



Εθνικό Μετσόβιο Πολυτεχνείο
Δ.Π.Μ.Σ Γεωπληροφορική

Σχολή Αγρονόμων & Τοπογράφων Μηχανικών
Τομέας Τοπογραφίας
Εργαστήριο Φωτογραμμετρίας

**Αυτόματος εντοπισμός και αναγνώριση σημάτων
οδικής κυκλοφορίας**

Μεταπτυχιακή εργασία

Αδάμ Κατερίνα

Αθήνα, Σεπτέμβριος 2014

Ευχαριστίες

Με το τέλος της παρούσας μεταπτυχιακής εργασίας, που σηματοδοτεί και το τέλος της παραμονής μου στη ΣΑΤΜ και το Εργαστήριο Φωτογραμμετρίας, δε θα μπορούσα να μην ευχαριστήσω ορισμένα άτομα, που στιγμάτισαν τη μακρόχρονη αυτή πορεία.

Καταρχάς, θα ήθελα να ευχαριστήσω θερμά τον επιβλέποντα καθηγητή μου, κ. Χ. Ιωαννίδη για την συμπαράσταση του όλο αυτό το χρονικό διάστημα. Επίσης, θα ήθελα να ευχαριστήσω τον καθηγητή του Εργαστηρίου Φωτογραμμετρίας, κ. Γ. Καρρά και την καθηγήτρια του ΤΕΙ Αθήνας, κ. Ε. Πέτσα, καθώς ήταν οι πρώτοι που μου έδειξαν τον δρόμο προς το πεδίο της Όρασης Υπολογιστών. Και οι τρεις τους, με ανέχτηκαν όλα αυτά τα χρόνια με προθυμία και υπομονή, μοιράστηκαν μαζί μου επιστημονικές αναζητήσεις και σε με συμβούλευαν –τόσο σε προσωπικό όσο και σε επαγγελματικό επίπεδο- πάντα με γνώμονα το προσωπικό μου συμφέρον. Ένα ευχαριστώ είναι λίγο για την καθοδήγησή τους, τον χρόνο και τις γνώσεις τους, που μοιράστηκαν απλόχερα μαζί μου και για όλα αυτά που μου έχουν προσφέρει μέχρι σήμερα. Η επαγγελματική μου πορεία και αντιμετώπιση, δίχως τη δικιά συμβολή τους, θα ήταν εντελώς διαφορετικές.

Θα ήθελα ακόμα να ευχαριστήσω όλα τα μέλη του Εργαστηρίου Φωτογραμμετρίας για τη συνεργασία μας και που με έκαναν να νιώσω το Εργαστήριο για μεγάλο χρονικό διάστημα σα δεύτερο σπίτι μου.

Ευχαριστώ επίσης ιδιαίτερα τους συμφοιτητές μου για όλα αυτά που μοιραστήκαμε, άγχη, αγωνίες αλλά και πολύ όμορφες στιγμές.

Τέλος, οφείλω ένα τεράστιο ευχαριστώ σε όλα τα μέλη του Εργαστηρίου Γνώσης, Πολυμέσων και Κοινωνικών Δικτύων του Εθνικού Κέντρου Έρευνας και Τεχνολογικής Ανάπτυξης για τη βοήθεια, έμπρακτη συμπαράσταση και υπομονή που επέδειξαν κατά τη διάρκεια εκπόνησης της παρούσας εργασίας.

Περίληψη

Η Όραση Υπολογιστών και η Μηχανική Μάθηση δύνανται να χρησιμοποιηθούν σε ένα ευρύ φάσμα ιδιαίτερα ενδιαφερουσών εφαρμογών. Μία χαρακτηριστική εφαρμογή συνιστά η δημιουργία Συστημάτων Υποβοήθησης Οδήγησης (Driving Assistance Systems), τα οποία ουσιαστικά στοχεύουν, με την εξέλιξή τους, στην πλήρως αυτοματοποιημένη οδήγηση. Για τη δημιουργία ενός τέτοιου συστήματος απαιτούνται διάφορες επιμέρους εφαρμογές, όπως συστήματα αυτόματου εντοπισμού εμποδίων, εντοπισμού λωρίδων οδικής κυκλοφορίας και εντοπισμού και κατηγοριοποίησης σημάτων οδικής κυκλοφορίας. Η παρούσα εργασία στοχεύει στη δημιουργία της τελευταίας εφαρμογής. Πιο συγκεκριμένα, εκμεταλλεζόμενη τεχνικές επεξεργασίας εικόνας και μηχανικής μάθησης φιλοδοξεί να δημιουργήσει έναν αλγόριθμο που είναι ικανός να αναγνωρίζει αυτόματα ποιές πινακίδες οδικής κυκλοφορίας συναντά ένα κινούμενο όχημα. Προς το παρόν, μία τέτοια εφαρμογή μπορεί να χρησιμοποιηθεί βοηθητικά (για παράδειγμα στέλνοντας προειδοποιητικούς ήχους όταν ο οδηγός παραβιάζει μία από τις υπάρχουσες πινακίδες), αλλά ακόμα μπορεί να αποτελέσει και μέρος ενός συστήματος αυτόματης καταγραφής της ακριβούς θέσης των πινακίδων. Πληθώρα προβλημάτων ανακύπτουν κατά τη διαδικασία, τα οποία ο αλγόριθμος οφείλει να υπερβεί προκειμένου να λειτουργεί ικανοποιητικά. Ανάμεσά τους συγκαταλέγονται οι έντονες εναλλαγές στις συνθήκες φωτεινότητας, οι οποίες οφείλονται όχι μόνο στις εναλλαγές ημέρας-νύχτας αλλά και στις διαφορετικές καιρικές συνθήκες. Ακόμα, πιθανά προβλήματα στην αυτόματη διαδικασία εντοπισμού και αναγνώρισης προκαλούν αποκρύψεις των πινακίδων, αλλοιώσεις στο χρώμα τους, λόγω της έκθεσης στον ήλιο, βανδαλισμοί που υφίστανται συχνά τα σήματα, καθώς και η ύπαρξη παρόμοιων αντικειμένων. Επίσης, καθώς συχνά οι πινακίδες στρέφονται και μετακινούνται, ο αλγόριθμος που θα αναπτυχθεί οφείλει να παρουσιάζει ανοχές σε ανάλογους μετασχηματισμούς.

Η δομή της παρούσας εργασίας είναι η ακόλουθη. Αρχικά πραγματοποιείται μία εκτενής αναφορά στις διάφορες μεθόδους, που έχουν χρησιμοποιηθεί κατά καιρούς για τον συγκεκριμένο σκοπό, καθώς ο αυτόματος εντοπισμός και κατηγοριοποίηση πινακίδων οδικής κυκλοφορίας συνιστά ένα ιδιαίτερα ευρύ και δραστήριο πεδίο έρευνας. Κατόπιν, αναλύεται και επεξηγείται ο αλγόριθμος που συντάχθηκε στο πλαίσιο της παρούσας εργασίας με έμφαση στο θεωρητικό υπόβαθρο των τεχνικών όρασης υπολογιστών και μηχανικής μάθησης που επιστρατεύτηκαν. Μοναδικό δεδομένο εισόδου για τον αλγόριθμο συνιστούν έγχρωμες εικόνες εισημμένες από φωτομηχανή χαμηλού κόστους, η οποία έχει τοποθετηθεί σε κινούμενο όχημα. Στις εικόνες πραγματοποιείται εντοπισμός βάσει χρώματος, προκειμένου να εντοπιστούν περιοχές ενδιαφέροντος, που πιθανώς περιέχουν πινακίδες οδικής κυκλοφορίας. Ακολουθεί ο κυκλικός μετασχηματισμός Hough, που εκμεταλλεύεται τις γεωμετρικές ιδιότητες των σημάτων και οι τελικές περιοχές ενδιαφέροντος αναπαρίστανται με τη χρήση των περιγραφέων HOG και εισέρχονται σε ταξινομητές (Μηχανές Δια-

νυματικής Στήριξης), που έχουν προηγουμένως εκπαιδευθεί. Για την εκπαίδευση του συστήματος δημιουργήθηκε μία βάση δεδομένων με εικόνες που περιέχουν τα ελληνικά σήματα οδικής κυκλοφορίας. Οι εικόνες ελήφθησαν κάτω από διαφορετικές καιρικές συνθήκες και συνθήκες φωτισμού και στη συνέχεια διεξήχθη σειρά πειραμάτων, ώστε να διαπιστωθεί η επιτυχία του αλγορίθμου που αναπτύχθηκε. Τα αποτελέσματα συνοψίζονται και σχολιάζονται στο τέλος της εργασίας, όπου ακόμα παρατίθενται και κάποια τελικά συμπεράσματα και προτάσεις για μελλοντική έρευνα, όσο και η χρησιμοποιηθείσα βιβλιογραφία.

Λέξεις Κλειδιά: Εντοπισμός, Κατηγοριοποίηση, Χρώμα, Όραση Υπολογιστών, Μηχανική Μάθηση

School of Rural and Surveying Engineering

National Technical University of Athens, Greece

Automatic road sign detection and recognition

Katerina Adam

Master Thesis

September 2014

Abstract

Computer Vision and Machine Learning techniques contribute to a wide range of interesting applications. A well known application among them is road sign detection and classification, which is a very active field of research. Although the first works in this area date back to the 1960s, significant progress has been made during the last years. Automatic road sign detection and classification has many practical implementations. The most widespread is the usage of the method in order to create Driver Assistance Systems (DAS). Such systems aspire to the fully automatic navigation of the car through auto pilot and include systems for automatic lane detection, obstacle detection in the vehicle path and road sign recognition. For the time being, road sign detection and classification shall be used in order to assist the driver and enhance safety, e.g., by sending warning signals revealing over speed or indicating the presence of a specific sign that the driver may not notice due to distraction and lack of attention. Another interesting application is the mapping of the traffic signs, in order to be used for automated road maintenance. For this purpose, the system developed shall be used in combination with a mobile mapping system, offering information about the exact location of each detected sign. Traffic sign recognition systems have to face many challenges. First of all, illumination conditions are not controllable. Depending on the time of the day and the weather conditions, illumination may vary dramatically. Secondly, traffic signs may be partially damaged, vandalized or with faded color due to long exposure to sunlight; problems that hamper the successful detection and recognition. Other common problems are occlusions and shadows occurring by other objects surrounding the traffic signs as well as the existence of similar objects, which may be detected as road signs. Finally, possible rotations and translation of the traffic signs occur, thus the system developed should be invariant under rotation and translation.

The thesis is organized as follows. Firstly, the state of the art methods for traffic sign detection and classification are described. Afterwards, the proposed the system, including the theoretical background of the techniques applied is presented and

analysed. The proposed system uses color-images acquired by a low cost camera mounted on a moving vehicle. Color based detection is used in order to locate regions of interest. Then, a circular Hough transform is applied to complete detection, taking advantage of the shape properties of the road signs. The regions of interest are finally represented using HOG descriptors and are fed into trained Support Vector Machines (SVMs) in order to be recognized. For the training procedure, a database with several training examples depicting Greek road signs has been developed. Many experiments have been conducted and are presented, to measure the efficiency of the proposed methodology especially under adverse weather conditions and poor illumination. The final results are presented and discussed and the conclusions of the thesis are presented, along with the cited references.

KEY WORDS: Detection, Classification, Colour, Automation, Vision, Learning

Περιεχόμενα

| | |
|----------------------------------------------------------------------------------------------|----|
| Ευχαριστίες | i |
| Περίληψη | ii |
| Abstract | iv |
| Περιεχόμενα | vi |
| 1. Εισαγωγή | 1 |
| 1.1 Στόχος της εργασίας | 1 |
| 1.2 Προβλήματα στη διαδικασία | 1 |
| 1.3 Τα σήματα οδικής κυκλοφορίας | 2 |
| 1.3.1 Τα σήματα οδικής κυκλοφορίας στον Ελληνικό χώρο | 3 |
| 1.4 Διάρθρωση της εργασίας | 4 |
| 2. Μέθοδοι και τεχνικές εντοπισμού και αναγνώρισης σημάτων | 5 |
| 2.1 Εντοπισμός | 5 |
| 2.1.1 Εντοπισμός με βάση το χρώμα | 6 |
| 2.1.1.1 Εντοπισμός στον χώρο RGB | 7 |
| 2.1.1.2 Εντοπισμός στους χώρους HSI και HSV | 11 |
| 2.1.1.3 Μετασχηματισμός σε άλλους χρωματικούς χώρους | 18 |
| 2.1.2 Εντοπισμός με βάση το σχήμα | 20 |
| 2.1.2.1 Απόσταση από Όρια (Distance to Borders) | 21 |
| 2.1.2.2 Μετασχηματισμός Hough | 23 |
| 2.1.2.3 Γρήγορος Ακτινικός Μετασχηματισμός Συμμετρίας (Fast Radial symmetry transform) | 28 |
| 2.1.2.4 Συνταύτιση προτύπων | 31 |
| 2.1.2.5 Αλγόριθμος Ransac | 33 |
| 2.2 Αναγνώριση | 35 |
| 2.2.1 Συνταύτιση προτύπων (Template matching) | 36 |
| 2.2.2 Νευρωνικά Δίκτυα (Neural Networks) | 36 |
| 2.2.3 Μηχανές Διανυσματικής Στήριξης (Support Vector Machines) | 42 |
| 3. Μηχανές Διανυσματικής Στήριξης (SVM) | 49 |
| 3.1 Γενίκευση των SVM | 49 |
| 3.1.1 Πραγματικός, εμπειρικός και διαρθρωτικός κίνδυνος | 49 |
| 3.1.2 Διάσταση VC | 51 |
| 3.1.3 Ελαχιστοποίηση δομικού κινδύνου | 52 |
| 3.2 Γραμμικές Μηχανές Διανυσμάτων Στήριξης | 53 |
| 3.3 Ελαστικό περιθώριο (soft margin) | 56 |

| | | |
|---------|-------------------------------------------------|-----|
| 3.4 | Μη γραμμικές μηχανές διανυσμάτων στήριξης | 57 |
| 3.5 | Συναρτήσεις πυρήνα..... | 59 |
| 3.5.1 | Θεώρημα του Mercer | 59 |
| 3.5.2 | Συνήθειες συναρτήσεις πυρήνα | 60 |
| 3.5.3 | Επιλογή κατάλληλης συνάρτησης πυρήνα..... | 61 |
| 3.6 | ν -SVM | 62 |
| 3.7 | Ταξινόμηση σε περισσότερες κατηγορίες | 63 |
| 4. | Προτεινόμενος αλγόριθμος..... | 65 |
| 4.1 | Εντοπισμός στον χρωματικό χώρο HSI..... | 67 |
| 4.1.1 | Ο χρωματικός χώρος HSI..... | 67 |
| 4.1.2 | Κατωφλίωση στον χρωματικό χώρο HSI..... | 69 |
| 4.2 | Φίλτρο μεσαίας τιμής | 70 |
| 4.3 | Διαστολή δυαδικής εικόνας..... | 71 |
| 4.4 | Επιλογή των τελικών περιοχών ενδιαφέροντος..... | 74 |
| 4.5 | Εντοπισμός κύκλων..... | 77 |
| 4.5.1 | Τελεστής ανίχνευσης ακμών (Canny) | 78 |
| 4.5.2 | Ο κυκλικός μετασχηματισμός Hough | 79 |
| 4.6 | Περιγραφείς HOG | 82 |
| 4.7 | Αναγνώριση..... | 85 |
| 5. | Αποτελέσματα..... | 86 |
| 5.1 | Αποτελέσματα εντοπισμού..... | 87 |
| 5.1.1 | Αποτελέσματα εντοπισμού με βάση το χρώμα | 87 |
| 5.1.2 | Αποτελέσματα εντοπισμού με βάση το σχήμα..... | 95 |
| 5.2 | Αναγνώριση..... | 97 |
| 5.2.1 | Εκπαίδευση..... | 97 |
| 5.2.1.1 | Εξαγωγή περιγραφικών χαρακτηριστικών | 98 |
| 5.2.1.2 | Τα δεδομένα εκπαίδευσης | 98 |
| 5.2.1.3 | Το λογισμικό SVMlight | 99 |
| 5.2.1.4 | Προετοιμασία δεδομένων εκπαίδευσης..... | 99 |
| 5.2.1.5 | Αποτελέσματα εκπαίδευσης..... | 100 |
| 5.2.2 | Αποτελέσματα αναγνώρισης | 104 |
| 6. | Συμπεράσματα | 108 |
| 6.1 | Σχόλια και παρατηρήσεις | 108 |
| 6.2 | Προτάσεις για το μέλλον | 109 |
| | Βιβλιογραφία | 111 |

| | |
|------------------|-----|
| Παράρτημα Α..... | 121 |
| Παράρτημα Β..... | 163 |

1

Εισαγωγή

1.1 Στόχος της εργασίας

Με την πάροδο των χρόνων και την πρόοδο της τεχνολογίας και της πληροφορικής, η ανάπτυξη συστημάτων υποβοήθησης οδήγησης που στοχεύουν στην αύξηση της οδικής ασφάλειας είναι πλέον μία πραγματικότητα. Στην ανάπτυξη ανάλογων συστημάτων συμπεριλαμβάνεται ο αυτόματος εντοπισμός λωρίδων κυκλοφορίας, ο εντοπισμός πεζών, η καταγραφή της κίνησης, η δημιουργία συστημάτων που διευκολύνουν τη στάθμευση και άλλα. Ένα ακόμα σημαντικό κομμάτι για ανάλογα συστήματα, με το οποίο ασχολείται η συγκεκριμένη διπλωματική εργασία, είναι ο αυτόματος εντοπισμός και κατηγοριοποίηση των σημάτων οδικής κυκλοφορίας, με μοναδικό δεδομένο βίντεο ή εικόνες που προέρχονται από κάμερα τοποθετημένη στο αυτοκίνητο. Ένα εγχείρημα αρκετά περίπλοκο, ειδικά εάν αναλογιστεί κανείς τα διάφορα προβλήματα που ανακύπτουν. Ταυτόχρονα όμως, η ανάπτυξη ενός τέτοιου συστήματος αποτελεί ένα ιδιαίτερα καινοτόμο και χρήσιμο προϊόν καθώς όχι μόνο θα εντοπίζει τα σήματα οδικής κυκλοφορίας, αλλά θα είναι σε θέση παράλληλα να κατηγοριοποιεί και τα αναγνωρίζει ώστε να προειδοποιεί τον οδηγό για την επιθυμητή αντίδρασή του. Το πρώτο στάδιο του εντοπισμού βασίζεται σε τεχνικές επεξεργασίας εικόνας με τελικό σκοπό την κατάτμησή της. Η διαδικασία αυτή εκμεταλλεύεται το χρώμα και το σχήμα των πινακίδων. Ακολουθεί το στάδιο της κατηγοριοποίησης των σημάτων με μεθόδους μηχανικής μάθησης. Μέχρι στιγμής έχουν πραγματοποιηθεί διάφορες ανάλογες προσπάθειες και έχουν συνταχθεί αλγόριθμοι που αναλύονται στην παρούσα εργασία.

1.2 Προβλήματα στη διαδικασία

Η διαδικασία αυτόματου εντοπισμού και κατηγοριοποίησης των σημάτων οδικής κυκλοφορίας είναι ιδιαίτερα δύσκολη λόγω διαφόρων προβλημάτων. Το σημαντικότερο ίσως από αυτά σχετίζεται με τις δυσμενείς εξωτερικές συνθήκες, όπως ή ύπαρξη ομίχλης, βροχής, χιονιού και οι διαφορετικοί φωτισμοί των λήψεων, οι οποίες έχουν άμεσο αντίκτυπο στις πινακίδες οδικής κυκλοφορίας. Για παράδειγμα, συχνό φαινόμενο σε μέρες με ηλιοφάνεια ή και κατά τη διάρκεια της νύχτας αποτελούν οι ανακλάσεις στις εικόνες. Η δημιουργία ενός αλγορίθμου ικανού να ανταποκρίνεται σε όλους τους διαφορετικούς φωτισμούς (ηλιοφάνεια, συννεφιά, νύκτα) αποτελεί μία πρόκληση. Ακόμα, συχνά εμφανίζονται θολές και κουνημένες εικόνες, λόγω της κίνησης και πιθανών αναταράξεων του αυτοκινήτου. Άλλη μία δυσκολία που πρέπει να αντιμετωπιστεί είναι οι πιθανές αποκρύψεις των πινακίδων από άλλα αντικείμενα,

καθώς και η ύπαρξη αντικειμένων στη λήψη που μοιάζουν με σήματα οδικής κυκλοφορίας. Επίσης, προβλήματα δημιουργούνται στον εντοπισμό πινακίδων που έχουν φθαρεί από την πάροδο του χρόνου και στον εντοπισμό και την κατηγοριοποίηση σημάτων κατεστραμμένων ή με αυτοκόλλητα. Στην Εικόνα 1.1 εμφανίζονται διάφορα παραδείγματα ανάλογων περιπτώσεων. Πιο συγκεκριμένα, στην πρώτη περίπτωση εμφανίζονται δύο πινακίδες οδικής κυκλοφορίας που έχουν φθαρεί από τοποθέτηση αυτοκόλλητων και έχουν μετατοπιστεί από τη συνήθη θέση. Η δεύτερη περίπτωση αποτελεί χαρακτηριστικό παράδειγμα ύπαρξης έντονων αποκρύψεων και στην τρίτη εικόνα ο διαχωρισμός της πινακίδας από το υπόβαθρο της εικόνας καθίσταται δύσκολος, λόγω παρόμοιων αποχρώσεων της πινακίδας με τον περιβάλλοντα χώρο.



Εικόνα 1.1: Παραδείγματα δυσμενών περιπτώσεων για εντοπισμό και κατηγοριοποίηση πινακίδων

1.3 Τα σήματα οδικής κυκλοφορίας

Πρωταρχικό στάδιο για την εκπόνηση ενός τέτοιου εγχειρήματος αποτελεί η μελέτη των σημάτων που ανήκουν στον κώδικα οδικής κυκλοφορίας. Σε όλη την Ευρώπη, πραγματοποιείται προσπάθεια για την ομοιογένεια των πινακίδων αυτών. Για τον σκοπό αυτό έχει συνταχθεί η συνθήκη της Βιέννης για τα σήματα οδικής κυκλοφορίας (1968) στην οποία και υπακούουν όλα τα μέλη της Ευρώπης (εκτός από την Μολδαβία, την Ιρλανδία και το Ηνωμένο Βασίλειο). Παρά τη συγκεκριμένη συνθήκη, πολλές παραμένουν οι διαφορές που εντοπίζονται μεταξύ διαφορετικών κρατών και εντοπίζονται κυρίως στη χρησιμοποιούμενη γραμματοσειρά, στα βέλη, στα εικονογράμματα και στη χρήση διαφορετικών χρωμάτων. Χαρακτηριστικό παράδειγμα αποτελεί η Εικόνα 1.2.



Εικόνα 1.2: Η διαφοροποίηση των σημάτων οδικής κυκλοφορίας σε διάφορες χώρες της Ευρώπης

Η παρούσα εργασία επικεντρώθηκε στα σήματα του ελληνικού κώδικα οδικής κυκλοφορίας τα οποία παρουσιάζονται συνοπτικά στο ακόλουθο υποκεφάλαιο.

1.3.1 Τα σήματα οδικής κυκλοφορίας στον Ελληνικό χώρο

Τα σήματα οδικής κυκλοφορίας, σύμφωνα με τον Ελληνικό κώδικα οδικής κυκλοφορίας [1], κατηγοριοποιούνται σε τέσσερις βασικές ομάδες: στις πινακίδες αναγγελίας κινδύνου (Κ), τις ρυθμιστικές πινακίδες της κυκλοφορίας (Ρ), τις πληροφοριακές πινακίδες (Π) και τις πρόσθετες πινακίδες (Πρ). Παραδείγματα για κάθε μια από τις ακόλουθες κατηγορίες εμφανίζονται ακολούθως.



Εικόνα 1.3α: Πινακίδα αναγγελίας



Εικόνα 1.3β Ρυθμιστική πινακίδα



Εικόνα 1.3γ: Πληροφοριακή Πινακίδα



Εικόνα 1.3δ: Πρόσθετη Πινακίδα

Κατόπιν μελέτης του Κ.Ο.Κ, γίνεται κατανοητό ότι τα σήματα που ανήκουν στις δύο πρώτες κατηγορίες, αποτελούν τα σημαντικότερα για την εξασφάλιση ασφαλούς οδήγησης. Συνεπώς, τα συστήματα αυτόματου εντοπισμού και κατηγοριοποίησης

πινακίδων επικεντρώνονται στις δύο αυτές κατηγορίες. Τα βασικά χρώματα των πινακίδων αυτών είναι το μπλε, κόκκινο, κίτρινο και λευκό, ενώ τα σχήματά τους είναι κατά βάση τριγωνικά, κυκλικά και τετράπλευρα.

1.4 Διάρθρωση της εργασίας

Η παρούσα εργασία διαρθρώνεται ακολούθως ως εξής. Στο **κεφάλαιο 2** πραγματοποιείται μία βιβλιογραφική ανασκόπηση των αλγορίθμων που έχουν εφαρμοστεί και χρησιμοποιηθεί μέχρι στιγμής. Στο **κεφάλαιο 3** πραγματοποιείται μία εκτενής αναφορά στις Μηχανές Διανυσματικής Στήριξης (SVM) που χρησιμοποιήθηκαν ως μέθοδος μηχανικής μάθησης για την εφαρμογή που αναπτύχθηκε στα πλαίσια της παρούσας εργασίας. Στο **κεφάλαιο 4** αναλύεται η εφαρμογή και περιγράφεται λεπτομερώς το κάθε στάδιό της. Ακολουθεί το **κεφάλαιο 5**, όπου συνοψίζονται και αναλύονται τα αποτελέσματα της μεθόδου και τέλος το **κεφάλαιο 6** με τα συμπεράσματα και ορισμένες σκέψεις για το μέλλον. Η μεταπτυχιακή εργασία ολοκληρώνεται με την παράθεση των βιβλιογραφικών αναφορών.

2

Μέθοδοι και τεχνικές εντοπισμού και αναγνώρισης σημάτων

Μέχρι στιγμής έχουν δημοσιευτεί αρκετές εργασίες και έχουν αναπτυχθεί πλήθος αλγόριθμοι που ασχολούνται με τον αυτόματο εντοπισμό και κατηγοριοποίηση πινακίδων οδικής κυκλοφορίας. Η πρώτη αναφορά σημειώνεται στην Ιαπωνία το 1984 από τον Paclik [2]. Από τότε πλήθος ερευνητών έχουν προσεγγίσει το θέμα, με δημιουργία λογισμικών που λειτουργούν ακόμα και σε πραγματικό χρόνο, ενσωματώνοντας αλγορίθμους από τα πεδία της όρασης υπολογιστών, της επεξεργασίας εικόνας και της μηχανικής μάθησης.

Τα απαραίτητα βήματα για ένα τέτοιο σύστημα είναι ο εντοπισμός της περιοχής ενδιαφέροντος και η αναγνώριση του σήματος. Το πρώτο βήμα βασίζεται στις χρωματικές και γεωμετρικές ιδιότητες των πινακίδων. Η συνηθέστερη τακτική που εκμεταλλεύεται το χρώμα των σημάτων οδικής κυκλοφορίας είναι η κατωφλίωση σε διάφορους χρωματικούς χώρους. Χρησιμοποιούνται ακόμα τεχνικές όπως η επέκταση περιοχής ενδιαφέροντος (region growing) και η εφαρμογή κανόνων ασαφούς λογικής (fuzzy rules). Από την άλλη μεριά, για την εκμετάλλευση των γεωμετρικών ιδιοτήτων έχουν εφαρμοστεί διάφοροι αλγόριθμοι, όπως ο μετασχηματισμός του Hough, η τεχνική Distance to Borders κ.α. Στο στάδιο της αναγνώρισης χρησιμοποιούνται νευρωνικά δίκτυα, Support Vector Machines και διάφορες ακόμα μέθοδοι που αναλύονται εκτενώς στη συνέχεια της εργασίας.

Σε αυτό κεφάλαιο κατηγοριοποιούνται και αναλύονται εν συντομία οι πιο διαδεδομένες μέθοδοι για τα στάδια του εντοπισμού και της ταξινόμησης.

2.1 Εντοπισμός

Ζητούμενο του πρώτου σταδίου της διαδικασίας είναι η εύρεση του τμήματος της εικόνας που περιέχει το προς ανίχνευση αντικείμενο. Το συγκεκριμένο τμήμα ονομάζεται «περιοχή ενδιαφέροντος» (Region of Interest – ROI). Κατά τον εντοπισμό προκύπτουν μία ή και περισσότερες περιοχές ενδιαφέροντος, που αποκόπτονται και τροφοδοτούν το στάδιο της αναγνώρισης. Ο εντοπισμός είναι ιδιαίτερα σημαντικός,

καθώς μάλιστα πληροφορία που απομακρύνεται κατά την εφαρμογή του, δεν μπορεί να ανακτηθεί αργότερα.

Ακολούθως αναλύονται οι βασικότερες μέθοδοι εντοπισμού με βάση το χρώμα ή το σχήμα και αναφέρονται ορισμένες δημοσιεύσεις που χρησιμοποιούν αυτές τις μεθόδους. Ο εντοπισμός με βάση το χρώμα είναι συνήθως λιγότερο απαιτητικός υπολογιστικά σε σύγκριση με τον εντοπισμό με βάση το σχήμα. Συχνά όμως, για πιο αξιόπιστα αποτελέσματα, χρησιμοποιείται συνδυασμός των δύο προσεγγίσεων.

2.1.1 Εντοπισμός με βάση το χρώμα

Ο εντοπισμός με βάση το χρώμα βασίζεται στη φωτεινότητα και το χρώμα των εικονοψηφίδων. Στόχος είναι να εντοπιστεί η περιοχή ενδιαφέροντος μέσω της ομαδοποίησης ομοιογενών περιοχών, μία διαδικασία που ονομάζεται κατάτμηση εικόνας (image segmentation). Η κατάτμηση επιτυγχάνεται μέσω επτά βασικών μεθόδων, σύμφωνα με την κατηγοριοποίηση που επιχειρείται στο [3]. Πιο συγκεκριμένα, οι συγγραφείς αναφέρονται στη μέθοδο των ομοιογενών περιοχών (region based technique), τη μέθοδο των συστάδων (clustering technique), τη μέθοδο διαίρεσης-ένωσης (split – merge technique), τη μέθοδο με γραφήματα (normalized cuts technique), την μέθοδο επαύξησης περιοχής ενδιαφέροντος (region growing), την κατωφλίωση και τη μέθοδο εντοπισμού ακμών. Ανάλογη προσπάθεια κατηγοριοποίησης πραγματοποιείται στο [4], όπου οι συγγραφείς προσθέτουν την κατάτμηση που βασίζεται στις μερικές διαφορικές εξισώσεις και την κατάτμηση με χρήση νευρωνικών δικτύων. Καθώς τα τελευταία χρόνια, είναι ραγδαία η ανάπτυξη αλγορίθμων προερχόμενων από το πεδίο της μηχανικής μάθησης (machine learning) και η εφαρμογή τους στην κατάτμηση εικόνας, χρήσιμο θα ήταν η τελευταία μέθοδος να γενικευθεί σε κατάτμηση εικόνας με χρήση τεχνικών μηχανικής μάθησης. Στις δημοσιεύσεις που ακολουθούν αναλύεται πώς οι παραπάνω μέθοδοι χρησιμοποιούνται για τον αυτόματο εντοπισμό σημάτων οδικής κυκλοφορίας.

Από όλες τις τεχνικές που αναφέρθηκαν, η πιο διαδεδομένη για το εν λόγω θέμα είναι η κατωφλίωση, λόγω της χαμηλής υπολογιστικής πολυπλοκότητας. Πρόκειται για μία διαδικασία ικανή να διαχωρίσει τις εικονοψηφίδες που ανήκουν στο αντικείμενο (foreground object) κι εκείνες που ανήκουν στο υπόβαθρο (background object), καθώς είναι μία διαδικασία επιλογής εικονοψηφίδων που εντοπίζονται ανάμεσα σε όρια (κατώφλια). Στο σύνολο των εικονοψηφίδων εντός των ορίων αποδίδεται η τιμή 1 και αυτά συνιστούν το αντικείμενο, ενώ στις υπόλοιπες εικονοψηφίδες, που αποτελούν το υπόβαθρο, αποδίδεται η τιμή 0. Έτσι, αποτέλεσμα της διαδικασίας κατωφλίωσης είναι μία δυαδική εικόνα. Τα κατώφλια είναι είτε προκαθορισμένα, σταθερά και καθολικά είτε προσαρμοστικά (adaptive) και καθορίζονται με βάση τις χρωματικές ιδιότητες του προς εντοπισμού αντικειμένου. Το σταθερό, καθολικό κατώφλι ορίζεται ενιαία για όλο το εύρος της εικόνας, όπως υπαγορεύει η ακόλουθη σχέση:

$$g(x, y) = \begin{cases} 0 & f(x, y) < T \\ 1 & f(x, y) > T \end{cases} \quad (2.1)$$

Μια παραλλαγή της μεθόδου συνιστά η χρήση διπλού κατωφλίου για τον ορισμό ενός φάσματος τιμών [5]. Σε αντίθεση με τα σταθερά κατώφλια, τα δυναμικά προσαρμόζονται για κάθε υποπεριοχή της εικόνας, ανάλογα με τη φωτεινότητά της. Ο καθορισμός των εν λόγω ορίων μπορεί να πραγματοποιηθεί είτε αυτόματα είτε από τον χρήστη. Στο [6] συνοψίζονται οι έξι βασικές κατηγορίες αλγορίθμων αυτόματου εντοπισμού κατωφλίων, οι οποίες είναι: αλγόριθμοι βάσει του σχήματος του ιστογράμματος (histogram shape based methods), αλγόριθμοι συστάδας (clusteringbased methods), αλγόριθμοι εντροπίας (entropy based methods), αλγόριθμοι που βασίζονται στις ιδιότητες του αντικειμένου (object attribute based methods), χωρικοί αλγόριθμοι (spatial based methods) και τοπικοί αλγόριθμοι (local methods). Στην πλειονότητα των ερευνητικών εργασιών, που αφορούν στην κατωφλίωση για εντοπισμό πινακίδων, τα όρια εντοπίζονται χειροκίνητα με μελέτη πολλών δειγμάτων, καθώς οι αυτόματοι αλγόριθμοι δεν είναι πλήρως αποτελεσματικοί. Σε προσπάθειες αυτοματοποίησης της επιλογής των κατωφλίων, που έχουν γίνει κατά καιρούς, χρησιμοποιείται κυρίως ο αλγόριθμος του Otsu, όπως αναλύεται και στα επόμενα.

Η ποιότητα των αποτελεσμάτων της κατωφλίωσης εξαρτάται από πλήθος παραμέτρων και κυρίως από τον θόρυβο στην εικόνα, τη φωτεινότητά της, την αντίθεση, το μέγεθος του προς εντοπισμού αντικείμενου αλλά και από την ορθή επιλογή των κατωφλίων, όπως αναφέρεται στο [6].

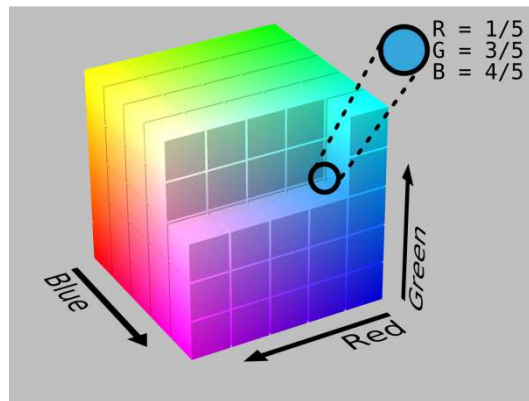
Σημαντικό στοιχείο στον εντοπισμό πινακίδων είναι ότι το λευκό και το μαύρο είναι ουδέτερα χρώματα (achromatic colors), αφού όλες οι αποχρώσεις (μήκη κύματος) συμμετέχουν ισομερώς για τον σχηματισμό τους (στο μαύρο δε συμμετέχει καμία απόχρωση και στο λευκό όλες), σε αντίθεση με τα κόκκινο, κίτρινο και μπλε (chromatic colors). Τα τελευταία εμφανίζουν μία κύρια απόχρωση και αυτό καθιστά ευκολότερο τον εντοπισμό τους. Στην περίπτωση των πινακίδων οδικής κυκλοφορίας, καθώς τα περιγράμματα αποτελούνται κατά το μεγαλύτερο ποσοστό από χρώματα (chromatic colors), η ερευνητική δραστηριότητα επικεντρώνεται κυρίως στον εντοπισμό τους, αν και έχουν προταθεί κατά καιρούς διάφορες μέθοδοι και για τον εντοπισμό του λευκού και του μαύρου.

Ακολούθως αναλύονται οι βασικότερες διαδικασίες κατάτμησης εικόνας με βάση το χρώμα, σε διάφορους χρωματικούς χώρους. Όπως επισημαίνεται στο [7] ο κάθε χρωματικός χώρος προκύπτει από ένα χρωματικό μοντέλο, δηλαδή ένα αφηρημένο μαθηματικό μοντέλο που αναπαριστά κάθε χρώμα ως συνδυασμό αριθμών. Οι πιο συνηθισμένες διαστάσεις των χρωματικών μοντέλων είναι τρεις ή τέσσερις. Για πιο αξιόπιστα αποτελέσματα, σε ορισμένες περιπτώσεις όπως στο [8], πραγματοποιείται εντοπισμός σε περισσότερους του ενός χρωματικού χώρου και τα αποτελέσματα γίνονται αποδεκτά μόνο αν ο εντοπισμός ισχύει και για όλες τις περιπτώσεις.

2.1.1.1 Εντοπισμός στον χώρο RGB

Ο χρωματικός χώρος RGB είναι ο βασικός χρωματικός χώρος. Ουσιαστικά ορίζεται από έναν κύβο στο καρτεσιανό σύστημα συντεταγμένων, όπου οι άξονες x , y και z

αναπαριστούν τα βασικά χρώματα R, G, B αντίστοιχα. Τα λοιπά χρώματα προκύπτουν από συνδυασμό των τριών βασικών χρωμάτων.



Εικόνα 2.1: Ο χρωματικός χώρος RGB

(Πηγή: http://en.wikipedia.org/wiki/RGB_color_space)

Ο εντοπισμός μπορεί να πραγματοποιηθεί απευθείας στα πρωτογενή δεδομένα, όπως περιγράφεται στα [9], [10], [11], [12], [13], [14], [15], [16]. Το συγκριτικό πλεονέκτημα του χώρου RGB είναι η υπολογιστική ταχύτητα αφού δεν απαιτείται μετασχηματισμός σε άλλο σύστημα.

Πιο συγκεκριμένα, στο [9] πραγματοποιείται κατωφλίωση στον χώρο RGB. Μειονέκτημα της κατωφλίωσης στον συγκεκριμένο χώρο συνιστά η χρήση περίπλοκων κατωφλίων. Στη δημοσίευση αυτή χρησιμοποιούνται όρια που ενσωματώνουν τους λόγους των καναλιών:

$$a_{min} * G < B < a_{max} * G$$

$$\beta_{min} * B < R < \beta_{max} * B$$

$$\gamma_{min} * B < G < \gamma_{max} * B \quad (2.2 a - c)$$

Οι ελάχιστες και μέγιστες τιμές για τις παραμέτρους a , β , γ καθορίζονται κατόπιν δοκιμών σε πλήθος εικόνες. Για καλύτερα αποτελέσματα, καθώς κατά τις πολύ πρωινές και απογευματινές ώρες η κατωφλίωση δεν είναι ικανοποιητική, αξιολογούνται οι συνθήκες φωτεινότητας και ακολουθεί χρωματική εξισορρόπηση. Για την αξιολόγηση των συνθηκών φωτεινότητας, συνήθως, εντοπίζεται μία επιφάνεια με λευκό χρώμα και συγκρίνονται οι τιμές φωτεινότητας της λήψης, με τις τιμές (255,255,255), που αποτελούν τις τιμές του απόλυτου λευκού. Καθώς στις εικόνες δρόμων δεν απεικονίζονται συχνά λευκές επιφάνειες, οι ερευνητές χρησιμοποίησαν ως επιφάνεια αναφοράς το οδόστρωμα και συνέκριναν τις τιμές φωτεινότητάς του στη λήψη με τις τιμές φωτεινότητας του οδοστρώματος σε ιδανικές συνθήκες. Μετά από αξιολόγηση ακολουθεί και η εξισορρόπηση της φωτεινότητας με διόρθωση της καμπύλης γάμμα (gamma-correction).



Εικόνα 2.2: Αρχική εικόνα (αριστερά) και Εικόνα μετά από χρωματική εξισορρόπηση (δεξιά)

(Πηγή:[9])

Στο [10] τα κατώφλια υπολογίζονται με τη βοήθεια μεταβλητής k , η οποία για κάθε εικονοψηφίδα ισούται με:

$$k = \frac{255}{\max(R, G, B)} \quad (2.3)$$

Για τον εντοπισμό του κόκκινου και του μπλε χρώματος χρησιμοποιούνται οι ακόλουθες σχέσεις αντίστοιχα:

$$\begin{aligned} & (\tilde{I}_{i,j}^R - \tilde{I}_{i,j}^G > 10) \text{ AND } (\tilde{I}_{i,j}^R - \tilde{I}_{i,j}^B > 10) \\ & M_{i,j} - \tilde{I}_{i,j}^R > 0.4 * \tilde{I}_{i,j}^R \end{aligned} \quad (2.4 a - b)$$

όπου:

$$\begin{aligned} \tilde{I}_{i,j}^R &= I_{i,j}^R * k \\ \tilde{I}_{i,j}^B &= I_{i,j}^B * k \\ \tilde{I}_{i,j}^G &= I_{i,j}^G * k \\ M_{i,j} &= \max(\tilde{I}_{i,j}^B, \tilde{I}_{i,j}^G) \end{aligned}$$

Στο [14] έχουν εντοπιστεί οι τιμές φωτεινότητας για τα βασικά χρώματα των πινακίδων σε λήψεις με ιδανικές συνθήκες και συγκρίνονται με τις τιμές φωτεινότητας κάθε εικονοψηφίδας, με το μέτρο της Ευκλείδειας απόστασης. Εάν η διαφορά κείται εντός συγκεκριμένων ορίων θεωρείται ότι η εικονοψηφίδα λαμβάνει την τιμή 1, διαφορετικά της αποδίδεται η τιμή 0.

Πιο συγκεκριμένα, κάθε χρώμα προκύπτει από γραμμικό μετασχηματισμό των τριών βασικών χρωμάτων (κόκκινο, πράσινο και μπλε), όπως υπαγορεύει κι η ακόλουθη σχέση:

$$c = c_1 * Red + c_2 * Green + c_3 * Blue \quad (2.5), \text{ όπου } 0 \leq c_i \leq 1$$

Έστω το χρώμα αναφοράς $r(r_1, r_2, r_3)$ και το μέτρο ομοιότητας μεταξύ τους:

$$d = \sqrt{(r_1 - c_1)^2 + (r_2 - c_2)^2 + (r_3 - c_3)^2} \quad (2.6) .$$

Η σύγκριση περιγράφεται μαθηματικά από την ακόλουθη σχέση:

$$F(c) = \begin{cases} (0,0,0) & \text{εάν } d > t \\ r & \text{εάν } d < t \end{cases} \quad (2.7)$$

όπου t : η τιμή του κατωφλίου

Στο [15] ο εντοπισμός για κόκκινες πινακίδες πραγματοποιείται με τη μέθοδο επέκτασης περιοχής ενδιαφέροντος (region growing). Αρχικά εντοπίζονται εικονοψηφίδες με συγκεκριμένες ιδιότητες, που αποτελούν τον πυρήνα (seed) και κατόπιν ελέγχεται η ομοιότητα όλων των γειτονικών εικονοψηφίδων με τον πυρήνα, με κάποιο μέτρο ομοιότητας, όπως η διαφορά φωτεινότητας. Εάν το μέτρο ομοιότητας έχει τιμή εντός συγκεκριμένων ορίων, η αρχική περιοχή ενδιαφέροντος επεκτείνεται με τις νέες εικονοψηφίδες. Η επιτυχία της μεθόδου εξαρτάται άμεσα από την αρχική επιλογή των εικονοψηφίδων του πυρήνα, το μέτρο ομοιότητας και τα όρια. Ειδικότερα, στη συγκεκριμένη δημοσίευση, οι συγγραφείς αρχικά εντοπίζουν όσες εικονοψηφίδες εμφανίζουν τιμές:

$$R > 128, \quad B > 0.8R, \quad G > 0.8R \quad (2.8 \text{ a} - \text{c})$$

και τις ορίζουν ως πυρήνες. Τα παραπάνω όρια, τα οποία υπολογίστηκαν κατόπιν δοκιμών, προκύπτουν, επειδή στις κόκκινες πινακίδες, η κόκκινη συνιστώσα εμφανίζει τιμή αρκετά υψηλή και μεγαλύτερη από τις άλλες δύο συνιστώσες. Στη συνέχεια, εξετάζονται παράθυρα μεγέθους 15x15 εικονοψηφίδων, γύρω από κάθε πυρήνα, για τον αριθμό των λευκών και κόκκινων εικονοψηφίδων που περιέχουν. Ανάλογα με τον αριθμό των λευκών και κόκκινων εικονοψηφίδων που υπάρχουν στο παράθυρο, αποφασίζεται εάν η περιοχή πρέπει να εξεταστεί περαιτέρω. Στον τελικό έλεγχο, σε κάθε εικονοψηφίδα της εξεταζόμενης περιοχής αποδίδεται η τιμή 1, εάν περιλαμβάνεται στο διάστημα από $-2\sigma+m$ μέχρι $2\sigma+m$, με m η μέση τιμή του κόκκινου ή του λευκού χρώματος. Διαφορετικά της αποδίδεται η τιμή 0. Εάν το σύνολο των εικονοψηφίδων με τιμή διαφορετική του 0, ξεπερνά το 95%, το παράθυρο προσαρτάται στην περιοχή ενδιαφέροντος.

Στο [16] χρησιμοποιούνται δυναμικά κατώφλια για την απομόνωση των περιοχών ενδιαφέροντος στις εικόνες, που προέρχονται από ένα σύστημα Mobile Mapping. Αρχικά χρησιμοποιήθηκε σταθερό κατώφλι, που προκύπτει από τη σχέση:

$$b_k(i, j) = \begin{cases} 1 & \text{για } \sigma_k - t_k \leq g_k(i, j) \leq \sigma_k + t_k \\ 0 & \text{για } \text{κάθε άλλη περίπτωση} \end{cases} \quad (2.9)$$

όπου σ_k οι φωτεινότητες ενός συγκεκριμένου χρώματος k , t_k η ανοχή απόκλισης από τη συγκεκριμένη φωτεινότητα και $g_k(i, j)$ οι φωτεινότητες του προς εξέταση εικονοστοιχείου.

Οι συγγραφείς εφάρμοσαν τη συγκεκριμένη σχέση αλλά παρατηρήθηκαν προβλήματα σε εικόνες ειλημμένες υπό από διαφορετικές συνθήκες φωτεινότητας, καθώς οι τιμές που ορίστηκαν δε λειτουργούσαν σε όλες τις περιπτώσεις. Μία λύση αποτελούσε η αύξηση της ανοχής απόκλισης, η οποία όμως εισήγαγε μεγάλα ποσοστά

θορύβου στη δυαδική εικόνα. Έτσι αποφασίστηκε η εισαγωγή δυναμικών κατωφλίων, σύμφωνα με τη σχέση:

$$b_k(i,j) = \begin{cases} 1 & \text{για } u_k - t_k \leq r_k(i,j) \leq u_k + t_k \\ 0 & \text{για } \text{κάθε άλλη περίπτωση} \end{cases} \quad (2.10)$$

Οι παράμετροι έχουν την ίδια σημασία όπως στη σχέση (2.9), με τη διαφορά ότι αναφέρονται στους λόγους μεταξύ καναλιών.

Στη δυαδική εικόνα που προκύπτει εφαρμόζεται ένα φίλτρο modal, με παράθυρο 7x7, για την απομάκρυνση του θορύβου. Το συγκεκριμένο φίλτρο υπολογίζει σε ένα παράθυρο, τον αριθμό των λευκών και μαύρων εικονοσηφίδων και όταν ο αριθμός των μαύρων εικονοσηφίδων είναι μικρότερος του 15, η κεντρική εικονοσηφίδα θεωρείται θόρυβος.

Τα αποτελέσματα της κατωφλίωσης συνιστούν διαφορετικούς πυρήνες ώστε να ακολουθήσει στη συνέχεια η τεχνική επέκτασης περιοχής ενδιαφέροντος.

2.1.1.2 Εντοπισμός στους χώρους HSI και HSV

Η πλειοψηφία των ερευνητικών εργασιών χρησιμοποιεί για το στάδιο του εντοπισμού είτε τον χώρο HSV (Hue, Saturation, Value – Απόχρωση, Κορεσμός, Τόνος), είτε τον χώρο HSI (Hue, Saturation, Intensity – Απόχρωση, Κορεσμός, Ένταση).

Ο χρωματικός χώρος HSV προσημειώνει αρκετά ικανοποιητικά την αντίληψη των χρωμάτων από τον άνθρωπο και αναπαριστάται σχηματικά στην Εικόνα 2.3. Η πρώτη συνιστώσα αναφέρεται τη χροιά του χρώματος, η οποία απεικονίζεται στον κυκλικό δίσκο της κορυφής και λαμβάνει τιμές από το 0° -360°. Η δεύτερη συνιστώσα αφορά τον κορεσμό, ο οποίος μειώνεται, όπως φαίνεται και στην ακόλουθη εικόνα, από τα άκρα προς το κέντρο και δηλώνει την καθαρότητα του χρώματος. Οι τιμές του κορεσμού κυμαίνονται μεταξύ 0 και 1. Τέλος, η τρίτη συνιστώσα αναφέρεται στο πόσο φωτεινό ή σκοτεινό είναι το εικονοστοιχείο. Όσο πιο κοντά στο λευκό είναι μία εικονοσηφίδα, τόσο μεγαλύτερη είναι η τρίτη συνιστώσα, η οποία έχει πεδίο ορισμού μεταξύ 0 και 1.

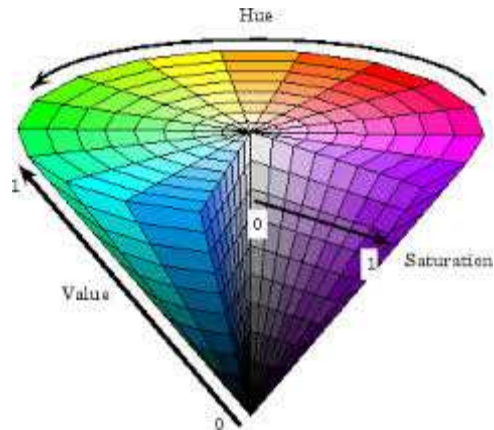
Για τη μετατροπή από τον χρωματικό χώρο RGB στον χρωματικό χώρο HSV χρησιμοποιούνται οι παρακάτω σχέσεις:

$$H = \begin{cases} \theta & \text{εάν } B \leq G \\ 360 - \theta & \text{εάν } B \geq G \end{cases}$$

$$S = 1 - \frac{3}{(R + G + B)} [\min(R, G, B)]$$

$$V = \frac{\max(R, G, B)}{255} \quad (2.11a - c)$$

$$\text{όπου } \theta = \cos^{-1} \left\{ \frac{\frac{1}{2}[(R-G)+(R-B)]}{[(R-G)^2+(R-B)(G-B)]^{1/2}} \right\}$$



Εικόνα 2.3: Ο χρωματικός χώρος HSV

(Πηγή: <http://www.mathworks.com/help/images/converting-color-data-between-color-spaces.html>)

Τα αποτελέσματα κατάτμησης στον συγκεκριμένο χρωματικό χώρο είναι ιδιαίτερα ικανοποιητικά, καθώς είναι δυνατόν να εξαλειφθούν οι επιπτώσεις της διαφορετικής φωτεινότητας, εάν δεν ληφθεί υπόψη το τρίτο κανάλι. Ουσιαστικά, η πληροφορία του χρώματος καταγράφεται στις δύο πρώτες συνιστώσες H και S του χώρου.

Παραδείγματα ερευνητικών εργασιών που χρησιμοποιούν τον χρωματικό χώρο HSV είναι οι [17], [18], [19], [20], [21], [22], [23], [24], [25], [26], [27], [28].

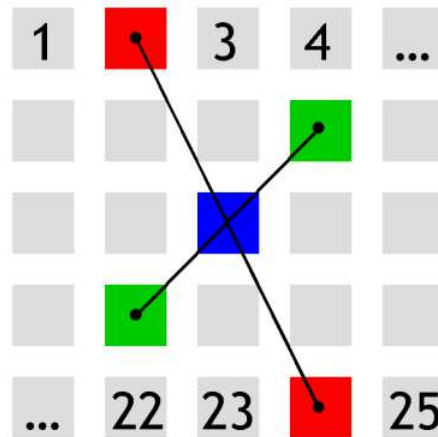
Στο [23] πραγματοποιείται εντοπισμός με κατωφλίωση στον χρωματικό χώρο HSV. Επιλέγονται μόνο τα εικονοστοιχεία με υψηλή τιμή για τη συνιστώσα V, καθώς μικρές τιμές της οδηγούν σε ασταθή αποτελέσματα. Τα καθολικά κατώφλια, που χρησιμοποιήθηκαν για κάθε χρώμα είναι:

$$H < 20 \text{ ή } H > 300 \text{ και } S > 0.3 \text{ για κόκκινο χρώμα}$$

$$20 < H < 70 \text{ και } S > 0.27 \text{ για κίτρινο χρώμα}$$

$$180 < H < 280 \text{ και } S > 0.25 \text{ για μπλε χρώμα (2.12 a - c)}$$

Στο [17] εφαρμόζεται αρχικά ένα φίλτρο SNN (Symmetric Nearest Neighbor) για απομάκρυνση του θορύβου. Το φίλτρο SNN [19] συγκρίνει τις τιμές των συμμετρικών εικοσηφίδων (όπως φαίνεται και στην Εικόνα 2.4) και επιλέγει τη τιμή που βρίσκεται κοντινότερα στη τιμή της κεντρικής εικοσηφίδας. Τέλος, αποδίδει στην κεντρική εικοσηφίδα τον μέσο όρο των τιμών φωτεινότητας των επιλεγμένων εικοσηφίδων. Σημαντικό πλεονέκτημα του φίλτρου, είναι ότι απομακρύνει τον θόρυβο, διατηρώντας όμως τις ακμές.



Εικόνα 2.4: Τα συμμετρικά εικονοστοιχεία

(Πηγή:http://www.subsurfwiki.org/wiki/Symmetric_nearest_neighbour_filter)

Ακολουθεί η μετατροπή στον χρωματικό χώρο HSV και επιλέγονται διάφοροι πυρήνες ανάλογα με τις τιμές H και S, που εμφανίζουν οι εικονοψηφίδες. Για κάθε μία υπολογίζεται μία μέση τιμή και το κριτήριο ομοιότητας, που καθορίζει εάν μία εικονοψηφίδα θα προσαρτηθεί στον πυρήνα, είναι η ευκλείδεια απόσταση μεταξύ της μέσης τιμής και της τιμής της εικονοψηφίδας στον κυλινδρικό χώρο HSV. Η τιμή αυτή πρέπει να είναι εντός ενός κατωφλίου, το οποίο είναι δυναμικό και ορίζεται κάθε φορά με βάση την τιμή κορεσμού στην περιοχή του πυρήνα, όπως ορίζει η ακόλουθη σχέση:

$$a = k - \sin(s_{seed}) \quad (2.13)$$

όπου a η τιμή του κατωφλίου, k μία παράμετρος κανονικοποίησης και s_{seed} η τιμή κορεσμού στον πυρήνα.

Στο [20] ο συγγραφέας εφαρμόζει ένα σεντ κανόνων ασαφούς λογικής προκειμένου να ορίσει τις περιοχές ενδιαφέροντος, δηλαδή τις περιοχές με κόκκινο, κίτρινο, μπλε, πράσινο και μαύρο χρώμα. Η εφαρμογή των κανόνων ασαφούς λογικής για κατάτμηση εικόνας είναι μία μέθοδος που εξετάζει κάθε εικονοψηφίδα ξεχωριστά, όπως και η κατωφλίωση, και αποδίδει σε κάθε εικονοψηφίδα ένα συγκεκριμένο χρώμα. Έχουν προταθεί διάφοροι τρόποι για αυτόματη δημιουργία των κανόνων, όπως αναφέρεται στο [21].

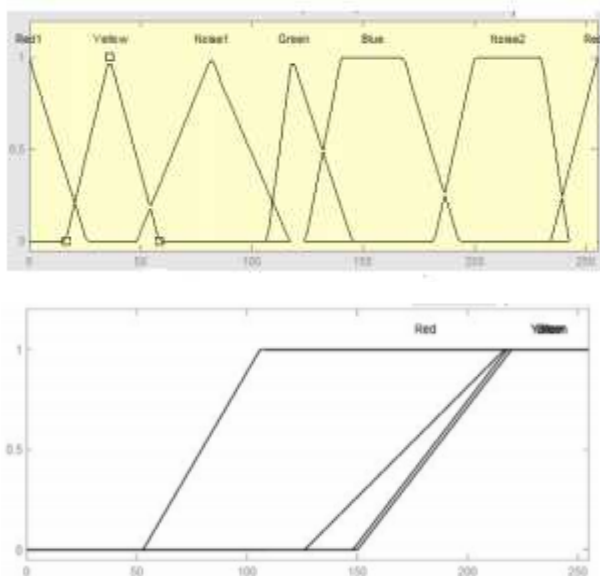
Στο [20] η εικόνα αρχικά μετασχηματίζεται στον χρωματικό χώρο HSV και εφαρμόζονται οι παρακάτω επτά κανόνες:

1. If (Hue is Red1) and (Saturation is Red) then (Result is Red)
2. If (Hue is Red2) and (Saturation is Red) then (Result is Red)
3. If (Hue is Yellow) and (Saturation is Yellow) then (Result is Yellow)
4. If (Hue is Green) and (Saturation is Green) then (Result is Green)
5. If (Hue is Blue) and (Saturation is Blue) then (Result is Blue)

6. If (Hue is Noise1) then (Result is Black)

7. If (Hue is Noise2) then (Result is Black)

Οι τιμές για τα Red1, Red2 κλπ προκύπτουν από τα ακόλουθα διαγράμματα, που αναπαριστούν τις συναρτήσεις συμμετοχής για τη χροιά και τον κορεσμό. Στην ασαφή λογική, οι συναρτήσεις συμμετοχής δείχνουν το βαθμό της αλήθειας για μία συγκεκριμένη τιμή μιας μεταβλητής.



Εικόνα 2.5: (α) Συνάρτηση συμμετοχής για χροιά (β) Συνάρτηση συμμετοχής για κορεσμό

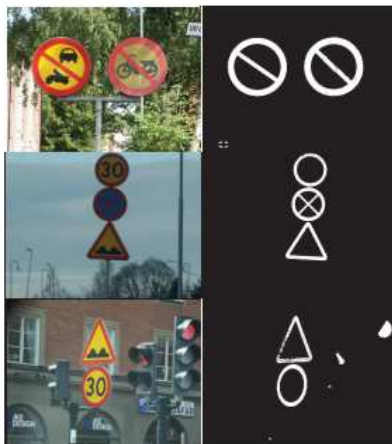
(Πηγή:[20])

Στο [28] η κατάτμηση εικόνας πραγματοποιείται με τον αλγόριθμο μηχανικής μάθησης Adaboost. Σκοπός του αλγορίθμου είναι να κατατάξει τα δεδομένα που δεχεται σε δύο κλάσεις. Επιστρέφει αποτέλεσμα 1 αν τα δεδομένα ανήκουν στην κλάση και -1 εάν δεν ανήκουν σε αυτή. Το όνομά του προέρχεται από το Adaptive Boosting, αφού ο αλγόριθμος είναι δυναμικός και χρησιμοποιεί πολλές επαναλήψεις με αδύναμους ταξινομητές (weak classifiers), προκειμένου να δημιουργηθεί ένας ισχυρός ταξινομητής (strong classifier). Οι αδύναμοι ταξινομητές έχουν χαμηλή συσχέτιση με το πραγματικό όριο μεταξύ των δύο κλάσεων και συνήθως αποτυγχάνουν να ταξινομήσουν επιτυχώς τα δεδομένα, σε αντίθεση με τους ισχυρούς ταξινομητές που πλησιάζουν κατά πολύ το πραγματικό όριο.

Στη συγκεκριμένη περίπτωση δεδομένα εισόδου αποτελούν οι εικονοψηφίδες της λήψης, οι οποίες έχουν μετασχηματιστεί στον χρωματικό χώρο HSV.

Ο αλγόριθμος είχε εκπαιδευτεί με πληθώρα λήψεων, οι οποίες κατατμήθηκαν χειροκίνητα. Οι περιοχές των σημάτων χρησιμοποιήθηκαν ως θετικά παραδείγματα (κλάση 1) και οι υπόλοιπες ως αρνητικά (κλάση -1). Όλες οι εικονοψηφίδες μεταφέρθηκαν στον χρωματικό χώρο HSV και δημιουργήθηκε ένας πίνακας που σε κάθε γραμμή περιείχε τις συντεταγμένες H και S και την κλάση στην οποία ανήκει η κάθε

εικονοψηφίδα. Με την εκπαίδευση του αλγορίθμου, προέκυψε ο τελικός δυνατός ταξινομητής, που εφαρμόζεται σε κάθε εικονοψηφίδα της λήψης, ώστε να τις κατατάξει σε μία από τις δύο κλάσεις. Στην τελική δυαδική εικόνα στις εικονοψηφίδες που ανήκουν στην κλάση 1 αποδίδεται η τιμή 1 και στις υπόλοιπες η τιμή 0. Τα αποτελέσματα του αλγορίθμου φαίνονται στην Εικόνα 2.6.



Εικόνα 2.6: Αποτέλεσμα εφαρμογής του αλγορίθμου Adaboost

(Πηγή:[28])

Παρόμοια λογική ακολουθείται στο [18], όπου δημιουργείται ένας αντίστοιχος πίνακας με τις τιμές για τη χροιά και την απόχρωση κάθε εικονοψηφίδας και η ταξινόμηση γίνεται με χρήση Support Vector Machines (SVM). Τα μοντέλα αυτά χρησιμοποιούνται στη μηχανική μάθηση ως αλγόριθμοι ταξινόμησης και αφού έχουν εκπαιδευτεί με ένα σύνολο δεδομένων N διαστάσεων, υπολογίζουν το βέλτιστο υπερεπίπεδο για το διαχωρισμό των κλάσεων (για περισσότερα βλ. Κεφάλαιο 3). Όπως και στο [28], η τεχνική εφαρμόζεται σε όλες τις εικονοψηφίδες για τον εντοπισμό των κατάλληλων χρωμάτων και την απομόνωση των περιοχών ενδιαφέροντος.

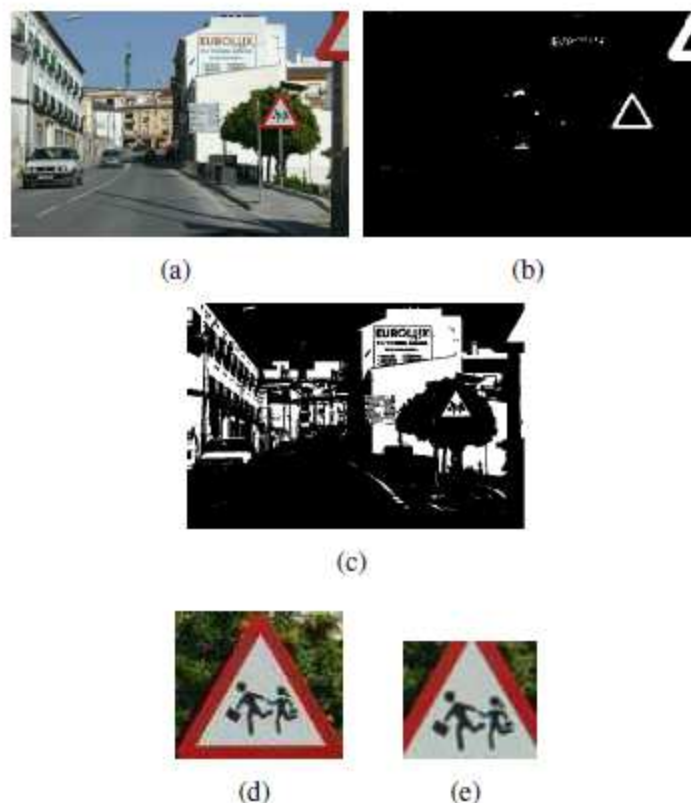
Ο χρωματικός χώρος HSI παρουσιάζει πολλές ομοιότητες με τον HSV και αναλύεται εκτενώς στο Κεφάλαιο 4, στα πλαίσια της εφαρμογής που αναπτύχθηκε. Μοναδική διαφορά μεταξύ των δύο χώρων αποτελεί η τρίτη συνιστώσα, η οποία υπολογίζεται διαφορετικά. Χαρακτηριστικά παραδείγματα εφαρμογών στον χρωματικό χώρο HSI αποτελούν τα [29], [30], [31], [32], [33], [34], [35], [36], [37],[38]. Παρακάτω αναλύονται ορισμένες από τις εργασίες αυτές.

Στο [29], οι Maldonado et al υπολόγισαν σταθερά κατώφλια για τα μπλε και κόκκινα σήματα μελετώντας τα ιστογράμματα πινακίδων οδικής κυκλοφορίας ελλημμένων με διαφορετικές καιρικές συνθήκες και συνθήκες φωτεινότητας καθώς και από διαφορετικές φωτομηχανές. Τα κατώφλια αφορούν τις δύο πρώτες συνιστώσες του χώρου HSI, ενώ για το λευκό χρώμα, που δεν είναι δυνατόν να εντοπιστεί με τον συγκεκριμένο τρόπο, χρησιμοποιούν την ακόλουθη συνάρτηση.

$$f = \frac{(|R - G| + |G - B| + |B - R|)}{3D} \quad (2.14)$$

όπου D ο βαθμός εξαγωγής ενός ουδέτερου χρώματος (στη δεδομένη περίπτωση τέθηκε ίσος με 20). Εάν η παραπάνω συνάρτηση έχει τιμή μεγαλύτερη της μονάδας, καταδεικνύεται η ύπαρξη λευκού χρώματος.

Παρόμοια μέθοδος για τον εντοπισμό του λευκού ακολουθείται και στο [30]. Πιο συγκεκριμένα, ο εντοπισμός των διάφορων χρωμάτων πραγματοποιείται ξεχωριστά από τον εντοπισμό του λευκού χρώματος και στο τέλος τα αποτελέσματα ενώνονται, όπως φαίνεται και στην Εικόνα 2.7. Με τον τρόπο αυτό, ο αλγόριθμος εντοπίζει τόσο την περίμετρο της πινακίδας, όσο και το εσωτερικό της, στο οποίο συνήθως σε μεγάλο ποσοστό συμμετέχει το λευκό χρώμα. Ο εντοπισμός των χρωμάτων πραγματοποιείται με κατωφλίωση στον χώρο HSI, χωρίς τη συμμετοχή της τρίτης συνιστώσας ώστε τα αποτελέσματα να είναι ανεξάρτητα της φωτεινότητας της εικόνας.

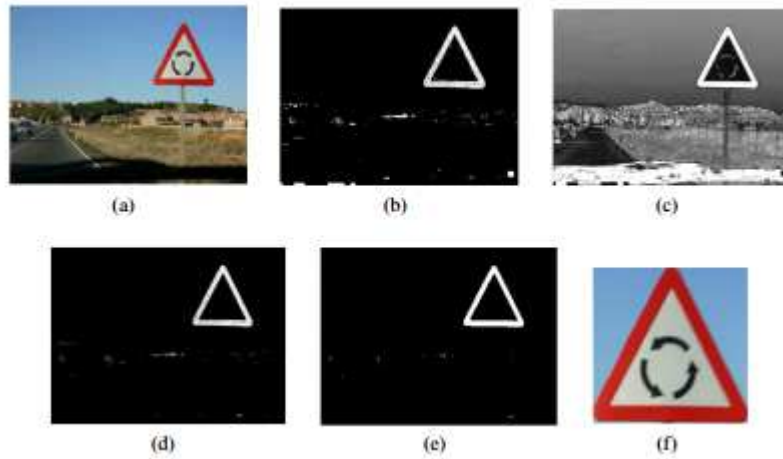


Εικόνα 2.7: (a) Η αρχική εικόνα, (b) ο εντοπισμός των χρωμάτων, (c) ο εντοπισμός του λευκού, (d) και (e) αποτελέσματα εντοπισμού

(Πηγή:[29])

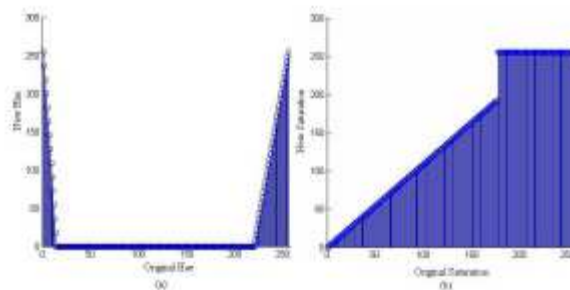
Στο [36], αφού η εικόνα έχει μετασχηματιστεί στον χώρο HSI, πραγματοποιείται ενίσχυση εικόνας (image enhancement) με τη βοήθεια ειδικών διαγραμμάτων (Look Up Tables). Η χρήση των LUTs είναι κοινή τεχνική στην επεξεργασία εικόνας για τον μη γραμμικό μετασχηματισμό των δεδομένων σε μία πιο επιθυμητή μορφή. Η εφαρμογή τους γίνεται για όλες τις εικονοπηφίδες και τα διαγράμματα ορίζουν την τιμή για κάθε εικονοπηφίδα στη βελτιωμένη εικόνα με βάση την τιμή της στην αρχική εικόνα.

Στη συγκεκριμένη δημοσίευση, οι συνιστώσες που μετασχηματίζονται είναι η χροιά και ο κορεσμός για τις εικονοψηφίδες με κόκκινο, κίτρινο και μπλε χρώμα. Σε κάθε εικόνα που αναπαριστά μία από τις δύο συνιστώσες του χώρου HSI εφαρμόζεται ξεχωριστά ο μετασχηματισμός και οι δύο εικόνες που προκύπτουν πολλαπλασιάζονται και κανονικοποιούνται στο διάστημα [0-255], όπως φαίνεται και στην Εικόνα 2.8. Η τελική περιοχή ενδιαφέροντος προκύπτει με εφαρμογή δυναμικών κατωφλίων.



Εικόνα 2.8: (a) Αρχική εικόνα, (b) Εικόνα χροιάς μετά από εφαρμογή LUT, (c) Εικόνα κορεσμού μετά από εφαρμογή LUT, (d) συνδυασμός των δύο εικόνων, (e) αποτέλεσμα κατωφλίωσης, (f) εξαγωγή ROI.

(Πηγή:[36])



Εικόνα 2.9: LUTs για χροιά και κορεσμό για το κόκκινο χρώμα

(Πηγή:[36])

Οι delaEscalera et. al στο [31] ακολουθούν παρόμοια τακτική για τον εντοπισμό των χρωμάτων κόκκινο, κίτρινο και μπλε. Εφαρμόζουν, όπως και στο [36], βελτίωση εικόνας με χρήση LUTs και στη συνέχεια κατωφλίωση. Για τον εντοπισμό του εσωτερικού των σημάτων, που αποτελούνται από ουδέτερα χρώματα, εισήγαγαν μία εξίσωση χρωματικής ενέργειας $E(A)$, η οποία είναι μία συνάρτηση κόστους:

$$E(Z) = \frac{1}{A_\alpha} \sum_{(x,y) \in S_\alpha} A(x,y) \quad (2.15)$$

όπου S_a η περιοχή του σήματος που έχει εντοπιστεί, A_a το εμβαδό της περιοχής και $A(x, y)$ η χρωματική εικόνα, η οποία μετρά την ύπαρξη ουδέτερου χρώματος.

Καθώς πρόκειται για μία συνάρτηση κόστους, στο εσωτερικό της πινακίδας (όπου εμφανίζονται τα ουδέτερα χρώματα) η συνάρτηση οφείλει να εμφανίζει χαμηλή τιμή. Για τη δημιουργία της χρωματικής εικόνας σημαντικό ρόλο παίζει η συνιστώσα του κορεσμού, καθώς όσο μικρότερος είναι, τόσο πιο ουδέτερο είναι το χρώμα. Για αποτελέσματα πιο σταθερά απέναντι στον θόρυβο, η εικόνα δε δημιουργείται με επιλογή του καναλιού του κορεσμού αλλά με τον ακόλουθο τρόπο. Αρχικά, υπολογίζεται η τιμή L για κάθε εικονοστοιχείο της αρχικής RGB εικόνας:

$$L(x, y) = 0.3 * R(x, y) + 0.6 * B(x, y) + 0.1 * G(x, y) \quad (2.16)$$

και στη συνέχεια, το κανονικοποιημένο σφάλμα για κάθε κανάλι:

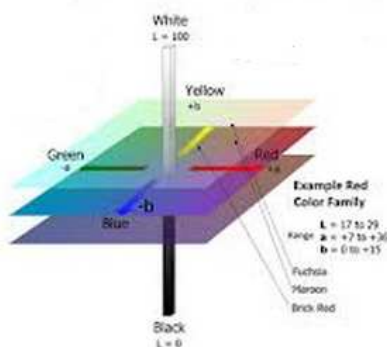
$$\begin{aligned} \varepsilon_R &= \frac{|L(x, y) - R(x, y)|}{L(x, y)} \\ \varepsilon_G &= \frac{|L(x, y) - G(x, y)|}{L(x, y)} \\ \varepsilon_B &= \frac{|L(x, y) - B(x, y)|}{L(x, y)} \end{aligned} \quad (2.17 \text{ a} - \text{c})$$

Όσο πιο μικρά είναι τα τρία κανονικοποιημένα σφάλματα τόσο πιο ουδέτερο είναι το χρώμα που φέρει η εικονοψηφίδα. Η τελική εικόνα προκύπτει με επιλογή της μεγαλύτερης από τις παραπάνω τιμές:

$$A(x, y) = \max\{\varepsilon_R, \varepsilon_G, \varepsilon_B\} \quad (2.18)$$

2.1.1.3 Μετασχηματισμός σε άλλους χρωματικούς χώρους

Τέλος, υπάρχουν ερευνητές που επιλέγουν τον μετασχηματισμό σε έναν λιγότερο συνηθισμένο χώρο. Στο [39] οι Siogas & Dermatas μετασχηματίζουν στον χρωματικό χώρο $L^*a^*b^*$, ο οποίος προσομοιάζει την ανθρώπινη όραση (όπως αντίστοιχα ο HSV). Ο χώρος αυτός είναι γνωστός και ως CIELab. Τα αρχικά CIE υποδηλώνουν ότι οι χώροι δημιουργήθηκαν από τη Διεθνή Ένωση Φωτεινότητας (International Commission on Illumination – CIE). Η συνιστώσα L^* περιγράφει τη φωτεινότητα της εικόνας και λαμβάνει τιμές από 0 μέχρι 100, ενώ οι a^* και b^* καταγράφουν τη χρωματική πληροφορία. Η a^* καταγράφει τον λόγο του κόκκινου προς το πράσινο (θετικές τιμές υποδεικνύουν την ύπαρξη κόκκινου χρώματος και αρνητικές του πράσινου) και η b^* τον λόγο μπλε προς κίτρινο (θετικές τιμές υποδεικνύουν το μπλε χρώμα και αρνητικές το κίτρινο). Ο χρωματικός χώρος είναι τρισδιάστατος, συνεπώς απεικονίζεται τρισδιάστατα, αν και συχνά εμφανίζονται δισδιάστατα διάγραμμα που αποτελούν τομές για μία συγκεκριμένη τιμή φωτεινότητας ([40]). Παράδειγμα αποτελεί η Εικόνα 2.10.

Εικόνα 2.10: Τομή για δεδομένη τιμή L^*

(Πηγή: http://coolroofs.org/newsletters/CRRC_Issue_2_2012.html)

Δεν υπάρχουν σχέσεις μετατροπής από τον RGB στον CIELab, καθώς ο πρώτος χώρος εξαρτάται από την συσκευή αναπαραγωγής (π.χ. από την οθόνη που απεικονίζει τα χρώματα) ενώ ο δεύτερος όχι, κάτι που αποτελεί σημαντικό πλεονέκτημα. Για την μετατροπή του RGB στον CIELab πρέπει ο πρώτος να μετασχηματιστεί σε ένα χρωματικά απόλυτο χώρο (δηλαδή ένα χώρο που δεν εξαρτάται από τη συσκευή αναπαραγωγής, για παράδειγμα AdobeRGB). Η πιο συνηθισμένη μετάβαση στον CIELab γίνεται από τον χρωματικό χώρο CIEXYZ (οι σχέσεις υπάρχουν στο [40]).

Έτσι, στο [39] οι συγγραφείς μελετούν το κανάλι L^* για να εξάγουν συμπεράσματα σχετικά με τη φωτεινότητα της εικόνας και τις συνθήκες υπό από τις οποίες αυτή ελήφθη. Τα συμπεράσματα αυτά είναι χρήσιμα, καθώς οι πινακίδες εμφανίζονται φωτεινότερες όταν ο φωτισμός δεν είναι επαρκής (π.χ. κατά τη διάρκεια της νύχτας). Στη συνέχεια της εργασίας, πραγματοποιούν κατωφλίωση στα θετικά και αρνητικά τμήματα των καναλιών a^* και b^* (αφού πρώτα έχουν πολλαπλασιάσει τα αρνητικά τμήματα -1), με κατώφλι που ορίζεται από τον αλγόριθμο του Otsu ([41]). Αποτέλεσμα της διαδικασίας είναι τέσσερις δυαδικές εικόνες (μία για κάθε τμήμα), που ενώνονται για να δημιουργηθεί η τελική εικόνα.

Ο αλγόριθμος του Otsu ανήκει στις αυτοματοποιημένες τεχνικές για τη δημιουργία δύο κλάσεων σε μία εικόνα και βασίζεται είτε στην μεγιστοποίηση της μεταβλητότητας των διαφορετικών κλάσεων, είτε στην ελαχιστοποίηση της μεταβλητότητας μεταξύ των ίδιων κλάσεων. Πιο συγκεκριμένα, για κάθε πιθανό κατώφλι (τιμές από 0 μέχρι 255) υπολογίζει την πιθανότητα ω και τη μεταβλητότητα σ_i^2 για κάθε μία από τις δύο κλάσεις και επιλέγεται τελικά η τιμή που ικανοποιεί ένα από τα δύο κριτήρια (αναλυτικότερη παρουσίαση στο [41]).

Στο [42] χρησιμοποιείται ο χώρος YCrCb για τον εντοπισμό των πινακίδων με κόκκινο χρώμα. Ο YCrCb αποτελείται από το κανάλι Y (Luma) που καταγράφει τη φωτεινότητα και από δύο χρωματικά κανάλια Cr και Cb, που αναφέρονται στο κόκκινο και μπλε κανάλι του χρωματικού χώρου RGB, αφού όμως έχει προηγηθεί χρωματική διόρθωση με τη συνάρτηση γ . Σχέσεις μετατροπής μεταξύ των δύο χώρων υπάρχουν στο [43].

Για τον εντοπισμό των κόκκινων πινακίδων στον YCrCb οι ερευνητές πραγματοποιούν δυναμική κατωφλίωση στο κανάλι Cr.

Τέλος, στο [44] υπάρχει συγκριτική μελέτη ως προς τα αποτελέσματα κατωφλίωσης στους χώρους RGB, HSV, L*a*b, YCbCr, CYMK και YIQ, που παρέχει περισσότερες πληροφορίες σχετικά με τον εντοπισμό στους διάφορους χρωματικούς χώρους.

2.1.2 Εντοπισμός με βάση το σχήμα

Εκτός των μεθόδων εντοπισμού που βασίζονται στις χρωματικές ιδιότητες των πινακίδων, έχουν αναπτυχθεί και μέθοδοι εντοπισμού με βάση τις γεωμετρικές τους ιδιότητες. Τα σήματα οδικής κυκλοφορίας αποτελούνται στην ουσία από τα βασικά, δισδιάστατα γεωμετρικά σχήματα, όπως κύκλος, τρίγωνο, οκτάγωνο και ορθογώνιο (Εικόνα 2.11). Σε κάθε περίπτωση, η γεωμετρία τους είναι αυστηρή και το σχήμα τους συμμετρικό, επιτρέποντας τον εντοπισμό βάσει αυτών των χαρακτηριστικών.



Εικόνα 2.11: Τα γεωμετρικά σχήματα των πινακίδων

Μέχρι στιγμής έχει αναπτυχθεί πληθώρα ανάλογων τεχνικών, οι οποίες αναλύονται ακολούθως. Οι τεχνικές αυτές μπορούν να χρησιμοποιηθούν είτε επικουρικά στον εντοπισμό με βάση το χρώμα είτε αυτόνομα. Ιδιαίτερη σημασία δίνεται στην ανοχή ως προς πιθανές προοπτικές παραμορφώσεις (που προκύπτουν είτε από στροφές της κάμερας του αυτοκινήτου είτε από στροφές της πινακίδας) και αλλαγές κλίμακας.

Ο εντοπισμός με βάση το σχήμα υπερτερεί σε σχέση με τον εντοπισμό με βάση το χρώμα σε ορισμένες περιπτώσεις. Για παράδειγμα, τα χρώματα των πινακίδων μεταβάλλονται ανάλογα με τη χώρα. Συνεπώς, για να στηριχθεί κανείς στις χρωματικές ιδιότητες των σημάτων χρειάζεται αρχικά μία μελέτη των χρωμάτων των πινακίδων και το σύστημα εντοπισμού που τελικώς αναπτύσσεται έχει εφαρμογή μόνο στη συγκεκριμένη χώρα. Ενώ, ο εντοπισμός με βάση τις γεωμετρικές ιδιότητες είναι πιο ανεκτικός, καθώς τα χαρακτηριστικά του σχήματος δεν παρουσιάζουν αντίστοιχες μεταβολές. Άλλο ένα πλεονέκτημα του εντοπισμού με βάση το σχήμα είναι ότι δεν εξαρτάται από τις συνθήκες φωτεινότητας της λήψης και τις καιρικές συνθήκες καθώς και από τη κατάσταση των χρωμάτων των σημάτων, τα οποία αλλοιώνονται με την πάροδο των χρόνων.

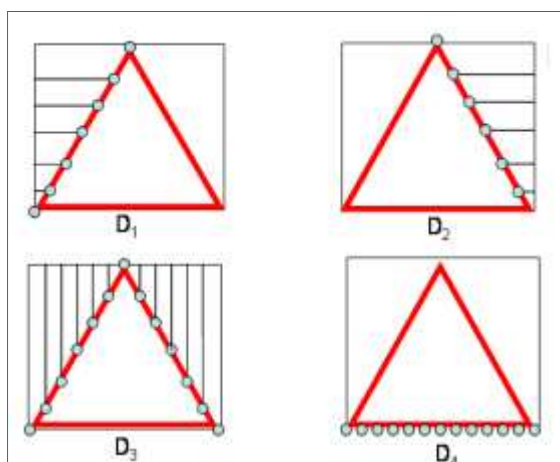
Σημαντικό μειονέκτημα όμως αποτελεί η απαίτηση περισσότερου χρόνου και υπολογιστικής ισχύος σε σχέση με την εντοπισμό που βασίζεται στο χρώμα.

Πιθανά προβλήματα στο συγκεκριμένο στάδιο ανακύπτουν λόγω της ύπαρξης αντικειμένων στην εικόνα με παρόμοια γεωμετρία, καθώς και λόγω πιθανών αποκρύψεων των σημάτων. Επίσης, πολλές φορές οι πινακίδες είναι κατεστραμμένες

στις γωνίες και το σχήμα τους αλλοιωμένο. Οι αλγόριθμοι που έχουν εφαρμοστεί μέχρι στιγμής και αναλύονται στη συνέχεια του κεφαλαίου προσπαθούν να αντιμετωπίσουν τα προαναφερθέντα προβλήματα.

2.1.2.1 Απόσταση από Όρια (Distance to Borders)

Μία από τις πιο διαδεδομένες μεθόδους που εμφανίζεται στα [29], [30], [33], [35] είναι η τεχνική Απόστασης από Όρια (Distance to Borders). Για κάθε περιοχή ενδιαφέροντος δημιουργείται ένα τετράπλευρο που εφάπτεται σε αυτή (bounding box). Η απόσταση της εξωτερικής ακμής της περιοχής ενδιαφέροντος από το τετράπλευρο μπορεί να δώσει πληροφορίες για το σχήμα της περιοχής. Η μέθοδος γίνεται πιο άμεσα κατανοητή με τη μελέτη του ακόλουθου σχήματος, που εμφανίζει τις αποστάσεις για μία τριγωνή πινακίδα. Η τεχνική αυτή χρησιμοποιείται συχνά καθώς είναι ανεξάρτητη από πιθανές μεταθέσεις. Ακόμα, αν το εφαπτόμενο τετράπλευρο στραφεί ώστε να ταυτιστεί με τον οριζόντιο άξονα, τα αποτελέσματα είναι ανεξάρτητα κι από ενδεχόμενες στροφές της περιοχής ενδιαφέροντος. Τέλος, με κανονικοποίηση των αποστάσεων με τρόπο ώστε να λαμβάνονται υπόψη οι διαστάσεις του τετραπλεύρου, η μέθοδος έχει ανοχή και ως προς τις διαφορές κλίμακας.

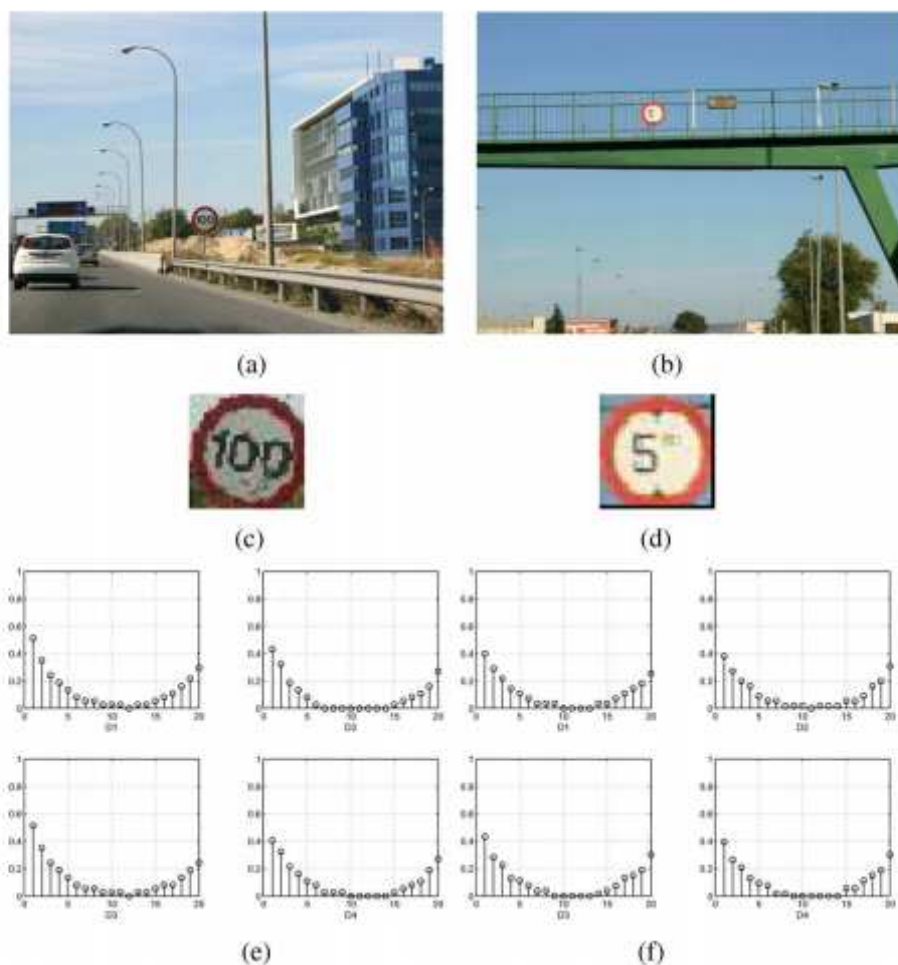


Εικόνα 2.12 : DtB για τριγωνικές πινακίδες

(Πηγή: [10])

Στο [29] τα δεδομένα από την εφαρμογή του DtB κατηγοριοποιούνται με τη χρήση SVM (λεπτομερής παρουσίαση των SVM γίνεται σε επόμενα κεφάλαια), τα οποία κατατάσσουν τα δεδομένα σε δύο κλάσεις $\{1, -1\}$. Το όριο μεταξύ των δύο κλάσεων είναι γραμμικό. Ο αλγόριθμος εκπαιδεύεται ξεχωριστά για κάθε μία από τις τέσσερις πλευρές του εφαπτόμενου τετραπλεύρου. Για την εκπαίδευση του χρησιμοποιείται ένας πίνακας που καταγράφει τις αποστάσεις πολλών δεδομένων καθώς και την κλάση που ανήκουν. Οι κάθετες αποστάσεις λαμβάνονται κατά μήκος της κάθε πλευράς σε σημεία που ισαπέχουν, με τέτοιο τρόπο ώστε να καταγράφονται είκοσι αποστάσεις για την κάθε πλευρά. Στη συνέχεια, για κάθε περιοχή ενδιαφέροντος που εντοπίστηκε υπολογίζονται τα δεδομένα DtB και εισάγονται στην SVM ώστε να υπολογιστεί η κλάση στην οποία ανήκουν. Αν εντοπιστούν παραπάνω από δύο θετικές ψήφοι (δηλαδή τουλάχιστον δύο πλευρές ανήκουν στην κλάση 1), για ένα

συγκεκριμένο σχήμα, η πινακίδα κατατάσσεται στην κατηγορία αυτή, διαφορετικά θεωρείται θόρυβος.



Εικόνα 2.13: (a-b) αρχικές εικόνες (c-d) αποτέλεσμα κατάτμησης με βάση το χρώμα (e-f) DtB για τις c και d αντίστοιχα

(Πηγή: [29])

Ανάλογα με το χρώμα που έχει εντοπιστεί χρησιμοποιούνται και οι ανάλογες SVM, καθώς τα πιθανά σχήματα είναι προκαθορισμένα βάσει του χρώματος των πινακίδων. Για παράδειγμα, στο κόκκινο χρώμα ελέγχεται εάν το σχήμα είναι κυκλικό ή τριγωνικό. Σε αυτή την περίπτωση, χρησιμοποιούνται δηλαδή, στο σύνολο οκτώ γραμμικά SVM (τέσσερα για κάθε πιθανό σχήμα) τα οποία επιστρέφουν αποτέλεσμα 1 εάν τα εξεταζόμενα δεδομένα ανήκουν στην κλάση και -1 σε κάθε άλλη περίπτωση.

Σημαντικό πλεονέκτημα της συγκεκριμένης τεχνικής είναι η ανοχή σε αποκρύψεις. Ακόμα κι αν υπάρχουν αποκρύψεις σε μία ή ακόμα και δύο πλευρές της πινακίδας, αυτή θα ταξινομηθεί σωστά με βάση τα αποτελέσματα του αλγορίθμου για τις δύο άλλες πλευρές.

Ανάλογη τακτική ακολουθείται και στα [30], [33],[35], όπου πριν τη ταξινόμηση τα εφαπτόμενα τετράπλευρα στρέφονται ως προς μία θέση αναφοράς, για ανεξαρτησία της μεθόδου από πιθανές στροφές.

2.1.2.2 Μετασχηματισμός Hough

Άλλη μία τεχνική που εφαρμόζεται για τον εντοπισμό πινακίδων με βάση τη γεωμετρία τους είναι ο Μετασχηματισμός Hough (Hough Transform), ο οποίος αρχικά χρησιμοποιήθηκε για τον εντοπισμό γραμμών (Hough lines) και στη συνέχεια επεκτάθηκε για εντοπισμό κύκλων (Κυκλικός Μετασχηματισμός Hough - Circular Hough Transform) και τυχαίων σχημάτων (Γενικευμένος Μετασχηματισμός Hough - Generalised Hough Transform).

Ο μετασχηματισμός εφαρμόζεται σε μια εικόνα κλίσεων. Όπως υπαγορεύεται και στο [45] οι κλίσεις της εικόνας (gradient) υποδηλώνουν τις αλλαγές της φωτεινότητας μεταξύ γειτονικών εικονοστοιχείων. Μαθηματικά, η κλίση είναι μία συνάρτηση δύο μεταβλητών που περιλαμβάνει τις πρώτες παραγώγους κατά την x και y συνιστώσα όπως φαίνεται και παρακάτω:

$$\nabla f(x, y) = \left[\frac{\partial f}{\partial x} \quad \frac{\partial f}{\partial y} \right]^T = [f_x \quad f_y]^T \quad (2.19)$$

Η εικόνα των κλίσεων μπορεί να προκύψει με εφαρμογή διάφορων τελεστών όπως Sobel και Prewitt (για περισσότερα βλ. [46] και [47] αντίστοιχα)

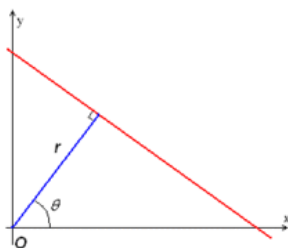
Μετά από περαιτέρω επεξεργασία της εικόνας των κλίσεων (π.χ. κατωφλίωση), εντοπίζονται οι ακμές, που αποτελούν τα σύνορα μεταξύ δύο ομοιογενών περιοχών με διαφορετική φωτεινότητα. Ο πιο συχνός τελεστής για εντοπισμό ακμών είναι ο Canny Edge Detector (για περισσότερα βλ. [48]). Η διαδικασία εντοπισμού ακμών αν και αποτελεί μία μέθοδο που βασίζεται στο χρώμα, της αναφέρεται και στο κεφάλαιο 2.1, τροφοδοτεί τον Μετασχηματισμό Hough, που αποτελεί μία μέθοδο εντοπισμού με βάση το σχήμα.

Ο Μετασχηματισμός Hough για τον εντοπισμό ευθειών στοχεύει στην αντίληψη της γραμμής στην εικόνα, όχι ως μεμονωμένα σημεία αλλά ως συνάρτηση συγκεκριμένων παραμέτρων. Αρχικά, χρησιμοποιήθηκε η παραμετροποίηση κλίσης-τομής (slope- intersection model):

$$y = ax + b \quad (2.20)$$

Όπου η μεταβλητή a εκφράζει την κλίση της ευθείας και b το σημείο τομής με τον άξονα των y . Σημαντικό μειονέκτημα της συγκεκριμένης αναπαράστασης είναι η περίπτωση κάθετων γραμμών, όπου η μεταβλητή a τείνει στο άπειρο. Για τον λόγο αυτό, προτιμάται η παραμετροποίηση της γραμμής με χρήση πολικών συντεταγμένων, οι οποίες καταγράφουν τη γωνία θ που σχηματίζει η κάθετη στην ευθεία με τον άξονα των x και την απόσταση r της ευθείας από την αρχή των αξόνων κατά μήκος της καθέτου (Εικόνα 2.14).

2. Μέθοδοι και τεχνικές εντοπισμού και αναγνώρισης σημάτων



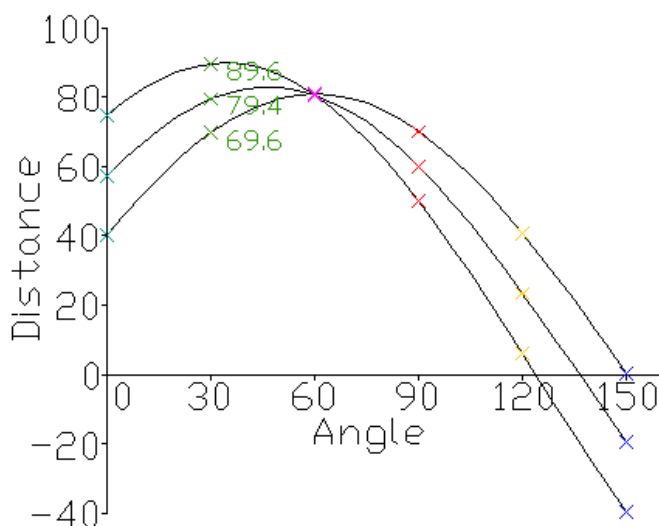
Εικόνα 2.14: Οι πολικές συντεταγμένες

(Πηγή: http://en.wikipedia.org/wiki/Hough_transform)

Η παραμετροποίηση με χρήση πολικών συντεταγμένων καταγράφεται μαθηματικά ως εξής:

$$x\cos\theta + y\sin\theta = \rho \quad (2.21)$$

Κάθε ευθεία στο επίπεδο x-y χαρακτηρίζεται από μοναδικό ζευγάρι τιμών (ρ, θ) και αναπαρίσταται ως ένα σημείο στο επίπεδο ρ - θ . Επίσης, κάθε ένα σημείο (x_0, y_0) που ανήκει στην ευθεία μετασχηματίζεται μέσω της σχέσης 2.21 σε μία ημιτονοειδή καμπύλη, στον χώρο των παραμέτρων. Όλα τα σημεία που ανήκουν στην ίδια ευθεία, θα αποτελούν ημιτονοειδείς καμπύλες που τέμνονται σε συγκεκριμένο σημείο (ρ, θ) , το οποίο δίνει τις πολικές συντεταγμένες της ευθείας, στην οποία ανήκουν τα σημεία. Ενώ αντίστοιχα, όλα τα σημεία που κείνται σε δεδομένη καμπύλη στο επίπεδο ρ - θ , αναπαριστούν ευθείες που διέρχονται από το ίδιο σημείο.



Εικόνα 2.15: Οι ημιτονοειδείς καμπύλες στον χώρο των παραμέτρων

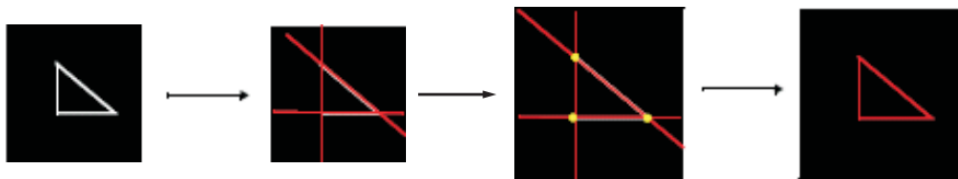
(Πηγή: http://en.wikipedia.org/wiki/Hough_transform)

Ο εντοπισμός των ευθειών μπορεί να γίνει, με βάση τα παραπάνω, εντοπίζοντας τομές των ημιτονοειδών καμπυλών στο επίπεδο ρ - θ . Επειδή όμως πρόκειται για μία αρκετά υπολογιστικά απαιτητική διαδικασία ο Hough ([49]) πρότεινε μία λίγο διαφορετική τεχνική, για την οποία ο χώρος ρ - θ δεν είναι πλέον συνεχής αλλά αποτελείται από διακριτές τιμές.

Αρχικά, υπολογίζονται ορισμένα διαστήματα για τις τιμές ρ και θ με συγκεκριμένο βήμα, ανάλογο του μέγιστου αποδεκτού σφάλματος. Όταν η γωνία $\theta \in [0, \pi]$, η απόσταση $\in \mathbb{R}$. Με βάση αυτά τα διαστήματα δημιουργείται ένας διδιάστατος πίνακας ψηφοφορίας (accumulator matrix) που περιέχει όλα τα πιθανά διαστήματα για τις παραμέτρους ρ , θ . Για κάθε ένα σημείο στην εικόνα των ακμών υπολογίζεται η ημιτονοειδής καμπύλη στον χώρο των παραμέτρων και αντίστοιχα στον διδιάστατο πίνακα ψηφοφορίας, προσαυξάνονται τα ανάλογα διαστήματα, στα οποία εμπίπτουν τα ρ και θ που υπολογίστηκαν για το σημείο. Στο τέλος, επιλέγονται τα ζευγάρια (ρ, θ) που έχουν συγκεντρώσει τις περισσότερες ψήφους, ως οι πιο επικρατούσες ευθείες στο επίπεδο $x-y$. Ο αριθμός των ψήφων για κάθε ζευγάρι τιμών αποτελεί στην ουσία τον αριθμό των σημείων που ανήκουν στη συγκεκριμένη ευθεία.

Πλεονεκτήματα της προσέγγισης είναι η ανοχή σε θόρυβο και εν μέρει αποκρύψεις, αλλά σημαντικό μειονέκτημα αποτελεί η εξάρτηση από την εικόνα ακμών (η οποία εξαρτάται από τη φωτεινότητα της αρχικής εικόνας).

Ο Μετασχηματισμός Hough χρησιμοποιείται και για εντοπισμό αντικειμένων στην λήψη (object detection). Παράδειγμα αποτελεί το επόμενο σχήμα, στο οποίο από την αρχική εικόνα ακμών εντοπίζονται οι ευθείες και με βάση τα σημεία τομής τους, εντοπίζεται τελικά ένα τρίγωνο. Με ανάλογο τρόπο χρησιμοποιείται ο αλγόριθμος για τον εντοπισμό σημάτων οδικής κυκλοφορίας με βάση τις γεωμετρικές τους ιδιότητες.



Εικόνα 2.16: Εντοπισμός τριγώνου από εικόνα ακμών

(Πηγή: *The Hough transform slides by Johan Carlson*)

Χαρακτηριστικά παραδείγματα εφαρμογής της τεχνικής για εντοπισμό πινακίδων αποτελούν τα [50], [51], [52].

Στο [50] εφαρμόζεται ο κλασικός Μετασχηματισμός Hough για τον εντοπισμό τριγώνων. Ο μετασχηματισμός εφαρμόζεται σε μία εικόνα ακμών, που προκύπτει από τον Canny Edge Detector. Από τις ακμές που εντοπίστηκαν γίνονται δεκτές μόνο όσες είναι «σχεδόν κλειστές» (closed contours) και το επαπτόμενο τετράπλευρο σε αυτές εμφανίζει λόγο ύψους-πλάτους εντός συγκεκριμένων ορίων. Κάθε αποδεκτή ακμή εξετάζεται με τον μετασχηματισμό και εάν εξαχθούν τρεις ευθείες με γωνίες τομής 60° , εντοπίζεται μία τριγωνική πινακίδα. Όπως φαίνεται και ακολούθως το τρίγωνο που θα εντοπιστεί με εφαρμογή του αλγορίθμου, μπορεί να αποτελεί είτε το εξωτερικό είτε το εσωτερικό τμήμα της πινακίδας.



Εικόνα 2.17: Εντοπισμός τριγωνικών πινακίδων

(Πηγή: [50])

Ίδια τακτική ακολουθείται και στο [51]. Σε κάθε περιοχή ενδιαφέροντος που έχει προκύψει από κατωφλίωση της εικόνας εφαρμόζεται ο Μετασχηματισμός Hough για εντοπισμό γραμμών. Στη συνέχεια υπολογίζονται οι γωνίες μεταξύ των ευθειών. Εάν εντοπιστούν τρεις γωνίες μεταξύ 50° - 70° , θεωρείται πως στη λήψη απεικονίζεται ένα τρίγωνο και υπολογίζεται το κέντρο βάρους του. Παρόμοια τακτική ακολουθείται και στο [52].

Αν και ο συγκεκριμένος μετασχηματισμός αρχικά ξεκίνησε για τον εντοπισμό γραμμών, στη συνέχεια επεκτάθηκε σε άλλες καμπύλες που μπορούν να παραμετροποιηθούν μαθηματικά όπως ο κύκλος και η έλλειψη.

Ο κύκλος μοντελοποιείται με τρεις παραμέτρους (δύο για το κέντρο κι έναν για την ακτίνα), όπως φαίνεται και στην ακόλουθη σχέση:

$$(x - x_o)^2 + (y - y_o)^2 = c^2 \quad (2.22)$$

Όπου x_o , y_o οι συντεταγμένες του κέντρου του κύκλου και c η ακτίνα. Για την παραλλαγή της μεθόδου, ο πίνακας ψηφοφορίας περιέχει πλέον τις τρεις προαναφερθείσες παραμέτρους και εντοπίζονται ως επικρατέστεροι κύκλοι οι συνδυασμοί με τις περισσότερες ψήφους, όπως αναφέρεται και στο [49] (για περισσότερα βλ. κεφάλαιο 4).

Αντίστοιχα, καθώς η έλλειψη παραμετροποιείται μαθηματικά με χρήση πέντε μεταβλητών, ο αλγόριθμος μπορεί να επεκταθεί και για τον εντοπισμό ελλείψεων. Ο πίνακας ψηφοφορίας στην δεδομένη παραλλαγή καταγράφει τα πιθανά διαστήματα και για τις πέντε μεταβλητές.

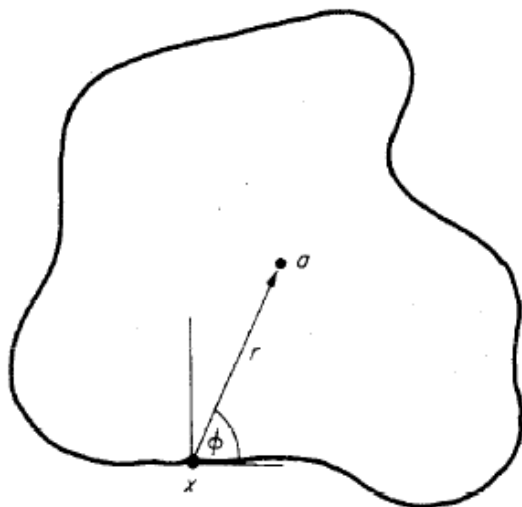
Η επέκταση του αλγορίθμου για τον κύκλο και την έλλειψη χρησιμοποιήθηκε για τον εντοπισμό κυκλικών πινακίδων οδικής κυκλοφορίας, όπως φαίνεται και στις ακόλουθες δημοσιεύσεις.

Στο [50] αφού έχει εφαρμοστεί ο κλασικός Μετασχηματισμός Hough για ευθείες, εφαρμόζεται σε κάθε κλειστή ακμή και ο μετασχηματισμός για εντοπισμό κύκλου.

Στα [51] και [52] οι συγγραφείς λαμβάνουν υπόψη τις προοπτικές παραμορφώσεις των κυκλικών σημάτων, τα οποία τελικά εμφανίζονται στην εικόνα ως ελλείψεις. Για τον λόγο αυτό, εφαρμόζουν τον Μετασχηματισμό Hough για εντοπισμό ελλείψεων.

Οι εντοπισθείσες ελλείψεις, θεωρούνται πλέον κύκλοι, με το κέντρο τους να ταυτίζεται με το κέντρο της έλλειψης και ακτίνα τους ίση με τον μεγάλο ημιάξονα της έλλειψης.

Τελικά, ο Μετασχηματισμός Hough επεκτάθηκε πλήρως, ώστε να εντοπίζει και σχήματα τα οποία δεν παραμετροποιούνται μαθηματικά. Ο νέος αλγόριθμος ονομάστηκε Γενικευμένος Μετασχηματισμός Hough (Generalised Hough Transform), όπως υπονοείται και στο [53] και εμφανίζει ομοιότητες με την τεχνική συνταύτισης προτύπων (template matching – βλ παρακάτω). Ζητούμενο του αλγορίθμου είναι ο υπολογισμός των τεσσάρων παραμέτρων (μετάθεση κατά τις συνιστώσες x και y , στροφή και κλίμακα) που περιγράφουν κατά τον βέλτιστο τρόπο τον διδιάστατο μετασχηματισμό ομοιότητας που έχει υποστεί το προς εντοπισμό αντικείμενο στη λήψη. Για να λειτουργήσει ο αλγόριθμος απαιτείται η *a priori* ύπαρξη μίας δυαδικής εικόνας με την ακμή του αντικειμένου που ζητείται να εντοπιστεί. Ορίζεται σε αυτό ένα σημείο αναφοράς \mathbf{x}_0 με συντεταγμένες x_c, y_c (ως σημείο αναφοράς χρησιμοποιείται συχνά το κέντρο βάρους του σχήματος). Για κάθε εικονοσηφίδα \mathbf{x} της ακμής ορίζεται η κλίση της ϕ (gradient) και η απόσταση r από το σημείο αναφοράς σε συνάρτηση με την κλίση (Εικόνα 2.18). Τα δεδομένα καταγράφονται σε ένα πίνακα με ονομασία \mathbf{R} (\mathbf{R} -table).



Εικόνα 2.18: Η απόσταση από το κέντρο βάρους ως συνάρτηση της κλίσης

(Πηγή: [53])

Στη συνέχεια, όπως και στις προαναφερθείσες περιπτώσεις Μετασχηματισμών Hough, δημιουργείται ένας πίνακας ψηφοφορίας με διαστήματα για τις τέσσερις παραμέτρους, που προαναφέρθηκαν. Στην προς εξέταση εικόνα εφαρμόζονται οι κατάλληλοι τελεστές ώστε να εντοπιστούν οι ακμές και για όλες τις εικονοσηφίδες κάθε ακμής υπολογίζεται η κλίση ϕ . Με βάση τον πίνακα \mathbf{R} προκύπτει η αντίστοιχη απόσταση που πρέπει να προστεθεί στο σημείο της ακμής για να προκύψει το σημείο αναφοράς \mathbf{x}_0 . Το αντίστοιχο διάστημα των συντεταγμένων του σημείου αναφοράς

στον πίνακα ψηφοφορίας, στο οποίο ανήκει το \mathbf{x}_0 που υπολογίστηκε, προσανξάνεται κατά μία ψήφο. Στην συγκεκριμένη περίπτωση που αναλύεται παραπάνω, ο μετασχηματισμός εκφυλίζεται σε μία απλή μετάθεση. Αντίστοιχα όμως, μπορούν να ληφθούν υπόψη οι επιπτώσεις από την αλλαγή κλίμακας s (ενιαίας και στις δύο διευθύνσεις) και στροφής θ . Οι δύο παράμετροι χωρίζονται σε συγκεκριμένα διαστήματα και κάθε εικονοψηφίδα προσανξάνει με την ψήφο της και αυτές τις δύο παραμέτρους. Το μέγιστο του πίνακα ψηφοφορίας στον χώρο των παραμέτρων αποτελεί τον βέλτιστο συνδυασμό των μεταβλητών του μετασχηματισμού ομοιότητας. Αν και ο Γενικευμένος Μετασχηματισμός Hough οδηγεί σε ικανοποιητικά αποτελέσματα όπως φαίνεται και στα [13], [55], [56] και παρουσιάζει ανοχή στις σχετικά μικρές, ανεκτές αποκρύψεις, δε χρησιμοποιείται ευρέως λόγω των υψηλών απαιτήσεων σε υπολογιστική ισχύ, που καθιστούν σχεδόν απαγορευτική τη χρήση του σε εφαρμογές πραγματικού χρόνου.

Στο [13] οι συγγραφείς χρησιμοποιούν μία παραλλαγή του Γενικευμένου Μετασχηματισμού Hough, όπως αυτή εισάγεται στο [54] για τον εντοπισμό σχημάτων με ίσες γωνίες. Λαμβάνουν λοιπόν υπόψη τις γεωμετρικές ιδιότητες των πινακίδων οδικής κυκλοφορίας, που αποτελούνται από σχήματα με ίσες γωνίες, για την εφαρμογή της συγκεκριμένης παραλλαγής. Οι κύκλοι θεωρούνται ως σχήματα με ίσες γωνίες μεταξύ των άπειρων πλευρών τους. Ο αλγόριθμος εφαρμόζεται σε εικόνες ακμών που προκύπτουν με χρήση κατάλληλων τελεστών στα καρτέ του βίντεο.

Στο [55] εφαρμόζεται η προαναφερθείσα παραλλαγή του αλγορίθμου και τα τελικά αποτελέσματα επεξεργάζονται περαιτέρω για να αυξηθεί η ακρίβειά τους καθώς ο Μετασχηματισμός Hough (και οι παραλλαγές του) είναι μία διαδικασία που οδηγεί σε πολλαπλές θετικές υποθέσεις, οι οποίες δεν είναι όλες αληθείς. Πρόκειται δηλαδή για ένα αλγόριθμο ανίχνευσης ιδιαίτερα ευαίσθητο. Για την βελτίωση της ακρίβειας των αποτελεσμάτων, οι ερευνητές θεωρούν τον πίνακα ψηφοφορίας σα μία συνάρτηση πιθανότητας και εφαρμόζουν στατιστικούς αλγορίθμους για τον εντοπισμό των μεγίστων και απομάκρυνση των λανθασμένων αποτελεσμάτων.

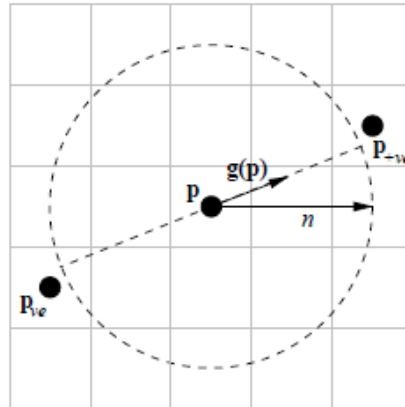
Τέλος, στο [56] αρχικά εφαρμόζεται ο κλασσικός Μετασχηματισμός Hough για εντοπισμό γραμμών και στα αποτελέσματα του τίθενται ορισμένοι κανόνες για τον εντοπισμό του σχήματος των πινακίδων. Για παράδειγμα, για τα τετράγωνα σήματα, τα ευθύγραμμα τμήματα που δημιουργούνται από τις τομές των ευθειών πρέπει να έχουν παρόμοιο μήκος. Στη συνέχεια, κάθε περιοχή ενδιαφέροντος τροφοδοτείται σε έναν Γενικευμένο Μετασχηματισμό Hough, ώστε να επαληθευτούν τα αποτελέσματα της προηγούμενης διαδικασίας.

2.1.2.3 Γρήγορος Ακτινικός Μετασχηματισμός Συμμετρίας (Fast Radial symmetry transform)

Παρόμοιος με τον Κυκλικό Μετασχηματισμό Hough είναι ο Γρήγορος Ακτινικός Μετασχηματισμός Συμμετρίας (*Fast Radial Symmetry Transform*), όπως αυτός αναλύεται στο ([57]). Η μέθοδος εντοπίζει αντικείμενα με έντονη συμμετρία μελετώντας τη συμβολή της κάθε εικονοψηφίδας σε μία ορισμένη απόσταση n .

Αρχικά, δημιουργούνται δύο εικόνες, η εικόνα του προσανατολισμού (orientation projection Image O_n) και η εικόνα των κλίσεων (gradient projection Image M_n). Κάθε εικονοψηφίδα ψηφίζει θετικά και αρνητικά σε ορισμένη απόσταση, κατά μήκος του διανύσματος της κλίσης της ($g(p)$). Οι εικονοψηφίδες που λαμβάνουν τελικά τις ψήφους είναι οι ακόλουθες, όπως φαίνεται στην Εικόνα 2.19:

$$p_{\pm ve} = p \pm \text{round} \left(\frac{g(p)}{\|g(p)\|} n \right) \quad (2.23)$$



Εικόνα 2.19: Θετικά και αρνητικά επηρεαζόμενες εικονοψηφίδες σε απόσταση n

(Πηγή:[57])

Για κάθε θετική ψήφο οι εικόνες O_n και M_n προσ αυξάνονται κατά μία τιμή και για κάθε αρνητική μειώνονται, όπως υπαγορεύουν κι ακόλουθες σχέσεις:

$$O_n(p \pm ve(p)) = O_n(p) \pm 1 \quad (2.24)$$

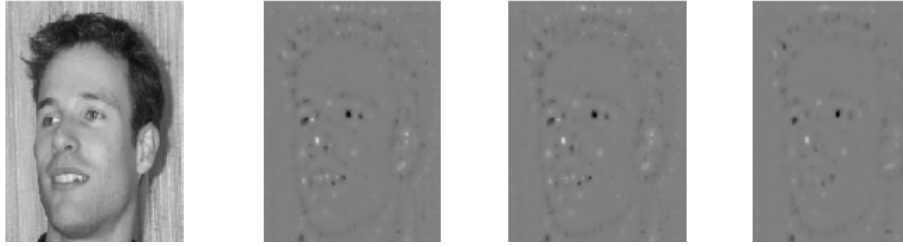
$$M_n(p \pm ve(p)) = M_n(p) \pm \|g(p)\| \quad (2.25)$$

Όπου $g(p)$ η τιμή της κλίσης της εξεταζόμενης εικονοψηφίδας.

Στη συνέχεια πραγματοποιούνται πράξεις μεταξύ των δύο εικόνων και προκύπτει η τελική εικόνα για κάθε απόσταση n . Ο αλγόριθμος ολοκληρώνεται με εφαρμογή της μεθόδου για διάφορες αποστάσεις n και συνένωση των αποτελεσμάτων. Στο κλασικό αλγόριθμο οι τιμές που λαμβάνει η μεταβλητή $n = 1, 2, 3, \dots, n$. Έρευνες όμως έδειξαν ότι τα αποτελέσματα είναι εξίσου ικανοποιητικά με χρήση λιγότερων αποστάσεων, ενώ ταυτόχρονα, με τον τρόπο αυτό, αυξάνεται η ταχύτητα εκτέλεσης του αλγορίθμου. Έτσι, στον Γρήγορο Μετασχηματισμό Ακτινικής Συμμετρίας χρησιμοποιούνται οι αποστάσεις $n = 1, 3, 5 \dots$. Τα αποτελέσματα εφαρμογής της τεχνικής φαίνονται στην Εικόνα 2.20.

Η συγκεκριμένη τεχνική είναι λιγότερο υπολογιστική απαιτητική σε σχέση με τον Μετασχηματισμό Hough, συνεπώς ενδείκνυται περισσότερο για εφαρμογές πραγματικού χρόνου. Χαρακτηριστικά παραδείγματα δημοσιεύσεων για τον εντοπισμό

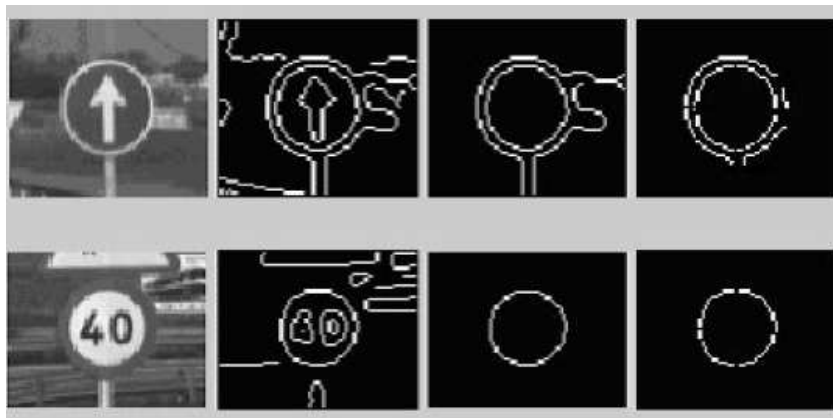
σημάτων οδικής κυκλοφορίας με τον συγκεκριμένο αλγόριθμο είναι οι [58], [59] και [60].



Εικόνα 2.20: (a) Αρχική εικόνα, (b) Εφαρμογή για $n = 1,2,3,4,5$, (c) Εφαρμογή για $n = 1,3,5$ και (d) Εφαρμογή για $n = 5$

(Πηγή:[57])

Στο [58] για κάθε εικόνα εξάγονται οι ακμές με χρήση του τελεστή Canny Edge Detector. Κάθε περιοχή ακμής συνιστά μία ξεχωριστή περιοχή ενδιαφέροντος, η οποία επεξεργάζεται περαιτέρω για μείωση του θορύβου. Στη συνέχεια, εφαρμόζεται ο Γρήγορος Ακτινικός Μετασχηματισμός Συμμετρίας για τον εντοπισμό των κυκλικών πινακίδων. Εάν ο μετασχηματισμός επιστρέψει αποτέλεσμα, η περιοχή ενδιαφέροντος τροφοδοτεί το στάδιο της αναγνώρισης, διαφορετικά απορρίπτεται. Ο Γρήγορος Ακτινικός Μετασχηματισμός Συμμετρίας μπορεί να βασιστεί είτε στον εξωτερικό είτε στον εσωτερικό κυκλικό δακτύλιο του σχήματος (εάν για παράδειγμα ο εξωτερικός δεν εμφανίζεται στην εικόνα), όπως φαίνεται και στο ακόλουθο σχήμα.



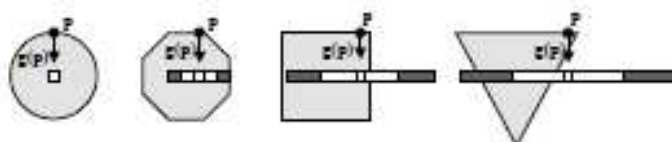
Εικόνα 2.21: (a) αρχική εικόνα, (b) αποτέλεσμα εφαρμογής CannyEdgeDetector, (c) αφαίρεση θορύβου, (d) αποτέλεσμα μετασχηματισμού

(Πηγή:[58])

Στο [59] οι ερευνητές μελετούν τον Γρήγορο Ακτινικό Μετασχηματισμό Συμμετρίας για τον εντοπισμό κυκλικών πινακίδων, με τιμές του n μεταξύ 5 και 9 εικονοψηφίδων. Η απόσταση n συνιστά ουσιαστικά την ακτίνα του κύκλου που πρόκειται να εντοπιστεί. Οι συγγραφείς μετά από πειράματα κατέληξαν ότι αυτές οι τιμές είναι οι πιο συνήθεις για την ακτίνα κυκλικών πινακίδων, στις διάφορες πιθανές κλίμακες. Κύκλοι με μικρότερες ακτίνες μπορεί να αποτελούν σήματα οδικής κυκλοφορίας

αλλά η απεικόνισή τους στη δεδομένη κλίμακα, δε θα επιτρέψει στη συνέχεια την αναγνώρισή τους. Για κάθε κύκλο που εντοπίζεται τίθενται ορισμένες επιπλέον δεσμεύσεις ώστε να απομακρυνθούν τα εσφαλμένα αποτελέσματα. Για παράδειγμα, απαιτείται η ακμή του να συγκεντρώνει αριθμό εικονοψηφίδων μεγαλύτερο από ένα κατώφλι και ο κύκλος να εντοπίζεται σε περισσότερα του ενός διαδοχικά καρέ, με ακτίνα που δεν έχει μεταβληθεί πάνω από δύο εικονοψηφίδες.

Τέλος, στο [60] επεκτείνεται ο Γρήγορος Ακτινικός Μετασχηματισμός Συμμετρίας προκειμένου να εντοπιστούν και οι υπόλοιπες πινακίδες, με τα λοιπά σχήματα. Ορίζεται ως «ακτίνα» του κάθε πολυγώνου η κάθετη απόσταση από την ακμή του στο κέντρο βάρους και κάθε εικονοψηφίδα «ψηφίζει» θετικά ή αρνητικά όχι μόνο για μία εικονοψηφίδα, που απέχει ορισμένη απόσταση, αλλά για μία ολόκληρη γραμμή επάνω στην οποία βρίσκεται το κέντρο βάρους. Στο επόμενο σχήμα απεικονίζεται η περίπτωση του κύκλου με ψήφο για μία μόνο εικονοψηφίδα και οι περιπτώσεις εξαγώνου, τετραγώνου και τριγώνου όπου η κάθε εικονοψηφίδα ψηφίζει για μία ολόκληρη γραμμή (με λευκό χρώμα αναπαρίστανται οι θετικές ψηφοί και με γκρι οι αρνητικές).



Εικόνα 2.22: Θετικές και αρνητικές ψηφοί της εικονοψηφίδας ρ για τα διάφορα σχήματα

(Πηγή:[60])

2.1.2.4 Συνταύτιση προτύπων

Άλλη μία συνήθης τεχνική για τον εντοπισμό αντικειμένων (object detection) είναι η συνταύτιση προτύπων (template matching), η οποία χρησιμοποιείται σε αρκετές δημοσιεύσεις για τον εντοπισμό με βάση το σχήμα.

Σκοπός της μεθόδου είναι ο εντοπισμός στην λήψη, ενός προτύπου (template), το οποίο είναι γνωστό a priori, όπως αναφέρεται και στο [61]. Η συνταύτιση προτύπων λαμβάνει χώρα είτε ως επιφανειακή συνταύτιση (template based approach) ή ως συνταύτιση χαρακτηριστικών (feature based approach). Στην πρώτη περίπτωση συγκρίνονται οι τιμές φωτεινότητας κάθε εικονοψηφίδας του προτύπου με τις αντίστοιχες τιμές κάθε εικονοψηφίδας της περιοχής που εξετάζεται στην εικόνα. Ενώ, στη δεύτερη περίπτωση συγκρίνονται χαρακτηριστικά (π.χ περιγραφέας SIFT, SURF), που έχουν εξαχθεί τόσο από το πρότυπο όσο και από την περιοχή ενδιαφέροντος. Η σύγκριση πραγματοποιείται μέσω ενός μέτρου (άθροισμα διαφορών τετραγώνων, συντελεστή συσχέτισης κ.α) το οποίο ελαχιστοποιείται ή μεγιστοποιείται ανάλογα, όταν εντοπιστεί το πρότυπο.

Για την επιτάχυνση της διαδικασίας συχνά χρησιμοποιούνται πυραμίδες εικόνων ή εικόνες ακμών, ενώ κατά καιρούς έχουν προταθεί διάφοροι αλγόριθμοι για να είναι η

συνταύτιση κατά το δυνατόν ανεκτική σε στροφές, αλλαγές κλίμακας, αλλαγές φωτεινότητας και αντίθεσης.

Παράδειγμα δημοσιεύσεων που εφαρμόζουν τη συνταύτιση πρότυπων για τον εντοπισμό σημάτων είναι τα [62], [63] και [9].

Στο [62] εντοπίζονται περιοχές ενδιαφέροντος και για κάθε μία εξάγεται η κόκκινη συνιστώσα στον χρωματικό χώρο RGB. Στη συνέχεια δύο πρότυπα (Εικόνα 2.23), ένα για τα τριγωνικά και ένα για τα κυκλικά σήματα μετακινούνται μέσα στην περιοχή ενδιαφέροντος και σε κάθε θέση υπολογίζεται ο αριθμός των κόκκινων εικονοψηφίδων που εντοπίζονται εντός του δακτυλίου. Τα πρότυπα που εφαρμόζονται έχουν διάφορα μεγέθη ώστε ο αλγόριθμος να μην εξαρτάται από την κλίμακα της λήψης. Εάν ο αριθμός των κόκκινων εικονοψηφίδων υπερβαίνει ένα προκαθορισμένο κατώφλι σε μία συγκεκριμένη θέση, επιλέγεται ως τελική περιοχή ενδιαφέροντος η περιοχή με κέντρο το κέντρο του προτύπου στη θέση αυτή και διαστάσεις ίσες με τις διαστάσεις του προτύπου.



Εικόνα 2.23: Τα διαφορετικά πρότυπα

(Πηγή:[62])

Αντίστοιχη τακτική ακολουθείται και στο [63] για εντοπισμό κυκλικών πινακίδων και απομάκρυνση των εσφαλμένων περιοχών ενδιαφέροντος που προέκυψαν από την επεξεργασία με βάση το χρώμα. Χρησιμοποιούνται συνολικά έξι πρότυπα διαφορετικών μεγεθών, το πιο μικρό με διαστάσεις 32x32 εικονοψηφίδες και το πιο μεγάλο με διαστάσεις 78x78 εικονοψηφίδες. Αρχικά ορίζεται μία μεταβλητή P ίση με μηδέν και για κάθε εικονοψηφίδα της περιοχής ενδιαφέροντος μελετάται εάν υπάρχουν κόκκινες εικονοψηφίδες στο κέντρο της πινακίδας, εάν οι κόκκινες εικονοψηφίδες δεν είναι συμμετρικά κατανομημένες, εάν ο αριθμός των λευκών εικονοψηφίδων είναι πολύ μικρός ή πολύ μεγάλος και αν υπάρχουν μαύρες εικονοψηφίδες στο κέντρο. Σε όλες τις παραπάνω περιπτώσεις η μεταβλητή P παραμένει σταθερή, διαφορετικά αυξάνεται κατά 1. Τελικά, εάν η μεταβλητή εμφανίζει τιμή μεγαλύτερη από ένα όριο, ο αλγόριθμος έχει εντοπίσει κυκλική πινακίδα στην περιοχή ενδιαφέροντος, στη θέση που η μεταβλητή έλαβε την υψηλότερη τιμή. Η διαδικασία επαναλαμβάνεται για όλα τα μεγέθη προτύπων και επιλέγεται το κατάλληλο με βάση την κλίμακα της λήψης.



Εικόνα 2.24: Διαφορετικά μεγέθη του ίδιου προτύπου

(Πηγή:[63])

Στο [9] χρησιμοποιούνται ως πρότυπα οι ακόλουθες πέντε δυαδικές εικόνες. Σε κάθε περιοχή ενδιαφέροντος πραγματοποιείται επανασύσταση της, ώστε να αποκτήσει διαστάσεις 50x50 εικονοψηφίδες για να ταυτίζεται με τις διαστάσεις του προτύπου και ο αλγόριθμος να είναι ανεκτικός σε αλλαγές κλίμακας, χωρίς να χρειάζεται να χρησιμοποιηθούν πρότυπα διαφορετικά μεγεθών. Το μέτρο ομοιότητας για τη συνταύτιση είναι ο λόγος των σωστά συνταυτισμένων εικονοψηφίδων μεταξύ προτύπου και περιοχής ενδιαφέροντος ως προς τον αριθμό των λευκών εικονοψηφίδων σε κάθε δυαδική εικόνα. Η συνταύτιση θεωρείται επιτυχής εάν ο λόγος υπερβαίνει ένα προκαθορισμένο όριο. Η περιοχή ενδιαφέροντος ταυτίζεται με το πρότυπο, το οποίο εμφανίζει τον μέγιστο λόγο κατά τη διαδικασία συνταύτισης.



Εικόνα 2.25: Οι δυαδικές εικόνες που χρησιμοποιούνται ως πρότυπα

(Πηγή:[9])

2.1.2.5 Αλγόριθμος Ransac

Άλλη μία τεχνική που έχει εφαρμοστεί για τον εντοπισμό με βάση το σχήμα είναι ο συνδυασμός του αλγορίθμου RANSAC (Random Sample Consensus), όπως αυτός περιγράφεται στο [64] με μία εικόνα ακμών.

Ο RANSAC είναι αλγόριθμος που υπολογίζει τις βέλτιστες τιμές για ένα πρόβλημα, από ένα σετ δεδομένων, που περιέχει και χονδροειδή σφάλματα, τα οποία ο αλγόριθμος κατά το δυνατόν απομακρύνει. Σε αντίθεση με τους περισσότερους αλγορίθμους τέτοιου είδους, ο RANSAC λειτουργεί με τα λιγότερα δεδομένα που απαιτούνται για τη λύση του εκάστοτε προβλήματος.

Κατ' αρχάς λοιπόν δημιουργούνται ορισμένα δείγματα. Το καθένα αποτελείται από τον μικρότερο δυνατό αριθμό δεδομένων. Από κάθε δείγμα υπολογίζεται μία λύση. Για την κάθε λύση ελέγχεται ο αριθμός των δεδομένων που είναι συμβατά με αυτή εντός ενός προκαθορισμένου ορίου (inliers). Τα υπόλοιπα δεδομένα θεωρούνται χονδροειδή σφάλματα (outliers).

Η παραπάνω διαδικασία επαναλαμβάνεται για συγκεκριμένο αριθμό δειγμάτων, ώστε να εξασφαλιστεί η επιδιωκόμενη ακρίβεια. Είναι προφανές ότι θα ήταν υπολογιστικά αδύνατο να πραγματοποιηθεί η διαδικασία για το σύνολο των δυνατών δειγμάτων. Όσο μικρότερα είναι τα δείγματα που χρησιμοποιούνται, τόσο μικρότερος είναι και ο αριθμός των επαναλήψεων που απαιτείται να δοκιμαστούν για να τερματιστεί επιτυχώς η διαδικασία. Ο αριθμός N των επαναλήψεων, που απαιτούνται ώστε κάποιο από τα δείγματα να περιλαμβάνει μόνο inliers με πιθανότητα p , δίνεται από την εξής σχέση:

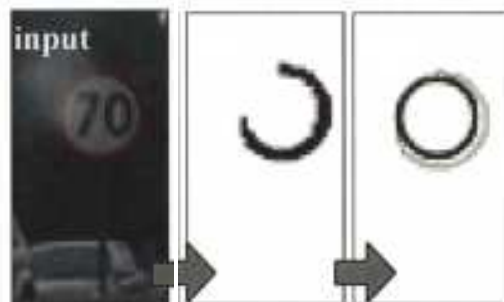
$$N = \frac{\log(1 - p)}{\log(1 - (1 - \varepsilon)^s)} \quad (2.26)$$

όπου s ο αριθμός των ελάχιστων σημείων που απαιτούνται σε κάθε δείγμα και ε η εκτίμηση του ποσοστού των χονδροειδών σφαλμάτων.

Η λύση με τα περισσότερα inliers είναι η επικρατέστερη και τελικά εφαρμόζεται ένας αλγόριθμος για τον υπολογισμό των βέλτιστων παραμέτρων του προβλήματος με δεδομένα εισόδου μόνο τα inliers

Χαρακτηριστικά παραδείγματα εφαρμογών του RANSAC σε συνδυασμό με εικόνες ακμών για εντοπισμό πινακίδων οδικής κυκλοφορίας με βάση το σχήμα είναι οι [42] και [22].

Στο [42] επεξεργάζονται κατάλληλα τα αποτελέσματα της κατωφλίωσης ώστε να προκύψουν διάφορες ακμές. Κάθε ακμή τροφοδοτείται στον RANSAC, ώστε να υπολογιστεί ο κύκλος που την περιγράφει με τον καλύτερο δυνατό τρόπο. Σύμφωνα με τα παραπάνω και δεδομένου ότι ένας κύκλος περιγράφεται πλήρως από τρία σημεία, ο αλγόριθμος δημιουργεί διάφορες υπο-ομάδες αποτελούμενες από τρία σημεία, προκειμένου να εντοπίσει τη βέλτιστη λύση, δηλαδή τον κύκλο που περιγράφει καλύτερα την εντοπισθείσα ακμή. Σημαντικό πλεονέκτημα συνιστά το ότι η προσέγγιση είναι ανθεκτική σε αποκρύψεις και επιστρέφει ένα πλήρη κύκλο ακόμα κι αν η ακμή που χρησιμοποιήθηκε δεν ήταν κλειστή, όπως φαίνεται και στο ακόλουθο σχήμα.



Εικόνα 2.26: (a) Αρχική εικόνα, (b) Αποτέλεσμα κατωφλίωσης, (c) Βέλτιστος κύκλος μετά από εφαρμογή του RANSAC

(Πηγή:[42])

Παρόμοια μέθοδος ακολουθείται και στο [22] με τη διαφορά ότι οι ερευνητές λαμβάνουν υπόψη και τις προοπτικές παραμορφώσεις της λήψης, που μετατρέπουν τις κυκλικές πινακίδες εμφανίζονται ως ελλείψεις. Από εικόνα ακμών που έχει προκύψει από την εφαρμογή του τελεστή Canny Edge Detector, λαμβάνουν τις συντεταγμένες των εικονοψηφίδων που απαρτίζουν την ακμή. Στη συνέχεια, δημιουργούν διάφορα δείγματα δεδομένων αποτελούμενα από τρία σημεία. Αν και η έλλειψη χρειάζεται πέντε σημεία για τον πλήρη καθορισμό της, οι ερευνητές εκμεταλλεύονται εδώ το θεώρημα του Pascal, σύμφωνα με το οποίο μία έλλειψη μπορεί να οριστεί πλήρως από τρία μόνο σημεία. Η κάθε υποομάδα δίνει μία λύση και ελέγχεται πόσα από τα σημεία που απαρτίζουν το σύνολο δεδομένων, απέχουν απόσταση από την έλλειψη που κείται εντός προκαθορισμένων ορίων. Τα σημεία αυτά αποτελούν τα *inliers*. Τελικά, επιλέγεται η λύση με τα περισσότερα συμβατά δεδομένα και βάσει αυτών υπολογίζεται η τελική έλλειψη. Εάν ολοκληρωθεί ο αριθμός των ελέγχων χωρίς να εντοπιστεί λύση με τον επιθυμητό αριθμό *inliers*, η περιοχή ενδιαφέροντος απορρίπτεται.

2.2 Αναγνώριση

Το δεύτερο βήμα για τη δημιουργία ενός συστήματος αυτόματου εντοπισμού και ταξινόμησης σημάτων οδικής κυκλοφορίας είναι η αναγνώρισή τους (*recognition*). Στο βήμα αυτό, κάθε περιοχή ενδιαφέροντος που έχει εντοπιστεί συγκρίνεται με δεδομένα που έχουν ήδη συλλέγει και ταξινομείται σε μία από τις πιθανές κατηγορίες. Αν η ταξινόμηση (*classification*) δεν καταστεί δυνατή, η περιοχή ενδιαφέροντος θεωρείται εσφαλμένη και απομακρύνεται ως θόρυβος.

Οι αλγόριθμοι ταξινόμησης προσεγγίζουν το πρόβλημα είτε με τη χρήση προτύπων (*template-based approach*) είτε με τη χρήση ταξινομητών (*classifier-based approach*) [65]. Στη δεύτερη περίπτωση, οι αλγόριθμοι προέρχονται από το πεδίο της μηχανικής μάθησης (*Machine Learning*). Όπως όρισε ο Samuel (1959), «το πεδίο της μηχανικής μάθησης είναι το πεδίο μελέτης που δίνει τη δυνατότητα στον υπολογιστή να μάθει, χωρίς να έχει προγραμματιστεί ρητά για αυτό» [66]. Σκοπός των αλγορίθμων που εντάσσονται στο συγκεκριμένο πεδίο είναι η δημιουργία ενός συστήματος που «μαθαίνει» από τα δεδομένα, ώστε στη συνέχεια να είναι ικανό να κατηγοριοποιήσει νέα δεδομένα [66]. Ο πιο διαδεδομένος ορισμός που έχει επικρατήσει για τη μηχανική μάθηση προέρχεται από τον Mitchell [67], σύμφωνα με τον οποίο: «Ένα πρόγραμμα υπολογιστή λέγεται ότι μαθαίνει από την εμπειρία *E* σε σχέση με μία κατηγορία διαδικασιών *T* και μέτρο απόδοσης *P*, εάν η απόδοση του στις κατηγορίες *T*, όπως μετρείται με το μέτρο *P*, βελτιώθηκε κατά εμπειρία *E*». Έτσι, στο συγκεκριμένο πρόβλημα εφαρμόζονται αλγόριθμοι που, με τη βοήθεια εικόνων σημάτων οδικής κυκλοφορίας, εκπαιδεύουν το σύστημα ώστε να κατηγοριοποιεί τα νέα δεδομένα. Κάθε ανάλογο σύστημα πρέπει να είναι σε θέση να «μάθει» ένα μεγάλο αριθμό κατηγοριών και να αποκτήσει όσο το δυνατόν περισσότερη *a priori* γνώση, ώστε τα αποτελέσματα της κατηγοριοποίησης να είναι τα βέλτιστα.

Οι τεχνικές που αναπτύσσονται για την ταξινόμηση των πινακίδων στις διάφορες ομάδες, είτε προέρχονται από το πεδίο της μηχανικής μάθησης είτε από αλλού,

πρέπει να έχουν τις λιγότερες δυνατές υπολογιστικές απαιτήσεις και να είναι ανθεκτικές στις διαφορές κλίμακας, στροφής και μετάθεσης αλλά και στον θόρυβο. Στη συνέχεια αναλύονται οι δημοφιλέστερες τεχνικές για την κατηγοριοποίηση των πινακίδων, όπως η συνταύτιση προτύπων, η χρήση νευρωνικών δικτύων και η χρήση μηχανών διανυσματικής στήριξης.

2.2.1 Συνταύτιση προτύπων (Template matching)

Η τεχνική συνταύτισης προτύπων, όπως αυτή αναλύθηκε προηγουμένως, στο υποκεφάλαιο 2.1, μπορεί να χρησιμοποιηθεί και στο βήμα της ταξινόμησης. Απαραίτητη προϋπόθεση είναι η ύπαρξη προτύπων, δηλαδή εικόνων με τα σήματα οδικής κυκλοφορίας. Κάθε περιοχή ενδιαφέροντος που έχει προκύψει από το στάδιο του εντοπισμού συγκρίνεται με τα διάφορα πρότυπα. Τελικά, η πινακίδα που απεικονίζεται στην περιοχή ενδιαφέροντος κατατάσσεται στην αντίστοιχη ομάδα του προτύπου με το οποίο εμφανίζει τη μέγιστη ομοιότητα (ή την ελάχιστη διαφορά). Χαρακτηριστικά παραδείγματα ερευνητικών εργασιών που χρησιμοποιούν τη μέθοδο αυτή είναι οι [39] και [69].

Στο [39] σε κάθε περιοχή ενδιαφέροντος πραγματοποιείται επανασύσταση εικόνας, ώστε να αποκτήσει διαστάσεις 64x64, ίδιες δηλαδή διαστάσεις με τα πρότυπα που πρόκειται να συγκριθεί. Ανάλογα με το είδος του χρώματος και του σχήματος της πινακίδας που έχει εντοπιστεί, χρησιμοποιούνται και τα αντίστοιχα πρότυπα, ώστε να αποφευχθούν επιπλέον συγκρίσεις που δε χρειάζονται και επιβαρύνουν υπολογιστικά και χρονικά τη διαδικασία αναγνώρισης. Μέτρο ομοιότητας της σύγκρισης είναι ο Κανονικοποιημένος Συντελεστής Συσχέτισης (Normalised Cross Correlation Coefficient). Το συγκεκριμένο μέτρο παρουσιάζει ανοχή στις γραμμικές διαφορές φωτεινότητας και αντίθεσης μεταξύ εικόνας και προτύπου, και ως γνωστόν μαθηματικά ισούται με:

$$p = \frac{1}{n} \sum_{x,y} \frac{(f(x,y) - \bar{f})(t(x,y) - \bar{t})}{\sigma_f \sigma_t} \quad (2.27)$$

όπου n ο αριθμός των εικονοψηφίδων στην εικόνα $f(x,y)$ και το πρότυπο $t(x,y)$, \bar{f} και \bar{t} η μέση τιμή φωτεινότητας για την εικόνα και το πρότυπο, αντίστοιχα, και σ_f , σ_t τυπική απόκλιση για την εικόνα και το πρότυπο.

Παρόμοια τακτική ακολουθείται και στο [69]. Εάν η μέγιστη τιμή του Κανονικοποιημένου Συντελεστή Συσχέτισης δεν ξεπερνά ένα προκαθορισμένο όριο, η περιοχή ενδιαφέροντος που μελετάται θεωρείται θόρυβος.

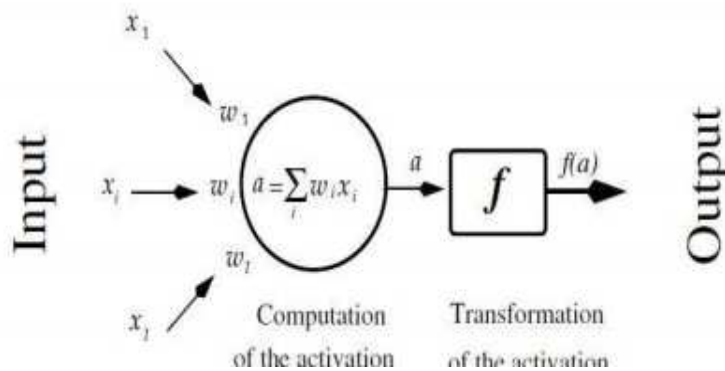
2.2.2 Νευρωνικά Δίκτυα (Neural Networks)

Ο όρος Νευρωνικό Δίκτυο (Neural Network) αναφέρεται σε ένα σύνολο κόμβων που συνεργάζονται προκειμένου να επιτελέσουν έναν σκοπό. Τα Τεχνητά Νευρωνικά Δίκτυα (ΤΝΔ) στην υπολογιστική νοημοσύνη προσομοιώνουν κατά το δυνατόν το ανθρώπινο νευρικό σύστημα και στοχεύουν στην επίλυση κάποιου υπολογιστικού προβλήματος. Η επεξεργασία της εισερχόμενης στο σύστημα πληροφορίας πραγμα-

τοποιείται μέσω ενός συστήματος πολλών, διασυνδεδεμένων νευρώνων, όπως ακριβώς συμβαίνει και στον ανθρώπινο εγκέφαλο. Ο τρόπος με τον οποίο οι διάφοροι νευρώνες διασυνδέονται μεταξύ τους ονομάζεται αρχιτεκτονική του ΤΝΔ. Τα περισσότερα ΤΝΔ οργανώνονται σε στρώματα (layers). Απαραίτητα είναι τα στρώματα εισόδου (input layer) και εξόδου (output layer), ενώ είναι πιθανόν να υπάρχουν και ενδιάμεσα, «αποκρυπτόμενα» στρώματα (hidden layer), οπότε το νευρωνικό δίκτυο ονομάζεται πολύστρωματικό (multilayer neural network). Κάθε στρώμα απαρτίζεται από ένα σύνολο κόμβων (νευρώνων). Οι νευρώνες μεταξύ τους ανταλλάσσουν πληροφορίες και οι σχέσεις μεταξύ των νευρώνων χαρακτηρίζονται από ένα βάρος. Αυτά τα βάρη πολλαπλασιάζονται με τις εισερχόμενες πληροφορίες και αθροίζονται με μία αθροιστική συνάρτηση. Το άθροισμα εισάγεται σε μία συνάρτηση ενεργοποίησης, που υλοποιείται εσωτερικά σε κάθε κόμβο, και το αποτέλεσμα της είναι τα δεδομένα εξόδου για τον συγκεκριμένο νευρώνα, τα οποία αποτελούν επίσης τα δεδομένα εισόδου για τον επόμενο νευρώνα. Η έξοδος κάθε νευρώνα k υπολογίζεται ως εξής [70]:

$$y_k = \varphi \left(\sum_{i=0}^N (x_{ki} * w_{ki}) \right) \quad (2.28)$$

όπου x_{ki} η i είσοδος του νευρώνα k , w_{ki} το αντίστοιχο βάρος και $\varphi(\cdot)$ η συνάρτηση ενεργοποίησης (Εικόνα 2.27). Οι νευρώνες κάθε στρώματος υποχρεούνται να χρησιμοποιούν την ίδια συνάρτηση ενεργοποίησης.

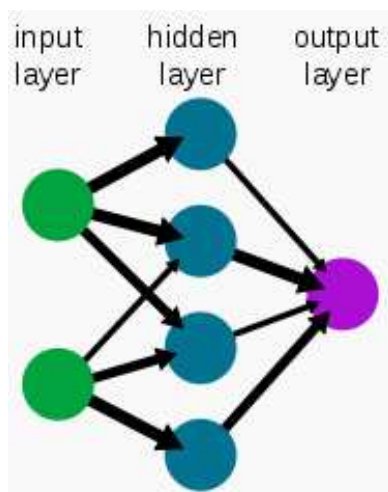


Εικόνα 2.27: Τεχνητός νευρώνας

(Πηγή:[72])

Η είσοδος στο νευρωνικό δίκτυο γίνεται μέσω του πρώτου στρώματος (“input layer”), το οποίο θα πολλαπλασιάσει την είσοδο με τα βάρη και θα περάσει τα τροποποιημένα πλέον δεδομένα στο επόμενο στρώμα. Αυτή η διαδικασία, κατά την οποία κάθε στρώμα τροποποιεί την είσοδό του και την “περνά” στο επόμενο, συνεχίζεται μέχρι το τελευταίο στρώμα (“output layer”), από το οποίο προκύπτουν τα δεδομένα εξόδου. Τα τεχνητά νευρωνικά δίκτυα διακρίνονται, εκτός από τον αριθμό των στρωμάτων τους, βάσει του τρόπου διάδοσης της πληροφορίας στους νευρώνες. Αν οι πληροφορίες διαδίδονται με τρόπο ώστε να μην υπάρχει νευρώνας που η έξοδος του να

είναι είσοδος κάποιου άλλου νευρώνα, του ίδιου ή του προηγούμενου στρώματος, τότε το δίκτυο ονομάζεται πρόσθιας τροφοδότησης (feed forward) ([71]). Στην περίπτωση αυτή δεν υπάρχουν βρόχοι ανατροφοδότησης και η πληροφορία μεταφέρεται μόνο προς μία κατεύθυνση. Επίσης, αν κάθε νευρώνας ενός στρώματος συνδέεται με όλους του νευρώνες του επόμενου στρώματος, το δίκτυο ονομάζεται πλήρως διασυνδεδεμένο (Εικόνα 2.28). Τέλος, εάν κάποιος νευρώνας συνδέεται με νευρώνες του ίδιου ή του προηγούμενου στρώματος, το δίκτυο παρουσιάζει ανάδραση (feedback).



Εικόνα 2.28: Ένα πλήρως διασυνδεδεμένο πολυστρωματικό δίκτυο

(Πηγή: http://el.wikipedia.org/wiki/%CE%9D%CE%B5%CF%85%CF%81%CF%89%CE%BD%CE%B9%CE%BA%CF%8C_%CE%B4%CE%AF%CE%BA%CF%84%CF%85%CE%BF)

Το κύριο χαρακτηριστικό των νευρωνικών δικτύων είναι η εγγενής ικανότητα μάθησης. Ως μάθηση μπορεί να οριστεί η σταδιακή βελτίωση της ικανότητας του δικτύου να διαχειρίζεται τα δεδομένα εισόδου. Έτσι, τα νευρωνικά δίκτυα είναι ικανά να «μάθουν» περίπλοκες, μη γραμμικές σχέσεις που συνδέουν τα δεδομένα εισόδου με τα δεδομένα εξόδου, μέσω της διαδικασίας της εκπαίδευσης, μίας επαναληπτικής διαδικασίας σταδιακής προσαρμογής των παραμέτρων του δικτύου (των βαρών του) σε τιμές κατάλληλες, ώστε να δίνει συγκεκριμένες απαντήσεις για συγκεκριμένες εισόδους ([71]). Αφού ένα δίκτυο εκπαιδευτεί, οι παράμετροί του συνήθως «παγώνουν» στις κατάλληλες τιμές και από εκεί κι έπειτα βρίσκονται σε κατάσταση λειτουργίας. Το ζητούμενο είναι το λειτουργικό δίκτυο να χαρακτηρίζεται από ικανότητα γενίκευσης, δηλαδή να καταλήγει σε ορθές εξόδους για δεδομένα εισόδου διαφορετικά από εκείνα με τα οποία εκπαιδεύτηκε. Η εκπαίδευση μπορεί να είναι επιβλεπόμενη (supervised learning), για παράδειγμα όταν κάποιος εξωτερικός εκπαιδευτής αποφασίζει τα δεδομένα εκπαίδευσης και το πότε η διαδικασία πρέπει να τερματιστεί. Τα δεδομένα εκπαίδευσης αποτελούνται από ζευγάρια δεδομένων εισόδου με την αντίστοιχη επιθυμητή έξοδο. Τα συγκεκριμένα δεδομένα εισάγονται στο σύστημα ως δεδομένα εισόδου και υπολογίζεται η έξοδος, η οποία συγκρίνεται με την επιθυμητή. Η αποχή μεταξύ τους χαρακτηρίζεται ως σφάλμα εκπαίδευσης και τα βάρη μεταβάλλονται ανάλογα, ώστε το σφάλμα εξόδου να μειωθεί, δηλαδή η νέα έξοδος να πλησιάζει περισσότερο την επιθυμητή έξοδο. Η μη επιβλεπόμενη εκ-

παίδευση (unsupervised learning) πραγματοποιείται απουσία κάποιου εξωτερικού παράγοντα, με τη χρήση των κατάλληλων αλγορίθμων, και οι παράμετροι του δικτύου τροποποιούνται ώστε να ανταποκρίνονται μόνο στα δεδομένα εισόδου [71].

Ο πιο συνηθισμένος αλγόριθμος εκπαίδευσης για τα δίκτυα πρόσθιας τροφοδότησης είναι ο αλγόριθμος ανάστροφης μετάδοσης σφάλματος (back propagation). Όπως προαναφέρθηκε, ο σκοπός της εκπαίδευσης είναι ο υπολογισμός των βέλτιστων τιμών των βαρών w που χαρακτηρίζουν τις σχέσεις μεταξύ των νευρώνων. Οι διορθώσεις των τιμών ακολουθούν την επόμενη σχέση:

$$w_{i+1} = w_i - a \frac{\partial E}{\partial w_i} \quad (2.29)$$

όπου: a παράμετρος που ορίζει το βήμα κάθε επανάληψης και
 E η ακόλουθη συνάρτηση κόστους:

$$E = \frac{\sum_{i=1}^p \|o_i - t_i\|^2}{2} \quad (2.30)$$

όπου: o_i τα δεδομένα εξόδου του δικτύου και
 t_i τα επιθυμητά δεδομένα εξόδου, όπως ορίζονται από τα δεδομένα εκπαίδευσης.

Για την εφαρμογή του αλγορίθμου εισάγονται αρχικά τα δεδομένα στο δίκτυο και υπολογίζονται τα δεδομένα εξόδου κάθε στρώματος. Επίσης, υπολογίζονται οι παράγωγοι των συναρτήσεων ενεργοποίησης σε κάθε κόμβο. Στη συνέχεια υπολογίζεται η τιμή $\frac{\partial E}{\partial w_i}$ από το τέλος προς την αρχή του δικτύου. Σε πρώτη φάση υπολογίζονται οι τιμές για το στρώμα εξόδου και μετά για το κάθε ενδιάμεσο στρώμα διαδοχικά. Αφού έχουν υπολογιστεί οι τιμές για όλα τα στρώματα, οι τιμές των βαρών ανανεώνονται ταυτόχρονα. Η παραπάνω διαδικασία επαναλαμβάνεται έως ότου το σφάλμα του δικτύου μειωθεί ικανοποιητικά. Ο αριθμός των επαναλήψεων δεν είναι γνωστός από την αρχή και εξαρτάται από το πλήθος και το είδος των δεδομένων του προβλήματος. Εκτενέστερη αναφορά γίνεται στα [70], [71], [72], [73].

Τα νευρωνικά δίκτυα, σύμφωνα με τα [74] και [75], χρησιμοποιούνται ευρέως στην ανάλυση εικόνας για πληθώρα εργασιών, όπως κατάτμηση εικόνας (image segmentation), αναγνώριση αντικειμένων (object detection), κατανόηση εικόνας (image understanding).

Στη παρούσα εφαρμογή, τα νευρωνικά δίκτυα χρησιμοποιούνται για ταξινόμηση των περιοχών ενδιαφέροντος που έχουν προκύψει από το στάδιο του εντοπισμού. Για να συμβεί αυτό απαιτείται η επιβλεπόμενη ταξινόμηση του συστήματος, ώστε τελικά η κάθε περιοχή ενδιαφέροντος να ταξινομηθεί σε μία από τις κατηγορίες που έχουν εκ των προτέρων οριστεί. Κάθε εικόνα που αποτελεί δεδομένο εισόδου εισάγεται στο σύστημα ως ένας πίνακας που την περιγράφει με μαθηματικό τρόπο. Όπως σημειώνεται στο [76], η μαθηματική περιγραφή μπορεί να είναι απλώς οι τιμές φω-

τεινότητας των εικονοψηφίδων (pixel data based approach) είτε, πιο σύνθετα, ορισμένα χαρακτηριστικά που έχουν εξαχθεί από τις εικόνες (feature based approach). Τα χαρακτηριστικά που συνήθως χρησιμοποιούνται επιλέγονται κατάλληλα, ώστε να εμφανίζουν ανοχή σε μετασχηματισμούς ομοιότητας. Αυτό αποτελεί σημαντικό πλεονέκτημα έναντι της πρώτης προσέγγισης, που για να εξασφαλίσει αυτή την ανεξαρτησία θα πρέπει να συμπεριλάβει στο σύνολο των δεδομένων πολλαπλές εικόνες του ίδιου αντικειμένου, υπό διαφορετικούς πιθανούς μετασχηματισμούς. Ακόμα, η δεύτερη μέθοδος συμπιέζει το μέγεθος των δεδομένων, μειώνοντας ριζικά την πολυπλοκότητα του συστήματος και τις απαιτήσεις σε χρόνο υπολογισμού και υπολογιστική ισχύ. Εν γένει, για τις περιπτώσεις που τα προς εντοπισμό αντικείμενα έχουν μεγάλο μέγεθος και ενδέχεται να παρουσιάζουν στροφές, μεταθέσεις και αλλαγές κλίμακας, η δημιουργία ενός συστήματος που βασίζεται στα χαρακτηριστικά είναι η βέλτιστη λύση. Η επιτυχία και η ακρίβεια της μεθόδου εξαρτώνται από τη σωστή επιλογή των χαρακτηριστικών που θα εξαχθούν από τα δεδομένα .

Στη συνέχεια παρουσιάζονται, ενδεικτικά, εφαρμογές όπου χρησιμοποιούνται τα ΤΝΔ για την ταξινόμηση των σημάτων οδικής κυκλοφορίας.

Στο [77] στην περιοχή ενδιαφέροντος, που έχει προκύψει από το πρώτο βήμα, πραγματοποιείται επανασύσταση με διγραμμική παρεμβολή ώστε να αποκτήσει διαστάσεις 50x50, καθώς όλες οι εικόνες που τροφοδοτούνται στο στάδιο της αναγνώρισης πρέπει να έχουν ενιαίες διαστάσεις. Οι διαστάσεις αυτές υπολογίζονται κατόπιν μελέτης των διαστάσεων των πινακίδων σε μία μέση κλίμακα για τη συγκεκριμένη φωτομηχανή που χρησιμοποιήθηκε. Η κάθε περιοχή ενδιαφέροντος, αφού υποστεί επεξεργασία που βασίζεται στο ιστόγραμμα της για καλύτερη ακρίβεια των αποτελεσμάτων ταξινόμησης, εισάγεται σε ένα πολυστρωματικό δίκτυο πρόσθιας τροφοδότησης. Στη συγκεκριμένη μέθοδο επιλέχθηκε να μην εξαχθούν χαρακτηριστικά, συνεπώς τα δεδομένα εισόδου στο ΤΝΔ είναι οι τιμές φωτεινότητας της περιοχής ενδιαφέροντος. Δημιουργούνται συνολικά έξι νευρωνικά δίκτυα, ένα για κάθε είδος πινακίδων (πληροφοριακές, απαγορευτικές, κίνδυνος, στοπ, απαγόρευση στάθμευσης και υποχρεωτικές). Κάθε στρώμα εισόδου αποτελείται από 2.500 νευρώνες (διαστάσεις ίσες με τον αριθμό εικονοψηφίδων της εικόνας με την οποία τροφοδοτείται στο δίκτυο) και τα στρώματα εξόδου σε κάθε νευρωνικό δίκτυο απαρτίζονται από τόσους νευρώνες όσα τα ξεχωριστά σήματα που ανήκουν σε κάθε κατηγορία. Πραγματοποιήθηκαν διάφορες δοκιμές για να αποφασιστεί εάν τα αποκρυπτόμενα στρώματα πρέπει να είναι ένα ή δύο. καθώς και για το πόσους νευρώνες πρέπει να έχει το κάθε αποκρυπτόμενο στρώμα. Ο τελικός αριθμός των κόμβων των στρωμάτων για κάθε ΤΝΔ, κατόπιν των δοκιμών, φαίνονται στην Εικόνα 2.29.

Νευρωνικό δίκτυο για περιοχές ενδιαφέροντος που δεν απεικονίζουν σήματα δε δημιουργήθηκε, αφού οι συγκεκριμένες περιοχές θα απομακρύνονται όταν παρουσιάζονται πολύ χαμηλές τιμές σε όλα τα στρώματα εξόδου.

| Net Name | # Hidden Layer | Geometry | # Output |
|--------------|----------------|----------|----------|
| Prohibitions | 2 | 115+65 | 37 |
| Information | 1 | 50 | 10 |
| Obligation | 2 | 90+40 | 26 |
| Danger | 1 | 175 | 42 |
| No-Parking | 1 | 40 | 3 |
| Stop | 1 | 80 | 3 |

Εικόνα 2.29: Η αρχιτεκτονική των ΤΝΔ που χρησιμοποιήθηκαν στο [77]

Τα ΤΝΔ εκπαιδεύτηκαν με τον αλγόριθμο ανάστροφης διάδοσης σφάλματος, με χρήση προσομοιωμένων και πραγματικών δεδομένων που απεικόνιζαν δεδομένα καλής ποιότητας, κακής ποιότητας (π.χ. με αποκρύψεις) και δείγματα που δεν περιείχαν σήματα οδικής κυκλοφορίας. Η διαδικασία της εκπαίδευσης πραγματοποιήθηκε για 50.000, 100.000 και 200.000 επαναλήψεις.

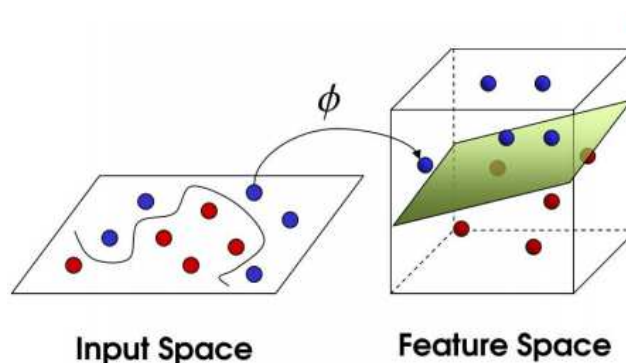
Στο [78] διερευνάται η αναγνώριση έξι σημάτων. Για κάθε διαφορετικό σήμα, δημιουργήθηκε ένα διαφορετικό ΤΝΔ, το οποίο δίνει δύο πιθανές απαντήσεις. Δηλαδή εξετάζει αν το σήμα ανήκει ή όχι στην κλάση που μελετάται, άρα πρόκειται για έναν δυαδικό ταξινομητή. Τα συγκριμένα ΤΝΔ είναι γνωστά με το όνομα “perceptrons” και αποτελούν την απλούστερη μορφή δικτύου πρόσθιας τροφοδότησης. Πλεονέκτημα είναι το σχετικά μικρό μέγεθός τους και η εύκολη επέκταση του συστήματος για τον εντοπισμό και άλλων σημάτων. Τα perceptrons που χρησιμοποιήθηκαν είναι πολυστρωματικά και πλήρως διασυνδεδεμένα. Εκπαιδεύτηκαν με τη βοήθεια του ελαστικού αλγορίθμου ανάστροφης μετάδοσης σφάλματος (resilient backpropagation), ο οποίος συγκλίνει πιο γρήγορα από τον κλασικό αλγόριθμο ανάστροφης διάδοσης σφάλματος, και με τον αλγόριθμο scale conjugate gradient, ώστε να συγκριθούν τα αποτελέσματα. Για την εκπαίδευσή τους χρησιμοποιήθηκαν διάφορες εικόνες που απεικονίζουν τα σήματα οδικής κυκλοφορίας όσο και εικόνες με αντικείμενα που δεν αποτελούν σήματα, ώστε να μπορούν τα δίκτυα να απομακρύνουν τις εσφαλμένες περιοχές ενδιαφέροντος. Τα δεδομένα εισαγωγής στα ΤΝΔ είναι οι τιμές χρώματος εικονοψηφίδων εικόνων με διαστάσεις 30x30. Κατά τη διαδικασία αναγνώρισης ιδιαίτερο ρόλο διαδραματίζουν οι κόμβοι των αποκρυπτόμενων στρωμάτων, που λειτουργούν ως ανιχνευτές των περιγραφικών χαρακτηριστικών που είναι ικανά να διαχωρίσουν τα δεδομένα μεταξύ τους, όπως ορίζεται και στο [79].

Τέλος, στο [42] επιχειρείται η κατηγοριοποίηση των πινακίδων με τα διαφορετικά όρια ταχύτητας. Μετά τη φάση του εντοπισμού, οι περιοχές ενδιαφέροντος κανονικοποιούνται και απομακρύνονται οι δακτύλιοι των κυκλικών σημάτων, ώστε στα δεδομένα εισόδου του νευρωνικού δικτύου να περιλαμβάνεται μόνο το εσωτερικό τμήμα της πινακίδας. Με τον τρόπο αυτό μειώνεται το μέγεθος των δεδομένων εισόδου. Στη συνέχεια, πραγματοποιείται επανασύσταση εικόνας ώστε οι εικόνες που τροφοδοτούνται στο ΤΝΔ να είναι διαστάσεων 20x20 (για να έχουν όλες οι περιοχές ενδιαφέροντες ενιαίες διαστάσεις, όπως αναφέρεται και στα ανωτέρω). Τα ΤΝΔ που χρησιμοποιήθηκαν είναι πολυστρωματικά δίκτυα πρόσθιας τροφοδότησης, με 400

νευρώνες στο στρώμα εισόδου (αφού οι εικόνες έχουν διαστάσεις 20x20), 30 νευρώνες για το αποκρυπτόμενο στρώμα και 12 νευρώνες στο στρώμα εξόδου (δηλαδή όσες και οι κατηγορίες που έχουν οριστεί). Το αποτέλεσμα κάθε νευρώνα για το στρώμα εξόδου είναι μία πιθανότητα, η οποία λαμβάνει τιμή μεταξύ 0 και 1. Ο νευρώνας για τον οποίο εντοπίζεται η μέγιστη πιθανότητα είναι εκείνος που καθορίζει το αποτέλεσμα. Για να θεωρηθεί έγκυρη η ταξινόμηση πρέπει η μέγιστη πιθανότητα να απέχει από την αμέσως επόμενη περισσότερο από 0.5. Επίσης, πραγματοποιήθηκαν διάφορες δοκιμές σχετικά με τον αριθμό των νευρώνων του αποκρυπτόμενου στρώματος (20, 30 και 40 νευρώνες) και με τον αριθμό των επαναλήψεων για την εκπαίδευση του δικτύου. Ο τελικός αριθμός επαναλήψεων που χρησιμοποιήθηκε είναι 150.

2.2.3 Μηχανές Διανυσματικής Στήριξης (Support Vector Machines)

Μία εξίσου διαδεδομένη τεχνική μηχανικής μάθησης είναι οι Μηχανές Διανυσματικής Στήριξης (Support Vector Machines – SVM). Οι SVM εκπαιδεύονται υπό επίβλεψη, με τη χρήση δεδομένων εκπαίδευσης. Με βάση τα συγκεκριμένα δεδομένα, που είναι εκ των προτέρων γνωστό σε ποια κλάση ανήκουν, επιχειρείται να δημιουργηθεί ένας αλγόριθμος που, στην περίπτωση της δυαδικής ταξινόμησης, κατατάσσει νέα δεδομένα σε μία από τις δύο πιθανές κλάσεις. Τα δεδομένα προβάλονται σε έναν διανυσματικό χώρο πολλών διαστάσεων ώστε να καθίσταται δυνατός ο γραμμικός διαχωρισμός τους από ένα υπερεπίπεδο. Ο αλγόριθμος στοχεύει ουσιαστικά στον υπολογισμό του υπερεπιπέδου με τέτοιο τρόπο ώστε να μεγιστοποιούνται τα περιθώρια (margins) από τις δύο κλάσεις (για περισσότερα βλ. κεφάλαιο 3). Η μεταφορά των δεδομένων στον «χώρο των χαρακτηριστικών», όπου αυτά διαχωρίζονται γραμμικά, γίνεται μέσω των συναρτήσεων πυρήνων (kernels) (Εικόνα 2.30). Η θεωρία των SVM επεκτάθηκε στην πορεία ώστε να μπορεί να κατηγοριοποιεί δεδομένα σε πολλαπλές κλάσεις (multiclass classification).



Εικόνα 2.30: Η προβολή των εισερχόμενων δεδομένων στον «χώρο των χαρακτηριστικών»

(Πηγή: <http://www.imtech.res.in/raghava/rbpred/svm.jpg>)

Οι SVM χρησιμοποιούνται ευρέως στην ανάλυση εικόνας με πολύ ικανοποιητικά αποτελέσματα. Όμοια με τα νευρωνικά δίκτυα, τα προς ταξινόμηση δεδομένα εισόδου μπορούν να είναι οι τιμές του χρώματος των εικονοσηφίδων (template based approach) ή χαρακτηριστικά που έχουν εξαχθεί από την εικόνα (feature based ap-

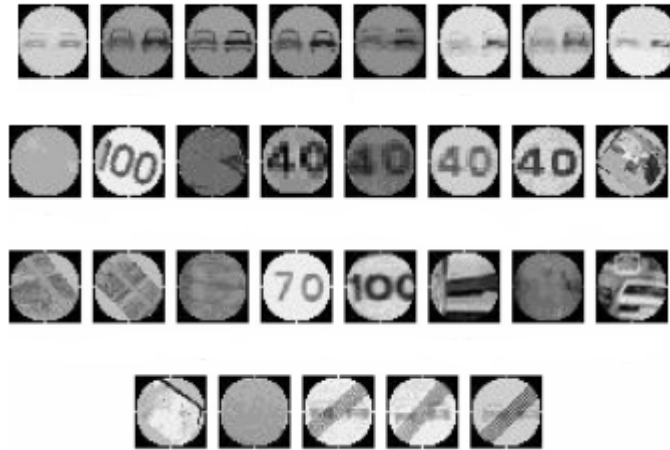
proach). Η επιλογή του τρόπου περιγραφής των δεδομένων που εισάγονται στο σύστημα ταξινόμησης θα πρέπει να γίνει έτσι ώστε να αποφευχθεί η χρήση πολλών παραμέτρων (overfitting of data), καθώς στην περίπτωση αυτή μειώνεται η ακρίβεια της ταξινόμησης. Αντιστοίχως, αν τα δεδομένα χαρακτηρίζονται από πολύ λίγες παραμέτρους, πιθανόν να μην είναι εφικτός ο επιτυχής διαχωρισμός τους. Για τον λόγο αυτό, θα πρέπει να μελετάται και να ελέγχεται με λεπτομέρεια ο τρόπος με τον οποίο θα περιγράφονται τα δεδομένα. Η ίδια λογική ισχύει εκτός από τις SVM και στα νευρωνικά δίκτυα.

Στην περίπτωση των υποβοηθούμενων συστημάτων οδήγησης χρησιμοποιούνται ευρέως οι SVM για την αναγνώριση των σημάτων οδικής κυκλοφορίας. Ακολούθως γίνεται ενδεικτική αναφορά σε σχετικές ερευνητικές εργασίες.

Στο [80] η ταξινόμηση των σημάτων οδικής κυκλοφορίας γίνεται με SVM. Τα δεδομένα εισόδου στο σύστημα είναι οι περιγραφείς HOG (Histogram of Oriented Gradient descriptors) που έχουν εξαχθεί από τις αντίστοιχες εικόνες. Πρόκειται, δηλαδή, για μία μέθοδο ταξινόμησης που βασίζεται στα χαρακτηριστικά. Οι περιγραφείς HOG υπολογίζουν σε σύνολα εικονοψηφίδων τα μέτρα των κλίσεων σε προκαθορισμένο αριθμό διευθύνσεων (για περισσότερα βλ. κεφάλαιο 4). Στη συγκεκριμένη δημοσίευση εφαρμόζεται μία ταξινόμηση σε έξι κατηγορίες, για εντοπισμό δύο σημάτων. Η κατηγορία 0 χρησιμοποιείται όταν δεν υπάρχει κάποια έγκυρη πινακίδα στην εικόνα. Αντίστοιχα, η κατηγορία 1 περιγράφει μία πινακίδα ΣΤΟΠ, η 2 πινακίδα ΣΤΟΠ με μερικές αποκρύψεις και η 3 σήματα ΣΤΟΠ που βρίσκονται σε μεγάλη απόσταση. Όμοιες με τις κατηγορίες 1, 2 και 3 είναι οι κατηγορίες 4, 5 και 6 για την πινακίδα που υπαγορεύει παραχώρηση προτεραιότητας. Ο αλγόριθμος εκπαιδεύεται με χρήση πολλαπλών εικόνων. Για κάθε εικόνα δημιουργείται ένα διάνυσμα που καταγράφει τον περιγραφέα HOG και σημειώνεται επίσης η κατηγορία στην οποία ανήκει η εικόνα. Το σύνολο των δεδομένων αυτών αποτελεί τα δεδομένα εκπαίδευσης του συστήματος.

Στα [30] και [33] οι περιοχές ενδιαφέροντος που έχουν προκύψει από το στάδιο του εντοπισμού υφίστανται επεξεργασία ώστε να αποκτήσουν διαστάσεις 31x31 και υπολογίζονται οι τιμές φωτεινότητας μόνο για τις εικονοψηφίδες του εσωτερικού των πινακίδων, ώστε να μειωθούν οι διαστάσεις των δεδομένων εισόδου. Οι SVM χρησιμοποιούν μία Γκαουσιανή συνάρτηση πυρήνα (Gaussian kernel) και το είδος κατηγοριοποίησης που χρησιμοποιείται είναι μία κατηγορία έναντι όλων των υπολοίπων (one versus all). Πρόκειται για ένα είδος ταξινόμησης πολλών κατηγοριών (multiclass classification), όπου τα δεδομένα που ανήκουν στην κατηγορία επιστρέφουν αποτέλεσμα 1 και τα δεδομένα που ανήκουν σε όλες τις υπόλοιπες κατηγορίες ή σε καμία από αυτές επιστρέφουν αποτέλεσμα -1. Ανάλογα με το σχήμα και το χρώμα της περιοχής ενδιαφέροντος που εντοπίστηκαν κατά το πρώτο στάδιο εφαρμόζονται οι αντίστοιχες SVM. Τα δείγματα που χρησιμοποιήθηκαν για ταξινόμηση κυμαίνονται μεταξύ 20 και 100 εικόνων για κάθε κατηγορία. Ανάμεσα σε αυτά συγκαταλέγονται τόσο αρνητικά όσο και θετικά παραδείγματα.

Ίδια μεθοδολογία ακολουθείται στο [29], όπου η εκπαίδευση του συστήματος πραγματοποιήθηκε με χρήση του λογισμικού LIBSVM, που διατίθεται ελεύθερα στο διαδίκτυο [81]. Στην Εικόνα 2.31 δίνονται ορισμένες από τις εικόνες που χρησιμοποιήθηκαν για την εκπαίδευση του συστήματος που αναγνωρίζει την πινακίδα απαγόρευσης προσπέρασης. Στην επάνω σειρά εμφανίζονται τα θετικά παραδείγματα. Στις υπόλοιπες σειρές εμφανίζονται τα αρνητικά, που περιλαμβάνουν άλλες πινακίδες και εικόνες με αντικείμενα που δεν αποτελούν σήματα οδικής κυκλοφορίας.



Εικόνα 2.31: Θετικά και αρνητικά παραδείγματα δεδομένων εκπαίδευσης

(Πηγή:[29])

Στα [81] και [20] η αναγνώριση πραγματοποιείται με χρήση SVM. Τα δεδομένα που εισάγονται για την ταξινόμηση είναι οι ροπές Zernike (Zernike moments), όπως αυτές περιγράφονται στο [83]. Οι ροπές Zernike, Legendre καθώς και οι ορθογωνικές ροπές Fourier-Melin, είναι σύνολα χαρακτηριστικών που εξάγονται από εικόνες και παρουσιάζουν ανοχή ως προς στροφές και θόρυβο. Οι ροπές Zernike προκύπτουν από την εφαρμογή ενός συνόλου πολύπλοκων πολυωνύμων, που καταλήγουν σε ένα πλήρες σύνολο χαρακτηριστικών μέσα σε έναν μοναδιαίο κύκλο. Οι δισδιάστατες ροπές τάξης p με επανάληψη q σε μία εικόνα με συνάρτηση φωτεινότητας $f(x,y)$ δίνονται από τον ακόλουθο τύπο:

$$Z_{pq} = \frac{p+1}{\pi} \int \int_{x^2+y^2 \leq 1} f(x,y) V_{pq}^*(x,y) dx dy \quad (2.31)$$

Όταν η εικόνα είναι διακριτή, η παραπάνω σχέση μετασχηματίζεται ως εξής:

$$Z_{pq} = \frac{p+1}{\pi} \sum_x \sum_y f(x,y) V_{pq}^*(x,y) \quad (2.32)$$

όπου $x^2 + y^2 \leq 1$ και $V_{pq}^*(x,y)$ το κυκλικό πολυώνυμο Zernike στον μοναδιαίο κύκλο, το οποίο ορίζεται μαθηματικά ως ακολούθως:

$$V_{pq}^*(x,y) = R_{pq}(r_{xy}) e^{jq\theta_{xy}} \quad (2.33)$$

όπου $r_{xy} = \sqrt{x^2 + y^2}$, $\theta_{xy} = \tan^{-1}\left(\frac{y}{x}\right)$ και

$$R_{pq}(r) = \sum_{k=0}^{(p-\frac{\|q\|}{2})} (-1)^k \frac{(p-k)!}{k! \left(\frac{p+|q|}{2}-k\right)! \left(\frac{p-|q|}{2}-k\right)!} r^{p-2q}$$

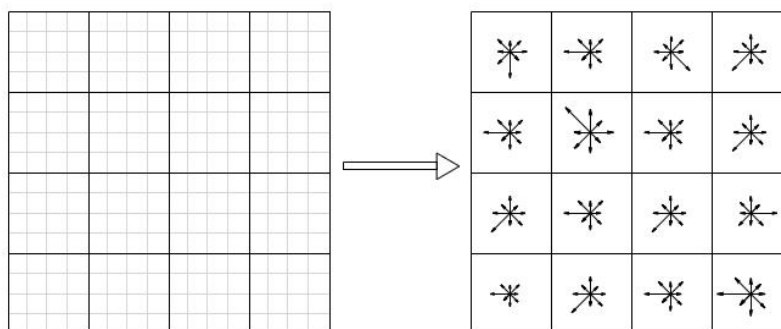
όπου $0 \leq |q| \leq p$ και ο αριθμός $p - |q|$ είναι περιττός.

Για την εξαγωγή των ροπών Zernike από μια πινακίδα οδικής κυκλοφορίας, αρχικά εντοπίζεται ο μικρότερος δυνατός κύκλος που περικλείει την πινακίδα και στη συνέχεια οι εικονοσυντεταγμένες μετασχηματίζονται με βάση αυτόν σε συντεταγμένες μοναδιαίου κύκλου. Τέλος, με χρήση των παραπάνω εξισώσεων υπολογίζονται οι ροπές.

Στο [81] το σύνολο των χαρακτηριστικών που εξήχθησαν χρησιμοποιήθηκαν ως δεδομένα εκπαίδευσης για διάφορες SVM, που μεταβάλλονται ως προς τη συνάρτηση πυρήνα. Για την εξαγωγή των ροπών οι τιμές (p, q) έλαβαν τιμές από (5,1) έως (12,12), ώστε για κάθε εικόνα το σύνολο των χαρακτηριστικών που εξάγονται να είναι ίσο με 40. Χρησιμοποιήθηκε γραμμική (linear), πολυωνυμική (polynomial), σιγμοειδής (sigmoid) και ακτινική (radialbasis function) συνάρτηση πυρήνα, ώστε να συγκριθεί η ακρίβεια της ταξινόμησης που παρέχει κάθε συνάρτηση. Ο τρόπος της ταξινόμησης ήταν one versus one. Στην περίπτωση αυτή, το σύνολο των δυαδικών ταξινομητών που πρέπει να δημιουργηθούν είναι ίσο με $k(k-1)/2$ (όπου k οι κατηγορίες προς ταξινόμηση).

Άλλη μία ερευνητική εργασία που χρησιμοποιεί τις SVM είναι η [84]. Το θέμα προσεγγίζεται με εξαγωγή χαρακτηριστικών από την εικόνα και τροφοδότηση τους στη μηχανή διανυσματικής στήριξης. Τα χαρακτηριστικά που χρησιμοποιούνται είναι οι περιγραφείς SIFT (Scale Invariant Feature Transform), οι οποίοι είναι ανθεκτικοί σε μετασχηματισμούς κλίμακας και στροφής, αλλαγές φωτεινότητας και ύπαρξη θορύβου [85]. Για τον υπολογισμό των διανυσμάτων περιγραφικών χαρακτηριστικών απαιτείται αρχικά ο εντοπισμός των σημείων ενδιαφέροντος. Ο αλγόριθμος λειτουργεί σε πυραμίδες εικόνων οι οποίες προκύπτουν από συνέλιξη της εικόνας με φίλτρο εξομάλυνσης Gauss. Για να δημιουργηθούν εικόνες με διαφορετική κλίμακα τα φίλτρα που εφαρμόζονται έχουν διαφορετικές τυπικές αποκλίσεις. Τα σημεία ενδιαφέροντος ανιχνεύονται με σύγκριση της τιμής κάθε εικονοψηφίδας με τις 8 γειτονικές, καθώς και με τις 9 τιμές που αντιστοιχούν στην αμέσως επόμενη και την αμέσως προηγούμενη εικόνα. Σε περίπτωση όπου το σημείο αποτελεί μέγιστο ή ελάχιστο ύστερα από τη σύγκριση με τις προαναφερθείσες τιμές και εάν η τιμή του είναι μεγαλύτερη από ένα προεπιλεγμένο κατώφλι, θεωρείται σημείο ενδιαφέροντος. Ακολουθεί ο ακριβής υπολογισμός της θέσης των σημείων ενδιαφέροντος με υποψηφιακή ακρίβεια. Κατόπιν απομακρύνονται τα σημεία που εντοπίζονται κατά μήκος ακμών και με τον τρόπο αυτό έχει ολοκληρωθεί ο εντοπισμός. Στη συνέχεια, γύρω από κάθε σημείο που έχει εντοπιστεί δημιουργείται ένα ιστόγραμμα ανά 10° , που επομένως χωρίζεται σε 36 τμήματα και σε αυτά απεικονίζει τη μεταβολή της έντασης. Το μέγιστο του ιστογράμματος αποτελεί τον προσανατολισμό του σημείου. Εάν υ-

πάρχουν περισσότερες της μίας κορυφής μεγαλύτερες από το 80% του μεγίστου του ιστογράμματος, δημιουργούνται περισσότερα του ενός σημεία ενδιαφέροντος, τα οποία διαφέρουν μόνο κατά προσανατολισμό. Ακολουθεί η δημιουργία ενός διανύσματος περιγραφικών χαρακτηριστικών για κάθε σημείο ενδιαφέροντος. Τα διανύσματα αυτά στρέφονται κατά τη γωνία του προσανατολισμού που έχει υπολογιστεί στο προηγούμενο βήμα. Για τον υπολογισμό του διανύσματος αυτού επιλέγεται περιοχή με διαστάσεις 4×4 περί το σημείο ενδιαφέροντος στην κλίμακα όπου αυτό έχει εντοπιστεί. Εκεί υπολογίζονται οι μεταβολές της έντασης σε 8 διευθύνσεις, με αποτέλεσμα ένα διάνυσμα με διαστάσεις $16 \times 8 = 128$ στοιχεία για κάθε σημείο ενδιαφέροντος (Εικόνα 2.32).



Εικόνα 2.32: Ο υπολογισμός των περιγραφικών χαρακτηριστικών

(Πηγή: [86])

Τα διανύσματα αυτά δεν πρέπει να είναι ευαίσθητα σε αλλαγές φωτεινότητας και για τον λόγο αυτό κανονικοποιούνται στο διάστημα $[0, 1]$. Με τον τρόπο αυτό, οποιαδήποτε γραμμική αλλαγή, όπως ο πολλαπλασιασμός κάθε pixel με σταθερά, δεν επηρεάζει τη συνταύτιση. Στη συνέχεια, στο νέο διάνυσμα εφαρμόζεται ένα μέγιστο κατώφλι ίσο με 0.2 και έπειτα κανονικοποιείται ξανά στο διάστημα $[0, 1]$. Έτσι, οι περιγραφείς που προκύπτουν είναι ανθεκτικοί και σε αλλαγές στη φωτεινότητα που έχουν προκύψει με μη γραμμικό τρόπο.

Οι περιγραφείς SIFT εμφανίζουν ομοιότητες με τους περιγραφείς HOG. Διαφοροποιούνται ως προς τη συχνότητα εξαγωγής των περιγραφικών χαρακτηριστικών, αφού οι πρώτοι αναφέρονται μόνο στα σημεία ενδιαφέροντος που έχουν ανιχνευτεί, ενώ οι δεύτεροι στο σύνολο της εικόνας. Για τον λόγο αυτό, και προκειμένου να είναι δυνατό οι περιγραφείς SIFT να χρησιμοποιηθούν για ανίχνευση αντικειμένων, υιοθετήθηκε το μοντέλο Bag of Words [87], [88], [89], σύμφωνα με το οποίο γύρω από κάθε σημείο ενδιαφέροντος λαμβάνεται μία περιοχή ενδιαφέροντος (Εικόνα 2.33). Οι περιοχές ενδιαφέροντος που προκύπτουν τελικά οργανώνονται σε συστάδες (clusters) με τον αλγόριθμο συσταδοποίησης k -μέσων (k -means clustering). Ο συγκεκριμένος αλγόριθμος είναι μία μη επιβλεπόμενη μέθοδος ταξινόμησης που προσπαθεί να ομαδοποιήσει τα όμοια δεδομένα σε k συστάδες (clusters), ελαχιστοποιώντας την Ευκλείδεια απόσταση. Επιλέγονται k συστάδες και τα δεδομένα κατατάσσονται στη συστάδα που βρίσκονται πιο κοντά (εκεί όπου ελαχιστοποιείται το μέτρο ομοιότητας).

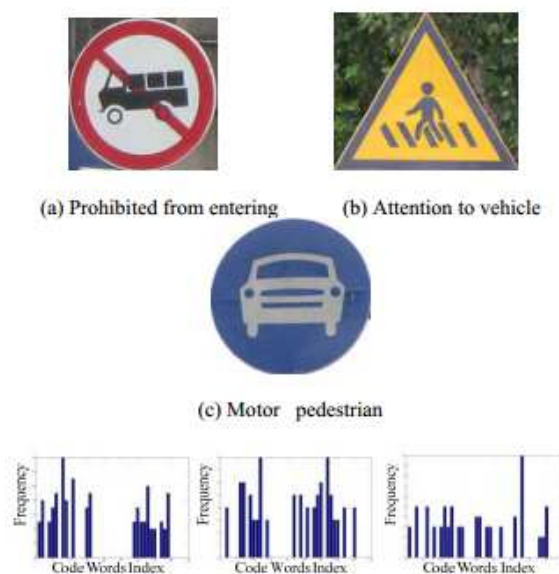
Όταν έχουν ταξινομηθεί όλα τα δεδομένα, υπολογίζονται εκ νέου τα κέντρα της κάθε συστάδας και η παραπάνω διαδικασία επαναλαμβάνεται έως ότου επιτευχθούν τα επιθυμητά όρια σύγκλισης. Οι συστάδες που τελικά δημιουργούνται μπορούν να παραλληλιστούν με λέξεις σε ένα κείμενο (codewords) που οργανώνονται ώστε να δημιουργήσουν ένα «λεξικό» (codebook), με μέγεθος n ίσο με τον αριθμό των διαφορετικών συστάδων.



Εικόνα 2.33: Το μοντέλο BoW

(Πηγή:[84])

Αφού έχουν υπολογιστεί οι τελικές συστάδες υπολογίζεται η συχνότητα εμφάνισης της κάθε μίας, όπως φαίνεται και στην Εικόνα 2.34. Τα ιστογράμματα αποτελούν τα δεδομένα που εισέρχονται στο σύστημα ταξινόμησης.



Εικόνα 2.34: Ιστογράμματα εμφάνισης κάθε «λέξης»

(Πηγή:[84])

Στη συγκεκριμένη έρευνα εκπαιδεύτηκαν τρεις SVM, για κάθε μία από τις τρεις κατηγορίες σημάτων απαγόρευσης, προειδοποίησης και υπόδειξης. Το σύνολο των δεδομένων εκπαίδευσης ήταν 130 εικόνες, ειλημμένες υπό διαφορετικές καιρικές συνθήκες και συνθήκες φωτισμού, με διαστάσεις 50x50. Χρησιμοποιήθηκε η ακτινική συνάρτηση πυρήνα, και η εκπαίδευση πραγματοποιήθηκε με χρήση του λογι-

σμικού LIBSVM.

Άλλες μέθοδοι, λιγότερο διαδεδομένες, που έχουν κατά καιρούς εφαρμοστεί για την αναγνώριση σημάτων οδικής κυκλοφορίας είναι η χρήση random forests [90], [91], [92] και η χρήση του ταξινομητή πλησιέστερου γείτονα (k-nearest neighbor) [93].

3

Μηχανές Διανυσματικής Στήριξης (SVM)

Για το δεύτερο στάδιο της διαδικασίας εντοπισμού και ταξινόμησης των σημάτων οδικής κυκλοφορίας, που αποτελεί ουσιαστικά την αναγνώριση των πινακίδων, απαιτείται η χρήση κάποιου εργαλείου από το πεδίο της Μηχανικής Μάθησης. Το εργαλείο που επιλέχθηκε είναι οι Μηχανές Διανυσματικής Στήριξης (SVM). Καθώς οι Μηχανές Διανυσματικής Στήριξης βασίζονται σε ένα ισχυρό θεωρητικό και μαθηματικό υπόβαθρο, κρίθηκε σκόπιμη η εκτενής περιγραφή και ανάλυσή τους σε ένα ξεχωριστό κεφάλαιο, προκειμένου να γίνουν βαθύτερα κατανοητές οι έννοιές τους, πριν αναπτυχθεί ο τρόπος και τα είδη Μηχανών Διανυσματικής Στήριξης που χρησιμοποιήθηκαν, στα πλαίσια της εφαρμογής που αναπτύχθηκε. Οι SVM ανήκουν στην κατηγορία των γραμμικών ταξινομητών, σε αντίθεση με τα νευρωνικά δίκτυα, που χρησιμοποιούνται ευρέως σε αντίστοιχα προβλήματα και αποτελούν μη γραμμικούς ταξινομητές. Η πρώτη αναφορά στις SVM πραγματοποιήθηκε το 1979 από τον Vapnik [94], η ευρεία διάδοση και χρήση τους ξεκίνησε όμως μετά το 1995 [95]. Στο παρόν κεφάλαιο πραγματοποιείται εκτενής αναφορά στον τρόπο λειτουργία τους, τα διάφορα είδη SVM και την ταξινόμηση με χρήση μηχανών διανυσματικής στήριξης.

3.1 Γενίκευση των SVM

Στις μεθόδους Μηχανικής Μάθησης η διαδικασία της γενίκευσης αναφέρεται στην εφαρμογή των συναρτήσεων σε νέα δεδομένα, ώστε τελικά να ταξινομηθούν. Σημαντικό στοιχείο για τις SVM είναι ότι η γενίκευση πραγματοποιείται με ελαχιστοποίηση του Δομικού Διαρθρωτικού Κινδύνου (Structural Risk Minimization) και όχι του Εμπειρικού Κινδύνου (Empirical Risk Minimization), όπως συμβαίνει στα νευρωνικά δίκτυα. Αυτή η διαφορά εξασφαλίζει καλύτερα αποτελέσματα ταξινόμησης. Η διαδικασία της γενίκευσης και ο ορισμός των κινδύνων αναλύεται στη συνέχεια.

3.1.1 Πραγματικός, εμπειρικός και διαρθρωτικός κίνδυνος

Τα δεδομένα εκπαίδευσης αποτελούνται από ένα σύνολο ιδιοτήτων (attributes) και από αποκρίσεις (labels), που καθορίζουν την τάξη στην οποία ανήκουν. Για δυαδική ταξινόμηση οι δύο πιθανές αποκρίσεις λαμβάνουν τιμές $\{+1, -1\}$. Έστω πίνακας με

ιδιότητες $\mathbf{x} \in \mathbb{R}^n$ και y η απόκριση του κάθε διανύσματος ιδιοτήτων. Υπάρχει μία άγνωστη κατανομή πιθανότητας $P(\mathbf{x}, y)$ η οποία καθορίζει την πιθανότητα σωστής ταξινόμησης του παραδείγματος [96]. Η κατανομή πιθανότητας διέπει τη γενική περίπτωση, της οποίας τμήμα είναι και η περίπτωση που εξετάζεται εδώ (όπου σε κάθε διάνυσμα ιδιοτήτων η απόκριση λαμβάνει προκαθορισμένες τιμές). Οι SVM στοχεύουν στον υπολογισμό μίας συνάρτησης $f(\mathbf{x}, \alpha)$, η οποία για κάθε διάνυσμα ιδιοτήτων υπολογίζει την απόκριση (όπου α ένα σύνολο παραμέτρων που έχουν οριστεί πλήρως μετά από την εκπαίδευση του συστήματος). Τη συνάρτηση αυτή θα τη «μάθει» το σύστημα με βάση τα ζεύγη δεδομένων $\mathbf{x}_i \rightarrow y_i$, που έχουν χρησιμοποιηθεί ως δεδομένα εκπαίδευσης. Ο αναμενόμενος κίνδυνος (expected risk) της απόκρισης για τα νέα δεδομένα ισούται με:

$$R(a) = \int \frac{1}{2} |y - f(\mathbf{x}, \alpha)| dP(\mathbf{x}, y) \quad (3.1)$$

Η σχέση 3.1 μπορεί να χαρακτηριστεί ακόμα ως πραγματικός κίνδυνος. Εκτός από τον πραγματικό κίνδυνο, ορίζεται και ο εμπειρικός κίνδυνος (empirical risk) ο οποίος ισούται με:

$$R(a)_{emp} = \frac{1}{2l} \sum_{i=1}^l |y_i - f(\mathbf{x}, \alpha)| \quad (3.2)$$

Ο εμπειρικός κίνδυνος ταυτίζεται με τον μέσο αριθμό λαθών που προκύπτουν κατά τη διαδικασία εκπαίδευσης και προκύπτει πάντα ο ίδιος για συγκεκριμένο σύνολο δεδομένων εκπαίδευσης και παραμέτρων α .

Η συνάρτηση που ποσοτικοποιεί το σφάλμα μεταξύ προβλεπόμενης τιμής και απόκρισης είναι μία συνάρτηση απώλειας (loss function), η οποία στη δεδομένη περίπτωση ισούται με:

$\frac{1}{2} |y - f(\mathbf{x})|$. Για τιμές απώλειας η για τις οποίες ισχύει $0 \leq \eta \leq 1$, ο πραγματικός κίνδυνος μπορεί να λάβει τιμές:

$$R(a) \leq R(a)_{emp} + \sqrt{\frac{h \left(\log \left(\frac{2l}{h} \right) + 1 \right) - \log \left(\frac{\eta}{4} \right)}{l}} \quad (3.3)$$

όπου h μη-μηδενικός ακέραιος που ονομάζεται διάσταση Vapnik Chernovekis (VC dimension).

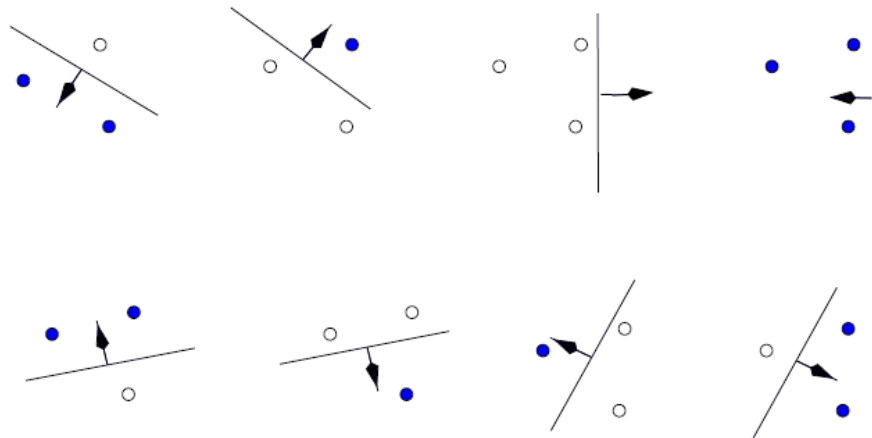
Η σχέση 3.3 ορίζει τον διαρθρωτικό κίνδυνο, που οφείλει να ελαχιστοποιείται σε μία μηχανή διανυσματικής στήριξης.

3.1.2 Διάσταση VC

Η διάσταση VC καθορίζεται από τον μέγιστο αριθμό των παραδειγμάτων που μπορούν να διαχωριστούν πλήρως (shattering). Πιο συγκεκριμένα, εάν σε μία δυαδική ταξινόμηση υπάρχουν τρία παραδείγματα, ο συνολικός αριθμός των πιθανών συνδυασμών (dichotomies) που προκύπτουν ανάλογα με την απόκριση του κάθε παραδείγματος είναι 2^3 :

$$\{(000), (100), (001), (010), (110), (011), (101), (111)\}$$

Εάν για όλες τις παραπάνω πιθανές περιπτώσεις υπάρχει συνάρτηση $f(\mathbf{x}, a)$ που να αποδίδει τις σωστές αποκρίσεις, τότε τα παραδείγματα διαχωρίζονται πλήρως και η διάσταση VC ισούται με 3 (δηλαδή με τον αριθμό των παραδειγμάτων που διαχωρίζονται). Το ότι υπάρχει μία ομάδα με αριθμό δεδομένων που ταξινομείται ορθά, σύμφωνα με τα παραπάνω, δεν συνεπάγεται αυτόματα ότι οποιαδήποτε διαφορετική ομάδα δεδομένων με αυτές τις διαστάσεις θα διαχωρίζεται πλήρως. Η διάσταση VC καθορίζεται στην πράξη από τη διάσταση των δεδομένων. Εάν τα δεδομένα έχουν διάσταση n , η διάσταση VC ισούται με $n+1$. Αυτό συμβαίνει γιατί το ένα σημείο ορίζεται ως αρχή και τα υπόλοιπα n σημεία, εάν είναι γραμμικά ανεξάρτητα, διαχωρίζονται από ένα προσανατολισμένο υπερεπίπεδο (oriented hyperplane). Ο διαχωρισμός $n+2$ σημείων δεν θα ήταν δυνατός αφού στον χώρο διαστάσεων n δεν υπάρχουν $n+1$ γραμμικά ανεξάρτητα σημεία. Ο προσανατολισμός του υπερεπίπεδου καθορίζει την πλευρά των θετικών αποκρίσεων. Στο ακόλουθο σχήμα (Εικόνα 3.1) εμφανίζονται οι 8 πιθανές περιπτώσεις και πώς ένα δισδιάστατο γραμμικό όριο μπορεί να τις διαχωρίσει. Τα δεδομένα ανήκουν στον χώρο \mathbb{R}^2 , άρα η διάσταση VC ισούται με $2+1 = 3$ (δηλαδή είναι ίση με τον μέγιστο αριθμό παραδειγμάτων που μπορούν να διαχωριστούν πλήρως)



Εικόνα 3.1: Τα τρία παραδείγματα που διαχωρίζονται πλήρως

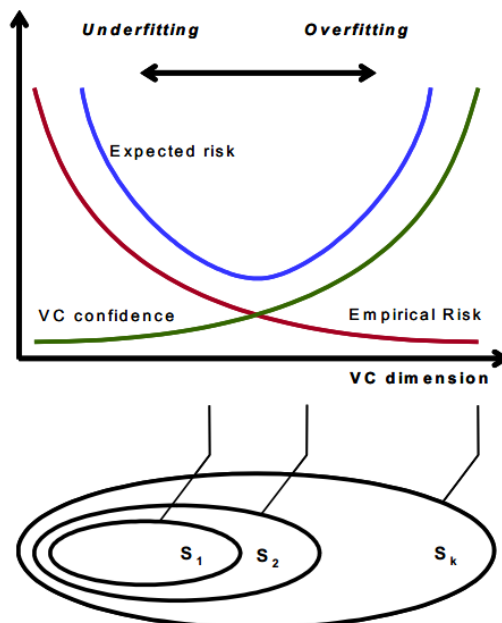
(Πηγή:[97])

Η διάσταση VC είναι ενδεικτική της ικανότητας μάθησης του συστήματος [97]. Η ικανότητα μάθησης πρέπει να επιλεγεί με τρόπον ώστε να μην είναι πολύ μεγάλη

αλλά ούτε πολύ μικρή, καθώς και στις δύο περιπτώσεις δημιουργούνται προβλήματα στη γενίκευση.

3.1.3 Ελαχιστοποίηση δομικού κινδύνου

Το αριστερό μέλος της Σχέσης 3.3 δεν μπορεί να υπολογιστεί, σε αντίθεση με το δεξί μέλος, το οποίο υπολογίζεται εύκολα εάν η διάσταση VC είναι γνωστή. Η επιλογή του η γίνεται με τρόπον ώστε να ελαχιστοποιείται η τιμή που μπορεί να λάβει ο πραγματικός κίνδυνος, και αυτό συνιστά τη βασική ιδέα της ελαχιστοποίησης του δομικού κινδύνου (structural risk minimization). Αυτή η αρχή βασίζεται στο γεγονός ότι ο ρυθμός λάθους μιας μηχανής μάθησης σε δεδομένα ελέγχου (ρυθμός λάθους γενίκευσης- πραγματικός κίνδυνος) έχει άνω όριο από το άθροισμα του ρυθμού λάθους εκπαίδευσης (εμπειρικός κίνδυνος) και ενός όρου που εξαρτάται από την διάσταση VC. Στην περίπτωση των γραμμικά διαχωρίσιμων προτύπων, το δίκτυο SVM μηδενίζει τον πρώτο όρο και ελαχιστοποιεί τον δεύτερο. Για το λόγο αυτό ένα δίκτυο SVM μπορεί να εμφανίζει μεγάλη ικανότητα γενίκευσης [98]. Για να ελαχιστοποιηθεί ο όρος στα δεξιά της ανισότητας για δεδομένο η , αρκεί να επιλεγεί το σύστημα με τη μικρότερη διάσταση VC. Στην περίπτωση που το εμπειρικό ρίσκο δεν είναι μηδενικό, πρέπει να βρεθεί μία ισορροπία μεταξύ του εμπειρικού κινδύνου και της διάστασης VC. Στο σημείο ισορροπίας ο διαρθρωτικός κίνδυνος ελαχιστοποιείται. Για τον σκοπό αυτό δημιουργούνται διάφορες υποομάδες δεδομένων εκπαίδευσης, με αυξανόμενη διαδοχικά τη διάσταση VC. Για κάθε υποομάδα εκπαιδεύεται ένα σύστημα ώστε να ελαχιστοποιηθεί το άθροισμα του ρυθμού λάθους εκπαίδευσης. Τελικά επιλέγεται το σύστημα που εξασφαλίζει το μικρότερο φραγμένο άνω όριο ως άθροισμα της διάστασης VC και του εμπειρικού κινδύνου. Τα παραπάνω εμφανίζονται σχηματικά στο διάγραμμα της Εικόνας 3.2.



Εικόνα 3.2:Ελαχιστοποίηση διαρθρωτικού κινδύνου

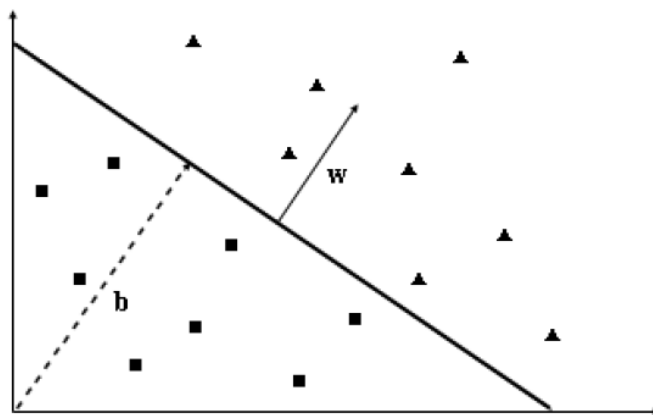
(Πηγή:[99])

3.2 Γραμμικές Μηχανές Διανυσμάτων Στήριξης

Η πιο εύκολη περίπτωση που συναντάται είναι οι γραμμικές μηχανές διανυσμάτων στήριξης, που εκπαιδεύονται σε δεδομένα τα οποία μπορούν να διαχωριστούν πλήρως στον χώρο των ιδιοτήτων (attributes). Έστω $x_i \in \mathbb{R}^d$ το i -διάνυσμα χαρακτηριστικών και $y_i \in \{-1, 1\}$ η απόκρισή του. Οι δύο κατηγορίες θα διαχωρίζονται από ένα υπερεπίπεδο, που χωρίζει τα παραδείγματα κατά τρόπον ώστε τα διανύσματα της ίδια κατηγορίας να μένουν στην ίδια πλευρά του. Το υπερεπίπεδο αποτελεί ουσιαστικά το σύνορο απόφασης για κάθε νέο δεδομένο. Δηλαδή, κάθε καινούριο διάνυσμα ιδιοτήτων ταξινομείται με βάση τη θέση του ως προς το υπερεπίπεδο, η μαθηματική εξίσωση του οποίου είναι:

$$\mathbf{w} \cdot \mathbf{x} + b = 0 \quad (3.4)$$

όπου \mathbf{x} ο πίνακας των ιδιοτήτων, \mathbf{w} το κάθετο διάνυσμα στο υπερεπίπεδο και b η τιμή μίας σταθεράς. Η διεύθυνση του κάθετου διανύσματος ορίζει την κατηγορία με θετική απόκριση (+1) και το μέγεθός του το βάρος του κάθε χαρακτηριστικού $i \in \mathbf{x}$. Ακόμα, η κάθετη απόσταση από το υπερεπίπεδο στην αρχή του χώρου των ιδιοτήτων ισούται με $|b|/\|\mathbf{w}\|$, όπου $\|\mathbf{w}\|$ είναι το μέτρο του διανύσματος.



Εικόνα 3.3: Σχηματική αναπαράσταση των b και w

(Πηγή:[100])

Καθώς υπάρχει πλήθος πιθανών υπερεπιπέδων που να διαχωρίζουν τα δεδομένα, πρέπει να επιλεγεί το βέλτιστο, το οποίο στην ουσία μεγιστοποιεί το περιθώριο (margin) ως προς τις δύο κλάσεις. Ως περιθώριο ορίζεται το άθροισμα της ελάχιστης απόστασης μεταξύ του υπερεπιπέδου και του κοντινότερου θετικού παραδείγματος με την ελάχιστη απόσταση μεταξύ του υπερεπιπέδου και του κοντινότερου αρνητικού παραδείγματος. Μαθηματικά, για το βέλτιστο υπερεπίπεδο ισχύουν τα εξής:

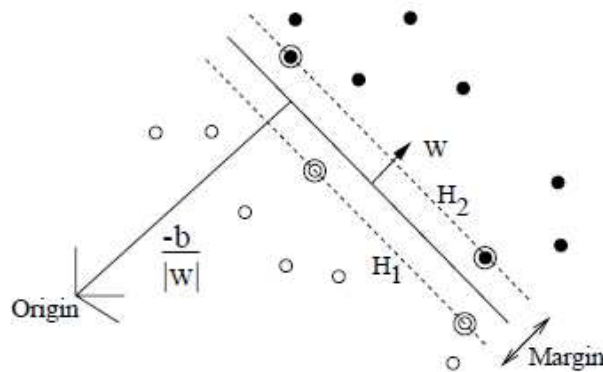
$$x_i \cdot \mathbf{w} + b \geq 1 \text{ για } y_i = +1$$

$$x_i \cdot \mathbf{w} + b \leq -1 \text{ για } y_i = -1 \quad (3.5 \text{ a} - b)$$

Οι δύο σχέσεις μπορούν να εκφραστούν ως μία με την ανισότητα:

$$y_i(\mathbf{x}_i \cdot \mathbf{w} + b) - 1 \geq 0 \quad \forall i \quad (3.6)$$

Για κάθε παράδειγμα που ικανοποιεί τη εξίσωση 3.5a η απόσταση από την αρχή του συστήματος ισούται με $|1-b/\|\mathbf{w}\|$, ενώ για κάθε παράδειγμα που ικανοποιεί την εξίσωση 3.5b η αντίστοιχη απόσταση ισούται με $|-1-b/\|\mathbf{w}\|$. Συνεπώς, η απόσταση των ακραίων παραδειγμάτων από το υπερεπίπεδο είναι ίση με $1/\|\mathbf{w}\|$ και το περιθώριο με $2/\|\mathbf{w}\|$. Για να μεγιστοποιηθεί το περιθώριο, αρκεί να ελαχιστοποιηθεί το $\|\mathbf{w}\|^2$, με ταυτόχρονη επιβολή της δέσμευσης που ορίζει η ανισότητα 3.6. Όσο μεγαλύτερο είναι το περιθώριο, τόσο καλύτερη είναι η ικανότητα γενίκευσης του ταξινομητή που αναπτύσσεται. Στην Εικόνα 3.4 επισημαίνεται το περιθώριο μεταξύ των δύο υπερεπίπεδων H_1 και H_2 και η απόσταση από την αρχή των αξόνων. Τα παραδείγματα που σημειώνονται με κύκλο είναι τα κοντινότερα σημεία στα υπερεπίπεδα και, εάν δεν υπήρχαν, η λύση του συστήματος θα ήταν διαφορετική. Τα παραδείγματα αυτά ονομάζονται διανύσματα στήριξης (support vectors). Εάν σε ένα σύνολο δεδομένων εκπαίδευσης αφαιρεθούν όλα τα δεδομένα, εκτός των διανυσμάτων υποστήριξης, η λύση που θα προέκυπτε μετά από την εκπαίδευση με τα νέα δεδομένα δε θα παρουσίαζε διαφορές σε σχέση με την αρχική.



Εικόνα 3.4: Σχηματική αναπαράσταση του περιθωρίου

(Πηγή: [97])

Για την ευκολότερη επίλυση του προβλήματος, και αφού η συνάρτηση $\Phi(\mathbf{w}) = \frac{1}{2} \|\mathbf{w}\|^2$ που πρέπει να ελαχιστοποιηθεί είναι κυρτή, εισάγονται οι πολλαπλασιαστές Lagrange. Για δεσμεύσεις που εμφανίζονται με τη μορφή ανισότητας $c_i \geq 0$, οι πολλαπλασιαστές Lagrange πολλαπλασιάζονται με την ανισότητα και αφαιρούνται από τη συνάρτηση.

Τελικά, προκύπτει η σχέση:

$$\begin{aligned} L_p &= \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^l a_i (y_i (\mathbf{x}_i * \mathbf{w} + b) - 1) \\ &= \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^l a_i y_i (\mathbf{x}_i * \mathbf{w} + b) + \sum_{i=1}^l a_i \quad (3.7) \end{aligned}$$

Για τον υπολογισμό των παραμέτρων πρέπει να ελαχιστοποιηθεί η L_p ως προς τα \mathbf{w} και b . Ταυτόχρονα, όλες οι παράγωγοι του L_p ως προς τα a_i πρέπει να είναι μηδενικές και επίσης να ισχύει $a_i \geq 0$. Το σύνολο των συγκεκριμένων δεσμεύσεων μπορεί να ονομαστεί c_1 . Η επίλυση του προβλήματος ισοδυναμεί με την επίλυση ενός κυρτού, τετραγωνικού προβλήματος. Για τον λόγο αυτό, μπορεί να διατυπωθεί και το δυϊκό του συγκεκριμένου προβλήματος, σύμφωνα με το θεώρημα δυϊκότητας Wolfe ([101]). Το δυϊκό πρόβλημα επιλύεται με μεγιστοποίηση του L_p , με ταυτόχρονη εφαρμογή των δεσμεύσεων μηδενισμού των παραγώγων του L_p ως προς \mathbf{w} και b και της δέσμευσης $a_i \geq 0$. Το συγκεκριμένο σύνολο των δεσμεύσεων μπορεί να ονομαστεί c_2 . Οι δύο διατυπώσεις του προβλήματος είναι ισοδύναμες και οδηγούν στις ίδιες τιμές \mathbf{w} , a και b .

Οι δεσμεύσεις c_2 διατυπώνονται μαθηματικά ως εξής:

$$\frac{\partial L_p}{\partial \mathbf{w}} = 0 \Rightarrow \mathbf{w} = \sum_{i=1}^l a_i y_i \mathbf{x}_i \quad (3.8)$$

$$\frac{\partial L_p}{\partial b} = 0 \Rightarrow \sum_i a_i y_i = 0 \quad (3.9)$$

Με υποκατάσταση των παραπάνω στη σχέση 3.7 προκύπτει:

$$L_D = \sum_i a_i y_i - \frac{1}{2} \sum_{i,j} a_i a_j y_i y_j \mathbf{x}_i \mathbf{x}_j \quad (3.10)$$

όπου $\mathbf{H}_{ij} = y_i y_j \mathbf{x}_i \mathbf{x}_j$ ο Εσσιανός πίνακας (Hessian Matrix). Για να έχει λύση το πρόβλημα θα πρέπει ο \mathbf{H} να μην εμφανίζει αρνητικές ιδιοτιμές.

Τελικά, ο υπολογισμός του βέλτιστου υπερεπιπέδου πραγματοποιείται από τη διατύπωση του δυϊκού (L_D) και όχι του πρωταρχικού προβλήματος (L_p). Πλεονεκτήματα της προσέγγισης μέσω του δυϊκού προβλήματος συνιστούν η εμφάνιση των δεδομένων εκπαίδευσης πάντα σε εσωτερικά γινόμενα ($\mathbf{x}_i \mathbf{x}_j$) και η εξάρτηση του προβλήματος από τον αριθμό των δεδομένων εκπαίδευσης (ένας πολλαπλασιαστής

για κάθε παράδειγμα). Συνεπώς, η διαδικασία εκπαίδευσης πραγματοποιείται με μεγιστοποίηση του L_D και ταυτόχρονη ισχύ της δεύτερης δέσμευσης. Αφού υπολογιστούν οι πολλαπλασιαστές Lagrange, το βέλτιστο υπερεπίπεδο υπολογίζεται από τη σχέση 3.8 και με επιλογή ενός τυχαίου διανύσματος και της σχέσης:

$$a_i(y_i(\mathbf{x} * \mathbf{w} + b) - 1) = 0 \quad (3.11)$$

μπορεί να υπολογιστεί και η τιμή b .

Αξίζει να σημειωθεί ότι οι πολλαπλασιαστές Lagrange που προκύπτουν είναι $a_i > 0$ για τα διανύσματα υποστήριξης και $a_i = 0$ για τα λοιπά παραδείγματα.

3.3 Ελαστικό περιθώριο (soft margin)

Στην περίπτωση που τα δεδομένα δεν είναι γραμμικά διαχωρίσιμα, λόγω θορύβου ή λόγω πολυπλοκότητας των πραγματικών δεδομένων, ο παραπάνω αλγόριθμος δε θα μπορέσει να εντοπίσει ένα υπερεπίπεδο. Αυτό συμβαίνει γιατί δε θα ικανοποιούνται οι δεσμεύσεις 3.5a και 3.5b. Εάν εισαχθεί ένα επιπλέον κόστος, όταν αυτό είναι απαραίτητο (δηλαδή όταν τα παραδείγματα δεν είναι γραμμικά διαχωρίσιμα), οι δεσμεύσεις «χαλαρώνουν» και μπορεί να εντοπιστεί το βέλτιστο υπερεπίπεδο. Αυτό μπορεί να γίνει με την εισαγωγή ψευδομεταβλητών (slack variables) ξ_i στις δεσμεύσεις ώστε τελικά να διατυπώνονται μαθηματικά ως εξής:

$$\mathbf{x}_i * \mathbf{w} + b \geq 1 - \xi_i \quad \text{για } y_i = +1 \quad (3.12)$$

$$\mathbf{x}_i * \mathbf{w} + b \leq -1 - \xi_i \quad \text{για } y_i = -1 \quad (3.13)$$

$$\forall \xi_i \geq 0$$

Με τον τρόπο αυτό μπορούν να γίνονται αποδεκτά δεδομένα που εντοπίζονται στη σωστή πλευρά αλλά εντός της περιοχής καθορισμού, εάν $\xi < 1$, όσο και δεδομένα που έχουν ταξινομηθεί λάθος, βρίσκονται δηλαδή στη λάθος πλευρά του ορίου απόφασης, εάν $\xi > 1$. Το άθροισμα $\sum_{i=1}^n \xi_i$ όλων των ψευδομεταβλητών που εισάγονται είναι ενδεικτικό του μέγιστου αποδεκτού σφάλματος. Στη συγκεκριμένη περίπτωση, δε ζητείται η ελαχιστοποίηση της $\frac{1}{2} \|\mathbf{w}\|^2$ αλλά η ελαχιστοποίηση της $\frac{1}{2} \|\mathbf{w}\|^2 + C(\sum \xi_i)^n$, όπου C μία μεταβλητή που ορίζεται από τον χρήστη και όσο αυξάνει η τιμή της τόσο αυξάνει και το μέγιστο αποδεκτό σφάλμα. Εάν εφαρμόσει κανείς τους πολλαπλασιαστές Lagrange και επιλέξει το δυϊκό πρόβλημα, η μαθηματική σχέση που πρέπει να μεγιστοποιηθεί είναι η:

$$L_D = \sum_i a_i - \frac{1}{2} \sum_{i,j} a_i a_j y_i y_j \mathbf{x}_i \mathbf{x}_j \quad (3.14)$$

με ταυτόχρονη επιβολή των δεσμεύσεων:

$$0 \leq a_i \leq C \quad (3.15)$$

$$\sum_i a_i y_i = 0 \quad (3.16)$$

Το βέλτιστο υπερεπίπεδο υπολογίζεται με την ίδια σχέση 3.8, λαμβάνοντας υπόψη μόνο τα διανύσματα στήριξης. Συνεπώς, η μόνη διαφορά που εντοπίζεται λόγω της εισαγωγής των ψευτομεταβλητών είναι το άνω φράγμα C στους πολλαπλασιαστές a_i ([97]).

3.4 Μη γραμμικές μηχανές διανυσμάτων στήριξης

Συχνά τα δεδομένα δεν είναι γραμμικά διαχωρίσιμα στον χώρο των ιδιοτήτων (attributes). Οι μηχανές διανυσμάτων στήριξης λειτουργούν όμως και σε αυτές τις περιπτώσεις μέσω προβολής των διανυσμάτων των ιδιοτήτων στον χώρο των χαρακτηριστικών (feature space). Δηλαδή εάν $\Phi(x)$ η συνάρτηση που προβάλλει τα διανύσματα ισχύει ότι:

$$\Phi : R^d \rightarrow H \quad (3.17).$$

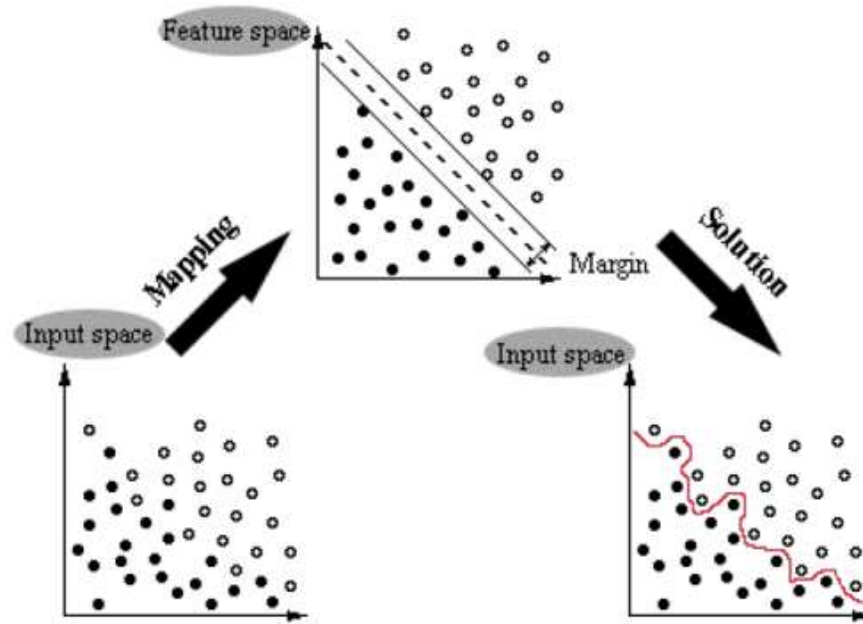
Ο χώρος των χαρακτηριστικών συνήθως έχει μεγαλύτερες διαστάσεις και εκεί τα δεδομένα μπορούν να διαχωριστούν από ένα υπερεπίπεδο, με τη διαδικασία που περιγράφεται παραπάνω. Σχηματικά η ταξινόμηση στην περίπτωση των μη γραμμικών μηχανών φαίνεται στην Εικ. 3.17. Το σύστημα μέσω της προβολής στον χώρο των χαρακτηριστικών είναι ικανό να εντοπίσει εκεί ένα γραμμικό όριο διαχωρισμού των δύο κατηγοριών, το οποίο δεν είναι γραμμικό στον χώρο των ιδιοτήτων.

Όταν κατά την εκπαίδευση χρησιμοποιούνται δεδομένα που έχουν προβληθεί στον χώρο των χαρακτηριστικών, η μόνη εξίσωση που μεταβάλλεται θα είναι η σχέση 3.10, η οποία πλέον είναι:

$$L_D = \sum_i a_i - \frac{1}{2} \sum_{i,j} a_i a_j y_i y_j \Phi(x_i) \Phi(x_j) \quad (3.18)$$

Κατά τη διαδικασία λοιπόν της εκπαίδευσης, τα προβεβλημένα διανύσματα δεν εμφανίζονται ποτέ μόνα τους, αλλά πάντα εμφανίζεται το εσωτερικό γινόμενο τους. Γίνεται έτσι κατανοητό ότι δεν απαιτείται η γνώση των $\Phi(x)$, αλλά η γνώση μίας συνάρτησης πυρήνα K (kernel function) για την οποία ισχύει:

$$K(x_i, x_j) = \Phi(x_i) * \Phi(x_j) \quad (3.19)$$



Εικόνα 3.5: Μη γραμμικός διαχωρισμός

(Πηγή:[100])

Τελικά, το υπερεπίπεδο που διαχωρίζει τα δεδομένα στον χώρο των χαρακτηριστικών εκφράζεται ως:

$$\mathbf{w} = \sum_{i=1}^{N_s} a_i y_i \Phi(s_i) \Phi(\mathbf{x}) + b = \sum_{i=1}^{N_s} a_i y_i K(s_i, \mathbf{x}) + b \quad (3.20)$$

όπου s_i τα διανύσματα στήριξης. Συνεπώς, δε χρειάζεται ούτε εδώ να υπολογιστεί η εξίσωση που προβάλλει τα διανύσματα στον χώρο των χαρακτηριστικών, παρά μόνο η γνώση της συνάρτησης πυρήνα K , η οποία προβάλλει το εσωτερικό γινόμενο στον χώρο των χαρακτηριστικών.

Όσον αφορά στις εξισώσεις πυρήνα, για την ίδια συνάρτηση K είναι δυνατόν να υπάρχουν περισσότερες από μία διαφορετικές συναρτήσεις $\Phi(\mathbf{x})$ που να την ικανοποιούν, και από αυτήν να εξαρτώνται και οι διαστάσεις του χώρου των χαρακτηριστικών. Χαρακτηριστικό παράδειγμα για διανύσματα ιδιοτήτων που ανήκουν στον \mathbb{R}^2 αποτελεί η συνάρτηση πυρήνα:

$$K(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i \cdot \mathbf{x}_j)^2 \quad (3.21)$$

και μπορεί να προκύψει από τη συνάρτηση:

$$\Phi(\mathbf{x}) = \begin{pmatrix} x_1^2 \\ \sqrt{2}x_1x_2 \\ x_2^2 \end{pmatrix} \quad (3.22)$$

που προβάλλει τα δεδομένα στον \mathbb{R}^3 ή από τη συνάρτηση:

$$\Phi(\mathbf{x}) = \begin{pmatrix} x_1^2 \\ x_1 x_2 \\ x_1 x_2 \\ x_2^2 \end{pmatrix} \quad (3.23)$$

που προβάλλει τα δεδομένα στον \mathbb{R}^4 .

3.5 Συναρτήσεις πυρήνα

3.5.1 Θεώρημα του Mercer

Προκειμένου μία συνάρτηση να αποτελεί συνάρτηση πυρήνα, θα πρέπει να αποτελεί εσωτερικό γινόμενο σε κάποιον χώρο. Αυτός ο έλεγχος μπορεί να πραγματοποιηθεί μέσω του θεωρήματος του Mercer. Σύμφωνα με αυτό, υπάρχει μία συνάρτηση προβολής $\Phi(\mathbf{x})$ και η αντίστοιχη συνάρτηση πυρήνα:

$$K(\mathbf{x}, \mathbf{y}) = \sum_i \Phi(\mathbf{x})_i \Phi(\mathbf{y})_i \quad (3.24)$$

εάν και μόνο εάν για κάθε $g(\mathbf{x})$ τέτοιο ώστε το

$\int g(\mathbf{x})^2 d\mathbf{x}$ να είναι πεπερασμένο ισχύει:

$$\int K(\mathbf{x}) g(\mathbf{x}) g(\mathbf{y}) d\mathbf{x} d\mathbf{y} \geq 0 \quad (3.25)$$

Επειδή η παραπάνω σχέση είναι πολλές φορές δύσκολο να αποδεχθεί, μπορεί η συνάρτηση πυρήνα να αντικατασταθεί με το άθροισμα των εσωτερικών γινόμενων υψωμένο σε μία θετική δύναμη, και η παραπάνω σχέση να αντικατασταθεί από την ακόλουθη:

$$\int \left(\sum_{i=1}^d \mathbf{x} \cdot \mathbf{y} \right)^p g(\mathbf{x}) g(\mathbf{y}) d\mathbf{x} d\mathbf{y} \geq 0 \quad (3.26)$$

Το θεώρημα του Mercer χρησιμοποιείται για να ελεγχθεί εάν μία συνάρτηση αποτελεί συνάρτηση πυρήνα, δεν παρέχει όμως καμία πληροφορία σχετικά με την $\Phi(\mathbf{x})$ ούτε σχετικά με τις διαστάσεις του χώρου των χαρακτηριστικών ([97]). Η ισχύς του θεωρήματος εξασφαλίζει ότι ο πίνακας $K(\mathbf{x}_i, \mathbf{x}_j) \forall i, j$ δεν έχει αρνητικές ιδιοτιμές, και συνεπώς το πρόβλημα έχει λύση στον χώρο των χαρακτηριστικών.

Για συναρτήσεις πυρήνα όπου η παραπάνω σχέση δεν ισχύει δεν εξασφαλίζεται η ύπαρξη λύσης στον εντοπισμό του βέλτιστου υπερεπιπέδου. Είναι πιθανό μία συνάρτηση να μην ικανοποιεί το ανωτέρω κριτήριο, αλλά να είναι ικανή να προβάλλει τα δεδομένα σε ένα χώρο χαρακτηριστικών όπου το βέλτιστο υπερεπιπέδο μπορεί να εντοπιστεί.

3.5.2 Συνήθεις συναρτήσεις πυρήνα

Έχουν προταθεί διάφορες συναρτήσεις πυρήνα που συγκεντρώνουν τα επιθυμητά χαρακτηριστικά. Οι πιο διαδεδομένες είναι η γραμμική, η σιγμοειδής, η Γκαουσιανή συνάρτηση ακτινικής βάσης, η πολυωνυμική και η εκθετική συνάρτηση ακτινικής βάσης. Άλλες πιο πολύπλοκες συναρτήσεις πυρήνα μπορούν να προκύψουν με πρόσθεση ή πολλαπλασιασμό των παραπάνω, καθώς και οι νέες συναρτήσεις θα συγκεντρώνουν τα επιθυμητά χαρακτηριστικά, όπως αναφέρεται στο [100]. Ακολούθως, αναλύονται οι πιο συχνά χρησιμοποιούμενες συναρτήσεις.

1. Γραμμική συνάρτηση πυρήνα (*linearkernel*)

Η γραμμική συνάρτηση πυρήνα αποτελεί την απλούστερη μορφή γραμμικού διαχωρισμού των δεδομένων. Ισούται με:

$$K(\mathbf{x}, \mathbf{y}) = \mathbf{x} \cdot \mathbf{y} \quad (3.27)$$

2. Πολυωνυμική συνάρτηση πυρήνα (*polynomial kernel*)

Η πολυωνυμική συνάρτηση πυρήνα αποτελεί μία από τις πιο ευρέως χρησιμοποιούμενες συναρτήσεις, όταν τα δεδομένα δεν είναι γραμμικά διαχωρίσιμα. Μαθηματικά εκφράζεται ως εξής:

$$K(\mathbf{x}, \mathbf{y}) = (\mathbf{x} \cdot \mathbf{y})^p \quad (3.28)$$

Αν και για καλύτερη διαχείριση των δεδομένων εκπαίδευσης συχνά διατυπώνεται και ως εξής:

$$K(\mathbf{x}, \mathbf{y}) = (\gamma(\mathbf{x} \cdot \mathbf{y}) + \delta)^p \quad (3.29)$$

όπου γ , δ , p οι παράμετροι της συνάρτησης. Για το γ ισχύει υποχρεωτικά $\gamma > 0$.

3. Γκαουσιανή συνάρτηση ακτινικής βάσης (*Gaussian radial basis function*)

Άλλη μία συνάρτηση πυρήνα που προβάλλει με μη γραμμικό τρόπο το εσωτερικό γινόμενο των διανυσμάτων σε έναν χώρο μεγαλύτερων διαστάσεων είναι η Γκαουσιανή συνάρτηση ακτινικής βάσης, η οποία είναι:

$$K(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{y}\|^2}{2\sigma^2}\right) \quad (3.30)$$

Η συγκεκριμένη συνάρτηση πυρήνα εξασφαλίζει πολλές φορές καλύτερα αποτελέσματα σε σχέση με την πολυωνυμική, καθώς η ακρίβεια κάθε μεθόδου εξαρτάται από τις τιμές των παραμέτρων που χρησιμοποιούνται. Για τη συγκεκριμένη συνάρτηση πρέπει να εντοπιστεί η βέλτιστη τιμή μόνο για την παράμετρο σ , σε αντίθεση με την πολυωνυμική συνάρτηση όπου πρέπει να εντοπιστεί ο βέλτιστος συνδυασμός για τις τρεις παραμέτρους. Γενικότερα, ο αριθμός των παραμέτρων επηρεάζει την πολυπλοκότητα του μοντέλου, και για αυτό θα πρέπει να χρη-

σιμοποιούνται κατά το δυνατόν λιγότερες παράμετροι. Τέλος, το εύρος τιμών της Γκαουσιανής συνάρτησης ακτινικής βάσης περιορίζεται στο διάστημα $[0,1]$, σε αντίθεση με την πολυωνυμική συνάρτηση που το εύρος τιμών της είναι το διάστημα $[0,\infty)$.

4. Εκθετική συνάρτηση ακτινικής βάσης (Exponential radial basis function)

Μία μικρή παραλλαγή της Γκαουσιανής συνάρτησης ακτινικής βάσης είναι η εκθετική συνάρτηση ακτινικής βάσης η οποία ισούται με:

$$K(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{y}\|}{2\sigma^2}\right) \quad (3.31)$$

5. Σιγμοειδής συνάρτηση (Sigmoid function)

Τέλος, μία συνήθης συνάρτηση πυρήνα είναι η σιγμοειδής συνάρτηση. Η χρήση της στις SVM ισοδυναμεί με τη χρήση ενός νευρωνικού δικτύου πρόσθιας τροφοδότησης, που αποτελείται από δύο στρώματα. Η σιγμοειδής συνάρτηση διατυπώνεται μαθηματικά ως εξής:

$$K(x, y) = \tanh(\gamma(x \cdot y) + \delta) \quad (3.32)$$

3.5.3 Επιλογή κατάλληλης συνάρτησης πυρήνα

Η επιλογή της καταλληλότερης συνάρτησης πυρήνα είναι ιδιαίτερα δύσκολη. Απαιτεί γνώση και κατανόηση του προβλήματος σε βάθος, καθώς και μελέτη της βιβλιογραφίας και των συναρτήσεων πυρήνα που έχουν ήδη χρησιμοποιηθεί. Συχνά χρησιμοποιούνται περισσότερες της μίας συναρτήσεις και η ακρίβεια της μεθόδου συγκρίνεται με τα δεδομένα ελέγχου. Τα δεδομένα ελέγχου αποτελούνται, όπως τα δεδομένα εκπαίδευσης, από διανύσματα ιδιοτήτων και τις αποκρίσεις τους. Εισάγονται στο σύστημα ως νέα δεδομένα και προβλέπεται η απόκριση για το κάθε ένα από αυτά, η οποία στη συνέχεια συγκρίνεται με την αντίστοιχη σωστή απόκριση. Ο λόγος των σωστά ταξινομημένων δεδομένων προς το σύνολο τους καλείται ακρίβεια του συστήματος.

Όταν έχει πλέον επιλεγεί η βέλτιστη συνάρτηση πυρήνα, για τον υπολογισμό των κατάλληλων παραμέτρων είναι χρήσιμο να πραγματοποιείται η διαδικασία της «εσωτερικής επαλήθευσης» (cross validation). Η διαδικασία αυτή εφαρμόζεται συχνά στις στατιστικές μεθόδους εκμάθησης (statistical learning methods) και στοχεύει στην πρόβλεψη της ακρίβειας της μεθόδου σε νέα δεδομένα. Η πιο συνηθισμένη διαδικασία cross validation είναι η k-fold cross validation ([101]). Σύμφωνα με την k-fold cross validation τα δεδομένα εκπαίδευσης υποδιαιρούνται σε K ομάδες με παρόμοιο μέγεθος. Για κάθε επανάληψη οι K-1 ομάδες χρησιμοποιούνται για εκπαίδευση της μεθόδου και τα υπόλοιπα δεδομένα, χρησιμοποιούνται σε δεδομένα ελέγχου, ώστε να προβλεφθεί η ακρίβεια της μεθόδου. Η διαδικασία πραγματοποιείται K φορές ώστε όλες οι υποομάδες να χρησιμοποιηθούν για επαλήθευση. Μία συνηθισμένη τιμή για το K σε αντίστοιχες εφαρμογές είναι 10 (συνεπώς το 90%

των δεδομένων χρησιμοποιείται για εκπαίδευση και το 10% για επαλήθευση). Σε κάθε επανάληψη δοκιμάζονται διαφορετικές τιμές για τις παραμέτρους που πρέπει να ορισθούν. Τελικά, επιλέγονται οι παράμετροι που εξασφαλίζουν τη μεγαλύτερη ακρίβεια.

Η παραπάνω μέθοδος μπορεί να χρησιμοποιηθεί και για επιλογή της κατάλληλης συνάρτησης πυρήνα. Για κάθε μέθοδο παρέχεται κάποιο μέτρο, όπως π.χ. η ακρίβεια που προαναφέρθηκε, που είναι ενδεικτικό της ικανότητας γενίκευσης του συστήματος. Για τον λόγο αυτό, η εσωτερική επαλήθευση χρησιμοποιείται όχι μόνο για σύγκριση διαφορετικών μεθόδων ή σύγκριση διαφορετικών τιμών μεταβλητών αλλά και ως εργαλείο πρόβλεψης της ικανότητας του συστήματος που εκπαιδεύεται.

3.6 v-SVM

Όσα αναλύθηκαν παραπάνω ισχύουν για την κατηγορία των ονομαζόμενων c-SVMs.

Το [102] αναφέρεται σε μία επέκταση αυτής, που αποτελεί την κατηγορία v-SVM. Στην κατηγορία αυτή εισάγεται μία μεταβλητή v με σκοπό να ελέγχει τον αριθμό των διανυσμάτων στήριξης και τον αριθμό των λαθών. Το πρόβλημα της δυαδικής ταξινόμησης διατυπώνεται πλέον ως ελαχιστοποίηση της σχέσης:

$$\frac{1}{2} \|\mathbf{w}\|^2 + \frac{1}{l} \left(\sum \xi_i \right) - \nu \rho \quad (3.33)$$

με ταυτόχρονη επιβολή των δεσμεύσεων:

$$y_i(\Phi(\mathbf{x}_i) * \mathbf{w} + b) + \xi_i \geq \rho \quad (3.34)$$

$$\xi_i, \rho \geq 0 \quad (3.52)$$

και το δυϊκό πρόβλημα διατυπώνεται πλέον ως μεγιστοποίηση της:

$$\frac{1}{2} \sum_{i,j} a_i a_j y_i y_j \Phi(\mathbf{x}_i) \Phi(\mathbf{x}_j) \quad (3.35)$$

με επιβολή των δεσμεύσεων:

$$\sum_i^l a_i y_i = 0 \text{ για } 0 \leq a_i \leq \frac{1}{l} \quad (3.36)$$

$$\sum_i^l a_i \geq \nu \quad (3.37)$$

3.7 Ταξινόμηση σε περισσότερες κατηγορίες

Όπως αναφέρεται και στην αρχή της ενότητας, οι SVM ξεκίνησαν ως δυαδικοί ταξινομητές, σύντομα όμως επεκτάθηκαν και για ταξινόμηση σε περισσότερες των δύο κατηγοριών (multiclass classification). Στο [103] συγκρίνονται οι διάφορες μέθοδοι αυτού του είδους ταξινόμησης.

Μία μέθοδος για τέτοιου είδους ταξινόμηση είναι η ταξινόμηση μίας κατηγορίας έναντι όλων (oneversus all classification). Στην περίπτωση αυτή εκπαιδεύονται συνολικά K μηχανές, όσες είναι και οι K διαφορετικές κατηγορίες. Τα δεδομένα εκπαίδευσης που ανήκουν στην κατηγορία για την οποία ο ταξινομητής εκπαιδεύεται σημειώνονται με θετική απόκριση και τα δεδομένα που ανήκουν στις υπόλοιπες κατηγορίες με αρνητική. Η διαδικασία εκπαίδευσης πραγματοποιείται σύμφωνα με όσα αναφέρθηκαν παραπάνω. Δηλαδή, επιδιώκεται η ελαχιστοποίηση της σχέσης:

$$\frac{1}{2} \|\mathbf{w}\|^2 + C \left(\sum \xi_i \right) \quad (3.38)$$

με ταυτόχρονη επιβολή των δεσμεύσεων:

$$\Phi(\mathbf{x}_i) * \mathbf{w} + b \geq 1 - \xi_i \quad (3.57) \text{ για τα δεδομένα που ανήκουν στην κατηγορία που εξετάζεται}$$

$$\Phi(\mathbf{x}_i) * \mathbf{w} + b \leq -1 + \xi_i \quad (3.58) \text{ για τα δεδομένα που ανήκουν σε όλες τις υπόλοιπες κατηγορίες}$$

$$\xi_i \geq 0 \quad (3.39)$$

Άλλη μέθοδος είναι η ταξινόμηση μίας κατηγορίας έναντι μίας άλλης κατηγορίας (one versus one). Στην περίπτωση αυτή αναπτύσσονται συνολικά $K(K-1)/2$ ταξινομητές για ταξινόμηση σε K κατηγορίες. Για την εκπαίδευση του κάθε ταξινομητή ως δεδομένα εκπαίδευσης, με θετική απόκριση ορίζονται τα δεδομένα που ανήκουν στην κατηγορία i και με αρνητική απόκριση τα δεδομένα που ανήκουν στην κατηγορία j . Έτσι, για κάθε λοιπόν ταξινομητή χρησιμοποιούνται δύο κατηγορίες και η διαδικασία ολοκληρώνεται όταν έχουν χρησιμοποιηθεί όλες οι κατηγορίες. Η διαδικασία πραγματοποιείται με ελαχιστοποίηση της συνάρτησης:

$$\frac{1}{2} \|\mathbf{w}\|^2 + C \left(\sum \xi_i \right) \quad (3.40)$$

με ταυτόχρονη επιβολή των δεσμεύσεων:

$$\Phi(\mathbf{x}_i) * \mathbf{w} + b \geq 1 - \xi_i \text{ για τα δεδομένα που ανήκουν στην κατηγορία } i$$

$$\Phi(\mathbf{x}_i) * \mathbf{w} + b \leq -1 + \xi_i \quad (3.41 \text{ a} - b) \text{ για τα δεδομένα που ανήκουν στην κατηγορία } j$$

$$\xi_i \geq 0 \quad (3.42)$$

Για την ταξινόμηση εφαρμόζεται το σύνολο των ταξινομητών και κάθε ταξινομητής «συλλέγει» ψήφους. Τα νέα δεδομένα τελικά κατατάσσονται στην κατηγορία για την οποία συγκεντρώνουν τις περισσότερες ψήφους

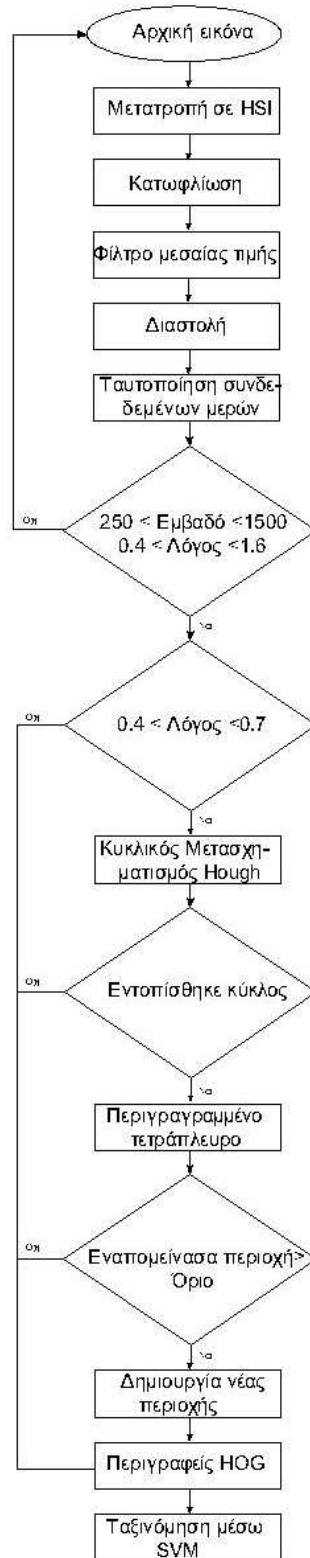
4

Προτεινόμενος αλγόριθμος

Στα πλαίσια της μεταπτυχιακής εργασίας συντάχθηκε ένας αλγόριθμος που εντοπίζει και ταξινομεί αυτόματα τις πινακίδες οδικής κυκλοφορίας, ως συστατικό μέρος ενός συστήματος υποβοήθησης οδήγησης. Τα δεδομένα εισόδου σε υπάρχοντα συστήματα μπορεί να είναι βίντεο, που αναλύονται σε καρέ και υφίστανται επεξεργασία όπως πχ. στα [29], [30], [33] ή εικόνες όπως πχ. στα [31], [32], [26]. Εδώ επιχειρείται η εξαγωγή πληροφοριών από δεδομένα μεμονωμένων εικόνων. Οι εικόνες λαμβάνονται ανά τακτά χρονικά διαστήματα από φωτογραφική μηχανή που έχει τοποθετηθεί στο αυτοκίνητο. Σε ορισμένες αντίστοιχες ερευνητικές εργασίες, όπως στην [104], περιορίζεται η περιοχή που εξετάζεται σε κάθε εικόνα. Οι πινακίδες οδικής κυκλοφορίας τοποθετούνται κατά κανόνα στα όρια του οδοστρώματος, σε προκαθορισμένα ύψη. Με βάση αυτό το χαρακτηριστικό, μπορούν να εντοπιστούν στην εικόνα περιοχές που εμφανίζουν την μεγαλύτερη πιθανότητα να συμπεριλαμβάνουν πινακίδες. Καθώς όμως υπάρχουν και διαφορετικοί τρόποι τοποθέτησης πινακίδων, για τον υπό ανάπτυξη αλγόριθμο επιλέχθηκε να μην περιοριστεί η περιοχή μελέτης και ο εντοπισμός να πραγματοποιείται με εξέταση της εικόνας στο σύνολο της.

Όπως αναφέρθηκε στο κεφάλαιο 2, τα βασικά βήματα για την ανάπτυξη ενός τέτοιου συστήματος είναι ο εντοπισμός και η αναγνώριση των σημάτων οδικής κυκλοφορίας. Ο αλγόριθμος που συντάχθηκε βασίζεται στα δύο αυτά βήματα. Αρχικά, μετατρέπει την εικόνα στον χρωματικό χώρο HSI και εντοπίζει περιοχές ενδιαφέροντος με χρήση ενός ενιαίου κατωφλίου. Στη συνέχεια, εφαρμόζεται στην εικόνα ένα φίλτρο για την απομάκρυνση θορύβου. Για τις περιοχές ενδιαφέροντος που έχουν εντοπιστεί χρησιμοποιείται η μέθοδος ταυτοποίησης συνδεδεμένων μερών (connected components labeling), ώστε να υπολογιστεί το περιγεγραμμένο τετράπλευρο κάθε ακμής (bounding box). Κάθε blob με εμβαδό μικρότερο ή μεγαλύτερο από ένα προκαθορισμένο όριο απομακρύνεται. Για κάθε ένα από τα εφαπτόμενα τετράπλευρα που εντοπίστηκαν υπολογίζεται ο λόγος των πλευρών τους και γίνονται αποδεκτά μόνο όσα εμφανίζουν τιμές εντός συγκεκριμένων κατωφλίων. Στις περιοχές ενδιαφέροντος με συγκεκριμένο λόγο μεταξύ των κάθετων πλευρών των εφαπτόμενων τετραπλεύρων, εφαρμόζεται ο κυκλικός μετασχηματισμός Hough και ακολουθεί το στάδιο της αναγνώρισης, κατά το οποίο εξάγονται οι περιγραφείς HOG για τις περιοχές ενδιαφέροντος και τροφοδοτούνται σε SVM, οι οποίες έχουν προηγουμένως εκπαιδευτεί. Σε κάθε περιοχή ενδιαφέροντος εφαρμόζονται οι κατάλληλοι ταξινομητές α-

νάλογα με το χρώμα και το σχήμα που έχει εντοπιστεί. Στο τελικό βήμα απομακρύνονται οι εσφαλμένες περιοχές ενδιαφέροντος και ταξινομούνται οι έγκυρες περιοχές στην κατηγορία της πινακίδας που απεικονίζουν. Ο αλγόριθμος περιγράφεται σχηματικά στην Εικόνα 4.1 αναλύεται στη συνέχεια του κεφαλαίου.



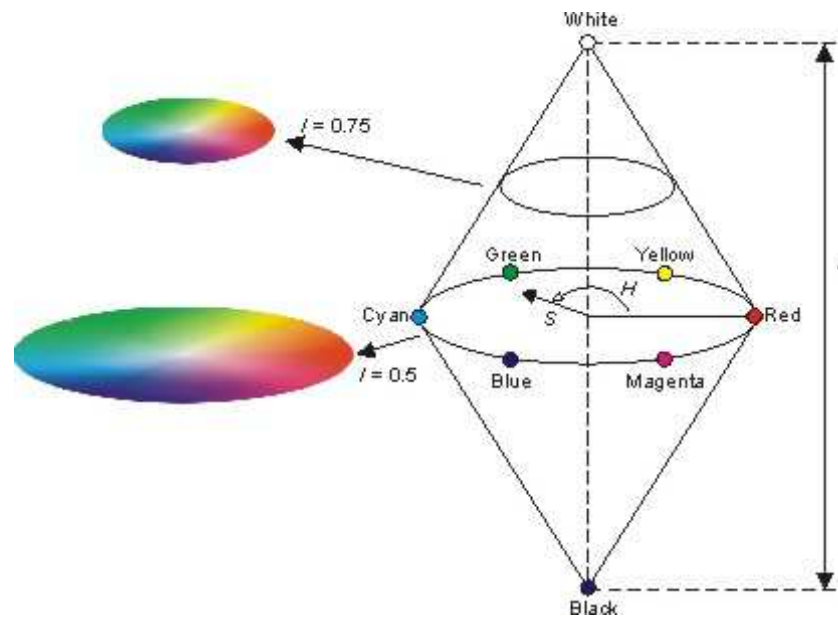
Εικόνα 4.1: Διάγραμμα ροής αλγορίθμου

4.1 Εντοπισμός στον χρωματικό χώρο HSI

4.1.1 Ο χρωματικός χώρος HSI

Το πρώτο βήμα για τη δημιουργία ενός συστήματος υποβοήθησης οδήγησης είναι ο εντοπισμός των πινακίδων οδικής κυκλοφορίας. Στον αλγόριθμο που αναπτύχθηκε, αποφασίστηκε να πραγματοποιηθεί ο εντοπισμός στον χρωματικό χώρο HSI με χρήση ενός ολικού καταφλίου, καθώς η μέθοδος αυτή εξασφαλίζει ικανοποιητικά αποτελέσματα, σύμφωνα με τη βιβλιογραφική επισκόπηση.

Αρχικά, η εικόνα μετασχηματίζεται από τον χρωματικό χώρο RGB στον χρωματικό χώρο HSI. Όπως αναφέρθηκε στο κεφάλαιο 2, ο χρωματικός χώρος HSI (Hue, Saturation, Intensity – Απόχρωση, Κορεσμός, Ένταση) προσομοιώνει ικανοποιητικά τον τρόπο αντίληψης των χρωμάτων από τον ανθρώπινο εγκέφαλο. Η πρώτη συνιστώσα H αναπαριστά την απόχρωση του εξεταζόμενου αντικειμένου, δηλαδή το πόσο αυτή πλησιάζει στα χρώματα κόκκινο, μπλε, πράσινο ή σε συνδυασμό τους και λαμβάνει τιμές από 0° - 360° . Η τιμή 0° υποδεικνύει το κόκκινο χρώμα, 60° είναι το κίτρινο, 120° το πράσινο, 180° το κυανό, 240° το μπλε και 300° το ματζέντα, όπως φαίνεται και στην Εικόνα 4.2. Εάν θεωρηθεί σημείο O στον χρωματικό χώρο, η αριθμητική τιμή της H προκύπτει από τη γωνία που σχηματίζεται μεταξύ του άξονα του κόκκινου και της γραμμής που ενώνει το κέντρο με το σημείο O .



Εικόνα 4.2:Ο χρωματικός χώρος HSI

(Πηγή:[105])

Η συνιστώσα του κορεσμού S περιγράφει το πόσο «καθαρό» είναι ένα χρώμα, δηλαδή το ποσοστό ανάμειξης με το λευκό φως και λαμβάνει τιμές μεταξύ $[0,1]$. Όσο μεγαλύτερη είναι η τιμή του κορεσμού τόσο καθαρότερο είναι το χρώμα. Στο ανωτέρω σχήμα, η αριθμητική τιμή του κορεσμού προκύπτει από το μήκος της ευθείας που ενώνει το σημείο O με το κέντρο. Τέλος, η ένταση I στο μοντέλο HSI

περιγράφει το ποσό του φωτός. Κανονικοποιείται ώστε να λαμβάνει τιμές μεταξύ [0,1] και η τιμή της ορίζεται σε σχέση με τη θέση που κατέχει πάνω σε μια κατακόρυφη γραμμή που περνά από το κέντρο της αναπαράστασης του χρωματικού χώρου. Η μικρότερη τιμή της έντασης αναπαριστά το μαύρο και η μεγαλύτερη το λευκό ([106]).

Οι σχέσεις μετασχηματισμού από τον χρωματικό χώρο RGB στον χρωματικό χώρο HSI είναι οι ακόλουθες:

$$H = \begin{cases} \theta & \text{εάν } B \leq G \\ 360 - \theta & \text{εάν } B \geq G \end{cases} \quad (4.1)$$

$$S = 1 - \frac{3}{(R + G + B)} [\min(R, G, B)] \quad (4.2)$$

$$I = \frac{1}{3}(R + G + B) \quad (4.3)$$

$$\text{Όπου } \theta = \cos^{-1} \left\{ \frac{\frac{1}{2}[(R-G)+(R-B)]}{[(R-G)^2+(R-B)(G-B)]^{1/2}} \right\} \quad (4.4)$$

Σημαντικό πλεονέκτημα των χώρων που περιέχουν τη συνιστώσα H, είναι η ανθεκτικότητά της σε γραμμικές αλλαγές φωτεινότητας και αντίθεσης (σύμφωνα με το [20]). Πιο συγκεκριμένα:

- Η απόχρωση H είναι ανθεκτική σε αλλαγές που προκύπτουν με πολλαπλασιασμό των τριών συνιστωσών του χώρου R,G, B με τη μεταβλητή a, δηλαδή $\text{Hue}(R,G,B) = \text{Hue}(aR, aG, aB)$ για όλες τις τιμές του a, υπό την προϋπόθεση ότι το $(aR, aG, aB) \in [0,255] \times [0,255] \times [0,255]$.
- Η απόχρωση H είναι ανθεκτική σε αλλαγές που προκύπτουν με μετατόπιση των τριών συνιστωσών του χώρου R,G, B κατά τη μεταβλητή a, δηλαδή $\text{Hue}(R,G,B) = \text{Hue}(a+R, a+G, a+B)$ για όλες τις τιμές του a, υπό την προϋπόθεση ότι το $(a+R, a+G, a+B) \in [0,255] \times [0,255] \times [0,255]$.
- Η απόχρωση H είναι ανθεκτική σε αλλαγές στον κορεσμό και την ένταση. Για τον λόγο αυτό είναι δυνατόν να εντοπιστεί μία συγκεκριμένη απόχρωση ενός αντικειμένου, ακόμα και υπό πολύ διαφορετικές συνθήκες φωτισμού.

Η χρήση του συγκεκριμένου χρωματικού χώρου εμφανίζει όμως και ορισμένα μειονεκτήματα όσο αφορά την απόχρωση:

- Όταν η ένταση λαμβάνει πολύ υψηλές ή πολύ χαμηλές τιμές, η απόχρωση δεν διαδραματίζει πλέον κανένα ρόλο.
- Όταν ο κορεσμός λαμβάνει τιμές κοντά στον 0, η συνιστώσα της απόχρωσης δεν επηρεάζει το αποτέλεσμα.
- Γενικότερα, όταν ο κορεσμός λαμβάνει μικρές τιμές, οι τιμές της απόχρωσης είναι ασταθείς.

4.1.2 Κατωφλίωση στον χρωματικό χώρο HSI

Αφού η εικόνα έχει μετασχηματιστεί στον χρωματικό χώρο HSI, εφαρμόζονται ενιαία κατώφλια για τον εντοπισμό του κόκκινου, κίτρινου και μπλε χρώματος, που είναι τα βασικά χρώματα των πινακίδων. Τα κατώφλια αναφέρονται μόνο στις συνιστώσες της απόχρωσης H και του κορεσμού S, ώστε τα αποτελέσματα του εντοπισμού να είναι ανεξάρτητα από τη φωτεινότητα της εικόνας. Οι δύο πρώτες συνιστώσες παρουσιάζουν πολύ μικρές μεταβολές, για το ίδιο χρώμα, ακόμα και εάν οι συνθήκες της λήψης μεταβάλλονται έντονα. Για τον καθορισμό των κατωφλίων απαιτείται η μελέτη του ιστογράμματος για τις συνιστώσες της απόχρωσης H και του κορεσμού S. Ελήφθησαν εικόνες υπό πολύ διαφορετικές καιρικές συνθήκες και συνθήκες φωτισμού, ώστε τα όρια που θα προκύψουν να λειτουργούν για το μεγαλύτερο ποσοστό των πιθανών περιπτώσεων λήψης. Από τις εικόνες αποκόπηκαν χειροκίνητα οι πινακίδες οδικής κυκλοφορίας, και ακολούθως εμφανίζονται ενδεικτικά ιστογράμματα που εξήχθησαν για τις δύο πρώτες συνιστώσες, για το μπλε, κόκκινο και κίτρινο χρώμα αντίστοιχα (Εικόνα 4.3). Τελικά, τα κατώφλια που υπολογίστηκαν είναι τα ακόλουθα:

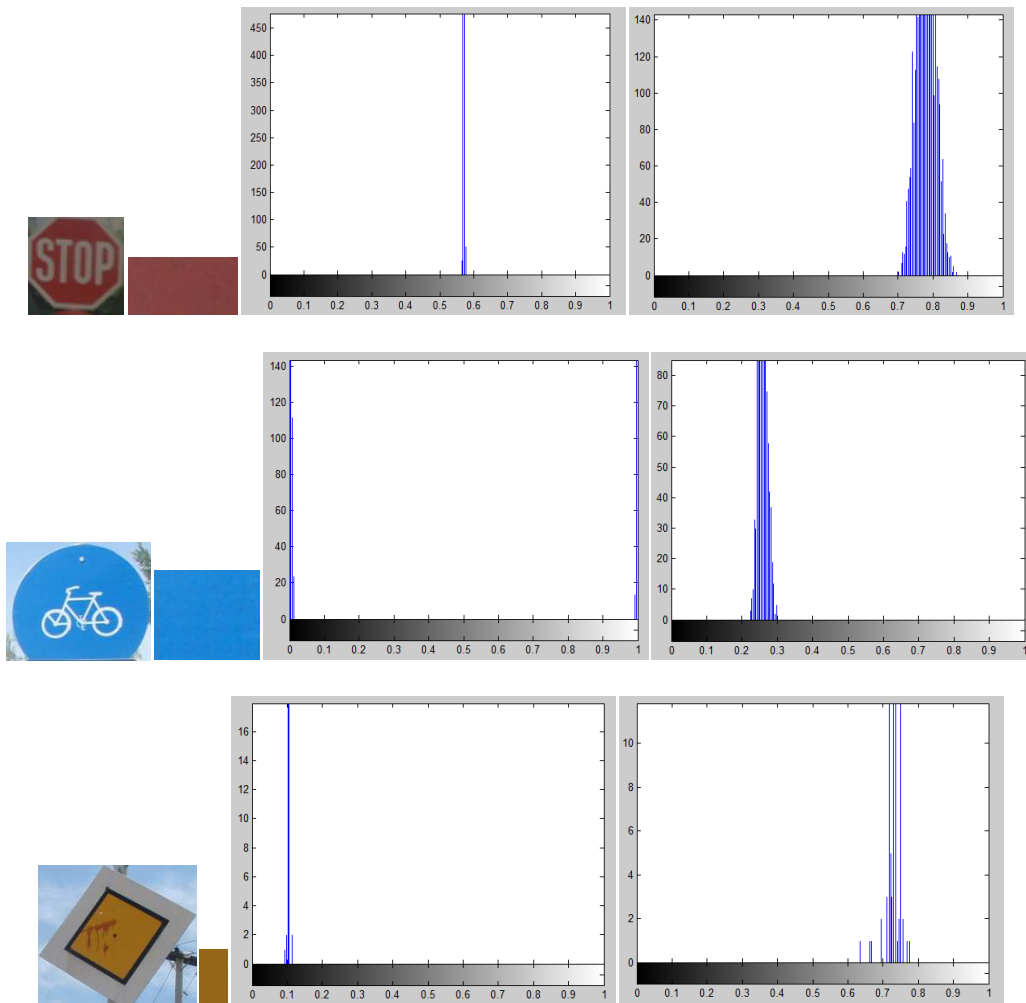
$180 < H < 290$ και $S > 0.4$ για το μπλε χρώμα

$H < 22$ ή $H > 306$ και $S > 0.15$ για το κόκκινο χρώμα

$18 < H < 72$ και $0.5 < S < 0.8$ για το κίτρινο χρώμα

Όσες εικονομηφίδες βρίσκονται εντός των προκαθορισμένων κατωφλίων λαμβάνουν τιμή 255 (λευκό χρώμα) και οι υπόλοιπες τιμή 0 (μαύρο χρώμα). Αποτέλεσμα της διαδικασίας της κατωφλίωσης είναι μία δυαδική εικόνα.

Δεν πραγματοποιήθηκε περαιτέρω μελέτη για την εξαγωγή των ουδέτερων χρωμάτων, γιατί οι πινακίδες στις οποίες επικεντρώνεται η παρούσα εργασία μπορούν να εντοπιστούν μόνο με εξαγωγή των τριών προαναφερθέντων χρωμάτων. Οι πινακίδες με μπλε χρώμα μπορούν να εντοπιστούν εξ ολοκλήρου με κατωφλίωση, ενώ για τις πινακίδες με κόκκινο δακτύλιο αρκεί ο εντοπισμός του εξωτερικού δακτυλίου. Μοναδική εξαίρεση αποτελούν τα σήματα ΣΤΟΠ και κίνησης σε δρόμο προτεραιότητας, καθώς ο εξωτερικός τους δακτύλιος είναι λευκός (Εικόνα 4.4). Οι πινακίδες όμως αυτές μπορούν να εντοπιστούν εύκολα, καθώς το εσωτερικό τους αποτελείται από ομοιογενές κόκκινο και κίτρινο χρώμα, αντίστοιχα.



Εικόνα 4.3: Πινάκίδα που αποκόπηκε χειροκίνητα, τμήμα της πινακίδας που χρησιμοποιήθηκε για την εξαγωγή ιστογραμμάτων και τα τελικά ιστογράμματα για τα κανάλια H (αριστερά) και S (δεξιά) για το μπλέ, κόκκινο και κίτρινο χρώμα.



Εικόνα 4.4:Οι δύο πινακίδες με λευκό δακτύλιο

4.2 Φίλτρο μεσαίας τιμής

Στην εικόνα που προκύπτει από την καταφλίσωση, εκτός από τις εντοπισμένες περιοχές των πινακίδων που είναι κατά τη μεγαλύτερη πλειοψηφία συμπαγείς, υπάρχουν και πολλές μεμονωμένες εικονοσηφίδες που λαμβάνουν την τιμή 255 και αποτελούν θόρυβο. Για την απομάκρυνσή τους χρησιμοποιήθηκε φίλτρο μεσαίας τιμής (median filter).

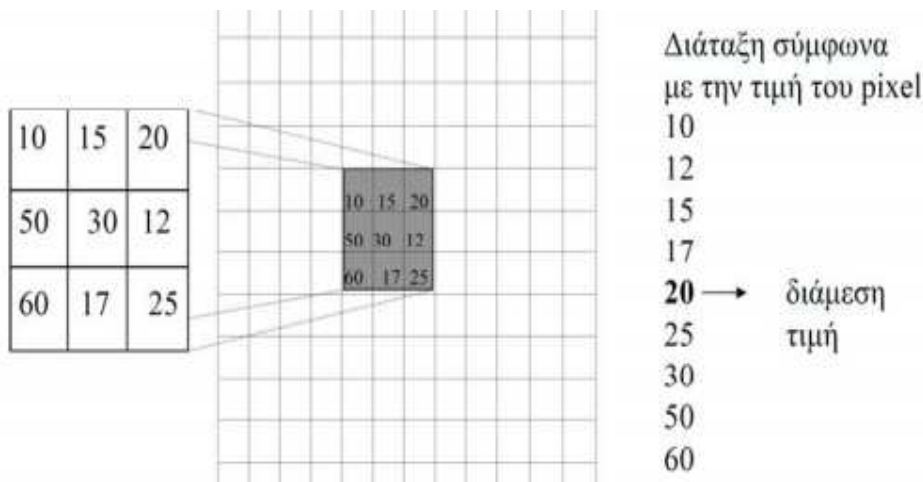
Τα φίλτρα χρησιμοποιούνται στην επεξεργασία σήματος, συχνά για την απομάκρυνση ορισμένων ανεπιθύμητων στοιχείων ([107]). Η εφαρμογή τους στην εικόνα πραγματοποιείται με τη διαδικασία της συνέλιξης μεταξύ της εικόνας και της μάσκας

(ενός τετραγωνικού παραθύρου). Η συνέλιξη αναπαρίσταται με το σύμβολο * και μαθηματικά ορίζεται ως εξής:

$$c[m, n] = a[m, n] * h[m, n] = \sum_{j=0}^{J-1} \sum_{k=0}^{K-1} h[j, k] a[m - j, n - k] \quad (4.5)$$

όπου h το φίλτρο και a η εικόνα.

Το φίλτρο μεσαίας τιμής είναι ένα, μη γραμμικό φίλτρο. Φιλτράρει δηλαδή κυρίως τις μεμονωμένες υψηλές συχνότητες που αποτελούν ενδείξεις θορύβου (outliers). Σύμφωνα με το [98], αποτέλεσμα εφαρμογής του φίλτρου μεσαίας τιμής είναι μία εικόνα ίδιων διαστάσεων με την αρχική όπου κάθε εικονοψηφίδα λαμβάνει τιμή ίση με τη διάμεσο του εξεταζόμενου τμήματος. Το τμήμα που εξετάζεται κάθε φορά εξαρτάται από τις διαστάσεις του φίλτρου (Εικόνα 4.5).



Εικόνα 4.5:Εφαρμογή φίλτρου μεσαίας τιμής

(Πηγή:[108])

Εδώ χρησιμοποιήθηκε φίλτρο μεσαίας τιμής με διαστάσεις 3x3 για την απομάκρυνση των μεμονωμένων εικονοψηφίδων που προέκυψαν κατά την κατωφλίωση. Το φίλτρο εφαρμόζεται στη δυαδική εικόνα, και η επιλογή των τελικών περιοχών ενδιαφέροντος προκύπτει με επιλογή των αντίστοιχων περιοχών στην αρχική, έγχρωμη εικόνα.

4.3 Διαστολή δυαδικής εικόνας

Στη δυαδική εικόνα που προκύπτει μετά από τη συνέλιξη με το φίλτρο μεσαίας τιμής εφαρμόζεται διαστολή (dilation), ώστε να ενωθούν κομμάτια της ίδια πινακίδας τα οποία αποκόπηκαν λόγω λανθασμένης κατάτμησης. Η λανθασμένη κατάτμηση μπορεί να οφείλεται σε αποκρύψεις ή σε πρόσθετα χαρακτηριστικά του σήματος οδικής κυκλοφορίας (για παράδειγμα, ύπαρξη κάποιου αυτοκόλλητου στην πινακίδα ή αλλοιωμένου χρώματος σε ένα μέρος του δακτυλίου σε βαθμό που οι τιμές του χρώματος να είναι εκτός των ορίων της κατωφλίωσης). Επίσης, η διαστολή χρησιμο-

ποιείται προκειμένου να διευρυνθούν τα όρια της περιοχής ενδιαφέροντος, ώστε να διευκολύνεται η διαδικασία της αναγνώρισης.

Η διαστολή της εικόνας είναι μία πράξη μαθηματικής μορφολογίας. Όπως αναφέρεται και στο [109], τα εργαλεία της μαθηματικής μορφολογίας έχουν ως στόχο την εξαγωγή χαρακτηριστικών από εικόνες. Τα εργαλεία αυτά αρχικά αναπτύχθηκαν για δυαδικές εικόνες, σύντομα όμως η χρήση τους γενικεύτηκε και σε λήψεις που καταγράφουν τους τόνους του γκριζού (gray-scale images). Ο κλάδος γνώρισε μεγάλη αναγνώριση όταν προτάθηκε η χρήση μορφολογικών φίλτρων συναρτήσεων κλίμακας για την επεξεργασία γεωτεχνικών εικόνων σε γεωτεχνικές εφαρμογές ([110]).

Οι μέθοδοι που συμπεριλαμβάνονται στη μαθηματική μορφολογία είναι μη γραμμικές, και με τη χρήση των κατάλληλων τελεστών και πράξεων μετασχηματίζουν τη μορφή ενός αντικειμένου ή οδηγούν στη εξαγωγή συγκεκριμένων χαρακτηριστικών από την εικόνα. Οι τελεστές εφαρμόζονται μέσω των δομικών στοιχείων (structuring element). Ως δομικό στοιχείο ορίζεται ένας πίνακας με τιμές 0 και 1, που μπορεί να έχει οποιοδήποτε σχήμα και μέγεθος. Οι διαστάσεις του πίνακα καθορίζουν το μέγεθος του δομικού στοιχείου, ενώ το πρότυπο των εικονοσηφίδων με τιμές 0 και 1 καθορίζουν το σχήμα του. Στην Εικόνα 4.6 εμφανίζονται χαρακτηριστικά παραδείγματα δομικών στοιχείων (με magenta σημειώνεται το σημείο αναφοράς του δομικού στοιχείου.)

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 |
| 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |

Εικόνα 4.6: Δομικά στοιχεία 5x5 σε σχήμα τετραγώνου, ρόμβου και σταυρού

(Πηγή:[111])

Τελικά, οι μορφολογικές πράξεις ορίζονται με τη βοήθεια κλασικών πράξεων και σχέσεων μεταξύ συνόλων.

Οι βασικές μέθοδοι της μαθηματικής μορφολογίας:

- διατηρούν τα πλήρη πλέγματα (complete lattices), δηλαδή τις σχέσεις διάταξης όπως αυτές ορίζονται από τη δομή του πεδίου εργασίας
- αντιστοιχούν στα βασικά στοιχεία των πλήρων πλεγμάτων το «supremum» (ελάχιστο άνω φράγμα) και το «infimum» (μέγιστο κάτω φράγμα) [112].

Το πλήρες πλέγμα είναι ένα σύνολο P για το οποίο ισχύει ότι:

- ορίζεται η μερική σχέση διάταξης \leq :

$$\begin{aligned} A &\leq A \\ A &\leq B, B \leq A \Rightarrow B = A \\ A &\leq B, B \leq C \Rightarrow A \leq C \quad (4.6 \text{ a} - \text{c}) \end{aligned}$$

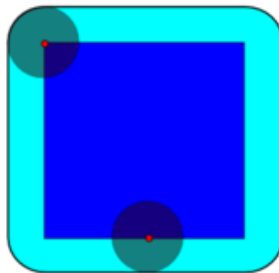
- για κάθε υποσύνολο του P υπάρχει ένα ελάχιστο άνω φράγμα και ένα μέγιστο κάτω φράγμα.

Οι βασικές λειτουργίες που διατηρούν τις σχέσεις διάταξης είναι η διαστολή (dilation) και η διάβρωση (erosion). Η διαστολή αντιστοιχεί στο ελάχιστο άνω φράγμα και προσθέτει λευκές εικονοψηφίδες στα όρια των αντικειμένων, ενώ η διάβρωση αντιστοιχεί στο μέγιστο κάτω φράγμα και αφαιρεί λευκές εικονοψηφίδες από τα όρια των αντικειμένων. Η διαστολή ενός δυαδικού σήματος A (εικόνας) ισοδυναμεί με την πρόσθεσή του κατά Minkowski με ένα άλλο δυαδικό σήμα B , το δομικό στοιχείο.

Μαθηματικά η πράξη της διαστολής ορίζεται ως εξής:

$$A \oplus B = \cup_{b \in B} A_b \quad (4.7)$$

Η διαστολή του A από το B μπορεί να κατανοηθεί ως το σύνολο των εικονοψηφίδων που καλύπτονται από το B , όταν το κέντρο του κινείται μέσα στην εικόνα A . Για παράδειγμα, το αποτέλεσμα της διαστολής ενός μπλε τετραγώνου με πλευρές 10 εικονοψηφίδων, με ένα κυκλικό δομικό στοιχείο ακτίνας δύο εικονοψηφίδων είναι ένα τετράγωνο με στρογγυλεμένες γωνίες και πλευρές με διαστάσεις 14 εικονοψηφίδες (Εικόνα 4.7).



Εικόνα 4.7: Διαστολή τετραγώνου με κυκλικό δομικό στοιχείο

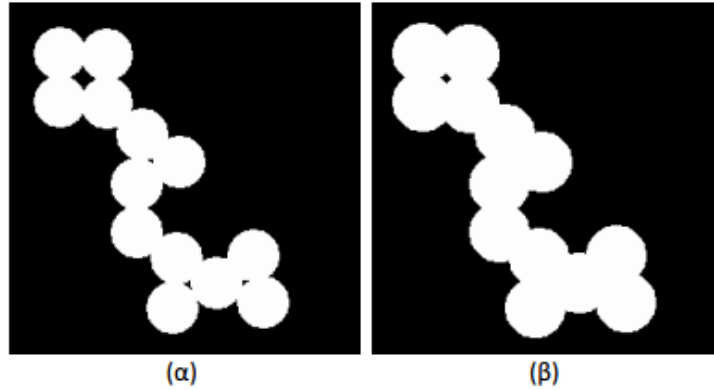
(Πηγή: [http://en.wikipedia.org/wiki/Dilation_\(morphology\)](http://en.wikipedia.org/wiki/Dilation_(morphology)))

Η διαστολή μπορεί ακόμα να οριστεί ως το σύνολο των μετατοπίσεων κατά b , ώστε τα σύνολα A και B^s (συμμετρικό σύνολο του B) να επικαλύπτονται κατά τουλάχιστον ένα μη μηδενικό στοιχείο. Δηλαδή μαθηματικά:

$$A \oplus B = \{z \in E \mid (B^s)_z \cap A \neq \emptyset\} \quad (4.8)$$

όπου $B^s = \{b \in E \mid -b \in B\}$

Αποτελέσματα εφαρμογής της διαστολής σε δυαδική εικόνα εμφανίζονται στην Εικόνα 4.8. Στην αρχική δυαδική εικόνα πραγματοποιήθηκε διαστολή με κυκλικό δομικό στοιχείο ακτίνας 4 εικονοψηφίδων.



Εικόνα 4.8:Αποτέλεσμα διαστολής

(Πηγή:[109])

Στα πλαίσια του αναπτυσσόμενου αλγορίθμου πραγματοποιήθηκαν δοκιμές για κυκλικά και τετραγωνικά δομικά στοιχεία διαστάσεων 3x3, 4x4 και 5x5. Τελικά, επιλέχθηκε το τετραγωνικό δομικό στοιχείο διαστάσεων 3x3, που εξασφάλισε τα καλύτερα αποτελέσματα. Ακόμα, με τη συγκεκριμένη επιλογή μεγέθους η τελική «πλάτυνση» των περιοχών ενδιαφέροντος δε δημιουργεί προβλήματα διαχωρισμού ξεχωριστών πινακίδων που απέχουν πολύ λίγο μεταξύ τους (χαρακτηριστικό παράδειγμα αποτελεί η Εικόνα 4.9).

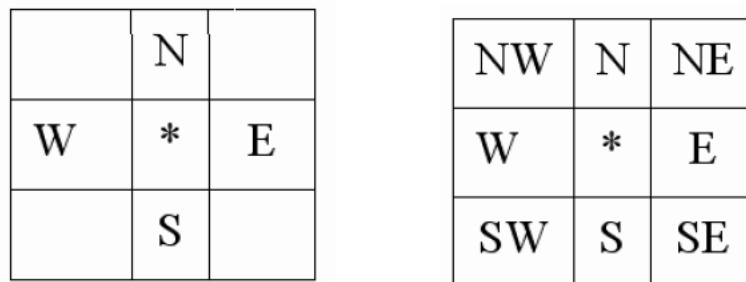


Εικόνα 4.9:Πινακίδες που απέχουν λίγα εκατοστά

4.4 Επιλογή των τελικών περιοχών ενδιαφέροντος

Η συνέχεια του αλγορίθμου περιλαμβάνει την τεχνική Ταυτοποίησης Συνδεδεμένων Μερών (Labeling Connected Components). Στη δυαδική εικόνα, τα αποτελέσματα της κατωφλίωσης συνιστούν τις διάφορες περιοχές ενδιαφέροντος. Για τον ανθρώπινο εγκέφαλο ο διαχωρισμός των περιοχών αυτών είναι εύκολη υπόθεση, για τον υπολογιστή όμως η δυαδική εικόνα αποτελεί μόνο ένα πίνακα που λαμβάνει τιμές 0

και 255. Για να μπορέσει να συγκροτήσει τις διάφορες περιοχές, δηλαδή να εντοπίσει τα αντικείμενα που απεικονίζονται, απαιτείται η ομαδοποίηση των εικονοψηφίδων, η οποία πραγματοποιείται με την τεχνική ταυτοποίησης συνδεδεμένων μερών. Σε μία δυαδική εικόνα δυο εικονοψηφίδες θεωρούνται συνδεδεμένα μέρη εάν υπάρχουν διαδοχικές εικονοψηφίδες που ορίζουν μία διαδρομή, ώστε να καθίσταται δυνατή η μετάβαση από τη μία στην άλλη. Τα διαδοχικά pixel της διαδρομής μπορεί να έχουν μεταξύ τους γειτονία 4 ή 8 pixel [113] (Εικόνα 4.10). Αποτέλεσμα της διαδικασίας είναι μία εικόνα όπου κάθε εικονοψηφίδα φέρει τον κωδικό που αντιστοιχεί στο συνδεδεμένο μέρος στο οποίο ανήκει. Για την αναπαράσταση των κωδικών χρησιμοποιούνται συνήθως θετικοί, ακέραιοι αριθμοί.



Εικόνα 4.10:Γειτονία 4 ή 8 εικονοψηφίδων

(Πηγή:[113])

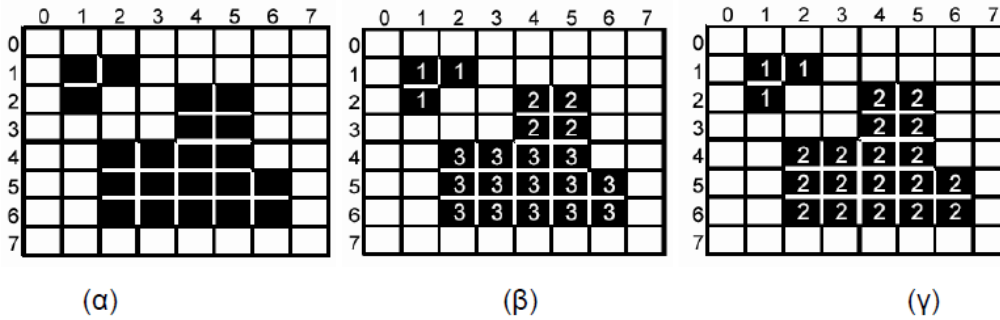
Υπάρχουν διάφοροι αλγόριθμοι για την ταυτοποίηση των συνδεδεμένων μερών όπως αναφέρεται και στο [114]. Εδώ, χρησιμοποιήθηκε ο αλγόριθμος που πραγματοποιεί δύο σαρώσεις στην εικόνα. Πιο συγκεκριμένα, τα βήματα του αλγορίθμου είναι τα ακόλουθα:

1) Η εικόνα σαρώνεται από πάνω προς τα κάτω και από αριστερά προς τα δεξιά. Για κάθε pixel p_i :

- Εξετάζονται οι 4 ή 8 γειτονικές εικονοψηφίδες (Εικόνα 4.10)
- Αν και οι 4 ή 8 εικονοψηφίδες ανήκουν στο υπόβαθρο, τότε το p_i λαμβάνει νέα τιμή συνδεδεμένου συστατικού.
- Αν μία από τις 4 ή 8 γειτονικές εικονοψηφίδες έχει τιμή 255, τότε το p_i λαμβάνει ως τιμή τον κωδικό της συγκεκριμένης εικονοψηφίδας.
- Αν περισσότερες από μία από τις 4 ή 8 γειτονικές εικονοψηφίδες έχουν τιμή 255 και ανήκουν σε συνδεδεμένο μέρος, το p_i λαμβάνει ως τιμή τον κωδικό του συνδεδεμένου μέρους.
- Αν περισσότερες από μία από τις 4 ή 8 γειτονικές εικονοψηφίδες έχουν τιμή 255, το p_i λαμβάνει έναν από τους πιθανούς κωδικούς και σημειώνεται ότι όλοι οι πιθανοί κωδικοί είναι ισοδύναμοι.

2) Η εικόνα σαρώνεται και όλες οι τιμές των συνδεδεμένων συστατικών που έχουν σημειωθεί ως ισοδύναμες αντικαθίσταται από μία κοινή τιμή.

Η εφαρμογή του αλγορίθμου εφαρμόζεται σχηματικά στην Εικόνα 4.11.



Εικόνα 4.11: (α) Αρχική εικόνα, (β) Αποτέλεσμα βήματος 1, (γ) Αποτέλεσμα βήματος 2

(Πηγή: [113])

Στη συνέχεια υπολογίζεται το εμβαδό για κάθε συνδεδεμένο μέρος και απομακρύνονται όσες περιοχές έχουν εμβαδό μικρότερο από 250 εικονοσηφίδες. Με τον τρόπο αυτό απομακρύνονται τόσο αντικείμενα που έχουν εντοπιστεί και συνιστούν θόρυβο, καθώς ενδεχομένως και πινακίδες οδικής κυκλοφορίας που έχουν εντοπιστεί αλλά απεικονίζονται σε πολύ μικρή κλίμακα με αποτέλεσμα να μην είναι δυνατή η αναγνώρισή τους. Ο καθορισμός του ελάχιστου εμβαδού προέκυψε κατόπιν μελέτης των περιοχών ενδιαφέροντος που εντοπίστηκαν στις διάφορες εικόνες συγκεκριμένης ανάλυσης, οι οποίες χρησιμοποιήθηκαν για να εξετασθεί η επιτυχία του αλγορίθμου, όπως αναφέρεται στο κεφάλαιο 5.

Για όσες περιοχές ενδιαφέροντος έγιναν τελικά δεκτές εντοπίζεται το Περιγεγραμμένο Τετράπλευρο (Bounding Box), το οποίο είναι το μικρότερο δυνατό τετράπλευρο που περικλείει ολόκληρη την περιοχή και οι πλευρές του εφάπτονται στα όρια του αντικειμένου που έχει εντοπιστεί. Μία ενδιαφέρουσα ιδιότητα του περιγεγραμμένου τετραπλεύρου είναι ο λόγος των πλευρών του. Στον Πίνακα 4.1 καταγράφονται οι ιδεατοί λόγοι των πλευρών των περιγεγραμμένων τετραπλεύρων για τα τρία βασικά σχήματα των πινακίδων κυκλοφορίας, τον κύκλο, το οκτάγωνο και το ισόπλευρο τρίγωνο:

| Σχήμα | Λόγος πλευρών |
|------------|---------------|
| Οκταγωνικό | 1 |
| Τριγωνικό | 1.1180 |
| Κυκλικό | 1 |

Πίνακας 4.1: Λόγοι πλευρών εφαπτόμενου τετραπλεύρου για διάφορα σχήματα

Συνεπώς, μπορούν να τεθούν ορισμένα όρια στις τιμές των λόγων των τετραπλεύρων που γίνονται τελικά αποδεκτά, προκείμενου να απομακρυνθούν επιπλέον εσφαλμένες περιοχές ενδιαφέροντος. Για τον αλγόριθμο που αναπτύχθηκε, μελετήθηκε ο λόγος της οριζόντιας προς την κατακόρυφη πλευρά του εφαπτόμενου τετραπλεύρου. Το ελάχιστο αποδεκτό όριο τέθηκε ίσο με 0.45 και το μέγιστο αποδεκτό όριο ίσο με 1.6. Τα όρια αυτά δίνουν ένα εύρος τιμών αρκετά μεγάλο, ώστε να μην απομακρυνθούν ως εσφαλμένες πινακίδες που έχουν μερικώς εντοπιστεί λόγω λανθασμένης κατάτμησης. Για παράδειγμα, εάν υπάρχει απόκρυψη στο πάνω μέρος της πινακίδας ή το χρώμα της είναι μερικώς αλλοιωμένο, η περιοχή που θα εντοπιστεί θα έχει σχήμα διαφορετικό από τα παραπάνω και ο λόγος των πλευρών του τετραπλεύρου θα διαφέρει σημαντικά από τις τιμές που σημειώνονται στον πίνακα. Ακόμα, ορισμένες πινακίδες είναι τοποθετημένες με τέτοιο τρόπο ώστε να μη διαχωρίζονται (Εικόνα 4.12). Στην περίπτωση αυτή, ο θεωρητικός λόγος των πλευρών του εφαπτόμενου τετραπλεύρου που περιέχει και τις δύο πινακίδες είναι 0.5. Με τα όρια που τέθηκαν, ακόμα και σε αυτή την περίπτωση η περιοχή ενδιαφέροντος γίνεται αποδεκτή και στη συνέχεια, όπως αναφέρεται και παρακάτω, διαιρείται σε δύο νέες περιοχές, μία για κάθε απεικονιζόμενη πινακίδα. Τέλος, οι περιοχές ενδιαφέροντος με εμβαδό περιγεγραμμένου τετραπλεύρου πάνω από 1000 εικονοψηφίδες δε γίνονται αποδεκτές. Η τιμή αυτή ορίστηκε με βάση τα πειραματικά δεδομένα που μελετήθηκαν και μελέτη των μέγιστων δυνατών διαστάσεων των σημάτων, όταν αυτά απεικονίζονται σε μεγάλη κλίμακα.



Εικόνα 4.12: Επικαλυπτόμενα σήματα οδικής κυκλοφορίας

4.5 Εντοπισμός κύκλων

Για τον πλήρη εντοπισμό των πινακίδων είναι χρήσιμο, εκτός από τον εντοπισμό με βάση το χρώμα, να πραγματοποιηθεί και εντοπισμός με βάση το σχήμα. Για τον λόγο αυτό, στις περιοχές ενδιαφέροντος που έχουν προκύψει κατά το προηγούμενο βήμα και εμφανίζουν λόγο πλευρών περιγεγραμμένου τετραπλεύρου μικρότερου του 0.7 εφαρμόζεται ο κυκλικός μετασχηματισμός Hough. Στόχος της διαδικασίας αυτής είναι να κατατμηθεί η περιοχή ενδιαφέροντος που περιέχει επικαλυπτόμενα σήματα σε δύο νέες περιοχές ενδιαφέροντος, μία για την κάθε πινακίδα. Επιλέγονται για περαιτέρω επεξεργασία μόνο οι περιοχές με λόγο μικρότερου του 0.7, οι οποίες παρουσιάζουν μεγάλη πιθανότητα να απεικονίζουν ένα συνδεδεμένο σήμα. Στις λοιπές περιοχές εν-

διαφέροντος δεν πραγματοποιείται εντοπισμός βάσει σχήματος για λόγους υπολογιστικής ταχύτητας.

4.5.1 Τελεστής ανίχνευσης ακμών (Canny)

Όπως αναφέρθηκε στο κεφάλαιο 2, προαπαιτούμενο για τη χρήση του μετασχηματισμού Hough και των διάφορων παραλλαγών του είναι η εφαρμογή κάποιου τελεστή ανίχνευσης ακμών. Για τον συγκεκριμένο αλγόριθμο χρησιμοποιήθηκε ο τελεστής ανίχνευσης ακμών Canny ([115]).

Ο Canny σημείωσε ότι τα χαρακτηριστικά ενός τελεστή ανίχνευσης ακμών είναι τα εξής:

- Υψηλό ποσοστό ανίχνευσης. Δηλαδή, το ποσοστό μη ανίχνευσης κάποιας ακμής ή εσφαλμένης ανίχνευσης θα πρέπει να είναι ιδιαίτερα μικρό.
- Εντοπισμός της ακριβούς θέσης της ακμής. Δηλαδή, ελαχιστοποίηση της απόστασης μεταξύ των εικονοψηφίδων που ανιχνεύονται από την τελεστή και του κέντρου της αληθούς ακμής.
- Μία απάντηση για μία ακμή, δηλαδή σε κάθε περιοχή ακμών ο αλγόριθμος θα πρέπει να επιστρέφει μία μόνο ακμή.

Για τη διαδικασία ανίχνευσης ακμών αρχικά απαιτείται η συνέλιξη με ένα φίλτρο Gauss. Η εικόνα μετά από τη συνέλιξη εμφανίζεται πιο θολή, αλλά ταυτόχρονα χαρακτηρίζεται από μικρότερο ποσοστό θορύβου. Κατόπιν, εφαρμόζεται ο τελεστής Sobel για την ανίχνευση των κλίσεων κατά τη x και y διεύθυνση. Οι μάσκες με τις οποίες πραγματοποιείται η συνέλιξη είναι:

$$G_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}$$

$$G_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} \quad (4.9a - b)$$

Τελικά, οι απόλυτες τιμές των δύο αποτελεσμάτων προστίθενται για να υπολογιστεί το μέτρο της κλίσης, (gradient magnitude) της κάθε εικονοψηφίδας, το οποία ισούται με:

$$m = \sqrt{G_x^2 + G_y^2} \quad (4.10)$$

καθώς και ο προσανατολισμός (angle of the gradient):

$$\theta = \text{atan}\left(\frac{G_y}{G_x}\right) \quad (4.11)$$

Στη συνέχεια, ο προσανατολισμός αποδίδεται στην πλησιέστερη διεύθυνση που μπορεί να εντοπιστεί στην εικόνα, δηλαδή στις διευθύνσεις των γειτονικών εικονοψηφίδων. Οι τιμές τους είναι ίσες με 0° , 45° , 90° και 135° . Αφού έχει ολοκληρωθεί η απόδοση των διευθύνσεων στην πλησιέστερη από τις 4 παραπάνω κατηγορίες, ακολουθεί η αφαίρεση των μη μεγίστων (non-maximum suppression). Ανάλογα με τη διεύθυνση της κλίσης εξετάζονται οι τρεις γειτονικές εικονοψηφίδες και διατηρούνται μόνο όσες έχουν τη μέγιστη τιμή κατά τη σύγκριση. Με τον τρόπο αυτό απομακρύνονται όσες εικονοψηφίδες δεν αντιστοιχούν σε ακμές. Τέλος, εφαρμόζεται κατωφλίωση υστέρησης (thresholding with hysteresis) με χρήση διπλού ορίου. Όσες εικονοψηφίδες έχουν κλίση μεγαλύτερη από ένα κατώφλι T_1 θεωρείται ότι ανήκουν σε ακμές. Οι εικονοψηφίδες αυτές δεν απομακρύνονται. Το ίδιο ισχύει και για όσες εικονοψηφίδες συνδέονται μαζί τους και έχουν κλίση μεγαλύτερη από ένα κατώφλι T_2 . Η χρήση διπλού κατωφλίου εξασφαλίζει πολύ καλύτερα αποτελέσματα σε σχέση με τη χρήση ενός μοναδικού, καθολικού κατωφλίου, καθώς λόγω θορύβου υπάρχουν εικονοψηφίδες που παρότι ανήκουν σε ακμές, η αριθμητική τιμή της κλίσης τους δεν υπερβαίνει το κατώφλι. Αποτέλεσμα αυτής της αδυναμίας είναι η ύπαρξη ασυνεχειών στις ακμές.

Στην εφαρμογή που πραγματοποιήθηκε το άνω όριο για την κατωφλίωση υστέρησης τέθηκε ίσο με 40 και το κάτω ίσο με 20.

4.5.2 Ο κυκλικός μετασχηματισμός Hough

Όπως περιγράφεται στο κεφάλαιο 2, ο μετασχηματισμός Hough διεξάγεται μέσω μίας διαδικασίας ψηφοφορίας στον χώρο των παραμέτρων. Όσες περισσότερες ψήφους έχει λάβει ένας συνδυασμός τιμών τόσο ισχυρότερες είναι οι ενδείξεις για την ύπαρξη στην εικόνα καμπύλης που χαρακτηρίζεται από τις συγκεκριμένες τιμές. Οι παράμετροι για τις οποίες διεξάγεται η ψηφοφορία ισοδυναμούν με τον αριθμό των μεταβλητών που χρειάζονται προκειμένου να ορισθεί πλήρως μία καμπύλη ([49]).

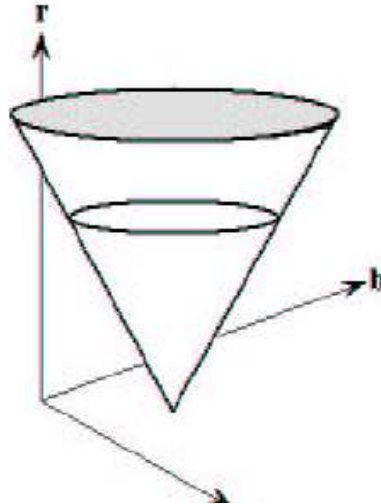
Καθώς η αναλυτική εξίσωση ενός κύκλου είναι:

$$(x - x_o)^2 + (y - y_o)^2 = r^2 \quad (4.12)$$

όπου x_o , y_o οι συντεταγμένες του κέντρου και r η ακτίνα, στη συγκεκριμένη περίπτωση για τον εντοπισμό των κύκλων, ο χώρος των παραμέτρων είναι τρισδιάστατος και αποτελείται από τις παραμέτρους x_o , y_o , r . Κάθε σημείο στην εικόνα απεικονίζεται στον χώρο των παραμέτρων ως κώνος και η τομή πολλαπλών κώνων στο ίδιο σημείο με συντεταγμένες x_{o1} , y_{o1} , r_1 υποδεικνύει την ύπαρξη κύκλου στην εικόνα με κέντρο το x_{o1} , y_{o1} και ακτίνα ίση με r_1 .

Αλγεβρικά, για την εφαρμογή του κυκλικού μετασχηματισμού Hough απαιτείται αρχικά ο ορισμός ενός ελάχιστου κι ενός μέγιστου ορίου της ακτίνας των κύκλων που πρόκειται να εντοπιστούν. Στη συνέχεια, δημιουργείται ένας τρισδιάστατος πίνακας ψηφοφορίας που περιέχει όλα τα πιθανά διαστήματα για τις συντεταγμένες του κέντρου και την ακτίνα. Τα διαστήματα ορίζονται με βάση το μέγιστο αποδεκτό

σφάλμα. Για κάθε εικονοψηφίδα που ανήκει σε ακμή, όπως αυτές έχουν ανιχνευθεί από τον τελεστή Canny, χαράσσονται τόσοι κύκλοι όσα και τα διαστήματα των πιθανών ακτινών. Ταυτόχρονα, για κάθε κύκλο που χαράσσεται οι αντίστοιχες τιμές ακτίνας και κέντρου προσαυξάνονται κατά μια ψήφο στον τελικό πίνακα ψηφοφορίας. Τελικά, ο εντοπισμός των κύκλων πραγματοποιείται για τις τιμές με τις περισσότερες ψήφους.



Εικόνα 4.13: Ο αναπαράσταση ενός σημείου x, y της εικόνας που ανήκει σε κύκλο με κέντρο a, b και ακτίνα r

(Πηγή:[116])

Επειδή η προαναφερθείσα διαδικασία παρουσιάζει μεγάλες απαιτήσεις σε χώρο αποθήκευσης, στο [53] παρατίθεται μία παραλλαγή της μεθόδου που μειώνει κατά πολύ τις απαιτήσεις τόσο στους χρόνους που απαιτούνται για την εκτέλεση του αλγορίθμου όσο και σε μνήμη. Η παραλλαγή αυτή λαμβάνει υπόψη και τις κλίσεις της εικόνας. Πιο συγκεκριμένα, σε κάθε κύκλο το κέντρο εντοπίζεται σε απόσταση ίση με την ακτίνα κατά μήκος της καθέτου στη διεύθυνση στην ακμή. Με τον τρόπο αυτό, η μόνη παράμετρος που πρέπει να εντοπιστεί για να οριστεί πλήρως ένας κύκλος είναι η ακτίνα του.

Η παραλλαγή της μεθόδου μπορεί να εφαρμοστεί γενικά στον μετασχηματισμό Hough για εντοπισμό αναλυτικών καμπυλών, που περιγράφονται πλήρως από ένα πλήθος μεταβλητών a .

Μαθηματικά, με βάση τη δέσμευση:

$$f(x, a) = 0 \quad (4.13)$$

όπου x το σημείο στην εικόνα και a το διάνυσμα των παραμέτρων, ο υπολογισμός των παραμέτρων που απαιτούνται για τον πλήρη ορισμό της καμπύλης μειώνονται κατά μία.

Αντίστοιχα, αν κάποιος εκμεταλλευτεί και την πληροφορία της κλίσης η οποία εκφράζεται μαθηματικά ως:

$$\frac{df}{dx}(\mathbf{x}, \mathbf{a}) = 0 \quad (4.14)$$

ο συνολικός αριθμός των παραμέτρων που απομένει να οριστούν μειώνεται κατά 2. Έτσι, εάν πρέπει να υπολογιστούν συνολικά n παράμετροι, λαμβάνοντας κανείς υπόψη τις παραπάνω δεσμεύσεις, ο αριθμός των παραμέτρων που πρέπει τελικά να υπολογίσει ισούται με $n-2$.

Στον αλγόριθμο που χρησιμοποιήθηκε για τον εντοπισμό των κύκλων στα πλαίσια της διπλωματικής εργασίας χρησιμοποιείται η προϋπάρχουσα γνώση σχετικά με τη διεύθυνση της κλίσης. Ακόμα, το κέντρο του κύκλου μπορεί να εντοπιστεί ουσιαστικά ως το σημείο τομής των κάθετων διευθύνσεων σε κάθε σημείο της ακμής. Ο αλγόριθμος εκμεταλλεύεται και αυτή την ιδιότητα για να μειώσει τις απαιτήσεις του σε υπολογιστική μνήμη και ισχύ ([117]). Χρησιμοποιείται τελικά ένας διδιάστατος πίνακας ψηφοφορίας, που καταγράφει τις πιθανές θέσεις των κέντρων. Η ακτίνα του κύκλου υπολογίζεται ως απόσταση κάθε σημείου ακμής από τα πιθανά κέντρα και δημιουργείται ένα αντίστοιχο ιστόγραμμα συχνοτήτων. Η επιλογή των μεγίστων του ιστογράμματος και του πίνακα ψηφοφορίας καταδεικνύουν τους κύκλους στην εικόνα.

Για τον ορισμό των κατάλληλων παραμέτρων του αλγορίθμου πραγματοποιήθηκαν διάφορες δοκιμές. Τελικά, το μέγιστο αποδεκτό σφάλμα, σύμφωνα με το οποίο υπολογίστηκαν τα διαστήματα, τέθηκε ίσο με την ανάλυση της εικόνας. Επίσης, η ελάχιστη ακτίνα κύκλου που μπορεί να εντοπιστεί ίση με 10 εικονοψηφίδες και η μέγιστη με 200. Τέλος, η ελάχιστη απόσταση μεταξύ των κέντρων δύο κύκλων τέθηκε ίση με το $1/8$ των σειρών της εικόνας που εξετάζεται.

Σε κάθε περιοχή ενδιαφέροντος που προκύπτει από το στάδιο του εντοπισμού, με λόγο πλευρών μικρότερο του 0.7, εφαρμόζεται ο κυκλικός μετασχηματισμός Hough. Εάν δεν εντοπιστεί κύκλος η εικόνα τροφοδοτείται αυτούσια στο στάδιο της αναγνώρισης. Εάν εντοπιστεί κύκλος υπολογίζεται το περιγεγραμμένο τετράπλευρο του κύκλου και ελέγχονται τα εναπομένοντα τμήματα της περιοχής ενδιαφέροντος. Εάν για το ύψος τους h_p ισχύει η σχέση:

$$\frac{h}{2} \leq h_p \leq 2h \quad (4.15)$$

όπου h το ύψος του εφαπτόμενου τετραπλεύρου του κύκλου, τα εναπομένοντα τμήματα θεωρούνται και αυτά περιοχές ενδιαφέροντος και τροφοδοτούνται στη διαδικασία αναγνώρισης.

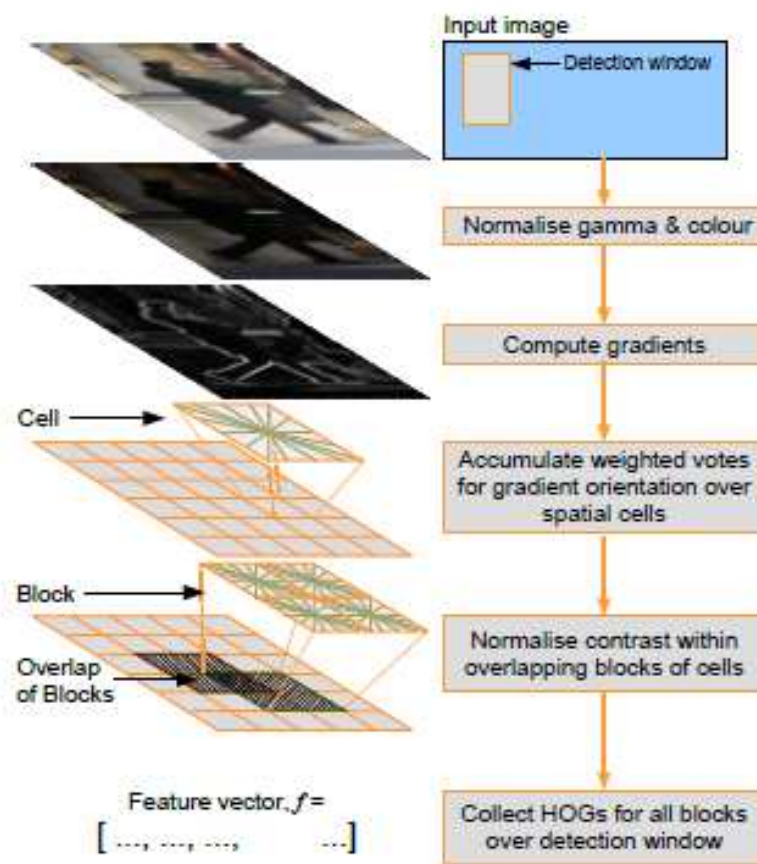
4.6 Περιγραφείς HOG

Η περιγραφή των δεδομένων εκπαίδευσης αποφασίστηκε να πραγματοποιηθεί όχι με βάση τις τιμές φωτεινότητας των εικονοψηφίδων (pixel based approach) αλλά με βάση ορισμένα χαρακτηριστικά (feature based approach) της εκάστοτε εικόνας. Οι περιγραφείς που επιλέχθηκαν είναι οι περιγραφείς HOG (HOGdescriptors – Histogram of Oriented Gradient), οι οποίοι προκύπτουν από συνδυασμό πολλών κανονικοποιημένων ιστογραμμάτων που ποσοτικοποιούν τις τοπικές κλίσεις σε κάθε τμήμα της εικόνας ([118]). Με τη χρήση των συγκεκριμένων περιγραφικών χαρακτηριστικών το σχήμα και η εμφάνιση κάθε απεικονιζόμενου αντικειμένου στην εικόνα περιγράφονται ικανοποιητικά από την κατανομή των τοπικών κλίσεων και από τη διεύθυνσή τους. Τα βήματα που ακολουθούνται για την εξαγωγή των περιγραφικών χαρακτηριστικών είναι:

- Το πρώτο βήμα περιλαμβάνει μία κανονικοποίηση της εικόνας (με κανονικοποίηση γάμμα – gamma normalisation), ώστε να μειωθούν οι επιπτώσεις της φωτεινότητάς της στο διάνυσμα των περιγραφικών χαρακτηριστικών. Εάν η εικόνα είναι έγχρωμη η κανονικοποίηση εφαρμόζεται σε κάθε κανάλι της.
- Στη συνέχεια υπολογίζονται οι κλίσεις της εικόνας (δηλαδή οι πρώτες παράγωγοι κατά τις διευθύνσεις x και y). Η μάσκα που εξασφάλισε τα καλύτερα αποτελέσματα για τον υπολογισμό των κλίσεων είναι η μονοδιάστατη μάσκα $[-1,0,1]$. Οι κλίσεις παρέχουν πληροφορίες σχετικά με την ύπαρξη ακμών και το σχήμα των απεικονιζόμενων αντικειμένων. Σε ορισμένες παραλλαγές του αλγορίθμου, εκτός από τις πρώτες παραγωγούς υπολογίζονται και οι δεύτερες παράγωγοι της εικόνας. Στην περίπτωση έγχρωμων εικόνων υπολογίζονται οι κλίσεις και στα τρία κανάλια και επιλέγεται η μεγαλύτερη τιμή.
- Στο τρίτο βήμα, η εικόνα χωρίζεται σε υποπεριοχές. Η κάθε υποπεριοχή (cell) απαρτίζεται από ένα συγκεκριμένο και προκαθορισμένο αριθμό εικονοψηφίδων και μπορεί να έχει σχήμα ορθογωνικό ή ακτινικό. Για κάθε μία περιοχή δημιουργείται ένα μονοδιάστατο ιστόγραμμα που καταγράφει τις κλίσεις του συνόλου των εικονοψηφίδων που ανήκουν σε αυτή. Η καταγραφή γίνεται σε συγκεκριμένες διευθύνσεις. Ειδικότερα, ανάλογα με τον προκαθορισμένο αριθμό διαστημάτων που χρησιμοποιούνται δημιουργείται ένα ιστόγραμμα με συγκεκριμένους προσανατολισμούς. Το εύρος του ιστογράμματος μπορεί να είναι από 0° - 180° (μη προσημασμένη κλίση: unsigned gradient) ή από 0° - 360° (προσημασμένη κλίση: signed gradient). Για παράδειγμα, εάν ο χώρος 0° - 180° υποδιαιρεθεί ανά 10° , δημιουργείται ένα ιστόγραμμα με δεκαοχτώ πιθανές διευθύνσεις κλίσεων, ενώ αν διαιρεθεί ο χώρος 0° - 360° , δημιουργείται ένα ιστόγραμμα με 36 πιθανές διευθύνσεις. Για κάθε εικονοψηφίδα υπολογίζεται η διεύθυνση (ή και η φορά) της κλίσης, και αυτή ψηφίζει με κάποιο βάρος για το αντίστοιχο διάστημα στο οποίο ανήκει.
- Στο τέταρτο βήμα, τα ιστογράμματα που έχουν προκύψει κανονικοποιούνται για μεγαλύτερη ανοχή των περιγραφικών χαρακτηριστικών σε σκιάσεις και αλλαγές φωτεινότητας. Η κανονικοποίηση πραγματοποιείται με τη αθροιστική

συσσώρευση ενός μέτρου που ποσοτικοποιεί την ενέργεια κάθε τοπικού ιστογράμματος σε ευρύτερες περιοχές (blocks). Οι μεγαλύτερες περιοχές αποτελούνται από έναν προκαθορισμένο αριθμό υποπεριοχών (cells). Κάθε υποπεριοχή που περιγράφεται από ένα τοπικό ιστόγραμμα συμπεριλαμβάνεται σε περισσότερες ευρύτερες περιοχές. Η κανονικοποίηση του ιστογράμματος εξαρτάται από την ευρύτερη περιοχή με βάση την οποία πραγματοποιήθηκε. Συνεπώς, για μία υποπεριοχή που συμμετέχει σε N ευρύτερες περιοχές θα προκύψουν στα συνολικά N κανονικοποιημένα ιστογράμματα, τα οποία διαφέρουν μεταξύ τους. Για την εξαγωγή των περιγραφών HOG χρησιμοποιούνται όλες οι πληροφορίες από το σύνολο των κανονικοποιημένων ιστογραμμάτων, καθώς η προσέγγιση αυτή εξασφαλίζει πιο σταθερά αποτελέσματα.

Σχηματικά, τα βήματα του αλγορίθμου εμφανίζονται στην Εικ. 4.14



Εικόνα 4.14:Εξαγωγή περιγραφών HOG

(Πηγή:[119])

Όπως έχει αναφερθεί και παραπάνω, η ευρύτερη περιοχή μπορεί να έχει σχήμα ορθογωνικό ή ακτινικό. Στην πρώτη περίπτωση, η ευρύτερη περιοχή ονομάζεται R-HOG και ο αλγόριθμος παρουσιάζει πολλές ομοιότητες με τον αλγόριθμο SIFT, όπως αυτός περιγράφηκε στο κεφάλαιο 2. Οι δύο αλγόριθμοι διαφέρουν ως προς τη

συχνότητα εξαγωγής των ιστογραμμάτων, αφού ο SIFT αποδίδει το διάνυσμα των περιγραφικών χαρακτηριστικών που προκύπτει από το ιστόγραμμα μόνο στα σημεία ενδιαφέροντος, ενώ οι περιγραφείς HOG εξάγονται για το σύνολο της εικόνας. Ακόμα, το διάνυσμα των περιγραφικών χαρακτηριστικών στον SIFT στρέφεται κατά την επικρατούσα διεύθυνση του σημείου ενδιαφέροντος, κάτι που δεν ισχύει για τους περιγραφείς HOG. Όσον αφορά στις ευρύτερες, ακτινικές περιοχές, που ονομάζονται C-HOG, για τη χρήση τους υπάρχουν δύο πιθανές επιλογές, όπως φαίνεται στην Εικόνα 4.15. Στην πρώτη περίπτωση, η κεντρική περιοχή είναι ενιαία και υποδιαιρούνται οι κυκλικοί δακτύλιοι, ενώ στη δεύτερη ολόκληρη η ακτινική ευρύτερη περιοχή χωρίζεται σε ίσο αριθμό υποπεριοχών.



Εικόνα 4.15: Οι C-HOG ευρύτερες περιοχές

(Πηγή:[118])

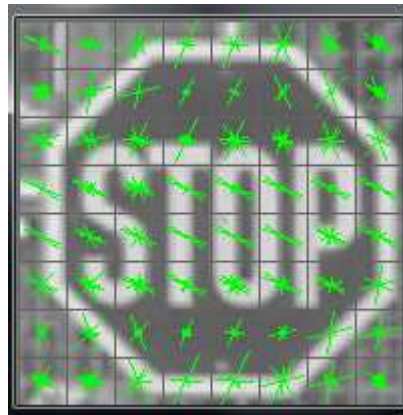
Για την αναπαράσταση των δεδομένων εκπαίδευσης επιλέχθηκαν οι περιγραφείς HOG, καθώς εμφανίζουν μεγάλο αριθμό πλεονεκτημάτων. Καταρχάς, η χρήση ιστογραμμάτων που καταγράφουν τις κλίσεις της εικόνας βοηθά στην ικανοποιητική περιγραφή του σχήματος των απεικονιζόμενων αντικειμένων. Ακόμα, οι περιγραφείς είναι ανθεκτικοί σε πιθανές στροφές και μεταθέσεις, καθώς και αλλαγές φωτεινότητας, όπως προαναφέρθηκε. Δηλαδή, τα περιγραφικά διανύσματα που χαρακτηρίζουν τα ίδια αντικείμενα, ειλημμένα υπό πολύ διαφορετικές συνθήκες, δε διαφέρουν σημαντικά. Τέλος, η δυνατότητα επιλογής του πλήθους των οικονομικών που απαρτίζουν τις περιοχές, του αριθμού των υποπεριοχών (cells) που συμμετέχουν στον σχηματισμό των ευρύτερων περιοχών (blocks) και των προκαθορισμένων διευθύνσεων των ιστογραμμάτων βοηθά στην καταγραφή της πληροφορίας με την «ακρίβεια» που απαιτεί ο χρήστης.

Στον αλγόριθμο που χρησιμοποιήθηκε για την εξαγωγή των περιγραφέων HOG των σημάτων οδικής κυκλοφορίας, η αρχική εικόνα μετατρέπεται σε εικόνα διαστάσεων 32x32, προκειμένου όλα τα δεδομένα εισόδου να έχουν ίδιες διαστάσεις. Η τιμή επιλέχθηκε με μελέτη των μέσων διαστάσεων των σημάτων στην ανάλυση των εικόνων που χρησιμοποιείται για εντοπισμό και ταξινόμησή τους και για την ελάχιστη δυνατή απόσταση λήψης που επιτρέπει την κατηγοριοποίηση των πινακίδων. Κατόπιν, η εικόνα μετατρέπεται σε grayscale και εφαρμόζεται διόρθωση γάμμα. Τέλος, πριν εξαχθούν οι περιγραφείς πραγματοποιείται συνέλιξη με φίλτρο Gauss για την απομάκρυνση του θορύβου. Το παράθυρο αναζήτησης έχει διαστάσεις 32x32, ίδιες με την εικόνα, συνεπώς τοποθετείται σε μία μόνο θέση. Οι υποπεριοχές (cells) έχουν διαστάσεις 4x4 εικονοπηφίδες και οι ευρύτερες περιοχές (blocks) είναι τετραγωνικές και αποτελούνται από 4x4 cells. Το ιστόγραμμα υπολογίζεται σε

συνολικά 6 διευθύνσεις. Για την κανονικοποίηση των διανυσμάτων των περιγραφικών χαρακτηριστικών εφαρμόζεται η κανονικοποίηση L2-Hys. Εάν v το μη κανονικοποιημένο διάνυσμα και v_k η k νόρμα του, το τελικό διάνυσμα προκύπτει από τη σχέση:

$$v = \frac{v}{\sqrt{\|v\|_2^2 + \varepsilon^2}} \quad (4.16)$$

όπου ε μία σταθερά, η οποία εδώ τέθηκε ίση με 0.2. Το προκύπτον διάνυσμα κανονικοποιείται ώστε η μέγιστη τιμή να είναι ίση με 0.2 και υπολογίζεται εκ νέου. Το συνολικό διάνυσμα των περιγραφικών χαρακτηριστικών που προκύπτει έχει διαστάσεις 1x384. Οι περιγραφείς HOG μπορούν και να οπτικοποιηθούν για τη βέλτιστη κατανόησή τους. Παράδειγμα για την πινακίδα του ΣΤΟΠ αποτελεί η Εικόνα 4.16.



Εικόνα 4.16: Περιγραφείς HOG για εικόνα STOP διαστάσεων 64x64. Το κάθε cell ισοδυναμεί εδώ με περιοχή 8x8 εικονοψηφίδων και το κάθε block με 16x16 cells. Τα ιστογράμματα δημιουργούνται σε 9 προκαθορισμένες διευθύνσεις.

4.7 Αναγνώριση

Για το τελικό στάδιο της αναγνώρισης, το περιγραφικό διάνυσμα που έχει εξαχθεί, όπως περιγράφεται αναλυτικά στην ενότητα, εισάγεται σε Μηχανές Διανυσματικής Στήριξης.

Το είδος της ταξινόμησης, που επιλέχθηκε ήταν One Against All, δημιουργήθηκαν συνεπώς N ταξινομητές, ίσοι με τον αριθμό των πινακίδων που πρόκειται να αναγνωρισθούν. Όπως αναφέρεται και στο κεφάλαιο 3, απαραίτητο για τη χρήση SVM είναι το στάδιο της εκπαίδευσης, το οποίο αναλύεται και παρουσιάζεται μαζί με τα δεδομένα εκπαίδευσης και τα αποτελέσματά του στο κεφάλαιο 5. Πραγματοποιήθηκαν δοκιμές με διάφορες συναρτήσεις πυρήνα και μελετήθηκαν στη συνέχεια τα ποσοστά επιτυχίας της αναγνώρισης, ώστε να επιλεγεί η βέλτιστη συνάρτηση. Το λογισμικό που χρησιμοποιήθηκε για την αναγνώριση ήταν το SVMlight [120].

Τα αποτελέσματα του εντοπισμού και της αναγνώρισης παρατίθενται και αναλύονται στο Κεφάλαιο 5.

5

Αποτελέσματα

Ο αλγόριθμος που περιγράφεται στο κεφάλαιο 4 αναπτύχθηκε με χρήση του προγραμματιστικού περιβάλλοντος Matlab και της γλώσσας C++. Ο προγραμματισμός στη C++ πραγματοποιήθηκε στο Microsoft Visual Studio 2010 και χρησιμοποιήθηκαν συναρτήσεις από την βιβλιοθήκη της OpenCV (Open Computer Vision Library) [121]. Η βιβλιοθήκη αυτή διανέμεται ελεύθερα και περιέχει πλήθος συναρτήσεων ανοικτού κώδικα, που στοχεύουν στην επεξεργασία εικόνας. Το project OpenCV ξεκίνησε το 1999 από την Intel και δόθηκε πρώτη φορά για λειτουργία στο ευρύ κοινό στο συνέδριο της IEEE για την Όραση Υπολογιστών και Αναγνώριση Προτύπων (IEEEon Computer Vision and Pattern Recognition), το 2000. Από τότε έχουν κυκλοφορήσει διάφορες νεότερες, βελτιωμένες εκδόσεις και πλέον τη βιβλιοθήκη διαχειρίζεται ένας μη κερδοσκοπικός οργανισμός με το όνομα OpenCV.org. Παρότι οι συναρτήσεις χρησιμοποιούν τη γλώσσα C++ έχουν πλέον αναπτυχθεί διεπαφές ώστε να είναι συμβατή με Matlab, Python και Java [122]. Η έκδοση της OpenCV που χρησιμοποιήθηκε στα πλαίσια της συγκεκριμένης διπλωματικής εργασίας είναι η 2.4.3

Ο αλγόριθμος που συντάχθηκε εφαρμόστηκε σε πραγματικά δεδομένα, τα οποία συλλέχθηκαν υπό πολύ διαφορετικές καιρικές συνθήκες και συνθήκες φωτεινότητας, προκειμένου να εξεταστεί πλήρως η λειτουργία του. Τα πραγματικά δεδομένα αποτελούνται από εικόνες που ελήφθησαν με τη φωτογραφική μηχανή Olympus μ 850 SW με αισθητήρα ανάλυσης 8MP και οπτικό zoom 6.7-20.1mm, η οποία τοποθετήθηκε σε κινούμενο όχημα.



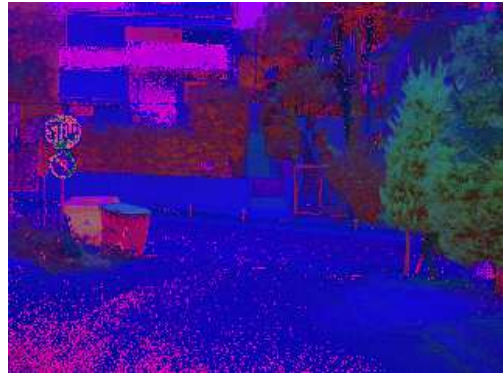
Εικόνα 5.1: Η φωτογραφική μηχανή που χρησιμοποιήθηκε

Ελήφθησαν εικόνες διαστάσεων 2560x1920 εικονοψηφίδων, οι οποίες μειώνονται από τον αλγόριθμο που συντάχθηκε, με στόχο τη γρηγορότερη εκτέλεσή του. Οι διαστάσεις των τελικών εικόνων είναι 680x510 εικονοψηφίδες, οι οποίες επεξεργάζονται κατάλληλα, προκειμένου να εντοπιστούν και αναγνωριστούν οι πινακίδες οδικής κυκλοφορίας. Τα αποτελέσματα παρουσιάζονται και αναλύονται στη συνέχεια του κεφαλαίου για το σύνολο των 42 εικόνων που χρησιμοποιήθηκαν προκειμένου να ελεγχθεί η επιτυχία της εφαρμογής που αναπτύχθηκε. Οι εικόνες ελήφθησαν κατά τη διάρκεια ημέρας και νύχτας, τόσο σε μέρες με ηλιοφάνεια, όπως και σε μέρες με βροχή και συννεφιά, για να καλυφθούν κατά το δυνατόν περισσότερες συνθήκες φωτεινότητας των λήψεων. Οι εικόνες που ελήφθησαν απεικονίζουν κατά το μεγαλύτερο μέρος πινακίδες ΣΤΟΠ και πινακίδες που υποδηλώνουν την ύπαρξη ποδηλατόδρομου, καθώς εκπαιδεύθηκαν συνολικά δύο ταξινομητές, για την αναγνώριση αυτών των δύο πινακίδων. Τα λοιπά σήματα που περιέχονταν στις εικόνες που συλλέχθηκαν, χρησιμοποιήθηκαν ως αρνητικά παραδείγματα για την εκπαίδευση των ταξινομητών, όπως αναφέρεται και στη συνέχεια του κεφαλαίου.

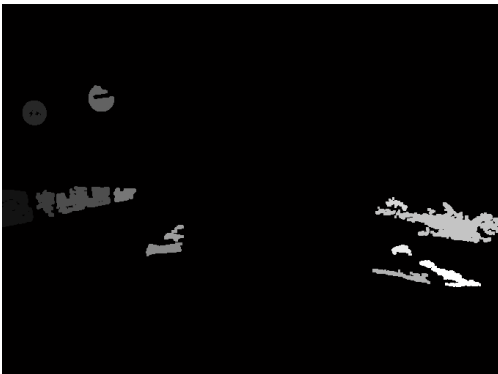
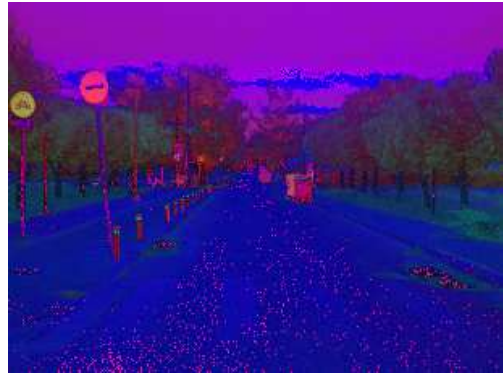
5.1 Αποτελέσματα εντοπισμού

5.1.1 Αποτελέσματα εντοπισμού με βάση το χρώμα

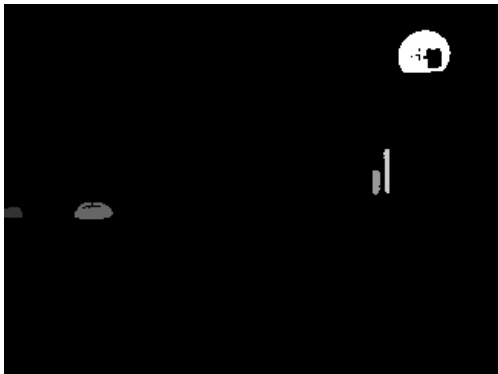
Για εντοπισμό με βάση το χρώμα πραγματοποιήθηκε η ροή των εργασιών, όπως αυτή περιγράφεται στο κεφάλαιο 4. Για την τελική επιλογή των καθολικών παραμέτρων, που εξασφαλίζουν τα βέλτιστα δυνατά αποτελέσματα στο σύνολο όλων των εικόνων, πραγματοποιήθηκαν διάφορες δοκιμές. Ακολουθώντας στην Εικόνα 5.2 παρατίθενται τέσσερα (4) χαρακτηριστικά αποτελέσματα εντοπισμού με βάση το χρώμα, με τις τιμές ορίων που αναφέρονται στο κεφάλαιο 4. Το σύνολο των αποτελεσμάτων για τις 42 εικόνες δίνεται στο Παράρτημα Α. Για κάθε εικόνα εμφανίζεται η μετατροπή στον χρωματικό χώρο HSI, τα αποτελέσματα της καταφλίωσης, τα αποτελέσματα εφαρμογής του φίλτρου μεσαίας τιμής και διαστολής με δομικό στοιχείο με σχήμα τετράγωνο, οι συνδεδεμένες περιοχές κάθε μία με διαφορετικό χρώμα, και οι περιοχές ενδιαφέροντος που έγιναν τελικά δεκτές με βάση το εμβαδό τους και τον λόγο του επαπτόμενου τετραπλεύρου. Οι περιοχές ενδιαφέροντος εμφανίζονταν όπως αυτές αποκόπηκαν από την αρχική εικόνα και θα τροφοδοτηθούν στα επόμενα βήματα του αλγορίθμου.



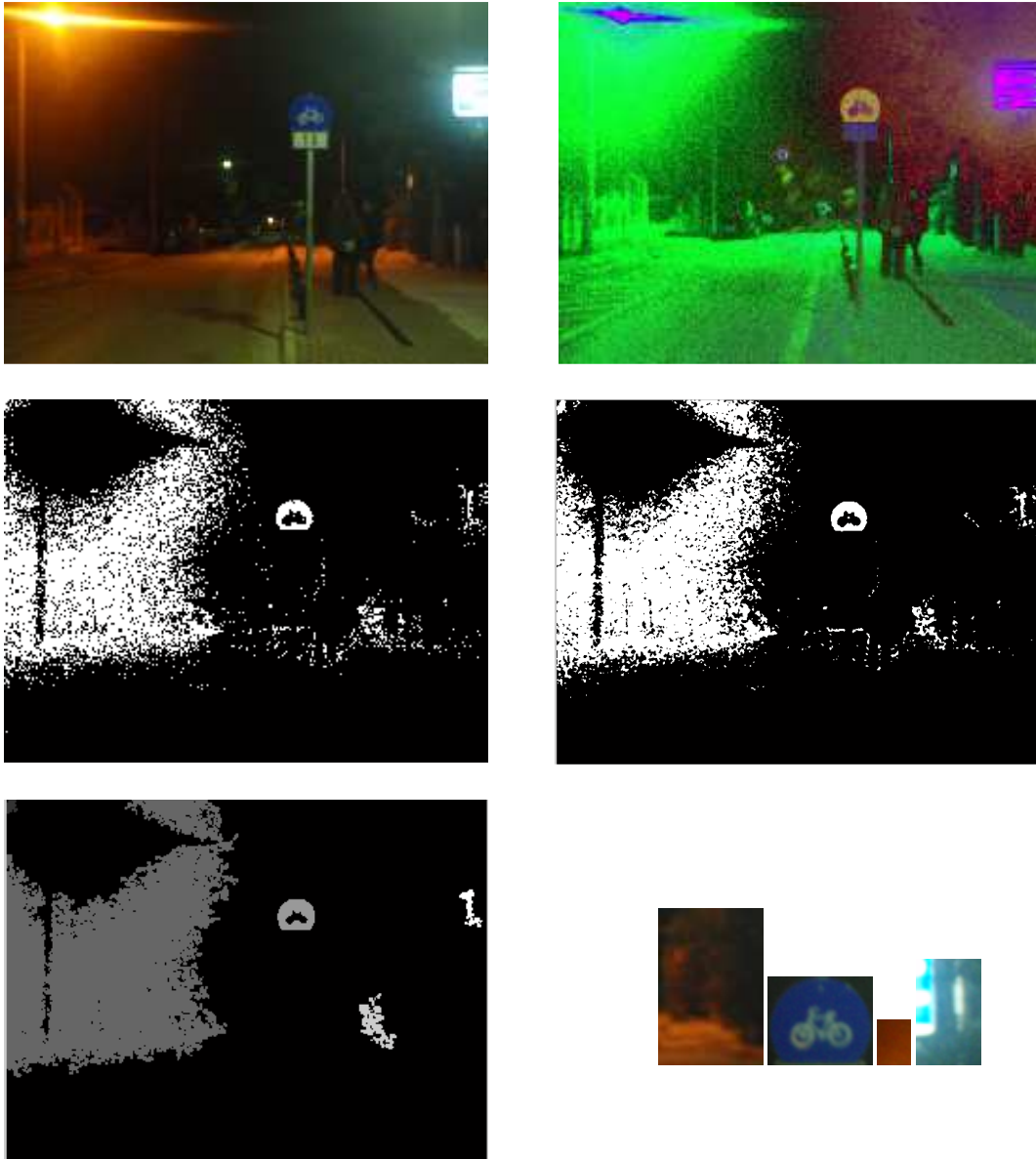
(α) Αποτελέσματα για εικόνα 1



(β) Αποτελέσματα για εικόνα 2



(γ) Αποτελέσματα για εικόνα 8



(δ) Αποτελέσματα για εικόνα 19

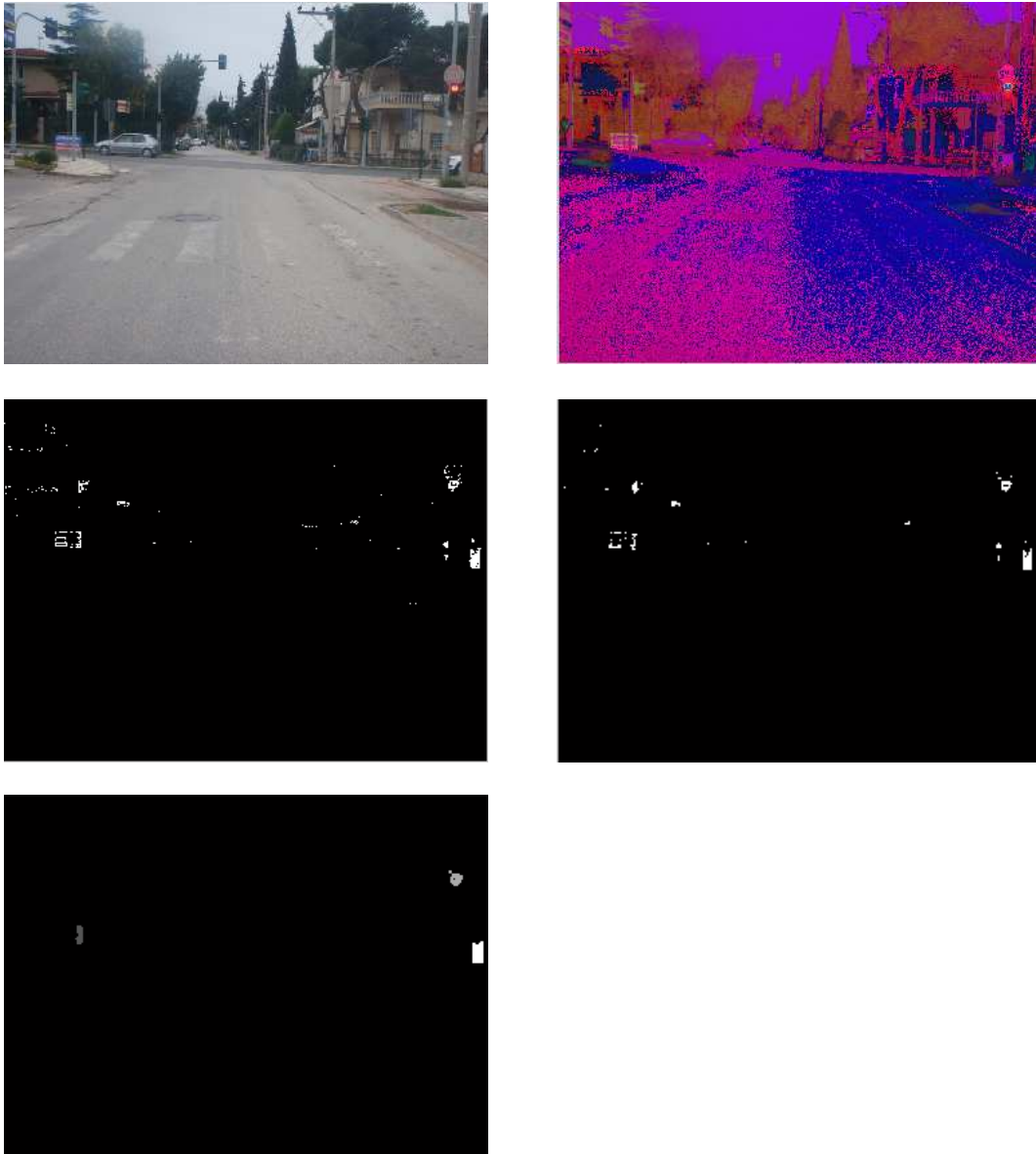
Εικόνα 5.2: Χαρακτηριστικά παραδείγματα αποτελεσμάτων εντοπισμού

Στα παραπάνω ενδεικτικά παραδείγματα που απεικονίζονται σήματα κατά τη διάρκεια της ημέρας και νύχτας αλλά και σε αρκετά διαφορετικές καιρικές συνθήκες, η απόδοση του αλγορίθμου είναι αρκετά ικανοποιητική καθώς εντοπίστηκαν όλες οι πινακίδες. Ακολουθώς αναφέρονται τρία παραδείγματα που ο αλγόριθμος απέτυχε στον εντοπισμό των σημάτων.



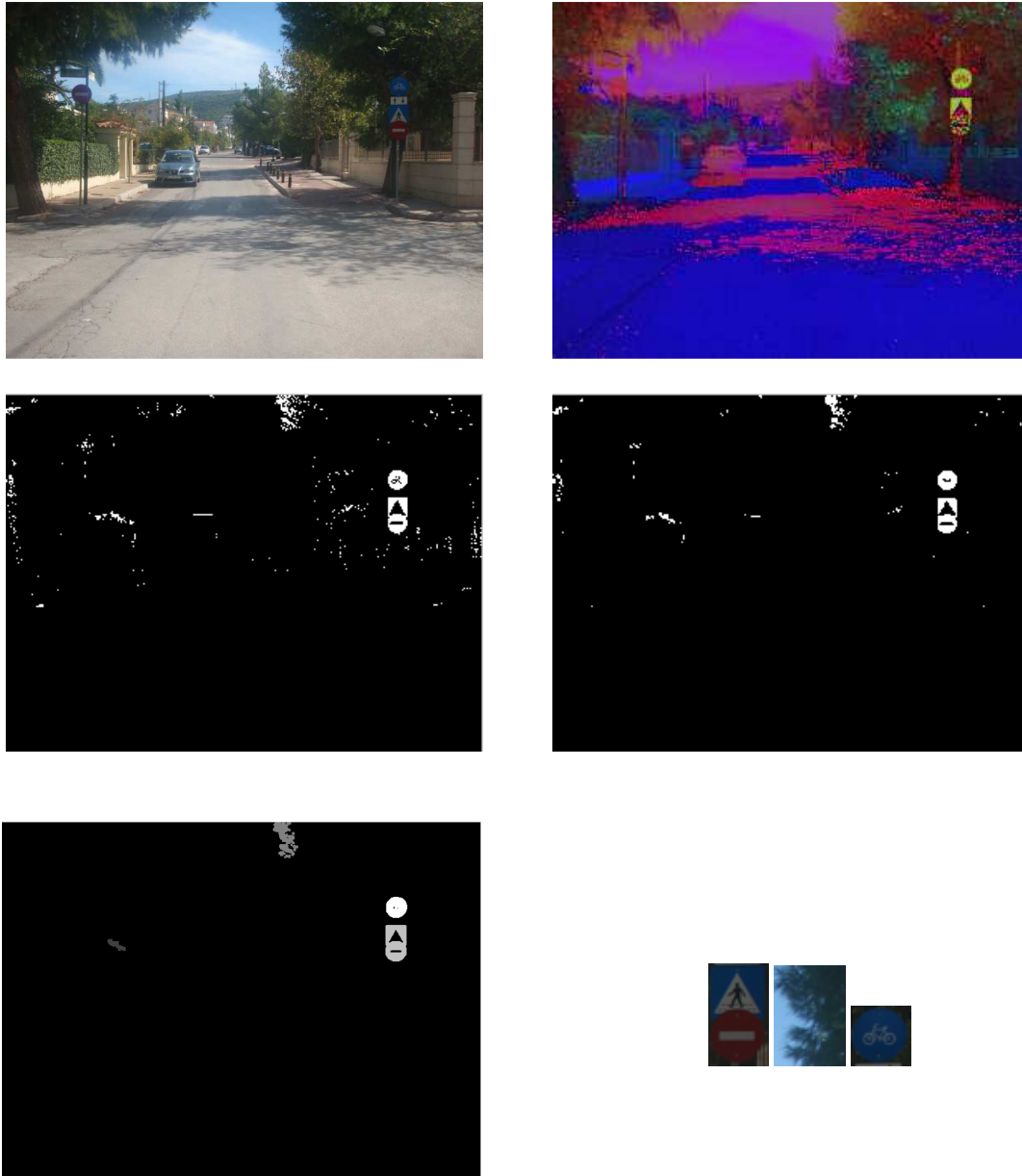
Εικόνα 5.3: Περίπτωση αποτυχίας εντοπισμού του σήματος

Στην περίπτωση της Εικόνα 5.3 από τις δύο πινακίδες εντοπίστηκε τελικά, μόνο η πινακίδα του ΣΤΟΠ και όχι η δεύτερη που υποδηλώνει την απαγόρευση στροφής στα δεξιά, λόγω του κόκκινου υποβάθρου της εικόνας πίσω από την πινακίδα. Η ύπαρξή του έχει ως αποτέλεσμα την επιλογή όχι μόνο του σήματος αλλά και του περιβάλλοντα χώρου της εικόνας ως περιοχή ενδιαφέροντος. Τελικά δημιουργείται ένα συνδεδεμένο μέρος, το οποίο στο τελικό κομμάτι του εντοπισμού απορρίπτεται λόγω του μεγάλου εμβαδού του. Η ύπαρξη παρόμοιων χρωμάτων με τα χρώματα των πινακίδων στο υπόβαθρο της εικόνας συνιστά καίριο πρόβλημα για τη δημιουργία συστημάτων αυτόματου εντοπισμού πινακίδων οδικής κυκλοφορίας, όπως επισημαίνεται κατά καιρούς σε πολλές αντίστοιχες ερευνητικές εργασίες [20].



Εικόνα 5.4: Περίπτωση αποτυχίας εντοπισμού του σήματος

Στην περίπτωση της Εικόνα 5.4 ο αλγόριθμος απέτυχε να εντοπίσει το σήμα λόγω του αλλοιωμένου χρώματός του. Η αλλοίωση ήταν τόσο έντονη που η πινακίδα δεν μπόρεσε να ανιχνευθεί κατά το πρώτο στάδιο της κατωφλίωσης.



Εικόνα 5.5: Περίπτωση αποτυχίας εντοπισμού των σημάτων

Η τρίτη περίπτωση (Εικόνα 5.5) που ο αλγόριθμος δεν ανίχνευσε την πινακίδα, που υποδηλώνει την υπαγόρευση της διέλευσης, οφείλεται σε ελλιπή φωτισμό σε αυτό το κομμάτι της λήψης.

Οι εσφαλμένες περιοχές ενδιαφέροντος που εξήχθησαν απεικονίζουν, σε μεγάλο ποσοστό, φανάρια αυτοκινήτων, κόκκινες περιφράξεις και σκεπές, καθώς και διαφημιστικές πινακίδες με τα προς εντοπισμό χρώματα. Οι περιοχές ενδιαφέροντος που περιέχουν κάποιο σήμα οδικής κυκλοφορίας, στην μεγαλύτερη πλειοψηφία περικλείουν ικανοποιητικά την πινακίδα, εκτός λίγων εξαιρέσεων. Χαρακτηριστικά παραδείγματα αποτελούν οι περιπτώσεις της Εικόνας 5.6 λόγω αποκρύψεων στην αρχική λήψη (για την πρώτη περίπτωση) και ύπαρξη κόκκινου χρώματος στο υπόβαθρο της αρχικής λήψης (δεύτερη περίπτωση).



Εικόνα 5.6: Μη ικανοποιητικός εντοπισμός

Για το σύνολο των 42 εικόνων εντοπίστηκαν, συνολικά, 145 περιοχές ενδιαφέροντος, που περιείχαν 59 πινακίδες οδικής κυκλοφορίας. Στο σύνολο των πειραματικών δεδομένων εντοπίζονται συνολικά 62 σήματα με μέγεθος ικανό να ανιχνευτούν και να αναγνωριστούν. Συνεπώς, ο αλγόριθμος του εντοπισμού λειτούργησε με ποσοστό επιτυχίας 95%.

5.1.2 Αποτελέσματα εντοπισμού με βάση το σχήμα

Όπως αναφέρθηκε στο κεφάλαιο 4, στην περίπτωση όπου εντοπιστούν επικαλυπτόμενα σήματα, επιχειρείται ο διαχωρισμός τους. Πιο συγκεκριμένα, εάν ο λόγος της οριζοντίου προς την κάθετη πλευρά της περιοχής ενδιαφέροντος είναι μικρότερος του 0.7 στην περιοχή εφαρμόζεται ένας κυκλικός μετασχηματισμός Hough και, εάν εντοπιστεί κάποιος κύκλος, δημιουργείται μία νέα περιοχή ενδιαφέροντος με διαστάσεις ίσες με το περιγεγραμμένο τετράγωνο του κύκλου. Η περιοχή αυτή αποκόπτεται από την αρχική περιοχή ενδιαφέροντος και εάν οι εναπομείνουσες περιοχές έχουν ύψος τουλάχιστον ίσο με το μισό της περιοχής που αποκόπηκε, οι περιοχές αυτές ορίζονται ως νέες περιοχές ενδιαφέροντος. Εάν δεν εντοπιστεί κύκλος στην περιοχή ενδιαφέροντος αυτή τροφοδοτείται χωρίς μεταβολή στο στάδιο της αναγνώρισης. Ο κυκλικός μετασχηματισμός Hough εφαρμόζεται με χρήση της συνάρτησης της OpenCV Hough Circles. Η συνάρτηση λαμβάνει ως ορίσματα την εικόνα με τους τόνους του γκριζου, που θα εφαρμοστεί ο μετασχηματισμός του Hough, ένα διάνυσμα με το όνομα Circles στο οποίο θα αποθηκευτούν οι τιμές για το κέντρο και την ακτίνα των κύκλων που εντοπίστηκαν, και την παράμετρο CV_HOUGH_GRADIENT που ορίζει τον ακριβή αλγόριθμο με τον οποίο θα πραγματοποιηθεί ο εντοπισμός (περιγράφεται αναλυτικά στο Κεφάλαιο 4). Άλλες παράμετροι που πρέπει να οριστούν είναι η `dp`, που συνιστά τον λόγο της ανάλυσης της εικόνας και η `min_dist` που αποτελεί την ελάχιστη απόσταση μεταξύ δύο κύκλων που θα εντοπιστούν. Τέλος, ο αλγόριθμος δέχεται ως όρισμα τις τιμές `param1` και `param2`, που είναι η ελάχιστη και μέγιστη τιμή κατωφλίου για την εφαρμογή του τελεστή ανίχνευσης ακμών Canny, και τις παραμέτρους `min_radius` και `max_radius` για τις αναμενόμενες ελάχιστες και μέγιστες ακτίνες των κύκλων που θα εντοπιστούν.

Στην Εικόνα 5.7 εμφανίζεται μία πινακίδα ΣΤΟΠ και μία πινακίδα απαγόρευσης στροφής που είναι επικαλυπτόμενες. Με κόκκινο χρώμα σημειώνεται ο κύκλος που εντοπίστηκε και επίσης εμφανίζονται οι δύο περιοχές ενδιαφέροντος που προέκυψαν.



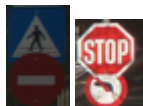
Εικόνα 5.7: Επιτυχής διαχωρισμός επικαλυπτόμενων σημάτων

Αντίστοιχα, στην Εικόνα 5.8 εμφανίζονται οι αρχικές και οι τελικές εικόνες που προέκυψαν για όλες τις περιπτώσεις που παρουσιάστηκαν στα δεδομένα.



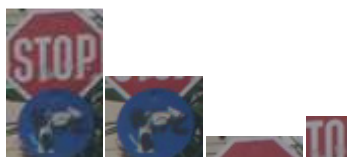
Εικόνα 5.8: Αποτελέσματα διαχωρισμού επικαλυπτόμενων σημάτων

Στην τελευταία περίπτωση η εφαρμογή του βήματος είναι ιδιαίτερα σημαντική καθώς όχι μόνο διαχώρισε τις δύο πινακίδες σε δύο περιοχές ενδιαφέροντος, αλλά ταυτόχρονα απομάκρυνε και την πληροφορία του υποβάθρου, που εσφαλμένα περιέχεται στην περιοχή ενδιαφέροντος. Υπήρχαν, όμως, περιπτώσεις στις οποίες το συγκεκριμένο βήμα δεν απέφερε τα επιθυμητά αποτελέσματα. Στις δύο περιπτώσεις, που φαίνονται στην Εικόνα 5.9, δεν εντοπίστηκε κάποιος κύκλος με αποτέλεσμα να μην τμηθεί η περιοχή ενδιαφέροντος και να τροφοδοτηθεί αυτούσια στο στάδιο της αναγνώρισης. Η κυριότερη αιτία που οδήγησε στην αποτυχία της μεθόδου στα συγκεκριμένα δύο παραδείγματα είναι η μη ικανοποιητική εικόνα ακμών που εξάγεται.



Εικόνα 5.9: Ανεπιτυχής διαχωρισμός επικαλυπτόμενων σημάτων

Τέλος, σε μία τρίτη περίπτωση (Εικόνα 5.10) ο κύκλος που εντοπίστηκε τελικά είναι λανθασμένος, συνεπώς και η κατάτμηση που τελικά προέκυψε δεν είναι ικανοποιητική.



Εικόνα 5.10: Ανεπιτυχής διαχωρισμός επικαλυπτόμενων σημάτων

Από τις δέκα περιπτώσεις που εφαρμόστηκε το συγκεκριμένο βήμα, είχε επιτυχή αποτελέσματα στις επτά. Η εφαρμογή του όμως είναι εξαιρετικά σημαντική, καθώς η τροφοδότηση στο στάδιο της αναγνώρισης αυτούσιας της περιοχής των επικαλυπτόμενων σημάτων οδηγεί σε ανεπιτυχή αποτελέσματα ταξινόμησης. Ο τελικός αριθμός των περιοχών ενδιαφέροντος που εισήχθη στο τελικό στάδιο της αναγνώρισης μετά το πέρας του συγκεκριμένου σταδίου ισούται με 156.

5.2 Αναγνώριση

5.2.1 Εκπαίδευση

Για την εκπαίδευση των SVM χρησιμοποιήθηκε το πρόγραμμα SVMight (βλ. παρακάτω). Πραγματοποιήθηκαν διάφορες δοκιμές με διαφορετικό αριθμό δεδομένων εκπαίδευσης και με διαφορετικές συναρτήσεις πυρήνων και η δυνατότητα γενίκευσης κάθε εκπαιδευμένου συστήματος ποσοτικοποιήθηκε με τους κατάλληλους δείκτες που παρέχει το λογισμικό.

Αρχικό στάδιο για την τροφοδότηση των δεδομένων εκπαίδευσης στο σύστημα συνιστά η αναπαράσταση της εικόνας με τους περιγραφείς HOG (βλ. Κεφάλαιο 4).

5.2.1.1 Εξαγωγή περιγραφικών χαρακτηριστικών

Αρχικά, για τα δεδομένα εκπαίδευσης εξήχθησαν οι περιγραφείς HOG με χρήση της κλάσης της OpenCV HOGDescriptor. Για τη χρήση της συνάρτησης μπορεί κανείς να επιλέξει τις παραμέτρους του περιγραφέα που επιθυμεί, οι οποίες είναι οι εξής:

- το μέγεθος του παραθύρου που τοποθετείται διαδοχικά στην εικόνα (win_size),
- το μέγεθος των ευρύτερων περιοχών (block_size),
- το βήμα για τη δημιουργία των ευρύτερων περιοχών (block_stride),
- το μέγεθος των υποπεριοχών (cell_size),
- ο αριθμός των διευθύνσεων του προσανατολισμού στο ιστόγραμμα (bin),
- το σ του Γκαουσιάνου φίλτρου με το οποίο θα πραγματοποιηθεί συνέλιξη (win_sigma) και
- ένα κατώφλι για την κανονικοποίηση του διανύσματος των περιγραφέων (threshold_L2hys).

Επίσης ο χρήστης καλείται να επιλέξει εάν θα πραγματοποιηθεί ή όχι γάμμα-διόρθωση (gamma_correction) και τον μέγιστο αριθμό ανίχνευσης (nlevels). Οι τιμές των παραμέτρων που χρησιμοποιήθηκαν στην εφαρμογή αναφέρονται λεπτομερώς στο κεφάλαιο 4.

Στη συνέχεια, με χρήση της συνάρτησης getDescriptors, που ανήκει στην ίδια κλάση, υπολογίζεται το διάνυσμα των περιγραφικών χαρακτηριστικών. Η συνάρτηση λαμβάνει ως ορίσματα την εικόνα (img), ένα διάνυσμα με το όνομα descriptors στο οποίο θα αποθηκευτεί το διάνυσμα των περιγραφικών χαρακτηριστικών και ένα διάνυσμα με το όνομα locations, στο οποίο αποθηκεύσει τις τοποθεσίες της εικόνας για τις οποίες υπολογίστηκαν οι περιγραφείς. Τέλος, λαμβάνει ως ορίσματα το winStride και το padding, που ορίζει πόσες επιπλέον εικονοψηφίδες χρειάζονται στα όρια της εικόνας, για τον υπολογισμό των περιγραφικών διανυσμάτων. Για κάθε εικόνα που ανήκει στα δεδομένα εκπαίδευσης εξάγεται ένα διάνυσμα περιγραφικών χαρακτηριστικών.

5.2.1.2 Τα δεδομένα εκπαίδευσης

Για την εκπαίδευση του συστήματος χρησιμοποιήθηκαν πραγματικά δεδομένα. Ελήφθησαν εικόνες των σημάτων με διαφορετικές συνθήκες φωτισμού και καιρού (συννεφιά – βροχή – ηλιοφάνεια). Από το σύνολο των εικόνων αποκόπηκαν χειροκίνητα οι πινακίδες που πρόκειται να ταξινομηθούν και χρησιμοποιήθηκαν ως θετικά παραδείγματα. Η ταξινόμηση πραγματοποιήθηκε με τη μέθοδο one versus all, συνεπώς αποκόπηκαν και εικόνες άλλων σημάτων, καθώς και περιοχών που δεν περιείχαν σήματα, οι οποίες χρησιμοποιήθηκαν ως αρνητικά παραδείγματα. Για το κυκλικό σήμα του ποδηλάτου υπήρξαν διαθέσιμα 134 θετικά και 83 αρνητικά παραδείγματα. Ενώ για την πινακίδα του ΣΤΟΠ χρησιμοποιήθηκαν 124 θετικά και 83 αρνητικά παραδείγματα. Χαρακτηριστικές εικόνες που αποτελούν δεδομένα εκπαίδευσης για την πινακίδα ΣΤΟΠ και την πινακίδα ύπαρξης ποδηλατοδρόμου παρουσιάζονται στις εικόνες 5.11 και 5.12 αντίστοιχα.



Εικόνα 5.11:Θετικά (επάνω) και αρνητικά (κάτω) παραδείγματα για την εκπαίδευση ταξινομητή για την πινακίδα ΣΤΟΠ



Εικόνα 5.12:Θετικά (επάνω) και αρνητικά (κάτω) παραδείγματα για την εκπαίδευση ταξινομητή για την πινακίδα σήμανσης ποδηλατοδρόμου

Για όλα τα δεδομένα εκπαίδευσης εξήχθησαν οι περιγραφείς HOG, σύμφωνα με τα παραπάνω, και σημειώθηκαν οι αντίστοιχες αποκρίσεις ανάλογα με την κατηγορία στην οποία ανήκουν.

5.2.1.3 Το λογισμικό SVMlight

Για την εκπαίδευση του συστήματος χρησιμοποιήθηκε το λογισμικό SVMlight [120]. Το συγκεκριμένο λογισμικό είναι γραμμένο σε γλώσσα προγραμματισμού C και χρησιμοποιείται για προβλήματα ταξινόμησης και παλινδρόμησης. Δίνει τη δυνατότητα εφαρμογής πολλών συναρτήσεων πυρήνα και επιτρέπει την εισαγωγή νέων συναρτήσεων που έχουν δημιουργηθεί από τον χρήστη. Επίσης, παρέχει μέτρα για την επιτυχία της κάθε μεθόδου, τα οποία αναλύονται στη συνέχεια του κεφαλαίου. Το SVMlight μπορεί να διαχειριστεί γρήγορα και αποτελεσματικά μεγάλο αριθμό διανυσμάτων στήριξης και δεδομένων εκπαίδευσης.

5.2.1.4 Προετοιμασία δεδομένων εκπαίδευσης

Για να εισαχθούν τα δεδομένα εκπαίδευσης στο πρόγραμμα SVMlight θα πρέπει να έχουν την ακόλουθη μορφή:

- 1 1:feat1 2:feat2 n:feat3 για δεδομένα που ανήκουν στην κατηγορία 1
 -1 1:feat1 2:feat2 n:feat3 για δεδομένα που ανήκουν στην κατηγορία -1

Συνεπώς για κάθε σύνολο δεδομένων δημιουργήθηκε αρχικά ένας πίνακας με διαστάσεις $m \times n$

όπου: m ο αριθμός των δεδομένων εκπαίδευσης και

n οι διαστάσεις του διανύσματος των περιγραφικών χαρακτηριστικών, δηλαδή 384 στη συγκεκριμένη περίπτωση.

Κατασκευάστηκαν έξι σύνολα δεδομένων εκπαίδευσης. Τρία για κάθε πινακίδα που πρόκειται να εντοπιστεί. Για τα δεδομένα του ΣΤΟΠ:

- το πρώτο σύνολο απαρτίζεται από 51 θετικά παραδείγματα και 25 αρνητικά, συνεπώς ο πίνακας έχει διαστάσεις 76×384
- το δεύτερο σύνολο δεδομένων περιέχει 92 θετικά παραδείγματα και 57 αρνητικά, άρα ο πίνακας που χρησιμοποιήθηκε για την εκπαίδευση έχει διαστάσεις 149×384
- το τρίτο σύνολο δεδομένων περιέχει 124 θετικά παραδείγματα και 83 αρνητικά και ο τελικός πίνακας έχει διαστάσεις 207×384 .

Για την περίπτωση του σήματος ποδηλάτου:

- το πρώτο σύνολο δεδομένων περιέχει 45 θετικά παραδείγματα και 26 αρνητικά, συνεπώς δημιουργείται πίνακας με διαστάσεις 71×384
- το δεύτερο σύνολο δεδομένων αποτελείται από 90 θετικά παραδείγματα και 60 αρνητικά, άρα ο πίνακας που περιέχει το σύνολο των δεδομένων είναι 150×384 και
- το τρίτο σύνολο δεδομένων περιέχει 134 θετικά παραδείγματα και 83 αρνητικά, συνεπώς ο πίνακας που τελικά δημιουργείται έχει διαστάσεις 217×384 .

5.2.1.5 Αποτελέσματα εκπαίδευσης

Όταν ζητηθεί από το SVMlight να κατασκευάσει μία SVM η οποία λειτουργεί ως δυαδικός ταξινομητής, ο χρήστης καλείται να ορίσει τη συνάρτηση πυρήνα $K(\mathbf{x})$ που θα χρησιμοποιηθεί και να καθορίσει τις τιμές των μεταβλητών που απαιτούνται, δηλαδή το C και όποιες άλλες τιμές απαιτούνται ανάλογα με τη συνάρτηση πυρήνα που χρησιμοποιείται. Εάν ο χρήστης δεν ορίσει τις τιμές, το λογισμικό τις θέτει ίσες με προκαθορισμένες τιμές (default values) και τελικά εκτιμά την ικανότητα γενίκευσης του με υπολογισμό ορισμένων δεικτών, οι οποίοι είναι οι εξής [123]:

- Ποσοστό σφάλματος (error rate). Για εκπαίδευση με χαλαρό περιθώριο το ποσοστό σφάλματος μαθηματικά ισούται με:

$$Err_{\xi\alpha}^n(h_L) = \frac{d}{n} \quad (5.1)$$

$$\text{με } d = |\{i: (\rho\alpha_i R_\Delta^2 + \xi_i \geq 1)\}|$$

όπου: ρ μία μεταβλητή που παίρνει συνήθως την τιμή 2,

α_i οι πολλαπλασιαστές Lagrange,

ξ_i οι ψευδομεταβλητές όπως ορίζονται από το δυϊκό πρόβλημα.

Το R_{Δ}^2 είναι το άνω όριο στο $K(\mathbf{x}, \mathbf{x}) - K(\mathbf{x}, \mathbf{x}') \forall \mathbf{x}, \mathbf{x}'$ και για τη γραμμική συνάρτηση πυρήνα ισχύει $R_{\Delta}^2=1$.

Το ποσοστό σφάλματος αποτελεί τον λόγο των παραδειγμάτων d για τα οποία ισχύει η ανισότητα ($\rho\alpha_i R_{\Delta}^2 + \xi_i \geq 1$), προς το σύνολο των παραδειγμάτων. Για κάθε παράδειγμα που ταξινομείται εσφαλμένα από ένα σύστημα που έχει εκπαιδευθεί με μέρος των δεδομένων εκπαίδευσης, η παραπάνω ανισότητα ισχύει και για το σύστημα που έχει δημιουργηθεί με εκπαίδευση με το σύνολο των δεδομένων εκπαίδευσης. Συνεπώς, το d είναι το άνω όριο για τον αριθμό των παραδειγμάτων που παράγουν σφάλμα, δηλαδή το άνω όριο για τα ψευδή αποτελέσματα.

- Ποσοστό ανάκλησης (recall), το οποίο ποσοτικοποιεί την πιθανότητα που έχει ένα παράδειγμα με απόκριση +1 να ταξινομηθεί ορθά. Στη συγκεκριμένη περίπτωση υπολογίζει την πιθανότητα που εμφανίζει μία πινακίδα π.χ. ΣΤΟΠ να ταξινομηθεί όντως ως πινακίδα ΣΤΟΠ. Μαθηματικά ισούται με:

$$Rec_{\xi\alpha}^n(h_L) = 1 - \frac{d_+}{n_+} \quad (5.2)$$

όπου d_+ το δεδομένα με θετική απόκριση που ικανοποιούν την ανισότητα $\rho\alpha_i R_{\Delta}^2 + \xi_i \geq 1$ και n_+ το σύνολο των παραδειγμάτων με θετική απόκριση. Δηλαδή το d_+ αποτελεί το άνω όριο των δεδομένων με θετική απόκριση, που υπάρχει περίπτωση να ταξινομηθούν εσφαλμένα.

Το ποσοστό ανάκλησης ποσοτικοποιεί πόσα από τα θετικά παραδείγματα επιλέχθηκαν ως έγκυρα (true positives) και δεν περιέχει πληροφορία σχετικά με τις λανθασμένες θετικές αποκρίσεις (false positives).

- Ποσοστό ορθότητας (precision) που ποσοτικοποιεί την πιθανότητα που έχει ένα παράδειγμα για το οποίο υπολογίστηκε $f(\mathbf{x})=1$ να έχει ταξινομηθεί σωστά. Δηλαδή, για τον συγκεκριμένο αλγόριθμο το ποσοστό ορθότητας υπολογίζει την πιθανότητα που εμφανίζει μία πινακίδα, η οποία έχει ταξινομηθεί ως σήμα ΣΤΟΠ, να αποτελεί πινακίδα ΣΤΟΠ στην πραγματικότητα. Μαθηματικά εκφράζεται ως:

$$Rec_{\xi\alpha}^n(h_L) = \frac{n_+ - d_+}{n_+ - d_+ + d_-} \quad (5.3)$$

όπου: n_+ το σύνολο των παραδειγμάτων με θετική απόκριση, δηλαδή $n_+ = |\{i: y_i = 1\}|$

d_+ το σύνολο των παραδειγμάτων με θετική απόκριση για τα οποία ισχύει $d_+ = |\{i: y_i = 1 \wedge (\rho\alpha_i R_{\Delta}^2 + \xi_i \geq 1)\}|$, δηλαδή το άνω όριο για τα ψευδώς θετικά αποτελέσματα

d_- το σύνολο των παραδειγμάτων με αρνητική απόκριση για τα οποία ισχύει $d_+ = |\{i: y_i = -1 \wedge (p\alpha_i R_\Delta^2 + \xi_i \geq 1)\}|$, δηλαδή το άνω όριο για τα ψευδώς αρνητικά αποτελέσματα.

Ο δείκτης της ορθότητας δεν εμπεριέχει πληροφορία για τα παραδείγματα που ανήκουν στην κατηγορία με απόκριση +1, αλλά ταξινομήθηκαν εσφαλμένα και ποσοτικοποιεί στην ουσία τον αριθμό των θετικών αποκρίσεων που ταξινομήθηκαν σωστά.

- Δείκτης F_1 ο οποίος συνδυάζει τα δύο παραπάνω μέτρα καθώς ορίζεται γεωμετρικά ως ο μέσος του ποσοστού ακρίβειας και ανάκλησης. Αλγεβρικά εκφράζεται ως:

$$F1_{\xi\alpha}^n(h_L) = \frac{2n_+ - 2d_+}{2n_+ - d_+ + d_-} \quad (5.4)$$

Σημαντικό πλεονέκτημα των παραπάνω δεικτών συνιστά ο ταχύς υπολογισμός τους, καθώς μπορούν να οριστούν πλήρως για ένα σύστημα που έχει εκπαιδευτεί (δηλαδή έχουν υπολογιστεί οι πολλαπλασιαστές Lagrange α_i και οι ψευδομεταβλητές ξ_i), χωρίς να απαιτείται η διαδικασία της εσωτερικής εξακρίβωσης.

Στη συγκεκριμένη περίπτωση, επειδή οι τιμές των διάφορων μεταβλητών δεν ήταν γνωστές a priori, για καλύτερα αποτελέσματα χρησιμοποιήθηκε η μέθοδος της εσωτερικής εξακρίβωσης (crossvalidation), όπως αναφέρεται στο κεφάλαιο 3. Η μέθοδος εσωτερικής εξακρίβωσης, που χρησιμοποιεί το SVMlight, είναι η leave-one-out, σύμφωνα με την οποία για ένα σύνολο N δεδομένων εκπαίδευσης η SVM εκπαιδεύεται N φορές. Σε κάθε επανάληψη, που πραγματοποιείται με διαφορετικές τιμές για τις μεταβλητές, ένα παράδειγμα αφήνεται έξω από τη διαδικασία εκπαίδευσης για να χρησιμοποιηθεί σε δεδομένο ελέγχου. Η διαδικασία εκπαίδευσης πραγματοποιείται με $N-1$ παραδείγματα και υπολογίζονται οι παραπάνω δείκτες που ποσοτικοποιούν την ικανότητα γενίκευσης. Τελικά, επιλέγεται η λύση που εξασφαλίζει τα καλύτερα αποτελέσματα. Στην περίπτωση υπολογισμού των δεικτών με τη διαδικασία εσωτερικής εξακρίβωσης, οι παραπάνω δείκτες χαρακτηρίζονται ως leave-one-out δείκτες. Οι δείκτες που προκύπτουν μετά τη διαδικασία εσωτερικής εξακρίβωσης είναι περισσότερο αμερόληπτοι σε σχέση με τους δείκτες Χι-άλφα.

Ακολούθως, στους Πίνακες 5.1 και 5.2 αναφέρονται οι τιμές για τους παραπάνω δείκτες, που προέκυψαν κατά την εκπαίδευση του συστήματος με τρία διαφορετικά σύνολα δεδομένων. Για κάθε πινακίδα που πρόκειται να εντοπιστεί χρησιμοποιήθηκαν δύο συναρτήσεις πυρήνα, γραμμική και πολυωνυμική, για να συγκριθούν μεταξύ τους τα ποσοστά επιτυχίας κάθε συνάρτησης.

Από τα αποτελέσματα που φαίνονται στον Πίνακα 5.1 παρατηρείται ότι όσο αυξάνει ο αριθμός των δεδομένων εκπαίδευσης βελτιώνεται η ικανότητα γενίκευσης του συστήματος, καθώς μειώνεται το ποσοστό σφάλματος και αυξάνεται το ποσοστό ορ-

θότητας. Δηλαδή, με εκπαίδευση του συστήματος με μεγαλύτερο αριθμό δεδομένων μειώνεται ο αριθμός των εσφαλμένων ταξινομήσεων και αυξάνεται η πιθανότητα μία πινακίδα που έχει ταξινομηθεί ως πινακίδα ποδηλατοδρόμου να αποτελεί στην πραγματικότητα ένα σήμα ύπαρξης ποδηλατοδρόμου. Ο μόνος δείκτης που επηρεάζεται ελαφρώς αρνητικά από την αύξηση του αριθμού των δεδομένων εκπαίδευσης είναι το ποσοστό ανάκλησης, το οποίο καταδεικνύει το ποσοστό των σημάτων ποδηλατοδρόμου, που τελικά εντοπίστηκαν. Επίσης, η χρήση της πολυωνυμικής συνάρτησης πυρήνα εξασφαλίζει καλύτερα αποτελέσματα, καθώς βελτιώνει αισθητά τους δείκτες τους σφάλματος και της ορθότητας, ειδικά στις περιπτώσεις που χρησιμοποιούνται λιγότερα δεδομένα εκπαίδευσης. Η λύση που χρησιμοποιεί 71 δεδομένα εκπαίδευσης και πολυωνυμική συνάρτηση πυρήνα εκτιμάται να παρουσιάζει καλύτερη ικανότητα γενίκευσης από τη λύση με τα 217 αποτελέσματα και τη γραμμική συνάρτηση. Γίνεται με τον τρόπο αυτό κατανοητή η μεγάλη σημασία επιλογής της κατάλληλης συνάρτησης πυρήνα.

| | Ποσοστό σφάλματος | | | Ποσοστό ανάκλησης | | | Ποσοστό ορθότητας | | |
|------------------------------|-------------------|----------|---------|-------------------|----------|---------|-------------------|----------|---------|
| | 217 εκπ. | 150 εκπ. | 71 εκπ. | 217 εκπ. | 150 εκπ. | 71 εκπ. | 217 εκπ. | 150 εκπ. | 71 εκπ. |
| Γραμμική συνάρτηση πυρήνα | 9.68% | 11.33% | 19.72% | 99.25% | 100% | 100% | 86.93% | 84.11% | 76.27% |
| Πολυωνυμική συνάρτηση πυρήνα | 4.15% | 4.67% | 7.04% | 97.01% | 98.89% | 100% | 96.30% | 93.68% | 90.00% |

Πίνακας 5.1: Δείκτες leave-one-out για την εκπαίδευση του ταξινομητή για την πινακίδα του ποδηλατοδρόμου, με διάφορα σύνολα δεδομένων

Στον Πίνακα 5.2 εμφανίζονται τα αντίστοιχα αποτελέσματα για την πινακίδα του ΣΤΟΠ.

| | Ποσοστό σφάλματος | | | Ποσοστό ανάκλησης | | | Ποσοστό ορθότητας | | |
|------------------------------|-------------------|----------|---------|-------------------|----------|---------|-------------------|----------|---------|
| | 207 εκπ. | 149 εκπ. | 76 εκπ. | 207 εκπ. | 149 εκπ. | 76 εκπ. | 207 εκπ. | 149 εκπ. | 76 εκπ. |
| Γραμμική συνάρτηση πυρήνα | 0.48% | 0.67% | 1.32% | 99.19% | 98.91% | 98.04% | 100% | 100% | 100% |
| Πολυωνυμική συνάρτηση πυρήνα | 0.48% | 0.67% | 1.32% | 99.19% | 98.91% | 98.04% | 100% | 100% | 100% |

Πίνακας 5.2: Δείκτες leave-one-out για την εκπαίδευση του ταξινομητή για την πινακίδα του ΣΤΟΠ, με διάφορα σύνολα δεδομένων

Για την αναγνώριση των πινακίδων του ΣΤΟΠ οι δείκτες γενίκευσης παρουσιάζουν πολύ πιο ικανοποιητικές τιμές σε σχέση με την πινακίδα υπαγόρευσης ποδηλατοδρόμου. Αυτό συμβαίνει γιατί τα δεδομένα του ΣΤΟΠ περιγράφονται πολύ ικανοποιητικά από τους περιγραφείς HOG που έχουν εξαχθεί. Αποτελούνται από μεγαλύτερα γράμματα στο εσωτερικό τους σε σχέση με την πινακίδα του ποδηλάτου, που το εικονόγραμμά της χαρακτηρίζεται από μεγαλύτερη λεπτομέρεια. Συνεπώς, το ακριβές σχήμα των σημάτων ΣΤΟΠ αναπαρίσταται πλήρως από τους περιγραφείς που προέκυψαν σύμφωνα με τις παραμέτρους που ορίστηκαν και έτσι τα δεδομένα είναι εύκολα διαχωρίσιμα. Για την καλύτερη απόδοση του συστήματος αναγνώρισης σημάτων με ποδήλατο, πιθανώς η αλλαγή ορισμένων παραμέτρων για την εξαγωγή

των περιγραφικών χαρακτηριστικών να ήταν ευνοϊκή, όπως για παράδειγμα η μείωση των διαστάσεων των υποπεριοχών (cells) και η αύξηση του αριθμού των προκαθορισμένων προσανατολισμών στο ιστόγραμμα (bins), ώστε να καταγράφεται το σχήμα με μεγαλύτερη λεπτομέρεια.

Όμως, ακόμη και στην περίπτωση των πινακίδων ΣΤΟΠ, η αύξηση του αριθμού των δεδομένων εκπαίδευσης βελτιώνει ελαφρώς την ικανότητα γενίκευσης του συστήματος. Είναι ενδιαφέρον ότι στην περίπτωση αυτή η χρήση πολυωνυμικού πυρήνα παρέχει ακριβώς τα ίδια αποτελέσματα με τη χρήση γραμμικού πυρήνα. Αυτό συμβαίνει διότι οι παράμετροι της πολυωνυμικής συνάρτησης πυρήνα μπορεί να οριστούν με τέτοιο τρόπο ώστε η συνάρτηση να ταυτίζεται με τη γραμμική συνάρτηση πυρήνα, και στη συγκεκριμένη περίπτωση τα δεδομένα διαχωρίζονται βέλτιστα γραμμικά. Για τον λόγο αυτό, οι τιμές που προέκυψαν από τη διαδικασία της εσωτερικής εξακρίβωσης είναι τέτοιες ώστε οι δύο συναρτήσεις πυρήνα να ταυτίζονται.

5.2.2 Αποτελέσματα αναγνώρισης

Για το στάδιο της αναγνώρισης πραγματοποιήθηκαν διάφορες δοκιμές με χρήση των ταξινομητών που δημιουργήθηκαν. Στο σύνολο των περιοχών ενδιαφέροντος, που προέκυψαν από το βήμα του εντοπισμού, επισημάνθηκαν οι αποκρίσεις ώστε οι περιοχές αυτές να χρησιμοποιηθούν ως δεδομένα ελέγχου. Με τον τρόπο αυτό γίνεται να καθοριστεί η ακρίβεια του κάθε συστήματος και να οριστούν οι δείκτες ορθότητας και ανάκλησης για κάθε συγκεκριμένο σύνολο δεδομένων εκπαίδευσης. Επιπλέον δείκτης που χρησιμοποιείται ώστε να ποσοτικοποιήσει την επιτυχία της αναγνώρισης είναι η ακρίβεια, που ισούται με τον αριθμό των σωστά ταξινομημένων παραδειγμάτων προς το σύνολο των παραδειγμάτων και μαθηματικά εκφράζεται:

$$Acc = \frac{TP + TN}{FP + FN + TP + TN} \quad (5.5)$$

όπου: TP = True Positives, ο αριθμός των σωστά ταξινομημένων παραδειγμάτων με θετική απόκριση

TN = True Negatives, ο αριθμός των σωστά ταξινομημένων παραδειγμάτων με αρνητική απόκριση

FP = False Positives, ο αριθμός των λανθασμένα ταξινομημένων παραδειγμάτων με θετική απόκριση

FN = False Negatives, ο αριθμός των λανθασμένα ταξινομημένων παραδειγμάτων με αρνητική απόκριση

Από τους παρακάτω πίνακες προκύπτει ότι η συνάρτηση πυρήνα που χρησιμοποιείται, ο αριθμός των δεδομένων εκπαίδευσης και το είδος των περιγραφικών χαρακτηριστικών αποτελούν παράγοντες αλληλένδετους για την επιτυχία της διαδικασίας αναγνώρισης. Για παράδειγμα, εάν τα περιγραφικά διανύσματα αναπαριστούν με μεγάλη λεπτομέρεια τα δεδομένα, τότε αυτά θα είναι γραμμικά διαχωρίσιμα και μπορεί να χρησιμοποιηθεί μία γραμμική συνάρτηση πυρήνα. Σε αυτή την περίπτωση απαιτείται μικρότερος αριθμός δεδομένων εκπαίδευσης. Αντίστροφα,

εάν τα δεδομένα δεν περιγράφονται ικανοποιητικά από το περιγραφικό διάγραμμα, πρέπει να χρησιμοποιηθεί μεγαλύτερος αριθμός δεδομένων εκπαίδευσης και μία μη γραμμική συνάρτηση πυρήνα.

| | Ποσοστό ακρίβειας | | | Ποσοστό ανάκλησης | | | Ποσοστό ορθότητας | | |
|------------------------------|-------------------|----------|---------|-------------------|----------|---------|-------------------|----------|---------|
| | 217 εκπ. | 150 εκπ. | 71 εκπ. | 217 εκπ. | 150 εκπ. | 71 εκπ. | 217 εκπ. | 150 εκπ. | 71 εκπ. |
| Γραμμική συνάρτηση πυρήνα | 88.05% | 60.38% | 23.27% | 95.24% | 95.24% | 100% | 52.63% | 24.93% | 14.69% |
| Πολυωνυμική συνάρτηση πυρήνα | 97.48% | 93.08% | 72.33% | 85.71% | 85.71% | 95.24% | 94.74% | 69.23% | 31.75% |

Πίνακας 5.3: Ποσοστά επιτυχίας αναγνώρισης του ταξινομητή για την πινακίδα του ποδηλατόδρομου, με διάφορα σύνολα δεδομένων

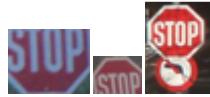
| | Ποσοστό ακρίβειας | | | Ποσοστό ανάκλησης | | | Ποσοστό ορθότητας | | |
|------------------------------|-------------------|----------|---------|-------------------|----------|---------|-------------------|----------|---------|
| | 207 εκπ. | 149 εκπ. | 76 εκπ. | 207 εκπ. | 149 εκπ. | 76 εκπ. | 207 εκπ. | 149 εκπ. | 76 εκπ. |
| Γραμμική συνάρτηση πυρήνα | 94.34% | 94.34% | 93.71% | 60.87% | 60.87% | 65.22% | 100% | 100% | 88.24% |
| Πολυωνυμική συνάρτηση πυρήνα | 93.71% | 93.71% | 94.74% | 56.52% | 56.52% | 65.22% | 100% | 100% | 100% |

Πίνακας 5.4: Ποσοστά επιτυχίας αναγνώρισης του ταξινομητή για την πινακίδα του ΣΤΟΠ, με διάφορα σύνολα δεδομένων

Πιο συγκεκριμένα, όπως κατέδειξαν και οι δείκτες leave-one-out και επιβεβαίωσαν τα ποσοστά επιτυχίας της αναγνώρισης, στην περίπτωση του σήματος ΣΤΟΠ, καθώς το εικονόγραμμα είναι αδρό, οι περιγραφείς HOG, με τις παραμέτρους που επιλέχθηκαν, είναι ικανοί να αναπαραστήσουν πλήρως την πληροφορία των πινακίδων, συνεπώς δεν παρατηρούνται μεγάλες διακυμάνσεις στα ποσοστά ανάκλησης, ορθότητας και ακρίβειας όταν χρησιμοποιούνται περισσότερα δεδομένα εκπαίδευσης ή η πολυωνυμική συνάρτηση πυρήνα. Ουσιαστικές διαφορές μεταξύ γραμμικής και πολυωνυμικής συνάρτησης πυρήνα παρατηρούνται μόνο όταν χρησιμοποιείται το σετ με τα 76 δεδομένα εκπαίδευσης. Αντίθετα, το εικονόγραμμα του σήματος που υποδηλώνει την ύπαρξη ποδηλατόδρομου είναι πιο λεπτομερές, με αποτέλεσμα το διάνυσμα HOG να μην περιγράφουν εξίσου ικανοποιητικά τα δεδομένα, με την πινακίδα του ΣΤΟΠ. Για τον λόγο αυτό, παρατηρούνται εμφανείς διαφορές στα ποσοστά επιτυχίας της μεθόδου ανάλογα με τον αριθμό των δεδομένων εκπαίδευσης και της συνάρτησης πυρήνα που χρησιμοποιείται. Η πιο επιτυχής αναγνώρισης επιτυγχάνεται για 217 δεδομένα εκπαίδευσης και πολυωνυμική συνάρτηση πυρήνα.

Ακόμα, εάν εξετασθούν πιο λεπτομερώς τα ποσοστά που ποσοτοκοποιούν την επιτυχία της αναγνώρισης για το σήμα ΣΤΟΠ επισημαίνονται ιδιαίτερα υψηλά ποσοστά ακρίβειας, που σημαίνει ότι η κατάταξη στις δύο κλάσεις ήταν αρκετά επιτυχής. Για τα δύο πρώτα σύνολα δεδομένων (207 και 149 δεδομένα εκπαίδευσης) και για τις δύο συναρτήσεις πυρήνα, τα ποσοστά ορθότητας ισούνται με 100%, που σημαίνει ότι όλες οι πινακίδες που αναγνωρίστηκαν ως ΣΤΟΠ αποτελούσαν στην πραγματικότητα πινακίδες αυτού του είδους, ενώ τα ποσοστά ανάκλησης είναι μικρότερα,

το οποίο σημαίνει ότι ορισμένες πινακίδες ΣΤΟΠ δεν αναγνωρίστηκαν επιτυχώς. Από τη μελέτη των τριών δεικτών προκύπτει ότι το σημαντικότερο πρόβλημα της αναγνώρισης είναι η αποτυχία αναγνώρισης ορισμένων πινακίδων, που λανθασμένα απόδόθηκαν στην κλάση -1. Στην Εικόνα 5.13 δίνονται ορισμένες περιπτώσεις που η αναγνώριση για το σήμα ΣΤΟΠ απέτυχε (στην περίπτωση που ο ταξινομητής εκπαιδεύθηκε με 207 δεδομένα και γραμμική συνάρτηση πυρήνα).



Εικόνα 5.13: Παραδείγματα πινακίδων ΣΤΟΠ που δεν αναγνωρίστηκαν επιτυχώς

Στις δύο πρώτες περιπτώσεις, η αναγνώριση αποτυγχάνει καθώς οι εικόνες που τροφοδοτούνται στη διαδικασία ταξινόμησης δεν περιλαμβάνουν ολόκληρη την πινακίδα οδικής κυκλοφορίας, ενώ στην τρίτη περίπτωση πρόκειται να αναγνωριστούν δύο επικαλυπτόμενα σήματα οδικής κυκλοφορίας και, όπως ήταν αναμενόμενο, η αναγνώριση αποτυγχάνει. Τονίζεται στο σημείο αυτό η σημασία του βήματος διαχωρισμού των επικαλυπτόμενων σημάτων, η οποία πραγματοποιείται με εντοπισμό με βάση το σχήμα, όπως αναλύεται ανωτέρω. Τα σχετικά χαμηλά ποσοστά ανάκλησης οφείλονται στο στάδιο του εντοπισμού. Καθώς, όπως διαφαίνεται από τα ποσοστά ακρίβειας και ορθότητας ο διαχωρισμός του ΣΤΟΠ από τα υπόλοιπα σήματα είναι αρκετά επιτυχής, μία πιο ακριβής διαδικασία εντοπισμού, θα βοηθούσε και στην αύξηση των ποσοστών ανάκλησης.

Αντίθετα, στην περίπτωση του σήματος που υποδηλώνει την ύπαρξη ποδηλατόδρομου, τα ποσοστά ανάκλησης είναι μεγαλύτερα από τα ποσοστά ορθότητας. Αυτό σημαίνει ότι το σύστημα επιστρέφει το μεγαλύτερο ποσοστό των σημάτων που είναι σχετικά με την αναζήτηση, αλλά ανάμεσα στα θετικά αποτελέσματα, συμπεριλαμβάνονται και αρκετές εικόνες που έχουν λανθασμένα αποδοθεί στην κατηγορία +1, δηλαδή φαίνεται να απεικονίζουν την πινακίδα του ποδηλατοδρόμου, ενώ στην πραγματικότητα δεν την περιέχουν. Από την εξέταση των διαφορών μεταξύ γραμμικής και πολυωνυμικής συνάρτησης πυρήνα, προκύπτει ότι στην πρώτη περίπτωση τα ποσοστά ανάκλησης είναι 100% και τα ποσοστά ορθότητας αρκετά χαμηλά, ενώ στη δεύτερη περίπτωση τα ποσοστά ανάκλησης χειροτερεύουν και τα ποσοστά ορθότητας αυξάνονται κατακόρυφα. Αυτό συμβαίνει γιατί η γραμμική συνάρτηση πυρήνα δε διαχωρίζει σωστά τα δεδομένα, με αποτέλεσμα να κατατάσσει πολλά δεδομένα λανθασμένα στην κατηγορία $f(x) = 1$ (false positives), ενώ η πολυωνυμική συνάρτηση πυρήνα διαχωρίζει πολύ καλύτερα τα δεδομένα, με αποτέλεσμα να αυξάνονται τα ποσοστά ορθότητας και ακρίβειας, αφαιρεί όμως ορισμένα από τα σωστά, θετικά ταξινομημένα δεδομένα της προηγούμενης περίπτωσης, και έτσι μειώνονται τα ποσοστά ανάκλησης.

Στην Εικόνα 5.14 εμφανίζονται δύο χαρακτηριστικές περιπτώσεις που ο ταξινομητής δεν αναγνώρισε επιτυχώς την πινακίδα οδικής κυκλοφορίας, λόγω σκιών και μειωμένης φωτεινότητας των εικόνων.



Εικόνα 5.14: Παραδείγματα πινακίδων ποδηλατοδρόμου που δεν αναγνωρίστηκαν επιτυχώς

Σε γενικές γραμμές, οι περιγραφείς HOG συνιστούν μία αρκετά ικανοποιητική λύση για την αναπαράσταση της πληροφορίας της εικόνας στο πρόβλημα της αυτόματης κατηγοριοποίησης των σημάτων οδικής κυκλοφορίας. Ακόμα και στην περίπτωση λεπτομερών εικονογραμμάτων (όπως στην περίπτωση του ποδηλατόδρομου) εάν συνδυαστούν με την κατάλληλη συνάρτηση πυρήνα και ο ταξινομητής εκπαιδευθεί με ικανό αριθμό δειγμάτων τα αποτελέσματα της αναγνώρισης είναι πολύ ικανοποιητικά. Επίσης, αποδείχθηκε πως οι συγκεκριμένοι περιγραφείς παρουσιάζουν ανοχή σε μικρές στροφές και έντονες αλλαγές φωτεινότητας καθώς τα σύνολα των εικόνων που χρησιμοποιήθηκαν για τα πειράματα ήταν ειλημμένα με πολύ διαφορετικές συνθήκες φωτεινότητας.

6

Συμπεράσματα

Στην παρούσα εργασία παρουσιάζεται μία νέα μέθοδος για τον αυτόματο εντοπισμό και κατηγοριοποίηση των σημάτων οδικής κυκλοφορίας. Βασικός στόχος είναι η μελέτη της επίδρασης των περιγραφέων HOG στη διαδικασία αναγνώρισης. Στο παρόν κεφάλαιο συνοψίζονται τα συμπεράσματα που προέκυψαν και αναφέρονται ορισμένες μελλοντικές επεκτάσεις του αλγόριθμου που αναπτύχθηκε.

6.1 Σχόλια και παρατηρήσεις

Ο αλγόριθμος που συντάχθηκε είναι ικανός να εντοπίζει και να κατηγοριοποιεί αυτόματα τις πινακίδες οδικής κυκλοφορίας. Μοναδικό δεδομένο εισόδου για τον αλγόριθμο συνιστούν έγχρωμες φωτογραφίες, που ελήφθησαν από μία χαμηλού κόστους φωτομηχανή, η οποία τοποθετήθηκε σε κινούμενο όχημα. Αρχικά, η εικόνα μεταφέρεται στον χρωματικό χώρο HSI, όπου πραγματοποιείται κατωφλίωση με καθολικό όριο. Κατόπιν, ένα φίλτρο μεσαίας τιμής εφαρμόζεται στη δυαδική εικόνα για απομάκρυνση του θορύβου και στη συνέχεια πραγματοποιείται διαστολή με δομικό στοιχείων διαστάσεων 3×3 . Ακολούθως, ταυτοποιούνται τα διασυνδεδεμένα μέρη και υπολογίζονται τα περιγεγραμμένα τετράπλευρα των περιοχών ενδιαφέροντος, που τελικά εντοπίστηκαν. Όσες περιοχές ενδιαφέροντος δεν εμφανίζουν εμβαδό και λόγο των κάθετων πλευρών του περιγεγραμμένου τετραπλεύρου, εντός συγκεκριμένων ορίων, απομακρύνονται ως εσφαλμένες. Ενώ, όσες από τις περιοχές ενδιαφέροντος εμφανίζουν λόγο κάθετων πλευρών μεταξύ των τιμών 0.4 και 0.7 τροφοδοτούνται σε ένα στάδιο εντοπισμού βάσει σχήματος, καθώς είναι πιθανόν να απεικονίζουν επικαλυπτόμενα σήματα οδικής κυκλοφορίας, τα οποία πρέπει να διαχωριστούν. Στο στάδιο αυτό, εφαρμόζεται ο κυκλικός μετασχηματισμός Hough και εάν εντοπιστεί κύκλος δημιουργείται μία νέα περιοχή ενδιαφέροντος. Οι έναπομείνουσες περιοχές ελέγχονται ως προς το εμβαδό τους, το οποίο καθορίζει ένα θα αποτελέσουν και αυτές νέες περιοχές ενδιαφέροντος. Το σύνολο των τελικών περιοχών που εξήχθησαν από την παραπάνω διαδικασία, αναπαρίστανται με τη βοήθεια των περιγραφέων HOG και εισέρχονται σε δυαδικούς ταξινομητές, που έχουν προηγουμένως εκπαιδευθεί, προκειμένου να αναγνωριστούν τελικά οι πινακίδες.

Στην παραπάνω περιγραφή συνοψίζονται οι προσθήκες και τροποποιήσεις που πραγματοποιήθηκαν σε ήδη υπάρχοντες αλγορίθμους που συναντώνται στη διεθνή βιβλιογραφία, ώστε να βελτιστοποιηθούν τα ποσοστά επιτυχίας της μεθόδου. Τα

αποτελέσματα ήταν ιδιαίτερα ικανοποιητικά καθώς το ποσοστό εντοπισμού της μεθόδου που αναπτύχθηκε ισούται με 95% και το μέσο ποσοστό ακρίβειας αναγνώρισης για τις δύο τάξεις με 95.91%. Σε πολλές περιπτώσεις τα συγκεκριμένα ποσοστά υπερίσχυαν των ποσοστών επιτυχίας γνωστών μέχρι τώρα μεθόδων, όπως αυτής που παρουσιάζεται στο [59], όπου τα ποσοστά εντοπισμού και αναγνώρισης ήταν 90% και 75% αντίστοιχα, στο [124], όπου τα ανάλογα ποσοστά είναι 78% και 88% και στο [77], όπου το ποσοστό εντοπισμού ισούται με 74% και το ποσοστό αναγνώρισης με 89.1%.

Συνοπτικά, τα ποσοστά αναγνώρισης του αλγόριθμου είναι ιδιαίτερα ικανοποιητικά. Σημαντική βελτίωση των αποτελεσμάτων παρατηρείται με την εφαρμογή του φίλτρου μεσαίας τιμής για την απομάκρυνση του θορύβου, το οποίο δεν έχει χρησιμοποιηθεί μέχρι στιγμή ευρέως στη διεθνή βιβλιογραφία. Μία ακόμη σημαντική συμβολή του αλγορίθμου στην ήδη υπάρχουσα έρευνα στο πεδίο αυτό, συνιστά το στάδιο που ενσωματώθηκε και στοχεύει στον διαχωρισμό των επικαλυπτόμενων σημάτων, πρόβλημα που αντιμετωπίζεται σπάνια. Το βήμα αυτό είναι ιδιαίτερα σημαντικό, καθώς πολύ συχνά συναντώνται επικαλυπτόμενα σήματα, τα όποια εάν δε διαχωρίσουν δεν είναι δυνατόν να κατηγοριοποιηθούν επιτυχώς.

Όσον αφορά στο στάδιο της αναγνώρισης, οι περιγραφείς HOG, οι οποίοι δεν είναι ιδιαίτερα διαδεδομένοι για το συγκεκριμένο θέμα, καθώς αρχικά χρησιμοποιήθηκαν για αναγνώριση ανθρώπων, συνιστούν τελικά ενδεδειγμένο τρόπο αναπαράστασης της πληροφορίας και για αυτό το πρόβλημα, αφού εξασφαλίζουν ιδιαίτερα υψηλά ποσοστά αναγνώρισης. Ακόμα, υπερνικούν σε αρκετές περιπτώσεις προβλήματα που ανακύπτουν στην ταξινόμηση οδικών πινακίδων και οφείλονται σε στροφές των σημάτων, δυσμενείς καιρικές συνθήκες και ύπαρξη αυτοκόλλητων στο σήμα. Τα αποτελέσματα της κατηγοριοποίησης είναι ιδιαίτερα ικανοποιητικά, όπως καταδεικνύουν και οι δείκτες που δίνονται στο κεφάλαιο 5. Τα κυριότερα προβλήματα, που προέκυψαν και οδήγησαν σε εσφαλμένη αναγνώριση πινακίδων, οφείλονται στον εν μέρει εντοπισμό του σήματος, πρόβλημα που μπορεί να αντιμετωπισθεί με βελτίωση της μεθόδου εντοπισμού αλλά και σε δυσμενείς συνθήκες φωτισμού, πρόβλημα που είναι δυνατόν να αντιμετωπισθεί με περαιτέρω προεπεξεργασία της εικόνας.

6.2 Προτάσεις για το μέλλον

Ορισμένες προτάσεις για επέκταση του αλγορίθμου στο μέλλον περιλαμβάνουν τη διεξαγωγή πειραμάτων με διαφορετική παραμετροποίηση για τους περιγραφείς HOG, ώστε να επιλεγθούν οι καταλληλότερες τιμές κατά το δυνατόν, αλλά και τη διεξαγωγή πειραμάτων που χρησιμοποιούν διαφορετική αναπαράσταση των εικόνων (π.χ. μέσω σημείων ενδιαφέροντος SIFT) σε συνδυασμό με διάφορες μεθόδους κωδικοποίησης (όπως Fischer Vectors και VLAD). Θα ήταν ακόμα ενδιαφέρον να μελετηθεί η απόδοση άλλων ταξινομητών (π.χ νευρωνικών δικτύων) σε συνδυασμό με την αναπαράσταση με περιγραφείς HOG, που χρησιμοποιήθηκε στην παρούσα εφαρμογή. Επίσης, ένας άλλος έλεγχος μπορεί να πραγματοποιηθεί εάν τα δεδομένα εισόδου του αλγορίθμου δεν είναι εικόνες αλλά βίντεο και συγκριθούν τελικά τα

ποσοστά επιτυχίας της μεθόδου, ως προς τον εντοπισμό και την αναγνώριση. Τέλος, χρήσιμο θα ήταν να ενσωματωθεί ένας αλγόριθμος παρακολούθησης (π.χ. KLT tracker), ώστε να απομακρύνονται ως εσφαλμένες όσες από τις περιοχές ενδιαφέροντος εντοπίστηκαν και παρουσιάζουν μεγάλη μετατόπιση από λήψη σε λήψη. Βέβαια, καθώς η μετατόπιση οφείλεται και στη σχετική κίνηση της φωτογραφικής μηχανής, θα πρέπει προηγουμένως να εφαρμοστεί μία τεχνική που αναιρεί την κίνηση που οφείλεται στη μηχανή (motion compensation).

Άλλες πιθανές προεκτάσεις της εργασίας στα ευρύτερα πλαίσια, συνιστούν ο τρισδιάστατος εντοπισμός των πινακίδων, ο οποίος μπορεί να εκμεταλλευτεί τη γεωμετρία των επίπεδων πινακίδων, και τα αποτελέσματα να συγκριθούν με τον δισδιάστατο εντοπισμό. Στα πλαίσια δε ενός ολοκληρωμένου συστήματος υποβοήθησης οδήγησης, θα ήταν ενδιαφέρον μαζί με τον προτεινόμενο αλγόριθμο να αναπτυχθεί ένας αλγόριθμος εντοπισμού των ορίων της λωρίδας οδικής κυκλοφορίας καθώς και εντοπισμού πιθανών εμποδίων στη διαδρομή κίνησης, ώστε να προσεγγισθεί η πλήρως αυτοματοποιημένη οδήγηση.

Βιβλιογραφία

- [1] http://www.astynomia.gr/images/stories/DOCS/KOK_iNomoi_2012Feb.pdf
(Πρόσβαση: 26 Αυγούστου 2014)
- [2] <http://euler.fd.cvut.cz/research/rs2/articles/itsp.html>.
(Πρόσβαση: 2 Αυγούστου 2014)
- [3] Sharma, N., Mishra, M., Shrivastava, M. (2012). Colour image segmentation techniques and issues: an approach. *International Journal of Scientific & Technology Research*, 4.
- [4] Dass, R., & Devi, S. (2002). Image Segmentation Techniques. *International Journal of Electronics & Communication Technology*, 3(1).
- [5] Νικήτα, Κ. (2002). *Στοιχεία επεξεργασίας Ιατρικών Σημάτων*. Σημειώσεις Σχολής ΗΜΜΥ, Αθήνα, ΕΜΠ.
- [6] Sezgin, M. (2004). Survey over image thresholding techniques and quantitative performance evaluation. *Journal of Electronic imaging*, 13(1), 146-168.
- [7] Brkic, K. (2010). An overview of traffic sign detection methods. *Department of Electronics, Microelectronics, Computer and Intelligent Systems Faculty of Electrical Engineering and Computing Unska*, 3, 10000.
- [8] Shadeed, W. G., Abu-Al-Nadi, D. I., & Mismar, M. J. (2003). Road traffic sign detection in color images. In *Electronics, Circuits and Systems, 2003. ICECS 2003. Proceedings of the 2003 10th IEEE International Conference on* (Vol. 2, pp. 890-893). IEEE.
- [9] Broggi, A., Cerri, P., Medici, P., Porta, P. P., & Ghisio, G. (2007). Real time road signs recognition. In *Intelligent Vehicles Symposium, 2007 IEEE* (pp. 981-986). IEEE.
- [10] Andrey, V., & Jo, K. H. (2006). Automatic detection and recognition of traffic signs using geometric structure analysis. In *SICE-ICASE, 2006. International Joint Conference* (pp. 1451-1456). IEEE.
- [11] Fleyeh, H. (2004). Color detection and segmentation for road and traffic signs. In *Cybernetics and Intelligent Systems, 2004 IEEE Conference on* (Vol. 2, pp. 809-814). IEEE.
- [12] Bahlmann, C., Zhu, Y., Ramesh, V., Pellkofer, M., & Koehler, T. (2005). A system for traffic sign detection, tracking, and recognition using color, shape, and motion information. In *Intelligent Vehicles Symposium, 2005. Proceedings. IEEE* (pp. 255-260). IEEE.

- [13] Ruta, A., Li, Y., & Liu, X. (2008). Detection, tracking and recognition of traffic signs from video input. In *Proc. ITSC* (pp. 55-60).
- [14] Ghica, D., Lu, S. W., & Yuan, X. (1995). Recognition of traffic signs by artificial neural network. In *Neural Networks, 1995. Proceedings., IEEE International Conference on* (Vol. 3, pp. 1444-1449). IEEE.
- [15] Yuille, A. L., Snow, D., & Nitzberg, M. (1998). Signfinder: using color to detect, localize and identify informational signs. In *Computer Vision, 1998. Sixth International Conference on* (pp. 628-633). IEEE.
- [16] Kim, G. H., Sohn, H. G., & Song, Y. S. (2006). Road Infrastructure Data Acquisition Using a Vehicle-Based Mobile Mapping System. *Computer-Aided Civil and Infrastructure Engineering*, 21(5), 346-356.
- [17] Vitabile, S., Pollaccia, G., Pilato, G., & Sorbello, F. (2001). Road signs recognition using a dynamic pixel aggregation technique in the HSV color space. In *Image Analysis and Processing, 2001. Proceedings. 11th International Conference on* (pp. 572-577). IEEE.
- [18] Silapachote, P., Hanson, A., & Weiss, R. (2005). A hierarchical approach to sign recognition. In *Application of Computer Vision, 2005. WACV/MOTIONS'05 Volume 1. Seventh IEEE Workshops on* (Vol. 1, pp. 22-28). IEEE.
- [19] http://www.subsurfwiki.org/wiki/Symmetric_nearest_neighbour_filter
(Πρόσβαση: 20 Ιανουαρίου 2014)
- [20] Fleyeh, H. (2008). *Traffic and road sign recognition*. Doctoral dissertation, Napier University, Edinburgh, Scotland.
- [21] Borji, A., & Hamidi, M. (2007). Evolving a fuzzy rule-base for image segmentation. *International Journal of Intelligent Systems and Technologies*, 28, 178-183.
- [22] Arlicot, A., Soheilian, B., & Paparoditis, N. (2009). Circular Road sign extraction from street level images using colour, shape and texture databases maps. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 205-210.
- [23] Wang, Y. Q., Liu, L. M., & Yang, Z. (2009). Traffic sign detection based on fixed color combination and intensity restraint. In *Computer Network and Multimedia Technology, 2009. CNMT 2009. International Symposium on* (pp. 1-5). IEEE.
- [24] Fleyeh, H. (2008). Traffic sign recognition by fuzzy sets. In *Intelligent Vehicles Symposium, 2008 IEEE* (pp. 422-427). IEEE.

- [25] Wu, W. Y., Hsieh, T. C., & Lai, C. S. (2007). Extracting road signs using the color information. *World Academy of Science, Engineering and Technology*, 32, 282-286.
- [26] Paulo, C. F., & Correia, P. L. (2007). Automatic detection and classification of traffic signs. In *Image Analysis for Multimedia Interactive Services, 2007. WIAMIS'07. Eighth International Workshop on* (pp. 11-11). IEEE.
- [27] Malik, R. A. B. I. A., Khurshid, J. A. V. A. I. D., & Ahmad, S. N. (2007). Road sign detection and recognition using colour segmentation, shape analysis and template matching. In *Machine Learning and Cybernetics, 2007 International Conference on* (Vol. 6, pp. 3556-3560). IEEE.
- [28] Fleyeh, H., Biswas, R., & Davami, E. (2013). Traffic sign detection based on AdaBoost color segmentation and SVM classification. In *EUROCON, 2013 IEEE* (pp. 2005-2010). IEEE.
- [29] Maldonado-Bascon, S., Lafuente-Arroyo, S., Gil-Jimenez, P., Gomez-Moreno, H., & López-Ferreras, F. (2007). Road-sign detection and recognition based on support vector machines. *Intelligent Transportation Systems, IEEE Transactions on*, 8(2), 264-278.
- [30] Lafuente-Arroyo, S., Maldonado-Bascón, S., Gil-Jiménez, P., Gomez-Moreno, H., & López-Ferreras, F. (2006). Road sign tracking with a predictive filter solution. *Proc. of IEEE IECON, Paris, France*, 3314-3319.
- [31] de la Escalera, A., Armingol, J. M., Pastor, J. M., & Rodríguez, F. J. (2004). Visual sign information extraction and identification by deformable models for intelligent vehicles. *Intelligent Transportation Systems, IEEE Transactions on*, 5(2), 57-68.
- [32] Kuo, W. J., & Lin, C. C. (2007). Two-stage road sign detection and recognition. In *Multimedia and Expo, 2007 IEEE International Conference on* (pp. 1427-1430). IEEE.
- [33] Lafuente-Arroyo, S., Maldonado-Bascon, S., Gil-Jimenez, P., Acevedo-Rodriguez, J., & Lopez-Sastre, R. J. (2007). A tracking system for automated inventory of road signs. In *Intelligent Vehicles Symposium, 2007 IEEE* (pp. 166-171). IEEE.
- [34] Maldonado-Bascon, S., Lafuente-Arroyo, S., Siegmann, P., Gomez-Moreno, H., & Acevedo-Rodriguez, F. J. (2008). Traffic sign recognition system for inventory purposes. In *Intelligent Vehicles Symposium, 2008 IEEE* (pp. 590-595). IEEE.
- [35] Lafuente-Arroyo, S., Gil-Jimenez, P., Maldonado-Bascon, R., Lopez-Ferreras, F., & Maldonado-Bascon, S. (2005). Traffic sign shape classification evaluation I: SVM using distance to borders. In *Intelligent Vehicles Symposium, 2005. Proceedings. IEEE* (pp. 557-562). IEEE.

- [36] Kiran, C. G., Prabhu, L. V., Abdu, R. V., & Rajeev, K. (2009). Traffic sign detection and pattern recognition using support vector machine. In *Advances in Pattern Recognition, 2009. ICAPR'09. Seventh International Conference on* (pp. 87-90). IEEE.
- [37] Gil-Jimenez, P., Lafuente-Arroyo, S., Gomez-Moreno, H., Lopez-Ferreras, F., & Maldonado-Bascon, S. (2005). Traffic sign shape classification evaluation. Part II. FFT applied to the signature of blobs. In *Intelligent Vehicles Symposium, 2005. Proceedings. IEEE* (pp. 607-612). IEEE.
- [38] Kiran, C. G., Prabhu, L. V., Abdu, R. V., & Rajeev, K. (2009). Traffic sign detection and pattern recognition using support vector machine. In *Advances in Pattern Recognition, 2009. ICAPR'09. Seventh International Conference on* (pp. 87-90). IEEE.
- [39] Siogkas, G. K., & Dermatas, E. S. (2006). Detection, tracking and classification of road signs in adverse conditions. In *Electrotechnical Conference, 2006. MELECON 2006. IEEE Mediterranean* (pp. 537-540). IEEE.
- [40] http://en.wikipedia.org/wiki/Lab_color_space
(Πρόσβαση: 22 Δεκεμβρίου 2013)
- [41] Otsu, N. (1975). A threshold selection method from gray-level histograms. *Automatica*, 11(285-296), 23-27.
- [42] Eichner, M. L., & Breckon, T. P. (2008,). Integrated speed limit detection and recognition from real-time video. In *Intelligent Vehicles Symposium, 2008 IEEE* (pp. 626-631). IEEE.
- [43] <http://en.wikipedia.org/wiki/YCbCr>
(Πρόσβαση: 23 Δεκεμβρίου 2013)
- [44] Zakir, U. (2011). *Automatic Road Sign Detection and Recognition*. Doctoral dissertation, Loughborough University, Loughborough, United Kingdom.
- [45] <http://en.wikipedia.org/wiki/Imagegradient>
(Πρόσβαση: 23 Δεκεμβρίου 2013)
- [46] http://en.wikipedia.org/wiki/Sobel_operator
(Πρόσβαση: 23 Δεκεμβρίου 2013)
- [47] http://en.wikipedia.org/wiki/Prewitt_operator
(Πρόσβαση: 23 Δεκεμβρίου 2013)
- [48] http://en.wikipedia.org/wiki/Canny_edge_detector
(Πρόσβαση: 26 Αυγούστου 2014)
- [49] Duda, R. O., & Hart, P. E. (1972). Use of the Hough transformation to detect lines and curves in pictures. *Communications of the ACM*, 15(1), 11-15.

- [50] García-Garrido, M. Á., Sotelo, M. Á., & Martín-Gorostiza, E. (2005). Fast road sign detection using Hough transform for assisted driving of road vehicles. In *Computer Aided Systems Theory—EUROCAST 2005* (pp. 543-548). Springer Berlin Heidelberg.
- [51] Hatzidimos, J. (2004). Automatic traffic sign recognition in digital images. In *Proceedings of the International Conference on Theory and Applications of Mathematics and Informatics-ICTAMI* (pp. 174-184).
- [52] Hegadi, R. (2011). Automatic Traffic Sign Recognition. In *Proceedings of International Conference on “Communication, Computation, Management & Nanotechnology*.
- [53] Ballard, D. H. (1981). Generalizing the Hough transform to detect arbitrary shapes. *Pattern recognition*, 13(2), 111-122.
- [54] Barnes, N., Loy, G., Shaw, D., & Robles-Kelly, A. (2005). Regular polygon detection. In *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on* (Vol. 1, pp. 778-785). IEEE.
- [55] Ruta, A., Porikli, F., Watanabe, S., & Li, Y. (2011). In-vehicle camera traffic sign detection and recognition. *Machine Vision and Applications*, 22(2), 359-375.
- [56] Barrile, V., Meduri, G. M., & Cuzzocrea, D. (2012). Automatic Recognition of Road Signs by Hough Transform: Road-GIS. *Journal of Earth Science and Engineering*, 2(1), 42-50.
- [57] Loy, G., & Zelinsky, A. (2002). A fast radial symmetry transform for detecting points of interest. In *Computer Vision—ECCV 2002* (pp. 358-368). Springer Berlin Heidelberg.
- [58] Escalera, S., & Radeva, P. (2004). Fast greyscale road sign model matching and recognition. *Recent Advances in Artificial Intelligence Research and Development*, 2(1), 69-76.
- [59] Barnes, N., & Zelinsky, A. (2004). Real-time radial symmetry for speed sign detection. In *Intelligent Vehicles Symposium, 2004 IEEE* (pp. 566-571). IEEE.
- [60] Loy, G., & Barnes, N. (2004). Fast shape-based road sign detection for a driver assistance system. In *Intelligent Robots and Systems, 2004.(IROS 2004). Proceedings. 2004 IEEE/RSJ International Conference on*(Vol. 1, pp. 70-75). IEEE.
- [61] http://en.wikipedia.org/wiki/Template_matching
(Πρόσβαση: 6 Ιανουαρίου 2014)

- [62] Hsu, S. H., & Huang, C. L. (2001). Road sign detection and recognition using matching pursuit method. *Image and Vision Computing*, 19(3), 119-129.
- [63] Torresen, J., Bakke, J. W., & Sekanina, L. (2004). Efficient recognition of speed limit signs. In *Intelligent Transportation Systems* (pp. 652-656).
- [64] Fischler, M. A., & Bolles, R. C. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6), 381-395.
- [65] Escalera, S., Baró, X., Pujol, O., Vitrià, J., & Radeva, P. (2011). Background on traffic sign detection and recognition. In *Traffic-Sign Recognition Systems* (pp. 5-13). Springer London.
- [66] http://en.wikipedia.org/wiki/Machine_learning#cite_notearthur_samuel_machine (Πρόσβαση: 17 Ιανουαρίου 2014)
- [67] Mitchell, T. M. (1997). Machine learning. 1997. *Burr Ridge, IL: McGraw Hill*, 45.
- [68] <http://en.wikipedia.org/wiki/Cross-correlation> (Πρόσβαση: 17 Ιανουαρίου 2014)
- [69] Piccioli, G., De Micheli, E., & Campani, M. (1994). A robust method for road sign detection and recognition. In *Computer Vision—ECCV'94* (pp. 493-500). Springer Berlin Heidelberg.
- [70] http://el.wikipedia.org/wiki/%CE%9D%CE%B5%CF%85%CF%81%CF%89%CE%BD%CE%B9%CE%BA%CF%8C_%CE%B4%CE%AF%CE%BA%CF%84%CF%85%CE%BF (Πρόσβαση: 17 Ιανουαρίου 2014)
- [71] http://en.wikipedia.org/wiki/Artificial_neural_network (Πρόσβαση: 17 Ιανουαρίου 2014)
- [72] Διπλάρης, Σ. (2009). *Επεξεργασία εικόνας και εξαγωγή χαρακτηριστικών γνωρισμάτων*. Μεταπτυχιακή εργασία, Πανεπιστήμιο Πατρών, Πάτρα, Ελλάδα.
- [73] Μακρής, Α., (2005). *Αυτόματη αναγνώριση των ειδών μουσικής με χρήση μεθόδων μάθησης*. Διπλωματική εργασία, Πανεπιστήμιο Πατρών, Πάτρα Ελλάδα.
- [74] Basu, J. K., Bhattacharyya, D., & Kim, T. H. (2010). Use of artificial neural network in pattern recognition. *International journal of software engineering and its applications*, 4(2).
- [75] Nigrin, A. (1993). *Neural networks for pattern recognition*. MIT press.

- [76] Egmont-Petersen, M., de Ridder, D., & Handels, H. (2002). Image processing with neural networks—a review. *Pattern recognition*, 35(10), 2279-2301.
- [77] Medici, P., Caraffi, C., Cardarelli, E., Porta, P. P., & Ghisio, G. (2008, September). Real time road signs classification. In *Vehicular Electronics and Safety, 2008. ICVES 2008. IEEE International Conference on* (pp. 253-258). IEEE.
- [78] Nguwi, Y. Y., & Kouzani, A. Z. (2008). Detection and classification of road signs in natural environments. *Neural computing and applications*, 17(3), 265-289.
- [79] Kumar, S. (2004). *Neural networks: a classroom approach*. Tata McGraw-Hill Education.
- [80] Chang, K., & Swinson, T. (2011). Using Support Vector Machines to Classify Traffic Signs for an Autonomous Vehicle.
- [81] <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>
(Πρόσβαση: 20 Ιανουαρίου 2014)
- [82] Shi, M., Wu, H., & Fleyeh, H. (2008, June). Support vector machines for traffic signs recognition. In *Neural Networks, 2008. IJCNN 2008. (IEEE World Congress on Computational Intelligence). IEEE International Joint Conference on* (pp. 3820-3827). IEEE.
- [83] Khotanzad, A., & Hong, Y. H. (1990). Invariant image recognition by Zernike moments. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 12(5), 489-497.
- [84] HUa, X., ZHUa, X., Lia, D., & LI, H. (2010). Traffic sign recognition using Scale invariant feature transform and SVM. In *A special joint symposium of ISPRS Technical Commission IV & AutoCarto in conjunction with ASPRS/ CaGIS Fall Specialty Conference November* (pp. 15-19).
- [85] Lowe, D. G. (1999). Object recognition from local scale-invariant features. In *Computer vision, 1999. The proceedings of the seventh IEEE international conference on* (Vol. 2, pp. 1150-1157). IEEE.
- [86] Καλησπεράκης, Η (2010). *Η επιπολική γεωμετρία στον προβολικό και στον ευκλείδειο χώρο*. Διδακτορική Διατριβή, Εθνικό Μετσόβιο Πολυτεχνείο, Αθήνα, Ελλάδα.
- [87] http://en.wikipedia.org/wiki/Bag-of-words_model_in_computer_vision
(Πρόσβαση: 22 Ιανουαρίου 2014)
- [88] Sivic, J., Russell, B. C., Efros, A. A., Zisserman, A., & Freeman, W. T. (2005). Discovering objects and their location in images. In *Computer Vision, 2005*.

- ICCV 2005. Tenth IEEE International Conference on* (Vol. 1, pp. 370-377). IEEE.
- [89] http://cs.nyu.edu/~fergus/teaching/vision_2012/9_BoW.pdf
(Πρόσβαση: 22 Ιανουαρίου 2014)
- [90] Kouzani, A. Z. (2007). Road-sign identification using ensemble learning. In *Intelligent Vehicles Symposium, 2007 IEEE* (pp. 438-443). IEEE.
- [91] Zaklouta, F., Stanculescu, B., & Hamdoun, O. (2011). Traffic sign classification using kd trees and random forests. In *Neural Networks (IJCNN), The 2011 International Joint Conference on* (pp. 2151-2155). IEEE.
- [92] Greenhalgh, J., & Mirmehdi, M. (2012). Traffic sign recognition using MSER and Random Forests. In *Signal Processing Conference (EUSIPCO), 2012 Proceedings of the 20th European* (pp. 1935-1939). IEEE.
- [93] Zheng, Y. J., Ritter, W., & Janssen, R. (1994). An adaptive system for traffic sign recognition. In *Intelligent Vehicles' 94 Symposium, Proceedings of the* (pp. 165-170). IEEE.
- [94] Vapnik, V., & Kotz, S. (2006). *Estimation of dependences based on empirical data*. Springer.
- [95] Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3), 273-297.
- [96] Harman, G., & Kulkarni, S. (2012). Statistical Learning Theory and Induction. In *Encyclopedia of the Sciences of Learning* (pp. 3186-3188). Springer US.
- [97] Burges, C. J. (1998). A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery*, 2(2), 121-167.
- [98] Σπύρου, Ε. (2003). *Μηχανές διανυσμάτων στήριξης με χρήση πυρήνα ασαφών βασικών συναρτήσεων*. Διπλωματική εργασία, Εθνικό Μετσόβιο Πολυτεχνείο, Αθήνα Ελλάδα
- [99] http://research.cs.tamu.edu/prism/lectures/pr/pr_121.pdf
(Πρόσβαση: 25 Ιανουαρίου 2014)
- [100] Shi, M. (2006). *Road and traffic sign recognition using support vector machines*. Master thesis, Dalarna University, Högscolegatan, Falun, Sweden.
- [101] http://www.cse.iitb.ac.in/~tarung/smt/papers_ppt/ency-cross-validation.pdf
(Πρόσβαση: 25 Ιανουαρίου 2014)
- [102] Schölkopf, B., Smola, A. J., Williamson, R. C., & Bartlett, P. L. (2000). New support vector algorithms. *Neural computation*, 12(5), 1207-1245.

- [103] Hsu, C. W., & Lin, C. J. (2002). A comparison of methods for multiclass support vector machines. *Neural Networks, IEEE Transactions on*, 13(2), 415-425.
- [104] Alefs, B., Eschemann, G., Ramoser, H., & Beleznai, C. (2007). Road sign detection from edge orientation histograms. In *Intelligent Vehicles Symposium, 2007 IEEE* (pp. 993-998). IEEE.
- [105] Καραγύργης, Α. (2006). *Μελέτη και ανάπτυξη αλγορίθμων για την ανάπτυξη βρογχοσκοπικών εικόνων σε πραγματικό χρόνο*. Διπλωματική εργασία, Εθνικό Μετσόβιο Πολυτεχνείο, Αθήνα, Ελλάδα.
- [106] Σκορδάς, Ε. (2009). *Αναγνώριση προπορευόμενου οχήματος με ψηφιακή επεξεργασία εικόνας*. Διπλωματική εργασία, Πανεπιστήμιο Πατρών, Πάτρα, Ελλάδα.
- [107] [http://en.wikipedia.org/wiki/Filter_\(signal_processing\)](http://en.wikipedia.org/wiki/Filter_(signal_processing))
(Πρόσβαση: 20 Ιανουαρίου 2014)
- [108] http://www.teiser.gr/icd/staff/simeonid/psee/Noise_Mean_Median.pdf
(Πρόσβαση: 21 Ιανουαρίου 2014)
- [109] Τελώνη, Ε. (2011). *Διερεύνηση ανίχνευσης ακτογραμμών από τηλεπισκοπικά δεδομένα και ανάπτυξη μεθοδολογίας με έμφαση σε προηγμένες τεχνικές ομαλοποίησης και τον τελεστή Canny*. Διπλωματική εργασία, Εθνικό Μετσόβιο Πολυτεχνείο, Αθήνα, Ελλάδα.
- [110] Najman, L., & Ronse, C. (Eds.). (2005). *Mathematical Morphology: 40 Years On*. Springer.
- [111] <https://www.cs.auckland.ac.nz/courses/compsci773s1c/lectures/Image-Processing-html/topic4.htm>
(Πρόσβαση: 23 Ιανουαρίου 2014)
- [112] Casas, J.R.. (1996). *Image Compression based on Perceptual Coding Techniques*. Doctoral dissertation, University of Catalunya, Barcelona, Spain.
- [113] Γάτος, Β. (2002). *Κατάτμηση σελίδας εγγράφων*. Σημειώσεις μαθήματος Ψηφιακή επεξεργασία και αναγνώριση εγγράφων, Εθνικό Μετσόβιο Πολυτεχνείο, Αθήνα, Ελλάδα.
- [114] <http://www.cse.msu.edu/~stockman/Book/ch3.pdf>
(Πρόσβαση: 23 Ιανουαρίου 2014)
- [115] Canny, J. (1986). A computational approach to edge detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, (6), 679-698.

- [116] Rhody, H. (2005). Hough Circle Transform. Notes, Chester F. Carlson Center for Imaging Science, Rochester Institute of Technology, Rochester, New York, U.S.A.
- [117] Yuen, H. K., Princen, J., Illingworth, J., & Kittler, J. (1990). Comparative study of Hough transform methods for circle finding. *Image and Vision Computing*, 8(1), 71-77.
- [118] Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on* (Vol. 1, pp. 886-893). IEEE.
- [119] Dalal, N. (2006). *Finding people in images and videos* Doctoral dissertation, Institut National Polytechnique de Grenoble-INPG, Grenoble, France.
- [120] Joachims, T. (1999). Making large scale SVM learning practical.
- [121] <http://opencv.org/>
(Πρόσβαση: 4 Μαρτίου 2014)
- [122] <http://en.wikipedia.org/wiki/OpenCV>
(Πρόσβαση: 4 Μαρτίου 2014)
- [123] Joachims, T. (2000). Estimating the generalization performance of a SVM efficiently.
- [124] Shneier, M. (2006). Road sign detection and recognition. In *Defense and Security Symposium* (pp. 623016-623016). International Society for Optics and Photonics.

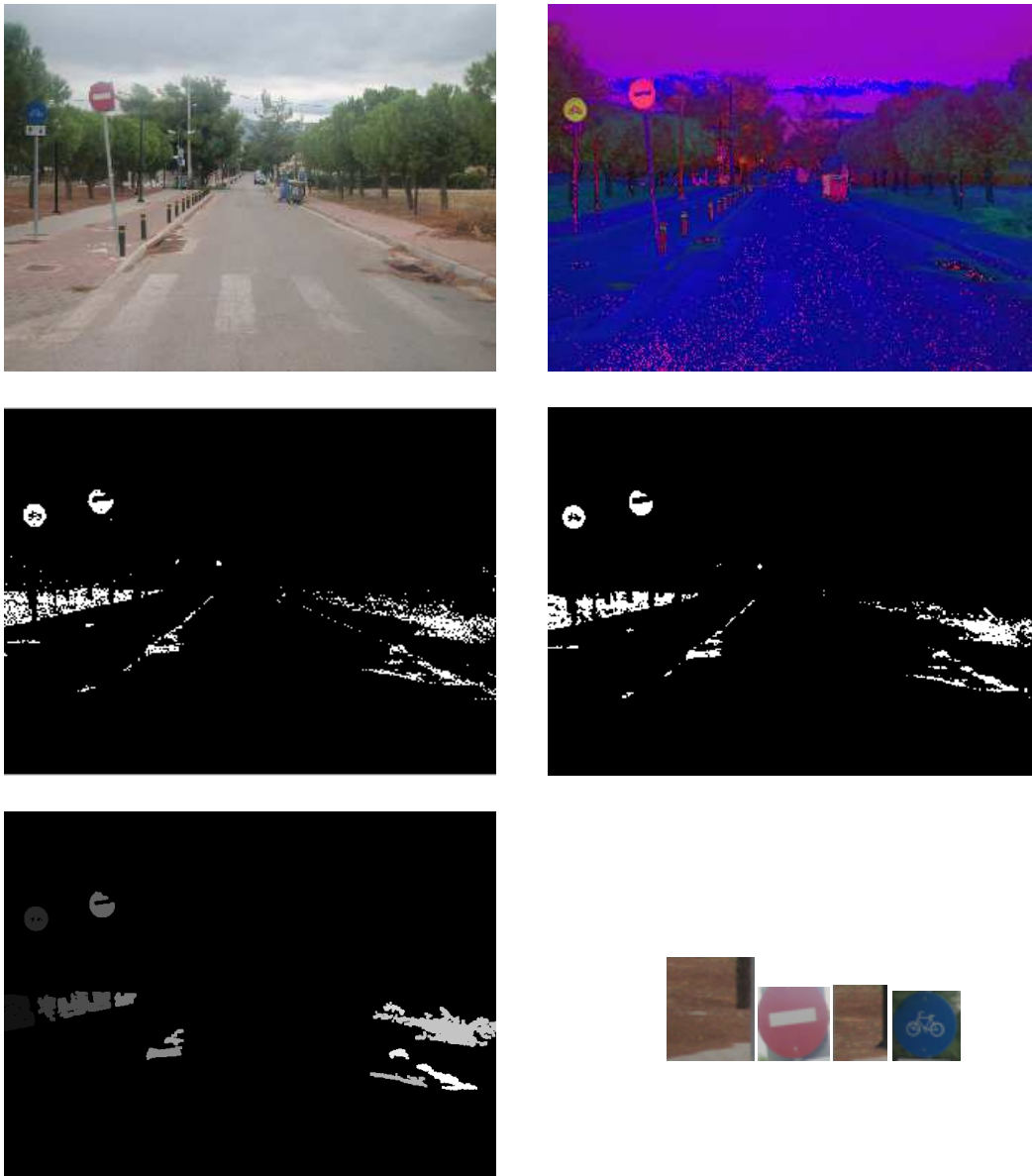
Παράρτημα Α

Αποτελέσματα εντοπισμού με βάση το χρώμα

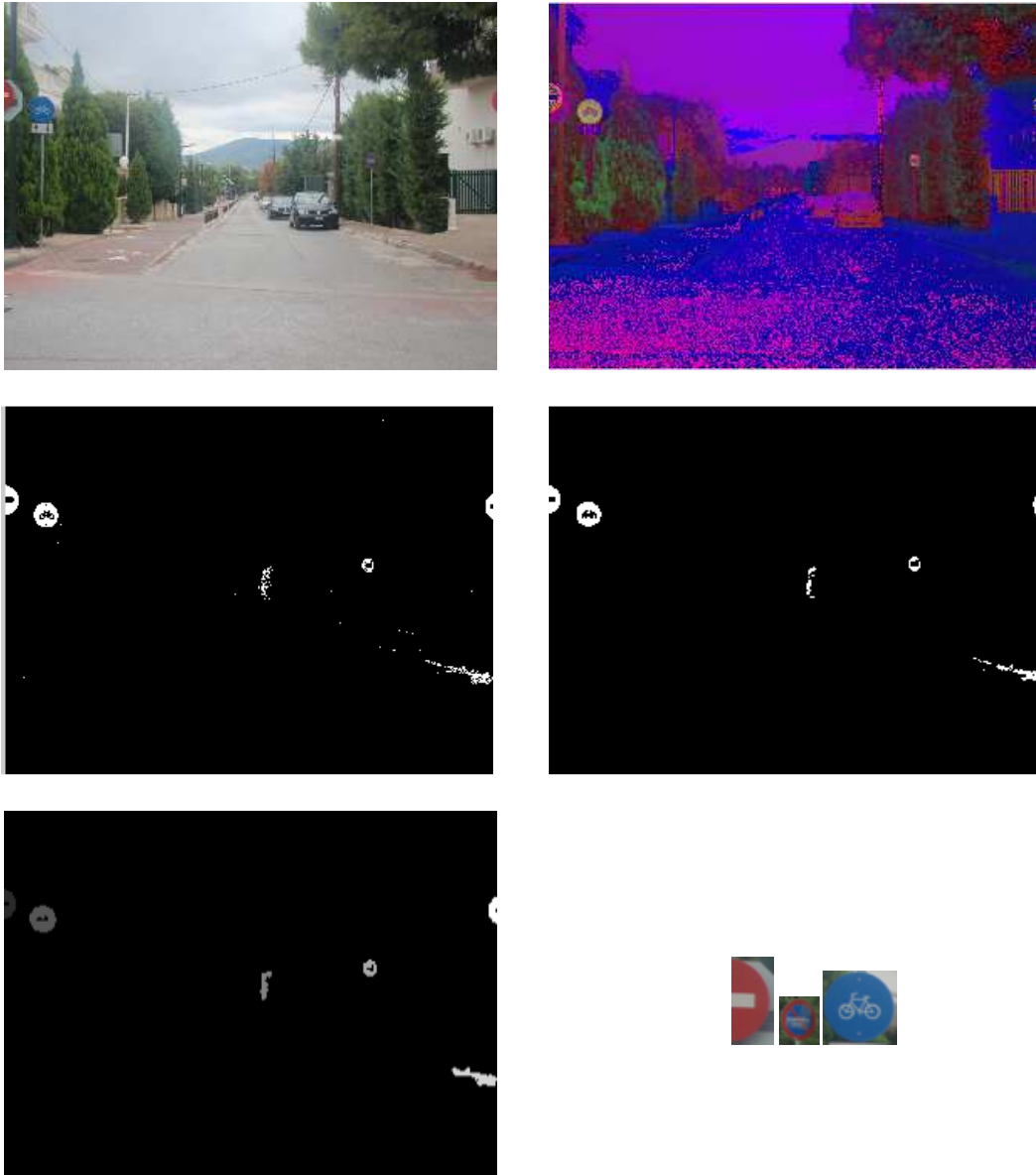
Παρατίθενται τα αποτελέσματα από τον εντοπισμό με βάση το χρώμα και για τις 42 εικόνες, που χρησιμοποιήθηκαν ως δεδομένα στην εφαρμογή (κεφάλαιο 5). Για κάθε εικόνα δίνεται η ίδια η εικόνα, η μετατροπή της στον χρωματικό χώρο HSI, τα αποτελέσματα της κατωφλίωσης, τα αποτελέσματα εφαρμογής του φίλτρου μεσαίας τιμής και διαστολής με δομικό στοιχείο με σχήμα τετράγωνο, οι συνδεδεμένες περιοχές κάθε μία με διαφορετικό χρώμα, και οι περιοχές ενδιαφέροντος που έγιναν τελικά δεκτές με βάση το εμβαδό τους και τον λόγο του επαπτόμενου τετραπλεύρου, όπως αυτές αποκόπηκαν από την αρχική εικόνα.



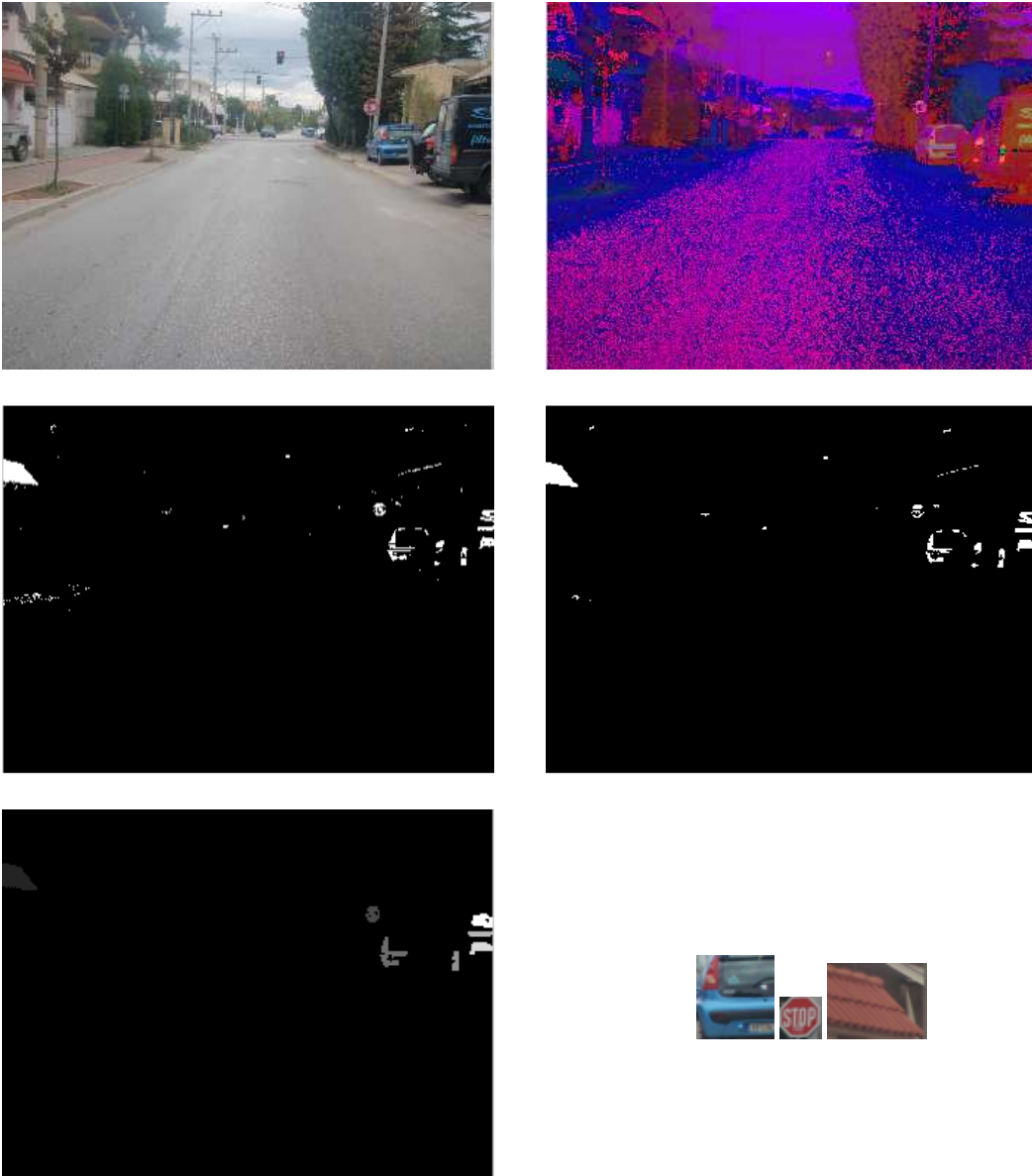
Εικόνα 1



Εικόνα 2



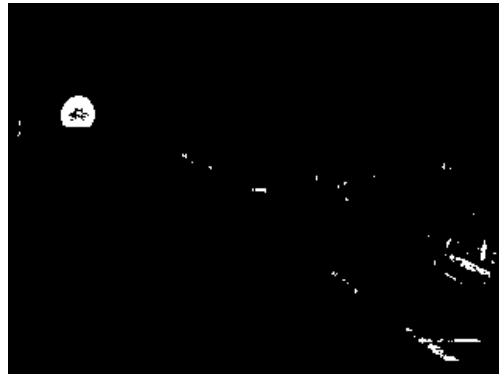
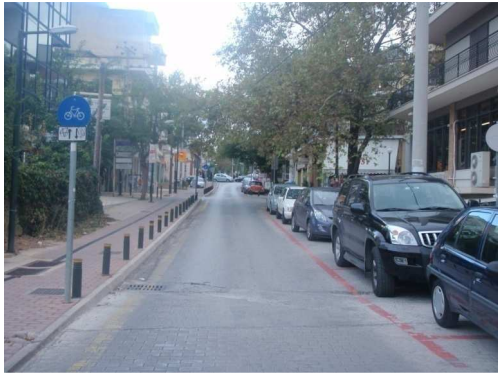
Εικόνα 3



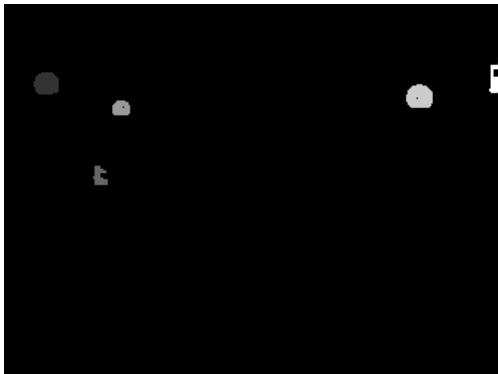
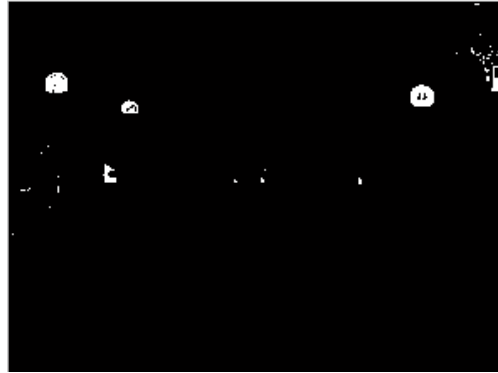
Εικόνα 4



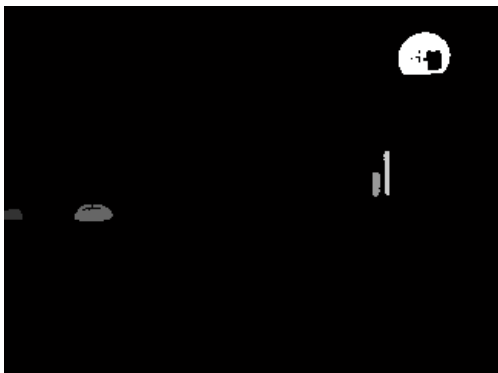
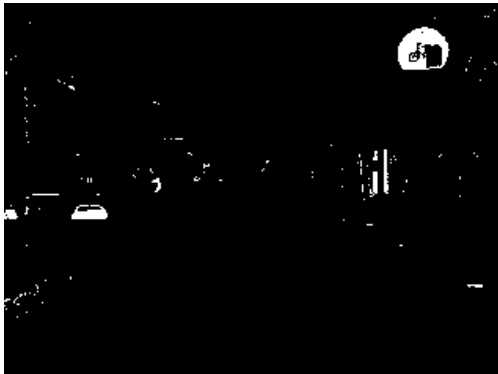
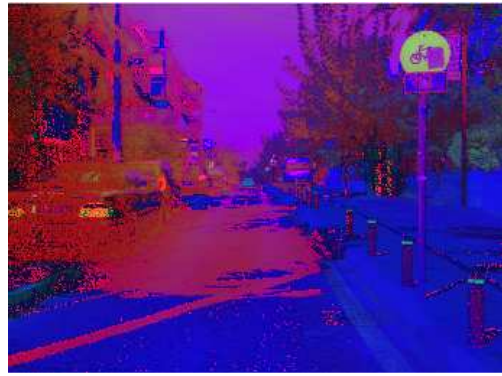
Εικόνα 5



Εικόνα 6



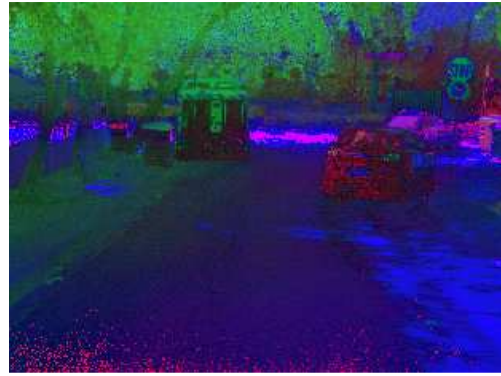
Εικόνα 7



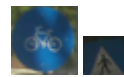
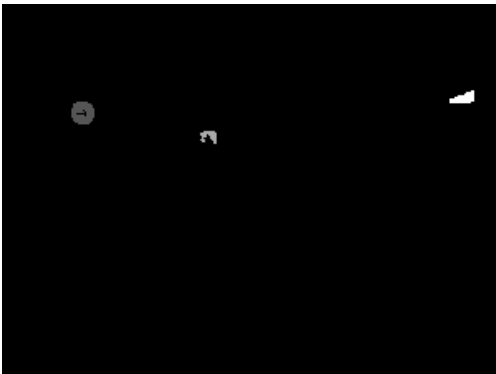
Εικόνα 8



Εικόνα 9



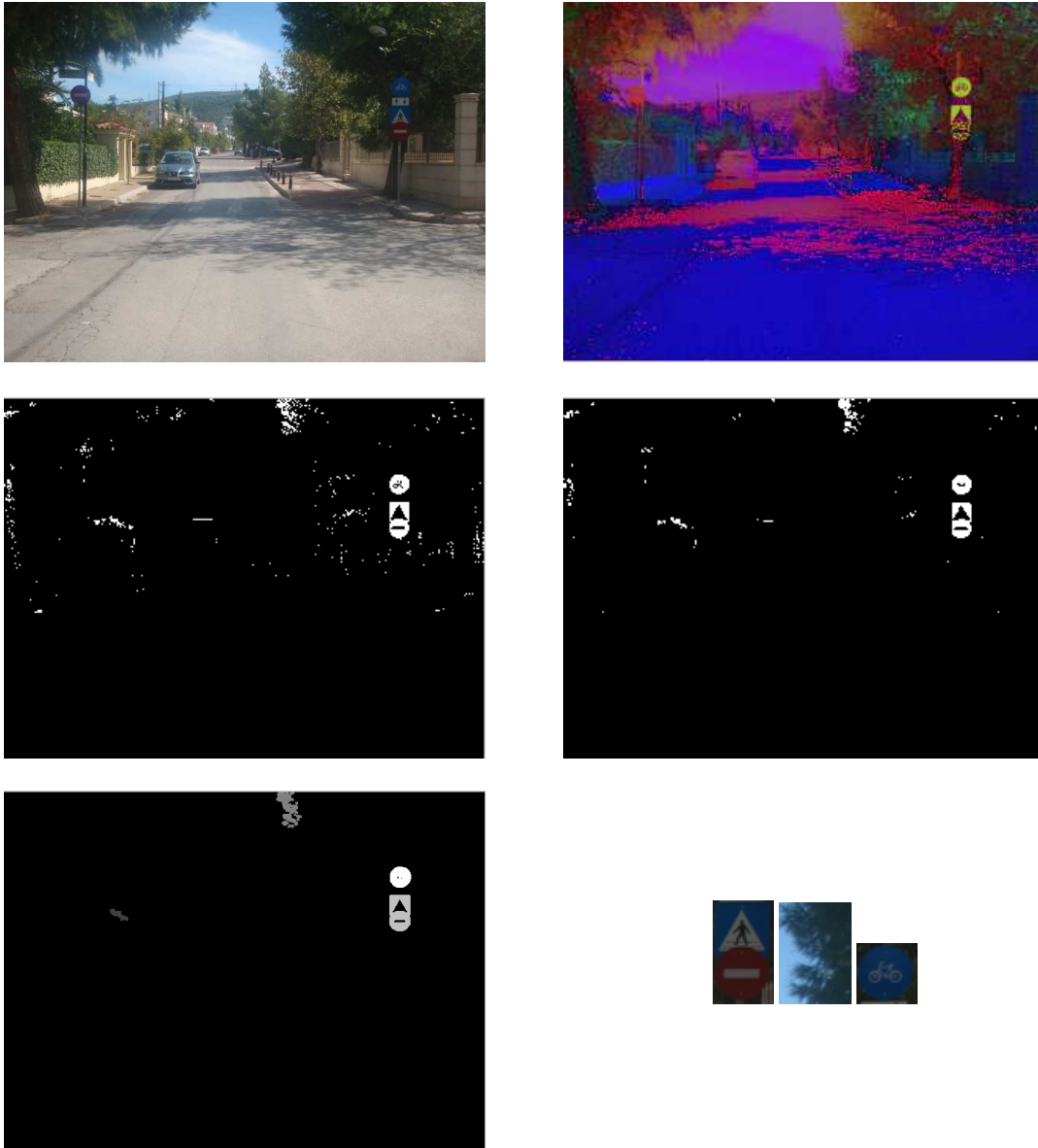
Εικόνα 10



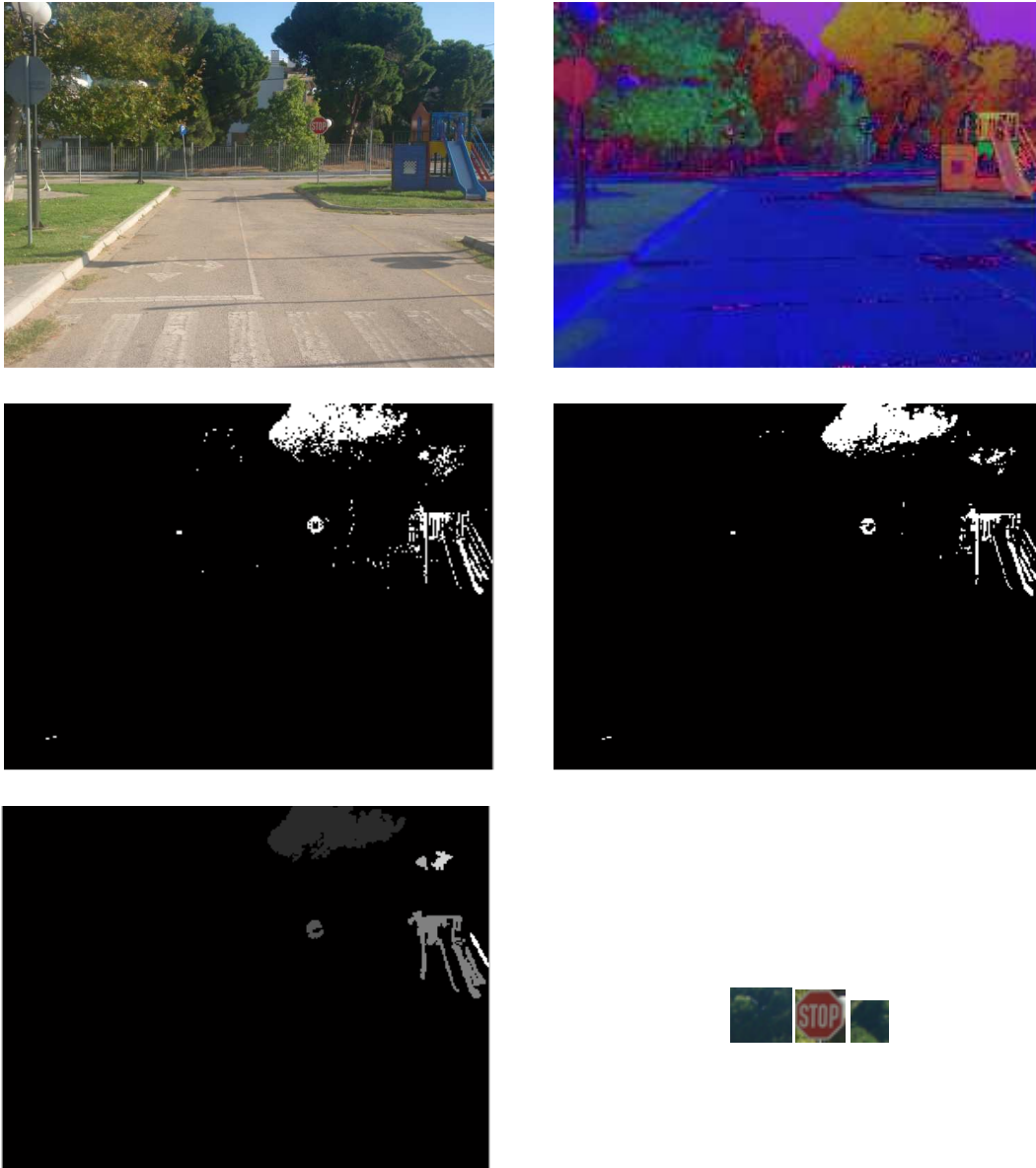
Εικόνα 11



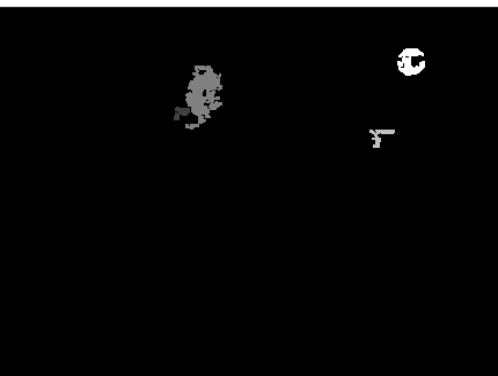
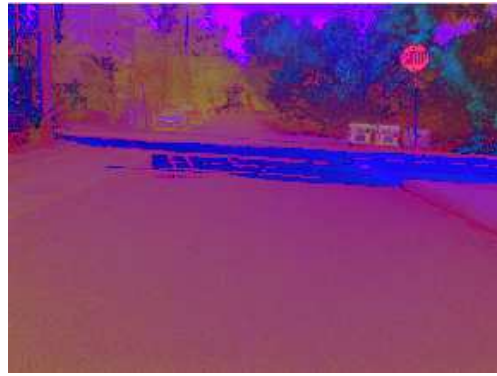
Εικόνα 12



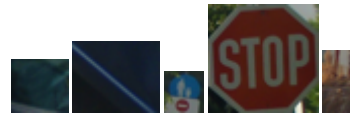
Εικόνα 13



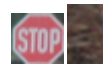
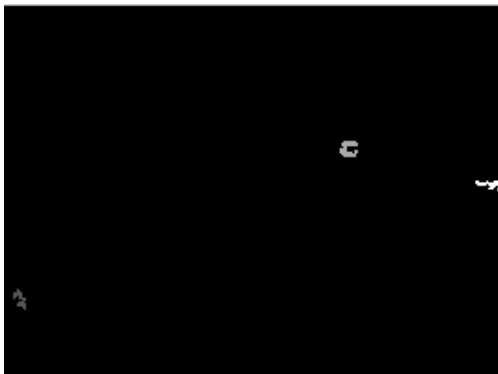
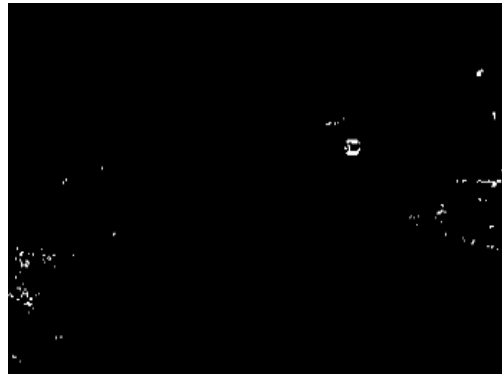
Εικόνα 14



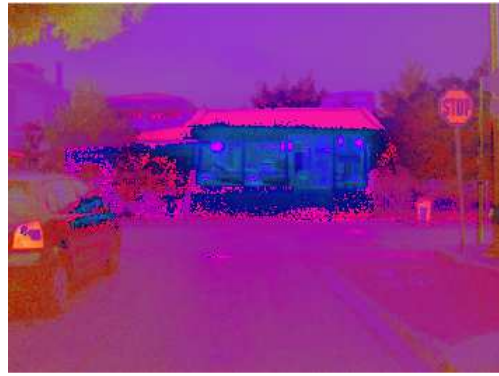
Εικόνα 15



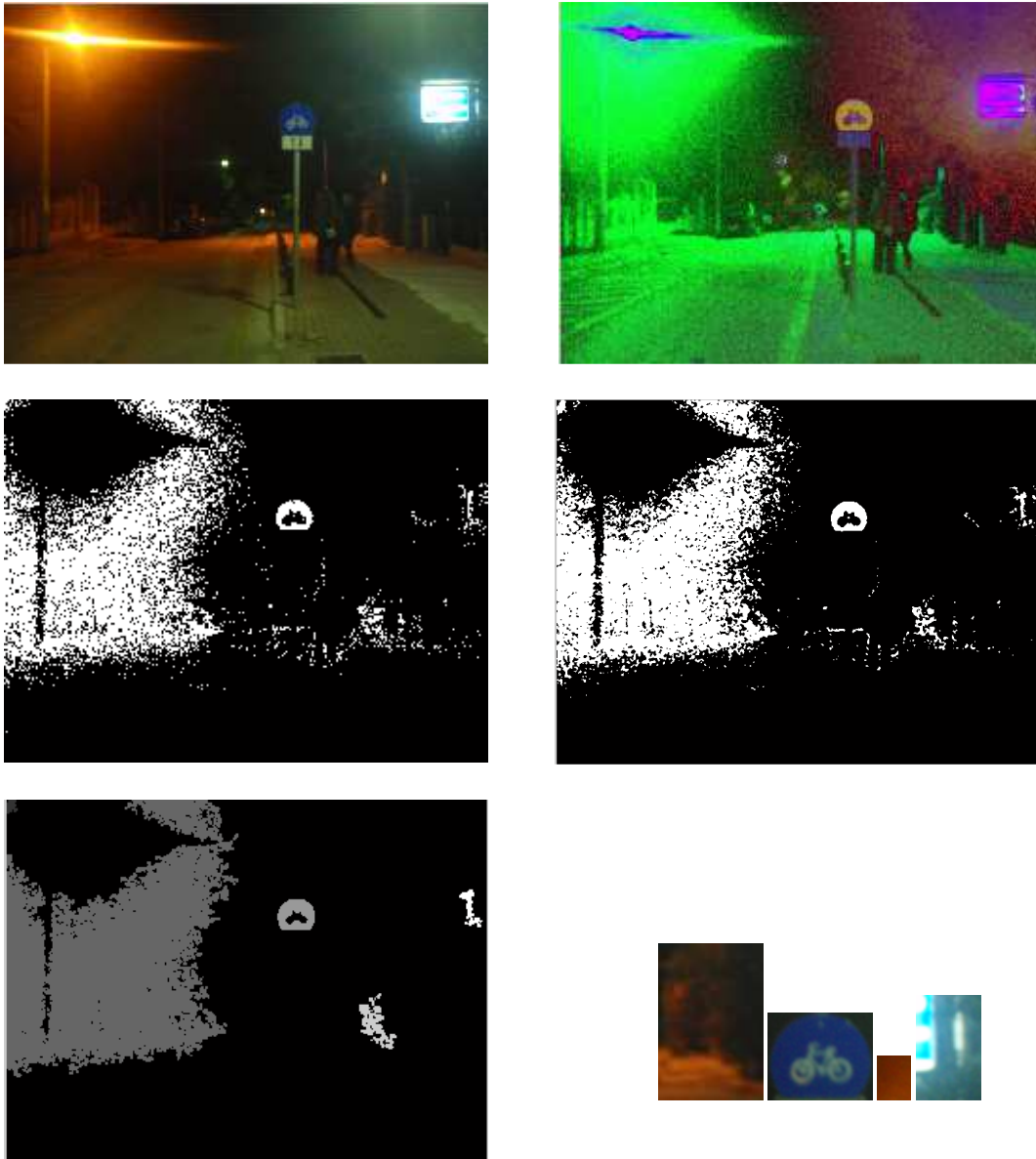
Εικόνα 16



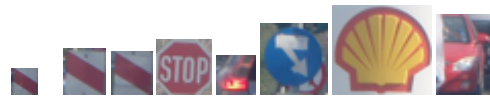
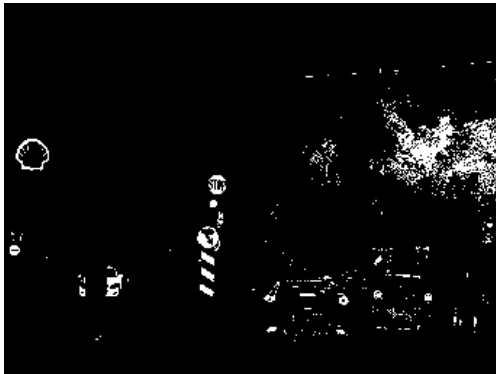
Εικόνα 17



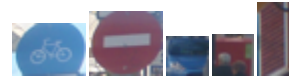
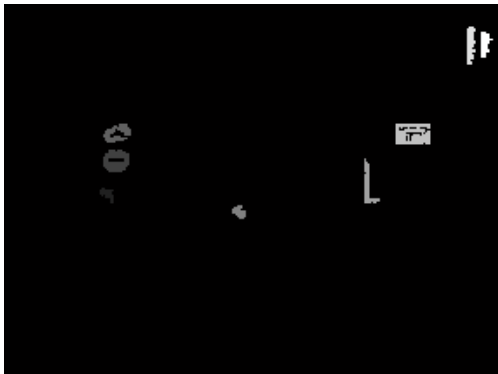
Εικόνα 18



Εικόνα 19



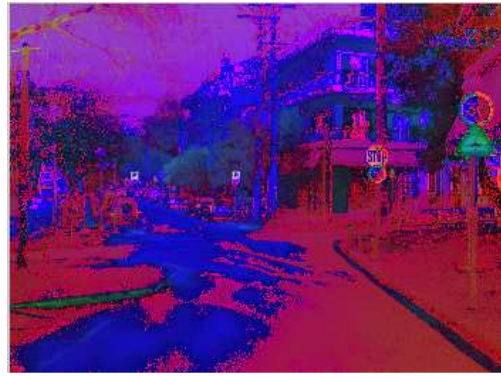
Εικόνα 20



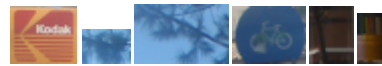
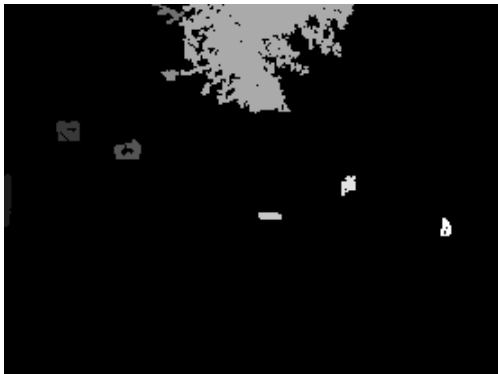
Εικόνα 21



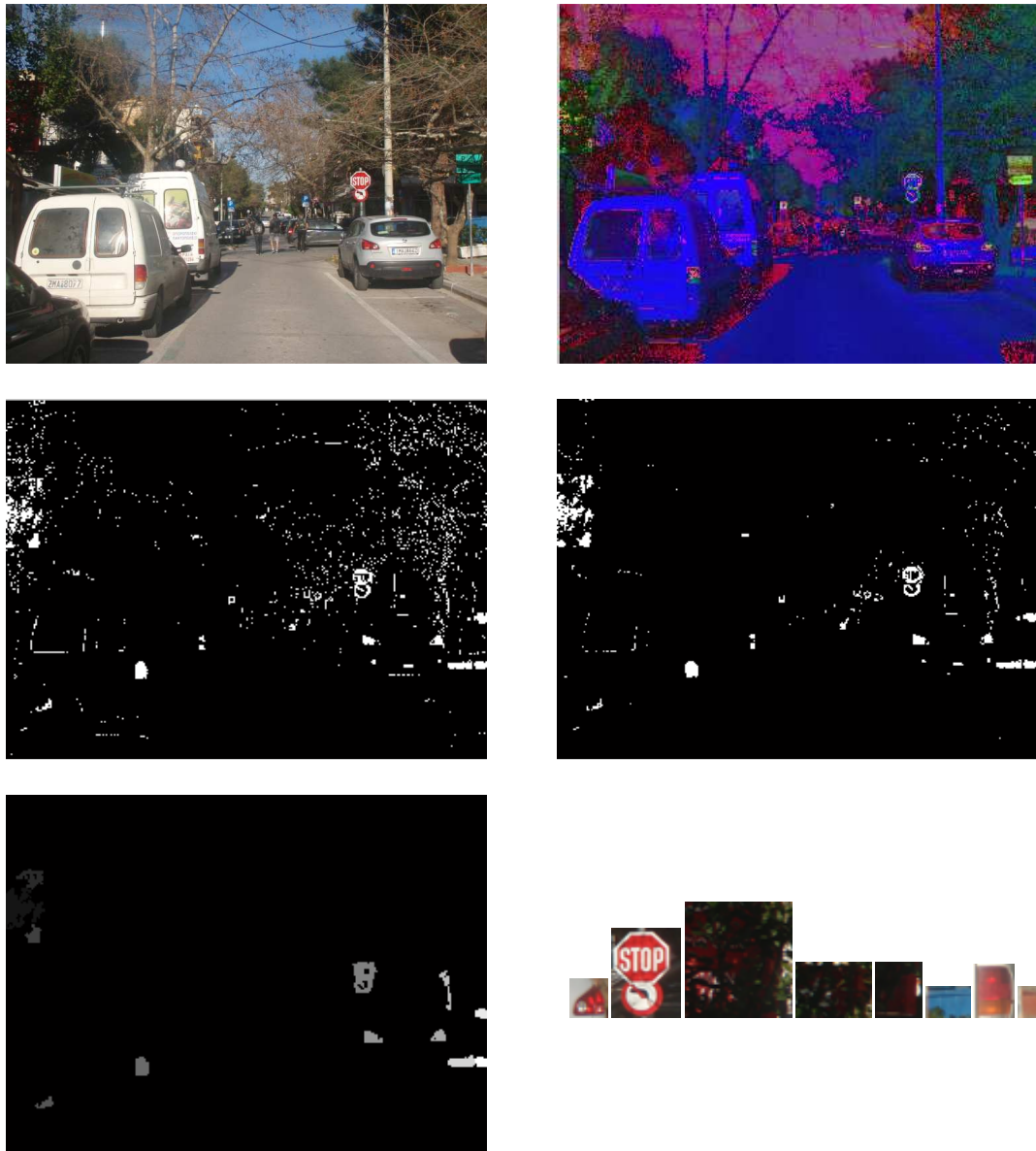
Εικόνα 22



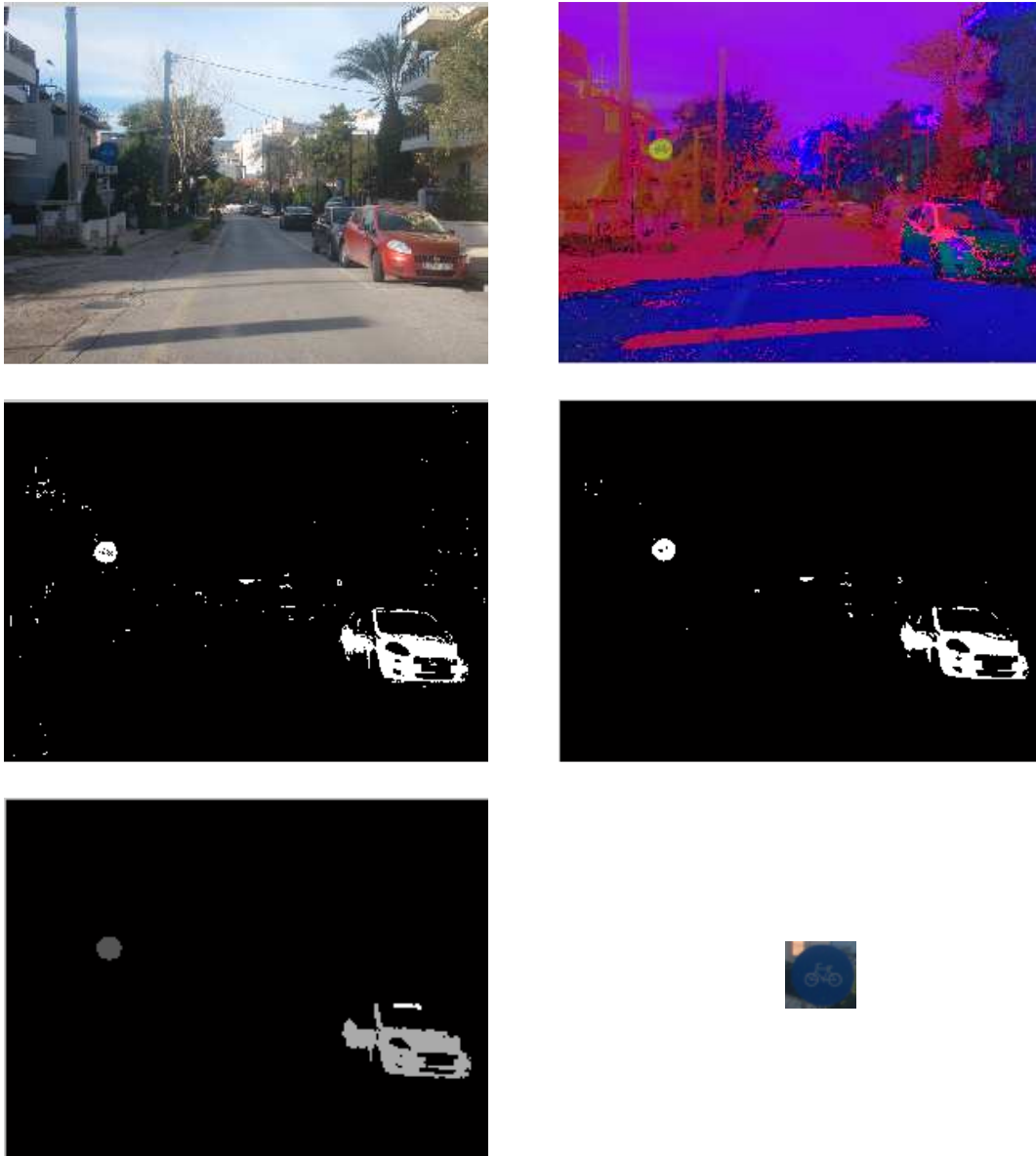
Εικόνα 23



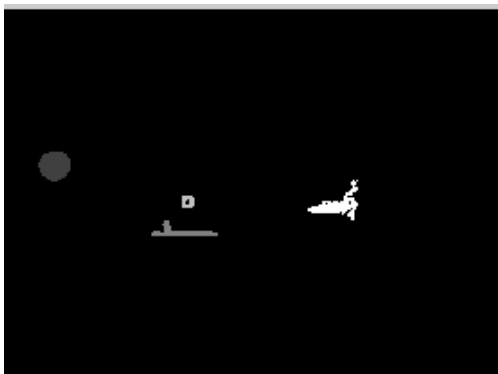
Εικόνα 24



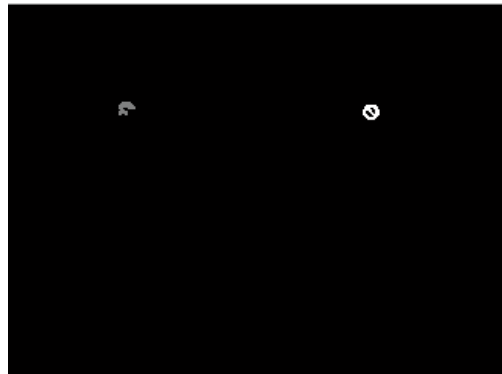
Εικόνα 25



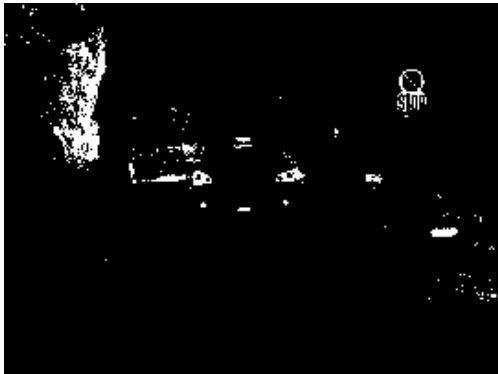
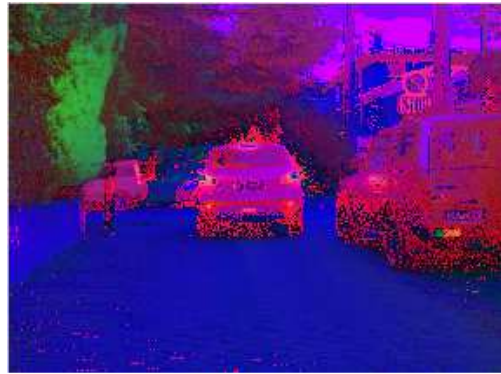
Εικόνα 26



Εικόνα 27



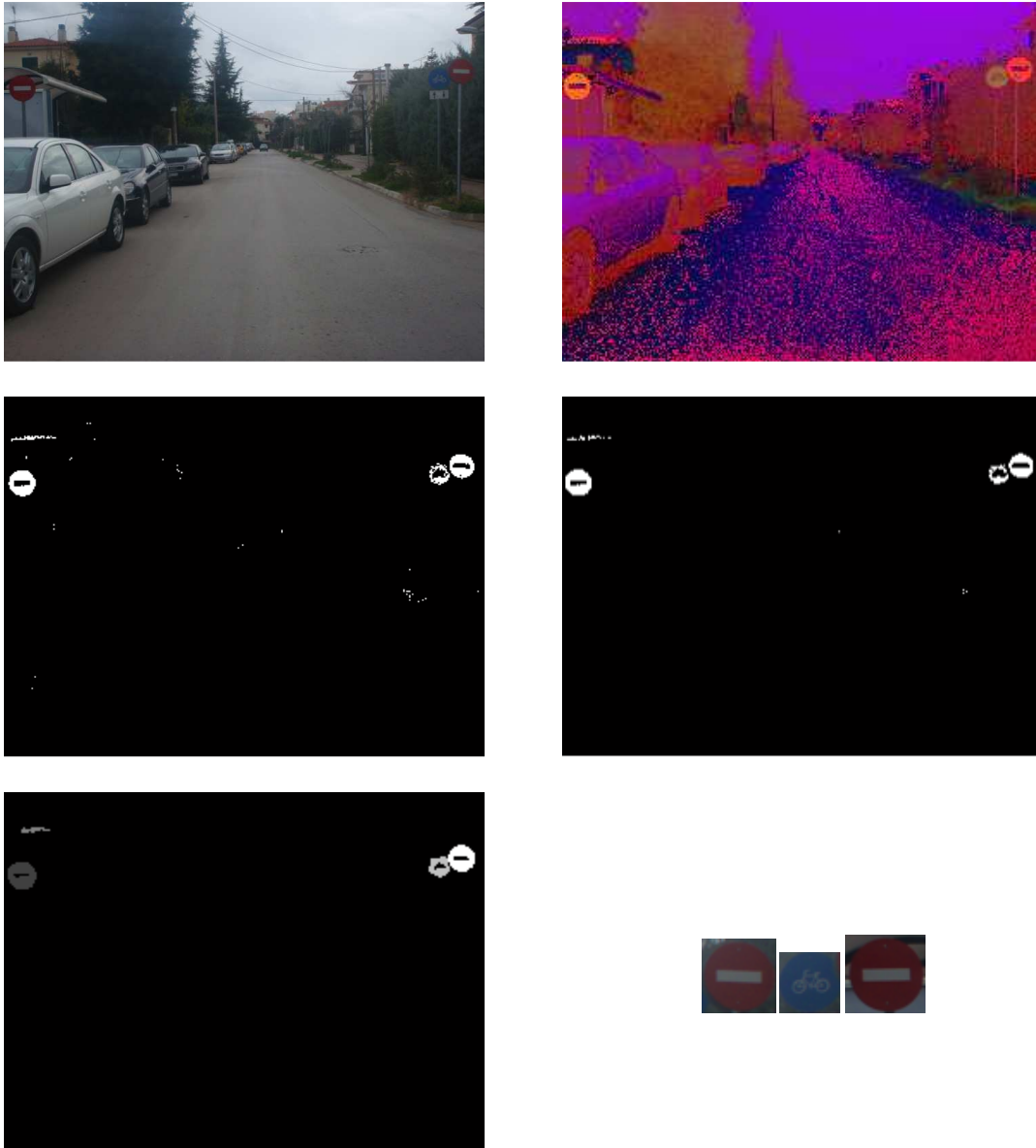
Εικόνα 28



Εικόνα 29



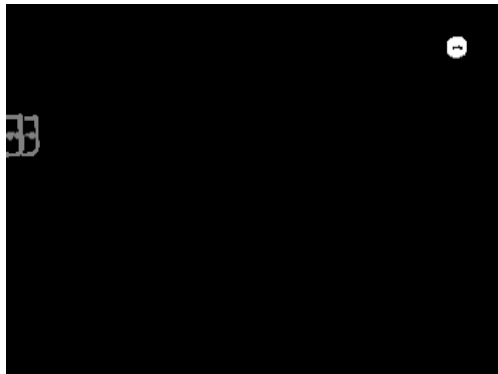
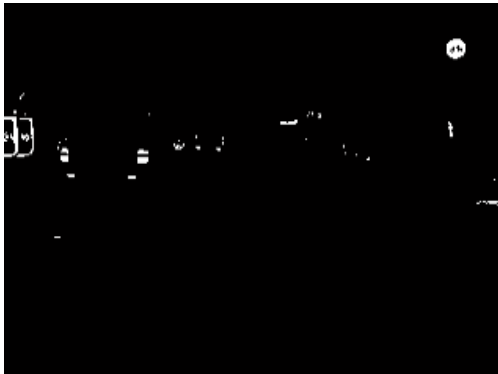
Εικόνα 30



Εικόνα 31



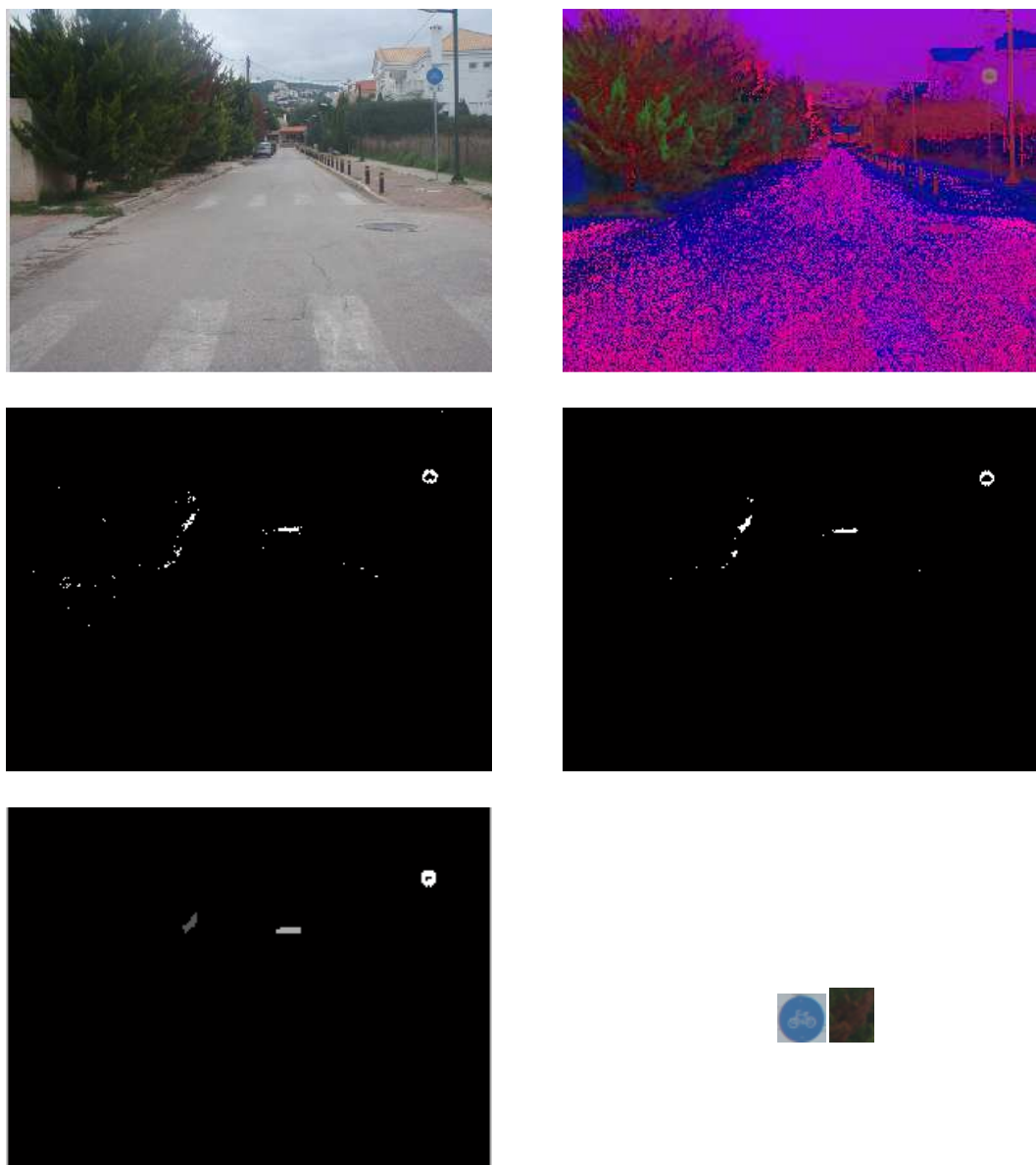
Εικόνα 32



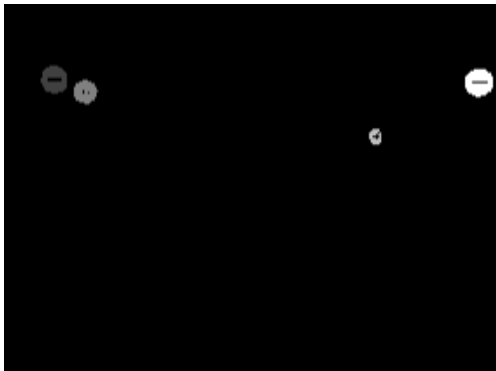
Εικόνα 33



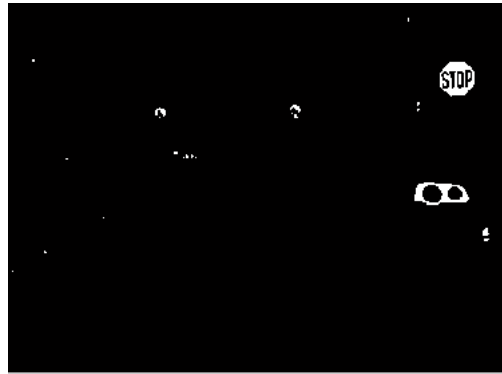
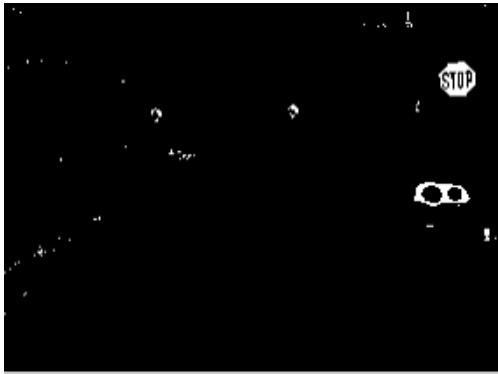
Εικόνα 34



Εικόνα 35



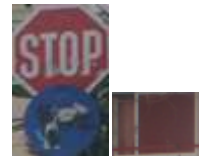
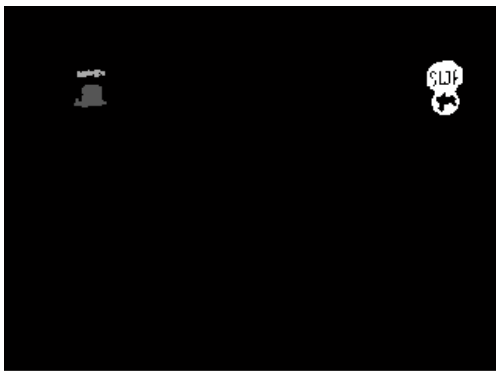
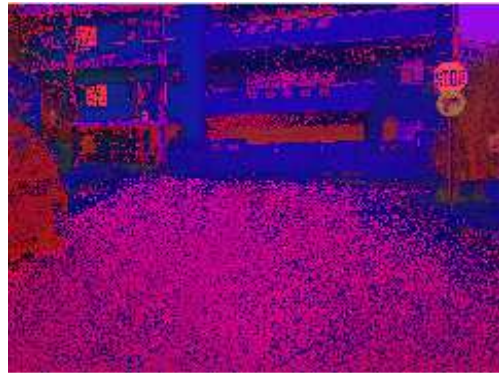
Εικόνα 36



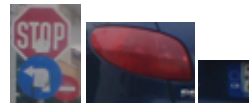
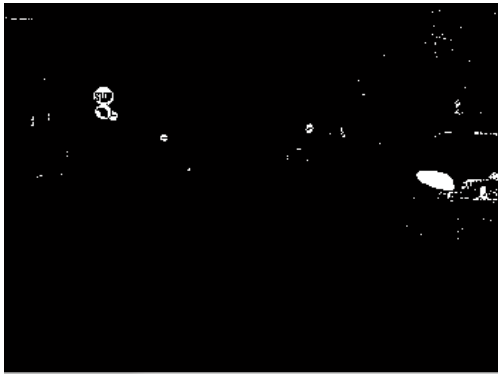
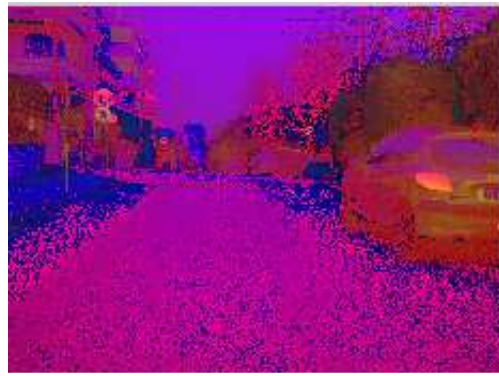
Εικόνα 37



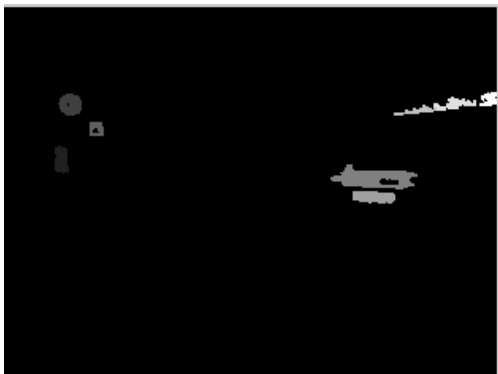
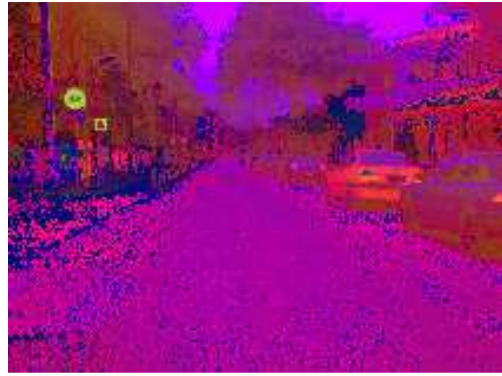
Εικόνα 38



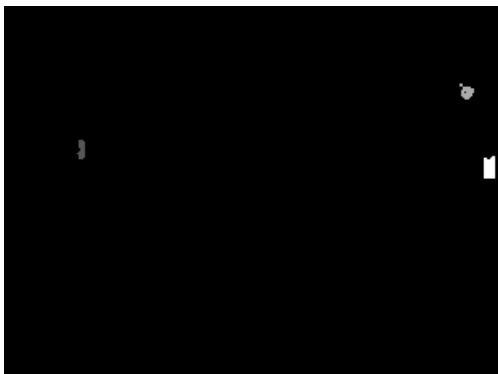
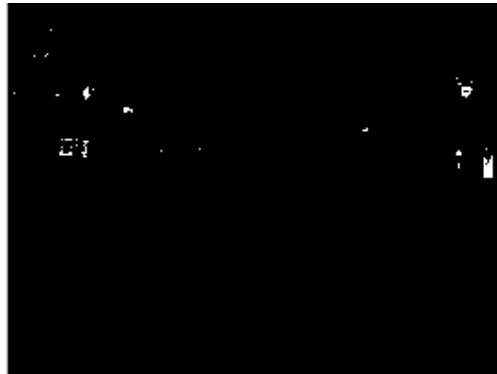
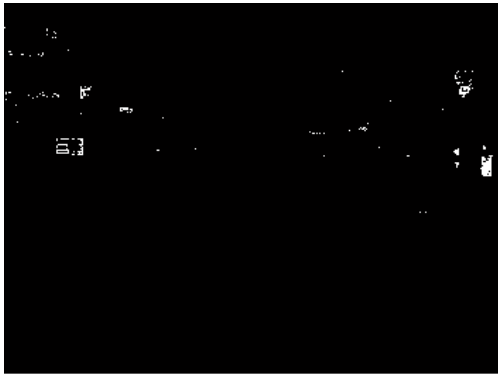
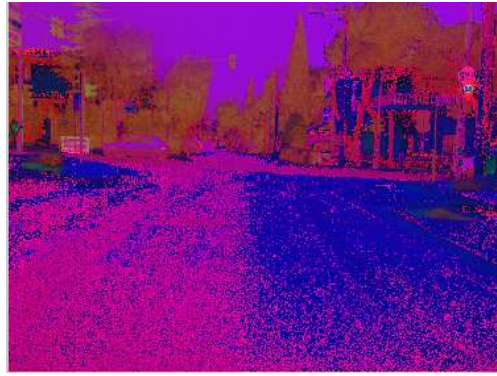
Εικόνα 39



Εικόνα 40



Εικόνα 41



Εικόνα 42

Παράρτημα Β

Αποτελέσματα εκπαίδευσης

Δίνονται αναλυτικά τα αποτελέσματα εκπαίδευσης για τα έξι σύνολα δεδομένων που χρησιμοποιήθηκαν και για τις δύο συναρτήσεις πυρήνα (γραμμική και πολυωνυμική). Για κάθε περίπτωση το λογισμικό επιστρέφει τον αριθμό των διανυσμάτων στήριξης (number of VC) και πόσα από αυτά εντοπίστηκαν στο άνω όριο (upper bound, δηλαδή $\alpha_i = C$).

Επιστρέφει επίσης την προκαθορισμένη τιμή για το κόστος C (default regularization parameter) για κάθε συνάρτηση πυρήνα, τον αριθμό των επαναλήψεων που πραγματοποιήθηκαν για τον υπολογισμό του βέλτιστου υπερεπίπεδου (iterations), τον αριθμό των παραδειγμάτων που κατηγοριοποιήθηκαν λάθος (misclassified) και τον χρόνο επεξεργασίας των δεδομένων (runtime in CPU).

Άλλα αποτελέσματα που καταγράφονται είναι η τιμή της συνάρτησης κόστους L_1 (L_1 loss), το μέτρο του διανύσματος $\|\mathbf{w}\|$ (norm of weight vector) καθώς και το μέτρο του διανύσματος για το πιο απομακρυσμένο παράδειγμα $\|\mathbf{x}\|$ (norm of longest example).

Τέλος, σημειώνονται το μέγεθος της διάστασης VC (estimated VC dimension) και οι δείκτες Χι-Άλφα (για τιμή του $\rho=1$ XiAlpha estimate of the error, XiAlpha estimate of the recall, XiAlpha estimate of the precision).

Όλα τα παραπάνω δεν απαιτούν τη διαδικασία της εσωτερικής εξακρίβωσης για να υπολογιστούν. Καθώς στα συγκεκριμένα παραδείγματα οι βέλτιστες τιμές των παραμέτρων υπολογίστηκαν με τη διαδικασία της εσωτερικής εξακρίβωσης καταγράφονται επίσης οι παραπάνω δείκτες όπως αυτοί προέκυψαν από τη διαδικασία της εσωτερικής εξακρίβωσης με τη μέθοδο Leave-one-out (Leave-one-out estimate of the error, Leave-one-out estimate of the recall, Leave-one-out estimate of the precision) και ο αριθμός των επαναλήψεων που πραγματοποιήθηκαν για τη διαδικασία της εσωτερικής εξακρίβωσης (Actual Leave-one-outs computed).

Αποτελέσματα για το σήμα του ποδηλατοδρόμου:

```
C:\Users\Katerina\Desktop\SUM_light>svm_learn -x 1 train_podilato_25.txt podilato_linear
Scanning examples...done
Reading examples into memory...100%.200%.OK. (217 examples read)
Setting default regularization parameter C=0.0157
Optimizing.....done. (31 iterations)
Optimization finished (13 misclassified, maxdiff=0.00088).
Runtime in cpu-seconds: 0.02
Number of SU: 148 (including 118 at upper bound)
L1 loss: loss=54.06542
Norm of weight vector: |w|=1.09000
Norm of longest example vector: |x|=7.97982
Estimated UCdim of classifier: UCdim<=76.65586
Computing X1alpha-estimates...done
Runtime for X1alpha-estimates in cpu-seconds: 0.00
X1alpha-estimate of the error: error<=54.38% (rho=1.00,depth=0)
X1alpha-estimate of the recall: recall=>58.21% (rho=1.00,depth=0)
X1alpha-estimate of the precision: precision=>55.71% (rho=1.00,depth=0)
```

```
Retrain on full problem.....done.
Leave-one-out estimate of the error: error=9.68%
Leave-one-out estimate of the recall: recall=99.25%
Leave-one-out estimate of the precision: precision=86.93%
Actual leave-one-outs computed: 105 (rho=1.00)
Runtime for leave-one-out in cpu-seconds: 0.28
Writing model file...done
```

1α. Με γραμμική συνάρτηση πυρήνα και 217 δεδομένα εκπαίδευσης

```
C:\Users\Katerina\Desktop\SUM_light>svm_learn -x 1 -t 1 train_podilato_25.txt podilato_polynomial
Scanning examples...done
Reading examples into memory...100%.200%.OK. (217 examples read)
Setting default regularization parameter C=0.0000
Optimizing.....done. (44 iterations)
Optimization finished (0 misclassified, maxdiff=0.00093).
Runtime in cpu-seconds: 0.03
Number of SU: 136 (including 67 at upper bound)
L1 loss: loss=18.00975
Norm of weight vector: |w|=0.01713
Norm of longest example vector: |x|=520.15107
Estimated UCdim of classifier: UCdim<=80.42513
Computing X1alpha-estimates...done
Runtime for X1alpha-estimates in cpu-seconds: 0.01
X1alpha-estimate of the error: error<=30.88% (rho=1.00,depth=0)
X1alpha-estimate of the recall: recall=>75.37% (rho=1.00,depth=0)
X1alpha-estimate of the precision: precision=>74.81% (rho=1.00,depth=0)
```

```
Retrain on full problem.....done.
Leave-one-out estimate of the error: error=4.15%
Leave-one-out estimate of the recall: recall=97.01%
Leave-one-out estimate of the precision: precision=96.30%
Actual leave-one-outs computed: 67 (rho=1.00)
Runtime for leave-one-out in cpu-seconds: 0.24
Writing model file...done
```

1β. Με πολυωνυμική συνάρτηση πυρήνα και 217 δεδομένα εκπαίδευσης

```
C:\Users\Katerina\Desktop\SUMlight2>svm_learn -x 1 train_podilato_25.txt podilato_linear
Scanning examples...done
Reading examples into memory...100..OK. (150 examples read)
Setting default regularization parameter C=0.0157
Optimizing.....done. (31 iterations)
Optimization finished (10 misclassified, maxdiff=0.00089).
Runtime in cpu-seconds: 0.01
Number of SV: 111 (including 83 at upper bound)
L1 loss: loss=38.11237
Norm of weight vector: |w|=0.95619
Norm of longest example vector: |x|=7.97982
Estimated UCdim of classifier: UCdim<=59.22049
Computing XiAlpha-estimates...done
Runtime for XiAlpha-estimates in cpu-seconds: 0.00
XiAlpha-estimate of the error: error<=55.33% (rho=1.00,depth=0)
XiAlpha-estimate of the recall: recall=>57.78% (rho=1.00,depth=0)
XiAlpha-estimate of the precision: precision=>53.61% (rho=1.00,depth=0)
```

```
Retrain on full problem.....done.
Leave-one-out estimate of the error: error=11.33%
Leave-one-out estimate of the recall: recall=100.00%
Leave-one-out estimate of the precision: precision=84.11%
Actual leave-one-outs computed: 73 (rho=1.00)
Runtime for leave-one-out in cpu-seconds: 0.36
Writing model file...done
```

2α. Με γραμμική συνάρτηση πυρήνα και 150 δεδομένα εκπαίδευσης

```
C:\Users\Katerina\Desktop\SUMlight2>svm_learn -x 1 -t 1 train_podilato_25.txt podilato_poly
Scanning examples...done
Reading examples into memory...100..OK. (150 examples read)
Setting default regularization parameter C=0.0000
Optimizing.....done. (47 iterations)
Optimization finished (0 misclassified, maxdiff=0.00007).
Runtime in cpu-seconds: 0.03
Number of SV: 113 (including 40 at upper bound)
L1 loss: loss=11.21606
Norm of weight vector: |w|=0.01525
Norm of longest example vector: |x|=520.15107
Estimated UCdim of classifier: UCdim<=63.90793
Computing XiAlpha-estimates...done
Runtime for XiAlpha-estimates in cpu-seconds: 0.00
XiAlpha-estimate of the error: error<=26.67% (rho=1.00,depth=0)
XiAlpha-estimate of the recall: recall=>82.22% (rho=1.00,depth=0)
XiAlpha-estimate of the precision: precision=>75.51% (rho=1.00,depth=0)
```

```
Retrain on full problem.....done.
Leave-one-out estimate of the error: error=4.67%
Leave-one-out estimate of the recall: recall=98.89%
Leave-one-out estimate of the precision: precision=93.68%
Actual leave-one-outs computed: 40 (rho=1.00)
Runtime for leave-one-out in cpu-seconds: 0.17
Writing model file...done
```

2β. Με πολωνυμική συνάρτηση πυρήνα και 150 δεδομένα εκπαίδευσης

```

Reading examples into memory...OK. (71 examples read)
Setting default regularization parameter C=0.0157
Optimizing.....done. (13 iterations)
Optimization finished (8 misclassified, maxdiff=0.00070).
Runtime in cpu-seconds: 0.02
Number of SU: 58 (including 43 at upper bound)
L1 loss: loss=22.06500
Norm of weight vector: |w|=0.65360
Norm of longest example vector: |x|=7.97950
Estimated VCdim of classifier: VCdim<=28.20037
Computing XiAlpha-estimates...done
Runtime for XiAlpha-estimates in cpu-seconds: 0.00
XiAlpha-estimate of the error: error<=60.56% (rho=1.00,depth=0)
XiAlpha-estimate of the recall: recall=>57.78% (rho=1.00,depth=0)
XiAlpha-estimate of the precision: precision=>52.00% (rho=1.00,depth=0)

```

```

Retrain on full problem.....done.
Leave-one-out estimate of the error: error=19.72%
Leave-one-out estimate of the recall: recall=100.00%
Leave-one-out estimate of the precision: precision=76.27%
Actual leave-one-outs computed: 35 (rho=1.00)
Runtime for leave-one-out in cpu-seconds: 0.08
Writing model file...done

```

3α. Με γραμμική συνάρτηση πυρήνα και 71 δεδομένα εκπαίδευσης

```

C:\Users\Katerina\Desktop\SUMlight1>svm_learn -x 1 -t 1 train_podilato_25.txt po
dilato_polynomial
Scanning examples...done
Reading examples into memory...OK. (71 examples read)
Setting default regularization parameter C=0.0000
Optimizing.....done. (20 iterations)
Optimization finished (1 misclassified, maxdiff=0.00098).
Runtime in cpu-seconds: 0.01
Number of SU: 56 (including 23 at upper bound)
L1 loss: loss=6.10306
Norm of weight vector: |w|=0.01140
Norm of longest example vector: |x|=520.08950
Estimated VCdim of classifier: VCdim<=36.15938
Computing XiAlpha-estimates...done
Runtime for XiAlpha-estimates in cpu-seconds: 0.00
XiAlpha-estimate of the error: error<=32.39% (rho=1.00,depth=0)
XiAlpha-estimate of the recall: recall=>80.00% (rho=1.00,depth=0)
XiAlpha-estimate of the precision: precision=>72.00% (rho=1.00,depth=0)

```

```

Retrain on full problem.....done.
Leave-one-out estimate of the error: error=7.04%
Leave-one-out estimate of the recall: recall=100.00%
Leave-one-out estimate of the precision: precision=90.00%
Actual leave-one-outs computed: 22 (rho=1.00)
Runtime for leave-one-out in cpu-seconds: 0.06
Writing model file...done

```

3β. Με πολυωνυμική συνάρτηση πυρήνα και 71 δεδομένα εκπαίδευσης

Αποτελέσματα για το σήμα του ΣΤΟΠ:

```

Scanning examples...done
Reading examples into memory...100..200..OK. (207 examples read)
Setting default regularization parameter C=0.0157
Optimizing.....done. (23 iterations)
Optimization finished (1 misclassified, maxdiff=0.00064).
Runtime in cpu-seconds: 0.01
Number of SV: 57 (including 31 at upper bound)
L1 loss: loss=5.45039
Norm of weight vector: |w|=0.78196
Norm of longest example vector: |x|=7.97950
Estimated UCdim of classifier: UCdim<=39.92876
Computing XiAlpha-estimates...done
Runtime for XiAlpha-estimates in cpu-seconds: 0.00
XiAlpha-estimate of the error: error<=14.98% (rho=1.00,depth=0)
XiAlpha-estimate of the recall: recall=>85.48% (rho=1.00,depth=0)
XiAlpha-estimate of the precision: precision=>89.08% (rho=1.00,depth=0)

```

```

Retrain on full problem.....done.
Leave-one-out estimate of the error: error=0.48%
Leave-one-out estimate of the recall: recall=99.19%
Leave-one-out estimate of the precision: precision=100.00%
Actual leave-one-outs computed: 30 (rho=1.00)
Runtime for leave-one-out in cpu-seconds: 0.10
Writing model file...done

```

1α. Με γραμμική συνάρτηση πυρήνα και 207 δεδομένα εκπαίδευσης

```

C:\Users\Katerina\Desktop\SVM_light>svm_learn -x 1 -t 1 train_stop_25.txt stop_p
oly
Scanning examples...done
Reading examples into memory...100..200..OK. (207 examples read)
Setting default regularization parameter C=0.0000
Optimizing.....done. (36 iterations)
Optimization finished (1 misclassified, maxdiff=0.00097).
Runtime in cpu-seconds: 0.02
Number of SV: 61 (including 3 at upper bound)
L1 loss: loss=1.32237
Norm of weight vector: |w|=0.00959
Norm of longest example vector: |x|=520.08950
Estimated UCdim of classifier: UCdim<=25.87724
Computing XiAlpha-estimates...done
Runtime for XiAlpha-estimates in cpu-seconds: 0.00
XiAlpha-estimate of the error: error<=1.45% (rho=1.00,depth=0)
XiAlpha-estimate of the recall: recall=>98.39% (rho=1.00,depth=0)

```

```

Retrain on full problem.....done.
Leave-one-out estimate of the error: error=0.48%
Leave-one-out estimate of the recall: recall=99.19%
Leave-one-out estimate of the precision: precision=100.00%
Actual leave-one-outs computed: 2 (rho=1.00)
Runtime for leave-one-out in cpu-seconds: 0.02
Writing model file...done

```

1β. Με πολωνυμική συνάρτηση πυρήνα και 207 δεδομένα εκπαίδευσης

```

C:\Users\Katerina\Desktop\SUMlight2>svm_learn -x 1 -t 0 train_stop_25.txt stop_1
inear
Scanning examples...done
Reading examples into memory...100..OK. (149 examples read)
Setting default regularization parameter C=0.0157
Optimizing.....done. (19 iterations)
Optimization finished (1 misclassified, maxdiff=0.00089).
Runtime in cpu-seconds: 0.00
Number of SV: 52 (including 31 at upper bound)
L1 loss: loss=5.66933
Norm of weight vector: |w|=0.72324
Norm of longest example vector: |x|=7.97950
Estimated UCdim of classifier: UCdim<=34.29239
Computing XiAlpha-estimates...done
Runtime for XiAlpha-estimates in cpu-seconds: 0.01
XiAlpha-estimate of the error: error<=20.81% (rho=1.00,depth=0)
XiAlpha-estimate of the recall: recall=>81.52% (rho=1.00,depth=0)
XiAlpha-estimate of the precision: precision=>84.27% (rho=1.00,depth=0)

```

```

Retrain on full problem.....done.
Leave-one-out estimate of the error: error=0.67%
Leave-one-out estimate of the recall: recall=98.91%
Leave-one-out estimate of the precision: precision=100.00%
Actual leave-one-outs computed: 30 (rho=1.00)
Runtime for leave-one-out in cpu-seconds: 0.16
Writing model file...done

```

2α. Με γραμμική συνάρτηση πυρήνα και 149 δεδομένα εκπαίδευσης

```

Reading examples into memory...100..OK. (149 examples read)
Setting default regularization parameter C=0.00000
Optimizing.....done. (29 iterations)
Optimization finished (1 misclassified, maxdiff=0.00092).
Runtime in cpu-seconds: 0.01
Number of SV: 52 (including 4 at upper bound)
L1 loss: loss=1.19235
Norm of weight vector: |w|=0.00919
Norm of longest example vector: |x|=520.08950
Estimated UCdim of classifier: UCdim<=23.81844
Computing XiAlpha-estimates...done
Runtime for XiAlpha-estimates in cpu-seconds: 0.00
XiAlpha-estimate of the error: error<=2.68% (rho=1.00,depth=0)
XiAlpha-estimate of the recall: recall=>96.74% (rho=1.00,depth=0)
XiAlpha-estimate of the precision: precision=>98.89% (rho=1.00,depth=0)

```

```

Retrain on full problem.....done.
Leave-one-out estimate of the error: error=0.67%
Leave-one-out estimate of the recall: recall=98.91%
Leave-one-out estimate of the precision: precision=100.00%
Actual leave-one-outs computed: 3 (rho=1.00)

```

2β. Με πολωνυμική συνάρτηση πυρήνα και 149 δεδομένα εκπαίδευσης

```

Optimizing.....done. (15 iterations)
Optimization finished (1 misclassified, maxdiff=0.00075).
Runtime in cpu-seconds: 0.01
Number of $U$: 40 (including 19 at upper bound)
L1 loss: loss=4.46513
Norm of weight vector: |w|=0.61190
Norm of longest example vector: |x|=7.97950
Estimated UCdim of classifier: UCdim<=24.84054
Computing XiAlpha-estimates...done
Runtime for XiAlpha-estimates in cpu-seconds: 0.00
XiAlpha-estimate of the error: error<=25.00% (rho=1.00,depth=0)
XiAlpha-estimate of the recall: recall=>84.31% (rho=1.00,depth=0)
XiAlpha-estimate of the precision: precision=>79.63% (rho=1.00,depth=0)

```

```

Leave-one-out estimate of the error: error=1.32%
Leave-one-out estimate of the recall: recall=98.04%
Leave-one-out estimate of the precision: precision=100.00%
Actual leave-one-outs computed: 18 (rho=1.00)
Runtime for leave-one-out in cpu-seconds: 0.06

```

3α. Με γραμμική συνάρτηση πυρήνα και 76 δεδομένα εκπαίδευσης

```

Reading examples into memory...OK. (76 examples read)
Setting default regularization parameter C=0.0000
Optimizing.....done. (22 iterations)
Optimization finished (0 misclassified, maxdiff=0.00073).
Runtime in cpu-seconds: 0.02
Number of $U$: 38 (including 4 at upper bound)
L1 loss: loss=1.03309
Norm of weight vector: |w|=0.00791
Norm of longest example vector: |x|=520.08950
Estimated UCdim of classifier: UCdim<=17.90058
Computing XiAlpha-estimates...done
Runtime for XiAlpha-estimates in cpu-seconds: 0.00
XiAlpha-estimate of the error: error<=5.26% (rho=1.00,depth=0)
XiAlpha-estimate of the recall: recall=>96.08% (rho=1.00,depth=0)
XiAlpha-estimate of the precision: precision=>96.08% (rho=1.00,depth=0)

```

```

Retrain on full problem.....done.
Leave-one-out estimate of the error: error=1.32%
Leave-one-out estimate of the recall: recall=98.04%
Leave-one-out estimate of the precision: precision=100.00%
Actual leave-one-outs computed: 4 (rho=1.00)
Runtime for leave-one-out in cpu-seconds: 0.04

```

3β. Με πολυωνυμική συνάρτηση πυρήνα και 76 δεδομένα εκπαίδευσης