ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΕΦΑΡΜΟΣΜΕΝΩΝ ΜΑΘΗΜΑΤΙΚΩΝ ΚΑΙ ΦΥΣΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΟΜΕΑΣ ΜΑΘΗΜΑΤΙΚΩΝ

Δ.Π.Μ.Σ. «ΜΑΘΗΜΑΤΙΚΗ ΠΡΟΤΥΠΟΠΟΙΗΣΗ ΣΕ ΣΥΓΧΡΟΝΕΣ ΤΕΧΝΟΛΟΓΙΕΣ ΚΑΙ ΤΗΝ ΟΙΚΟΝΟΜΙΑ»

# Ανάλυση και Αναζήτηση Εικόνων με Μεθόδους Ανίχνευσης Τοπικών Χαρακτηριστικών

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Γεώργιος Π. Κούμουλος

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

Αθήνα, Μάρτιος 2009

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΕΦΑΡΜΟΣΜΕΝΩΝ ΜΑΘΗΜΑΤΙΚΩΝ ΚΑΙ ΦΥΣΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΟΜΕΑΣ ΜΑΘΗΜΑΤΙΚΩΝ

Δ.Π.Μ.Σ. «ΜΑΘΗΜΑΤΙΚΗ ΠΡΟΤΥΠΟΠΟΙΗΣΗ ΣΕ ΣΥΓΧΡΟΝΕΣ ΤΕΧΝΟΛΟΓΙΕΣ ΚΑΙ ΤΗΝ ΟΙΚΟΝΟΜΙΑ»

# Ανάλυση και Αναζήτηση Εικόνων με Μεθόδους Ανίχνευσης Τοπικών Χαρακτηριστικών

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Γεώργιος Π. Κούμουλος

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

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την ….η Μαρτίου 2009.

| ..... | ..... | ..... |
|---|---|---|
| Στέφανος Κόλλιας<br>Καθηγητής Ε.Μ.Π. | Ανδρέας-Γεώργιος Σταφυλοπάτης<br>Καθηγητής Ε.Μ.Π. | Γεώργιος Στάμου<br>Λέκτορας Ε.Μ.Π. |

Αθήνα, Μάρτιος 2009

.....

**ΓΕΩΡΓΙΟΣ Π. ΚΟΥΜΟΥΛΟΣ**

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

# Περίληψη

Ο μεγάλος όγκος οπτικών δεδομένων και η εύκολη πρόσβαση σε αυτά μέσω του διαδικτύου έχει οδηγήσει στην ανάγκη για την σύντομη περιγραφή του σημασιολογικού περιεχομένου κάθε εικόνας. Οι τεχνικές για τη γρήγορη αναζήτηση εικόνων αποτελούν αντικείμενο έρευνας που συνεχώς εξελίσσεται, ενώ τα τελευταία χρόνια τα οφέλη από τη χρήση των τοπικών χαρακτηριστικών γίνονται όλο και πιο εμφανή.

Στην παρούσα εργασία αρχικά μελετώνται οι πιο γνωστές μέθοδοι ανίχνευσης τοπικών χαρακτηριστικών και εξαγωγής τοπικών περιγραφέων. Στη συνέχεια κατασκευάζεται ένα σύστημα τοπικών χαρακτηριστικών με συνδυασμό των μεθόδων και ακολουθεί η πειραματική σύγκριση των επιδόσεών τους με βάση αντικειμενικά κριτήρια. Πιο συγκεκριμένα, με τη βοήθεια ενός συνόλου από εικόνες διαφόρων σκηνών που υπόκεινται σε μετασχηματισμούς (σύνολο αναφοράς), υπολογίζονται οι επιδόσεις των μεθόδων ως προς την επαναληψιμότητα των σημείων τους, την ακρίβεια εντοπισμού τους και την ικανότητα ταιριάσματος των περιγραφέων τους.

Η αποδοτικότητα των τοπικών χαρακτηριστικών εξετάζεται μέσω ενός συστήματος αναζήτησης εικόνων σε μεγάλες βάσεις δεδομένων, απ' όπου προκύπτει μία ποσοτική σύγκριση των τεχνικών ως προς τα ποσοστά επιτυχίας του συστήματος. Στο σύστημα αυτό, με μηχανισμό ανάλογο της αναζήτησης κειμένου, δημιουργείται ένα οπτικό λεξικό και το περιεχόμενο κάθε εικόνας αναπαρίσταται από ένα διάνυσμα. Η αναζήτηση γίνεται με την εκτίμηση της ομοιότητας μεταξύ των διανυσμάτων αναπαράστασης. Τα οπτικά λεξικά που έχουν προκύψει από τις διαφορετικές μεθόδους τοπικών χαρακτηριστικών συγκρίνονται με βάση τα μέτρα αξιολόγησης της αναζήτησης εικόνων, όπως είναι τα μέτρα ακριβείας, ανάκτησης και μέσης ακριβείας (mean Average Precision - mAP).

## Λέξεις – κλειδιά:

τοπικά χαρακτηριστικά, σημείο – περιοχή ενδιαφέροντος, τοπικός περιγραφέας, Harris-Affine, Hessian-Affine, MSER, SIFT, SURF, αναγνώριση αντικειμένων, αναζήτηση εικόνων, δεικτοδότηση εικόνων, οπτικό λεξικό, μέτρο μέσης ακριβείας (mAP)

# Abstract

The large amount of optical information and the easy access to available data through the internet has led to the emerging need for efficient description of image content. Many techniques for fast image retrieval have been proposed in literature, but in the recent years the use local features has come to maturity because of their efficiency.

In this thesis work, the most known methods of local detectors and descriptors are firstly studied. Next, an integrated system of local invariant features is implemented, with the use of already tested techniques, and different methods are combined in order to compose a powerful tool for image analysis. The experimental evaluation follows, which is done over a standard set (benchmark) of images under various transformations (photometric and geometric). All previously analyzed methods are compared via objective criteria: repeatability score, accuracy of detectors (localization), matching score and performance of descriptors.

The efficiency of local features is testified in the image retrieval system with the use of large image databases. The experimental procedure provides a quantitative comparison of the aforementioned techniques. The main image retrieval mechanism is related to text retrieval methods: a visual vocabulary is created and a model vector is constructed for each image, which represents its semantic content. The image retrieval procedure is done through vector similarity measures. Different visual vocabularies (generated by various local feature methods) are compared with respect to image retrieval evaluation criteria, like precision, recall and mean Average Precision (mAP).

## Keywords:

local features, interest point – region, keypoint, local detector – descriptor, Harris-Affine, Hessian-Affine, MSER, SIFT, SURF, image matching, object recognition, image retrieval, image indexing, visual vocabulary, mean Average Precision (mAP)

# Ευχαριστίες

Η παρούσα μεταπτυχιακή διπλωματική εργασία εκπονήθηκε κατά το ακαδημαϊκό έτος 2008-2009 στο Εργαστήριο Ψηφιακής Επεξεργασίας Εικόνας, Βίντεο και Πολυμέσων (IVML) του Εθνικού Μετσόβιου Πολυτεχνείου. Θα ήθελα να ευχαριστήσω τον επιβλέποντα Καθηγητή κ. Στέφανο Κόλλια για την εμπιστοσύνη που μου έδειξε αναθέτοντάς μου την εργασία αυτή και για τη δυνατότητα που μου έδωσε να ασχοληθώ με το συγκεκριμένο ενδιαφέρον θέμα. Θα ήθελα επίσης να ευχαριστήσω τον Ερευνητή ΕΠΙΣΕΥ – ΕΜΠ Δρ Ιωάννη Αβρίθη και τον Ερευνητή Δρ Κωνσταντίνο Ραπαντζίκο, για την συνεχή καθοδήγησή τους, τις πολύτιμες συμβουλές τους και τον χρόνο που μου αφιέρωσαν. Τέλος, θέλω να ευχαριστήσω τον φίλο Γιάννη Καλαντίδη για τη συμβολή του με την υλοποίηση και τις χρήσιμες υποδείξεις του σε ό,τι αφορούσε το σύστημα αναζήτησης εικόνων.

# Περιεχόμενα

# ΚΕΦΑΛΑΙΟ 1

# Εισαγωγή

## 1.1 Αντικείμενο της μελέτης

Οι πληροφορίες που περιέχονται σε μια εικόνα, είτε αυτή είναι φωτογραφία του πραγματικού κόσμου είτε είναι κάποιο τεχνητό σχέδιο, παίζουν πολύ σημαντικό ρόλο σε όλες τις πτυχές του σύγχρονου τρόπου ζωής καθώς χρησιμοποιούνται συχνά σε επαγγελματικά, ενημερωτικά, εκπαιδευτικά ή και ψυχαγωγικά θέματα. Επιπλέον, η εύκολη πρόσβαση σε μεγάλο όγκο δεδομένων μέσω του διαδικτύου έχει οδηγήσει στην ανάγκη για την κατάλληλη περιγραφή του περιεχομένου κάθε εικόνας με σύντομο αλλά ουσιαστικό τρόπο, ούτως ώστε η ομαδοποίηση των πληροφοριών αλλά και η αναζήτηση των ζητούμενων εικόνων να γίνεται αποδοτικά αλλά και αξιόπιστα.

Ένας απλός τρόπος για την ανάκτηση εικόνων από μία μεγάλη βάση δεδομένων είναι η αναζήτηση με βάση το κείμενο. Ο χρήστης δίνει στο ερώτημά του ορισμένες λέξεις-κλειδιά που αντιστοιχούν σε έννοιες και αντικείμενα τα οποία επιθυμεί να περιλαμβάνονται στο αποτέλεσμα της αναζήτηση. Η διαδικασία είναι παρόμοια με τις μεθόδους αναζήτησης κειμένου. Γίνεται δηλαδή ανάκτηση εικόνων στην περίπτωση που αυτές περιέχουν τις συγκεκριμένες λέξεις, είτε στο γύρω χώρο της εικόνας ως λεζάντα ή σχόλια είτε σε δεδομένα που συνοδεύουν την εικόνα (μεταδεδομένα). Αυτή είναι και η μέθοδος που χρησιμοποιείται ευρύτατα στις σύγχρονες μηχανές αναζήτησης εικόνων στο διαδίκτυο, έχει όμως ορισμένα πολύ σημαντικά μειονεκτήματα. Το αποτέλεσμα της αναζήτησης εξαρτάται από το λεκτικό περιεχόμενο, το οποίο καθορίζεται από τον ανθρώπινο παράγοντα και για τον λόγο αυτό δεν είναι πάντοτε αντικειμενικό και αξιόπιστο. Επιπλέον, ο όγκος του διαθέσιμου πολυμεσικού περιεχομένου αυξάνεται με συνεχείς ρυθμούς καθιστώντας αδύνατη την πλήρη λεκτική περιγραφή του.

Από τα παραπάνω γίνεται φανερό ότι η δυνατότητα αυτόματης περιγραφής κάθε διαθέσιμης εικόνας, και μάλιστα με βάση το σημασιολογικό της περιεχόμενο, συνεισφέρει στη σωστή εκμετάλλευση της υπάρχουσας πληροφορίας. Αυτό μπορεί να επιτευχθεί εξάγοντας από την εικόνα κάποια χαρακτηριστικά που περιέχουν την σημαντικότερη πληροφορία της και έτσι να αναπαρασταθεί σύντομα και περιεκτικά. Τα χαρακτηριστικά αυτά ανήκουν σε τρεις κατηγορίες: μπορεί να είναι γεωμετρικά, ολικά ή τοπικά. Όσον αφορά την πρώτη κατηγορία, πρώτα μοντελοποιούνται τα αντικείμενα με συγκεκριμένα τρισδιάστατα σχήματα, όπως είναι οι γραμμές, οι κορυφές και οι ελλείψεις, και στη συνέχεια αναζητούνται αυτά τα γεωμετρικά χαρακτηριστικά μέσα στην εικόνα με σκοπό τον εντοπισμό των αντικειμένων. Το κύριο μειονέκτημα της μεθοδολογίας αυτής είναι ότι απαιτεί την ύπαρξη αυτών των συγκεκριμένων γεωμετρικών ιδιοτήτων μέσα στην εικόνα, κάτι που δεν συμβαίνει απόλυτα σε εικόνες του φυσικού κόσμου. Για τον λόγο αυτό οι τεχνικές αυτές ενώ χρησιμοποιήθηκαν την δεκαετία του 1980 εγκαταλείφθηκαν

αργότερα και έτσι στις αρχές της δεκαετίας του 1990 το ενδιαφέρον στράφηκε στα λεγόμενα ολικά χαρακτηριστικά της εικόνας (global features), τα οποία δεν αναζητούν συγκεκριμένα σχέδια, αλλά χρησιμοποιούν ό,τι περιέχεται πραγματικά στην εικόνα. Έτσι αρχικά χρησιμοποιούνται ιστογράμματα χρώματος ή φωτεινότητας τα οποία κατασκευάζονται από τα pixels ολόκληρης της εικόνας και με τον τρόπο αυτό προκύπτει η "υπογραφή" φωτεινότητας της εικόνας, δηλαδή η συνολική αναπαράσταση της πληροφορίας της.

Παρόλο που τα ολικά χαρακτηριστικά όπως τα ιστογράμματα είναι αρκετά σταθερά και μένουν ανεπηρέαστα από μικρές μεταβολές του περιεχομένου, από σφάλματα και θόρυβο, παρουσιάζουν και αυτά πολλά μειονεκτήματα και προκύπτουν ακατάλληλα για το συγκεκριμένο πρόβλημα που καλούνται να αντιμετωπίσουν. Αποτυγχάνουν να αναπαραστήσουν επαρκώς το περιεχόμενο μιας εικόνας στην περίπτωση που υπάρχει μερική επικάλυψη αντικειμένων (partial visibility, occlusion), στην περίπτωση πολλών "εξωγενών" περιττών χαρακτηριστικών και αποπροσανατολισμού λόγω συγκεχυμένου περιβάλλοντος (cluttered images), καθώς και στην περίπτωση που η εικόνα περιέχει επιμέρους αντικείμενα (ή κομμάτια τους) που ανήκουν σε πολλές διαφορετικές εικόνες της βάσης δεδομένων. Οι παραπάνω συνθήκες ισχύουν ως επί των πλείστων σε συστήματα ανάκτησης εικόνων με βάση το περιεχόμενο και γι' αυτό ο πιο κατάλληλος τρόπος για να αναπαρασταθεί ικανοποιητικά η εικόνα είναι ο τοπικός υπολογισμός της πληροφορίας της. Τα τοπικά χαρακτηριστικά είναι συγκεκριμένα σημεία της εικόνας που περιέχουν πολύ σημαντική πληροφορία και από την γύρω περιοχή τους εξάγονται ο τοπικοί περιγραφείς που αναπαριστούν με αποδοτικό τρόπο το οπτικό περιεχόμενο της εικόνας. Η ανάλυση των εικόνων με τις μεθόδους τοπικών χαρακτηριστικών (local features) έχει εφαρμογές, εκτός από τα προβλήματα αναζήτησης εικόνων ανάμεσα σε τεράστιο όγκο διαθέσιμου υλικού (image retrieval), επίσης σε θέματα ανακατασκευής του τρισδιάστατου μοντέλου της σκηνής (scene reconstruction), σε ταίριασμα επιμέρους εικόνων από διάφορες οπτικές γωνίες (wide baseline image matching image stitching), κατασκευή πανοράματος τοπίων (panorama building), καθώς και σε προβλήματα κατηγοριοποίησης και αναγνώρισης αντικειμένων (object categorization and recognition).

## 1.2 Τεχνικές εξαγωγής τοπικών χαρακτηριστικών

Τα τοπικά χαρακτηριστικά είναι τμήματα της εικόνας τα οποία διαφέρουν από το άμεσο γειτονικό τους περιβάλλον. Δεδομένου ότι πρέπει να περιέχουν σημαντική πληροφορία, σχετίζονται συχνά με αλλαγές σε κάποια ιδιότητα της εικόνας (π.χ. ένταση), ενώ μπορεί να βρίσκονται είτε ακριβώς στο σημείο της αλλαγής είτε πολύ κοντά σε αυτήν. Τα τοπικά χαρακτηριστικά εμφανίζονται σε διάφορες μορφές: μπορούν να είναι σημεία, οπότε ονομάζονται σημεία ενδιαφέροντος (interest points) ή σημεία-κλειδιά (keypoints), να είναι περιοχές που αποτελούνται από περισσότερα pixels και λέγονται περιοχές ενδιαφέροντος (interest regions), ή ακόμα να είναι οι ακμές της εικόνας (edges). Μετά των εντοπισμό των τοπικών αυτών πληροφοριών ακολουθεί η κατασκευή ενός διανύσματος που να περιγράφει κατάλληλα το περιεχόμενο στο συγκεκριμένο σημείο. Η εξαγωγή αυτού του διανύσματος γίνεται είτε ακριβώς από την εντοπισμένη περιοχή ενδιαφέροντος είτε από μία περιοχή της εικόνας κεντραρισμένη στο χαρακτηριστικό αλλά που να περιλαμβάνει και την αλλαγή που βρίσκεται κοντά του. Όλα μαζί τα διανύσματα χαρακτηριστικών που υπολογίζονται σε όλα τα τοπικά χαρακτηριστικά της εικόνας αποτελούν και την αναπαράσταση του σημασιολογικού της περιεχομένου.

Η ανίχνευση των ακμών αποτελεί μία πολύ διαδεδομένη μορφή επεξεργασίας της εικόνας, που οδηγεί συχνά σε επίλυση αρκετών προβλημάτων, όπως για παράδειγμα στην επιστήμη της τηλεπισκόπησης όπου σε διάφορες αεροφωτογραφίες οι ακμές αντιστοιχούν σε δρόμους, γραμμές τραίνων, ηλεκτρικά δίκτυα, κορυφογραμμές, ακτογραμμές και άλλα χρήσιμα

χαρακτηριστικά. Επίσης μία άλλη τεχνική είναι η εξαγωγή σημείων από το περίγραμμα των αντικειμένων, είτε από διασταυρώσεις ακμών ακολουθώντας την αλυσίδα των pixels κάθε ακμής, είτε από περιοχές όπου παρατηρείται μεγάλη τιμή στην καμπυλότητα [Langridge, 1982], είτε ακόμα από περιοχές όπου η παράγωγος αυξάνεται και παρατηρείται αλλαγή στην κατεύθυνση [Kitchen et al., 1982]. Ενώ όμως οι ακμές αποτελούν ένα ισχυρό χαρακτηριστικό υπό συνθήκες διαφόρων μεταβολών και θορύβου, δεν μπορούν να καλύψουν το γενικότερο πρόβλημα στο οποίο τα αντικείμενα δεν εμφανίζονται απαραίτητα με σαφή και καθαρά όρια.

Πάνω στην ίδια ιδέα υπολογισμού της πρώτης παραγώγου του πεδίου της φωτεινότητας της εικόνας βασίζονται και οι έρευνες που ακολούθησαν με σκοπό την εξαγωγή γωνιών (corners), οι οποίες και αντιπροσωπεύουν καλύτερα (σε σχέση με τις ακμές) τις τοπικές ιδιότητες που απαιτούνται στις διάφορες εφαρμογές. Ως γωνιακό χαρακτηρίζεται ένα σημείο της εικόνας στο οποίο παρατηρείται τοπικά μεγάλη μεταβολή της έντασης ταυτόχρονα και σε δύο κατευθύνσεις (διδιάστατη εικόνα). Η αρχική ιδέα ήταν η χρήση ενός παραθύρου το οποίο μετακινείται τοπικά για λίγα pixels σε διάφορες κατευθύνσεις για να μετρήσει τη μέση μεταβολή της έντασης. Στη συνέχεια η ιδέα από τις πρώτες παραγώγους προχώρησε στην κατασκευή του λεγόμενου πίνακα ροπών δεύτερης τάξης και εξελίχθηκε από τους Harris και Stephens [Harris et al., 1988], οι οποίοι όρισαν μία αναλυτική έκφραση η οποία αποτελεί ένα μέτρο μεταβολής της έντασης στις δύο κατευθύνσεις (ιδιοτιμές του πίνακα). Η τελευταία αποτελεί την πιο διαδεδομένη και αποδοτική μέθοδο εύρεσης γωνιών σε εικόνες, που συνήθως ονομάζεται ανιχνευτής γωνιών του Harris (Harris corner detector).

Μία άλλη μέθοδος για τον εντοπισμό γωνιών είναι αυτή που προτάθηκε από τους Smith και Brady και η οποία είναι γνωστή με το όνομα SUSAN [Smith et al., 1997]. Η μέθοδος αυτή είναι αρκετά γρήγορη, αφού στην κυκλική γειτονιά κάθε σημείου της εικόνας υπολογίζει το ποσοστό των pixels που έχουν παρόμοια τιμή στη φωτεινότητα με το κέντρο και με τη χρήση κατάλληλης τιμής κατωφλίου επιλέγει γωνιακά σημεία ως τοπικά ελάχιστα του ποσοστού αυτού. Μία παρόμοια τεχνική συγκρίνει μόνο σημεία που βρίσκονται πάνω σε έναν κύκλο γύρω από το υποψήφιο γωνιακό σημείο, και όχι σε ολόκληρη τη γειτονιά του. Για την βελτίωση της ταχύτητας της μεθόδου προτείνεται ο πιο πρόσφατος ανιχνευτής γωνιών FAST [Rosten et al., 2006], στον οποίο ένα σημείο προσδιορίζεται ως γωνιακό αν υπάρχει ένα αρκετά μεγάλο σύνολο σημείων, τα οποία να βρίσκονται πάνω σε κύκλο ορισμένης ακτίνας και να έχουν τιμές έντασης πολύ μεγαλύτερες ή πολύ μικρότερες από το κέντρο.

Προς τα τέλη της δεκαετίας του 1990 έγινε από τους Schmid και Mohr μία πολύ σημαντική δημοσίευση, που επηρέασε έντονα τις εξελίξεις στο ερευνητικό πεδίο της αναζήτησης εικόνων και της αναγνώρισης αντικειμένων. Χρησιμοποιώντας τα σημεία ενδιαφέροντος από τη μέθοδο του Harris, κατασκευάζουν ένα ολοκληρωμένο σύστημα ανάκτησης εικόνων από μεγάλες βάσεις δεδομένων [Schmid et al., 1997]. Η καινούργια ιδέα δεν ήταν φυσικά η χρήση τοπικών χαρακτηριστικών, αλλά η επισήμανση των πολύ καλών επιδόσεών τους κάτω από συνθήκες μερικής επικάλυψης, αλλαγής κλίμακας ή παρουσίας ξένων περιττών αντικειμένων. Με λίγα λόγια, εξάγοντας σταθερά χαρακτηριστικά (όπως ένα σύνολο από παραγώγους που ονομάζεται local jet) επί του σημείου ενδιαφέροντος, επιτυγχάνουν μία σύντομη αναπαράσταση της εικόνας. Στη συνέχεια μέσω της δεικτοδότησης όλων των εικόνων με την τοπική περιγραφή τους γίνεται ταίριασμα και ανάκτηση της σωστής εικόνας από την βάση, χρησιμοποιώντας επίσης και κάποιους ημιτοπικούς περιορισμούς που βελτιώνουν το αποτέλεσμα.

Τα γωνιακά σημεία του Harris δεν επηρεάζονται από την περιστροφή της εικόνας. Όμως για να μένουν ανεπηρέαστα και από την αλλαγή κλίμακας στην προηγούμενη δημοσίευση παρουσιάζεται η προσέγγιση της πολυκλιμακωτής ανάλυσης, όπου δηλαδή ο εντοπισμός των σημείων γίνεται σε διάφορες κλίμακες. Το μειονέκτημα αυτής της μεθόδου είναι ότι προκύπτουν πολλά επαναλαμβανόμενα σημεία, δηλαδή σημεία που ουσιαστικά αντιπροσωπεύουν το ίδιο περιεχόμενο της εικόνας. Για την αποφυγή αυτού του προβλήματος οι Mikolajczyk και Schmid προτείνουν τα σημεία Harris Laplace τα οποία μένουν ανεπηρέαστα από την αλλαγή κλίμακας [Mikolajczyk et al., 2001]. Ο εντοπισμός των σημείων γίνεται με μία παραλλαγή της συνάρτησης

που χρησιμοποιούν οι Harris et al. και στη συνέχεια ακολουθεί η επιλογή εκείνων μόνο των σημείων στα οποία παρουσιάζεται τοπικό μέγιστο για μία κατάλληλα επιλεγμένη συνάρτηση που μεταβάλλεται με την αλλαγή της κλίμακας. Στην εργασία τους [Mikolajczyk et al., 2002] προχωρούν ένα βήμα πιο πέρα, με την ανεξαρτησία ως προς τις αφινικές μεταβολές της εικόνας. Ξεκινώντας από τα σημεία της προηγούμενης μεθόδου (scale invariant interest points), ακολουθείται μια επαναληπτική διαδικασία "αφινικής" προσαρμογής της περιοχής (γύρω από το γωνιακό σημείο) βασισμένη στον πίνακα ροπών δεύτερης τάξης και έτσι προκύπτουν τα σημεία Harris Affine, τα οποία ορίζουν μια γειτονιά γύρω τους η οποία δεν επηρεάζεται από αφινικούς μετασχηματισμούς (affine invariant interest points).

Μετά τα γωνιακά σημεία, ένα άλλο σημαντικό είδος τοπικών χαρακτηριστικών είναι οι κηλίδες (blobs), οι οποίες μπορούν να θεωρηθούν ως συμπληρωματικές των γωνιών αφού εντοπίζονται σε διαφορετικές περιοχές της εικόνας (για παράδειγμα μεταξύ των ακμών και όχι πάνω στις ενώσεις τους). Μία μέθοδος ανίχνευσης κηλίδων είναι η χρήση των στοιχείων του χεσσιανού πίνακα των παραγώγων δεύτερης τάξης, και ειδικότερα του ίχνους και της ορίζουσάς του, απ' όπου προκύπτουν οι ιδιότητες της τοπικής δομής της εικόνας. Οι Mikolajczyk και Schmid σε νέα εργασία τους παρουσιάζουν μαζί με τα τροποποιημένα σημεία Harris και τα αντίστοιχα σημεία κηλίδων Hessian Laplace και Hessian Affine [Mikolajczyk et al., 2004]. Η διαδικασία που ακολουθείται είναι παρόμοια με τα αντίστοιχα Harris, δηλαδή ξεκινώντας από τις εντοπισμένες κηλίδες, υπολογίζεται το τοπικό ελάχιστο κατάλληλης συνάρτησης ώστε να προκύψουν σημεία ανεξάρτητα της κλίμακας και στη συνέχεια επαναλαμβάνεται μια διαδικασία προσαρμογής οπότε και προκύπτουν σημεία σταθερά σε αφινικούς μετασχηματισμούς.

Μία άλλη περισσότερο θεωρητική προσέγγιση του θέματος εντοπισμού δομών κηλίδας αποτελεί η χρήση ενός μέτρου για την τοπική σημασία (saliency) μέσα στην εικόνα. Ένα τέτοιο μέτρο θα μπορούσε να είναι η μεταβλητότητα εντός μιας περιοχής της εικόνας. Έτσι η τοπική σημαντικότητα ορίζεται μέσω της πολυπλοκότητας του σήματος χρησιμοποιώντας την έννοια της εντροπίας από τη θεωρία του Shannon. Ως σημεία ενδιαφέροντος επιλέγονται εκείνα στα οποία υπάρχει γειτονιά με υψηλή συγκέντρωση πληροφορίας, δηλαδή μεγάλες μεταβολές στην εντροπία ενός ιστογράμματος φωτεινότητας γύρω από αυτά. Η μέθοδος αυτή επεκτάθηκε από τους Kadir και Brady [Kadir et al., 2001] για να μπορούν να επιλέγονται σημεία ανεξάρτητα της αλλαγής κλίμακας και στη συνέχεια προσαρμόστηκε ώστε να υπολογίζονται περιοχές γύρω από αυτά που θα εξακολουθούν να υπάρχουν μετά από αφινικούς μετασχηματισμούς της εικόνας [Kadir et al., 2004].

Υπάρχουν και κάποιες άλλες τεχνικές, που υπολογίζουν ταυτόχρονα τη θέση του σημείου ή της περιοχής ενδιαφέροντος (interest point/region), το μέγεθός δηλαδή την κλίμακα στην οποία εμφανίζεται (scale invariant), καθώς και το αφινικό σχήμα της τοπικής δομής (affine invariance). Ένα παράδειγμα είναι η μέθοδος που βασίζεται σε ακμές και ονομάζεται σύντομα EBR (edge-based regions), η οποία ξεκινώντας από τα γωνιακά σημεία Harris και τις προσκείμενες τεμνόμενες ακμές κατασκευάζει παραλληλόγραμμα τα οποία ουσιαστικά καλύπτουν την εσωτερική περιοχή των γωνιών [Tuytelaars et al., 1999]. Δύο σημεία διατρέχουν τις δύο ακμές μέχρι να βρεθεί τοπικό μέγιστο σε κάποια φωτομετρική ποσότητα, οπότε το παραλληλόγραμμο ορίζεται από τα δύο τελικά αυτά σημεία και την κορυφή. Η περιοχή ενδιαφέροντος που προκύπτει δεν επηρεάζεται από γεωμετρικούς μετασχηματισμούς, αλλά το κυριότερο μειονέκτημά της είναι ότι προϋποθέτει την ύπαρξη καθαρών συνεχών ακμών καθώς και τη συγκεκριμένη δομή παραλληλογράμμου στην εικόνα.

Μία εναλλακτική μέθοδος του ίδιου τύπου τοπικών χαρακτηριστικών (αφινικών περιοχών) δημοσιεύεται από τους ίδιους συγγραφείς στην εργασία τους [Tuytelaars et al., 2000], η οποία σύντομα ονομάζεται IBR (intensity-based regions). Ξεκινώντας από τοπικά ακρότατα της έντασης και προχωρώντας κατά μήκος συγκεκριμένων ακτινών που ξεκινούν από το ακρότατο (πιθανό σημείο ενδιαφέροντος), υπολογίζεται κάποιο μέτρο της αλλαγής της έντασης και σημαδεύεται το σημείο στο οποίο η αλλαγή μεγιστοποιείται. Ενώνοντας όλα αυτά τα σημεία από όλες τις ακτίνες δημιουργείται μία περιοχή στην οποία αν αντιστοιχηθεί μία έλλειψη, προκύπτει η

ζητούμενη αφινική περιοχή. Οι δύο μέθοδοι εντοπισμού περιοχών δημοσιεύονται από κοινού στην πιο πρόσφατη εργασία [Tuytelaars et al., 2004], όπου περιλαμβάνεται μελέτη για τη βελτίωση του ταιριάσματος εικόνων και σχετικά πειράματα.

Παρόμοια με την προηγούμενη προσέγγιση είναι η μέθοδος MSER που προτείνεται από τους Matas et al. [Matas et al., 2002], κατά την οποία ακολουθείται μια διαδικασία που θυμίζει κατάτμηση εικόνας με τη μέθοδο του πλημμυρισμού (watershed segmentation). Τα πιθανά σημεία ενδιαφέροντος είναι και εδώ τα τοπικά ακρότατα της έντασης και το αποτέλεσμα καθορίζεται από ένα κατώφλι που ξεχωρίζει περιοχές όπου το περίγραμμα έχει την μεγαλύτερη (ή την μικρότερη) ένταση και το εσωτερικό έχει παρόμοιες τιμές έντασης. Ως περιοχές ενδιαφέροντος επιλέγονται αυτές που παραμένουν σταθερές σε ένα ευρύ φάσμα τιμών κατωφλίου (από αυτή την ιδιότητα προκύπτει και το όνομα της μεθόδου: maximally stable extremal regions). Οι περιοχές ακανόνιστου σχήματος που προκύπτουν πλησιάζουν αρκετά στη δομή τις περιοχές της μεθόδου IBR και μπορούν και αυτές να αντικατασταθούν από ελλείψεις. Μάλιστα πρέπει να σημειωθεί ότι αυτές οι δύο τεχνικές παράγουν περιοχές ενδιαφέροντος που είναι εκ κατασκευής ανεξάρτητες των γεωμετρικών μετασχηματισμών (affine invariant), αντιστοιχούν σε δομές που μοιάζουν με κηλίδες, δηλαδή κεντρική περιοχή που διαφέρει έντονα στην ένταση από το περιβάλλον γύρω της, και όπως αποδεικνύεται ταιριάζουν καλύτερα σε εικόνες με ευδιάκριτες δομές (structured images).

Από τη σκοπιά της υλοποίησης αποδοτικότερων και ταχύτερων μεθόδων τοπικών χαρακτηριστικών, πρέπει να αναφέρουμε την πολύ σημαντική εργασία του Lowe [Lowe, 1999] στην οποία επιδιώκεται η επίλυση του προβλήματος αναγνώρισης αντικειμένων (αναζήτησης) σε εικόνες με πολύπλοκο περιβάλλον. Μέσω μιας διαδικασίας διαδοχικού φιλτραρίσματος της αρχικής εικόνας, τα σημεία ενδιαφέροντος εντοπίζονται ως τοπικά ακρότατα σε έναν χώρο κλίμακας που προσεγγίζει την λαπλασιανή της εικόνας, δηλαδή το ίχνος του χεσσιανού πίνακα. Στη συνέχεια από τη γειτονιά των σημείων αυτών εξάγεται ένα διάνυσμα χαρακτηριστικών που περιέχει τα στοιχεία του ιστογράμματος των κατευθύνσεων των πρώτων παραγώγων και αποκαλείται διάνυσμα SIFT. Όλα τα διανύσματα χαρακτηριστικών χρησιμοποιούνται για το ταίριασμα των εικόνων με κάποια δεδομένα μοντέλα αντικειμένων και μέσω μιας δεικτοδότησης κοντινότερου γείτονα αποφασίζεται η παρουσία ενός αντικειμένου μέσα σε μια εικόνα.

Προσπαθώντας να βελτιώσουν ακόμα περισσότερο την ταχύτητα των μεθόδων, οι Bay et al. προτείνουν τη χρήση ολοκληρωτικών εικόνων (integral images) προκειμένου να υπολογίσουν προσεγγιστικά τον χεσσιανό πίνακα και στη συνέχεια να εντοπίσουν τα σημεία ενδιαφέροντος σε έναν κατάλληλο χώρο κλίμακας [Bay et al., 2006]. Στη συνέχεια κατασκευάζουν ένα διάνυσμα χαρακτηριστικών για τετραγωνικές γειτονιές γύρω από τα σημεία, το οποίο αντιπροσωπεύει τις κατανομές των παραγώγων ως προς τον προσανατολισμό τους (ιστόγραμμα το οποίο χρησιμοποιεί και πάλι ολοκληρωτικές εικόνες) και ονομάζεται συντομογραφικά SURF. Οι δύο παραπάνω τεχνικές ανίχνευσης σημείων (SIFT και SURF) παράγουν τοπικά χαρακτηριστικά σε περιοχές κηλίδων συνήθως, είναι ανεξάρτητα από περιστροφή ή αλλαγή κλίμακας της εικόνας και πειραματικά αποδεικνύονται κατάλληλα για προβλήματα ανάκτησης και αναγνώρισης αντικειμένων.

Για να περιγραφεί το περιεχόμενο της εικόνας τοπικά, είναι απαραίτητο να κατασκευαστεί ένα διάνυσμα χαρακτηριστικών από την γειτονιά του σημείου ή της περιοχής ενδιαφέροντος. Στη βιβλιογραφία μπορεί να συναντήσει κανείς ποικίλες μεθόδους εξαγωγής χαρακτηριστικών, όπως είναι το σχηματικό περιβάλλον (shape context), τα κατευθυνόμενα φίλτρα (steerable filters), τα μιγαδικά φίλτρα (complex filters), σταθερά χαρακτηριστικά παραγώγων (differential invariants) ή ροπών (moment invariants) και η ετεροσυσχέτιση των τιμών δειγματοληπτούμενων pixels (cross-correlation). Όλες οι παραπάνω τεχνικές αναλύονται και συγκρίνονται εκτενώς μαζί με τον περιγραφέα SIFT στην πολύ σημαντική δημοσίευση [Mikolajczyk et al., 2003], απ' όπου και προκύπτει ότι το διάνυσμα χαρακτηριστικών SIFT είναι καταλληλότερο για να περιγράψει την τοπική δομή και το σημασιολογικό περιεχόμενο της εικόνας. Αυτό μπορεί να εξηγηθεί από το γεγονός ότι το ιστόγραμμα προσανατολισμού παραγώγων, το οποίο αποτελεί την κεντρική ιδέα

του περιγραφέα αυτού, μένει ανεπηρέαστο από μικρές μετατοπίσεις μερικών σημείων της εικόνας (σφάλματα και θόρυβος).

## 1.3 Σκοπός της εργασίας

Όπως είδαμε στην προηγούμενη παράγραφο, η χρήση των μεθόδων τοπικών χαρακτηριστικών αποτελεί αντικείμενο έρευνας στον χώρο της ανάλυσης εικόνας εδώ και τρεις δεκαετίες περίπου. Τα σημεία ενδιαφέροντος αρχικά αντιμετωπίζονταν ως ένα υποσύνολο των σημείων της εικόνας, που οδηγούσε σε μείωση της πολυπλοκότητας και αύξηση της ταχύτητας επεξεργασίας. Με την πάροδο του χρόνου όμως όλο και περισσότερο βάρος δινόταν στη σημασία των ιδιοτήτων της εικόνας στην περιοχή ενδιαφέροντος, και έτσι αναπτύχθηκαν τεχνικές ανίχνευσης με βάση διαφορετικά τοπικά γνωρίσματα (γωνίες, ακμές, κηλίδες). Τα τοπικά χαρακτηριστικά πλέον χρησιμοποιούνται ως μία σύντομη αναπαράσταση του σημασιολογικού περιεχομένου της εικόνας, που ακόμα και ένα υποσύνολό τους επιτρέπει την ικανοποιητική περιγραφή σκηνών και αντικειμένων. Η διαδικασία εντοπισμού σημείων ή περιοχών μπορεί πλέον να διαχειρίζεται κατάλληλα τις γεωμετρικές αλλαγές στην εικόνα, όπως είναι για παράδειγμα η αλλαγή κλίμακας (μεγέθυνση ή σμίκρυνση των αντικειμένων) ή οι αφινικές μεταβολές, και οι σύγχρονες μελέτες στηρίζονται κυρίως στο είδος και στο βαθμό της ανεξαρτησίας που απαιτεί το κάθε πρόβλημα. Τα τελευταία χρόνια λοιπόν ο επιστημονικός κλάδος της ανάλυσης εικόνας έχει οδηγηθεί στη συστηματικότερη μελέτη των μεθόδων αυτών, καθώς τα τοπικά χαρακτηριστικά και οι περιγραφείς που εξάγονται από αυτά αποτελούν ένα σημαντικό στάδιο κατά την επίλυση προβλημάτων όπως είναι η αναζήτηση εικόνων (image retrieval) ή η κατηγοριοποίηση αντικειμένων (object class recognition).

Στην παρούσα εργασία αρχικά αναλύονται οι πιο γνωστές μέθοδοι τοπικών χαρακτηριστικών, ως προς το θεωρητικό τους υπόβαθρο αλλά και το πρακτικό τους αποτέλεσμα, ενώ στη συνέχεια επιχειρείται η πειραματική σύγκριση των επιδόσεών τους με βάση αντικειμενικά κριτήρια που έχουν δημοσιευθεί σε σχετικές εργασίες. Με τη βοήθεια ενός συνόλου από εικόνες συγκεκριμένων σκηνών που υπόκεινται σε διάφορους μετασχηματισμούς, υπολογίζονται οι επιδόσεις των μεθόδων ως προς την επαναληψιμότητα των σημείων τους, την ακρίβεια εντοπισμού τους και την ικανότητα ταιριάσματος των περιγραφέων τους. Τα αποτελέσματα της σύγκρισης αυτής είναι περισσότερο ποιοτικά, καθώς στόχος της παραπάνω μελέτης είναι η εξέταση της καταλληλότητας των μεθόδων ανάλογα με το περιεχόμενο της εικόνας, το είδος του μετασχηματισμού που έχουν υποστεί και τη μορφή του προβλήματος που πρέπει να επιλυθεί.

Αμέσως μετά παρουσιάζεται μια εφαρμογή της ανίχνευσης τοπικών χαρακτηριστικών σε ένα σύστημα αναζήτησης εικόνων σε διαφορετικά σύνολα φωτογραφιών (μεγάλες βάσεις δεδομένων), απ' όπου προκύπτει μία ποσοτική σύγκριση των τεχνικών ως προς τα ποσοστά επιτυχίας του συστήματος. Στο σύστημα αυτό, με μηχανισμό ανάλογο της αναζήτησης κειμένου, δημιουργείται ένα οπτικό λεξικό ομαδοποιώντας τα σημεία των περιγραφέων και στη συνέχεια, μέσω των κέντρων των ομάδων που αποτελούν τις οπτικές λέξεις, κατασκευάζεται ένα διάνυσμα αναπαράστασης της εικόνας, με στοιχεία τη συχνότητα εμφάνισης για κάθε λέξη. Η αναζήτηση στη βάση παρόμοιων εικόνων με ένα δείγμα εικόνας του χρήστη γίνεται με την εκτίμηση της ομοιότητας των διανυσμάτων αναπαράστασης. Τα οπτικά λεξικά που έχουν προκύψει από τις διαφορετικές μεθόδους τοπικών χαρακτηριστικών συγκρίνονται με βάση τα μέτρα αξιολόγησης της αναζήτησης εικόνων, όπως είναι τα μέτρα ακριβείας (precision), ανάκτησης (recall) και μέσης ακριβείας (mean Average Precision - mAP).

Βασικός σκοπός λοιπόν της εργασίας είναι η επισκόπηση των μεθόδων που εμφανίζονται στη βιβλιογραφία και η ανάδειξη των πλεονεκτημάτων των τοπικών χαρακτηριστικών, μέσω της διεξαγωγής πειραματικών συγκρίσεων, τόσο ποιοτικών όσο και ποσοτικών, και συγκεκριμένα της

αξιολόγησης των επιδόσεών τους στο ταίριασμα εικόνων διαφορετικών όψεων της ίδιας σκηνής και στην αναζήτηση εικόνων παρόμοιου περιεχομένων από μεγάλες βάσεις δεδομένων.

## 1.4 Οργάνωση της αναφοράς

Στο Κεφάλαιο 2 παρατίθενται αρχικά οι ορισμοί και οι βασικές έννοιες των τοπικών χαρακτηριστικών. Στη συνέχεια αναλύονται οι θεμελιώδεις αρχές διαφόρων μεθόδων ανίχνευσης και περιγραφής τοπικών σημείων ή περιοχών, οι οποίες χρησιμοποιούνται ευρέως στη βιβλιογραφία. Πιο συγκεκριμένα μελετάται η ανίχνευση γωνιών του Harris, οι παραλλαγές αυτής προσαρμοσμένες σε αλλαγές κλίμακας ή αφινικές μεταβολές (Harris Laplace/Affine), η ανίχνευση κηλίδων Hessian μέσω του χεσσιανού πίνακα και οι παραλλαγές της (Hessian Laplace/Affine), ο ανιχνευτής περιοχών MSER και οι ανιχνευτές σημείων SIFT και SURF με ανεξαρτησία από αλλαγές στην κλίμακα, καθώς και οι αντίστοιχοι τοπικοί περιγραφείς όπως παρουσιάζονται στις ίδιες δημοσιεύσεις (οι τελευταίες μέθοδοι περιλαμβάνουν υπολογιστικά αποδοτικές υλοποιήσεις).

Στο Κεφάλαιο 3 περιλαμβάνονται τα πειράματα και οι συγκρίσεις των μεθόδων τοπικών χαρακτηριστικών, αφού πρώτα δοθεί η αναλυτική περιγραφή του συνόλου των πειραματικών εικόνων και των μέτρων επίδοσης. Παρατίθενται τα διαγράμματα ξεχωριστά για κάθε μέθοδο και για κάθε είδος μετασχηματισμού της εικόνας και επίσης η αναλυτική αξιολόγηση των αποτελεσμάτων.

Στο Κεφάλαιο 4 παρουσιάζεται το σύστημα αναζήτησης εικόνων και η επίδραση των προηγούμενων τεχνικών στα αποτελέσματά του. Περιγράφονται αναλυτικά τα δομικά στοιχεία και ο τρόπος σχεδίασης του συστήματος, καθώς και τα κριτήρια επίδοσης στην ανάκτηση εικόνων, όπως χρησιμοποιούνται στη βιβλιογραφία, και στη συνέχεια εφαρμόζονται τρεις διαφορετικές μέθοδοι τοπικών χαρακτηριστικών για την ανάκτηση εικόνων, με σκοπό την αξιολόγηση της επίδοσής τους σε πραγματικά προβλήματα αναζήτησης σε τρεις γνωστές συλλογές εικόνων.

Στο Κεφάλαιο 5 παρουσιάζονται τα συνολικά συμπεράσματα που προκύπτουν από τη μελέτη των μεθόδων τοπικών χαρακτηριστικών και την εφαρμογή τους στην αναζήτηση εικόνων, καθώς και προτάσεις εναλλακτικής χρήσης τους ή βελτίωσης των επιδόσεών τους.

Τέλος, στο Παράρτημα υπάρχει μία σύντομη περιγραφή του συστήματος τοπικών χαρακτηριστικών που χρησιμοποιήθηκε για τα πειράματα και την αξιολόγησή τους, δηλαδή περιγράφεται το περιβάλλον και οι βιβλιοθήκες προγραμμάτων που χρησιμοποιήθηκαν, η δομή της υλοποίησης και ο τρόπος χρήσης του συστήματος και της εφαρμογής (application interface).

# ΚΕΦΑΛΑΙΟ 2

# Μελέτη Μεθόδων Τοπικών Χαρακτηριστικών

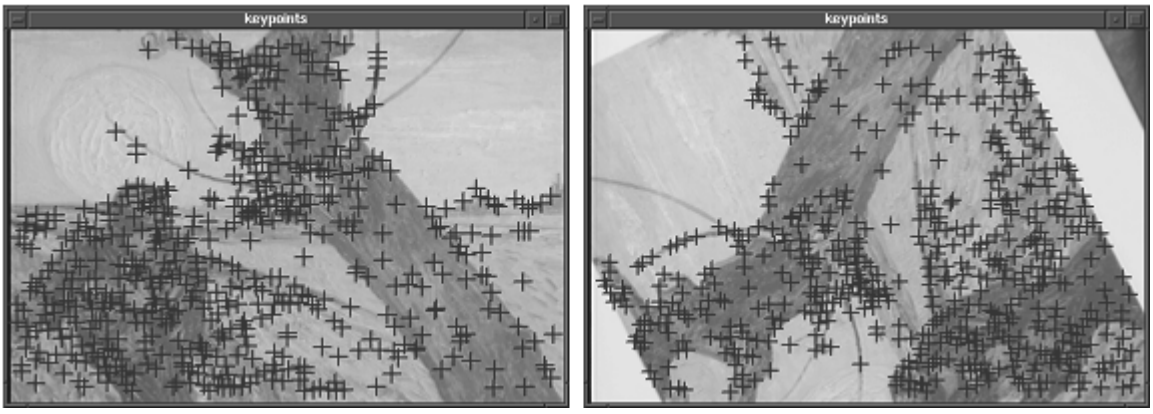
## 2.1 Εισαγωγικές έννοιες

Τα τοπικά χαρακτηριστικά είναι σημεία ή μικρές περιοχές της εικόνας που διαφέρουν από τη γύρω γειτονιά τους και συνήθως ονομάζονται σημεία ή περιοχές ενδιαφέροντος (interest points – regions). Σχετίζονται με αλλαγές σε κάποια ή κάποιες ιδιότητες της εικόνας, όπως είναι η φωτεινότητα (ένταση), το χρώμα, η υφή, οι ακμές. Για να αναπαρασταθεί το περιεχόμενο κοντά στο σημείο ενδιαφέροντος, συνήθως από μια μικρή περιοχή γύρω από αυτό εξάγεται ένα διάνυσμα χαρακτηριστικών το οποίο ονομάζεται τοπικός περιγραφέας. Το σύνολο των τοπικών χαρακτηριστικών μαζί με τους περιγραφείς τους αποτελούν μια αναπαράσταση του περιεχομένου της εικόνας, χωρίς να πρέπει να αναλυθεί το τι ακριβώς υπάρχει σε κάθε περιοχή ενδιαφέροντος (διαδικασία που απαιτεί υψηλού επιπέδου επεξεργασία). Ένα παράδειγμα ανίχνευσης τοπικών χαρακτηριστικών φαίνεται στις δύο εικόνες του σχήματος 2.1.

Για τη μέθοδο που εντοπίζει τα τοπικά χαρακτηριστικά, συνήθως χρησιμοποιείται ο όρος τοπικός ανιχνευτής (detector), αν και ο όρος εξαγωγέας (extractor) ταιριάζει καλύτερα στη διαδικασία που λαμβάνει χώρα. Για τη μέθοδο της κατασκευής του διανύσματος χαρακτηριστικών χρησιμοποιείται ο όρος τοπικός περιγραφέας. Όταν η ανίχνευση τοπικών χαρακτηριστικών οδηγεί σε σημεία, τότε μιλάμε για σημεία ενδιαφέροντος, ενώ αν οδηγεί σε υποπεριοχές της εικόνας, τότε πρόκειται για περιοχές ενδιαφέροντος. Το διάνυσμα χαρακτηριστικών στην πρώτη περίπτωση εξάγεται από μία κατάλληλη τοπική γειτονιά κεντραρισμένη στο σημείο ενδιαφέροντος (local patch), ενώ στη δεύτερη περίπτωση εξάγεται είτε από την ίδια την περιοχή ενδιαφέροντος (detected – distinguished region) είτε από μία μεγαλύτερη περιοχή που αποτελεί μία μεγέθυνση της αρχικής (measurement region).

Μία άλλη πολύ σημαντική έννοια είναι η ανεξαρτησία (invariance) ή σταθερότητα του τοπικού χαρακτηριστικού ως προς κάποιες μεταβολές της εικόνας. Για παράδειγμα όταν έχουμε μετατόπιση ενός αντικειμένου μέσα στην εικόνα, τα σημεία που εντοπίστηκαν στην αρχική εικόνα περιμένουμε να εντοπιστούν και στη νέα, και ιδανικά ακριβώς στην ίδια θέση, αφού η δομή εντός του αντικειμένου δεν μεταβλήθηκε. Το ίδιο ισχύει και σε περίπτωση περιστροφής του αντικειμένου ή ολόκληρης της εικόνας. Ένα τέτοιο παράδειγμα όπου φαίνεται η περιστροφή της εικόνας και ο εντοπισμός σημείων σε αντίστοιχες θέσεις βρίσκεται στο σχήμα 2.1. Σύμφωνα με τα παραπάνω, ταιριάζει καλύτερα ο όρος συμμεταβλητότητα (covariance) των τοπικών χαρακτηριστικών σε συνάρτηση με τις μεταβολές της εικόνας, αφού όπως φαίνεται ο εντοπισμός

τους δεν γίνεται στην ίδια θέση, αλλά σε νέα τοποθεσία που εξαρτάται από την συγκεκριμένη μεταβολή (μετατόπιση, περιστροφή κτλ.). Επειδή όμως έχει επικρατήσει ο όρος ανεξαρτησία ή σταθερότητα (invariance), από εδώ και στο εξής θα μιλάμε για μεθόδους ανίχνευσης ανεξάρτητες ή αμετάβλητες από μετασχηματισμούς στην εικόνα. Από τη σκοπιά του τοπικού περιγραφέα, πρέπει να τονίσουμε ότι μετά από ένα στάδιο κανονικοποίησης της περιοχής ενδιαφέροντος, το διάνυσμα των χαρακτηριστικών εξάγεται με τρόπο ανεξάρτητο κάθε μετασχηματισμού, και αυτό θα γίνει περισσότερο κατανοητό παρακάτω όπου περιγράφεται η κατασκευή των περιγραφέων.




**Σχ. 2.1:** Σημεία ενδιαφέροντος σε δύο εικόνες που σχετίζονται με μία περιστροφή ([Schmid et al., 1997])

Μία από τις ιδιότητες που επιδιώκεται να έχουν τα τοπικά χαρακτηριστικά είναι η ανεξαρτησία από γεωμετρικούς κυρίως μετασχηματισμούς. Πρέπει επίσης να μένουν ανεπηρέαστα και από άλλες παραμορφώσεις της εικόνας, όπως είναι οι αλλαγές στη φωτεινότητα (photometric changes), το θόλωμα (blurring), ο θόρυβος (noise), η μερική επικάλυψη αντικειμένων (partial visibility, occlusion) και γενικά η σύγχυση του περιεχομένου (cluttered image). Ένας καλός ανιχνευτής θα πρέπει να έχει υψηλή επαναληψιμότητα, δηλαδή μεγάλο ποσοστό των τοπικών χαρακτηριστικών να εντοπίζεται και στις δύο εικόνες στο κοινό τους κομμάτι, και μάλιστα σε όσο το δυνατόν αντίστοιχη θέση, γεγονός που καθορίζει την ακρίβεια του ανιχνευτή.

Γενικότερα, τα σημεία και οι περιοχές ενδιαφέροντος θα πρέπει να χαρακτηρίζονται από τοπικότητα, δηλαδή τα χαρακτηριστικά να αντιστοιχούν σε μικρές περιοχές της εικόνας οπότε να μπορεί να αντιμετωπιστεί ένα πρόβλημα επικάλυψης μέρους ενός αντικειμένου ή ένα πρόβλημα διαφορετικών γωνιών λήψης της ίδιας σκηνής. Επίσης, είναι πολύ βασικό πλεονέκτημα το κάθε τοπικό χαρακτηριστικό να προσφέρει τη δυνατότητα να διακρίνεται εύκολα η συγκεκριμένη περιοχή από τις υπόλοιπες και να μπορεί αποτελεσματικά να ταιριάξει με την αντίστοιχη σε μια άλλη όψη της εικόνας. Όσο περισσότερα τα χαρακτηριστικά, τόσο περισσότερο καλά μπορεί να περιγραφεί το περιεχόμενο της εικόνας, ακόμα και για μικρά αντικείμενα. Βέβαια, το κατάλληλο πλήθος των χαρακτηριστικών εξαρτάται από το περιεχόμενο της εικόνας αλλά και από τις απαιτήσεις του εκάστοτε προβλήματος. Πρέπει να σημειώσουμε ότι η αξία μιας μεθόδου τοπικού ανιχνευτή κρίνεται και από την ταχύτητά της, ειδικά σε εφαρμογές όπου απαιτείται η εξαγωγή χαρακτηριστικών ως πρώτο βήμα για να γίνουν στη συνέχεια πολλές άλλες διεργασίες, όπως ταίριασμα εικόνων ή εντοπισμός αντικειμένων.

## 2.2 Ανίχνευση τοπικών χαρακτηριστικών

### 2.2.1 Εισαγωγικά

Στο σημείο αυτό του κεφαλαίου γίνεται μια σύντομη μελέτη των βασικών αρχών διαφόρων μεθόδων τοπικών ανιχνευτών. Ανάλογα με το είδος των χαρακτηριστικών που παράγουν, οι μέθοδοι χωρίζονται συνήθως σε μεθόδους σημείων ενδιαφέροντος και μεθόδους περιοχών ενδιαφέροντος. Η πρώτη κατηγορία αφορά είτε σημεία γωνιών είτε σημεία κηλίδων, ενώ η δεύτερη κατηγορία αφορά γενικότερα περιοχές της εικόνας, οι οποίες σε κάποιες περιπτώσεις μπορούν να θεωρηθούν γειτονιές γύρω από κηλίδες. Στις επόμενες υποπαραγράφους παρουσιάζονται διαδοχικά οι μέθοδοι ανεξαρτήτως κατηγορίας, με την παρατήρηση στο τέλος καθεμίας σχετικά με τη μορφή των τοπικών τους χαρακτηριστικών τους. Οι συγκεκριμένες μέθοδοι επιλέχθηκαν από τη βιβλιογραφία για δύο βασικούς λόγους: έχουν παρουσιάσει υψηλές επιδόσεις σε συγκριτικά πειράματα [Mikolajczyk et al., 2005b] και έχουν δημοσιευθεί αποδοτικές υλοποιήσεις τους [Lowe, 2004], [Bay et al., 2006], γεγονός που τις καθιστά πολύτιμες για συστήματα ανάκτησης εικόνων από μεγάλες βάσεις δεδομένων.

### 2.2.2 Ανιχνευτής Harris Laplace

Η μέθοδος γωνιών Harris δημοσιεύθηκε το 1988 στην πολύ γνωστή εργασία του [Harris et al., 1988], η οποία και έπαιξε σπουδαίο ρόλο στην εξέλιξη των μεθόδων τοπικών χαρακτηριστικών. Βασίζεται στον υπολογισμό του πίνακα ροπών δεύτερης τάξης M, ο οποίος μπορεί να θεωρηθεί ότι είναι ο πίνακας αυτοσυσχέτισης του σήματος (έντασης) της εικόνας σε ένα μικρό τοπικό παράθυρο. Ο πίνακας αυτός περιγράφει την τοπική δομή της εικόνας, δηλαδή τον τρόπο κατανομής των πρώτων παραγώγων στη γειτονιά ενός σημείου, και για ένα τυχαίο διδιάστατο σημείο z=(x,y) της εικόνας ορίζεται με την εξίσωση 2.1 παρακάτω.

$$M = \begin{bmatrix} \mu_{11} & \mu_{12} \\ \mu_{21} & \mu_{22} \end{bmatrix} = \sigma_D^2 \cdot g(\sigma_I) * \begin{bmatrix} I_x^2(z,\sigma_D) & I_x(z,\sigma_D)\cdot I_y(z,\sigma_D) \\ I_x(z,\sigma_D)\cdot I_y(z,\sigma_D) & I_y^2(z,\sigma_D) \end{bmatrix} \quad \textbf{(εξ. 2.1)}$$

όπου οι πρώτες παράγωγοι της έντασης δίνονται από τις συνελίξεις:

$$I_x(z,\sigma_D) = \frac{\partial}{\partial x} g(\sigma_D) * I(z) \quad \textbf{(εξ. 2.2α)}$$

$$I_y(z,\sigma_D) = \frac{\partial}{\partial y} g(\sigma_D) * I(z) \quad \textbf{(εξ. 2.2β)}$$

και g η ισοτροπική γκαουσιανή συνάρτηση δύο μεταβλητών:

$$g(\sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \quad \textbf{(εξ. 2.3)}$$

Με τον τρόπο αυτό επιτυγχάνεται ταυτόχρονα και η παραγώγιση του σήματος I ως προς τις δύο κατευθύνσεις μέσω της συνέλιξης με την γκαουσιανή με κλίμακα παραγώγισης $\sigma_D$ (differentiation scale) και η ολοκλήρωση του πίνακα σε ένα παράθυρο γύρω από το σημείο z με γκαουσιανή συνέλιξη (ομαλοποίηση) κλίμακας ολοκλήρωσης $\sigma_I$ (integration scale), που είναι ουσιαστικά ένας ομαλοποιημένος μέσος όρος των τιμών του πίνακα για μικρές μετατοπίσεις του z και βοηθάει στην αντιμετώπιση προβλημάτων θορύβου, που προκύπτουν αν απλά προσθέσουμε τις τιμές σε ένα τετραγωνικό παράθυρο (μέσος όρος).

Οι ιδιοτιμές του πίνακα M είναι οι κύριες καμπυλότητες στη γειτονιά του σημείου z, δηλαδή αντιστοιχούν στις μεταβολές του σήματος σε δύο κάθετες κατευθύνσεις. Έτσι, ένα υποψήφιο σημείο αποτελεί γωνιακό σημείο, αν οι τιμές των ιδιοτιμών του M είναι και οι δύο πολύ μεγάλες, δηλαδή το σήμα μεταβάλλεται ταυτόχρονα και στις δύο κατευθύνσεις. Με διαφορετική προσέγγιση οι Harris et al. προτείνουν τη χρήση ενός μέτρου "γωνιότητας" (cornerness), το οποίο συνδυάζει τις ιδιοτιμές σε μία λιγότερο πολύπλοκη υπολογιστικά σχέση. Αν οι ιδιοτιμές του M είναι $\lambda_1$ και $\lambda_2$, τότε η ορίζουσα (det) και το ίχνος (tr) του πίνακα είναι αντίστοιχα:

$$\det(M) = \lambda_1 \cdot \lambda_2 \qquad tr(M) = \lambda_1 + \lambda_2 \qquad \textbf{(εξ. 2.4)}$$

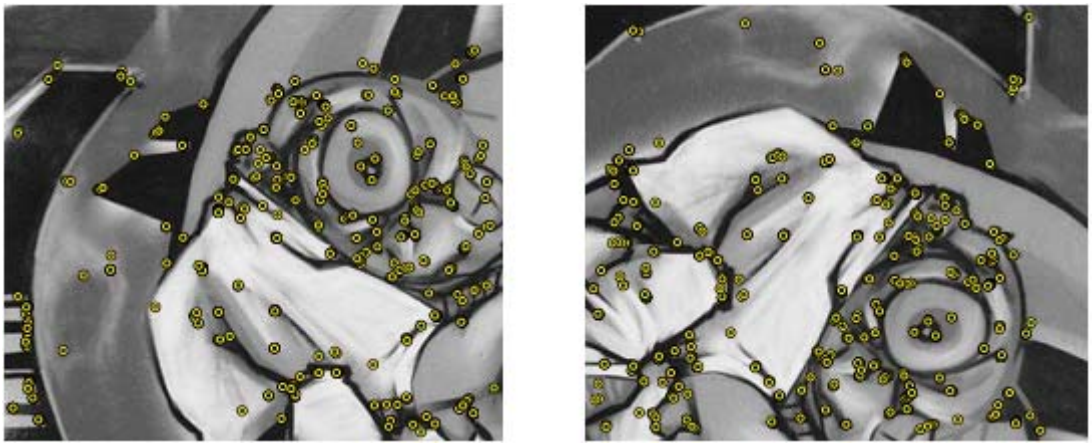
και το μέτρο cornerness δίνεται εναλλακτικά από τις σχέσεις:

$$c = \det(M) - k \cdot tr(M)^2 \qquad \textbf{(εξ. 2.5α)}$$

$$c = (\lambda_1 \cdot \lambda_2) - k \cdot (\lambda_1 + \lambda_2)^2 \qquad \textbf{(εξ. 2.5β)}$$

Στις προηγούμενες σχέσεις k είναι μία σταθερά που υπολογίζεται πειραματικά και συνήθως χρησιμοποιείται η τιμή 0.04. Από την τελευταία σχέση φαίνεται ότι για υψηλές τιμές των ιδιοτιμών προκύπτει υψηλή τιμή του μέτρου γωνιότητας, οπότε για να εντοπιστεί ένα γωνιακό σημείο θα πρέπει να οριστεί ένα κατώφλι (threshold) έστω $t_1$, ώστε η συνθήκη για να εντοπιστεί ένα σημείο ως γωνιακό να είναι:

$$c > t_1 \qquad \textbf{(εξ. 2.6)}$$



**Σχ. 2.2:** Σημεία γωνιών Harris στην εικόνα graffiti για δύο διαφορετικές όψεις (περιστροφή)

Στο σχήμα 2.2 φαίνονται τα γωνιακά σημεία Harris, όπως προκύπτουν για δύο εικόνες του ίδιου αντικειμένου (graffiti) για δύο διαφορετικές όψεις, που συνδέονται με μία τυχαία

περιστροφή. Παρατηρώντας τις δύο εικόνες, γίνεται εύκολα αντιληπτό ότι μεγάλο ποσοστό των γωνιών εμφανίζεται σε αντίστοιχα σημεία, γεγονός που δείχνει την ανεξαρτησία της μεθόδου από την περιστροφή της εικόνας, καθώς και από κάθε μετατόπιση του περιεχομένου της εικόνας (όπως εύκολα μπορεί να προκύψει). Η μέθοδος παράγει αποτέλεσμα εκτός από γωνίες και σε άλλα σημεία που έχουν υψηλή καμπυλότητα, όπως είναι οι διασταυρώσεις σχήματος Τ. Επειδή οι παράγωγοι δεν επηρεάζονται από συνολικές αλλαγές της έντασης, τα σημεία Harris είναι αμετάβλητα υπό διάφορες συνθήκες φωτεινότητας. Σε ένα βασικό συγκριτικό πείραμα διάφορων τοπικών ανιχνευτών [Schmid et al., 2000] οι γωνίες Harris αποδείχτηκαν τα σημεία ενδιαφέροντος με το μεγαλύτερο ποσοστό επαναληψιμότητας και με μεγάλο ποσοστό πληροφορίας.

Παρά τα παραπάνω πλεονεκτήματα, η μέθοδος του Harris επηρεάζεται από τις αλλαγές στην κλίμακα. Μία λύση σε αυτό το πρόβλημα θα ήταν μία πολυκλιμακωτή ανάλυση (multiscale analysis), η οποία όμως δίνει σαν αποτέλεσμα πολλαπλά σημεία ενδιαφέροντος (σε πολλές συνεχόμενες κλίμακες) που αντιστοιχούν στην ίδια γωνία. Μία βελτίωση της τεχνικής προτείνεται στην εργασία [Mikolajczyk et al., 2001] όπου κατασκευάζεται η μέθοδος Harris-Laplace που είναι αμετάβλητη από αλλαγές κλίμακας. Αρχικά χρησιμοποιώντας έναν χώρο πολλών κλιμάκων (διαδοχικές συνελίξεις τη εικόνας με 2D γκαουσιανή) γίνεται εντοπισμός των γωνιών με το μέτρο του Harris (cornerness) σε κάθε επίπεδο και στη συνέχεια επιλέγονται εκείνα τα σημεία που ανήκουν σε κάποια χαρακτηριστική κλίμακα. Η ιδέα βασίζεται στην επιλογή κλίμακας (scale selection) όπως προτείνεται στη δημοσίευση [Lindeberg, 1998]. Θεωρώντας μία συγκεκριμένη συνάρτηση για κάθε σημείο που μεταβάλλεται συναρτήσει της κλίμακας και επιλέγοντας τα σημεία στα οποία παρουσιάζει τοπικό μέγιστο προκύπτει η χαρακτηριστική κλίμακα (characteristic scale) για αυτό το σημείο, δηλαδή η κλίμακα στην οποία "ταιριάζει" καλύτερα το τοπικό χαρακτηριστικό. Η συνάρτηση που χρησιμοποιείται είναι η λαπλασιανή, όπως προδίδει και το όνομα της μεθόδου, είναι το ίδιο με ένα φιλτράρισμα της εικόνας (συνέλιξη) με το φίλτρο LoG (laplacian of gaussian), και δίνεται παρακάτω:

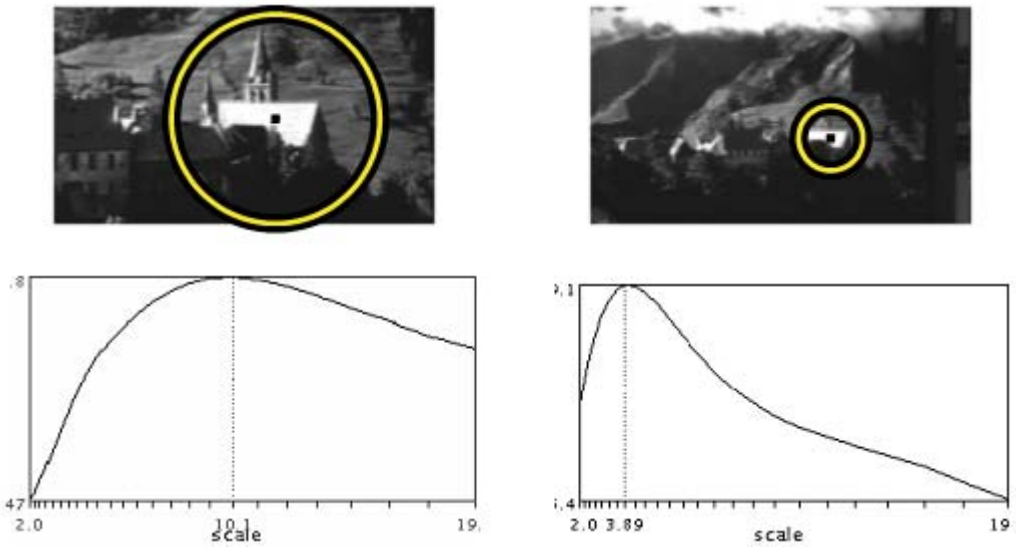
$$LoG = \sigma^2 \cdot \left[ I_{xx}(z,\sigma) + I_{yy}(z,\sigma) \right] \qquad \textbf{(εξ. 2.7)}$$

όπου:

$$I_{xx}(z,\sigma) = \frac{\partial}{\partial x^2} g(\sigma) * I(z) \qquad \textbf{(εξ. 2.8α)}$$

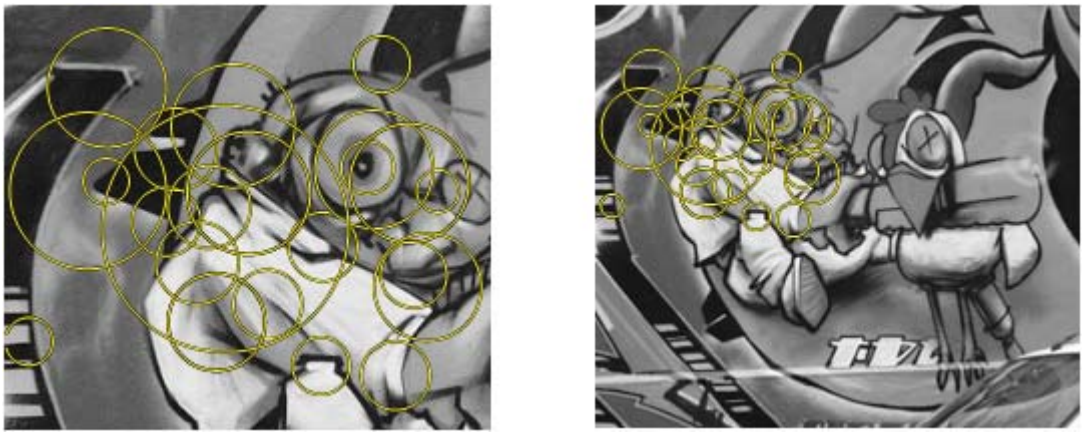
$$I_{yy}(z,\sigma) = \frac{\partial}{\partial y^2} g(\sigma) * I(z) \qquad \textbf{(εξ. 2.8β)}$$

Έτσι, ανεξάρτητα από το μέγεθος των αντικειμένων σε διαφορετικές εικόνες, η περιοχή γύρω από το τοπικό χαρακτηριστικό μένει αμετάβλητη και καθορίζεται από έναν κύκλο με ακτίνα ανάλογη της χαρακτηριστικής κλίμακας. Στο σχήμα 2.3 φαίνεται η χαρακτηριστική κλίμακα που εντοπίζεται με τη μέθοδο για δύο εικόνες με την ίδια σκηνή σε διαφορετική κλίμακα. Είναι εμφανές ότι εντός των δύο κύκλων περιλαμβάνεται το ίδιο κομμάτι της εικόνας και επίσης ο λόγος των δύο χαρακτηριστικών κλιμάκων είναι και ο λόγος της μεγέθυνσης της ανάλυσης μεταξύ των δύο εικόνων (περίπου ίσος με 2).




**Σχ. 2.3:** Χαρακτηριστική κλίμακα και διάγραμμα των τιμών της λαπλασιανής ([Mikolajczyk et al., 2001])

Στο σχήμα 2.4 παρουσιάζονται τα σημεία που ανιχνεύει η μέθοδος Harris-Laplace σε δύο διαφορετικές όψεις της εικόνας graffiti οι οποίες συνδέονται με μια συγκεκριμένη αλλαγή στην κλίμακα. Όπως μπορούμε να παρατηρήσουμε σε κάθε γωνία της εικόνας εντοπίζεται ένα μοναδικό σημείο ενδιαφέροντος, το οποίο αντιστοιχεί στην χαρακτηριστική κλίμακα και μεταξύ των δύο όψεων εντοπίζεται η ίδια περιοχή. Το τελευταίο αυτό γεγονός είναι πολύ σημαντικό, καθώς από αυτήν την περιοχή θα κατασκευαστεί αργότερα το διάνυσμα χαρακτηριστικών.



**Σχ. 2.4:** Σημεία γωνιών Harris-Laplace στην εικόνα graffiti για δύο διαφορετικές όψεις (αλλαγή κλίμακας)

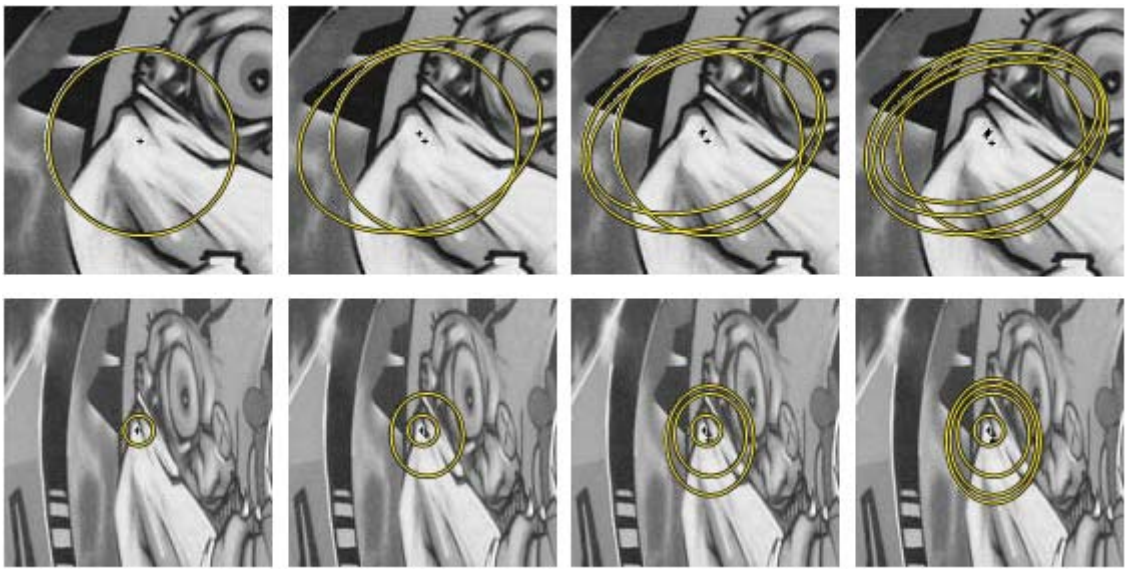
### 2.2.3 Ανιχνευτής Harris Affine

Οι ίδιοι συγγραφείς Mikolajczyk και Schmid σε επόμενη εργασία τους [Mikolajczyk et al., 2002] προτείνουν μία άλλη παραλλαγή της μεθόδου του Harris, η οποία είναι ανεξάρτητη από αφινικές μεταβολές. Ξεκινώντας από τις γωνίες Harris της πολυκλιμακωτής ανάλυσης και με μία κατάλληλη επαναληπτική διαδικασία, παράγουν σημεία μαζί με μία ελλειπτική γειτονιά, τα οποία

είναι ανεξάρτητα από αφινικές μεταβολές. Συνήθως τα σημεία σε πολλές κλίμακες που αντιστοιχούν στην ίδια γωνία θα πρέπει να συγκλίνουν στο ίδιο αφινικό σημείο, αλλά πολλές φορές αυτό δεν είναι εφικτό με αποτέλεσμα να προκύπτουν πολλαπλά τελικά σημεία. Σε μία άλλη διαφορετική εκδοχή της μεθόδου, ως αρχικά σημεία λαμβάνονται σημεία ανεξάρτητα της κλίμακας, οπότε δεν υπάρχει το πρόβλημα των πολλαπλών τελικών σημείων [Mikolajczyk et al., 2004]. Η κεντρική ιδέα είναι ο αρχικός εντοπισμός των γωνιών στην χαρακτηριστική τους κλίμακα με τη μέθοδο Harris-Laplace και στη συνέχεια η εφαρμογή του επαναληπτικού αλγορίθμου που υπολογίζει αφινικές περιοχές, όπως προτείνεται στη δημοσίευση [Lindeberg et al., 1997]. Ο υπολογισμός αυτός γίνεται με τον πίνακα ροπών δεύτερης τάξης, ο οποίος περιγράφει την τοπική δομή της εικόνας, και έτσι η κυκλική περιοχή γύρω από το σημείο μετασχηματίζεται σε μία ελλειπτική περιοχή που μεταβάλλεται ανάλογα με τον αφινικό μετασχηματισμό της εικόνας.

Τα βασικά βήματα του επαναληπτικού αλγορίθμου είναι τα εξής:

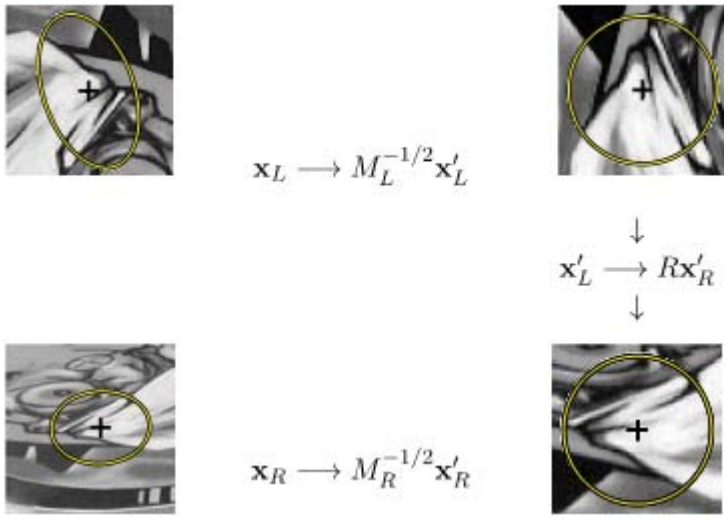
1. Εντοπισμός των αρχικών κυκλικών περιοχών Harris-Laplace.
2. Υπολογισμός του πίνακα ροπών δεύτερης τάξης και της αφινικής δομής της περιοχής.
3. Κανονικοποίηση της ελλειπτικής περιοχής σε κύκλο.
4. Επαναπροσδιορισμός θέσης και κλίμακας του σημείου στη νέα περιοχή.
5. Αν οι ιδιοτιμές του πίνακα ροπών στο νέο σημείο δεν είναι ίσες, τότε πήγαινε στο βήμα 2.



**Σχ. 2.5:** Επαναλήψεις του αλγορίθμου αφινικής προσαρμογής στην εικόνα graffiti για δύο διαφορετικές οπτικές γωνίες

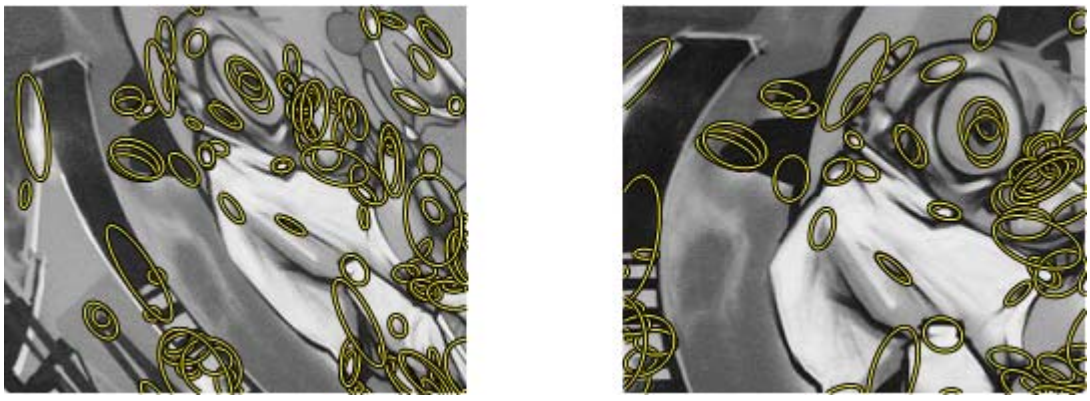
Σημειώνουμε εδώ ότι ο πίνακας ροπών δεύτερης τάξης υπολογίζεται στην περιοχή που καθορίζεται από την χαρακτηριστική κλίμακα s, δηλαδή χρησιμοποιείται γκαουσιανή με παράμετρο σίγμα ίση με s για τη συνέλιξη με την ένταση. Οι ιδιοτιμές του πίνακα ροπών δεύτερης τάξης δείχνουν την τοπική δομή γύρω από το σημείο, δηλαδή αναπαριστούν το αφινικό της σχήμα (έλλειψη). Οι τιμές τους αντιστοιχούν στους άξονες της έλλειψης και γι' αυτό κατά την εξέλιξη του αλγορίθμου πλησιάζουν μεταξύ τους, καθώς η περιοχή προσεγγίζει έναν κύκλο. Η συνθήκη οι ιδιοτιμές να είναι (σχεδόν) ίσες χρησιμοποιείται ως κριτήριο τερματισμού (ώστε η δομή της τελικής περιοχής να αντιστοιχεί σε κύκλο). Οι επαναλήψεις του αλγορίθμου και τα στάδια εξέλιξης της ελλειπτικής περιοχής φαίνονται στο σχήμα 2.5, σε δύο διαφορετικές οπτικές γωνίες της ίδιας σκηνής.

Εφόσον οι ιδιοτιμές του πίνακα ροπών δεύτερης τάξης προσδιορίζουν το αφινικό σχήμα της περιοχής, μπορούμε να καθορίσουμε τον μετασχηματισμό της γειτονιάς του σημείου από ελλειπτική σε κυκλική, δηλαδή σε μία περιοχή της οποίας οι ιδιοτιμές είναι ίσες. Με αυτόν τον τρόπο κάθε σημείο, ακόμα και αν ανήκει σε περιοχές που διαφέρουν κατά μία αφινική μεταβολή, μπορεί να μετασχηματιστεί σε μία κανονικοποιημένη περιοχή (βλέπε σχήμα 2.6). Ως πίνακας του μετασχηματισμού αυτού λαμβάνεται η τετραγωνική ρίζα του πίνακα ροπών δεύτερης τάξης που υπολογίστηκε στο τελευταίο βήμα της επαναληπτικής μεθόδου. Όπως φαίνεται στο σχήμα 2.6, τα δύο σημεία μετασχηματίζονται σε δύο κανονικοποιημένες περιοχές οι οποίες διαφέρουν μόνο κατά μία περιστροφή (πίνακας περιστροφής R), η οποία μπορεί εύκολα να υπολογιστεί από την κατανομή των κατευθύνσεων των παραγώγων (ιστόγραμμα παραγώγων) και γίνεται στο στάδιο της εξαγωγής του τοπικού περιγραφέα.




**Σχ. 2.6:** Αφινικός μετασχηματισμός των περιοχών για τις δύο όψεις στην εικόνα graffiti.
Οι τελικές περιοχές διαφέρουν μόνο ως προς μία περιστροφή R ([Mikolajczyk et al., 2005b])

Έτσι με την παραπάνω επαναληπτική διαδικασία προσαρμογής του αφινικού σχήματος της περιοχής προκύπτουν σημεία ενδιαφέροντος που δεν επηρεάζονται από αφινικούς μετασχηματισμούς της εικόνας, ούτε φυσικά από μετατόπιση, περιστροφή ή αλλαγή κλίμακας. Η μέθοδος ονομάζεται Harris-Affine και ένα παράδειγμα των σημείων που παράγει φαίνεται στο σχήμα 2.7.



**Σχ. 2.7:** Σημεία γωνιών Harris-Affine στην εικόνα graffiti για δύο διαφορετικές οπτικές γωνίες

### 2.2.4 Ανιχνευτές Hessian Laplace/Affine

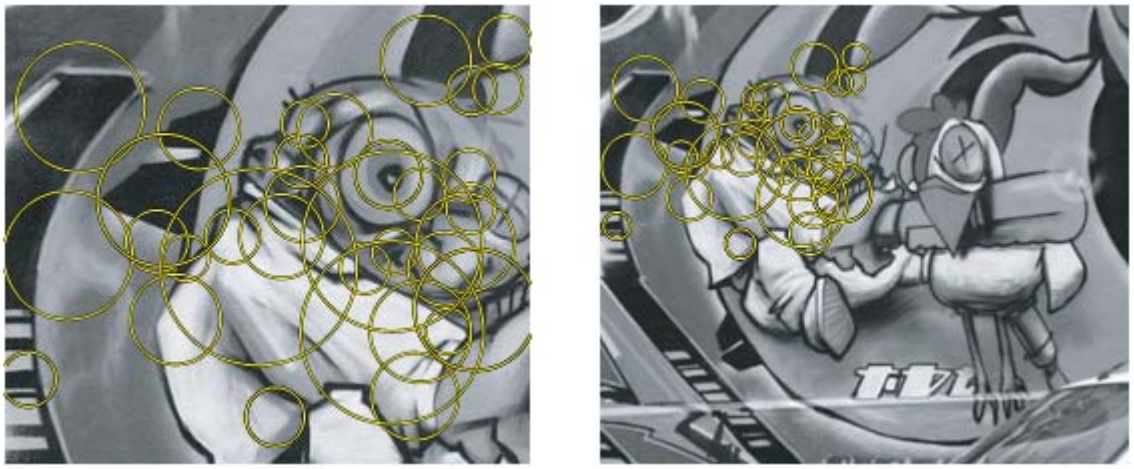
Η μέθοδος Hessian ανίχνευσης σημείων ενδιαφέροντος, όπως λέει και το όνομά της, βασίζεται στον χεσσιανό πίνακα (ο οποίος προέρχεται από το ανάπτυγμα Taylor της έντασης της εικόνας):

$$H = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} = \begin{bmatrix} I_{xx}(z,\sigma_D) & I_{xy}(z,\sigma_D) \\ I_{xy}(z,\sigma_D) & I_{yy}(z,\sigma_D) \end{bmatrix} \qquad \textbf{(εξ. 2.9)}$$

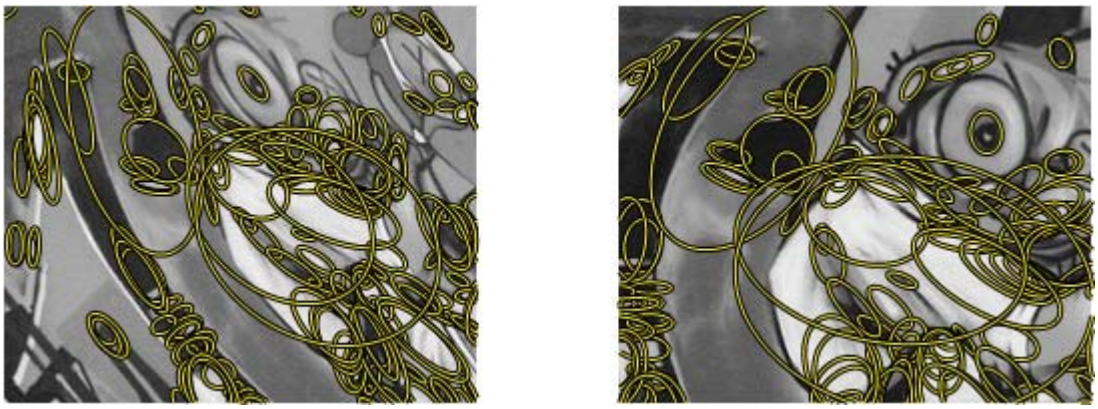
όπου οι ομαλοποιημένες παράγωγοι δίνονται από τις εξισώσεις 2.8 α και β. Ο πίνακας αυτός περιέχει πληροφορίες για την τοπική δομή της εικόνας, δείχνοντας το πώς αλλάζει η κάθετος σε μια ισοδυναμική επιφάνεια. Πιο συγκεκριμένα, χρησιμοποιείται συχνά το ίχνος ή η ορίζουσα του πίνακα Η για να εντοπιστούν σημαντικές αλλαγές στο σχήμα της εικόνας. Τα σημεία ενδιαφέροντος μπορούν να εντοπιστούν ως τοπικά μέγιστα της ορίζουσας του Η και είναι περιοχές κυρίως με δομή κηλίδας. Το κριτήριο ανίχνευσης σημείου-κηλίδας είναι προφανώς μια σχέση της μορφής (όπου το $t_2$ είναι ένα κατάλληλο κατώφλι):

$$\det(H) > t_2 \qquad \textbf{(εξ. 2.10)}$$

Προκειμένου να προκύψουν σημεία ανεξάρτητα των αλλαγών κλίμακας και των αφινικών μετασχηματισμών, ακολουθώντας την ίδια διαδικασία με τα αντίστοιχα σημεία γωνιών Harris (επιλογή χαρακτηριστικής κλίμακας και επαναληπτικός αλγόριθμος αφινικής προσαρμογής), κατασκευάζονται οι δύο μέθοδοι Hessian-Laplace και Hessian-Affine αντίστοιχα, με τη διαφορά ότι ως αρχικά σημεία θεωρούνται τα τοπικά μέγιστα της ορίζουσας του Η. Παραδείγματα σημείων ενδιαφέροντος των δύο αυτών μεθόδων παρουσιάζονται στα σχήματα 2.8 και 2.9. Συνήθως είναι τα σημεία της μεθόδου Hessian είναι συμπληρωματικά των γωνιών, αφού αντιστοιχούν σε διαφορετικές δομές της εικόνας.



**Σχ. 2.8:** Σημεία γωνιών Hessian-Laplace στην εικόνα graffiti για διαφορετικές όψεις (αλλαγή κλίμακας)



**Σχ. 2.9:** Σημεία γωνιών Hessian-Affine στην εικόνα graffiti για διαφορετικές οπτικές γωνίες

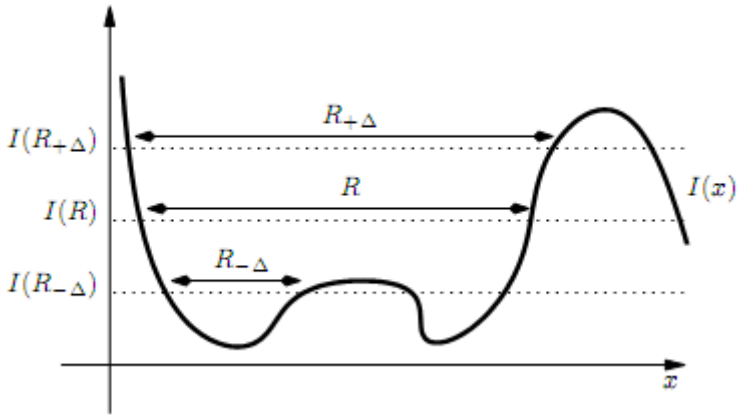
### 2.2.5 Ανιχνευτής MSER

Η μέθοδος ανίχνευσης MSER (Maximally Stable Extremal Regions, πολύ σταθερές ακραίες περιοχές) προτάθηκε στην εργασία [Matas et al., 2002] και από τότε έχει γίνει πολύ δημοφιλής, κυρίως λόγω της αποτελεσματικότητάς της. Η διαδικασία βασίζεται στη διαδοχική κατωφλίωση της εικόνας και σε κάθε βήμα της μεθόδου μία περιοχή MSER είναι μία κατάλληλη συνεκτική συνιστώσα (connected component) της εικόνας. Ο όρος "ακραία" (extremal) αναφέρεται σε υποπεριοχές της εικόνας στις οποίες όλα τα εσωτερικά pixels έχουν ένταση μικρότερη ή μεγαλύτερη από τα συνοριακά pixels. Η περιοχή τότε ονομάζεται μέγιστη ή ελάχιστη αντίστοιχα (maximal or minimal). Ο όρος "πολύ σταθερή" (maximally stable) αναφέρεται στην ιδιότητα μιας ακραίας περιοχής να μένει σταθερή κατά τη διαδικασία αύξησης του κατωφλίου.

Πρώτα θα δώσουμε τις βασικές αρχές του αλγορίθμου για την περίπτωση των μέγιστων περιοχών (maximal), αφού η αντίστοιχη διαδικασία για τις ελάχιστες περιοχές (minimal) είναι παρόμοια, με τη μόνη διαφορά ότι αντιστρέφεται η ένταση σε ολόκληρη την εικόνα πριν την εφαρμογή του αλγορίθμου. Υποθέτουμε ότι οι τιμές τις έντασης είναι ακέραιες 0-255. Κατασκευάζουμε τα επίπεδα της εικόνας τα οποία δημιουργούνται με μία κατωφλίωσή της διαδοχικά με κατώφλι t = 0, 1, 2, …, 255. Σε κάθε επίπεδο τα pixels τις εικόνας που έχουν τιμή έντασης μικρότερη από το κατώφλι θεωρείται ότι ανήκουν σε μία περιοχή. Αυτό σημαίνει ότι ο αλγόριθμος αρχικά θα ξεκινήσει από τα τοπικά ελάχιστα της εικόνας με τιμή έντασης ίση με 0 (για t = 0). Στη συνέχεια και καθώς αυξάνεται το κατώφλι, καινούργιες περιοχές θα δημιουργούνται, ενώ στις υπάρχουσες θα προστίθενται νέα pixels και θα διευρύνονται. Υπάρχει περίπτωση επίσης δύο μικρές περιοχές που ξεκίνησαν από δύο διαφορετικά τοπικά ελάχιστα να συγχωνευθούν σε κάποιο επίπεδο και από εκεί και πέρα να εξελίσσονται σαν μία περιοχή. Σε κάθε επίπεδο κατωφλίωσης οι ακραίες περιοχές είναι οι συνεκτικές συνιστώσες των pixels της εικόνας που έχουν ένταση μικρότερη από το κατώφλι. Από το σύνολο όλων των ακραίων περιοχών για όλα τα επίπεδα, θα πρέπει να επιλεγούν ορισμένες που είναι περισσότερο σταθερές σε ένα εύρος τιμών κατωφλίου. Έστω R μία ακραία περιοχή της εικόνας που εξελίσσεται κατά τη διαδικασία αύξησης του κατωφλίου. Αν Δ συμβολίσουμε μία παράμετρο της μεθόδου που αντιπροσωπεύει το εύρος των επιπέδων που επιθυμούμε να μένει σταθερή μια MSER, τότε $R_{-\Delta}$ και $R_{+\Delta}$ είναι η αντίστοιχη περιοχή σε Δ επίπεδα πριν και Δ επίπεδα μετά και το μέτρο μεταβολής της περιοχής R μπορεί να οριστεί από την ακόλουθη σχέση:

$$q(R) = \frac{\left|R_{+\Delta} - R_{-\Delta}\right|}{\left|R\right|} \qquad \textbf{(εξ. 2.11)}$$

Επομένως από ένα σύνολο ακραίων περιοχών που εξελίσσονται στην κατωφλίωση ως περισσότερο σταθερή μπορεί να επιλεγεί η περιοχή R στην οποία παρουσιάζεται τοπικό ελάχιστο για την ποσότητα q(R). Για να απλοποιήσουμε την παρουσίαση του αλγορίθμου, υποθέτουμε ότι η ένταση της εικόνας είναι συνάρτηση μιας μεταβλητής x, οπότε οι ακραίες περιοχές θα είναι υποσύνολα του ημιάξονα των x. Στο σχήμα 2.10 βλέπουμε την εξέλιξη μιας περιοχής σε διάφορα επίπεδα. Η ένταση I της εικόνας παριστάνεται ως καμπύλη και τα pixels της βρίσκονται διατεταγμένα στον άξονα των τετμημένων x, ενώ οι σημειωμένες τιμές στον άξονα των τεταγμένων αντιστοιχούν στα κατώφλια στα οποία προέκυψαν οι αντίστοιχες ακραίες περιοχές που σημειώνονται με τα βέλη.

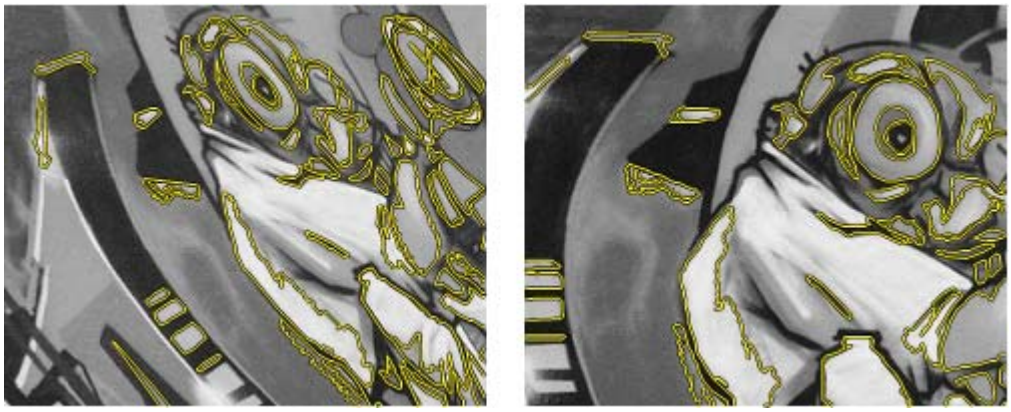



**Σχ. 2.10:** Η εξέλιξη μιας ακραίας περιοχής R σε μεταβολή του κατωφλίου I(.) και οι περιοχές $R_{-\Delta}$ και $R_{+\Delta}$ για τα αντίστοιχα κατώφλια ([VLFeat])

Ο τρόπος κατασκευής του συνόλου των περιοχών MSER της εικόνας οδηγεί σε πολλά πλεονεκτήματα. Καταρχάς μία μονοτονική μεταβολή στην ένταση των pixels δεν επηρεάζει τις ακραίες περιοχές της. Επίσης αλλαγές στην κλίμακα ενός αντικειμένου και πάλι δεν θα επηρεάσει την περιοχή που εξάγεται πάνω σε αυτό. Γενικότερα, οποιαδήποτε γεωμετρική αφινική μεταβολή του περιεχομένου της εικόνας δεν θα επηρεάσει τη μέθοδο, αφού τα pixels που αντιστοιχούν σε μια MSER περιοχή απλά θα μετατοπιστούν κατά τον αφινικό μετασχηματισμό και θα ανήκουν μια άλλη συνεκτική συνιστώσα, η οποία θα επιλεγεί και πάλι ως πολύ σταθερή (MSER). Από την πλευρά της υπολογιστικής πολυπλοκότητας, η διαδικασία μπορεί να επιταχυνθεί αν τα pixels της εικόνας τοποθετηθούν με αύξουσα τιμή έντασης και στη συνέχεια υπολογιστούν όλες οι συνεκτικές συνιστώσες με έναν αποδοτικό αλγόριθμο union-find. Η χρονική πολυπλοκότητα του πρώτου βήματος είναι O(n) και του δεύτερου O(nloglogn), όπου n το πλήθος των pixels της εικόνας, γεγονός που αποδεικνύει την ταχύτητα της μεθόδου. Η ταχύτητα του αλγορίθμου MSER μπορεί να μειωθεί ακόμα περισσότερο όπως παρουσιάζεται σε μια πιο πρόσφατη υλοποίηση [Nister et al., 2008], στην οποία χρησιμοποιείται διαφορετικός τρόπος διάταξης των pixels και σε κάθε στάδιο του αλγορίθμου εξετάζεται μία μόνο συνεκτική συνιστώσα.

Ο παραπάνω αλγόριθμος έχει κοινά σημεία με έναν αλγόριθμο κατάτμησης πλημμυρισμού (watershed segmentation algorithm) αλλά η βασική διαφορά τους είναι ότι ενώ στον watershed αλγόριθμο η κατάτμηση γίνεται στα σημεία όπου δύο μέτωπα (watersheds) συναντώνται, δηλαδή εκεί που συμβαίνει μεγάλη αλλαγή και οι δύο περιοχές συγχωνεύονται, αντίθετα στον MSER αλγόριθμο τα σημεία όπου οι περιοχές παραμένουν σχεδόν αμετάβλητες

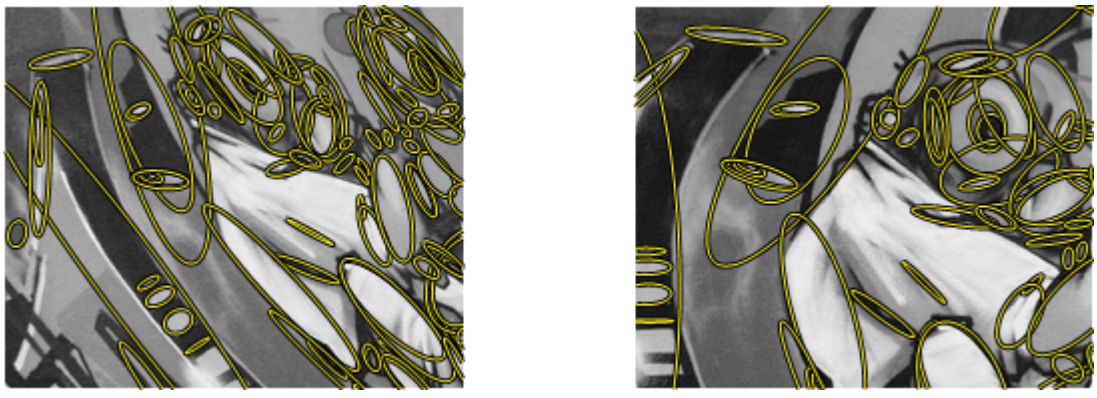
επιλέγονται ως ενδιαφέροντα. Πρέπει ιδιαίτερα να τονιστεί ότι δεν αναζητείται κάποιο βέλτιστο κατώφλι, αλλά όλες οι τιμές και όλα τα επίπεδα εξετάζονται ώστε να εντοπιστούν οι πιο σταθερές συνεκτικές συνιστώσες. Το αποτέλεσμα της μεθόδου δεν είναι ούτε μια κατάτμηση της εικόνας ούτε και μια δυαδικοποίηση (συνολική μάσκα), αλλά πολλές περιοχές μερικές από τις οποίες μπορεί να περιέχονται σε κάποιες άλλες (και μερικά pixels να ανήκουν σε περισσότερες της μιας σταθερές περιοχές MSER). Ένα παράδειγμα εξόδου του αλγορίθμου φαίνεται στο σχήμα 2.11.



**Σχ. 2.11:** Ανιχνευμένες περιοχές MSER στην εικόνα graffiti για διαφορετικές οπτικές γωνίες

Για τις περισσότερες μεθόδους που ανιχνεύουν σημεία με τρόπο ανεξάρτητο από αφινικούς μετασχηματισμούς, συνήθως εξάγεται εκτός από το σημείο και μία ελλειπτική περιοχή γύρω από αυτό. Παρόλ' αυτά στην περίπτωση της μεθόδου MSER το αποτέλεσμα είναι κάποιες περιοχές ακανόνιστου σχήματος, στις οποίες όμως μπορεί να προσαρμοστεί μία έλλειψη τέτοια που να έχει τις ίδιες ροπές πρώτης και δεύτερης τάξης με το σχήμα της αρχικής ανιχνευμένης περιοχής. Τέτοιες ελλείψεις φαίνονται στις εικόνες του σχήματος 2.12. Αυτή η διαδικασία είναι χρήσιμη, έτσι ώστε η περιοχή να μετασχηματιστεί σε κύκλο (με τον ίδιο τρόπο που κανονικοποιούνται και οι Harris/Hessian Affine περιοχές) και στη συνέχεια να κατασκευαστεί το διάνυσμα χαρακτηριστικών από την κανονικοποιημένη περιοχή.

Πρέπει τέλος να τονίσουμε ότι η τεχνική MSER εξάγει αφινικές περιοχές της εικόνας (affine invariant), οι οποίες σε κάποιες περιπτώσεις μπορούν αν θεωρηθούν ότι μοιάζουν με κηλίδες. Στη γενικότερη περίπτωση όμως αντιστοιχούν σε μεγαλύτερες περιοχές και κυρίως εντοπίζονται σε τμήματα αντικειμένων που ξεχωρίζουν από το περιβάλλον τους με μεγάλες μεταβολές στην ένταση. Γι' αυτό το λόγο η μέθοδος είναι πολύ αποτελεσματική σε εικόνες που περιέχουν σκηνές με δομημένα αντικείμενα και ξεκάθαρα όρια περιοχών (π.χ. graffiti).



**Σχ. 2.12:** Ελλειπτικές περιοχές MSER στην εικόνα graffiti για διαφορετικές οπτικές γωνίες

### 2.2.6 Ανιχνευτής Difference of Gaussian (DoG)

Στις προηγούμενες μεθόδους συναντήσαμε τον υπολογισμό πολύπλοκων μαθηματικών παραστάσεων, που καθιστούν τη διαδικασία ανίχνευσης απαιτητική σε χρόνο. Στην παράγραφο αυτή παρουσιάζεται μία αποδοτική μέθοδος ανίχνευσης σημείων ενδιαφέροντος που βασίζεται στην προσέγγιση του φίλτρου LoG (laplacian of gaussian) και προτάθηκε από τον Lowe to 1999 στην πολύ γνωστή πλέον δημοσίευση [Lowe, 1999], όπου παρουσιάζεται ο ανιχνευτής DoG μαζί με τον πλέον διαδεδομένο περιγραφέα SIFT.

Όπως ήδη αναφέρθηκε, για να προκύψουν σημεία ανεξάρτητα της κλίμακας, συνηθίζεται να χρησιμοποιείται μία συνάρτηση η οποία μεταβάλλεται με την κλίμακα και παρουσιάζει τοπικό μέγιστο στη χαρακτηριστική κλίμακα, σε αυτή δηλαδή που ταιριάζει το κάθε χαρακτηριστικό. Για να γίνει η αναζήτηση σε όλες τις δυνατές κλίμακες θα πρέπει να κατασκευαστεί ένας χώρος κλίμακας με φιλτράρισμα της εικόνας και η πιο κατάλληλη συνάρτηση για να χρησιμοποιηθεί ως πυρήνας είναι η γκαουσιανή (βλέπε σχετική βιβλιογραφία του [Lindeberg, 1994]). Ο γκαουσιανός χώρος κλίμακας L προκύπτει από διαδοχική συνέλιξη γκαουσιανών συναρτήσεων (αυξανόμενου σ) με τη διδιάστατη εικόνα I:

$$L(x, y, \sigma) = g(x, y, \sigma) * I(x, y) \qquad \textbf{(εξ. 2.12)}$$

όπου η γκαουσιανή είναι:

$$g(x, y, \sigma) = \frac{1}{2\pi\sigma^2} \exp\{-(x^2 + y^2) / 2\sigma^2\} \qquad \textbf{(εξ. 2.13)}$$

Ο Lindeberg είχε προτείνει τη χρήση της κανονικοποιημένης με την κλίμακα λαπλασιανής (laplacian of gaussian) LoG, δηλαδή της σχέσης:

$$LoG = \sigma^2 \cdot \nabla^2 g = \sigma^2 \cdot (\frac{\partial^2}{\partial x^2} g + \frac{\partial^2}{\partial y^2} g) \qquad \textbf{(εξ. 2.14)}$$

ως φίλτρο για τη συνέλιξη με την εικόνα (βλέπε εξίσωση 2.7), ούτως ώστε να προκύψει ανεξαρτησία από την κλίμακα. Αν αντί της λαπλασιανής χρησιμοποιηθεί μια προσέγγισή της, τότε οδηγούμαστε σε μία ταχύτερη μέθοδο εντοπισμού της χαρακτηριστικής κλίμακας (αφού θα παρακάμψουμε τον υπολογισμό των δευτέρων παραγώγων). Με αυτή την ιδέα και ξεκινώντας από την ισοτροπική εξίσωση διάχυσης θερμότητας, η οποία είναι:

$$\frac{\partial g}{\partial \sigma} = \sigma \cdot \nabla^2 g \qquad \textbf{(εξ. 2.15)}$$

ο Lowe παρατηρεί [Lowe, 2004] ότι μπορεί να υπολογίσει προσεγγιστικά την λαπλασιανή από τη διαφορά μεταξύ των συνελίξεων της εικόνας δύο διαδοχικών πολύ κοντινών επιπέδων, έστω σ και kσ, ως εξής:

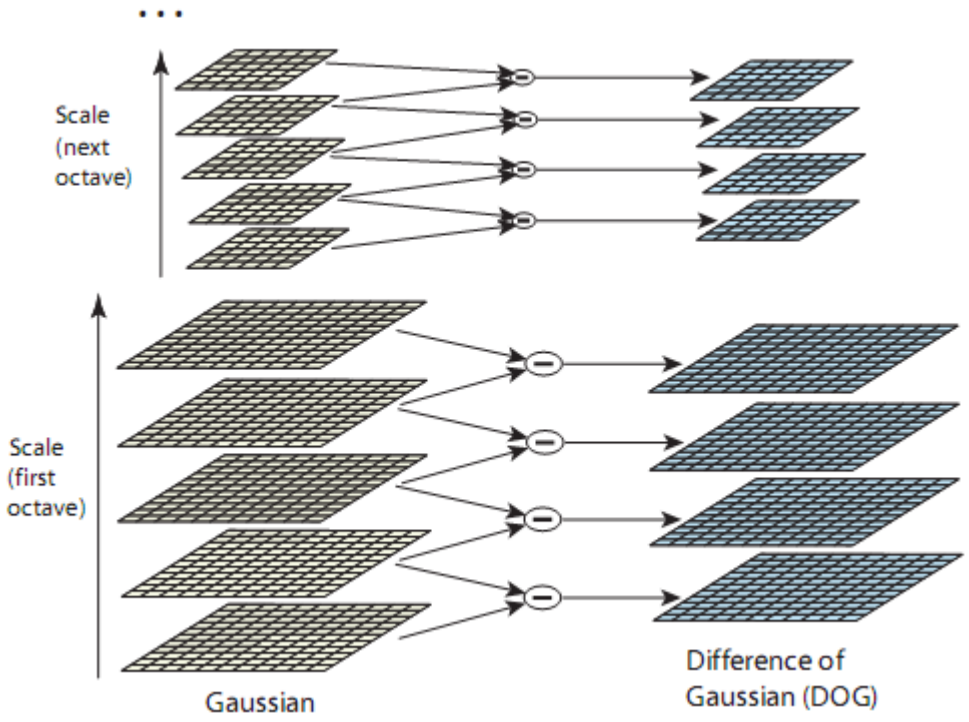
$$\sigma \cdot \nabla^2 g = \frac{\partial g}{\partial \sigma} \approx \frac{g(x, y, k\sigma) - g(x, y, \sigma)}{k\sigma - \sigma} \qquad \textbf{(εξ. 2.16)}$$
$$\Rightarrow g(x, y, k\sigma) - g(x, y, \sigma) \approx (k-1) \cdot \sigma^2 \cdot \nabla^2 g$$

Επομένως μπορεί να χρησιμοποιηθεί ένας προσεγγιστικός χώρος κλίμακας D αντί του περισσότερο πολύπλοκου λαπλασιανού, ο χώρος διαφορών γκαουσιανών (difference of gaussian)

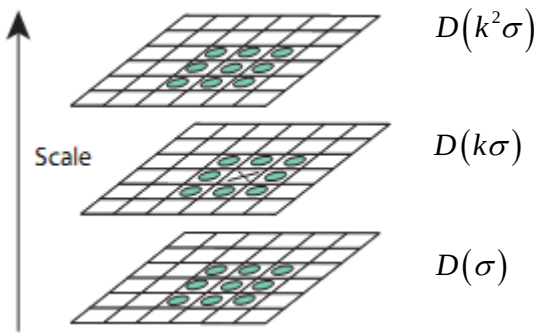
ο οποίος προκύπτει από τη διαφορά μεταξύ διαδοχικών κλιμάκων του γκαουσιανού χώρου ως εξής:

$$D(x,y,\sigma) = [g(x,y,k\sigma) - g(x,y,\sigma)] * I(x,y) = L(x,y,k\sigma) - L(x,y,\sigma) \quad \textbf{(εξ. 2.17)}$$

Για την κατασκευή του χώρου κλίμακας, υπολογίζονται διαδοχικά οι συνελίξεις της εικόνας με γκαουσιανά φίλτρα παραμέτρου σ, έστω για s+1 φορές, οπότε ολοκληρώνεται μία οκτάβα γκαουσιανά ομαλοποιημένων εικόνων (Gaussian) με διαφορά μεταξύ των επιπέδων της $k = 2^{1/s}$. Στη συνέχεια αφαιρούνται οι εικόνες των διαδοχικών επιπέδων ανά δύο, οπότε προκύπτουν s επίπεδα για τον χώρο των γκαουσιανών διαφορών (Difference of Gaussian - DoG), ο οποίος προσεγγίζει τον λαπλασιανό. Για την κατασκευή της επόμενης οκτάβας, η εικόνα υποδειγματοληπτείται ώστε να μειωθεί η πολυπλοκότητα των υπολογισμών. Η όλη διαδικασία δημιουργίας των χώρων κλίμακας φαίνεται στο σχήμα 2.13. Στη συνέχεια, στον τρισδιάστατο χώρο κλίμακας DoG που προκύπτει εντοπίζονται τα σημεία που αποτελούν τοπικό ακρότατο ταυτόχρονα και στο επίπεδο (x,y) αλλά και στον κατακόρυφο άξονα της κλίμακας, εντοπίζεται δηλαδή ταυτόχρονα και η θέση αλλά και η χαρακτηριστική κλίμακα του σημείου ενδιαφέροντος. Κάθε υποψήφιο σημείο συγκρίνεται στο επίπεδο με τα οκτώ γειτονικά του σημεία, και στη συνέχεια στο χώρο με τα εννέα σημεία στο προηγούμενο και στο επόμενο επίπεδο κλίμακας, όπως φαίνεται στο σχήμα 2.14 η σύγκριση σε έναν κύβο 3×3×3.




**Σχ. 2.13:** Κατασκευή διαδοχικών επιπέδων του γκαουσιανού χώρου κλίμακας και του προσεγγιστικού χώρου κλίμακας DoG ([Lowe, 2004])

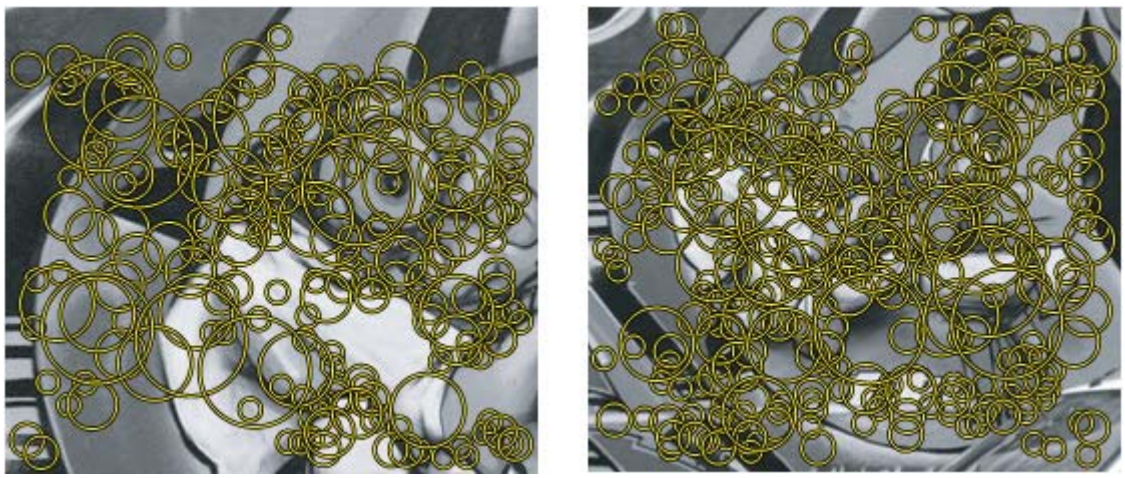



**Σχ. 2.14:** Εντοπισμός σημείων ενδιαφέροντος σε τοπικά ακρότατα του χώρου DoG ([Lowe, 2004])

Ο ανιχνευτής DoG εντοπίζει σημεία ενδιαφέροντος ανεξάρτητα από αλλαγές της κλίμακας (scale invariant). Λόγω της μορφής του φίλτρου LoG (ζωνοπερατό), είναι λογικό η λαπλασιανή μέθοδος να έχει μεγάλη απόκριση σε σημεία κηλίδων, δηλαδή σε περιοχές οι οποίες έχουν κυκλική μορφή και διαφέρουν έντονα από τη γύρω γειτονιά τους (για παράδειγμα σκούρες κηλίδες σε φωτεινό περιβάλλον ή το αντίθετο). Επειδή όμως ταυτόχρονα δίνει ως αποτέλεσμα σημεία που βρίσκονται κοντά σε ακμές, προτείνεται ένα επιπλέον στάδιο φιλτραρίσματος των εντοπισμένων σημείων [Lowe, 2004]. Αυτό βασίζεται στον υπολογισμό των ιδιοτιμών του χεσσιανού πίνακα H, που όμως δεν επιβαρύνει την διαδικασία γιατί δεν υπολογίζεται σε ολόκληρη την εικόνα παρά μόνο στα συγκεκριμένα τοπικά ακρότατα που έχουν ήδη ανιχνευθεί. Όταν η περιοχή βρίσκεται κοντά σε ακμή οι ιδιοτιμές θα διαφέρουν κατά μεγάλο ποσοστό. Μία σχέση που περιλαμβάνει τον λόγο των ιδιοτιμών r είναι η εξής:

$$\frac{tr(H)}{\det(H)} = \frac{(r+1)^2}{r} \qquad \textbf{(εξ. 2.18)}$$

οπότε χρησιμοποιώντας για κατώφλι μία τιμή r = 10 (προκύπτει πειραματικά) απορρίπτονται σημεία που έχουν λόγο ίχνους προς ορίζουσα μεγαλύτερο από $(r+1)^2 / r = 12.1$. και κρατούνται τα σημεία που είναι πραγματικές κηλίδες. Ένα παράδειγμα ανίχνευσης σημείων ενδιαφέροντος DoG (όπως θα αποκαλούνται σύντομα) δίνεται στο σχήμα 2.15.



**Σχ. 2.15:** Σημεία ενδιαφέροντος DoG στην εικόνα graffiti για διαφορετικές όψεις (αλλαγή κλίμακας)

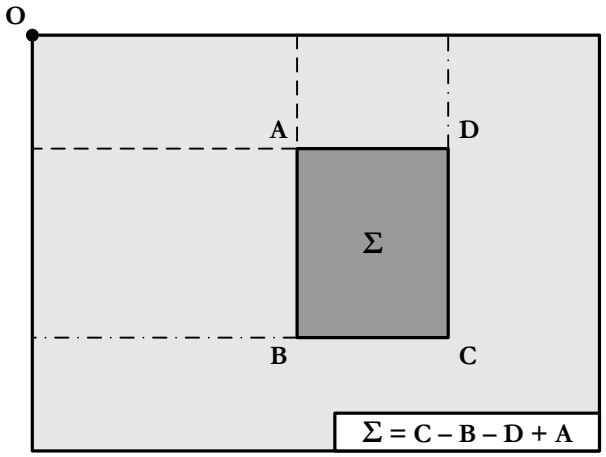
### 2.2.7 Ανιχνευτής Fast Hessian (FastH)

Προκειμένου να βελτιώσουν ακόμη περισσότερο την ταχύτητα του ανιχνευτή τοπικών χαρακτηριστικών οι Bay et al. [Bay et al., 2006] προτείνουν τον προσεγγιστικό υπολογισμό του χεσσιανού πίνακα με τη χρήση ολοκληρωτικών εικόνων (integral images), στα πλαίσια της νέας μεθόδου τοπικών χαρακτηριστικών που ονομάζεται SURF (speeded up robust features).

Μια ολοκληρωτική εικόνα $I_\Sigma$ είναι μια εικόνα ίδιων διαστάσεων με την αρχική εικόνα I, αλλά κάθε pixel της στο σημείο z=(x,y) έχει τιμή ίση με το άθροισμα των τιμών όλων των pixels της αρχικής εικόνας που περιέχονται σε ένα ορθογώνιο που ορίζεται από την πάνω αριστερή γωνία της εικόνας και από το σημείο z. Οι τιμές της integral image υπολογίζονται μέσω της εξίσωσης:

$$I_\Sigma(x,y) = \sum_{i=0}^{i\le x}\sum_{j=0}^{j\le y} I(i,j) \qquad \textbf{(εξ. 2.19)}$$

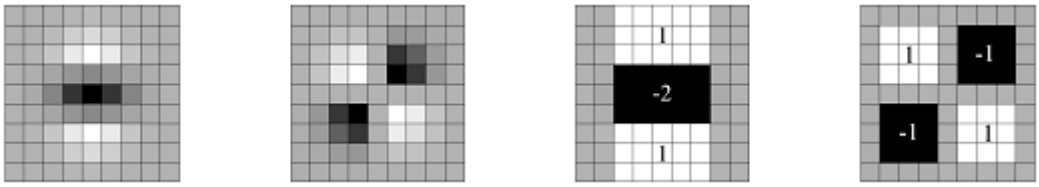
Αφού πρώτα κατασκευάσουμε την integral image $I_\Sigma$, για τον υπολογισμό του αθροίσματος των εντάσεων όλων των pixels οποιουδήποτε ορθογωνίου της αρχικής εικόνας I απαιτούνται μόλις τέσσερις προσθέσεις, ανεξαρτήτως του μεγέθους της περιοχής. Αυτό επιτυγχάνεται με τη χρήση των τιμών της $I_\Sigma$ στις κορυφές του ορθογωνίου όπως γίνεται εύκολα αντιληπτό από το σχήμα 2.16. Με αυτόν τον τρόπο επιταχύνεται κατά πολύ ο υπολογισμός των κυματιδίων Haar (Haar wavelets) και γενικά κάθε φίλτρου ή συνέλιξης που έχει μορφή κουτιού (box filter).




**Σχ. 2.16:** Υπολογισμός του αθροίσματος Σ εντός ορθογωνίου μέσω της integral image

Ο ανιχνευτής σημείων ενδιαφέροντος Fast Hessian, όπως προδίδει και το όνομά του, χρησιμοποιεί τον χεσσιανό πίνακα H (βλέπε εξίσωση 2.9), όπως ήδη έχουμε δει σε άλλες μεθόδους. Στην περίπτωση αυτή όμως χρησιμοποιείται το ίδιο μέτρο, η ορίζουσα του πίνακα H, για τον εντοπισμό της θέσης του σημείου ενδιαφέροντος αλλά και την εκτίμησης της κλίμακάς του. Ο χεσσιανός πίνακας προσεγγίζεται χρησιμοποιώντας ένα σύνολο από φίλτρα-κουτιά και δεν χρησιμοποιείται ομαλοποίηση κατά τη διαδικασία της εξέτασης των διαφόρων κλιμάκων, έτσι ώστε να υπολογιστεί η θέση του σημείου με περισσότερη ακρίβεια (ως γνωστόν το θόλωμα της εικόνας με την γκαουσιανή οδηγεί σε σημαντικά σφάλματα μετατόπισης των χαρακτηριστικών). Η συνέλιξη με τις γκαουσιανές μπορεί να είναι κατάλληλη για την ανάλυση του χώρου κλίμακας,

αλλά διακριτοποιώντας τις τιμές τους οδηγούμαστε στην συσσώρευση σφαλμάτων. Γι' αυτό το λόγο οι συγγραφείς του SURF προτείνουν τη χρήση των box-filters για να προσεγγίσουν τις δεύτερες παραγώγους γκαουσιανών συναρτήσεων, δεδομένου ότι αυτό δεν θα επηρεάσει πολύ το αποτέλεσμα αφού ήδη έχουν προστεθεί σφάλματα στη διαδικασία. Χρησιμοποιώντας την integral image της αρχικής εικόνας (η οποία και κατασκευάζεται μία φορά μόνο στην αρχή) οι δεύτερες παράγωγοι υπολογίζονται πολύ γρήγορα, ανεξαρτήτως μεγέθους.




**Σχ. 2.17:** Αριστερά: οι διακριτοποιημένες γκαουσιανές δεύτερες παράγωγοι (ως προς y και xy)
Δεξιά: οι προσεγγίσεις των παραγώγων αυτών με box-filters ([Bay et al., 2006])

Στο σχήμα 2.17 παρουσιάζονται οι γκαουσιανές παράγωγοι δεύτερης τάξης με μέγεθος 9×9. Στο αριστερό μέρος του σχήματος φαίνεται η διακριτοποίηση των παραγώγων σε κάθε pixel, που σημαίνει ότι εισάγονται αναγκαστικά σφάλματα υπολογισμών, αφού το αποτέλεσμα της συνέλιξης υπόκειται αναγκαστικά σε υποδειγματοληψία στη συνέχεια. Επίσης η διατύπωση ότι δεν δημιουργούνται νέες δομές καθώς προχωράμε σε μεγαλύτερες κλίμακες έχει αποδειχθεί για την περίπτωση μονοδιάστατων σημάτων, αλλά δεν είναι προφανής η γενίκευσή της για τη διδιάστατη περίπτωση.

Στο δεξί μέρος του σχήματος 2.17 φαίνονται οι προσεγγίσεις των προηγούμενων παραγώγων με box filters, τα οποία υπολογίζονται εύκολα και γρήγορα με τη μέθοδο του σχήματος 2.16 (το γκρίζο κομμάτι της εικόνας αντιστοιχεί στην τιμή μηδέν). Τα 9×9 φίλτρα που παρουσιάζονται εδώ είναι και τα πιο μικρά φίλτρα που χρησιμοποιούνται για να υπολογιστεί η πιο λεπτομερής κλίμακα σ = 1.2. Συμβολίζουμε τις προσεγγίσεις των δευτέρων παραγώγων ως προς ως προς x, ως προς y και ως προς xy με $D_{xx}$, $D_{yy}$ και $D_{xy}$ αντίστοιχα. Στην έκφραση υπολογισμού της προσεγγιστικής ορίζουσας του χεσσιανού πίνακα H εισάγονται βάρη τα οποία συμβάλλουν στην εξισορρόπηση μεταξύ των γκαουσιανών και των προσεγγιστικών γκαουσιανών πυρήνων (οι τιμές των βαρών για λόγους ταχύτητας θα πρέπει να είναι απλές). Για να ρυθμιστούν κατάλληλα τα βάρη στην έκφραση της προσεγγιστικής ορίζουσας χρησιμοποιείται η εξίσωση:

$$w = \frac{|I_{xy}(1.2)|_F |D_{yy}(9)|_F}{|I_{yy}(1.2)|_F |D_{xy}(9)|_F} = 0.912... \approx 0.9$$, όπου $|.|_F$ η Frobenius νόρμα

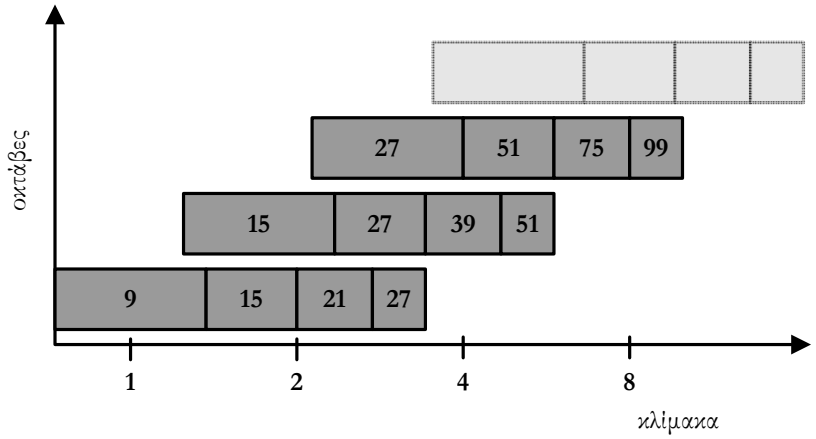
Οπότε η προσεγγιστική σχέση που δίνει την ορίζουσα του χεσσιανού πίνακα είναι:

$$\det(H_{approx}) = D_{xx} \cdot D_{yy} - (0.9 \cdot D_{xy})^2 \qquad \textbf{(εξ. 2.20)}$$

Ένας χώρος κλίμακας αποτελείται συνήθως από πυραμίδες, όπου η αρχική εικόνα ομαλοποιείται διαδοχικά με γκαουσιανά φίλτρα και στη συνέχεια γίνεται υποδειγματοληψία καθώς προχωράμε προς τα ανώτερα επίπεδα κλίμακας. Η τελευταία συμβαίνει για τη μείωση των υπολογισμών, καθώς οι όλο και πιο "θολωμένες" εικόνες περιέχουν μικρότερη πληροφορία και άρα μπορούν να αναπαρασταθούν με λιγότερα pixels. Με τη χρήση των box-filters (integral images) δεν χρειάζεται πλέον η επίπονη κατασκευή των επιπέδων της πυραμίδας, αφού για τη δημιουργία των διαφόρων κλιμάκων απλά εφαρμόζεται στην αρχική εικόνα ένα φίλτρο αυξανόμενης διάστασης (με το ίδιο πολύ μικρό κόστος υπολογισμού κάθε φορά).

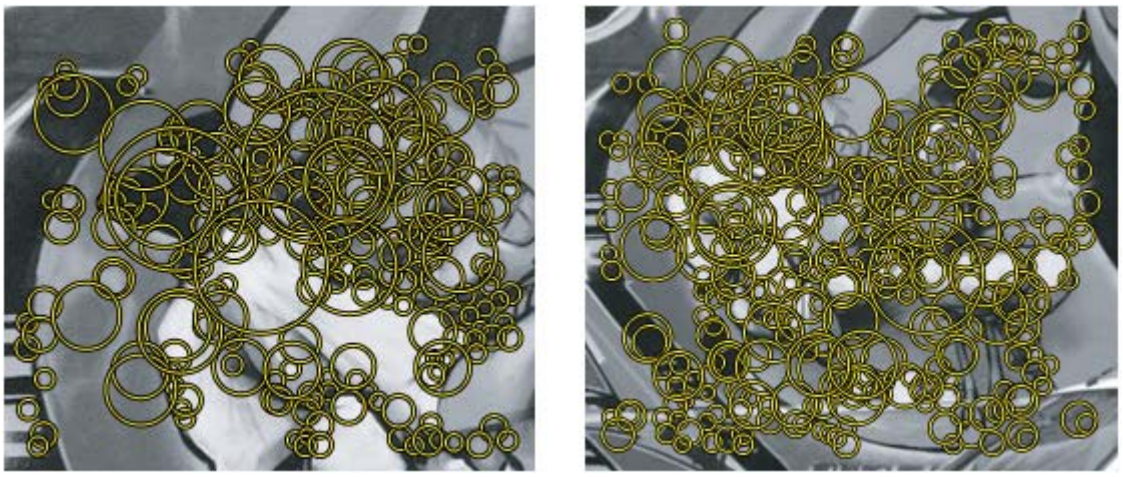
Η ανάλυση του χώρου κλίμακας ξεκινάει από το 9×9 φίλτρο όπως είπαμε, το οποίο αντιστοιχεί στο αρχικό επίπεδο κλίμακας s = 1.2, και συνεχίζει με φίλτρα μεγαλύτερων

διαστάσεων, κατάλληλα για να ταιριάζουν με τη μορφή των διακριτών προσεγγιστικών box-filters. Έτσι τα φίλτρα στα επόμενα επίπεδα της αρχικής οκτάβας είναι 15×15, 21×21, 27×27. Στις επόμενες οκτάβες το βήμα μεταξύ των επιπέδων αυξάνεται ανάλογα, οπότε στη δεύτερη οκτάβα το βήμα γίνεται ίσο με 12 (από 6 που ήταν στην αρχική) κ.ο.κ. Στο σχήμα 2.18 φαίνονται οι τρεις πρώτες οκτάβες και τα αντίστοιχα μεγέθη των φίλτρων που χρησιμοποιούνται (ο λογαριθμικός άξονας των τετμημένων αναπαριστάνει την κλίμακα). Να σημειώσουμε ότι η κλίμακα αυξάνεται ανάλογα με το μέγεθος του φίλτρου, δηλαδή για το φίλτρο 27×27 η κλίμακα της εικόνας είναι s = 3×1.2 = 3.6.




**Σχ. 2.18:** Μεγέθη των φίλτρων που χρησιμοποιούνται για τις τρεις πρώτες οκτάβες του χώρου κλίμακας

Τα σημεία ενδιαφέροντος εντοπίζονται ταυτόχρονα και στο επίπεδο της εικόνας (θέση x,y) και στον χώρο κλίμακας (κλίμακα s), και είναι όπως ακριβώς και στη μέθοδο DoG τα τοπικά μέγιστα και ελάχιστα σε έναν κύβο 3×3×3. Χρησιμοποιείται επίσης μία μέθοδος παρεμβολής για την ακριβή εκτίμηση θέσης και κλίμακας [Brown et al., 2002]. Η προσεγγιστική ορίζουσα του πίνακα Hessian (εξίσωση 2.20), οι τιμές της οποίας αποτελούν τον τρισδιάστατο χώρο αναζήτησης, αποκρίνεται κυρίως σε κηλίδες, είναι ανεξάρτητη από αλλαγές κλίμακας και είναι περίπου πέντε φορές ταχύτερη από τη μέθοδο DoG. Ένα παράδειγμα ανίχνευσης Fast Hessian φαίνεται στο σχήμα 2.19.



**Σχ. 2.19:** Σημεία ενδιαφέροντος Fast Hessian στην εικόνα graffiti για διαφορετικές όψεις (αλλαγή κλίμακας)

## 2.3 Τοπικοί περιγραφείς

### 2.3.1 Επισκόπηση τοπικών περιγραφέων

Μετά την εξαγωγή της κατάλληλης περιοχής ενδιαφέροντος (είτε ως γειτονιά γύρω από σημείο ενδιαφέροντος είτε ως περιοχή που ανιχνεύθηκε από την μέθοδο), το δεύτερο πολύ σημαντικό βήμα είναι η κατάλληλη περιγραφή του σημασιολογικού τοπικού περιεχομένου με ένα διάνυσμα σε ένα χώρο χαρακτηριστικών υψηλής διάστασης. Στη βιβλιογραφία μπορούν να αναζητηθούν πολλά είδη περιγραφέων, οι οποίοι διαχωρίζονται σε βασικές κατηγορίες, όπως: σε περιγραφείς βασισμένους σε κατανομές (ιστογράμματα), για παράδειγμα ένα απλό ιστόγραμμα των εντάσεων των pixels της περιοχής ή ένα ιστόγραμμα κατανομής των παραγώγων (όπως το SIFT ή το SURF που θα αναλυθούν παρακάτω), σε περιγραφείς χωρικής συχνότητας (όπως φίλτρα Gabor ή wavelets), σε περιγραφείς βασισμένους σε παραγώγους, όπως σταθερά χαρακτηριστικά παραγώγων (differential invariants) που αποτελούν το λεγόμενο local jet [Schmid et al., 1997], κατευθυνόμενα φίλτρα (steerable filters), ή μιγαδικά φίλτρα (complex filters) [Baumberg, 2000], και σε πολλούς άλλους περιγραφείς που δεν ανήκουν σε κάποια κατηγορία από τις προηγούμενες, όπως για παράδειγμα τα χαρακτηριστικά ροπών (moment invariants).

Πολλές τεχνικές εξαγωγής τοπικών περιγραφέων αναλύονται με συγκριτικά πειράματα [Mikolajczyk et al., 2003], απ' όπου προκύπτει το διάνυσμα χαρακτηριστικών SIFT καλύτερο ως προς τις επιδόσεις του (συνδυασμό μέτρων ακρίβειας και ανάκτησης, precision-recall) σε σχέση με όλα τα υπόλοιπα. Επίσης δύο παραλλαγές του SIFT, οι μέθοδοι PCA-SIFT και GLOH, που χρησιμοποιούν αρχικά περισσότερα στοιχεία στο διάνυσμα χαρακτηριστικών, αλλά μειώνουν τη διάσταση του χώρου με την ανάλυση σε κύριες συνιστώσες (τεχνική PCA, principal component analysis), έχουν παρόμοιες ή και καλύτερες επιδόσεις με την αρχική μέθοδο. Εξαιτίας όμως της αυξημένης χρονικής πολυπλοκότητας της ανάλυσης PCA και την ανάγκη ορισμού ενός πίνακα συνδιασποράς (διαφορετικές σε κάθε σύνολο εικόνων με διαφορετικό περιεχόμενο), δεν εξετάζονται εδώ. Στην παράγραφο αυτή θα αναλυθεί η τεχνική SIFT και επίσης η πολύ ταχύτερη τεχνική SURF, που βασίζεται στην ίδια ιδέα κατανομής των κατευθύνσεων των παραγώγων.

### 2.3.2 Κανονικοποίηση περιοχής ενδιαφέροντος

Πριν προχωρήσουμε παρακάτω, θα πρέπει να προσδιορίσουμε την περιοχή από την οποία θα προέλθει η πληροφορία για την κατασκευή του διανύσματος χαρακτηριστικών, την οποία θα ονομάζουμε περιοχή μέτρησης (measurement region). Στην περίπτωση των κυκλικών περιοχών ενδιαφέροντος, η περιοχή μέτρησης είναι και αυτή κυκλική και μάλιστα είναι ανάλογη της χαρακτηριστικής κλίμακας που έχει υπολογιστεί από τον ανιχνευτή. Συνήθως χρησιμοποιείται κυκλική περιοχή ακτίνας $3\times\sigma$ (όπου $\sigma$ η χαρακτηριστική κλίμακα), έτσι ώστε να περιλαμβάνεται και πρόσθετη πληροφορία για την περιοχή από τη γειτονιά του σημείου. Στην περίπτωση των ανιχνευτών που επιστρέφουν μία περιοχή ανεξάρτητη από αφινικούς μετασχηματισμούς της εικόνας, το σχήμα είναι μία έλλειψη. Είτε πρόκειται για την περίπτωση των Harris/Hessian Affine είτε για τις περιοχές MSER, η έλλειψη καθορίζεται από έναν συμμετρικό πίνακα ροπών δεύτερης τάξης, έστω M με:

$$M = \begin{bmatrix} a & b \\ b & c \end{bmatrix} \qquad \textbf{(εξ. 2.21)}$$

Οι ιδιοτιμές και τα ιδιοδιανύσματα του πίνακα M δείχνουν την τοπική δομή γύρω από το σημείο, δηλαδή αναπαριστούν το αφινικό της σχήμα (έλλειψη). Τα ιδιοδιανύσματα $u_1$ και $u_2$ του πίνακα M δίνουν τις κατευθύνσεις των αξόνων της έλλειψης, ενώ οι ιδιοτιμές του $\lambda_1$ και $\lambda_2$ δίνουν τα μήκη των αξόνων της. Αν $x = [x_1\ x_2]^T$ είναι τα σημεία στο διδιάστατο επίπεδο της εικόνας, τότε η εξίσωση της έλλειψης είναι:

$$x^T \cdot M \cdot x = 1 \qquad \textbf{(εξ. 2.22)}$$

Πρέπει να σημειώσουμε ότι στην περίπτωση που το κέντρο της έλλειψης βρίσκεται στο σημείο $[o_1\ o_2]$, τότε η παραπάνω σχέση ισχύει για $x = [x_1\text{-}o_1\ x_2\text{-}o_2]^T$, αλλά χωρίς βλάβη της γενικότητας θα δεχτούμε ότι το κέντρο της είναι το (0, 0), όπως φαίνεται και στο σχήμα 2.20.

Επειδή ο πίνακας M είναι συμμετρικός, διαγωνοποιείται από έναν ορθογώνιο πίνακα Q ($Q^{-1}=Q^T$) με διαγώνιο πίνακα τον πίνακα V των ιδιοτιμών ως εξής:

$$V = Q^T \cdot M \cdot Q \Leftrightarrow M = Q \cdot V \cdot Q^T \qquad \textbf{(εξ. 2.23)}$$

Έτσι αντικαθιστώντας στην εξίσωση 2.22 έχουμε:

$$x^T \cdot (Q \cdot V \cdot Q^T) \cdot x = 1 \Rightarrow$$
$$(Q^T x)^T \cdot V \cdot (Q^T x) = 1 \Rightarrow$$
$$w^T \cdot V \cdot w = 1$$

Δηλαδή μετασχηματίζοντας την έλλειψη σε ένα άλλο σύστημα συντεταγμένων μέσω της σχέσης:

$$x = Q \cdot w \Leftrightarrow w = Q^T \cdot x \qquad \textbf{(εξ. 2.24)}$$
$$(\text{αφού } Q^{-1} = Q^T)$$

έχουμε τη νέα μορφή της εξίσωσης της έλλειψης:

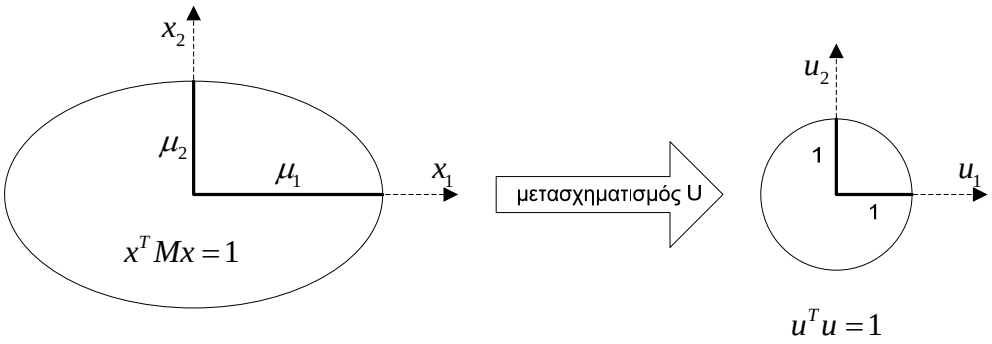
$$w^T \cdot V \cdot w = 1 \Rightarrow [w_1\ w_2] \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = 1 \Rightarrow \lambda_1 {w_1}^2 + \lambda_2 {w_2}^2 = 1 \qquad \textbf{(εξ. 2.25)}$$

από την οποία γίνεται αμέσως φανερή η σχέση μεταξύ των ιδιοτιμών του πίνακα M και των μηκών των αξόνων της έλλειψης ($2\mu_1$ και $2\mu_2$):

$$\mu_1 = {\lambda_1}^{-1/2} \text{ και } \mu_2 = {\lambda_2}^{-1/2} \qquad \textbf{(εξ. 2.26)}$$

Για να πάρουμε την περιοχή μέτρησης, για παράδειγμα περιοχή κ φορές μεγαλύτερη της αρχικής, αρκεί να διαιρεθεί ο πίνακας M με το τετράγωνο του κ, οπότε η εξίσωση 2.22 θα δίνει τα σημεία της νέας έλλειψης. Με την παραπάνω πράξη διαιρούνται οι ιδιοτιμές του M με $\kappa^2$, άρα σύμφωνα με την 2.26 πολλαπλασιάζονται οι άξονες της έλλειψης με κ (rescaling). Επίσης, το εμβαδόν της νέας έλλειψης είναι αυξημένο επί $\kappa^2$ σε σχέση με την αρχική, αφού γενικά το εμβαδόν E μιας έλλειψης υπολογίζεται σύμφωνα με τον τύπο:

$$E = \pi \cdot \mu_1 \cdot \mu_2 \qquad \textbf{(εξ. 2.27)}$$




**Σχ. 2.20:** Κανονικοποίηση έλλειψης σε μοναδιαίο κύκλο

Στη συνέχεια, πρέπει να καθοριστεί ο μετασχηματισμός μέσω του οποίου τα σημεία της έλλειψης M θα απεικονιστούν σε έναν μοναδιαίο κύκλο. Αυτό γίνεται μέσω του πίνακα μετασχηματισμού U:

$$U = M^{-1/2} \qquad \textbf{(εξ. 2.28)}$$

ο οποίος χρησιμοποιείται για να συνδέει σημεία της έλλειψης x με σημεία του κύκλου u μέσω της σχέσης:

$$x = U \cdot u \;\; \Leftrightarrow \;\; x = M^{-1/2} \cdot u \qquad \textbf{(εξ. 2.29)}$$

οπότε η σχέση 2.22 μέσω της 2.29 δίνει την εξίσωση του ζητούμενου μοναδιαίου κύκλου:

$$u^T \cdot u = 1 \qquad \textbf{(εξ. 2.30)}$$

Για τον υπολογισμό του πίνακα μετασχηματισμού U απαιτείται ο υπολογισμός της τετραγωνικής ρίζας του M, η οποία ορίζεται από τη σχέση:

$$M = (M^{1/2})^T \cdot (M^{1/2}) \qquad \textbf{(εξ. 2.31)}$$

Ο υπολογισμός της ρίζας μπορεί έυκολα να γίνει αφού πρώτα κατασκευάσουμε μία διαγωνοποίηση του M, όπως αυτή της σχέσης 2.23. Επειδή ο M είναι συμμετρικός, υπάρχει πάντοτε μια διαγωνοποίηση του και μάλιστα μέσω ορθοκανονικής βάσης. Αν Q είναι ο ορθοκανονικός πίνακας των κανονικοποιημένων ιδιοδιανυσμάτων του M και V είναι ο διαγώνιος πίνακας με στοιχεία διαγωνίου τις ιδιοτιμές του M, τότε από την ορθοκανονική διαγωνοποίηση του M προκύπτει η τετραγωνική του ρίζα ως εξής:

$$V = Q^T \cdot M \cdot Q \Rightarrow M = Q \cdot V \cdot Q^T \Rightarrow M^{1/2} = Q \cdot V^{1/2} \cdot Q^T \qquad \textbf{(εξ. 2.32)}$$

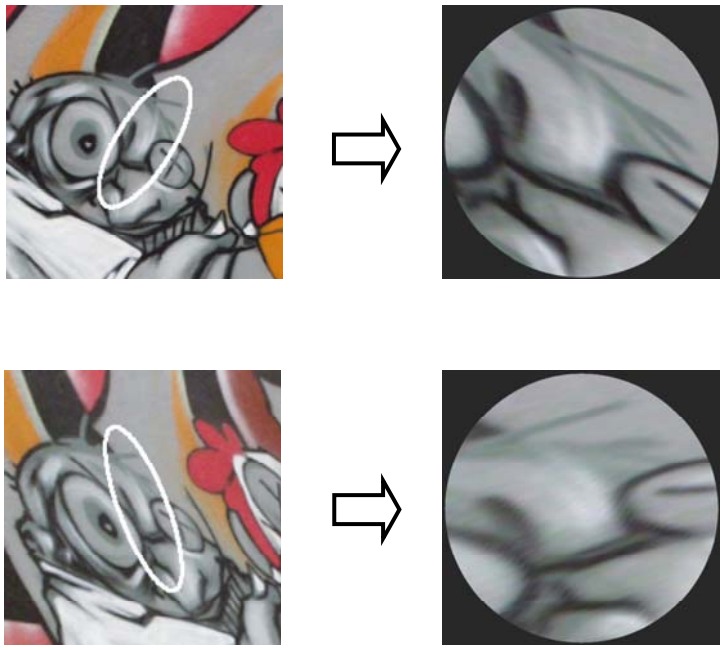
όπου η τετραγωνική ρίζα του V υπολογίζεται πολύ απλά ως εξής:

$$V = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \Rightarrow V^{1/2} = \begin{bmatrix} \lambda_1^{1/2} & 0 \\ 0 & \lambda_2^{1/2} \end{bmatrix} \qquad \textbf{(εξ. 2.33)}$$

Τέλος, επειδή η κανονικοποιημένη περιοχή αφορά μοναδιαίο κύκλο, όπως φαίνεται στο σχήμα 2.20, θα πρέπει να διευρυνθεί ώστε να έχει ακτίνα ίση με ρ, δηλαδή να περιέχεται σε τετράγωνο πλευράς 2ρ. Αυτό γίνεται απλά διαιρώντας τον πίνακα μετασχηματισμού με ρ, οπότε η τελική εξίσωση του κύκλου ακτίνας ρ προκύπτει:

$$u^T \cdot u = \rho^2 \qquad \textbf{(εξ. 2.34)}$$

Στην περίπτωση των πειραμάτων που διενεργούνται σε όλη την υπόλοιπη εργασία, οι τιμές για τις παραμέτρους είναι $\kappa = 3$ και $\rho = 20.5$, το οποίο σημαίνει ότι ως περιοχή μέτρησης λαμβάνεται η τριπλάσια της αρχικής (9 φορές μεγαλύτερη σε εμβαδόν) και η κανονικοποιημένη περιοχή ενδιαφέροντος είναι κύκλος που βρίσκεται εντός τετραγώνου διαστάσεων 41×41 pixels. Όπως φαίνεται στο σχήμα 2.21 για την περίπτωση της ίδιας σκηνής υπό διαφορετικές οπτικές γωνίες, οι δύο ελλείψεις μετασχηματίζονται σε δύο αντίστοιχους κύκλους και το διάνυσμα των χαρακτηριστικών εξάγεται από την κανονικοποιημένη περιοχή. Για λόγους ευκρίνειας η κανονικοποιημένη περιοχή περιέχεται σε τετράγωνο πλευράς μεγαλύτερης από 41 pixels. Αν κατευθύνουμε κατάλληλα το σύστημα συντεταγμένων (με τον κατάλληλο πίνακα περιστροφής R), τα διανύσματα χαρακτηριστικών θα είναι παρόμοια για τις δύο περιοχές.



**Σχ. 2.21:** Κανονικοποίηση ελλειπτικών περιοχών σε κυκλικές για την εξαγωγή του περιγραφέα

### 2.3.3 Περιγραφέας SIFT

Στην ίδια δημοσίευση όπου ο Lowe διατύπωσε την μέθοδο DoG για την ανίχνευση σημείων ενδιαφέροντος [Lowe, 1999], κατασκεύασε και ένα νέο τοπικό περιγραφέα, ο οποίος είχε πολύ καλές επιδόσεις σε ποικίλα πειράματα ταιριάσματος εικόνων. Ο περιγραφέας SIFT αποτελείται από ένα τρισδιάστατο ιστόγραμμα της θέσης και του προσανατολισμού των παραγώγων της εικόνας, στην τοπική γειτονιά γύρω από το σημείο ενδιαφέροντος (local image patch). Από τότε που πρωτοδιατυπώθηκε, έχει επιφέρει σημαντική επιρροή στην εξέλιξη των πεδίων αναγνώρισης αντικειμένων και αναζήτησης εικόνων και είναι πλέον ο πιο συχνά χρησιμοποιούμενος περιγραφέας. Η διαδικασία που ακολουθείται χωρίζεται σε δύο στάδια: στην εκτίμηση του προσανατολισμού (orientation) της γειτονιάς ενδιαφέροντος και στην κατασκευή του περιγραφέα για το συγκεκριμένο σημείο (descriptor), τα οποία και αναλύονται παρακάτω.

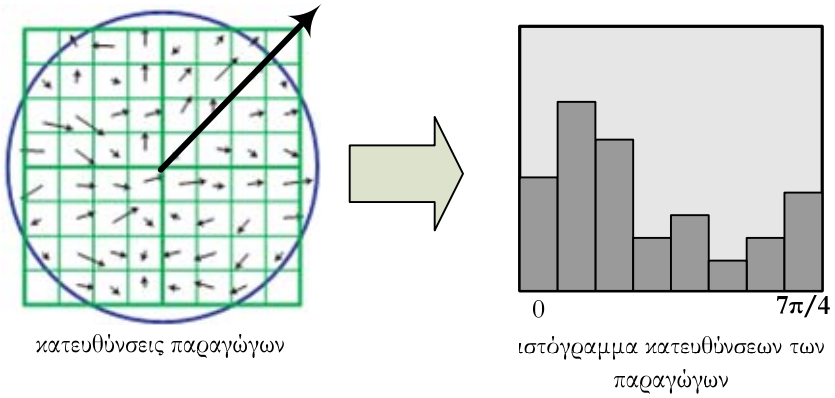
Επειδή μία κυκλική εν γένει περιοχή της εικόνας μπορεί να διαφέρει από μια άλλη κυκλική περιοχή κατά ένα μετασχηματισμό περιστροφής, αλλά να παριστάνει το ίδιο περιεχόμενο με αυτή (βλέπε σχήμα 2.21), είναι επιθυμητό να συσχετιστεί κάθε περιοχή με μία συγκεκριμένη κατεύθυνση σύμφωνα με την τοπική της δομή. Τότε το διάνυσμα χαρακτηριστικών μπορεί να κατασκευαστεί λαμβάνοντας υπ' όψη την κύρια αυτή κατεύθυνση και έτσι η περιγραφή κάθε περιοχής να είναι ανεξάρτητη από την τυχαία περιστροφή της (rotation invariant).

Χρησιμοποιώντας τη χαρακτηριστική κλίμακα s που έχει υπολογιστεί στο στάδιο του ανιχνευτή, λαμβάνεται η ομαλοποιημένη εικόνα L μέσω της εξίσωσης 2.12 (για την πιο κοντινή κλίμακα) και όλοι οι υπολογισμοί γίνονται ανεξάρτητοι από αλλαγές κλίμακας. Για κάθε σημείο (x,y) της γειτονιάς γύρω από το σημείο ενδιαφέροντος (image patch) υπολογίζεται το μέτρο των πρώτων παραγώγων m (magnitude) και ο προσανατολισμός τους θ (orientation) μέσω των σχέσεων [Lowe, 2004]:

$$m(x,y) = \sqrt{(L(x+1,y) - L(x-1,y))^2 + (L(x,y+1) - L(x,y-1))^2} \qquad \textbf{(εξ. 2.35)}$$

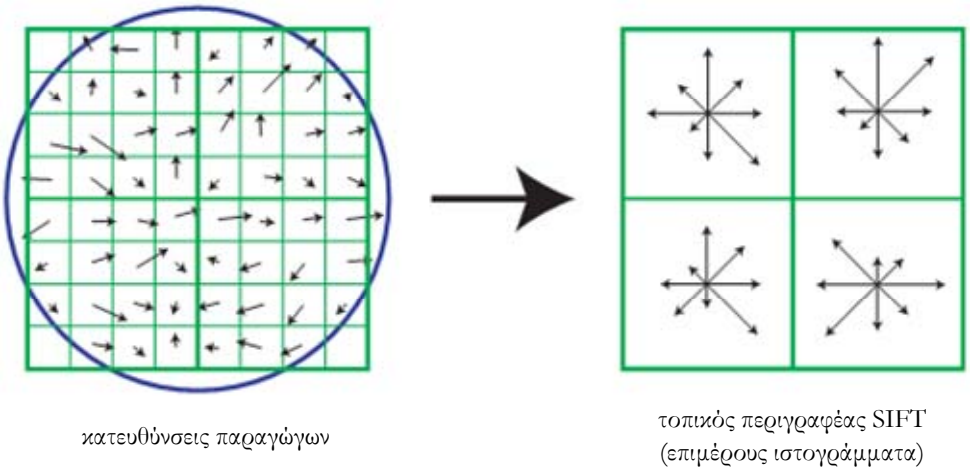
$$\theta(x,y) = \tan^{-1}\frac{L(x,y+1) - L(x,y-1)}{L(x+1,y) - L(x-1,y)} \qquad \textbf{(εξ. 2.36)}$$

Στη συνέχεια κατασκευάζεται ένα ιστόγραμμα των κατευθύνσεων των παραγώγων (orientation histogram) για αυτό το patch με χρησιμοποιώντας 36 στοιχεία (bins) τα οποία καλύπτουν συνολικά τις $360^o$ (ανά 10 μοίρες). Για την κατασκευή του ιστογράμματος χρησιμοποιούνται οι τιμές προσανατολισμού θ για όλα τα σημεία του patch, αλλά το κάθε σημείο προστίθεται στο κατάλληλο στοιχείο του ιστογράμματος με βάρος την τιμή του m που του αντιστοιχεί πολλαπλασιασμένη με την τιμή γκαουσιανής που έχει παράμετρο σ ίση 1.5×s, όπου s η χαρακτηριστική κλίμακα. Με αυτόν τον τρόπο τα σημεία που είναι πιο κοντά στο κέντρο ή εκείνα που έχουν μεγάλες τιμές παραγώγων συμβάλλουν περισσότερο στον προσδιορισμό της κύριας κατεύθυνσης (dominant orientation). Ένα παράδειγμα ιστογράμματος κατευθύνσεων των παραγώγων με 7 bins φαίνεται στο σχήμα 2.22 για το συγκεκριμένο patch. Εκτός από την κύρια κατεύθυνση, θα μπορούσαμε να κρατήσουμε και δευτερεύουσες κατευθύνσεις, εκεί όπου το ιστόγραμμα έχει τιμή πάνω από το 80% της μέγιστης τιμής (peak). Με τον τρόπο αυτό σε μερικά σημεία ενδιαφέροντος αντιστοιχούν περισσότερες βασικές κατευθύνσεις και έτσι εξάγονται περισσότεροι του ενός περιγραφείς. Αυτό πολλές φορές βελτιώνει την απόδοση στο στάδιο του ταιριάσματος εικόνων, αλλά δεν θα χρησιμοποιηθεί στα πειράματα του επόμενου κεφαλαίου.




**Σχ. 2.22:** Ιστόγραμμα των κατευθύνσεων των παραγώγων και κύρια κατεύθυνση

Το δεύτερο βήμα στη διαδικασία είναι η κατασκευή του διανύσματος χαρακτηριστικών από την περιοχή που "κεντράρεται" στο σημείο ενδιαφέροντος και προσανατολίζεται με βάση την κύρια κατεύθυνση. Στο σημείο αυτό αρκεί να γίνει η σύμβαση ότι ο ημιάξονας που ορίζεται από την κύρια κατεύθυνση θα είναι ο οριζόντιος θετικός ημιάξονας για το σύστημα συντεταγμένων στο οποίο θα υπολογιστούν τα στοιχεία του περιγραφέα. Στην ομαλοποιημένη εικόνα L της χαρακτηριστικής κλίμακας έχουν ήδη υπολογιστεί τα μέτρα και οι προσανατολισμοί των παραγώγων από τις εξισώσεις 2.35 και 2.36, για μία τετραγωνική περιοχή 16×16 pixels.

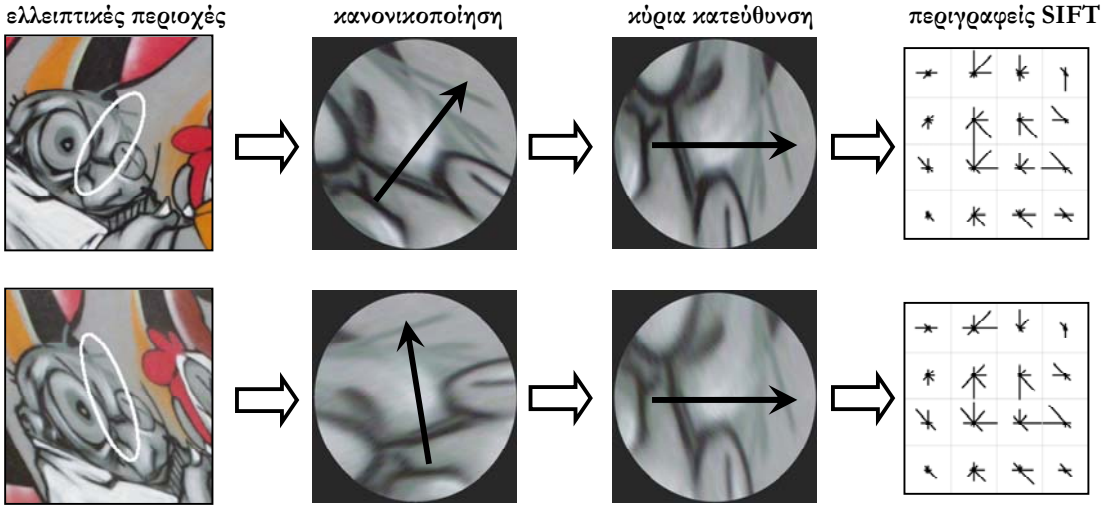



**Σχ. 2.23:** Κατασκευή του διανύσματος χαρακτηριστικών από ένα δείγμα 8×8 pixels (4 υποπεριοχές)
Ο περιγραφέας έχει διάσταση 32 σε αυτή την περίπτωση ([Lowe, 2004])

Στα μέτρα των παραγώγων τοποθετούνται βάρη με μία διδιάστατη γκαουσιανή που έχει παράμετρο σ ίση με το μισό του εύρους του τετραγωνικού παραθύρου, όπως φαίνεται από τον κύκλο τους σχήματος 2.23. Στη συνέχεια η περιοχή γύρω από το σημείο ενδιαφέροντος χωρίζεται σε 4×4 υποπεριοχές (καθεμία από τις οποίες αποτελείται από ένα μικρότερο τετράγωνο 4×4 pixels) και για κάθε τέτοια περιοχή δημιουργείται ένα ιστόγραμμα κατευθύνσεων με 8 στοιχεία [Lowe, 2004]. Στο σχήμα 2.23 απεικονίζονται τα στοιχεία αυτών των ιστογραμμάτων με βέλη, μόνο που για μεγαλύτερη σαφήνεια έχει σχεδιαστεί ένα κομμάτι με 2×2 υποπεριοχές (συνολικά 8×8 pixels). Ο τρόπος κατασκευής των ιστογραμμάτων αυτών επιτρέπουν

την μικρή μετατόπιση των παραγώγων χωρίς να επηρεάζεται το ιστόγραμμά τους. Το τελικό διάνυσμα περιγραφής του σημείου ενδιαφέροντος αποτελείται από τις τιμές των ιστογραμμάτων όλων των υποπεριοχών του τετραγώνου, δηλαδή αποτελείται από 4×4×8 = 128 στοιχεία. Έτσι κάθε σημείο απεικονίζεται στον χώρο των χαρακτηριστικών (128 διαστάσεων) με ένα συγκεκριμένο διάνυσμα, που προκύπτει με τον περιγραφέα SIFT (scale invariant feature transform). Σημεία ενδιαφέροντος που έχουν το ίδιο σημασιολογικό περιεχόμενο αναπαρίστανται από παρόμοια διανύσματα χαρακτηριστικών.

Το διάνυσμα κανονικοποείται ώστε να έχει μοναδιαίο μήκος. Μία αλλαγή στην αντίθεση είναι ένας πολλαπλασιασμός των τιμών των εντάσεων στα τα pixels της εικόνας με μια σταθερά, άρα και οι παράγωγοι πολλαπλασιάζονται με την ίδια σταθερά και έτσι η επίδραση της αλλαγής εξαλείφεται με την κανονικοποίηση του διανύσματος. Επίσης, μια αλλαγή στη φωτεινότητα, που είναι η πρόσθεση μιας σταθεράς στα pixels της εικόνας, δεν επηρεάζει καθόλου το ιστόγραμμα, αφού οι παράγωγοι υπολογίζονται ως διαφορές. Δηλαδή γενικά οι τιμές του διανύσματος δεν επηρεάζονται από αφινικές μεταβολές της φωτεινότητας (δηλαδή μεταβολές της μορφής I' = α * I + β).

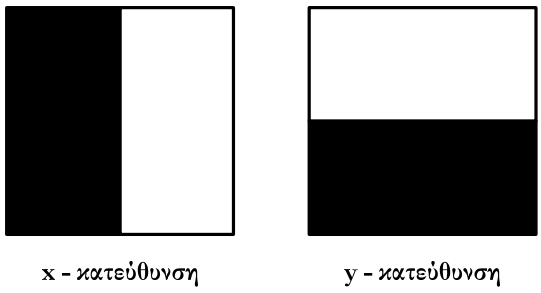



**Σχ. 2.24:** Ελλειπτικές (αφινικές) περιοχές, κανονικοποίηση, κύρια κατεύθυνση και εξαγωγή του περιγραφέα SIFT (4×4×8=128 bins)

### 2.3.4 Περιγραφέας SURF

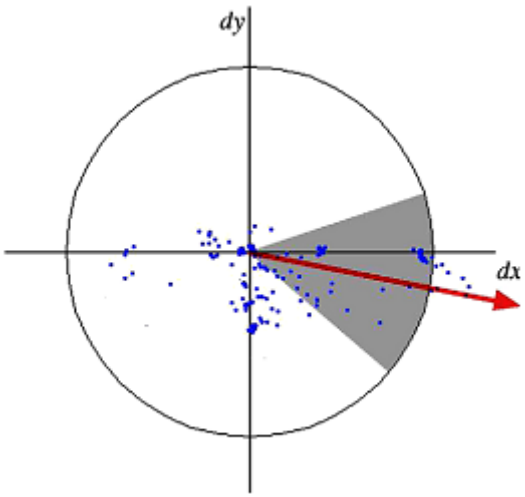
Δεδομένης της επιτυχίας του περιγραφέα SIFT, της διακριτικής του ικανότητας και των πολύ καλών επιδόσεων που έδωσε πειραματικά [Mikolajczyk et al., 2003], [Mikolajczyk et al., 2005a] σε σχέση με τις υπόλοιπες τεχνικές, οι συγγραφείς Bay et al. που πρότειναν την ανίχνευση Fast Hessian διατυπώνουν και ένα νέο περιγραφέα, ο οποίος βασίζεται σε παρόμοιες τοπικές ιδιότητες με τον SIFT αλλά παράγεται με πολύ πιο γρήγορους υπολογισμούς και ονομάζεται SURF (speeded-up robust features). Η διαδικασία χωρίζεται και εδώ σε δύο στάδια, στην εκτίμηση της κύριας κατεύθυνσης (προσανατολισμού) από μια κυκλική γειτονιά γύρω από το σημείο ενδιαφέροντος και στην κατασκευή του διανύσματος χαρακτηριστικών από ένα τετράγωνο προσανατολισμένο κατά την κύρια κατεύθυνση.

Για τον υπολογισμό της κύριας κατεύθυνσης χρησιμοποιούνται οι αποκρίσεις των φίλτρων Haar (Haar wavelets) σε δύο κατευθύνσεις x και y σε μία κυκλική γειτονιά του σημείου ενδιαφέροντος ακτίνας 6×s (s η χαρακτηριστική κλίμακα του σημείου). Στην κατεύθυνση της κλίμακας το βήμα δειγματοληψίας είναι ίσο με s, ενώ οι πλευρές των τετραγωνικών φίλτρων (κυματιδίων Haar) είναι ίσες με 4×s. Τα κυματίδια Haar που χρησιμοποιούνται δίνονται στο σχήμα 2.25. Για τον γρήγορο υπολογισμό των αποκρίσεων των φίλτρων χρησιμοποιούνται integral images (όπως στην περίπτωση του ανιχνευτή Fast Hessian), οπότε απαιτούνται μόλις έξι πράξεις για να υπολογιστεί η απόκριση σε οποιοδήποτε σημείο.



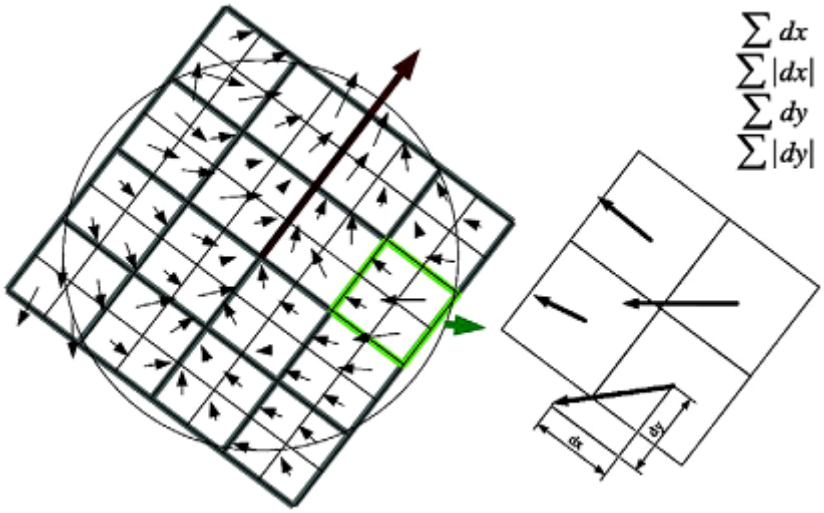

**Σχ. 2.25:** Φίλτρα Haar wavelets που χρησιμοποιούνται στον περιγραφέα SURF (κατευθύνσεις x, y)

Μόλις υπολογιστούν οι αποκρίσεις Haar, προστίθενται βάρη με μία γκαουσιανή συνάρτηση με $\sigma = 2\times s$ και απεικονίζονται σε έναν χώρο δύο διαστάσεων με οριζόντιο άξονα την απόκριση dx (για την κατεύθυνση x) και κατακόρυφο άξονα την απόκριση dy (για την κατεύθυνση y), όπως απεικονίζεται στο σχήμα 2.26. Στο διάγραμμα αυτό υπολογίζεται το άθροισμα όλων αποκρίσεων (x και y) των σημείων εντός μια γωνίας εύρους $\pi/3$ η οποία περιστρέφεται περί της αρχής των αξόνων, και έτσι προκύπτει ένα διάνυσμα προσανατολισμού με μέτρο το άθροισμα αυτό και κατεύθυνση τη διχοτόμο της κυλιόμενης γωνίας. Ως κύρια κατεύθυνση της περιοχής του σημείου ενδιαφέροντος λαμβάνεται η κατεύθυνση του μέγιστου διανύσματος προσανατολισμού [Bay et al., 2008].




**Σχ. 2.26:** Υπολογισμός της κύριας κατεύθυνσης για το σημείο ενδιαφέροντος ([Bay et al., 2008])

Στη συνέχεια, για την εξαγωγή του διανύσματος χαρακτηριστικών, κατασκευάζουμε μία τετραγωνική γειτονιά γύρω από το σημείο ενδιαφέροντος, προσανατολισμένη κατά την κύρια κατεύθυνση που μόλις υπολογίστηκε. Η πλευρά του τετραγωνικού παραθύρου είναι ίση με 20×s. Χωρίζοντας το τετράγωνο αυτό σε 4×4 υποπεριοχές, υπολογίζουμε τα χαρακτηριστικά σε όλα τα σημεία κάθε τέτοιας περιοχής, και συγκεκριμένα σε 5×5 σημεία δειγματοληψίας (συμμετρικά διατεταγμένα). Τα χαρακτηριστικά που υπολογίζονται είναι οι αποκρίσεις των Haar κυματιδίων (πλευράς 2×s), τις οποίες συμβολίζουμε με dx για την οριζόντια διεύθυνση και με dy για την κατακόρυφη διεύθυνση. Στο σχήμα 2.27 παρουσιάζεται ο τρόπος υπολογισμού των αποκρίσεων, με 2×2 σημεία δειγματοληψίας για λόγους ευκρίνειας.



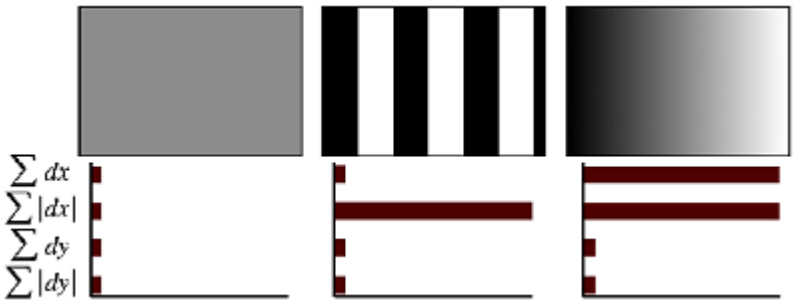

**Σχ. 2.27:** Εξαγωγή του διανύσματος χαρακτηριστικών (4 στοιχεία) από κάθε υποπεριοχή του παραθύρου γύρω από το σημείο ενδιαφέροντος ([Bay et al., 2008])

Οι αποκρίσεις αυτές πολλαπλασιάζονται με κατάλληλα βάρη (που είναι οι τιμές γκαουσιανής με σ = 3.3×s), ούτως ώστε να αντιμετωπίζονται μικρές γεωμετρικές μεταβολές, και στη συνέχεια αθροίζονται σε καθεμία υποπεριοχή. Εκτός από τις αποκρίσεις στις δύο κατευθύνσεις (dx και dy), χρησιμοποιούνται και τα αθροίσματα των απολύτων τιμών τους |dx|

και |dy|, τεχνική που επιτρέπει να διατηρηθεί περισσότερη πληροφορία σχετικά με τις μεταβολές της έντασης (για μια εξήγηση της χρησιμότητας των όρων αυτών βλέπε σχήμα 2.28). Έτσι για κάθε υποπεριοχή προκύπτει ένα διάνυσμα τεσσάρων στοιχείων (τοπικών αθροισμάτων):

$$v = \left( \sum dx, \sum dy, \sum |dx|, \sum |dy| \right) \qquad \textbf{(εξ. 2.37)}$$

το οποίο περιγράφει την τοπική δομή των μεταβολών της έντασης, όπως φαίνεται στο σχήμα 2.27. Οπότε για το σημείο ενδιαφέροντος το διάνυσμα χαρακτηριστικών SURF, που περιλαμβάνει τα διανύσματα χαρακτηριστικών v από όλες τις υποπεριοχές της γειτονιάς του, θα έχει διάσταση 4×4×4 = 64. Οι αποκρίσεις των φίλτρων Haar δεν επηρεάζονται από μεταβολές της φωτεινότητας, ενώ για να είναι η περιγραφή SURF ανεξάρτητη και από αλλαγές στην αντίθεση, το διάνυσμα κανονικοποιείται ώστε να έχει μοναδιαίο μέτρο.




**Σχ. 2.28:** Παρουσίαση της χρησιμότητας των στοιχείων του διανύσματος SURF ([Bay et al., 2008])

# ΚΕΦΑΛΑΙΟ 3

# Πειράματα και Σύγκριση των Μεθόδων

## 3.1 Εισαγωγικά

### 3.1.1 Περιγραφή των πειραματικών δεδομένων

Οι μέθοδοι τοπικών χαρακτηριστικών που αναλύθηκαν στο προηγούμενο κεφάλαιο συγκρίνονται ως προς την αποτελεσματικότητά τους μέσω μίας συγκεκριμένης πειραματικής διαδικασίας, η οποία στη βιβλιογραφία θεωρείται συχνά σημείο αναφοράς για τις μετρήσεις των επιδόσεων τέτοιων τεχνικών. Για όλα τα πειράματα του παρόντος κεφαλαίου χρησιμοποιείται ένα σύνολο από 48 εικόνες, οι οποίες ανά 6 αναπαριστούν την ίδια σκηνή ή το ίδιο αντικείμενο, υπό διαφορετικές όμως συνθήκες λήψης ή αποθήκευσης της εικόνας. Κάθε τέτοια ακολουθία από τις 6 εικόνες περιλαμβάνει διαδοχικές φωτομετρικές ή γεωμετρικές μεταβολές [VGG dataset].

Σχετικά με τις δυνατές μεταβολές που συμβαίνουν συνήθως σε μια εικόνα, μελετώνται πέντε διαφορετικοί τύποι μετασχηματισμών: αλλαγή στην οπτική γωνία (viewpoint change), αλλαγή κλίμακας (scale change), θόλωμα της εικόνας (blur), αλλαγές στην φωτεινότητα (illumination) ή παραμορφώσεις που προκαλούνται από τη συμπίεση της εικόνας (για παράδειγμα συμπίεση JPEG). Όσον αφορά στους τρεις πρώτους τύπους που αναφέρθηκαν, για τον ίδιο μετασχηματισμό συμπεριλαμβάνονται δύο διαφορετικοί τύποι σκηνών στα πειραματικά δεδομένα. Ο ένας τύπος αποτελείται από εικόνες με ομοιογενείς περιοχές οι οποίες έχουν ευδιάκριτα όρια μεταξύ τους (ακμές). Τέτοιες είναι για παράδειγμα οι εικόνες που περιέχουν κτίρια, γκράφιτι, κτλ. και θα αναφέρονται από εδώ και στο εξής ως εικόνες με δομή (structured scene). Ο δεύτερος τύπος περιλαμβάνει εικόνες με επαναλαμβανόμενα μοτίβα, δηλαδή περιοχές που έχουν μια συγκεκριμένη υφή, χωρίς όμως να έχουν συγκεκριμένα όρια που να τις διαχωρίζουν (όπως για παράδειγμα εικόνες με ύφασμα, γρασίδι, δέντρα), και στο εξής θα ονομάζονται εικόνες υφής (textured scene). Με αυτόν τον διαχωρισμό των εικόνων σε δύο κατηγορίες με βάση τη μορφή τους είναι δυνατή η σύγκριση των μεθόδων και ως προς τον τύπο της σκηνής αλλά και ως προς το είδος μετασχηματισμού της εικόνας.

Για να επιτευχθούν οι αλλαγές στην οπτική γωνία, η μηχανή λήψης στρέφεται ως προς τον κατακόρυφο άξονά της, οπότε από τον αρχικό προσανατολισμό με τον φακό παράλληλο στο επίπεδο της σκηνής, διαδοχικά και με βήμα $20^o$, η γωνία λήψης (ορίζεται ως γωνία του άξονα της κάμερας) μεταβάλλεται μέχρι και τις $60^o$. Οι δύο ακολουθίες εικόνων που παρουσιάζουν τέτοιο

μετασχηματισμό είναι η ακολουθία Graffiti και η ακολουθία Wall, που κατατάσσονται στις σκηνές δομής και υφής αντίστοιχα. Η αλλαγή στην κλίμακα προκύπτει με μεταβολή του zoom της μηχανής, οπότε προκύπτουν οι ακολουθίες Boat και Bark (με δομή και με υφή αντίστοιχα), που περιλαμβάνουν μεγέθυνση των αντικειμένων τους έως και 4 φορές επί το αρχικό μέγεθος, αλλά επίσης και περιστροφή γύρω από τον κάθετο στο επίπεδο της εικόνας άξονα. Για το θόλωμα της εικόνας χρησιμοποιείται η δυνατότητα χειροκίνητης εστίασης (focus), οπότε οι αντίστοιχες εικόνες με διάφορα επίπεδα θόλωσης περιλαμβάνονται στα σύνολα Bikes και Trees, που περιέχουν σκηνές με δομή και με υφή (όπως είναι φανερό). Τέλος, η ακολουθία Leuven περιέχει εικόνες με μεταβαλλόμενη φωτεινότητα (μέσω του διαφράγματος) και η ακολουθία UBC περιέχει εικόνες που έχουν υποστεί συμπίεση JPEG, με παράμετρο ποιότητας από 40% έως και 2%. Όλες οι παραπάνω φωτογραφίες είναι ενδιάμεσου μεγέθους, από 800×640 pixels έως το πολύ 1000×700 pixels. Μερικές από τις εικόνες των πειραματικών δεδομένων φαίνονται στο σχήμα 3.1.

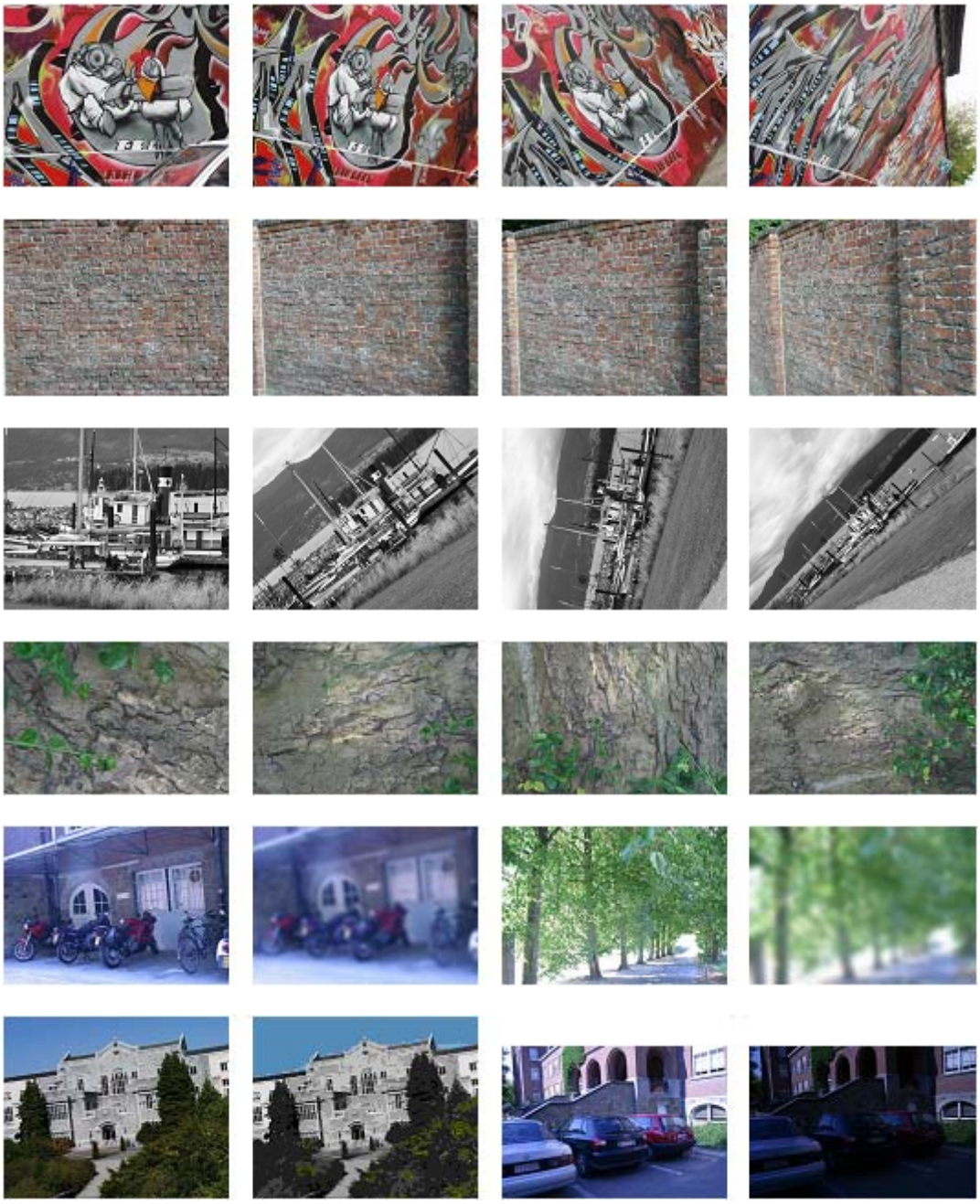
Οι εικόνες του συνόλου των πειραμάτων έχουν την εξής ιδιότητα: είτε αποτελούνται από επίπεδες σκηνές είτε η κάμερα παραμένει σταθερή κατά τις διαφορετικές λήψεις του ίδιου περιεχομένου. Αυτό σημαίνει ότι οι εικόνες που ανήκουν στην ίδια ακολουθία συνδέονται μεταξύ τους με κάποια ομογραφία, δηλαδή με έναν μετασχηματισμό επίπεδης προβολής (που απεικονίζει σημεία σε σημεία και ευθείες σε ευθείες). Υπολογίζοντας τον μετασχηματισμό αυτόν μπορούμε να κατασκευάσουμε το σύνολο αληθείας (ground truth), δηλαδή την αντιστοιχία των σημείων μεταξύ δύο εικόνων. Ο 3×3 πίνακας της ομογραφίας H χρησιμοποιείται για την απεικόνιση σημείου x της δεύτερης εικόνας σε σημείο x' της πρώτης μέσω της σχέσης:

$$x' = H \cdot x \qquad \textbf{(εξ. 3.1)}$$

Οι ομογραφίες που θα υπολογιστούν θα αναφέρονται πάντοτε στην απεικόνιση σημείων κάθε εικόνας σε σημεία της πρώτης εικόνας της ακολουθίας, που θα ονομάζεται από εδώ και στο εξής ως εικόνα αναφοράς. Ο υπολογισμός τους γίνεται σε δύο βήματα: πρώτα υπολογίζεται μία προσεγγιστική ομογραφία μέσω σημείων που επιλέγονται χειροκίνητα. Στη συνέχεια η μία εικόνα μετασχηματίζεται μέσω της προσέγγισης αυτής και με κάποιον κατάλληλο αλγόριθμο για μικρές μετατοπίσεις εκτιμάται η ακριβής ομογραφία που απομένει, ούτως ώστε τελικά το σφάλμα υπολογισμού μεταξύ των δύο εικόνων να είναι μικρότερο από 1 pixel. Με τον τρόπο αυτόν, ανεξάρτητα από τις μεθόδους ανίχνευσης σημείων, προκύπτουν οι πίνακες H για όλες τις εικόνες και χρησιμοποιούνται ως σύνολο αληθείας [VGG dataset]. Στον πίνακα 3.1 δίνεται το είδος του μετασχηματισμού και οι διαστάσεις για τις εικόνες του κάθε υποσυνόλου του dataset.

**Πίνακας 3.1:** Διαστάσεις και μετασχηματισμός για κάθε ακολουθία εικόνων

| ΑΚΟΛΟΥΘΙΑ ΕΙΚΟΝΩΝ | ΔΙΑΣΤΑΣΕΙΣ | ΕΙΔΟΣ ΜΕΤΑΣΧΗΜΑΤΙΣΜΟΥ |
|---|---|---|
| **BARK** | 765×512 | zoom + rotation |
| **BIKES** | 1000×700 | increasing blur |
| **BOAT** | 850×680 | zoom + rotation |
| **GRAF** | 800×640 | viewpoint change |
| **LEUVEN** | 900×600 | decreasing light |
| **TREES** | 1000×700 | increasing blur |
| **UBC** | 800×640 | JPEG compression |
| **WALL** | 880×680 | viewpoint change |



**Σχ. 3.1:** Μερικές εικόνες από το σύνολο που χρησιμοποιείται στα πειράματα. Στην πρώτη και δεύτερη γραμμή περιλαμβάνονται εικόνες από την ακολουθία Graffiti και την ακολουθία Wall (αλλαγές στην οπτική γωνία). Οι δύο επόμενες γραμμές περιλαμβάνουν περιστροφή και αλλαγή κλίμακας (ακολουθίες Boat και Bark), στην πέμπτη γραμμή υπάρχει σταδιακό θόλωμα των εικόνων (Bikes και Trees) και στο τέλος συμπίεση JPEG και μεταβολές φωτεινότητας (ακολουθίες UBC και Leuven αντίστοιχα).

### 3.1.2 Παρατηρήσεις σχετικά με την πειραματική διαδικασία

Όλες οι μέθοδοι ανίχνευσης σημείων και περιοχών ενδιαφέροντος που αναπτύχθηκαν στο προηγούμενο κεφάλαιο θα συγκριθούν με κοινά κριτήρια, και μάλιστα θα συνδυαστούν με τους δύο περιγραφείς SIFT και SURF, ώστε να εντοπιστεί ο σωστός συνδυασμός σε κάθε περίπτωση.

Πρέπει στο σημείο αυτό να σημειώσουμε ότι για τις παραμέτρους των μεθόδων χρησιμοποιήθηκαν κοινές τιμές σε ολόκληρο το σύνολο των εικόνων. Παρόλ' αυτά ο αριθμός των περιοχών ενδιαφέροντος που εξάγει η κάθε μέθοδος στις επιμέρους σκηνές κυμαίνεται μεταξύ διαφορετικών επιπέδων, είτε όσον αφορά τα σημεία του ίδιου ανιχνευτή για διαφορετικούς τύπους εικόνων, είτε όσον αφορά τη συμπεριφορά των διαφόρων ανιχνευτών για μία συγκεκριμένη σκηνή. Η πρώτη περίπτωση αποδεικνύει ότι το αποτέλεσμα των τοπικών χαρακτηριστικών συνήθως εξαρτάται από το περιεχόμενο της εικόνας και η δεύτερη επιβεβαιώνει το γεγονός ότι οι μέθοδοι ανίχνευσης αποκρίνονται σε διαφορετικά χαρακτηριστικά των εικόνων (γωνίες, κηλίδες, ακμές, ευδιάκριτες περιοχές).

**Πίνακας 3.2:** Μέσος αριθμός σημείων-περιοχών ενδιαφέροντος ανά εικόνα για κάθε μέθοδο. Πρώτα δίνεται ο αριθμός των σημείων για κάθε είδος εικόνας ξεχωριστά και στη συνέχεια ο μέσος όρος για ολόκληρο το σύνολο δεδομένων.

| **ΑΚΟΛΟΥΘΙΑ ΕΙΚΟΝΩΝ** | **ΜΕΘΟΔΟΙ ΑΝΙΧΝΕΥΣΗΣ** | | | | |
|---|---|---|---|---|---|
| | **DOG** | **FASTH** | **MSER** | **HAR-AFF** | **HES-AFF** |
| **BARK** | 1632 | 787 | 703 | 233 | 202 |
| **BIKES** | 1034 | 591 | 268 | 773 | 624 |
| **BOAT** | 1079 | 1484 | 2327 | 2704 | 2710 |
| **GRAF** | 1167 | 1721 | 799 | 2788 | 2585 |
| **LEUVEN** | 645 | 680 | 813 | 925 | 866 |
| **TREES** | 3123 | 2816 | 2820 | 5649 | 4844 |
| **UBC** | 1224 | 1264 | 1799 | 2127 | 2130 |
| **WALL** | 1950 | 2005 | 3085 | 1536 | 1153 |
| **ΜΕΣΟΣ ΑΡΙΘΜΟΣ ΣΗΜΕΙΩΝ** | **1482** | **1419** | **1577** | **2092** | **1889** |

Στον πίνακα 3.2 παρουσιάζεται το πλήθος των σημείων που εντοπίζει η κάθε μέθοδος για κάθε ακολουθία εικόνων ξεχωριστά, αλλά και για ολόκληρο το σύνολο των δεδομένων. Σε κάθε περίπτωση ο αριθμός των σημείων αντιστοιχεί σε ένα μέσο όρο που υπολογίζεται είτε για τις 6 εικόνες κάθε σκηνής, είτε για τις 48 συνολικά εικόνες. Οι συντομογραφίες των μεθόδων είναι προφανείς από τις ονομασίες των ανιχνευτών (Difference Of Gaussians, Fast Hessian, MSER, Harris Affine, Hessian Affine). Όπως βλέπουμε, η μέθοδος DOG εξάγει λίγο περισσότερα από 1000 σημεία ανά εικόνα, με εξαίρεση τις εικόνες Leuven όπου είναι περίπου τα μισά και την ακολουθία Trees όπου είναι τριπλάσια. Η μέθοδος FASTH εξάγει σε τρεις σκηνές έως 1000 σημεία, φτάνοντας σχεδόν τα διπλά στις υπόλοιπες (και σχεδόν 3000 στην Trees). Η μέθοδος MSER παράγει την ίδια κατανομή σημείων, με λίγα σημεία στις μισές σκηνές και πολλά στις υπόλοιπες. Οι δύο τελευταίες μέθοδοι, παρόλο που ακολουθούν παρόμοια συμπεριφορά με τις προηγούμενες, εμφανίζουν πολύ περισσότερα σημεία, κυρίως στην ακολουθία Trees. Παρατηρώντας συνολικά τον πίνακα, βλέπουμε ότι η μέθοδος DOG παράγει συνήθως 1000 με 2000 σημεία ανεξάρτητα από το είδος της εικόνας, ενώ οι υπόλοιπες μέθοδοι εντοπίζουν λίγα σημεία για κάποιες συγκεκριμένες ακολουθίες (Bark, Bikes, Leuven), γεγονός που εξηγείται από το περιεχόμενο των σκηνών. Οι τρεις πρώτες παρουσιάζουν τον μεγαλύτερο αριθμό σημείων για σκηνές textured με υφή (Trees, Wall), ενώ οι δύο τελευταίες έχουν γενικά αποτέλεσμα πολλά

σημεία, ειδικότερα δε σε structured σκηνές με δομή (Boat, Graf) αλλά και στην ακολουθία Trees.

Στη συνέχεια επιχειρείται μια σύγκριση του χρόνου υπολογισμού που απαιτείται από κάθε μέθοδο για την ανίχνευση των σημείων – περιοχών ενδιαφέροντος. Στον πίνακα 3.3 φαίνονται οι μέσοι χρόνοι υπολογισμού σε milliseconds, όπως υπολογίστηκαν σε ένα φορητό υπολογιστή με επεξεργαστή Intel Core 2 Duo 2.2GHz και μνήμη 2048MB. Σε κάθε ακολουθία εικόνων αναγράφεται ο μέσος χρόνος υπολογισμού για τα σημεία μιας από τις 6 εικόνες, ενώ στην τελευταία γραμμή του πίνακα παρουσιάζεται ο μέσος χρόνος κάθε μεθόδου υπολογισμένος σε ολόκληρο το πειραματικό σύνολο. Από εδώ γίνεται εμφανής η ταχύτητα της μεθόδου FASTH, η οποία είναι σχεδόν 4 φορές γρηγορότερη από την αμέσως επόμενη DOG, ενώ ακολουθεί η μέθοδος MSER με τον ικανοποιητικό χρόνο περίπου 1 sec ανά εικόνα. Τέλος, οι μέθοδοι HAR-AFF και HES-AFF χρειάζονται σχεδόν 3 δευτερόλεπτα για να παράγουν τα σημεία ενδιαφέροντος, γεγονός που εξηγείται από την αυξημένη πολυπλοκότητα της διαδικασίας επιλογής χαρακτηριστικής κλίμακας και αφινικής προσαρμογής. Οι χρόνοι υπολογισμού επιβεβαιώνουν το πλεονέκτημα της προσέγγισης του χώρου κλίμακας λαπλασιανών με διαφορές γκαουσιανών επιπέδων (DOG scale space) αλλά και την σημαντική βελτίωση στην ταχύτητα με τη χρήση των box filters (Fast Hessian).

**Πίνακας 3.3:** Μέσοι χρόνοι υπολογισμού (ms) ανά εικόνα για όλες τις μεθόδους

| **ΑΚΟΛΟΥΘΙΑ ΕΙΚΟΝΩΝ** | **ΜΕΘΟΔΟΙ ΑΝΙΧΝΕΥΣΗΣ** | | | | |
|---|---|---|---|---|---|
| | **DOG** | **FASTH** | **MSER** | **HAR-AFF** | **HES-AFF** |
| **BARK** | 614 | 152 | 659 | 780 | 675 |
| **BIKES** | 1042 | 196 | 981 | 1860 | 1569 |
| **BOAT** | 869 | 242 | 1228 | 3484 | 3453 |
| **GRAF** | 785 | 260 | 909 | 3387 | 3253 |
| **LEUVEN** | 857 | 170 | 842 | 1760 | 1617 |
| **TREES** | 1180 | 405 | 1678 | 6200 | 5427 |
| **UBC** | 825 | 217 | 835 | 2799 | 2730 |
| **WALL** | 1014 | 309 | 1556 | 2428 | 1954 |
| **ΜΕΣΟΣ ΧΡΟΝΟΣ** | **876 ms** | **242 ms** | **1085 ms** | **2813 ms** | **2566 ms** |

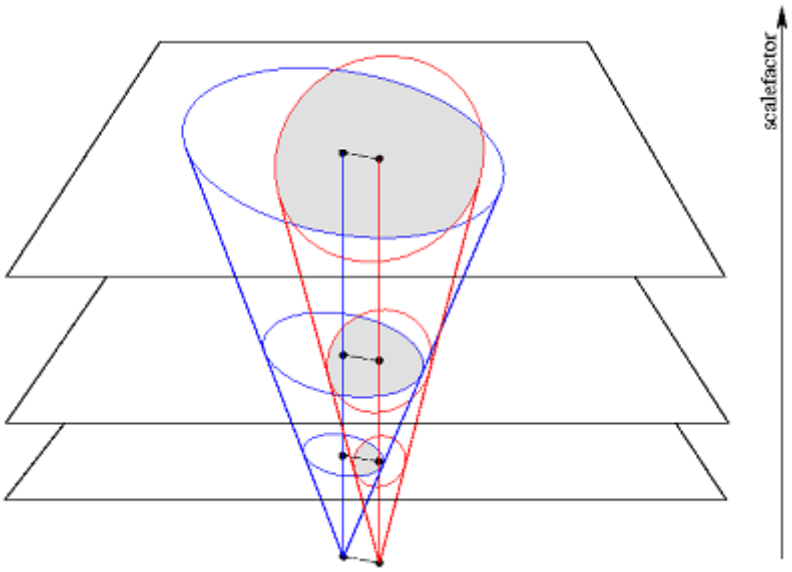
Τέλος, σχετικά με τον τρόπο εξαγωγής των περιγραφέων από τις ανιχνευθείσες περιοχές ενδιαφέροντος, θα πρέπει να σημειώσουμε μερικές λεπτομέρειες. Όπως ήδη έχει αναφερθεί, στην περίπτωση των ανιχνευτών σημείων ενδιαφέροντος Difference Of Gaussians και Fast Hessian, η περιοχή απ' όπου εξάγονται τα στοιχεία του διανύσματος χαρακτηριστικών είναι ένα τετραγωνικό χωρίο (local patch) προσανατολισμένο κατά την κύρια κατεύθυνση και με πλευρά ανάλογη της κλίμακας στην οποία εντοπίστηκε το σημείο. Στην περίπτωση των ανιχνευτών ανεξάρτητων από αφινικές μεταβολές, όπως είναι οι Harris-Hessian Affine και MSER, οι περιοχή ενδιαφέροντος καθορίζεται από μια έλλειψη. Αυτή θα ονομάζεται αρχική ή διακεκριμένη περιοχή (distinguished region). Από την περιοχή αυτή, με αφινικό τρόπο (πολλαπλασιασμό και των δύο αξόνων της έλλειψης με τον ίδιο παράγοντα), μπορεί να κατασκευαστεί μία άλλη μεγαλύτερη περιοχή (έλλειψη), γεγονός που συνήθως είναι επιθυμητό ώστε να περιληφθεί περισσότερη πληροφορία

που πιθανώς βρίσκεται έξω από την αρχική έλλειψη. Αυτό εξηγείται από το ότι η διαδικασία ανίχνευσης εντοπίζει περιοχές στα όρια των οποίων συμβαίνει αλλαγή στην ένταση (περιεχόμενο) της εικόνας. Γι' αυτό μεγεθύνοντας την περιοχή χρησιμοποιούμε αυτήν την πρόσθετη πληροφορία που βρίσκεται στη "γειτονιά" ενδιαφέροντος. Η νέα περιοχή θα ονομάζεται από εδώ και στο εξής περιοχή μέτρησης (measurement region). Στη συνέχεια, όπως είναι ήδη γνωστό, ακολουθείται μία κοινή διαδικασία κανονικοποίησης της περιοχής μέτρησης (normalization), οπότε τα σημεία της απεικονίζονται σε μία κυκλική περιοχή συγκεκριμένου μεγέθους (normalized region). Από την τελική αυτή περιοχή εξάγονται τα στοιχεία του περιγραφέα (ανεξαρτήτως της χρησιμοποιούμενης μεθόδου ανίχνευσης). Σε όλα τα πειράματα αυτού του κεφαλαίου, για την εξαγωγή του περιγραφέα χρησιμοποιείται περιοχή μέτρησης 3 φορές μεγαλύτερη της αρχικής περιοχής ενδιαφέροντος και κανονικοποιημένη κυκλική περιοχή που περιέχεται σε ένα τετραγωνικό χωρίο διαστάσεων 41×41 pixels [Ke et al., 2004], [Mikolajczyk et al., 2005a]. Οι διαστάσεις αυτές, εκτός του ότι συναντώνται συχνά στη βιβλιογραφία μεταξύ παρόμοιων πειραμάτων, αποδείχθηκε πειραματικά ότι δίνουν τις καλύτερες τιμές για τα μέτρο αξιολόγησης του ταιριάσματος μεταξύ των εικόνων (όπως αυτό ορίζεται στην επόμενη παράγραφο).

## 3.2 Κριτήρια αξιολόγησης των μεθόδων

Οι μέθοδοι τοπικών χαρακτηριστικών πρέπει να συγκριθούν με κάποιο αντικειμενικό κριτήριο, το οποίο μεταξύ δύο εικόνων της ίδιας σκηνής να δείχνει πόσο συχνά εντοπίζονται παρόμοιες περιοχές και πόσο κοντά βρίσκονται αυτές οι περιοχές στις "πραγματικές" περιοχές (στα ιδανικά τοπικά χαρακτηριστικά). Δύο είναι λοιπόν οι βασικοί παράγοντες που θα πρέπει να μελετήσουμε: η επαναληψιμότητα, δηλαδή το πλήθος των περιοχών που αντιστοιχούν στο ίδιο κομμάτι της εικόνας, και η ακρίβεια εντοπισμού, δηλαδή το πόσο καλά προσεγγίζεται η πραγματική περιοχή. Η μέτρησή τους γίνεται με το σχετικό ποσοστό επικάλυψης των περιοχών ανάμεσα σε δύο εικόνες της ίδιας σκηνής, που συνδέονται με κάποιο είδος μετασχηματισμού (φωτομετρικού ή γεωμετρικού). Όπως ήδη αναφέραμε, ως σύνολο αληθείας (ground truth) χρησιμοποιούνται οι ομογραφίες που συνδέουν τις εικόνες μιας ακολουθίας, από τις οποίες προκύπτουν οι πραγματικές περιοχές, αυτές δηλαδή που θα έπρεπε ιδανικά η μέθοδος να εντοπίσει. Οι περιοχές κάθε εικόνας μέσω της γνωστής ομογραφίας απεικονίζονται στην εικόνα αναφοράς (η πρώτη κάθε ακολουθίας), η οποία θεωρείται ότι περιέχει τις σωστές περιοχές που θα πρέπει να εντοπίζονται και στις υπόλοιπες μετασχηματισμένες εικόνες.

Καθώς η επικάλυψη μετράται με το εμβαδόν των δύο περιοχών πάνω στην εικόνα αναφοράς, είναι προφανές ότι οι ευρύτερες περιοχές θα έχουν και μεγαλύτερη πιθανότητα να αλληλοκαλύπτονται. Παίρνοντας λοιπόν μεγαλύτερες περιοχές μέτρησης, θα μπορούσαμε να αυξήσουμε το ποσοστό επικάλυψης. Αυτό εξηγείται στο σχήμα 3.2, στο οποίο φαίνονται διάφορες ελλειπτικές περιοχές μέτρησης που προκύπτουν από την αρχική έλλειψη με κάποια αλλαγή κλίμακας (μεγέθυνση), πάνω στο επίπεδο της εικόνας. Μεταβάλλοντας με συνεχή τρόπο την κλίμακα (κατακόρυφα), προκύπτει γεωμετρικά ένας κώνος στον τρισδιάστατο χώρο (παράλληλα επίπεδα της εικόνας), ο οποίος στην τομή του με κάποιο επίπεδο s δίνει την αντίστοιχη έλλειψη. Για τις δύο περιοχές (την πραγματική από την εικόνα αναφοράς και την μετασχηματισμένη από τη δεύτερη εικόνα μέσω ομογραφίας) προκύπτουν δύο κώνοι, οι οποίοι όπως φαίνεται δίνουν ελλείψεις που η τομή τους συνεχώς αυξάνεται με την κλίμακα, ενώ για μικρότερες κλίμακες είναι δυνατόν ακόμα και να μην τέμνονται. Οι κορυφές των κώνων αντιστοιχούν στις αποστάσεις μεταξύ των κέντρων των ελλείψεων, δηλαδή στο σφάλμα εντοπισμού της θέσης σου σημείου ενδιαφέροντος.




**Σχ. 3.2:** Επίδραση του μεγέθους των περιοχών στο ποσοστό επικάλυψής τους ([Mikolajczyk et al., 2005b])

Εξαιτίας των παραπάνω, θα πρέπει να λάβουμε υπόψη μας την επίδραση του μεγέθους των περιοχών ενδιαφέροντος στο ποσοστό επικάλυψης, άρα και στα μέτρα σύγκρισης, γι' αυτό χρησιμοποιούνται οι αρχικές περιοχές (και όχι οι περιοχές μέτρησης). Παρόλ' αυτά και πάλι υπάρχει πλεονέκτημα για τους ανιχνευτές που εξάγουν γενικά περιοχές μεγάλου μεγέθους, γι' αυτό χρησιμοποιείται η τεχνική της κανονικοποίησης σε ένα συγκεκριμένο μέγεθος. Πριν υπολογιστεί δηλαδή το εμβαδόν επικάλυψης μεταξύ των δύο περιοχών, κανονικοποιείται η περιοχή αναφοράς σε συγκεκριμένη κλίμακα και αντίστοιχα μεταβάλλεται και η δεύτερη περιοχή.

Ως επικάλυψη (overlap) δύο περιοχών εννοούμε το εμβαδόν της τομής τους, όταν αυτές απεικονίζονται στην ίδια εικόνα αναφοράς. Μπορούμε λοιπόν να ορίσουμε το σφάλμα επικάλυψης OE (overlap error), το οποίο θα είναι το ποσοστό του εμβαδού που δεν ανήκει στην τομή τους. Θα ισούται δηλαδή με το λόγο του εμβαδού του μη κοινού τμήματος των περιοχών προς το εμβαδόν της ένωσής τους:

$$OE = \frac{|R_A \cup R_B{}' - R_A \cap R_B{}'|}{|R_A \cup R_B{}'|} \qquad \textbf{(εξ. 3.2)}$$

Στην προηγούμενη σχέση ο τελεστής |.| αποτελεί το μέτρο μια επίπεδης περιοχής, δηλαδή το εμβαδόν της. Αν $M_A$ είναι ο πίνακας που ορίζει την ελλειπτική περιοχή στην εικόνα αναφοράς (τα σημεία της δίνονται από την εξίσωση $x^TM_Ax = 1$), με $R_A$ συμβολίζουμε την περιοχή αυτή. Αν $M_B$ είναι ο πίνακας που ορίζει την ελλειπτική περιοχή στην δεύτερη εικόνα ($x^TM_Bx = 1$), τότε γνωρίζοντας την ομογραφία H που συνδέει τις δύο εικόνες, απεικονίζουμε τη δεύτερη έλλειψη στην εικόνα αναφοράς και η νέα μετασχηματισμένη περιοχή συμβολίζεται με $R_B{}'$ και ορίζεται από τον πίνακα $M_B{}'=H^TM_BH$.

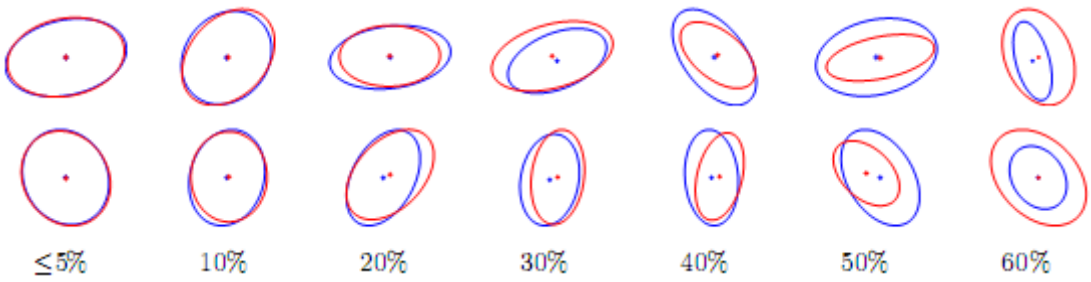
Στο σημείο αυτό θα πρέπει να βρεθεί μία συνθήκη ώστε να οριστεί το μέτρο επαναληψιμότητας, ένα κριτήριο δηλαδή για την ικανότητα του ανιχνευτή να εντοπίζει τις ίδιες περιοχές της σκηνής υπό διάφορες συνθήκες μεταβολών. Μπορούμε να θεωρήσουμε ότι δύο περιοχές αντιστοιχούν μεταξύ τους (corresponding regions), αν το σφάλμα επικάλυψής τους είναι μικρότερο από ένα συγκεκριμένο ποσοστό, το οποίο καθορίζεται ως κατώφλι. Η συνθήκη αυτή δίνεται στην επόμενη σχέση:

$$OE = 1 - \frac{|R_A \cap R_B'|}{|R_A \cup R_B'|} < \varepsilon_O \qquad \textbf{(εξ. 3.3)}$$

Το μέτρο επαναληψιμότητας RS (repeatability score) για δύο συγκεκριμένες εικόνες ορίζεται ως ο λόγος του πλήθους των αντίστοιχων περιοχών προς το πλήθος των συνολικών περιοχών (στην εικόνα με τις λιγότερες περιοχές) [Mikolajczyk et al., 2005b]:

$$RS = \frac{\#corresponding\ regions}{\#total\ regions} \qquad \textbf{(εξ. 3.4)}$$

Οι περιοχές που αντιστοιχούν καθορίζονται από το κριτήριο της εξίσωσης 3.3 και στα πειράματα που θα ακολουθήσουν χρησιμοποιείται κατώφλι $\varepsilon_O$ ίσο με 40%. Λαμβάνονται υπόψη μόνο οι περιοχές που βρίσκονται στο κοινό τμήμα των δύο εικόνων και επίσης γίνεται κανονικοποίηση της περιοχής με τον τρόπο που αναφέρθηκε στην προηγούμενη παράγραφο (απεικονίζεται η περιοχή αναφοράς σε συγκεκριμένου μεγέθους περιοχή και αντίστοιχα μεταβάλλεται και η δεύτερη περιοχή). Μερικές περιπτώσεις επικάλυψης ελλειπτικών περιοχών παρουσιάζονται στο σχήμα 3.3, για διάφορες τιμές του σφάλματος επικάλυψης. Όπως είναι εμφανές, ακόμα και περιοχές που παρουσιάζουν σφάλμα 50% μπορούν να χρησιμοποιηθούν για το ταίριασμα μεταξύ εικόνων, αφού είναι πιθανό οι περιγραφείς τους να ταιριάζουν σε μεγάλο βαθμό.




**Σχ. 3.3:** Επικάλυψη περιοχών για διάφορες τιμές του σφάλματος $\varepsilon_O$ ([Mikolajczyk et al., 2005b])

Εκτός όμως από τη θεωρητική προσέγγιση των επιδόσεων των μεθόδων μέσω του σφάλματος επικάλυψης και του μέτρου επαναληψιμότητας, θα πρέπει οι μέθοδοι να συγκριθούν και σε πρακτικά προβλήματα ταιριάσματος εικόνων, δηλαδή να μελετηθεί η διακριτική τους ικανότητα με βάση τους περιγραφείς που εξάγονται από τις περιοχές τους. Το διάνυσμα χαρακτηριστικών (SIFT ή SURF) εξάγεται από τη γειτονιά του σημείου ή από την κανονικοποιημένη περιοχή 41×41 pixels, όπως συζητήσαμε στο τέλος της παραγράφου 3.1.2, και πάντοτε με προσανατολισμό ως προς την κύρια κατεύθυνση.

Μέσω του διανύσματος αυτού κάθε σημείο ή περιοχή της εικόνας περιγράφεται από ένα σημείο στον χώρο των χαρακτηριστικών και το ταίριασμα μεταξύ αυτών των σημείων γίνεται με την ευκλείδεια απόσταση. Δύο περιοχές δημιουργούν ένα ταίριασμα (match) αν αποτελούν τους κοντινότερους γείτονες στον χώρο των χαρακτηριστικών. Έτσι κάθε περιοχή έχει ένα το πολύ ταίριασμα. Εναλλακτικά, μπορεί να χρησιμοποιηθεί ένα κριτήριο ομοιότητας (similarity), οπότε δύο περιοχές να θεωρείται ότι ταιριάζουν, αν η απόστασή τους είναι μικρότερη από ένα συγκεκριμένο κατώφλι. Με τον δεύτερο τρόπο κάθε περιοχή μπορεί να ταιριάζει με περισσότερες από μία περιοχές της άλλης εικόνας. Στα σχετικά πειράματα που θα ακολουθήσουν θα χρησιμοποιηθεί η δεύτερη προσέγγιση για το ταίριασμα μεταξύ των περιοχών.

Ένα ταίριασμα μιας συγκεκριμένης περιοχής Α με μια περιοχή Β από την άλλη εικόνα θα θεωρείται ότι είναι σωστό (correct match), αν το σφάλμα επικάλυψης όπως ορίστηκε προηγουμένως (βλέπε σχέσεις 3.2 και 3.3) είναι ελάχιστο για αυτές τις δύο περιοχές και επίσης μικρότερο του 40% [Mikolajczyk et al., 2005b]. Εννοούμε ελάχιστο μεταξύ κάθε άλλου ζεύγους περιοχών που ταιριάζουν, σε περίπτωση φυσικά που έχουμε επιλέξει τον δεύτερο τρόπο ταιριάσματος (βλέπε παραπάνω) και υπάρχουν και άλλες περιοχές $C_i$ που ταιριάζουν με την περιοχή Α. Με αυτόν τον τρόπο ορίζεται ένα μόνο σωστό ταίριασμα για κάθε περιοχή, άρα μεταξύ των δύο εικόνων θα υπάρχουν μοναδικά ζεύγη περιοχών που ταιριάζουν σωστά και επίσης περιοχές που δεν ταιριάζουν καθόλου (απομακρυσμένες από όλες τις άλλες στον χώρο των χαρακτηριστικών) ή που ταιριάζουν λανθασμένα (τα σημεία περιγραφής τους βρίσκονται κοντά αλλά δεν ικανοποιείται η συνθήκη του σφάλματος επικάλυψης).

Ως μέτρο ταιριάσματος MS (matching score) για δύο συγκεκριμένες εικόνες ορίζεται ο λόγος του πλήθους των σωστών ταιριασμάτων (κοινό και για τις δύο εικόνες, αφού πρόκειται για αντιστοιχία "1-1") προς το μικρότερο πλήθος των συνολικών περιοχών για τις δύο εικόνες:

$$MS = \frac{\#correct\ matches}{\#total\ regions} \qquad \textbf{(εξ. 3.5)}$$

Το μέτρο αυτό είναι ενδεικτικό της διακριτικής ικανότητας των μεθόδων, δηλαδή του πόσο καλά μπορούν να ταιριάξουν δύο εικόνες της ίδιας σκηνής, λαμβάνοντας υπόψη της τιμές του περιγραφέα που εξάγονται από τις περιοχές των ανιχνευτών. Ειδικότερα ο περιγραφέας εξάγεται από την περιοχή μέτρησης, η οποία στην παρούσα εργασία θα θεωρείται πάντοτε τριπλάσια της αρχικής, ακριβώς όπως προτείνεται στη βιβλιογραφία και έχει πειραματικά αποδειχθεί κατάλληλη [Ke et al., 2004], [Mikolajczyk et al., 2005a], [Mikolajczyk et al., 2005b]. Ο πολλαπλασιαστικός παράγοντας που σχετίζει την περιοχή μέτρησης με την αρχική περιοχή θα πρέπει να επιλέγεται προσεκτικά, ειδικά σε προβλήματα ανάκτησης εικόνων και αναγνώρισης αντικειμένων, όπου μπορεί να συμβαίνουν συχνά μερικές επικαλύψεις αντικειμένων. Η εξαγωγή πληροφορίας από μεγαλύτερη περιοχή εισάγει το ρίσκο της χρήσης περιττής ή και επιζήμιας πληροφορίας.

Όπως προκύπτει από τον ορισμό του, το μέτρο ταιριάσματος είναι πάντοτε μικρότερο από το μέτρο επαναληψιμότητας, αφού έχει ως αριθμητή το πλήθος των σωστών ταιριασμάτων με άνω φράγμα τις περιοχές που αντιστοιχούν μεταξύ τους (όλες οι αντιστοιχίες με σφάλμα επικάλυψης έως και 40%):

$$MS \leq RS \qquad \textbf{(εξ. 3.6)}$$

Επειδή κάθε έλλειψη μπορεί να έχει περισσότερες από μία περιοχές που να της αντιστοιχούν, γίνεται εμφανές ότι το πλήθος των σωστών ταιριασμάτων είναι πολύ μικρότερο από το πλήθος των αντίστοιχων περιοχών. Αυτό σημαίνει ότι μία διαφορά, ας πούμε της τάξης του 20% (όπως θα φανεί και παρακάτω), μεταξύ των μέτρων επαναληψιμότητας και ταιριάσματος θα οφείλεται περισσότερο στην εξ' ορισμού απόκλισή τους και λιγότερο στο προβληματικό ταίριασμα του περιγραφέα. Επίσης, πρέπει σημειώσουμε ότι λόγω της πολλαπλής αντιστοιχίας μεταξύ περιοχών, οι μέθοδοι που γενικά παράγουν μεγάλο αριθμό σημείων (όπως οι Harris-Hessian Affine, βλέπε πίνακα 3.2) ευνοούνται διότι θα έχουν συστηματικά αυξημένα ποσοστά επαναληψιμότητας σε σχέση με τις υπόλοιπες. Με την αύξηση της πυκνότητας των περιοχών, αυξάνεται και η πιθανότητα επικάλυψης μεγαλύτερου ποσοστού σε εμβαδόν.

## 3.3 Σύγκριση των μεθόδων ανίχνευσης και περιγραφής

### 3.3.1 Σύγκριση ως προς την επαναληψιμότητα

Στα πειράματα αυτής της παραγράφου επιχειρείται η σύγκριση των μεθόδων τοπικών χαρακτηριστικών (ανιχνευτών) με βάση το μέτρο επαναληψιμότητας. Πιο συγκεκριμένα γίνεται μια εκτίμηση του βαθμού στον οποίο επηρεάζεται το πλήθος των αντίστοιχων περιοχών (κριτήριο 3.3) σε σχέση με τους διάφορους σταδιακούς μετασχηματισμούς (ανάλογα με την ακολουθία εικόνων). Μια ιδανική μέθοδος θα απέδιδε επαναληψιμότητα ίση με 100% για οποιαδήποτε μεταβολή της εικόνας, αλλά όπως είναι αναμενόμενο αυτό δεν είναι εφικτό, είτε λόγω σφαλμάτων του ανιχνευτή, είτε λόγω περιεχομένου της εικόνας είτε και λόγω συνδυασμού των δύο αυτών παραγόντων. Επίσης, για τρεις από τους πέντε συνολικά μετασχηματισμούς, επιχειρείται να εξεταστεί η απόδοση των μεθόδων και ως προς τον τύπο της σκηνής (με δομή και με υφή).

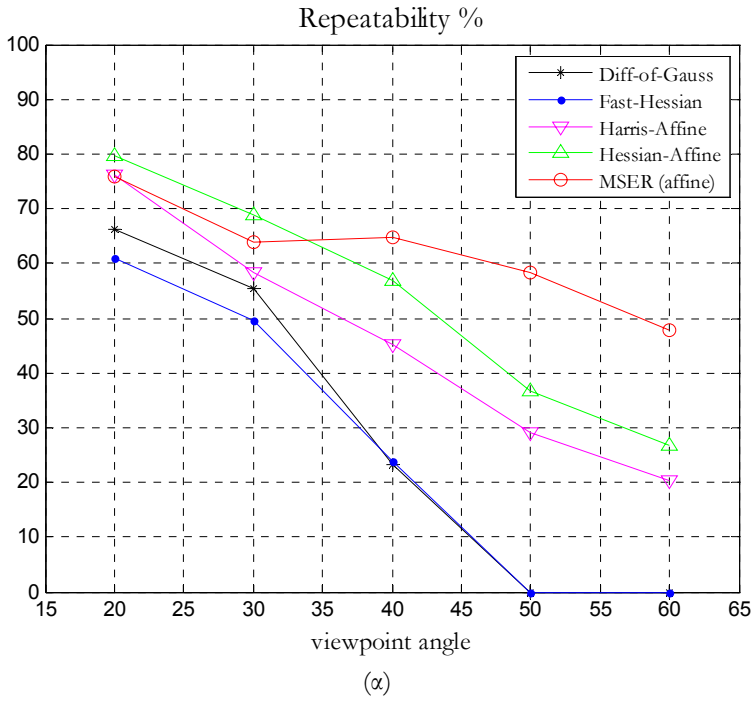
Στο σχήμα 3.4 περιλαμβάνονται τα διαγράμματα του ποσοστού της επαναληψιμότητας για τις ακολουθίες που αφορούν μεταβολές στη γωνία λήψης (viewpoint change), για δύο διαφορετικούς τύπους εικόνων, την ακολουθία Graffiti με structured εικόνες και την ακολουθία Wall με textured εικόνες (επαναλαμβανόμενο μοτίβο). Όπως φαίνεται από το πρώτο διάγραμμα του σχήματος, καλύτερη από όλες τις μεθόδους προκύπτει η τεχνική MSER, καθώς παρουσιάζει πολύ μικρή κλίση ως προς τη μεταβολή της γωνίας. Μάλιστα για γωνίες $20^o$ και $30^o$ η μέθοδος Hessian Affine έχει πολύ καλό ποσοστό (ξεκινάει από 80%), αλλά στη συνέχεια για μεγαλύτερες γωνίες, αυτή και η μέθοδος Harris Affine μειώνονται πολύ πιο απότομα από την MSER, η οποία για την πιο ακραία εικόνα δεν πέφτει πολύ κάτω από 50%. Οι μέθοδοι DoG και Fast Hessian όπως είναι λογικό (αφού δεν παράγουν σημεία ανεξάρτητα από αφινικές μεταβολές), έχουν σχετικά καλά ποσοστά μόνο για τις μικρές γωνίες, ενώ για πολύ μεγάλη κλίση της κάμερας τα ποσοστά τους μηδενίζονται. Αυτό εξηγείται από τον τρόπο καθορισμού της περιοχής ενδιαφέροντος (κυκλικής), όπου ενώ πολλά σημεία εντοπίζονται σωστά, η περιοχή (που καθορίζεται από την χαρακτηριστική τους κλίμακα) δεν αντιστοιχεί πλήρως στον αφινικό μετασχηματισμό και το σφάλμα επικάλυψης είναι υψηλό. Όσον αφορά την ακολουθία Wall με υφή, βλέπουμε ότι όλοι οι ανιχνευτές παρουσιάζουν παρόμοια συμπεριφορά στην αρχή και μέχρι τη γωνία των $40^o$, αλλά για τις τελευταίες δύο γωνίες η μέθοδος MSER είναι και πάλι καλύτερη (αν και με μικρότερα ποσοστά), ακολουθούμενη από τις Harris-Hessian Affine και τελευταίες τις μη αφινικές μεθόδους.

Στο σχήμα 3.5 φαίνονται τα διαγράμματα για τις σκηνές στις οποίες εφαρμόζεται περιστροφή και ταυτόχρονη αλλαγή κλίμακας, δηλαδή στις ακολουθίες Boat (εικόνες με δομή) και Bark (εικόνες με υφή). Και στους δύο τύπους εικόνων οι μέθοδοι Harris-Hessian Affine έρχονται πρώτες, αν και στην δεύτερη εικόνα όπως βλέπουμε και από τον πίνακα 3.2 οι τεχνικές αυτές παράγουν πολύ λίγα σημεία, με αποτέλεσμα να μειώνεται η πιθανότητα σφάλματος και άρα να ευνοούνται στα διαγράμματα. Οι DoG και Fast Hessian τεχνικές ακολουθούν ενδιάμεση πορεία, ενώ η μέθοδος MSER παρόλο που έρχεται τελευταία, φαίνεται ότι ακολουθεί τη μεταβολή των υπολοίπων.

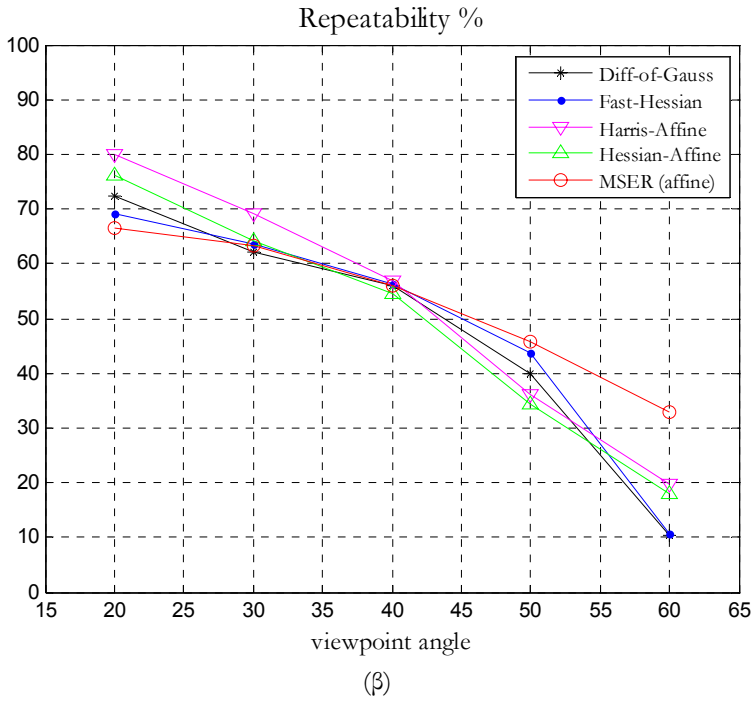
Τα διαγράμματα στο σχήμα 3.6 αφορούν τις εικόνες με σταδιακό θόλωμα, δηλαδή τις ακολουθίες Bikes και Trees (structured και textured scene αντίστοιχα). Με μια γενική ματιά φαίνεται ότι και οι πέντε τεχνικές δεν επηρεάζονται από τον μετασχηματισμό αυτόν, αφού παρουσιάζουν πολύ μικρή κλίση καθώς αυξάνεται το θόλωμα της εικόνας. Στο πρώτο διάγραμμα οι Harris-Hessian Affine μεταβάλλονται μαζί, ενώ οι Fast Hessian και MSER έχουν σχεδόν σταθερό μέτρο επαναληψιμότητας για τις διάφορες μεταβολές. Η μέθοδος DoG έρχεται τελευταία, όμως στο δεύτερο διάγραμμα είναι η μόνη που διατηρεί το ποσοστό της σχεδόν

αμετάβλητο, ενώ όλες οι υπόλοιπες παρουσιάζουν σημαντική υστέρηση σε σχέση με το πρώτο διάγραμμα, σχεδόν κατά 20%. Η μέθοδος MSER είναι η χειρότερη, γεγονός που εξηγείται από το ότι η εικόνα δεν έχει πλέον ομοιογενείς περιοχές (είναι εικόνα υφής) και οι ευδιάκριτες ακμές της έχουν χαθεί (θόλωμα), οπότε η διαδικασία εντοπισμού σταθερών περιοχών δεν μπορεί να λειτουργήσει σωστά. Το γεγονός όμως ότι για όλες τις μεθόδους το ποσοστό επαναληψιμότητας είναι ανεξάρτητο από τον βαθμό του μετασχηματισμού σημαίνει ότι αυτό που επηρέασε την επαναληψιμότητα είναι ο συνδυασμός τύπου σκηνής και θολώματος και όχι από μόνο του το θόλωμα της εικόνας.

Τέλος, στο σχήμα 3.7 φαίνονται τα ποσοστά επαναληψιμότητας για τις ακολουθίες UBC και Leuven που αφορούν συμπίεση JPEG και αλλαγές στο φωτισμό της εικόνας (γραμμική μεταβολή $I'=\alpha I+\beta$). Για την πρώτη ακολουθία είναι φανερό ότι οι Harris-Hessian Affine αποδίδουν με τον καλύτερο έως τώρα τρόπο, ξεκινώντας από ποσοστά μεγαλύτερα του 95%, ακολουθούνται από την Fast Hessian μέθοδο που δεν πέφτει κάτω από 50% και την DoG λίγο πιο κάτω, ενώ τελευταία η MSER φαίνεται να μην ταιριάζει στα artifacts που δημιουργούνται με την συμπίεση JPEG και να επηρεάζεται σε μέγιστο βαθμό. Όσον αφορά την ακολουθία Leuven, βλέπουμε ότι οι μεταβολές της φωτεινότητας δεν επηρεάζουν τις μεθόδους, που έχουν σχεδόν σταθερό μέτρο επαναληψιμότητας και μάλιστα με την MSER να είναι καλύτερη με σταθερό (ή και αυξανόμενο) ποσοστό.

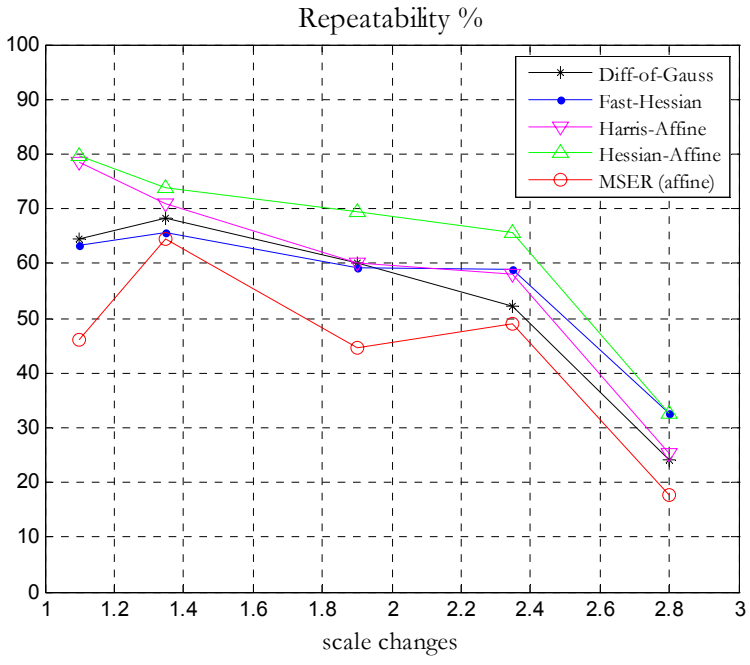



(α)

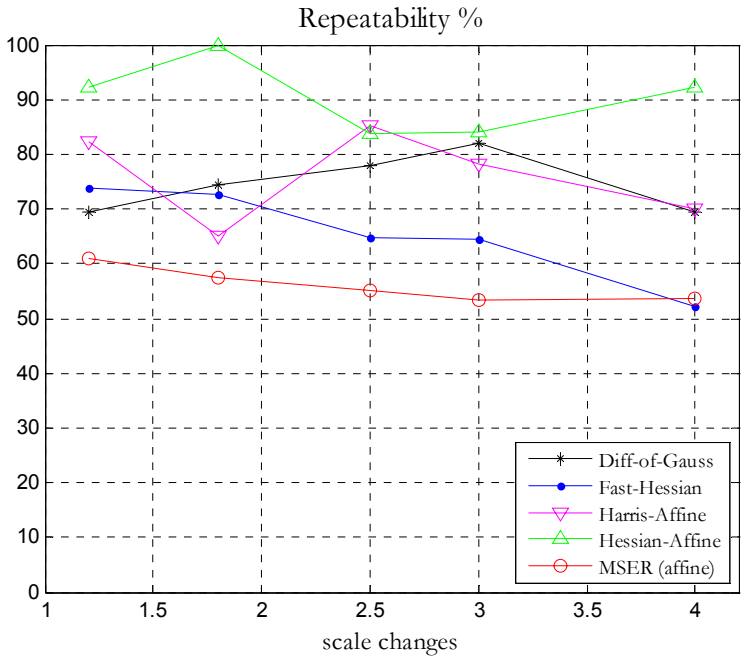



(β)

**Σχ. 3.4:** Το μέτρο επαναληψιμότητας (% ποσοστό) για τις ακολουθίες εικόνων (α) Graffiti και (β) Wall με μετασχηματισμό της οπτικής γωνίας (viewpoint change)

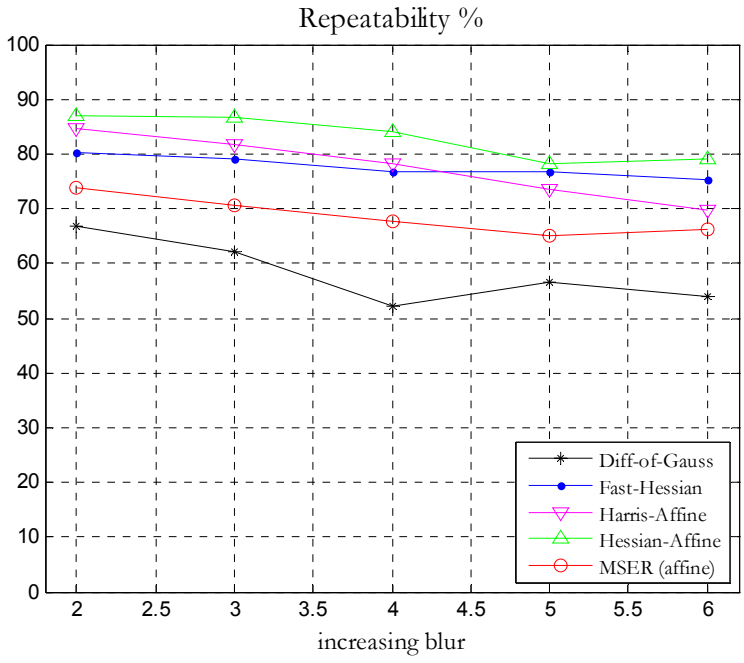



(α)

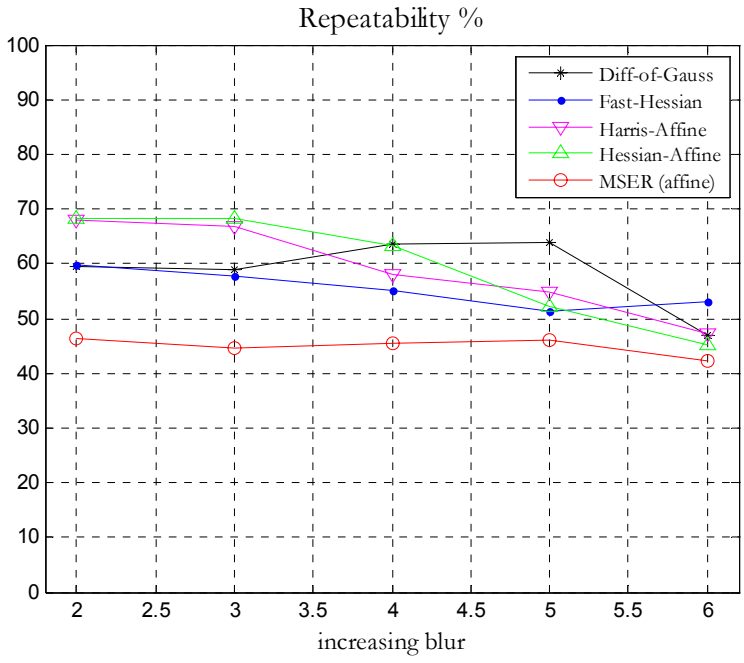



(β)

**Σχ. 3.5:** Το μέτρο επαναληψιμότητας (% ποσοστό) για τις ακολουθίες εικόνων (α) Boat και (β) Bark με περιστροφή και αλλαγή κλίμακας (scale change)

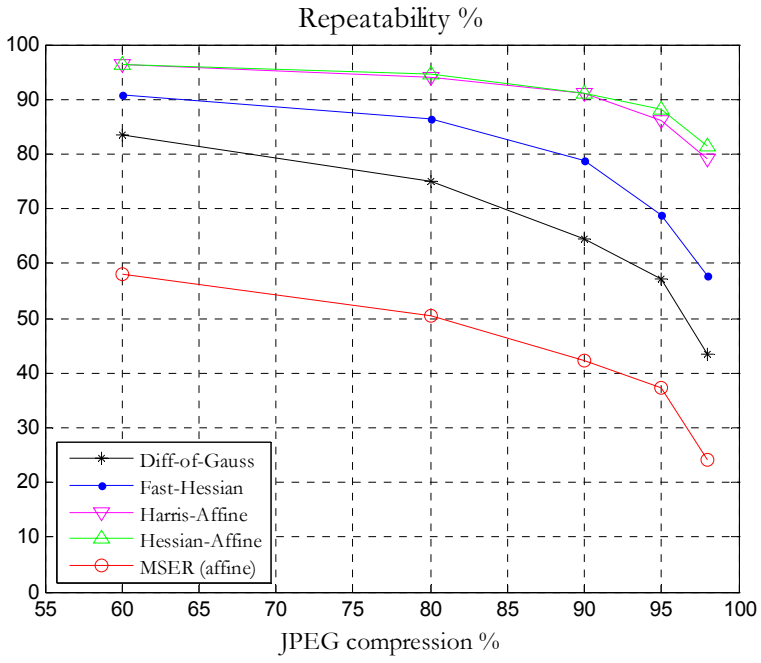



(α)

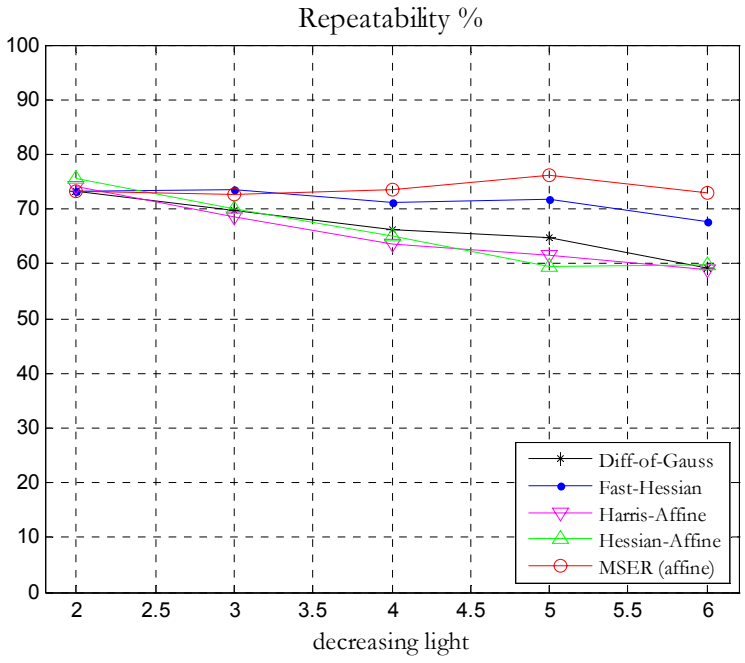



(β)

**Σχ. 3.6:** Το μέτρο επαναληψιμότητας (% ποσοστό) για τις ακολουθίες εικόνων (α) Bikes και (β) Trees με διαδοχικό θόλωμα της εικόνας (increasing blur)




(α)




(β)

**Σχ. 3.7:** Το μέτρο επαναληψιμότητας (% ποσοστό) για τις ακολουθίες εικόνων (α) UBC και (β) Leuven με συμπίεση JPEG και διαδοχική μείωση της φωτεινότητας (decreasing light) αντίστοιχα

### 3.3.2 Σύγκριση ως προς την ακρίβεια ανίχνευσης

Στην παράγραφο αυτή γίνεται μία πιο λεπτομερής ανάλυση της απόδοσης των μεθόδων μετρώντας την ακρίβεια με την οποία εντοπίζουν τα σημεία ή τις περιοχές ενδιαφέροντος. Αυτό μπορεί να γίνει μεταβάλλοντας το κατώφλι με το οποίο καθορίζεται το μέγιστο αποδεκτό σφάλμα επικάλυψης (σχέση 3.3) και παρατηρώντας τη μορφή της καμπύλης επαναληψιμότητας. Στα σχήματα 3.8 – 3.11 περιλαμβάνονται τα διαγράμματα του μέτρου επαναληψιμότητας σε συνάρτηση με τη μεταβολή του σφάλματος επικάλυψης. Όλα τα διαγράμματα αφορούν την επαναληψιμότητα μεταξύ της πρώτης και της τέταρτης εικόνας κάθε ακολουθίας. Όσο το κατώφλι για το σφάλμα διευρύνεται, τόσο περισσότερες περιοχές ικανοποιούν το κριτήριο και αντιστοιχούν με άλλες, οπότε το μέτρο επαναληψιμότητας για όλες τις μεθόδους αυξάνει.

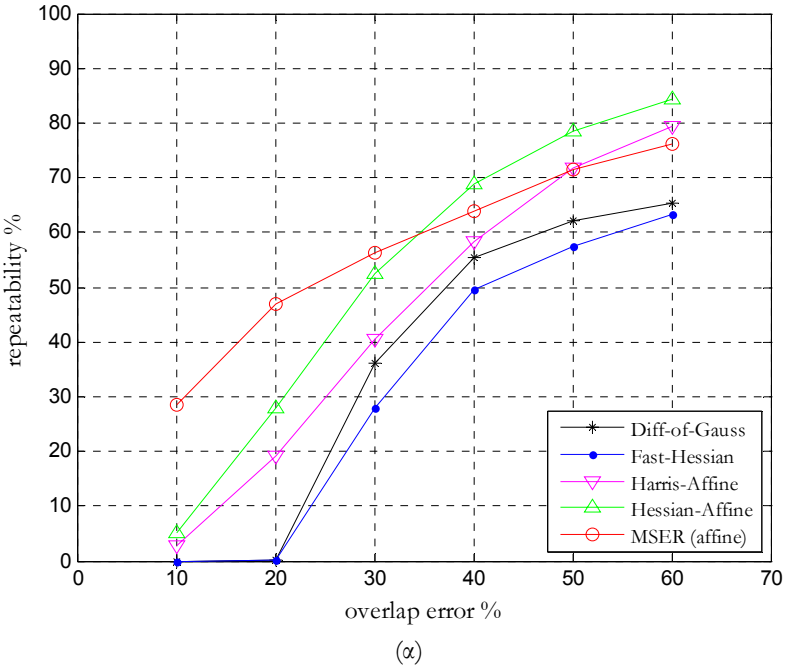
Μία μέθοδος ανίχνευσης με την ιδανικότερη ακρίβεια θα έπρεπε να έχει ποσοστό επαναληψιμότητας κοντά στο 100%, για οποιαδήποτε τιμή σφάλματος επικάλυψης. Επειδή όμως αυτό δεν είναι εφικτό, ως καλύτερη μέθοδος θεωρείται εκείνη που ξεκινάει με υψηλά ποσοστά επαναληψιμότητας για τις μικρότερες τιμές σφάλματος επικάλυψης και στη συνέχεια παρουσιάζει αύξηση με όσο το δυνατόν μικρότερη κλίση (ιδανικά επίπεδη καμπύλη). Αυτό σημαίνει ότι ένας ανιχνευτής υψηλής ακρίβειας θα εντοπίζει σωστά τη θέση των σημείων-περιοχών ενδιαφέροντος ανεξάρτητα από το πόσο μεγάλη περιοχή χρησιμοποιείται για το μέτρο επαναληψιμότητας, δηλαδή χωρίς να ευνοείται από την ανοχή σε επικάλυψη μέσω του κριτηρίου 3.3.

Στο σχήμα 3.8 βλέπουμε τα διαγράμματα για την περίπτωση της αλλαγής στην οπτική γωνία. Για την ακολουθία Graffiti είναι φανερή η επικράτηση της μεθόδου MSER ως προς την ακρίβεια, αφού ξεκινάει με μεγαλύτερες τιμές και στη συνέχεια αυξάνεται με τη μικρότερη κλίση. Οι δύο μη αφινικές μέθοδοι όπως είναι λογικό ξεκινούν από μηδενικές τιμές και στη συνέχεια αυξάνονται, όχι όμως τόσο απότομα όσο οι Harris-Hessian Affine οι οποίες είναι οι λιγότερο ακριβείς. Στην περίπτωση των εικόνων Wall όλες οι μέθοδοι ακολουθούν παρόμοια συμπεριφορά στην καμπύλη.

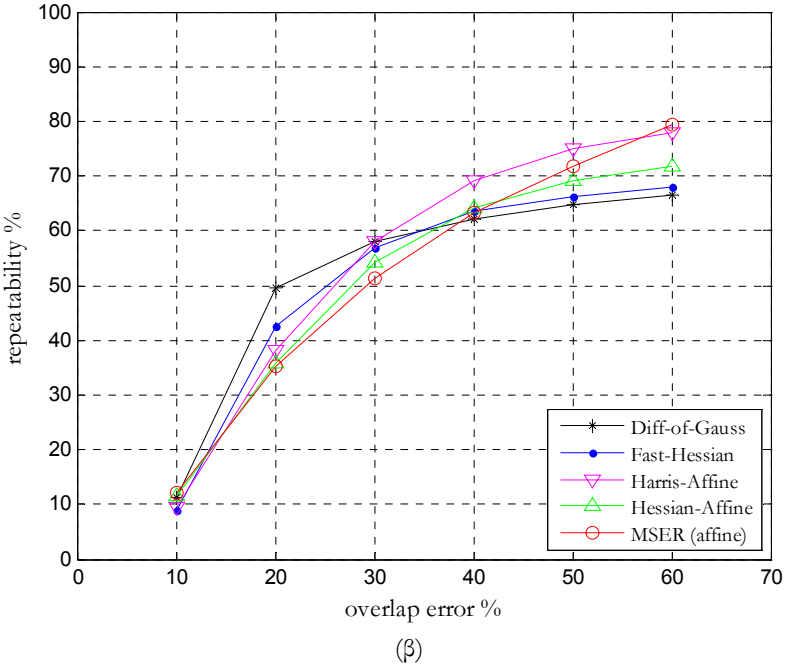
Στο σχήμα 3.9 τα διαγράμματα αφορούν τις εικόνες με μεταβολή της κλίμακας και περιστροφή. Από το πρώτο διάγραμμα για την ακολουθία Boat προκύπτει η μέθοδος DoG ως πιο ακριβής, ενώ όλες οι υπόλοιπες μέθοδοι παρουσιάζουν μεγάλη κλίση. Στις εικόνες Bark οι μέθοδοι που είναι προσανατολισμένες να εντοπίζουν σημεία ανεξάρτητα από αλλαγές κλίμακας είναι πράγματι πιο ακριβείς, ακολουθούνται όμως από την MSER.

Στα διαγράμματα του σχήματος 3.10 για τις εικόνες με θόλωμα, γίνεται άμεσα αντιληπτό ότι οι μέθοδοι DoG και Fast Hessian έχουν την καλύτερη ακρίβεια, ενώ όλες οι αφινικές μέθοδοι έχουν παρόμοια μορφή στις γραφικές τους παραστάσεις και για τους δύο τύπους σκηνών. Αυτό δείχνει ότι η επικράτηση των Harris-Hessian Affine ως προς την επαναληψιμότητα στα διαγράμματα του σχήματος 3.6 δεν οφειλόταν τόσο στην ικανότητα των μεθόδων, όσο στη μεγάλη ανοχή σε σφάλματα επικάλυψης.

Τέλος, στο σχήμα 3.11 περιέχονται τα διαγράμματα για τη συμπίεση JPEG και την μεταβολή της φωτεινότητας. Από την καμπύλη για την ακολουθία UBC φαίνεται ότι οι Harris-Hessian Affine έχουν πολύ καλή ακρίβεια (σχεδόν επίπεδη καμπύλη με τιμές κοντά στο 100%), ακολουθούνται από τις Fast Hessian και DoG, ενώ η μέθοδος MSER είναι η χειρότερη για αυτόν τον τύπο μετασχηματισμού. Για την ακολουθία Leuven οι μέθοδοι Fast Hessian και DoG παρουσιάζουν πολύ καλά αποτελέσματα, με τις υπόλοιπες αφινικές να συμπεριφέρονται ικανοποιητικά.

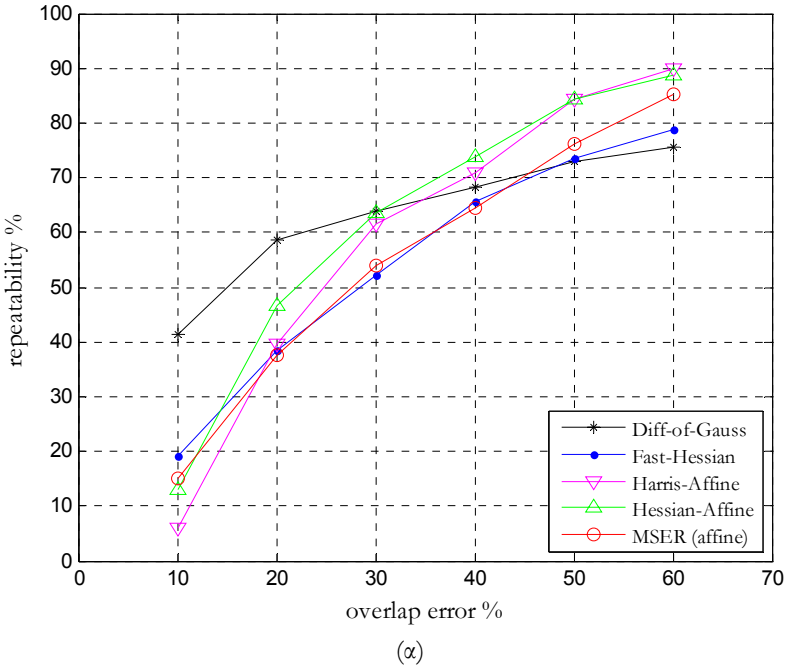



(α)

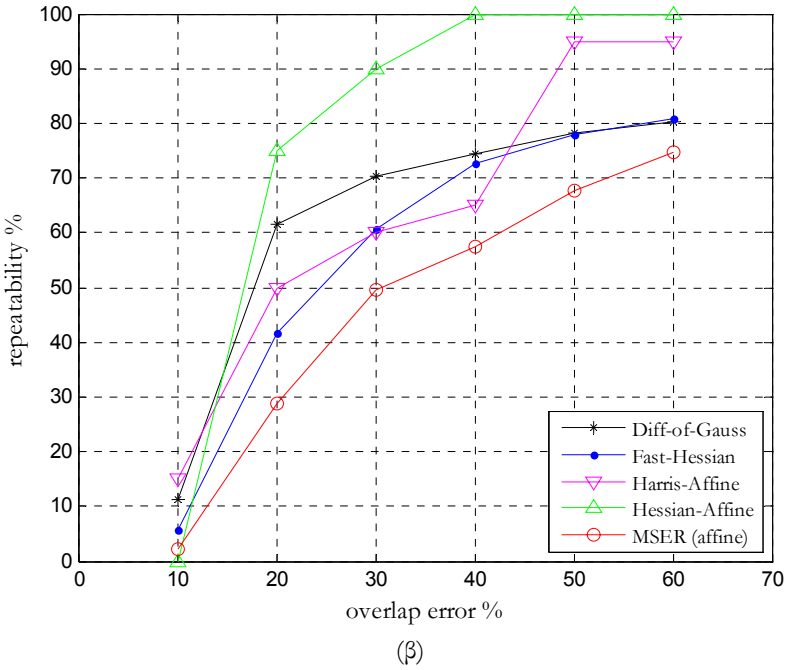



(β)

**Σχ. 3.8:** Ακρίβεια των ανιχνευτών (επίδραση του σφάλματος επικάλυψης στην επαναληψιμότητα) για τις ακολουθίες εικόνων (α) Graffiti και (β) Wall με μετασχηματισμό της οπτικής γωνίας (viewpoint change)

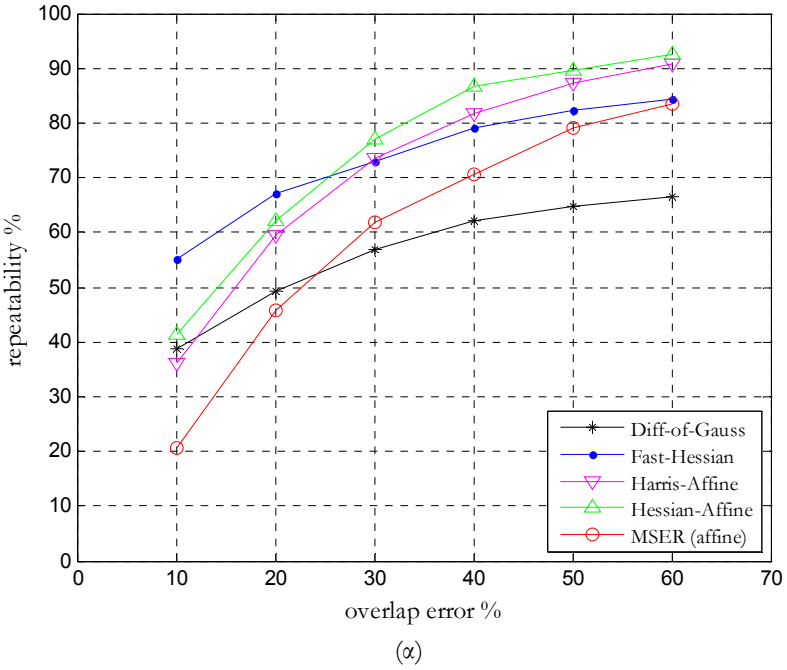



(α)

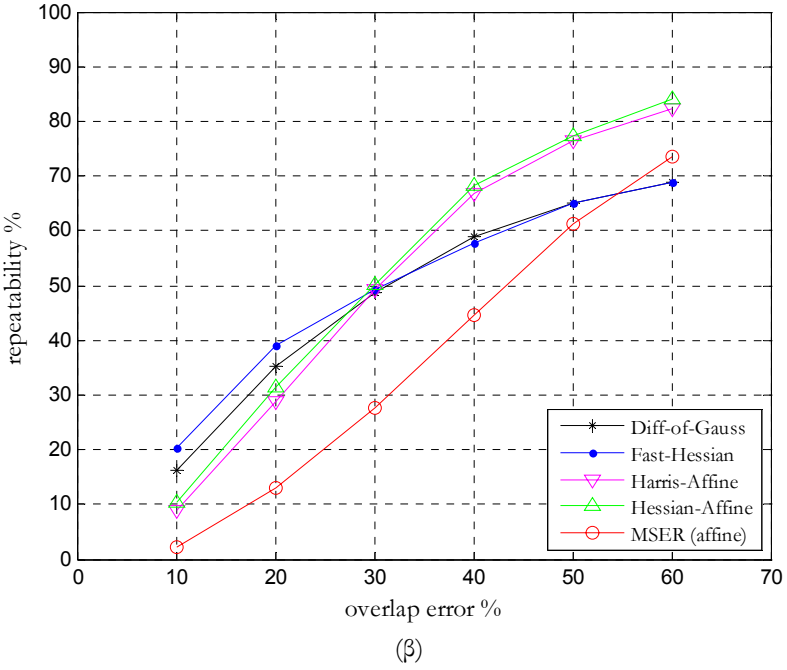



(β)

**Σχ. 3.9:** Ακρίβεια των ανιχνευτών (επίδραση του σφάλματος επικάλυψης στην επαναληψιμότητα) για τις ακολουθίες εικόνων (α) Boat και (β) Bark με περιστροφή και αλλαγή κλίμακας (scale change)

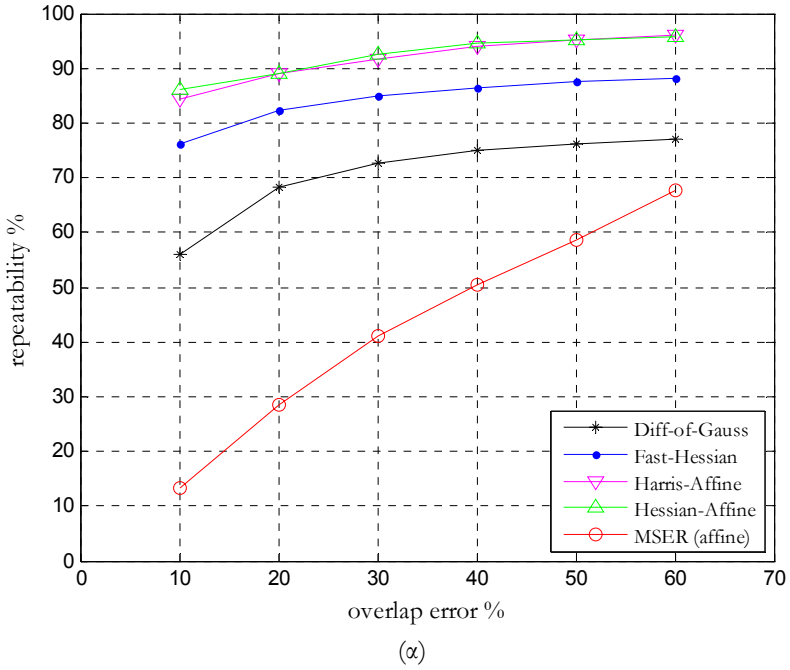



(α)

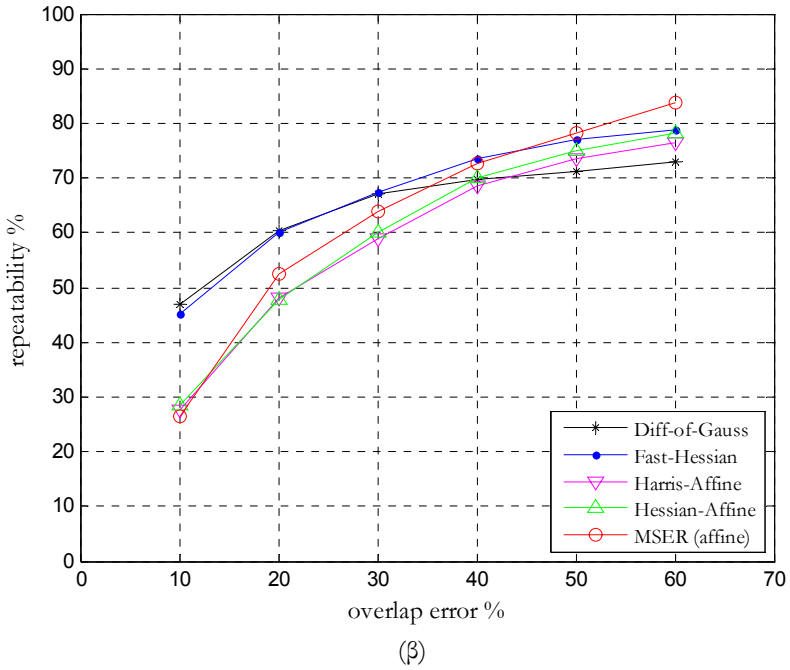



(β)

**Σχ. 3.10:** Ακρίβεια των ανιχνευτών (επίδραση του σφάλματος επικάλυψης στην επαναληψιμότητα) για τις ακολουθίες εικόνων (α) Bikes και (β) Trees με διαδοχικό θόλωμα της εικόνας (increasing blur)




(α)




(β)

**Σχ. 3.11:** Ακρίβεια των ανιχνευτών (επίδραση του σφάλματος επικάλυψης στην επαναληψιμότητα) για τις ακολουθίες εικόνων (α) UBC και (β) Leuven με συμπίεση JPEG και διαδοχική μείωση της φωτεινότητας (decreasing light) αντίστοιχα

### 3.3.3 Σύγκριση ως προς την ικανότητα ταιριάσματος

Στα διαγράμματα αυτής της παραγράφου παρουσιάζεται το μέτρο ταιριάσματος μεταξύ των εικόνων όπως ορίζεται από τη σχέση 3.5 ως προς το βαθμό του μετασχηματισμού για κάθε ακολουθία ξεχωριστά. Τα διαγράμματα αυτά είναι περισσότερο ποιοτικά παρά ποσοτικά. Δείχνουν την ικανότητα ταιριάσματος περιοχών – σημείων στο χώρο των χαρακτηριστικών. Όπως συζητήθηκε ήδη στο τέλος της παραγράφου 3.2 το ποσοστό ταιριάσματος αναμένεται να είναι μικρότερο από το ποσοστό επαναληψιμότητας, και αυτό εξαιτίας του τρόπου ορισμού του. Παρόλ' αυτά τα διαγράμματα ταιριάσματος θα πρέπει να ακολουθούν σε μορφή τα αντίστοιχα διαγράμματα επαναληψιμότητας. Όταν δεν συμβαίνει κάτι τέτοιο τότε συμπεραίνουμε πως ο συγκεκριμένος ανιχνευτής παράγει περιοχές οι οποίες δεν είναι αρκετά διαχωρίσιμες από τις υπόλοιπες ούτως ώστε να δίνουν το κατάλληλο ταίριασμα με τις αντίστοιχες από την άλλη εικόνα. Αν τα διαγράμματα επαναληψιμότητας αφορούν τη θεωρητική μέτρηση της απόδοσης των μεθόδων ανίχνευσης, τότε το μέτρο ταιριάσματος αποτελεί το κριτήριο για την καταλληλότητα της μεθόδου σε πρακτικά προβλήματα.

Οι γραφικές παραστάσεις λοιπόν θα αναλυθούν ενδεικτικά, περισσότερο ως προς την μεταβολή τους συναρτήσει του βαθμού παραμόρφωσης (περιστροφή, θόλωμα, κτλ.) παρά ως προς τις τιμές τους. Για το ταίριασμα των περιοχών χρησιμοποιούνται ξεχωριστά τα δύο είδη περιγραφέων που αναλύθηκαν στο κεφάλαιο 2. Οι συνεχόμενες γραμμές αφορούν ταίριασμα με διάνυσμα χαρακτηριστικών SIFT ενώ οι διακεκομμένες γραμμές είναι για τα αποτελέσματα ταιριάσματος με διάνυσμα χαρακτηριστικών SURF. Σε όλα τα πειράματα χρησιμοποιείται κατώφλι 40% για το σφάλμα επικάλυψης, το οποίο καθορίζει τις περιοχές που αντιστοιχούν η μία στην άλλη.

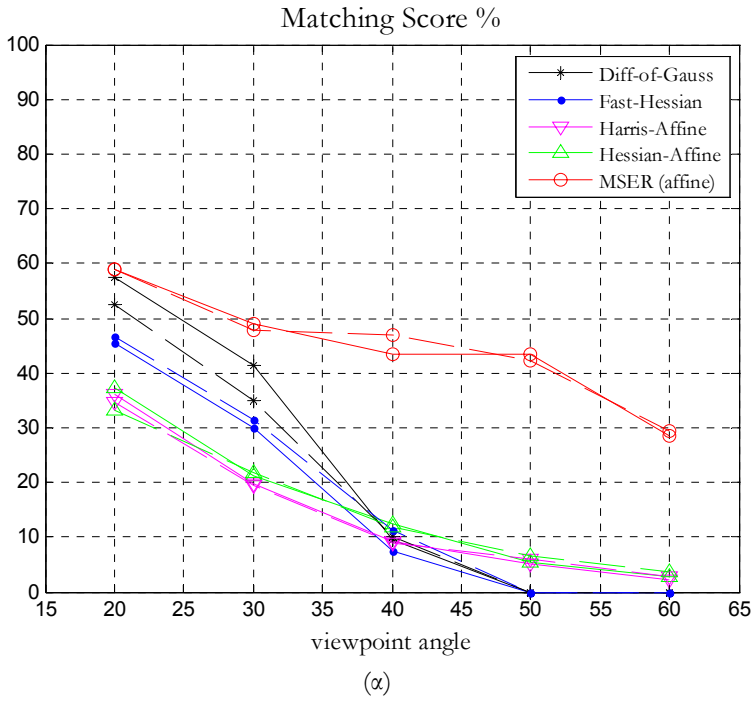
Στο σχήμα 3.12 περιλαμβάνονται τα διαγράμματα για τις σκηνές με αλλαγή στην οπτική τους γωνία. Η διάταξη αλλά και η πορεία των γραφημάτων είναι ίδια με εκείνη του σχήματος 3.4, που αφορά το μέτρο επαναληψιμότητας, εκτός από τις μεθόδους Harris-Hessian Affine, οι οποίες και για τους δύο τύπους σκηνών δίνουν πολύ χαμηλά ποσοστά ταιριάσματος. Οι περιοχές δηλαδή που ανιχνεύουν δεν είναι κατάλληλες για να αντιστοιχηθούν οι δύο εικόνες μέσω των περιγραφέων των σημείων τους. Η μέθοδος MSER φαίνεται ότι είναι πολύ καλύτερη ως προς τις υπόλοιπες για τέτοιου είδους μετασχηματισμούς, όχι λόγω των τιμών του ποσοστού ταιριάσματος, αλλά κυρίως λόγω της μικρής αρνητική κλίσης καθώς ο μετασχηματισμός αυξάνεται.

Στο σχήμα 3.13 βρίσκονται τα διαγράμματα για τις ακολουθίες Boat και Bark που υπόκεινται σε περιστροφή και αλλαγή κλίμακας. Όπως και προηγουμένως, η εξέλιξη των καμπυλών ως προς τον μετασχηματισμό είναι ίδια με την αντίστοιχη της επαναληψιμότητας (σχήμα 3.5), αλλά με τις μεθόδους Harris-Hessian Affine να αποδίδουν και πάλι σε χαμηλότερα επίπεδα σε αντίθεση με την υψηλή τους επαναληψιμότητα.

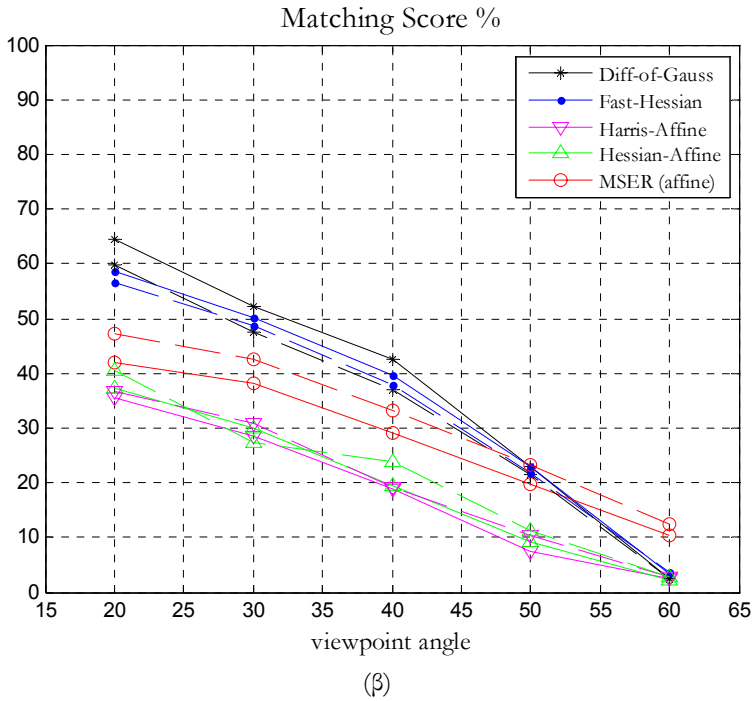
Τα διαγράμματα του σχήματος 3.14 αφορούν το αυξανόμενο θόλωμα. Στην ακολουθία Bikes (σκηνή με δομή) οι μέθοδοι Fast Hessian και DoG ταιριάζουν πολύ καλά τις σωστές περιοχές, όμως για την ακολουθία Trees (σκηνή με υφή) τα λάθη στο ταίριασμα (false matches) αυξάνονται κατά πολύ. Οι μέθοδοι MSER και Harris-Hessian Affine και για τους δύο τύπους σκηνών δεν παράγουν σωστά ταιριάσματα με αποτέλεσμα το μέτρο ταιριάσματος να πέφτει σχεδόν 40% (για τις δύο τελευταίες) κάτω από το μέτρο επαναληψιμότητας, ειδικά για την ακολουθία των textured εικόνων με τα δέντρα. Παρόλο που η κλίση για αυτές τις καμπύλες δεν είναι μεγάλη, η πολύ μικρή τιμή στα ποσοστά ακόμα και για τις πρώτες εικόνες του μετασχηματισμού μας οδηγεί στο συμπέρασμα ότι δεν είναι κατάλληλες για τα προβλήματα στα οποία συμβαίνει θόλωμα των εικόνων.

Τέλος, στο σχήμα 3.15 βλέπουμε τα διαγράμματα ως προς τη συμπίεση JPEG και την μεταβολή της φωτεινότητας. Στην ακολουθία UBC (πρώτο διάγραμμα) είναι η μόνη φορά που οι Harris-Hessian Affine επιτυγχάνουν (επιβεβαιώνοντας το θεωρητικό μέτρο επαναληψιμότητας)

και μαζί με την Fast Hessian δίνουν υψηλά ποσοστά επιτυχίας. Στην περίπτωση της ακολουθίας Leuven, ενώ θεωρητικά όλες οι μέθοδοι είχαν παρόμοια συμπεριφορά, μόνο οι Fast Hessian και DoG διατηρούν τα επίπεδα επιτυχίας τους, ακολουθούνται από την MSER, με τελευταίες τις Harris-Hessian Affine με τη μεγαλύτερη κλίση στα γραφήματά τους.

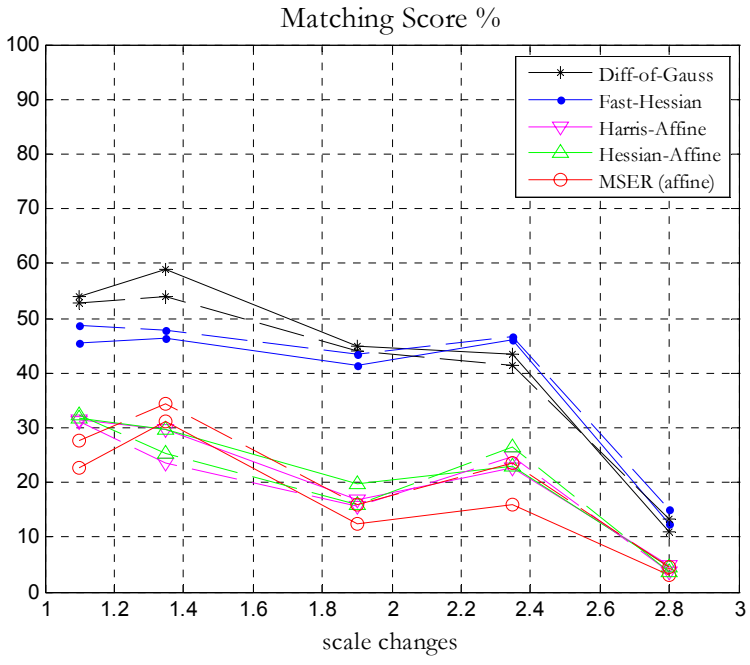



(α)

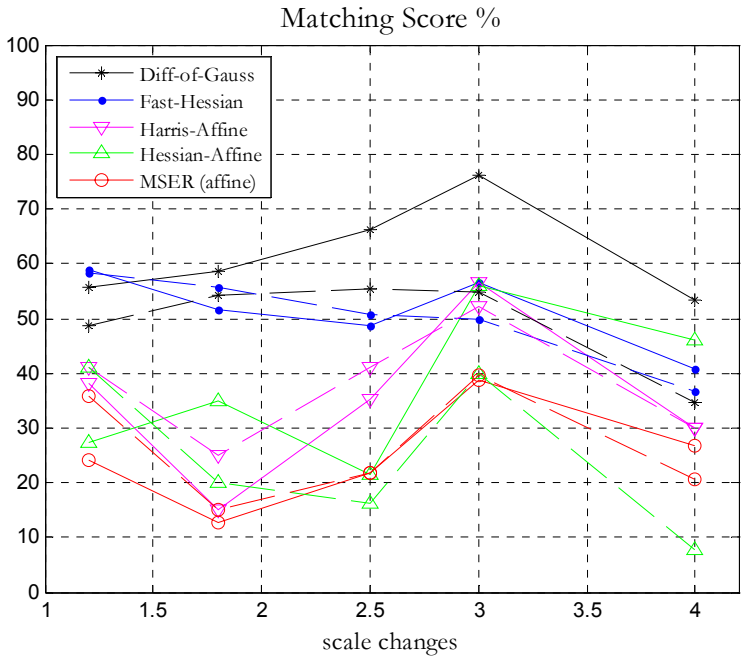



(β)

**Σχ. 3.12:** Το μέτρο ταιριάσματος (% ποσοστό) για τις ακολουθίες εικόνων (α) Graffiti και (β) Wall με μετασχηματισμό της οπτικής γωνίας (viewpoint change)

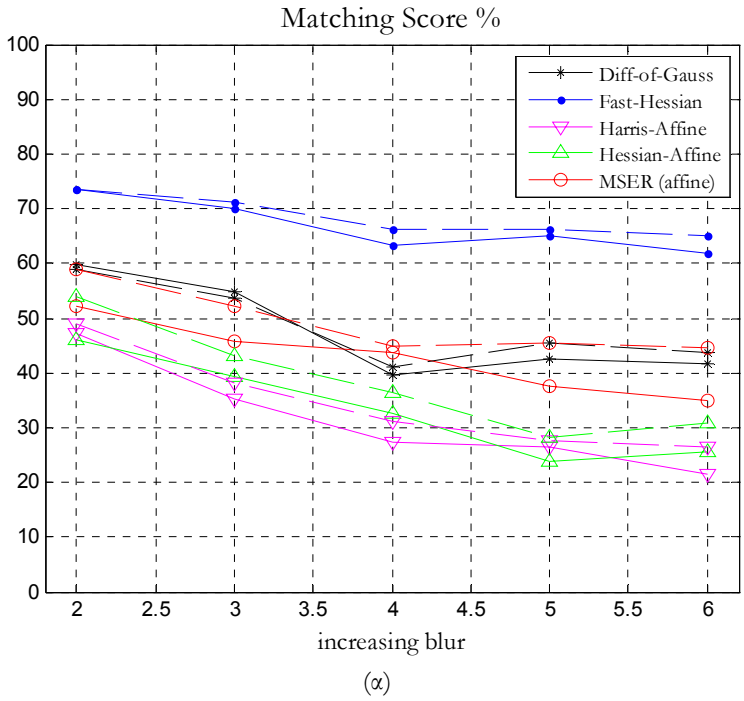



(α)

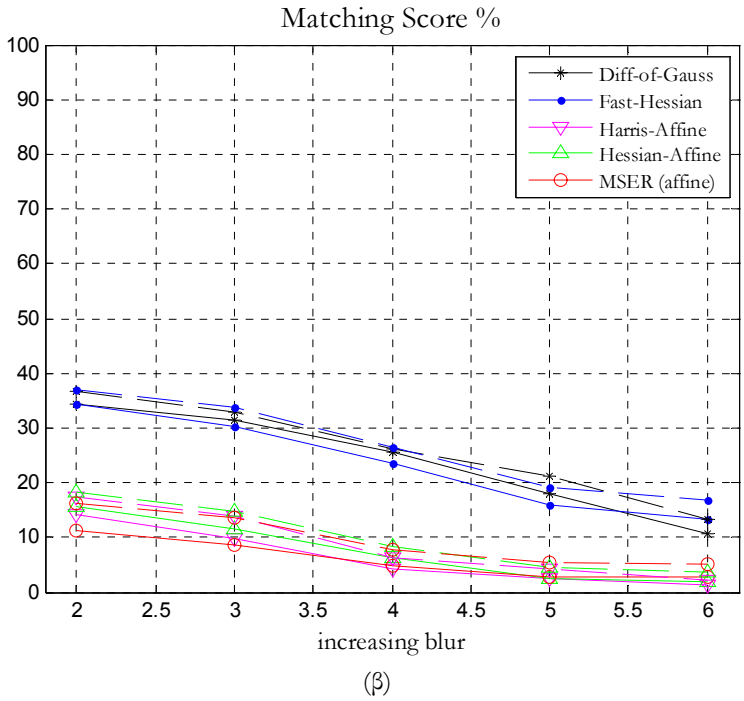



(β)

**Σχ. 3.13:** Το μέτρο ταιριάσματος (% ποσοστό) για τις ακολουθίες εικόνων (α) Boat και (β) Bark με περιστροφή και αλλαγή κλίμακας (scale change)

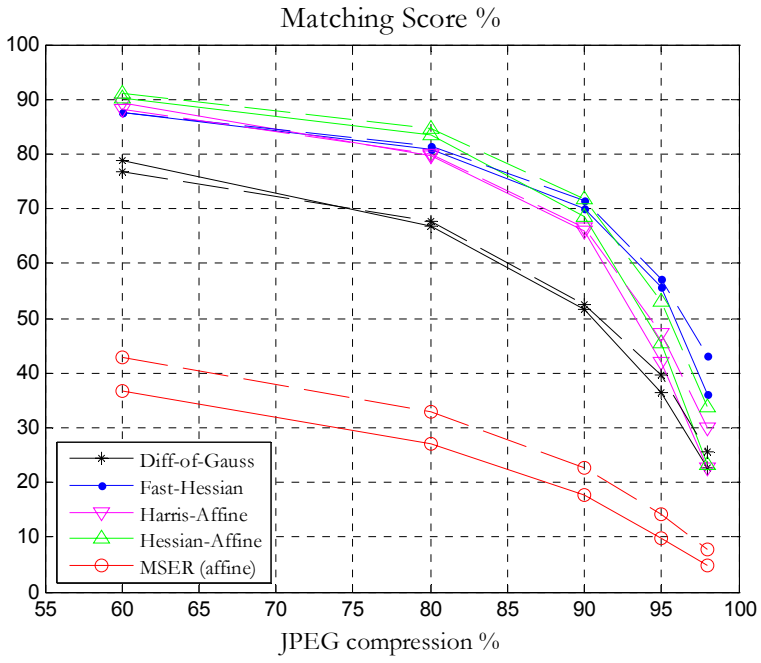



(α)

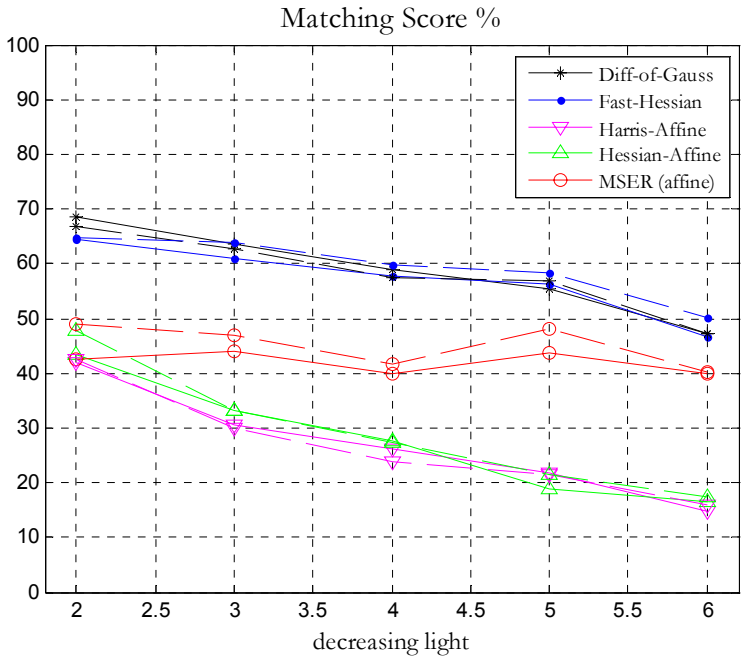



(β)

**Σχ. 3.14:** Το μέτρο ταιριάσματος (% ποσοστό) για τις ακολουθίες εικόνων (α) Bikes και (β) Trees με διαδοχικό θόλωμα της εικόνας (increasing blur)




(α)




(β)

**Σχ. 3.15:** Το μέτρο ταιριάσματος (% ποσοστό) για τις ακολουθίες εικόνων (α) UBC και (β) Leuven με συμπίεση JPEG και διαδοχική μείωση της φωτεινότητας (decreasing light) αντίστοιχα

### 3.3.4 Σύγκριση ως προς την επίδοση των περιγραφέων

Στη συνέχεια των πειραμάτων και για την ολοκλήρωση της σύγκριση των μεθόδων, θα επιχειρηθεί η εκτίμηση της επίδοσης των δύο περιγραφέων SIFT και SURF σύμφωνα με τον τρόπο που έχει προταθεί στα [Mikolajczyk et al., 2003], [Mikolajczyk et al., 2005a]. Το κριτήριο βασίζεται στο πλήθος των σωστών και των εσφαλμένων ταιριασμάτων (correct and false matches). Έστω οι δύο εικόνες A και B, οι οποίες αναπαριστούν την ίδια σκηνή υπό διαφορετικές γεωμετρικές ή φωτομετρικές συνθήκες. Όπως ήδη είπαμε, δύο περιοχές $R_A$ και $R_B$ από τις δύο εικόνες (αντίστοιχα) θα θεωρούνται ότι ταιριάζουν αν η απόσταση μεταξύ των περιγραφέων τους $D_A$ και $D_B$ (αντίστοιχα) είναι μικρότερη από ένα δεδομένο κατώφλι t. Οι περιγραφείς είναι σημεία που απεικονίζονται στον χώρο των χαρακτηριστικών και η σχετική τους θέση μετράται μέσω της ευκλείδειας απόστασης. Αυτό το κριτήριο ονομάζεται κριτήριο ομοιότητας-κατωφλίου, αλλά μπορούν να χρησιμοποιηθούν και άλλα κριτήρια (συνήθως πιο αυστηρά), όπως το κριτήριο του κοντινότερου γείτονα ή το κριτήριο κοντινότερου γείτονα και επιπλέον συνθήκη για την απόσταση με τον δεύτερο κοντινότερο γείτονα. Εδώ θα χρησιμοποιηθεί το πρώτο κριτήριο ομοιότητας-κατωφλίου, από το οποίο προκύπτει ότι κάθε περιοχή μπορεί να έχει περισσότερα από ένα ταιριάσματα στην άλλη εικόνα.

Ακριβώς όπως και πριν, οι περιοχές που αντιστοιχούν μεταξύ τους αλλά και τα σωστά ταιριάσματα των περιοχών ορίζονται μέσω της σχέσης 3.3 για το σφάλμα επικάλυψης. Στην παράγραφο αυτή χρησιμοποιείται σφάλμα $\varepsilon_O$ ίσο με 50% ώστε να καθοριστούν οι αντίστοιχες περιοχές (corresponding regions). Ένα ταίριασμα μιας συγκεκριμένης περιοχής $R_A$ με μια περιοχή $R_B$ θα θεωρείται ότι είναι σωστό (correct match), αν το σφάλμα επικάλυψης είναι ελάχιστο για αυτές τις δύο, σε σχέση με κάθε άλλο ζεύγος που δίνει σφάλμα μικρότερο από 50%. Τα υπόλοιπα ταιριάσματα θεωρούνται ως λανθασμένα (false matches). Για τη σύγκριση των επιδόσεων χρησιμοποιούνται τα μέτρα ανάκτησης και ακρίβειας, τα οποία ορίζονται στη συνέχεια [Mikolajczyk et al., 2005a].

Ως μέτρο ανάκτησης μεταξύ δύο εικόνων της ίδιας σκηνής ορίζεται το πλήθος των σωστών ταιριασμάτων των περιοχών τους προς το πλήθος των αντίστοιχων περιοχών τους:

$$recall = \frac{\#correct\ matches}{\#correspondences} \qquad \textbf{(εξ. 3.7)}$$

Ως μέτρο ακριβείας μεταξύ δύο εικόνων της ίδιας σκηνής ορίζεται το πλήθος των σωστών ταιριασμάτων των περιοχών τους προς το πλήθος των συνολικών ταιριασμάτων (σωστών και εσφαλμένων μαζί):

$$precision = \frac{\#correct\ matches}{\#total\ matches} \qquad \textbf{(εξ. 3.8)}$$

Για να κατασκευαστούν τα διαγράμματα στα σχήματα που ακολουθούν, μεταβάλλεται το κατώφλι t που δίνει τα ταιριάσματα των περιοχών μεταξύ των δύο εικόνων. Έτσι αυξάνονται και τα συνολικά ταιριάσματα (και το μέτρο ανάκτησης), αλλά αυξάνεται και η πιθανότητα θορύβου, αφού συμπεριλαμβάνονται και περιοχές οι οποίες δεν έχουν το ίδιο περιεχόμενο. Λαμβάνονται λοιπόν πολλές διαφορετικές τιμές που συμβάλλουν τελικά στην καμπύλη ανάκτησης – ακριβείας όπως θα φανεί στα σχήματα. Η καμπύλη αυτή για τον ιδανικό περιγραφέα θα είχε μέτρο ανάκτησης ίσο με ένα για όλες τις τιμές ακριβείας. Άρα καλύτερος θεωρείται ένας περιγραφέας που έχει όσο το δυνατόν πιο υψηλό μέτρο ανάκτησης. Αν ο μέτρο ανάκτησης ακολουθεί οριζόντια καμπύλη, τότε αυτό σημαίνει ότι επιτυγχάνεται πολύ καλή ανάκτηση με πολύ καλή ακρίβεια. Αντίθετα, όταν το μέτρο ανάκτησης είναι υψηλό μόνο για μικρές τιμές στην ακρίβεια

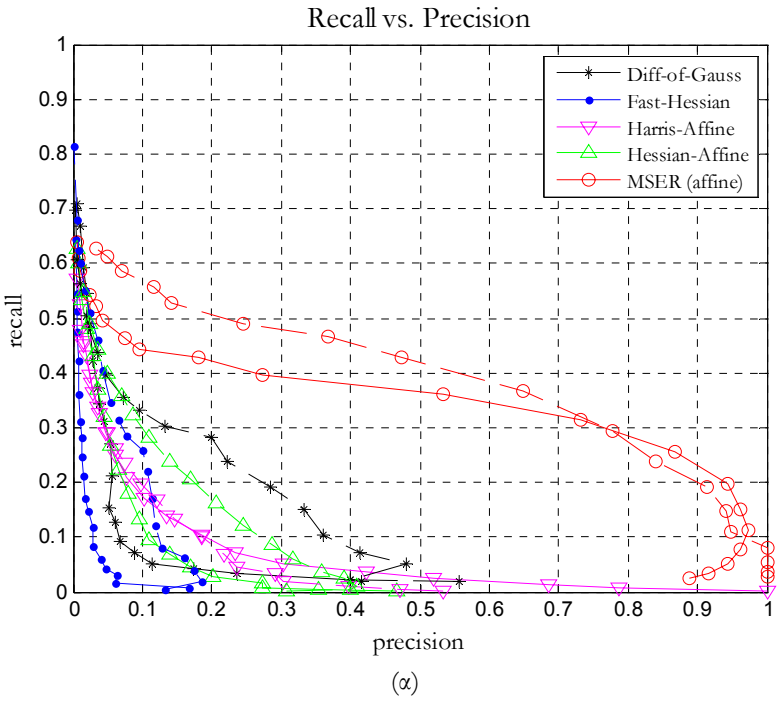
και μετά μειώνεται, αυτό σημαίνει ότι απαιτούνται πάρα πολλά ταιριάσματα για να εντοπιστούν τα σωστά και άρα πρακτικά ο περιγραφέας δεν έχει μεγάλη διακριτική ικανότητα.

Τα σχήματα 3.16 – 3.19 παρουσιάζονται οι καμπύλες ανάκτησης-ακριβείας για τη σύγκριση των δύο περιγραφέων SIFT και SURF, όπως αυτοί έχουν υπολογιστεί πάνω στις περιοχές που ανιχνεύονται ξεχωριστά από κάθε μέθοδο, μεταξύ της πρώτης και της τέταρτης εικόνας από κάθε ακολουθία. Η καμπύλη που αφορά τον περιγραφέα SIFT παριστάνεται με συνεχόμενη γραμμή ενώ η καμπύλη του περιγραφέα SURF παριστάνεται με διακεκομμένη γραμμή. Όπως είναι φυσικό, αφού οι δύο τύποι περιγραφέων μοντελοποιούν με διαφορετικό τρόπο την ίδια πληροφορία (κατευθύνσεις των τοπικών παραγώγων της έντασης), τα διαγράμματα αναμένεται να είναι παρόμοια. Θα σχολιάσουμε αυτά τα διαγράμματα παρατηρώντας την επίδοση για κάθε περιγραφέα συνολικά και όχι ειδικά για κάθε είδος μετασχηματισμού.

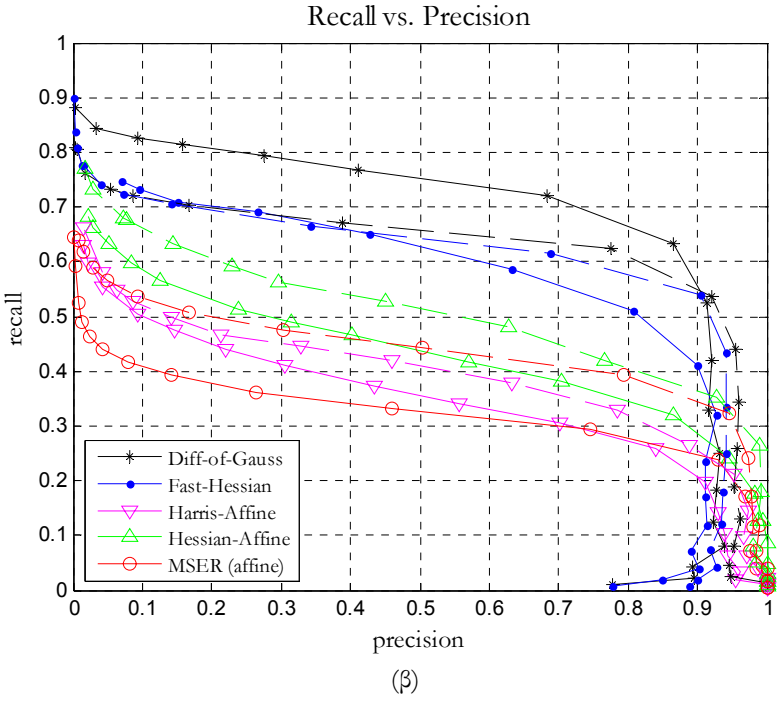
Ήδη από τα διαγράμματα των σχημάτων 3.12 – 3.15 μπορούμε να εξάγουμε συμπεράσματα για την απόδοση των περιγραφέων σε σχέση με το είδος του ανιχνευτή που χρησιμοποιείται. Εύκολα μπορούμε να δούμε ότι γενικότερα για όλες τις μεθόδους και για όλες τις πειραματικές ακολουθίες, ο περιγραφέας SURF δίνει υψηλότερα ποσοστά ταιριάσματος, εκτός από τις περιπτώσεις όπου ο ανιχνευτής DoG εμφανίζεται καλύτερος από τους υπόλοιπους (τουλάχιστον από τον Fast Hessian) και στην οποία τα ποσοστά είναι καλύτερα για τον περιγραφέα SIFT. Στα διαγράμματα ανάκτησης-ακριβείας δεν θα μελετήσουμε όλες τις καμπύλες, παρά μονό αυτές για τις οποίες τα ποσοστά ταιριάσματος για τους ανιχνευτές ήταν από τα πρώτα στα αντίστοιχα διαγράμματα.

Για τον περιγραφέα SIFT βλέπουμε στο δεύτερο διάγραμμα του σχήματος 3.16 και συγκεκριμένα για τον ανιχνευτή DoG (μαύρο χρώμα) ότι οι τιμές ανάκτησης είναι μεγαλύτερες σε σχέση με εκείνες που δίνει ο περιγραφέας SURF. Μία άλλη περίπτωση που θα πρέπει να μελετήσουμε τον περιγραφέα SIFT είναι το σχήμα 3.17 στο οποίο και πάλι για τον ανιχνευτή DoG οι καμπύλες ανάκτησης-ακριβείας είναι καλύτερες. Άλλη μία περίπτωση όπου ο SIFT παρουσιάζεται προνομιούχος είναι η ακολουθία Bikes στο σχήμα 3.18. Αυτό σημαίνει ότι συνολικά αν κάποιος θέλει να ταιριάξει εικόνες που περιέχουν σκηνές και μετασχηματισμούς όπως textured σκηνή και αλλαγή στην οπτική γωνία, ή structured σκηνή και θόλωμα, ή ακόμα εικόνες με περιστροφή και αλλαγή κλίμακας ανεξαρτήτως τύπου σκηνής, ο συνδυασμός ανιχνευτή DoG και περιγραφέα SIFT καλύπτει μεγάλο μέρος των αναγκών του προβλήματος.

Στη συνέχεια παρατηρούμε για τον περιγραφέα SURF ότι δίνει γενικά καλύτερες καμπύλες απ' ότι ο SIFT για τον ίδιο ανιχνευτή, παρόλο που οι τιμές του προκύπτει με προσεγγιστικό τρόπο. Πιο συγκεκριμένα, στην περίπτωση του Graffiti όπου η τεχνική MSER φαίνεται κατάλληλη για εντοπισμό αφινικών περιοχών, το διάνυσμα SURF δίνει μεγαλύτερες πιθανότητες για σωστό ταίριασμα, όπως συμβαίνει και για τον ανιχνευτή Fast Hessian στις εικόνες Wall αλλά και στις εικόνες Boat, με περιστροφή και αλλαγή κλίμακας. Επίσης, για την ακολουθία Trees με textured σκηνή και διαδοχικό θόλωμα, όπου οι αφινικές μέθοδοι αποτυγχάνουν, το διάνυσμα χαρακτηριστικών SURF υπολογισμένο σε σημεία του ανιχνευτή DoG ή του ανιχνευτή Fast Hessian, δίνει ένα ικανοποιητικό μέτρο ανάκτησης (ακόμα και για μεγαλύτερες τιμές ακρίβειας). Τέλος, στις περιπτώσεις της συμπίεσης JPEG και της μεταβολής στη φωτεινότητα, η ανάκτηση είναι και πάλι πολύ καλή για τα σημεία DoG και Fast Hessian. Ακόμα και για τις υπόλοιπες μεθόδους ανίχνευσης, που δεν ακολουθούν οριζόντιες καμπύλες, ο περιγραφέας SURF δίνει ελαφρώς μεγαλύτερες τιμές ανάκτησης από τον SIFT, γεγονός που μπορεί να παίξει σημαντικό ρόλο σε ένα πρόβλημα ταιριάσματος, όταν τα σημεία ενδιαφέροντος είναι λίγα και τα σωστά ταιριάσματα περιοχών ακόμα πιο περιορισμένα.

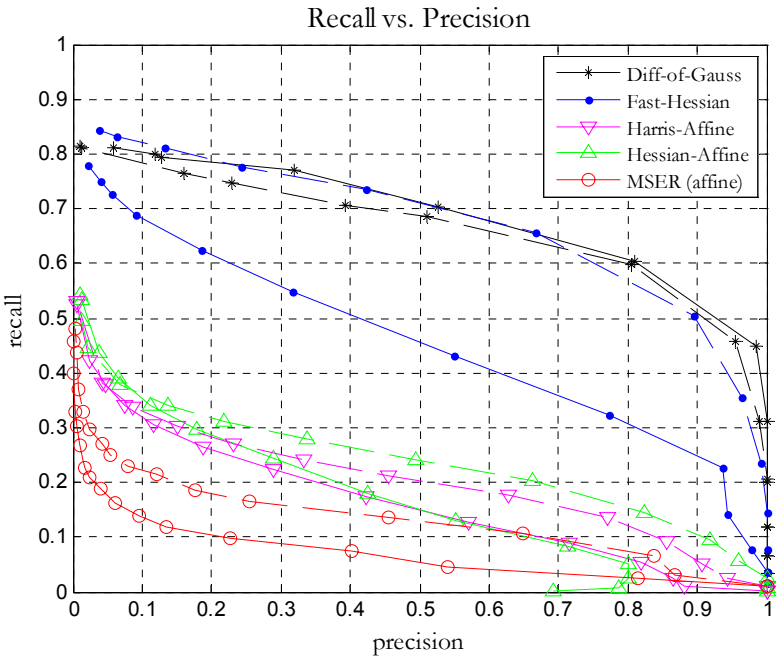



(α)

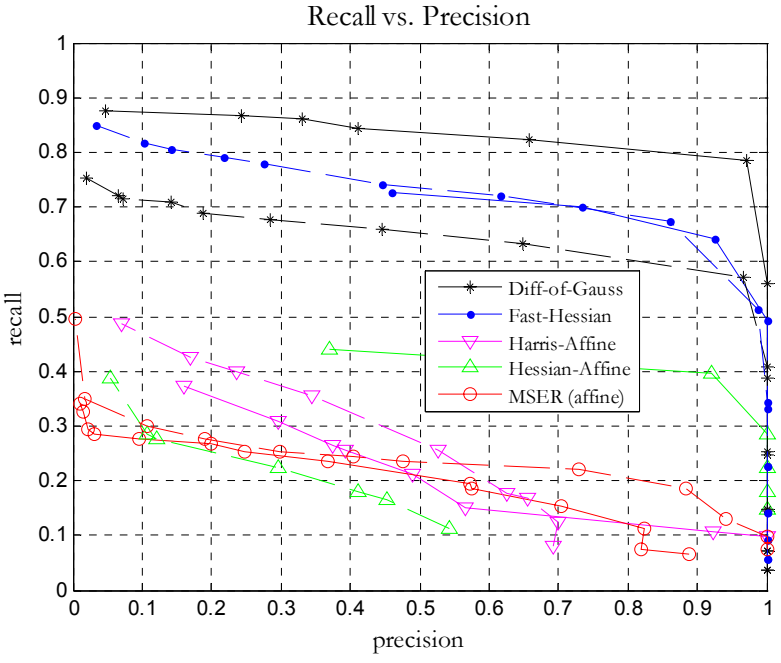



(β)

**Σχ. 3.16:** Καμπύλες ανάκτησης-ακριβείας των περιγραφέων SIFT- SURF -- για τις ακολουθίες εικόνων (α) Graffiti και (β) Wall με μετασχηματισμό της οπτικής γωνίας (viewpoint change)

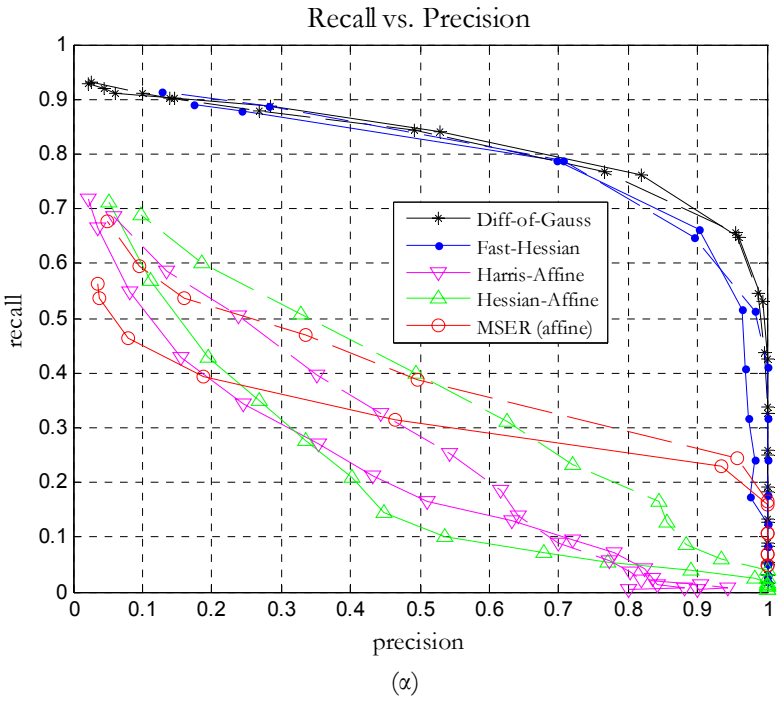



(α)

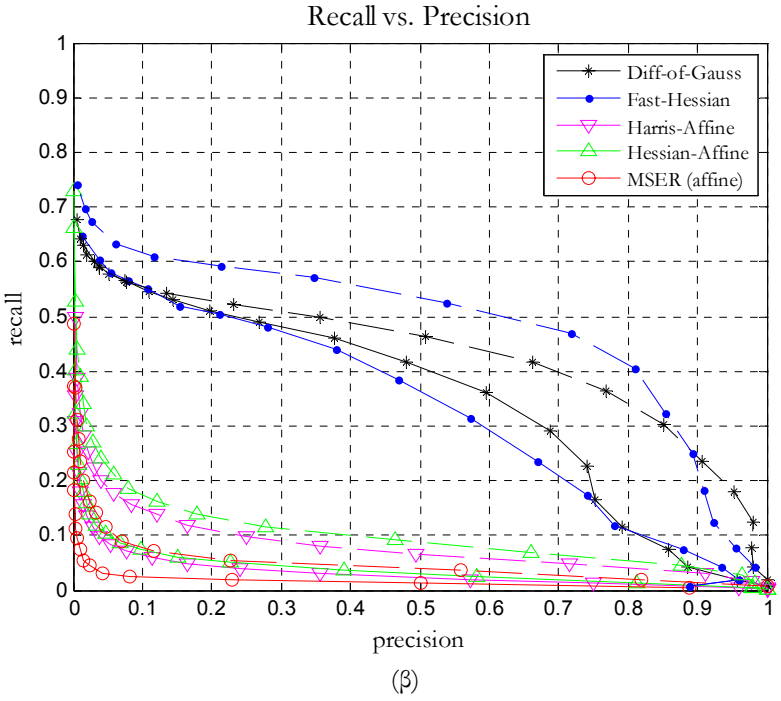



(β)

**Σχ. 3.17:** Καμπύλες ανάκτησης-ακριβείας των περιγραφέων SIFT- SURF -- για τις ακολουθίες εικόνων (α) Boat και (β) Bark με περιστροφή και αλλαγή κλίμακας (scale change)

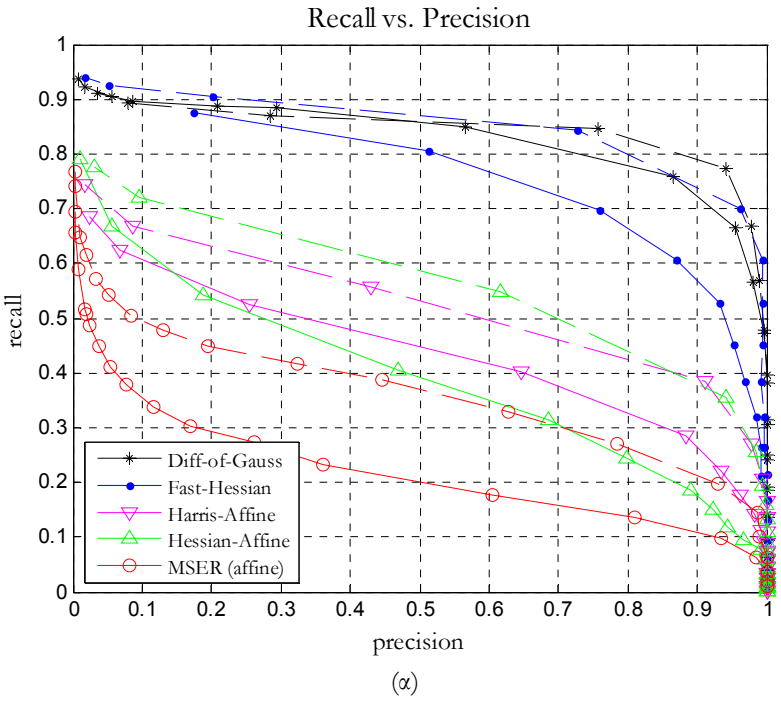



(α)

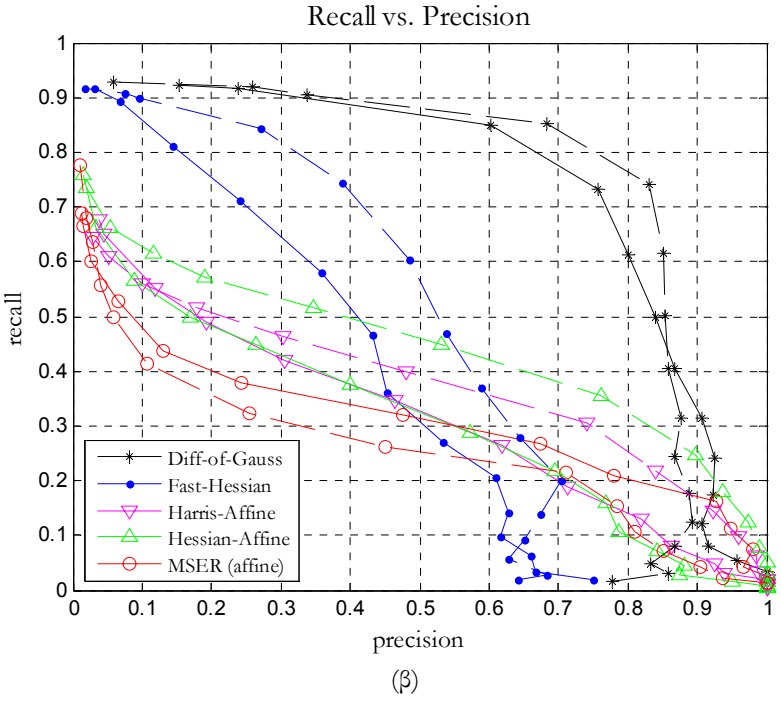



(β)

**Σχ. 3.18:** Καμπύλες ανάκτησης-ακριβείας των περιγραφέων SIFT- SURF -- για τις ακολουθίες εικόνων (α) Bikes και (β) Trees με διαδοχικό θόλωμα της εικόνας (increasing blur)




(α)




(β)

**Σχ. 3.19:** Καμπύλες ανάκτησης-ακριβείας των περιγραφέων SIFT- SURF -- για τις ακολουθίες εικόνων (α) UBC και (β) Leuven με συμπίεση JPEG και διαδοχική μείωση της φωτεινότητας (decreasing light) αντίστοιχα

## 3.4 Γενικά συμπεράσματα

Από ολόκληρη την πειραματική διαδικασία που προηγήθηκε μπορούν να εξαχθούν κάποια γενικά συμπεράσματα ώστε οι μέθοδοι τοπικών ανιχνευτών να συνδυαστούν με τον κατάλληλο περιγραφέα και να χρησιμοποιηθούν σε κάποιο πρακτικό πρόβλημα. Η μέθοδος ανίχνευσης περιοχών ενδιαφέροντος MSER έχει υψηλά ποσοστά στο μέτρο επαναληψιμότητας σε σκηνές που περιλαμβάνουν αλλαγή στην οπτική γωνία, και μάλιστα με πολύ καλή ακρίβεια. Σε συνδυασμό με ένα διάνυσμα χαρακτηριστικών, φαίνεται ότι μπορεί να ανταποκριθεί ακόμα και σε πολύ μεγάλες γωνίες λήψης. Αυτό εξηγείται από τη φύση της μεθόδου, αφού είναι κατασκευασμένη ώστε να εντοπίζει περιοχές ανεξάρτητα από αφινικές μεταβολές. Επίσης, η μέθοδος ανταποκρίνεται πολύ καλά σε σκηνές με δομή (structured), αφού σε αυτές υπάρχουν ομοιογενείς περιοχές και ευδιάκριτα όρια μεταξύ των αντικειμένων (Graffiti), οπότε ο εντοπισμός των πιο σταθερών συνεκτικών συνιστωσών αντιστοιχεί σε πλήρη περιγραφή του περιεχομένου. Αντίθετα, σε σκηνές με υφή (Wall) η μέθοδος MSER παρουσιάζει πολύ χαμηλότερες επιδόσεις, διότι εκεί δεν υπάρχουν ομοιογενείς περιοχές και η έξοδος του ανιχνευτή είναι πολλές μικρές ελλείψεις που μοιάζουν μεταξύ τους ως προς το περιεχόμενο και άρα δεν μπορούν να ταιριάξουν με τον περιγραφέα.

Οι μέθοδοι Harris-Hessian Affine εξάγουν πάρα πολλά σημεία σε σχέση με τις υπόλοιπες, γεγονός που είναι επιθυμητό σε θέματα ανάκτησης και αναγνώρισης αντικειμένων (αφού για παράδειγμα αντιμετωπίζεται έτσι το πρόβλημα της μερικής επικάλυψης), αλλά συνήθως μειώνει την ακρίβεια της μεθόδου (πολλά false matches). Οι δύο αυτές μέθοδοι ανταποκρίνονται καλύτερα σε εικόνες με συμπίεση JPEG και σε μεταβολή της γωνίας λήψης. Παρόλ' αυτά η περιοχή τους δεν είναι χαρακτηριστική σε ικανοποιητικό βαθμό ώστε ο περιγραφέας της να μπορεί να χρησιμοποιηθεί για το ταίριασμα με την αντίστοιχη περιοχή μιας δεύτερης εικόνας (βλέπε διαγράμματα ταιριάσματος). Επίσης, πρέπει να σημειώσουμε ότι η ακρίβεια των τεχνικών αυτών γενικά είναι οι χειρότερη μεταξύ των μεθόδων, όπως φαίνεται από τα αντίστοιχα διαγράμματα. Οι υψηλές τιμές επαναληψιμότητας δεν οφείλονται στην ακρίβειά τους, αλλά στο γεγονός ότι εξάγουν πάρα πολλές ελλείψεις αυξάνοντας έτσι την πιθανότητα επικάλυψης.

Οι μέθοδοι DoG και Fast Hessian έχουν την καλύτερη ακρίβεια μεταξύ των υπολοίπων. Εντοπίζουν δηλαδή τα σημεία που αντιστοιχούν με τα ίδια τοπικά χαρακτηριστικά (κηλίδες) σε δύο διαφορετικές όψεις της ίδιας σκηνής και δεν επηρεάζονται από τις μεταβολές της εικόνας. Οι περιγραφείς που εξάγονται γύρω από αυτά εμφανίζουν μεγάλα ποσοστά στην επαναληψιμότητα και στο μέτρο ταιριάσματος, ειδικά σε περιπτώσεις εικόνων με αλλαγές κλίμακας. Η μόνη περίπτωση που δεν μπορούν να ανταγωνιστούν την επιτυχία ταιριάσματος MSER είναι όταν συμβαίνουν πολύ μεγάλες αλλαγές στην οπτική γωνία (περίπου $50^o$-$60^o$), και αυτό διότι δεν είναι ανεξάρτητες από αφινικές μεταβολές.

Έτσι, για να κλείσουμε το κεφάλαιο με μια γενική παρατήρηση, στην περίπτωση που στο πρόβλημά μας περιέχονται εικόνες που έχουν υποστεί περιστροφή, αλλαγή κλίμακας ή θόλωμα, είτε ακόμα εικόνες με σκηνές υφής, τότε είναι προτιμότερο να χρησιμοποιηθεί ο συνδυασμός ανιχνευτή σημείων DoG με περιγραφέα SIFT ή ο ανιχνευτή Fast Hessian με περιγραφέα SURF. Αν στις εικόνες του προβλήματος συμβαίνουν αλλαγές στην οπτική γωνία (αφινικές μεταβολές), τότε μπορεί να χρησιμοποιηθούν MSER περιοχές με διάνυσμα SURF και το αποτέλεσμα σύμφωνα με τα πειραματικά δεδομένα θα είναι ικανοποιητικό. Αυτοί είναι και οι τρεις συνδυασμοί που θα χρησιμοποιηθούν στα πειράματα του επόμενου κεφαλαίου σε ανάκτηση εικόνων από μεγάλες βάσεις δεδομένων.

# ΚΕΦΑΛΑΙΟ 4

# Αναζήτηση Εικόνων μέσω Τοπικών Χαρακτηριστικών

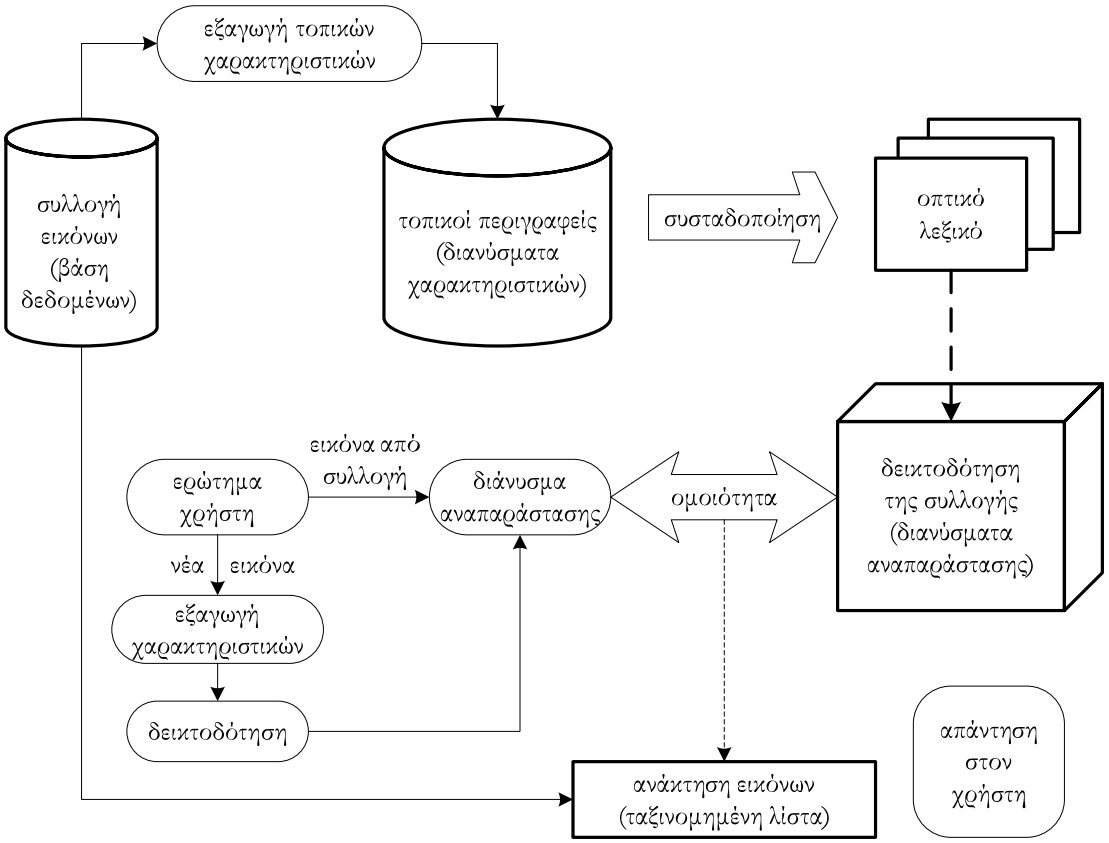
## 4.1 Εισαγωγικά

Τα τοπικά χαρακτηριστικά παρουσιάζουν πολλά πλεονεκτήματα σε σχέση με άλλες εναλλακτικές μεθόδους περιγραφής του περιεχομένου της εικόνας, όπως τα χαρακτηριστικά από ολόκληρη την εικόνα (περιγραφείς MPEG7, ιστόγραμμα). Καταρχάς με τη χρήση τους αποφεύγεται η διαδικασία της κατάτμησης της εικόνας σε αντικείμενα, η οποία είναι υπολογιστικά πολύπλοκη και απαιτεί ανώτερου επιπέδου γνώση η οποία δεν είναι πάντα διαθέσιμη. Τα σημεία ενδιαφέροντος παρουσιάζουν υψηλά ποσοστά επαναληψιμότητας, δηλαδή μεγάλο ποσοστό τους εντοπίζεται επιτυχώς σε εικόνες που αφορούν την ίδια σκηνή υπό διαφορετικές όψεις ή συνθήκες φωτισμού. Με μια κατάλληλη τοπική περιγραφή όπως είναι το ιστόγραμμα των παραγώγων στη γειτονιά ενδιαφέροντος τα τοπικά χαρακτηριστικά αποτελούν ισχυρό εργαλείο για να εντοπίζονται τα ζητούμενα αντικείμενα ακόμα κι αν η εικόνα έχει υποστεί ακραίους μετασχηματισμούς (περιστροφή, αλλαγή κλίμακας ή οπτικής γωνίας κτλ.). Επειδή σε ένα αντικείμενο ανιχνεύονται συνήθως περισσότερα από ένα σημεία ενδιαφέροντος, η μερική επικάλυψη του αντικειμένου σε κάποια εικόνα δεν θα εμποδίσει το ταίριασμα για τους περιγραφείς του ορατού του μέρους, παρά την παρουσία άλλων "περιττών" αντικειμένων. Η αποτελεσματικότητα αλλά και η σύγκριση μεταξύ των μεθόδων τοπικών χαρακτηριστικών θα εξεταστεί σε ένα σύστημα αναζήτησης εικόνων από μεγάλες βάσεις δεδομένων.

Το σύστημα αναζήτησης που θα χρησιμοποιηθεί έχει πολλές ομοιότητες με τις τεχνικές αναζήτησης κειμένου που είναι ευρέως διαδεδομένες στο διαδίκτυο (βλέπε Google). Στις τεχνικές αυτές η αναζήτηση γίνεται με βάση λεξικογραφικούς όρους και το αποτέλεσμά της εξάγεται πολύ γρήγορα με ένα σύστημα ανεστραμμένων αρχείων. Η απάντηση προς τον χρήστη είναι μία λίστα από έγγραφα (ιστοσελίδες, στην περίπτωση του Google) ταξινομημένη με βάση τη σχετικότητα του περιεχομένου τους. Από την άλλη, ένα αντικείμενο σε μια εικόνα μπορεί να περιγραφεί με τα τοπικά του χαρακτηριστικά (σημεία ή περιοχές) και το κατάλληλο διάνυσμα χαρακτηριστικών (περιγραφέας). Κάθε εικόνα της βάσης δεδομένων αναπαριστάται από όλους τους τοπικούς της περιγραφείς, δηλαδή από ένα σύνολο σημείων στον χώρο των χαρακτηριστικών. Η αναζήτηση των αντίστοιχων αντικειμένων μεταξύ των εικόνων μπορεί να γίνει μέσω της απόστασης των σημείων των περιγραφέων τους. Σε αυτό το στάδιο μελετάται η δυνατότητα εφαρμογής του λεκτικού τρόπου αναζήτησης για την αναγνώριση του αντικειμένου, και για να γίνει αυτό θα πρέπει να βρεθεί ένας ανάλογος λεκτικός όρος.

Το περιεχόμενο της εικόνας αναπαριστάται με όλους τους περιγραφείς της, άρα το λεκτικό ανάλογο θα μπορούσε να προκύψει με την κβαντοποίηση των διανυσμάτων χαρακτηριστικών [Sivic et al., 2003]. Ομαδοποιώντας τα όμοια χαρακτηριστικά προκύπτουν οι οπτικοί όροι, σε πλήρη αντιστοιχία με τους λεκτικούς όρους ενός κειμένου, όπου πολλές παρόμοιες λέξεις αντιμετωπίζονται με μία γενική έννοια, όπως για παράδειγμα οι λέξεις "περπατάω, περπάτημα, περιπατητής" αντικαθίστανται από την κοινή τους ρίζα "περπατάω". Έτσι έχοντας εκ των προτέρων υπολογίσει και ομαδοποιήσει (με κάποιο αλγόριθμο συσταδοποίησης) τα διανύσματα των περιγραφέων, δημιουργείται ένα οπτικό λεξικό, σύμφωνα με το οποίο κάθε εικόνα της συλλογής θα αναπαρασταθεί από τις οπτικές λέξεις φράσεις τις οποίες περιέχει. Στο στάδιο της δεικτοδότησης του περιεχομένου κατασκευάζεται για κάθε έγγραφο ένα διάνυσμα αναπαράστασης, το οποίο περιέχει τους οπτικούς όρους και τη συχνότητα με την οποία αυτοί εμφανίζονται μέσα στην εικόνα. Στα στοιχεία του διανύσματος αυτού μπορούν να προστεθούν βάρη, όπως για παράδειγμα ένας συντελεστής αντίστροφης συχνότητας εμφάνισης ώστε να μειωθεί η επίδραση των λέξεων που εμφανίζονται συχνά σε ολόκληρη τη συλλογή. Επιπλέον των βαρών αυτών, μπορεί να εφαρμοστεί και μια λίστα τερματισμού (stop list), η οποία θα απορρίπτει οπτικές λέξεις που εμφανίζονται συχνά στις περισσότερες εικόνες και άρα δεν βοηθούν στη διάκριση μεταξύ του περιεχομένου (όπως αντίστοιχα στη μηχανή αναζήτησης Google απορρίπτονται οι πολύ κοινές λέξεις "και, αν, οι").

Τέλος, τα διανύσματα αναπαράστασης για όλη τη συλλογή των εικόνων οργανώνονται σε ένα ανεστραμμένο αρχείο. Ως ανεστραμμένο αρχείο ορίζεται μία ταξινομημένη δομή των οπτικών λέξεων του λεξικού, δηλαδή ένας τύπος ευρετηρίου στο οποίο κάθε καταχώρηση περιλαμβάνει και μία λίστα των εγγράφων-εικόνων μέσα στα οποία εμφανίζονται αυτές οι λέξεις (περιγραφείς) και την ακριβή θέση στην οποία εμφανίζονται. Σε κάθε ερώτημα του χρήστη (παράδειγμα εικόνας) υπολογίζεται το διάνυσμα αναπαράστασης (αν δεν έχει ήδη υπολογιστεί για την εικόνα) και στη συνέχεια εκτιμάται η ομοιότητά του με κάθε διάνυσμα της δεικτοδοτημένης συλλογής. Έχοντας υπολογίσει από πριν τα διανύσματα αναπαράστασης για ολόκληρη τη συλλογή και με τη χρήση των ανεστραμμένων αρχείων (inverted files) η ανάκτηση των πιο κοντινών εικόνων γίνεται πολύ γρήγορα, και μάλιστα επιστρέφεται ως απάντηση μια λίστα με σειρά σχετικότητας, από την πιο συναφή στην πιο ασήμαντη. Μια γενική δομή του συστήματος αναζήτησης εικόνων φαίνεται στο σχήμα 4.1.




**Σχ. 4.1:** Γενικό διάγραμμα του συστήματος αναζήτησης εικόνων

## 4.2 Περιγραφή του συστήματος αναζήτησης εικόνων

### 4.2.1 Εξαγωγή τοπικών περιγραφέων

Το πρώτο βήμα στο σύστημα αναζήτησης εικόνων είναι η ανίχνευση των τοπικών χαρακτηριστικών και η εξαγωγή των περιγραφέων τους, δηλαδή των σημείων στο χώρο των χαρακτηριστικών που αναπαριστούν το σημασιολογικό περιεχόμενο της εικόνας. Χρησιμοποιούνται οι μέθοδοι που αναλύθηκαν και συγκρίθηκαν στα προηγούμενα κεφάλαια, και συγκεκριμένα οι παρακάτω τρεις συνδυασμοί μεθόδων ανιχνευτών και περιγραφέων:

- ανίχνευση σημείων Difference of Gaussian και εξαγωγή περιγραφέα SIFT
- ανίχνευση σημείων Fast Hessian και εξαγωγή περιγραφέα SURF
- ανίχνευση περιοχών MSER και εξαγωγή περιγραφέα SURF

Από εδώ και στο εξής θα αναφερόμαστε με τον όρο DoG-SIFT για τον πρώτο συνδυασμό, με τον όρο FastH-SURF για τον δεύτερο και με τον όρο MSER-SURF για τον τρίτο. Ένα παράδειγμα με τις αρχικές περιοχές που ανιχνεύονται από τη μέθοδο MSER σε μια τυχαία εικόνα φαίνεται στο σχήμα 4.2 (α), ενώ στο σχήμα 4.2 (β) φαίνονται για την ίδια εικόνα οι περιοχές μέτρησης (αλλαγή κλίμακας επί 3), από τις οποίες εξάγεται ο περιγραφέας SURF.

**Σχ. 4.2:** Παράδειγμα τοπικών χαρακτηριστικών MSER σε εικόνες από τα τρία διαφορετικά σύνολα δεδομένων που χρησιμοποιούνται στα πειράματα

Κάθε εικόνα μέσω των διανυσμάτων χαρακτηριστικών της αναπαριστάται από ένα σύνολο σημείων στον χώρο των χαρακτηριστικών, ο οποίος έχει διάσταση ίση με 128 για την περίπτωση του περιγραφέα SIFT και 64 για την περίπτωση του περιγραφέα SURF. Γίνεται δηλαδή ένας μετασχηματισμός του περιεχομένου από το πεδίο των pixels της εικόνας (διδιάστατη συνάρτηση της έντασης) στον πολυδιάστατο χώρο των χαρακτηριστικών.

Στη συνέχεια, για να συγκριθούν δύο εικόνες αρκεί να συγκριθούν όλα τα διανύσματα χαρακτηριστικών τους ένα προς ένα. Αυτή η διαδικασία μπορεί να γίνει αναζητώντας τον κοντινότερο γείτονα για κάθε διάνυσμα ενός μέτρου για την απόσταση στον χώρο αυτόν. Ένας τρόπος είναι η χρήση της ευκλείδειας απόστασης:

$$d_E(d_1, d_2) = \sum_{i=1}^{D} \left[ d_1(i) - d_2(i) \right]^{1/2} \qquad \textbf{(εξ. 4.1)}$$

όπου $d_1$, $d_2$ είναι δύο διανύσματα στον χώρο των χαρακτηριστικών, ο οποίος έχει διάσταση D.

Ένας άλλος τρόπος υπολογισμού της απόστασης δύο διανυσμάτων είναι η απόσταση Mahalanobis με τη βοήθεια ενός πίνακα συνδιακύμανσης Σ. Ο πίνακας αυτός υπολογίζεται από ολόκληρο το σύνολο δεδομένων και συμβάλλει ώστε να περιοριστούν τα στοιχεία των περιγραφέων με τον περισσότερο θόρυβο και να αφαιρεθεί η μικρή συσχέτιση που τυχόν υπάρχει. Η απόσταση Mahalanobis δίνεται από τον τύπο:

$$d_M(d_1, d_2) = \sqrt{(d_1 - d_2)^T \cdot \Sigma^{-1} \cdot (d_1 - d_2)} \qquad \textbf{(εξ. 4.2)}$$

Έχοντας υπολογίσει τα πιο κοντινά διανύσματα μεταξύ των δύο εικόνων, η αναγνώριση του ίδιου αντικειμένου ή της ίδιας σκηνής γίνεται με ένα κατώφλι t θεωρώντας τις αποστάσεις που είναι μικρότερες από t ως ταίριασμα. Η διαδικασία αυτή γίνεται για ολόκληρη τη συλλογή και οι εικόνες με τα πιο πολλά ταιριάσματα θεωρείται ότι περιέχουν το ζητούμενο αντικείμενο. Μπορεί επίσης να χρησιμοποιηθεί εκ των υστέρων και κάποιο κριτήριο χωρικής διάταξης, π.χ. μεταξύ δύο εικόνων τα κοντινά σημεία που ταιριάζουν να ικανοποιούν κάποια εξίσωση μετασχηματισμού (χωρική συνάφεια).

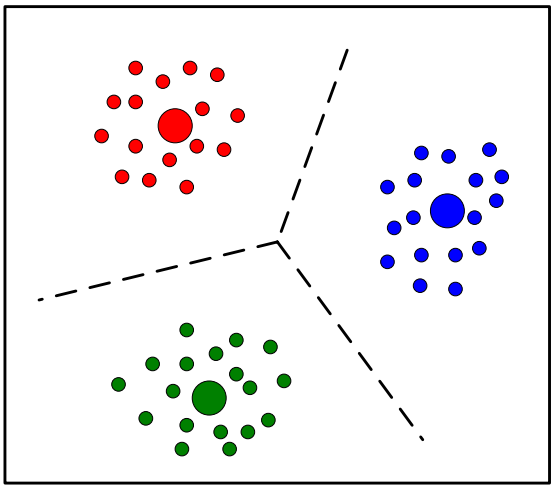
Το μεγαλύτερο πρόβλημα είναι η χρονική πολυπλοκότητα που εισάγεται κατά τη διαδικασία αναζήτησης εικόνων όμοιων με ένα παράδειγμα που δίνει ο χρήστης. Αυτό συμβαίνει διότι πρέπει να υπολογιστούν οι αποστάσεις για όλα τα ζεύγη των εικόνων και για όλους τους συνδυασμούς των περιγραφέων τους (περίπου 1000 διανύσματα περιγραφέων ανά εικόνα). Για ερώτημα με 1 εικόνα και ένα σύνολο δεδομένων από 1000 εικόνες με 1000 σημεία-διανύσματα ανά εικόνα, πρέπει να γίνουν $10^9$ υπολογισμοί αποστάσεων, δηλαδή περίπου $10^9 \times D^2$ πράξεις (D = 64 ή 128). Η πολυπλοκότητα αυτή είναι απαγορευτική για την ταχύτητα που θα έπρεπε να έχει ένα σύστημα αναζήτησης. Γι' αυτό το λόγο αναπτύσσεται η τεχνική με το οπτικό λεξικό στην επόμενη παράγραφο.

### 4.2.2 Δημιουργία οπτικού λεξικού

Τα διανύσματα χαρακτηριστικών κβαντοποιούνται σε πολλές ομάδες οι οποίες αντιστοιχούν στην ίδια "οπτική έννοια". Με τον τρόπο αυτό κάθε διάνυσμα χαρακτηριστικών αντιστοιχίζεται στην κοντινότερη ομάδα και έτσι σε κάθε εικόνα δεν χρειάζεται η πληροφορία για

όλα τα στοιχεία όλων των διανυσμάτων χαρακτηριστικών, παρά μόνο τα δεδομένα για το ποιες οπτικές λέξεις περιέχονται (και πού).

Η κβαντοποίηση του χώρου των χαρακτηριστικών γίνεται με τον αλγόριθμο συσταδοποίησης K-μέσων (K-means clustering), που βασίζεται στην ελαχιστοποίηση του τετραγωνικού σφάλματος ομοιότητας μεταξύ των δεδομένων. Ξεκινώντας από έναν τυχαίο ορισμό των αρχικών k ομάδων (clusters), κάθε σημείο των δεδομένων αντιστοιχίζεται σε μια ομάδα μέσω της ελάχιστης απόστασής τους από τα κέντρα των ομάδων (εδώ χρησιμοποιούμε την ευκλείδεια απόσταση). Στη συνέχεια υπολογίζονται ξανά τα κέντρα των νέων ομάδων (centroids) ως μέσος όρος των σημείων που τους ανήκουν και η διαδικασία επαναλαμβάνεται μέχρι να ικανοποιηθεί το κριτήριο σύγκλισης, δηλαδή έως όταν δεν υπάρχουν σημεία που να ανταλλάσσονται μεταξύ των ομάδων.

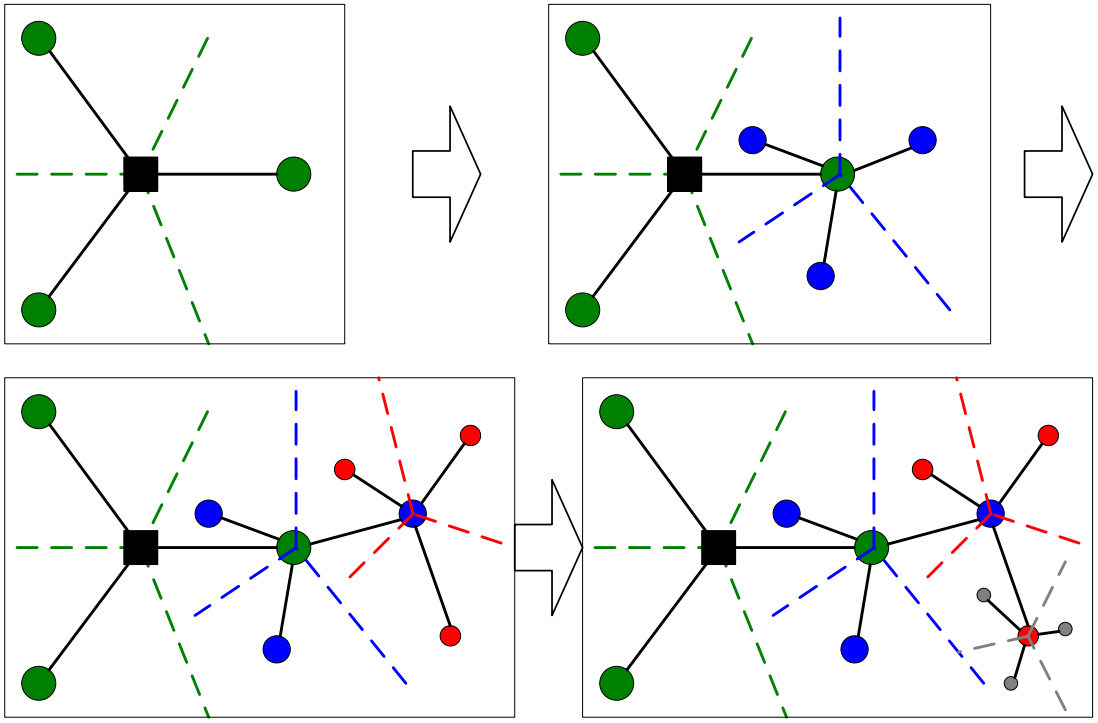


**Σχ. 4.3:** Συσταδοποίηση k-means διδιάστατων σημείων σε k=3 συστάδες ([VLFeat])

Η πιο γνωστή υλοποίηση της συσταδοποίησης K-means είναι ο αλγόριθμος του Lloyd, κατά τον οποίο χρησιμοποιείται ένας ευριστικός τρόπος για την αρχικοποίηση των ομάδων και στη συνέχεια σε κάθε βήμα υπολογίζονται τα λεγόμενα διαγράμματα Voronoi. Ένα παράδειγμα συσταδοποίησης διδιάστατων σημείων σε k = 3 ομάδες φαίνεται στο σχήμα 4.3. Τα διαφορετικά χρώματα προσδιορίζουν σημεία που ανήκουν σε τρεις τελικές ομάδες στις οποίες συγκλίνει ο αλγόριθμος και οι διακεκομμένες γραμμές αντιπροσωπεύουν τα όρια διαχωρισμού των ομάδων αυτών. Οι περιοχές που καθορίζονται από τα όρια μεταξύ των διαφορετικά χρωματισμένων σημείων αποτελούν τα κελιά Voronoi, ενώ οι μεγαλύτεροι κύκλοι είναι τα κέντρα των ομάδων.

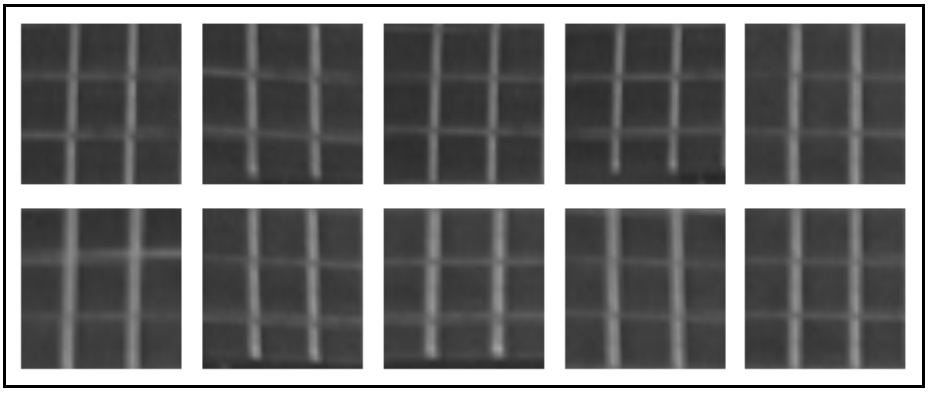
Στην περίπτωση των πολυδιάστατων διανυσμάτων χαρακτηριστικών (D = 64, 128) και λόγω του μεγάλου πλήθους των σημείων ανά εικόνα της συλλογής, ο αλγόριθμος k-means απαιτεί πολύ χρόνο μέχρι να συγκλίνει στην τελική λύση. Γι' αυτό το λόγο έχει προταθεί και χρησιμοποιείται και εδώ η τεχνική του ιεραρχικού αλγορίθμου k-means, που βασίζεται στην διάσπαση του προβλήματος ομαδοποίησης σε πολλά μικρότερα με τη χρήση ενός ιεραρχικού δέντρου [Philbin et al., 2007]. Όπως έχει αρχικά προταθεί, στο πρώτο επίπεδο του δέντρου τα σημεία ομαδοποιούνται σε πολύ μικρό αριθμό συστάδων (για παράδειγμα k = 3) με πολύ μικρό κόστος και στο επόμενο επίπεδο το σύνολο των σημείων κάθε ομάδας ομαδοποιείται ξεχωριστά και πάλι σε k ομάδες. Η διαδικασία συνεχίζεται μέχρι και το τελευταίο επίπεδο n, οπότε ο τελικός αριθμός ομάδων είναι $L = k^n$ όσα δηλαδή και τα φύλλα του δέντρου. Ο αλγόριθμος έχει ως αποτέλεσμα μία προσεγγιστική λύση και όχι τη βέλτιστη κατανομή των σημείων σε ομάδες, όμως από πειράματα αποδεικνύεται η καταλληλότητά του για τη δημιουργία ενός μεγάλου οπτικού λεξικού, το οποίο προσφέρει πολύ καλύτερη περιγραφή των οπτικών εννοιών από ένα

μικρότερο λεξικό (όπως αναγκαστικά κατασκευάζεται με τον απλό αλγόριθμο k-means). Στο σχήμα 4.4 φαίνεται ένα παράδειγμα ιεραρχικής συσταδοποίησης και διαδοχικής κβαντοποίησης των σημείων στο χώρο χαρακτηριστικών, με k = 3 και διάσταση του χώρου ίση με 2 για λόγους ευκρίνειας του σχήματος. Εδώ τα διαφορετικά χρώματα αντιπροσωπεύουν τα 4 διαφορετικά επίπεδα του ιεραρχικού δέντρου (πράσινο, μπλε, κόκκινο, γκρι αντίστοιχα). Από κάθε επίπεδο συνεχίζουμε τη συσταδοποίηση για μία μόνο από τις τρεις ομάδες και φτάνουμε μέχρι το 4ο επίπεδο, και πάλι για λόγους οπτικής σαφήνειας. Οι διακεκομμένες γραμμές είναι τα όρια και οι χρωματισμένοι κύκλοι τα κέντρα των ομάδων σε κάθε επίπεδο, ενώ οι συνεχείς μαύρες γραμμές είναι οι κλάδοι του δέντρου (το μαύρο τετράγωνο είναι η ρίζα του δέντρου).



**Σχ. 4.4:** Ιεραρχική συσταδοποίηση k-means για τα 4 πρώτα επίπεδα σε k=3 ομάδες, με σημεία χαρακτηριστικών στο διδιάστατο επίπεδο ([Nister et al., 2006])

Στο σχήμα 4.5 φαίνονται περιοχές MSER μετά την κανονικοποίησή τους, οι οποίες μέσω των διανυσμάτων χαρακτηριστικών τους (SURF) θα αντιστοιχηθούν στην ίδια τελική ομάδα, δηλαδή σε ένα σημείο του χώρου χαρακτηριστικών (μέσος όρος των σημείων της ομάδας τους) το οποίο θα αποτελεί ένα συγκεκριμένο οπτικό όρο του λεξικού. Πρέπει να σημειώσουμε εδώ ότι η οπτική λέξη δεν είναι ακριβώς το ίδιο με έναν γλωσσικό όρο, διότι ο όρος σε ένα κείμενο αντιστοιχεί σε μία συγκεκριμένη έννοια. Σε μια εικόνα τα αντικείμενα που περιέχουν μερικές ίδιες λέξεις δεν είναι πάντοτε όμοια, δηλαδή η οπτική λέξη δεν αντιστοιχεί απαραίτητα σε έννοια, αλλά αποτελεί το δομικό στοιχείο για κάθε ευρύτερη "οπτική έννοια".



**Σχ. 4.5:** Παράδειγμα κανονικοποιημένων περιοχών MSER που αντιστοιχούν στην ίδια οπτική λέξη

Τέλος πρέπει να σημειώσουμε ότι το βασικό μειονέκτημα του αλγορίθμου k-means είναι ο προκαθορισμένος αριθμός των k ομάδων. Απαιτείται δηλαδή εκ των προτέρων γνώση του πλήθους των οπτικών όρων που θα χρησιμοποιηθούν στο λεξικό. Πολλά πειράματα έχουν γίνει στο θέμα αυτό και συνήθως επιλέγεται εμπειρικά η τιμή που δίνει τα καλύτερα αποτελέσματα ανάκτησης εικόνων, η οποία πολλές φορές εξαρτάται από τον μέγεθος της βάσης δεδομένων αλλά και από το είδος του περιεχομένου των εικόνων. Θα πρέπει το μέγεθος του λεξικού να είναι τέτοιο ώστε μετά την κβαντοποίηση των διανυσμάτων χαρακτηριστικών να μπορούν να ανακτηθούν σωστά οι ίδιες εικόνες, να διατηρείται δηλαδή η χρήσιμη και να απορρίπτεται η περιττή πληροφορία των περιγραφέων [Philbin et al., 2007].

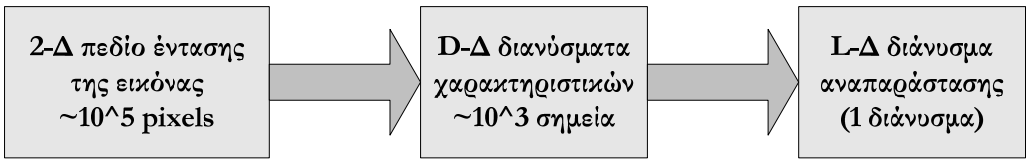
### 4.2.3 Δεικτοδότηση των εικόνων

Στο στάδιο της δεικτοδότησης, όπως ήδη είπαμε, μέσω του οπτικού λεξικού που κατασκευάστηκε, κάθε διάνυσμα χαρακτηριστικών σε κάθε εικόνα αντιστοιχίζεται σε μία από τις οπτικές λέξεις. Αυτό γίνεται με τον υπολογισμό του κοντινότερο γείτονα, διαδικασία που επιταχύνεται με τη χρήση των k-d trees [Friedman et al., 1977], δυαδικά δέντρα τα οποία μειώνουν κατά πολύ τον χρόνο αναζήτησης αφού η πολυπλοκότητα των συγκρίσεων πέφτει στο logn. Αντί να αποθηκεύονται όλα τα διανύσματα περιγραφέων για όλα τα τοπικά χαρακτηριστικά που έχουν ανιχνευθεί στην εικόνα, διατηρείται πλέον μόνο η σημαντική πληροφορία που είναι το ποιες και πόσες οπτικές λέξεις περιέχονται σε αυτήν. Η συχνότητα εμφάνισης f κάθε λέξης j μέσα σε κάθε εικόνα i της συλλογής αποθηκεύεται στο λεγόμενο διάνυσμα αναπαράστασης v (model vector), όπου L το μέγεθος του οπτικού λεξικού:

$$\boldsymbol{v_i} = [\, f_{i1} \quad f_{i2} \quad \ldots \quad f_{ij} \quad \ldots \quad f_{iL} \,]^T \qquad \textbf{(εξ. 4.3)}$$

Με τον παραπάνω τρόπο κάθε εικόνα μπορεί να δεικτοδοτηθεί με το διάνυσμα αναπαράστασής της και μόνο. Το αποτέλεσμα είναι ότι μειώσαμε την πολυπλοκότητα του προβλήματος περιγραφής του περιεχομένου και ταυτόχρονα δημιουργήσαμε μια πιο περιεκτική πληροφορία για την αναζήτηση με βάση το περιεχόμενο. Από την ακατέργαστη πληροφορία της έντασης της διδιάστατης εικόνας (όλα τα pixels) υπολογίσαμε το διάνυσμα χαρακτηριστικών στον χώρο διάστασης D για όλα σημεία ενδιαφέροντος (της τάξης του $10^3$ ανά εικόνα) και μετά ομαδοποιήσαμε τα δεδομένα αυτά φτιάχνοντας το οπτικό λεξικό με διάσταση L. Κάθε εικόνα τώρα περιγράφεται από ένα διάνυσμα αναπαράστασης στον χώρο των οπτικών λέξεων διάστασης

L. Ο μετασχηματισμός της πληροφορίας της εικόνας διαδοχικά από τον χώρο της έντασης στον χώρο των χαρακτηριστικών και τέλος στον χώρο του λεξικού παρουσιάζεται στο σχήμα 4.6.




**Σχ. 4.6:** Διαδοχικός μετασχηματισμός της πληροφορίας των εικόνων

### 4.2.4 Αναζήτηση με βάση την ομοιότητα

Στο τελευταίο στάδιο του συστήματος αναζήτησης εικόνων θα πρέπει να εξηγήσουμε τον τρόπο με τον οποίο οι όμοιες εικόνες θα επιστρέφονται στον χρήστη. Για να ξεκινήσει η αναζήτηση θα πρέπει πρώτα ο χρήστης να δώσει ως ερώτημα ένα παράδειγμα εικόνας, είτε από το υπάρχον σύνολο της βάσης είτε νέα εικόνα. Στη δεύτερη περίπτωση τα τοπικά χαρακτηριστικά εντοπίζονται στη νέα εικόνα και από τις τιμές των περιγραφέων της κατασκευάζεται το διάνυσμα αναπαράστασης. Έτσι, με δεδομένη τη δεικτοδοτημένη συλλογή, απομένει να καθορίσουμε ένα μέτρο ομοιότητας μεταξύ των εικόνων με βάση τα διανύσματα αναπαράστασης.

Ως μέτρο ομοιότητας (similarity) θα μπορούσαμε να χρησιμοποιήσουμε την ευκλείδεια απόσταση των σημείων στον L-D χώρο του λεξικού. Επειδή όμως δεν γνωρίζουμε κάποιο άνω φράγμα για τα μέτρα των διανυσμάτων v (σε μια εικόνα θεωρητικά μπορεί να περιέχονται πάρα πολλές οπτικές λέξεις ενός είδους), το μέτρο αυτό δεν είναι κατάλληλο για να μας δώσει τη σχετική ομοιότητα που θέλουμε. Καταρχάς θα πρέπει πρώτα να κανονικοποιήσουμε τα διανύσματα αναπαράστασης και αυτό μπορεί να γίνει διαιρώντας το καθένα με το μέτρο του υπολογισμένο είτε με την L1 είτε με την L2 νόρμα. Το μέτρο ενός διανύσματος $v_i$ δίνεται από τους επόμενους τύπους (L1 και L2 αντίστοιχα):

$$|\boldsymbol{v_i}|_1 = \sum_{j=1}^{L} |f_{ij}| \qquad \textbf{(εξ. 4.4)}$$

$$|\boldsymbol{v_i}|_2 = \left[ \sum_{j=1}^{L} f_{ij}^{\,2} \right]^{1/2} \qquad \textbf{(εξ. 4.5)}$$

Στη συνέχεια η ομοιότητα μεταξύ των κανονικοποιημένων διανυσμάτων των δύο εικόνων μπορεί να υπολογιστεί από την απόσταση Manhattan ή την ευκλείδεια απόσταση, αλλά και πάλι αυτές οι τιμές ομοιότητας δεν είναι σχετικές αλλά απόλυτες (εξαρτώνται δηλαδή από τον αριθμό των σημείων ενδιαφέροντος που θα ανιχνευθούν) και δεν προσφέρονται ως γενικό μέτρο σύγκρισης.

Η κβαντοποίηση του χώρου των χαρακτηριστικών οδηγεί στην κατασκευή ενός ιστογράμματος, του οποίου κάθε στοιχείο αντιστοιχεί σε κάθε οπτική λέξη. Το διάνυσμα αναπαράστασης για κάθε εικόνα είναι το ιστόγραμμα των οπτικών όρων της, αφού κάθε στοιχείο του είναι ουσιαστικά η συχνότητα εμφάνισης της αντίστοιχης λέξης. Η απόσταση μεταξύ ιστογραμμάτων εκφράζεται από την τομή τους, δηλαδή ένα διάνυσμα με στοιχεία τα ελάχιστα των αντίστοιχων στοιχείων των δύο ιστογραμμάτων. Ένα μέτρο ομοιότητας λοιπόν που θα μπορούσε να χρησιμοποιηθεί είναι η νόρμα L1 της τομής των δύο ιστογραμμάτων h (histogram

intersection) των διανυσμάτων αναπαράστασης. Κανονικοποιώντας πρώτα τα διανύσματα με τη νόρμα L1, το μέτρο ομοιότητας $s_1$ (similarity) θα είναι:

$$s_1 = \left| h\left( \frac{\boldsymbol{v_1}}{|\boldsymbol{v_1}|_1}, \frac{\boldsymbol{v_2}}{|\boldsymbol{v_2}|_1} \right) \right|_1 = \sum_{j=1}^{L} \min\left( \frac{\boldsymbol{v_{1j}}}{|\boldsymbol{v_1}|_1}, \frac{\boldsymbol{v_{2j}}}{|\boldsymbol{v_2}|_1} \right) \quad \textbf{(εξ. 4.6)}$$

Ένας άλλος τρόπος είναι να χρησιμοποιηθεί ένα μέτρο από τη διανυσματική ανάλυση όπως το κανονικοποιημένο εσωτερικό γινόμενο, το οποίο ουσιαστικά ισούται με το συνημίτονο της γωνίας των δύο διανυσμάτων. Επειδή μάλιστα τα στοιχεία των διανυσμάτων είναι συχνότητες εμφάνισης και άρα πάντοτε θετικά (στο πρώτο τεταρτημόριο, αν σκεφτούμε την υποπερίπτωση του διδιάστατου επιπέδου), η γωνία που σχηματίζουν μεταξύ τους θα ανήκει στο διάστημα [0, π/2] και άρα το συνημίτονο θα είναι στο διάστημα [0, 1]. Χρησιμοποιώντας την κανονικοποίηση με τη νόρμα L2, το μέτρο ομοιότητας $s_2$ (similarity) είναι:

$$s_2 = \cos(\boldsymbol{v_1}, \boldsymbol{v_2}) = \frac{\boldsymbol{v_1} \cdot \boldsymbol{v_2}}{|\boldsymbol{v_1}|_2 \cdot |\boldsymbol{v_2}|_2} \quad \textbf{(εξ. 4.7)}$$

Τα δύο μέτρα που ορίστηκαν στις σχέσεις 4.6 και 4.7 ($s_1$ και $s_2$ αντίστοιχα) είναι κανονικοποιημένα και άρα είναι κατάλληλα για να χρησιμοποιηθούν ως μέτρα της ομοιότητας μεταξύ δύο εικόνων, με τον συγκεκριμένο συνδυασμό κανονικοποίησης και μέτρου που αναφέρθηκε: L1-νόρμα και τομή ιστογράμματος ή L2-νόρμα και εσωτερικό γινόμενο. Στο τέλος, αφού υπολογιστούν οι τιμές ομοιότητας μεταξύ του ερωτήματος και κάθε εικόνας από τη συλλογή δεδομένων, κατασκευάζεται ένα διάνυσμα σχετικότητας των εικόνων (με διάσταση το μέγεθος όλης της συλλογής) και αφού ταξινομηθεί με σειρά φθίνουσας ομοιότητας, επιστρέφεται ως απάντηση στον χρήστη. Στις πρώτες θέσεις του διανύσματος αυτού περιέχονται οι πιο σχετικές με το ερώτημα εικόνες, με βάση το περιεχόμενο τους. Σε όλα τα πειράματα της επόμενης παραγράφου χρησιμοποιείται το μέτρο ομοιότητας $s_2$ (cosine) το οποίο είναι κανονικοποιημένο στο διάστημα [0, 1] και έδωσε καλύτερα αποτελέσματα σε σχέση με το $s_1$.

## 4.3 Πειράματα

### 4.3.1 Περιγραφή των πειραματικών δεδομένων

Στο σύστημα αναζήτησης εικόνων που αναλύθηκε προηγουμένως επιχειρείται μία πειραματική διαδικασία κατά την οποία θα αξιολογηθούν οι επιδόσεις των τριών διαφορετικών μεθόδων τοπικών χαρακτηριστικών. Τα πειράματα θα γίνουν σε τρία μεγάλα σύνολα δεδομένων από περισσότερες από 1000 εικόνες το καθένα, τα οποία χρησιμοποιούνται συχνά στη βιβλιογραφία ως μέτρο σύγκρισης και σημείο αναφοράς για διάφορες μετρήσεις.

Η πρώτη συλλογή που θα χρησιμοποιηθεί είναι η συλλογή Zurich Buildings Dataset ή συντομογραφικά ZuBuD, η οποία περιέχει 1005 έγχρωμες εικόνες που περιλαμβάνουν 201 διαφορετικά τυχαία επιλεγμένα κτίρια της πόλης της Ζυρίχης, το καθένα από τα οποία υπάρχει σε 5 διαφορετικές όψεις [ZuBud]. Κάθε εικόνα έχει διαστάσεις 640×480 pixels, είναι αποθηκευμένη στη μορφή PNG και έχει ληφθεί υπό διαφορετικές συνθήκες γωνίας, φωτισμού,

καιρού, ακόμα και από δύο διαφορετικές φωτογραφικές μηχανές. Σε μερικές εικόνες σκόπιμα υπάρχουν αντικείμενα όπως δέντρα που καλύπτουν μερικώς τα κτίρια και έτσι μπορεί να επιβεβαιωθεί η ανεξαρτησία των τοπικών χαρακτηριστικών από την μερική επικάλυψη των κτιρίων. Μερικές μόνο εικόνες από το σύνολο ZuBuD φαίνονται στο σχήμα 4.7, όπου για κάθε κτίριο περιλαμβάνονται και οι 5 διαφορετικές όψεις του, ώστε να γίνουν αντιληπτές οι συνθήκες λήψεις.

**Σχ. 4.7:** Μερικές ενδεικτικές εικόνες από τη συλλογή ZuBuD

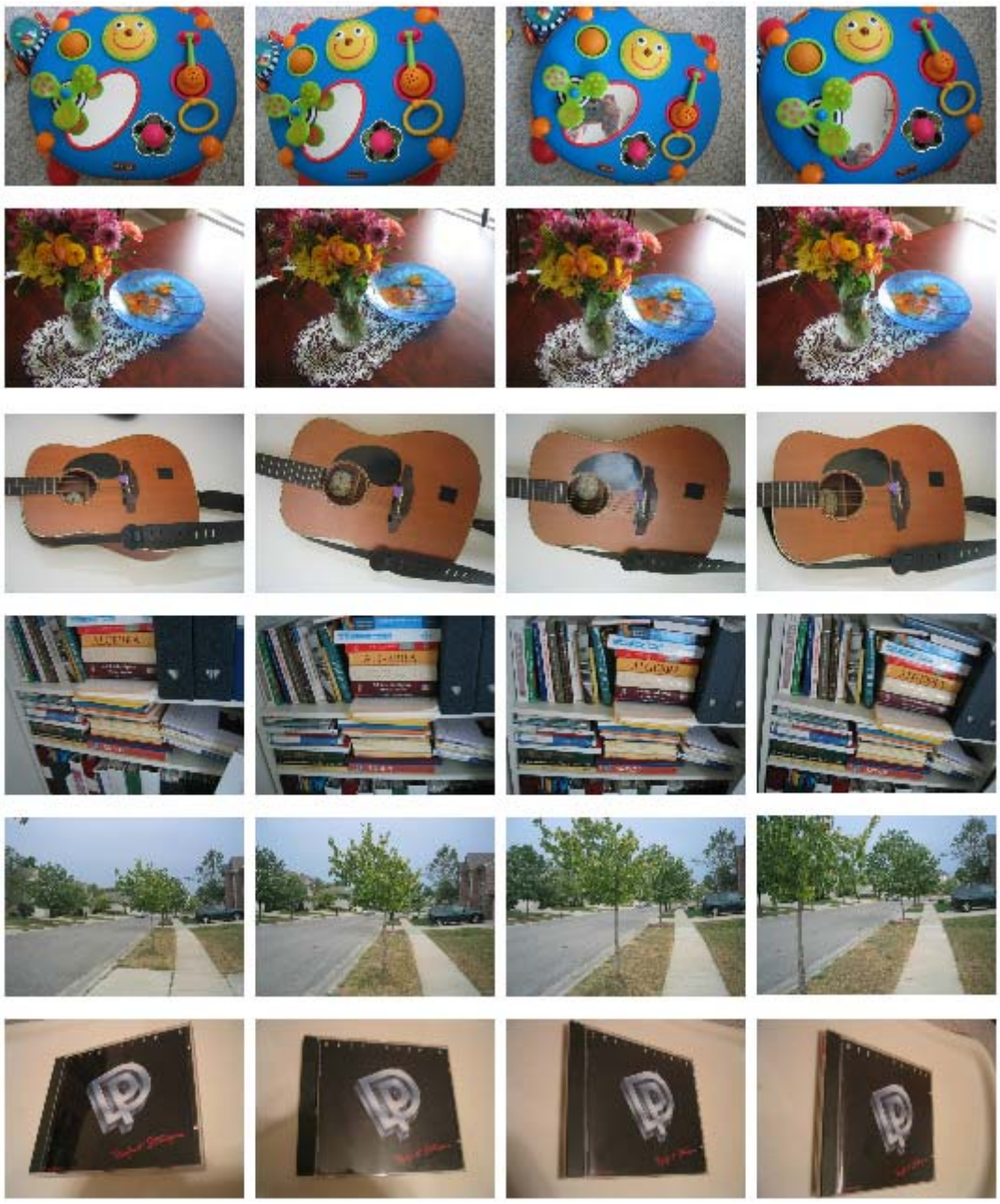
Η επόμενη συλλογή που χρησιμοποιείται είναι η Oxford Buildings Dataset ή απλά Oxford Buildings, η οποία αποτελείται από 5063 εικόνες με διάφορα κτίρια του Λονδίνου [OxBuD]. Οι εικόνες έχουν ενδιάμεσες διαστάσεις περίπου 1024×768 pixels και είναι συμπιεσμένες και αποθηκευμένες με το πρότυπο JPEG. Ολόκληρο το σύνολο δεδομένων έχει προέλθει από τον ιστότοπο [FlickR], όπου με κατάλληλες αναζητήσεις βρέθηκαν εικόνες που αφορούν 11 συγκεκριμένα παλιά κτίρια του Λονδίνου. Στη συνέχεια οι εικόνες έχουν μελετηθεί και σχολιαστεί κατάλληλα ώστε να προκύψει το σύνολο αληθείας (ground truth), χωρισμένο σε άλλα επιμέρους σύνολα: ένα άριστο σύνολο με υψηλά ποσοστά εμφάνισης των κτιρίων, ένα καλό σύνολο με αρκετές εμφανίσεις των κτιρίων αλλά μαζί με άλλα αντικείμενα, και ένα κακό σύνολο το οποίο έχει πολύ μικρά μέρη των αντικειμένων που ενδιαφέρουν και πολλά αντικείμενα που μπερδεύουν. Για τις ερωτήσεις στο στάδιο των πειραμάτων υπάρχουν ξεχωριστές εικόνες, 5 για κάθε ένα από τα 11 μνημεία του συνόλου, αλλά δεν θα χρησιμοποιηθούν εδώ. Επίσης, πρέπει να σημειώσουμε ότι πολλές από τις υπόλοιπες εικόνες που δεν ανήκουν στα παραπάνω σύνολα

αληθείας περιέχουν "θορυβώδη" αντικείμενα, τα οποία αναμένεται να μειώσουν το συνολικό ποσοστό επιτυχίας. Εκτός δηλαδή από εικόνες με κτίρια, υπάρχουν και φωτογραφίες εσωτερικών χώρων, αρχαιολογικών ευρημάτων και μνημείων, πινάκων, ανθρώπων κτλ. Παραδείγματα εικόνων από τη συλλογή Oxford Buildings φαίνονται στο σχήμα 4.8.

**Σχ. 4.8:** Μερικές ενδεικτικές εικόνες από τη συλλογή Oxford Buildings

Τέλος, η πιο μεγάλη συλλογή που χρησιμοποιείται είναι η UKBench, η οποία αποτελείται από 10200 εικόνες, στις οποίες υπάρχουν 2550 αντικείμενα σε 4 διαφορετικές όψεις το καθένα [UKBench]. Κάθε εικόνα έχει διαστάσεις 640×480 pixels και είναι συμπιεσμένη και αποθηκευμένη με το πρότυπο JPEG. Οι εικόνες περιέχουν διάφορα καθημερινά αντικείμενα, είτε εσωτερικού χώρου, όπως βιβλία, CDs, ρούχα κτλ., αλλά και τοπία από εξωτερικούς χώρους, όπως δρόμους, αυτοκίνητα, δέντρα, κτλ. Για τα πειράματα μπορεί να χρησιμοποιηθεί ολόκληρο το

σύνολο δεδομένων ή μικρότερα υποσύνολά του. Πρέπει να λάβουμε υπόψη ότι η επιτυχία των αλγορίθμων εξαρτάται από το είδος του υποσυνόλου που θα πάρουμε, αφού μερικά αντικείμενα είναι από μόνα τους πιο δύσκολο να αναζητηθούν σε σχέση με άλλα (όπως τα λουλούδια σε σχέση με τα CDs), ενώ εξαρτάται και από την ποιότητα ορισμένων φωτογραφιών, αφού μερικές είναι θολές, θορυβώδεις ή δεν έχουν καλό φωτισμό. Μερικές εικόνες του συνόλου UKBench dataset φαίνονται στο σχήμα 4.9.



**Σχ. 4.9:** Μερικές ενδεικτικές εικόνες από τη συλλογή UKBench

Τέλος, πρέπει να αναφέρουμε το σύνολο για κάθε συλλογή το οποίο θεωρούμε σαν δεδομένη αλήθεια (ground truth). Για τα σύνολα ZuBuD και UKBench, στα οποία υπάρχουν 5 και 4 αντίστοιχα οπτικές γωνίες του ίδιου αντικειμένου, ως σύνολο αληθείας για κάθε κτίριο ή αντικείμενο θεωρούμε τις διαφορετικές του όψεις. Για τη συλλογή Oxford Buildings ως σύνολο

αληθείας για τις 11 διαφορετικές τοποθεσίες μνημείων δεχόμαστε τις εικόνες που ορίζονται στο σύνολο του άριστου ground truth το οποίο δημοσιεύεται στην ιστοσελίδα [OxBuD]. Στη συνέχεια, για να γίνουν τα πειράματα και να εξαχθούν συγκριτικά αποτελέσματα, πρέπει να επιλεγούν τα ερωτήματα προς το σύστημα. Ως ερώτημα επιλέγεται κάθε διαφορετική όψη για κάθε αντικείμενο που υπάρχει στο ground truth. Ως σωστή απάντηση θεωρείται κάθε άλλη όψη του ίδιου αντικειμένου, και αυτό γίνεται και για τα τρία σύνολα πειραματικών δεδομένων. Αυτό σημαίνει ότι για τα ZuBuD και UKBench ως ερωτήματα χρησιμοποιούνται όλες οι εικόνες του συνόλου, αλλά για το Oxford Buildings χρησιμοποιούνται μόνο οι εικόνες που ανήκουν στο άριστο σύνολο των 11 μνημείων (δηλαδή ένα πολύ μικρό υποσύνολο του dataset, περίπου 300 εικόνες) και αυτό κάνει ακόμα πιο δύσκολη την αναζήτηση στο σύνολο αυτό.

### 4.3.2 Κριτήρια αξιολόγησης

Εδώ πρέπει να ορίσουμε τα μέτρα με τα οποία θα αξιολογήσουμε το σύστημα αναζήτησης και κατά συνέπεια θα συγκρίνουμε και τις επιδόσεις των διαφορετικών μεθόδων τοπικών χαρακτηριστικών. Τα δύο πιο διαδεδομένα μέτρα που χρησιμοποιούνται σε προβλήματα αναζήτησης πληροφορίας είναι το μέτρο ακριβείας (precision) και το μέτρο ανάκτησης (recall). Έστω ότι γνωρίζουμε το σύνολο δεδομένης αλήθειας, ground truth, δηλαδή το σύνολο των εικόνων που ιδανικά θα έπρεπε να επιστρέψει το σύστημα για ένα συγκεκριμένο ερώτημα. Συμβολίζουμε με GT το πλήθος αυτών των εικόνων. Από τις συνολικές εικόνες που επιστρέφει το σύστημα ως απάντηση (positives), κάποιες είναι σωστές και ανήκουν στο σύνολο αληθείας (true positives) και κάποιες άλλες δεν θα έπρεπε να επιστραφούν ως σωστές (false positives), γιατί δεν ανήκουν στο σύνολο αληθείας και πολύ απλά δεν είναι όμοιες με το ερώτημα-εικόνα. Μετρώντας τις σωστές και λάθος απαντήσεις, των οποίων το πλήθος συμβολίζουμε με TP και FP αντίστοιχα, οι συνολικές απαντήσεις P είναι: P = TP + FP.

Το μέτρο ακριβείας ορίζεται ως ο λόγος των σωστών εικόνων που επεστράφησαν προς τις συνολικές εικόνες που επεστράφησαν και είναι:

$$precision = \frac{TP}{TP + FP} \qquad \textbf{(εξ. 4.8)}$$

Το μέτρο ανάκτησης ορίζεται ως ο λόγος των σωστών εικόνων που επεστράφησαν προς τις συνολικές αναμενόμενες εικόνες που θα έπρεπε να επιστραφούν και είναι:

$$recall = \frac{TP}{GT} \qquad \textbf{(εξ. 4.9)}$$

Στο σημείο αυτό όμως θα πρέπει να κάνουμε μια επισήμανση. Έστω N το πλήθος των εικόνων του συνόλου δεδομένων. Το σύστημα αναζήτησης σε κάθε ερώτηση του χρήστη επιστρέφει ως απάντηση ολόκληρη τη συλλογή, ταξινομημένη με φθίνουσα σειρά σχετικότητας. Αυτό σημαίνει ότι η επιστρεφόμενες απαντήσεις δεν θα είναι ούτε μία εικόνα ούτε ένα συγκεκριμένο πλήθος, για παράδειγμα ίσο με 4 ή 5, αλλά και οι N εικόνες. Επειδή λοιπόν μέσα στην απάντηση περιέχονται όλες οι εικόνες του ground truth, δεν μπορεί κανείς να βγάλει σωστά συμπεράσματα, αφού η ανάκτηση θα είναι πάντοτε ίση με ένα και η ακρίβεια σχεδόν ίση με μηδέν.

Γι' αυτόν το λόγο χρησιμοποιείται η έννοια της εμβέλειας (scope), δηλαδή ενός παραθύρου πάνω στο σύνολο της απάντησης (έστω διάνυσμα μήκους m), από το οποίο προκύπτει

ως τρέχουσα απάντηση το υποσύνολο από τις πρώτες m εικόνες. Από το υποσύνολο αυτό μπορούμε να εξάγουμε το τρέχον μέτρο ακριβείας και ανάκτησης (scope precision και scope recall) και στη συνέχεια, μεταβάλλοντας το εύρος του παραθύρου για όλες τις δυνατές τιμές από 1 μέχρι N, μπορούμε να υπολογίσουμε ένα μέσο όρο των μέτρων ακριβείας (average precision), λαμβάνοντας υπόψη όμως μόνο τα παράθυρα εκείνα στα οποία στην τελευταία τους θέση εμφανίζεται μία νέα εικόνα από το ground truth.

Το τρέχον μέτρο ακριβείας $p_m$ (scope precision) υπολογισμένο στο παράθυρο scope με εύρος m όπως καταλαβαίνουμε θα δίνεται από τη σχέση:

$$p_m = \frac{TP(m)}{TP(m) + FP(m)} = \frac{TP(m)}{m} \qquad \textbf{(εξ. 4.10)}$$

Με τον ίδιο τρόπο ορίζεται και το τρέχον μέτρο ανάκτησης (scope recall) $r_m$ μέσω της σχέσης:

$$r_m = \frac{TP(m)}{GT} \qquad \textbf{(εξ. 4.11)}$$

Τότε για το συγκεκριμένο ερώτημα ο μέσος όρος ακριβείας που θα συμβολίζεται με AP (average precision), προκύπτει από την τρέχουσα ακρίβεια που υπολογίζεται για όλες τις δυνατές τιμές του scope αλλά μόνο για εκείνες τις περιπτώσεις όπου συναντάται μία νέα σωστή εικόνα στο τέλος του παραθύρου (δηλαδή μόνο όταν αλλάζει το μέτρο ανάκτησης $r_m$), θα δίνεται από τον εξής τύπο:

$$AP = \frac{\sum_{m=1}^{N} t_m \cdot p_m}{GT} = \frac{1}{GT} \sum_{m=1}^{N} t_m \cdot \frac{TP(m)}{m} \qquad \textbf{(εξ. 4.12)}$$

όπου ο συντελεστής $t_m$ είναι ίσος με 1 μόνο αν η εικόνα στη θέση m ανήκει στο ground truth (θεωρείται σωστή ανάκτηση), αλλιώς είναι ίσος με 0.

Με τον τρόπο αυτό προκύπτει ένα μέτρο average precision το οποίο λαμβάνει υπόψη τις σωστές αναμενόμενες εικόνες αλλά και τη θέση τους μέσα στο διάνυσμα της απάντησης. Οι σωστές εικόνες που βρίσκονται στις πρώτες θέσεις του διανύσματος της απάντησης συνεισφέρουν πολύ περισσότερο στο μέτρο ακριβείας, ενώ οι ανόμοιες εικόνες στις μεσαίες και τελευταίες θέσεις δεν επηρεάζουν αρνητικά τις τιμές του. Για να προκύψει το συνολικό μέτρο αξιολόγησης για ολόκληρο το dataset, υπολογίζονται τα διάφορα μέτρα AP για όλα τα διαθέσιμα ερωτήματα και στη συνέχεια υπολογίζεται η μέση τιμή τους. Η ποσότητα αυτή ονομάζεται μέση ακρίβεια (mean average precision) και θα συμβολίζεται ως mAP. Έστω ότι έχουμε συνολικά M ερωτήματα στο πείραμα, τότε η μέση ακρίβεια θα δίνεται από τον τύπο:

$$mAP_1 = \frac{1}{M} \sum_{i=1}^{M} AP(i) \qquad \textbf{(εξ. 4.13)}$$

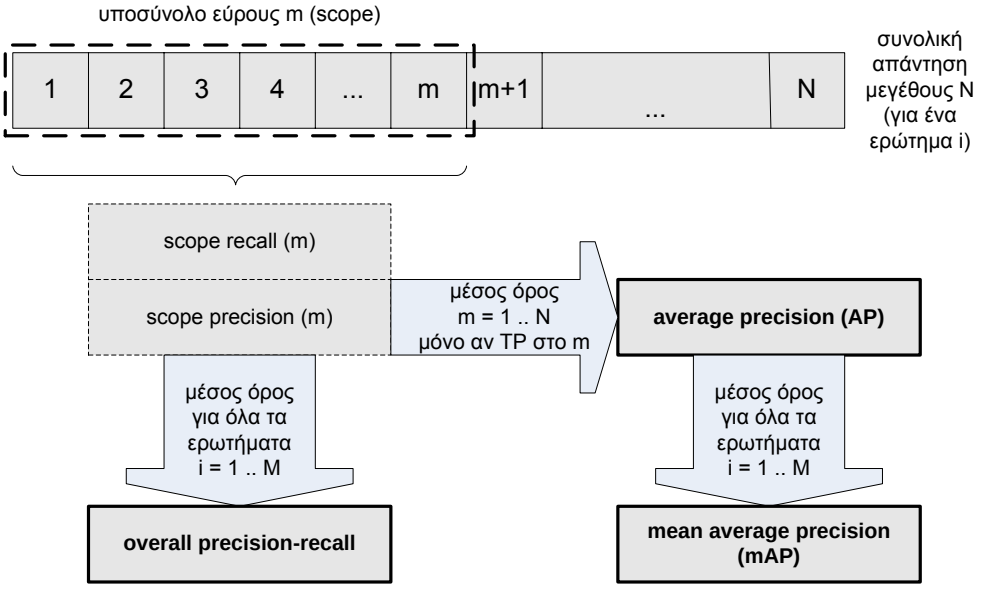
Συμβολίσαμε το παραπάνω μέτρο ως $mAP_1$ γιατί εδώ θα ορίσουμε και ένα δεύτερο μέτρο $mAP_2$, το οποίο εκφράζει την ίδια κατανομή των σωστών εικόνων μέσα στο διάνυσμα της απάντησης. Τα δύο μέτρα $p_m$ και $r_m$ (βλέπε εξισώσεις 4.10 και 4.11) ως συναρτήσεις της παραμέτρου m αποτελούν τις συνιστώσες διδιάστατης καμπύλης και μπορούν να αναπαρασταθούν γραφικά και να δώσουν ένα διάγραμμα ακριβείας-ανάκτησης για το τρέχον ερώτημα. Ως μέτρο μέσης ακριβείας ορίζεται το εμβαδόν κάτω από την καμπύλη αυτή, όπως έχει

χρησιμοποιηθεί σε πολλές πρόσφατες δημοσιεύσεις [Chum et al., 2007], [Philbin et al., 2008]. Οπότε τελικά η μέση ακρίβεια για όλα τα υπάρχοντα ερωτήματα είναι η μέση τιμή όλων αυτών των εμβαδών, σύμφωνα με τον τύπο:

$$mAP_2 = \frac{1}{M}\sum_{i=1}^{M} a(i) \qquad \textbf{(εξ. 4.14)}$$

όπου α το εμβαδόν της καμπύλης, που δίνεται από τη σχέση:

$$a(i) = \sum_{m=1}^{N-1} \frac{p_{m+1}(i) + p_m(i)}{2}[r_{m+1}(i) - r_m(i)] \qquad \textbf{(εξ. 4.15)}$$




**Σχ. 4.10:** Μεταβολή του εύρους m του scope, τρέχον μέτρο ακριβείας και ανάκτησης (scope m), και υπολογισμός των μέτρων AP, mAP, $op_m$ και $or_m$ (βλέπε εξισώσεις 4.11-4.13, 4.16-4.17)

Τα μέτρα μέσης ακριβείας $mAP_1$ και $mAP_2$ είναι μια εκτίμηση της συνολικής απόδοσης του συστήματος στην ανάκτηση των σωστών εικόνων. Για να πάρουμε μια ιδέα της κατανομής των σωστών εικόνων μέσα στο διάνυσμα της απάντησης και μια πιο λεπτομερή περιγραφή των επιδόσεων των μεθόδων τοπικών χαρακτηριστικών που μας ενδιαφέρουν, πρέπει να κατασκευάσουμε μια καμπύλη ακριβείας-ανάκτησης. Από τις τιμές των μέτρων $p_m$ και $r_m$ θα μπορούσαμε να σχεδιάσουμε μια καμπύλη ξεχωριστή για κάθε ερώτημα, αλλά αυτό δεν είναι πρακτικό. Έτσι, υπολογίζοντας και πάλι έναν μέσο όρο για αυτά τα μέτρα πάνω σε όλα τα ερωτήματα (από 1 μέχρι M), προκύπτει ένα συνολικό μέτρο ακριβείας $op_m$ (overall precision) και ένα συνολικό μέτρο ανάκτησης $or_m$ (overall recall), τα οποία και αυτά είναι συναρτήσεις του scope m:

$$op_m = \frac{1}{M}\sum_{i=1}^{M} p_m(i) \qquad \textbf{(εξ. 4.16)}$$

$$or_m = \frac{1}{M}\sum_{i=1}^{M} r_m(i) \qquad \textbf{(εξ. 4.17)}$$

Στο σχήμα 4.10 φαίνεται το διάνυσμα της απάντησης (μεγέθους N), το πώς ακριβώς λαμβάνεται το scope ως υποσύνολο αυτού και πώς υπολογίζονται τα τρέχοντα μέτρα ακριβείας και ανάκτησης ($p_m$ και $r_m$), το μέτρο ακριβείας AP (μέσος όρος) και μέσης ακριβείας mAP και τέλος τα συνολικά μέτρα ακριβείας και ανάκτησης ($op_m$ και $or_m$).

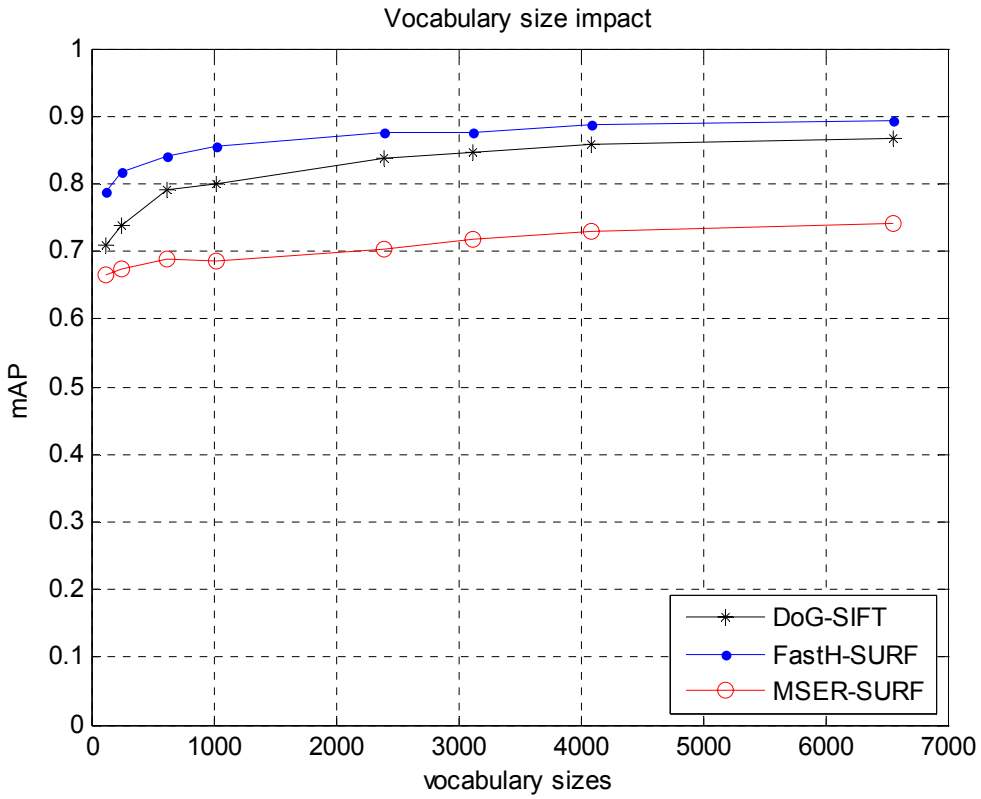
Τέλος, οι τιμές των συνολικών αυτών μέτρων αναπαρίστανται σε γραφήματα που αφορούν όλα τα διαθέσιμα ερωτήματα από ολόκληρο το σύνολο δεδομένης αλήθειας και χρησιμοποιούνται μαζί με τις μέσες ακρίβειες (mAP1 και mAP2) για την εξαγωγή συμπερασμάτων, όπως θα δούμε στη συνέχεια.

### 4.3.3 Πειράματα και παρατηρήσεις

Όπως είπαμε, θα διεξάγουμε πειράματα για να συγκρίνουμε την απόδοση των τριών μεθόδων τοπικών χαρακτηριστικών DoG-SIFT, FastH-SURF και MSER-SURF σε τρεις διαφορετικές συλλογές εικόνων: ZuBuD, Oxford Buildings και UKBench. Καταρχάς με τη μέθοδο της ιεραρχικής συσταδοποίησης k-means κατασκευάζουμε διαφόρων μεγεθών λεξικά, και συγκεκριμένα για τις παρακάτω τιμές k και n, προκύπτουν και τα αντίστοιχα μεγέθη L:

k = 5 n = 3 : L = 125
k = 4 n = 4 : L = 256
k = 5 n = 4 : L = 625
k = 4 n = 5 : L = 1024
k = 7 n = 4 : L = 2401
k = 5 n = 5 : L = 3125
k = 8 n = 4 : L = 4096
k = 9 n = 4 : L = 6561

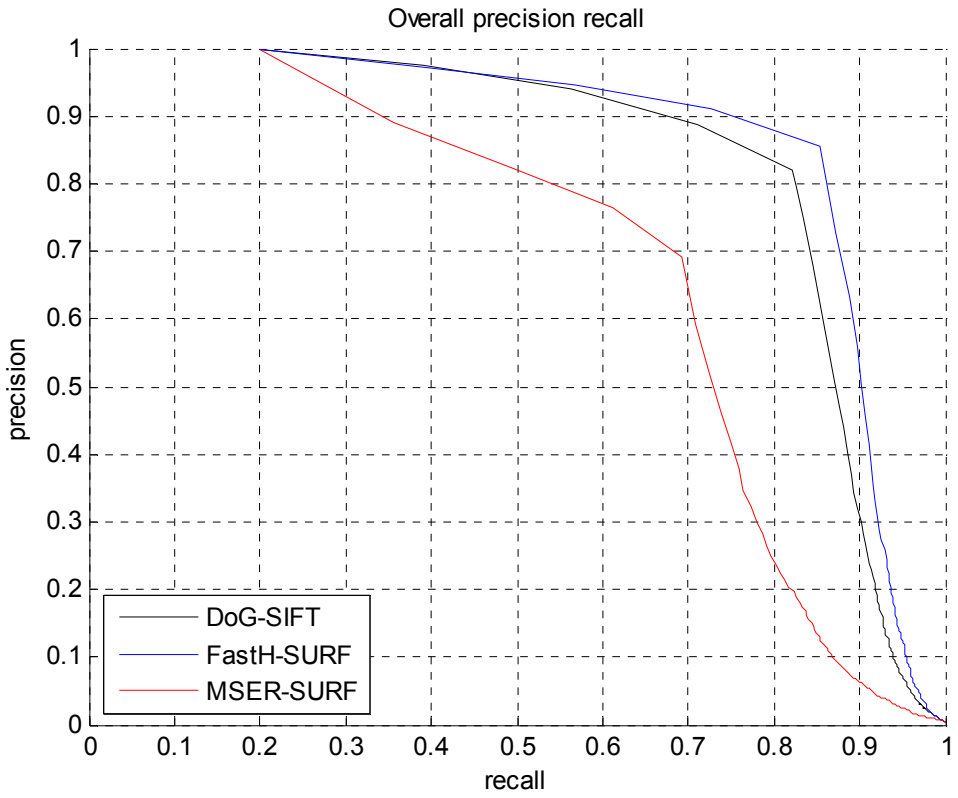
Από εδώ και στο εξής για την αξιολόγηση του συστήματος αναζήτησης και των μεθόδων θα χρησιμοποιούμε το $1^ο$ μέτρο μέσης ακριβείας $mAP_1$. Σε σχετικά πειράματα τα δύο μέτρα που ορίζονται στις σχέσεις 4.13 και 4.14 έδωσαν παρόμοια αποτελέσματα με το πρώτο να έχει ελαφρώς μεγαλύτερες τιμές. Επίσης όλα τα διαγράμματα που αφορούν μέτρα ακρίβειας-ανάκτησης αναφέρονται στα συνολικά μέτρα που ορίζονται στις σχέσεις 4.16 και 4.17.



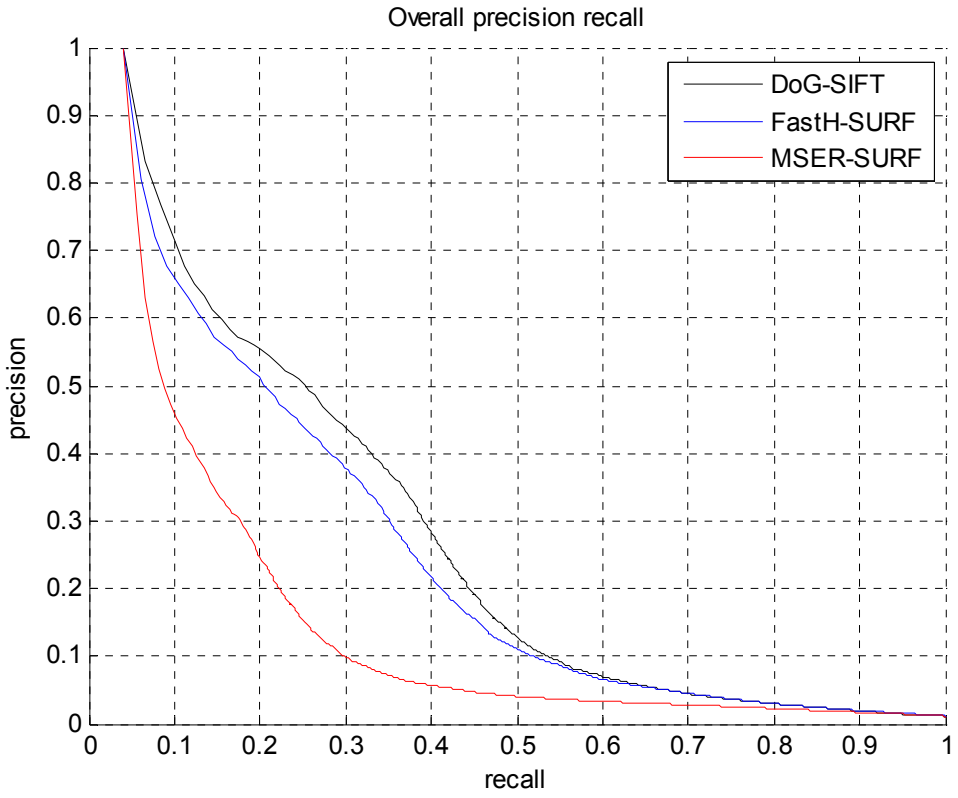

**Σχ. 4.11:** Επίδραση της διάστασης του οπτικού λεξικού στο μέτρο μέσης ακριβείας mAP (ZuBuD)

Στη συνέχεια, για να αποφασίσουμε για το κατάλληλο μέγεθος λεξικού, υπολογίζουμε το μέτρο μέσης ακριβείας mAP για κάθε λεξικό που κατασκευάσαμε. Οι τιμές του mAP για το σύνολο εικόνων ZuBuD φαίνονται στο διάγραμμα του σχήματος 4.11. Για όλες τις μεθόδους η καμπύλη αυξάνεται απότομα στην αρχή και στη συνέχεια ακολουθεί σταθερή αύξηση, με την τάση προς το τέλος της να παραμείνει οριζόντια. Αυτό μας οδηγεί στο συμπέρασμα ότι η περαιτέρω αύξηση του πλήθους των οπτικών λέξεων δεν θα ακολουθηθεί από δραματική αύξηση της μέσης ακριβείας. Από εδώ και στο εξής λοιπόν σε όλα τα πειράματα χρησιμοποιούνται λεξικά με μέγεθος 6.5K οπτικές λέξεις. Στα σύνολα ZuBuD και UKBench οι εικόνες από τις οποίες έγινε η ομαδοποίηση επιλέχθηκαν με κριτήριο το περιεχόμενό τους, για παράδειγμα επιλέχτηκε μία από τις 4 ή 5 διαφορετικές όψεις του ίδιου αντικειμένου. Στο σύνολο Oxford Buildings οι εικόνες που συνέβαλλαν στη δημιουργία του λεξικού ήταν οι εικόνες του άριστου ground truth (good).

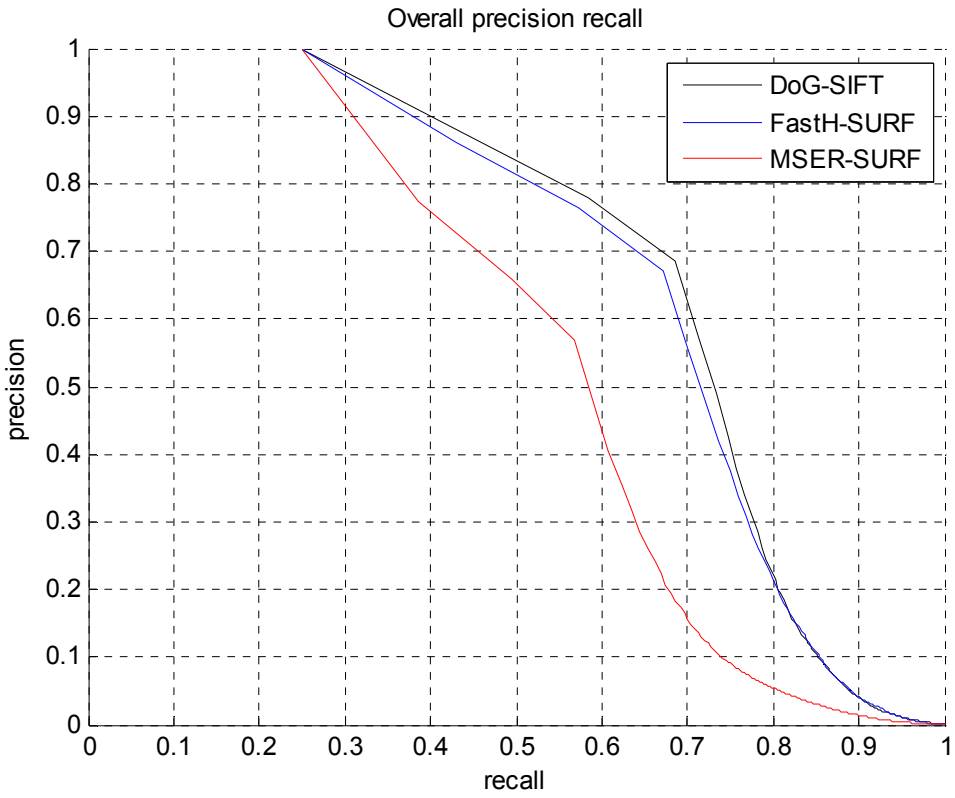
Στη συνέχεια σε κάθε συλλογή εικόνων, αφού πρώτα κατασκευάσουμε και αποθηκεύσουμε τα διανύσματα αναπαράστασης, εκτελούμε τη διαδικασία της ερώτησης στο σύστημα, για όλες τις εικόνες του συνόλου δεδομένης αλήθειας. Επαναλαμβάνουμε τη διαδικασία αυτή και για τις τρεις μεθόδους και υπολογίζουμε τα συνολικά μέτρα ακριβείας και ανάκτησης (overall precision-recall). Τα διαγράμματα ακριβείας-ανάκτησης παρουσιάζονται στα σχήματα 4.12 – 4.16, για τις συλλογές ZuBuD, Oxford Buildings και UKBench διαδοχικά. Οι καμπύλες τους προκύπτουν από τα σημεία που αντιστοιχούν σε διαφορετικές τιμές της εμβέλειας m (scope). Για την τελευταία συλλογή UKBench των 10200 εικόνων έγιναν τρία διαφορετικά πειράματα, χρησιμοποιώντας ως σύνολο δεδομένων το ένα τέταρτο της συλλογής με 2548 εικόνες, τη μισή συλλογή με 5100 εικόνες, και ολόκληρη τη συλλογή με 10200 εικόνες, με τα αντίστοιχα διαγράμματα στα σχήματα 4.14 – 4.16.



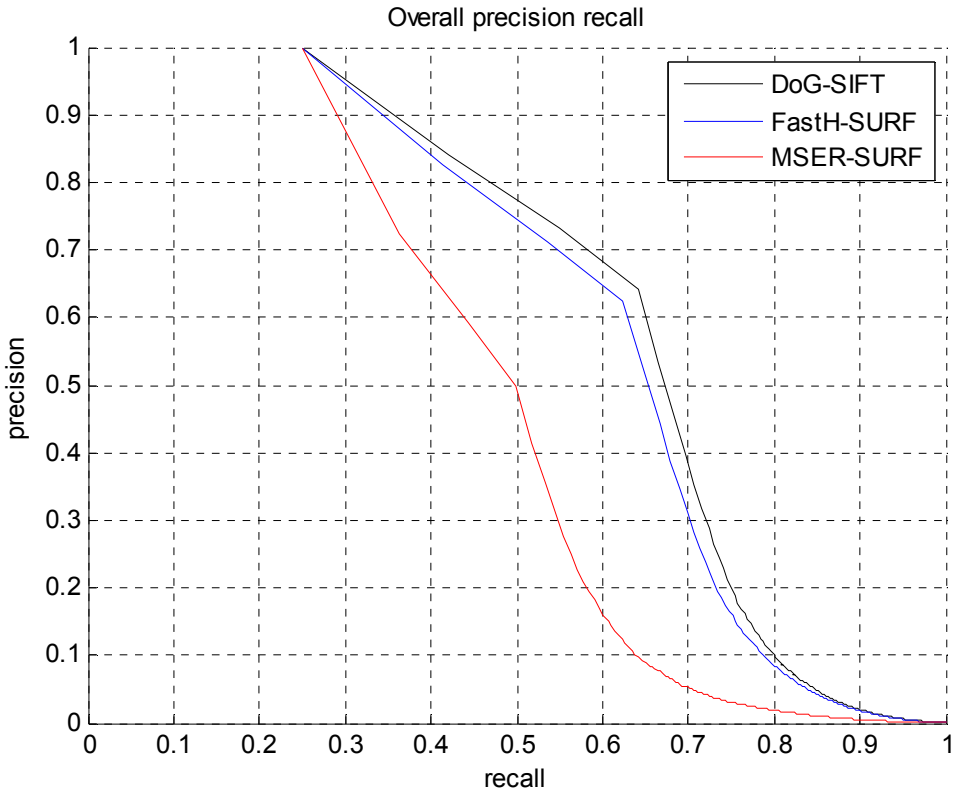

**Σχ. 4.12:** Καμπύλη precision-recall για τη συλλογή ZuBuD των 1005 εικόνων κτιρίων



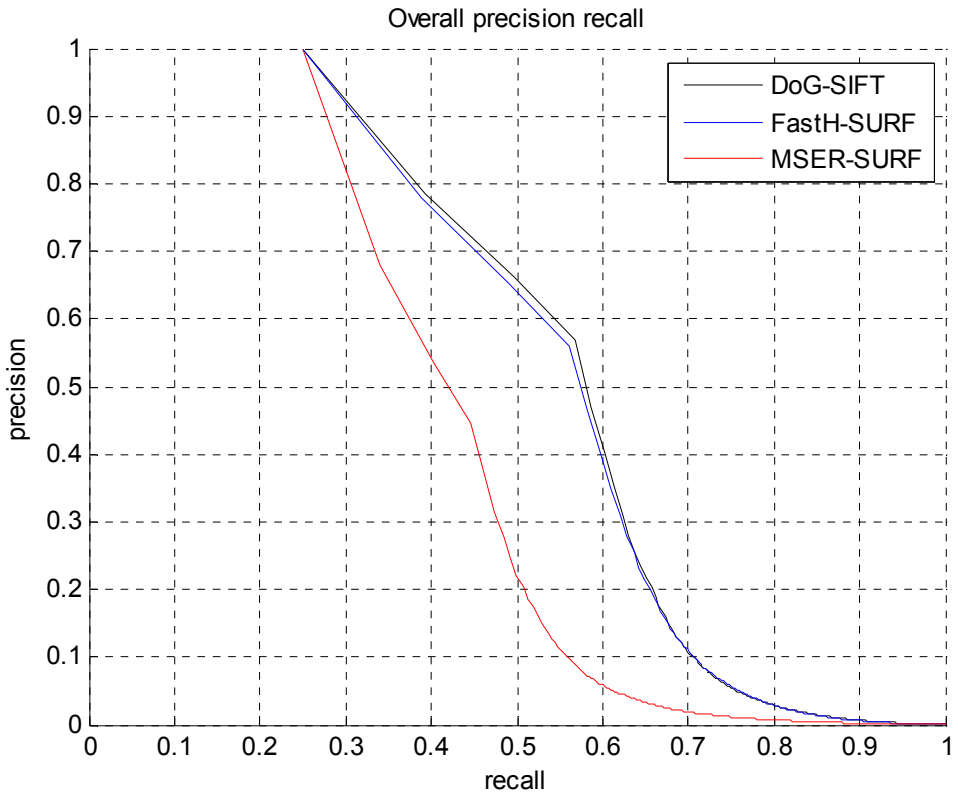

**Σχ. 4.13:** Καμπύλη precision-recall για τη συλλογή Oxford Buildings των 5063 εικόνων μνημείων




**Σχ. 4.14:** Καμπύλη precision-recall για ένα υποσύνολο της συλλογής UKBench με 2548 εικόνες




**Σχ. 4.15:** Καμπύλη precision-recall για ένα υποσύνολο της συλλογής UKBench με 5100 εικόνες




**Σχ. 4.16:** Καμπύλη precision-recall για τη συνολική συλλογή UKBench των 10200 εικόνων

Όπως φαίνεται από τα διαγράμματα οι μέθοδοι DoG-SIFT και FastH-SURF αποδίδουν πολύ καλύτερα απ' ότι ο συνδυασμός MSER-SURF. Συγκεκριμένα για τη συλλογή ZuBuD ο περιγραφέας SURF υπολογισμένος πάνω σε σημεία Fast Hessian έχει καλύτερη απόδοση από τον SIFT. Σε όλα τα υπόλοιπα πειράματα ο συνδυασμός ανιχνευτή Difference of Gaussians και διανύσματος χαρακτηριστικών SIFT έρχεται πρώτος μεταξύ των υπολοίπων. Ο περιγραφέας SURF υπολογισμένος στις κανονικοποιημένες περιοχές MSER έχει αρκετά χειρότερα αποτελέσματα από τους άλλους δύο συνδυασμούς. Παρόλ' αυτά διατηρεί αρκετά υψηλά ποσοστά ακριβείας-ανάκτησης για τη συλλογή ZuBuD, μπορεί δηλαδή να παράγει ικανοποιητικά αποτελέσματα. Επίσης, ο περιγραφέας SURF υπερτερεί σε σχέση με τον SIFT υπολογισμένο στις ίδιες περιοχές (έγιναν σχετικά πειράματα αλλά δεν παρουσιάζονται για λόγους απλότητας).

Κάτι που πρέπει να παρατηρήσουμε σε όλα τα προηγούμενα διαγράμματα είναι το εξής. Η ιδανική καμπύλη ακρίβειας-ανάκτησης θα είχε τιμή ανάκτησης ίση με 1 για όλες τις τιμές της ακριβείας, δηλαδή μια σχεδόν κατακόρυφη γραμμή που να ξεκινάει από το σημείο 1 του άξονα των τετμημένων. Αυτό φυσικά δεν είναι εφικτό στην πράξη, αλλά εκείνο που παρατηρούμε σε όλες τις καμπύλες (εκτός από τη συλλογή Oxford Buildings όπου τα ποσοστά είναι πολύ χαμηλά) είναι η παρουσία ενός έντονου σημείου καμπής. Αυτό το σημείο αντιστοιχεί στην τιμή m της εμβέλειας η οποία ισούται με το πλήθος των σωστών εικόνων του συνόλου αληθείας (για παράδειγμα στο σύνολο ZuBuD είναι η τιμή m = 5, ενώ στο σύνολο UKBench είναι η τιμή m = 4). Αν μειώσουμε και άλλο την εμβέλεια (αυξάνοντας την ακρίβεια), επειδή ακριβώς σε περίπτωση καλής επίδοσης όπως στο ZuBuD στις πρώτες θέσεις περιέχονται οι περισσότερες σωστές εικόνες, το ποσοστό ανάκτησης μειώνεται απότομα και έτσι δημιουργείται το σκέλος της καμπύλης πάνω από το σημείο καμπής. Τέλος, να σημειώσουμε ότι η μικρότερη τιμή για το μέτρο ανάκτησης είναι εκείνη που αντιστοιχεί σε εμβέλεια m = 1 όπου λαμβάνεται υπόψη μόνο η πρώτη εικόνα της απάντησης, που είναι ίδια με την εικόνα του ερωτήματος άρα πάντοτε σωστή.

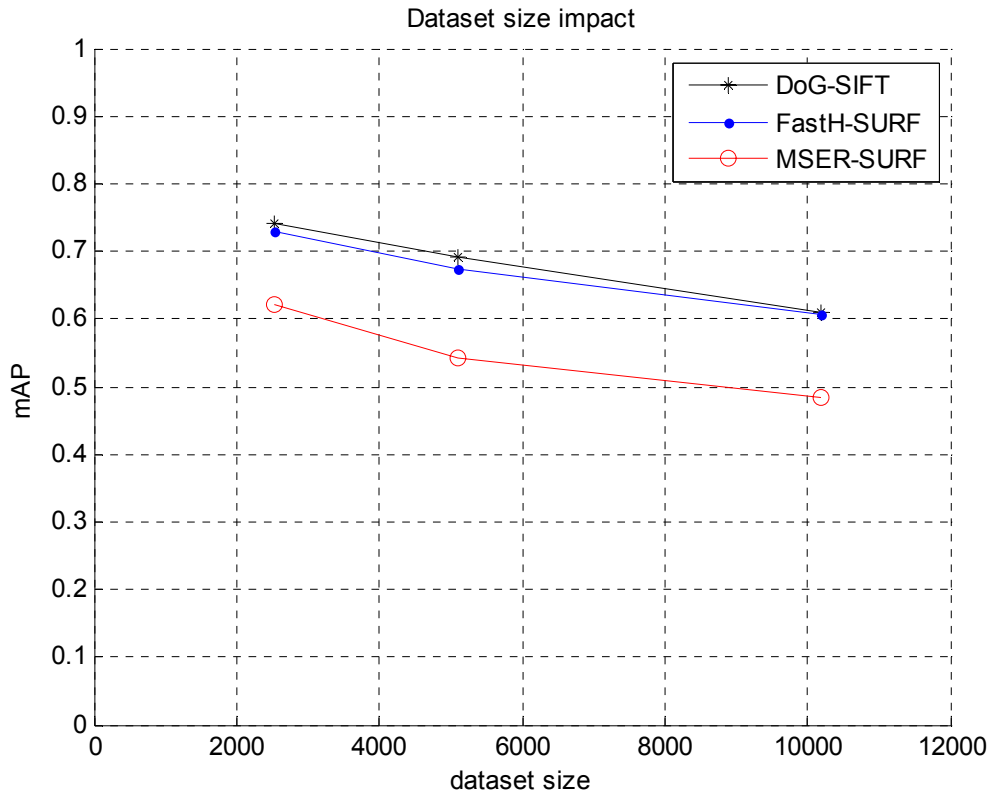
Για την περίπτωση του ZuBuD το μέτρο αυτό είναι ίσο με 1/5 = 0.20 ενώ για το UKBench είναι 1/4 = 0.25.

Στον πίνακα 4.1 περιέχονται οι τιμές του μέτρου για τη μέση ακρίβεια mAP όπως προκύπτει με εφαρμογή ξεχωριστά κάθε μεθόδου και για τις πέντε περιπτώσεις πειραματικών δεδομένων. Οι συμβολισμοί UKBench-2548, UKBench-5100 και UKBench-10200 αναφέρονται στα αντίστοιχα σύνολα-υποσύνολα της συλλογής UKBench. Οι τιμές του μέτρου mAP επιβεβαιώνουν τα διαγράμματα, ότι δηλαδή ο συνδυασμός DoG-SIFT είναι ελαφρώς καλύτερος από τον FastH-SURF, εκτός από το πρώτο πείραμα, και ο συνδυασμός MSER-SIFT δεν αποδίδει τόσο καλά όσο θα περιμέναμε, ειδικά στις εικόνες αυτών των συνόλων όπου παρατηρούνται μεταβολές στην οπτική γωνία, μετασχηματισμό τον οποίο οι περιοχές MSER μπορούν αν αντιμετωπίσουν. Επίσης μπορούμε να παρατηρήσουμε τα πολύ μικρά ποσοστά επιτυχίας για τη συλλογή Oxford Buildings. Αυτό οφείλεται στο περιεχόμενο των εικόνων αυτών. Περιέχουν πολλά ξένα αντικείμενα, ενώ τα ερωτήματα του ground truth είναι πολύ συγκεκριμένα, οπότε χωρίς ελέγχους χωρικής συνάφειας και απόρριψης εσφαλμένων ταιριασμάτων, η αναζήτηση ολόκληρων σκηνών δυσκολεύει αρκετά.

**Πίνακας 4.1:** Μέση ακρίβεια mAP για όλες τις συλλογές και τις μεθόδους τοπικών χαρακτηριστικών

| **Μέτρο μέσης ακριβείας (mAP)** | | | | | |
|---|---|---|---|---|---|
| **ΣΥΛΛΟΓΗ** | **ZuBuD** | **Oxford Buildings** | **UKBench - 2548** | **UKBench - 5100** | **UKBench - 10200** |
| **λεξικό DoG** | 0.866291 | 0.340777 | 0.739785 | 0.690924 | 0.610655 |
| **λεξικό FastH** | 0.89271 | 0.317686 | 0.729084 | 0.673415 | 0.605416 |
| **λεξικό MSER** | 0.742348 | 0.202121 | 0.622034 | 0.543712 | 0.482769 |

Στη συνέχεια παραθέτουμε και ένα διάγραμμα με τις τιμές της μέσης ακριβείας, για τις τρεις διαφορετικές περιπτώσεις πειραμάτων πάνω στο σύνολο UKBench, δηλαδή για ολόκληρο το σύνολο των 10200 εικόνων και για υποσύνολά του με 2548 και 5100 εικόνες. Από το σχήμα 4.17 μπορούμε να παρατηρήσουμε ότι και για τις τρεις διαθέσιμες μεθόδους, η κλίση των καμπύλων είναι παρόμοια και σχεδόν ευθεία. Αυτό σημαίνει ότι η ακρίβεια του συστήματος μεταβάλλεται γραμμικά σε σχέση με το μέγεθος της βάσης δεδομένων που θα χρησιμοποιηθεί. Επομένως από τις μεθόδους αυτές δεν υπάρχει κάποια περισσότερο ευαίσθητη από τις υπόλοιπες στην αύξησης του μεγέθους της συλλογής, αλλά όλες επηρεάζονται το ίδιο και μειώνονται με μικρό ρυθμό ως προς την αύξηση των εικόνων της συλλογής. Είναι λογικό να υπάρχει μία τέτοια μείωση, αφού με τις περισσότερες εικόνες προστίθεται νέα πληροφορία και οι σχετικές ομοιότητες αναδιατάσσονται στο διάνυσμα της απάντησης.




**Σχ. 4.17:** Επίδραση του μεγέθους του συνόλου δεδομένων στη μέση ακρίβεια mAP (UKBench)

Τέλος, παρατηρώντας την μέτρια απόδοση της μεθόδου MSER ακόμα και στη συλλογή ZuBuD όπου οι άλλες δύο έχουν μέση ακρίβεια κοντά στο 90%, επιχειρούμε να συνδυάσουμε τις μεθόδους και τα διαφορετικά λεξικά που προκύπτουν από αυτές και να αξιολογήσουμε το αποτέλεσμα. Δύο διαφορετικά λεξικά μπορούν να θεωρηθούν ότι είναι δύο διαφορετικοί τρόποι περιγραφής της ίδιας έννοιας. Ο συνδυασμός τους μπορούμε να πούμε ότι ισοδυναμεί με τη χρήση δύο εγκυκλοπαιδικών πηγών για την σύνταξη ενός κειμένου. Θα χρησιμοποιήσουμε δύο συνδυασμούς λεξικών: τα λεξικά από τις μεθόδους DoG και MSER και τα λεξικά από τις μεθόδους FastH και MSER.

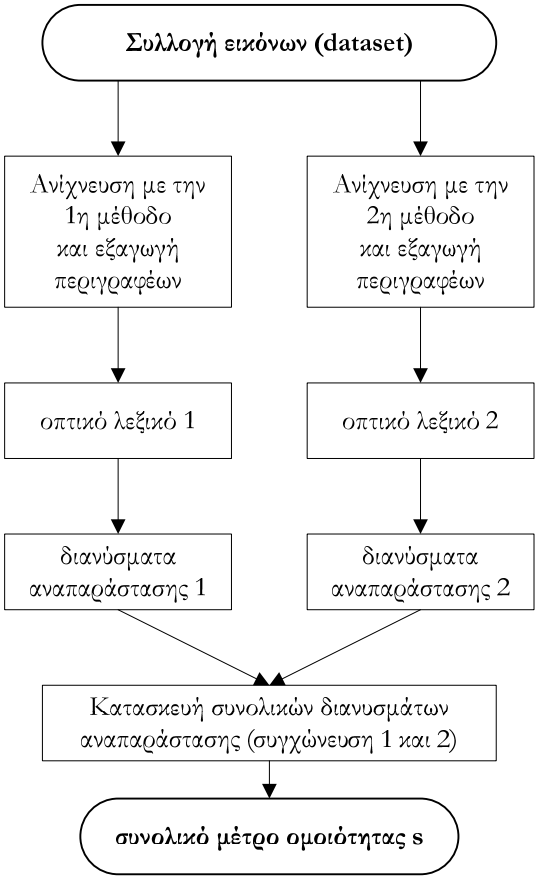
Η διαδικασία που ακολουθείται είναι η ίδια με την περίπτωση του ενός λεξικού: εξάγονται τα τοπικά χαρακτηριστικά και οι περιγραφείς τους με την κάθε μέθοδο, στη συνέχεια ομαδοποιούνται ξεχωριστά ώστε να κατασκευαστούν τα δύο οπτικά λεξικά και τέλος υπολογίζονται τα διανύσματα αναπαράστασης. Για να υπολογιστεί όμως η ομοιότητα μεταξύ του ερωτήματος και κάθε εικόνας από τη βάση θα πρέπει να δημιουργηθεί ένα τελικό διάνυσμα αναπαράστασης από τη συνένωση των δύο διανυσμάτων που προκύπτουν για τα δύο λεξικά. Αν $p_1$ και $p_2$ είναι τα δύο διανύσματα αναπαράστασης της εικόνας από τη βάση δεδομένων και $q_1$ και $q_2$ είναι τα δύο διανύσματα αναπαράστασης της εικόνας του ερωτήματος, τότε κατασκευάζονται τα συνολικά διανύσματα αναπαράστασης p και q ως εξής:

$$\boldsymbol{p} = [\boldsymbol{p}_1^T \quad \boldsymbol{p}_2^T]^T$$
$$\boldsymbol{q} = [\boldsymbol{q}_1^T \quad \boldsymbol{q}_2^T]^T \qquad \textbf{(εξ. 4.18)}$$

Οπότε το μέτρο ομοιότητας μεταξύ των δύο εικόνων θα υπολογιστεί μέσω του κανονικοποιημένου εσωτερικού γινομένου ως εξής:

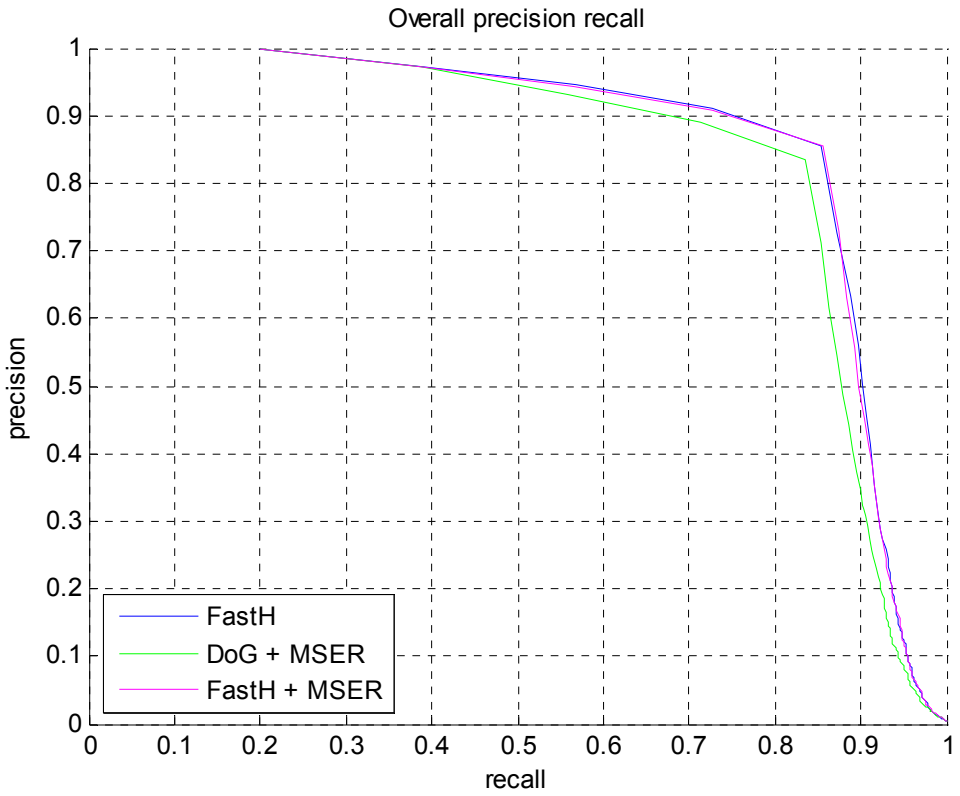
$$s = \cos(\boldsymbol{p},\boldsymbol{q}) = \frac{\boldsymbol{p}\cdot\boldsymbol{q}}{|\boldsymbol{p}|_2 \cdot |\boldsymbol{q}|_2} = \frac{\boldsymbol{p}_1\cdot\boldsymbol{q}_1 + \boldsymbol{p}_2\cdot\boldsymbol{q}_2}{|\boldsymbol{p}|_2 \cdot |\boldsymbol{q}|_2} \quad \textbf{(εξ. 4.19)}$$

Το τελικό μέτρο ομοιότητας s λαμβάνει υπόψη του ταυτόχρονα την ομοιότητα με βάση το πρώτο λεξικό και την ομοιότητα με βάση το δεύτερο λεξικό. Όταν η ομοιότητα είναι μεγάλη είτε ως προς το ένα είτε ως προς το άλλο λεξικό, τότε αυτό αρκεί για να συσχετισθούν οι δύο εικόνες αφού το μέτρο s θα παίρνει μεγάλες τιμές. Όταν για κανένα από τα δύο λεξικά δεν προκύπτουν κοντινά διανύσματα αναπαράστασης (δηλαδή ούτε τα $p_1$ και $q_1$ είναι κοντινά, ούτε τα $p_2$ και $q_2$), τότε μόνο η ομοιότητα παίρνει χαμηλές τιμές. Με αυτόν τον τρόπο αν δύο εικόνες της ίδιας σκηνής δεν δίνουν ομοιότητα κοντά στο 1 σύμφωνα με το πρώτο λεξικό, τότε υπάρχει μεγάλη πιθανότητα να είναι όμοιες σύμφωνα με το δεύτερο λεξικό και αντίστροφα, οπότε ενδέχεται να μειώσουμε τα σφάλματα και να βελτιώσουμε τα αποτελέσματα της αναζήτησης. Η διαδικασία της αναζήτησης με το συνδυασμό δύο λεξικών φαίνεται συνοπτικά στο σχήμα 4.18.

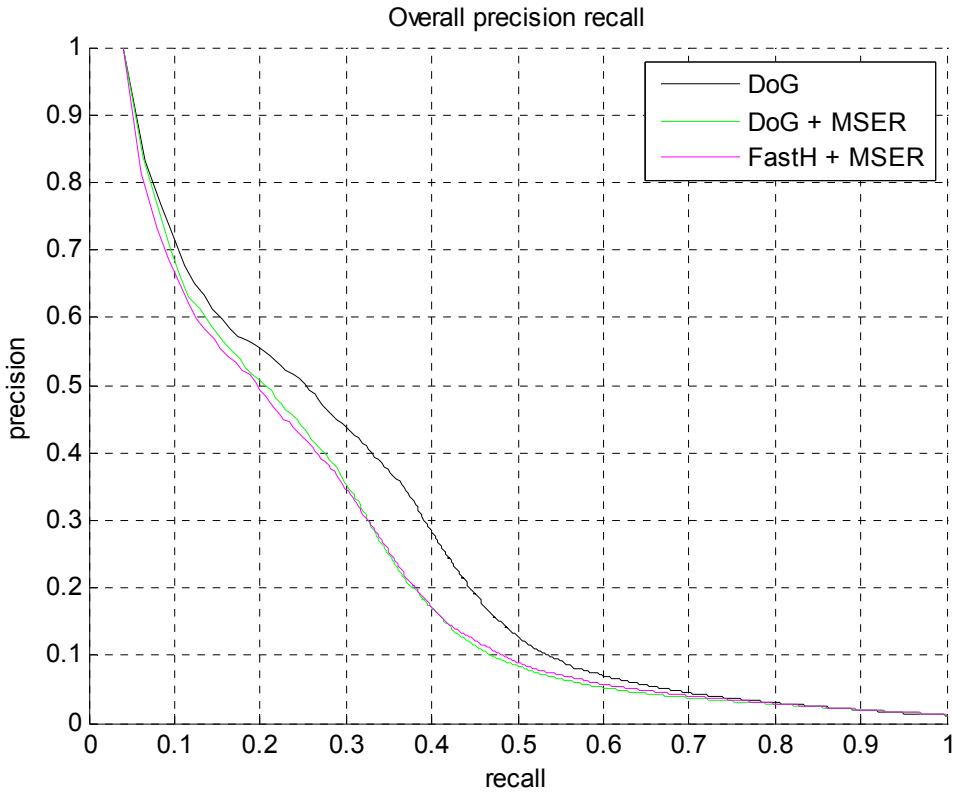



**Σχ. 4.18:** Κατασκευή και συνδυασμός δύο οπτικών λεξικών για την αναζήτηση εικόνων

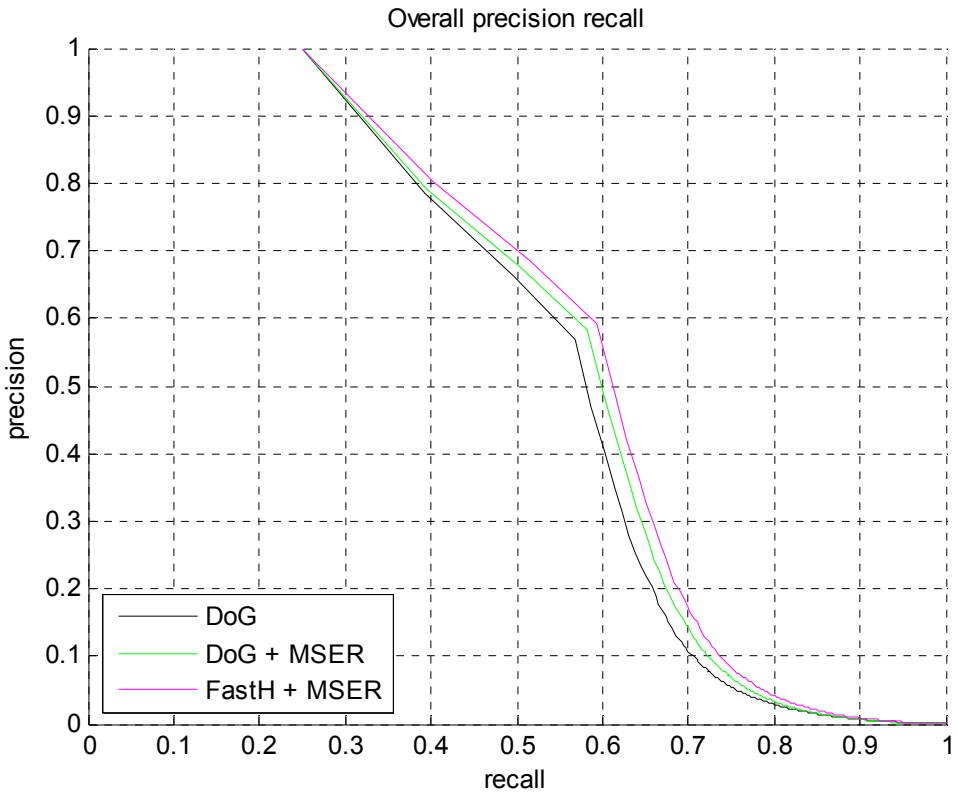
Με την παραπάνω διαδικασία διεξάγουμε τα πειράματα αναζήτησης εικόνων με τα ίδια σύνολα αληθείας στις συλλογές ZuBuD, Oxford Buildings και UKBench (σε ολόκληρη τη συλλογή UKBench). Τα αντίστοιχα διαγράμματα συνολικής ακριβείας-ανάκτησης περιλαμβάνονται στα σχήματα 4.19 – 4.21. Μαζί με τις καμπύλες για τα δύο λεξικά παρουσιάζεται και η καλύτερη καμπύλη που προέκυψε από τη χρήση ενός μόνο λεξικού, για την ευκολία των συγκρίσεων.




**Σχ. 4.19:** Καμπύλη precision-recall για τη συλλογή ZuBuD με συνδυασμό λεξικών




**Σχ. 4.20:** Καμπύλη precision-recall για τη συλλογή Oxford Buildings με συνδυασμό λεξικών




**Σχ. 4.21:** Καμπύλη precision-recall για τη συλλογή UKBench με συνδυασμό λεξικών

Όπως παρατηρούμε από τα διαγράμματα precision-recall στην περίπτωση του συνδυασμού των δύο διαφορετικών λεξικών, δεν παρατηρείται βελτίωση της καμπύλης για το σύνολο ZuBuD σε σχέση με τη χρήση του λεξικού από τη μέθοδο FastH. Στο σύνολο Oxford Buildings στο οποίο ήδη τα αποτελέσματα ήταν πολύ κάτω από μέτρια ο συνδυασμός όχι μόνο δεν βελτίωσε το αποτέλεσμα, αλλά επιδείνωσε τις μέτριες επιδόσεις σε σχέση με τη μέθοδο DoG και το λεξικό της. Στη συλλογή UKBench όμως, η βελτίωση της απόδοσης είναι αισθητή και για τους δύο συνδυασμούς λεξικών, παρόλ' αυτά χωρίς πολύ μεγάλες διαφορές με τις αρχικές. Η βελτίωση απόδοσης του συστήματος φαίνεται και από τον πίνακα 4.2, συγκρίνοντας τα μέτρα μέσης ακριβείας mAP με εκείνα του πίνακα 4.1 για τη συλλογή UKBench.

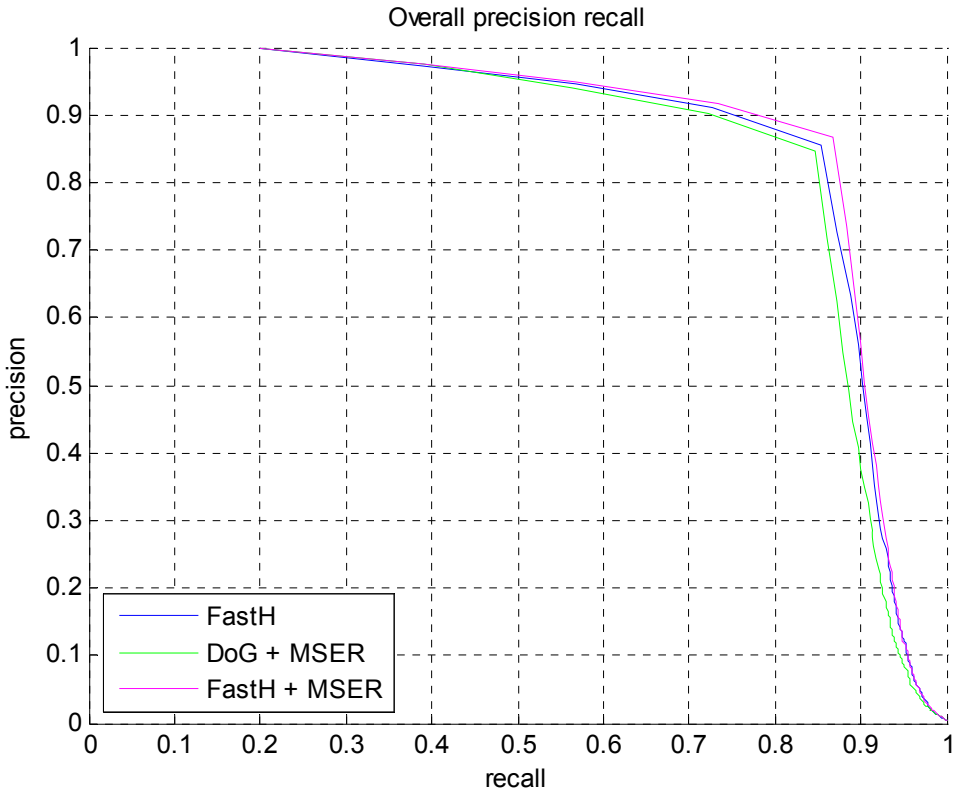
**Πίνακας 4.2:** Μέση ακρίβεια mAP για το συνδυασμό των μεθόδων και των οπτικών λεξικών

| **Μέτρο μέσης ακριβείας (mAP) για συνδυασμό λεξικών** | | | |
|---|---|---|---|
| **ΣΥΛΛΟΓΗ** | **ZuBuD** | **Oxford Buildings** | **UKBench** |
| **λεξικό DoG-MSER** | 0.8741 | 0.301204 | 0.625999 |
| **λεξικό FastH-MSER** | 0.892208 | 0.300637 | 0.637681 |

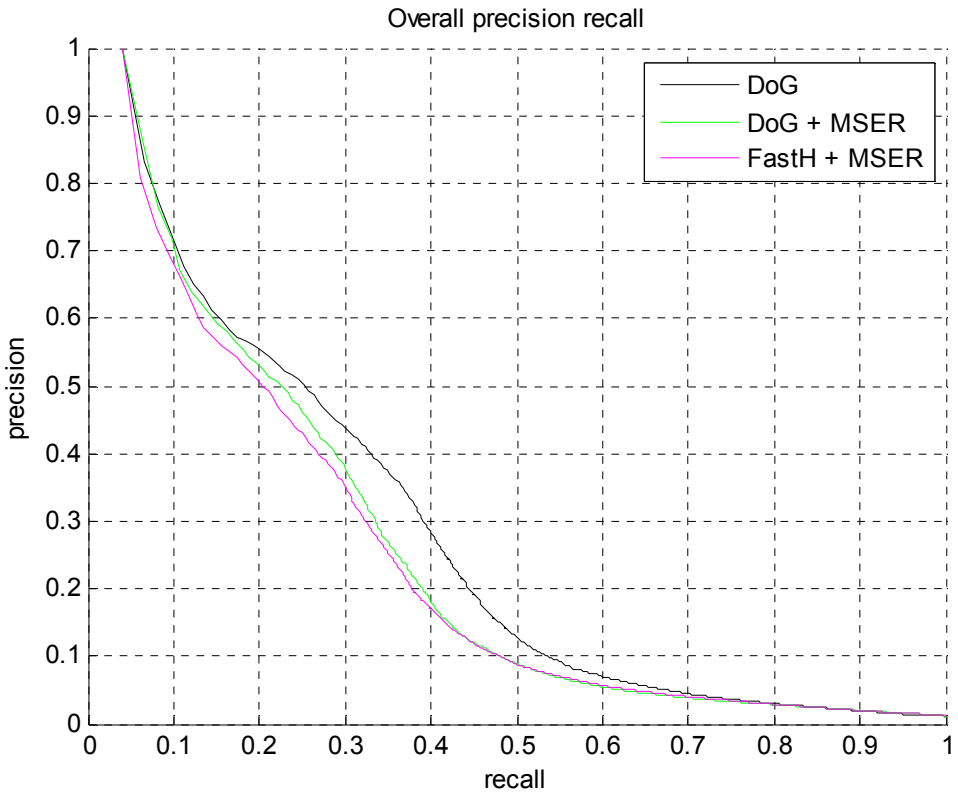
Η προηγούμενη προσέγγιση με την κατασκευή ενός συνολικού διανύσματος αναπαράστασης έχει ως αποτέλεσμα τον υπολογισμό του συνολικού μέτρου ομοιότητας s, το οποίο λαμβάνει μεγάλες τιμές όταν η ομοιότητα είναι μεγάλη είτε ως προς το ένα είτε ως προς το άλλο λεξικό. Αυτό ισοδυναμεί με την πράξη "OR" μεταξύ των δύο μέτρων ομοιότητας, δηλαδή στο νέο σύστημα προκύπτουν εικόνες με μεγαλύτερη ομοιότητα, και αυτό επηρεάζει την απόδοση του συστήματος αναδιατάσσοντας το διάνυσμα της απάντησης. Με βάση την ίδια λογική, μπορούμε να υλοποιήσουμε το συμμετρικό ανάλογο του παραπάνω, δηλαδή την πράξη "AND" μεταξύ των δύο οπτικών λεξικών και των ομοιοτήτων που προκύπτουν από αυτά, υπολογίζοντας την τελική ομοιότητα ως γινόμενο των επιμέρους μέτρων. Με την πράξη αυτή προκύπτει ένα πιο "αυστηρό" κριτήριο το οποίο απαιτεί και τα δύο μέτρα ομοιότητας να έχουν μεγάλες τιμές προκειμένου δύο εικόνες να θεωρηθούν όμοιες. Η διαδικασία συνδυασμού των δύο λεξικών είναι ίδια με αυτή του σχήματος 4.18, εκτός από το προτελευταίο στάδιο, ενώ το τελικό μέτρο ομοιότητας ορίζεται από τη σχέση:

$$s = s_1 * s_2 \qquad \textbf{(εξ. 4.20)}$$

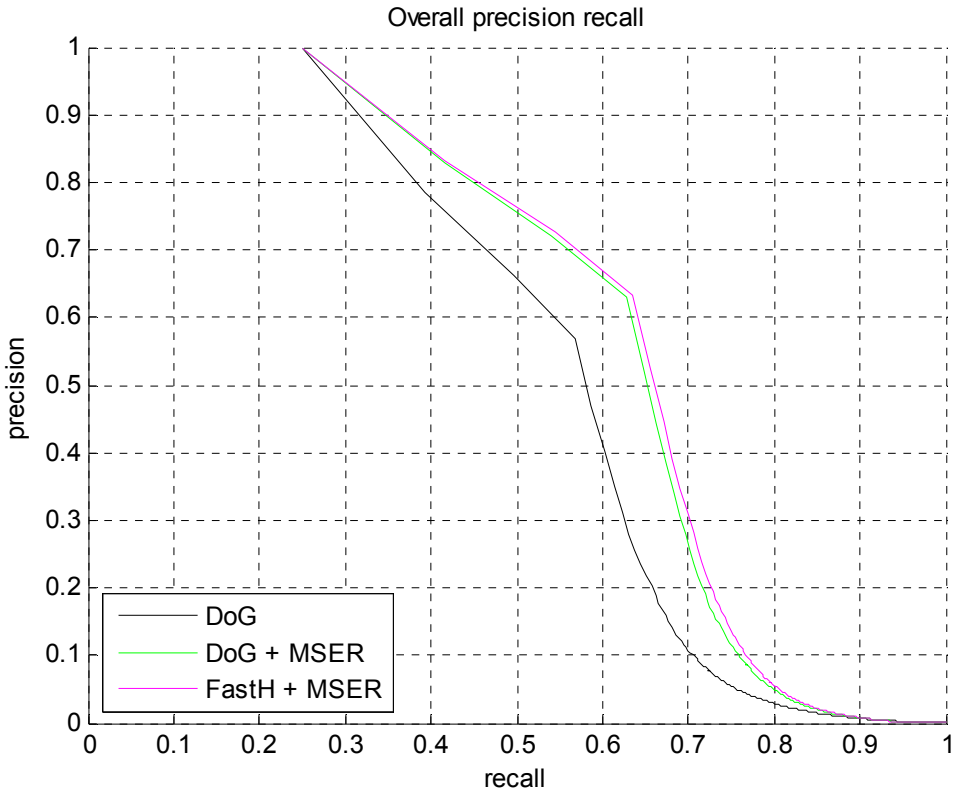
Διεξάγουμε τα ίδια πειράματα αναζήτησης εικόνων στις συλλογές ZuBuD, Oxford Buildings και UKBench (σε ολόκληρη τη συλλογή UKBench). Τα αντίστοιχα διαγράμματα συνολικής ακριβείας-ανάκτησης περιλαμβάνονται στα σχήματα 4.22 – 4.24. Μαζί με τις καμπύλες για τα δύο λεξικά παρουσιάζεται και η καλύτερη καμπύλη που προέκυψε από τη χρήση ενός μόνο λεξικού, για την ευκολία των συγκρίσεων.




**Σχ. 4.22:** Καμπύλη precision-recall για τη συλλογή ZuBuD με συνδυασμό λεξικών ($s = s_1 * s_2$)




**Σχ. 4.23:** Καμπύλη precision-recall για τη συλλογή Oxford Buildings με συνδυασμό λεξικών ($s = s_1 * s_2$)




**Σχ. 4.24:** Καμπύλη precision-recall για τη συλλογή UKBench με συνδυασμό λεξικών ($s = s_1 * s_2$)

Σε αντίθεση με τον προηγούμενο τρόπο συνδυασμού των λεξικών, παρατηρείται μικρή βελτίωση της καμπύλης precision-recall με συνδυασμό FastH-MSER για το σύνολο ZuBuD σε σχέση με τη χρήση του λεξικού από τη μέθοδο FastH. Στο σύνολο Oxford Buildings και πάλι δεν βελτίωσε το αποτέλεσμα, αλλά στη συλλογή UKBench η βελτίωση της απόδοσης είναι πολύ μεγαλύτερη απ' ότι στον προηγούμενο συνδυασμό. Αυτό φαίνεται και από τον πίνακα 4.3, συγκρίνοντας τα μέτρα μέσης ακριβείας mAP με εκείνα του πίνακα 4.2 για τη συλλογή UKBench.

**Πίνακας 4.3:** Μέση ακρίβεια mAP για το συνδυασμό των μεθόδων και των οπτικών λεξικών ($s = s_1 * s_2$)

| **Μέτρο μέσης ακριβείας (mAP) για συνδυασμό λεξικών** | | | |
|---|---|---|---|
| **ΣΥΛΛΟΓΗ** | **ZuBuD** | **Oxford Buildings** | **UKBench** |
| **λεξικό DoG-MSER** | 0.882421 | 0.309485 | 0.669674 |
| **λεξικό FastH-MSER** | 0.899947 | 0.301652 | 0.67674 |

Χρησιμοποιώντας το λεξικό που προέκυψε από την μέθοδο MSER έχουμε πρόσθετη πληροφορία για τους αφινικούς μετασχηματισμούς των εικόνων. Αυτή η πληροφορία χρησιμοποιείται μαζί με τα λεξικά των δύο άλλων μεθόδων που αντιμετωπίζουν με μεγάλη επιτυχία αλλαγές στην κλίμακα και έτσι το συνολικό αποτέλεσμα είναι περισσότερο εύρωστο ως προς τις δύο αυτές μεταβολές, οι οποίες είναι οι πιο συχνές σε εικόνες της καθημερινής ζωής. Η βελτίωση παρατηρήθηκε στο σύνολο UKBench και μάλιστα σε πολύ μεγαλύτερο ποσοστό όταν χρησιμοποιήθηκε η προσέγγιση της πράξης "AND" και του πολλαπλασιασμού των μέτρων ομοιότητας. Αυτό δείχνει ότι η χρήση του αυστηρότερου μέτρου οδηγεί στην απόρριψη μερικών αρχικών ταιριασμάτων που περιείχαν σφάλματα. Μόνο οι εικόνες για τις οποίες και τα δύο λεξικά δίνουν μεγάλο μέτρο ομοιότητας γίνονται δεκτές ως ταίριασμα με την εικόνα-ερώτημα. Ένα τελευταίο συμπέρασμα που μπορούμε να εξάγουμε είναι ότι ο συνδυασμός δύο διαφορετικών λεξικών με τον δεύτερο τρόπο που περιγράφηκε (αυστηρότερο μέτρο) βελτιώνει την επίδοση τους συστήματος αναζήτησης εικόνων, με την προϋπόθεση ότι οι αρχικές τιμές ακρίβειας-ανάκτησης δεν πέφτουν κάτω από 0.5, διότι τότε τα μέτρα ομοιότητας περιλαμβάνουν λάθη (θόρυβο) και άρα δεν μπορεί να υπολογιστεί μέσω αυτών ένα αξιόπιστο τελικό μέτρο ομοιότητας.

# ΚΕΦΑΛΑΙΟ 5

# Συμπεράσματα και Μελλοντικές Επεκτάσεις

## 5.1 Συνεισφορά

Στην εργασία αυτή αναλύθηκαν διάφορες μέθοδοι ανίχνευσης τοπικών χαρακτηριστικών, οι οποίες χρησιμοποιούνται ευρέως στη βιβλιογραφία για ταίριασμα μεταξύ εικόνων, αναζήτηση αντικειμένων, κατασκευή πανοράματος και πολλές άλλες εφαρμογές. Αφού πρώτα μελετήθηκε το θεωρητικό τους υπόβαθρο, στη συνέχεια κατασκευάστηκε ένα σύστημα στο οποίο ενσωματώθηκαν υπάρχουσες υλοποιήσεις των πιο γνωστών μεθόδων ανίχνευσης σημείων/περιοχών ενδιαφέροντος και τοπικής περιγραφής, με κριτήριο την ταχύτητα επεξεργασίας, τη δυνατότητα ενσωμάτωσης νέων τεχνικών και την ευκολία συνδυασμού μεταξύ των μεθόδων. Το σύστημα αυτό μπορεί να χρησιμοποιηθεί σε πειράματα αξιολόγησης των επιδόσεων των διαφόρων μεθόδων, σε συστήματα αναζήτησης εικόνων και γενικότερα σε προβλήματα αναγνώρισης αντικειμένων. Περισσότερες λεπτομέρειες για τη δομή και τη λειτουργία του συστήματος τοπικών χαρακτηριστικών (Local Invariant Features) βρίσκονται στο Παράρτημα.

Το σύστημα που κατασκευάστηκε εφαρμόστηκε στα συγκριτικά πειράματα όπου μετρήθηκαν οι επιδόσεις των μεθόδων. Το πλεονέκτημα του συστήματος είναι ότι προσφέρει έναν κοινό τρόπο χειρισμού των δεδομένων που σχετίζονται με τα τοπικά χαρακτηριστικά και τους περιγραφείς. Έτσι διευκολύνεται η πειραματική μεθοδολογία για τους διάφορους συνδυασμούς και παρέχεται μία απλή διαδικασία για μελλοντικές δοκιμές νέων μεθόδων και πειραμάτων. Τα πειράματα που έλαβαν χώρα χρησίμευσαν να σχηματιστεί μια γενική άποψη για τις επιδόσεις των πιο γνωστών τεχνικών και να γίνει η επιλογή των πιο κατάλληλων μεθόδων ανίχνευσης και περιγραφής για την εφαρμογή σε προβλήματα αναζήτησης εικόνων. Τέλος, οι συνδυασμοί ανιχνευτών και περιγραφέων που επελέγησαν χρησιμοποιήθηκαν για την δόμηση του συστήματος αναζήτησης εικόνων. Κατασκευάστηκαν οπτικά λεξικά μεγάλου μεγέθους (6.5K οπτικών λέξεων) και έγιναν πειράματα αναζήτησης πάνω σε μεγάλες βάσεις δεδομένων, με πλήθος εικόνων από 1K έως 10K. Με αυτόν τον τρόπο έγινε η σύγκριση των μεθόδων απευθείας ως προς τα αποτελέσματά τους στο πρόβλημα της αναζήτησης εικόνων, κάτι που έχει μεγάλη πρακτική αξία καθώς επιβεβαιώνεται η αξία των τοπικών χαρακτηριστικών. Επιχειρήθηκε επίσης μία βελτίωση των αποτελεσμάτων με τον συνδυασμό διαφορετικών οπτικών λεξικών, που όμως δεν οδήγησε σε εμφανείς θετικές αλλαγές.

## 5.2 Συμπεράσματα

Οι μέθοδοι ανίχνευσης τοπικών χαρακτηριστικών και εξαγωγής περιγραφέων ελέγχθηκαν ως προς τις επιδόσεις τους σε ένα συγκεκριμένο σύνολο εικόνων που θεωρείται σημείο αναφοράς στη βιβλιογραφία. Τα πειράματα είχαν σκοπό τη σύγκριση των μεθόδων ως προς την επαναληψιμότητα, την ακρίβεια και το ταίριασμα των αντίστοιχων περιοχών για κάθε ζεύγος από εικόνες του συνόλου, για το οποίο υπάρχει το σύνολο αληθείας (ground truth) με τη μορφή της ομογραφίας. Τα αντικειμενικά κριτήρια που χρησιμοποιήθηκαν βοηθούν στην εξαγωγή συμπερασμάτων σχετικά με την απόδοση των μεθόδων σε σχέση με το είδος της σκηνής που αναπαριστά η εικόνα, αλλά και σε σχέση με τον μετασχηματισμό μεταξύ των διαφορετικών εικόνων (διαφορετικές γεωμετρικές και φωτομετρικές συνθήκες). Ως αποτέλεσμα της σύγκρισης προκύπτει ότι στις περισσότερες περιπτώσεις, όπως σε περιστροφή της εικόνας, σε αλλαγή κλίμακας, ή σε θόλωμα (μεταβολές που συμβαίνουν συχνά), οι καλύτερες τεχνικές είναι οι Difference Of Gaussians και Fast Hessian, με τη δεύτερη να υπερτερεί και εξαιτίας της ταχύτητάς της. Οι τεχνικές αυτές είναι κατασκευασμένες για να χειρίζονται αλλαγές στην κλίμακα. Στην περίπτωση όμως που έχουμε αλλαγή στην οπτική γωνία, οπότε οι δύο εικόνες συνδέονται με έναν αφινικό μετασχηματισμό, η κατάλληλη μέθοδος για να εντοπίσει τις ίδιες περιοχές ενδιαφέροντος με μικρό σφάλμα επικάλυψης είναι η MSER, όπως προκύπτει από την επαναληψιμότητα των περιοχών αυτών.

Στη συνέχεια, οι τρεις επικρατέστερες μέθοδοι εφαρμόστηκαν σε ένα σύστημα αναζήτησης εικόνων, το οποίο χρησιμοποιεί ένα μηχανισμό ανάλογο με την αναζήτηση κειμένου. Ένα οπτικό λεξικό δημιουργείται ομαδοποιώντας τα διανύσματα των περιγραφέων και στη συνέχεια κατασκευάζεται ένα διάνυσμα αναπαράστασης για κάθε εικόνα, με τη συχνότητα εμφάνισης κάθε οπτικής λέξης. Για να ταιριάξει μία εικόνα-ερώτημα με κάποια ή κάποιες εικόνες από τη βάση, εκτιμάται η ομοιότητα των διανυσμάτων αναπαράστασης και έτσι ολόκληρη η συλλογή ταξινομείται με σειρά φθίνουσας ομοιότητας. Ορίζοντας μέτρα αξιολόγησης της αναζήτησης εικόνων, όπως είναι τα μέτρα ακριβείας, ανάκτησης και μέσης ακριβείας (mAP), τα οπτικά λεξικά που προέκυψαν από τις διάφορες μεθόδους τοπικών χαρακτηριστικών συγκρίθηκαν ως προς τις επιδόσεις τους για τρία διαφορετικά σύνολα πολλών εικόνων, που έχουν χρησιμοποιηθεί σε παρόμοια πειράματα στη βιβλιογραφία. Η επιτυχία των μεθόδων Difference Of Gaussians και Fast Hessian στις δύο συλλογές ZuBuD και UKBench δεν ακολουθήθηκε εξίσου από τη μέθοδο MSER. Επίσης, στην τρίτη συλλογή Oxford Buildings η επίδοση όλων των μεθόδων ήταν χαμηλή και αυτό πιο πολύ οφείλεται στο περιεχόμενο των εικόνων, όπου εκτός από τα κτίρια που αναζητούνται στα ερωτήματα, υπάρχουν μέσα στις εικόνες πολλά επικαλυπτόμενα αντικείμενα ή πολύπλοκες σκηνές που εμποδίζουν την αναγνώριση των κτιρίων. Τέλος, πραγματοποιήθηκε ένας συνδυασμός των λεξικών που προήλθαν από τις δύο μεθόδους που είναι ανεξάρτητες από αλλαγές κλίμακας (Difference Of Gaussians, Fast Hessian) και από την αφινική μέθοδο MSER. Με την ταυτόχρονη αναπαράσταση του περιεχομένου των εικόνων στις δύο διαφορετικές περιγραφές (δύο οπτικά λεξικά), τα αποτελέσματα της ανάκτησης φάνηκαν ελαφρώς να βελτιώνονται. Χρησιμοποιήθηκαν δύο διαφορετικοί τρόποι για τον συνδυασμό των μέτρων ομοιότητας, ένας περισσότερο επιεικής (OR) και ένας αυστηρότερος (AND), με τον δεύτερο να δίνει αισθητά μεγαλύτερη βελτίωση. Η απόδοση του συστήματος με τον συνδυασμό λεξικών αυξάνεται κυρίως στη συλλογή UKBench (σε σχέση με τις υπόλοιπες), στην οποία τα μέτρα επίδοσης βρίσκονται σε ενδιάμεσα επίπεδα. Η κατάλληλη προσέγγιση για την εκτίμηση της τελικής ομοιότητας απαιτεί περισσότερη έρευνα.

## 5.3 Μελλοντικές επεκτάσεις

Στα πειράματα του Κεφαλαίου 3 για τις επιδόσεις των μεθόδων και συγκεκριμένα στα σύνολα στα οποία έχουμε μετασχηματισμούς στην οπτική γωνία, παρατηρείται συχνά ότι οι αφινικές μέθοδοι ενώ ως προς την επαναληψιμότητα (repeatability) είναι κατά πολύ ανώτερες από τις υπόλοιπες, οι επιδόσεις τους μειώνονται κατά πολύ όταν υπολογιστεί το μέτρο ταιριάσματος (matching score). Για τη βελτίωση της απόδοσης της μεθόδου MSER, αλλά και των υπόλοιπων αφινικών, ως προς το ταίριασμα των εικόνων, θα πρέπει να μελετηθεί καλύτερα η περιοχή της εικόνας απ' όπου εξάγεται ο περιγραφέας. Η κανονικοποιημένη αυτή περιοχή περιέχει την πληροφορία που χρειάζεται για να περιγραφεί τοπικά η εικόνα και γι' αυτό ο τρόπος υπολογισμού της είναι πολύ σημαντικός. Μπορούν να χρησιμοποιηθούν διάφορα μεγέθη περιοχών μέτρησης (με κλίμακα πολλαπλάσια της αρχικής, για παράδειγμα ×2, ×3, ×4) και μέσω πειραμάτων να εντοπιστεί το κατάλληλο μέγεθος, το οποίο μπορεί να εξαρτάται και από το περιεχόμενο της εικόνας.

Στο σύστημα της αναζήτησης εικόνων, βλέπουμε ότι παρά την πολύ καλή απόδοση όλων των μεθόδων στη συλλογή ZuBuD με τις 1005 εικόνες, η επιτυχία τους μειώνεται αρκετά στην περίπτωση μεγαλύτερων συνόλων, όπως το UKBench. Γι' αυτό θα πρέπει να κατασκευαστούν μεγαλύτερα λεξικά (>6.5K) τα οποία θα περιέχουν περισσότερες οπτικές λέξεις, άρα θα μπορούν εν γένει να μοντελοποιήσουν καλύτερα ένα μεγαλύτερο εύρος από αντικείμενα. Επίσης, μία άλλη παρατήρηση σε όλα τα σύνολα δεδομένων είναι ότι η μέθοδος MSER έχει αρκετά χαμηλότερες επιδόσεις σε σχέση με τις υπόλοιπες. Μάλιστα οι περισσότερες εικόνες στις βάσεις αυτές περιέχουν μετασχηματισμούς στην οπτική γωνία (αφινικούς μετασχηματισμούς), κάτι που η μέθοδος MSER είναι κατάλληλη να αντιμετωπίζει, άρα θα αναμέναμε τουλάχιστον παρόμοιες αν όχι καλύτερες επιδόσεις. Προς την κατεύθυνση της αντιμετώπισης των αφινικών μετασχηματισμών θα μπορούσαμε να χρησιμοποιήσουμε μια μέθοδο πλήρως αναλλοίωτη από αφινικούς μετασχηματισμούς, όπου ο πίνακας μετασχηματισμού U δεν θα είναι πλέον συμμετρικός αλλά τυχαίος (δηλαδή με έναν παραπάνω βαθμό ελευθερίας). Για τον σκοπό αυτό δεν αρκεί ο συμμετρικός πίνακας των ροπών δεύτερης τάξης, αλλά θα πρέπει να ακολουθηθεί μία άλλη γεωμετρική προσέγγιση μέσω της τυχαίας περιοχής (συνεκτικής συνιστώσας) που προκύπτει από τη μέθοδο MSER, όπως για παράδειγμα στα [Obdrzalek et al., 2002a], [Obdrzalek et al., 2002b].

Τέλος, περισσότερα πειράματα αναζήτησης πρέπει να γίνουν με τη χρήση διαφορετικών οπτικών λεξικών, δεδομένης της βελτίωσης των αποτελεσμάτων που επιτεύχθηκε στο σύνολο UKBench. Χρησιμοποιώντας το λεξικό που προέκυψε από την μέθοδο MSER έχουμε πρόσθετη πληροφορία για τους αφινικούς μετασχηματισμούς των εικόνων. Με την ίδια λογική μπορούμε να συνδυάσουμε περισσότερες μεθόδους που εξάγουν διαφορετικά χαρακτηριστικά (συνδυάζοντας για παράδειγμα τη μέθοδο γωνιών Harris και τις μεθόδους ανίχνευσης κηλίδων DoG ή FastH) για να βελτιώσουμε ακόμα περισσότερο τα αποτελέσματα. Παρόλ' αυτά, εκτενέστερη μελέτη θα πρέπει να γίνει ως προς τον μηχανισμό συνδυασμού των δύο οπτικών λεξικών και να εξεταστούν διαφορετικοί τρόποι υπολογισμού της τελικής ομοιότητας μέσω των διαφορετικών διανυσμάτων αναπαράστασης (πότε και γιατί βελτιώνεται η απάντηση του συστήματος αναζήτησης).

# Παράρτημα

Στο Παράρτημα αυτό περιγράφεται το σύστημα τοπικών χαρακτηριστικών LIF (Local Invariant Features) που κατασκευάστηκε και χρησιμοποιήθηκε σε όλα τα πειράματα της παρούσας εργασίας. Στο σύστημα αυτό ενσωματώθηκαν όλες οι μέθοδοι ανίχνευσης τοπικών χαρακτηριστικών οι οποίες αναλύθηκαν θεωρητικά στο Κεφάλαιο 2 και μελετήθηκαν πειραματικά στο Κεφάλαιο 3. Επίσης, το σύστημα LIF, εξάγοντας τα διανύσματα χαρακτηριστικών (περιγραφείς) από κάθε εικόνα, παρέχει τα απαραίτητα δεδομένα για είσοδο στο επόμενο σύστημα για αναζήτηση εικόνων του Κεφαλαίου 4. Στη συνέχεια παρατίθενται με τη σειρά: η γλώσσα και το περιβάλλον προγραμματισμού, καθώς και οι βιβλιοθήκες που χρησιμοποιήθηκαν, η περιγραφή του συστήματος LIF ως προς τη δομή του (δομές δεδομένων, συναρτήσεις, περιβάλλον διεπαφής) και τέλος η μεθοδολογία χρήσης του.

## Π.1 Γλώσσα/περιβάλλον προγραμματισμού και βιβλιοθήκες

Για την υλοποίηση του συστήματος χρησιμοποιήθηκε η γλώσσα προγραμματισμού C++ κυρίως γιατί προσφέρει ταχύτητα στο στάδιο της εκτέλεσης, κάτι που είναι πολύ σημαντικό στη χαμηλού επιπέδου επεξεργασία (low-level processing) που απαιτούν τα τοπικά χαρακτηριστικά της εικόνας. Ένα άλλο πλεονέκτημα της χρήσης της C++ είναι η εκμετάλλευση πολλών βιβλιοθηκών και έτοιμων προγραμμάτων που είναι γραμμένα σε C και C++, αλλά και η εύκολη ενσωμάτωση σε υλοποιήσεις στο περιβάλλον προγραμματισμού MATLAB, που προσφέρει χρήσιμα εργαλεία οπτικοποίησης και παρουσίασης των αποτελεσμάτων. Επίσης, παρόλο που ο κώδικας C++ δεν είναι άμεσα μεταφέρσιμος μεταξύ διαφορετικών λειτουργικών συστημάτων (cross platform, Windows, Linux, Macintosh), με κάποιες μικρές συγκεκριμένες αλλαγές μπορεί εύκολα να χρησιμοποιηθεί από διαφορετικές πλατφόρμες. Η εφαρμογή αναπτύχθηκε σε λειτουργικό σύστημα Microsoft Windows, οπότε ως περιβάλλον προγραμματισμού χρησιμοποιήθηκε το Microsoft Visual C++ 2005, ένα ολοκληρωμένο περιβάλλον ανάπτυξης προγραμμάτων (IDE – Integrated Development Environment).

Για όλες τις διεργασίες χαμηλού επιπέδου στην εικόνα, όπως για παράδειγμα το διάβασμα και η αποθήκευση της εικόνας ή ο μετασχηματισμός των σημείων της, χρησιμοποιείται η βιβλιοθήκη OpenCV [OpenCV]. Η βιβλιοθήκη αυτή παρέχει πολλές αποδοτικές διεργασίες γραμμένες σε C και C++, που καλύπτουν τις περισσότερες ανάγκες για την ανάλυση εικόνας και βίντεο. Για τους ανιχνευτές σημείων DoG (Difference of Gaussians) και περιοχών MSER, όπως και για την εξαγωγή του περιγραφέα SIFT χρησιμοποιήθηκε ο κώδικας από τη βιβλιοθήκη VLFeat, η οποία περιλαμβάνει πολλές λειτουργίες που σχετίζονται με ανάλυση της εικόνας με τοπικά χαρακτηριστικά [VLFeat]. Για τον ανιχνευτή FastH (Fast Hessian) όπως και για τον περιγραφέα SURF χρησιμοποιείται το πακέτο με την δυναμική βιβλιοθήκη κλειστού κώδικα (dll – dynamic link library) από τους συγγραφείς των σχετικών δημοσιεύσεων [SURF]. Για την

ανίχνευση των σημείων Harris Affine και Hessian Affine καλείται το εκτελέσιμο αρχείο που παρέχεται στην ιστοσελίδα των συγγραφέων [VGG affine]. Τέλος σημειώνουμε ότι έγινε χρήση της βιβλιοθήκης IVL η οποία αναπτύσσεται εντός του εργαστηρίου IVML.

## Π.2 Δομή του συστήματος τοπικών χαρακτηριστικών

Στην παράγραφο αυτή περιγράφονται οι δομές και συναρτήσεις που αναπτύχθηκαν και χρησιμοποιήθηκαν για τις λειτουργίες του συστήματος LIF. Καταρχήν θα πρέπει να πούμε ότι εκτός των προκαθορισμένων τύπων και δομών δεδομένων της C++ (STL – standard template library) χρησιμοποιήθηκαν δύο δομές της βιβλιοθήκης IVL, η κλάση array που μοντελοποιεί έναν πίνακα με στοιχεία οποιουδήποτε (αλλά του ιδίου) τύπου και η κλάση image που μοντελοποιεί πολλές λειτουργίες που σχετίζονται με την εικόνα. Για το σύστημα LIF οι δύο βασικές δομές δεδομένων είναι η κλάση interest_point, που περιλαμβάνει όλες τις πληροφορίες που σχετίζονται με ένα σημείο ενδιαφέροντος, και η κλάση local_features που μοντελοποιεί τις διαδικασίες των τοπικών χαρακτηριστικών,. δηλαδή το διάβασμα της εικόνας, τις απαραίτητες μετατροπές τύπων, την ανίχνευση σημείων ενδιαφέροντος και την εξαγωγή περιγραφέων για τα σημεία αυτά. Παρακάτω παρουσιάζεται η δομή των δύο αυτών κλάσεων, των δεδομένων-μελών (data members) και των πιο βασικών μελών-συναρτήσεών τους (member functions), ακολουθούμενα από μία σύντομη περιγραφή.

***Κλάση* interest_point**

***Δεδομένα-μέλη***

| | |
|---|---|
| abscissa/ordinate | : συντεταγμένες του σημείου ενδιαφέροντος |
| scale | : χαρακτηριστική κλίμακα του σημείου |
| orientation | : γωνία της κύρια κατεύθυνσης |
| strength | : βαθμός εγκυρότητας της ανίχνευσης του σημείου |
| affine_a/b/c/d | : στοιχεία του πίνακα M (αφινικό σχήμα) |
| feature_vector | : διάνυσμα χαρακτηριστικών (με 64 ή 128 στοιχεία) |

***Μέλη-συναρτήσεις***

| | |
|---|---|
| clear | : καθαρισμός όλων των δεδομένων του σημείου |
| estimate_circle | : υπολογισμός της ακτίνας του κύκλου γύρω από σημείο |
| estimate_scale | : υπολογισμός της κλίμακας μια ελλειπτικής περιοχής |

***Κλάση* local_features**

***Δεδομένα-μέλη***

| | |
|---|---|
| intensity_image | : οι τιμές της έντασης I της εικόνας (αποθηκευμένες στη δομή image) |
| float_data | : οι τιμές της έντασης σε πίνακα από float (32 bits, εύρος 0.0-255.0) |
| double_data | : οι τιμές της έντασης σε πίνακα double (64 bits, εύρος 0.0-1.0) |
| uchar_data | : οι τιμές της έντασης σε πίνακα unsigned char (8 bits, εύρος 0-255) |
| detector_method | : η χρησιμοποιούμενη μέθοδος ανίχνευσης (ακεραίου τύπου) |
| descriptor_method | : η χρησιμοποιούμενη μέθοδος περιγραφής (ακεραίου τύπου) |
| fvector_size | : το μήκος του διανύσματος χαρακτηριστικών |
| num_of_points | : το πλήθος των σημείων/περιοχών ενδιαφέροντος που εντοπίστηκαν |

interest_array : πίνακας που περιέχει όλα τα σημεία/περιοχές ενδιαφέροντος μαζί με τα διανύσματα χαρακτηριστικών (αποθηκευμένα στη δομή interest_point)

***Μέλη-συναρτήσεις***

detector : ανίχνευση των σημείων/περιοχών ενδιαφέροντος και αποθήκευσή τους στον πίνακα interest_array για περαιτέρω επεξεργασία

descriptor : κατασκευή του διανύσματος χαρακτηριστικών για κάθε σημείο του πίνακα interest_array και αποθήκευση των τιμών στον πίνακα

extract : ανίχνευση και περιγραφή των σημείων/περιοχών ενδιαφέροντος και αποθήκευσή τους στον πίνακα interest_array

load_txt_features : εισάγει τα τοπικά χαρακτηριστικά από αρχείο κειμένου και αποθήκευσή τους στη δομή του πίνακα interest_array

save_txt_features : εξάγει σε αρχείο κειμένου τα τοπικά χαρακτηριστικά που έχουν υπολογιστεί και βρίσκονται στον πίνακα interest_array

load_bin_features : εισάγει τα τοπικά χαρακτηριστικά από δυαδικό αρχείο και αποθήκευσή τους στη δομή του πίνακα interest_array

save_bin_features : εξάγει σε δυαδικό αρχείο τα τοπικά χαρακτηριστικά που έχουν υπολογιστεί και βρίσκονται στον πίνακα interest_array

**Σημείωση 1:** Χρησιμοποιούνται τρεις διαφορετικοί τύποι για τις τιμές των δεδομένων της έντασης I της εικόνας, καθώς σε κάθε μέθοδο απαιτείται διαφορετική είσοδος ως εξής:
MSER: τιμές unsigned char στο διάστημα 0 έως 255
DoG και SIFT: τιμές float στο διάστημα 0 έως 255
FastH και SURF: τιμές double στο διάστημα 0 έως 1

**Σημείωση 2:** Για την αποθήκευση των τοπικών χαρακτηριστικών και των περιγραφέων σε αρχεία υπάρχουν κάποιες συγκεκριμένες μορφές διάταξης, ανάλογα με το είδος του αρχείου. Επίσης για τον προσδιορισμό των τύπων των μεθόδων χρησιμοποιούνται προκαθορισμένες ακέραιες τιμές. Όλα τα παραπάνω αναφέρονται εκτενώς στην επόμενη παράγραφο.

Εκτός από τις δύο προηγούμενες δομές δεδομένων του συστήματος LIF, για την ανίχνευση των τοπικών χαρακτηριστικών και την κατασκευή του περιγραφέα χρησιμοποιείται μια σειρά από συναρτήσεις, στις οποίες περιλαμβάνονται οι διαδικασίες με την κατάλληλη χρήση των βιβλιοθηκών που προαναφέρθηκαν. Οι συναρτήσεις αυτές παρουσιάζονται εδώ με συνοπτική περιγραφή της λειτουργίας τους.

detect_diff_of_gauss : ανίχνευση των σημείων Difference Of Gaussian (DoG)

compute_sift : υπολογισμός του διανύσματος χαρακτηριστικών (περιγραφέα) SIFT σε σημεία ενδιαφέροντος που έχουν ήδη ανιχνευθεί

extract_sift : ανίχνευση των σημείων DoG και εξαγωγή του περιγραφέα SIFT

detect_fast_hessian : ανίχνευση των σημείων Fast Hessian (FastH)

compute_surf : υπολογισμός του διανύσματος χαρακτηριστικών (περιγραφέα) SURF σε σημεία ενδιαφέροντος που έχουν ήδη ανιχνευθεί

extract_surf : ανίχνευση των σημείων FastH και εξαγωγή του περιγραφέα SURF

detect_mser : ανίχνευση των περιοχών MSER (προσαρμοσμένων ελλείψεων)

detect_haraff : ανίχνευση των σημείων Harris Affine

detect_hesaff : ανίχνευση των σημείων Hessian Affine

compute_affine_features : κανονικοποίηση της περιοχής ενδιαφέροντος (από έλλειψη σε κύκλο) και στη συνέχεια εξαγωγή του ζητούμενου περιγραφέα

## Π.3 Χρήση του συστήματος τοπικών χαρακτηριστικών

Για την διαδικασία ανίχνευσης τοπικών χαρακτηριστικών και εξαγωγής περιγραφέα σε μια δεδομένη εικόνα υπάρχουν τρεις τρόποι χρήσης του συστήματος LIF.

***1. Χρήση των δομών δεδομένων του LIF:***

Ο πρώτος τρόπος είναι να δηλωθεί ένα αντικείμενο της κλάσης local_features με δεδομένη την εικόνα που θα αναλυθεί (είσοδος ως όνομα αρχείου ή απευθείας εικόνα του τύπου image). Στη συνέχεια καλείται η συνάρτηση extract με παραμέτρους τις κατάλληλες τιμές για ανιχνευτή και περιγραφέα. Οι διαθέσιμες ακέραιες τιμές και οι αντίστοιχοι τύποι τοπικών ανιχνευτών και περιγραφέων είναι:

***Τοπικοί ανιχνευτές (detectors):***
0 – κανένας
1 – Difference Of Gaussian
2 – Fast Hessian
3 – MSER
4 – Haris Affine
5 – Hessian Affine

***Τοπικοί περιγραφείς (descriptors):***
0 – κανένας
1 – SIFT
2 – SURF

**Σημείωση:** Οι ανιχνευτές Haris Affine και Hessian Affine χρησιμοποιούν το εκτελέσιμο αρχείο από την ερευνητική ομάδα [VGG affine], το οποίο απαιτεί εγκατεστημένο το περιβάλλον cygwin [Cygwin] για τη σωστή λειτουργία του.

Στη συνέχεια τα σημεία ενδιαφέροντος μαζί με τους υπολογισμένους περιγραφείς μπορούν να μεταφερθούν και να σωθούν σε αρχείο με τις συναρτήσεις save_txt_features ή save_bin_features (για αρχείο κειμένου ή δυαδικό αρχείο αντίστοιχα).

Η μορφή που έχουν τα περιεχόμενα των δύο ειδών αρχείων είναι η εξής:

***Αρχεία κειμένου (1ο είδος):***
Η πρώτη γραμμή περιέχει το μήκος του διανύσματος του περιγραφέα και η δεύτερη γραμμή το πλήθος των τοπικών χαρακτηριστικών που έχουν ανιχνευθεί στη εικόνα.
Στη συνέχεια κάθε γραμμή περιέχει τα δεδομένα για ένα συγκεκριμένο τοπικό χαρακτηριστικό και τον περιγραφέα του, με τη δομή: συντεταγμένες (x και y), στοιχεία του πίνακα M αφινικού σχήματος (a, b, c) και στο τέλος τις τιμές του διανύσματος του περιγραφέα.

***Δυαδικά αρχεία (2ο είδος):***
Τα δυαδικά αρχεία αποθηκεύονται ανά ζεύγη: ένα αρχείο (αρχείο 1) με τα στοιχεία των τοπικών χαρακτηριστικών (σημεία/περιοχές ενδιαφέροντος) και ένα αρχείο (αρχείο 2) με τις τιμές του περιγραφέα για τοπικά χαρακτηριστικά. Και πάλι οι δύο πρώτες γραμμές περιέχουν πληροφορίες για το μέγεθος των αρχείων, δηλαδή η πρώτη γραμμή περιέχει το πλήθος των στοιχείων σε κάθε γραμμή και η δεύτερη γραμμή το πλήθος των τοπικών χαρακτηριστικών που έχουν ανιχνευθεί στη εικόνα.

Στο αρχείο 1 η διάταξη των πληροφοριών σε κάθε γραμμή είναι: συντεταγμένες (x και y), χαρακτηριστική κλίμακα (scale), κύρια κατεύθυνση (orientation), βαθμός εγκυρότητας (strength) και τέλος τα στοιχεία του πίνακα M αφινικού σχήματος (a, b, c, d), όπου b = c.
Στο αρχείο 2 σε κάθε γραμμή περιλαμβάνονται οι τιμές για τον περιγραφέα του αντίστοιχου τοπικού χαρακτηριστικού.

***2. Χρήση των συναρτήσεων του LIF:***

Ο δεύτερος τρόπος για τη χρήση του συστήματος LIF (ανίχνευση τοπικών χαρακτηριστικών και εξαγωγή περιγραφέα) είναι μέσω των υλοποιημένων συναρτήσεων που εκτελούν εσωτερικά την όλη διαδικασία και επιστρέφουν στο χρήστη τη δομή που επιθυμεί (wrappers). Μερικές από αυτές τις χρήσιμες συναρτήσεις παραθέτουμε με μία απλή τεκμηρίωση στη συνέχεια. Στην περιγραφή των συναρτήσεων χρησιμοποιούνται δύο "wildcards" s και t, τα οποία αντιστοιχούν σε περισσότερα από ένα ονόματα συναρτήσεων ως εξής:

s: array, vector, list (για πίνακα, διάνυσμα ή λίστα αντίστοιχα)
t: txtfile, binfile (για αρχείο κειμένου ή δυαδικό αρχείο αντίστοιχα)

***ΧΡΗΣΙΜΕΣ ΣΥΝΑΡΤΗΣΕΙΣ***

- **extract_features_s**

ΕΙΣΟΔΟΣ : εικόνα (με όνομα αρχείου ή δομή image)
ΔΙΑΔΙΚΑΣΙΑ : ανίχνευση τοπικών χαρακτηριστικών και εξαγωγή περιγραφέα
ΕΞΟΔΟΣ : τοπικά χαρακτηριστικά και περιγραφείς στη δομή που καθορίζεται στο s

- **export_features_t**

ΕΙΣΟΔΟΣ : εικόνα (με όνομα αρχείου ή δομή image)
ΔΙΑΔΙΚΑΣΙΑ : ανίχνευση τοπικών χαρακτηριστικών και εξαγωγή περιγραφέα
ΕΞΟΔΟΣ : τοπικά χαρακτηριστικά και περιγραφείς σε αρχείο είδους t (κείμενο ή δυαδικό)

- **import_t_features_s**

ΕΙΣΟΔΟΣ : αρχείο με τοπικά χαρακτηριστικά και περιγραφείς (t το είδος του αρχείου)
ΔΙΑΔΙΚΑΣΙΑ : διάβασμα των περιεχομένων των αρχείων
ΕΞΟΔΟΣ : τοπικά χαρακτηριστικά και περιγραφείς στη δομή που καθορίζεται στο s

- **extract_descriptor_s**

ΕΙΣΟΔΟΣ : εικόνα (με όνομα αρχείου ή δομή image)
ΔΙΑΔΙΚΑΣΙΑ : ανίχνευση τοπικών χαρακτηριστικών και εξαγωγή περιγραφέα
ΕΞΟΔΟΣ : μόνο περιγραφείς στη δομή που καθορίζεται στο s

Συναρτήσεις για τον καθορισμό των παραμέτρων των μεθόδων:

- **set_dog_sift_params**
- **set_fasth_surf_params**
- **set_mser_params**
- **set_haraff_params**
- **set_hesaff_params**

***3. Χρήση του εκτελέσιμου αρχείου LIF:***

Τέλος, υπάρχει και ένα εκτελέσιμο αρχείο "lif.exe", το οποίο παρέχει σε περιβάλλον κελύφους των Windows μια διεπαφή χρήσης API (application programming interface) του

συστήματος LIF, απλά δίνοντας σαν είσοδο μια εικόνα ή ένα φάκελο με εικόνες και τις παραμέτρους που καθορίζουν τις επιθυμητές μεθόδους και παίρνοντας ως έξοδο τα τοπικά χαρακτηριστικά σε αρχεία που αποθηκεύονται σε συγκεκριμένο φάκελο. Ακολουθούν μερικές χρήσιμες οδηγίες για το εκτελέσιμο αυτό.

Υπόδειξη κλήσης του εκτελέσιμου αρχείου:

lif.exe [input_type] [input] [output_type] [detector_type] [descriptor_type]

Το όρισμα [input] καθορίζει την είσοδο για τη διαδικασία, δηλαδή το όνομα του αρχείου της εικόνας ή το όνομα του φακέλου που περιέχει τις εικόνες τις οποίες επιθυμούμε να αναλύσουμε.

Το όρισμα [input_type] καθορίζει το είδος της εισόδου, δηλαδή αν είναι:
"input_type" –i : τότε η είσοδος είναι ένα αρχείο εικόνας
"input_type" –f : τότε η είσοδος είναι ένας φάκελος με εικόνες

Το όρισμα [output_type] καθορίζει το είδος του αρχείου εξόδου, που θα περιέχει τα τοπικά χαρακτηριστικά και τους περιγραφείς, και συγκεκριμένα:
"output_type" -b : τα χαρακτηριστικά και οι περιγραφείς αποθηκεύονται σε δυαδικό αρχείο
"output_type" -t : τα χαρακτηριστικά και οι περιγραφείς αποθηκεύονται σε αρχείο κειμένου

Τα ορίσματα [detector_type] και [descriptor_type] προσδιορίζουν τις επιθυμητές μεθόδους τοπικών ανιχνευτών και περιγραφέων, οι οποίες δίνονται σε ακέραιες τιμές (0, 1, 2,...). Αν δεν προσδιοριστούν τύποι για τις μεθόδους, τότε κατασκευάζονται και εκτιμώνται όλοι οι συνδυασμοί, διαδικασία χρήσιμη κυρίως για περιπτώσεις πειραμάτων, όπως στο Κεφάλαιο 3.

**Σημείωση 1:** η μορφή για τα αρχεία εξόδου και οι τύποι των μεθόδων (ακέραιες τιμές) είναι ακριβώς όπως περιγράφηκαν παραπάνω.

**Σημείωση 2:** τα αρχεία κειμένου έχουν κατάληξη ανάλογη με τις μεθόδους ανίχνευσης και περιγραφής που θα χρησιμοποιηθούν, ενώ τα δυαδικά αρχεία (ζεύγος) έχουν κατάληξη ".par" και ".des" για τα χαρακτηριστικά και τους περιγραφείς αντίστοιχα.

# Βιβλιογραφία


[Agrawal et al., 2008] M. Agrawal, K. Konolige and M.R. Blas, "CenSurE: Center Surround Extremas for Realtime Feature Detection and Matching", European Conference on Computer Vision (ECCV 2008), pp. 102-115, 2008.

[Baumberg, 2000] A. Baumberg, "Reliable Feature Matching across Widely Separated Views", in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, pp.774-781, 2000.

[Bay et al., 2006] H. Bay, T. Tuytelaars and L. Van Gool, "SURF: Speeded Up Robust Features", in 9th European Conference on Computer Vision, pp. 404-417, Graz, Austria, 7-13 May, 2006.

[Bay et al., 2008] H. Bay, A. Ess, T. Tuytelaars, L. Van Gool, "Speeded-Up Robust Features (SURF)", Computer Vision and Image Understanding (CVIU), vol. 110, no. 3, pp. 346-359, 2008.

[Brown et al., 2002] M. Brown and D.G. Lowe, "Invariant features from interest point groups", In British Machine Vision Conference, Cardiff, Wales, pp. 656-665, 2002.

[Chum et al., 2007] O. Chum, J. Philbin, J. Sivic, M. Isard and A. Zisserman, "Total Recall: Automatic Query Expansion with a Generative Feature Model for Object Retrieval", in the IEEE 11th International Conference on Computer Vision, pp. 1-8, 14-21 October 2007.

[Friedman et al., 1977] J.H. Friedman, J.L. Bentley, and R.A. Finkel, "An Algorithm for Finding Best Matches in Logarithmic Expected Time", in ACM Transactions on Mathematical Software, vol. 3, no. 3, pp. 209-226, September 1977.

[Harris et al., 1988] C. Harris and M. Stephens, "A Combined Corner and Edge Detector", in Alvey Vision Conference, pp. 147–151, 1988.

[Kadir et al., 2001] T. Kadir and M. Brady, “Scale, saliency and image description”, in International Journal of Computer Vision, vol. 45, no. 2, pp. 83-105, 2001.

[Kadir et al., 2004] T. Kadir, A. Zisserman, M. Brady, "An Affine Invariant Salient Region Detector", in European Conference on Computer Vision, pp. 228-241, 2004.

[Ke et al., 2004] Y. Ke and R. Sukthankar, "PCA-SIFT: A more distinctive representation for local image descriptors", In Proc. of the IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 506-513, 27 June-2 July, 2004.

[Kitchen et al., 1982] L. Kitchen and A. Rosenfeld, "Gray-level corner detection", in Pattern Recognition Letters, vol. 1, pp. 95-102, 1982.

[Langridge, 1982] D.J. Langridge, "Curve encoding and detection of discontinuities", in Computer Graphics Image Processing, vol. 20, pp. 58-71, 1982.

[Lazebnik et al., 2006] S. Lazebnik, C. Schmid, J. Ponce, "Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories", in IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, pp.2169-2178, 2006.

[Lindeberg et al., 1997] T. Lindeberg and J. Garding, "Shape-adapted smoothing in estimation of 3-D shape cues from a?ne deformations of local 2-D brightness structure", Image and Vision Computing, vol. 15, no. 6, pp. 415-434, 1997.

[Lindeberg, 1994] T. Lindeberg, "Scale-space theory: A basic tool for analysing structures at different scales", in Journal of Applied Statistics, vol. 21, no. 2, pp. 224-270, 1994.

[Lindeberg, 1998] T. Lindeberg, "Feature Detection with Automatic Scale Selection", International Journal of Computer Vision, vol. 30, no. 2, pp. 79-116, November 1998.

[Lowe, 1999] D.G. Lowe, "Object Recognition from Local Scale-Invariant Features", In Proceedings of the Seventh IEEE International Conference on Computer Vision (ICCV'99), vol. 2, pp. 1150-1157, Corfu, Greece, September 1999.

[Lowe, 2001] D. Lowe, "Local feature view clustering for 3D object recognition", In Proc. of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 682-688, December 2001.

[Lowe, 2004] D.G. Lowe, "Distinctive image features from scale-invariant keypoints", International Journal of Computer Vision, Vol.60, No.2, pp.91-110, 2004.

[Matas et al., 2002] J. Matas, O. Chum, M. Urban, T. Pajdla, "Robust wide baseline stereo from maximally stable extremal regions", In Proceedings of the British Machine Vision Conference, vol. 1, pp. 384-393, UK, September 2002.

[Matas et al., 2004] J. Matas, O. Chum, M. Urban, T. Pajdla, "Robust wide-baseline stereo from maximally stable extremal regions", in Image and Vision Computing, vol. 22, no. 10, pp. 761-767, 1 September, 2004.

[Mikolajczyk et al., 2001] K. Mikolajczyk and C. Schmid, "Indexing based on scale invariant interest points", In Proceedings of the 8th International Conference on Computer Vision, pp. 525-531, 2001.

[Mikolajczyk et al., 2002] K. Mikolajczyk and C. Schmid, "An affine invariant interest point detector", In Proceedings of the 7th European Conference on Computer Vision, vol. 1, pp. 128-142, 2002.

[Mikolajczyk et al., 2003] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors", In Proc. IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 257-263, 2003.

[Mikolajczyk et al., 2004] K. Mikolajczyk and C. Schmid, "Scale & Affine Invariant Interest Point Detectors", International Journal of Computer Vision, vol. 60, no. 1, pp. 63-86, Oct. 2004.

[Mikolajczyk et al., 2005a] K. Mikolajczyk, C. Schmid, "A performance evaluation of local descriptors", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 10, pp. 1615-1630, October 2005.

[Mikolajczyk et al., 2005b] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, L. Van Gool, " A Comparison of Affine Region Detectors", International Journal of Computer Vision, vol. 65, no. 1-2, pp. 43-72, November 2005.

[Muller et al, 2002] H. Muller, S. Marchand-Maillet, and T. Pun, "The Truth about Corel - Evaluation in Image Retrieval", in Proceedings of the International Conference on Image and Video Retrieval, Lecture Notes In Computer Science, vol. 2383, pp. 38-49, July 18 - 19, 2002.

[Nister et al., 2006] D. Nister and H. Stewenius, "Scalable Recognition with a Vocabulary Tree", IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 2161-2168, 2006.

[Nister et al., 2008] D. Nister and H. Stewenius, "Linear Time Maximally Stable Extremal Regions", European Conference on Computer Vision (ECCV 2008), pp. 183-196, 2008.

[Obdrzalek et al., 2002a] S. Obdrzalek and J. Matas, "Local Affine Frames for Image Retrieval", in Proceedings of the International Conference on Image and Video Retrieval, Lecture Notes In Computer Science, vol. 2383, 2002.

[Obdrzalek et al., 2002b] S. Obdrzalek and J. Matas, "Object Recognition using Local Affine Frames on Distinguished Regions", in Proceedings of the British Machine Vision Conference, 2002.

[Philbin et al., 2007] J. Philbin, O. Chum, M. Isard, J. Sivic and A. Zisserman, "Object retrieval with large vocabularies and fast spatial matching", in IEEE Conference on Computer Vision and Pattern Recognition, pp. 1-8, 17-22 June 2007.

[Philbin et al., 2008] J. Philbin, O. Chum, M. Isard, J. Sivic, A. Zisserman, "Lost in Quantization: Improving Particular Object Retrieval in Large Scale Image Databases", In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2008.

[Rosten et al., 2006] E. Rosten and T. Drummond, "Machine Learning for High-Speed Corner Detection", in European Conference on Computer Vision, pp. 430-443, 2006.

[Rui et al, 1999] Y. Rui, T.S. Huang and S. Chang, "Image Retrieval - Current Techniques, Promising Directions, and Open Issues", in Journal of Visual Communication and Image Representation, vol. 10, pp. 39-62, 1999.

[Schaffalitzky et al., 2002] F. Schaffalitzky and A. Zisserman, "Multi-view Matching for Unordered Image Sets, or How Do I Organize My Holiday Snaps?", In Proceedings of the 7th European Conference on Computer Vision-Part I, pp. 414-431, May 28 - 31, 2002.

[Schmid et al., 1997] C. Schmid and R. Mohr, "Local Greyvalue Invariants for Image Retrieval", In IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 5, pp.530-535, May 1997.

[Schmid et al., 2000] C. Schmid, C. Mohr and C. Bauckhage, "Evaluation of Interest Point Detectors", International Journal of Computer Vision, vol. 37, no. 2, pp. 151-172, 2000.

[Shao et al., 2003] H. Shao, T. Svoboda and T. Tuytelaars, "HPAT Indexing for Fast Object/Scene Recognition Based on Local Appearance", in International Conference on Image and Video Retrieval, no. 2, vol. 2728, pp. 71-80, 2003.

[Sivic et al., 2003] J. Sivic and A. Zisserman, "Video Google: A Text Retrieval Approach to Object Matching in Videos", In Proceedings of the Ninth IEEE International Conference on Computer Vision (ICCV'03), vol. 2, pp. 1470-1477, October 2003.

[Smeulders et al, 2000] A.W. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-Based Image Retrieval at the End of the Early Years", in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 12, pp. 1349-1380, December 2000.

[Smith et al., 1997] S.M. Smith and J.M. Brady, "SUSAN - A New Approach to Low Level Image Processing", in International Journal of Computer Vision, vol. 23, no. 1, pp. 45-78, May 1997.

[Squire et al., 2000] D.M. Squire, W. Muller, H. Muller, T. Pun, "Content-based query of image databases: inspirations from text retrieval", in Pattern Recognition Letters, vol. 21, no. 13, pp. 1193-1198, December 2000.

[Tuytelaars et al., 1999] T. Tuytelaars and L. Van Gool, "Content-Based Image Retrieval Based on Local Affinely Invariant Regions", in Proceedings of the 3rd International Conference on Visual Information and Information Systems, pp. 493-500, Amsterdam, The Netherlands, June 1999.

[Tuytelaars et al., 2000] T. Tuytelaars and L. Van Gool, "Wide Baseline Stereo Matching Based on Local, Affinely Invariant Regions", in Proceedings of the 11th British Machine Vision Conference, pp. 412-425, September 2000.

[Tuytelaars et al., 2004] T. Tuytelaars and L. Van Gool, "Matching Widely Separated Views Based on Affine Invariant Regions", International Journal of Computer Vision, vol. 59, no. 1, pp. 61-85, August 2004.

[Tuytelaars et al., 2008] T. Tuytelaars and K. Mikolajczyk, "Local Invariant Feature Detectors: A Survey", Foundations and Trends® in Computer Graphics and Vision, vol. 3, no. 3, pp. 177-280, 2008.

[Cygwin] http://www.cygwin.com/

[FlickR] http://www.flickr.com/

[OpenCV] http://sourceforge.net/projects/opencvlibrary

[OxBuD] http://www.robots.ox.ac.uk/~vgg/data/oxbuildings/index.html

[SURF] http://www.vision.ee.ethz.ch/~surf/

[UKBench] http://vis.uky.edu/~stewe/ukbench/

[VGG affine] http://www.robots.ox.ac.uk/~vgg/research/affine/index.html

[VLFeat] http://www.vlfeat.org/

[ZuBud] http://www.vision.ee.ethz.ch/datasets/index.en.html