



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

ΑΝΙΧΝΕΥΣΗ ΚΑΙ ΑΝΑΓΝΩΡΙΣΗ ΠΙΝΑΚΙΔΩΝ ΚΥΚΛΟΦΟΡΙΑΣ
ΜΕ ΧΡΗΣΗ ΤΟΠΙΚΩΝ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Γεώργιος Π. Περάκης

Επιβλέπων : Στέφανος Κόλλιας
Καθηγητής ΣΗΜΜΥ ΕΜΠ

Αθήνα, Οκτώβριος 2014



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

Ανίχνευση και Αναγνώριση Πινακίδων Κυκλοφορίας με χρήση Τοπικών Χαρακτηριστικών

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Γεωργίου Π. Περάκη

Επιβλέπων : Στέφανος Κόλλιας

Καθηγητής ΣΗΜΜΥ ΕΜΠ

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 17^η Οκτωβρίου 2014.

.....

Στέφανος Κόλλιας,
Καθηγητής, Μέλος Δ.Ε.Π

.....

Ανδρέας-Γεώργιος
Σταφυλοπάτης,
Καθηγητής, Μέλος Δ.Ε.Π

.....

Γεώργιος Στάμου,
Επίκουρος Καθηγητής,
Μέλος Δ.Ε.Π

Αθήνα, Οκτώβριος 2014

.....
Γεώργιος Π. Περάκης

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

Copyright © Γεώργιος Περάκης, 2014.

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

Ευχαριστίες

Πριν την έναρξη του κυρίως σώματος της εργασίας, θα ήθελα να ευχαριστήσω τον καθηγητή κ. Στέφανο Κόλλια που μου εμπιστεύθηκε την εκπόνηση διπλωματικής εργασίας στο εργαστήριο Ψηφιακής Επεξεργασίας Σημάτων και Εικόνων. Θα ήθελα επίσης να ευχαριστήσω ιδιαίτερα τον κ. Γιάννη Αβρίθη για την καθοδήγησή του καθ' όλη την περίοδο της προετοιμασίας και της συγγραφής αυτής της εργασίας, και τον κ. Κώστα Ραπαντζίκο για τις πολύ χρήσιμες υποδείξεις του σε κρίσιμα σημεία της, ώστε αυτή να λάβει την τελική της μορφή. Όμως χρωστάω και ένα μεγάλο ευχαριστώ στην οικογένειά μου και τους φίλους μου που με στήριξαν ηθικά σε αυτήν την προσπάθεια, και ιδιαίτερα στην αδερφή μου, Ειρητίνα, για τη βοήθειά της στην καταγραφή των πινακίδων από τα σύνολα δεδομένων και στη διόρθωση των ορθογραφικών και εκφραστικών λαθών, και στην Αφροδίτη για την πολύτιμη βοήθειά της σε κάθε στάδιο της εργασίας, από την καταγραφή των πινακίδων μέχρι τη συγγραφή και την προετοιμασία της παρουσίασης.

Περίληψη

Η αναγνώριση πινακίδων κυκλοφορίας είναι αναπόσπαστο κομμάτι ενός αυτόματου συστήματος ελέγχου της κυκλοφορίας με ολοένα και περισσότερες εφαρμογές. Στόχος της παρούσης εργασίας είναι η ανάπτυξη ενός πρωτότυπου συστήματος ανίχνευσης και αναγνώρισης πινακίδων κυκλοφορίας, με χρήση τοπικών χαρακτηριστικών της εικόνας. Γίνεται περιγραφή των σημαντικότερων ως τώρα μεθόδων και για τα δύο σκέλη του προβλήματος και σχολιασμός των αποτελεσμάτων τους. Παρουσιάζονται τα προβλήματα που προέκυψαν κατά την υλοποίηση της μεθόδου και οι λύσεις που επιλέχθηκαν για την αντιμετώπισή τους. Επίσης, επισημαίνεται η ανάγκη για δημιουργία διεθνούς βάσης δεδομένων για την αναγνώριση πινακίδων κυκλοφορίας, αφού ελλείψει αυτής δεν μπορεί να γίνει αξιόπιστη σύγκριση των προτεινόμενων λύσεων και επιλογή της βέλτιστης. Στη διαθέσιμη βάση δεδομένων ελληνικών πινακίδων, το ποσοστό ανίχνευσης ανήλθε στο 94%, και το ποσοστό αναγνώρισης των χαρακτήρων της πινακίδας με χρήση ελεύθερης πλατφόρμας στο 32,3%.

Λέξεις-κλειδιά: αναγνώριση πινακίδων κυκλοφορίας, ΑΠΚ, ανίχνευση, τοπικά χαρακτηριστικά, tesseract-OCR

Abstract

License plate recognition is an integral part of intelligent traffic control systems, with ever more applications. The goal of this diploma thesis is the creation of a prototype system for license plate detection and recognition with the use of local descriptors of the image. The most important methods so far for both parts of the problem are described. Problems that arose during the implementation of the method are presented, as well as the chosen solutions against them. Furthermore, the need to create a global data set for license plate recognition is noted, since without it there cannot be a reliable comparison among the suggested solutions and selection of the best. For the data set of greek license plates that was available, the detection rate was 94% and the recognition rate 32.3%, with the use of a free platform.

Keywords: license plate recognition, LPR, detection, local descriptors, tesseract-OCR

Περιεχόμενα

Κεφάλαιο 1: Εισαγωγή.....	9
1.1 Εφαρμογές της Αναγνώρισης Πινακίδων Κυκλοφορίας (ΑΠΚ).....	9
1.2 Η ανάπτυξη συστημάτων ΑΠΚ.....	11
1.3 Στόχος της εργασίας.....	11
1.4 Δομή της εργασίας.....	12
Κεφάλαιο 2: Θεωρητικό Υπόβαθρο.....	13
2.1 Προσεγγίσεις του προβλήματος ανίχνευσης.....	13
2.1.1 Μέθοδοι ανάλυσης ακμών, μορφολογικές μέθοδοι.....	13
2.1.2 Μέθοδοι ανάλυσης χρώματος και υφής.....	15
2.1.3 Μέθοδοι υπολογιστικής μάθησης.....	17
2.1.4 Άλλες μέθοδοι.....	19
2.2 Ο διαχωρισμός των χαρακτήρων.....	21
2.3 Η αναγνώριση των χαρακτήρων.....	23
2.3.1 Ταύτιση προτύπων.....	24
2.3.2 Νευρωνικά δίκτυα.....	26
2.3.3 Άλλες μέθοδοι.....	27
Κεφάλαιο 3: Μέθοδος ανίχνευσης της πινακίδας.....	29
3.1 Οι μέγιστα σταθερές ακρότατες περιοχές (MSER).....	29
3.2 Περιγραφή του αλγορίθμου εύρεσης της πινακίδας.....	33
3.2.1 Προεπεξεργασία εικόνας.....	33
3.2.2 Εύρεση περιοχών MSER.....	34
3.2.3 Μείωση των υποψήφιων περιοχών με βάση τις ιδιότητες των MSER+.....	37
3.2.4 Μείωση τα των περιοχών με βάση τις ιδιότητες των περιοχών MSER-.....	38
3.2.5 Επιλογή της τελικής υποψήφιας περιοχής.....	40
Κεφάλαιο 4: Η αναγνώριση των χαρακτήρων.....	43
4.1 Tesseract OCR.....	43
4.2 Προεπεξεργασία της εικόνας της πινακίδας.....	44
4.2.1 Περιστροφή της εικόνας αποκοπή του πλαισίου της πινακίδας.....	44
4.2.2 Προσαρμογή του μεγέθους της εικόνας.....	45
4.2.3 Φιλτράρισμα της εικόνας.....	46

4.2.4	<i>Κατωφλίωση της εικόνας</i>	47
4.3	<i>Προσαρμογή των παραμέτρων του tesseract-OCR</i>	47
Κεφάλαιο 5: Πειράματα		49
5.1	<i>Η πειραματική διαδικασία</i>	49
5.2	<i>Σύνολα δεδομένων</i>	49
5.3	<i>Αποτελέσματα Ανίχνευσης της Πινακίδας</i>	50
5.3.1	<i>Μεταβολή των παραμέτρων των MSER</i>	50
5.3.2	<i>Μετρήσεις για διαφορετικό ποσοστό επικάλυψης εμβადού</i>	54
5.3.3	<i>Παραδείγματα λανθασμένης ανίχνευσης</i>	55
5.3.4	<i>Εφαρμογή των παραμέτρων στα μικρότερα σύνολα δεδομένων</i>	57
5.4	<i>Αποτελέσματα της αναγνώρισης χαρακτήρων</i>	58
5.4.1	<i>Δοκιμές στο στάδιο της προεπεξεργασίας</i>	58
5.4.2	<i>Δοκιμές στις παραμέτρους του tesseract-OCR</i>	61
5.4.3	<i>Παραδείγματα αποτυχημένης αναγνώρισης</i>	61
5.4.4	<i>Τελική επιλογή παραμέτρων – εφαρμογή στα μικρότερα σύνολα δεδομένων</i>	62
5.5	<i>Σύγκριση με άλλες μεθόδους</i>	63
Κεφάλαιο 6: Συμπεράσματα		67
Παράρτημα		73
	<i>Συναρτήσεις της OpenCV που χρησιμοποιήθηκαν για την υλοποίηση</i>	73

Κεφάλαιο 1

Εισαγωγή

Η αναγνώριση πινακίδων κυκλοφορίας (ΑΠΚ, license plate recognition – LPR ή automatic number plate recognition – ANPR) είναι μία από τις πιο σημαντικές λειτουργίες ενός έξυπνου κυκλοφοριακού συστήματος, δηλαδή ενός συστήματος ελέγχου της κυκλοφορίας των οχημάτων και των σχετικών εφαρμογών. Σκοπός της είναι η ανίχνευση της πινακίδας κυκλοφορίας ενός οχήματος από μια εικόνα εισόδου και η ανάγνωσή της για αυτόματη ταυτοποίηση του οχήματος. Η λήψη της εικόνας μπορεί να γίνει με κάμερες διαχείρισης κυκλοφορίας που είναι ήδη εγκατεστημένες στις άκρες των δρόμων. Οι κάμερες αυτές εποπτεύουν την κυκλοφοριακή ροή, την κίνηση των πεζών, τα οδικά έργα και ατυχήματα. Όμως η ΑΠΚ πρέπει να δίνει σε κάθε περίπτωση αξιόπιστα αποτελέσματα, ανεξάρτητα από τις περιβαλλοντικές συνθήκες (φωτισμός, καιρικές συνθήκες) και την ταχύτητα του οχήματος, η οποία μπορεί να προκαλέσει θόλωση της εικόνας. Για αυτό είναι πιο αποτελεσματικό να χρησιμοποιούνται ειδικές κάμερες τοποθετημένες σε κατάλληλο σημείο, οι οποίες περιλαμβάνουν τεχνολογίες βελτίωσης της λήψης της εικόνας.

1.1 *Εφαρμογές της Αναγνώρισης Πινακίδων Κυκλοφορίας (ΑΠΚ)*

Οι εφαρμογές ενός τέτοιου συστήματος είναι ποικίλες και καθεμία μπορεί να διευκολύνει την κίνηση ενός συνεχώς αυξανόμενου αριθμού αυτοκινήτων στις μεγαλουπόλεις. Ακόμη, μπορεί να υποβοηθήσει το έργο της τροχαίας για μεγαλύτερη ασφάλεια στους δρόμους και των αστυνομικών αρχών για την προστασία της ιδιωτικής ασφάλειας. Τέλος, είναι ικανή να μειώσει το χρόνο κάποιων καθημερινών λειτουργιών, όπως η πληρωμή διοδίων ή η είσοδος σε χώρους στάθμευσης.

Για τη μεγάλη χρησιμότητα της ΑΠΚ σε περιπτώσεις διαλεύκανσης εγκληματικών ενεργειών υπάρχει το πρόσφατο παράδειγμα του φόνου της αστυνομικού Sharon Beshenivsky. Το 2005 στο Bradford της Αγγλίας [1], αμέσως μετά το φόνο της αστυνομικού, ένα σύστημα καμερών κλειστού κυκλώματος μπόρεσε να παρακολουθήσει αυτόματα την πορεία του αυτοκινήτου με το οποίο είχαν διαφύγει οι δράστες, οδηγώντας στη σύλληψή τους. Η αναζήτηση ενός οχήματος σε τόσο πολλές κάμερες από ανθρώπους χωρίς τη χρήση του συστήματος

αυτόματης αναγνώρισης πινακίδων κυκλοφορίας θα ήταν πολύ πιο χρονοβόρα και πιθανόν να μην είχε το ίδιο επιτυχές αποτέλεσμα.

Μια ακόμη εφαρμογή της στην αντιμετώπιση εγκλημάτων, η οποία μάλιστα δε χρησιμοποιεί σταθερή κάμερα αλλά κινητή, εντάσσεται στο σύστημα MoveOn [2] [3], το οποίο υποβοηθά σημαντικά το έργο των δικυκλιστών αστυνομικών και μειώνει το φόρτο εργασίας τους. Μέρος του συστήματος είναι και η αυτόματη αναγνώριση πινακίδων κυκλοφορίας, καθώς μια κάμερα ενσωματωμένη στην κορυφή του κράνους μπορεί να εντοπίσει την πινακίδα ενός προπορευόμενου οχήματος και να αποστείλει τον αριθμό της στο επιχειρησιακό κέντρο. Αυτό με τη σειρά του μπορεί να ενημερώσει εντός δευτερολέπτων τον αστυνομικό για το αν το όχημα είναι κλεμμένο ή ύποπτο.

Όσον αφορά την υποβοήθηση του έργου της τροχαίας, και εδώ οι εφαρμογές είναι πολλές. Σε αρκετές ευρωπαϊκές χώρες χρησιμοποιούνται συστήματα που μετρούν τη μέση ταχύτητα ενός οχήματος σε μια μεγάλη απόσταση για να υπολογιστεί αν έχει ξεπεράσει τα όρια ταχύτητας. Με αυτόν τον τρόπο αποφεύγεται το φαινόμενο να φρενάρουν απότομα τα οχήματα όταν πλησιάζουν ένα γνωστό σημείο ελέγχου της στιγμιαίας ταχύτητας και να επιταχύνουν μετά. Όμως η μέτρηση της μέσης ταχύτητας μπορεί να χρησιμοποιηθεί και για τη διαχείριση της κυκλοφορίας. Πιο συγκεκριμένα, μπορεί να υπολογιστεί ο μέσος χρόνος που απαιτείται για την κάλυψη μιας απόστασης μέσα στην πόλη σε ώρες αιχμής και αυτός να αναγράφεται σε πινακίδες μεταβλητού μηνύματος. Επομένως, δίνεται η δυνατότητα στους οδηγούς να επιλέξουν μια συντομότερη διαδρομή.

Η ταυτοποίηση των οχημάτων μπορεί να χρησιμοποιηθεί για τον έλεγχο των οχημάτων που εισέρχονται σε έναν ιδιωτικό χώρο ή έναν χώρο αυξημένης ασφαλείας, καθώς και όσων περνούν από συνοριακούς σταθμούς. Επίσης μπορεί να χρησιμοποιηθεί για τη συλλογή τελών διοδίων χωρίς να χρειάζεται να σταματά το όχημα στο σταθμό διοδίων, μειώνοντας έτσι το χρόνο του ταξιδιού. Ήδη σε αρκετούς αυτοκινητοδρόμους στην Ευρώπη, τις ΗΠΑ και την Αυστραλία λειτουργούν τέτοια συστήματα που στηρίζονται μερικώς ή εξ ολοκλήρου στην αυτόματη αναγνώριση πινακίδων κυκλοφορίας.

Άλλες εφαρμογές περιλαμβάνουν τη μέτρηση του χρόνου παραμονής σε χώρους στάθμευσης, για στατιστικούς λόγους ή για τον επακόλουθο υπολογισμό του αντιτίμου. Επίσης διευκολύνουν πολύ τις μελέτες συγκοινωνιολόγων, δίνοντας τη δυνατότητα στατιστικής ανάλυσης κυκλοφοριακών μετρήσεων για διάφορες παραμέτρους. Αυτό πραγματοποιείται λόγω του μεγάλου όγκου δεδομένων που μπορούν να παρέχουν γρήγορα και αξιόπιστα, με στόχο το σχεδιασμό λύσεων για τις περιπτώσεις που χρήζουν βελτίωσης. Τέλος, μια άλλη εφαρμογή αφορά την εξαγωγή μοτίβων συμπεριφοράς των καταναλωτών που εισέρχονται με τα οχήματά τους σε χώρους στάθμευσης για εμπορικούς σκοπούς.

1.2 Η ανάπτυξη συστημάτων ΑΠΚ

Ένα γενικό σύστημα αναγνώρισης πινακίδων κυκλοφορίας αποτελείται από τρία κύρια μέρη: Την ανίχνευση της πινακίδας μέσα στην εικόνα, το διαχωρισμό των χαρακτήρων της πινακίδας και την αναγνώρισή τους. Ο διαχωρισμός των χαρακτήρων μέσα σε μια προκαθορισμένη περιοχή και η αναγνώρισή τους, ακόμα κι αν χρειάζεται κάποια διαδικασία εκπαίδευσης, είναι πεδία που έχουν πολλές εφαρμογές και η έρευνα έχει προχωρήσει με μεγάλο βαθμό επιτυχίας. Όμως ο εντοπισμός της πινακίδας πρέπει να μπορεί να γίνει σε μη ελεγχόμενα περιβάλλοντα. Για παράδειγμα, οι διαφορετικές συνθήκες καιρού και φωτισμού μπορούν να προκαλέσουν χαμηλή αντίθεση εικόνας ή σκιάσεις εντός των πινακίδων, η μεγάλη ταχύτητα του οχήματος οδηγεί σε θολές εικόνες, ενώ και οι πινακίδες μπορεί να μην είναι καθαρές λόγω σκόνης, χρώματος ή να έχουν γίνει επιτηδευμένες προσπάθειες για κάλυψή της. Έτσι η ανίχνευση της πινακίδας μέσα στην εικόνα είναι το πιο δύσκολο και σημαντικό μέρος αυτής της διαδικασίας.

Η ΑΠΚ αποτελεί αντικείμενο έρευνας ήδη από τη δεκαετία του '80 και στα πολλά έτη έρευνας πάνω σε αυτόν τον τομέα έχουν προταθεί διάφορες προσεγγίσεις για την επίτευξή της, οι οποίες θα αναλυθούν στο επόμενο κεφάλαιο. Σε κάθε περίπτωση χρησιμοποιούνται διάφορα χαρακτηριστικά της πινακίδας κυκλοφορίας για την ανίχνευσή της, όπως η ένταση των ακμών της, ο τύπος των χαρακτήρων που περιέχει, το ορθογώνιο σχήμα της, η σχέση πλάτους προς ύψος, το χρώμα, υφή, η τοπική μεταβλητότητα της φωτεινότητας κ.ά.

1.3 Στόχος της εργασίας

Στόχος της παρούσης εργασίας είναι η εξ ολοκλήρου δημιουργία ενός πρωτότυπου συστήματος ανίχνευσης και αναγνώρισης πινακίδων κυκλοφορίας, ώστε να δίνει ικανοποιητικά αποτελέσματα σε χρόνους εκτέλεσης κατάλληλους για εφαρμογές πραγματικού χρόνου. Δόθηκε μεγαλύτερη έμφαση στην ανίχνευση της θέσης της πινακίδας μέσα στην εικόνα, επειδή αυτό είναι το βασικότερο βήμα στη διαδικασία αναγνώρισης. Η ανάγνωση των χαρακτήρων προϋποθέτει την γνώση της ακριβούς τοποθεσίας τους, ανεξάρτητα από τις διαφορετικές συνθήκες που καλείται να αντιμετωπίσει κάθε φορά το σύστημα. Προτιμήθηκε η χρήση μεθόδων που χρησιμοποιούν ιδιότητες περιοχών της εικόνας, και συγκεκριμένα στις μέγιστα σταθερές ακρότατες περιοχές (maximally stable extremal regions – MSER) [4], αφού αποτελούν μια σχετικά πρόσφατη μέθοδο, και θεωρούνται από τους καλύτερους περιγραφείς περιοχών [5].

Η βασική ιδέα της εργασίας βασίστηκε στην προσέγγιση της εργασίας των Donoser, Arth, Bischof [6], στην οποία όμως έγιναν πολλές τροποποιήσεις, τόσο λόγω των περιορισμών που επέβαλε το σύνολο δεδομένων που χρησιμοποιήθηκε, όσο και λόγω δοκιμής άλλων μεθόδων

που έδιναν καλύτερα αποτελέσματα. Επίσης η εργασία εκείνη εκμεταλλεύεται και ιδιότητες συνεχόμενων καρτέ από βίντεο, κάτι που όμως δεν ήταν διαθέσιμο για τη παρούσα εργασία. Γίνεται σύγκριση με άλλες προσεγγίσεις. Η υλοποίηση έγινε σε γλώσσα προγραμματισμού C++ με χρήση της βιβλιοθήκης OpenCV, με κάποιες τροποποιήσεις υπάρχουσων συναρτήσεων και μεθόδων, ώστε να είναι δυνατή η υλοποίηση του αλγορίθμου.

Για το κομμάτι της οπτικής αναγνώρισης των χαρακτήρων επιχειρήθηκε η χρήση έτοιμης και ανοιχτής βιβλιοθήκης, της tesseract-OCR, η οποία θεωρείται από τις καλύτερες διαθέσιμες βιβλιοθήκες για αναγνώριση κειμένου. Η εξειδίκευσή της όμως σε αναγνώριση κειμένου αντί για μεμονωμένους χαρακτήρες δημιούργησε κάποιες δυσκολίες, οι οποίες αναλύονται στην περιγραφή της υλοποίησης. Σε κάθε περίπτωση επιχειρήθηκε η εύρεση της βέλτιστης δυνατής λύσης ώστε να επιτευχθούν τα καλύτερα αποτελέσματα. Επιπλέον, η υλοποίηση ενός εξειδικευμένου συστήματος οπτικής αναγνώρισης χαρακτήρων θα ξέφευγε από τα χρονικά όρια για την ολοκλήρωση της διπλωματικής εργασίας.

Για την αποτίμηση των αποτελεσμάτων έγιναν μετρήσεις πάνω σε δύο σύνολα δεδομένων, ένα ελληνικό [7] και ένα κροατικό [8], δοκιμάζοντας διαφορετικές παραμέτρους και μεθόδους για τη βελτίωση των αποτελεσμάτων. Της αποτίμησης όμως προηγήθηκε η καταγραφή των δεδομένων για τη θέση και το περιεχόμενο των πινακίδων (annotation) και στα δύο σύνολα, αφού οι ερευνητές που τις διαθέτουν δεν ήταν σε θέση να διαθέσουν, πέρα από τις εικόνες, και τα δεδομένα που είναι απαραίτητα για την αποτίμηση του αλγορίθμου. Έτσι αναδείχθηκε και ένα πρόβλημα της έρευνας για την αυτόματη αναγνώριση πινακίδων κυκλοφορίας, αφού παρά τις πολλές εφαρμογές της ΑΠΚ και του μεγάλου αριθμού δημοσιεύσεων πάνω σε αυτό το θέμα, δεν υπάρχει κάποια διεθνής τεκμηριωμένη βάση δεδομένων.

1.4 Δομή της εργασίας

Η δομή της εργασίας έχει ως εξής. Αρχικά θα γίνει μια σύντομη ανάλυση των διάφορων μεθοδολογιών αναγνώρισης πινακίδων που έχουν αναπτυχθεί και τα πλεονεκτήματα και μειονεκτήματα της καθεμιάς, ξεχωριστά για τη διαδικασία εντοπισμού της πινακίδας και της ανάγνωσης του περιεχομένου της. Στη συνέχεια θα αναλυθεί η μεθοδολογία που αναπτύχθηκε, στα πλαίσια της οποίας θα γίνει παρουσίαση των MSER και της βιβλιοθήκης tesseract που χρησιμοποιήθηκε για την οπτική αναγνώριση των χαρακτήρων. Ακολουθεί η περιγραφή της προγραμματιστικής διαδικασίας που χρησιμοποιήθηκε και οι παράμετροι που επιλέχθηκαν ώστε να επιτευχθούν καλύτερα αποτελέσματα. Τέλος θα γίνει παρουσίαση των αποτελεσμάτων και σύγκρισή τους με άλλες μεθόδους, καθώς και προτάσεις για τη βελτίωσή τους με μελλοντική εργασία.

Κεφάλαιο 2

Θεωρητικό Υπόβαθρο

Το πρόβλημα της αναγνώρισης πινακίδων κυκλοφορίας περιλαμβάνει τρία κύρια μέρη: Τον εντοπισμό της πινακίδας εντός της εικόνας, την κατάτμηση των χαρακτήρων και την αναγνώρισή τους. Συχνά τα δύο τελευταία ενοποιούνται ως αναγνώριση χαρακτήρων. Σε αυτό το κεφάλαιο θα γίνει παρουσίαση των τεχνικών που έχουν χρησιμοποιηθεί σε διάφορες εργασίες για την επίλυση των επιμέρους προβλημάτων.

2.1 Προσεγγίσεις του προβλήματος ανίχνευσης

Όπως αναφέρθηκε και στην εισαγωγή, το πιο δύσκολο και πιο σημαντικό στάδιο της ΑΠΚ είναι και το πρώτο χρονικά, δηλαδή ο εντοπισμός της πινακίδας κυκλοφορίας μέσα στην εικόνα. Πρέπει δηλαδή να εξαχθούν με αρκετή ακρίβεια οι συντεταγμένες των κορυφών της, ώστε να είναι δυνατή η απομόνωση των χαρακτήρων που περιέχει και στη συνέχεια η ανάγνωσή τους. Για αυτό το κομμάτι του προβλήματος έχουν αναπτυχθεί κατά καιρούς διάφορες μεθοδολογίες, με καλά ή υποσχόμενα αποτελέσματα, ώστε να επιτυγχάνεται σε ένα καλό ποσοστό η ανίχνευση της θέσης της πινακίδας (συνήθως γύρω στο 97-98%) με διάρκεια που να επιτρέπει τη χρήση του συστήματος σε εφαρμογές πραγματικού χρόνου.

2.1.1 Μέθοδοι ανάλυσης ακμών, μορφολογικές μέθοδοι

Μία από τις πρώτες μεθόδους που χρησιμοποιήθηκαν για το εξεταζόμενο πρόβλημα κάνει χρήση της ιδιότητας των πινακίδων κυκλοφορίας να παρουσιάζουν πολλές κάθετες ακμές με μικρά κενά ανάμεσά τους εντός της πινακίδας, λόγω της παρουσίας των χαρακτήρων εντός της. Έτσι, με τη βοήθεια ενός ανιχνευτή της έντασης κάθετων ακμών, συνήθως του τελεστή Sobel, γίνεται εντοπισμός των κάθετων ακμών. Στη συνέχεια με διάφορες μεθόδους γίνεται κατηγοριοποίηση των ακμών που βρέθηκαν με σκοπό να ευρεθεί η περιοχή που έχει τις περισσότερες ακμές με επιθυμητά χαρακτηριστικά, και η οποία πιθανότατα θα αντιστοιχεί στην πινακίδα κυκλοφορίας.

Όμως για τη βελτίωση του σημαντικού τμήματος της εικόνας χρησιμοποιείται συνήθως κάποια προεπεξεργασία της. Έτσι, οι Zheng κ.συν. [9] υπολογίζουν τη μέση τιμή

φωτεινότητας και την τυπική της απόκλιση σε block από pixel και ενισχύουν αυτά που έχουν τιμές των δύο αυτών παραμέτρων εντός κάποιων ορίων. Από την άλλη, οι Hongliang και Changping [10] προχωρούν σε απλή κανονικοποίηση της φωτεινότητας σε ολόκληρη την εικόνα, για να βελτιώσουν την αντίθεση σε πολύ φωτεινές ή πολύ σκοτεινές εικόνες. Οι Sharipo κ.συν. [11] προχωρούν ακόμη περισσότερο, αφού λογαριθμούν τη φωτεινότητα, ώστε οι σκούρες περιοχές με μικρές τιμές φωτεινότητας, που συνήθως αντιστοιχούν στους χαρακτήρες, να διαχωρίζονται πιο έντονα από τις περιοχές με μέση ή μεγάλη φωτεινότητα, δηλαδή την υπόλοιπη πινακίδα. Τέλος οι Mahini κ.συν. [12] χρησιμοποιούν το μορφολογικό φίλτρο closing και αφαιρούν από το αποτέλεσμα την αρχική εικόνα (τελεστής Bot Hat), ενισχύοντας έτσι τις σημαντικές περιοχές.

Στη συνέχεια γίνεται η εύρεση των ακμών. Αυτή μπορεί να γίνει με χρήση κάποιου τελεστή, όπως Sobel [9] [12] [13], Roberts [11] ή Mexican Hat [14], ή με μαθηματικό τρόπο, υπολογίζοντας τη μερική παράγωγο ως προς την οριζόντια διεύθυνση [15] [16]. Ύστερα είτε το ίδιο το μέτρο της έντασής τους χρησιμοποιείται σαν δείκτης για την εύρεση της περιοχής της πινακίδας [13] [14] είτε γίνεται κατώφλιωση των εντάσεων των ακμών και προκύπτει δυαδική εικόνα. Το κατώφλι μπορεί να επιλεγεί τυχαία, αλλά έχουν χρησιμοποιηθεί και μορφολογικές μέθοδοι, όπως ταξικό φίλτρο (rank filter) [11]. Το ταξικό φίλτρο εφαρμόζεται στην εικόνα των ακμών στην κλίμακα του γκρι και αντικαθιστά την τιμή φωτεινότητας του κάθε pixel με το 80% της μεγαλύτερης τιμής που εντοπίζεται σε ένα μικρό παράθυρο γύρω του. Ακολουθεί απομάκρυνση των ακμών με πολύ μικρό ή πολύ μεγάλο μήκος, με βάση το αναμενόμενο μέγεθος της πινακίδας ή εφαρμόζεται τελεστής closing [12].

Για την εύρεση της πινακίδας χρησιμοποιούνται διάφορες μέθοδοι. Η πρώτη προσέγγιση είναι η μέτρηση των κάθετων ακμών ανά γραμμή και η προβολή τους στον κάθετο άξονα, η οποία δημιουργεί ένα ιστόγραμμα προβολών. Η περιοχή αυτού του ιστογράμματος με τις περισσότερες γραμμές με πλήθος ακμών πάνω από ένα κατώφλι θεωρείται υποψήφια περιοχή πινακίδας [11] [13] [15] [16]. Στη συνέχεια αναζητούνται στον οριζόντιο άξονα οι περιοχές με κοντινές κάθετες ακμές, σχηματίζοντας έτσι το περίβλημα της πινακίδας ορθογώνιου σχήματος.

Μια δεύτερη προσέγγιση είναι η μέτρηση των ακμών εντός ενός κυλιόμενου παραθύρου, συνήθως μεγέθους όσο το αναμενόμενο μέγεθος της πινακίδας. Δηλαδή τα παράθυρα που περιέχουν μεγάλο αριθμό από pixel που αντιστοιχούν σε κάθετες ακμές θεωρούνται υποψήφιας περιοχής πινακίδας [9]. Εναλλακτικά μπορεί να εφαρμοστεί συνέλιξη με περιοχή σχήματος ορθογώνιου ώστε να ευρεθούν οι περιοχές με μεγάλη πυκνότητα ακμών [16]. Στην εργασία [10] γίνεται στατιστική ανάλυση των σημείων που ανήκουν σε κάθετες ακμές και τα ομαδοποιούν σε γραμμές ανάλογα με την οριζόντια απόστασή τους και σε ορθογώνια με βάση την κάθετη απόσταση μεταξύ δύο ομάδων αποδεκτών σημείων. Αποδεκτά γίνονται τα ορθογώνια που έχουν μήκος μεγαλύτερο από ένα κατώφλι και που δεν επικαλύπτονται με άλλα.

Αφού έχουν βρεθεί οι υποψήφιες περιοχές πινακίδας, το επόμενο βήμα είναι να επιλεγεί μία περιοχή ανά εικόνα. Γι' αυτό το σκοπό επιστρατεύονται διάφορες ευριστικές. Η πρώτη αφορά τις διαστάσεις της πινακίδας, καθώς αυτές είναι σταθερές για κάθε τύπο πινακίδας και στην εικόνα εξαρτώνται από την απόσταση μεταξύ κάμερας και οχήματος κατά τη λήψη και τη σχετική τους γωνία. Όμως επειδή αυτή δεν είναι πάντα γνωστή ούτε σταθερή, πιο αξιόπιστη θεωρείται η σχέση μεταξύ πλάτους και ύψους του παραλληλογράμμου που την περικλείει, καθώς και η γωνία αυτού με τον οριζόντιο άξονα.

Η πρώτη αυτή μέθοδος δίνει αρκετά καλά αποτελέσματα σε μικρό υπολογιστικό χρόνο, αλλά τα σύνολα δεδομένων που χρησιμοποιούνται για τη δοκιμή των αλγορίθμων περιέχουν συνήθως εικόνες χωρίς σύνθετο φόντο ή διαφορετικά περιβάλλοντα. Για την εξάλειψη της επίδρασης τέτοιων συνθηκών χρησιμοποιείται σχεδόν πάντα κάποια προεπεξεργασία της εικόνας, ενώ μια πιο αξιόπιστη μέθοδος εξάλειψης του περιβαλλοντικού θορύβου είναι η χρησιμοποίηση διαδοχικών καρτέ από βίντεο. Όποτε αυτό είναι διαθέσιμο, μπορεί με διάφορες μεθόδους, όπως ανάλυση block [13], να εξαχθεί η εικόνα του οχήματος και να γίνει η αναζήτηση ακμών μόνο εντός της εικόνας του οχήματος. Συχνά όμως δεν υπάρχει αυτή η δυνατότητα λόγω του μεγάλου κόστους αποθήκευσης διαδοχικών καρτέ ενός οχήματος σε πραγματικά συστήματα.

2.1.2 Μέθοδοι ανάλυσης χρώματος και υφής

Μια δεύτερη προσέγγιση που έχει χρησιμοποιηθεί εκμεταλλεύεται τις χρωματικές ιδιότητες ή την υφή της περιοχής της πινακίδας, για να την ανιχνεύσει μέσα στην εικόνα. Γενικά δίνει εξίσου καλά αποτελέσματα με την πρώτη μέθοδο, αλλά ο χρόνος εκτέλεσης είναι συνήθως μεγαλύτερος, ενώ σε κάποιες περιπτώσεις χρησιμοποιείται και κάποιο σχήμα εκπαίδευσης για την κατηγοριοποίηση της υφής. Παρακάτω παρουσιάζονται κάποιες χαρακτηριστικές εργασίες.

Σε χώρες της Άπω Ανατολής είναι σύνηθες να χρησιμοποιούνται και άλλα χρώματα για το φόντο ή τους χαρακτήρες της πινακίδας, πέρα από το άσπρο και το μαύρο. Τα διαφορετικά χρώματα μπορεί να αντιστοιχούν σε διαφορετικές γεωγραφικές περιοχές ή σε διαφορετικό τύπο οχήματος. Συνήθως όμως ο συνδυασμός χρωμάτων που επιλέγεται δεν εμφανίζεται συχνά στο περιβάλλον, άρα ένας αλγόριθμος μπορεί να εκμεταλλευτεί αυτό το γεγονός. Για παράδειγμα, οι Shi κ.συν. [17] δημιουργούν κατηγορίες χρωμάτων στο χώρο RGB, όπως σκούρο μπλε, ανοιχτό μπλε, ανοιχτό γκρι, άσπρο, μαύρο κτλ. Μπορεί να χρησιμοποιηθεί και ο χρωματικός χώρος HLS, που κάνει καλύτερη κατηγοριοποίηση των χρωμάτων, αλλά ο χρόνος εκτέλεσης σχεδόν διπλασιάζεται. Στη συνέχεια διατρέχουν την εικόνα ψάχνοντας γειτονικά pixel με συνδυασμούς χρωμάτων όπως μπλε-άσπρο, μαύρο-κίτρινο, διαδοχικά για κάθε ζεύγος ανάλογα με την αναμενόμενη συχνότητα εμφάνισης (πχ στα μικρά Ι.Χ. οχήματα χρησιμοποιείται το μπλε-άσπρο, άρα αυτός είναι ο πρώτος συνδυασμός που αναζητείται). Οι περιοχές που εμφανίζουν μεγάλη τέτοια συχνότητα ενοποιούνται και επιβεβαιώνεται η

υποψηφιότητά τους ως περιοχές πινακίδας με βάση το λόγο πλάτους προς ύψος. Η εκτέλεση του αλγορίθμου διαρκεί λίγο πάνω από 100 ms για εικόνες 512x389.

Όμως, όπως αναφέρθηκε, οι περισσότερες εργασίες που χρησιμοποιούν αυτή την προσέγγιση εκπαιδεύουν κάποιον ταξινομητή της υφής και του χρώματος για τον εντοπισμό των υποψήφιων περιοχών πινακίδων. Έτσι οι K.I.Kim κ.συν. [18] εκπαιδεύουν μια μηχανή υποστηρικτικών διανυσμάτων (support vector machine – SVM) με περιοχές που έχουν υφή πράσινου και λευκού χρώματος, αφού οι κορεάτικες πινακίδες έχουν λευκούς χαρακτήρες σε πράσινο φόντο. Στη συνέχεια ερευνούν ένα μικρό παράθυρο γύρω από κάθε pixel και το ταξινομούν με τη βοήθεια του SVM σε περιοχή με υφή χρώματος όπως της πινακίδας ή σε αδιάφορη περιοχή. Όμως η ταξινόμηση αυτή προϋποθέτει και τη χρήση κυλιόμενων παραθύρων διαφορετικών μεγεθών, αυξάνοντας το υπολογιστικό κόστος με αποτέλεσμα να χρειάζονται 1-9 sec για την εύρεση της πινακίδας.

Στις περισσότερες περιοχές του κόσμου όμως οι πινακίδες αποτελούνται από μαύρους χαρακτήρες σε λευκό ή κίτρινο φόντο. Έτσι η αναζήτηση γίνεται σε εικόνες με κλίμακα του γκρι, κάτι που επίσης μειώνει τον χρόνο εκτέλεσης. Αν οι πινακίδες κοιταχθούν μακροσκοπικά, παρουσιάζουν μια χαρακτηριστική υφή, η οποία τις κάνει ορατές ακόμα και σε εικόνες χαμηλής ανάλυσης. Διάφοροι ερευνητές έχουν προσπαθήσει να εκμεταλλευτούν αυτή την ιδιότητα για την εύρεση της πινακίδας κυκλοφορίας. Για παράδειγμα, μία πολύ καλή περιγραφή της υφής είναι τα φίλτρα Gabor. Οι Kahraman κ.συν. [19] δημιουργούν 12 φίλτρα Gabor, για διαφορετικά μεγέθη και σε διαφορετικές κατευθύνσεις. Η αρχική εικόνα συνελίσσεται με καθένα από αυτά τα φίλτρα ώστε να προκύψει ενίσχυση των περιοχών με υφή όμοια με αυτήν της πινακίδας. Οι περιοχές αυτές ενοποιούνται με μορφολογικές μεθόδους και έτσι προκύπτει η περιοχή της πινακίδας. Η μέθοδος αυτή παρουσιάζει πολυπλοκότητα ανάλογη του γινομένου του μεγέθους της εικόνας με το μέγεθος του φίλτρου, άρα είναι γρήγορη μόνο για εικόνες μικρού μεγέθους και με μικρή περιοχή πινακίδας.

Όμως υπάρχουν και απλούστεροι περιγραφείς υφής. Για παράδειγμα, η μέση τιμή της φωτεινότητας και η τυπική της απόκλιση σε μια περιοχή γύρω από κάθε pixel είναι μεγέθη που στην περιοχή της πινακίδας λαμβάνουν διαφορετικές τιμές σε σχέση με την υπόλοιπη εικόνα. Το γεγονός αυτό εκμεταλλεύονται οι Αναγνωστόπουλος κ.συν. [20] για την επίλυση του προβλήματος ανίχνευσης. Έτσι, αφού υπολογίζουν την τιμή των δύο αυτών παραμέτρων για κάθε pixel, στη συνέχεια γίνεται σάρωση της εικόνας ταυτόχρονα με δύο ομόκεντρα ορθογώνια διαφορετικού μεγέθους, και συγκρίνονται οι δύο αυτές παράμετροι για τα δύο ορθογώνια επιλέγονται οι περιοχές στις οποίες αυτές παρουσιάζουν στατιστική διαφοροποίηση πάνω από ένα όριο. Στη συνέχεια βρίσκουν τον αριθμό αντικειμένων μέσα σε κάθε υποψήφια περιοχή ώστε να επιλέξουν αυτή που πιθανότατα αντιστοιχεί σε πινακίδα.

Μια ακόμη μέθοδος για την ενίσχυση περιοχών εντονότερης υφής είναι η χρήση κυματιδίων (wavelets). Οι Kumar κ.συν. [21] χρησιμοποιούν ολικώς ταιριασμένα κυματίδια (globally matched wavelets) για την κατηγοριοποίηση των υφών. Η μέθοδός τους χρησιμοποιείται γενικά για την ανίχνευση κειμένου, όχι μόνο σε προβλήματα εύρεσης πινακίδων

κυκλοφορίας. Δημιουργούν ένα διάνυσμα 6 διαστάσεων και με βάση αυτό εκπαιδεύουν ένα SVM για την κατηγοριοποίηση των περιοχών και την εύρεση των περιοχών κειμένου. Από την άλλη, οι Hsieh κ.συν. [22] χρησιμοποιούν για το μετασχηματισμό κυματιδίων συνάρτηση κλιμάκωσης Haar, και επιλέγουν τις περιοχές όπου η μετασχηματισμένη εικόνα παίρνει τις μέγιστες τιμές. Στη συνέχεια επιλέγεται η περιοχή με τη σωστή αναλογία πλάτους-ύψους και τη μεγαλύτερη μέση τιμή εντός της ως περιοχή πινακίδας.

2.1.3 Μέθοδοι υπολογιστικής μάθησης

Χαρακτηριστικό των δύο προηγούμενων προσεγγίσεων είναι η χρήση μεθόδων που χρησιμοποιούν εμπειρικές τιμές για τα κατώφλια ή χρησιμοποιούν κάποιον απλό ταξινομητή όπως τα SVM. Όμως, καθώς πληθαίνει η ανάγκη για σχεδόν τέλεια αποτελέσματα σε μικρούς χρόνους για συγκεκριμένες εφαρμογές, η υπολογιστική νοημοσύνη χρησιμοποιείται συχνότερα. Αυτό συμβαίνει γιατί οι ταξινομητές δίνουν πολύ καλά αποτελέσματα σε γρήγορους χρόνους, όταν έχει προηγηθεί η εκπαίδευσή τους με ένα σύνολο δεδομένων παρόμοιο με αυτό στο οποίο θα κληθούν να ανιχνεύσουν την πινακίδα κυκλοφορίας. Και πάλι όμως τα χαρακτηριστικά που χρησιμοποιούνται είναι οι ακμές στην περιοχή της πινακίδας και η υφή της. Ο πιο συχνός ταξινομητής που χρησιμοποιείται είναι τα νευρωνικά δίκτυα, που εκπαιδεύονται με διάφορες παραμέτρους της εικόνας στην περιοχή της πινακίδας. Από άλλους ερευνητές έχουν χρησιμοποιηθεί γενετικοί αλγόριθμοι και ασαφής λογική για την εκπαίδευση των ταξινομητών.

Μια απλή χρήση των νευρωνικών δικτύων γίνεται από τους E. R. Lee κ.συν. [23], οι οποίοι αναζητούν πινακίδες με βάση το χρώμα φόντου, που δεν είναι λευκό, και έτσι εκπαιδεύουν ένα νευρωνικό δίκτυο ώστε να αναγνωρίζει αυτό το χρώμα. Αυτό έχει 24 κόμβους εισόδου, οι οποίοι αντιπροσωπεύουν τις τρεις τιμές του χρώματος σε χώρο HLS για τα οχτώ γειτονικά pixel κάθε σημείου, έχει 30 ενδιάμεσους κόμβους και 4 κόμβους εξόδου, που αντιστοιχούν στα χρώματα πράσινο, κόκκινο, λευκό και άλλο (νευρωνικό δίκτυο 24-30-4). Για την εκπαίδευσή του χρησιμοποιείται αλγόριθμος οπισθοδιάδοσης σφάλματος. Έτσι εντοπίζονται οι περιοχές με επιθυμητό χρώμα, και επιλέγονται αυτές που έχουν σχέση πλάτους προς ύψος μέσα στα σωστά όρια.

Οι Kim και Chien [24] χρησιμοποιούν γενικευμένο μετασχηματισμό συμμετρίας (generalized symmetry transform – GST) για να σαρώσουν υπό διάφορες γωνίες την εικόνα, στην οποία έχουν εντοπιστεί οι ακμές με χρήση τελεστή Sobel και να βρουν τους άξονες συμμετρίας της. Έτσι εντοπίζουν τις τέσσερις γωνίες της πινακίδας, από τις οποίες βρίσκουν και τον προσανατολισμό της, με σκοπό να την περιστρέψουν και να αλλάξουν το μέγεθός της με προοπτικό μετασχηματισμό. Με αυτό τον τρόπο κανονικοποιούν τις διαστάσεις της πινακίδας και χρησιμοποιούν την εικόνα των ακμών της για να φτιάξουν ένα πλέγμα 10x20 σημείων, τα οποία μπαίνουν στην είσοδο ενός νευρωνικού δικτύου 200-5-2 που αποφασίζει αν η εικόνα αντιστοιχεί σε πινακίδα κυκλοφορίας ή όχι.

Από την άλλη, οι K. K. Kim κ.συν. [25] εντοπίζουν πρώτα το όχημα σε μια ακολουθία από καρέ και εκπαιδεύουν δύο TDNN (time delay neural network) για τη μοντελοποίηση του χρώματος και της υψής της πινακίδας, ένα για την οριζόντια και ένα για την κάθετη διεύθυνση. Οι περιοχές με επιθυμητές τιμές για αυτές τις παραμέτρους και στις δύο διευθύνσεις ενοποιούνται και διατηρείται αυτή με το καταλληλότερο σχήμα, μέγεθος και λόγο πλάτους προς ύψος.

Τα νευρωνικά δίκτυα υπάρχουν σε πολλές παραλλαγές, και η χρήση τους διαφέρει μεταξύ των ερευνών που παρουσιάζονται. Οι ter Brugge κ.συν. [26] χρησιμοποιούν DTCNN (discrete-time cellular NN) για να εξετάσουν τη φωτεινότητα και την υφή στην περιοχή κάθε pixel, ώστε να επιλεγούν αυτά που έχουν τιμές εντός κάποιων ορίων για την κάθε παράμετρο. Στη συνέχεια εφαρμόζεται closing και επιλέγονται οι περιοχές στις οποίες συμπίπτουν σωστές τιμές και για τις δύο παραμέτρους. Τέλος, επιλέγονται οι περιοχές σωστού μεγέθους και σχήματος. Οι Chacon και Zimmermann [27], από την άλλη, χρησιμοποιούν PCNN (pulse coupled NN) για να εντοπίσουν περιοχές της εικόνας τις οποίες αξιολογούν εφαρμόζοντας μετασχηματισμό Fourier και εξάγοντας στατιστικά μεγέθη από αυτόν εντός της περιοχής.

Μια άλλη μέθοδος υπολογιστικής νοημοσύνης που χρησιμοποιείται είναι η ασαφής λογική (fuzzy logic). Δηλαδή παράμετροι της εικόνας στην περιοχή της πινακίδας περιγράφονται με ασαφείς μεταβλητές, κάνοντας τον αλγόριθμο πιο εύρωστο σε διαφορετικές συνθήκες λήψης της εικόνας. Για παράδειγμα, οι S.-L. Chang κ.συν. [28] δημιουργούν έναν ασαφή ταξινομητή, εκμεταλλευόμενοι τη χρωματική διαφορά ανάμεσα στην πινακίδα και το φόντο της. Εξετάζουν τέσσερις παραμέτρους σε κάθε pixel: τις τρεις τιμές του στο χρωματικό χώρο HSI και το αν ανήκει σε χρωματική ακμή (δηλαδή ακμή λευκού-μαύρου, λευκού-κόκκινου ή λευκού-πράσινου). Η χρήση ασαφούς λογικής μοντελοποιεί τις ατέλειες της εικόνας, αφού για την ένταξη του pixel σε περιοχή πινακίδας γίνεται άθροιση των τιμών των ασαφών μεταβλητών στη γειτονιά του, κάτι που σε απλή λογική δεν θα έδινε εξίσου καλά αποτελέσματα.

Οι Nijhuis κ.συν. [29] μοντελοποιούν τους ασαφή ορισμούς της περιοχής της πινακίδας, δηλαδή ότι αυτή είναι μια κίτρινη περιοχή που περιέχει κάποιους μαύρους χαρακτήρες και ότι αυτή βρίσκεται στο κέντρο του κάτω μέρους της όψης του αυτοκινήτου. Για τον πρώτο, εκπαιδεύουν τον ταξινομητή με βάση τη συχνότητα εμφάνισης των συνδυασμών RGB για το κίτρινο χρώμα του φόντου της πινακίδας σε εικόνες εκπαίδευσης, ενώ για το δεύτερο ορισμό χρησιμοποιείται η υφή, η οποία προκύπτει από τις τιμές φωτεινότητας των 8 γειτόνων κάθε pixel. Οι συνεκτικές περιοχές που προκύπτουν αξιολογούνται ως προς το μέγεθος και τις διαστάσεις και επιλέγονται οι υποψήφιες περιοχές πινακίδας.

Ένα τρίτο παράδειγμα μεθόδου υπολογιστικής νοημοσύνης είναι η χρήση γενετικών αλγορίθμων. Οι γενετικοί ή εξελικτικοί αλγόριθμοι μιμούνται τη διαδικασία της φυσικής επιλογής, ώστε από τις διάφορες υποψήφιες περιοχές να επιλεγούν αυτές που έχουν τα επιθυμητά χαρακτηριστικά και αυτά να συνδυαστούν ώστε να παράξουν ακόμη καλύτερες περιοχές. Οι S. K. Kim κ.συν. [30] χρησιμοποιούν κατανεμημένο γενετικό αλγόριθμο που

χρησιμοποιεί ως χαρακτηριστικό τις τιμές χρώματος HLS στη γειτονιά του κάθε ενός από 16 τυχαία pixel της εικόνας. Η επιλογή των pixel μετακινείται κάθε φορά προς την κατεύθυνση του πιο ισχυρού γείτονα και ενοποιείται με αυτόν μέχρι να υπάρξει σύγκλιση, οπότε γίνεται κατάτμηση της εικόνας σε 16 περιοχές, που κάθε μία έχει παρόμοιες τιμές HLS. Στη συνέχεια επιλέγεται η περιοχή της πινακίδας με βάση το χρώμα (όπως προαναφέρθηκε, οι πινακίδες στην Κορέα έχουν πράσινο φόντο) και το λόγο πλάτους προς ύψος.

Μία διαφορετική χρήση γενετικών αλγορίθμων είναι αυτή των Yoshimori κ.συν. [31], όπου αυτοί χρησιμοποιούνται για την προσαρμογή του ταξινομητή. Συγκεκριμένα, το άνω και το κάτω όριο της τιμής φωτεινότητας που πρέπει να έχει ένα pixel ώστε να ανήκει σε περιοχή πινακίδας προκύπτουν ως αποτέλεσμα του γενετικού αλγορίθμου μέσα στην περιοχή του οχήματος, το οποίο έχει ήδη εντοπιστεί συγκρίνοντας διαδοχικά καρέ μεταξύ τους. Έτσι ο αλγόριθμος προσαρμόζει τους ταξινομητές σε διαφορετικές συνθήκες λήψης της εικόνας. Στη συνέχεια γίνεται εντοπισμός της περιοχής της πινακίδας με τη βοήθεια ιδιοτήτων συμμετρίας.

2.1.4 Άλλες μέθοδοι

Στην υποενότητα αυτή εντάσσονται οι μέθοδοι εύρεσης περιοχών εικόνας, οι οποίες αποτελούν τη βάση του μεγαλύτερου μέρους της παρούσας εργασίας. Μία από τις πρώτες εργασίες που χρησιμοποιούν αυτή την προσέγγιση είναι αυτή των Naito κ.συν. [32], στην οποία οι περιοχές της εικόνας εντοπίζονται με δυναμική κατωφλίωση. Γίνεται δηλαδή μετατροπή της εικόνας σε δυαδική με κατώφλι που δεν είναι ίδιο για όλη την εικόνα, αλλά προσαρμόζεται ανάλογα με την τοπική μέση φωτεινότητα [33]. Οι περιοχές που εντοπίζονται με αυτόν τον τρόπο απαριθμούνται, και εξετάζεται η σχετική τους θέση για τον εντοπισμό της πινακίδας. Συγκεκριμένα, πρέπει τουλάχιστον τρεις περιοχές να βρίσκονται στην ίδια ευθεία, να έχουν ίδιο μέγεθος και να έχουν γύρω τους άλλες περιοχές όπως παύλες, τελείες και διακριτικά της πινακίδας.

Όμως κατά της τελευταία δεκαετία έχουν αναπτυχθεί πιο αποτελεσματικοί αλγόριθμοι εύρεσης περιοχών σε μια εικόνα. Μία από αυτές, που παρουσιάζει πολλά πλεονεκτήματα, τα οποία θα αναλυθούν σε επόμενο κεφάλαιο, είναι οι μέγιστα σταθερές περιοχές ακροτάτων (maximally stable extremal regions – MSER). Η πρώτη τους χρήση για εντοπισμό της πινακίδας γίνεται από τους Matas, Zimmermann [34], οι οποίοι, αφού εντοπίσουν τις περιοχές της εικόνας, τις ταξινομούν ως χαρακτήρες ή μη χαρακτήρες με βάση ένα νευρωνικό δίκτυο, το οποίο έχουν εκπαιδεύσει με μεγάλο αριθμό χαρακτήρων (1600), σε διάφορα μεγέθη και υπό διάφορες γωνίες λήψης. Η πινακίδα εντοπίζεται εκεί όπου υπάρχουν τουλάχιστον τρεις περιοχές χαρακτήρων όμοιου μεγέθους, οι οποίες είναι συνευθειακές, κάτι που επιβεβαιώνεται με χρήση μετασχηματισμού Hough. Ο συγκεκριμένος αλγόριθμος είναι πολύ αποδοτικός σε συνθήκες μερικής κάλυψης της πινακίδας από εμπόδιο, κάτι όμως που σπάνια αποτελεί πρόβλημα σε πραγματικά συστήματα.

Μια άλλη προσέγγιση της χρήσης MSER γίνεται από τους Donoser, Arth, Bischof [6], και είναι αυτή πάνω στην οποία βασίστηκε μεγάλο μέρος της παρούσης εργασίας. Το πλεονέκτημά της είναι ότι δεν απαιτεί κάποιο σχήμα εκπαίδευσης. Συγκεκριμένα, η διαδικασία εντοπισμού της πινακίδας ξεκινά εντοπίζοντας τις φωτεινές περιοχές με σκούρο περίγραμμα (MSER+) και τις σκούρες περιοχές με ανοιχτόχρωμο περίγραμμα (MSER-). Αναζητούνται περιοχές MSER+ που περιέχουν έναν αριθμό MSER- στο εσωτερικό τους, τα οποία έχουν παρόμοιο μέγεθος μεταξύ τους και με την περιοχή που τα περικλείει και βρίσκονται σε ευθεία γραμμή.

Έχουν χρησιμοποιηθεί όμως και άλλες μέθοδοι που δεν εντάσσονται στις παραπάνω κατηγορίες. Μια ενδιαφέρουσα μέθοδος που αφορά το πεδίο της συχνότητας αναπτύχθηκε από τους Parisi κ.συν. [35], οι οποίοι εκμεταλλεύονται το γεγονός της περίπου σταθερής απόστασης μεταξύ των χαρακτήρων εντός της πινακίδας. Συγκεκριμένα, υπολογίζουν το διακριτό μετασχηματισμό Fourier μίας διάστασης (1D-DFT) σε κάθε γραμμή της εικόνας και αναζητούν τις γραμμές αυτές στις οποίες η αρμονική που αντιστοιχεί στην αναμενόμενη απόσταση μεταξύ των χαρακτήρων λαμβάνει μέγιστη τιμή. Στο πεδίο του χώρου αυτό προϋποθέτει σταθερό μέγεθος πινακίδας, δηλαδή και σταθερή απόσταση μεταξύ του οχήματος και της κάμερας. Επομένως, μπορεί να εφαρμοστεί μόνο σε περιπτώσεις που η θέση του οχήματος είναι εκ των προτέρων γνωστή με σχετική ακρίβεια, όπως στην περίπτωση των διοδίων ή της εισόδου σε χώρο στάθμευσης.

Ο L. Diagnekon [36] χρησιμοποιεί τη μέθοδο adaptive boosting (AdaBoost), η οποία χρησιμοποιείται με μεγάλη επιτυχία στον εντοπισμό ανθρώπινων προσώπων εντός της εικόνας. Η μέθοδος αυτή στηρίζεται στο συνδυασμό πολλών ασθενών ταξινομητών για την εύρεση της σωστής περιοχής. Δηλαδή δεν επιδιώκει να κατασκευάσει έναν ταξινομητή με ποσοστό επιτυχίας άνω του 90%, αλλά εκπαιδεύει απλούς ταξινομητές της μορφής που φαίνονται στην εικόνα. Ο κάθε ένας από αυτούς έχει ποσοστό επιτυχίας λίγο μεγαλύτερο από 50%, που δίνει όμως γρήγορα αποτελέσματα ταξινόμησης και αν συνδυαστεί με ικανό αριθμό από άλλους όμοιους ταξινομητές μπορεί να επιτύχει υψηλά ποσοστά εγκυρότητας.

Οι Duan κ.συν. [37], αφού πραγματοποιούν αναζήτηση ακμών με τελεστή Sobel, στη συνέχεια αναζητούν τις κλειστές καμπύλες και με χρήση μετασχηματισμού Hough αναζητούν ποιες από αυτές τις καμπύλες περιέχουν παράλληλες ευθείες. Έτσι καταλήγουν με έναν μικρό αριθμό παραλληλόγραμμων σχημάτων, και για να επιλέξουν την πινακίδα μετρούν τις κάθετες ακμές στον άξονά τους.

Οι Cano και Perez-Cortez [38], αφού ομαλοποιήσουν την εικόνα και βρουν τις ακμές με τελεστή Sobel, εκπαιδεύουν τον ταξινομητή με ανάλυση κύριων συνιστωσών (principal component analysis – PCA) στην περιοχή κάθε pixel της εικόνας των ακμών, για να μειωθεί η διάσταση του ταξινομητή.

Οι Abolghasemi και Ahmadyfard [39] συνδυάζουν διάφορες μεθόδους. Αρχικά ενισχύουν την εικόνα στις περιοχές με μεγάλη πυκνότητα ακμών (χαρακτηριστικό υφής), και ύστερα ενοποιούν με μορφολογικό φίλτρο τις περιοχές με πολλές κάθετες ακμές ώστε να εξάγουν τις

υποψήφιες περιοχές. Αποκλείουν κάποιες από αυτές με βάση το σχήμα και την αναλογία των διαστάσεών τους, ενώ η τελική επιλογή γίνεται με υπολογισμό της συσχέτισης της κάθε περιοχής με το μοντέλο μιας περιοχής πινακίδας (στατιστικό χαρακτηριστικό)

Μια ενδιαφέρουσα μέθοδος αναπτύχθηκε από τους Zunino και Rovetta [40], οι οποίοι χρησιμοποιούν κβαντισμό φορέα (vector quantization) στην εικόνα, έχοντας ως διπλό όφελος την κωδικοποίηση. Αυτή μπορεί να οδηγήσει στη συμπίεση του μεγέθους της εικόνας, αλλά και στην εξαγωγή χαρακτηριστικών για κάθε κομμάτι της εικόνας, τα οποία αν αξιολογηθούν σωστά μπορούν να αποτελέσουν τη βάση ταξινομητή για τον εντοπισμό της πινακίδας κυκλοφορίας.

2.2 *Ο διαχωρισμός των χαρακτήρων*

Ο διαχωρισμός των χαρακτήρων αποτελεί το πρώτο μέρος της διαδικασίας αναγνώρισής τους, αφού, ακόμα και σε σύγχρονα εργαλεία οπτικής αναγνώρισης χαρακτήρων, τα οποία αναγνωρίζουν τους χαρακτήρες στο πλαίσιο της λέξης στην οποία ανήκουν, προηγείται μια εκτίμηση για τους μεμονωμένους χαρακτήρες. Επίσης, σε αρκετά προβλήματα, όπως και στο παρόν της αναγνώρισης πινακίδων κυκλοφορίας, η δυνατότητα αναγνώρισης ολόκληρης της λέξης δεν είναι χρήσιμη, και μάλιστα οδηγεί σε περισσότερα σφάλματα σε σχέση με την απλή αναγνώριση των χαρακτήρων. Άρα, ένα ενδιάμεσο στάδιο διαχωρισμού των χαρακτήρων είναι απαραίτητο. Στις εργασίες που έχουμε συζητήσει ως τώρα έχουν χρησιμοποιηθεί διαφορετικές μέθοδοι, οι οποίες θα παρουσιαστούν στην παρούσα ενότητα.

Το αρχικό στάδιο του διαχωρισμού των χαρακτήρων της πινακίδας, τα όρια της οποίας έχουν ήδη εντοπιστεί, είναι η μετατροπή της εικόνας σε δυαδική (binarization). Ιδανικά πρέπει το περίγραμμα των χαρακτήρων και η περιοχή εντός να έχει μία απόχρωση, συνήθως μαύρη, και το φόντο, το υπόβαθρο δηλαδή της πινακίδας, λευκό. Παρά την προσπάθεια να διακρίνονται ξεκάθαρα οι χαρακτήρες από το φόντο στις πινακίδες όλων των χωρών, μεταξύ άλλων μέσω χρήσης ανακλαστικού υλικού και ανάγλυφης επιφάνειας, το οπτικό αποτέλεσμα δεν είναι πάντα ξεκάθαρο. Αιτία μπορεί να είναι σκιάσεις από μέρη του ίδιου του οχήματος ή από αντικείμενα όπως δέντρα και κολώνες στην άκρη του δρόμου. Ακόμα μπορεί να φταίει η παρουσία σκόνης στην επιφάνεια που μειώνει την ανακλαστικότητά της, η ύπαρξη ανακλάσεων φωτός ή έντονο φως από τους προβολείς, ή και η σκόπιμη μείωση της ανακλαστικότητας της πινακίδας με χρήση πλαστικών, διάφανων καλυμμάτων.

Για τη μετατροπή της εικόνας σε δυαδική επιλέγεται ένα κατώφλι φωτεινότητας, και όσα pixel έχουν φωτεινότητα μεγαλύτερη από αυτό καταχωρούνται ως λευκά και όσα έχουν μικρότερη ως μαύρα. Το κατώφλι μπορεί να είναι σταθερό σε ολόκληρη την εικόνα της, επιλέγοντας ως τιμή του τη μέση τιμή φωτεινότητας εντός της πινακίδας [16], [41] ή κάποια άλλη που προκύπτει με βάση την κατανομή φωτεινότητας στο ιστόγραμμα [17], [42].

Μπορεί όμως και να είναι και προσαρμοζόμενο, δηλαδή να μην έχει σταθερή τιμή, αλλά να εξαρτάται από την τοπική μέση φωτεινότητα, ή τη φωτεινότητα κοντά σε ακμές [11], [14], [20], [28], [32], [35].

Υπάρχει και η επιλογή να γίνει η κατάτμηση με βάση τις αρχικές τιμές φωτεινότητας, χωρίς δηλαδή να γίνει κατωφλιώση, όπως για παράδειγμα στην εργασία [25], οι οποίοι υπολογίζουν το άθροισμα των φωτεινότητων των pixel της κάθε στήλης της εικόνας της πινακίδας, μαζί με τη μεταβλητότητά τους, ώστε να προκύψει κατάτμηση εκεί όπου το άθροισμα παρουσιάζει τοπικό ελάχιστο. Εφόσον όμως η εικόνα έχει μετατραπεί σε δυαδική, υπάρχουν δύο βασικές προσεγγίσεις της κατάτμησης των χαρακτήρων.

Η πρώτη εκμεταλλεύεται το ιστόγραμμα προβολών, το οποίο έχει χρησιμοποιηθεί και για τον εντοπισμό της ίδιας της πινακίδας, στη μέθοδο που περιγράφεται στην ενότητα 2.1.1. Εδώ όμως δεν προβάλλουμε την εικόνα των ακμών στον κάθετο άξονα, αλλά την εικόνα των ίδιων των χαρακτήρων στον οριζόντιο άξονα. Έτσι, οι στήλες που αντιστοιχούν σε χαρακτήρες θα περιέχουν πολλά μαύρα pixel, ενώ τα κενά ανάμεσά τους λιγότερα. Με αυτό τον τρόπο μπορεί να γίνει ενοποίηση των στηλών του κάθε χαρακτήρα μεταξύ τους, και διαχωρισμός τους από τους άλλους χαρακτήρες. Ένα πρόβλημα που ανακύπτει σε αυτή τη μέθοδο είναι η επιλογή του κατωφλίου στο ιστόγραμμα προβολής, ώστε μια στήλη να θεωρηθεί ότι περιέχει χαρακτήρα. Εάν αυτό επιλεγεί μικρό μπορεί να ενώσει κάποιους χαρακτήρες ως έναν αν υπάρχει ανάμεσά τους ένας αριθμός μαύρων pixel λόγω θορύβου που δεν έχει αφαιρεθεί, ενώ αν επιλεγεί μεγάλο μπορεί να διαχωρίσει έναν χαρακτήρα που περιέχει στένωση, όπως το Κ ή το Η, σε δύο χαρακτήρες. Επίσης, αν η πινακίδα δεν είναι τελείως οριζόντια αλλά έχει κάποια κλίση, οι προβολές δύο διαφορετικών χαρακτήρων θα επικαλύπτονται, κάνοντας αδύνατο το διαχωρισμό τους, άρα πριν την έναρξη της διαδικασίας πρέπει η πινακίδα να έχει ευθυγραμμιστεί.

Η δεύτερη βασική προσέγγιση περιλαμβάνει τη χρήση συνδεδεμένων συνιστωσών (connected components), δηλαδή συνόλων pixel που αντιστοιχούν σε μια περιοχή ίδιας φωτεινότητας και είναι κοντά το ένα στο άλλο. Κάθε τέτοια συνιστώσα ή περιοχή αποτελείται από το σύνορο και το περιεχόμενό της. Αν η εικόνα έχει μετατραπεί σωστά σε δυαδική, κάθε τέτοια περιοχή θα αντιστοιχεί σε έναν χαρακτήρα ή σε κάποιο άλλο σημείο της πινακίδας, όπως παύλες, οπές για τις βίδες, σφραγίδες κτλ. Το πλεονέκτημα αυτών των περιοχών είναι ότι μπορούν σχετικά γρήγορα να εξαχθούν από αυτές πληροφορίες όπως το μέγεθος, το ύψος, το σχήμα κ.ά., οι οποίες μπορούν να εκτιμηθούν ώστε να διατηρηθούν μόνο οι περιοχές που αντιστοιχούν σε χαρακτήρες. Επίσης εξετάζεται αν είναι οι περιοχές ευθυγραμμισμένες, συνήθως με χρήση μετασχηματισμού Hough.

Όμως αν η εικόνα δεν είναι καθαρή ή έχει επιλεγεί λάθος κατώφλι στο προηγούμενο βήμα, ή κάποιο αντικείμενο όπως η βίδα στερέωσης της πινακίδας παρεμβάλλεται μεταξύ δύο χαρακτήρων, μπορεί μια περιοχή να αντιστοιχεί σε δύο ή περισσότερους χαρακτήρες, να έχουμε δηλαδή λανθασμένη κατάτμηση. Αυτό επιχειρείται να διορθωθεί με χρήση μορφολογικών φίλτρων όπως erosion, opening, thickening κ.ά. για την απομάκρυνση των

περιοχών που δημιουργούν θόρυβο και την κατάτμηση στο σημείο στένωσης μεταξύ δύο ενωμένων χαρακτήρων [43] ή με εκμετάλλευση της πρότερης γνώσης του πλήθους των χαρακτήρων της πινακίδας, όπως και του μεγέθους τους και της απόστασης μεταξύ τους (π.χ. [29], [28], [37]).

Υπάρχουν όμως και άλλες μέθοδοι που έχουν αναπτυχθεί για το πρόβλημα του διαχωρισμού των χαρακτήρων. Για παράδειγμα, οι Lee, Chen και Wang [13] χρησιμοποιούν κάποιες εικόνες από πινακίδες για εκπαίδευση του συστήματος διαχωρισμού των χαρακτήρων, στις οποίες σημειώνουν τα όρια των χαρακτήρων, αφού το μέγεθος και η απόστασή τους είναι προκαθορισμένα σε κάθε χώρα. Έτσι το σύστημα αυτό γνωρίζει πού πρέπει να διαχωρίσει την πινακίδα, ώστε σε κάθε τμήμα να υπάρχει ένας χαρακτήρας. Στη συνέχεια κανονικοποιείται το μέγεθος της πινακίδας η οποία ανευρέθη, ώστε το περίγραμμά της να συμπίπτει με το περίγραμμα των εικόνων εκπαίδευσης, άρα και οι χαρακτήρες εντός της να διαχωριστούν σωστά. Για να λειτουργήσει όμως αυτή η μέθοδος, πρέπει το όριο της πινακίδας να έχει βρεθεί με μεγάλη ακρίβεια, και να είναι αυτή εντελώς ορθογώνια και οριζόντια.

Άλλες, πιο σύγχρονες μέθοδοι περιλαμβάνουν τη χρήση τυχαίων πεδίων Markov (Markov random fields – MRF) ή διανυσματικού κβαντισμού (vector quantization – VQ). Στην πρώτη μέθοδο που εφαρμόζεται από τους Cui και Huang [44] γίνεται χρήση πιθανοτικών μεθόδων και γενετικών αλγορίθμου ώστε να κατασκευαστεί το MRF και να γίνει διαχωρισμός των χαρακτήρων, με βάση διαδοχικές εικόνες του κάθε οχήματος. Η δεύτερη μέθοδος χρησιμοποιείται από τους Kahraman κ.συν. [19], οι οποίοι επιχειρούν να ταξινομήσουν κάθε pixel της εικόνας της πινακίδας σε ένα από δύο διανύσματα. Αυτά έχουν προκύψει έτσι ώστε το μέσο τετραγωνικό σφάλμα ανάμεσα στην τιμή φωτεινότητας των pixel τα οποία τα αποτελούν και τη μέση τιμή τους να γίνει ελάχιστη. Επιλέγονται δύο διανύσματα, ώστε το ένα να αντιστοιχεί στους χαρακτήρες και το άλλο στο φόντο. Τα pixel των χαρακτήρων στη συνέχεια διαχωρίζονται σε συνδεδεμένες συνιστώσες, ώστε κάθε συνιστώσα να αντιστοιχεί σε ένα γράμμα.

2.3 Η αναγνώριση των χαρακτήρων

Το τρίτο βήμα της αναγνώρισης της πινακίδας είναι η ανάγνωση των χαρακτήρων που περιέχει. Το πρόβλημα της οπτικής αναγνώρισης χαρακτήρων (optical character recognition – OCR) έχει γίνει αντικείμενο μελέτης από παλιά και συνεχίζει να είναι ένα πολύ δραστήριο πεδίο έρευνας. Οι εφαρμογές της είναι αμέτρητες, ενώ συνεχώς εμφανίζονται νέες με χρήση σε φορητές συσκευές. Πολλές από αυτές επικεντρώνονται στην ανάγνωση κειμένων από σαρωμένες εικόνες από έντυπα, ενώ πιο σύγχρονες περιλαμβάνουν την εύρεση του κειμένου σε σύνθετες εικόνες, ή την ανάγνωση χειρόγραφου κειμένου. Χαρακτηριστικό των λύσεων είναι η πολυπλοκότητά τους, καθώς θα πρέπει να αντιμετωπίσουν, πέρα από τις διαφορετικές

συνθήκες φωτισμού και αντίθεσης της εικόνας, προβλήματα θορύβου, περιστραμμένου κειμένου, διαφορετικών γραμματοσειρών, ξένων συμβόλων κ.ά.

Η ανάγνωση του περιεχομένου της πινακίδας όμως παρέχει κάποιες διευκολύνσεις σε αυτό το πρόβλημα. Συγκεκριμένα, τα όρια του κειμένου έχουν βρεθεί από πριν, ενώ και η γραμματοσειρά είναι συνήθως συγκεκριμένη για ένα σύνολο εικόνων, καθώς είναι προκαθορισμένη η γραμματοσειρά που χρησιμοποιείται σε κάθε χώρα. Αν ο διαχωρισμός των χαρακτήρων έχει γίνει σωστά, με κάποιον από τους τρόπους που περιγράφηκαν στην προηγούμενη υποενότητα, υπάρχουν δύο κύριες προσεγγίσεις για την αναγνώριση των χαρακτήρων. Η πρώτη χρησιμοποιεί την ταύτιση προτύπων (template matching), δηλαδή δημιουργεί ένα πρότυπο για κάθε χαρακτήρα, και στη συνέχεια μετρά την ομοιότητα των προς διερεύνηση χαρακτήρων με κάθε πρότυπο. Η δεύτερη χρησιμοποιεί νευρωνικά δίκτυα, που στην είσοδό τους δέχονται κάποια χαρακτηριστικά της εικόνας, και στην έξοδό τους δίνουν την εκτίμησή τους για τον χαρακτήρα στον οποίο αντιστοιχεί η κάθε εικόνα.

2.3.1 Ταύτιση προτύπων

Η μέθοδος αυτή ακολουθεί την εξής ροή εκτέλεσης. Αρχικά εξάγονται κάποια χαρακτηριστικά από πρότυπες εικόνες για κάθε χαρακτήρα. Αυτές οι εικόνες μπορεί να έχουν προκύψει από παραδείγματα από το σύνολο δεδομένων ή με χρήση κάποιας γραμματοσειράς, στην οποία έχουν γίνει προσαρμογές ώστε να λαμβάνει υπ' όψιν πιθανές παραμορφώσεις της εικόνας. Τα χαρακτηριστικά μπορεί να είναι απλά, όπως η τιμή φωτεινότητας σε κάθε pixel, ή πιο σύνθετα, όπως κατευθυντήριες παράγωγοι. Στη συνέχεια εξάγονται τα ίδια χαρακτηριστικά για κάθε προς διερεύνηση χαρακτήρα και ύστερα γίνεται σύγκρισή τους με τα χαρακτηριστικά του προτύπου. Το πρότυπο με το οποίο η εικόνα παρουσιάζει τη μεγαλύτερη ομοιότητα επιλέγεται ως χαρακτήρας αναγνώρισης.

Για να γίνει σύγκριση μεταξύ των τιμών φωτεινότητας μεταξύ της εικόνας του χαρακτήρα προς αναγνώριση και του προτύπου, γίνεται πρώτα εξίσωση του μεγέθους των εικόνων. Στην περίπτωση της απλής κατωφλίωσης της εικόνας, όπως κάνουν οι Lee κ.συν. [23], αρκεί να βρεθεί η τιμή Jaccard για κάθε πρότυπο, και να επιλεγεί αυτό με τη μεγαλύτερη τέτοια τιμή.

Αν μιλάμε για δυαδικές εικόνες, η τιμή Jaccard ισούται με $\frac{n_{11}}{n_{11}+n_{10}+n_{01}}$, όπου n_{ij} ο αριθμός των pixel που έχουν την τιμή i στη μία εικόνα και την τιμή j στην άλλη.

Όμως το περίγραμμα κάθε χαρακτήρα μπορεί να είναι αρκετά διαφορετικό σε διάφορες απεικονίσεις του, και έτσι η απλή σύγκριση των τιμών φωτεινότητας θα οδηγεί σε μικρά ποσοστά αναγνώρισης. Μια προσέγγιση που χρησιμοποιούν οι Yung κ.συν. [45] κάνει χρήση μορφολογικών τελεστών για τη μείωση της επίδρασης τέτοιων αλλοιώσεων. Τα πρότυπα που δημιουργούν δεν έχουν απλά τιμή 1 εκεί που η εικόνα είναι τμήμα του χαρακτήρα και 0 εκτός του, αλλά προχωρούν σε thinning του χαρακτήρα, ώστε να έχει τιμή 1 μόνο σε μια στενή περιοχή, και στη συνέχεια επεκτείνουν τα όρια του χαρακτήρα με dilation και αναθέτουν εκεί

ενδιάμεσες τιμές φωτεινότητας. Προκύπτει δηλαδή ένα πρότυπο, που έχει τιμή 1 σε μια στενή λωρίδα κατά μήκος του χαρακτήρα, και ενδιάμεσες τιμές φωτεινότητας μέχρι και λίγο έξω από τα αρχικά όριά του. Έξω και από αυτήν την περιοχή, η εικόνα έχει την τιμή 0. Για τη σύγκριση κάθε χαρακτήρα με το αντίστοιχο πρότυπο λαμβάνονται υπ' όψιν μόνο οι περιοχές με τιμές 1 και 0, άρα υπάρχει το περιθώριο κάποιος χαρακτήρας να είναι ελαφρώς περιεστραμμένος ή τα όριά του να έχουν υποστεί παραμόρφωση. Στην ίδια γραμμή κινούνται και οι Shapiro κ.συν. [11], οι οποίοι χρησιμοποιούν επιπροσθέτως κάποια στοιχεία για τη θέση του χαρακτήρα μέσα στην πινακίδα, ώστε να καθορίσουν αν αυτή είναι γράμμα ή αριθμός, βελτιώνοντας τα ποσοστά αναγνώρισης για αμφιλεγόμενες περιπτώσεις.

Και οι Shi κ.συν. [17] χρησιμοποιούν αυτή τη γνώση, αφού σε κάθε χώρα είναι καθορισμένο σε ποιες θέσεις μέσα στην πινακίδα βρίσκονται αριθμοί και σε ποιες γράμματα της αλφαβήτου. Έτσι, αν για παράδειγμα ένας χαρακτήρας έχει μεγάλο λόγο ύψους προς πλάτος μπορούν αμέσως να συμπεράνουν ότι αντιστοιχεί σε I ή 1, ανάλογα αν βρίσκεται σε θέση αριθμού ή γράμματος. Για τους υπόλοιπους χαρακτήρες μετράται η ευκλείδεια απόσταση με κάθε πρότυπο και επιλέγεται αυτό με τη μικρότερη απόσταση. Επίσης, για χαρακτήρες όπως οι B και 8 ή οι O και D, που μοιάζουν αρκετά, γίνεται επιπλέον μελέτη του πάνω και κάτω αριστερού τμήματος, όπου εμφανίζεται η μικρότερη ομοιότητα, ώστε να γίνει η διάκριση μεταξύ τους.

Οι Naito κ.συν. [32] χρησιμοποιούν στοιχεία για το χαρακτήρα, όπως η θέση, το πλάτος και το ύψος ως επιπρόσθετη πληροφορία. Επίσης, δημιουργούν πολλαπλά πρότυπα για κάθε χαρακτήρα, για περιστροφές της εικόνας από -50° ως 50° . Οι Wang και Lee [14] χωρίζουν την εικόνα του χαρακτήρα σε 64 τμήματα ώστε κάθε τμήμα να έχει περίπου ίδιο αριθμό μαύρων pixel. Επίσης μετρά την απόσταση των ακμών του χαρακτήρα από το περιθώριο δηλαδή τα λευκά pixel στην οριζόντια και την κάθετη απόσταση. Αυτά τα χαρακτηριστικά συγκρίνονται με αντίστοιχα χαρακτηριστικά από πρότυπους χαρακτήρες και επιλέγεται αυτός με την μεγαλύτερη ομοιότητα. Τα ίδια χαρακτηριστικά χρησιμοποιούν και οι Lee κ.συν. [13], μαζί με πληροφορίες για την κατεύθυνση των ακμών.

Οι Rahman κ.συν. [46] εκπαιδεύουν τα πρότυπα με βάση τη γραμματοσειρά που μοιάζει περισσότερο με αυτή των πινακίδων, την Arial. Χωρίζουν την εικόνα σε 15 τμήματα και υπολογίζουν το ιστόγραμμα της σε κάθε ένα από αυτά. Τα ίδια ιστογράμματα υπολογίζονται και σε κάθε εικόνα προς διερεύνηση. Για να επιταχυνθεί η σύγκριση με κάθε πρότυπο εξετάζονται αρχικά κάποια από τα ιστογράμματα για να αποκλειστούν πολλές επιλογές και η τελική σύγκριση γίνεται με τα υπόλοιπα.

Οι Εμίρης και Κουλουριώτης [47] εκπαιδεύουν τον ταξινομητή με περίπου 15 εικόνες για κάθε χαρακτήρα που έχουν ληφθεί από διαφορετικές αποστάσεις και υπό διαφορετικές συνθήκες. Εντοπίζουν πρώτα την παύλα στην πινακίδα για να ξεχωρίσουν τους αριθμούς από τα γράμματα, και αφού γίνει κατάτμηση της εικόνας υπολογίζουν την συσχέτιση του κάθε χαρακτήρα με τα πρότυπα.

Ένα άλλο χαρακτηριστικό που χρησιμοποιείται πολύ είναι η προβολή της εικόνας στον οριζόντιο ή στον κάθετο άξονα. Για παράδειγμα οι Yan κ.συν. [15] υπολογίζουν τις προβολές αυτές καθώς και την περίμετρο του χαρακτήρα και την πυκνότητα της πληροφορίας σε κάθε κατεύθυνση για τη δημιουργία των προτύπων και τη σύγκριση κάθε χαρακτήρα με αυτά. Και εδώ χρησιμοποιείται ως κριτήριο η ελάχιστη ευκλείδεια απόσταση με κάθε πρότυπο.

Τέλος, οι Wang κ.συν. [48] υπολογίζουν επίσης τις προβολές στους δυο άξονες και χρησιμοποιούν την τεχνική που ονομάζεται δυναμική στρέβλωση προβόλων για την εύρεση της ελαχίστης απόστασης των προβόλων κάθε εικόνας με τα αντίστοιχα πρότυπα.

2.3.2 Νευρωνικά δίκτυα

Τα νευρωνικά δίκτυα στις διάφορες μορφές τους έχουν χρησιμοποιηθεί κατά κόρον στην αναγνώριση χαρακτήρων. Είναι ταξινομητές που μπορούν να εκφράσουν τα χαρακτηριστικά με μη διαισθητικούς για τον άνθρωπο τρόπους που όμως δίνουν καλά αποτελέσματα ταξινόμησης από τον υπολογιστή. Οι βασικές μορφές που χρησιμοποιούνται είναι perceptron πολλαπλών επιπέδων και χάρτες αυτό-οργάνωσης.

Για παράδειγμα, οι Nijhuis κ.συν. [29] εξάγουν κάποια χαρακτηριστικά από την εικόνα του χαρακτήρα, συγκεκριμένα την οριζόντια και την κάθετη προβολή καθώς και τον αριθμό των στοιχείων σε κάθε μια από τις δύο κατευθύνσεις. Δημιουργούν έτσι ένα perceptron 24 εισόδων. Ομοίως, οι Parisi [35] δημιουργούν ένα ασαφές νευρωνικό δίκτυο με 60 εισόδους, 35 εξόδους και 30 νευρώνες στο κρυφό επίπεδο. Η εκπαίδευσή του γίνεται με τη μέθοδο των ελαχίστων τετραγώνων. Ο Draghici [42] διαιρεί απλά την εικόνα σε 8x16 τμήματα και υπολογίζει τη φωτεινότητα στο καθένα. Έτσι προκύπτουν οι 128 είσοδοι του νευρωνικού δικτύου, το οποίο έχει 36 εξόδους και 20 νευρώνες στο κρυφό επίπεδο. Επίσης, το νευρωνικό δίκτυο επιστρέφει και το επίπεδο βεβαιότητας για κάθε ταξινόμηση. Αν αυτό δεν είναι ικανοποιητικό, γίνεται νέα κατάτμηση των χαρακτήρων και νέα προσπάθεια ταξινόμησης.

Οι Αναγνωστόπουλος κ.συν. [20] χρησιμοποιούν πιθανοτικό νευρωτικό δίκτυο. Η είσοδος περιέχει τις τιμές φωτεινότητας των $9 \times 12 = 108$ τμημάτων της εικόνας εισόδου. Επίσης, στο δίκτυο είναι ενσωματωμένος ένας μηχανισμός ανταγωνισμού μεταξύ των κόμβων του κρυφού επιπέδου. Υπολογίζεται το αποτέλεσμα μιας συνάρτησης και με βάση αυτό αποφασίζεται ο κόμβος του οποίου το βάρος θα αναβαθμιστεί. Η συνάρτηση κόστους που χρησιμοποιείται είναι η ευκλείδεια απόσταση.

Οι χάρτες αυτό-οργάνωσης Kohonen αποτελούνται από κόμβους οι όποιοι αλλάζουν θέση στο χώρο κατά τη διαδικασία εκπαίδευσής τους. Οι Adorni κ.συν. [16] χρησιμοποιούν ένα τέτοιο χάρτη, τροποποιημένο ώστε να δίνει και ποσοστό βεβαιότητας. Στη συνέχεια χρησιμοποιούν νευρωσασφή έλεγχο για τον επανέλεγχο των αποτελεσμάτων με μικρό ποσοστό βεβαιότητας.

Οι Chang κ.συν. [28] δίνουν πιο αναλυτική περιγραφή του χάρτη Kohonen που χρησιμοποιούν. Αρχικά απορρίπτουν τις περιοχές που έχουν μεγαλύτερο πλάτος από ύψος και στη συνέχεια ευθυγραμμίζουν τους χαρακτήρες με χρήση μετασχηματισμού hough. Μετά, με βάση τη θέση του χαρακτήρα μέσα στην πινακίδα καθορίζουν αν πρόκειται για γράμμα ή για αριθμό και μετρούν στον καθένα χαρακτηριστικές περιοχές ακμών για παράδειγμα σχήματος “T” ή “+” καθώς και τον αριθμό των οπών. Ο χάρτης Kohonen για κάθε χαρακτήρα αποτελείται από 40 σημεία από το εξωτερικό όριο του σχήματος της κάθε περιοχής. Σε κάθε χαρακτήρα που έρχεται προς διερεύνηση λαμβάνονται πάλι 40 τέτοια σημεία και για κάθε πρότυπο μετράται η μεταβολή τους μέχρι να αποκτήσουν το σχήμα του πρότυπου χαρακτήρα. Το πρότυπο για το οποίο η μετακίνηση που χρειάζεται είναι η μικρότερη είναι το αποτέλεσμα της ταξινόμησης. Τέλος, για αμφιλεγόμενα ζεύγη χαρακτήρων όπως τα 8/B και O/D γίνεται επιπλέον μελέτη των περιοχών στα όποια αυτά διαφέρουν.

2.3.3 Άλλες μέθοδοι

Οι μηχανές υποστηρικτικών διανυσμάτων (SVM) αποτελούν επίσης ένα χρήσιμο ταξινομητή για την αναγνώριση πινακίδων. Οι Donoser χρησιμοποιούν πολλά δυαδικά SVM δηλαδή ένα SVM για κάθε ζεύγος χαρακτήρων κάτι που όμως δίνει κακά αποτελέσματα αναγνώρισης χαρακτήρων. Όμως αν ληφθούν πολλές εικόνες της ίδιας πινακίδας από διαδοχικά καρέ ενός βίντεο, τα αποτελέσματα βελτιώνονται θεαματικά. Οι Kim κ.συν. [25] διαιρούν την εικόνα σε 64 block και χρησιμοποιούν επίσης πολλαπλά SVM για τα γράμματα και τους αριθμούς που θέλουν να ταξινομήσουν.

Μια άλλη προσέγγιση γίνεται από τους Duan κ.συν. [37], οι οποίοι χρησιμοποιούν κρυφά μαρκοβιανά μοντέλα. Συγκεκριμένα, η εικόνα σαρώνεται με παράθυρο 9x9 pixel, μέσα στο οποίο υπολογίζεται το ποσοστό των μαύρων pixel προς τα ολικά σε κάθε θέση. Έτσι προκύπτουν 196 τιμές για κάθε χαρακτήρα. Το μοντέλο εκπαιδεύεται με περίπου 60 εικόνες για κάθε χαρακτήρα.

Σε αυτό το σημείο αξίζει να αναφέρουμε και δύο πιο πρόσφατες προσπάθειες για βελτιωμένα συστήματα αναγνώρισης χαρακτήρων. Οι Lee κ.συν. [49] εξάγουν χαρακτηριστικά χαμηλού επιπέδου. Συγκεκριμένα το ιστόγραμμα των κατευθυντήριων παραγώγων σε 6 κατευθύνσεις, η ένταση των ακμών και το χρώμα λαμβάνονται για εικόνες διαφορετικών μεγεθών και παραμορφώσεων. Ένα SVM χρησιμοποιείται για την επιλογή των χαρακτηριστικών εκείνων, που δίνουν σε κάθε χαρακτήρα σημαντικές διαφοροποιήσεις από τους υπόλοιπους. Με αυτόν τον τρόπο διατηρούνται μόνο τα σημαντικά χαρακτηριστικά.

Τέλος, οι Cong κ.συν. [50] εξάγουν strokelets, δηλαδή χαρακτηριστικά μέσου επιπέδου για κάθε χαρακτήρα. Πρόκειται για περιοχές στις οποίες ο χαρακτήρας εμφανίζει μεγαλύτερες διαφοροποιήσεις από τους υπόλοιπους. Για την εξαγωγή τους χρησιμοποιούνται διαδοχικά σχήματα ομαδοποίησης και επανεκπαίδευσης. Πέρα από τα strokelets, για την αναγνώριση

χρησιμοποιείται και το ιστόγραμμα κατευθυντήριων παραγώγων (histogram of oriented gradients - HOG)

Κεφάλαιο 3

Μέθοδος ανίχνευσης της πινακίδας

Το πρώτο μέρος της αναγνώρισης πινακίδων κυκλοφορίας είναι η ανίχνευση των ορίων της εντός της εικόνας. Σε αυτό το κεφάλαιο θα γίνει περιγραφή της μεθόδου που χρησιμοποιήθηκε για το σκέλος αυτό, η ιδέα για την οποία στηρίχθηκε στην εργασία [6].

3.1 Οι μέγιστα σταθερές ακρότατες περιοχές (MSER)

Πριν ξεκινήσει η περιγραφή του αλγορίθμου, θα ήταν χρήσιμο να γίνει μια παρουσίαση της μεθόδου που επιλέχθηκε για την εύρεση περιοχών. Τα MSER (maximally stable extremal regions) παρουσιάστηκαν για πρώτη φορά από τους Matas, Chum, Urban, και συν. το 2002 [4]. Μια απόπειρα μετάφρασής τους είναι Μέγιστα Σταθερές Ακρότατες Περιοχές. Επινοήθηκαν στα πλαίσια του προβλήματος στερεοσκοπικής ανάλυσης, δηλαδή την ανάγκη να βρεθούν τα κοινά σημεία ενός αντικειμένου, σε εικόνες του από διαφορετικές οπτικές γωνίες. Είναι πολύ ενδιαφέρουσες περιοχές επειδή κατέχουν τις εξής ιδιότητες:

- Είναι αναλλοίωτες ως προς αφινικούς μετασχηματισμούς της φωτεινότητας
- Έχουν υψηλή συνδιακύμανση (covariance) με συνεχείς μετασχηματισμούς στο πεδίο της εικόνας
- Είναι σταθερές, για τους λόγους που θα αναλυθούν παρακάτω
- Πραγματοποιούν ανίχνευση σε πολλές κλίμακες μεγέθους ταυτόχρονα
- Μπορούν να απαριθμηθούν με πολυπλοκότητα $O(n \log \log n)$, όπου n ο αριθμός των pixel της εικόνας.

Η ερμηνεία του προσδιορισμού «ακρότατες» (extremal) έχει να κάνει με την ιδιότητα όλων των pixel εντός της περιοχής να έχουν τιμή φωτεινότητας μικρότερη ή μεγαλύτερη από την τιμή των pixel που αποτελούν το εξωτερικό όριό της. Ανάλογα με το τι από τα δύο ισχύει, γίνεται διάκριση των περιοχών σε θετικές (MSER+), αν το περιεχόμενο της περιοχής είναι πιο φωτεινό από το όριό της, και σε αρνητικές (MSER-) αν ισχύει το αντίθετο.

Για την εύρεση των περιοχών αυτών, ακολουθείται η εξής διαδικασία. Η γκρι εικόνα κατωφλιοποιείται με διαφορετικά κατώφλια, με αποτέλεσμα να δημιουργούνται διάφορες συνεκτικές περιοχές. Η κάθε περιοχή αριθμείται και παρακολουθείται η εξέλιξή της όσο μεταβάλλεται το κατώφλι. Αν ο ρυθμός μεταβολής του εμβαδού της παρουσιάσει τοπικό

ελάχιστο, τότε η περιοχή θεωρείται «μέγιστα σταθερή» (maximally stable), οπότε καταχωρείται ως MSER. Πιο συγκεκριμένα, αν $I(f)$ η τιμή φωτεινότητας του pixel f , και αν Q μια συνεκτική περιοχή της εικόνας και ∂Q το εξωτερικό σύνορό της, δηλαδή τα pixel που εφάπτονται στην περιοχή αλλά δεν ανήκουν σε αυτήν, τότε:

Η περιοχή ονομάζεται **ακρότατη** αν για όλα τα $p \in Q, q \in \partial Q : I(p) > I(q)$ (περιοχή μέγιστης φωτεινότητας) ή $I(p) > I(q)$ (περιοχή ελάχιστης φωτεινότητας).

Έστω μια ακολουθία «φωλιασμένων» ακρότατων περιοχών $Q_1, \dots, Q_{i-1}, Q_i, \dots$, όπου $Q_i \subset Q_{i+1}$. Η ακρότατη περιοχή Q_{i^*} ονομάζεται **μέγιστα σταθερή** αν και μόνο αν η $q(i) = \frac{|Q_{i+\Delta} \setminus Q_{i-\Delta}|}{|Q_i|}$ έχει τοπικό ελάχιστο στο σημείο i^* , όπου ο τελεστής $|\cdot|$ δείχνει το πλήθος των pixel κάθε περιοχής.

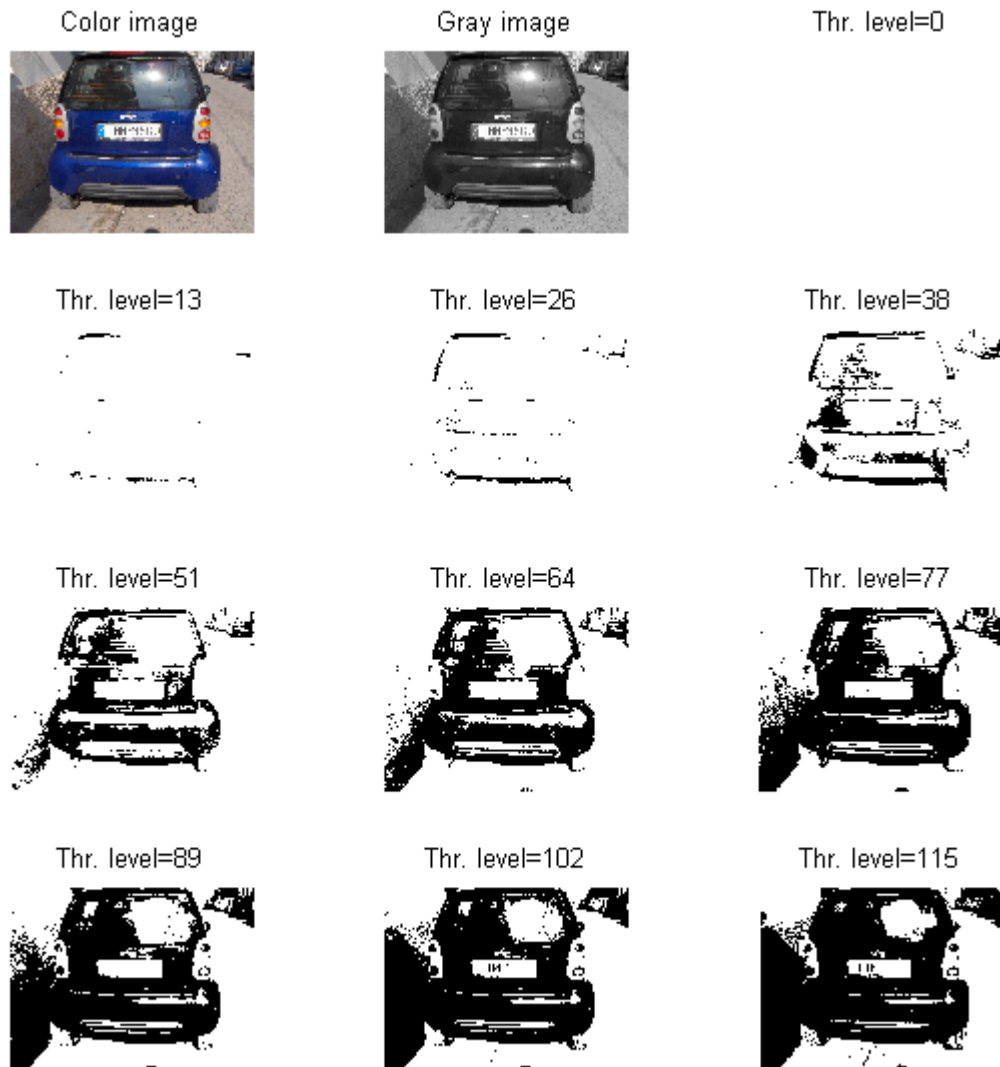
Η παράμετρος Δ που εμφανίζεται στην παραπάνω εξίσωση δείχνει τα επίπεδα φωτεινότητας πάνω και κάτω από το παρόν επίπεδο, για τα οποία γίνεται ο υπολογισμός της μεταβολής του μεγέθους της περιοχής. Η συνάρτηση $q(i)$, που ονομάζεται και *variation*, υπολογίζει το λόγο αυτής της μεταβολής προς το εμβαδόν της περιοχής στο παρόν επίπεδο. Κάποιες ακρότατες περιοχές μπορεί να θεωρούνται μέγιστα σταθερές για κάποιο συνδυασμό των δύο αυτών τιμών και για άλλο συνδυασμό όχι. Έτσι ο σωστός καθορισμός τους είναι σημαντικός για την επίτευξη του στόχου της βέλτιστης κατάτμησης.

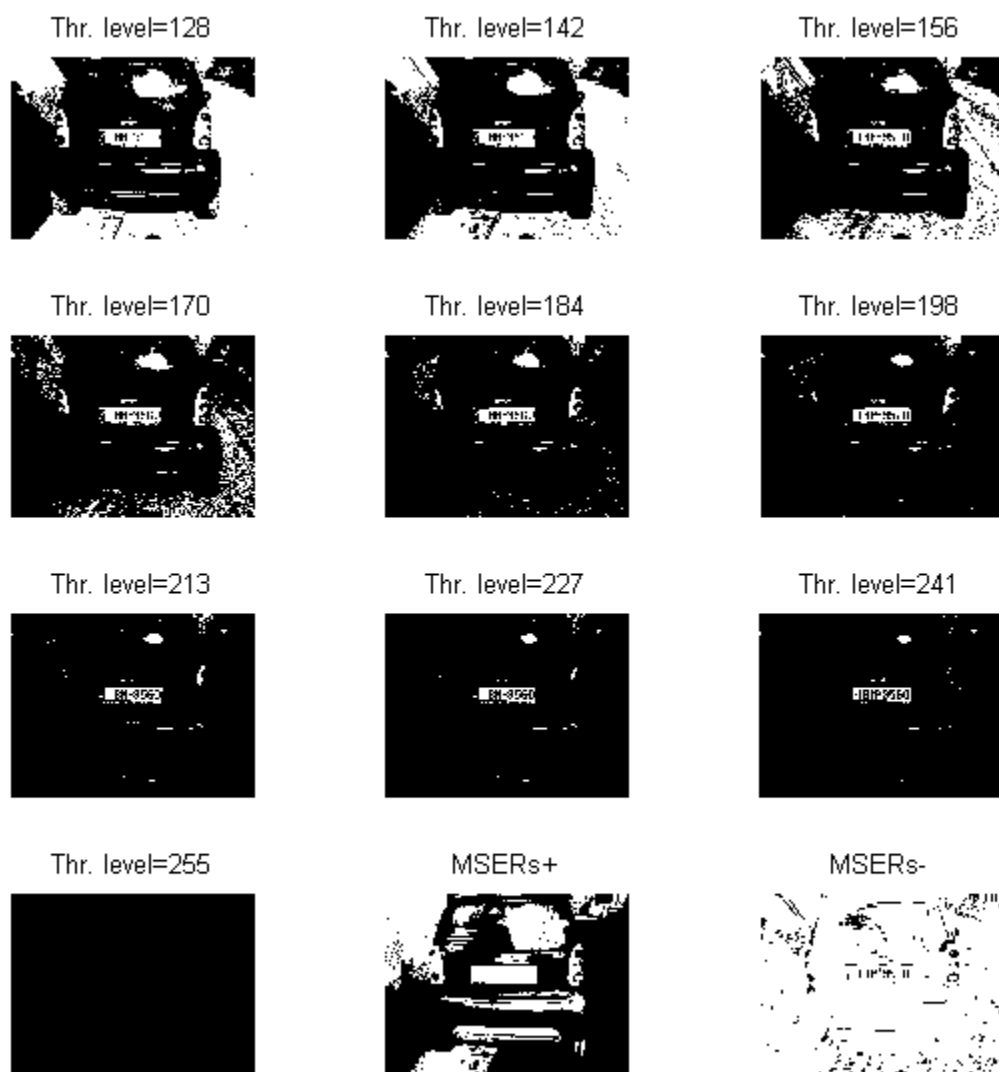
Όμως ακόμα και αν ισχύει η παραπάνω σχέση για μια περιοχή, αυτή μπορεί και πάλι να απορριφθεί, εάν το μέγεθός της είναι πολύ μικρό ή πολύ μεγάλο, ή αν έχει καταχωρηθεί ως MSER μια περιοχή στην ίδια θέση, που είναι λίγο μικρότερη ή λίγο μεγαλύτερη από αυτήν που μελετάται. Η παράμετρος αυτή ονομάζεται *diversity* και ορίζεται από τον τύπο $diversity = \frac{|Q_{i+k} \setminus Q_i|}{|Q_i|}$, όπου Q_{i+k} μια περιοχή που έχει επιλεγεί ως MSER και ισχύει

$Q_i \subset Q_{i+k}$. Πρέπει δηλαδή η διαφορά μεταξύ δύο περιοχών που προέρχονται από το ίδιο σύνολο φωλιασμένων περιοχών να έχουν διαφορά μεγέθους πάνω από ένα κατώφλι.

Στο παρακάτω σχήμα φαίνεται ένα παράδειγμα εύρεσης MSER για μια από τις εικόνες του συνόλου δεδομένων που χρησιμοποιήθηκε. Παρατηρείται ότι αρχικά όλη η εικόνα αποτελεί μια περιοχή από λευκά pixel, αφού όλα τα pixel της εικόνας έχουν φωτεινότητα πάνω από αυτό το κατώφλι. Όσο ανεβαίνει το κατώφλι αυτή διαχωρίζεται σε μικρότερες που στη συνέχεια μικραίνουν κι άλλο μέχρι να εξαφανιστούν. Αν κοιτάξει κανείς την ακολουθία ανάποδα, δηλαδή ξεκινήσει από το μέγιστο κατώφλι και εξετάσει τις περιοχές που αποτελούνται από μαύρα pixel, παρατηρείται το ίδιο φαινόμενο.

Είναι φανερό ότι κάποιες περιοχές μένουν σχεδόν αμετάβλητες καθώς αλλάζουν τα επίπεδα φωτεινότητας. Αυτές οι περιοχές αποτελούν μέγιστα σταθερές ακρότατες περιοχές. Ανάλογα με το αν πρόκειται για λευκές περιοχές ή μαύρες προκύπτουν αντίστοιχα θετικά ή αρνητικά MSER. Στα MSER+ συμπεριλαμβάνεται η περιοχή της πινακίδας και στα MSER- οι χαρακτήρες της. Οι παράμετροι που έχουν χρησιμοποιηθεί στο παράδειγμα είναι $\Delta = 3$ και $variation = 0,5$ για τα MSER+ και $\Delta = 2$ και $variation = 0,4$ για τα MSER-.

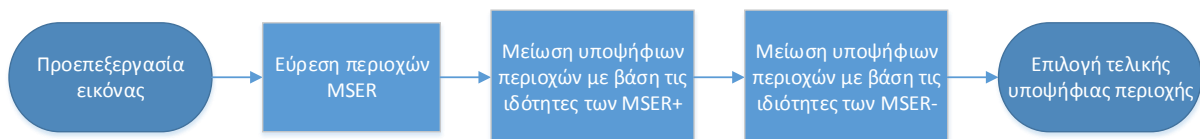




Εικόνα 3.1. Κατωφλίωση της εικόνας σε διάφορα επίπεδα φωτεινότητας. Στις δύο τελευταίες εικόνες φαίνονται τα MSER, δηλαδή οι περιοχές που μένουν σχεδόν αμετάβλητες για πολλά επίπεδα φωτεινότητας.

3.2 Περιγραφή του αλγορίθμου εύρεσης της πινακίδας

Η μέθοδος αυτή στηρίζεται αδρά στην εργασία των Donoser, Arth, Bischof [6], κυρίως ως προς την ιδέα της χρήσης ιδιοτήτων περιοχών και των MSER ως περιγραφέων τους. Όμως έγιναν πολλές προσαρμογές με βάση τα προβλήματα που εμφανίζονταν στην αποτίμησή της πάνω στη βάση δεδομένων ελληνικών πινακίδων κυκλοφορίας. Η γενική δομή του αλγορίθμου φαίνεται στο παρακάτω διάγραμμα. Αποτελείται από την προεπεξεργασία, όπου γίνεται προσπάθεια για βελτίωση της εικόνας πριν την εξαγωγή των περιοχών. Ύστερα γίνεται επιλογή της περιοχής της πινακίδας με βάση διάφορες παραμέτρους των υποψήφιων περιοχών, καθώς και των υποψήφιων χαρακτήρων. Στις επόμενες ενότητες θα γίνει ανάλυση των επιμέρους βημάτων.



3.2.1 Προεπεξεργασία εικόνας

Το πρώτο βήμα είναι η ανάγνωση της εικόνας και η μετατροπή της σε κλίμακα του γκρι. Μας ενδιαφέρουν μόνο οι αποχρώσεις του γκρι, αφού οι πινακίδες στις περισσότερες περιπτώσεις περιέχουν μόνο τα χρώματα λευκό και μαύρο, ενώ έτσι μειώνεται και το υπολογιστικό κόστος. Δεν εμφανίστηκε η ανάγκη για περαιτέρω επεξεργασία της εικόνας, αφού τα όρια της πινακίδας και των χαρακτήρων είναι συνήθως αρκετά ευδιάκριτα, άρα δεν είναι απαραίτητη η χρήση κάποιου υπερπαρατού φίλτρου. Επίσης, η κατανομή της φωτεινότητας στην περιοχή της πινακίδας δεν εμφανίζει τόσο σημαντικές διακυμάνσεις ώστε να χρειάζεται βαθυπερατό φίλτρο. Τέλος, ένας ακόμη λόγος είναι ότι οι τιμές φωτεινότητας που μας ενδιαφέρουν ανήκουν στα δύο άκρα του εύρους της, αφού οι χαρακτήρες έχουν σχεδόν μηδενική φωτεινότητα, και η περιοχή της πινακίδας έχει τιμή φωτεινότητας κοντά στη μέγιστη.



Εικόνα 3.2. Η αρχική εικόνα στην οποία θα γίνει αναζήτηση της πινακίδας.

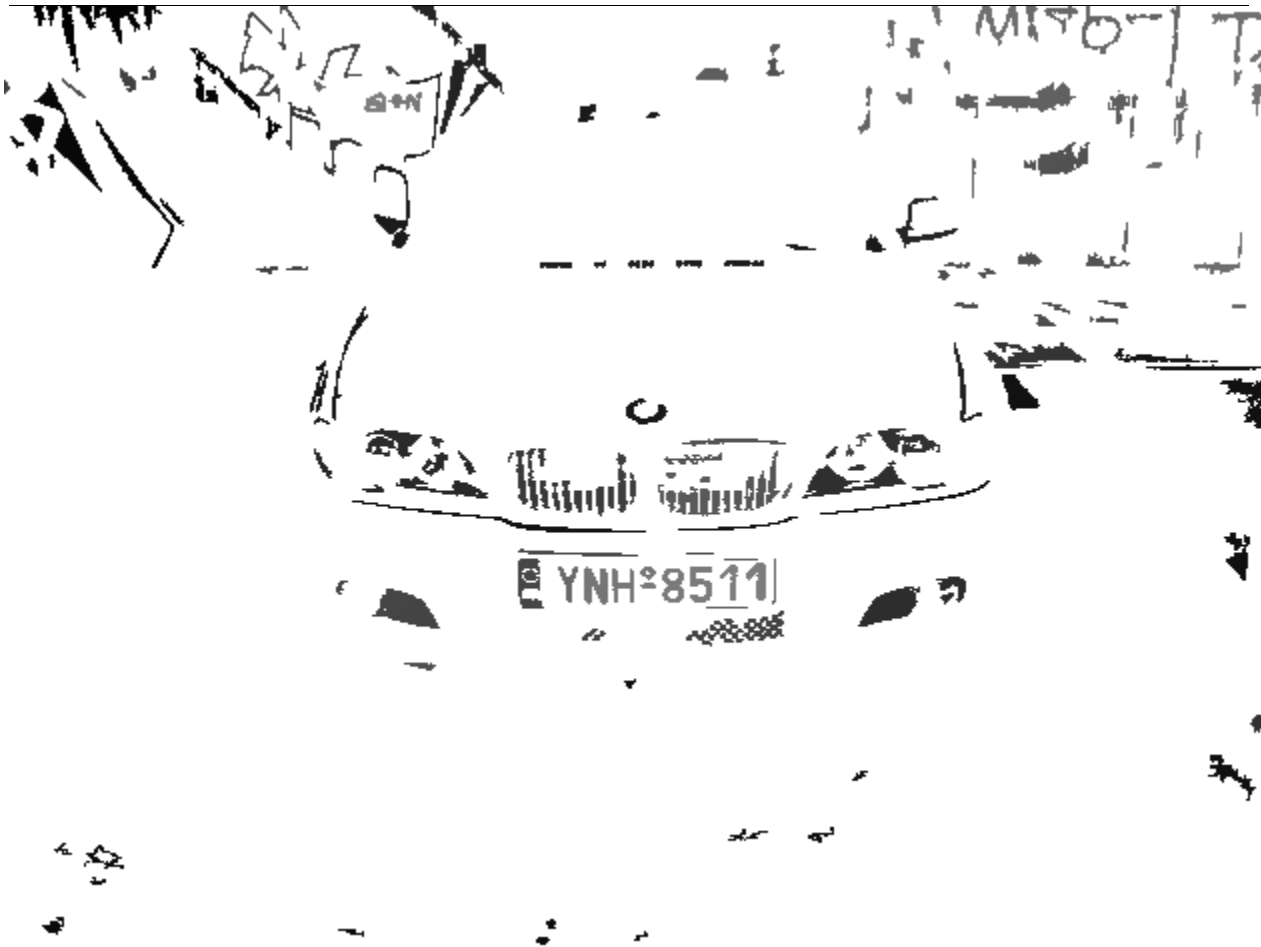
3.2.2 *Εύρεση περιοχών MSER*

Η επεξεργασία ξεκινά με την εύρεση των MSER, τόσο των θετικών (φωτεινές περιοχές με σκούρο περίγραμμα – MSER+) όσο και των αρνητικών (σκούρες περιοχές με φωτεινό περίγραμμα – MSER-). Σε αυτό το στάδιο τίθεται ο περιορισμός για το αποδεκτό μέγεθος των περιοχών. Έχοντας υπ’ όψη το ελάχιστο μέγεθος που πρέπει να έχει η πινακίδα ώστε να είναι ευανάγνωστη, αλλά και το μέγιστο μέγεθος των χαρακτήρων σε σχέση με αυτήν, επιλέγονται οι περιοχές με μέγεθος μέχρι το 1/8 του μεγέθους της εικόνας ως υποψήφιες περιοχές πινακίδας, και μέχρι το 1/100 ως περιοχές χαρακτήρων.



Εικόνα 3.3. Οι περιοχές MSER+ της εικόνας, με μόνο περιορισμό το μέγεθός τους. Κάθε περιοχή εμφανίζεται με διαφορετική τιμή φωτεινότητας

Στη συνέχεια εξάγονται χαρακτηριστικά των περιοχών, τα οποία θα χρησιμοποιηθούν στα επόμενα βήματα για την επιλογή της πιθανότερης ως περιοχής πινακίδας. Για την απλοποίηση των υπολογισμών, αφού οι περιοχές έχουν αυθαίρετο σχήμα, επιλέγεται σε πολλές περιπτώσεις αυτές να περιγράφονται με βάση το ελάχιστο περιγεγραμμένο ορθογώνιο γύρω από την κάθε περιοχή. Επιλέγεται ορθογώνιο και όχι έλλειψη, επειδή τόσο η πινακίδα, όσο και οι χαρακτήρες, προσομοιάζουν περισσότερο με ορθογώνιο ως προς το περίγραμμά τους. Αυτό μάλιστα είναι συνήθως περιστραμένο, αφού ελάχιστες περιοχές είναι παράλληλες με τον οριζόντιο ή τον κάθετο άξονα.



Εικόνα 3.4. Οι περιοχές MSER- της εικόνας, με μόνο περιορισμό το μέγεθός τους. Κάθε περιοχή εμφανίζεται με διαφορετική τιμή φωτεινότητας

Με βάση λοιπόν την ίδια την περιοχή, αλλά και το ελάχιστο ορθογώνιο, εξάγονται τα ακόλουθα χαρακτηριστικά:

- οι συντεταγμένες των pixel που αποτελούν την περιοχή
- το εμβαδόν της περιοχής, δηλαδή το πλήθος των pixel που την αποτελούν,
- το ύψος της περιοχής,
- οι συντεταγμένες των κορυφών του περιστραμμένου περιγεγραμμένου ορθογωνίου,
- η διεύθυνσή του, δηλαδή η γωνία περιστροφής του και
- ο λόγος πλάτους προς ύψος
- η λίστα που περιέχει τα διανύσματα με τις περιοχές MSER- που περιέχονται εντός της περιοχής.

3.2.3 Μείωση των υποψήφιων περιοχών με βάση τις ιδιότητες των MSER+

Τώρα που έχουν εξαχθεί τα κατάλληλα χαρακτηριστικά, μπορεί να γίνει αποτίμηση της κάθε περιοχής και να αποφασιστεί αν μπορεί να είναι υποψήφια περιοχή πινακίδας ή αν αυτή απορρίπτεται. Όπως έχει αναφερθεί, αναζητείται μια ανοιχτόχρωμη περιοχή (το φόντο της πινακίδας) που να περιέχει έναν ικανό αριθμό μικρότερων σκουρόχρωμων περιοχών (οι χαρακτήρες που περιέχει). Άρα ζητείται μια περιοχή MSER+ που περιέχει κάποιες περιοχές MSER-.

Όμως στις περισσότερες φωτογραφίες το φόντο είναι περίπλοκο, με αποτέλεσμα να υπάρχουν και άλλες περιοχές πέρα από την πινακίδα με βάση αυτό το χαρακτηριστικό, και έτσι τα κριτήρια που θα χρησιμοποιηθούν πρέπει να είναι περισσότερα. Επίσης λόγω του μεγάλου αριθμού περιοχών MSER+ που ευρίσκονται στο πρώτο βήμα, της τάξης συνήθως των 20-100 στο ελληνικό σύνολο δεδομένων, η απόρριψη των περιοχών που δεν αντιστοιχούν σε πινακίδες γίνεται σε τρία στάδια.

Στο πρώτο στάδιο εξετάζονται ιδιότητες μόνο των περιοχών MSER+, και αφού απορριφθούν όσες με βάση αυτές δεν μπορούν να αντιστοιχούν σε πινακίδα, τότε μόνο γίνεται εξέταση των περιοχών MSER- που περιέχουν. Γίνεται δηλαδή προσπάθεια οι πιο χρονοβόροι υπολογισμοί να γίνουν σε μικρότερο αριθμό περιοχών, και να απορριφθούν οι περισσότερες με βάση απλά χαρακτηριστικά τους. Όμως λόγω ανακριβειών στην εύρεση των MSER-, που θα αναλυθούν παρακάτω, η απόρριψη υποψήφιων περιοχών πινακίδων γίνεται επίσης σε δύο στάδια, για τη μείωση του υπολογιστικού κόστους, αφού οι περιοχές που φτάνουν μέχρι το τελευταίο βήμα είναι συνήθως μεταξύ μίας και δέκα.

Ο πρώτος γύρος λοιπόν απόρριψης μη υποψήφιων περιοχών γίνεται με βάση τα εξής κριτήρια, σχετικά με τις περιοχές MSER+:

- Η γωνία του περικλειόμενου ορθογωνίου με τον οριζόντιο άξονα να είναι μεταξύ -45° και 45° . Ο λόγος που δόθηκε τόσο μεγάλο περιθώριο δεν είναι ότι μπορεί μια πινακίδα να είναι τοποθετημένη τόσο στραβά σε σχέση με το έδαφος, αλλά ότι η λήψη της εικόνας δε γίνεται σχεδόν ποτέ ακριβώς μπροστά από το όχημα στα πραγματικά συστήματα, αφού αυτό βρίσκεται συνήθως σε κίνηση ή έχει πρόθεση να κινηθεί μπροστά και η κάμερα θα το εμπόδιζε. Όμως οι περιοχές που έχουν προσανατολισμό με γωνία μεγαλύτερη από αυτές που επιλέχθηκαν είναι σχεδόν βέβαιο ότι δεν αντιστοιχούν σε πινακίδα.
- Η σχέση πλάτους προς ύψος πρέπει να είναι μεταξύ 3 και 8. Όπως φαίνεται στο παρακάτω σχήμα, οι παλαιού τύπου ελληνικές πινακίδες κυκλοφορίας έχουν πλάτος 42,5 cm και ύψος 12 cm, ενώ οι νέου τύπου πλάτος 52 cm και ύψος 11 cm. Άρα οι πραγματικοί λόγοι πλάτους προς ύψος θα είναι 3,54 και 4,72 αντίστοιχα. Όμως λόγω παραμόρφωσης από τη γωνία λήψης, όπως και πριν, ο λόγος αυτός μπορεί να μεταβληθεί σημαντικά, ενώ και τα όρια των περιοχών μπορεί να μην έχουν βρεθεί με ακρίβεια, οπότε είναι απαραίτητο να δοθεί ένα τέτοιο περιθώριο.



Εικόνα 3.5. Διαστάσεις ελληνικών πινακίδων κυκλοφορίας μεταξύ 1983-2004 (ΦΕΚ 455B/1982)



Εικόνα 3.6. Διαστάσεις ελληνικών πινακίδων κυκλοφορίας μετά το 2004 (ΦΕΚ 192B/2004)

Στο πρώτο αυτό στάδιο, το οποίο έχει πολύ μικρό υπολογιστικό κόστος, απορρίπτεται το 40-93% των περιοχών, ανάλογα με την εικόνα, άρα ο φόρτος υπολογισμού για τα επόμενα βήματα περιορίζεται κατά δύο τρίτα με τρία τέταρτα περίπου. Η σημαντική μείωση των περιοχών μετά από αυτούς τους απλούς υπολογισμούς φαίνεται στην εικόνα 3.7.

3.2.4 Μείωση τα των περιοχών με βάση τις ιδιότητες των περιοχών MSER-

Το επόμενο βήμα στις περιοχές που απομένουν είναι η εύρεση των αρνητικών περιοχών που περιέχουν. Για το σκοπό αυτό μελετάται για κάθε περιοχή MSER- αν κάποιο σημείο της βρίσκεται εντός του περιγεγραμμένου ορθογωνίου μιας από τις περιοχές MSER+ που προέκυψαν από το προηγούμενο βήμα.

Θεωρήθηκε ότι αρκεί η εκτίμηση για ένα μόνο σημείο, αφού είναι πολύ σπάνιο να βρίσκεται μια περιοχή MSER- εν μέρει εντός και εν μέρει εκτός μιας περιοχής MSER+. Επίσης, η χρήση του περιγεγραμμένου ορθογωνίου αντί των πραγματικών ορίων της περιοχής προσθέτει κάποια ανακρίβεια, η οποία όμως είναι αμελητέα για τη συνέχεια. Στην εικόνα 3.8 φαίνονται οι περιεχόμενες περιοχές για την εικόνα του παραδείγματος.

Έχοντας λοιπόν τη λίστα με τα περιεχόμενα MSER-, εξετάζονται τα εξής κριτήρια για κάθε μία από τις υποψηφίες περιοχές πινακίδας:

- Το πλήθος των περιεχόμενων MSER- να είναι τουλάχιστον 4 και
- Το πλήθος των περιοχών των οποίων το ύψος να ισούται με τουλάχιστον το μισό του ύψος του περιγεγραμμένου ορθογωνίου να είναι τουλάχιστον 3.



Εικόνα 3.7. Οι περιοχές MSER+ που έμειναν προς διερεύνηση μετά την εφαρμογή περιορισμών ως προς τις ιδιότητές τους.

Όπως είναι γνωστό, οι πινακίδες περιέχουν 7 χαρακτήρες και τουλάχιστον 3 περιοχές ακόμα (παύλα ανάμεσα στα γράμματα και τους αριθμούς, σπές για τις βίδες, σφραγίδα κτλ). Όμως το όριο των περιεχόμενων περιοχών για να γίνει αποδεκτή η πινακίδα είναι τόσο μικρό, επειδή η εύρεση των περιοχών MSER- δεν είναι πολύ ακριβής για την αρχική εικόνα, λόγω του μικρού μεγέθους και της συχνά κακής ανάλυσης της περιοχής της πινακίδας. Και πάλι όμως επιτυγχάνεται περεταίρω μείωση των υποψήφιων περιοχών, σε μόλις 1-12 δηλαδή 3-15% των συνολικών περιοχών. Αν έχει ευρεθεί μία ή και καμία πινακίδα, η αναζήτηση σταματά εδώ, και είτε η πινακίδα που βρέθηκε είναι η τελική υποψήφια είτε η αναζήτηση κηρύσσεται άκαρπη, δηλαδή δε βρέθηκε καμία πινακίδα. Στην εικόνα 3.9 φαίνονται οι περιοχές που δεν απορρίφθηκαν σε αυτό το στάδιο.



Εικόνα 3.8. Περιοχές MSER- που περιέχονται στα MSER+ της εικόνας 3.7.

3.2.5 *Επιλογή της τελικής υποψήφιας περιοχής*

Συνήθως όμως απομένουν τουλάχιστον δύο περιοχές, οπότε ακολουθεί το δεύτερο στάδιο της αναζήτησης με βάση τις ιδιότητες των περιεχόμενων MSER-. Εδώ γίνεται πιο διεξοδική ανάλυσή τους. Ένα χαρακτηριστικό της διαδικασίας εύρεσης των περιοχών που παρατηρήθηκε είναι η πολλαπλή καταγραφή της ίδιας περιοχής ως διαφορετικές περιοχές. Ο βασικός λόγος για αυτό είναι η ύπαρξη διαφορετικών αποχρώσεων του γκρι εντός του χαρακτήρα, οι οποίες μπορούν να οδηγήσουν σε εύρεση μικρότερων μέγιστα σταθερών περιοχών εντός των ορίων του χαρακτήρα.



Εικόνα 3.9. Οι περιοχές MSER+ που έμειναν προς διερεύνηση μετά την εφαρμογή περιορισμών ως προς τις ιδιότητες των περιεχόμενων MSER-.

Αυτές δεν είναι χρήσιμες για τη συνέχεια, ενώ οδηγούν και σε λανθασμένη καταμέτρηση των περιοχών που περιέχονται σε κάθε υποψήφια περιοχή πινακίδας. Επίσης, κάποιες ιδιότητες, όπως το αν η περιοχή έχει όρθιο προσανατολισμό και σωστές αναλογίες ύψους προς πλάτος, μπορούν να εξαχθούν μόνο αν μελετήσουμε τα περιγεγραμμένα ορθογώνια της κάθε περιεχόμενης περιοχής. Έτσι, το πρώτο βήμα αυτού του σταδίου είναι η εύρεση αυτού του ορθογωνίου για κάθε περιοχή εντός της υποψήφιας περιοχής πινακίδας. Οι ιδιότητες που εξετάζονται είναι οι εξής:

- Το κέντρο του χαρακτήρα πρέπει να απέχει τουλάχιστον 3 pixel σε όλες τις κατευθύνσεις από τα κέντρα όλων των χαρακτήρων που έχουν γίνει δεκτοί μέχρι τώρα,
- Να έχει όρθιο προσανατολισμό, δηλαδή η γωνία του με τον οριζόντιο άξονα να είναι μεταξύ 45° και 135° ,
- Ο λόγος ύψους προς πλάτος να είναι μεταξύ 1 και 5. Έτσι εξαλείφονται περιοχές όπως κάγκελα στο φόντο ή κάθετες σχάρες εισόδου του καλοριφέρ του αυτοκινήτου,

- Το κέντρο του να μην απέχει πολύ από τον άξονα της πινακίδας, ο οποίος υπολογίζεται ως η μεσοπαράλληλος των δύο μεγάλων πλευρών του περιγεγραμμένου ορθογωνίου,
- Τέλος, εξετάζεται ξανά το ύψος της περιοχής, το οποίο τώρα πρέπει να είναι μεγαλύτερο από 0,6 επί το ύψος του ορθογωνίου της πινακίδας

Να σημειωθεί ότι σε αυτό το στάδιο δεν υπήρξαν ποσοτικοί περιορισμοί, δηλαδή δε γινόταν δεκτή μια υποψήφια πινακίδα αν είχε έναν συγκεκριμένο αριθμό από περιεχόμενα MSER-που να πληρούν τις παραπάνω προϋποθέσεις, αλλά έγινε σύγκριση μεταξύ των υποψήφιων περιοχών και επελέγη αυτή η οποία είχε τους περισσότερους τέτοιους χαρακτήρες. Στην περίπτωση που δύο περιοχές είχαν ίσο αριθμό τέτοιων περιοχών ή καμία περιοχή δεν περιείχε κατάλληλες περιοχές, τότε επελέγη αυθαίρετα η πρώτη.

Το τελευταίο στάδιο της διαδικασίας εύρεσης της περιοχής πινακίδας είναι η καταγραφή των συντεταγμένων των κορυφών του οριζόντιου περιγεγραμμένου ορθογωνίου της περιοχής που επελέγη ώστε να μπορέσει στη συνέχεια να γίνει οπτική αναγνώριση των χαρακτήρων εντός της, αλλά και σύγκριση με τα δεδομένα που εξάχθηκαν για κάθε εικόνα «με το χέρι», για αποτίμηση των αποτελεσμάτων του αλγορίθμου.



Εικόνα 3.10. Οι περιοχές MSER- που έμειναν μετά τη μείωσή τους σύμφωνα με τα χαρακτηριστικά τους.



Εικόνα 3.11. Το περίγραμμα της επιλεγμένης περιοχής πάνω στην αρχική εικόνα.

Κεφάλαιο 4

Η αναγνώριση των χαρακτήρων

Μετά την εύρεση της πινακίδας του οχήματος εντός της εικόνας, το δεύτερο τμήμα της παρούσης εργασίας είναι η ανάγνωση των χαρακτήρων που περιέχει. Πρόκειται δηλαδή για προσπάθεια επίλυσης του προβλήματος οπτικής αναγνώρισης χαρακτήρων (optical character recognition – OCR). Επιλέχθηκε η χρήση της μηχανής tesseract-OCR, η οποία είναι ελεύθερα διαθέσιμη, και η οποία θεωρείται από τις καλύτερες ελεύθερες μηχανές αναγνώρισης χαρακτήρων. Ένα άλλο της βασικό πλεονέκτημα είναι η συμβατότητά της για ενσωμάτωση σε πρόγραμμα γλώσσας C++, η οποία είναι η γλώσσα που χρησιμοποιήθηκε και για την υλοποίηση του πρώτου μέρους της εργασίας.

Στο παρόν κεφάλαιο θα γίνει αρχικά παρουσίαση του tesseract-OCR και στη συνέχεια θα παρουσιαστούν τα βήματα που ακολουθήθηκαν για την επεξεργασία της εικόνας με τους χαρακτήρες οι οποίοι πρέπει να αναγνωριστούν.

4.1 Tesseract OCR

Για την οπτική αναγνώριση των χαρακτήρων της πινακίδας επιλέχθηκε η μηχανή tesseract-OCR [51]. Αυτή αναπτύχθηκε αρχικά ως εμπορικό πρόγραμμα από την Hewlett Packard στο Bristol της Αγγλίας και στο Greeley του Κολοράντο από το 1985 ως το 1994 σε γλώσσα προγραμματισμού C. Στη συνέχεια έγιναν κάποιες αλλαγές ώστε να είναι συμβατή με λειτουργικό σύστημα Windows και ο κώδικάς της να μπορεί να μεταγλωττιστεί με compiler της C++, ενώ οι προσθήκες στον κώδικα γίνονταν πλέον σε γλώσσα C++. Η εξέλιξη σταμάτησε στα μέσα της δεκαετίας του 1990, και το 2005 δόθηκε ως πρόγραμμα ελεύθερου κώδικα από την Hewlett Packard και το πανεπιστήμιο της Νεβάδας (UNLV). Από το 2006 η εξέλιξή του χρηματοδοτείται από την Google. Σήμερα διατίθεται ως ελεύθερο λογισμικό με άδεια Apache 2.0 [52].

Είναι σχεδιασμένο να τρέχει από γραμμή εντολών, ενώ έχουν αναπτυχθεί από τρίτους προγράμματα για αλληλεπίδραση με το χρήστη (GUI). Επίσης υπάρχει αλληλεπίδραση και με την OpenCV, η οποία χρησιμοποιήθηκε στην παρούσα εργασία. Η μηχανή tesseract-OCR

θεωρείται μία από τις καλύτερες ελεύθερες στην οπτική αναγνώριση χαρακτήρων, ενώ μπορεί να αναγνωρίσει κείμενο σε περισσότερες από 35 γλώσσες. Ένας τρόπος για να το επιτύχει αυτό είναι το ενσωματωμένο λεξιλόγιο, που της επιτρέπει να αναγνωρίζει σωστά λέξεις, ακόμα και αν κάποιος από τους χαρακτήρες θα αναγνωριζόταν λανθασμένα ως μεμονωμένος. Στο παρόν πρόβλημα όμως, η συγκεκριμένη ιδιότητα αποτελεί μειονέκτημα, αφού οι χαρακτήρες που περιέχει η πινακίδα κυκλοφορίας δεν αντιστοιχούν σε κάποια λέξη, άρα πρέπει να γίνει απενεργοποίησή της. Θα ήταν σωστότερο λοιπόν να χαρακτηριστεί ως μηχανή αναγνώρισης κειμένου, αφού ο βασικός στόχος με τον οποίο δημιουργήθηκε ήταν η ψηφιοποίηση τυπωμένου και σκαναρισμένου κειμένου.

4.2 Προεπεξεργασία της εικόνας της πινακίδας

Όπως αναφέρεται στις συνήθεις ερωτήσεις στην ιστοσελίδα του tesseract [53], η αναγνώριση χαρακτήρων εμφανίζει καλά αποτελέσματα μόνο αν το ύψος των μικρών χαρακτήρων είναι τουλάχιστον 20 pixel και το κείμενο δεν είναι περιστραμμένο ή με πλάγιους χαρακτήρες. Επίσης, πρέπει να προηγηθεί απομάκρυνση των ανομοιομορφιών στη φωτεινότητα του υποβάθρου (θόρυβος χαμηλής συχνότητας), κάτι που μπορεί να επιτευχθεί με χρήση υπερερατού φίλτρου ή σωστή καταωφλίωση της εικόνας ώστε να προκύψει δυαδική, στην οποία οι χαρακτήρες θα έχουν τη μία τιμή φωτεινότητας και το φόντο την άλλη. Τέλος, πρέπει να γίνει από πριν απομάκρυνση των γραμμών και των πλαισίων γύρω από το κείμενο, αφού αυτά οδηγούν σε λανθασμένη αναγνώριση, καθώς και η όσο το δυνατόν εκμηδένιση κάθε άλλου θορύβου. Σε αυτό το κεφάλαιο παρουσιάζεται η προσπάθεια που έγινε ώστε να ικανοποιηθούν όλες αυτές οι συνθήκες.

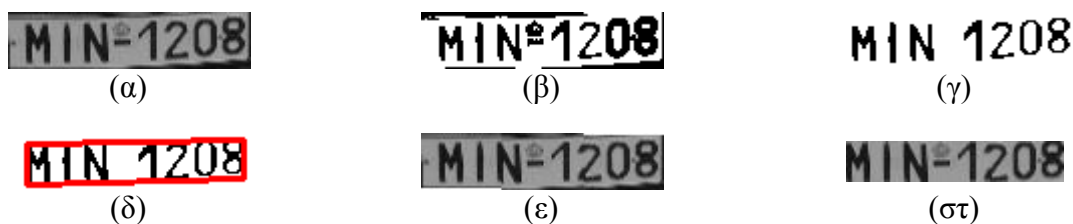
4.2.1 Περιστροφή της εικόνας αποκοπή του πλαισίου της πινακίδας

Όπως έχει αναφερθεί, η πινακίδα εμφανίζεται λίγο ως πολύ περιστραμμένη σε σχέση με τον οριζόντιο άξονα, λόγω της δυσκολίας να ληφθεί εικόνα του οχήματος από ακριβώς μπροστά, όταν αυτό βρίσκεται σε κίνηση. Όμως αυτή η περιστροφή μπορεί να οδηγήσει σε μικρότερη πιθανότητα αναγνώρισης των χαρακτήρων. Γι' αυτό πρέπει με κάποιον τρόπο, να βρεθεί η γωνία περιστροφής της πινακίδας, έχοντας ως μοναδικό δεδομένο την ίδια την εικόνα της. Μια λύση που έχει προταθεί για αυτό το πρόβλημα είναι η χρήση μετασχηματισμού Hough για την εύρεση των κυριότερων γραμμικών συνιστωσών της εικόνας. Όμως ο μετασχηματισμός αυτός είναι αρκετά χρονοβόρος.

Επειδή και στο κομμάτι της ανίχνευσης της πινακίδας έγινε ανάπτυξη διάφορων μεθόδων για την εξαγωγή χαρακτηριστικών από συνεκτικές περιοχές της εικόνας, έγινε και εδώ επιλογή να χρησιμοποιηθούν κυρίως αυτές οι μέθοδοι. Η διαδικασία είναι η εξής. Αν γίνει δυνατό να απομονωθούν οι χαρακτήρες και να βρεθεί το ελάχιστο περιγεγραμμένο ορθογώνιο που τους περικλείει, τότε είναι εύκολο να βρεθεί και η γωνία περιστροφής τους. Στη συνέχεια μπορεί

να περιστραφεί η αρχική εικόνα κατά την αντίθετη γωνία, και το κείμενο να καταλήξει να είναι οριζόντιο. Επίσης, με βάση το ίδιο ορθογώνιο μπορεί να γίνει αποκοπή των περιθωρίων, ώστε στην εικόνα να βρίσκονται μόνο οι χαρακτήρες χωρίς το πλαίσιο της πινακίδας.

Για την εύρεση των περιοχών επιλέχθηκαν και πάλι τα MSER, τα οποία δίνουν τις σταθερές περιοχές της εικόνας, όπως είναι και οι χαρακτήρες. Στη συνέχεια αφαιρούνται οι περιοχές που δεν πληρούν τα κριτήρια που παρουσιάζονται στην ενότητα 3.2.5, με αποτέλεσμα να μείνουν μόνο οι περιοχές που αντιστοιχούν σε χαρακτήρες. Έτσι μπορεί να βρεθεί το ελάχιστο ορθογώνιο που περιέχει αυτές τις περιοχές. Με αυτό γίνεται γνωστή η γωνία περιστροφής και τα όρια των χαρακτήρων. Ο αντίστροφος μετασχηματισμός, δηλαδή η περιστροφή με την αντίθετη γωνία, εφαρμόζεται στην αρχική εικόνα της πινακίδας. Επίσης, γίνεται αποκοπή της περιοχής πέρα από τους χαρακτήρες ώστε στο στάδιο της αναγνώρισης να μην υπάρχουν άλλες γραμμές εκτός από τους χαρακτήρες. Η πορεία της διαδικασίας αυτής παρουσιάζεται στις παρακάτω εικόνες για μία από τις πινακίδες.



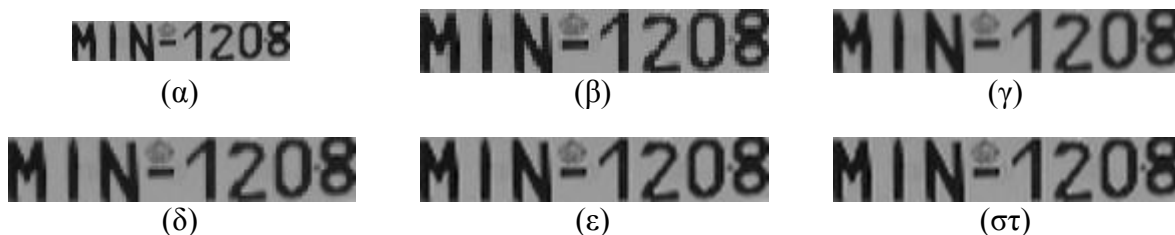
Εικόνα 4.1. Τα στάδια περιστροφής της εικόνας:

(α) Αρχική εικόνα, (β) τα MSER- της εικόνας, (γ) οι περιοχές μετά την αφαίρεση των μη αποδεκτών περιοχών, (δ) το ορθογώνιο που περικλείει τους χαρακτήρες, (ε) η αρχική εικόνα μετά από περιστροφή κατά $1,4^\circ$, (στ) αποκοπή των περιθωρίων της εικόνας

4.2.2 Προσαρμογή του μεγέθους της εικόνας

Τώρα που η εικόνα περιστράφηκε ώστε να είναι όσο το δυνατόν όρθιοι οι χαρακτήρες, και πριν γίνει η κατωφλίωση, πρέπει να γίνει προσαρμογή του μεγέθους τους. Εφόσον τα μικρά γράμματα πρέπει να έχουν ύψος τουλάχιστον 20 pixel, πρέπει να εξασφαλιστεί ότι τα κεφαλαία θα είναι μεγαλύτερα από αυτό το μέγεθος. Έτσι, γίνεται προσαρμογή του μεγέθους της πινακίδας, ώστε όλες οι πινακίδες να έχουν ύψος τουλάχιστον 32 pixel, μετά την περιστροφή και την αποκοπή του περιθωρίου τους. Το πλάτος τους διαμορφώνεται αναλόγως, ώστε μετά τη μεταβολή του μεγέθους να υπάρχει η ίδια αναλογία ύψους προς πλάτος.

Σε αυτό το στάδιο είναι κρίσιμο να επιλεγεί η σωστή μέθοδος παρεμβολής, για να αλλοιωθεί όσο γίνεται λιγότερο η αρχική εικόνα. Στις παρακάτω εικόνες φαίνονται τα αποτελέσματα της αλλαγής του μεγέθους με διάφορες μεθόδους παρεμβολής.



Εικόνα 4.2. Οι διάφοροι τρόποι παρεμβολής κατά την αλλαγή μεγέθους. (α) περιστραμμένη εικόνα χωρίς το πλαίσιο της πινακίδας, (β) παρεμβολή με τη μέθοδο του κοντινότερου γειτονικού pixel, (γ) γραμμική παρεμβολή, (δ) παρεμβολή με βάση το εμβαδόν των pixel, (ε) δικυβική παρεμβολή, (στ) παρεμβολή με τη μέθοδο Lanczos

4.2.3 Φιλτράρισμα της εικόνας

Για τη βελτίωση των αποτελεσμάτων αναγνώρισης με το tesseract, προτείνεται η χρήση φίλτρων εξομάλυνσης του θορύβου χαμηλής συχνότητας, που όμως δεν αλλοιώνουν την ένταση των ακμών, κάτι που πετυχαίνει το υπερερατό φίλτρο. Επίσης είναι χρήσιμο να εξαλειφθεί και ο θόρυβος που προέρχεται από μικρές λεπτομέρειες της εικόνας που δεν ανήκουν στους χαρακτήρες, όπως σημάδια από χρώμα πάνω στην πινακίδα, μικρές σκιάσεις και οι οπές για τις βίδες. Έτσι πέρα από το υπερερατό φίλτρο, δοκιμάστηκαν και φίλτρα εξομάλυνσης. Σε αυτά ανήκουν τα φίλτρα μέσου όρου, διαμέσου και Gauss. Επειδή η εικόνα που χρησιμοποιήθηκε στις προηγούμενες φωτογραφίες είναι αρκετά καθαρή, θα παρουσιαστούν και παραδείγματα άλλων εικόνων ώστε να φαίνεται πιο καθαρά η επίδραση του κάθε φίλτρου.



Εικόνα 4.3. i. αρχική εικόνα , ii. υπερερατό (bilateral) φίλτρο , iii. φίλτρο μέσου όρου (mean) μεγέθους 3 , iv. φίλτρο διαμέσου (median) 5, v. γκαουσιανό φίλτρο (Gaussian) με $\sigma=0,35$

4.2.4 Κατωφλίωση της εικόνας

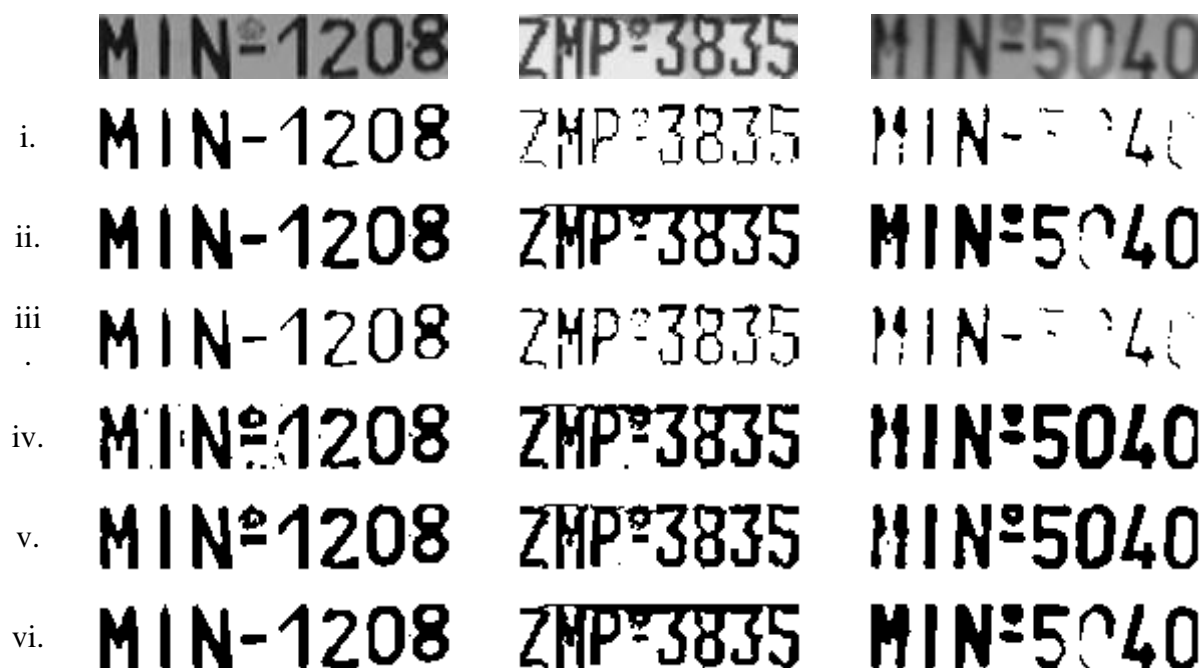
Το τελευταίο βήμα της επεξεργασίας της εικόνας πριν την προσπάθεια ανάγνωσής της είναι η μετατροπή της σε δυαδική. Στη μηχανή OCR είναι ενσωματωμένο το βήμα της κατωφλίωσης, άρα δεν είναι απαραίτητη προϋπόθεση αυτή να γίνει από πριν. Όμως δεν είναι γνωστή η μέθοδος που χρησιμοποιείται, άρα ο μόνος τρόπος να επιλεγεί η σωστή μέθοδος για καλύτερα αποτελέσματα είναι να δοθεί η εικόνα ως δυαδική με την χρήση του κατάλληλου κατωφλίου.

Πρώτα δοκιμάστηκε το οικουμενικό κατώφλι, δηλαδή μια σταθερή τιμή κατωφλίωσης για όλες τις εικόνες. Μετά υπολογίστηκε η μέγιστη και η ελάχιστη τιμή φωτεινότητας κάθε εικόνας και το κατώφλι τέθηκε στο 50% και στο 25% της διαφοράς τους. Ύστερα δοκιμάστηκαν προσαρμοστικές μέθοδοι, που επιλέγουν διαφορετικό κατώφλι σε διάφορες περιοχές της εικόνας: το προσαρμοστικό κατώφλι μέσης τιμής, που χρησιμοποιεί ως τιμή κατωφλίωσης τη μέση τιμή σε μια περιοχή της εικόνας, και το προσαρμοστικό φίλτρο με συντελεστές Gauss, που πριν την εξαγωγή της μέσης τιμής σε μια περιοχή, πολλαπλασιάζει τη φωτεινότητα με κατανομή Gauss γύρω από κάθε σημείο. Τέλος, δοκιμάστηκε η μέθοδος Otsu [33], η οποία αναφέρεται από πολλούς συγγραφείς στη βιβλιογραφία, η οποία υπολογίζει το ιστόγραμμα φωτεινότητας της εικόνας και επιλέγει ως τιμή κατωφλίωσης την τιμή ακριβώς ανάμεσα στις δύο κύριες κορυφές του. Στην εικόνα 4.4 παρουσιάζεται και πάλι η επίδραση της κάθε μεθόδου στη βασική εικόνα των προηγούμενων παραδειγμάτων, καθώς και σε δύο δυσκολότερες περιπτώσεις. Σε όλες τις περιπτώσεις δεν χρησιμοποιήθηκε φίλτρο πριν την κατωφλίωση.

Στη συνέχεια πραγματοποιείται προσθήκη λευκού πλαισίου γύρω από την κατωφλιωμένη εικόνα, για να αποδίδεται πιο σωστά το σχήμα του πρώτου και του τελευταίου χαρακτήρα. Τέλος, πραγματοποιείται αφαίρεση των περιοχών μικρού εμβαδού που οφείλονται σε θόρυβο και εφαρμόζονται μορφολογικά φίλτρα με σκοπό τη βελτίωση της εικόνας που οδηγείται προς αναγνώριση από το tesseract.

4.3 Προσαρμογή των παραμέτρων του tesseract-OCR

Η μηχανή tesseract δίνει στο χρήστη τη δυνατότητα να προσαρμόσει κάποιες παραμέτρους, ώστε να βελτιώσει τα αποτελέσματα αναγνώρισης για το συγκεκριμένο πρόβλημα. Κατά τη διαδικασία αρχικοποίησης της εφαρμογής, γίνεται φόρτωση κάποιας γλώσσας και του αντίστοιχου λεξικού, δηλαδή των λέξεων που περιέχει. Όμως, εδώ δεν είναι στόχος η αναγνώριση λέξεων, αλλά μεμονωμένων χαρακτήρων, άρα είναι προτιμότερο να απενεργοποιηθεί η φόρτωση του λεξικού. Ακόμη, μπορεί να προσδιοριστεί αν στην προς ανάγνωση εικόνα υπάρχει μια παράγραφος κειμένου, μια γραμμή κειμένου, μια λέξη ή ένας μεμονωμένος χαρακτήρας.



Εικόνα 4.4. i. οικουμενικό κατώφλι , ii. κατώφλι στο 50% του (maxValue-minValue) κάθε εικόνας , iii. κατώφλι στο 25% του (maxValue-minValue) κάθε εικόνας, iv. προσαρμοστικό κατώφλι μέσης τιμής , v. προσαρμοστικό κατώφλι με συντελεστές Gauss, vi. μέθοδος Otsu

MIN 1208

(α)

MIN 1208

(β)

Εικόνα 4.5. (α) Η εικόνα που προκύπτει μετά την αφαίρεση των μικρών σε ύψος περιοχών, έχοντας επιλέξει γκαουσιανό φίλτρο. (β) Η τελική εικόνα μετά τη χρήση μορφολογικού φίλτρου opening.

Επίσης, δίνεται η δυνατότητα να προσδιοριστεί ένα υποσύνολο των χαρακτήρων, οι οποίοι θα είναι οι μόνοι που θα γίνονται δεκτοί ως αποτέλεσμα αναγνώρισης, η ονομαζόμενη «λευκή λίστα» (whitelist). Στην περίπτωση της ελληνικής βάσης δεδομένων, η συγκεκριμένη δυνατότητα είναι σημαντική, αφού στις ελληνικές πινακίδες χρησιμοποιείται ένα υποσύνολο μόνο της αλφαβήτου, και συγκεκριμένα μόνο οι χαρακτήρες που είναι κοινοί στο ελληνικό και στο αγγλικό αλφάβητο (A,B,E,Z,H,I,K,M,N,O,P,T,Y,X). Επιπλέον, μπορεί να γίνει διαχωρισμός, ώστε στην αρχή να αναγνωρίζει μόνο γράμματα και μετά μόνο αριθμούς, κάτι που επίσης μπορεί να βελτιώσει την αναγνώριση. Τα αποτελέσματα των παραπάνω προσαρμογών παρουσιάζονται στο επόμενο κεφάλαιο.

Κεφάλαιο 5

Πειράματα

5.1 *Η πειραματική διαδικασία*

Ο αλγόριθμος που περιγράφηκε παραπάνω υλοποιήθηκε σε γλώσσα προγραμματισμού C++, με χρήση της βιβλιοθήκης OpenCV. Πρόκειται για ένα σύνολο μεταβλητών και συναρτήσεων που έχουν σκοπό να διευκολύνουν την υλοποίηση εφαρμογών επεξεργασίας εικόνας και όρασης υπολογιστών. Η δημιουργία της ξεκίνησε το 1999 στην Intel Ρωσίας, με σκοπό να κάνει ελεύθερα διαθέσιμο ένα σημαντικό φορητό εργαλείο, με κορυφαία απόδοση, για τις αναρίθμητες εφαρμογές της επεξεργασίας εικόνας.

Το λογισμικό που χρησιμοποιήθηκε είναι το Visual Studio 2010 της Microsoft, σε λειτουργικό σύστημα Microsoft Windows 7 – 64 bit, σε υπολογιστή Intel Core i5 στα 2.3 GHz, με μνήμη RAM 6GB. Όλες οι μετρήσεις που αναφέρονται παρακάτω έγιναν σε αυτό το σύστημα, οπότε μπορεί να γίνει αποτίμηση και του χρόνου εκτέλεσης για τις διάφορες παραμέτρους που δοκιμάστηκαν.

Στη συνέχεια του κεφαλαίου θα γίνει περιγραφή των data set που χρησιμοποιήθηκαν, καθώς και οι ενέργειες που χρειάστηκαν για να γίνει σωστή χρήση τους στην αποτίμηση του αλγορίθμου. Στη συνέχεια θα παρουσιαστούν οι μετρήσεις που έγιναν με διάφορες παραμέτρους, αρχικά στο κομμάτι του εντοπισμού της πινακίδας, και στη συνέχεια σε αυτό της ανάγνωσης του περιεχομένου της. Τέλος, θα γίνει σύγκριση των αποτελεσμάτων με τα αποτελέσματα των μεθόδων που περιγράφονται στο κεφάλαιο 2.

5.2 *Σύνολα δεδομένων*

Μια βασική δυσκολία που αντιμετωπίζουν όλες οι έρευνες για το πρόβλημα της αναγνώρισης πινακίδων κυκλοφορίας είναι η απουσία ενός διεθνούς συνόλου δεδομένων (data set), με το οποίο να μπορούν όλοι να αξιολογήσουν τις λύσεις που βρίσκουν και να τις συγκρίνουν με άλλες, όπως συμβαίνει με άλλα προβλήματα της όρασης υπολογιστών. Βασικές αιτίες αποτελούν η δυσκολία ελεύθερης διάθεσης ενός συνόλου πραγματικών εικόνων αυτοκινήτων

με τις πινακίδες κυκλοφορίας τους, λόγω θεμάτων προστασίας προσωπικών δεδομένων, καθώς και η μη ομοιομορφία των πινακίδων κυκλοφορίας στις διάφορες χώρες, τόσο ως προς το αλφάβητο που χρησιμοποιούν, όσο και ως προς την ίδια τη μορφή της πινακίδας, δηλαδή οι διαστάσεις της, τη γραμματοσειρά, το χρώμα κτλ.

Για αυτόν το λόγο, στην παρούσα εργασία χρησιμοποιούνται δύο σύνολα εικόνων, τα οποία διατίθενται ελεύθερα στο διαδίκτυο, ένα ελληνικό [7] και ένα κροατικό [8]. Και στις δύο περιπτώσεις ζητήθηκε από τους συγγραφείς το αρχείο που περιέχει τις θέσεις των πινακίδων μέσα στην εικόνα και το περιεχόμενό τους, αλλά και στις δύο περιπτώσεις αυτό δεν ήταν διαθέσιμο. Έτσι η καταγραφή αυτή έγινε με το χέρι από εμένα και άλλους που αναφέρονται στις ευχαριστίες, ώστε να μπορεί να γίνει ανάγνωση των στοιχείων από το πρόγραμμα για την αποτίμηση των αποτελεσμάτων.

Οι μετρήσεις για την επιλογή των καλύτερων παραμέτρων έγινε στο ελληνικό σύνολο δεδομένων, λόγω της αναγκαιότητας να δουλεύει για ελληνικές πινακίδες. Όμως και το ίδιο το σύνολο παρουσιάζει μεγαλύτερη ποικιλία ως προς την ποιότητα των εικόνων που περιέχει, καθώς έχει γίνει προσπάθεια να υπάρχουν εικόνες με διαφορετικές συνθήκες φωτισμού, σκίασης και καθαρότητας της πινακίδας. Επίσης υπάρχουν θολές εικόνες και εικόνες με πολύπλοκο φόντο, και έτσι ο ταξινομητής μπορεί να προσαρμοστεί καλύτερα ως προς τις παραμέτρους τους για δύσκολες περιπτώσεις εικόνων.

5.3 *Αποτελέσματα Ανίχνευσης της Πινακίδας*

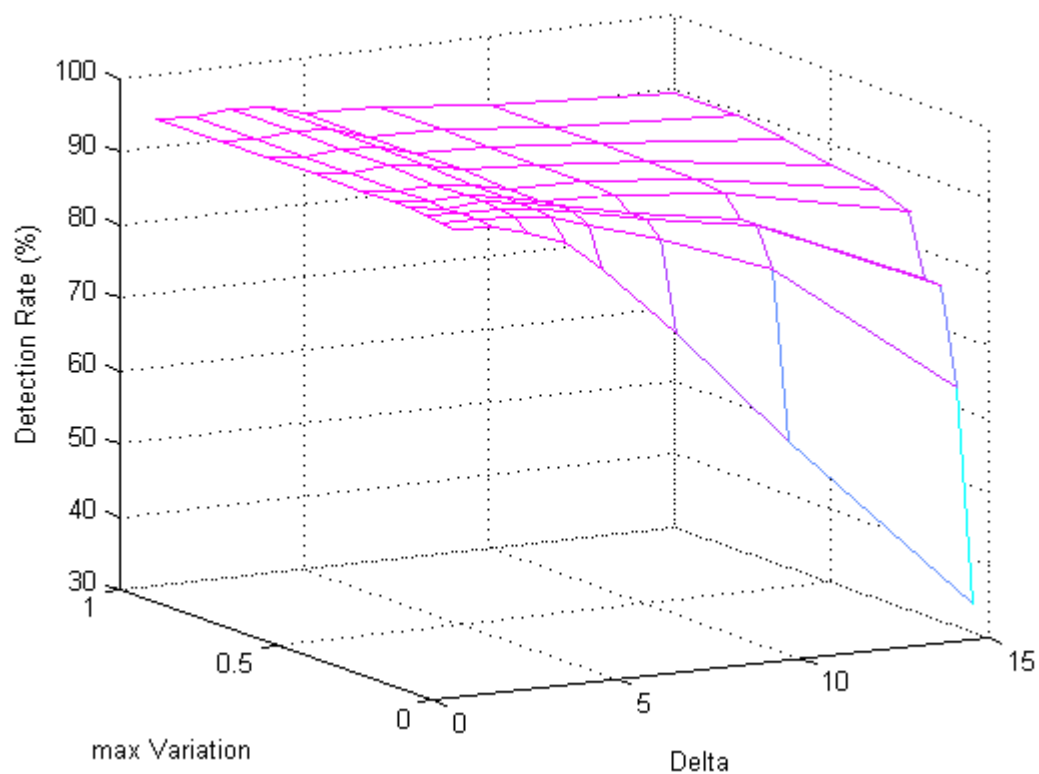
Κατά τη διαδικασία ανίχνευσης της πινακίδας σε κάθε εικόνα, ένα από τα σημαντικότερα βήματα είναι η εύρεση των MSER. Όπως αναλύεται στην ενότητα 3.1, για την εύρεσή τους χρησιμοποιούνται κάποιες παράμετροι, οι οποίες καθορίζουν πόσο σταθερά πρέπει να είναι τα όρια της περιοχής, πόση διαφορά πρέπει να έχει μια περιοχή από αυτήν στην οποία περιέχεται και ποιο πρέπει να είναι το ελάχιστο και το μέγιστο μέγεθός της.

5.3.1 *Μεταβολή των παραμέτρων των MSER*

Όπως είναι γνωστό, ο αλγόριθμος εύρεσης των MSER υπολογίζει το εμβαδόν των συνεκτικών περιοχών της εικόνας που προκύπτουν από κατωφλίωσή της. Το πόσο σταθερή είναι μια περιοχή καθορίζεται από τις παραμέτρους Delta και Maximum Variation. Η πρώτη δείχνει τον αριθμό των επιπέδων φωτεινότητας ανάμεσα στα οποία γίνεται η σύγκριση των εμβαδών, και η δεύτερη το πόσο μεγαλύτερο μπορεί να είναι το ένα από το άλλο ώστε να θεωρηθεί η περιοχή αρκετά σταθερή. Επειδή αυτές οι παράμετροι είναι αλληλοεξαρτώμενες, έγιναν μετρήσεις με διάφορους συνδυασμούς τους, οι οποίοι φαίνονται στους πίνακες I και III.

Delta mVar	1	2	3	4	5	7	10	15
0.05	93.23	93.06	91.45	89.52	85.48	75.65	59.03	33.87
0.1	93.39	93.39	93.06	92.42	90.81	87.58	81.77	62.74
0.15	93.55	93.39	93.22	92.74	91.77	89.68	87.1	75.97
0.2	94.03	93.23	93.55	92.74	91.94	89.68	87.1	75.97
0.25	94.03	93.39	93.71	92.9	91.77	91.45	90	84.52
0.35	94.03	93.39	93.71	93.06	91.94	91.77	90.48	86.29
0.5	94.03	93.55	94.03	93.39	92.26	91.77	90.65	87.26
0.65	94.03	93.55	94.03	93.71	92.26	91.94	90.81	88.55
0.8	94.03	93.55	94.03	93.71	92.26	92.1	90.81	89.35
1	94.03	93.55	94.03	93.71	92.42	92.1	90.65	89.35
1.5	94.03	93.55	94.03	93.87	92.26	92.1	91.13	89.19
3	94.03	93.55	94.03	93.87	92.74	92.26	90.97	89.68
10	94.03	93.55	94.03	93.87	92.74	92.42	91.29	89.52

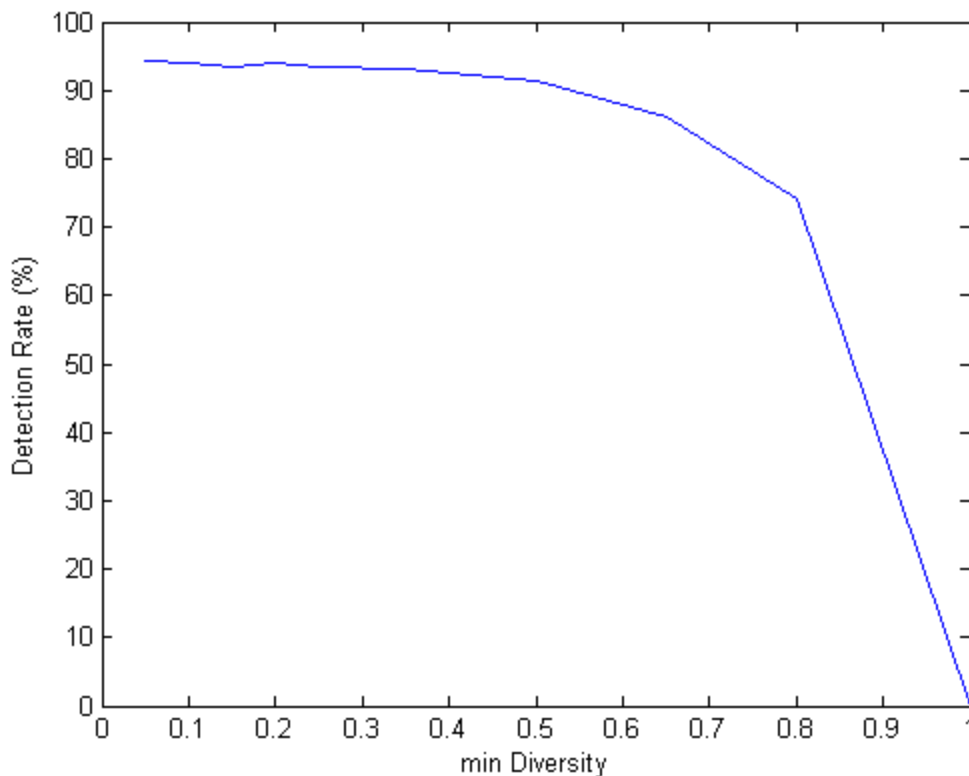
Πίνακας Ι. Ποσοστό ανίχνευσης (%) για τιμές των Delta και Max Variation για τα MSER+, για ποσοστό επικάλυψης 50%



Σχήμα Α. Τριδιάστατη αναπαράσταση του Πίνακα Ι (MSER+)

Επίσης, υπάρχει η παράμετρος Minimum Diversity, η οποία δείχνει πόση διαφορά πρέπει να έχει μια σταθερή περιοχή από μια άλλη στην οποία περιέχεται. Αυτή η παράμετρος είναι απαραίτητη, γιατί μπορεί κάποια περιοχή να μην έχει ομαλή κλίση της φωτεινότητας στις ακμές της. Δηλαδή μπορεί η κλίση να παρουσιάζει περισσότερα τοπικά ελάχιστα σε μικρή απόσταση μεταξύ τους, και κάθε ένα από αυτά να δίνει μια αρκετά σταθερή περιοχή. Τότε θα γίνουν πολλές καταχωρήσεις της ίδιας σχεδόν περιοχής ως MSER, κάτι που αυξάνει το κόστος υπολογισμού χωρίς σημαντικό όφελος. Και για αυτήν την παράμετρο έγιναν δοκιμές με διάφορες τιμές, των οποίων τα αποτελέσματα φαίνονται στους πίνακες II και IV.

min Diversity	Detection Rate (%)
0.05	94.19
0.1	94.03
0.15	93.55
0.2	94.03
0.25	93.55
0.35	93.23
0.5	91.45
0.65	86.13
0.8	74.03
1	0
1.5	0

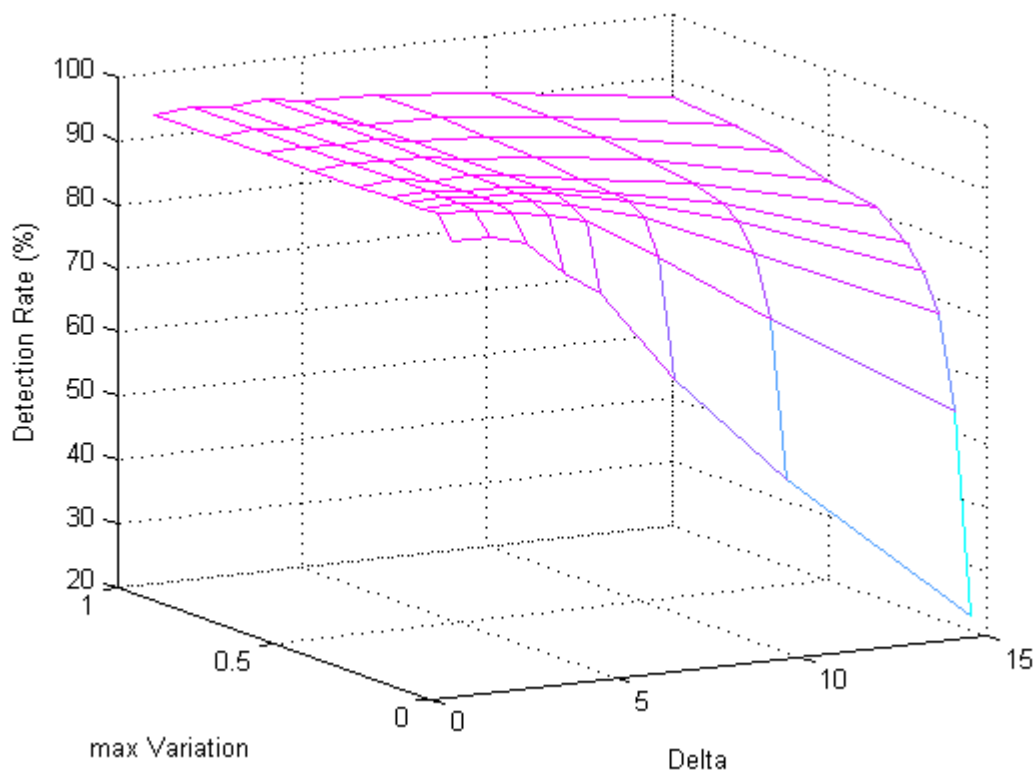


Πίνακας II. – Σχήμα Β. Ποσοστό ανίχνευσης (%) για τιμές του Min Diversity για τα MSER+ και η αντίστοιχη γραφική παράσταση

Τέλος, οι παράμετροι Minimum Area και Maximum Area καθορίζουν το ελάχιστο και το μέγιστο, αντίστοιχα, εμβαδόν που πρέπει να έχει μια περιοχή για να καταχωρηθεί ως MSER. Για τις παραμέτρους αυτές χρησιμοποιήθηκαν συγκεκριμένες τιμές, δίνοντας μεγάλα περιθώρια, αφού το μέγεθος της πινακίδας και των χαρακτήρων μέσα στην εικόνα δεν ήταν σταθερό, αφού οι φωτογραφίες ήταν τραβηγμένες από διαφορετικές αποστάσεις. Όμως και το μέγεθος των εικόνων σε pixel δεν ήταν σταθερό, γι' αυτό πριν την εύρεση των MSER προηγήθηκε αναπροσαρμογή του μεγέθους τους, ώστε όλες οι εικόνες να έχουν το ίδιο.

Delta mVar	1	2	3	4	5	7	10	15
0.05	90.16	90.16	88.39	83.23	79.35	64.52	46.94	22.26
0.1	93.71	93.23	92.42	91.29	89.68	82.9	70.97	53.39
0.15	93.39	93.55	93.06	92.74	91.45	88.06	80.64	67.9
0.2	93.55	93.71	93.06	93.06	92.26	89.68	84.84	73.71
0.25	93.39	93.71	93.23	93.23	92.58	90.32	86.61	77.26
0.35	93.39	93.87	93.06	93.23	92.58	90.81	88.23	81.13
0.5	93.39	93.87	93.39	93.55	92.42	91.61	89.36	82.58
0.65	93.39	93.87	93.39	93.71	92.58	92.1	90	84.68
0.8	93.39	93.87	93.39	93.87	92.74	92.42	90.48	85.97
1	93.39	94.03	93.23	93.87	92.9	92.42	90.81	86.94
1.5	93.39	94.03	93.23	93.87	93.23	92.1	91.13	88.71
3	93.39	94.03	93.06	94.03	93.39	92.58	92.58	90.48
10	93.39	94.03	93.06	94.03	93.71	92.74	92.58	91.61

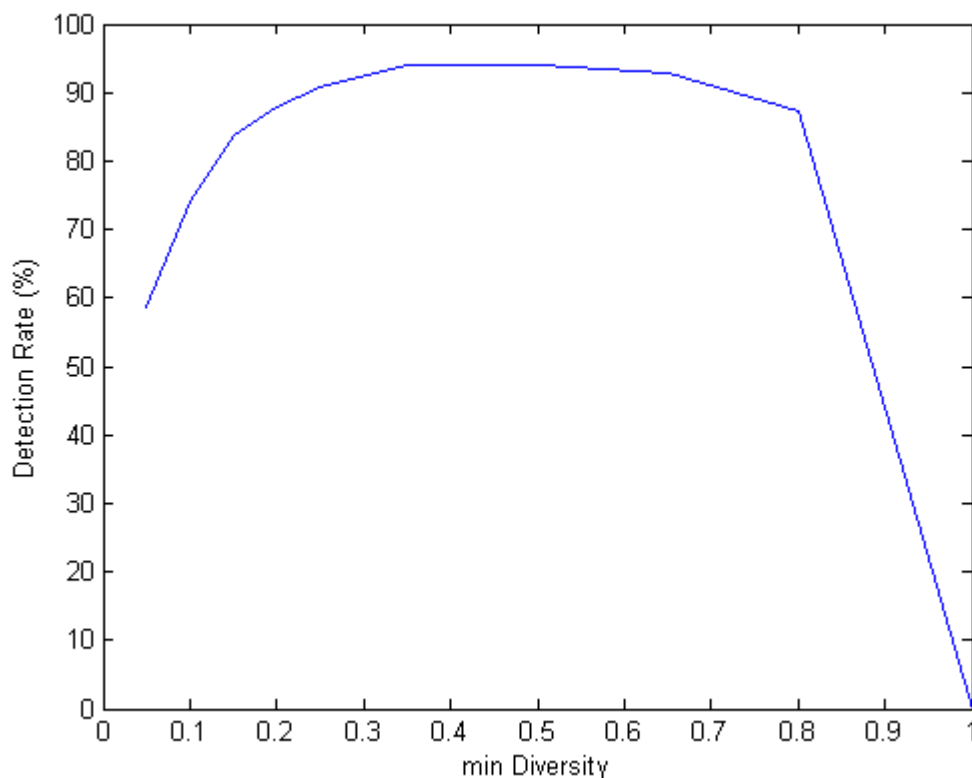
Πίνακας III. Ποσοστό ανίχνευσης (%) για τιμές των Delta και Max Variation για τα MSER-, για ποσοστό επικάλυψης 50%



Σχήμα Γ. Τριδιάστατη αναπαράσταση του Πίνακα III (MSER-)

Τα αποτελέσματα που φαίνονται σε αυτούς τους πίνακες αφορούν όλα το ελληνικό σύνολο δεδομένων, το οποίο αποτελείται από 620 εικόνες διαφορετικών μεγεθών και ποιότητας. Οι τριδιάστατες απεικονίσεις έγιναν με χρήση του προγράμματος MatLab R2010a. Όλες οι εικόνες κανονικοποιούνται κατά την επεξεργασία σε μέγεθος με 480 ή 500 γραμμές. Για αυτό το μέγεθος εικόνας, ο χρόνος εκτέλεσης είναι περίπου 87,6 δευτερόλεπτα για το σύνολο των εικόνων, δηλαδή περίπου 140 msec/εικόνα. Επίσης, σε όλους τους πίνακες, το αποτέλεσμα αφορά επικάλυψη κατά 50% του πραγματικού ορθογωνίου της πινακίδας με το ευρισκόμενο.

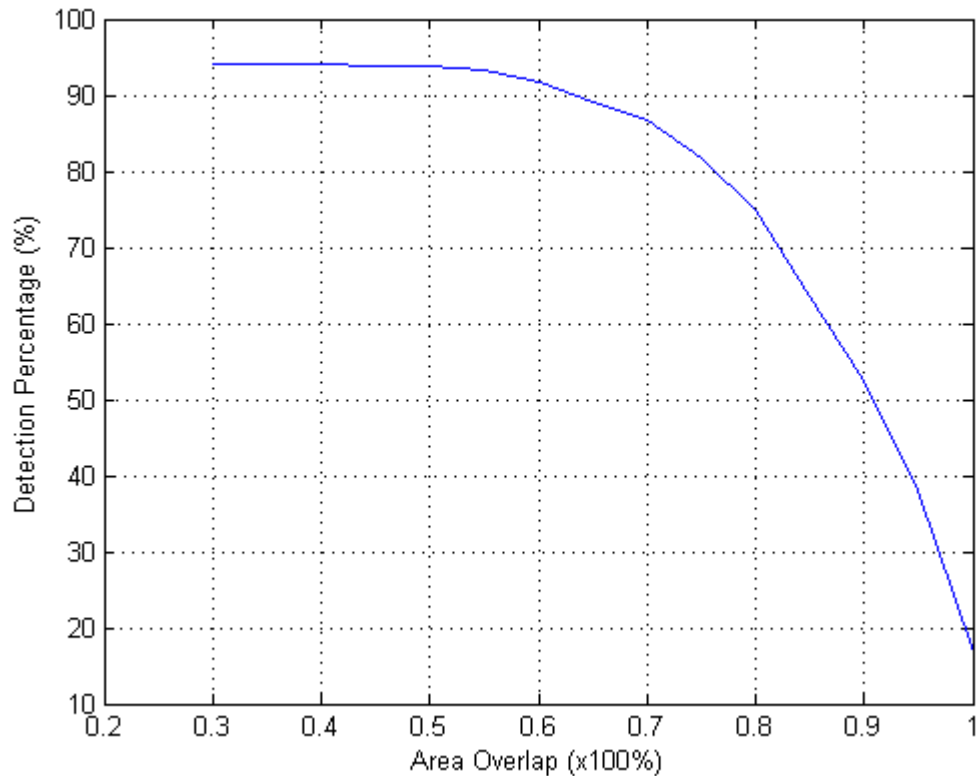
min Diversity	Detection Rate (%)
0.05	58.55
0.1	74.03
0.15	83.87
0.2	87.9
0.25	90.81
0.35	93.87
0.5	94.03
0.65	92.74
0.8	87.42
1	0
1.5	0



Πίνακας IV. – Σχήμα Δ . Ποσοστό ανίχνευσης για τιμές του Min Diversity για τα MSER- και η αντίστοιχη γραφική παράσταση

5.3.2 Μετρήσεις για διαφορετικό ποσοστό επικάλυψης εμβαδού

Για την καλύτερη εποπτεία του αποτελέσματος της ανίχνευσης, παρατίθεται και γράφημα με το ποσοστό ανίχνευσης για διαφορετικά ποσοστά επικάλυψης μεταξύ του ορθογωνίου που ανιχνεύθηκε ότι περιέχει τη πινακίδα, και του αντίστοιχου ορθογωνίου που καταγράφηκε με το χέρι. Το αποτέλεσμα της ανίχνευσης για διαφορετικά ποσοστά επικάλυψης απεικονίζονται στο σχήμα Ε. Ακόμα και για ποσοστό επικάλυψης 90%, οι πινακίδες ανιχνεύονται σωστά σε πάνω από τις μισές εικόνες.



Σχήμα Ε. Ποσοστό αναγνώρισης για διάφορα ποσοστά επικάλυψης εμβαδού.

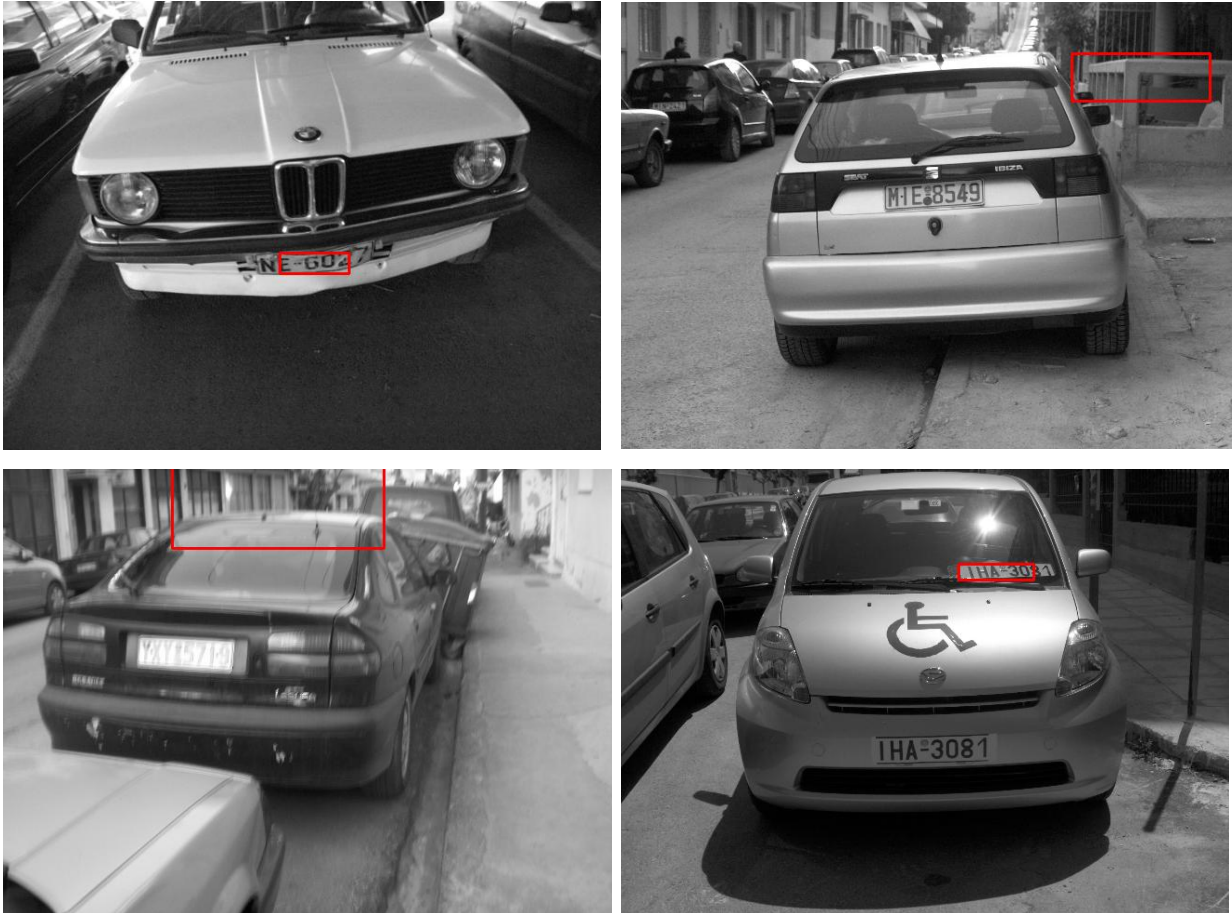
5.3.3 Παραδείγματα λανθασμένης ανίχνευσης

Οι παρακάτω εικόνες είναι παραδείγματα αποτυχημένης ανίχνευσης της πινακίδας, όπου δεν ανιχνεύθηκε καμία πινακίδα, δηλαδή δεν βρέθηκαν MSER+ που να πληρούν τα κριτήρια σχετικά με τις περιεχόμενες περιοχές τους. Στην πρώτη περίπτωση, απουσία αναγνωρισμένης πινακίδας οφείλεται στη χαμηλή αντίθεση ανάμεσα στην πινακίδα και το χρώμα του αυτοκινήτου. Στη δεύτερη περίπτωση υπάρχει ακμή λόγω σκίασης που είναι πιο έντονη από τα όρια της πινακίδας. Στην τρίτη περίπτωση η εικόνα είναι θολή και στην τελευταία υπάρχει εμπόδιο μπροστά στην πινακίδα το οποίο αλλοιώνει τα όρια της ανιχνευόμενης περιοχής.



Εικόνα 5.1. Εικόνες στις οποίες δεν ανιχνεύθηκε η πινακίδα

Στο επόμενο σύνολο εικόνων παρουσιάζονται περιπτώσεις στις οποίες έγινε επιλογή κάποιας περιοχής ως περιοχής πινακίδας, η οποία όμως ήταν λανθασμένη. Στην πρώτη περίπτωση εντοπίζεται η περιοχή της πινακίδας αλλά το πλαίσιο που επιλέγεται είναι μικρότερο του 50% της πραγματικής επιφάνειας της πινακίδας. Στη δεύτερη περίπτωση υπάρχει πολύπλοκο φόντο στην εικόνα που περιέχει περιοχές ίδιου μεγέθους και αναλογίας με τους χαρακτήρες, ενώ στην τρίτη εικόνα υπάρχει επιπλέον πολύ μικρή αντίθεση εντός της πινακίδας. Τέλος, στην τέταρτη εικόνα υπάρχει και δεύτερη πινακίδα που επιλέγεται αντί της επιθυμητής.



Εικόνα 5.2. Εικόνες στις οποίες ανιχνεύθηκε λανθασμένα η πινακίδα

5.3.4 Εφαρμογή των παραμέτρων στα μικρότερα σύνολα δεδομένων

Έχοντας υπολογίσει τις παραμέτρους των MSER που δίνουν τα καλύτερα αποτελέσματα για το σύνολο δεδομένων στο οποίο αξιολογήθηκαν, μπορεί να γίνει δοκιμή της μεθόδου και σε «άγνωστο» σύνολο δεδομένων. Δηλαδή σε σύνολο δεδομένων με το οποίο δεν έχουν γίνει δοκιμές των παραμέτρων, ώστε να φανεί πόσο εύρωστο είναι το σύστημα που αναπτύχθηκε. Γι' αυτό το σκοπό χρησιμοποιείται το κροατικό σύνολο δεδομένων, το οποίο αποτελείται από 502 εικόνες. Έγινε δοκιμή του αλγορίθμου με τις παραμέτρους που έδωσαν τα βέλτιστα αποτελέσματα και το ποσοστό ανίχνευσης για 50% επικάλυψη ήταν 87,65%. Ένας βασικός λόγος που το ποσοστό ήταν μικρότερο από το αντίστοιχο για το ελληνικό σύνολο δεδομένων, είναι η ύπαρξη περιπτώσεων που δεν συναντώνται στις ελληνικές πινακίδες, όπως αυτές που φαίνονται στις εικόνες 5.3.

Τέλος, έγινε δοκιμή και για ένα υποσύνολο του ελληνικού συνόλου δεδομένων, το οποίο αποτελούνταν μόνο από πινακίδες νέας γενιάς, δηλαδή μετά το 2004, οι οποίες ήταν συνήθως πιο καθαρές, αλλά χωρίς να απουσιάζουν οι σκιάσεις εντός των πινακίδων. Το αποτέλεσμα της ανίχνευσης για τις 112 τέτοιες εικόνες ήταν 97,29%.



(α)

(β)

Εικόνα 5.3. Παραδείγματα εικόνων του κροατικού συνόλου δεδομένων.

(α) Ο λόγος πλάτους προς ύψος της πινακίδας είναι αρκετά διαφορετικός από τις ελληνικές πινακίδες, (β) Τα χρώματα εντός της πινακίδας είναι πολύ διαφορετικά

5.4 Αποτελέσματα της αναγνώρισης χαρακτήρων

Για τις δοκιμές της αναγνώρισης χαρακτήρων με χρήση του tesseract-OCR χρησιμοποιήθηκε η διεπαφή του προγράμματος για σύνδεση με την OpenCV. Έγιναν δοκιμές για βελτιώση του αποτελέσματος, τόσο κατά το στάδιο της προεπεξεργασίας, όσο και κατά το στάδιο της ανάγνωσης της τελικής εικόνας από το tesseract. Σε αυτή την ενότητα παρουσιάζονται τα αποτελέσματα αυτών των δοκιμών

5.4.1 Δοκιμές στο στάδιο της προεπεξεργασίας

Αρχικά έγιναν δοκιμές με τις παραμέτρους της προεπεξεργασίας, οι οποίες παρουσιάζονται στην ενότητα 4.2 . Τα αποτελέσματα που παρατίθενται στον παρακάτω πίνακα προέκυψαν με τις βέλτιστες παραμέτρους του tesseract, για την εικόνα της πινακίδας από την κατάτμηση που έγινε με το χέρι, και όχι με το αποτέλεσμα της ανίχνευσης πινακίδας.

Πιο συγκεκριμένα, έγινε πρώτα αναγνώριση της πινακίδας με περιορισμό να αναγνωρίζονται μόνο γράμματα που εμφανίζονται στις ελληνικές πινακίδες, και στη συνέχεια δεύτερη αναγνώριση με περιορισμό να αναγνωρίζονται μόνο αριθμοί. Στη συνέχεια λήφθηκαν οι τρεις πρώτοι χαρακτήρες της συμβολοσειράς που προέκυψε με την πρώτη αναγνώριση και οι τέσσερις τελευταίοι της δεύτερης. Αφού αφαιρέθηκαν τα κενά, έγινε η σύγκριση με τη συμβολοσειρά της κάθε πινακίδας που καταγράφηκε με το χέρι. Το αποτέλεσμα ανάγνωσης

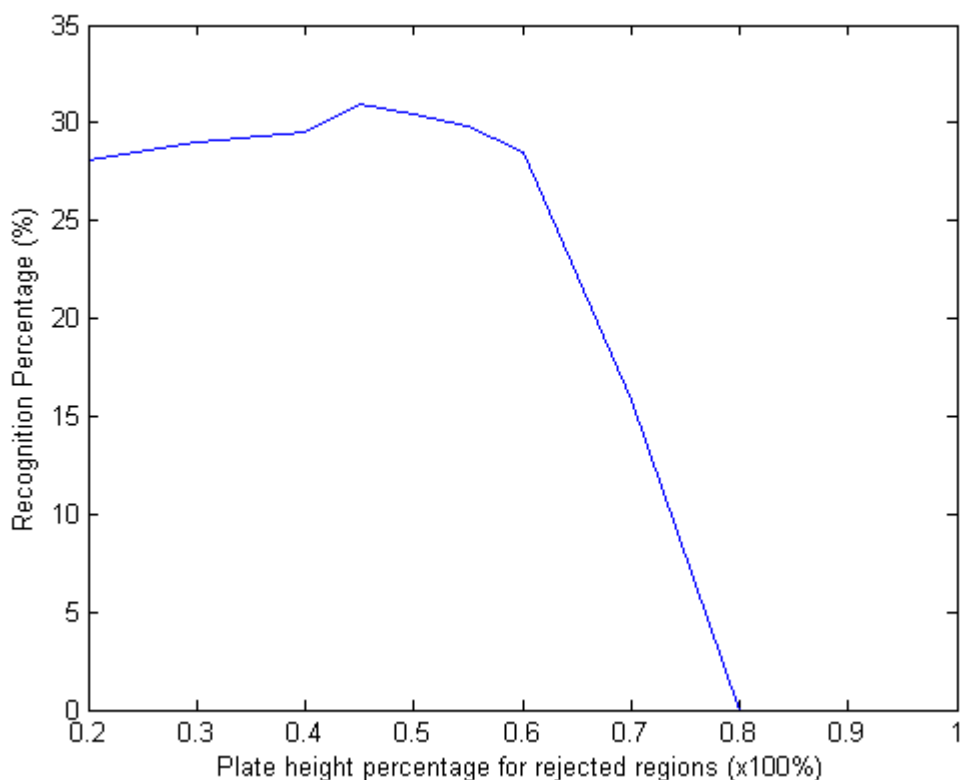
αναφέρεται σε σωστή αναγνώριση όλων των χαρακτήρων της πινακίδας, δηλαδή των τριών γραμμμάτων και των τεσσάρων αριθμών της πινακίδας.

Μέθοδος παρεμβολής κατά τη μεγέθυνση	Μέθοδος φιλτραρίσματος	Μέθοδος κατωφλίωσης	Αποτέλεσμα ανάγνωσης ολόκληρης πινακίδας
πλησιέστερου γείτονα	γκαουσιανό	προσαρμοστικό με συντελεστές Gauss	19,19%
γραμμική	γκαουσιανό	προσαρμοστικό με συντελεστές Gauss	26,61%
εμβαδόν pixel	γκαουσιανό	προσαρμοστικό με συντελεστές Gauss	24,68%
δικυβική	γκαουσιανό	προσαρμοστικό με συντελεστές Gauss	28,06%
Lanczos	γκαουσιανό	προσαρμοστικό με συντελεστές Gauss	27,10%
χωρίς μεγέθυνση	γκαουσιανό	προσαρμοστικό με συντελεστές Gauss	20,32%
δικυβική	υπερατό	προσαρμοστικό με συντελεστές Gauss	27,10%
δικυβική	μέσου όρου	προσαρμοστικό με συντελεστές Gauss	24,03%
δικυβική	διαμέσου	προσαρμοστικό με συντελεστές Gauss	25,65%
δικυβική	χωρίς φίλτρο	προσαρμοστικό με συντελεστές Gauss	28,55%
δικυβική	γκαουσιανό	οικουμενικό	14,03%
δικυβική	γκαουσιανό	50% του εύρους φωτεινότητας κάθε εικόνας	23,06%
δικυβική	γκαουσιανό	25% του εύρους φωτεινότητας κάθε εικόνας	15,81%
δικυβική	γκαουσιανό	προσαρμοστικό μέσης τιμής	22,10%
δικυβική	γκαουσιανό	μέθοδος Otsu	23,55%
δικυβική	γκαουσιανό	χωρίς κατωφλιοποίηση	0,0%

Πίνακας V. Αποτελέσματα αναγνώρισης ολόκληρης της πινακίδας με συνδυασμό διαφόρων παραμέτρων κατά την προεπεξεργασία της εικόνας.

Φαίνεται δηλαδή ότι τόσο η μεγέθυνση με τη σωστή μέθοδο παρεμβολής, όσο και η κατωφλίωση της εικόνας, βελτιώνουν σημαντικά τα αποτελέσματα αναγνώρισης. Το φίλτράρισμα, αν και οπτικά φαίνεται να βελτιώνει την ποιότητα κάποιων εικόνων, τελικά δεν βελτιώνει τα αποτελέσματα αναγνώρισης, με μόνο το γκαουσιανό φίλτρο να τα διατηρεί στο ίδιο επίπεδο.

Στις παραπάνω μετρήσεις, μετά την εφαρμογή των παραπάνω ενεργειών και πριν την αναγνώριση του κειμένου από τη μηχανή OCR, μεσολάβησε η απομάκρυνση των περιοχών μικρού ύψους, που συνήθως αντιστοιχούν σε παύλες, σφραγίδες και οπές για βίδες επί της πινακίδας. Το ύψος των περιοχών, κάτω από το οποίο αυτές απορρίπτονται, είναι ακόμη ένας παράγοντας που επηρεάζει το τελικό αποτέλεσμα, με τον εξής τρόπο: αν επιλεγεί μικρό δεν θα αφαιρέσει όλες τις ανεπιθύμητες περιοχές, ενώ αν είναι πολύ μεγάλο θα αφαιρέσει και κάποια τμήματα χαρακτήρων που ίσως έχουν χωριστεί σε κομμάτια κατά την κατωφλίωση. Έτσι έγιναν μετρήσεις και για αυτήν την παράμετρο, με τα αποτελέσματα να φαίνονται στο παρακάτω σχήμα. Οι παράμετροι που χρησιμοποιήθηκαν σε αυτό το βήμα ήταν δικυβική παρεμβολή, γκαουσιανό φίλτρο και προσαρμοστική κατωφλίωση με συντελεστές Gauss.



Σχήμα Στ . Διάγραμμα του ποσοστού αναγνώρισης σε σχέση με ελάχιστο ύψος που πρέπει να έχει μια περιοχή ώστε να μην απορριφθεί, σε σχέση με το ύψος της πινακίδας

Τέλος, επιχειρήθηκε η βελτίωση των αποτελεσμάτων με χρήση μορφολογικών φίλτρων, όπου διαπιστώθηκε ότι η χρήση μορφολογικού φίλτρου opening 3x3 στην κατωφλιωμένη εικόνα βελτίωσε λίγο το ποσοστό αναγνώρισης.

5.4.2 Δοκιμές στις παραμέτρους του tesseract-OCR

Το δεύτερο κομμάτι όπου μπορούν να γίνουν προσαρμογές για τη βελτίωση των αποτελεσμάτων της αναγνώρισης των χαρακτήρων είναι η αλλαγή των παραμέτρων αρχικοποίησης του tesseract. Όπως αναφέρθηκε, ένα πρώτο βήμα είναι η μη φόρτωση του λεξικού της αγγλικής γλώσσας, η οποία βελτιώνει τα αποτελέσματα. Μια δεύτερη σημαντική παράμετρος είναι η λεγόμενη «λευκή λίστα» (“whitelist”), η οποία περιέχει τους χαρακτήρες που είναι πιθανό να βρίσκονται στην εικόνα, απορρίπτοντας όλους τους άλλους. Έγιναν δοκιμές χωρίς χρήση λίστας, με χρήση ενιαίας λίστας για όλη την πινακίδα, και με ξεχωριστές λίστες για τα γράμματα και τους αριθμούς.

Τέλος, ένας ακόμη τρόπος να βελτιωθούν τα αποτελέσματα είναι η αφαίρεση των κενών από τη συμβολοσειρά της αναγνώρισης, τα οποία μπορεί να είναι λανθασμένα, πριν τη σύγκρισή της με την πραγματική συμβολοσειρά. Τα αποτελέσματα που φαίνονται παρακάτω έγιναν για όριο απόρριψης περιοχών το 50% του ύψους της πινακίδας, και χωρίς χρήση φίλτρου opening. Στην πρώτη γραμμή παρατίθεται η περίπτωση κατά την οποία δεν γίνεται φόρτωση του λεξικού, χρησιμοποιούνται δύο ξεχωριστές λευκές λίστες και αφαιρούνται τα κενά κατά τη σύγκριση με την πραγματική συμβολοσειρά. Στις υπόλοιπες γραμμές φαίνονται οι προσαρμογές που έγιναν και το αποτέλεσμα της καθεμιάς.






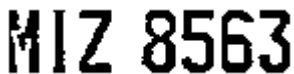
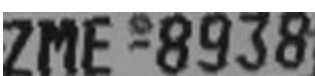

Προσαρμογή	Ποσοστό αναγνώρισης γραμμάτων	Ποσοστό αναγνώρισης αριθμών	Ολικό ποσοστό αναγνώρισης
με δύο ξεχωριστές whitelist	55%	44,19%	30,48%
με κοινή whitelist	49,52%	39,68%	25,16%
χωρίς whitelist	45,81%	36,61%	22,58%
με φόρτωση λεξικού	52,58%	44,19%	29,84%
χωρίς αφαίρεση κενών πριν την αποτίμηση	54,84%	43,39%	30,32%

Πίνακας VI. Αποτέλεσμα αναγνώρισης ολόκληρης της πινακίδας με δοκιμή σε διάφορες παραμέτρους του tesseract-OCR

5.4.3 Παραδείγματα αποτυχημένης αναγνώρισης

Στον παρακάτω πίνακα παρουσιάζονται ενδεικτικά κάποιες περιπτώσεις αποτυχημένης αναγνώρισης χαρακτήρων. Στην πρώτη περίπτωση η εικόνα είναι πολύ θολή, ενώ στη δεύτερη γίνεται λανθασμένη κατάτμηση μετά την περιστροφή της εικόνας. Στην τρίτη

περίπτωση, η είσοδος στο σύστημα αναγνώρισης είναι αρκετά καλής ποιότητας αλλά γίνεται λανθασμένη ανάγνωση, και στην τέταρτη δε γίνεται σωστή απομάκρυνση του πλαισίου της πινακίδας, με αποτέλεσμα να αναγνωριστεί ένας χαρακτήρας παραπάνω από τους υπάρχοντες.

Περιστραμμένη, κατατμημένη εικόνα	Είσοδος του συστήματος οπτικής αναγνώρισης χαρακτήρων	Αποτέλεσμα αναγνώρισης
		IHB 5 17
		MIB 79
		MIZ 8553
		ZNE 89381

Πίνακας VII. Περιπτώσεις πινακίδων στις οποίες έγινε λανθασμένη αναγνώριση

5.4.4 Τελική επιλογή παραμέτρων – εφαρμογή στα μικρότερα σύνολα δεδομένων

Με βάση τις μετρήσεις που έγιναν και παρουσιάστηκαν στις δύο προηγούμενες υποενότητες, μπορεί τώρα να γίνει η τελική επιλογή των βέλτιστων παραμέτρων για την επίτευξη του καλύτερου ποσοστού αναγνώρισης. Για να επιτευχθεί αυτό πρέπει να επιλεγεί η δικυβική παρεμβολή κατά τη μεγέθυνση της εικόνας της πινακίδας, ύστερα να εφαρμοστεί γκαουσιανό φίλτρο και η κατωφλίωση να γίνει με την προσαρμοστική μέθοδο με συντελεστές Gauss. Ύστερα πρέπει να γίνει αφαίρεση των περιοχών που δεν υπερβαίνουν το 45% του ύψους της πινακίδας και να εφαρμοστεί στην εικόνα φίλτρο opening 3x3, πριν αυτό παραδοθεί στη μηχανή tesseract.

Εκεί πρέπει να μην φορτωθεί το λεξικό, να γίνει χρήση ξεχωριστής «λευκής λίστας» για τα γράμματα και τους αριθμούς, και να αφαιρεθούν τα κενά της τελικής συμβολοσειράς. Το ποσοστό αναγνώρισης με αυτές τις παραμέτρους φτάνει το 32,26%. Τα γράμματα αναγνωρίζονται σωστά στο 54,52% των πινακίδων και οι αριθμοί στο 48,23%. Στο 67% των πινακίδων αναγνωρίζονται σωστά τουλάχιστον οι τρεις από τους τέσσερις αριθμούς. Τέλος, στο 11,3% των πινακίδων δεν αναγνωρίζεται σωστά κανένας από τους χαρακτήρες. Ο χρόνος εκτέλεσης στον υπολογιστή που χρησιμοποιήθηκε είναι περίπου 150 msec/εικόνα.

Με αυτές τις παραμέτρους, γίνεται αρχικά αποτίμηση ολόκληρου του αλγορίθμου. Δηλαδή, ως είσοδοι του συστήματος αναγνώρισης χαρακτήρων δίνονται οι συντεταγμένες των ορίων

των πινακίδων, όπως αυτές καταγράφηκαν με το χέρι, αλλά οι συντεταγμένες που προέκυψαν από την ανίχνευση της πινακίδας με τη μέθοδο των MSER. Το συνολικό αποτέλεσμα της ανίχνευσης και αναγνώρισης πινακίδων για το σύνολο δεδομένων είναι 29,68% και ο συνολικός χρόνος είναι περίπου 240 msec/εικόνα.

Επίσης δοκιμάστηκε να δοθούν ως είσοδος στην αναγνώριση μόνο τα MSER που περιέχονταν στην πινακίδα που αναγνωρίστηκε, με μόνη επεξεργασία την περιστροφή τους ώστε να είναι οριζόντια. Οι παράμετροι του tesseract είναι οι ίδιοι που έδωσαν τα καλύτερα αποτελέσματα στις προηγούμενες περιπτώσεις. Το αποτέλεσμα της αναγνώρισης σε αυτήν την περίπτωση είναι 6,77%. Το αποτέλεσμα αυτό κυρίως οφείλεται στο ότι για να βρεθεί η περιοχή της πινακίδας, χρειάστηκε να απορριφθούν πολλές περιοχές MSER-, κάποιες εκ των οποίων αντιστοιχούσαν σε χαρακτήρες, οι οποίοι δεν είναι δυνατό να ανακτηθούν.

Με τις βέλτιστες αυτές παραμέτρους έγινε δοκιμή και για υποσύνολο των ελληνικών πινακίδων που είναι κατασκευασμένες μετά το 2004, το οποίο έχει γενικά καλύτερης ποιότητας εικόνες. Το ποσοστό αναγνώρισης για προσδιορισμό των ορίων της πινακίδας με το χέρι είναι 46,85% (71% για τα γράμματα, 56% για τους αριθμούς). Έτσι, σε συνδυασμό με το αποτέλεσμα της ανίχνευσης που αναφέρεται στην ενότητα 5.3.4, το ποσοστό σωστής ανίχνευσης της πινακίδας και αναγνώρισής της είναι 31,53%.

5.5 Σύγκριση με άλλες μεθόδους

Στον παρακάτω πίνακα παρουσιάζεται επιγραμματικά η απόδοση των διάφορων μεθόδων που έχουν χρησιμοποιηθεί και έχουν αναλυθεί στο κεφάλαιο 2, τόσο για την ανίχνευση της πινακίδας, όσο και για την αναγνώριση των χαρακτήρων. Μέρος του πίνακα έχει ληφθεί από την εργασία [20] και ο υπόλοιπος συμπληρώθηκε με βάση τα στοιχεία που παρέχουν οι ερευνητές στις εργασίες τους.

Εργασία	ποσοστό αναγνώρισης: (pl) ανίχνευση πινακίδας, (ch) αναγνώριση χαρακτήρων, (ον) συνολικό ποσοστό	πλατφόρμα και επεξεργαστής	χρόνος (sec)	μεθοδολογία. P: μέθοδος ανίχνευσης πινακίδας, R: μέθοδος αναγνώρισης χαρακτήρων
παρούσα εργασία	94.03% (pl), 32.26% (ch - ολόκληρη πινακίδα), 29,68% (ον)	C++, Visual Studio 2010, Intel Core i5 2.3 GHz	0.14 (ανίχνευση), 0.15 (αναγνώριση), ~0.3 συνολικά)	P: MSER, R: tesseract-OCR, τεστ: 620 εικόνες
[6]	80.74%-97.16% (ch), 23.23%-94.65% (ch για ολόκληρη την πινακίδα)	δεν αναφέρεται	0.07 (ανίχνευση), 0.081 (συνολικά)	P: ακολουθία MSER, R: SVM
[9]	βλ. τελευταία στήλη	P IV, 2.4 GHz	~0.5	P: στατιστικά οριζόντιων - κάθετων ακμών. 96,8% (πρώτο σετ), 99.6% (δεύτερο σετ), 99.7% (τρίτο σετ).
[10]	99.60%	P IV, 1.7GHz	~0.1	P: στατιστικά ακμών και μορφολογία. τεστ:~10000 εικόνες.
[11]	81.2% (όλες), 96.3% μόνο καλής ποιότητας (pl), 85.2%-92.3% (εμπορικό ocr), 58.6%-97.7% (ελεύθερο ocr)	δεν αναφέρεται	δεν αναφέρεται	P: ταξικό φίλτρο, R: έτοιμο πρόγραμμα
[12]	96,5% (pl)	C++, P IV, 512MB RAM	0.3	P: μορφολογικό+Sobel
[13]	92.3%(pl), 95.7%(ch)	δεν αναφέρεται	~1.6	P: σύγκριση με εικόνα-πρότυπο με τεχνικές με χρήση block
[14]	94.2%(pl), 98.6%(ch)	C++, P II, 300MHz	δεν αναφέρεται	P: κάθετες ακμές και γεωμετρικά χαρακτηριστικά. τεστ: 104 εικόνες
[15]	97%(ον την ημέρα) 90%(ον τη νύχτα)	δεν αναφέρεται	0.1-0.5	υπολογιστής υψηλής απόδοσης για ΑΠΚ
[16]	δεν αναφέρεται	P III, 600 MHz	0.15	P: γενετικοί αλγόριθμοι.
[17]	90%(ον)	P III, 1 GHz	~0.3	P: χρώμα και σχήμα.

[18]	89%(pl)	C++, P III, 800MHz	~1.3	P: SVM, συνεχής προσαρμοζόμενος αλγόριθμος meanshift (CAMShift) για χρωματιστές εικόνες.
[19]	98%(pl)	P IV, 1.4GHz	3.12	P: μετασχηματισμός Gabor και κβαντοποίηση διανυσμάτων.
[20]	96.5%(pl) 89.1%(ch) 86.0% (ov)	Microsoft's Visual C++ 6.0, Pentium IV 3.0 GHz με 512-MB RAM	0.111 det, 0.037 proc, 0.128 rec, 276 tot	P: κινούμενο ομόκεντρα παράθυρα. R: πιθανοτικά ΝΔ (σχεδόν ίδιο σύνολο δεδομένων με την παρούσα εργασία)
[21]	δεν αναφέρεται	δεν αναφέρεται	δεν αναφέρεται	P: κυματίδια
[22]	92.4%(pl)	Borland C++, P IV 1.6GHz	δεν αναφέρεται	P: κυματίδια
[23]	91.25% (pl), 98.8% (ch), 90.61% (ov)	C, PC 486 50MHz	0.09(ανίχνευση), 2ms(αναγνώριση)	P: ΝΔ, R: ταύτιση προτύπων
[24]	93.6%(pl)	Visual C++, AMD Athlon 900 MHz	~1.3	P: γενικευμένος μετασχηματισμός συμμετρίας και image warping.
[25]	100%(pl), 94.7%(ch)	Visual C++, PII	~1	P:2 ΝΔ χρονοκαθυστερήσης (TDNNs). R: μηχανή υποστηρικτικών διανυσμάτων (SVM). τεστ: 1000 ακολουθίες βίντεο.
[26]	85%(pl)	δεν αναφέρεται	δεν αναφέρεται	P: κυψελωτό ΝΔ διακριτού χρόνου (DTCNN). R: perceptron πολλών επιπέδων.
[27]	85%(pl)	Pulse-Coupled NN Processor	δεν αναφέρεται	P: Pulse Coupled Neural Network (PCNN).
[28]	97.6%(pl) 95.6%(ch)	P IV, 1.6 GHz	~2.5	P: ασαφείς κανόνες για το χρώμα. R: αυτοοργανούμενο ΝΔ. τεστ: 1065 εικόνες από 71 οχήματα.
[29]	75.4%(pl) 98.5%(ch)	δεν αναφέρεται	δεν αναφέρεται	P:ασαφής λογική. R: ΝΔ (Perceptron πολλών επιπέδων). τεστ:10000 εικόνες.
[30]	92.80%	C++	δεν αναφέρεται	P: γενετικοί αλγόριθμοι.
[31]	80.6%(pl) κατά μέσο όρο	P III, 500 MHz	0.18	P: γενετικοί αλγόριθμοι.

[32]	97.0-99.3%	δεν αναφέρεται	δεν αναφέρεται	P: σύστημα μετάδοσης με μεγάλο δυναμικό εύρος. τεστ: 1000 εικόνες.
[35]	98.7% (ch), 90% (ov)	δεν αναφέρεται	δεν αναφέρεται	P: 1D-DFT, R: Ασαφές ΝΔ
[36]	95.6%(pl)	δεν αναφέρεται	δεν αναφέρεται	P:AdaBoost (adaptive boosting).
[37]	98.8%(pl), 95.2%(ch)	Visual C++, P IV 1.4Ghz	0.65(plates) 0.1(per character)	P:μετασχηματισμός Hough. R:κρυφά μοντέλα Markov.
[38]	98.5%(pl)	AMD Athlon 1.2 GHz	~34	P: επιτηρούμενος ταξινομητής υφής. R: "κοντινότερος γείτονας". τεστ: 131 εικόνες.
[39]	92.3% 1 rep, 97.1% 7 reps. (pl)	Matlab	5sec, 7sec	P: μορφολογικό + ιδιότητες σχημάτων
[40]	98%(pl)	C++, P I, 200Mhz	0.2	P:κβαντισμός διανυσμάτων (VQ) για συμπίεση εικόνας και ΑΠΚ.
[42]	80%(ov)	Intel 486/66Mhz	~15	P: επεξεργασία γκρι εικόνας. R: ΝΔ.
[43]	84.5% (ch segmentation)	δεν αναφέρεται	δεν αναφέρεται	χωρισμός χαρακτήρων: εξαγωγή διανύσματος χαρακτηριστικών και μορφολογία
[44]	δεν αναφέρεται	δεν αναφέρεται	δεν αναφέρεται	P: στοχαστικό πεδίο Markov και γενετικοί αλγόριθμοι σε ακολουθία βίντεο.
[45]	~95%(ch)	δεν αναφέρεται	δεν αναφέρεται	R: ταύτιση προτύπων
[46]	δεν αναφέρεται	C++	δεν αναφέρεται	R: ταύτιση προτύπων
[47]	72%-91% (ch)	δεν αναφέρεται	δεν αναφέρεται	R: ταύτιση προτύπων
[48]	93.1%(ov)	δεν αναφέρεται	δεν αναφέρεται	P: χωρική επεξεργασία σε κλίμακα του γκρι, R: dynamic projection warping.
[50]	80.2% (ch)	δεν αναφέρεται	δεν αναφέρεται	R: HOG + Strokelets

Πίνακας VIII. Συνοπτική παρουσίαση των αποτελεσμάτων άλλων μεθόδων

Κεφάλαιο 6

Συμπεράσματα

Από την υλοποίηση της μεθόδου και τη σύγκρισή της με τις υπόλοιπες, παρατηρεί κανείς ότι η απόδοσή της είναι ανταγωνιστική των καλύτερων μεθόδων που έχουν αναπτυχθεί. Ο χρόνος εκτέλεσής της (περίπου 4 εικόνες ανά δευτερόλεπτο) επιτρέπει τη χρήση της για εφαρμογές πραγματικού χρόνου, αρκεί να βρεθεί μια καλύτερη μέθοδος για την αναγνώριση των χαρακτήρων. Επίσης, οι μετρήσεις πραγματοποιήθηκαν σε ένα αρκετά απαιτητικό σύνολο δεδομένων, ενώ από παραδείγματα εικόνων σε άλλες εργασίες φάνηκε ότι οι εικόνες που επεξεργάστηκαν ήταν αρκετά πιο καθαρές.

Από το γεγονός αυτό αναδεικνύεται ένα σημαντικό πρόβλημα της ανάπτυξης μεθόδων για την αναγνώριση πινακίδων κυκλοφορίας: Η μη ύπαρξη ενός διεθνούς τεκμηριωμένου συνόλου δεδομένων. Αυτό ίσως έχει ως αιτία λόγους προστασίας προσωπικών δεδομένων ή το γεγονός ότι οι πινακίδες στις διάφορες χώρες διαφέρουν αρκετά μεταξύ τους στο σχήμα και τη μορφή των χαρακτήρων, ώστε να μην έχει νόημα μια τέτοια προσπάθεια. Όμως η σύγκριση των αποτελεσμάτων μεταξύ των εργασιών δε γίνεται σε κοινά δεδομένα, και αυτό δυσκολεύει την απόφαση για την καταλληλότερη μέθοδο.

Σχετικά με τα αποτελέσματα της αναγνώρισης χαρακτήρων παρατηρεί κανείς εύκολα ότι αυτά απέχουν αρκετά από τα επιθυμητά ενός τέτοιου συστήματος, ακόμα και με την καλύτερη δυνατή βελτίωση της εικόνας. Γίνεται δηλαδή φανερό, ότι η μηχανή tesseract-OCR, που θεωρείται από τις καλύτερες ελεύθερες στην ανάγνωση κειμένου, δεν δίνει καλά αποτελέσματα στην αναγνώριση πινακίδων. Η καλή της φήμη ίσως οφείλεται στην ικανότητά της να αναγνωρίζει σωστά λέξεις στηριζόμενη όχι μόνο στην καλή αναγνώριση των μεμονωμένων χαρακτήρων, αλλά και των γειτονικών της στην ίδια ή σε διπλανές λέξεις, δηλαδή από τα συμφραζόμενα.

Επίσης, ένας άλλος πιθανός λόγος είναι ότι η γραμματοσειρά που χρησιμοποιείται στις ελληνικές πινακίδες προ του 2004 δεν είναι συνηθισμένη σε κείμενα, και ίσως μια εκπαίδευση της μηχανής με βάση αυτήν να βελτιώνει τα αποτελέσματα. Για να εξεταστούν τα αποτελέσματα κάτω από καλύτερες συνθήκες ως προς την ποιότητα των εικόνων, απομονώθηκαν από το σύνολο δεδομένων οι φωτογραφίες με ελληνικές πινακίδες κυκλοφορίας κατασκευασμένες μετά το 2004, και τα αποτελέσματα ήταν καλύτερα, αλλά και πάλι όχι αποδεκτά για ένα τέτοιο σύστημα.

Ένας τρόπος βελτίωσης θα ήταν η χρήση πολλαπλών εικόνων για κάθε πινακίδα, που θα μπορούσαν να προκύψουν για παράδειγμα από ακολουθία βίντεο. Η χρήση βίντεο βελτιώνει επίσης την αντίληψη, αφού μπορεί να βρεθεί πρώτα η περιοχή του κινούμενου οχήματος και να γίνει αναζήτηση της πινακίδας μόνο εντός της, περιορίζοντας έτσι τις λανθασμένες ανιχνεύσεις λόγω πολύπλοκου φόντου. Όμως δεν βρέθηκε κάποιο διαθέσιμο βίντεο για να δοκιμαστούν τεχνικές πάνω σε αυτό.

Ως μελλοντική έρευνα προτείνεται η αναζήτηση μεθόδων που θα βελτιώνουν την είσοδο για το σύστημα αναγνώρισης χαρακτήρων, ώστε να μεγιστοποιηθεί η χρήση των δυνατοτήτων του. Επίσης, προτείνεται η ανάπτυξη τεχνικών για την δημιουργία ενός εξειδικευμένου συστήματος αναγνώρισης χαρακτήρων που περιέχονται στις ελληνικές πινακίδες. Στις πινακίδες που κατασκευάζονται μετά το 2004 χρησιμοποιείται μια διεθνώς προτυποποιημένη γραμματοσειρά, η DIN 1451, οπότε μπορεί να εκπαιδευτεί ένα σύστημα για να την αναγνωρίζει σωστά, είτε αυτό στηρίζεται σε ταύτιση προτύπων ή σε μεθόδους μηχανικής μάθησης. Γι' αυτό το σκοπό παρατίθενται εδώ οι πρότυπες εικόνες



Εικόνα 6.1. Η γραμματοσειρά που χρησιμοποιείται στις νέες ελληνικές πινακίδες κυκλοφορίας

Βιβλιογραφία

- [1] "bbc.co.uk," 21 November 2005. [Online]. Available: http://news.bbc.co.uk/2/hi/uk_news/england/bradford/4455918.stm.
- [2] Business Partners Online, April 2011. [Online]. Available: <http://bponline.amcham.gr/?p=730>.
- [3] University of Reading - Intelligent Systems Research Laboratory, 2009. [Online]. Available: <http://www.imss.reading.ac.uk/index.php/projects/79-none/99-moveon>.
- [4] J. Matas, O. Chum, M. Urban and T. Pajdla, "Robust wide baseline stereo from maximally stable extremal regions," *BMVC02*, vol. 1, p. 384–393, 2002.
- [5] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 27, no. 10, p. 1615–1630, 2005.
- [6] M. Donoser, c. Arth and H. Bischof, "Detecting, Tracking and Recognizing License Plates," *Asian Conference on Computer Vision*, pp. 447-456, 2007.
- [7] Εργαστήριο Τεχνολογίας Πολυμέσων ΣΗΜΜΥ ΕΜΠ, "<http://www.medialab.ntua.gr/research/LPRdatabase.html>".
- [8] ZEMRIS, "<http://www.zemris.fer.hr/projects/LicensePlates/english/results.shtml>".
- [9] D. Zheng, Y. Zhao and J. Wang, "An efficient method of license plate detection," *Pattern Recognition Letters*, vol. 26, no. 15, p. 2431–2438, 2005.
- [10] B. Hongliang and L. Changping, "A hybrid license plate extraction method based on edge statistics and morphology," in *17th Int. Conf. Pattern Recognition (ICPR'04)*, 2004.
- [11] V. Shapiro, G. Gluhchev and D. Dimov, "Towards a multinational car license plate recognition system," *Machine Vision and Applications*, no. 17, p. 173–183, 2006.
- [12] H. Mahini, S. Kasaei, F. Dorri and F. Dorri, "An Efficient Features–Based License Plate Localization Method," in *Proc. IEEE Int.Conf Pattern Recognition ICPR*, 2006.
- [13] H.-J. Lee, S.-Y. Chen and S.-Z. Wang, "Extraction and recognition of license plates of motorcycles and vehicles on highways," in *Proc ICPR*, 2004.
- [14] S. Z. Wang and H. M. Lee, "Detection and recognition of license plate characters with different appearances," in *Proc. Conf. Intell. Transp. Syst.*, 2003.
- [15] D. Yan, M. Hongqing, L. Jilin and L. Langang, "A high performance license plate recognition system based on the web technique," in *Proc. Conf. Intell. Transp. Syst.*, 2001.
- [16] G. Adorni, S. Cagnoni, M. Gori and M. Mordonini, "Access control system with neuro-fuzzy supervision," in *Proc. Conf. Intell. Transp. Syst.*, 2001.
- [17] X. Shi, W. Zhao and Y. Shen, "Automatic license plate recognition system based on color image processing," *Lecture Notes on Computer Science*, vol. 3483, p. 1159–1168, 2005.
- [18] K. I. Kim, K. Jung and J. H. Kim, "Color texture-based object detection: An application to license plate localization," *Lecture Notes on Computer Science*, vol. 2388, p. 293–309.

- [19] F. Kahraman, B. Kurt and M. Gökmen, "License plate character segmentation based on the gabor transform and vector quantization," *Lecture Notes on Computer Science*, vol. 2869, p. 381–388, 2003.
- [20] C.-N. E. Anagnostopoulos, I. E. Anagnostopoulos, V. Loumos and E. Kayafas, "A License Plate-Recognition Algorithm for Intelligent Transportation System Applications," *IEEE Trans. Intelligent Transportation Systems*, vol. 7, no. 3, September 2006.
- [21] S. Kumar, N. Khanna, S. Chaudhury and S. D. Joshi, "Locating Text in Images using Matched Wavelets," in *Proc. Eight International Conference on Document Analysis and Recognition (ICDAR'05)*, 2005.
- [22] C.-T. Hsieh, Y.-S. Juan and K.-M. Hung, "Multiple license plate detection for complex background," in *Proc. Int. Conf. AINA*, 2005.
- [23] E. R. Lee, P. K. Kim and H. J. Kim, "Automatic Recognition of A Car License Plate Using Color Image Processing," in *Proceeding of International Conference on Image Processing*, 1994.
- [24] D.-S. Kim and S.-I. Chien, "Automatic car license plate extraction using modified generalized symmetry transform and image warping," in *Proc. ISIE*, 2001.
- [25] K. K. Kim, K. I. Kim, J. B. Kim and H. J. Kim, "Learning-based approach for license plate recognition," in *Proc. IEEE Signal Process. Soc. Workshop, Neural Netw. Signal Process.*, 2000.
- [26] M. H. ter Brugge, J. H. Stevens, J. A. G. Nijhuis and L. Spaanenburg, "License plate recognition using DTCNNs," in *Proc. IEEE Int. Workshop Cellular Neural Netw. and Appl.*, 1998.
- [27] M. I. Chacon and A. Zimmermann, "License plate location based on a dynamic PCNN scheme," in *Proc. Int. Joint Conf. Neural Netw.*, 2003.
- [28] S.-L. Chang, L.-S. Chen, Y.-C. Chung and S.-W. Chen, "Automatic License Plate Recognition," in *Proc. IEEE Int. Conf Intelligent Transportation Systems*, 2004.
- [29] J. A. G. Nijhuis, M. H. ter Brugge, K. A. Helmholt, J. P. W. Pluim, L. Spaanenburg, R. S. Venema and M. A. Westenberg, "Car license plate recognition with neural networks and fuzzy logic," in *Proc. IEEE Int. Conf. Neural Netw.*, 1995.
- [30] S. K. Kim, D. W. Kim and H. J. Kim, "A recognition of vehicle license plate using a genetic algorithm based segmentation," in *Proc. Int. Conf. Image Processing*, 1996.
- [31] S. Yoshimori, Y. Mitsukura, M. Fukumi and N. Akamatsu, "License plate detection using hereditary threshold determine method," *Lecture Notes in Artificial Intelligence*, vol. 2773, p. 585–593, 2003.
- [32] T. Naito, T. Tsukada, K. Yamada, K. Kozuka and S. Yamamoto, "Robust license plate recognition method for passing vehicles under outside environment," *IEEE Trans. Veh. Technol.*, vol. 49, no. 6, p. 2309–2319, Nov 2000.
- [33] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst., Man, Cybern.*, Vols. SMC-9, p. 62–66, Jan 1979.
- [34] J. Matas and K. Zimmermann, "Unconstrained License Plate and Text Localization and Recognition," in *Intelligent Transportation Systems*, 2005.

- [35] R. Parisi, E. D. D. Claudio, G. Lucarelli and G. Orlandi, "Car plate recognition by neural networks and image processing," in *Proc. IEEE Int. Symp. Circuits and Systems*, 1998.
- [36] L. Dlagnekov, "License Plate Detection Using AdaBoost," *Computer Science and Engineering Department, San Diego*, 2004.
- [37] T. D. Duan, T. L. Hong Du, T. V. Phuoc and N. V. Duan, "Building an automatic vehicle license plate recognition system," in *Proc. Int. Conf. Comput. Sci. RIVF*, 2005.
- [38] J. Cano and J. C. Perez-Cortez, "Vehicle license plate segmentation in natural images," *Lecture Notes on Computer Science*, vol. 2652, p. 142–149, 2003.
- [39] V. Abolghasemi and A. Ahmadyfard, "Local Enhancement of Car Image for License Plate Detection," in *15th European Signal Processing Conference (EUSIPCO)*, 2007.
- [40] R. Zunino and S. Rovetta, "Vector quantization for license plate location and image coding," *IEEE Trans. Industrial Electron.*, vol. 47, no. 1, p. 159–167, Feb 2000.
- [41] N. H. C. Yung, K. H. Au and A. H. S. Sai, "Recognition of vehicle registration mark on moving vehicles in an outdoor environment," in *Proc. IEEE Int. Conf. Intelligent Transportation Systems*, 1999.
- [42] S. Draghici, "A neural network based artificial vision system for license plate recognition," *Int. J. Neural Systems*, vol. 8, p. 113–126, 1997.
- [43] S. Nomura, K. Yamanaka, O. Katai, H. Kawakami and T. Shiose, "A novel adaptive morphological approach for degraded character image segmentation," *Pattern Recognition*, vol. 38, no. 11, p. 1961–1975, November 2005.
- [44] Y. Cui and Q. Huang, "Extracting characters of license plates from video sequences," *Mach. Vis. Appl.*, vol. 10, no. 5/6, p. 308–320, Apr 1998.
- [45] N. H. C. Yung, K. H. Au and A. H. S. Lai, "Recognition of vehicle registration mark on moving vehicles in an outdoor environment," in *Proc. IEEE Conf Intelligent Transportation Systems*, 1999.
- [46] C. Rahman, B. W and A. Radmanesh, "A real time vehicle's license plate recognition system," in *AVSS. Proceedings of the IEEE Conference on Advanced Video and Signal Based Surveillance*, Los Alamitos, 2003.
- [47] D. M. Emiris and D. E. Koulouriotis, "Automated optic recognition of alphanumeric content in car license plates in a semi-structured environment," in *IEEE Image Processing, 2001. Proceedings. 2001 International Conference*, 2001.
- [48] T. H. Wang, F. C. Ni, K. T. Li and Y. P. Chen, "Robust license plate recognition based on dynamic projection warping," in *Proc. IEEE Int. Conf. Netw., Sensing and Control*, 2004.
- [49] C.-Y. Lee, A. Bhardwaj, W. Di and R. Piramanthu, "Region-based Discriminative Feature Pooling for Scene Text Recognition".
- [50] Y. Cong, B. Xiang, S. Baoguang and L. Wenyu, "A Learned Multi-Scale Representation for Scene Text Recognition".
- [51] Google, "code.google.com/p/tesseract-ocr/".
- [52] Apache, "<http://www.apache.org/licenses/LICENSE-2.0>," 2004.

[53] Google, "<https://code.google.com/p/tesseract-ocr/wiki/FAQ>," [Online].

[54] D. Nistér and H. Stewenius, "Linear time maximally stable extremal regions," in *Computer Vision - ECCV 2008*, Springer, 2008, pp. 183-196.

Παράρτημα

Συναρτήσεις της OpenCV που χρησιμοποιήθηκαν για την υλοποίηση

Για την εύρεση των MSER χρησιμοποιήθηκε η βιβλιοθήκη “mser.h”, που δημιουργήθηκε από τον Liu Liu το 2009 με βάση την εργασία των Nistér, Stewénius [54], και έχει γραμμική πολυπλοκότητα, ενώ έχει ενσωματωθεί στις βιβλιοθήκες της OpenCV. Χρειάστηκε όμως να γίνει μια αλλαγή στον κώδικά της, προσθέτοντας μια παράμετρο στην κλήση της συνάρτησης εύρεσης των MSER, ώστε να υπολογίζει ξεχωριστά τα MSER+ και τα MSER-, αφού ο αυθεντικός αλγόριθμος τα έδινε και τα δύο στην έξοδο χωρίς διάκριση μεταξύ τους. Όμως αυτή η διάκριση είναι ζωτικό κομμάτι της μεθόδου που επιλέχθηκε, άρα ήταν απαραίτητη αυτή η μετατροπή για να προχωρήσει η υλοποίησή της.

Οι περιοχές που βρέθηκαν αποθηκεύονται σε δυναμικές δομές τύπου vector της C++, αφού δεν είναι από πριν γνωστό το μέγεθός τους. Επειδή όμως τα σημεία που αποτελούν το κάθε τέτοιο διάνυσμα είναι τυχαία ταξινομημένα εντός του, δεν μπορούν να εξαχθούν πολλά συμπεράσματα από αυτή τη δομή, πέρα από το μέγεθός της, δηλαδή το εμβαδόν της περιοχής σε αριθμό pixel, ή μια αδρή εκτίμηση της θέσης της μέσα στην εικόνα. Ένα πολύ χρήσιμο εργαλείο που παρέχει η OpenCV για την εξαγωγή περισσότερων συμπερασμάτων είναι το περιγεγραμμένο ορθογώνιο (bounding rectangle) της περιοχής, για το οποίο υπάρχει συνάρτηση εύρεσής του με βάση το διάνυσμα των σημείων που την αποτελούν. Μάλιστα, παρέχεται η δυνατότητα της εύρεσης του περιγεγραμμένου ορθογωνίου μικρότερου εμβαδού, το οποίο συνήθως είναι περιστραμμένο, αλλά και του μικρότερου περιγεγραμμένου ορθογωνίου με πλευρές κάθετες στους άξονες, δηλαδή οριζοντίου ή καθέτου.

Όμως και αυτή η πληροφορία δεν αρκεί από μόνη της, αφού χρειάζονται δεδομένα όπως η γωνία περιστροφής του ορθογωνίου ή η σχέση πλάτους προς ύψος, οι οποίες θα χρησιμοποιηθούν κατ’ επανάληψη στη διαδικασία απομόνωσης της σωστής περιοχής, και οι οποίες δεν παρέχονται αμέσως μέσω κάποιας έτοιμης συνάρτησης. Για αυτόν το λόγο δημιουργήθηκε η κλάση (δομή class της C++) με το όνομα regionProps, στην οποία εντάχθηκαν όλες οι πληροφορίες που χρειάστηκαν σχετικά με τις περιοχές, καθώς και οι συναρτήσεις υπολογισμού και ανάκτησής τους. Συνολικά δημιουργήθηκαν 15 συναρτήσεις-μέλη (member functions), οι οποίες υπολογίζουν και επεξεργάζονται οχτώ βασικά μεγέθη.

Η εύρεση των περιεχόμενων περιοχών γίνεται με χρήση της συνάρτησης contains της OpenCV, η οποία είναι συνάρτηση-μέλος της κλάσης των περιγεγραμμένων ορθογωνίων. Εφαρμόζεται δηλαδή στο περιγεγραμμένο ορθογώνιο της περιοχής και δέχεται ως όρισμα ένα σημείο μιας περιοχής MSER-, επιστρέφοντας τιμή αληθείας true αν το σημείο αυτό βρίσκεται εντός του ορθογωνίου και false αν είναι εκτός του.