασκευή των α -σχημάτων (απλών, σταθμισμένων ή ανισοτροπικά σταθμισμένων). Σε αυτή την ενότητα θα περιγράψουμε τη διαδικασία ομαδοποίησης και παρακολούθησης της εξέλιξης των α -σχημάτων, καθώς και της επιλογής των τελικών χαρακτηριστικών.

3.4.1 Συνεκτικές συνιστώσες και γειτνίαση

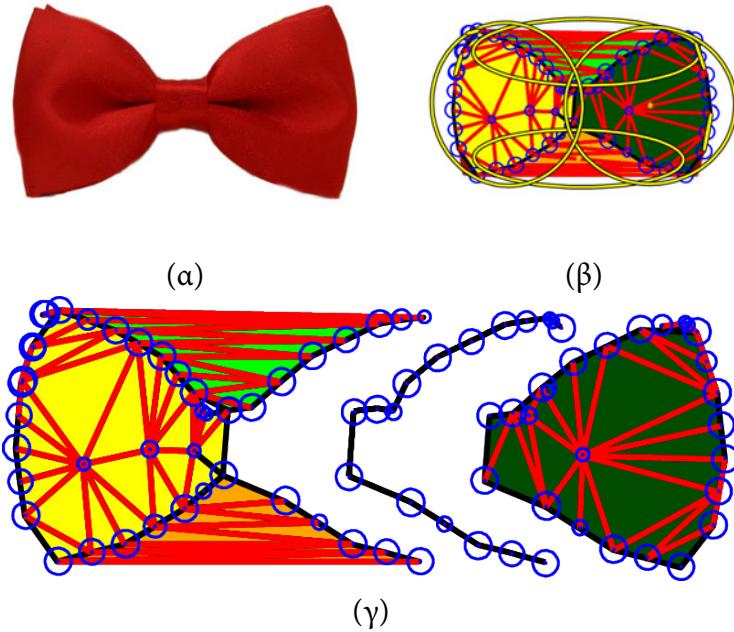
Όλα τα άπλοκα του συμπλόκου απλοειδών \mathcal{K} (βλ. ενότητα 3.3), τυπικά διατάσσονται σε αύξουσα σειρά ως προς το μέγεθός τους, ώστε να κατασκευαστεί η σταθμισμένη διήθηση [27]. Στην παρούσα εργασία αποκλίνουμε από αυτή τη συνήθη πρακτική με δύο τρόπους. Πρώτον, χρησιμοποιούμε μόνο τα άπλοκα διάστασης 1 και 2 (ακμές και τρίγωνα), απορρίπτοντας τα σημεία. Με αυτό τον τρόπο, κατασκευάζουμε το σύμπλοκο

$$\mathcal{K}' = \{\sigma_T \in \mathcal{K} : |T| \geq 2\}. \quad (3.22)$$

Στην επιλογή αυτή μας οδηγεί το γεγονός ότι το μέγεθος των ακμών ελέγχει την συνεκτικότητα μεταξύ τριγώνων όπως θα δούμε παρακάτω, ενώ τα σημεία δεν την επηρεάζουν. Δεύτερον, σε αντίθεση με την κατασκευή που ορίστηκε από την (3.16), επιλέγουμε το άνω α -σύμπλοκο

$$\bar{\mathcal{K}}_\alpha = \mathcal{K}' \setminus \mathcal{K}_\alpha = \{\sigma_T \in \mathcal{K}' : \rho_T \geq \alpha\} \quad (3.23)$$

όπου η διάταξη των άπλοκων γίνεται με φθίνουσα σειρά μεγέθους. Όπως και στην εργασία [27], θα χρειαστεί να εξετάσουμε μόνο ένα πεπερασμένο σύνολο από τιμές για το α . Συγκεκριμένα, ταξινομούμε όλα τα άπλοκα $\sigma_T \in \mathcal{K}'$ με φθίνουσα σειρά μεγέθους ρ_T , δημιουργώντας την ακολουθία $(\sigma_1, \dots, \sigma_n)$, όπου $n = |\mathcal{K}'|$. Αν ρ_i είναι το μέγεθος του σ_i για $i = 1, \dots, n$, τότε $\rho_1 \geq \dots \geq \rho_n$. Ακόμα, αν $K_i = \{\sigma_1, \dots, \sigma_i\}$, δημιουργείται η άνω



Εικόνα 3.6: Παράδειγμα της σχέσης γειτνίασης μεταξύ άπλοκων. Για μια χαμηλή τιμή του α βλέπουμε τις διαφορετικές συνεκτικές συνιστώσες της τριγωνοποίησης (β), για μια εικόνα εισόδου (α). Οι ακμές διατηρούν ασύνδετες τις συνιστώσες μέχρις ότου συμπεριληφθούν και οι ίδιες στο α -σχήμα (γ). Τα τελικά χαρακτηριστικά που ανιχνεύονται φαίνονται στο (β).

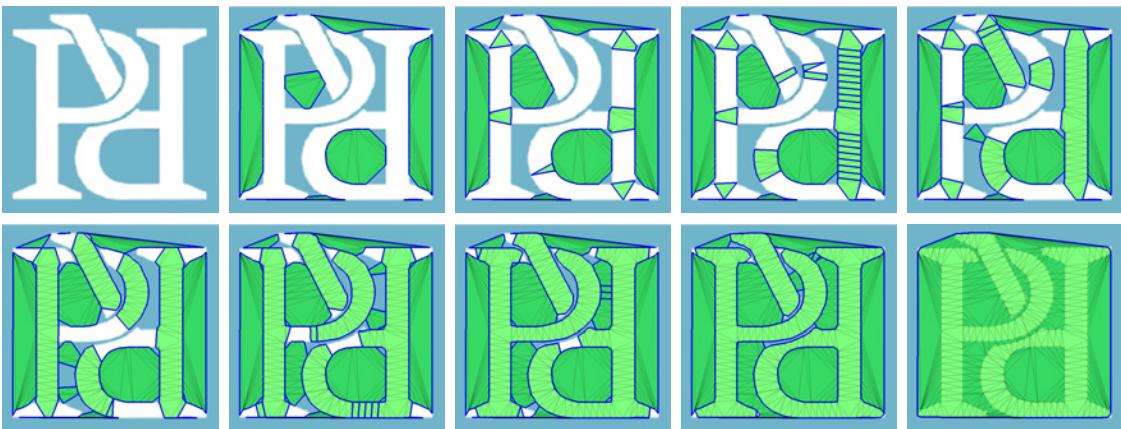
α -διήθηση

$$\emptyset = K_0 \subseteq \dots \subseteq K_n = \mathcal{K}'. \quad (3.24)$$

Ξεκινώντας από το μεγαλύτερο άπλοκο σ_1 μεγέθους ρ_1 και μειώνοντας την τιμή του α μέχρι το ρ_n , το άνω α -σύμπλοκο μοντελοποιεί την ανάπτυξη κοιλοτήτων στο σχήμα μας. Για να συλλάβουμε την εξέλιξη της τοπολογίας του σχήματος, κατασκευάζουμε ένα δέντρο συνιστωσών, όπως και στο [81].

Για να ορίσουμε την συνδετικότητα επί των άπλοκων του συμπλόκου, θα πρέπει να ορίσουμε τη σχέση γειτνίασης μεταξύ άπλοκων. Ορίζουμε ως γείτονες ενός τριγώνου $\sigma_T \in \mathcal{K}'$ (με $|T| = 3$) τις τρεις ακμές του, ενώ ως γείτονες μιας ακμής σ_T (με $|T| = 2$) τα δύο τρίγωνα της τριγωνοποίησης στα οποία ανήκει. Συμβολίζουμε τη γειτονιά ενός άπλοκου $\sigma \in \mathcal{K}'$ με $N(\sigma)$. Δεδομένης της φθίνουσας σειράς ως προς το μέγεθος, και πως μια ακμή σε μια κανονική τριγωνοποίηση δε μπορεί να έχει μέγεθος μεγαλύτερο από το μέγεθος των δυο γειτονικών της τριγώνων, η διαισθητική περιγραφή του συστήματος γειτνίασης που προτείνουμε είναι πως μια ακμή θα διατηρεί τα γειτονικά της τρίγωνα ασύνδετα, μέχρι να μειωθεί η τιμή του α ώστε να συμπεριληφθεί και η ακμή στο σχήμα (βλ. Εικόνα 3.6). Τελικά, ο χρόνος επίσκεψης κάθε ακμής και τριγώνου εξαρτάται από το τοπικό σχήμα στο άπλοκο που εξετάζουμε, καθώς και από την κλίση της φωτεινότητας της εικόνας.

Δεδομένου του συστήματος γειτνίασης που περιγράψαμε, ο αλγόριθμος ξεκινά θεωρώντας όλα τα άπλοκα του \mathcal{K}' ως ξεχωριστές συνιστώσες. Επισκεπτόμαστε κάθε άπλοκο



Εικόνα 3.7: Παράδειγμα της άνω α -διήθησης. Από επάνω αριστερά προς τα δεξιά και έπειτα στη δεύτερη σειρά το α μειώνεται, εξάγοντας διαφορετικά στιγμιότυπα της άνω α -διήθησης. Οι χρωματισμένες μπλε ακμές δεν έχουν επεξεργαστεί για την τρέχουσα τιμή του α (δεν ανήκουν στο α -σύμπλοκο), διατηρώντας τις συνιστώσες ασύνδετες.

με φθίνουσα σειρά μεγέθους, ενώνοντάς το με τα γειτονικά του άπλοκα τα οποία έχουμε επισκεφθεί ήδη. Έτσι, τα τρίγωνα της τριγωνοποίησης θα σχηματίζουν ξεχωριστές συνδεδεμένες συνιστώσες, μέχρι να επισκεφθούμε τις μεταξύ τους ακμές, οι οποίες θα τα ενώνουν σε συνδεδεμένες συνιστώσες. Η διαδικασία περιγράφεται στον Αλγόριθμο 1.

Για κάθε συνιστώσα κ_T στο δέντρο συνιστωσών, ορίζουμε ένα μέγεθος ρ_T . Το μέγεθος αυτό είναι ίσο με το μέγεθος του τελευταίου άπλοκου σ_T που προστέθηκε στη συνιστώσα (το μικρότερο που ανήκει στη συνιστώσα) και προκάλεσε μεταβολή του εμβαδού της. Σε περίπτωση που ένα άπλοκο προστίθεται σε μια συνιστώσα χωρίς να μεταβάλλεται το εμβαδό της, τότε το άπλοκο είναι εσωτερική ακμή της συνιστώσας και δε μεταβάλει το μέγεθος ρ_T . Εναλλακτικά, το μέγεθος της συνιστώσας ισούται με το μέγεθος του πρώτου άπλοκου (ακμής) που προστέθηκε στη συνιστώσα και προκάλεσε τη συνένωση δύο ξένων μεταξύ τους συνιστωσών.

Στην Εικόνα 3.7 βλέπουμε ένα παράδειγμα της άνω α -διήθησης, σε συνδυασμό με το σύστημα γειτνίασης που προτείνουμε. Για $\alpha = \infty$ ξεκινάμε με το κενό σύνολο. Μειώνοντας την τιμή του α , προστίθενται τρίγωνα και ακμές στο άνω α -σύμπλοκο, και για $\alpha = 0$ το α -σχήμα ταυτίζεται με το κυρτό περιβλήμα των σημείων εισόδου. Σε κάθε περίπτωση, οι ακμές που είναι χρωματισμένες μπλε δεν ανήκουν στο α -σύμπλοκο, αλλά σχηματίζουν το σύνορό του. Χρησιμοποιώντας το σύστημα γειτνίασης που προτείνουμε, οι ακμές αυτές αποτρέπουν την ένωση των ασύνδετων συνιστωσών. Με τον τρόπο αυτό οι συνιστώσες παραμένουν ασύνδετες για περισσότερες τιμές του α , περιγράφοντας με μεγαλύτερη ακρίβεια τα τμήματα των αντικειμένων της εικόνας.

3.4.2 Επιλογή χαρακτηριστικών

Ξεκινώντας από μεγάλες τιμές του α και παρακολουθώντας την εξέλιξη του άνω α-σύμπλοκου από το κενό σύνολο ως το κυρτό περίβλημα της τριγωνοποίησης για $\alpha = 0$ (που καλύπτει το σύνολο της εικόνας), σχηματίζονται διαφορετικές συνδεδεμένες συνιστώσες. Οι συνιστώσες αυτές εμφανίζονται πάνω από περιοχές της εικόνας με διακριτά όρια. Η σημαντικότητα μιας συνιστώσας, που μετράται επί της α -διήθησης, εξαρτάται από τη σταθερότητα της αντίστοιχης περιοχής της εικόνας

Για να μετρήσουμε τη σημαντικότητα, υπολογίζουμε ένα μέτρο σημαντικότητας για κάθε συνιστώσα. Συγκεκριμένα, θεωρούμε τη συνιστώσα κ_U μεγέθους ρ_U , που είναι ένα σύνολο άπλοκων (τριγώνων και ακμών) μεγέθους μεγαλύτερου ή ίσου με ρ_U . Όταν η συνιστώσα κ_U ενώνεται μέσω μιας ακμής σ_T που βρίσκεται στο σύνορό της και έχει μέγεθος ρ_T με μια άλλη συνιστώσα (έστω $\kappa_{U'}$) για να σχηματίσουν μια νέα συνιστώσα κ_T , τότε υπολογίζουμε τη σημαντικότητα της κ_U (όπως επίσης και της $\kappa_{U'}$). Μπορούμε να δούμε το κ_U ως μια συνδεδεμένη συνιστώσα από τρίγωνα και ακμές της τριγωνοποίησης που η περιοχή της εικόνας που της αντίστοιχεί περικλείεται από ακμές της εικόνας και το μήκος του μεγαλύτερου ανοίγματος επί του συνόρου της είναι ίσο με $\sqrt{\rho_T}$.

Επιλέγουμε να εξετάσουμε μέτρα σημαντικότητας που απαιτούν το ελάχιστο δυνατό υπολογιστικό κόστος, εστιάζοντας στις πληροφορίες που έχουμε ήδη διαθέσιμες στο δέντρο συνιστωσών, ή μπορούμε να αποκτήσουμε με απλές αριθμητικές πράξεις από τα άπλοκα και την τριγωνοποίηση. Εξετάζουμε διαφορετικά μέτρα σημαντικότητας που αξιοποιούν διαφορετικές οπτικές και γεωμετρικές ιδιότητες των υποψήφιων περιοχών της εικόνας.

Το πρώτο μέτρο, βασίζεται στο εμβαδό μιας συνιστώσας, όπως αυτό συγκρίνεται με το μεγαλύτερο άνοιγμα επί του συνόρου της:

$$s(\kappa_U) = \frac{a(\kappa_U)}{\rho_T}, \quad (3.25)$$

όπου $a(\kappa_U)$ είναι το εμβαδό της συνιστώσας κ_U · για την ακρίβεια, είναι το είναι το άθροισμα των εμβαδών όλων των άπλοκων που ανήκουν στο κ_U . Δεδομένου ότι επισκεπτόμαστε τα άπλοκα με φθίνουσα σειρά μεγέθους, το ρ_T αντίστοιχεί, όπως έχουμε δει, στο μεγαλύτερο άνοιγμα επί του συνόρου της συνιστώσας κ_U . Αυτό το μέτρο σημαντικότητας, από κατασκευής ευνοεί τις μεγάλες συνιστώσες με μικρά (ή καθόλου) ανοίγματα στο σύνορό τους, οπότε το ονομάζουμε κλειστότητα. Το μέγεθος ρ ενός άπλοκου μετρά μήκος στο τετράγωνο (επιφάνεια), άρα το μέτρο αυτό είναι αδιάστατο. Για να επιλέξουμε μια συνιστώσα σαν τοπικό χαρακτηριστικό, συγκρίνουμε τη σημαντικότητά της με ένα κατώφλι τ .

Ένα δεύτερο μέτρο σημαντικότητας που είναι υπολογιστικά φθηνό, σχετίζεται με τη διάρκεια ζωής κάθε συνιστώσας. Η διάρκεια ζωής μετριέται ως η διαφορά μεταξύ του μεγέ-

θους ρ_U της συνιστώσας και του μεγέθους ρ_T της ακμής που προστίθεται στη συνιστώσα και σχηματίζει τη νέα συνιστώσα κ_T . Αν η προστιθέμενη ακμή είναι εσωτερική, το σχήμα και το εμβαδό της συνιστώσας παραμένει αμετάβλητο, οπότε το μέτρο δεν υπολογίζεται.

Το μέτρο σημαντικότητας διάρκειας ζωής ορίζεται ως

$$s(\kappa_U) = \rho_U - \rho_T, \quad (3.26)$$

το οποίο, λόγω της φθίνουσας σειράς μεγέθους, είναι πάντα θετικό. Και σε αυτή την περίπτωση, οι συνιστώσες με διάρκεια ζωής μεγαλύτερη από ένα κατώφλι τ επιλέγονται ως χαρακτηριστικά.

Ένα τρίτο μέτρο που εξετάσαμε είναι αντίστοιχο με αυτό που χρησιμοποιείται στον ανιχνευτή MSER. Ο MSER δημιουργεί συνιστώσες κατευθείαν από τα εικονοστοιχεία, όπου η σειρά επίσκεψης των εικονοστοιχείων καθορίζεται από το επίπεδο φωτεινότητας. Το μέτρο σημαντικότητας που χρησιμοποιεί ορίζεται ως η σχετική αλλαγή του εμβαδού μιας συνιστώσας για μια δεδομένη αλλαγή της φωτεινότητας. Στην περίπτωσή μας, όταν η συνιστώσα κ_U ενώνεται με μια άλλη για τη δημιουργία της συνιστώσας κ_T , το μέτρο θα είναι

$$s(\kappa_U) = \frac{a(\kappa_U) - a(\kappa_T)}{a(\kappa_U)},$$

το οποίο είναι επίσης θετικό. Ωστόσο, καθώς οι αλλαγές στο εμβαδό των συνιστωσών στην περίπτωσή μας δεν είναι τόσο ομαλές όσο στην περίπτωση του MSER που δουλεύει με τα εικονοστοιχεία, το μέτρο είναι αρκετά θορυβώδες. Γι' αυτό το λόγο, προτείνουμε ένα μέτρο σημαντικότητας που συνδυάζει και τη διάρκεια ζωής με τα προηγούμενα:

$$s(\kappa_U) = (\rho_U - \rho_T) \frac{a(\kappa_U)}{a(\kappa_U) - a(\kappa_T)}. \quad (3.27)$$

Το μέτρο σημαντικότητας το ονομάζουμε **σταθερότητα**, καθώς από τη μορφή του (βλ. εξίσωση 3.27) βλέπουμε πως είναι αντιστρόφως ανάλογο προς το σχετικό ρυθμό αλλαγής του εμβαδού της συνιστώσας. Όπως και στον MSER, μια συνιστώσα είναι σταθερή αν το μέτρο της σημαντικότητας παρουσιάζει τοπικό μέγιστο στη συγκεκριμένη συνιστώσα καθώς διατρέχουμε τη διήθηση. Μια σταθερή συνιστώσα επιλέγεται σαν χαρακτηριστικό αν η τιμή της σημαντικότητας στο τοπικό μέγιστο ξεπερνά τις γειτονικές τιμές κατά τουλάχιστον ένα κατώφλι τ .

Σαν τελευταίο μέτρο σημαντικότητας εξετάζουμε ένα μέτρο ελλειπτικότητας, για να αποτιμήσουμε τη φυσική σύνδεση με το σχήμα των περιοχών που προσπαθούμε να ανιχνεύσουμε. Αρχικά υπολογίζουμε τις ροπές αδράνειας έως 2^{n_s} τάξης (δηλ. $m_{00}, m_{10}, m_{01}, m_{11}, m_{20}$ και m_{02}) για όλα τα τρίγωνα της τριγωνοποίησης. Οι ροπές υπολογίζονται αναλυτικά, χρησιμοποιώντας μόνο τις συντεταγμένες των τριών κορυφών κάθε τριγώνου. Έπειτα, κατά την κατασκευή του δέντρου συνιστωσών, υπολογίζουμε τις ροπές κάθε συνιστώσας με αυξητικό τρόπο, χρησιμοποιώντας τις ροπές των συνιστωσών που ενώνονται. Η διαδικασία απαιτεί μόνο μερικές προσθέσεις σε κάθε κόμβο του δέντρου.

Μετά την κατασκευή του δέντρου συνιστωσών, θα πρέπει να διατρέξουμε μια φορά όλο το δέντρο για να υπολογίσουμε τις κεντρικές ροπές μ σε κάθε κόμβο από τις ροπές m , όπως περιγράφεται στο [45]. Ταυτόχρονα, υπολογίζουμε τη πιο απλή αφινικά αμετάβλητη ποσότητα ροπών, όπως περιγράφεται στο [37]:

$$I_1 = \frac{\mu_{20}\mu_{02} - \mu_{11}^2}{\mu_{00}^4}.$$

Τελικά, η ελλειπτικότητα ενός σχήματος ορίζεται ως

$$s(\kappa_U) = \begin{cases} 16\pi^2 I_1 & \text{εάν } I_1 \leq \frac{1}{16\pi^2}, \\ \frac{1}{16\pi^2 I_1} & \text{αλλιώς.} \end{cases} \quad (3.28)$$

Αυτό το μέτρο παίρνει τιμές στο διάστημα $[0, 1]$, και μεγιστοποιείται για μια τέλεια έλλειψη (ή κύκλο) [97]. Η ελλειπτικότητα συγκρίνεται με ένα κατώφλι τ ώστε να επιλέξουμε μια συνιστώσα ως χαρακτηριστικό.

3.4.3 Αλγόριθμος

Ο ψευδοκώδικας ολόκληρης της προτεινόμενης μεθόδου φαίνεται στον Αλγόριθμο 1. Εεκινάμε με τη δειγματοληψία της εικόνας, που περιγράφεται στη Διεργασία Δειγματοληψία και την ενότητα 3.1. Η έξοδος της διεργασίας είναι ένα σύνολο δειγμάτων P με τις απαραίτητες πληροφορίες που χρειάζονται για τα παρακάτω βήματα του αλγορίθμου (βάρη, πίνακας τοπικού σχήματος M).

Τα δείγματα P αποτελούν τους κόμβους της τριγωνοποίησης που κατασκευάζεται στο επόμενο βήμα. Η τριγωνοποίηση \mathcal{R} μπορεί να είναι Delaunay, Delaunay με περιορισμούς ή κανονική, όπως περιγράφηκε στην ενότητα 3.2. Δεδομένης της τριγωνοποίησης, εξάγουμε το απλοειδές σύμπλοκο K και κατασκευάζουμε το χάρτη μεγέθους των συμπλόκων ρ . Για κάθε άπλοκο αποθηκεύουμε τους γείτονές του σε ένα χάρτη N , ενώ το σύμπλοκο K' κατασκευάζεται όπως περιγράφηκε από την εξίσωση 3.22.

Για τη χρήση ανισοτροπικά σταθμισμένων α-σχημάτων, η τριγωνοποίηση θα πρέπει να είναι κανονική, ενώ η δειγματοληψία θα πρέπει να είναι μη ομοιόμορφη, ώστε να εξάγουμε το μετρικό τανυστή για κάθε δείγμα. Όταν επιλέγουμε ανισοτροπικά σταθμισμένα α-σχήματα με ομοιόμορφη δειγματοληψία, θα πρέπει αφού εξάγουμε τα δείγματα από τις ακμές, να εκτελέσουμε τη διαδικασία προσαρμογής στο τοπικό σχήμα για κάθε δείγμα ώστε να εξάγουμε τους αντίστοιχους μετρικούς τανυστές. Συνολικά προκύπτουν οι συνδυασμοί τριγωνοποίησεων και α-σχημάτων που φαίνονται στον πίνακα 3.1. Σε κάθε τέτοιο συνδυασμό, μπορούμε να έχουμε είτε ομοιόμορφη δειγματοληψία (U), είτε ανομοιόμορφη (NU).

Για την παρακολούθηση των συνδεδεμένων συνιστωσών χρησιμοποιούμε δύο δομές, όπως στο [81]. Η πρώτη είναι ένα δάσος όπου κάθε κόμβος αντιστοιχεί σε ένα άπλοκο του

Αλγόριθμος 1: Ανίχνευση τοπικών χαρακτηριστικών WαSH

είσοδος: μονόχρωμη εικόνα f

έξοδος : σύνολο τοπικών χαρακτηριστικών F

- 1 $P \leftarrow \text{Δειγματοληψία}(f, m)$ ▷ δειγματοληψία
 - 2 $\mathcal{R} \leftarrow \text{Τριγωνοποίηση}(P; \text{Delaunay} \mid \text{με περιορισμούς} \mid \text{κανονική})$
 - 3 $(\mathcal{K}, \rho) \leftarrow \text{Σύμπλοκο}(\mathcal{R}; \text{ισοτροπικό} \mid \text{ανισοτροπικό})$ ▷ απλοειδές σύμπλοκο + μεγέθη
 - 4 $N \leftarrow \text{Γείτονες}(\mathcal{K}')$ ▷ σύστημα γειτνίασης
 - 5 $F \leftarrow \emptyset$
 - 6 **για κάθε** $\sigma_T \in \mathcal{K}'$ ▷ αρχικοποίηση κάθε απλοκου
 - 7 | Δημιουργία Συνόλου(σ_T) ▷ σαν ανεξάρτητη συνιστώσα
 - 8 | $\sigma_T.\rho i\zeta\alpha \leftarrow \sigma_T$
 - 9 **για κάθε** $\sigma_T \in \mathcal{K}'$ σε φθίνουσα σειρά ρ_T ▷ τρέχον απλοκό
 - 10 | $\kappa_T \leftarrow \text{Αναζήτηση}(\sigma_T)$ ▷ τρέχουσα συνιστώσα κ_T
 - 11 | $r_T \leftarrow \kappa_T.\rho i\zeta\alpha$
 - 12 **για κάθε** $\sigma_U \in N(\sigma_T)$ τ.ω. $\rho_U \geq \rho_T$ ▷ γειτονικό, που έχουμε ήδη επισκεφτεί
 - 13 | | $\kappa_U \leftarrow \text{Αναζήτηση}(\sigma_U)$ ▷ γειτονική συνιστώσα κ_U
 - 14 | | $r_U \leftarrow \kappa_U.\rho i\zeta\alpha$
 - 15 | | **εάν** $\kappa_T \neq \kappa_U$ **τότε** ▷ αν ανήκουν σε διαφορετικές συνιστώσες
 - 16 | | | **εάν** $|U| = 3 \wedge \Sigma_{\text{ημαντικότητα}}(U) > \tau$ **τότε** ▷ επιλέγουμε το κ_U αν είναι τρίγωνο ή σημαντικό
 - 17 | | | | $F \leftarrow F \cup r_U$ ▷ επιλέγουμε το κ_U αν είναι τρίγωνο ή σημαντικό
 - 18 | | | | $r_T.\text{ΠροσθήκηΠαιδιού}(r_U)$ ▷ το προσθέτουμε κάτω από το κ_T
 - 19 | | | | $r_T.\text{εμβαδό} \leftarrow r_T.\text{εμβαδό} + r_U.\text{εμβαδό}$ ▷ προσθέτουμε τα εμβαδά
 - 20 | | | | $\kappa_T \leftarrow \text{Ένωση}(\kappa_T, \kappa_U)$ ▷ ενώνουμε τα ξένα σύνολα
 - 21 | | | | $\kappa_T.\rho i\zeta\alpha \leftarrow r_T$
-

Διεργασία Δειγματοληψία(f, m)

είσοδος: μονόχρωμη εικόνα f ;

μέθοδος δειγματοληψίας m

έξοδος : σύνολο δειγμάτων P

- 1 $\hat{G} \leftarrow \|\nabla f\| / \max\{\|\nabla f\|\}$ ▷ κανονικοποιημένη κλίση
 - 2 $E \leftarrow \text{CANNY}(\hat{G})$ ▷ ανίχνευση ακμών
 - 3 **εάν** ($m == \text{ομοιόμορφη δειγματοληψία}$) **τότε** ▷ ομοιόμορφη δειγματοληψία
 - 4 | $P \leftarrow \text{Ομοιόμορφη Δειγματοληψία}(f, s)$ ▷ σταθερό βήμα s
 - 5 **αλλιώς** ▷ ανομοιόμορφη δειγματοληψία
 - 6 | $P \leftarrow \text{Ανομοιόμορφη Δειγματοληψία}(f, s(p))$ ▷ μεταβλητό βήμα $s(p)$
-

Ονομασία	Τριγωνοποίηση	α -σχήματα
Delaunay	Delaunay	απλά
Constrained	Delaunay με περιορισμούς	απλά, $\rho_T = 0$ στους περιορισμούς
Regular	κανονική	σταθμισμένα ανισοτροπικά
Anisotropic	κανονική	σταθμισμένα

Πίνακας 3.1: Συνδυασμοί τριγωνοποιήσεων και α -σχημάτων.

οποίου το υποδέντρο περιέχει όλα τα μεγαλύτερα σε μέγεθος άπλοκα της ίδιας συνιστώσας. Διατηρούμε μια λίστα με τα παιδιά κάθε άπλοκου, χρησιμοποιώντας τη συνάρτηση ΠροσθήκηΠαιδιού, ενώ αρχικά όλα τα άπλοκα θεωρούνται φύλλα. Η δεύτερη δομή είναι ένα σύνηθες δάσος ασύνδετων συνόλων (disjoint set forest), όπου τα άπλοκα δείχνουν μόνο τον γονιό τους. Τη δομή αυτή τη διαχειρίζόμαστε με τις συναρτήσεις Δημιουργία-Συνόλου, Αναζήτηση και Ένωση [18].

Οι δύο δομές επικοινωνούν μέσω του δείκτη ρ . Η δεύτερη χρησιμοποιείται για ταχύτητα, ενώ η πρώτη για να συλλέγουμε πληροφορίες. Συγκεκριμένα, για κάθε επιλεγμένη συνιστώσα συλλέγουμε το σύνολο των άπλοκων που την απαρτίζουν με μία κατά πλάτος αναζήτηση στο υποδέντρο του άπλοκου- ρ , και προσαρμόζουμε στο κυρτό περίβλημά της μια έλλειψη. Η πολυπλοκότητα κατασκευής του δέντρου συνιστωσών είναι οιονεί γραμμική ως προς n , δηλαδή γραμμική για τη συνήθη περίπτωση [81].

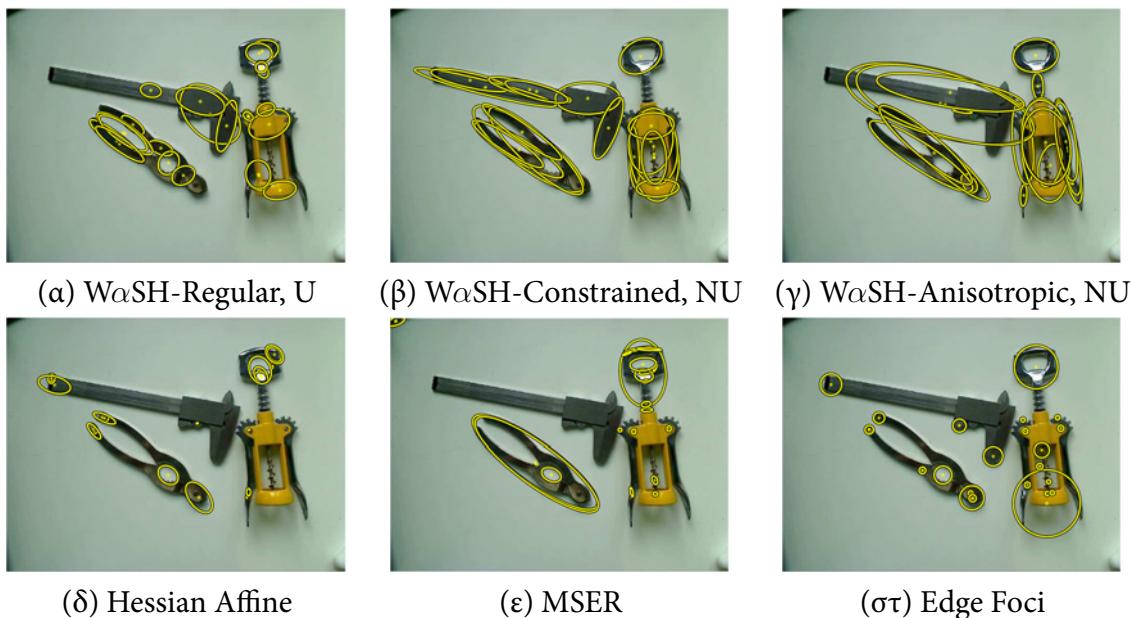
Παραδείγματα

Για να περιγράψουμε πιο ποιοτικά τις ιδιότητες των τοπικών χαρακτηριστικών που ανιχνεύουμε, στις επόμενες εικόνες δείχνουμε παραδείγματα ανιχνεύσεων του προτεινόμενου αλγορίθμου WαSH, καθώς και ανταγωνιστικών μεθόδων. Στις Εικόνες 3.8 και 3.9 βλέπουμε τα τοπικά χαρακτηριστικά που ανιχνεύονται σε μια εικόνα του συνόλου δεδομένων με εικόνες χωρίς υφή των Tombari et al. [108]. Στην εικόνα φαίνονται τρία εργαλεία σε λευκό φόντο. Αρχικά ρυθμίζουμε τις παραμέτρους των ανιχνευτών ώστε να εξάγουμε περίπου τον ίδιο αριθμό χαρακτηριστικών ανά εικόνα και βλέπουμε τα αποτελέσματα στην Εικόνα 3.8. Στην Εικόνα 3.9 βλέπουμε τα εξαγόμενα χαρακτηριστικά όταν χρησιμοποιούμε τις προκαθορισμένες τιμές παραμέτρων για κάθε ανιχνευτή.

Ο ανιχνευτής WαSH εντοπίζει περιοχές με ευδιάκριτα όρια, καλύπτοντας όλα τα αντικείμενα ενδιαφέροντος. Χρησιμοποιώντας τα ανισοτροπικά α -σχήματα εντοπίζουμε καλύτερα τις επιμήκεις δομές, συγκριτικά με τη σταθμισμένη και την περιορισμένη εκδοχή του αλγορίθμου. Αυξάνοντας τον αριθμό των εξαγόμενων χαρακτηριστικών εντοπίζονται όλα τα δομικά μέρη των αντικειμένων, όπως φαίνεται στην Εικόνα 3.9.

Ο ανιχνευτής Hessian-Affine εξάγει πολλαπλά χαρακτηριστικά με μεγάλες επικαλύ-

Ανίχνευση χαρακτηριστικών



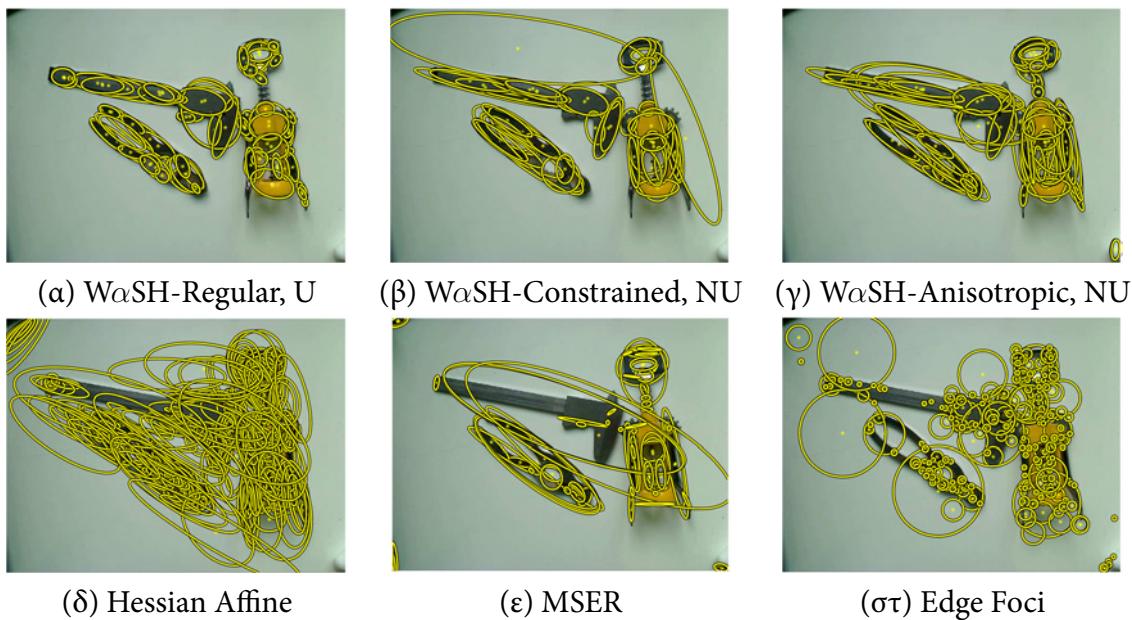
Εικόνα 3.8: Ανίχνευση χαρακτηριστικών σε αντικείμενα χωρίς υφή. Οι παράμετροι των ανιχνεύτων καθορίστηκαν έτσι ώστε ο αριθμός των χαρακτηριστικών για κάθε ανιχνευτή να περιοριστεί στα περίπου 20 ανά εικόνα. W α SH-Regular, U: 23, W α SH-Constrained, NU: 20, W α SH-Anisotropic, NU: 21, Hessian-Affine: 21, MSER: 22, Edge Foci: 23 χαρακτηριστικά. Όπου U χρησιμοποιούμε ομοιόμορφη δειγματοληψία, ενώ όπου NU ανομοιόμορφη.

ψεις σε διαφορετικές κλίμακες, αποτυγχάνοντας να εντοπίσει σημαντικές λεπτομέρειες των αντικειμένων όταν ο αριθμός των χαρακτηριστικών είναι περιορισμένος. Χρησιμοποιώντας τις προκαθορισμένες παραμέτρους εντοπίζονται όλες οι λεπτομέρειες των αντικειμένων, αλλά ο αριθμός των χαρακτηριστικών αυξάνεται δραματικά.

Ο ανιχνευτής EF, χρησιμοποιώντας τις προεπιλεγμένες παραμέτρους, ανιχνεύει μικρές λεπτομέρειες των αντικειμένων με υψηλή διακριτικότητα. Εντούτοις, λόγω του περιορισμού των τεσσάρων παραμέτρων που εκμεταλλεύεται στους μετασχηματισμούς, εντοπίζει χαρακτηριστικά με χαμηλή επαναληψιμότητα σε επιμήκεις περιοχές. Ο ανιχνευτής MSER ανιχνεύει αραιές περιοχές με καλή προσαρμογή στο σχήμα των αντικειμένων, λόγω της επεξεργασίας σε επίπεδο εικονοστοιχείων σε συνδυασμό με τη σταθερότητα στη μεταβολή της φωτεινότητας. Παρόλα αυτά, δεν εντοπίζει λεπτομέρειες του σχήματος σε μικρή κλίμακα, που θα μπορούσαν να βοηθήσουν σε ένα σύστημα ανίχνευσης αντικειμένων ή αναζήτησης εικόνων.

Στις Εικόνες 3.10 και 3.11 βλέπουμε τα χαρακτηριστικά που ανιχνεύθηκαν σε μία εξωτερική και μία εσωτερική σκηνή. Για να συγκρίνουμε πιο δίκαια τους αλγορίθμους, καθορίζουμε και πάλι τον αριθμό των ανιχνεύσεων ανά εικόνα. Στην Εικόνα 3.10 ο ανιχνευτής W α SH (με όλες τις παραλλαγές του) πετυχαίνει μεγάλη κάλυψη στην εικόνα, εξάγοντας χαρακτηριστικά στα περισσότερα αντικείμενα της εικόνας.

Στην Εικόνα 3.11, χρησιμοποιώντας τα ανισοτροπικά α -σχήματα εντοπίζουμε την πλειο-

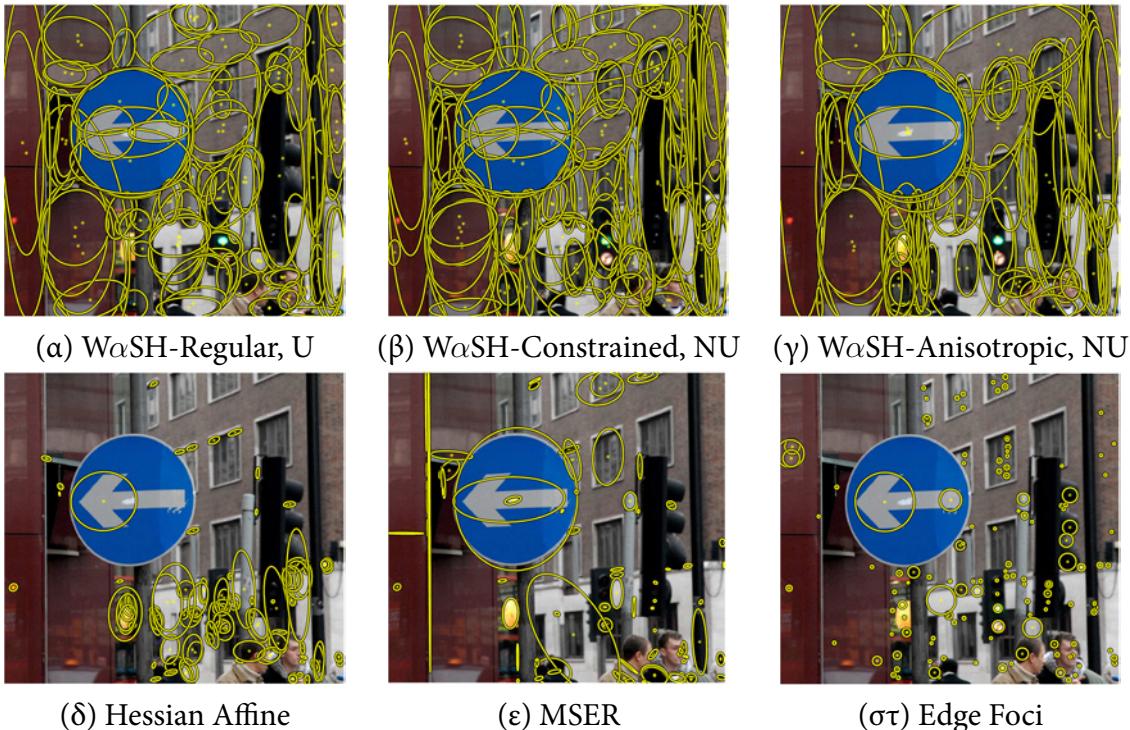


Εικόνα 3.9: Ανίχνευση χαρακτηριστικών σε αντικείμενα χωρίς υφή, χρησιμοποιώντας τις προεπιλεγμένες παραμέτρους κάθε ανιχνευτή. *WαSH-Regular, U*: 85, *WαSH-Constrained, NU*: 53, *WαSH-Anisotropic, NU*: 94, *Hessian-Affine*: 808, *MSER*: 88, *Edge Foci*: 284 χαρακτηριστικά.

ψηφία των εξεχουσών περιοχών της εικόνας, όπως τους σχεδόν κυκλικούς χρωματιστούς δακτύλιους. Ο αλγόριθμός μας είναι ο μόνος που ανιχνεύει την επαναλαμβανόμενη δομή πλακιδίων στη σκηνή. Ο ανιχνευτής MSER αποδίδει καλά, αλλά εντοπίζει αρκετές επιμήκεις περιοχές κατά μήκος των ορίων των πλακιδίων, τα οποία έχουν πολύ χαμηλή διακριτική δύναμη. Ο ανιχνευτής EF είναι σχεδιασμένος ώστε να ανιχνεύει κυκλικές περιοχές, και περιοχές που έχουν κάποια κυκλική συμμετρία, έστω και μερική. Παρά το γεγονός ότι υπάρχουν πολλές ομόκεντρες κυκλικές δομές στην Εικόνα 3.11, ο ανιχνευτής EF δεν τις ανιχνεύει.

3.5 Πειραματική αξιολόγηση

Στο κεφάλαιο αυτό θα αποτιμήσουμε πειραματικά την απόδοση του αλγορίθμου ανίχνευσης τοπικών χαρακτηριστικών WαSH που προτείνουμε. Θα εξετάσουμε την επίδραση των διαφορετικών μεθόδων κάθε βήματος (δειγματοληψία, τριγωνοποίηση, α-σχήματα), ενώ ταυτόχρονα θα συγκρίνουμε την απόδοση του ανιχνευτή με τους ανταγωνιστικούς ανιχνευτές της βιβλιογραφίας. Αρχικά, θα αποτιμήσουμε την απόδοση του αλγορίθμου χρησιμοποιώντας τα διαφορετικά μέτρα επιλογής χαρακτηριστικών που προτείνουμε, ώστε να επιλέξουμε το καταλληλότερο. Έπειτα, θα εξετάσουμε την επίδραση των προτεινόμενων μεθόδων δειγματοληψίας καθώς χρησιμοποιούμε τις διαφορετικές τριγωνοποιήσεις και α-σχήματα. Από την πειραματική αποτίμηση θα φανεί στην πράξη η καταλληλότητα κάθε μεθόδου σε κάθε βήμα του αλγορίθμου.



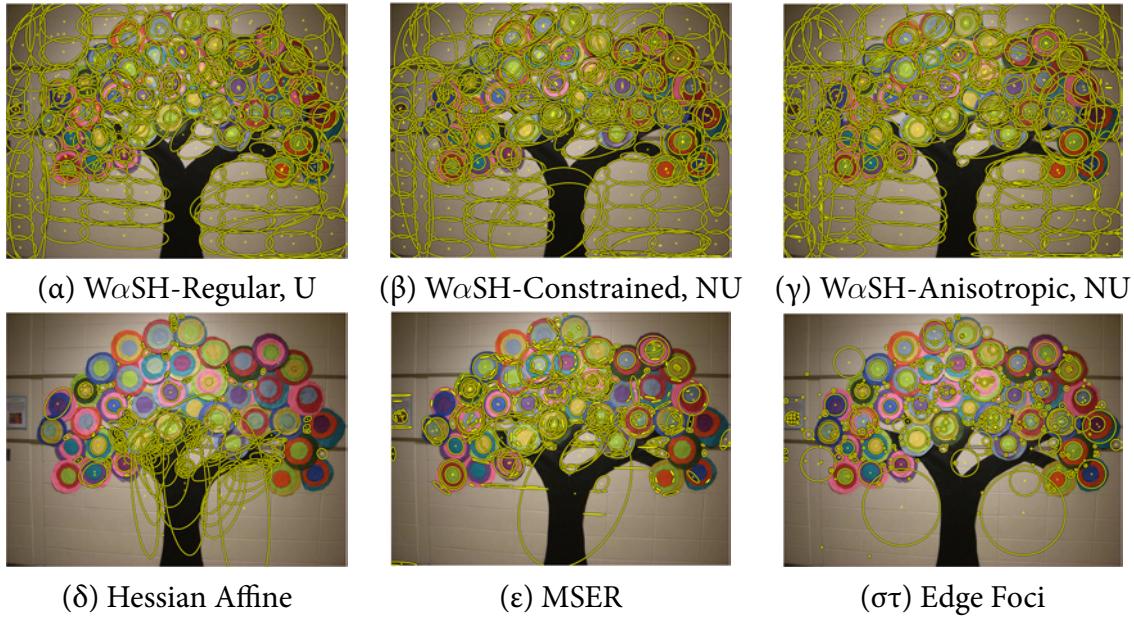
Εικόνα 3.10: Ανίχνευση χαρακτηριστικών σε εξωτερική σκηνή. Τα κατώφλια των ανιχνευτών έχουν προσαρμοστεί ώστε να εξαχθούν περίπου 130 χαρακτηριστικά ανά εικόνα. W α SH-Regular, U: 122, W α SH-Constrained, NU: 136, W α SH-Anisotropic, NU: 114, Hessian-Affine: 137, MSER: 134, Edge Foci: 135 χαρακτηριστικά.

Για την πειραματική αποτίμηση των ανιχνευτών χαρακτηριστικών μπορούμε είτε να ελέγξουμε τις ιδιότητες των περιοχών που ανιχνεύονται (π.χ. επαναληψιμότητα, μοναδικότητα, διακριτικότητα), είτε να χρησιμοποιήσουμε τα εξαγόμενα χαρακτηριστικά ως στάδιο ενός άλλου αλγορίθμου της όρασης υπολογιστών (π.χ. ανίχνευσης και αναγνώρισης αντικειμένων, ταξινόμησης και αναζήτησης εικόνων) και να εξετάσουμε τις επιπτώσεις στην απόδοση του χρησιμοποιούμενου αλγορίθμου. Στο κεφάλαιο αυτό θα χρησιμοποιήσουμε δύο μεθόδους για την αποτίμηση του ανιχνευτή W α SH, ένα πείραμα ταιριάσματος εικόνων και ένα μεγάλης κλίμακας πείραμα ανάκτησης εικόνων.

Ταίριασμα εικόνων

Το πείραμα ταιριάσματος εικόνων προτάθηκε από τους Mikolajczyk et al. [78] και αποτιμά την απόδοση ενός αλγορίθμου ταιριάσματος εικόνων με χρήση τοπικών χαρακτηριστικών. Περιλαμβάνει ένα σύνολο εικόνων που χωρίζονται σε ακολουθίες, οι οποίες εξετάζουν την επίδραση διάφορων μετασχηματισμών και αλλοιώσεων των εικόνων στην απόδοση του ταιριάσματος. Κάθε ακολουθία αποτελείται από έξι εικόνες, όπου το ταίριασμα της πρώτης με τις υπόλοιπες πέντε παρουσιάζει αυξανόμενη δυσκολία.

Με τις ακολουθίες *bark* και *boat* εξετάζουμε την επίδραση της αλλαγής κλίμακας και



Εικόνα 3.11: Ανίχνευση χαρακτηριστικών σε εσωτερική σκηνή. Τα κατώφλια των ανιχνευτών έχουν προσαρμοστεί ώστε να εξαχθούν περίπου 400 χαρακτηριστικά ανά εικόνα. *WaSH-Regular*, *U*: 413, *WaSH-Constrained*, *NU*: 380, *WaSH-Anisotropic*, *NU*: 405, *Hessian-Affine*: 425, *MSER*: 432, *Edge Foci*: 413 χαρακτηριστικά. Παρατηρήστε τα πλακίδια στο κάτω μέρος της εικόνας που ανιχνεύονται από τον προτεινόμενο αλγόριθμο.

περιστροφής στο ταίριασμα των εικόνων, ενώ με τις ακολουθίες *graffiti* και *wall* την επίδραση αφινικών μετασχηματισμών στις εικόνες, από αλλαγή της γωνίας θέασης. Οι ακολουθίες *bikes* και *trees* εξετάζουν την επίδραση του θολώματος εικόνων, ενώ στην ακολουθία *leuven* εξετάζεται η επίδραση της αλλαγής φωτεινότητας. Τέλος, η ακολουθία *ubc* εξετάζει την επίδραση της συμπίεσης κατά JPEG.

Στο συγκεκριμένο πείραμα, η ποιότητα του ανιχνευτή χαρακτηριστικών αποτιμάται με τη μέτρηση της επαναληψιμότητας (*repeatability*), του αριθμού των αντιστοιχισμένων περιοχών (*number of correspondencies*), της απόδοσης ταιριάσματος (*matching score*), και του αριθμού των σωστών ταιριασμάτων (*correct matches*). Για να μετρήσουμε τις ποσότητες που αναφέραμε παραπάνω θα πρέπει να γνωρίζουμε τον ακριβή μετασχηματισμό H μεταξύ δύο εικόνων a και b τις οποίες εξετάζουμε. Η έλλειψη που αντιστοιχεί σε ένα χαρακτηριστικό συμβολίζεται με R_μ και ορίζεται από την εξίσωση $x^\tau \mu x = 1$. Η προβολή της έλλειψης R_{μ_b} από την εικόνα b στην εικόνα a συμβολίζεται με $R_{H^\tau \mu_b H}$. Η ένωση των δύο έλλειψεων συμβολίζεται με $R_{\mu_a} \cup R_{H^\tau \mu_b H}$, η τομή τους με $R_{\mu_a} \cap R_{H^\tau \mu_b H}$, ενώ το $|\cdot|$ συμβολίζει το εμβαδό μιας περιοχής. Το σφάλμα επικάλυψης μεταξύ δύο έλλειψεων ορίζεται ως:

$$1 - \frac{|R_{\mu_a} \cap R_{H^\tau \mu_b H}|}{|R_{\mu_a} \cup R_{H^\tau \mu_b H}|}. \quad (3.29)$$

Για να θεωρήσουμε ότι δύο έλλειψεις ταυτίζονται, θα πρέπει το σφάλμα επικάλυψης μεταξύ τους να είναι μικρότερο από ϵ_0 . Στο [78] επιλέχθηκε ότι δύο έλλειψεις ταυτίζονται όταν

$\epsilon_0 = 40\%$, επιλογή που ακολουθούμε και σε όλα μας τα πειράματα.

Η επαναληψιμότητα μεταξύ ενός δεδομένου ζεύγους εικόνων a και b , ορίζεται ως ο λόγος των ελείψεων που ταυτίζονται στις δύο εικόνες (αντιστοιχισμένων περιοχών), προς τον μικρότερο από τους αριθμούς χαρακτηριστικών στις δύο εικόνες. Για τη μέτρηση αυτή λαμβάνονται υπόψιν μόνο τα χαρακτηριστικά που βρίσκονται στο κοινό τμήμα των δύο εικόνων.

Η απόδοση ταιριάσματος είναι μέτρηση αντίστοιχη της επαναληψιμότητας, όπου λαμβάνεται υπόψιν και ο περιγραφέας κάθε τοπικού χαρακτηριστικού. Ο ορισμός είναι αντίστοιχος με αυτόν της επαναληψιμότητας, όμως για να ταυτίζονται δύο ελλείψεις θα πρέπει, επιπλέον, οι περιγραφείς τους να είναι οι κοντινότεροι γείτονες μεταξύ όλων των περιγραφέων των χαρακτηριστικών των δύο εικόνων. Ισούται, δηλαδή, με τον αριθμό των σωστών ταιριασμάτων προς τον αριθμό των χαρακτηριστικών στο κοινό τμήμα των δύο εικόνων. Για τον υπολογισμό των σωστών ταιριασμάτων και της απόδοσης ταιριάσματος εξάγουμε περιγραφείς SIFT [70] για κάθε χαρακτηριστικό, διάστασης 128. Για τα χαρακτηριστικά KAZE εξάγουμε περιγραφείς SURF [8], καθώς αυξάνουν την απόδοσή τους και προτείνονται από τους συγγραφείς. Για να εξασφαλίσουμε την ακρίβεια και ορθότητα των αποτελεσμάτων, χρησιμοποιούμε για την πραγματοποίηση του πειράματος τη βιβλιοθήκη VL-Benchmarks που προτάθηκε πρόσφατα από τους Lenc et al. [61].

Μεγάλης κλίμακας ανάκτηση εικόνων

Το πείραμα ανάκτησης εικόνων αφορά το σύνολο εικόνων Oxford 5K που προτάθηκε από τους Philbin et al. [89] και περιλαμβάνει 5062 εικόνες του δικτυακού τόπου Flickr¹ που προέκυψαν από αναζήτηση συγκεκριμένων ορόσημων της Οξφόρδης. Αποτελείται από 55 εικόνες ερωτήματα και περιλαμβάνει σωστές και λανθασμένες απαντήσεις, οι οποίες έχουν επισημειωθεί από τους συγγραφείς του [89].

Για όλους τους ανιχνευτές χαρακτηριστικών που συμμετέχουν στο πείραμα εξάγουμε περιγραφείς SIFT, όπως και στο πείραμα ταιριάσματος εικόνων. Για τους ανιχνευτές SURF και KAZE εξάγουμε περιγραφείς SURF, καθώς μεγιστοποιούν την απόδοση των συγκεκριμένων ανιχνευτών και προτείνονται από τους αντίστοιχους συγγραφείς (όμοια με το πείραμα ταιριάσματος εικόνων). Από ένα τυχαίο υποσύνολο των περιγραφέων κατασκευάζουμε ένα οπτικό λεξικό με τη μέθοδο του προσεγγιστικού k -means [89]. Το μέγεθος του λεξικού επηρεάζει την τελική απόδοση του συστήματος, κι έτσι πειραματιζόμαστε με διαφορετικά μεγέθη.

Για την αναπαράσταση των εικόνων από τις εμφανίσεις των οπτικών λέξεων χρησιμοποιούμε το μοντέλο *Bag-of-Words* (*BoW*). Σταθμίζουμε τις εμφανίσεις των οπτικών λέξεων σύμφωνα με τη μέθοδο *tf-idf* και χρησιμοποιούμε ανεστραμμένο αρχείο για τη δεικτοδό-

¹<http://www.flickr.com/>

τηση. Για το ταίριασμα των ιστογραμμάτων του μοντέλου BoW χρησιμοποιούμε τη μέθοδο της τομής ιστογραμμάτων, μετά από κανονικοποίησή τους ως προς την ℓ_1 νόρμα. Επιπλέον, αποτιμούμε την απόδοση του συστήματος όταν γίνεται χρήση του συστήματος γρήγορου χωρικού ταιριάσματος *FastSM* (*fast spatial matching*) [89] για τη χωρική επιβεβαίωση των αποτελεσμάτων. Ο ελάχιστος αριθμός χαρακτηριστικών για τη χωρική επιβεβαίωση τίθεται ίσος με 7. Η μετρική αποτίμησης είναι το μέτρο μέσης ακρίβειας *mAP* (*mean Average Precision*).

3.5.1 Μέτρα σημαντικότητας

Σε αυτή την ενότητα συγκρίνουμε τα μέτρα σημαντικότητας της ενότητας 3.4.2 και επιλέγουμε το πιο κατάλληλο. Η επιλογή βασίζεται στη μέτρηση της επαναληψιμότητας και του βαθμού ταιριάσματος σε ένα πείραμα μικρής κλίμακας. Για κάθε διαφορετικό μέτρο σημαντικότητας, προσαρμόζουμε το κατώφλι επιλογής ώστε να εξάγουμε κατά προσέγγιση τον ίδιο αριθμό χαρακτηριστικών ανά εικόνα.

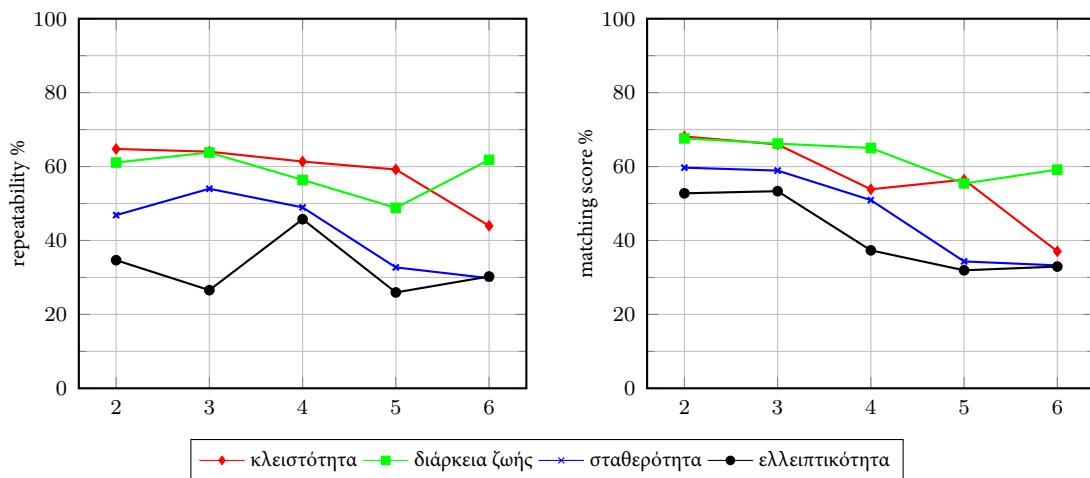
Χρησιμοποιούμε έξι ακολουθίες εικόνων από το σύνολο δεδομένων των Mikolajczyk et al., ώστε να αποτιμήσουμε την επίδραση της αλλαγής κλίμακας και περιστροφής (ακολουθία *boat*), αλλαγών στη γωνία θέασης (*wall* και *graffiti*), τη φωτεινότητα (*leuven*) και το θόλωμα (*bikes and trees*). Στην Εικόνα 3.12 βλέπουμε την απόδοση του ανιχνευτή μας χρησιμοποιώντας όλα τα διαφορετικά μέτρα σημαντικότητας κάτω από περιστροφή, θόλωμα, αλλαγές στην κλίμακα, τη φωτεινότητα, καθώς και αφινικούς μετασχηματισμούς. Οι μετρήσεις αποτελούν το μέσο όρο για κάθε εικόνα από όλες τις διαφορετικές ακολουθίες. Για το πείραμα αυτό χρησιμοποιήσαμε ομοιόμορφη δειγματοληψία επί ακμών, κανονική τριγωνοποίηση και σταθμισμένα α-σχήματα (ισοτροπικά).

Ενώ ο αριθμός των χαρακτηριστικών ανά εικόνα είναι κατά προσέγγιση ίδιος, παρατηρούμε μεγάλη διακύμανση στην απόδοση χρησιμοποιώντας διαφορετικά μέτρα σημαντικότητας, γεγονός που είναι ενδεικτικό της σημασίας τους. Γενικά, με τα μέτρα κλειστότητας και διάρκειας ζωής λαμβάνουμε τις καλύτερες αποδόσεις και στις δύο μετρήσεις. Για όλα τα υπόλοιπα πειράματα χρησιμοποιούμε το μέτρο κλεισμάτος.

3.5.2 Δειγματοληψία, τριγωνοποίησεις και α-σχήματα

Σε αυτή την ενότητα ελέγχουμε την απόδοση των διαφορετικών τριγωνοποιήσεων και α-σχημάτων σε συνδυασμό με τις διαφορετικές δειγματοληψίες επί ακμών. Συγκεκριμένα για τις δύο μεθόδους δειγματοληψίας επί ακμών (ομοιόμορφη και ανομοιόμορφη), εξετάζουμε όλες τις επιλογές για την τριγωνοποίηση (Delaunay, Delaunay με περιορισμούς, κανονική) σε συνδυασμό με τα α-σχήματα που περιγράψαμε στην ενότητα 3.3 (σταθμισμένα και ανισοτροπικά σταθμισμένα α-σχήματα σε κανονική τριγωνοποίηση).

Πειραματική αξιολόγηση



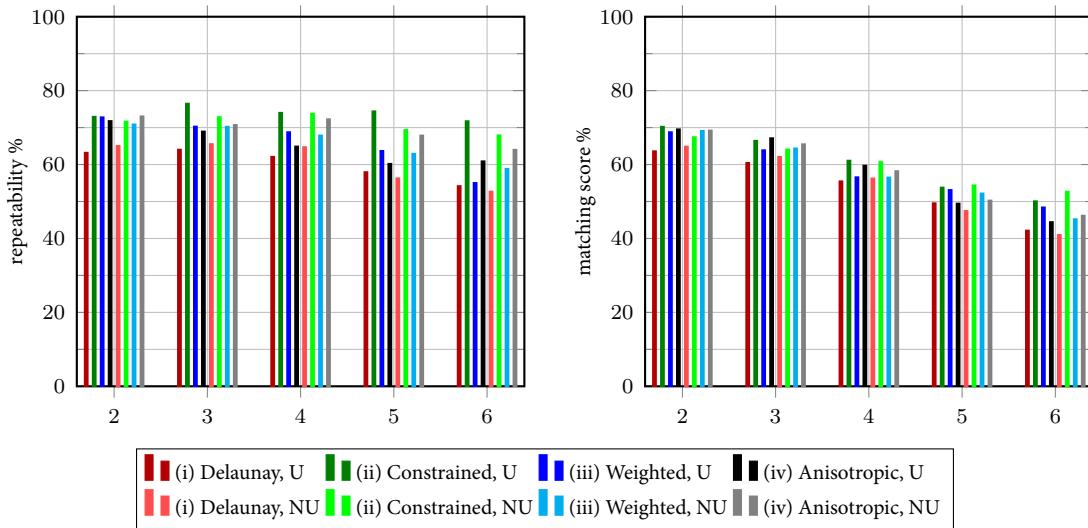
Εικόνα 3.12: Αποτίμηση των προτεινόμενων μέτρων σημαντικότητας. Διατηρώντας τον αριθμό των χαρακτηριστικών σταθερό για όλα τα μέτρα σημαντικότητας, εξετάζουμε (α) την επαναληψιμότητα και (β) το βαθμό ταιριάσματος. Σε κάθε εικόνα (2 έως 6) ανξένεται ο βαθμός δυσκολίας στο ταιριάσμα, χρησιμοποιώντας πιο έντονο μετασχηματισμό. Οι μετρήσεις είναι ο μέσος όρος των έξι διαφορετικών ακολουθιών εικόνων.

Επαναληψιμότητα και απόδοση ταιριάσματος

Αρχικά, όπως και στην ενότητα 3.5.1, για κάθε διαφορετική επιλογή μετράμε τη μέση επαναληψιμότητα και απόδοση ταιριάσματος στις ίδιες ακολουθίες εικόνων. Έπειτα παρουσιάζουμε αναλυτικά την απόδοση επιλεγμένων συνδυασμών σε επαναληψιμότητα και απόδοση ταιριάσματος.

Τα πειραματικά αποτελέσματα για τους οχτώ συνδυασμούς φαίνονται στην εικόνα 3.13. Χρησιμοποιώντας την τριγωνοποίηση Delaunay σε συνδυασμό είτε με την ομοιόμορφη, είτε με την ανομοιόμορφη δειγματοληψία επί ακμών, τα αποτελέσματα είναι κατώτερα των άλλων περιπτώσεων. Το γεγονός αυτό είναι αναμενόμενο, καθώς η περίπτωση αυτή χρησιμοποιεί συγκριτικά τη λιγότερη πληροφορία από το πεδίο της εικόνας. Με χρήση της τριγωνοποίησης Delaunay με περιορισμούς παίρνουμε τα καλύτερα αποτελέσματα, ενώ τα ανισοτροπικά σταθμισμένα α -σχήματα με την κανονική τριγωνοποίηση δίνουν ανταγωνιστικά αποτελέσματα. Για τα αναλυτικά αποτελέσματα που θα ακολουθήσουν, παρουσιάζουμε μόνο τις δύο αυτές επιλογές με την καλύτερη απόδοση, για λόγους ευκρίνειας των γραφικών παραστάσεων.

Στην ομοιόμορφη δειγματοληψία, θέτουμε εμπειρικά το βήμα s ίσο με 11. Σε ορισμένες περιπτώσεις, η ομοιόμορφη δειγματοληψία ξεπερνά την απόδοση της ανομοιόμορφης: σε αυτές τις περιπτώσεις το σταθερό βήμα είναι κοντά στο ιδανικό. Στην ανομοιόμορφη δειγματοληψία, έπειτα από ένα σύνολο ποιοτικών πειραμάτων σε εικόνες μεταβλητής λεπτομέρειας, θέτουμε το άνω όριο εκκεντρότητας k ίσο με 3 (βλ. ενότητα 3.1.3). Επιλέγουμε το αρχικό μέγεθος της γειτονιάς p κάθε δείγματος p να είναι ίσο με 11×11 εικονοστοι-



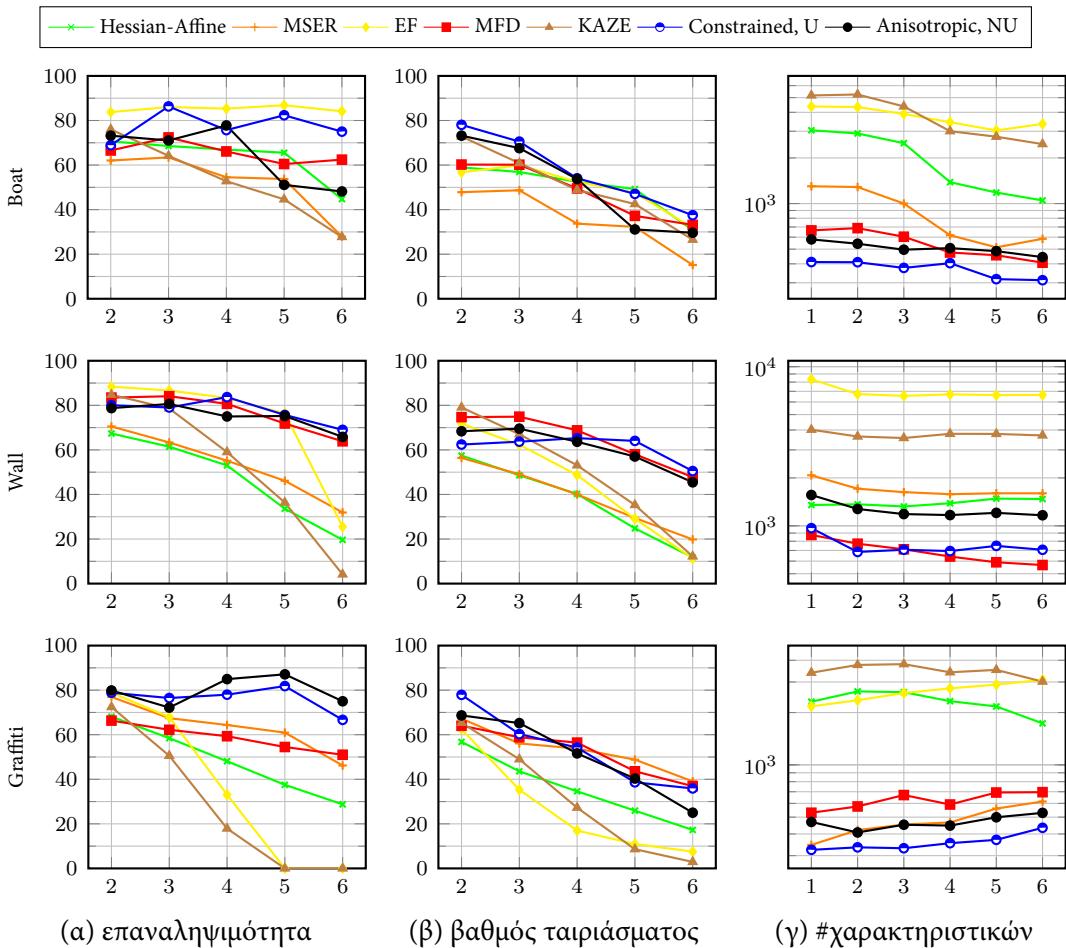
Εικόνα 3.13: Μέση επαναληψιμότητα και βαθμός ταιριάσματος για διαφορετικές τριγωνοποιήσεις, α-σχήματα και δειγματοληψία. Εξετάζουμε α-σχήματα με τριγωνοποίηση (i) Delaunay και (ii) Delaunay με περιορισμούς, κανονική τριγωνοποίηση με (iii) σταθμισμένα α-σχήματα και (iv) ανισοτροπικά σταθμισμένα α-σχήματα. Όπου U χρησιμοποιούμε ομοιόμορφη δειγματοληψία, ενώ όπου NU ανομοιόμορφη. Οι μετρήσεις είναι ο μέσος όρος των έξι διαφορετικών ακολουθιών εικόνων.

χεία. Το αρχικό μέγεθος της γειτονιάς δεν είναι κρίσιμο, καθώς αυτή θα προσαρμοστεί στο τοπικό σχήμα μέσω της επαναληπτικής διαδικασίας. Έτσι, επιλέγουμε μια σχετικά μικρή αρχική γειτονιά για να μειώσουμε το υπολογιστικό κόστος της μεθόδου δειγματοληψίας.

Εστιάζοντας στις δύο περιπτώσεις με την καλύτερη απόδοση, όπως είδαμε στην Εικόνα 3.13, συγκρίνουμε αναλυτικά τον ανιχνευτή μας με τις ανταγωνιστικές μεθόδους της βιβλιογραφίας, στο ίδιο πείραμα ταιριάσματος εικόνων. Επιλέξαμε τους ανιχνευτές με την καλύτερη απόδοση από το [78] (Hessian-Affine και MSER), καθώς και τρεις πρόσφατους ανιχνευτές της βιβλιογραφίας: MFD, EF και KAZE. Τα χαρακτηριστικά των τριών αυτών ανιχνευτών εξάγονται από εκτελέσιμα που παρέχουν δημόσια οι συγγραφείς, και τα έχουμε ενσωματώσει στη βιβλιοθήκη VL-Benchmarks. Για όλους τους ανιχνευτές χρησιμοποιούμε τις προεπιλεγμένες τιμές παραμέτρων.

Τα αποτελέσματα για όλες τις ακολουθίες εικόνων φαίνονται στις Εικόνες 4.8 και 4.9, μαζί με τον αριθμό χαρακτηριστικών ανά εικόνα. Ο προτεινόμενος ανιχνευτής WαSH (και με τις δύο διαφορετικές τριγωνοποιήσεις) προσφέρει εξαιρετική ισορροπία μεταξύ απόδοσης και αριθμού χαρακτηριστικών. Η απόδοσή του βρίσκεται σταθερά μεταξύ των καλύτερων σε όλες τις περιπτώσεις, ενώ ο αριθμός των χαρακτηριστικών διατηρείται χαμηλά. Η υψηλή του απόδοση αποδεικνύει την μεγάλη ανοχή των χαρακτηριστικών σε όλους τους εξεταζόμενους μετασχηματισμούς.

Πειραματική αξιολόγηση



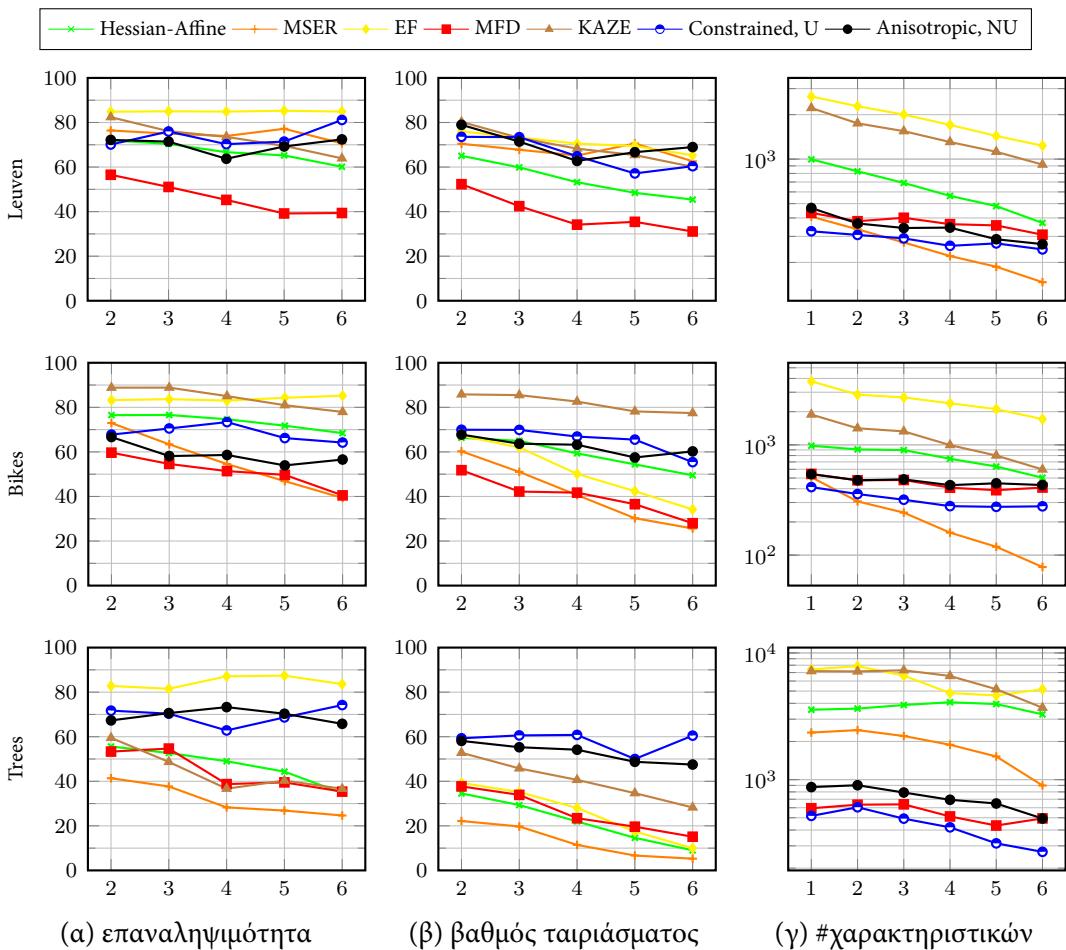
Εικόνα 3.14: Επαναληψιμότητα και βαθμός ταιριάσματος. Σύγκριση του προτεινόμενου ανιχνευτή (Delaunay με περιορισμούς με ομοιόμορφη δειγματοληψία και ανισοτροπικά σταθμισμένα α -σχήματα με ανομοιόμορφη δειγματοληψία) με τη βιβλιογραφία στα σύνολα εικόνων boat, wall και graffiti.

Ανάκτηση εικόνων

Στο πείραμα ανάκτησης εικόνων συγκρίνουμε όλους τους συνδυασμούς του ανιχνευτή χαρακτηριστικών $W\alpha SH$ που εξετάσαμε στο πείραμα ταιριάσματος εικόνων, με τους υπόλοιπους ανταγωνιστικούς ανιχνευτές, καθώς και τους ανιχνευτές SIFT και SURF, που χρησιμοποιούνται σε πληθώρα υλοποιημένων εφαρμογών αναζήτησης εικόνων.

Για κάθε ανταγωνιστικό ανιχνευτή χρησιμοποιούνται, και πάλι, οι προεπιλεγμένες παράμετροι. Για κάθε ανιχνευτή, εξετάζουμε την απόδοση με οπτικά λεξικά μεγέθους 50K και 200K λέξεων. Στην Εικόνα 3.16 βλέπουμε παραδείγματα των χαρακτηριστικών που εξάγει ο αλγόριθμός μας, καθώς και οι ανταγωνιστικοί ανιχνευτές, σε μια εικόνα του συνόλου εικόνων Oxford 5K με τις παραμέτρους που χρησιμοποιήθηκαν στο πείραμα.

Στους Πίνακες 3.2 και 3.3 βλέπουμε αναλυτικά αποτελέσματα, συμπεριλαμβανομένου του συνολικού αριθμού χαρακτηριστικών, του μέσου χρόνου εξαγωγής χαρακτηριστικών ανά εικόνα, του μεγέθους του ανεστραμμένου αρχείου, του μέσου χρόνου ερωτήματος και της μέτρησης mAP για κάθε ανιχνευτή, για τα δύο διαφορετικά οπτικά λεξικά. Ο συνολι-

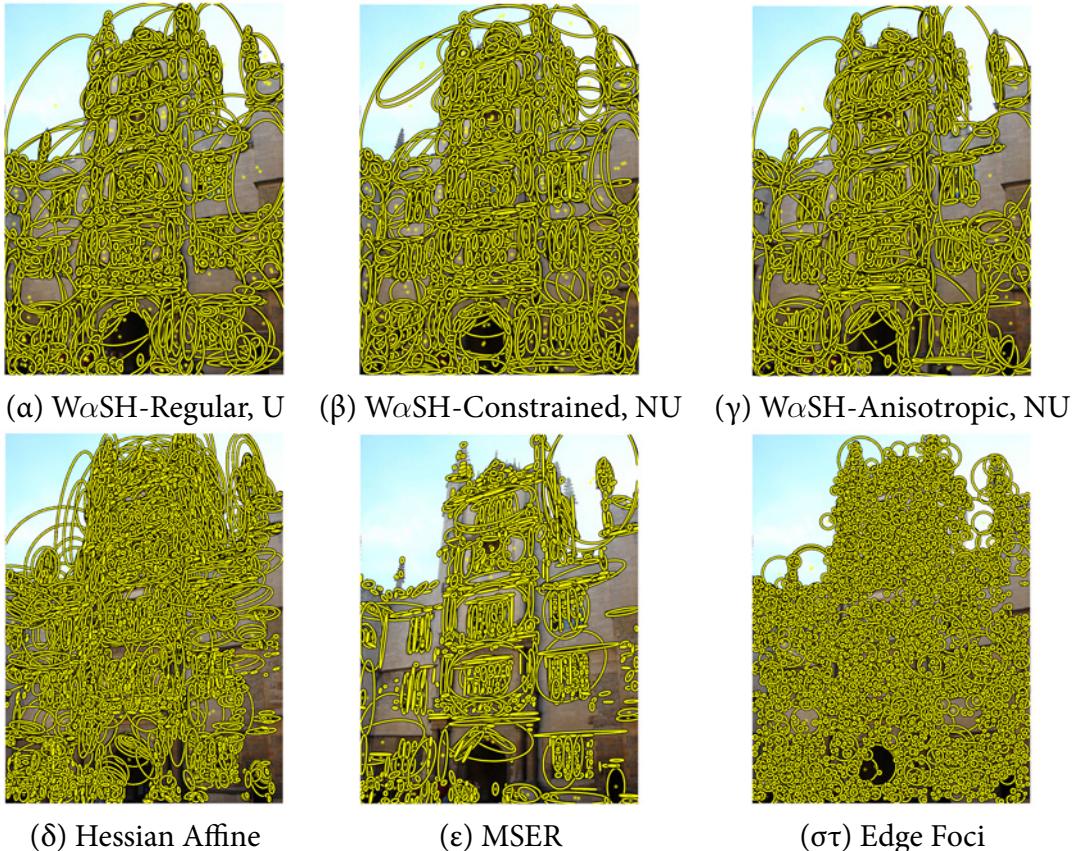


Εικόνα 3.15: Επαναληψιμότητα και βαθμός ταιριάσματος. Σύγκριση του προτεινόμενου ανίχνευσης Delaunay με περιορισμούς με ομοιόμορφη δειγματοληψία επί ακμών και ανισοτροπικά σταθμισμένα α -σχήματα με ανομοιόμορφη δειγματοληψία) με τη βιβλιογραφία στα σύνολα εικόνων leuven, bikes και trees.

κός αριθμός χαρακτηριστικών και ο μέσος χρόνος εξαγωγής χαρακτηριστικών ανά εικόνα ταυτίζονται στους δύο πίνακες, αλλά παρατίθενται για καλύτερη σύγκριση.

Η απόδοση όλων των παραλλαγών του προτεινόμενου αλγορίθμου είναι καλύτερη ή αντίστοιχη των καλύτερων της βιβλιογραφίας, παρά το γεγονός ότι χρησιμοποιείται πολύ μικρότερος αριθμός χαρακτηριστικών συγκριτικά με τις ανταγωνιστικές μεθόδους. Συγκεκριμένα, χρησιμοποιώντας μόλις το 1/4 των χαρακτηριστικών του ανίχνευτή Hessian-Affine, πετυχαίνουμε καλύτερη απόδοση. Το όφελος από τη μείωση του χρόνου απάντησης ερωτήματος είναι επίσης σημαντικό. Συγκρίνοντας τους Πίνακες 3.2 και 3.3 βλέπουμε ότι η αύξηση του μεγέθους του λεξικού βελτιώνει την απόδοση για όλους τους ανίχνευτές.

Η χρήση των ανισοτροπικά σταθμισμένων α -σχημάτων επί της κανονικής τριγωνοποίησης δίνει τα καλύτερα αποτελέσματα, ξεπερνώντας τις καλύτερες ανταγωνιστικές μεθόδους της βιβλιογραφίας. Ακολουθούν τα α -σχήματα με τριγωνοποίηση Delaunay με περιορισμούς και τα σταθμισμένα α -σχήματα με κανονική τριγωνοποίηση, επιβεβαιώντας τα αποτελέσματα του πειράματος ταιριάσματος εικόνων. Η υψηλή απόδοση του ανι-



Εικόνα 3.16: Ανιχνευμένα χαρακτηριστικά σε εικόνα του συνόλου Oxford 5K, χρησιμοποιώντας τις παραμέτρους του πειράματος αναζήτησης εικόνων. WαSH-Regular, U: 1453, WαSH-Constrained, NU: 1250, WαSH-Anisotropic, NU: 1315 χαρακτηριστικά. Hessian-Affine: 6921, MSER: 1229, Edge Foci: 4294 χαρακτηριστικά.

στροπικού ανιχνευτή έχει το μειονέκτημα του αυξημένου χρονικού κόστους στην μέση ανίχνευση ανά εικόνα, που οφείλεται στην παραμόρφωση του χώρου για κάθε άπλοκο (βλ. ενότητα 3.3.2) και στην εξαγωγή του μετρικού τανυστή που περιγράφει το τοπικό σχήμα για κάθε σημείο της δειγματοληψίας (βλ. ενότητα 3.1.3).

Η χρήση της ανομοιόμορφης δειγματοληψίας επί ακμών σε αρκετές περιπτώσεις επιφέρει μικρή μείωση της απόδοσης. Σε συνδυασμό με τα σταθμισμένα α -σχήματα, η μείωση της απόδοσης είναι σημαντική. Η μείωση αυτή συμβαίνει καθώς όταν το βήμα δειγματοληψίας s αυξάνει κατά μήκος μιας ευθείας ακμής της εικόνας, αυξάνεται αντίστοιχα και το βάρος των δειγμάτων σύμφωνα με την εξίσωση 3.1. Ως συνέπεια αυτού, ο κύκλος του σταθμισμένου σημείου αυξάνεται ισοτροπικά, με αποτέλεσμα το μέγεθος ρ των ακμών της τριγωνοποίησης που είναι κάθετες στην ακμή της εικόνας να μειώνεται δραστικά. Σε ακραία αύξηση του βάρους ενός δείγματος, ο κύκλος του μπορεί να αυξηθεί τόσο, ώστε να καλύπτει πλήρως τον κύκλο ενός άλλου δείγματος, με συνέπεια την εξαφάνιση του τελευταίου (βλ. ενότητα 3.2.1).

Η ανομοιόμορφη δειγματοληψία επί ακμών αυξάνει το υπολογιστικό κόστος της με-

Κεφάλαιο 3. Ανίχνευση χαρακτηριστικών με α-σχήματα

Ανίχνευτής								
	Πλήθος χαρακτηριστικών ($\times 10^6$)							
HessAff	29.02	6.54	128.8	1.61	6.10	0.578	0.608	
MSER	13.33	0.40	78.8	0.88	2.20	0.568	0.593	
SIFT	11.13	5.24	84.0	0.95	5.29	0.494	0.516	
SURF	6.84	0.43	53.5	0.64	3.45	0.575	0.591	
EF	19.72	13.63	146.2	1.81	4.69	0.528	0.566	
KAZE	13.82	6.59	99.6	1.67	1.91	0.487	0.541	
MFD	7.64	2.98	58.4	0.68	0.93	0.600	0.600	
ομοιόμορφη	Delaunay	7.54	1.38	52.1	0.88	1.11	0.580	0.577
	Constrained	7.17	1.57	50.3	0.84	1.01	0.588	0.590
	Weighted	6.85	2.01	48.1	0.83	1.16	0.595	0.594
	Anisotropic	7.00	3.90	48.6	0.85	1.08	0.621	0.615
ανομοιόμορφη	Delaunay	7.09	2.89	50.4	0.84	0.96	0.592	0.592
	Constrained	7.27	3.52	50.9	0.86	0.99	0.610	0.597
	Weighted	7.71	3.98	53.6	0.88	1.07	0.557	0.560
	Anisotropic	7.47	6.60	50.4	0.89	1.08	0.602	0.594

Πίνακας 3.2: Αποτελέσματα του πειράματος αναζήτησης εικόνων για διαφορετικά α-σχήματα σε λεξικό 200K λέξεων. Το πλήθος των χαρακτηριστικών αναφέρεται σε όλο το σύνολο εικόνων. Ο χρόνος ανίχνευσης είναι ο μέσος χρόνος ανά εικόνα, ενώ ο χρόνος ερωτήματος είναι ο μέσος χρόνος ανά εικόνα-ερώτημα.

Θόδου, λόγω της διαδικασίας προσαρμογής στο τοπικό σχήμα σε κάθε δείγμα. Ωστόσο, αφαιρώντας το βήμα δειγματοληψίας από τις παραμέτρους του ανίχνευτή, καταργούμε την ανάγκη για εύρεση της βέλτιστης παραμέτρου s . Η μέθοδος αυτή ενδείκνυται για πειράματα μεγάλης κλίμακας, όπου δεν είναι χρονικά εφικτή η εύρεση βέλτιστων παραμέτρων με δοκιμές.

Ο ανιχνευτής MFD των Αβρίθη και Ραπαντζίκου [6] έχει απόδοση αντίστοιχη με του

Αναχνευτής	Πλήθος χαρακτηριστικών ($\times 10^6$)	Χρόνος ανίχνευσης (s)	Ανεστραμένο αρχείο (MB)	Χρόνος ερωτήματος με BoW (s)	Χρόνος ερωτήματος με FastSM (s)	Μέση ακρίβεια με BoW (mAP)	Μέση ακρίβεια με FastSM (mAP)	
HessAff	29.02	6.54	116.2	2.71	25.17	0.489	0.516	
MSER	13.33	0.40	71.2	1.32	6.57	0.489	0.524	
SIFT	11.13	5.24	75.9	1.51	8.35	0.422	0.446	
SURF	6.84	0.43	47.8	0.88	3.75	0.466	0.497	
EF	19.72	13.63	132.1	3.11	26.01	0.455	0.500	
KAZE	13.82	6.59	89.4	2.62	7.30	0.403	0.464	
MFD	7.64	2.98	51.9	0.94	2.45	0.531	0.540	
ομοιόμορφη	Delaunay	7.54	1.38	47.0	1.15	1.60	0.521	0.537
	Constrained	7.17	1.57	45.3	1.20	1.27	0.541	0.553
	Weighted	6.85	2.01	43.2	1.05	1.37	0.544	0.566
	Anisotropic	7.00	3.90	43.6	1.12	1.32	0.553	0.567
ανομοιόμορφη	Delaunay	7.09	2.89	45.4	1.08	1.43	0.514	0.526
	Constrained	7.27	3.52	45.9	1.11	1.45	0.551	0.567
	Weighted	7.71	3.98	48.3	1.13	1.51	0.476	0.465
	Anisotropic	7.47	6.60	45.4	1.13	1.20	0.532	0.551

Πίνακας 3.3: Αποτελέσματα του πειράματος αναζήτησης εικόνων για διαφορετικά α-σχήματα σε λεξικό 50K λέξεων. Το πλήθος των χαρακτηριστικών αναφέρεται σε όλο το σύνολο εικόνων. Ο χρόνος ανίχνευσης είναι ο μέσος χρόνος ανά εικόνα, ενώ ο χρόνος ερωτήματος είναι ο μέσος χρόνος ανά εικόνα-ερώτημα.

προτεινόμενου αλγορίθμου WαSH, χρησιμοποιώντας περίπου τον ίδιο αριθμό χαρακτηριστικών. Για να συγκρίνουμε περαιτέρω τις δύο μεθόδους, πραγματοποιήσαμε ένα επιπλέον πείραμα, όπου ο αριθμός χαρακτηριστικών μειώνεται δραστικά (από περίπου 7×10^6 σε περίπου 3×10^6). Στόχος του πειράματος είναι να διερευνήσουμε αν οι μέθοδοι μπορούν να διαχειριστούν ακόμα μεγαλύτερης κλίμακας σύνολα εικόνων, χωρίς την αντίστοιχη αύξηση σε απαιτήσεις μνήμης και χρόνου. Μειώνοντας τον αριθμό των χαρακτηριστικών,

Ανιχνευτής		Πλήθος χαρακτηριστικών ($\times 10^6$)	Ανεστροφικό αρχείο (MB)		Μέση ακρίβεια με BoW (mAP)		Μέση ακρίβεια με FastSM (mAP)	
			50K	100K	50K	100K	50K	100K
MFD		2.59	18.8	20.4	0.516	0.534	0.517	0.537
ομοιόμορφη	Delaunay	3.13	20.8	22.2	0.530	0.544	0.523	0.537
	Constrained	3.06	20.3	21.7	0.537	0.552	0.524	0.548
	Weighted	3.09	20.5	21.9	0.527	0.546	0.520	0.543
	Anisotropic	3.27	21.4	22.8	0.532	0.552	0.537	0.563
ανομοιόμορφη	Delaunay	3.19	21.6	23.0	0.522	0.532	0.530	0.538
	Constrained	3.07	20.5	21.8	0.539	0.549	0.529	0.542
	Weighted	2.96	20.0	21.4	0.469	0.486	0.460	0.476
	Anisotropic	3.10	20.1	21.5	0.511	0.531	0.518	0.531

Πίνακας 3.4: Αποτελέσματα του πειράματος αναζήτησης εικόνων για διαφορετικά α-σχήματα συγκρίνοντας με τον ανιχνευτή MFD. Ο αριθμός των χαρακτηριστικών είναι χαμηλότερος (περίπου 3×10^6 χαρακτηριστικά για όλο το σύνολο εικόνων). Κατασκευάζουμε οπτικά λεξικά με μικρότερο μεγέθος (50K και 100K λέξεις) για να αποφύγουμε την υπερπροσαρμογή.

μειώνουμε και τα μεγέθη των οπτικών λεξικών που εξετάζουμε, ώστε να αποφύγουμε την υπερπροσαρμογή στα δεδομένα (overfitting) κατά τη συσταδοποίηση. Τα λεξικά που κατασκευάζουμε είναι μεγέθους 50K και 100K οπτικών λέξεων. Η απόδοση όλων των παραλλαγών του προτεινόμενου αλγορίθμου είναι υψηλή, ειδικά με το λεξικό των 100K οπτικών λέξεων, όπως φαίνεται στον Πίνακα 3.4. Και πάλι τα ανιστροπικά σταθμισμένα α-σχήματα ξεπερνούν σε απόδοση τις υπόλοιπες μεθόδους, καθώς και τον ανιχνευτή MFD, επιβεβαιώνοντας τα προηγούμενα αποτελέσματα.

Κεφάλαιο 4

Δειγματοληψία εικόνων με διάχυση σφάλματος

Σε αυτό το κεφάλαιο παρουσιάζουμε μια νέα μέθοδο για δειγματοληψία εικόνων, που βασίζεται σε κατωφλίωση και διάχυση σφάλματος σε συνεχείς συναρτήσεις του πεδίου της εικόνας, πέρα από τη φωτεινότητα. Εξάγουμε δείγματα πάνω σε περιγράμματα εικόνων που αντιστοιχούν στα σχήματα που απεικονίζονται σε αυτές. Η πυκνότητα δειγματοληψίας δεν είναι σταθερή, ούτε καθορίζεται ρητά από κάποια εξωτερική παράμετρο. Αντιθέτως, καθορίζεται από τις συναρτήσεις του πεδίου εικόνας που χρησιμοποιούμε, όπως η κλίση και η απόκριση της Hessian. Τις προτεινόμενες μεθόδους δειγματοληψίας τις χρησιμοποιούμε σαν αρχικό στάδιο του αλγορίθμου εξαγωγής χαρακτηριστικών WaSH. Ελέγχουμε την επιρροή των παραμέτρων των προτεινόμενων μεθόδων στην απόδοση του WaSH σε ένα σύνολο πειραματικών αποτιμήσεων, και συγκρίνουμε την τελική απόδοση του WaSH με τους ανταγωνιστικούς ανιχνευτές της βιβλιογραφίας.

4.1 Εισαγωγή

Στο κεφάλαιο 3 περιγράψαμε δύο μεθόδους για δειγματοληψία σημείων σε εικόνες, διασχίζοντας δυαδικές ακμές με σταθερό ή μεταβλητό βήμα. Η χρήση σταθερού βήματος είναι αρκετά γρήγορη διαδικασία, αλλά απαιτεί τον προσδιορισμό του βήματος με μια παράμετρο εισόδου. Η δειγματοληψία και κατά συνέπεια τα αποτελέσματα του αλγορίθμου WaSH επηρεάζονται σημαντικά από την τιμή του βήματος. Χρησιμοποιώντας το μεταβλητό βήμα ξεπερνάμε τον περιορισμό αυτό και καταφέρνουμε να καθορίζουμε την πυκνότητα δειγματοληψίας με βάση το τοπικό σχήμα, αποδίδοντας με ακρίβεια μικρές αλλαγές του σχήματος, χωρίς περιττά σημεία σε ομοιόμορφες περιοχές. Ωστόσο, η διαδικασία της ανομοιόμορφης δειγματοληψίας επί ακμών έχει αυξημένο υπολογιστικό κόστος (βλ. Πίνακα 3.2).

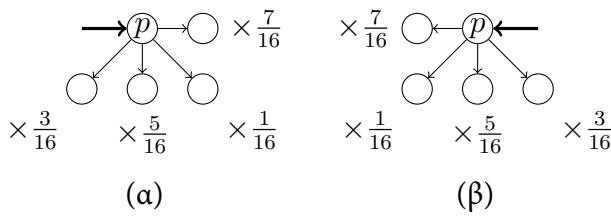
Για να ξεπεράσουμε το πρόβλημα της σταθερής πυκνότητας δειγματοληψίας, και να διατηρήσουμε το υπολογιστικό κόστος χαμηλό, προτείνουμε μια νέα προσέγγιση για τη δειγματοληψία σημείων σε εικόνες. Η μέθοδος βασίζεται στη μεθοδολογία κβαντισμού εικόνων και στη διάχυση σφάλματος με τον αλγόριθμο Floyd-Steinberg [36].

Ο αλγόριθμος Floyd-Steinberg ήταν ο πρώτος αλγόριθμος για διάχυση σφάλματος, όπου στόχος είναι η μετατροπή μονόχρωμων εικόνων σε δυαδικές, με τη δημιουργία μοτίβων από εικονοστοιχεία που διατηρούν σταθερή τη μέση φωτεινότητα στις περιοχές της εικόνας από την αρχική στην τελική εικόνα. Ο αλγόριθμος συγκρίνει τις τιμές φωτεινότητας των εικονοστοιχείων με ένα σταθερό κατώφλι και διαχέιτει το σφάλμα που προκύπτει από τη μετατροπή των αρχικών τιμών, σε μια γειτονιά εικονοστοιχείων, χρησιμοποιώντας προκαθορισμένα βάρη. Ο αλγόριθμος έχει χαμηλό υπολογιστικό κόστος, ενώ η ποιότητα των αποτελεσμάτων είναι αρκετά καλή.

Ο αλγόριθμος Floyd-Steinberg έχει μελετηθεί εκτεταμένα στη βιβλιογραφία. Ενδεικτικά, ο Ostromoukhov [84] και οι Zhuand και Fang [125] αντιμετωπίζουν μειονεκτήματα του αρχικού αλγορίθμου, όπως τα επαναλαμβανόμενα μοτίβα που δημιουργούνται σε ιδιαίτερα φωτεινές ή σκοτεινές περιοχές ομοιόμορφης φωτεινότητας, μεταβάλλοντας τα βάρη της διάχυσης σφάλματος με τιμές που εξαρτώνται από τη φωτεινότητα σε κάθε θέση. Πρόσφατα, οι Pang et al. [85] πρότειναν έναν επαναληκτικό αλγόριθμο για “dithering” εικόνων που αναλύει τις δομές τις εικόνας και διατηρεί την τοπική υφή, ενώ περιλαμβάνει ένα βήμα βελτιστοποίησης με απαγορευτικό, όμως, υπολογιστικό κόστος. Στη μέθοδο που προτείνουμε, χρησιμοποιούμε τον αρχικό αλγόριθμο λόγω του χαμηλού υπολογιστικού κόστους και της φύσης του εξεταζόμενου προβλήματος, που δεν είναι η απόδοση εικόνων μισού τόνου, αλλά η δειγματοληψία. Επίσης, εφαρμόζουμε τη διάχυση σφάλματος σε συναρτήσεις της φωτεινότητας, και όχι στις αρχικές τιμές φωτεινότητας, όπου τα επαναλαμβανόμενα μοτίβα περιορίζονται.

Η εργασία μας σχετίζεται και με την εργασία των Gu et al. [43], οι οποίοι ανιχνεύουν τοπικά χαρακτηριστικά από τα τοπικά μέγιστα και ελάχιστα της β -σταθερής Λαπλασιανής (β -stable Laplacian). Συνδυάζουν τα τοπικά χαρακτηριστικά για να κατασκευάσουν μια αναπαράσταση υψηλότερου επιπέδου, που έχει ομοιότητες με το μοντέλο αστερισμού (constellation model) [43, 34]. Ωστόσο, στην προτεινόμενη μέθοδο δεν εξάγουμε δείγματα από εικόνες σαν χαρακτηριστικά σημεία της εικόνας, αλλά εξετάζουμε τη δειγματοληψία σημείων για την αρχικοποίηση του αλγορίθμου WaSH.

Στις επόμενες ενότητες παρουσιάζουμε δύο μεθόδους δειγματοληψίας εικόνων με μεταβλητή πυκνότητα, πάνω σε συνεχείς συναρτήσεις της φωτεινότητας. Εξετάζουμε την επίδραση των διαφορετικών παραμέτρων των μεθόδων στο μέγεθος του συνόλου σημείων, καθώς και στην ποιότητα της δειγματοληψίας, μετρώντας την απόδοση του αλγορίθμου WaSH. Επίσης, εξετάζουμε τη βελτίωση που προσφέρει η εξαγωγή βαρών στα ση-



Εικόνα 4.1: Συντελεστές διάχυσης σφάλματος του αλγορίθμου Floyd-Steinberg, κατά την προσπέλαση των εικονοστοιχείων (α) από αριστερά προς τα δεξιά, και (β) από δεξιά προς τα αριστερά.

μεία της δειγματοληψίας για την κατασκευή της τριγωνοποίησης, χρησιμοποιώντας De-launay και κανονική τριγωνοποίηση, όπως στην ενότητα 3.2.

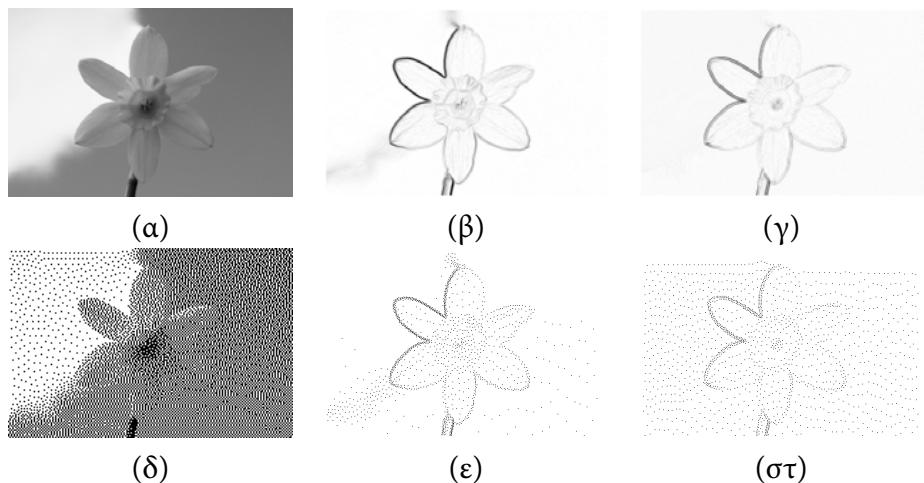
4.1.1 Κβαντισμός εικόνων με διάχυση σφάλματος

Ο κβαντισμός εικόνων με διάχυση σφάλματος είναι μια συνήθης τεχνική που χρησιμοποιείται για τη μετατροπή εικόνων με πολλές τιμές φωτεινότητας σε δυαδική μορφή. Περιλαμβάνει ένα βήμα κατωφλίωσης, που ακολουθείται από διάχυση του σφάλματος για κάθε κατωφλιωμένο εικονοστοιχείο. Οποιοσδήποτε αλγόριθμος της μορφής αυτής επισκέπτεται κάθε εικονοστοιχείο (x, y) του λάχιστον μία φορά με συγκεκριμένη σειρά και συγκρίνει την τιμή φωτεινότητας $f(x, y)$ του εικονοστοιχείου με ένα κατώφλι θ , έτσι ώστε να προκύψει η νέα τιμή

$$f'(x, y) = \begin{cases} 1 & \text{if } f(x, y) > \theta, \\ 0 & \text{otherwise.} \end{cases} \quad (4.1)$$

Το σφάλμα $e(x, y) = f(x, y) - f'(x, y)$ που προκύπτει από τη διαδικασία κβαντισμού μεταξύ της εξόδου f' και της εισόδου f διαχέεται σε μια γειτονιά του εικονοστοιχείου (x, y) .

Ο αλγόριθμος που εισήγαγαν οι Floyd και Steinberg [36] χρησιμοποιείται πλέον στην πλειοψηφία των σχετικών εφαρμογών που χρησιμοποιούν την τεχνική dithering. Ο αλγόριθμος επισκέπτεται κάθε εικονοστοιχείο μόνο μία φορά, με σειρά σερπαντίνας, δηλ. από τα αριστερά προς τα δεξιά και από τα δεξιά προς τα αριστερά με εναλλαγή ανά γραμμή της εικόνας. Το σφάλμα $e(x, y)$ διαχέεται στα 4 εικονοστοιχεία της 8-συνδεδεμένης γειτονιάς του εικονοστοιχείου (x, y) που δεν έχουν επεξεργασθεί ακόμα. Για τη διάχυση σφάλματος στις δύο διαφορετικές κατευθύνσεις επεξεργασίας, χρησιμοποιούνται οι συντελεστές που φαίνονται στο σχήμα της Εικόνας 4.1. Ο αλγόριθμος Floyd-Steinberg επιλέχθηκε για την απλότητα της υλοποίησης και το χαμηλό υπολογιστικό κόστος και για την ταυτόχρονη παραγωγή ικανοποιητικών αποτελεσμάτων.



Εικόνα 4.2: Δειγματοληψία με διάχυση σφάλματος. (α) Εικόνα εισόδου και (δ) εικόνα μετά τη χρήση του αλγορίθμου Floyd-Steinberg. (β) Κανονικοποιημένο μέτρο κλίσης \hat{G} και (ε) δειγματοληψία επί του \hat{G} . (γ) Απόκριση της Hessian $\hat{\lambda}_1$ και (στ) δειγματοληψία επί της $\hat{\lambda}_1$.

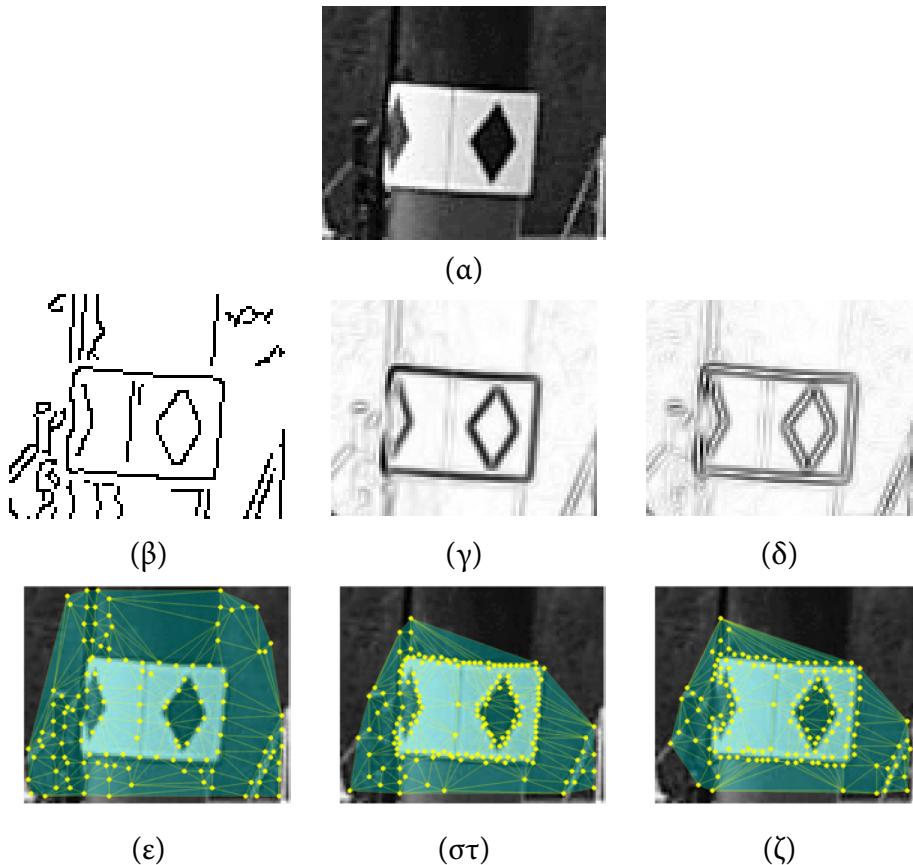
4.2 Δειγματοληψία με διάχυση σφάλματος

Σε αυτή την ενότητα προτείνουμε δύο μεθόδους δειγματοληψίας που βασίζονται στη διάχυση σφάλματος. Ο στόχος είναι να προσαρμόσουμε τη χωρική πυκνότητα των δειγμάτων επί της εικόνας και να πετύχουμε μια αραιή αναπαράσταση που θα μπορέσει να αναπαραστήσει με ακρίβεια τις δομές της εικόνας. Αίροντας τον περιορισμό τα δείγματα να βρίσκονται πάνω σε δυαδικές ακμές, αναμένουμε να λάβουμε μια λιγότερο αυστηρή αναπαράσταση των αντικειμένων μέσω της προκύπτουσας τριγωνοποίησης, που όμως θα παραμένει αρκετά περιγραφική.

Για την διάχυση σφάλματος χρησιμοποιούμε τον αλγόριθμο των Floyd και Steinberg [36], που είναι αρκετά γρήγορος καθώς απαιτεί μόνο ένα πέρασμα από τα εικονοστοιχεία και προσφέρει ικανοποιητικά αποτελέσματα. Στην περίπτωσή μας, η διάχυση σφάλματος δε θα εφαρμοστεί στη φωτεινότητα της εικόνας, αλλά σε μια βαθμωτή συνάρτηση $h(x, y)$ με πεδίο ορισμού το πεδίο της εικόνας. Οι δύο μέθοδοι που προτείνουμε βασίζονται σε διαφορετικές επιλογές της συνάρτησης h . Και στις δύο περιπτώσεις, εξάγουμε δείγματα στα σημεία που η δυαδική έξοδος του αλγορίθμου Floyd–Steinberg λαμβάνει μη μηδενικές τιμές.

Σε κάθε δείγμα $p(x, y)$ αναθέτουμε μια τιμή βάρους, ανάλογη της τιμής της συνάρτησης $h(x, y)$. Τα βάρη αυτά θα χρησιμοποιηθούν για την κατασκευή της κανονικής τριγωνοποίησης και των σταθμισμένων α -σχημάτων. Επίσης, θα εξετάσουμε και την περίπτωση μηδενικών βαρών, δηλαδή της αναπαράστασης με Delaunay τριγωνοποίηση και απλά α -σχήματα.

Για τη χρήση των νέων μεθόδων δειγματοληψίας στον αλγόριθμο του ανιχνευτή WαSH, απαιτείται μόνο η αλλαγή της διαδικασίας δειγματοληψίας, καθώς όλα τα υπόλοιπα στά-



Εικόνα 4.3: Παράδειγμα των διαφορετικών μεθόδων δειγματοληψίας με τις αντίστοιχες τριγωνοποιήσεις. (α) Εικόνα εισόδου, λεπτομέρεια της πρώτης εικόνας από το σύνολο εικόνων boat από το [78]. (β) Δυαδικός χάρτης ακμών και (ε) ομοιόμορφη δειγματοληψία επί ακμών στη (β). (γ) Κανονικοποιημένο μέτρο κλίσης και (στ) διάχυση σφάλματος στη (γ). (δ) Απόκριση της Hessian και (ζ) διάχυση σφάλματος στη (δ). (Για βελτίωση της απεικόνισης, στα (β,γ,δ) οι υψηλές αποκρίσεις φαίνονται με μαύρο χρώμα.)

δια του αλγορίθμου παραμένουν ανεπηρέαστα. Η νέα διαδικασία δειγματοληψίας φαίνεται στη Διεργασία Δειγματοληψία.

4.2.1 Διάχυση σφάλματος στην κλίση

Όπως είδαμε και στην ενότητα 3.1.1, παίρνουμε το μέτρο της κλίσης G της φωτεινότητας μιας εικόνας f από τη συνέλιξη της εικόνας με την κλίση ενός Gaussian φίλτρου $g(\sigma)$ με τυπική απόκλιση σ ,

$$G = |\nabla g(\sigma) * f|. \quad (4.2)$$

Έπειτα, όμοια με την εργασία των Yang et al. [120], αν $\hat{G}(x, y)$ είναι το μέτρο της κίσης στο σημείο (x, y) κανονικοποιημένο στο διάστημα $[0, 1]$, εφαρμόζουμε τη μη-γραμμική συνάρτηση

$$h(x, y) = \hat{G}(x, y)^{\gamma} \quad (4.3)$$

Διεργασία Δειγματοληψία(f, m)

είσοδος: μονόχρωμη εικόνα f ;
μέθοδος δειγματοληψίας m

έξοδος : σύνολο δειγμάτων P

1	εάν ($m == \text{δειγματοληψία επί ακμών}$) τότε	$\triangleright \text{δειγματοληψία επί ακμών}$
2	$\hat{G} \leftarrow \ \nabla f\ / \max\{\ \nabla f\ \}$	$\triangleright \text{κανονικοποιημένη κλίση}$
3	$E \leftarrow \text{CANNY}(\hat{G})$	$\triangleright \text{ανίχνευση ακμών}$
4	εάν ($m == \text{ομοιόμορφη δειγματοληψία}$) τότε	$\triangleright \text{ομοιόμορφη δειγματοληψία}$
5	$P \leftarrow \text{ΟμοιόμορφηΔειγματοληψία}(f, s)$	$\triangleright \text{σταθερό βήμα } s$
6	αλλιώς	$\triangleright \text{ανομοιόμορφη δειγματοληψία}$
7	$P \leftarrow \text{ΑνομοιόμορφηΔειγματοληψία}(f, s(p))$	$\triangleright \text{μεταβλητό βήμα } s(p)$
8	αλλιώς	$\triangleright \text{δειγματοληψία με διάχυνση σφάλματος}$
9	εάν ($m == \text{επί της κλήσης}$) τότε	$\triangleright \text{επί της κλήσης}$
10	$\hat{G} \leftarrow \ \nabla f\ / \max\{\ \nabla f\ \}$	$\triangleright \text{κανονικοποιημένη κλίση}$
11	$P \leftarrow \Delta\text{ιάχυσηΣφάλματος}(g^\gamma)$	
12	αλλιώς	$\triangleright \text{επί της Hessian}$
13	$H = \text{Hessian}(f)$	
14	$\hat{\lambda}_1 = \text{eig}(H) / \max\{\text{eig}(H)\}$	$\triangleright \text{κανονικοποιημένο μέτρο μέγιστης ιδιοτιμής}$
15	$P \leftarrow \Delta\text{ιάχυσηΣφάλματος}(\hat{\lambda}_1^\gamma)$	

για να αναπαραστήσουμε τα περιγράμματα των αντικειμένων της εικόνας, όπου το γ είναι μια θετική σταθερά. Εφαρμόζουμε διάχυνση σφάλματος με τον αλγόριθμο Floyd-Steinberg [36] επί της $h(x, y)$, αντί της φωτεινότητας της εικόνας $f(x, y)$. Τα μη μηδενικά στοιχεία που θα προκύψουν αποτελούν τα δείγματα που θα χρησιμοποιήσουμε για είσοδο του αλγορίθμου ανίχνευσης. Αυξάνοντας την τιμή του γ , οι τιμές της $h(x, y)$ μειώνονται, και τα τελικά δείγματα γίνονται πιο αραιά.

Σε ομαλές περιοχές της εικόνας, π.χ. στο εσωτερικό αντικειμένων ή σε ομαλό φόντο, η G παίρνει χαμηλές τιμές κι έτσι τα δείγματα είναι αραιά. Αυτό έχει ως αποτέλεσμα η τριγωνοποίηση των σημείων να έχει μεγάλα τρίγωνα σε αυτές τις ομαλές περιοχές. Αντίθετα, κοντά σε ακμές της εικόνας και γωνίες, η G παίρνει υψηλές τιμές, τα δείγματα γίνονται πιο πυκνά, και κατ' επέκταση η τριγωνοποίηση θα είναι μια πιο λεπτομερής διαμέριση του χώρου που θα μπορεί να συλλάβει σημαντικές λεπτομέρειες του σχήματος. Με τον τρόπο αυτό η πυκνότητα των δειγμάτων εξαρτάται από τη μορφολογία της εικόνας, ενώ μπορούμε να μεταβάλλουμε έμμεσα την συνολική πυκνότητα των δειγμάτων.

4.2.2 Διάχυση σφάλματος στην Hessian

Αντί της χρήσης του μέτρου της κλίσης ως εισόδου στον αλγόριθμο διάχυσης σφάλματος, οι Yang et al. [120] χρησιμοποιούν το μέτρο της μεγαλύτερης ιδιοτιμής του πίνακα Hessian σε κάθε σημείο. Εξετάζουμε, λοιπόν, και αυτή την παραλλαγή για τη δειγματοληψία.

Έπειτα από το φιλτράρισμα της εικόνας με φίλτρο Gaussian $g(\sigma)$, παίρνουμε τον πίνακα Hessian $H(x, y)$ στο σημείο (x, y) . Έστω ότι $\lambda_1(x, y)$ είναι το μέτρο της μεγαλύτερης ιδιοτιμής του $H(x, y)$, το οποίο αντιστοιχεί στη μεγαλύτερη τιμή κατευθυνόμενης παραγώγου δεύτερης τάξης της φωτεινότητας f . Όμοια με την εξίσωση 4.3, αν $\hat{\lambda}_1(x, y)$ είναι το μέτρο της μεγαλύτερης ιδιοτιμής κανονικοποιημένο στο διάστημα $[0, 1]$, εφαρμόζουμε τη διάχυση σφάλματος στη συνάρτηση

$$h(x, y) = \hat{\lambda}_1(x, y)^\gamma \quad (4.4)$$

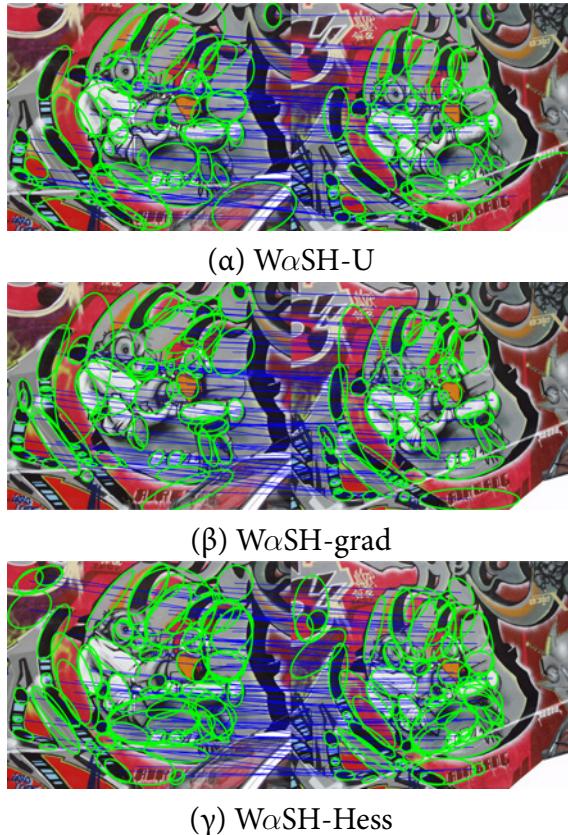
για να αναπαραστήσουμε τα όρια αντικειμένων.

Το μέτρο των παραγώγων δεύτερης τάξης αυξάνεται κοντά σε ακμές της εικόνας, με συνέπεια ο αλγόριθμος διάχυσης σφάλματος να ευνοεί την πυκνή δειγματοληψία σε τέτοιες περιοχές. Ωστόσο, τα δείγματα θα εμφανίζονται διασκορπισμένα εκατέρωθεν των ακμών, κάνοντας την τριγωνοποίηση πιο πολύπλοκη. Σε ομαλές περιοχές, η δειγματοληψία θα είναι και πάλι αραιή, αλλά επειδή οι παράγωγοι δεύτερης τάξης και κατ' επέκταση η Hessian είναι πιο ευαίσθητες στο θόρυβο, μπορεί να εμφανιστεί μια δειγματοληψία τύπου πλέγματος (βλ. Εικόνα 4.2στ). Σε σύγκριση με τη δειγματοληψία με διάχυση σφάλματος στην κλίση, συχνά ο αριθμός των τελικών χαρακτηριστικών είναι μικρότερος (βλ. ενότητα 4.3).

4.2.3 Παραδείγματα – Συζήτηση

Στην Εικόνα 4.2 βλέπουμε ένα παράδειγμα της διαδικασίας δειγματοληψίας με διάχυση σφάλματος. Στις Εικόνες 4.2β,ε βλέπουμε το κανονικοποιημένο μέτρο κλίσης \hat{G} και τη δειγματοληψία που προκύπτει. Παρατηρούμε πως τα δείγματα είναι αραιά σε ομαλές περιοχές και πυκνά σε περιοχές με δομές όπως ακμές και γωνίες. Στις εικόνες 4.2γ,στ βλέπουμε την απόκριση της Hessian $\hat{\lambda}_1$ και την προκύπτουσα δειγματοληψία. Σε αυτή την περίπτωση βλέπουμε ότι κάποιες αχνές ακμές χάνονται μέσα στο θόρυβο του φόντου. Σε όλα τα παραδείγματα έχουμε θέσει $\gamma = 1$.

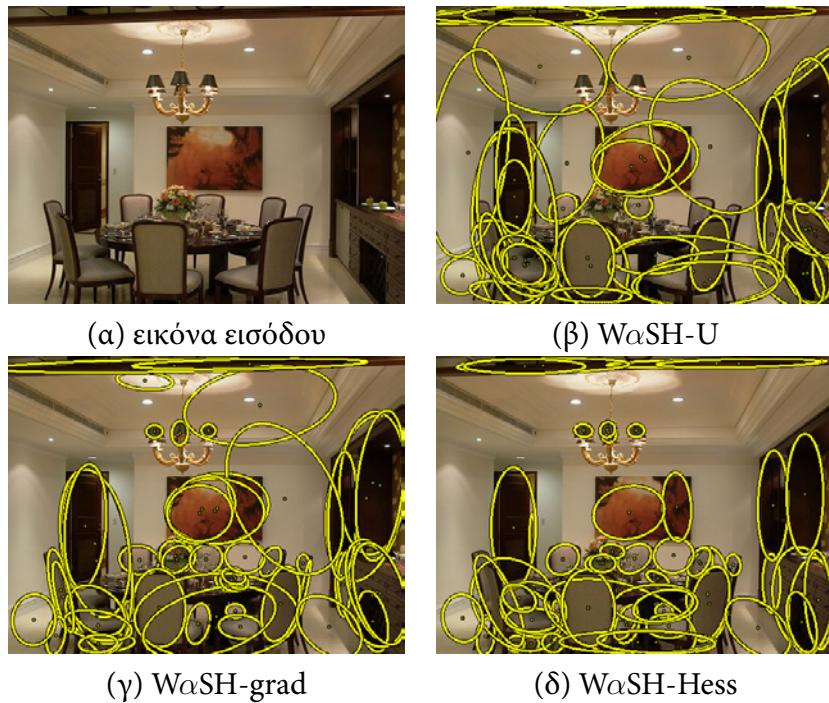
Στην Εικόνα 4.3 βλέπουμε ένα παράδειγμα διαφορετικών δειγματοληψιών και των αντίστοιχων κανονικών τριγωνοποιήσεων, σε μια λεπτομέρεια μιας εικόνας. Χρησιμοποιώντας την ομοιόμορφη δειγματοληψία επί ακμών, παίρνουμε αραιά δείγματα τα οποία είναι καλά κατανεμημένα στις ακμές, αλλά χάνουμε κάποιες λεπτομέρειες σε γωνίες και



Εικόνα 4.4: Ταίριασμα τοπικών χαρακτηριστικών στο σύνολο δεδομένων graffiti, χρησιμοποιώντας τα χαρακτηριστικά WaSH με: (α) ομοιόμορφη δειγματοληψία ακμών (*U*), (β) δειγματοληψία επί της κλίσης (*grad*), και (γ) δειγματοληψία επί της *Hessian* (*Hess*).

τμήματα ακμών με μεγάλη καμπυλότητα. Αντίθετα, χρησιμοποιώντας τις μεθόδους με διάχυση σφάλματος παίρνουμε πιο πυκνά δείγματα, αλλά διατηρούμε την πληροφορία για το σχήμα των αντικειμένων της εικόνας καλύτερα. Στη μέθοδο με διάχυση σφάλματος στην *Hessian* βλέπουμε δείγματα εκατέρωθεν των ακμών, γεγονός που σε κάποιες εφαρμογές μπορεί να αποδειχθεί χρήσιμο στην καλύτερη ανακατασκευή της εικόνας αντλώντας πληροφορία μόνο από τα δείγματα. Αντίθετα, στη μέθοδο με διάχυση σφάλματος στην κλίση εξάγονται δείγματα σε θέσεις με έντονη κλίση, που αντιστοιχούν σε ακμές τις εικόνας και όρια αντικειμένων. Αναμένουμε η δειγματοληψία στην κλίση να ταιριάξει καλύτερα με τον ανιχνευτή περιοχών WaSH, καθώς ο ανιχνευτής εντοπίζει ομοιόμορφες περιοχές που περικλείονται από όρια αντικειμένων.

Στις Εικόνες 4.4 και 4.5 βλέπουμε παραδείγματα των τελικών χαρακτηριστικών που εξάγουμε με τον ανιχνευτή μας, χρησιμοποιώντας είτε την ομοιόμορφη δειγματοληψία επί ακμών, είτε τις δύο δειγματοληψίες με διάχυση σφάλματος. Σε κάθε παράδειγμα, ο αριθμός των εξαγώμενων χαρακτηριστικών για κάθε μέθοδο είναι περίπου ο ίδιος (350 χαρακτηριστικά για το παράδειγμα της Εικόνας 4.4 και 50 για το παράδειγμα της Εικόνας 4.5). Στην Εικόνα 4.4 βλέπουμε τα χαρακτηριστικά που έχουν ταιριάξει μεταξύ δύο εικόνων του συνόλου δεδομένων *graffiti* του [78]. Για το ταίριασμα των χαρακτηριστικών χρησιμοποι-



Εικόνα 4.5: Παράδειγμα τοπικών χαρακτηριστικών $W\alpha SH$ με διαφορετικές δειγματοληψίες. (α) Εικόνα εισόδου και χαρακτηριστικά $W\alpha SH$ με: (β) ομοιόμορφη δειγματοληψία ακμών (U), (γ) δειγματοληψία επί της κλίσης ($grad$), και (δ) δειγματοληψία επί της Hessian ($Hess$).

ήθηκαν περιγραφείς SIFT. Χρησιμοποιώντας τη δειγματοληψία με διάχυση σφάλματος στην κλίση ανιχνεύονται πιο λεπτομερείς περιοχές, οι οποίες τελικά ταιριάζονται μεταξύ των εικόνων. Με τη δειγματοληψία με διάχυση σφάλματος στην Hessian, πετυχαίνουμε περισσότερα ταιριάσματα μεγαλύτερων ομογενών περιοχών (blobs). Στο παράδειγμα της Εικόνας 4.5, η εικόνα εισόδου ανήκει στο σύνολο δεδομένων PASCAL VOC 2007 [33], το οποίο χρησιμοποιείται εκτενώς για την πειραματική αποτίμηση αλγορίθμων αναγνώρισης αντικειμένων. Και σε αυτό το παράδειγμα βλέπουμε ότι χρησιμοποιώντας δειγματοληψία με διάχυση σφάλματος εντοπίζονται πιο λεπτομερείς περιοχές ενδιαφέροντος (βλ. το φωτιστικό στην οροφή και τις καρέκλες), που βοηθούν στη βελτίωση της απόδοσης σε εφαρμογές αναγνώρισης αντικειμένων.

4.3 Πειραματική αξιολόγηση

Σε αυτή την ενότητα θα αποτιμήσουμε την επίπτωση που έχουν στην απόδοση του αλγορίθμου ανίχνευσης τοπικών χαρακτηριστικών $W\alpha SH$ οι προτεινόμενες μέθοδοι δειγματοληψίας με διάχυση σφάλματος που προτείνουμε. Πραγματοποιούμε τα δύο πειράματα που παρουσιάσαμε στην ενότητα 3.5 (ταίριασμα και ανάκτηση εικόνων), όπου συγκρίνουμε και με τους καλύτερους ανιχνευτές της βιβλιογραφίας. Σε όλα τα πειράματα της ενότητας, χρησιμοποιούμε κανονική τριγωνοποίηση και σταθμισμένα α -σχήματα.

Αρχικά ελέγχουμε την επίδραση στην απόδοση του αλγορίθμου WαSH, αλλάζοντας τις τιμές των παραμέτρων των προτεινόμενων μεθόδων δειγματοληψίας. Για τη ρύθμιση των παραμέτρων δειγματοληψίας χρησιμοποιούμε το πείραμα ανάκτησης εικόνων στο σύνολο δεδομένων Oxford 5K [89]. Ξεκινώντας από το μη γραμμικό συντελεστή γ , ελέγχουμε την επίδρασή του: α) στον αριθμό των εξαγώμενων δειγμάτων, β) στο χρόνο που απαιτεί ο συνολικός αλγόριθμος για να ανιχνεύσει τοπικά χαρακτηριστικά, και γ) στην απόδοση του αλγορίθμου στο πείραμα ανάκτησης εικόνων. Δεδομένης της βέλτιστης τιμής για το γ , ελέγχουμε την επίδραση του κατωφλίου θ του βήματος της διάχυσης σφάλματος. Για να συγκρίνουμε με την ομοιόμορφη δειγματοληψία επί ακμών που παρουσιάστηκε στην ενότητα 3.1.2, αποτιμούμε την απόδοση του αλγορίθμου για διαφορετικές τιμές του βήματος δειγματοληψίας s . Επίσης, εξετάζουμε και την επίδραση των βαρών στα δείγματα, χρησιμοποιώντας και Delaunay τριγωνοποίηση με απλά α -σχήματα (μηδενίζοντας τα βάρη των δειγμάτων).

Δεδομένων των βέλτιστων τιμών των παραμέτρων των μεθόδων δειγματοληψίας, συγκρίνουμε την απόδοση του αλγορίθμου με τους καλύτερους ανιχνευτές της βιβλιογραφίας, χρησιμοποιώντας και τα δύο πειράματα. Στο πείραμα ανάκτησης εικόνων χρησιμοποιούμε το σύνολο δεδομένων Oxford 5K, καθώς και το Paris [90], το οποίο περιλαμβάνει εικόνες κτηρίων ως ερωτήματα, όπως και το Oxford 5K. Το σύνολο Paris χρησιμοποιείται μόνο για τον έλεγχο της απόδοσης του αλγορίθμου, καθώς όλες οι παράμετροι έχουν βελτιστοποιηθεί στο σύνολο Oxford 5K. Οι υπόλοιπες παράμετροι του πειράματος (αριθμός χαρακτηριστικών, περιγραφείς και μέθοδος ταιριάσματος) παραμένουν ίδιες με τα πειράματα της ενότητας 3.5.

4.3.1 Βελτιστοποίηση παραμέτρων

Για να διερευνήσουμε τον αντίκτυπο των διαφορετικών παραμέτρων των μεθόδων δειγματοληψίας με διάχυση σφάλματος, αποτιμούμε την απόδοση του αλγορίθμου ανίχνευσης στο πείραμα ανάκτησης εικόνων στο σύνολο Oxford 5K. Χρησιμοποιούμε λεξικό 200K οπτικών λέξεων και αναπαριστούμε τις εικόνες με το μοντέλο BoW.

Ο συντελεστής μη γραμμικότητας γ των εξισώσεων 4.3 και 4.4 επηρεάζει τις τιμές της συνάρτησης $h(x, y)$. Για $\gamma > 1$, οι χαμηλές τιμές της $h(x, y)$ μειώνονται περαιτέρω, καθώς οι $\hat{G}(x, y)$ και $\hat{\lambda}_1(x, y)$ έχουν κανονικοποιηθεί στο διάστημα $[0, 1]$. Το γεγονός αυτό έχει ως αποτέλεσμα η συνάρτηση $h(x, y)$ να γίνει πιο επιλεκτική και λιγότερο ομαλή, μοιάζοντας περισσότερο με συνάρτηση δυαδικών ακμών. Συνολικά εξάγονται λιγότερα δείγματα για δεδομένο κατώφλι θ , τα οποία συγκεντρώνονται στις ακμές της εικόνας. Αντίθετα, για $\gamma < 1$ οι χαμηλές τιμές της $h(x, y)$ αυξάνονται, με αποτέλεσμα να γίνεται πιο ομαλή η $h(x, y)$ και να εξάγονται περισσότερα δείγματα σε όλο το επίπεδο της εικόνας.

Για διαφορετικές τιμές του $\gamma \in [0.5, 2.0]$, εξετάζουμε το μέσο αριθμό δειγμάτων ανά

εικόνα του συνόλου δεδομένων (βλ. Εικόνα 4.6α), καθώς και το χρόνο που απαιτείται για να εξάγει ο αλγόριθμος τα τελικά χαρακτηριστικά (βλ. Εικόνα 4.6β). Για κάθε τιμή του συντελεστή γ πραγματοποιύμε το πείραμα ανάκτησης εικόνων, εξάγοντας χαρακτηριστικά και περιγραφείς για κάθε εικόνα, και κατασκευάζοντας λεξικά οπτικών λέξεων.¹ Έτσι, αποτιμούμε την επίδραση της παραμέτρου στην τελική απόδοση του αλγορίθμου ανάκτησης εικόνων (βλ. Εικόνα 4.6γ).

Η παράμετρος γ επηρεάζει σημαντικά τον αριθμό των δειγμάτων, και ως συνέπεια το χρόνο εξαγωγής των χαρακτηριστικών, ενώ η επίδραση στο χρόνο δειγματοληψίας είναι μικρή. Χαμηλές τιμές της παραμέτρου γ οδηγούν στη δειγματοληψία μεγαλύτερου αριθμού σημείων και την κατασκευή πιο σύνθετων αναπαραστάσεων (τριγωνοποιήσεων και α -σχημάτων).

Από τη διαδικασία προκύπτει ότι η απόδοση του αλγορίθμου, όπως αυτή αποτιμάται στο πείραμα ανάκτησης εικόνων στο σύνολο Oxford από τη μετρική mAP, μεγιστοποιείται για $\gamma = 1$. Αυτή η τιμή αντιστοιχεί στη γραμμική περίπτωση, όπου η πυκνότητα της αναπαράστασης και η πολυπλοκότητα του αλγορίθμου βρίσκονται σε ισορροπία. Τα αποτελέσματα βρίσκονται σε συμφωνία για τις περιπτώσεις δειγματοληψίας με διάχυση σφάλματος στην κλίση και στην Hessian. Στα υπόλοιπα πειράματα θέτουμε $\gamma = 1$ και για τις δύο περιπτώσεις. Παρόλα αυτά, θεωρούμε ότι ο μη-γραμμικός συντελεστής είναι χρήσιμος, καθώς μπορεί να βελτιώσει την απόδοση σε άλλο σύνολο εικόνων, ή σε περιπτώσεις δειγματοληψίας για άλλες εφαρμογές, πέρα από το συγκεκριμένο ανιχνευτή τοπικών χαρακτηριστικών.

Το κατώφλι θ του αλγορίθμου διάχυσης σφάλματος ελέγχει άμεσα τον αριθμό των δειγμάτων που θα εξαχθούν, μικραίνοντας το σύνολο καθώς το θ αυξάνεται (βλ. Εικόνα 4.6δ). Η πυκνότητα των δειγμάτων επηρεάζει σημαντικά το χρόνο εξαγωγής χαρακτηριστικών (βλ. Εικόνα 4.6ε).

Χρησιμοποιώντας τη δειγματοληψία επί της Hessian, η απόδοση του WαSH πέφτει καθώς μειώνεται η πυκνότητα των δειγμάτων (βλ. Εικόνα 4.6στ). Επιλέγουμε την τιμή $\theta = 0.5$ ώστε να διατηρήσουμε την υψηλή απόδοση, παρά το υψηλό υπολογιστικό κόστος που προκύπτει για την εξαγωγή χαρακτηριστικών. Στη δειγματοληψία επί της κλίσης η απόδοση του αλγορίθμου δεν επηρεάζεται σημαντικά μέχρι την τιμή $\theta = 5$, παρουσιάζοντας μέγιστο για $\theta = 2$, τιμή που χρησιμοποιούμε στα υπόλοιπα πειράματα.

Στην Εικόνα 4.7 βλέπουμε τα δείγματα που εξάγονται για διαφορετικές τιμές του κατωφλίου θ , του βήματος δειγματοληψίας s και της μέγιστης εκκεντρότητας k , για δειγματοληψία: (α) με διάχυση σφάλματος στην κλίση, (β) διάχυση σφάλματος στην Hessian, (γ) ομοιόμορφη επί ακμών, και (δ) ανομοιόμορφη επί ακμών. Η εικόνα είναι μια λεπτομέρεια από εικόνα του συνόλου Oxford 5K. Για όλες τις μεθόδους η πυκνότητα δειγματοληψίας αυξάνεται από τα αριστερά προς τα δεξιά. Με τη δειγματοληψία επί της κλίσης, οι δομές

της εικόνας παραμένουν ιδιαίτερα εμφανείς ακόμα και σε αραιές δειγματοληψίες. Επίσης, με την ανομοιόμορφη δειγματοληψία βλέπουμε τα δείγματα να αραιώνουν σε ευθύγραμμα τμήματα ακμών, ενώ διατηρούν αυξημένη πυκνότητα σε ακμές με μεγάλη καμπυλότητα.

Σύγκριση με τη δειγματοληψία επί ακμών. Με στόχο να συγκρίνουμε τις μεθόδους δειγματοληψίας με διάχυση σφάλματος με τη δειγματοληψία επί ακμών, εξετάζουμε την απόδοση του αλγορίθμου WαSH χρησιμοποιώντας ομοιόμορφη δειγματοληψία επί ακμών (U) με διαφορετικά πλήθη δειγμάτων. Για να επηρεάσουμε τον αριθμό δειγμάτων μεταβάλλουμε το βήμα δειγματοληψίας s στο διάστημα [1, 17].

Ο αριθμός των δειγμάτων μειώνεται εκθετικά ως προς το βήμα δειγματοληψίας s (βλ. Εικόνα 4.6ζ). Παρά το γεγονός ότι ο χρόνος δειγματοληψίας παραμένει σχετικά αμετάβλητος (βλ. Εικόνα 4.6η), ο συνολικός χρόνος εξαγωγής τοπικών χαρακτηριστικών μεταβάλλεται εκθετικά, καθώς η πολυπλοκότητα των κατασκευών εξαρτάται από τον αριθμό των δειγμάτων. Η απόδοση του αλγορίθμου παραμένει σταθερή για βήματα δειγματοληψίας έως $s = 11$, και μειώνεται για μεγαλύτερες τιμές (βλ. Εικόνα 4.6θ). Επιλέγουμε την τιμή $s = 11$ για τα υπόλοιπα πειράματα, ώστε να διατηρήσουμε την υψηλή απόδοση του αλγορίθμου σε συνδυασμό με τη γρήγορη εξαγωγή των χαρακτηριστικών.

Στις Εικόνες 4.6ι,α συγκρίνουμε τις μεθόδους δειγματοληψίας, εξετάζοντας την επιρροή της αλλαγής του αριθμού των δειγμάτων στην απόδοση του αλγορίθμου WαSH, καθώς και το χρόνο εξαγωγής χαρακτηριστικών. Η δειγματοληψία επί της κλίσης ξεπερνά σε απόδοση τη δειγματοληψία επί της Hessian, καθώς και την ομοιόμορφη επί ακμών, παρουσιάζοντας χαμηλό υπολογιστικό κόστος. Η δειγματοληψία επί της Hessian παρουσιάζει τη χαμηλότερη απόδοση και είναι σημαντικά πιο χρονοβόρα. Στην Εικόνα 4.6ιβ εξετάζουμε την απόδοση του αλγορίθμου, σε σχέση με το χρόνο που απαιτεί για την εξαγωγή χαρακτηριστικών. Παρατηρούμε ότι η καλύτερη σχέση μεταξύ απόδοσης και κόστους σε χρόνο επιτυγχάνεται χρησιμοποιώντας τη δειγματοληψία επί της κλίσης.

Απλά και σταθμισμένα α-σχήματα. Στον Πίνακα 4.1 βλέπουμε αποτελέσματα του πειράματος αναζήτησης εικόνων χρησιμοποιώντας τις διαφορετικές μεθόδους δειγματοληψίας σε συνδυασμό με τη Delaunay και την κανονική (Regular) τριγωνοποίηση, και τα απλά και σταθμισμένα α-σχήματα αντίστοιχα (βλ. κεφάλαιο 3). Μηδενίζοντας τα βάρη των σημείων που εξάγουμε, η τριγωνοποίηση από κανονική γίνεται Delaunay. Χρησιμοποιώντας τα σταθμισμένα σημεία και κανονική τριγωνοποίηση αυξάνεται η πολυπλοκότητα των γεωμετρικών κατασκευών (τριγωνοποίηση και α-σχήματα), απαιτώντας περισσότερους υπολογισμούς ανά σημείο. Ως αποτέλεσμα, το υπολογιστικό κόστος του αλγορίθμου εξαγωγής χαρακτηριστικών αυξάνεται κατά περίπου 5%. Ταυτόχρονα όμως, η αναπαράσταση των αντικειμένων με σταθμισμένα σημεία περιέχει περισσότερη πληροφορία από το πεδίο της εικόνας, γεγονός που αυξάνει την απόδοση του αλγορίθμου κατά περίπου 2% στη δειγματοληψία επί της κλίσης και μέχρι 4% στην ομοιόμορφη δειγματοληψία

ανιχνευτής	χαρ/κά ($\times 10^6$)	χρόνος ανίχν. (s)	Bag-of-Words (mAP)			ReRanking (mAP)			
			50K	100K	200K	50K	100K	200K	
HessAff	29.02	6.54	0.483	0.539	0.573	0.518	0.577	0.607	
MSER	13.33	0.40	0.487	0.534	0.565	0.519	0.569	0.595	
SIFT	11.13	5.24	0.422	0.465	0.495	0.441	0.486	0.517	
SURF	6.84	0.43	0.465	0.526	0.574	0.509	0.573	0.603	
Delaunay	WaSH, U	7.41	3.04	0.507	0.547	0.583	0.507	0.552	0.581
	WaSH, grad	7.42	2.91	0.537	0.569	0.598	0.539	0.565	0.591
	WaSH, Hess	7.30	14.14	0.506	0.545	0.569	0.499	0.535	0.564
Regular	WaSH, U	7.66	3.14	0.542	0.583	0.591	0.530	0.573	0.590
	WaSH, grad	7.59	3.07	0.532	0.575	0.605	0.543	0.581	0.599
	WaSH, Hess	7.30	15.14	0.507	0.559	0.582	0.515	0.555	0.570

Πίνακας 4.1: Αποτελέσματα του πειράματος αναζήτησης στο σύνολο εικόνων Oxford, χρησιμοποιώντας 3 διαφορετικά οπτικά λεξικά, το μοντέλο Bag-of-Words και χωρική επιβεβαίωση των αποτελεσμάτων. Η αποτίμηση γίνεται με τη μετρική μέσης ακρίβειας mAP.

ανιχνευτής	χαρ/κά ($\times 10^6$)	Bag-of-Words (mAP)			ReRanking (mAP)		
		50K	100K	200K	50K	100K	200K
HessAff	36.14	0.467	0.491	0.507	0.479	0.500	0.517
MSER	17.33	0.465	0.485	0.497	0.480	0.499	0.503
SIFT	25.54	0.476	0.492	0.492	0.457	0.457	0.476
SURF	8.56	0.458	0.479	0.487	0.471	0.486	0.493
WaSH, U	9.01	0.454	0.459	0.457	0.449	0.455	0.455
WaSH, grad	9.35	0.497	0.509	0.511	0.498	0.506	0.510
WaSH, Hess	9.32	0.477	0.474	0.478	0.468	0.469	0.476

Πίνακας 4.2: Αποτελέσματα του πειράματος αναζήτησης στο σύνολο εικόνων Paris. Η αποτίμηση γίνεται με τη μετρική μέσης ακρίβειας mAP.

επί ακμών.

4.3.2 Σύγκριση με τη βιβλιογραφία

Σε αυτή την ενότητα αποτιμούμε την απόδοση του ανιχνευτή σε σύγκριση με τους καλύτερους ανιχνευτές της βιβλιογραφίας χρησιμοποιώντας τις διαφορετικές μεθόδους δειγματοληψίας και στα δύο διαφορετικά πειράματα.

Ταίριασμα εικόνων

Αποτιμούμε την απόδοση του WaSH στο πείραμα ταιριάσματος εικόνων χρησιμοποιώντας τις διαφορετικές μεθόδους δειγματοληψίας, ενώ ταυτόχρονα τον συγκρίνουμε με ανταγωνιστικούς ανιχνευτές της βιβλιογραφίας. Συγκεκριμένα χρησιμοποιούμε τους ανιχνευτές Hessian-Affine, MSER, SIFT και SURF, για τους οποίους χρησιμοποιούμε τις υλοποιήσεις των αντίστοιχων συγγραφέων και τις τυπικές τιμές παραμέτρων. Εξαίρεση

Πειραματική αξιολόγηση

αποτελεί ο ανιχνευτής SIFT για τον οποίο χρησιμοποιούμε την υλοποίηση ανοιχτού κώδικα της βιβλιοθήκης VLFeat [114]. Χρησιμοποιούμε σύνολα εικόνων που ελέγχουν την απόδοση σε αλλαγές στη γωνία θέασης, την περιστροφή, την κλίμακα, τη φωτεινότητα και το θόλωμα. Για το ταίριασμα των εικόνων χρησιμοποιούμε περιγραφείς SIFT 128 διαστάσεων για όλους τους ανιχνευτές, εκτός του SURF που παρουσιάζει καλύτερη απόδοση με τον περιγραφέα SURF. Σε όλα τα πειράματα, ο αλγόριθμος WαSH χρησιμοποιεί κανονική τριγωνοποίηση και σταθμισμένα α -σχήματα (βλ. κεφάλαιο 3).

Τα αποτελέσματα της αποτίμησης παρουσιάζονται στις Εικόνες 4.8 και 4.9. Στην τελευταία γραμμή της Εικόνας 4.9 παρουσιάζεται ο μέσος όρος των αποτελεσμάτων για τα 6 σύνολα εικόνων. Για την αποτίμηση της απόδοσης εξετάζουμε τις τιμές της επαναληψιμότητας, του βαθμού ταιριάσματος και τον αριθμό των χαρακτηριστικών που εξάγονται από τον κάθε ανιχνευτή. Ο προτεινόμενος ανιχνευτής WαSH, παρουσιάζει την καλύτερη απόδοση μεταξύ όλων των ανιχνευτών χρησιμοποιώντας τη δειγματοληψία επί της κλίσης, ακολουθούμενος από τον ανιχνευτή Hessian-Affine. Και οι δύο δειγματοληψίες με διάχυση σφάλματος βελτιώνουν την απόδοση του WαSH σε σχέση με τις δειγματοληψίες επί ακμών (ομοιόμορφη και ανομοιόμορφη).

Ανάκτηση εικόνων

Σε αυτή την ενότητα συγκρίνουμε τις προτεινόμενες παραλλαγές του αλγορίθμου WαSH με τους ανταγωνιστικούς ανιχνευτές της βιβλιογραφίας. Χρησιμοποιούμε το πείραμα ανάκτησης εικόνων μεγάλης κλίμακας με τα σύνολα εικόνων Oxford 5K και Paris. Όπως και στα προηγούμενα πειράματα, συγκρίνουμε με τους ανιχνευτές Hessian-affine, MSER, SIFT και SURF. Για τους ανταγωνιστικούς ανιχνευτές χρησιμοποιούμε τα διαθέσιμα εκτελέσιμα και τυπικές παραμέτρους. Οι περιγραφείς που εξάγουμε είναι SIFT για όλους τους ανιχνευτές, εκτός από τον SURF.

Όμοια με το πείραμα βελτιστοποίησης παραμέτρων της ενότητας 4.3.1, προσαρμόζουμε τα κατώφλια επιλογής χαρακτηριστικών ώστε οι παραλλαγές του αλγορίθμου να εξάγουν περίπου τον ίδιο αριθμό χαρακτηριστικών με τον αρχικό αλγόριθμο. Για κάθε έναν από τους ανιχνευτές δημιουργούμε 3 διαφορετικά οπτικά λεξικά που περιλαμβάνουν 50K, 100K και 200K οπτικές λέξεις. Συγκρίνουμε την απόδοση με τη μέθοδο Bag-of-Words (BoW) καθώς και με χωρικό ταίριασμα των αποτελεσμάτων, τα οποία φαίνονται στον Πίνακα 4.1.

Ο αριθμός των χαρακτηριστικών που εξάγει ο κάθε ανιχνευτής είναι κρίσιμος για εφαρμογές ανάκτησης εικόνων μεγάλης κλίμακας, καθώς επηρεάζει το χρόνο δεικτοδότησης και το μέγεθος της μνήμης που απαιτείται για την αποθήκευση του ανεστραμμένου αρχείου, ενώ στις περισσότερες περιπτώσεις μειώνοντας τον αριθμό των χαρακτηριστικών η απόδοση πέφτει. Ο ανιχνευτής SURF εξήγαγε το μικρότερο αριθμό χαρακτηριστικών,

ακολουθούμενος από τις παραλλαγές του προτεινόμενου αλγορίθμου. Παρά τον σημαντικά μικρότερο αριθμό χαρακτηριστικών, η απόδοση των SURF και WaSH είναι συγκρισιμη με την απόδοση του Hessian-affine. Η αύξηση του μεγέθους του λεξικού οδήγησε σε αύξηση της απόδοσης για όλους τους ανιχνευτές. Χρησιμοποιώντας τη δειγματοληψία επί της κλίσης στον αλγόριθμο WaSH πετυχαίνουμε την καλύτερη απόδοση στο πείραμα, χρησιμοποιώντας το χωρικό ταίριασμα των αποτελεσμάτων, γεγονός που επιβεβαιώνει τα ευρήματα του προηγούμενου πειράματος. Με τη μέθοδο BoW η απόδοση της δειγματοληψίας επί της κλίσης είναι συγκρισιμη με αυτή της ομοιόμορφης δειγματοληψίας επί ακμών.

Τελικά, συγκρίνουμε όλους τους ανιχνευτές χρησιμοποιώντας το σύνολο εικόνων *Paris*. Όλες οι τιμές για τις παραμέτρους των ανιχνευτών καθώς και τα κατώφλια επιλογής χαρακτηριστικών διατηρήθηκαν ίδιες με αυτές που χρησιμοποιήθηκαν στο σύνολο εικόνων *Oxford*. Επίσης, για την εξαγωγή οπτικών λέξεων χρησιμοποιήθηκαν τα λεξικά που κατασκευάστηκαν από τις εικόνες του συνόλου *Oxford*. Η αποτίμηση γίνεται μόνο με κανονική τριγωνοποίηση με σταθμισμένα α -σχήματα, και τα αποτελέσματα παρουσιάζονται στον Πίνακα 4.2. Τα αποτελέσματα συνάδουν με αυτά των προηγούμενων πειραμάτων και συνόλων δεδομένων. Το πλήθος των χαρακτηριστικών που εξήγαγε ο ανιχνευτής SURF ήταν το μικρότερο, ακολουθούμενο από τον προτεινόμενο αλγόριθμο με μικρή διαφορά. Η δειγματοληψία επί της κλίσης παρουσίασε τα καλύτερα αποτελέσματα, ξεπερνώντας τους ανταγωνιστικούς ανιχνευτές της βιβλιογραφίας, ενώ η δειγματοληψία επί της Hessian ξεπέρασε την ομοιόμορφη δειγματοληψία επί ακμών.

4.4 Συμπεράσματα

Σε αυτό το κεφάλαιο παρουσιάσαμε μια νέα μέθοδο για δειγματοληψία σημείων από εικόνες, που στηρίζεται σε ιδέες από την απόδοση εικόνων μισού τόνου. Η δειγματοληψία χρησιμοποιεί διάχυση σφάλματος σε ομαλές συναρτήσεις της φωτεινότητας της εικόνας, και προτείνοντας τη χρήση δύο διαφορετικών συναρτήσεων προκύπτουν δύο παραλλαγές της μεθόδου.

Η προτεινόμενη μέθοδος δειγματοληψίας συνδυάστηκε με τον αλγόριθμο εξαγωγής τοπικών χαρακτηριστικών WaSH που προτείναμε στο κεφάλαιο 3. Η δειγματοληψία εικόνων είναι το πρώτο βήμα του αλγορίθμου WaSH και επηρεάζει τις ιδιότητες των χαρακτηριστικών που ανιχνεύονται, καθώς και την συνολική απόδοση του ανιχνευτή σε εφαρμογές ταιριάσματος και ανάκτησης εικόνων.

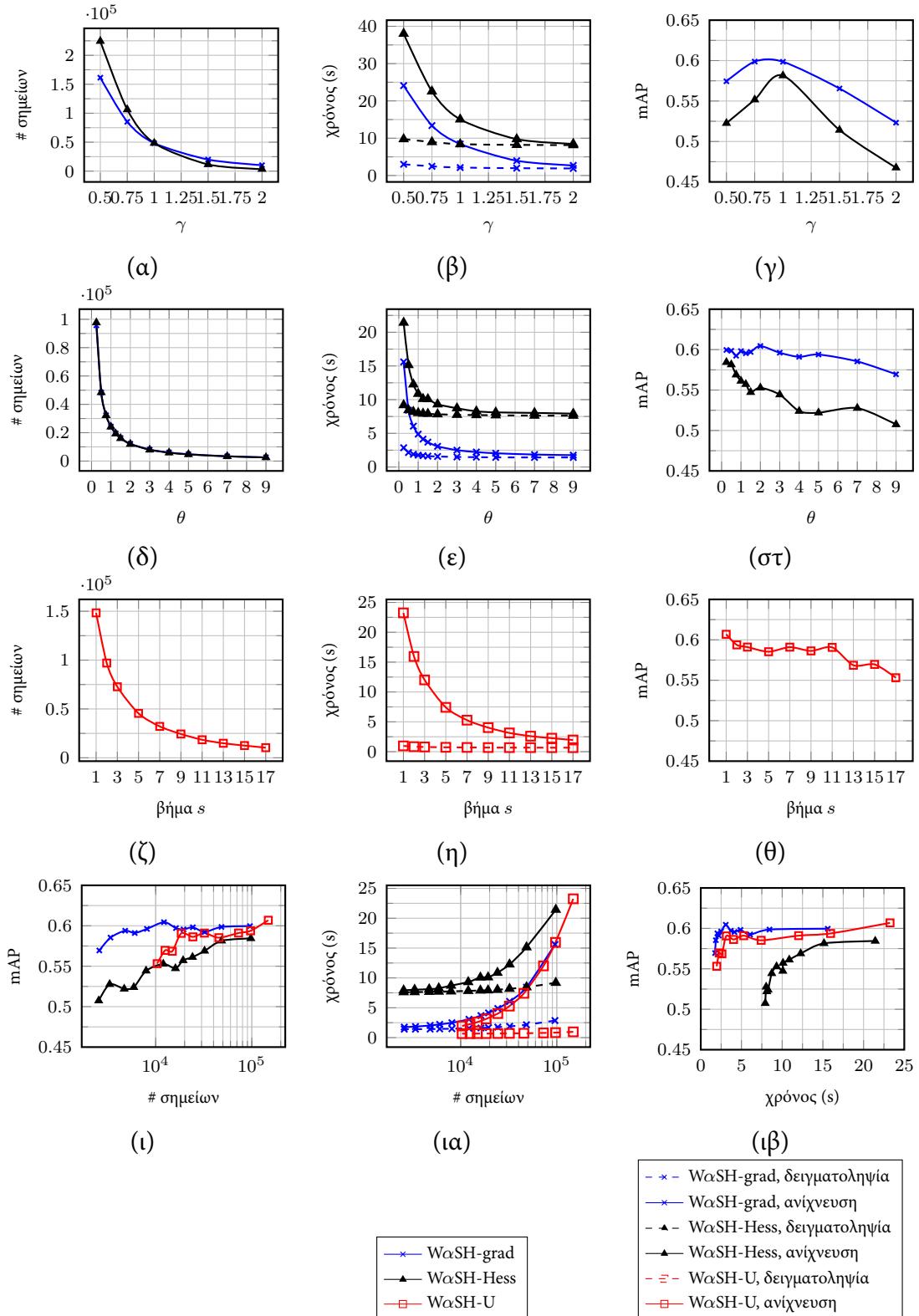
Η ποιότητα και οι ιδιότητες της προτεινόμενης δειγματοληψίας ελέγχθηκαν διεξοδικά μέσω της απόδοσης του WaSH, προσαρμόζοντας τις παραμέτρους της δειγματοληψίας και συγκρίνοντας με τους καλύτερους ανταγωνιστικούς αλγορίθμους εξαγωγής τοπικών

Συμπεράσματα

χαρακτηριστικών της βιβλιογραφίας. Η δειγματοληψία βοηθά ώστε να κατασκευαστεί μια πιο ακριβής αναπαράσταση των δομών που εμφανίζονται στις εικόνες. Ο ανιχνευτής εντοπίζει πιο λεπτομερείς δομές, ενώ διατηρεί την υψηλή κάλυψη των εικόνων.

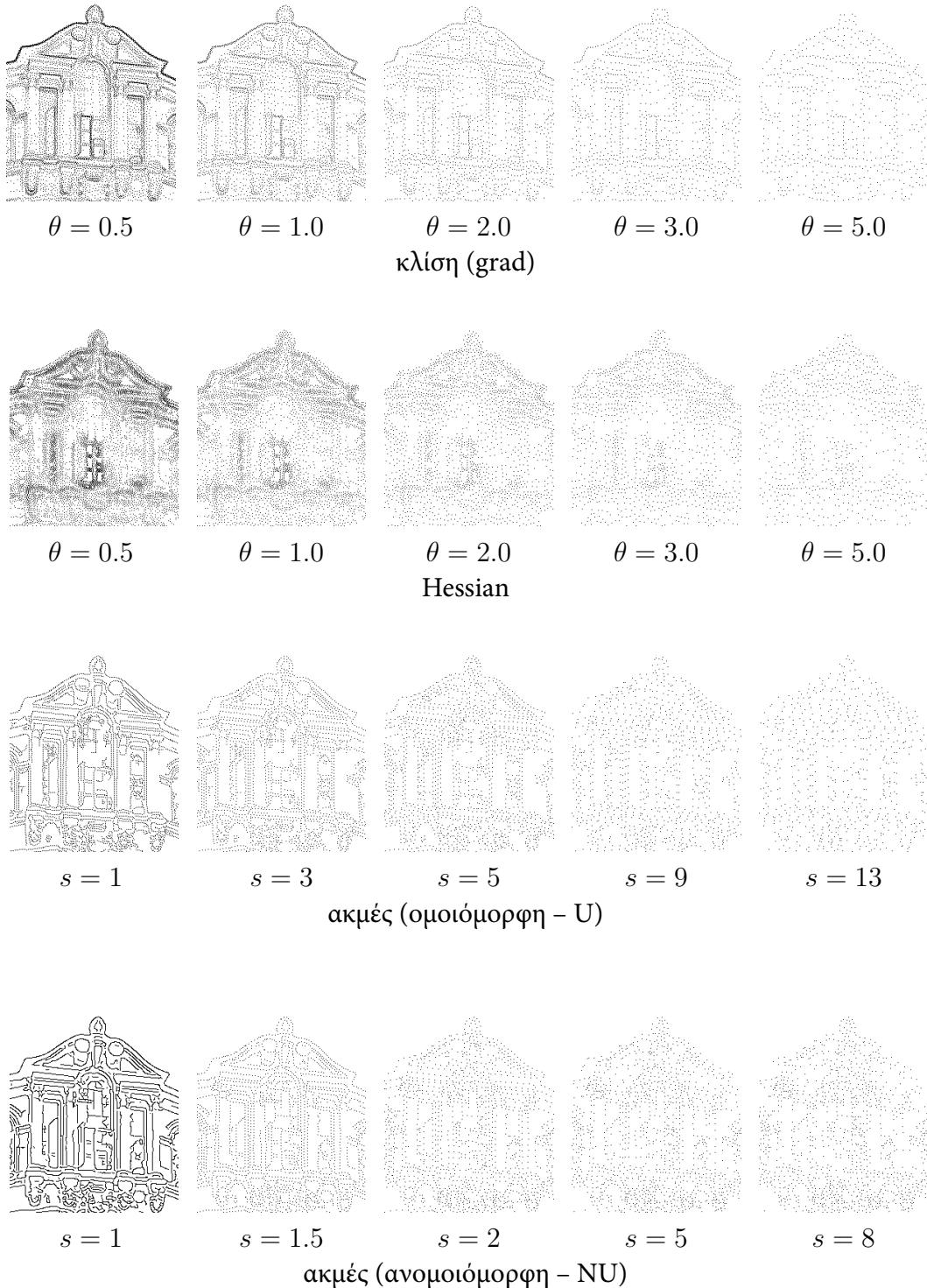
Χρησιμοποιώντας το μέτρο της κλίσης στη δειγματοληψία, εξάγουμε δείγματα σε περιοχές υψηλής κλίσης που αντιστοιχούν σε όρια αντικειμένων. Η επιλογή αυτή ταιριάζει με τον αλγόριθμο WαSH, γεγονός που αυξάνει την απόδοσή του χωρίς να αυξάνει το υπολογιστικό κόστος. Με την απόκριση της Hessian, αυξάνεται το υπολογιστικό κόστος για την εξαγωγή δειγμάτων, τα οποία διασκορπίζονται εκατέρωθεν των περιοχών υψηλής κλίσης. Η επιλογή αυτή ταιριάζει περισσότερο σε μεθόδους αναπαράστασης εικόνων με πλέγματα, παρά με τον αλγόριθμο εξαγωγής χαρακτηριστικών. Τα σημεία μπορούν να αντλήσουν οπτική πληροφορία κοντά σε κρίσιμες περιοχές, με αποτέλεσμα την ακριβή αναπαράσταση χρώματος και φωτεινότητας κοντά σε ακμές.

Κεφάλαιο 4. Δειγματοληψία εικόνων με διάχυση σφάλματος

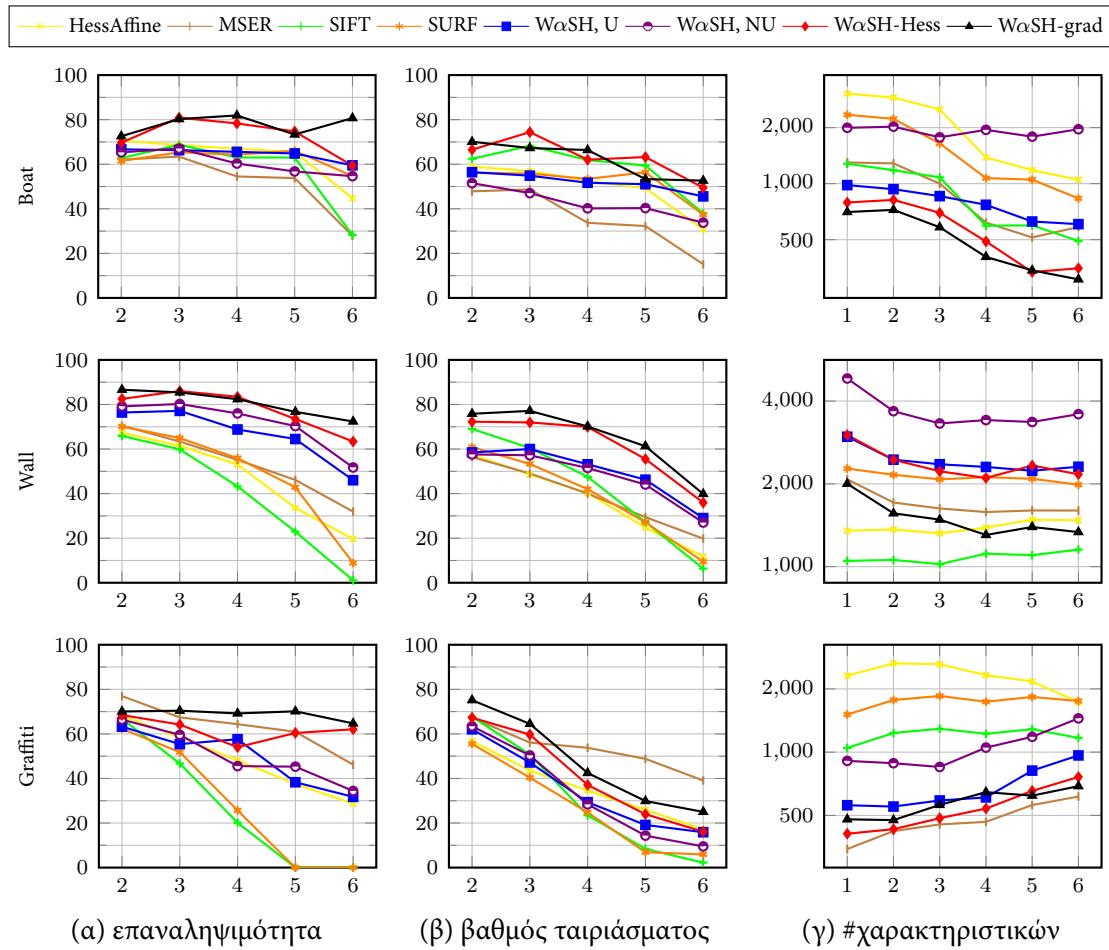


Εικόνα 4.6: Ρύθμιση των παραμέτρων των διαφορετικών μεθόδων δειγματοληψίας. Στις μεθόδους δειγματοληψίας με διάχυση σφάλματος ελέγχουμε την επίδραση της παραμέτρου γ (πρώτη γραμμή) και του κατωφλίου θ (δεύτερη γραμμή), ενώ για την ομοιόμορφη δειγματοληψία επί ακμών ελέγχουμε το βήμα δειγματοληψίας s (τρίτη γραμμή). Στην τελευταία γραμμή συγκρίνουμε την απόδοση και τη χρονική πολυπλοκότητα των διαφορετικών μεθόδων δειγματοληψίας, βάσει του αριθμού των δειγμάτων.

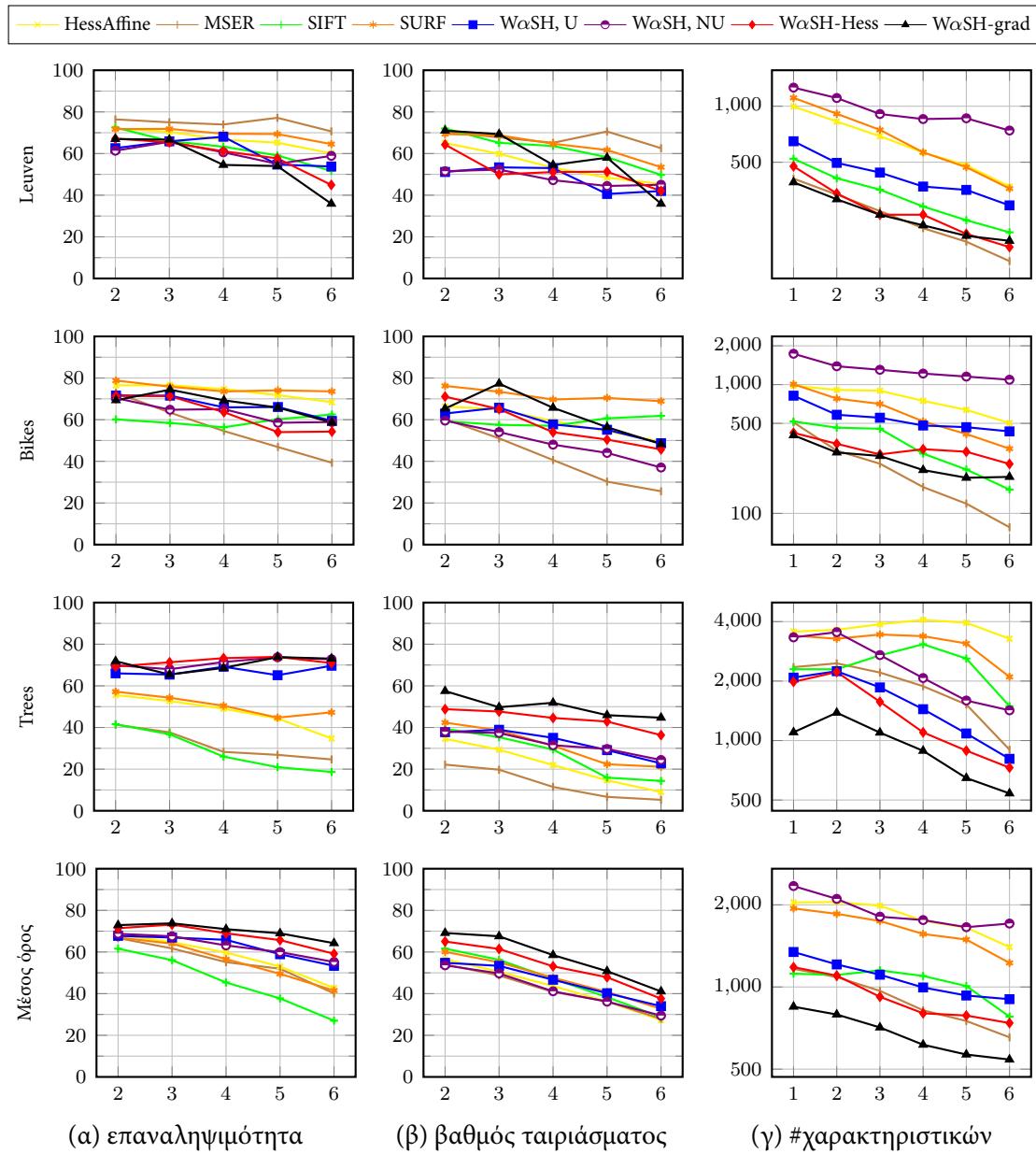
Συμπεράσματα



Εικόνα 4.7: Παραδείγματα δειγματοληψιών χρησιμοποιώντας τις προτεινόμενες μεθόδους για διαφορετικές τιμές κατωφλίων. Στην ομοιόμορφη δειγματοληψία επί ακμών μεταβάλουμε την τιμή του βήματος δειγματοληψίας s , ενώ στην ανομοιόμορφη την τιμή της μέγιστης εκκεντρότητας k .



Εικόνα 4.8: Επαναληψιμότητα και βαθμός ταιριάσματος για διαφορετικές δειγματοληψίες (U: ομοιόμορφη επί ακμών, NU: ανομοιόμορφη επί ακμών, Hess: με διάχυση σφάλματος επί της Hessian και grad: διάχυση σφάλματος επί της κλίσης), στις ακολουθίες boat, wall και graffiti. Συγκρίνονται με τους ανιχνευτές Hessian-Affine και MSER. #χαρακτηριστικών: αριθμός χαρακτηριστικών που εξάγονται σε κάθε εικόνα.



Εικόνα 4.9: Επαναληψιμότητα και βαθμός ταιριάσματος για διαφορετικές δειγματοληψίες (U: ομοιόμορφη επί ακμών, NU: ανομοιόμορφη επί ακμών, Hess: με διάχυση σφάλματος επί της Hessian και grad: διάχυση σφάλματος επί της κλίσης), στις ακολουθίες leuven, bikes και trees. Στην τελευταία γραμμή φαίνεται ο μέσος όρος όλων των ακολουθιών για κάθε μετρική. Συγκρίνουμε με τους ανιχνευτές Hessian-Affine και MSER. #χαρακτηριστικών: αριθμός χαρακτηριστικών που εξάγονται σε κάθε εικόνα.

Κεφάλαιο 5

Περιγραφή και ανάκτηση βίντεο

Σε αυτό το κεφάλαιο της διατριβής προτείνουμε ένα πλαίσιο για ανάκτηση χειρουργικών βίντεο, καθώς και ένα πρωτόκολλο για την αποτίμηση των αποτελεσμάτων της διαδικασίας. Η μέθοδος ανάκτησης περιλαμβάνει αρχικά την κατάτμηση των βίντεο σε πλάνα, λαμβάνοντας υπ' όψιν τις ιδιαιτερότητες του περιεχομένου. Έπειτα εξάγονται περιγραφείς από κάθε πλάνο, χρησιμοποιώντας βαθιά νευρωνικά δίκτυα για την εξαγωγή περιγραφέων. Για την αποτίμηση των αποτελεσμάτων προτείνουμε νέα κριτήρια ομοιότητας μεταξύ των διαφορετικών πλάνων. Τα αποτελέσματα αποδεικνύουν την υπεροχή της προτεινόμενης μεθόδου έναντι της χρήσης τοπικών χαρακτηριστικών.

5.1 Ανάκτηση σε ιατρικά βίντεο

Για την ανάκτηση ιατρικών εικόνων με βάση το περιεχόμενο έχουν προταθεί πολλές μέθοδοι [80], ενώ για την ανάκτηση βίντεο η βιβλιογραφία είναι περιορισμένη. Ενδεικτικές εργασίες περιλαμβάνουν: την επισήμανση βίντεο κολονοσκόπησης για ανίχνευση πολυπόδων [86] και την ανάκτηση βίντεο ενδομικροσκοπίας [5]. Στο συναφές πεδίο της ενδοσκοπίας με κάψουλα, έχει πραγματοποιηθεί εκτεταμένη έρευνα για την κατάτμηση βίντεο [72], καθώς και για την ανάκτηση εικόνων για ανίχνευση αλλοιώσεων και ταξινόμηση ως προς την ανατομία [13].

Οι μέθοδοι που έχουν προταθεί για την ανίχνευση και ανάκτηση χειρουργικών γεγονότων σε βίντεο ελάχιστα επεμβατικής χειρουργικής (ΕΕΧ) είναι περιορισμένες. Οι δυνατότητες της κατάτμησης εγχειρήσεων με ανάλυση βίντεο εξετάστηκε στο [10], όπου ο στόχος ήταν η εξαγωγή των φάσεων της εγχειρήσης. Η μέση ακρίβεια αναγνώρισης ήταν της τάξης του 60%, χρησιμοποιώντας Κρυφά Μαρκοβιανά Μοντέλα (Hidden Markov models – HMMs) και Δυναμική Παραμόρφωση Χρόνου (Dynamic Time Warping – DTW). Παρά το γεγονός ότι τα αποτελέσματα δεν ήταν ιδιαίτερα ενθαρρυντικά, η συγκεκριμένη εργασία παρουσιάζει το μέγεθος του προβλήματος της ανάλυσης βίντεο ΕΕΧ.

Οι Lo et al. [68] προτείνουν μια προσέγγιση για ανίχνευση γεγονότων σε βίντεο EEX που βασίζεται στη συνένωση πολλαπλών οπτικών χαρακτηριστικών, όπως το σχήμα, η παραμόρφωση και η αντανάκλαση φωτός σε ένα Bayesian πλαίσιο. Οι Primus et al. [91] προτείνουν μια μέθοδο για την κατάτμηση βίντεο EEX, ανιχνεύοντας έντονες διαφορές στην κίνηση, με ανάλυση των διαφορών διανυσμάτων κίνησης στο χώρο και το χρόνο. Οι Γιάνναρου και Yang [40] περιγράφουν ένα καινοτόμο πλαίσιο για την αναπαράσταση χειρουργικών σκηνών με βάση το περιεχόμενο. Ξεκινώντας από την εξαγωγή διανυσμάτων κίνησης, εντοπίζουν εξέχοντα κινούμενα αντικείμενα, με στόχο τη μοντελοποίηση της κίνησης οργάνων και ιστών, και την αναγνώριση χειρουργικών γεγονότων.

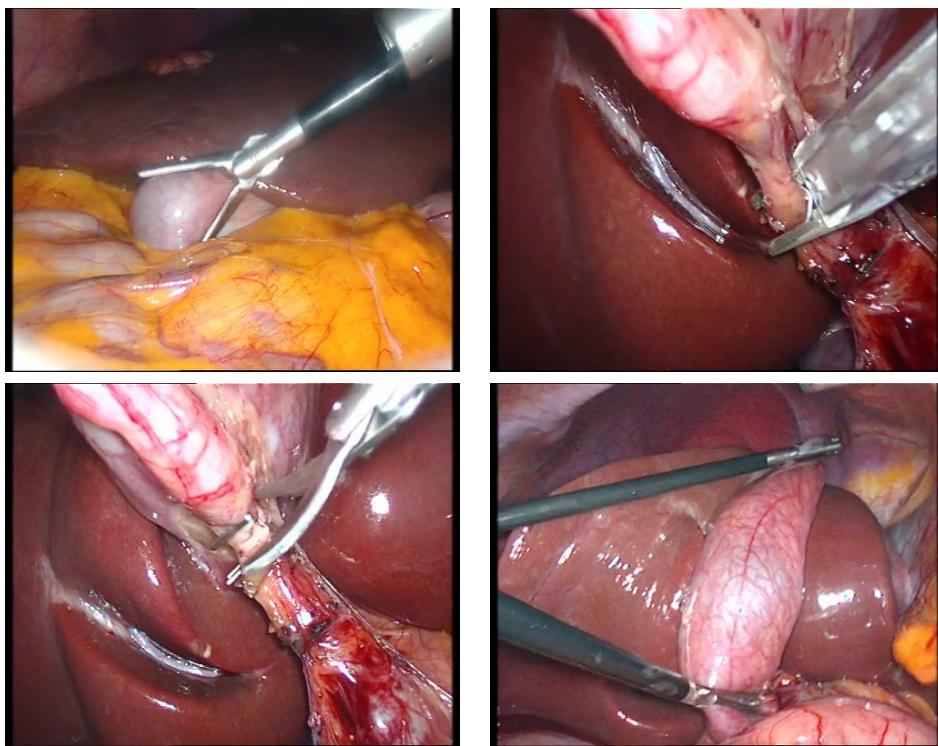
Πρόσφατα, οι Λουκάς και Γεωργίου [69] υποστήριξαν τη σημασία της επισήμανσης βίντεο EEX με βάση τα γεγονότα και ασχολήθηκαν με το πρόβλημα της ανίχνευσης καυτηριασμού ιστών. Οι Lux et al. [71] προτείνουν μια καινοτόμο προσέγγιση για τη δημιουργία περιλήψεων από αρθροσκοπικά βίντεο, χρησιμοποιώντας μια συλλογή χαρακτηριστικών καρέ. Για την επιλογή των χαρακτηριστικών καρέ, δημιουργούν συστάδες από όλα τα καρέ του βίντεο σε ένα χώρο χαρακτηριστικών, που δημιουργείται από συνδυασμό ιστογραμμάτων χρώματος, υφής και ακμών.

Πρόσφατα, η ανάκτηση βίντεο σε χειρουργική αμφιβληστροειδούς έχει προσελκύσει το ενδιαφέρον ερευνητών [26, 92]. Σε αυτόν τον τομέα, στόχο αποτελεί η ανάκτηση ολοκληρωτών ροών βίντεο από συλλογές ψηφιακών αρχείων, παρά η ανάκτηση συγκεκριμένων χειρουργικών γεγονότων. Ωστόσο, αυτές οι μέθοδοι είναι προσαρμοσμένες στο συγκεκριμένο οπτικό περιεχόμενο, όπου δεν υπάρχουν διακυμάνσεις στη φωτεινότητα, έντονες κινήσεις της κάμερας, και παραμόρφωση των προβαλλόμενων αντικειμένων (ιστών).

Τα βίντεο EEX επεμβάσεων δεν είναι δυνατόν να κατατμηθούν σε πλάνα αναζητώντας για κοψίματα ή ομαλές μεταβάσεις με αλλαγή φωτεινότητας, καθώς η κάμερα επικεντρώνεται συνεχώς σε μια περιοχή ενδιαφέροντος και κινείται και περιστρέφεται ελεύθερα στο χώρο. Αναπόφευκτα, συνήθεις πρακτικές για ανίχνευση πλάνων οδηγούν σε υπερ- ή υποκατάτμηση. Για να αντιμετωπίσουμε αυτό το πρόβλημα, προτείνουμε μια καινοτόμο προσέγγιση για ανίχνευση πλάνων, που εκμεταλλεύεται τις χωρο-χρονικές αλλαγές στα καρέ του βίντεο, υπολογίζοντας ένα καθολικό και εύρωστο μέτρο μεταβολών που οδηγεί σε μια ουσιαστική χρονική κατάτμηση του βίντεο.

5.2 Σύνολο δεδομένων

Η συλλογή και αποθήκευση βίντεο από επεμβάσεις EEX για τη δημιουργία συνόλων δεδομένων δεν απαιτεί καμία αλλαγή στη χειρουργική διαδικασία, καθώς κατά τη διάρκεια της επέμβασης απαιτείται η ύπαρξη ενδοσκοπικής κάμερας. Ωστόσο, πολυπλοκότητα της διαδικασίας έχει άμεση επίπτωση στο οπτικό περιεχόμενο που συλλαμβάνει η κάμερα κατά



Εικόνα 5.1: Καρέ από βίντεο λαπαροσκοπικής χολοκυστεκτομής, όπου φαίνονται τα διαφορετικά λαπαροσκοπικά εργαλεία να έρχονται σε επαφή με τη χοληδόχο κύστη.

τη διάρκεια της επέμβασης, και οφείλεται: α) στο γεγονός ότι ο χειριστής πρέπει να παρακολουθεί συνεχώς το αντικείμενο ενδιαφέροντος, προκαλώντας απότομες κινήσεις της κάμερας και μη ομαλές αλλαγές της γωνίας λήψης, και β) στο γεγονός ότι η περιοχή ενδιαφέροντος αλλάζει δυναμικά λόγω εσωτερικών κινήσεων (π.χ. σε κάθε καρδιακό παλμό), διακυμάνσεων της φωτεινότητας που παρέχει η σημειακή πηγή φωτός και επικαλύψεων μεταξύ οργάνων. Το πλήθος των διαθέσιμων βίντεο από επεμβάσεις αυξάνεται διαρκώς, χάρη στην ραγδαία αύξηση της χωρητικότητας των μέσων αποθήκευσης καθώς και των τεχνολογιών διαμοιρασμού αρχείων, οι οποίες χρησιμοποιούνται και σε αυτό το πεδίο χρηστών. Βίντεο επεμβάσεων EEX είναι διαθέσιμα σε ιστότοπους διαμοιρασμού βίντεο γενικού περιεχομένου, όπως το youtube, καθώς και σε εξειδικευμένους ιστότοπους όπως το websurg¹.

Για να εξετάσουμε το πρόβλημα της ανάκτησης σε βίντεο ιατρικού περιεχομένου, δημιουργήσαμε μια βάση από βίντεο EEX, συγκεκριμένα λαπαροσκοπικής χολοκυστεκτομής (ΛΧ). Η ΛΧ είναι μια τυπική χειρουργική επέμβαση EEX με μεγάλη εκπαιδευτική αξία· συνήθως είναι η πρώτη λαπαροσκοπική επέμβαση που πραγματοποιούν οι ειδικευόμενοι χειρουργοί. Η επέμβαση αποτελείται από τρεις κύριες φάσεις που εκτελούνται ακολουθιακά: α) αφαίρεση προσκολλήσεων μεταξύ της χοληδόχου κύστης και γειτονικών οργάνων (τομή της χοληδόχου κύστης), β) τομή της κυστικής αρτηρίας, και γ) διαχωρισμός

¹<http://www.websurg.com/>

της χοληδόχου κύστης από το ήπαρ και εξαγωγή.

Κατά τη διάρκεια της επέμβασης χρησιμοποιούνται 6 διαφορετικά χειρουργικά εργαλεία (π.χ. ψαλίδια, διαχωριστές ιστών, λαβίδες). Για τη χρησιμοποίηση των βίντεο στην πειραματική αξιολόγηση του αλγορίθμου ανάκτησης, πραγματοποιήθηκε επισημείωση των πλάνων των βίντεο με κριτήριο τον τύπο των εργαλείων που εμφανίζονται σε κάθε πλάνο. (βλ. Εικόνα 5.1). Κατά την επισημείωση λάβαμε υπ' όψιν και δύο επιπλέον κατηγορίες αντικειμένων: α) τα *trocar* (επιμήκης σωλήνας που χρησιμοποιείται για την εισαγωγή των εργαλείων), και β) τα *clips* (μεταλλικά εξαρτήματα που τοποθετούνται από τον χειρουργό). Κάθε πλάνο επισημαίνεται με όλα τα εργαλεία που εμφανίζονται σε αυτό. Στο βίντεο που επισημειώσαμε, εμφανίζονται από 1 μέχρι και 5 εργαλεία σε ένα πλάνο (με διάμεσο τα 2 εργαλεία).

5.3 Ανίχνευση πλάνων σε βίντεο

Διαισθητικά, ένα βίντεο EEX αποτελείται από ένα μόνο πλάνο, καθώς η περιοχή που συλλαμβάνεται από την κάμερα δεν αλλάζει (η κάμερα εστιάζει συνεχώς στην περιοχή κάτω από το ήπαρ). Ωστόσο, στο βίντεο εντοπίζονται αλλαγές που αντιστοιχούν σε σημαντικά για το πεδίο γεγονότα, όπως η εισαγωγή ή η εξαγωγή ενός εργαλείου, ο χειρισμός της χοληδόχου κύστης, ή η αλλαγή της γωνίας λήψης (π.χ. από το περιβάλλοντα ιστό στη χοληδόχο κύστη).

Για την ανίχνευση πλάνων επικεντρωνόμαστε στα γεγονότα της εισαγωγής και εξαγωγής εργαλείων. Για τον εντοπισμό εργαλείων και την σήμανση αλλαγών στα πλάνα, εκπαιδεύουμε ένα μοντέλο της μεθόδου *Objectness*. Χρησιμοποιούμε την προσέγγιση των Alexe et al. [3], λόγω των καλών αποτελεσμάτων στον εντοπισμό αντικειμένων και της χρονικής απόδοσης του αλγορίθμου. Αρχικά, δημιουργήσαμε ένα σύνολο εικόνων των διαφορετικών χειρουργικών εργαλείων που εμφανίζονται στα βίντεο EEX, επισημειώνοντας τις περιοχές των καρέ στις οποίες εμφανίζεται κάθε εργαλείο. Το σύνολο εικόνων το χρησιμοποιούμε για την εκμάθηση των παραμέτρων της μεθόδου *Objectness*. Στη συνέχεια προσαρμόζουμε τη μέθοδο, συγχωνεύοντας τα αποτελέσματα σε κάθε καρέ, ώστε να υπολογίσουμε μια αριθμητική τιμή που χαρακτηρίζει το συνολικό μέτρο “*objectness*” του καρέ, αθροίζοντας τις τιμές κάθε περιοχής του.

Προτείνουμε πως οι αλλαγές στο καθολικό μέτρο “*objectness*” των καρέ των βίντεο σχετίζονται με σημαντικές αλλαγές πλάνων, σχετικές με την εμφάνιση και την απόκρυψη εργαλείων. Παρακολουθούμε τη διακύμανση του μέτρου σε όλη τη διάρκεια του βίντεο, και οριοθετούμε ως αλλαγές πλάνων τα καρέ στα οποία εμφανίζεται έντονη πτώση ή αύξηση του μέτρου.

5.4 Περιγραφή βίντεο

Για την περιγραφή των πλάνων, εξάγουμε έναν περιγραφέα υψηλής διάστασης, ο οποίος θα χρησιμοποιηθεί κατά την ανάκτηση με αναζήτηση πλησιέστερου γείτονα. Η αναπαράσταση που προτείνουμε είναι συμπαγής, και μπορεί να χρησιμοποιηθεί για γρήγορη ανάκτηση βίντεο σε σύνολα δεδομένων μεγάλης κλίμακας. Για κάθε πλάνο αρχικά εξάγουμε όλα τα καρέ με σταθερό χρονικό βήμα μεταξύ καρέ, ίσο με $200ms$ για όλα τα πειράματα που πραγματοποιούμε.

5.4.1 Περιγραφή βίντεο με τοπικά χαρακτηριστικά

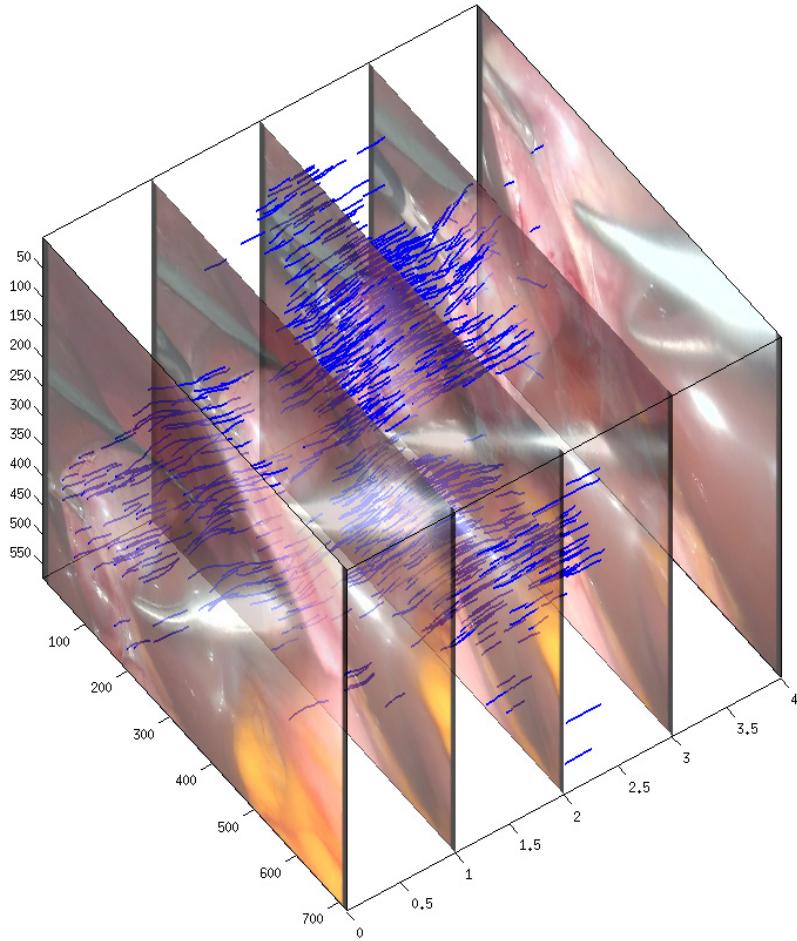
Αρχικά εξετάζουμε την εξαγωγή περιγραφών για πλάνα βίντεο χρησιμοποιώντας τοπικά χαρακτηριστικά από εικόνες. Για κάθε καρέ εντοπίζουμε χαρακτηριστικά SIFT [70], μιαζί με τους αντίστοιχους περιγραφείς. Κατασκευάζουμε ένα οπτικό λεξικό $100K$ λέξεων, από τη συσταδοποίηση ενός τυχαίου υποσυνόλου των περιγραφών που εξάγαμε, με χρήση του αλγόριθμου προσεγγιστικού k-means [89], ώστε να κβαντίσουμε τους περιγραφείς σε οπτικές λέξεις.

Επιλέγουμε μόνο τα χαρακτηριστικά τα οποία έχουν ταιριάζει μεταξύ διαδοχικών καρέ σύμφωνα με κάποιο γεωμετρικό μοντέλο, για να απορρίψουμε τα θορυβώδη χαρακτηριστικά στην κατασκευή της αναπαράστασης. Τα χαρακτηριστικά που παραμένουν, αντιστοιχούν σε κοινά τμήματα των αντικειμένων στα διαφορετικά καρέ. Οι οπτικές λέξεις που αντιστοιχούν στα χαρακτηριστικά που επιλέχθηκαν σε κάθε καρέ συναθροίζονται, και κατασκευάζουμε ένα ιστόγραμμα εμφάνισης των οπτικών λέξεων για κάθε πλάνο. Για την αναζήτηση χρησιμοποιούμε το μοντέλο bag-of-words (BoW), δηλαδή το κανονικοποιημένο με την απόσταση L^2 ιστόγραμμα των λέξεων. Το μέγεθος του τελικού περιγραφέα ενός πλάνου, είναι ίσο με το μέγεθος του οπτικού λεξικού, δηλαδή $100K$.

Οι διαφορετικές παραλλαγές που εξετάζουμε με χρήση τοπικών χαρακτηριστικών και ταιριασμα μεταξύ καρέ είναι οι ακόλουθες:

Βασικό (R): Το βασικό σύστημα συναθροίζει τις οπτικές λέξεις των χαρακτηριστικών που ταιριάζαν μεταξύ διαδοχικών καρέ για την κατασκευή του ιστογράμματος. Πραγματοποιούμε το χωρικό ταίριασμα των διαδοχικών καρέ με χρήση του αλγορίθμου RANSAC [35], αναζητώντας τον βέλτιστο μετασχηματισμό ομογραφίας μεταξύ των καρέ, ώστε να επιλέξουμε επαναλαμβανόμενα χαρακτηριστικά και να απορρίψουμε τα υπόλοιπα.

Επαναληπτικός RANSAC (R_{iter}): Για να ξεπεράσουμε τον περιορισμό που θέτει ο αλγόριθμος RANSAC, ότι δηλαδή το ταίριασμα περιλαμβάνει τα σημεία που βρίσκονται σε ένα μοναδικό επίπεδο (ένας μετασχηματισμός ομογραφίας μεταξύ των καρέ), εφαρμόζουμε μια απλή, αλλά αποτελεσματική προσέγγιση. Όπως προτείνεται και στα [117, 50], ταιριάζουμε τα διαδοχικά καρέ εκτελώντας τον αλγόριθμο RANSAC επαναληπτικά, αναζητώ-



Εικόνα 5.2: Τροχιές τοπικών χαρακτηριστικών: επιλογή των τοπικών χαρακτηριστικών που ταιριάζουν μεταξύ τουλάχιστον 5 διαδοχικών καρέ του βίντεο.

ντας πολλαπλούς μετασχηματισμούς ομογραφίας μεταξύ των καρέ. Σε κάθε επανάληψη απορρίπτουμε τα σημεία που έχουν ήδη ταιριάζει, έως ότου δε μπορεί να βρεθεί ικανοποιητικός αριθμός σημείων για ταιριασμα μεταξύ των καρέ. Με αυτή τη μέθοδο μπορούμε να μειώσουμε την ανοχή σε χωρικό σφάλμα σε κάθε επανάληψη, ενώ ταυτόχρονα, λόγω των επαναλήψεων, επιλέγονται διαφορετικοί μετασχηματισμοί και χαρακτηριστικά που δε θα είχαν επιλεχθεί από το βασικό σύστημα.

Τροχιές τοπικών χαρακτηριστικών (TTX): Κατασκευάζουμε τροχιές τοπικών χαρακτηριστικών (local feature trajectories – *LFT*) ταιριάζοντας χαρακτηριστικά με χρήση είτε του απλού αλγορίθμου RANSAC, είτε του επαναληπτικού. Επιλέγουμε μόνο τις τροχιές που εκτείνονται σε τουλάχιστον 5 διαδοχικά καρέ του βίντεο. Με αυτή τη μέθοδο απορρίπτουμε περισσότερα θορυβώδη χαρακτηριστικά που είτε εμφανίζονται για πολύ μικρό χρονικό διάστημα (έντονες επικαλύψεις), είτε εσφαλμένα ταιριάζονται από τον αλγόριθμο χωρικού ταιριασμάτος (απλού ή επαναληπτικού RANSAC) (βλ. Εικόνα 5.2).

5.4.2 Περιγραφή βίντεο με νευρωνικά δίκτυα

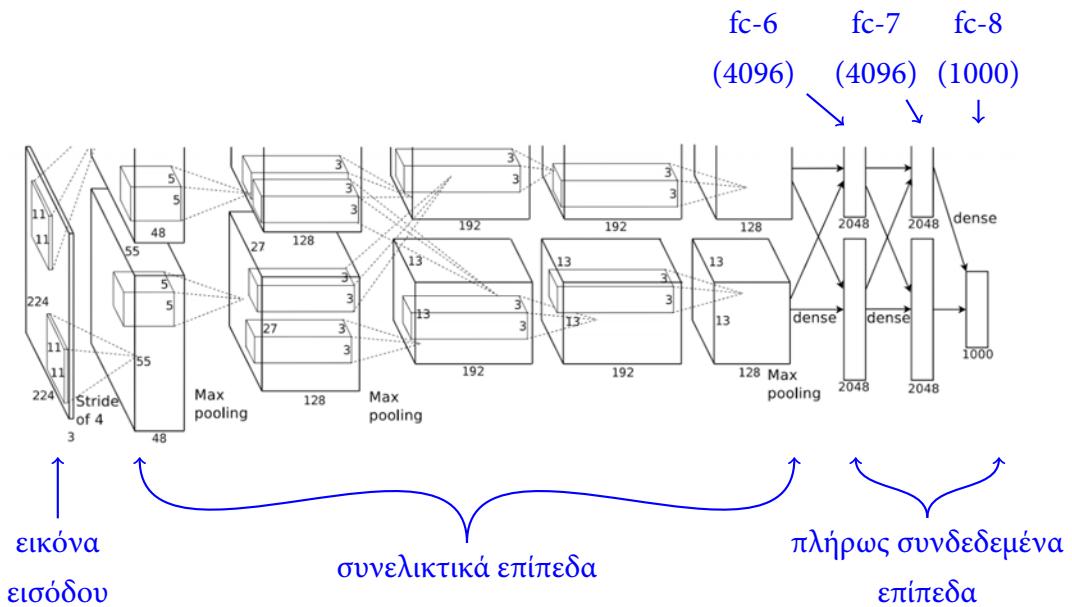
Οι τοπολογίες βαθιών συνελικτικών νευρωνικών δικτύων που έχουν προταθεί στη βιβλιογραφία (βλ. ενότητα 2.3) περιλαμβάνουν στα πρώτα επίπεδα συνελικτικά και συγκεντρωτικά επίπεδα. Τα συνελικτικά επίπεδα αντιστοιχούν σε οπτικά φίλτρα, που συλλαμβάνουν μοτίβα στις εικόνες, είτε χρωματικά, είτε υφής (βλ. Εικόνα 5.4α,β). Τα συγκεντρωτικά επίπεδα μεταφέρουν τις έντονες αποκρίσεις των φίλτρων προς επόμενα επίπεδα, εντοπίζοντας εξέχουσες δομές, ενώ ταυτόχρονα μειώνουν τις διαστάσεις του δικτύου και το υπολογιστικό κόστος των επιπέδων. Τα τελευταία επίπεδα των πιο επιτυχημένων τοπολογιών είναι πλήρως συνδεδεμένα, συγχωνεύοντας τις εξέχουσες αποκρίσεις στα οπτικά φίλτρα, δημιουργώντας αναπαραστάσεις εικόνων με αυξημένη περιγραφικότητα.

Τα καρέ που εξάγουμε από τα βίντεο αλλάζουν μέγεθος ώστε να ταιριάζουν με το μέγεθος εισόδου του νευρωνικού δικτύου, και δίνονται ως είσοδοι σε αυτό. Χρησιμοποιούμε τις αποκρίσεις των πλήρως συνδεδεμένων επιπέδων (χάρτες χαρακτηριστικών) ως περιγραφείς των καρέ. Οι περιγραφείς είναι διανύσματα διαστάσεων $d_l \times 1$, όπου l είναι το επίπεδο που επιλέγουμε. Συνενώνουμε τους περιγραφείς των καρέ στη διάσταση του χρόνου, οπότε για ένα πλάνο διάρκειας n καρέ, προκύπτει ένας περιγραφέας διαστάσεων $d_l \times n$, και εφαρμόζουμε ένα μονοδιάστατο φίλτρο διαμέσου στη χρονική διάσταση για να ομαλοποιήσουμε τις διακυμάνσεις από καρέ σε καρέ. Για την κατασκευή ενός περιγραφέα ανεξάρτητου της χρονικής διάρκειας του πλάνου, συναθροίζουμε τους περιγραφείς στη χρονική διάσταση με ένα επίπεδο συγκέντρωσης μεγίστου. Το τελικό μέγεθος του διανύσματος περιγραφέα του πλάνου θα είναι ίσο με το μέγεθος του πλήρως συνδεδεμένου δικτύου που επιλέγουμε για την εξαγωγή των χαρτών χαρακτηριστικών. Στις τοπολογίες που εξετάζουμε το μέγεθος αυτό είναι ίσο με 4096 για τα επίπεδα “fc-6” και “fc-7”, και 1000 για το επίπεδο “fc-8” (βλ. Εικόνα 5.3).

5.5 Πειραματική αξιολόγηση

Σε ένα βίντεο EEX διάρκειας δύο ωρών ανιχνεύσαμε 186 πλάνα, με την μέθοδο που περιγράψαμε στην ενότητα 5.3. Κάθε πλάνο επισημειώθηκε από έναν ειδικό του πεδίου, εστιάζοντας στα εργαλεία που εμφανίζονται στα πλάνα (όπως περιγράψαμε στην ενότητα 5.2).

Πρώτα αποτιμάμε την απόδοση των προσεγγίσεων με χρήση τοπικών χαρακτηριστικών SIFT. Αποτιμάμε την απόδοση του βασικού συστήματος με τοπικά χαρακτηριστικά, που χρησιμοποιεί τον αλγόριθμο RANSAC (R) για το τοπικό ταίριασμα. Ελέγχουμε τη βελτίωση στην απόδοση του συστήματος όταν εφαρμόζουμε τον επαναληπτικό RANSAC (R_{iter}) και εξάγοντας τροχιές χαρακτηριστικών (TTX – LFT). Οι τροχιές χαρακτηριστικών μπορούν να εξαχθούν είτε χρησιμοποιούμε τον απλό, είτε τον επαναληπτικό RANSAC.



Εικόνα 5.3: Τοπολογία δικτύου Alexnet [54].

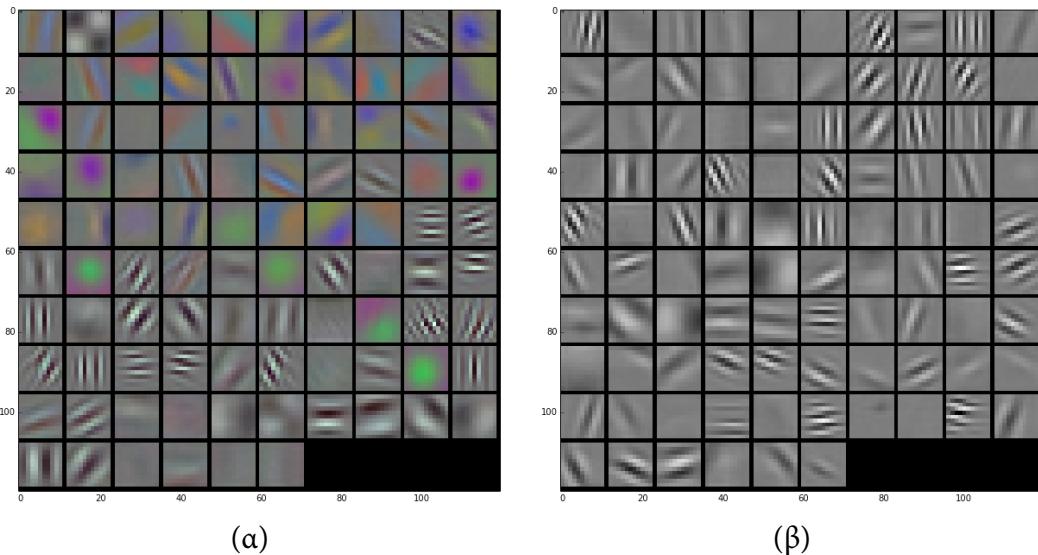
Τα αποτελέσματα της αποτίμησης φαίνονται στον Πίνακα 5.1. Η χρήση του επαναληπτικού RANSAC βελτιώνει τα αποτελέσματα κατά περίπου 2%, ενώ η χρήση τροχιών χαρακτηριστικών βελτιώνει τα αποτελέσματα κατά περίπου 19%.

Αποτιμάμε την προτεινόμενη αναπαράσταση με χρήση νευρωνικών δικτύων, χρησιμοποιώντας μοντέλα προ-εκπαιδευμένα στο σύνολο δεδομένων Imagenet [23], της βιβλιοθήκης Caffe [49]. Συγκεκριμένα, χρησιμοποιούμε τα μοντέλα δικτύων AlexNet [54], CaffeNet (το οποίο διαφέρει ελάχιστα από το AlexNet), και τα δίκτυα VGG [104].

Όλες οι τοπολογίες δικτύων που εξετάζουμε χρησιμοποιούν έγχρωμες εικόνες σαν είσοδο, σε αντίθεση με τις μεθόδους με βάση τα τοπικά χαρακτηριστικά, όπου ο ανιχνευτής και ο περιγραφέας SIFT χρησιμοποιούν μόνο τη φωτεινότητα των εικόνων. Για να εξετάσουμε την επίδραση του χρώματος στην περιγραφή των καρέ και την ανάκτηση, δημιουργούμε την τοπολογία συνελικτικού δικτύου *Alexnet-gray* που βασίζεται στην τοπολογία του *AlexNet* (βλ. Εικόνα 5.3), αλλά η είσοδός του είναι ασπρόμαυρες εικόνες. Η εκπαίδευση του δικτύου έγινε χρησιμοποιώντας τις εικόνες του συνόλου Imagenet, απορρίπτοντας τη χρωματική πληροφορία. Στην Εικόνα 5.4 βλέπουμε τις παραμέτρους του πρώτου

Τροχιές	Χωρικό Ταίριασμα	Ακρίβεια
—	R	64.6
	R_{iter}	65.9
<i>LFT</i>	R	76.8
	R_{iter}	78.3

Πίνακας 5.1: Πειραματικά αποτελέσματα ανάκτησης βίντεο με χρήση τοπικών χαρακτηριστικών για την αναπαράσταση πλάνων.



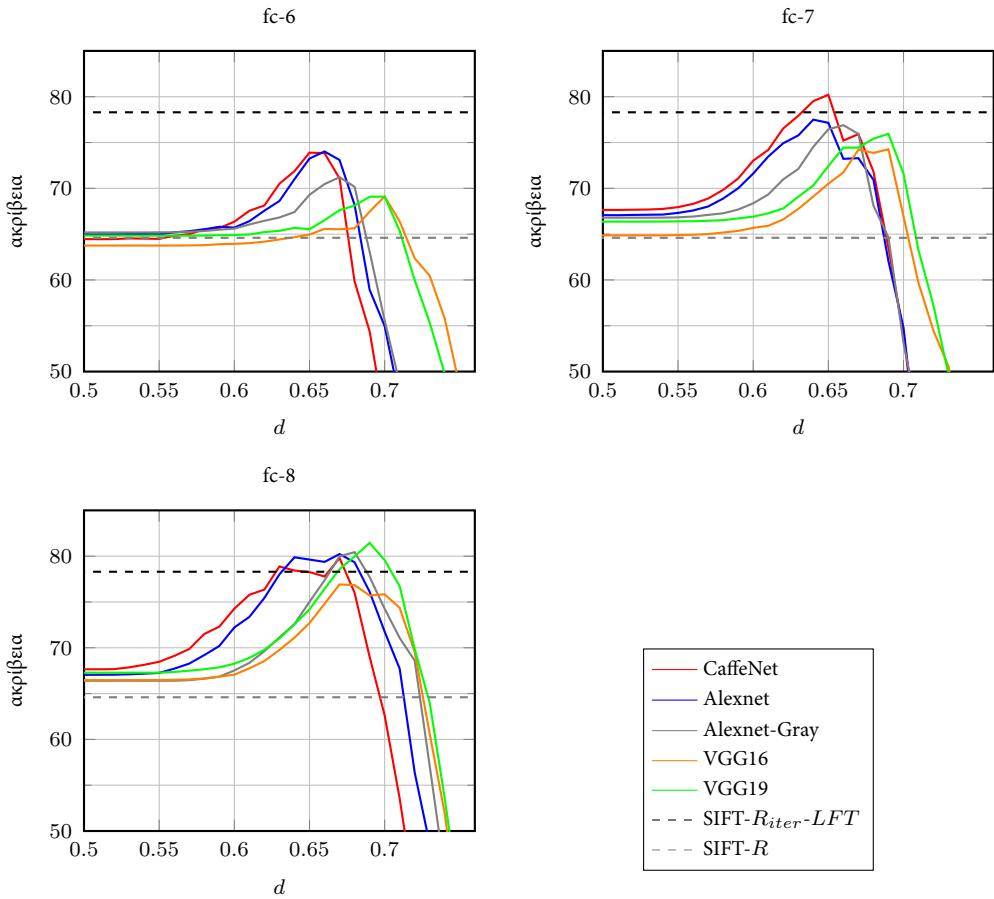
Εικόνα 5.4: Απεικόνιση των παραμέτρων του πρώτου συνελικτικού επιπέδου: (α) του νευρωνικού δικτύου *AlexNet*, και (β) του δικτύου *Alexnet-gray*.

συνελικτικού επιπέδου των δικτύων AlexNet και Alexnet-gray, που είναι 96 φίλτρα μεγέθους 11×11 , μετά τη διαδικασία εκμάθησης. Στο δίκτυο AlexNet (Εικόνα 5.4α), τα φίλτρα του πρώτου επιπέδου μοιράζονται μεταξύ αυτών που εντοπίζουν χρωματικά πρότυπα και αυτών που εντοπίζουν πρότυπα φωτεινότητας, όπως ακμές σε διάφορες κατευθύνσεις αλλά και πιο σύνθετα φίλτρα που προσεγγίζουν φίλτρα Gabor. Στο δίκτυο Alexnet-gray όλα τα φίλτρα εντοπίζουν πρότυπα φωτεινότητας, με αποτέλεσμα την εκμάθηση περισσότερων σύνθετων φίλτρων σε περισσότερες διαφορετικές γωνίες περιστροφής.

Τα αποτελέσματα της αποτίμησης φαίνονται στην Εικόνα 5.5. Ο άξονας x στα γραφήματα αντιστοιχεί στο κατώφλι που εφαρμόζουμε στην ομοιότητα μεταξύ των περιγραφέων για την επιλογή των σχετικών μεταξύ τους πλάνων, ενώ η μετρική της αποτίμησης είναι η μέση ακρίβεια. Σε όλα τα δίκτυα που εξετάζουμε, το επίπεδο “fc-6” είναι το πρώτο πλήρως συνδεδεμένο επίπεδο του δικτύου, που συγχωνεύει τις αποκρίσεις των συνελικτικών επιπέδων. Το επίπεδο “fc-7” είναι το δεύτερο πλήρως συνδεδεμένο επίπεδο, ενώ το “fc-8” είναι το τελευταίο επίπεδο του δικτύου.

Τα αποτελέσματα που παίρνουμε χρησιμοποιώντας το επίπεδο “fc-6” για την εξαγωγή περιγραφέων καρέ είναι καλύτερα από αυτά της βασικής μεθόδου, αλλά χειρότερα από το καλύτερο αποτέλεσμα που πήραμε χρησιμοποιώντας τοπικά χαρακτηριστικά (βλ. Πίνακα 5.1). Χρησιμοποιώντας τα επίπεδα “fc-7” ή “fc-8” ξεπερνάμε την απόδοση των μεθόδων με τοπικά χαρακτηριστικά. Η χρήση πολλαπλών πλήρως συνδεδεμένων επιπέδων οδηγεί στην εξαγωγή καλύτερων περιγραφέων, καθώς περισσότεροι νευρώνες χρησιμοποιούνται για την συγχώνευση των αποκρίσεων των φίλτρων των συνελικτικών επιπέδων. Η βέλτιστη απόδοση επιτυγχάνεται χρησιμοποιώντας το επίπεδο “fc-8” του δικτύου VGG19, που αποτελείται από συνολικά 19 επίπεδα και είναι το πιο βαθύ δίκτυο της πειρα-

Πειραματική αξιολόγηση



Εικόνα 5.5: Πειραματικά αποτελέσματα ανάκτησης βίντεο χρησιμοποιώντας περιγραφείς νευρωνικών δικτύων.

ματικής αποτίμησης. Το επίπεδο “fc-8” είναι το τελευταίο επίπεδο του δικτύου, το οποίο για τα δίκτυα που χρησιμοποιήσαμε έχει εκπαιδευτεί με στόχο την ταξινόμηση εικόνων στις 1000 κατηγορίες του συνόλου εικόνων Imagenet. Το γεγονός ότι χρησιμοποιώντας αυτό το επίπεδο πετυχαίνουμε τα καλύτερα αποτελέσματα οφείλεται στο μέγεθος του συνόλου Imagenet και στη διαφορετικότητα των εικόνων που περιλαμβάνονται σε αυτό.

Συγκρίνοντας την απόδοση του δικτύου AlexNet και του Alexnet-gray βλέπουμε ότι η χρωματική πληροφορία βοηθάει στην αύξηση της απόδοσης όταν χρησιμοποιούμε το πρώτο πλήρως συνδεδεμένο δίκτυο (“fc-6”). Ωστόσο, η διαφορά στην απόδοση εξαλείφεται χρησιμοποιώντας περισσότερα πλήρως συνδεδεμένα επίπεδα, όπως φαίνεται στις περιπτώσεις “fc-7” και “fc-8”. Σε περιπτώσεις που οι εικόνες εισόδου είναι ασπρόμαυρες, το δίκτυο Alexnet-gray αναμένεται να έχει καλύτερη απόδοση, καθώς χρησιμοποιεί πιο σύνθετα φίλτρα στα συνελικτικά επίπεδα, και δε στηρίζεται στη χρωματική πληροφορία για την αποδοτική περιγραφή των καρέ.

Κεφάλαιο 6

Συμπεράσματα

Στην παρούσα διατριβή ασχοληθήκαμε με τη δημιουργία συμπαγών και αποδοτικών αναπαραστάσεων οπτικής πληροφορίας από εικόνες και βίντεο. Οι αναπαραστάσεις που προτείναμε μπορούν να χρησιμοποιηθούν σε πληθώρα προβλημάτων της ανάλυσης εικόνων και βίντεο και όρασης υπολογιστών, όπως την κατηγοριοποίηση, την ανίχνευση αντικειμένων και την ανάκτηση με βάση το περιεχόμενο. Ως πεδίο εφαρμογής επιλέξαμε το πρόβλημα ανάκτησης εικόνων και βίντεο, στο οποίο εξετάσαμε και την απόδοση των αναπαραστάσεων που προτείναμε.

Η διατριβή συνδυάζει διαφορετικές ερευνητικές περιοχές της ανάλυσης εικόνων και όρασης υπολογιστών με την υπολογιστική γεωμετρία και τη μηχανική μάθηση. Στα πλαίσια της διατριβής μελετήσαμε εκτενώς τη βιβλιογραφία των ερευνητικών περιοχών που εξετάσαμε, και προτείναμε νέους αλγορίθμους για την ανίχνευση τοπικών χαρακτηριστικών, τη δημιουργία α -σχημάτων, τη δειγματοληψία εικόνων, και την περιγραφή βίντεο.

Η εκπόνηση της διατριβής ξεκίνησε εξετάζοντας τη δημιουργία ενός υπολογιστικά αποδοτικού ανιχνευτή τοπικών χαρακτηριστικών, με έμφαση στον εντοπισμό σημαντικών δομών των αντικειμένων. Η εφαρμογή των α -σχημάτων σε εικόνες έδειξε ότι προσφέρει μια πολύ περιγραφική αναπαράσταση, που θα μπορούσε να οδηγήσει στην εξαγωγή χαρακτηριστικών. Συνδυάζοντας τα α -σχήματα με την φωλιασμένη αναπαράσταση των δέντρων συνιστωσών, κατασκευάσαμε μια ιεραρχία σχημάτων που συλλαμβάνει εξελισσόμενες δομές της εικόνας, τις οποίες θεωρήσαμε υποψήφιες περιοχές ενδιαφέροντος. Το αποτέλεσμα αυτής της διαδικασίας ήταν η πρόταση του νέου αλγορίθμου ανίχνευσης χαρακτηριστικών W_αSH. Εξετάσαμε αναλυτικά κάθε στάδιο του αλγορίθμου W_αSH, και προτείναμε νέες μεθόδους για κάθε βήμα του, με σημαντικότερη την πρόταση νέων μεθόδων δειγματοληψίας εικόνων και την κατασκευή α -σχημάτων με προσαρμογή στο τοπικό αφινικό σχήμα.

Ο εντοπισμός τοπικών χαρακτηριστικών σε εικόνες με 2Δ α -σχήματα που προτείναμε στην παρούσα διατριβή, μπορεί να μεταφερθεί στις 3 διαστάσεις για τον εντοπισμό χωρο-

Συνεισφορά

χρονικών περιοχών ενδιαφέροντος. Οι περιοχές αυτές αντιστοιχούν σε χωρο-χρονικές κοιλότητες των 3Δ α-σχημάτων, που διαισθητικά περιγράφονται σαν περιοχές των καρέ του βίντεο ομοιόμορφης φωτεινότητας (blobs) τα οποία δεν παραμορφώνονται ιδιαίτερα στο χρόνο. Στην πορεία της διατριβής εξετάσαμε την εξαγωγή τέτοιων χωρο-χρονικών χαρακτηριστικών, με δειγματοληψία σε καρέ βίντεο και κατασκευή 3Δ α-σχημάτων. Ωστόσο, οι διαφορετικές κινήσεις των αντικειμένων στα βίντεο οδήγησαν σε θορυβώδεις αναπαραστάσεις, παρά τις διαφορετικές δειγματοληψίες που εξετάσαμε, και κατ' επέκταση σε πολύπλοκες 3Δ δομές (τριγωνοποιήσεις και α-σχήματα). Ως αποτέλεσμα, οι χωρο-χρονικές περιοχές που εντοπίσαμε σε ιεραρχικές δομές των 3Δ α-σχημάτων (δέντρα συνιστωσών), με κριτήρια αντίστοιχα αυτών που χρησιμοποιήσαμε στις 2 διαστάσεις δεν οδήγησε στον εντοπισμό εύρωστων περιοχών ενδιαφέροντος.

Η ραγδαία ανάπτυξη των Συνελικτικών Νευρωνικών Δικτύων τα τελευταία χρόνια μας οδήγησε στο τελευταίο μέρος της διατριβής να μελετήσουμε το συγκεκριμένο ερευνητικό πεδίο. Εξετάσαμε τη χρήση βαθιών ΣΝΔ για την κατηγοριοποίηση εικόνων και τον εντοπισμό αντικειμένων, και οδηγηθήκαμε στην εξαγωγή περιγραφέων από εικόνες με χρήση ΣΝΔ. Επεκτείνοντας υπάρχουσες μεθόδους, προτείναμε ένα νέο αλγόριθμο για την κατάτμηση βίντεο σε πλάνα και την αποδοτική αναπαράστασή τους με χρήση περιγραφέων ΣΝΔ, και με σκοπό την ανάκτηση βίντεο μεγάλης κλίμακας.

6.1 Συνεισφορά

Στα προηγούμενα κεφάλαια περιγράψαμε με λεπτομέρεια τη συνεισφορά της διατριβής στα διαφορετικά πεδία της ανάλυσης εικόνων και βίντεο με τα οποία ασχοληθήκαμε. Συνοπτικά, η συνεισφορά της διατριβής έγκειται στα εξής:

- Για την περιγραφή εικόνων, προτείναμε τον νέο αλγόριθμο ανίχνευσης τοπικών χαρακτηριστικών *WaSH* (*Weighted α-SHapes detector*), που συνδυάζει μεθόδους από το επιστημονικό πεδίο της επεξεργασίας εικόνων και της υπολογιστικής γεωμετρίας. Ξεκινώντας από δυαδικές ακμές της εικόνας, εξάγαμε ένα χαρακτηριστικό σύνολο δειγμάτων της εικόνας το οποίο χρησιμοποιήσαμε για την δημιουργία γεωμετρικών κατασκευών. Συγκεκριμένα, χρησιμοποιήσαμε διαφορετικές τριγωνοποιήσεις και τα α-σχήματα για να περιγράψουμε τις δομές που εμφανίζονται από τα αρχικά σημεία της δειγματοληψίας.

Ο ανιχνευτής που περιγράψαμε είναι αποδοτικός ως προς το χρόνο εξαγωγής των χαρακτηριστικών και το πλήθος τους, που καθορίζει το μέγεθος της τελικής αναπαράστασης της εικόνας. Τα χαρακτηριστικά που εξάγονται είναι αναλλοίωτα ως προς πολλούς μετασχηματισμούς (αλλαγή κλίμακας, περιστροφή και αφινικούς μετασχηματισμούς), αλλαγές της φωτεινότητας και θόλωμα της εικόνας. Το μικρό,

συγκριτικά, πλήθος εξαγόμενων χαρακτηριστικών, καθώς και το αναλλοίωτο σε μετασχηματισμούς καθιστά τον προτεινόμενο ανιχνευτή ιδιαίτερα κατάλληλο για εφαρμογές μεγάλης κλίμακας (ανάκτηση εικόνων, ανίχνευση με πολλές κατηγορίες κ.α.). Η πειραματική αποτίμηση της απόδοσης του αλγορίθμου επαληθεύει τους ισχυρισμούς μας, καθώς ο ανιχνευτής συγκρίθηκε με τους καλύτερους ανιχνευτές της βιβλιογραφίας σε δύο διαφορετικά πειράματα.

- Προτείναμε δύο διαφορετικές οικογένειες μεθόδων δειγματοληψίας εικόνων: δειγματοληψία επί δυαδικών ακμών και δειγματοληψία με διάχυση σφάλματος σε συνεχείς συναρτήσεις της φωτεινότητας. Οι δειγματοληψίες εμφανίζουν διαφορετικές μεταξύ τους ιδιότητες: η δειγματοληψία επί ακμών με σταθερό βήμα και η δειγματοληψία επί της κλίσης έχουν χαμηλό υπολογιστικό κόστος, η δειγματοληψία επί ακμών με ανίχνευση τοπικού σχήματος δημιουργεί σύνολα σημείων μεταβλητής πυκνότητας, ενώ η δειγματοληψία επί της Hessian μπορεί να οδηγήσει σε καλύτερη ανακατασκευή της αρχικής εικόνας.
- Προτείναμε τα ανισοτροπικά σταθμισμένα α-σχήματα, τα οποία συνδυάζουν την απόδοση και σταθερότητα της κανονικής τριγωνοποίησης, με μια πλούσια αναπαράσταση για κάθε άπλοκο (ακμή ή τρίγωνο), που εκμεταλλευόμενη το τοπικό σχήμα ξεπερνά τα μειονεκτήματα των α-σχημάτων. Η κατασκευή είναι ντετερμινιστική και δεν εξαρτάται από τη σειρά επίσκεψης στα άπλοκα, αλλά μπορεί να οδηγήσει σε μη συνεπείς καθολικά κατασκευές, λόγω της τοπικής ανίχνευσης του σχήματος.
- Για την περιγραφή βίντεο χρησιμοποιήσαμε βαθιά νευρωνικά δίκτυα, τα οποία χρησιμοποιούνται ευρέως τα τελευταία χρόνια για την κατηγοριοποίηση εικόνων. Η μέθοδος που προτείναμε για ανάκτηση βίντεο χωρίζει τα βίντεο σε πλάνα εξετάζοντας τη μεταβολή της καθολικής πιθανότητας εμφάνισης αντικειμένων στα καρέ. Η εξαγωγή περιγραφέων από τα πλάνα γίνεται με χρήση βαθιών συνελικτικών νευρωνικών δικτύων. Οι έξοδοι των τελευταίων σταδίων των νευρωνικών δικτύων χρησιμοποιούνται ως καθολικοί περιγραφείς των καρέ των βίντεο, και μετά από μια διαδικασία συνάθροισης προκύπτει ο περιγραφέας του πλάνου. Η αναπαράσταση είναι πολύ συμπαγής, καθώς δημιουργείται ένα διάνυσμα σταθερού και σχετικά μικρού μήκους ανεξάρτητα από τη χρονική διάρκεια του πλάνου, με αποτέλεσμα να είναι ιδανική για προβλήματα ανάκτησης σε μεγάλες συλλογές βίντεο. Η πειραματική αποτίμηση σε ένα πείραμα ανάκτησης πλάνων σε χειρουργικά βίντεο επιβεβαίωσε την καταλληλότητα της προτεινόμενης μεθόδου στην περιγραφή βίντεο.

6.2 Μελλοντικές επεκτάσεις

Στη διατριβή εξετάσαμε θέματα περιγραφής εικόνων και βίντεο με διαφορετικές μεθόδους και εργαλεία, προτείνοντας νέους αλγορίθμους. Από τη διατριβή προκύπτουν μελλοντικές κατευθύνσεις έρευνας από τα θέματα με τα οποία ασχοληθήκαμε, όπως:

- **Συνδυασμός τοπικών χαρακτηριστικών με περιγραφείς ΣΝΔ.** Πολύ πρόσφατα πρόταθηκε η εξαγωγή περιγραφέων με χρήση ΣΝΔ από υποπαράθυρα της εικόνας που προκύπτουν από ανιχνευτές τοπικών χαρακτηριστικών [88]. Η εξαγωγή τοπικών περιγραφέων με ΣΝΔ φαίνεται πολύ ενδιαφέρουσα προσέγγιση, όπου θα μπορούσαν να χρησιμοποιηθούν τα τοπικά χαρακτηριστικά που εξάγει ο αλγόριθμος WαSH. Επίσης, θα μπορούσαμε να εξετάσουμε την χρησιμοποίηση συναρτήσεων της φωτεινότητας ως είσοδο στα ΣΝΔ, αντί της χρωματικής πληροφορίας, όπως προτείναμε για τη δειγματοληψία εικόνων στο κεφάλαιο 4.
- **Κατάτμηση βίντεο με ΣΝΔ.** Στο κεφάλαιο 5 προτείναμε τη χρήση ΣΝΔ για την εξαγωγή περιγραφέων από βίντεο, και εξετάσαμε την εφαρμογή της μεθόδου στο πρόβλημα της ανάκτησης βίντεο. Η κατάτμηση των βίντεο σε πλάνα έγινε με κριτήριο την πιθανότητα εμφάνισης αντικειμένων στα καρέ. Ενδιαφέρον θα είχε η διερεύνηση της κατάτμησης παρακολουθώντας τις διακυμάνσεις στους καθολικούς περιγραφείς που εξάγουμε με ΣΝΔ. Το αποτέλεσμα αναμένεται να οδηγήσει σε πιο εύρωστους περιγραφείς, καθώς η διακύμανσή τους θα είναι μικρότερη εντός του πλάνου.
- **Περιγραφή βίντεο με δίκτυα με ανατροφοδότηση.** Για την περιγραφή βίντεο μπορεί να εξεταστεί και η χρήση δικτύων με ανατροφοδότηση και μνήμη, ώστε να γίνεται καλύτερη εκμετάλλευση της συνέχειας στο χρόνο. Ήδη έχουν εμφανιστεί εργασίες με χρήση βαθιών νευρωνικών δικτύων με ανατροφοδότηση για την εξαγωγή λεκτικών περιγραφών των δράσεων [116, 115]. Η εξαγωγή περιγραφέων σταθερού μήκους κατευθείαν από το νευρωνικό δίκτυο πιστεύουμε ότι θα μπορούσε να οδηγήσει σε καλύτερα αποτελέσματα από την συνάθροιση με σταθερή συνάρτηση (π.χ. μέσου όρου ή μεγίστου).

Αποδόσεις Ξένων Όρων

additively weighted height function	:	συνάρτηση ύψους με προστιθέμενα βάρη
anisotropically weighted alpha-shapes	:	ανισοτροπικά σταθμισμένα α -σχήματα
back-propagation	:	οπισθοδιάδοση σφάλματος
benchmark	:	πρότυπο συγκριτικής αξιολόγησης
cavity	:	εσοχή
circulant matrix	:	κυκλοειδής πίνακας
closure	:	κλειστότητα
complex	:	σύμπλοκο
component tree	:	δέντρο συνιστωσών
conformal	:	σύμμορφος
connected component	:	συνεκτική συνιστώσα
connectedness	:	συνδετικότητα
convex hull	:	κυρτό περίβλημα
edge	:	ακμή
edge foci	:	εστίες ακμών
face	:	όψη
facet	:	έδρα
filtration	:	διήθηση
flip	:	αναστροφή
gradient	:	κλίση
incident	:	προσπίπτων
incremental	:	αυξητικός
indexing	:	δεικτοδότηση
integral image	:	ολοκληρωτική εικόνα
level set	:	επιπεδοσύνολο
local features	:	τοπικά χαρακτηριστικά
lower hull	:	κάτω περίβλημα
matching score	:	απόδοση ταιριάσματος
max pooling	:	συγκέντρωση μεγίστου
medial axis transform	:	μετασχηματισμός διάμεσου άξονα
moment of inertia	:	ροπή αδράνειας
normal vector	:	κανονικοποιημένο κάθετο διάνυσμα
overfitting	:	υπερπροσαρμογή
pocket	:	κοιλότητα
polytope	:	πολύτοπο
power diagram	:	διάγραμμα δύναμης

Μελλοντικές επεκτάσεις

radical center	:	ριζικό κέντρο
recurrent network	:	δίκτυο με ανατροφοδότηση
regular triangulation	:	κανονική τριγωνοποίηση
regularised diffusion matrix	:	κανονικοποιημένος πίνακας διάχυσης
repeatability	:	επαναληψιμότητα
ridge	:	ράχη
salient	:	εξέχων
second-moment matrix	:	πίνακας ροπών 2 ^{ης} τάξης
simplex	:	άπλοκο
simplicial	:	απλοειδές
simplicial complex	:	απλοειδές σύμπλοκο
site	:	εστία
vertex	:	κορυφή
watershed	:	πλημμυρισμός
weighted α -shapes	:	σταθμισμένα α -σχήματα

Βιβλιογραφία

- [1] Pablo Fernández Alcantarilla, Adrien Bartoli, and Andrew J Davison. Kaze features. In *European Conference on Computer Vision*, pages 214–227. Springer, 2012.
- [2] Pablo Fernández Alcantarilla, Jesus Nuevo, and Adrien Bartoli. Fast explicit diffusion for accelerated features in nonlinear scale spaces. In *British Machine Vision Conference (BMVC)*, pages 13.1–13.11. BMVA Press, 2013.
- [3] Bogdan Alexe, Thomas Deselaers, and Vittorio Ferrari. Measuring the objectness of image windows. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 34(11):2189–2202, 2012.
- [4] A. Almansa and T. Lindeberg. Fingerprint enhancement by shape adaptation of scale-space operators with automatic scale selection. *Image Processing, IEEE Transactions on*, 9(12):2027–2042, 2000.
- [5] Barbara André, Tom Vercauteren, Anna M Buchner, Michael B Wallace, and Nicholas Ayache. A smart atlas for endomicroscopy using automated video retrieval. *Medical image analysis*, 15(4):460–476, 2011.
- [6] Y. Avrithis and K. Rapantzikos. The medial feature detector: Stable regions from image boundaries. In *International Conference on Computer Vision (ICCV)*, pages 1724–1731, nov. 2011.
- [7] Artem Babenko, Anton Slesarev, Alexandr Chigorin, and Victor Lempitsky. Neural codes for image retrieval. In *European Conference on Computer Vision (ECCV)*, pages 584–599. Springer, 2014.
- [8] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool. Speeded-up robust features (SURF). *Computer Vision and Image Understanding (CVIU)*, 110:346–359, June 2008.
- [9] P.R. Beaudet. Rotationally invariant image operators. In *International Joint Conference on Pattern Recognition*, pages 579–583, 1978.

- [10] Tobias Blum, Hubertus Feußner, and Nassir Navab. Modeling and segmentation of surgical workflow from laparoscopic video. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2010*, pages 400–407. Springer, 2010.
- [11] Michael Calonder, Vincent Lepetit, Christoph Strecha, and Pascal Fua. BRIEF: Binary robust independent elementary features. In *European Conference on Computer Vision (ECCV)*, pages 778–792. Springer, 2010.
- [12] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 8(6):679–698, 1986.
- [13] Yu Cao, Danyu Liu, Wallapak Tavanapong, Johnny Wong, JungHwan Oh, and Piet C De Groen. Computer-aided detection of diagnostic and therapeutic operations in colonoscopy videos. *Biomedical Engineering, IEEE Transactions on*, 54(7):1268–1279, 2007.
- [14] F. Cazals, J. Giesen, M. Pauly, and A. Zomorodian. Conformal alpha shapes. In *Point-Based Graphics. Eurographics/IEEE VGTC Symposium Proc.*, pages 55–61. IEEE, 2005.
- [15] Kai-Yueh Chang, Tyng-Luh Liu, Hwann-Tzong Chen, and Shang-Hong Lai. Fusing generic objectness and visual saliency for salient object detection. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 914–921. IEEE, 2011.
- [16] Ken Chatfield, Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. Return of the devil in the details: Delving deep into convolutional nets. In *Proceedings of the British Machine Vision Conference*. BMVA Press, 2014.
- [17] Ming-Ming Cheng, Ziming Zhang, Wen-Yan Lin, and Philip Torr. Bing: Binarized normed gradients for objectness estimation at 300fps. In *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pages 3286–3293. IEEE, 2014.
- [18] T. H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein. *Introduction to Algorithms*. MIT Press, 3rd edition, 2009.
- [19] H. S. M. Coxeter and S. L. Greitzer. *Geometry Revisited*. The Mathematical Association of America, 1995.
- [20] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 886–893. IEEE, 2005.

- [21] Navneet Dalal, Bill Triggs, and Cordelia Schmid. Human detection using oriented histograms of flow and appearance. In *European Conference on Computer Vision (ECCV)*, pages 428–441. Springer, 2006.
- [22] Jesús A. De Loera, Jörg Rambau, and Francisco oSantos. *Triangulations: Structures for algorithms and applications*, volume 25. Springer, 2010.
- [23] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 248–255. IEEE, 2009.
- [24] Jeff Donahue, Yangqing Jia, Oriol Vinyals, Judy Hoffman, Ning Zhang, Eric Tzeng, and Trevor Darrell. Decaf: A deep convolutional activation feature for generic visual recognition. In *Proceedings of The 31st International Conference on Machine Learning*, pages 647–655, 2014.
- [25] Matthijs Douze, Jérôme Revaud, Jakob Verbeek, Hervé Jégou, and Cordelia Schmid. Circulant temporal encoding for video retrieval and temporal alignment. *International Journal of Computer Vision*, pages 1–16, 2015.
- [26] Zakarya Droueche, Mathieu Lamard, Guy Cazuguel, Gwénolé Quellec, Christian Roux, and Béatrice Cochener. Motion-based video retrieval with application to computer-assisted retinal surgery. In *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, pages 4962–4965. IEEE, 2012.
- [27] H. Edelsbrunner. Alpha shapes — a survey. In *Tessellations in the Sciences: Virtues, Techniques and Applications of Geometric Tilings*. Springer Verlag, 2010.
- [28] H. Edelsbrunner, D. Kirkpatrick, and R. Seidel. On the shape of a set of points in the plane. *IEEE Transactions on Information Theory*, 29(4):551–559, 1983.
- [29] Herbert Edelsbrunner. *Weighted Alpha Shapes*. University of Illinois at Urbana-Champaign, Department of Computer Science, 1992.
- [30] Herbert Edelsbrunner and Nimish R Shah. Incremental topological flipping works for regular triangulations. *Algorithmica*, 15(3):223–241, 1996.
- [31] Ioannis Z Emiris, Elias P Tsigaridas, and George M Tzoumas. The predicates for the exact voronoi diagram of ellipses under the euclidean metric. *International Journal of Computational Geometry & Applications*, 18(06):567–597, 2008.

- [32] M. Everingham, S. M. A. Eslami, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The pascal visual object classes challenge: A retrospective. *International Journal of Computer Vision*, 111(1):98–136, January 2015.
- [33] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The PASCAL Visual Object Classes Challenge 2007 (VOC2007) Results. <http://www.pascal-network.org/challenges/VOC/voc2007/workshop/index.html>, 2003.
- [34] Robert Fergus, Pietro Perona, and Andrew Zisserman. Object class recognition by unsupervised scale-invariant learning. In *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, volume 2, pages II–264. IEEE, 2003.
- [35] Martin A Fischler and Robert C Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.
- [36] Robert W. Floyd and Louis Steinberg. An adaptive algorithm for spatial gray-scale. In *Proceedings of the Society of Information Display*, volume 17, pages 75–77, 1976.
- [37] Jan Flusser and Toms Suk. Pattern recognition by affine moment invariants. *Pattern Recognition*, 26(1):167 – 174, 1993.
- [38] Kunihiko Fukushima. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological cybernetics*, 36(4):193–202, 1980.
- [39] Anthony G Gallagher, E Matt Ritter, Andrew B Lederman, David A McClusky, and C Daniel Smith. Video-assisted surgery represents more than a loss of three-dimensional vision. *The American journal of surgery*, 189(1):76–80, 2005.
- [40] Stamatia Giannarou and Guang-Zhong Yang. Content-based surgical workflow representation using probabilistic motion modeling. In *Medical Imaging and Augmented Reality*, pages 314–323. Springer, 2010.
- [41] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Computer Vision and Pattern Recognition (CVPR)*, June 2014.
- [42] Georgia Gkioxari and Jitendra Malik. Finding action tubes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 759–768, 2015.

- [43] Steve Gu, Ying Zheng, and Carlo Tomasi. Critical nets and beta-stable features for image matching. In *European Conference on Computer Vision*, pages 663–676. Springer Berlin Heidelberg, 2010.
- [44] Chris Harris and Mike Stephens. A combined corner and edge detector. In *Alvey Vision Conference*, pages 147–151, 1988.
- [45] Ming-Kuei Hu. Visual pattern recognition by moment invariants. *Information Theory, IRE Transactions on*, 8(2):179–187, 1962.
- [46] David H Hubel and Torsten N Wiesel. Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex. *The Journal of physiology*, 160(1):106–154, 1962.
- [47] Manan Jain, Jan Van Gemert, Hervé Jégou, Patrick Bouthemy, and Cees GM Snoek. Action localization with tubelets from motion. In *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pages 740–747. IEEE, 2014.
- [48] Shuiwang Ji, Wei Xu, Ming Yang, and Kai Yu. 3d convolutional neural networks for human action recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(1):221–231, 2013.
- [49] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell. Caffe: Convolutional architecture for fast feature embedding. *arXiv preprint arXiv:1408.5093*, 2014.
- [50] Yasushi Kanazawa and Hiroshi Kawakami. Detection of planar regions with uncalibrated stereo using distributions of feature points. In *BMVC*, pages 1–10. Citeseer, 2004.
- [51] Andrej Karpathy, George Toderici, Sanketh Shetty, Thomas Leung, Rahul Sukthankar, and Li Fei-Fei. Large-scale video classification with convolutional neural networks. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 1725–1732, 2014.
- [52] Alexandre Karpenko and Parham Aarabi. Tiny videos: a large data set for nonparametric video retrieval and frame classification. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33(3):618–630, 2011.
- [53] Philipp Krähenbühl and Vladlen Koltun. Geodesic object proposals. In *European Conference on Computer Vision (ECCV)*, pages 725–739. Springer, 2014.

- [54] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C.J.C. Burges, L. Bottou, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc, 2012.
- [55] Francois Labelle and Jonathan Richard Shewchuk. Anisotropic voronoi diagrams and guaranteed-quality anisotropic mesh generation. In *Proceedings of the nineteenth annual symposium on Computational geometry*, pages 191–200. ACM, 2003.
- [56] Ivan Laptev, Marcin Marszałek, Cordelia Schmid, and Benjamin Rozenfeld. Learning realistic human actions from movies. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE, 2008.
- [57] Quoc Le, Marc'Aurelio Ranzato, Rajat Monga, Matthieu Devin, Kai Chen, Greg Corrado, Jeff Dean, and Andrew Ng. Building high-level features using large scale unsupervised learning. In John Langford and Joelle Pineau, editors, *Proceedings of the 29th International Conference on Machine Learning (ICML-12)*, ICML ’12, pages 81–88, New York, NY, USA, July 2012. Omnipress.
- [58] Quoc V Le, Will Y Zou, Serena Y Yeung, and Andrew Y Ng. Learning hierarchical invariant spatio-temporal features for action recognition with independent subspace analysis. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 3361–3368. IEEE, 2011.
- [59] Yann LeCun, Bernhard Boser, John S Denker, Donnie Henderson, Richard E Howard, Wayne Hubbard, and Lawrence D Jackel. Backpropagation applied to handwritten zip code recognition. *Neural computation*, 1(4):541–551, 1989.
- [60] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.
- [61] K. Lenc, V. Gulshan, and A. Vedaldi. VLFeat benchmarks. <http://www.vlfeat.org/benchmarks/xsxs>, 2011.
- [62] Stephan Leutenegger, Margarita Chli, and Roland Siegwart. Brisk: Binary robust invariant scalable keypoints. In *International Conference on Computer Vision (ICCV)*, pages 2548–2555, Nov 2011.
- [63] Rainer W Lienhart. Comparison of automatic shot boundary detection algorithms. In *Electronic Imaging’99*, pages 290–301. International Society for Optics and Photonics, 1998.

- [64] Min Lin, Qiang Chen, and Shuicheng Yan. Network in network. *Arxiv Preprint Arxiv:1312.4400*, 2013.
- [65] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *Computer Vision–ECCV 2014*, pages 740–755. Springer, 2014.
- [66] Tony Lindeberg. Feature detection with automatic scale selection. *International Journal of Computer Vision (IJCV)*, 30(2):79–116, 1998.
- [67] Tony Lindeberg and Jonas Gårding. Shape-adapted smoothing in estimation of 3-d shape cues from affine deformations of local 2-d brightness structure. *Image and Vision Computing*, 15(6):415–434, 1997.
- [68] Benny PL Lo, Ara Darzi, and Guang-Zhong Yang. Episode classification for the analysis of tissue/instrument interaction with multiple visual cues. In *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2003*, pages 230–237. Springer, 2003.
- [69] Constantinos Loukas and Evangelos Georgiou. Smoke detection in endoscopic surgery videos: a first step towards retrieval of semantic events. *The International Journal of Medical Robotics and Computer Assisted Surgery*, 11(1):80–94, 2015.
- [70] David G Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision (IJCV)*, 60(2):91–110, 2004.
- [71] Mathias Lux, Oge Marques, Klaus Schöffmann, Laszlo Böszörmenyi, and Georg Lajtai. A novel tool for summarization of arthroscopic videos. *Multimedia Tools and Applications*, 46(2-3):521–544, 2010.
- [72] Michal Mackiewicz, Jeff Berens, and Mark Fisher. Wireless capsule endoscopy color video segmentation. *Medical Imaging, IEEE Transactions on*, 27(12):1769–1781, 2008.
- [73] Elmar Mair, Gregory D Hager, Darius Burschka, Michael Suppa, and Gerhard Hirzinger. Adaptive and generic corner detection based on the accelerated segment test. In *European Conference on Computer Vision*, pages 183–196. Springer, 2010.
- [74] Santiago Manen, Matthieu Guillaumin, and Luc Gool. Prime object proposals with randomized prim’s algorithm. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2536–2543, 2013.

- [75] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide-baseline stereo from maximally stable extremal regions. *Image and Vision Computing*, 22(10):761–767, 2004.
- [76] H. Meine, U. Köthe, and P. Stelldinger. A topological sampling theorem for robust boundary reconstruction and image segmentation. *Discrete Applied Mathematics*, 157(3):524–541, 2009.
- [77] K. Mikolajczyk and C. Schmid. An affine invariant interest point detector. In *European Conference on Computer Vision (ECCV)*, pages 128–142. Springer, 2002.
- [78] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L.V Gool. A comparison of affine region detectors. *International Journal of Computer Vision (IJCV)*, 65(1):43–72, 2005.
- [79] K. Mikolajczyk, A. Zisserman, and C. Schmid. Shape recognition with edge-based features. In *British Machine Vision Conference (BMVC)*, volume 2, pages 779–788, 2003.
- [80] Henning Müller, Nicolas Michoux, David Bandon, and Antoine Geissbuhler. A review of content-based image retrieval systems in medical applications—clinical benefits and future directions. *International journal of medical informatics*, 73(1):1–23, 2004.
- [81] L. Najman and M. Couplie. Building the component tree in quasi-linear time. *IEEE Transactions on Image Processing*, 15(11):3531–3539, 2006.
- [82] Štěpán Obdržálek and Jiří Matas. Object recognition using local affine frames on maximally stable extremal regions. In *Toward Category-Level Object Recognition*, volume 4170 of *Lecture Notes in Computer Science*, pages 83–104. Springer, 2006.
- [83] Maxime Oquab, Leon Bottou, Ivan Laptev, and Josef Sivic. Learning and transferring mid-level image representations using convolutional neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1717–1724, 2014.
- [84] Victor Ostromoukhov. A simple and efficient error-diffusion algorithm. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pages 567–572. ACM, 2001.
- [85] Wai-Man Pang, Yingge Qu, Tien-Tsin Wong, Daniel Cohen-or, and Pheng-Ann Heng. Structure-aware halftoning. In *ACM Transactions on Graphics (TOG)*, volume 27, page 89. ACM, 2008.

- [86] Sun Young Park, Dustin Sargent, Inbar Spofford, Kirby G Vosburgh, et al. A colon video analysis framework for polyp detection. *Biomedical Engineering, IEEE Transactions on*, 59(5):1408–1418, 2012.
- [87] Nilesh V Patel and Ishwar K Sethi. Video shot detection and characterization for video databases. *Pattern Recognition*, 30(4):583–592, 1997.
- [88] Mattis Paulin, Matthijs Douze, Zaid Harchaoui, Julien Mairal, Florent Perronnin, and Cordelia Schmid. Local convolutional features with unsupervised training for image retrieval. In *The IEEE International Conference on Computer Vision (ICCV)*, December 2015.
- [89] James Philbin, Ondrej Chum, Michael Isard, Josef Sivic, and Andrew Zisserman. Object retrieval with large vocabularies and fast spatial matching. In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, pages 1–8. IEEE, 2007.
- [90] James Philbin, Ondrej Chum, Michael Isard, Josef Sivic, and Andrew Zisserman. Lost in quantization: Improving particular object retrieval in large scale image databases. In *IEEE Conference on Computer Vision and Pattern Recognition, (CVPR)*, pages 1–8. IEEE, 2008.
- [91] Manfred Jurgen Primus, Klaus Schoeffmann, and Laszlo Boszormenyi. Segmentation of recorded endoscopic videos by detecting significant motion changes. In *Content-Based Multimedia Indexing (CBMI), 2013 11th International Workshop on*, pages 223–228. IEEE, 2013.
- [92] Gwénolé Quellec, Mathieu Lamard, Guy Cazuguel, Zakarya Droueche, Christian Roux, and Béatrice Cochener. Real-time retrieval of similar videos with application to computer-aided retinal surgery. In *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*, pages 4465–4468. IEEE, 2011.
- [93] Esa Rahtu, Juho Kannala, and Matthew Blaschko. Learning a category independent object detection cascade. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 1052–1059. IEEE, 2011.
- [94] K. Rapantzikos, Y. Avrithis, and S. Kollias. Detecting regions from single scale edges. In *Intern. Workshop on Sign, Gesture and Activity (SGA), European Conference on Computer Vision (ECCV)*, volume 6553 of *Lecture Notes in Computer Science*, pages 298–311. Springer Berlin Heidelberg, September 2010.
- [95] Ali Razavian, Hossein Azizpour, Josephine Sullivan, and Stefan Carlsson. Cnn features off-the-shelf: an astounding baseline for recognition. In *Proceedings of the*

- IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 806–813, 2014.
- [96] Frank Rosenblatt. *The perceptron, a perceiving and recognizing automaton Project Para*. Cornell Aeronautical Laboratory, 1957.
- [97] Paul L. Rosin. Measuring shape: Ellipticity, rectangularity, and triangularity. *Machine Vision and Applications*, 14(3):172–184, 2003.
- [98] Edward Rosten and Tom Drummond. Fusing points and lines for high performance tracking. In *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, volume 2, pages 1508–1515. IEEE, 2005.
- [99] Edward Rosten and Tom Drummond. Machine learning for high-speed corner detection. In *European Conference on Computer Vision (ECCV)*, pages 430–443, May 2006.
- [100] Edward Rosten, Reid Porter, and Tom Drummond. Faster and better: A machine learning approach to corner detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(1):105–119, 2010.
- [101] Ethan Rublee, Vincent Rabaud, Kurt Konolige, and Gary Bradski. Orb: an efficient alternative to sift or surf. In *International Conference on Computer Vision (ICCV)*, pages 2564–2571. IEEE, Nov 2011.
- [102] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning representations by back-propagating errors. *Cognitive modeling*, 5(3):1, 1988.
- [103] Pierre Sermanet, David Eigen, Xiang Zhang, Michael Mathieu, Rob Fergus, and Yann Lecun. Overfeat: Integrated recognition, localization and detection using convolutional networks. In *International Conference on Learning Representations (ICLR 2014)*. Cbls, April 2014.
- [104] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In *in Proceedings of International Conference on Learning Representations*, 2015.
- [105] Ju Sun, Xiao Wu, Shuicheng Yan, Loong-Fah Cheong, Tat-Seng Chua, and Jintao Li. Hierarchical spatio-temporal context modeling for action recognition. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 2004–2011. IEEE, 2009.

- [106] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–9, 2015.
- [107] M. Teichmann and M. Capps. Surface reconstruction with anisotropic density-scaled alpha shapes. In *IEEE Visualization*, pages 67–72. IEEE, 1998.
- [108] Federico Tombari, Alessandro Franchi, and Luigi Di Stefano. Bold features to detect texture-less objects. In *International Conference on Computer Vision (ICCV)*, pages 1265–1272, Dec 2013.
- [109] Du Tran, Lubomir Bourdev, Rob Fergus, Lorenzo Torresani, and Manohar Paluri. Learning spatiotemporal features with 3d convolutional networks. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4489–4497, 2015.
- [110] T. Tuytelaars and L. Van Gool. Matching widely separated views based on affine invariant regions. *International Journal of Computer Vision*, 59(1):61–85, 2004.
- [111] Tinne Tuytelaars. Dense interest points. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 2281–2288. IEEE, 2010.
- [112] Jasper RR Uijlings, Koen EA van de Sande, Theo Gevers, and Arnold WM Smeulders. Selective search for object recognition. *International journal of computer vision*, 104(2):154–171, 2013.
- [113] Michael Van den Bergh, Gemma Roig, Xavier Boix, Santiago Manen, and Luc Van Gool. Online video seeds for temporal window objectness. In *Computer Vision (ICCV), 2013 IEEE International Conference on*, pages 377–384. IEEE, 2013.
- [114] A. Vedaldi and B. Fulkerson. VLFeat: An open and portable library of computer vision algorithms. <http://www.vlfeat.org/>, 2008.
- [115] Subhashini Venugopalan, Marcus Rohrbach, Jeff Donahue, Raymond Mooney, Trevor Darrell, and Kate Saenko. Sequence to sequence – video to text. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2015.
- [116] Subhashini Venugopalan, Huijuan Xu, Jeff Donahue, Marcus Rohrbach, Raymond Mooney, and Kate Saenko. Translating videos to natural language using deep recurrent neural networks. In *NAACL HLT*, 2015.

- [117] Etienne Vincent and Robert Laganiére. Detecting planar homographies in an image pair. In *Image and Signal Processing and Analysis, 2001. ISPA 2001. Proceedings of the 2nd International Symposium on*, pages 182–187. IEEE, 2001.
- [118] Heng Wang, Alexander Kläser, Cordelia Schmid, and Cheng-Lin Liu. Dense trajectories and motion boundary descriptors for action recognition. *International journal of computer vision*, 103(1):60–79, 2013.
- [119] Heng Wang, Dan Oneata, Jakob Verbeek, and Cordelia Schmid. A robust and efficient video representation for action recognition. *International Journal of Computer Vision*, pages 1–20, July 2015.
- [120] Y. Yang, M.N. Wernick, and J.G Brankov. A fast approach for accurate content-adaptive mesh generation. *Image Processing, IEEE Transactions on*, 12(8):866–881, 2003.
- [121] Jason Yosinski, Jeff Clune, Yoshua Bengio, and Hod Lipson. How transferable are features in deep neural networks? In Z. Ghahramani, M. Welling, C. Cortes, N.D. Lawrence, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 3320–3328. Curran Associates, Inc., 2014.
- [122] Jinhui Yuan, Huiyi Wang, Lan Xiao, Wujie Zheng, Jianmin Li, Fuzong Lin, and Bo Zhang. A formal study of shot boundary detection. *Circuits and Systems for Video Technology, IEEE Transactions on*, 17(2):168–186, 2007.
- [123] Joe Yue-Hei Ng, Matthew Hausknecht, Sudheendra Vijayanarasimhan, Oriol Vinyals, Rajat Monga, and George Toderici. Beyond short snippets: Deep networks for video classification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4694–4702, 2015.
- [124] Qiyang Zhao, Zhibin Liu, and Baolin Yin. Cracking bing and beyond. In *British Machine Vision Conference (BMVC)*, 2014.
- [125] Bingfeng Zhou and Xifeng Fang. Improving mid-tone quality of variable-coefficient error diffusion using threshold modulation. In *ACM Transactions on Graphics (TOG)*, volume 22, pages 437–444. ACM, 2003.
- [126] C.L. Zitnick and K. Ramnath. Edge foci interest points. In *International Conference on Computer Vision (ICCV)*, pages 359–366, 2011.
- [127] A. Zomorodian, L. Guibas, and P. Koehl. Geometric filtering of pairwise atomic interactions applied to the design of efficient statistical potentials. *Computer-Aided Geometric Design*, 23(6):531–544, 2006.

Κατάλογος Δημοσιεύσεων

- [1] Christos Varytimidis, Konstantinos Rapantzikos, Yannis Avrithis, and Stefanos Kollias. α -shapes for local feature detection. *Pattern Recognition*, 50:56 – 73, 2016.
- [2] Christos Varytimidis, Konstantinos Rapantzikos, Yannis Avrithis, and Stefanos Kollias. Dithering-based sampling and weighted α -shapes for local feature detection. *IPSJ Transactions on Computer Vision and Applications*, 7:189–200, 2015.
- [3] Christos Varytimidis, Konstantinos Rapantzikos, and Yannis Avrithis. W α sh: Weighted α -shapes for local feature detection. In *European Conference on Computer Vision (ECCV 2012)*, Florence, Italy, October 2012.
- [4] Christos Varytimidis, Konstantinos Rapantzikos, Yannis Avrithis, and Stefanos Kollias. Improving local features by dithering-based image sampling. In *in Proceedings of Asian Conference on Computer Vision (ACCV 2014)*, Singapore, November 2014, (oral presentation).
- [5] Christos Varytimidis, Konstantinos Rapantzikos, Constantinos Loukas, and Stefanos Kollias. Surgical video retrieval using deep neural networks. In *International Conference on Image Processing (ICIP 2016)*, 2016, (submitted).
- [6] Christos Varytimidis, Konstantinos Rapantzikos, and Stefanos Kollias. W α shing visual repositories: searching europeana using local features. In *International Conference on Digital Signal Processing (DSP 2013)*, July 2013, (oral presentation).
- [7] Christina Lakka, Spiros Nikolopoulos, Christos Varytimidis, and Ioannis Kompatsiaris. A bayesian network modeling approach for cross media analysis. *Signal Processing: Image Communication*, 26(3):175–193, 2011.
- [8] Spiros Nikolopoulos, Christina Lakka, Yiannis Kompatsiaris, Christos Varytimidis, Konstantinos Rapantzikos, and Yannis Avrithis. Visual image retrieval and localization. In *Compound document analysis by fusing evidence across media (CBMI 2009)*, Chania, Greece, June 2009.