

ΜΕΘΟΔΟΙ ΤΑΞΙΝΟΜΗΣΗΣ ΕΙΚΟΝΑΣ

Η ΜΕΤΑΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ ΕΞΕΙΔΙΚΕΥΣΗΣ

Υποβάλλεται στην

ορισθείσα από την Γενική Συνέλευση Ειδικής Σύνθεσης
του Τμήματος Πληροφορικής
Εξεταστική Επιτροπή

από την

Ειρήνη Μπράχου

ως μέρος των Υποχρεώσεων

για τη λήψη

του

ΜΕΤΑΠΤΥΧΙΑΚΟΥ ΔΙΠΛΩΜΑΤΟΣ ΣΤΗΝ ΠΛΗΡΟΦΟΡΙΚΗ

ΜΕ ΕΞΕΙΔΙΚΕΥΣΗ ΣΤΙΣ ΤΕΧΝΟΛΟΓΙΕΣ-ΕΦΑΡΜΟΓΕΣ

Ιούνιος 2010



ΒΙΒΛΙΟΘΗΚΗ
ΠΑΝΕΠΙΣΤΗΜΙΟΥ ΙΩΑΝΝΙΝΩΝ



026000321443



ΑΦΙΕΡΩΣΗ

Αφιερώνεται στην οικογένειά μου.



ΕΥΧΑΡΙΣΤΙΕΣ

Με την ευκαιρία της ολοκλήρωσης της Μεταπτυχιακής μου εργασίας θα ήθελα να ευχαριστήσω ορισμένα άτομα, τα οποία συνέβαλαν στην περάτωση αυτής της εργασίας.

Καταρχήν θα ήθελα να ευχαριστήσω τον επιβλέπων της εργασίας, κ. Αριστείδη Λύκα, Αναπληρωτή Καθηγητή του Τμήματος Πληροφορικής Πανεπιστημίου Ιωαννίνων (Π.Ι) για την καθοδήγηση και την υπομονή του.

Στη συνέχεια, θα ήθελα να ευχαριστήσω ορισμένα σημαντικά πρόσωπα, τα οποία γνώρισα είτε ως προπτυχιακή είτε ως μεταπτυχιακή φοιτήτρια στο Π.Ι και είχα την τιμή να συνεργαστώ μαζί τους. Αρχικά, ευχαριστώ τα μέλη του Εργαστηρίου Πυρηνικής Φυσικής, του Τμήματος Φυσικής Π.Ι, με τα οποία είχα μια άψογη συνεργασία. Επίσης, ένα μεγάλο ευχαριστώ απευθύνεται και στους «προσωρινούς» μου συναδέλφους στο Τμήμα Μηχανογράφησης & Τεκμηρίωσης Κεντρικής Βιβλιοθήκης Π.Ι- για την δημιουργία ενός εξαιρετικού και ιδανικού εργασιακού περιβάλλοντος-κλίματος.

Ένα ιδιαίτερο ευχαριστώ σε όλους τους συμφοιτητές και τους φίλους που απέκτησα κατά τη διάρκεια της επταετούς παραμονής μου στο Π.Ι και όχι μόνο, που χωρίς τις συμβουλές και την υποστήριξή τους, το σίγουρο είναι ότι αυτή η εργασία δεν θα είχε ολοκληρωθεί.

Θα ήταν μεγάλη παράλειψη να μην ευχαριστήσω την κ. Φωτεινή Ευθυμίου, φιλόλογο-υποψήφια διδάκτωρ-Βιβλιοθηκονόμο της Κεντρικής Βιβλιοθήκης-Κέντρου Πληροφόρησης του Π.Ι, η οποία προσφέρθηκε με ιδιαίτερη προθυμία να κάνει την φιλολογική επιμέλεια του κειμένου της Μεταπτυχιακής εργασίας.



Τέλος και πάνω από όλα ευχαριστώ θερμά τα μέλη της οικογένειας μου, τα οποία είναι πάντα δίπλα μου, με στηρίζουν και μου δίνουν κουράγιο για να προχωράω...



ΠΕΡΙΕΧΟΜΕΝΑ

	Σελ
ΑΦΙΕΡΩΣΗ	iii
ΕΥΧΑΡΙΣΤΙΕΣ	iv
ΠΕΡΙΕΧΟΜΕΝΑ	vi
ΕΥΡΕΤΗΡΙΟ ΠΙΝΑΚΩΝ	ix
ΕΥΡΕΤΗΡΙΟ ΣΧΗΜΑΤΩΝ	x
ΠΕΡΙΛΗΨΗ	xiii
EXTENDED ABSTRACT IN ENGLISH	xv

ΚΕΦΑΛΑΙΟ 1. ΕΙΣΑΓΩΓΗ

1.1 Εισαγωγή στο πρόβλημα της ταξινόμησης εικόνας	1
1.2 Ορισμός του προβλήματος	2
1.3 Παρόμοια προβλήματα	4
1.4 Εφαρμογές του προβλήματος	5
1.5 Προκλήσεις του προβλήματος	7
1.6 Διάρθρωση της εργασίας	11

ΚΕΦΑΛΑΙΟ 2. ΑΝΑΠΑΡΑΣΤΑΣΗ ΕΙΚΟΝΑΣ

2.1 Αναπαράσταση εικόνας	13
2.2 Εξαγωγή χαρακτηριστικών	14
2.2.1 Αλγόριθμος SIFT	14
2.2.2 Αλγόριθμος SPIN	21
2.2.3 Αλγόριθμος RIFT	21
2.2.4 Αλγόριθμος SURF	21
2.3 Περιγραφή σχήματος	22
2.3.1 Αλγόριθμος Canny	22
2.3.2 Πυραμίδα από ιστογράμματα προσανατολισμού ακμών-PHOG	23

ΚΕΦΑΛΑΙΟ 3. ΜΗΧΑΝΕΣ ΔΙΑΝΥΣΜΑΤΩΝ ΥΠΟΣΤΗΡΙΞΗΣ ΜΕ ΧΡΗΣΗ ΑΠΛΟΥ ΠΥΡΗΝΑ



3.1 Γενικά	28
3.2 Εκπαίδευση ταξινομητών SVM	28
3.3 Συναρτήσεις πυρήνα	30
3.4 Πως υπολογίζεται η έξοδος του SVM και πως γίνεται η ταξινόμηση σε πολλές κατηγορίες	32

ΚΕΦΑΛΑΙΟ 4. ΑΛΓΟΡΙΘΜΟΣ SIMPLEMKL

4.1 Γενικά	34
4.2 Χώρος του γραμμικού συνδυασμού πυρήνων	35
4.3 Το πρωτεύον πρόβλημα MKL.....	35
4.4 Το δυικό πρόβλημα MKL	36
4.5 Περιγραφή του αλγόριθμου SimpleMKL.....	38
4.6 Συνθήκες Βελτιστότητας.....	41
4.7 Ο αλγόριθμος SimpleMKL για πολλές κατηγορίες	43
4.8 Σύγκλιση του αλγόριθμου SimpleMKL	

ΚΕΦΑΛΑΙΟ 5. ΜΕΘΟΔΟΙ ΤΑΞΙΝΟΜΗΣΗΣ ΕΙΚΟΝΑΣ

5.1 Ταξινόμηση εικόνας με θύλακες από χαρακτηριστικά	46
5.1.2 Θύλακες χαρακτηριστικών με χρήση βαρών.....	48
5.2 Συνδυασμός χαρακτηριστικών και πυρήνων	49
5.2.1 Συνδυασμός χαρακτηριστικών στο πρόβλημα της ταξινόμησης εικόνας	50
5.2.2 Συνδυασμός χαρακτηριστικών με χρήση του αλγόριθμου SimpleMKL.....	52

ΚΕΦΑΛΑΙΟ 6. ΠΕΙΡΑΜΑΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ

6.1 Γενικά	53
6.2 Το σύνολο δεδομένων CalTech-101.....	54
6.3 Πειραματική εκτίμηση της μεθόδου ταξινόμησης με θύλακες από σημεία-κλειδιά	54
6.4 Πειραματική αξιολόγηση της μεθόδου με χρήση σχήματος	60
6.5 Πειραματική αξιολόγηση του συνδυασμού εμφάνισης και σχήματος	62
6.6 Πειραματική αξιολόγηση του συνδυασμού συνιστωσών με τον αλγόριθμο SimpleMKL	63
6.7 Συμπεράσματα	69
ΑΝΑΦΟΡΕΣ.....	71





ΕΥΡΕΤΗΡΙΟ ΠΙΝΑΚΩΝ

• Πίνακας 1.1 Αντιστοίχιση True/False σε κάθε εικόνα	3
--	---



ΕΥΡΕΤΗΡΙΟ ΣΧΗΜΑΤΩΝ

Σχήμα 1.1 Σύνολο Εικόνων $I=\{\text{Εικόνα1}, \text{Εικόνα2}, \text{Εικόνα3}\}$ για ταξινόμηση	3
Σχήμα 1.2 Σχηματική Παράσταση της Διαδικασίας Ανάκτησης Εικόνας με Βάση το Περιεχόμενο.....	5
Σχήμα 1.3 Φωτογράφιση του Ίδιου Αντικειμένου με Διαφορετική Οπτική Γωνία. 8	
Σχήμα 1.4 Αντικείμενα της Ίδιας Κατηγορίας με Δομικές Διαφορές.....	8
Σχήμα 1.5 Πρώτη γραμμή: Εικόνες με αρπύχορδο. Δεύτερη γραμμή: Εικόνες με πιάνο.	9
Σχήμα 1.6 Αλλαγές στις Συνθήκες Φωτισμού.....	10
Σχήμα 1.7 Αλλαγές στην Κλίμακα.....	10
Σχήμα 1.8. Εικόνες με Θόρυβο.	11
Σχήμα 2.1 (α) Αρχική Εικόνα. (β) Εικόνα με τα 219 Patches που Έχουν Εξαχθεί από τον Αλγόριθμο SIFT.	14
Σχήμα 2.2 (αριστερά) Η Αρχική Εικόνα Αλλάζει Κλίμακα και Γίνεται Συνέλιξη των Εικόνων που Παράγονται με το Φίλτρο Gauss. (δεξιά) «Γειτονικές» Εικόνες Αφαιρούνται και Παράγεται η Συνάρτηση Διαφοράς του Gauss.	16
Σχήμα 2.3 Το Pixel που Σημειώνεται με X Συγκρίνεται με τα 8 Γειτονικά του στην Ίδια Κλίμακα και με τα 18 Pixels στις «Γειτονικές» του Εικόνες.	17
Σχήμα 2.4 Περιγραφή του Patch.....	20
Σχήμα 2.5 Αναπαράσταση Εικόνας με Ιστογράμμο Προσανατολισμού Ακμών. Πρώτη Γραμμή: Διαίρεση της Αρχικής Εικόνας σε Επίπεδα και Κελιά. Παρακάτω: Υπολογισμός Ιστογράμματος για Κάθε Εικόνα και Κάθε Επίπεδο.....	24
Σχήμα 2.6 Σύγκριση Απόστασης Chamfer (Πράσινες Ράβδοι) με την χ^2 Απόσταση των Διανυσμάτων PHOG (γκρι ράβδοι). Στον Κάθετο Άξονα η Τιμή 1 Αντιστοιχεί σε Τέλεια Ομοιότητα, ενώ η Τιμή 0 και Τέλεια Ανομοιότητα.....	26



Σχήμα 2.7 Σύγκριση Απόστασης Chamfer (Πράσινες Ράβδοι) με την χ^2 Απόσταση των Διανυσμάτων PHOG (Γκρι Ράβδοι). Ομοιότητα μιας Εικόνας με Άλλες Εικόνες που Ανήκουν στο Σύνολο Caltech.	27
Σχήμα 5.1 Ταξινόμηση Εικόνας σε Κατηγορία με Θύλακες από Σημεία-Κλειδιά ...	
Σχήμα 6.1 Εικόνες από το Σύνολο Caltech-101.....	47
Σχήμα 6.2 Γραφική Παράσταση για Σύγκριση της Μεθόδου για Δύο, Τρεις και Δέκα Κατηγορίες σε Συνάρτηση με την Τιμή της Παραμέτρου k	56
Σχήμα 6.3 Γραφική Παράσταση για Σύγκριση της Μεθόδου με 10 και 15 Παραδείγματα/Κατηγορία ως Σύνολο Εκπαίδευσης.	57
Σχήμα 6.4 Γραφική Παράσταση για Σύγκριση του Πυρήνα Rbf και του Γραμμικού Πυρήνα.	58
Σχήμα 6.5 Γραφική Παράσταση για Σύγκριση του Πυρήνα Rbf και του χ^2 Πυρήνα.....	59
Σχήμα 6.6 Γραφική Παράσταση- Θύλακες από Σημεία-Κλειδιά με Βάρη.....	59
Σχήμα 6.7 Γραφική Παράσταση-Σύγκριση του PHOG για $l=0,1,2,3,4$ και για Γραμμικό Πυρήνα και Πυρήνα rbf.....	61
Σχήμα 6.8 Γραφική Παράσταση- Χρήση Βαρών στο PHOG.....	62
Σχήμα 6.9 Γραφική Παράσταση- Σύγκριση Χρήσης Σχήματος με τον Συνδυασμό Εμφάνισης και Σχήματος.....	63
Σχήμα 6.10 Γραφική Παράσταση- Σύγκριση Απλού και Σταθμισμένου Αθροίσματος Πυρήνων.....	65
Σχήμα 6.11 Γραφική Παράσταση-Σύγκριση του Συνδυασμού Πυρήνων με SimpleMKL σε Σχέση με τη Χρήση Απλού Πυρήνα για Τέσσερα Δυναδικά Προβλήματα.	65
Σχήμα 6.12 Γραφική Παράσταση-Σύγκριση της Χρήσης Πολυωνυμικού και rbf Πυρήνα στην Περίπτωση του Συνδυασμού Πυρήνων με SimpleMKL σε Σχέση με τη Χρήση Απλού Πυρήνα.....	66
Σχήμα 6.13 Γραφική Παράσταση-Σύγκριση της Επίδοσης του Συνδυασμού Χαρακτηριστικών με τη Χρήση του SVM με Απλό Πυρήνα. (Πολυωνυμικός Πυρήνας).....	67
Σχήμα 6.14 Σύγκριση της Επίδοσης του Συνδυασμού Χαρακτηριστικών με τη Χρήση του SVM με Απλό Πυρήνα. (Πυρήνας rbf).....	68
Σχήμα 6.15 Γραφική Παράσταση-Επίδοση της Μεθόδου Συνδυασμού Συνιστωσών με Τιμές που Έχουν Προσδιοριστεί Από τη Δική μας Αναζήτηση Πλέγματος.....	69



ΠΕΡΙΛΗΨΗ

Ειρήνη Μπράχου του Βασιλείου και της Μαργωλίτσας

MSc, Τμήμα Πληροφορικής, Πανεπιστήμιο Ιωαννίνων, Ιούνιος, 2010

Μέθοδοι Ταξινόμησης Εικόνας

Επιβλέπων : Αριστείδης Λύκας

Το πρόβλημα της ταξινόμησης εικόνων σε κατηγορία (image classification) αποτελεί σημαντικό ερευνητικό πρόβλημα στο πεδίο της τεχνητής όρασης (computer vision). Πρόκειται για ιδιαίτερα δύσκολο πρόβλημα, εξαιτίας των ιδιαιτεροτήτων που συναντά κανείς σε ρεαλιστικές εικόνες, όπως για παράδειγμα οι μεταβολές στο φωτισμό, στο μέγεθος των αντικειμένων, στην οπτική γωνία και στην απόσταση λήψης, στην ύπαρξη εμποδίων που υπερκαλύπτουν τμήμα του ουσιώδους περιεχομένου της εικόνας κλπ.

Οι μέθοδοι που έχουν προταθεί για το πρόβλημα ταξινόμησης εικόνας στην πλειοψηφία τους περιλαμβάνουν δύο στάδια. Στο πρώτο στάδιο γίνεται ανάλυση του περιεχομένου των εικόνων και εξάγονται συνοπτικές περιγραφές για διάφορα χαρακτηριστικά π.χ. χρώμα, υφή, σχήμα. Με βάση αυτές τις περιγραφές δημιουργούνται διανυσματικές αναπαραστάσεις των εικόνων οι οποίες στο δεύτερο στάδιο χρησιμοποιούνται από αλγορίθμους μηχανικής μάθησης για την κατασκευή συστημάτων ταξινόμησης με χρήση παραδειγμάτων εκπαίδευσης.

Στην παρούσα εργασία μελετώνται διάφοροι συνδυασμοί μεθόδων για τα δύο στάδια του προβλήματος. Για το στάδιο της ανάλυσης εικόνας μελετώνται i) οι περιγραφείς εμφάνισης SIFT (Scale Invariant Feature Transform) οι οποίοι στη συνέχεια ομαδοποιούνται για τη δημιουργία πιο συνοπτικών περιγραφών που ονομάζονται



οπτικές λέξεις (visual words) και ii) οι περιγραφείς σχήματος PHOG (Pyramid Histogram of Orientation Gradient). Για το στάδιο της ταξινόμησης χρησιμοποιήθηκε καταρχήν ο ταξινομητής SVM. Στη συνέχεια μελετήθηκε η μέθοδος SimpleMKL για εκπαίδευση ταξινομητών SVM με πολλαπλούς πυρήνες (kernels) με αυτόματη επιλογή της βαρύτητας που υπολογίζεται για κάθε πυρήνα. Η μέθοδος αυτή αφενός παρέχει τη δυνατότητα σταθμισμένου συνδυασμού περιγραφέων και αφετέρου επιτρέπει την επιλογή των αποδοτικών συνιστωσών (feature selection) από τους επιμέρους περιγραφείς. Οι μέθοδοι που μελετήθηκαν αξιολογούνται πειραματικά με χρήση του γνωστού συνόλου δεδομένων CalTech-101.



EXTENDED ABSTRACT IN ENGLISH

Brachou, Eirini, V.

MSc. Computer Science Department. University of Ioannina, Greece. June, 2010.

Image Classification Methods.

Supervisor: Aristidis Likas..

Image classification has been an important research problem in the field of computer vision. It is considered a particularly difficult problem due to challenges in real world images, such as illumination, scaling, camera-view point variations, occlusion e.t.c.

The majority of image classification methods, that have been proposed, include two stages. In the first one, the contents of the images are analyzed and compact descriptors of various features such as color, shape, texture are extracted. Based on these descriptors, vector representations of the images are then produced, which machine learning algorithms make use of in the second stage to construct classification systems using training examples.

Various combinations of methods for both stages of the problem are studied in the present thesis. For the image analysis stage both i) SIFT (Scale Invariant Feature Transform) appearance descriptors, which are then quantized to form more dense descriptors called visual words and ii) PHOG (Pyramid Histogram of Orientation Gradients) shape descriptors are studied. For the classification stage the SVM classifier was used in the first place. Then, the SimpleMKL method, which is used for training multikernel SVM classifiers and automatically selecting the weight for each



kernel, was studied. This method both provides a weighted combination of descriptors and enables efficient feature rather than descriptor selection.

The methods that were studied are experimentally evaluated using the well known CalTech-101 dataset.



ΚΕΦΑΛΑΙΟ 1. ΕΙΣΑΓΩΓΗ

1.1 Εισαγωγή στο πρόβλημα της ταξινόμησης εικόνας

1.2 Ορισμός του προβλήματος

1.3 Παρόμοια προβλήματα

1.4 Εφαρμογές του προβλήματος

1.5 Προκλήσεις του προβλήματος

1.6 Διάρθρωση της εργασίας

1.1. Εισαγωγή στο πρόβλημα της ταξινόμησης εικόνας

Τα τελευταία χρόνια, η χρήση των ψηφιακών φωτογραφικών μηχανών έχει αυξηθεί κατακόρυφα. Αυτό είναι απολύτως φυσικό, αν αναλογιστεί κανείς ότι είναι ενσωματωμένες στα περισσότερα κινητά τηλέφωνα επιτρέποντας στο χρήστη του κινητού τηλεφώνου να απαθανατίσει οποιαδήποτε σκηνή επιλέξει, με το πάτημα ενός πλήκτρου. Επίσης, το συνεχώς μειούμενο κόστος τους, σε συνδυασμό με τις ευκολίες που παρέχουν (π.χ δυνατότητα αποθήκευσης εκατοντάδων φωτογραφιών στο σκληρό δίσκο ενός οικιακού υπολογιστή, δυνατότητα διαγραφής ανεπιθύμητων φωτογραφιών, δυνατότητα επιλεκτικής εκτύπωσης φωτογραφιών) καθιστούν την χρήση τους ιδιαίτερα ελκυστική, ακόμη και από άτομα που δεν έχουν εξειδικευμένη τεχνολογική κατάρτιση.

Η μαζική χρήση των ψηφιακών φωτογραφικών μηχανών δημιουργεί ένα σημαντικό πρόβλημα: την παραγωγή τεράστιων συλλογών από ψηφιακές φωτογραφίες. Ενδεικτικά αναφέρεται ότι το Associated Press συλλέγει και αρχειοθετεί καθημερινά περίπου 1000 φωτογραφίες! Ελλείψει κάποιας οργάνωσης αυτού του τεράστιου όγκου από ψηφιακές πληροφορίες καθίσταται δύσκολη οποιαδήποτε λειτουργία,



όπως αναζήτηση μιας συγκεκριμένης φωτογραφίας, διαγραφή κάποιας άλλης κ.ό.κ. Είναι επομένως κατανοητή η ανάγκη οργάνωσης των φωτογραφικών συλλογών με κάποιο τρόπο, ώστε η διαχείρισή τους να πραγματοποιείται με άνεση, εξοικονομώντας πολύτιμο χρόνο και αποφεύγοντας την ανάγκη, για παράδειγμα, σειριακής αναζήτησης, ώστε να βρεθεί η επιθυμητή φωτογραφία.

Μια προσέγγιση που προτάθηκε για την επίλυση του παραπάνω προβλήματος ήταν η περιγραφή κάθε εικόνας με κείμενο (textual description). Με αυτόν τον τρόπο κάποιος μπορούσε να χρησιμοποιήσει μια απλή περιγραφή για να εντοπίσει την φωτογραφία που τον ενδιέφερε από την φωτογραφική συλλογή (text-based image retrieval). Ωστόσο, η παραπάνω τακτική ήταν αναποτελεσματική, γιατί λόγω του μεγάλου, συνήθως, μεγέθους μιας φωτογραφικής συλλογής, η λεκτική περιγραφή είναι αφενός μεν χρονοβόρα και επίπονη διαδικασία, αφετέρου δε υποκειμενική και πολύ συχνά ασαφής, μη πλήρης.

Μια άλλη προσέγγιση που προτάθηκε ήταν η κατηγοριοποίηση των φωτογραφιών με κάποιο κριτήριο. Η παρούσα διατριβή εστιάζει στο πρόβλημα της ταξινόμησης εικόνας με βάση το περιεχόμενο (visual categorization) δηλαδή στην ανάθεσή της σε μια κατηγορία από ένα σύνολο προκαθορισμένων κατηγοριών.

Οποιοσδήποτε άνθρωπος, που έχει αποκτήσει οπτικές εμπειρίες με τη βοήθεια του εγκεφάλου και του συστήματος οράσεως, μπορεί να ταξινομήσει εικόνες εύκολα και γρήγορα. Ωστόσο, στην παρούσα εργασία στόχος είναι η διαδικασία της ταξινόμησης να γίνει αυτόματα με τη βοήθεια μεθόδων, ώστε να απαλλάσσεται ο άνθρωπος από την επίπονη και πολλές φορές επιρρεπή σε λάθη διαδικασία της χειρωνακτικής ταξινόμησης. Επίσης, στόχος είναι η ταξινόμηση να γίνεται με όσο το δυνατόν μεγαλύτερη ακρίβεια και λιγότερα λάθη.

1.2. Ορισμός του προβλήματος

Σε αυτή την παράγραφο δίνεται ένας τυπικός ορισμός της ταξινόμησης εικόνας σε κατηγορία με βάση το περιεχόμενο. Έστω $I = \{I_1, I_2, I_3, \dots, I_N\}$ ένα σύνολο από N εικόνες



και έστω $C=\{C_1,C_2,C_3,...C_M\}$ ένα σύνολο από M προκαθορισμένες κατηγορίες. Στη διαδικασία της ταξινόμησης σε κάθε ζεύγος $\langle I,C_i \rangle$ $I \times C$, ανατίθεται μια λογική τιμή (True/False). Η ανάθεση της τιμής True (Αληθής) σημαίνει ότι η εικόνα I_i ανήκει στην κατηγορία C_i , ενώ η ανάθεση της τιμής False (Ψευδής) σημαίνει ότι η εικόνα I_i δεν ανήκει στην κατηγορία C_i . Ουσιαστικά επιδιώκεται η δημιουργία μιας συνάρτησης απεικόνισης $\Phi: D \times C \rightarrow \{T,F\}$ η οποία ονομάζεται ταξινομητής (classifier), για την προσέγγιση μιας άγνωστης συνάρτησης στόχου (target function) $\Phi': D \times C \rightarrow \{T,F\}$, η οποία περιγράφει πως θα έπρεπε να ταξινομηθούν οι εικόνες. Φυσικά είναι επιθυμητό η συνάρτηση στόχος και ο ταξινομητής να συμπίπτουν όσο το δυνατόν περισσότερο.

Για παράδειγμα, αν το σύνολο εικόνων $I=\{\text{Εικόνα1}, \text{Εικόνα2}, \text{Εικόνα3}\}$, που φαίνονται στο σχήμα 1.1 πρέπει να ταξινομηθεί στο σύνολο από προκαθορισμένες κατηγορίες. $C=\{\text{αεροπλάνο}, \text{μοτοσικλέτα}\}$, με χειρωνακτική ταξινόμηση, προκύπτει ο πίνακας 1.1.



Σχήμα 1.1 Σύνολο Εικόνων $I=\{\text{Εικόνα1}, \text{Εικόνα2}, \text{Εικόνα3}\}$ για ταξινόμηση

Πίνακας 1.1 Αντιστοίχιση True/False σε κάθε εικόνα.

	Αεροπλάνο	Μοτοσικλέτα
Εικόνα1	True	False
Εικόνα2	True	False
Εικόνα3	False	True

Το παραπάνω πρόβλημα αποτελεί ένα δυαδικό πρόβλημα ταξινόμησης (binary classification problem), δεδομένου ότι οι εικόνες πρέπει να ταξινομηθούν σε δύο προκαθορισμένες κατηγορίες. Στην περίπτωση που οι προκαθορισμένες κατηγορίες είναι περισσότερες από δύο έχουμε πρόβλημα με πολλές κατηγορίες (multiclass classification problem).

1.3. Παρόμοια προβλήματα

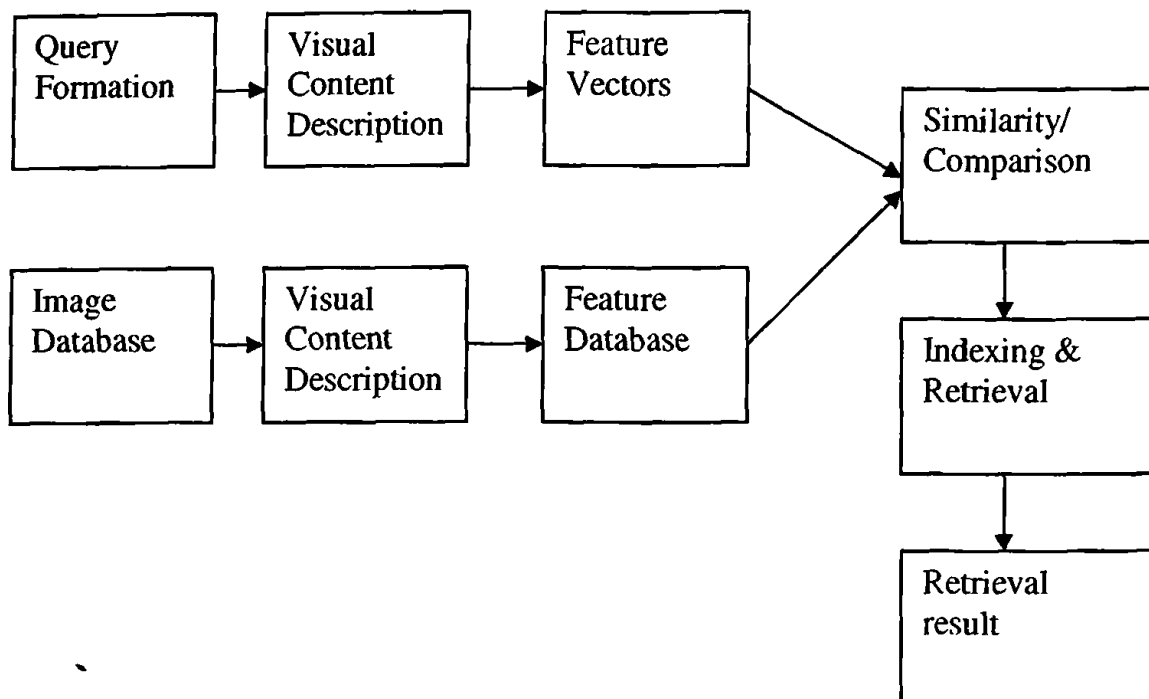
Υπάρχουν προβλήματα, όπως τα τρία παρακάτω, τα οποία παρουσιάζουν αναλογίες με το πρόβλημα της ταξινόμησης εικόνας με βάση το περιεχόμενο. Στη συνέχεια δίνεται μια σύντομη περιγραφή αυτών των προβλημάτων.

1. Ανάκτηση εικόνας με βάση το περιεχόμενο (Content based image Retrieval).
2. Εντοπισμός Αντικειμένου σε εικόνα (Object Detection).
3. Αναγνώριση Αντικειμένου σε εικόνα (Object Recognition).

1. Ανάκτηση εικόνας με βάση το περιεχόμενο. (Content based image Retrieval). Έχει να κάνει με την ανάκτηση μιας εικόνας από μια βάση δεδομένων με εικόνες. Σε αυτή τη βάση δεδομένων κάθε εικόνα έχει καταχωρηθεί σαν ένα διάνυσμα μεγάλης διάστασης (feature vector), με βάση τα οπτικά της περιεχόμενα (visual contents), δηλαδή το χρώμα, την υφή, το σχήμα των αντικειμένων που απεικονίζονται σε αυτή. Ουσιαστικά, η βάση δεδομένων δεν περιέχει τις εικόνες αποθηκευμένες στη φυσική τους μορφή, αλλά περιέχει κωδικοποιημένες εικόνες, με τη βοήθεια των διανυσμάτων που αντιστοιχούν σε αυτές (feature database).

Για την ανάκτηση εικόνας, ο χρήστης παρέχει στο σύστημα ανάκτησης μια εικόνα-παράδειγμα, αντιπροσωπευτική της κατηγορίας εικόνων που τον ενδιαφέρει. Για παράδειγμα, αν κάποιος επιθυμεί να ανακτήσει όλες τις εικόνες που απεικονίζουν κούπα, τότε πρέπει να τροφοδοτήσει τη βάση δεδομένων με μια εικόνα-παράδειγμα, που στη συγκεκριμένη περίπτωση απεικονίζει κούπα. Στη συνέχεια, το σύστημα ανάκτησης μετατρέπει την εικόνα-παράδειγμα σε διάνυσμα και υπολογίζει την απόσταση αυτού του διανύσματος από τα διανύσματα των εικόνων που έχουν καταχωρηθεί στη βάση δεδομένων, προκειμένου να υπολογίσει τις ομοιότητες/ διαφορές. Από αυτόν τον υπολογισμό προκύπτουν τα αποτελέσματα της ανάκτησης, μεταξύ των οποίων συγκαταλέγονται εικόνες που εμφανίζουν μεγάλη ομοιότητα με την εικόνα-παράδειγμα.





Σχήμα 1.2 Σχηματική Παράσταση της Διαδικασίας Ανάκτησης Εικόνας με Βάση το Περιεχόμενο.

2. Εντοπισμός αντικειμένου (Object Detection) Αυτό το πρόβλημα σχετίζεται με τον εντοπισμό σε μια εικόνα ενός αντικειμένου μιας δεδομένης κατηγορίας. Ειδικότερα τα τελευταία χρόνια η ερευνητική προσπάθεια εστιάζεται σε εντοπισμό ανθρώπινου προσώπου (face detection), αυτοκινήτων, πεζών [1]-[2] κ.λ.π.

3. Αναγνώριση αντικειμένου (Object Recognition). Η αναγνώριση αντικειμένου στην τεχνητή όραση είναι η διαδικασία εντοπισμού ενός δεδομένου αντικειμένου σε μια εικόνα ή βίντεο.

1.4. Εφαρμογές του προβλήματος

Η ταξινόμηση εικόνας σε κατηγορία με βάση το περιεχόμενο μπορεί να φανεί χρήσιμη σε διάφορες εφαρμογές, όπως αυτές που ακολουθούν.

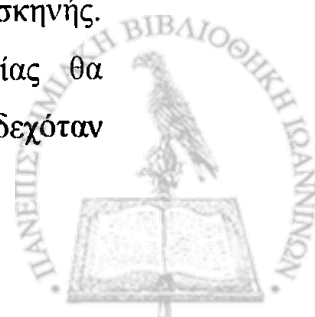
1. Αναζήτηση εικόνας (Image search). Η αναζήτηση εικόνας είναι η πιο άμεση εφαρμογή, όταν οι άνθρωποι μιλάνε για ταξινόμηση εικόνας σε κατηγορία με βάση

το περιεχόμενο. Με αυτήν την έννοια, μπορεί κανείς να σκεφτεί την αναζήτηση εικόνας στη μεγαλύτερη βάση δεδομένων του κόσμου, με την κατασκευή μηχανών αναζήτησης εικόνων στο διαδίκτυο ή απλώς τη δημιουργία εφαρμογών για αναζήτηση εικόνων σε ένα προσωπικό υπολογιστή. Στις μέρες μας, η κακή επίδοση που παρουσιάζουν οι μηχανές αναζήτησης στο διαδίκτυο οφείλεται στο ότι πραγματοποιούν αναζήτηση όχι με βάση το περιεχόμενο της εικόνας, αλλά χρησιμοποιούν πληροφορίες, όπως το όνομα του αρχείου ή τον κώδικα HTML που πλαισιώνει την εικόνα. Ωστόσο, ο φυσικός τρόπος εντοπισμού μιας εικόνας είναι η οπτική αναζήτηση, και αυτός είναι ο στόχος των μεθόδων τεχνητής όρασης.

2. Αναζήτηση βίντεο (Video Search). Τα τελευταία χρόνια έχουν παραχθεί πολλές διαφημίσεις και δεδομένα βίντεο. Τα αρχεία βίντεο συνηθίζεται να αποθηκεύονται σε βάσεις δεδομένων με πληροφορίες μεταδεδομένων (metadata). Θα ήταν χρήσιμο να μπορούσε κανείς να ανακτήσει ένα αρχείο βίντεο με βάση το περιεχόμενό του. Επίσης, οι παραγωγοί ή οι σκηνοθέτες ταινιών θα ενδιαφερόταν να ανακτήσουν πλάνα ταινιών στα οποία πρωταγωνιστεί ένας συγκεκριμένος ηθοποιός ή που εξελίσσονται σε ένα συγκεκριμένο τόπο.

3. Ιατρικές εφαρμογές (Medical Applications). Στον τομέα της ιατρικής καθημερινά παράγονται πολλές εικόνες (π.χ ακτινογραφίες, μαστογραφίες κ.λ.π). Θα ήταν χρήσιμο για τους γιατρούς να έχουν στη διάθεσή τους εργαλεία για να ταξινομούν τις ιατρικές εικόνες σε κατηγορία (π.χ εικόνες αξονικής τομογραφίας με καρκίνο, εικόνες αξονικής τομογραφίας χωρίς καρκίνο) και να μην τις εξετάζουν ανά περίπτωση, όπως κάνουν.

4. Συμπίεση Βίντεο (Video Compression). Λόγω του περιορισμένου εύρου ζώνης σημαντικών καναλιών επικοινωνίας (π.χ ασύρματων, υποθαλάσσιων κ.λ.π) η μετάδοση βίντεο σε αυτά τα κανάλια απαιτεί ουσιαστική συμπίεση των δεδομένων που μεταδίδονται. Μια πολλά υποσχόμενη προσέγγιση θα ήταν η συμπίεση των δεδομένων με βάση το περιεχόμενο της σκηνής, με την εφαρμογή διαφορετικών επιπέδων συμπίεσης ανάλογα με το πόσο σημαντικά είναι τα αντικείμενα της σκηνής. Για παράδειγμα, οι ηθοποιοί που είναι αντικείμενα μεγάλης σημασίας θα διατηρούσαν υψηλή οπτική ποιότητα, ενώ αντικείμενα του φόντου θα δεχόταν



υψηλότερη κωδικοποίηση με σκοπό να καταλαμβάνουν λιγότερα bytes. Σε αυτή την εφαρμογή, η τεχνητή όραση θα συνεισέφερε στο να γίνει αρχικά κατάτμηση του πλάνου στα αντικείμενα που το απαρτίζουν και στη συνέχεια ταξινόμησή τους σε κατηγορίες.

5. Παρακολούθηση (Surveillance). Τα συστήματα παρακολούθησης δεν έχουν εξελιχθεί ακόμη αρκετά και απαιτείται η παρουσία του ανθρώπινου παράγοντα για να εντοπίσει ύποπτους ανθρώπους και ασυνήθιστα περιστατικά. Τα προχωρημένα συστήματα επιδιώκουν να ανιχνεύσουν αυτόματα τέτοιου είδους γεγονότα, όπως π.χ ένα διαπληκτισμό σε ένα ποδοσφαιρικό αγώνα που διεξάγεται σε στάδιο.

6. Εναέριες εικόνες (Aerial images). Κάθε χρόνο δαπανώνται χιλιάδες ευρώ από τις εθνικές υπηρεσίες χαρτογραφίας, προκειμένου να ενημερωθούν οι χάρτες. Η διαδικασία της ανανέωσης είναι συχνά χρονοβόρα και απαιτεί κάποιο άτομο για τη σύγκριση του τρέχοντος χάρτη με την πιο πρόσφατη δορυφορική εικόνα υψηλής ανάλυσης. Τεχνικές για ταξινόμηση εικόνας θα βοηθούσαν στον εντοπισμό των αλλαγών στο τοπίο, χωρίς την ανθρώπινη παρέμβαση.

7. Ρομποτική (Robotics). Η παροχή όρασης σε ένα ρομπότ είναι ίσως ένας από τους πιο φιλόδοξους στόχους στο πεδίο της τεχνητής όρασης. Με αυτόν τον τρόπο ένα εντελώς αυτόνομο ρομπότ θα μπορούσε να χρησιμοποιηθεί για να εντοπίσει συγκεκριμένα αντικείμενα ενδιαφέροντος και να αντικαταστήσει τον ανθρώπινο παράγοντα σε επικίνδυνες καταστάσεις (π.χ πυρκαγιά, υποθαλάσσια εξερεύνηση κ.λ.π).

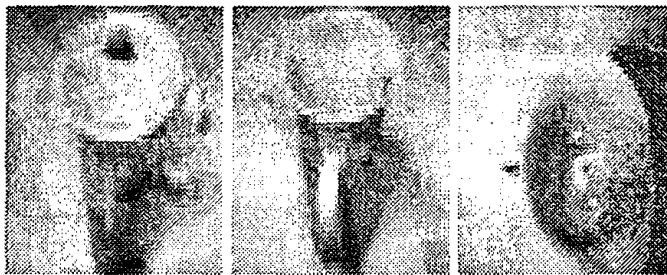
1.5. Προκλήσεις του προβλήματος της ταξινόμησης εικόνας

Αν και από το 2000 έχει γίνει ερευνητική δουλειά στο πρόβλημα της ταξινόμησης εικόνας σε κατηγορία με βάση το περιεχόμενο, το πρόβλημα θεωρείται ένα από τα πιο δύσκολα στο πεδίο της τεχνητής όρασης, γιατί παρουσιάζει ορισμένες ιδιαιτερότητες, που σχετίζονται με μεταβολές στην εμφάνιση αντικειμένων της ίδιας κατηγορίας. Για να είναι αποτελεσματική μια μέθοδος ταξινόμησης θα πρέπει να



μπορεί να χειρίζεται ικανοποιητικά τις μεταβολές αυτές. Ακολουθούν οι σημαντικότερες προκλήσεις.

1. Μεταβολές που οφείλονται στην οπτική γωνία λήψης. (Camera viewpoint variations). Αφού η εικόνα είναι η προβολή του τρισδιάστατου κόσμου (με τρισδιάστατα αντικείμενα) σε ένα χώρο δύο διαστάσεων, η οπτική γωνία φωτογράφισης του αντικειμένου μπορεί να επηρεάσει αισθητά την εμφάνισή του στην εικόνα, παρόλο που τα χαρακτηριστικά του παραμένουν αμετάβλητα. Στο σχήμα 1.3 το ίδιο αντικείμενο έχει φωτογραφηθεί από διαφορετική οπτική γωνία, ωστόσο και οι τρεις εικόνες θα πρέπει να ταξινομηθούν στην ίδια κατηγορία.



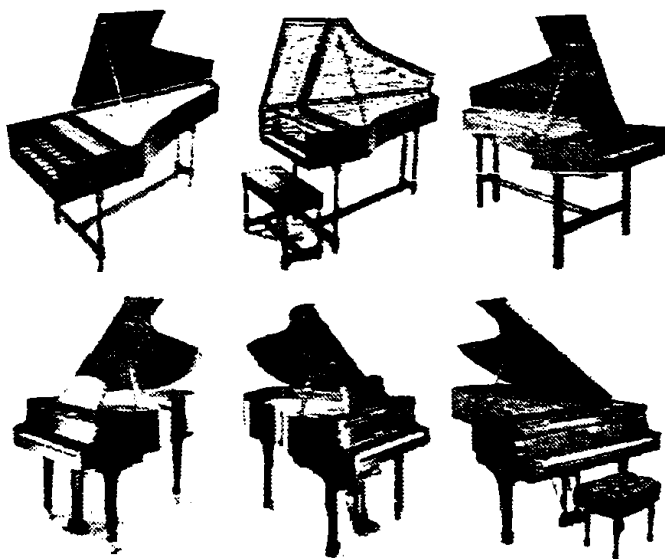
Σχήμα 1.3 Φωτογράφιση του Ίδιου Αντικειμένου με Διαφορετική Οπτική Γωνία.

2. Μεταβολές που σχετίζονται με εσωτερικά χαρακτηριστικά αντικειμένων της ίδιας κατηγορίας (Intraclass variations). Ακόμη και αν οι συνθήκες και η οπτική γωνία φωτογράφισης παραμένουν ίδιες, είναι πιθανόν αντικείμενα της ίδιας κατηγορίας να έχουν δομικές διαφορές και συνεπώς η εμφάνισή τους να διαφέρει σημαντικά. Στο σχήμα 1.4 τα αντικείμενα έχουν διαφορετικά εσωτερικά χαρακτηριστικά, διαφορετικό σχήμα, διαφορετική διάμετρο. Ωστόσο, είναι επιθυμητό να ανατεθούν στην ίδια κατηγορία από τη μέθοδο ταξινόμησης.



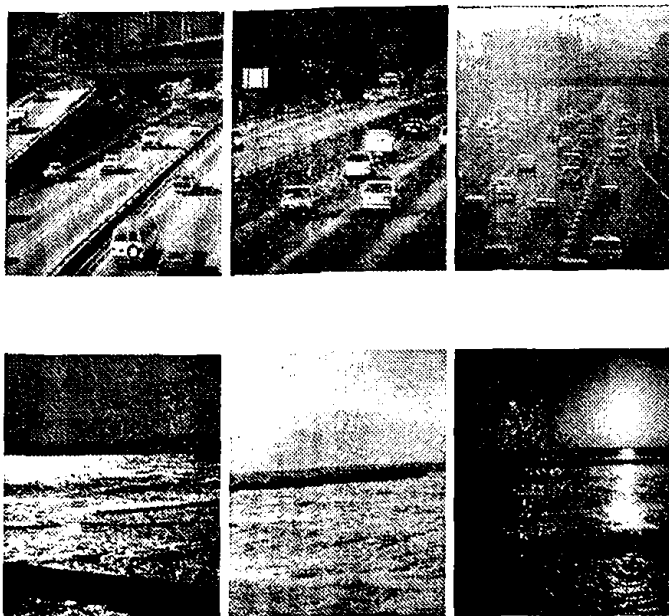
Σχήμα 1.4 Αντικείμενα της Ίδιας Κατηγορίας με Δομικές Διαφορές.

3. Μεταβολές που σχετίζονται με εσωτερικά χαρακτηριστικά αντικειμένων διαφορετικών κατηγοριών (Interclass variations). Μια σημαντική πρόκληση του προβλήματος είναι ότι είναι πιθανόν, εικόνες από διαφορετικές εικόνες να μοιάζουν τόσο ώστε να συγχέονται. Για παράδειγμα οι εικόνες στο σχήμα 1.5 απεικονίζουν αντίστοιχα αρπίχορδα και πιάνο. Αν και αυτά τα μουσικά όργανα μοιάζουν στην εξωτερική τους εμφάνιση και σχήμα, ανήκουν σε διαφορετικές κατηγορίες και το σύστημα θα πρέπει να τις κατατάσσει σε δύο κατηγορίες.



Σχήμα 1.5 Πρώτη γραμμή: Εικόνες με Αρπίχορδο. Δεύτερη γραμμή: Εικόνες με Πιάνο.

4. Συνθήκες φωτισμού (Illumination). Μια σημαντική παράμετρος του προβλήματος είναι οι συνθήκες φωτισμού στις εικόνες. Για παράδειγμα στο σχήμα 1.6 υπάρχουν 2 ομάδες εικόνων, από 3 εικόνες η καθεμία από το ίδιο τοπίο, το οποίο έχει φωτογραφηθεί κάτω από διαφορετικές συνθήκες. Το σύστημα ταξινόμησης θα πρέπει να αναθέτει αυτές τις εικόνες στην ίδια κατηγορία.



Σχήμα 1.6 Αλλαγές στις Συνθήκες Φωτισμού.

5. Μεταβολές στην κλίμακα (Scaling). Αυτή είναι μια σημαντική παράμετρος που πρέπει να ληφθεί υπόψη στο πρόβλημα της ταξινόμησης. Μπορούμε να έχουμε μια εικόνα με ένα βουνό μπροστά μας ή εικόνες με ένα βουνό στο φόντο της εικόνας, ωστόσο και οι δύο εικόνες θα πρέπει να ταξινομηθούν στην ίδια κατηγορία. Το σχήμα 1.7 δείχνει εικόνες που αφορούν σε αυτό το πρόβλημα.



Σχήμα 1.7 Αλλαγές στην Κλίμακα

6. Λοιπές μεταβολές. Στις ρεαλιστικές εικόνες, είναι συχνό το φαινόμενο ο φωτογράφος να επιθυμεί να φωτογραφίσει ένα αντικείμενο, αλλά στο πίσω μέρος της σκηνής να υπάρχουν και άλλα αντικείμενα τα οποία να μην μπορεί να αποφύγει (background clutter). Επίσης είναι εξίσου συχνό να παρεμβάλλονται διάφορα αντικείμενα ανάμεσα στο φωτογράφο και το αντικείμενο που επιθυμεί να φωτογραφίσει, επισκιάζοντας ή καλύπτοντας τμήμα του αντικειμένου που τον ενδιαφέρει (occlusion, shadows). Τα παραπάνω φαινόμενα θεωρούνται, εν γένει,

θόρυβος (noise). Στην κατασκευή μιας μεθόδου ταξινόμησης ο θόρυβος είναι μια ανεπιθύμητη κατάσταση, ωστόσο ορισμένες φορές είναι αναπόφευκτος. Στο σχήμα 1.8 δίνονται παραδείγματα εικόνων με θόρυβο.



Σχήμα 1.8. Εικόνες με Θόρυβο.

1.6. Διάρθρωση της εργασίας

Στα κεφάλαια που ακολουθούν μελετώνται μέθοδοι ταξινόμησης εικόνας και εκτιμάται πειραματικά η επίδοσή τους. Ειδικότερα, το κεφάλαιο 2 περιλαμβάνει μια εκτενή ανάλυση για την αναπαράσταση εικόνας είτε με χρήση σχήματος είτε με χρήση οπτικών χαρακτηριστικών. Μελετώνται αλγόριθμοι που χρησιμοποιούνται ευρέως στο πεδίο της τεχνητής όρασης όπως ο SIFT (Scale Space Invariant Feature), ο SURF(Speeded Up Robust Feature), ο PHOG (Pyramid Histogram of Orientation Gradients) και άλλοι. Στο κεφάλαιο 3 παρουσιάζεται το θεωρητικό υπόβαθρο στο οποίο στηρίζονται οι ταξινομητές SVM. Μελετάται ο αλγόριθμος για την εκπαίδευση ταξινομητή SVM καθώς και το πώς χρησιμοποιείται για να κάνει ταξινόμηση είτε σε δύο είτε σε πολλές κατηγορίες. Στο κεφάλαιο 4 αναλύεται ο αλγόριθμος SimpleMKL, ο οποίος χρησιμοποιείται για την κατασκευή ταξινομητών SVM με πολλαπλούς πυρήνες. Σε αυτό το κεφάλαιο μελετάται το πρόβλημα της κατασκευής ταξινομητή SVM με τη χρήση όχι απλού πυρήνα, αλλά γραμμικού συνδυασμού πυρήνων. Στο κεφάλαιο 5 παρουσιάζονται οι ακόλουθες μέθοδοι ταξινόμησης εικόνας: i) ταξινόμηση με χρήση θυλάκων από σημεία-κλειδιά, ii) ταξινόμηση με χρήση θυλάκων από σημεία-κλειδιά με βάρη, iii) ταξινόμηση με χρήση συνδυασμού χαρακτηριστικών και iv) ταξινόμηση με συνδυασμό χαρακτηριστικών με χρήση του αλγόριθμου SimpleMKL.

Στο κεφάλαιο 6 οι μέθοδοι ταξινόμησης αξιολογούνται πειραματικά με χρήση του γνωστού συνόλου δεδομένων CalTech-101. Η εργασία ολοκληρώνεται με συμπεράσματα που προκύπτουν από την εκτέλεση των πειραμάτων.



ΚΕΦΑΛΑΙΟ 2. ΑΝΑΠΑΡΑΣΤΑΣΗ ΕΙΚΟΝΑΣ

2.1 Αναπαράσταση εικόνας

2.2 Εξαγωγή χαρακτηριστικών

2.2.1 Αλγόριθμος SIFT

2.2.2 Αλγόριθμος SPIN

2.2.3 Αλγόριθμος RIFT

2.2.4 Αλγόριθμος SURF

2.3 Περιγραφή σχήματος

2.3.1 Αλγόριθμος Canny

2.3.2 Πυραμίδα από ιστογράμματα προσανατολισμού ακμών-PHOG

2.1. Αναπαράσταση εικόνας

Προκειμένου να εφαρμοστούν οι αλγόριθμοι ταξινόμησης είναι συνήθως αναγκαίο τα δεδομένα να αναπαρασταθούν ως διανύσματα. Ειδικά, στην περίπτωση των εικόνων οι αναπαραστάσεις ονομάζονται διανύσματα περιγραφής (image descriptors), που αφορούν είτε την εμφάνιση (appearance) είτε το σχήμα (shape). Για καθένα από τα παραπάνω χαρακτηριστικά έχουν προταθεί διάφοροι αλγόριθμοι εξαγωγής και υπολογισμού καθένas από τους οποίους έχει διαφορετικά πλεονεκτήματα και γι' αυτό βρίσκουν εφαρμογή σε διαφορετικές περιπτώσεις.

Για την εμφάνιση, τα τελευταία χρόνια χρησιμοποιείται ευρέως ο αλγόριθμος SIFT (Scale Invariant Feature Transform) [6] και λιγότερο άλλοι αλγόριθμοι όπως οι SURF (Speeded Up Robust Features) [7], SPIN και RIFT (Rotation Invariant Feature Transform) [25]. Οι παραπάνω αλγόριθμοι εντοπίζουν στις εικόνες σημεία-κλειδιά



(keypoints), και στη συνέχεια περιγράφονται με διανύσματα τα χαρακτηριστικά της εικόνας σε περιοχές με κέντρα τα σημεία-κλειδιά (patches).

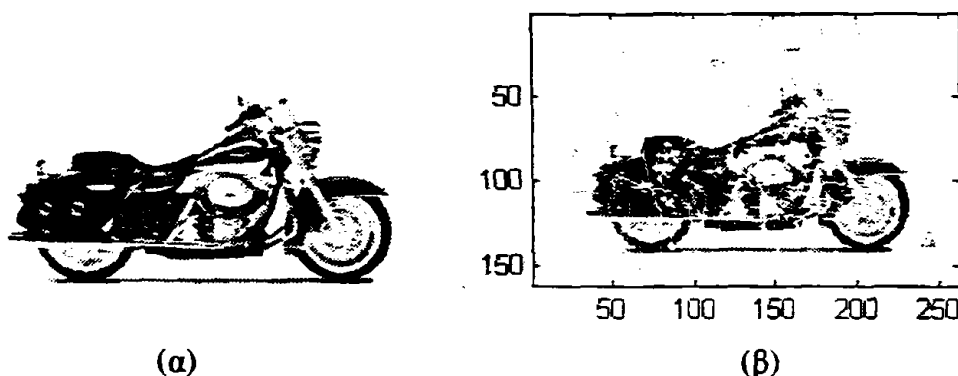
Για την περιγραφή του σχήματος οι πιο απλοί αλγόριθμοι χρησιμοποιούν τους αλγόριθμους εντοπισμού ακμών Canny και Sobel/Prewitt. Άλλοι πιο σύνθετοι αλγόριθμοι που προτάθηκαν ήταν ο HOG (Histogram of Orientation Gradient) [13], καθώς και μια επέκτασή του ο PHOG (Pyramid Histogram of Orientation Gradients) [4]. Στις επόμενες ενότητες περιγράφονται αυτοί οι αλγόριθμοι.

2.2. Εξαγωγή χαρακτηριστικών

2.2.1 Αλγόριθμος SIFT

Για τον εντοπισμό και την περιγραφή της εμφάνισης μιας εικόνας έχουν αναπτυχθεί πολλοί αλγόριθμοι στο πεδίο της τεχνητής όρασης. Ένας από αυτούς είναι ο αλγόριθμος SIFT (Scale Invariant Feature Transform) [6], ο οποίος εξάγει από την εικόνα ορισμένα σημεία-κλειδιά, και ορίζει περιοχές ενδιαφέροντος patches.

Στο παρακάτω σχήμα δίνεται ένα παράδειγμα με τα patches, που εντοπίζει ο αλγόριθμος SIFT. Συγκεκριμένα, στο σχήμα (α) δίνεται η αρχική εικόνα, ενώ στο σχήμα (β) απεικονίζονται (με μπλε βέλη) τα patches που εξάγει ο αλγόριθμος από αυτήν την αρχική εικόνα.



Σχήμα 2.1 (α) Αρχική Εικόνα. (β) Εικόνα με τα 219 Patches που Έχουν Εξαχθεί από τον Αλγόριθμο SIFT.

Ο αλγόριθμος SIFT στηρίζεται σε μια προσέγγιση διαδοχικών φίλτρων, δηλαδή αρχικά χρησιμοποιεί αποτελεσματικούς αλγόριθμους για να εντοπίσει τοποθεσίες στην εικόνα, όπου είναι πιθανό να υπάρχουν patches, και στη συνέχεια τις εξετάζει με μεγαλύτερη λεπτομέρεια για να εντοπίσει ακριβώς τα patches. Ο αλγόριθμος αποτελείται από τα ακόλουθα τέσσερα βήματα:

- ✓ Ανίχνευση ακρότατων στο χώρο της κλιμάκωσης (Scale-space extrema detection)
- ✓ Εντοπισμός σημείων-κλειδιών (Patch Localization)
- ✓ Ανάθεση προσανατολισμού (Orientation Assignment)
- ✓ Περιγραφή των σημείων-κλειδιών (Patch Description)

Ανίχνευση ακρότατων στο χώρο της κλιμάκωσης

Δεδομένης μιας εικόνας $I(x, y)$, για τον εντοπισμό τοποθεσιών όπου είναι πιθανό να υπάρχουν patches, εφαρμόζεται ένα φίλτρο σε κάθε pixel της εικόνας, ώστε να μειωθεί ο θόρυβος. Το φίλτρο που χρησιμοποιεί ο αλγόριθμος SIFT είναι Gaussian

$$G(x, y, s) = \frac{1}{2\pi s^2} e^{- (x^2 + y^2) / 2s^2}, \text{ όπου } s \text{ είναι η τυπική απόκλιση. Παίρνοντας τη}$$

συνέλιξη (convolution) (σχέση 1) του φίλτρου με την αρχική εικόνα, για τιμές του $s=1$ ως $s=2$ παράγεται μια «οικογένεια» από εικόνες στις οποίες έχει μειωθεί ο θόρυβος (Gaussian blurred images) .

$$L(x, y, s) = G(x, y, s) * I(x, y) \quad (1)$$

Στις εικόνες του πραγματικού κόσμου η εμφάνιση ενός αντικειμένου και συγκεκριμένα η κλίμακά του επηρεάζεται σημαντικά από την απόσταση του από το φωτογράφο. Δηλαδή, όταν η απόσταση αντικειμένου-φωτογράφου είναι μικρή, το αντικείμενο φαίνεται μεγάλο, ενώ όταν η απόσταση αντικειμένου-φωτογράφου είναι μεγάλη, το ίδιο αντικείμενο φαίνεται μικρό. Ωστόσο, τα patches που εντοπίζει ο αλγόριθμος θα πρέπει να είναι τα ίδια, ανεξάρτητα από την κλίμακα του αντικειμένου. Επειδή ένα σύστημα τεχνητής όρασης δεν είναι δυνατόν να γνωρίζει



την κλίμακα του αντικειμένου εκ των προτέρων, χρησιμοποιείται μια αναπαράσταση της ίδιας εικόνας, αλλά σε διαφορετική κλίμακα. (scale space representation).

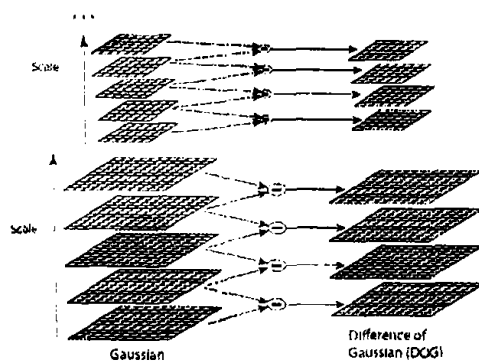
Για να είναι εφικτό το παραπάνω, δηλαδή για να παραχθεί μια σειρά φωτογραφιών με διαφορετική κλίμακα από την αρχική εικόνα, εφαρμόζουμε το φίλτρο Gauss με ένα πολλαπλασιαστικό παράγοντα $k \geq 0$, $G(x, y, ks)$ και παίρνουμε τη συνέλιξη του με την αρχική εικόνα $I(x, y)$, οπότε προκύπτει:

$$L(x, y, ks) = G(x, y, ks) * I(x, y) \quad (2)$$

Στη συνέχεια χρησιμοποιείται η συνάρτηση διαφοράς του Gauss (difference-of-Gauss) $D(x, y, \sigma)$ και προκύπτει η σχέση (3):

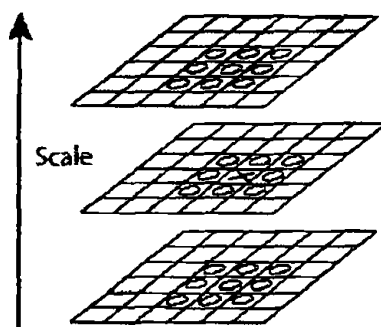
$$\begin{aligned} D(x, y, s) &= (G(x, y, ks) - G(x, y, s)) * I(x, y) \\ &= L(x, y, ks) - L(x, y, s) \end{aligned}$$

Από την παραπάνω σχέση γίνεται φανερό ότι αφαιρούμε «γειτονικές» εικόνες, δηλαδή εικόνες που διαφέρουν κατά τον πολλαπλασιαστικό παράγοντα k και παράγεται ένα νέο σύνολο από εικόνες, όπως φαίνεται και από το παρακάτω σχήμα.



Σχήμα 2.2 (αριστερά) Η Αρχική Εικόνα Αλλάζει Κλίμακα και Γίνεται Συνέλιξη των Εικόνων που Παράγονται με το Φίλτρο Gauss. (δεξιά) «Γειτονικές» Εικόνες Αφαιρούνται και Παράγεται η Συνάρτηση Διαφοράς του Gauss.

Το επόμενο βήμα αφορά στον εντοπισμό του τοπικού ελαχίστου και μεγίστου της συνάρτησης διαφοράς του Gauss. Για να γίνει αυτό, συγκρίνουμε κάθε pixel της εικόνας με τα 8 γειτονικά της, καθώς και με τα 26 κοντινότερα pixel των «γειτονικών» της εικόνων όπως φαίνεται στο παρακάτω σχήμα. Επιλέγεται το pixel του οποίου η συνάρτηση διαφορών του Gauss έχει τιμή μεγαλύτερη ή μικρότερη από τη συνάρτηση διαφορών του Gauss των υπόλοιπων pixel με τα οποία συγκρίνεται.



Σχήμα 2.3 Το Pixel που Σημειώνεται με X Συγκρίνεται με τα 8 Γειτονικά του στην Ίδια Κλίμακα και με τα 18 Pixels στις «Γειτονικές» του Εικόνες.

Εντοπισμός σημείων-κλειδιών

Για να γίνει ο εντοπισμός των σημείων-κλειδιών, αρχικά χρησιμοποιείται για κάθε υποψήφιο σημείο-κλειδί παρεμβολή των κοντινών σημείων. Η παρεμβολή υπολογίζεται με τετραγωνικό ανάπτυγμα Taylor της συνάρτησης Διαφοράς του Gauss $D(x,y,\sigma)$ ως εξής:

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

Όπου η συνάρτηση D και οι παράγωγοί της υπολογίζονται για το υποψήφιο σημείο-κλειδί και $x = (x, y, \sigma)$ είναι η απόσταση από το σημείο κλειδί. Η τοποθεσία του ακρότατου x' καθορίζεται υπολογίζοντας την παράγωγο ως προς x και θέτοντάς την ίση με 0.

Εάν η απόσταση του ακρότατου είναι μεγαλύτερη από 0.5 σε οποιαδήποτε διάσταση, αυτό είναι ένδειξη ότι το ακρότατο βρίσκεται πιο κοντά σε άλλο σημείο-κλειδί. Σε αυτήν την περίπτωση το υποψήφιο σημείο-κλειδί αλλάζει και η παρεμβολή υπολογίζεται στο σημείο αυτό. Αλλιώς, η απόσταση προστίθεται στο υποψήφιο σημείο-κλειδί για να πάρουμε την τοποθεσία του ακρότατου.

Για να απορρίψουμε τα σημεία-κλειδιά με μικρή αντίθεση υπολογίζεται το ανάπτυγμα Taylor δεύτερης τάξης στο ακρότατο. Αν η τιμή είναι μικρότερη από 0.03 τότε το σημείο-κλειδί απορρίπτεται. Αλλιώς διατηρείται με τελική θέση που δίνεται από τον τύπο:

$$y + x'$$

και κλίμακα σ , όπου y είναι η αρχική τοποθεσία και του σημείου-κλειδιού σε κλίμακα σ .

Η συνάρτηση Διαφοράς του Gauss έχει υψηλή τιμή κατά μήκος των ακμών, ακόμη και αν το υποψήφιο σημείο-κλειδί δεν είναι ανθεκτικό σε μικρά επίπεδα θορύβου. Επομένως για να αυξηθεί η σταθερότητα πρέπει να περιοριστούν τα σημεία-κλειδιά τα οποία έχουν μεγάλη απόκριση, αλλά δεν έχει προσδιοριστεί επαρκώς η θέση τους.

Για αυτό το λόγο υπολογίζονται τα ιδιοδιανύσματα του Εσσιανού πίνακα δεύτερης τάξης:

$$H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix}$$

Ανάθεση προσανατολισμού

Σε αυτό το βήμα απαιτείται να ανατεθεί προσανατολισμός σε κάθε patch που εντοπίστηκε. Ο προσανατολισμός είναι απαραίτητος, γιατί είναι πιθανόν σε δύο εικόνες το ίδιο αντικείμενο να έχει φωτογραφηθεί με διαφορετικές γωνίες περιστροφής. Ωστόσο, ο αλγόριθμος θα πρέπει να είναι σε θέση να εντοπίζει patches



σε ένα αντικείμενο, ακόμη και αν αυτό είναι περιστραμμένο. Για την ανάθεση του προσανατολισμού ακολουθείται η διαδικασία που περιγράφεται στη συνέχεια.

Δεδομένης μιας Gaussian blurred εικόνας $L(x, y)$ στην οποία έχουμε εντοπίσει το patch, για κάθε pixel αυτής υπολογίζουμε το μέγεθος κλίσης $m(x, y)$ και τον προσανατολισμό $q(x, y)$ χρησιμοποιώντας τα τέσσερα γειτονικά pixels. Ο υπολογισμός γίνεται με τις ακόλουθες σχέσεις:

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2}$$
$$q(x, y) = \tan^{-1}((L(x, y+1) - L(x, y-1)) / (L(x+1, y) - L(x-1, y)))$$

Στη συνέχεια επιλέγονται ορισμένα pixels γύρω από το patch που έχει εντοπίσει και για αυτά τα pixels φτιάχνεται ένα ιστόγραμμα με τον προσανατολισμό τους. Κάθε pixel που εισάγεται στο ιστόγραμμα, πολλαπλασιάζεται με το μέγεθος κλίσης που αντιστοιχεί σε αυτό.

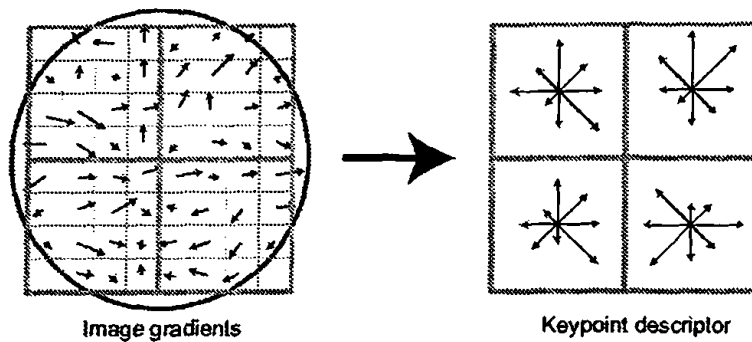
Τέλος, από το ιστόγραμμα επιλέγεται ο προσανατολισμός με την υψηλότερη συχνότητα, καθώς και οι προσανατολισμοί που αντιστοιχούν στο 80% της υψηλότερης συχνότητας και ανατίθενται στο patch.

Περιγραφή των σημείων-κλειδιών

Το τελικό στάδιο είναι η περιγραφή κάθε patch με ένα διάνυσμα, το οποίο να περιλαμβάνει πληροφορία για την θέση, τον προσανατολισμό και την κλίμακά του.

Αρχικά υπολογίζεται το μέγεθος κλίσης και ο προσανατολισμός για κάθε pixel, που είναι γειτονικό του patch που θέλουμε να περιγράψουμε, όπως φαίνεται και στο παρακάτω σχήμα (αριστερά) και στις υπολογιζόμενες τιμές εφαρμόζεται γκαουσιανό παράθυρο (μπλε κύκλος). Στη συνέχεια για κάθε μία από τις 4 υποπεριοχές κατασκευάζεται ένα ιστόγραμμα με τους προσανατολισμούς και αυτό έχει ως αποτέλεσμα να παράγεται ένα διάνυσμα περιγραφής διάστασης 2x2, όπως φαίνεται και δεξιά.





Σχήμα 2.4 Περιγραφή του Patch

Θα πρέπει να σημειωθεί ότι ο αλγόριθμος SIFT στην πραγματικότητα υπολογίζει μέγεθος κλίσης και για 8×8 περιοχή που είναι γειτονική του patch. Αυτό έχει ως αποτέλεσμα η διάσταση του υπολογιζόμενου διανύσματος να είναι 128.

Πλεονεκτήματα αλγόριθμου SIFT

Ο αλγόριθμος SIFT παρουσιάζει ορισμένα σημαντικά πλεονεκτήματα, σε σχέση με άλλους αλγόριθμους τεχνητής όρασης, όπως ο Harris detector [26] για εξαγωγή χαρακτηριστικών.

Τα 128-διάστατα διανύσματα για την περιγραφή των patches παρέχουν πληρέστερη περιγραφή σε σχέση με τα 12-διάστατα ή 16-διάστατα διανύσματα άλλων αλγορίθμων.

Τα εξαγόμενα patches είναι «ανθεκτικά» σε θόρυβο, σε συνθήκες φωτισμού και, μερικώς «ανθεκτικά» στην οπτική γωνία φωτογράφισης ενός αντικειμένου.

Σε τυπικές εικόνες εξάγεται μεγάλος αριθμός από patches και αυτό έχει ως αποτέλεσμα να εξάγονται patches και από αντικείμενα που καταλαμβάνουν μικρό τμήμα της σκηνής ή καλύπτονται, μερικώς, από άλλο αντικείμενο.

Οι απαιτούμενοι υπολογισμοί είναι αποτελεσματικοί και μπορεί να εξαχθεί μεγάλος αριθμός από patches από μια πραγματική εικόνα με μεγάλη ταχύτητα και χωρίς αυξημένες απαιτήσεις σε υλικό.

Τέλος, πρέπει να αναφερθεί ότι οι Mikolajczyk et al [8] σύγκριναν αλγόριθμους για εξαγωγή χαρακτηριστικών με τον αλγόριθμο SIFT και απέδειξαν ότι ο τελευταίος έχει καλύτερες επιδόσεις.

2.2.2 Αλγόριθμος SPIN

Ο αλγόριθμος SPIN [5] παράγει ένα διδιάστατο ιστόγραμμα με τις τιμές της φωτεινότητας της εικόνας σε μια περιοχή της. Οι διαστάσεις του ιστογράμματος είναι το d , δηλαδή η απόσταση από το κέντρο καθώς και I η ένταση της φωτεινότητας. Το ζευγάρι (d,i) είναι η πιθανότητα να υπάρχουν εικονοστοιχεία σε απόσταση d από το κέντρο του patch με τιμή φωτεινότητας i .

2.2.3 Αλγόριθμος RIFT

Ο αλγόριθμος RIFT [5] είναι επέκταση του αλγόριθμου SIFT, με τη διαφορά ότι τα patches που εξάγονται έχουν αμεταβλητότητα σε τυχόν περιστροφές. Ο αλγόριθμος RIFT διαιρεί την εικόνα σε ομόκεντρους δακτύλιους ίσου πλάτους και υπολογίζει για κάθε δακτύλιο ένα ιστόγραμμα προσανατολισμών. Για να επιτευχθεί η αμεταβλητότητα σε περιστροφές ο προσανατολισμός μιας ακμής υπολογίζεται σε κάθε σημείο, το οποίο είναι σε κατεύθυνση εκτός του κέντρου του κύκλου.

2.2.4 Αλγόριθμος SURF

Ο αλγόριθμος SURF (Speeded-Up Robust Features) υλοποιεί όπως και ο SIFT, έναν ανιχνευτή για τον προσδιορισμό σημείων ενδιαφέροντος σε μία εικόνα. Επίσης



αναπτύσσει μια νέα μέθοδο για την δημιουργία ενός περιγραφέα. Στο πρώτο βήμα, γίνεται η εύρεση σημείων ενδιαφέροντος μέσα σε μια εικόνα. Τα σημεία μοναδικότητας παραμένουν αμετάβλητα σε μεταβολές όπως η αλλαγή κλίμακας μεγέθους, η περιστροφή, η αλλαγή φωτισμού ή ο θόρυβος εικόνας. Έπειτα, η γειτονιά κάθε ενδιαφέροντος σημείου αναπαρίσταται από ένα χαρακτηριστικό διάνυσμα. Κάθε χαρακτηριστικό διάνυσμα χρησιμοποιείται στην διαδικασία ταύτισης κοινών σημείων ανάμεσα σε δύο εικόνες. Στο επόμενο βήμα, τα χαρακτηριστικά διανύσματα μια εικόνας συγκρίνονται με αυτά μιας άλλης ώστε να βρεθούν κοινά σημεία ενδιαφέροντος. Ταυτίζοντας τα κοινά στοιχεία ανάμεσα σε δύο ή περισσότερες εικόνες μπορεί να υλοποιηθεί αναγνώριση και ταξινόμηση αντικειμένων.

Στην αρχή, ο ανιχνευτής σημείων ενδιαφέροντος χρησιμοποιεί τον Γρήγορο Εσσιανό πίνακα (Fast Hessian matrix), ο οποίος βασίζεται στους υπολογισμούς του Εσσιανού πίνακα για κάθε ένα σημείο στην εικόνα (Bay [7]). Επιπλέον, από τον πίνακα εξάγεται η τοποθεσία και ο χώρος κλιμάκωσης. Έπειτα δημιουργούνται τετραγωνικά φίλτρα μεγέθους 9×9 , τα οποία προσεγγίζουν την δευτέρας τάξης παράγωγο της συνάρτησης Gauss και αναπαριστούν την χαμηλότερη κλίμακα μεγέθους. Η δημιουργία του περιγραφέα υλοποιείται σε δύο βήματα. Το πρώτο περιγράφεται από την εύρεση ενός αποτελεσματικού προσανατολισμού βασισμένο στην γύρω περιοχή του ενδιαφέροντος σημείου. Ο ιδανικός προσανατολισμός υπολογίζεται από τη σύγκριση των αθροισμάτων των αποκρίσεων που περιγράφονται από παράθυρα. Το μεγαλύτερο άθροισμα από τις συγκρινόμενες αντιστοιχίες δίνει τον προσανατολισμό του στο σημείο ενδιαφέροντος. Στη συνέχεια, για την εξαγωγή του κατασκευάζεται μια τετραγωνική περιοχή (γύρω από την περιοχή των σημείων ενδιαφέροντος) και ευθυγραμμίζεται με τον προσανατολισμό που έχει επιλεγεί στο προηγούμενο βήμα. Τελικά, παράγεται το διάνυσμα του περιγραφέα το οποίο αποτελείται από έναν πίνακα 64 στοιχείων.

2.3 Περιγραφή σχήματος

2.3.1 Αλγόριθμος Canny



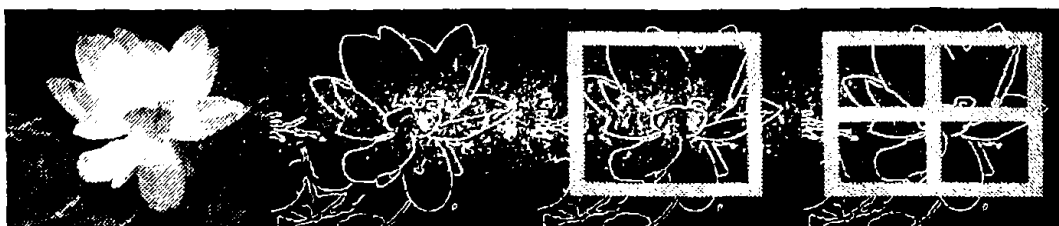
Ο αλγόριθμος Canny [10] χρησιμοποιείται για τον εντοπισμό των ακμών σε εικόνες.

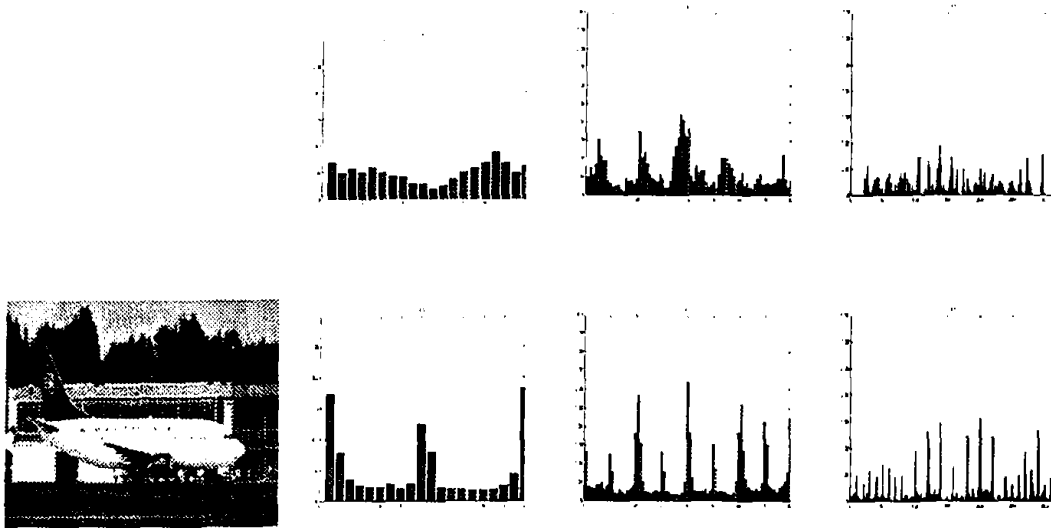
Τα βήματα που κάνει ο αλγόριθμος είναι τα ακόλουθα:

- ✓ Απομάκρυνση του θορύβου.
- ✓ Υπολογισμός της κλίσης της φωτεινότητας.
- ✓ Εφαρμογή διαδικασίας κατωφλίωσης της κλίσης φωτεινότητας για να δοθεί έμφαση στις μέγιστες τιμές της πρώτης παραγώγου της εικόνας που αντιστοιχούν σε ακμές.
- ✓ Λέπτυνση των ακμών ακολουθώντας τη διεύθυνση της ακμής και θέτοντας ίσο με 0 κάθε εικονοστοιχείο που δεν αντιστοιχεί σε μέγιστη τιμή.

2.3.2 Πυραμίδα από ιστογράμματα προσανατολισμού ακμών PHOG (Pyramid of Histograms of Orientation Gradient)

Η μέθοδος PHOG (Pyramid of Histograms of Orientation Gradients) αποτελεί μια μέθοδο αναπαράστασης μιας εικόνας και προτάθηκε από τους A. Bosch et al [4] και η οποία στηρίζεται τόσο στην αναπαράσταση εικόνας με πυραμίδα που προτάθηκε από S. Lazebnik et al [5] αλλά και στο ιστόγραμμα με προσανατολισμούς ακμών (Histogram of Orientation Gradients-HOG) των Dalal και Triggs [11]. Η μέθοδος εξάγει από την εικόνα πληροφορίες σχετικά με το σχήμα των αντικειμένων που περιέχει και το αποτέλεσμα της είναι ένα διάνυσμα περιγραφής σχήματος (shape descriptor) για την εικόνα. Δεδομένου ότι το σχήμα ενός αντικειμένου είναι η συνισταμένη όλων των ακμών που το αποτελούν, ο αλγόριθμος υπολογίζει ένα ιστόγραμμα με τους προσανατολισμούς των ακμών. Η ιδέα στην οποία στηρίζεται ο αλγόριθμος είναι ότι εικόνες με παρόμοιο σχήμα θα έχουν παρόμοιο ιστόγραμμα ακμών. Κατά συνέπεια η απόσταση μεταξύ των διανυσμάτων περιγραφής σχήματος είναι ένα μέτρο της ομοιότητας των εικόνων.





Σχήμα 2.5 Αναπαράσταση Εικόνας με Ιστόγραμμα Προσανατολισμού Ακμών.

Πρώτη Γραμμή: Διαίρεση της Αρχικής Εικόνας σε Επίπεδα και Κελιά. Παρακάτω:

Υπολογισμός Ιστογράμματος για Κάθε Εικόνα και Κάθε Επίπεδο.

Η μέθοδος PHOG επεξεργάζεται την εικόνα σε διαφορετικά επίπεδα (levels) $l=0,1,2..$ και σε κάθε ένα από αυτά τη διαιρεί σε κελιά (cells), όπως φαίνεται και στην εικόνα 2.5. Στο επίπεδο $l=0$ ολόκληρη η εικόνα αποτελεί ένα κελί. Στο επίπεδο $l=1$, η εικόνα διαιρείται σε 4 κελιά, στο $l=2$ διαιρείται σε 16 κελιά και στο επίπεδο l διαιρείται σε 2^l κελιά, με αποτέλεσμα να σχηματίζεται μια πυραμίδα από κελιά.

Ο αλγόριθμος επεξεργάζεται το κάθε κελί κάθε επιπέδου μεμονωμένα ως εξής: αρχικά εντοπίζει τις ακμές που υπάρχουν στο κελί, στη συνέχεια υπολογίζει τον προσανατολισμό τους και με βάση αυτόν τις κατανέμει σε K κάδους (bins) ενός ιστογράμματος και στη συνέχεια υπολογίζει το τελικό διάνυσμα, με συνένωση όλων των ιστογραμμάτων. Δηλαδή, για κάθε κελί υπολογίζεται ένα ιστόγραμμα με προσανατολισμούς των ακμών (HOG- Histogram of Orientation Gradients) και η συνένωση των παραπάνω ιστογραμμάτων οδηγεί σε μια πυραμίδα από ιστογράμματα προσανατολισμού ακμών. Κατά συνέπεια το επίπεδο $l=0$ αναπαρίσταται από ένα K -διάστατο διάνυσμα., το $l=1$ από ένα $4K$ -διάστατο διάνυσμα, το $l=2$ από ένα $16K$ διάνυσμα και τα L επίπεδα της εικόνας από ένα διάνυσμα με διάσταση $4K \sum_{l \in L} 4^l$.

Στην πράξη και στα πειράματα που έχουν πραγματοποιηθεί στην τρέχουσα εργασία χρησιμοποιείται μέχρι $L=3$, λόγω υπολογιστικού κόστους. Επίσης, το τελικό διάνυσμα υφίσταται κανονικοποίηση ώστε όλες οι συνιστώσες του να αθροίζουν στη μονάδα. Έτσι αποφεύγεται οι εικόνες που έχουν περισσότερες ακμές να έχουν μεγαλύτερες τιμές στους αντίστοιχους κάδους, σε σχέση με εικόνες με λιγότερες ακμές.

Για τον υπολογισμό των ακμών χρησιμοποιείται ο αλγόριθμος εντοπισμού ακμών Canny [10] και για τον υπολογισμό των προσανατολισμών των ακμών χρησιμοποιείται μια 3×3 Sobel μάσκα.

Για παράδειγμα, αν επεξεργαστούμε μια εικόνα σε 2 επίπεδα και κατανείμουμε τις ακμές σε ιστόγραμμα με 20 bins, από το επίπεδο $l=0$ προκύπτει ένα διάνυσμα διάστασης 20 και από το επίπεδο $l=1$, ένα διάνυσμα διάστασης 80. Τελικώς η εικόνα περιγράφεται από ένα διάνυσμα διάστασης 100.

Το σχήμα 2.5 [4] δείχνει ότι εικόνες της ίδιας κατηγορίας έχουν παρόμοιο διάνυσμα PHOG και εικόνες διαφορετικής κατηγορίας έχουν διαφορετικό διάνυσμα PHOG.

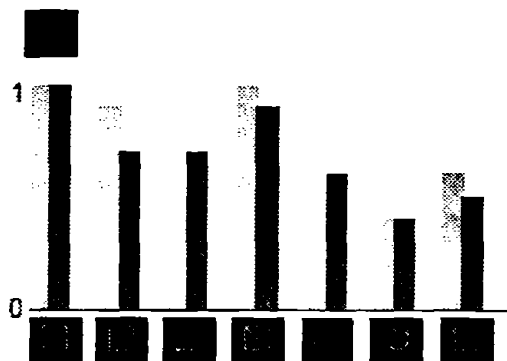
Στο σημείο αυτό θα γίνει μια σύγκριση της απόστασης διανυσμάτων PHOG από διαφορετικές εικόνες σε σχέση με την απόσταση Chamfer [32]. Η τελευταία είναι μια μέθοδος η οποία έχει χρησιμοποιηθεί σε προβλήματα, όπως η αναγνώριση αντικειμένων.

Η απόσταση Chamfer υπολογίζει την απόσταση μεταξύ δύο ακμών, υπολογίζοντας το μέσο όρο των κοντινότερων σημείων τους. Δεδομένων δύο ακμών που περιγράφονται αντίστοιχα από τα σύνολα σημείων $\{x_m\}$, $m \geq 0$ και $\{x_i\}$, $i \geq 0$ η απόσταση Chamfer [32] είναι:

$$Chamfer = \frac{1}{Nm} \sum_m \min_{x_i} \|x_m - x_i\|$$



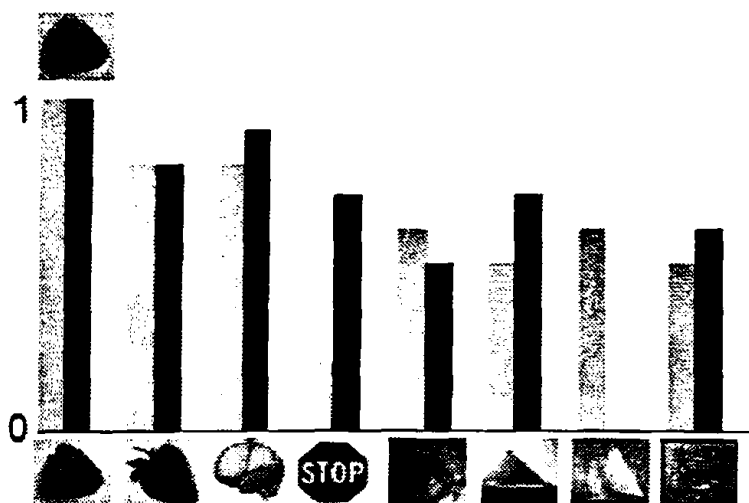
Το σχήμα 2.6 [4] συγκρίνει την απόσταση Chamfer με την απόσταση διανυσμάτων PHOG, η οποία έχει υπολογιστεί με την απόσταση χ^2 , σε ένα σύνολο από τεχνητές εικόνες. Από το σχήμα προκύπτει ότι και οι δύο αποστάσεις έχουν ανοχή σε ακμές που λείπουν (στο 2^ο και 3^ο παράδειγμα λείπει μια ακμή) καθώς και σε θόρυβο στο πίσω μέρος της εικόνας (στο 4^ο παράδειγμα έχει προστεθεί μια ακμή). Επίσης, ο αλγόριθμος PHOG χειρίζεται καλύτερα τις περιστροφές, λόγω του ότι υπολογίζει την απόσταση σε περιοχές εικόνων, ενώ η απόσταση Chamfer όχι.



Σχήμα 2.6 Σύγκριση Απόστασης Chamfer (Πράσινες Ράβδοι) με την χ^2 Απόσταση των Διανυσμάτων PHOG (γκρι ράβδοι). Στον Κάθετο Άξονα η Τιμή 1 Αντιστοιχεί σε Τέλεια Ομοιότητα, ενώ η Τιμή 0 και Τέλεια Ανομοιότητα.

Το σχήμα 2.7 [4] συγκρίνει επίσης την απόσταση Chamfer με την χ^2 απόσταση των διανυσμάτων PHOG, σε εικόνες του συνόλου δεδομένων Caltech-101. Από το σχήμα επίσης προκύπτει ότι και οι δύο αποστάσεις έχουν παρόμοια συμπεριφορά.





Σχήμα 2.7 Σύγκριση Απόστασης Chamfer (Πράσινες Ράβδοι) με την χ^2 Απόσταση των Διανυσμάτων PHOG (Γκρι Ράβδοι). Ομοιότητα μιας Εικόνας με Άλλες Εικόνες που Ανήκουν στο Σύνολο Caltech.

Αν και οι δύο αποστάσεις έχουν παρόμοια συμπεριφορά, ο αλγόριθμος PHOG έχει ορισμένα πλεονεκτήματα σε σχέση με την απόσταση Chamfer, όπως μεγαλύτερη ανθεκτικότητα σε περιστροφές. Επίσης, ο αλγόριθμος PHOG παρέχει ένα διάνυσμα περιγραφής, το οποίο μπορεί να χρησιμοποιηθεί σε αλγόριθμους μάθησης που χρησιμοποιούν πυρήνες.

ΚΕΦΑΛΑΙΟ 3. ΤΑΞΙΝΟΜΗΤΕΣ SVM ΜΕ ΧΡΗΣΗ ΑΠΛΟΥ ΠΥΡΗΝΑ

3.1 Γενικά

3.2 Εκπαίδευση ταξινομητών SVM

3.3 Συναρτήσεις πυρήνα

3.4 Πώς υπολογίζεται η έξοδος του SVM και πώς γίνεται η ταξινόμηση σε πολλές κατηγορίες

3.1 Γενικά

Η μέθοδος των Μηχανών Διατυμάτων Υποστήριξης SVM (Support Vector Machine) [23], έχει εδραιωθεί ως μια από τις πιο διαδεδομένες μεθόδους ταξινόμησης, αποτελώντας συνήθως τη βέλτιστη επιλογή για προβλήματα, όπως η ταξινόμηση κειμένων (text categorization), η αναγνώριση γραφής (handwriting recognition) και η ταξινόμηση δεδομένων έκφρασης γονιδίων (gene expression data). Η μέθοδος χρησιμοποιείται τόσο για δυαδικά προβλήματα (binary) όσο και για προβλήματα πολλών κατηγοριών (multiclass).

3.2 Εκπαίδευση ταξινομητών SVM

Στο στάδιο της εκπαίδευσης το SVM προσπαθεί να βρει ένα υπερεπίπεδο απόφασης για να διαχωρίσει τα δεδομένα του συνόλου εκπαίδευσης. Μάλιστα, η ιδιαιτερότητα του SVM έγκειται στο ότι προσπαθεί να βρει το καλύτερο υπερεπίπεδο, το οποίο



περιγράφεται με τον όρο υπερεπίπεδο μέγιστου περιθωρίου (maximum margin hyperplane) και είναι εκείνο που διαχωρίζει τα δεδομένα με τη μεγαλύτερη δυνατή απόσταση.

Δεδομένου ενός συνόλου $\{(x_i, y_i)\}_{i=1}^l$ και δύο κατηγοριών $y_i \in \{-1, 1\}$ η επιφάνεια απόφασης έχει τη μορφή

$$a^T x + b = 0$$

όπου a και b είναι οι παράμετροι του μοντέλου και το περιθώριο (margin) δίνεται από τον τύπο $margin = \frac{2}{\|a\|}$

Δεδομένου ότι πρέπει τα παραδείγματα να ταξινομούνται σωστά και το υπερεπίπεδο να έχει το μεγαλύτερο εύρος, το πρόβλημα ελαχιστοποίησης που επιχειρεί να επιλύσει το SVM είναι:

$$\min \frac{\|a\|^2}{2}$$

$$\text{subject to } y_i(a^T x_i + b) \geq 1, \forall i = 1, 2, \dots, l$$

Το παραπάνω είναι ένα πρόβλημα βελτιστοποίησης με περιορισμούς (constraint optimization problem), το οποίο πρέπει να επιλυθεί, δηλαδή, ουσιαστικά, απαιτείται προσδιορισμός των παραμέτρων (a, b) . Το παραπάνω πρόβλημα επιλύεται με πολλαπλασιαστές Lagrange και η διαχωριστική επιφάνεια απόφασης είναι η

$$\left(\sum_{i=1}^l \lambda_i y_i x_i\right)^T x + b = 0$$

Επειδή σε πραγματικές συνθήκες τα δεδομένα δεν είναι γραμμικώς διαχωρίσιμα, ο αλγόριθμος που συζητήθηκε παραπάνω δεν μπορεί να χρησιμοποιηθεί για την εύρεση της επιφάνειας απόφασης. Αντίθετα, χρησιμοποιείται μια λίγο διαφορετική προσέγγιση στην εκπαίδευση του SVM, η οποία προσπαθεί να βρει το υπερεπίπεδο



χαλαρού περιθωρίου (soft margin hyperplane), δηλαδή την επιφάνεια απόφασης που διαχωρίζει τα δεδομένα κάνοντας τα λιγότερα λάθη. Σε αυτή την περίπτωση χρησιμοποιούνται οι βοηθητικές μεταβλητές $\xi_i \geq 0$ (slack variables).

Σε κάθε πρότυπο του συνόλου εκπαίδευσης αντιστοιχεί μια τιμή ξ_i , ανάλογα με τη θέση του ως προς τη διαχωριστική επιφάνεια. Συγκεκριμένα υπάρχουν τρεις περιπτώσεις: Όταν ένα πρότυπο βρίσκεται στη σωστή πλευρά του υπερεπιπέδου και σε απόσταση μεγαλύτερη ή ίση της ελάχιστης απαιτούμενης, τότε $\xi_i = 0$. Αν το πρότυπο βρίσκεται στη σωστή πλευρά του υπερεπιπέδου, αλλά όχι σε ικανοποιητική απόσταση, τότε $0 < \xi_i < 1$. Τέλος, αν το πρότυπο βρίσκεται στη λάθος πλευρά του υπερεπιπέδου, τότε $\xi_i \geq 1$. Το χαλαρό λάθος (soft error) είναι το άθροισμα των ξ_i , των προτύπων του συνόλου εκπαίδευσης, δηλαδή:

$$\sum_i \xi_i$$

Για την εύρεση του υπερεπιπέδου χαλαρού εύρους, απαιτείται η επίλυση του εξής προβλήματος βελτιστοποίησης με περιορισμούς (constraint optimization problem):

$$\min\left(\frac{\|a\|^2}{2} + C \sum_{i=1}^l \xi_i\right)$$

$$\text{subject to } \xi_i \geq 0 \text{ and } y_i(a^T x_i + b) \geq 1 - \xi_i, \forall i = 1, 2, \dots, l$$

δηλαδή, απαιτείται ο προσδιορισμός των παραμέτρων ξ_i και της μεταβλητής C , η οποία καθορίζει το πόσο αυστηροί είμαστε με τα λάθη.

Το παραπάνω πρόβλημα βελτιστοποίησης λύνεται και αυτό με τη βοήθεια των πολλαπλασιαστών Lagrange $\lambda_i \geq 0, i = 1, 2, \dots, l$ και η επιφάνεια απόφασης που προκύπτει είναι:

$$\left(\sum_{i=1}^l \lambda_i y_i x_i^T x\right) + b = 0$$



3.3 Συναρτήσεις πυρήνα

Στην περίπτωση που τα δεδομένα του συνόλου εκπαίδευσης είναι μη γραμμικά διαχωρίσιμα, αυτά απεικονίζονται σε ένα χώρο μεγαλύτερης διάστασης με μια συνάρτηση $\Phi(x)$ και σε αυτόν τον χώρο γίνεται προσπάθεια για γραμμικό διαχωρισμό.

Σε αυτήν την περίπτωση η γραμμική επιφάνεια απόφασης είναι της μορφής:

$$a^T \Phi(x) + b = 0$$

Για την εύρεση της διαχωριστικής επιφάνειας απαιτείται η επίλυση του εξής προβλήματος ελαχιστοποίησης με περιορισμούς:

$$\min \left(\frac{\|a\|^2}{2} + C \sum_{i=1}^l \xi_i \right)$$

$$\text{subject to } \xi_i \geq 0 \text{ and } y_i (a^T \Phi(x_i) + b) \geq 1 - \xi_i, \forall i$$

το οποίο καταλήγει στο εξής δυϊκό πρόβλημα (dual problem):

$$Ll = \sum_{i=1}^l \lambda_i - \frac{1}{2} \sum_{i,j} \lambda_i \lambda_j y_i y_j \Phi(x_i)^T \Phi(x_j)$$

$$\text{subject to } \xi_i \geq 0 \text{ and } \sum_{i=1}^l \lambda_i y_i = 0$$

Μετά τον προσδιορισμό των λ_i , χρησιμοποιούνται οι KKT συνθήκες για να μετατραπούν οι ανισότητες σε ισότητες και εν συνεχεία υπολογίζονται οι παράγωγοι ως προς a , b και μηδενίζονται. Έτσι προκύπτουν οι ακόλουθες δύο σχέσεις:



$$a = \sum_{i=1}^l \lambda_i y_i \Phi(x_i)$$

$$\lambda_i (y_i (a \cdot \Phi(x_i) + b) - 1 + \xi_i) = 0 \Rightarrow \lambda_i (y_i (\sum_{j=1}^l \lambda_j y_j \Phi(x_j) \cdot \Phi(x_i) + b - 1 + \xi_i) = 0$$

Η επιφάνεια απόφασης που προκύπτει έχει την παρακάτω μορφή:

$$(\sum_{i=1}^l \lambda_i y_i \Phi(x_i) \cdot \Phi(x)) + b = 0$$

Το εσωτερικό γινόμενο στο χώρο $\Phi(x)$ ονομάζεται συνάρτηση πυρήνας (kernel function). Στην περίπτωση των προβλημάτων τεχνητής όρασης οι βασικοί πυρήνες (baseline kernels) που συνήθως χρησιμοποιούνται και συνδυάζονται είναι οι ακόλουθοι:

- ✓ Γραμμικός (linear): $k(x_i, x_j) = x_i \bullet x_j$
- ✓ Πολυωνυμικός πυρήνας (polynomial): $k(x_i, x_j) = (\gamma x_i \bullet x_j + r)^p, \gamma > 0$
- ✓ Ακτινικής Συνάρτησης Βάσης (Radial Basis Function):
 $k(x_i, x_j) = e^{-\gamma d(x_i, x_j)}$, όπου $d(x, x')$ η απόσταση των διανυσμάτων.
- ✓ Πυρήνας χ^2 (χ^2 kernel): $k(x_i, x_j) = \sum_i \frac{(x_i - y_j)^2}{x_i + y_j}$ για $x_i \geq 0, y_j \geq 0$

3.4 Πως υπολογίζεται η έξοδος του SVM και πως γίνεται η ταξινόμηση σε πολλές κατηγορίες

Σε αυτήν την παράγραφο εξετάζεται η περίπτωση του προβλήματος ταξινόμησης με πολλές προκαθορισμένες κατηγορίες. Σε αυτήν την περίπτωση, υπάρχουν δύο προσεγγίσεις: η προσέγγιση «έναντι - όλων» (one-against-all approach) και η προσέγγιση ένας-εναντίον-ένα (one-against-one approach).

Στην προσέγγιση «έναντι-εναντίον-όλων» δεδομένου ενός συνόλου M προκαθορισμένων κατηγοριών $C = \{C_1, C_2, \dots, C_M\}$, για το διαχωρισμό των προτύπων



απαιτείται εκπαίδευση M δυαδικών SVM, με τρόπο ώστε κάθε SVM να βρίσκει μια επιφάνεια απόφασης, που διαχωρίζει τα πρότυπα της κατηγορίας i από τα πρότυπα των υπόλοιπων $M-1$ κατηγοριών. Ουσιαστικά, το πρόβλημα των πολλών κατηγοριών διαιρείται σε M δυαδικά προβλήματα όπου για την εύρεση του υπερεπιπέδου απόφασης κάθε δυαδικό SVM χρησιμοποιεί τους αλγόριθμους εκπαίδευσης που περιγράφηκαν στις παραπάνω παραγράφους.

Η απόφαση για την κατηγορία του προτύπου προκύπτει με ένα είδος «ψηφοφορίας»: υπολογίζεται η έξοδος για κάθε δυαδικό SVM και το πρότυπο ταξινομείται στην κατηγορία C_j , αν η έξοδος του ταξινομητή j είναι μεγαλύτερη από τις εξόδους των υπόλοιπων ταξινομητών.

Στην περίπτωση της προσέγγισης «ένας-εναντίον-ένα», κατασκευάζεται ένας ταξινομητής SVM για κάθε ζεύγος κατηγοριών. Δεδομένου ενός συνόλου M προκαθορισμένων κατηγοριών $C=\{C_1, C_2, \dots, C_M\}$ κατασκευάζονται $M(M-1)/2$ SVM ταξινομητές και κάθε ταξινομητής διακρίνει τα παραδείγματα της μιας κατηγορίας από τα παραδείγματα κάθε άλλης κατηγορίας. Για τη λήψη της απόφασης ισχύει το σύστημα της ψηφοφορίας, όπου κάθε SVM ψηφίζει για μια κατηγορία.



ΚΕΦΑΛΑΙΟ 4. ΑΛΓΟΡΙΘΜΟΣ SIMPLEMKL

- 4.1 Γενικά
 - 4.2 Χώρος του γραμμικού συνδυασμού πυρήνων
 - 4.3 Το πρωτεύον πρόβλημα MKL
 - 4.4 Το δυικό πρόβλημα MKL
 - 4.5 Περιγραφή του αλγόριθμου SimpleMKL
 - 4.6 Συνθήκες Βελτιστότητας
 - 4.7 Ο αλγόριθμος SimpleMKL για πολλές κατηγορίες
 - 4.8 Σύγκλιση του αλγόριθμου SimpleMKL
-

4.1 Γενικά

Ο αλγόριθμος SimpleMKL προτάθηκε από τους Rakotomamonjy et al [15], με σκοπό την αποτελεσματική επίλυση του προβλήματος της μάθησης με πολλαπλούς πυρήνες (Multiple Kernel Learning-MKL). Ο αλγόριθμος χρησιμοποιεί μεθόδους βελτιστοποίησης, με τις οποίες προσδιορίζονται τόσο τα βάρη των επιμέρους πυρήνων, όσο και οι βέλτιστες παράμετροι για τον ταξινομητή SVM.

Δεδομένου ενός συνόλου από M θετικά ορισμένους πυρήνες K_m , $m=1,2,\dots,M$ το πρόβλημα MKL συνίσταται στην εύρεση του βέλτιστου γραμμικού συνδυασμού των πυρήνων

$$k^*(x, x') = \sum_{m=1}^F d_m k_m(x, x'), d_m \geq 0, \sum_m d_m = 1.$$



4.2 Χώρος του γραμμικού συνδυασμού πυρήνων

Ορισμός: Δεδομένου ενός συνόλου από M θετικά ορισμένους πυρήνες K_m , $m=1,2,\dots,M$, ο καθένας από τους οποίους συνδυάζεται με ένα Hilbert χώρο H_m , με εσωτερικό γινόμενο $\langle \cdot, \cdot \rangle_m$, ο χώρος του γραμμικού συνδυασμού των πυρήνων

$$k^*(x, x') = \sum_{m=1}^M d_m k_m(x, x'), d_m \geq 0, \sum_{m=1}^M d_m = 1 \text{ είναι } H = \bigoplus_{m=1}^M H'_m, \text{ όπου } H'_m \text{ είναι χώρος}$$

Hilbert ο οποίος παράγεται από τον χώρο H_m ως εξής:

$$H'_m = \{f \mid f \in H_m : \frac{\|f\|_{H_m}}{d_m} < \infty\} \text{ και διαθέτει το εσωτερικό γινόμενο}$$

$$\langle f, g \rangle_{H'_m} = \frac{1}{d_m} \langle f, g \rangle_m.$$

Απόδειξη

$$\forall f \in H'_m \subseteq H_m, f(x) = \langle f(\cdot), K_m(x, \cdot) \rangle_m$$

$$\begin{aligned} \text{Ισχύει} \quad &= \frac{1}{d_m} \langle f(\cdot), d_m K_m(x, \cdot) \rangle_m \\ &= \langle f(\cdot), d_m K_m(x, \cdot) \rangle_{H'_m} \end{aligned}$$

M

Αν ορίσουμε $H = \bigoplus_{m=1}^M H'_m$ το άθροισμα των επιμέρους χώρων, τότε ο H είναι

$$\text{χώρος Hilbert με πυρήνα } K(x, x') = \sum_{m=1}^M d_m K_m(x, x') \text{ [27].}$$

4.3 Το πρωτεύον πρόβλημα MKL

Δεδομένου ενός συνόλου $\{(x_i, y_i)\}_{i=1}^l$, όπου x_i αντιστοιχεί σε παράδειγμα και $y_i \in \{-1, 1\}$ είναι η αντίστοιχη κατηγορία στην περίπτωση του SVM με απλό πυρήνα, το πρόβλημα ελαχιστοποίησης συνίσταται σε:



$$\begin{aligned} \min_{f, b, \xi} \quad & \frac{1}{2} \|f\|_H^2 + C \sum_i \xi_i \\ \text{s.t} \quad & 1 - \xi_i \leq y_i(f(x_i) + b) \quad \forall i \\ & 0 \leq \xi_i \end{aligned}$$

$$\text{που } f(x) = \sum_{i=1}^l a_i^* K(x, x_i) + b^*$$

Στόχος της εκπαίδευσης είναι να προσδιοριστούν οι παράμετροι a_i^* και b^* .

Στην περίπτωση του MKL, το υπερεπίπεδο απόφασης δίνεται από τον τύπο $f(x) + b = \sum_m f_m(x) + b$, όπου καθεμία από τις επιμέρους συναρτήσεις f_m συνδυάζεται με ένα πυρήνα K_m . Ο αλγόριθμος SimpleMKL στηρίζεται στο παραπάνω θεωρητικό υπόβαθρο από το οποίο προκύπτει και το πρωτεύον πρόβλημα MKL, το οποίο ορίζεται ακολούθως:

$$\begin{aligned} \min_{\{f_m\}, b, \xi, d} \quad & \frac{1}{2} \sum_m \frac{1}{d_m} \|f_m\|_{H_m}^2 + C \sum_i \xi_i \\ \text{s.t} \quad & 1 - \xi_i \leq y_i \sum_m f_m(x_i) + y_i b \quad \forall i \\ & \xi_i \geq 0 \quad \forall i \\ & \sum_m d_m = 1, d_m \geq 0 \quad \forall m \end{aligned}$$

Στο παραπάνω πρόβλημα παρατηρεί κανείς ότι στα επιμέρους βάρη των πυρήνων επιβάλλεται ο περιορισμός να αθροίζουν στη μονάδα. Αυτό έχει ως αποτέλεσμα ορισμένα από τα βάρη να μηδενίζονται, με αποτέλεσμα και οι αντίστοιχοι πυρήνες να εξαλείφονται και να μην συμμετέχουν στο άθροισμα.

4.4 Το δυνικό πρόβλημα MKL



Από το πρωτεύον πρόβλημα που παρουσιάστηκε παραπάνω, με χρήση των πολλαπλασιαστών Lagrange προκύπτει το αντίστοιχο δυικό.

$$L = \frac{1}{2} \sum_m \frac{1}{d_m} \|f_m\|_{H_m}^2 + C \sum_i \xi_i + \sum_i a_i (1 - \xi_i - y_i \sum_m f_m(x_i) - y_i b) - \sum_i v_i \xi_i + \lambda (\sum_m d_m - 1) - \sum_m \eta_m d_m$$

Στό παραπάνω πρόβλημα οι πολλαπλασιαστές Lagrange a_i και v_i προκύπτουν από τους γνωστούς περιορισμούς του SVM, ενώ οι πολλαπλασιαστές λ και η_m , προκύπτουν από τους περιορισμούς των βαρών.

Με μηδενισμό της παραγώγου της παραπάνω σχέσης προκύπτει το ακόλουθο:

$$(a) \frac{1}{d_m} f_m(.) = \sum_i a_i y_i K_m(., x_i), \forall m$$

$$(b) \sum_i a_i y_i = 0$$

$$(c) C - a_i - v_i = 0, \forall i$$

$$(d) -\frac{1}{2} \frac{\|f_m\|_{H_m}^2}{d_m^2} + \lambda - \eta_m = 0, \forall m$$

Στις παραπάνω σχέσεις παρατηρείται ότι όταν το $f_m(.)$ μηδενίζεται, μηδενίζεται επίσης και το d_m . Εφαρμόζοντας το παραπάνω προκύπτει το εξής δυικό πρόβλημα:

$$\max_{a_i, \lambda} \sum_i a_i - \lambda$$

$$s.t \sum_i a_i y_i = 0$$

$$0 \leq a_i \leq C, \forall i$$

$$\frac{1}{2} \sum_{i,j} a_i a_j y_i y_j K_m(x_i, x_j) \leq \lambda, \forall m$$

Το παραπάνω δυικό πρόβλημα είναι δύσκολο να επιλυθεί εξαιτίας του τελευταίου περιορισμού. Αυτός ο περιορισμός μπορεί να απομακρυνθεί, αλλά σε αυτή την



περίπτωση το πρόβλημα γίνεται μη διαφορίσιμο και προκαλεί νέες δυσκολίες. Για αυτό το λόγο ο αλγόριθμος SimpleMKL, ο οποίος περιγράφεται στην ακόλουθη ενότητα, είναι μια προσέγγιση που βασίζεται στην βελτιστοποίηση του πρωτεύοντος MKL προβλήματος.

4.5 Περιγραφή του αλγόριθμου SimpleMKL

Το πρόβλημα βελτιστοποίησης το οποίο επιχειρεί να λύσει ο αλγόριθμος SimpleMKL είναι το ακόλουθο:

$$\min_d J(d) \text{ such that } \sum_{m=1}^M d_m = 1, d_m \geq 0 \text{ όπου}$$

$$J(d) = \left\{ \begin{array}{ll} \min_{\{f\}, b, \xi} & \frac{1}{2} \sum_m \frac{1}{d_m} \|f_m\|_{H_m}^2 + C \sum_i \xi_i \quad \forall i \\ \text{s.t} & 1 - \xi_i \leq y_i \sum_m f_m(x_i) + y_i b \\ & \xi_i \geq 0 \quad \forall i \end{array} \right\}$$

Η συνάρτηση με πολλαπλασιαστές Lagrange του παραπάνω προβλήματος είναι η ακόλουθη:

$$L = \frac{1}{2} \sum_m \frac{1}{d_m} \|f_m\|_{H_m}^2 + C \sum_i \xi_i + \sum_i a_i (1 - \xi_i - y_i \sum_m f_m(x_i) - y_i b) - \sum_i v_i \xi_i$$

Μηδενίζοντας τις παραγώγους του παραπάνω προβλήματος, προκύπτουν οι ακόλουθες συνθήκες:

$$(a) \frac{1}{d_m} f_m(\cdot) = \sum_i a_i y_i K_m(\cdot, x_i), \forall m$$

$$(b) \sum_i a_i y_i = 0$$

$$(c) C - a_i - v_i = 0, \forall i$$



από τις οποίες προκύπτει και το ακόλουθο δυικό πρόβλημα:

$$\begin{aligned} \max_a \quad & -\frac{1}{2} \sum_{i,j} a_i a_j y_i y_j \sum_m d_m K_m(x_i, x_j) + \sum_i a_i \\ \text{with} \quad & \sum_i a_i y_i = 0 \\ & 0 \leq a_i \leq C \quad \forall i \end{aligned}$$

Το παραπάνω πρόβλημα είναι το πρόβλημα βελτιστοποίησης του απλού SVM, στο οποίο απλώς έχει χρησιμοποιηθεί ο συνδυασμός πυρήνων

$$K(x_i, x_j) = \sum_m d_m K_m(x_i, x_j).$$

Η ύπαρξη και ο υπολογισμός της συνάρτησης $J(d)$ βασίζεται στο Θεώρημα των Bonnans και Shaphiro [16] και με παραγωγή της συνάρτησης $J(d)$ ως προς το d_m

προκύπτει η ακόλουθη σχέση $\frac{\partial J}{\partial d_m} = -\frac{1}{2} \sum_{i,j} a_i^* a_j^* y_i y_j K_m(x_i, x_j) \forall m$.

Δεδομένου ότι έχει υπολογιστεί η παράγωγος της $J(d)$ το διάνυσμα d των βαρών ενημερώνεται χρησιμοποιώντας της μέθοδο της απότομης καθόδου (gradient descent) και εξασφαλίζοντας ότι όλοι οι περιορισμοί ικανοποιούνται.

Ο περιορισμός $\sum_m d_m = 1$ εξασφαλίζεται υπολογίζοντας το reduced gradient.

Δεδομένου ότι d_μ είναι η μη μηδενική και μεγαλύτερη τιμή του διανύσματος d η reduced gradient τιμή του $J(d)$, η οποία είναι η $\nabla_{red} J$ είναι η ακόλουθη:

$$[\nabla_{red} J]_\mu = \frac{\partial J}{\partial d_m} - \frac{\partial J}{\partial d_\mu} \forall m \neq \mu \text{ και } [\nabla_{red} J]_\mu = \sum_{m \neq \mu} \left(\frac{\partial J}{\partial d_m} - \frac{\partial J}{\partial d_\mu} \right).$$



Ο περιορισμός όλα τα βάρη να είναι θετικά, επίσης λαμβάνεται υπόψη. Εφόσον είναι επιθυμητό να ελαχιστοποιηθεί η συνάρτηση $J(\cdot)$, το $-\nabla_{red} J|_m$ είναι στην κατεύθυνση της καθόδου. Ωστόσο αν κάποια τιμή βάρους d_m είναι μηδενική και $\nabla_{red} J|_m > 0$, αυτό θα παραβίαζε τον περιορισμό της θετικότητας για τα βάρη.

Επομένως για την συγκεκριμένη συνιστώσα, η κατεύθυνση της καθόδου μηδενίζεται.

Με δεδομένες τις παραπάνω παρατηρήσεις, η κατεύθυνση της καθόδου για τα βάρη είναι η ακόλουθη:

$$Dm = \begin{cases} 0 & d_m = 0 \quad \frac{\partial J}{\partial d_m} - \frac{\partial J}{\partial d_\mu} > 0 \\ -\frac{\partial J}{\partial d_m} + \frac{\partial J}{\partial d_\mu} & d_m > 0 \quad m \neq \mu \\ -\sum_{g \neq \mu, d_g > 0} \left(\frac{\partial J}{\partial d_g} - \frac{\partial J}{\partial d_\mu} \right) & m = \mu \end{cases}$$

Στον αλγόριθμο απότομης καθόδου (gradient descent) η ενημέρωση των βαρών γίνεται με βάση το σχήμα $d \leftarrow d + \gamma D$, όπου γ είναι το βήμα της καθόδου. Ωστόσο, στον αλγόριθμο SimpleMKL υπάρχει η εξής διαφοροποίηση: μόλις υπολογιστεί η κατεύθυνση καθόδου D , γίνεται αναζήτηση για να προσδιοριστεί το μέγιστο επιτρεπτό βήμα καθόδου για το οποίο η αντικειμενική τιμή μειώνεται ή όχι.

Με βάση τα παραπάνω ακολουθεί ο αλγόριθμος SimpleMKL σε μορφή ψευδοκώδικα.

Αλγόριθμος SimpleMKL

Θέσε $d_m = \frac{1}{M}, m = 1 \dots M$

Όσο το κριτήριο ελαχιστοποίησης δεν ικανοποιείται

Υπολογισμός της τιμής $J(d)$ χρησιμοποιώντας SVM με $K = \sum_m d_m K_m$



Υπολογισμός $\frac{\partial J}{\partial d_m}, m=1, \dots, M$ καθώς και της κατεύθυνσης καθόδου D

$$\mu = \arg \max_m d_m, J' = J, d' = d, D' = D$$

Όσο $J' < J$

$$d' = d, D' = D$$

$$v = \arg \min_{\{m|D_m < 0\}} -d_m / D_m, \gamma_{\max} = -d_v / D$$

$$d' = d + \gamma_{\max} D, D'_\mu = D_\mu - D_v, D'_v = 0$$

Υπολογισμός J' με χρήση SVM και με $K = \sum_m d'_m K_m$

Τέλος Όσο

Γραμμική αναζήτηση στο D για $\gamma \in [0, \gamma_{\max}]$

$$d \leftarrow d + \gamma D$$

Τέλος Όσο

Τα κριτήρια ελαχιστοποίησης στον αλγόριθμο SimpleMKL μπορεί να είναι είτε το δυικό χάσμα (duality gap), είτε οι συνθήκες KKT (Karush–Kuhn–Tucker), είτε η απόκλιση (variation) του d ανάμεσα σε δύο διαδοχικά βήματα ή ακόμη και ένας μέγιστος αριθμός επαναλήψεων. Η υλοποίηση του SimpleMKL που χρησιμοποιήσαμε στηρίζεται στο δυικό χάσμα και ακολουθεί στη συνέχεια.

4.6 Συνθήκες Βελτιστότητας

Από το πρωτεύον και δυικό πρόβλημα που παρουσιάστηκαν ήδη, προκύπτει ο ακόλουθος τύπος για το δυικό χάσμα:

$$DualGap = J(d^*) - \sum_i a_i^* + \frac{1}{2} \max_m \sum_{i,j} a_i^* a_j^* y_i y_j K_m(x_i, x_j)$$



όπου d^* και $\{a_i^*\}$ είναι οι βέλτιστες μεταβλητές του πρωτεύοντος και του δυικού προβλήματος και το $J(d^*)$ εξαρτάται από τις βέλτιστες τιμές του πρωτεύοντος προβλήματος $\{f_m^*\}$, b^* και $\{\xi_i^*\}$.

Το δυικό χάσμα στην περίπτωση του SVM με ένα πυρήνα είναι το ακόλουθο:

$$DG_{SVM} = J(d^*) + \frac{1}{2} \sum_{i,j} a_i^* a_j^* y_i y_j \sum_m d_m^* K_m(x_i, x_j) - \sum_i a_i^*$$

$$DualGap = DG_{SVM} - \frac{1}{2} \sum_{i,j} a_i^* a_j^* y_i y_j \sum_m d_m^* K_m(x_i, x_j) + \frac{1}{2} \max_m \sum_{i,j} a_i^* a_j^* y_i y_j K_m(x_i, x_j)$$

Επομένως το δυικό χάσμα στην περίπτωση του SimpleMKL έχει ένα επιπλέον κόστος σε σχέση με το δυικό χάσμα σε σχέση με το απλό SVM με πυρήνα.

Στην περίπτωση που χρησιμοποιηθεί κάποιο άλλο κριτήριο τερματισμού, όπως μια τιμή ε χρησιμοποιείται το ακόλουθο κριτήριο:

$$\max_m \sum_{i,j} a_i^* a_j^* y_i y_j K_m(x_i, x_j) - \sum_{i,j} a_i^* a_j^* y_i y_j \sum_m d_m^* K_m(x_i, x_j) \leq \varepsilon$$

Επειδή για ορισμένα προβλήματα η παραπάνω σχέση είναι δύσκολο να υπολογιστεί, χρησιμοποιούνται οι συνθήκες KKT, ως κριτήριο τερματισμού. Για το πρόβλημα που επιλύει ο SimpleMKL οι πρώτης τάξεως συνθήκες βελτιστότητας που προκύπτουν από τις συνθήκες KKT είναι:

$$\begin{aligned} \frac{\partial J}{\partial d_m} + \lambda - \eta_m &= 0, \forall m \\ \eta_m d_m &= 0, \forall m \end{aligned}$$



όπου λ , η_m είναι οι πολλαπλασιαστές Lagrange που προκύπτουν από τους περιορισμούς του προβλήματος. Αυτοί οι περιορισμοί μετασχηματίζονται στις ακόλουθες σχέσεις:

$$\begin{aligned}\frac{\partial J}{\partial d_m} &= -\lambda, d_m > 0 \\ \frac{\partial J}{\partial d_m} &\geq -\lambda, d_m = 0\end{aligned}$$

Ωστόσο, επειδή ο αλγόριθμος SimpleMKL δεν περιλαμβάνει τον υπολογισμό της παραμέτρου λ , ορίζουμε τις ποσότητες $dJ_{\min}$ και $dJ_{\max}$ ως εξής:

$$dJ_{\min} = \min_{\{d_m \mid d_m > 0\}} \frac{\partial J}{\partial d_m}$$

$$dJ_{\max} = \max_{\{d_m \mid d_m > 0\}} \frac{\partial J}{\partial d_m}$$

Οι συνθήκες βελτιστότητας προσεγγίζονται από τις ακόλουθες συνθήκες τερματισμού:

$$|dJ_{\min} - dJ_{\max}| \leq \varepsilon \text{ και } \frac{\partial J}{\partial d_m} \geq dJ_{\max} \text{ αν } d_m = 0$$

4.7 Ο αλγόριθμος SimpleMKL για πολλές κατηγορίες

Στην περίπτωση που το πρόβλημα ταξινόμησης αφορά σε πολλές κατηγορίες, συνδυάζονται δυαδικοί SVM ταξινομητές. Οι δύο προσεγγίσεις είναι η «ένανς-εναντίον-όλων» (one against all) και «ένανς-εναντίον-ένα» (one-against-one).

Δεδομένου ενός προβλήματος με P κατηγορίες, στην προσέγγιση ένας-εναντίον-όλων απαιτείται να εκπαιδευτούν P δυαδικοί ταξινομητές, όπου ο p -οστός ταξινομητής



θεωρεί θετικά τα παραδείγματα της κατηγορίας p και αρνητικά τα παραδείγματα των υπόλοιπων κατηγοριών. Στην προσέγγιση ένας-εναντίον-ένα, εκπαιδεύονται $P(P-1)/2$ δυαδικοί SVM ταξινομητές.

Ο αλγόριθμος SimpleMKL για πολλές κατηγορίες διαφέρει από τον αλγόριθμο SimpleMKL για δύο κατηγορίες, όσον αφορά στη συνάρτηση κόστους $J(d)$, η οποία στην περίπτωση των πολλών κατηγοριών βελτιστοποιεί από κοινού όλα τα ζευγάρια κατηγοριών από τον ακόλουθο τύπο:

$$J(d) = \sum_{p \in P} J_p(d)$$

όπου P είναι το σύνολο όλων των ζευγών κατηγοριών και $J_p(d)$ είναι η τιμή του δυαδικού SVM για το ζευγάρι κατηγοριών p .

Ο υπολογισμός της παραγώγου της συνάρτησης $J(d)$ δίνεται από τον ακόλουθο τύπο:

$$\frac{\partial J}{\partial d_m} = -\frac{1}{2} \sum_{p \in P} \sum_{i,j} a_{i,p}^* a_{j,p}^* y_i y_j K_m(x_i, x_j) \forall m$$

όπου $a_{j,p}$ είναι ο πολλαπλασιαστής Lagrange για το j -οστό παράδειγμα που εμπλέκεται στην p -οστή συνάρτηση απόφασης. Οι πολλαπλασιαστές Lagrange μπορούν να υπολογιστούν ανεξάρτητα για κάθε ζευγάρι.

Η προσέγγιση που περιγράφηκε παραπάνω στοχεύει στην εύρεση ενός συνδυασμού πυρήνων, ο οποίος αποτελεί λύση για όλα τα επιμέρους δυαδικά προβλήματα. Μια άλλη προσέγγιση θα ήταν η ανεξάρτητη εφαρμογή του SimpleMKL σε κάθε επιμέρους δυαδικό πρόβλημα. Ωστόσο, αν εφαρμοζόταν αυτό, θα προέκυπταν τόσοι συνδυασμοί πυρήνων όσα και τα ζεύγη δυαδικών προβλημάτων.

4.8 Σύγκλιση του αλγόριθμου SimpleMKL



Ο αλγόριθμος SimpleMKL είναι μια μέθοδος η οποία συγκλίνει. Δεδομένου ότι το

$$\text{πρόβλημα } J(d) = \left\{ \begin{array}{ll} \min_{\{f\}, b, \xi} & \frac{1}{2} \sum_m \frac{1}{d_m} \|f_m\|_{H_m}^2 + C \sum_i \xi_i \quad \forall i \\ \text{s.t} & 1 - \xi_i \leq y_i \sum_m f_m(x_i) + y_i b \\ & \xi_i \geq 0 \quad \forall i \end{array} \right\}$$

έχει πάντα λύση, δηλαδή το δυικό χάσμα είναι 0, ο υπολογισμός της παραγώγου

$$\frac{\partial J}{\partial d_m} = -\frac{1}{2} \sum_{i,j} a_i^* a_j^* y_i y_j K_m(x_i, x_j) \quad \forall m$$

είναι πάντα ακριβής και επομένως ο αλγόριθμος SimpleMKL εφαρμόζει τη μέθοδο της απότομης καθόδου σε μια συνεχώς διαφορούμενη συνάρτηση $J(\cdot)$, που ορίζεται από το simplex $\{d \mid \sum_m d_m = 1, d_m \geq 0\}$ και οδηγεί στο ολικό ελάχιστο της συνάρτησης J [19].

Ωστόσο, στην πράξη το πρόβλημα που ελαχιστοποιούμε δεν έχει πάντα λύση, διότι τα περισσότερα SVM τερματίζουν όταν το δυικό χάσμα είναι μικρότερο μιας τιμής ε . Σε αυτήν την περίπτωση, η επίλυση του SVM οδηγεί σε ένα ε -subgradient [17, 18].



5.1 Ταξινόμηση εικόνας με θύλακες από χαρακτηριστικά

5.1.2 Θύλακες χαρακτηριστικών με χρήση βαρών

5.2 Συνδυασμός χαρακτηριστικών και πυρήνων

5.2.1 Συνδυασμός χαρακτηριστικών στο πρόβλημα της ταξινόμησης εικόνας

5.2.2 Συνδυασμός χαρακτηριστικών με χρήση του αλγόριθμου SimpleMKL

5.1 Ταξινόμηση εικόνας με θύλακες από χαρακτηριστικά

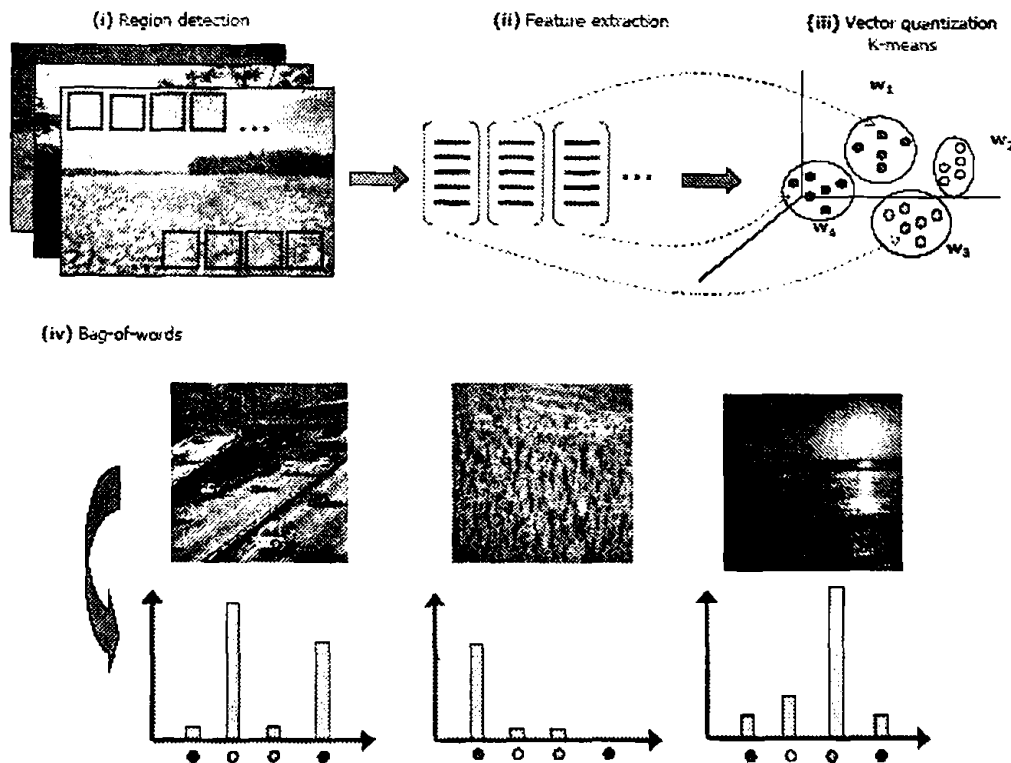
Για το πρόβλημα της ταξινόμησης εικόνας σε κατηγορία με βάση το περιεχόμενο έχουν προταθεί διάφορες μέθοδοι, μεταξύ των οποίων και οι θύλακες από χαρακτηριστικά (Bag of Features-BOF), οι οποίες πρωτοχρησιμοποιήθηκαν για ταξινόμηση κειμένων. Αυτές οι μέθοδοι προεπεξεργάζονται τις εικόνες, εξάγοντας από αυτές κάποιο χαρακτηριστικό και δημιουργώντας ιστόγραμμα, ενώ ακολούθως κατασκευάζουν μια συνάρτηση-ταξινομητή.

Η ερευνητική δουλειά που έχει πραγματοποιηθεί μέχρι στιγμής εστιάζει στη δημιουργία θυλάκων από σημεία-κλειδιά και εκμεταλλεύεται πληροφορίες που σχετίζονται με την εμφάνιση των εικόνων (visual categorization with bags of keypoints) [3].

Το μοντέλο ταξινόμησης που χρησιμοποιεί θύλακες από σημεία-κλειδιά συνίσταται στα ακόλουθα βήματα:

- ✓ Εξαγωγή χαρακτηριστικών (π.χ χρώμα, υφή, εμφάνιση κ.λ.π) από το σύνολο δεδομένων.
- ✓ Δημιουργία ενός οπτικού λεξικού (visual vocabulary), με χρήση ενός αλγόριθμου ομαδοποίησης (clustering).
- ✓ Για κάθε εικόνα, κατασκευή ενός ιστογράμματος, το οποίο εκφράζει τη συχνότητα των οπτικών χαρακτηριστικών στην εικόνα.
- ✓ Κατασκευή ενός μοντέλου ταξινόμησης, με βάση τα υπολογισθέντα ιστογράμματα.

Από την εικόνα εξάγονται οπτικές λέξεις και η αναπαράσταση των εικόνων ονομάζεται θύλακες από σημεία-κλειδιά (bag-of-keypoints). Το παρακάτω σχήμα δείχνει τη διαδικασία με την οποία εξάγονται θύλακες από σημεία-κλειδιά από τις εικόνες.



Σχήμα 5.1 Ταξινόμηση Εικόνας σε Κατηγορία με Θύλακες από Σημεία-Κλειδιά.

Η μέθοδος με τους θύλακες είναι «άτακτη» (orderless) με την έννοια ότι χρησιμοποιεί μόνο τη συχνότητα εμφάνισης των χαρακτηριστικών στην εικόνα και δεν εκμεταλλεύεται πληροφορίες που σχετίζονται για παράδειγμα με τη φυσική θέση του χαρακτηριστικού πάνω στην εικόνα ή με τη θέση ενός χαρακτηριστικού ως προς ένα άλλο.

Περαιτέρω επεκτάσεις της μεθόδου, οι οποίες έχουν προταθεί, περιλαμβάνουν δειγματοληψία (sampling)[24] για μείωση του χρόνου προεπεξεργασίας των εικόνων.

5.1.2 Θύλακες χαρακτηριστικών με χρήση βαρών

Μια επέκταση της μεθόδου των θυλάκων από σημεία-κλειδιά είναι η χρήση βαρών, ώστε ο ταξινομητής να εκμεταλλεύεται και πληροφορία, σχετικά με τη σημασία μιας οπτικής λέξης για μια δεδομένη κατηγορία.

Δοθείσης μιας λέξης w_i , η σημασία της λέξης για μια κατηγορία C_j , προσδιορίζεται από τον ακόλουθο τύπο:

Weight (w_i)=σύνολο εμφανίσεων της λέξης στην κατηγορία C_j / σύνολο εμφανίσεων της λέξης σε όλες τις κατηγορίες

Από τον παραπάνω τύπο προκύπτει ότι οι λέξεις που εμφανίζονται συχνά σε μια κατηγορία έχουν μεγάλη τιμή βάρους, ενώ οι λέξεις που είναι πιο σπάνιες έχουν μικρή τιμή βάρους.

Μετά τον προσδιορισμό των βαρών, η κάθε συνιστώσα του διανύσματος περιγραφής κάθε εικόνας πολλαπλασιάζεται με το αντίστοιχο βάρος που έχει υπολογιστεί στο προηγούμενο βήμα. Αυτό έχει ως αποτέλεσμα, οι συνιστώσες που έχουν μεγάλο βάρος και είναι σημαντικές τελικά να αποκτούν μεγάλη τιμή, ενώ οι συνιστώσες με μικρό βάρος αποκτούν μικρή τιμή.



Στο επόμενο στάδιο της μεθόδου, κατασκευάζεται ο ταξινομητής με χρήση των προεπεξεργασμένων διανυσμάτων.

5.2 Συνδυασμός χαρακτηριστικών και πυρήνων

Τα τελευταία χρόνια η τάση που επικρατεί στο πεδίο της τεχνητής όρασης είναι ο συνδυασμός διαφορετικών χαρακτηριστικών (feature combination) και κατά συνέπεια και ο συνδυασμός διαφορετικών συναρτήσεων-πυρήνων (kernel combination) [20]. Ειδικά για το πρόβλημα της ταξινόμησης εικόνας σε κατηγορία, υπάρχουσες μέθοδοι συνδυάζουν διαφορετικά χαρακτηριστικά των εικόνων, όπως σχήμα, εμφάνιση, χρώμα, υφή κ.λπ.

Στο πρόβλημα της ταξινόμησης εικόνας, η συνάρτηση-πυρήνας είναι ένα μέτρο της ομοιότητας των εικόνων. Στην περίπτωση του συνδυασμού χαρακτηριστικών, για κάθε χαρακτηριστικό υπολογίζεται μια συνάρτηση πυρήνας. Δεδομένου ενός χαρακτηριστικού $m=1,2,\dots,F$, η συνάρτηση πυρήνας για αυτό το χαρακτηριστικό είναι k_m , και η ομοιότητα δύο εικόνων x και x' ως προς το χαρακτηριστικό f_m είναι $k_m(x, x') = k(f_m(x), f_m(x'))$. Επί παραδείγματι, αν το χαρακτηριστικό που έχουμε στη διάθεσή μας αφορά την υφή μιας εικόνας, η συνάρτηση πυρήνα παραπάνω υπολογίζει την ομοιότητα των εικόνων ως προς το χαρακτηριστικό αυτό και μόνο.

Οι απλούστεροι τρόποι με τους οποίους μπορούν να συνδυαστούν πυρήνες είναι ο υπολογισμός ενός αριθμητικού μέσου όρου (averaging) και ο υπολογισμός ενός γεωμετρικού μέσου όρου (product) των βασικών πυρήνων (baseline kernels).

Ο αριθμητικός μέσος όρος θεωρεί ότι οι πυρήνες είναι ισοδύναμοι, και ορίζεται ως ακολούθως:

$$k^*(x, x') = \frac{1}{F} \sum_{m=1}^F k_m(x, x')$$

Ο γεωμετρικός μέσος όρος θεωρεί επίσης ότι οι πυρήνες συμμετέχουν ισοδύναμα στο υπολογισμό του αποτελέσματος, που δίνεται από την έκφραση:



$$k^*(x, x') = (\prod_{m=1}^F k_m(x, x'))^{1/F}$$

Οι παραπάνω δύο μέθοδοι συνδυασμού πυρήνων έχουν το μειονέκτημα ότι όλα τα χαρακτηριστικά συμμετέχουν στον υπολογισμό του τελικού αποτελέσματος, ενώ ορισμένα από αυτά μπορεί να είναι εντελώς άχρηστα. Για παράδειγμα, το χρώμα είναι ένα χαρακτηριστικό το οποίο πιθανόν να έχει καλή επίδοση σε ταξινόμηση εικόνων με καμηλοπαρδάλεις και ζέβρες, ενώ το σχήμα στις παραπάνω δύο κατηγορίες είναι παραπλήσιο. Για αυτό το λόγο προτάθηκε ο παρακάτω συνδυασμός των βασικών πυρήνων με βάρη, ο οποίος στην περίπτωση που κάποια από τα βάρη είναι μηδενικά, ουσιαστικά κάνει επιλογή χαρακτηριστικών (feature selection). Η παρούσα διατριβή εστιάζει το ενδιαφέρον της σε αυτό το συνδυασμό πυρήνων. [18]

$$k^*(x, x') = \sum_{m=1}^F d_m k_m(x, x'), d_m \geq 0, \sum_m d_m = 1$$

5.2.1 Συνδυασμός χαρακτηριστικών στο πρόβλημα της ταξινόμησης εικόνας

Στην περίπτωση των προβλημάτων τεχνητής όρασης οι βασικοί πυρήνες (baseline kernels) που συνήθως χρησιμοποιούνται και συνδυάζονται είναι οι ακόλουθοι:

- ✓ Γραμμικός (linear): $k(x, x') = x \bullet x'$
- ✓ Πολυωνυμικός πυρήνας (polynomial): $k(x, x') = (\gamma x \bullet x' + r)^p, \gamma > 0$
- ✓ Ακτινικής Συνάρτησης Βάσης (Radial Basis Function):
 $k(x, x') = e^{-\gamma d(x, x')}$, όπου $d(x, x')$ η απόσταση των διανυσμάτων.
- ✓ Πυρήνας χ^2 (χ^2 kernel): $k(x, x') = \sum_i \frac{(x_i - y_i)^2}{x_i + y_i}$ για $x_i, y_i \geq 0$

Στο πεδίο της τεχνητής όρασης και ειδικά για το πρόβλημα της ταξινόμησης εικόνας σε κατηγορία έχουν πραγματοποιηθεί πειράματα στα οποία συνδυάζεται το σχήμα και η εμφάνιση με σκοπό την κατασκευή συστήματος ταξινόμησης. Ο συνδυασμός



χαρακτηριστικών είχε καλύτερα αποτελέσματα σε σύγκριση με τη χρήση ενός μόνο χαρακτηριστικού [4].

Ειδικότερα, προτάθηκαν δύο μέθοδοι, στις οποίες ο τελικός πυρήνας προέκυψε από συνδυασμό εμφάνισης και σχήματος.

Η πρώτη μέθοδος είναι γραμμικός συνδυασμός των πυρήνων σχήματος και εμφάνισης και ο τελικός πυρήνας προκύπτει από τη σχέση:

$$K(x, y) = \alpha K_A(x_{app}, y_{app}) + \beta K_S(x_{shp}, y_{shp})$$

όπου K_A είναι ο πυρήνας ο οποίος υπολογίζεται για τα διανύσματα περιγραφής εμφάνισης x_{app} και K_S είναι ο πυρήνας για τα διανύσματα περιγραφής σχήματος x_{shp} .

Για τον προσδιορισμό των παραμέτρων α, β στον παραπάνω πυρήνα έχουν προταθεί δύο προσεγγίσεις: η εύρεση των παραμέτρων από κοινού για όλες τις κατηγορίες (Global Feature Weights-GFW) και η εύρεση των παραμέτρων χωριστά για κάθε κατηγορία (Class-Specific Feature Weights-CFW). Στην πρώτη προσέγγιση προσδιορίζονται οι παράμετροι α και β , οι οποίες είναι κοινές για όλες τις κατηγορίες. Στην δεύτερη προσέγγιση, για κάθε μια κατηγορία C χωριστά προσδιορίζεται ένα ζευγάρι παραμέτρων (α_c, β_c) . Τα πραγματοποιηθέντα πειράματα [4] απέδειξαν ότι η δεύτερη προσέγγιση έχει καλύτερη επίδοση από την πρώτη.

Η δεύτερη μέθοδος που προτάθηκε επιλέγει τη μέγιστη τιμή από το σχήμα ή την εμφάνιση προκύπτει από τη σχέση:

$$K(x, y) = \max[K_A(x_{app}, y_{app}), K_S(x_{shp}, y_{shp})]$$

5.2.2 Συνδυασμός συνιστωσών χαρακτηριστικών με χρήση του αλγόριθμου SimpleMKL



Για την ταξινόμηση εικόνας, στα πλαίσια της παρούσας εργασίας υλοποιήθηκε μια μέθοδος, η οποία αφορά στον αλγόριθμο SimpleMKL που περιγράψαμε στο προηγούμενο κεφάλαιο για το συνδυασμό συνιστωσών χαρακτηριστικών. Το χαρακτηριστικό του οποίου συνδυάστηκαν οι συνιστώσες ήταν το σχήμα.

Δεδομένου ότι στην ταξινόμηση εικόνας, υπάρχει περίπτωση κάποιες συνιστώσες χαρακτηριστικών να είναι σημαντικές και κάποιες άλλες όχι, σε αυτή την μέθοδο χρησιμοποιήσαμε ένα γραμμικό συνδυασμό με τον περιορισμό το άθροισμα των επιμέρους συντελεστών του αθροίσματος να είναι ένα. Με τον τρόπο αυτό, οι σημαντικές συνιστώσες έχουν μεγάλο συντελεστή και οι ασήμαντες μικρό συντελεστή. Για τον βέλτιστο προσδιορισμό των συντελεστών χρησιμοποιήθηκε ο αλγόριθμος SimpleMKL [15] που προτάθηκε για την εκπαίδευση ταξινομητών SVM με γραμμικό συνδυασμό πολλών πυρήνων.

Ορισμός: Δεδομένου ενός συνόλου εικόνων $\{(x_i, y_i)\}_{i=1, \dots, N}$, όπου $x_i \in X$ και $y_i \in C$, και δοθέντων F συνιστωσών χαρακτηριστικών, ο συνδυασμός χαρακτηριστικών έγκειται στον υπολογισμό ενός πυρήνα ως γραμμικού συνδυασμού των επιμέρους

$$k^*(x, x') = \sum_{m=1}^F d_m k_m(x, x'), d_m \geq 0, \sum_m d_m = 1$$

Ο συνδυασμός συνιστωσών χαρακτηριστικών με χρήση του αλγόριθμου SimpleMKL περιλαμβάνει τα ακόλουθα βήματα:

- ✓ Εξαγωγή σχήματος από το σύνολο δεδομένων.
- ✓ Επιλογή του κατάλληλου πυρήνα-παραμέτρων.
- ✓ Εκπαίδευση του ταξινομητή SVM με το σύνολο εκπαίδευσης- και προσδιορισμός d_m , $m=1, 2, \dots, F$ με χρήση του αλγόριθμου SimpleMKL του γραμμικού αθροίσματος πυρήνων συνιστωσών χαρακτηριστικών.
- ✓ Υπολογισμός της επίδοσης ταξινόμησης με τη χρήση του συνόλου ελέγχου.



ΚΕΦΑΛΑΙΟ 6. ΠΕΙΡΑΜΑΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ

6.1 Γενικά

6.2 Το σύνολο δεδομένων CalTech-101

6.3 Πειραματική εκτίμηση της μεθόδου ταξινόμησης με θύλακες από σημεία-κλειδιά

6.4 Πειραματική αξιολόγηση της μεθόδου με χρήση σχήματος

6.5 Πειραματική αξιολόγηση του συνδυασμού εμφάνισης και σχήματος

6.6 Πειραματική αξιολόγηση του συνδυασμού συνιστωσών με τον αλγόριθμο SimpleMKL

6.7 Συμπεράσματα

6.1 Γενικά

Σε αυτό το κεφάλαιο αναφέρονται τα αποτελέσματα τα οποία προέκυψαν από την εκτέλεση πειραμάτων σε όλη τη διάρκεια εκπόνησης της παρούσας εργασίας. Το σύνολο δεδομένων που χρησιμοποιήθηκε σε όλα τα πειράματα είναι το Caltech-101 [31] και ο ταξινομητής που χρησιμοποιήθηκε ήταν το λογισμικό LIBSVM [29-30].

Στα πλαίσια της εργασίας έγιναν τα τρία ακόλουθα σύνολα πειραμάτων: α) πειράματα για τη μέθοδο ταξινόμησης με θύλακες από σημεία κλειδιά, β) πειράματα στα οποία εξήχθησαν πληροφορίες από τις εικόνες με τον αλγόριθμο PHOG, γ) πειράματα στα οποία συνδυάστηκε σχήμα και εμφάνιση.

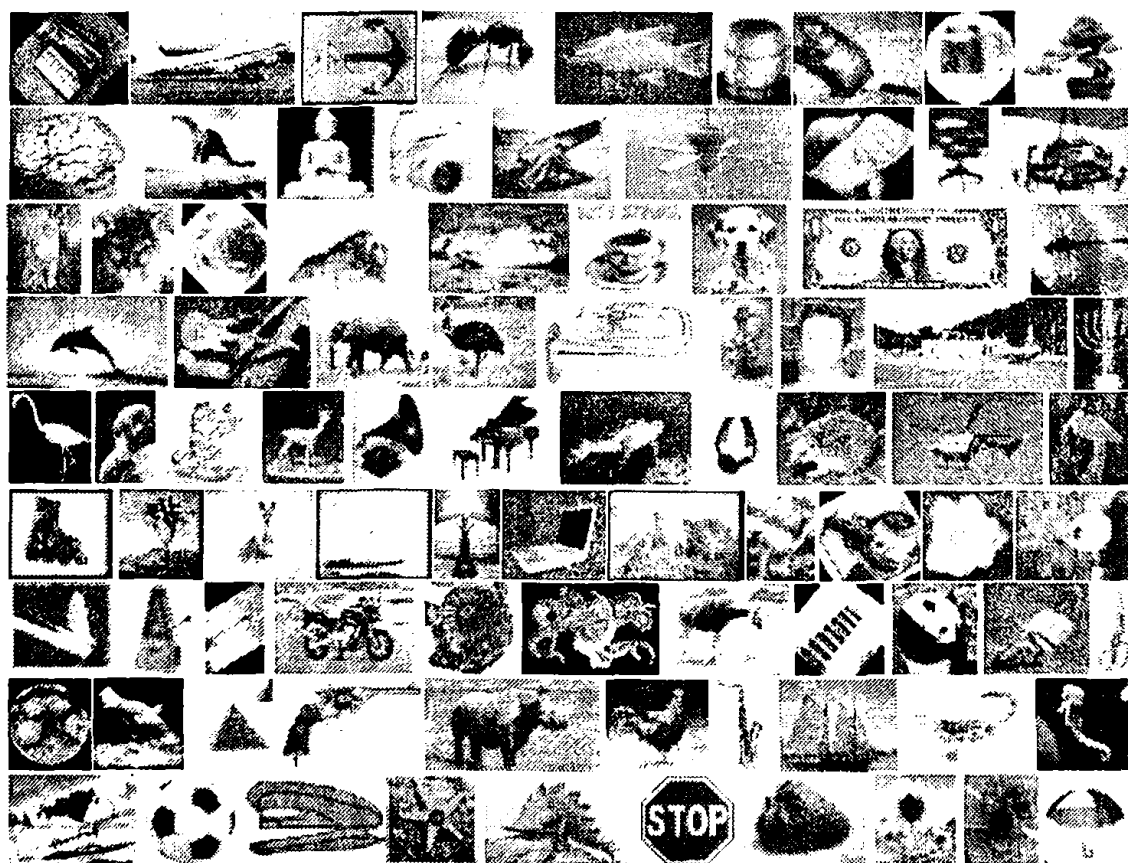
Η μετρική που χρησιμοποιήθηκε για την εκτίμηση της μεθόδου ήταν το ποσοστό επιτυχούς ταξινόμησης στο σύνολο ελέγχου, δηλαδή το ποσοστό των παραδειγμάτων



που η πραγματική τους κατηγορία (ground truth) ταυτίζεται με την κατηγορία στην οποία το ανέθεσε η εκάστοτε μέθοδος.

6.2 Το σύνολο δεδομένων CalTech-101

Το σύνολο δεδομένων Caltech-101 είναι ένα πειραματικό σύνολο από εικόνες που συλλέχθηκαν από τους Fei Fei et al [31] και χρησιμοποιείται ευρέως σε ταξινόμηση εικόνων. Περιλαμβάνει 101 κατηγορίες εικόνων και κάθε κατηγορία έχει 31 ως 800 εικόνες, μέτριας ανάλυσης 300x300 pixels. Οι εικόνες περιλαμβάνουν λίγο θόρυβο, ενώ τα αντικείμενα βρίσκονται στο κέντρο της εικόνας και καταλαμβάνουν το μεγαλύτερο μέρος της. Επίσης, τα περισσότερα αντικείμενα έχουν φωτογραφηθεί από την ίδια οπτική γωνία. Το σχήμα 6.1 δείχνει εικόνες από το σύνολο Caltech-101.



Σχήμα 6.1 Εικόνες από το Σύνολο Caltech-101.

6.3 Πειραματική εκτίμηση της μεθόδου ταξινόμησης με θύλακες από σημεία-κλειδιά



Αρχικά πραγματοποιήθηκαν πειράματα για τη μέθοδο των θυλάκων από σημεία κλειδιά. Χρησιμοποιήθηκε ο αλγόριθμος SIFT για την εξαγωγή των σημείων-κλειδιών, ο αλγόριθμος k-means για την ομαδοποίησή τους και στη συνέχεια ο ταξινομητής SVM.

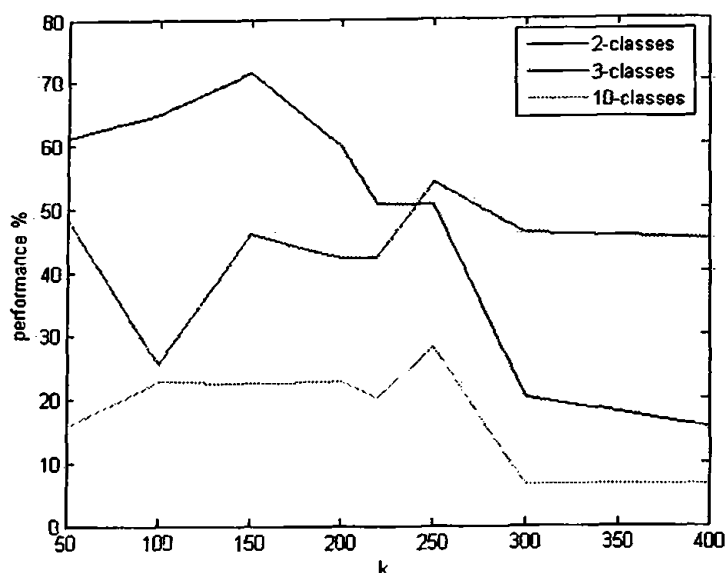
Επίδραση της παραμέτρου k του αλγόριθμου k-means στα πειραματικά αποτελέσματα.

Με σκοπό να εκτιμηθεί η επίδραση της παραμέτρου k του αλγόριθμου ομαδοποίησης στο ποσοστό επιτυχούς ταξινόμησης έγιναν πειράματα σε δύο, τρεις και δέκα κατηγορίες οι οποίες επιλέχθηκαν τυχαία από το Caltech-101. Χρησιμοποιήθηκε ο αλγόριθμος SIFT, ο οποίος στην περίπτωση των δύο κατηγοριών εντόπισε 13554 patches, ενώ στην περίπτωση των τριών και δέκα κατηγοριών εντοπίστηκαν 19062 και 53370 patches αντίστοιχα.

Για να πετύχουμε την ταχύτερη επεξεργασία των δεδομένων, επιλέχθηκαν τυχαία από κάθε κατηγορία δέκα παραδείγματα για το σύνολο εκπαίδευσης, ενώ τα υπόλοιπα παραδείγματα κάθε κατηγορίας χρησιμοποιήθηκαν ως σύνολο ελέγχου. Αυτό είχε ως αποτέλεσμα να έχουμε μέγεθος συνόλου εκπαίδευσης-μέγεθος συνόλου ελέγχου: 20 παραδείγματα-77 παραδείγματα, 30 παραδείγματα-109 παραδείγματα, 100 παραδείγματα-627 παραδείγματα για τις δύο, τρεις και δέκα κατηγορίες αντίστοιχα. Στα πειράματα όπου χρησιμοποιήθηκε ο gbf πυρήνας, ο προσδιορισμός του βέλτιστου ζεύγους παραμέτρων (C, g) του SVM έγινε χρησιμοποιώντας 10-fold cross validation.

Στα πειράματα ξεκινήσαμε από μια μικρή τιμή του k και την αυξήσαμε σταδιακά. Στο παρακάτω σχήμα απεικονίζεται στον κατακόρυφο άξονα το ποσοστό επιτυχούς ταξινόμησης και στον οριζόντιο η τιμή του k . Οι τρεις γραμμές αντιστοιχούν σε 2 κατηγορίες (μπλε χρώμα), 3 κατηγορίες (κόκκινο χρώμα) και 10 κατηγορίες (πράσινο χρώμα) και σε κάθε μία περίπτωση τὸ μεγαλύτερο πρόβλημα είναι υπερσύνολο του μικρότερου προβλήματος.





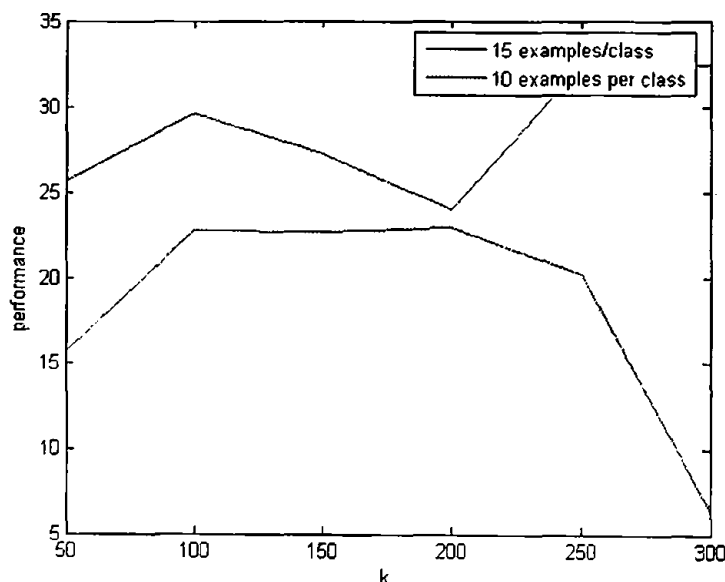
Σχήμα 6.2 Γραφική Παράσταση για Σύγκριση της Μεθόδου για Δύο, Τρεις και Δέκα Κατηγορίες σε Συνάρτηση με την Τιμή της Παραμέτρου k .

Από το γράφημα εξάγονται δύο σημαντικά αποτελέσματα. Αρχικά, για τις δύο κατηγορίες τα πειραματικά αποτελέσματα είναι εμφανώς καλύτερα σε σύγκριση με τις τρεις και τις δέκα κατηγορίες. Αυτό είναι αναμενόμενο δεδομένου ότι καθώς αυξάνεται το πλήθος των κατηγοριών, αυξάνεται και η πολυπλοκότητα του προβλήματος και κατά συνέπεια είναι πιο δύσκολο να κατασκευαστεί μια συνάρτηση απόφασης που θα διαχωρίζει τα παραδείγματα. Επίσης, όταν η τιμή της παραμέτρου k αποκτήσει μεγάλη τιμή η απόδοση της μεθόδου πέφτει. Καθώς αυξάνεται η τιμή της παραμέτρου k , μέχρι κάποιο σημείο το οπτικό λεξικό που κατασκευάζουμε είναι επαρκές για να συμπεριλάβει όλα τα σημεία κλειδιά που υπάρχουν στις εικόνες και κατά συνέπεια το ποσοστό επιτυχούς ταξινόμησης αυξάνεται. Όταν η τιμή της παραμέτρου είναι πολύ μικρή, η επίδοση δεν είναι υψηλή, γιατί διαφορετικά σημεία κλειδιά καταχωρούνται στην ίδια λέξη στο οπτικό λεξικό. Όταν η τιμή της παραμέτρου είναι πολύ μεγάλη το οπτικό λεξικό είναι πολύ μεγάλο και ίσως να περιλαμβάνει και κενές καταχωρίσεις.

Πλήθος παραδειγμάτων του συνόλου εκπαίδευσης.

Επειδή τα πειραματικά αποτελέσματα της μεθόδου δεν ήταν υψηλά, κρίθηκε σκόπιμο να μελετήσουμε το αν το πλήθος των παραδειγμάτων που επιλέγουμε από κάθε

κατηγορία για να σχηματίσουμε το σύνολο εκπαίδευσης είναι σημαντικό. Για αυτό το λόγο κάναμε το ακόλουθο πείραμα στις παραπάνω δέκα κατηγορίες. Αρχικά επιλέξαμε από κάθε κατηγορία από δεκαπέντε παραδείγματα (επομένως το σύνολο εκπαίδευσης αποτελείτο από εκατόν πενήντα παραδείγματα) και στη συνέχεια επιλέξαμε από κάθε κατηγορία από δέκα παραδείγματα (το μέγεθος του συνόλου εκπαίδευσης ήταν εκατό εικόνες). Στην πρώτη περίπτωση τα patches που εντόπισε ο αλγόριθμος SIFT ήταν 78440, ενώ στην δεύτερη περίπτωση 53370 όπως και παραπάνω. Και στις δύο περιπτώσεις χρησιμοποιήθηκε ο πυρήνας `rbf` όπως και παραπάνω. Στην παρακάτω γραφική παράσταση απεικονίζονται οι δύο περιπτώσεις για τις ίδιες τιμές της παραμέτρου k .



Σχήμα 6.3 Γραφική Παράσταση για Σύγκριση της Μεθόδου με 10 και 15 Παραδείγματα/Κατηγορία ως Σύνολο Εκπαίδευσης

Όπως φαίνεται και στην γραφική παράσταση, για κάθε τιμή του k εκτελέστηκαν δύο πειράματα: ένα πείραμα όπου επιλέχθηκαν 15 εικόνες από κάθε κατηγορία ως σύνολο εκπαίδευσης και ένα όπου επιλέχθηκαν 10 εικόνες από κάθε κατηγορία. Από τη γραφική παράσταση προκύπτει εμφανώς ότι το μεγαλύτερο σύνολο εκπαίδευσης έχει καλύτερη επίδοση, γεγονός που δικαιολογείται δεδομένου ότι ο ταξινομητής έχει στην διάθεσή του περισσότερη πληροφορία για να διαχωρίσει τα δεδομένα.

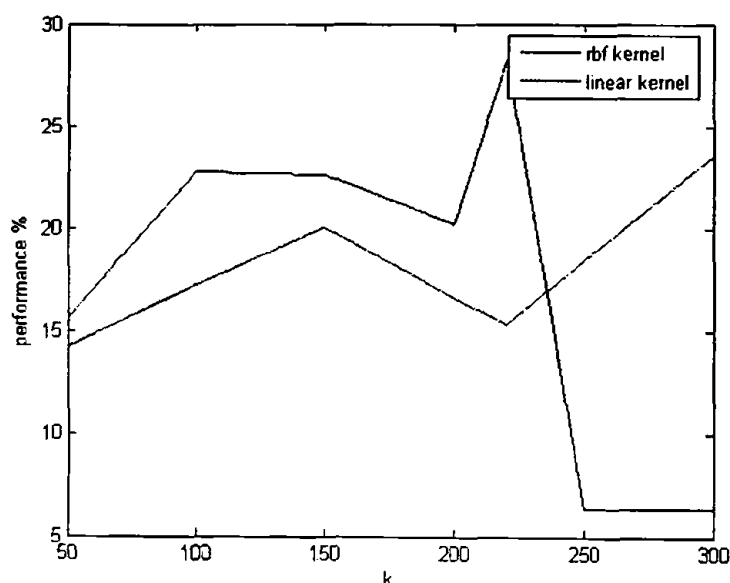
Από την άλλη όμως πλευρά, η χρήση επιπλέον παραδειγμάτων έχει το μειονέκτημα ότι απαιτεί αφενός μεν υπολογιστικό σύστημα με μεγαλύτερη μνήμη για την



αποθήκευση των patches, αφετέρου δε γρηγορότερο επεξεργαστή για την κατασκευή του λεξικού. Στη συνέχεια των πειραμάτων υιοθετήθηκε η επιλογή δεκαπέντε εικόνων ανά κατηγορία, προκειμένου να κρατηθεί ισορροπία ανάμεσα στους παράγοντες υπολογιστική ισχύ και επίδοση.

Επίδραση του πυρήνα στο ποσοστό επιτυχούς ταξινόμησης.

Γιά να εκτιμήσουμε την σημασία του πυρήνα στο τελικό αποτέλεσμα πραγματοποιήθηκαν τα εξής πειράματα. Αρχικά συγκρίναμε τον πυρήνα rbf και τον γραμμικό πυρήνα για τις δέκα κατηγορίες που επιλέξαμε και παραπάνω. Τα αποτελέσματα που προέκυψαν απεικονίζονται στην παρακάτω γραφική παράσταση.

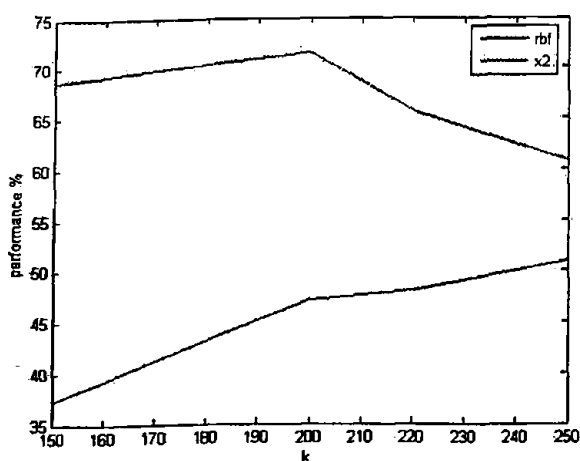


Σχήμα 6.4 Γραφική Παράσταση για Σύγκριση του Πυρήνα Rbf και του Γραμμικού Πυρήνα

Από την παραπάνω γραφική παράσταση προκύπτει ότι ο πυρήνας rbf έχει καλύτερη επίδοση σε σύγκριση με γραμμικό πυρήνα στα περισσότερα από τα πειράματα που πραγματοποιήθηκαν.

Στη συνέχεια πραγματοποιήθηκαν πειράματα για να συγκριθεί ο πυρήνας rbf με έναν άλλο πυρήνα που χρησιμοποιείται ευρέως σε προβλήματα της τεχνητής όρασης, τον χ^2 . Τα αποτελέσματα που προέκυψαν απεικονίζονται στην ακόλουθη γραφική παράσταση.



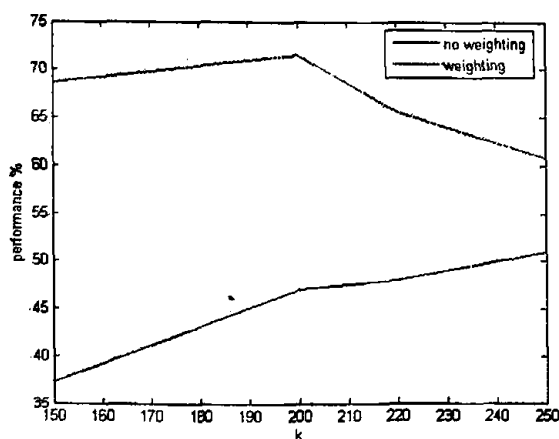


Σχήμα 6.5 Γραφική Παράσταση για Σύγκριση του Πυρήνα Rbf και του χ^2 Πυρήνα

Θύλακας από σημεία-κλειδιά με βάρη

Προκειμένου να εκτιμηθεί η απλή μέθοδος ταξινόμησης με θύλακες από σημεία-κλειδιά σε σχέση με την μέθοδο ταξινόμησης με θύλακες από σημεία κλειδιά χρησιμοποιώντας βάρη, που περιγράφηκε σε προηγούμενο κεφάλαιο, εκτελέστηκαν πειράματα.

Τα πειράματα εκτελέστηκαν σε τρεις κατηγορίες και ο πυρήνας που χρησιμοποιήθηκε ήταν ο rbf με τις βέλτιστες παραμέτρους κάθε φορά. Στην ακόλουθη γραφική παράσταση απεικονίζονται τα αποτελέσματα που προέκυψαν με βάρη και χωρίς βάρη.



Σχήμα 6.6 Γραφική Παράσταση- Θύλακες από Σημεία-Κλειδιά με Βάρη

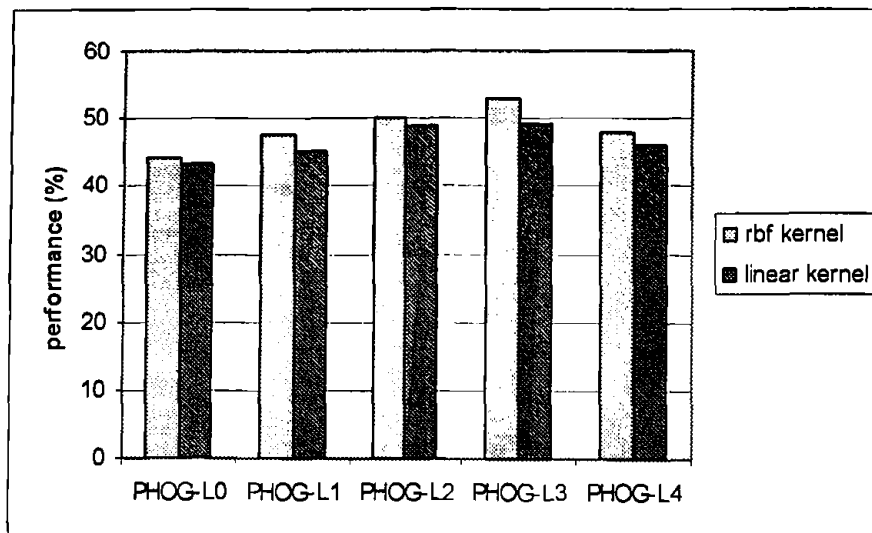
Από το γράφημα προκύπτει ότι η χρήση βαρών πετυχαίνει καλύτερα αποτελέσματα σε σχέση με την απλή μέθοδο. Επομένως η συχνότητα εμφάνισης μιας λέξης είναι κάτι που ενισχύει τη μέθοδο ταξινόμησης. Αυτό είναι λογικό, δεδομένου ότι αν μια λέξη επαναλαμβάνεται αυτό σημαίνει ότι είναι καθοριστικής σημασίας για την κατηγορία. Για παράδειγμα, το μάτι είναι ένα στοιχείο το οποίο επαναλαμβάνεται σε όλες τις εικόνες ανθρώπων.

6.4 Πειραματική αξιολόγηση της μεθόδου με χρήση σχήματος (PHOG)

Στο δεύτερο σύνολο πειραμάτων πραγματοποιήθηκαν πειράματα με σκοπό να εκτιμηθεί τι επίδοση έχει η χρήση του σχήματος στο πρόβλημα της ταξινόμησης εικόνας. Για το σκοπό αυτό εξάγαμε από τρεις κατηγορίες εικόνων (όπως παραπάνω) πληροφορίες για το σχήμα τους χρησιμοποιώντας τη μέθοδο PHOG. Σε όλα τα πειράματα που πραγματοποιήθηκαν χρησιμοποιήθηκε ένα ιστόγραμμα εμφάνισης προσανατολισμού ακμών με 36 κάδους, έτσι ώστε κάθε κάδος να έχει εύρος 10 μοίρες.

Για λόγους υπολογιστικής ισχύος εκτιμήσαμε την επίδοση του SVM ταξινομητή για τα επίπεδα $L=0,1,2,3$, με σκοπό να εκτιμήσουμε ποιο από τα ακόλουθα χαρακτηριστικά PHOG-L0, PHOG-L1, PHOG-L2, PHOG-L3 δίνει καλύτερα αποτελέσματα. Πραγματοποιήθηκαν πειράματα τόσο με πυρήνα rbf όσο και με γραμμικό πυρήνα, τα οποία και απεικονίζονται στην ακόλουθη γραφική παράσταση.





Σχήμα 6.7 Γραφική Παράσταση-Σύγκριση του PHOG για $l=0,1,2,3,4$ και για Γραμμικό Πυρήνα και Πυρήνα rbf

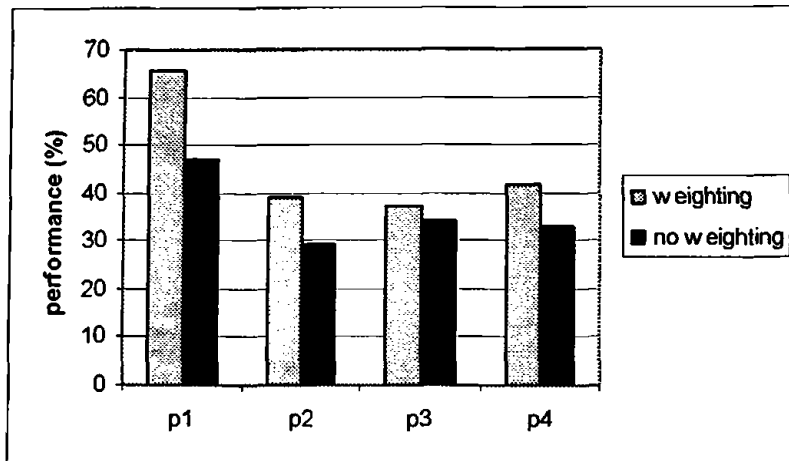
Από το παραπάνω γράφημα προκύπτει ότι η επίδοση ταξινόμησης είναι καλύτερη στην περίπτωση του PHOG-L3 και στην περίπτωση του γραμμικού και στην περίπτωση του rbf πυρήνα. Έγιναν δοκιμές και για την περίπτωση του $L=4$, ωστόσο αυτό αποδείχθηκε αφενός μεν χρονοβόρο, αφετέρου δε δεν επιτεύχθηκαν καλύτερα αποτελέσματα σε σχέση με το $L=3$. Το συμπέρασμα αυτό μας ώθησε στο να χρησιμοποιήσουμε τον αλγόριθμο PHOG με $L=3$ για τα υπόλοιπα πειράματα της παρούσας εργασίας.

Χρήση βαρών στο PHOG

Ακολουθώντας την ίδια πολιτική με την περίπτωση της εμφάνισης, εφαρμόσαμε την εισαγωγή βαρών στην περίπτωση του σχήματος με σκοπό να εκτιμηθεί αν αυτό έχει καλύτερα αποτελέσματα όπως και στην περίπτωση της εμφάνισης με τον αλγόριθμο SIFT. Τα αποτελέσματα που προέκυψαν από την εκτέλεση των πειραμάτων απεικονίζονται στην ακόλουθη γραφική παράσταση.

Εκτελέσαμε πειράματα σε τέσσερα διαφορετικά προβλήματα ταξινόμησης με τρεις κατηγορίες.





Σχήμα 6.8 Γραφική Παράσταση- Χρήση Βαρών στο PHOG

Η γραφική παράσταση παραπάνω απεικονίζει τα αποτελέσματα που προέκυψαν σε αυτά τα προβλήματα. Γίνεται είναι φανερό ότι στην περίπτωση που χρησιμοποιήθηκε βάρος, τα αποτελέσματα ήταν σαφώς καλύτερα και μάλιστα σε όλες τις περιπτώσεις.

6.5 Πειραματική αξιολόγηση του συνδυασμού εμφάνισης και σχήματος

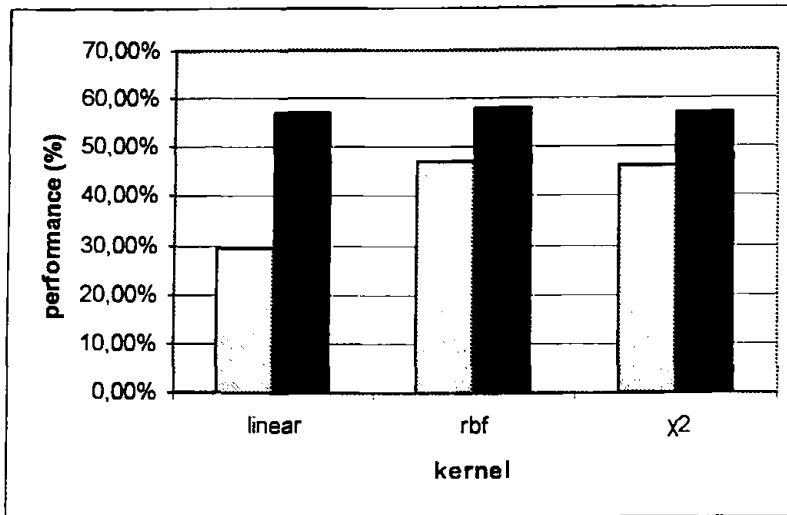
Με σκοπό να εκτιμήσουμε αν ο συνδυασμός εμφάνισης και σχήματος είναι πιο αποδοτικός σε σχέση με μόνο την εμφάνιση ή το σχήμα, εκτελέστηκαν πειράματα.

Τα πειράματα εκτελέστηκαν σε τρεις κατηγορίες, από τις οποίες επιλέχθηκαν 15 εικόνες/κατηγορία ως σύνολο εκπαίδευσης και οι υπόλοιπες εικόνες για έλεγχο. Στα πειράματά μας δημιουργήσαμε ένα πυρήνα ο οποίος προέκυψε από το άθροισμα των επιμέρους πυρήνων, δίνοντας με αυτόν τον τρόπο ίδια βαρύτητα στο σχήμα και στην εμφάνιση. Οι επιμέρους πυρήνες ήταν rbf πυρήνες και ο τελικός πυρήνας προέκυψε από το άθροισμα αυτών:

$$K_{train} = K_{sift} + K_{phog}$$

Τα αποτελέσματα που προέκυψαν απεικονίζονται στην ακόλουθη γραφική παράσταση.





Σχήμα 6.9 Γραφική Παράσταση- Σύγκριση Χρήσης Σχήματος με τον Συνδυασμό Εμφάνισης και Σχήματος

Από τα αποτελέσματα προκύπτει ότι ο συνδυασμός σχήματος και εμφάνισης (μπλε ράβδοι) είχε καλύτερη απόδοση σε σχέση με την απλή χρήση σχήματος (γκρι ράβδοι).

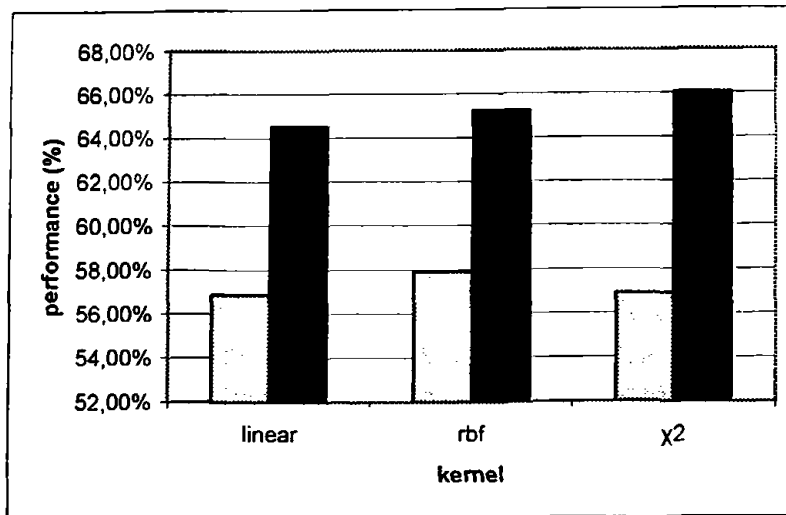
Επίσης, στα πλαίσια της παρούσας εργασίας, συνδυάστηκε σχήμα και εμφάνιση με χρήση του αλγόριθμου SimpleMKL. Ο τελικός πυρήνας που χρησιμοποιήθηκε για την κατασκευή του μοντέλου ταξινόμησης ήταν ο ακόλουθος:

$$K_{train} = d_1 K_{sift} + d_2 K_{phog}, \text{ όπου}$$

$$d_1 + d_2 = 1$$

$$d_1, d_2 \geq 0$$

Ο αλγόριθμος SimpleMKL προσδιόρισε τις βέλτιστες τιμές των παραμέτρων d_1, d_2 . Τα πειραματικά αποτελέσματα που προέκυψαν απεικονίζονται στην ακόλουθη γραφική παράσταση:

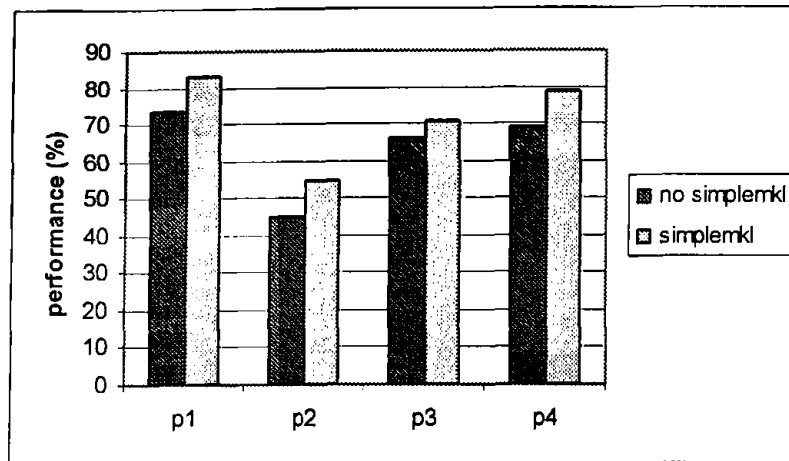


Σχήμα 6.10 Γραφική Παράσταση- Σύγκριση Απλού και Σταθμισμένου Αθροίσματος Πυρήνων

Από τη γραφική παράσταση προκύπτει ότι η χρήση σταθμισμένου αθροίσματος πυρήνων (μπλε ράβδοι) σε κάθε περίπτωση πυρήνα (γραμμικού, Rbf, χ^2) είχε καλύτερη επίδοση σε σχέση με το απλό άθροισμα πυρήνων (γκρι ράβδοι).

6.6 Πειραματική αξιολόγηση του συνδυασμού συνιστωσών με τον αλγόριθμο SimpleMKL

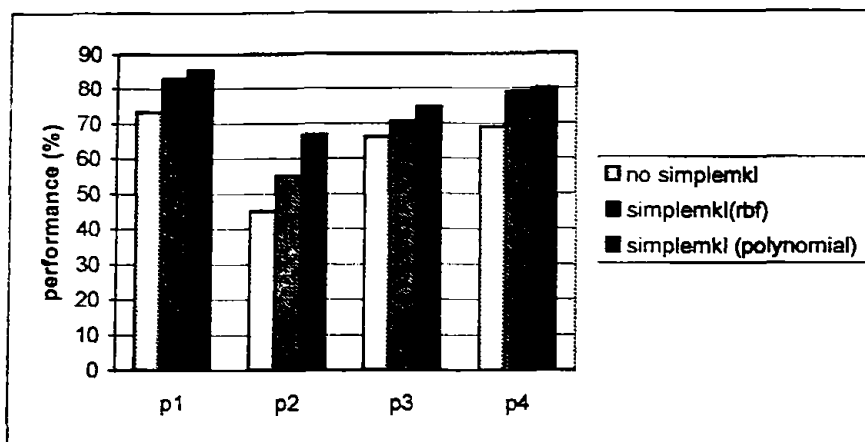
Σε ένα άλλο σύνολο πειραμάτων χρησιμοποιήθηκε ο αλγόριθμος SimpleMKL προκειμένου να εκτιμήσουμε την αποτελεσματικότητα της μεθόδου. Σε αυτήν την περίπτωση χρησιμοποιήθηκε η μέθοδος PHOG-L3 για την εξαγωγή πληροφοριών για το σχήμα. Αυτό είχε ως αποτέλεσμα να προκύψει ένα διάνυσμα διάστασης 3060 για κάθε εικόνα. Δεδομένου ότι θεωρούμε ένα πυρήνα για κάθε συνιστώσα, η μέθοδος SimpleMKL συνδυάζει 3060 πυρήνες. Αρχικά θεωρήσαμε ότι είχαμε στη διάθεσή μας τέσσερα προβλήματα των δύο κατηγοριών και εκτελέσαμε πειράματα για να εκτιμήσουμε τη συμπεριφορά του δυαδικού ταξινομητή. Σε όλες τις περιπτώσεις χρησιμοποιήθηκε ο πυρήνας rbf.



Σχήμα 6.11 Γραφική Παράσταση-Σύγκριση του Συνδυασμού Πυρήνων με SimpleMKL σε Σχέση με τη Χρήση Απλού Πυρήνα για Τέσσερα Δυαδικά Προβλήματα.

Ο αλγόριθμος SimpleMKL με συνδυασμό χαρακτηριστικών όπως φαίνεται και στη γραφική παράσταση δίνει καλύτερα αποτελέσματα σε σύγκριση με την απλή χρήση σχήματος ως πληροφορία για την ταξινόμηση εικόνας. Δεδομένου ότι κάποιες συνιστώσες είναι πιο σημαντικές από άλλες, δηλαδή κάποιοι προσανατολισμοί ακμών συναντώνται πιο συχνά σε σχέση με άλλους, με τον αλγόριθμο που προτείνεται αυτές οι συνιστώσες αποκτούν μεγαλύτερο βάρος σε σχέση με τις άλλες που είναι πιο σπάνιες.

Δεδομένου ότι τα πειράματα που συναντήσαμε στη βιβλιογραφία πάνω στον αλγόριθμο SimpleMKL χρησιμοποιούν συχνά πολωνυμικό πυρήνα, αυτό μας ώθησε στο να εκτελέσουμε πειράματα στον αλγόριθμό μας με τη χρήση πολωνυμικού πυρήνα. Έτσι στα παραπάνω τέσσερα δυαδικά προβλήματα ταξινόμησης συγκρίναμε τον πολωνυμικό πυρήνα με τον πυρήνα *rbf*. Τα αποτελέσματα απεικονίζονται στην ακόλουθη γραφική παράσταση.

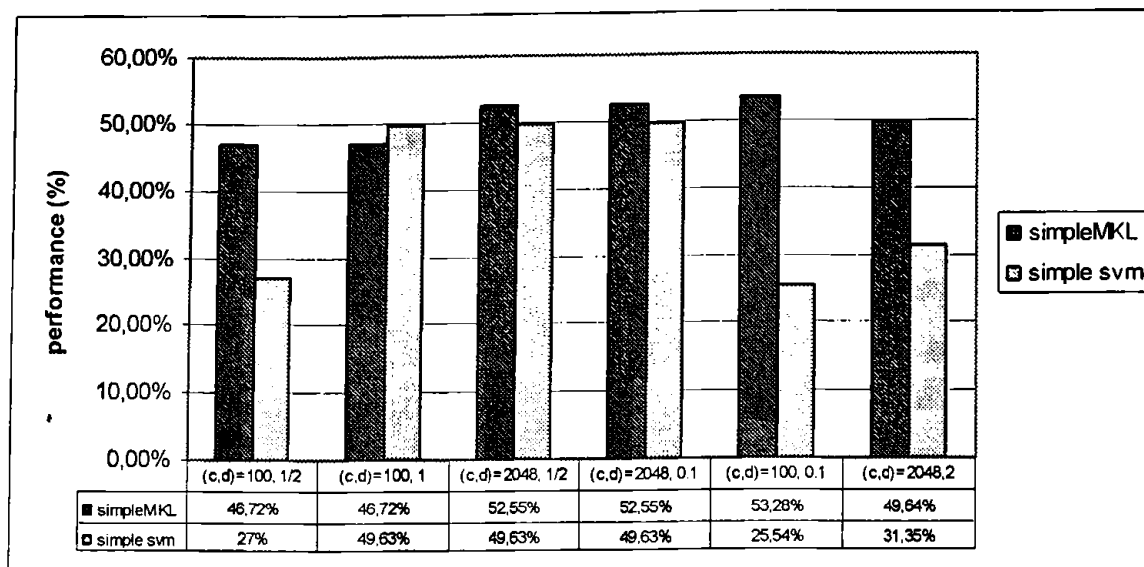


Σχήμα 6.12 Γραφική Παράσταση- Σύγκριση της Χρήσης Πολυωνυμικού και rbf Πυρήνα στην Περίπτωση του Συνδυασμού Πυρήνων με SimpleMKL σε Σχέση με τη Χρήση Απλού Πυρήνα.

Από την γραφική παράσταση παραπάνω προκύπτει ότι σε όλες τις περιπτώσεις ο πολυωνυμικός πυρήνας είχε καλύτερη επίδοση σε σχέση με τον πυρήνα rbf. Για αυτό το λόγο για το υπόλοιπο των πειραμάτων μας χρησιμοποιήθηκε ο πολυωνυμικός πυρήνας.

Στη συνέχεια αξιολογήθηκε η χρήση συνδυασμού συνιστωσών στο πρόβλημα των περισσότερων κατηγοριών και για την ακρίβεια σε 5 κατηγορίες. Επιλέχθηκαν 30 εικόνες από κάθε κατηγορία για το σύνολο εκπαίδευσης και οι υπόλοιπες χρησιμοποιήθηκαν ως σύνολο ελέγχου. Για την εξαγωγή του σχήματος χρησιμοποιήθηκε η μέθοδος PHOG-L3, και για τον ταξινομητή SVM χρησιμοποιήθηκε η μέθοδος «ένας-εναντίον-ένα».

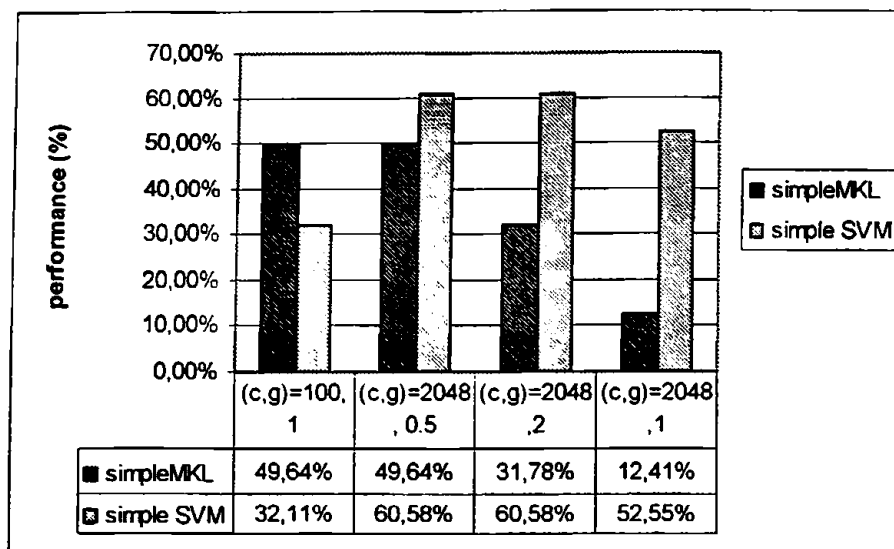
Στο πρώτο σύνολο πειραμάτων εκτελέστηκαν πειράματα στα οποία συνδυάστηκαν πολυωνυμικοί πυρήνες, ένας για κάθε συνιστώσα, και με τον αλγόριθμο SimpleMKL προσδιορίστηκαν οι συντελεστές στο γραμμικό συνδυασμό των πυρήνων. Τα πειραματικά αποτελέσματα απεικονίζονται στην ακόλουθη γραφική παράσταση, η οποία συγκρίνει τη επίδοση του συνδυασμού συνιστωσών σε σχέση με την χρήση απλού ταξινομητή για ταξινόμηση.



Σχήμα 6.13 Γραφική Παράσταση-Σύγκριση της Επίδοσης του Συνδυασμού Χαρακτηριστικών με τη Χρήση του SVM με Απλό Πυρήνα. (Πολυωνυμικός Πυρήνας)

Στον οριζόντιο άξονα δίνονται οι τιμές των παραμέτρων για τους πυρήνες που συνδυάστηκαν (ο βαθμός του πολυωνύμου-d), καθώς και η τιμή της παραμέτρου C, ενώ στον κατακόρυφο άξονα η επίδοση της ταξινόμησης. Από την παραπάνω γραφική παράσταση προκύπτει ότι στην πλειοψηφία των περιπτώσεων ο συνδυασμός χαρακτηριστικών έχει καλύτερη επίδοση σε σχέση με τη χρήση απλού πυρήνα.

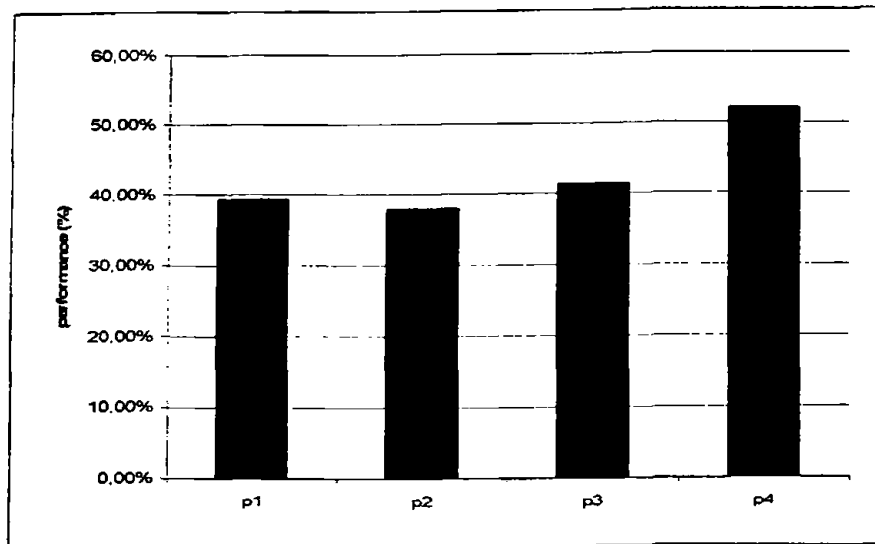
Στο δεύτερο σύνολο πειραμάτων, με την ίδια πειραματική υποδομή όπως παραπάνω, εκτελέστηκαν πειράματα στα οποία συνδυάστηκαν πυρήνες rbf. Οι παράμετροι των πυρήνων αυτών απεικονίζονται στον οριζόντιο άξονα της ακόλουθης γραφικής παράστασης, ενώ η επίδοση στον κατακόρυφο άξονα της γραφικής παράστασης.



Σχήμα 6.14 Γραφική Παράσταση- Σύγκριση της Επίδοσης του Συνδυασμού Χαρακτηριστικών με τη Χρήση του SVM με Απλό Πυρήνα. (Πυρήνας rbf)

Μια επιπλέον επέκταση που πραγματοποιήθηκε σχετικά με τον συνδυασμό συνιστωσών με SimpleMKL ήταν ο προσδιορισμός των βέλτιστων παραμέτρων των πυρήνων με χρήση αναζήτησης πλέγματος σε διάστημα που καθορίσαμε εμείς. Για το σκοπό αυτό δημιουργήθηκε κώδικας σε MATLAB για αναζήτηση του βέλτιστου ζεύγους παραμέτρων (C, d) , όπου C είναι η παράμετρος του SVM ταξινομητή και $d \geq 0$ είναι ο βαθμός των πολωνυμικών πυρήνων που συνδυάζονται. Το διάστημα στο οποίο έγινε η αναζήτηση πλέγματος, με βήμα 2, είναι το $[-5, 15]$ για το C και το $[3, 15]$ για το d και σε κάθε συνδυασμό (C, d) χρησιμοποιήθηκαν δυνάμεις του 2.

Τα πειράματα πραγματοποιήθηκαν σε 4 προβλήματα τριών κατηγοριών και τα αποτελέσματα τα οποία προέκυψαν απεικονίζονται στην ακόλουθη γραφική παράσταση.



Σχήμα 6.15 Γραφική Παράσταση- Επίδοση της Μεθόδου Συνδυασμού Συνιστωσών με Τιμές που Έχουν Προσδιοριστεί Από τη Δική μας Αναζήτηση Πλέγματος

6.7 Συμπεράσματα

Από την εκτέλεση των πειραμάτων έχουν προκύψει ορισμένα χρήσιμα συμπεράσματα. Αρχικά, στη μέθοδο της ταξινόμησης με θύλακες από σημεία-κλειδιά παίζουν καθοριστικό ρόλο τόσο το μέγεθος του συνόλου δεδομένων που έχουμε στη διάθεσή μας όσο και οι τιμές των παραμέτρων της μεθόδου, όπως ο αριθμός των ομάδων, ο πυρήνας και η παράμετρος C του ταξινομητή SVM. Βελτίωση στην απλή μέθοδο με θύλακες αποτελεί η χρήση βαρών.

Στα πειράματα που πραγματοποιήθηκαν με χρήση σχήματος, τα καλύτερα αποτελέσματα προέκυψαν στην περίπτωση του PHOG-L3. Η χρήση βαρών και σε αυτήν την περίπτωση αποδείχθηκε ωφέλιμη και σε αυτήν την περίπτωση.

Ο απλός συνδυασμός εμφάνισης και σχήματος είχε καλύτερα αποτελέσματα στα πειράματά μας, σε σχέση με τη χρήση μόνο σχήματος.

Αντίθετα, ο συνδυασμός εμφάνισης και σχήματος με χρήση του αλγόριθμου SimpleMKL είχε καλύτερα αποτελέσματα. Αυτό είναι αναμενόμενο, δεδομένου ότι ο αλγόριθμος SimpleMKL δεν δίνει ίδια βαρύτητα στο σχήμα και στην εμφάνιση, αλλά

διαφορετικές ανάλογα με το αν το σχήμα ή η εμφάνιση έχει μεγαλύτερη σημασία για το διαχωρισμό των εικόνων που ταξινομούμε.

Ο συνδυασμός συνιστωσών, που πρωτοεφαρμόστηκε στην παρούσα εργασία, επίσης αποδείχθηκε ενθαρρυντικός και μάλιστα στην περίπτωση του πολυωνυμικού πυρήνα, σε σχέση με την χρήση ενός ταξινομητή SVM απλού πυρήνα. Αυτό είναι λογικό, εφόσον κάποιες συνιστώσες οι οποίες δεν συμβάλλουν στον διαχωρισμό των δεδομένων εξαλείφονται από τον αλγόριθμο, γιατί τους δίνει μικρή βαρύτητα. Αντίθετα, οι σημαντικές συνιστώσες ενισχύονται από τον αλγόριθμο, που τους δίνει μεγάλη βαρύτητα.



ΑΝΑΦΟΡΕΣ

- [1] Osuna, E., Freund, R. And Girosi. **Training support vector machines: An application to face detection.** *CVPR (Computer Vision and Pattern recognition)*, 1997
- [2] Ronfard, R., Schmid, C., Triggs, B. **Learning to parse pictures of people,** *ECCV*, 2002.
- [3] Csurka, G., Bray, C., Dance, C., and Fan, L. **Visual categorization with bags of keypoints,** *In ECCV workshop on Statistical Learning in Computer Vision*, pp 59-74.
- [4] A.Bosch, A.Zisserman, X.Muñoz. **Representing Shape with a Spatial Pyramid Kernel.** *International Conference on Image and Video Retrieval.* Amsterdam, The Netherlands. July 2007
- [5] S. Lazebnik, C. Schmid, and J. Ponce. **Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories.** *In Proc. CVPR*, 2006.
- [6] Lowe, D. **Distinctive Image Features from Scale-Invariant Keypoints.** *International Journal of Computer Vision*, 60, 2 , pp 91-110 (2004).
- [7] H. and Bay. **Surf: Speeded up robust features.** *In European Conference on Computer Vision*, volume 1, pages 404–417, Graz, Austria, 2006.
- [8] Mikolajczyk, K and Schmid, C. 2005. **A performance evaluation of local descriptors.** *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27(10): 1615-1630
- [9] D. Haussler. **Convolution kernels on discrete structures.** Technical Report UCSCRL-99-10, 1999.
- [10] J. Canny. **A computational approach to edge detection.** *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8:679–698, 1986.
- [11] N. Dalal and B. Triggs. **Histograms of oriented gradients for human detection.** *In IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 886–893, San Diego, California, June 2005.



- [12] J. Zhang and M. Marszalek, S. Lazebnik, C. Schmid, **Local Features and Kernels for Classification of Texture and Object Categories: A Comprehensive Study**, *International Journal of Computer Vision* 73(2), 213–238, 2007
- [13] N. Dalal and B Triggs. **Histogram of oriented gradients for human detection**. *In Proc. CVPR*, 2005.
- [14] G. R. G. Lanckriet, N. Cristianini, P. Bartlett, L. E. Ghaoui, and M. I. Jordan. **Learning the kernel matrix with semi definite programming**. *JMLR*, 5:27–72, 2004.
- [15] A. Rakotomamonjy, F. Bach, S. Canu, Y. Grandvalet, **SimpleMKL** *Journal of Machine Learning Research*, 9, 2491-2521, 2008
- [16] J.F Bonnans. A. Shapiro. **Optimization problems with perturbation : A guided tour**. *SIAM Review*, 40 (2): 202-227, 1998
- [17] J.F Bonnans, J.C Gilbert, C. Lemarechal, C.A Sagastizbal. **Numerical Optimization Theoretical and Practical Aspects**. Springer 2003.
- [18] F. Bach, G. Lanckriet, M. Jordan. **Multiple Kernel Learning, conic duality and the SMO algorithm**. In Proceedings of the 21st International Conference on Machine Learning, pages 41-48, 2004
- [19] D. Luenberg. **Linear and non linear programming**. Addison-Wesley, 1984.
- [20] Gehler P. Nowozin S. **On Feature Combination for multiclass Object Classification** Proceedings of the Twelfth IEEE International Conference on Computer Vision (ICCV 2009). 1-8, IEEE Computer Society, Los Alamitos, CA, USA (10 2009)
- [21] E. Nowak et al. **Sampling strategies for bag-of-features image classification**. *In ECCV*, 2006.
- [22] C. Chang and C. Lin. **LIBSVM: a library for support vector machines**, 2001.
- [23] Vapnik, V, Cortes, C. **Support Vector Networks**, Machine Learning, vol. 20, no 3, pp 273-297, 1995.
- [24] E. Nowak et al. **Sampling strategies for bag-of-features image classification**. *ECCV*, 2006.
- [25] S. Lazebnik, C. Schmid, and J. Ponce. **A sparse texture representation using affine invariant regions**. IEEE Computer Society Conference on Computer



Vision and Pattern Recognition, volume 2, pages 319–324, Madison, Wisconsin, June 2003.

[26] C. Harris and M. Stephens. **A combined corner and edge detector**. In Alvey Vision Conference, pages 147–151, Manchester, UK, 1988.

[27] N. Aronszajn. **Theory of reproducing kernels**. Trans. Am. Math. Soc., (68):337–404, 1950.

[28] S. Sonnenburg, G. Ratsch, and C. Schafer. **A general and efficient algorithm for multiple kernel learning**. In Advances in Neural Information Processing Systems, volume 17, pages 1–8, 2005.

[29] Chang, C.-C and C.-J. Lin (2001), **LIBSVM : a library for Support Vector Machines**.

[30] Hsu, C.-W, Chang, C.-C , Lin, C.-J. **A Practical Guide to Support Vector Classification**.

[31] L. Fei-Fei, R. Fergus and P. Perona. **Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories**. IEEE. CVPR 2004, Workshop on Generative-Model Based Vision. 2004

[32] H.G. Barrow, J.M. Tanenbaum, R.c. Bolles, and H.C. Wolf. **Parametric correspondence and chamfer matching: Two new techniques for image matching**. In International Joint Conference on Artificial Intelligence, pages 659–663, Cambridge, 1977.



ΣΥΝΤΟΜΟ ΒΙΟΓΡΑΦΙΚΟ

Η Ειρήνη Μπράχου γεννήθηκε στα Ιωάννινα το 1985. Το 2003 εισήχθη στο Τμήμα Πληροφορικής του Πανεπιστημίου Ιωαννίνων και ολοκλήρωσε τις προπτυχιακές της σπουδές το 2007. Παρακολούθησε το Πρόγραμμα Μεταπτυχιακών Σπουδών στο ίδιο τμήμα, από το οποίο αποφοίτησε τον Ιούνιο 2010.

