Εθνικο Μετσοβιο Πολυτεχνειο
Σχολη Ηλεκτρολογων Μηχανικων και Μηχανικων Υπολογιστων
Τομεας Τεχνολογιας Πληροφορικης και Υπολογιστων
Εργαστηριο Επεξεργασιας Εικονας Βιντεο και Πολυμεσων

Κατηγοριοποίηση Εικόνων Με Τεχνικές Χωρικού Ταιριάσματος Και Δεικτοδότησης

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

της

Αγνής Δ. Δελβινιώτη

Επιβλέπων: Στέφανος Κόλλιας
Καθηγητής Ε.Μ.Π.

Αθήνα, Οκτώβριος 2012

Εθνικο Μετσοβιο Πολυτεχνειο
Σχολη Ηλεκτρολογων Μηχανικων και Μηχανικων Υπολογιστων
Τομεας Τεχνολογιας Πληροφορικης και Υπολογιστων
Εργαστηριο Επεξεργασιας Εικονας Βιντεο και Πολυμεσων

Κατηγοριοποίηση Εικόνων Με Τεχνικές Χωρικού Ταιριάσματος Και Δεικτοδότησης

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

της

Αγνής Δ. Δελβινιώτη

Επιβλέπων: Στέφανος Κόλλιας
Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 2η Οκτωβρίου 2012.

(Υπογραφή)	*(Υπογραφή)*	*(Υπογραφή)*
.....	.....	.....
Στέφανος Κόλλιας	Ανδρέας-Γεώργιος Σταφυλοπάτης	Γεώργιος Στάμου
Καθηγητής Ε.Μ.Π.	Καθηγητής Ε.Μ.Π.	Επίκουρος Καθηγητής Ε.Μ.Π.

Αθήνα, Οκτώβριος 2012

.....

Αγνή Δ. Δελβινιώτη
Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

Ευχαριστίες

Η παρούσα διπλωματική εργασία εκπονήθηκε το ακαδημαϊκό έτος 2011-2012 στο εργαστήριο Ψηφιακής Επεξεργασίας Εικόνας, Βίντεο και Πολυμέσων του Εθνικού Μετσόβιου Πολυτεχνείου. Θα ήθελα να ευχαριστήσω τον καθηγητή Στέφανο Κόλλια για την εμπιστοσύνη που μου έδειξε με την ανάθεση αυτής της εργασίας, δίνοντάς μου παράλληλα τη δυνατότητα να έρθω σε επαφή με έναν χώρο που προάγει την επιστήμη και την έρευνα σε κλίμα συνεργασίας. Επίσης, αισθάνομαι τυχερή και ευγνώμων απέναντι στο Δρ Ιωάννη Αβρίθη που με ενθάρρυνε από την πρώτη στιγμή στην ανάληψη ενός θέματος πρωτότυπου με καθαρά ερευνητικό χαρακτήρα και μου ενέπνευσε τη δική του χαρά και αφοσίωση στη μελέτη και την έρευνα. Ευχαριστώ ακόμη, τον υποψήφιο Δρ Γεώργιο Τόλια για την καθοδήγηση και υποστήριξη καθ' όλη τη διάρκεια της διπλωματικής. Τέλος, θα ήθελα να ευχαριστήσω τους γονείς μου για την αμέριστη συμπαράστασή τους κάθε στιγμή, όλα αυτά τα χρόνια.

Περίληψη

Στο πλαίσιο αυτής της Διπλωματικής εργασίας, εισάγουμε μια νέα μέθοδο κατηγοριοποίησης εικόνων, η οποία ενσωματώνει το χωρικό ταίριασμα και τη δεικτοδότηση στη διαδικασία ταξινόμησης. Το χωρικό ταίριασμα βασίζεται στο *ταίριασμα πυραμίδας Hough (Hough pyramid matching* (HPM), η δεικτοδότηση βασίζεται στη δομή του ανεστραμμένου αρχείου όπως στην ανάκτηση εικόνων και η ταξινόμηση πραγματοποιείται με μια *μηχανή διανυσμάτων υποστήριξης (support vector machine* (SVM) ως ταξινομητή πολλών κλάσεων.

Χρησιμοποιούμε την τεχνική HPM ως μέτρο ομοιότητας και κάνοντας λογικές υποθέσεις δείχνουμε ότι αποτελεί πυρήνας Mercer. Στην κατεύθυνση αυτή, το εκφράζουμε σαν ένα εσωτερικό γινόμενο σε έναν χώρο πολλών διαστάσεων, όπου οι εικόνες διαθέτουν μια κβαντισμένη αναπαράσταση των τοπικών χαρακτηριστικών τους και των περιγραφέων τους. Στη συνέχεια χρησιμοποιούμε αυτόν τον πυρήνα στην εκπαίδευση με SVM αντί για κάποιο γραμμικό πυρήνα, ο οποίος είναι η τυπική επιλογή με βάση το μοντέλο *"σάκος οπτικών λέξεων" (bag of words)* (BoW). είναι η πρώτη φορά που μια συνάρτηση πυρήνας λαμβάνει υπόψιν τη χωρική διάταξη, διατηρώντας το αναλλοίωτο ως προς τη μετατόπιση, την κλίμακα και την περιστροφή. Στις περισσότερες περιπτώσεις, τεχνητές μεταβολές είναι ο μόνος τρόπος να επιτευχθεί το γεωμετρικό αναλλοίωτο, με μια εκθετική αύξηση του χρόνου εκπαίδευσης.

Εκπαιδεύουμε ένα δυαδικό SVM ταξινομητή για κάθε κατηγορία ακολουθώντας την προσέγγιση "ένα έναντι των υπολοίπων" (one-versus-the-rest) και στη συνέχεια συνδυάζουμε τους μεμονωμένους ταξινομητές σε έναν ταξινομητή πολλών κλάσεων. Συγκριτικά με τον ταξινομητή του "πιο κοντινού γείτονα" (nearest neighbor) που χρησιμοποιούν για παράδειγμα οι μέθοδοι ανάκτησης εικόνων, εκμεταλλευόμαστε την αραιή αναπαράσταση των SVM: σε χρόνο ταξινόμησης, η εικόνα αναζήτησης ταιριάζεται με HPM με βάση μόνο τα διανύσματα υποστήριξης. Παρόλα αυτά, το ταίριασμα δε χρειάζεται να είναι εξαντλητικό. Τα διανύσματα υποστήριξης δεικτοδοτούνται με βάση ένα ανεστραμμένο αρχείο, και ο HPM εφαρμόζεται μόνο σε ένα μικρό υποσύνολο, το οποίο έχει την υψηλότερη κατάταξη με βάση κάποιο βαθμωτό μέτρο, όπως για παράδειγμα με βάση το BoW. Η μέθοδος επομένως εφαρμόζεται εύκολα σε ταξινόμηση μεγάλης κλίμακας, ενώ η εκπαίδευση νέων κλάσεων δεν απαιτεί επανεκπαίδευση των ήδη υπαρχόντων.

Λόγω της φύσης των τοπικών χαρακτηριστικών και της χρήσης τους σε ταίριασμα που διατηρεί το αναλλοίωτο, η μέθοδος είναι η πιο κατάλληλη για αναγνώριση συγκεκριμένων αντικειμένων. Εμείς την εφαρμόζουμε σε αναγνώριση αξιοθέατων, διεξάγοντας πειράματα σε δικό μας σύνολο δεδομένων, το οποίο έχει κατασκευαστεί από το σύνολο δεδομένων World cities μέσω μιας ημιαυτόνομης διαδικασίας, η οποία συνδυάζει οπτική και γεωγραφική συσταδοποίηση. Συγκρίνουμε με έναν ταξινομητή αναφοράς (baseline), ο οποίος χρησιμοποιεί BoW και πετυχαίνουμε περισσότερο από διπλάσια αύξηση σε ακρίβεια για πειράματα μέχρι

και 68 αξιοθέατων.

Λέξεις κλειδιά

Κατηγοριοποίηση εικόνων, εκμάθηση με χρήση πυρήνων, χωρικό ταίριασμα, δεικτοδότηση, ανάκτηση εικόνων, αναγνώριση αξιοθέατων

Abstract

In the framework of this Diploma thesis we introduce a new image categorization method, which integrates spatial matching and indexing in the classification process. Spatial matching is based on *Hough pyramid matching* (HPM); indexing is based on an inverted file structure as in image retrieval; and classification is carried out with a multiclass *support vector machine* (SVM) classifier.

We use HPM as an image similarity measure and we show that under reasonable assumptions it is a Mercer kernel. We do so by explicitly expressing it as an inner product in a high dimensional space where images lie given an appropriate quantized representation of their local features and descriptors. We then use this kernel for SVM training instead of a linear kernel, which is a typical choice under the *bag of words* (BoW) model. It is the first time that a kernel function takes spatial configuration into account while being invariant to translation, scale and rotation. In most cases, artificial perturbations are the only way to achieve geometric invariance, with an exponential increase of training time.

We train one binary SVM classifier for each category following an one-versus-the -rest strategy and then combine individual classifiers into one multiclass classifier. Comparing to nearest-neighbor classifier using e.g. image retrieval methods, we exploit the sparse representation of SVMs: at classification time, the query image is matched via HPM against the chosen support vectors only. However, matching need not be exhaustive. Support vectors are indexed into an inverted file, and HPM may be applied only to a small subset that is top-ranking according to any scalar similarity measure, e.g. based on BoW. The method therefore easily applies to large scale classification, while training for unseen classes does not require re-training for existing ones.

Due to the nature of local features and their use in invariant matching, the method is most appropriate for specific object recognition. We apply it to landmark recognition, conducting experiments on our own dataset, constructed from the World cities dataset via a semi-automatic process that combines visual and geographical clustering. We compare to a baseline classifier using a BoW representation and achieve more than a twofold increase in accuracy on experiments of up to 68 landmarks.

Keywords

Image categorization, kernel learning, spatial matching, indexing, image retrieval, landmark recognition

Περιεχόμενα

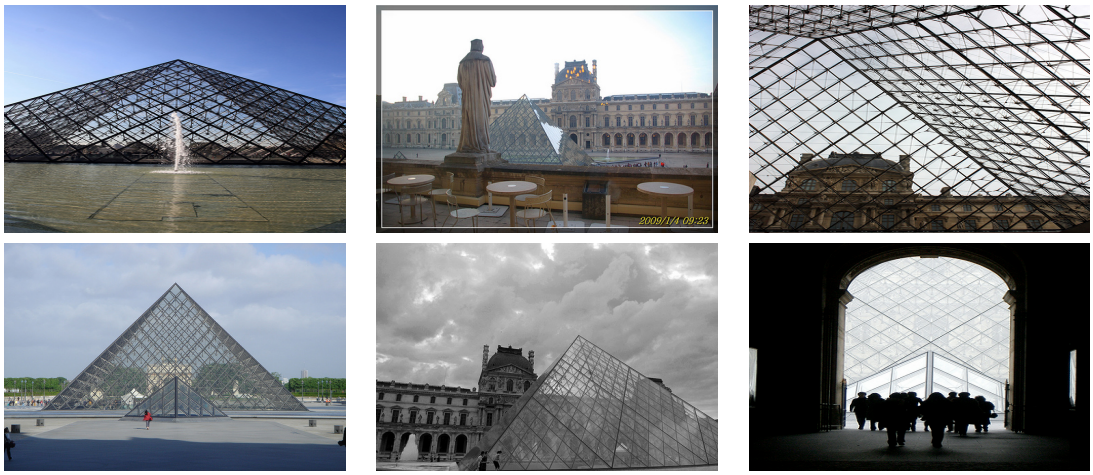
Κατάλογος σχημάτων

Κεφάλαιο 1

Εισαγωγή

1.1 Περιγραφή προβλήματος

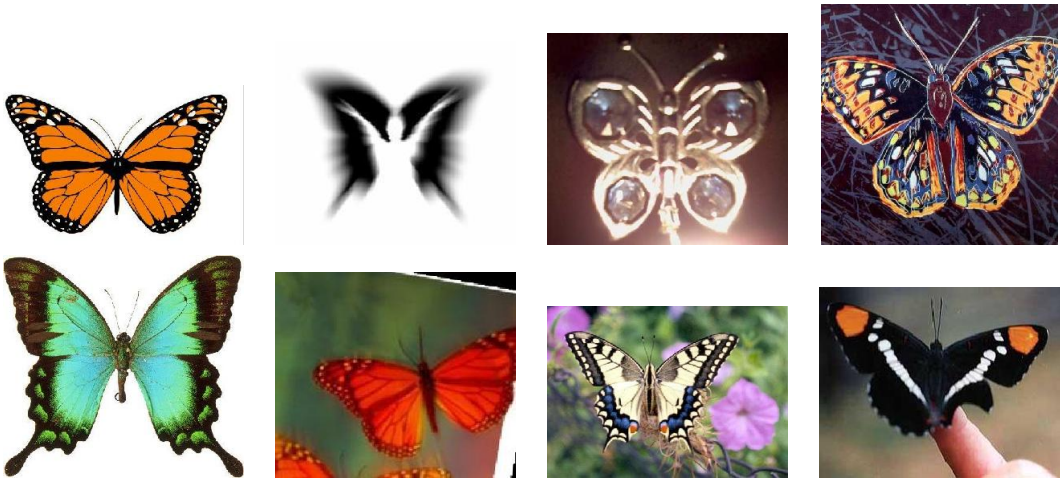
Πληθώρα εφαρμογών, σε διαφόρους επιστημονικούς κλάδους στη σημερινή πραγματικότητα αναδεικνύουν την αναγκαιότητα για μηχανές αναγνώρισης οπτικού περιεχομένου σε εικόνες. Η ευρεία εξάπλωση των έξυπνων τηλεφώνων, καθώς και η ραγδαία ανάπτυξη εφαρμογών διαχείρισης και αναζήτησης εικόνων, και κυρίως ιστοσελίδων κοινοποίησης εικόνων στον παγκόσμιο ιστό, όπως το Facebook και το Flickr, συντελούν στη δημιουργία ολοένα περισσότερων και ποικιλόμορφων συλλογών ψηφιακών εικόνων. Η ανάγκη διαχείρισης των συλλογών αυτών, ωθεί στην αναζήτηση τεχνικών για την αξιοποίηση της σημαντικής σχετικά με το περιεχόμενο των εικόνων πληροφορίας.



Σχήμα 1.1: Δείγμα εικόνων του αξιοθέατου Louvre Pyramid από το δικό μας σύνολο δεδομένων, οι οποίες χαρακτηρίζονται από διαφορετική οπτική γωνία λήψης, κλίμακα, φωτισμό, ύπαρξη οπτικών εμποδίων και αταξία περιβάλλοντος .

Κλασσικό πρόβλημα στην όραση υπολογιστών, στην επεξεργασία εικόνας και τη μηχανική όραση αποτελεί η ανάγκη να αποφασίσουμε αν μια εικόνα περιλαμβάνει ένα συγκεκριμένο αντικείμενο, χαρακτηριστικό ή δραστηριότητα και περιγράφεται με τον όρο *αναγνώριση*

(recognition). Μια διαδικασία αυθόρμητη και εύκολη για τον άνθρωπο, μοιάζει πολύπλοκη για μια μηχανή, ιδιαίτερα στη γενική περίπτωση όπου αναζητούμε αυθαίρετα αντικείμενα, 1) σε αυθαίρετες θέσεις σε σχέση με το φωτογραφικό φακό/ποικιλία οπτικής γωνίας (viewpoint variation), 2) σε αυθαίρετη κλίμακα (scale), 3) αυθαίρετες συνθήκες φωτισμού (illumination), 4) οπτικών εμποδίων (occlusion), 5) αταξίας περιβάλλοντος (background clutter) και 6) παραμόρφωσης (deformation) (σχήμα 1.1).



Σχήμα 1.2: Πεταλούδες από το σύνολο δεδομένων Caltech-101 ως παράδειγμα αναγνώρισης κατηγορίας αντικειμένων.

Στη βιβλιογραφία αναφέρονται διάφορα προβλήματα αναγνώρισης. Η διαφοροποίησή τους έγκειται σε δύο βασικά χαρακτηριστικά, α) στο αν υπάρχει ένα αντικείμενο ή χαρακτηριστικό σε μια εικόνα σε σχέση με το που και β) στο αν θέλουμε να αναγνωρίσουμε συγκεκριμένα αντικείμενα ή ολόκληρες κατηγορίες αντικειμένων. Το πρώτο στοιχείο διαφοροποίησης αφορά στο είδος της απάντησης που αναζητά το πρόβλημα ναί/όχι σε σχέση με τη θέση του αντικειμένου στην εικόνα, ενώ το δεύτερο στοιχείο προσδιορίζει το βαθμό γενίκευσης του δεδομένου προβλήματος αναγνώρισης (σχήμα 1.3). Αναφέρουμε τρία βασικά προβλήματα αναγνώρισης.

- *Ανίχνευση (Detection)*, όπου προσπαθούμε να αναγνωρίσουμε ένα μεμονωμένο αντικείμενο ή χαρακτηριστικό το οποίο καλύπτει ολόκληρη την εικόνα. Στην περίπτωση αυτή για παράδειγμα, μια πεταλούδα με δεδομένο σχήμα διαφοροποιείται ως αντικείμενο από μια άλλη με έστω και ελάχιστα διαφορετικό σχήμα. Αντίθετα ένα πρόβλημα κατηγοριοποίησης θα ανέθετε τις δύο πεταλούδες στην ίδια κατηγορία (σχήματα 1.2, 1.4).

- *Αναγνώριση κατηγοριών (Category recognition)*, όπου καλούμαστε να αναγνωρίσουμε εκπαιδευμένα αντικείμενα, κατηγορίες αντικειμένων, όπως αυτοκίνητα, πρόσωπα, οπουδήποτε στην εικόνα. Σκοπός είναι να απαντήσουμε καταφατικά ή αρνητικά (σχήμα 1.2).

- *Ανάκτηση (Retrieval)*, όπου επιδιώκουμε να ταιριάξουμε την εικόνα αναζήτησης με κάποιες εικόνες βάσης σε μια λογική *ταιριάσματος των πιο κοντινών "γειτόνων" (nearest neighbours matching)*, με σκοπό στη συνέχεια την ανάκτηση τους ως ομοίων. Κατά τη διαδικασία εκτίμησης του κατά πόσο δύο εικόνες ταιριάζουν, αναζητώνται αντικείμενα ή κατηγορίες αντικειμένων οπουδήποτε στην εικόνα.

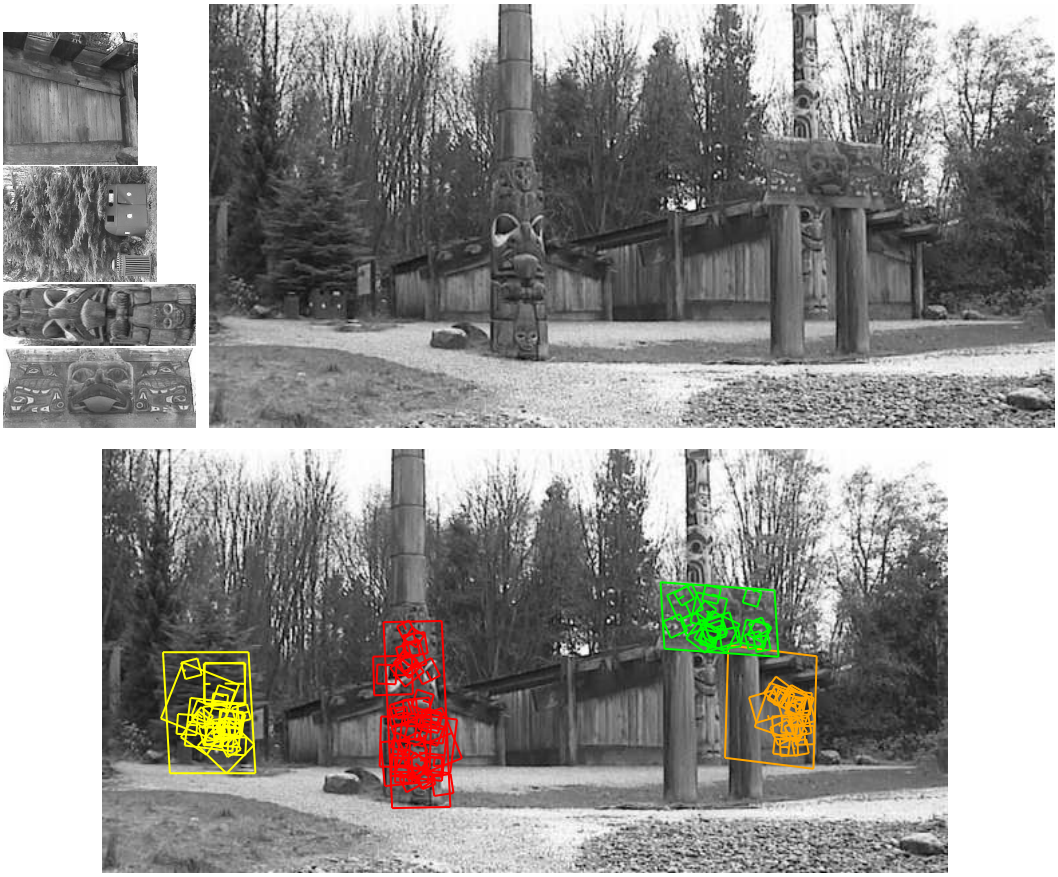
(α) (β)

Σχήμα 1.3: Διαφοροποίηση προβλημάτων αναγνώρισης. (α) Αναγνώριση συγκεκριμένων αντικειμένων (β) Αναγνώριση κατηγοριών http://cs.nyu.edu/ fergus/icml_tutorial/.

Οι υπάρχουσες τεχνικές κατηγοριοποίησης εικόνων παρουσιάζουν διάφορα προβλήματα. Συχνά απαιτούν ακριβή ευθυγράμμιση των εκπαιδευόμενων εικόνων και το διαχωρισμό τους σε διαφορετικές οπτικές γωνίες, διαδικασία χρονοβόρα που δυσχεραίνει την επιδίωξη για γενίκευση. Άλλες περιλαμβάνουν τον προσδιορισμό κάποιων καθολικών χαρακτηριστικών της εικόνας, διαδικασία δύσκολη σε εικόνες με έντονη ποικιλομορφία, ενώ κάποιες στερούνται του αναλλοίωτου όσον αφορά στην ποικιλία των συνθηκών που αναφέραμε στην προηγούμενη παράγραφο, στην προσπάθεια ενσωμάτωσης της χωρικής πληροφορίας της εικόνας (SPM).

Σχετικά με την ανάκτηση εικόνων, οι πιο πολλές τεχνικές ταιριάσματος εξετάζουν ένα σύνολο υποθέσεων ομομορφικών μετασχηματισμών επί χαρακτηριστικών πολλών διαστάσεων των εικόνων, και επιλέγουν κάθε φορά τη βέλτιστη υπόθεση με βάση κάποιο κριτήριο απονομής ψήφων. Ο συνολικός αριθμός των ψήφων καθορίζει την κατάταξη των εικόνων προς ανάκτηση σχετικά με την εικόνα αναζήτησης (RANSAC, LO-RANSAC, FSM). Στο πλαίσιο αυτό παρατηρούνται σημαντικά προβλήματα στον επιτυχή καθορισμό του συνόλου των υποθέσεων και σημαντικό υπολογιστικό κόστος στην εξέταση τους προς εύρεση της βέλτιστης αυτών, με αποτέλεσμα οι χρόνοι αναζήτησης να παραμένουν μεγάλοι. Επιπλέον, ο βαθμός ομοιότητας των εικόνων εξάγεται με βάση χαμηλού επιπέδου χαρακτηριστικά, τα οποία δεν παρουσιάζουν άμεση συσχέτιση με το σημασιολογικό περιεχόμενο της εικόνας. Το γεγονός αυτό καθιστά αναγκαία την ενσωμάτωση της χωρικής πληροφορίας σε δεύτερο στάδιο, με βασική επιδίωξη παράλληλα τη διατήρηση του αναλλοίωτου ως προς την κλίμακα, την περιστροφή και τη μετατόπιση και κατ'επέκταση της διακριτικής ικανότητας.

Παρά τις δυσκολίες και τα προβλήματα που παρουσιάζει, η ανάκτηση εικόνων παρέχει ένα ισχυρό και αποτελεσματικό πλαίσιο, όσον αφορά στην εξαγωγή του βαθμού γειτνίασης δύο εικόνων, τη στιγμή που οι τεχνικές εκμάθησης αδυνατούν να ανταποκριθούν σε ακρίβεια, στην αναγνώριση των εκάστοτε κατηγοριών, δεδομένου ενός βαθμού γενίκευσης. Γεννάται λοιπόν



Σχήμα 1.4: Παράδειγμα ανίχνευσης αντικειμένων σε μια πολύπλοκη σκηνή [15]. Οι εικόνες εκπαίδευσης φαίνονται πάνω αριστερά και η 640×315 εικόνα ελέγχου, η οποία έχει ληφθεί από διαφορετική οπτική γωνία πάνω δεξιά. Οι περιοχές που έχουν ανιχνευτεί φαίνονται στην κάτω εικόνα, με τα "σημεία κλειδιά"(keypoints) να αναπαρίστανται με τετράγωνα και ένα εξωτερικό παραλληλόγραμμο να δείχνει τα όρια των εκπαιδευμένων εικόνων με βάση τον αφινικό μετασχηματισμό που χρησιμοποιήθηκε κατά την αναγνώριση.

το ερώτημα αν θα μπορούσαμε να εκμεταλλευτούμε την αποτελεσματικότητα της διαδικασίας ανάκτησης σχετικά με τον τρόπο συσχέτισης των εικόνων, στην κατεύθυνση της βελτίωσης της κατηγοριοποίησης εικόνων. Μια τέτοια προσέγγιση φαίνεται να υπόσχεται την εκπαίδευση μοντέλων ταξινόμησης, τα οποία αποτυπώνουν με μεγαλύτερη ακρίβεια τα χαρακτηριστικά των διαφόρων κατηγοριών "παντρεύοντας" κατά κάποιο τρόπο γενίκευση εκμάθησης και διακριτική ικανότητα, συμβάλλοντας στην αποτελεσματικότητα και στην αποδοτικότητα της κατηγοριοποίησης εικόνων.

Το πρόβλημα λοιπόν που θα μας απασχολήσει στην προκειμένη εργασία είναι η κατηγοριοποίηση εικόνων στα πλαίσια της διαδικασίας ανάκτησης εικόνων σε μηχανές αναζήτησης από μεγάλες βάσεις εικόνων. Όταν αναφερόμαστε στην κατηγοριοποίηση εικόνων βασική επιδίωξη αποτελεί η εύρεση τεχνικών, οι οποίες είναι αρκετά γενικές ώστε να λειτουργούν αποτελεσματικά παρουσία πολλών κατηγοριών αντικειμένων ταυτόχρονα, ενώ παράλληλα μπορούν εύκολα να επεκταθούν, ώστε να συμπεριλαμβάνουν νέες κατηγορίες αντικειμένων. Πέρα από

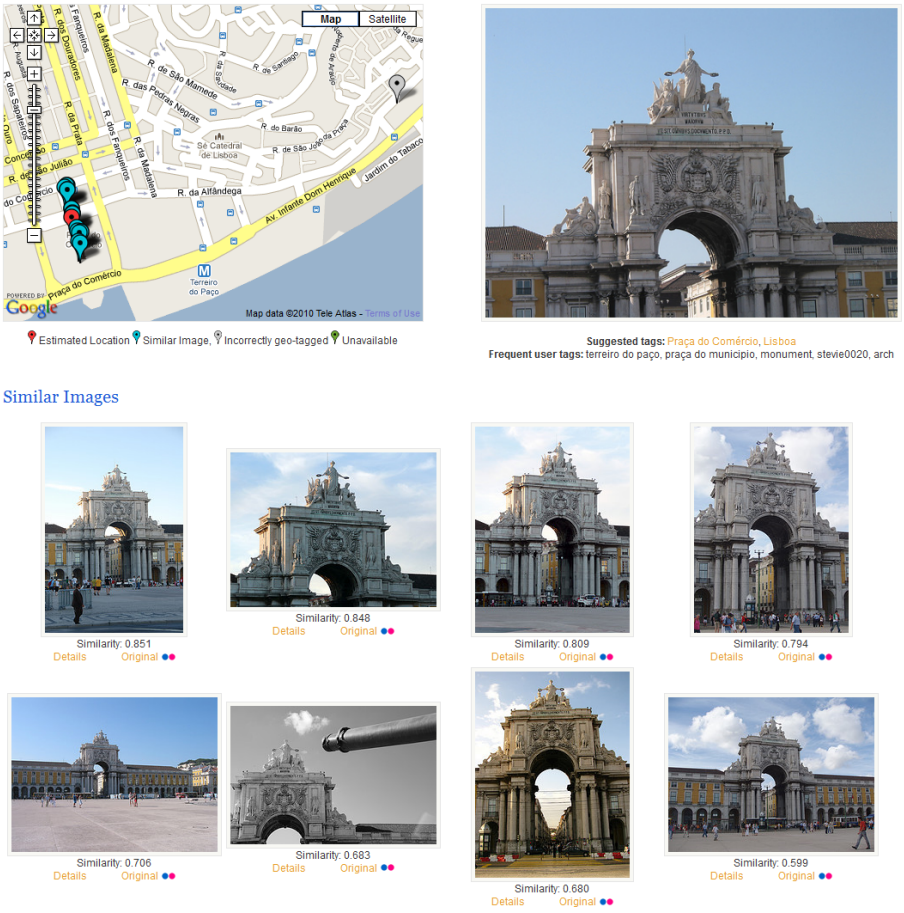
τη δυνατότητα γενίκευσης έναντι στις παραλλαγές των αντικειμένων μιας κλάσης, αυτές οι τεχνικές είναι σημαντικό να αντιμετωπίζουν ικανοποιητικά και τις έξι δυσκολίες που προαναφέραμε, φαινόμενα εγγενή στο φυσικό κόσμο.

1.2 Συνεισφορά εργασίας

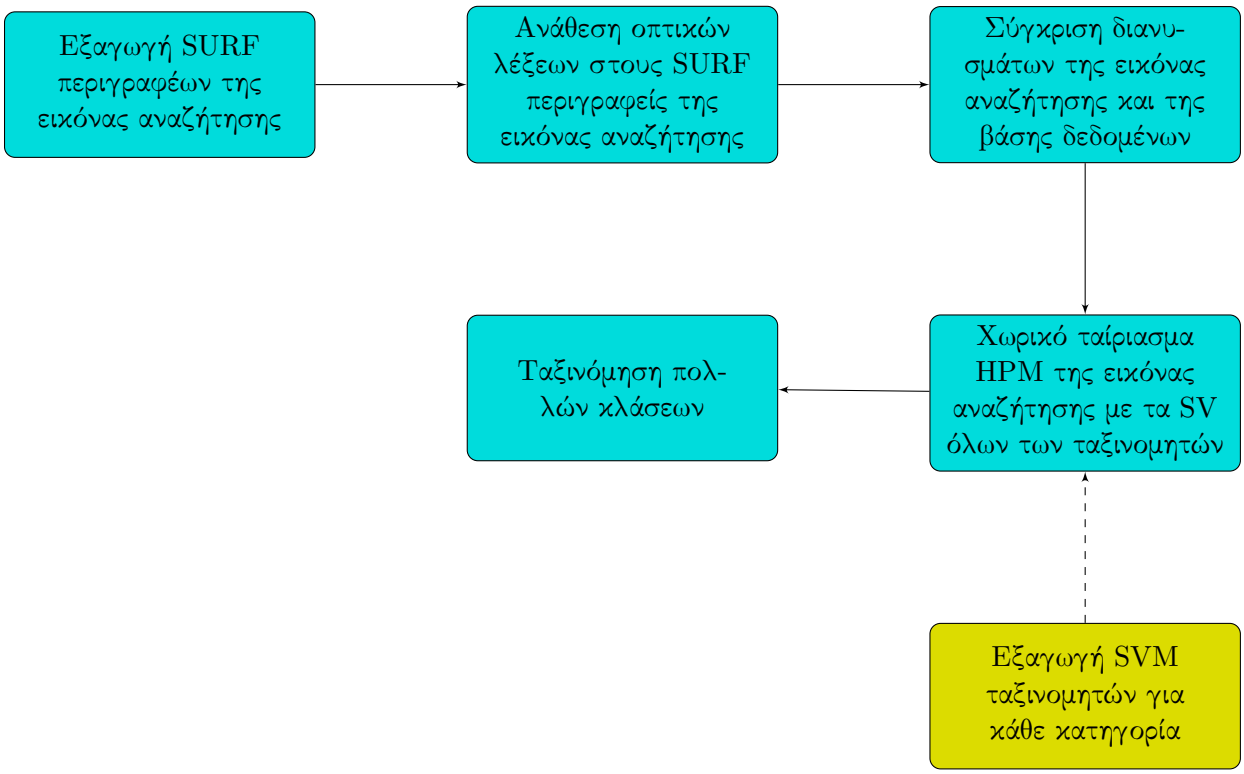
Η λειτουργία μιας μηχανής κατηγοριοποίησης εικόνων αξιολογείται με βάση δύο κύρια κριτήρια, την *αποτελεσματικότητα* της, δηλαδή πόσο ορθά αποδίδει τα εκάστοτε δείγματα ελέγχου στις υποψήφιες κατηγορίες, και την *αποδοτικότητα* της, δηλαδή πόσο υπολογιστικό χώρο και χρόνο απαιτεί για να παρέχει μια απάντηση στο χρήστη. Έχοντας κατά νού τα δύο αυτά κριτήρια, επιχειρούμε να συνδυάσουμε δύο αυτόνομες θεωρίες προκειμένου να επωφεληθούμε των πλεονεκτημάτων τους σε ένα ενοποιημένο πλαίσιο, την εκμάθηση και το χωρικό ταίριασμα εικόνων, όπως αυτό έχει ενσωματωθεί σε μια διαδικασία χωρικής επαλήθευσης και δεικτοδότησης στη διαδικασία ανάκτησης εικόνων.

Ήδη υπάρχουσες τεχνικές επιδιώκουν την επίλυση της κατηγοριοποίησης εικόνων αποσυνδεδεμένα συνήθως από τη διαδικασία είτε δεικτοδότησης, είτε χωρικού ταιριάσματος στις μηχανές αναζήτησης εικόνων. Οι Csurka και Dance στο [8] εισάγουν μια μέθοδο *οπτικής κατηγοριοποίησης (visual categorization)* με βάση τη θεωρία *"σάκος" από σημεία κλειδιά (bag of keypoints)* και εφαρμόζουν ταξινόμηση πολλών κλάσεων με μηχανές διανυσμάτων υποστήριξης (SVM) και γραμμικό πυρήνα. Οι Lazenbnik, Schmid και Ponce στο [13] αναπτύσσουν μια μέθοδο χωρικού ταιριάσματος πυραμίδας και εισάγουν το αποτέλεσμα της υπό τη μορφή πυρήνα σε SVM προκειμένου να εφαρμόσουν στη συνέχεια ταξινόμηση πολλών κλάσεων. Ακόμη οι Li, Crandall και Huttenlocher στο [14] εφαρμόζουν με τον ίδιο τρόπο ταξινόμηση πολλών κλάσεων, αξιοθέατων για την ακρίβεια, είτε αξιοποιώντας παράλληλα κάποια πληροφορία κειμένου που συνοδεύει τις εκπαιδευόμενες εικόνες, είτε εισάγοντας χρονικούς περιορισμούς στο πρόβλημα εκμάθησης σχετικά με ακολουθίες εικόνων που παρουσιάζουν κάποια χρονική συσχέτιση και έχουν κοινοποιηθεί από τον ίδιο χρήστη.

Οι Tolias και Avrithis στο [20] εισάγουν τον αλγόριθμο *ταιριάσματος πυραμίδας Hough* (HPM), μια τεχνική χαλαρού, χωρικού ταιριάσματος, η οποία είναι ανεξάρτητη από μετασχηματισμούς ομοιότητας και δεν απαιτεί τη χρονοβόρα διαδικασία βελτιστοποίησης και απαρίθμησης ψήφων, που περιλαμβάνουν πολλές υπάρχουσες μέθοδοι χωρικού ταιριάσματος. Επιπλέον, παρουσιάζει ευελιξία, καθώς διατηρεί τη μη άκαμπτη κίνηση και το πολλαπλό ταίριασμα επιφανειών και αντικειμένων, χωρίς όμως να χάνει και το αναλλοίωτο σε κλίμακα, μετατόπιση και περιστροφή. Όλα αυτά τα χαρακτηριστικά συνθέτουν μια μέθοδο που εκμεταλλεύεται τη χωρική πληροφορία με αποδοτικό τρόπο, δίνοντας παράλληλα τη δυνατότητα ενσωμάτωσης γρήγορων και εύκολα υλοποιήσιμων αλγορίθμων για την ανακατάταξη εικόνων σε μηχανές αναζήτησης.



Σχήμα 1.5: Παράδειγμα ανάκτησης εικόνων από την εφαρμογή VIRaL (http://viral.image.ntua.gr) [12]. Αποτελέσματα επιτυχούς οπτικού ταιριάσματος, τοποθεσίας και αναγνώρισης αξιοθέατου. **Πάνω αριστερά**: χάρτης που απεικονίζει τις ακριβείς θέσεις των παρόμοιων εικόνων (similar images) (μπλέ ενδείξεις) και η εκτιμώμενη τοποθεσία της εικόνας αναζήτησης (κόκκινες ενδείξεις). **Πάνω δεξιά**: η εικόνα αναζήτησης μαζί με τα σύνολα των πιο συχνών και προτεινόμενων ετικετών (tags). **Κάτω**: οπτικά παρόμοιες εικόνες.



Σχήμα 1.6: Κατηγοριοποίηση εικόνων με SVM και πυρήνα πυραμίδας Hough

Στην παρούσα διπλωματική εργασία αναπτύσσουμε μια μέθοδο κατηγοριοποίησης εικόνων με SVM, η οποία ενσωματώνει την τεχνική HPM, ενώ συνδέεται άμεσα με τη διαδικασία δεικτοδότησης της ανάκτησης εικόνων με ανεστραμμένο αρχείο. Αφού αποδείξουμε ότι το αποτέλεσμα της τεχνικής HPM αποτελεί πυρήνας, εκπαιδεύουμε μεμονωμένους ταξινομητές κατηγοριών εικόνων SVM, οι οποίοι στη συνέχεια συνδυάζονται σε μια διαδικασία ταξινόμησης πολλών κλάσεων με βάση την προσέγγιση "ένα έναντι των υπολοίπων" (one versus the rest) την οποία και θα αναπτύξουμε σε επόμενο κεφάλαιο 3.7. Παράλληλα η μέθοδος που εισάγουμε, χρησιμοποιεί το ανεστραμμένο αρχείο για τον υπολογισμό των τιμών του πυρήνα HPM, επιταχύνοντας με αυτό τον τρόπο σημαντικά τη διαδικασία εκπαίδευσης. Τα βήματα της μεθόδου κατηγοριοποίησης απεικονίζονται συνοπτικά στο γράφημα 1.6.

Συνοψίζουμε διαδοχικά τα εξής:

- Για πρώτη φορά εισάγεται σε SVM πυρήνας πέραν του γραμμικού, ο οποίος να συνδυάζει χωρική πληροφορία και το αναλλοίωτο. Συγκεκριμένα, πετυχαίνουμε την εκμετάλλευση των αραιών και συμπαγών αναπαραστάσεων των λύσεων που παρέχουν τα SVM, σε συνδυασμό με την αποδοτική αξιοποίηση της χωρικής πληροφορίας που προσφέρει η τεχνική HPM, δεδομένου ότι διατηρείται το αναλλοίωτο σε κλίμακα, μετατόπιση και περιστροφή.

- Συνδυάζεται η διαδικασίας ανάκτησης εικόνων με την διαδικασία εκμάθησης με σκοπό τη βελτίωση της κατηγοριοποίησης εικόνων, σε μια λογική εξισορρόπησης διακριτικής ικανότητας και προσπάθειας για γενίκευση.

- Εισάγουμε μια αποτελεσματική μέθοδο κατηγοριοποίησης και χωρικού ταιριάσματος,

υποσχόμενη καλύτερους υπολογιστικούς χρόνους και απαιτήσεις σε μνήμη στις μηχανές ανάκτησης εικόνων από μεγάλες βάσεις εικόνων, η οποία ταυτόχρονα παρέχει την προοπτική ενσωμάτωσης σημασιολογικού περιεχομένου στη διαδικασία ανάκτησης εικόνων.

Όσον αφορά στην υλοποίηση, στο πλαίσιο της διπλωματικής εργασίας και για την κάλυψη των απαιτήσεων της πειραματικής διαδικασίας ασχολήθηκα εκτενώς με πληθώρα εφαρμογών και την ανάπτυξη κώδικα.

Αρχικά διεξήχθησαν αρκετά πειράματα στο περιβάλλον Matlab για την εξοικείωση με τα SVM και τις διάφορες μοντελοποιήσεις τους σε συνθετικά κυρίως δεδομένα και με συνήθεις πυρήνες όπως ο γραμμικός και ο Gaussian. Σε αυτό το πρώτο στάδιο πραγματοποιήθηκε παράλληλα και μιας μορφής αξιολόγηση των δεδομένων εργαλείων εκπαίδευσης, καθώς και μια προσπάθεια βελτιστοποίησης των διαφόρων παραμέτρων για την εξαγωγή υψηλών τιμών ακρίβειας ταξινόμησης των δειγμάτων ελέγχου.

Στη συνέχεια, ασχολήθηκα με τις εικόνες σε πρώτη φάση υλοποιώντας το γραμμικό πυρήνα BoW σε C++, στο προγραμματιστικό περιβάλλον Visual Studio, όπως αυτός υλοποιείται στο πλαίσιο της διαδικασίας ανάκτησης εικόνων με ανεστραμμένο αρχείο, με σκοπό τόσο την εξοικείωση με τη γλώσσα προγραμματισμού, όσο και τη βαθύτερη κατανόηση της διαδικασίας. Ακολούθως, χρησιμοποιήσαμε έτοιμους πυρήνες HPM υπολογισμένους σε εξυπηρετητή του εργαστηρίου, προς σύγκριση με το γραμμικό πυρήνα BoW και έλεγχο της αρχικής ιδέας σε μικρή κλίμακα, δεδομένου του ερευνητικού χαρακτήρα της εργασίας.

Με θετικά λοιπόν αποτελέσματα από τη μέχρι τότε διαδικασία, προχωρήσαμε σε μεγαλύτερης κλίμακας πειράματα στον εξυπηρετητή του εργαστηρίου σε μεγαλύτερη βάση εικόνων. Για τις ανάγκες των πειραμάτων στον εξυπηρετητή εξοικειώθηκα με τη γλώσσα προγραμματισμού Bash shell programming σε περιβάλλον Linux/Unix και όλοι οι υπολογισμοί πραγματοποιήθηκαν σε συνδυασμό και με εκτελέσιμα αρχεία προερχόμενα από κώδικα σε C++.

Να σημειώσουμε ότι όλη η διαδικασία εκπαίδευσης τόσο στο περιβάλλον Matlab όσο και με τη μορφή εκτελέσιμων στον εξυπηρετητή, πραγματοποιήθηκε με βάση τη βιβλιοθήκη LIBSVM ([1]). Επίσης χρησιμοποιήθηκαν έτοιμα εκτελέσιμα αρχεία υλοποίησης του αλγορίθμου HPM, όπως αυτά είχαν υλοποιηθεί και για τη διεξαγωγή των πειραμάτων στο [20].

1.3 Δομή διπλωματικής

Η παρούσα διπλωματική εργασία δομείται στο πλαίσιο δύο βασικών θεμάτων, των μηχανών διανυσμάτων υποστήριξης (SVM) και του χωρικού ταιριάσματος εικόνων σε μηχανές αναζήτησης, ως υπόβαθρο για την εισαγωγή μιας νέας τεχνικής κατηγοριοποίησης εικόνων. Στο **κεφάλαιο** 2, περιγράφονται βασικές έννοιες της θεωρίας εκμάθησης και της θεωρίας των πυρήνων, ενώ παράλληλα ορίζεται το χωρικό ταίριασμα στο πλαίσιο της διαδικασίας ανάκτησης εικόνων από μεγάλες βάσεις δεδομένων σε μηχανές αναζήτησης. Στο **κεφάλαιο** 3, επικεντρωνόμαστε στις μηχανές διανυσμάτων υποστήριξης (SVM), τις διάφορες περιγραφές και επεκτάσεις τους με την ενσωμάτωση πυρήνων, με σκοπό την ανάδειξη των πλεονεκτημάτων τους στην κατεύθυνση του προβλήματος κατηγοριοποίησης. Στο **κεφάλαιο** 4 γίνεται μια ανασκόπηση διαφόρων τεχνικών χωρικού ταιριάσματος εικόνων και των αλγορίθμων που

[1] LIBSVM tool

περιλαμβάνουν, με έμφαση στον αλγόριθμο *ταιριάσματος πυραμίδας Hough* (HPM) και τη λειτουργία του, ο οποίος αποτελεί και δομικό στοιχείο της τεχνικής που εισάγουμε. Στο **κεφάλαιο** 5 αποδεικνύουμε ότι ο αλγόριθμος HPM είναι πυρήνας, προκειμένου στη συνέχεια να χρησιμοποιηθεί στην εκπαίδευση SVM, με σκοπό την εκμάθηση κατηγοριών εικόνων. Στο **κεφάλαιο** 6 παρατίθενται τα αποτελέσματα της νέας μεθόδου κατηγοριοποίησης εικόνων σε σύνολα δεδομένων από αξιοθέατα μαζί με το σχολιασμό τους και ακολούθως γίνεται σύγκριση με τις ήδη υπάρχουσες τεχνικές που προαναφέρθηκαν. Τέλος, στο **κεφάλαιο** 7 συνοψίζουμε τα συμπεράσματα επί της μεθόδου που αναπτύξαμε και παρουσιάζουμε νέες ιδέες στην προσπάθεια που αυτή τροφοδοτεί για περαιτέρω έρευνα.

Κεφάλαιο 2

Γενικό θεωρητικό υπόβαθρο

2.1 Εισαγωγή

Σε αυτό το κεφάλαιο αναπτύσσουμε τις δύο βασικές θεωρίες, στις οποίες βασίζεται αλλά και τις οποίες επεκτείνει η παρούσα διπλωματική εργασία στην κατεύθυνση της αποτελεσματικής αλλά και αποδοτικής κατηγοριοποίησης εικόνων σε μεγάλες βάσεις δεδομένων. Αποτελεσματικής όσον αφορά στην ορθότητα ταξινόμησης των δειγμάτων εκπαίδευσης και αποδοτικής, όσον αφορά στις απαιτήσεις σε μνήμη και χρόνο υλοποίησης. Η πρώτη θεωρία αφορά γενικά την εκμάθηση και ταξινόμηση δεδομένων, ενώ η δεύτερη το χωρικό ταίριασμα εικόνων, καθώς και όλα τα επιμέρους στάδια επεξεργασίας που αυτό περιλαμβάνει. Προτού ξεκινήσουμε, αναφέρουμε ότι η ανάλυση της πρώτης θεωρίας βασίζεται στα συγγράμματα [7], [4] και [16].

2.2 Εκμάθηση

Ακολουθώντας την παραδοσιακή προγραμματιστική προσέγγιση προκειμένου να λύσουμε ένα πρόβλημα, υποδεικνύουμε στον υπολογιστή τη μέθοδο επίλυσης με βάση τα δεδομένα εισόδου, βήμα προς βήμα. Σε κάποιες περιπτώσεις προβλημάτων όμως, η παραπάνω προσέγγιση δεν είναι εφικτή, δημιουργώντας εύλογα το ερώτημα αν θα μπορούσαμε εναλλακτικά να εκπαιδεύσουμε τον υπολογιστή, όσον αφορά τη συσχέτιση εισόδου/εξόδου, βασιζόμενοι σε παραδείγματα. Αυτός είναι άλλωστε και ο τρόπος που μαθαίνουν οι άνθρωποι από παιδιά να ξεχωρίζουν αντικείμενα,να διαβάζουν κ.λ.π. Αυτός ο τρόπος επίλυσης προβλημάτων με παραδείγματα ορίζεται ως *εκμάθηση (learning)*.

Στην περίπτωση που τα παραδείγματα προς εκμάθηση αποτελούν ζεύγη εισόδου/εξόδου αναφερόμαστε στον όρο *εποπτευόμενη εκμάθηση (supervised learning)*, ενώ τα παραδείγματα ζευγών εισόδου/εξόδου αναφέρονται ως *δεδομένα εκπαίδευσης (training data)*. Όταν δεν υπάρχουν τιμές εξόδου, η εκμάθηση αφορά την κατανόηση της διαδικασίας που παρήγαγε τα δεδομένα, και αναφέρεται ως *μη εποπτευόμενη εκμάθηση (unsupervised learning)*. Επιπλέον, όταν τα δεδομένα εκπαίδευσης δίδονται στη μηχανή εκπαίδευσης από την αρχή της διαδικασίας εκμάθησης αναφερόμαστε στον όρο *εκμάθηση "παρτίδας" (batch learning)*.

(α)

(β)

Σχήμα 2.1: Επιδιώκουμε να αναγνωρίσουμε πρότυπα αντικειμένων στο σύνολο των δεδομένων (α), με βάση κάποιο χαρακτηριστικό που προσδιορίζει κάθε κατηγορία και τη διαφοροποιεί από την άλλη. Στο συγκεκριμένο παράδειγμα θέλουμε να μάθουμε να διακρίνουμε τα λούτρινα ζώα από τα πραγματικά (β) http://cs.nyu.edu/ fergus/icml_tutorial/.

Η σχέση των ζευγών εισόδου/εξόδου, όταν είναι προσδιορίσιμη, ονομάζεται *συνάρτηση στόχος (target function)*. Η συνάρτηση στόχος αποτελεί και τη *λύση του προβλήματος εκμάθησης (solution of the learning problem)* και επιλέγεται από ένα χώρο υποψήφιων συναρτήσεων, ο οποίος αναφέρεται ως *χώρος υποθέσεων (hypothesis space)*. Ο αλγόριθμος που δέχεται ως είσοδο τα δεδομένα εκπαίδευσης και επιλέγει μια υπόθεση από το χώρο υποθέσεων αναφέρεται ως *αλγόριθμος εκμάθησης (learning algorithm)*. Η επιλογή του χώρου των υποθέσεων και ο αλγόριθμος εκμάθησης αποτελούν τα σημαντικότερα συστατικά της στρατηγικής εκμάθησης. Να σημειώσουμε ότι όλα τα συστήματα εκμάθησης υλοποιούν κάποια πρότερη υπόθεση Bayesian τύπου, η οποία ονομάζεται *προκατάληψη εκμάθησης (learning bias)*, δεδομένης της ελευθερίας επιλογής του χώρου των υποθέσεων και της εκτίμησης των πρότερων πιθανοτήτων τους.

Μια υποπερίπτωση προβλήματος εποπτευόμενης εκμάθησης αποτελεί η *ταξινόμηση*

(classification), κατά την οποία η μηχανή εκπαίδευσης καλείται να αναγνωρίσει σε ποιά κατηγορία ανήκει μια νέα παρατήρηση/είσοδος, δεδομένου ενός συνόλου δεδομένων εκπαίδευσης, το οποίο περιλαμβάνει ήδη ταξινομημένες παρατηρήσεις. Η συνάρτηση στόχος στην περίπτωση της ταξινόμησης αναφέρεται και ως *συνάρτηση απόφασης (decision function)*.

Ένα πρόβλημα ταξινόμησης με δυαδικές εξόδους, που είναι και το πιο απλό, αναφέρεται ως *δυαδικό πρόβλημα ταξινόμησης (binary classification problem)*, ενώ όταν ο αριθμός των κατηγοριών είναι πεπερασμένος *πρόβλημα ταξινόμησης πολλών κλάσεων (multi-class classification problem)*.

Η μεθοδολογία εκμάθησης υπόσχεται μείωση της πολυπλοκότητας επίλυσης των προβλημάτων σε υπολογιστικό χώρο και χρόνο συγκριτικά με την παραδοσιακή προγραμματιστική προσέγγιση, όπως θα δούμε και στη συνέχεια, παράλληλα όμως παρουσιάζει τις εξής δυσκολίες:

- Ο αλγόριθμος εκμάθησης μπορεί να αποδειχτεί ανεπαρκής, όπως για παράδειγμα σε τοπικά ελάχιστα.
- Ο χώρος των υποθέσεων μπορεί συχνά να γίνει πολύ μεγάλος και μη πρακτικός.
- Αν ο αριθμός των δεδομένων εκπαίδευσης είναι μικρός και οι υποθέσεις που επιλέγονται είναι αρκετά πολύπλοκες, ώστε να ταιριάζουν απόλυτα στα δεδομένα εκπαίδευσης, ενδέχεται να οδηγηθούμε σε *υπερταίριασμα (overfitting)* στα δεδομένα εκπαίδευσης και *φτωχή/ανεπαρκή γενίκευση (poor generalisation)* της λύσης του προβλήματος, δηλαδή σε λύσεις ανεπαρκώς ανταποκρινόμενες σε δεδομένα που δεν ανήκουν στο σύνολο δεδομένων εκπαίδευσης.
- Συχνά ο αλγόριθμος εκμάθησης ελέγχεται από μεγάλο αριθμό παραμέτρων που απαιτούν την επιλογή συγκεκριμένων *χειρισμών (heuristics)* και καθιστούν το σύστημα αναξιόπιστο και δύσχρηστο.

Στην παρούσα εργασία θα χρησιμοποιήσουμε μια συγκεκριμένη κατηγορία μηχανών εκμάθησης, τις *μηχανές εκμάθησης διανυσμάτων υποστήριξης (ΜΔΥ) (support vector machines* (SVM), προκειμένου να αξιοποιήσουμε τα πλεονεκτήματα που αυτή παρέχει στο πλαίσιο της κατηγοριοποίησης εικόνων, που μας ενδιαφέρει.

2.3 Μοντελοποίηση και ταίριασμα εικόνων

Πληθώρα εφαρμογών ωθούν στην ανάπτυξη αποτελεσματικών τεχνικών *ταιριάσματος εικόνων (matching)*. Εμείς θα επικεντρωθούμε στο *χωρικό ταίριασμα εικόνων (spatial matching)*, όπου με τον όρο *χωρικό* αναφερόμαστε στην χρησιμοποίηση πληροφορίας που αφορά τη χωρική διάταξη των χαρακτηριστικών που εξάγονται από τις εικόνες των οποίων θέλουμε να υπολογίσουμε την ομοιότητα. Χαρακτηριστικά παραδείγματα αποτελούν η αναγνώριση κατηγοριών εικόνων, η ανίχνευση εικόνων με κυρίαρχο την ανάκτηση εικόνων σε μεγάλη κλίμακα. Διακριτική ικανότητα, γεωμετρικό αναλλοίωτο (geometry invariance), περιορισμοί ακαμψίας και αντιστοίχησης (rigidity and mapping constraints), υποθέσεις όσον αφορά στους υποκείμενους περιγραφείς ή χαρακτηριστικά της εικόνας και φυσικά χαμηλή υπολογιστική πολυπλοκότητα περιλαμβάνονται στις προδιαγραφές μιας αποτελεσματικής μηχανής αναζήτησης.

Η ανάκτηση εικόνων αποτελεί βασική εφαρμογή που αναδεικνύει τη σημασία του χωρικού ταιριάσματος εικόνων. Ως *ανάκτηση εικόνων (image retrieval)*, ορίζουμε το εξής πρόβλημα: Δοθείσης μιας *εικόνας αναζήτησης (query)*, θέλουμε η μηχανή αναζήτησης να ανακτήσει όλες τις εικόνες από μια μεγάλη βάση δεδομένων που ταιριάζουν με αυτή. Στις επόμενες ενότητες του κεφαλαίου θα αναπτύξουμε εν συντομία τα βασικά στάδια της διαδικασίας ανάκτησης εικόνων.

2.3.1 Εξαγωγή χαρακτηριστικών

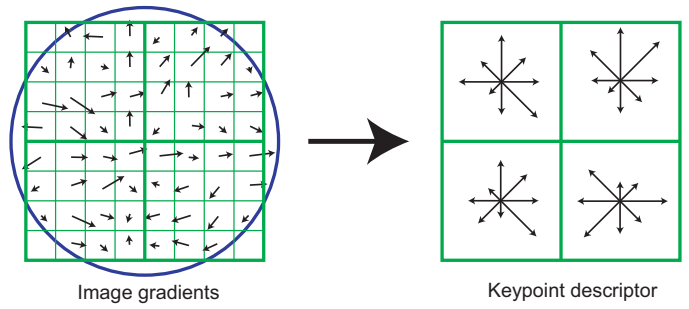
Γενικά υπάρχουν δύο είδη αναπαραστάσεων/περιγραφών των εικόνων. Η μια κατηγορία είναι οι *dense (πυκνές)* περιγραφές και αφορούν χαρακτηριστικά της εικόνας εξαγόμενα με διάφορες μεθόδους στο σύνολο της εικόνας σε κάποιο πλέγμα. Οι περιγραφές αυτές είναι πιο κατάλληλες στην κατάτμηση της εικόνας σε ομοιόμορφες περιοχές με διαφορετική υφή και στην αναγνώριση αντικειμένων με δεδομένο σχήμα από το περιβάλλον τους. Η δεύτερη κατηγορία περιγραφών είναι η ανίχνευση *σημείων ενδιαφέροντος (interest points)* με διάφορες *καθολικές (global)* τεχνικές, αραιές περιγραφές. Να αναφέρουμε ότι συνήθως οποιαδήποτε περιγραφή αφορά τη μονοχρωματική εκδοχή των εικόνων.

(α) (β)

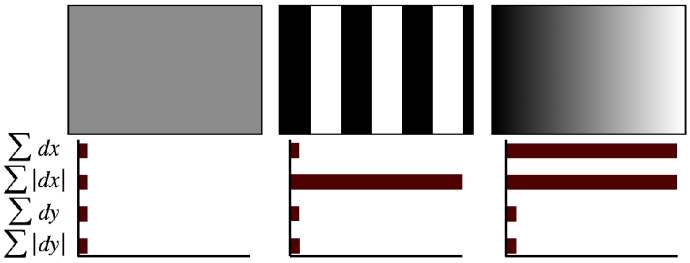
Σχήμα 2.2: Διαφοροποίηση περιγραφών των εικόνων. (α) Πυκνές περιγραφές (http://cs.nyu.edu/ fergus/icml_tutorial/) (β) Ανίχνευση σημείων ενδιαφέροντος (SIFT).

Όσον αφορά στην ανάκτηση εικόνων, η εξαγωγή χαρακτηριστικών περιλαμβάνει δύο διαδικασίες: την *ανίχνευση σημείων ενδιαφέροντος (interest point detector)* και την *εξαγωγή αντίστοιχων περιγραφέων τους (descriptors)*, με κύριο κριτήριο σε κάθε περίπτωση την επαναληπτικότητα (repeatability), δηλαδή τα συγκεκριμένα σημεία να ανιχνεύονται σε οποιεσδήποτε οπτικές συνθήκες, το αναλλοίωτο από κλίμακα και περιστροφή (scale and rotation invariance) και επομένως τη διακριτικότητα και την ευρωστία.

Σχετικά με την πρώτη διαδικασία, ανιχνεύουμε συνήθως σημεία ενδιαφέροντος σε συγκεκριμένες περιοχές της εικόνας, όπως γωνίες, blobs, T-junctions, οι οποίες συνήθως είναι αφινικά ανεξάρτητες. Να σημειώσουμε ότι σε κάθε εικόνα διάστασης 1024 ανιχνεύονται 3300 περίπου περιοχές. Πολύ δημοφιλείς είναι οι *Hessian ανιχνευτές*.



(α)



(β)

Σχήμα 2.3: (α) **SIFT**: Ένας περιγραφέας εξάγεται υπολογίζοντας αρχικά το μέτρο και τον προσανατολισμό της κλίσης (gradient) γύρω από κάθε σημείο δείγμα της εικόνας σε μια περιοχή όπως φαίνεται αριστερά. Οι ποσότητες αυτές σταθμίζονται με βάση ένα Gaussian παράθυρο, το οποίο επισημαίνεται με τον υπερκείμενο κύκλο. Τα δείγματα αυτά αθροίζονται σε ιστογράμματα προσανατολισμών επί 4×4 υποπεριοχών, όπως φαίνεται δεξιά, με το μήκος κάθε βέλους να αντιστοιχεί στο άθροισμα του μεγέθους των κλίσεων κοντά σε κάθε κατεύθυνση εντός της περιοχής αυτής. Στο σχήμα απεικονίζεται ένας 2×2 περιγραφέας που έχει εξαχθεί από 8×8 σύνολα δειγμάτων, ενώ πιο συχνά χρησιμοποιούνται 4×4 περιγραφείς από 16×16 πίνακες δειγμάτων [15]. (β) **SURF**: Οι τιμές του περιγραφέα για μια υποπεριοχή εκφράζονται με την παρακάτω ανάλυση έντασης. **Αριστερά** Σε περίπτωση μιας ομογενούς περιοχής, όλες οι τιμές είναι σχετικά χαμηλές. **Μέση** Όταν υπάρχουν συχνότητες στην κατεύθυνση x, η τιμή του $\sum|d_x|$ είναι υψηλή, αλλά όλες οι άλλες παραμένουν χαμηλές. Αν η ένταση αυξάνεται σταδιακά στην κατεύθυνση x, και οι δύο τιμές $\sum d_x, \sum|d_x|$ είναι υψηλές [3].

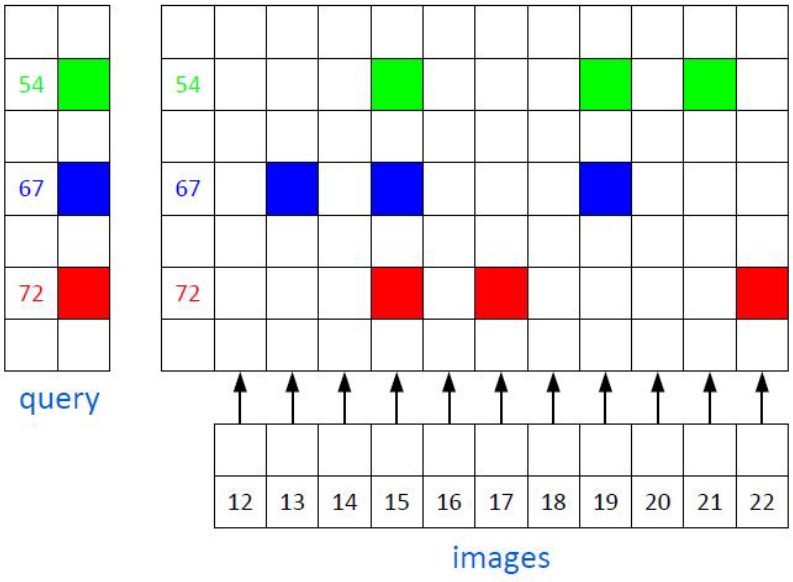
Στη συνέχεια, η γειτονιά κάθε σημείου ενδιαφέροντος περιγράφεται με ένα διάνυσμα χαρακτηριστικών. Δύο πολλοί γνωστοί περιγραφείς είναι οι *Μετασχηματισμός Χαρακτηριστικών Ανεξάρτητος Κλίμακας (Scale Invariant Feature Transform* (SIFT) [15] και τα *Γρήγορα Εύρωστα Χαρακτηριστικά (Speeded Up Robust Features* (SURF) [3]. Οι SIFT περιγραφείς περιλαμβάνουν ένα διάνυσμα 128 διαστάσεων, το οποίο υπολογίζεται με βάση ένα ιστόγραμμα προσανατολισμένων κλίσεων (gradients) (8 κατευθύνσεις για κάθε 4×4 υποπεριοχές). Οι SURF περιλαμβάνουν ένα διάνυσμα 64 διαστάσεων στην πιο απλή περίπτωση, το οποίο εξάγεται με βάση αθροίσματα *κυματίου Harris (Harris wavelet)* αποκρίσεων στην κατεύθυνση x και y με έναν επικρατών προσανατολισμό. Ο προσανατολισμός αυτός εξάγεται βασιζόμενος σε πληροφορία από μια κυκλική περιοχή γύρω από το σημείο ενδιαφέροντος (4 αθροίσματα για κάθε 4×4 τετραγωνικές υποπεριοχές). Να σημειώσουμε ότι οι περιγραφείς SURF είναι ιδιαίτερα γρήγοροι στον υπολογισμό τους, λόγω της εξαγωγής τους με τη χρήση ολοκληρωτικών

(integral) εικόνων (υπολογισμός αθροισμάτων).

2.3.2 Οπτικό λεξικό και ανεστραμμένο αρχείο

Για την κβάντιση των περιγραφέων της πληθώρας των εικόνων σε μια μηχανή αναζήτησης, έχουν αναπτυχθεί διάφορες τεχνικές φιλτραρίσματος με επικρατέστερη την τεχνική *"σάκος" οπτικών λέξεων (bag of visual words)* [19]. Σύμφωνα με την τεχνική αυτή, τα διανύσματα κβάντισης προκύπτουν από την εφαρμογή σε ένα σύνολο εικόνων εκπαίδευσης κάποιου αλγορίθμου συσταδοποίησης (clustering), όπως ο k-means, προσεγγιστικός (approximate) k-means (AKM), ιεραρχικός (hierarchical) k-means (HPM). Τα κέντρα των συστάδων που προκύπτουν, αποτελούν ουσιαστικά τις *λέξεις του οπτικού λεξικού (visual words)*.

Στη συνέχεια κάθε εικόνα της βάσης αναζήτησης αναπαριστάται σαν ένα αραιό διάνυσμα των συχνοτήτων των οπτικών λέξεων. Για την αναζήτηση εικόνων στη βάση, συγκρίνουμε το διάνυσμα της εικόνας αναζήτησης με τα διανύσματα των εικόνων της βάσης με τη χρήση διαφόρων μετρικών, όπως L1/L2 απόσταση (distance), διασταύρωση (intersection). Για λόγους ταχύτητας, κατασκευάζουμε ένα *ανεστραμμένο αρχείο (inverted file)* των εικόνων της βάσης, το οποίο περιέχει μία καταχώριση για κάθε οπτική λέξη του λεξικού, ενώ κάθε καταχώριση ακολουθείται από μία λίστα με τις εικόνες της βάσης που περιέχουν αυτήν τη λέξη. Ένα απλό παράδειγμα ανεστραμμένου αρχείου παρουσιάζεται στο σχήμα 2.4. Η αριστερή στήλη του μεγάλου πίνακα περιέχει τις λέξεις του λεξικού, ενώ κάθε λέξη ακολουθείται από μία λίστα με τις εικόνες που την εμπεριέχουν.



Σχήμα 2.4: Σε ένα ανεστραμμένο αρχείο (inverted file) κάθε λέξη του οπτικού λεξικού ακολουθείται από μία λίστα με εικόνες που την περιλαμβάνουν.

Η υιοθέτηση ενός σχήματος με βάρη όπως το *συχνότητας λέξεων-συχνότητας του ανεστραμμένου εγγράφου (tf-idf)* [2], μας επιτρέπει την υποβάθμιση της συνεισφοράς των πιο συχνών οπτικών λέξεων, δίνοντας παράλληλα μεγαλύτερη βαρύτητα στις πιο σπάνιες. Τα βάρη tf-idf υπολογίζονται ως εξής: Υποθέτουμε ότι έχουμε ένα λεξικό με k οπτικές λέξεις, οπότε κάθε

εικόνα αναπαριστάται με ένα διάνυσμα διάστασης k, $V_d = (t_1, ...t_i, ...t_k)$ από σταθμισμένες συχνότητες των λέξεων με στοιχεία:

$$t_i = \frac{n_{id}}{n_d} \log \frac{N}{n_i}$$

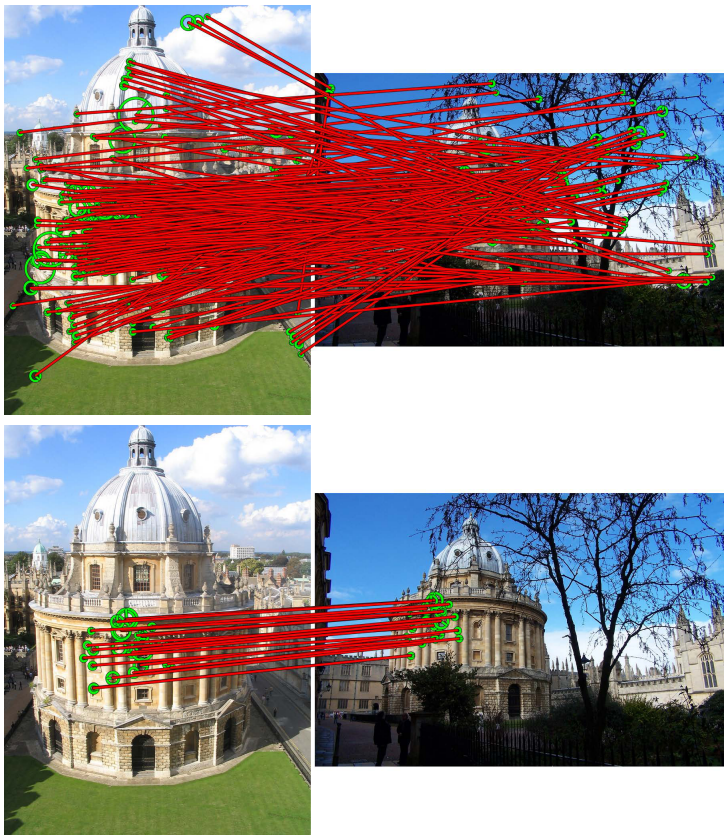
όπου n_{id} είναι ο αριθμός των εμφανίσεων της λέξης i στην εικόνα d, n_d είναι ο συνολικός αριθμός λέξεων στην εικόνα d, n_i ο αριθμός των εμφανίσεων της λέξης i σε όλη τη βάση και N ο αριθμός όλων των εικόνων της βάσης. Ο σταθμικός όρος είναι γινόμενο δύο όρων: της *συχνότητας λέξεων (tf)* n_{id}/n_d και της *συχνότητας του ανεστραμμένου εγγράφου (idf)* $\log N/N_i$. Ο όρος tf ενισχύει τη βαρύτητα των συχνών οπτικών λέξεων σε μια εικόνα, ενώ ο όρος idf μειώνει τη βαρύτητα των πιο συχνών στη βάση δεδομένων οπτικών λέξεων, δεδομένου ότι δεν συνεισφέρουν θετικά στη διακριτικότητα και ταυτοποίηση των εικόνων. Στη χειρότερη περίπτωση, η υπολογιστική πολυπλοκότητα αναζήτησης στο ανεστραμμένο αρχείο είναι γραμμική με το μέγεθος της βάσης εικόνων, ενώ για αραιά διανύσματα αναζήτησης η χρονική βελτίωση της διαδικασίας είναι ουσιαστική, καθώς ελέγχεται η ομοιότητα μόνο με εικόνες που περιέχουν λέξεις που εμπεριέχει η εικόνα αναζήτησης.

Κατά την ανάκτηση εικόνων, οι εικόνες της βάσης κατατάσσονται (κατάταξη (ranking)) με βάση το κανονικοποιημένο, κλιμακωτό γινόμενο (τη γωνία του συνημιτόνου) του διανύσματος V_q της εικόνας αναζήτησης και όλων των διανυσμάτων V_d των εικόνων της βάσης.

Σχήμα 2.5: Οι οπτικές λέξεις δεν αντιπροσωπεύουν άμεσα χαρακτηριστικά μέρη του προσώπου, αλλά αποτελούν μια αφηρημένη περιγραφή επί χαρακτηριστικών, τα οποία έχουν εξαχθεί από την εικόνα και τα οποία βρίσκονται κοντά με βάση μια προσέγγιση συσταδοποίησης, αλλά απουσία ενσωμάτωσης κάποιας χωρικής πληροφορίας της εικόνας .

2.3.3 Χωρικό ταίριασμα και ανακατάταξη εικόνων

Στο προηγούμενο στάδιο της διαδικασίας ανάκτησης εικόνων, είχαμε θεωρήσει τα features που περιγράφουν τις εικόνες ως συλλογή τοπικών χαρακτηριστικών, απουσία οποιασδήποτε χωρικής διάταξης (σχήμα 2.5). Για το λόγο αυτό οι *εικόνες με υψηλή κατάταξη (top ranked)*, υπόκεινται σε *ανακατάταξη (re-ranking)*, με τη χρήση χωρικών περιορισμών. Η διαδικασία της χωρικής επαλήθευσης εκτιμά έναν μετασχηματισμό μεταξύ της περιοχής της εικόνας αναζήτησης και της κάθε εικόνας βάσης προς ανακατάταξη, βασιζόμενο στο πόσο καλά οι περιοχές των χαρακτηριστικών προβλέπονται από τον εκτιμώμενο μετασχηματισμό. Κατά αυτό τον τρόπο, οι προκείμενες εικόνες βάσης ανακατατάσσονται με βάση τη διακριτική ικανότητα



Σχήμα 2.6: **Δεξιά** Το σύνολο των αντιστοιχίσεων μεταξύ των χαρακτηριστικών των δύο εικόνων χωρίς χωρική πληροφορία. **Αριστερά** Το αποτέλεσμα της χωρικής επαλήθευσης, στο συγκεκριμένο παράδειγμα μετά την εφαρμογή του αλγορίθμου LO-RANSAC 4.2.2.

των χωρικά επαληθευμένων οπτικών λέξεων. Η παραπάνω διαδικασία χωρικής επαλήθευσης περιλαμβάνει και το χωρικό ταίριασμα εικόνων.

Έχουν αναπτυχθεί πληθώρα προσεγγίσεων όσον αφορά το χωρικό ταίριασμα εικόνων, ιδίως σε μηχανές αναζήτησης μεγάλης κλίμακας, στην κατεύθυνση κάλυψης διαφόρων επιδιώξεων για τη διατήρηση του αναλλοίωτου, καθολική γεωμετρική επαλήθευση και χαμηλές απαιτήσεις σε χώρο. Οι μέθοδοι αυτές εφαρμόζονται είτε απευθείας στον χώρο των οπτικών λέξεων, είτε σε ζεύγη *αντιστοιχιών (correspondences)* χαρακτηριστικών, δεδομένης μιας λογικής *ένα προς ένα αντιστοίχισης (one to one mapping)* των χαρακτηριστικών, είτε σε κάποιο χώρο καθολικού μετασχηματισμού ενός πλέγματος της εικόνας, το οποίο ενσωματώνει και πληροφορία αναφορικά με το σχήμα των τοπικών χαρακτηριστικών, όπως κλίμακα και προσανατολισμός. Αποτέλεσμα του χωρικού ταιριάσματος είναι η εξαγωγή *μεμονωμένων αντιστοιχιών (single correspondences)* με μεγάλο βαθμό ομοιότητας (σχήμα 2.6). Θα αναφερθούμε εκτενώς σε κάποιες μεθόδους χωρικού ταιριάσματος στο επόμενο κεφάλαιο.

2.3.4 Χωρικό ταίριασμα και ΜΔΥ

Τα στάδια της ανάκτησης εικόνων που αναφέρθηκαν στις ενότητες 2.3.1 και 2.3.2 έχουν ενσωματωθεί σε προηγούμενες εργασίες στο πλαίσιο εκμάθησης με SVM ([8] και [14]) και αποτελούν παράλληλα τη βάση σύγκρισης (baseline) για τη μέθοδο που εισάγουμε. Η *νέα ιδέα* είναι να εκμεταλλευτούμε και το τελευταίο στάδιο της χωρικής επαλήθευσης (ενότητα 2.3.3), αξιοποιώντας και την χωρική πληροφορία για την αποτελεσματική κατηγοριοποίηση εικόνων. Απαραίτητη προϋπόθεση στην κατεύθυνση αυτή αποτελεί η απόδειξη ότι η διαδικασία χωρικού ταιριάσματος που επιθυμούμε να ενσωματώσουμε στα SVM παράγει τιμές πυρήνα, δηλαδή εσωτερικού γινομένου δύο διανυσμάτων, που περιγράφουν τις εκάστοτε δύο εικόνες των οποίων υπολογίζουμε την ομοιότητα. Ένα επιπλέον στοιχείο που πρέπει να λάβουμε υπόψιν είναι η διαδικασία χωρικού ταιριάσματος να είναι γρήγορη και αποδοτική και να μην επιβαρύνει την όλη διαδικασία εκμάθησης σε κάποιο επίπεδο. Αυτά τα δύο στοιχεία, 1) η απόδειξη της διαδικασίας χωρικού ταιριάσματος ως πυρήνα και 2) η απαίτηση για ταχύτητα και απόδοση θα μας καθοδηγήσουν στην σύνθεση της μεθόδου, αλλά και την ανάλυσή της στη συνέχεια.

Κεφάλαιο 3

Μηχανές διανυσμάτων υποστήριξης

3.1 Εισαγωγή

Οι *μηχανές διανυσμάτων υποστήριξης (ΜΔΥ) (support vector machines)* (SVM) αποτελούν συστήματα εκμάθησης με χώρο υποθέσεων που περιλαμβάνει γραμμικές συναρτήσεις, δέχονται ως εισόδους δείγματα από χώρους χαρακτηριστικών πολλών διαστάσεων και εκπαιδεύονται με βάση έναν αλγόριθμο εκμάθησης στο πλαίσιο της θεωρίας βελτιστοποίησης. Ο αλγόριθμος αυτός υλοποιεί μία *δεδομένη προκατάληψη εκμάθησης (learning bias)*, η οποία εξάγεται με βάση τη θεωρία στατιστικής εκμάθησης. Τα SVM αναπτύχθηκαν ως στρατηγική εκμάθησης από τον Vapnik και τους συνεργάτες του.

Συγκεκριμένα, όσον αφορά στην *ταξινόμηση διανυσμάτων υποστήριξης (support vector classification)*, βασική λειτουργία των SVM είναι η εύρεση υπολογιστικά αποδοτικού τρόπου εκμάθησης *υπερεπιπέδων (hyperplanes)* "καλού" διαχωρισμού, σε χώρους χαρακτηριστικών πολλών διαστάσεων, όπου ως υπερεπίπεδα "καλού" διαχωρισμού θεωρούμε αυτά που βελτιστοποιούν τα *όρια γενίκευσης (generalisation bounds)* και ως υπολογιστικά αποδοτικούς αλγορίθμους εκμάθησης, αυτούς που ανταποκρίνονται σε μεγέθη συνόλων δεδομένων της τάξης των 100000 δειγμάτων.

Τα SVM αντιμετωπίζουν επιτυχώς τα προβλήματα αποδοτικότητας όσον αφορά στην εκπαίδευση, στον *έλεγχο νέων δεδομένων (testing)*, στο *υπερταίριασμα (overfitting)* των μηχανών εκμάθησης και στην αναξιοπιστία των εκάστοτε *χειρισμών (heuristics)*, όπως προκύπτει από τα παρακάτω:

- Λόγω των συνθηκών του Mercer για τους *πυρήνες (kernels)* που χρησιμοποιούνται στην εκμάθηση, όπως θα εξηγήσουμε παρακάτω, τα αντίστοιχα προβλήματα βελτιστοποίησης είναι κυρτά και συνεπώς δεν εμφανίζονται τοπικά ελάχιστα και κάθε τοπική λύση αποτελεί και ολικό βέλτιστο (global optimum). Επομένως, εξασφαλίζεται η ύπαρξη λύσεων ακόμη και για μεγάλα σύνολα δεδομένων εκπαίδευσης.

- Η δυνατότητα χρήσης *συναρτήσεων πυρήνων (kernel functions)* αποτελεί το κλειδί για την αποδοτική χρήση των SVM σε χώρους πολλών διαστάσεων.

- Η θεωρία γενίκευσης που εφαρμόζουν, εξασφαλίζει τον έλεγχο της χωρητικότητας και συνεπώς την αποφυγή του overfitting, το οποίο είναι εγγενές σε χώρους πολλών διαστάσεων, ελέγχοντας τα μέτρα του *περιθωρίου (margin)* των υπερεπιπέδων διαχωρισμού, ενώ παράλληλα η θεωρία βελτιστοποίησης παρέχει τις μαθηματικές τεχνικές που απαιτούνται για την εύρεση των υπερεπιπέδων που βελτιστοποιούν αυτά τα μέτρα.

- Παρέχουν *συμπαγείς και αραιές δυικές αναπαραστάσεις (compact and sparse dual representations)* της εκάστοτε υπόθεσης, μειώνοντας των αριθμό των παραμέτρων εκμάθησης, ευνοώντας την εξαγωγή αποδοτικών αλγορίθμων εκμάθησης. Αυτό οφείλεται στις *Karush-Kuhn-Tucker συνθήκες* που ικανοποιεί η λύση, όπως θα αναλύσουμε παρακάτω.

3.2 Γραμμική ταξινόμηση

Έστω $X \subseteq \mathbb{R}^n$ ο *χώρος εισόδου (input space)* και Y ο αντίστοιχος *χώρος εξόδου (output domain)*, που στη γενική περίπτωση της ταξινόμησης πολλών κλάσεων (multiclass classification) είναι $Y = \{1, 2, \ldots, m\}$, ενώ στη δυαδική περίπτωση συχνά ορίζεται ως $Y = \{-1, 1\}$.

Ορισμός 3.2.1. *Ένα* σύνολο δεδομένων εκπαίδευσης (training set) *ορίζεται ως*

$$S = ((\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)) \subseteq (X \times Y)^n, \tag{3.1}$$

όπου n ο αριθμός των δειγμάτων. Τα $\mathbf{x}_i$ αναφέρονται ως δείγματα (examples/instances) *και τα y_i ως* labels (ετικέτες).

Να σημειώσουμε ότι τα διανύσματα εισόδου, καθώς και τα διανύσματα βαρών θα είναι διανύσματα στήλη.

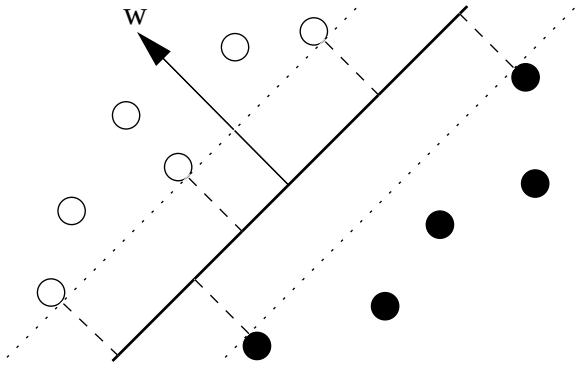
Στην περίπτωση της δυαδικής ταξινόμησης, ορίζουμε μια πραγματική συνάρτηση $f : X \to Y$, έτσι ώστε η είσοδος $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)'$ κατατάσσεται στη θετική κλάση, αν $f(\mathbf{x}) \geq 0$, διαφορετικά στην αρνητική κλάση, δηλαδή σε κάθε περίπτωση ισχύει: $yf(\mathbf{x}) > 0$, όπου $y \in Y$. Στην γραμμική περίπτωση λοιπόν θεωρούμε τη συνάρτηση

$$\begin{aligned} f(\mathbf{x}) &= \langle \mathbf{w}, \mathbf{x} \rangle + b \\ &= \sum_{i=1}^{n} \mathbf{w}_i \mathbf{x}_i + b, \end{aligned} \tag{3.2}$$

όπου $(\mathbf{w}, b) \in \mathbb{R}^n \times \mathbb{R}$ το *διάνυσμα βαρών (weight vector)* και το bias, οι παράμετροι που ελέγχουν τη συνάρτηση απόφασης και ο *κανόνας απόφασης (decision rule)* ορίζεται από το πρόσημο $sgn(f(\mathbf{x}))$, όπου υποθέτουμε $sgn(0) = 1$. Κάποιες φορές το $-b$ συμβολίζεται με ϑ, ποσότητα γνωστή ως *κατώφλι (threshold)*.

Η γεωμετρική ερμηνεία αυτού του τύπου υπόθεσης είναι ότι, ο χώρος εισόδων χωρίζεται σε δύο μέρη από έναν αφινικό υποχώρο διαστάσεων $n-1$, ένα *υπερεπίπεδο (hyperplane)* που ορίζεται από την εξίσωση: $f(\mathbf{x}) = 0$. Τα δύο μέρη του χώρου εισόδων που δημιουργούνται, αντιστοιχούν στις εισόδους των δύο διακριτών κλάσεων (Σχήμα 3.1). Το διάνυσμα $\mathbf{w}$ ορίζει μια κατεύθυνση κάθετη στο υπερεπίπεδο, ενώ μεταβολή του b συνεπάγεται παράλληλη μετακίνηση του υπερεπιπέδου.

Να σημειώσουμε ότι η γραμμική αυτή υπόθεση (3.2) εμπεριέχει έναν εγγενή βαθμό ελευθερίας, δεδομένου ότι η συνάρτηση δεν αλλάζει αν κλιμακοποιήσουμε το υπερεπίπεδο ως $(\lambda\mathbf{w}, \lambda b)$ για κάποιο $\lambda \in \mathbb{R}^+$.



Σχήμα 3.1: Ένα υπερεπίπεδο διαχωρισμού $(\mathbf{w}, b)$ για ένα σύνολο εκπαίδευσης δύο διαστάσεων. Το υπερεπίπεδο αποτελεί η μαύρη γραμμή με τη θετική περιοχή στην πάνω πλευρά και την αρνητική στην κάτω, ενώ η απόσταση των διακεκομμένων γραμμών από το υπερεπίπεδο υποδεικνύει το μέγεθος της παραμέτρου b [18].

Στη συνέχεια παραθέτουμε κάποιους ορισμούς που θα χρησιμοποιηθούν ευρύτερα και στη συνέχεια.

Ορισμός 3.2.2. *Ορίζουμε* συναρτησιακό περιθώριο(functional margin) *ενός δείγματος* $(\mathbf{x}_i, y_i)$, *όσον αφορά στο υπερεπίπεδο* $(\mathbf{w}, b)$ *την ποσότητα*

$$\gamma_i = y_i(\langle \mathbf{w}, \mathbf{x}_i \rangle + b), \tag{3.3}$$

όπου $\gamma_i > 0$ *συνεπάγεται σωστή ταξινόμηση του δείγματος.*

Αν αντικαταστήσουμε το συναρτησιακό περιθώριο με το *γεωμετρικό (geometric)*, εξάγουμε την ισοδύναμη ποσότητα για την κανονικοποιημένη γραμμική συνάρτηση $(\frac{1}{\|\mathbf{w}\|}, \frac{b}{\|\mathbf{w}\|})$, η όποια εκφράζει την Ευκλείδια απόσταση των σημείων από το *όριο απόφασης (decision boundary)* στο χώρο εισόδου.

Ορισμός 3.2.3. *Ορίζουμε* περιθώριο (margin) ενός συνόλου δεδομένων εκπαίδευσης *την ποσότητα*

$$\gamma = \arg\max_{\mathbf{w},b} \left(\frac{1}{\|\mathbf{w}\|} \min_i \gamma_i \right), \tag{3.4}$$

δηλαδή το μέγιστο γεωμετρικό περιθώριο όλων των υπερεπιπέδων και θα είναι θετικό για γραμμικά διαχωρίσιμα (linearly separable) *σύνολα δεδομένων εκπαίδευσης, δηλαδή μόνο αν υπάρχει υπερεπίπεδο που χωρίζει σωστά τα δεδομένα. Το υπερεπίπεδο το οποίο εξάγει αυτό το περιθώριο ονομάζεται* υπερεπίπεδο μέγιστου περιθωρίου (maximal margin hyperplane).

Να σημειώσουμε ότι, προκειμένου να εκπαιδεύσουμε μη γραμμικές σχέσεις με μια γραμμική μηχανή εκπαίδευσης, θα πρέπει ανάλογα με τη 3.2, να χρησιμοποιήσουμε υποθέσεις της μορφής

$$f(\mathbf{x}) = \sum_{i=1}^{n} \mathbf{w}_i \phi(\mathbf{x}_i) + b, \tag{3.5}$$

όπου ϕ μια μη γραμμική απεικόνιση από το χώρο εισόδου $\mathbf{x}$ σε ένα χώρο χαρακτηριστικών εσωτερικού γινομένου.

3.3 Πυρήνες

Όπως προαναφέραμε στην εισαγωγή του κεφαλαίου, ένα σημαντικό χαρακτηριστικό των SVM είναι η δυνατότητα χρήσης πυρήνων. Οι *αναπαραστάσεις πυρήνων (kernel representations)* παρέχουν μια εναλλακτική για την εξαγωγή πιο πολύπλοκων υποθέσεων από τις γραμμικές συναρτήσεις σχετικά με τα προβλήματα εκμάθησης, προβάλλοντας τα δεδομένα σε χώρους χαρακτηριστικών πολλών διαστάσεων, αυξάνοντας έτσι την υπολογιστική δύναμη των γραμμικών μηχανών εκμάθησης. Παράλληλα, η μέθοδος των πυρήνων παρέχει την αποσύνδεση των αλγορίθμων εκμάθησης και της θεωρίας από τις ιδιαιτερότητες των εφαρμογών, καθώς ανάγει την εκμάθηση στο σχεδιασμό μιας κατάλληλης συνάρτησης πυρήνα, υπό τη συνθήκη ότι ο πυρήνας υπολογίζει το εσωτερικό γινόμενο των διανυσμάτων χαρακτηριστικών που αντιστοιχούν σε δύο εισόδους. Στη συνέχεια παραθέτουμε κάποιους ορισμούς και θεωρήματα, όσον αφορά τη θεωρία πυρήνων.

Οι ποσότητες που χρησιμοποιούμε για να περιγράψουμε τα δεδομένα ονομάζονται *χαρακτηριστικά (features)*, ενώ τα αρχικά δεδομένα συχνά αναφέρονται ως *γνωρίσματα (attributes)*. Η διαδικασία επιλογής της πιο κατάλληλης αναπαράστασης, ονομάζεται *επιλογή χαρακτηριστικών (feature selection)*. Ο χώρος $F = \{\phi(\mathbf{x}) : \mathbf{x} \in X\}$ ονομάζεται *χώρος χαρακτηριστικών (feature space)*.

Αναφορικά με την επιλογή χαρακτηριστικών, συχνά ψάχνουμε να προσδιορίσουμε το μικρότερο δυνατό σύνολο χαρακτηριστικών, το οποίο εμπεριέχει όλη την απαραίτητη πληροφορία για τα attributes. Η διαδικασία αυτή αναφέρεται ως *μείωση διαστατικότητας (dimensionality reduction)* και μπορεί να αποβεί πολύ σημαντική, καθώς η υπολογιστική επίδοση και η επίδοση γενίκευσης χειροτερεύουν όσο ο αριθμός των χαρακτηριστικών αυξάνει, φαινόμενο γνωστό ως *κατάρα της διαστατικότητας (curse of dimensionality)*.

Ορισμός 3.3.1. Πυρήνας (kernel) *ονομάζεται μια συνάρτηση τέτοια ώστε* $\forall \mathbf{x}, \mathbf{x}' \in X$

$$k(\mathbf{x}, \mathbf{x}') = \langle \phi(\mathbf{x}), \phi(\mathbf{x}') \rangle, \tag{3.6}$$

όπου ϕ μια απεικόνιση από το χώρο εισόδου X σε ένα χώρο χαρακτηριστικών εσωτερικού γινομένου F.

Από τον παραπάνω ορισμό βλέπουμε ότι ο πυρήνας είναι μια συμμετρική συνάρτηση ως προς τα ορίσματά της, έτσι ώστε $k(\mathbf{x}, \mathbf{x}') = k(\mathbf{x}', \mathbf{x})$. Υπάρχουν πολλές κατηγορίες πυρήνων. Η πιο απλή περίπτωση είναι αυτή του *γραμμικού πυρήνα (linear kernel)*, όπου $\phi(\mathbf{x}) = \mathbf{x}$ και $k(\mathbf{x}, \mathbf{x}') = \mathbf{x}^T \mathbf{x}'$. Ενδεικτικά αναφέρουμε τους *στάσιμους πυρήνες (stationary kernels)*, όπου $k(\mathbf{x}, \mathbf{x}') = k(\mathbf{x} - \mathbf{x}')$, οι οποίοι είναι ανεξάρτητοι από μεταφορές (translations) στο χώρο εισόδου και τους *ομογενείς πυρήνες (homogeneous kernels)* ή διαφορετικά *ακτινικές συναρτήσεις βάσης (radial basis functions)*, όπου $k(\mathbf{x}, \mathbf{x}') = k(\|\mathbf{x} - \mathbf{x}'\|)$.

Με βάση τις ιδιότητες των χώρων εσωτερικού γινομένου και δεδομένου του ορισμού 3.3.1, μια συνάρτηση αποτελεί πυρήνας αν ικανοποιεί τις παρακάτω συνθήκες

$$\begin{aligned} k(\mathbf{x},\mathbf{x}') &= \langle\phi(\mathbf{x}),\phi(\mathbf{x}')\rangle = \langle\phi(\mathbf{x}'),\phi(\mathbf{x})\rangle = k(\mathbf{x}',\mathbf{x}) \\ k(\mathbf{x},\mathbf{x}')^2 &= \langle\phi(\mathbf{x}),\phi(\mathbf{x}')\rangle^2 \leq \|\phi(\mathbf{x})\|^2\|\phi(\mathbf{x}')\|^2 \\ &= \langle\phi(\mathbf{x}),\phi(\mathbf{x})\rangle\langle\phi(\mathbf{x}'),\phi(\mathbf{x}')\rangle = k(\mathbf{x},\mathbf{x})k(\mathbf{x}',\mathbf{x}') \end{aligned} \tag{3.7}$$

Δηλαδή θα πρέπει η $k(\mathbf{x},\mathbf{x}')$ να είναι συμμετρική και επιπλέον ο πίνακας Gram K με στοιχεία $k(x_i,x_j)$ να είναι θετικά ημιορισμένος για όλες τις δυνατές επιλογές του συνόλου $\{x_i\}$, ώστε να ισχύει και η δεύτερη συνθήκη, της ανισότητας Cauchy-Schwartz. Πρακτικά θα πρέπει όλα τα στοιχεία του πίνακα Gram να είναι μη αρνητικά. Η συνθήκη αυτή συνοψίζεται και στο θεώρημα Mercer (Θεώρημα 3.6 [7]).

Μπορούμε να κατασκευάσουμε πυρήνες εύκολα χρησιμοποιώντας απλούστερους πυρήνες ως δομικά συστατικά, χρησιμοποιώντας τις παρακάτω ιδιότητες [4].

Ορισμός 3.3.2. *Δοθέντων των πυρήνων $k_1(\mathbf{x},\mathbf{x}')$, $k_2(\mathbf{x},\mathbf{x}')$ ορίζονται οι παρακάτω πυρήνες:*

$$k(\mathbf{x},\mathbf{x}') = ck_1(\mathbf{x},\mathbf{x}') \tag{3.8}$$
$$k(\mathbf{x},\mathbf{x}') = f(\mathbf{x})k_1(\mathbf{x},\mathbf{x}')f(\mathbf{x}') \tag{3.9}$$
$$k(\mathbf{x},\mathbf{x}') = q(k_1(\mathbf{x},\mathbf{x}')) \tag{3.10}$$
$$k(\mathbf{x},\mathbf{x}') = exp(k_1(\mathbf{x},\mathbf{x}')) \tag{3.11}$$
$$k(\mathbf{x},\mathbf{x}') = k_1(\mathbf{x},\mathbf{x}') + k_2(\mathbf{x},\mathbf{x}') \tag{3.12}$$
$$k(\mathbf{x},\mathbf{x}') = k_1(\mathbf{x},\mathbf{x}')k_2(\mathbf{x},\mathbf{x}') \tag{3.13}$$
$$k(\mathbf{x},\mathbf{x}') = k_3(\phi(\mathbf{x}),\phi(\mathbf{x}')) \tag{3.14}$$
$$k(\mathbf{x},\mathbf{x}') = \mathbf{x}A\mathbf{x}' \tag{3.15}$$
$$k(\mathbf{x},\mathbf{x}') = k_\alpha(\mathbf{x}_\alpha,\mathbf{x}'_\alpha) + k_b(\mathbf{x}_b,\mathbf{x}'_b) \tag{3.16}$$
$$k(\mathbf{x},\mathbf{x}') = k_\alpha(\mathbf{x}_\alpha,\mathbf{x}'_\alpha)k_b(\mathbf{x}_b,\mathbf{x}'_b), \tag{3.17}$$

όπου $c>0$ μια σταθερά, f μια συνάρτηση, q ένα πολυώνυμο με μη αρνητικούς συντελεστές, $\phi(\mathbf{x})$ είναι μια συνάρτηση του $\mathbf{x}$ στον $\mathbb{R}^m$, $k_3(\mathbf{x},\mathbf{x}')$ ένας πυρήνας στον $\mathbb{R}^m$, A ένας συμμετρικός, θετικά ημιορισμένος πίνακας, $\mathbf{x}_\alpha$ και $\mathbf{x}_b$ είναι μεταβλητές (όχι αναγκαστικά ανεξάρτητες) με $\mathbf{x}=(\mathbf{x}_\alpha,\mathbf{x}_b)$ και k_α,k_b πυρήνες στους αντίστοιχους χώρους.

Στη συνέχεια θα δείξουμε ότι, χρησιμοποιώντας πυρήνες, μπορούμε να απεικονίσουμε τα δεδομένα σε ένα χώρο χαρακτηριστικών, όπου εξασφαλίζεται η γραμμική διαχωρισιμότητα και επομένως μπορούμε να χρησιμοποιήσουμε γραμμικές μηχανές εκπαίδευσης.

3.4 Ταξινομητής μεγίστου περιθωρίου

Ο *ταξινομητής μεγίστου περιθωρίου (maximal margin classifier)* δουλεύει μόνο για δεδομένα γραμμικά διαχωρίσιμα στο χώρο των χαρακτηριστικών. Παρόλα αυτά αναφερόμαστε σε αυτόν, καθώς αποτελεί δομικό συστατικό πιο πολύπλοκων SVM ταξινομητών, τους οποίους θα εξετάσουμε στη συνέχεια.

Ο ταξινομητής μεγίστου περιθωρίου βελτιστοποιεί ένα *σφάλμα γενίκευσης (generalisation error)* (Θεώρημα 4.18 [7]), το οποίο εμπεριέχει το γεωμετρικό margin, χωρίζοντας τα δεδο-

μένα με το υπερεπίπεδο μεγίστου περιθωρίου και ταυτόχρονα το όριο απόφασης που παράγει δεν εξαρτάται από τη διάσταση του χώρου εισόδων.

Αν $\mathbf{w}$, το διάνυσμα βαρών που παράγει συναρτησιακό περιθώριο ίσο με 1 στο σημείο της θετικής κλάσης $\mathbf{x}_{+1}$ και στο σημείο της αρνητικής κλάσης $\mathbf{x}_{-1}$, με βάση την υπόθεση 3.2 θα έχουμε

$$\begin{aligned} \langle \mathbf{w}, \phi(\mathbf{x}_{+1}) \rangle + b = 1 \\ \langle \mathbf{w}, \phi(\mathbf{x}_{-1}) \rangle + b = -1 \end{aligned} \tag{3.18}$$

Κανονικοποιώντας το διάνυσμα $\mathbf{w}$ εκφράζουμε το γεωμετρικό margin γ ως

$$\begin{aligned} \gamma &= \frac{1}{2}\left(\left\langle \frac{\mathbf{w}}{\|\mathbf{w}\|_2}, \phi(\mathbf{x}_{+1}) \right\rangle - \left\langle \frac{\mathbf{w}}{\|\mathbf{w}\|_2}, \phi(\mathbf{x}_{-1}) \right\rangle \right) \\ &= \frac{1}{2\|\mathbf{w}\|_2}(\langle \mathbf{w}, \phi(\mathbf{x}_{+1}) \rangle - \langle \mathbf{w}, \phi(\mathbf{x}_{-1}) \rangle) \\ &= \frac{1}{\|\mathbf{w}\|_2} \end{aligned} \tag{3.19}$$

Ορισμός 3.4.1. *Δεδομένου ενός γραμμικά διαχωρίσιμου συνόλου δεδομένων εκπαίδευσης S (3.1), το υπερεπίπεδο $(\mathbf{w}, b)$ αποτελεί λύση του προβλήματος βελτιστοποίησης*

$$\begin{aligned} \min_{\mathbf{w},b} \quad & \frac{1}{2}\langle \mathbf{w}, \mathbf{w} \rangle \\ \text{subject to} \quad & y_i(\langle \mathbf{w}, \phi(\mathbf{x}_i) \rangle + b) \geq 1 \\ & 1 \leq i \geq l \end{aligned} \tag{3.20}$$

Σύμφωνα με τη θεωρία βελτιστοποίησης (Κεφάλαιο 5,[7]), η πρωτογενής (primal) Lagrangian είναι:

$$L(\mathbf{w}, b, \alpha) = \frac{1}{2}\langle \mathbf{w}, \mathbf{w} \rangle - \sum_{i=1}^{l} \alpha_i [y_i(\langle \mathbf{w}, \phi(\mathbf{x}_i) \rangle + b) - 1] \tag{3.21}$$

όπου $\alpha_i \geq 0$ οι πολλαπλασιαστές Lagrange.

Όσον αφορά το ισοδύναμο δυϊκό πρόβλημα, παραγωγίζοντας τη 3.21 ως προς $\mathbf{w}$ και b έχουμε

$$\begin{aligned} \mathbf{w} &= \sum_{i=1}^{l} y_i \alpha_i \phi(\mathbf{x}_i) \\ 0 &= \sum_{i=1}^{l} y_i \alpha_i \end{aligned} \tag{3.22}$$

Με βάση τις παραπάνω σχέσεις, η πρωτογενής Lagrangian γίνεται

$$L(\mathbf{w}, b, \alpha) = \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{l} y_i y_j \alpha_i \alpha_j \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle \tag{3.23}$$

Επομένως προκύπτει το ισοδύναμο δυϊκό πρόβλημα.

Ορισμός 3.4.2. *Δεδομένου ενός γραμμικά διαχωρίσιμου συνόλου δεδομένων εκπαίδευσης S (3.1), και ότι οι παράμετροι α^* αποτελούν λύση του τετραγωνικού προβλήματος βελτιστοποίησης*

$$\begin{aligned} \max_{\alpha} \; & W(\alpha) = \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{l} y_i y_j \alpha_i \alpha_j k(\mathbf{x}_i, \mathbf{x}_j) \\ \text{subject to } & \sum_{i=1}^{l} y_i \alpha_i = 0 \\ & \alpha_i \geq 0, i = 1 \dots l, \end{aligned} \tag{3.24}$$

όπου $k(\mathbf{x}_i, \mathbf{x}_j) = \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle$, το διάνυσμα βαρών $\mathbf{w}^* = \sum_{i=1}^{l} y_i \alpha_i^* \phi(\mathbf{x}_i)$ παράγει το υπερεπίπεδο μεγίστου περιθωρίου με γεωμετρικό περιθώριο

$$\gamma = \frac{1}{\|\mathbf{w}^*\|_2} \tag{3.25}$$

Η τιμή του b προσδιορίζεται με βάση το πρωτογενές πρόβλημα ως

$$b^* = -\frac{\max_{y_i=-1} \langle \mathbf{w}_*, \phi(\mathbf{x}_i) \rangle + min_{y_i=1} \langle \mathbf{w}_*, \phi(\mathbf{x}_i) \rangle}{2} \tag{3.26}$$

Με βάση το θεώρημα 5.21 και από τις συνθήκες Karush-Kuhn-Tucker (Κεφάλαιο 5.2 [7]) προκύπτει ότι η βέλτιστη λύση $\alpha^*, (\mathbf{w}^*, b^*)$ θα πρέπει να ικανοποιεί τη σχέση

$$\alpha_i^* [y_i (\langle \mathbf{w}^*, \phi(\mathbf{x}_i) \rangle + b^*) - 1] = 0 \tag{3.27}$$

Η παραπάνω σχέση είναι πολύ σημαντική, καθώς ορίζει τα *διανύσματα υποστήριξης (support vectors (SVs))* ως τα σημεία για τα οποία το συναρτησιακό περιθώριο είναι ίσο με 1 και γι αυτό βρίσκονται πιο κοντά στο υπερεπίπεδο διαχωρισμού, καθώς και ότι τα αντίστοιχα σε αυτά α_i του δυϊκού προβλήματος είναι τα μόνα μη μηδενικά. Αυτό σημαίνει ότι αυτά τα σημεία εμπεριέχουν όλη τη σημαντική πληροφορία για την ευστάθεια της βέλτιστης λύσης.

Με βάση τα παραπάνω το γεωμετρικό margin της σχέσης 3.25 διαμορφώνεται ως

$$\gamma = \left(\sum_{i \in sv} \alpha_i^* \right)^{-\frac{1}{2}} \tag{3.28}$$

Από τη σχέση 3.24 παρατηρούμε ότι, τα δεδομένα εκπαίδευσης εμφανίζονται στη λύση μόνο στον πυρήνα . Επομένως, η δυϊκή αναπαράσταση απαλλάσσει την εξαγωγή της βέλτιστης λύσης από τη δυσκολία που εισάγει η διάσταση του χώρου των χαρακτηριστικών, παρέχοντας τη δυνατότητα χρήσης του λεγόμενου "κόλπου πυρήνα" (kernel trick) ή υποκατάσταση πυρήνα (kernel substitution). Αν λοιπόν το διάνυσμα εισόδου $\mathbf{x}$ εμφανίζεται στη λύση μόνο με τη μορφή εσωτερικών γινομένου, μπορούμε να αντικαταστήσουμε το γινόμενο με κάποιο κατάλληλα επιλεγμένο πυρήνα.

Με βάση την παραπάνω ανάλυση, μπορούμε να εξάγουμε τα παρακάτω συμπεράσματα αναφορικά με τα μεγίστου περιθωρίου SVM:

- Το margin έχει διπλό ρόλο, αφενός η μεγιστοποίηση του εξασφαλίζει *μικρή διάσταση θρυμματισμού (low fat-shattering)*, δηλαδή σύμφωνα με τη θεωρία γενίκευσης απαιτείται ο προσδιορισμός ενός μικρού υποσυνόλου σημείων για το διαχωρισμό των δεδομένων με δεδομένο περιθώριο, και αφετέρου αυτό οδηγεί στην εξαγωγή αραιής λύσης μέσω τον περιορισμών που επιβάλλει σε αυτήν.
- Δεδομένων των SV, μπορούμε να ανακατασκευάσουμε το υπερεπίπεδο μεγίστου περιθωρίου, το οποίο ταξινομεί σωστά όλο το σύνολο των δεδομένων εκπαίδευσης, συμπέρασμα που βασίζεται στη λεγόμενη *αραιότητα (sparseness)* της λύσης (ένα υποσύνολο μόνο των πολ/στών Lagrange είναι μη μηδενικό).
- Με την κατάλληλη επιλογή πυρήνα συνήθως μπορούμε να καταστήσουμε τα δεδομένα διαχωρίσιμα. Εντούτοις, αυτό μπορεί να οδηγήσει σε overfitting, ειδικά παρουσία θορύβου στα δεδομένα εκπαίδευσης. Σε αυτήν την περίπτωση, όμως επειδή τυχόν σημεία με μεγάλες τιμές/θορυβώδη (outliers) θα χαρακτηρίζονται από μεγάλους πολ/στες Lagrange ως δύσκολα στην εκπαίδευση, θα μπορούσαμε να χρησιμοποιήσουμε την όλη διαδικασία για αποθορυβοποίηση των δεδομένων.

3.5 Ταξινομητής χαλαρού περιθωρίου

3.5.1 C-ΜΔΥ

Ο ταξινομητής μεγίστου περιθωρίου που εξετάσαμε στην προηγούμενη ενότητα, παρέχει ακριβή διαχωρισμό των δεδομένων εκπαίδευσης στον αρχικό χώρο εισόδου, εξασφαλίζοντας γραμμική διαχωρισιμότητα των δεδομένων σε κάποιο χώρο χαρακτηριστικών με τη χρήση πυρήνων, ακόμη κι αν τελικά το όριο απόφασης είναι μη γραμμικό. Στην πράξη όμως, ενδέχεται οι κατανομές των κλάσεων των δεδομένων να επικαλύπτονται και ακριβής διαχωρισμός των δεδομένων εκπαίδευσης να οδηγεί σε "φτωχή"/ανεπαρκή γενίκευση (poor generalisation). Για το λόγο αυτό τροποποιούμε τα SVM, ώστε να επιτρέπεται κάποια σημεία να μην ταξινομούνται σωστά, αλλά με κάποια *ποινή (penalty)*, η οποία όμως αυξάνεται γραμμικά με την απόσταση από το υπερεπίπεδο διαχωρισμού. Έτσι προκύπτει ο *ταξινομητής χαλαρού περιθωρίου (soft margin classifier)*.

Προκειμένου να παραβιάζονται οι περιορισμοί που θέτει το περιθώριο, εισάγουμε *χαλαρές μεταβλητές (slack variables)* (Bennett, 1992 ,Cortes and Vapnik, 1995) για τις οποίες ισχύει

$$\xi_i = |y_i - f(\mathbf{x}_i)|, i = 1 \dots l \tag{3.29}$$

Με βάση την παραπάνω σχέση, $\xi_i = 0$ για τα σημεία εντός ή επί του συνόρου επιτρεπτού margin, $\xi_i = 1$ για τα σημεία επί του συνόρου διαχωρισμού και $\xi_i \geq 1$ για τα λάθος ταξινομημένα. Δηλαδή η συνθήκη για τα επιτρεπόμενα σημεία εντός του περιθωρίου είναι: $0 < \xi_i \leq 1$.

Εισάγοντας τις μεταβλητές ξ στο πρόβλημα βελτιστοποίησης 3.20 έχουμε

$$\begin{aligned} \min_{\xi,\mathbf{w},b} \ & \frac{1}{2}\langle \mathbf{w}, \mathbf{w} \rangle + C \sum_{i=1}^{l} \xi_i \\ \textit{subject to} \ & y_i(\langle \mathbf{w}\phi(\mathbf{x}_i) \rangle + b) \geq 1 - \xi_i, i = 1 \dots l \\ & \xi_i \geq 0, i = 1 \dots l \end{aligned} \tag{3.30}$$

Για ένα δεδομένο πρόβλημα, η επιλογή μιας τιμής για τη μεταβλητή C, ισοδυναμεί με την επιλογή μιας τιμής για το $\|\mathbf{w}\|_2$ και στη συνέχεια την ελαχιστοποίηση του $\|\xi\|_2$ για το δεδομένο $\mathbf{w}$. Λόγω της παρουσίας και του ρόλου αυτής της παραμέτρου θα αναφερόμαστε στο εξής στη συγκεκριμένη κατηγορία SVM ως *C-ΜΔΥ* (C-SVM).

Να σημειώσουμε ότι, στην παραπάνω ανάλυση υποθέσαμε *χαλαρό περιθώριο νόρμας* 1 *(1-norm soft margin), περιορισμός κουτιού (the box constraint)*, καθώς αυτό χρησιμοποιήσαμε και στην πειραματική διαδικασία. Απλά αναφέρουμε ότι συχνά χρησιμοποιείται και το *χαλαρό περιθώριο (Ευκλείδιας) νόρμας* 2 *(2-norm soft margin), σταθμίζοντας τη διαγώνιο (weighting the diagonal).*

Για το πρωτογενές πρόβλημα βελτιστοποίησης (3.30), η συνάρτηση Lagrangian είναι

$$L(\mathbf{w}, b, \xi, \alpha, \mathbf{r}) = \frac{1}{2}\langle \mathbf{w}, \mathbf{w}\rangle + C\sum_{i=1}^{l}\xi_i - \sum_{i=1}^{l}\alpha_i[y_i(\langle \mathbf{w}, \phi(\mathbf{x}_i)\rangle + b) - 1 + \xi_i] - \sum_{i=1}^{l} r_i\xi_i, \quad (3.31)$$

όπου $\alpha_i \geq 0$ και $\xi_i \geq 0$ οι πολλαπλασιαστές Lagrange. Όσον αφορά στο ισοδύναμο δυϊκό πρόβλημα, παραγωγίζοντας τη σχέση 3.31 ως προς $\mathbf{w}, \xi$ και b έχουμε

$$\begin{aligned} \mathbf{w} &= \sum_{i=1}^{l} y_i\alpha_i\phi(\mathbf{x}_i) \\ \alpha_i &= C - r_i \\ 0 &= \sum_{i=1}^{l} y_i\alpha_i \end{aligned} \quad (3.32)$$

Με βάση τις παραπάνω σχέσεις, η πρωτογενής Lagrangian γίνεται

$$L(\mathbf{w}, b, \xi, \alpha, \mathbf{r}) = \sum_{i=1}^{l}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{l} y_iy_j\alpha_i\alpha_j\langle\phi(\mathbf{x}_i), \phi(\mathbf{x}_j)\rangle, \quad (3.33)$$

η οποία είναι ίδια με αυτή του μεγίστου περιθωρίου. Η μόνη διαφορά είναι η συνθήκη $\alpha_i \leq C$ που προκύπτει συνδυάζοντας τη δεύτερη συνθήκη (3.32) με τη $r_i \geq 0$. Οι συνθήκες Karush-Kuhn-Tucker (Κεφάλαιο 5.2 [7]) διαμορφώνονται ως

$$\begin{aligned} \alpha_i^*[y_i(\langle \mathbf{w}^*, \phi(\mathbf{x}_i)\rangle + b^*) - 1 - \xi_i] &= 0 \\ \xi_i(\alpha_i - C) &= 0 \end{aligned} \quad (3.34)$$

Από τις παραπάνω συνθήκες παρατηρούμε ότι, μη μηδενικές χαλαρές μεταβλητές έχουμε όταν $\alpha_i = C$, με γεωμετρικό περιθώριο μικρότερο από $1/\|\mathbf{w}\|$. Επομένως προκύπτει το ισοδύναμο δυϊκό πρόβλημα.

Ορισμός 3.5.1. *Δεδομένου ενός συνόλου δεδομένων εκπαίδευσης S (3.1), θεωρώντας το χώρο χαρακτηριστικών που ορίζεται από τον πυρήνα $k(\mathbf{x}, \mathbf{x}')$ και ότι οι παράμετροι α^* αποτελούν λύση του τετραγωνικού προβλήματος βελτιστοποίησης*

$$\begin{aligned} \max_{\alpha} \ & W(\alpha) = \sum_{i=1}^{l}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{l} y_iy_j\alpha_i\alpha_j k(\mathbf{x}_i, \mathbf{x}_j) \\ \text{subject to } & \sum_{i=1}^{l} y_i\alpha_i = 0 \\ & C \geq \alpha_i \geq 0, i = 1 \ldots l, \end{aligned} \quad (3.35)$$

ο κανόνας απόφασης δίδεται από το $sgn(f(\mathbf{x}))$*, όπου*

$$f(\mathbf{x}) = \sum_{i=1}^{l} y_i \alpha_i^* k(\mathbf{x}_i, \mathbf{x}) + b^*, \tag{3.36}$$

το b^* *επιλέγεται ώστε να ισχύει:* $y_i f(\mathbf{x}_i) = 1, \forall i$ *με* $C > \alpha_i^* > 0$*,*

$$b^* = \frac{1}{|I|} \sum_{i \in I} \left(y_i - \sum_{j=1}^{l} y_j \alpha_j k(\mathbf{x}_i, \mathbf{x}j) \right), \tag{3.37}$$

όπου $I := \{i : 0 < \alpha_i < C\}$*, το σύνολο των SV, και οι χαλαρές μεταβλητές ορίζονται με σύμφωνα με το γεωμετρικό περιθώριο*

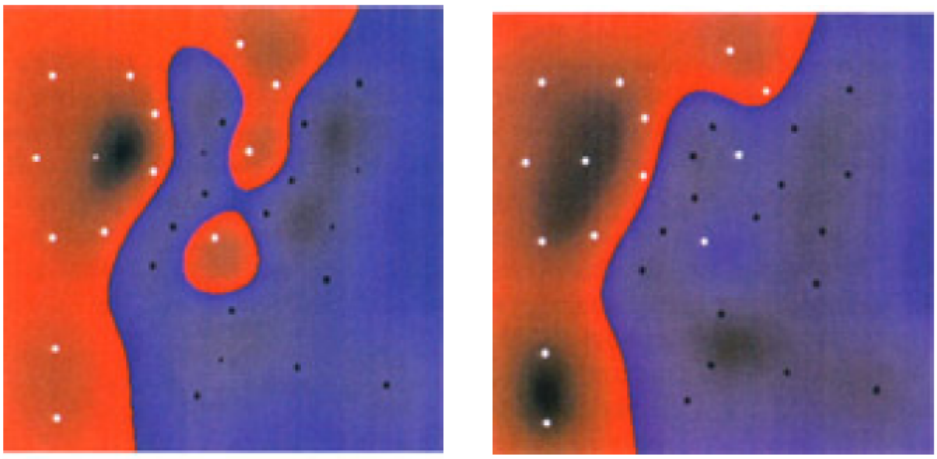
$$\gamma = \left(\sum_{i,j \in sv} y_i y_j \alpha_i^* \alpha_j^* k(\mathbf{x}_i, \mathbf{x}_j) \right)^{-1/2} \tag{3.38}$$

Συνολικά παρατηρούμε τα εξής αναφορικά με τον ταξινομητή χαλαρού περιθωρίου νόρμας 1:

- Η παράμετρος $C \in [1/l, \infty)$, ισοδυναμεί σε ένα συντελεστή ρύθμισης, καθώς ελέγχει το *ισοζύγιο (trade-off)* μεταξύ των επιτρεπόμενων λαθών στα εκπαιδευόμενα δεδομένα και συνεπώς της ακρίβειας (accuracy), και της πολυπλοκότητας του μοντέλου, δεδομένου ότι ελέγχει το μέγεθος των συντελεστών α_i.

- Το *χαλαρού περιθωρίου υπερεπίπεδο (soft margin hyperplane)* είναι ισοδύναμο με αυτό του μεγίστου περιθωρίου, με τον πρόσθετο περιορισμό ότι τα α_i είναι άνω φραγμένα από την ποσότητα C, γι αυτό και ο χαρακτηρισμός *περιορισμός κουτιού (box constraint)*. Ο περιορισμός αυτός εξασφαλίζει αφενός την ευστάθεια της βέλτιστης λύσης και αφετέρου περιορίζει την επιρροή των outliers και επομένως του θορύβου που πιθανότατα εμπεριέχουν τα δεδομένα.

Το παρακάτω σχήμα διασαφηνίζει τη διαφοροποίηση των ταξινομητών μέγιστου και χαλαρού περιθωρίου.



(α) (β)

Σχήμα 3.2: Παράδειγμα ταξινομητών μεγίστου (α) και χαλαρού (β) περιθωρίου [7].

3.5.2 ν-ΜΔΥ

Ένα από τα προβλήματα των C-SVM είναι ο προσδιορισμός της παραμέτρου C, καθώς η κλίμακα της επηρεάζεται από την εκάστοτε επιλογή του χώρου των χαρακτηριστικών. Εντούτοις, αποδεικνύεται ότι, αν θεωρήσουμε $C = 1/(\nu l)$, οι λύσεις που εξάγονται για διαφορετικές τιμές της παραμέτρου είναι οι ίδιες με εκείνες που εξάγονται, καθώς το ν μεταβάλλεται μεταξύ 0 και 1 στο παρακάτω πρόβλημα βελτιστοποίησης:

Ορισμός 3.5.2. *Δεδομένου ενός συνόλου δεδομένων εκπαίδευσης S (3.1), θεωρώντας το χώρο χαρακτηριστικών που σιωπηρά ορίζεται από τον πυρήνα $k(\mathbf{x}, \mathbf{x}')$ και ότι οι παράμετροι α^* αποτελούν λύση του τετραγωνικού προβλήματος βελτιστοποίησης*

$$\begin{aligned} \max_{\alpha} \ & W(\alpha) = -\frac{1}{2}\sum_{i,j=1}^{l} y_i y_j \alpha_i \alpha_j k(\mathbf{x}_i \mathbf{x}_j) \\ \text{subject to } & \sum_{i=1}^{l} y_i \alpha_i = 0 \\ & \sum_{i=1}^{l} y_i \alpha_i \geq \nu \\ & 1/l \geq \alpha_i \geq 0, i = 1 \ldots l, \end{aligned} \tag{3.39}$$

ο κανόνας απόφασης δίδεται από το $sgn(f(\mathbf{x}))$, όπου $f(\mathbf{x})$ δίδεται από τη σχέση 3.36, το b^* επιλέγεται ώστε να ισχύει: $y_i f(\mathbf{x}_i) = 1, \forall i$, ενώ το ξ_i ορίζεται:

$$\xi_i = \max\left(0, 1 - y_i\left(\sum_{j=1}^{l} y_j \alpha_j k(\mathbf{x}_j, \mathbf{x}_i) + b\right)\right) \tag{3.40}$$

Να σημειώσουμε ότι στην πράξη, η παράμετρος ν δεν κυμαίνεται σε όλο το διάστημα

$(0, 1]$ (Chang and Lin (2001)), αλλά η λύση είναι εφικτή αν και μόνο αν

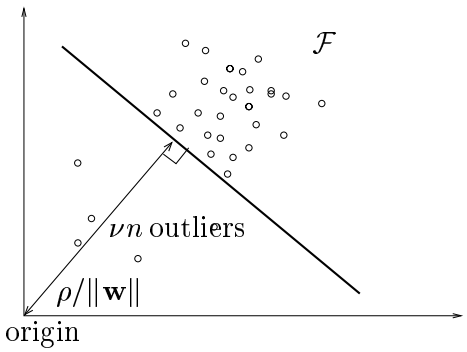
$$\nu \leq \left(\frac{2 \min(N_+, N_-)}{l} \right) \leq 1, \tag{3.41}$$

όπου N_+, N_- τα μεγέθη των δύο κλάσεων και l η διάσταση του χώρου των χαρακτηριστικών.

Με βάση τα παραπάνω, το ν θέτει ένα κάτω φράγμα στο άθροισμα των α_i με αποτέλεσμα ο γραμμικός όρος να εξαλείφεται από την αντικειμενική συνάρτηση. Αποδεικνύεται ότι, το κλάσμα του συνόλου των εκπαιδευόμενων δεδομένων είναι άνω φραγμένο από το ν, ενώ παράλληλα η ίδια παράμετρος αποτελεί κάτω φράγμα για το συνολικό αριθμό των SV. Κατά συνέπεια, η δεδομένη παραμετροποίηση είναι πιο διαφανής, καθώς δεν επηρεάζεται από την κλίμακα του χώρου των χαρακτηριστικών παρά μόνο από το επίπεδο θορύβου στα δεδομένα.

3.6 ΜΔΥ-μονής κλάσης

Τα *ΜΔΥ-μονής κλάσης* (single class-SVM) αποτελούν αποτελούν μηχανές μη εποπτευόμενης εκπαίδευσης με στόχο την εκτίμηση της πυκνότητας των εκπαιδευόμενων δεδομένων. Τα single-class SVM αντί να μοντελοποιούν την πυκνότητα των δεδομένων, στοχεύουν στην εύρεση ενός ομαλού συνόρου που να περικλείει μια περιοχή υψηλής πυκνότητας. Το σύνορο αυτό επιλέγεται, ούτως ώστε να αναπαριστά ένα ποσοστό της πυκνότητας, δηλαδή η πιθανότητα ένα σημείο δεδομένων εξαγόμενο από την κατανομή να βρεθεί μέσα σε αυτή την περιοχή δίδεται από ένα καθορισμένο αριθμό μεταξύ 0 και 1, δηλαδή έχει οριστεί εκ των προτέρων. Γενικά έχουν αναπτυχθεί δύο προσεγγίσεις όσον αφορά στο παραπάνω πρόβλημα. Η προσέγγιση που χρησιμοποιήσαμε (Schölkopf 2001), υπολογίζει ένα υπερεπίπεδο στο χώρο των χαρακτηριστικών τέτοιο ώστε, ένα καθορισμένο κλάσμα των δειγμάτων εκπαίδευσης να κείται πάνω από αυτό, ενώ την ίδια στιγμή να έχει μέγιστη απόσταση (περιθώριο) από την αρχή των αξόνων. Αντίθετα η εναλλακτική προσέγγιση (Tax και Duin (1999)) αναζητά τη μικρότερη δυνατή σφαίρα στο χώρο των χαρακτηριστικών, η οποία περιέχει ένα κλάσμα των σημείων δεδομένων. Για πυρήνες οι οποίοι είναι συναρτήσεις βαθμωτών γινομένων, οι δύο αλγόριθμοι είναι ισοδύναμοι.



Σχήμα 3.3: Ένα υπερεπίπεδο διαχωρισμού $(\mathbf{w}, \rho)$ για ένα σύνολο εκπαίδευσης δύο διαστάσεων στο χώρο χαρακτηριστικών F, το οποίο μεγιστοποιεί την απόσταση από την αρχή, επιτρέποντας παράλληλα ν outliers.

Σε αντιστοιχία με τις προηγούμενες προσεγγίσεις χαλαρού περιθωρίου νόρμας 1 ταξινόμησης, προκύπτει το ακόλουθο τετραγωνικό πρόβλημα

$$\begin{aligned} \min_{\xi,\mathbf{w},\rho} \ & \frac{1}{2}\langle \mathbf{w}, \mathbf{w} \rangle + \frac{1}{\nu l} \sum_{i=1}^{l} \xi_i - \rho \\ \textit{subject to } \ & \langle \mathbf{w}, \phi(\mathbf{x}_i) \rangle \geq \rho - \xi_i, i = 1 \dots l \\ & \xi_i \geq 0, i = 1 \dots l, \end{aligned} \tag{3.42}$$

Η παράμετρος $\nu \in (0, 1]$, είναι αντίστοιχη με αυτή της προηγούμενης ενότητας. Επίσης, η συνάρτηση απόφασης έχει τη μορφή

$$f(\mathbf{x}) = \langle \mathbf{w}, \phi(\mathbf{x}) \rangle - \rho \tag{3.43}$$

Ορισμός 3.6.1. *Δεδομένου ενός συνόλου δεδομένων εκπαίδευσης S (3.1), θεωρώντας το χώρο χαρακτηριστικών που ορίζεται από τον πυρήνα $k(\mathbf{x}, \mathbf{x}')$ και ότι οι παράμετροι α^* αποτελούν λύση του τετραγωνικού προβλήματος βελτιστοποίησης*

$$\begin{aligned} \max_{\alpha} \ & W(\alpha) = -\frac{1}{2} \sum_{i,j=1}^{l} y_i y_j \alpha_i \alpha_j k(\mathbf{x}_i, \mathbf{x}_j) \\ \text{subject to } \ & \sum_{i=1}^{l} y_i \alpha_i = 1 \\ & 0 \geq \alpha_i \geq 1/\nu l, i = 1 \dots l, \end{aligned} \tag{3.44}$$

ο κανόνας απόφασης δίδεται από το $sgn(f(\mathbf{x}))$, όπου $f(\mathbf{x})$ δίδεται από τη σχέση 3.43 και το ρ^* δίνεται ως

$$\begin{aligned} \rho^* &= \langle \mathbf{w}^*, \phi(\mathbf{x}_i) \rangle \\ &= \sum_j \alpha_j^* k(\mathbf{x}_i, \mathbf{x}_j) \end{aligned} \tag{3.45}$$

Ο ρόλος του ν στην παραπάνω μοντελοποίηση συνοψίζεται ως εξής: Όταν το ν πλησιάζει το 0, το πρόβλημα ανάγεται σε hard margin, καθώς η ποινή των λαθών γίνεται άπειρη (3.42). Αποδεικνύεται ότι, αν τα δεδομένα είναι διαχωρίσιμα, ο παραπάνω αλγόριθμος βρίσκει το ενιαίο υπερεπίπεδο υποστήριξης με την ιδιότητα ότι διαχωρίζει τα δεδομένα ως προς την αρχή και ότι η απόστασή του από την αρχή είναι μέγιστη μεταξύ όλων των αντίστοιχων δυνατών υπερεπιπέδων. Από την άλλη πλευρά, αν το ν ισούται με 1, όλα τα α_i ισούται με το άνω φράγμα $1/\nu l$. Στην περίπτωση αυτή, για πυρήνες με ολοκλήρωμα 1, η συνάρτηση απόφασης αντιστοιχεί σε ένα εκτιμητή παραθύρου Parzen με κατώφλι.

Θεώρημα 3.6.2 (ν-Ιδιότητα)**.** *Υποθέτουμε ότι η λύση της σχέσης 3.45 ικανοποιεί $\rho \neq 0$. Τότε ισχύουν τα παρακάτω:*

(i) Το ν είναι ένα άνω φράγμα του κλάσματος των outliers.

(ii) Το ν είναι ένα κάτω φράγμα του κλάσματος των SV.

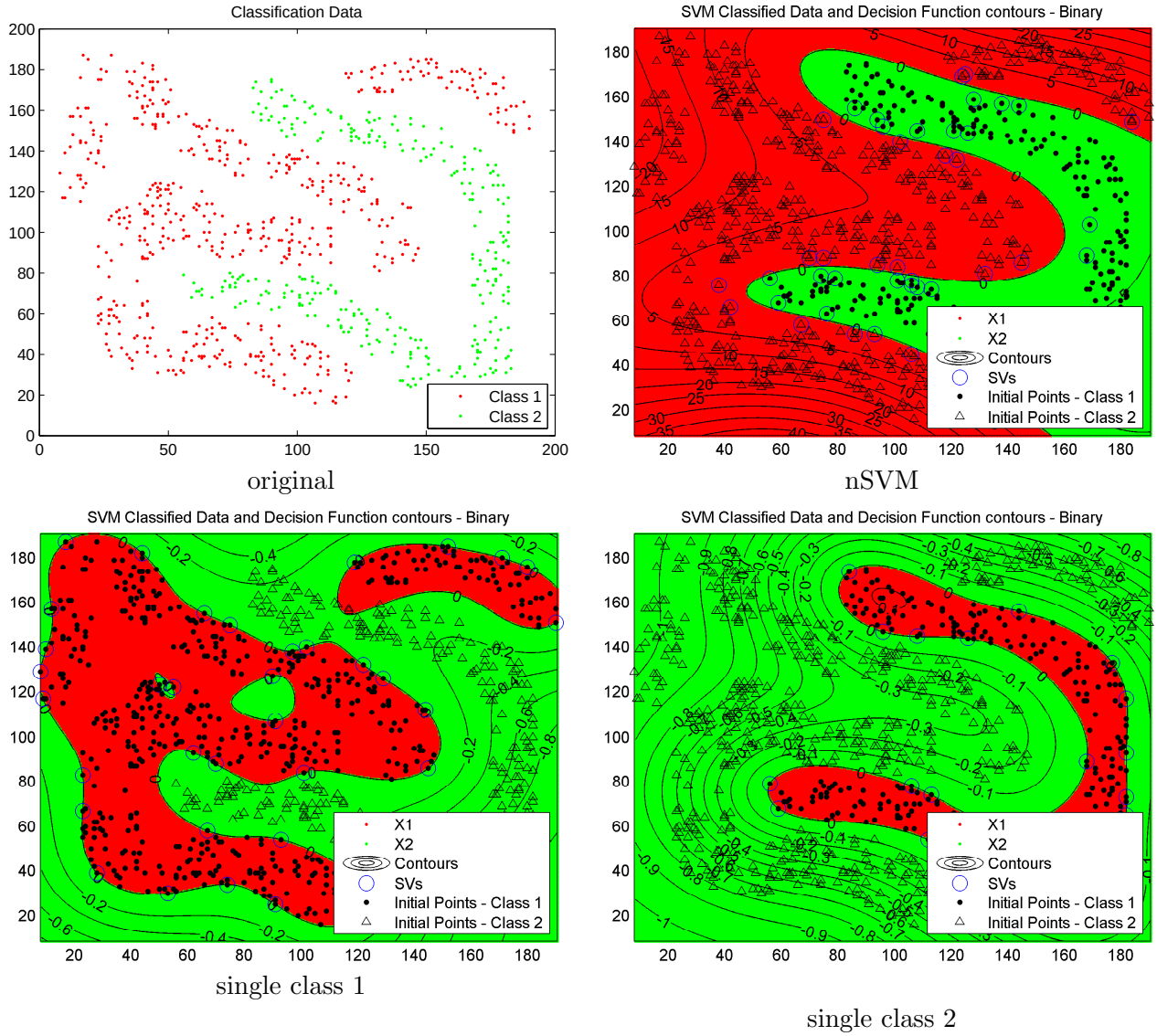
(iii) Υποθέτουμε ότι τα δεδομένα έχουν δημιουργηθεί ανεξάρτητα από μια κατανομή $P(\mathbf{x})$, η οποία δεν περιλαμβάνει διακριτά στοιχεία και επιπλέον ότι ο πυρήνας είναι αναλυτικός και μη σταθερός. Όταν ο αριθμός l των δειγμάτων τείνει στο άπειρο, με πιθανότητα 1, το ν ισούται με το κλάσμα τόσο των SV, όσο και των outliers.

Συνολικά προκύπτουν τα εξής συμπεράσματα αναφορικά με τον αλγόριθμο των single-class SVM.

- Υπολογίζουν μια "καλή περιοχή", η οποία περικλείει ένα κλάσμα των δειγμάτων εκπαίδευσης, όπου ο χαρακτηρισμός "καλή" εκφράζει το ότι αντιστοιχεί σε μια μικρή τιμή της νόρμας $\|\mathbf{w}\|^2$ και επομένως η υποκείμενη συνάρτηση θα είναι ομαλή (smooth).
- Υποθέτοντας ότι το εκτιμώμενο υπερεπίπεδο έχει μικρή τιμή $\|\mathbf{w}\|^2$ και διαχωρίζει μέρος των εκπαιδευόμενων δεδομένων από την αρχή με έναν συγκεκριμένο περιθώριο $\rho/\|\mathbf{w}\|$, η πιθανότητα τα νέα δείγματα που προέρχονται από την ίδια κατανομή να βρεθούν εκτός μιας ελαφρά ευρύτερης περιοχής δε θα είναι πολύ μεγαλύτερη από το κλάσμα των outliers των δεδομένων εκπαίδευσης.

Το παράδειγμα 3.4 συνοψίζει τη διαφοροποίηση των διαφόρων κατηγοριών SVM, δεδομένου του συνόλου δεδομένων *fourclass (LIBSVM Data)* [1].

[1] http://www.csie.ntu.edu.tw/ cjlin/libsvmtools/datasets/



Σχήμα 3.4: Παράδειγμα ταξινόμησης διαφόρων SVM.

3.7 ΜΔΥ-πολλών κλάσεων

Έχουν αναπτυχθεί διάφορες μέθοδοι για το συνδυασμό πολλών δυαδικών SVM, με στόχο τη δημιουργία *ΜΔΥ-πολλών κλάσεων)*. Η πιο συχνά χρησιμοποιούμενη προσέγγιση (Vapnik, 1998) είναι η κατασκευή K μεμονωμένων SVM, στα οποία το k-οστό μοντέλο $y_k(\mathbf{x})$ εκπαιδεύεται χρησιμοποιώντας δεδομένα από την κλάση C_k ως θετικά δείγματα και δεδομένα από τις υπόλοιπες $k-1$ ως αρνητικά δείγματα. Αυτή είναι γνωστή ως *μια έναντι των υπολοίπων (one-versus-the-rest)* προσέγγιση.

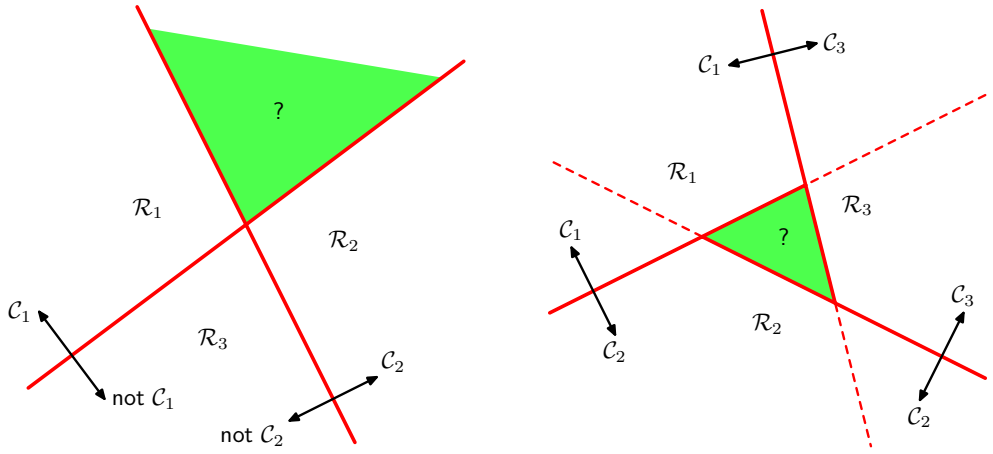
Παρόλα αυτά, η εκπαίδευση πολλών μεμονωμένων ταξινομητών μπορεί να οδηγήσει σε αντιφατικά αποτελέσματα, καθώς μια είσοδος μπορεί να αποδοθεί σε πολλές κλάσεις ταυ-

τόχρονα. Αυτό το πρόβλημα μερικές φορές διευθετείται κάνοντας προβλέψεις για τις νέες εισόδους $\mathbf{x}$ χρησιμοποιώντας τη σχέση

$$y(\mathbf{x}) = \max_k y_k \tag{3.46}$$

Δυστυχώς αυτός ο χειρισμός, παρουσιάζει το πρόβλημα ότι διαφορετικοί ταξινομητές έχουν εκπαιδευτεί σε διαφορετικά προβλήματα, και έτσι δεν υπάρχει καμία εγγύηση ότι, οι πραγματικές ποσότητες y_k για τους διαφορετικούς ταξινομητές θα έχουν την ίδια κλίμακα. Ένα ακόμη πρόβλημα είναι ότι τα εκπαιδευόμενα σύνολα δεδομένων δεν είναι ισοκατανεμημένα όσον αφορά στην αναλογία θετικών, αρνητικών δειγμάτων, με αποτέλεσμα να χάνεται η συμμετρία του αρχικού προβλήματος.

Μια εναλλακτική μέθοδος είναι να εισάγουμε $K(K-1)/2$ συναρτήσεις δυαδικής ταξινόμησης, μια για κάθε ζεύγος κλάσεων και είναι γνωστή ως *μια έναντι μιας (one-versus-one)* προσέγγιση. Κάθε δείγμα τότε ταξινομείται με βάσει μια ψήφο πλειοψηφίας μεταξύ των συναρτήσεων. Και αυτή η προσέγγιση παρουσιάζει προβλήματα όπως φαίνεται στο παρακάτω σχήμα.



Σχήμα 3.5: Η προσπάθεια ταξινόμησης K κλάσεων με βάση μεμονωμένους δυαδικούς ταξινομητές αντιμετωπίζει το πρόβλημα των αμφιλεγόμενων περιοχών που φαίνονται με πράσινο. **Αριστερά** παράδειγμα δύο ταξινομητών, οι οποίοι διακρίνουν την κλάση C_k από τις υπόλοιπες. **Δεξιά** παράδειγμα τριών ταξινομητών, καθένας από τους οποίους διακρίνει ένα ζεύγος κλάσεων C_k, C_j κάθε φορά.

Στη μέθοδο που εισάγουμε, για την εφαρμογή ταξινόμησης πολλών κλάσεων χρησιμοποιήσαμε την προσέγγιση μια έναντι των υπολοίπων. Υπάρχουν διάφορες μέθοδοι αντιμετώπισης του προβλήματος της κλίμακας των επιμέρους ταξινομητών. Παρόλα αυτά δε χρησιμοποιήθηκε κάποια, δεδομένης της ικανοποιητικής λειτουργίας του συνδυαστικού ταξινομητή πολλών κλάσεων.

Κεφάλαιο 4

Χωρικό ταίριασμα

4.1 Εισαγωγή

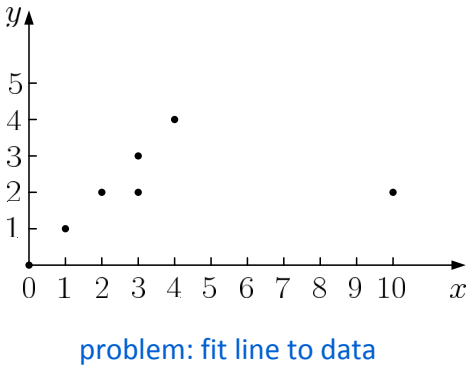
Με βάση το χωρικό ταίριασμα που ορίσαμε στο προηγούμενο κεφάλαιο 2.3, κύριο θέμα του κεφαλαίου είναι η εισαγωγή μιας ειδικής κατηγορίας χωρικού ταιριάσματος, η οποία παρουσιάζει ιδιαίτερο ενδιαφέρον για λόγους που θα αναπτυχθούν στη συνέχεια. Με τον όρο *ευέλικτο και χαλαρό χωρικό ταίριασμα (flexible and relaxed spatial matching)*, αναφερόμαστε σε μια προσέγγιση της διαδικασίας χωρικού ταιριάσματος, η οποία παρουσιάζει δύο βασικά χαρακτηριστικά. Ο πρώτος όρος, "ευέλικτο" προκύπτει από το γεγονός ότι επιτρέπει την μη άκαμπτη κίνηση και το πολλαπλό ταίριασμα επιφανειών και αντικειμένων και ο δεύτερος όρος, "χαλαρό" δηλώνει την εξαγωγή κατανομών επί ιεραρχικών διαμερίσεων αντί για υπολογισμούς ανά ζεύγη. Η προσέγγιση αυτή αποδεικνύεται ως πιο αποδοτική, καθώς διατηρεί όλη τη χωρική πληροφορία, ενώ παράλληλα συμβάλλει στη βελτίωση της επίδοσης των αλγορίθμων ταιριάσματος απλοποιώντας σημαντικά τη διαδικασία σε επίπεδο υπολογισμών. Εισάγει λοιπόν, τη δυνατότητα ανάπτυξης μεθόδων κατάλληλων για τη βελτίωση της επίδοσης της ανακατάταξης εικόνων στις μηχανές αναζήτησης, όσον αφορά τόσο στην ακρίβεια, όσο και στην ταχύτητα. Στη συνέχεια αναπτύσσουμε κάποιες τεχνικές χωρικού ταιριάσματος εικόνων ως εισαγωγή σε μια μέθοδο ευέλικτου, χαλαρού χωρικού ταιριάσματος στην οποία θα αναφερθούμε αναλυτικά.

4.2 Τεχνικές χωρικού ταιριάσματος εικόνων

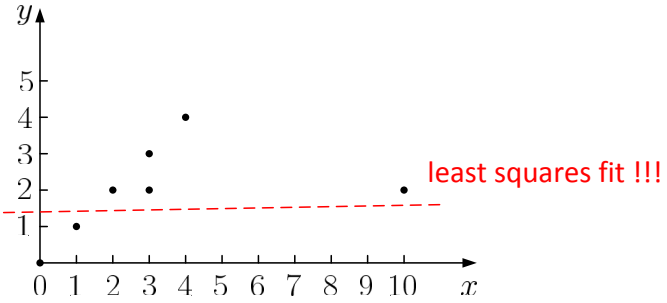
4.2.1 Ομοφωνία τυχαίων δειγμάτων

Την πιο διαδεδομένη μέθοδο χωρικού ταιριάσματος αποτελεί ο αλγόριθμος *Ομοφωνία Τυχαίων Δειγμάτων (RANdom SAmple Consensus)* (RANSAC) [11], ο οποίος περιλαμβάνει τη εξαγωγή υποθέσεων μετασχηματισμού χρησιμοποιώντας τον ελάχιστο αριθμό αντιστοιχιών και στη συνέχεια την αξιολόγηση κάθε υπόθεσης με βάση τον αριθμό των *έγκυρων σημείων (inliers)* μεταξύ των χαρακτηριστικών της δεδομένης υπόθεσης. Πιο συγκεκριμένα στην πιο απλή περίπτωση, δεδομένου ενός συνόλου δισδιάστατων σημείων (σχήμα 4.1) αναζητούμε την ευθεία που ελαχιστοποιεί το άθροισμα των τετραγώνων των κάθετων αποστάσεων, υπό τη συνθήκη ότι κανένα από τα inliers, δεν αποκλίνει από τη δεδομένη ευθεία περισσότερο από κάποια *τιμή κατωφλίου (threshold)* t. Επομένως, προκύπτει ένα πρόβλημα δύο κατευθύνσεων:

η εύρεσης γραμμής που να ταιριάζει στα δεδομένα και η ταξινόμηση των σημείων σε inliers και σε outliers.



(α′) Σύνολο δεδομένων στο χώρο δύο διαστάσεων.



(β′) Η γραμμή που ταιριάζει στα δεδομένα με τη χρήση ελάχιστων τετραγώνων.

Σχήμα 4.1: Πρόβλημα εύρεσης γραμμής που ταιριάζει στα δεδομένα (Γιάννης Αβρίθης).

Η ιδέα είναι πολύ απλή. Επιλέγουμε τυχαία δύο σημεία και θεωρούμε την ευθεία που ορίζουν. Η *υποστήριξη (support)* για τη δεδομένη ευθεία είναι ο αριθμός των inliers. Η ίδια διαδικασία επαναλαμβάνεται μερικές φορές και η ευθεία με τη μεγαλύτερη υποστήριξη επιλέγεται ως βέλτιστη. Ο αλγόριθμος RANSAC συνοψίζεται παρακάτω 1.

Όσον αφορά στο στάδιο ανακατάταξης στην ανάκτηση εικόνων, αρχικά βρίσκουμε τα *αντίστοιχα χαρακτηριστικά (corresponding features)*, δηλαδή τα χαρακτηριστικά με την ίδια οπτική λέξη, και στη συνέχεια για κάθε τέτοιο ζεύγος θεωρούμε μια υπόθεση ομογραφικού μετασχηματισμού. Κάθε υπόθεση μετασχηματισμού αξιολογείται με βάση τον αριθμό των inliers που εμφανίζει, αποθηκεύουμε τον αντίστοιχο ομογραφικό μετασχηματισμό και επαναλαμβάνουμε τη διαδικασία μέχρις ότου να πετύχουμε ένα μέγιστο αριθμό inliers.

Θεωρούμε ότι η χωρική επιβεβαίωση είναι επιτυχημένη αν στο χωρικό ταίριασμα της εικόνας αναζήτησης και των εικόνων βάσης που βρίσκονται στην κορυφή της λίστας κατάταξης, ανιχνεύσουμε έναν ελάχιστο αριθμό inliers, ο οποίος για παράδειγμα στην περίπτωση επίπεδης ομογραφίας δεδομένων με αντιστοιχίες δύο διαστάσεων είναι 4. Στη συνέχεια, ανακατατάσσουμε τις εικόνες με βάση το άθροισμα των τιμών idf των λέξεων που είναι inliers. Οι εικόνες που δεν επιβεβαιώθηκαν γεωμετρικά μπαίνουν στο τέλος της λίστας.

Algorithm 1 RANSAC

Στόχος: Εύρωστο ταίριασμα ενός μοντέλου σε ένα σύνολο δεδομένων S που περιέχει outliers:

1: Επέλεξε ένα τυχαίο δείγμα σημείων δεδομένων s από το σύνολο S και κατασκεύασε το μοντέλο με αυτές τις παραμέτρους.
2: Προσδιόρισε το σύνολο των σημείων δεδομένων S_i, τα οποία βρίσκονται εντός ενός κατωφλίου απόστασης t από το μοντέλο. Το σύνολο S_i είναι το σύνολο *"ομοφωνίας" του δείγματος (consensus set)* και ορίζει τους inliers του S.
3: Αν το μέγεθος του S_i (ο αριθμός των inliers) είναι μεγαλύτερο από κάποιο κατώφλι T, τότε επανεκτίμησε το μοντέλο χρησιμοποιώντας όλα τα σημεία του S και τερμάτισε.
4: Αν το μέγεθος του S_i είναι μικρότερο από T, επέλεξε ένα νέο δείγμα και επανάλαβε τα παραπάνω βήματα.
5: Μετά από N δοκιμές, το μεγαλύτερο σύνολο ομοφωνίας S_i επιλέγεται και το μοντέλο επανεκτιμάται χρησιμοποιώντας όλα τα σημεία του υποσυνόλου S_i

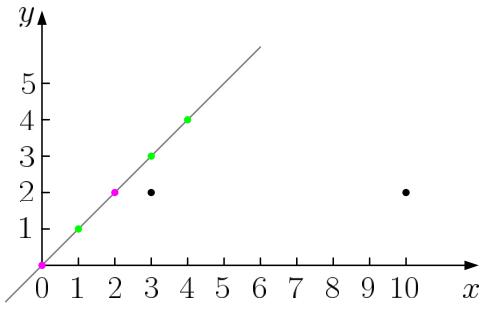
Συνοψίζοντας, αναφέρουμε την πρόταση των Fischler and Bolles [9] ότι η διαδικασία που περιλαμβάνει ο αλγόριθμος RANSAC είναι αντίθετη αυτής του συμβατικών τεχνικών εξομάλυνσης. Αντί να χρησιμοποιεί όσο το δυνατόν περισσότερα δεδομένα για την εξαγωγή μιας αρχικής λύσης και να επιδιώκει ακολούθως την αφαίρεση των μη έγκυρων δεδομένων, χρησιμοποιεί όσο το δυνατόν λιγότερα δεδομένα και επεκτείνει αυτό το σύνολο με συνεπή δεδομένα, όταν αυτό είναι δυνατό. Σημαντικό μειονέκτημα της τεχνικής αυτής είναι η χαμηλή απόδοση, όταν το ποσοστό των inliers είναι πολύ μικρό.

4.2.2 Τοπικά βελτιστοποιημένη ομοφωνία τυχαίων δειγμάτων

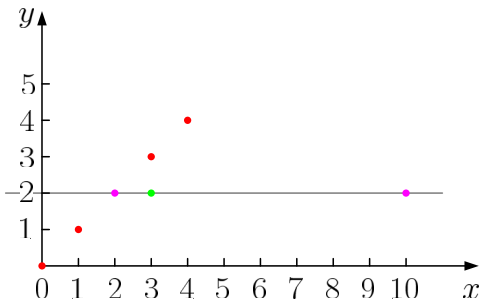
Η μέθοδος *Τοπικά Βελτιστοποιημένη Ομοφωνία Τυχαίων Δειγμάτων (Locally Optimised RANdom SAmple Consensus)* (Lo-RANSAC) [6] αποτελεί επέκταση της μεθόδου RANSAC, με την προσθήκη ενός βήματος βελτιστοποίησης ενός γενικευμένου μοντέλου (LO βήμα) το οποίο εφαρμόζεται μόνο σε μοντέλα με βαθμολογία καλύτερη από όλα τα προηγούμενα. Στον απλό RANSAC γίνεται η υπόθεση ότι αν ένα δείγμα σημείων δεν περιέχει outliers, το μοντέλο που προκύπτει με βάση το δεδομένο δείγμα θα είναι συνεπές για όλα τα inliers. Στην πραγματικότητα όμως αυτό δεν ισχύει λόγω πιθανής παρουσίας θορύβου και κακής θεώρησης συνθηκών (poor conditioning). Ο LO-RANSAC λοιπόν βασίζεται στην παρατήρηση ότι ιδανικά όλα τα εκτιμώμενα μοντέλα με βάση δείγματα ελαχίστου μεγέθους, περιλαμβάνουν μεγάλο ποσοστό inliers εντός της υποστήριξής τους. Γι αυτό, εισάγεται μια διαδικασία βελτιστοποίησης, δεδομένου του βέλτιστου υποτιθέμενου μοντέλου. Με αυτό τον τρόπο, ο αλγόριθμος που προκύπτει συμφωνεί σε μεγαλύτερο βαθμό με τη ζητούμενη, θεωρητική επίδοση, ενώ είναι πιο εύρωστος καθώς παρουσιάζει χαμηλότερη ευαισθησία στο θόρυβο και τη θεώρηση ανεπαρκών συνθηκών.

Έστω P η πιθανότητα να επιλεγεί τυχαία από ένα σύνολο U από N σημεία δεδομένων, ένα δείγμα που δεν περιέχει outliers. Τότε ισχύει:

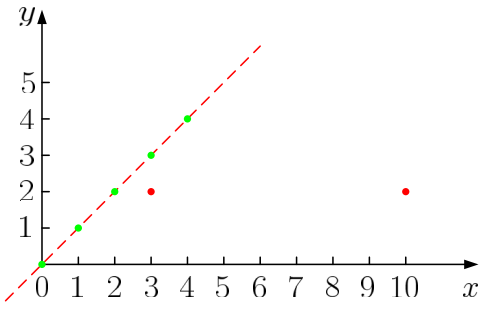
$$P = \frac{\binom{I}{m}}{\binom{N}{m}} = \prod_{j=0}^{m-1} \frac{I-j}{N-j} \approx \epsilon^m \tag{4.1}$$



(α′) Η υποθετική ευθεία ορίζεται από τα δύο *ρόζ* σημεία. Με *πράσινο* συμβολίζουμε τους *inliers* και με *κόκκινο* τους *outliers*.



(β′) Μία άλλη υποθετική ευθεία μαζί με τους *inliers* και *outliers*.



(γ′) Μετά από μερικές επαναλήψεις καταλήξαμε ότι η διακεκομμένη ευθεία ταιριάζει καλύτερα στα δεδομένα μας.

Σχήμα 4.2: Εφαρμογή RANSAC στην εύρεση ευθείας που ταιριάζει στα δεδομένα (Γιάννης Αβρίθης).

όπου ϵ το ποσοστό των inliers $\epsilon = 1/N$ και I το μέγεθος του δείγματος. Παρακάτω συνοψίζεται ο αλγόριθμος LO-RANSAC:

Algorithm 2 LO-RANSAC

Επανάλαβε μέχρι η πιθανότητα P εύρεσης ενός μοντέλου με υποστήριξη μεγαλύτερη από I^* στο k βήμα γίνεται μικρότερη από ένα προκαθορισμένο κατώφλι η_0:

1: Επέλεξε ένα τυχαίο δείγμα ελάχιστου μεγέθους m από το U.
2: Εκτίμησε τις παραμέτρους του μοντέλου που είναι συνεπές με το δεδομένο δείγμα.
3: Υπολόγισε την υποστήριξη I_k του μοντέλου, δηλαδή τα σημεία δεδομένων με σφάλμα μικρότερο από ένα προκαθορισμένο κατώφλι ϑ. Αν $I_k > I_j, \forall i < k$, δηλαδή όταν ένα νέο μέγιστο έχει επιτευχθεί, τότε τρέξε: **LO βήμα**. Εφάρμοσε βελτιστοποίηση. Αποθήκευσε το καλύτερο μοντέλο που βρέθηκε και την υποστήριξή του $I^*(I^* \geq I_k)$ λόγω της βελτιστοποίησης.

Για το βήμα βελτιστοποίησης έχουν προταθεί και δοκιμαστεί διάφορες μέθοδοι [5]. Για λόγους πληρότητας τις αναφέρουμε συνοπτικά.

1. **Απλή**. Θεωρεί όλα τα σημεία δεδομένων με σφάλμα μικρότερο από ένα κατώφλι ϑ και χρησιμοποιεί ένα γραμμικό αλγόριθμο για την υπόθεση των παραμέτρων του νέου μοντέλου.

2. **Επαναληπτική**. Θεωρεί όλα τα σημεία με σφάλμα μικρότερο από $K \cdot \vartheta$ και χρησιμοποιεί ένα γραμμικό αλγόριθμο για τον υπολογισμό των παραμέτρων του νέου μοντέλου. Μειώνει το κατώφλι και επαναλαμβάνει μέχρι το κατώφλι να γίνει ίσο με ϑ.

3. **Εσωτερικός RANSAC**. Υλοποιείται μια νέα δειγματοληπτική διαδικασία. Επιλέγονται δείγματα μόνο από I_k σημεία δεδομένων, τα οποία είναι συνεπή με το υποτιθέμενο μοντέλο του k βήματος του RANSAC. Νέα μοντέλα επιβεβαιώνονται έναντι ολόκληρου του συνόλου των σημείων δεδομένων. Όσο η δειγματοληψία γίνεται σε inliers, δε χρειάζεται το μέγεθος του δείγματος να είναι ελάχιστο. Αντίθετα, το μέγεθος του δείγματος επιλέγεται να ελαχιστοποιεί το σφάλμα των παραμέτρων του εκτιμώμενου μοντέλου.

4. **Επαναληπτικός Εσωτερικός RANSAC**. Η μέθοδος αυτή είναι παρόμοια με την προηγούμενη, με τη διαφορά ότι κάθε δείγμα του εσωτερικού RANSAC υπόκειται στη μέθοδο 2.

Η τελευταία μέθοδος αναφέρεται ότι παρουσιάζει την καλύτερη επίδοση. Να σημειώσουμε ότι ο αλγόριθμος φαίνεται να εγγυάται ότι το LO βήμα εφαρμόζεται τόσο σπάνια, ώστε η επίδραση στο χρόνο εκτέλεσης να είναι μικρή.

Όσον αφορά στην ανάκτηση εικόνων, στο χωρικό ταίριασμα ο υπολογισμός του μετασχηματισμού μεταξύ της εικόνας αναζήτησης και την εικόνων της βάσης, πραγματοποιείται χρησιμοποιώντας τον Επαναληπτικό Εσωτερικό RANSAC, στο LO βήμα του αλγορίθμου LO-RANSAC θεωρώντας ως σύνολο δεδομένων ζεύγη χαρακτηριστικών με την ίδια οπτική λέξη (corresponding features) [17]. Πιο συγκεκριμένα, με βάση τις κοινές οπτικές λέξεις μεταξύ της εικόνας αναζήτησης και των εικόνων βάσης με το μεγαλύτερο βάρος tf-idf θεωρούμε *πιθανές αντιστοιχίες σημείων (tentative correspondences)*. Με δεδομένα λοιπόν δύο αντίστοιχα χαρακτηριστικά, υποθέτουμε κάποιο προσεγγιστικό μοντέλο και στη συνέχεια εφαρμόζουμε τον αλγόριθμο LO-RANSAC, όπως περιγράφεται παραπάνω.

4.2.3 Χωρικό ταίριασμα πυραμίδας

Η τεχνική *Χωρικό Ταίριασμα Πυραμίδας (Spatial Pyramid Matching)* (SPM) [13] αναφέρεται σε καθολικές μη ανεξάρτητες αναπαραστάσεις, βασισμένες σε αθροιστικά στατιστικά τοπικών χαρακτηριστικών επί καθορισμένων υποπεριοχών. Πιο συγκεκριμένα, τμηματοποιεί την εικόνα σε αυξανόμενα λεπτομερείς (fine) υποπεριοχές και υπολογίζει ιστογράμματα τοπικών χαρακτηριστικών που βρίσκονται σε κάθε υποπεριοχή. Η *χωρική πυραμίδα (spatial pyramid)* που προκύπτει, αποτελεί μια απλή και υπολογιστικά αποδοτική επέκταση της αναπαράστασης εικόνας χωρίς διάταξη, *"σάκος" χαρακτηριστικών (bag-of-features)*, καθώς μπορεί να αναδείξει σημαντικά από άποψη αντίληψης χαρακτηριστικά και παρουσιάζει σημαντικά βελτιωμένη επίδοση σε δύσκολα προβλήματα κατηγοριοποίησης σκηνών (scene categorization).

Η βασική ιδέα είναι ο συνδυασμός με έναν αξιωματικό τρόπο πολλών επιπέδων ανάλυσης της εικόνας στο πλαίσιο της λογικής *"subdivide and disorder (υποδιαίρεσε και ανατάραξε)"*. Το αποτέλεσμα της όλης διαδικασίας είναι μια μέθοδος *προσεγγιστικού γεωμετρικού ταιριάσματος (approximate geometrical matching)*, με πολλών διαστάσεων αναπαραστάσεις, οι οποίες εμπεριέχουν περισσότερη πληροφορία.

Έστω , δύο σύνολα διανυσμάτων σε ένα χώρο χαρακτηριστικών d διαστάσεων. Ας κατασκευάσουμε μια ακολουθία πλεγμάτων σε ανάλυσης $0 \dots L$, τέτοια ώστε σε κάθε επίπεδο l το πλέγμα να έχει 2^l κελιά σε κάθε διάσταση, άρα συνολικά $D = 2^{dl}$ κελιά. Έστω H_X^ℓ, H_Y^ℓ τα ιστογράμματα των , στο συγκεκριμένο επίπεδο/ανάλυση, έτσι ώστε $H_X^\ell(i), H_Y^\ell(i)$ ο αριθμός των σημείων από τα X, Y, τα οποία βρίσκονται στο i κελί του πλέγματος. Τότε ο αριθμός των ταιριασμάτων στο επίπεδο l δίνεται από τη συνάρτηση διατομής/διασταύρωσης (intersection) ιστογράμματος:

$$I(H_X^\ell, H_Y^\ell) = \sum_{i=1}^{D} \min(H_X^\ell(i), H_Y^\ell(i)) \tag{4.2}$$

Για συντομία αναφέρουμε στη συνέχεια το $I(H_X^\ell, H_Y^\ell)$ ως I^ℓ. Ο αριθμός των νέων ταιριασμάτων στο επίπεδο l είναι $I^\ell - I^{\ell+1}$ για $\ell = 0, \dots, L-1$. Δεδομένου ότι θέλουμε να δώσουμε μικρότερη βαρύτητα σε ταιριάσματα που βρίσκονται σε μεγαλύτερα κελιά, καθώς τα μεγαλύτερα κελιά περιλαμβάνουν περισσότερο ανόμοια χαρακτηριστικά, θέτουμε το βάρος όσον αφορά στο κάθε επίπεδο ℓ ίσο με $\frac{1}{2^{L-l}}$, δηλαδή αντιστρόφως ανάλογο του πλάτους του κελιού του αντίστοιχου επιπέδου. Συνοψίζοντας λοιπόν, παίρνουμε τον ακόλουθο ορισμό του *πυρήνα ταιριάσματος πυραμίδας (pyramid match kernel)*:

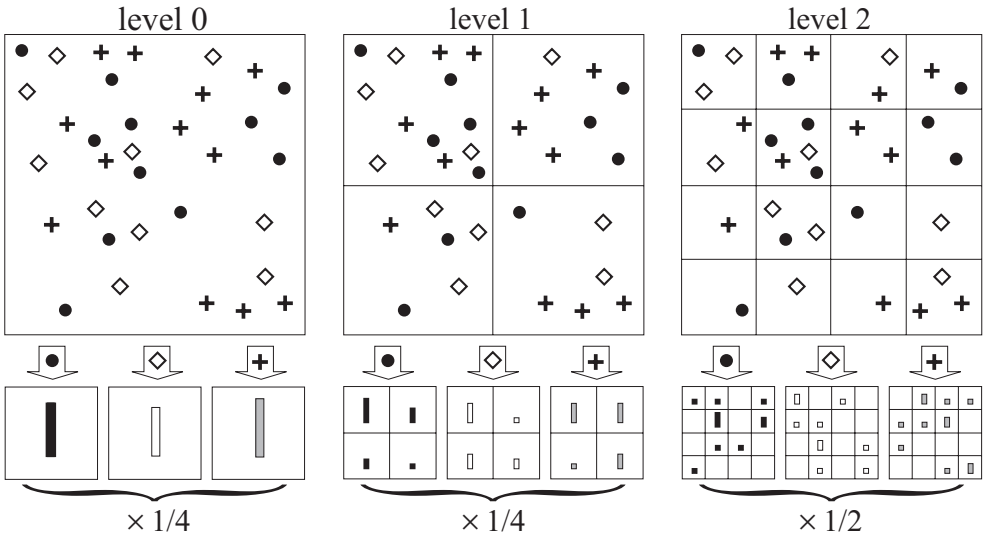
$$\begin{aligned} k^L(X,Y) &= I^L + \sum_{l=0}^{L-1} \frac{1}{2^{L-l}} (I^\ell - I^{\ell+1}) \\ &= \frac{1}{2^L} I^0 + \sum_{\ell=1}^{L} \frac{1}{2^{L-\ell+1}} I^\ell \end{aligned} \tag{4.3}$$

Τόσο η διασταύρωση ιστογράμματος όσο και ο πυρήνας ταιριάσματος πυραμίδας αποτελούν πυρήνες Mercer.

Αναφορικά και με την ανάκτηση εικόνων, η τεχνική SPM περιλαμβάνει μια "ορθογώνια" προσέγγιση ως εξής: εφαρμόζει ταίριασμα πυραμίδας στο χώρο δύο διαστάσεων της εικόνας

και χρησιμοποιεί παραδοσιακές τεχνικές συσταδοποίησης (clustering) στο χώρο των χαρακτηριστικών. Πιο συγκεκριμένα, έχοντας κατασκευάσει το οπτικό λεξικό M λέξεων, κάθε κανάλι m που αντιστοιχεί σε μια από τις οπτικές λέξεις, μας δίνει δύο σύνολα διανυσμάτων δύο διαστάσεων, X_m, Y_m, τα οποία αντιπροσωπεύουν τις συντεταγμένες των χαρακτηριστικών που περιέχουν τη λέξη m στις δύο εικόνες προς σύγκριση. Ο τελικός πυρήνας διαμορφώνεται ως το άθροισμα των πυρήνων των χωριστών καναλιών:

$$K^L(X, Y) = \sum_{m=1}^{M} k^L(X_m, Y_m) \tag{4.4}$$



Σχήμα 4.3: Η εικόνα έχει τρία είδη χαρακτηριστικών, που επισημαίνονται ως κύκλοι, διαμάντια και σταυροί. Στο πάνω μέρος υποδιαιρούμε την εικόνα σε τρία διαφορετικά επίπεδα ανάλυσης. Στη συνέχεια για κάθε επίπεδο ανάλυσης και κάθε κανάλι μετράμε τα χαρακτηριστικά που βρίσκονται μέσα σε κάθε χωρικό κελί. Τέλος, σταθμίζουμε κάθε χωρικό ιστόγραμμα με βάση τη σχέση 4.3

Το θετικό της παραπάνω προσέγγισης είναι ότι είναι συνεπής με τη θεωρία των οπτικών λεξικών (ισοδυναμεί με το bag-of-words για $L = 0$). Επίσης, επειδή ο πυρήνας ταιριάσματος πυραμίδας συνίσταται απλά από ένα σταθμισμένο άθροισμα διασταυρώσεων ιστογραμμάτων, και $c\min(\alpha, b) = \min(c\alpha, cb)$ για θετικούς αριθμούς, μπορούμε να υλοποιήσουμε το K^L ως ένα μεμονωμένο ιστόγραμμα διασταύρωσης διανυσμάτων μεγάλων διαστάσεων ενώνοντας τα κατάλληλα σταθμισμένα ιστογράμματα για όλα τα κανάλια και όλες τις αναλύσεις. Κατά αυτό τον τρόπο, για L επίπεδα και M κανάλια, το διάνυσμα που προκύπτει έχει διαστάσεις $M\sum_{\ell=0}^{L} 4^\ell = M\frac{1}{3}(4^{\ell+1} - 1)$. Επειδή τα διανύσματα ιστογραμμάτων είναι πολύ αραιά, η υπολογιστική πολυπλοκότητα του πυρήνα δεν επιβαρύνεται από την ύπαρξη πολλών διαστάσεων, αλλά αποδεικνύεται ότι είναι γραμμική στον αριθμό των χαρακτηριστικών.

Επιπρόσθετα, για μέγιστη υπολογιστική απόδοση, κανονικοποιούμε όλα τα ιστογράμματα με το συνολικό βάρος όλων των χαρακτηριστικών στην κάθε εικόνα, επιβάλλοντας ο αριθμός των χαρακτηριστικών σε όλες τις εικόνες να είναι ο ίδιος.

Συνολικά λοιπόν εξάγουμε τα ακόλουθα συμπεράσματα όσον αφορά στην τεχνική SPM:

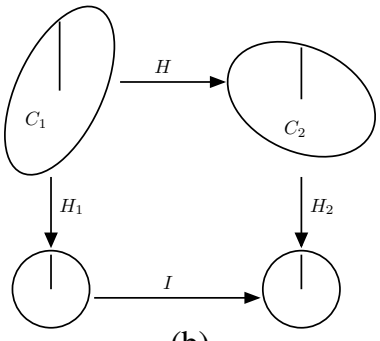
- Ο συνδυασμός πολλαπλών αναλύσεων της εικόνας με έναν αξιωματικό τρόπο, εξασφαλίζει ευρωστία αναφορικά με λάθη σε μεμονωμένα επίπεδα.
- Ένας καλύτερος τρόπος να εκμεταλλευτούμε τη δομή τόσο στην εικόνα, όσο και στο χώρο των χαρακτηριστικών ίσως είναι να συνδυάσουμε την πολυεπίπεδη προσέγγιση πυραμίδας με τα οπτικά λεξικά σε ένα ενιαίο πλαίσιο.
- Επειδή οι χωρικές πυραμίδες βασίζονται σε χαρακτηριστικά υπολογισμένα στην αρχική ανάλυση της εικόνας, φαίνεται να εντοπίζουν την οργάνωση βασικών εικονογραφικών στοιχείων ή ομοιόμορφες περιοχές (blobs), καθώς και την κατευθυντικότητα των επικρατουσών ευθειών και ακμών και γενικά λεπτομερειών υψηλής συχνότητας.
- Μειονέκτημα αποτελεί η δυσκολία εύρεσης καθολικών χαρακτηριστικών σε εικόνες με έντονη γεωμετρική ποικιλομορφία.

4.2.4 Γρήγορο χωρικό ταίριασμα

Η μέθοδος *Γρήγορο Χωρικό Ταίριασμα (Fast Spatial Matching)* [17] αποτελεί ουσιαστικά μια περίπτωση εφαρμογής του αλγορίθμου LO-RANSAC με βήμα τοπικής βελτιστοποίησης LO Επαναληπτικό Εσωτερικό RANSAC που αναλύσαμε σε προηγούμενη ενότητα. Συγκεκριμένα στην ανάκτηση εικόνων που μελετάμε, περιλαμβάνει την εξαγωγή υποθέσεων ενός προσεγγιστικού μοντέλου και στη συνέχεια επανεκτιμά επαναληπτικά νέες υποθέσεις με βάση το σύνολο των σημείων δεδομένων. Επιλέγοντας έναν περιορισμένο αριθμό μετασχηματισμών για την εξαγωγή υποθέσεων και εκμεταλλευόμενοι πληροφορία όσον αφορά στο σχήμα στις αφινικά ανεξάρτητες περιοχές μπορούμε να δημιουργήσουμε υποθέσεις μόνο με μεμονωμένα ζεύγη αντίστοιχων χαρακτηριστικών. Γι αυτό το λόγο απαριθμούμε όλες τις υποθέσεις, αφαιρώντας έτσι την τυχαιότητα από τον αλγόριθμο. Να σημειώσουμε ότι τα χαρακτηριστικά αποτελούν οπτικές λέξεις εξαγόμενες με τον αλγόριθμο συσταδοποίησης *AKM (Approximate k-means)*.

Transformation	dof	Matrix
translation + isotropic scale	3	$\begin{bmatrix} a & 0 & t_x \\ 0 & a & t_y \end{bmatrix}$
translation + anisotropic scale	4	$\begin{bmatrix} a & 0 & t_x \\ 0 & b & t_y \end{bmatrix}$
translation + vertical shear	5	$\begin{bmatrix} a & 0 & t_x \\ b & c & t_y \end{bmatrix}$

(a)



(b)

Σχήμα 4.4: Τα τρία αφινικά υποσύνολα που χρησιμοποιούνται στη χωρική ανακατάταξη και ο υπολογισμός του H ως $H_2^{-1}H_1$ για την 5-dof περίπτωση.

Οι περιπτώσεις των αφινικών υποθέσεων που εξετάζονται φαίνονται στον σχήμα 4.4. Πιο συγκεκριμένα αναφέρουμε κάποια στοιχεία για κάθε υπόθεση.

1. **3 βαθμοί ελευθερίας (3-dof)**. Υπολογίζεται σε μια περιοχή γύρω από κάθε αντιστοιχία χρησιμοποιώντας το κεντροειδές της περιοχής για την εκτίμηση της μεταφοράς

(translation) και την κλίμακα κάθε περιοχής για την εκτίμηση της ισοτροπικής αλλαγής κλίμακας μεταξύ της εικόνας αναζήτησης και της εκάστοτε υψηλά σε κατάταξη εικόνας βάσης (ιδανικός για περιπτώσεις αλλαγής στη μεγέθυνση (zoom) της κάμερας ή της απόστασής της από τη σκηνή).

2. **4 βαθμοί ελευθερίας (4-dof)**. Από μια περιοχή γύρω από μια μεμονωμένη αντιστοιχία, η κλίμακα στην κατεύθυνση x υπολογίζεται από το λόγο των ορίων της περιοχής στη διάσταση x. Όμοια και για την κατεύθυνση y (ιδανικός για βράχυνση με οριζόντια ή κατακόρυφη κλιμάκωση).

3. **5 βαθμοί ελευθερίας (5-dof)**. Εκτιμάται από τη μεμονωμένη αντιστοιχία δύο ελλειπτικών περιοχών, $C_1 \leftrightarrow C_2$, για τις οποίες θεωρούμε τους μετασχηματισμούς H_1, H_2, οι οποίοι μετασχηματίζουν τις ελλείψεις σε μοναδιαίους κύκλους, έτσι ώστε ο προσανατολισμός του μοναδιαίου διανύσματος στην κατεύθυνση y να διατηρείται ($(0,1)^T$) αποτελεί ένα ιδιοδιάνυσμα του μετασχηματισμού). Ο συνολικός μετασχηματισμός δίνεται ως $H_2^{-1}H_1$ (διατηρεί την κατακόρυφη διεύθυνση και επιτρέπει την ανισοτροπική κλιμάκωση και την κατακόρυφη διάτμηση).

Σε κάθε περίπτωση στο επαναληπτικό βήμα επανεκτίμησης του LO-RANSAC χρησιμοποιείται ένας γενικός *μετασχηματισμός 6 βαθμών ελευθερίας (6-dof)*, ο οποίος εκτιμάται γύρω από κεντροειδή του τρέχοντος συνόλου των inliers, τα οποία προσδιορίζονται με τους προηγούμενους πιο απλούς μετασχηματισμούς.

Για τον υπολογισμό των inliers χρησιμοποιείται το *σφάλμα μεταφοράς δύο κατευθύνσεων με κατώφλι κλίμακας (two-way transfer error with scale threshold)*. Το ζητούμενο είναι να βρούμε το χωρικά πιο κοντινό χαρακτηριστικό που ανήκει στην ίδια οπτική λέξη με κάποιο αντίστοιχο χαρακτηριστικό της εικόνας αναζήτησης και να ελέγξουμε αν αυτή η μεμονωμένη απόσταση είναι μικρότερη από το κατώφλι. Η υλοποίηση της διαδικασίας αυτής περιλαμβάνει την κατασκευή ενός δισδιάστατου kd δέντρου, για την εξασφάλιση λογαριθμικού χρόνου αναζήτησης. Επίσης, θεωρούμε ότι η χωρική επαλήθευση είναι επιτυχής αν βρούμε έναν μετασχηματισμό με τουλάχιστον 4 inliers αντιστοιχίες. Στη συνέχεια ανακατατάσσουμε τις εικόνες με βάση το άθροισμα των idf τιμών των inliers οπτικών λέξεων και τοποθετούμε τις χωρικά επαληθευμένες πιο ψηλά σε κατάταξη από τις υπόλοιπες.

Συνολικά λοιπόν ο αλγόριθμος FSM βελτιώνει σημαντικά την απόδοση των μηχανών ανάκτησης εικόνων, με λίγο μικρότερο περιθώριο για μεγάλα λεξικά, καθώς ήδη παρέχουν υψηλή διακριτικότητα με αποτέλεσμα την επιβάρυνση του υπολογιστικού χρόνου χωρίς αντίστοιχη βελτίωση της επίδοσης. Επίσης, η απόδοση των διαφορετικών υποθέσεων δε διαφοροποιείται σημαντικά λόγου του επαναληπτικού βήματος που είναι κοινό και πιο καθοριστικό, με καλύτερη την 5-dof υπόθεση.

4.3 Ταίριασμα πυραμίδας Hough

Μέχρι τώρα όλες οι μέθοδοι χωρικού ταιριάσματος που εξετάστηκαν προσπαθούν να ανιχνεύσουν inliers, είτε βάσει υποθέσεων μετασχηματισμού, οι οποίες αξιοποιούν το τοπικό σχήμα των χαρακτηριστικών, είτε επιβάλλοντας περιορισμούς κατά ζεύγη για την εξαγωγή αντιστοιχιών (ευέλικτα μοντέλα (flexible models)), όπου οι χρόνοι επεξεργασίας παραμένουν ακόμη απαγορευτικοί για τη βελτίωση της επίδοσης της ανακατάταξης εικόνων στις μηχανές αναζήτησης.

Η τεχνική *Ταίριασμα Πυραμίδας Hough (Hough Pyramid Matching)* [20] αναπτύσσεται στο πλαίσιο ενός μοντέλου *χαλαρού, χωρικού ταιριάσματος (relaxed spatial matching)*, το οποίο βασίζεται στο σχήμα των τοπικών χαρακτηριστικών για την εξαγωγή ψήφων (votes) και είναι ανεξάρτητο των μετασχηματισμών ομοιότητας, δεν απαιτεί επαλήθευση καταμέτρησης των inliers και άρα προσαρμογή μοντέλου (model fitting) και είναι γραμμικό στον αριθμό των αντιστοιχιών. Επίσης εφαρμόζει ένα προς ένα αντιστοίχιση (one to one mapping) και είναι ευέλικτο, επιτρέποντας μη άκαμπτη κίνηση (non rigid motion) και πολλαπλό ταίριασμα επιφανειών και αντικειμένων.

4.3.1 Διατύπωση προβλήματος

Υποθέτουμε ότι μια εικόνα αναπαρίσταται από ένα σύνολο P τοπικών χαρακτηριστικών. Για κάθε χαρακτηριστικό $p \in P$, ορίζεται ένας μετασχηματισμός ομοιότητας αναφορικά με κάποια κανονικοποιημένη περιοχή (patch) της εικόνας, με τοπικό σχήμα και κλίμακα, που δίνεται από τον 3×3 πίνακα

$$F(p) = \begin{bmatrix} M(p) & \mathbf{t}(p) \\ \mathbf{0}^T & 1 \end{bmatrix}, \tag{4.5}$$

όπου $M(p) = \sigma(p)R(p)$ και $\sigma(p)$,$R(p)$,$\mathbf{t}(p)$ αντιστοιχούν στην ισοτροπική κλίμακα, τον προσανατολισμό και τη θέση, και ο $R(p)$ είναι ένας ορθογώνιος 2×2 πίνακας με $\det R(p) = 1$, που αναπαρίσταται από μια γωνία $\vartheta(p)$.

Δοθέντων δύο εικόνων P, Q, *μια ανάθεση (assignment) ή αντιστοιχία (correspondence)* $c = (p, q)$ είναι ένα ζεύγος χαρακτηριστικών $p \in P$, $q \in Q$. Ο σχετικός μετασχηματισμός από το p στο q είναι πάλι ένας μετασχηματισμός ομοιότητας, ο οποίος δίδεται ως

$$F(c) = F(p)F(q)^{-1} = \begin{bmatrix} M(c) & \mathbf{t}(c) \\ \mathbf{0}^T & 1 \end{bmatrix}, \tag{4.6}$$

όπου $M(c) = \sigma(c)R(c)$, $\mathbf{t}(c) = \mathbf{t}(q) - M(c)\mathbf{t}(p)$ και $\sigma(c) = \sigma(q)/\sigma(p)$, $R(c) = R(q)R(p)^{-1}$ είναι η σχετική κλίμακα και προσανατολισμός από το p στο q. Συνεπώς οι αναθέσεις αποτελούν σημεία σε ένα χώρο μετασχηματισμού $\mathcal{F}$ τεσσάρων διαστάσεων και περιγράφονται από ένα διάνυσμα παραμέτρων 4-dof μετασχηματισμού

$$f(c) = (x(c), y(c), \sigma(c), \vartheta(c)), \tag{4.7}$$

όπου $[x(c)y(c)]^T = \mathbf{t}(c)$ και $\vartheta(c) = \vartheta(q) - \vartheta(p)$.

Θεωρούμε ότι δύο χαρακτηριστικά είναι αντίστοιχα, όταν ανατίθενται στην ίδια οπτική λέξη:

$$C = \{(p, q) \in P \times Q : u(p) = u(q)\}, \tag{4.8}$$

όπου $u(p)$ είναι η κωδική λέξη ή οπτική λέξη του p. Αναφερόμαστε λοιπόν σε *μια πολλά προς πολλά αντιστοίχιση (many-to-many mapping)*. Δεδομένης μια ανάθεσης $c = (p, q)$, ορίζουμε την κωδική της λέξη $u(c) = u(p) = u(q)$.

Σε κάθε αντιστοιχία $c = (p, q) \in C$ αποδίδεται ένας συντελεστής βαρύτητας μέτρησης της σχετικής της σημασίας $w(c)$, που τυπικά εδώ είναι το idf της οπτικής λέξης. Δοθέντος ενός ζεύγους αναθέσεων $c, c' \in C$, θεωρούμε μια *βαθμολογία γειτνίασης (affinity score)* $\alpha(c, c')$, η οποία μετρά την ομοιότητα ως μια μη αύξουσα συνάρτηση της απόστασής στο

χώρο μετασχηματισμού. Επιπλέον, δύο αναθέσεις $c = (p, q), c' = (p', q')$ ονομάζονται *συμβατές (corresponding)* αν $p \neq p'$ και $q \neq q'$, και *συγκρουόμενες (conflicting)* διαφορετικά.

Το πρόβλημα *δυαδικού τετραγωνικού προγραμματισμού (binary quadratic programming)* προς επίλυση είναι η εύρεση ενός υποσυνόλου ζευγών συμβατών αναθέσεων, τα οποία μεγιστοποιούν τη συνολική σταθμισμένη βαθμολογία γειτνίασης των ζευγών.

4.3.2 Διαδικασία ταιριάσματος

Θεωρούμε ότι οι παράμετροι μετασχηματισμού είναι κανονικοποιημένοι ή μη γραμμικά αντιστοιχισμένοι στο $[0, 1]$. Οπότε ο χώρος μετασχηματισμού είναι $\mathcal{F} = [0, 1]^d$. Κατασκευάζουμε μια *ιεραρχική διαμέριση* $\mathcal{B} = B_0, \ldots, B_{L-1}$ του χώρου $\mathcal{F}$ σε L επίπεδα. Κάθε $B_\ell \in \mathcal{B}$ διαμερίζει τον $\mathcal{F}$ σε 2^{kd} *κάδους (bins)*(υπερκύβους), όπου $k = L - 1 - \ell$. Τα bins εξάγονται κβαντίζοντας ομοιόμορφα κάθε παράμετρο, ή διαμερίζοντας κάθε διάσταση σε 2^k ίσα διαστήματα μήκους 2^{-k}. B_0 είναι το πιο λεπτομερές επίπεδο (πυθμένας), ενώ το B_{L-1} είναι το πιο τραχύ επίπεδο (κορυφή). Κάθε διαμέριση B_ℓ αποτελεί μια λεπτή διαμέριση του $B_{\ell+1}$.

Ξεκινώντας με ένα σύνολο υποθετικών αντιστοιχιών των εικόνων P, Q, κατανείμουμε τις αντιστοιχίες σε bins με ένα *ιστόγραμμα πυραμίδας (histogram pyramid)*. Δοθέντος ενός bin b, θεωρούμε

$$h(b) = \{c \in C : f(c) \in b\} \tag{4.9}$$

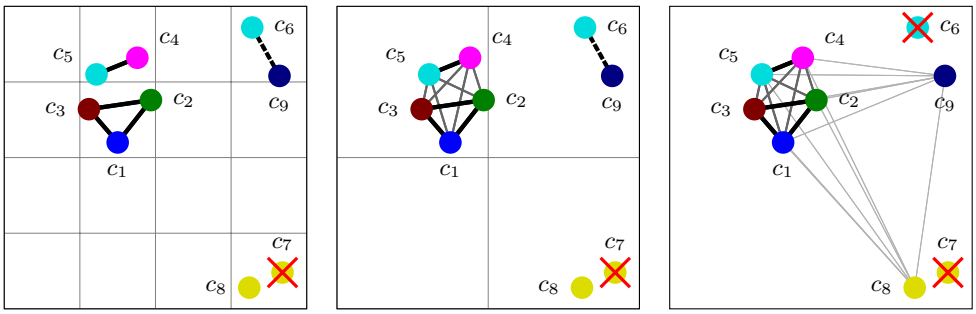
το σύνολο των αντιστοιχιών με διανύσματα παραμέτρων να εμπίπτουν στο b, και $|h(b)|$ το πλήθος του.

Στη συνέχεια περιγράφουμε τα στάδια της διαδικασίας χωρικού ταιριάσματος, η οποία συνοψίζεται στον υπολογισμό μιας *ενίσχυσης (strength)*, όσον αφορά στο σύνολο C των εικόνων P, Q, και κατ' επέκταση της ομοιότητάς τους. Χωρίζουμε αναδρομικά τις αντιστοιχίες σε bins με βάση μια *από την κορυφή και προς τα κάτω λογική (top-down-fashion)* και στη συνέχεια τα ομαδοποιούμε πάλι αναδρομικά με βάση μια *από τον πυθμένα προς τα πάνω λογική (bottom-up-fashion)*. Προκειμένου να επιβάλλουμε μια ένα προς ένα αντιστοίχιση, σε κάθε επίπεδο διαγράφουμε μια από τις συγκρουόμενες αντιστοιχίες με τρόπο που θα εξηγήσουμε πιο κάτω. Έστω X_ℓ το σύνολο των αντιστοιχιών που έχουν διαγραφεί μέχρι το επίπεδο ℓ. Αν $b \in B_\ell$ ένα bin στο επίπεδο ℓ, το σύνολο των αντιστοιχιών που κρατάμε στο b είναι $\hat{h}(b) = h(b) \setminus X_\ell$. Συνεπώς, ορίζουμε ως *πλήθος ομάδας* ενός bin b την ποσότητα

$$g(b) = \max\{0, |\hat{h}(b)| - 1\} \tag{4.10}$$

Έστω $b_0 \subseteq \ldots \subseteq b_\ell$ μια ακολουθία από bins που περιέχουν την αντιστοιχία c σε διαδοχικά επίπεδα μέχρι το επίπεδο ℓ, έτσι ώστε $b_k \in B_k$ για $k = 0, \ldots, \ell$. Για κάθε k, εκτιμούμε τη γειτνίαση (affinity) $\alpha(c, c')$ του c ως προς κάθε άλλη αντιστοιχία $c' \in b_k$, ως μια μη αρνητική και μη αύξουσα συνάρτηση *γειτνίασης επιπέδου (level affinity)* του k, την $\alpha(k) = 2^{-k}$, ούτως ώστε η γειτνίαση να είναι αντιστρόφως ανάλογη του μεγέθους του bin, βασιζόμενοι στη λογική ότι τα πιο μικρά bins σε πιο χαμηλά επίπεδα είναι πιο πιθανόν να περιέχουν αντιστοιχίες που είναι πράγματι inliers. Επιπλέον, οι νέες αντιστοιχίες που ενσωματώνονται στη c σε μια ομάδα στο επίπεδο k είναι $g(b_k) - g(b_{k-1})$. Συνεπώς η ενίσχυση του c μέχρι το επίπεδο ℓ θα είναι

$$s_\ell(c) = g(b_0) + \sum_{k=1}^{\ell} 2^{-k}\{g(b_k) - g(b_{k-1})\} \tag{4.11}$$



Σχήμα 4.5: Ταίριασμα 9 αναθέσεων μιας πυραμίδας 3 επιπέδων στο δισδιάστατο χώρο. Τα χρώματα υποδεικνύουν τις οπτικές λέξεις, και το πάχος των ακμών τη γειτνίαση. Η διακεκομμένη γραμμή μεταξύ των c_6, c_9 υποδεικνύει μια ομάδα η οποία έχει σχηματιστεί στο επίπεδο 0 και διασπάστηκε στη συνέχεια στο επίπεδο 2, αφότου διαγράφηκε η c_6.

Συνολικά λοιπόν η ενίσχυση του c είναι απλά η ενίσχυση στο κορυφαίο επίπεδο, δηλαδή $s(c) = s_{L-1}(c)$. Αφαιρώντας και τις αναθέσεις που έχουν διαγραφεί $X = X_{L-1}$ και λαμβάνοντας υπόψιν και τους συντελεστές βαρύτητας, ορίζουμε τη *βαθμολογία ομοιότητας (similarity score)* των εικόνων P, Q ως

$$s(C) = \sum_{c \in C \setminus X} w(c)s(c) \tag{4.12}$$

Στο σημείο αυτό είμαστε σε θέση να ορίσουμε τα κριτήρια διαγραφής των συγκρουόμενων αντιστοιχιών , δηλαδή αυτών με την ίδια κωδική λέξη και άρα τον ίδιο συντελεστή βαρύτητας. Διαγράφουμε σε κάθε επίπεδο τη συγκρουόμενη αντιστοιχία με τη χαμηλότερη μέχρι το δεδομένο επίπεδο ενίσχυση. Στην περίπτωση του πιο χαμηλού επιπέδου (του 0) ή ίσων ενισχύσεων διαγράφουμε με τυχαίο τρόπο επιλογής.

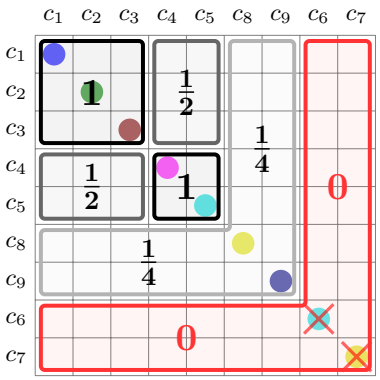
Στο σχήμα 4.5 παραθέτουμε ένα παράδειγμα με παιχνίδια για την κατανόηση της μεθόδου υποθέτοντας μια διάταξη των χαρακτηριστικών/αναθέσεων. Επίσης, στο σχήμα 4.6 αποσαφηνίζεται η διαμόρφωση της βαθμολογίας ομοιότητας των δύο εικόνων P, Q, με βάση τις ενισχύσεις μεμονωμένων αναθέσεων, καθώς και ο τρόπος με τον οποίο οι συνεισφορές όλων των επιπέδων στην ενίσχυση κάθε αντιστοιχίας μπορούν να οργανωθούν σε έναν $n \times n$ πίνακα γειτνίασης A, ώστε το άθροισμα ανά γραμμή του πίνακα A να ισούται με την ενίσχυση της αντίστοιχης ανάθεσης. Επίσης η αναπαράσταση αυτή καταδεικνύει την κατά ζεύγη προσέγγιση της μεθόδου.

4.3.3 Αλγόριθμος

Στη συνέχεια παραθέτουμε τον αλγόριθμο υλοποίησης της μεθόδου 3, όπου υλοποιείται αναδρομικά η διαδικασία του από την κορυφή προς τα κάτω χωρίσματος, και στη συνέχεια η ομαδοποίηση από τον πυθμένα προς τα πάνω των αντιστοιχιών. Γενικά η αποθήκευση στα bins είναι αραιή και γραμμική στο πλήθος $|C|$. Επίσης όσον αφορά στη διαδικασία διαγραφής, για όλες τις αναθέσεις που περιέχονται σε ένα bin, βρίσκουμε το σύνολο των κοινών οπτικών λέξεων και κρατάμε την κωδική λέξη με την πιο ισχυρή ανάθεση (τη μεγαλύτερη ενίσχυση), διαγράφουμε τις υπόλοιπες και ανανεώνουμε το X. Δεδομένου ότι πράξεις που εκτελούνται σε κάθε αναδρομική κλήση σε ένα bin είναι γραμμικές στο $|h(b)|$ και ότι ανά επίπεδο ℓ είναι

	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			$\frac{1}{4}6w(c_9)$

(α)



(β)

Σχήμα 4.6: (α)Ετικέτες αναθέσεων, χαρακτηριστικά και βαθμολογίες ενίσχυσης αναφορικά με το σχήμα 4.5. Εδώ κορυφές και ακμές υποδεικνύουν τα χαρακτηριστικά (στις εικόνες P, Q) και τις αναθέσεις αντίστοιχα. Οι αναθέσεις c_5, c_6 είναι συγκρουόμενες, όντας της μορφής $(p, q), (p, q')$. Ομοίως για τις c_7, c_8. Οι αναθέσεις $c_1, \ldots, c_5$ μετέχουν σε ομάδες στο επίπεδο 0; c_8, c_9 στο επίπεδο 2; και οι c_6, c_7 διαγράφονται. (β) Πίνακας γειτνίασης ισοδύναμος με τις ενισχύσεις του σχήματος 4.5. Οι αναθέσεις έχουν αναδιαταχθεί έτσι ώστε, οι ομάδες να εμφανίζονται σε συνεχόμενα κουτιά (blocks). Στις ομάδες που σχηματίστηκαν στα επίπεδα $0, 1, 2$ αποδίδεται γειτνίαση $1, \frac{1}{2}, \frac{1}{4}$ αντίστοιχα. Οι βαθμολογίες ομοιότητας του σχήματος (α) μπορούν να εξαχθούν αθροίζοντας τις γειτνιάσεις ανά γραμμή και πολλαπλασιάζοντας με συντελεστές βαρύτητας ανάθεσης.

γραμμικές στο $n = |C|$, ο αλγόριθμος έχει πολυπλοκότητα $O(nL)$.

4.3.4 Παρατηρήσεις

Συνοψίζουμε λοιπόν τα παρακάτω συμπεράσματα σχετικά με τη μέθοδο HPM.

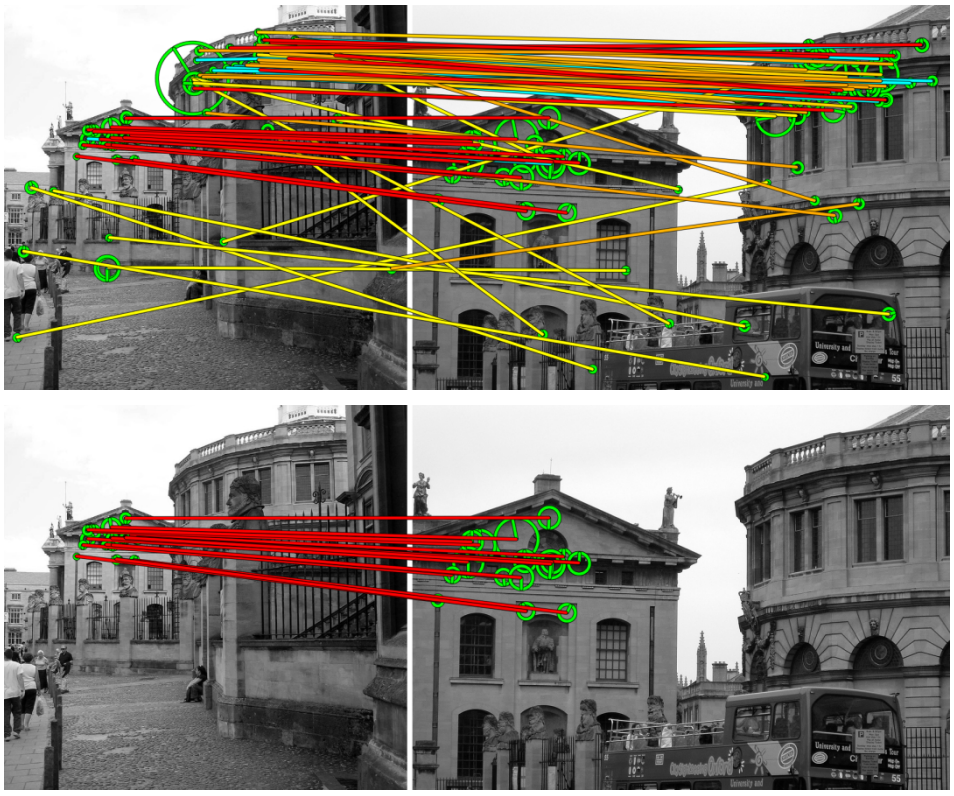
- Εφαρμόζει ένα επί ένα αντιστοίχιση επιτρέποντας παράλληλα ομαδοποίηση των αντιστοιχίσεων με βάση τη γειτνίαση τους, ενώ παράλληλα διατηρεί την ανεξαρτησία από μετασχηματισμούς ομοιότητας. Έτσι δεν απαιτεί εκμάθηση και την επιλογή κάποιας βέλτιστης υπόθεσης.
- Με βάση τη λογική ταιριάσματος πυραμίδας, προσεγγίσει τη γειτνίαση με το μέγεθος των bin, έτσι δεν απαιτείται η απαρίθμηση ζευγών αντιστοιχιών.

Τα παραπάνω χαρακτηριστικά καταδεικνύουν τον HPM ως μια μέθοδο που εκμεταλλεύεται τη χωρική πληροφορία με βέλτιστο και αποδοτικό τρόπο, δίνοντας παράλληλα τη δυνατότητα ενσωμάτωσης γρήγορων και εύκολα υλοποιήσιμων αλγορίθμων για την ανακατάταξη εικόνων σε μηχανές αναζήτησης.

Algorithm 3 Hough Pyramid Matching

procedure HPM(assignments C, levels L)
  $X \leftarrow \emptyset$; $\mathcal{B} \leftarrow$ PARTITION(L)
  for all $c \in C$ **do** $s(c) \leftarrow 0$
  HPM-REC($C, L-1$)
  return score $\sum_{c \in C \setminus X} w(c)s(c)$
end procedure

procedure HPM-REC(assignments C, level ℓ)
  if $|C| < 2 \vee \ell < 0$ **return**
  for all $b \in \mathcal{B}_\ell$ **do** $h(b) \leftarrow \emptyset$
  for all $c \in C$ **do** $h(\beta_\ell(c)) \leftarrow h(\beta_\ell(c)) \cup c$
  for all $b \in \mathcal{B}_\ell$ **do** HPM-REC($h(b)$,$\ell - 1$)
  for all $b \in \mathcal{B}_\ell$
    $X \leftarrow X \cup$ ERASE $(h(b))$
    $h(b) \leftarrow h(b) \cup \setminus X$
    if $h(b) < 2$ **continue**
    if $\ell = L - 1$ **then** $\alpha \leftarrow 2$ **else** $\alpha \leftarrow 1$
    for all $c \in h(b)$ **do** $s(c) \leftarrow s(c) + \alpha 2^{-\ell}|h(b)|$
    end for
end procedure



Σχήμα 4.7: **Πάνω**: HPM ταίριασμα δύο εικόνων από το Oxford σύνολο δεδομένων. Απεικονίζονται όλες οι πιθανές αντιστοιχίσεις. Οι γαλάζιες διαγράφονται. Οι υπόλοιπες χρωματίζονται με βάση την ενίσχυσή τους με το κόκκινο να περιγράφεις τις ισχυρότερες και το κίτρινο τις πιο ασχθενείς. **Κάτω**: Inliers που έχουν βρεθεί με ένα μετασχηματισμό $4 - dof$ FSM και αφινικό μοντέλο LO-RANSAC.

Κεφάλαιο 5

ΜΔΥ με πυρήνα πυραμίδας Hough

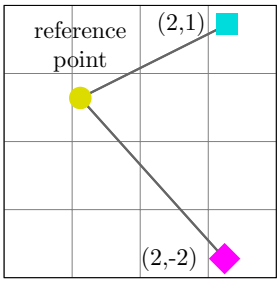
5.1 Εισαγωγή

Το πρόβλημα που καλούμαστε να λύσουμε είναι η αποδοτική κατηγοριοποίηση εικόνων σε μεγάλες βάσεις δεδομένων, στο πλαίσιο γρήγορων μηχανών αναζήτησης. Τα SVM που αναλύσαμε προηγουμένως παρέχουν αποδοτικούς αλγορίθμους εκμάθησης με πυρήνες ακόμη και μη διαχωρίσιμων δεδομένων, σε χώρους πολλών διαστάσεων, ενώ παράλληλα ο αλγόριθμος χαλαρού και ευέλικτου χωρικού ταιριάσματος HPM συντελεί στην αποτελεσματική και αποδοτική ενσωμάτωση των χωρικών χαρακτηριστικών των εικόνων κατά τη διαδικασία ανάκτησής τους σε μια μηχανή αναζήτησης. Γεννάται λοιπόν το ερώτημα αν θα μπορούσαμε να εκμεταλλευτούμε την ισχυρή διακριτική ικανότητα των SVM σε χώρους πολλών διαστάσεων, εισάγοντας σε αυτά τον αλγόριθμο HPM ως πυρήνα, ο οποίος ενισχύει γενικά την αποτελεσματικότητα εκμάθησης των όμοιων εικόνων λόγω της χωρικής συσχέτισης που παρέχει, παραμένοντας ανεξάρτητος του χωρικού μετασχηματισμού. Αυτός είναι και ο βασικός άξονας αυτού του κεφαλαίου, που παράλληλα συνοψίζει τη μέθοδο που εισάγουμε για την κατηγοριοποίηση εικόνων.

5.2 Πυρήνας πυραμίδας Hough

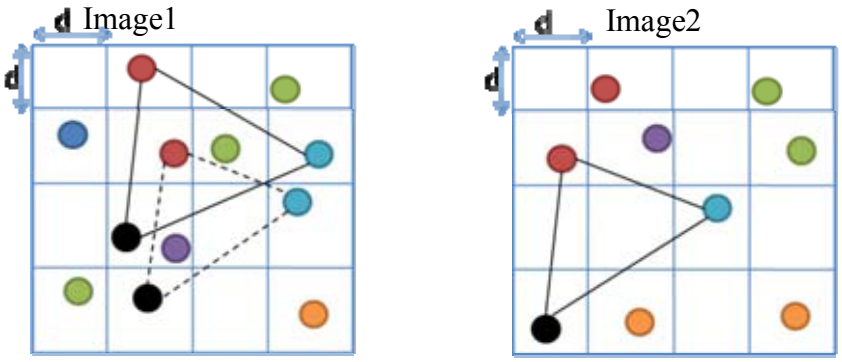
Για την υλοποίηση της παραπάνω ιδέας καταρχήν καλούμαστε να αποδείξουμε ότι ο αλγόριθμος HPM είναι πυρήνας. Θα προσπαθήσουμε λοιπόν να εκφράσουμε τη συνολική ενίσχυση που υπολογίσαμε για το σύνολο των αντιστοιχιών δύο εικόνων με τον αλγόριθμο HPM, ως μιας μορφής εσωτερικό γινόμενο χαρακτηριστικών των δύο εικόνων.

Ορίζουμε ως *χαρακτηριστικά n-οστής τάξης (n^{th} order features)* [21] τα χαρακτηριστικά που έχουν n κοινές οπτικές λέξεις. Αν $\mathcal{U}$ το οπτικό λεξικό, ορίζουμε το χαρακτηριστικό n-οστής τάξης f_n, ως μια n-άδα οπτικών λέξεων από το $\mathcal{U}$ με συγκεκριμένη χωρική διάταξη. Έστω $\mathcal{F}_n$ το σύνολο όλων των δυνατών συνδυασμών από n οπτικές λέξεις και όλες τις δυνατές διατάξεις, με $|\mathcal{F}_n|$ το πλήθος του. Ουσιαστικά $\mathcal{F}_n = \{\mathcal{U} \times F\}$, όπου F ο χώρος των δυνατών διατάξεων των τοπικών χαρακτηριστικών της εικόνας.



Σχήμα 5.1: Διάταξη ομάδας αντιστοιχιών.

Προκειμένου να ορίσουμε μια διάταξη αναφορικά με μια ομάδα αντιστοιχιών σε μια εικόνα, επιλέγουμε ένα από τα τοπικά χαρακτηριστικά της ομάδας ως σημείο αναφοράς, για παράδειγμα το τοπικό χαρακτηριστικό με το μικρότερο αύξοντα αριθμό οπτικής λέξης, και στη συνέχεια υπολογίζουμε τη θέση των υπολοίπων της ιδίας ομάδας με βάση ένα πλαίσιο *σχετικής κβάντισης* (σχήμα 5.1). Ο όρος *σχετική κβάντιση (relative quantization)* εισάγεται στην προκείμενη ανάλυση σε επίπεδο αντιστοιχιών και αφορά ζεύγη τοπικών χαρακτηριστικών, ενώ διαφοροποιείται από την *απόλυτη κβάντιση (absolute quantisation)*, η οποία εκφράζει το απόλυτο του ορισμού της θέσης σε επίπεδο τοπικών χαρακτηριστικών.



Σχήμα 5.2: Παράδειγμα διάταξης στο χώρο των τοπικών χαρακτηριστικών της εικόνας. Τα χρώματα δηλώνουν διαφορετικές οπτικές λέξεις [21].

Αν κβαντίσουμε το χώρο της εικόνας με βάση μια σταθερή απόσταση d, τότε το σχήμα κβάντισης διατηρείται αναλλοίωτο ως προς τη μετατόπιση στο χώρο των αντιστοιχιών. Στο χώρο Hough όμως που αναφερόμαστε εμείς, θα πρέπει με κάποιο τρόπο να λάβουμε υπόψιν τόσο την κλίμακα, όσο και την περιστροφή όσον αφορά στο αναλλοίωτο των διαφόρων περιγραφών, καθώς περιλαμβάνει 4 βαθμούς ελευθερίας. Μια λύση αποτελεί να απεικονίσουμε το χώρο Hough σε κάποιον άλλο 2 διαστάσεων και στη συνέχεια να θεωρήσουμε νέες αντιστοιχίες, οπότε θα εξασφαλίζεται το αναλλοίωτο. Άλλη προσέγγιση θα ήταν να λάβουμε υπόψιν, στον ορισμό των δυνατών διατάξεων, πέραν της θέσης την κλίμακα και την περιστροφή, και κατά αυτό τον τρόπο θα θεωρούσαμε χώρους αντιστοιχιών πολλών διαστάσεων. Το αποτέλεσμα αυτό είναι ιδιαίτερα σημαντικό καθώς για πρώτη φορά παρέχεται ένα πλαίσιο διατήρησης

του αναλλοίωτου των αντιστοιχιών στο χώρο Hough.

Με βάση την ανάλυση του αλγορίθμου HPM που προηγήθηκε 4.3.1, έστω C ο χώρος των αντιστοιχιών των εικόνων με σύνολο τοπικών χαρακτηριστικών P, Q, $u(c)$ η οπτική λέξη της αντιστοιχίας $c \in C$ και $w(c) = idf(u(c))$ ο συντελεστής βαρύτητας μέτρησης της σχετικής σημασίας της $u(c)$. Επιπλέον, ορίζουμε $h(c) \subset C$ την ομάδα στην οποία ανήκει η αντιστοιχία $c \in C$ με

$$\mathcal{H} = \{h(c) | c \in C\}, \tag{5.1}$$

όλες τις ομάδες στο χώρο C. Για κάθε ομάδα $h \in \mathcal{H}$, ορίζουμε το *πλήθος ομάδας (group count)* ως

$$\eta(h) = \max\{0, |h| - 1\}, h \subset C \tag{5.2}$$

Έστω F_c ο χώρος όλων των δυνατών διατάξεων στο χώρο C, όπως αυτός προκύπτει με κάποια μέθοδο σχετικής κβάντισης, δεδομένου ότι περιλαμβάνει διατάξεις που ορίζονται επί αντιστοιχιών τοπικών χαρακτηριστικών. Ορίζουμε $f(h)$ το σύνολο των διατάξεων της ομάδας $h \in \mathcal{H}$ και $\mathcal{H}(f)$ το σύνολο όλων των ομάδων που έχουν ανατεθεί στο σύνολο διατάξεων $f \in F_c$.

Για λόγους απλούστευσης θεωρούμε αρχικά ότι η πυραμίδα Hough έχει ένα μόνο επίπεδο, δηλαδή $\ell = 0$, όπου ℓ ο δείκτης του επιπέδου, ότι δε διαγράφουμε τις συγκρουόμενες αντιστοιχίες και ότι αναφερόμαστε σε ένα μόνο σύνολο διατάξεων. Στην περίπτωση αυτή η σχέση 4.11 ανάγεται στην $s(c) = g(c)$ και σε αντιστοιχία με τη σχέση 4.12 έχουμε

$$\begin{aligned} s(C) &= \sum_{c \in C} w(c) s(c) \\ &= \sum_{h \in \mathcal{H}} \eta(h) \sum_{c \in h} w(c) \end{aligned} \tag{5.3}$$

Αν θεωρήσουμε όλα τα δυνατά μεγέθη ομάδων n, το οποίο εκφράζει και τον αριθμό των κοινών οπτικών λέξεων των τοπικών χαρακτηριστικών πλέον που μετέχουν σε μια ομάδα, η ενίσχυση του συνόλου των αντιστοιχιών και βαθμολογία ομοιότητας των δύο εικόνων διαμορφώνεται ως

$$s(C) = \sum_{n=1}^{\infty} (n-1) \sum_{\substack{h \in \mathcal{H} \\ |h|=n}} \sum_{c \in h} w(c) \tag{5.4}$$

Παρατηρούμε δηλαδή, ότι η βαθμολογία ομοιότητας των δύο εικόνων ανάγεται, με βάση μια πιο γενική λογική αντιστοίχισης των τοπικών χαρακτηριστικών τους, σε ένα άθροισμα εξαρτώμενο πλέον από ομάδες αντιστοιχιών.

Αν συμπεριλάβουμε και όλα τα δυνατά σύνολα διατάξεων για κάθε ομάδα, η σχέση 5.4 γίνεται

$$\begin{aligned} s(C) &= \sum_{n=1}^{\infty} (n-1) \sum_{\substack{f \in F_c \\ |f|=n}} \sum_{\substack{h \in \mathcal{H} \\ |h|=n}} \sum_{c \in h} w(c) \\ &= \sum_{f \in F_c} (|f| - 1) \left[\sum_{\substack{h \in \mathcal{H} \\ f(h)=f}} \sum_{c \in h} w(c) \right] \\ &= \sum_{f \in F_c} (|f| - 1) w(f) |\mathcal{H}(f)|, \end{aligned} \tag{5.5}$$

όπου $w(f) = \sum_{c \in h} w(c)$. Καταλήγουμε πλέον σε μια έκφραση, η οποία εξαρτάται μόνο από το σύνολο διατάξεων f, δεδομένου ότι σε κάθε ομάδα αντιστοιχεί ένα σύνολο διατάξεων, με πλήθος ίσο με n.

Έστω H_x^f, H_y^f τα ιστογράμματα ως προς f όλων των ομάδων που σχηματίζονται στο επίπεδο αντιστοιχιών C των τοπικών χαρακτηριστικών δύο εικόνων x, y. Τότε θα ισχύει:

$$|\mathcal{H}(f)| = (H_x^f)' H_y^f \tag{5.6}$$

Στο σημείο αυτό, προκειμένου να ολοκληρώσουμε την απόδειξη του πυρήνα HPM, καλούμαστε να δείξουμε ότι το $|\mathcal{H}(f)|$ ισούται με κάποιο εσωτερικό γινόμενο των αρχικών τοπικών χαρακτηριστικών των δύο εικόνων. Στην κατεύθυνση αυτή, επιστρέφουμε στα χαρακτηριστικά n-οστής τάξης, που είχαμε ορίσει στην αρχή της ενότητας και θεωρούμε ένα n-οστής τάξης διάνυσμα μετασχηματισμού $\phi^n(x)$ της εικόνας x, με συντεταγμένες $\phi_f^n(x)$ τέτοια ώστε αν $f_n \in \mathcal{F}_n$ και F_x το σύνολο των διαφορετικών χαρακτηριστικών f_n που εντοπίζονται στο χώρο P της εικόνας,

$$\phi_f^n(x) = h_{f_n}(x), \tag{5.7}$$

όπου $h_{f_n}(x)$ η συχνότητα εμφάνισης του στοιχείου $f_n \in F_x$. Επομένως ουσιαστικά μετράμε πόσες φορές εμφανίζεται ένα δεδομένο χαρακτηριστικό f_n στο σύνολο των οπτικών λέξεων και των διατάξεων που περιγράφουν την εικόνα x. Κατασκευάζουμε κατά αυτό τον τρόπο τον πυρήνα

$$\begin{aligned} K_n(x,y) &= \langle \phi^n(x), \phi^n(y) \rangle \\ &= \sum_{f_n \in \mathcal{F}_n} \langle \phi_f^n(x), \phi_f^n(y) \rangle \\ &= \sum_{f_n \in \mathcal{F}_n} h_{f_n}(x) \times h_{f_n}(y) \end{aligned} \tag{5.8}$$

Θεωρούμε ότι το σχήμα σχετικής και απόλυτης κβάντισης στο χώρο F και F_c δε διαφοροποιείται σημαντικά και έχουμε κατά νού ότι στα πολλά επίπεδα, οι αποκλίσεις κβάντισης αντισταθμίζονται. Τότε οι δύο τελευταίες σχέσεις εκφράζουν ακριβώς την ίδια ποσότητα, τον αριθμό δηλαδή όλων των ομάδων τοπικών χαρακτηριστικών των δύο εικόνων για ένα δεδομένο σύνολο διατάξεων f, ο οποίος ισούται με τον αριθμό όλων των χαρακτηριστικών $f_n \in \mathcal{F}_n$ που εμφανίζονται στις δύο εικόνες. Επομένως η σχέση 5.5 γράφεται ως

$$\begin{aligned} s(C) &= \sum_{f \in F} (|f| - 1) w(f) \langle \phi_f^n(x), \phi_f^n(y) \rangle \\ &= \sum_{f \in F} (|f| - 1) w(f) K_{|f|}(x,y) \end{aligned} \tag{5.9}$$

όπου $w(f) = \sum_{c \in h} w(c)$ και $K_{|f|}(x,y)$ είναι ισοδύναμο με τον πυρήνα $K_n(x,y)$ δεδομένου ότι $|f| = n$ και ότι το σύνολο των δυνατών διατάξεων ομάδων τοπικών χαρακτηριστικών και το σύνολο δυνατών διατάξεων των χαρακτηριστικών n-οστής τάξης ταυτίζεται.

Η παραπάνω ανάλυση γενικεύεται εύκολα και για περισσότερα επίπεδα, δεδομένου ότι τα επίπεδα συμμετέχουν στη συνολική βαθμολογία ομοιότητας, με το άθροισμα των ενισχύσεων τους απλά με τη διαφοροποίηση κάποιου συντελεστή βαρύτητας εξαρτώμενο από το επίπεδο (Παράρτημα Α′).

Συνολικά λοιπόν και με βάση τις ιδιότητες των πυρήνων 3.3.2, εκφράσαμε τη βαθμολογία ομοιότητας του αλγορίθμου HPM ως εσωτερικό γινόμενο χαρακτηριστικών των δύο εικόνων και επομένως αποτελεί πυρήνας και μάλιστα πυρήνας Hough. Στο σημείο αυτό μπορούμε να ενσωματώσουμε τον πυρήνα HPM σε SVM.

5.3 Κατηγοριοποίηση εικόνων με ΜΔΥ και πυρήνα πυραμίδας Hough

Με βάση την απόδειξη της προηγούμενης ενότητας, εισάγουμε τον αλγόριθμο HPM ως πυρήνα σε ν-SVM προκειμένου να εκπαιδεύσουμε μεμονωμένους ταξινομητές για κάθε κατηγορία. Στη συνέχεια συνδυάζουμε τις εξόδους τους με βάση μια λογική ταξινόμησης πολλών κλάσεων, *μια κλάση έναντι των υπολοίπων* , όπως περιγράψαμε στην ενότητα 3.7.

Η μέθοδος κατηγοριοποίησης εικόνων με πυρήνα Hough πυραμίδας υλοποιείται στα εξής στάδια:

- Χωρίζουμε τα δεδομένα σε *σύνολα εκπαίδευσης (train sets)* και σε *σύνολα ελέγχου (test sets)*. Τα σύνολα ελέγχου των μεμονωμένων ταξινομητών περιλαμβάνουν ίσο αριθμό *θετικών (positives)* και *αρνητικών (negatives) δειγμάτων* για λόγους συμμετρίας και όπως προέκυψε από πειράματα βελτιστοποίησης του αριθμού των αρνητικών στα σύνολα ελέγχου. Επίσης όσον αφορά στο σύνολο ελέγχου για το στάδιο των πολλών κλάσεων, παίρνουμε ίσο αριθμό δειγμάτων από κάθε κλάση πάλι για λόγους συμμετρίας.

- Υπολογίζουμε τις τιμές του πυρήνα HPM των δεδομένων ελέγχου και εκπαίδευσης, με βάση το ανεστραμμένο αρχείο για λόγους ευκολίας και ταχύτητας. Δηλαδή φορτώνουμε κάθε φορά στη βάση τις εικόνες με τις οποίες θέλουμε να υπολογίσουμε την ομοιότητα, τις εικόνες εκπαίδευσης, και θεωρούμε ως εικόνες αναζήτησης τις εικόνες για τις οποίες θέλουμε να υπολογίσουμε τις τιμές του HPM κάθε φορά, είτε εκπαίδευσης είτε ελέγχου επί των εικόνων βάσης. Ο αλγόριθμος HPM σε συνδυασμό με το ανεστραμμένο αρχείο μας επιστρέφει τις πιο υψηλά σε κατάταξη εικόνες βάσης και τις αντίστοιχες τιμές ομοιότητας σε κάθε περίπτωση. Οι πιο υψηλά σε κατάταξη εικόνες προκύπτουν εφαρμόζοντας BOW φιλτράρισμα, δηλαδή επιστρέφονται οι βαθμολογίες ομοιότητας των εικόνων βάσης με τις πιο πολλές κοινές οπτικές λέξεις με την εικόνα αναζήτησης. Στην περίπτωση μας, ανακτούμε όλες τις εικόνες βάσης δεδομένου του περιορισμένου μεγέθους των εικόνων βάσης.

- Προτού εισάγουμε τους υπολογισμένους πυρήνες HPM εκπαίδευσης στα μεμονωμένα *nu*-SVM, εφαρμόζουμε κάποια επεξεργασία, ώστε να εξασφαλίζεται 1) η συμμετρία τους, 2) η αντιμετώπιση του *προβλήματος της επικράτησης των στοιχείων της διαγωνίου (diagonal dominance)* και 3) η κανονικοποίηση των τιμών τους. Πιο συγκεκριμένα, 1) το πρόβλημα της επικράτησης των στοιχείων της διαγωνίου συνίσταται στην εμφάνιση υψηλών τιμών στη διαγώνιο συγκριτικά με το μέσο όρο του πυρήνα, στην περίπτωση υπολογισμού της ομοιότητας των δεδομένων εκπαίδευσης με τον εαυτό τους και αντιμετωπίζεται με διάφορες μεθόδους [10]. Εμείς χρησιμοποιούμε ως βέλτιστη τη *μέθοδο μετακίνησης της διαγωνίου (diagonal shift)*, η οποία περιλαμβάνει την αφαίρεση μιας σταθεράς από τα στοιχεία της διαγωνίου, και συγκεκριμένα της μέσης τιμής των στοιχείων της διαγωνίου. 2) Σχετικά με το πρόβλημα της συμμετρίας που εμφανίζεται τόσο

στον πυρήνα HPM των δεδομένων εκπαίδευσης, όσο και ελέγχου, υπολογίζουμε τον αντίστροφο πυρήνα HPM των εικόνων βάσης ως προς τις εικόνες αναζήτησης και κάθε στοιχείο του διορθωμένου πίνακα προκύπτει ως το άθροισμα των αντίστοιχων στοιχείων και προς τις δύο κατευθύνσεις. 3) Για κάθε μεμονωμένο ταξινομητή βελτιστοποιούμε μια παράμετρο κανονικοποίησης p ως προς την ακρίβεια εκπαίδευσης, η οποία αντιστοιχεί σε μια συνάρτηση κανονικοποίησης των τιμών του πυρήνα HPM της μορφής $f(x) = 1 - \exp(-px)$.

- Εκπαιδεύουμε τους μεμονωμένους ν-SVM ταξινομητές έχοντας βελτιστοποιήσει προηγουμένως ως προς την ακρίβεια εκπαίδευσης την τιμή της παραμέτρου ν.

- Εφαρμόζουμε ταξινόμηση πολλών κλάσεων 3.46, με τον εξής κανόνα απόφασης: επιλέγουμε κάθε φορά τη μέγιστη των εξόδων των μεμονωμένων ταξινομητών, ως την έξοδο της κατηγορίας στην οποία ανήκει το εκάστοτε δείγμα ελέγχου. Ο αύξων αριθμός της κατηγορίας αποτελεί και την *πρόβλεψη (ετικέτα πρόβλεψης)*, σχετικά με το δείγμα ελέγχου.

Να σημειώσουμε ότι κάθε διαδικασία βελτιστοποίησης περιλαμβάνει την εξαγωγή βέλτιστης τιμής παραμέτρου με βάση τη μέση τιμή της ακρίβειας σε έναν αριθμό διαφορετικών χωρισμάτων των δεδομένων σε σύνολα εκπαίδευσης και ελέγχου.

Κεφάλαιο 6

Πειραματική Αξιολόγηση

6.1 Εισαγωγή

Το τελευταίο αλλά και πολύ σημαντικό στάδιο της αποτίμησης κάθε ερευνητικής διαδικασίας είναι ουσιαστικά ο έλεγχος του πόσο καλά με βάση κάποια κριτήρια, μια νέα μέθοδος ανταποκρίνεται στην επίλυση του προβλήματος που επιδιώκει να επιλύσει. Εισάγοντας λοιπόν και μελετώντας κάποιους δείκτες αξιολόγησης είμαστε σε θέση να εξάγουμε συμπεράσματα σχετικά με την αποτελεσματικότητα της υποκείμενης μεθόδου, μέσω των πλεονεκτημάτων και μειονεκτημάτων που παρουσιάζει και σε σύγκριση με ήδη υπάρχουσες μεθόδους. Η αποτίμηση δε αυτή μπορεί να αποβεί ιδιαίτερα βοηθητική στη διαμόρφωση μια μελλοντικής ερευνητικής κατεύθυνσης.

6.2 Σύνολο Δεδομένων

Για τη διεξαγωγή των πειραμάτων κατηγοριοποίησης, κατασκευάσαμε ένα δικό μας σύνολο δεδομένων με βάση το σύνολο δεδομένων World cities [1] [20], το οποίο περιλαμβάνει εικόνες από το Flickr[2] και αποτελείται από 927 εικόνες από το κέντρο της Βαρκελώνης, οι οποίες έχουν επισημανθεί ως προς το περιεχόμενό τους και 2 εκατομμύρια εικόνες από 38 πόλεις ως *σύνολο προς απόσπαση (distractors set)*. Με βάση τη μέθοδο ανάκτησης εικόνων που εισάγουν οι Αβρίθης, Καλαντίδης και Τόλιας στο [1], συλλέξαμε εικόνες οι οποίες κατηγοριοποιήθηκαν σε αξιοθέατα συνδυάζοντας τεχνικές οπτικής και γεωγραφικής συσταδοποίησης, τα αποτελέσματα των οποίων στη συνέχεια επαληθεύτηκαν και με βάση την ιστοσελίδα [3]. Στην πειραματική διαδικασία χρησιμοποιούμε εικόνες αξιοθέατων από τις πόλεις: Αθήνα (6 αξιοθέατα), Λονδίνο (16 αξιοθέατα), Νέα Υόρκη (34 αξιοθέατα) και Παρίσι (12 αξιοθέατα). Ως κατηγορίες θεωρούμε τα αξιοθέατα από τις τέσσερις πόλεις, 68 στο σύνολο. Επιπλέον, επιλέγουμε τυχαία το μέγιστο 200 εικόνες από κάθε κατηγορία. Πιο κάτω παραθέτουμε κάποια χαρακτηριστικά του συνόλου δεδομένων και ενδεικτικά κάποια τυχαία δείγματα εικόνων ανά πόλη και ανά αξιοθέατο. Παράλληλα για να δείξουμε τη δυσκολία του συνόλου δεδομένων, παραθέτουμε και κάποια δείγματα εικόνων από δύο συγκεκριμένα αξιοθέατα, το MoMA στη

[1] http://image.ntua.gr/iva/datasets/world_cities
[2] http://www.flickr.com
[3] http://wikipedia.org

Νέα Υόρκη και το Big Ben στο Λονδίνο.

Athens	London	New York	Paris	Total
4004	14791	22014	12056	52865

Πίνακας 6.1: Κατανομή εικόνων και μέγεθος συνόλου δεδομένων.

Κατά τη διαδικασία εκπαίδευσης των ταξινομητών των αξιοθέατων, το *σύνολο εκπαίδευσης (training set)* κάθε ταξινομητή αποτελείται από 100 *θετικές εικόνες (positives)*, που προέρχονται δηλαδή από τη συγκεκριμένη κατηγορία που θέλουμε να εκπαιδεύσουμε και άλλες 100 *αρνητικές εικόνες (negatives)*,οι οποίες είναι τυχαία επιλεγμένες από τις εικόνες των υπολοίπων κατηγοριών. Αντίστοιχα δομείται και το *σύνολο ελέγχου (testing set)* κάθε ταξινομητή. Όσον αφορά στο σύνολο ελέγχου της ταξινόμησης όλων των κλάσεων, κατασκευάζουμε ένα σύνολο ελέγχου με 10 εικόνες από κάθε αξιοθέατο σε κάθε περίπτωση. Να σημειώσουμε ότι σε κάθε περίπτωση όλα τα σύνολα δειγμάτων επιλέγονται τυχαία και οι δείκτες αξιολόγησης εξάγονται σε μέσους όρους επί διαφορετικών χωρισμάτων σε σύνολα εκπαίδευσης και ελέγχου.

IDs	Landmarks	Random images
57 58 59 60	Arc de Triomphe Centre Georges Pompidou Eiffel Tower Louvre	
61 62 63 64	Louvre Pyramid Montmartre Notre Dame de Paris Palais Garnier	
65 66 67 68	Parc des Princes Place du Tertre Pont Notre Dame Sainte Chapelle	

Σχήμα 6.1: Τυχαία επιλεγμένα μνημεία από το Παρίσι.

IDs	Landmarks	Random images
1 2 3	Acropolis Athens Ancient Agora of Athens Erechtheum	
4 5 6	Odeon of Herodes Atticus Parthenon Temple of Hephaestus	

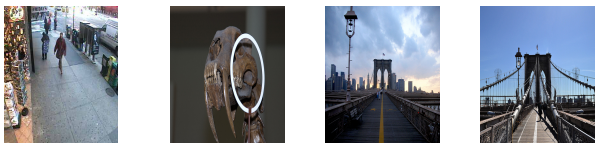
Σχήμα 6.2: Τυχαία επιλεγμένα μνημεία από την Αθήνα.

IDs	Landmarks	Random images
7 8 9 10	Big Ben British Museum Buckingham Palace London Eye	
11 12 13 14	National Gallery Palace of Westminster Piccadilly Piccadilly Circus	
15 16 17 18	Shoreditch Tate Modern Tower Bridge Tower of London	
19 20 21 22	Trafalgar Square Westminster Westminster Abbey Westminster Bridge	

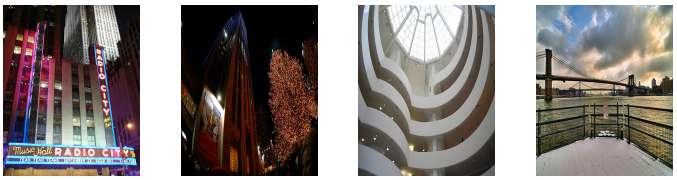
Σχήμα 6.3: Τυχαία επιλεγμένα μνημεία από το Λονδίνο.

Χρησιμοποιούμε SURF χαρακτηριστικά και περιγραφείς τα οποία έχουν εξαχθεί από κάθε εικόνα, θέτοντας *κατώφλι ενίσχυσης (strength threshold)* ίσο με 2.0 στον ανιχνευτή. Για τη διαδικασία δεικτοδότησης, διαθέτουμε ένα γενικό οπτικό λεξικό μεγέθους $100K$, το οποίο έχει κατασκευαστεί από ένα σύνολο $2M$ εικόνων με βάση τον αλγόριθμο συσταδοποίησης

προσεγγιστικό k-means (AKM) [20].

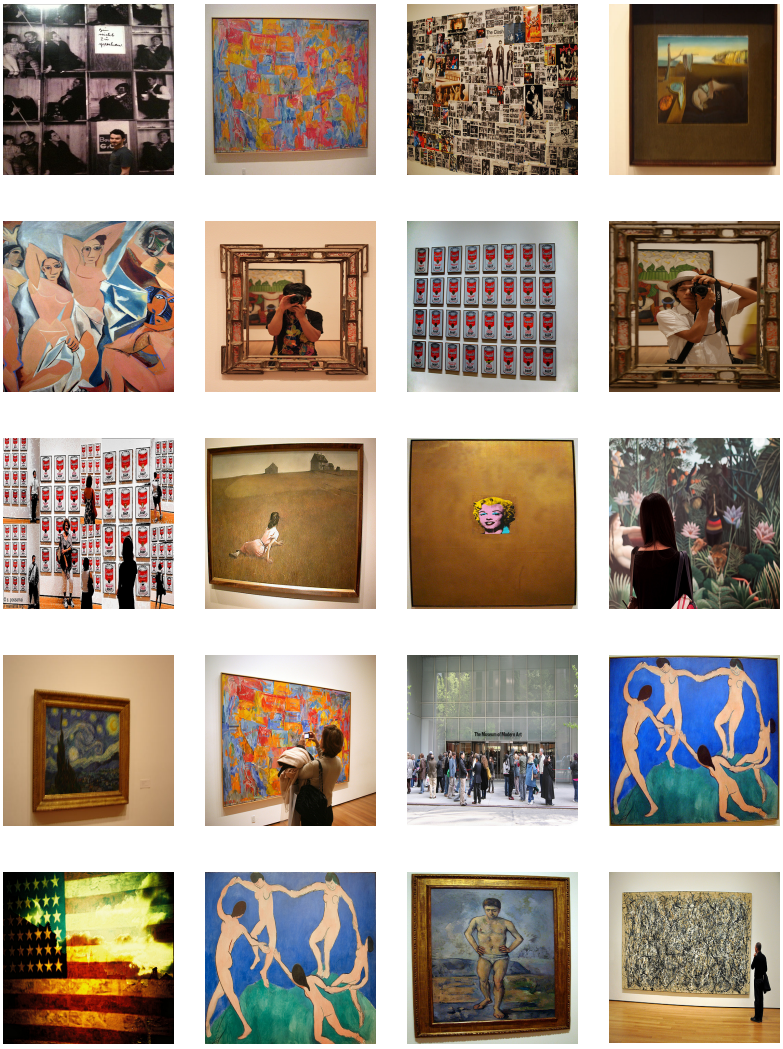
IDs	Landmarks	Random images
23 24 25 26	6 Times Square American Museum of Natural History Brooklyn Brooklyn Bridge	
27 28 29 30	Brooklyn Bridge Park Central Park Central Park Zoo Chrysler Building	
31 32 33 34	Citi Field Coney Island DUMBO Industrial District Empire State Building	
35 36 37 38	Flatiron Building GE Building Grand Central Terminal Helmsley Building	
39 40 41 42	Liberty Island Manhattan Manhattan Bridge Metropolitan Museum of Art	

Σχήμα 6.4: Τυχαία επιλεγμένα μνημεία από τη Νέα Υόρκη.

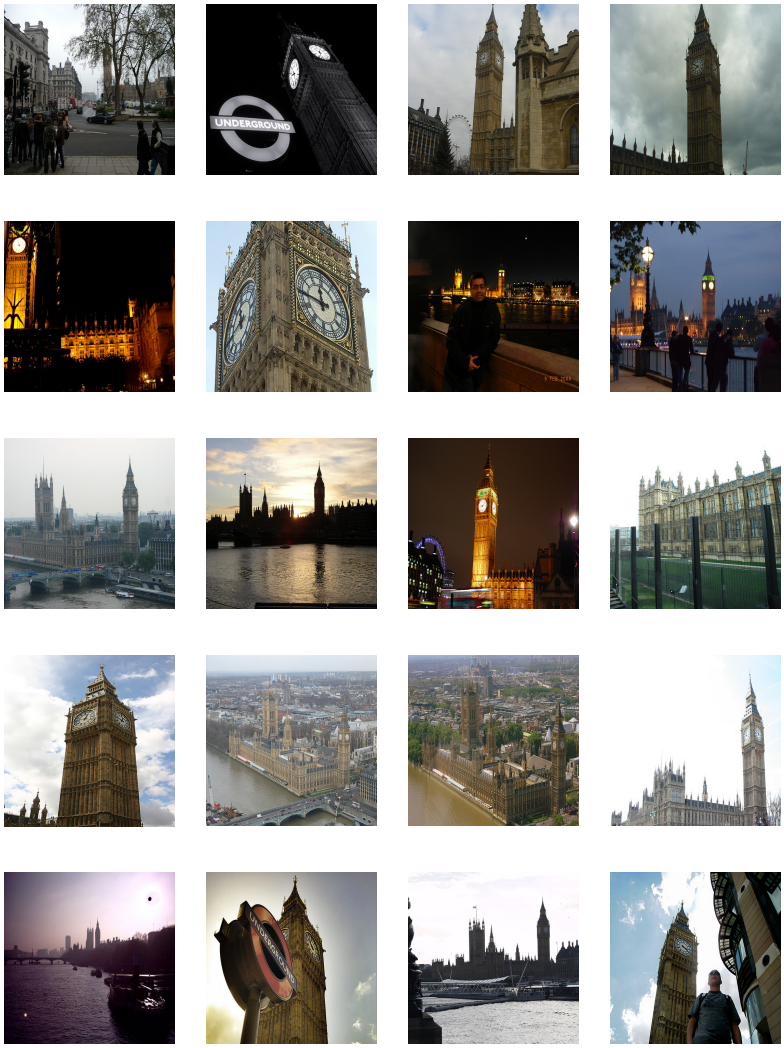
IDs	Landmarks	Random images
43 44 45 46	Museum of Modern Art New York City New Yorker Hotel Radio City Music Hall	
47 48 49 50	Rockefeller Center Rockefeller Center Christmas Tree Solomon R. Guggenheim Museum South Street Seaport	
51 52 53 54	Statue of Liberty Stork Club The Art of This Century gallery Times Square	
55 56	Wall Street Yankee Stadium	

Σχήμα 6.5: Τυχαία επιλεγμένα μνημεία από τη Νέα Υόρκη.

Στους πίνακες 6.2 και 6.3 παραθέτουμε τις κατηγορίες αξιοθέατων με τις αντίστοιχες πόλεις από τις οποίες προέρχονται και τα ονόματά τους, καθώς και την τιμή της παραμέτρου κανονικοποίησης του πυρήνα HPM p, όπως έχει βελτιστοποιηθεί για κάθε κατηγορία. Τα βέλτιστα p έχουν εξαχθεί με βάσει το μέσο όρο της ακρίβειας ταξινόμησης πολλών κλάσεων επί 5 διαφορετικών χωρισμάτων των δεδομένων σε σύνολα εκπαίδευσης και ελέγχου, για όλες τις κατηγορίες.



Σχήμα 6.6: Τυχαία επιλεγμένα μνημεία από το Museum of Modern Art (Μουσείο Μοντέρνας Τέχνης) στη Νέα Υόρκη.



Σχήμα 6.7: Τυχαία επιλεγμένα μνημεία από το Big Ben στο Λονδίνο.

ID	City	Landmark	Images number	p
1	Athens	Acropolis Athens	1662	1.4
2	Athens	Ancient Agora of Athens	246	0.7
3	Athens	Erechtheum	344	0.6
4	Athens	Odeon of Herodes Atticus	233	0.4
5	Athens	Parthenon	1312	0.9
6	Athens	Temple of Hephaestus	207	1.5
7	London	Big Ben	2321	0.6
8	London	British Museum	958	0.9
9	London	Buckingham Palace	974	0.9
10	London	London Eye	1719	1.2
11	London	National Gallery	375	1.2
12	London	Palace of Westminster	470	0.8
13	London	Piccadilly	597	1.3
14	London	Piccadilly Circus	575	0.7
15	London	Shoreditch	217	0.6
16	London	Tate Modern	678	0.6
17	London	Tower Bridge	1349	0.9
18	London	Tower of London	849	1.2
19	London	Trafalgar Square	1059	1.2
20	London	Westminster	1552	1.5
21	London	Westminster Abbey	882	1.4
22	London	Westminster Bridge	216	1.3
23	New York	6 Times Square	1497	1.1
24	New York	American Museum of Natural History	205	1.3
25	New York	Brooklyn	1427	1.4
26	New York	Brooklyn Bridge	1315	1.1
27	New York	Brooklyn Bridge Park	485	1.4
28	New York	Central Park	589	1.4
29	New York	Central Park Zoo	295	1.4
30	New York	Chrysler Building	247	1.5
31	New York	Citi Field	220	0.6
32	New York	Coney Island	247	1.4
33	New York	DUMBO Industrial District	252	1.1
34	New York	Empire State Building	1849	0.6

Πίνακας 6.2: Σύνολο δεδομένων και βέλτιστοι παράμετροι κανονικοποίησης για κάθε μνημείο (1).

6.3 Δείκτες αξιολόγησης

Για την αξιολόγηση των μεθόδων χρησιμοποιούμε δύο δείκτες επίδοσης, το *μέσο όρο της ακρίβειας ταξινόμησης (average classification accuracy)*, όπως αυτός διαμορφώνεται επί ενός αριθμού διαφορετικών χωρισμάτων των εικόνων σε σύνολα εκπαίδευσης και ελέγχου και τον *πίνακα "σύγχυσης" (confusion matrix)*. Ο δείκτης *ακρίβειας ταξινόμησης (classification accuracy)* ορίζεται ως

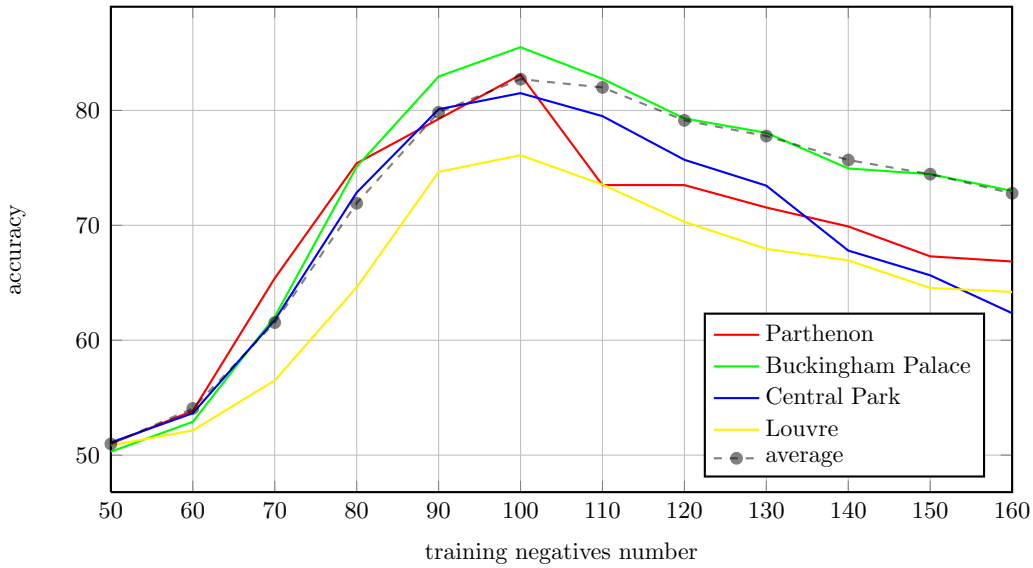
$$\text{ακρίβεια ταξινόμησης} = \frac{\text{σωστά ταξινομημένες εικόνες ελέγχου}}{\text{σύνολο εικόνων ελέγχου}} \tag{6.1}$$

Ακολούθως, ο πίνακας "σύγχυσης" [8] ορίζεται ως

$$M_{ij} = \frac{|\{I_k \in C_j : h(I_k) = i|\}}{|C_j|}, \tag{6.2}$$

όπου $i, j \in \{1, \ldots, N_c\}$, C_j το σύνολο ελέγχου της κατηγορίας j και $h(I_k)$ είναι η κατηγορία, η οποία έλαβε τη μέγιστη τιμή απόφασης ταξινομητή 3.46 για την εικόνα I_k. Συνήθως οι τιμές του confusion matrix εκφράζονται σε ποσοστά επί τις εκατό (%).

6.4 Πειραματικά αποτελέσματα



Σχήμα 6.8: Καμπύλες ακρίβειας συναρτήσει του αριθμού των αρνητικών εικόνων εκπαίδευσης για 4 αξιοθέατα και ο μέσος όρος όλων των αξιοθέατων του dataset.

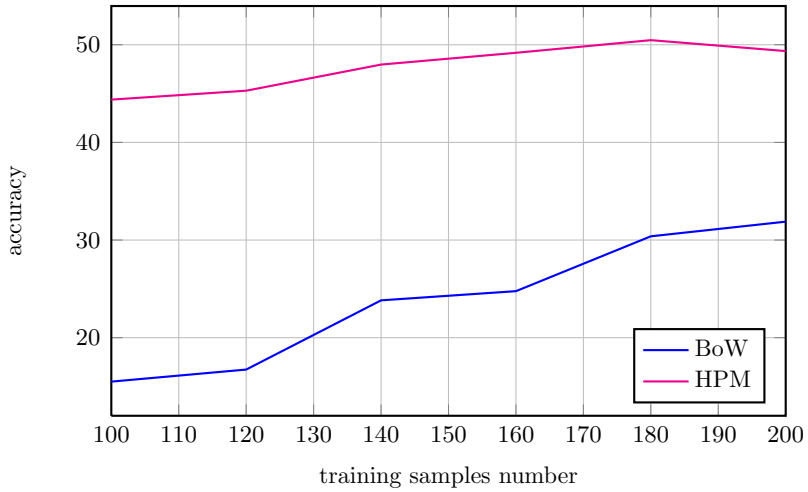
Στα γραφήματα που ακολουθούν παραθέτουμε τα αποτελέσματα των διαφόρων πειραμάτων που διεξήχθησαν. Να σημειώσουμε ότι η αξιολόγηση συντελείται σε πρώτη φάση αναφορικά με την ίδια τη μέθοδο και την αποτελεσματικότητα της, και τον τρόπο με τον οποίο επιδρούν σε αυτή οι διάφοροι παράμετροι (γράφημα 6.8), ενώ σε δεύτερο στάδιο γίνονται συγκρίσεις

ID	City	Landmark	Images number	p
35	New York	Flatiron Building	205	0.5
36	New York	GE Building	475	1.2
37	New York	Grand Central Terminal	716	0.4
38	New York	Helmsley Building	205	0.8
39	New York	Liberty Island	210	0.5
40	New York	Manhattan	1331	1.5
41	New York	Manhattan Bridge	435	1.4
42	New York	Metropolitan Museum of Art	500	0.4
43	New York	Museum of Modern Art	447	1.5
44	New York	New York City	831	1.4
45	New York	New Yorker Hotel	307	1.5
46	New York	Radio City Music Hall	310	0.5
47	New York	Rockefeller Center	2071	1.1
48	New York	Rockefeller Center Christmas Tree	490	0.8
49	New York	Solomon R. Guggenheim Museum	251	1.4
50	New York	South Street Seaport	209	0.8
51	New York	Statue of Liberty	765	1.0
52	New York	Stork Club	291	0.4
53	New York	The Art of This Century gallery	355	1.2
54	New York	Times Square	2122	1.4
55	New York	Wall Street	437	0.4
56	New York	Yankee Stadium	420	1.5
57	Paris	Arc de Triomphe	1188	0.4
58	Paris	Centre Georges Pompidou	267	1.5
59	Paris	Eiffel Tower	2829	0.6
60	Paris	Louvre	1642	1.3
61	Paris	Louvre Pyramid	346	1.1
62	Paris	Montmartre	1037	0.8
63	Paris	Notre Dame de Paris	2712	0.5
64	Paris	Palais Garnier	213	1.2
65	Paris	Parc des Princes	476	0.3
66	Paris	Place du Tertre	221	0.9
67	Paris	Pont Notre Dame	720	0.8
68	Paris	Sainte Chapelle	405	1.2

Πίνακας 6.3: Σύνολο δεδομένων και βέλτιστοι παράμετροι κανονικοποίησης p για κάθε μνημείο (2).

με βάση τη γραμμική προσέγγιση πυρήνα της τεχνικής BoW, η οποία αποτελεί το baseline (γραφήματα 6.9, 6.10, 6.11.

Με βάση λοιπόν το γράφημα 6.8, παρατηρούμε ότι γενικά εκπαιδεύουμε καλύτερα ισοκατανεμημένα σύνολα εκπαίδευσης, όσον αφορά στην αναλογία θετικών και αρνητικών δειγμάτων εικόνων. Δηλαδή, η ακρίβεια ταξινόμησης επηρεάζεται σημαντικά από τη συμμετρία του συνόλου εκπαίδευσης. Πιο συγκεκριμένα, για τα συγκεκριμένα πειράματα παρατηρούμε ότι ο μέσος όρος της ακρίβειας όλων των μεμονωμένων ταξινομητών μεγιστοποιείται για αριθμό αρνητικών εικόνων ίσο με **100**, ακριβώς δηλαδή ίσος με τον αριθμό των θετικών.



Σχήμα 6.9: Καμπύλες ακρίβειας συναρτήσει του αριθμού των εικόνων εκπαίδευσης του πυρήνα BoW και HPM για το σύνολο των 68 αξιοθέατων.

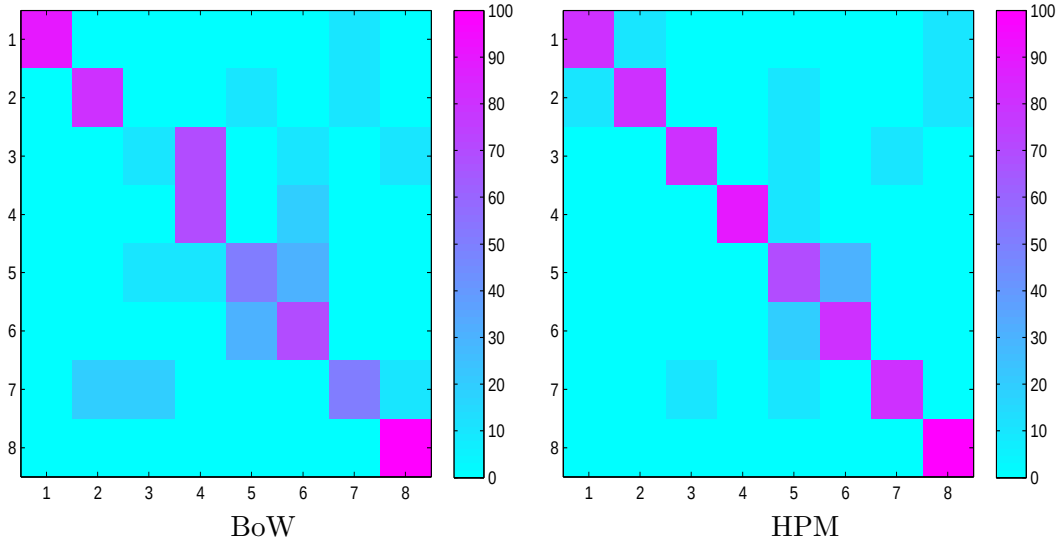
Το γράφημα 6.9 καταδεικνύει την υπεροχή του πυρήνα HPM έναντι του BoW ακόμη και για πολύ λίγα δείγματα εκπαίδευσης, όπου μάλιστα το προβάδισμα είναι **5**%, ενώ στο τελικό πείραμα παραμένει σημαντικό και ίσο με **17.4706**%. Να σημειώσουμε ότι το γράφημα αναφέρεται στο σύνολο των 68 αξιοθέατων. Και σε απόλυτα μεγέθη για συνολικό αριθμό δειγμάτων εκπαίδευσης 200,

Method	**classification accuracy** (%)
BoW	34.3824
HPM	**50.7646**

Πίνακας 6.4: Ακρίβεια ταξινόμησης για συνολικό αριθμό δειγμάτων εκπαίδευσης 200 και από τα 68 αξιοθέατα.

Categories	4	2	15	7	26	41	60	65
4	90	0	0	0	0	0	10	0
2	0	80	0	0	10	0	10	0
15	0	0	10	70	0	10	0	10
7	0	0	0	70	0	20	0	0
26	0	0	10	10	50	30	0	0
41	0	0	0	0	30	70	0	0
60	0	20	20	0	0	0	50	10
65	0	0	0	0	0	0	0	100

Categories	4	2	15	7	26	41	60	65
4	80	10	0	0	0	0	0	10
2	10	80	0	0	10	0	0	10
15	0	0	80	0	10	0	10	0
7	0	0	0	90	10	0	0	0
26	0	0	0	0	70	30	0	0
41	0	0	0	0	20	80	0	0
60	0	0	10	0	10	0	80	0
65	0	0	0	0	0	0	0	100

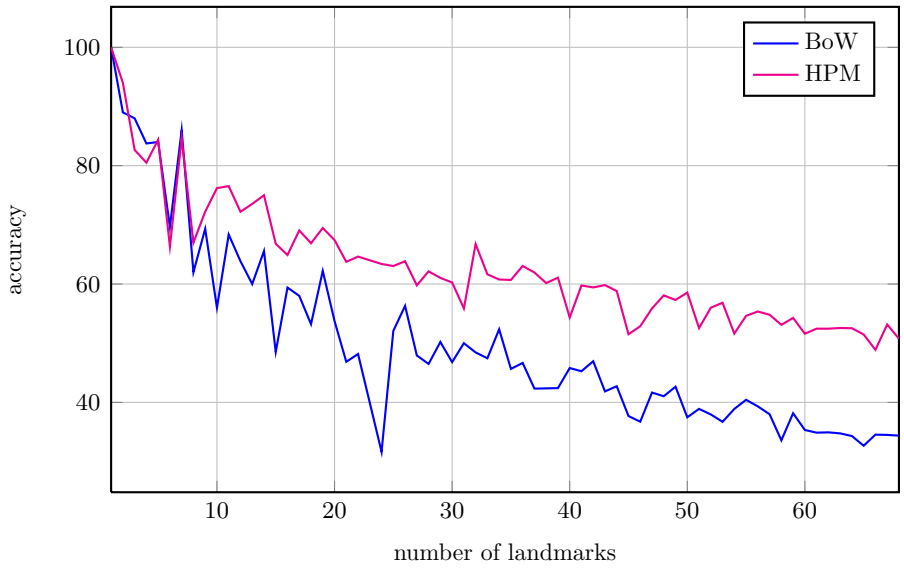


Σχήμα 6.10: Παρατηρούμε ότι η τεχνική HPM παρουσιάζει πάντα σημαντικά μεγάλες τιμές στη διαγώνιο συγκριτικά με το BoW, γεγονός που δείχνει ότι ταιριάζει τις εικόνες της ίδιας κατηγορίας σε μεγαλύτερο ποσοστό, κάνοντας λιγότερα λάθη και κατ'επέκταση τις μαθαίνει καλύτερα.

Ακόμη, με βάση το γράφημα 6.11 διαπιστώνουμε τα εξής:

- Γενικά καταδεικνύεται η δυσκολία ταξινόμησης πολλών κατηγοριών, δηλαδή ότι όσο αυξάνεται ο αριθμός των αξιοθέατων πέφτει η ακρίβεια ταξινόμησης και ο ταξινομητής δυσκολεύεται να δώσει απάντηση.

- Όταν ο αριθμός των αξιοθέατων είναι μικρός, η απόδοση των δύο τεχνικών είναι συγκρίσιμη με την HPM να υπερέχει ελαφρά κατά μέσο όρο (**5**%)και σ' αυτή την περίπτωση. Όσο όμως το πείραμα μεγαλώνει και αυξάνει σε κλίμακα, τόσο γίνεται πιο σαφής η υπεροχή του πυρήνα HPM έναντι του BoW, με την πρώτη τεχνική να παρουσιάζει μεγαλύτερη ακρίβεια ταξινόμησης (**16.3822**%), πολύ καλύτερη διακριτική συμπεριφορά και κατά συνέπεια πιο αποτελεσματική κατηγοριοποίηση των εικόνων.



Σχήμα 6.11: Καμπύλες ακρίβειας συναρτήσει του αριθμού των κατηγοριών των αξιοθέατων για τον πυρήνα BoW και HPM.

Κεφάλαιο 7

Συζήτηση

Ήρθε η ώρα να συνοψίσουμε το αποτέλεσμα της εργασίας, να το αξιολογήσουμε και να διερευνήσουμε πιθανές προοπτικές για περαιτέρω μελέτη και έρευνα.

Τα πειραματικά αποτελέσματα που παρουσιάσαμε και σχολιάσαμε στην προηγούμενη παράγραφο οδηγούν στη θετική αξιολόγηση της αρχικής ιδέα σε πολλά επίπεδα. Καταφέραμε να αναπτύξουμε μια μέθοδο που βελτιώνει τη μέχρι τώρα αποτελεσματικότητα και απόδοση των τεχνικών κατηγοριοποίησης εικόνων (γραμμικό πυρήνα BoW) συνδυάζοντας με επιτυχία σε ένα ενοποιημένο πλαίσιο τρία στοιχεία: 1) τα SVM, 2) την αποδεδειγμένα αποτελεσματική τεχνική χωρικού ταιριάσματος HPM και 3) το ανεστραμμένο αρχείο (δεικτοδότηση) της διαδικασίας ανάκτησης εικόνων. Κοινή συνισταμένη των τριών αυτών στοιχείων είναι ο αρμονικός συνδυασμός εκμάθησης και χωρικού ταιριάσματος, με αποτέλεσμα την εξισορρόπηση της απαίτησης για γενίκευση με τη διακριτική ικανότητα αντίστοιχα, και τη διατήρηση της αναλλοίωτου ως προς μετατόπιση, κλίμακα και περιστροφή μέσω της χρήσης του HPM. Επιπλέον, τα SVM μαζί με την τεχνική δεικτοδότησης συνεισφέρουν σε αραιές και συμπαγείς αναπαραστάσεις υλοποιήσιμες με χαμηλές απαιτήσεις σε υπολογιστικό χώρο και χρόνο, δημιουργώντας το υπόβαθρο για την ανάπτυξη αποδοτικών αλγορίθμων κατηγοριοποίησης εικόνων.

Δυσκολίες παρουσιάστηκαν όσον αφορά σε heuristics κάποιων παραμέτρων, πιο συγκεκριμένα της κανονικοποίησης των δειγμάτων του πυρήνα p, όπου παρατηρήθηκε σημαντική εξάρτηση από την τυχαιότητα των χωρισμάτων κάθε φορά των δεδομένων. Παρόλα αυτά η μέθοδος ανταποκρίνεται με επιτυχία στο πρόβλημα που καλούμαστε να λύσουμε. Ακόμη αξίζει να αναφέρουμε ότι επιχειρήσαμε να συνδυάσουμε σε αντίστοιχη λογική ταξινόμησης πολλών κλάσεων και μεμονωμένους ταξινομητές μονής κλάσης χωρίς αξιόλογα όμως αποτελέσματα, με βάση βέβαια και τα δεδομένα, έτοιμα εργαλεία υλοποίησης (LIBSVM tool). Ένα ακόμη μειονέκτημα που παρατηρήθηκε αποτελεί το γεγονός ότι λόγω της δυσκολίας του συνόλου δεδομένων απαιτείται σημαντικός αριθμός ακόμη SV για την εκπαίδευση των μεμονωμένων ταξινομητών.

Στη μέθοδο που περιγράφουμε αποδώσαμε στην ανάκτηση εικόνων βοηθητικό ρόλο, με στόχο τη βελτίωση της διαδικασίας κατηγοριοποίησης εικόνων. Αντίστοιχα, αναφέρουμε ότι, παρουσιάζεται μια εναλλακτική υλοποίησης της διαδικασίας ανάκτησης εικόνων με βάση τα ήδη εκπαιδευμένα μοντέλα σε μια λογική ταξινόμησης πολλών κλάσεων, απαιτώντας μόνο τον υπολογισμό των τιμών πυρήνα HPM της εικόνας αναζήτησης μόνο με τις εικόνες της

βάσης, οι οποίες αποτελούν τα SV των ταξινομητών που έχουμε εκπαιδεύσει. Κατά αυτό τον τρόπο η εικόνα αναζήτησης θα αποδίδεται σε μια συγκεκριμένη κατηγορία εικόνων, οπότε θα ανακτώνται άμεσα όλες οι εικόνες της δεδομένης κατηγορίας από τη βάση χωρίς πλέον να απαιτούνται ένα προς ένα συγκρίσεις όλων των εικόνων, εισάγοντας κατά αυτό τον τρόπο και κάποια διάσταση σημασιολογικού περιεχομένου στην όλη διαδικασία. Επιπλέον θα πετυχαίναμε εξοικονόμηση υπολογιστικού χώρου και χρόνου σε λειτουργία πραγματικού χρόνου, δεδομένων των αραιών λύσεων που παρέχουν τα SVM, εισάγοντας παρόλα αυτά ένα επιπλέον φόρτο "offline" εκπαίδευσης των επιμέρους ταξινομητών των κατηγοριών των εικόνων.

Για την πληρέστερη αξιολόγηση της μεθόδου θα μπορούσαμε να μεγαλώσουμε την κλίμακα των πειραμάτων, όσον αφορά στον αριθμό των κατηγοριών, καθώς και να διεξάγουμε πειράματα σε περισσότερα και διαφορετικής μορφής σύνολα δεδομένων.

Μια άλλη προσέγγιση που θα μπορούσε να διερευνηθεί είναι να εφαρμοστεί το ταίριασμα εικόνων απλά με τη λογική των nearest neighbours χωρίς τη διαδικασία εκμάθησης. Η απόφαση ταξινόμησης θα λαμβάνεται με βάση το μεγαλύτερο από ένα σταθμισμένο άθροισμα των βαθμολογιών ομοιότητας των υψηλά κατατεταγμένων εικόνων για κάθε αξιοθέατο/κατηγορία.

Με βάση τα αρνητικά που παρατηρήθηκαν, περαιτέρω μελέτη θα μπορούσε να εστιαστεί στην αναζήτηση πιο αποτελεσματικών περιγραφών μηχανών εκμάθησης προς ενσωμάτωση στη διαδικασία κατηγοριοποίησης. Πιθανότατα, η διερεύνηση μεθόδων εκμάθησης κάθε κατηγορίας με βάση μόνο τα θετικά δείγματα σε μια προσέγγιση προσδιορισμού της κατανομής της κάθε κατηγορίας (distribution estimation) και όχι τόσο διαχωρισμού των δεδομένων, να αποτελεί ενδιαφέρουσα εναλλακτική στην κατεύθυνση της καλύτερης και αποδοτικότερης αντιμετώπισης της κατηγοριοποίησης εικόνων.

Παράρτημα Α΄

Πυρήνας πυραμίδας Hough

Αποδεικνύεται ότι στην περίπτωση των πολλών επιπέδων ισχύουν διαδοχικά τα εξής:

$$\begin{aligned} s(C) &= \sum_{c\in C} w(c)s(c) \\ &= \sum_{\ell=0}^{L-1} \alpha 2^{-\ell} \sum_{h\in\mathcal{H}} \eta(h) \sum_{c\in h} w(c) \\ &= \sum_{\ell=0}^{L-1} \alpha 2^{-\ell} \sum_{n=1}^{\infty} (n-1) \sum_{\substack{h\in\mathcal{H} \\ |h|=n}} \sum_{c\in h} w(c) \end{aligned} \tag{Α΄.1}$$

$$\begin{aligned} s(C) &= \sum_{\ell=0}^{L-1} \alpha 2^{-\ell} \sum_{n=1}^{\infty} (n-1) \sum_{\substack{f\in F \\ |f|=n}} \sum_{\substack{h\in\mathcal{H} \\ |h|=n}} \sum_{c\in h} w(c) \\ &= \sum_{\ell=0}^{L-1} \alpha 2^{-\ell} \sum_{f\in F} (|f|-1) \left[\sum_{\substack{h\in\mathcal{H} \\ f(h)=f}} \sum_{c\in h} w(c) \right] \\ &= \sum_{\ell=0}^{L-1} \alpha 2^{-\ell} \sum_{f\in F} (|f|-1) w(f) |\mathcal{H}(f)|, \end{aligned} \tag{Α΄.2}$$

$$\begin{aligned} s(C) &= \sum_{\ell=0}^{L-1} \alpha 2^{-\ell} \sum_{f\in F} (|f|-1) w(f) \langle \phi_f^n(x), \phi_f^n(y) \rangle \\ &= \sum_{\ell=0}^{L-1} \alpha 2^{-\ell} \sum_{f\in F} (|f|-1) w(f) K_{|f|}(x,y) \end{aligned} \tag{Α΄.3}$$

Βιβλιογραφία

[1] Y. Avrithis, Y. Kalantidis, G. Tolias, and E. Spyrou. Retrieving landmark and non-landmark images from community photo collections. In *Proceedings of the international conference on Multimedia*, pages 153–162. ACM, 2010.

[2] R. Baeza-Yates, B. Ribeiro-Neto, et al. *Modern information retrieval*, volume 82. Addison-Wesley New York, 1999.

[3] H. Bay, T. Tuytelaars, and L. Van Gool. Surf: Speeded up robust features. *Computer Vision–ECCV 2006*, pages 404–417, 2006.

[4] C.M. Bishop and SpringerLink (Service en ligne). *Pattern recognition and machine learning*, volume 4. Springer New York, 2006.

[5] O. Chum, J. Matas, and J. Kittler. Locally optimized ransac. *Pattern Recognition*, pages 236–243, 2003.

[6] O. Chum, J. Matas, and S. Obdrzalek. Enhancing ransac by generalized model optimization. In *Proc. of the ACCV*, volume 2, pages 812–817, 2004.

[7] N. Cristianini and J. Shawe-Taylor. *An introduction to support Vector Machines: and other kernel-based learning methods*. Cambridge Univ Pr, 2000.

[8] G. Csurka, C. Dance, L. Fan, J. Willamowski, and C. Bray. Visual categorization with bags of keypoints. In *Workshop on statistical learning in computer vision, ECCV*, volume 1, page 22, 2004.

[9] 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.

[10] D. Greene and P. Cunningham. Practical solutions to the problem of diagonal dominance in kernel document clustering. In *Proceedings of the 23rd international conference on Machine learning*, pages 377–384. ACM, 2006.

[11] R. Hartley, A. Zisserman, and Inc ebrary. *Multiple view geometry in computer vision*, volume 2. Cambridge Univ Press, 2003.

[12] 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*, 51(2):555–592, 2011.

[13] S. Lazebnik, C. Schmid, and J. Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 2, pages 2169–2178. IEEE, 2006.

[14] Y. Li, D.J. Crandall, and D.P. Huttenlocher. Landmark classification in large-scale image collections. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 1957–1964. IEEE, 2009.

[15] D.G. Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004.

[16] K.R. Muller, S. Mika, G. Ratsch, K. Tsuda, and B. Scholkopf. An introduction to kernel-based learning algorithms. *Neural Networks, IEEE Transactions on*, 12(2):181–201, 2001.

[17] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman. Object retrieval with large vocabularies and fast spatial matching. In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, pages 1–8. IEEE, 2007.

[18] J. Shawe-Taylor and N. Cristianini. *Kernel methods for pattern analysis*. Cambridge university press, 2004.

[19] J. Sivic and A. Zisserman. Video google: A text retrieval approach to object matching in videos. In *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pages 1470–1477. IEEE, 2003.

[20] G. Tolias and Y. Avrithis. Speeded-up, relaxed spatial matching. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 1653–1660. IEEE, 2011.

[21] Y. Zhang and T. Chen. Efficient kernels for identifying unbounded-order spatial features. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 1762–1769. IEEE, 2009.