



**Εθνικό Μετσόβιο Πολυτεχνείο  
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ  
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ**

ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ & ΥΠΟΛΟΓΙΣΤΩΝ

**Αναζήτηση εικόνων με γεωμετρικά κριτήρια**

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

του

**ΓΕΩΡΓΙΟΥ Κ. ΤΟΛΙΑ**

Διπλωματούχου Ηλεκτρολόγου Μηχανικού &  
Μηχανικού Υπολογιστών Ε.Μ.Π. (2007)

Αθήνα, Οκτώριος 2013





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

**Αναζήτηση εικόνων με γεωμετρικά κριτήρια**

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

του

**ΓΕΩΡΓΙΟΥ Κ. ΤΟΛΙΑ**

Διπλωματούχου Ηλεκτρολόγου Μηχανικού &  
Μηχανικού Υπολογιστών Ε.Μ.Π. (2007)

Συμβουλευτική Επιτροπή:  
Στέφανος Κόλλιας  
Ανδρέας Σταφυλοπάτης  
Γεώργιος Στάμου

Εγκρίθηκε από την επταμελή επιτροπή την 16η Οκτωβρίου 2013.

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

...  
Πέτρος Μαραγκός                  Ιωάννης Εμίρης                  Άλλος ένας  
Καθηγητής Ε.Μ.Π.                  Καθηγητής Ε.Κ.Π.Α.                  Καθηγητής

...  
allos  
Καθηγητής Ε.Μ.Π.  
Αθήνα, Οκτώριος 2013

ΓΕΩΡΓΙΟΣ Κ. ΤΟΛΙΑΣ

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

© 2013 - All rights reserved

# Περιεχόμενα

<b>1</b>	<b>Εισαγωγή</b>	<b>1</b>
1.1	Γενικό υπόβαθρο.....	2
1.2	Συνεισφορά.....	3
<b>2</b>	<b>Χάρτες χαρακτηριστικών</b>	<b>7</b>
2.1	Εισαγωγή και σχετική βιβλιογραφία.....	7
2.2	Υπόβαθρο .....	10
2.2.1	Ταίριασμα σχήματος .....	10
2.2.2	Αντιστοιχίες περιοχών .....	11
2.2.3	Bag-of-words .....	11
2.2.4	Προς τον αλγόριθμο RANSAC.....	12
2.3	Χάρτες χαρακτηριστικών .....	13
2.3.1	Τοπικές περιοχές.....	13
2.3.2	Υπόθεση μονών αντιστοιχιών .....	13
2.3.3	Αναμόρφωση του συνόλου χαρακτηριστικών .....	14
2.3.4	Κβαντισμός.....	15
2.3.5	Χάρτες χαρακτηριστικών .....	17
2.4	Χάρτες χαρακτηριστικών και κατακερματισμός .....	18
2.4.1	Τοπικά εναίσθητος κατακερματισμός .....	19
2.4.2	Τυχαίες αντιμεταθέσεις .....	19
2.4.3	Ταίριασμα σκαριφημάτων χαρτών.....	20
2.4.4	Λεπτομέρειες υλοποίησης.....	21
2.5	Πειράματα - χάρτες και κατακερματισμός .....	24
2.5.1	Βάσεις εικόνων .....	24
2.5.2	Πρωτόκολλο αξιολόγησης .....	24
2.5.3	Βελτιστοποίηση παραμέτρων .....	25
2.5.4	Αποτελέσματα .....	27
2.6	Επιλογή χαρακτηριστικών .....	29
2.6.1	Πηγές χαρτών .....	32

2.6.2	Αναμορφωμένα χαρακτηριστικά .....	33
2.6.3	Μοναδικές εικόνες.....	35
2.6.4	Λεπτομέρειες υλοποίησης.....	36
2.7	Πειράματα - χάρτες κι επιλογή χαρακτηριστικών.....	38
2.7.1	Σύνολα εικόνων .....	38
2.7.2	Πρωτόκολλο αξιολόγησης .....	38
2.7.3	Βελτιστοποίηση παραμέτρων .....	39
2.7.4	Αποτελέσματα .....	41
<b>3</b>	<b>Πυραμίδες Hough</b>	<b>45</b>
3.1	Συνεισφορά και σχετική βιβλιογραφία .....	45
3.2	Ορισμός προβλήματος .....	47
3.3	Ταίριασμα πυραμίδας Hough .....	49
3.3.1	Διαδικασία ταιριάσματος .....	49
3.3.2	Παραδείγματα και συζήτηση .....	51
3.3.3	Ο αλγόριθμος .....	52
3.4	Ταίριασμα με χρήση πολλαπλής ανάθεσης .....	56
3.5	Υλοποίηση.....	58
3.5.1	Δεικτοδότηση κι ανακατάταξη .....	58
3.5.2	Κβαντισμός.....	59
3.5.3	Πρότερη γνώση περιστροφής .....	59
3.6	Πειράματα .....	60
3.6.1	Πρωτόκολλο πειραμάτων.....	60
3.6.2	Πείραμα ταιριάσματος ζευγαριών .....	60
3.6.3	Πειράματα ανακατάταξης.....	61
3.6.4	Ανακατάταξη με την χρήση πολλαπλής ανάθεσης .....	67
<b>4</b>	<b>Επιλογή χαρακτηριστικών μέσω εντοπισμού συμμετριών</b>	<b>71</b>
4.1	Εισαγωγή .....	71
4.2	Σχετική βιβλιογραφία .....	72
4.3	Επιλογή χαρακτηριστικών .....	74
4.3.1	Αναπαράσταση .....	75
4.3.2	Γεωμετρικό αυτο-ταίριασμα .....	77
4.3.3	Ανάστροφο ταίριασμα .....	80
4.3.4	Χαλαρό γεωμετρικό αυτο-ταίριασμα .....	83
4.4	Πειράματα .....	86
4.4.1	Σύνολα εικόνων .....	87
4.4.2	Πρωτόκολλο .....	87

4.4.3	Βελτιστοποίηση παραμέτρων .....	88
4.4.4	Συγκρίσεις .....	91
<b>5</b>	<b>Επέκταση ερωτήματος με δυαδικούς περιγραφείς</b>	<b>93</b>
5.1	Εισαγωγή και σχετική βιβλιογραφία.....	93
5.2	Το βασικό σύστημα αναζήτησης .....	95
5.3	Root-SIFT και Shift-SIFT .....	96
5.4	Hamming Embedding με επέκταση ερωτήματος .....	97
5.4.1	Επιλογή αξιόπιστων εικόνων .....	97
5.4.2	Επέκταση του συνόλου χαρακτηριστικών .....	100
5.5	Επέκταση ερωτήματος με συνάθροιση περιγραφέων .....	101
5.6	Γεωμετρική πληροφορία .....	103
5.7	Πειράματα .....	104
5.7.1	Σύνολα εικόνων και πρωτόκολλο πειραμάτων .....	104
5.7.2	Επίδραση των παραμέτρων .....	105
5.7.3	Σύγκριση με γνωστές μεθόδους.....	108
5.7.4	Πολυπλοκότητα: χρόνος και μέγεθος ερωτήματος .....	109
<b>6</b>	<b>Συμπληρωματικές εργασίες</b>	<b>111</b>
6.1	Χάρτες Σκηνών .....	111
6.2	Επιλεκτικός πυρήνας συνάθροισης.....	113
6.3	VIRaL .....	114
<b>7</b>	<b>Συμπεράσματα</b>	<b>117</b>
	<b>Βιβλιογραφία</b>	<b>121</b>



## ΠΡΟΛΟΓΟΣ

Η ανάλυση εικόνας κέρδισε το ενδιαφέρον μου από νωρίς στα προπτυχιακά μου έτη και στην συνέχεια ακόμα περισσότερο με την εκπόνηση τις διπλωματικής μου εργασίας πάνω σε θέματα εντοπισμού αντικειμένων σε εικόνες. Στην διατριβή μου ασχολήθηκα με το πρόβλημα της αναζήτησης παρόμοιων εικόνων. Τα τελευταία χρόνια η αναζήτηση πληροφορίας σε κείμενα και ιστοσελίδες στο διαδίκτυο έχει γίνει αναπόσπαστο κομμάτι της προσωπικής ζωής αλλά και της εργασίας. Ισχυρά εργαλεία αναζήτησης κειμένου είναι διαθέσιμα σε πολλούς και για πρώτη φορά η αναζήτηση της γνώσης και της αλήθειας είναι δυνατή με πιο εύκολο και άμεσο τρόπο. Οι μέθοδοι αυτόματου ταιριάσματος εικόνων έρχονται να προσθέσουν ισχυρά εργαλεία σε αυτή την αναζήτηση η οποία θα μπορεί να γίνεται και σε οπτικό υλικό το οποίο είναι σήμερα διαθέσιμο σε τεράστιο όγκο.

Θα ήθελα να ευχαριστήσω τον καθηγητή Στέφανο Κόλλια ο οποίος με εμπιστεύθηκε δίνοντας μου την ευκαιρία να ασχοληθώ με όλα αυτά τα ενδιαφέροντα θέματα και ο οποίος ήταν πάντα στο πλευρό μου. Καθ' όλη την διάρκεια της εκπόνησης της διατριβής είχα στενή συνεργασία με τον ερευνητή Γιάννη Αβρίθη. Τον ευχαριστώ ιδιαιτέρως για την πολύτιμη καθοδήγηση του και την εξαιρετική συνεργασία. Επίσης εξαιρετική συνεργασία είχα με τον υποψήφιο διδάκτωρ Γιάννη Καλαντίδη σε μεγάλο μέρος της ερευνητικής μας δουλειάς. Οι άνθρωποι οι οποίοι στελεχώνουν την ευρύτερη ερευνητική ομάδα διαδραμάτισαν όλοι τους σημαντικό ρόλο στην διεκπεραίωση του ερευνητικού μας έργου. Τέλος θα ήθελα να ευχαριστήσω όλους εκείνους που υπήρξαν στο πλευρό μου κατά την διάρκεια των πέντε ετών ενασχόλησης με το αντικείμενο της διατριβής. Όλους εκείνους που καταλάβαιναν και όλους εκείνους που χαριτωμένα δεν καταλάβαιναν τι ακριβώς σημαίνει να ασχολείται κανείς με το πρόβλημα της αναζήτησης εικόνων.

*“It remains completely unknown to us what objects may be in themselves and apart from the receptivity of our senses. We know nothing but our manner of perceiving them.”*

—Immanuel Kant



## ΠΕΡΙΛΗΨΗ

Ένα μεγάλο εύρος από ιδιότητες και υποθέσεις καθορίζουν το πιο κατάλληλο μοντέλο γεωμετρικού ταιριάσματος για εφαρμογές όπως η αναγνώριση κι ο εντοπισμός αντικειμένων, η ευθυγράμμιση εικόνων, ή η ανάκτηση από μεγάλες βάσεις εικόνων. Πιο σημαντικές από αυτές είναι το γεωμετρικά αναλλοίωτο, οι περιορισμοί ακαμψίας, οι περιορισμοί στις αντιστοιχίες, οι υποθέσεις σχετικά με τα χρησιμοποιούμενα τοπικά χαρακτηριστικά ή περιγραφείς, η διακριτική τους ικανότητα και φυσικά το υπολογιστικό κόστος. Στην παρούσα διατριβή θα ασχοληθούμε με το πρόβλημα της ανάκτησης από μεγάλη βάση εικόνων με χρήση εικόνας ερωτήματος. Σκοπός είναι η εύρεση εικόνων οι οποίες απεικονίζουν το ίδιο αντικείμενο από διαφορετικές οπτικές γωνίες ακόμα κι αν αυτό δεν είναι πλήρως ορατό. Για τον σκοπό αυτό προτείνουμε δύο μεθόδους ανάκτησης με χρήση γεωμετρικής πληροφορίας είτε στην δομή δεικτοδότησης είτε σε ένα δεύτερο στάδιο ανακατάξης των εικόνων με γεωμετρικά κριτήρια. Επίσης προτείνουμε τρόπους με τους οποίους το γεωμετρικό ταίριασμα μπορεί να χρησιμοποιηθεί για επιλογή χαρακτηριστικών. Τέλος αναπτύσσουμε μία μέθοδο για επέκταση ερωτήματος η οποία εμπλουτίζει την τοπική αναπάρασταση της εικόνας, επιτυγχάνοντας υψηλή απόδοση και σε χαμηλούς χρόνους.

Παρουσιάζουμε μία νέα μέθοδο για δεικτοδότηση και ανάκτηση εικόνων, η οποία ενσωματώνει την εμφάνιση των τοπικών χαρακτηριστικών μαζί με πληροφορία καθολικής γεωμετρίας στην διαδικασία δεικτοδότησης. Χρησιμοποιούμε τις παραμέτρους σχήματος των τοπικών χαρακτηριστικών για να πάρουμε μία εκτίμηση της ευθυγράμμισης των εικόνων από μονές αντιστοιχίες χαρακτηριστικών. Για κάθε χαρακτηριστικό κατασκευάζουμε έναν αραιό χωρικό χάρτη ο οποίος κωδικοποιεί την κανονικοποιημένη θέση όλων των υπολοίπων χαρακτηριστικών. Μία εικόνα αναπαρίσταται τελικά από ένα σύνολο τέτοιων χαρτών χαρακτηριστικών και τελικά ένα ταίριασμα εικόνων παρόμοιο με εκείνο του αλγόριθμου RANSAC ανάγεται σε έναν αριθμό από εσωτερικά γινόμενα ή τομές συνόλων. Χρησιμοποιούμε την μέθοδο των ανεξάρτητων αντιμεταθέσεων και προτείνουμε ένα μέτρο ομοιότητας για συλλογές χαρτών χαρακτηριστικών. Εκτός από την τυχαία επιλογή χαρακτηριστικών προχωράμε και σε επιλογή τους με βάση το γεωμετρικό ταίριασμα ανάμεσα σε διαφορετικές όψεις του ίδιου αντικειμένου. Αυτή η τεχνική μας επιτρέπει να κατορθώσουμε δεικτοδότηση γεωμετρίας για 1M εικόνες. Εκμεταλλευόμενοι την αραιότητα

της αναπαράστασής μας χρησιμοποιούμε ανάστροφο αρχείο ως δομή δεικτοδότησης μετατρέποντας την αναζήτηση σε διαδικασία υπο-γραμμική ως προς το μέγεθος της βάσης εικόνων.

Επιπλέον προτείνουμε ένα πολύ απλό μοντέλο γεωμετρικού ταιριάσματος (ταίριασμα πυραμίδων Hough) με έμπνευση από τον μετασχηματισμό Hough στον χώρο των γεωμετρικών μετασχηματισμών. Οι ψήφοι Hough, προκύπτουν και πάλι από μονές αντιστοιχίες χαρακτηριστικών. Ο αλγόριθμός μας χρησιμοποιεί χαλαρό γεωμετρικό ταίριασμα το οποίο επιτρέπει το ταίριασμα πολλαπλών αντικειμένων ή επιφανειών, επιβάλλοντας παράλληλα αντιστοιχίση ένα προς ένα με πολυπλοκότητα γραμμική ως προς το πλήθος των αρχικών αντιστοιχιών. Τον εφαρμόζουμε σε ανάκτηση από βάση 2 εκατομμυρίων εικόνων επιδεικνύοντας υψηλότερη απόδοση αλλά και ταχύτερη εκτέλεση από σύγχρονες μεθόδους.

Επεκτείνουμε και χρησιμοποιούμε την μέθοδο ταιριάσματος πυραμίδων Hough για τον εντοπισμό συμμετρικών κι επαναλαμβανόμενων δομών μέσα στην ίδια εικόνα. Υποστηρίζουμε πως τοπικά χαρακτηριστικά τα οποία συμμετέχουν σε τέτοιες δομές έχουν μεγαλύτερη πιθανότητα να ταιριάζουν και μεταξύ διαφορετικών εικόνων. Το αποτέλεσμα του γεωμετρικού ταιριάσματος της εικόνας με τον εαυτό της και το αντικατοπτρισμένο είδωλο της μας δίνει την πληροφορία για την επιλογή χαρακτηριστικών από μοναδικές εικόνες.

Σε αντίθεση με όλες τις προηγούμενες μεθόδους με τις οποίες ασχοληθήκαμε και προτείναμε, οι οποίες χρησιμοποιούν μόνο τις οπτικές λέξεις για ταίριασμα μεταξύ τοπικών χαρακτηριστικών, ασχολούμαστε και με την μέθοδο Hamming Embedding (HE) η οποία κάνει επιπλέον χρήση πληροφορίας του τοπικού περιγραφέα. Κάθε τοπικό χαρακτηριστικό περιγράφεται από μία οπτική λέξη αλλά και μία δυαδική υπογραφή. Προτείνουμε μία μέθοδο για επέκταση ερωτήματος η οποία χρησιμοποιεί την βασική αναπαράσταση της μεθόδου HE και είναι εναρμονισμένη με αυτή. Καταφέρνουμε να βελτιώσουμε την απόδοση ακόμα και χωρίς την χρήση γεωμετρικής πληροφορίας και να πετύχουμε έτσι πολύ χαμηλούς χρόνους εκτέλεσης. Τελικά δείχνουμε πως με επιπλέον γεωμετρικό ταίριασμα η απόδοση βελτιώνεται ακόμα περισσότερο.

## ABSTRACT

A wide range of properties and assumptions determine the most appropriate *spatial matching* model for an application, e.g. recognition, detection, registration, or large scale image retrieval. Most notably, these include discriminative power, geometric invariance, rigidity constraints, mapping constraints, assumptions made on the underlying features or descriptors and, of course, computational complexity.

We present a new approach to image indexing and retrieval, which integrates appearance with global image geometry in the indexing process, while enjoying robustness against viewpoint change, photometric variations, occlusion, and background clutter. We exploit shape parameters of local features to estimate image alignment via a single correspondence. Then, for each feature, we construct a sparse spatial map of all remaining features, encoding their normalized position and appearance, typically vector quantized to visual word. An image is represented by a collection of such *feature maps* and RANSAC-like matching is reduced to a number of set intersections. We use min-wise independent permutations and derive a similarity measure for feature map collections. In addition to random selection, we have further exploited multiple view matching for feature selection. This allows us to scale geometry indexing up to 1M images. We then exploit sparseness to build an inverted file whereby the retrieval process is sub-linear in the total number of images, ideally linear in the number of relevant ones.

We further present a very simple model inspired by Hough voting in the transformation space, where votes arise from single feature correspondences. A *relaxed matching* process allows for multiple matching surfaces or non-rigid objects under *one-to-one* mapping, yet is *linear* in the number of correspondences. We apply it to geometry re-ranking in a search engine, yielding superior performance with the same space requirements but a dramatic *speed-up* compared to the state of the art.

We further extend and use our relaxed spatial matching for self-matching and symmetry detection. We assume that features participating in symmetric and repeating structures have higher probability to be matched between different views of the same object. Information from geometric self-matching and matching of the image with its mirrored counterpart is used for feature selection of single images.

In contrast to the previous methods that we discussed or proposed, which all use only visual word information to perform feature matching, we further exploit the Hamming (HE) Embedding technique, which further use descriptor information. HE employs each feature with visual word and a binary signature which allows more precise feature matching. We develop a novel query expansion strategy which is aligned with the HE representation. We achieve to improve performance even without geometry matching, in contrast to previous query expansion methods, along with low query times. We finally show that combining our scheme with geometry matching can further boost performance and outperform state of the art methods.

# Κεφάλαιο 1

## Εισαγωγή

Η χρήση γεωμετρικής πληροφορίας είναι σημαντική σε πολλά προβλήματα της όρασης υπολογιστών όπως η αντιστοιχία χαρακτηριστικών, η ευθυγράμμιση εικόνων, η αναγνώριση αντικειμένων και η ανάκτηση εικόνων. Αρχικά όταν τα τοπικά χαρακτηριστικά δεν είχαν διακριτική ικανότητα (π.χ. απλά  $2\Delta$  σημεία) η γεωμετρία ήταν ιδιαιτέρως απαραίτητη. Με την έλευση χαρακτηριστικών και περιγραφέων με διακριτική ικανότητα [53, 8], το να αγνοήσει κανείς την γεωμετρία ήταν ένας εύκολος τρόπος να αντιμετωπισθεί η αλλαγή οπτικής γωνίας αντικειμένων καθώς και η μερική ορατότητα τους (occlusion). Η μεγάλη επιτυχία της μεθόδου σάκος λέξεων (bag-of-words) βασίζεται κυρίως στο χαμηλό υπολογιστικό της κόστος, για παράδειγμα στην δουλεία των Sivic και Zisserman [73].

Παρόλα αυτά για να βελτιωθεί η απόδοση σε αναζήτηση μεγάλης κλίμακας, η εφαρμογή γεωμετρίας είναι απαραίτητη. Στην παρούσα διατριβή θα ασχοληθούμε με το πρόβλημα ανάκτησης από μεγάλη βάση εικόνων με χρήση εικόνας ερωτήματος (query image). Σκοπός είναι να ταξινομηθούν οι εικόνες της βάσης σύμφωνα με την ομοιότητα τους με την εικόνα ερωτήματος. Η ομοιότητα στην περίπτωση μας ορίζεται ως απεικόνιση των ίδιων αντικειμένων από διαφορετικές οπτικές γωνίες, υπό διαφορετικές συνθήκες φωτισμού, διαφορετικό υπόβαθρο φωτογραφίας (background) και μερική ορατότητα του αντικειμένου. Μέθοδοι ανάκτησης εικόνων που μελετήθηκαν ή προτάθηκαν εφαρμόστηκαν σε συστήματα αυτόματου γεωγραφικού εντοπισμού εικόνων καθώς και αυτόματου χαρακτηρισμού του περιεχόμενου τους [41][75][42][43].

Οι σύγχρονες μέθοδοι βασίζονται κυρίως στην περιγραφή της εμφάνισης των τοπικών χαρακτηριστικών, με την χρήση περιγραφέων, στο πρώτο στάδιο της αναζήτησης. Το πρώτο αυτό στάδιο ονομάζουμε στάδιο φιλτραρίσματος εικόνων καθώς παρέχεται μία πρώτη ταξινόμηση της βάσης εικόνων με βάση την ομοιότητα τους με την εικόνα ερωτήματος. Κλασικό παράδειγμα μεθόδου φιλτραρίσματος είναι η μέθοδος Bag-of-Words (BoW) σε συνδυασμό με την χρήση ανάστροφου αρχείου [73].

Η εφαρμογή της γεωμετρίας έρχεται σε ένα δεύτερο στάδιο εφαρμοζόμενη ανάζευγάρι

εικόνων και μόνο σε όσες κατατάχθηκαν σε υψηλές θέσεις κατά το στάδιο φιλτραρίσματος. Αυτή η δεύτερη διαδικασία ονομάζεται αναδιάταξη εικόνων, καθώς όσες επιβεβαιώθουν γεωμετρικά καταλαμβάνουν τις πρώτες θέσεις στις απαντήσεις της μηχανής αναζήτησης. Αυτή είναι και η περίπτωση στην δουλεία των Philbin *et al.* [66], όπου αναγνωρίζεται και η ανάγκη για να ενσωματωθεί γεωμετρική πληροφορία στην δομή δεικτοδότησης. Σε πρόσφατες δουλειές αυτό έχει επιτευχθεί μόνο με την μορφή χαλαρών γεωμετρικών περιορισμών όπως στην δουλεία των Jegou *et al.* [34], ή τοπικής γεωμετρίας, των Chum *et al.* [17]. Από την άλλη, η δεικτοδότηση καθολικής γεωμετρίας είναι τουλάχιστον όσο παλιά όσο και ο γεωμετρικός κατακερματισμός (geometric hashing) των Lamdan και Wolfson [47]. Παρόλα αυτά δεν έχει αναφερθεί μέθοδος η οποία να ενσωματώνει στην διαδικασία δεικτοδότησης τόσο την εμφάνιση των τοπικών χαρακτηριστικών όσο και την γεωμετρική πληροφορία.

Η χρήση γεωμετρικής πληροφορίας έχει επίσης κριθεί απαραίτητη σε μεθόδους επέκτασης ερωτήματος (query expansion) [18, 16]. Αυτές είναι ένα είδος αυτόματης ανάδρασης όπου μέσω γεωμετρικού ταιριάσματος αναγνωρίζονται οι εικόνες οι οποίες απεικονίζουν ίδια αντικείμενα με την εικόνα ερωτήματος. Αυτές χρησιμοποιούνται για να δημιουργηθεί ένα νέο εμπλουτισμένο ερώτημα. Με κόστος τον υπολογιστικό χρόνο, τέτοιες μέθοδοι και με χρήση γεωμετρίας έχουν επιτύχει τις υψηλότερες αποδόσεις στο πρόβλημα της αναζήτησης εικόνων.

## 1.1 Γενικό υπόβαθρο

Κάθε εικόνα αναπαρίσταται από ένα σύνολο τοπικών χαρακτηριστικών ή τοπικών περιοχών  $P$ , οι οποίες προέρχονται από αντίστοιχους ανιχνευτές περιοχών [8][53][3][84]. Η κάθε περιοχή  $p \in P$  συνοδεύεται από συντεταγμένες του κέντρου της  $(x, y)$ , κλίμακα ( $\sigma$ ) και περιστροφή ( $\theta$ ) εάν πρόκειται για περιοχές αναλλοίωτες ως προς την κλιμάκωση και την περιστροφή, ή έναν πίνακα

$$F(p) = \begin{bmatrix} a & b & x \\ c & d & y \\ 0 & 0 & 1 \end{bmatrix}, \quad (1.1)$$

εάν πρόκειται για αφινικά αναλλοίωτες περιοχές. Παραδείγματα τέτοιου συνόλου τοπικών χαρακτηριστικών φαίνονται στο Σχήμα 1.1. Κάθε χαρακτηριστικό  $p$  συνοδεύεται επίσης από έναν τοπικό περιγραφέα  $d = d(p) \in \mathbb{R}^n$  ο οποίος περιγράφει την περιοχή της εικόνας από μία τετράγωνη ή ορθογώνια περιοχή της εικόνας γύρω από το εντοπισμένο χαρακτηριστικό. Τυπικό παράδειγμα είναι μία χωρική περιγραφή της κατανομής των ακμών.

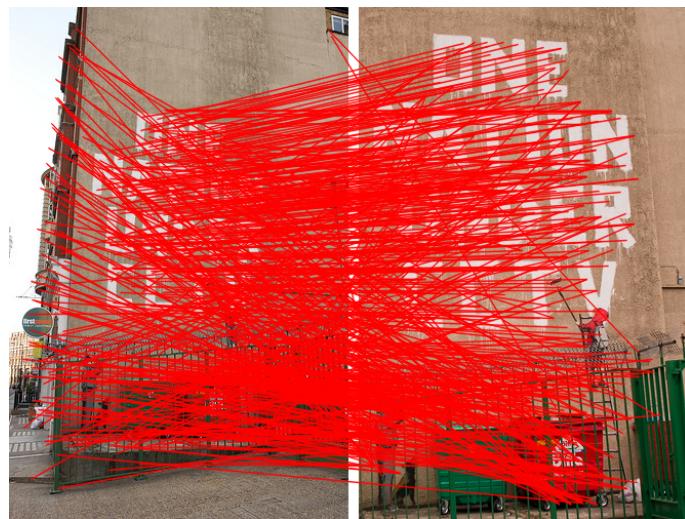


**Σχήμα 1.1:** Αριστερά: Σύνολο τοπικών χαρακτηριστικών SURF, αναλλοίωτων ως προς κλιμάκωση και περιστροφή. Δεξιά: Σύνολο τοπικών χαρακτηριστικών Hessian Affine, αφινικά αναλλοίωτων.

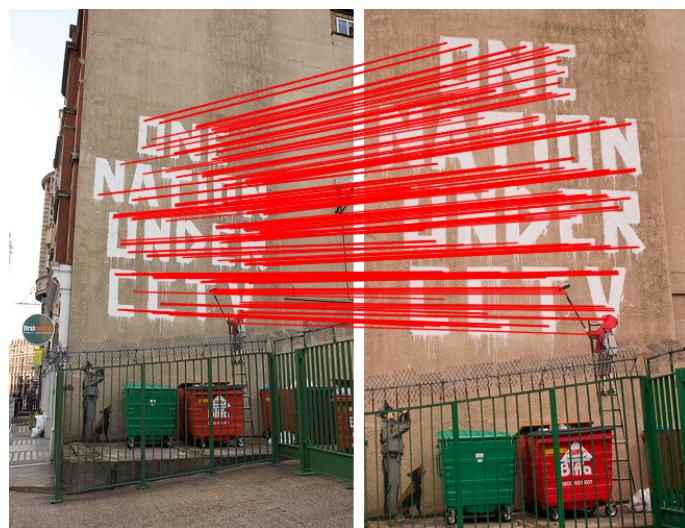
Η εγγύτητα των τοπικών περιγραφέων στον χώρο των περιγραφέων δηλώνει και ομοιότητα ανάμεσα στις αντίστοιχες περιοχές των εικόνων. Με αυτόν τον τρόπο προκύπτουν και αντιστοιχίες ανάμεσα σε τοπικά χαρακτηριστικά. Στην ανάκτηση εικόνων λόγω της ανάγκης γρήγορων υπολογισμών, έχει επικρατήσει η χρήση οπτικού λεξικού (visual vocabulary)  $\mathcal{V} \subset \mathbb{R}^n$ , το οποίο προκύπτει μέσω συσταδοποίησης (clustering) σε ένα μεγάλο σύνολο εκμάθησης από διανύσματα τοπικών περιγραφέων. Το οπτικό λεξικό αποτελείται από  $|\mathcal{V}| = k$  οπτικές λέξεις (visual word) και το κάθε διάνυσμα περιγραφέα ανήκει σε μία από αυτές. Άν  $v(p)$  είναι η οπτική λέξη του πειγραφέα  $d(p)$ , τότε το  $p$  είναι σε αντιστοιχία με το χαρακτηριστικό  $q \in \mathcal{Q}$  μιας δεύτερης εικόνας μόνο εάν ανήκουν στην ίδια οπτική λέξη, δηλαδή  $v(p) = v(q)$ . Στο Σχήμα 1.2 φαίνονται οι αντιστοιχίες τοπικών χαρακτηριστικών με την χρήση λεξικού. Στην συνέχεια οι συμβολισμοί για τα τοπικά χαρακτηριστικά θα ορίζονται εκ νέου ανάλογα με τις ανάγκες κάθε φορά. Σκοπός της εφαρμογής γεωμετρικής πληροφορίας είναι να φιλτραριστούν όλες οι αντιστοιχίες οι οποίες έχουν προέλθει από λανθασμένο ταίριασμα των τοπικών περιγραφέων και να διατηρηθούν όλες εκείνες οι οποίες είναι σύμφωνες γεωμετρικά μεταξύ τους, όπως φαίνεται και στο Σχήμα 1.3 μετά την εφαρμογή γεωμετρικού ταιριάσματος μέσω του αλγορίθμου RANSAC [26].

## 1.2 Συνεισφορά

Στην πρώτη προτεινόμενη μέθοδο (Κεφάλαιο 2) της παρούσας διατριβής [6][81] γίνεται μία προσπάθεια για ενσωμάτωση γεωμετρικής πληροφορίας στην δομή δεικτοδότησης. Έμπνευση μας είναι το γεωμετρικό ταίριασμα [66], το οποίο είναι παραλλαγή του αλγο-



Σχήμα 1.2: Αντιστοιχίες χαρακτηριστικών με την χρήση οπτικού λεξικού.



Σχήμα 1.3: Αντιστοιχίες οι οποίες ανιχνεύθηκαν ως *inliers* έπειτα από γεωμετρικό ταίριασμα.

ρίθμου RANSAC [26]. Οι παράμετροι του σχήματος των τοπικών χαρακτηριστικών χρησιμοποιούνται για να παραχθεί η κάθε μία από τις γεωμετρικές υποθέσεις με βάση μονές αντιστοιχίες χαρακτηριστικών. Παραδοσιακά στον αλγόριθμο RANSAC χρησιμοποιούνται τριάδες ή τετράδες αντιστοιχιών για την δημιουργία υπόθεσης ανάλογα με το χρησιμοποιούμενο κάθε φορά μετασχηματισμό. Όταν τα τοπικά χαρακτηριστικά συνοδεύονται από πληροφορία θέσης ( $x, y$ ) τότε χρειάζονται 2, 3 και 4 αντιστοιχίες για εκτίμηση μετασχηματισμού ομοιότητας, αφινικού μετασχηματισμού και ομογραφίας αντίστοιχα. Ενώ όταν συνοδεύονται επιπλέον από πληροφορία κλίμακας ( $\sigma$ ) και περιστροφής ( $\theta$ ) μία αντιστοιχία είναι αρκετή για εκτίμηση μετασχηματισμού ομοιότητας.

Εμείς πάμε ένα βήμα παραπέρα, κωδικοποιούμε για κάθε σημείο την κανονικοποιημένη θέση και την εμφάνιση όλων των υπόλοιπων σε ένα αραιό ιστόγραμμα το οποίο

ονομάζουμε χάρτη χαρακτηριστικών (feature map). Έπειτα χρησιμοποιούμε τις ανεξάρτητες αντιμεταθέσεις (min-wise independent permutations) [10] είτε μία μέθοδο χωρίς επίβλεψη για επιλογή χαρακτηριστικών και δημιουργούμε ένα μέτρο ομοιότητας για σύνολα χαρτών. Με την χρήση ενός ανάστροφου αρχείου, η διαδικασία αναζήτησης γίνεται υπογραμμική ως προς τον συνολικό αριθμό εικόνων. Επιπλέον, οι επιστρεφόμενες εικόνες σχετίζονται και με μία πρόχειρη εκτίμηση του σχετικού γεωμετρικού μετασχηματισμού. Στον ίδιο χρόνο εκτέλεσης, αυξάνουμε τις εικόνες στις οποίες εφαρμόζουμε γεωμετρική επιβεβαίωση κατά μία τάξη μεγέθους.

Εκμεταλλευόμενοι τις παραμέτρους σχήματος των τοπικών χαρακτηριστικών για να εξάγουμε σχετικούς μετασχηματισμούς, εκτός από το ότι είναι δυνατό να δημιουργήσουμε γεωμετρικές υποθέσεις για τον αλγόριθμο RANSAC μέσω μονών αντιστοιχιών [66], μπορούμε επίσης να δούμε τις αντιστοιχίες ως σημεία (*Hough votes*) σε κάποιον χώρο μετασχηματισμών [53]. Στην πρώτη περίπτωση κανείς πρέπει να μετρήσει inliers για κάθε υπόθεση, και αυτή είναι μια διαδικασία τετραγωνικής πολυπλοκότητας ως προς τον αριθμό των αντιστοιχιών. Στην δεύτερη περίπτωση η καταμέτρηση ομάδων όμοιων σημείων μπορεί μετά από κβαντισμό του χώρου των μετασχηματισμών να γίνει γραμμική. Παρόλα αυτά επιπλέον επιβεβαίωση με καταμέτρηση inliers φαίνεται να είναι αναπόφευκτη.

Εύκαμπτα χωρικά μοντέλα είναι πιο συνηθισμένα στην αναγνώριση αντικειμένων, τα οποία όμως συνήθως είτε δεν είναι αναλλοίωτα ως προς γεωμετρικούς μετασχηματισμούς, είτε χρησιμοποιούν περιορισμούς ανά ζευγάρια αντιστοιχιών για να εντοπίσουν inliers [50]. Η πολυπλοκότητα αυτών των μοντέλων είναι τουλάχιστον τετραγωνική ως προς τον πλήθος των αντιστοιχιών και ο πρακτικός χρόνος εκτέλεσης τους είναι απαγορευτικός εάν ο στόχος μιας είναι η ανακατάταξη εικόνων κι εφαρμογή σε εκατοντάδες εικόνες ανά δευτερόλεπτο.

Σαν δεύτερη συνεισφορά της παρούσας διατριβής (Κεφάλαιο 3), αναπτύσσουμε μια μέθοδο χαλαρού χωρικού ταιριάσματος, στην οποία όμοια με την διάσημη μέθοδο ταιριάσματος πυραμίδας [30], χρησιμοποιεί την κατανομή των αντιστοιχιών σε μία ιεραρχική διαμέριση του χώρου των μετασχηματισμών [78][4]. Χρησιμοποιώντας το σχήμα των τοπικών χαρακτηριστικών για να δημιουργήσουμε ψήφους, η μέθοδος είναι αναλλοίωτη ως προς τον μετασχηματισμό ομοιότητας, απαλλαγμένη από καταμέτρηση inliers και γραμμική ως προς τον αριθμό των αντιστοιχιών. Εφαρμόζει ταίριασμα ένα προς ένα το οποίο είναι εύκαμπτο, επιτρέποντας έτσι ταίριασμα πολλαπλών αντικειμένων ή επιφανειών.

Σε αναζήτηση μεγαλύτερης κλίμακας οι απατήσεις μνήμης φαίνεται να μετατρέπονται σε περιορισμό. Η επιλογή χαρακτηριστικών είναι πιθανόν η μόνη πρακτική εναλλακτική λύση. Μία ήδη διαδεδομένη ιδέα είναι κανείς να επιλέξει χαρακτηριστικά με βάση το ταίριασμα μεταξύ διαφορετικών εικόνων που απεικονίζουν το ίδιο αντικείμενο [82], παραρομοίως με την μέθοδο την οποία προτείνουμε κι εμείς για επιλογή χαρακτηριστικών σε

## Συνεισφορά

συνδυασμό με τους χάρτες χαρακτηριστικών. Αυτό εάν υποθέσουμε ότι υπάρχουν όντως πολλαπλές όψεις του κάθε αντικειμένου στην βάση. Στην πράξη, παρόλα αυτά, οι περισσότερες εικόνες είναι οι μοναδικές απεικονίσεις ενός αντικειμένου και δεν υπάρχει κάποια άλλη εικόνα στην βάση με την οποία να μπορούν να ταιριάξουν. Εμείς υποθέτουμε πως τοπικά χαρακτηριστικά τα οποία επαναλαμβάνονται μέσα στην ίδια εικόνα και συμμετέχουν σε επαναλαμβανόμενες και συμμετρικές δομές είναι πιθανό να είναι ικανά να ταιριάξουν και με άλλες όψεις του ίδιου αντικειμένου. Αναπτύσσουμε δύο μεθόδους για αυτό τον σκοπό (Κεφάλαιο 4) οι οποίες βασίζονται σε επιτυχημένες μεθόδους για γεωμετρικό ταίριασμα, την *fast spatial matching* (FSM) των Philbin *et al.* [66], και την δική μας *Hough pyramid matching* (HPM). Τις εφαρμόζουμε είτε ανάμεσα στην εικόνα και τον εαυτό της, είτε ανάμεσα στην εικόνα και το αντικατοπτρισμένο είδωλο της για να επιτύχουμε την επιλογή χαρακτηριστικών ανεξάρτητα από την κάθε εικόνα μόνη της. Τελικά καταφέρνουμε να μειώσουμε τις απαιτήσεις μνήμης και σε μεγάλη κλίμακα να επιτύχουμε την ίδια απόδοση με την περίπτωση που γίνεται χρήση όλων των διαθέσιμων περιοχών.

Στην τελευταία συνεισφορά της διατριβής (Κεφάλαιο 5) προτείνουμε μία νέα μέθοδο για επέκταση ερωτήματος (ένα είδος αυτόματης ανάδρασης) η οποία εμπλουτίζει και ραφινάρει την τοπική αναπαράσταση των τοπικών χαρακτηριστικών. Αυτό έρχεται σε αντίθεση με τις μέχρι τώρα μεθόδους για επέκταση ερωτήματος στις οποίες χρησιμοποιείται και εμπλουτίζεται η ολική αναπαράσταση της εικόνας, η οποία είναι το ιστόγραμμα εμφάνισης των οπτικών λέξεων. Η προτεινόμενη μέθοδος χρησιμοποιεί την ίδια αναπαράσταση και το ίδιο βασικό σύστημα αναζήτησης με την μέθοδο Hamming Embedding (HE) [37]. Χρησιμοποιούμε τέτοια κριτήρια για τον εντοπισμό των αξιόπιστων εικόνων ώστε να μην υπάρχει η ανάγκη για γεωμετρική επιβεβαίωση, με αποτέλεσμα να καταλήγουμε σε πολύ χαμηλούς χρόνους εκτέλεσης. Επίσης ένα σημαντικό συστατικό της μεθόδου μας είναι η συνάθροιση των τοπικών χαρακτηριστικών ανά οπτική λέξη την οποία πραγματοποιούμε κατά την δημιουργία της εμπλουτισμένης αναπαράστασης από την επέκταση ερωτήματος. Αυτό το στάδιο μειώνει πολύ την πολυπλοκότητα του νέου ερωτήματος αλλά ταυτόχρονα βελτιώνει και τα αποτελέσματα. Κατά τα όσα γνωρίζουμε είναι η πρώτη μέθοδος για επέκταση ερωτήματος κατά την οποία δεν είναι απαραίτητη η γεωμετρική επιβεβαίωση για να είναι αποτελεσματική. Παρόλα αυτά στα πειράματα μας δείχνουμε ότι εάν επιπλέον συνδυάσουμε την μέθοδο μας με γεωμετρική επιβεβαίωση τα αποτελέσματα βελτιώνονται ακόμα περισσότερο.

# Κεφάλαιο 2

## Χάρτες χαρακτηριστικών

### 2.1 Εισαγωγή και σχετική βιβλιογραφία

Η έλευση των τοπικών χαρακτηριστικών με διακριτική ικανότητα, λόγω της εξαγωγής τοπικών περιγραφέων [53][8], έχει κάνει δυνατή την δεικτοδότηση της εμφάνισης των χαρακτηριστικών για μεγάλες συλλογές εικόνων και σε μικρούς χρόνους εκτέλεσης ενός ερωτήματος. Αντιθέτως, λιγότερα είναι τα επιτεύγματα στην δεικτοδότηση της γεωμετρικής πληροφορίας, καθώς συνήθως μία προτιμότερη λύση είναι η ανακατάταξη των εικόνων σε ένα δεύτερο στάδιο το οποίο ως πιο κοστοβόρο εφαρμόζεται μόνο στις πιο όμοιες εικόνες. Όμως η χρήση γεωμετρικής πληροφορίας φαίνεται απαραίτητη.

Η εκμετάλλευση του τοπικού σχήματος των ανιχνευθέντων περιοχών [8][53][3][84], π.χ. κλίμακα, περιστροφή, αφινικό σχήμα, μπορεί να οδηγήσει σε λύσεις οι οποίες βασίζονται σε μονές αντιστοιχίες μεταξύ χαρακτηριστικών για την εκτίμηση γεωμετρικών μετασχηματισμών ανάμεσα σε ζευγάρια εικόνων ή και μεταξύ χαρακτηριστικών της ίδιας εικόνας.

Το σημείο εκκίνησης για αυτή την δουλεία είναι το [66] όπου το γεωμετρικό ταίριασμα γίνεται με μία ειδική περίπτωση του αλγορίθμου RANSAC [26]. Το σχήμα των περιοχών σε αντιστοιχία χρησιμοποιείται για να δημιουργήσει υποθέσεις από μονές αντιστοιχίες σε αντίθεση με τις δυάδες, τριάδες ή τετράδες αντιστοιχιών που χρειάζεται ο παραδοσιακός αλγόριθμος, ανάλογα με τον μοντέλο μετασχηματισμού σε κάθε περίπτωση. Αυτή η ιδέα προέρχεται και από το [53] κι έχει διερευνηθεί περισσότερο, π.χ. από τους Köser *et al.* [46]. Εμείς θα κανονικοποιήσουμε εκ των προτέρων τις θέσεις των χαρακτηριστικών ως προς το τοπικό σχήμα όλων των υπολοίπων. Χρησιμοποιούμε αυτή την πληροφορία σε μία κατάλληλη δομή δεικτοδότησης ώστε να είναι διαθέσιμη για γρήγορο ταίριασμα κατά την στιγμή του ερωτήματος. Ονομάζουμε την αναπαράσταση που δημιουργούμε χάρτη χαρακτηριστικών και μπορεί κανείς να την δει ως έναν τοπικό περιγραφέα ο οποίος περιγράφει ολικά το σύνολο χαρακτηριστικών σε ένα τοπικό σύστημα αναφοράς. Το ταί-

ριασμα μεταξύ τέτοιων αναπαραστάσεων τελικά ανάγεται σε απλό εσωτερικό γινόμενο. Χρησιμοποιούμε την μέθοδο των τυχαίων αντιμεταθέσεων για να καταλήξουμε σε μία πιο συμπαγή αναπαράσταση αλλά και προτείνουμε έναν τρόπο για επιλογή χαρακτηριστικών χωρίς επίβλεψη ο οποίος θα μας επιτρέψει να εφαρμόσουμε την μέθοδο των χαρτών σε πολύ μεγαλύτερη κλίμακα.

**Γεωμετρικό ταίριασμα** Η κανονικοποίηση ενός συνόλου  $2\Delta$  σημείων σε ένα σύστημα αναφοράς το οποίο ορίζεται από ένα σύνολο σημείων αναφοράς είναι πολύ συνηθισμένη. Παραδείγματα είναι οι συντεταγμένες *Bookstein* και *Kendall* [24], όπου τα πρώτα δύο σημεία επιλέγονται τυχαία σαν αναφορά, με αποτέλεσμα να αφαιρούνται μετασχηματισμοί ομοιότητας. Για την αντιμετώπιση των outliers, στο *geometric hashing* [47] συμβαίνει το ίδιο, για κάθε όμως πιθανό συνδυασμό σημείων αναφοράς από το αρχικό σύνολο. Μεγάλα σύνολα σημείων αναφοράς χρησιμοποιούνται για να αφαιρεθούν πιο πολύπλοκοι μετασχηματισμοί, π.χ. συνδυασμοί τριών σημείων για αφινικούς μετασχηματισμούς. Οι θέσεις κβαντίζονται, όπως και στην δική μας μέθοδο. Η πολυπλοκότητα όμως είναι τέτοια, με αποτέλεσμα να εφαρμόζεται μόνο σε έναν μικρο αριθμό προτύπων στην αναγνώριση αντικειμένων.

Μία και μοναδική περιοχή είναι αρκετή στην μέθοδο μας για να ορίσει ένα σύστημα αναφοράς, με αποτέλεσμα να μπορούμε να αποσυνθέσουμε όλες τις εικόνες της βάσεις, το ίδιο και την εικόνα ερώτημα κατά την διάρκεια του ερωτήματος. Οι Chum και Matas [14] χρησιμοποιούν μία περιοχή για να ορίσουν το σύστημα αναφοράς, αλλά για κάθε άλλη περιοχή κωδικοποιούν το σχήμα και όχι την εμφάνιση του. Ισχυριζόμαστε ότι είναι αρκετό να χρησιμοποιήσει κανείς το τοπικό σχήμα μόνο όταν αναμορφώνουμε – ενώ ενοποιούμε την εμφάνιση στο κοινό λεξικό, εξάγοντας έναν χάρτη χαρακτηριστικών με μεγάλη διακριτική ικανότητα. Ο χάρτης χαρακτηριστικών, αν τον δει κανείς σαν τοπικό περιγραφέα, είναι μία έννοια πολύ κοντά στο *shape context* [9], στο γεγονός ότι οι θέσεις όλων των γειτονικών σημείων κβαντίζονται σε πολικές συντεταγμένες. Παρόλα αυτά η ιδιότητα του γεωμετρικά αναλλοίωτου βασίζεται σε ολικές μετρήσεις.

Οι Philbin *et al.* [66] προσεγγίζουν το αλγόριθμο RANSAC βασισμένοι σε μονές αντιστοιχίες. Η γεωμετρική επαλήθευση που προτείνουν εφαρμόζεται για την αναδιάταξη 1000 εικόνων. Αντίθετα, εμείς υπολογίζουμε εκ των προτέρων τους αναμορφωμένους χάρτες χαρακτηριστικών και τα αντίστοιχα σκαριφήματα και τα ενσωματώνουμε στην δομή δεικτοδότησης, έτσι ώστε οι επιστρεφόμενες εικόνες ως όμοιες να είναι ήδη γεωμετρικά επιβεβαιωμένες. Εστιάζοντας στην αποδοτικότητα ως προς την μνήμη, οι Perdoch *et al.* [61] κβαντίζουν το τοπικό σχήμα των περιοχών χωρίς απώλειες στην απόδοση. Οι Jegou *et al.* [35] επίσης εστιάζουν στις ανάγκες μνήμης προσφέροντας πολύ μεγάλη συμπίεση, αλλά η απόδοση θυσιάζεται.

Οι Jegou *et al.* [34] κάνουν μία ακόμα προσπάθεια να ενσωματώσουν γεωμετρικά χα-

ρακτηριστικά στην δομή δεικτοδότησης μέσων του *weak geometric consistency* (WGC). Επεκτείνουν το μοντέλο *bag-of-words* (BoW), καταγράφοντας ξεχωριστά τις διαφορές κλίμακας και περιστροφής ανάμεσα στις περιοχές. Έτσι, το τοπικό σχήμα παίζει ρόλο στην ομοιότητα μεταξύ εικόνων, αλλά δεν επεκτείνεται για να αντιμετωπίσει αφινικούς μετασχηματισμούς. Επίσης η θέση των περιοχών έχει αμεληθεύει. Οι μέθοδοι BoW και WGC είναι οι μέθοδοι με τις οποίες συγκρίνουμε στα πειράματα μας στην ενότητα 2.5.

Υπάρχει μία πληθώρα μεθόδων οι οποίες χρησιμοποιούν hashing στην αναζήτηση εικόνων, όπως το *pyramid match hashing* [29] και το *random histograms* [23], χωρίς όμως να χρησιμοποιούν κάποιας μορφής αναπαράστασης για την γεωμετρία. Εξαίρεση αποτελεί η δουλειά των Chum *et al.* [17], οι οποίοι αναπαριστούν την γεωμετρία μέσω του *geometric min-hashing* (GmH). Το GmH χρησιμοποιεί την αμοιβαία θέση την περιοχών μόνο για να επαληθεύσει/επιβεβαιώσει τις προσκρούσεις το οποίο φυσικά δεν είναι υπο-γραμμικό. Η γεωμετρία αξιοποιείται μόνο σε τοπικές γειτονιές, ενώ στην δική μας μέθοδο είναι ολική και κωδικοποιείται απευθείας στα στοιχεία του σκαριφήματος. Οι κεντρικές περιοχές διαλέγονται τυχαία στο [17] και το GmH εφαρμόζεται σε συσταδοποίηση εικόνων και για εύρεση μικρών αντικειμένων. Σε αντίθεση, εμείς εστιάζουμε σε γενική αναζήτηση εικόνων και διαλέγουμε σαν κέντρα, περιοχές οι οποίες τείνουν να ευθυγραμμίζονται καλύτερα, όπως περιγράφεται στην ενότητα 2.4.4.

**Επιλογή χαρακτηριστικών** Η επιλογή τοπικών χαρακτηριστικών έχει γίνει ένας δημοφιλής τρόπος για την μείωση των αναγκών μνήμης στην αναζήτηση εικόνων. Οι Schindler *et al.* [71] και οι Li και Kosecka [51] πρότειναν τις πρώτες μεθόδους οι οποίες είχαν μάλιστα εφαρμογή την αναγνώριση τοποθεσίας. Δεδομένης μίας βάσης εικόνων με γεωγραφικό στίγμα, οι Schindler *et al.* χρησιμοποιούν την έννοια του κέρδους πληροφορίας για να επιλέξουν τα χαρακτηριστικά με την περισσότερη πληροφορία, π.χ. χαρακτηριστικά τα οποία εμφανίζονται σε όλες τις εικόνες μία συγκεκριμένης τοποθεσίας, αλλά σπάνια ή ποτέ αλλού. Ομοίως, οι Li και Kosecka αποκτούν μία πιθανότητα του πληροφοριακού περιεχομένου για κάθε χαρακτηριστικό με γνώμονα την αναγνώριση τοποθεσίας.

Οι Knopp *et al.* [45] επίσης χρησιμοποιούν συλλογή φωτογραφιών με γεωγραφικό στίγμα και εκμεταλλεύονται το γεγονός πως φωτογραφίες που βρίσκονται σε μεγάλη απόσταση μεταξύ τους δεν θα πρέπει να ταιριάζουν. Οι Gammeter *et al.* [27] αρχίζουν από ομάδες εικόνων οι οποίες δημιουργήθηκαν με βάση την γεωγραφική εγγύτητα και την οπτική ομοιότητα. Έπειτα χρησιμοποιούν τα στατιστικά των χαρακτηριστικών που ταιριάζουν για να εκτιμήσουν ένα ορθογώνιο πλαίσιο γύρω από κάθε αντικείμενο. Παρά το ότι κρατάνε και δεικτοδοτούν μόνο όσα χαρακτηριστικά βρίσκονται μέσα σε αυτό το πλαίσιο, τελικά το 66% αυτών διατηρούνται.

Όλες οι παραπάνω μέθοδοι χρειάζονται ενός είδους επίβλεψη η οποία προέρχεται από το γεωγραφικό στίγμα. Το μοντέλο το οποίο θα προτείνουμε εμείς για επιλογή χαρακτηρι-

στικών είναι χωρίς επίβλεψη και η έμπνευση μας προέρχεται από την δουλειά των Turcot και Lowe [82]. Εκείνοι χρησιμοποιούν κάθε εικόνα της βάσης ως ερώτημα και επιλέγουν όσα τοπικά χαρακτηριστικά θα εντοπισθούν ως inliers ανάμεσα στην εικόνα ερωτήματος και σε κάποια εικόνα της βάσης. Οι Tolias *et al.* [80] (Κεφάλαιο 4) πρότειναν μία μέθοδο για επιλογή χαρακτηριστικών για τις εικόνες οι οποίες δεν ταιριάζουν με καμία άλλη στην βάση κι έτσι η μέθοδος των Turcot και Lowe δεν μπορεί να εφαρμοσθεί. Ταιριάζουν γεωμετρικά κάθε εικόνα με τον εαυτό της και το αντικατοπτρισμένο είδωλο της κι επιλέγουν με βάση τις εντοπισμένες συμμετρίες και επαναλαμβανόμενες δομές.

## 2.2 Υπόβαθρο

Ξεκινώντας θα εξετάσουμε μερικά απλά μοντέλα ταιριάσματος συνόλου χαρακτηριστικών, τα οποία βασίζονται στην γεωμετρία, την εμφάνιση, ή και στα δύο. Παρατηρούμε πως αυτά τα μοντέλα παρέχουν λύσεις για ευθυγράμμιση, αντιστοίχιση, και για εντοπισμό outliers, και καταλήγουμε σε ένα μοντέλο το οποίο θα προσπαθήσουμε να απλοποιήσουμε στην συνέχεια.

Θεωρούμε πως μία εικόνα αναπαρίσταται από ένα σύνολο χαρακτηριστικών. Για κάθε χαρακτηριστικό  $x$  στην εικόνα  $X$ , συμβολίζουμε με  $p(x) \in \mathbb{R}^2$  την θέση του στην εικόνα, και με  $d(x)$  τον τοπικό περιγραφέα που του αντιστοιχεί και κωδικοποιεί την εμφάνιση του. Στα ακόλουθα μοντέλα τα τοπικά χαρακτηριστικά μπορεί να έχουν θέση, περιγραφέα ή και τα δύο.

### 2.2.1 Ταίριασμα σχήματος

Έστω  $X$ , δύο εικόνες. Τα χαρακτηριστικά θεωρούνται χωρίς διακριτική ικανότητα, δηλαδή δεν συνοδεύονται από κάποιο τοπικό περιγραφέα, οπότε μόνο οι θέσεις τους είναι γνωστές. Υποθέτουμε προς το παρόν ότι  $|X| = |Y|$  και ότι υπάρχει μία γνωστή 1-1 αντιστοιχία  $\pi : X \rightarrow Y$ . Στην στατιστική θεωρία σχήματος (Dryden και Mardia [24]), ένα από τα πιο γνωστά προβλήματα είναι η εκτίμηση του βέλτιστου γεωμετρικού μετασχηματισμού ο οποίος ευθυγραμμίζει τα δύο σύνολα

$$S_T(X, Y; r) = \max_{B, t} \sum_{x \in X} r(Bp(x) + t, p(\pi(x))), \quad (2.1)$$

όπου  $r : \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow [0, 1]$  είναι ένα οποιοδήποτε μέτρο χωρικής ομοιότητας (εγγύτητας), και  $B \in \mathbb{R}^{2 \times 2}$ ,  $t \in \mathbb{R}^2$  είναι ένας μετασχηματισμός, ο οποίος εδώ θα θεωρήσουμε ότι είναι αφινικός και η μετατόπιση, αντίστοιχα. Για να εφαρμοσθεί ο μετασχηματισμός, το κάθε σημείο στο  $\mathbb{R}^2$  θεωρείται ως ένα  $2 \times 1$  διάνυσμα. Το πρόβλημα σε αυτή την μορφή δεν εμφανίζεται στην περίπτωση μας, λόγω των άγνωστων αντιστοιχιών ανάμεσα στα χαρακτηριστικά καθώς και λόγω παρουσίας outliers.

## 2.2.2 Αντιστοιχίες περιοχών

Τώρα, φεύγοντας από την περίπτωση της γνωστής 1-1 αντιστοιχίας υποθέτουμε ότι τα χαρακτηριστικά έχουν διακριτική ικανότητα, η οποία προέρχεται από το πεπερασμένο σύνολο περιγραφέων. Επίσης ξεχνάμε την υπόθεση σύμφωνα με την οποία  $|X| = |Y|$ . Αγνοώντας τις θέσεις προσωρινά, το ακόλουθο πρόβλημα αντιστοιχίσης (*assignment*) μπορεί να αντιμετωπίσει άγνωστες αντιστοιχίες και μερικώς τους outliers:

$$S_A(X, Y; s) = \max_{\{a\}} \sum_{x \in X} \sum_{y \in Y} a_{x,y} s(x, y) \quad (2.2)$$

$$\text{s.t.} \quad \sum_{x \in X} a_{x,y} \leq 1, \quad \forall y \in Y \quad (2.3)$$

$$\sum_{y \in Y} a_{x,y} \leq 1, \quad \forall x \in X \quad (2.4)$$

$$a_{x,y} \in \{0, 1\}, \quad \forall x \in X, y \in Y \quad (2.5)$$

όπου  $s : X \times Y \rightarrow [0, 1]$  είναι επίσης ένα οποιοδήποτε μέτρο ομοιότητας στον χώρο των περιγραφέων, και το  $a$  μπορεί να θεωρηθεί είτε σαν  $|X| \times |Y|$  πίνακας με μονάδες και μηδενικά ή μία αντιστοιχία  $a : X \times Y \rightarrow \{0, 1\}$ . Με το  $a$ , κάθε σημείο  $x \in X$  μπορεί να αντιστοιχηθεί με μέχρι ένα σημείο  $y \in Y$ , και αντίστροφα. Παρά την απώλεια της γεωμετρίας, αυτό είναι ένα πολύ σημαντικό πρόβλημα επειδή μπορεί να δουλέψει πολύ καλά στην πράξη εάν τα χαρακτηριστικά συνοδεύονται από περιγραφείς με υψηλή διακριτική ικανότητα.

## 2.2.3 Bag-of-words

Επιπλέον, ορίζουμε ένα οπτικό λεξικό  $\mathcal{V}$  ως ένα πεπερασμένο υποσύνολο του χώρου των τοπικών περιγραφέων με  $|\mathcal{V}| = k_v$  στοιχεία ή οπτικές λέξεις, προερχόμενες για παράδειγμα από κβαντοποίηση σε ένα σύνολο εκπαίδευσης. Έστω  $v(x)$  η κβαντισμένη έκδοση του περιγραφέα  $d(x)$  ενός χαρακτηριστικού  $x \in X$ . Μπορεί έτσι να κατασκευασθεί η αναπαράσταση του μοντέλου *bag-of-words*, δηλαδή ένα ιστόγραμμα του  $X$  πάνω στο  $\mathcal{V}$  θεωρώντας πως  $H_b(X) = \{x \in X : v(x) = b\}$  είναι τα χαρακτηριστικά του  $X$  τα οποία προβάλλονται στην οπτική λέξη (κυτίο)  $b \in \mathcal{V}$ , και  $h_b(X) = |H_b(X)|/|X|$  η συχνότητα τους. Το ιστόγραμμα, το οποίο συμβολίζεται ως  $h_{\mathcal{V}}(X)$ , είναι ένα διάνυσμα στο  $\mathbb{R}^{|\mathcal{V}|}$ , και μπορεί να αναπαρασταθεί ως

$$h_{\mathcal{V}}(X) = \sum_{b \in \mathcal{V}} h_b(X) \mathbf{e}_b = \frac{1}{|X|} \sum_{x \in X} \mathbf{e}_{v(x)}, \quad (2.6)$$

όπου  $\{\mathbf{e}_b \in \mathbb{R}^{|\mathcal{V}|} : b \in \mathcal{V}\}$  είναι η βάση του  $\mathbb{R}^{|\mathcal{V}|}$ . Είναι έπειτα φυσικό να ορισθεί το διακριτό μέτρο ομοιότητας

$$s_{\mathcal{V}}(x, y) = \delta_{v(x), v(y)} = \begin{cases} 1, & \text{if } v(x) = v(y) \\ 0, & \text{otherwise,} \end{cases} \quad (2.7)$$

έτσι ώστε δύο χαρακτηριστικά  $x, y$  σε δύο εικόνες  $X, Y$  να είναι όμοια εάν και μόνο εάν προβάλλονται στην ίδια οπτική λέξη. Είναι εύκολο να δούμε ότι

$$S_A(X, Y; s_{\mathcal{V}}) = \sum_{b \in \mathcal{V}} \min(h_b(X), h_b(Y)) \quad (2.8)$$

είναι η τομή ιστογραμμάτων των οπτικών λέξεων για τα σύνολα  $X$  και  $Y$ . Με ανάλογο τρόπο, μπορούμε να αντικαταστήσουμε το σχήμα 1-1 αντιστοιχίας των (2.2)-(2.5) με ένα σχήμα ένα προς πολλά

$$S_M(X, Y; s) = \sum_{x \in X} \sum_{y \in Y} s(x, y) \quad (2.9)$$

και να επιβεβαιώσουμε ότι η ομοιότητα της αναπαράστασης με οπτικές λέξεις είναι ισοδύναμη με ένα εσωτερικό γινόμενο (Jegou *et al.* [34]),

$$S_M(X, Y; s_{\mathcal{V}}) = \sum_{b \in \mathcal{V}} h_b(X)h_b(Y) = \langle h_{\mathcal{V}}(X), h_{\mathcal{V}}(Y) \rangle. \quad (2.10)$$

Όταν τα ιστογράμματα είναι κανονικοποιημένα, αυτό είναι η πολύ γνωστή ομοιότητα συνημιτόνου (*cosine similarity*). Σε συνδυασμό με ένα ανάστροφο αρχείο για να εκμεταλλευτούμε την αραιότητα των διανυσμάτων, είναι ένας πολύ απλός και γρήγορος τρόπος για ανάκτηση εικόνων.

#### 2.2.4 Προς τον αλγόριθμο RANSAC

Το ταίριασμα ένα με πολλά μπορεί να δώσει απρόσμενα αποτελέσματα. Είναι παρόλα αυτά πιο εύκολο να εκτιμηθεί, ιδιαίτερα όταν γίνεται χρήση λεξικού. Ας ορίσουμε ένα σύνολο αρχικών αντιστοιχιών (*tentative correspondences*), οι οποίες ορίζονται με βάση πλησιέστερους γείτονες, είτε ώς

$$\mathcal{X}(X, Y; s) = \{(x, y) \in X \times Y : s(x, y) > \delta_s\}, \quad (2.11)$$

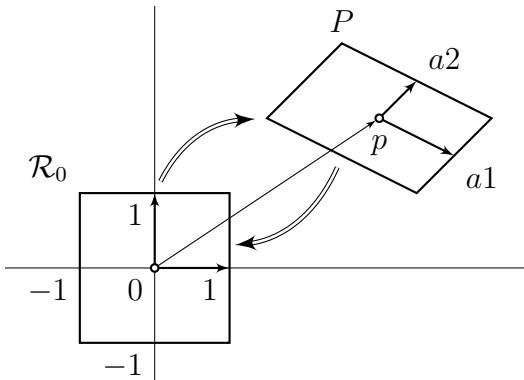
Δεδομένου ενός συγκεκριμένου συνόλου αντιστοιχιών  $\mathcal{C}$ , επιστρέφουμε στην σχέση (2.1) και μεγιστοποιούμε ως προς τον μετασχηματισμό  $(B, t)$  πάνω σε ένα πεπερασμένο σύνολο γεωμετρικών υποθέσεων (*hypotheses*)  $\mathcal{H}$ :

$$S_R(X, Y; \mathcal{C}, r) = \max_{(B, t) \in \mathcal{H}} \sum_{(x, y) \in \mathcal{C}} r(Bp(x) + t, p(y)). \quad (2.12)$$

Όταν οι υποθέσεις επιλέγονται τυχαία ακολουθώντας μια συγκεκριμένη στρατηγική και η χωρική ομοιότητα ορίζεται από έναν ομοιόμορφο πυρήνα

$$r_{\epsilon}(p, q) = \begin{cases} 1, & \text{if } \|p - q\|_2 < \epsilon \\ 0, & \text{αλλιώς,} \end{cases} \quad (2.13)$$

με  $p, q \in \mathbb{R}^2$  ο οποίος απλά καταμετρά τους inliers, τότε το παραπάνω αποτέλεσμα δεν είναι πολύ διαφορετικό από τον αλγόριθμο RANSAC. Με τις κατάλληλες αντιστοιχίες, μπορεί από κοινού να λύσει το πρόβλημα ως προς την ευθυγράμμιση και τους outliers.



Σχήμα 2.1: Αρχικές και αναμορφωμένες περιοχές εικόνας. Οι άξονες είναι χωρικοί.

## 2.3 Χάρτες χαρακτηριστικών

### 2.3.1 Τοπικές περιοχές

Θεωρούμε εδώ ότι κάθε τοπικό χαρακτηριστικό  $x$  είναι συσχετισμένο με μία τοπική περιοχή  $P(x)$  της εικόνας, το οποίο αναπαριστά το τοπικό σχήμα και την θέση. Συχνά οι έννοιες χαρακτηριστικό και περιοχή θα χρησιμοποιούνται εναλλάξ και ισάξια. Σύμφωνα με τους Rothganger *et al.* [70], η τοπική περιοχή είναι ένα παραλληλόγραμμο που αναπαρίσταται από έναν πίνακα

$$P(x) = \begin{bmatrix} a1 & a2 & p \\ 0 & 0 & 1 \end{bmatrix}, \quad (2.14)$$

όπου  $p(x)$  είναι τώρα το κέντρο και  $a1(x), a2(x) \in \mathbb{R}^2$  είναι τα διανύσματα από το  $p(x)$  στα μέσα των δύο πλευρών, όπως φαίνεται στο Σχήμα 2.1. Η αναμορφωμένη περιοχή  $\mathcal{R}_0$  αναπαρίσταται από τον μοναδιαίο πίνακα  $I_3$  και μετασχηματίζεται στην τοπική περιοχή μέσω του  $P$ , ενώ η περιοχή αναμορφώνεται στο  $\mathcal{R}_0$  μέσω του  $P^{-1}$ . Έτσι το  $P$  συμβολίζει και την περιοχή αλλά και έναν αφινικό μετασχηματισμό. Ο παραπάνω συμβολισμός είναι ισοδύναμος με τα *local affine frames* [14].

### 2.3.2 Υπόθεση μονών αντιστοιχιών

Δεδομένης μία αντιστοιχίας ανάμεσα σε τοπικά χαρακτηριστικά-περιοχές  $P(x) \leftrightarrow P(y)$ , ανάμεσα στα χαρακτηριστικά  $x, y$  των εικόνων  $X, Y$ , ο μετασχηματισμός από την μία περιοχή στην άλλη είναι  $P(y)P(x)^{-1}$ . Εάν οι δύο εικόνες σχετίζονται με μετασχηματισμό ομογραφίας, κι εφόσον από την εξαγωγή τους οι περιοχές είναι αφινικά αναλλοίωτες, οι Kōser *et al.* [46] διαχέουν τον μετασχηματισμό σε όλη την εικόνα και εξετάζουν την ευθυγράμμιση μεταξύ δύο εικόνων μέσω μονών αντιστοιχιών μεταξύ αφινικών περιοχών.

Οι Philbin *et al.* [66] επιπλέον παρατηρούν ότι η κάθε αντιστοιχία προσφέρει μία υπόθεση μετασχηματισμού. Οι υποθέσεις είναι έτσι  $O(n)$  και μπορούν να απαριθμηθούν όλες:

$$\mathcal{H}(\mathcal{C}) = \{P(y)P(x)^{-1} : (x, y) \in \mathcal{C}\}. \quad (2.15)$$

Για την ακρίβεια, λέμε πως η υπόθεση  $P(y)P(x)^{-1} \in \mathcal{H}(\mathcal{C})$  παράγεται από την αντιστοιχία  $(x, y) \in \mathcal{C}$ . Το ταίριασμα δύο εικόνων τελικά καταλήγει στην εύρεση της υπόθεσης με την μεγαλύτερη υποστήριξη (μεγαλύτερο αριθμό από inliers),

$$S_H(X, Y; \mathcal{C}, r) = \max_{A \in \mathcal{H}(\mathcal{C})} \sum_{(x, y) \in \mathcal{C}} r(A\mathbf{p}(x), \mathbf{p}(y)). \quad (2.16)$$

Εδώ το  $\mathbf{p}(x) = [p(x)^T 1]^T$  συμβολίζει την θέση του  $x$  σε ομογενείς συντεταγμένες, αυτό είναι ένα  $3 \times 1$  διάνυσμα στον προβολικό χώρο  $\mathbb{P}^2$ , ενώ ο  $A \in \mathbb{R}^{3 \times 3}$  είναι ένας αφινικός μετασχηματισμός ο οποίος περιέχει και μετατόπιση, αντίθετα με το  $B$  στην σχέση (2.12).

### 2.3.3 Αναμόρφωση του συνόλου χαρακτηριστικών

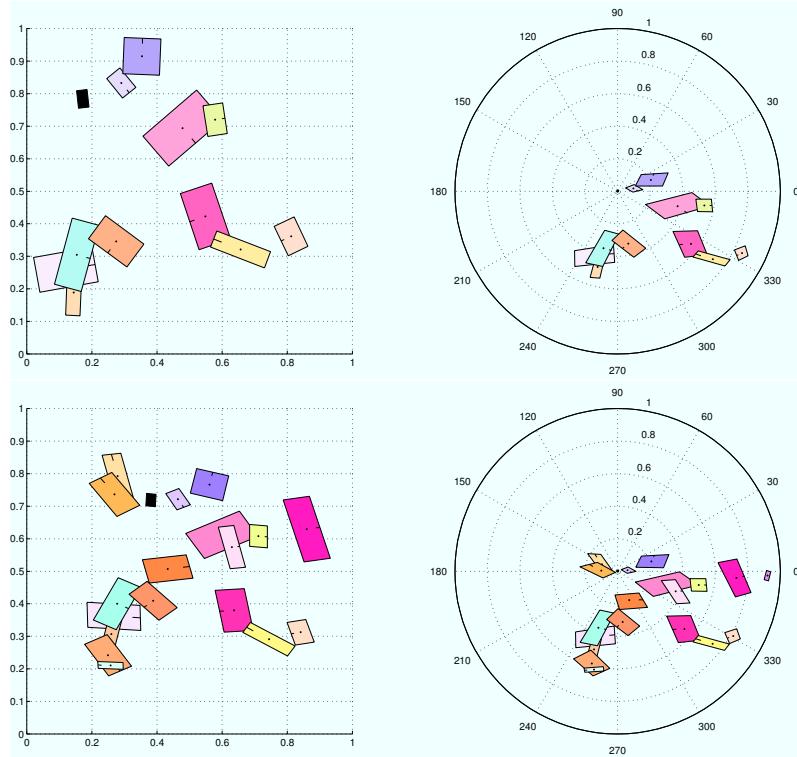
Αντί να κατασκευάσουμε τους μετασχηματισμούς  $P(y)P(x)^{-1}$  από όλες τις αντιστοιχίες και να εκτελέσουμε γεωμετρικό ταίριασμα κατά την στιγμή του ερωτήματος για την αναζήτηση εικόνων όπως οι Philbin *et al.* [66], μπορούμε να διαχύσουμε τον κάθε τοπικό μετασχηματισμό σε όλη την εικόνα και να αναμορφώσουμε τα υπόλοιπα χαρακτηριστικά εκ των προτέρων. Για την ακρίβεια, δεδομένου του χαρακτηριστικού  $\hat{x}$  στην εικόνα, το οποίο ονομάζουμε πηγή, αναμορφώνουμε όλα τα χαρακτηριστικά  $x \in X$  με βάση το  $\hat{x}$  ως ακολούθως.

Έστω ότι  $p^{(\hat{x})}(x) \in \mathbb{R}^2$  το Ευκλείδειο αντίστοιχο (το  $2 \times 1$  διάνυσμα με τα δύο πρώτα στοιχεία) του  $P^{-1}(\hat{x})\mathbf{p}(x)$ . Λέμε έτσι πως το χαρακτηριστικό  $x$  αναμορφώνεται ως προς το  $\hat{x}$  και είναι ένα νέο χαρακτηριστικό  $x^{(\hat{x})}$  με θέση  $p(x^{(\hat{x})}) = p^{(\hat{x})}(x)$  και περιγραφέα  $d(x^{(\hat{x})}) = d(x)$ . Επίσης λέμε πως το  $x^{(\hat{x})}$  είναι ένα αναμορφωμένο χαρακτηριστικό. Δεν συσχετίζουμε το  $x^{(\hat{x})}$  με κάποια περιοχή. Ο ρόλος του  $P(x)$  περιορίζεται στο να χρησιμοποιηθεί το  $x$  ως πηγή. Τελικά, ορίζουμε ως  $X^{(\hat{x})} = \{x^{(\hat{x})} : x \in X\}$  το τελικό σύνολο των χαρακτηριστικών του  $X$  αναμορφωμένα ως προς την περιοχή  $\hat{x}$ . Το Σχήμα 2.2 δείχνει ένα τυχαίο σύνολο περιοχών, κι μία μετασχηματισμένη και παραμορφωμένη έκδοση του, και τα αντίστοιχα αναμορφωμένα σύνολα ως προς 2 περιοχές πηγές που βρίσκονται σε αντιστοιχία — παρατηρούμε πως τα δύο τελευταία είναι κάπως ευθυγραμμισμένα.

Το ίδιο σύνολο των αντιστοιχιών  $\mathcal{C}$ , που αποκτήθηκε από ταίριασμα των περιγραφέων, χρησιμοποιείται και για ευθυγράμμιση αλλά και για καταμέτρηση των inliers:

$$\hat{S}_H(X, Y; \mathcal{C}, r) = \max_{(\hat{x}, \hat{y}) \in \mathcal{C}} \sum_{(x, y) \in \mathcal{C}} r(p(x^{(\hat{x})}), p(y^{(\hat{y})})) \quad (2.17)$$

$$= \max_{(\hat{x}, \hat{y}) \in \mathcal{C}} I(\mathcal{C}; \hat{x}, \hat{y}, r), \quad (2.18)$$



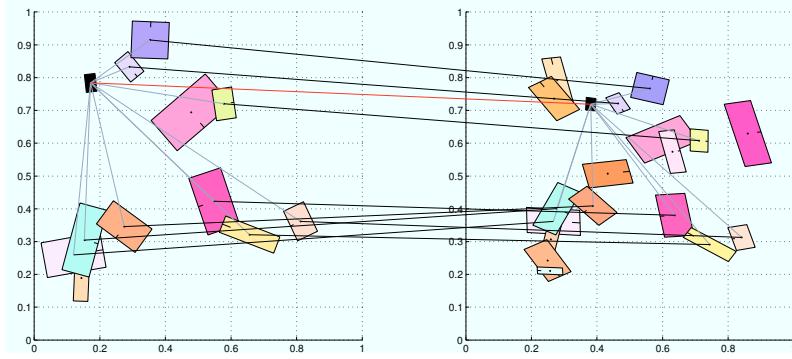
**Σχήμα 2.2:** Πάνω αριστερά: Ένα τυχαίο σύνολο περιοχών. Κάτω αριστερά: Το ίδιο σύνολο μετασχηματισμένο αφινικά, όπου οι θέσεις των περιοχών και το σχήμα τους έχουν διαστρεβλωθεί, και έχουν εισαχθεί νέες περιοχές. Δεξιά: Τα αντίστοιχα αναμορφωμένα σύνολα. Οι πηγές είναι οι δύο μαύρες περιοχές στα αριστερά. Το πολικό πλέγμα ορίζει τα χωρικά κυτία για τους χάρτες χαρακτηριστικών, με  $\tau = 0.95$ ,  $k_\rho = 5$  και  $k_\theta = 12$  (βλέπε ενότητα 2.4.4).

όπου  $I(\cdot; \hat{x}, \hat{y}, \cdot)$  είναι ο συνολικός αριθμός των inliers για την υπόθεση  $(\hat{x}, \hat{y})$ . Το μέτρο ομοιότητας αυτό δεν είναι το ίδιο με της σχέσης (2.16), αλλά στην πραγματικότητα η απόσταση τώρα καταμετρείται στο αναμορφωμένο σύστημα αναφοράς.

### 2.3.4 Κβαντισμός

Αντίθετα με την σχέση (2.16), η σχέση (2.17) υποθέτει ευθυγραμμισμένες περιοχές και θυμίζει μέτρο υπέρθεσης. Φαίνεται πως θα μπορούσαμε να χρησιμοποιήσουμε κάποια μορφή χωρικού κβαντισμού στα αναμορφωμένα συστήματα αναφοράς. Υιοθετούμε το οπτικό λεξικό κι επιπλέον ορίζουμε ένα χωρικό λεξικό  $\mathcal{U} \subseteq \mathbb{R}^2$  με  $|\mathcal{U}| = k_u$  κυτία (bins). Ο κβαντισμός μπορεί να είναι ομοιόμορφος. Παρόλα αυτά, η κωδικοποίηση όλων των θέσεων σε ένα πεπερασμένο σύνολο δεν είναι τετριμένη, βλέπε ενότητα 2.4.4.

Έστω ότι για κάθε χαρακτηριστικό  $x$ ,  $u(x)$  είναι η κβαντισμένη εκδοχή της θέσης  $p(x)$ , και  $w(x) = (v(x), u(x))$  το από κοινού κυτίο του χαρακτηριστικού. Επιπλέον, ορίζουμε το συνδυαστικό λεξικό  $\mathcal{W} = \mathcal{V} \times \mathcal{U}$  με  $|\mathcal{W}| = k_v k_u = k$  κυτία. Έπειτα, για οποιοδήποτε αναμορφωμένο σύνολο χαρακτηριστικών  $\hat{X}$  (με οποιαδήποτε πηγή), κατασκευάζουμε το από



**Σχήμα 2.3:** *Inliers* μεταξύ δύο συνόλων τοπικών χαρακτηριστικών. Ο κάθε ένας αντιστοιχεί σε ένα μη μηδενικό όρο του εσωτερικού γινομένου των αντίστοιχων χαρτών χαρακτηριστικών. Οι μαύρες γραμμές ενώνουν τους *inliers*. Οι κόκκινες γραμμές ενώνουν τις πηγές. Οι γκρίζες γραμμές ενώνουν τις πηγές με τους *inliers*.

κοινού οπτικο-χωρικό ιστόγραμμα πάνω στο  $\mathcal{W}$ , ορίζοντας  $H_b(\hat{X}) = \{x \in \hat{X} : w(x) = b\}$  είναι το σύνολο των αναμορφωμένων χαρακτηριστικών τα οποία προβάλλονται στο κυτίο  $b \in \mathcal{W}$ , και  $h_b(\hat{X}) = |H_b(\hat{X})|/|\hat{X}|$  το πλήθος τους. Το από κοινού ιστόγραμμα, το οποίο συμβολίζεται με  $h_{\mathcal{W}}(\hat{X})$ , είναι ένα διάνυσμα στο  $\mathbb{R}^k$ , το οποίο αναπαρίσταται ομοίως με το ιστόγραμμα της μεθόδου bag-of-words της σχέσης (2.6).

Ομοίως με την διακριτή οπτική ομοιότητα (2.7), ορίζουμε την διακριτή χωρική ομοιότητα  $r_{\mathcal{U}}(x, y) = \delta_{u(x), u(y)}$ , έτσι ώστε τα  $x, y$  ταιριάζουν μόνο αν έχουν αντιστοιχηθεί στο ίδιο κυτίο. Έπειτα, το μοντέλο ταιριάσματος (2.17) γίνεται

$$\hat{S}_H(X, Y; \mathcal{C}, r_{\mathcal{U}}) = \max_{(\hat{x}, \hat{y}) \in \mathcal{C}} \sum_{w \in \mathcal{W}} h_w(X^{(\hat{x})}) h_w(Y^{(\hat{y})}). \quad (2.19)$$

Η χρήση ενός οπτικού λεξικού σημαίνει πως οι αντιστοιχίες  $\mathcal{C}$  είναι μεταξύ χαρακτηριστικών  $x, y$  τα οποία ανήκουν στην ίδια οπτική λέξη,

$$\mathcal{C}(X, Y) = \{(x, y) \in X \times Y : v(x) = v(y)\}; \quad (2.20)$$

αυτό εξειδικεύει την αρχικές αντιστοιχίες (2.11) για την διακριτή οπτική ομοιότητα (2.7). Τώρα, έστω πως  $V(X) = \{b \in \mathcal{V} : H_b(X) \neq \emptyset\}$  είναι το σύνολο των οπτικών λέξεων οι οποίες εμφανίζονται στο σύνολο χαρακτηριστικών  $X$  και  $V(X, Y) = V(X) \cap V(Y)$  είναι οι κοινές οπτικές λέξεις των συνόλων  $X, Y$ . Έπειτα, εάν  $f^{(\hat{x})}(X) = h_{\mathcal{W}}(X^{(\hat{x})})$  είναι το ιστόγραμμα του  $X$  το οποίο όμως έχει αναμορφωθεί με βάση το  $\hat{x}$ , ο τελικός βαθμός ομοιότητας γίνεται

$$\hat{S}_F(X, Y) = \max_{b \in V(X, Y)} \max_{\substack{\hat{x} \in H_b(X) \\ \hat{y} \in H_b(Y)}} \langle f^{(\hat{x})}(X), f^{(\hat{y})}(Y) \rangle. \quad (2.21)$$

Έχουμε έτσι παράγει μία νέα αναπαράσταση εικόνας και μία αντίστοιχη διαδικασία ταιριάσματος όπως αυτή εκφράζεται από την σχέση (2.21).

### 2.3.5 Χάρτες χαρακτηριστικών

Ονομάζουμε  $f^{(\hat{x})}(X)$  τον χάρτη χαρακτηριστικών του  $X$  με πηγή  $\hat{x}$ . Το σύνολο  $F(X) = \{f^{(\hat{x})}(X) : \hat{x} \in X\}$ , είναι η συλλογή χαρτών χαρακτηριστικών του  $X$ . Οπτικά, ένας χάρτης χαρακτηριστικών μπορεί να κατανοηθεί ως η καταχώρηση των αναμορφωμένων χαρακτηριστικών σε χωρικά κυτία, όπως δεξιά στο Σχήμα 2.2. Η ακριβής διάταξη των χωρικών κυτίων συζητείται στην ενότητα 2.4.4. Υπάρχει ένας διαφορετικός χάρτης για κάθε πηγή: μπορούμε να φανταστούμε τον χάρτη της κάθε πηγής σαν ένα τοπικό περιγραφέα, ο οποίος κωδικοποιεί ολικά το σύνολο χαρακτηριστικών αναμορφωμένο σε ένα τοπικό σύστημα συντεταγμένων. Καλά ευθυγραμμισμένα σύνολα περιοχών είναι πιθανόν να έχουν χάρτες με μεγάλο βαθμό επικάλυψης.

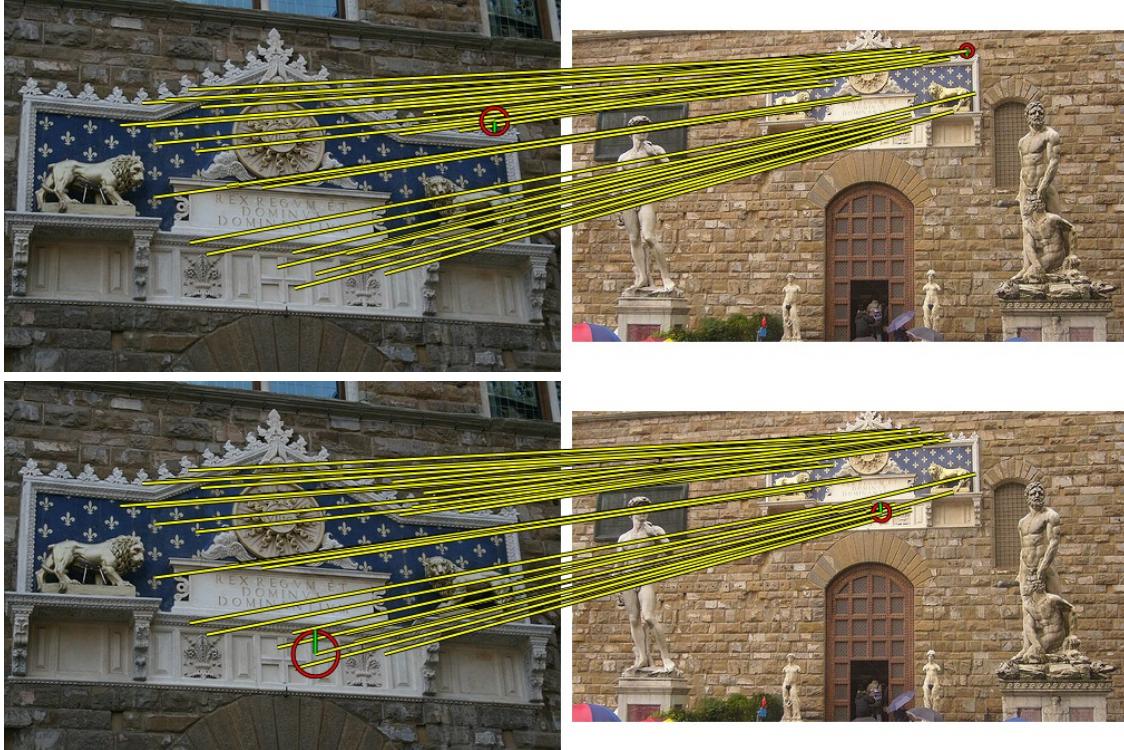
Επιστρέφοντας στο παράδειγμα του Σχήματος 2.2, οι inliers των δύο αναμορφωμένων συνόλων είναι εκείνα τα χαρακτηριστικά τα οποία πέφτουν στο ίδιο κυτίο του συνδυαστικού ιστογράμματος. Αυτοί οι inliers φαίνονται σαν αντιστοιχίες με μαύρο χρώμα στο Σχήμα 2.3. Ο κάθε inlier αντιστοιχεί σε έναν μην μηδενικό όρο του εσωτερικού γινομένου της σχέσης (2.21). Ουσιαστικά το Σχήμα 2.3 απεικονίζει όλες τις αντιστοιχίες για τους χάρτες που έχουν σαν πηγές τα δύο μαύρα χαρακτηριστικά. Μεγιστοποιώντας ως προς όλες τις πηγές  $\hat{x}, \hat{y}$  καταλήγουμε στο μέτρο ομοιότητας μας ανάμεσα σε δύο εικόνες (2.21), όπου οι πιθανοί χάρτες σε αντιστοιχία περιορίζονται με το να έχουν ίδια οπτική λέξη για την πηγή τους. Ο τελικός βαθμός ομοιότητας είναι το εσωτερικό γινόμενο για τους δύο πιο ευθυγραμμισμένους χάρτες των δύο εικόνων.

Παρατηρούμε στην σχέση (2.21) έναν σαφή διαχωρισμό ανάμεσα (α) στην ευθυγράμμιση και την καταμέτρηση inliers, βασισμένη στην χωρική πληροφορία και (β) στις αντιστοιχίες με βάση την οπτική πληροφορία. Η κάθε καταμέτρηση inlier γράφεται ως εσωτερικό γινόμενο ανάμεσα σε δύο χάρτες. Αυτή η λειτουργία είναι απομεινάρι της επιλογής μας για το ένα προς πολλά σχήμα στην σχέση (2.11). Θα μπορούσαμε εξίσου να χρησιμοποιήσουμε την τομή ιστογραμμάτων. Από την άλλη, το σχήμα 1-1 αυξάνει τις υποθέσεις και τις πιθανότητες για ευθυγράμμιση. Μία επιλογή για ταχύτητα κι ευρωστία είναι να χρησιμοποιηθούν ως πηγές  $\hat{x}, \hat{y}$  μόνο χαρακτηριστικά τα οποία είναι τα μοναδικά που προβάλλονται σε μία οπτική λέξη, όπως και στο [17]. Συγκεκριμένα, προσθέτουμε τον περιορισμό  $h_b(X) = h_b(Y) = 1$  στο εξωτερικό maximum της σχέσης (2.21).

Θα αναφερόμαστε στην σχέση (2.21) ως ομοιότητα χαρτών χαρακτηριστικών (feature map similarity - FMS) μεταξύ δύο εικόνων. Εάν  $n$  είναι το μέσο πλήθος χαρακτηριστικών ανά εικόνα, τότε ο χρόνος που χρειάζεται για να υπολογιστεί η τομή ιστογραμμάτων είναι ανάλογη του μεγέθους των χαρτών, δηλαδή  $O(n)$ . Όταν το οπτικό λεξικό είναι αρκετά μεγάλο και χρησιμοποιούμε πηγές όπως ορίστηκε παραπάνω, τότε η μεγιστοποίηση γίνεται πάνω σε  $O(j)$  συνδυασμούς, όπου  $j$  είναι ο μέσος αριθμός κοινών λέξεων. Η συνολική διαδικασία έχει πολυπλοκότητα  $O(nj)$ , και  $O(n^2)$  στην χειρότερη περίπτωση. Οι

## Χάρτες χαρακτηριστικών και κατακερματισμός

ανάγκες αποθήκευσης είναι  $O(n)$  για κάθε χάρτη χαρακτηριστικών και  $O(n^2)$  για μια συλλογή στην χειρότερη περίπτωση. Οικονομία μπορεί να επιτευχθεί με περιορισμούς στην χωρική εγγύτητα μέσω μίας παραμέτρου βεληνεκούς  $\tau$ , ή με επιλογή πηγών. Στην ενότητα 2.4.4 σχολιάζονται και οι δύο λύσεις. Η οπτικοποίηση του αλγορίθμου fast spatial matching (FastSM) [66] και του FMS με αντιστοιχίες από τοπικά χαρακτηριστικά SURF παρουσιάζονται στο Σχήμα 2.4.



**Σχήμα 2.4:** Πάνω: Inliers με το FastSM και μονές αντιστοιχίες. Κάτω: Το ίδιο για το FMS. Υπάρχουν 35 and 32 inliers, αντίστοιχα. Οι πηγές φαίνονται με κόκκινους κύκλους με την κλίμακα και την περιστροφή της περιοχής. Οι inliers φαίνονται με κίτρινες γραμμές.

## 2.4 Χάρτες χαρακτηριστικών και κατακερματισμός

Η σχέση (2.21) παρέχει έναν πολύ γρήγορο τρόπο για να ταιριάξει κανείς δύο εικόνες. Παρόλα αυτά, δεν είναι αρκετό για δεικτοδότηση. Το εσωτερικό γινόμενο ή η τομή ιστογράμματος έχει πολυπλοκότητα  $O(n)$  με αραιή αναπαράσταση και θα χρειαστούμε μία συνάρτηση σκαριφήματος (sketch function) για να παράγουμε ένα προσεγγιστικό μέτρο ομοιότητας. Πριν προχωρήσουμε, θα χαλαρώσουμε την μεγιστοποίηση πάνω σε όλους τους συνδυασμούς των χαρακτηριστικών. Δεδομένων δύο συλλογών χαρτών χαρακτηριστικών  $F = F(X)$ ,  $G = G(X)$  που αναπαριστούν δύο εικόνες, και θεωρώντας ότι

υπάρχει 1-1 αντιστοιχία μεταξύ χαρακτηριστικών και χαρτών, η σχέση (2.21) δίνει

$$\hat{S}_F(F, G) = \max_{f \in F} \max_{g \in G} s_F(f, g), \quad (2.22)$$

όπου  $s_F(f, g)$  είναι είτε εσωτερικό γινόμενο είτε τομή. Παρατηρούμε την ομοιότητα με το ένα προς πολλά σχήμα της σχέσης (2.9), όπου το άθροισμα έχει αντικατασταθεί με τον μέγιστο. Αυτό υποδηλώνει ότι η διαδικασία για το ταίριασμα συνόλου χαρακτηριστικών μπορεί να προσαρμοστεί για σύνολα χαρτών χαρακτηριστικών. Ο τοπικά ευαίσθητος κατακερματισμός (locality sensitive hashing) παρέχει μία γρήγορη και χωρίς επίβλεψη λύση. Δίνουμε παρακάτω μία σύντομη περιγραφή, αναδεικνύουμε τα προβλήματα και παρουσιάζουμε την λύση μας.

### 2.4.1 Τοπικά ευαίσθητος κατακερματισμός

Όπως ορίζεται στο [13], δεδομένου ενός χώρου χαρακτηριστικών  $\mathbb{F}$ , με τον όρο *συνάρτηση κατακερματισμού* (hash function) αναφερόμαστε σε μία τυχαία προβολή  $h : \mathbb{F} \rightarrow \mathbb{H}$  τέτοια ώστε η πιθανότητα δύο αντικείμενα από το  $\mathbb{F}$  να προβάλλονται στην ίδια τιμή κατακερματισμού στον χώρο  $\mathbb{H}$  να αντικατοπτρίζει την ομοιότητα τους. Ένα σχήμα τοπικά ευαίσθητου κατακερματισμού (LSH) είναι μία κατανομή σε μία οικογένεια  $\mathcal{F}$  τέτοιο προβολών, ώστε για όλα τα  $f, g \in \mathbb{F}$ ,

$$\Pr_{h \in \mathcal{F}}[h(f) = h(g)] = s_{\mathcal{F}}(f, g), \quad (2.23)$$

όπου  $s_{\mathcal{F}}(\cdot, \cdot) \in [0, 1]$  είναι ένα μέτρο ομοιότητας. Στην περίπτωση μας είναι εσωτερικό γινόμενο ή τομή, κανονικοποιημένο ανάλογα κάθε φορά. Για μία οποιαδήποτε συνάρτηση κατατεμαχισμού  $h \in \mathcal{F}$ , ορίζουμε το μέτρο ομοιότητας  $s_h(f, g) = \mathbb{1}_{h(f)=h(g)}$  τέτοιο ώστε  $\mathbb{E}_{h \in \mathcal{F}}[s_h(f, g)] = s_{\mathcal{F}}(f, g)$ .

### 2.4.2 Τυχαίες αντιμεταθέσεις

Ένας χάρτης χαρακτηριστικών είναι ένα εξαιρετικά αραιό διάνυσμα. Ο αριθμός των χαρακτηριστικών σε ένα κουτί του χάρτη είναι μία τυχαία μεταβλητή η οποία με ομοιόμορφη κατανομή (μία λογική υπόθεση τουλάχιστον για το χωρικό κομμάτι), δίνεται από την διωνυμική κατανομή  $\text{Bi}(\cdot; n, k^{-1})$ . Για  $n, k_v$  και  $k_u$  της τάξης  $10^3, 10^5$  και  $10^2$ , αντίστοιχα, η προσδοκώμενη τιμή είναι της τάξης  $10^{-4}$ . Έτσι, το πλήθος των χαρακτηριστικών ενός κυτίου είναι συνήθως  $\{0, 1\}$ . Σε αυτό το σημείο, δεδομένου ενός χάρτη  $f$ , ορίζουμε το  $\bar{f} \subset \mathcal{W}$  να περιέχει μόνο εκείνα τα στοιχεία του  $\mathcal{W}$  για τα οποία το αντίστοιχο κυτίο είναι μη-κενό:  $\bar{f} = \{w \in \mathcal{W} : f^T \mathbf{e}_w \neq 0\}$ . Ο χώρος χαρακτηριστικών τώρα είναι το δυναμοσύνολο του  $\mathcal{W}$ ,  $\mathbb{F} = \mathcal{P}(\mathcal{W})$ , το σύνολο όλων των υποσυνόλων του  $\mathcal{W}$ . Σε αυτή

την περίπτωση, το εσωτερικό γινόμενο και η τομή δύο χαρτών  $f, g$  προσεγγίζονται και τα δύο πολύ καλά από  $|\bar{f} \cap \bar{g}|$ .

Αυτό ανοίγει τον δρόμο για την χρήση της μεθόδου *ανεξάρτητων αντιμεταθέσεων* (min-wise independent permutations) [10]. Η συνάρτηση κατακερματισμού εδώ προβάλλει τα αντικείμενα πίσω στο  $\mathcal{W}$ , δηλαδή  $\mathbb{H} = \mathcal{W}$  και  $h : \mathcal{P}(\mathcal{W}) \rightarrow \mathcal{W}$ . Δεδομένου ενός χάρτη  $\bar{f} \subset \mathcal{W}$ , η συνάρτηση ορίζεται ως  $h(\bar{f}) = \min\{\pi(\bar{f})\}$ , όπου  $\pi$  είναι μία αντιμετάθεση (permutation) επιλεγμένη ομοιόμορφα από το σύνολο όλων των αντιμεταθέσεων  $|\mathcal{W}|$  στοιχείων. Έπειτα για όλα τα  $\bar{f}, \bar{g} \subset \mathcal{W}$ ,

$$s_{\mathcal{F}}(\bar{f}, \bar{g}) = \frac{|\bar{f} \cap \bar{g}|}{|\bar{f} \cup \bar{g}|} = J(\bar{f}, \bar{g}), \quad (2.24)$$

αυτό σημαίνει ότι, τα  $\bar{f}, \bar{g}$  προβάλλονται στο ίδιο κυτίο με πιθανότητα η οποία αντικατοπτρίζει την ομοιότητα τους με τον συντελεστή Jaccard.

#### 2.4.3 Ταίριασμα σκαριφημάτων χαρτών

Στην πράξη, μπορούμε να εκτιμήσουμε το  $s_{\mathcal{F}}(\bar{f}, \bar{g})$  απλά προσεγγίζοντας στατιστικά το  $\mathbb{E}_{h \in \mathcal{F}}[s_h(\bar{f}, \bar{g})]$ . Χρειάζεται να κατασκευάσουμε ένα σύνολο  $\Pi = \{\pi_i : i = 1, \dots, m\}$  από  $m$  ανεξάρτητες τυχαίες αντιμεταθέσεις και να αναπαραστήσουμε τον κάθε χάρτη  $\bar{f}$  με ένα σκαρίφημα χάρτη  $\mathbf{f} \in \mathcal{W}^m$ ,

$$\mathbf{f} = \mathbf{f}(\bar{f}) = [\min\{\pi_1(\bar{f})\}, \dots, \min\{\pi_m(\bar{f})\}]^T. \quad (2.25)$$

Ορίζουμε την ομοιότητα σκαριφημάτων απλά ως  $s_K(\mathbf{f}, \mathbf{g}) = m - \|\mathbf{f} - \mathbf{g}\|_0$ , το οποίο είναι ο αριθμός των στοιχείων τα οποία τα σκαριφήματα  $\mathbf{f}, \mathbf{g}$  έχουν κοινά. Όταν υπάρχει τουλάχιστον έναν τέτοιο στοιχείο, λέμε ότι τα σκαριφήματα *συγκρούονται*. Εάν  $\mathbf{F} = \mathbf{F}(F) = \{\mathbf{f}(\bar{f}) : f \in F\}$  είναι μία συλλογή σκαριφημάτων χαρτών του  $F$ , τότε το μέτρο ομοιότητας δύο εικόνων γίνεται μέτρο ομοιότητας σκαριφημάτων

$$S_M(\mathbf{F}, \mathbf{G}) = \max_{\mathbf{f} \in \mathbf{F}} \max_{\mathbf{g} \in \mathbf{G}} s_K(\mathbf{f}, \mathbf{g}). \quad (2.26)$$

Είναι πλέον άμεσο να επαναφέρουμε τους περιορισμούς της σχέσης (2.21) στην (2.26) και να μεγιστοποιήσουμε πάνω στο περιορισμένο υποσύνολο του  $\mathbf{F} \times \mathbf{G}$  που αντιστοιχεί στα χαρακτηριστικά των δύο εικόνων που προβάλλονται στην ίδια, μοναδική οπτική λέξη. Αυτό που μένει να γίνει, είναι για κάθε πηγή μοναδική οπτικής λέξης  $\hat{v}$  να επισυνάψουμε το  $\hat{v}$  σε κάθε στοιχείο  $w \in \mathcal{W}$  του αντίστοιχου σκαριφήματος. Για να συγκρουστούν δύο σκαριφήματα τώρα πρέπει και οι πηγές τους να είναι σε αντιστοιχία. Εφόσον το κάθε στοιχείο  $w$  σχετίζεται επίσης με μία αντιμετάθεση  $\pi$ , αναπαρίσταται τώρα από την τριάδα  $(\hat{v}, w, \pi)$ .

Εφόσον  $m \ll n$ , η σχέση (2.26) δίνει έναν εξαιρετικά γρήγορο τρόπο για να προσεγγίσουμε την ομοιότητα εικόνων, με χρόνο εκτέλεσης  $O(mj)$ . Οι ανάγκες χώρου σε αυτή

την περίπτωση είναι  $O(mn)$  και εξοικονόμηση μπορεί να γίνει με επιλογή πηγών. Το πιο σημαντικό, όταν το  $m$  είναι αρκετά μικρό, για κάθε  $\mathbf{f}, \mathbf{g}$  σε ένα ζευγάρι μη όμοιων εικόνων, το  $s_K(\mathbf{f}, \mathbf{g})$  — το ίδιο και το  $S_M(\mathbf{f}, \mathbf{g})$  — είναι μηδέν με μεγάλη πιθανότητα. Αυτό συμβαίνει επειδή η πιθανότητα όλες οι  $m$  τιμές κατατεμαχισμού των δύο χαρτών να είναι διαφορετικές είναι  $(1 - p_j)^m$ , όπου  $p_j$  συντελεστής Jaccard των δύο χαρτών. Ένα γραμμικό πέρασμα σε όλες τις εικόνες της βάσης θα έδινε μία πολύ αραιή απάντηση. Αυτό δίνει την δυνατότητα χρήσης ενός ανάστροφου αρχείου για υπό-γραμμική δεικτοδότηση, όπως περιγράφεται στην ενότητα 2.4.4. Από την άλλη, μπορεί να δειχθεί ότι η η πιθανότητα σύγκρουσης αυξάνεται για τις όμοιες εικόνες.

Συγκρούσεις μπορούν να εμφανιστούν για πολλά ζευγάρια χαρτών ανάμεσα σε δύο εικόνες. Αυτό το γεγονός δεν αποτυπώνεται από τον τελεστή  $\text{max}$  στην σχέση (2.26) η οποία κρατά τον αριθμό των συγκρούσεων μόνο για το καλύτερα ευθυγραμμισμένο ζευγάρι. Επομένως αναμένουμε ο τελεστής  $\text{sum}$  πάνω στις συγκρούσεις όλων των ζευγαριών χαρτών να διακρίνει καλύτερα τις όμοιες από τις ανόμοιες εικόνες και να προσφέρει μία καλύτερη τελική ταξινόμηση των εικόνων.

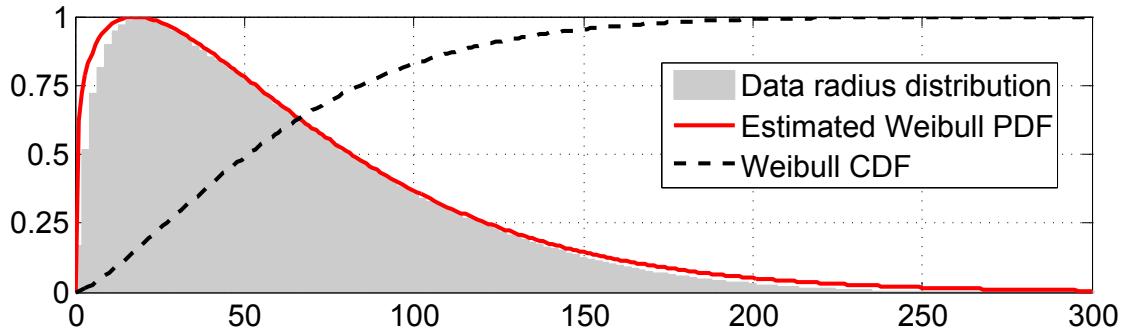
$$S_K(\mathbf{F}, \mathbf{G}) = \sum_{\mathbf{f} \in \mathbf{F}} \sum_{\mathbf{g} \in \mathbf{G}} s_K(\mathbf{f}, \mathbf{g}). \quad (2.27)$$

Θα αναφερόμαστε σε αυτό το μέτρο ομοιότητας (2.27) ως *κατατεμαχισμό χαρτών χαρακτηριστικών* (feature map hashing - FMH). Όπως και στη σχέση (2.12) και το ταίριασμα με τον αλγόριθμο RANSAC, έτσι και η σχέση (2.26) κρατά μόνο την καλύτερη υπόθεση μετασχηματισμού για να καταμετρήσει inliers. Με την χρήση του αθροίσματος, η σχέση (2.27) γίνεται όμοια με την (2.9) και το ένα προς πολλά σχήμα. Παρόλα αυτά με την χρήση πηγών που έχουν οπτική λέξη που εμφανίζεται μία μοναδική φορά μέσα στην εικόνα τότε γίνεται όμοια με το 1-1 σχήμα. Σαν αποτέλεσμα, επιτρέπουμε σε κάθε χάρτη του ενός συνόλου να ταιριάξει το πολύ με έναν χάρτη του άλλου συνόλου, όπως στην σχέση (2.26), αλλά χρησιμοποιούμε το άθροισμα της ομοιότητας όλων των σκαριφημάτων (2.27).

#### 2.4.4 Λεπτομέρειες υλοποίησης

**Χωρικός κβαντισμός** Σε ένα αναμορφωμένο σύστημα αναφοράς, αναπαριστούμε τις θέσεις των περιοχών σε πολικές συντεταγμένες  $(\rho, \theta)$ . Για να είναι η ευαισθησία ως προς την κλίμακα και την περιστροφή της περιοχής πηγής ανεξάρτητη από την απόσταση από την πηγή, οι λογαριθμικές-πολικές (log-polar) συντεταγμένες συνηθίζονται, όπως στο [9]. Στην περίπτωση μας, λόγω της αραιότητας των διανυσμάτων, είναι σημαντικό να διασφαλίσουμε ομοιόμορφη κατανομή ως προς το  $\rho$ . Όπως φαίνεται στο Σχήμα 2.5, η κατανομή του  $\rho$  φαίνεται πειραματικά κοντά στην κατανομή Weibull  $Wb(\cdot; \lambda, \kappa)$  με  $\lambda$  και  $\kappa$  να είναι οι παράμετροι κλίμακας και σχήματος αντίστοιχα, η οποία εκτιμήθηκε μέσω μέγιστης πιθανοφάνειας (maximum likelihood) [20]. Επειτα κάνοντας μη γραμμικό μετασχηματισμό

## Χάρτες χαρακτηριστικών και κατακερματισμός



**Σχήμα 2.5:** Η κατανομή της ακτίνας  $\rho$  πάνω σε 40K αναμορφωμένα σύνολα χαρακτηριστικών από 200 εικόνες της βάσης εικόνων European Cities (βλέπε ενότητα 2.5), οι οποίες περιέχουν 8M αναμορφωμένες περιοχές. Ταίριασμα μέγιστης πιθανοφάνειας της κατανομής Weibull δίνει  $\lambda = 1.23$  και  $\kappa = 68.7$ .

μέσω της Weibull CDF  $\hat{\rho} = 1 - e^{-(\rho/\lambda)^\kappa}$  η κατανομή του  $\hat{\rho}$  γίνεται περίπου ομοιόμορφη στο  $[0, 1]$ .

Εφόσον όλες οι μεγάλες τιμές του  $\rho$  προβάλλονται κοντά στο  $\hat{\rho} = 1$  και είναι χωρίς ιδιαίτερη πληροφορία, επιλέγουμε να τις αγνοήσουμε ορίζοντας  $\bar{\rho} = (\hat{\rho}/\tau) \mathbb{1}_{\hat{\rho} \in [0, \tau]}$  και αγνοώντας τα χαρακτηριστικά με  $\bar{\rho} = 0$ . Η παράμετρος βεληνεκούς  $\tau \in [0, 1]$  ελέγχει την ισορροπία μεταξύ τοπικής και ολικής γεωμετρίας. Τελικά κβαντίζουμε ομοιόμορφα τα  $\bar{\rho}$  και  $\theta$  σε  $k_\rho$  και  $k_\theta$  κυτία στο  $[0, 1]$  και  $[0, 2\pi]$  αντίστοιχα, έτσι ώστε  $k_\rho k_\theta = k_u$ . Η χωρική προβολή  $(\bar{\rho}, \theta)$  οπτικοποιείται στα δεξιά του Σχήματος 2.2, όπου η μη-γραμμική παραμόρφωση κοντά στο  $\bar{\rho} = 1$  είναι ορατή.

**Επιλογή πηγών κι απαιτήσεις μνήμης** Χρησιμοποιώντας σαν πηγές μόνο χαρακτηριστικά τα οποία είναι τα μοναδικά που προβάλλονται σε μία οπτική λέξη όχι μόνο κάνουμε οικονομία μνήμης αλλά κι επιταχύνουμε της διαδικασία ταιριάσματος. Επιπλέον συμπίεση μπορεί να επιτευχθεί επιλέγοντας έναν περιορισμένο αριθμό από χαρακτηριστικά σαν πηγές. Ο ρόλος των πηγών είναι να παρέχουν γεωμετρική ευθυγράμμιση ανάμεσα σε δύο εικόνες. Επομένως κρατάμε εκείνα τα χαρακτηριστικά σαν πηγές τα οποία προβάλλονται σε οπτικές λέξεις οι οποίες έχουμε δει ότι τείνουν να ανήκουν σε χαρακτηριστικά τα οποία ευθυγραμμίζονται καλά. Η ευθυγράμμιση έχει μετρηθεί ανά οπτική λέξη σε μία διαδικασία εκμάθησης. Όπως περιγράφεται στην ενότητα 2.5 μεταβάλλουμε το ποσοστό των λέξεων τις οποίες κρατάμε ώστε η απόδοση να μην επηρεάζεται πολύ. Ορίζουμε ως  $\nu$  το μέσο αριθμό πηγών (ή χαρτών) ανά εικόνα σύμφωνα με αυτή την μέθοδο επιλογής.

Οι ανάγκες μνήμης συνοψίζονται στον Πίνακα 2.1. Μία συλλογή σκαριφημάτων χαρτών, η οποία είναι η τελική αναπαράσταση μιας εικόνας, έχει  $\nu$  σκαριφήματα χαρτών, με  $m$  στοιχεία το καθένα. Συνεπώς μία συλλογή σκαριφημάτων χαρτών, χρειάζεται 40 Kbytes για να αποθηκευθεί. Η κάθε εικόνα απαιτεί  $\nu m$  καταχωρήσεις για να αποθηκευθεί σε ένα ανάστροφο αρχείο και συνεπώς 40 Kbytes μνήμης.

Αναπαράσταση εικόνας		Ανάστροφο αρχείο	
spatial bin id	5 bits	image id	16 bits
visual word id	18 bits	origin id	10 bits
joint bin id	4 bytes	both ids	4 bytes
map sketch	200 bytes	total	40Kbytes
map collection	40Kbytes		

Πίνακας 2.1: Χρήση μνήμης για την αναπαράσταση εικόνας με χρήση συλλογής σκαριφημάτων χαρτών και για ανάστροφο αρχείο. Η μνήμη υπολογίζεται για  $k_\rho = 4$ ,  $k_\theta = 6$ ,  $k_v = 200$ K,  $\nu = 200$ ,  $m = 50$  και μέγιστο μέγεθος βάσης εικόνων 55K.

**Δεικτοδότηση και φιλτράρισμα** Για να καταφέρουμε υπο-γραμμική πρόσβαση στις εικόνες, προ-κατασκευάζουμε όλες τις συλλογές σκαριφημάτων χαρτών και τις αποθηκεύουμε σε μία δομή ανάστροφου αρχείου. Για κάθε συνδυασμό οπτικής λέξης πηγής  $\hat{v} \in \mathcal{V}$ , κυτίου  $w \in \mathcal{W}$ , και αντιμετάθεσης  $\pi \in \Pi$ , αποθηκεύουμε μία προβολή από την τριάδα  $(\hat{v}, w, \pi)$  σε μία λίστα (posting list) όλων των σχετικών χαρτών και εικόνων που βρέθηκαν στην βάση.

Κατά την στιγμή του ερωτήματος, υπολογίζουμε τα σκαριφήματα χαρτών για την εικόνα ερώτημα, εξάγουμε τις τριάδες  $(\hat{v}, w, \pi)$ , κι έτσι έχουμε πρόσβαση στις σχετικές λίστες και κατασκευάζουμε ένα αραιό διάνυσμα από τους χάρτες και τις εικόνες που βρέθηκαν, μαζί με τις αντίστοιχες ψήφους. Σαν αποτέλεσμα για κάθε σκαριφήματα του ερωτήματος  $f$  και σκαριφήματα της βάσης  $g$ , υπολογίζουμε την ομοιότητα  $S_K(F, G)$  χωρίς ποτέ να υπολογίζουμε ρητά κανένα μηδενικό στοιχείο των όρων  $s_K(f, g)$  της σχέσης (2.27). Κατά την διαδικασία αυτή, κρατάμε επίσης το ζευγάρι χαρτών  $(f, g)$  το οποίο έδωσε την μέγιστη ομοιότητα για κάθε εικόνα της βάσης. Για μία βάσης 50K εικόνων και  $m = 50$  αντιμεταθέσεις, μόνο περίπου 2K εικόνες ανακτώνται κατά την διάρκεια του ερωτήματος και μάλιστα σε περίπου μόνο 50ms, κατά μέσο όρο.

**Τοπική βελτιστοποίηση και ανακατάταξη** Το καλύτερο ζευγάρι  $(f, g)$  ανάμεσα στην εικόνα ερώτημα και σε εικόνα της βάσης δίνει μία αντίστοιχιά μεταξύ περιοχών  $L \leftrightarrow R$ , επομένως και μία εκτίμηση για τον μετασχηματισμό  $RL^{-1}$  από την μία εικόνα στην άλλη. Ακόμα κι αν η εκτίμηση είναι κάπως χονδροειδής, π.χ. μετασχηματισμός ομοιότητας με χρήση SURF χαρακτηριστικών, μπορούμε ακόμα να ανακτήσουμε τον σωστό αφινικό μετασχηματισμό δεδομένων τουλάχιστον τριών αντίστοιχιών. Χρησιμοποιούμε την αρχική εκτίμηση ως είσοδο για ένα και μοναδικό βήμα της μεθόδου 3 του αλγορίθμου of LO-RANSAC [15]. Επανεκτιμούμε τις παραμέτρους του μετασχηματισμού με την μέθοδο ελαχίστων τετραγώνων σε όλο το σύνολο των inliers που βρίσκουμε σε κάθε επανάληψη. Είδαμε στα πειράματα μας πως 3 επαναλήψεις είναι αρκετές. Αναδιατάσσουμε της λίστα των φιλτραρισμένων εικόνων σύμφωνα με τον αριθμό των inliers που βρέθηκαν. Αυτή η διαδικασία είναι μία τάξη μεγέθους πιο γρήγορη από το fast spatial matching (FastSM) [66].

## Πειράματα - χάρτες και κατακερματισμός

Το τελευταίο είναι αυτό που χρησιμοποιούμε για να αναδιατάξουμε τις λίστες των μεθόδων BoW και WGC στις συγκρίσεις μας στην ενότητα 2.5, εφόσον δεν υπάρχει αρχική εκτίμηση σε αυτή την περίπτωση. Ο χρόνος εκτέλεσης για τον τρόπο αναδιάταξης που προτείνουμε είναι κατά μέσο όρο 0.5ms ανά εικόνα, με το FastSM να είναι στα 8ms.

## 2.5 Πειράματα - χάρτες και κατακερματισμός

### 2.5.1 Βάσεις εικόνων

Έχουμε διεξάγει πειράματα σε δύο δημόσια διαθέσιμες βάσεις εικόνων, την *Oxford Buildings*<sup>1</sup> και την *INRIA Holidays*<sup>2</sup>, όπως επίσης και στην δική μας *European Cities*<sup>3</sup>. Οι δύο πρώτες είναι μικρές (5K και 1.4K εικόνες αντίστοιχα) και συνήθως συνδυάζονται με μεγάλα σύνολα άσχετων εικόνων περίσπασης που έχουν συλλεχθεί από το Flickr μέσω συνηθισμένων ετικετών/λέξεων περιγραφής των χρηστών. Η βάση *European Cities* αποτελείται από 50778 εικόνες με γεωγραφικό στίγμα από 14 Ευρωπαϊκές πόλεις, οι οποίες συλλέχθηκαν από το Flickr με χρήση γεωγραφικών ερωτημάτων που καλύπτουν ένα παράθυρο γύρω από το κέντρο της πόλης. Ένα υποσύνολο 778 εικόνων από τις 9 πόλεις έχουν ομαδοποιηθεί σε 20 ομάδες εικόνων που απεικονίζουν την ίδια σκηνή, κτίριο ή αξιοθέατο. Εφόσον δεν είναι όλα αξιοθέατα η προσημείωση δεν μπορεί να βασιστεί μόνο στις ετικέτες των χρηστών, αλλά είναι ένας συνδυασμός επέκτασης του ερωτήματος (query expansion) στην ανάκτηση εικόνας και χειροκίνητου ξεκαθαρίσματος. Πέντε εικόνες από κάθε ομάδα έχουν επιλεχθεί σαν ερώτημα, με αποτέλεσμα συνολικά να έχουμε 100 ερωτήματα. Οι υπόλοιπες 50K εικόνες από τις άλλες 5 πόλεις είναι οι εικόνες περίσπασης. Οι περισσότερες απεικονίζουν αστικές σκηνές όπως επίσης και η βάση με τις προσημειωμένες εικόνες, δημιουργώντας έτσι ένα πιο δύσκολο κι απαιτητικό σύνολο εικόνων περίσπασης. Σύνολα εικόνων ερωτήματος απεικονίζονται στο Σχήμα 2.6, ενώ αντιπροσωπευτικές εικόνες από κάθε ομάδα απεικονίζονται στο Σχήμα 2.7. Δείγματα εικόνων από τις εικόνες περίσπασης παρουσιάζονται στο Σχήμα 2.8.

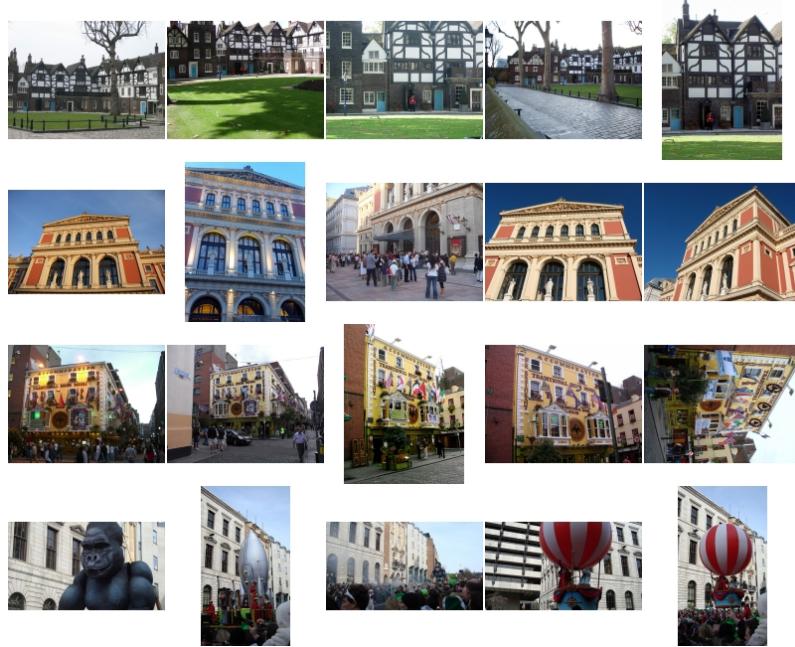
### 2.5.2 Πρωτόκολλο αξιολόγησης

Εστιάζουμε στην προσπάθεια να δείξουμε το όφελος από την δεικτοδότηση της γεωμετρίας. Τα πειράματα μας περιέχουν συγκρίσεις με την βασική μέθοδο bag-of-words κι άλλες μεθόδους για δεικτοδότηση και για ανακατάταξη με βάση την γεωμετρία. Σε όλα

<sup>1</sup> <http://www.robots.ox.ac.uk/~vgg/data/oxbuildings/>

<sup>2</sup> <http://lear.inrialpes.fr/~jegou/data.php>

<sup>3</sup> <http://image.ntua.gr/iva/datasets/>



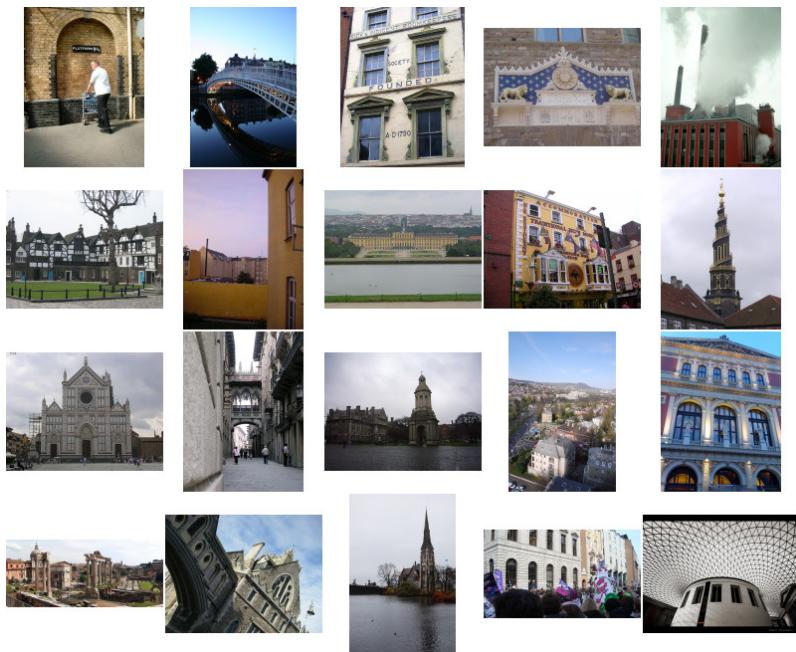
Σχήμα 2.6: Επιλεγμένες εικόνες ερωτήματα από 4 ομάδες της βάσης European Cities.

τα πειράματα, οι εικόνες μας έχουν μέγεθος με μέγιστη διάσταση  $500 \times 500$  pixels. Έχουμε εξάγει SURF χαρακτηριστικά [8] και κρατήσει μέγιστο αριθμό  $n = 1000$  χαρακτηριστικών ανά εικόνα. Χρησιμοποιούμε λεξικό  $k_v = 200K$  οπτικών λέξεων το οποίο έχει εκπαιδευθεί σε σύνολο εικόνων που δεν είναι μέρος του συνόλου που χρησιμοποιούμε για αξιολόγηση. Χρησιμοποιήσαμε τον προσεγγιστικό αλγόριθμο k-μέσων (approximate k-means) [66] για την δημιουργία του λεξικού. Η υλοποίηση μας για το BoW χρησιμοποιεί το εσωτερικό γινόμενο σε  $L_2$ -κανονικοποιημένα διανύσματα με την χρήση βαρών τύπου tf-idf. Η υλοποίηση μας για το WGC δεν χρησιμοποιεί πρότερη γνώση για την σχετική κλίμακα και περιστροφή. Για την αξιολόγηση των μεθόδων χρησιμοποιούμε το μέτρο μέσης ακρίβειας (mean average precision - mAP).

### 2.5.3 Βελτιστοποίηση παραμέτρων

Πειραματιζόμενοι στις 778 εικόνες της βάσης *European Cities*, είδαμε πως το μήκος σκαριφήματος  $m = 50$  είναι ένας καλός συμβιβασμός ανάμεσα στην υψηλή ανάκληση και στις αραιές απαντήσεις. Οι μετρήσεις mAP με τα 100 ερωτήματα και με  $\tau = 0.95$  στον Πίνακα 2.2 δείχνουμε την απόδοση για διαφορετικά επίπεδα χωρικού κβαντισμού. Μεγάλα κυτία χαλαρώνουν στο χωρικό ταίριασμα οπότε ανακτούν περισσότερες ανόμοιες εικόνες, ενώ πολύ μικρά αυξάνουν την ευαισθησία ως προς την ευθυγράμμιση. Επιλέγουμε τελικά  $k_p = 4$  και  $k_\theta = 6$ . Παρόμοια πειράματα για την παράμετρο βεληνεκούς  $\tau$  υποδεικνύουν σταθερή απόδοση για τιμές στο [.5, .95]. Επιλέγουμε  $\tau = 0.7$  σαν καλό συμβιβασμό ανάμεσα στην ολική γεωμετρία και τις ανάγκες μνήμης.

## Πειράματα - χάρτες και κατακερματισμός



**Σχήμα 2.7:** Αντιπροσωπευτικές εικόνες από κάθε προσημειωμένη ομάδα της βάσης European Cities.

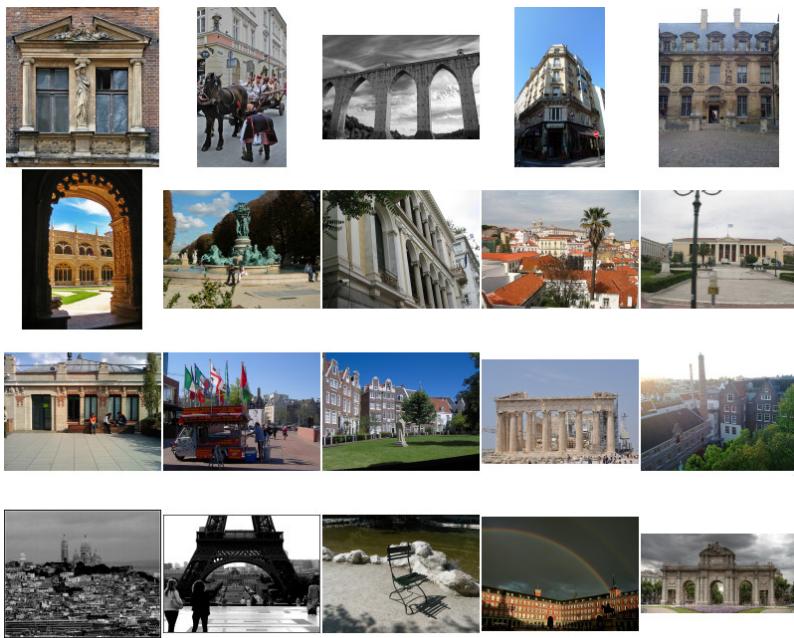
$k_\rho \times k_\theta$	2 × 3	4 × 6	8 × 12	16 × 24
mAP	0.663	0.689	0.648	0.618

**Πίνακας 2.2:** Μέση ακρίβεια (mAP) για διαφορετικά επίπεδα χωρικού κβαντισμού του FMH πάνω στην βάση European Cities με  $m = 50$  και  $\tau = 0.95$ .

Χρησιμοποιούμε τις 778 προσημειωμένες εικόνες από την βάση European Cities για να μετρήσουμε την ευθυγράμμιση ανά οπτική λέξη όπως περιγράφηκε στην ενότητα 2.4.4 και μεταβάλλουμε το ποσοστό των λέξεων που κρατάμε. Μετράμε το mAP στα 100 ερωτήματα. Ο πίνακας 2.3 δείχνει το mAP για μεταβλητό αριθμό πηγών  $\nu$ . Όταν χρησιμοποιούμε για πηγές μόνο χαρακτηριστικά με μοναδική οπτική λέξη, χωρίς καμία άλλη μέθοδο για επιλογή, τότε το  $\nu$  είναι κατά μέσο όρο 720. Τελικά υποχρεώνουμε το  $\nu$  να είναι ίσο με 200 επιλέγοντας τον κατάλληλο αριθμό από οπτικές λέξεις, σαν καλό συμβιβασμό ανάμεσα στην απόδοση και στις απαιτήσεις μνήμης.

$\nu$	720	500	300	200	100
mAP	0.687	0.682	0.678	0.676	0.642

**Πίνακας 2.3:** Μέση ακρίβεια (mAP) για μεταβλητό αριθμό πηγών  $\nu$  στην βάση European Cities.

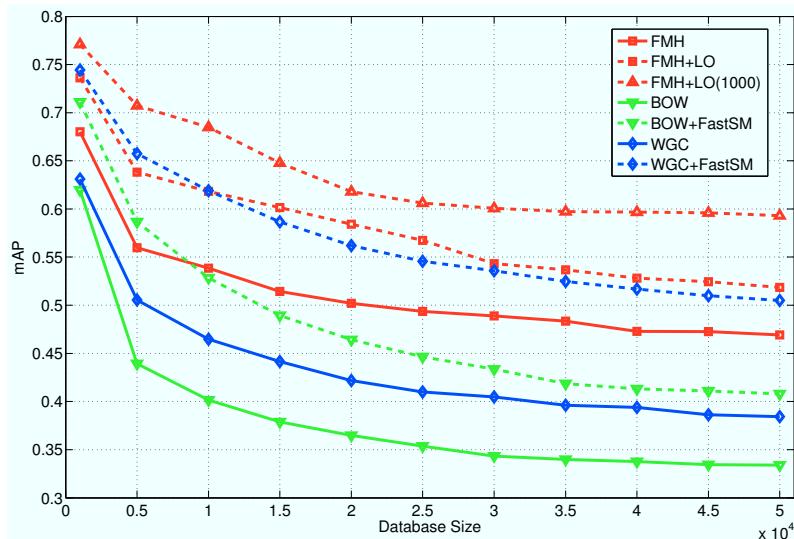


**Σχήμα 2.8:** Δείγματα εικόνων περίσπασης από την βάση European Cities.

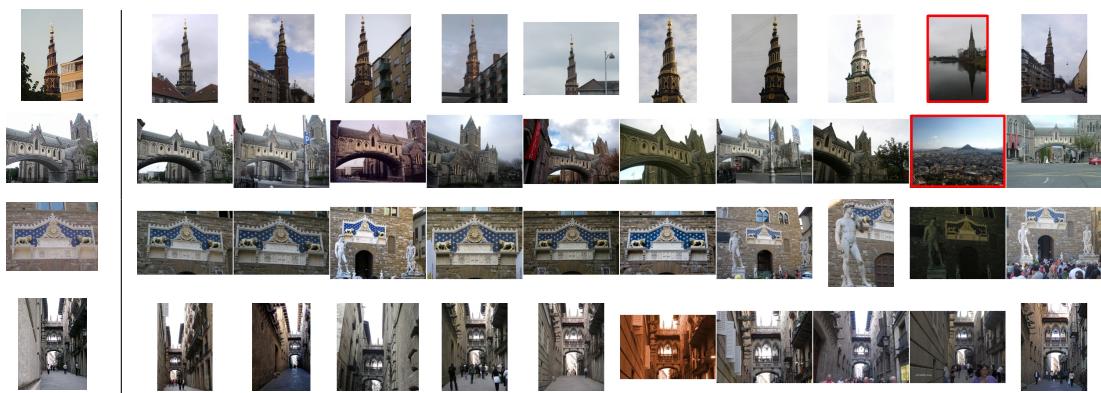
#### 2.5.4 Αποτελέσματα

Το σχήμα 2.9 παρουσιάζει την σύγκριση της προτεινόμενης μεθόδου με το (BoW) και το weak geometric consistency (WGC)[34] στην βάση *European Cities* με και χωρίς ανακατάταξη, για μεταβλητό αριθμό εικόνων περίσπασης. Το BoW με το FastSM είναι ακριβώς η μέθοδος που προτείνεται στο [66]. Το FMH αισθητά ξεπερνά τις άλλες μεθόδους με το όφελος από την δεικτοδότηση της γεωμετρίας, ιδιαίτερα σε μεγάλη κλίμακα. Είναι κάπως απρόσμενο ότι η απόδοση σταθεροποιείται κοντά στην πλήρη χρήση της βάσης. Για την ανακατάταξη έχουμε επιτρέψει να γίνεται στις 1000 πρώτες εικόνες με το local optimization (LO) και το FMH, το οποίο γίνεται σε λιγότερο χρόνο από ότι 100 εικόνες με το FastSM και το BoW ή το WGC. Αυτός ο χρόνος είναι κατά μέσο όρο 500ms και 800ms αντίστοιχα. Το Σχήμα 2.10 δείχνει παραδείγματα εικόνων ερωτημάτων και τις αντίστοιχες ανακτημένες εικόνες από την βάση *European Cities* με 50K εικόνες περίσπασης χρησιμοποιώντας το FMH χωρίς ανακατάταξη. Με χρήση του BoW χωρίς καμία χρήση γεωμετρικής πληροφορίας λανθασμένες εικόνες επιστρέφονται σε υψηλή θέση όπως φαίνεται στο Σχήμα 2.11. Παρά την χρήση της γεωμετρίας στο FMS λανθασμένες αντιστοιχίες μπορεί να εμφανιστούν εξαιτίας θορυβωδών οπτικών λέξεων, όπως δείχνουμε στο Σχήμα 2.12. Παρόλα αυτά, αυτό δεν είναι πάντα πρόβλημα στην διαδικασία ανάκτησης μας. Μία λάθος εικόνα θα έπρεπε να πάρει ψήφο ομοιότητας και μετά τον κατατεμαχισμό (FMH). Επίσης η διαδικασία ανακατάταξης μπορεί να φιλτράρει ένα τέτοιο λάθος παράδειγμα.

Ο πίνακας 2.4 συνοψίζει παρόμοια αποτελέσματα για τις βάσεις *Holidays* και *Oxford*. Χωρίς εικόνες περίσπασης, το FMH αποδίδει λίγο καλύτερα στο *Holidays* αλλά ξεπερνιέ-



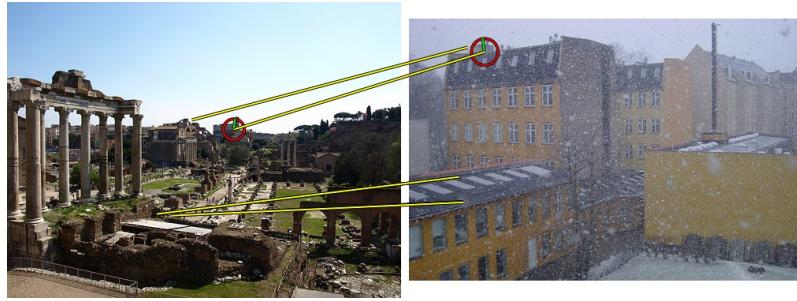
Σχήμα 2.9: Μέση ακρίβεια ( $mAP$ ) για μεταβλητό μέγεθος βάσης στο European Cities για τα BoW, WGC και FMH, με και χωρίς ανακατάταξη.



Σχήμα 2.10: Παραδείγματα εικόνων ερωτήματος και οι ανακτημένες εικόνες από την βάση European Cities με 50K εικόνες περισπασης για το FMH χωρίς ανακατάταξη. Οι λάθος εικόνες απεικονίζονται με κόκκινο πλαίσιο.



Σχήμα 2.11: Παραδείγματα εικόνων ερωτημάτων και οι ανακτημένες εικόνες από την βάση European Cities με 50K εικόνες περισπασης με το BOW χωρίς ανακατάταξη. Οι λάθος εικόνες απεικονίζονται με κόκκινο πλαίσιο..



**Σχήμα 2.12:** Παράδειγμα λανθασμένου ταιριάσματος με το FMS. Οι πηγές φαίνονται με κόκκινους κύκλους με την κλίμακα και την περιστροφή της περιοχής. Οι inliers φαίνονται με κίτρινες γραμμές.

Dataset	Holidays		Oxford		
	Method	1.4K	51.4K	5K	55K
BOW	0.583	0.492	0.372	0.329	
WGC	0.591	0.510	<b>0.375</b>	0.333	
FMH	<b>0.610</b>	<b>0.542</b>	0.362	<b>0.362</b>	
BOW+FastSM	0.622	0.537	0.421	0.356	
WGC+FastSM	0.626	0.542	<b>0.436</b>	0.388	
FMH+LO(100)	<b>0.639</b>	0.556	0.422	0.391	
FMH+LO(1000)	-	<b>0.571</b>	0.431	<b>0.410</b>	

**Πίνακας 2.4:** Μέση ακρίβεια (mAP) για τις βάσεις INRIA Holidays και Oxford Buildings, με και χωρίς εικόνες περίσπασης. Το FastSM εφαρμόζεται στις 100 πρώτες εικόνες. Ενώ το LO στις 100 και 1000.

ται από το WGC στο *Oxford*. Αντιθέτως, το FMH ολοφάνερα ξεπερνά όλες τις μεθόδους και στις δύο βάσεις με την παρουσία 50K εικόνων περίσπασης. με και χωρίς ανακατάταξη. Αυτό φαίνεται καλύτερα στην βάση *Oxford*, πιθανόν εξαιτίας του ίδιου τύπου αστικών σκηνών με το σύνολο εικόνων περίσπασης.

Το αποτέλεσμα μας για την μέθοδο BoW στην βάση *Oxford* (0.372) δεν είναι άμεσα συγκρίσιμο με το αποτέλεσμα (0.618) που έχει επιτευχθεί στο [66], όπου χρησιμοποιείται λεξικό το οποίο έχει κατασκευαστεί από τις ίδιες εικόνες της βάσης. Είναι περισσότερο συγκρίσιμο με το αποτέλεσμα στο [69] (0.403) όπου το λεξικό έχει κατασκευαστεί από άλλες εικόνες, όπως και στην περίπτωση μας. Επίσης περισσότερος απώλειες υπάρχουν κι από το γεγονός ότι κρατάμε τα καλύτερα 1000 χαρακτηριστικά για κάθε εικόνα. Τα αποτελέσματα μας για τις μεθόδους BoW και WGC στην βάση *Holidays* είναι συγκρίσιμα με αυτά του [34], (0.572) και (0.611) αντίστοιχα, όπου χρησιμοποιείται επίσης ένα γενικό λεξικό.

## 2.6 Επιλογή χαρακτηριστικών

Το μέγεθος ενός χάρτη χαρακτηριστικών είναι τετραγωνικό ως προς τον αριθμό των χαρα-

## Επιλογή χαρακτηριστικών

κτηριστικών σε μια εικόνα. Αυτό φυσικά ισχύει εάν όλα τα χαρακτηριστικά χρησιμοποιούνται και ως πηγές και ως αναμορφωμένα σημεία. Για αυτό τον λόγο, στην ενότητα 2.4 αναπτύξαμε την μεθοδολογία κατά την οποία γίνεται επιλογή με μεθόδους κατακερματισμού (FMH) [6]. Τα πειράματα τα οποία διεξάγουμε για αυτή την μέθοδο δείχνουν πως οι ανάγκες μνήμης είναι ο βασικός περιορισμός της μεθόδου, τα οποία φτάνουν μέχρι την δεικτοδότηση μιας βάσης 50K εικόνων. Αξίζει να παρατηρήσουμε πως η μέθοδος των τυχαίων αντιμεταθέσεων χρησιμοποιείται και την μέθοδο *geometric min-hashing* (GmH) [17] για να επιλέξει και τα κεντρικά σημεία αλλά και τα γειτονικά. Αυτή η επιλογή επιτρέπει την δεικτοδότηση περισσότερων εικόνων αλλά υποφέρει από χαμηλή απόδοση στην ανάκτηση εικόνων.

Σε αυτή την ενότητα παρουσιάζουμε ένα νέο σχήμα για επιλογή χαρακτηριστικών. Το ενδιαφέρον είναι πως η μεθοδολογία μας για την επιλογή είναι προσαρμοσμένη ακριβώς στις ανάγκες των χαρτών χαρακτηριστικών, με την έννοια ότι εισάγουμε διαφορετικά κριτήρια για τις πηγές και για τα αναμορφωμένα σημεία. Αυτά τα κριτήρια βασίζονται σε μία εκ των προτέρων διαδικασία χωρίς επίβλεψη η οποία εφαρμόζεται σε όλη την βάση των εικόνων που πρόκειται να δεικτοδοτηθεί. Κάθε χαρακτηριστικό επιλέγεται ανεξάρτητα, είτε ως πηγή είτε ως αναμορφωμένο σημείο.

Ο σκοπός αυτής της διαδικασίας εκμάθησης είναι να ανιχνευθούν εκείνα τα χαρακτηριστικά τα οποία είναι πιο ικανά για να ταιριάξουν ανάμεσα σε εικόνες του ίδιου αντικειμένου. Για αυτό τον σκοπό στήνουμε ένα σύστημα αναζήτησης εικόνων το οποίο βασίζεται στην μέθοδο bag-of-words και στο οποίο γίνεται γεωμετρικό ταίριασμα των πιο όμοιων εικόνων με την μέθοδο FastSM [66]. Έπειτα χρησιμοποιούμε κάθε εικόνα  $X$  από την βάση ως εικόνα ερωτήματος και παίρνουμε μία απάντηση  $\mathcal{R}(X)$ . Η απάντηση θεωρούμε πως είναι όλες εκείνες οι εικόνες οι οποίες απεικονίζουν το ίδιο αντικείμενο με την εικόνα ερωτήματος κάτω από διαφορετικές οπτικές γωνίες. Επιλέγουμε τα χαρακτηριστικά από την εικόνα  $X$  σύμφωνα με τις δομές οι οποίες επαναλαμβάνονται στο  $\mathcal{R}(X)$ . Η όλη στρατηγική είναι όμοια με εκείνη των Turcot και Lowe [82], αλλά τα κριτήρια επιλογής είναι διαφορετικά στην περίπτωση μας.

Η επιλογή η οποία γίνεται είναι εφικτή μόνο για αντικείμενα τα οποία απεικονίζονται σε τουλάχιστον δύο εικόνες στην συλλογή μας, αυτό σημαίνει εικόνες για τις οποίες το  $\mathcal{R}(X)$  είναι μη κενό. Αυτές τις ονομάζουμε εικόνες που ταιριάζανε με επιτυχία, όλες τις υπόλοιπες τις αποκαλούμε μοναδικές εικόνες, οι οποίες είναι οι μοναδικές στην συλλογή οι οποίες απεικονίζουν ένα συγκεκριμένο αντικείμενο ή σκηνή. Σε αυτές τις εικόνες η επιλογή γίνεται με βάση πληροφορία από τον ανιχνευτή χαρακτηριστικών.

Επειδή αρκετά ζευγάρια χαρτών μπορούν να ευθυγραμμισθούν κι επειδή αυτό δεν καταγράφεται από τον τελεστή max, στην σχέση (2.21) όπως και στην ενότητα 2.4 με την χρήση της μεθόδου κατακερματισμού θα αθροίσουμε τις συνεισφορές από όλα τα ζευγά-



**Σχήμα 2.13:** Αριστερά: Αρχικές αντιστοιχίες με χρήση των χαρακτηριστικών SURF και λεξικό μεγέθους 100K. Υπάρχουν δύο διαφορετικές επιφάνειες οι οποίες ταιριάζουν, και δύο σχετικοί μετασχηματισμοί. Δεξιά: Inliers με την μέθοδο FastSM, αποτυπώνοντας την μία μόνο επιφάνεια.



**Σχήμα 2.14:** Inliers με την χρήση της μεθόδου FMS για διαφορετικά ζευγάρια πηγών, για το παραδειγμα του Σχήματος 2.13. Οι πηγές φαίνονται με κόκκινο χρώμα. Οι inliers με κίτρινο. Κάθε πηγή συμβάλει στο άθροισμα της σχέσης (2.28).

ρια χαρτών ορίζοντας της ομοιότητα των χαρτών χαρακτηριστικών (*feature map similarity* - FMS) ως

$$S_F(X, Y) = \sum_{b \in V(X, Y)} \sum_{\substack{\hat{x} \in H_b(X) \\ \hat{y} \in H_b(Y)}} \langle f^{(\hat{x})}(X), f^{(\hat{y})}(Y) \rangle. \quad (2.28)$$

Στην μέθοδο που περιγράψαμε στην ενότητα 2.4 [6], οι πηγές επιλέγονται ανάμεσα από τα χαρακτηριστικά τα οποία είναι τα μοναδικά τα οποία προβάλλονται σε κάποια συγκεκριμένη οπτική λέξη. Συγκεκριμένα προσθέσαμε τον περιορισμό  $|H_v(X)| = |H_v(Y)| =$

1 στο εξωτερικό  $\max$  της σχέσης (2.21). Στην μέθοδο που θα περιγράψουμε τώρα δεν υπάρχει αυτός ο περιορισμός κι επιτρέπουμε κάθε χαρακτηριστικό να μπορεί να επιλεχθεί ως πηγή. Τέλος χρησιμοποιούμε και την αντίστροφη συχνότητα *inverse document frequency - idf* στο σχήμα ομοιότητας μας προσαρμόζοντας την δημιουργία του ιστογράμματος. Δεδομένου ενός αναμορφωμένου συνόλου σημείων  $\hat{X}$ , ορίζουμε

$$h_{\mathcal{W}}(\hat{X}) = \sum_{w=(v,u) \in \mathcal{W}} h_w(\hat{X}) \text{idf}(v) \mathbf{e}_w, \quad (2.29)$$

όπου  $\text{idf}(v)$  είναι η τιμή του  $\text{idf}$  για την οπτική λέξη  $v$  του οπτικο-χωρικού κυτίου  $w = (v, u)$ .

### 2.6.1 Πηγές χαρτών

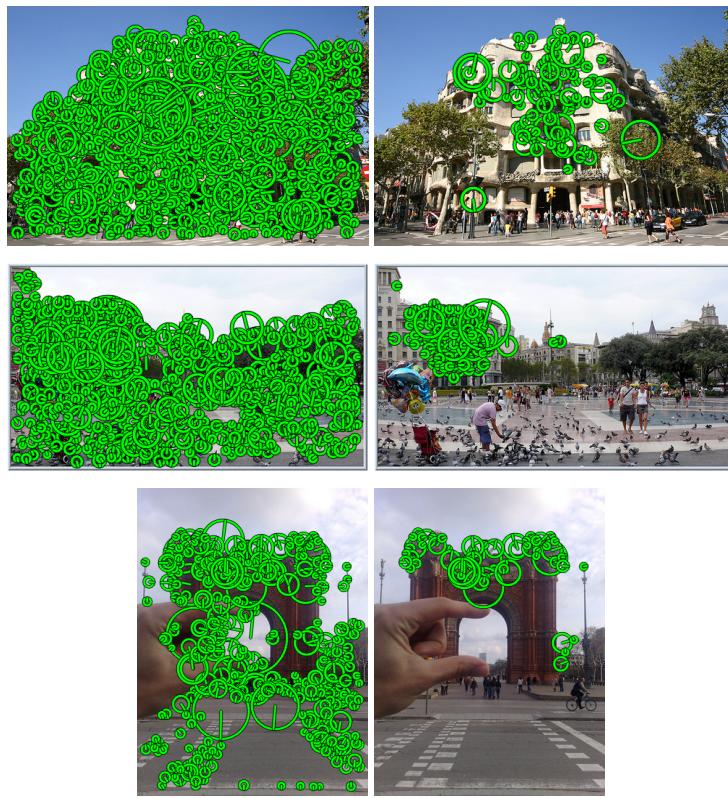
Στην υπόθεση της μονής αντιστοιχίας (2.16), καθώς και στο ταίριασμα των αναμορφωμένων συνόλων (2.17), η κάθε αντιστοιχία  $(\hat{x}, \hat{y}) \in \mathcal{C}$  καθορίζει μία υπόθεση μετασχηματισμού. Για κάθε υπόθεση γίνεται καταμέτρηση των inliers,

$$I(\mathcal{C}; \hat{x}, \hat{y}, r) = \sum_{(x,y) \in \mathcal{C}} r(p(x^{(\hat{x})}), p(y^{(\hat{y})})), \quad (2.30)$$

αυτό σημαίνει, με βάση τις αντιστοιχίες  $(x, y) \in \mathcal{C}$  ποια χαρακτηριστικά  $x, y$  ευθυγραμμίζονται όταν τα  $\hat{x}, \hat{y}$  χρησιμοποιούνται ως πηγές. Υψηλός αριθμός από inliers δηλώνει πως οι πηγές  $\hat{x}, \hat{y}$  είναι ευθυγραμμισμένες τόσο ως προς το τοπικό τους σχήμα όσο και ως προς την θέση τους, όπως αυτά περιγράφονται από τις περιοχές  $P(\hat{x}), P(\hat{y})$ . Αντίθετα οι inliers  $x, y$  είναι ευθυγραμμισμένοι μόνο ως προς την θέση τους. Φαίνεται πως το πρώτο είναι κατάλληλο κριτήριο για τις πηγές ενώ το δεύτερο για τα αναμορφωμένα σημεία. Η χρήση της εμφάνισης τους είναι επιθυμητή και στις δύο περιπτώσεις, εφόσον οι αρχικές αντιστοιχίες  $\mathcal{C}$  δημιουργούνται με βάση τις οπτικές λέξεις (2.7) από την χρήση ενός οπτικού λεξικού.

Εστιάζοντας στην επιλογή πηγών με την χρήση οπτικού λεξικού, έστω ότι  $\mathcal{C}(X, Y)$  είναι οι αρχικές αντιστοιχίες ανάμεσα στις εικόνες  $X, Y$ , όπως αυτές δίνονται από την σχέση (2.20). Κάθε χαρακτηριστικό  $z$  της εικόνας  $X$  μπορεί να συμμετέχει σε πολλές υποθέσεις ανάμεσα στην εικόνα ερωτήματος και στις εικόνες  $Y \in \mathcal{R}(X)$  οι οποίες βρέθηκαν να απεικονίζουν το ίδιο αντικείμενο με την εικόνα  $X$ . Θέλουμε να το επιλέξουμε ως πηγή εάν συμμετέχει τουλάχιστον σε έναν προκαθορισμένο αριθμό από επιτυχημένες υποθέσεις. Οπότε ορίζουμε τον μέγιστον αριθμό από inliers πάνω σε όλες τις υποθέσεις στις οποίες συμμετέχει ένα χαρακτηριστικό τον βαθμό υποστήριξης της πηγής  $z \in X$  ως

$$\alpha_X(z) = \max_{Y \in \mathcal{R}(X)} \max_{\substack{(\hat{x}, \hat{y}) \in \mathcal{C}(X, Y) \\ \hat{x} = z}} I(\mathcal{C}(X, Y); \hat{x}, \hat{y}), \quad (2.31)$$



**Σχήμα 2.15:** Αριστερά: Δείγματα εικόνων και τα εντοπισμένα χαρακτηριστικά SURF. Δεξιά: Τα χαρακτηριστικά τα οποία επιλέχθηκαν ως πηγές για κάθε εικόνα.

όπου παραλείπουμε το  $r$  κι έχουμε υποθέσει ότι η χωρική εγγύτητα μετρείται από τον ομοιόμορφο πυρήνα της σχέσης (2.13). Αντί να μετράμε απευθείας την ευθυγράμμιση δύο περιοχών μετράμε τον αριθμό των inliers τους οποίους φέρνουν. Αυτό αντικατοπτρίζει καλύτερα το ταίριασμα των αναμορφωμένων χαρακτηριστικών.

Δεδομένης μία εικόνας  $X$ , επιλέγουμε τις πηγές σύμφωνα με την υποστήριξη τους:

$$\alpha(X) = \{z \in X : \alpha_X(z) > \tau_\alpha\}. \quad (2.32)$$

Έτσι η συλλογή χαρτών χαρακτηριστικών γίνεται  $F(X) = \{f^{(\hat{x})}(X) : \hat{x} \in \alpha(X)\}$ . Το Σχήμα 2.15 απεικονίζει δείγματα εικόνων και τα αντίστοιχα χαρακτηριστικά που επιλέχθηκαν ως πηγές. Αυτά ως επί τω πλείστων εμφανίζονται πάνω στα αντικείμενα ενδιαφέροντος και όχι στο υπόβαθρο.

## 2.6.2 Αναμορφωμένα χαρακτηριστικά

Είναι σημαντικό για ένα αναμορφωμένο τοπικό χαρακτηριστικό να εμφανίζεται ως inlier για κάποιες υποθέσεις μεταξύ δύο εικόνων. Δεδομένων των επιλεγμένων πηγών  $\alpha(X), \alpha(Y)$  για τις εικόνες  $X, Y$ , έστω ότι  $\mathcal{A}(X, Y) = \mathcal{C}(\alpha(X), \alpha(Y))$  είναι το σύνολο των αντιστοιχιών μεταξύ πηγών με βάση ένα οπτικό λεξικό και σύμφωνα με την σχέση (2.20). Κάθε

## Επιλογή χαρακτηριστικών

χαρακτηριστικό  $z \in X$  μπορεί να καταλήξει ως inlier για πολλαπλές υποθέσεις οι οποίες έχουν παραχθεί από τις αντιστοιχίες  $\mathcal{A}(X, Y)$ , με εικόνες από τις  $Y \in \mathcal{R}(X)$ . Θέλουμε να το επιλέξουμε ως αναμορφωμένο χαρακτηριστικό εάν εμφανίζεται ως inlier σε τουλάχιστον μία υπόθεση. Σύμφωνα με τον ορισμό των inliers (2.30), θεωρούμε την ελάχιστη απόσταση ανάμεσα στο  $z$  και οποιοδήποτε αντίστοιχο χαρακτηριστικό  $y$  σε οποιαδήποτε εικόνα  $Y \in \mathcal{R}(X)$ , αφού έχει αυτό αναμορφωθεί με βάση οποιαδήποτε υπόθεση στο  $\mathcal{A}(X, Y)$ ,

$$\delta_X(z) = \min_{Y \in \mathcal{R}(X)} \min_{\substack{(\hat{x}, \hat{y}) \in \mathcal{A}(X, Y) \\ (x, y) \in \mathcal{C}(X, Y) \\ x=z}} \|p(x^{(\hat{x})}) - p(y^{(\hat{y})})\|_2, \quad (2.33)$$

Ο λόγος για αυτό είναι πως ένα χαρακτηριστικό  $z$  με μικρή απόσταση  $\delta_X(z)$  δηλώνει πως το  $z$  έχει ευθυγραμμιστεί τουλάχιστον με ένα χαρακτηριστικό από κάποια άλλη εικόνα, κι αυτό είναι στοιχείο πως το  $z$  μπορεί όντως να είναι inlier σε κάποια υπόθεση μιας νέας εικόνας ερωτήματος. Στην συνέχεια ορίζουμε τον βαθμό υποστήριξης για inlier ενός χαρακτηριστικού  $z \in X$ ,

$$i_X(z) = \exp \left\{ -\frac{\delta_X(z)^2}{2\sigma_i^2} \right\}, \quad (2.34)$$

ως μία φθίνουσα συνάρτηση της ελάχιστης απόστασης  $\delta_X(z)$ . Η παράμετρος  $\sigma_i$  είναι ένα μέτρο της χωρικής εγγύτητας των δύο χαρακτηριστικών που ταιριάζουν στα αναμορφωμένα πλαίσια, οπότε σχετίζεται άμεσα με το κατώφλι για τους inliers  $\epsilon$  της σχέσης (2.13), και με το μέγεθος του χωρικού κυτίου.

Δεύτερος σκοπός είναι να ενθαρρύνουμε την επιλογή χαρακτηριστικών κοντά στην πηγή. Ένας λόγος είναι πως κοντινά χαρακτηριστικά είναι πιο πιθανόν να ανήκουν στην ίδια επιφάνεια ή αντικείμενο, οπότε και πιο πιθανό να ακολουθούν τον ίδιο μετασχηματισμό. Ένας δεύτερος λόγος είναι πως οι υποθέσεις μετασχηματισμών δημιουργούνται από το τοπικό σχήμα, οπότε είναι μία εκτίμηση του πραγματικού μετασχηματισμού, η οποία είναι πιθανότερο να είναι σωστή κοντά στην πηγή. Έτσι, δεδομένης της πηγής  $\hat{x}$  της εικόνας  $X$ , η ισχύς τοπικότητας  $\ell^{(\hat{x})}(z)$  του χαρακτηριστικού  $z \in X$  με βάση το  $\hat{x}$  είναι μία φθίνουσα συνάρτηση της απόστασης  $\rho^{(\hat{x})}(z)$  του  $z$  από την πηγή όταν αυτό έχει αναμορφωθεί. Για την ακρίβεια, επιλέγουμε

$$\ell^{(\hat{x})}(z) = \exp \left\{ -\frac{\rho^{(\hat{x})}(z)^2}{2\sigma_\ell^2} \right\}, \quad (2.35)$$

όπου η παράμετρος  $\sigma_\ell$  καθορίζει την ισορροπία ανάμεσα στην τοπική και ολική γεωμετρία. Ο πλήρης βαθμός υποστήριξης αναμορφωμένου χαρακτηριστικού του  $z \in X$  με βάση την πηγή  $\hat{x} \in X$  είναι

$$\beta_X^{(\hat{x})}(z) = i_X(z)\ell^{(\hat{x})}(z), \quad (2.36)$$



**Σχήμα 2.16:** Ζευγάρια χαρτών χαρακτηριστικών που ταίριαζαν και οι αντίστοιχοι *inliers* μετά από την επιλογή των αναμορφωμένων χαρακτηριστικών, ακολουθώντας το παράδειγμα των Σχημάτων 2.13 και 2.14. Οι *inliers* είναι λιγότεροι από εκείνους του Σχήματος 2.14, αλλά αρκετοί για να ταιριάζουν τις δύο εικόνες, ειδικά εφόσον οι *inliers* αθροίζονται πάνω σε όλους τους χάρτες χαρακτηριστικών. Σημειώνουμε πως αυτοί δεν είναι οι μοναδικοί χάρτες για τους οποίους εμφανίστηκαν *inliers* ανάμεσα στις απεικονιζόμενες εικόνες.

ευνοώντας έτσι χαρακτηριστικά με υψηλή υποστήριξη ως *inlier* και τα οποία βρίσκονται κοντά στην πηγή. Αυτό το σχήμα επιλογής παρέχει μία ευκαμψία διότι επιλέγονται διαφορετικά χαρακτηριστικά σε κάθε χάρτη. Δεδομένης της εικόνας  $X$  και της πηγής  $\hat{x} \in X$ , πρώτα επιλέγουμε χαρακτηριστικά σύμφωνα με τον βαθμό υποστήριξης των αναμορφωμένων χαρακτηριστικών, ως

$$\beta^{(\hat{x})}(X) = \{z \in X : \beta_X^{(\hat{x})}(z) > \tau_\beta\}, \quad (2.37)$$

κι έπειτα κατασκευάζουμε τους αντίστοιχους χάρτες  $f^{(\hat{x})}(\beta^{(\hat{x})}(X))$ . Ένα παράδειγμα των *inliers* με χρήση χαρτών και της επιλογής χαρακτηριστικών παρουσιάζεται στο Σχήμα 4.1.

### 2.6.3 Μοναδικές εικόνες

Εικόνες για τις οποίες η απάντηση με τις όμοιες τους ( $R(X)$ ) είναι κενή, είναι εικόνες οι οποίες είναι οι μοναδικές σε όλη την συλλογή που απεικονίζουν κάποια όψη ενός αντικειμένου. Δυστυχώς τέτοιες εικόνες είναι η πλειοψηφία στην πράξη. Χωρίς να υπάρχει κάποια άλλη πηγή πληροφορίας να βασιστούμε για την επιλογή, στρεφόμαστε προς την ισχύ του κάθε χαρακτηριστικού η οποία προκύπτει από τον ανιχνευτή χαρακτηριστικών. Αυτή η επιλογή έρχεται σε αντίθεση με το [82], όπου η επιλογή γίνεται με βάση την κλί-

μακα των περιοχών. Για πηγές κρατάμε έναν προκαθορισμένο αριθμό χαρακτηριστικών, αυτά με την μεγαλύτερη ισχύ του ανιχνευτή. Για τα αναμορφωμένα χαρακτηριστικά, χρησιμοποιούμε μία φθίνουσα συνάρτηση της ισχύς του ανιχνευτή για να αντικαταστήσουμε τον βαθμό υποστήριξης για inlier της σχέσης (2.34), κι έπειτα την συνδυάζουμε με την ισχύ τοπικότητας όπως στην σχέση (2.36). Μία συγκεκριμένη επιλογή αυτής της συνάρτησης παρουσιάζεται στην ενότητα 2.6.4. Πάλι, κρατάμε έναν προκαθορισμένο αριθμό χαρακτηριστικών, αυτά με τον μεγαλύτερο βαθμό υποστήριξης.

#### 2.6.4 Λεπτομέρειες υλοποίησης

Σε αυτή την ενότητα παρουσιάζουμε λεπτομέρειες σχετικά με την κανονικοποίηση της θέσης των αναμορφωμένων χαρακτηριστικών ως προς την πηγή και σχετικά με την επιλογή των χαρακτηριστικών. Ανακατάταξη δεν εφαρμόζεται όπως στην περίπτωση των σκαριφημάτων χαρτών καθώς πειραματικά είδαμε πως δεν είναι απαραίτητη για να βελτιώσει τα αποτελέσματα. Οι λεπτομέρειες δεικτοδότησης είναι όμοιες με εκείνες που παρουσιάστηκαν στην ενότητα 2.4.4.

**Χωρικός κβαντισμός** Στο αναμορφωμένο πλαίσιο, κωδικοποιούμε τις θέσεις σε πολικές συντεταγμένες  $(\rho, \theta)$  ομοίως με την ενότητα 2.4.4. Εφαρμόζουμε και πάλι μη-γραμμικό μετασχηματισμό στην απόσταση  $\rho$  με την συνάρτηση κατανομή πιθανότητα της Weibull

$$F_{\lambda,\kappa}(\rho) = 1 - e^{-(\rho/\lambda)^\kappa} \quad (2.38)$$

η οποία μετατρέπει την κατανομή πολύ κοντά σε ομοιόμορφη στο  $[0, 1]$ .

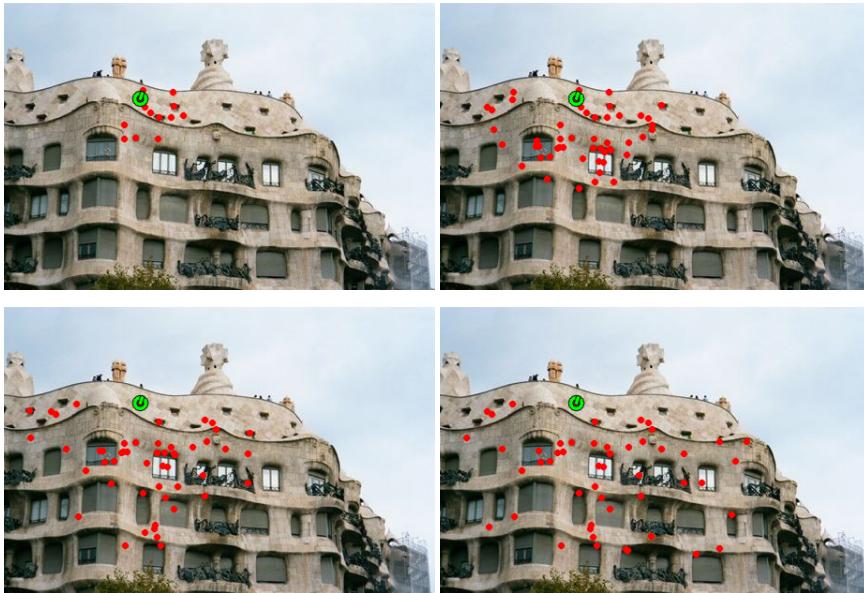
Ας θεωρήσουμε τώρα το χαρακτηριστικό  $z \in X$  αναμορφωμένο ως προς το  $\hat{x}$ . Ο βαθμός υποστήριξης του ως inlier είναι  $i_X(z) \leq 1$  από την σχέση (2.34), οπότε εάν το  $z$  πρόκειται να επιλεχθεί ως αναμορφωμένο χαρακτηριστικό, οι Σχέσεις (2.36) και (2.37) υπονοούν ότι  $\tau_\beta < \beta^{(\hat{x})}(x) \leq \ell^{(\hat{x})}(z)$ , έτσι ώστε  $\rho^{(\hat{x})}(z) < (-2\sigma_\ell^2 \ln \tau_\beta)^{1/2}$  από την σχέση (2.35). Αυτό το άνω όριο μας παρακινεί να περικόψουμε την ακτίνα  $\rho$ , κωδικοποιώντας την ως

$$\bar{\rho} = \begin{cases} \frac{1}{\tau} F_{\lambda,\kappa}(\rho), & \text{if } F_{\lambda,\kappa}(\rho) \in [0, \tau) \\ 0, & \text{αλλιώς,} \end{cases} \quad (2.39)$$

όπου

$$\tau = F_{\lambda,\kappa}((-2\sigma_\ell^2 \ln \tau_\beta)^{1/2}) \quad (2.40)$$

είναι μία παράμετρος βεληνεκούς, και τελικά επίσης απορρίπτουμε όλα τα χαρακτηριστικά με  $\bar{\rho} = 0$ . Σημειώνουμε πως το  $\tau$  στην ενότητα 2.4.4 χρησιμοποιήθηκε για να ελέγχει απευθείας την ισορροπία ανάμεσα σε ολική και τοπική γεωμετρία. Αντίθετα, τώρα κρατάμε το  $\tau_\beta$  σταθερό, κι ελέγχουμε το  $\tau$  μέσω της παραμέτρου  $\sigma_\ell$ . Με αυτό τον τρόπο έχουμε κοινό έλεγχο για τον βαθμό υποστήριξης τοπικότητας και για inliers.



**Σχήμα 2.17:** Επιλεγμένα αναμορφωμένα χαρακτηριστικά για μία συγκεκριμένη πηγή και διαφορετικές τιμές της παραμέτρου  $\tau = 0.2, 0.4, 0.6, 0.8$ . Η πηγή απεικονίζεται με πράσινο χρώμα με κλίμακα και περιστροφή, ενώ τα αναμορφωμένα σημεία με κόκκινο.

Τελικά κβαντίζουμε ομοιόμορφα τα  $\bar{\rho}$  και  $\theta$  σε  $k_\rho$  και  $k_\theta$  κυτίο στο  $[0, 1]$  και  $[0, 2\pi]$  αντίστοιχα, έτσι ώστε  $k_\rho k_\theta = k_u$ . Ο κβαντισμός αυτός απεικονίζεται στα δεξιά του Σχήματος 2.2. Τα επιλεγμένα αναμορφωμένα χαρακτηριστικά για μία συγκεκριμένη πηγή και διαφορετικές τιμές της παραμέτρου  $\tau$  απεικονίζονται στο Σχήμα 2.17.

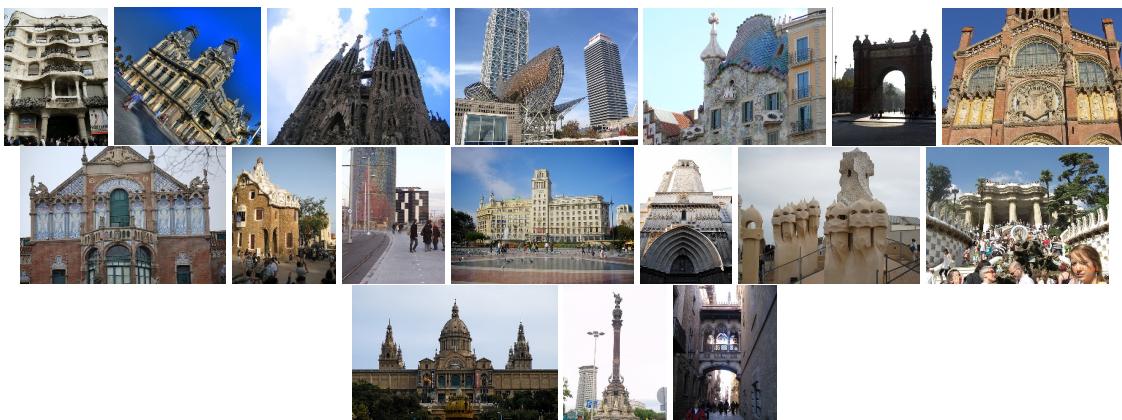
**Επιλογή χαρακτηριστικών** Εφόσον ο σκοπός μας είναι να μειώσουμε τις ανάγκες μνήμης, επιπλέον περιορίζουμε τον αριθμό των πηγών ανά εικόνα και τον αριθμό των αναμορφωμένων περιοχών ανά χάρτη σε  $n_\alpha$  και  $n_\beta$ , αντίστοιχα. Για την ακρίβεια, ταξινομούμε τις πηγές  $\alpha(X)$  με φθίνουσα σειρά με βάση τον βαθμό υποστήριξης ως πηγή (2.31) και κρατάμε τα  $n_\alpha$  πρώτα όταν  $|\alpha(X)| > n_\alpha$ . Ομοίως, για κάθε χάρτη  $\hat{x}$ , ταξινομούμε τα αναμορφωμένα χαρακτηριστικά  $\beta^{(\hat{x})}(X)$  με φθίνουσα σειρά ως προς τον βαθμό υποστήριξης ως αναμορφωμένες περιοχές (2.36), και κρατάμε τα  $n_\beta$  πρώτα όταν  $|\beta^{(\hat{x})}(X)| > n_\beta$ .

Για τις μοναδικές εικόνες, κρατάμε τα  $n_\alpha^s$  πιο ισχυρά χαρακτηριστικά ως πηγές. Τα αναμορφωμένα χαρακτηριστικά επιλέγονται με βάση τον ίδιο βαθμό υποστήριξης  $\beta_X^{(\hat{x})}(z)$  ο οποίος δίνεται από την σχέση (2.36), όμως η απόσταση  $\delta_X(z)$  στην σχέση (2.33) αντικαθίσταται από μία φθίνουσα συνάρτηση της ισχύς του ανιχνευτή χαρακτηριστικών. Για την ακρίβεια, επιλέγουμε  $\delta_X(z) = 1 / \log g(z)$ , όπου  $g(z)$  είναι η ισχύς του ανιχνευτή για το χαρακτηριστικό  $z$ . Πάλι, μόνο τα πρώτα  $n_\beta^s$  αναμορφωμένα χαρακτηριστικά διατηρούνται.

## 2.7 Πειράματα - χάρτες κι επιλογή χαρακτηριστικών

### 2.7.1 Σύνολα εικόνων

Διεξάγουμε πειράματα στην συλλογή εικόνων *European Cities 1M*<sup>4</sup>[5]. Το σύνολο ελέγχου αποτελείται από 927 εικόνες οι οποίες απεικονίζουν αξιοθέατα στην πόλη της Βαρκελώνης. Θα αναφερόμαστε σε αυτό επίσης ως σύνολο ελέγχου *Barcelona*. Δείγματα από το σύνολο ελέγχου παρουσιάζονται στο Σχήμα 2.18. Το σύνολο των εικόνων περίσπασης αποτελείται από 908.859 εικόνες κατεβασμένες από το Flickr. Επιπλέον διεξάγουμε πειράματα στην δημόσια διαθέσιμη βάση *Oxford Buildings*<sup>5</sup>.



Σχήμα 2.18: Αντιπροσωπευτικές εικόνες από όλες τις 17 ομάδες της συλλογής European Cities 1M.

### 2.7.2 Πρωτόκολλο αξιολόγησης

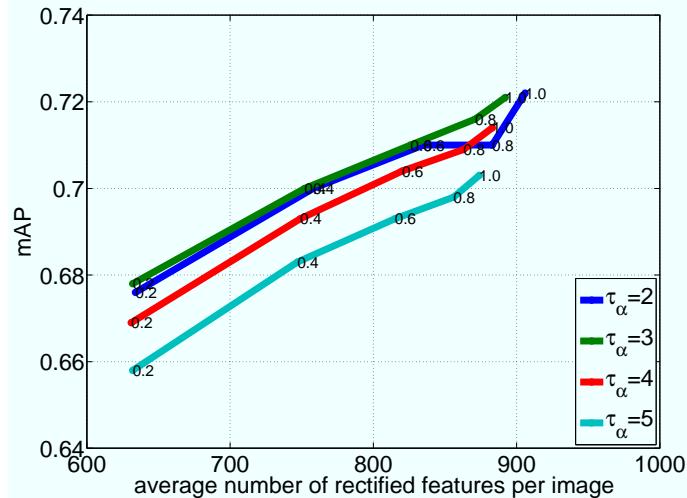
Σε όλα τα πειράματα, εξάγουμε SURF χαρακτηριστικά και περιγραφείς [8] και χρησιμοποιούμε ένα λεξικό μεγέθους  $k_v = 100K$  οπτικών λέξεων το οποίο έχει δημιουργηθεί με τον προσεγγιστικό αλγόριθμο k-means (AKM) [66] σε ένα σύνολο εικόνων το οποίο δεν είναι μέρος των συνόλων αξιολόγησης. Οι μέθοδοι με τις οποίες συγκρίνουμε είναι η bag-of-words (BoW), η weak geometric consistency (WGC) [32] και η επιλογή χρήσιμων χαρακτηριστικών (useful feature selection - UF) [82]. Συνοδεύουμε τις μεθόδους BoW και WGC από ανακατάταξη με γεωμετρική επιβεβαίωση κάνοντας χρήση της μεθόδου FastSM [66]. Η ανακατάταξη γίνεται στις 500 και 100 πρώτες εικόνες κατά την διαδικασία επιλογής χαρακτηριστικών και κατά την εκτέλεση του τελικού ερωτήματος αντίστοιχα. Εικόνες με τουλάχιστον 4 inliers θεωρούνται επιβεβαιωμένες.

Στην υλοποίηση μας για την μέθοδο UF κρατάμε τα 300 χαρακτηριστικά με την μεγαλύτερη ισχύ του ανιχνευτή, σε αντίθεση με το [82] όπου χρησιμοποιείται η κλίμακα.

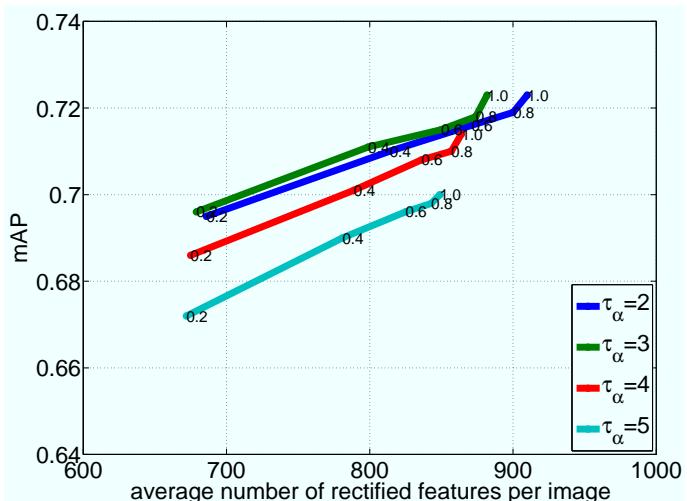
<sup>4</sup><http://image.ntua.gr/iva/datasets/ec1m/>

<sup>5</sup><http://www.robots.ox.ac.uk/~vgg/data/oxbuildings/>

Για τον χωρικό κβαντισμό στην μέθοδο FMS επιλέγουμε  $k_\rho = 4$ ,  $k_\theta = 6$  σύμφωνα με προηγούμενα πειράματα [6].



(a')  $n_\alpha = 50, n_\beta = 100$



(b')  $n_\alpha = 100, n_\beta = 50$

**Σχήμα 2.19:** Μέσος αριθμός αναμορφωμένων σημείων ανά εικόνα και απόδοση (*mAP*) στην συλλογή εικόνων European Cities 1M με 50,000 εικόνες περίσπασης. Η τιμή της παραμέτρου  $\tau$  εμφανίζεται πάνω στα σημεία της κάθε καμπύλης.

### 2.7.3 Βελτιστοποίηση παραμέτρων

Διεξάγουμε πειράματα στον σύνολο ελέγχου μαζί με εικόνες περίσπασης για να βρούμε τις πιο κατάλληλες τιμές για τις παραμέτρους του μοντέλου μας. Το να κρατάει κανείς περισσότερα γεωμετρικά επιβεβαιωμένα χαρακτηριστικά αυξάνει την απόδοση δεδομένου ότι υπάρχουν οι κατάλληλοι υπολογιστικοί πόροι, ειδικά η μνήμη λόγω των αυξημένων αναγκών των δομών δεικτοδότησης. Στην πράξη χρειαζόμαστε έναν συμβιβασμό ανάμεσα στα δύο. Η απόδοση μετρείται μέσω mean average precision (*mAP*). Οι ανάγκες μνήμης

μετρώνται μέσω του μέσου αριθμού των αναμορφωμένων χαρακτηριστικών αθροισμένα από όλους τους χάρτες μίας εικόνας. Αυτό είναι ίσο με τον μέσο αριθμό εισαγωγών ανά εικόνα στην δομή δεικτοδότησης.

Ο μέγιστος αριθμός πηγών ανά εικόνα,  $n_\alpha$ , και αναμορφωμένων χαρακτηριστικών ανά χάρτη,  $n_\beta$ , δίνουν ένα άνω όριο στις συνολικές ανάγκες μνήμης. Στην πράξη αυτό περιορίζεται από τις τιμές κατωφλίωσης στον βαθμό υποστήριξης για πηγή και αναμορφωμένο χαρακτηριστικό,  $\tau_\alpha$  και  $\tau_\beta$  αντίστοιχα. Επηρεάζεται επίσης από την παράμετρο της ισχύς τοπικότητας  $\sigma_\ell$  ή ισοδύναμα από την παράμετρο βεληνεκούς  $\tau$ , όπως ορίζονται στις Σχέσεις (2.35) και (2.40), αντίστοιχα. Έτσι ελέγχεται και η ισορροπία ανάμεσα σε τοπική και ολική γεωμετρία, ενώ η τιμή του  $\tau_\beta$  παραμένει σταθερή. Για την ακρίβεια, κρατάμε έως δύο φορές την τυπική απόκλιση με βάση την τοπικότητα θέτοντας  $\tau_\beta = e^{-2}$ , έτσι ώστε η σχέση (2.40) να γίνει  $\tau = F_{\kappa,\lambda}(2\sigma_\ell)$ . Ως αποτέλεσμα, μεταβάλουμε το  $\sigma_\ell$  αλλά μετράμε το  $\tau$  στα πειράματα μας, το οποίο είναι πιο διαισθητικό διότι  $\tau \in [0, 1]$ . Αυτή η συγκεκριμένη επιλογή για το  $\tau_\beta$  σημαίνει πως κρατάμε επίσης μέχρι δύο φορές την τυπική απόκλιση με βάση τον βαθμό υποστήριξης για inlier (2.34), όπου η παράμετρος  $\sigma_i$  έχει σταθερή τιμή ίση με 0.05.

Το Σχήμα 2.19 δείχνει τις ανάγκες μνήμης και την απόδοση στην συλλογή εικόνων European Cities 1M με ένα υποσύνολο από 50,000 εικόνες περίσπασης. Κάθε καμπύλη αντιστοιχεί σε διαφορετική τιμή της παραμέτρου  $\tau_\alpha$ . Κατά μήκος κάθε καμπύλης μεταβάλλεται το  $\tau$  από 0.2 έως 1.0. Η μέγιστη απόδοση επιτυγχάνεται για  $\tau_\alpha = 3$ . Ενδιαφέρον είναι το γεγονός ότι για  $\tau_\alpha = 2$  αυξάνεται η μνήμη αλλά πέφτει η απόδοση. Αυτή η περίπτωση αντιστοιχεί στην ευθυγράμμιση τουλάχιστον τριών χαρακτηριστικών, ένα εκ των οποίων είναι το χαρακτηριστικό πηγή. Οι πραγματικοί inliers είναι μόνο δύο, το οποίο δεν είναι αρκετά έγκυρο στοιχείο για γεωμετρική επιβεβαίωση. Εξοικονόμηση μνήμης μπορεί τώρα να γίνει χρησιμοποιώντας μεγαλύτερες τιμές για την παράμετρο  $\tau_\alpha$  ή μικρότερες για την παράμετρο  $\tau$ . Το Σχήμα 2.19 αποκαλύπτει ότι προτιμότερη είναι η τελευταία επιλογή, εφόσον επιτυγχάνεται υψηλότερη απόδοση με την ίδια μνήμη. Ως συμβιβασμό επιλέγουμε  $\tau_\alpha = 3$  και  $\tau = 0.6$ .

Έπειτα από την διεξαγωγή ενός αριθμού δοκιμών πάνω στον μέγιστο αριθμό πηγών κι αναμορφωμένων χαρακτηριστικών, δείχνουμε στο Σχήμα 2.19(α') και 2.19(β') τα αποτελέσματα για  $n_\alpha = 50$ ,  $n_\beta = 100$  και  $n_\alpha = 100$ ,  $n_\beta = 50$  αντίστοιχα. Διαλέγουμε τον δεύτερο συνδυασμό, όπως και πριν, για μεγαλύτερη απόδοση με την ίδια μνήμη. Φαίνεται πως είναι καλύτερη επιλογή το να διαλέξει κανείς περισσότερες πηγές καθώς η ύπαρξη ευθυγραμμισμένων πηγών είναι πολύ σημαντική για την μέθοδο μας, ενώ τα αναμορφωμένα σημεία είναι πιο εύκολο να ταιριάζουν. Ομοίως μετά από μερικές δοκιμές, θέτουμε τον μέγιστο αριθμό πηγών ανά εικόνα,  $n_\alpha^s$ , και των αναμορφωμένων σημείων ανά χάρτη,  $n_\beta^s$ , ίσα με 30 και 20 αντίστοιχα σε όλα τα πειράματα μας.

Method	FMS			FMH [6]
	Single	Matched	Total	Total
Images	652104	257682	909786	-
Features b.s.	483.1	562.6	505.6	-
Origins	30.0	55.4	37.2	200
Rectified features	18.5	43.2	25.5	50
Features a.s.	557.6	2401.9	1079.9	10000.0

Πίνακας 2.5: Στατιστικά στοιχεία για τις μεθόδους FMS και FMH. Οι μετρήσεις για την μέθοδο FMS έγιναν στην συλλογή European Cities 1M, ενώ για την μέθοδο FMH είναι σταθερά. Συνολικός αριθμός εικόνων, μέσος αριθμός χαρακτηριστικών ανά εικόνα πριν την επιλογή (b.s.), μέσος αριθμός πηγών ανά εικόνα, μέσος αριθμός αναμορφωμένων περιοχών ανά πηγή και μέσος αριθμός στοιχείων ανά εικόνα μετά την επιλογή (a.s.).

	Single	Matched	Total
Images	652104	257682	909786
Features b.s.	483.1	562.6	505.6
Features a.s.	270.1	70.9	213.7

Πίνακας 2.6: Στατιστικά στοιχεία για την επιλογή χαρακτηριστικών με την μέθοδο UF στην συλλογή εικόνων European Cities 1M: συνολικός αριθμός εικόνων, μέσος αριθμός χαρακτηριστικών ανά εικόνα πριν (b.s.) και μετά την επιλογή (a.s.).

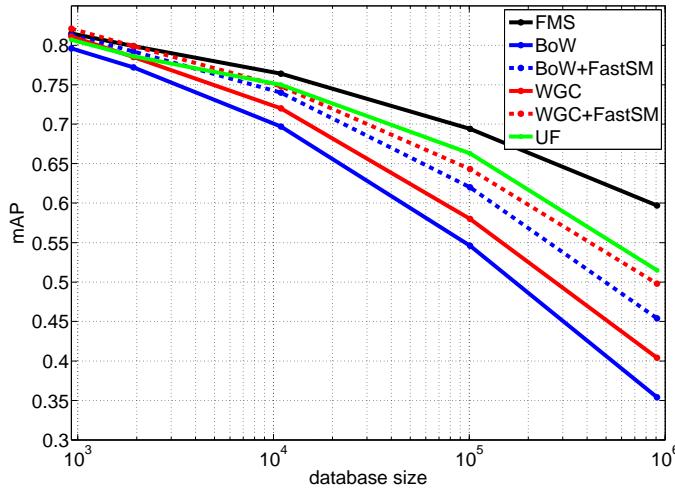
#### 2.7.4 Αποτελέσματα

Για τις τιμές των παραμέτρων που επιλέχθηκαν, ο Πίνακας 2.5 δίνει τον μέσο αριθμό χαρακτηριστικών ανά εικόνα πριν την επιλογή, των πηγών ανά εικόνα, των αναμορφωμένων περιοχών ανά χάρτη και τον συνολικό αριθμό στοιχείων (αναμορφωμένων περιοχών) ανά εικόνα. Οι μετρήσεις παρουσιάζονται ξεχωριστά για τις εικόνες που ταίριαζαν με επιτυχία, για τις μοναδικές αλλά και συνολικά. Ο αριθμός των στοιχείων ανά εικόνα, είναι και ο πραγματικός αριθμός εισαγωγών στην δομή δεικτοδότησης ανά εικόνα. Η προτεινόμενη μέθοδος με την επιλογή χαρακτηριστικών έχει μία τάξη μεγέθους λιγότερες απαιτήσεις μνήμης από την μέθοδο με την χρήση μεθόδων κατακερματισμού (FMH). Παρόμοιες μετρήσεις φαίνονται στον Πίνακα 2.6 και για την μέθοδο UF. Παρατηρούμε πως η μέθοδος μας η οποία ενσωματώνει στην δομή δεικτοδότησης γεωμετρική πληροφορία χρειάζεται μόνο δύο φορές περισσότερη μνήμη από την μέθοδο BoW, η οποία δεν έχει καθόλου γεωμετρική πληροφορία. Συγκριτικά με την μέθοδο UF, χρειάζεται περίπου 5 φορές περισσότερη μνήμη.

Στο Σχήμα 2.20 γίνεται σύγκριση της προτεινόμενης μεθόδου με την μέθοδο BoW,

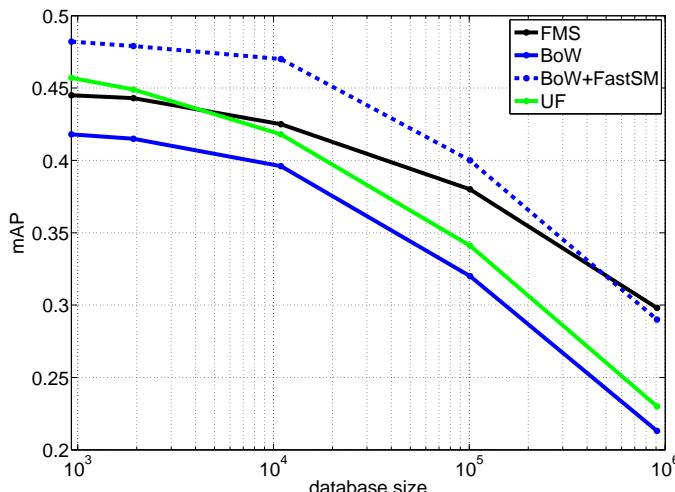
BoW	BoW+FastSM	WGC	WGC+FastSM	UF	FMS
88	1167	3221	4294	54	286

Πίνακας 2.7: Μέσος χρόνος εκτέλεσης σε ms για όλες τις μεθόδους.



**Σχήμα 2.20:** Mean average precision για μεταβλητό αριθμό εικόνων περίσπασης από την συλλογή European Cities 1M για τις μεθόδους BoW, WGC, UF και FMS. Οι BoW και WGC συνδυάζονται επίσης με ανακατάταξη των 100 πρώτων εικόνων από την μέθοδο FastSM.

την μέθοδο weak geometric consistency (WGC) [32] και την μέθοδο UF [82] στην συλλογή εικόνων European Cities 1M για μεταβλητό αριθμό από εικόνες περίσπασης. Μαζί με την χρήση των μεθόδων BoW και WGC εφαρμόζουμε και ανακατάταξη εικόνων με την χρήση της μεθόδου FastSM στις εικόνες που κατατάχθηκαν στις 100 πρώτες θέσεις. Η μέθοδος FMS υπερέχει έναντι των υπολοίπων, δείχνοντας το όφελος από την δεικτοδότηση γεωμετρίας, ειδικά σε μεγαλύτερη κλίμακα πειράματος.



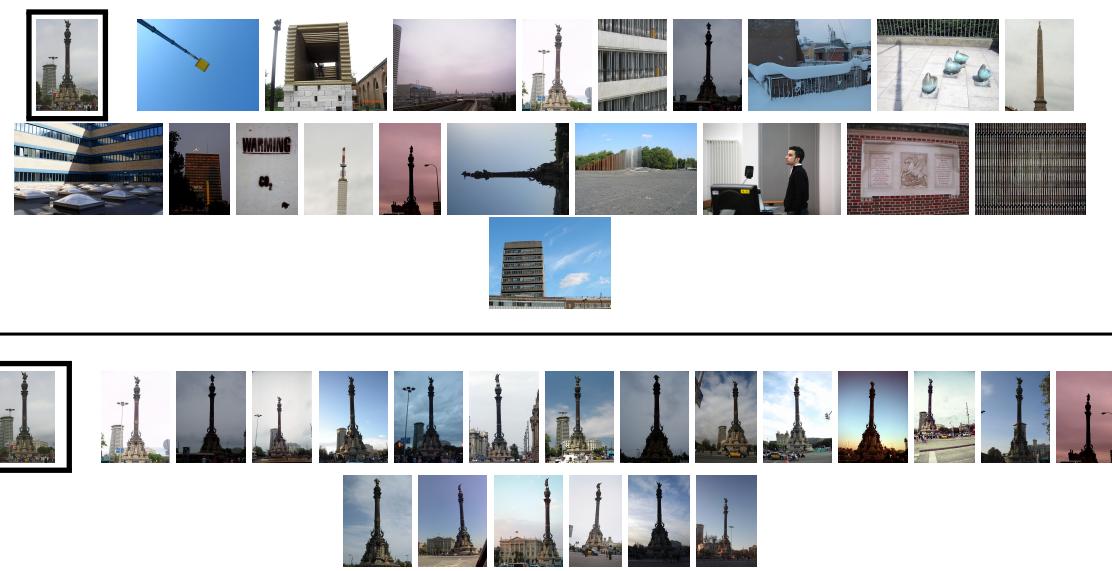
**Σχήμα 2.21:** Mean average precision για μεταβλητό μέγεθος εικόνων περίσπασης από την συλλογή Oxford Buildings για τις μεθόδους BoW, UF και FMS. Η μέθοδος BoW συνδυάζεται επίσης με ανακατάταξη των 100 πρώτων εικόνων από την μέθοδο FastSM.

Η γεωμετρική επιβεβαίωση κατά την ανακατάταξη είναι συνήθως το πιο κοστοβόρο στάδιο. Όπως παρουσιάζεται στον Πίνακα 2.7, οι χρόνοι του FMS είναι συγκρίσιμοι με

εκείνους του BoW και πολύ χαμηλότεροι από του BoW με ανακατάταξη. Όπως φαίνεται στα πειράματα μας και σε συμφωνία με το [34], η μέθοδος WGC αυξάνει την απόδοση αλλά επίσης αυξάνει δραματικά τον χρόνο εκτέλεσης. Η μέθοδος WGC μαζί με ανακατάταξη έχει απαγορευτικούς χρόνους εκτέλεσης. Η μέθοδος UF, επιλέγει τα χαρακτηριστικά με βάση γεωμετρικά κριτήρια και καταφέρνει να έχει υψηλή απόδοση σε χαμηλούς χρόνους. Η μέθοδος μας υπερέχει της UF με το να ενσωματώνει γεωμετρία όχι μόνο στο στάδιο της επιλογής χαρακτηριστικών αλλά και της δεικτοδότησης.

Τα ίδια πειράματα διεξήχθησαν και στην δημόσια διαθέσιμη βάση εικόνων *Oxford Buildings* ως σύνολο ελέγχου, με τις εικόνες περίσπασης της βάσης European Cities 1M. Τα αποτελέσματα παρουσιάζονται στο Σχήμα 2.21. Σε αυτή την περίπτωση το όφελος από την δεικτοδότηση γεωμετρίας φαίνεται σε μεγαλύτερη κλίμακα, καθώς προσθέτουμε όλο και περισσότερες εικόνες περίσπασης.

Τέλος, παρουσιάζουμε παραδείγματα ανάκτησης για μία συγκεκριμένη εικόνα ερωτήματος στο Σχήμα 2.22 για τις μεθόδους BoW και FMS. Πολλές πραγματικές εικόνες χάνονται από την μέθοδο BoW, ενώ φαίνεται να έρχονται στις πρώτες θέσεις με την μέθοδο FMS.



**Σχήμα 2.22:** Δείγμα εικόνας ερωτήματος (μαύρο πλαίσιο) και οι ανακτημένες εικόνες. Αποτελέσματα για την μέθοδο BoW (πάνω) και FMS (κάτω).



# Κεφάλαιο 3

## Πυραμίδες Hough

### 3.1 Συνεισφορά και σχετική βιβλιογραφία

Δεδομένου ενός αριθμού αντιστοιχιών ανάμεσα σε ένα ζευγάρι εικόνων, ο αλγόριθμος RANSAC [26], μαζί με αρκετές παραλλαγές του, είναι ακόμα ο πιο δημοφιλής τρόπος για γεωμετρική επιβεβαίωση. Χρησιμοποιεί σαν στοιχείο επιβεβαίωσης τον αριθμό των inliers σε ένα γεωμετρικό μοντέλο (π.χ. ομογραφία, αφινικός μετασχηματισμός, ομοιότητα). Οι υποθέσεις παράγονται από τυχαία σύνολα αντιστοιχιών, το πλήθος των οποίων εξαρτάται από την πολυπλοκότητα του μοντέλου. Παρόλα αυτά, η απόδοση του είναι χαμηλή όταν ο λόγος των inliers προς όλες τις αντιστοιχίες είναι πολύ χαμηλός. Οι Philbin *et al.* [66] παράγουν τις υποθέσεις από μονές αντιστοιχίες εκμεταλλευόμενοι το σχήμα των τοπικών χαρακτηριστικών. Το ταίριασμα των εικόνων τότε γίνεται ντετερμινιστικό με το να απαριθμηθούν όλες οι υποθέσεις. Η διαδικασία αυτή όμως παραμένει με πολυπλοκότητα τετραγωνική ως προς τον αριθμό των αντιστοιχιών.

Οι ομάδες από σύμφωνες αντιστοιχίες μπορούν πρώτα να βρεθούν στον χώρο των μετασχηματισμών με τον γενικευμένο μετασχηματισμό Hough (generalized Hough transform) [7]. Αυτό κάνει ο Lowe [53], αλλά μόνο σαν βήμα πριν την επιβεβαίωση. Οι αρχικές αντιστοιχίες βρίσκονται με βάση την γειτνίαση στον χώρο των περιγραφέων και χρησιμοποιούνται για να παράγουν ψήφους στον χώρο των μετασχηματισμών. Οι αντιστοιχίες είναι μονές λόγω του τοπικού σχήματος, ακριβώς όπως στο [66]. Η προβολή των αντιστοιχιών στα Hough κυτία είναι γραμμική ως προς το πλήθος τους και η απόδοση εξαρτάται από τον αριθμό των inliers και όχι από τον λόγο τους. Παρόλα αυτά, πολλαπλές ομάδες αντιστοιχιών πρέπει να επιβεβαιωθούν για inliers και αυτή η διαδικασία μπορεί να έχει τετραγωνική πολυπλοκότητα στην χειρότερη περίπτωση.

Οι Jégou *et al.* χρησιμοποιούν πιο ασθενή γεωμετρικά μοντέλα [34] όπου οι ομάδες αντιστοιχιών συμφωνούν μόνο στην σχετική τους κλίμακα και περιστροφή. Οι αντιστοιχίες βρίσκονται με οπτικό λεξικό. Η κλίμακα και η περιστροφή των τοπικών χαρακτηρι-

στικών κβαντίζονται και αποθηκεύονται σε ανάστροφο αρχείο. Για αυτό τον λόγο, ασθενείς γεωμετρική περιορισμοί ενοποιούνται στην διαδικασία φιλτραρίσματος σε μία μηχανή αναζήτησης εικόνων. Παρόλα αυτά, το μοντέλο αυτό δεν απαλλάσσεται από την ανακατάταξη με βάση την γεωμετρία.

Περισσότερο εύκαμπτα μοντέλα χρησιμοποιούνται τυπικά στην αναγνώριση αντικειμένων. Για παράδειγμα, πολλαπλές ομάδες αντιστοιχιών που συμφωνούν μεταξύ τους αναγνωρίζονται με το εύκαμπτο, ημι-τοπικό μοντέλο των Carneiro και Jepson [12], επιστρατεύοντας δυαδικές σχέσεις μεταξύ των αντιστοιχιών κι επιτρέποντας μη-άκαμπτες παραμορφώσεις. Όμοια οι Leordeanu και Hebert [50] δημιουργούν έναν αραιό πίνακα γειτνίασης των αντιστοιχιών και άπληστα ανακτούν τους inliers με βάση το κύριο ιδιοδιάνυσμα του. Αυτή η φασματική μέθοδος μπορεί επιπλέον να ενσωματώσει διαφορετικούς περιορισμούς αντιστοιχιών, όπως το 1-1.

Το σχήμα αντιστοιχίας 1-1 είναι μάλλον απομεινάρι των πρώτων μεθόδων σε αντιστοιχίες πάνω σε χαρακτηριστικά χωρίς διακριτική ικανότητα, αλλά μπορεί να είναι πολύ σημαντικό όταν τα λεξικά είναι μικρά, κάτω από την παρουσία επαναλαμβανόμενων δόμων, ή π.χ. με τα μοντέλα χαλαρής αντιστοιχίσης (soft assignment) όπως οι Philbin *et al.* [69]. Πιο πρόσφατα, οι Enqvist *et al.* [25] σχηματίζουν έναν γράφο με τις αντιστοιχίες σαν κόμβους και τις ασυμβατότητες ως ακμές. Το σχήμα αντιστοιχίας 1-1 ενσωματώνεται εύκολα με το να έχουμε τις πολλαπλές αντιστοιχίες ενός χαρακτηριστικού να σχηματίζουν μία κλίκα πάνω στον γράφο. Τα περισσότερα εύκαμπτα μοντέλα είναι επαναληπτικά και τουλάχιστον τετραγωνικής πολυπλοκότητας ως προς τον αριθμό των αντιστοιχιών.

Χαλαρές διαδικασίες ταιριάσματος όπως των Vedaldi και Soatto [86] προσφέρουν μία πολύ ελκυστική εναλλακτική πρόταση όσον αφορά την πολυπλοκότητα με το να χρησιμοποιούν κατανομές πάνω σε *ιεραρχικές καταταμήσεις* αντί για δυαδικούς υπολογισμούς. Το πλέον γνωστό είναι από τους Grauman και Darell [30], οι οποίοι προβάλουν τα χαρακτηριστικά σε ιστογράμματα πολλαπλών αναλύσεων στον χώρο των περιγραφέων, κι έπειτα γίνεται ταίριασμα σε μία από κάτω προς τα πάνω διαδικασία. Το όφελος προέρχεται κυρίως από την προσέγγιση της ομοιότητας με βάση το μέγεθος των κυτίων. Οι Lazebnik *et al.* [49] εφαρμόζουν την ίδια ιδέα στον 2Δ χώρο της εικόνας αλλά με τέτοιον τρόπο που το αναλλοίωτο ως προς τον γεωμετρικούς μετασχηματισμούς χάνεται.

**Συνεισφορά.** Ενώ οι παραπάνω χαλαρές μέθοδοι εφαρμόζονται σε δύο σύνολα χαρακτηριστικών, αντίθετα εμείς εφαρμόζουμε την ίδια ιδέα σε ένα σύνολο αντιστοιχιών (ζευγάρια χαρακτηριστικών) με σκοπό την *ομαδοποίηση* σύμφωνα με την εγγύτητα τους. Αυτό το πρόβλημα θυμίζει το *mode seeking* [85], αλλά η λύση μας είναι μία μη-επαναληπτική διαδικασία ομαδοποίησης που δεν σχετίζεται με καμία παράμετρο κλίμακας. Αναπαριστούμε τις αντιστοιχίες στον χώρο των μετασχηματισμών χρησιμοποιώντας στο σχήμα των τοπικών χαρακτηριστικών όπως και στο [53], αλλά κατασκευάζουμε τις αντιστοιχίες

με χρήση οπτικού λεξικού, όπως στα [66][34]. Όπως και στην μέθοδο *pyramid match* [30], προσεγγίζουμε την ομοιότητα με το μέγεθος των κυτίων, χωρίς να απαριθμούμε ποτέ όλα τα ζευγάρια αντιστοιχιών.

Επιπλέον εφαρμόζουμε το σχήμα αντιστοιχιών 1-1, έτσι ώστε κάθε χαρακτηριστικό της μίας εικόνας να ταιριάζει το πολύ με ένα χαρακτηριστικό στην άλλη εικόνα. Αυτό κάνει το πρόβλημα μας όμοιο με εκείνο του [50], με την έννοια ότι επιλέγουμε με άπληστο τρόπο ένα δυαδικά συμβατό υποσύνολο αντιστοιχιών, το οποίο μεγιστοποιεί έναν μη-αρνητικό, συμμετρικό πίνακα ομοιοτήτων. Παρόλα αυτά, επιτρέπουμε πολλαπλές ομάδες αντιστοιχιών. Αντίθετα με τα [25][50], το μοντέλο μας είναι μη-επαναληπτικό και γραμμικό ως προς τον αριθμό των αντιστοιχιών.

Συνοψίζοντας την συνεισφορά μας, προτείνουμε ένα σχήμα εύκαμπτου γεωμετρικού ταιριάσματος όπου όλες οι αρχικές αντιστοιχίες συνεισφέρουν, με ανάλογο βάρος, στην τελική ομοιότητα δύο εικόνων. Αυτό που είναι πιο αξιοσημείωτο είναι ότι δεν χρειάζεται επιβεβαίωση, εύρεση καλύτερου μετασχηματισμού ή καταμέτρηση των inlier, όπως στα [53], [66] or [12]. Πέρα από το σημαντικό κέρδος στην απόδοση, υπάρχει και μία εξαιρετική μείωση του χρόνου εκτέλεσης. Το σχήμα 1-1 και το χαλαρό ταίριασμα βελτιώνουν την απόδοση. Το αποτέλεσμα είναι ένας απλός αλγόριθμος ο οποίος πολύ εύκολα ενσωματώνεται σε μία διαδικασία ανάκτησης εικόνων.

## 3.2 Ορισμός προβλήματος

Υποθέτουμε ότι μία εικόνα αναπαρίσταται από ένα σύνολο τοπικών χαρακτηριστικών  $P$ , και ότι για κάθε χαρακτηριστικό  $p \in P$  έχουμε τον περιγραφέα του, την θέση του και τοπικό σχήμα της περιοχής του. Περιορίζουμε την συζήτηση σε χαρακτηριστικά αναλλοίωτα ως προς την κλίμακα και την περιστροφή. έτσι ώστε το σχήμα και η θέση της περιοχής του  $p$  να δίνονται από τον  $3 \times 3$  πίνακα

$$F(p) = \begin{bmatrix} M(p) & \mathbf{t}(p) \\ \mathbf{0}^T & 1 \end{bmatrix}, \quad (3.1)$$

όπου  $M(p) = \sigma(p)R(p)$  και  $\sigma(p), R(p), \mathbf{t}(p)$  είναι η ισοτροπική κλίμακα, η περιστροφή και η θέση αντίστοιχα.  $R(p)$  είναι ένας ορθογώνιος  $2 \times 2$  πίνακας με  $\det R(p) = 1$ , ο οποίος αναπαρίσταται από την γωνία  $\theta(p)$ . Σαν αποτέλεσμα, ο  $F(p)$  ορίζει έναν μετασχηματισμό ομοιότητας με βάση μία κανονικοποιημένη περιοχή, π.χ. κεντραρισμένη στην αρχή των αξόνων, με κλίμακα  $\sigma_0 = 1$  και περιστροφή  $\theta_0 = 0$ .

Δεδομένων δύο εικόνων  $P, Q$ , μία αντιστοιχία  $c = (p, q)$  είναι ένα ζευγάρι χαρακτηριστικών  $p \in P, q \in Q$ . Ο σχετικός μετασχηματισμός από το  $p$  στο  $q$  είναι και πάλι ένας

μετασχηματισμός ομοιότητας, ο οποίος δίνεται από

$$F(c) = F(q)F(p)^{-1} = \begin{bmatrix} M(c) & \mathbf{t}(c) \\ \mathbf{0}^T & 1 \end{bmatrix}, \quad (3.2)$$

όπου  $M(c) = \sigma(c)R(c)$  και  $\mathbf{t}(c) = \mathbf{t}(q) - M(c)\mathbf{t}(p)$ . Επίσης  $\sigma(c) = \sigma(q)/\sigma(p)$  και  $R(c) = R(q)R(p)^{-1}$  είναι η σχετική κλίμακα και περιστροφή αντίστοιχα από το  $p$  στο  $q$ . Αυτός είναι ένας μετασχηματισμός με 4 βαθμούς ελευθερίας κι αναπαρίσταται από το διάνυσμα παραμέτρων

$$f(c) = (x(c), y(c), \sigma(c), \theta(c)), \quad (3.3)$$

όπου  $[x(c) \ y(c)]^T = \mathbf{t}(c)$  and  $\theta(c) = \theta(q) - \theta(p)$ . Επομένως οι αντιστοιχίες μπορούν να είναι αντιληπτές ως σημεία σε έναν  $d$ -διάστατο χώρο μετασχηματισμών  $\mathcal{F}$ ,  $d = 4$  στην περίπτωση μας, ενώ για αφινικά αναλλοίωτες περιοχές θα ήταν  $d = 6$ .

Ένα σύνολο υποψήφιων ή αρχικών αντιστοιχιών  $C$  κατασκευάζεται σύμφωνα με την εγγύτητα των χαρακτηριστικών στον χώρο των περιγραφέων. Υπάρχουν διαφορετικά κριτήρια, όπως η αναζήτηση πλησιέστερων γειτόνων ή τα οπτικά λεξικά. Εμείς θεωρούμε την απλούστερη, του οπτικού λεξικού, όπου δύο χαρακτηριστικά είναι σε αντιστοιχία όταν τους έχει επισυναφθεί η ίδια οπτική λέξη:

$$C = \{(p, q) \in P \times Q : u(p) = u(q)\}, \quad (3.4)$$

όπου  $u(p)$  είναι η οπτική λέξη του  $p$ . Αυτό είναι ένα σχήμα αντιστοιχίας πολλά με πολλά, όπου κάθε χαρακτηριστικό στο  $P$  μπορεί να έχει πολλαπλές αντιστοιχίες σε χαρακτηριστικά του  $Q$ , κι αντίστροφα. Δεδομένης της αντιστοιχίας  $c = (p, q)$ , ορίζουμε την οπτική της λέξη  $u(c)$  ως την κοινή οπτική λέξη  $u(p) = u(q)$ .

Κάθε αντιστοιχία  $c = (p, q) \in C$  έχει μία ισχύ  $w(c)$  η οποία μετρά την σχετική της σημασία. Χρησιμοποιούμε την αντίστροφη συχνότητα εγγράφου (inverse document frequency - *idf*) της οπτικής λέξης. Δεδομένου ενός ζευγαριού αντιστοιχιών  $c, c' \in C$ , ορίζουμε έναν βαθμό ομοιότητας  $\alpha(c, c')$  ο οποίος προκύπτει από μία γνησίως φθίνουσα συνάρτηση της απόστασης τους στον χώρο των μετασχηματισμών. Θέλουμε να ομαδοποιήσουμε τις αντιστοιχίες σύμφωνα με την ομοιότητα τους, χωρίς όμως ποτέ να απαριθμήσουμε όλα τα ζευγάρια. Τέλος λέμε ότι δύο αντιστοιχίες  $c = (p, q)$ ,  $c' = (p', q')$  είναι συμβατές εάν  $p \neq p'$  και  $q \neq q'$ , αλλιώς ασύμβατες. Για παράδειγμα, οι  $c, c'$  είναι ασύμβατες εάν προβάλουν δύο χαρακτηριστικά του  $P$  στο ίδιο χαρακτηριστικό του  $Q$ .

Το πρόβλημα μας είναι να αναγνωρίσουμε το υποσύνολο των δυαδικά συμβατών αντιστοιχιών οι οποίες μεγιστοποιούν την τελική σταθμισμένη ομοιότητα.

### 3.3 Ταιριασμα πυραμίδας Hough

Υποθέτουμε ότι οι παράμετροι ενός μετασχηματισμού μπορούν να κανονικοποιηθούν στο  $[0, 1]$  (βλέπε ενότητα 3.5). Επομένως ο χώρος των μετασχηματισμών είναι  $\mathcal{F} = [0, 1]^d$ . Ενώ διατυπώσαμε το πρόβλημα μας με  $d = 4$ , η διαδικασία του ταιριάσματος μπορεί να εφαρμοστεί σε χώρο οσονδήποτε διαστάσεων (βαθμών ελευθερίας).

Κατασκευάζουμε μία ιεραρχική κατάτμηση  $\mathcal{B} = \{B_0, \dots, B_{L-1}\}$  του  $\mathcal{F}$  σε  $L$  επίπεδα. Κάθε  $B_\ell \in \mathcal{B}$  κατατεμαχίζει το  $\mathcal{F}$  σε  $2^{kd}$  κυτία (υπερκύβους), όπου  $k = L - 1 - \ell$ . Τα κυτία δημιουργούνται με ομοιόμορφο κβαντισμό της κάθε παραμέτρου, ή με κατάτμηση της κάθε διάστασης σε  $2^k$  ίσα διαστήματα μήκους  $2^{-k}$ . Το  $B_0$  είναι το λεπτομερές (κατώτερο) επίπεδο, ενώ το  $B_{L-1}$  είναι το χοντροκομμένο (ανώτερο) επίπεδο κι έχει μόνο ένα κυτίο. Η κάθε κατάτμηση  $B_\ell$  είναι ένα ραφινάρισμα του  $B_{\ell+1}$ . Αντίθετα, το κάθε κυτίο  $B_\ell$  είναι η ένωση  $2^\ell$  κυτίων του  $B_{\ell-1}$ .

Αρχίζοντας με το σύνολο  $C$  των υποψήφιων αντιστοιχιών των εικόνων  $P, Q$ , διανέμουμε τις αντιστοιχίες στα κυτία του ιστογράμματος πυραμίδας. Δεδομένου ενός κυτίου  $b$ , έστω

$$h(b) = \{c \in C : f(c) \in b\} \quad (3.5)$$

το σύνολο των αντιστοιχιών με διανύσματα παραμέτρων τα οποία πέφτουν στο  $b$ , και  $|h(b)|$  το πλήθος τους. Χρησιμοποιούμε αυτό το πλήθος για να προσεγγίσουμε τις ομοιότητες μέσα στα κυτία της ιεραρχίας, παίρνοντας άπληστες αποφάσεις για τις ασυμβατότητες, για να υπολογίσουμε την τελική ομοιότητα ανάμεσα στα  $P, Q$ . Αυτό ο υπολογισμός είναι γραμμικού χρόνου ως προς τον αριθμό των υποψήφιων αντιστοιχιών  $n = |C|$ .

#### 3.3.1 Διαδικασία ταιριάσματος

Αναδρομικά διαχωρίζουμε τις αντιστοιχίες στα κυτία σε μια διαδικασία από πάνω προς τα κάτω, κι έπειτα τα ομαδοποιούμε πάλι σε μια διαδικασία από κάτω προς τα πάνω. Αναμένουμε να βρούμε τις ομάδες των πιο συμβατών αντιστοιχιών στο κατώτερο επίπεδο, αλλά πάμε μέχρι το ανώτερο επίπεδο για λόγους ευκαμψίας-ευελιξίας. Μεγάλες ομάδες αντιστοιχιών οι οποίες σχηματίστηκαν στα κατώτερα επίπεδα είναι πιο πιθανόν να είναι πραγματικές, δηλαδή *inliers*. Αντίθετα, απομονωμένες αντιστοιχίες ή ομάδες που σχηματίζονται στα ανώτερα επίπεδα αναμένουμε να είναι λανθασμένες, δηλαδή *outliers*. Φαίνεται πως κάθε αντιστοιχία θα πρέπει να συνεισφέρει στον τελικό βαθμό ομοιότητας σύμφωνα με το μέγεθος των ομάδων στις οποίες συμμετέχει και το επίπεδο στο οποίο αυτές σχηματίστηκαν. Χρησιμοποιούμε το πλήθος των αντιστοιχιών ενός κυτίου για να εκτιμήσουμε το μέγεθος της ομάδας του, και το επίπεδο του για να εκτιμήσουμε την ομοιότητα ανάμεσα σε ζευγάρια της ομάδας: πράγματι, το μέγεθος του κυτίου (επομένως και οι αποστάσεις

μέσα σε ένα κυτίο) αυξάνονται όσο αυξάνεται το επίπεδο, ενώ η ομοιότητα φθίνει.

Προκειμένου να εφαρμόσουμε περιορισμούς 1-1, εντοπίζουμε τις μη συμβατές αντιστοιχίες σε κάθε επίπεδο και επιλέγουμε με άπληστο τρόπο την καλύτερη για να κρατήσουμε. Οι υπόλοιπες σημαδεύονται σαν διαγραμμένες. Έστω ότι  $X_\ell$  συμβολίζει το σύνολο των διαγραμμένων αντιστοιχιών μέχρι το επίπεδο  $\ell$ . Εάν  $b \in B_\ell$  είναι ένα κυτίο στο επίπεδο  $\ell$ , τότε το σύνολο των αντιστοιχιών που έχουμε κρατήσει στο  $b$  είναι  $\hat{h}(b) = h(b) \setminus X_\ell$ . Προφανώς, μία μεμονωμένη αντιστοιχία σε ένα κυτίο δεν δημιουργεί ομάδα, ενώ κάθε αντιστοιχία συνδέεται με  $m - 1$  αντιστοιχίες σε μία ομάδα από  $m$  αντιστοιχίες, για  $m > 1$ . Για αυτό τον λόγο, ορίζουμε ως το πλήθος της ομάδα ενός κυτίου  $b$  ως

$$g(b) = \max\{0, |\hat{h}(b)| - 1\}. \quad (3.6)$$

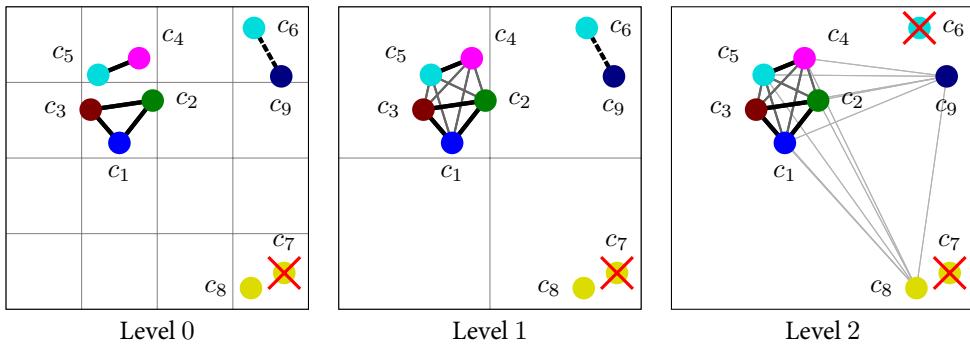
Έστω ότι  $b_0 \subseteq \dots \subseteq b_\ell$  είναι η ακολουθία των κυτίων τα οποία περιλαμβάνουν την αντιστοιχία  $c$  σε διαδοχικά επίπεδα, μέχρι το επίπεδο  $\ell$  έτσι ώστε  $b_k \in B_k$  για  $k = 0, \dots, \ell$ . Για κάθε  $k$ , προσεγγίζουμε την ομοιότητα  $\alpha(c, c')$  της  $c$  με κάθε άλλη αντιστοιχία  $c' \in b_k$  με μία σταθερή ποσότητα. Αυτή η ποσότητα είναι μια μη-αρνητική, γνησίως φθίνουσα συνάρτηση του  $k$ , έστω  $\alpha(k)$ . Εστιάζουμε εδώ στην μορφή  $\alpha(k) = 2^{-k}$ . Από την άλλη μεριά, υπάρχουν  $g(b_k) - g(b_{k-1})$  νέες αντιστοιχίες που συνδέονται με την  $c$  σε μία ομάδα του επιπέδου  $k$ . Ομοίως με το [30], προκύπτει η ισχύς του  $c$  μέχρι το επίπεδο  $\ell$ :

$$s_\ell(c) = g(b_0) + \sum_{k=1}^{\ell} 2^{-k} \{g(b_k) - g(b_{k-1})\}. \quad (3.7)$$

Είμαστε τώρα σε θέση να ορίσουμε τον βαθμό ομοιότητας ανάμεσα στις εικόνες  $P, Q$ . Πράγματι, η συνολική ισχύς μίας αντιστοιχίας  $c$  είναι απλά η ισχύς της στο ανώτερο επίπεδο,  $s(c) = s_{\ell-1}(c)$ . Έπειτα, εξαιρώντας όλες τις διαγραμμένες αντιστοιχίες  $X = X_{\ell-1}$ , ορίζουμε τον τελικό βαθμό ομοιότητας από το άθροισμα

$$s(C) = \sum_{c \in C \setminus X} w(c)s(c). \quad (3.8)$$

Από την άλλη μεριά, είμαστε σε θέση να επιλέξουμε την καλύτερη αντιστοιχία σε περίπτωση μη συμφωνίας και να επιβάλλουμε αντιστοιχία 1-1. Συγκεκριμένα, έστω ότι  $c = (p, q), c' = (p', q')$  είναι δύο αντιστοιχίες που δεν συμφωνούν. Εξ ορισμού (3.4), όλα τα τέσσερα χαρακτηριστικά  $p, p', q, q'$  μοιράζονται την ίδια οπτική λέξη, έτσι οι  $c, c'$  έχουν την ίδια ισχύ με βάση τις οπτικές λέξεις:  $w(c) = w(c')$ . Τώρα, έστω ότι  $b \in B_\ell$  είναι το πρώτο (κατώτερο) κυτίο στην ιεραρχία όπου  $c, c' \in b$ . Συνεπώς οι (3.7) και (3.8) σημαίνουν ότι η συνεισφορά τους στον τελικό βαθμό ομοιότητας μπορεί να διαφέρει μόνο μέχρι το επίπεδο  $\ell - 1$ . Επομένως επιλέγουμε το ισχυρότερο μέχρι το επίπεδο  $\ell - 1$  σύμφωνα με την ισχύ (3.7). Σε περίπτωση ισοβαθμίας, διαλέγουμε ένα τυχαία.



**Σχήμα 3.1:** Ταίριασμα 9 αντιστοιχιών σε μία πυραμίδα τριών επιπέδων στον 2Δ χώρο. Τα χρώματα συμβολίζουν τις διαφορετικές οπτικές λέξεις, και τα πάχη των ακμών τον βαθμό ομοιότητας. Οι διακεκομμένες γραμμές μεταξύ των  $c_6$ ,  $c_9$  δηλώνουν μία ομάδα που σχηματίστηκε στο επίπεδο 0 και μετά διαλύθηκε στο επίπεδο 2, εφόσον το  $c_6$  διαγράφηκε.

	$p$	$q$	similarity score
$c_1$			$(2 + \frac{1}{2}2 + \frac{1}{4}2)w(c_1)$
$c_2$			$(2 + \frac{1}{2}2 + \frac{1}{4}2)w(c_2)$
$c_3$			$(2 + \frac{1}{2}2 + \frac{1}{4}2)w(c_3)$
$c_4$			$(1 + \frac{1}{2}3 + \frac{1}{4}2)w(c_4)$
$c_5$			$(1 + \frac{1}{2}3 + \frac{1}{4}2)w(c_5)$
$c_6$			0
$c_7$			0
$c_8$			$\frac{1}{4}6w(c_8)$
$c_9$			$\frac{1}{4}6w(c_9)$

**Σχήμα 3.2:** Αντιστοιχίες και βαθμοί ομοιότητας αναφερόμενα στο Σχήμα 3.1. Εδώ οι κόμβοι και οι ακμές δηλώνουν χαρακτηριστικά (στις εικόνες  $P, Q$ ) και αντιστοιχίες, αντίστοιχα. Οι αντιστοιχίες  $c_5, c_6$  είναι ασύμβατες, όντας της μορφής  $(p, q), (p, q')$ . Ομοίως για τις  $c_7, c_8$ . Οι αντιστοιχίες  $c_1, \dots, c_5$  σχηματίζουν ομάδες στο επίπεδο 0, οι  $c_8, c_9$  στο επίπεδο 2 και οι  $c_6, c_7$  διαγράφονται.

### 3.3.2 Παραδείγματα και συζήτηση

Ένα συνθετικό παράδειγμα της διαδικασίας ταιριάσματος με τον προτεινόμενο αλγόριθμο απεικονίζεται στα Σχήματα 3.1, 3.2, 3.3. Υποθέτουμε ότι οι αντιστοιχίες είναι σε ασυμφωνία όταν μοιράζονται την ίδια οπτική λέξη, όπως αυτή ορίζεται από το χρώμα. Όπως φαίνεται στο Σχήμα 3.1, τρεις ομάδες αντιστοιχιών σχηματίζονται στο επίπεδο 0:  $\{c_1, c_2, c_3\}$ ,  $\{c_4, c_5\}$  και  $\{c_6, c_9\}$ . Οι πρώτες δύο ενώνονται στο επίπεδο 1. Οι αντιστοιχίες  $c_7, c_8$  είναι σε ασυμφωνία, και η  $c_7$  διαγράφεται τυχαία. Οι αντιστοιχίες  $c_5, c_6$  είναι επίσης σε ασυμφωνία, αλλά συγκρίνονται μόνο στο επίπεδο 2 όπου και πέφτουν στο ίδιο κυτίο. Σύμφωνα με την σχέση (3.7), το  $c_5$  είναι πιο δυνατό εφόσον συμμετέχει σε ομάδα των 5. Επομένως η ομάδα  $\{c_6, c_9\}$  διασπάται, το  $c_6$  διαγράφεται και τελικά τα  $c_8, c_9$  ενώνονται με τα  $c_1, \dots, c_5$ .

σε μία ομάδα των 7 στο επίπεδο 2.

Εκτός από την διάταξη χαρακτηριστικών/αντιστοιχιών στις εικόνες  $P, Q$ , το Σχήμα 3.2 απεικονίζει επίσης πως σχηματίζεται ο βαθμός ομοιότητας της σχέσης (3.8). Για παράδειγμα, οι αντιστοιχίες  $c_1, \dots, c_5$  έχουν συνεισφορές κι από τα 3 επίπεδα, ενώ οι  $c_8, c_9$  μόνο από το επίπεδο 2. Το Σχήμα 3.3 δείχνει πως αυτές οι συνεισφορές ορίζονται σε έναν θεωρητικό πίνακα ομοιοτήτων  $A$  διάστασης  $n \times n$ . Στην πραγματικότητα, το άθροισμα σε μία γραμμή του  $A$  ισοδυναμεί με την ισχύ της συγκεκριμένης αντιστοιχίας, σημειωτέον ότι η διαγώνιος αγνοείται λόγω της σχέσης (3.6).

Παρατηρούμε ότι το άνω τριγωνικό μέρος του  $A$ , υπεύθυνο για τον μισό βαθμό ομοιότητας της σχέσης (3.8), αντιστοιχεί στο σύνολο των ακμών μεταξύ των αντιστοιχιών που φαίνονται στο Σχήμα 3.1, όπου το βάρος της ακμής είναι ανάλογο της ομοιότητας. Αυτό αποκαλύπτει την δυαδική φύση της μεθόδου [12][50] και το γεγονός ότι μία μοναδική αντιστοιχία δεν σχηματίζει ομάδα από μόνη της.

Ένα άλλο παράδειγμα είναι αυτό του Σχήματος 3.5, όπου ταιριάζουμε δύο πραγματικές εικόνες της ίδια σκηνής από διαφορετικές οπτικές γωνίες. Είναι εμφανές ότι οι ισχυρότερες αντιστοιχίες, οι οποίες συνεισφέρουν και περισσότερο στον βαθμό ομοιότητας, είναι οι πραγματικοί inliers. Η γεωμετρία της σκηνής είναι τέτοια ώστε ούτε ένας μετασχηματισμός ομογραφίας δεν μπορεί να αποτυπώσει όλες τις ορατές επιφάνειες. Πράγματι οι inliers σε ένα αφινικό μοντέλο με το RANSAC είναι μόνο ένα μικρό ποσοστό αυτών που φαίνονται εδώ.

Σε αναλογία με το συνθετικό παράδειγμα του Σχήματος 3.1, το Σχήμα 3.4 απεικονίζει το ταίριασμα των αντιστοιχιών στον χώρο *Hough*. Παρατηρούμε πως οι αντιστοιχίες γίνονται πιο δυνατές και συνεισφέρουν πιο πολύ σύμφωνα με την εγγύτητα τους με άλλες.

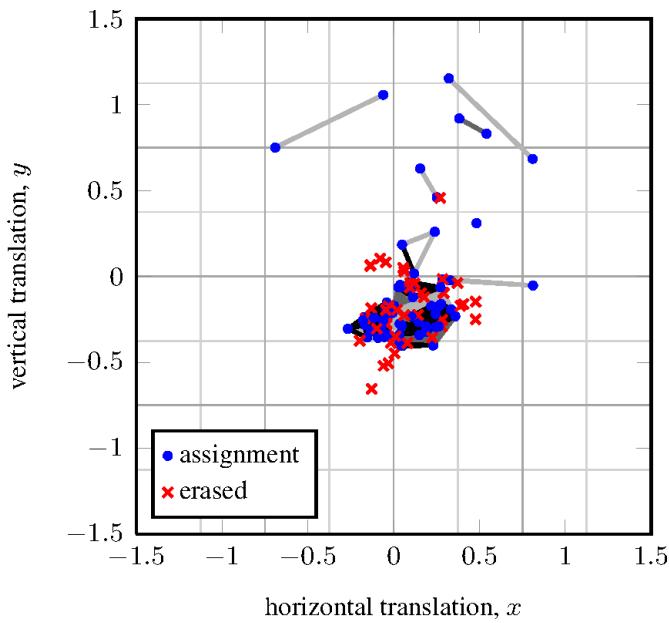
### 3.3.3 Ο αλγόριθμος

Η διαδικασία ταιριάσματος περιγράφεται πιο μεθοδικά στον αλγόριθμο 1. Προκύπτει μία αναδρομική υλοποίηση: ο κώδικας πριν την αναδρομική κλήση της γραμμής 5 σχετίζεται με την διαδικασία διαχωρισμού από πάνω προς τα κάτω, ενώ μετά από αυτή, με την διαδικασία ένωσης από κάτω προς τα πάνω. Για συντομία, μεταβλητές ή συναρτήσεις που χρησιμοποιούνται χωρίς ορισμό θεωρούνται καθολικές (global). Για παράδειγμα τα βάρη με βάση τις οπτικές λέξεις  $w(c)$  είναι καθολικές στην συνάρτηση HPM, τα βάρη  $s(c)$  είναι καθολικά στις συναρτήσεις HPM-REC και ERASE, παρομοίως π.χ. για τα  $B, L, X$ . Αντιθέτως, το  $C$  ορίζεται τοπικά. Στην πραγματικότητα, το  $|C|$  φθίνει όσο κατεβαίνουμε την ιεραρχία.

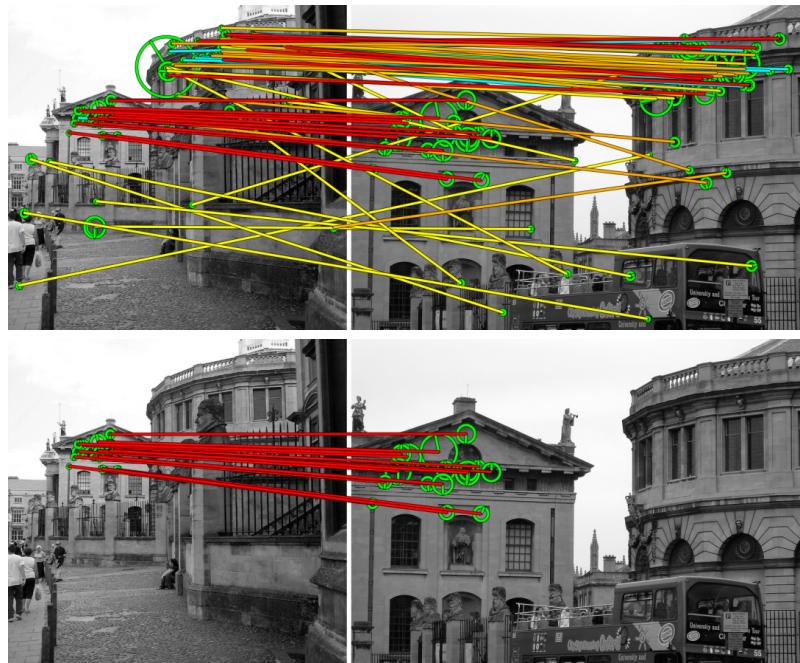
Η καταχώρηση των αντιστοιχιών σε κυτία γίνεται σε γραμμικό χρόνο ως προς το  $|C|$ , στην γραμμή 4. Το  $B_\ell$  κατατεμαχίζει το  $\mathcal{F}$  για κάθε επίπεδο  $\ell$ , έτσι δεδομένης μιας αντιστοιχίας  $c$  υπάρχει ένα μοναδικό κυτίο  $b \in B_\ell$  τέτοιο ώστε  $f(c) \in b$ . Έπειτα ορίζονται

	$c_1$	$c_2$	$c_3$	$c_4$	$c_5$	$c_8$	$c_9$	$c_6$	$c_7$
$c_1$	●								
$c_2$		●							
$c_3$			●						
$c_4$				●	●				
$c_5$					●				
$c_8$						●			
$c_9$							●		
$c_6$								●	
$c_7$									●

**Σχήμα 3.3:** Πίνακας με βαθμούς ομοιότητας ισοδύναμους με το Σχήμα 3.2 και σύμωνα με την σχέση (3.7). Οι αντιστοιχίες έχουν αντιμετατεθεί ώστε οι ομάδες να σχηματίζουν συνεχόμενα μπλοκς. Οι ομάδες οι οποίες σχηματίζονται στα επίπεδα 0, 1, 2 λαμβάνουν ομοιότητα ίση με  $1, \frac{1}{2}, \frac{1}{4}$  αντίστοιχα. Ο βαθμός ομοιότητας του Σχήματος 3.2 μπορεί να αποκτηθεί προσθέτοντας ανά γραμμή και πολλαπλασιάζοντας με την ισχύ της κάθε αντιστοιχίας.



**Σχήμα 3.4:** Οι αντιστοιχίες του παραδείγματος στο Σχήμα 3.5 σαν σημεία στον  $4\Delta$  χώρο των μετασχηματισμών. Απεικονίζουμε μία προβολή σε  $2\Delta$ , αυτήν της μετατόπισης  $(x, y)$ . Η μετατόπιση είναι κανονικοποιημένη με βάση την μέγιστη διάσταση εικόνες. Υπάρχουν  $L = 5$  επίπεδα κι εμείς εστιάζουμε στα κεντρικά  $8 \times 8$  κυτία. Οι ακμές αναπαριστούν τους συνδέσμους μεταξύ αντιστοιχιών που ομαδοποιούνται στα επίπεδα 0, 1, 2 μόνο. Η ομοιότητα α αναπαρίσταται με 3 διαφορετικούς τόνους του γκρι, με το μαύρο να αντιστοιχεί στο  $\alpha(0) = 1$ .



**Σχήμα 3.5:** Πάνω: Ταίριασμα με τον HPM δύο εικόνων της βάσης Oxford, σε 0.6ms. Όλες οι αρχικές αντιστοιχίες απεικονίζονται. Με γαλάζιο χρώμα είναι εκείνες που διαγράφηκαν. Οι υπόλοιπες χρωματίζονται ανάλογα με την ισχύ τους, με το κόκκινο (κίτρινο) να είναι οι πιο δυνατές (αδύναμες). Κάτω: Inliers με την μέθοδο FSM με 4 βαθμούς ελευθερίας και αφινικό μοντέλο στο LO-RANSAC, σε 7ms.

την προβολή  $\beta_\ell : c \mapsto b$  κβαντίζοντας το διάνυσμα παραμέτρων  $f(c)$  στο επίπεδο  $\ell$ . Η αποθήκευση στα κυτία είναι αραιή και γραμμικών απαιτήσεων ως προς το πλήθος  $|C|$ . Οι πλήρεις κατατμήσεις  $B_\ell$  των χώρου δεν κατασκευάζονται ποτέ στην πραγματικότητα.

Δεδομένου ενός συνόλου αντιστοιχιών σε ένα κυτίο, ο βέλτιστος εντοπισμός των ασυμβατοτήτων μπορεί να είναι δύσκολο πρόβλημα. Στην συνάρτηση ERASE, ακολουθούμε μία πολύ απλή προσέγγιση όπου δύο αντιστοιχίες είναι σε ασυμβατότητα εάν μοιράζονται την ίδια οπτική λέξη. Αυτό έχει ιδιαίτερο νόημα σε ένα πολύ μεγάλο λεξικό, όπου σε πολλές οπτικές λέξεις ανήκει μόνο ένα χαρακτηριστικό από κάθε εικόνα (π.χ. 92% των χαρακτηριστικών για κάθε εικόνα στα σύνολα πειραμάτων μας).

Όπως φαίνεται και στον αλγόριθμο 2, για όλες τις αντιστοιχίες  $h(b)$  του κυτίου  $b$  πρώτα κατασκευάζουμε το σύνολο  $U$  των κοινών οπτικών λέξεων (γραμμή 3), όπου  $u(c)$  είναι η οπτική λέξη στην οποία έχει αντιστοιχηθεί η αντιστοιχία  $c$ . Έπειτα, στις γραμμές 4-5, ορίζουμε για κάθε οπτική λέξη  $u \in U$  την κατηγορία οπτικής λέξης  $e(u)$ , δηλαδή όλες τις αντιστοιχίες οι οποίες έχουν αντιστοιχηθεί στην οπτική λέξη  $u$ . Σύμφωνα με την παραδοχή μας, όλες οι αντιστοιχίες που ανήκουν στην ίδια κατηγορία είναι ασύμβατες ανά δύο μεταξύ τους. Έτσι, κρατάμε την πιο ισχυρή αντιστοιχία σε κάθε κατηγορία, διαγράφουμε τις υπόλοιπες και ανανεώνουμε το  $X$ .

---

**Algorithm 1** Hough pyramid matching (HPM)

---

```

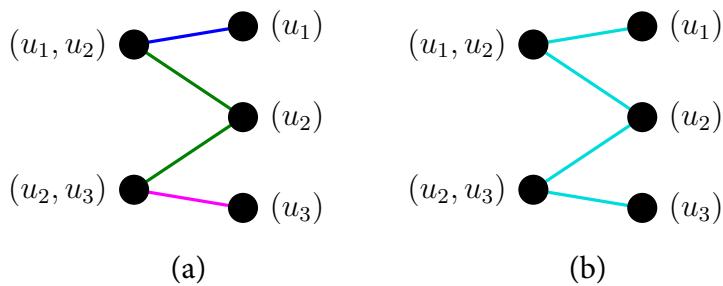
procedure HPM(assignments  $C$ , levels  $L$ )
begin
     $X \leftarrow \emptyset; \mathcal{B} \leftarrow \text{PARTITION}(L)$                                  $\triangleright$  partition space in  $L$  levels
    for  $c \in C$  do  $s(c) \leftarrow 0$                                                $\triangleright$  initialize strengths
    2 HPM-REC( $C, L - 1$ )                                                  $\triangleright$  recurse at top
    3 return score  $\sum_{c \in C \setminus X} w(c)s(c)$ 

procedure HPM-REC(assignments  $C$ , level  $\ell$ )
begin
    if  $|C| < 2 \vee \ell < 0$  then return
    for  $b \in B_\ell$  do  $h(b) \leftarrow \emptyset$                                           $\triangleright$  initialize histogram
    4 for  $c \in C$  do  $h(\beta_\ell(c)) \leftarrow h(\beta_\ell(c)) \cup c$            $\triangleright$  quantize
    5 for  $b \in B_\ell$  do HPM-REC( $h(b), \ell - 1$ )                          $\triangleright$  recurse down
    6 for  $b \in B_\ell$  do
        7      $X \leftarrow X \cup \text{ERASE}(h(b))$                                 $\triangleright$  exclude erased
         $h(b) \leftarrow h(b) \setminus X$                                           $\triangleright$  exclude isolated
        8     if  $|h(b)| < 2$  then continue
        9     if  $\ell = L - 1$  then  $a \leftarrow 2$  else  $a \leftarrow 1$ 
        10    for  $c \in h(b)$  do  $s(c) \leftarrow s(c) + a2^{-\ell}|h(b)|$             $\triangleright$  (3.7)

```

---

Είναι προφανές ότι όλες οι πράξεις σε κάθε αναδρομική κλήση του κυτίου  $b$  είναι γραμμικές ως προς το  $|h(b)|$ . Εφόσον το  $B_\ell$  κατατεμαχίζει το  $\mathcal{F}$  για όλα  $\ell$ , οι συνολικές πράξεις ανά επίπεδο είναι γραμμικές ως προς το  $n = |C|$ . Επομένως η χρονική πολυπλοκότητα του HPM είναι  $O(nL)$ .



**Σχήμα 3.6:** Εντοπισμός των ασυμβατοτήτων βάσει (a) κατηγοριών οπτικών λέξεων και (b) κατηγοριών συνιστωσών, με πολλαπλή ανάθεση οπτικών λέξεων. Οι κόμβοι στα αριστερά (δεξιά) αναπαριστούν τα χαρακτηριστικά της εικόνας ερωτήματος (της βάσης). Οι επικεφαλίδες συμβολίζουν τις οπτικές λέξεις, οι οποίες είναι πολλαπλές για τα τοπικά χαρακτηριστικά της εικόνας ερωτήματος. Το ίδιο χρώμα συμβολίζει αντιστοιχίες σε ασυμβατότητα.

---

**Algorithm 2** Διαγραφή με χρήση κατηγοριών οπτικών λέξεων

---

```

1: procedure ERASE(assignments  $C$ )
2:    $x \leftarrow \emptyset; U \leftarrow \emptyset;$ 
3:   for all  $c \in C$  do  $U \leftarrow U \cup u(c)$  ▷ common visual words
4:   for all  $u \in U$  do  $e(u) \leftarrow \emptyset$  ▷ visual word classes
5:   for all  $c \in C$  do  $e(u(c)) \leftarrow e(u(c)) \cup c$ 
6:   for all  $u \in U$  do  $x \leftarrow x \cup e(u) \setminus \arg \max_{c \in e(u)} s(c)$ 
7:   return erased assignments  $x$  ▷ all but strongest
8: end procedure

```

---

**Algorithm 3** Διαγραφή με χρήση κατηγοριών συνεκτικών συνιστωσών

---

```

1: procedure ERASE-CC(connected components  $Z$ )
2:    $x \leftarrow \emptyset;$ 
3:   for all  $z \in Z$  do  $e(z) \leftarrow E(z)$  ▷ component classes
4:   for all  $z \in Z$  do  $x \leftarrow x \cup e(z) \setminus \arg \max_{c \in e(z)} s(c)$ 
5:   return erased assignments  $x$  ▷ all but strongest
6: end procedure

```

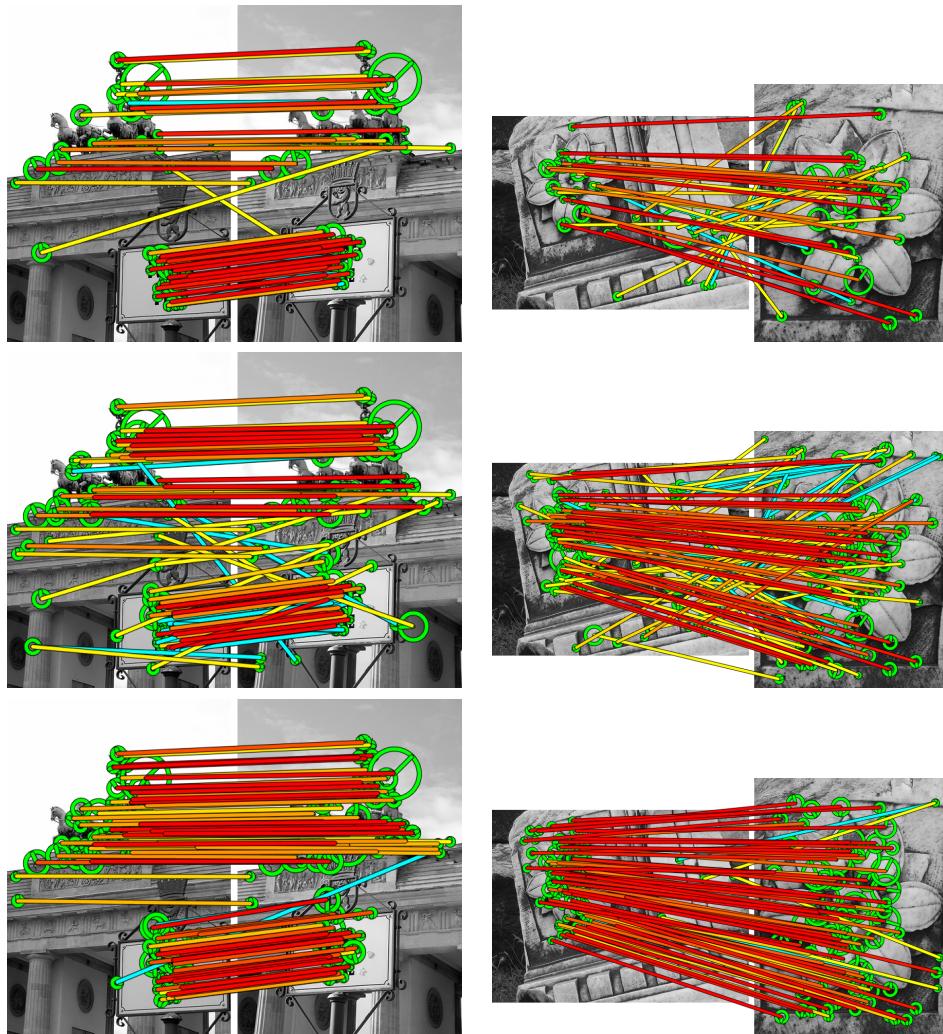
---

### 3.4 Ταίριασμα με χρήση πολλαπλής ανάθεσης

Η χρήση ενός οπτικού λεξικού πάντα εισάγει απώλειες. Μία συνηθισμένη στρατηγική για να βελτιώσει κανείς αυτή την κατάσταση είναι να αναθέτει τα τοπικά χαρακτηριστικά σε πολλαπλές οπτικές λέξεις, στις κοντινότερες στον περιγραφέα τους [69]. Αυτό είναι προτιμότερο να γίνεται από την μεριά της εικόνας ερωτήματος μόνο καθώς δεν αυξάνονται έτσι οι ανάγκες μνήμης της δομής δεικτοδότησης [34]. Διερευνάμε εδώ την λύση της πολλαπλής ανάθεσης μαζί με την χρήση της μεθόδου HPM, ακολουθώντας τον τρόπο αυτό τόσο στην διαδικασία φιλτραρίσματος όσο και σε εκείνη της ανακατάταξης.

Στην περίπτωση της απλής ανάθεσης, ο εντοπισμός των ασυμβατοτήτων έγινε με την παρατήρηση ότι ένα υψηλό ποσοστό τοπικών χαρακτηριστικών είναι τα μοναδικά τα οποία αντιστοιχίζονται με μία συγκεκριμένη οπτική λέξη. Αυτό όμως δεν ισχύει και για την περίπτωση της πολλαπλής ανάθεσης. Το ποσοστό αυτό πέφτει πολύ, π.χ. 85% (80%) για 3 (5) πλησιέστερους γείτονες, δηλαδή αντίστοιχο αριθμό αναθέσεων σε οπτικές λέξεις.

Το Σχήμα 3.6(a) απεικονίζει τον εντοπισμό των ασυμβατοτήτων με την χρήση κατηγοριών οπτικών λέξεων όταν γίνεται πολλαπλή ανάθεση. Η εικόνα της βάσης έχει τρία τοπικά χαρακτηριστικά σε αυτό το παράδειγμα, τα οποία αντιστοιχίζονται σε τρεις διαφορετικές οπτικές λέξεις  $u_1, u_2, u_3$  αντίστοιχα, ενώ η εικόνα ερωτήματος έχει δύο τοπικά χαρακτηριστικά κι έπειτα από πολλαπλή ανάθεση αντιστοιχίζονται στις οπτικές λέξεις  $u_1, u_2$  και  $u_2, u_3$  αντίστοιχα. Υπάρχουν τρεις κατηγορίες οπτικών λέξεων, και διατηρώντας μία



**Σχήμα 3.7:** Παραδείγματα ταιριάσματος με την μέθοδο HPM. (Πάνω) απλή ανάθεση, κατηγορίες οπτικών λέξεων. (Μέση) Πολλαπλή ανάθεση με 3 πλησιέστερους γείτονες, κατηγορίες συνιστωσών. (Κάτω) απευθείας ταιριασμα με χρήση οπτικών περιγραφέων και χρήση του ελέγχου λόγου (ratio test) [53], κατηγορίες συνιστωσών.

αντιστοιχία από την κάθε μία (αλγόριθμος 2) οδηγούμαστε σε σχήμα ταιριάσματος ένα με πολλά, έχοντας ένα τοπικό χαρακτηριστικό από την εικόνα της βάσης να αντιστοιχεί σε δύο χαρακτηριστικά της εικόνας της βάσης.

Για αυτό τον λόγο προτείνουμε μία εναλλακτική λύση η οποία πάντα διατηρεί το ταιριασμα ένα προς ένα, χωρίς να επηρεάζει πολύ την πολυπλοκότητα. Δεδομένων δύο εικόνων  $P, Q$ , το σύνολο αποτελούμενο από την ένωση των χαρακτηριστικών  $V = P \cup Q$  και το σύνολο των αντιστοιχιών  $C$  μπορούν να θεωρηθούν ως ένας μη κατευθυντικός γράφος  $G = (V, C)$ , όπου κάθε χαρακτηριστικό είναι ένας κόμβος και κάθε αντιστοιχία είναι μία ακμή.

Με αυτή την αναπαράσταση, υπολογίζουμε το σύνολο  $Z$  των συνεκτικών συνιστωσών του  $G$ . Στην πράξη αυτό γίνεται με την χρήση του αλγορίθμου *union find*. Οι αντιστοιχίες

## Υλοποίηση

οι οποίες ανήκουν σε κάθε συνιστώσα χρησιμοποιούνται για να ορίζουν τις κατηγορίες συνιστώσων. Οι κατηγορίες συνιστωσών αντικαθιστούν τις κατηγορίες οπτικών λέξεων στον αλγόριθμο 2. Συγκεκριμένα, όλες οι αντιστοιχίες στην ίδια συνιστώσα θεωρούνται μη συμβατές και μόνο μία θα κρατηθεί.

Το Σχήμα 3.6(b) απεικονίζει τις ασυμβατότητες χρησιμοποιώντας αυτό το σχήμα για το παράδειγμα του Σχήματος 3.6(a). Υπάρχει μόνο μία συνεκτική συνιστώσα σε αυτή την περίπτωση, μόνο μία αντιστοιχία διατηρείται και τελικά εφαρμόζεται σχήμα ένα προς ένα. Ισχύει γενικά ότι όλες οι αντιστοιχίες ενός τοπικού χαρακτηριστικού ανήκουν στην ίδια συνιστώσα. Το να κρατήσει κανείς μία αντιστοιχία από κάθε συνιστώσα συνεπάγεται να κρατάει μία αντιστοιχία για κάθε χαρακτηριστικό άρα το σχήμα ένα προς ένα ισχύει πάντα. Φυσικά ο τρόπος του παραδείγματος στο Σχήμα 3.6(b) είναι πιο αυστηρός από το απαραίτητο. Χάνει για παράδειγμα μία αντιστοιχία η οποία θα μπορούσε να είναι έγκυρη. Από την άλλη είναι ευεργετικός γιατί αφαιρεί πολλές αντιστοιχίες οι οποίες πιθανόν να αντιστοιχούν σε τοπικά χαρακτηριστικά που προέρχονται από το υπόβαθρο.

Η τροποποιημένη συνάρτηση ERASE-CC συνοψίζεται στον Αλγόριθμο 3, όπου οι κατηγορίες οπτικών λέξεων  $e(u)$  έχουν αντικατασταθεί από τις κατηγορίες συνεκτικών συνιστωσών  $e(z)$ . Το  $E(z)$  στην γραμμή 3 συμβολίζει το σύνολο των ακμών του  $z$  (το οποίο είναι κι αυτό ένας γράφος). Αντιμετωπίζουμε και τις δύο αυτές κατηγορίες σαν κατηγορίες ισοδυναμίας όπου μόνο ένας αντιπρόσωπος επιλέγεται από την κάθε μία. Όπως και στο παράδειγμα του Σχήματος 3.6, αυτός ο τρόπος μπορεί να γίνει πολύ αυστηρός, ειδικά όταν το λεξικό δεν είναι αρκετά μεγάλο. Από την άλλη, εφαρμόζεται ακόμα και όταν δεν γίνεται χρήση λεξικού, όπως στην περίπτωση που έχουμε σχηματισμό αντιστοιχιών με απευθείας ταίριασμα τοπικών περιγραφέων.

Το ταίριασμα μεταξύ δύο εικόνων με χρήση διαφορετικού τύπου αντιστοιχιών και διαφορετικών μεθόδων διαγραφής απεικονίζεται στο Σχήμα 3.7. Είναι εμφανές το πως η μέθοδος HPM δουλεύει πολύ καλύτερα στην περίπτωση που οι αντιστοιχίες δημιουργούνται από απευθείας ταίριασμα περιγραφέων. Με την χρήση λεξικού και πολλαπλής ανάθεσης οι ισχυρές αντιστοιχίες είναι πιο λίγες, ενώ με την απλή ανάθεση ακόμα λιγότερες.

## 3.5 Υλοποίηση

### 3.5.1 Δεικτοδότηση κι ανακατάταξη

Το HPM φαίνεται ότι είναι αρκετά γρήγορο για να το ενσωματώσουμε στην αναζήτηση εικόνων από μεγάλες βάσεις και να το χρησιμοποιήσουμε για γεωμετρική επιβεβαίωση κατά το στάδιο της ανακατάταξης. Κατασκευάζουμε ένα ανάστροφο αρχείο δεικτοδοτημένο ανά οπτική λέξη, όπου για κάθε χαρακτηριστικό της βάσης αποθηκεύουμε την κβα-

ντισμένη θέση, κλίμακα, περιστροφή και τον αύξοντα αριθμό της κάθε εικόνας. Δεδομένης μίας εικόνας ερωτήματος, αυτή η πληροφορία είναι αρκετή για να διεκπεραιώσουμε το αρχικό υπο-γραμμικό φιλτράρισμα, είτε με την μέθοδο *bag of words* (BoW) [73] είτε με την μέθοδο *weak geometric consistency* (WGC) [34]. Ένας προκαθορισμένος αριθμός εικόνων από τις πιο όμοιες σημαδεύονται για ανακατάταξη. Για κάθε χαρακτηριστικό της εικόνες ερωτήματος, ανακτάμε τις υποψήφιες αντιστοιχίες από το ανάστροφο αρχείο μία φορά ακόμα, αλλά αυτή την φορά μόνο για τις σημαδεμένες εικόνες. Για κάθε αντιστοιχία *c* που βρίσκουμε, υπολογίζουμε το διάνυσμα παραμέτρων  $f(c)$  του σχετικού μετασχηματισμού και το αποθηκεύουμε σε μία συλλογή δεικτοδοτημένη ανά αριθμό εικόνας. Δεδομένων όλων των αντιστοιχιών και των διανυσμάτων παραμέτρων ταιριάζουμε κάθε εικόνα με την εικόνα ερώτημα χρησιμοποιώντας το HPM. Τέλος, κανονικοποιούμε την ομοιότητα με την BoW  $\ell_2$  κανονικοποίηση των εικόνων της βάσης και προχωράμε σε ανακατάταξη.

image id	$x$	$y$	$\log \sigma$	$\theta$	total
16	4	4	4	4	32

**Πίνακας 3.1:** Χρήση σε μνήμη του ανάστροφου αρχείου ανά χαρακτηριστικό, σε bits. Χρησιμοποιούμε κωδικοποίηση run-length για τον αύξοντα αριθμό της κάθε εικόνας, οπότε 2 bytes είναι αρκετά.

### 3.5.2 Κβαντισμός

Συμπεριφερόμαστε ξεχωριστά σε κάθε παράμετρο των σχετικών μετασχηματισμών  $x, y, \sigma, \theta$  —βλέπε σχέση (3.3). Η μετατόπιση  $t(c)$  στην σχέση (3.2) αναφέρεται στο τοπικό σύστημα αξόνων της εικόνας ερωτήματος,  $Q$ . Εάν  $r$  είναι η μέγιστη διάσταση του  $Q$  σε pixels, κρατάμε μόνο τις αντιστοιχίες με οριζόντια ή κάθετη μετατόπιση  $x, y \in [-3r, 3r]$ . Επίσης κρατάμε μόνο αντιστοιχίες για τις οποίες ισχύει  $\sigma \in [1/\sigma_m, \sigma_m]$ , όπου το  $\sigma_m = 10$  είναι πάνω από τις δυνατότητες κάθε ανιχνευτή χαρακτηριστικών. Υπολογίζουμε την λογαριθμική κλίμακα, κανονικοποιούμε στο  $[0, 1]$  και κβαντίζουμε τις παραμέτρους ομοιόμορφα. Επίσης κβαντίζουμε τις παραμέτρους του τοπικού σχήματος των χαρακτηριστικών: με 5 επίπεδα, κάθε παράμετρος κβαντίζεται σε 16 κυτία. Οι ανάγκες τις μεθόδου μας σε μνήμη ανά χαρακτηριστικό, συνοψίζονται στον Πίνακα 3.1, και είναι ακριβώς οι ίδιες με αυτές του [34]. Οι παράμετροι της εικόνας ερωτήματος δεν κβαντίζονται.

### 3.5.3 Πρότερη γνώση περιστροφής

Επειδή οι περισσότερες εικόνες στο διαδίκτυο απεικονίζουν τα αντικείμενα είτε οριζόντια είτε κάθετα, προηγούμενες μέθοδοι χρησιμοποιούν αυτή την γνώση για την σχετική περιστροφή των μοντέλων τους [66][34]. Χρησιμοποιούμε κι εμείς το μοντέλο πρότερης γνώ-

## Πειράματα

σης με βάρη όπως και στο WGC [34]. Αυτό γίνεται με την μορφή επιπλέον βαρών στην σχέση (3.8).

### 3.6 Πειράματα

Σε αυτή την ενότητα αξιολογούμε το HPM απέναντι στην μέθοδο *fast spatial matching* (FSM) [66] στο ταίριασμα ζευγαριών εικόνων και στην ανακατάταξη εικόνων στην ανάκτηση εικόνων από μεγάλες βάσεις. Στην δεύτερη περίπτωση, χρησιμοποιούμε δύο διαφορετικές τεχνικές για το πρώτο στάδιο του φιλτραρίσματος των εικόνων, την μέθοδο *bag-of-words* (BoW) [73] και την *weak geometric consistency* (WGC) [34].

#### 3.6.1 Πρωτόκολλο πειραμάτων

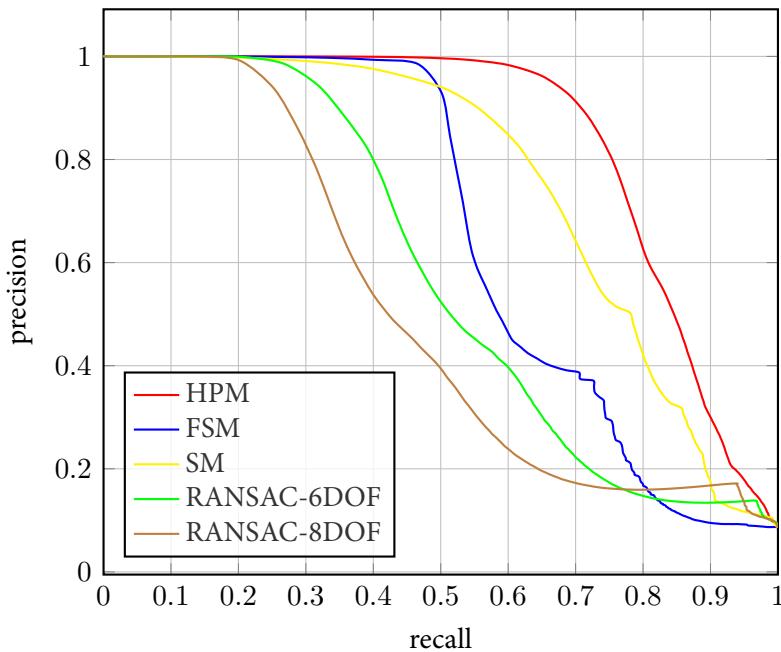
**Βάσεις εικόνων.** Διεξάγουμε πειράματα σε δύο δημόσια διαθέσιμες βάσεις εικόνων, την *Oxford Buildings* [66] και την *Paris* [69], καθώς και την δική μας *World Cities*<sup>1</sup>. Η τελευταία έχει εικόνες που έχουν συλλεχθεί από το *Flickr* κι αποτελείται από 927 προσημειωμένες εικόνες οι οποίες έχουν τραβηγθεί στο κέντρο της πόλης της Βαρκελώνης και 2 εκατομμύρια εικόνες από 38 πόλεις για να χρησιμοποιηθούν ως εικόνες περίσπασης. Οι προσημειωμένες εικόνες είναι διαχωρισμένες σε 17 ομάδες, από τις οποίες η κάθε μία απεικονίζει το ίδιο κτήριο ή σκηνή. Έχουμε επιλέξει 5 εικόνες ερωτήματα από κάθε ομάδα, δημιουργώντας συνολικά 85 εικόνες ερωτήματα για την αξιολόγηση. Αναφερόμαστε στις βάσεις *Oxford Buildings*, *Paris* και στο δικό μας προσημειωμένο σύνολο εικόνων ως σύνολα ελέγχου.

**Τοπικά χαρακτηριστικά κι οπτικά λεξικά.** Εξάγουμε από κάθε εικόνα τοπικά χαρακτηριστικά και περιγραφείς SURF [8], θέτοντας το κατώφλι της ισχύς των χαρακτηριστικών ίσο με 2.0 για τον ανιχνευτή. Κατασκευάζουμε τα οπτικά λεξικά με τον προσεγγιστικό αλγόριθμο  $k$ -μέσων (approximate  $k$ -means - AKM) [66] και χρησιμοποιούμε κυρίως γενικό λεξικό μεγέθους 100K λέξεων το οποίο προέρχεται από υποσύνολο των 2M εικόνων περίσπασης κι όχι από τα προσημειωμένα σύνολα εικόνων. Αυτό είναι πιο κοντά στην κατάσταση ενός πραγματικού συστήματος ανάκτησης εικόνων. Παρόλα αυτά, για λόγους συγκρίσεων χρησιμοποιούμε και ειδικά λεξικά τα οποία έχουν κατασκευασθεί από τα σύνολα ελέγχου. Χρησιμοποιούμε τον γενικό λεξικό όπου δεν δηλώνεται κάτι διαφορετικό.

#### 3.6.2 Πείραμα ταιριάσματος ζευγαριών

Απαριθμώντας όλα τα δυνατά ζευγάρια εικόνων στο σύνολο ελέγχου *World Cities*, υπάρχουν 74,075 ζευγάρια εικόνων τα οποία απεικονίζουν το ίδιο κτήριο. Ο βαθμός ομοιότητας πρέπει να είναι υψηλός για αυτά τα ζευγάρια ενώ χαμηλός για τα υπόλοιπα 785, 254.

<sup>1</sup>[http://image.ntua.gr/iva/datasets/world\\_cities](http://image.ntua.gr/iva/datasets/world_cities)



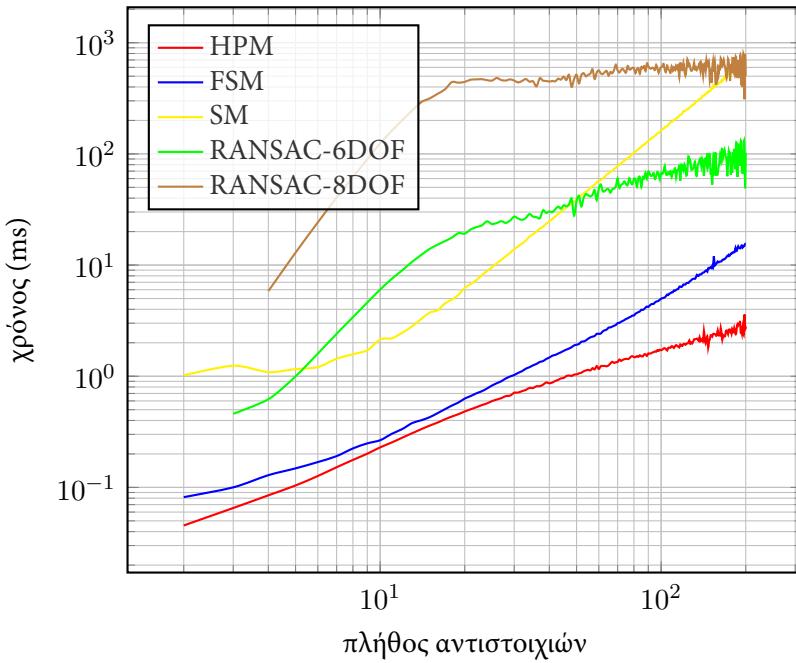
Σχήμα 3.8: Καμπύλες ακρίβειας-ανάκλησης σε όλα τα ζευγάρια του συνόλου ελέγχου World Cities χωρίς εικόνες περίσπασης.

Επομένως εφαρμόζουμε διαφορετικά κατώφλια για να ταξινομήσουμε τις εικόνες σε όμοιες κι ανόμοιες, και συγκρίνουμε με την προσημείωση. Ταιριάζουμε όλα τα πιθανά ζευγάρια με τον αλγόριθμο RANSAC με 6 βαθμούς ελευθερίας, με τον αλγόριθμο RANSAC με 8 βαθμούς ελευθερίας, τον FSM με 4 βαθμούς ελευθερίας (μετατόπιση, κλιμάκωση και περιστροφή), με τον αλγόριθμο (spectral matching) SM [50] και με το HPM. Στους αλγόριθμους RANSAC και FSM εφαρμόζουμε κι ένα τελευταίο στάδιο του LO-RANSAC όπως στο [66] για να ανακτήσουμε αφινικό μετασχηματισμό (ομογραφίας για τον RANSAC με 8 βαθμούς ελευθερίας), και υπολογίζουμε την τελική ομοιότητα από τις τιμές του *idf* των inliers. Στο HPM ο βαθμός ομοιότητας δίνεται από την σχέση (3.8). Ταξινομούμε τα ζευγάρια σύμφωνα με την ομοιότητα τους και κατασκευάζουμε καμπύλες ακρίβειας-ανάκλησης (precision-recall) οι οποίες φαίνονται στο Σχήμα 3.8, όπου το HPM φαίνεται να ξεπερνά όλες τις μεθόδους.

Για το ίδιο πείραμα δείχνουμε στο Σχήμα 3.9 τους χρόνους εκτέλεσης ως προς τον αριθμό των αντιστοιχιών ανάμεσα στις δύο εικόνες. Το HPM είναι πιο γρήγορο από τις άλλες μεθόδους και φαίνεται και η γραμμική σχέση του χρόνου με τον πλήθος των αντιστοιχιών.

### 3.6.3 Πειράματα ανακατάταξης

Διεξάγουμε πειράματα στην ανάκτηση εικόνων με το BoW και το WGC με  $\ell_2$  κανονικοποίηση για το πρώτο στάδιο του φιλτραρίσματος. Συνδυάζουμε και τις δύο μέθοδος με



**Σχήμα 3.9:** Χρόνος εκτέλεσης ως προς τον πλήθος των αντιστοιχιών για όλα τα ζευγάρια εικόνων του συνόλου Barcelona.

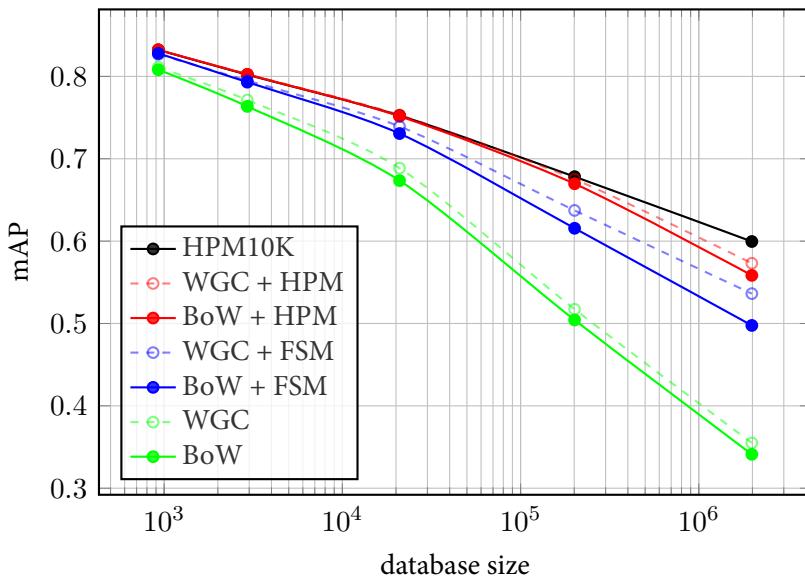
το HPM και το 4-dof FSM για ανακατάταξη σε δεύτερο στάδιο με χρήση της γεωμετρίας. Μετράμε την απόδοση μέσω της μέσης ακρίβειας (mean Average Precision - mAP). Επιπλέον συγκρίνουμε τους χρόνους που χρειάζεται η κάθε μέθοδος ανακατάταξης. Όλοι οι χρόνοι έχουν μετρηθεί σε έναν επεξεργαστή 2GHz quad core με τις δικές μας υλοποιήσεις σε C++.

$L$	2	3	4	5	6
πυραμίδα	0.473	0.498	0.536	0.556	<b>0.559</b>
επίπεδο	0.448	0.485	0.524	<b>0.534</b>	0.509

**Πίνακας 3.2:** Απόδοση mAP για επίπεδο και πυραμιδικό ταίριασμα σε διαφορετικά επίπεδα  $L$  στην βάση World Cities με 2M εικόνες περίσπασης. Το φιλτράρισμα γίνεται με την μέθοδο BoW και οι πρώτες 1K εικόνες ανακατατάσσονται.

**Επίπεδα.** Κβαντίζοντας τις παραμέτρους των τοπικών χαρακτηριστικών σε 6 επίπεδα στον ανάστροφο αρχείο, μετράμε την απόδοση του HPM ως προς τα επίπεδα  $L$  της πυραμίδας, όπως φαίνεται στον Πίνακα 3.2. Εφαρμόζουμε επίσης ανακατάταξη σε ένα και μοναδικό επίπεδο της πυραμίδας, το κατώτερο, για κάθε  $L$ . Αναφερόμαστε σε αυτό τον τρόπο ως ανακατάταξη με επίπεδο ταίριασμα. Παρατηρούμε πως η βελτίωση πηγαίνοντας από τα 5 στα 6 επίπεδα είναι πολύ μικρή, ενώ στο επίπεδο ταίριασμα η απόδοση πέφτει. Επομένως η επιλογή μας για  $L = 5$  έχει ιδιαίτερο νόημα, εκτός από την εξοικονόμηση χώρου—βλέπε ενότητα 3.5. Για το ίδιο πείραμα, το mAP είναι 0.341 και 0.497 για την μέ-

θοδο BoW και για την BoW+FSM αντίστοιχα. Είναι ενδιαφέρον να παρατηρήσουμε πως ακόμα και για το επίπεδο ταίριασμα υπάρχει αισθητή βελτίωση. Αυτό συμβαίνει λόγω της ευκαμψίας του μοντέλου μας και την ικανότητα του να χειριστεί πολλαπλούς μετασχηματισμούς, το οποίο είναι μία διαφοροποίηση από το [53]. Το ταίριασμα με την πυραμίδα βελτιώνει επιπλέον την απόδοση.

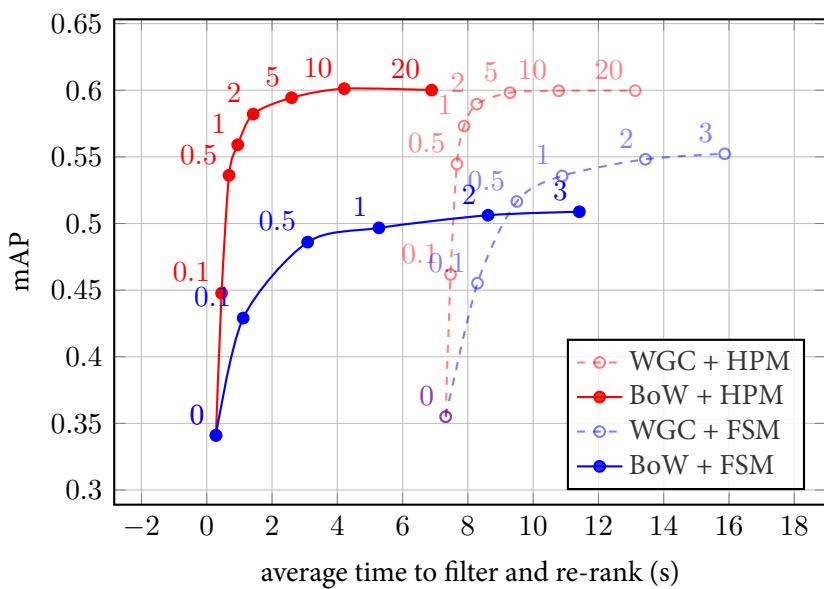


**Σχήμα 3.10:** Σύγκριση  $mAP$  για μεταβλητό μέγεθος βάσης World Cities με μέχρι 2M εικόνες περίσπασης. Το φιλτράρισμα εφαρμόζεται με την μέθοδο BoW ή WGC και η ανακατάταξη στις πρώτες 1K εικόνες με το FSM ή το HPM, εκτός από την περίπτωση HPM10K όπου οι καμπύλες BoW και WGC συμπίπτουν.

**Εικόνες περίσπασης.** Το Σχήμα. 3.10 συγκρίνει το HPM με το FSM και τις βασικές μεθόδους, για διαφορετικό αριθμό εικόνων περίσπασης μέχρι και 2M. Οι μέθοδοι BoW και WGC χρησιμοποιούνται για φιλτράρισμα. Το HPM φαίνεται να ξεπερνά σε απόδοση το FSM σε όλες τις περιπτώσεις. Επίσης με το HPM ανακατατάσσουμε 10K εικόνες, εφόσον αυτό παίρνει λιγότερο χρόνο από ότι 1K εικόνες με το FSM. Σε αυτή την περίπτωση προκύπτει και η καλύτερη απόδοση, ιδιαίτερα με την παρουσία εικόνων περίσπασης. Ενδιαφέρον είναι πως δεν υπάρχει διαφορά εάν το φιλτράρισμα γίνεται με το BoW ή με το WGC σε αυτή την περίπτωση. Στον Πίνακα 3.3 συνοψίζουμε τα αποτελέσματα από το ίδιο πείραμα με πρότερη γνώση περιστροφής στο WGC και το HPM. Όταν αυτά χρησιμοποιούνται μαζί τότε η πρότερη γνώση εφαρμόζεται και στα δύο. Ακόμα, οι μέθοδοι BoW και WGC είναι σχεδόν ισοδύναμες στην περίπτωση HPM10K. Η χρήση της πρότερης γνώσης περιστροφής γενικότερα αυξάνει την απόδοση, αλλά εξαρτάται από την βάση εικόνων. Το ανεπιθύμητο αποτέλεσμα σε αυτή την περίπτωση είναι η περιορισμένη αν αλλοίωση ως προς την περιστροφή.

Μέθοδος	όχι εικ. περισπ.		2M εικ. περισπ.	
	όχι προτ.	προτ.	όχι προτ.	προτ.
WGC+HPM10K	–	–	0.599	0.612
BoW+HPM10K	–	–	<b>0.601</b>	<b>0.613</b>
WGC+HPM	<b>0.832</b>	<b>0.851</b>	<b>0.573</b>	<b>0.599</b>
BoW+HPM	0.832	0.837	0.558	0.565
WGC+FSM	0.826	0.846	0.536	0.572
BoW+FSM	0.827	–	0.497	–
WGC	0.811	0.843	0.355	0.447
BoW	0.808	–	0.341	–

Πίνακας 3.3: Σύγκριση mAP στην βάση World Cities με και χωρίς χρήση πρότερης γνώσης περιστροφής. Η ανακατάταξη γίνεται στις πρώτες 1K εικόνες, εκτός από την περίπτωση HPM10K.



Σχήμα 3.11: mAP και συνολικός χρόνος εκτέλεσης του ερωτήματος (φιλτράρισμα + ανακατάταξη) για μεταβαλλόμενο αριθμό εικόνων στις οποίες εφαρμόζεται η ανακατάταξη. Οι τελευταίες φαίνονται με κειμενικές επιγραφές δίπλα στα σημεία, σε χιλιάδες. Τα αποτελέσματα είναι στην βάση World Cities με 2M εικόνες περίσπασης.

**Μετρήσεις χρόνων.** Μεταβάλλοντας τον αριθμό των εικόνων στις οποίες εφαρμόζεται ανακατάταξη, μετράμε το mAP και τον χρόνο εκτέλεσης του ερωτήματος για το FSM και το HPM. Για μία ακόμα φορά, χρησιμοποιούμε και το BoW και το WGC για φιλτράρισμα. Το συνδυαστικό διάγραμμα δίνεται στο Σχήμα 3.11. Το HPM φαίνεται να μπορεί να εφαρμοστεί σε 10 φορές περισσότερες εικόνες σε λιγότερο χρόνο από ότι το FSM. Με το BoW, το mAP του είναι 10% υψηλότερο από εκείνο του FSM για τον ίδιο χρόνο, κατά μέσο όρο. Με το κόστος 7 επιπλέον δευτερολέπτων για φιλτράρισμα, το FSM τελικά επωφελείται από το WGC, ενώ το HPM δεν επηρεάζεται. Πράγματι, μετά από 3.3 δευτερόλεπτα, το

mAP του BoW+HPM φθάνει στον κορεσμό με ανακατάταξη 7K εικόνων, ενώ το WGC φαίνεται να μην συνεισφέρει.

Μέθοδος	Μέγεθος λεξικού			
	100K	200K	500K	700K
BoW+HPM+P	0.640	0.683	0.701	0.690
BoW+HPM	0.622	<b>0.669</b>	<b>0.692</b>	<b>0.686</b>
BoW+FSM	<b>0.631</b>	0.642	0.677	0.653
BoW	0.545	0.575	0.619	0.614

**Πίνακας 3.4:** Σύγκριση mAP στην βάση Oxford για ειδικά λεξικά μεταβλητού μεγέθους, χωρίς εικόνες περίσπασης. Το φιλτράρισμα γίνεται με το BoW και η ανακατάταξη στις 1K πρώτες εικόνες με το FSM και το HPM.  $P = prior$ .

**Ειδικά λεξικά.** Ο Πίνακας 3.4 συνοψίζει την απόδοση στην βάση Oxford για ειδικά λεξικά μεταβλητού μεγέθους, τα οποία έχουν δημιουργηθεί από όλες τις εικόνες της βάσης Oxford. Το HPM και πάλι έχει καλύτερη απόδοση σε όλες τις περιπτώσεις εκτός από το λεξικό μεγέθους 100K. Η υψηλότερη μας απόδοση χωρίς πρότερη γνώση περιστροφής (0.692) μπορεί επίσης να συγκριθεί με την υψηλότερη απόδοση (0.664) η οποία έχει επιτευχθεί από το 5-dof FSM και ειδικό λεξικό στο [66], παρά το γεγονός ότι εκεί γίνεται χρήση λεξικού μεγέθους 1M και διαφορετικών τοπικών χαρακτηριστικών.

**Επιπλέον βάσεις εικόνων.** Επιστρέφοντας στην περίπτωση του γενικού λεξικού, διεκπεραιώνουμε πειράματα μεγάλης κλίμακας στις βάσεις Oxford και Paris. Εστιάζοντας σε πρακτικούς χρόνους εκτέλεσης χρησιμοποιούμε μόνο το BoW για φιλτράρισμα. Το HPM ξεπερνά ξεκάθαρα το FSM, ενώ η ανακατάταξη 10K εικόνων αισθητά μεγαλώνει την διαφορά σε μεγάλη κλίμακα. Η υψηλότερη απόδοση μας στην βάση Oxford (0.522) μπορεί να συγκριθεί με την αντίστοιχη (0.460) η οποία έχει επιτευχθεί από το FSM στο [69] με γενικό λεξικό μεγέθους 1M το οποίο έχει δημιουργηθεί από την βάση Paris.

Μέθοδος	Oxford		Paris	
	0	2M	0	2M
BoW+HPM10K+P	–	<b>0.418</b>	–	<b>0.419</b>
BoW+HPM10K	–	0.403	–	0.418
BoW+HPM+P	0.546	0.381	0.595	0.402
BoW+HPM	<b>0.522</b>	<b>0.372</b>	<b>0.581</b>	<b>0.397</b>
BoW+FSM	0.503	0.317	0.542	0.336
BoW	0.430	0.201	0.539	0.282

**Πίνακας 3.5:** Σύγκριση mAP στις βάσεις Oxford και Paris με γενικό λεξικό μεγέθους 100K, με και χωρίς 2M εικόνες περίσπασης.  $P = prior$ .

method	Oxford5k	Oxford105k
HPM (this work)	<b>0.789</b>	<b>0.730</b>
Shen <i>et al.</i> [72]	0.752	0.729
Zhang <i>et al.</i> [87] [87]+RANSAC	0.696 0.713	- -
Cao <i>et al.</i> [11] [11]+RANSAC	0.656 0.661	- -
Perdoch <i>et al.</i> [61]	<b>0.789</b>	0.726
FSM [66]	0.647	0.541

**Πίνακας 3.6:** Σύγκριση απόδοσης μέσω mAP με άλλα μοντέλα γεωμετρικού ταιριάσματος στα σύνολα εικόνων Oxford5k και Oxford105k. Τόσο για την δική μας όσο και για τις άλλες μεθόδους έχει δημιουργηθεί ένα ειδικό λεξικό μεγέθους 1M για τις εικόνες του συνόλου Oxford5k.

**Σύγκριση με άλλα μοντέλα.** Προκειμένου να είναι η μέθοδος μας άμεσα συγκρίσιμη και με άλλα δημοσιευμένα μοντέλα για γεωμετρικό ταίριασμα, διεξάγουμε ένα πείραμα χρησιμοποιώντας τον τροποποιημένο ανιχνευτή χαρακτηριστικών Hessian-Affine [61], όπου χρησιμοποιείται η υπόθεση του διανύσματος βαρύτητας (gravity vector assumption) για την εκτίμηση της κύριας κατεύθυνσης των χαρακτηριστικών. Ομοίως με τις δουλείες των [61] και [72], δεν χρησιμοποιούμε την περιστροφή στο μοντέλο μας και διεξάγουμε την διαδικασία του ταιριάσματος πυραμίδας σε έναν χώρο 3 διαστάσεων. Χρησιμοποιούμε ειδικό λεξικό 1M οπτικών λέξεων το οποίο δημιουργήσαμε από όλες τις εικόνες της βάσης Oxford 5k.

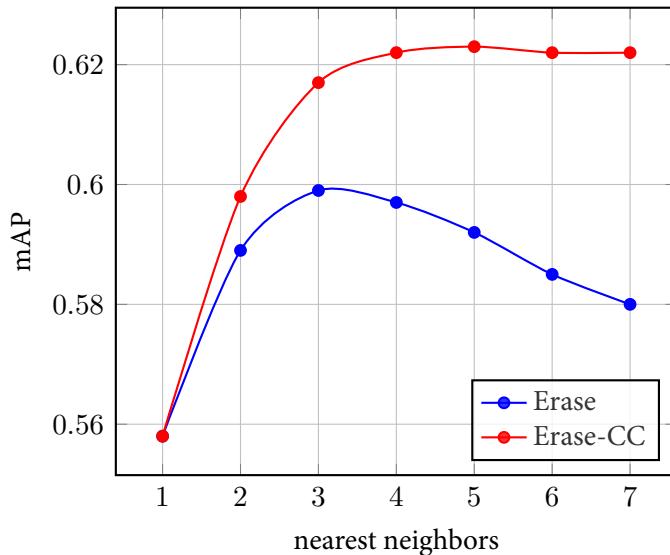
Ο Πίνακας 3.6 παρουσιάζει την απόδοση μέσω mAP για έναν αριθμό μοντέλων γεωμετρικού ταιριάσματος για την βάση Oxford 5k και την Oxford 105k. Η μέθοδος HPM καταλήγει σε πολύ καλά αποτελέσματα και στα δύο σύνολα εικόνων. Ο χρόνος μας για ανακατάταξη 1000 εικόνων είναι 210ms με υλοποίηση ενός πυρήνα, ενώ ο χρόνος που αναφέρεται στο [61] είναι 238ms σε 4 πυρήνες. Αυτός ο χρόνος για το HPM είναι χωρίς την χρήση της πρόωρης εγκατάλειψης της ανακατάταξης επιπλέον εικόνων του [66] ή του [61]. Ο χρόνος ο οποίος αναφέρεται στο [72] είναι 89ms στο Oxford 5k. Όμως εφόσον αυτή είναι μία μέθοδος εξαντλητική πάνω σε όλες τις εικόνες, η πολυπλοκότητα της είναι γραμμική ως προς τον αριθμό των εικόνων στην βάση.

**Επίδραση του σχήματος ένα προς ένα.** Ο βαθμός ομοιότητας που έχουμε ορίσει είναι ένα σταθμισμένο άθροισμα πάνω σε όλες τις αντιστοιχίες δύο εικόνων. Οι αντιστοιχίες με μετασχηματισμούς οι οποίοι πέφτουν εκτός ορίων (έξω από τα όρια του χώρου που κβαντίζεται για το HPM) απορρίπτονται, όπως περιγράφεται στην ενότητα 3.5. Τα αποτελέσματα του Πίνακα 3.7 δείχνουν πως απλά αφαιρώντας τέτοιες αντιστοιχίες υπάρχει ένα μικρό όφελος στην απόδοση. Σε αυτή την περίπτωση, το σύνολο  $X$  της σχέσης (3.8) περιέχει μόνο τις αντιστοιχίες εκτός ορίων, ενώ η ισχύς  $w(c)$  είναι ίση με 1.0. Στην συνέχεια

method	strength	mAP
BoW only	-	0.341
out-of-bound removed	1.0	0.373
one-to-one only	1.0	0.503
HPM	as in (3.7)	0.558

Πίνακας 3.7: Σύγκριση απόδοσης μέσω mAP στο σύνολο εικόνων Barcelona με 2M εικόνες περίσπασης από το World Cities, απεικονίζοντας την θετική επίδραση της διαγραφής των εκτός ορίων αντιστοιχιών, της εφαρμογής του σχήματος ένα προς ένα και της χρήσης της ισχύς η οποία προκύπτει από το ταίριασμα των πυραμίδων Hough. Το τελευταίο είναι ακριβώς ο αλγόριθμος HPM.

εφαρμόζουμε το σχήμα ένα προς ένα με την διαδικασία διαγραφής που χρησιμοποιούμε, αλλά διατηρούμε την ισχύ ίση με 1.0. Το σχήμα ένα προς ένα παρουσιάζεται να βελτιώνει εντυπωσιακά την απόδοση, εφόσον πολλές λανθασμένες αντιστοιχίες δεν θα συμπεριληφθούν καθόλου στον τελικό βαθμό ομοιότητας. Τελικά, με την χρήση της ισχύς η οποία δίνεται από την σχέση (3.7) υπάρχει μία επιπλέον βελτίωση στην απόδοση.



Σχήμα 3.12: Σύγκριση διαφορετικών τρόπων για τον εντοπισμό ασυμβατοτήτων με την μέθοδο HPM. Γίνεται χρήση των κατηγοριών οπτικών λέξεων (ERASE) και των κατηγοριών συνιστωσών (ERASE-CC). Η απόδοση mAP παρουσιάζεται για διαφορετικό αριθμό πλησιέστερων γειτόνων στην πολλαπλή ανάθεση. Αποτελέσματα στην βάση εικόνων Barcelona με 2M εικόνες περίσπασης από το σύνολο World Cities.

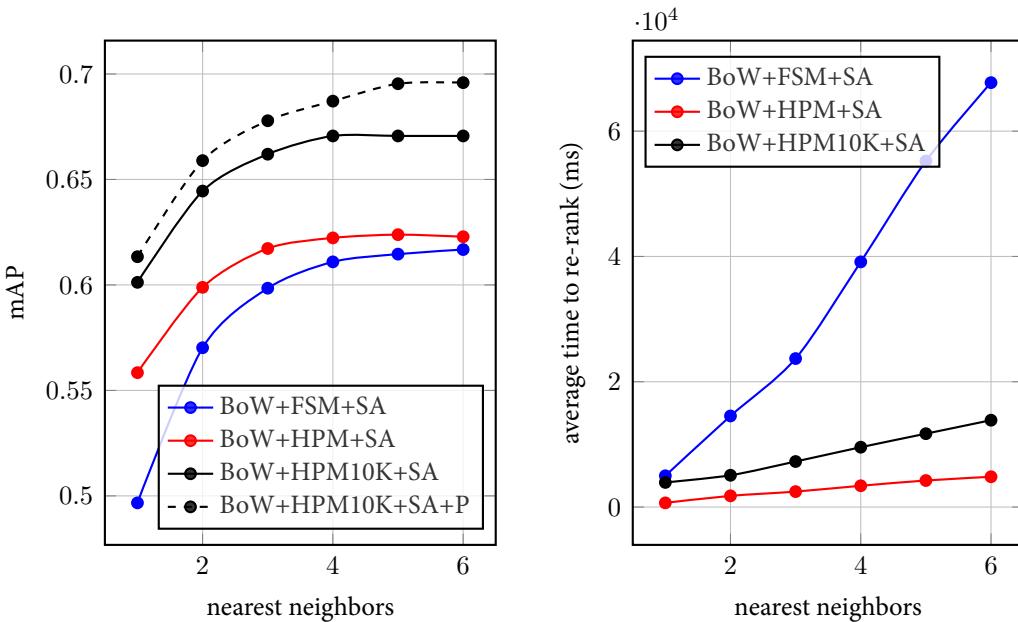
### 3.6.4 Ανακατάταξη με την χρήση πολλαπλής ανάθεσης

Διεξάγουμε πειράματα στην ανάκτηση εικόνων με χρήση πολλαπλής ανάθεσης οπτικών λέξεων [69] από την μεριά της εικόνας ερωτήματος [34]. Διεξάγουμε πειράματα μεγάλης

## Πειράματα

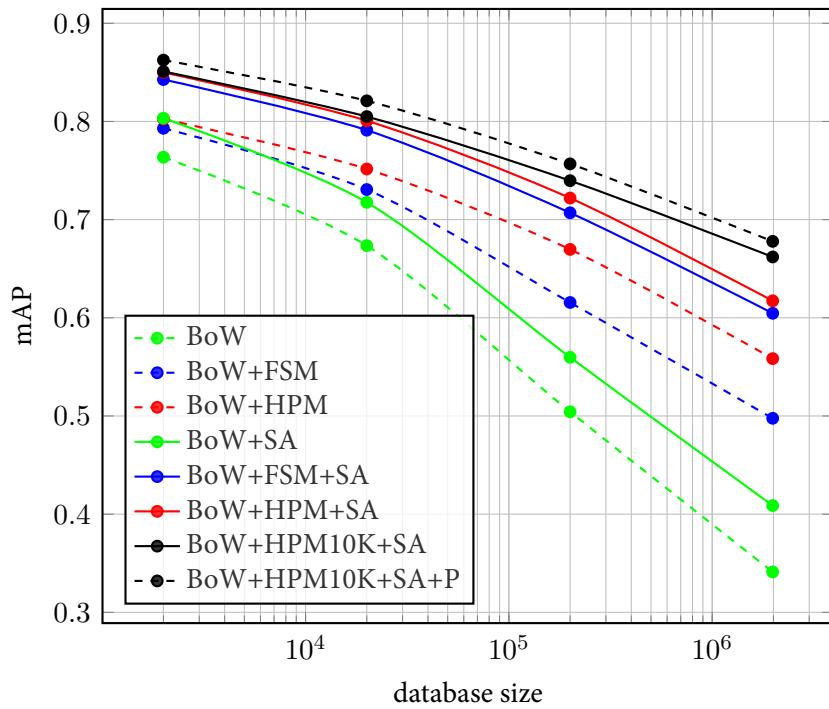
κλίμακας και συγκρίνουμε τις μεθόδους HPM και FSM με την πολλαπλή ανάθεση και στις δύο. Η μέθοδος BoW χρησιμοποιείται για το αρχικό φιλτράρισμα στα παρακάτω πειράματα.

**Εντοπισμός ασυμβατοτήτων.** Συγκρίνουμε τις δύο μεθόδους για εντοπισμό ασυμβατοτήτων, δηλαδή την μία που βασίζεται σε κατηγορίες οπτικών λέξεων (ERASE) και την άλλη που βασίζεται σε κατηγορίες συνιστωσών (ERASE-CC). Τα αποτελέσματα φαίνονται στο Σχήμα 3.12 για διαφορετικό αριθμό πλησιέστερων γειτόνων στην πολλαπλή ανάθεση. Η χρήση των κατηγοριών συνιστωσών ξεπερνά σε απόδοση τις κατηγορίες οπτικών λέξεων. Για την ακρίβεια η απόδοση των τελευταίων πέφτει από ένα σημείο και μετά, γεγονός το οποίο αποδίδεται στην μη τήρηση του σχήματος ένα προς ένα. Στα υπόλοιπα πειράματα στα οποία χρησιμοποιείται πολλαπλή ανάθεση θα χρησιμοποιούμε τις κατηγορίες συνιστωσών.

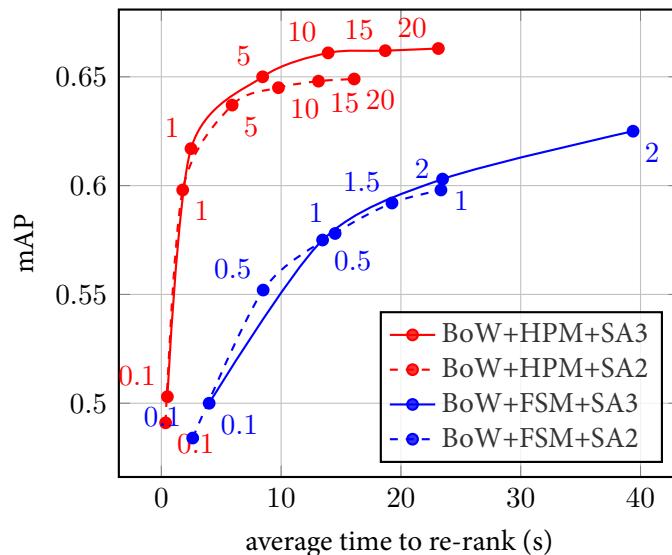


**Σχήμα 3.13:** Απόδοση mAP (αριστερά) και μέσος χρόνος εκτέλεσης ανά εικόνα ερωτήματος (δεξιά) έναντι του αριθμού των οπτικών λέξεων στην πολλαπλή ανάθεση (πλησιέστερων γειτόνων). Μετρήσεις για το σύνολο εικόνων Barcelona μαζί με 2M είκονες περίσπασης από το σύνολο World Cities. SA = soft assignment, P = prior.

**Πλησιέστεροι γείτονες.** Συγκρίνουμε τις μεθόδους HPM και FSM ως προς την απόδοση μέσω mAP και του χρόνου εκτέλεσης για διαφορετικό πλήθος οπτικών λέξεων πολλαπλής ανάθεσης, όπως φαίνεται στο Σχήμα 3.13. Οι δύο μέθοδοι φαίνεται να καταλήγουν σε παρόμοια απόδοση όταν αυξάνουμε τον αριθμό των πλησιέστερων γειτόνων για ανακατάταξη 1K εικόνων. Παρόλα αυτά, η μέθοδος HPM υπερέχει όταν γίνεγαι ανακατάταξη 10K εικόνων, κι ακόμα περισσότερο με την χρήση της πρότερης γνώσης για την περιστρφή. Το HPM παραμένει μία τάξη μεγέθους πιο γρήγορο, ενώ το FSM καταλήγει σε



**Σχήμα 3.14:** Σύγκριση απόδοσης  $mAP$  ως προς το μέγεθος της βάσης, για μεταβλητό αριθμό εικόνων περίσπασης. Μέτρηση στο σύνολο Barcelona με έως 2M εικόνες περίσπασης από το σύνολο World Cities. SA = soft assignment, με 3 πλησιέστερους γείτονες για όλες τις μεθόδους.



**Σχήμα 3.15:** Απόδοση  $mAP$  ως προς τον χρόνο εκτέλεσης για μεταβλητό αριθμό εικόνων ανακατάταξης. Οι τελευταίες απεικονίζονται με κειμενικές επιγραφές δίπλα από τα σημεία της κάθε καμπύλης, σε χιλιάδες. Αποτελέσματα από το σύνολο εικόνων Barcelona με 2M εικόνες περίσπασης από το World Cities, με πολλαπλή ανάθεση και 3 πλησιέστερους γείτονες για όλες τις μεθόδους.

απογορευτικούς χρόνους ακόμα και για 3 πλησιέστερους γείτονες. Ο αριθμός των αντι-

## Πειράματα

στοιχιών είναι σχεδόν γραμμικός ως προς τον αριθμό των οπτικών λέξεων πολλαπλής ανάθεσης όταν αυτή γίνεται από την μεριά της εικόνας ερωτήματος μόνο [69], κι έτσι επιβεβαιώνεται για ακόμα μία φορά ότι η μέθοδος HPM έχει γραμμική πολυπλοκότητα ως προς τον αριθμό των αντιστοιχιών.

**Εικόνες περίσπασης, χρόνος εκτέλεσης.** Ομοίως με την απλή ανάθεση, συγκρίνουμε τις δύο μεθόδους (HPM και FSM) για μεταβλητό αριθμό εικόνων περίσπασης, μέχρι 2M. Η απόδοση μέσω mAP παρουσιάζεται στο Σχήμα 3.14 όπου χρησιμοποιούμε πολλαπλή ανάθεση στις 3 πλησιέστερες οπτικές λέξεις. Φαίνεται και πάλι η υπεροχή της μεθόδου HPM. Η απόδοση mAP έναντι του χρόνου εκτέλεσης για διαφορετικό αριθμό εικόνων οι οποίες ανακατατάσσονται φαίνεται στο Σχήμα 3.15. Ομοίως με την απλή ανάθεση η μέθοδος HPM κάνει ανακατάταξη μια τάξη μεγέθους περισσότερες εικόνες στον ίδιο χρόνο. Γενικά, όλες οι μέθοδοι επωφελούνται 7-13% από την χρήση της πολλαπλής ανάθεσης.

# Κεφάλαιο 4

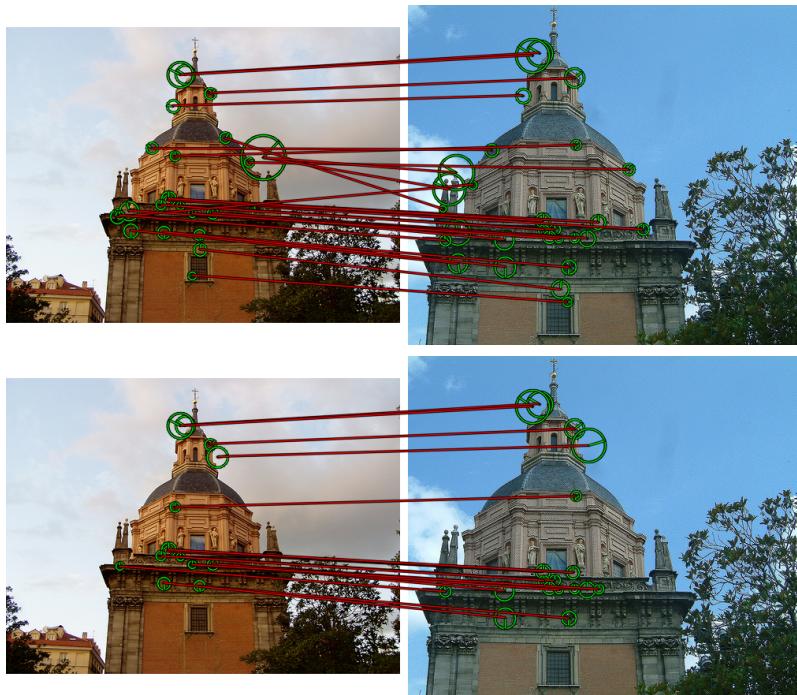
## Επιλογή χαρακτηριστικών μέσω εντοπισμού συμμετριών

### 4.1 Εισαγωγή

Η μέθοδος *bag-of-words* (BoW) με την χρήση τοπικών χαρακτηριστικών και περιγραφέων, μαζί με την χρήση γεωμετρικού ταιριάσματος στην αναζήτηση εικόνων έχει γνωρίσει μεγάλη επιτυχία [73][66]. Σε αναζήτηση μεγαλύτερης κλίμακας οι απατήσεις μνήμης φαίνεται να μετατρέπονται σε περιορισμό και πολύ περισσότερο μετά από τις πρόσφατες προσπάθειες για δεικτοδότηση γεωμετρίας [34]. Μία λύση είναι τελικά κανείς να εγκαταλείψει την αναπαράσταση της μεθόδου BoW και να χρησιμοποιήσει περισσότερο καθολικές αναπαραστάσεις όπως *Fisher kernels* [64][35], αλλά αυτές δεν είναι συμβατές με την εφαρμογή γεωμετρικού ταιριάσματος [66][61], οπότε δεν καταφέρνουν να φτάσουν στο ίδιο επίπεδο απόδοσης.

Έως τώρα, η επιλογή χαρακτηριστικών είναι πιθανόν η μόνη πρακτική εναλλακτική λύση. Μία ήδη διαδεδομένη ιδέα είναι κανείς να επιλέξει χαρακτηριστικά με βάση το ταιριάσμα μεταξύ διαφορετικών εικόνων που απεικονίζουν το ίδιο αντικείμενο [82]. Υποθέτοντας ότι υπάρχουν όντως πολλαπλές όψεις του κάθε αντικειμένου στην βάση, αυτή η ιδέα έχει γίνει δημοφιλής και σε άλλα προβλήματα όπως η εκμάθηση λεξικού [28][57], ο εντοπισμός τοποθεσίας κι αξιοθέατων [71][27] και η 3Δ ανακατασκευή [1]. Στην πράξη, παρόλα αυτά, οι περισσότερες εικόνες είναι οι μοναδικές απεικονίσεις ενός αντικειμένου και δεν υπάρχει κάποια άλλη εικόνα στην βάση με την οποία να μπορούν να ταιριάξουν. Ή μήπως υπάρχει;

Εμείς θα ασχοληθούμε ακριβώς με αυτές τις μοναδικές όψεις, χρησιμοποιώντας την ιδέα της αυτο-ομοιότητας, πόσο δηλαδή μοιάζει η κάθε εικόνα με τον εαυτό της. Αναπτύσσουμε δύο μεθόδους για αυτο-ταιριάσμα, εμπνευσμένες από γνωστές μεθόδους γεωμετρικού ταιριάσματος: *fast spatial matching* (FSM) των Philbin *et al.* [66], και της δικής μας



**Σχήμα 4.1:** Αρχικές αντιστοιχίες με χρήση όλων (πάνω) και των επιλεγμένων (κάτω) τοπικών χαρακτηριστικών. Η επιλογή έχει εφαρμοσθεί μόνο στην εικόνα στα δεξιά.

*Hough pyramid matching* (HPM)(Κεφάλαιο 3) [78]. Τις εφαρμόζουμε είτε ανάμεσα στην εικόνα και τον εαυτό της, είτε ανάμεσα στην εικόνα και το αντικατοπτρισμένο είδωλο της. Τελικά εντοπίζουμε επαναλαμβανόμενες δομές και συμμετρίες. Έτσι επιλέγουμε να κρατήσουμε τις περιοχές οι οποίες συμμετέχουν σε αυτές τις ομάδες. Υποθέτουμε πως χαρακτηριστικά τα οποία κατάφεραν να ταιριάξουν μέσα στην ίδια εικόνα θα είναι ικανά να ταιριάξουν και με άλλες όψεις του ίδιου αντικειμένου όταν αυτό εμφανισθεί σε μία εικόνα ερωτήματος.

Στο παράδειγμα του Σχήματος 4.1, η εικόνα στα αριστερά (δεξιά) θεωρείται εικόνα ερώτημα (βάσης). Η επιλογή χαρακτηριστικών έχει εφαρμοσθεί μόνο στην δεξιά εικόνα. Υπάρχουν αρκετές συμμετρίες οι οποίες βοηθούν να επιλεχθούν χαρακτηριστικά τα οποία θα δώσουν τελικά τις κατάλληλες αντιστοιχίες, οι οποίες φαίνονται και πιο καθαρές από outliers.

## 4.2 Σχετική βιβλιογραφία

Η επιλογή τοπικών χαρακτηριστικών έχει πρόσφατα γίνει δημοφιλής τρόπος ώστε να μειώσει κανείς τις ανάγκες μνήμης των δομών δεικτοδότησης οι οποίες χρησιμοποιούνται στην αναζήτηση εικόνων. Αυτό είναι τυπικά μία διαδικασία η οποία εξαρτάται από την εφαρμογή, και δύο προηγούμενες μέθοδοι έχουν εφαρμοσθεί στην αναγνώριση γεωγραφικής τοποθεσίας. Δεδομένης μιας βάσης εικόνων με γεωγραφικό στίγμα μεγάλης πυκνότητας,

οι Schindler *et al.* [71] επιλέγουν τα τοπικά χαρακτηριστικά με την περισσότερη πληροφορία, π.χ. τοπικά χαρακτηριστικά τα οποία εμφανίζονται συνήθως σε εικόνες μίας συγκεκριμένης γεωγραφικής τοποθεσίας. Παρομοίως, οι Li και Kosecka [51] αποκτούν μία πιθανότητα του πληροφοριακού περιεχομένου για κάθε χαρακτηριστικό με βάση την αναγνώριση τοποθεσίας. Τα χαρακτηριστικά ταξινομούνται και στις δύο μεθόδους ανάλογα και μόνο ένα υποσύνολο τους χρησιμοποιείται.

Οι Knopp *et al.* [45] υπολογίζουν με πυκνό τρόπο ένα τοπικό βαθμό σύγχυσης για περιοχές της εικόνας μίας βάσης εικόνων με γεωγραφικό στίγμα. Αυτό γίνεται με την χρήση ενός παραθύρου ολίσθησης κι έπειτα αφαίρεση των χαρακτηριστικών με υψηλό βαθμό σύγχυσης. Οι Gammeter *et al.* [27] ξεκινάμε από ομάδες εικόνων οι οποίες δημιουργήθηκαν από εγγύτητα στον γεωγραφικό χώρο και οπτική ομοιότητα. Στην συνέχεια χρησιμοποιούν την πληροφορία από το ταίριασμα των εικόνων για να εκτιμήσουν ένα παράθυρο το οποίο να περικλείει το κύριο αντικείμενο της κάθε εικόνας. Παρά το γεγονός ότι διατηρούν μόνο τα χαρακτηριστικά μέσα σε αυτό το παράθυρο, τα περισσότερα εκ των χαρακτηριστικών τελικά θα διατηρηθούν. Οι Avrithis *et al.* [5](Κεφάλαιο 6) επιπλέον κατασκευάζουν έναν χάρτη σκηνών από τα επιλεγμένα χαρακτηριστικά των ευθυγραμμισμένων όψεων, όμως τελικά το αποτέλεσμα είναι μία αύξηση της απόδοσης κι όχι τόσο η συμπίεση μνήμης.

Όλα τα προηγούμενα μοντέλα περιλαμβάνουν επίβλεψη καθώς χρησιμοποιούν την γεωγραφική πληροφορία των εικόνων. Σε αντίθεση οι Turcot και Lowe [82], όπως κι εμείς στην επιλογή περιοχών για τους χάρτες χαρακτηριστικών(Κεφάλαιο 2, προτείνουν μία μέθοδο χωρίς επίβλεψη για να επιλέξουν τα χρήσιμα τοπικά χαρακτηριστικά. Αυτά είναι εκείνα που βρέθηκαν σαν inliers κατά την διαδικασία γεωμετρικού ταιριάσματος, όπου κάθε εικόνα της βάσης χρησιμοποιείται σαν ερώτημα σε ένα σύστημα αναζήτησης. Οι Naikal *et al.* [59] αποφεύγουν την χρήση γεωμετρικού ταιριάσματος με το να χρησιμοποιήσουν ένα σύνολο εικόνων εκπαίδευσης ανά κατηγορία, εφαρμόζουν την μέθοδο PCA στον πίνακα συμμεταβλητών των ιστογραμμάτων BoW. Παρόλα αυτά, η μέθοδος τους εφαρμόζεται σε λεξικά μεγέθους μέχρι  $10^3$ , κάνοντας την λύση αυτή ακατάλληλη για αναζήτηση εικόνων.

Όλα τα μοντέλα που αναφέρθηκαν μέχρι στιγμής απαιτούν πολλαπλές όψεις του ίδιο αντικειμένου ή σκηνής. Για τις εικόνες οι οποίες είναι οι μοναδικές όψεις ενός αντικειμένου, οι Turcot and Lowe [82] διατηρούν τα τοπικά χαρακτηριστικά με την μεγαλύτερη κλίμακα, το οποίο είναι ισοδύναμο με το να μειώνει κανείς την διάσταση της εικόνας. Στην ενότητα 5.7 δείχνουμε ότι αυτή η μέθοδος αποτυγχάνει με την παρουσία πολλών εικόνων περίσπασης. Στην δική μας προσέγγιση εντοπίζουμε τις αυτο-ομοιότητες, τις επαναλαμβανόμενες δομές και τις συμμετρίες για να επιλέξουμε χαρακτηριστικά από αυτές τις μοναδικές όψεις.

Προηγούμενες μέθοδοι για εντοπισμό συμμετριών θεωρούν όλη την εικόνα σαν ένα σήμα και αναζητούν έναν ή περισσότερους άξονες συμμετρίας [76][44]. Πιο πρόσφατες μέθοδοι δίνουν έμφαση στην χρήση τοπικών χαρακτηριστικών. Οι Tuytelaars *et al.* [83] ξεκινούν με αναλλοίωτες γειτονιές κι εφαρμόζουν διαδοχικούς μετασχηματισμούς Hough για να εντοπίσουν συνευθειακά σημεία τομής, μέθοδος με μεγάλη πολυπλοκότητα. Οι Cornelius *et al.* [21] χρησιμοποιούν επίσης ψήφους στον χώρο του μετασχηματισμού Hough αλλά χωρίς να απαιτούν τοπικούς περιγραφείς, εντοπίζουν *local affine frames* (LAFs) από μία εικόνα και το αντικατοπτρισμένο είδωλο της, τις ταιριάζουν με χρήση τοπικής γεωμετρίας, κι έπειτα επιτρέπουν κάθε ταιριασμένο ζευγάρι LAF να ψηφίσει για έναν άξονα συμμετρίας. Ομοίως, οι Loy and Eklundh [54] εντοπίζουν συμμετρικούς αστερισμούς από τοπικά χαρακτηριστικά με χρήση μετασχηματισμού Hough.

Οι Lazebnik *et al.* [48] εξαπλώνουν τους σχηματισμούς των τοπικών χαρακτηριστικών προς γεωμετρικά αναλλοίωτα τμήματα 3Δ αντικειμένων. Για να περιορίσουν τον εκθετικό αριθμό των υποθέσεων εφαρμόζουν αυστηρούς γεωμετρικούς περιορισμούς στους αρχικούς σχηματισμούς. Παρόλα αυτά η πολυπλοκότητα παραμένει τετραγωνική ως προς τον αριθμό των αντιστοιχιών. Δείχνουμε ότι το πρόβλημα του εντοπισμού συμμετριών μπορεί να λυθεί με οποιαδήποτε μέθοδο γεωμετρικού ταιριάσματος, κι έτσι αποκομίζουμε τις δύο προτεινόμενες μεθόδους. Η μία με τετραγωνική [66], και η άλλη με γραμμική πολυπλοκότητα [78] ως προς τον αριθμό των αντιστοιχιών.

## 4.3 Επιλογή χαρακτηριστικών

Παρουσιάζουμε εδώ δύο εναλλακτικές μεθόδους για επιλογή χαρακτηριστικών, οι οποίες εστιάζουν στις γεωμετρικά συμβατές ομάδες που εμφανίζονται μέσα στην ίδια εικόνα και αντιστοιχούν σε επαναλαμβανόμενες δομές ή τοπικές συμμετρίες. Οι δύο μέθοδοι αναλύονται στις ενότητες 4.3.2 και 4.3.4. Κατά την εφαρμογή και των δύο μεθόδων, η κάθε εικόνα ταιριάζεται με τον εαυτό της και με το αντικατοπτρισμένο είδωλο της, ώστε να διαχειριστούμε ευθείς αλλά κι ανεστραμμένους μετασχηματισμούς, αντίστοιχα. Οι δύο διαδικασίες στις οποίες θα αναφερόμαστε ως ευθύ και ανάστροφο ταίριασμα, παρουσιάζονται στις ενότητες 4.3.1 και 4.3.3, αντίστοιχα.

Όλη η μέθοδος μας βασίζεται στην παρατήρηση ότι ο εντοπισμός συμμετριών ή επαναλαμβανόμενων δομών σε μία εικόνα δεν διαφέρει και πολύ από το γεωμετρικό ταίριασμα μεταξύ δύο εικόνων. Στην πραγματικότητα, εντοπίζουμε δύο θέματα που απαιτούν προσοχή:

- Το γεωμετρικό ταίριασμα συνήθως (αλλά όχι πάντα) υποθέτει αντιστοιχίες ένα προς ένα ανάμεσα στα τοπικά χαρακτηριστικά δύο εικόνων. Για αυτό ακριβώς τον λόγο

περιορισμοί όπως ο έλεγχος λόγου (ratio test) [53] χρησιμοποιούνται όταν ταιριάζουμε τοπικούς περιγραφείς. Εμείς αντιμετωπίζουμε διαφορετική κατάσταση καθώς μία δομή μπορεί να επαναλαμβάνεται περισσότερες από μία φορές.

2. Αναζητώντας ευρωστία ως προς τους outliers, συνήθως χρησιμοποιείται η υπόθεση ενός μοναδικού γεωμετρικού μοντέλου, π.χ. μετασχηματισμός ομοιότητας ή ομογραφίας. Αυτός δεν είναι ο κανόνας καθώς η μέθοδος HPM [78] για παράδειγμα μπορεί να εντοπίσει πολλαπλούς μετασχηματισμούς. Για κάθε συμμετρία ή επαναλαμβανόμενη δομή χρειαζόμαστε όντως διαφορετικό μετασχηματισμό: η μέθοδος HPM είναι πιο κατάλληλη με αυτή την έννοια, αλλά επίσης επεκτείνουμε την μέθοδο FSM [66] προς αυτή την κατεύθυνση.

### 4.3.1 Αναπαράσταση

Υποθέτουμε ότι μία εικόνα αναπαρίσταται από ένα σύνολο τοπικών χαρακτηριστικών  $X$ . Το κάθε τοπικό χαρακτηριστικό  $x \in X$  σχετίζεται με έναν  $D$ -διάστατο τοπικό περιγραφέα  $d(x) \in \mathbb{R}^D$ , ο οποίος κωδικοποιεί την τοπική εμφάνιση της εικόνας, κι ο οποίος έπειτα κβαντίζεται στην οπτική λέξη  $w(x) \in W$ , όπου  $W$  είναι ένα δεδομένο οπτικό λεξικό. Υποθέτουμε ότι τα τοπικά χαρακτηριστικά είναι αναλλοίωτα ως προς την κλιμάκωση και την περιστροφή, τότε κάθε χαρακτηριστικό σχετίζεται επίσης και με την θέση  $p(x) \in \mathbb{R}^2$  στο επίπεδο της εικόνας, την λογαριθμιμένη κλίμακα  $\sigma(x) \in \mathbb{R}$  και την περιστροφή  $\theta(p) \in (-\pi, \pi]$ . Όλη αυτή η πληροφορία αναπαρίσταται με βολικό τρόπο στον διάνυσμα

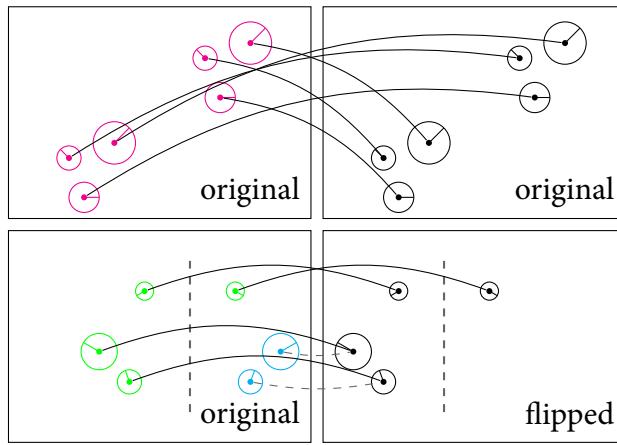
$$g(x) = [p(x)^T \ \sigma(x) \ \theta(x)]^T \quad (4.1)$$

κωδικοποιώντας έτσι την τοπική γεωμετρία. Η λογαριθμιμένη κλίμακα χρησιμοποιείται έτσι ώστε η σχετική κλιμάκωση να εκφράζεται σαν διαφορά κι όχι σαν λόγος. Η περιστροφή αναπαρίσταται επίσης από τον ορθογώνιο πίνακα  $R(x) \in \mathbb{R}^{2 \times 2}$  με  $\det R(x) = 1$ . Τελικά, η θέση αναπαρίσταται επίσης από τις ομογενείς συντεταγμένες με το διάνυσμα  $p(x)$  στον προβολικό χώρο  $\mathbb{P}^2(\mathbb{R})$ .

Μία αντιστοιχία χαρακτηριστικών σε μία εικόνα  $X$  είναι ένα ζευγάρι χαρακτηριστικών  $c \in X^2$ . Δεδομένης της αντιστοιχίας  $c = (x, y)$ , είναι πιθανό να ορίσουμε τον μετασχηματισμό ομοιότητας ο οποίος ευθυγραμμίζει τις δύο περιοχές  $x, y \in X$  στο επίπεδο της εικόνας. Φαίνεται πως [78] αυτός ο μετασχηματισμός μπορεί επίσης να αναπαρασταθεί από το διάνυσμα

$$g(x, y) = g(c) = [p(c)^T \ \sigma(c) \ \theta(c)]^T \quad (4.2)$$

όπου  $\sigma(c) = \sigma(y) - \sigma(x)$ ,  $\theta(c) = \theta(y) - \theta(x)$  είναι η σχετική κλιμάκωση και σχετική περιστροφή αντίστοιχα, και  $p(c) = p(y) - M(c)p(x)$  είναι η σχετική μετατόπιση, με  $M(c) = \sigma(c)R(c)$  και  $R(c) = R(y)R(x)^{-1}$ .



**Σχήμα 4.2:** (Πάνω) επαναλαμβανόμενη δομή η οποία εντοπίζεται από ευθύ ταίριασμα, ματζέντα: επιλογή με ευθύ ταίριασμα. (Κάτω) συμμετρία η οποία εντοπίζεται με ανάστροφο ταίριασμα, πράσινο: επιλογή με ανάστροφο ταίριασμα, κυανό: επιλογή μετά από προβολή στην αρχική εικόνα. (Αριστερά) αρχική εικόνα. (Δεξιά) εικόνα προς ταίριασμα, όλα τα χαρακτηριστικά σε μαύρο χρώμα. Συνεχείς γραμμές: αρχικές αντιστοιχίες, διακεκομμένες γραμμές: προβολή πίσω στην αρχική εικόνα + άξονες συμμετρίας.

Είναι τώρα δυνατόν να χρησιμοποιήσουμε την νόρμα  $\|g(x, y)\|$  για να μετρήσουμε την γεωμετρική εγγύτητα των περιοχών  $x, y$ . Επειδή οι ανιχνευτές τοπικών χαρακτηριστικών συχνά απαντάνε με πολλαπλές αλληλεπικαλυπτόμενες περιοχές, χρειάζεται να εξαρέσουμε κοντινά τοπικά χαρακτηριστικά από αυτό το μέτρο γεωμετρικής εγγύτητας και την αναζήτηση μας για συμμετρίες. Συγκεκριμένα, λέμε πως το  $(x, y)$  είναι ένα έγκυρο ζευγάρι και γράφουμε  $v(x, y)$  έαν και μόνο εάν  $\|g(x, y)\| \geq \rho$ , με  $\rho > 0$ . Έπειτα, δεδομένης της εικόνας  $X$ , ορίζουμε το σύνολο των έγκυρων αντιστοιχιών της

$$C_v(X) = \{(x, y) \in X^2 : v(x, y)\}. \quad (4.3)$$

Από το σύνολο αυτό επίσης εξαιρούνται οι τετριμμένες αντιστοιχίες  $(x, x)$  του χαρακτηριστικού  $x$  με τον εαυτό του, εφόσον  $\|g(x, x)\| = 0$ .

Τώρα οι αντιστοιχίες χαρακτηριστικών βασίζονται στην εμφάνιση τους. Ο απλούστερος τρόπος είναι με την χρήση των οπτικών λέξεων, όπου δύο τοπικά χαρακτηριστικά θα βρίσκονται σε αντιστοιχία όταν αντιστοιχίζονται στην ίδια οπτική λέξη, η οποία είναι η κβαντισμένη εκδοχή του περιγραφέα τους. Η χρήση ενός οπτικού λεξικού εισάγει κι απώλειες κβαντισμού, κι επειδή η επιλογή χαρακτηριστικών είναι μία διαδικασία η οποία εφαρμόζεται εκ των προτέρων, έτσι κι εμείς χρησιμοποιούμε απευθείας ταίριασμα των περιγραφέων για να δημιουργήσουμε τις αντιστοιχίες. Επιπλέον, εφόσον δεν έχει γίνει ανάθεση οπτικών λέξεων πριν την επιλογή χαρακτηριστικών, είναι πιο γρήγορο να ψάξει κανείς για τους πλησιέστερους γείτονες μέσα στο ίδιο το σύνολο  $X$  παρά σε ολόκληρο το λεξικό το οποίο είναι μεγέθους της τάξεως του  $10^6$ .

Συγκεκριμένα, ορίζουμε την ανομοιότητα  $d(x, y)$  των χαρακτηριστικών  $x, y \in X$  ως την απόσταση τους  $\|d(x) - d(y)\|$  στον χώρο περιγραφέων  $\mathbb{R}^D$ . Λέμε πως τα χαρακτηριστικά  $x, y$  είναι όμοια εάν  $d(x, y) \leq \delta$ , με κατώφλι ομοιότητας  $\delta > 0$ . Τώρα, δεδομένου ενός χαρακτηριστικού  $x \in X$ , έστω η γειτονιά του  $N(x)$  να είναι το σύνολο των όμοιων χαρακτηριστικών τα οποία περιορίζονται από το σύνολο  $\mathcal{N}_X^k(x)$  των  $k$ -πλησιέστερων γειτόνων στο  $X$  σύμφωνα με την ανομοιότητα  $d$ ,

$$N(x) = \{y \in X : y \in \mathcal{N}_X^k(x) \wedge d(x, y) \leq \delta\}. \quad (4.4)$$

Χρησιμοποιούμε τους πλησιέστερους γείτονες για να μειώσουμε τον αριθμό των αντιστοιχιών, κάνοντας έτσι την επακόλουθη διαδικασία ταιριάσματος πιο γρήγορη. Έπειτα, το σύνολο των αντιστοιχιών βασισμένων στους περιγραφείς για την εικόνα  $X$  περιέχει όλα τα ζευγάρια  $x \in X$  με τους γείτονες τους,

$$C_d(X) = \{(x, y) \in X^2 : y \in N(x)\}, \quad (4.5)$$

Παρατηρούμε πως η σχέση  $C_d(X)$  δεν είναι συμμετρική. Εφόσον ψάχνουμε για επαναλαμβανόμενες δομές δεν εφαρμόζουμε ένα προς ένα αντιστοιχίες. Οι ορισμοί (4.4), (4.5) επιτρέπουν μέχρι  $k$  επαναλήψεις.

Τελικά το σύνολο των αρχικών αντιστοιχιών του  $X$  περιέχει ζευγάρια όμοιων χαρακτηριστικών τα οποία είναι κι έγκυρα,

$$C_t(X) = C_d(X) \cap C_v(X). \quad (4.6)$$

Όλοι οι ορισμοί αντιστοιχιών μέχρι τώρα αναφέρονται σε ευθύ ταίριασμα, και θα τροποποιηθούν κατάλληλα για το ανάστροφο ταίριασμα στην ενότητα 4.3.3. Το ευθύ ταίριασμα απεικονίζεται στο Σχήμα 4.2(πάνω), δείχνοντας μία δομή τριών χαρακτηριστικών η οποία επαναλαμβάνεται δύο φορές ενώ έχει υποστεί μετατόπιση. Παρατηρούμε πως οι αρχικές αντιστοιχίες οι οποίες είναι έγκυρες εξαιρούν εκείνες μεταξύ του χαρακτηριστικού και του εαυτού του στις δύο εικόνες. Οι δύο ομάδες των τριών αντιστοιχιών είναι ουσιαστικά οι ίδιες, όντας της μορφής  $(x, y)$  και  $(y, x)$ . Παρόλα αυτά, οι αντιστοιχίες δεν είναι πάντα συμμετρικές. Η δομή θα μπορούσε να επαναλαμβάνεται τρεις οι περισσότερες φορές, ενώ όλοι οι μετασχηματισμοί ομοιότητας επιτρέπονται.

### 4.3.2 Γεωμετρικό αυτο-ταίριασμα

Προκειμένου να εντοπίσουμε τις τοπικές συμμετρίες και τις επαναλαμβανόμενες δομές, χρειαζόμαστε περισσότερες ενδείξεις από μεμονωμένες αντιστοιχίες. Ακολουθούμε δύο διαφορετικές προσεγγίσεις και οι δύο εμπνευσμένες από μεθόδους γεωμετρικού ταιριάσματος μεταξύ δύο διαφορετικών εικόνων, τις οποίες παρόλα αυτά εφαρμόζουμε στο ταίριασμα μίας εικόνας με τον εαυτό της. Η πρώτη προσέγγιση μας αναζητεί ομάδες από

γεωμετρικά επιβεβαιωμένες αντιστοιχίες, τους inliers, και είναι εμπνευσμένη από το γρήγορο γεωμετρικό ταίριασμα (*fast spatial matching* - FSM) [66].

Δεδομένης μίας εικόνας  $X$ , κάθε αντιστοιχία  $c = (x, y) \in C_t(X)$  εγείρει έναν μετασχηματισμό ομοιότητας ο οποίος αναπαρίσταται από το διάνυσμα  $g(c)$  όπως ορίζεται στην σχέση (4.2). Αυτός ο μετασχηματισμός έχει τέσσερις βαθμούς ελευθερίας και μπορεί επίσης να αναπαρασταθεί από έναν πίνακα  $t(c) \in \mathbb{R}^{3 \times 3}$ , με

$$t(x, y) = t(c) = \begin{bmatrix} M(c) & p(c) \\ \mathbf{0}^T & 1 \end{bmatrix}. \quad (4.7)$$

Αυτή η εναλλακτική αναπαράσταση είναι χρήσιμη όταν η θέση αναπαρίσταται σε ομογενείς συντεταγμένες. Δεδομένου του μετασχηματισμού  $h = t(c) = t(x, y)$ , η θέση  $\mathbf{p}(z) \in \mathbb{R}^3$  του  $z \in X$  μετασχηματίζεται στην  $h\mathbf{p}(z)$ .

Η μέθοδος FSM είναι μία διαδικασία ταιριάσματος παρόμοια με του RANSAC. Δεδομένης της εικόνας  $X$  και του αρχικού συνόλου αντιστοιχιών  $C = C_t(X)$ , κάθε  $c \in C$  ορίζει μία υπόθεση μετασχηματισμού  $h = t(c)$  η οποία επιβεβαιώνεται ελέγχοντας για inliers  $I_C(h)$  ανάμεσα σε όλες τις αντιστοιχίες  $(x, y) \in C$ , με

$$I_C(h) = \{(x, y) \in C : \|\mathbf{p}(y) - h\mathbf{p}(x)\| < \epsilon\} \quad (4.8)$$

με  $h \in \mathbb{R}^{3 \times 3}$ . Το σύνολο των inliers βασίζεται στο κατώφλι  $\epsilon > 0$ , το οποίο δίνεται σε pixels. Ανάμεσα σε όλες τις υποθέσεις  $t(C) = \{t(c) : c \in C\}$ , το FSM αναζητά την υπόθεση  $h \in t(C)$  με τους περισσότερους inliers  $|I_C(h)|$ . Εδώ είναι και το σημείο στο οποίο διαφοροποιούμαστε: αναζητούμε την καλύτερη υπόθεση για κάθε ξεχωριστό inlier.

Συγκεκριμένα, για κάθε αντιστοιχία  $c = (x, y) \in C$ , ορίζουμε το σύνολο των υποθέσεων με τις οποίες αυτή συσχετίζεται  $H_C(c) = H_C(x, y) \subseteq t(C)$  και την ευθυγραμμίζουν ως inlier,

$$H_C(x, y) = \{h \in t(C) : \|\mathbf{p}(y) - h\mathbf{p}(x)\| < \epsilon\}. \quad (4.9)$$

μπορούμε τώρα να ορίσουμε την ισχύ ενός inlier  $\alpha(c)$  της αντιστοιχίας  $c$  ως τον μεγαλύτερο αριθμό inlier  $|I_C(h)|$  από όλες τις συσχετισμένες με αυτήν υποθέσεις  $h \in H_C(c)$ ,

$$\alpha_C(c) = \max\{|I_C(h)| : h \in H_C(c)\}. \quad (4.10)$$

Όλη η διαδικασία για το αυτο-ταίριασμα συνοψίζεται στον Αλγόριθμο 4, στον οποίο θα αναφερόμαστε ως χωρικό αυτο-ταίριασμα (spatial self-matching - SSM). Ο παραδοσιακός αλγόριθμος [66] έχει τετραγωνική πολυπλοκότητα ως προς το πλήθος των αντιστοιχιών. Για να επιταχύνουμε αυτή την διαδικασία ταιριάσματος παραλείπουμε τις υποθέσεις οι οποίες προκύπτουν από αντιστοιχίες οι οποίες έχουν ήδη θεωρηθεί ως inliers σε προηγούμενες υποθέσεις (γραμμή 16). Παρατηρήσαμε πως αυτό δεν επηρεάζει την αποτελεσματικότητα της επιλογής χαρακτηριστικών στην πράξη. Τώρα η διαδικασία έχει τετραγωνική πολυπλοκότητα στην χειρότερη περίπτωση, όταν δηλαδή δεν εντοπίζονται καθόλου inliers, αλλά στην πράξη έχουν υπολογιστικό πλεονέκτημα.

---

**Algorithm 4** Spatial self-matching (SSM)

---

**procedure**  $\alpha \leftarrow \text{SSM}(C, t; \tau_\alpha)$

```

input      : correspondences  $C$ , transformations  $t$ 
parameter: inlier threshold  $\tau_\alpha$ 
output    : inlier strengths  $\alpha$ 

11 for  $c \in C$  do                                 $\triangleright$  initialize
12    $inlier(c) \leftarrow \text{FALSE}$                  $\triangleright$  mark as outlier
13    $\alpha(c) \leftarrow 0$                            $\triangleright$  zero strength

14 for  $c \in C$  do                                 $\triangleright$  for all hypotheses
15   if  $inlier(c)$  then continue             $\triangleright$  skip hypothesis?
16
17    $h \leftarrow t(c)$                              $\triangleright$  current hypothesis
18    $I \leftarrow I_C(h)$                            $\triangleright$  current inliers (4.8)
19   if  $|I| < \tau_\alpha$  then continue         $\triangleright$  verified hypothesis?

20
21   for  $c' \in I$  do                       $\triangleright$  for all inliers
22      $inlier(c') \leftarrow \text{TRUE}$            $\triangleright$  mark as inlier
23      $\alpha(c') \leftarrow \max(\alpha(c'), |I|)$    $\triangleright$  update strength

24 return  $\alpha$                                  $\triangleright$  inlier strengths

```

---

Όταν έχει υπολογισθεί η ισχύς για όλες τις αντιστοιχίες, το σύνολο των γεωμετρικά επιβεβαιωμένων αντιστοιχιών  $\alpha(C) \subseteq C$  είναι

$$\alpha(C) = \{c \in C : \alpha_C(c) \geq \tau_\alpha\}, \quad (4.11)$$

με το κατώφλι  $\tau_\alpha > 0$ . Τελικά, δεδομένης μίας εικόνας  $X$  με αρχικές αντιστοιχίες  $C = C_t(X)$ , επιλέγουμε εκείνα τα χαρακτηριστικά  $x \in X$  τα οποία συμμετέχουν σε αρκετές επιβεβαιωμένες υποθέσεις στο  $\alpha(C)$ ,

$$\alpha_d(X) = \pi^1(\alpha(C)) \cup \pi^2(\alpha(C)), \quad (4.12)$$

όπου, για  $i = 1, 2$ ,  $\pi^i(S)$  είναι η  $i$ -οστή προβολή της δυαδικής σχέσης  $S \subseteq X_1 \times X_2$ , η οποία συλλέγει το  $i$ -οστό στοιχείο όλα της τα ζευγάρια,

$$\pi^i(S) = \{x^i \in X_i : (x^1, x^2) \in S\}. \quad (4.13)$$

Ονομάζουμε το  $\alpha_d(X)$  επιλογή χαρακτηριστικών στο  $X$  με ευθύ ταίριασμα.

### 4.3.3 Ανάστροφο ταίριασμα

Μέχρι τώρα, έχουμε μόνο θεωρήσει τους ευθείς μετασχηματισμούς ομοιότητας, αυτό σημαίνει πως υπάρχει μία υπόθεση  $h$  με  $\det h > 0$ . Τί συμβαίνει όμως με τους μετασχηματισμούς με  $\det h < 0$ , όπως οι αντικατοπτρισμοί; Στην πραγματικότητα, όταν αντικατοπτρίζεται μία εικόνα, τότε και η κάθε τοπική περιοχή αντικατοπτρίζεται, κι ο κάθε περιγραφέας δεν είναι ίδιος πια, εκτός εάν η εικόνα είναι εκ φύσεως συμμετρική. Έτσι το να αναστρέψει κανείς την πληροφορία για την τοπική γεωμετρία (4.1) δεν είναι αρκετό: πρέπει να αντικατοπτρίσουμε ολόκληρη την εικόνα και να εξάγουμε εξαρχής τοπικά χαρακτηριστικά και περιγραφείς.

Οποιοσδήποτε αντικατοπτρισμός θα ήταν αρκετός. Επίσης επιλέγουμε αντικατοπτρισμό ως προς τον κατακόρυφο άξονα. Έστω  $Y$  το σύνολο των τοπικών χαρακτηριστικών το οποίο έχει εξαχθεί από την αντικατοπτρισμένη εικόνα. Υποθέτουμε πως το κάθε χαρακτηριστικό  $y \in Y$  έχει ένα αντικατοπτρισμένο αντίστοιχο  $y'$ , το οποίο έχει προβληθεί στην αρχική εικόνα:

$$g(y') = [w - p^1(y) \ p^2(y) \ \sigma(y) \ \pi - \theta(y)]^T, \quad (4.14)$$

όπου  $w$  είναι το πλάτος της εικόνας και  $p^i(y), i = 1, 2$  είναι οι συντεταγμένες του  $y$  πάνω στην εικόνα. Δεν υπάρχει διαθέσιμος περιγραφέας για αυτές τις περιοχές στην αρχική εικόνα πριν την επιλογή που θα ακολουθήσει.

Οι αρχικές αντιστοιχίες σχηματίζονται ακριβώς όπως στο αυτό-ταίριασμα, αλλά τώρα ορίζονται ανάμεσα στα τοπικά χαρακτηριστικά  $x \in X, y \in Y$  της αρχικής και της αντικατοπτρισμένης εικόνας. Τα σύνολα των έγκυρων, των βασισμένων στην εμφάνιση και των αρχικών αντιστοιχιών (4.3), (4.5), (4.6) τροποποιούνται τώρα ως εξής

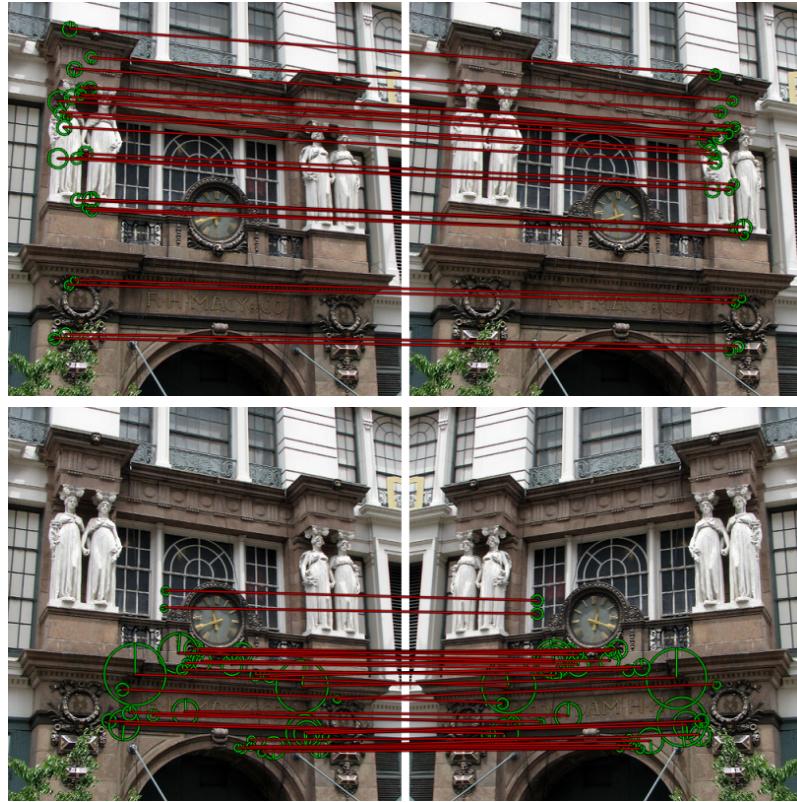
$$C_v(X, Y) = \{(x, y) \in X \times Y : v(x, y')\}, \quad (4.15)$$

$$C_d(X, Y) = \{(x, y) \in X \times Y : y \in N(x)\}, \quad (4.16)$$

$$C_t(X, Y) = C_d(X, Y) \cap C_v(X, Y). \quad (4.17)$$

Παρατηρούμε πως τα έγκυρα ζευγάρια τώρα αναφέρονται στην ίδια εικόνα, οπότε χρησιμοποιούμε την περιοχή  $y'$  η οποία έχει προβληθεί πίσω στην αρχική εικόνα αντί για την  $y$  στην σχέση (4.15).

Δεδομένου του συνόλου  $C_f = C_t(X, Y)$  των αρχικών αντιστοιχιών ανάμεσα στα σύνολα  $X$  και  $Y$ , η διαδικασία SSM που περιγράφηκε στην ενότητα 4.3.2 παραμένει ακριβώς η ίδια. Έστω πως  $\alpha(C_f)$  είναι το τελικό σύνολο των γεωμετρικά επιβεβαιωμένων αντιστοιχιών. Το δεύτερο σύνολο χαρακτηριστικών το οποίο επιλέγουμε είναι μέσω της αντικατοπτρισμένης επιλογής, και περιέχει αυτά τα τοπικά χαρακτηριστικά του  $X$  τα οποία συμμετέχουν σε μία επιβεβαιωμένη αντιστοιχία στο  $\alpha(C_f)$  και δεν βρίσκονται κοντά σε κάποιο



**Σχήμα 4.3:** Πάνω: δείγμα από ομάδα *inliers* οι οποίοι εντοπίσθηκαν με τον αλγόριθμο SSM στην αρχική εικόνα, και αντιστοιχούν σε επαναλαμβανόμενη δομή. Κάτω: δείγμα από ομάδα *inliers* οι οποίοι εντοπίσθηκαν με τον αλγόριθμο SSM ανάμεσα στην αρχική εικόνα και την αντικατοπτρισμένη.

από τα χαρακτηριστικά του  $\alpha_d(X)$  το οποίο να έχει επιλεχθεί,

$$\alpha_f(X) = \pi^1(\alpha(C_f)) \setminus_v \alpha_d(X). \quad (4.18)$$

Με  $A \setminus_v B$  συμβολίζουμε εκείνα τα χαρακτηριστικά του  $A$  τα οποία είναι έγκυρα όσον αφορά το  $B$ ,

$$A \setminus_v B = \{a \in A : v(a, b) \text{ for all } b \in B\}. \quad (4.19)$$

Το τρίτο σύνολο χαρακτηριστικών, εκείνων που επιλέγονται μέσω της προβολής πίσω στην αρχική εικόνα, περιέχει εκείνα τα τοπικά χαρακτηριστικά του  $Y$  τα οποία συμμετέχουν σε μία επιβεβαιωμένη αντιστοιχία. Στην πραγματικότητα χρησιμοποιούμε το αντίστοιχο χαρακτηριστικό που έχει προβληθεί πίσω σε αυτή την περίπτωση, αγνοώντας και πάλι τα κοντινά επιλεγμένα χαρακτηριστικά στο  $\alpha_d(X) \cup \alpha_f(X)$ ,

$$\alpha'_f(X) = [\pi^2(\alpha(C_f))]' \setminus_v (\alpha_d(X) \cup \alpha_f(X)), \quad (4.20)$$

όπου  $A' = \{y' : y \in A\}$  συμβολίζει την προβολή του συνόλου  $A \subseteq Y$  πίσω στην αρχική εικόνα. Είναι πλέον η στιγμή να εξάγουμε τοπικούς περιγραφείς από την αρχική εικόνα και

## Επιλογή χαρακτηριστικών



**Σχήμα 4.4:** Αρχικά (αριστερά) κι επιλεγμένα (δεξιά) τοπικά χαρακτηριστικά με τον αλγόριθμο SSM. Οι αρχικές, αντικατοπτρισμένες και με προβολή στην αρχική εικόνα επιλογές απεικονίζονται με κόκκινο, πράσινο και μπλε χρώμα αντίστοιχα.

για τα επιλεγμένα τοπικά χαρακτηριστικά τα οποία προβλήθηκαν πίσω σε αυτή. Τελικά, το ολοκληρωμένο σύνολο των επιλεγμένων χαρακτηριστικών περιέχει τα χαρακτηριστικά μέσω ευθείας, ανάστροφης και με προβολή στην αρχική εικόνα επιλογής,

$$\alpha(X) = \alpha_d(X) \cup \alpha_f(X) \cup \alpha'_f(X). \quad (4.21)$$

Θα λέμε απλώς ότι το  $\alpha(X)$  είναι το σύνολο των επιλεγμένων χαρακτηριστικών από το  $X$ . Είναι τα μόνα χαρακτηριστικά τα οποία πρέπει να αντιστοιχηθούν σε οπτική λέξη και να δεικτοδοτηθούν.

Το ανάστροφο ταίριασμα απεικονίζεται στο Σχήμα 4.2(κάτω), δείχνοντας μία δομή με έξι περιοχές οι οποίες είναι συμμετρικές ως προς κατακόρυφο άξονα συμμετρίας. Τέσσερις από τις περιοχές (πράσινες) εντοπίστηκαν στην αρχική εικόνα, σχηματίζοντας την ανάστροφη επιλογή, και τέσσερα (μαύρα) στην αντικατοπτρισμένη εικόνα, οι δύο ομάδες έχουν δύο κοινές περιοχές οι οποίες εντοπίσθηκαν και στις δύο εικόνες. Οι δύο περιοχές οι οποίες εντοπίσθηκαν μόνο στην ανάστροφη εικόνα δίνουν την επιλογή μέσω προβολής πίσω στην αρχική εικόνα (κυανό).

Το Σχήμα 4.3 απεικονίζει το ταίριασμα μέσω της διαδικασίας SSM ανάμεσα στην εικόνα  $X$  και τον εαυτό της, αλλά και το αντικατοπτρισμένο είδωλο της  $Y$ , εντοπίζοντας μία ομάδα περιοχών η οποία αντιστοιχεί σε επαναλαμβανόμενη δομή και μία συμμετρία αντίστοιχα, σε μία πραγματική εικόνα. Υπάρχουν περισσότερες εντοπισμένες ομάδες οι οποίες δεν απεικονίζονται στο παράδειγμα. Για την ίδια εικόνα, το Σχήμα 4.4 απεικονίζει όλα τα αρχικά χαρακτηριστικά, καθώς και τις τρεις ομάδες επιλογών που έγιναν μέσω της διαδικασίας SSM, την ευθεία ( $\alpha_d$ ), ανάστροφη ( $\alpha_f$ ) και μέσω προβολής στην αρχική εικόνα ( $\alpha'_f$ ) επιλογή.

#### 4.3.4 Χαλαρό γεωμετρικό αυτο-ταίριασμα

Δεδομένου ενός συνόλου αντιστοιχιών  $C$ , το γεωμετρικό αυτο-ταίριασμα το οποίο περιγράφηκε στην ενότητα 4.3.2 βασίζεται στον αλγόριθμο FSM [66] κι έχει τετραγωνική πολυπλοκότητα ως προς τον αριθμό των αντιστοιχιών,  $|C|$ , στην χειρότερη περίπτωση. Από την άλλη, η μέθοδος *Hough pyramid matching* (HPM) [78] είναι μία μέθοδος χαλαρού γεωμετρικού ταιριάσματος της οποίας η πολυπλοκότητα είναι γραμμική ως προς το  $|C|$  κι έχει δειχθεί ότι ξεπερνά σε απόδοση την μέθοδο FSM όταν εφαρμοσθεί στην ανακατάταξη εικόνων. Η μέθοδος HPM δεν είναι απλώς πιο γρήγορη επειδή δεν απαιτεί καταμέτρηση inliers, αλλά επίσης είναι απελευθερωμένη από κάποιο κατώφλι το οποίο να ορίζει τι είναι ένας inlier, όπως το  $\epsilon$  στην σχέση (4.8) ή (4.9). Επιπλέον, υπολογίζει μία ισχύ για κάθε αντιστοιχία, η οποία αντικατοπτρίζει το πόσο συμβατή γεωμετρικά είναι με όλες τις υπόλοιπες. Αυτό πραγματοποιείται χωρίς ποτέ να απαριθμηθούν όλα τα ζευγάρια αντιστοιχιών.

Θεωρούμε την μέθοδο HPM ως την βάση για ένα πιο γρήγορο κι εναλλακτικό τρόπο αυτο-ταιριάσματος με σκοπό την επιλογή χαρακτηριστικών. Αντί για ένα συνολικό βαθμό ομοιότητας ανάμεσα στην εικόνα και τον εαυτό της (το οποίο δεν θα είχε νόημα), χρησιμοποιούμε την ισχύ που έχει υπολογισθεί για κάθε αντιστοιχία. Κατά αυτή την έννοια, η μέθοδος HPM είναι πιο εύκολο να χρησιμοποιηθεί για αυτο-ταίριασμα από ότι η μέθοδος FSM. Παρόλα αυτά υπάρχουν μερικές παραλλαγές ως προς την παραδοσιακή μέθοδο [78].

Δεδομένου ενός συνόλου αρχικών αντιστοιχιών  $C$ , είτε ένα υποσύνολο του  $X^2$  (αυτο-ταίριασμα) ή  $X \times Y$  (ανάστροφο ταίριασμα), κάθε αντιστοιχία  $c = (x, y) \in C$  παράγει έναν μετασχηματισμό τεσσάρων βαθμών ελευθερίας, ο οποίος τώρα αναπαρίσταται από το διάνυσμα  $g(c) \in \mathbb{R}^4$  που δίνεται από την σχέση (4.2). Βλέποντας αυτά τα διανύσματα ως σημεία σε έναν 4-διάστατο χώρο μετασχηματισμών, το πρόβλημα μας τότε γίνεται το να εντοπισθούν περιοχές μεγάλης πυκνότητας σε αυτόν τον χώρο.

Δημιουργείται μία ιεραρχική διαμέριση του χώρου των μετασχηματισμών σε  $L$  επίπεδα ως αλληλουχία των διαμερίσεων  $B_0, \dots, B_{L-1}$ . Το  $B_0$  είναι το κατώτερο επίπεδο με την πιο λεπτομερή διαμέριση, ενώ το  $B_{L-1}$  είναι το ανώτερο με την χονδροειδή διαμέριση το οποίο έχει ένα και μοναδικό κυτίο. Κάθε κυτίο του  $B_\ell$  χωρίζεται σε  $2^4$  κυτία στο  $B_{\ell+1}$ . Έπειτα κατασκευάζεται ένα ιστόγραμμα πυραμίδας κατανέμοντας τις αντιστοιχίες στα κυτία όλων των επιπέδων της πυραμίδας. Δεδομένου ενός κυτίου  $b$ , έστω  $F(b) = \{c \in C : g(c) \in b\}$  το σύνολο των αντιστοιχιών των οποίων το διάνυσμα παραμέτρων πέφτει στο κυτίο  $b$ , και  $f(b) = |F(b)|$  το πλήθος τους, το οποίο αναπαριστά την συχνότητα ιστογράμματος του κυτίου.

Οι αντιστοιχίες οι οποίες πέφτουν μέσα στο ίδιο κυτίο ομαδοποιούνται, και περιμένουμε να βρούμε τις ομάδες οι οποίες είναι γεωμετρικά πιο συμβατές στα κατώτερα επίπεδα της ιεραρχίας, με τις μεγάλες ομάδες να θεωρούνται πιο συμβατές. Μία μόνη της

---

**Algorithm 5** Hough pyramid self-matching (HPSM)

---

**procedure**  $\beta \leftarrow \text{HPSM}(C, L)$

```

input : correspondences  $C$ , levels  $L$ 
output: strengths  $\beta$ 

25 begin
26    $B \leftarrow \text{PARTITION}(L)$                                  $\triangleright$  partition space in  $L$  levels
27   for  $c \in C$  do  $\beta(c) \leftarrow 0$                           $\triangleright$  initialize strengths
28
29   HPSM-REC( $\beta, C, L - 1, B$ )                                 $\triangleright$  recurse at top
30   return  $\beta / \max(\beta)$                                       $\triangleright$  normalize

31 procedure HPSM-REC( $\beta, C, \ell, B$ )
32   in/out : strengths  $\beta$ 
33   input : correspondences  $C$ , level  $\ell$ , partition map  $B$ 
34   begin
35     if  $\ell < 0$  then return
36     for  $b \in B_\ell$  do  $F(b) \leftarrow \emptyset$                        $\triangleright$  initialize histogram
37     for  $c \in C$  do                                               $\triangleright$  populate histogram
38        $F(q_\ell(c)) \leftarrow F(q_\ell(c)) \cup c$                           $\triangleright \dots$  by quantizing
39     for  $b \in B_\ell$  do
40        $F \leftarrow F(b)$                                                   $\triangleright$  correspondences in  $b$ 
41       if  $|F| < 2$  then continue                                          $\triangleright$  exclude singles
42     HPSM-REC( $\beta, F, \ell - 1, B$ )                                       $\triangleright$  recurse down
43     if  $\ell = L - 1$  then  $m \leftarrow 2$ 
44     else  $m \leftarrow 1$ 
45     for  $c \in F$  do                                               $\triangleright$  update strengths in  $b$ 
46        $\beta(c) \leftarrow \beta(c) + 2^{-\ell}m|F|$                             $\triangleright \dots$  as in (4.22)

```

---

αντιστοιχία δεν δίνει κάποια στοιχεία ομοιότητας, και δεν μπορεί να σχηματίσει ομάδα, έτσι το μέγεθος μίας ομάδας ορίζεται ως  $s(b) = [f(b) - 1]_+ = \max(0, f(b) - 1)$ . Έστω  $b_0 \subseteq \dots \subseteq b_\ell$  η αλληλουχία των κυτίων τα οποία περιέχουν την αντιστοιχία  $c$  μέχρι το επίπεδο  $\ell$ . Η ισχύς μέχρι το επίπεδο  $\ell$  για κάθε αντιστοιχία  $c \in C$ , η οποία αντικατοπτρίζει



**Σχήμα 4.5:** Ανάστροφο ταίριασμα με την μέθοδο HPSM και  $L = 5$  επίπεδα. (Πάνω) αντιστοιχίες σε ένα κυτίο του επιπέδου 0, οι οποίες αποκαλύπτουν μία ομάδα συμμετρικών περιοχών. (Κάτω) όλες οι αντιστοιχίες, με τις κόκκινες (κίτρινες) να είναι οι πιο ισχυρές (αδύναμες).

τις παραπάνω παρατηρήσεις, δίνεται από

$$\beta_\ell(c) = s(b_0) + \sum_{i=1}^{\ell} 2^{-i} \{s(b_i) - s(b_{i-1})\}. \quad (4.22)$$

Η συνολική ισχύς της αντιστοιχίας  $c$  μέχρι το ανώτερο επίπεδο  $L - 1$  είναι απλά  $\beta_{L-1}(c)$ , την οποία μετατρέπουμε σε σχετική ισχύ, κανονικοποιώντας με την μέγιστη ισχύ από όλες τις αντιστοιχίες του  $C$ :

$$\beta(c) = \beta_{L-1}(c) / \max_{a \in C} \beta_{L-1}(a). \quad (4.23)$$

Το προηγούμενο σχήμα κανονικοποίησης είναι μία παραλλαγή του [78], δίνοντας την ευκαιρία στις πιο ισχυρές αντιστοιχίες να συμμετέχουν στην επιλογή ακόμα κι αν η ισχύς τους δεν είναι μεγάλη ως απόλυτη τιμή. Μία άλλη παραλλαγή είναι πως δεν εφαρμόζουμε ένα προς ένα αντιστοιχία, εφόσον αυτό θα ήταν αντίθετο με τον εντοπισμό επαναλαμβανόμενων δομών στην ίδια εικόνα. Η διαδικασία ταιριάσματος απλοποιείται αρκετά σε σχέση με εκείνη του [78], όπως συνοψίζεται στον Αλγόριθμο 5. Θα αναφερόμαστε σε αυτό ως *αυτο-ταίριασμα πυραμίδας Hough (HPSM)*. Η προβολή του  $q_\ell : C \rightarrow B_\ell$  η οποία εμφανίζεται στην γραμμή 37 χρησιμοποιείται για να κβαντίσει μία αντιστοιχία σε ένα κυτίο

## Πειράματα



**Σχήμα 4.6:** Εικόνες δείγματα, μαζί με τα επιλεγμένα χαρακτηριστικά με χρήση της μεθόδου HPSM και  $L = 5$  επίπεδα, χρωματισμένα όπως στο Σχήμα 4.4.

του επιπέδου  $\ell$ . Η αναπαράσταση ιστογράμματος είναι αραιή και ο αλγόριθμος παραμένει με γραμμική πολυπλοκότητα ως προς το  $|C|$ .

Τώρα, αντικαθιστώντας το  $\alpha$  με  $\beta$ , μπορούμε να χρησιμοποιήσουμε την μέθοδο HPSM για να βρούμε τις επιβεβαιωμένες αντιστοιχίες  $\beta(C) \subseteq C$ , τροποποιώντας την σχέση (4.10) ως

$$\beta(C) = \{c \in C : \beta(c) \geq \tau_\beta\}, \quad (4.24)$$

με κατώφλι  $\tau_\beta \in [0, 1]$ . Ομοίως, έστω πως  $\beta(C_f)$  είναι οι επιβεβαιωμένες αντιστοιχίες μετά από ανάστροφο ταίριασμα με την μέθοδο HPSM. Έπειτα τα επιλεγμένα χαρακτηριστικά με ευθεία ( $\beta_d$ ), ανάστροφη ( $\beta_f$ ), με προβολή πίσω στην αρχική εικόνα ( $\beta'_f$ ) και πλήρη επιλογή ( $\beta$ ) ορίζονται ομοίως με τις Σχέσεις 4.3.2, 4.3.3, απλά αντικαθιστώντας το  $\alpha$  με το  $\beta$  στις Σχέσεις (4.12), (4.18), (4.20), (4.21) αντίστοιχα. Θα αναφερόμαστε στο  $\beta(X)$  ως επιλεγμένο σύνολο χαρακτηριστικών για την εικόνα  $X$  σε αυτή την περίπτωση.

Το Σχήμα 4.5 απεικονίζει το ανάστροφο ταίριασμα με χρήση της διαδικασίας ταιριάσματος του HPSM, για την ίδια εικόνα με εκείνη των Σχημάτων 4.3, 4.4. Σε αυτή την περίπτωση δεν υπάρχουν ομάδες από αντιστοιχίες οι οποίες είναι inliers σε έναν και μοναδικό μετασχηματισμό. Κάποιος μπορεί να θεωρήσει σαν ομάδα τις αντιστοιχίες που πέφτουν στο ίδιο κυτίο. Όσο πιο χαμηλά βρίσκεται το επίπεδο ταίριασμάτος, τόσο πιο καλό είναι το ταίριασμα με τον μετασχηματισμό. Κάθε αντιστοιχία λαμβάνει την ισχύ της ανεξάρτητα, ανάλογα με το κατά πόσο είναι γεωμετρικά συμβατή με τις υπόλοιπες αντιστοιχίες. Οι ισχυρότερες είναι εκείνες που θα επιλεχθούν.

Στο Σχήμα 4.6 παρουσιάζονται επιπλέον παραδείγματα για μερικές εικόνες, απεικονίζοντας το πλήρες σύνολο επιλεγμένων περιοχών με χρήση τις διαδικασίας HPSM, περιλαμβάνοντας και τις τρεις κατηγορίες επιλογής.

## 4.4 Πειράματα

Σε αυτή την ενότητα διερευνούμε την συμπεριφορά των διαδικασιών SSM και HPSM με χρήση διαφορετικών παραμέτρων και συγκρίνουμε την απόδοση τους και τις απαιτήσεις



Σχήμα 4.7: Δείγματα εικόνων από το σύνολο SymCity.

μνήμης με την περίπτωση που γίνεται χρήση του πλήρους συνόλου χαρακτηριστικών. Επίσης συγκρίνουμε τις προτεινόμενες μεθόδους με άλλες προσεγγίσεις για επιλογή χαρακτηριστικών από μοναδικές εικόνες, οι οποίες βασίζονται σε πληροφορία που είναι διαθέσιμη από τους ανιχνευτές χαρακτηριστικών, όπως η κλίμακα τους [82] και η ισχύς τους.

#### 4.4.1 Σύνολα εικόνων

Χρησιμοποιούμε δύο σύνολα εικόνων, το νέο *SymCity* και το *World Cities*<sup>1</sup> [78]. Το *SymCity* είναι ένα σύνολο το οποίο αποτελείται από 953 προσημειωμένες εικόνες από το *Flickr*, χωρισμένες σε 299 μικρές ομάδες. Από κάθε ομάδα, μία μοναδική εικόνα εισάγεται στην βάση, ενώ οι υπόλοιπες 654 χρησιμοποιούνται ως ερωτήματα. Για κάθε ερώτημα επιθυμούμε να ανακτήσουμε την μία και μοναδική εικόνα από την βάση και η απόδοση εξαρτάται από την θέση αυτής και μόνο. Το σύνολο των εικόνων περίσπασης που χρησιμοποιούμε είναι το *World Cities* το οποίο αποτελείται από 2M εικόνες, ενώ εμείς θα χρησιμοποιήσουμε μόνο το πρώτο 1M εικόνων του. Δείγματα εικόνων από το σύνολο *SymCity* παρουσιάζονται στο Σχήμα 4.7.

#### 4.4.2 Πρωτόκολλο

Εξάγουμε τοπικά χαρακτηριστικά και περιγραφείς SURF [8] για κάθε εικόνα και τα αντιστοιχούμε σε οπτικές λέξεις γενικού οπτικού λεξικού μεγέθους 100K με την χρήση του λογισμικού FLANN [58] και κρατώντας μόνο τον έναν πλησιέστερο γείτονα με χρήση της Ευκλείδειας απόστασης. Το λεξικό δημιουργείται με χρήση του προσεγγιστικού αλγορίθμου *k-means* (AKM) [66] σε ένα ανεξάρτητο σύνολο 15K εικόνων.

Κατά την διαδικασία επιλογής χαρακτηριστικών, οι αρχικές αντιστοιχίες δημιουργού-

<sup>1</sup><http://image.ntua.gr/iva/datasets/wc/>

νται με εύρεση των πλησιέστερων γειτόνων στον χώρο των τοπικών περιγραφέων. Αυτό γίνεται και πάλι με την χρήση του λογισμικού FLANN [58]. Κατά την αποτυχία του εντοπισμού συμμετριών, επιλέγουμε κάποια χαρακτηριστικά να κρατήσουμε με βάση πληροφορία από τον ανιχνευτή χαρακτηριστικών. Για την ακρίβεια, η μέθοδος SSM αποτυγχάνει όταν δεν υπάρχουν καθόλου επιβεβαιωμένες αντιστοιχίες, ενώ η μέθοδος HPSM όταν επιλέγει λιγότερα από  $\tau_\alpha$  χαρακτηριστικά, τα οποία είναι ίσα με τον ελάχιστο αριθμό από inliers που απαιτούνται για να επιβεβαιωθεί μία υπόθεση. Σε κάθε περίπτωση, επιλέγουμε ένα ποσοστό  $\lambda = 15\%$  των χαρακτηριστικών με την μεγαλύτερη ισχύ που παρέχεται από τον ανιχνευτή χαρακτηριστικών.

Για την ανάκτηση, δημιουργούμε ανάστροφο αρχείο και για κάθε ερώτημα κατατάσσουμε τις εικόνες με βάση το  $\ell^2$ -κανονικοποιημένο εσωτερικό γινόμενο ανάμεσα στα ιστογράμματα κατά BoW και το *tf-idf* βάρος [73].

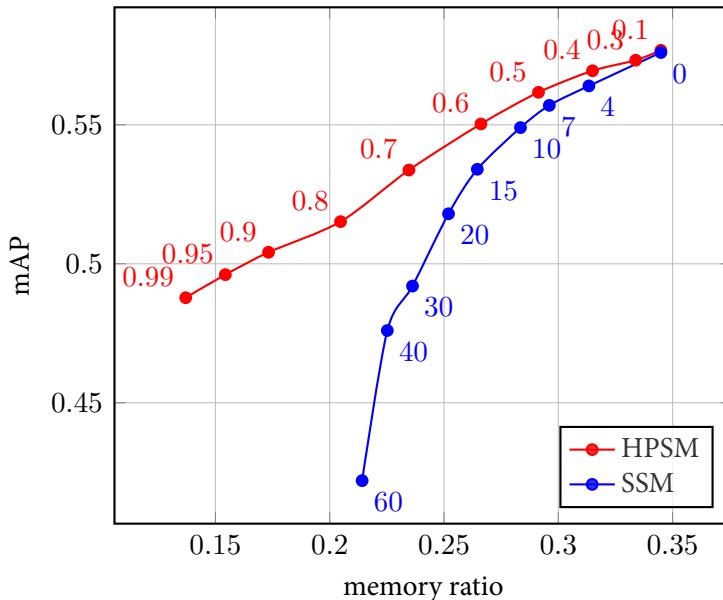
Μετράμε την απόδοση μέσω του μέτρου *mean Average Precision* (mAP) πάνω σε όλες τις εικόνες ερωτήματος. Μετράμε το μέγεθος της δομής δεικτοδότησης από τον αριθμό των εισαγωγών ανά εικόνα, ο οποίος είναι ο αριθμός των οπτικών λέξεων που εμφανίζονται στην εικόνα τουλάχιστον μία φορά. Για κάθε μέθοδο ορίζουμε τον λόγο μνήμης ως τον λόγο του μεγέθους της δομής δεικτοδότησης της μεθόδου προς εκείνο του πλήρους συνόλου χαρακτηριστικών.

Ο χρόνος εκτέλεσης αναφέρεται σε ολόκληρη την διαδικασία της επιλογής χαρακτηριστικών περιέχοντας και το ευθύ και το ανάστροφο ταίριασμα, αλλά εξαιρώντας τον χρόνο που απαιτείται για την δημιουργία των αρχικών αντιστοιχιών οι οποίες είναι ίδιες για κάθε μέθοδο. Σε όλα τα πειράματα μας, εφαρμόζουμε την επιλογή χαρακτηριστικών μόνο στις 299 εικόνες του SymCity και όχι στις εικόνες περίσπασης. Αυτό γίνεται για να επιτρέψουμε μια πιο δίκαιη σύγκριση όταν για κάθε περίπτωση χρησιμοποιείται ίδια ποσότητα χαρακτηριστικών η οποία προέρχεται από εικόνες περίσπασης.

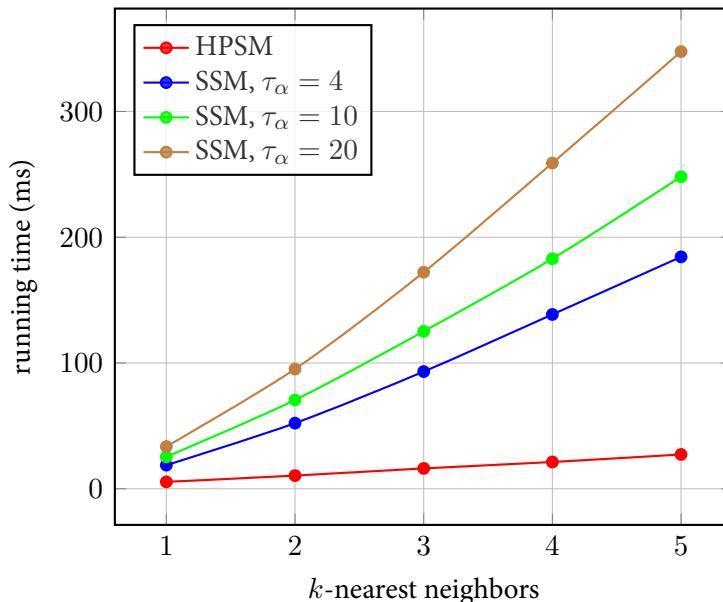
#### 4.4.3 Βελτιστοποίηση παραμέτρων

Αρχικά θέτουμε το κατώφλι ομοιότητας (4.4)  $\delta = 0.1$  και το κατώφλι για τους inliers του SSM (4.8)  $\epsilon = 7$  pixels. Το πρώτο επιλέγεται συγκεκριμένα για περιγραφείς SURF και το δεύτερο δεν είναι απαραίτητο για το HPSM. Επίσης θέτουμε  $L = 5$  επίπεδα για το HPSM, όπως στο [78].

Με αυτές τις παραμέτρους, διεξάγουμε επιπλέον συντονισμό παραμέτρων για το σύννολο εικόνων SymCity, χρησιμοποιώντας 100K εικόνες περίσπασης από το *World Cities*. Στο Σχήμα 4.8 απεικονίζουμε το mAP και τον λόγο μνήμης για το SSM και το HPSM με μεταβλητά κατώφλια επιλογής  $\tau_\alpha$  (4.11) και  $\tau_\beta$  (4.24) αντίστοιχα. Παρατηρούμε πως και οι δύο μέθοδοι αποδίδουν το ίδιο καλά για υψηλούς λόγους μνήμης, το οποίο αντιστοιχεί στην περίπτωση που επιλέγονται όλα τα χαρακτηριστικά τα οποία συμμετέχουν σε κά-



Σχήμα 4.8: Απόδοση καταμετρημένη μέσω  $mAP$  για τις διαδικασίες SSM και HPSM για μεταβλητά κατώφλια επιλογής  $\tau_\alpha$  και  $\tau_\beta$  αντίστοιχα, έναντι του λόγου μνήμης και με την παρουσία 100K εικόνων περίσπασης, με  $k = 3$  πλησιέστερους γείτονες. Το κατώφλι επιλογής απεικονίζεται δίπλα στα σημεία των γραφήματος.



Σχήμα 4.9: Μέσος χρόνος εκτέλεσης ( $ms$ ) για τις διαδικασίες HPSM και SSM, έναντι του αριθμού των  $k$ -πλησιέστερων γειτόνων για τις αρχικές αντιστοιχίες, όπου  $\tau_\alpha$  μεταβάλλεται για το SSM.

ποια αντιστοιχία. Παρόλα αυτά, όταν γινόμαστε πιο επιλεκτικοί, το SSM καταρρέει ενώ το HPSM φθίνει πιο ομαλά.

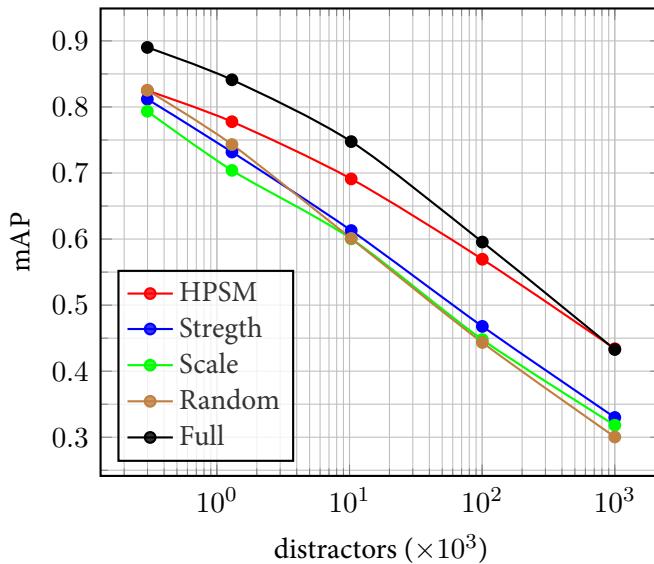
Στο Σχήμα 4.9 απεικονίζουμε τον καταμετρημένο μέσο χρόνο και για τις δύο μεθό-

$k$	1	2	3	4	5
$\tau_\beta = 0.4$	0.545	0.566	<b>0.569</b>	0.566	0.568
$\tau_\beta = 0.6$	0.522	0.538	0.550	0.551	0.547
$\tau_\beta = 0.8$	0.484	0.511	0.515	0.524	0.529

Πίνακας 4.1: Απόδοση για το HPSM μετρημένη μέσω mAP έναντι του αριθμού των  $k$ -πλησιέστερων γειτόνων για μεταβλητό κατώφλι  $\tau_\beta$  και με την παρουσία 100K εικόνων περίσπασης.

δους, για μεταβλητό αριθμό  $k$  πλησιέστερων γειτόνων οι οποίοι χρησιμοποιούνται για την δημιουργία των αρχικών αντιστοιχιών στην σχέση (4.4). Ο χρόνος εκτέλεσης για την μέθοδο SSM εξαρτάται επίσης και από την τιμή του κατωφλίου  $\tau_\alpha$ . Χαμηλότερο κατώφλι θα επιτρέψει σε περισσότερες υποθέσεις να παραληφθούν. Η μέθοδος HPSM είναι κατά μέσο όρο 5 με 15 φορές πιο γρήγορη από την SSM. Έτσι δεδομένης και της υψηλότερης απόδοσης έχει ξεκάθαρο πλεονέκτημα. Στην συνέχεια των πειραμάτων επιλέγουμε να χρησιμοποιήσουμε μόνο αυτή.

Στον Πίνακα 4.1 παρουσιάζεται η σύγκριση της απόδοσης της μεθόδου HPSM για μεταβλητό αριθμό  $k$  πλησιέστερων γειτόνων. Η απόδοση σταθεροποιείται και σχεδόν φθίνει για περισσότερους από 3 γείτονες. Επιλέγουμε να χρησιμοποιήσουμε  $k = 3$  για την συνέχεια των πειραμάτων μας. Επίσης επιλέγουμε  $\tau_\beta = 0.4$ , παρά το γεγονός ότι υπάρχει επιπλέον πείραμα με μεταβλητό  $\tau_\beta$ . Ο μέσος χρόνος εκτέλεσης του HPSM είναι 16.2ms.

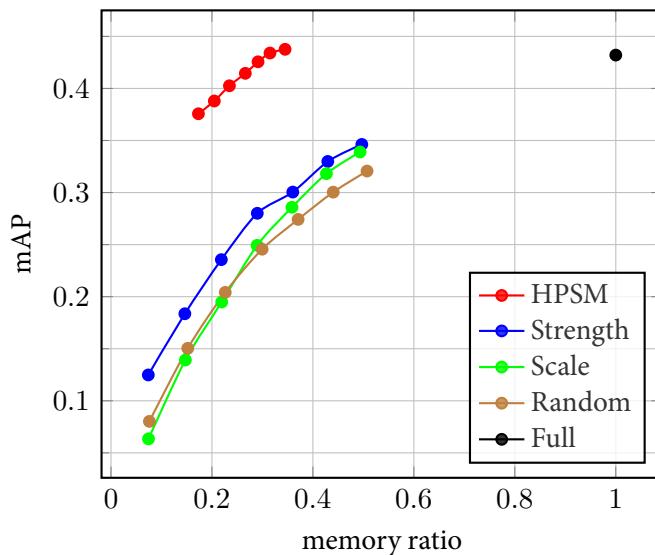


Σχήμα 4.10: Σύγκριση mAP έναντι του αριθμού των εικόνων περίσπασης, με  $\tau_\beta = 0.4$  για το HPSM κι ένα σταθερό  $n = 300$  για τα υπόλοιπα κριτήρια επιλογής.

#### 4.4.4 Συγκρίσεις

Πηγαίνοντας σε μεγαλύτερη κλίμακα, μέχρι 1M εικόνες περίσπασης από το *World Cities*, συγκρίνουμε την δική μας μέθοδο επιλογής με την χρήση του πλήρους συνόλου χαρακτηριστικών. Επίσης συγκρίνουμε με τρεις εναλλακτικούς τρόπους για επιλογή, οι οποίοι βασίζονται σε απλά κριτήρια. Κρατάμε έναν σταθερό αριθμό από  $n$  χαρακτηριστικά, εκείνα τα οποία έχουν την μεγαλύτερη ισχύ, ή την μεγαλύτερη κλίμακα σ' όπως στο [82], είτε με τυχαία επιλογή από το πλήρες σύνολο περιοχών.

Στο Σχήμα 4.10 παρουσιάζουμε την απόδοση mAP με μεταβλητό αριθμό εικόνων περίσπασης. Ενώ η επιλογή μέσω του HPSM δεν υπερέχει σε μικρή κλίμακα, το οποίο δεν μας ξαφνιάζει, όμως τα καταφέρνει να φτάσει την απόδοση του πλήρους συνόλου σε μεγάλη κλίμακα. Αυτό το πείραμα εκτελέσθηκε με κατώφλι επιλογής  $\tau_\beta = 0.4$ , το οποίο καταλήγει σε λόγο μνήμης 0.31 στο προσημειωμένο σύνολο εικόνων, το οποίο αντιστοιχεί σε επιλογή 284 περιοχών ανά εικόνα. Για να έχουμε μία δίκαιη σύγκριση ως προς τις άλλες μεθόδους, θέτουμε  $n = 300$  ώστε να γίνεται χρήση περίπου του ίδιου αριθμού περιοχών σε κάθε περίπτωση. Τελικά το HPSM υπερέχει των άλλων μεθόδων επιλογής.



**Σχήμα 4.11:** Σύγκριση mAP έναντι του λόγου μνήμης για μεταβλητό κατώφλι  $\tau_\beta$ ,  $n$  με την παρουσία 1M εικόνων περίσπασης.

Η χρήση ενός κατωφλίου  $\tau_\beta$  θα έκανε την διαδικασία πολύ πιο επιλεκτική και θα μείωνε τον λόγο μνήμης, με κόστος χαμηλότερο mAP. Αυτός είναι ένας τρόπος να ρυθμιστεί η ισορροπία ανάμεσα στις ανάγκες μνήμης και την απόδοση. Το Σχήμα 4.11 απεικονίζει αυτή την σχέση ανάμεσα σε μνήμη κι απόδοση με την χρήση 1M εικόνων περίσπασης, δείχνοντας πως το HPSM μπορεί να ξεπεράσει την χρήση του πλήρους συνόλου με λόγο μνήμης περίπου 35%. Επίσης η απόδοση του είναι πολύ καλύτερη από τις τρεις άλλες μεθόδους επιλογής όταν χρησιμοποιείται η ίδια ακριβώς μνήμη.



# Κεφάλαιο 5

## Επέκταση ερωτήματος με δυαδικούς περιγραφείς

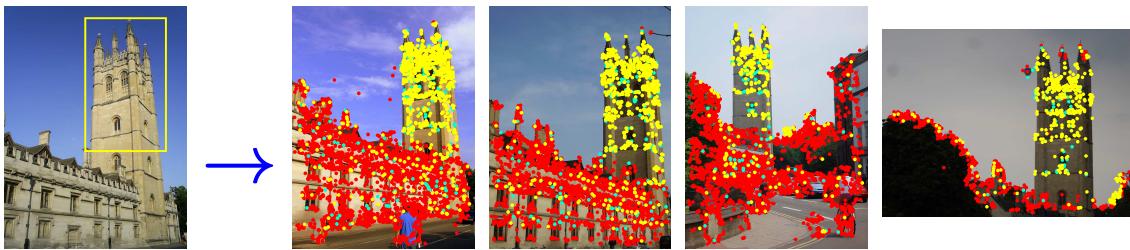
### 5.1 Εισαγωγή και σχετική βιβλιογραφία

Μία από τις πιο διαδεδομένες τεχνικές στην αναζήτηση πληροφορίας είναι η επέκταση ερωτήματος (query expansion - QE) [55], η οποία είναι ένα είδος αυτόματης ανάδρασης. Η γενική ιδέα είναι να εκμεταλλευτεί κανείς τα πιο αξιόπιστα από τα αποτελέσματα που κατατάχθηκαν στις πρώτες θέσεις μετά από την αρχική αναζήτηση, και να τα χρησιμοποιήσει για να παράγει μια εμπλουτισμένη αναπαράσταση. Εάν τα αξιόπιστα αποτελέσματα είναι αρκετά τότε μπορεί ο νέος γύρος αναζήτησης να ανακτήσει πολύ περισσότερα σωστά αποτελέσματα.

Η επέκταση ερωτήματος εισήχθηκε στην όραση υπολογιστών από τους Chum et al. [19], οι οποίοι πρότειναν μία τεχνική η οποία υλοποιεί την ιδέα της επέκτασης ερωτήματος προσαρμοσμένη στην οπτική αναζήτηση. Αρχικά εφαρμόζεται γεωμετρικό ταίριασμα στην λίστα των αποτελεσμάτων από το αρχικό ερώτημα. Οι γεωμετρικά επιβεβαιωμένες εικόνες παρέχουν χαρακτηριστικά για τον εμπλουτισμό της αναπαράστασης. Η επικρατέστερη από τις μεθόδους που προτείνονται είναι εκείνη η οποία δημιουργεί τα ιστογράμματα BoW με βάση τα χαρακτηριστικά τα οποία προβάλλονται πίσω στην εικόνα ερωτήματος με επιτυχία. Τελικά η νέα αναπαράσταση είναι το μέσο ιστόγραμμα εξ αυτών.

Αρκετές παραλλαγές έχουν προταθεί [40, 16, 2], όπως η χρήση ενός ταξινομητή για την δημιουργία του εμπλουτισμένου νέου ερωτήματος [2]. Αν και αυτές οι παραλλαγές έχουν βελτιώσει την απόδοση, υποφέρουν από δύο μειονεκτήματα τα οποία επηρεάζουν την πολυπλοκότητα και την ποιότητα της αναζήτησης:

- Προϋποθέτουν την χρήση κοστοβόρου γεωμετρικού ταιριάσματος ώστε να επιλεχθούν εκείνες οι εικόνες οι οποίες απεικονίζουν το ίδιο αντικείμενο με την εικόνα ερωτήματος.



**Σχήμα 5.1:** Εικόνα ερωτήματος (αριστερά) και τα επιλεγμένα χαρακτηριστικά (κίτρινα+κυανά) για να εμπλουτίσουν το νέο ερώτημα από τις εικόνες που ανακτήθηκαν. Τα χαρακτηριστικά με κόκκινο χρώμα απορρίπτονται. Τα χαρακτηριστικά με κυανό χρώμα είναι εκείνα που αντιστοιχούν σε οπτικές λέξεις οι οποίες υπάρχουν και στην εικόνα ερωτήματος, ενώ τα κίτρινα σε εκείνες που δεν υπάρχουν. Το συγκεκριμένο παράδειγμα δεν περιλαμβάνει καθόλου γεωμετρική πληροφορία.

- Η νέα εμπλουτισμένη αναπαράσταση περιέχει πολύ περισσότερα μη μηδενικά στοιχεία, άρα θα επισκεφθεί πολύ περισσότερες λίστες του ανάστροφου αρχείου με αποτέλεσμα να γίνεται πιο αργή η αναζήτηση.

Έχουν προταθεί κι άλλες μέθοδοι για συλλογές εικόνων σταθερού μεγέθους [2, 82]. Βασίζονται σε μία εκ των προτέρων διαδικασία γεωμετρικού ταιριάσματος όλων των εικόνων της βάσης με όλες. Αυτό το στάδιο είναι πολύ κοστοβόρο και απαγορευτικό για πολύ μεγάλες βάσεις. Μία σχετική μέθοδος [22] δημιουργεί έναν γράφο που συνδέει όλες τις όμοιες εικόνες και χρησιμοποιεί τους αμοιβαίους πλησιέστερους γείτονες για να ορίσει μία νέα μετρική ομοιότητας με την οποία ανακατατάσσονται οι ανακτημένες εικόνες. Το κόστος δημιουργίας ενός τέτοιου γράφου είναι και πάλι πολύ μεγάλο.

Υπάρχουν επίσης αρκετές τεχνικές που στοχεύουν να αντιμετωπίσουν τις απώλειες κβαντισμού που συνδέονται με την μέθοδο BoW. Τέτοιες είναι η πολλαπλή ανάθεση σε οπτικές λέξεις [39] ή η χαλαρή ανάθεση [68]. Με συμπληρωματικό τρόπο η τεχνική Hamming Embedding (HE) [37] βελτιώνει δραστικά την ποιότητα της αναζήτησης με το να χρησιμοποιεί για κάθε περιγραφέα, όχι μόνο την οπτική λέξη αλλά κι έναν δυαδικό περιγραφέα (δυαδική υπογραφή) για κάθε τοπικό χαρακτηριστικό. Η μέθοδος HE δεν είναι συμβατή με τις υπάρχουσες τεχνικές για επέκταση ερωτήματος καθώς αυτές υποθέτουν ότι υπάρχει ένα ιστόγραμμα που περιγράφει όλη την εικόνα, το οποίο είναι το ιστόγραμμα εμφάνισης των οπτικών λέξεων. Μία εξαίρεση είναι η μεταβατική επέκταση ερωτήματος [19] η οποία δεν χρησιμοποιεί την εκάστοτε αναπαράσταση εικόνας, όμως η απόδοση της και ο χρόνος εκτέλεσης δεν είναι ικανοποιητικά.

Σε αυτό το κεφάλαιο θα παρουσιάσουμε μια μέθοδο για επέκταση ερωτήματος η οποία χρησιμοποιεί την ίδια αναπαράσταση και το ίδιο βασικό σύστημα αναζήτησης με την μέθοδο HE. Η τοπική αναπαράσταση τελικά εμπλουτίζεται και ραφινάρεται, με τέτοιο τρόπο ώστε να μην είναι απαραίτητη η γεωμετρική επιβεβαίωση των εικόνων. Στο Σχήμα 5.1 απεικονίζεται ένα παράδειγμα εικόνων και τοπικών χαρακτηριστικών τα οποία έχουν επιλεχθεί ως αξιόπιστα από την μέθοδο μας για να εμπλουτιστεί η αρχική αναπαράσταση.

## 5.2 Το βασικό σύστημα αναζήτησης

Σε αυτή την ενότητα περιγράφουμε το σύστημα αναζήτησης το οποίο βασίζεται στην μέθοδο Hamming Embedding (HE), πάνω στην οποία χτίζουμε και την δική μας μέθοδος για επέκταση ερωτήματος. Η βασική μας μέθοδος αναζήτησης βασίζεται στην επιτυχημένη μέθοδο HE [37], η οποία περιγράφεται στη συνέχεια. Μία εικόνα αναπαρίσταται από ένα σύνολο  $\mathcal{P}$  τοπικών περιγραφέων SIFT [52] οι οποίοι έχουν εξαχθεί από περιοχές Hessian-Affine [56].

**BoW και Hamming Embedding.** Οι τοπικοί περιγραφείς κβαντίζονται με τον παραδοσιακό αλγόριθμο  $k$ -means, όπου το  $k$  καθορίζει το μέγεθος του οπτικού λεξικού. Έτσι ένας περιγραφές  $p \in \mathcal{P}$  αναπαρίσταται από έναν ακέραιο δείκτη, ο οποίος όπως έχει ήδη αναφερθεί ονομάζεται  $v(p)$ . Ο υπολογισμός και η κανονικοποίηση του ιστογράμματος των οπτικών λέξεων παράγει την αναπαράσταση της μεθόδου BoW. Αυτό ισοδυναμεί με μία μέθοδο ταιριάσματος κατά την οποία όλοι οι περιγραφείς που ανήκουν στην ίδια οπτική λέξη ταιριάζουν. Συγκεκριμένα ταιριάζουν με βαθμό ομοιότητας ίσο με ένα βάρος που σχετίζεται με την αντίστροφη συχνότητα εμφάνισης της οπτικής λέξης [74, 37].

Για να προσδιορίσει κανείς καλύτερα την ποιότητα του ταιριάσματος και για να παρέχει πιο αξιόπιστα βάρη στο κάθε ταίριασμα, με την χρήση της τεχνικής HE [37] ο κάθε οπτικός περιγραφέας  $p$  εκτός από την οπτική λέξη τελικά αντιστοιχίζεται και με ένα δυαδικό διάνυσμα ή αλλιώς δυαδική υπογραφή  $b(p)$ , το οποίο παρέχει έναν καλύτερο προσδιορισμό της θέσης του κάθε περιγραφέα, διαιρώντας το κάθε κελί κβαντισμού  $v(p)$ . Στην μέθοδο HE η σύγκριση μεταξύ δύο περιγραφέων  $q$  και  $p$  που ανήκουν στην ίδια οπτική λέξη  $v(p) = v(q)$  γίνεται μέσω της απόστασης Hamming  $h(q, p) = \|b(q) - b(p)\|_1$  ανάμεσα στα δυαδικές τους υπογραφές. Εάν η απόσταση Hamming είναι πάνω από ένα κατώφλι  $h_t$ , τότε οι περιγραφείς, αντίστοιχα και οι περιοχές από τις οποίες προήλθαν οι περιγραφείς, θεωρείται ότι δεν ταιριάζουν και τους αντιστοιχίζεται ομοιότητα ίση με μηδέν. Το ίδιο φυσικά γίνεται και για περιγραφείς που ανήκουν σε διαφορετική οπτική λέξη. Μία σημαντική βελτίωση [37, 33] στην απόδοση μπορεί να επιτευχθεί εάν ο κάθε βαθμός ομοιότητας σταθμιστεί από μία συνάρτηση φθίνουσα ως προς την απόσταση Hamming. Εμείς θα νιοθετήσουμε την Γκαουσιανή συνάρτηση η οποία χρησιμοποιήθηκε και σε προηγούμενες δουλειές [33].

**Το φαινόμενο ομοβροντίας (burstiness)** στις εικόνες παρουσιάστηκε και αντιμετωπίσθηκε πρώτα από τους Jegou *et al.* [33]. Ασχολείται με τους περιγραφείς οι οποίοι διεγείρουν πολλαπλές αντιστοιχίες με συγκεκριμένες εικόνες, το οποίο συνήθως σχετίζεται με επαναλαμβανόμενες δομές ή τοπικά χαρακτηριστικά τα οποία είναι περιέργως πολύ συχνά σε πολλές εικόνες της βάσης. Από τους τρόπους κανονικοποίησης για αντιμετώπιση του φαινομένου που προτάθηκαν, εμείς νιοθετούμε εκείνον ο οποίος σταθμίζει κάθε ταίριασμα

$\checkmark$	$-\mu$	mAP/BoW	mAP/HE	IF
		$47.7 \pm 0.8$	$67.1 \pm 0.6$	$1.200 \pm 0.003$
x		$47.7 \pm 0.5$	$69.5 \pm 0.8$	$1.290 \pm 0.003$
x	x	$48.1 \pm 0.7$	$69.6 \pm 0.8$	$1.238 \pm 0.003$

Πίνακας 5.1: Αξιολόγηση με την μετρική mAP (απόδοση) και IF (βαθμός ισορροπίας) διαφόρων διαδικασιών επεξεργασίας των περιγραφέων SIFT: RootSIFT [2, 31] (συμβολίζεται με  $\checkmark$ ) και κεντράρισμα (συμβολίζεται με  $-\mu$ ). Εκτελέσαμε 10 μετρήσεις για το κάθε πείραμα στην βάση Oxford5k με διαφορετικά λεξικά ( $k=16k$ ) και αναφέρουμε μέση τιμή και τυπική απόκλιση.

με την τετραγωνική ρίζα του αριθμού των ταιριασμάτων που σχετίζονται με τον συγκεκριμένο περιγραφέα της εικόνας ερωτήματος και την κάθε εικόνα βάσης [33].

**Πολλαπλή ανάθεση (Multiple assignment - MA).** Οι μέθοδοι BoW και HE διαχειρίζονται τους τοπικούς περιγραφείς οι οποίοι έχουν αντιστοιχηθεί στην ίδια οπτική λέξη. Παρόλα αυτά μπορούν να υπάρξουν απώλειες όταν περιγραφείς οι οποίοι θα έπρεπε να ταιριάζουν αντιστοιχίζονται σε διαφορετικές οπτικές λέξεις. Αυτό το πρόβλημα έχει αντιμετωπισθεί με το να γίνεται αντιστοίχιση του κάθε περιγραφέα σε πολλαπλές οπτικές λέξεις [39, 68]. Εφαρμόζουμε την πολλαπλή ανάθεση (MA) μόνο για την εικόνα ερωτήματος καθώς αυτό δεν επηρεάζει τις ανάγκες μνήμης της δομής δεικτοδότησης [37]. Η απλή ανάθεση θα αναφέρεται ως SA (single assignment) για να διακρίνεται από την πολλαπλή (MA).

### 5.3 Root-SIFT και Shift-SIFT

Πρόσφατα αποδείχθηκε πειραματικά [2, 31] ότι το να εφαρμόσει κανείς την τετραγωνική ρίζα στοιχείο προς στοιχείο στα διανύσματα των τοπικών περιγραφέων SIFT βελτιώνει την απόδοση της αναζήτησης. Αυτό επιτυγχάνεται με  $L_1$ -κανονικοποίηση του SIFT περιγραφέα [2] πριν από την εφαρμογή της τετραγωνικής ρίζας ή, ισοδύναμα [31], με την εφαρμογή τετραγωνικής ρίζας και την  $L_2$  κανονικοποίηση του διανύσματος. Αυτή η διαδικασία ισοδυναμεί με τον υπολογισμό της απόστασης Hellinger αντί της Ευκλείδειας. Η επίδραση αυτής της διαδικασίας αξιολογείται πειραματικά στον Πίνακα 5.1 στην βάση εικόνων Oxford5k [67] για τις μεθόδους BoW και HE, χωρίς χρήση της κανονικοποίησης ομοβροντίας. Ακολουθώντας το συνηθισμένο πρωτόκολλο, μετράμε mean average precision (mAP). Για να αντιμετωπίσουμε την διακύμανση των αποτελεσμάτων λόγω της τυχαιότητας της αρχικοποίησης του αλγορίθμου  $k$ -means, εκτελούμε το κάθε πείραμα 10 φορές με διαφορετικά λεξικά και αναφέρουμε μέση τιμή και τυπική απόκλιση. Η βελτίωση η οποία παρέχεται από την χρήση της τετραγωνικής ρίζας είναι σημαντική όταν χρησιμοποιείται με την μέθοδο HE.

Παρόλα αυτά, παρατηρούμε σαν παρενέργεια ότι η διαδικασία Root-SIFT καταλήγει

σε λιγότερο καλά ισορροπημένες λίστες των οπτικών λέξεων. Αυτό το φαινόμενο παρατηρήθηκε πρώτη φορά από τους Nister et al. [60] και συνήθως μετριέται με τον βαθμό ισορροπίας (imbalance factor - IF) [37].

**Shift-SIFT.** Για να μειώσουμε αυτό το ανεπιθύμητο φαινόμενο, εισάγουμε μία ακόμα διαδικασία επεξεργασίας των περιγραφέων SIFT η οποία αναφέρεται ως κεντράρισμα (shift-SIFT). Είναι εμπνευσμένη από την μέθοδο που προτάθηκε για τα ιστογράμματα της μεθόδου BoW [36], η οποία στοχεύει στον να διαχειριστεί τα στοιχεία απουσίας των οπτικών λέξεων με το να κεντράρει τα ιστογράμματα και τελικά να τα κανονικοποιεί με  $L_2$  κανονικοποίηση.

Ο Πίνακας 5.1 δείχνει το ενδιαφέρον αυτής της επεξεργασίας των περιγραφέων SIFT. Το όφελος στην απόδοση είναι σημαντικό μόνο για την μέθοδο BoW. Όμως, το κεντράρισμα παρέχει πιο ισορροπημένες λίστες του ανάστροφου αρχείου και μειώνει το κόστος αναζήτησης κατά 4%, όπως αυτό αντικατοπτρίζεται από το μέτρο IF. Η ερμηνεία μας είναι πως με αυτή την επεξεργασία η κατανομή των διανυσμάτων στον χώρο των χαρακτηριστικών είναι πιο ομοιόμορφη. Ο χώρος τώρα είναι η πλήρης  $L_2$  σφαίρα και όχι μόνο το θετικό της τεταρτοσφαίριο. Χρησιμοποιούμε την μέθοδο κεντραρίσματος μαζί με την εφαρμογή τετραγωνικής ρίζας σε όλα τα πειράματα μας.

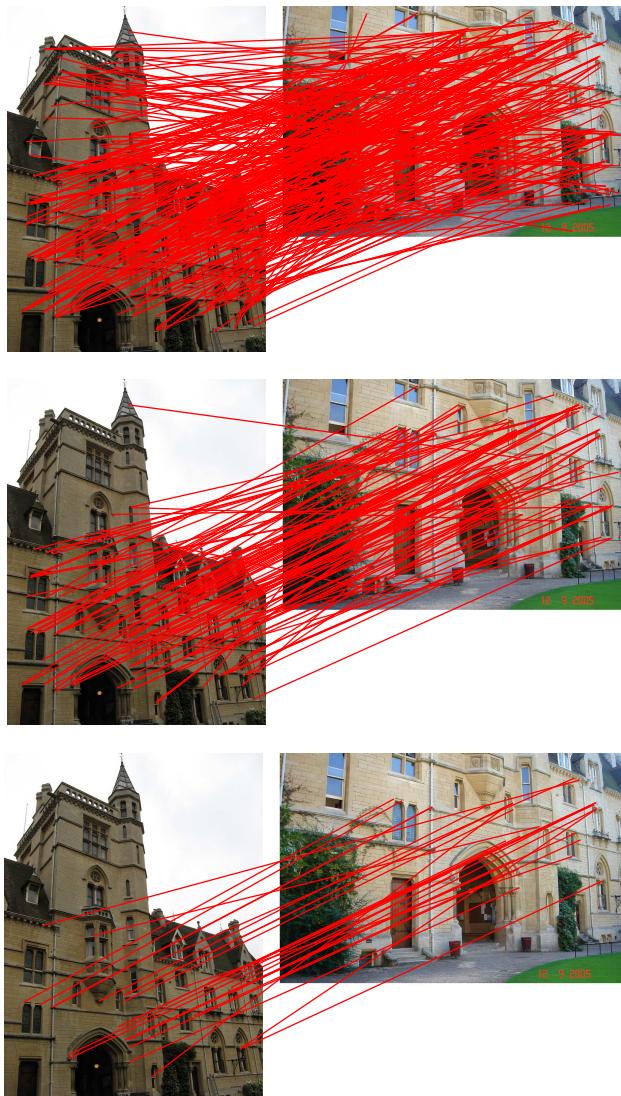
## 5.4 Hamming Embedding με επέκταση ερωτήματος

Σε αυτή την ενότητα ορίζουμε την μέθοδο η οποία βασίζεται στην μέθοδο HE για να κάνει επέκταση ερωτήματος χωρίς να περιλαμβάνει κανένα στάδιο γεωμετρικού ταιριάσματος. Αρχικά περιγράφουμε πως επιλέγονται οι αξιόπιστες εικόνες από το αρχικό σύνολο αποτελεσμάτων. Έπειτα αναλύουμε τον τρόπο με τον οποίο εμπλουτίζουμε την αναπαράσταση της εικόνας ερωτήματος αντλώντας χρήσιμη πληροφορία από τις αξιόπιστες εικόνες.

### 5.4.1 Επιλογή αξιόπιστων εικόνων

Όπως και σε όλες τις μεθόδους για επέκταση ερωτήματος [19, 40, 16, 2], γίνεται ένα αρχικό ερώτημα με το βασικό σύστημα αναζήτησης. Τα αποτελέσματα αναλύονται με σκοπό να αναγνωριστεί ένα υποσύνολο των εικόνων που είναι ταξινομημένες στις πρώτες θέσεις, ως σύνολο αξιόπιστων εικόνων. Αυτές είναι πιθανό να απεικονίζουν το αντικείμενο που φαίνεται στην εικόνα ερωτήματος. Αυτές οι εικόνες μπορούν να παρέχουν χρήσιμη πληροφορία για τον εμπλουτισμό της αναπαράστασης της εικόνας ερωτήματος.

Στην συνέχεια, θα συμβολίζουμε με  $\mathcal{Q}$  το σύνολο των οπτικών χαρακτηριστικών της εικόνας ερωτήματος, και εκείνο της εικόνας της βάσης με  $\mathcal{P}$ . Ως κριτήριο καθορισμού των

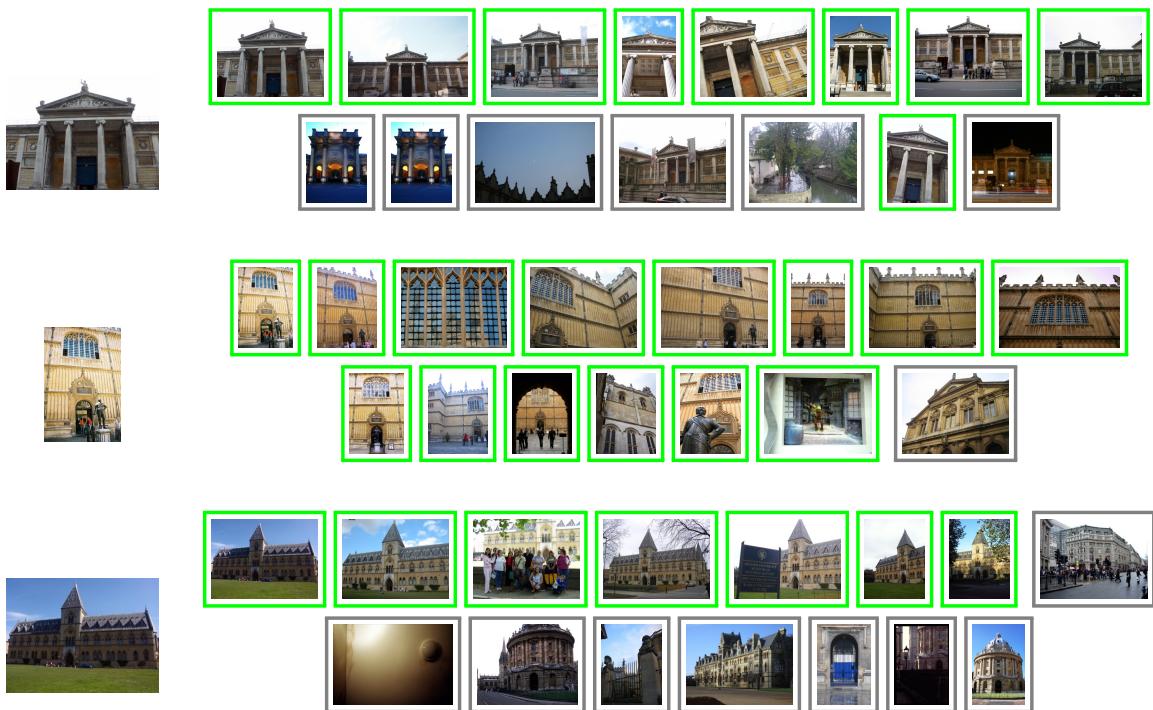


**Σχήμα 5.2:** Ταίριασμα χαρακτηριστικών με την μέθοδο BoW (πάνω), HE με  $h_t = 24$  (μέση) και HE με  $h_t = 16$  (κάτω).

αξιόπιστων εικόνων, μετράμε τον αριθμό  $C(\mathcal{Q}, \mathcal{P})$  των αυστηρών αντιστοιχιών ανάμεσα σε δύο εικόνες. Αυτό δίνεται από

$$C(\mathcal{Q}, \mathcal{P}) = |\{(q, p) \in \mathcal{Q} \times \mathcal{P} : h(q, p) \leq h_t^*\}|, \quad (5.1)$$

όπου το κατώφλι  $h_t^*$  είναι πιο χαμηλό από το τυπικό κατώφλι της μεθόδου HE  $h_t$  το οποίο χρησιμοποιείται για την αρχική κατάταξη των εικόνων. Ένα τέτοιο χαμηλό κατώφλι επιτρέπει έναν υψηλότερο λόγο πραγματικών προς λανθασμένων αντιστοιχιών [37]. Παρέχει έναν αυστηρό τρόπο για να μετρήσει κανείς αντιστοιχίες με έναν τρόπο ο οποίος θυμίζει τον αριθμό των inliers από την μέθοδο RANSAC που χρησιμοποιείται συχνά για γεωμετρική επιβεβαίωση εικόνων [67]. Είναι σαφώς λιγότερο ακριβές από την μέθοδο RANSAC, αλλά έχει το πλεονέκτημα ότι δεν χρησιμοποιεί καθόλου γεωμετρική πληροφορία, συνεπώς είναι και πολύ πιο γρήγορο.



**Σχήμα 5.3:** Παραδείγματα εικόνων ερωτήματος (αριστερά) και τα αντίστοιχα αποτελέσματα αναζήτησης με το βασικό σύστημα αναζήτησης. Οι εικόνες που (δεν) επιλέχθηκαν ως αξιόπιστες παρουσιάζονται με (γκρι) πράσινο πλαίσιο.

Το Σχήμα 5.2 αναπαριστά, για ένα ζευγάρι εικόνων, τις αντιστοιχίες οι οποίες αποκτώνται με χρήση της μεθόδου BoW και HE. Χρησιμοποιούμε δύο διαφορετικές τιμές για το κατώφλι για να δείξουμε την επίδραση του αυστηρού κατωφλίου  $h_t^* = 16$  στον τελικό αριθμό αντιστοιχιών. Παρατηρούμε ότι η μέθοδος HE φιλτράρει πολλές λανθασμένες αντιστοιχίες σε σχέση με την μέθοδο BoW. Με το χαμηλό κατώφλι, το αποτέλεσμα δεν είναι πολύ διαφορετικό από ότι θα έπαιρνε κανείς με την χρήση κάποιας μεθόδου γεωμετρικού ταιριάσματος.

Μία εικόνα χαρακτηρίζεται ως αξιόπιστη εάν έχει τουλάχιστον  $c_t$  αντιστοιχίες, το οποίο μας οδηγεί στον να ορίζουμε το σύνολο των αξιόπιστων εικόνων ως

$$\mathcal{L}_{\mathcal{Q}} = \{\mathcal{P} : C(\mathcal{Q}, \mathcal{P}) \geq c_t\}. \quad (5.2)$$

Πρακτικά, μόνο οι εικόνες που έρχονται πρώτες σε κατάταξη θα θεωρηθούν ως υποψήφιες για αξιόπιστες. Στα πειράματα μας, μετράμε τον αριθμό των αντιστοιχιών με την σχέση 5.1 μόνο για τις 100 πρώτες εικόνες. Το Σχήμα 5.3 δείχνει παραδείγματα από εικόνες ερωτήματος και τις αντίστοιχες αξιόπιστες εικόνες. Παρά το γεγονός ότι επιλέγονται κάποιες λάθος εικόνες και κάποιες σωστές δεν επιλέγονται, το αποτέλεσμα δεν είναι πολύ μακριά από αυτό που θα έπαιρνε κανείς εάν χρησιμοποιούσε κάποια μέθοδο γεωμετρικής επιβεβαίωσης. Αυτό υποστηρίζει το επιχείρημα, πως η μέθοδος HE με ένα χαμηλό κατώφλι μπορεί να χρησιμοποιηθεί για τις ανάγκες τις επέκτασης ερωτήματος.



**Σχήμα 5.4:** Δείγματα από αξιόπιστες εικόνες και τα τοπικά χαρακτηριστικά που αντιστοιχίζονται σε αξιόπιστες οπτικές λέξεις, όταν δεν χρησιμοποιείται γεωμετρική πληροφορία. Αριστερά: Εικόνα ερωτήματος. Πάνω: Τοπικά χαρακτηριστικά που αντιστοιχίζονται σε αξιόπιστες οπτικές λέξεις οι οποίες υπάρχουν και στην εικόνα ερωτήματος. Κάτω: Τοπικά χαρακτηριστικά από το σύνολο των οπτικών λέξεων επαύξησης. Σημείωση: απεικονίζουμε μόνο ένα υποσύνολο των πραγματικών αξιόπιστων οπτικών λέξεων. Κάθε χρώμα συμβολίζει μία διαφορετική οπτική λέξη.

#### 5.4.2 Επέκταση του συνόλου χαρακτηριστικών

Πρώτα ας θυμηθούμε ότι ο κάθε τοπικός περιγραφέας σχετίζεται με μία οπτική λέξη και με μία δυαδική υπογραφή. Η στρατηγική μας για επαύξηση της τωρινής αναπαράστασης, π.χ. πως να εισάγουμε νέα τοπικά χαρακτηριστικά στην αναπαράσταση, βασίζεται κυρίως στην επιλογή οπτικών λέξεων οι οποίες δεν εμφανίζονται στην εικόνα ερωτήματος.

Εφόσον μία μεγάλη μεγάλη μερίδα των αξιόπιστων εικόνων απεικονίζει το ίδιο αντικείμενο, τότε οι οπτικές λέξεις οι οποίες εμφανίζονται συχνά στις αξιόπιστες εικόνες  $\mathcal{L}_Q$  είναι πολύ πιθανό να σχετίζονται με το αντικείμενο ενδιαφέροντος στην εικόνα ερωτήματος κι όχι με το υπόβαθρο. Η διαδικασία επιλογής που προτείνουμε είναι απλή και αποτελείται από την επιλογή των πιο συχνά εμφανιζόμενων οπτικών λέξεων στο  $\mathcal{L}_Q$ . Συγκεκριμένα, ταξινομούμε τις οπτικές λέξεις οι οποίες περιέχονται στις εικόνες του  $\mathcal{L}_Q$  με βάση τον αριθμό των αξιόπιστων εικόνων στις οποίες εμφανίζονται. Οι λέξεις που ταξινομούνται στις πρώτες θέσεις επιλέγονται και ορίζουν το σύνολο των αξιόπιστων οπτικών λέξεων  $N$ , το οποίο μπορεί να περιέχει είτε λέξεις οι οποίες εμφανίζονται στην εικόνα ερωτήματος είτε κι εκείνες που λείπουν από αυτήν. Ονομάζουμε τις τελευταίες ως οπτικές λέξεις επαύξησης. Το πλήθος τους καθορίζεται από την παράμετρο  $\alpha$  η οποία φροντίζει ο αριθμός των αξιόπιστων οπτικών λέξεων να είναι ανάλογος του αρχικού αριθμού οπτικών λέξεων στην

εικόνα ερωτήματος:

$$|\mathcal{N} \setminus \mathcal{N}_{\mathcal{Q}}| = \alpha \cdot |\mathcal{N}_{\mathcal{Q}}|, \quad (5.3)$$

όπου  $\mathcal{N}_{\mathcal{Q}}$  είναι το σύνολο των οπτικών λέξεων που εμφανίζονται στην εικόνα ερωτήματος. Η παράμετρος  $\alpha$  τυπικά παίρνει τιμές στο διάστημα  $(0, 2]$ .

Το αρχικό σύνολο χαρακτηριστικών εμπλουτίζεται με χαρακτηριστικά των αξιόπιστων εικόνων τα οποία έχουν αντιστοιχηθεί με μία από τις αξιόπιστες οπτικές λέξεις. Ορίζονται ως

$$\mathcal{G} = \{p \in \mathcal{P} : \mathcal{P} \in \mathcal{L}_{\mathcal{Q}} \wedge v(p) \in \mathcal{V}\} \quad (5.4)$$

την ένωση όλων των χαρακτηριστικών κάποιας αξιόπιστης εικόνας που έχει αντιστοιχηθεί σε κάποια αξιόπιστη λέξη. Ορίζεται έτσι το σύνολο των χαρακτηριστικών που χρησιμοποιούνται για να εμπλουτισθεί η αρχική αναπαράσταση. Με άλλα λόγια, αυτό το σύνολο ενώνεται με το αρχικό σύνολο χαρακτηριστικών για να κατασκευασθεί η επαυξημένη αναπαράσταση ως

$$\mathcal{Q}_E = \mathcal{Q} \cup \mathcal{G}. \quad (5.5)$$

Το Σχήμα 5.4 απεικονίζει μερικά χαρακτηριστικά αξιόπιστων εικόνων που έχουν αντιστοιχηθεί σε αξιόπιστες οπτικές λέξεις. Παρατηρούμε ότι παρά το γεγονός ότι δεν έχει χρησιμοποιηθεί γεωμετρική πληροφορία, τα επιλεγμένα τοπικά χαρακτηριστικά προέρχονται από το αντικείμενο ενδιαφέροντος κι όχι από το υπόβαθρο. Επίσης, κάθε οπτική λέξη αντιστοιχεί όχι απλώς σε παρόμοιες περιοχές της εικόνας, αλλά ακριβώς στο ίδιο σημείο του αντικείμενου ενδιαφέροντος, όπως ακριβώς εάν είχε χρησιμοποιηθεί γεωμετρικό ταίριασμα.

Ένας απλός τρόπος για να κατασκευασθεί μία εμπλουτισμένη αναπαράσταση που θα αποτελέσει το νέο ερώτημα είναι να χρησιμοποιηθεί το επαυξημένο σύνολο χαρακτηριστικών. Όμως αυτό, όπως και σε υπάρχουσες τεχνικές για επέκταση ερωτήματος, οδηγεί σε πολύ μεγαλύτερη πολυπλοκότητα του νέου ερωτήματος, καθώς ο αριθμός των χαρακτηριστικών αυξάνεται δραματικά. Στην επόμενη ενότητα προτείνουμε μία μέθοδο για την αντιμετώπιση αυτού του φαινομένου με την οποία θα μειώσουμε το μέγεθος του επαυξημένου συνόλου χαρακτηριστικών.

## 5.5 Επέκταση ερωτήματος με συνάθροιση περιγραφέων

Η τεχνική για επέκταση ερωτήματος που προτείνεται στο [19] παίρνει μέσο όρο των ιστογραμμάτων της μεθόδου BoW για να παράγει το νέο ερώτημα. Στην δική μας περίπτωση, το επαυξημένο σύνολο περιέχει πολλαπλά χαρακτηριστικά τα οποία αναπαριστούν την ίδια περιοχή, σε διαφορετικές όμως εικόνες, είτε στην εικόνα ερωτήματος είτε



**Σχήμα 5.5:** Χαρακτηριστικά που επιλέχθηκαν για το επαυξημένο σύνολο μίας συγκεκριμένης εικόνας ερωτήματος (αριστερά) χωρίς (μέση) και με την χρήση γεωμετρίας (δεξιά). Με την χρήση γεωμετρικού ταιριάσματος: χαρακτηριστικά που προβάλλονται εκτός της περιοχής ερωτήματος απορρίπτονται (κόκκινα), ενώ τα υπόλοιπα (μπλε και πράσινα) διατηρούνται. Αυτά που έχουν αντιστοιχηθεί σε αξιόπιστες οπτικές λέξεις απεικονίζονται με πράσινο. Χωρίς την χρήση γεωμετρικού ταιριάσματος: τα χαρακτηριστικά τα οποία έχουν αντιστοιχηθεί σε αξιόπιστες οπτικές λέξεις απεικονίζονται με κυανό ή κίτρινο χρώμα, με κίτρινο να είναι όσα έχουν αντιστοιχηθεί συγκεκριμένα σε οπτικές λέξεις επαύξησης. Όσα απορρίπτονται απεικονίζονται με κόκκινο.

όχι. Περιγραφείς οι οποίοι σχετίζονται με την ίδια περιοχή είναι πιθανό να έχουν και παρόμοιες δυαδικές υπογραφές. Η στρατηγική που περιγράφεται παρακάτω χρησιμοποιεί αυτή την ιδιότητα για να παράγει ένα νέο σύνολο περιγραφέων με λιγότερες επαναλήψεις και λιγότερη φλυαρία.

Αρχικά να σημειώσουμε ότι ο τρόπος με τον οποίο επιλέχθηκαν οι αξιόπιστες εικόνες και τα αξιόπιστα χαρακτηριστικά μπορούν να εισάγουν και κάποια λάθη. Αυτό είναι το τίμημα το οποίο πληρώνεται για την μη χρήση γεωμετρικής πληροφορίας. Η διαδικασία συνάθροισης που περιγράφεται παρακάτω, έρχεται να συμπληρώσει την επιλογή των αξιόπιστων χαρακτηριστικών και να διορθώσει τυχόν τέτοια λάθη. Η διαδικασία αυτή είναι εύρωστη σε τέτοια λάθος επιλεγμένα χαρακτηριστικά, σε αντίθεση με την συνάθροιση των ιστογραμμάτων [19] η οποία δεν είναι.

Η διαδικασία συνάθροισης η οποία προτείνουμε είναι εμπνευσμένη από μεθόδους [65, 38] όπως η τεχνική VLAD, κατά την οποία οι περιγραφείς συναθροίζονται ανά οπτική λέξη. Στην δική μας περίπτωση, συναθροίζουμε τα χαρακτηριστικά του  $Q_E$  τα οποία έχουν αντιστοιχηθεί στην ίδια οπτική λέξη. Επομένως παράγουμε ακριβώς μία δυαδική υπογραφή ανά οπτική λέξη του  $Q_E$ .

Το κίνητρό μας είναι το γεγονός ότι οι σωστά επιλεγμένοι περιγραφείς θα υπερισχύουν από τους λανθασμένους. Αυτό πρακτικά συμβαίνει επειδή οι πραγματικές αντιστοιχίες είναι περισσότερες και σχετίζονται με πιο συμβατές μεταξύ τους δυαδικές υπογραφές.

Για κάθε οπτική λέξη  $v$  η οποία εμφανίζεται στο  $\mathcal{Q}_E$ , μία νέα δυαδική υπογραφή  $b(v)$  παράγεται υπολογίζοντας την ενδιάμεση τιμή όλων των δυαδικών διανυσμάτων στο  $\mathcal{Q}_E$  που σχετίζονται με το  $v$ . Σε περίπτωση ισοπαλίας μεταξύ 0 και 1, επιλέγουμε τυχαία. Το νέο σύνολο χαρακτηριστικών αποτελείται από μία δυαδική υπογραφή ανά οπτική λέξη και λειτουργεί σαν νέο ερώτημα στο σύστημα.

Θα αναφερόμαστε στην μέθοδο αυτή ως επέκταση ερωτήματος Hamming (Hamming Query Expansion - HQE) στην συνέχεια.

## 5.6 Γεωμετρική πληροφορία

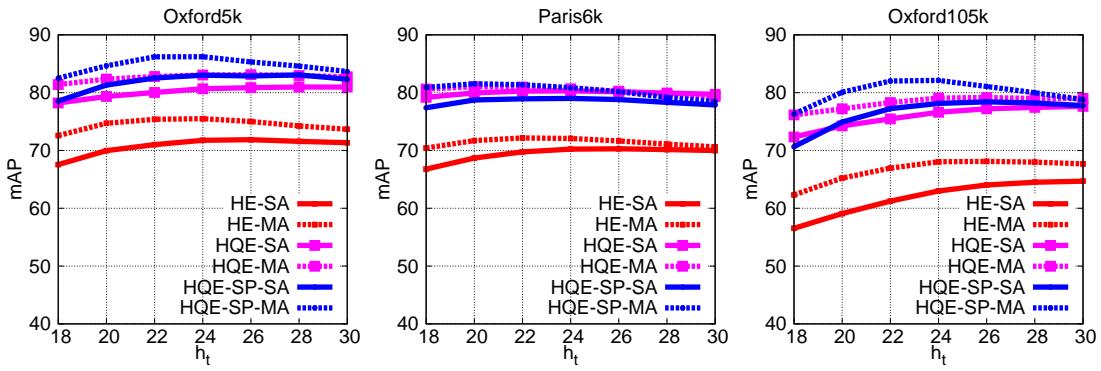
Σε αυτή την ενότητα προτείνουμε μια παραλλαγή της μέχρι τώρα μεθόδου, με την οποία ελαχιστοποιούμε τις λάθος επιλογές αξιόπιστων χαρακτηριστικών με το να εισάγουμε γεωμετρική πληροφορία στην διαδικασία αυτή. Για αυτό τον λόγο και όπως θα δείξουμε στα πειράματα, δεν είναι τόσο γρήγορη όσο η μέθοδος HQE που περιγράφηκε στις ενότητες 5.4 και 5.5. Παρόλα αυτά προκαλεί μία επιπλέον βελτίωση της απόδοσης κι έτσι παραμένει ενδιαφέροντα.

Η διαδικασία έχει ως εξής. Οι αντιστοιχίες συλλέγονται μέσω της μεθόδου HE, π.χ., επιστρέφονται με το αρχικό ερώτημα. Αντί να υπολογίσουμε τον αριθμό των αντιστοιχιών με την σχέση 5.1, βασιζόμαστε στον αριθμό  $C_g(\mathcal{Q}, \mathcal{P})$  των inliers που βρίσκονται μέσω μιας διαδικασίας γεωμετρικού ταιριάσματος. Για αυτό τον σκοπό χρησιμοποιήσαμε το γεωμετρικό ταίριασμα των Philbin *et al.* [67].

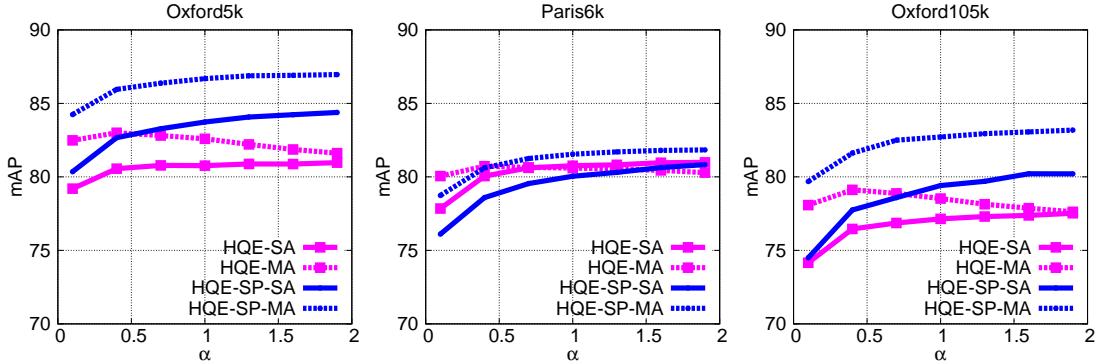
Ομοίως με άλλες μεθόδους γεωμετρικού ταιριάσματος, αυτή η διαδικασία εφαρμόζεται μόνο στις εικόνες που κατατάχθηκαν στις πρώτες θέσεις. Μία εικόνα χαρακτηρίζεται ως αξιόπιστη εάν ο αριθμός των inliers ξεπερνά ένα προκαθορισμένο κατώφλι. Η εκτίμηση ενός αφινικού μετασχηματισμού μέσα από την διαδικασία γεωμετρικού ταιριάσματος χρησιμοποιείται για να φιλτράρει το σύνολο χαρακτηριστικών της εικόνας της βάσης. Όπως προτείνανε οι Chum *et al.* [19], οι περιοχές των εικόνων της βάσεις προβάλλονται πίσω στην εικόνα ερωτήματος και όσες βρίσκονται εκτός της περιοχής ερωτήματος απορρίπτονται.

Τα υπόλοιπα βήματα παραμένουν όμοια με την μέθοδο HQE όπως περιγράφηκε στις ενότητες 5.4 και 5.5. Η μόνη διαφορά είναι πως το σύνολο των αξιόπιστων χαρακτηριστικών είναι διαφορετικό. Επομένως, πρώτα επιλέγουμε τις αξιόπιστες οπτικές λέξεις και εκτελούμε την επαύξηση του συνόλου χαρακτηριστικών. Η συνάθροιση εφαρμόζεται παρομοίως.

## Πειράματα



Σχήμα 5.6: Επίδραση της παραμέτρου  $h_t$  στην απόδοση των μεθόδων HE και HQE.



Σχήμα 5.7: Επίδραση της παραμέτρου  $\alpha$ . Απόδοση της μεθόδου HQE για μεταβλητό αριθμό νέων οπτικών λέξεων στο επαυξημένο ερώτημα.

Το Σχήμα 5.5 απεικονίζει τα χαρακτηριστικά που επιλέγονται με ή χωρίς χρήση γεωμετρίας. Παρατηρούμε πως ακόμα και χωρίς γεωμετρία τα επιλεγμένα χαρακτηριστικά εντοπίζονται πάνω στο αντικείμενο ενδιαφέροντος. Με την χρήση γεωμετρίας όμως φιλτράρονται μερικά χαρακτηριστικά τα οποία εντοπίζονται εκτός του αντικειμένου ενδιαφέροντος.

## 5.7 Πειράματα

### 5.7.1 Σύνολα εικόνων και πρωτόκολλο πειραμάτων

**Σύνολα εικόνων και μετρικές αξιολόγησης.** Οι τεχνικές για επέκταση ερωτήματος είναι αποτελεσματικές μόνο όταν υπάρχουν αρκετές όμοιες εικόνες με την εικόνα ερωτήματος. Αξιολογούμε την προτεινόμενη μέθοδο σε δύο δημόσια διαθέσιμα σύνολα εικόνων τέτοιου τύπου. Το Oxford5k Buildings [67] και το Paris [68], αλλά επίσης και σε ένα το οποίο έχει λίγες όμοιες εικόνες με την κάθε εικόνα ερωτήματος, αυτό είναι το UKB [60]. Η αξιολόγηση γίνεται μέσω του μέτρου mean Average Precision (mAP) για τα πρώτα δύο σύνολα και το ειδικά σχεδιασμένο μέτρο για το UKB: ο μέσος αριθμός των σωστών εικόνων που κατατάχθηκαν στις πρώτες 4 θέσεις (από 0 έως 4). Το πείραμα μεγαλύτερης κλίμα-

κας που διεξάγουμε γίνεται με την προσθήκη 100k εικόνων περίσπασης στο Oxford5k, το οποίο αποτελεί το Oxford105k.

**Τοπικά χαρακτηριστικά και πρωτόκολλο.** Για τα σύνολα Oxford5k και Paris, χρησιμοποιήσαμε τον τροποποιημένο ανιχνευτή Hessian-Affine που προτάθηκε από τους Perdoch *et al.* [62] για να εντοπίσουμε τοπικά χαρακτηριστικά. Οι SIFT περιγραφείς που εξάγονται επεξεργάζονται με τον τρόπο που περιγράφηκε στην ενότητα 5.2. Για το UKB, χρησιμοποιήσαμε τα χαρακτηριστικά που παρέχονται από τα [39, 22]. Ακολουθούμε την διαδικασία κατά την οποία κανείς φτιάχνει το λεξικό από ανεξάρτητο σύνολο εικόνων.

Εκτός αν αναφέρεται κάτι διαφορετικό, χρησιμοποιούμε λεξικό  $k = 65, 536$  οπτικών λέξεων, δυαδικές υπογραφές με 64 διαστάσεις, και χρησιμοποιούμε την μέθοδο HE με βάρη και κανονικοποίηση για ομοβροντία. Σε όλα τα πειράματα μας οι αξιόπιστες εικόνες επιλέγονται ανάμεσα από εκείνες που κατατάχθηκαν στις 100 πρώτες θέσεις. Όταν γίνεται χρήση πολλαπλής ανάθεσης (MA), αυτό εφαρμόζεται μόνο από την μεριά της εικόνας ερωτήματος και για τις 3 πλησιέστερες οπτικές λέξεις.

Με την πολλαπλή ανάθεση δημιουργούνται περισσότερες αντιστοιχίες, οπότε η πιθανότητα να βρεθεί κάποια λανθασμένη αντιστοιχία αυξάνεται ακόμα και με το γεωμετρικό ταίριασμα, οπότε οι αντίστοιχες παράμετροι θα πρέπει να προσαρμοστούν [37]. Ορίζουμε τον ελάχιστον αριθμό αντιστοιχιών ως  $c_t = 4$  με SA και  $c_t = 5$  με MA.

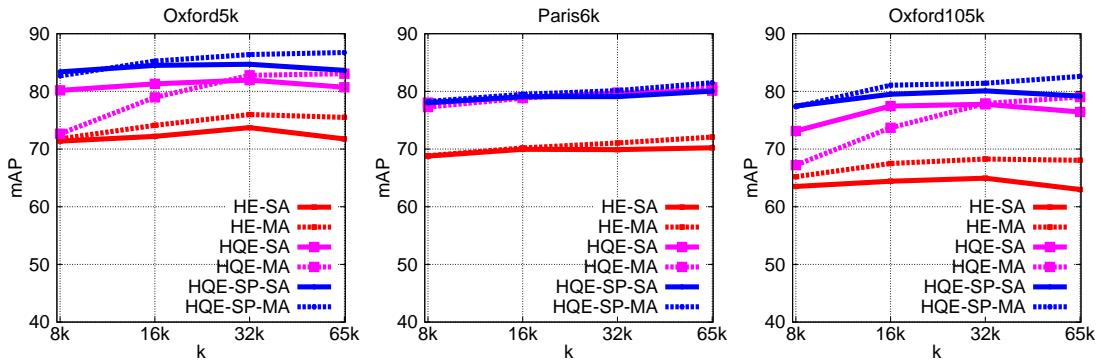
Υπάρχουν δύο παράγοντες που εισάγουν τυχαιότητα στην μέθοδο μας: Ο τυχαίος πίνακας προβολής (της μεθόδου HE) και η τυχαία απόφαση για τις ισοπαλίες όταν κάνουμε συνάθροιση. Για αυτό τον λόγο, εκτελούμε το κάθε πείραμα 5 φορές και αναφέρουμε μέση τιμή και τυπική απόκλιση.

### 5.7.2 Επίδραση των παραμέτρων

**Κατώφλια.** Το αυστηρό κατώφλι  $h_t^*$  είναι σταθερό και το ορίζουμε ίσο με 16 σε κάθε περίπτωση. Το Σχήμα 5.6 δείχνει την επίδραση της παραμέτρου  $h_t$ . Η απόδοση δεν είναι τόσο ευαίσθητη γύρω από την βέλτιστη τιμή  $h_t = 22$  ή  $h_t = 24$ . Παρατηρούμε ότι ήδη η μέθοδος HQE δίνει μία αισθητή βελτίωση ως προς την μέθοδο HE. Στα υπόλοιπα πειράματα θέτουμε  $h_t = 24$ . Σε αυτό το αρχικό πείραμα είχαμε θέσει  $\alpha = 0.5$ , το οποίο σημαίνει πως το μέγεθος του νέου ερωτήματος είναι το πολύ 1.5 μεγαλύτερο από το αρχικό. Πρακτικά είναι πολύ μικρότερο λόγω της συνάθροισης. Για παράδειγμα στον Πίνακα 5.3 συγκρίνουμε τον αριθμό των χαρακτηριστικών στο αρχικό σύνολο και το επαυξημένο.

Η παράμετρος  $\alpha$  ελέγχει το μέγεθος του επαυξημένου ερωτήματος. Το Σχήμα 5.7 παρουσιάζει την επίδραση του στην απόδοση. Η μέθοδος HQE χωρίς γεωμετρία πετυχαίνει μία μέγιστη τιμή και ύστερα φθίνει. Αυτό σημαίνει ότι δεν πρέπει να επιλεχθούν πολλές οπτικές λέξεις όταν δεν χρησιμοποιείται γεωμετρία, καθώς εισάγονται λάθη. Σε αντίθεση, με

## Πειράματα



Σχήμα 5.8: Επίδραση των μεγέθους του λεξικού στην απόδοση των μεθόδων HE, HQE και HQE με χρήση γεωμετρίας.

W burst MA HQE SP	Oxford5k	Paris6k	Oxford105k
X	66.9	65.7	55.5
X X	70.4	68.4	59.6
X X X	71.7	70.2	62.9
X X X X	75.4	72.0	68.0
X X X X X	83.0	80.6	79.0
BoW	86.8	81.5	82.6
BoW	53.3	54.8	44.2

Πίνακας 5.2: Απόδοση (mAP) για τα διαφορετικά συστατικά από τα οποία αποτελείται η προτεινόμενη μέθοδος. Η αρχική μέθοδος είναι το παραδοσιακό Hamming Embedding χωρίς βάρη. Με W συμβολίζουμε την χρήση βαρών.

χρήση γεωμετρίας κανείς μπορεί πιο εύκολα να προσθέτει επιπλέον οπτικές λέξεις και η απόδοση να αυξάνεται. Μια ισορροπημένη επιλογή ανάμεσα στην απόδοση και στην πολυπλοκότητα είναι να θέσουμε  $\alpha = 0.5$  και  $1.0$  χωρίς και με γεωμετρία αντιστοίχως.

Το μέγεθος του λεξικού  $k$  είναι σημαντικό στις περισσότερες των μεθόδων που βασίζονται στην μέθοδο BoW. Το Σχήμα 5.8 δείχνει ότι αυτό δεν είναι ίδιο για μεθόδους που βασίζονται στην μέθοδο HE.

**Βάρη, ομοβροντία και HQE.** Ο Πίνακας 5.2 συνοψίζει τις συνεισφορές του κάθε συστατικού από τα οποία αποτελείται η προτεινόμενη μέθοδος.

**Συνάθροιση.** Ο Πίνακας 5.3 αποκαλύπτει το διπλό όφελος από την συνάθροιση χαρακτηριστικών που γίνεται στην μέθοδο μας. Ωφελεί τόσο την απόδοση, όσο και την πολυπλοκότητα. Η συνάθροιση δυαδικών υπογραφών μειώνει το μέγεθος του επαυξημένου ερωτήματος κι έτσι έχει θετική επίδραση στην πολυπλοκότητα, όπως φαίνεται και στον Πίνακα 5.3. Η συνάθροιση μειώνει το μέγεθος του επαυξημένου ερωτήματος κατά μία τάξη μεγέθους, το οποίο γίνεται συγκρίσιμο με το μέγεθος του αρχικού ερωτήματος.

Επιπλέον, φαίνεται πως η συνάθροιση βελτιώνει την απόδοση σε κάθε περίπτωση. Για να το δείξουμε αυτό συγκρίναμε την μέθοδο HQE (π.χ., με συνάθροιση) με μία μέθοδο η

Dataset	Method	mAP		$ Q $	
		SA	MA	SA	MA
Oxford5k	HE	71.7	75.4	1,362	4,089
	HQE/b.a.	79.0	82.0	11,937	27,345
	HQE	80.7	<b>83.0</b>	1,810	<b>5,030</b>
Paris6k	HE	70.2	72.0	1,460	4,382
	HQE/b.a.	76.6	77.3	35,982	66,665
	HQE	80.2	<b>80.6</b>	1,843	<b>5,045</b>
Oxford105k	HE	62.9	68.0	1,362	4,088
	HQE/b.a.	73.5	76.5	12,176	28,699
	HQE	75.6	<b>79.0</b>	1,810	<b>5,030</b>

Πίνακας 5.3: Απόδοση και μέσος αριθμός χαρακτηριστικών του ερωτήματος  $|Q|$  για την βασική μέθοδο HE, το HQE και το ίδιο ερώτημα πριν την συνάθροιση (HQE/b.a.). Η συνάθροιση φαίνεται σημαντική όχι μόνο εξαιτίας της μείωσης της πολυπλοκότητας αλλά και λόγω της αύξησης της απόδοσης.

Building	GT	HE	HQE		HQE-SP	
		mAP	$ \mathcal{L}_Q $	mAP	$ \mathcal{L}_Q $	mAP
All Souls	183	78.2	47.0	94.6	44.8	97.3
Ashmolean	56	63.7	10.9	76.1	9.9	80.8
Balliol	30	72.7	15.8	81.0	8.0	82.1
Bodleian	54	66.4	33.4	94.5	19.8	86.9
Christ Church	211	74.9	39.5	75.7	45.1	90.7
Cornmarket	22	69.5	9.6	64.9	6.4	71.4
Hertford	55	87.7	41.5	95.0	43.5	98.3
Keble	18	93.0	9.5	96.5	7.6	99.5
Magdalen	157	29.9	15.6	36.5	8.8	48.6
Pitt Rivers	16	100.0	9.7	99.7	7.0	100.0
Radcliffe	569	93.9	97.1	98.5	96.0	98.7

Πίνακας 5.4: Oxford5k: Αναλυτικές λεπτομέρειες ως προς κτίριο για τον αριθμό των σωστών εικόνων, των αξιόπιστων εικόνων και της απόδοσης για τις μεθόδους HE, HQE και HQE με γεωμετρία.

οποία χρησιμοποιεί ως επαυξημένο ερώτημα το σύνολο το οποίο ορίζεται στην σχέση 5.5, π.χ., πριν την συνάθροιση. Η συνάθροιση θα μειώσει την επίδραση των από λάθος επιλεγμένων χαρακτηριστικών και παράλληλα θα αφαιρέσει την φλυαρία από τον σύνολο αυτό.

**Αναλυτική απόδοση στο Oxford5k.** Ο Πίνακας 5.4 παρουσιάζει αναλυτικά στατιστικά για την βάση Oxford5k. Η διαδικασία με την οποία επιλέγουμε τις αξιόπιστες εικόνες φαίνεται να προσαρμόζεται εύκολα στα διαφορετικά κτίρια παρά την ποικιλία στον αριθμό των σωστών εικόνων στην βάση. Αυτό ισχύει είτε με είτε χωρίς την χρήση γεωμετρίας.

Method	SP	MA	Oxford5k	Paris6k	Oxford105k
Perdoch [62]	×		78.4	N/A	72.8
Perdoch [62]	×	×	82.2	N/A	77.2
Mikulik [57]	×	×	<b>84.9</b>	<b>82.4</b>	<b>79.5</b>
Chum [16]	×		82.7	80.5	76.7
Arandjelovic [2]	×		80.9	76.5	72.2
HQE			$80.7 \pm 0.9$	$80.2 \pm 0.2$	$76.6 \pm 1.1$
HQE-SP	×		$83.7 \pm 0.7$	$80.0 \pm 0.2$	$79.4 \pm 0.6$
HQE		×	$83.0 \pm 0.9$	$80.6 \pm 0.2$	$79.0 \pm 1.0$
HQE-SP	×	×	$86.8 \pm 0.3$	$81.5 \pm 0.3$	$82.6 \pm 0.4$
HQE 128bits			$81.0 \pm 0.5$	$81.5 \pm 0.2$	$76.9 \pm 0.6$
HQE-SP 128bits	×		$85.3 \pm 0.4$	$81.3 \pm 0.3$	$80.8 \pm 0.5$
HQE 128bits		×	$83.8 \pm 0.3$	$82.8 \pm 0.1$	$80.4 \pm 0.5$
HQE-SP 128bits	×	×	$88.0 \pm 0.3$	$82.8 \pm 0.2$	$84.0 \pm 0.2$

Πίνακας 5.5: Σύγκριση της απόδοσης με δημοσιευμένα αποτελέσματα για τις βάσεις Oxford5k, Paris6k και Oxford105k. Η τυπική απόκλιση προέρχεται από την εκτέλεση 5 ανεξάρτητων πειραμάτων.

Jégou [33]	Jégou [39]	Danfeng [22]	HE-MA	HQE-MA
3.64	<b>3.68</b>	<b>3.67</b>	3.59	<b>3.67</b>

Πίνακας 5.6: Σύγκριση απόδοσης στην βάση UKB.

### 5.7.3 Σύγκριση με γνωστές μεθόδους

**Oxford5k, Paris6k και Oxford105k.** Ο Πίνακας 5.5 συγκρίνει την προτεινόμενη μέθοδο με δημοσιευμένα αποτελέσματα άλλων μεθόδων για επέκταση ερωτήματος. Για να είναι η σύγκριση δίκαιη, αναφέρουμε τα αποτελέσματα με χρήση του ίδιου ακριβώς ανιχνευτή τοπικών χαρακτηριστικών [62]. Παρουσιάζουμε αποτελέσματα ακόμα και για δυαδικές υπογραφές με διάσταση 128.

Είναι ενδιαφέρον το γεγονός ότι η μέθοδος μας ακόμα και χωρίς την χρήση γεωμετρίας ξεπερνά όλες τις άλλες μεθόδους στα Oxford105k και Paris6k. Ενώ η HQE-SP τις ξεπερνά σε κάθε περίπτωση.

Το UKB [60] είναι μία βάση με λίγες όμοιες εικόνες στην βάση για κάθε εικόνα ερωτήματος (4, συμπεριλαμβανομένης της εικόνας ερωτήματος). Συνεπώς οι τεχνικές για επέκταση ερωτήματος δεν αναμένεται να αποδώσουν καλά. Για αυτό το σύνολο εικόνων και μόνο θεωρούμε πως αξιόπιστες είναι πάντα οι 2 εικόνες που έχουν καταταχθεί στις πρώτες θέσεις. Ο Πίνακας 5.6 παρουσιάζει την βελτίωση που παρέχει η τεχνική μας. Έχει απόδοση το ίδιο καλή με άλλες μεθόδους οι οποίες απαιτούν να έχει κανείς εκ των προτέρων ταιριάζει όλες τις εικόνες της βάσης μεταξύ τους, το οποίο έχει ένα μεγάλο υπολογιστικό κόστος.

Method	HE	HQE	HQE-SP
SA	30 ms	79 ms	731 ms
MA	76 ms	204 ms	955 ms

Πίνακας 5.7: Μέσος χρόνος ανά ερώτημα για την βασική μέθοδο HE και τις δικές μας στην βάση Oxford105k όταν χρησιμοποιείται υλοποίηση ενός πυρήνα.

#### 5.7.4 Πολυπλοκότητα: χρόνος και μέγεθος ερωτήματος

Το αρχικό ερώτημα περιέχει μία δυαδική υπογραφή ανά χαρακτηριστικό και πολλά χαρακτηριστικά μπορεί να αντιστοιχηθούν στην ίδια οπτική λέξη, ιδιαίτερα με χρήση πολλαπλής ανάθεσης. Επιπλέον, το επαυξημένο σύνολο ερωτήματος, όπως ορίζεται πριν την συνάθροιση, είναι πολύ μεγαλύτερο καθώς πολλές εικόνες συνεισφέρουν σε αυτό με τα αξιόπιστα χαρακτηριστικά τους, όπως είδαμε στον Πίνακα 5.3. Χάρη στην μέθοδο HQE, μόνο μία δυαδική υπογραφή ανά οπτική λέξη διατηρείται. Αυτό ευνοεί πολύ την πολυπλοκότητα της μεθόδου κι έτσι το επαυξημένο ερώτημα είμαι απλά λίγο μεγαλύτερο από το αρχικό.

Στον Πίνακα 5.7 αναφέρουμε τον μέσο χρόνο αναζήτησης στην βάση Oxford105k. Έχουν μετρηθεί με υλοποίηση που χρησιμοποιεί έναν πυρήνα σε επεξεργαστή 3.2 Ghz. Όπως ήταν αναμενόμενο οι χρόνοι για την μέθοδο χωρίς γεωμετρία είναι πολύ ανταγωνιστικοί. Ως σημείο σύγκρισης αναφέρουμε ότι ο καλύτερος χρόνος που έχει αναφερθεί [62] είναι 509ms στο Oxford105k με επεξεργαστή  $4 \times 3.0\text{Ghz}$ .



# Κεφάλαιο 6

## Συμπληρωματικές εργασίες

Σε αυτό το κεφάλαιο αναφέρουμε επιπλέον μεθόδους οι οποίες αναπτύχθηκαν και προτάθηκαν κατά την διάρκεια της διατριβής αλλά δεν περιγράφονται αναλυτικά στα προηγούμενα κεφάλαια.

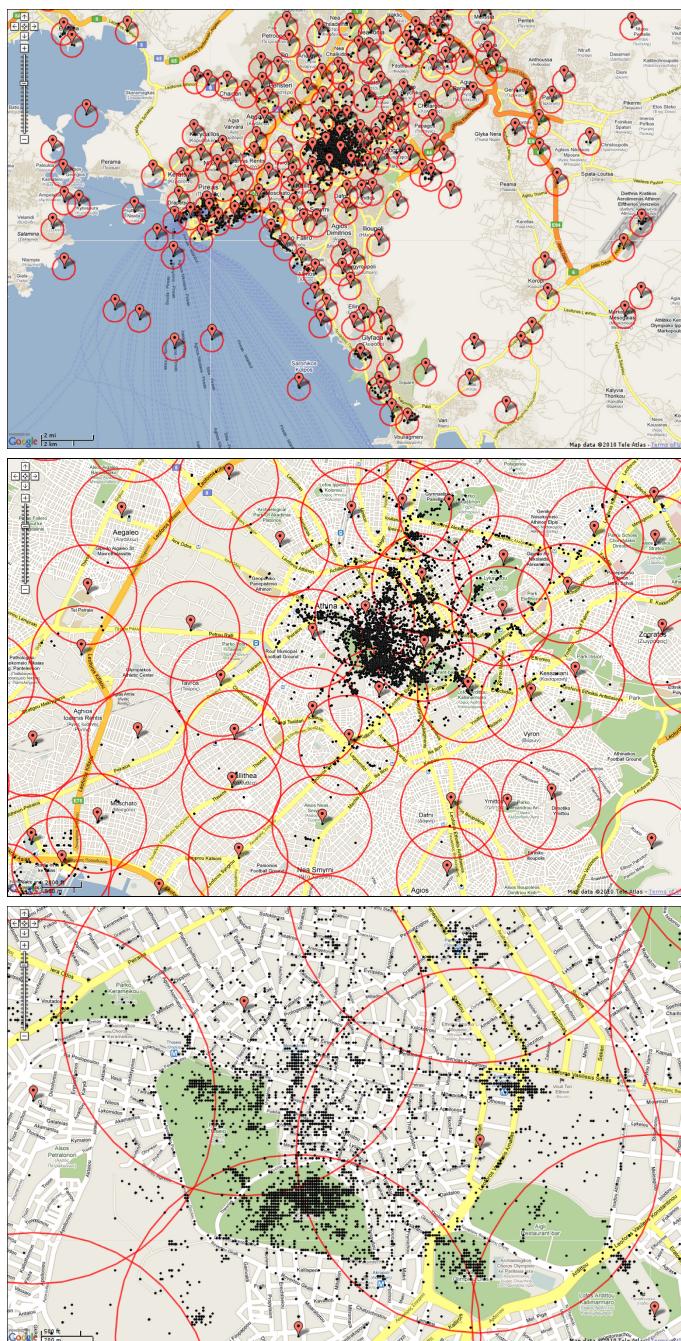
### 6.1 Χάρτες Σκηνών

Εκμεταλλευόμαστε τις συλλογές εικόνων οι οποίες είναι δημόσια διαθέσιμες μαζί με προσημείωση, όπως το γεωγραφικό στίγμα, και προτείνουμε μία μέθοδο για ομαδοποίηση εικόνων [5]. Η μέθοδος μας είναι μια διαδικασία την οποία μπορεί κανείς να δει ως κβαντισμό διανύσματος, με τον οποίο συμπιέζεται μία ομάδα από οπτικά όμοιες εικόνες ενώ παρέχεται ένα όριο στην διαστρέβλωση. Αυτό μας επιτρέπει για παράδειγμα να αναπαραστήσουμε μερικές εκατοντάδες εικόνων οι οποίες απεικονίζουν τον Παρθενώνα, με μόνο μερικούς χάρτες σκηνών.

Ξεκινώντας από μία συλλογή εικόνων οι οποίες συνοδεύονται από γεωγραφικό στίγμα, τις ομαδοποιούμε πρώτα με βάση την γεωγραφική τους εγγύτητα κι έπειτα με βάση την οπτική ομοιότητα. Για να εντοπίσουμε εικόνες οι οποίες πιθανόν να απεικονίζουν κάποιο όμοιο αντικείμενο, βασιζόμαστε σε μία μέγιστη γεωγραφική απόσταση, για παράδειγμα δύο εικόνες οι οποίες έχουν τραβηγθεί 2km η μία μακριά από την άλλη δεν περιμένουμε να απεικονίζουν το ίδιο αντικείμενο. Το Σχήμα 6.1 απεικονίζει τις γεωγραφικές ομάδες από την πόλη της Αθήνας σε διαφορετικά επίπεδα λεπτομέρειας.

Για την οπτική ομοιότητα βασιζόμαστε σε ένα σύστημα αναζήτησης με βάση την μέθοδο BoW και την χρήση ανάστροφου αρχείου. Γεωμετρική επιβεβαίωση γίνεται τελικά με την μέθοδο Fast Spatial Matching [66] και από εκεί προκύπτει και ο βαθμός οπτικής ομοιότητας ο οποίος χρησιμοποιείται για την ομαδοποίηση. Εικόνες οι οποίες ανήκουν στην ίδια οπτική ομάδα, ευθυγραμμίζονται, όπως φαίνεται στο Σχήμα 6.2. Τελικά δημιουργείται ένας χάρτης σκηνών για τον οποίο διατηρείται ένας αντιπρόσωπος από τα χαρακτηριστικά τα

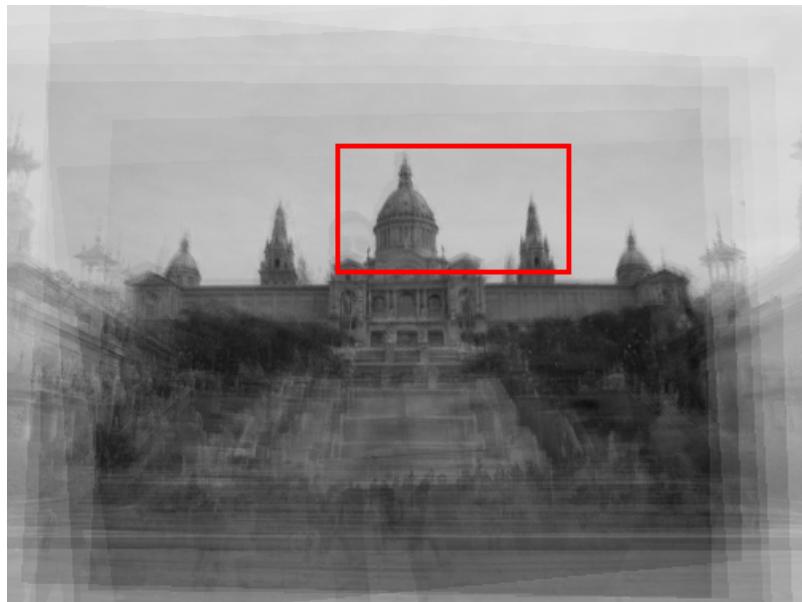
## Χάρτες Σκηνών



**Σχήμα 6.1:** Χάρτης της Αθήνας ο οποίος απεικονίζει τις γεωγραφικές ομάδες σε τρεις κλίμακες.

οποία αντιστοιχούν στο ίδιο φυσικό σημείο ανάμεσα στις διαφορετικές εικόνες. Η γεωγραφική και οπτική ομαδοποίηση αλλά και εκείνη των χαρακτηριστικών σε έναν χάρτη σκηνών γίνεται με χρήση του αλγορίθμου *kernel vector quantization* (KVQ) των Tipping και Schölkopf [77].

Οι χάρτες σκηνών προσφέρουν έναν βαθμό συμπίεσης στην αναζήτηση εικόνων με χρήση ανάστροφου αρχείου αλλά και ταυτόχρονα καλύτερη απόδοση από την χρήση της βασικής μεθόδου BoW.

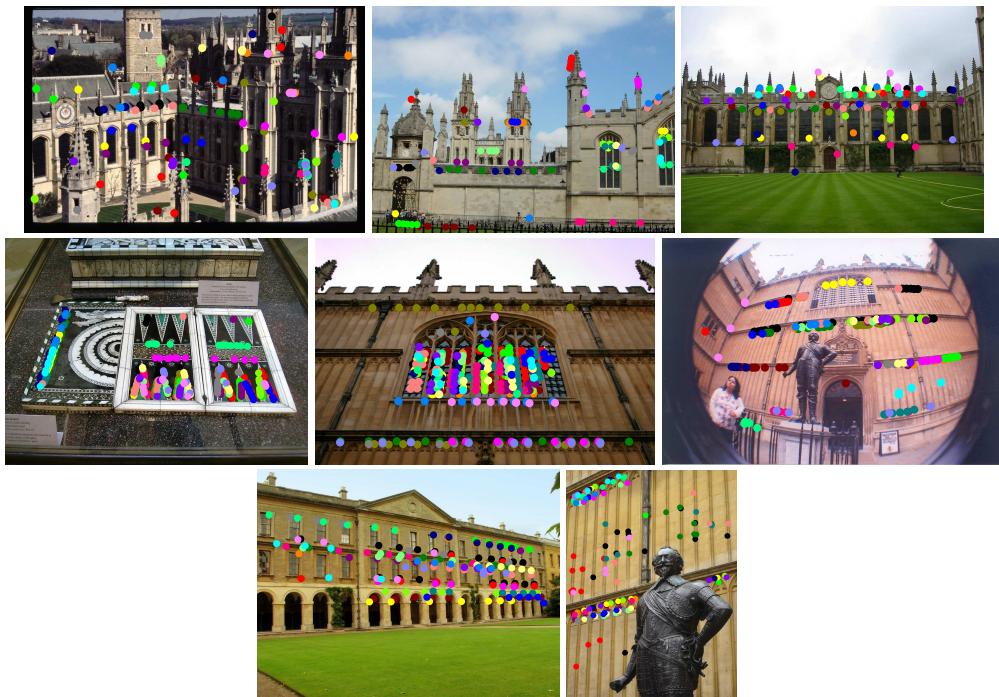


**Σχήμα 6.2:** Χάρτης σκηνών ο οποίος κατασκευάστηκε από 10 εικόνες του ορόσημου Palau Nacional, Montjuic, Barcelona.

## 6.2 Επιλεκτικός πυρήνας συνάθροισης

Σε αυτή την δουλειά [79] εργαζόμαστε με μία οικογένεια μετρικών οι οποίες χρησιμοποιούνται για την σύγκριση δύο εικόνων με χρήση των τοπικών περιγραφέων. Εμπεριέχει τον περιγραφέα VLAD [35] και τεχνικές ταιριάσματος όπως η Hamming Embedding [37]. Γεφυρώνοντας αυτές τις μεθόδους καταλήγουμε στον να προτείνουμε έναν πυρήνα ταιριάσματος ο οποίος κρατάει τα καλύτερα χαρακτηριστικά από κάθε μέθοδο με το να συνδυάσει μία διαδικασία συνάθροισης και έναν επιλεκτικό πυρήνα ταιριάσματος. Τελικά, η αναπαράσταση η οποία χρησιμοποιείται σε αυτόν τον πυρήνα προσεγγίζεται, δίνοντας έτσι την δυνατότητα να γίνει αναζήτηση σε μεγάλη κλίμακα και με καλή ακρίβεια.

Παρουσιάζουμε ένα πλαίσιο το οποίο ενσωματώνει τεχνικές ταιριάσματος όπως η μέθοδος BoW, HE και η πρόσφατη αναπαράσταση των VLAD. Για αυτό τον σκοπό παρουσιάζουμε μία κατηγορία πυρήνων ταιριάσματος οι οποίοι περιέχουν υπάρχουσες τεχνικές ταιριάσματος κι εκείνες που εμπεριέχουν συνάθροιση τοπικών περιγραφέων. Συζητάμε και αναλύουμε τις δύο βασικές διαφορές των διαφορετικών μεθόδων. Πρώτα, αναλύουμε την επιλεκτικότητα της συνάρτησης ταιριάσματος, δηλαδή την ιδιότητα πως μία αντιστοιχία συνεισφέρει στην οπτική ομοιότητα μόνο εάν η βεβαιότητα της είναι αρκετά μεγάλη. Δεύτερον, η συνάθροιση η οποία χρησιμοποιείται στις μεθόδους όπως η BoW, VLAD και τα διανύσματα Fisher [65][64], δεν χρησιμοποιείται σε μεθόδους ταιριάσματος όπως η HE. Δείχνουμε το όφελος από τον συνδυασμό των δύο, και συζητάμε την σχέση του με άλλες μεθόδους (π.χ. , [33, 63]) οι οποίες εισήχθησαν για να αντιμετωπίσουν το φαινόμενο ομοβροντίας (burstiness) [33].



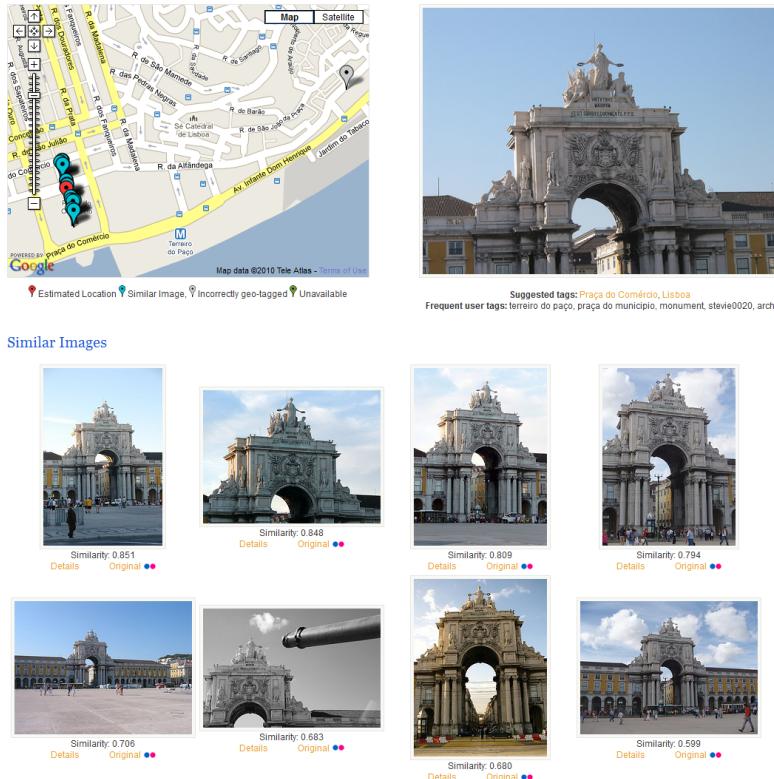
**Σχήμα 6.3:** Παραδείγματα από τοπικά χαρακτηριστικά τα οποία προβάλλονται στην ίδια οπτική λέξη και τελικά συναθροίζονται.

Οδηγούμαστε στο συμπέρασμα ότι κανένα από τα υπάρχοντα σχήματα δεν μας οδηγεί στην καλύτερη απόδοση και τελικά προτείνουμε μία νέα μέθοδο η οποία συνδυάζει τα πιο χρήσιμα συστατικά από τις υπάρχουσες τεχνικές. Συνδυάζει την τεχνική της συνάθροισης για χαρακτηριστικά τα οποία προβάλλονται στην ίδια οπτική λέξη (Σχήμα 6.3) μαζί με έναν επιλεκτικό πυρήνα ταιριάσματος.

### 6.3 VIRaL

Με χρήση γνωστών μεθόδων οι οποίες έχουν περιγραφεί ή προταθεί στην παρούσα διατριβή δημιουργήσαμε την μηχανή αναζήτησης, αυτόματου χαρακτηρισμού και γεωεντοπισμού εικόνων VIRal<sup>1</sup>(Σχήμα 6.4 [41]. Δεδομένης μίας εικόνας ερωτήματος εντοπίζονται εκείνες οι εικόνες της βάσης οι οποίες απεικονίζουν το ίδιο αντικείμενο με εκείνη. Οι εικόνες της βάσης συνοδεύονται από κειμενικό χαρακτηρισμό τον οποίο επεξεργαζόμαστε για να αναγνωρίσουμε τις πιο συχνά χρησιμοποιούμενες περιγραφές από τους χρήστες ανάμεσα στις όμοιες εικόνες. Με αυτό τον τρόπο προκύπτει αυτόματη πρόταση για κειμενική περιγραφή. Στην συνέχεια με ομαδοποίηση των γεωγραφικών συντεταγμένων των όμοιων εικόνων ανιχνεύονται μία προτεινόμενη τοποθεσία για την εικόνα ερωτήματος. Τέλος γίνεται μία διασταύρωση των αξιοθέατων τα οποία εντοπίζονται στην γειτονική περιοχή με βάση τις λίστες που παρέχονται από την Wikipedia, με αποτέλεσμα να

<sup>1</sup><http://viral.image.ntua.gr>



**Σχήμα 6.4:** Στιγμιότυπο από αναζήτηση με εικόνα ερωτήματος στην μηχανή αναζήτησης VIRaL.

παρέχεται μία επιτυχημένη και με βεβαιότητα ανίχνευση αξιοθέατου.

Για την αναπαράσταση των εικόνων έχουν χρησιμοποιηθεί τοπικά χαρακτηριστικά και περιγραφείς SURF [8]. Δημιουργούμε την αναπαράσταση της μεθόδου BoW με την χρήση ενός οπτικού λεξικού μεγέθους 100K οπτικών λέξεων. Η ανακατάταξη και γεωμετρική επιβεβαίωση γίνεται με την μέθοδο Fast Spatial Matching [66], αλλά υπάρχει και μία παραλλαγή του συστήματος<sup>2</sup> η οποία χρησιμοποιεί τον δικό μας αλγόριθμο Hough Pyramid Matching [78].

Η μέθοδος της ομαδοποίησης που χρησιμοποιήθηκε για τους χάρτες σκηνής έχει εφαρμοσθεί στην εφαρμογή VIRaL-explore<sup>3</sup> η οποία απεικονίζει όλη την συλλογή εικόνων πάνω στον παγκόσμιο χάρτη. Για εύκολη πλοιήγηση σε κάθε βαθμίδα ανάλυσης εντοπίζονται οι πιο σημαντικές ομάδες και αντιπροσωπευτικές εικόνες εξ αυτών απεικονίζονται πάνω στον χάρτη.

<sup>2</sup><http://viral-beta.image.ntua.gr>

<sup>3</sup><http://viral.image.ntua.gr/explore.php>



## Κεφάλαιο 7

### Συμπεράσματα

Σύμφωνα με τα όσα γνωρίζουμε, η μέθοδος των χαρτών χαρακτηριστικών ήταν από τις πρώτες οι οποίες ενσωμάτων την εμφάνιση των τοπικών χαρακτηριστικών καθώς και καθολική γεωμετρία στην δεικτοδότηση εικόνων, ενώ παραμένει αναλλοίωτη ως προς αφινικούς μετασχηματισμούς και εύρωστη σε μερική ορατότητα των αντικειμένων. Θεωρούμε τα πειράματα μας επιτυχημένα καθώς καταφέρνουμε να κάνουμε γεωμετρικό ταίριασμα σε μεγάλη κλίμακα.

Θεωρούμε ότι η απόδοση περιορίζεται κυρίως από την ίδια την υπόθεση η οποία καθιστά και την δεικτοδότηση γεωμετρίας εφικτή: ότι μία μοναδική αντιστοιχία είναι αρκετή για να πραγματοποιηθεί ευθυγράμμιση εικόνων. Βλέπουμε σαν πρόκληση για τους μελλοντικούς ανιχνευτές τοπικών χαρακτηριστικών την επίτευξη καλύτερης ευθυγράμμισης. Αναπτύξαμε την μεθοδολογία μας για αφινικές περιοχές, αλλά η επέκταση σε χρήση ομογραφίας είναι απλή, με την χρήση κατάλληλων χαρακτηριστικών.

Θεωρούμε την αναπαράσταση με τους χάρτες χαρακτηριστικών (Κεφάλαιο 2) σαν την πιο σημαντική συνεισφορά της μεθόδου FMS. Είναι εκείνη που ανάγει το γεωμετρικό ταίριασμα σε απλό εσωτερικό γινόμενο της νέας αναπαράστασης κι επιτρέπει την χρήση ενός απλού ανάστροφου αρχείου για δεικτοδότηση και γρήγορη αναζήτηση. Λόγω των υψηλών απαιτήσεων μνήμης η επιλογή τοπικών χαρακτηριστικών με βάση το γεωμετρικό ταίριασμα ανάμεσα σε διαφορετικές όψεις του ίδιο αντικειμένου κρίθηκε απαραίτητη. Επιτρέπει την δεικτοδότηση μέχρι 1M εικόνων με τέτοιο τρόπο ώστε η απόδοση να διατηρείται σε υψηλά επίπεδα καθώς γίνεται επιλογή εκείνων την χαρακτηριστικών που έχουν την μεγαλύτερη πιθανότητα να ταιριάζουν με μία νέα εικόνα ερωτήματος.

Όσον αφορά την αναπαράσταση την χαρτών χαρακτηριστικών μελλοντικές κατευθύνσεις μπορεί να είναι η εφαρμογή τους σε προβλήματα αναγνώρισης κι εντοπισμού αντικειμένων, καθώς κι εκεί η γεωμετρική συμφωνία των τοπικών χαρακτηριστικών είναι σημαντική όπως και στην ανάκτηση εικόνων.

Επίσης στην μέθοδο HPM (Κεφάλαιο 3) αναπτύξαμε έναν απλό αλγόριθμο χωρικού

ταιριάσματος ο οποίος μπορεί με απλό τρόπο να ενσωματωθεί σε οποιαδήποτε μηχανή αναζήτησης εικόνων. Βελτιώνει την απόδοση με το να επιτρέπει εύκαμπτο ταίριασμα καθώς και ταίριασμα πολλαπλών αντικειμένων κι επιφανειών. Ένας ακόμα λόγος για την βελτίωση της απόδοσης φαίνεται να είναι απελευθέρωση από την χρήση ενός αυστηρού γεωμετρικού μετασχηματισμού, με το οποίο γίνεται δυαδική απόφαση για τον αν μία αντιστοιχία είναι inlier ή όχι. Στην περίπτωση μας κάθε αντιστοιχία συμμετέχει με τον κατάλληλο βαθμό στο συνολικό μέτρο ομοιότητας. Οι πραγματικοί inliers θα συμμετέχουν με μεγάλο βαθμό.

Εκτός από την υψηλότερη απόδοση επίτευγμα για τον αλγόριθμο HPM είναι και το ταχύτερο ταίριασμα. Με την απαλλαγή από την ανάγκη καταμέτρησης inliers σύμφωνα με κάποιες υποθέσεις γεωμετρικών μοντέλων, το HPM αποφέρει σημαντική επιτάχυνση σε σύγκριση με μεθόδους που βασίζονται στην ιδέα του RANSAC. Επιπλέον, η μη ανάγκη για εκτίμηση των παραμέτρων ενός γεωμετρικού μετασχηματισμού προσφέρει υπολογιστικό κέρδος. Η απόδοση της μεθόδου μας φθάνει στον κορεσμό μέσα σε 3 δευτερόλεπτα, το οποίο είναι και λογικός χρόνος εκτέλεσης ενός ερωτήματος.

Μία πολύ ενδιαφέρουσα ερώτηση είναι κατά πόσο μπορούμε ακόμα να επωφεληθούμε από την δεικτοδότηση γεωμετρίας. Πειράματα σε κλίμακα 50K εικόνων έχουν δώσει κάποιες ενδείξεις [6]. Φαίνεται όμως ότι ένα τελικό στάδιο γεωμετρικού ταιριάσματος στις εικόνες οι οποίες έχουν καταταχθεί πρώτες είναι απαραίτητο, και το HPM αποτελεί ένα καλό εργαλείο σε αυτή την περίπτωση. Από την άλλη, σε συνδυασμό με την επιλογή χαρακτηριστικών με γεωμετρικά κριτήρια φάνηκε πως η ανακατάταξη δεν έχει κάτι επιπλέον να προσφέρει [81]. Είναι σίγουρο ότι αυτό αξίζει να διερευνηθεί σε ακόμα μεγαλύτερη κλίμακα (10M ή 100M εικόνων) καθώς οι συμπεριφορές σε τέτοιες περιπτώσεις δεν καθίστανται εύκολο να προβλεφθούν. Η ανακατάταξη εικόνων όμως παραμένει ένας εύκολος και αποδοτικός τρόπος για εφαρμογή γεωμετρίας με τρόπο ο οποίος δεν έχει γραμμική πολυπλοκότητα ως προς το μέγεθος της βάσης, αντίθετα με την δεικτοδότηση.

Η χρήση γεωμετρίας με παρόμοιο τρόπο και με μικρή παραλλαγή της μεθόδου HPM μας επέτρεψε να αναπτύξουμε μία μέθοδο για επιλογή χαρακτηριστικών η οποία βασίζεται στο πολύ γενικό κριτήριο της συμμετρίας ή της αυτο-ομοιότητας (Κεφάλαιο 4). Κατά τα όσα γνωρίζουμε είμαστε οι πρώτοι που εφαρμόζουμε τον εντοπισμό συμμετριών για την επιλογή χαρακτηριστικών. Επίσης δείχνουμε πως ο εντοπισμός συμμετριών και επαναλαμβανόμενων δομών δεν είναι και πολύ διαφορετικός από το γενικό γεωμετρικό ταίριασμα μεταξύ δύο εικόνων. Αυτή η μέθοδος σε συνδυασμό με την μέθοδο που εφαρμόσαμε στους χάρτες χαρακτηριστικών για την επιλογή περιοχών δίνουν μία πλήρη λύση ως προς την επιλογή περιοχών σε ένα πραγματικό σύστημα αναζήτησης εικόνων. Η μία παρέχει τρόπους για επιλογή από μοναδικές εικόνες και η άλλη από εκείνες οι οποίες κατά ομάδες απεικονίζουν το ίδιο αντικείμενο. Φυσικά στις τελευταίες μπορεί επιπλέον να εφαρμοσθεί

η επιλογή με βάση τις συμμετρίες.

Εκτός από την χρήση γεωμετρίας, υπάρχουν πολλοί άλλοι τρόποι με τους οποίους μπορεί να βελτιωθεί η απόδοση της ανάκτησης εικόνων, όπως με την *επέκταση ερωτήματος* (query expansion) [18] και την μέθοδο HQE (Κεφάλαιο 5). Αυτό έγινε εφικτό για πρώτη φορά αικόμα και χωρίς την χρήση γεωμετρίας. Επίσης για πρώτη φορά εμπλουτίστηκε η τοπική πληροφορία κι όχι η ολική, ιδιότητα στην οποία προσδίδεται η υψηλή απόδοση της τεχνικής. Ο συνδυασμός της μεθόδου με τεχνικές συνάθροισης περιγραφέων μείωσε δραστικά την πολυπλοκότητα του ερωτήματος, κρατώντας την σε κοντινά επίπεδα με εκείνα του αρχικού ερωτήματος. Σαν μελλοντική επέκταση μπορεί να δοκιμασθεί η χρήση πιο χαλαρών γεωμετρικών μοντέλων [78, 37], όπως το HPM, σαν συμβιβασμός ανάμεσα στις κοστοβόρες μεθόδους γεωμετρικού ταιριάσματος και στην μη χρήση γεωμετρίας. Η μη χρήση γεωμετρίας προσφέρει εξαιρετική επιτάχυνση στον χρόνο εκτέλεσης, ενώ τελικά με τον συνδυασμό με γεωμετρική πληροφορία η απόδοση βελτιώνεται αικόμα περισσότερο.



# Βιβλιογραφία

- [1] Sameer Agarwal, Noah Snavely, Ian Simon, Steven M. Seitz, and Richard Szeliski. Building Rome in a day. In *ICCV*, 2009.
- [2] R. Arandjelovic and A. Zisserman. Three things everyone should know to improve object retrieval. In *IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2012.
- [3] Y. Avrithis and K. Rapantzikos. The medial feature detector: Stable regions from image boundaries. In *International Conference on Computer Vision (ICCV 2011)*, Barcelona, Spain, November 2011.
- [4] Y. Avrithis and G. Tolias. Hough pyramid matching: speeded-up geometric re-ranking in large scale image retrieval. *International Journal of Computer Vision (IJCV) – submitted*, 2012.
- [5] Yannis Avrithis, Yannis Kalantidis, Giorgos Tolias, and Evangelos Spyrou. Retrieving landmark and non-landmark images from community photo collections. In *ACM Multimedia*, Firenze, Italy, October 2010.
- [6] Yannis Avrithis, Giorgos Tolias, and Yannis Kalantidis. Feature map hashing: Sub-linear indexing of appearance and global geometry. In *ACM Multimedia*, Firenze, Italy, October 2010.
- [7] D.H Ballard. Generalizing the hough transform to detect arbitrary shapes. *Pattern Recognition*, 1981.
- [8] H. Bay, T. Tuytelaars, and L. Van Gool. SURF: Speeded up robust features. In *ECCV*, 2006.
- [9] S. Belongie, J. Malik, and J. Puzicha. Shape context: A new descriptor for shape matching and object recognition. In *NIPS*, volume 12, pages 831–827, 2000.
- [10] A.Z Broder. Identifying and filtering near-duplicate documents. In *Symposium on Combinatorial Pattern Matching*, 2000.

- [11] Y. Cao, C. Wang, Z. Li, L. Zhang, and L. Zhang. Spatial-bag-of-features. In *CVPR*, pages 3352–3359, 2010.
- [12] G. Carneiro and A.D Jepson. Flexible spatial configuration of local image features. *PAMI*, pages 2089–2104, 2007.
- [13] M.S Charikar. Similarity estimation techniques from rounding algorithms. In *ACM Symposium on Theory of Computing*, pages 380–388, 2002.
- [14] O. Chum and J. Matas. Geometric hashing with local affine frames. In *CVPR*, volume 1, pages 879–884, 2006.
- [15] O. Chum, J. Matas, and J. Kittler. Locally optimized RANSAC. In *DAGM*, page 236. Springer Verlag, 2003.
- [16] O. Chum, A. Mikulik, M. Perdoch, and J. Matas. Total recall II: Query expansion revisited. In *IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2011.
- [17] O. Chum, M. Perdoch, and J. Matas. Geometric min-hashing: Finding a (thick) needle in a haystack. In *CVPR*, 2009.
- [18] 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 *ICCV*, 2007.
- [19] 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 *International Conference on Computer Vision*, Oct. 2007.
- [20] A.C Cohen. Maximum likelihood estimation in the weibull distribution based on complete and on censored samples. *Technometrics*, 7(4):579–588, 1965.
- [21] H. Cornelius, M. Perdoch, J. Matas, and G. Loy. Efficient symmetry detection using local affine frames. In *ECIA*, 2007.
- [22] Q. Danfeng, S. Gammeter, L. Bossard, T. Quack, and L. Van Gool. Hello neighbor: Accurate object retrieval with k-reciprocal nearest neighbors. In *IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2011.
- [23] Wei Dong, Zhe Wang, Moses Charikar, and Kai Li. Efficiently matching sets of features with random histograms. In *ACM MM*, pages 179–188, 2008.
- [24] I. L. Dryden and Kanti V. Mardia. *Statistical Shape Analysis*. WileyBlackwell, July 1998.

- [25] O. Enqvist, K. Josephson, and F. Kahl. Optimal correspondences from pairwise constraints. In *ICCV*, 2009.
- [26] M.A. Fischler and R.C Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.
- [27] Stephan Gammeter, Lukas Bossard, Till Quack, and Luc V. Gool. I know what you did last summer: Object-level auto-annotation of holiday snaps. In *ICCV*, 2009.
- [28] E. Gavves, C. G. M. Snoek, and A. W. M. Smeulders. Visual synonyms for landmark image retrieval. *CVIU*, 2012.
- [29] K. Grauman and T. Darrell. Pyramid match hashing: Sub-linear time indexing over partial correspondences. In *CVPR*, 2007.
- [30] K. Grauman and T. Darrell. The pyramid match kernel: Efficient learning with sets of features. *Journal of Machine Learning Research*, 8:725–760, 2007.
- [31] M. Jain, R. Benmokhtar, P. Gros, and H. Jégou. Hamming embedding similarity-based image classification. In *ICMR*, Jun. 2012.
- [32] H. Jegou, M. Douze, and C. Schmid. Hamming embedding and weak geometric consistency for large scale image search. In *ECCV*, 2008.
- [33] H. Jegou, M. Douze, and C. Schmid. On the burstiness of visual elements. In *IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2009.
- [34] H. Jegou, M. Douze, and C. Schmid. Improving bag-of-features for large scale image search. *IJCV*, 87(3):316–336, 2010.
- [35] H. Jegou, M. Douze, C. Schmid, and P. Perez. Aggregating local descriptors into a compact image representation. In *CVPR*, 2010.
- [36] Hervé Jegou and Ondrej Chum. Negative evidences and co-occurrences in image retrieval: The benefit of PCA and whitening. In *European Conference on Computer Vision*, Oct. 2012.
- [37] Hervé Jegou, Matthijs Douze, and Cordelia Schmid. Improving bag-of-features for large scale image search. *International Journal of Computer Vision*, 87(3):316–336, Feb. 2010.

- [38] Hervé Jégou, Florent Perronnin, Matthijs Douze, Jorge Sánchez, Patrick Pérez, and Cordelia Schmid. Aggregating local descriptors into compact codes. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Sep. 2012.
- [39] Hervé Jégou, Cordelia Schmid, Hedi Harzallah, and Jakob Verbeek. Accurate image search using the contextual dissimilarity measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(1):2–11, Jan. 2010.
- [40] A. Joly and O. Buisson. Logo retrieval with a contrario visual query expansion. In *ACM International conference on Multimedia*, Oct. 2009.
- [41] Y. Kalantidis, G. Tolias, Y. Avrithis, M. Phinikettos, E. Spyrou, P. Mylonas, and S. Kollias. Viral: Visual image retrieval and localization. *Multimedia Tools and Applications*, January 2011.
- [42] Y. Kalantidis, G. Tolias, E. Spyrou, Ph. Mylonas, and Y. Avrithis. Visual image retrieval and localization. In *in Proceedings of 7th International Workshop on Content-Based Multimedia Indexing (CBMI 2009)*, Chania, Greece, June 2009.
- [43] Y. Kalantidis, G. Tolias, E. Spyrou, Ph. Mylonas, Y. Avrithis, and S. Kollias. Visual image retrieval and localization. In *in Proceedings of 3ο Πανελλήνιο Συνέδριο Φοιτητών Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών*, Thessaloniki, Greece, April 2009.
- [44] Y. Keller and Y. Shkolnisky. An algebraic approach to symmetry detection. *ICPR*, pages 186–189, 2004.
- [45] Jan Knopp, Josef Sivic, and Tomas Pajdla. Avoiding confusing features in place recognition. In *ECCV*, 2010.
- [46] Kevin Köser, Christian Beder, and Reinhard Koch. Conjugate rotation: Parameterization and estimation from an affine feature correspondence. In *CVPR*, 2008.
- [47] Y. Lamdan and HJ Wolfson. Geometric hashing: A general and efficient model-based recognition scheme. In *ICCV*, pages 238–249, 1988.
- [48] S. Lazebnik, C. Schmid, and J. Ponce. Semi-local affine parts for object recognition. In *BMVC*, 2004.
- [49] S. Lazebnik, C. Schmid, and J. Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *CVPR*, volume 2, page 1, 2006.

- [50] M. Leordeanu and M. Hebert. A spectral technique for correspondence problems using pairwise constraints. In *ICCV*, volume 2, pages 1482–1489, 2005.
- [51] F. Li and J. Kosecka. Probabilistic location recognition using reduced feature set. In *ICRA*, 2006.
- [52] D. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, Nov. 2004.
- [53] D.G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004.
- [54] Gareth Loy and Jan-Olof Eklundh. Detecting symmetry and symmetric constellations of features. In *ECCV*, 2006.
- [55] C. D. Manning, P. Raghavan, and H. Schütze. *Introduction to Information Retrieval*, chapter Relevance feedback & query expansion. Cambridge University Press, 2008.
- [56] K. Mikolajczyk and C. schmid. Scale and affine invariant interest point detectors. *International Journal of Computer Vision*, 60(1):63–86, Oct. 2004.
- [57] Andrej Mikulik, Michal Perdoch, Ondrej Chum, and Jiri Matas. Learning a fine vocabulary. In *ECCV*. Springer, 2010.
- [58] M. Muja and D.G Lowe. Fast approximate nearest neighbors with automatic algorithm configuration. In *ICCV*, 2009.
- [59] N. Naikal, A.Y. Yang, and S. Shankar Sastry. Informative feature selection for object recognition via sparse pca. In *ICCV*, 2011.
- [60] D. Nistér and H. Stewénius. Scalable recognition with a vocabulary tree. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2161–2168, Jun. 2006.
- [61] Michal Perdoch, Ondrej Chum, and Jiri Matas. Efficient representation of local geometry for large scale object retrieval. In *CVPR*, 2009.
- [62] Michal Perdoch, Ondrej Chum, and Jiri Matas. Efficient representation of local geometry for large scale object retrieval. In *IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2009.
- [63] F. Perronnin, J.Sánchez, and T. Mensink. Improving the Fisher kernel for large-scale image classification. In *European Conference on Computer Vision*, Sep. 2010.

- [64] F. Perronnin, Y. Liu, J. Sanchez, and H. Poirier. Large-scale image retrieval with compressed Fisher vectors. In *IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2010.
- [65] Florent Perronnin and Christopher R. Dance. Fisher kernels on visual vocabularies for image categorization. In *IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2007.
- [66] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman. Object retrieval with large vocabularies and fast spatial matching. In *CVPR*, 2007.
- [67] 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*, Jun. 2007.
- [68] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman. Lost in quantization: Improving particular object retrieval in large scale image databases. In *IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2008.
- [69] James Philbin, Ondrej Chum, Josef Sivic, Michael Isard, and Andrew Zisserman. Lost in quantization: Improving particular object retrieval in large scale image databases. In *CVPR*, 2008.
- [70] F. Rothganger, S. Lazebnik, C. Schmid, and J. Ponce. 3D object modeling and recognition using local affine-invariant image descriptors and multi-view spatial constraints. *IJCV*, 66(3):231–259, 2006.
- [71] G. Schindler, M. Brown, and R. Szeliski. City-scale location recognition. In *CVPR*, 2007.
- [72] X. Shen, Z. Lin, J. Brandt, S. Avidan, and Y. Wu. Object retrieval and localization with spatially-constrained similarity measure and k-nn re-ranking. In *CVPR*. IEEE, 2012.
- [73] J. Sivic and A. Zisserman. Video Google: A text retrieval approach to object matching in videos. In *ICCV*, pages 1470–1477, 2003.
- [74] J. Sivic and A. Zisserman. Video Google: A text retrieval approach to object matching in videos. In *International Conference on Computer Vision*, Oct. 2003.
- [75] E. Spyrou, Y. Kalantidis, G. Tolias, Ph. Mylonas, and S. Kollias. Intelligent content retrieval using a visual vocabulary and geometric constraints. In *in Proceedings of FUZZ-IEEE 2010, in conjunction with The 2010 IEEE World Congress on Computational Intelligence (IEEE WCCI 2010)*, July 2010.

- [76] C. Sun and D. Si. Fast reflectional symmetry detection using orientation histograms. *Real Time Imaging*, 5(1):63–74, 1999.
- [77] M. Tipping and B. Schölkopf. A kernel approach for vector quantization with guaranteed distortion bounds. In *Artificial Intelligence and Statistics*, pages 129–134, 2001.
- [78] G. Tolias and Y. Avrithis. Speeded-up, relaxed spatial matching. In *International Conference on Computer Vision (ICCV 2011)*, Barcelona, Spain, November 2011.
- [79] G. Tolias, Y. Avrithis, and H. Jégou. To aggregate or not to aggregate: Selective match kernels for image search. In *ICCV*, 2013.
- [80] G. Tolias, Y. Kalantidis, and Y. Avrithis. Symcity: Feature selection by symmetry for large scale image retrieval. In *ACM Multimedia*, Nara, Japan, October 2012. ACM.
- [81] G. Tolias, Y. Kalantidis, and Y. Avrithis. Towards large scale geometry indexing. *Computer Vision and Image Understanding (CVIU) – submitted*, 2012.
- [82] P. Turcot and D.G Lowe. Better matching with fewer features: the selection of useful features in large database recognition problems. In *International Conference on Computer Vision Workshop*, Oct. 2009.
- [83] T. Tuytelaars, A. Turina, and L. Van Gool. Noncombinatorial detection of regular repetitions under perspective skew. *PAMI*, 2003.
- [84] C. Varytimidis, K. Rapantzikos, and Y. Avrithis. Wash: Weighted a-shapes for local feature detection. In *European Conference on Computer Vision (ECCV 2012)*, Florence, Italy, October 2012.
- [85] A. Vedaldi and S. Soatto. Quick shift and kernel methods for mode seeking. In *ECCV*, 2008.
- [86] A. Vedaldi and S. Soatto. Relaxed matching kernels for robust image comparison. In *CVPR*, 2008.
- [87] Y. Zhang, Z. Jia, and T. Chen. Image retrieval with geometry-preserving visual phrases. In *CVPR*, pages 809–816. IEEE, 2011.



# Κατάλογος Δημοσιεύσεων

G. Tolias, and Y. Avrithis and H. Jégou, To aggregate or not to aggregate: Selective match kernels for image search, In *International Conference on Computer Vision (ICCV 2013)*, 2013.

G. Tolias, and H. Jégou Local visual query expansion: Exploiting an image collection to refine local descriptors, *INRIA Technical Report, July, 2013*

Y. Avrithis, and G. Tolias, Hough Pyramid Matching: Speeded-up geometry reranking for large scale image retrieval, *International Journal of Computer Vision (IJCV). accepted.*

G. Tolias, Y. Kalantidis, and Y. Avrithis, Towards large scale geometry indexing, *Computer Vision and Image Understanding (CVIU). submitted.*

Y. Kalantidis, G. Tolias, Y. Avrithis, M. Phinikettos, E. Spyrou, P. Mylonas, and S. Kollias, Viral: Visual image retrieval and localization, *Multimedia Tools and Applications (MTAP)*, 2011.

G. Tolias, Y. Kalantidis, and Y. Avrithis, SymCity: Feature Selection by Symmetry for Large Scale Image Retrieval, In *ACM Multimedia (ACM MM 2012), Nara, Japan, October 2012.*

G. Tolias and Y. Avrithis, Speeded-up, relaxed spatial matching, In *International Conference on Computer Vision (ICCV 2011)*, 2011.

Y. Avrithis, Y. Kalantidis, G. Tolias, and E. Spyrou, Retrieving landmark and non-landmark images from community photo collections, In *ACM Multimedia (ACM MM 2010), Firenze, Italy, October 2010.*

Y. Avrithis, G. Tolias, and Y. Kalantidis, Feature map hashing: Sub-linear indexing of appearance and global geometry, In *ACM Multimedia (ACM MM 2010), Firenze, Italy, October 2010.*

2010.

S. Papadopoulos, C. Zigkolis, G. Tolias, Y. Kalantidis, P. Mylonas, Y. Kompatsiaris, and A. Vakali, Image clustering through community detection on hybrid image similarity graphs, In *International Conference on Image Processing (ICIP 2010), Hong Kong, September 2010*.

E. Spyrou, Y. Kalantidis, G. Tolias, Ph. Mylonas, and S. Kollias, Intelligent content retrieval using a visual vocabulary and geometric constraints, In *FUZZ-IEEE 2010, in conjunction with The 2010 IEEE World Congress on Computational Intelligence (WCCI 2010), July 2010*.

Y. Kalantidis, G. Tolias, E. Spyrou, Ph. Mylonas, and Y. Avrithis, Visual image retrieval and localization, In *International Workshop on Content-Based Multimedia Indexing (CBMI 2009), Chania, Greece, June 009*.

Y. Kalantidis and G. Tolias and E. Spyrou and Ph. Mylonas and Y. Avrithis and S. Kollias, Visual Image Retrieval and Localization, In *Proceedings of 3ο Πανελλήνιο Συνέδριο Φοιτητών Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών, April, Thessaloniki, Greece, 2009*.

□