

Ταξινόμηση Εικόνων με Τυχαία Δάση

Αγγελική Τσιάρα

ΜΕΤΑΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ ΕΞΕΙΔΙΚΕΥΣΗΣ

— ♦ —

Ιωάννινα, Ιανουάριος 2012



ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΙΩΑΝΝΙΝΩΝ

DEPARTMENT OF COMPUTER SCIENCE
UNIVERSITY OF IOANNINA



ΒΙΒΛΙΟΘΗΚΗ
ΠΑΝΕΠΙΣΤΗΜΙΟΥ ΙΩΑΝΝΙΝΩΝ



026000305392



ΤΑΞΙΝΟΜΗΣΗ ΕΙΚΟΝΩΝ ΜΕ ΤΥΧΑΙΑ ΔΑΣΗ

Η
ΜΕΤΑΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ ΕΞΕΙΔΙΚΕΥΣΗΣ

Υποβάλλεται στην

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

από την

Αγγελική Τσιάρα

ως μέρος των Υποχρεώσεων

για τη λήψη

του

ΜΕΤΑΠΤΥΧΙΑΚΟΥ ΔΙΠΛΩΜΑΤΟΣ ΣΤΗΝ ΠΛΗΡΟΦΟΡΙΚΗ
ΜΕ ΕΞΕΙΔΙΚΕΥΣΗ ΣΤΙΣ ΤΕΧΝΟΛΟΓΙΕΣ-ΕΦΑΡΜΟΓΕΣ

Ιανουάριος 2012



ΑΦΙΕΡΩΣΗ

Στα νεογέννητα ανιψάκια μου.



ΕΥΧΑΡΙΣΤΙΕΣ

Θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή μου κ. Κωνσταντίνο Μπλέκα, Επίκουρο Καθηγητή του Τμήματος Πληροφορικής του Πανεπιστημίου Ιωαννίνων για την καθοδήγησή του κατά την εκπόνηση της παρούσας εργασίας. Ένα μεγάλο ευχαριστώ στην οικογένεια μου για την κατανόηση και τη στήριξή τους.



ΠΕΡΙΕΧΟΜΕΝΑ

ΑΦΙΕΡΩΣΗ	Σελ ii
ΕΥΧΑΡΙΣΤΙΕΣ	iii
ΠΕΡΙΕΧΟΜΕΝΑ	iv
ΕΥΡΕΤΗΡΙΟ ΠΙΝΑΚΩΝ	vi
ΕΥΡΕΤΗΡΙΟ ΣΧΗΜΑΤΩΝ	vii
ΠΕΡΙΛΗΨΗ	ix
EXTENDED ABSTRACT IN ENGLISH	x
ΚΕΦΑΛΑΙΟ 1. ΕΙΣΑΓΩΓΗ	1
1.1. Το Πρόβλημα της Ταξινόμησης Εικόνων	1
1.2. Συναφή Προβλήματα	3
1.3. Εφαρμογές του Προβλήματος	5
1.4. Στόχοι της Εργασίας	7
1.5. Δομή της Εργασίας	8
ΚΕΦΑΛΑΙΟ 2. ΣΥΝΔΥΑΣΤΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ	10
2.1. Τεχνικές Μάθησης	10
2.1.1 Μάθηση χωρίς Επίβλεψη (Unsupervised Learning)	11
2.1.2 Επιβλεπόμενη Μάθηση (Supervised Learning)	13
2.2. Μέθοδοι Ταξινόμησης	15
2.2.1 Ταξινομητές Κοντινότερων Γειτόνων (Nearest Neighbors Classifier)	16
2.2.2 Δέντρα Απόφασης και Κανόνες Απόφασης (Decision Trees and Rules)	16
2.2.3 Μπεϋζιανοί Ταξινομητές (Bayesian Classifiers)	18
2.2.4 Διακρίνουσες Συναρτήσεις (Discriminant Functions)	19
2.2.5 Μηχανές Διανυσμάτων Υποστήριξης (Support Vector Machines-SVM)	19
2.2.6 Τεχνητά Νευρωνικά Δίκτυα (Artificial Neural Networks-ANN)	23
2.3. Συνδυαστικές Τεχνικές Μεθόδων Μηχανικής Μάθησης	23
2.3.1 Τρόποι Συνδυασμού των Προβλέψεων Διαφορετικών Υποθέσεων	24
2.4. Κατασκευή Συνδυαστικών Ταξινομητών (Ensemble Classifiers)	25
2.4.1 Η μέθοδος Bagging	26
2.4.2 Η μέθοδος Boosting	26
2.5. Τυχαία Δάση (Random Forests)	29
2.5.1 Σύγκλιση του Σφάλματος Γενίκευσης	30
2.5.2 Πλεονεκτήματα των Τυχαίων Δασών	31
ΚΕΦΑΛΑΙΟ 3. ΤΥΧΑΙΑ ΔΑΣΗ ΓΙΑ ΤΑΞΙΝΟΜΗΣΗ ΕΙΚΟΝΩΝ	33
3.1. Μέθοδοι Ταξινόμησης Εικόνων με Βάση το Περιεχόμενο	33
3.1.1 Γενικά	33
3.1.2 Η Μέθοδος Bag of Keypoints	40
3.2. Η Προτεινόμενη Μέθοδος	41
3.2.1 Εκπαίδευση του Μοντέλου Ταξινόμησης	43



3.2.1.1 Δειγματοληψία Σημείων της Εικόνας και Εξαγωγή Διανυσμάτων Χαρακτηριστικών	44
3.2.1.2 Κατασκευή Στοχαστικών Δασών Δυαδικών Δέντρων Απόφασης	46
3.2.1.3 Εξαγωγή Γνώσης από τους Τερματικούς Κόμβους των Δέντρων	48
3.2.2 Κατασκευή του Μηχανισμού Απόφασης	49
3.2.2.1 Δειγματοληψία Σημείων από την Εικόνα Ελέγχου και Εξαγωγή ενός Διανύσματος Χαρακτηριστικών	50
3.2.2.2 Η Διαδικασία της "Ψηφοφορίας"	50
3.2.2.3 Ταξινόμηση Πολλών Κατηγοριών (Multi-class Problem)	57
ΚΕΦΑΛΑΙΟ 4. ΠΕΙΡΑΜΑΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ	60
4.1. Γενικά	60
4.2. Πειραματική Εκτίμηση της Απόδοσης της Προτεινόμενης Μεθόδου	60
4.2.1 Πειραματική Διαδικασία	60
4.2.2 Εκτίμηση της Απόδοσης της Μεθόδου	62
4.2.3 Επίδραση του Πλήθους των Δυαδικών Δέντρων Απόφασης σε κάθε Δάσος	64
4.2.4 Σύγκριση της Απόδοσης της Προτεινόμενης Μεθόδου με τη Μέθοδο Bag of Keypoints	65
4.2.5 Σύγκριση της Απόδοσης της Προτεινόμενης Μεθόδου σε Διαφορετικά Σύνολα Δεδομένων	68
4.3. Πειραματικά Δεδομένα	69
4.3.1 Το Σύνολο Δεδομένων TU Darmstadt Database (παλαιότερα ETHZ Database)	69
4.3.2 Το Σύνολο Δεδομένων TUD Pedestrians	71
4.3.3 Εικόνες από τις Βάσεις Pascal VOC 2006 Database και Caltech	71
4.4. Διερεύνηση της Δυνατότητας Χρήσης της Μεθόδου για Ανάκτηση Εικόνων	72
ΚΕΦΑΛΑΙΟ 5. ΣΥΜΠΕΡΑΣΜΑΤΑ - ΜΕΛΛΟΝΤΙΚΗ ΕΡΓΑΣΙΑ	73
ΑΝΑΦΟΡΕΣ	75
ΣΥΝΤΟΜΟ ΒΙΟΓΡΑΦΙΚΟ	79



ΕΥΡΕΤΗΡΙΟ ΠΙΝΑΚΩΝ

Πίνακας	Σελ
Πίνακας 1.1 Χρήση Ενός Ταξινομητή Πολλών Κατηγοριών ($K=3$)	8
Πίνακας 1.2 Χρήση Τριών Δυαδικών Ταξινομητών	8
Πίνακας 3.1 Μορφή του Πίνακα R-Table	51
Πίνακας 4.1 Confusion Matrix και η Ακρίβεια Ταξινόμησης ανά Κατηγορία (για 18 Χαρακτηριστικά)	62
Πίνακας 4.2 Confusion Matrix και η Ακρίβεια Ταξινόμησης ανά Κατηγορία (για 6 Χαρακτηριστικά)	63
Πίνακας 4.3 Περιγραφή των Συνόλων Δεδομένων	68
Πίνακας 4.4 Ακρίβεια Ταξινόμησης της Προτεινόμενης Μεθόδου σε Διαφορετικά Σύνολα Δεδομένων	69



ΕΥΡΕΤΗΡΙΟ ΣΧΗΜΑΤΩΝ

Σχήμα	Σελ
Σχήμα 1.1 Δύο Δομικά Διακριτά Μέλη της Κατηγορίας «Αυτοκίνητο»	4
Σχήμα 1.2 Διαδικασία Ανάκτησης Εικόνων	4
Σχήμα 1.3 Παραδείγματα Εικόνων Πολλών Κατηγοριών	8
Σχήμα 2.1 Η Διαδικασία της Μάθησης Ενός Γνωστικού Συστήματος	10
Σχήμα 2.2 Η Ταξινόμηση ως Διαδικασία Αντιστοίχισης από τον Χώρο των Χαρακτηριστικών στον Χώρο των Κατηγοριών	14
Σχήμα 2.3 Η Διαδικασία της Ταξινόμησης	14
Σχήμα 2.4 Η Δομή Ενός Δυαδικού Δέντρου Απόφασης	17
Σχήμα 2.5 Η Διαδικασία Κατασκευής των Random Forests	29
Σχήμα 3.1 Διαδικασία Ταξινόμησης με Βάση το Περιεχόμενο των Εικόνων	34
Σχήμα 3.2 Ταξινόμηση Εικόνων με Χρήση Θυλάκων από Σημεία-Κλειδιά	40
Σχήμα 3.3 Διαδικασία Εκπαίδευσης του Μοντέλου Ταξινόμησης	43
Σχήμα 3.4 Οριοθέτηση Αντικειμένου μέσω Μασκών Κατάτμησης Εικονοστοιχείων και Πλαισίων Οριοθέτησης	44
Σχήμα 3.5 Δειγματοληψία των Patches. Το Βέλος Αντιστοιχεί στο Διάνυσμα Απόκλισης (Offset Vector) του Object Patch	46
Σχήμα 3.6 Δείγματα των Patches ενός Τερματικού Κόμβου. Τα Μπλε Αντιστοιχούν σε Object Patches για την Κατηγορία «Cars»	47
Σχήμα 3.7 Δείγματα Τερματικών Κόμβων σε ένα Δέντρο	48
Σχήμα 3.8 Η Διαδικασία Ταξινόμησης μιας Εικόνας	49
Σχήμα 3.9 (α) Επιλογή Σημείων Περιγράμματος (β) Εικόνα Τυχαίων Σχημάτων (γ) Τα Σημεία Ακμής του Σχήματος S (στην εικόνα) Αυξάνουν το Ίδιο Σημείο στον Συσσωρευτή ενώ τα Σημεία Ακμής Διαφορετικών Σχημάτων Αυξάνουν Σκόρπια Σημεία	51
Σχήμα 3.10 Τα Διανύσματα Απόκλισης των Object Patches Παρέχουν την Πληροφορία Κατεύθυνσης για τη Διαδικασία της Ψηφοφορίας	52
Σχήμα 3.11 Διάσχιση των Δέντρων Ενός Δάσους από Ένα Test Patch	54
Σχήμα 3.12 Εκχώρηση των Ψήφων σε Όλες τις Πιθανές Θέσεις Εμφάνισης του Αντικειμένου	56
Σχήμα 3.13 Για καθένα από τα 3 Patches που Σημειώθηκαν στην Εικόνα (α) Υπολογίζονται από το Δάσος της Κατηγορίας «Pedestrians» οι Πιθανοτικοί Ψήφοι για τις Πιθανές Θέσεις του Κεντροειδούς του Πεζού (β). Συναθροίζονται οι Ψήφοι Όλων των Patches σε μια Εικόνα Hough (γ) Η Θέση του Πεζού Ανιχνεύεται με Βάση τη Μέγιστη Τιμή της Εικόνας Hough (δ)	57
Σχήμα 3.14 Οι Εικόνες Hough που Υπολογίστηκαν από το Δάσος κάθε Κατηγορίας. Το Πράσινο Αντιστοιχεί στην Εικόνα Hough της Κατηγορίας «Cars»	58
Σχήμα 3.15 Διαδικασία Απόφασης	59
Σχήμα 4.1 Γραφική Παράσταση για την Εκτίμηση της Επίδρασης του Πλήθους	



των Δυαδικών Δέντρων Απόφασης σε κάθε Δάσος στην Ακρίβεια της Ταξινόμησης	64
Σχήμα 4.2 Γραφική Παράσταση για την Σύγκριση της Απόδοσης της Προτεινόμενης Μεθόδου με τη Μέθοδο Bag of Keypoints στο Σύνολο των Εικόνων Εκπαίδευσης για Διαφορετικό Πλήθος Εικόνων Εκπαίδευσης	66
Σχήμα 4.3 Γραφική Παράσταση για την Σύγκριση της Απόδοσης της Προτεινόμενης Μεθόδου με τη Μέθοδο Bag of Keypoints στο Σύνολο των Εικόνων Ελέγχου για Διαφορετικό Πλήθος Εικόνων Εκπαίδευσης	67
Σχήμα 4.4 Το Σύνολο Δεδομένων για την Κατηγορία «Cars»	70
Σχήμα 4.5 Το Σύνολο Δεδομένων για την Κατηγορία «Cows»	70
Σχήμα 4.6 Το Σύνολο Δεδομένων για την Κατηγορία «Motorbikes»	70
Σχήμα 4.7 Το Σύνολο Εικόνων TUD Pedestrians	71
Σχήμα 4.8 Εικόνες από Pascal VOC 2006 και Caltech	72



ΠΕΡΙΛΗΨΗ

Αγγελική Τσιάρα του Κωνσταντίνου και της Ζωίτσας.

MSc, Τμήμα Πληροφορικής, Πανεπιστήμιο Ιωαννίνων, Ιανουάριος, 2012.

Ταξινόμηση Εικόνων με Τυχαία Δάση.

Επιβλέπων: Κωνσταντίνος Μπλέκας.

Η παρούσα μεταπτυχιακή εργασία μελετά το πρόβλημα της ταξινόμησης εικόνων με βάση το περιεχόμενο σε μια από τις προκαθορισμένες κατηγορίες. Το συγκεκριμένο πρόβλημα όσον αφορά τις φυσικές εικόνες παρουσιάζει αρκετές δυσκολίες οι οποίες οφείλονται αφενός μεν στις μεταβαλλόμενες και μερικές φορές μη ελεγχόμενες συνθήκες απεικόνισης αφετέρου δε στο χάσμα ανάμεσα στους πίνακες των αριθμών που αναπαριστούν τις εικόνες και στο σημασιολογικό περιεχόμενο των εικόνων. Στην εργασία δίνεται μια συνοπτική περιγραφή των πιο γνωστών μεθόδων ταξινόμησης και αναλύονται μέθοδοι συνδυαστικών ταξινομητών όπως τα Τυχαία Δάση (Random Forests). Παρουσιάζονται τα βασικά στάδια των μεθόδων που έχουν προταθεί για την επίλυση του προβλήματος και περιγράφεται συνοπτικά μια από τις πιο γνωστές μεθόδους ταξινόμησης εικόνας, η μέθοδος Bag of Keypoints. Επίσης, προτείνεται μια μέθοδος η οποία στηρίζεται στην κατασκευή ενός συνδυαστικού μοντέλου ταξινόμησης με χρήση δυαδικών δέντρων απόφασης. Περιγράφεται αναλυτικά η διαδικασία εκπαίδευσης του μοντέλου καθώς και η κατασκευή ενός στοχαστικού μηχανισμού απόφασης. Η απόδοση της προτεινόμενης μεθόδου εκτιμάται πειραματικά και συγκρίνεται με την απόδοση της μεθόδου Bag of Keypoints.



EXTENDED ABSTRACT IN ENGLISH

Tsiara, Aggeliki, K.

MSc, Computer Science Department, University of Ioannina, Greece. January, 2012.

Image Classification with Random Forests.

Thesis Supervisor: Konstantinos Blekas.

This master thesis studies the problem of content-based image classification. As far as natural images are concerned, this problem has several difficulties due firstly to the changing and sometimes uncontrolled conditions e.g. illumination, occlusion, background clutter etc., and secondly the gap between arrays of numbers that represent images and the semantic content of these images. A technique that has been proposed and used for image classification as well as for image retrieval, is based on describing images with text, however this is not efficient because descriptions are generally vague and subjective. The term “content-based” refers to the fact that classification is based on the actual image content and not on some metadata such as keywords, tags or descriptions associated with the image.

Features such as color, texture and shape are often used to describe image content. The majority of content-based image classification methods that have been proposed include two main stages, vectorial representation of images and classification model construction. In the first stage, feature descriptors are extracted from images. These descriptors are used to form vector representations which are subsequently used to construct classification models. One of the most known image classification methods, namely the method “Bag of Keypoints” is briefly described. This method is based on vector quantization of affine invariant descriptors of image patches and can be implemented using either SVM or Naive Bayes classifiers.



The method proposed in this study is based on the construction of an ensemble classification model using binary decision trees. Although there are many ways to construct an ensemble of base classifiers, we choose to create multiple training sets by resampling the original data according to the uniform sampling distribution. Each training set is used for growing a different decision tree. The procedure of tree induction includes the application of the clustering algorithm K-means for the binary partition of each non-terminal node. Specifically, for each predefined category a forest of binary decision trees is being constructed. Every such forest represents a stochastic mapping e.g. a “voting” procedure based on the idea of the generalized Hough transform proposed by Ballard [33]. This “voting” procedure along with directional information is used to map any part of the image to a probabilistic vote on the existence of an object of interest in the image. In each forest, votes made by all image parts are aggregated to form a decision mechanism called Hough image. A “new” image is assigned to one of the predefined categories by combining the predictions made by all classifiers according to a majority voting scheme.

The performance of the proposed method is estimated experimentally and compared to the performance of the “Bag of Keypoints” method.



ΚΕΦΑΛΑΙΟ 1. ΕΙΣΑΓΩΓΗ

1.1 Το Πρόβλημα της Ταξινόμησης Εικόνων

1.2 Συναφή Προβλήματα

1.3 Εφαρμογές του Προβλήματος

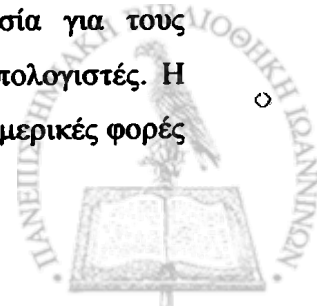
1.4 Στόχοι της Εργασίας

1.5 Δομή της Εργασίας

1.1. Το Πρόβλημα της Ταξινόμησης Εικόνων

Η ανάπτυξη της τεχνολογίας έχει επιτρέψει την κατασκευή και διάθεση μηχανών παραγωγής και επεξεργασίας εικόνας με χαμηλό κόστος ώστε να είναι προσιτές σε μεγάλο μέρος του πληθυσμού. Η αυξανόμενη χρήση τέτοιων συσκευών έχει ως αποτέλεσμα τη διόγκωση της οπτικής πληροφορίας σε προσωπικές συλλογές και κυρίως στο Διαδίκτυο και την ανάγκη δημιουργίας συστημάτων για τη σωστή διαχείριση της. Το συγκεκριμένο πρόβλημα αποτελεί ένα σημαντικό αντικείμενο έρευνας σε πολλά διαφορετικά επιστημονικά πεδία του κλάδου της Πληροφορικής, όπως η επεξεργασία εικόνας, η υπολογιστική όραση, η αναγνώριση προτύπων, η μηχανική μάθηση και η εξόρυξη δεδομένων. Τα τελευταία χρόνια γίνονται εκτεταμένες προσπάθειες για την ανάπτυξη αποτελεσματικών και υπολογιστικά αποδοτικών μεθόδων κατηγοριοποίησης και ανάκτησης εικόνων με βάση το περιεχόμενό τους.

Ο όρος ταξινόμηση (ή κατηγοριοποίηση) εικόνας αναφέρεται στην ταυτοποίηση εικόνων αγνώστου προέλευσης από ένα σύνολο προκαθορισμένων (συνήθως) κατηγοριών. Αν και αυτό δεν αποτελεί μια πολύ δύσκολη εργασία για τους ανθρώπους, έχει αποδειχθεί ότι είναι εξαιρετικά δύσκολο για τους υπολογιστές. Η δυσκολία αυτή οφείλεται ως επί των πλείστον στις μεταβαλλόμενες και μερικές φορές



μη ελεγχόμενες συνθήκες απεικόνισης όπως επίσης και στο χάσμα ανάμεσα στους πίνακες των αριθμών που αναπαριστούν τις φυσικές εικόνες και στο σημασιολογικό περιεχόμενο των εικόνων που αντιλαμβάνονται οι άνθρωποι. Συνεπώς ένα σύστημα κατηγοριοποίησης εικόνων για να είναι αποδοτικό θα πρέπει να είναι εύρωστο, να μπορεί να διαχειριστεί τις μεταβολές των συνθηκών απεικόνισης καθώς αυτές αποτελούν σημαντικές παραμέτρους του προβλήματος και επηρεάζουν σε μεγάλο βαθμό την εμφάνιση των εικόνων. Συνοπτικά, ένα ιδανικό μοντέλο ταξινόμησης πρέπει να «θεραπεύει» αποτελεσματικά τις παρακάτω δυσκολίες οι οποίες προέρχονται από τη φύση του προβλήματος:

(α) Διαφορές που εντοπίζονται στα εσωτερικά χαρακτηριστικά των αντικειμένων της ίδιας κατηγορίας (*intra-class differences*). Αντικείμενα που ανήκουν στην ίδια κατηγορία μπορεί να εμφανίζουν σημαντικές διαφορές στη δομή και την εμφάνισή τους, παρόλα αυτά το μοντέλο ταξινόμησης πρέπει να τα αναθέσει στην ίδια κατηγορία.

(β) Μεταβολές που αφορούν την οπτική γωνία λήψης (*view-point variations*). Η εμφάνιση ενός αντικειμένου σε μια εικόνα μεταβάλλεται όταν αλλάζει η οπτική γωνία παρατήρησης.

(γ) Μεταβολές στις συνθήκες φωτισμού (*illumination*). Η μεταβολή του φωτισμού αλλάζει την εμφάνιση του αντικειμένου.

(δ) Ποικιλομορφία του φόντου (*background clutter*). Σε πραγματικές συνθήκες κανένα αντικείμενο δεν εμφανίζεται σε μια άδεια σκηνή, γι' αυτό το λόγο οι αλγόριθμοι επεξεργασίας πρέπει να διαχωρίζουν ικανοποιητικά το αντικείμενο ενδιαφέροντος από το φόντο (αφαίρεση φόντου).

(ε) Μεταβολές στην κλίμακα (*scale*). Το μέγεθος ή το βάθος του αντικειμένου στην εικόνα δεν αλλάζει το γεγονός ότι πρόκειται για το ίδιο αντικείμενο.

(στ) Απόκρυψη μέρους του αντικειμένου (*occlusion*). Ο αλγόριθμος πρέπει να λαμβάνει υπόψη του ότι μπροστά από το αντικείμενο ενδιαφέροντος μπορεί να παρεμβάλλονται άλλα αντικείμενα. Το μοντέλο ταξινόμησης θα πρέπει να αντιμετωπίζει αποτελεσματικά και τις περιπτώσεις που υπάρχει θόρυβος στην εικόνα.

Με άλλα λόγια, η πολυπλοκότητα του προβλήματος της ταξινόμησης εικόνων έγκειται στην ποικιλομορφία που παρουσιάζει το περιεχόμενο των εικόνων ακόμα και



όταν αυτές απεικονίζουν το ίδιο αντικείμενο. Ο όρος αντικείμενο στο φυσικό κόσμο, στον οποίο ζούμε, αναφέρεται σε κάτι το οποίο είναι ορατό, απτό και έχει σχετικά σταθερή μορφή. Ένα αντικείμενο σε μια εικόνα ορίζεται ως μια περιοχή, όχι απαραίτητα συνδεδεμένη, η οποία αποτελεί μια προβολή ενός αντικειμένου του φυσικού κόσμου.

Μια τεχνική που έχει προταθεί για την επίλυση του προβλήματος είναι η περιγραφή των εικόνων με κείμενο το οποίο να είναι ενδεικτικό του περιεχομένου τους. Χρησιμοποιώντας αυτές τις περιγραφές, μπορούμε να ταξινομήσουμε ένα σύνολο εικόνων σε ένα διακριτό σύνολο προκαθορισμένων κατηγοριών (*text-based image classification*). Η συγκεκριμένη τεχνική είναι η μέθοδος που ακολουθείται και από τις πιο γνωστές μηχανές αναζήτησης για την ανάκτηση εικόνων μιας συγκεκριμένης κατηγορίας. Παρόλη όμως την ευρεία χρήση της, η συγκριμένη μέθοδος παρουσιάζει σημαντικά μειονεκτήματα ως προς την εφαρμογή και την απόδοσή της. Αφενός μεν η διαδικασία επισήμανσης κάθε εικόνας με μια λεκτική περιγραφή είναι χρονοβόρα, αφετέρου δε η περιγραφή αυτή είναι συνήθως ασαφής και υποκειμενική. Η παρούσα εργασία εστιάζει στο πρόβλημα της ταξινόμησης εικόνας με βάση το περιεχόμενο (*content-based classification*). Ο όρος *content-based* αναφέρεται στο γεγονός ότι η ταξινόμηση γίνεται με βάση το πραγματικό περιεχόμενο της εικόνας και όχι σύμφωνα με κάποια μετα-δεδομένα όπως λέξεις-κλειδιά, ετικέτες ή περιγραφές που σχετίζονται με την εικόνα.

1.2. Συναφή Προβλήματα

Το πρόβλημα της ταξινόμησης εικόνων συχνά μελετάται σε συνδυασμό με άλλα προβλήματα με τα οποία εμφανίζει και αρκετές αναλογίες όπως ο εντοπισμός και η αναγνώριση αντικειμένων σε εικόνες και η ανάκτηση εικόνων με βάση το περιεχόμενο. Στη συνέχεια δίνεται μια σύντομη περιγραφή αυτών των προβλημάτων:

(α) Αναγνώριση αντικειμένων (*object recognition*). Αφορά τον εντοπισμό συγκεκριμένου στιγμιότυπου ενός αντικειμένου σε μια εικόνα. Το συγκεκριμένο πρόβλημα διαφέρει από αυτό της ταξινόμησης στο ότι δεν αναζητάμε ένα οποιοδήποτε μέλος μιας δεδομένης κατηγορίας αλλά ένα συγκεκριμένο στιγμιότυπο

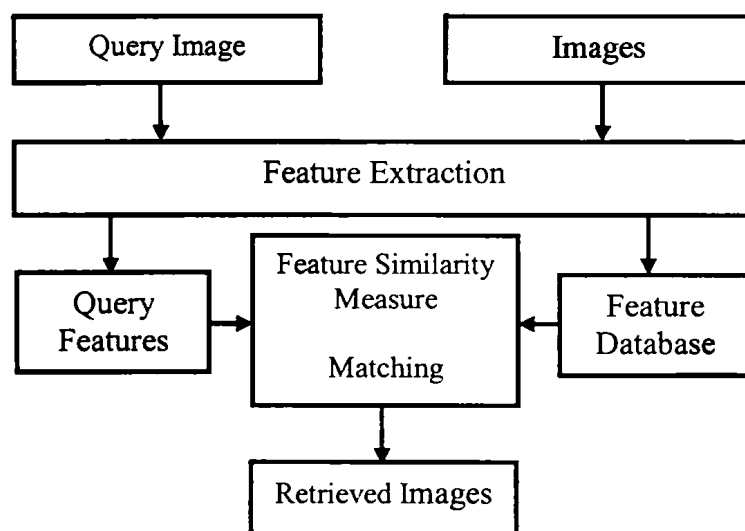


της. Για παράδειγμα, το πρόβλημα της αναγνώρισης διακρίνει μεταξύ τους τις εικόνες που απεικονίζουν δυο δομικά διακριτά αυτοκίνητα (Σχήμα 1.1), ενώ το πρόβλημα της ταξινόμησης δεν τις διακρίνει αλλά τις τοποθετεί στην ίδια κατηγορία.



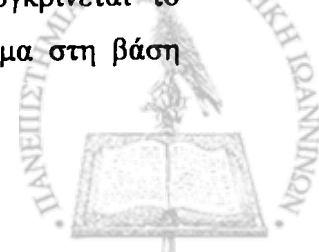
Σχήμα 1.1 Δύο Δομικά Διακριτά Μέλη της Κατηγορίας «Αυτοκίνητο»

- (β) Ανάκτηση εικόνας με βάση το περιεχόμενο (*content-based image retrieval*). Αναφέρεται στη διαδικασία ανάκτησης εικόνων συγκεκριμένης κατηγορίας από μια βάση εικόνων [1]. Το παρακάτω σχήμα απεικονίζει την αρχιτεκτονική ενός CBIR συστήματος [2].



Σχήμα 1.2 Διαδικασία Ανάκτησης Εικόνων

Κάθε εικόνα στη βάση δεδομένων αναπαρίσταται με ένα διάνυσμα χαρακτηριστικών (*feature vector*). Όταν μια εικόνα-ερώτημα (*query image*), ενδεικτική της κατηγορίας των εικόνων που μας ενδιαφέρει, παρέχεται στο σύστημα τότε συγκρίνεται το διάνυσμα χαρακτηριστικών που αντιστοιχεί σε αυτή με κάθε διάνυσμα στη βάση



χαρακτηριστικών (*feature database*) που έχει δημιουργηθεί. Τα διανύσματα που έχουν μεγαλύτερη ομοιότητα με το διάνυσμα της εικόνας ανακτούνται, συνεπακόλουθα και οι αντίστοιχες εικόνες.

(γ) Ανίχνευση αντικειμένου σε εικόνα (*object detection*). Αναφέρεται στον εντοπισμό (της θέσης) ενός αντικειμένου μιας συγκεκριμένης κατηγορίας σε μια εικόνα. Πρόσφατες ερευνητικές εργασίες έχουν ως αντικείμενο μελέτης την ανίχνευση αυτοκινήτων, προσώπων, πεζών κτλ [3, 4, 5, 6]. Το πρόβλημα της ταξινόμησης μπορεί να επιλυθεί εφαρμόζοντας έναν ανιχνευτή για κάθε κατηγορία ενδιαφέροντος σε μια δεδομένη εικόνα. Αυτή η προσέγγιση είναι αναποτελεσματική όταν το πλήθος των προκαθορισμένων κατηγοριών είναι μεγάλο.

1.3. Εφαρμογές του Προβλήματος

Εξαιτίας του μεγάλου όγκου των ψηφιακών εικόνων που υπάρχουν πλέον σε προσωπικούς υπολογιστές, ψηφιακές βιβλιοθήκες και ιδιαίτερα στο Διαδίκτυο έχει καταστεί αναγκαία η οργάνωση των εικόνων αυτών σε κατηγορίες ώστε να διευκολύνεται η διαδικασία ανάκτησής τους. Οι μέθοδοι ταξινόμησης εικόνας με βάση το περιεχόμενο που έχουν αναπτυχθεί βρίσκουν εφαρμογή σε ποικίλα προβλήματα της καθημερινής ζωής καθώς και σε άλλα πιο εξειδικευμένα. Κάποιες από τις πιο σημαντικές εφαρμογές αναφέρονται παρακάτω:

(α) Αναζήτηση εικόνας: Η αναζήτηση εικόνων είναι η πιο συνηθισμένη εφαρμογή του προβλήματος της ταξινόμησης εικόνων. Η ύπαρξη κάποιας οργάνωσης των διαθέσιμων εικόνων είναι απαραίτητη είτε αναφερόμαστε σε προσωπικές συλλογές είτε σε μεγάλες ψηφιακές βιβλιοθήκες. Ο πιο φυσικός τρόπος κατηγοριοποίησης μιας εικόνας είναι με βάση το περιεχόμενο της καθώς η εύρεση μιας σαφούς και αντικειμενικής λεκτικής περιγραφής αποτελεί μια δύσκολη και λιγότερο αποδοτική διαδικασία. Σε αυτό οφείλεται και η χαμηλή απόδοση των μηχανών αναζήτησης οι οποίες αντλούν πληροφορίες όχι από τις ίδιες τις εικόνες αλλά από τις πληροφορίες κειμένου που τις πλαισιώνουν (*metadata*) [7]. Με τον ίδιο τρόπο γίνεται και ανάκτηση αρχείων βίντεο στο Διαδίκτυο.



(β) Παρακολούθηση (*surveillance*): Το πρόβλημα της ανίχνευσης κίνησης έχει κερδίσει το ενδιαφέρον στην κοινότητα της υπολογιστικής όρασης εξαιτίας της δυνατότητας χρησιμοποίησης του σε ένα μεγάλο πλήθος εφαρμογών. Τα τελευταία χρόνια έχουν αναπτυχθεί αρκετές μέθοδοι που εστιάζουν στην αναγνώριση κίνησης σε φυσικά, χωρίς περιορισμούς βίντεο όπως είναι οι ακολουθίες από ταινίες μεγάλου μήκους [8], οι μεταδόσεις αθλητικών εκδηλώσεων [9] και το YouTube [10]. Το συγκεκριμένο πρόβλημα θεωρείται μια μορφή εντοπισμού αντικειμένου σε μεγαλύτερη διάσταση. Πρόσφατες ερευνητικές εργασίες επεκτείνουν μεθόδους που έχουν προταθεί για τον εντοπισμό της θέσης του αντικειμένου σε εικόνες (όπως τα μοντέλα Hough Forest, Bag-of-words κ.α.) στο χωρο-χρονικό πεδίο [11, 12].

(γ) Ιατρικές εικόνες: Η ταχεία ανάπτυξη σύγχρονων ιατρικών συσκευών οπτικοποίησης έχει ως αποτέλεσμα την αυξανόμενη διάθεση ιατρικών εικόνων. Αυτό έχει οδηγήσει σε μια αυξημένη απαίτηση για ανάπτυξη μεθόδων ευρετηριοποίησης, ανάλυσης, σύγκρισης και σχολιασμού (*annotation*) των εικόνων [13]. Συνήθως οι σχετικές βάσεις δεδομένων είναι προσβάσιμες μέσω μιας περιγραφής και μεταδεδομένων. Αυτές οι πληροφορίες εισάγονται είτε αυτόματα από το ψηφιακό σύστημα είτε χειροκίνητα από τους γιατρούς ή ακτινολόγους κατά τη διαδικασία τεκμηρίωσης. Στόχος του αυτόματου σχολιασμού εικόνων είναι η περιγραφή του περιεχομένου μιας εικόνας βάση των χαρακτηριστικών της με έναν επίσημο και γενικευμένο τρόπο χρησιμοποιώντας μεθόδους αναγνώρισης προτύπων και δομικής ανάλυσης. Αυτή η περιγραφή μπορεί να χρησιμοποιηθεί στη συνέχεια για τη σύγκριση μιας νέας «άγνωστης» εικόνας με ένα γνωστό σύνολο δεδομένων προκαθορισμένων κατηγοριών και την ανάθεσή της στη σωστή κατηγορία.

(δ) Τηλεπισκόπηση (*remote sensing*): Η τηλεπισκόπηση είναι η ανίχνευση και η μέτρηση της ηλεκτρομαγνητικής ενέργειας που προέρχεται από απομακρυσμένα αντικείμενα τα οποία είναι κατασκευασμένα από διάφορα υλικά. Η ταξινόμηση εικόνων τηλεπισκόπησης στηρίζεται σε πληροφορίες φάσματος καθώς διαφορετικά αντικείμενα έχουν διαφορετικές φασματικές υπογραφές [14]. Η τεχνολογία της τηλεπισκόπησης μπορεί να χρησιμοποιηθεί για τη ανάκτηση πληροφοριών σχετικά με την επιφάνεια της γης, την ατμόσφαιρα ή τις διαστημικές πλατφόρμες.



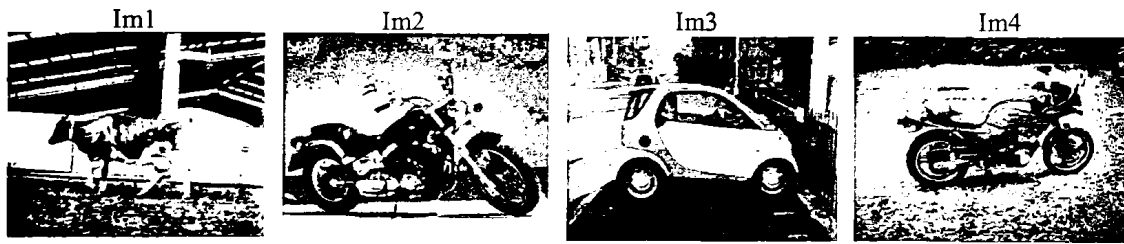
1.4. Στόχοι της Εργασίας

Η παρούσα εργασία επικεντρώνεται στη μελέτη του προβλήματος της ταξινόμησης εικόνων με βάση το περιεχόμενο τους μέσω της χρήσης στοχαστικών δασών (random forests) ως μοντέλου ταξινόμησης. Προτείνεται μια μεθοδολογία κατασκευής στοχαστικών δασών χρησιμοποιώντας τον αλγόριθμο ομαδοποίησης K-means για τη δυαδική διαμέριση κάθε ενδιάμεσου κόμβου. Η διαδικασία απόφασης του μοντέλου προκύπτει μέσω μιας στοχαστικής διαδικασίας ψηφοφορίας η οποία εκμεταλλεύεται την πληροφορία των τερματικών κόμβων. Βασικός στόχος της εργασίας είναι η πειραματική μελέτη των στοχαστικών δασών ως συνδυαστικών ταξινομητών και της απόδοσής τους στο πρόβλημα της ταξινόμησης εικόνων.

- Γενικά, το πρόβλημα της ταξινόμησης των εικόνων με βάση το περιεχόμενο (π.χ. αυτοκίνητα, πεζοί κ.α.) μπορεί να διατυπωθεί ως εξής: Δοθέντος ενός συνόλου N εικόνων $I = \{I_1, I_2, \dots, I_N\}$ και ενός πεπερασμένου συνόλου M κατηγοριών $C = \{C_1, C_2, \dots, C_M\}$, η διαδικασία της ταξινόμησης εικόνων συνίσταται στην ανάθεση μιας λογικής τιμής (*True/False*) σε κάθε ζεύγος της μορφής (I_i, C_j) του συνόλου $I \times C$. Η ανάθεση της τιμής *True* σημαίνει ότι η εικόνα I_i ανήκει στην κατηγορία C_j , ενώ η ανάθεση της τιμής *False* σημαίνει ότι η εικόνα I_i δεν ανήκει στην κατηγορία C_j . Συνεπώς επιδιώκεται η εκπαίδευση ενός μοντέλου ταξινόμησης το οποίο θα μπορεί να προβλέπει με ακρίβεια την κατηγορία (*class label*) νέων «άγνωστων» εικόνων.

Αν έχουμε μια κατηγορία ενδιαφέροντος τότε το πρόβλημα της ταξινόμησης είναι δυαδικό (*binary classification*) καθώς το σύστημα πρέπει να δώσει μια από δύο τιμές *True/False* ($K=2$). Αν οι προκαθορισμένες κατηγορίες είναι περισσότερες από δύο ($K>2$) τότε έχουμε ταξινόμηση πολλών κατηγοριών (*multi-class classification*). Παρόλα αυτά το πρόβλημα της ταξινόμησης πολλών κατηγοριών μπορεί να θεωρηθεί ως ένα πρόβλημα πολλών (ίσο με K) δυαδικών προβλημάτων ταξινόμησης, ένα για κάθε κατηγορία. Για παράδειγμα, αν το παρακάτω σύνολο εικόνων (Σχήμα 1.3) πρέπει να ταξινομηθεί σε $K=3$ προκαθορισμένες κατηγορίες $C = \{Cars, Cows, Motorbikes\}$, ένας ιδανικός μηχανισμός ταξινόμησης έχει την μορφή που φαίνεται στους δυο πίνακες παρακάτω:





Σχήμα 1.3 Παραδείγματα Εικόνων Πολλών Κατηγοριών

Πίνακας 1.1 Χρήση Ενός Ταξινομητή Πολλών Κατηγοριών (K=3)

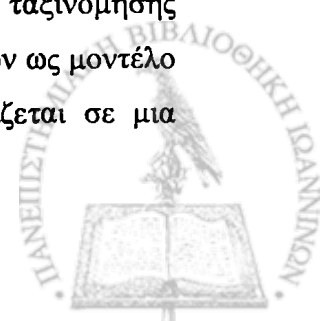
	Cars	Cows	Motorbikes
Im1	False	True	False
Im2	False	False	True
Im3	True	False	False
Im4	False	False	True

Πίνακας 1.2 Χρήση Τριών Δυαδικών Ταξινομητών

	Cars	Other		Cows	Other		Motorbikes	Other
Im1	False	True	Im1	True	False	Im1	False	True
Im2	False	True	Im2	False	True	Im2	True	False
Im3	True	False	Im3	False	True	Im3	False	True
Im4	False	True	Im4	False	True	Im4	True	False

1.5. Δομή της Εργασίας

Η παρούσα εργασία αποτελείται από 5 κεφάλαια. Το πρώτο κεφάλαιο είναι εισαγωγικό, στο δεύτερο κεφάλαιο ορίζεται το πρόβλημα της ταξινόμησης και παρουσιάζονται συνοπτικά οι πιο γνωστές τεχνικές μηχανικής μάθησης για την επίλυσή του. Παρουσιάζονται γνωστές μέθοδοι κατασκευής συνδυαστικών ταξινομητών και ειδικότερα των στοχαστικών δασών. Στο τρίτο κεφάλαιο περιγράφεται η εφαρμογή των στοχαστικών δασών στο πρόβλημα της ταξινόμησης εικόνων. Παρουσιάζονται τα στάδια κατασκευής των στοχαστικών δασών ως μοντέλο ταξινόμησης καθώς και ενός μηχανισμού απόφασης ο οποίος στηρίζεται σε μια



στοχαστική διαδικασία «ψηφοφορίας». Στο τέταρτο κεφάλαιο περιγράφεται αναλυτικά η διαδικασία διεξαγωγής των πειραμάτων τα οποία πραγματοποιήθηκαν για την εκτίμηση της απόδοσης της προτεινόμενης μεθόδου και παρουσιάζονται τα αποτελέσματα τους, ενώ στο πέμπτο κεφάλαιο παρουσιάζονται τα συμπεράσματα που προέκυψαν από την εργασία καθώς και η μελλοντική εργασία που θα μπορούσε να γίνει για τη βελτίωση της μεθόδου.



ΚΕΦΑΛΑΙΟ 2. ΣΥΝΔΥΑΣΤΙΚΟΙ ΤΑΞΙΝΟΜΗΤΕΣ

2.1 Τεχνικές Μάθησης

2.2 Μέθοδοι Ταξινόμησης

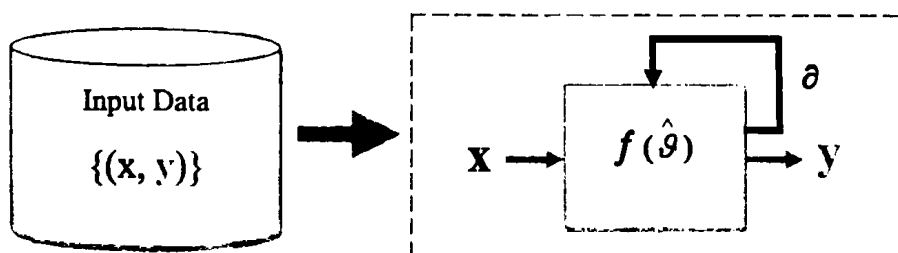
2.3 Συνδυαστικές Τεχνικές Μεθόδων Μηχανικής Μάθησης

2.4 Κατασκευή Συνδυαστικών Ταξινομητών

2.5 Τυχαία Δάση (Random Forests)

2.1. Τεχνικές Μάθησης

Η έννοια της μάθησης σε ένα γνωστικό σύστημα (*cognitive system*), δηλ. σε ένα φυσικό ή τεχνητό σύστημα επεξεργασίας πληροφοριών με δυνατότητες αντίληψης, μάθησης και λήψης απόφασης, συνδέεται με δύο βασικές ιδιότητες: (α) την ικανότητά του για πρόσκτηση γνώσης μέσω της αλληλεπίδρασή του με το περιβάλλον και (β) την ικανότητά του να βελτιώνει με τον τρόπο με τον οποίο εκτελεί μία ενέργεια (συνεπώς και την απόδοσή του) μέσω της επανάληψης. Για να είναι όμως αυτό εφικτό θα πρέπει να προσδιοριστούν με ακρίβεια οι παράμετροι (θ) που συντελούν σ' αυτή τη βελτίωση.



Σχήμα 2.1 Η Διαδικασία της Μάθησης Ενός Γνωστικού Συστήματος



Ο άνθρωπος προσπαθεί να κατανοήσει το περιβάλλον του παρατηρώντας το και δημιουργώντας μια απλοποιημένη (αφαιρετική) εκδοχή του που ονομάζεται μοντέλο. Η γνώση που προκύπτει από ένα τέτοιο μοντέλο ονομάζεται επαγωγική μάθηση (*inductive learning*). Επιπλέον, ο άνθρωπος έχει τη δυνατότητα να οργανώνει και να συσχετίζει τις εμπειρίες και τις παραστάσεις του δημιουργώντας νέες δομές που ονομάζονται πρότυπα. Η δημιουργία μοντέλων ή προτύπων από ένα σύνολο δεδομένων, από ένα υπολογιστικό σύστημα, ονομάζεται μηχανική μάθηση (*machine learning*).

Έχουν αναπτυχθεί πολλές τεχνικές μηχανικής μάθησης (ταξινόμησης, παλινδρόμησης, ομαδοποίησης, κανόνων συσχέτισης, διαφορικών εξισώσεων) οι οποίες χρησιμοποιούνται ανάλογα με τη φύση του προβλήματος και εμπίπτουν σε ένα από τα παρακάτω δυο είδη:

- Μάθηση χωρίς επίβλεψη (*unsupervised learning*) ή μάθηση από παρατήρηση.
- Επιβλεπόμενη μάθηση (*supervised learning*) ή μάθηση με παραδείγματα.

Στη επιβλεπόμενη μάθηση το σύστημα καλείται να «μάθει» επαγωγικά μέσω ενός συνόλου δεδομένων $\{x, y\}$ μια συνάρτηση f , η οποία αποτελεί την περιγραφή ενός μοντέλου. Υπάρχει πάντα κάποιος «επιβλέπων» ο οποίος παρέχει τη σωστή τιμή εξόδου y της συνάρτησης για τα δεδομένα που εξετάζονται. Αντίθετα, στη μάθηση χωρίς επίβλεψη το σύστημα δημιουργεί πρότυπα ανακαλύπτοντας συσχετίσεις ή ομάδες σε ένα σύνολο δεδομένων για τα οποία δεν η τιμή εξόδου y της συνάρτησης είναι γνωστή. Το αποτέλεσμα είναι ένα σύνολο προτύπων - περιγραφών, κάθε ένα από τα οποία περιγράφει ένα μέρος των δεδομένων.

2.1.1 Μάθηση χωρίς Επίβλεψη (*Unsupervised Learning*)

Η συσταδοποίηση ή ομαδοποίηση (*clustering*) είναι μια από τις πιο δημοφιλής μεθόδους μάθησης χωρίς επίβλεψη. Στόχος του αλγορίθμου μάθησης είναι να καθορίσει συνεκτικά υποσύνολα (συστάδες, *clusters*) των παραδειγμάτων εκπαίδευσης. Το πλήθος των συστάδων είτε δίνεται εκ των προτέρων ως μέρος της γνώσης που προϋπάρχει είτε καθορίζεται από τον αλγόριθμο μάθησης. Επομένως ο αλγόριθμος πρέπει να καθορίζει ένα σχετικά μικρό πλήθος συνεκτικών συστάδων δηλ. υποσυνόλων παρόμοιων παραδειγμάτων. Η επιλογή του μέτρου ομοιότητας



(απόστασης) είναι το πιο σημαντικό κομμάτι της προϋπάρχουσας γνώσης και υψίστης σημασίας για μια επιτυχή και ουσιαστική ομαδοποίηση. Οι πιο γνωστές προσεγγίσεις για ομαδοποίηση είναι οι ακόλουθες:

- **Ιεραρχική ομαδοποίηση (*hierarchical clustering*):** Στη συσσωρευτική (*bottom-up*) ιεραρχική ομαδοποίηση, κάθε παράδειγμα αποτελεί αρχικά μια ξεχωριστή συστάδα. Σε κάθε βήμα του αλγόριθμου, οι πιο όμοιες συστάδες συγχωνεύονται, σχηματίζοντας με αυτόν τον τρόπο ένα δέντρο ομαδοποιήσεων (δενδρόγραμμα). Συνήθως η συγχώνευση γίνεται κατά ζεύγη και οδηγεί στο σχηματισμό ενός δυαδικού δέντρου. Η συγχώνευση συνεχίζεται μέχρις ότου όλα τα παραδείγματα να ανήκουν σε μια μοναδική συστάδα. Τελικά ο αλγόριθμος ή ο τελικός χρήστης επιλέγουν το πιο κατάλληλο επίπεδο συσταδοποίησης από το δέντρο που έχει κατασκευαστεί. Στη διαιρετική (*top-down*) ιεραρχική ομαδοποίηση, όλα τα παραδείγματα ανήκουν αρχικά σε μια συστάδα. Το πλήθος των συστάδων αυξάνεται σε κάθε βήμα του αλγόριθμου διαιρώντας μια υπάρχουσα συστάδα σε (δυο συνήθως) υπό-συστάδες. Η διαδικασία συνεχίζεται μέχρι να σχηματιστεί ένα κατάλληλο πλήθος συστάδων.

- **Ομαδοποίηση με βάση τη διαμέριση (*partitional clustering*):** Αρχικά, πρέπει να είναι γνωστό το πλήθος c των απαιτούμενων ξένων συστάδων. Δοθέντος ενός κριτηρίου μέτρησης της καταλληλότητας της διαμέρισης των παραδειγμάτων σε c συστάδες, ο αλγόριθμος αναζητά τις βέλτιστες διαμερίσεις των παραδειγμάτων. Η διαδικασία ξεκινά με μια αρχική ομαδοποίηση (η οποία συνήθως σχηματίζεται με τυχαία επιλογή c παραδειγμάτων), και ένα σύνολο μετασχηματισμών για την αλλαγή μίας διαμέρισης σε μια άλλη διαμέριση. Ο αλγόριθμος μάθησης τροποποιεί μια διαμέριση μέχρις ότου κανένας επιτρεπτός μετασχηματισμός να μη μπορεί να βελτιώσει το δοθέν κριτήριο για την καταλληλότητα της διαμέρισης. Μια από τις πιο γνωστές τεχνικές ομαδοποίησης που ανήκει σ' αυτήν την κατηγορία είναι ο αλγόριθμος *K-means*, οποίος αναζητά K κέντρα τα οποία αντιπροσωπεύουν το σύνολο των αρχικών σημείων.



Συγκεκριμένα, δοθέντος ενός συνόλου σημείων, στόχος του αλγόριθμου K-means [22, 23] είναι να χωρίσει τα σημεία σε K συστάδες. Αρχικά, επιλέγονται K σημεία-κέντρα, όσες και οι απαιτούμενες συστάδες, και καθένα από τα υπόλοιπα σημεία ανατίθενται στο «κοντινότερο» κέντρο. Στη συνέχεια το κέντρο κάθε συστάδας ενημερώνεται με βάση τα σημεία που του έχουν ανατεθεί. Τα δύο αυτά βήματα επαναλαμβάνονται έως ότου κανένα σημείο να μην αλλάζει συστάδα.

Ως «κοντινότερο» νοείται το κέντρο που έχει την μικρότερη απόσταση από το σημείο. Για τον προσδιορισμό της έννοιας του κοντινότερου, ο αλγόριθμος χρησιμοποιεί ένα μέτρο απόστασης ανάλογα με το είδος των δεδομένων (π.χ. Euclidean distance, cosine similarity, Hamming, Manhattan). Κάθε μέτρο απόστασης δίνει διαφορετική ομαδοποίηση. Συνήθως το κέντρο της συστάδας ορίζεται ως το μέσο (mean) των σημείων που ανήκουν στη συστάδα. Ο στόχος του αλγόριθμου είναι να βρεθεί η βέλτιστη ομαδοποίηση. Η αντικειμενική συνάρτηση για την εκτίμηση της ποιότητας της ομαδοποίησης ορίζεται σύμφωνα με το άθροισμα των αποστάσεων των σημείων από τα κέντρα των συστάδων στα οποία ανήκουν.

2.1.2 Επιβλεπόμενη Μάθηση (*Supervised Learning*)

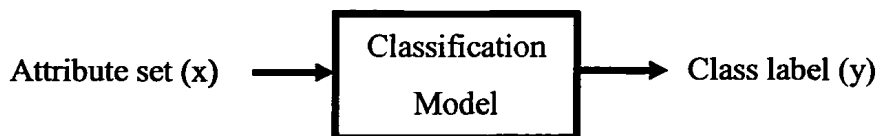
Οι μέθοδοι που εφαρμόζουν επιβλεπόμενη μάθηση συνήθως χρησιμοποιούνται για ταξινόμηση. Για να ορίσουμε το πρόβλημα της ταξινόμησης ας υποθέσουμε ότι έχουμε ένα αντικείμενο, το οποίο έχει περιγραφεί με διάφορα χαρακτηριστικά (γνωρίσματα, ιδιότητες). Κάθε τέτοιο αντικείμενο μπορεί να ανατεθεί σε μία ακριβώς κατηγορία από ένα πεπερασμένο σύνολο πιθανών κατηγοριών. Τα χαρακτηριστικά είναι ανεξάρτητες παρατηρούμενες μεταβλητές, συνεχείς ή διακριτές. Η κατηγορία είναι μια εξαρτημένη διακριτή μεταβλητή και η τιμή της καθορίζεται από τις τιμές των αντίστοιχων ανεξάρτητων μεταβλητών. Οι μέθοδοι μηχανικής μάθησης χρησιμοποιούνται μεταξύ των άλλων και για την κατασκευή των ταξινομητών. Στόχος των ταξινομητών είναι να καθορίζουν σε ποια κατηγορία πρέπει να ανατεθεί κάθε αντικείμενο.

Ένα τυπικό παράδειγμα ταξινόμησης αποτελεί η ιατρική διάγνωση: κάθε ασθενής περιγράφεται με συνεχή χαρακτηριστικά (π.χ. ηλικία, ύψος, θερμοκρασία σώματος, καρδιακή συχνότητα, πίεση) και με διακριτά χαρακτηριστικά (π.χ. φύλο, θέση του



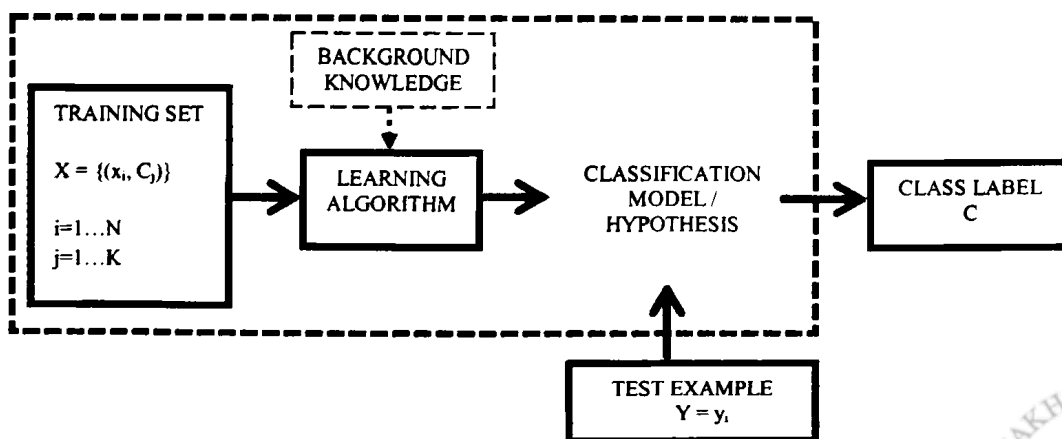
πόνου). Στόχος του ταξινομητή είναι να παράγει μια διάγνωση (π.χ. υγιής, γρίπη, πνευμονία).

Για να καθορίσει την κατηγορία, ένας ταξινομητής χρειάζεται να περιγράψει μια διακριτή συνάρτηση, δηλ. μια αντιστοίχιση από τον χώρο των χαρακτηριστικών στον χώρο των κατηγοριών.



Σχήμα 2.2 Η Ταξινόμηση ως Διαδικασία Αντιστοίχισης από τον Χώρο των Χαρακτηριστικών στον Χώρο των Κατηγοριών.

Αυτή η συνάρτηση μπορεί να δίνεται εκ των προτέρων ή μπορεί να προκύψει από τα παραδείγματα εκπαίδευσης, τα οποία περιγράφουν παλαιότερα λυμένα προβλήματα. Για παράδειγμα, ας θεωρήσουμε την περίπτωση της ιατρικής διάγνωσης. Παλαιότερα λυμένα προβλήματα αποτελούν οι ιατρικές εγγραφές, οι οποίες περιλαμβάνουν τις διαγνώσεις για όλους τους ασθενείς οι οποίοι νοσηλεύτηκαν σε ένα νοσοκομείο. Στόχος του αλγόριθμου μάθησης είναι να καθορίσει την αντιστοίχιση μαθαίνοντας από το σύνολο των ασθενών με γνωστές διαγνώσεις. Αυτή η αντιστοίχιση (η οποία αναπαρίσταται ως συνάρτηση, κανόνας κτλ) μπορεί να χρησιμοποιηθεί στη συνέχεια για τη διάγνωση νέων ασθενών.



Σχήμα 2.3 Η Διαδικασία της Ταξινόμησης



Οι εισοδοί σε έναν αλγόριθμο μηχανικής μάθησης είναι το σύνολο δεδομένων εκπαίδευσης (*training set*) και το γνωστικό υπόβαθρο (*background knowledge*) που υπάρχει για το συγκεκριμένο πρόβλημα [15]. Η έξοδος του αλγόριθμου είναι μια περιγραφή (μοντέλο, υπόθεση) η οποία περιγράφει και εξηγεί τα δεδομένα και τις προηγούμενες γνώσεις. Αντί για τη λέξη *μοντέλο* χρησιμοποιούμε πολλές φορές τη λέξη *υπόθεση* καθώς στην περίπτωση της μηχανικής μάθησης ένα μοντέλο ταξινόμησης αποτελεί ένα αφαιρετικό μοντέλο των δεδομένων. Το γνωστικό υπόβαθρο είναι τις περισσότερες φορές ο χώρος των πιθανών μοντέλων και ένα κριτήριο βελτιστοποίησης. Για αυτό το λόγο ο αλγόριθμος μάθησης αναζητά ένα μοντέλο το οποίο να ικανοποιεί το κριτήριο βελτιστοποίησης.

Επειδή όμως ο χώρος των πιθανών μοντέλων είναι πολύ μεγάλος η αναζήτηση της βέλτιστης λύσης είναι ανέφικτη, γι' αυτό αναζητάμε τις καλύτερες υποβέλτιστες λύσεις. Ως εκ τούτου χρειαζόμαστε κριτήρια τα οποία μετράνε την ποιότητα των μοντέλων. Ένα γνωστό κριτήριο είναι η αρχή *Occam's razor*, σύμφωνα με την οποία το πιο απλό μοντέλο είναι και το πιο αξιόπιστο. Η συγκεκριμένη αρχή μπορεί να γενικευτεί στην αρχή του *ελάχιστου μήκους περιγραφής*. Για δεδομένη ακρίβεια πρόβλεψης, το πιο πιθανό μοντέλο είναι εκείνο που είναι το μικρότερο σε μέγεθος και ταιριάζει καλύτερα στα δεδομένα. Επειδή το μοντέλο είναι το μικρότερο, είναι και το πιο κατανοητό αφού χρησιμοποιεί το ελάχιστο πλήθος παραμέτρων (χαρακτηριστικών) και εξασφαλίζει με αυτόν τον τρόπο την πιο γρήγορη πρόβλεψη και μεγάλη γενικευτική ικανότητα.

2.2 Μέθοδοι ταξινόμησης

Όπως αναφέρθηκε και προηγουμένως, σε μεθόδους μηχανικής μάθησης οι οποίες εφαρμόζουν επίβλεψη, ο αλγόριθμος μάθησης λαμβάνει ως εισόδους την προηγούμενη γνώση που υπάρχει για το πρόβλημα και τα δεδομένα εκπαίδευσης, εξετάζει τον χώρο της υπόθεσης και επιστρέφει ως αποτέλεσμα την τελική υπόθεση (μοντέλο). Ως εκ τούτου είναι απαραίτητο η προηγούμενη γνώση, τα δεδομένα εκπαίδευσης και ο χώρος υπόθεσης να αναπαριστώνται αποδοτικά ώστε να επιτρέπουν την αποδοτική χρήση και παραγωγή της νέας γνώσης. Κατά την εξέταση του χώρου της υπόθεσης, ο αλγόριθμος χρησιμοποιεί κατάλληλους τελεστές για να



προχωρήσει στον χώρο της αναζήτησης δηλ. για να μεταβάλλει την τρέχουσα υπόθεση. Διαφορετικοί αλγόριθμοι μάθησης αναπαριστούν την γνώση με διαφορετικό τρόπο. Οι πιο συνηθισμένοι αλγόριθμοι ταξινόμησης είναι τα δέντρα απόφασης, οι κανόνες απόφασης, οι ταξινομητές κοντινότερου γείτονα, οι μηχανές διανυσμάτων υποστήριξης, τα Μπεϋζιανά δίκτυα πεποίθησης, οι ταξινομητές Naive Bayes, οι γραμμικές διακρίνουσες συναρτήσεις και τα τεχνητά νευρωνικά δίκτυα. Παρακάτω δίνεται μια συνοπτική περιγραφή αυτών των μεθόδων ταξινόμησης [15].

2.2.1 Ταξινομητές Κοντινότερων Γειτόνων (*Nearest Neighbors classifiers*)

Ο συγκεκριμένος ταξινομητής είναι η απλούστερη μέθοδος ταξινόμησης καθώς είναι μη παραμετρική. Ο ταξινομητής k - κοντινότερων γειτόνων (k -NN) παράγει τη βάση της γνώσης του αποθηκεύοντας όλα τα δεδομένα εκπαίδευσης. Για την ταξινόμηση ενός νέου παραδείγματος χρησιμοποιούνται τα αποθηκευμένα δεδομένα για να βρεθεί ένα συγκεκριμένο πλήθος (k) των πιο όμοιων παραδειγμάτων εκπαίδευσης (κοντινότεροι γείτονες), σύμφωνα με μια μετρική απόστασης. Το νέο παράδειγμα ανατίθεται στην κατηγορία που είναι επικρατέστερη μεταξύ των κοντινότερων γειτόνων του. Ο αριθμός k είναι συνήθως ένας μικρός περιττός αριθμός (π.χ. 5 ή 7).

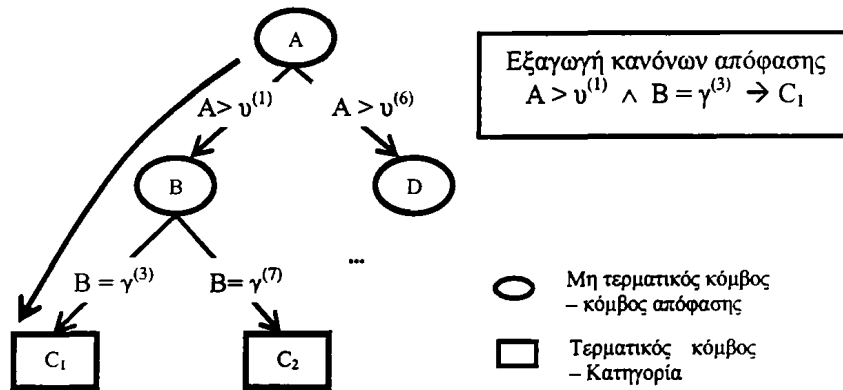
2.2.2 Δέντρα Απόφασης και Κανόνες Απόφασης (*Decision Trees and Rules*)

Οι αλγόριθμοι για την κατασκευή των δέντρων και των κανόνων απόφασης επιλέγουν χαρακτηριστικά και κατάλληλα υποσύνολα των τιμών τους και τα χρησιμοποιούν ως δομικά στοιχεία της υπόθεσης (πρότερο) ενός συνδυαστικού κανόνα. Το τμήμα της απόφασης (απότοκο) περιλαμβάνει την κατηγορία.

Δοθέντος ενός χαρακτηριστικού A και των τιμών του $v^{(1)} \dots v^{(m)}$, τα δέντρα και οι κανόνες απόφασης σχηματίζουν προτάσεις όπως για παράδειγμα $A = v^{(j)}$ ή $A \in S \subset \{v^{(1)}, \dots, v^{(m)}\}$. Τα συνεχή χαρακτηριστικά είτε διακριτοποιούνται εκ των προτέρων ή, ακόμα καλύτερα, οι όροι όπως $A > v^{(j)}$, $A \leq v^{(j)}$, ή $v^{(1)} \geq A > v^{(2)}$ σχηματίζονται κατά τη διάρκεια της εκπαίδευσης. Το πρότερο ενός κανόνα σχηματίζεται συνήθως από τη σύζευξη τέτοιων προτάσεων π.χ. αν το A είναι διακριτό χαρακτηριστικό τότε οι



προτάσεις σχηματίζονται ως εξής $D_i = (A_i \in S)$, όπου $S \subset \{v^{(1)}, \dots, v^{(m)}\}$. Ένας κανόνας απόφασης έχει την εξής μορφή $D_{i1} \wedge D_{i2} \wedge \dots \wedge D_{in} \rightarrow C_0$.



Σχήμα 2.4 Η Δομή ενός Δυαδικού Δέντρου Απόφασης

Ένα δέντρο απόφασης αποτελεί μια ειδική κατηγορία ενός συνόλου κανόνων απόφασης. Αποτελείται από εσωτερικούς κόμβους οι οποίοι αντιστοιχούν στα χαρακτηριστικά (*attributes*), ακμές οι οποίες αντιστοιχούν σε υποσύνολα τιμών των χαρακτηριστικών και τερματικούς κόμβους οι οποίοι αντιστοιχούν στις κατηγορίες (*class labels*). Κάθε μονοπάτι το οποίο ξεκινάει από τη ρίζα του δέντρου και καταλήγει σε ένα τερματικό κόμβο αντιστοιχεί σε έναν κανόνα απόφασης. Οι υποθέσεις (χαρακτηριστικό-σύνολο τιμών χαρακτηριστικού) που υπολογίζονται σε κάθε εσωτερικό κόμβο συνδέονται μέσω σύζευξης (*conjunction*). Για παράδειγμα, το μονοπάτι που έχει επισημανθεί Σχήμα 2.3 αντιστοιχεί στον ακόλουθο κανόνα: $A > v^{(1)} \wedge B = \gamma^{(3)} \rightarrow C_1$. Κάθε δέντρο απόφασης μπορεί να μετασχηματιστεί σε τόσους κανόνες όσα και τα φύλλα του δέντρου. Αντίθετα, δεν μπορεί κάθε σύνολο κανόνων απόφασης να μετατραπεί σε ένα δέντρο απόφασης. Η αναπαράσταση της γνώσης με δέντρα απόφασης μπορεί να γενικευτεί αξιολογώντας σε κάθε κόμβο μια αυθαίρετη συνάρτηση διάφορων χαρακτηριστικών αντί για ένα μοναδικό χαρακτηριστικό. Αν κάθε εσωτερικός κόμβος έχει το πολύ δύο «παιδιά» (*successors*), τότε το δέντρο καλείται δυαδικό. Με βάση την αρχή της μηχανικής μάθησης, προτιμάμε τα απλούστερα δέντρα γιατί οι κανόνες που περιγράφουν είναι πιο γενικοί και ως εκ τούτου έχουν καλύτερη γενικευτική ικανότητα.



2.2.3 Μπεϋζιανοί Ταξινομητές (Bayesian Classifiers)

Οι Μπεϋζιανοί ταξινομητές ή ταξινομητές κατά Bayes βασίζονται στον υπολογισμό της εκ των υστέρων πιθανότητας όπως υπολογίζεται από τον κανόνα του Bayes και μοντελοποιούν την πιθανοτική σχέση μεταξύ του συνόλου χαρακτηριστικών και της κατηγορίας. Συγκεκριμένα, δοθέντος των τιμών των χαρακτηριστικών ενός νέου παραδείγματος, στόχος του Μπεϋζιανού ταξινομητή είναι να υπολογίσει τις υπό συνθήκη πιθανότητες για όλες τις πιθανές κατηγορίες. Ο Μπεϋζιανός ταξινομητής ελαχιστοποιεί την αναμενόμενη πιθανότητα σφάλματος (*error rate*) και υπό αυτήν την έννοια είναι βέλτιστος. Καθώς όμως οι ακριβείς ταξινομητές είναι δύσκολο να εξασφαλιστούν (εκτός αν τα δεδομένα εκπαίδευσης καλύπτουν πλήρως τον χώρο των χαρακτηριστικών και δεν υπάρχει καθόλου θόρυβος), συγκεκριμένες προσεγγίσεις και υποθέσεις πρέπει να χρησιμοποιηθούν κατά τον υπολογισμό των υπό συνθήκη πιθανοτήτων.

Ο ταξινομητής Naive Bayes υποθέτει την υπό συνθήκη ανεξαρτησία των χαρακτηριστικών δεδομένης της κατηγορίας. Αυτή η ισχυρή υπόθεση συνήθως εξασφαλίζει αξιόπιστες εκτιμήσεις των υπό συνθήκη πιθανοτήτων οι οποίες απαιτούνται για ταξινόμηση, ακόμα και από πολύ μικρά σύνολα δεδομένων. Οι υλοποιήσεις του ταξινομητή Naive Bayes συχνά υποθέτουν ότι χρησιμοποιούνται μόνο διακριτά χαρακτηριστικά, επομένως τα συνεχή χαρακτηριστικά πρέπει να διακριτοποιηθούν εκ των προτέρων.

Μια γενίκευση του ταξινομητή Naive Bayes αποτελούν τα Μπεϋζιανά δίκτυα πεποίθησης (*Bayesian belief networks*). Η συγκεκριμένη μέθοδος χρησιμοποιεί άκυκλους γράφους για να μοντελοποιήσει τις εξαρτήσεις ανάμεσα στα χαρακτηριστικά των παραδειγμάτων και τις κατηγορίες. Υποθέτει έμμεσα την υπό συνθήκη ανεξαρτησία μεταξύ κόμβων (χαρακτηριστικών) οι οποίοι δεν συνδέονται απευθείας. Η τοπολογία ενός Μπεϋζιανού δικτύου μπορεί να δίνεται εκ των προτέρων (ως τμήμα του υπόβαθρου των γνώσεων για το πρόβλημα), ή μπορεί να κατασκευαστεί από τα δεδομένα. Οι υπό συνθήκη εξαρτήσεις υπολογίζονται και αυτές από τα δεδομένα, ενώ η τοπολογία άκυκλου γράφου εξασφαλίζει τον αποτελεσματικό υπολογισμό των υπό συνθήκη πιθανοτήτων για όλες τις κατηγορίες.



Τα Μπεϋζιανά δίκτυα πεποιθήσης μπορούν να χρησιμοποιηθούν για να μοντελοποιήσουν κατανομές πιθανότητας.

2.2.4 Διακρίνουσες Συναρτήσεις (*Discriminant Functions*)

Στόχος των μεθόδων ταξινόμησης που βασίζονται σε παραμετρικές διακρίνουσες συναρτήσεις είναι η εκτίμηση των παραμέτρων των διακρινουσών συναρτήσεων των οποίων η δομή είναι εκ των προτέρων καθορισμένη. Η συνάρτηση αυτή είναι ουσιαστικά ένα υπερεπίπεδο το οποίο διαχωρίζει κατάλληλα τον χώρο των χαρακτηριστικών των δειγμάτων. Αν έχουμε περισσότερες από δύο κατηγορίες, χρειαζόμαστε μια ξεχωριστή υπερεπιφάνεια για κάθε ζεύγος κατηγοριών. Οι διακρίνουσες συναρτήσεις μπορούν να είναι γραμμικές, τετραγωνικές, πολυωνυμικές κτλ. Στην περίπτωση των γραμμικών συναρτήσεων η αντίστοιχη υπερεπιφάνεια είναι ένα υπερεπίπεδο που διχοτομεί μεταξύ δύο κατηγοριών. Οι συντελεστές του υπερεπιπέδου καθορίζονται με τέτοιο τρόπο ώστε να μειώνεται το ποσοστό σφάλματος ταξινόμησης.

Συχνά χρησιμοποιείται η γραμμική διακρίνουσα συνάρτηση Fischer, η οποία υποθέτει ότι τα παραδείγματα εκπαίδευσης κάθε κατηγορίας ακολουθούν την κανονική κατανομή. Μεγιστοποιεί την (Ευκλείδεια) απόσταση ανάμεσα στα παραδείγματα των δύο κατηγοριών και υπολογίζει ένα βέλτιστο όριο ταξινόμησης μεταξύ των κατηγοριών.

Εναλλακτικά μπορούν να χρησιμοποιηθούν συναρτήσεις πυρήνα (*kernel functions*) οι οποίες έχουν ως στόχο τον μετασχηματισμό των δειγμάτων σε ένα νέο χώρο ορίζοντας νέες συναρτήσεις βάσης (*kernel space*) όπου η διακριτική ικανότητα του μηχανισμού ταξινόμησης να είναι σημαντικά αυξημένη.

2.2.5 Μηχανές Διανυσμάτων Υποστήριξης (*Support Vector Machines - SVM*)

Οι Μηχανές Διανυσμάτων Υποστήριξης αποτελούν μία από τις πιο ακριβείς προσεγγίσεις διακρινουσών συναρτήσεων για ταξινόμηση. Ο ταξινομητής SVM προσπαθεί να βρει ένα υπερεπίπεδο απόφασης το οποίο να διαχωρίζει το σύνολο των



παραδειγμάτων εκπαίδευσης με τέτοιο τρόπο ώστε τα παραδείγματα που ανήκουν στην ίδια κατηγορία να είναι στη ίδια πλευρά του υπερεπιπέδου. Μεταξύ όλων των πιθανών υπερεπιπέδων αναζητά εκείνο για το οποίο η απόσταση από το κοντινότερο παράδειγμα είναι μέγιστη, δηλ. αναζητά υπερεπίπεδο μέγιστου περιθωρίου (*maximal margin hyperplane*) [15, 16].

Το υπερεπίπεδο απόφασης για ένα σύνολο N παραδειγμάτων εκπαίδευσης $\{(x_i, y_i)\}_{i=1}^N$ και δύο κατηγορίες $y_i \in \{-1, 1\}$ ορίζεται ως εξής $w \bullet x + b = 0$, όπου w και b είναι οι παράμετροι του μοντέλου και το $x_i = (x_{i1}, x_{i2}, \dots, x_{id})$ αντιστοιχεί στο σύνολο χαρακτηριστικών του i -οστού παραδείγματος εκπαίδευσης. Το περιθώριο του υπερεπιπέδου υπολογίζεται ως εξής:

$$d = \frac{2}{\|w\|}$$

Γενικά, ο ταξινομητής SVM είναι μια μέθοδος βελτιστοποίησης πολλαπλών κριτηρίων:

- Μεγιστοποιεί την απόσταση μεταξύ των διανυσμάτων υποστήριξης (*support vectors*) και ενός υπερεπιπέδου απόφασης. Τα διανύσματα υποστήριξης είναι τα παραδείγματα εκπαίδευσης που βρίσκονται πιο κοντά στο υπερεπίπεδο και καθορίζουν το περιθώριό του (*margin*). Η μέθοδος μεγιστοποιεί το περιθώριο, το οποίο αποτελεί μέτρο της γενικευτικής ικανότητας του ταξινομητή, καθώς ταξινομητές που παράγουν όρια απόφασης με μικρά περιθώρια είναι ευάλωτοι σε φαινόμενα υπερεκπαίδευσης (*model overfitting*).

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

$$\text{subject to } y_i(w \bullet x_i + b) \geq 1, \quad i = 1, 2, \dots, N$$

- Στα περισσότερα προβλήματα ταξινόμησης τα παραδείγματα εκπαίδευσης δεν είναι γραμμικά διαχωρίσιμα. Σ' αυτή την περίπτωση, τα SVM απεικονίζουν το



αρχικό σύνολο χαρακτηριστικών σε ένα σύνολο μεγαλύτερης διάστασης χρησιμοποιώντας μια συνάρτηση $\Phi(x)$.

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

$$\text{subject to } y_i(w \bullet \Phi(x_i) + b) \geq 1, \quad i = 1, 2, \dots, N$$

Το πρόβλημα βελτιστοποίησης επιλύεται με την χρήση πολλαπλασιαστών Lagrange και καταλήγει στο ακόλουθο δυικό πρόβλημα βελτιστοποίησης χωρίς περιορισμούς.

$$L_D = \sum_{i=1}^N \lambda_i - \frac{1}{2} \sum_{i,j} \lambda_i \lambda_j y_i y_j \Phi(x_i) \bullet \Phi(x_j)$$

Η συνάρτηση απόφασης για κάθε παράδειγμα ελέγχου x μπορεί να εκφραστεί ως εξής:

$$f(x) = \text{sign}\left(\sum_{i=1}^N \lambda_i y_i \Phi(x_i) \bullet \Phi(x) + b\right)$$

Το εσωτερικό γινόμενο στον χώρο $\Phi(x)$ καλείται συνάρτηση πυρήνας (*kernel function*).

Η επιτυχία της μεθόδου εξαρτάται από τη σωστή επιλογή της συνάρτησης πυρήνα. Οι πιο γνωστοί πυρήνες είναι οι ακόλουθοι:

- Γραμμικός (*linear*): $k(x_i, x_j) = x_i \bullet x_j$
- Πολυωνυμικός (*polynomial*): $k(x_i, x_j) = (ax_i \bullet x_j + r)^p, a > 0$
- Ακτινικής Συνάρτησης Βάσης (*radial basis function*):
 $k(x_i, x_j) = e^{-g\|x_i - x_j\|^2}$, όπου $g = 1/2\sigma^2$ και σ που εκφράζει το εύρος του πυρήνα.

- Πυρήνας χ^2 (*kernel χ^2*): $k(x_i, x_j) = \sum_i \frac{(x_i - y_j)^2}{x_i + y_j}$



- Επειδή το πρόβλημα της ταξινόμησης μπορεί να επιλυθεί χρησιμοποιώντας σύνθετες διακρίνουσες συναρτήσεις, τα SVM επιδιώκουν να ελαχιστοποιούν το μέγεθος της λύσης (το άθροισμα των βαρών των χαρακτηριστικών).
- Παρόλο που ο μετασχηματισμός των χαρακτηριστικών είναι μη γραμμικός, κάποια από τα παραδείγματα δεν ταξινομούνται σωστά. Τα SVM τείνουν να μειώνουν το πλήθος των εσφαλμένων ταξινομήσεων. Σ' αυτή την περίπτωση αναζητάνε ένα υπερεπίπεδο «χαλαρού» περιθωρίου, δηλ. μια επιφάνεια απόφασης η οποία διαχωρίζει τα δεδομένα εκπαίδευσης κάνοντας τα λιγότερα λάθη. Οι βοηθητικές μεταβλητές $\xi_i \geq 0$ (slack variables) αποτελούν μια εκτίμηση του σφάλματος του επιπέδου απόφασης για το i -οστό παράδειγμα εκπαίδευσης και η μεταβλητή C καθορίζει πόσο αυστηροί είμαστε με τα λάθη.

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

$$\text{subject to } \xi_i \geq 0 \text{ and } y_i(w \cdot \Phi(x_i) + b) \geq 1 - \xi_i, \quad i = 1, 2, \dots, N$$

Τα SVM χρησιμοποιούν διάφορους αλγόριθμους βελτιστοποίησης ώστε να ικανοποιούνται όλα τα κριτήρια ταυτόχρονα, παρόλα αυτά χρειάζεται να σταθμίσουμε κατάλληλα όλα τα παραπάνω κριτήρια ώστε να πετύχουμε την καλύτερη ταξινόμηση.

Για προβλήματα ταξινόμησης πολλών κατηγοριών υπάρχουν δύο προσεγγίσεις: (α) ένας-εναντίον-όλων (*one-against-all*) και (β) ένας-εναντίον-ένα (*one-against-one*). Στην πρώτη προσέγγιση για ένα σύνολο K κατηγοριών $C = \{C_1, C_2, \dots, C_K\}$ απαιτείται η εκπαίδευση K δυαδικών SVM. Κάθε SVM υπολογίζει ένα υπερεπίπεδο απόφασης το οποίο διαχωρίζει τα παραδείγματα της κατηγορίας i από τα παραδείγματα των υπολοίπων $K-1$ κατηγοριών. Ένα παράδειγμα ελέγχου x ανατίθεται στην κατηγορία C_i αν η έξοδος του ταξινομητή i είναι μεγαλύτερη από τις εξόδους των υπόλοιπων SVM. Στη δεύτερη προσέγγιση για ένα σύνολο K κατηγοριών $C = \{C_1, C_2, \dots, C_K\}$ απαιτείται η εκπαίδευση $K(K-1)/2$ δυαδικών SVM. Οι ταξινομητές υπολογίζουν ένα υπερεπίπεδο απόφασης το οποίο διαχωρίζει τα παραδείγματα της κατηγορίας i από τα παραδείγματα κάθε μίας από τις υπόλοιπες κατηγορίες. Κάθε παράδειγμα ελέγχου x ανατίθεται σε μια από τις κατηγορίες με βάση το αποτέλεσμα μιας «ψηφοφορίας»,



δηλ. ανατίθεται στην κατηγορία η οποία εμφανίζεται πιο πολλές φορές στην έξοδο των δυαδικών ταξινομητών.

2.2.6 Τεχνητά Νευρωνικά Δίκτυα (*Artificial Neural Networks - ANN*)

Τα τεχνητά νευρωνικά δίκτυα μιμούνται τα βιολογικά νευρωνικά δίκτυα αναθέτοντας τις λειτουργίες των νευρών σε ένα απλό στοιχείο (νευρώνας) το οποίο είναι ικανό μόνο να αθροίζει την είσοδό του και να κανονικοποιεί την έξοδό του. Οι νευρώνες είναι διασυνδεδεμένοι σε αυθαίρετα σύνθετα τεχνητά νευρωνικά δίκτυα. Για ταξινόμηση συνήθως χρησιμοποιούνται νευρωνικά δίκτυα εμπρόσθιας τροφοδότησης πολλών επιπέδων. Οι νευρώνες οργανώνονται σε επίπεδα: ένα επίπεδο εισόδου (το οποίο αντιστοιχεί στα χαρακτηριστικά), ένα ή περισσότερα κρυμμένα επίπεδα και ένα επίπεδο εξόδου (το οποίο αντιστοιχεί στις κατηγορίες). Στόχος του αλγορίθμου μάθησης είναι να καθορίσει τα βάρη των συνδέσεων μεταξύ των νευρώνων (τα οποία χρησιμοποιούνται για να υπολογίσουμε σταθμισμένα αθροίσματα σε κάθε νευρώνα) με στόχο να μειώσει το ποσοστό σφάλματος ταξινόμησης.

Για την ταξινόμηση ενός νέου παραδείγματος οι τιμές των χαρακτηριστικών εφαρμόζονται στις εισόδους στους νευρώνες εισόδου του ANN. Αυτές οι τιμές σταθμίζονται σύμφωνα με συνδέσεις μεταξύ των νευρώνων και τα σταθμισμένα αθροίσματά τους υπολογίζονται σε κάθε νευρώνα του επόμενου επιπέδου νευρώνων. Τα κανονικοποιημένα αποτελέσματα στους νευρώνες εξόδου καθορίζουν το αποτέλεσμα της ταξινόμησης (κατηγορία).

2.3. Συνδυαστικές Τεχνικές Μεθόδων Μηχανικής Μάθησης

Η αρχή του ελάχιστου μήκους περιγραφής, όπως αναφέρθηκε και προηγουμένως, απαιτεί από τον αλγόριθμο μάθησης την εύρεση ενός συμβιβασμού μεταξύ της πολυπλοκότητας και της ακρίβειας του μοντέλου. Μια υπόθεση (μοντέλο) πρέπει να είναι όσο πιο απλή γίνεται και να παρέχει μια καλή μοντελοποίηση των δεδομένων. Από την άλλη πλευρά, η αρχή των πολλαπλών επεξηγήσεων απαιτεί από τον αλγόριθμο πρόβλεψης να χρησιμοποιεί διαφορετικές υποθέσεις για τα δεδομένα, καθώς τα καλύτερα αποτελέσματα επιτυγχάνονται με κατάλληλο συνδυασμό όλων



των παραγόμενων υποθέσεων. Ο αλγόριθμος μάθησης πρέπει να εξετάσει και τα δύο κριτήρια και να παράγει διαφορετικές υποθέσεις (μοντέλα) κατά τη φάση της εκπαίδευσης. Υπάρχουν δύο παραλλαγές:

- Οι υποθέσεις να παράγονται από διαφορετικούς αλγορίθμους μάθησης χρησιμοποιώντας τα ίδια δεδομένα. Σε ένα πρόβλημα ταξινόμησης θα μπορούσαμε για παράδειγμα, να εκπαιδεύσουμε έναν ταξινομητή Naive Bayes και ένα πολυεπίπεδο νευρωνικό δίκτυο, να αποθηκεύσουμε τα δεδομένα εκπαίδευσης για έναν αλγόριθμο k-κοντινότερων γειτόνων και να υπολογίσουμε μια συνάρτηση διακρίνουσας χρησιμοποιώντας τη μέθοδο SVM.
- Οι υποθέσεις να παράγονται από τον ίδιο αλγόριθμο είτε εκτελώντας τον με διαφορετικές ρυθμίσεις των παραμέτρων ή μεταβάλλοντας το σύνολο εκπαίδευσης του. Σ' αυτή την ομάδα μεθόδων ανήκουν και οι δημοφιλείς μέθοδοι *bagging* και *boosting* καθώς και τα τυχαία δάση.

2.3.1 Τρόποι Συνδυασμού των Προβλέψεων Διαφορετικών Υποθέσεων

Όταν κατασκευάζονται πολλαπλές υποθέσεις (μοντέλα), πρέπει να συνδυαστούν με βάση την αρχή των πολλαπλών επεξηγήσεων σε μια τελική υπόθεση (ταξινομητή). Οι προβλέψεις διαφορετικών ταξινομητών μπορούν να συνδυαστούν με τους παρακάτω τρόπους:

Ψηφοφορία (Voting): κάθε ταξινομητής δίνει μια ψήφο για μία συγκεκριμένη κατηγορία. Κάθε παράδειγμα ανατίθεται στην κατηγορία η οποία έχει συγκεντρώσει τις περισσότερες ψήφους.

Ψηφοφορία με βάρη (Weighted voting): κάθε ταξινομητής συνεισφέρει σύμφωνα με μια κατανομή πιθανότητας κατηγορίας. Οι πιθανότητες κατηγορίας ενεργούν ως βάρη. Κάθε παράδειγμα ανατίθεται στην κατηγορία η οποία έχει συγκεντρώσει το μεγαλύτερο άθροισμα βαρών.

Σταθμισμένη ψηφοφορία με βάση την αξιοπιστία (Reliability-weighted voting): η αξιοπιστία μιας συγκεκριμένης πρόβλεψης εκτιμάται με κάποια μέθοδο (π.χ. με transduction) ή υπολογίζοντας τη συνολική αξιοπιστία του ταξινομητή (π.χ. με cross-validation). Οι εκτιμήσεις αξιοπιστίας χρησιμοποιούνται μετέπειτα ως βάρη σε μια σταθμισμένη ψηφοφορία.



Τοπικά σταθμισμένη ψηφοφορία (Locally weighted voting): Για κάθε νέο παράδειγμα το οποίο πρέπει να ταξινομηθεί, επιλέγονται k - κοντινότεροι γείτονες από το σύνολο εκπαίδευσης. Για κάθε ταξινομητή, υπολογίζεται μια μέση (τοπική) απόδοση χρησιμοποιώντας τα επιλεγμένα k παραδείγματα για έλεγχο. Οι ταξινομήσεις για το κάθε νέο παράδειγμα είναι είτε σταθμισμένες με την τοπική απόδοσή τους ή χρησιμοποιείται μόνο ο καλύτερος τοπικός ταξινομητής.

Στοιβά (Stacking): αυτή η προσέγγιση απαιτεί ένα ξεχωριστό σύνολο επικύρωσης (validation set). Οι ταξινομήσεις των παραδειγμάτων επικύρωσης χρησιμοποιούνται ως χαρακτηριστικά για μετα-εκπαίδευση. Η συνάρτηση συνδυασμού δεν δίνεται εκ των προτέρων αλλά κατασκευάζεται από τα δεδομένα.

Μπεϋζιανή ψηφοφορία (Bayesian voting): κάθε ταξινομητής συνεισφέρει σύμφωνα με μια κατανομή πιθανότητας κατηγορίας. Χρησιμοποιείται η εξίσωση της μεθόδου Naive Bayes για τον υπολογισμό πιθανοτήτων για όλες τις κατηγορίες. Στην εξίσωση κάθε ταξινομητής αντιμετωπίζεται ως ένα διακριτό χαρακτηριστικό (η υπό συνθήκη πιθανότητα κατηγορίας δεδομένου του ταξινομητή ισούται με την πιθανότητα που παρέχει ο ταξινομητής).

Δυναμική επιλογή ταξινομητών: Όπως και στην περίπτωση της τοπικά σταθμισμένης ψηφοφορίας, για κάθε νέο παράδειγμα που πρέπει να ταξινομηθεί επιλέγονται οι k - κοντινότεροι γείτονες από το σύνολο εκπαίδευσης. Κάθε ταξινομητής συνεισφέρει σύμφωνα με μια κατανομή πιθανότητας κατηγορίας για κάθε νέο παράδειγμα και για τους k - κοντινότερους γείτονες. Ένα μέτρο χρησιμοποιείται για τον υπολογισμό των αποστάσεων μεταξύ των κατανομών πιθανοτήτων καθενός από τους γείτονες και του νέου παραδείγματος. Για την εκτίμηση της απόδοσης του ταξινομητή χρησιμοποιούμε τους γείτονες που έχουν παρόμοιες κατανομές πιθανότητας. Για την ταξινόμηση νέων παραδειγμάτων είτε χρησιμοποιούνται οι καλύτεροι τοπικοί ταξινομητές είτε χρησιμοποιούνται όλοι με κάποιο βάρος ανάλογα με την τοπική απόδοσή τους.

2.4. Κατασκευή Συνδυαστικών Ταξινομητών (Ensemble Classifiers)

Στην προηγούμενη ενότητα αναφέρθηκαν περιληπτικά τρόποι με τους οποίους οι αλγόριθμοι μάθησης μπορούν να συνδυαστούν παράλληλα. Σ' αυτήν την περίπτωση οι αλγόριθμοι εκτελούνται ανεξάρτητα και μόνο οι προβλέψεις τους συνδυάζονται με κάποιον τρόπο. Οι αλγόριθμοι μάθησης μπορούν να κατασκευαστούν όμως και



σειριακά. Ο πρώτος αλγόριθμος μαθαίνει την συνάρτηση στόχου, ο δεύτερος μαθαίνει να προβλέπει τα σφάλματα που πρώτου κ.ο.κ. Λόγω του ότι υπάρχει πιθανότητα υπερεκπαίδευσης (*overfitting*), συνήθως το πλήθος των υποθέσεων που κατασκευάζονται δεν είναι πολύ μεγάλο. Παρακάτω περιγράφονται κάποιοι από τους πιο γνωστούς συνδυαστικούς αλγόριθμους μάθησης. Το κοινό χαρακτηριστικό αυτών των αλγόριθμων είναι ότι παράγουν σύνολα ταξινομητών βάσης και διασφαλίζουν την ποικιλομορφία τους παρέχοντας τους διαφορετικά σύνολα παραδειγμάτων εκπαίδευσης.

2.4.1 Η Μέθοδος Bagging

Ο όρος *bagging* είναι συντομογραφία του *bootstrap aggregating* [15, 16]. Bootstrap είναι η μέθοδος που χρησιμοποιείται για την αναπαραγωγή των παραδειγμάτων εκπαίδευσης στην περίπτωση όπου το σύνολο εκπαίδευσης είναι μικρό. Δεδομένου ότι το συνολικό πλήθος των παραδειγμάτων εκπαίδευσης είναι n , κάθε νέο σύνολο εκπαίδευσης παράγεται επιλέγοντας n φορές ένα παράδειγμα από το αρχικό σύνολο. Η επιλογή γίνεται με επανατοποθέτηση σύμφωνα με την ομοιόμορφη κατανομή. Με αυτόν τον τρόπο κάποια παραδείγματα εκπαίδευσης μπορεί να εμφανίζονται περισσότερες φορές και άλλα να μην εμφανίζονται καθόλου στο σύνολο εκπαίδευσης (περίπου το 36.8% των παραδειγμάτων) [17, 18]. Κάθε νέο τέτοιο σύνολο χρησιμοποιείται ως είσοδος στον αλγόριθμο μάθησης. Με αυτόν τον τρόπο δημιουργείται μια σειρά διαφορετικών μοντέλων τα οποία χρησιμοποιούνται για την πρόβλεψη της τιμής της εξαρτημένης μεταβλητής συνδυάζοντας τις επί μέρους προβλέψεις όλων των μοντέλων. Η μέθοδος είναι ιδιαίτερα αποδοτική όταν χρησιμοποιούνται ασταθείς ταξινομητές βάσης με μεγάλη διακύμανση (ταξινομητές οι οποίοι μεταβάλλονται αρκετά σε μικρές διαταραχές του συνόλου εκπαίδευσης) όπως τα δέντρα απόφασης. Η μέθοδος δεν είναι επιρρεπής σε φαινόμενα υπερεκπαίδευσης καθώς αυξάνεται το πλήθος των παραγόμενων υποθέσεων.

2.4.2 Η Μέθοδος Boosting

Βασική ιδέα της μεθόδου αποτελεί η στάθμιση των παραδειγμάτων εκπαίδευσης σύμφωνα με τη δυσκολία της ταξινόμησης τους [15, 16, 17, 19]. Περιλαμβάνει μια



επαναληπτική διαδικασία η οποία προσαρμόζει την κατανομή των παραδειγμάτων εκπαίδευσης με τέτοιο τρόπο ώστε οι ταξινομητές βάσης να εστιάζουν στα παραδείγματα που ταξινομούνται δύσκολα. Τα βάρη μπορούν να χρησιμοποιηθούν είτε ως κατανομή δειγματοληψίας για να καθορίσουν το σύνολο εκπαίδευσης είτε να χρησιμοποιηθούν από τον ταξινομητή βάσης για την εκπαίδευση ενός μοντέλου προσανατολισμένου σε παραδείγματα με μεγάλο βάρος.

Συγκεκριμένα, στην πρώτη επανάληψη ο αλγόριθμος μάθησης κατασκευάζει ένα μοντέλο από όλα τα παραδείγματα εκπαίδευσης καθώς έχουν ίσο βάρος. Το μοντέλο αυτό προβλέπει την τιμή της εξαρτημένης μεταβλητής (κατηγορία) για κάθε παράδειγμα εκπαίδευσης. Στη συνέχεια όλα τα βάρη προσαρμόζονται. Για τα παραδείγματα που ταξινομήθηκαν σωστά τα βάρη μειώνονται ενώ για τα παραδείγματα με εσφαλμένες προβλέψεις τα βάρη αυξάνονται. Έστω e το κανονικοποιημένο σφάλμα ενός μοντέλου για ένα συγκεκριμένο παράδειγμα. Το βάρος του συγκεκριμένου παραδείγματος αναπροσαρμόζεται πολλαπλασιαζόμενο με $e/(1-e)$. Μικρότερα λάθη οδηγούν σε μικρότερα βάρη. Οι επόμενες επαναλήψεις εστιάζουν στα παραδείγματα που ταξινομούνται δύσκολα. Η διαδικασία επαναλαμβάνεται έως ότου το ολικό σφάλμα γίνει αρκετά μικρό ή αρκετά μεγάλο, το τελευταίο σημαίνει ότι τα υπολειπόμενα δύσκολα παραδείγματα δεν μπορούν να ταξινομηθούν. Αυτά τα μοντέλα απορρίπτονται, όπως επίσης και εκείνα για τα οποία το ολικό σφάλμα είναι μηδέν, το οποίο σημαίνει ότι τείνουν να εμφανίζουν φαινόμενα υπερεκπαίδευσης. Τα υπόλοιπα μοντέλα χρησιμοποιούνται για τις τελικές προβλέψεις. Κάθε μοντέλο έχει ως βάρος την απόδοσή του, γι' αυτό το λόγο οι ακριβείς ταξινομητές έχουν μεγαλύτερα βάρη. Η μέθοδος boosting μπορεί να χρησιμοποιηθεί και με ευσταθείς αλγόριθμους οι οποίοι παρουσιάζουν μικρή διακύμανση και γενικά εμφανίζει καλύτερη απόδοση από μέθοδο bagging. Παρακάτω περιγράφεται μια υλοποίηση του αλγόριθμου Boosting η οποία καλείται AdaBoost (*Adaptive Boosting*):

Έστω ένα σύνολο N παραδειγμάτων εκπαίδευσης $D = \{(x_j, y_j) \mid j=1,2,\dots,N\}$. Αρχικοποιούμε τα βάρη των όλων παραδειγμάτων w_j ($w_j=1/N$). Ο αλγόριθμος AdaBoost σε κάθε επανάληψη, δημιουργεί ένα σύνολο εκπαίδευσης D_i με δειγματοληψία (με επανατοποθέτηση) από το αρχικό σύνολο D σύμφωνα με τα βάρη



w και εκπαιδεύει έναν ταξινομητή C_i πάνω στο σύνολο D_i . Εφαρμόζει τον ταξινομητή C_i σε όλα τα παραδείγματα του συνόλου D και υπολογίζει την πιθανότητα σφάλματος (*error rate*) του ταξινομητή ως εξής:

$$\varepsilon_i = \frac{1}{N} \left[\sum_{j=1}^N w_j I(C_i(x_j) \neq y_j) \right]$$

όπου $I(p)=1$ αν το κατηγορήμα p είναι αληθές, αλλιώς είναι 0. Αν το error rate του ταξινομητή είναι μικρότερο από 50%, τότε όλα τα βάρη τίθενται στην αρχική τους τιμή, αλλιώς τα βάρη ενημερώνονται με τρόπο ώστε να «τιμωρούνται» οι ταξινομητές που εμφανίζουν χαμηλή ακρίβεια.

$$w_{j+1} = w_j \cdot \exp(a_i \cdot I(y_j \neq C_i(x_j)))$$

όπου $a_i = \ln((1 - \varepsilon_i) / \varepsilon_i)$. Η παράμετρος a_i δείχνει πόσο σημαντικός είναι ένας ταξινομητής. Παρατηρούμε ότι το a_i παίρνει μια μεγάλη θετική τιμή όταν η πιθανότητα σφάλματος είναι κοντά στο 0 και μια μεγάλη αρνητική τιμή όταν η πιθανότητα σφάλματος είναι κοντά στο 1.

Η τελική πρόβλεψη του συνδυαστικού ταξινομητή υπολογίζεται από το σταθμισμένο μέσο όρο των προβλέψεων όλων των ταξινομητών βάσης.

$$C^*(x) = \text{sign} \left[\sum_{j=1}^T \alpha_j I(C_j(x) = y) \right]$$

Αν το error rate ενός ταξινομητή βάσης είναι μικρότερο από 50% τότε μπορεί να γραφεί ως $\varepsilon_i = 0.5 - \gamma_i$, όπου γ_i είναι ένα μέτρο για το πόσο καλύτερος είναι ο ταξινομητής από την τυχαία πρόβλεψη. Το όριο του σφάλματος εκπαίδευσης ορίζεται ως εξής:

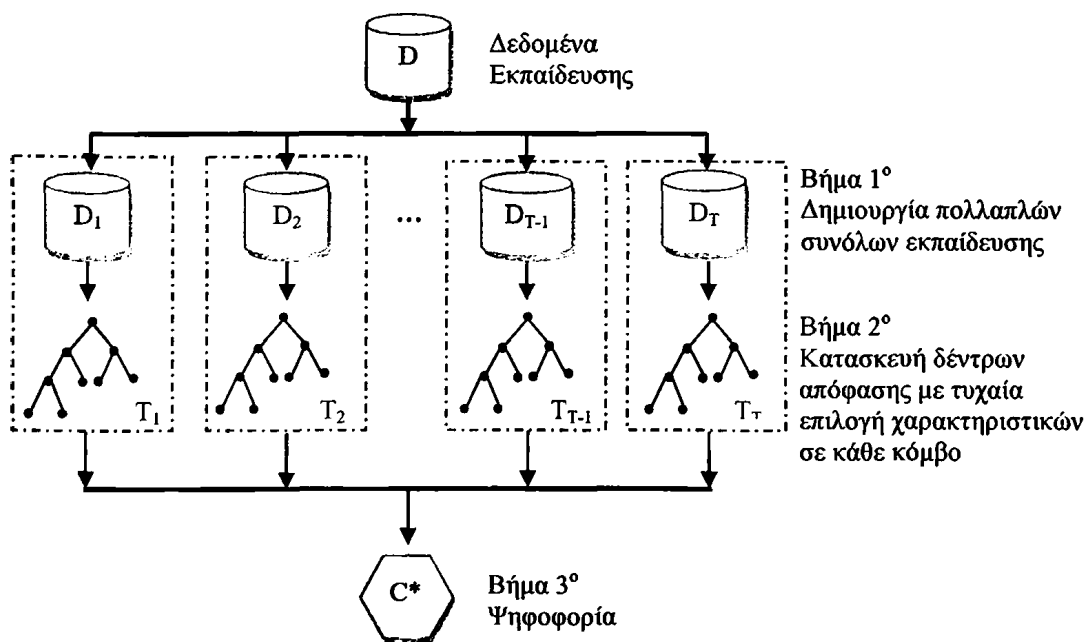
$$e_{ensemble} \leq \prod_i \sqrt{1 - 4\gamma_i^2} \leq \exp\left(-2 \sum_i \gamma_i^2\right)$$



Αν $\gamma_i < \gamma^*$ για όλα τα i , τότε η τεχνική Boosting τείνει να μειώνει εκθετικά το σφάλμα εκπαίδευσης, το οποίο οδηγεί σε γρήγορη σύγκλιση του αλγορίθμου. Παρόλα αυτά όμως είναι επιρρεπής σε φαινόμενα υπερεκπαίδευσης λόγω του ότι εστιάζει σε παραδείγματα που δεν ταξινομούνται σωστά.

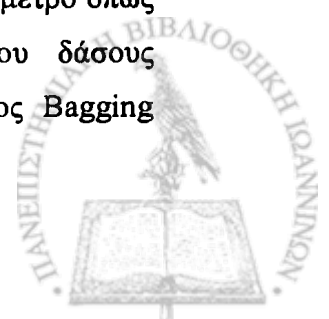
2.5. Τυχαία Δάση (Random Forests)

Τα τυχαία δάση [20] αποτελούν μια ειδική κατηγορία των συνδυαστικών μεθόδων ταξινόμησης η οποία χρησιμοποιεί για ταξινομητές δέντρα απόφασης [21].



Σχήμα 2.5 Η Διαδικασία Κατασκευής των Random Forests

Για την κατασκευή ενός δέντρου απόφασης ανατίθεται αρχικά στη ρίζα του το σύνολο των δειγμάτων εκπαίδευσης. Κάθε ενδιάμεσος κόμβος περιέχει υποσύνολο των δειγμάτων το οποίο μέσω της εφαρμογής ενός κατάλληλου ελέγχου διαχωρίζεται σε δυο ή περισσότερα μικρότερα υποσύνολα (παιδιά) στο επόμενο επίπεδο. Ο έλεγχος συνήθως αφορά ένα υποσύνολο των χαρακτηριστικών των δειγμάτων εκπαίδευσης. Η επιλογή του καλύτερου διαχωρισμού γίνεται σύμφωνα με ένα κατάλληλο μέτρο όπως π.χ. Gini index, εντροπία, misclassification error. Τα δέντρα του δάσους αναπτύσσονται στο μέγιστο μέγεθός τους, χωρίς κλάδεμα. Η μέθοδος Bagging



χρησιμοποιώντας για ταξινομητές δέντρα απόφασης αποτελεί μια ειδική κατηγορία των Random Forests (Σχήμα 2.3). Σ' αυτή την περίπτωση η τυχαιότητα ενσωματώνεται στο μοντέλο και μέσω της τυχαίας επιλογής N παραδειγμάτων εκπαίδευσης, με επανατοποθέτηση, από το αρχικό σύνολο εκπαίδευσης.

Η διαδικασία ταξινόμησης «άγνωστων» παραδειγμάτων πραγματοποιείται μέσω της διάσχισης των δέντρων του δάσους ξεκινώντας από τη ρίζα και καταλήγοντας σε ένα από τα φύλλα του δέντρου και στη συνέχεια συνδυάζοντας τις προβλέψεις των ταξινομητών σύμφωνα με ένα πλειοψηφικό σύστημα ψηφοφορίας (*majority voting scheme*). Κάθε παράδειγμα ανατίθεται στην κατηγορία με τη μεγαλύτερη συχνότητα.

2.5.1 Σύγκλιση του σφάλματος γενίκευσης

Θεωρητικά έχει αποδειχθεί ότι το σφάλμα γενίκευσης (*generalization error*) των τυχαίων δασών συγκλίνει στην ακόλουθη έκφραση όταν ο αριθμός των δέντρων είναι αρκετά μεγάλος [16].

$$\text{Generalization_error} \leq \frac{\bar{\rho}(1-s^2)}{s^2}$$

όπου $\bar{\rho}$ είναι η μέση συσχέτιση μεταξύ των ταξινομητών και s μια ποσότητα η οποία υπολογίζει τη «δύναμη» των ταξινομητών. Η δύναμη (*strength*) ενός συνόλου ταξινομητών αναφέρεται στη μέση απόδοση των ταξινομητών $M(X, Y)$. Η απόδοση υπολογίζεται πιθανοτικά ως το περιθώριο του ταξινομητή $M(X, Y) = P(\hat{Y}_g = Y) - \max_{Z \neq Y} P(\hat{Y}_g = Z)$, όπου \hat{Y}_g είναι η προβλεπόμενη κατηγορία του X σύμφωνα με έναν ταξινομητή ο οποίος έχει κατασκευαστεί από κάποιο τυχαίο διάνυσμα θ . Όσο πιο μεγάλο είναι το περιθώριο, τόσο πιο πιθανό είναι ο ταξινομητής να προβλέψει σωστά την κατηγορία του παραδείγματος X . Δηλ. όσο μεγαλώνει η συσχέτιση μεταξύ των δέντρων ή αντίστοιχα όσο η δύναμη του δάσους μειώνεται, το όριο του σφάλματος γενίκευσης τείνει να αυξάνεται. Η τυχαιότητα βοηθάει στη μείωση της συσχέτισης μεταξύ των δέντρων απόφασης και κατ' επέκταση στη βελτίωση του σφάλματος γενίκευσης του συνδυαστικού μοντέλου. Η παραπάνω σχέση εξηγεί το λόγο για τον οποίο τα Random Forests δεν εμφανίζουν φαινόμενα



υπερεκπαίδευσης κατά την προσθήκη περισσότερων δέντρων αλλά αντιθέτως τείνουν να περιορίζουν το σφάλμα γενίκευσης.

2.5.2 Πλεονεκτήματα των Τυχαίων Δασών

Παρακάτω αναφέρονται συνοπτικά οι ιδιότητες και τα κύρια πλεονεκτήματα των Random Forests:

- Μπορούν να εκπαιδευτούν σε σύνολα δεδομένων υψηλής διάστασης όπως είναι τα κείμενα και οι εικόνες, χωρίς να εμφανίσουν σημαντικό βαθμό υπερεκπαίδευσης.
- Εξαιτίας του μεγάλου πλήθους δέντρων στο δάσος, το σφάλμα γενίκευσης είναι περιορισμένο. Αυτό έχει ως αποτέλεσμα τη μη εμφάνιση φαινομένων υπερεκπαίδευσης.
- Μη επαναληπτική διαδικασία εκπαίδευσης, ο αλγόριθμος ολοκληρώνεται σε σταθερό αριθμό βημάτων.
- Η τυχαία επιλογή ενός υποσυνόλου των χαρακτηριστικών για τη διαμέριση των παραδειγμάτων κάθε ενδιάμεσου κόμβου ελαττώνει τη συσχέτιση ανάμεσα στα δέντρα και διατηρεί την πόλωση (*bias*) σε χαμηλά επίπεδα καθώς τα δέντρα αναπτύσσονται χωρίς κλάδεμα. Χρησιμοποιώντας ένα σύνολο δέντρων απόφασης μειώνεται και η διακύμανση (*variance*).
- Η διάσχιση ενός δέντρου από ένα παράδειγμα ξεκινώντας από τη ρίζα και καταλήγοντας σε έναν από τους τερματικούς κόμβους γίνεται σε λογαριθμικό ως προς το πλήθος των φύλλων του.
- Παρουσιάζουν ανεκτικότητα ως προς το θόρυβο και αριθμητικών σφαλμάτων στα δεδομένα εκπαίδευσης (π.χ. απόκρυψη μέρους του αντικειμένου, ελλιπή δεδομένα).
- Για την επαγωγή κάθε δέντρου περίπου το 1/3 των παραδειγμάτων δεν επιλέγεται για εκπαίδευση. Αυτά τα παραδείγματα καλούνται Out-of-Bag παραδείγματα και μπορούν να χρησιμοποιηθούν για την εκτίμηση της πιθανότητας σφάλματος, εξαλείφοντας την ανάγκη ύπαρξης ενός συνόλου ελέγχου ή εφαρμογής της τεχνικής cross-validation.
- Παράγει μια εσωτερική αμερόληπτη εκτίμηση του σφάλματος γενίκευσης καθώς εξελίσσεται η διαδικασία κατασκευής του δέντρου.



- Υπάρχει η δυνατότητα παράλληλης επαγωγής των δέντρων, σε αντίθεση με την μέθοδο Boosting.
- Αναζητά το καλύτερο διαχωρισμό σε ένα μικρό υποσύνολο των χαρακτηριστικών και δεν κάνει εξαντλητική αναζήτηση όπως ο αλγόριθμος Boosting.
- Μπορεί να χρησιμοποιηθεί για ομαδοποίηση.
- Επιτρέπει τη δημιουργία παραλλαγών της βασικής τεχνικής ως προς την κατασκευή του μοντέλου ταξινόμησης π.χ. χρήση διαφορετικών τεχνικών διαμέρισης των παραδειγμάτων των ενδιάμεσων κόμβων.

Τα τυχαία δάση παρουσιάζουν όμως και κάποια σημαντικά μειονεκτήματα ως προς την εφαρμογή τους τα οποία αναφέρονται συνοπτικά παρακάτω:

- Υψηλό υπολογιστικό κόστος.
- Υπάρχει σημαντικό πλήθος ελεύθερων παραμέτρων τις οποίες πρέπει να προσδιορίσει ο χρήστης π.χ. πλήθος δέντρων, βαθμός κόμβων, πλήθος παραδειγμάτων εκπαίδευσης, συνθήκη τερματισμού διαμέρισης των κόμβων.
- Για την επέκταση ενός μοντέλου με στόχο την εισαγωγή μιας ακόμα κατηγορίας απαιτείται η κατασκευή του μοντέλου από την αρχή.
- Κάθε νέο παράδειγμα πρέπει να διασχίσει όλα τα δέντρα του δάσους για την εκτίμηση της κατηγορίας του.



ΚΕΦΑΛΑΙΟ 3. ΤΥΧΑΙΑ ΔΑΣΗ ΓΙΑ ΤΑΞΙΝΟΜΗΣΗ ΕΙΚΟΝΩΝ

3.1 Μέθοδοι Ταξινόμησης Εικόνων με Βάση το Περιεχόμενο

3.2 Ανάλυση της Προτεινόμενης Μεθόδου

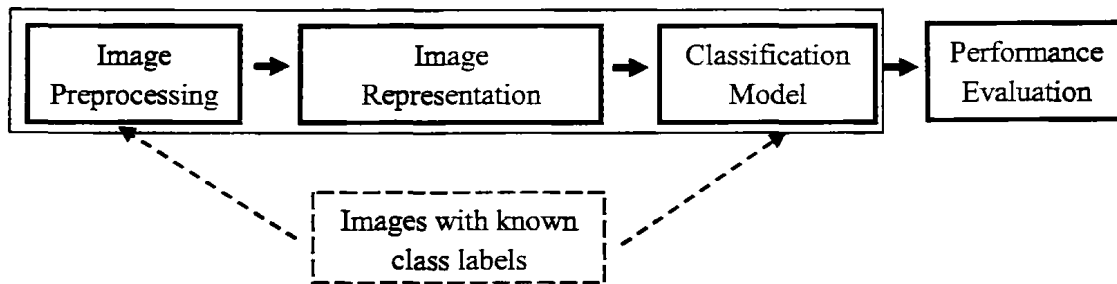
3.1. Μέθοδοι Ταξινόμησης Εικόνων με Βάση το Περιεχόμενο

3.1.1 Γενικά

Όπως έχει αναφερθεί και στο πρώτο κεφάλαιο, οι μέθοδοι ταξινόμησης εικόνας που έχουν προταθεί διακρίνονται γενικά σε δύο κατηγορίες με βάση το είδος της πληροφορίας που χρησιμοποιούν για την επίλυση του προβλήματος. Υπάρχουν μέθοδοι οι οποίες χρησιμοποιούν λεκτικές περιγραφές ενδεικτικές του περιεχομένου των εικόνων (*text-based image classification*) [24] και μέθοδοι οι οποίες χρησιμοποιούν το πραγματικό περιεχόμενο της εικόνας (*content-based classification*) και όχι κάποια μεταδεδομένα που σχετίζονται με την εικόνα. Η παρούσα εργασία εξετάζει αποκλειστικά το πρόβλημα της ταξινόμησης εικόνας με βάση το πραγματικό της περιεχόμενο.

Μια ολοκληρωμένη μέθοδος ταξινόμησης εικόνας με βάση το περιεχόμενο αποτελείται από 4 βασικά στάδια όπως φαίνεται στο Σχήμα 3.1: 1) το στάδιο προεπεξεργασίας των εικόνων, 2) το στάδιο διανυσματικής αναπαράστασης των εικόνων, 3) το στάδιο επιλογής της μεθόδου ταξινόμησης και κατασκευής του μοντέλου ταξινόμησης και 4) το στάδιο αξιολόγησης του ταξινομητή. Στη συνέχεια περιγράφονται αναλυτικά τα παραπάνω στάδια.





Σχήμα 3.1 Διαδικασία Ταξινόμησης με Βάση το Περιεχόμενο των Εικόνων

1) Στάδιο Προεπεξεργασίας (Preprocessing): Το στάδιο αυτό αφορά την προεπεξεργασία των εικόνων. Περιλαμβάνει το φιλτράρισμα, την κανονικοποίηση, την κατάτμηση της εικόνας σε περιοχές, αντικείμενα ή σχέσεις, και την αναγνώριση του αντικειμένου στην εικόνα.

2) Στάδιο Διανυσματικής Αναπαράστασης Εικόνας (Image Representation): Για την εφαρμογή οποιουδήποτε αλγόριθμου ταξινόμησης, είναι απαραίτητη η μετατροπή των δεδομένων (εικόνων) σε μια πιο κατάλληλη μορφή, συνήθως διανυσματική. Ένα μεγάλο μέρος των ερευνητικών εργασιών έχει αφιερωθεί στην εύρεση κατάλληλων χαρακτηριστικών για την αναπαράσταση των εικόνων. Η διαδικασία εξαγωγής χαρακτηριστικών περιλαμβάνει την παραγωγή ενός συνόλου χαρακτηριστικών με το οποίο πετυχαίνουμε την αναπαράσταση της εικόνας. Η συγκεκριμένη διαδικασία είναι ιδιαίτερα σημαντική καθώς τα χαρακτηριστικά που θα διατεθούν (στο επόμενο στάδιο) στον αλγόριθμο εκπαίδευσης επηρεάζουν άμεσα την αποτελεσματικότητα του μοντέλου ταξινόμησης.

Γενικά, τα χαρακτηριστικά μπορούν να οριστούν ως μια συνάρτηση μιας ή περισσότερων μετρήσεων οι οποίες καθορίζουν ποσοτικά ορισμένα σημαντικά χαρακτηριστικά της εικόνας. Για να είναι ένα χαρακτηριστικό κατάλληλο θα πρέπει να παρουσιάζει τις ακόλουθες ιδιότητες:

- Να έχει μικρό υπολογιστικό κόστος, το οποίο είναι υψίστης σημασίας λόγω της μεγάλης διάστασης των δεδομένων.
- Να έχει μεγάλη διακριτική ικανότητα.



- Να είναι όσο γίνεται πιο αμετάβλητο στο θόρυβο ή σε ένα συγκεκριμένο σύνολο μετασχηματισμών.

Τα χαρακτηριστικά διακρίνονται αδρά σε δύο κατηγορίες: στα χαρακτηριστικά χαμηλού επιπέδου (*low-level features*) τα οποία μπορούν να εξαχθούν απευθείας από τις εικόνες και στα χαρακτηριστικά υψηλού επιπέδου (*high-level features*) η εξαγωγή των οποίων βασίζεται στα χαρακτηριστικά χαμηλού επιπέδου. Τα χαρακτηριστικά χαμηλού επιπέδου έχουν μικρή σχέση με το σημασιολογικό περιεχόμενο της εικόνας αλλά είναι πιο εύκολο να υπολογιστούν. Χαρακτηριστικά όπως το σχήμα, η υφή, το χρώμα κ.α. μπορούν να χρησιμοποιηθούν για να περιγράψουν το περιεχόμενο μιας εικόνας.

Χρώμα: Το χαρακτηριστικό που χρησιμοποιείται συνήθως σε μεθόδους ταξινόμησης εικόνας είναι το χρώμα. Μια ψηφιακή εικόνα αποτελείται από διακριτά στοιχεία τα οποία καλούνται εικονοστοιχεία ή pixels. Κάθε εικονοστοιχείο σχετίζεται με έναν (αν η εικόνα είναι grayscale) ή περισσότερους αριθμούς (τρεις ή τέσσερις αν η εικόνα είναι έγχρωμη), οι οποίοι κυμαίνονται συνήθως από 0 έως 255 και αναπαριστούν τη φωτεινότητα μιας σχετικά μικρής περιοχής στην εικόνα. Γενικά, το χρώμα μιας εικόνας αναπαρίσταται με τη βοήθεια κάποιου χρωματικού μοντέλου το οποίο προσδιορίζεται συνήθως μέσω ενός τρισδιάστατου συστήματος συντεταγμένων και ενός υποχώρου εντός του συστήματος όπου κάθε χρώμα αναπαρίσταται από ένα μοναδικό σημείο. Το πιο συχνά χρησιμοποιούμενο χρωματικό μοντέλο είναι το RGB (κόκκινο, πράσινο, μπλε). Συνεπώς, το χρωματικό περιεχόμενο μιας εικόνας χαρακτηρίζεται μέσω των τριών καναλιών ενός χρωματικού μοντέλου.

Ο πιο απλός τρόπος για να αναπαραστήσουμε την εμφάνιση μιας εικόνας είναι μέσω της bitmap αναπαράστασής της. Ωστόσο ο συγκεκριμένος τρόπος δεν είναι αποδοτικός λόγω των μεγάλων απαιτήσεων σε μνήμη. Για παράδειγμα, αν μια έγχρωμη εικόνα έχει διάσταση $h \times w$, το διάνυσμα αναπαράστασής της θα έχει διάσταση $3 \times h \times w$. Ένας τρόπος για να μειώσουμε τη διάσταση είναι να υπολογίσουμε στατιστικά μεγέθη, όπως η μέση τιμή και η τυπική απόκλιση, για κάθε χρώμα και στη συνέχεια να αναπαραστήσουμε την εικόνα με ένα διάνυσμα που θα



περιλαμβάνει για κάθε χαρακτηριστικό τόσες τιμές όσα και τα μεγέθη που χρησιμοποιούμε.

Τα ιστογράμματα χρώματος είναι μια εξίσου απλή και χαμηλού επιπέδου μέθοδος αναπαράστασης του χρωματικού περιεχομένου μιας εικόνας, η οποία είναι αρκετά αποδοτική στην πράξη. Για μια εικόνα τριών καναλιών, έχουμε τρία διαφορετικά ιστογράμματα τα οποία περιγράφουν τη σχετική συχνότητα σε διαφορετικούς κάδους (*bins*), σε μια προσπάθεια να απεικονιστεί πιο αδρά το περιεχόμενο της εικόνας και να μειωθεί η διάσταση των δεδομένων. Το ιστόγραμμα H μιας εικόνας περιγράφει την κατανομή των χρωμάτων στην εικόνα και ορίζεται ως $H = \{h(1), h(2), \dots, h(i), \dots, h(N)\}$, όπου i ένα χρώμα του ιστογράμματος, $h(i)$ το πλήθος αυτού του χρώματος στην εικόνα και N το πλήθος των κάδων του ιστογράμματος. Για να μπορούμε να συγκρίνουμε εικόνες διαφορετικών διαστάσεων πρέπει τα ιστογράμματα χρώματος να είναι κανονικοποιημένα. Οι ροπές χρώματος πρώτης (*mean*), δεύτερης (*variance*) και τρίτης (*skewness*) τάξης έχει αποδειχθεί ότι παρέχουν έναν αποτελεσματικό και αποδοτικό τρόπο αναπαράστασης της κατανομής χρώματος μιας εικόνας. Επειδή χρησιμοποιούν μόνο 9 αριθμούς (3 ροπές για κάθε συστατικό χρώματος) για την αναπαράσταση του χρωματικού περιεχομένου κάθε εικόνας, οι ροπές αποτελούν μια πολύ συμπαγή αναπαράσταση συγκριτικά με άλλα χαρακτηριστικά χρώματος. Η συνάρτηση ομοιότητας που χρησιμοποιείται για ταξινόμηση είναι το σταθμισμένο άθροισμα των απόλυτων διαφορών μεταξύ των κατάλληλων ροπών.

Γενικά, αν H και G αναπαριστούν δύο ιστογράμματα χρώματος, η ομοιότητά τους δίνεται μέσω της τομής τους η οποία υπολογίζεται ως εξής:

$$I(H, G) = \sum_{i=1}^N \min(H_i, G_i)$$

Τα ιστογράμματα δεν περιλαμβάνουν χωρικές και σημασιολογικές πληροφορίες, γεγονός που εξασφαλίζει την πλήρη αμεταβλητότητα στις περιστροφές και στις μεταβολές στην κλιμάκωση των απεικονιζόμενων αντικειμένων. Από την άλλη πλευρά όμως η απώλεια των συγκεκριμένων πληροφοριών αποτελεί και μειονέκτημα



της μεθόδου καθώς δύο εικόνες με όμοια ιστογράμματα μπορεί να έχουν τελείως διαφορετικό περιεχόμενο.

Υφή: Ένα άλλο σημαντικό χαρακτηριστικό το οποίο ορίζεται για συγκεκριμένες περιοχές της εικόνας είναι η υφή (*texture*). Συνηθίζεται να ορίζουμε την υφή ως ένα οπτικό μοτίβο το οποίο χαρακτηρίζεται από την επανάληψη μερικών βασικών αρχέτυπων ή *textons*. Τα *textons* είναι θεμελιώδεις μικρο-δομές των φυσικών εικόνων και θεωρούνται τα βασικά στοιχεία της προ-προσεκτικής (*pre-attentive*) ανθρώπινης οπτικής αντίληψης [25]. Η προ-προσεκτική όραση κατευθύνει την προσεκτική όραση σε περιοχές όπου υπάρχουν διαβαθμίσεις των *textons*. Το συγκεκριμένο χαρακτηριστικό δεν έχει την ικανότητα να βρίσκει όμοιες εικόνες αλλά μπορεί να χρησιμοποιηθεί για να διακρίνει εικόνες με υφή από εικόνες που δεν έχουν κάποια συγκεκριμένη υφή και στη συνέχεια να συνδυαστεί με κάποιο άλλο οπτικό χαρακτηριστικό, όπως το χρώμα, για να καταστεί η ταξινόμηση πιο αποδοτική. Οι μέθοδοι αναπαράστασης υφής μπορούν να διακριθούν σε δύο κατηγορίες: στις δομικές (*structural*) και στις στατιστικές (*statistical*) [26]. Οι στατιστικές μέθοδοι, συμπεριλαμβανομένων των μεθόδων φασμάτων ισχύος του Fourier (*Fourier power spectra*), shift-invariant principal components analysis (SPCA), μετασχηματισμών Gabor και wavelet, χαρακτηρίζουν την υφή μέσω της στατιστικής κατανομής της έντασης της εικόνας.

Σχήμα: Το σχήμα είναι από τα βασικά χαρακτηριστικά περιγραφής του περιεχομένου μιας εικόνας. Αλγόριθμοι όπως ο Canny, Sobel και άλλοι πιο σύνθετοι όπως ο HOG (*Histogram of Oriented Gradients*) [27] παρέχουν πληροφορίες σχετικά με το σχήμα των αντικειμένων που περιέχει μια εικόνα με τη μορφή ενός διανύσματος περιγραφής σχήματος. Δεδομένου ότι το σχήμα ενός αντικειμένου είναι συνισταμένη των ακμών που το αποτελούν, είναι αναμενόμενο ότι εικόνες με παρόμοιο περιεχόμενο θα έχουν και παρόμοιο ιστογράμμο ακμών. Ο αλγόριθμος HOG για τον υπολογισμό ενός ιστογράμματος χωρίζει την εικόνα σε $C \times C$ (π.χ. $C=5$) κελιά και εντοπίζει τις ακμές που υπάρχουν σε κάθε κελί. Υπολογίζει τον προσανατολισμό τους και με βάση αυτόν τις κατανέμει σε m κάδους (*bins*). Το τελικό διάνυσμα υπολογίζεται από την συνάθροιση των επιμέρους ιστογραμμάτων.



Καθολικότητα και Τοπικότητα των Χαρακτηριστικών: Τα χαρακτηριστικά που περιγράφηκαν παραπάνω μπορούν να υπολογιστούν είτε για το σύνολο της εικόνας (*global features*) είτε για μικρότερα τμήματά της (*local features*) π.χ. για μικρές περιοχές της εικόνας σταθερής διάστασης (*patches*) ή περιοχές μεγάλης ομοιογένειας σε χρώμα και υφή. Τα καθολικά χαρακτηριστικά έχουν τη δυνατότητα να περιγράφουν μια εικόνα μέσω ενός μοναδικού διανύσματος χαρακτηριστικών, ως εκ τούτου η χρήση τους σε μεθόδους ταξινόμησης είναι απλή και μπορούν να χρησιμοποιηθούν με οποιονδήποτε ταξινομητή. Τα χαρακτηριστικά αυτά είναι ελκυστικά καθώς παράγουν πολύ συμπαγείς αναπαραστάσεις των εικόνων, όπου κάθε εικόνα αντιστοιχεί σε ένα σημείο σε έναν υψηλής διάστασης χώρο χαρακτηριστικών. Από την άλλη πλευρά, τα καθολικά χαρακτηριστικά είναι ευαίσθητα σε περιπτώσεις μερικής απόκρυψης του αντικειμένου και «ακαταστασίας» του φόντου. Για αυτό το λόγο, υποθέτουμε είτε ότι η εικόνα περιλαμβάνει ένα μοναδικό αντικείμενο ενδιαφέροντος είτε ότι είναι εύκολη η απομάκρυνση του φόντου από το αντικείμενο.

Σε αντίθεση με τα καθολικά χαρακτηριστικά, τα τοπικά χαρακτηριστικά υπολογίζονται σε πολλαπλά σημεία της εικόνας συνεπώς είναι περισσότερο ανθεκτικά σε περιπτώσεις θορύβου ενώ παράλληλα δεν απαιτούν την απομάκρυνση του φόντου. Σ' αυτή την περίπτωση η αναπαράσταση μιας εικόνας περιλαμβάνει ένα σύνολο διανυσμάτων χαρακτηριστικών. Ο εντοπισμός των σημείων ενδιαφέροντος σε μια εικόνα πραγματοποιείται συνήθως σε πολλαπλές κλίμακες και επιλέγονται σημεία τα οποία επαναλαμβάνονται σε διαφορετικές όψεις του αντικειμένου. Το διάνυσμα χαρακτηριστικών για κάθε σημείο περιγράφει ένα τμήμα της εικόνας (*image patch*) γύρω από το σημείο ενδιαφέροντος. Αυτή η τεχνική χρησιμοποιείται από τον αλγόριθμο SIFT [28, 29], ο οποίος χρησιμοποιεί τα τοπικά μέγιστα της συνάρτησης διαφορών του Gauss ως σημεία ενδιαφέροντος και τα ιστογράμματα προσανατολισμών κλίσης (*gradient orientations*) που υπολογίζονται γύρω από τα σημεία ως περιγραφείς της εικόνας. Επίσης, ο εντοπισμός των σημείων ενδιαφέροντος μπορεί να γίνει και μέσω δειγματοληψίας σημείων από την εικόνα σύμφωνα με κάποια κατανομή. Γενικά, τα τοπικά χαρακτηριστικά αναπαριστούν συνήθως την υφή ή το χρώμα μιας περιοχής της εικόνας, ενώ τα καθολικά χαρακτηριστικά περιλαμβάνουν αναπαραστάσεις περιγράμματος (*contours*), περιγραφείς σχήματος και χαρακτηριστικά υφής.



Η προτεινόμενη μέθοδος, η οποία περιγράφεται στην ενότητα 3.2, ανήκει στις μεθόδους οι οποίες εντοπίζουν σημεία ενδιαφέροντος στην εικόνα μέσω δειγματοληψίας και εξάγουν τοπικά χαρακτηριστικά χρώματος από μια μικρή περιοχή γύρω από το σημείο. Από την άλλη πλευρά, υπάρχουν μέθοδοι οι οποίες εκμεταλλεύονται και τις δύο παραπάνω τεχνικές, υπολογίζουν δηλ. καθολικά χαρακτηριστικά μέσω του υπολογισμού τοπικών χαρακτηριστικών. Παράδειγμα μιας τέτοιας μεθόδου αποτελεί η μέθοδος Bag of Keyoints [30, 31] η οποία περιγράφεται στην επόμενη ενότητα.

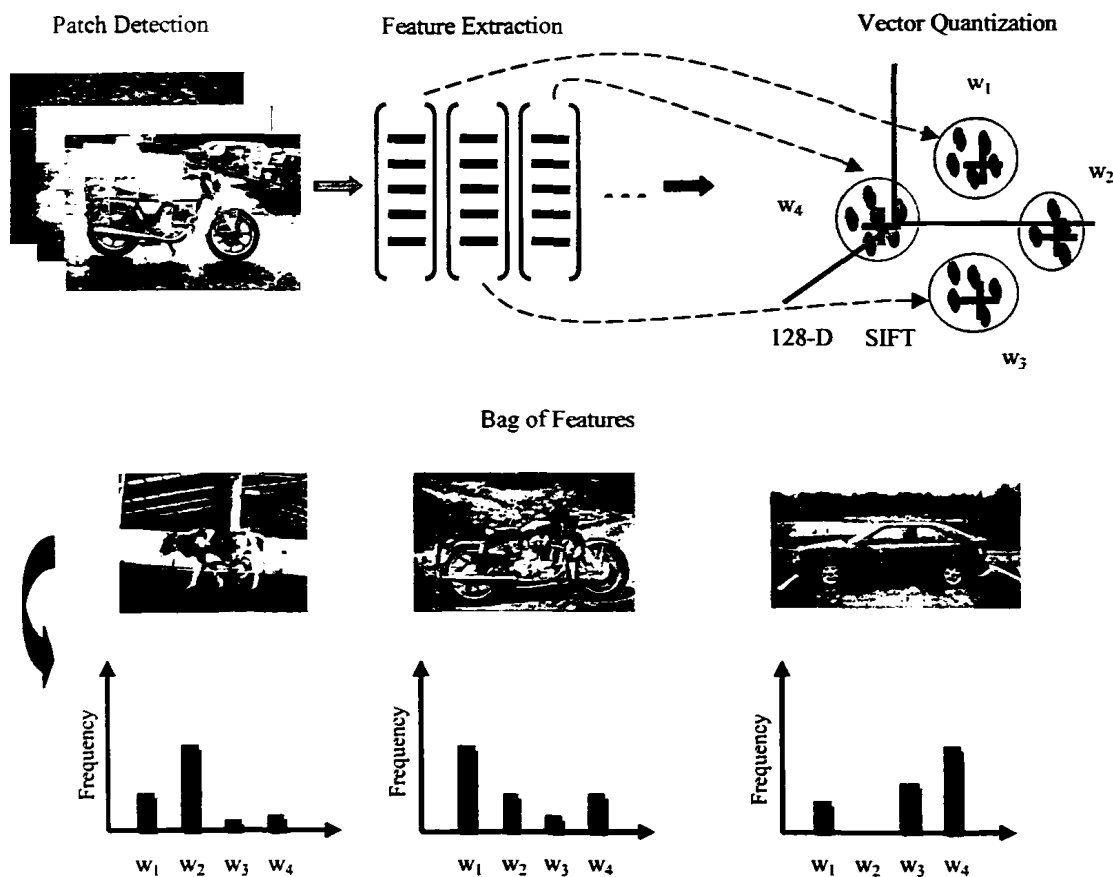
3) Στάδιο Επιλογής Μεθόδου Ταξινόμησης/ Κατασκευής Μοντέλου Ταξινόμησης (Classifier Construction): Στο δεύτερο κεφάλαιο έγινε μια συνοπτική παρουσίαση των κυριότερων μεθόδων ταξινόμησης. Γενικά, δεν μπορούμε να ισχυριστούμε ότι μια μέθοδος ταξινόμησης είναι αποδοτικότερη από τις υπόλοιπες παρά μόνο λαμβάνοντας υπόψη την διαθέσιμη γνώση για το συγκεκριμένο πρόβλημα και τη δομή των δεδομένων εκπαίδευσης π.χ. τη διάσταση των δεδομένων, το πλήθος των διαθέσιμων παραδειγμάτων εκπαίδευσης, αν υπάρχει θόρυβος στα δεδομένα, αν είναι γνωστή η κατηγορία (*class label*) των training patches, αν η αναπαράσταση των εικόνων εκπαίδευσης γίνεται με καθολικά ή τοπικά χαρακτηριστικά κτλ. Οι πιο συχνά χρησιμοποιούμενοι ταξινομητές σε προβλήματα ταξινόμησης εικόνων είναι τα δέντρα απόφασης και τα SVM.

4) Στάδιο Αξιολόγησης του Ταξινομητή (Performance Evaluation): Στόχος των αλγόριθμων μάθησης είναι η κατασκευή μοντέλων ταξινόμησης με καλή γενικευτική ικανότητα δηλ. μοντέλων τα οποία προβλέπουν με ακρίβεια την κατηγορία νέων εικόνων. Οι μετρικές που χρησιμοποιούνται για την εκτίμηση της απόδοσης είναι η ακρίβεια (*accuracy*) και η πιθανότητα σφάλματος (*error rate*). Για τον υπολογισμό τους απαιτείται ένα σύνολο εικόνων (*test images*) διαφορετικό από αυτό που έχει χρησιμοποιηθεί για την εκπαίδευση του μοντέλου. Η εκτίμηση της απόδοσης είναι αξιόπιστη όταν το σύνολο των εικόνων ελέγχου είναι αρκετά μεγάλο, σε διαφορετική περίπτωση η διαδικασία εκπαίδευσης-ελέγχου επαναλαμβάνεται πολλές φορές εφαρμόζοντας *bootstrap* μεθόδους ή *cross-validation*.



3.1.2 Η Μέθοδος Bag of Keypoints

Η μέθοδος *Bag of Keypoints* έχει εμπνευστεί από ανάλογες μεθόδους ταξινόμησης κειμένου (*bag of words*) [30, 31]. Η διαδικασία κατασκευής του μοντέλου ταξινόμησης βασίζεται στην παραδοχή ότι κάθε αντικείμενο αποτελείται από ένα σύνολο διακριτών ως προς την εμφάνισή τους περιοχών (*patches*), χωρίς να μας ενδιαφέρει η θέση τους στο χώρο.



Σχήμα 3.2 Ταξινόμηση Εικόνων με Χρήση Θυλάκων από Σημεία-Κλειδιά

Τα βασικά βήματα της μεθόδου για την εκπαίδευση του μοντέλου απεικονίζονται στο Σχήμα 3.2 και περιγράφονται συνοπτικά παρακάτω:

- *Εξαγωγή χαρακτηριστικών από ένα σύνολο εικόνων εκπαίδευσης:* Για την περιγραφή της εμφάνισης κάθε εικόνας χρησιμοποιείται ο αλγόριθμος SIFT (*Scale Invariant Feature Transform*) [28, 29] ο οποίος εξάγει από την εικόνα ένα σύνολο από σημεία-κλειδιά. Κάθε σημείο-κλειδί περιγράφεται από ένα διάνυσμα διάστασης 128.



- *Δημιουργία ενός οπτικού λεξικού:* Το σύνολο των διανυσμάτων που έχουν παραχθεί από τον αλγόριθμο SIFT για το σύνολο των εικόνων εκπαίδευσης διαμερίζεται σε k υποσύνολα-ομάδες χρησιμοποιώντας τον αλγόριθμο ομαδοποίησης *K-means* [22]. Τα κέντρα των ομάδων αποτελούν τις «λέξεις» του οπτικού λεξικού.
- *Κατασκευή ενός ιστογράμματος για κάθε εικόνα εκπαίδευσης:* Η κατασκευή του ιστογράμματος γίνεται με βάση τη συχνότητα εμφάνισης των οπτικών λέξεων του λεξικού σε κάθε εικόνα.
- *Κατασκευή ενός μοντέλου ταξινόμησης:* Ο ταξινομητής SVM [32] χρησιμοποιείται για την εύρεση ενός υπερεπιπέδου απόφασης το οποίο να διαχωρίζει το σύνολο των ιστογραμμάτων με τη μορφή διανυσμάτων συχνοτήτων.

Για τη εκτίμηση της απόδοσης του μοντέλου ταξινόμησης, η μέθοδος ακολουθεί τα εξής βήματα για κάθε εικόνα ελέγχου:

- *Εξαγωγή χαρακτηριστικών και κβάντιση σε οπτικές λέξεις*
- *Κατασκευή ενός ιστογράμματος με βάση τη συχνότητα εμφάνισης των οπτικών λέξεων*
- *Εκτίμηση της κατηγορίας με χρήση του ταξινομητή SVM*

3.2. Η Προτεινόμενη Μέθοδος

Στην παρούσα εργασία προτείνεται μια μέθοδος η οποία επιλύει το πρόβλημα της ταξινόμησης εικόνων με βάση το περιεχόμενο σε μια από τις προκαθορισμένες κατηγορίες εφαρμόζοντας την προσέγγιση «winner takes all», δηλ. βρίσκει την κατηγορία του αντικειμένου που είναι πιο πιθανό να απεικονίζεται στην εικόνα. Η μέθοδος αυτή στηρίζεται στην κατασκευή τυχαίων δασών αποτελούμενων από δυαδικά δέντρα απόφασης. Παρόμοιες μέθοδοι έχουν προταθεί τα τελευταία χρόνια για την επίλυση άλλων προβλημάτων όπως η ανίχνευση αντικειμένων σε φυσικές εικόνες (*object detection*) [3, 4] και η αναγνώριση κίνησης σε βίντεο [5, 12]. Η μέθοδος που προτείνεται περιλαμβάνει την κατασκευή ενός διακριτού δάσους για καθεμία από τις προκαθορισμένες κατηγορίες του προβλήματος.



Γενική Περιγραφή της Μεθόδου

Η μέθοδος αρχικά επεξεργάζεται ένα σύνολο εικόνων εκπαίδευσης. Το συγκεκριμένο σύνολο αποτελείται από εικόνες όλων των προκαθορισμένων κατηγοριών με ίση αναλογία μεταξύ τους. Για τη διανυσματική περιγραφή των εικόνων εξάγουμε τοπικά χαρακτηριστικά χρώματος από μικρά, σταθερής διάστασης, τμήματα της εικόνας (*training patches*). Το σύνολο των διανυσμάτων χαρακτηριστικών που προκύπτει από τα patches χρησιμοποιείται στη συνέχεια από τη μέθοδο ταξινόμησης για την κατασκευή ενός μοντέλου ταξινόμησης. Η προτεινόμενη μέθοδος χρησιμοποιεί δυαδικά δέντρα απόφασης (*binary decision trees*) ως ταξινομητές. Για κάθε κατηγορία δημιουργείται ένα διαφορετικό σύνολο δυαδικών δέντρων απόφασης, ένα διαφορετικό τυχαίο δάσος. Κάθε δυαδικό δέντρο απόφασης κατασκευάζεται μέσα από μια επαναληπτική διαδικασία ξεκινώντας από τη ρίζα προς τα φύλλα (*top-down*). Κατά την επαγωγή ενός δέντρου σε κάθε μη τερματικό κόμβο ανατίθεται ένα υποσύνολο των training patches τα οποία διαμερίζονται στο επόμενο στάδιο με την εφαρμογή του αλγόριθμου k-means. Στους τερματικούς κόμβους αποθηκεύεται η πληροφορία για τα training patches τα οποία έχουν ανατεθεί σε αυτούς.

Για την ταξινόμηση «άγνωστης» ταυτότητας εικόνων εξάγουμε χαρακτηριστικά χρώματος από patches τα οποία επιλέγουμε κάνοντας δειγματοληψία σύμφωνα με την ομοιόμορφη κατανομή. Καθένα από τα patches διασχίζει όλα τα δέντρα του δάσους μιας συγκεκριμένης κατηγορίας, ξεκινώντας από τη ρίζα τους και καταλήγοντας σε έναν από τα φύλλα. Με βάση την πληροφορία που έχει αποθηκευτεί στους τερματικούς κόμβους (κατά την φάση εκπαίδευσης του μοντέλου) υπολογίζεται για το patch μια πιθανοτική ψήφος η οποία ανατίθεται σε διαφορετικά σημεία μιας αρχικά μαύρης εικόνας (εικόνα Hough). Από τη συνάθροιση των ψήφων όλων των test patches κατασκευάζεται η τελική εικόνα Hough η μέγιστη τιμή της οποίας αποτελεί μέτρο εμπιστοσύνης για την απεικόνιση του αντικειμένου ενδιαφέροντος στην εικόνα. Στην περίπτωση που έχουμε να επιλύσουμε ένα πρόβλημα ταξινόμησης πολλών κατηγοριών είναι απαραίτητο κάθε test patch να διασχίσει τα δέντρα όλων των δασών και να υπολογίσουμε μια διαφορετική εικόνα Hough για κάθε κατηγορία. Στη συνέχεια εξετάζουμε τις μέγιστες τιμές των εικόνων Hough και αναθέτουμε την εικόνα στην κατηγορία εκείνης της εικόνας Hough που έδωσε τη μεγαλύτερη τιμή. Παρακάτω περιγράφεται αναλυτικά η διαδικασία που ακολουθεί η μέθοδος για την

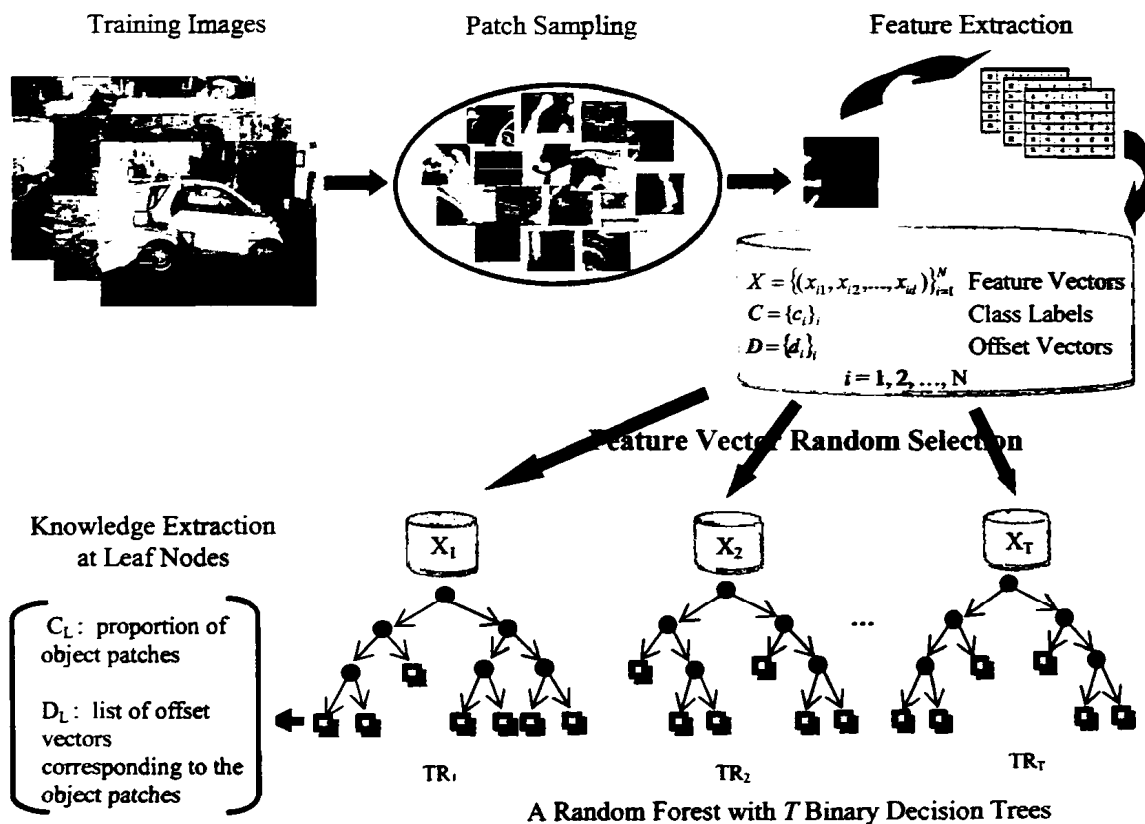


κατασκευή του μοντέλου ταξινόμησης καθώς και για την κατασκευή του μηχανισμού απόφασης.

3.2.1 Εκπαίδευση του Μοντέλου Ταξινόμησης

Όπως φαίνεται στο Σχήμα 3.3 τα βασικά βήματα για την κατασκευή ενός μοντέλου με την χρήση της προτεινόμενης μεθόδου είναι τα εξής:

- Δειγματοληψία σημείων της εικόνας και εξαγωγή διανυσμάτων χαρακτηριστικών από ένα σύνολο εικόνων εκπαίδευσης.
- Κατασκευή τυχαίων υποσυνόλων διανυσμάτων χαρακτηριστικών.
- Κατασκευή στοχαστικών δασών δυαδικών δέντρων απόφασης.
- Εξαγωγή γνώσης από τους τερματικούς κόμβους των δέντρων.



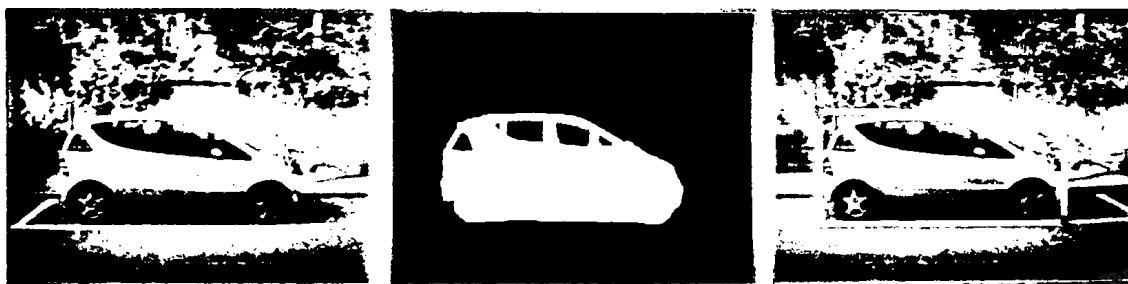
Σχήμα 3.3 Διαδικασία Εκπαίδευσης του Μοντέλου Ταξινόμησης



3.2.1.1 Δειγματοληψία Σημείων της Εικόνας και Εξαγωγή Διανυσμάτων Χαρακτηριστικών

Για τη διανυσματική αναπαράσταση των εικόνων του συνόλου εκπαίδευσης, όπως αναφέρθηκε και παραπάνω, εξάγουμε τοπικά χαρακτηριστικά χρώματος από μικρά τμήματα των εικόνων (patches) σταθερής διάστασης $M \times M$ pixels. Συγκεκριμένα, επιλέγουμε σημεία πάνω στις εικόνες κάνοντας δειγματοληψία (με επανατοποθέτηση) σύμφωνα με την ομοιόμορφη κατανομή. Τα σημεία αυτά αποτελούν τα κέντρα των patches. Στη προτεινόμενη μέθοδο τα patches που χρησιμοποιούνται για την εκπαίδευση του μοντέλου (*training patches*) διακρίνονται γενικά σε δύο κατηγορίες: (α) *object patches* και (β) *background patches*. Ως *object patches* θεωρούνται τα τμήματα της εικόνας τα οποία απεικονίζουν ένα μέρος του αντικειμένου ενδιαφέροντος, ενώ ως *background patches* θεωρούνται τα τμήματα της εικόνας τα οποία είτε δεν απεικονίζουν κάποιο συγκεκριμένο αντικείμενο δηλ. είναι μέρος του φόντου είτε απεικονίζουν μέρος ενός αντικειμένου διαφορετικής κατηγορίας από την κατηγορία ενδιαφέροντος.

Συνεπώς, για τη δειγματοληψία των patches είναι απαραίτητο να γνωρίζουμε τα όρια του αντικειμένου ώστε να διακρίνουμε τα *object* από τα *background patches*. Η πληροφορία των ορίων του αντικειμένου μπορεί να υπολογιστεί (όπως φαίνεται στο Σχήμα 3.4) είτε σε επίπεδο πλαισίων οριοθέτησης (*bounding boxes*) π.χ. $(x_{\min}, y_{\min}) - (x_{\max}, y_{\max}) : (14, 45) - (238, 111)$ είτε μέσω μασκών κατάτμησης εικονοστοιχείων (*pixel segmentation masks*). Με αυτόν τον τρόπο γνωρίζουμε εκ των προτέρων την κατηγορία στην οποία ανήκει κάθε patch και μπορούμε να χρησιμοποιήσουμε την γνώση αυτή ώστε να εκπαιδεύσουμε το μοντέλο ταξινόμησης με τη μέγιστη δυνατή επίβλεψη (*supervised learning*).



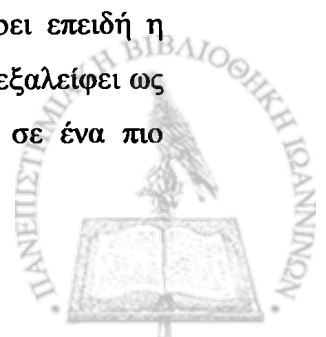
Σχήμα 3.4 Οριοθέτηση Αντικειμένου μέσω Μασκών Κατάτμησης Εικονοστοιχείων και Πλαισίων Οριοθέτησης



Το σύνολο των training patches ορίζεται ως εξής: $\{P_i=(I_i, c_i, d_i)\}$, όπου I_i είναι η εμφάνιση του patch η οποία ορίζεται από τα εξαγόμενα χαρακτηριστικά του. Η συγκεκριμένη μέθοδος χρησιμοποιεί μόνο την πληροφορία χρώματος για να παράγει ένα διάνυσμα χαρακτηριστικών για κάθε patch. Αν για παράδειγμα χρησιμοποιήσουμε την πληροφορία των φωτεινότητων των pixels (R,G,B) και των απόλυτων τιμών των πρώτων x - και y - παραγώγων τους η εμφάνιση ενός training patch περιγράφεται από ένα διάνυσμα της μορφής $I_i = (I^r, I^g, I^b, I^r_x, I^r_y, I^g_x, I^g_y, I^b_x, I^b_y)$, όπου I_i είναι μια εικόνα $M \times M$ για τα I^r, I^g, I^b , $M-1 \times M$ για τα I^r_y, I^g_y, I^b_y και $M \times M-1$ για τα I^r_x, I^g_x, I^b_x . Για να αυξήσουμε την αμεταβλητότητα σε περίπτωση θορύβου και μεγάλων διακυμάνσεων στην δομή των επιμέρους τμημάτων του αντικειμένου εκτός από τα παραπάνω χαρακτηριστικά μπορούμε να χρησιμοποιήσουμε τις απόλυτες τιμές των δεύτερων x - και y - παραγώγων των φωτεινότητων καθώς και χαρακτηριστικά σχήματος όπως HOG-like κανάλια χαρακτηριστικών.

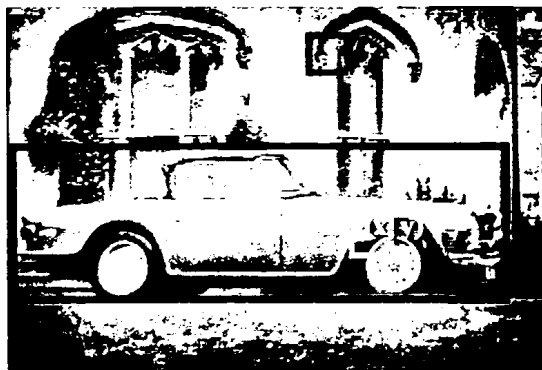
Πριν την εξαγωγή των χαρακτηριστικών, εφαρμόζεται ένα στάδιο προεπεξεργασίας στα κανάλια των φωτεινότητων ώστε τα αρχικά ακατέργαστα δεδομένα να μετασχηματιστούν σε μια καταλληλότερη μορφή για μετέπειτα επεξεργασία. Συγκεκριμένα, οι τιμές των φωτεινότητων κανονικοποιούνται στο $[0,1]$ με στόχο να μειώσουμε το εύρος των τιμών τους.

Από την άλλη πλευρά, όσο αυξάνουμε το πλήθος των εξαγόμενων γνωρισμάτων, τόσο αυξάνεται και η διάσταση των δεδομένων που πρέπει να επεξεργαστεί ο αλγόριθμος εκπαίδευσης. Αν τα δεδομένα εκπαίδευσης είναι υψηλής διάστασης τότε το μοντέλο που θα προκύψει θα είναι αρκετά πολύπλοκο. Για να μειωθεί η διάσταση των δεδομένων και να απεικονίσουμε πιο αποδοτικά τη σημαντική πληροφορία μιας εικόνας, κρατάμε για κάθε patch το μέσο όρο και την τυπική απόκλιση των τιμών κάθε εξαγόμενου χαρακτηριστικού. Επομένως, αν το πλήθος των εξαγόμενων γνωρισμάτων είναι d , κάθε patch περιγράφεται από ένα διάνυσμα διάστασης $1 \times D$, όπου $D=2*d$. Το σημαντικότερο πλεονέκτημα από τη μείωση της διάστασης είναι ότι οι περισσότεροι αλγόριθμοι εξόρυξης δεδομένων είναι πιο αποδοτικοί όταν το πλήθος των χαρακτηριστικών των δεδομένων είναι μικρό. Αυτό ισχύει εν μέρει επειδή η μείωση της διάστασης οδηγεί σε μείωση του θορύβου στα δεδομένα και εξαλείφει ως ένα βαθμό μη σχετικά χαρακτηριστικά. Επίσης, μπορεί να οδηγήσει σε ένα πιο



κατανοητό μοντέλο ενώ παράλληλα η χρονική διάρκεια και η μνήμη που απαιτεί ο αλγόριθμος περιορίζονται.

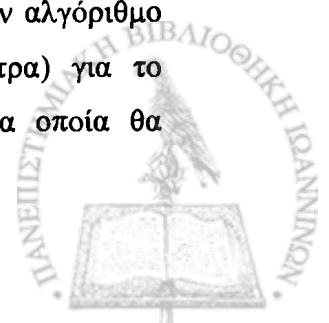
Το τελικό αποτέλεσμα της εξαγωγής χαρακτηριστικών από N patches (το πλήθος αυτό καθορίζεται από τον χρήστη) είναι η δημιουργία μιας βάσης δεδομένων η οποία αποτελείται από N διανύσματα χαρακτηριστικών της μορφής $x_i=(x_1, x_2, \dots, x_D)$, όπου $i=1,2, \dots,N$. Επειδή η εκπαίδευση του μοντέλου ταξινόμησης γίνεται με επίβλεψη για κάθε patch εκτός από το διάνυσμα χαρακτηριστικών παρέχουμε στον αλγόριθμο εκπαίδευσης την κατηγορία c_i του patch (object ή background). Τα σημεία που λαμβάνονται από το εσωτερικό του πλαισίου οριοθέτησης του αντικειμένου ενδιαφέροντος ορίζουν patches με $c_i=1$ (object patches), όλα τα υπόλοιπα θεωρούνται σημεία τα οποία ανήκουν στο φόντο και ορίζουν patches με $c_i=0$ (background patches). Επιπλέον παρέχουμε στον αλγόριθμο για κάθε object patch ένα δισδιάστατο διάνυσμα απόκλισης (offset vector) d_i . Το διάνυσμα αυτό ορίζεται από το κεντροειδές του πλαισίου οριοθέτησης του αντικειμένου στο κέντρο του patch δηλ. $d_i = (y-BB_y, x-BB_x)$. Για τα background patches το d_i δεν ορίζεται.



Σχήμα 3.5 Δειγματοληψία των Patches. Το Βέλος Αντιστοιχεί στο Διάνυσμα Απόκλισης (Offset Vector) του Object Patch

3.2.1.2 Κατασκευή Στοχαστικών Δασών Δυαδικών Δέντρων Απόφασης

Για την κατασκευή ενός δυαδικού δέντρου απόφασης η προτεινόμενη μέθοδος χρησιμοποιεί επαναληπτικά μια διαδικασία διαμέρισης των δειγμάτων που ανήκουν σ' αυτό σε δύο ξένα υποσύνολα. Αυτό επιτυγχάνεται εφαρμόζοντας τον αλγόριθμο ομαδοποίησης K -means [22] (για $K=2$ επειδή έχουμε δυαδικά δέντρα) για το διαχωρισμό των δειγμάτων του κόμβου σε δύο ξένα υποσύνολα, τα οποία θα



αποτελέσουν τα «παιδιά» του κόμβου. Στην περίπτωση των patches επειδή τα δεδομένα είναι (σχετικά) υψηλής διάστασης και έχουν μη αρνητικές τιμές χρησιμοποιούμε την απόσταση συνημίτονου (*cosine similarity*), η οποία υπολογίζει το συνημίτονο της γωνίας δυο διανυσμάτων A, B που αντιστοιχούν σε patches του κόμβου.

$$\cos(A, B) = \frac{A^T \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}}$$

Τα κεντροειδή (*centroids*) που υπολογίζονται για κάθε κόμβο θα χρησιμοποιηθούν κατά τη φάση του ελέγχου, στη διάσχιση των δέντρων από τα test patches.

Όπως έχει αναφερθεί και στην ενότητα 2.2.2 προτιμάμε γενικά τα απλούστερα δέντρα από τα πιο σύνθετα καθώς τα πρώτα είναι λιγότερο επιρρεπή σε φαινόμενα υπερεκπαίδευσης και συνεπώς εμφανίζουν καλύτερη γενικευτική ικανότητα. Όσον αφορά την προτεινόμενη μέθοδο τα δέντρα δεν αναπτύσσονται στο μέγιστο μέγεθός τους, αντίθετα χρησιμοποιούμε συνθήκες πρόωρου τερματισμού. Οι συνθήκες αυτές αφορούν το μέγιστο βάθος (d_{max}) και το ελάχιστο πλήθος των patches (N_{min}) του κόμβου. Αν ένας κόμβος ικανοποιεί τουλάχιστον μια από τις συνθήκες δηλώνεται ως τερματικός.



Σχήμα 3.6 Δείγματα των Patches ενός Τερματικού Κόμβου. Τα Μπλε Αντιστοιχούν σε Object Patches για την Κατηγορία «Cars»

Η εφαρμογή του αλγόριθμου K-means σε κάθε μη τερματικό κόμβο εξασφαλίζει ως ένα βαθμό ότι τα patches των τερματικών κόμβων θα έχουν μικρές αποκλίσεις όσον αφορά τα χαρακτηριστικά τους.

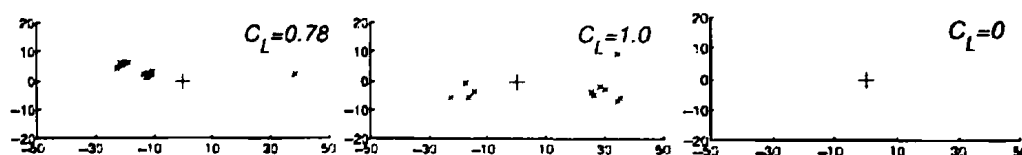
Για την κατασκευή ενός τυχαίου δάσους κατασκευάζουμε T δυαδικά δέντρα απόφασης. Κάθε δέντρο εκπαιδεύεται χρησιμοποιώντας ίσο πλήθος από object και



background patches. Για την επίλυση ενός προβλήματος ταξινόμησης πολλών κατηγοριών το μοντέλο ταξινόμησης που κατασκευάζει η μέθοδος αποτελείται από C στοχαστικά δάση, όπου C το πλήθος των προκαθορισμένων κατηγοριών του προβλήματος. Όλα τα δάση αποτελούνται από ίσο πλήθος δυαδικών δέντρων απόφασης. Κάθε δάσος κατασκευάζεται με τέτοιο τρόπο ώστε να επιλύει το δυαδικό πρόβλημα ταξινόμησης των εικόνων για μια από τις C κατηγορίες του προβλήματος. Για παράδειγμα, αν έχουμε το σύνολο των κατηγοριών {Cars, Cows, Pedestrians}, η μέθοδος κατασκευάζει τρία τυχαία δάση, ένα για κάθε κατηγορία. Για την κατασκευή ενός δάσους για την κατηγορία Cars επιλέγουμε object patches από τις εικόνες της κατηγορίας Cars και συγκεκριμένα από το εσωτερικό των ορίων των αυτοκινήτων και background patches από όλες τις εικόνες του συνόλου εκπαίδευσης, από τμήματα του φόντου ή τμήματα που απεικονίζουν διαφορετικά αντικείμενα π.χ. πεζούς, αγελάδες.

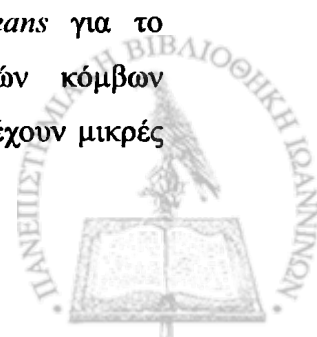
3.2.1.3 Εξαγωγή Γνώσης από τους Τερματικούς Κόμβους των Δέντρων

Σε κάθε τερματικό κόμβο ενός δέντρου αποθηκεύεται η πληροφορία για τα training patches τα οποία έχουν ανατεθεί σ' αυτόν. Πιο συγκεκριμένα, αποθηκεύεται η αναλογία C_L των object patches προς τα συνολικά patches του κόμβου (π.χ. $C_L=1$ σημαίνει ότι όλα τα patches του κόμβου είναι object patches) και μια λίστα $D_L = \{d_i\}_i$ με τα διανύσματα απόκλισης τα οποία αντιστοιχούν στα object patches (για τα background patches δεν ορίζεται d_i). Στο Σχήμα 3.7 απεικονίζεται η πληροφορία τριών τερματικών κόμβων ενός δέντρου. Οι πράσινοι σταυροί αντιστοιχούν στα διανύσματα απόκλισης των object patches στη λίστα D_L του φύλλου.



Σχήμα 3.7 Δείγματα Τερματικών Κόμβων Ενός Δέντρου

Όπως αναφέρθηκε και παραπάνω, η χρήση του αλγόριθμου *K-means* για το διαχωρισμό των παραδειγμάτων εκπαίδευσης των μη τερματικών κόμβων εξασφαλίζει ως ένα βαθμό ότι τα patches των τερματικών κόμβων θα έχουν μικρές

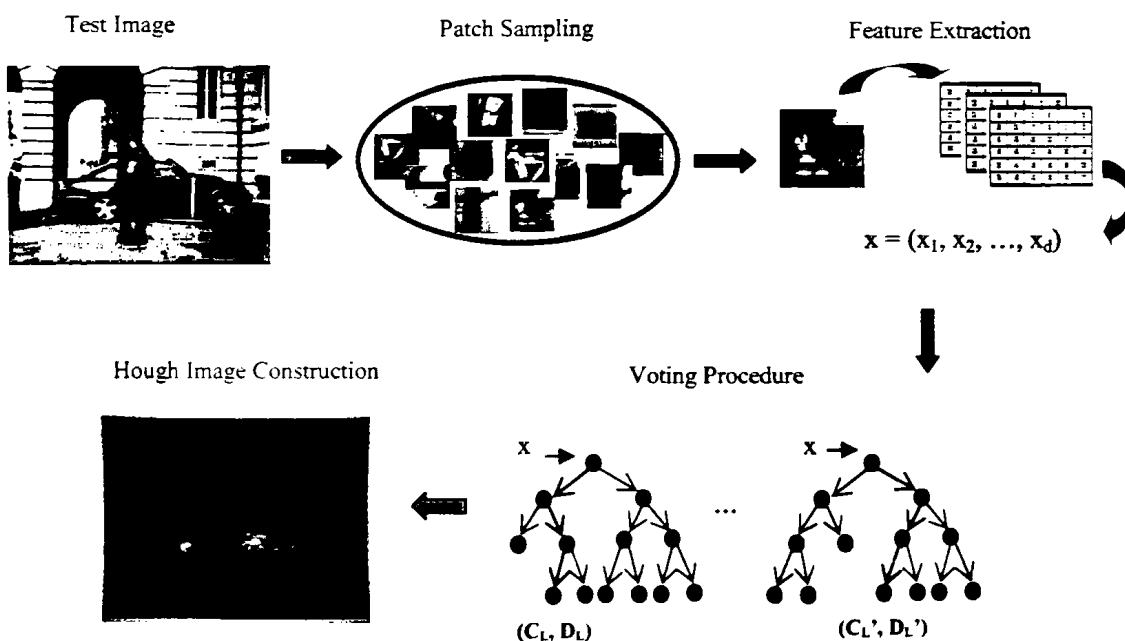


αποκλίσεις όσον αφορά τα χαρακτηριστικά τους. Ως αποτέλεσμα το C_L των φύλλων τείνει να είναι είτε κοντά στο 0 όταν ο κόμβος αποτελείται κυρίως από background patches, είτε κοντά στο 1 όταν ο κόμβος αποτελείται κυρίως από object patches. Επιδιώκουμε οι τερματικοί κόμβοι να έχουν μικρές αποκλίσεις όσον αφορά το class label.

Η πληροφορία των φύλλων θα χρησιμοποιηθεί κατά τη φάση του έλεγχου για να υπολογιστούν οι πιθανοτικές ψήφοι των test patches αναφορικά με την ύπαρξη του αντικειμένου στην εικόνα. Επειδή οι ψήφοι υπολογίζονται με βάση τις πληροφορίες των τερματικών κόμβων, κόμβοι με μικρό C_L , δηλ. κόμβοι που περιέχουν κυρίως background patches δίνουν μικρές ψήφους ενώ κόμβοι με μεγάλο C_L δηλ. κόμβοι που περιέχουν κυρίως object patches, δίνουν μεγάλες ψήφους.

3.2.2 Κατασκευή του Μηχανισμού Απόφασης

Στόχος κάθε αλγόριθμου εκπαίδευσης είναι η κατασκευή ενός μοντέλου ταξινόμησης με καλή γενικευτική ικανότητα δηλ. ενός μοντέλου το οποίο θα προβλέπει με ακρίβεια την κατηγορία «άγνωστων» εικόνων. Η διαδικασία της ταξινόμησης μίας εικόνας απεικονίζεται σχηματικά παρακάτω:



Σχήμα 3.8 Η Διαδικασία Ταξινόμησης μίας Εικόνας



Όπως φαίνεται στο Σχήμα 3.8 τα βασικά βήματα της προτεινόμενης μεθοδολογίας για την ταξινόμηση μιας εικόνας σε μία από τις προκαθορισμένες κατηγορίες είναι τα εξής:

- Δειγματοληψία σημείων από την εικόνα ελέγχου και εξαγωγή διανυσμάτων χαρακτηριστικών.
- Εφαρμογή μιας διαδικασίας «ψηφοφορίας»
- Κατασκευή ενός μηχανισμού απόφασης - εικόνα Hough

3.2.2.1 Δειγματοληψία Σημείων από την Εικόνα Ελέγχου και Εξαγωγή ενός Διανύσματος Χαρακτηριστικών

Για την ταξινόμηση μιας εικόνας ελέγχου με βάση το περιεχόμενό της σε μια από τις προκαθορισμένες κατηγορίες του προβλήματος επιλέγουμε σημεία από όλη την εικόνα με επανατοποθέτηση σύμφωνα με την ομοιόμορφη κατανομή. Τα σημεία αυτά αποτελούν τα κέντρα των test patches (τα patches όπως και στη φάση της εκπαίδευσης είναι σταθερής διάστασης $M \times M$ pixels). Για κάθε test patch είναι γνωστή μόνο η εμφάνιση του $I(y)$ όπου y το κέντρο του patch. Για τη διανυσματική περιγραφή των test patches ακολουθούμε τη διαδικασία που περιγράφηκε στην ενότητα 3.2.1.1 για τα training patches. Είναι απαραίτητο από τα test patches να εξάγουμε τα ίδια χαρακτηριστικά με αυτά που εξάγαμε από τα training patches κατά τη φάση της εκπαίδευσης καθώς η διαδικασία της ταξινόμησης, όπως θα περιγραφεί αναλυτικά στις επόμενες ενότητες, περιλαμβάνει τη σύγκριση των test patches με τα training patches που έχουν ανατεθεί στους μη τερματικούς κόμβους των δέντρων.

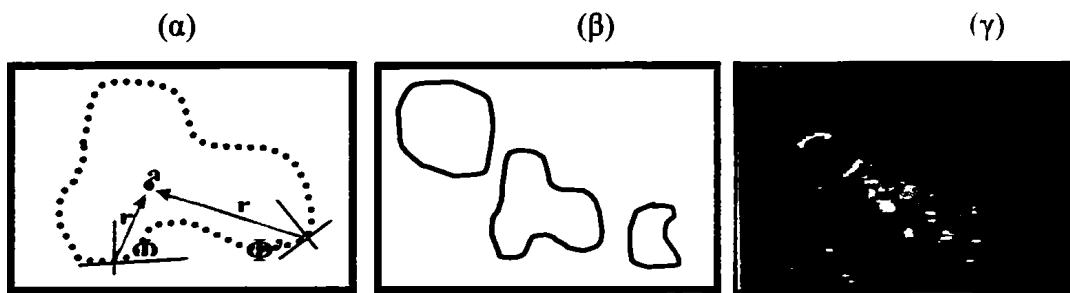
3.2.2.2 Η Διαδικασία της «Ψηφοφορίας»

Η διαδικασία της ψηφοφορίας που εφαρμόζει η μέθοδος στηρίζεται ουσιαστικά στην ιδέα του γενικευμένου μετασχηματισμού Hough [33]. Ο συγκεκριμένος μετασχηματισμός είναι μια τεχνική η οποία έχει προταθεί [34, 35] για τον εντοπισμό καμπυλών συγκεκριμένου σχήματος σε μία εικόνα χρησιμοποιώντας σημεία ακμών. Ως εκ τούτου η εφαρμογή του προϋποθέτει τη μετατροπή της εικόνας σε αναπαράσταση ακμών όπου κάθε σημείο της δεν περιέχει πληροφορία χρώματος αλλά αποτελεί ένα μέγεθος των τοπικών αλλαγών της έντασης του χρώματος. Η



σημασία που έχει δοθεί στα σχήματα των περιγραμμάτων οφείλεται σε πειράματα που έχουν γίνει στο σύστημα ανθρώπινης όρασης και έχουν δείξει ότι αδρές απεικονίσεις των ορίων είναι επαρκείς για την αναγνώριση αντικειμένων [36].

Βασικά σημεία του γενικευμένου αλγόριθμου Hough αποτελούν η χρήση της πληροφορίας κατεύθυνσης και η εφαρμογή μιας διαδικασίας ψηφοφορίας. Συγκεκριμένα, ο εντοπισμός ενός τυχαίου σχήματος S , απαιτεί την χρήση ενός (αρχικά μηδενικού) πίνακα A , ο οποίος καλείται συσσωρευτής (*accumulator*).



Σχήμα 3.9 (α) Επιλογή Σημείων Περιγράμματος (β) Εικόνα Τυχαίων Σχημάτων. (γ) Τα Σημεία Ακμής του Σχήματος S (στην εικόνα) Αυξάνουν το Ίδιο Σημείο στον Συσσωρευτή ενώ τα Σημεία Ακμής Διαφορετικών Σχημάτων Αυξάνουν Σκόρπια Σημεία

Ο αλγόριθμος επιλέγει τυχαία ένα σημείο αναφοράς a για το σχήμα. Όπως φαίνεται και στο Σχήμα 3.9α για κάθε σημείο x του περιγράμματος του, υπολογίζει την κατεύθυνση κλίσης $\varphi(x)$ και μια απόσταση $r=a-x$. Οι πληροφορίες αυτές αποθηκεύονται σε ένα πίνακα, ο οποίος καλείται R-table.

Πίνακας 3.1 Η Μορφή του Πίνακα R-Table

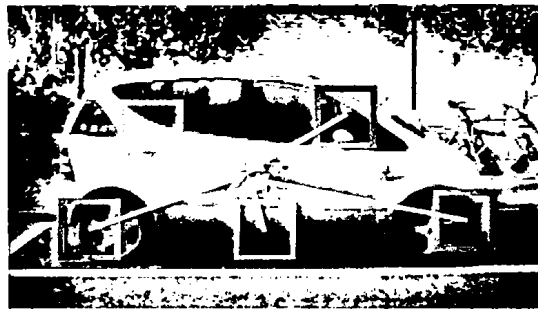
i	φ_i	Γ_{φ_i}
0	0	$\{r \mid a-r=x, x \text{ in Bound}, \varphi(x)=0\}$
1	φ'	$\{r \mid a-r=x, x \text{ in Bound}, \varphi(x)=\varphi'\}$
2	φ''	$\{r \mid a-r=x, x \text{ in Bound}, \varphi(x)=\varphi''\}$
...



Για την ίδια τιμή κλίσης ϕ μπορεί να υπάρχουν περισσότερες από μια καταχωρήσεις στον πίνακα. Ο πίνακας R-table (Πίνακας 3.1) αποτελεί μια πλήρη περιγραφή του σχήματος-στόχου και μπορεί να χρησιμοποιηθεί στη συνέχεια για την ανίχνευση στιγμιότυπων σχήματος S σε μια εικόνα.

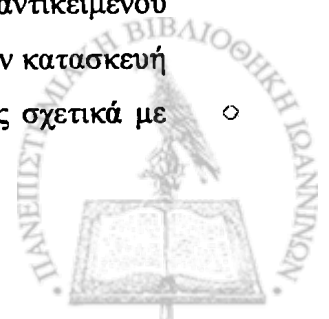
Η διαδικασία ανίχνευσης έχει ως εξής: για κάθε σημείο ακμής x στην εικόνα (β) Σχήμα 3.9, υπολογίζουμε την κατεύθυνση κλίσης $\phi(x)$ και καθορίζουμε σε ποιες γραμμές του πίνακα R-table αντιστοιχεί. Για κάθε τέτοια καταχώρηση στο πίνακα, αυξάνουμε κατά 1 όλα τα αντίστοιχα σημεία $x+r$ στον πίνακα του συσσωρευτή (δίνουμε μια ψήφο). Τα μέγιστα σημεία του συσσωρευτή αντιστοιχούν σε πιθανά στιγμιότυπα του σχήματος S στην εικόνα.

Η προτεινόμενη μέθοδος εφαρμόζει μια διαδικασία ψηφοφορίας αντίστοιχη του μετασχηματισμού Hough, χρησιμοποιώντας ως σημείο αναφοράς το κεντροειδές των αντικείμενων ενδιαφέροντος.



Σχήμα 3.10 Τα Διανύσματα Απόκλισης των Object Patches Παρέχουν την Πληροφορία Κατεύθυνσης για τη Διαδικασία της Ψηφοφορίας

Όπως φαίνεται στο Σχήμα 3.10, η πληροφορία κατεύθυνσης για τη διαδικασία της ψηφοφορίας παρέχεται μέσω των διανυσμάτων απόκλισης (*offset vectors*) των object patches (Σχήμα 3.5). Η πληροφορία αυτή έχει αποθηκευτεί στους τερματικούς κόμβους των δέντρων κατά τη φάση της εκπαίδευσης του μοντέλου ταξινόμησης. Συνεπώς, το σύνολο των τερματικών κόμβων ενός δάσους μπορεί να ιδωθεί ως ένα «κωδικό βιβλίο» το οποίο παρέχει μια πλήρη περιγραφή του αντικείμενου ενδιαφέροντος. Η διαδικασία της ψηφοφορίας έχει ως αποτέλεσμα για την κατασκευή ενός στοχαστικού μηχανισμού απόφασης ο οποίος παρέχει πληροφορίες σχετικά με



την απεικόνιση ή μη ενός αντικειμένου μιας συγκεκριμένης κατηγορίας στην εικόνα ελέγχου. Παρακάτω περιγράφονται αναλυτικά τα βασικά βήματα της διαδικασίας.

1^ο) Διάσχιση των δέντρων ενός δάσους από τα test patches

Κάθε test patch διασχίζει ένα δυαδικό δέντρο απόφασης ξεκινώντας από τη ρίζα και καταλήγει σ' έναν από τους τερματικούς κόμβους του δέντρου. Η μεταβίβαση του patch από έναν κόμβο σε ένα από τα παιδιά του γίνεται συγκρίνοντας τα χαρακτηριστικά του συγκεκριμένου patch με τα χαρακτηριστικά των training patches που έχουν εκχωρηθεί στα παιδιά του αρχικού κόμβου δηλ. υπολογίζεται η απόστασή του test patch από τα κέντρα των παιδιών (τα κέντρα έχουν υπολογιστεί και αποθηκευτεί στον κόμβο από τον αλγόριθμο ομαδοποίησης κατά τη φάση της εκπαίδευσης του μοντέλου). Το μέτρο ομοιότητας που χρησιμοποιείται είναι η απόσταση συνημίτονου. Γενικά, για τη σύγκριση των patches μπορούμε να χρησιμοποιήσουμε ένα υποσύνολο των γνωρισμάτων, έστω K όπου $K \leq D$. Η μέθοδος χρησιμοποιεί το σύνολο των γνωρισμάτων. Το patch μεταβαίνει στο παιδί το κέντρο του οποίου βρίσκεται πιο κοντά σ' αυτό.

2^ο) Υπολογισμός μιας διαφορετικής πιθανοτικής ψήφου για κάθε patch

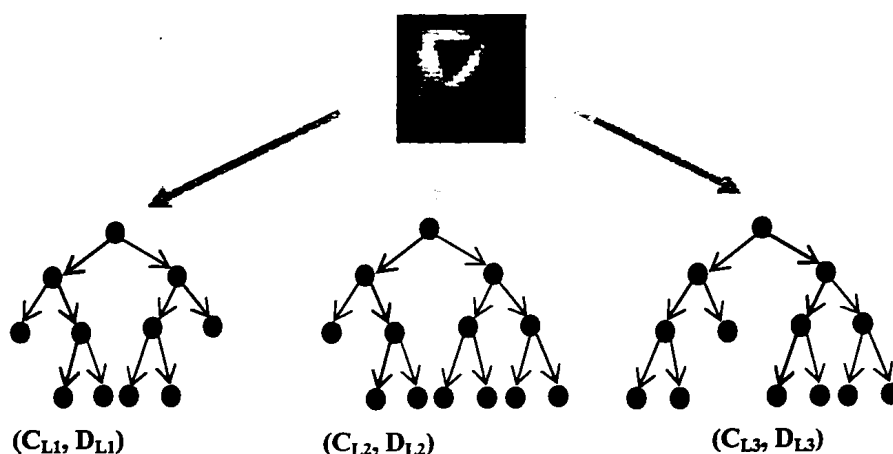
Η πληροφορία (C_L, D_L) του τερματικού κόμβου στον οποίο «φθάνει» τελικά το patch χρησιμοποιείται για τον υπολογισμό μιας πιθανοτικής ψήφου. Για κάθε δέντρο η ψήφος για ένα patch με κέντρο y υπολογίζεται ως εξής [2]:

$$p(E(x) | I(y); T) = \frac{1}{|D_L|} \cdot C_L$$

όπου το $E(x)$ δηλώνει το τυχαίο γεγονός που αντιστοιχεί στην ύπαρξη του κέντρου του αντικειμένου ενδιαφέροντος στη θέση x της εικόνας ελέγχου.

Το μοντέλο ταξινόμησης που κατασκευάζει η μέθοδος περιλαμβάνει ένα δάσος από T δυαδικά δέντρα απόφασης για κάθε κατηγορία. Όπως φαίνεται στο Σχήμα 3.11, κάθε test patch πρέπει να διασχίσει όλα τα δέντρα του δάσους ξεκινώντας από τη ρίζα και καταλήγοντας σε ένα τερματικό κόμβο.





Σχήμα 3.11 Διάσχιση των Δέντρων Ενός Δάσους από Ένα Test Patch

Η αποθηκευμένη πληροφορία (C_L, D_L) των τερματικών κόμβων διαφέρει, συνεπώς κάθε δέντρο δίνει διαφορετική ψήφο για το ίδιο patch. Η πιθανοτική ψήφος του patch υπολογίζεται σ' αυτήν την περίπτωση ως ο μέσος όρος των ψήφων όλων των δέντρων του δάσους. Ο υπολογισμός του μέσου όρου έχει ως αποτέλεσμα τη μείωση της διακύμανσης των ψήφων που δίνουν τα δέντρα ενός δάσους για ένα συγκεκριμένο patch λόγω θορύβου. Για το σύνολο του δάσους η ψήφος ενός patch με κέντρο y υπολογίζεται ως εξής:

$$p(E(x) | I(y); \{T_i\}_{i=1}^T) = \frac{1}{T} \sum_{i=1}^T p(E(x) | I(y); T_i)$$

Γενικά, μεγαλύτερες ψήφους δίνουν τα patches που λαμβάνονται από το «εσωτερικό» του αντικειμένου για το δάσος της κατηγορίας στην οποία ανήκει το αντικείμενο. Αν για παράδειγμα, έχω μια εικόνα η οποία απεικονίζει ένα αυτοκίνητο και πάρω ένα patch το οποίο απεικονίζει μέρος του αυτοκινήτου, τότε αυτό το patch θα έχει μικρή απόσταση από τα object patches που χρησιμοποιήθηκαν για την κατασκευή του δάσους της κατηγορίας «αυτοκίνητο» και επομένως αναμένεται να φθάσει σε αυτό το δάσος σε φύλλα με μεγάλο CL, ενώ για τα δάση των υπόλοιπων κατηγοριών θα έχει μικρότερη απόσταση από τα background patches. Αντίστοιχα, τα background patches κατά τη διάσχιση των δέντρων είναι πιο πιθανό να καταλήξουν σε φύλλο όπου το CL είναι μικρό (κοντά στο 0), δηλ. σε φύλλο όπου τα περισσότερα παραδείγματα



εκπαίδευσης που κατάληξαν σ' αυτό κατά τη φάση της εκπαίδευσης ήταν background patches.

Τα στοχαστικά δάση τα οποία αντιστοιχούν κάθε τμήμα μιας εικόνας σε μια πιθανοτική ψήφο σχετικά με την πιθανή θέση του κέντρου του αντικειμένου αναφέρονται στη βιβλιογραφία ως *Hough Forests* [3]. Η μόνη διαφορά τους από τα Random Forests έγκειται στο γεγονός ότι αποθηκεύουν στα φύλλα των δέντρων εκτός από την κατανομή της κατηγορίας (*class distribution*) και τα διανύσματα απόκλισης των patches (*offset vectors*), τα οποία και χρησιμοποιούν στη διαδικασία της ψηφοφορίας (*voting procedure*).

- Τα κύρια πλεονεκτήματα της χρήσης των Hough Forests (εκτός από αυτά που έχουν αναφερθεί για τα Random Forests και ισχύουν και εδώ λόγω της ομοιότητας τους) είναι τα εξής: (α) τα δέντρα κατασκευάζονται με τέτοιο τρόπο ώστε να βελτιστοποιούν τη διαδικασία της ψηφοφορίας, δηλ. οι ψήφοι που δίνουν τα φύλλα να έχουν την μικρότερη δυνατή αβεβαιότητα, (β) η διαδικασία της εκπαίδευσης γίνεται με επίβλεψη, για κάθε training patch υπάρχει εκ των προτέρων η γνώση για το εάν είναι τμήμα του αντικειμένου ή του φόντου, (γ) παρουσιάζουν ανεκτικότητα σε σημαντικό βαθμό θορύβου και λαθών στα δεδομένα εκπαίδευσης όπως π.χ. απόκρυψη τμήματος του αντικειμένου.

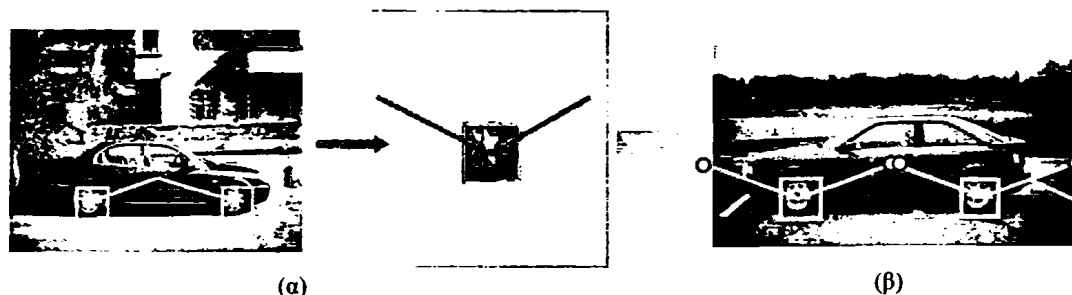
3^ο) Εκχώρηση των ψήφων σε διαφορετικά σημεία μιας αρχικά μαύρης εικόνας

Αφού υπολογιστεί η ψήφος ενός patch με κέντρο y , η τιμή της προστίθεται σε όλα τα pixels, μιας αρχικά μαύρης εικόνας, τα οποία ορίζονται από το σύνολο $\{y - d \mid d \in D_L\}$. Όπως φαίνεται και στο Σχήμα 3.12(α), κατά την εκπαίδευση του μοντέλου ταξινόμησης για κάθε training patch αποθηκεύουμε όλες τις πιθανές εμφανίσεις του καθώς και τα διανύσματα απόκλισης που αντιστοιχούν σε αυτές.

Κατά τη φάση του ελέγχου για κάθε «άγνωστη» εικόνα οι ψήφοι προστίθενται σε όλες τις πιθανές θέσεις εμφάνισης του αντικειμένου όπως φαίνεται στο Σχήμα 3.12(β). Στην περίπτωση που έχουμε ένα δάσος δυαδικών δέντρων η λίστα D_L είναι η ένωση των λιστών όλων των τερματικών κόμβων στους όποιους «έφθασε» το patch.

Αν θεωρήσουμε ότι έχουμε το δάσος του σχήματος 3.11 τότε $D_L = D_{L1} \cup D_{L2} \cup D_{L3}$.





Σχήμα 3.12 Εκχώρηση των Ψήφων σε Όλες τις Πιθανές Θέσεις Εμφάνισης του Αντικειμένου

4^ο) Συναθροίση των ψήφων όλων των patches

Από τις ψήφους όλων των test patches κατασκευάζεται ένας μηχανισμός απόφασης. Ο μηχανισμός αυτός είναι μια εικόνα η οποία έχει την ίδια διάσταση με την εικόνα ελέγχου και ονομάζεται *εικόνα Hough* [3, 4]. Η εικόνα αυτή υπολογίζεται συναθροίζοντας τις ψήφους όλων των test patches.

$$V(x) = \sum_y p(E(x) | I(y); \{T_i\}_{i=1}^T)$$

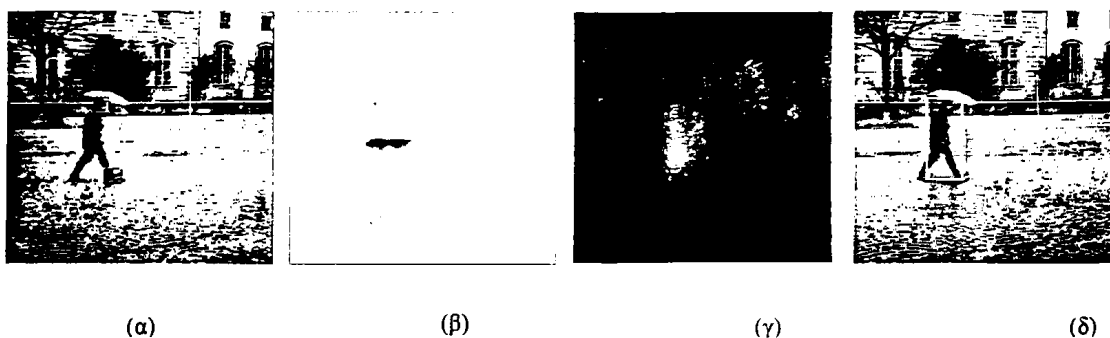
Για να βελτιώσουμε την ακρίβεια της ανίχνευσης μπορούμε να μειώσουμε τον θόρυβο της εικόνας Hough εφαρμόζοντας φίλτρο Gauss σε κάθε pixel της εικόνας, όπου σ είναι η τυπική απόκλιση της γκαουσιανής κατανομής.

$$p(E(x) | I(y); T) = \left[\frac{1}{|D_L|} \sum_{d \in D_L} \frac{1}{2\pi\sigma^2} \exp\left(-\frac{\|(y-x)-d\|^2}{2\sigma^2}\right) \right] \cdot C_L$$

Εύρεση της μέγιστης τιμής της εικόνας Hough

Οι τιμές των pixels της εικόνας Hough δείχνουν την πιθανότητα ύπαρξης του αντικειμένου ενδιαφέροντος σε κάθε σημείο. Σύμφωνα με τους Gall *et al.* [3] οι θέσεις των μεγίστων της εικόνας Hough αποτελούν υποθέσεις ανίχνευσης του κεντροειδούς του αντικειμένου και οι τιμές τους αποτελούν μέτρο εμπιστοσύνης για κάθε υπόθεση.





Σχήμα 3.13 Για καθένα από τα 3 Patches που Σημειώθηκαν στην Εικόνα (α) Υπολογίζονται από το Δάσος της Κατηγορίας «Pedestrians» οι Πιθανοτικοί Ψήφοι για τις Πιθανές Θέσεις του Κεντροειδούς του Πεζού (β). Συναθροίζονται οι Ψήφοι Όλων των Patches σε μια Εικόνα Hough (γ) Η Θέση του Πεζού Ανιχνεύεται με Βάση τη Μέγιστη Τιμή της Εικόνας Hough (δ)

3.2.2.3 Ταξινόμηση Πολλών Κατηγοριών (Multi-class Problem)

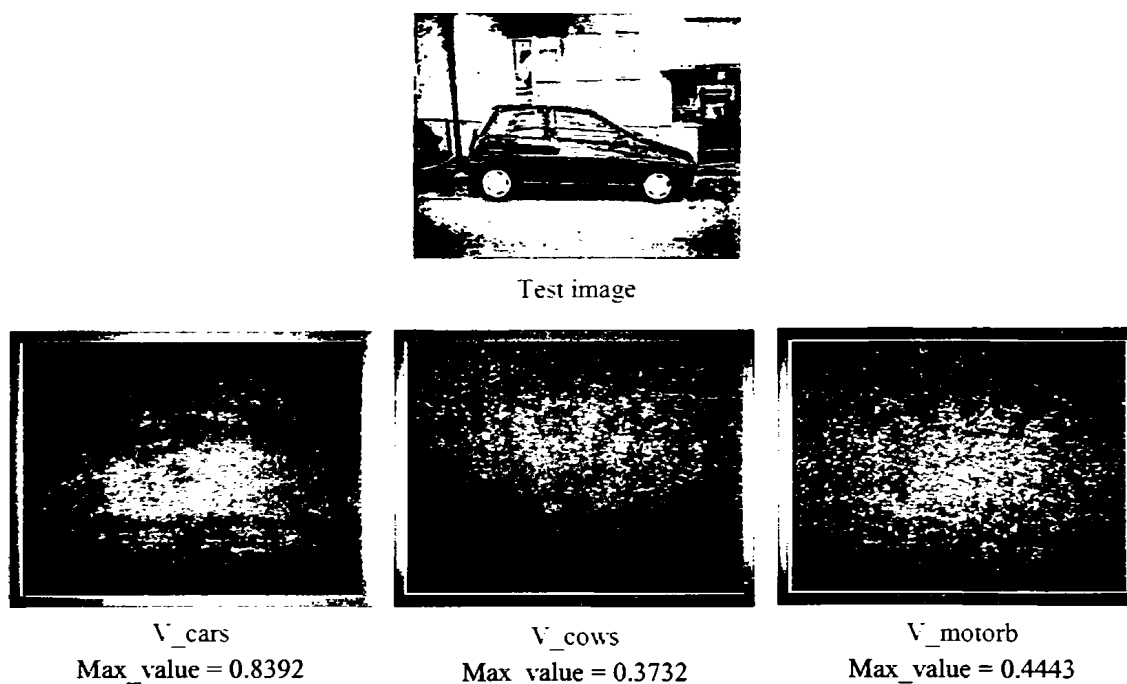
Σε περίπτωση που η απεικόνιση του αντικειμένου στην εικόνα δεν είναι βέβαιη τότε εκτός από τον εντοπισμό του κεντροειδούς του αντικειμένου, η μέγιστη τιμή παρέχει πληροφορίες και για την ύπαρξη ή μη του αντικειμένου στην εικόνα (όπως αποδεικνύεται πειραματικά).

Στην περίπτωση του προβλήματος ταξινόμησης σε πολλές κατηγορίες, κατασκευάζουμε μια εικόνα Hough από κάθε δάσος. Συνεπώς, για μια εικόνα ελέγχου κατασκευάζουμε τόσες εικόνες Hough όσες και οι προκαθορισμένες κατηγορίες του προβλήματος και κατ' επέκταση όσα και τα δάση του μοντέλου ταξινόμησης. Για την πρόβλεψη της κατηγορίας στην οποία ανήκει η εικόνα χρησιμοποιούμε τις μέγιστες τιμές των εικόνων Hough. Η εικόνα ελέγχου εκχωρείται στην κατηγορία η οποία δίνει τη μεγαλύτερη τιμή από τις μέγιστες τιμές των εικόνων Hough.

Ένα παράδειγμα φαίνεται στο Σχήμα 3.14. Παρατηρούμε ότι στην εικόνα Hough της κατηγορίας «Cars» τα σημεία είναι περισσότερο συγκεντρωμένα κοντά στο σημείο του κεντροειδούς του αντικειμένου. Αντίθετα, στις εικόνες των άλλων κατηγοριών τα σημεία είναι διάσπαρτα και δεν εντοπίζονται σε μια συγκεκριμένη περιοχή. Αυτό έχει



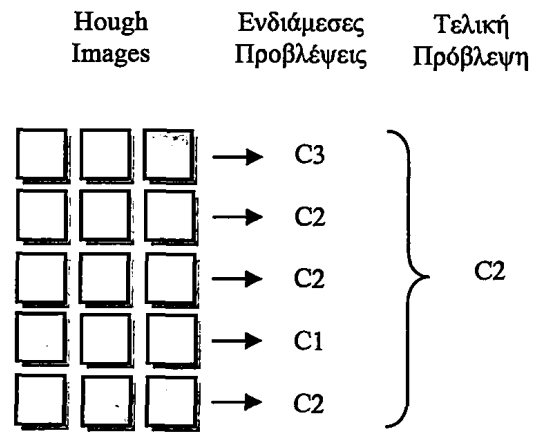
ως αποτέλεσμα η συνάθροιση των ψήφων των test patches να δίνει μεγαλύτερες τιμές στην εικόνα Hough της κατηγορίας ενδιαφέροντος.



Σχήμα 3.14 Οι Εικόνες Hough που Υπολογίστηκαν από το Δάσος κάθε Κατηγορίας.
Το Πράσινο Αντιστοιχεί στην Εικόνα Hough της Κατηγορίας «Cars»

Για να βελτιώσουμε την ακρίβεια της πρόβλεψης, κάνουμε λαμβάνουμε test patches περισσότερες από μια φορές και κάθε φορά κατασκευάζουμε ένα διαφορετικό σύνολο εικόνων Hough. Η τελική πρόβλεψη δίνεται με βάση την επικρατούσα κατηγορία δηλ. την κατηγορία η οποία έδωσε τις περισσότερες φορές τη μεγαλύτερη τιμή συγκρίνοντας τις μέγιστες τιμές των εικόνων Hough. Για παράδειγμα, αν έχω ένα σύνολο κατηγοριών $C = \{C1, C2, C3\}$ και κάνω δειγματοληψία 5 φορές θα δημιουργηθεί ένα σύνολο 15 εικόνων Hough. Από κάθε δειγματοληψία δημιουργείται ένα σύνολο 3 εικόνων Hough, μια από κάθε δάσος-κατηγορία, και προκύπτει ένα ενδιάμεσο αποτέλεσμα. Η εικόνα ανατίθεται τελικά στην κατηγορία που εμφανίζεται περισσότερες φορές. Αν δυο κατηγορίες ισοψηφούν τότε δεν μπορούμε να αποφανθούμε για την κατηγορία της εικόνας. Στο παρακάτω σχήμα απεικονίζεται σχηματικά η διαδικασία απόφασης.





Σχήμα 3.15 Διαδικασία Απόφασης

ΚΕΦΑΛΑΙΟ 4. ΠΕΙΡΑΜΑΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ

4.1 Γενικά

4.2 Πειραματική Εκτίμηση της Απόδοσης της Προτεινόμενης Μεθόδου

4.3 Πειραματικά Δεδομένα

4.4 Διερεύνηση της Δυνατότητας Χρήσης της Μεθόδου για Ανάκτηση Εικόνων

4.1. Γενικά

Τα πειράματα που πραγματοποιήθηκαν είχαν ως στόχο την εκτίμηση της απόδοσης της προτεινόμενης μεθόδου σε πρόβλημα ταξινόμησης με βάση το περιεχόμενο σε μια από τις προκαθορισμένες κατηγορίες (*multi-class classification*) καθώς και τη σύγκριση της απόδοσης της μεθόδου με μία από τις πιο γνωστές μεθόδους ταξινόμησης εικόνων, τη μέθοδο Bag of Keyoints. Το μέτρο εκτίμησης της μεθόδου είναι το ποσοστό επιτυχούς ταξινόμησης στο σύνολο ελέγχου, δηλ. το ποσοστό των παραδειγμάτων που η πραγματική τους κατηγορία ταυτίζεται με εκτιμώμενη κατηγορία.

4.2. Πειραματική Εκτίμηση της Απόδοσης της Προτεινόμενης Μεθόδου

4.2.1 Πειραματική Διαδικασία

Για την εκτίμηση της απόδοσης της μεθόδου χρησιμοποιήθηκε το σύνολο δεδομένων TU Darmstadt Database (η περιγραφή του συνόλου δίνεται στην ενότητα 4.3). Το μοντέλο ταξινόμησης που κατασκευάστηκε περιλαμβάνει τρία διακριτά δάση δυαδικών δέντρων απόφασης, ένα για καθεμία από τις κατηγορίες {Cars, Cows, Motorbikes}. Για τον ορισμό της θέσης του αντικειμένου ενδιαφέροντος στις εικόνες εκπαίδευσης χρησιμοποιήθηκε για την κατηγορία «Motorbikes» η πληροφορία των



σημείων που ορίζουν το αντίστοιχο πλαίσιο οριοθέτησης και για τις άλλες δυο κατηγορίες χρησιμοποιήθηκαν οι αντίστοιχες μάσκες κατάτμησης εικονοστοιχείων. Κάθε δέντρο εκπαιδεύτηκε χρησιμοποιώντας 15000 object και 15000 background patches. Τα object patches επιλέχθηκαν από τις εικόνες εκπαίδευσης της κατηγορίας ενδιαφέροντος με ίση αναλογία μεταξύ των εικόνων π.χ. αν το σύνολο εκπαίδευσης περιλαμβάνει 30 εικόνες της κατηγορίας ενδιαφέροντος τότε επιλέγουμε 500 patches από κάθε εικόνα. Τα background patches επιλέγονται από όλες τις εικόνες του συνόλου εκπαίδευσης με ίση αναλογία μεταξύ τους. Για τις εικόνες της κατηγορίας ενδιαφέροντος λαμβάνω background patches από το εξωτερικό του πλαισίου οριοθέτησης του αντικειμένου. Η διάσταση των patches είναι 16×16 και το πλήθος των δυαδικών δέντρων σε κάθε δάσος είναι 10. Ένας κόμβος δηλώνεται ως τερματικός αν βρίσκεται σε βάθος $d_{max}=15$ ή αν το πλήθος των patches που του έχουν ανατεθεί είναι $N_{min} \leq 20$. Σε κάθε μη τερματικό κόμβο εφαρμόζονται 20 εκτελέσεις με διαφορετική αρχικοποίηση του αλγόριθμου k-means με μέτρο απόστασης την απόσταση συνημίτονου. Οι τιμές των παραμέτρων για τη διάσταση των patches, το μέγιστο βάθος των δέντρων και το πλήθος των training patches έχουν προσδιοριστεί πειραματικά.

Για τη διανυσματική περιγραφή ενός patch χρησιμοποιήθηκαν 9 χαρακτηριστικά: τα τρία κανάλια χρώματος (R, G, B) και οι απόλυτες τιμές των πρώτων x - και y -παραγώγων για κάθε κανάλι. Η περιγραφή των patches έγινε με ένα διάνυσμα διάστασης 1×18 το οποίο υπολογίστηκε χρησιμοποιώντας το μέσο όρο και την τυπική απόκλιση για καθένα από τα εξαγόμενα χαρακτηριστικά.

Για την ταξινόμηση μιας εικόνας από το σύνολο ελέγχου κάναμε δειγματοληψία 10 φορές, επιλέγοντας κάθε φορά 5000 σημεία της εικόνας. Δημιουργούμε με αυτόν τον τρόπο ένα σύνολο 30 (3×10) εικόνων Hough. Η εικόνα ανατίθεται στην κατηγορία που δίνει τις περισσότερες φορές τη μεγαλύτερη τιμή από τις μέγιστες τιμές των εικόνων Hough (η διαδικασία απόφασης απεικονίζεται στο Σχήμα 3.13).



4.2.2 Εκτίμηση της Απόδοσης της Μεθόδου

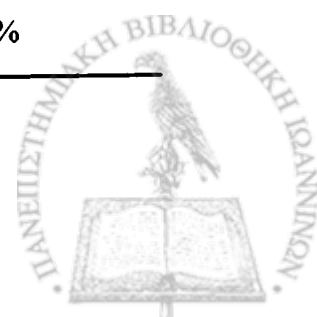
Η εκτίμηση της απόδοσης της μεθοδολογίας έγινε χρησιμοποιώντας 10-fold cross validation. Από το σύνολο των 300 εικόνων εκπαίδευσης (100 από κάθε κατηγορία) χρησιμοποιήσαμε 30 εικόνες για έλεγχο (10 από κάθε κατηγορία) και τις υπόλοιπες για εκπαίδευση. Η συνολική απόδοση του ταξινομητή είναι 90.33% και υπολογίζεται από το παρακάτω τύπο:

$$acc = \frac{1}{10} \sum_{i=1}^{10} \frac{num_of_correct_predictions}{total_num_of_predictions}$$

Στον Πίνακα 4.1 φαίνεται αναλυτικά το πλήθος των σωστών και εσφαλμένων προβλέψεων για κάθε κατηγορία και η αντίστοιχη ακρίβεια ταξινόμησης όπως υπολογίστηκαν από την παραπάνω διαδικασία. Παρατηρούμε ότι για τις κατηγορίες Cars και Cows το ποσοστό των σωστών προβλέψεων είναι μεγαλύτερο από την κατηγορία Motorbikes. Αυτό οφείλεται στην μεγαλύτερη ποικιλομορφία που παρουσιάζουν οι εικόνες της συγκεκριμένης κατηγορίας ως προς το φωτισμό, την κλίμακα και τον προσανατολισμό των αντικείμενων καθώς και στην παρουσία θορύβου σε κάποιες από αυτές. Επίσης παρατηρούμε ότι οι εικόνες που ανήκουν στην κατηγορία Cows είναι περισσότερο διακριτές από τις εικόνες των άλλων δυο κατηγοριών.

Πίνακας 4.1 Confusion Matrix και η Ακρίβεια Ταξινόμησης ανά Κατηγορία (για 18 Χαρακτηριστικά)

Προβλεπόμενες Κατηγορίες →	Cars	Cows	Motorbikes	Ακρίβεια Ταξινόμησης/ Κατηγορία
Cars	91	0	9	91%
Cows	0	97	3	97%
Motorbikes	16	1	83	83%
Συνολική Ακρίβεια				90.33%



Για να μελετήσουμε την επίδραση του πλήθους των χαρακτηριστικών στην απόδοση της μεθόδου ακολουθήσαμε την πειραματική διαδικασία που περιγράφεται στην ενότητα 4.2.1 χρησιμοποιώντας μικρότερο πλήθος χαρακτηριστικών. Συγκεκριμένα, χρησιμοποιήσαμε μόνο τα τρία κανάλια χρώματος (R,G,B). Η περιγραφή των patches έγινε με ένα διάνυσμα διάστασης 1×6 το οποίο υπολογίστηκε χρησιμοποιώντας το μέσο όρο και την τυπική απόκλιση για καθένα από τα εξαγόμενα χαρακτηριστικά. Η συνολική απόδοση του ταξινομητή σ' αυτή την περίπτωση είναι 86%.

Πίνακας 4.2 Confusion Matrix και η Ακρίβεια Ταξινόμησης ανά Κατηγορία (για 6 Χαρακτηριστικά)

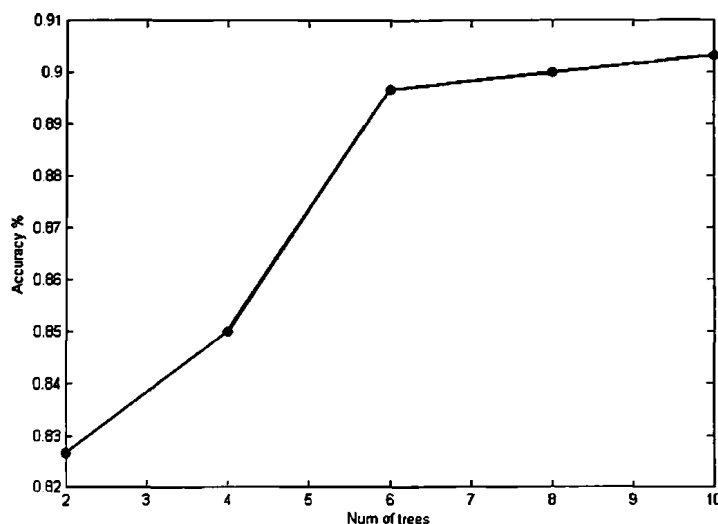
Προβλεπόμενες Κατηγορίες →	Cars	Cows	Motorbikes	Ακρίβεια Ταξινόμησης/Κατηγορία
Cars	91	0	9	91%
Cows	0	100	0	100%
Motorbikes	31	2	67	67%
			Συνολική Ακρίβεια	86%

Στον Πίνακα 4.2 φαίνεται αναλυτικά το πλήθος των σωστών και εσφαλμένων προβλέψεων για κάθε κατηγορία και η αντίστοιχη ακρίβεια ταξινόμησης. Παρατηρούμε ότι για τις κατηγορίες Cars και Cows το ποσοστό των σωστών προβλέψεων και σ' αυτή την περίπτωση είναι μεγαλύτερο από την κατηγορία Motorbikes. Οι εικόνες που ανήκουν στην κατηγορία Cows είναι περισσότερο διακριτές από τις εικόνες των άλλων δυο κατηγοριών. Παρατηρούμε επίσης ότι η χρήση λιγότερων χαρακτηριστικών έχει ως αποτέλεσμα την μείωση της συνολικής απόδοσης της μεθόδου η οποία οφείλεται στη μείωση της ακρίβειας ταξινόμησης για την κατηγορία Motorbikes. Συμπεραίνουμε λοιπόν ότι η χρήση επιπλέον χαρακτηριστικών είναι σημαντική σε περιπτώσεις όπου υπάρχει μεγάλη ποικιλομορφία στις εικόνες μιας κατηγορίας. Στα πειράματα που ακολουθούν για τη διανυσματική περιγραφή των patches χρησιμοποιήσαμε τα 18 χαρακτηριστικά (R,G,B , x - και y - πρώτες παραγώγους) όπως περιγράφηκαν παραπάνω.



4.2.3 Επίδραση του Πλήθους των Δυαδικών Δέντρων Απόφασης σε κάθε Δάσος

Για την εκτίμηση της επίδρασης του πλήθους των δυαδικών δέντρων απόφασης χρησιμοποιήσαμε το μοντέλο που περιγράφηκε στην ενότητα 4.2.1. Χρησιμοποιήσαμε 10-fold cross validation για διαφορετικές τιμές πλήθους δέντρων. Συγκεκριμένα, για 2, 4, 6, 8 και 10 δυαδικά δέντρα απόφασης σε κάθε δάσος.



Σχήμα 4.1 Γραφική Παράσταση για την Εκτίμηση της Επίδρασης του Πλήθους των Δυαδικών Δέντρων Απόφασης σε κάθε Δάσος στην Ακρίβεια της Ταξινόμησης

Από το Σχήμα 4.1 παρατηρούμε ότι η αύξηση του πλήθους των δυαδικών δέντρων απόφασης σε κάθε δάσος έχει ως αποτέλεσμα τη βελτίωση της απόδοσης της μεθόδου. Πιο συγκεκριμένα, για 2 δυαδικά δέντρα σε κάθε δάσος η ακρίβεια ταξινόμησης είναι 82.7% και για τιμή πλήθους 6 φθάνει στο 89.7%. Για τιμές πλήθους 8 και 10 δέντρων η απόδοση της μεθόδου είναι 90% και 90.3% αντίστοιχα. Υπάρχει δηλ. βελτίωση όχι όμως τόσο σημαντική. Συνεπώς, η απόδοση της προτεινόμενης μεθόδου είναι υψηλή ακόμα και στην περίπτωση που τα δάση τα οποία χρησιμοποιούμε περιλαμβάνουν μόνο 2 δυαδικά δέντρα απόφασης και με την αύξηση του πλήθους των δέντρων η απόδοση αυξάνεται χωρίς να παρατηρούνται φαινόμενα υπερεκπαίδευσης.

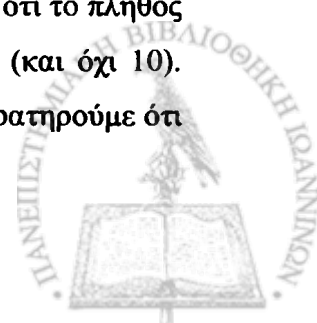


4.2.4 Σύγκριση της Απόδοσης της Προτεινόμενης Μεθόδου με τη Μέθοδο Bag of Keypoints

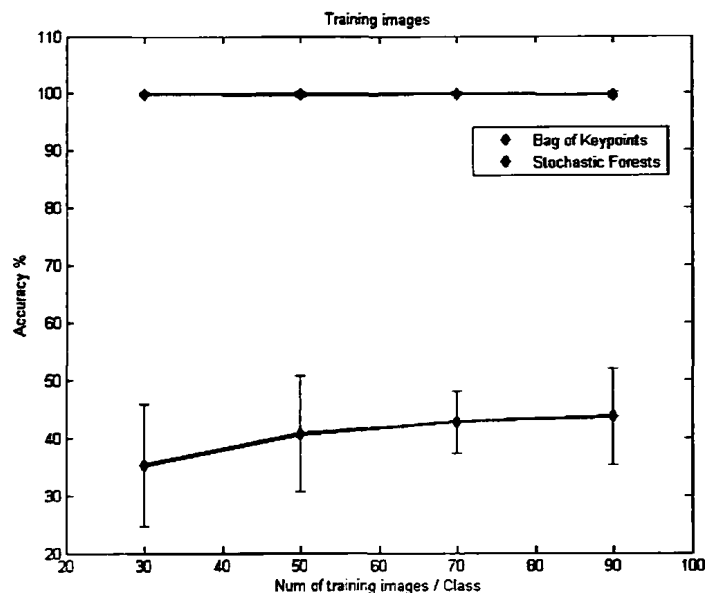
Η απόδοση των μεθόδων ελέγχθηκε σε ένα πρόβλημα ταξινόμησης τριών κατηγοριών. Το σύνολο των εικόνων εκπαίδευσης που χρησιμοποιήθηκε και για τις δυο μεθόδους είναι το TU Darmstadt Database. Για να υπολογίσουμε την απόδοση των δύο μεθόδων κατασκευάστηκαν μοντέλα ταξινόμησης για διαφορετικό πλήθος εικόνων εκπαίδευσης. Συγκεκριμένα, χρησιμοποιήθηκαν 30, 50, 70 και 90 εικόνες για εκπαίδευση από κάθε κατηγορία, ενώ σε κάθε περίπτωση οι υπόλοιπες εικόνες, από το αρχικό σύνολο των 100 εικόνων ανά κατηγορία, χρησιμοποιήθηκαν για την εκτίμηση της ακρίβεια ταξινόμησης (συγκεκριμένα 70, 50, 30, και 10 εικόνες). Για να πάρουμε πιο αξιόπιστα αποτελέσματα, για κάθε διαφορετικό πλήθος εικόνων κατασκευάσαμε 10 διαφορετικά μοντέλα ταξινόμησης επιλέγοντας κάθε φορά το σύνολο εκπαίδευσης με τυχαίο τρόπο. Για καθένα από αυτά τα 10 μοντέλα υπολογίσαμε την απόδοσή τους (τόσο στο σύνολο εκπαίδευσης όσο και στο σύνολο ελέγχου) και στη συνέχεια υπολογίστηκαν οι μέσοι όροι και οι τυπικές αποκλίσεις των επί μέρους αποδόσεων.

Για την κατασκευή των μοντέλων της μεθόδου Bag of Keypoints χρησιμοποιήσαμε τον αλγόριθμο SIFT [28] ο οποίος εντόπισε περίπου 60000 patches στην περίπτωση των 30 εικόνων εκπαίδευσης ανά κατηγορία, περίπου 100000 patches για τις 50 εικόνες εκπαίδευσης ανά κατηγορία, περίπου 140000 patches για τις 70 εικόνες εκπαίδευσης ανά κατηγορία και περίπου 180000 patches για τις 90 εικόνες εκπαίδευσης ανά κατηγορία. Τα patches ομαδοποιήθηκαν με τον αλγόριθμο k-means σχηματίζοντας ένα λεξικό 100 οπτικών λέξεων. Χρησιμοποιήθηκε το λογισμικό LIBSVM [37, 38] το οποίο υποστηρίζει την τεχνική Ένας-Εναντίον-Ενός. Επιλέξαμε να χρησιμοποιήσουμε τον πυρήνα RBF καθώς έχει αποδειχθεί ότι στις περισσότερες περιπτώσεις δίνει καλύτερα αποτελέσματα από τον γραμμικό [40]. Η επιλογή των βέλτιστων παραμέτρων (C , g) έγινε χρησιμοποιώντας 10-fold cross validation.

Για την κατασκευή των μοντέλων της προτεινόμενης μεθόδου χρησιμοποιήσαμε τη διαδικασία που περιγράφεται στην ενότητα 4.2.1. Η μόνη διαφορά είναι ότι το πλήθος των δυαδικών δέντρων που χρησιμοποιήσαμε για κάθε δάσος είναι 6 (και όχι 10). Χρησιμοποιήσαμε αυτό το πλήθος δέντρων καθώς από το Σχήμα 4.1 παρατηρούμε ότι



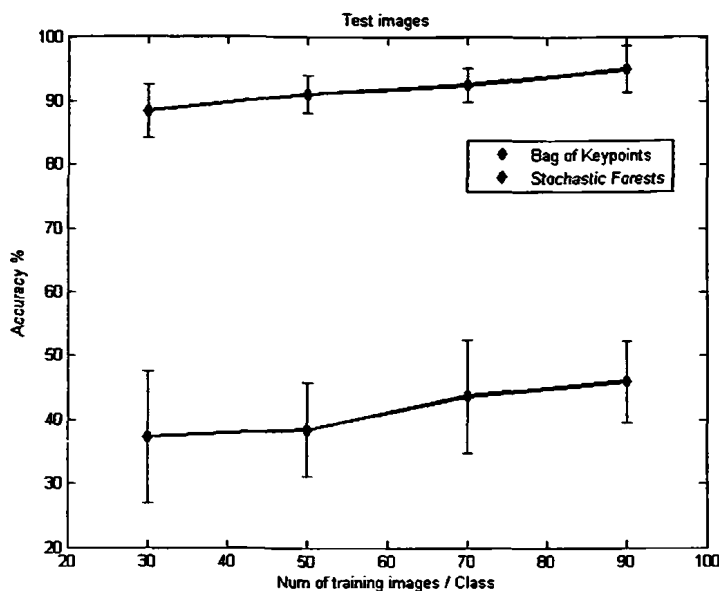
η απόδοση της μεθόδου για αυτή την τιμή είναι υψηλή, ενώ αν χρησιμοποιήσουμε 10 δέντρα για κάθε δάσος η απόδοση δεν βελτιώνεται σημαντικά ενώ αυξάνεται η απαιτούμενη μνήμη και ο χρόνος υπολογισμού.



Σχήμα 4.2 Γραφική Παράσταση για την Σύγκριση της Απόδοσης της Προτεινόμενης Μεθόδου με τη Μέθοδο Bag of Keypoints στο Σύνολο των Εικόνων Εκπαίδευσης για Διαφορετικό Πλήθος Εικόνων Εκπαίδευσης

Από το παραπάνω σχήμα παρατηρούμε ότι για σύνολα εκπαίδευσης πλήθους 30, 50, 70 και 90 εικόνων η προτεινόμενη μέθοδος εμφανίζει πολύ πιο υψηλή απόδοση από την μέθοδο Bag of Keypoints στα σύνολα εκπαίδευσης. Για την προτεινόμενη μέθοδο η απόδοση για κάθε πλήθος είναι κοντά στο 99.8% και η τυπική απόκλιση σχεδόν μηδενική. Συμπεραίνουμε λοιπόν ότι το μοντέλο που κατασκευάζει η μέθοδος έχει «μάθει» καλά τα δεδομένα εκπαίδευσης καθώς παρουσιάζει μεγάλη ακρίβεια στις εικόνες εκπαίδευσης και ως εκ τούτου μικρή πιθανότητα σφάλματος (training error). Από την άλλη πλευρά η μέθοδος Bag of Keypoints παρουσιάζει πολύ μικρότερη απόδοση και μεγαλύτερη τυπική απόκλιση. Η απόδοση αυτής της μεθόδου αυξάνεται με το πλήθος των εικόνων εκπαίδευσης ξεκινώντας από το 35% με τυπική απόκλιση 10.6 για πλήθος 30 εικόνων εκπαίδευσης και φθάνει στο 44% με τυπική απόκλιση 8.4 για πλήθος 90 εικόνων, ενώ τη μικρότερη διακύμανση 5.4 παρουσιάζει για πλήθος 70 εικόνων.





Σχήμα 4.3 Γραφική Παράσταση για την Σύγκριση της Απόδοσης της Προτεινόμενης Μεθόδου με τη Μέθοδο Bag of Keypoints στο Σύνολο των Εικόνων Ελέγχου για Διαφορετικό Πλήθος Εικόνων Εκπαίδευσης

Από το σχήμα παρατηρούμε ότι η προτεινόμενη μέθοδος εμφανίζει πολύ πιο υψηλή απόδοση από την μέθοδο Bag of Keypoints και στα αντίστοιχα σύνολα ελέγχου. Και για τις δύο μεθόδους η απόδοση αυξάνεται καθώς αυξάνεται το πλήθος των εικόνων εκπαίδευσης. Συγκεκριμένα, για την προτεινόμενη μέθοδο η απόδοση κυμαίνεται από 88.5% για 30 εικόνες εκπαίδευσης και φθάνει στο 95% για 90 εικόνες εκπαίδευσης ενώ η τυπική απόκλιση είναι κοντά στο 3.5. Για την μέθοδο Bag of Keypoints, η απόδοση κυμαίνεται από 37% για 30 εικόνες εκπαίδευσης και φθάνει στο 46% για 90 εικόνες εκπαίδευσης ενώ η τυπική απόκλιση είναι κοντά στο 8.

Όπως αναφέρθηκε και στην ενότητα 3.2.2.3 για να βελτιώσουμε την ακρίβεια πρόβλεψης της προτεινόμενης μεθοδολογίας η διαδικασία απόφασης περιλαμβάνει την κατασκευή πολλαπλών μηχανισμών απόφασης δηλ. διαφορετικών συνόλων εικόνων Hough (όπως απεικονίζεται στο Σχήμα 3.13). Αυτό όμως έχει ως αποτέλεσμα, εκτός από την αύξηση της απόδοσης, και την αύξηση του χρόνου ταξινόμησης μιας εικόνας. Αντίθετα, ο χρόνος ταξινόμησης για τη μέθοδο Bag of Keypoints είναι πολύ μικρότερος.



4.2.5 Σύγκριση της Απόδοσης της Προτεινόμενης Μεθόδου σε Διαφορετικά Σύνολα Δεδομένων

Στην ενότητα αυτή συγκρίνεται η απόδοση της μεθόδου σε προβλήματα ταξινόμησης δύο και τριών κατηγοριών. Στον Πίνακα 4.3 περιγράφονται τα σύνολα δεδομένων που χρησιμοποιήθηκαν. Οι εικόνες που περιλαμβάνονται στα παρακάτω σύνολα δεδομένων ανήκουν στις βάσεις TU Darmstadt Database [39] και TUD Pedestrians [11]. Για να υπολογίσουμε την απόδοση της μεθόδου κατασκευάσαμε 10 διαφορετικά μοντέλα ταξινόμησης για κάθε σύνολο επιλέγοντας κάθε φορά με τυχαίο τρόπο 90 εικόνες/κατηγορία για το σύνολο εκπαίδευσης και 10 εικόνες (διαφορετικές από αυτές του συνόλου εκπαίδευσης) για το σύνολο ελέγχου. Κάθε μοντέλο ταξινόμησης που κατασκευάστηκε περιλαμβάνει ένα δάσος δυαδικών δέντρων απόφασης για καθεμία από τις κατηγορίες του συνόλου δεδομένων. Για την κατασκευή των δασών και την ταξινόμηση των εικόνων ελέγχου ακολουθήθηκε η διαδικασία που περιγράφεται στην ενότητα 4.2.1.

Πίνακας 4.3 Περιγραφή των Συνόλων Δεδομένων

Σύνολο δεδομένων	Πλήθος εικόνων		
	Cars	Cows	Pedestrians
Cars-Pedestrians210	100	-	210
Cars-Pedestrians400	100	-	400
Cows-Pedestrians210	-	100	210
Cows-Pedestrians400	-	100	400
Cars-Cows-Pedestrians210	100	100	210
Cars-Cows-Pedestrians400	100	100	400

Στον Πίνακα 4.4 απεικονίζεται η απόδοση της μεθόδου για κάθε σύνολο δεδομένων. Παρατηρούμε ότι τα σύνολα δεδομένων που περιλαμβάνουν το σύνολο Pedestrians210 εμφανίζουν μικρότερη απόδοση από τα αντίστοιχα σύνολα δεδομένων που περιλαμβάνουν το Pedestrians400. Επίσης, παρατηρούμε ότι υπάρχει μεγαλύτερη διάκριση μεταξύ των κατηγοριών Cows και Pedestrians από ότι μεταξύ των



κατηγοριών Cars και Pedestrians. Αυτό οφείλεται κυρίως στο ότι το χρωματικό περιεχόμενο των εικόνων της κατηγορίας Cows διαφέρει σημαντικά από το χρωματικό περιεχόμενο των άλλων κατηγοριών καθώς και στο ότι υπάρχουν εικόνες της κατηγορίας Pedestrians στο σύνολο εκπαίδευσης στις οποίες εκτός από το αντικείμενο ενδιαφέροντος (πεζός) απεικονίζονται και τμήματα αυτοκινήτων.

Πίνακας 4.4 Ακρίβεια Ταξινόμησης της Προτεινόμενης Μεθόδου σε Διαφορετικά Σύνολα Δεδομένων

Σύνολο δεδομένων	Cars	Cows	Pedestrians	Συνολική Ακρίβεια Ταξινόμησης
Cars-Pedestrians210	100%	-	78%	89%
Cars-Pedestrians400	100%	-	96.5%	96.5%
Cows-Pedestrians210	-	100%	93%	96.5%
Cows-Pedestrians400	-	100%	96.5%	99%
Cars-Cows-Pedestrians210	99%	95%	46%	80%
Cars-Cows-Pedestrians400	97%	99%	86%	94%

4.3. Πειραματικά Δεδομένα

4.3.1 Το Σύνολο Δεδομένων TU Darmstadt Database (παλαιότερα ETHZ Database)

Η συγκεκριμένη βάση ανήκει στη συλλογή εικόνων “The PASCAL Object Recognition Database Collection” [39] και περιέχει εικόνες τριών κατηγοριών $C=\{Cars, Cows, Motorbikes\}$.

Συγκεκριμένα, η βάση περιέχει πλάγιες όψεις 50 αυτοκινήτων τα οποία έχουν αντικατοπτριστεί για να προκύψουν συνολικά 100 εικόνες. Όλα τα αυτοκίνητα έχουν περίπου το ίδιο μέγεθος και εμφανίζονται στο κέντρο της εικόνας. Υπάρχει μόνο ένα αυτοκίνητο ανά εικόνα, ενώ όλες οι εικόνες έχουν σχεδόν την ίδια (σχετικά μικρή) ανάλυση (περίπου 250x200).





Σχήμα 4.4 Το Σύνολο Δεδομένων για την Κατηγορία «Cars»

Για τη δεύτερη κατηγορία, το αντικείμενο ενδιαφέροντος (αγελάδες) έχει σε όλες τις εικόνες περίπου το ίδιο μέγεθος και τον ίδιο προσανατολισμό (πλαϊνή όψη, προς τα αριστερά). Υπάρχουν 111 εικόνες με αγελάδες (από τις οποίες χρησιμοποιήθηκαν 100) με μόνο 3 διαφορετικά περιβάλλοντα. Πολλές από τις εικόνες είναι αρκετά παρόμοιες με τουλάχιστον μια άλλη εικόνα στη βάση δεδομένων, ενώ όλες έχουν σχεδόν την ίδια ανάλυση (περίπου 250x200).



Σχήμα 4.5 Το Σύνολο Δεδομένων για την Κατηγορία «Cows»

Οι 115 εικόνες με μηχανάκια που παρέχει η βάση (από τις οποίες χρησιμοποιήθηκαν 100) παρουσιάζουν περισσότερη ποικιλία όσον αφορά το φωτισμό, την κλίμακα και τον προσανατολισμό των μηχανών. Απεικονίζουν καθημερινές σκηνές ανθρώπων με τα μηχανάκια τους. Σε κάποιες από τις εικόνες μπορεί να υπάρχει μερική απόκρυψη ή να εμφανίζονται πολλαπλά αντικείμενα (περισσότερα από ένα μηχανάκια). Η ανάλυση των εικόνων κυμαίνεται από 200x150 έως 640x480.



Σχήμα 4.6 Το Σύνολο Δεδομένων για την Κατηγορία «Motorbikes»



Η θέση των αντικειμένων δίνεται για τις εικόνες και των τριών κατηγοριών μέσω σημείων τα οποία ορίζουν τα πλαίσια οριοθέτησης των αντικειμένων, ενώ για τις εικόνες των κατηγοριών Cars και Cows υπάρχουν και οι αντίστοιχες μάσκες κατάτμησης εικονοστοιχείων. Η συλλογή εικόνων PASCAL παρέχει και εργαλεία για την επισήμανση των αντικειμένων ενδιαφέροντος σε εικόνες και τη φόρτωση των πληροφοριών επισήμανσης στην MATLAB.

4.3.2 Το Σύνολο Δεδομένων TUD Pedestrians

Η συγκεκριμένη βάση [11] παρέχει δύο σύνολα εκπαίδευσης. Το ένα σύνολο αποτελείται από 210 εικόνες ενώ το άλλο από 400 εικόνες (το πρώτο σύνολο εικόνων είναι υποσύνολο του δεύτερου). Οι εικόνες απεικονίζουν πεζούς οι οποίοι είναι πλήρως ορατοί, σε πλάγια όψη, σε τέσσερα διαφορετικά περιβάλλοντα. Η θέση των πεζών δίνεται μέσω μασκών κατάτμησης εικονοστοιχείων. Η ανάλυση των εικόνων κυμαίνεται από 321x257 έως 582x466.

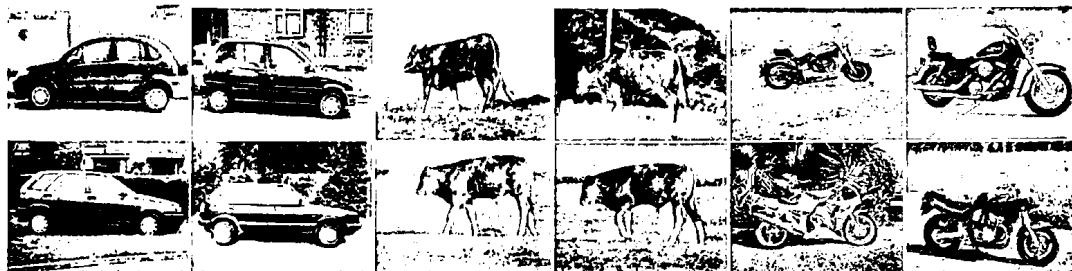


Σχήμα 4.7 Το Σύνολο Εικόνων TUD Pedestrians

4.3.2 Εικόνες από τις Βάσεις Pascal VOC 2006 Database και Caltech

Οι βάσεις VOC 2006 και Caltech ανήκουν στη συλλογή εικόνων “The PASCAL Object Recognition Database Collection” [39]. Περιλαμβάνουν εικόνες διαφόρων κατηγοριών όπως μηχανάκια, αεροπλάνα, πρόσωπα κ.α. Για κάθε εικόνα δίνεται η πληροφορία της θέσης του αντικειμένου ενδιαφέροντος μέσω των σημείων που ορίζουν το αντίστοιχο πλαίσιο οριοθέτησης. Το αντικείμενο ενδιαφέροντος είναι πάντα ένα σε κάθε εικόνα και εμφανίζεται στο κέντρο της. Η ανάλυση των εικόνων που χρησιμοποιήσαμε είναι περίπου 640x480.





Σχήμα 4.8 Εικόνες από Pascal VOC 2006 και Caltech

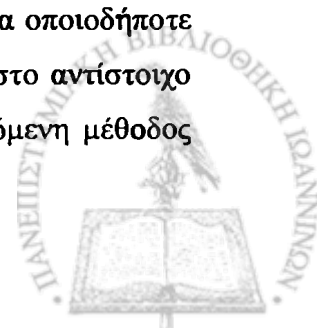
4.4. Διερεύνηση της Δυνατότητας Χρήσης της Μεθόδου για Ανάκτηση Εικόνων

Στα πλαίσια της εργασίας μελετήθηκε και η δυνατότητα χρήσης της μεθόδου για ανάκτηση εικόνων. Συγκεκριμένα, διερευνήθηκε η δυνατότητα εύρεσης μιας τιμής κατωφλίου (threshold) η οποία θα μπορούσε να διαχωρίσει τις εικόνες μιας κατηγορίας από τις υπόλοιπες. Η τιμή αυτή αναφέρεται στις μέγιστες τιμές των εικόνων Hough. Η μέθοδος όπως έχει αναφερθεί, χρησιμοποιεί για την αναπαράσταση της εικόνας μόνο την πληροφορία χρώματος. Συνήθως, εικόνες που ανήκουν στην ίδια κατηγορία παρουσιάζουν μεγάλη ομοιότητα ως προς το χρωματικό τους περιεχόμενο. Για να εξετάσουμε το παραπάνω ζήτημα κατασκευάσαμε ένα μοντέλο χρησιμοποιώντας 50 εικόνες από τη βάση TU Darmstadt Database για κάθε κατηγορία καθώς και 20 επιπλέον εικόνες για κάθε κατηγορία από τις βάσεις Pascal VOC2006 και Caltech. Η διαδικασία κατασκευής του μοντέλου ακολουθεί τον τρόπο που περιγράφηκε στην ενότητα 4.2.1. Το ίδιο σύνολο εικόνων χρησιμοποιήθηκε στη συνέχεια ως σύνολο ελέγχου για να εξετάσουμε αν μπορεί να βρεθεί ένα threshold το οποίο να αποτελεί κάτω φράγμα για τις μέγιστες τιμές των εικόνων Hough μιας δεδομένης κατηγορίας. Παρατηρήσαμε ότι η ύπαρξη μιας τέτοιας τιμής είναι εφικτή μόνο στην περίπτωση που η απεικόνιση του αντικειμένου ενδιαφέροντος (κλίμακα, χρώμα κτλ) στις εικόνες δεν παρουσιάζει μεγάλες διαφοροποιήσεις. Συμπεραίνουμε λοιπόν ότι χρησιμοποιώντας μόνο την πληροφορία χρώματος για την αναπαράσταση των εικόνων, δεν είναι δυνατή η εύρεση μιας τιμής κατωφλίου η οποία να δίνει ενδείξεις για το αν απεικονίζεται το αντικείμενο ενδιαφέροντος στην εικόνα.

ΚΕΦΑΛΑΙΟ 5. ΣΥΜΠΕΡΑΣΜΑΤΑ - ΜΕΛΛΟΝΤΙΚΗ ΕΡΓΑΣΙΑ

Στην παρούσα εργασία μελετήθηκε το πρόβλημα της ταξινόμησης εικόνων με βάση το περιεχόμενο σε μία από τις προκαθορισμένες κατηγορίες. Η μέθοδος που προτάθηκε για την επίλυσή του στηρίζεται στην χρήση στοχαστικών δασών δυαδικών δέντρων απόφασης. Παρουσιάζεται μια μεθοδολογία κατασκευής στοχαστικών δασών χρησιμοποιώντας τον αλγόριθμο ομαδοποίησης K-means για τη δυαδική διαμέριση κάθε ενδιάμεσου κόμβου. Η συγκεκριμένη μέθοδος εκμεταλλεύεται τα πλεονεκτήματα των συνδυαστικών μεθόδων ταξινόμησης όπως τα Bagging Trees και τα Τυχαία Δάση για να επιτύχει υψηλή απόδοση. Σημαντικό ρόλο στην ακρίβεια της μεθόδου παίζει η εξαγωγή τοπικών χαρακτηριστικών χρώματος σε συνδυασμό με την χρήση πληροφορίας κατεύθυνσης. Για την εκτίμηση της απόδοσής της χρησιμοποιήθηκαν γνωστά σύνολα εικόνων. Από τα πειραματικά αποτελέσματα παρατηρούμε ότι η απόδοση της μεθόδου επηρεάζεται από το πλήθος των χαρακτηριστικών που χρησιμοποιούνται για την περιγραφή των patches. Συγκεκριμένα, χρησιμοποιώντας εκτός από τα τρία κανάλια χρώματος (R, G, B) και τις πρώτες x - και y - παραγώγους τους, η απόδοση αυξάνεται. Επίσης, παρατηρούμε ότι η μέθοδος παρουσιάζει υψηλή απόδοση ακόμα και στην περίπτωση όπου τα δάση αποτελούνται μόνο από 2 δυαδικά δέντρα απόφασης, σ' αυτή την περίπτωση η απόδοση είναι 82,7%. Η απόδοση βελτιώνεται με την αύξηση του πλήθους των δυαδικών δέντρων απόφασης σε κάθε δάσος και φθάνει για πλήθος 10 δυαδικών δέντρων στο 90,3%.

Συγκρίνοντας την προτεινόμενη μέθοδο με τη μέθοδο Bag of Keypoints παρατηρούμε ότι η πρώτη παρουσιάζει πολύ υψηλότερη απόδοση από τη δεύτερη για οποιοδήποτε πλήθος εικόνων εκπαίδευσης τόσο στο σύνολο εκπαίδευσης όσο και στο αντίστοιχο σύνολο ελέγχου. Επομένως τα μοντέλα που κατασκευάζει η προτεινόμενη μέθοδος



είναι πιο αποδοτικά καθώς παρουσιάζουν μικρότερο σφάλμα εκπαίδευσης και μικρότερο σφάλμα γενίκευσης. Από την άλλη πλευρά ο χρόνος ταξινόμησης μιας εικόνας είναι πολύ μεγαλύτερος για την προτεινόμενη μέθοδο από τη μέθοδο Bag of Keypoints.

Συνήθως, εικόνες που ανήκουν στην ίδια κατηγορία παρουσιάζουν μεγάλη ομοιότητα ως προς το χρωματικό τους περιεχόμενο. Εξετάζοντας τη δυνατότητα χρήσης της μεθόδου για ανάκτηση εικόνων με βάση το περιεχόμενο αποδείχθηκε πειραματικά ότι η μέθοδος δεν μπορεί να χρησιμοποιηθεί για την επίλυση του συγκεκριμένου προβλήματος σε περιπτώσεις όπου υπάρχουν μεγάλες διαφοροποιήσεις στο χρωματικό περιεχόμενο των εικόνων της ίδιας κατηγορίας.

Στην συγκεκριμένη εργασία για τη διανυσματική αναπαράσταση των εικόνων εκμεταλλευτήκαμε μόνο την πληροφορία του χρώματος, μιας και αυτό το χαρακτηριστικό χρησιμοποιείται από τις περισσότερες μεθόδους που έχουν προταθεί για την επίλυση του προβλήματος. Θα μπορούσε μελλοντικά να εξεταστεί η δυνατότητα χρήσης της μεθόδου για ανάκτηση εικόνων χρησιμοποιώντας εκτός από τα χαρακτηριστικά χρώματος και χαρακτηριστικά υφής ή και σχήματος. Επίσης θα μπορούσαν να εξεταστούν διαφορετικές τεχνικές κατασκευής των δυαδικών δέντρων απόφασής.

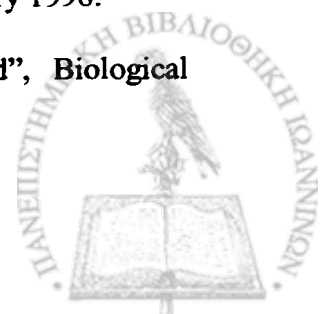


ΑΝΑΦΟΡΕΣ

- [1] R. Datta, D. Joshi, J. Li, J. Wang, "Image Retrieval: Ideas, Influences, and Trends of the New Age", *ACM Computing Surveys*, Vol. 40 (2), Article 5, April 2008.
- [2] R. S. Choras, "Image Feature Extraction Techniques and Their Applications for CBIR and Biometrics Systems", *International Journal of Biology and Biomedical Engineering*, Vol.1, pp. 6-16, 2007.
- [3] J. Gall, V. Lempitsky, "Class-Specific Hough Forests for Object Detection", In *Int' l Conf. Computer Vision and Pattern Recognition*, pp. 1022-1029, 2009.
- [4] V. Kumar, I. Patras, "A Discriminative Voting Scheme for Object Detection using Hough Forests", *BMVA Press*, September 2010.
- [5] J. Gall, A. Yao, N. Razavi, L. V. Gool, V. Lempitsky, "Hough Forests for Object Detection, Tracking, and Action Recognition", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 33, No 11, 2188-2201, 2011.
- [6] M-H Yang, D. Kriegman, N. Ahuja, "Detecting Faces in Images: A Survey", *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, Vol. 24, no. 1, pp. 34-58, 2002.
- [7] B. Tripti, "Image Retrieval from Web", *Journal of Computing*, Vol 2, Issue 12, December 2010.
- [8] I. Laptev, M. Marszalek, C. Schmid, B. Rozenfeld, "Learning realistic human actions from movies", *IEEE Conference on Computer Vision and Pattern Recognition*, 2008.
- [9] M. D. Rodriguez, J. Ahmed, M. Shah, "Action mach a spatio-temporal maximum average correlation height filter for action recognition", *IEEE Conference on Computer Vision and Pattern Recognition*, 2008.
- [10] J. Liu, J. Luo, M. Shad, "Recognition realistic actions from videos 'in the wild' ". *IEEE Conference on Computer Vision and Pattern Recognition*, 2009.



- [11] M. Andriluka, S. Roth, B. Shiele, "People-Tracking-by-Detection and People-Detection-by-Tracking", IEEE Conference on Computer Vision and Pattern Recognition, June 2008.
<http://www.mis.tu-darmstadt.de/node/382>
- [12] A. Yao, J. Gall, L. Van Gool, "A Hough Transform-Based Voting Framework for Action Recognition", IEEE Conference on Computer Vision and Pattern Recognition, 2010.
- [13] H. Pourghassem, H. Ghassemian, "Content-based medical image classification using a new hierarchical merging scheme". *Comp. Med. Image and Graph*, pp. 651-661, 2008.
- [14] P. Maheshwary, N. Srivastava, "Retrieval of Remote Sensing Images Using Color, Texture and Spectral Features", *International Journal of Engineering Science and Technology*, Vol.2(9), pp. 4306-4311, 2010.
- [15] I. Kononenko, M. Kukar, "Machine Learning and Data Mining: Introduction to Principles and Algorithms", Horwood, 2007.
- [16] P-N. Tan, M. Steinbach, V. Kumar, "Introduction to Data Mining", Addison-Wesley, May 2005.
- [17] D. Opitz, R. Maclin, "Popular ensemble methods: An empirical study", *Journal of Artificial Intelligence Research* Vol.11, pp. 169-198, 1999.
- [18] L. Breiman, "Bagging predictors", *Machine Learning*, Vol. 24, pp. 124-140, 1996.
- [19] E. Bauer, R. Kohavi, "An empirical comparison of voting classification algorithms: bagging, boosting and variants", *Machine Learning*, Vol. 36, pp. 1545-1588, 1999.
- [20] L. Breiman, "Random forests", *Machine Learning*, 45(1), pp. 5-32, 2001.
- [21] J. R. Quinlan, "Induction of decision trees", *Machine Learning*, 1(1), pp. 81-106, 1986.
- [22] J. B. MacQueen, "Some Methods for classification and Analysis of Multivariate Observations", *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, pp. 281-297, 1967.
- [23] J. A. Hartigan, M. A. Wong, "A K-means Clustering Algorithm", *Journal of Royal Statistical Society, Series C (Applied Statistics)*, 28(1), pp. 100-108, 1979.
- [24] A. Bjarnestam, "Text-based Hierarchical Image Classification and Retrieval of Stock Photography", *Challenge of Image Retrieval*, University of Northumbria, Newcastle on Tyne, BCS Electronics Workshops in Computing, February 1998.
- [25] B. Julesz, "Texton Gradients: The Texton Theory Revisited", *Biological Cybernetics*, Vol. 54 (4-5), pp. 245-251, 1986.



- [26] R. Haralick, "Statistical and Structural Approaches to Texture", Proceedings IEEE, Vol 67 (5), May 1979.
- [27] N. Dalal and B. Triggs. "Histogram of oriented gradients for human detection". In Proc. CVPR, 2005.
- [28] D. Lowe, "Distinctive image features from scale-invariant keypoints", IJCV, Corfu, Greece, pp. 1150-1157, 1999. (www.cs.ubc.ca/~lowe/keypoints/)
- [29] D. Lowe, "Object recognition from scale-invariant features", IJCV 60 (2), pp. 91-110, 2004.
- [30] G. Csurka, C. R. Dance, L. Fan, J. Willamowski, C. Bray, "Visual Categorization with Bags of Keypoints", In ECCV workshop on Statistical Learning in Computer Vision, pp.59-74, 2004.
- [31] E. Nowak, F. Jurie and B. Triggs. "Sampling Strategies for Bag-of-Features Image Classification". ECCV, 2006.
- [32] S. Tong, D. Koller, "Support vector machine active learning with applications to text classification", ICML, 2000.
- [33] D. H. Ballard, "Generalizing the Hough transform to detect arbitrary shapes", Pattern Recognition, 13(2), pp.111-122, 1981.
- [34] R. O. Duda, P. E. Hart, "Use of the Hough transform to detect lines and curves in pictures", Communications of Association for Computing Machinery, 15 (1), pp. 11-15, 1972.
- [35] P. V. C. Hough, "Method and means for recognizing complex patterns", U. S. Patent 3069654, 1962.
- [36] F. Attneave, "Some informational aspects of visual perception", Psychol. Rev 61, pp. 183-93, 1954.
- [37] C-W. Hsu, C-C. Chang, C-J. Lin, "A Practical Guide to Support Vector Classification".
- [38] C-C. Chang, C-J. Lin, "LIBSVM: a library for Support Vector Machines". ACM Transactions on Intelligent Systems and Technology, pp. 2:27:1-27:27, 2011. <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [39] The PASCAL Object Recognition Database Collection. <http://pascallin.ecs.soton.ac.uk/challenges/VOC/databases.html>
- [40] O. Chapelle, P. Haffner, V. Vapnik, "SVMs for Histogram-Based Image Classification", Neural Networks, IEEE Transactions on, Vol. 10, No. 5, pp. 1055-1064, 1999.



ΣΥΝΤΟΜΟ ΒΙΟΓΡΑΦΙΚΟ

Η Αγγελική Τσιάρα γεννήθηκε στα Ιωάννινα το 1983. Το 2001 εισήχθη στο Τμήμα Πληροφορικής του Πανεπιστημίου Ιωαννίνων και ολοκλήρωσε τις προπτυχιακές της σπουδές τον Ιούνιο του 2006. Από τον Οκτώβριο του 2009 έως τον Ιανουάριο του 2012 παρακολούθησε το Μεταπτυχιακό Πρόγραμμα Σπουδών του ίδιου Τμήματος.

