



ΤΕΙ ΠΕΙΡΑΙΑ

ΣΧΟΛΗ ΤΕΧΝΟΛΟΓΙΚΩΝ ΕΦΑΡΜΟΓΩΝ

ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΗΛΕΚΤΡΟΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΙΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

Οπτική αναγνώριση χαρακτήρων με συσκευή Android



Σταμάτιος Αρμένης

Εισηγητής: Ιωάννης Έλληνας

ΑΙΓΑΛΕΩ

ΙΑΝΟΥΑΡΙΟΣ 2014



ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

Οπτική αναγνώριση χαρακτήρων με συσκευή Android

Σταμάτιος Αρμένης

A.M. 39201

Εισηγητής: Ιωάννης Έλληνας

ΠΕΡΙΕΧΟΜΕΝΑ

ΕΙΣΑΓΩΓΗ.....	8
ΚΕΦΑΛΑΙΟ 1 - ΕΠΕΞΕΡΓΑΣΙΑ ΕΙΚΟΝΑΣ	
Εισαγωγή.....	9
1.1 Ψηφιακή Εικόνα.....	10
1.2 Απομωσαϊκοποίηση (Demosaicing).....	12
1.3 Χρωματικά Μοντέλα.....	14
1.4 Είδη Εικόνων.....	17
1.5 Βάθος Χρώματος.....	22
1.6 Ανάλυση Εικόνας.....	24
1.7 Μέγεθος αρχείων ψηφιακών εικόνων.....	26
1.8 Συμπύεση εικόνων.....	26
1.9 Ιστόγραμμα.....	29
1.10 Θόρυβος.....	31
1.10.1 Ομοιόμορφος Θόρυβος.....	32
1.10.2 Κανονικός Θόρυβος.....	34
1.10.3 Σημειακός Θόρυβος ή Αλατιού και Πιπεριού.....	35
1.11 Φίλτρα απολοιφής θορύβου.....	37
1.11.1 Παράδειγμα Εφαρμογής Χωρικού Φίλτρου.....	37
1.11.2 Φίλτρο Αριθμητικής Μέσης Τιμής.....	39
1.11.3 Φίλτρο Γεωμετρικής Μέσης Τιμής.....	41
1.11.4 Φίλτρο Μεσαίας Τιμής.....	43
1.11.5 Αμφίπλευρο Φίλτρο.....	45
1.12 – Δυαδική Μετατροπή.....	47
1.12.1 – Καθολική Κατωφλίωση.....	51
1.12.2 – Προσαρμοσμένη Κατωφλίωση.....	57
1.13 Μορφολογία Εικόνας.....	61
1.13.1 Επέκταση.....	61
1.13.2 Συρρίκνωση.....	63
1.13.3 Συνδυασμοί Μορφολογικών Χειρισμών.....	64

ΚΕΦΑΛΑΙΟ 2 - ΟΠΤΙΚΗ ΑΝΑΓΝΩΡΙΣΗ ΧΑΡΑΚΤΗΡΩΝ

Εισαγωγή.....	66
2.1 Ιστορία.....	67
2.2 Η Σημασία της για τους Τυφλούς.....	70
2.3 Η Σημασία της για τους Οργανισμούς και τις Επιχειρήσεις.....	71
2.4 Είδη Εφαρμογών και οι Δυνατότητες τους.....	73
2.5 Αναγνώριση με Τεχνητά Νευρωνικά Δίκτυα.....	76
2.6 Αναγνώριση με Στατιστικές Παρατηρήσεις.....	81

ΚΕΦΑΛΑΙΟ 3 - ΕΦΑΡΜΟΓΗ

Εισαγωγή.....	84
3.1 Εγκατάσταση Λογισμικού.....	84
3.2 Παρουσίαση της Εφαρμογής.....	91
3.2.1 Διαδικασία Επεξεργασίας.....	93

ΒΙΒΛΙΟΓΡΑΦΙΑ.....	101
--------------------------	------------

ΕΙΣΑΓΩΓΗ

Σκοπός της εργασίας αυτής είναι η ανάπτυξη μιας εφαρμογής αναγνώρισης χαρακτήρων για συσκευές Android. Συγκεκριμένα θα πρέπει να μπορεί να αναγνωρίζει απο κάποια φωτογραφία τον αριθμό που αναγράφεται στα ρολόγια της ΔΕΗ και να υπολογίζει το ποσό που θα πρέπει να καταβληθεί στην ΔΕΗ.

Αρχικά θα αναφερθούμε στο πως βγαίνει μια ψηφιακή φωτογραφία, δηλαδή τι συμβαίνει όταν εμεις πατάμε το κουμπί της ψηφιακής μηχανής, θα δούμε τους διάφορους τύπους συμπίεσης μιας εικόνας και φυσικά τους τύπους εικόνας. Στην συνέχεια θα πούμε μερικά προβλήματα που παρουσιάζονται στις ψηφιακές φωτογραφίες, όπως ο θόρυβος. Θα αναλύσουμε τους τύπους θορύβου και πως αυτοι μπορούν να προκληθούν. Επίσης θα δούμε πώς μπορούμε να εξαλείψουμε τον θόρυβο με τα φίλτρα, τα οποία και θα δούμε αναλυτικά την λειτουργία τους. Στο τέλος του πρώτου κεφαλαίου θα δούμε διάφορες τεχνικές δυαδικοποιήσεις μιας εικόνας για να μπορέσουμε στην συνέχεια να εξαγάουμε διάφορες πληροφορίες.

Στο δεύτερο κεφάλαιο θα πούμε μερικά πράγματα για την ιστορία της οπτικής αναγνώρισης χαρακτήρων, θα δούμε την σημασία και την βοήθεια που μπορεί να προσφέρει σε άτομα με προβλήματα όρασης όπως επίσης και σε πολλές εταιρίες και οργανισμούς μέσω των διαφόρων εφαρμογών. Τέλος θα δούμε δύο βασικούς τρόπος αναγνώρισης χαρακτήρων που χρησιμοποιούνται σήμερα.

Στο τρίτο και τελευταίο κεφάλαιο υπάρχει πλήρης οδηγός της εγκατάστασης των διαφόρων εργαλείων που χρησιμοποιήθηκαν για την υλοποίηση αυτής της εφαρμογής και βεβαίως η επεξήγηση ολόκληρης της διαδικασίας αναγνώρισης των χαρακτήρων βασισμένη σε όσα έχουν προαναφερθεί στα προηγούμενα κεφάλαια.

ΚΕΦΑΛΑΙΟ 1

ΕΠΕΞΕΡΓΑΣΙΑ ΕΙΚΟΝΑΣ

Εισαγωγή

Οι αισθητήριες συσκευές τείνουν να υποβαθμίσουν την ποιότητα των ψηφιακών εικόνων εισάγοντας θόρυβο, γεωμετρική παραμόρφωση, ή θάμπωμα εξ' αιτίας της κίνησης ή της λανθασμένης εστίασης της κάμερας. Ένας από τους κύριους σκοπούς της ψηφιακής επεξεργασίας εικόνας είναι να βελτιώσει την ποιότητα της εικόνας και να μετριάσει την υποβάθμιση που εισάγεται από τις αισθητήριες συσκευές και τις συσκευές καταγραφής της εικόνας. Οι τεχνικές αποκατάστασης εικόνας αφορούν κυρίως την ανάκτηση εικόνας που έχει παραμορφωθεί. Οι τεχνικές βελτίωσης της ποιότητας ψηφιακής εικόνας βελτιώνουν την υποκειμενική ποιότητα της εικόνας τονίζοντας ορισμένα χαρακτηριστικά της και μειώνοντας το θόρυβο.

Για να επιτευχθεί οπτική αναγνώριση χαρακτήρων σε μία εικόνα η οποία μπορεί να είναι ένα έγγραφο που έχουμε σαρώσει, η φωτογραφία μιάς πινακίδας κυκλοφορίας και στην περίπτωση μας μία φωτογραφία ενός ρολογιού της ΔΕΗ θα πρέπει να υποστεί κάποια επεξεργασία έτσι ώστε να μπορέσουμε να εξάγουμε την απαραίτητη πληροφορία. Επίσης υπάρχει η αναγνώριση αποτυπωμάτων (fingerprint recognition) και η αναγνώριση προσώπων (facial recognition).

Στην συνέχεια θα δούμε τι είναι ουσιαστικά μια εικόνα, θα μελετήσουμε τα χαρακτηριστικά και τις ιδιότητες των ψηφιακών φίλτρων καθώς επίσης θα περιγράψουμε ορισμένες τεχνικές αποκατάστασης και βελτίωσης της ποιότητας εικόνας και θα παρουσιάσουμε τα είδη θορύβου και τα κριτήρια επιλογής ψηφιακού φίλτρου. Τέλος θα εξετάσουμε τις διάφορες μεθόδους δυαδικοποίησης μιας εικόνας και θα δούμε βασικούς μορφολογικούς χειρισμούς των δυαδικών εικόνων.

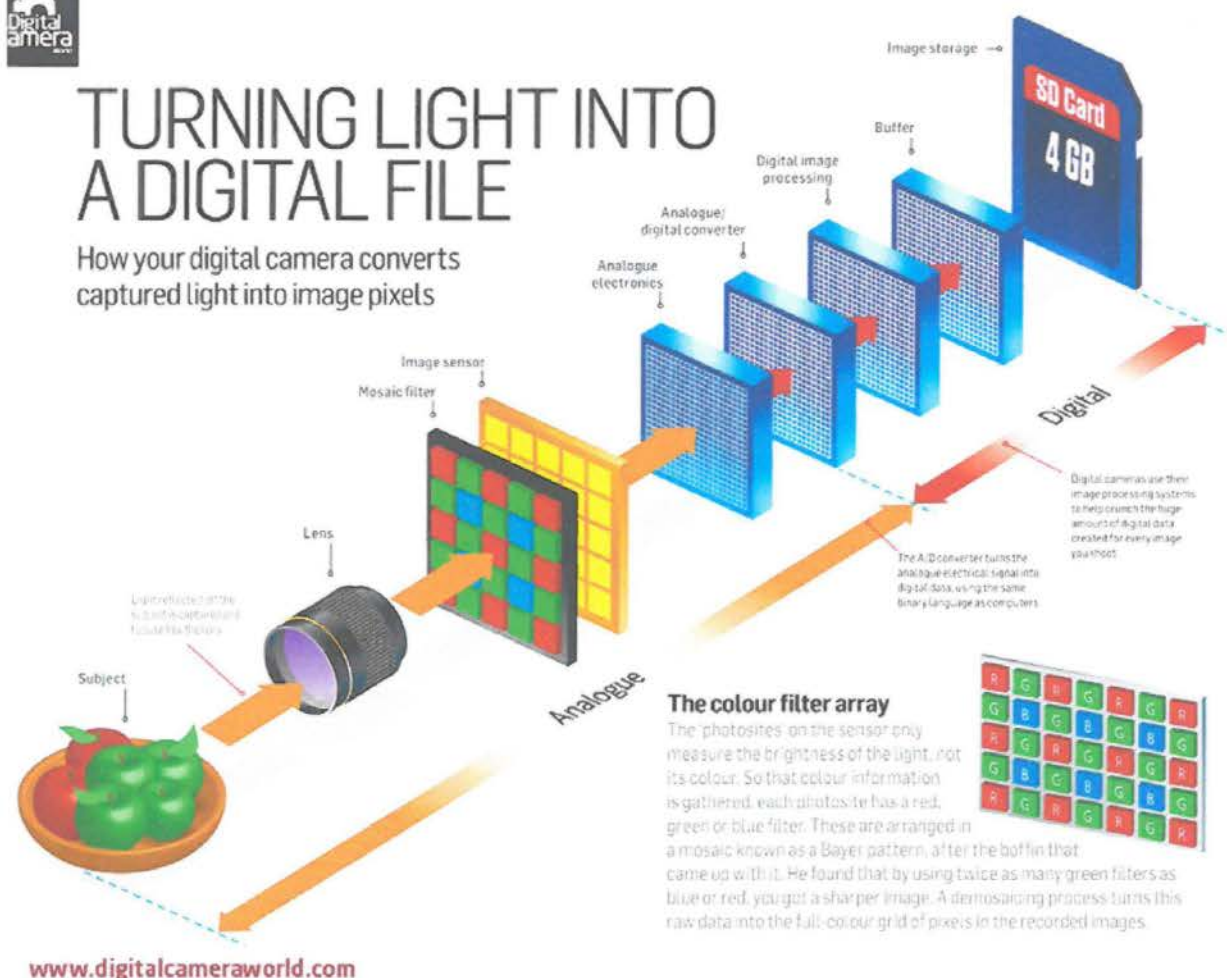
1.1 Ψηφιακή Εικόνα

Το αισθητήριο, στην προκειμένη περίπτωση η κάμερα του κινητού τηλεφώνου, το οποίο να ένα αναλογικό μέσο μετατρέπει το φως σε ηλεκτρικό ρεύμα το οποίο εγματοληπτείται, κβαντίζεται και ψηφιοποιείται μέσω ενός μετατροπέα Α/Ψ (Αναλογικό σε Ήφιακό – Analog to Digital Converter). Στην συνέχεια γίνεται επεξεργασία της εικόνας πτωετικά για να μειωθεί ο τεράστιος όγκος της ψηφιακής πληροφορίας που ημιουργήθηκε.



TURNING LIGHT INTO A DIGITAL FILE

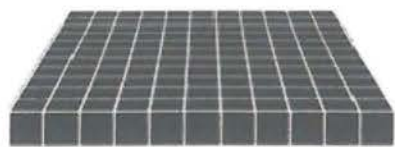
How your digital camera converts captured light into image pixels



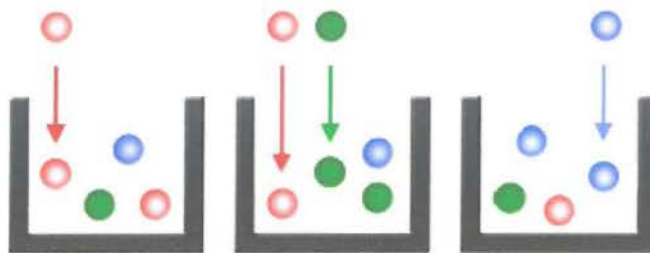
Εικόνα 1.1.1 – Λήψη Ψηφιακής Φωτογραφίας

αισθητήρας αποτελείται από εκατομμύρια φωτοευαίσθητες μονάδες (Light Cavities or 'photosites') – (Εικόνα 1.1.2) όπου η κάθε μία δημιουργεί το δικό της ηλεκτρικό σήμα

ανάλογα με την φωτεινότητα. Η κάθε μία ξεχωριστή μονάδα δεν βλέπει χρώματα παρα μόνο φωτεινότητες.

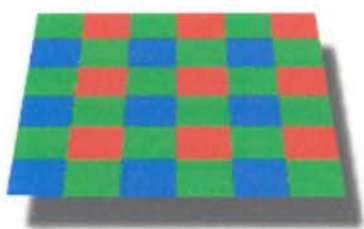


Εικόνα 1.1.2 - Φωτοευαίσθητος Πίνακας

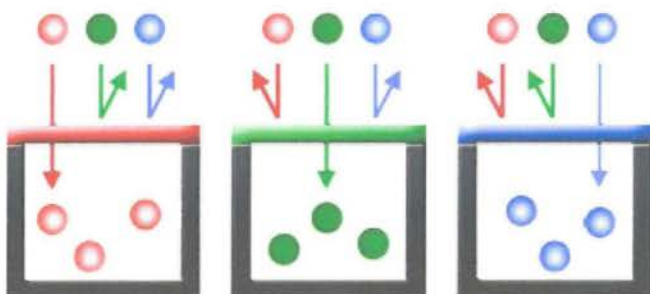


Εικόνα 1.1.3 - Φωτοευαίσθητες Μονάδες

Με αυτόν τον τρόπο μπορούμε να παράγουμε μόνο ασπρόμαυρες εικόνες (Grayscale) διότι η μονάδες δεν μπορούν να ξεχωρίσουν τι ποσότητα απο το κάθε χρώμα περιέχουν. Για να παράξουμε έγχρωμες εικόνες πρέπει να βάλουμε ένα φίλτρο πάνω απο κάθε μονάδα που θα επιτρέψει την είσοδο σε ένα μονο χρώμα. Για να έχουμε πλήρες χρώμα σε κάθε εικονοστοιχείο (Picture Element - pixel) η κάμερα θα πρέπει να προσδιορίσει τα άλλα δύο χρώματα. Το ποιό γνωστό φίλτρο χρώματος είναι το Bayer Array(Εικόνα 1.1.4).



Εικόνα 1.1.4 – Φίλτρο Χρώματος



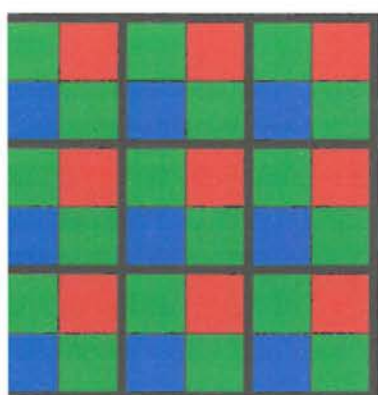
Εικόνα 1.1.5 - Φωτοευαίσθητες Μονάδες με Φίλτρο

Ένα Bayer Array αποτελείται εναλλασσόμενες σειρές κόκκινα-πράσινα και πράσινα-μπλέ φίλτρα. Παρατηρούμε ότι ο πίνακας περιέχει δύο φορές περισσότερους πράσινους παρά μπλέ και κόκκινους αισθητήρες, αυτό συμβαίνει διότι το ανθρώπινο μάτι είναι πιο ευαίσθητο στο πράσινο φως. Ως αποτέλεσμα η εικόνα που δημιουργείται έχει λιγότερο θόρυβο και περισσότερη λεπτομέρεια απ' ότι εάν συμπεριφερόμασταν σε κάθε χρώμα ισάξια. Για να δημιουργήσουμε την εικόνα πρέπει να ακολουθήσουμε μια μέθοδο η οποία ονομάζεται Απομωσαϊκοποίηση (Demosaicing). Ο αλγόριθμος αυτός πραγματοποιεί τον διαχωρισμό των τριών χρωματικών συνιστωσών απο την εικόνα υπέρθεσης που

πτέγραψε ο αισθητήρας. Επιπλέον, συνήθως πραγματοποιούνται ειδικές βελτιώσεις, όπως αποθορύβωση (Denoising), ενώ υλοποιείται και ο αλγόριθμος συμπίεσης.

2 Απομωσαϊκοποίηση (Demosaicing)

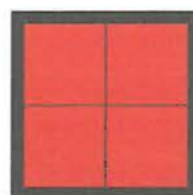
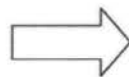
Πιο εύκολος και γρήγορος τρόπος είναι αυτός της παρεμβολής του κοντινότερου γείτονα (Nearest-Neighbour Interpolation) όπου προσπαθούμε να φτιάξουμε τρεις διαφορετικούς πίνακες ίδιου μέγεθος με αυτόν του Bayer που θα περιέχουν μόνο ένα χρώμα (Εικόνα 1.2.4). Από το φίλτρο Bayer (Εικόνα 1.2.1) εξάγουμε μία γειτονία 2x2 (Εικόνα 1.2.2) τώρα ως παράδειγμα θα πρέπει όποιο pixel δεν είναι κόκκινο να πάρει την τιμή του κοντινότερου κόκκινου και θα έχουμε μία κόκκινη γειτονία 2x2 (Εικόνα 1.2.3), στην περίπτωση του πράσινου που έχουμε δύο πράσινα pixels ακολουθούμε είτε τον οριζόντιο είτε τον διαγώνιο τρόπο. Στο τέλος οι τρεις διαφορετικοί πίνακες θα ενωθούν για να μας δώσουν την τελική μας εικόνα.



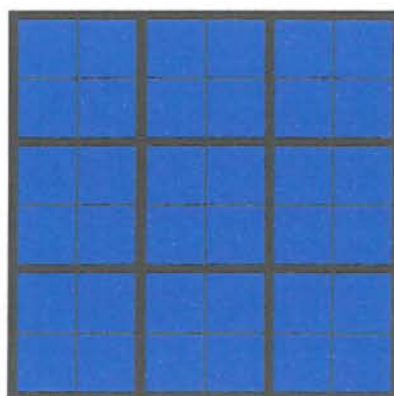
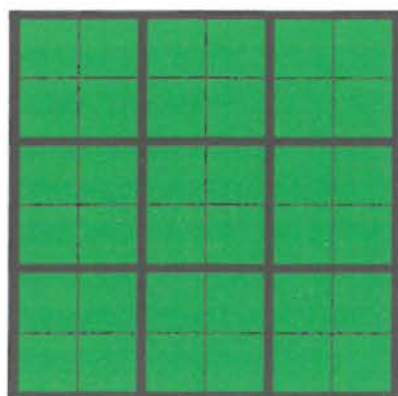
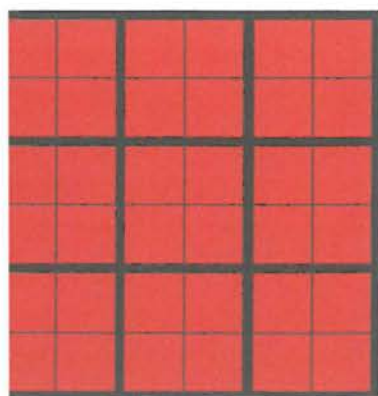
Εικόνα 1.2.1 – Bayer Filter



Εικόνα 1.2.2 – Γειτονία 2x2



Εικόνα 1.2.3



Εικόνα 1.2.4 - Πίνακες RGB

Η συγκεκριμένη μέθοδος εμφανίζει πολλά χρωματικά προβλήματα (False Color) ειδικά στις ακμές και δεν ενδείκνυται για εφαρμογές που είναι αναγκαία η ποιότητα. Βέβαια επειδή δεν γίνονται υπολογισμοί, αυτή η μέθοδος είναι ιδανική σε εφαρμογές όπου η ταχύτητα είναι απαραίτητη. Στις παρακάτω εικόνες βλέπουμε την αρχική μας (Εικόνα 1.2.5) και δίπλα την εικόνα με χρωματικά προβλήματα (Εικόνα 1.2.6), κυρίως πάνω στον φράχτη.



Εικόνα 1.2.5 – Αρχική



Εικόνα 1.2.6 – False Color

Ένας άλλος απλός τρόπος απομωσαϊκοποίησης είναι αυτός της διγραμμικής παρεμβολής (Bilinear Interpolation), όπου η τιμή του κόκκινου ενός pixel άλλου χρώματος υπολογίζεται από τον μέσο όρο των δύο ή τεσσάρων γειτονικών κόκκινων pixel, παρομοίως για το μπλέ και πράσινο. Υπάρχουν και πιο πολύπλοκοι μέθοδοι οι οποίες παρεμβάλουν μεμονωμένα σε κάθε χρωματικό πεδίο, όπως αυτός της δικυβικής παρεμβολής (Bicubic Interpolation). Όλες οι παραπάνω μέθοδοι έχουν καλά αποτελέσματα σε ομοιογενείς περιοχές μιας εικόνας, αλλά έχουν την τάση να κάνουν "λάθη" σε περιοχές με ακμές, όπως είδαμε και παραπάνω.

Φυσικά υπάρχουν και ποιο εξεζητημένοι αλγόριθμοι απομωσαϊκοποίησης που εκμεταλλεύονται το χωρικό (Spatial) και/ή φασματικό (Spectral) συσχετισμό των pixel μέσα σε μία έγχρωμη εικόνα. Χωρικός συσχετισμός (Spatial correlation) είναι η τάση των pixel να

ουν παρόμοιες χρωματικές τιμές μέσα σε μία μικρή ομοιογενή περιοχή της εικόνας. Ο χρωματικός συσχετισμός (Spectral correlation) είναι η εξάρτηση ανάμεσα στις τιμές των κει διαφορετικών χρωματικών πεδίων μέσα σε μια μικρή περιοχή της εικόνας.

3 Χρωματικά Μοντέλα

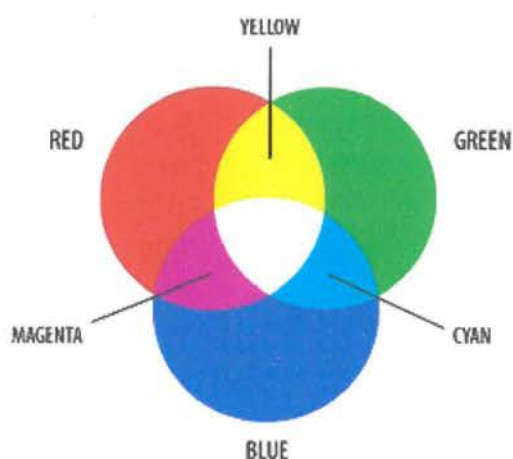
σιαστικά είναι η μετατροπή των χρωμάτων σε αριθμούς με μαθηματικό τύπο. Υπάρχουν πολλοί τρόποι να γίνει αυτό, και γ'αυτο υπάρχουν αρκετά χρωματικά μοντέλα και το θέμα είναι χρήσιμο με διαφορετικό τρόπο.

Προσθετικό Μοντέλο (RGB)



Λέγεται προσθετικό διότι προσθετοντας το φως απο τα τρία βασικά χρώματα κόκκινο (red), πράσινο (green) και μπλέ (blue) μπορούμε να δημιουργήσουμε όλα τα υπόλοιπα. Η ανάμειξη και των τριών χρωμάτων στην πλήρη ένταση τους δημιουργεί το λευκό χρώμα/φως. Χρησιμοποιείται

εις ψηφιακές φωτογραφικές μηχανές η οποίες παράγουν RGB εικόνες και σε όλες τις όνες.



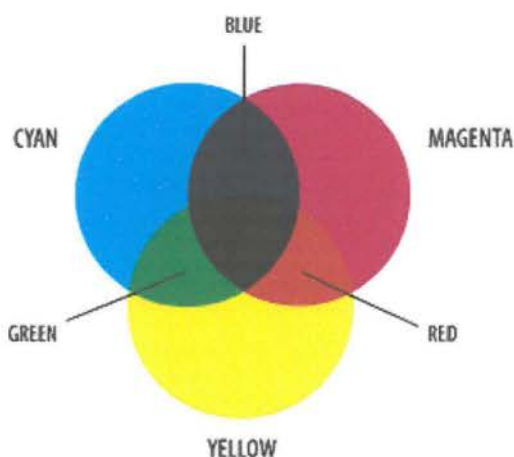
Εικόνα 1.3.1 – RGB

Αφαιρετικό Μοντέλο (CMY)



Αυτο το μοντέλο χρησιμοποιείται εκτεταμένα στον χώρο των εκτυπώσεων. Διότι το χαρτί αποδίδει ανακλόμενο φως. Έχοντας σαν βάση το λευκό χαρτί (το λευκό περιέχει όλο το χρωματικό φάσμα) όταν προστίθεται ένα χρώμα στο λευκό, αυτό που στην ουσία συμβαίνει είναι ότι αφαιρούνται

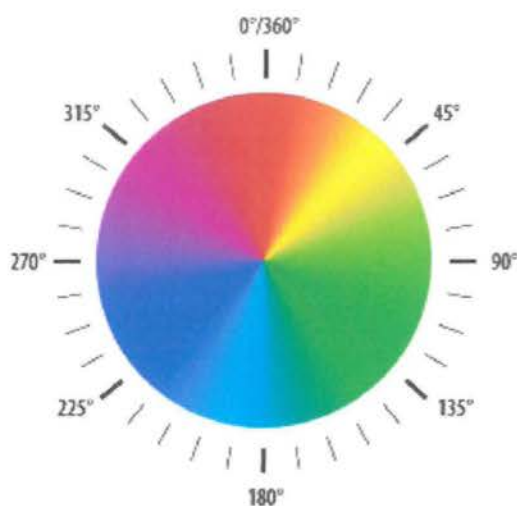
κάποια μήκη κύματος οπότε γίνεται ορατό το επιθυμητό χρώμα. Αν προσέξουμε την παρακάτω εικόνα θα δούμε ότι αν αναμίξουμε τα τρία δεν παίρνουμε το μαύρο χρώμα, αλλά ένα σκούρο καφέ. Γι' αυτό τον λόγο χρησιμοποιούμε το μαύρο ως τέταρτο χρώμα για να αποδίδουμε καλύτερα τις μαύρες περιοχές στις εκτυπώσεις μας. Και το όνομα του μοντέλου αλλάζει σε CMYK όπου C=Cyan (Κυανό) , M=Magenta (Πορφυρό), Y=Yellow (Κίτρινο) και K=Key (Μαύρο) αντί για B=Black να μην μπερδεύεται με το B=Blue στο RGB μοντέλο.



Εικόνα 1.3.2 – CMY

Μοντέλο HSB/HSL

Το όνομα βγαίνει απο τα αρχικά των Απόχρωση (Hue), Ένταση (Saturation) και Φωτεινότητα (Brightness ή Lightness). Η απόχρωση εκφράζεται σαν αριθμός απο το 0 έως τις 360 μοίρες ξεκινώντας και τελειώνοντας στο κόκκινο και διατρέχει το πράσινο, το μπλέ και όλα τα ενδιάμεσα χρώματα, όπως το πορτοκαλί, το μώβ, κτλ.



Εικόνα 1.3.3 – Χρωματικός Κύκλος

ένταση μας δείχνει τον βαθμό στον οποίο η απόχρωση διαφέρει από ένα ουδέτερο γκρι. Τιμές ξεκινούν από το 0%, δηλαδή καθόλου χρωματική ένταση, μέχρι 100%, δηλαδή την γνήσια ένταση μίας απόχρωσης.



Εικόνα 1.3.4 – Χρωματική Ένταση

Φωτεινότητα μας δείχνει το επίπεδο του φωτισμού. Οι τιμές ξεκινούν από το 0%, όπου γίνεται μαύρο (καθόλου φως), μέχρι 100%, όπου φαίνεται άσπρο (πλήρης φωτισμός). Τα χρώματα σε ποσοστά μικρότερα του 50% φαίνονται πιο σκούρα (σκοτεινά), ενώ σε ποσοστά μεγαλύτερα του 50% φαίνονται πιο ανοιχτά (φωτεινά).

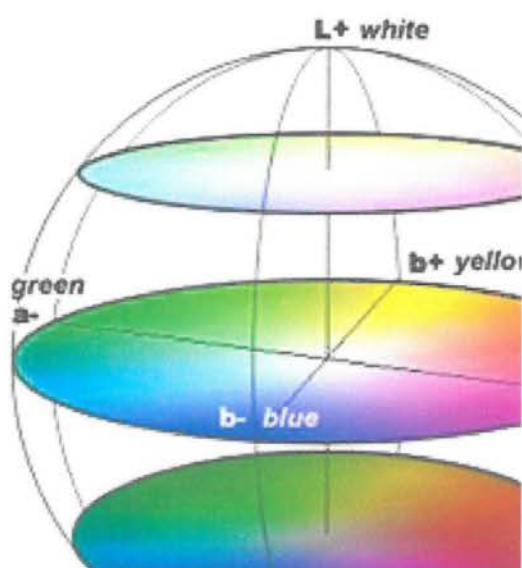


Εικόνα 1.3.5 - Φωτεινότητα

Μοντέλο CIELAB

Στις αρχές του 20^{ου} αιώνα ένας διεθνής οργανισμός, ο Commission International de L'Eclairage (CIE), δημιούργησε ένα διεθνές πρότυπο με στόχο να μπορεί να μετρηθεί κάθε χρώμα του ορατού φάσματος με βάση την αντιληπτική ικανότητα του ανθρώπινου ματιού και όχι με πιθανά μέσα στα οποία απεικονίζεται ή εκτυπώνεται. Το 1976 έγινε αναθεώρηση του CIE. Το χρωματικό μοντέλο που προέκυψε με βάση αυτή είναι το CIE $L^*a^*b^*$.

Στηρίζεται στην αρχή ότι ένα χρώμα δεν μπορεί να είναι συγχρόνως πράσινο και κόκκινο ή κίτρινο και μπλέ. Έχει τρεις μεταβλητές μία για φωτεινότητα (Lightness) και δύο για χρωματικές τιμές. Έτσι το L^* συμβολίζει την φωτεινότητα, το a^* την τιμή από πράσινο μέχρι κόκκινο και το b^* από κίτρινο έως μπλέ.

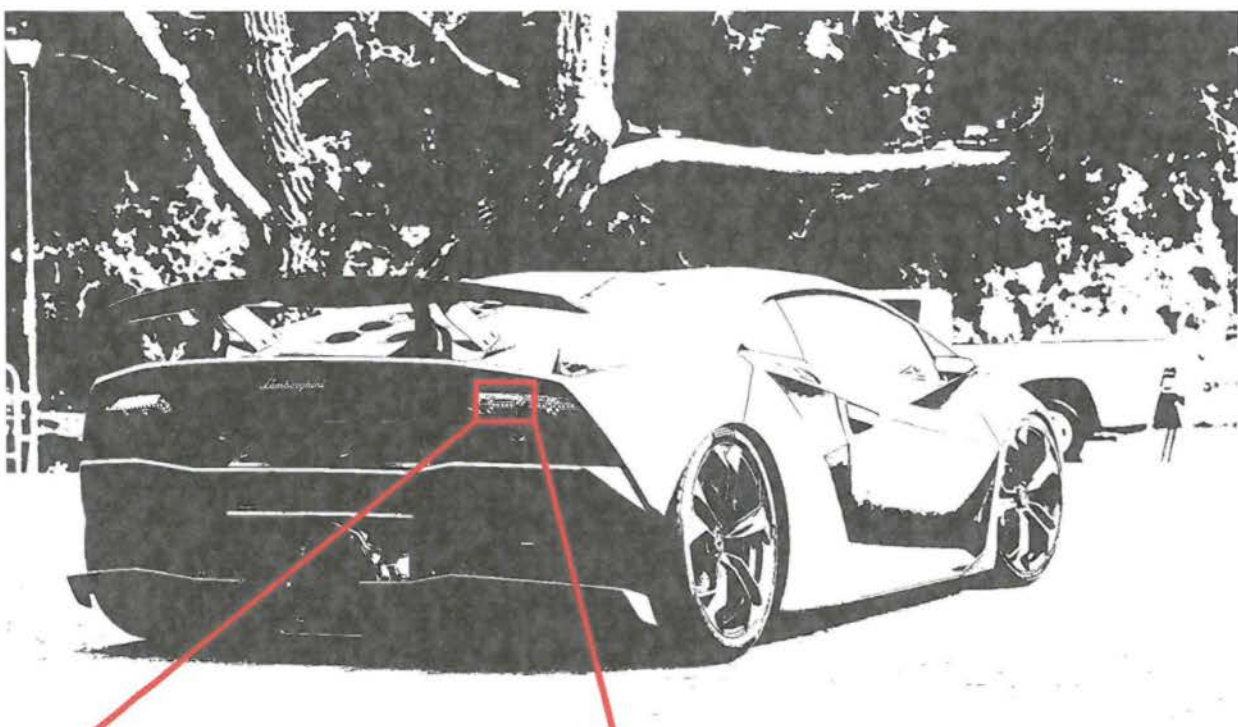


Εικόνα 1.3.6 - Μοντέλο CIELAB

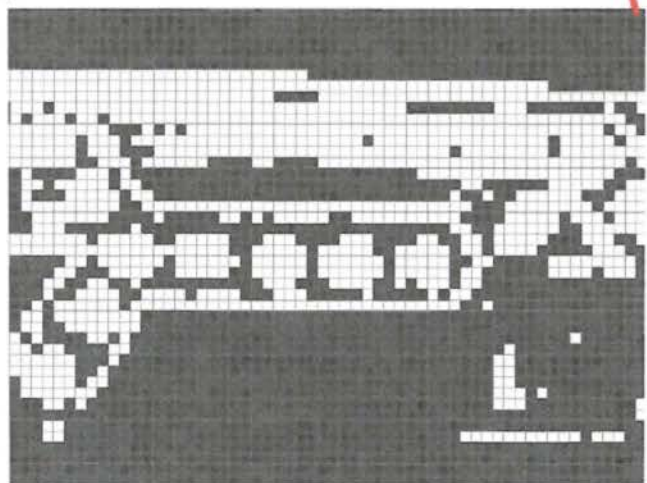
1.4 Είδη Εικόνων (Image Types)

Δυαδικές (binary) εικόνες

Σε μια δυαδική εικόνα, κάθε ρικελ μπορεί να πάρει μια από τις δύο διακριτές τιμές (0 ή 1). Η τιμή 0 αντιστοιχεί στο μαύρο και η τιμή 1 αντιστοιχεί στο λευκό. Μια δυαδική εικόνα αποθηκεύεται ως ένας δισδιάστατος πίνακας μηδενικών και μονάδων (Εικόνα 1.4.3). Μια δυαδική εικόνα μπορεί να θεωρηθεί ως ένα ειδικό είδος ασπρόμαυρης εικόνας, η οποία όμως περιέχει μόνο μαύρο και άσπρο.



Εικόνα 1.4.1 – Διαδική Εικόνα



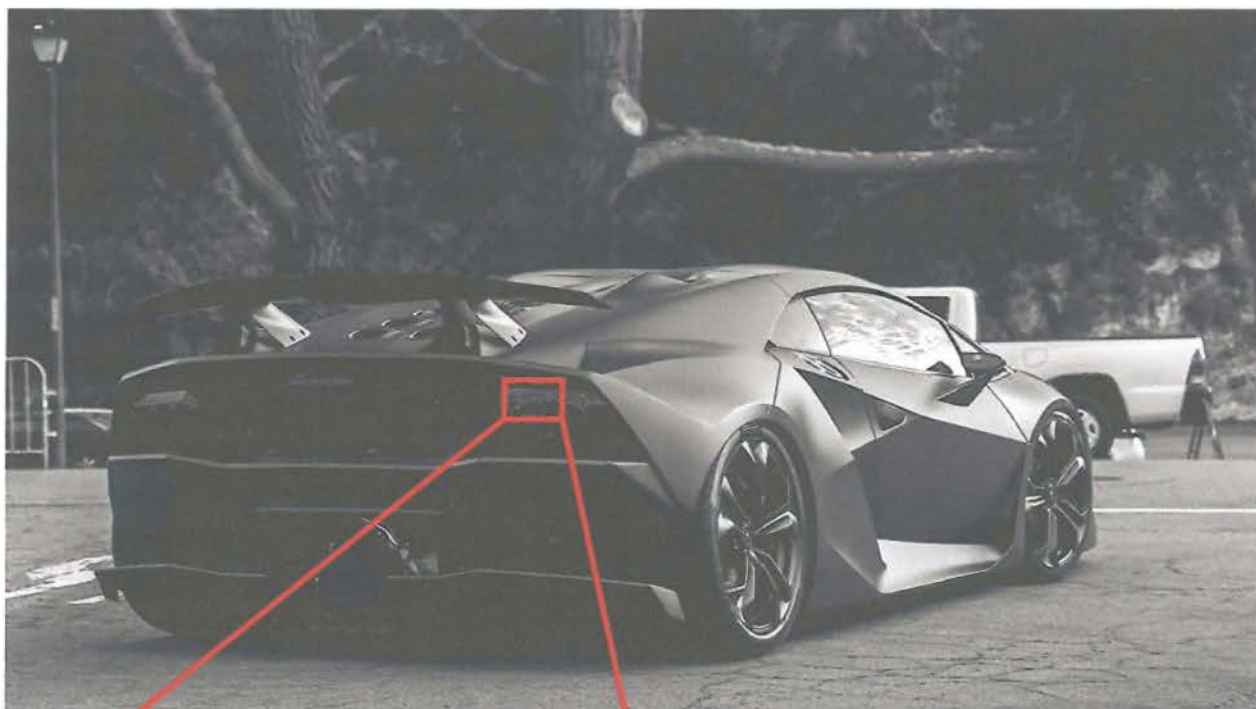
Εικόνα 1.4.2 – Εστιασμένη (Zoomed)

0	0	1	1	1
1	1	1	1	1
1	0	1	1	1
0	0	1	1	1
0	0	0	0	0

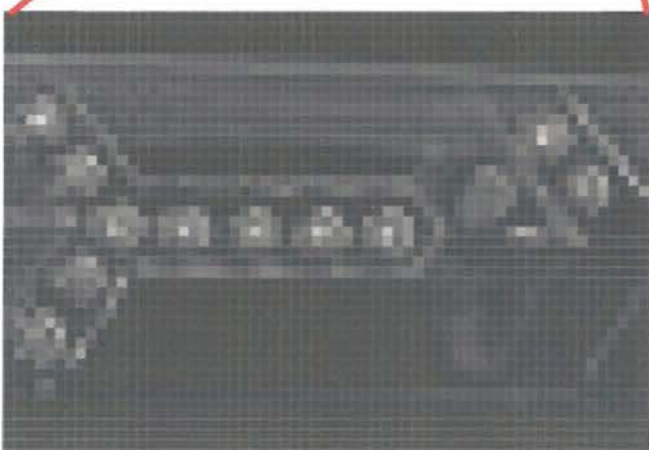
Εικόνα 1.4.3 – Δισδιάστατος Πίνακας

Ασπρόμαυρες (Grayscale) Εικόνες

Μια ασπρόμαυρη εικόνα (εικόνα έντασης) είναι ένας πίνακας δεδομένων του οποίου οι τιμές αναπαριστούν την ένταση της φωτεινότητας. Μια εικόνα έντασης, είναι ένας δισδιάστατος πίνακας (Εικόνα 1.4.6), του οποίου κάθε στοιχείο αναφέρεται σε ένα και μοναδικό pixel. Τα στοιχεία του πίνακα αναπαριστούν διάφορες εντάσεις του φωτός (επίπεδα του γκρι), όπου η τιμή 0 αναπαριστά το μαύρο και η τιμή 255 αναπαριστά το λευκό.



Εικόνα 1.4.4 - Ασπρόμαυρη Εικόνα



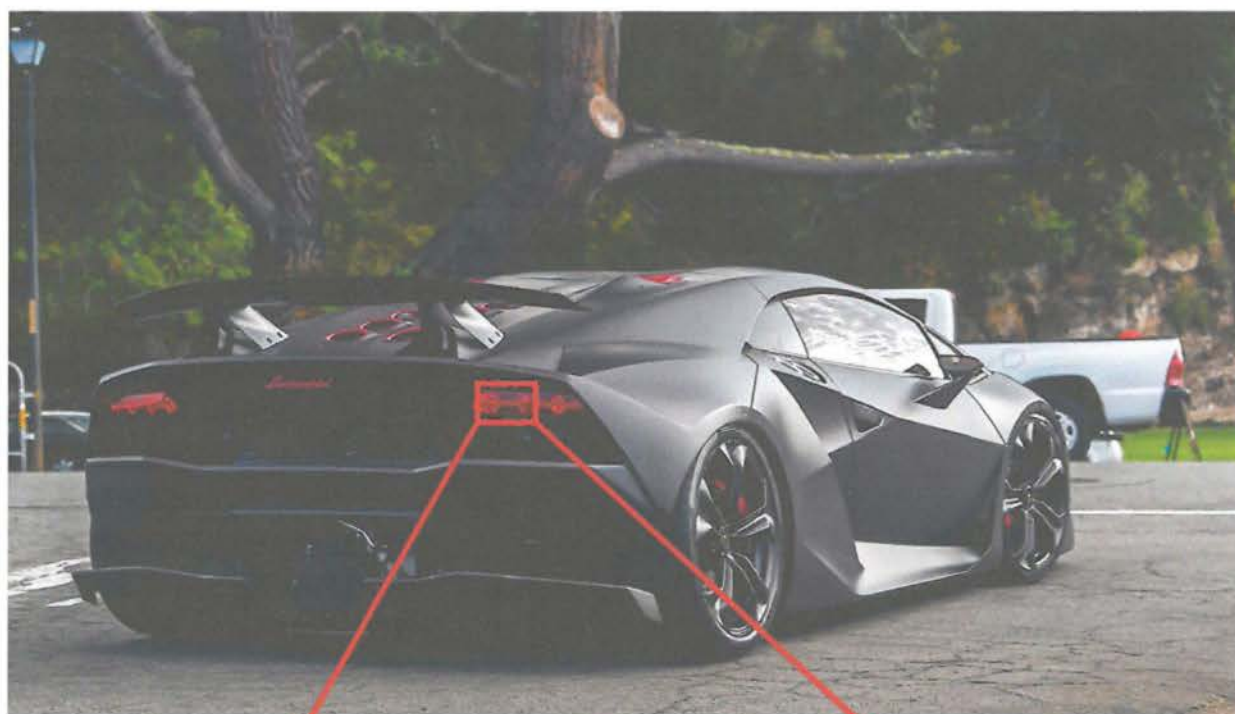
Εικόνα 1.4.5 – Εστιασμένη

44	32	34	33	40
12	21	16	32	36
22	64	88	94	124
73	117	109	122	123
49	94	104	176	108

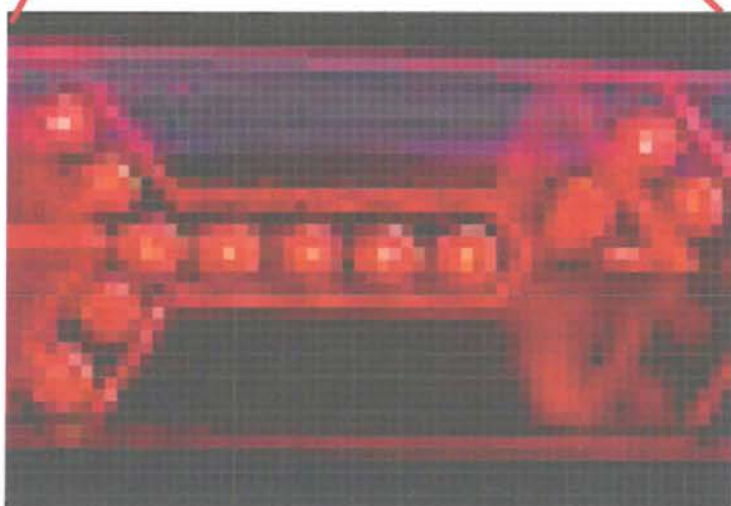
Εικόνα 1.4.6 – Δισδιάστατος Πίνακας

Έγχρωμες Εικόνες

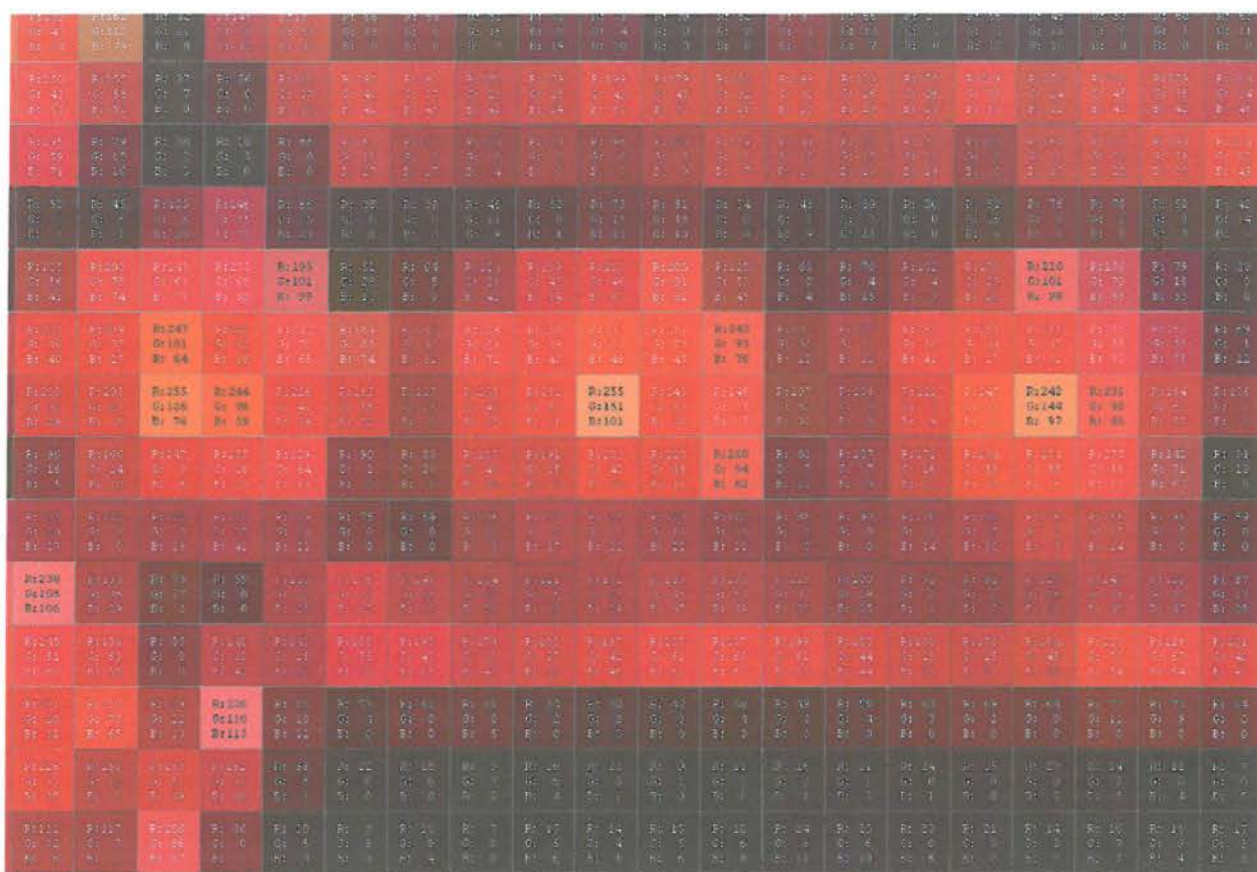
Οι έγχρωμες εικόνες αποτελούνται από 3 διαφορετικούς πίνακες ίδιων διαστάσεων, όπου ο κάθε πίνακας αντιπροσωπεύει ένα από τα τρία βασικά χρώματα. Κάθε pixel αποτελείται από 3 συνιστώσες (κόκκινη, πράσινη, μπλε), και η τιμή κάθε μιας από αυτές περιγράφει την «ποσότητα» (ένταση) του κόκκινου, πράσινου και μπλε αντίστοιχα που χρησιμοποιείται στη δημιουργία του τελικού χρώματος.



Εικόνα 1.4.7 – Έγχρωμη Εικόνα



Εικόνα 1.4.8 – Εστιασμένη (Zoom)



Εικόνα 1.4.9 – Εστιασμένη με RGB τιμές

Όπως είπαμε και παραπάνω κάθε ριχελ περιέχει τρεις τιμές. Οι τιμές αυτές είναι οι φωτεινότητες (ένταση) των βασικών χρωμάτων, έτσι ώστε να δημιουργηθεί του χρώμα του συγκεκριμένου ριχελ. Για να γίνει πιο κατανοητό και πιο ευδιάκριτο έχουμε απομονώσει ένα κομμάτι τις παραπάνω εικόνας (Εικόνα 1.4.10).

R:130 G: 6 B: 17	R:104 G: 0 B: 4	R:114 G: 0 B: 0	R: 98 G: 6 B: 0	R:123 G: 5 B: 0
R: 39 G: 0 B: 0	R: 46 G: 11 B: 9	R: 52 G: 0 B: 1	R: 73 G: 15 B: 13	R: 81 G: 18 B: 13
R: 64 G: 5 B: 0	R:124 G: 38 B: 41	R:189 G: 43 B: 54	R:203 G: 46 B: 53	R:205 G: 91 B: 81
R:142 G: 46 B: 32	R:229 G: 69 B: 71	R:253 G: 49 B: 40	R:247 G: 80 B: 48	R:251 G: 73 B: 45
R:113 G: 25 B: 3	R:205 G: 49 B: 37	R:232 G: 53 B: 31	R:255 G:151 B:101	R:249 G: 57 B: 26

Εικόνα 1.4.10 – Δισδιάστατος Πίνακας

ώρα για παράδειγμα εάν μηδενίσουμε τις τιμές των άλλων δύο βασικών χρωμάτων θα ούμε ότι η εικόνα μας θα έχει μόνο κόκκινες, πράσινες ή μπλέ αποχρώσεις. Για να το ούμε καλύτερα πέρνουμε το ίδιο κομμάτι (Εικόνα 1.4.10) και απεικονίζουμε τις τρεις υνιστώσεις, μία για κάθε χρώμα (Εικόνα 1.4.11).

			R:130 G: 0 B: 0	R:104 G: 0 B: 0	R:114 G: 0 B: 0	R: 98 G: 0 B: 0	R:123 G: 0 B: 0
	R: 0 G: 5 B: 0	R: 39 G: 0 B: 0	R: 46 G: 0 B: 0	R: 52 G: 0 B: 0	R: 73 G: 0 B: 0	R: 81 G: 0 B: 0	
R: 0 G: 0 B: 17	R: 0 G: 0 B: 0	R: 64 G: 0 B: 0	R:124 G: 0 B: 0	R:189 G: 0 B: 0	R:203 G: 0 B: 0	R:205 G: 0 B: 0	
R: 0 G: 0 B: 0	R: 0 G: 5 B: 0	R:142 G: 0 B: 0	R:229 G: 0 B: 0	R:253 G: 0 B: 0	R:247 G: 0 B: 0	R:251 G: 0 B: 0	
R: 0 G: 0 B: 0	R: 0 G: 46 B: 0	R:113 G: 0 B: 0	R:205 G: 0 B: 0	R:202 G: 0 B: 0	R:255 G: 0 B: 0	R:240 G: 0 B: 0	
R: 0 G: 0 B: 32	R: 0 G: 25 B: 0	R: 0 G: 49 B: 0	R: 0 G: 53 B: 0	R: 0 G:151 B: 0	R: 0 G: 57 B: 0		
R: 0 G: 0 B: 0	R: 0 G: 0 B: 37	R: 0 G: 0 B: 31	R: 0 G: 0 B:101	R: 0 G: 0 B: 26			

Εικόνα 1.4.11 – Συνιστώσες της RGB Εικόνας

1.5 Βάθος Χρώματος (Color Depth)

ο βάθος χρώματος μετράει το πόση χρωματική πληροφορία είναι διαθέσιμη για πεικόνιση ή εκτύπωση σε κάθε pixel μίας εικόνας. Μεγαλύτερο βάθος χρώματος σημαίνει ερισσότερα διαθέσιμα χρώματα και ακριβέστερη χρωματική απεικόνιση σε μια ψηφιακή κόνια. Κάθε πληροφορία στους υπολογιστές είναι σε δυαδικό σύστημα (0 και 1) έτσι η κρότερη δυνατή πληροφορία είναι ένα δυαδικό ψηφίο δηλαδή 1 bit (με τιμή 0 ή με τιμή 1). πομένως ένα pixel έχει χρωματική πληροφορία ή αλλιώς Bit Depth τουλάχιστον ενός bit.

Το δυαδικό σύστημα έχει για βάση το 2, επομένως έχουμε τους εξής συνδυασμούς:

$2^1 = 2$ συνδυασμοί αριθμών του 1 bit: 0, 1



$2^2 = 4$ συνδυασμοί αριθμών των 2 bit: 00, 01, 10, 11



$2^3 = 8$ συνδυασμοί αριθμών των 3 bit: 000, 001, 010, 011, 100, 101, 110, 111



$2^8 = 256$ συνδυασμοί αριθμών των 8 bit ή 1 byte: 0000 0000, 0000 0001, ..., 1111 1111



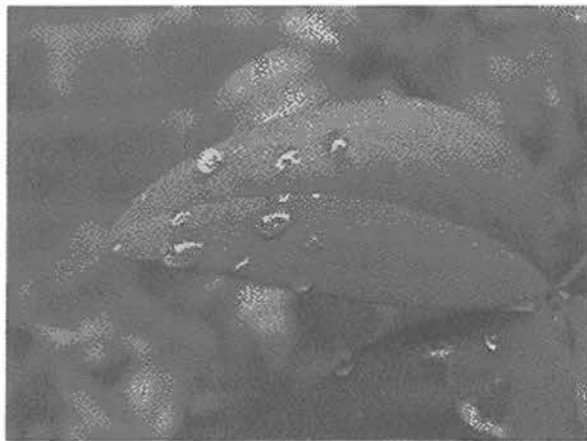
$2^{16} = 65536$ συνδυασμοί αριθμών των 16 bit ή 2 byte



Στην πράξη δεν χρησιμοποιούνται όλοι οι πιθανοί συνδυασμοί αριθμών, αλλά μερικοί από αυτούς, του οποίους τους βλέπουμε στις παρακάτω εικόνες.



Εικόνα 1.5.1 – 1 bit (2 χρώματα)



Εικόνα 1.5.2 – 2 bit (4 χρώματα)



Εικόνα 1.5.3 – 4 bit (16 χρώματα)



Εικόνα 1.5.4 – 8 bit (256 χρώματα)



Εικόνα 1.5.5 – 24 bit (16,777,216 χρώματα, "True color")

6 Ανάλυση Εικόνας (Image Resolution)

Ανάλυση (resolution) της εικόνας είναι ο αριθμός των εικονοστοιχείων της εικόνας σε κάθε διάστασή της και μετριέται σε κουκκίδες ανά ίντσα, dots per inch (dpi) ή pixels per inch (ppi). Όσο αυξάνει ο αριθμός των κουκκίδων αυξάνει η ποιότητα και η ευκρίνεια της εικόνας, καθώς προστίθενται περισσότερες λεπτομέρειες. Ενώ ο συνολικός αριθμός των pixels που περιέχει μια εικόνα παραμένει σταθερός, αν μεγαλώσουμε την εικόνα, θα αυξηθεί η ανάλυσή της και τα pixels της θα μεγαλώσουν σε μέγεθος και θα γίνουν ορατά. Αντίθετα, αν μικραίνουμε την εικόνα, θα μειωθεί η ανάλυση και η ευκρίνειά της.

Αν θέλουμε να καταλάβουμε καλύτερα την έννοια της ανάλυσης, έχουμε πάρει μια εικόνα και την έχουμε μικρύνει κατά δύο και κατά πέντε φορές και την έχουμε απεικονίσει στις αρχικές διαστάσεις. Λόγω των μειωμένων pixels, προσπαθώντας να την απεικονίσουμε στις αρχικές διαστάσεις, παραμορφώνεται η εικόνα.



Εικόνα 1.6.1 – Αρχική Εικόνα (356 x 200)



Εικόνα 1.6.2 – 2 φορές μικρότερη (178 x 100)



Εικόνα 1.6.3 – 5 φορές μικρότερη (72 x 40)

1.7 Μέγεθος αρχείων ψηφιακών εικόνων

Το μέγεθος των αρχείων των υπολογιστών εκφράζεται σε bytes:

1 byte = 8 bit

1 kilobyte (KB) = 1024 bytes (1024 x 8 = 8192 bits)

1 megabyte (MB) = 1024 KB

1 gigabyte (GB) = 1024 MB

1 terabyte (TB) = 1024 GB

Για να υπολογίσουμε το μέγεθος του αρχείου της εικόνας μας πρέπει να ξέρουμε τα bit πληροφορίας που περιέχει. Έστω λοιπόν ότι έχουμε μια εικόνα με διαστάσεις 900 x 1500 και βάθος χρώματος 24 bit. Το μέγεθος της θα είναι :

$$900 \times 1500 \times 24 = 32400000 \text{ bit} = 4050000 \text{ bytes} \approx 3955 \text{ Kb} \approx 3.86 \text{ Mb}$$

Το μέγεθος των ασυμπίεστων εικόνων αυξάνει δραματικά με την αύξηση των διαστάσεων της ή του βάθους χρώματος

1.8 Συμπίεση εικόνων (Image Compression)

Συμπίεση (compression) μιας ακολουθίας δεδομένων, ονομάζουμε την ελάττωση του μεγέθους της ακολουθίας, ώστε να χρειάζεται λιγότερος χώρος για την αποθήκευση ή τη μετάδοσή της. Η διαδικασία της συμπίεσης εφαρμόζεται συστηματικά στα υπολογιστικά συστήματα που χρησιμοποιούν και επεξεργάζονται μεγάλο όγκο ψηφιακών δεδομένων.

Όταν έχουμε την ακολουθία των δεδομένων σε συμπιεσμένη μορφή, πρέπει να εφαρμοστεί η αντίστροφη διαδικασία της αποσυμπίεσης (decompression, extraction) προκειμένου τα δεδομένα να μπορούν να χρησιμοποιηθούν και πάλι. Η διαδικασία αυτή της αποσυμπίεσης των δεδομένων καθορίζει και τις κατηγορίες των μεθόδων συμπίεσης. Έτσι, υπάρχουν δύο κατηγορίες αλγορίθμων συμπίεσης, οι απωλεστικοί (lossy) και οι μη απωλεστικοί (lossless) αλγόριθμοι.

Στους απωλεστικούς αλγορίθμους, όταν γίνει η συμπίεση και μετά ακολουθήσει αποσυμπίεση των δεδομένων, η τελική ακολουθία των δεδομένων διαφέρει από την αρχική. Αντίθετα στους μη απωλεστικούς αλγορίθμους, η διαδικασία συμπίεσης και αποσυμπίεσης επαναφέρει την αρχική ακολουθία. Αν πρέπει να μεταφερθούν δεδομένα με

απόλυτη ακρίβεια, πιστότητα, χωρίς να αλλοιωθεί το περιεχόμενό τους, πρέπει να εφαρμοστεί μια μη απωλεστική μέθοδος συμπίεσης. Εφαρμογές ή συστήματα που μεταδίδουν αναλλοίωτες πληροφορίες από το ένα μέσο στο άλλο, για παράδειγμα κάρτες δικτύου ή modem, χρησιμοποιούν τεχνικές μη απωλεστικής συμπίεσης. Υπάρχουν όμως εφαρμογές όπου η μικρή διαφοροποίηση από την αρχική μορφή των δεδομένων δεν επιφέρει σημαντικές αλλαγές. Έτσι, οι περισσότερες εφαρμογές που έχουν να κάνουν με σύνθετες μορφές δεδομένων όπως είναι ο ήχος, η εικόνα, το video, όπου το τελικό αποτέλεσμα αξιολογείται από τον ανθρώπινο παράγοντα (αυτί, μάτι), μπορούν να κάνουν απωλεστική συμπίεση χωρίς πολλές φορές να υπάρχουν εμφανείς αλλοιώσεις στην ποιότητα των δεδομένων.

.JPG (Joint Photographic Experts Group)

Ο πιο γνωστός τύπος αρχείων εικόνας, που χρησιμεύει για την μετατροπή και προβολή εικόνων σε χαμηλή ανάλυση μέσω internet και ψηφιακών φωτογραφικών μηχανών. Η εικόνα κατά τη συμπίεση μικραίνει ως προς το μέγεθος του αρχείου, χάνοντας ταυτόχρονα ένα ποσοστό από την ποιότητά της, αλλά η προβολή της παραμένει (ανάλογα με το ποσοστό συμπίεσης) αναλλοίωτη. Αυτό γίνεται γιατί ο συγκεκριμένος αλγόριθμος συμπίεσης απορρίπτει τις πληροφορίες μιας εικόνας που αδυνατεί να πιάσει το ανθρώπινο μάτι. Το ποσοστό συμπίεσης είναι ελεγχόμενο, κατά τη δημιουργία του αρχείου, από ένα πεδίο που λέγεται JPG Quality. Η συμπίεση και η ανάλυση λειτουργούν ως δυο τιμές αντιστρόφως ανάλογες. Δηλαδή, όσο μικρότερο γίνεται το ποσοστό της συμπίεσης τόσο υψηλότερη παραμένει η ανάλυση και αντίστοιχα όσο αυξάνεται η συμπίεση τόσο χαμηλότερη γίνεται η ανάλυση άρα και η ποιότητα της εικόνας καθώς και το μέγεθος του αρχείου. Αυτός ο τύπος αρχείου χρησιμοποιείται κυρίως σε περιπτώσεις που το μικρό μέγεθος αρχείου είναι σημαντικότερο από την ποιότητα της εικόνας. Τέτοιες περιπτώσεις είναι η προβολή εικόνων μέσω internet, emails, memory cards ή η χρήση εικόνων σε παρουσιάσεις μέσω projectors ή video walls κ.λπ.

.PNG (Portable Network Graphics)

Το format αυτό δημιουργήθηκε αποκλειστικά για την προβολή εικόνων μέσω internet. Υποστηρίζει από 1-bit έως 48 bit βάθος χρώματος, τις χρωματικές παλέτες RGB και είναι το καλύτερο για την προβολή κειμένου και γραφικών online. Δεν συνιστάται η μεταβολή του μεγέθους του μετά την δημιουργία του αρχείου.

.GIF (Graphics Interchange Format)

Αυτός ο τύπος αρχείου εικόνας δημιουργήθηκε από την CompuServe πριν τον τύπο JPG, για να υποστηρίξει την μετάδοση 8-bit video μέσω συνδέσεων dial up modems. Όμως ακόμα και σήμερα είναι ένας τύπος αρχείου που χρησιμοποιείται μέσω internet. Μπορεί να αναπαράγει μέχρι 256 χρώματα και παρέχει τη δυνατότητα animation και επιτρέπει την συμπίεση των αρχείων.

.TIFF (Tagged Image File Format)

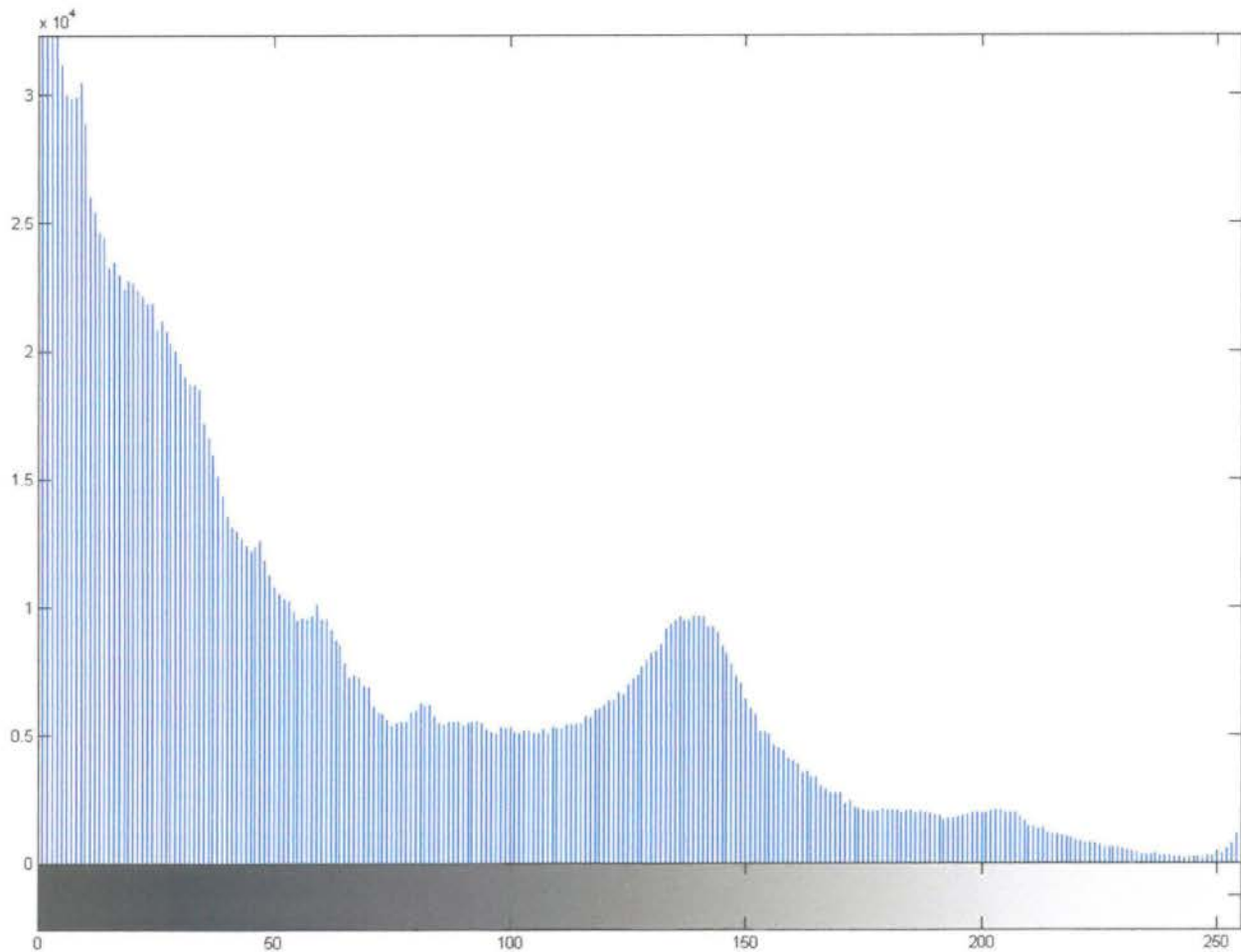
Το format TIF υποστηρίζει την υψηλή ποιότητα των φωτογραφιών και μάλιστα χωρίς απώλειες. Συμπεριλαμβάνει ως επιλογή τη συμπίεση LZW, η οποία θεωρείται η υψηλότερη μορφή στην ποιότητα των εικόνων. Το format TIF χρησιμοποιείται σε πολλές εφαρμογές εκτός από το internet. Υποστηρίζει από 1-bit σε 48 bit βάθος χρώματος, τις χρωματικές παλέτες RGB, CMYK, CIELAB κ.ά. και υψηλές αναλύσεις όπως 150dpi, 300dpi, 600dpi ή 1200dpi. Μέσα στα αρχεία αυτά αποθηκεύονται εικόνες που αποτελούνται από μάσκες, Alpha channels, φίλτρα, layers, κείμενο κ.λπ. Είναι ο τύπος φωτογραφιών που χρησιμοποιείται κυρίως σε έντυπα που πρόκειται να εκτυπωθούν με την offset ή με την ψηφιακή μέθοδο.

.BMP (Bitmap Image File)

Ένας τύπος αρχείου που χρησιμοποιούνταν περισσότερο στο παρελθόν για να αποθηκεύει ψηφιακές εικόνες, οι οποίες εξαρτιόνταν από τις τεχνικές προδιαγραφές της κάρτας γραφικών και του λειτουργικού συστήματος, κυρίως των Microsoft Windows και OS/2. Οι εικόνες που δημιουργούνταν με βάση αυτόν τον τύπο μπορούσε να είναι οποιοδήποτε μεγέθους, ανάλυσης ή χρωματικής παλέτας, αλλά η ασυμπίεστη ουσιαστικά μορφή τους δημιουργούσε πολύ μεγάλα και δύσχρηστα αρχεία.

1.9 Ιστόγραμμα (Histogram)

Το ιστόγραμμα για μια ψηφιακή εικόνα είναι ένα γράφημα που απεικονίζει τις τονικές τιμές των pixels μέσα στην εικόνα. Το ιστόγραμμα δείχνει τον αριθμό των pixels στην εικόνα με μια συγκεκριμένη φωτεινότητα, η κλίμακα της φωτεινότητας εκτείνεται στον οριζόντιο άξονα x και κυμαίνεται από 0 - 255, το εύρος της κλίμακας σχετίζεται με το βάθος χρώματος 8 bit. Ο αριθμός 0 στην αριστερή πλευρά αντιπροσωπεύει το αμιγώς μαύρο χρώμα και ο αριθμός 255 στη δεξιά πλευρά αντιπροσωπεύει το αμιγώς λευκό χρώμα. Ο κατακόρυφος άξονας y αποτελεί τον αριθμό των pixels που περιλαμβάνονται σε ένα συγκεκριμένο τόνο. Όσο υψηλότερη είναι η τιμή στον κατακόρυφο άξονα τόσο περισσότερα είναι τα pixels που έχει ένας συγκεκριμένος τόνος.



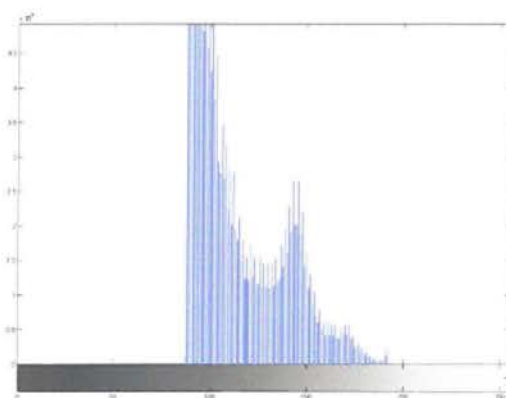
Εικόνα 1.9.1 – Ιστόγραμμα

Εξισορρόπηση ιστογράμματος (Histogram Equalization)

Η εξισορρόπηση ιστογράμματος έχει σκοπό την ομοιόμορφη κατανομή των επιπέδων φωτεινότητας και χρησιμοποιείται σε περιπτώσεις που το ιστόγραμμα μιας εικόνας είναι συγκεντρωμένο σε μία στενή περιοχή (Εικόνα 1.9.3).



Εικόνα 1.9.2 – Εικόνα με Χαμηλή Αντίθεση

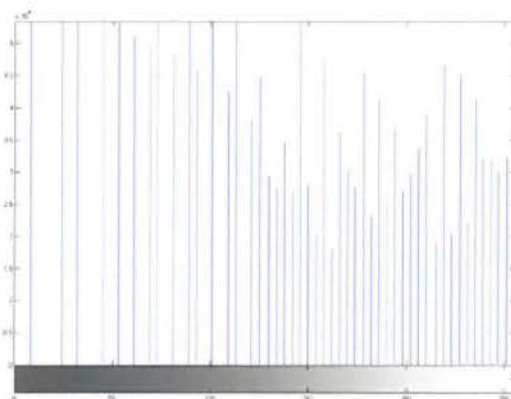


Εικόνα 1.9.3 – Αρχικό Ιστόγραμμα

Ο σκοπός της τεχνικής αυτής είναι να εμπλουτίσει την εικόνα και με άλλες αποχρώσεις του χρώματος, έτσι ώστε το ιστόγραμμά της να είναι περισσότερο ομοιόμορφο σε ολόκληρη την περιοχή.



Εικόνα 1.9.4 – Εξισορροπημένη Εικόνα



Εικόνα 1.9.4 – Εξισορροπημένο Ιστόγραμμα

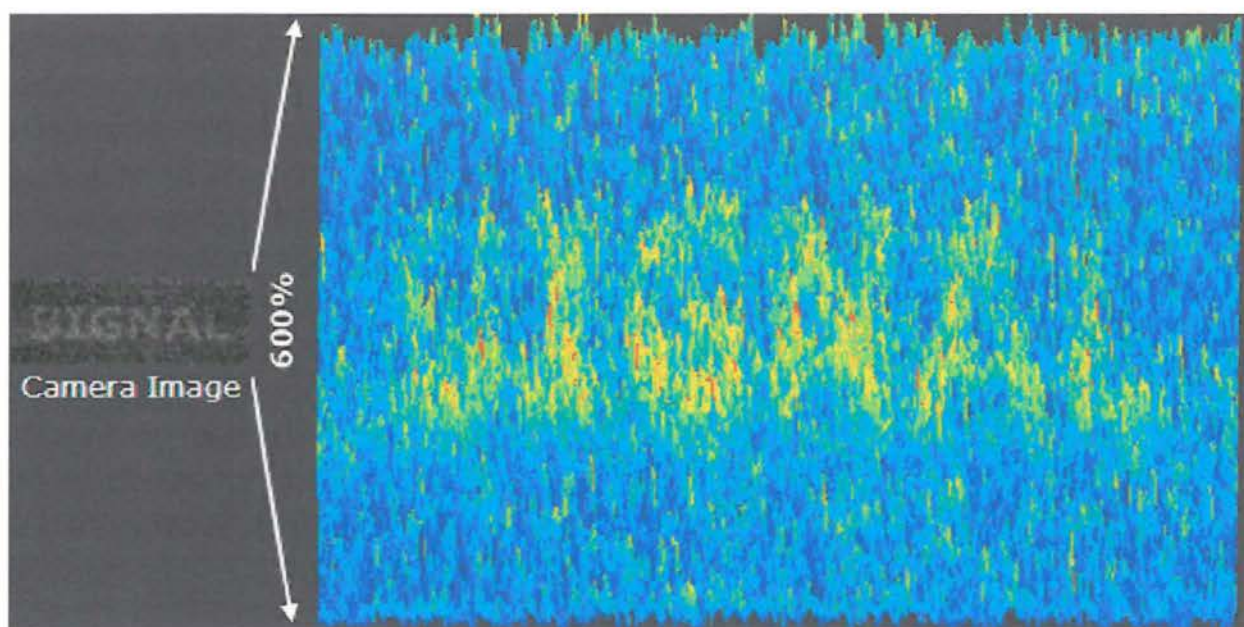
1.10 Θόρυβος (Noise)

Κάποιος βαθμός θορύβου υπάρχει πάντα σε κάθε ηλεκτρονική συσκευή η οποία στέλνει ή λαμβάνει ένα «σήμα». Στις ψηφιακές φωτογραφικές μηχανές, το σήμα είναι το φως που πέφτει στους αισθητήρες της μηχανής. Αν και ο θόρυβος είναι αναπόφευκτος, μπορεί να γίνει τόσο μικρός σε σχέση με το σήμα που μοιάζει ανύπαρκτος. Η σχέση σήματος-θορύβου (Signal to Noise Ratio - SNR) είναι ένας χρήσιμος και παγκόσμιος τρόπος να συγκρίνουμε την σχετική ποσότητα σήματος και θορύβου για κάθε ηλεκτρονικό σύστημα. Όταν ο λόγος σήματος – θορύβου είναι μεγάλος τότε ο θόρυβος φένεται πολύ λίγο, ενώ όταν ο λόγος είναι μικρός συμβαίνει το ανάποδο. Παρακάτω βλέπουμε την παραγωγή μιας εικόνας με πολύ θόρυβο όπου σε ένα λείο φόντο αναγράφεται η λέξη «signal» (= σήμα). Βλέπουμε την εικόνα που δημιουργήθηκε και επίσης την βλέπουμε σε τρισδιάστατη (3D) μεγέθυνση όπου η λέξη βρίσκεται πάνω απο την τον θόρυβο (Εικόνα 1.10.1).



Εικόνα 1.10.1 – 3D αναπαράσταση της εικόνας απο την μηχανή (Camera Image) με μεγάλο λόγο σήματος-θορύβου (SNR)

Η παραπάνω εικόνα που δημιουργήθηκε έχει ικονοποιητικά μεγάλο SNR με αποτέλεσμα να διαχωριστεί η πληροφορία της εικόνας απο τον θόρυβο. Εάν είχαμε μικρό SNR τότε είχαμε το σήμα και ο θόρυβος θα ήταν περισσότερο συγκρίσιμα με αποτέλεσμα να είναι δυσκολότερο να διακρίνουμε το ένα απο το άλλο. Το αποτέλεσμα φένεται παρακάτω (Εικόνα 1.10.2).



Εικόνα 1.10.2 - 3D αναπαράσταση της εικόνας από την μηχανή (Camera Image) με μικρό λόγο σήματος-θορύβου (SNR)

1.10.1 Ομοιόμορφος Θόρυβος (Uniform Noise)

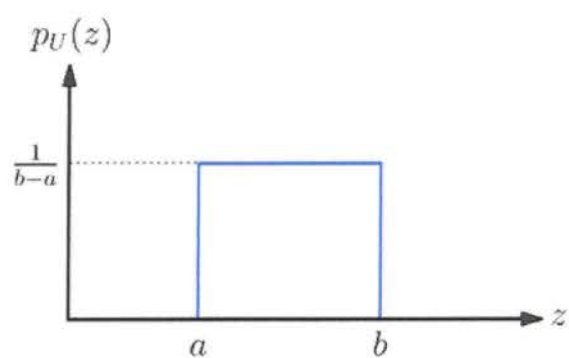
Ο τύπος θορύβου δεν δημιουργείται πρακτικά, αλλά είναι χρήσιμος για την προσομοίωση αλγορίθμων απαλοιφής θορύβου. Στην εικόνα 1.10.3 βλέπουμε την κατανομή του θορύβου σύμφωνα με τον τύπο :

$$p_U(z) = \begin{cases} \frac{1}{b-a} & \text{for } a \leq z \leq b \\ 0 & \text{for otherwise} \end{cases}$$

Η μέση τιμή (μ) (mean) και διασπορά (σ^2) (variance) του είναι :

$$\mu = \frac{a+b}{2}$$

$$\sigma^2 = \frac{(b-a)^2}{12}$$



Εικόνα 1.10.3 – Κατανομή Ομοιόμορφου Θορύβου



Εικόνα 1.10.4 – Αρχική Εικόνα



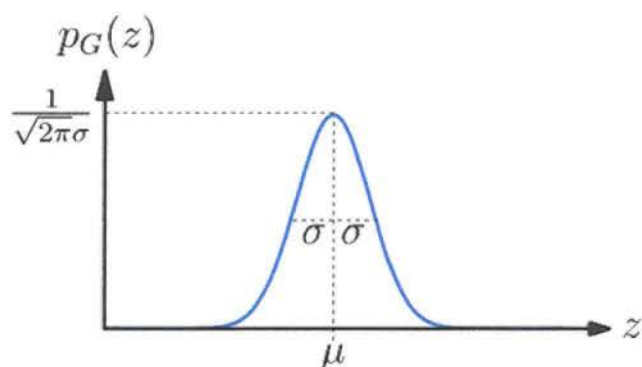
Εικόνα 1.10.5 – Εικόνα με Ομοιόμορφο Θόρυβο

1.10.2 Κανονικός Θόρυβος (Gaussian Noise)

Ο κανονικός θόρυβος υπάρχει σε μία εικόνα κυρίως λόγω της απόδοσης του αισθητηρίου ήψης κάτω από συνθήκες μη επαρκούς φωτισμού ή υψηλής θερμοκρασίας. Στην εικόνα 1.10.6 βλέπουμε την κατανομή του θορύβου, σύμφωνα με την ακόλουθη σχέση:

$$p_G(z) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(z-\mu)^2/(2\sigma^2)}$$

όπου (μ) είναι η μέση τιμή (mean), (σ) είναι η τυπική απόκλιση και (σ^2) είναι η διασπορά (variance).



Εικόνα 1.10.6 – Κατανομή Κανονικού Θορύβου

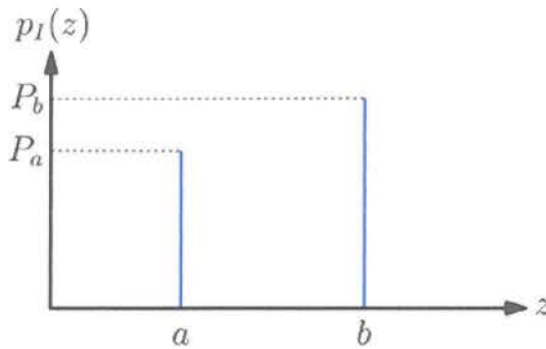


Εικόνα 1.10.7 – Εικόνα με Κανονικό Θόρυβο

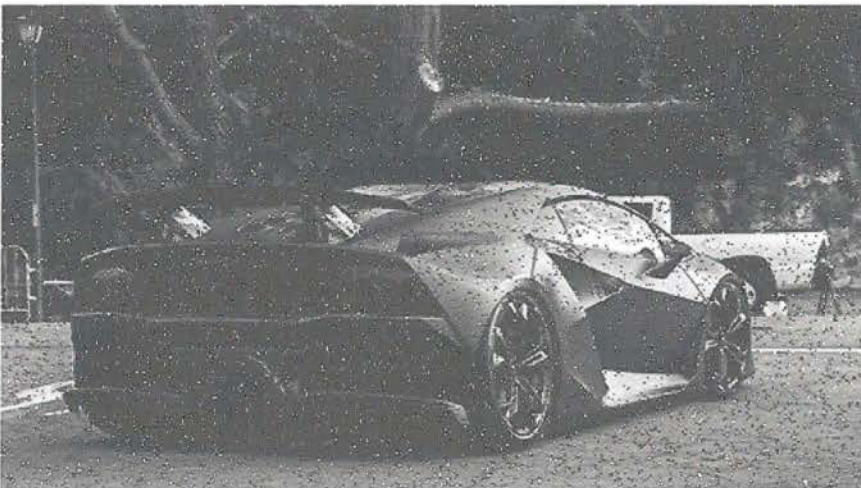
1.10.3 Σημειακός Θόρυβος ή Αλατιού και Πιπεριού (Impulse or Salt and Pepper Noise)

Ο θόρυβος αυτού του τύπου δημιουργείται συνήθως από το on/off διακοπών (μηχανικών ή ηλεκτρονικών) στη διαδικασία λήψης ή μετάδοσης μιας εικόνας. Στην εικόνα 1.10.8 απεικονίζεται η κατανομή του θορύβου, σύμφωνα με την ακόλουθη σχέση:

$$p_I(z) = \begin{cases} P_a & \text{for } z = a \\ P_b & \text{for } z = b \\ 0 & \text{otherwise} \end{cases}$$



Εικόνα 1.10.8 – Κατανομή Σημειακού Θορύβου



Εικόνα 1.10.9 – Εικόνα με Σημειακό Θόρυβο

Υπάρχουν και άλλα είδη θορύβου, όπως: ο θόρυβος Rayleigh, ο θόρυβος Erlang (Gamma) και ο εκθετικός θόρυβος (exponential noise).



Εικόνα 1.10.10 – Εικόνα με Rayleigh θόρυβο



Εικόνα 1.10.11 – Εικόνα με Erlang (Gamma) θόρυβο



Εικόνα 1.10.12 – Εικόνα με Εκθετικό θόρυβο

1.11 Φίλτρα απολοιφής θορύβου

Τα χωρικά φίλτρα (spatial filters) απαλοιφής θορύβου μεταβάλλουν την τιμή ενός pixel της εικόνας με κάποια μαθηματική σχέση η οποία επενεργεί στην γειτονία του, η οποία λέγεται μάσκα (mask).

1.11.1 Παράδειγμα Εφαρμογής Χωρικού Φίλτρου

Έχουμε την παρακάτω εικόνα χωρίς θόρυβο και εφαρμόζουμε το φίλτρο μέσης τιμής, μετά την εφαρμογή του η εικόνα μας θα είναι θολωμένη. Αυτό συμβαίνει διότι όπως θα δούμε και παρακάτω η εικόνα μας εξομαλύνεται.



Εικόνα 1.11.1 – Αρχική Εικόνα



Εικόνα 1.11.2 – Εικόνα μετά το φίλτρο μέσης τιμής 5x5

Στην εικόνα 1.11.3 βλέπουμε μια περιοχή 5x5 της αρχικής μας εικόνας, όπου θα εφαρμόσουμε το φίλτρο μας (Εικόνα 1.11.4) με σκοπό να αλλάξουμε την τιμή του μεσαίου pixel της περιοχής μας. Αφού γίνει ο πολλαπλασιασμός της περιοχής με την μάσκα,

ημιουργείται ένας πίνακας (Εικόνα 1.11.5) με τις καινούριες τιμές του κάθε ρixel με την οήθεια των οποίων, όταν τις αθροίσουμε, θα βρούμε τον μέση τιμή της περιοχής μας που θα γίνει και η καινούργια τιμή του μεσαίου ρixel (Εικόνα 1.11.6). Αυτή η διαδικασία γίνεται για κάθε ρixel της εικόνας που θέλουμε να φιλτράρουμε.

64	88	94	124	77
117	109	136	133	136
94	104	176	108	127
83	66	96	91	130
38	44	43	49	38

Εικόνα 1.11.3 – Περιοχή 5x5 Αρχικής Εικόνας

0.1111	0.1111	0.1111	0.1111	0.1111
0.1111	0.1111	0.1111	0.1111	0.1111
0.1111	0.1111	0.1111	0.1111	0.1111
0.1111	0.1111	0.1111	0.1111	0.1111
0.1111	0.1111	0.1111	0.1111	0.1111

Εικόνα 1.11.4 – Μάσκα Φίλτρου Μέσης Τιμής 5x5

2.5600	3.5200	3.7600	4.9600	3.0800
4.6800	4.3600	5.0400	4.9200	5.4400
3.7600	4.1600	7.0400	4.3200	5.0800
3.3200	2.6400	3.8400	3.6400	5.2000
1.5200	1.7600	1.9600	1.9600	1.5200

Εικόνα 1.11.5 – Τιμή του κάθε ρixel μετά την εφαρμογή του φίλτρου

66	71	79	74	67
70	78	90	80	73
74	82	94	86	76
74	79	89	82	75
70	74	82	80	76

Εικόνα 1.11.6 – Εικόνα μετά το Φίλτρο Μέσης Τιμής 5x5

Με την ίδια λογική εφαρμόζονται όλα τα χωρικά φίλτρα, το μόνο που διαφέρει είναι η μάσκα η οποία εφαρμόζεται στην εικόνα μας.

1.11.2 Φίλτρο Αριθμητικής Μέσης Τιμής (Arithmetic Mean Filter)

$$f^{\wedge}(x,y) = \frac{1}{mn} \sum_{(s,t) \in A_{xy}} g(s,t)$$

Το $g(s,t)$ είναι η εικόνα με θόρυβο, $f^{\wedge}(x,y)$ είναι η έξοδος του φίλτρου και A_{xy} είναι η μάσκα διαστάσεων $m \times n$. Όπως αναφέραμε και στο παράδειγμα, η αριθμητική μέση τιμή προκαλεί την εξομάλυνση μιας περιοχής και η αίσθηση μείωσης του θορύβου προκαλείται από το θολό (blur) αποτέλεσμα στην περιοχή.



Εικόνα 1.11.7 – Εικόνα με Κανονικό Θόρυβο (Gaussian)



Εικόνα 1.11.8 – Εικόνα μετά το Φίλτρο Μέσης Τιμής 5x5



Εικόνα 1.11.9 – Εικόνα με θόρυβο Salt & Pepper



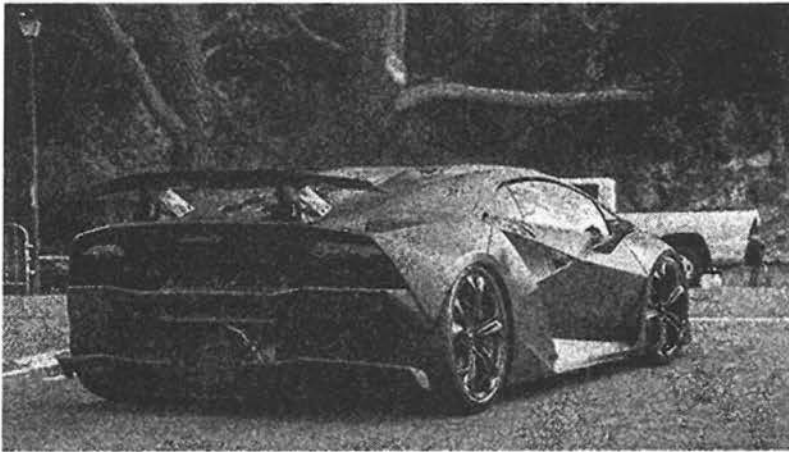
Εικόνα 1.11.10 – Εικόνα μετά το Φίλτρο Μέσης Τιμής 3x3

Παρατηρούμε ότι το φίλτρο της μέσης τιμής δεν είναι αποτελεσματικό για θόρυβο Αλατιού και Πιπεριού (Salt & Pepper). Τα μαύρα pixels έχουν σχεδόν καλυφθεί ενώ τα λευκά δεν έχουν μειωθεί, λόγω του θολώματος μπορούμε να πούμε ότι η εικόνα μας είναι λίγο καλύτερη απο πριν.

1.11.3 Φίλτρο Γεωμετρικής Μέσης Τιμής (Geometric Mean Filter)

$$\hat{f}(x,y) = \left[\prod_{(s,t) \in A_{xy}} g(s,t) \right]^{\frac{1}{mn}}$$

Το κάθε pixel της εικόνας με θόρυβο αντικαθιστάται από το γινόμενο των pixels στη γειτονία του, υψωμένο στη δύναμη $1/mn$. Με αυτό το φίλτρο παρατηρούμε λιγότερο θόλωμα (blurring) από το προηγούμενο φίλτρο.



Εικόνα 1.11.11 – Εικόνα με Κανονικό Θόρυβο (Gaussian)



Εικόνα 1.11.12 – Εικόνα μετά το Φίλτρο Γεωμετρικής Μέσης Τιμής 5x5



Εικόνα 1.11.13 – Εικόνα με θόρυβο Salt & Pepper



Εικόνα 1.11.14 – Εικόνα μετά το Φίλτρο Γεωμετρικής Μέσης Τιμής 3x3

Όπως και το φίλτρο της μέσης τιμής, έτσι και της γεωμετρικής μέσης τιμής δεν είναι αποτελεσματικό για θόρυβο Αλατιού και Πιπεριού (Salt & Pepper). Εδώ συμβαίνει το ανάποδο, τα λευκά pixels έχουν σχεδόν καλυφθεί ενώ τα μαύρα δεν έχουν μειωθεί και μάλιστα έχουν γίνει και πιο έντονα καλύπτοντας σχεδόν όλη την εικόνα.

1.11.4 Φίλτρο Μεσαίας Τιμής (Median Filter)

$$\hat{f}(x,y) = \underset{(s,t) \in A_{xy}}{\text{median}} \ g(s,t)$$

Το κάθε pixel της εικόνας με θόρυβο αντικαθιστάται από τη μεσαία τιμή των pixels στην γειτονία του. Όπως θα δούμε παρακάτω, αυτός ο τύπος φίλτρου είναι ιδιαίτερα αποτελεσματικός στην απομάκρυνση θορύβου Salt & Pepper.



Εικόνα 1.11.15 – Εικόνα με Κανονικό Θόρυβο (Gaussian)



Εικόνα 1.11.16 – Εικόνα μετά το Φίλτρο Μεσαίας Τιμής



Εικόνα 1.11.17 – Εικόνα με θόρυβο Salt & Pepper



Εικόνα 1.11.18 – Εικόνα μετά το Φίλτρο Μεσαίας Τιμής

Παρατηρούμε ότι το ο θόρυβος έχει εξαλειφθεί και η εικόνας μας είναι σχεδόν όπως η αρχική μας, με την διαφορά ότι κυρίως οι ακμές δεν είναι πια λείες και ομοιόμορφες.

1.11.5 Αμφίπλευρο Φίλτρο (Bilateral Filter)

Στο αμφίπλευρο φιλτράρισμα, η τιμή του pixel εξόδου εξαρτάται από τον σταθμισμένο συνδυασμό των τιμών των γειτονικών pixel.

$$g(i,j) = \frac{\sum_{k,l} f(k,l) w(i,j,k,l)}{\sum_{k,l} w(i,j,k,l)}$$

Ο συντελεστής βάρους/στάθμης $w(i, j, k, l)$ εξαρτάται από το γινόμενο του domain kernel και του range kernel.

$$w(i,j,k,l) = \exp\left(-\frac{(i-k)^2 + (j-l)^2}{2\sigma_d^2} - \frac{\|f(i,j) - f(k,l)\|^2}{2\sigma_r^2}\right)$$

Το φίλτρο αυτό ελέγχεται από δύο παραμέτρους : σ_d και σ_r

Όσο αυξάνεται η παράμετρος του εύρους (range - σ_r), το αμφίπλευρο φίλτρο σταδιακά πλησιάζει την Gaussian συνέλιξη επειδή το εύρος Gaussian διευρίνεται (widens) και γίνεται πιο επίπεδο (flattens), που σημαίνει ότι παραμένει σχεδόν σταθερό σε όλη την εικόνα.

Αυξάνοντας την χωρική παράμετρο (spatial / domain - σ_d), τα μεγαλύτερα χαρακτηριστικά γίνονται λεία (smooth).

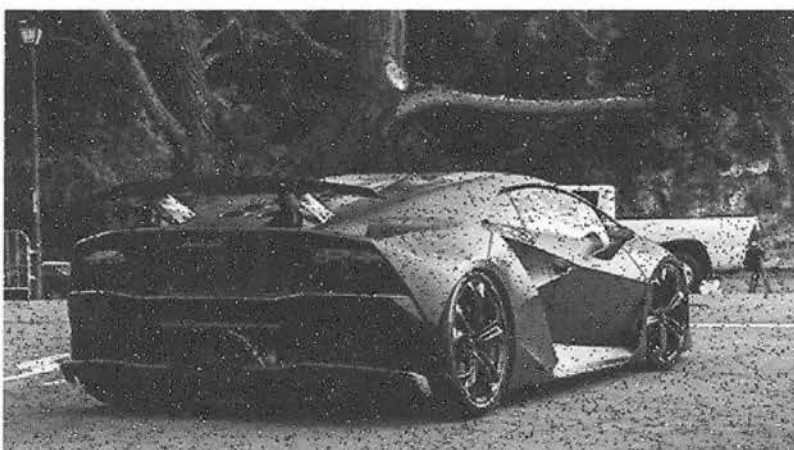


Εικόνα 1.11.19 – Εικόνα με Κανονικό Θόρυβο (Gaussian)



Εικόνα 1.11.20 – Εικόνα μετά το Αμφίπλευρο Φίλτρο

εικόνα μας θυμίζει πίνακα ζωγραφικής ή θα μπορούσαμε να πούμε ότι θυμίζει cartoon, στο λέγεται το φαινόμενο της σκάλας (Staircase Effect).



Εικόνα 1.11.21 – Εικόνα με θόρυβο Salt & Pepper



Εικόνα 1.11.22 – Εικόνα μετά το Αμφίπλευρο Φίλτρο

1.12 – Διαδική Μετατροπή (Binarization - Thresholding)

Η διαδική μετατροπή των εικόνων είναι το πρώτο βήμα των περισσότερων συστημάτων ανάλυσης και επεξεργασίας εγγράφων και αναφέρεται στην μετατροπή των gray scale εικόνων σε ασπρόμαυρες. Η διαδική μετατροπή αποτελεί βασικό στοιχείο των συστημάτων επεξεργασίας εγγράφων γιατί μία βέλτιστη διαδική μετατροπή αποτελεί την βάση για σωστή κατάτμηση και αναγνώριση των εγγράφων. Οι βασικοί λόγοι που το στάδιο της διαδικής μετατροπής είναι απαραίτητο είναι οι παρακάτω:

- Εξοικονόμηση αποθηκευτικών μέσων.
- Βελτίωση της εμφάνισης λόγω αφαίρεσης θορύβου.
- Διευκόλυνση και επιτάχυνση αλγορίθμων κατάτμησης και αναγνώρισης.

Έστω μία εικόνα με 256 στάθμες του γκρι όπου οι τιμές κοντά στο 0 είναι οι σκούρες περιοχές που ανήκουν στο κείμενο ενώ οι τιμές κοντά στο 255 είναι οι πιο ανοιχτές περιοχές και ανήκουν στο υπόβαθρο της εικόνας. Ο πιο εύκολος και συνηθισμένος τρόπος μετατροπής μιας εικόνας πολλών σταθμών του γκρι σε ασπρόμαυρη είναι η επιλογή ενός κατωφλιού (threshold), σύμφωνα με το οποίο όλες οι γκρι στάθμες κάτω από αυτό μετατρέπονται σε 1 (κείμενο) ενώ εκείνες που είναι πάνω από το κατώφλι μετατρέπονται σε 0 (υπόβαθρο - background).

$$f(x_{i,j}) = \begin{cases} 1, & x_{i,j} \leq T \\ 0, & x_{i,j} > T \end{cases}$$

Οι μέθοδοι που χρησιμοποιούν μία τιμή κατωφλιού T για την μετατροπή της εικόνας σε ασπρόμαυρη σύμφωνα με τον παραπάνω τύπο ονομάζονται μέθοδοι καθολικής κατωφλίωσης (global thresholding methods). Ένα παράδειγμα καθολικής κατωφλίωσης φαίνεται στην εικόνα 1.12.1. Το ιστόγραμμα H μιας gray scale εικόνας εκφράζει την κατανομή των αποχρώσεων του γκρι της εικόνας. Στην εικόνα 1.12.3 δίδεται ένα παράδειγμα ιστογράμματος μιας απλής gray scale εικόνας. Στο παράδειγμα αυτό εμφανίζονται μόνο 3 επίπεδα φωτεινότητας (0,2,9) τα οποία αντιστοιχούν σε 133, 48 και 53 pixels αντίστοιχα. Σε περιπτώσεις καθαρών εγγράφων το ιστόγραμμα H εμφανίζει δύο βασικές κατανομές που αντιστοιχούν στις περιοχές του υποβάθρου της εικόνας και του κειμένου (Εικόνα 1.12.3). Το κατώφλι T που χρησιμοποιείται στις μεθόδους καθολικής κατωφλίωσης θα πρέπει να διαχωρίζει ικανοποιητικά τις δύο βασικές κατανομές του ιστογράμματος. Η σωστή επιλογή του κατωφλιού είναι σημαντική τόσο για τον περιορισμό του θορύβου όσο και για την ποιότητα και το πάχος των χαρακτήρων της τελικής ασπρόμαυρης εικόνας. Στην Εικόνα 1.12.4 φαίνονται οι διαφορετικές τελικές εικόνες που προκύπτουν για διαφορετικές τιμές του κατωφλιού T . Παρατηρούμε ότι για μικρές τιμές του T ($T = 60, 80$) δεν έχουμε θόρυβο, όμως έχουμε ασυνέχειες στο σώμα των χαρακτήρων. Αντίθετα για μεγάλες τιμές του T ($T = 140, 160$) έχουμε εισαγωγή θορύβου, αλλά η ποιότητα των χαρακτήρων είναι καλύτερη.

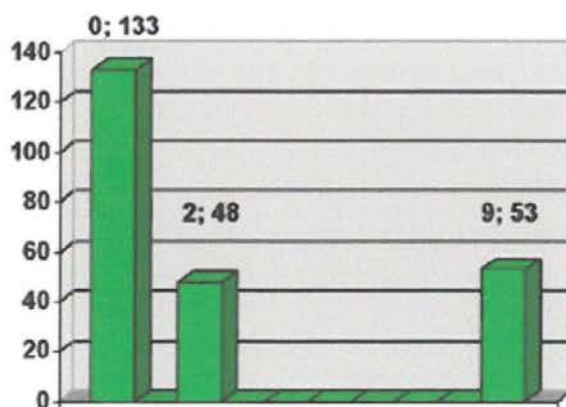
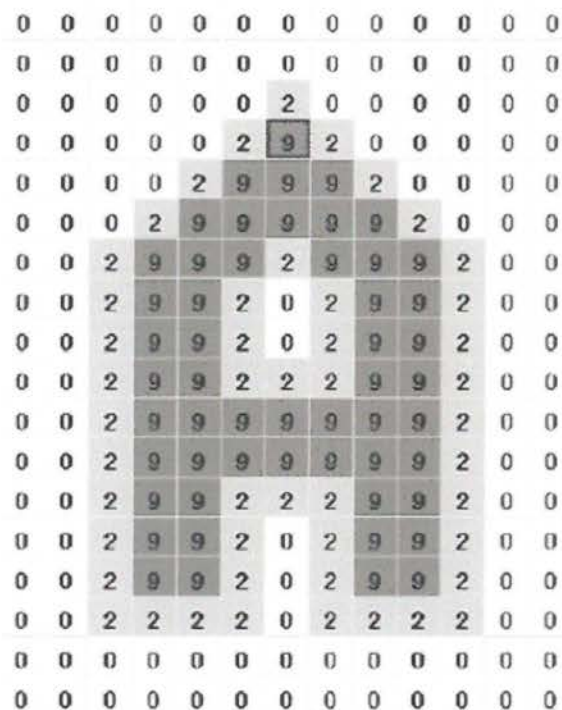
We are truly happy to be able thus heartily to commend the first official document of our new Executive. From the day of his nomination at Baltimore, to the hour of his inauguration at Washington, President Pierce has not uttered a word, nor written a line, at which any good citizen could take exception. We honestly hope that we may be able to repeat this high commendation four years hence.

The *Mirror* is not, never has been, and never will be, in our hands, a *partizan journal*. It nominated Gen. Taylor as an INDEPENDENT CANDIDATE, "without distinction of party." It opposed the election of Gen. Scott, on the ground of his adoption by abolition Whigs. It will support the administration of President Pierce, so long and so far as it can do so honestly and consistently—the cardinal points of our political creed being—*Free Trade with all the world;—and the Union of the States at all sacrifices.*

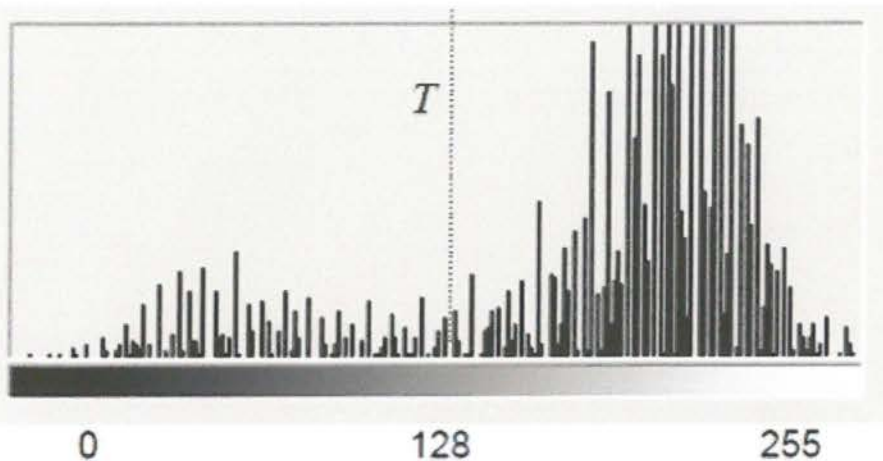
We are truly happy to be able thus heartily to commend the first official document of our new Executive. From the day of his nomination at Baltimore, to the hour of his inauguration at Washington, President Pierce has not uttered a word, nor written a line, at which any good citizen could take exception. We honestly hope that we may be able to repeat this high commendation four years hence.

The *Mirror* is not, never has been, and never will be, in our hands, a *partizan journal*. It nominated Gen. Taylor as an INDEPENDENT CANDIDATE, "without distinction of party." It opposed the election of Gen. Scott, on the ground of his adoption by abolition Whigs. It will support the administration of President Pierce, so long and so far as it can do so honestly and consistently—the cardinal points of our political creed being—*Free Trade with all the world;—and the Union of the States at all sacrifices.*

Εικόνα 1.12.1 – Αρχική Εικόνα και Εικόνα Μετα την Διαδική Μετατροπή



Εικόνα 1.12.2 – Η Grayscale εικόνα και το ιστόγραμμα της με 3 επίπεδα φωτεινότητας



Εικόνα 1.12.3 – Ιστόγραμμα της Εικόνας 1.12.1 και η τιμή T κατωφλίου

We are truly happy to be able thus heartily to commend the first official document of our new Executive. From the day of his nomination at Baltimore, to the hour of his inauguration at Washington, President Pierce has not uttered a word, nor written a line, at which any good citizen could take exception. We honestly hope that we may be able to repeat this high commendation four years hence.

The *Mirror* is not, never has been, and never will be, in our hands, a *partizan journal*. It nominated Gen. Taylor as an INDEPENDENT CANDIDATE, "without distinction of party." It opposed the election of Gen. Scott, on the ground of his adoption by abolition Whigs. It will support the administration of President Pierce, so long and so far as it can do so honestly and consistently—the cardinal points of our political creed being—*Free Trade with all the world;—and the Union of the States at all sacrifices.*

(α)

We are truly happy to be able thus heartily to commend the first official document of our new Executive. From the day of his nomination at Baltimore, to the hour of his inauguration at Washington, President Pierce has not uttered a word, nor written a line, at which any good citizen could take exception. We honestly hope that we may be able to repeat this high commendation four years hence.

The *Mirror* is not, never has been, and never will be, in our hands, a *partizan journal*. It nominated Gen. Taylor as an INDEPENDENT CANDIDATE, "without distinction of party." It opposed the election of Gen. Scott, on the ground of his adoption by abolition Whigs. It will support the administration of President Pierce, so long and so far as it can do so honestly and consistently—the cardinal points of our political creed being—*Free Trade with all the world;—and the Union of the States at all sacrifices.*

(γ)

We are truly happy to be able thus heartily to commend the first official document of our new Executive. From the day of his nomination at Baltimore, to the hour of his inauguration at Washington, President Pierce has not uttered a word, nor written a line, at which any good citizen could take exception. We honestly hope that we may be able to repeat this high commendation four years hence.

The *Mirror* is not, never has been, and never will be, in our hands, a *partizan journal*. It nominated Gen. Taylor as an INDEPENDENT CANDIDATE, "without distinction of party." It opposed the election of Gen. Scott, on the ground of his adoption by abolition Whigs. It will support the administration of President Pierce, so long and so far as it can do so honestly and consistently—the cardinal points of our political creed being—*Free Trade with all the world;—and the Union of the States at all sacrifices.*

(β)

We are truly happy to be able thus heartily to commend the first official document of our new Executive. From the day of his nomination at Baltimore, to the hour of his inauguration at Washington, President Pierce has not uttered a word, nor written a line, at which any good citizen could take exception. We honestly hope that we may be able to repeat this high commendation four years hence.

The *Mirror* is not, never has been, and never will be, in our hands, a *partizan journal*. It nominated Gen. Taylor as an INDEPENDENT CANDIDATE, "without distinction of party." It opposed the election of Gen. Scott, on the ground of his adoption by abolition Whigs. It will support the administration of President Pierce, so long and so far as it can do so honestly and consistently—the cardinal points of our political creed being—*Free Trade with all the world;—and the Union of the States at all sacrifices.*

(δ)

We are truly happy to be able thus heartily to commend the first official document of our new Executive. From the day of his nomination at Baltimore, to the day of his inauguration at Washington, President Pierce has not uttered a word, nor written a line, at which any good citizen could take exception. We honestly hope that we may be able to repeat this high commendation four years hence.

The *Mirror* is not, never has been, and never will be, in our hands, a *partizan journal*. It nominated Gen. Taylor as an INDEPENDENT CANDIDATE, "without distinction of party." It opposed the election of Gen. Scott, on the ground of his adoption by abolition Whigs. It will support the administration of President Pierce, so long and so far as it can do so honestly and consistently—the cardinal points of our political creed being—*Free Trade with all the world;—and the Union of the States at all sacrifices.*

(ε)

We are truly happy to be able thus heartily to commend the first official document of our new Executive. From the day of his nomination at Baltimore, to the day of his inauguration at Washington, President Pierce has not uttered a word, nor written a line, at which any good citizen could take exception. We honestly hope that we may be able to repeat this high commendation four years hence.

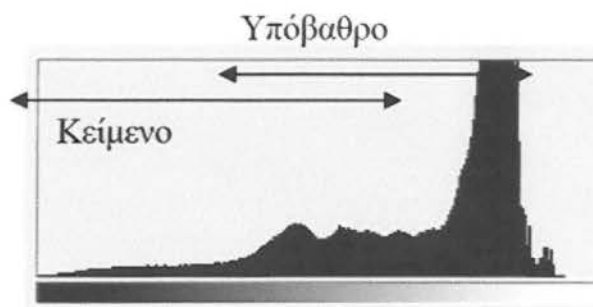
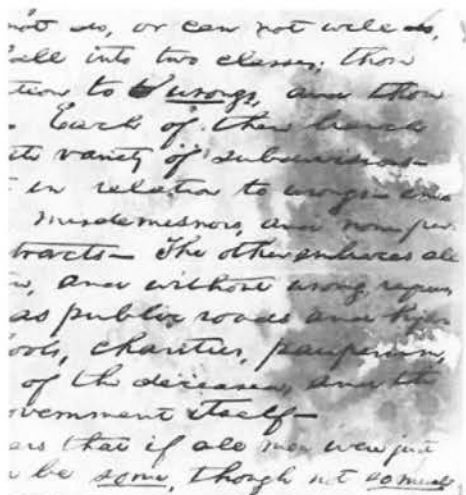
The *Mirror* is not, never has been, and never will be, in our hands, a *partizan journal*. It nominated Gen. Taylor as an INDEPENDENT CANDIDATE, "without distinction of party." It opposed the election of Gen. Scott, on the ground of his adoption by abolition Whigs. It will support the administration of President Pierce, so long and so far as it can do so honestly and consistently—the cardinal points of our political creed being—*Free Trade with all the world;—and the Union of the States at all sacrifices.*

(ζ)

Εικόνα 1.12.4 - Ασπρόμαυρες εικόνες που προκύπτουν από την εικόνα 1.12.1 για εφαρμογή καθολικής κατωφλίωσης επιλέγοντας διαφορετικές τιμές για το κατώφλι T .

(α) $T = 60$, (β) $T = 80$, (γ) $T = 100$, (δ) $T = 120$, (ε) $T = 140$, (ζ) $T = 160$.

Οι μέθοδοι καθολικής κατωφλίωσης μπορούν να εφαρμοστούν μόνο στις περιπτώσεις που υπάρχει σαφής διάκριση των περιοχών του κειμένου από το υπόβαθρο. Αυτό όμως δεν συμβαίνει στις περισσότερες περιπτώσεις των εγγράφων. Ιδίως στα παλιά έγγραφα έχουμε χαμηλή ποιότητα χαρακτήρων, σκιές, μη ομοιόμορφη φωτεινότητα υπόβαθρου, ήχο θόρυβο κ.λ.π. Στο ιστόγραμμα αυτών των εικόνων οι τιμές του γκρι που αντιστοιχούν στο κείμενο και στο υπόβαθρο επικαλύπτονται (Εικόνα 1.12.5). Για τους παραπάνω λόγους αναπτύχθηκαν οι μέθοδοι προσαρμοσμένης ή τοπικής κατωφλίωσης (local or regional or adaptive thresholding methods) στις οποίες χρησιμοποιούνται πολλές τιμές κατωφλίων ανάλογα με την τοπική πληροφορία της εικόνας. Έτσι επιτυγχάνεται η βέλτιστη δυαδική μετατροπή ακόμα και στις περιπτώσεις που δεν υπάρχει σαφής διάκριση των περιοχών του κειμένου από το υπόβαθρο (σχήμα 2.6).



Εικόνα 1.12.5 – Ιστορικό Έγγραφο και Ιστόγραμμα

not so, or can not well do,)
 fall into two classes, those
 which are to be used, and those
 to be left of their hands
 with variety of subdivisions
 in relation to usage - some
 as, moderation, and non
abstract - the other embraces all
 law, and will be among, regimen
 as public roads and high
roads, charities, pauperism,
 & of the diseases, and the
government itself
 that if all non
can be some, though not some,

not so, or can not well do,)
 fall into two classes, those
 which are to be used, and those
 to be left of their hands
 with variety of subdivisions
 in relation to usage - some
 as, moderation, and non
abstract - the other embraces all
 law, and will be among, regimen
 as public roads and high
roads, charities, pauperism,
 & of the diseases, and the
government itself
 that if all non
can be some, though not some,

Εικόνα 1.12.6 - Είναι αδύνατη η επιλογή ενός κατωφλιού για επιτυχή δυαδική μετατροπή.

1.12.1 – Καθολική Κατωφλίωση (Global Thresholding)

Παρακάτω θα δούμε κάποια χαρακτηριστικά παραδείγματα μεθοδολογιών κατωφλίωσης. Ζητούμενο όλων των μεθοδολογιών είναι ο υπολογισμός του καθολικού κατωφλιού ώστε να επιτυγχάνεται βέλτιστη δυαδική μετατροπή εικόνας.

Χρήση σημείων ακμών – Weszka

Τα σημεία ακμών είναι σημεία ανάμεσα στο αντικείμενο και στο υπόβαθρο της εικόνας ή μεταξύ δύο αντικειμένων. Το ιστόγραμμα της εικόνας που προκύπτει μόνο από τα σημεία αυτά μας δίνει καλύτερη διαχωριστικότητα κειμένου-υποβάθρου σε σχέση με το ιστόγραμμα της συνολικής εικόνας. Η ιδέα αυτή χρησιμοποιήθηκε από τον (Weszka 1974) ο οποίος πρότεινε μία μέθοδο καθολικής κατωφλίωσης βασισμένη στην Laplacian της αρχικής εικόνας, η οποία είναι μέθοδος προσδιορισμού ακμών. Πρώτα υπολογίζεται η Laplacian της εικόνας των σταθμών του γκρι. Αυτό μπορεί να γίνει με την εφαρμογή της παρακάτω μάσκας στην επιφάνεια της εικόνας:

$$\begin{matrix}
 0 & 1 & 0 \\
 1 & -4 & 1 \\
 0 & 1 & 0
 \end{matrix}$$

Στη συνέχεια προσδιορίζεται το ιστόγραμμα της αρχικής εικόνας λαμβάνοντας υπόψη μόνο τα σημεία που έχουν υψηλές Laplacian τιμές. Στην εικόνα 1.12.8 φαίνεται η Laplacian της αρχικής εικόνας καθώς και το ιστόγραμμα που προκύπτει από την συνολική εικόνα καθώς και από τα pixels που έχουν υψηλές Laplacian τιμές. Στην δεύτερη περίπτωση υπάρχει μεγαλύτερη διαχωριστικότητα κειμένου-υποβάθρου. Για τον εντοπισμό των δύο κορυφών του ιστογράμματος οι δύο τιμές της φωτεινότητας i_1 και i_2 μπορούν να εντοπιστούν ως εξής:

$$\begin{aligned}
 i_1: \text{Hist}(i) &= \max \text{ για } i = i_1 \\
 i_2: \text{Hist}(i) * (i - i_1)^2 &= \max \text{ για } i = i_2
 \end{aligned}$$

Εντονισμένη εικόνα. Σε τον αλγόριθμο αυτό εξασφαλίζεται ότι η δεύτερη κορυφή του ιστογράμματος απέχει από την πρώτη (Εικόνα 1.12.9).

Grey-level segmentation, or *thresholding*, is a conversion between a grey-level image and a *bilevel* (or *monochrome*, or *black and white*) image. A bilevel image should contain all of the essential information concerning the number, position, and shape of objects while containing a lot less information. The essential reason for classifying pixels by grey level is that pixels with similar levels in a nearby region usually belong to the same object, and reducing the complexity of the data simplifies many recognition and classification procedures. Thresholding is almost essential before thinning, vectorization, and morphological operations.

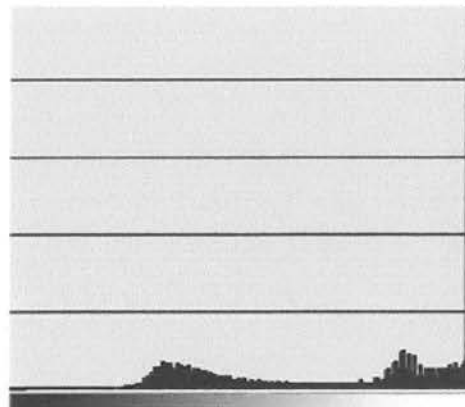
The most common way to convert between grey-level and bilevel images is to select a single threshold value. All of the grey levels below this value will be classified as black (0), and those above will be white (1). The segmentation problem becomes one of selecting the proper value for the threshold T . What is being assumed here is that the pixels in an image I belong to one of two classes based on their grey level. The first class is the collection of black pixels, which will be given the value one, and for this class:

$$I(i,j) < T \quad (\text{EQ 3.1})$$

The other class consists of those pixels that will become white:

$$I(i,j) \geq T \quad (\text{EQ 3.2})$$

This assumption is only true in some real images because of noise and illumination effects. It is not generally true that a single threshold can be used to



Εικόνα 1.12.7 – Αρχική εικόνα και το Ιστόγραμμα της

Grey-level segmentation, or *thresholding*, is a conversion between a grey-level image and a *bilevel* (or *monochrome*, or *black and white*) image. A bilevel image should contain all of the essential information concerning the number, position, and shape of objects while containing a lot less information. The essential reason for classifying pixels by grey level is that pixels with similar levels in a nearby region usually belong to the same object, and reducing the complexity of the data simplifies many recognition and classification procedures. Thresholding is almost essential before thinning, vectorization, and morphological operations.

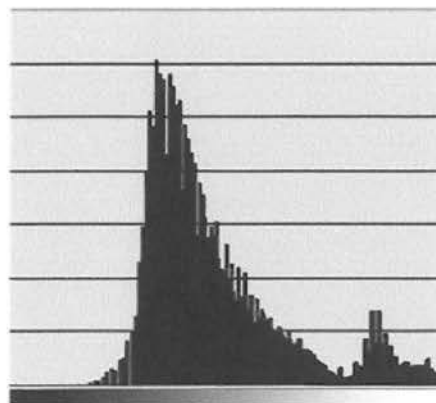
The most common way to convert between grey-level and bilevel images is to select a single threshold value. All of the grey levels below this value will be classified as black (0), and those above will be white (1). The segmentation problem becomes one of selecting the proper value for the threshold T . What is being assumed here is that the pixels in an image I belong to one of two classes based on their grey level. The first class is the collection of black pixels, which will be given the value one, and for this class:

$$I(i,j) < T \quad (\text{EQ 3.1})$$

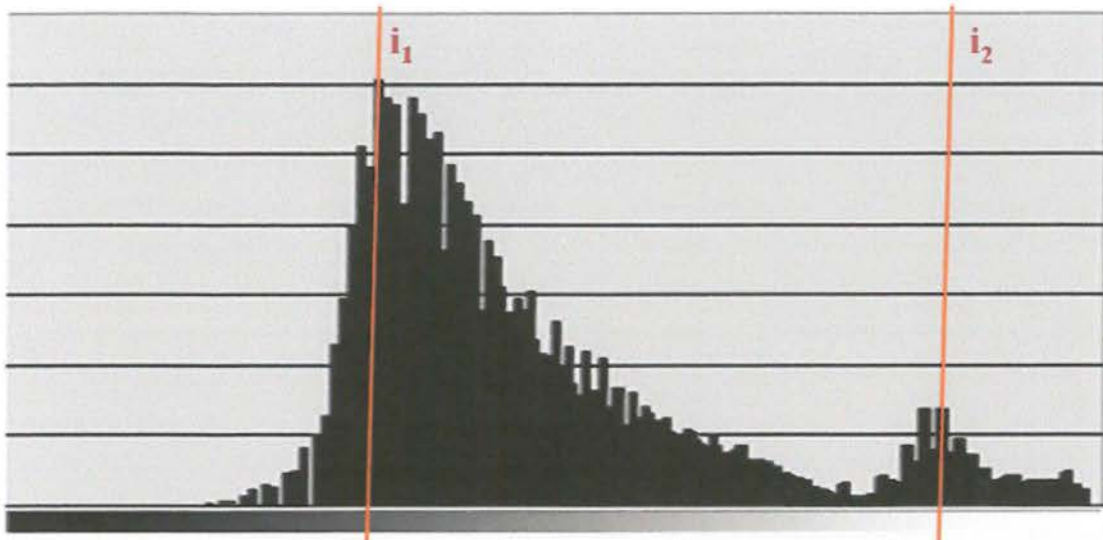
The other class consists of those pixels that will become white:

$$I(i,j) \geq T \quad (\text{EQ 3.2})$$

This assumption is only true in some real images because of noise and illumination effects. It is not generally true that a single threshold can be used to



Εικόνα 1.12.8 – Laplacian της Αρχικής Εικόνας και το Ιστόγραμμα της αρχικής εικόνας εντονισμένης υπόψη μόνο τα σημεία που έχουν υψηλές Laplacian τιμές.



Εικόνα 1.12.9 - Εύρεση των δύο κορυφών του ιστογράμματος

Επιλογή μετά απο επανάληψη – Ridler

Η επιλογή μετά από επανάληψη (Ridler 1978) είναι μια μεθοδολογία η οποία ξεκινά από μία αρχική εκτίμηση για την τιμή του κατωφλιού και στη συνέχεια γίνεται διόρθωση αυτής της τιμής λαμβάνοντας υπόψη τις περιοχές κειμένου και υποβάθρου που προκύπτουν κάθε φορά. Η αρχική εκτίμηση του κατωφλιού είναι η μέση τιμή των gray scale τιμών. Αυτό το κατώφλι χρησιμοποιείται για να υπολογιστούν στατιστικές για τις περιοχές κειμένου και υποβάθρου που προκύπτουν. Υπολογίζεται η μέση τιμή των gray scale τιμών των pixels κάτω από το αρχικό κατώφλι και αποθηκεύεται στην μεταβλητή T_b . Αντίστοιχα, υπολογίζεται και η μέση τιμή των gray scale τιμών των pixels πάνω από το αρχικό κατώφλι και αποθηκεύεται στην μεταβλητή T_o . Στη συνέχεια υπολογίζεται μία νέα προσέγγιση του κατωφλιού στην τιμή $(T_b + T_o)/2$ και η διαδικασία επαναλαμβάνεται μέχρι να μην υπάρχει αλλαγή στο κατώφλι μεταξύ δύο διαδοχικών επαναλήψεων (Εικόνα 1.12.10).

Grey-level segmentation, or *thresholding*, is a conversion between a grey-level image and a *bilevel* (or *monochrome*, or *black and white*) image. A bilevel image should contain all of the essential information concerning the number, position, and shape of objects while containing a lot less information. The essential reason for classifying pixels by grey level is that pixels with similar levels in a nearby region usually belong to the same object, and reducing the complexity of the data simplifies many recognition and classification procedures. Thresholding is almost essential before thinning, vectorization, and morphological operations.

The most common way to convert between grey-level and bilevel images is to select a single threshold value. All of the grey levels below this value will be classified as black (0), and those above will be white (1). The segmentation problem becomes one of selecting the proper value for the threshold T . What is being assumed here is that the pixels in an image I belong to one of two classes based on their grey level. The first class is the collection of black pixels, which will be given the value one, and for this class:

$$I(i,j) < T \quad (EQ 3.1)$$

The other class consists of those pixels that will become white:

$$I(i,j) \geq T \quad (EQ 3.2)$$

This assumption is only true in some real images because of noise and illumination effects. It is not generally true that a single threshold can be used

(α)

Grey-level segmentation, or *thresholding*, is a conversion between a grey-level image and a *bilevel* (or *monochrome*, or *black and white*) image. A bilevel image should contain all of the essential information concerning the number, position, and shape of objects while containing a lot less information. The essential reason for classifying pixels by grey level is that pixels with similar levels in a nearby region usually belong to the same object, and reducing the complexity of the data simplifies many recognition and classification procedures. Thresholding is almost essential before thinning, vectorization, and morphological operations.

The most common way to convert between grey-level and bilevel images is to select a single threshold value. All of the grey levels below this value will be classified as black (0), and those above will be white (1). The segmentation problem becomes one of selecting the proper value for the threshold T . What is being assumed here is that the pixels in an image I belong to one of two classes based on their grey level. The first class is the collection of black pixels, which will be given the value one, and for this class:

$$I(i,j) < T \quad (EQ 3.1)$$

The other class consists of those pixels that will become white:

$$I(i,j) \geq T \quad (EQ 3.2)$$

This assumption is only true in some real images because of noise and illumination effects. It is not generally true that a single threshold can be used

(β)

Grey-level segmentation, or *thresholding*, is a conversion between a grey-level image and a *bilevel* (or *monochrome*, or *black and white*) image. A bilevel image should contain all of the essential information concerning the number, position, and shape of objects while containing a lot less information. The essential reason for classifying pixels by grey level is that pixels with similar levels in a nearby region usually belong to the same object, and reducing the complexity of the data simplifies many recognition and classification procedures. Thresholding is almost essential before thinning, vectorization, and morphological operations.

The most common way to convert between grey-level and bilevel images is to select a single threshold value. All of the grey levels below this value will be classified as black (0), and those above will be white (1). The segmentation problem becomes one of selecting the proper value for the threshold T . What is being assumed here is that the pixels in an image I belong to one of two classes based on their grey level. The first class is the collection of black pixels, which will be given the value one, and for this class:

$$I(i,j) < T \quad (\text{EQ 3.1})$$

The other class consists of those pixels that will become white:

$$I(i,j) \geq T \quad (\text{EQ 3.2})$$

This assumption is only true in some real images because of noise and illumination effects. It is not generally true that a single threshold can be used to

(γ)

Grey-level segmentation, or *thresholding*, is a conversion between a grey-level image and a *bilevel* (or *monochrome*, or *black and white*) image. A bilevel image should contain all of the essential information concerning the number, position, and shape of objects while containing a lot less information. The essential reason for classifying pixels by grey level is that pixels with similar levels in a nearby region usually belong to the same object, and reducing the complexity of the data simplifies many recognition and classification procedures. Thresholding is almost essential before thinning, vectorization, and morphological operations.

The most common way to convert between grey-level and bilevel images is to select a single threshold value. All of the grey levels below this value will be classified as black (0), and those above will be white (1). The segmentation problem becomes one of selecting the proper value for the threshold T . What is being assumed here is that the pixels in an image I belong to one of two classes based on their grey level. The first class is the collection of black pixels, which will be given the value one, and for this class:

$$I(i,j) < T \quad (\text{EQ 3.1})$$

The other class consists of those pixels that will become white:

$$I(i,j) \geq T \quad (\text{EQ 3.2})$$

This assumption is only true in some real images because of noise and illumination effects. It is not generally true that a single threshold can be used to

(δ)

Εικόνα 1.12.10 - Καθολική κατωφλίωση με χρήση της επιλογής μετά από επανάληψη για την εικόνα του σχήματος 2.6α. (α) 1η επανάληψη, κατώφλι = 238. (β) 2η επανάληψη, κατώφλι = 209. (γ) 3η επανάληψη, κατώφλι = 189. (δ) 4η, 5η επανάληψη, κατώφλι = 182.

Επεξεργασία του Ιστογράμματος – Otsu

Η μέθοδος του Otsu (Otsu 1979) βασίζεται στην επεξεργασία του ιστογράμματος της εικόνας και στον προσδιορισμό του κατωφλιού βάσει του κριτηρίου της μεγιστοποίησης της διαχωρισιμότητας μεταξύ των περιοχών κειμένου και υποβάθρου. Αρχικά, υπολογίζουμε το ολικό τετράγωνο της τυπικής απόκλισης (global variance) των επιπέδων του γκρι της εικόνας σ_t^2 :

$$\sigma_t^2 = \sum_{i=0}^{255} (i - u_t)^2 p(i)$$

όπου $p(i)$ η πιθανότητα εμφάνισης της i στάθμης της εικόνας και u_t το συνολικό μέσο επίπεδο γκρι της εικόνας και N το συνολικό πλήθος των pixels της εικόνας :

$$p(i) = \frac{H(i)}{N}, \quad u_t = \sum_{i=0}^{255} i p(i)$$

Για κάθε υποψήφιο κατώφλι t το ιστογράμμα της εικόνας χωρίζεται σε δύο περιοχές - κλάσεις. Μπορούμε να υπολογίζουμε το τετράγωνο της τυπικής απόκλισης μεταξύ των δύο κλάσεων (between classes variance) σ_b^2 το οποίο είναι το τετράγωνο της τυπικής απόκλισης των μέσων τιμών κάθε κλάσης από την συνολική μέση τιμή όλων των pixels και υπολογίζεται από τον παρακάτω τύπο:

$$\sigma_b^2 = \omega_0 \omega_1 (\mu_1 - \mu_0)^2$$

όπου τα ω_0 , ω_1 , μ_0 και μ_1 υπολογίζονται από τους τύπους:

$$\omega_0 = \sum_{i=0}^t p(i), \quad \omega_1 = 1 - \omega_0, \quad \mu_0 = \frac{\sum_{i=0}^t i p(i)}{\omega_0}, \quad \mu_1 = \frac{\sum_{i=t+1}^{255} i p(i)}{\omega_1}$$

Σύμφωνα με τον Otsu, η βέλτιστη τιμή για το κατώφλι t επιτυγχάνεται όταν μεγιστοποιείται ο λόγος του τετραγώνου της τυπικής απόκλισης μεταξύ των δύο κλάσεων (between classes variance) προς το ολικό τετράγωνο της τυπικής απόκλισης (global variance) των επιπέδων του γκρι της εικόνας:

$$n(t) = \frac{\sigma_b^2}{\sigma_t^2}$$

Η μέθοδος του Otsu έχει αξιολογηθεί σαν μία από τις καλύτερες τεχνικές εύρεσης καθολικού κατωφλιού. Στην εικόνα 1.12.11 φαίνεται το αποτέλεσμα εύρεσης κατωφλιού με την παραπάνω μεθόδο της εικόνας 1.12.7.

Grey-level segmentation, or *thresholding*, is a conversion between a grey-level image and a *bilevel* (or *monochrome*, or *black and white*) image. A bilevel image should contain all of the essential information concerning the number, position, and shape of objects while containing a lot less information. The essential reason for classifying pixels by grey level is that pixels with similar levels in a nearby region usually belong to the same object, and reducing the complexity of the data simplifies many recognition and classification procedures. Thresholding is almost essential before thinning, vectorization, and morphological operations.

The most common way to convert between grey-level and bilevel images is to select a single threshold value. All of the grey levels below this value will be classified as black (0), and those above will be white (1). The segmentation problem becomes one of selecting the proper value for the threshold T . What is being assumed here is that the pixels in an image I belong to one of two classes based on their grey level. The first class is the collection of black pixels, which will be given the value one, and for this class:

$$I(i,j) < T \quad (\text{EQ 3.1})$$

The other class consists of those pixels that will become white:

$$I(i,j) \geq T \quad (\text{EQ 3.2})$$

This assumption is only true in some real images because of noise and illumination effects. It is not generally true that a single threshold can be used to

Εικόνα 1.12.11 - Διαδική μετατροπή της εικόνας 1.12.7 εφαρμόζοντας την μέθοδο του Otsu.

Χρησιμοποιώντας την Εντροπία – Pun, Karur

Η εντροπία είναι μία μετρική που σχετίζεται με το περιεχόμενο της πληροφορίας. Αν έχουμε n πιθανά σύμβολα (π.χ. γράμματα ή ψηφία) και το σύμβολο i εμφανίζεται με πιθανότητα $p(x_i)$, η εντροπία που σχετίζεται με τα σύμβολα X είναι:

$$H(X) = - \sum_{i=1}^n p(x_i) \log(p(x_i))$$

όπου η εντροπία μετριέται σε bits/σύμβολο.

Η εικόνα μπορεί να θεωρηθεί σαν ένα σύνολο συμβόλων ή τιμών του γκρι. Θεωρώντας ένα κατώφλι t , η εντροπία που σχετίζεται με τα pixels κειμένου/φόντου μπορεί να δοθεί από τον τύπο (Pun 1981):

$$H_b = - \sum_{i=1}^t p_i \log(p_i)$$

όπου p_i είναι η πιθανότητα εμφάνισης της στάθμης i . Παρόμοια, η εντροπία των pixels του υποβάθρου της εικόνας είναι:

$$H_w = - \sum_{i=t+1}^{255} p_i \log(p_i)$$

Ο προτεινόμενος αλγόριθμος (Pun 1981) προσπαθεί να βρει την βέλτιστη τιμή κατωφλιού t η οποία μεγιστοποιεί την $H = H_b + H_w$. Μία παραλλαγή του παραπάνω αλγορίθμου προτάθηκε από τον (Karur 1985) ο οποίος ορίζει τις πιθανότητες των δύο βασικών κατανομών ως εξής:

$$A: \frac{p_0}{P_t}, \frac{p_1}{P_t}, \dots, \frac{p_t}{P_t}, \quad B: \frac{p_{t+1}}{1-P_t}, \frac{p_{t+2}}{1-P_t}, \dots, \frac{p_{255}}{1-P_t}$$

όπου P_t η αθροιστική πιθανότητα εμφάνισης μέχρι την στάθμη t :

$$P_t = \sum_{i=0}^t p_i$$

Οι εντροπίες των περιοχών κειμένου/φόντου και υποβάθρου υπολογίζονται ανάλογα:

$$H_b = - \sum_{i=1}^t \frac{p_i}{P_t} \log\left(\frac{p_i}{P_t}\right), \quad H_w = - \sum_{i=t+1}^{255} \frac{p_i}{1-P_t} \log\left(\frac{p_i}{1-P_t}\right)$$

Grey-level segmentation, or *thresholding*, is a conversion between a grey-level image and a *bilevel* (or *monochrome*, or *black and white*) image. A bilevel image should contain all of the essential information concerning the number, position, and shape of objects while containing a lot less information. The essential reason for classifying pixels by grey level is that pixels with similar levels in a nearby region usually belong to the same object, and reducing the complexity of the data simplifies many recognition and classification procedures. Thresholding is almost essential before thinning, vectorization, and morphological operations.

The most common way to convert between grey-level and bilevel images is to select a single threshold value. All of the grey levels below this value will be classified as black (0), and those above will be white (1). The segmentation problem becomes one of selecting the proper value for the threshold T . What is being assumed here is that the pixels in an image I belong to one of two classes based on their grey level. The first class is the collection of black pixels, which will be given the value one, and for this class:

$$I(i,j) < T \quad (\text{EQ 3.1})$$

The other class consists of those pixels that will become white:

$$I(i,j) \geq T \quad (\text{EQ 3.2})$$

This assumption is only true in some real images because of noise and illumination effects. It is not generally true that a single threshold can be used to

Εικόνα 1.12.12 - Δυαδική μετατροπή της εικόνας 1.12.7 εφαρμόζοντας την μέθοδο της Εντροπίας.

1.12.2 – Προσαρμοσμένη Κατωφλίωση (Adaptive Thresholding)

Παρακάτω θα δούμε κάποια χαρακτηριστικά παραδείγματα μεθόδων προσαρμοσμένης κατωφλίωσης. Σύμφωνα με αυτές αναζητούνται πολλές τιμές κατωφλιών ανάλογα με την τοπική πληροφορία της εικόνας και έτσι επιτυγχάνεται η βέλτιστη δυαδική μετατροπή ακόμα και στις περιπτώσεις που δεν υπάρχει σαφής διάκριση των περιοχών του κειμένου από το υπόβαθρο.

Διαίρεση της Εικόνας – Chow

Σύμφωνα με την μέθοδο του (Chow 1972) η εικόνα διαιρείται σε επικαλυπτόμενες περιοχές. Για παράδειγμα για μία εικόνα 256x256 προτείνεται η διαίρεση σε 49 επικαλυπτόμενες περιοχές οι οποίες αποτελούνται από 64x64 pixels. Καταρχήν, υπολογίζεται το ιστόγραμμα για κάθε περιοχή και γίνεται έλεγχος ύπαρξης δύο βασικών κατανομών. Στην περίπτωση που υπάρχουν οι δύο κατανομές, θα υπάρχουν και δύο Gaussian καμπύλες που θα προσεγγίζουν τα δεδομένα χρησιμοποιώντας την μέθοδο των ελαχίστων τετραγώνων και κατά συνέπεια θα προσεγγίζεται και το βέλτιστο κατώφλι ανάμεσα στις δύο κατανομές. Στην περίπτωση που δεν υπάρχουν δύο κατανομές τιμών στο ιστόγραμμα των περιοχών, θα γίνεται προσεγγιστικός υπολογισμός (με παρεμβολή) του κατωφλιού χρησιμοποιώντας τα κατώφλια των γειτονικών περιοχών. Τέλος, σε κάθε pixel αντιστοιχίζεται μία τιμή κατωφλιού κάνοντας μία παρεμβολή με τις τιμές των γειτονικών pixels. Η μέθοδος αυτή είναι ο πρόδρομος των μεθόδων προσαρμοσμένης κατωφλίωσης και προτάθηκε κυρίως για την βελτίωση της ποιότητας των εικόνων από καρδιακές ακτινογραφίες.

Χρήση Ακμών - Yanowitz & Bruckstein

Σύμφωνα με την μέθοδο αυτή (Yanowitz 1989) πρώτα εντοπίζονται οι ακμές της εικόνας (σημεία που ανήκουν στα όρια των αντικειμένων). Οι αντίστοιχες στάθμες του γκρι για τα σημεία των ακμών θεωρούνται ενδεικτικές και για τα υπόλοιπα σημεία των αντικειμένων στο οποίο ανήκουν οι ακμές. Στη συνέχεια παράγεται μία επιφάνεια η οποία ταιριάζει στα επίπεδα φωτεινότητας των ακμών και δίνει προσεγγιστικές τιμές για τα υπόλοιπα σημεία των αντικειμένων που δεν ανήκουν στις ακμές της εικόνας. Τα σημεία που έχουν φωτεινότητα σημαντικά πάνω από την επιφάνεια αυτή ανήκουν στο υπόβαθρο, αλλιώς στο κείμενο/φόντο. Η μέθοδος αυτή έχει κυρίως εφαρμογή σε εικόνες όπου η φωτεινότητα του υποβάθρου μεταβάλλεται.

Χρήση Παραθύρου - Niblack, Sauvola

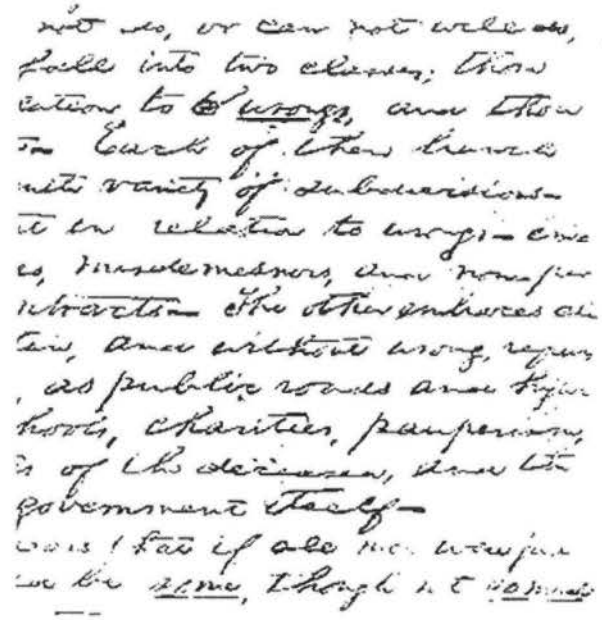
Ο αλγόριθμος του Niblack (Niblack 1986) υπολογίζει ένα κατώφλι για κάθε pixel κάνοντας χρήση ενός παραθύρου το οποίο μετακινείται σε όλη την εικόνα. Το κατώφλι T του κεντρικού pixel του παραθύρου υπολογίζεται χρησιμοποιώντας την μέση τιμή m και την τυπική απόκλιση s των τιμών των επιπέδων φωτεινότητας στο παράθυρο:

$$T = m + k s$$

όπου k είναι σταθερά ίση με -0.2. Η τιμή του k εξαρτάται από το ποσοστό του ορίου των χαρακτήρων που θεωρείται σαν μέρος των χαρακτήρων (σχήμα 2.10). Η μέθοδος αυτή μπορεί να ξεχωρίσει το αντικείμενο από το φόντο ακόμα και όταν οι φωτεινότητές τους δεν απέχουν αρκετά. Τα αποτελέσματα δεν είναι ιδιαίτερα ευαίσθητα στο μέγεθος του παραθύρου αρκεί το παράθυρο να περιλαμβάνει τουλάχιστον 1-2 χαρακτήρες. Όμως, τυχόν θόρυβος που υπάρχει στο υπόβαθρο της αρχικής εικόνας μπορεί να παραμείνει και στην τελική ασπρόμαυρη εικόνα. Ο αλγόριθμος του Sauvola (Sauvola 2000) δίνει λύση σε αυτό το πρόβλημα εισάγοντας την υπόθεση στα επίπεδα φωτεινότητας του κειμένου-φόντου και του υποβάθρου (τα pixels κειμένου έχουν στάθμες του γκρι κοντά στο 0 ενώ τα pixels του υποβάθρου έχουν στάθμες του γκρι κοντά στο 255). Η εξίσωση υπολογισμού του κατωφλιού γίνεται:

$$T = m + (1 - k (1 - s/R))$$

όπου R είναι το δυναμικό εύρος της τυπικής απόκλισης ίσο με 128 και k σταθερά ίση με 0.5.



Εικόνα 1.12.13 - Διαδική μετατροπή της εικόνας 1.12.5 με Niblack για $\kappa=-0.2$ (αριστερά) και Sauvola για $\kappa=0.5$ (δεξιά) με παράθυρο 3×3 .



(α)



(β)

not so, or can not will as
fall into two classes, those
entirely to strong, and those
in each of them, hence
with variety of subdivisions
it is relative to strong - one
is involvement and non in-
traction - the other embraces all
in and without wrong requir-
as public roads and high-
ways, charities, pauperism
of the deceased, and the
government itself -
was, that if all now would put
in be some though not so much

(γ)

not so, or can not will as
fall into two classes, those
entirely to strong, and those
in each of them, hence
with variety of subdivisions
it is relative to strong - one
is involvement and non in-
traction - the other embraces all
in and without wrong requir-
as public roads and high-
ways, charities, pauperism
of the deceased, and the
government itself -
was, that if all now would put
in be some though not so much

(δ)

not so, or can not will as
fall into two classes, those
entirely to strong, and those
in each of them, hence
with variety of subdivisions
it is relative to strong - one
is involvement and non in-
traction - the other embraces all
in and without wrong requir-
as public roads and high-
ways, charities, pauperism
of the deceased, and the
government itself -
was, that if all now would put
in be some though not so much

(ε)

Εικόνα 1.12.14 - Διαδική μετατροπή της εικόνας 1.12.5 με Niblack (α) $\kappa=-0.5$ (β) $\kappa=-0.8$ με παράθυρο 3x3 και (γ) παράθυρο 5x5 (δ) παράθυρο 7x7 με $\kappa=-0.2$ και τέλος (ε) $\kappa=-0.8$ και παράθυρο 7x7.

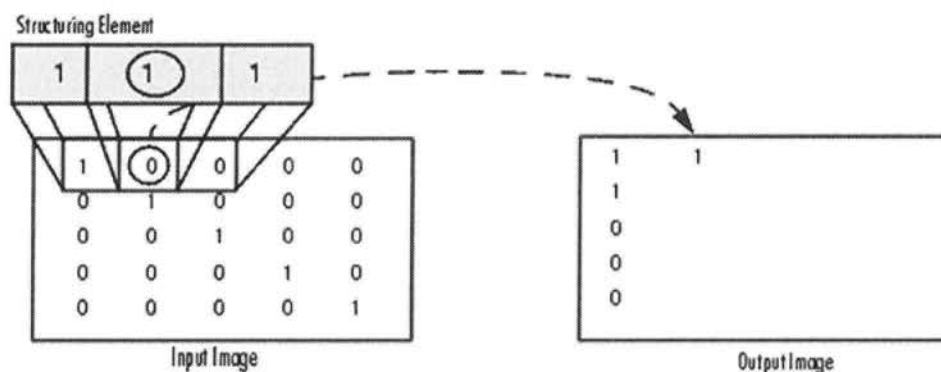
1.13 Μορφολογία Εικόνας

Η μορφολογία των εικόνων περιλαμβάνει την επεξεργασία δυαδικών εικόνων για την ανάδειξη διάφορων χαρακτηριστικών τους, όπως τα γεωμετρικά χαρακτηριστικά των αντικειμένων που εμπεριέχονται σε μία εικόνα ή τον χειρισμό των αντικειμένων τους. Η επεξεργασία των δυαδικών εικόνων καταλαμβάνει μικρή χωρητικότητα και οι χειρισμοί τους δεν απαιτούν μεγάλο χρόνο εκτέλεσης.

1.13.1 Επέκταση (Dilation)

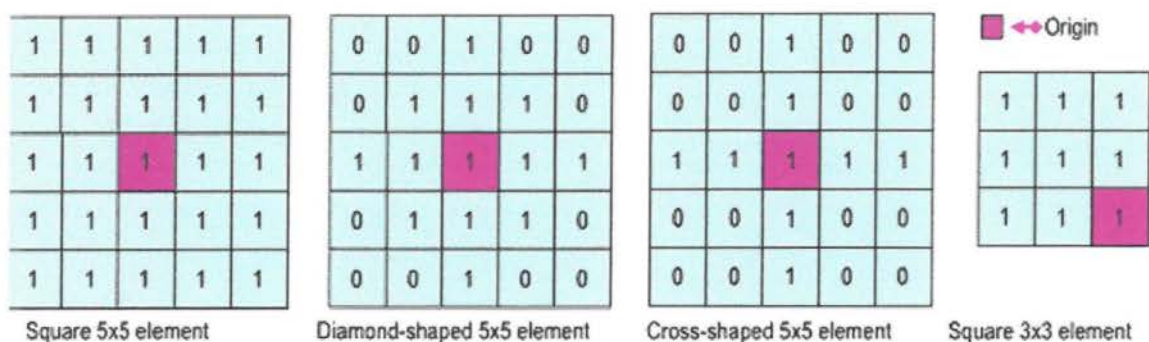
Η επέκταση είναι μία διαδικασία κατά την οποία προστίθενται pixels στα όρια των αντικειμένων, με αποτέλεσμα να αυξάνει το πάχος τους. Η διαδικασία αυτή γίνεται σε κάθε στοιχείο της δυαδικής εικόνας και η τιμή του στοιχείου μεταβάλλεται σύμφωνα με κάποιον κανόνα που ισχύει στη γειτονία του στοιχείου. Η γειτονία ορίζεται από τα στοιχεία ενός πίνακα, με τιμές 0 και 1, τα οποία έχουν τιμή 1. Η γειτονία ονομάζεται δομικό στοιχείο (Structuring Element) και μπορεί να έχει οποιοδήποτε σχήμα. Το κεντρικό pixel του δομικού στοιχείου τοποθετείται στη θέση του εξεταζόμενου pixel της δυαδικής εικόνας και εφαρμόζεται ο κανόνας για την λειτουργία της επέκτασης:

Το pixel της εξεταζόμενης δυαδικής εικόνας γίνεται 1 εάν οποιοδήποτε από τα pixel της γειτονίας του είναι 1, διαφορετικά γίνεται 0. Η γειτονία του ορίζεται από τα pixels του δομικού στοιχείου που είναι 1.



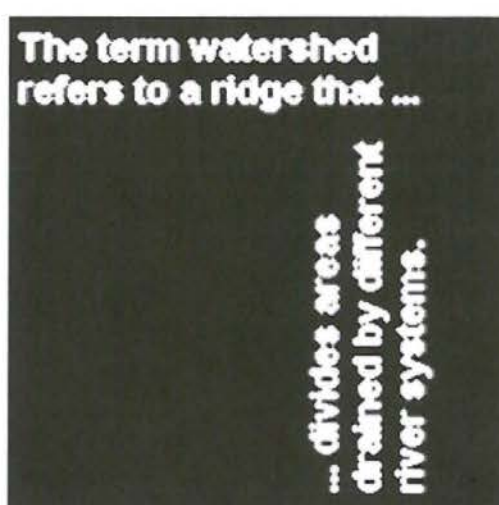
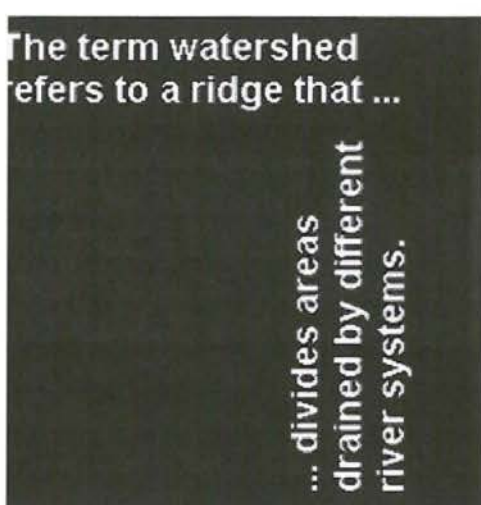
Εικόνα 1.13.1 – Μορφολογική Επέκταση Δυαδικής Εικόνας

Το παραπάνω δομικό στοιχείο κανονικά είναι 3x3, αλλά επειδή στην ουσία τα μηδενικά αγνοούνται τα έχουμε παραλήψη. Στην εικόνα 1.13.2 βλέπουμε μερικούς τύπους δομικών στοιχείων.



Εικόνα 1.13.2 – Τύποι Δομικών Στοιχείων

αρακάτω θα δούμε την επέκταση μίας δυαδικής εικόνας με δομικό στοιχείο 3x3 τύπου σταυρού (Cross-shaped). Παρατηρούμε ότι το αποτέλεσμα της επέκτασης είναι η πάχυνση των χαρακτήρων οριζόντια και κάθετα λόγω του σχήματος του δομικού στοιχείου.



Εικόνα 1.13.3 – Επέκταση με δομικό στοιχείο 3x3 σχήματος Σταυρού. Η Αρχική εικόνα στα αριστερά και εικόνα μετά την επέκταση στα δεξιά.

αρακάτω θα δούμε την επέκταση της ίδιας εικόνας με ένα δομικό στοιχείο το οποίο είναι μία γραμμή που αποτελείται από επτά pixels οριζόντια και με ένα δομικό στοιχείο που αποτελείται από επτά pixels κάθετα.



Εικόνα 1.13.3 – Επέκταση με δομικό στοιχείο: Οριζόντια γραμμή (αριστερά) και Κάθετη γραμμή (δεξιά).

1.13.2 Συρρίκνωση (Erosion)

Η συρρίκνωση είναι μία ανάλογη διαδικασία η οποία συρρικνώνει τα αντικείμενα μίας δυαδικής εικόνας σύμφωνα με κάποιο δομικό στοιχείο. Το κεντρικό pixel του δομικού στοιχείου τοποθετείται στη θέση του εξεταζόμενου pixel της δυαδικής εικόνας και εφαρμόζεται ο κανόνας για την λειτουργία της συρρίκνωσης:

Το pixel της εξεταζόμενης δυαδικής εικόνας γίνεται 1 εάν κάθε pixel της γειτονίας του είναι 1, διαφορετικά γίνεται 0. Η γειτονία του ορίζεται από τα pixels του δομικού στοιχείου που είναι 1.

Παρακάτω θα δούμε την συρρίκνωση μίας δυαδικής εικόνας με δομικό στοιχείο 3x3 τύπου Σταυρού. Παρατηρούμε ότι το αποτέλεσμα της συρρίκνωσης είναι η λέπτυνση των χαρακτήρων οριζόντια και κάθετα λόγω του σχήματος του δομικού στοιχείου.

The term watershed refers to a ridge that ...

... divides areas drained by different river systems.

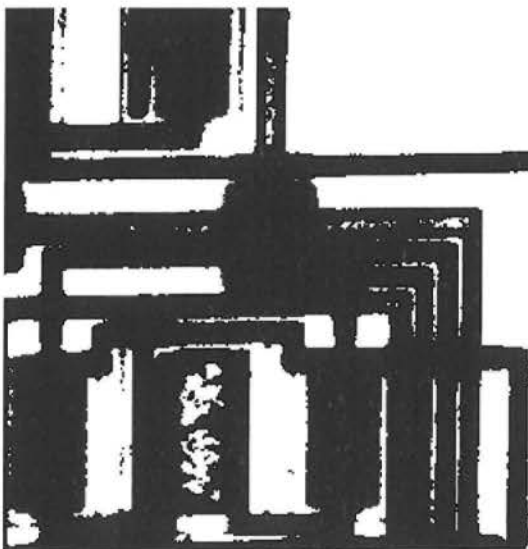
The term watershed refers to a ridge that ...

... divides areas drained by different river systems.

Εικόνα 1.13.4 – Συρρίκνωση με δομικό στοιχείο 3x3 σχήματος Σταυρού. Η Αρχική εικόνα στα αριστερά και εικόνα μετά την συρρίκνωση στα δεξιά.

1.13.3 Συνδυασμοί Μορφολογικών Χειρισμών

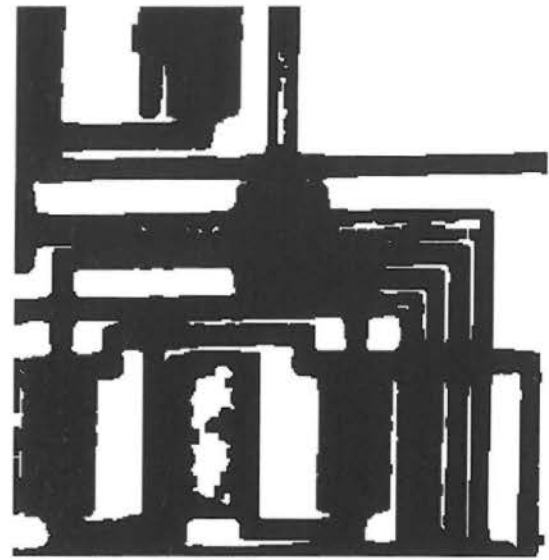
Οι μορφολογικοί χειρισμοί της επέκτασης και τη συρρίκνωσης είναι δυνατόν να συνδυαστούν για την δημιουργία διαφόρων άλλων χειρισμών όπως είναι το άνοιγμα (opening) και το κλείσιμο (closing). Το άνοιγμα αποτελείται από τη διαδικασία της συρρίκνωσης όταν ακολουθείται από τη διαδικασία της επέκτασης με το ίδιο δομικό στοιχείο, ενώ το κλείσιμο είναι η αντίστροφη διαδικασία. Η διαδικασία του ανοίγματος απομακρύνει τα μικρά αντικείμενα με την λειτουργία της συρρίκνωσης και στην συνέχεια επαναφέρει τα υπόλοιπα αντικείμενα στο αρχικό τους σχήμα με την λειτουργία της επέκτασης. Αντίστροφα, η διαδικασία του κλεισίματος καλύπτει κενά με την λειτουργία της επέκτασης και στην συνέχεια επαναφέρει τα αντικείμενα στο αρχικό τους σχήμα με την λειτουργία της συρρίκνωσης, καθώς επίσης απομακρύνει τυχόν δημιουργηθέντα στίγματα.



(α)



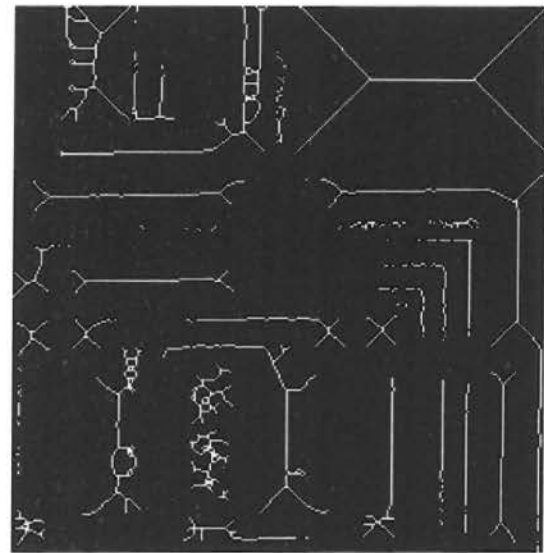
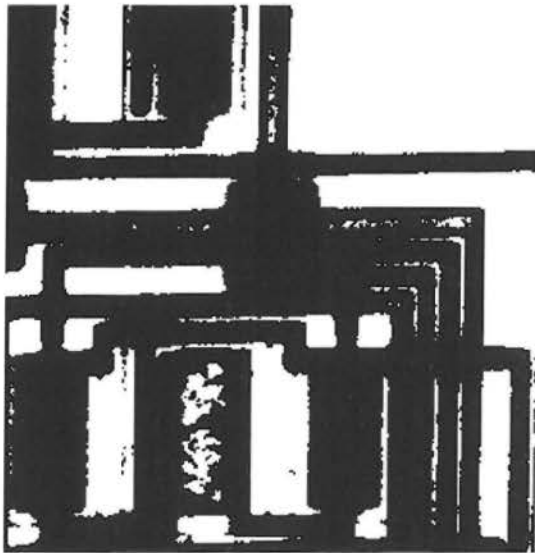
(β)



(γ)

Εικόνα 1.13.5 – Άνοιγμα και κλείσιμο με τετράγωνο δομικό στοιχείο 5x5. (α) Αρχική Εικόνα (β) Άνοιγμα και (γ) Κλείσιμο.

Παρακάτω θα δούμε την δημιουργία του σκελετού της ίδιας δυαδικής εικόνας. Με αυτήν την λειτουργία είναι δυνατόν να δημιουργηθεί ο σκελετός των αντικειμένων μίας δυαδικής εικόνας δίνοντας πληροφορίες για το σχήμα τους.



Εικόνα 1.13.6 – Δημιουργία του σκελετού της δυαδικής εικόνας. Αρχική εικόνα (αριστερά) και Σκελετος (δεξιά).

ΚΕΦΑΛΑΙΟ 2

ΟΠΤΙΚΗ ΑΝΑΓΝΩΡΙΣΗ ΧΑΡΑΚΤΗΡΩΝ

Εισαγωγή

Η Οπτική Αναγνώριση Χαρακτήρων (Optical Character Recognition - OCR) είναι η μηχανική ή ηλεκτρονική μετατροπή σαρωμένων εικόνων χειρόγραφου ή τυποποιημένου κειμένου σε κωδικοποιημένη μορφή χαρακτήρων κειμένου ηλεκτρονικού υπολογιστή.

Χρησιμοποιείται ευρέως ως ένας τρόπος εξαγωγής δεδομένων από κάποιας μορφής έντυπο, όπως έγγραφα, αποδείξεις πωλήσεων, γράμματα αλληλογραφίας, βιβλία κ.τ.λ. Είναι δηλαδή μια μέθοδος ψηφιοποίησης (digitizing) έντυπων κειμένων έτσι ώστε να μπορούν να προσπελαστούν ηλεκτρονικά, να αποθηκευτούν σε μικρότερο χώρο, να είναι προσβάσιμα από τον παγκόσμιο ιστό και να μπορούν να χρησιμοποιηθούν σε περαιτέρω διαδικασίες όπως είναι η αυτόματη μετάφραση (machine translation), η εξόρυξη δεδομένων (data mining) ή η μετατροπή κειμένου σε φωνή από τον ηλεκτρονικό υπολογιστή (text-to-speech). Επίσης σε πολλές χώρες χρησιμοποιείται για την αναγνώριση των πινακίδων κυκλοφορίας (Automatic Number Plate Recognition - ANPR) μέσω συστημάτων παρακολούθησης.

Η οπτική αναγνώριση χαρακτήρων αποτελεί πεδίο έρευνας για τους τομείς της αναγνώρισης προτύπων (pattern recognition), της τεχνητής νοημοσύνης (artificial intelligence) και της κατανόησης των εικόνων από τον ηλεκτρονικό υπολογιστή (computer vision).

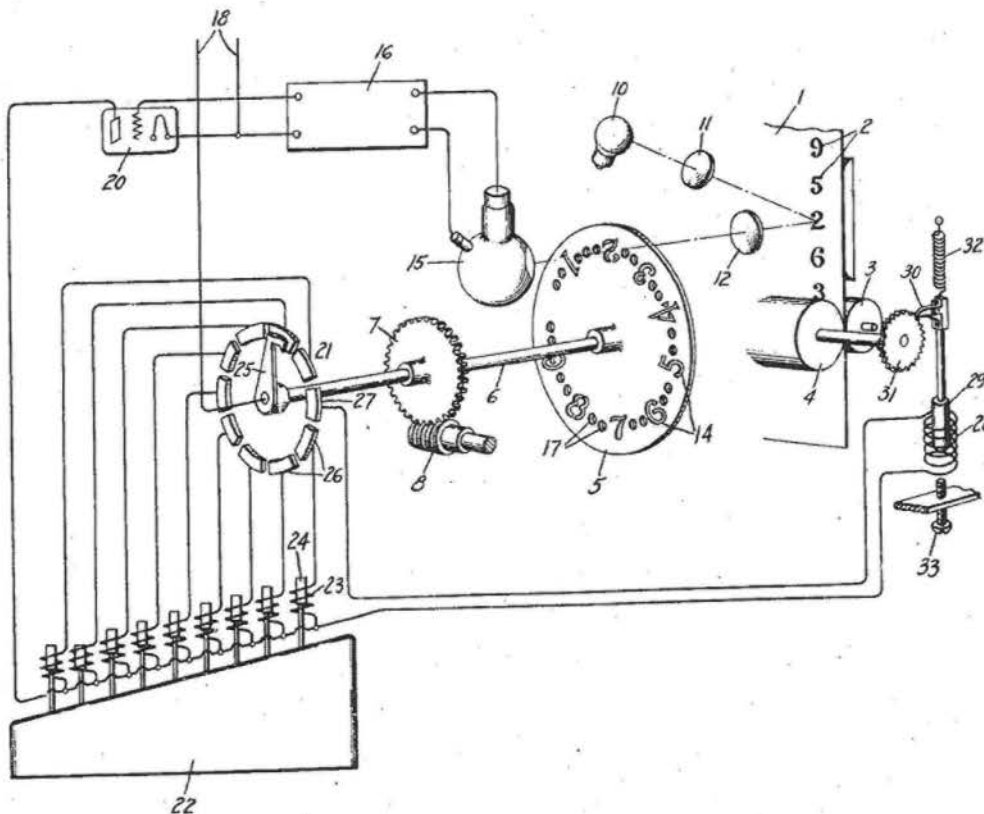
Οι πρώιμες εφαρμογές του είδους βασίζονταν στην απομνημόνευση φωτογραφιών ενός-ενός των γραμμάτων, τα οποία και έψαχναν μεμονωμένα μέσα στα προς έλεγχο κείμενα. Με τον καιρό αναπτύχθηκαν πιο έξυπνα συστήματα τα οποία εξετάζουν τις λέξεις σαν σύνολα και τα οποία προχωρούν σε «συλλογιστικά άλματα» λαμβάνοντας υπ' όψιν τους και το νόημα της λέξης στόχου στην προσπάθειά τους να κάνουν τη σωστή αναγνώριση (Για παράδειγμα μπορεί να χρησιμοποιούν λεξικό για να μαντέψουν τη σωστή λέξη, ακόμη και αν έχει αναγνωρισθεί λάθος κάποιο γράμμα της). Μερικά συστήματα είναι ικανά ακόμη και αναπαράγουν σε κάποιο βαθμό τη διαμόρφωση (print layout) της πρωτότυπης σελίδας (Δηλαδή τις εικόνες, τις στήλες ή άλλο γραφιστικό και στυλιστικό περιεχόμενο).

2.1 Ιστορία

Η εκκίνηση της ενασχόλησης με την οπτική αναγνώριση χαρακτήρων μπορεί να αναχθεί στην επέκταση της τηλεγραφίας και στις προσπάθειες ανάπτυξης συσκευών ανάγνωσης για τους τυφλούς.

Το 1914 ο Emanuel Goldberg ανέπτυξε μια μηχανή η οποία διάβαζε χαρακτήρες και τους μετέτρεπε σε κώδικα μορς. Την ίδια περίπου περίοδο, ο Edmund Fournier d'Albe ανέπτυξε το Orthophone, έναν φορητό σαρωτή εικόνας ο οποίος όταν κινούνταν κατά μήκος μια εκτυπωμένης γραμμής κειμένου παρήγαγε ηχητικούς τόνους που αντιστοιχούσαν στους χαρακτήρες που συναντούσε.

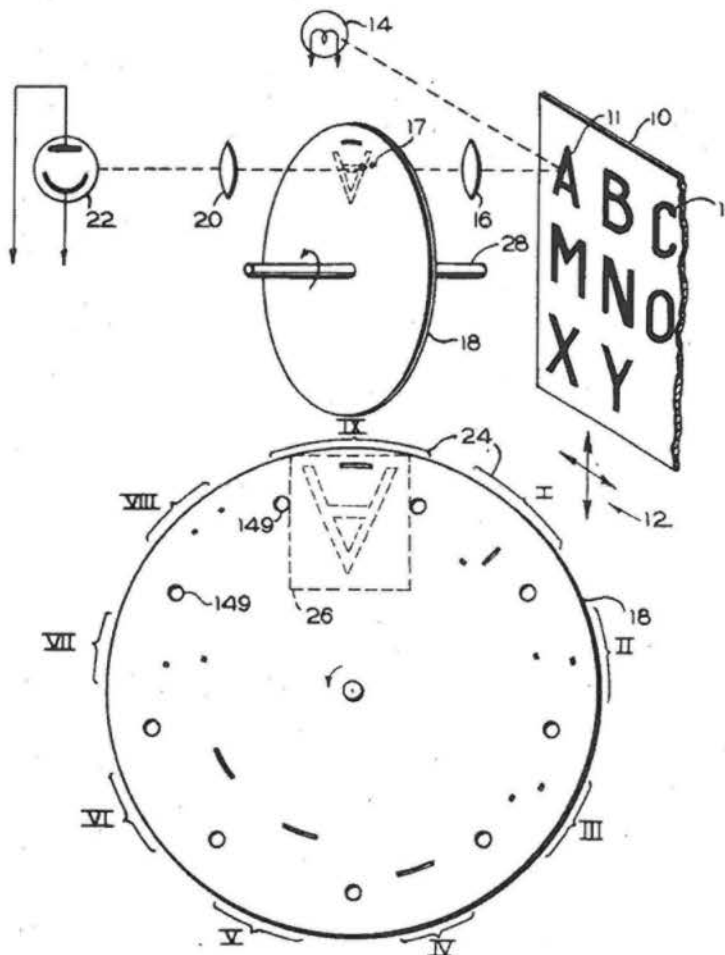
Ο Goldberg συνέχισε με την ανάπτυξη τεχνολογίας οπτικής αναγνώρισης χαρακτήρων για εισαγωγή δεδομένων. Αργότερα, πρότεινε μια μέθοδο φωτογράφισης δεδομένων τα οποία μετά, με τη χρήση φωτοκύτταρων, αντιστοιχίζονταν σε κάποια πρότυπα αναγνώρισης. Το 1929 ο Gustav Tauschek αποδείχθηκε ότι είχε παρόμοιες ιδέες και εξασφάλισε μια πατέντα οπτικής αναγνώρισης χαρακτήρων στη Γερμανία. Το ίδιο και ο Paul W.Handel, ο οποίος εξασφάλισε κι αυτός μια πατέντα ερασιτεχνίας που αφορούσε μια τεχνική αντιστοίχισης προτύπων στις Ηνωμένες Πολιτείες το 1933 (U.S. Patent 1,915,993). Το 1935 και ο Tauschek εξασφάλισε τελικά αμερικάνικη πατέντα για τη μεθόδου του (U.S. Patent 2,026,329).



Εικόνα 2.1.1 - U.S. Patent 1,915,993 - Paul W.Handel

Το 1949, ερευνητές της RCA (Radio Corporation of America) εργάστηκαν για πρώτη φορά πάνω σε μια πρώιμη μορφή ενός συστήματος οπτικής αναγνώρισης χαρακτήρων με τη βοήθεια ηλεκτρονικού υπολογιστή για λογαριασμό της Γενικής Γραμματείας Απόστρατων των Ηνωμένων Πολιτειών Αμερικής, το οποίο όμως αντί να μετατρέπει το κείμενο σε κείμενο υπολογιστή είχε ως στόχο να εκφωνεί τα γράμματα. Ήταν δηλαδή μία πρώιμη εφαρμογή text-to-speech, αλλά αποδείχθηκε πολύ ακριβή και ριζοσπαστική για την εποχή της ώστε να καθιερωθεί.

Το 1950 ο David H. Shepard, ένας εκ των κορυφαίων κρυπταναλυτών της Υπηρεσίας Ασφαλείας του Αμερικανικού Στρατού, ασχολήθηκε με το πρόβλημα της μετατροπής εκτυπωμένων μηνυμάτων σε γλώσσα μηχανής για την επεξεργασία των δεδομένων τους από ηλεκτρονικό υπολογιστή και κατασκεύασε μία μηχανή που την ονόμασε «Gismo». Έλαβε πατέντα για την κατασκευή του αυτή το 1953 (U.S. Patent 2,663,758). Ο «Gismo» μπορούσε να διαβάσει 23 από τα 26 γράμματα του λατινικού αλφαβήτου, να καταλάβει κώδικα μορς και μουσικά σύμβολα και να εκφωνήσει και να αντιγράψει εκτυπωμένες σελίδες. Ο Shepard προχώρησε ιδρύοντας την εταιρία IMR (Intelligent Machines Research Corporation) η οποία σε λίγο καιρό ανέπτυξε τα πρώτα συστήματα οπτικής αναγνώρισης χαρακτήρων για εμπορική χρήση.



Εικόνα 2.1.2 – «Gismo» - U.S. Patent 2,663,758 - David H. Shepard

Το 1955, το πρώτο εμπορικό σύστημα οπτικής αναγνώρισης χαρακτήρων εγκαταστάθηκε στα γραφεία του περιοδικού Reader's Digest και χρησιμοποιήθηκε για να εισάγει αναφορές πωλήσεων σε έναν ηλεκτρονικό υπολογιστή. Μετέτρεπε τις δακτυλογραφημένες αναφορές σε διάτρητες κάρτες τις οποίες αργότερα το προσωπικό του τμήματος συνδρομών έβαζε σ' έναν υπολογιστή που εκτελούσε την επεξεργασία των δεδομένων που αφορούσαν τις 15-20 εκατομμύρια αποστολές εντύπων που εκτελούσαν κάθε χρόνο. Το δεύτερο σύστημα πουλήθηκε στην Standard Oil και χρησιμοποιήθηκε για την ανάγνωση αποδείξεων συναλλαγών πιστωτικών καρτών. Άλλα συστήματα που δημιουργήθηκαν και πουλήθηκαν από την IMR περιελάμβαναν έναν αναγνώστη τιμολογίων για την τηλεφωνική εταιρία Bell του Ohio κι έναν σαρωτή σελίδων για την αμερικάνικη πολεμική αεροπορία ο οποίος χρησιμοποιήθηκε για την τηλεγράφιση δακτυλογραφημένων μηνυμάτων. Τα επόμενα χρόνια ο Shepard πούλησε τις πατέντες του σε διάφορες εταιρίες, μεταξύ των οποίων και στην IBM.

Γύρω στο 1965, το Reader's Digest και η RCA συνεργάστηκαν για την κατασκευή ενός αναγνώστη εγγράφων που είχε ως στόχο την ψηφιοποίηση των σειριακών αριθμών κουπονιών του περιοδικού που επιστράφηκαν από διαφημίσεις. Τα γράμματα που χρησιμοποιούταν στα έγγραφα αυτά είχαν εκτυπωθεί από έναν εκτυπωτή τυμπάνου (drum printer) της ίδιας εταιρίας με μια γραμματοσειρά ειδικά σχεδιασμένη για τους σκοπούς της οπτικής αναγνώρισης (OCR-A font). Μάλιστα ο αναγνώστης αυτός ήταν απ' ευθείας συνδεδεμένος με έναν υπολογιστή RCA 301, έναν από τους πρώτους ηλεκτρονικούς υπολογιστές τεχνολογίας ημιαγωγών (transistor technology). Ένας άλλος ειδικός αναγνώστης που επίσης αναπτύχθηκε σε συνεργασία, τοποθετήθηκε στα γραφεία της αεροπορικής εταιρίας TWA και χρησιμοποιήθηκε για την επεξεργασία του αποθέματος των καρτών επιβίβασης. Ο αναγνώστης αυτός διάβαζε έγγραφα με το ρυθμό των 1500 σελίδων το λεπτό απορρίπτοντας τις σελίδες που δεν μπορούσε να διαβάσει πλήρως. Το προϊόν αυτό μπήκε τελικά στην κανονική γραμμή παραγωγής της RCA όπου και προωθήθηκε εμπορικά τα επόμενα χρόνια για διάφορες εργασίες αντίστοιχου τύπου.

Το αμερικάνικο ταχυδρομείο χρησιμοποιούσε μηχανές οπτικής αναγνώρισης χαρακτήρων ήδη από το 1965 βασιζόμενο σε τεχνολογία του μεγάλου επιστήμονα κι εφευρέτη Jacob Rabinow. Η πρώτη χρήση τεχνολογίας οπτικής αναγνώρισης χαρακτήρων στην Ευρώπη έγινε από το βρετανικό ταχυδρομείο. Το 1965 ξεκίνησε το σχεδιασμό ενός ολόκληρου τραπεζικού συστήματος, το «National Giro» το οποίο θα χρησιμοποιούσε στον πυρήνα των συναλλαγών του τεχνολογία οπτικής αναγνώρισης χαρακτήρων, μια ιδέα ρηξικέλυθη η οποία άλλαξε από τότε τη μορφή των τραπεζικών συναλλαγών στη Βρετανία.

Το καναδικό ταχυδρομείο χρησιμοποιούσε τεχνολογία αυτόματης αναγνώρισης χαρακτήρων από το 1971. Το σύστημα διάβαζε το όνομα και τη διεύθυνση του παραλήπτη και εκτύπωνε πάνω στο φάκελο έναν γραμμικό κωδικό (bar code) με οδηγίες δρομολόγησης ανάλογα με τον ταχυδρομικό κώδικα του προορισμού. Για να μην συγχέονται οι γραμμικοί κωδικοί με σύμβολα που μπορεί να σχεδιάσει ο αποστολέας οπουδήποτε πάνω στο φάκελο η εκτύπωση αυτή γινόταν με ειδικό πορτοκαλί μελάνι το οποίο είχε πολύ υψηλά ανακλαστικά χαρακτηριστικά κάτω από υπεριώδες φως. Από αυτό το στάδιο και πέρα ήταν εύκολο τα γράμματα να ταξινομηθούν και να δρομολογηθούν μηχανικά από κάποιο σύστημα ελέγχου γραμμικών κωδικών.

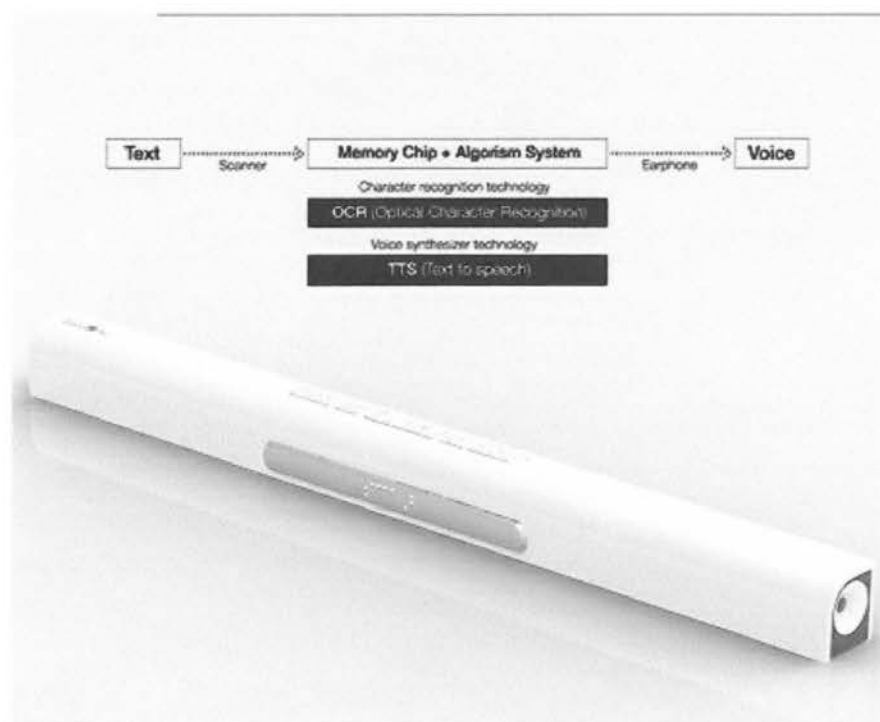
2 Η Σημασία της για τους Τυφλούς

Το 1974 ο Ray Kurzweil ίδρυσε την εταιρία Kurzweil Computer Products Inc και συνέχισε την ανάπτυξη προϊόντων οπτικής αναγνώρισης χαρακτήρων τα οποία λειτουργούσαν ανεξαρτήτως γραμματοσειράς. Αποφάσισε πως το καλύτερο προϊόν αυτής της τεχνολογίας θα ήταν αυτό που θα βοηθούσε τους τυφλούς ανθρώπους να διαβάζουν. Δηλαδή ένα προϊόν με το οποίο ένας ηλεκτρονικός υπολογιστής θα μπορούσε να εκφωνεί κείμενα τους τυφλούς.

Για να υλοποιηθεί μια τέτοια συσκευή ήταν απαραίτητη η ανακάλυψη δύο ακόμα τεχνολογιών. Του επίπεδου σαρωτή CCD (CCD flatbed scanner) και του ηλεκτρονικού συνθέτη φωνής (voice synthesizer). Στις 13 Ιανουαρίου 1976, το τελειωμένο προϊόν προκαλύφθηκε στο κοινό με μεγάλη δημοσιότητα κατά τη διάρκεια μιας κοινής συνέντευξης τύπου του Kurzweil και της αμερικανικής ένωσης τυφλών.

Το 1978 ξεκίνησε ανοικτά η διάθεση του προγράμματος στην αγορά. Ένας από τους πρώτους αγοραστές του προγράμματος ήταν η εταιρία LexisNexis η οποία το αγόρασε για να ανεβάσει επίσημα έγγραφα και αποκόμματα εφημερίδων στη εκκολαπτόμενη τότε online βάση δεδομένων τους.

Το 1980, ο Kurzweil πούλησε την εταιρία του στη Xerox την οποία ενδιέφερε έντονα εκείνη την περίοδο η περαιτέρω εμπορευματοποίηση της μετατροπής εντύπων σε αρχεία ηλεκτρονικού υπολογιστή. Τελικά η Xerox απεμπόλησε τα δικαιώματα της μεταβιβάζοντας τα στην εταιρία Scansoft, η οποία στη συνέχεια συγχωνεύτηκε με την εταιρία Nuance Communications όπου και παραμένουν μαζί μέχρι και σήμερα.



Εικόνα 2.2.1 – Voice Stick , φορητή συσκευή ανάγνωσης για τυφλούς

2.3 Η Σημασία της για τους Οργανισμούς και τις Επιχειρήσεις

Τα πλεονεκτήματα που προσφέρει η χρήση εφαρμογών οπτικής αναγνώρισης χαρακτήρων σε μια επιχείρηση ή έναν οργανισμό είναι πολλά. Οι διαδικασίες που καθίστανται δυνατές, ή που βελτιώνονται, από την ύπαρξη αυτής της τεχνολογίας αναμφισβήτητα αυξάνουν την παραγωγικότητα και την αποδοτικότητα της ανθρώπινης εργασίας, ανεξαρτήτως του πεδίου δραστηριοποίησης της εταιρίας ή του οργανισμού όπου εφαρμόζονται. Σε γενικές γραμμές, οι εφαρμογές οπτικής αναγνώρισης χαρακτήρων λειτουργούν επικουρικά βελτιώνοντας και σε κάποιες περιπτώσεις αποκλειστικά καθιστώντας την εφικτή, την ηλεκτρονική οργάνωση μιας επιχείρησης, κάποια από τα πλεονεκτήματα της οποίας είναι τα εξής:

- Μειωμένα έξοδα λειτουργίας - Ελαχιστοποιούνται τα έξοδα αρχειοθέτησης και μακρόχρονης αποθήκευσης εγγράφων, αφού καθίσταται δυνατό να διατηρούνται ή να παράγονται επιτόπου σε έντυπη μορφή μόνο τα απαραίτητα σε αυτή τη μορφή ζητούμενα από το κράτος και τους πελάτες έγγραφα. Εξαιτίας αυτής της τεχνολογίας έγινε πρακτικά εφικτή και οικονομικά βιώσιμη η μετατροπή ολόκληρου του αρχείου εγγράφων ενός οργανισμού σε ηλεκτρονική μορφή. Ο χειροκίνητος τρόπος εισαγωγής των εγγράφων στο ηλεκτρονικό σύστημα είναι εξ' αρχής αποτρεπτικός αφού για ένα ολόκληρο αρχείο θ' απαιτούσε τεράστια ανθρώπινη προσπάθεια η οποία μεταφράζεται άμεσα σε χρόνο και χρήμα. Υπολογίζεται ότι η ανθρώπινη αναζήτηση σ' ένα μεγάλο φυσικό αρχείο κοστίζει για μια εταιρία μέχρι και 20\$ ανά έγγραφο, εάν το έγγραφο ήταν αποθηκευμένο στη σωστή θέση και 120\$ αν είχε αποθηκευθεί σε λάθος θέση, ενώ κοστίζει περίπου 250\$ η αναπαραγωγή με τρίτους τρόπους, όπως διασταύρωση άλλων εγγράφων κ.τ.λ., ενός χαμένου εγγράφου.
- Αύξηση αποδοτικότητας – Μία από τις περισσότερο επαναλαμβανόμενες και λιγότερο αποδοτικές εργασίες γραφείου σήμερα είναι η εισαγωγή δεδομένων φορμών στα ηλεκτρονικά συστήματα των οργανισμών. Έχει υπολογιστεί στις Ηνωμένες Πολιτείες ότι ο μέσος εργαζόμενος χρειάζεται περίπου 12 λεπτά για να συμπληρώσει μία ηλεκτρονική φόρμα από δεδομένα που διαβάζει επιτόπου σε κάποιο έγγραφο.
- Βελτίωση της προσβασιμότητας των εγγράφων – Εφ' όσων υπάρχουν σε ηλεκτρονική μορφή, τα έγγραφα μπορούν να προσπελαστούν ταυτόχρονα από πολλούς εργαζόμενους, εξαλείφοντας την ανάγκη για φωτοτυπίες. Η ελαχιστοποίηση της χρήσης των φωτοτυπιών έχει και περιβαλλοντικό αντίκτυπο αφού μειώνει την κατανάλωση χαρτιού και μελάνης, όπως και τις εκπομπές διοξειδίου του άνθρακα εξαιτίας της χρήσης των φωτοτυπικών μηχανημάτων.
- Αύξηση παραγωγικότητας - Η ικανότητα άμεσης αναζήτησης στα περιεχόμενα ενός εγγράφου, ιδιαίτερα σε χώρους όπου γίνεται κατά συρροή χρήση εντύπων, μπορεί να είναι καθοριστική. Η ικανότητα ηλεκτρονικής τροποποίησης εγγράφων, αντιγραφής τμημάτων τους και γενικά όλα τα πλεονεκτήματα που πηγάζουν από την ύπαρξη σε ηλεκτρονική μορφή ολόκληρης της βάσης δεδομένων των εγγράφων ενός οργανισμού καθίστανται ικανά εξ' αιτίας της ύπαρξης αυτής της τεχνολογίας.
- Ανάκτηση πολύτιμου χώρου – Οι αναλογικά μεγάλοι χώροι αποθήκευσης εγγράφων καθίστανται πλέον αχρείαστοι, οπότε μπορεί να δημιουργηθεί άμεσα, χωρίς

επιπλέον έξοδα, πολύτιμος εργασιακός χώρος. Ένα απλό DVD μπορεί να χωρέσει έγγραφα που σε έντυπη μορφή θα γέμιζαν 15 αρχειοθήκες.

- Αποφυγή καταστροφών – Εξαιτίας της ευκολίας τήρησης αρχείων ανάκτησης (back up), αυξάνονται σημαντικά οι πιθανότητες να μην χαθεί σημαντικό υλικό εξαιτίας μιας καταστροφής όπως πυρκαγιά, πλημμύρα κ.τ.λ. Πλέον με την ανάπτυξη των τεχνολογιών cloud καθίσταται δυνατό η φυσική αποθήκευση του αρχείου να μην γίνεται καν στο χώρο εργασίας αλλά οπουδήποτε μέσα στο internet, όπου και πρακτικά είναι άφθαρτο.
- Ασφάλεια – Εξαιτίας του ηλεκτρονικού τρόπου της πρόσβασης, μετά την αρχική τους είσοδο στο σύστημα και την καταστροφή του φυσικού έγγραφου, είναι δυνατό να ασφαλίσουμε τα ευαίσθητα αρχεία με κωδικούς, ή να επιτρέπουμε την πρόσβαση σε αυτά επιλεκτικά, μόνο στους χρήστες που επιθυμούμε, μειώνοντας τις πιθανότητες να πέσουν εταιρικές πληροφορίες σε λάθος χέρια.
- Συνέπεια – Υπολογίζεται ότι περίπου το 3% με 7,5% των εγγράφων ενός οργανισμού έχει χαθεί, δεν μπορεί να ανακτηθεί, ή είναι αρχειοθετημένο σε λάθος θέση. Εξαιτίας της άμεσης ανταπόκρισης που έχει μια τέτοια οργάνωση σε σχέση με τη χειροκίνητη εκδοχή της κάνει τις επιχειρήσεις που την χρησιμοποιούν να μοιάζουν σ' έναν εξωτερικό παρατηρητή πιο οργανωμένες και επαγγελματικές.

εβαίως, η ηλεκτρονική οργάνωση μια επιχείρησης ή ενός οργανισμού ανοίγει το δρόμο σε νέες, μοντέρνες προκλήσεις ασφαλείας, οι οποίες πάντως κατά γενική ομολογία είναι συγκρίσιμες, αν όχι μικρότερες, από τις αντίστοιχες της φυσικής.



2.4 Είδη Εφαρμογών και οι Δυνατότητες τους

Η τεχνολογία οπτικής αναγνώρισης χαρακτήρων ενσωματώνεται σε προϊόντα με διάφορες μορφές.

Desktop εφαρμογές – Τα συστήματα οπτικής αναγνώρισης χαρακτήρων είναι αναλυτικές μηχανές τεχνητής νοημοσύνης (artificial intelligence) που εξετάζουν σειρές χαρακτήρων αντί για ολόκληρες λέξεις ή φράσεις. Στηρίζονται στην ανάλυση του σχήματος των χαρακτήρων και βασίζονται σε λεξικά για να κάνουν τις καλύτερες δυνατές προβλέψεις. Είναι οι πρώτες εφαρμογές που κυκλοφόρησαν ευρέως και χρησιμοποιούνται εδώ και πολλά χρόνια για επαγγελματικές και καθημερινές εργασίες. Αρκετές εφαρμογές τέτοιου είδους συμπεριλαμβάνονται συχνά στο λογισμικό εγκατάστασης συσκευών σάρωσης, εκτύπωσης ή συνδυασμού.

Online εφαρμογές – Με την ανάπτυξη της πληροφορικής, η χρήση της τεχνολογίας οπτικής αναγνώρισης χαρακτήρων μεταφέρθηκε από έναν υπολογιστή σε μια πληθώρα από πλατφόρμες όπως ο παγκόσμιος ιστός, το διαδικτυακό νέφος (cloud computing) και οι φορητές συσκευές (mobile devices). Μετά από 30 χρόνια ανάπτυξης για επιτραπέζιους υπολογιστές, οι εφαρμογές οπτικής αναγνώρισης χαρακτήρων άρχισαν να προσαρμόζονται στη νέα πραγματικότητα. Η διαδικτυακή οπτική αναγνώριση χαρακτήρων (Web OCR, Online OCR, Web Based OCR) έχει γίνει η νέα τάση του χώρου για την κάλυψη των αναγκών μεγαλύτερων ομάδων χρηστών. Η ευρεία εξάπλωση των ευρυζωνικών συνδέσεων διαδικτύου κατέστησε δυνατή τη δημιουργία εφαρμογών οπτικής αναγνώρισης χαρακτήρων σε αποκλειστικά διαδικτυακή μορφή για επιχειρήσεις και απλούς χρήστες. Από το 2000 και μετά ένας σημαντικός αριθμός νέων εταιριών μπήκε στο παιχνίδι της προσφοράς διαδικτυακών εφαρμογών οπτικής αναγνώρισης χαρακτήρων, με πολλές από αυτές να τις προσφέρουν ακόμα και δωρεάν στους τελικούς χρήστες, αντλώντας τα έσοδα τους από διαφημίσεις.

Εφαρμογές προσαρμοσμένες στην περίπτωση (application oriented OCR) – Από τη στιγμή που η τεχνολογία οπτικής αναγνώρισης χαρακτήρων απέκτησε ευρεία αποδοχή και άρχισε να χρησιμοποιείται σε πολλές διαφορετικές συνθήκες και έργα, προέκυψε η ανάγκη για συστήματα περισσότερο εξειδικευμένα στις ανάγκες κάποιων πεδίων. Για παράδειγμα, περίπλοκα φόντα σελίδων, φωτογραφίες φθαρμένες, τσαλακωμένες, λερωμένες, χαμηλής ανάλυσης, θολές, κείμενα με ορολογίες που δεν συναντώνται στο τυπικό λεξικό, ειδικές γραμματοσειρές, όλα αυτά είναι συνθήκες που συναντώνται συχνά σε πραγματικά περιβάλλοντα εργασίας και καθιστούν πολύ δύσκολη τη σωστή λειτουργία τυποποιημένων προγραμμάτων γενικής χρήσης. Έτσι αναπτύχθηκαν εφαρμογές βελτιστοποιημένες για συγκεκριμένες εισόδους όπως τιμολόγια (invoice OCR), επαγγελματικές κάρτες (business-card OCR), στιγμιότυπα οθόνης (screenshot OCR), ταυτότητες (ID card OCR), διπλώματα οδήγησης (driver license OCR) κ.τ.λ.

Ιδρυμένο από το υπουργείο ενέργειας των Ηνωμένων Πολιτειών Αμερικής (DOE), το ινστιτούτο έρευνας της επιστήμης της πληροφορίας (ISRI) είχε αναλάβει υπό τη σκέπη του την εξέλιξη της τεχνολογίας της αυτόματης κατανόησης εκτυπωμένων κειμένων από τον ηλεκτρονικό υπολογιστή και ήταν το καθ' ύλην αρμόδιο όργανο για τη διεξαγωγή μετρήσεων ακριβείας των λογισμικών οπτικής αναγνώρισης χαρακτήρων από το 1992 μέχρι το 1996.

Η αναγνώριση κειμένου γραμμένου με λατινικούς χαρακτήρες δεν είναι ακόμη και σήμερα 100% ακριβείς, από οποιοδήποτε σύστημα, ακόμη και όταν χρησιμοποιούνται καθαρές εικόνες. Μια μελέτη που χρησιμοποίησε ως δείγματα άρθρα από καλοδιατηρημένες εφημερίδες του 19ου και των αρχών του 20ου αιώνα αποφάνθηκε ότι η επιτυχία αναγνώρισης των εμπορικών προγραμμάτων οπτικής αναγνώρισης χαρακτήρων ποικίλει μεταξύ 71% και 98%. 100% αναγνώριση μπορεί να επιτευχθεί μόνο με τη συμμετοχή ανθρώπου στη διαδικασία. Τομείς όπως η αναγνώριση χειρόγραφων «ρέουσας» γραφής (μονοκονδυλιές - cursive handwriting) ή η αναγνώριση χαρακτήρων ανατολικοασιατικού τύπου (east asian), όπου το κάθε γράμμα μπορεί να αποτελείται από πολλά κομμάτια, αποτελούν ακόμη και σήμερα ενεργό πεδίο έρευνας.

Μετρήσεις ακρίβειας μπορούν να γίνουν με διαφορετικούς τρόπους οι οποίοι επηρεάζουν σημαντικά τα αποτελέσματα. Για παράδειγμα, εάν δεν επιτρέπεται στο προς μέτρηση πρόγραμμα να χρησιμοποιεί λεξικό για τη διόρθωση των λέξεων τις οποίες βρήκε σχεδόν σωστά, επιτυχία αναγνώρισης της τάξης του 95% θα μπορούσε σε πραγματικές συνθήκες να αντιστοιχεί σε επίδοση της τάξης του 99%. Ο λόγος που χρησιμοποιούνται και τέτοιες μετρήσεις είναι ώστε να βγάλουν οι ερευνητές συμπεράσματα για την ποιότητα των επιμέρους μηχανισμών που εμπλέκονται στη διαδικασία της αναγνώρισης. Δηλαδή να γνωρίζουν ποια εφαρμογή διαχειρίζεται πιο άρτια το λεξικό, αλλά να γνωρίζουν και ποια έχει τον ακριβέστερο μηχανισμό «ωμής αναγνώρισης» (raw recognition). Η γνώση από αυτήν την ανάλυση θα μπορούσε κάλλιστα να οδηγήσει στη συνδυαστική κατασκευή ενός ανώτερου προγράμματος το οποίο θα χρησιμοποιούσε τον ισχυρό μηχανισμό ωμής αναγνώρισης μιας εφαρμογής και τις έξυπνες μετά-διορθώσεις μιας άλλης.

Η τεχνολογία απευθείας αναγνώρισης χαρακτήρων (On-line Character Recognition – On-line CR) συχνά συγχέεται με την οπτική αναγνώριση χαρακτήρων, όμως δεν είναι έτσι. Η οπτική αναγνώριση χαρακτήρων, ακόμη και στη διαδικτυακή της μορφή (Online OCR), είναι μια μορφή «στατικής» αναγνώρισης (off-line character recognition), δηλαδή αναγνώρισης του σχήματος ενός στατικού στιγμιότυπου μιας μορφής (fixed static shape), ενώ η απευθείας αναγνώριση χαρακτήρων αναγνωρίζει δυναμικά τη μορφή που σχηματίζεται, όχι τόσο από το σχήμα, αλλά από τις αλλαγές κατεύθυνσης του μέσου γραφής (dynamic motion during handwriting). Για παράδειγμα, η απευθείας αναγνώριση χαρακτήρων που χρησιμοποιείται στις «χειρονομίες» (gestures) των ταμπλετών (tablet PC) αναγνωρίζει εκτός της μορφής και την κατεύθυνση που τρέχει το στυλό, όπως για παράδειγμα από αριστερά προς τα δεξιά ή το αντίστροφο. Η απευθείας αναγνώριση χαρακτήρων συναντάται συχνά και με άλλες ονομασίες όπως δυναμική αναγνώριση χαρακτήρων (dynamic character recognition), αναγνώριση χαρακτήρων πραγματικού χρόνου (realtime character recognition) ή ευφυής αναγνώριση χαρακτήρων (intelligent character recognition - ICR).

Προγράμματα απευθείας αναγνώρισης χαρακτήρων έχουν γίνει ευρέως γνωστά τα τελευταία χρόνια με την έκρηξη της αγοράς των έξυπνων κινητών τηλεφώνων (smartphones) και των ταμπλετών. Οι αλγόριθμοι που χρησιμοποιούνται σε αυτές τις συσκευές εκμεταλλεύονται το γεγονός ότι τη στιγμή της εισαγωγής των δεδομένων γίνονται γνωστά στη συσκευή η κίνηση, η ταχύτητα και η κατεύθυνση της γραφίδας σχεδίασης και ότι ο χρήστης μπορεί να εκπαιδευτεί να σχεδιάζει τα σύμβολα που περιμένει η εφαρμογή με τον τρόπο που βολεύει την αναγνώριση. Αυτά, δυστυχώς δεν μπορούν να εφαρμοστούν στην αναγνώριση σαρωμένων εγγράφων οπότε η οπτική αναγνώριση χειρόγραφων

χαρακτήρων είναι ακόμη ένα αρκετά ανοικτό πρόβλημα. Ακρίβεια της τάξης του 80% με 90% είναι δυνατόν να επιτευχθεί για καθαρούς, καλούς, χειρόγραφους χαρακτήρες κάτι που όμως και πάλι μεταφράζεται σε δεκάδες σφάλματα ανά σελίδα, κάτι που καθιστά μια τέτοια εφαρμογή δύσχρηστη σε πραγματικές συνθήκες εργασίας.

Η αναγνώριση «ρέουσας» γραφής, δηλαδή γραφής όπου η κάθε λέξη γίνεται με μία κίνηση του μέσου γραφής (μονοκονδυλιά), είναι ένα πεδίο έρευνας που βρίσκεται αυτή τη στιγμή στην ακμή του με τυπικά ποσοστά αναγνώρισης αρκετά μικρότερα από την αναγνώριση σαρωμένου χειρόγραφου εγγράφου (το οποίο έχει γραφτεί με διακριτά τα γράμματα – όπως π.χ. όταν συμπληρώνουμε φόρμες με κουτάκια). Σε αυτό τον τύπο γραφής, προτιμάται να εξετάζεται σαν ενιαίο δείγμα ολόκληρη η λέξη, αντί να επιχειρείται η τμηματοποίηση της και εξαγωγή των γραμμάτων της ένα ένα. Επίσης είναι συχνή η χρήση λεξικού προσαρμοσμένου στις ανάγκες της συγκεκριμένης εφαρμογής.

Για παράδειγμα, στο πρόβλημα της ανάγνωσης των στοιχείων μιας επιταγής η βελτίωση της αναγνώρισης που συντελείται με τη χρήση λεξικού είναι πολύ μεγάλη, αφού είναι πολύ λίγες οι λέξεις που μπορεί να γράψει κάποιος πάνω σε μια επιταγή. Η γνώση γραμματικής και συντακτικού από τον αλγόριθμο βοηθά κατακόρυφα στην κατανόηση της «ρέουσας» γραφής γιατί έχει αποδειχθεί ότι αποκλειστικά και μόνο από τη χρήση σχεδιαστικής πληροφορίας δεν είναι δυνατόν να επιτευχθεί ακρίβεια μεγαλύτερη του 98% για τη γραφή ενός μέσου ανθρώπου. Κατά κάποιο τρόπο, μέσος όρος ακρίβειας της τάξης του 98% είναι το ανώτατο θεωρητικό όριο που μπορεί να επιτύχει ένας τέλειος αλγόριθμος αναγνώρισης «ρέουσας» γραφής χωρίς να χρησιμοποιεί γραμματική και συντακτική πληροφορία. Αυτό συμβαίνει γιατί οι άνθρωποι έχουμε την τάση όταν γράφουμε με «ρέοντα» τρόπο μία λέξη, είτε να μη την ολοκληρώνουμε πλήρως, είτε να τη γράφουμε παραλείποντας γράμματα. Επίσης, είναι συχνό να προσθέτουμε κοσμητικά στοιχεία στους χαρακτήρες (π.χ. τραβηγμένες μεγάλες κυματιστές γραμμές στο τελικό 'ς', μεγάλες γυριστές ουρές που υπερκαλύπτουν τα υπόλοιπα γράμματα στο 'λ' κ.τ.λ.), όπως κι επίσης είναι γενικά αποδεκτό ότι ενώ τα «καλά», διακριτά, γράμματα μας μοιάζουν αρκετά με το πρότυπο, ο γρήγορος «ρέοντας» τρόπος γραφής μας απέχει σημαντικά από άνθρωπο σε άνθρωπο. Σε τέτοιες περιπτώσεις, εάν το πρόγραμμα δε ξέρει τι περίπου να περιμένει είναι αδύνατο να αναγνωρίσει τα πάντα.

Μια τεχνική η οποία έχει αξιοσημείωτη επιτυχία στην αναγνώριση δύσκολων λέξεων, από έγγραφα γενικά ακατάλληλα για αυτόματη αναγνώριση, είναι η αυτόματη αποστολή των μη αναγνωρισμένων από το πρόγραμμα λέξεων σε ανθρώπους για περεταίρω αναγνώριση, όπως γίνεται στο σύστημα reCAPTCHA.



2.5 Αναγνώριση με Τεχνητά Νευρωνικά Δίκτυα

Για την οπτική αναγνώριση χαρακτήρων χρησιμοποιούνται κατά κόρων αλγόριθμοι που υλοποιούν διατάξεις τεχνητών νευρωνικών δικτύων (artificial neural networks). Υπάρχουν πολλών ειδών τεχνητά νευρωνικά δίκτυα τα οποία στις πραγματικές εφαρμογές συνήθως συνδυάζονται έχοντας το καθένα δικαίωμα μιας «ζυγισμένης» ψήφου (weighted vote) στο τελικό αποτέλεσμα αλλά όλων η λογική είναι παρεμφερής και παρουσιάζεται παρακάτω.

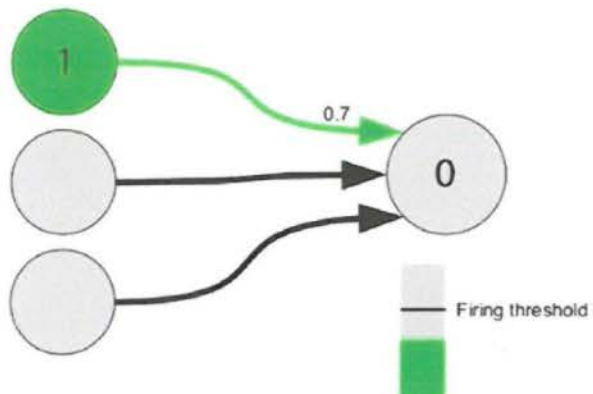
Τα τεχνητά νευρωνικά δίκτυα είναι μαθηματικοί ταξινομητές (classifiers) των οποίων η έμπνευση προήλθε από ένα μοντέλο λειτουργίας των βιολογικών νευρώνων του εγκεφάλου. Οι νευρώνες είναι βιολογικές οντότητες οι οποίες σε λογικό επίπεδο μπορούν να περιγραφούν σαν «πύλες» (logic gates) με από μία έως πολλές εισόδους και μία έξοδο. Οι συνδέσεις των νευρώνων λέγονται συνάψεις. Εκτός από απευθείας εισόδους κι έξοδο, οι νευρώνες μπορούν να έχουν σαν είσοδο την έξοδο ενός άλλου νευρώνα και αυτοί με τη σειρά τους να αποτελούν είσοδο για άλλους, σχηματίζοντας με αυτόν τον τρόπο δίκτυα.

Στα έμβια όντα αυτά τα δίκτυα μπορεί να αποτελούνται από δισεκατομμύρια νευρώνες συνδεδεμένους μεταξύ τους με πολλαπλάσιες συνάψεις. Υπάρχουν ενδείξεις ότι η πολυπλοκότητα του νευρωνικού δικτύου συνδέεται με το νοητικό επίπεδο του οργανισμού που το φέρει. Είναι άλωση εύλογο αφού ο ανθρώπινος εγκέφαλος έχει το πολυπλοκότερο γνωστό νευρωνικό δίκτυο που αποτελείται κατά μέσο όρο από 100.000.000.000 νευρώνες. Τα σήματα μεταξύ των νευρώνων στον εγκέφαλο, σε αντίθεση με αυτό που πιστεύαμε παλιότερα επηρεασμένοι από τον τρόπο λειτουργίας των ηλεκτρικών κυκλωμάτων, δεν είναι ηλεκτρικά αλλά χημικά.

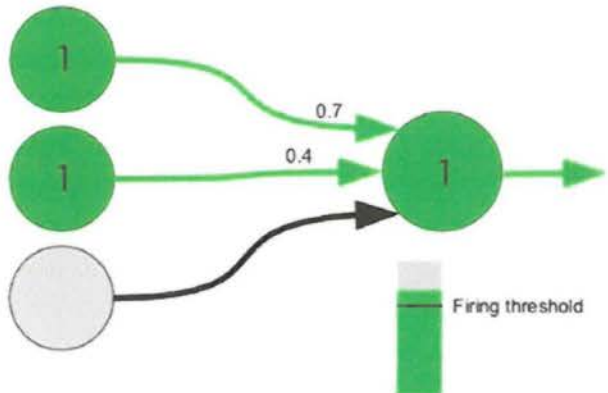
Οι χημικές ουσίες που διακινούνται μεταξύ των νευρώνων ονομάζονται νευροδιαβιβαστές (neurotransmitters). Βεβαίως, οι δυνάμεις που αναπτύσσονται και οι οποίες είναι υπεύθυνες για την εξαπόλυση, τη πρόσληψη, τη μετακίνηση και τη συνοχή αυτών των ουσιών είναι ηλεκτρικές αλλά το νόημα είναι ότι οι νευρώνες δε διακινούν ηλεκτρόνια, όπως τα ηλεκτρικά κυκλώματα, αλλά μόρια χημικών νευροδιαβιβαστών κάποιοι από τους οποίους είναι η σεροτονίνη (serotonin), η νορεπινεφρίνη (norepinephrine), η ντοπαμίνη (dopamine) και άλλοι. Μάλιστα εικάζεται ότι η ελαττωμένη σε σχέση με το φυσιολογικό ποσότητα αυτών των ουσιών στις συνάψεις των νευρώνων είναι η αιτία αρκετών ψυχικών διαταραχών. Αν και κανείς δεν έχει κάνει ως τώρα κάποια μέτρηση που να το αποδεικνύει σ' έναν εν λειτουργία εγκέφαλο έχει αναπτυχθεί ολόκληρη κατηγορία φαρμάκων που έχουν ως στόχο την αύξηση της ποσότητας των νευροδιαβιβαστών που πλέουν μεταξύ των συνάψεων. Η κατηγορία αυτών των φαρμάκων λέγεται «επιλεκτικοί αναστολείς επαναπρόσληψης νευροδιαβιβαστή» (συνήθως σεροτονίνης ή νορεπινεφρίνης) (Selective Serotonin Reuptake Inhibitor – SSRI and Selective Norepinephrine Reuptake Inhibitor – SNRI) και θεωρούνται το παρόν της ψυχιατρικής καθώς είναι πιο αποτελεσματικά και πολύ πιο ελαφριά σε παρενέργειες από τα φάρμακα παλαιού τύπου που αντικαθιστούν.

Η λογική λειτουργίας του βιολογικού νευρώνα είναι η εξής. Ο νευρώνας έχει έναν αριθμό εισόδων από τις οποίες δέχεται χημικά σήματα. Τα σήματα αυτά ανάλογα με το από ποια είσοδο (σύναψη) προέρχονται λαμβάνονται υπ' όψιν με διαφορετικό συντελεστή βαρύτητας. Δηλαδή μπορεί ο νευρώνας να είναι πιο ευαίσθητος σε σήματα που προέρχονται για παράδειγμα από τα μάτια και λιγότερο ευαίσθητος σε σήματα που προέρχονται από τη μύτη. Όταν το άθροισμα της έντασης των σημάτων, όπως αυτά υπολογίστηκαν ανάλογα με τα βάρη τους το καθένα, ξεπεράσει μια τιμή που λέγεται

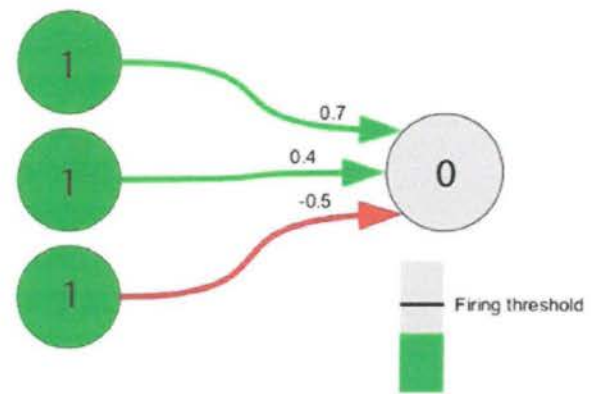
«κατώφλι πυροδότησης» (firing threshold) τότε ο νευρώνας ενεργοποιεί την έξοδό του, δηλαδή στέλνει το δικό του χημικό σήμα παρακάτω. Όσο το άθροισμα των σημάτων που λαμβάνει σαν είσοδο είναι κάτω από αυτό το όριο δεν στέλνει κανένα σήμα.



Εικόνα 2.5.1 - Η είσοδος 1 επι το βάρος της 0.7 ($1 \times 0.7 = 0.7$) δεν ξεπερνάει το κατώφλι πυροδότησης.



Εικόνα 2.5.2 - Οι είσοδοι επι το βάρος τους αθροιστικά ($1 \times 0.7 + 1 \times 0.4 = 1.1$) ξεπερνάνε το κατώφλι πυροδότησης



Εικόνα 2.5.3 - Οι είσοδοι μπορούν να έχουν αρνητικό βάρος

Το πώς αυτή η διάταξη μπορεί να οικοδομεί λογική γίνεται λίγο πιο προφανές με το παρακάτω παράδειγμα. Έστω ένας νευρώνας που έχει δύο μόνο εισόδους, κατώφλι πυροδότησης 1 και τα σήματα που λαμβάνει στις εισόδους του μπορεί να είναι 0 ή 1. Αν τα βάρη των εισόδων του οριστούν στο 0,5 για να ενεργοποιηθεί η έξοδός του θα πρέπει να ισχύει η ανίσωση:

$$(Είσοδος A \times 0.5) + (Είσοδος B \times 0.5) \geq 1$$

Αυτό γίνεται μόνο αν και οι δύο εισοδοί του είναι ενεργοποιημένοι (1), οπότε ένας νευρώνας σε τέτοια «ρύθμιση» λειτουργεί σαν λογική πύλη AND (AND logic gate). Αν στον ίδιο νευρώνα τα βάρη ήταν 1 και 1 τότε ο νευρώνας θα λειτουργούσε σαν λογική πύλη OR (OR logic gate). Αν στη μία είσοδο του δίνουμε πάντα 1 με βάρος 1 και αλλάζουμε το βάρος της άλλης εισόδου σε -1 τότε ο νευρώνας θα λειτουργούσε σαν μια λογική πύλη NOT (NOT logic gate). Είναι αποδεδειγμένο ότι έχοντας σαν πρώτες ύλες αυτές τις τρεις πύλες μπορείς να φτιάξεις οποιαδήποτε διάταξη λογικής.

Φεύγοντας από το ανθρώπινο σώμα κι επιστρέφοντας στο προγραμματιστικό μας περιβάλλον, η ιδέα είναι πως μπορούμε να κατασκευάσουμε μία μαθηματική διάταξη η οποία χρησιμοποιώντας την παραπάνω λογική μπορεί να κατηγοριοποιεί αντικείμενα. Ο παραλληλισμός έχει ως εξής. Έστω ένας αριθμός από αντικείμενα τα οποία ανήκουν σε τρεις κατηγορίες. Στην κατηγορία Α, στην κατηγορία Β και στην κατηγορία Γ. Έστω ότι τα αντικείμενα αυτά έχουν από 5 χαρακτηριστικά το καθένα. Για παράδειγμα βάρος, ύψος, χρώμα, υφή και ηλικία. Οι συνδυασμοί των χαρακτηριστικών καθορίζουν σε ποια από τις 3 κατηγορίες ανήκει το κάθε αντικείμενο. Να σημειωθεί εδώ ότι προφανώς η ίδια λογική μπορεί να επεκταθεί σε οσοδήποτε κατηγορίες και σε οσαδήποτε χαρακτηριστικά εισόδου.

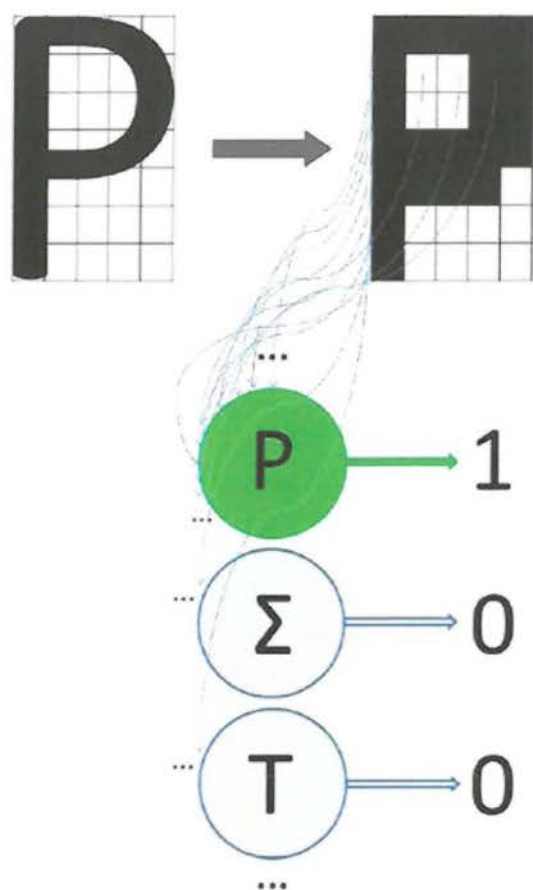
Φτιάχνουμε 3 νευρώνες, όσοι και οι κατηγορίες, και αναθέτουμε στον καθένα αφού εξετάσει όλα τα αντικείμενα να αποφασίσει ποια από αυτά ανήκουν στην κατηγορία που εσπεύει. Δηλαδή να πάρει σαν εισόδους τις 5 τιμές του κάθε αντικειμένου, να τις «ζυγίσει» ανάλογα με το πόσο θεωρεί ότι συμβάλει η κάθε μία στο αποτέλεσμα και αν το άθροισμα ξεπερνάει την τιμή ενεργοποίησης του να πει, ναι, το αντικείμενο αυτό ανήκει στην κατηγορία που ελέγχω. Η λογική λειτουργεί, το πρόβλημα είναι πώς να γνωρίζει ο νευρώνας τι βαρύτητα έχει η κάθε είσοδος στο αποτέλεσμα. Δηλαδή, μπορεί να παίζει πολύ μεγάλο ρόλο το ύψος στο αν ένα αντικείμενο ανήκει σε μια κατηγορία, ενώ η υφή να μη μετράει καθόλου (για παράδειγμα αν έχει να εξετάσει αν τα αντικείμενα ανήκουν στην κατηγορία «μεγάλα αντικείμενα»).

Τη συμβολή της κάθε εισόδου στο αποτέλεσμα την μαθαίνουν οι νευρώνες με μία διαδικασία που λέγεται εκπαίδευση (training). Μαζεύουμε μια λίστα με αντικείμενα των οποίων ξέρουμε την κατηγορία στην οποία ανήκουν. Τα δίνουμε στο νευρώνα ένα ένα και του ζητάμε να αποφασίσει για όλα αν ανήκουν στην κατηγορία που ελέγχει. Προφανώς στην αρχή που δεν ξέρει τίποτα οι απαντήσεις του θα είναι τυχαίες. Στο τέλος κάθε γύρου ελέγχου όλων των προτύπων το οποίο ονομάζεται εποχή (training epoch), ανάλογα με τι λάθη έκανε στις αποφάσεις του μεταβάλλει λίγο τα βάρη των εισόδων του με τη προσδοκία στον επόμενο γύρο να αποφασίσει σωστά για περισσότερα αντικείμενα. Ξεκινάει ο επόμενος γύρος εκπαίδευσης στον οποίο γίνονται τα ίδια πράγματα. Εάν η μεταβολή των βαρών που είχε κάνει πριν ήταν προς τη σωστή κατεύθυνση και σε αυτό το γύρο βρίσκει

σωστά περισσότερα αντικείμενα τότε συνεχίζει τη μεταβολή των βαρών προς την ίδια κατεύθυνση, αν όχι αλλάζει. Αυτό γίνεται παράλληλα για όλους τους νευρώνες και τερματίζεται μόλις εκπληρωθεί κάποια συνθήκη η οποία συνήθως περιλαμβάνει ένα μέτρο ποιότητας (για παράδειγμα αν βρήκε σωστά πάνω από το 80% των προτύπων) κι ένα μέτρο χρόνου (αν η εκπαίδευση ξεπέρασε τις 1000 εποχές κ.τ.λ.). Αυτό συνιστά τη φάση της εκπαίδευσης. Αξίζει να σημειωθεί ότι υπάρχουν πολλοί διαφορετικοί αλγόριθμοι εκπαίδευσης, άλλοι πιο ακριβείς αλλά χρονοβόροι, άλλοι λιγότερο αλλά γρήγοροι, πάντως όλοι στηρίζονται στην ίδια λογική της «δοκιμής και σφάλματος» (trial-and-error).

Τώρα που οι νευρώνες ξέρουν σε αρκετά καλό βαθμό κατά πόσο να λαμβάνουν υπ' όψιν τους το κάθε χαρακτηριστικό ώστε ν' αποφασίσουν τι ανήκει και τι όχι στην κατηγορία που εξετάζουν, μπορούμε να τους τροφοδοτούμε με άγνωστα αντικείμενα των οποίων ξέρουμε μόνο τα χαρακτηριστικά και να μας λένε με μια καλή προσέγγιση σε ποια κατηγορία πρέπει να ανήκουν. Αυτή η φάση ονομάζεται «ανάκληση» (recall) και είναι στην ουσία ο λόγος για τον οποίο κατασκευάζουμε τέτοια συστήματα. Για να μας λένε τι πρέπει να είναι αυτό που δεν ξέρουμε τι είναι.

Επιστρέφοντας στο πρόβλημα της οπτικής αναγνώρισης χαρακτήρων η παραπάνω λογική εφαρμόζεται ως εξής. Μαζεύουμε φωτογραφίες από το κάθε γράμμα του αλφαβήτου, των ψηφίων ή γενικά όλων των συμβόλων που μας ενδιαφέρει να αναγνωρίζει η εφαρμογή. Η πρώτη επιλογή που έχουμε να κάνουμε είναι με ποιο τρόπο θα μετατρέψουμε μια εικόνα ενός συμβόλου σε χαρακτηριστικά κατάλληλα για είσοδο ενός νευρώνα, όπως αυτά που περιγράψαμε πριν. Συνήθως αυτό γίνεται με τον εξής τρόπο. Φτιάχνουμε ένα ορθογώνιο πλέγμα από κουτάκια, έστω 7x5, και προβάλουμε πάνω του το σύμβολο του οποίου τα χαρακτηριστικά θέλουμε να εξάγουμε. Αξίζει να σημειωθεί ότι σε όσο πιο πολλά κουτάκια επιλέξουμε να χωρίζουμε τον κάθε χαρακτήρα, τόσο πιο ακριβής είναι η ανάκληση αλλά και τόσο πιο πολύ καθυστερεί η διαδικασία της εκπαίδευσης, ειδικά για κάποιους αλγορίθμους. Αν ένα κουτάκι περιέχει μέσα του τμήμα του χαρακτήρα που βρίσκεται από κάτω, τότε λέμε ότι το κουτάκι αυτό έχει την τιμή 1. Αν δεν περιέχει, έχει την τιμή 0. Οι τιμές όλων των κουτακιών αυτού του πλέγματος συνιστούν τα χαρακτηριστικά αυτής της μορφής και εισάγονται ως είσοδοι στους νευρώνες. Οπότε για ένα πρόβλημα στο οποίο θα είχαμε να αναγνωρίσουμε μόνο ελληνικά κεφαλαία γράμματα θα χρειαζόμασταν ένα δίκτυο από 24 νευρώνες, ένας για κάθε γράμμα, 35 εισόδων ο καθένας.



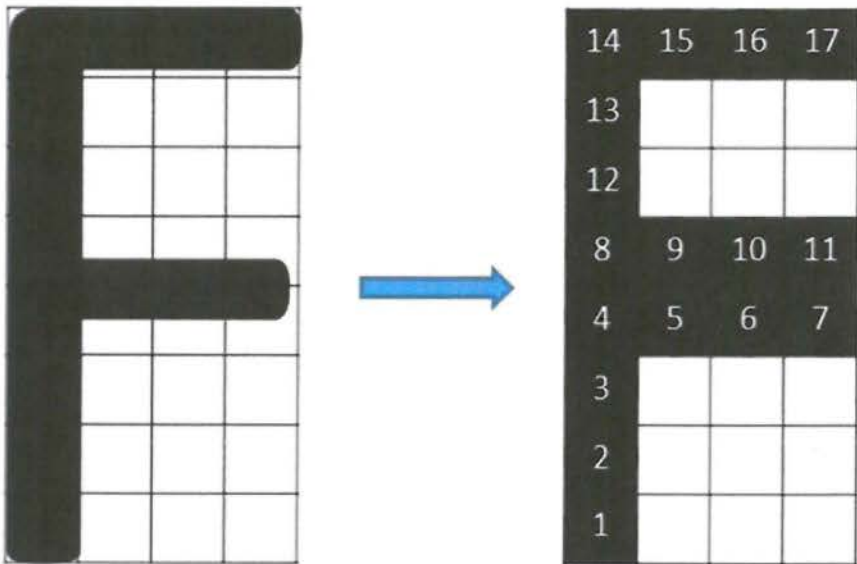
Εικόνα 2.5.4 – Όλοι οι νευρώνες παίρνουν τα ίδια κουτάκια, αλλά μόνο ο P πυροδοτεί.

εκπαίδευση και η ανάκληση γίνεται με τον ίδιο τρόπο που αναφέραμε παραπάνω. βαια, το παράδειγμα στο οποίο αναφερόμαστε είναι για εκπαιδευτικούς λόγους γλοποιημένο. Οι πραγματικές εφαρμογές όπως είπαμε και προηγουμένως μπορεί να ησιμοποιούν κατά την ανάκληση λεξικά για να αυξάνουν την ακρίβεια των γοτελεσμάτων, μπορεί να χρησιμοποιούν πολλά παράλληλα δίκτυα με διαφορετική παίδευση το καθένα και να συνδυάζουν τα αποτελέσματά τους, και γενικά να ησιμοποιούν πολλών ειδών προσαρμογές και βελτιστοποιήσεις οι οποίες χφοροποιούν και χαρακτηρίζουν την κάθε εφαρμογή.

2.6 Αναγνώριση με Στατιστικές Παρατηρήσεις

Αν και όπως είπαμε η οπτική αναγνώριση χαρακτήρων σχεδόν μονοπωλείται από τα τεχνητά νευρωνικά δίκτυα, έχουν αναπτυχθεί κατά καιρούς και διαφορετικοί μέθοδοι αναγνώρισης, με από κακές έως και καλές επιδόσεις, οι οποίες στηρίζουν τη λειτουργία τους στην εξαγωγή κάποιων στατιστικών παρατηρήσεων πάνω στις μορφές που προσπαθούν να αναγνωρίσουν. Κάποιοι από αυτούς παρουσιάζονται παρακάτω.

Ένας πρώτος, απλοϊκός τρόπος, είναι αφού απομονώσουμε τα γράμματα μιας λέξης να τα βάλουμε ένα ένα σ' ένα ορθογώνιο πλέγμα προκαθορισμένου μεγέθους και να μετρήσουμε τον αριθμό των κουτακιών που καταλαμβάνουν. Η λογική είναι ότι για αρκετές γραμματοσειρές και γι' αρκετά μεγέθη, τυχαίνει κανένας χαρακτήρας να μην σχηματίζεται με τον ακριβώς ίδιο αριθμό εικονοστοιχείων μ' έναν άλλον. Για παράδειγμα, σ' ένα υποθετικό πλέγμα 15x10 το Α μπορεί να καταλαμβάνει 63 κουτάκια, το Β που είναι πιο «γεμάτο» 69 κουτάκια, ενώ το «λεπτό» Ι μόλις 24. Έχοντας κρατήσει το σύστημα κατά την εκπαίδευσή του μετρήσεις από όλα τα γράμματα επιχειρεί να αναγνωρίσει μια καινούρια μορφή με τον εξής τρόπο. Βάζει την προς εξέταση μορφή μέσα στο ίδιο πλέγμα που χρησιμοποιήθηκε στην εκπαίδευση και μετράει τον αριθμό των κουτακιών που καταλαμβάνει. Έπειτα, κοιτάζει τη βάση δεδομένων του και βλέπει πιο από τα αποθηκευμένα γράμματα έχει την πλησιέστερη τιμή σε αυτό που μέτρησε τώρα. Με αυτό τον τρόπο αποφασίζει σε ποιο γράμμα αντιστοιχεί η μορφή που μέτρησε. Στο παράδειγμά μας, έστω ότι μετράει μια μορφή και βλέπει ότι αποτελείται από 64 κουτάκια. Αυτή η τιμή, είναι πιο κοντά στην τιμή του Α (63) απ' ότι στο αμέσως επόμενο Β (96), οπότε αποφασίζει ότι είναι το Α.



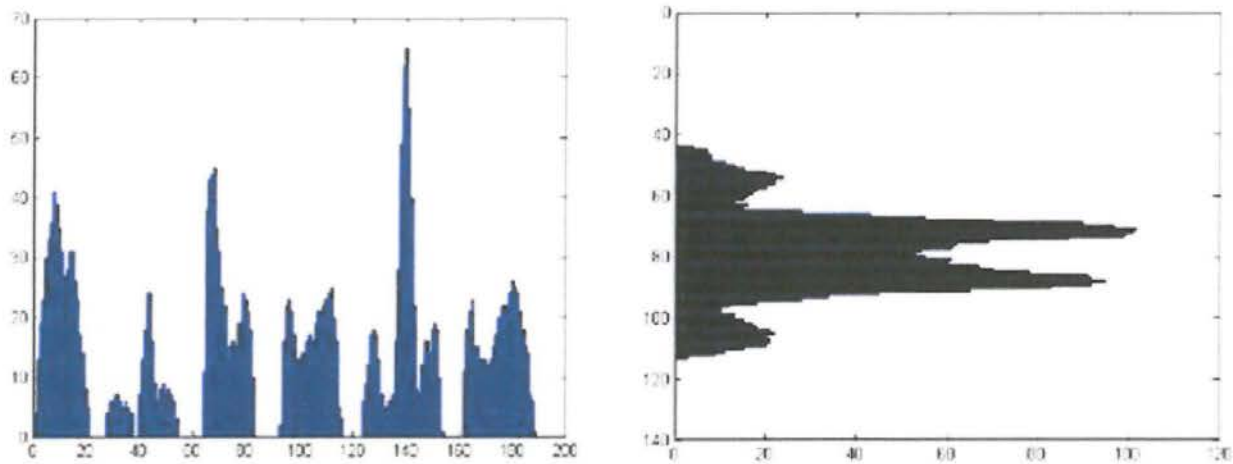
Εικόνα 2.6.1 – Το γράμμα F καταλαμβάνει 17 κουτάκια.

Το πρόβλημα με την παραπάνω διαδικασία, είναι ότι στηρίζεται στην υπόθεση ότι δε θα τύχουν για οποιαδήποτε γραμματοσειρά χαρακτήρες που για το ίδιο μέγεθος θα αποτελούνται από τον ίδιο αριθμό εικονοστοιχείων. Αυτό σε πραγματικές συνθήκες και για μια ευρεία γκάμα γραμματοσειρών δεν συμβαίνει. Υπάρχουν αρκετές γραμματοσειρές για παράδειγμα που τα γράμματα M και W, ή τα 6 και 9 ή άλλα, αποτελούνται από τον ίδιο ακριβώς αριθμό εικονοστοιχείων. Είναι και λογικό, αφού αυτά τα ζευγάρια χαρακτήρων είναι ουσιαστικά αναστροφές της ίδια μορφής. Εμπλουτίζοντας την παραπάνω λογική, έχει προταθεί να χωρίζεται ο κάθε χαρακτήρας στη μέση και να παίρνονται 2 μετρήσεις. Μία για το άνω μισό και μία για το κάτω. Οπότε ο συνδυασμός των δύο αυτών μετρήσεων να αντιπροσωπεύει πιο μονοσήμαντα έναν χαρακτήρα. Ακόμη καλύτερα, χώρισε τον χαρακτήρα σε 4 μέρη και πάρε τέσσερις διαφορετικές μετρήσεις. Η λογική αυτή μπορεί να συνεχιστεί, πάντως έχει μάλλον αποφασιστεί ότι για γενικές συνθήκες δεν μπορεί να αποδώσει ικανοποιητικά κι έτσι εγκαταλείφτηκε για μόνο κάποιες περιπτώσεις αναγνώρισης ειδικών ομάδων χαρακτήρων, σχεδιασμένων ειδικά για την περίπτωση. Το πλεονέκτημα πάντως αυτού του αλγορίθμου είναι η ευκολία των υπολογισμών του και άρα η ταχύτητα εκτέλεσής του σε πιθανές συσκευές και γι' αυτό η λογική των γραμμικών κωδικών (bar codes) βρίσκεται πολύ κοντά σε αυτόν (εκεί μετράμε το πάχος των γραμμών).

Η μόνη (μέχρι σήμερα) βιώσιμη και εμπορικά εφαρμόσιμη λύση αυτής της κατηγορίας αναγνωρίσεων είναι μια μέθοδος που βασίζεται στην κατασκευή ιστογραμμάτων (histograms). Μάλιστα, αυτές οι τεχνικές φαίνεται ότι λειτουργούν καλύτερα για ανατολικά σύνολα χαρακτήρων, όπως κινέζικα και γιαπωνέζικα ιδεογράμματα (εκεί που η δυτικού τύπου οπτική αναγνώριση χαρακτήρων δε λειτουργεί και πολύ καλά) και γι αυτό στα μέρη εκείνα κυκλοφορούν κάποια εμπορικά προϊόντα που βασίζονται σ' αυτές τις τεχνικές.

Η λογική του αλγορίθμου είναι η εξής. Παίρνουμε τον χαρακτήρα που θέλουμε να ελέγξουμε και «πατικώσουμε» τα pixels του από πάνω προς τα κάτω. Ανάλογα με το πλήθος των pixels που έχει ο κάθε χαρακτήρας σε κάθε νοητή κάθετο που περνάει από μέσα του σχηματίζεται στη βάση ένα ιστόγραμμα το οποίο είναι χαρακτηριστικό του. Έτσι, αφού έχουμε αποθηκεύσει ιστογράμματα από όλους τους χαρακτήρες που μας ενδιαφέρει να αναγνωρίζει το σύστημά μας, μπορούμε να εξετάζουμε άγνωστους χαρακτήρες κατασκευάζοντας και συγκρίνοντας το ιστόγραμμα τους με τα' αποθηκευμένα. Για μεγαλύτερη ακρίβεια, η μέθοδος αυτή μπορεί να επεκταθεί και στη λήψη δύο ιστογραμμάτων από τον κάθε χαρακτήρα, ένα στον οριζόντιο κι ένα στον κάθετο άξονα.

έγραψα



Εικόνα 2.6.2 – Κάθετο και Οριζόντιο ιστογράμμο της λέξης «έγραψα».

Ένα εν γένει πλεονέκτημα αυτής της μεθόδου, πέραν της απλότητας των υπολογισμών της, χαρακτηριστικό της οικογένειας των στατιστικών μεθόδων αναγνώρισης χαρακτήρων, είναι ότι παρέχει έναν ευθύ και άμεσο τρόπο, χωρίς την ανάγκη περεταίρω υπολογισμών, διαχωρισμού των χαρακτήρων μέσα σε μια λέξη ή μια πρόταση. Λαμβάνοντας μία μέτρηση στο μήκος μιας ολόκληρης γραμμής, σε εκτυπωμένους μη επικαλυπτόμενους χαρακτήρες τουλάχιστον, είναι εύκολο να βρεις τα σημεία διαχωρισμού των χαρακτήρων ή των λέξεων αναζητώντας τα μηδενικά σημεία του ιστογράμματος. Γι αυτό το λόγο συνήθως η δειγματοληψία γίνεται μαζικά, σ' επίπεδο γραμμής, και όχι μεμονωμένα καθώς δεν υπάρχει κάποιο ιδιαίτερο πλεονέκτημα από αυτό (τουλάχιστον σε πρώτο στάδιο, σε περίπτωση που χρειάζεται περεταίρω έρευνα για την αποκάλυψη της ταυτότητας κάποιου χαρακτήρα μπορεί μεμονωμένα το πρόγραμμα να παίρνει και κάθετο ιστογράμμο από τον αμφιλεγόμενο χαρακτήρα).

ΚΕΦΑΛΑΙΟ 3

ΕΦΑΡΜΟΓΗ

Εισαγωγή

Όταν να δημιουργήσουμε μία εφαρμογή για Android θα χρειαστούμε την βοήθεια ενός IDE (Integrated Development Environment), δηλαδή ενός λογισμικού που θα μας παρέχει όλα τα εργαλεία που χρειαζόμαστε για να γράψουμε κώδικα. Ένα IDE μας παρέχει έναν Source Code Editor για να γράψουμε το κώδικα και επιπλέον μπορεί να ακολουθείτε από το σύστημα αυτόματης συμπλήρωσης (Intelligent Code Completion) το οποίο είτε ολοκληρώνει τις εντολές που γράφουμε ή τις διορθώνει και επίσης μπορεί να μας «προτείνει» πιθανές εντολές. Τέλος ένα ακόμη χρήσιμο εργαλείο είναι ο Debugger, με την βοήθεια του οποίου μπορούμε να βρούμε πολύ πιο εύκολα πιθανά λογικά λάθη στον κώδικα μας. Ο βασικότερος λόγος είναι ότι ένα IDE υποστηρίζει πάρα πολλά Plugins, δηλαδή πρόσθετα εργαλεία, αυτό που θα χρησιμοποιήσουμε εμείς είναι το ADT (Android Developer Tools). Με αυτό το εργαλείο μπορούμε πολύ εύκολα να σχεδιάσουμε το GUI (Graphical User Interface) της εφαρμογής μας, όπως επίσης να την ελέγξουμε ότι δουλεύει σωστά με τον Εξομοιωτή (Emulator) που μας παρέχει πρώτου την δοκιμάσουμε σε κάποια συσκευή.

1 Εγκατάσταση Λογισμικού

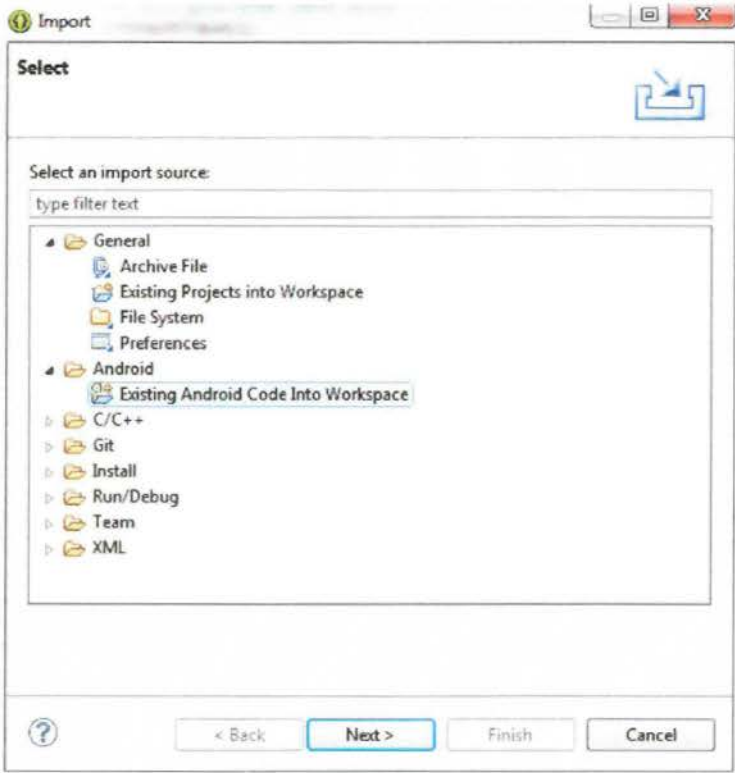
Πρώτα θα εγκαταστήσουμε ένα πακέτο που ονομάζεται Android SDK (Software Development Kit), το οποίο περιλαμβάνει το IDE μας, το οποίο λέγεται Eclipse, μαζί με το ADT plugin και κάποια πρόσθετα εργαλεία.



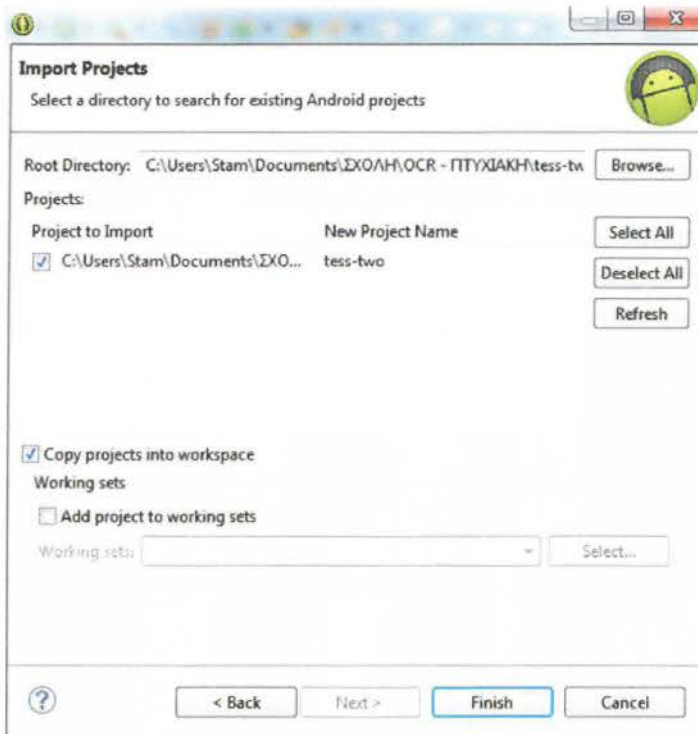
Στην συνέχεια θα χρειαστούμε δύο βιβλιοθήκες μία για την αναγνώριση των χαρακτήρων και μία για την επεξεργασία της εικόνας. Για την αναγνώριση θα χρησιμοποιήσουμε το Tesseract και για την επεξεργασία το OpenCV. Για να μην χρειαστεί να κτίσουμε (built) των κώδικα του Tesseract και ύστερα να τον εισάγουμε στο Eclipse μας θα κατεβάσουμε την βιβλιοθήκη tess-two, η οποία είναι έτοιμη για να την εισάγουμε στο Eclipse.

3.1.1 – Tesseract (tess-two library)

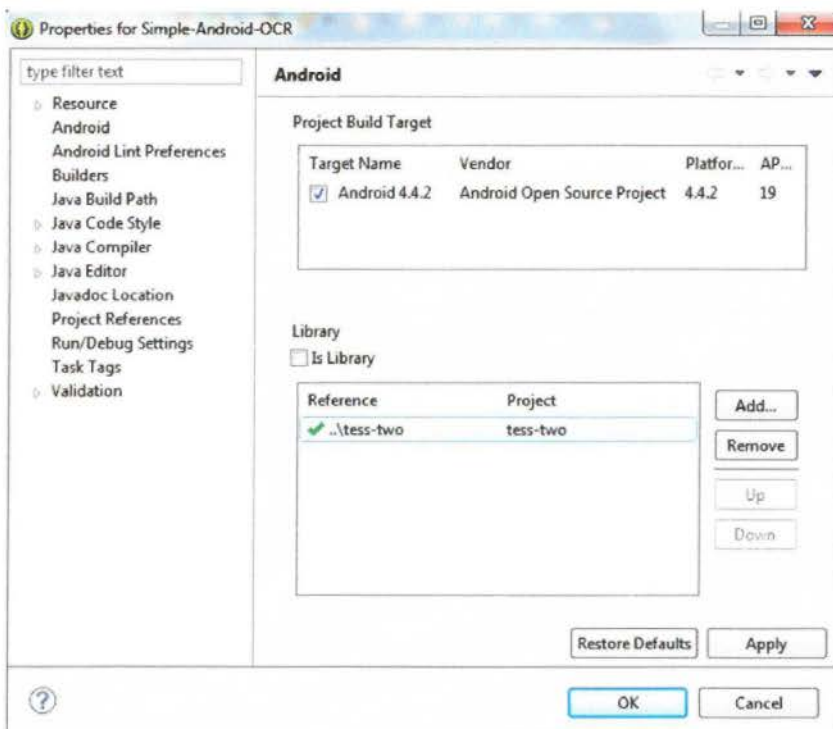
Τρέχουμε το Eclipse που βρίσκετε μέσα στον φάκελο eclipse, ο οποίος βρίσκεται μέσα στον φάκελο adt-bundle-windows-x86_64-20131030. Για να εισάγουμε την βιβλιοθήκη tess-two στο Eclipse , πάμε *File > Import...* μετα επιλέγουμε το *Existing Android Code Into Workspace*.



Μετά πατάμε το *Browse...* και βρίσκουμε τον φάκελο tess-two και πατάμε OK. Ύστερα το επιλέγουμε στα *Projects to Import* σε περίπτωση που δεν είναι ήδη επιλεγμένο και επίσης επιλέγουμε το *Copy projects into Workspace* έτσι ώστε ολόκληρη η βιβλιοθήκη να αντιγραφεί στο workspace μας και να φορτώνεται πιο γρήγορα.

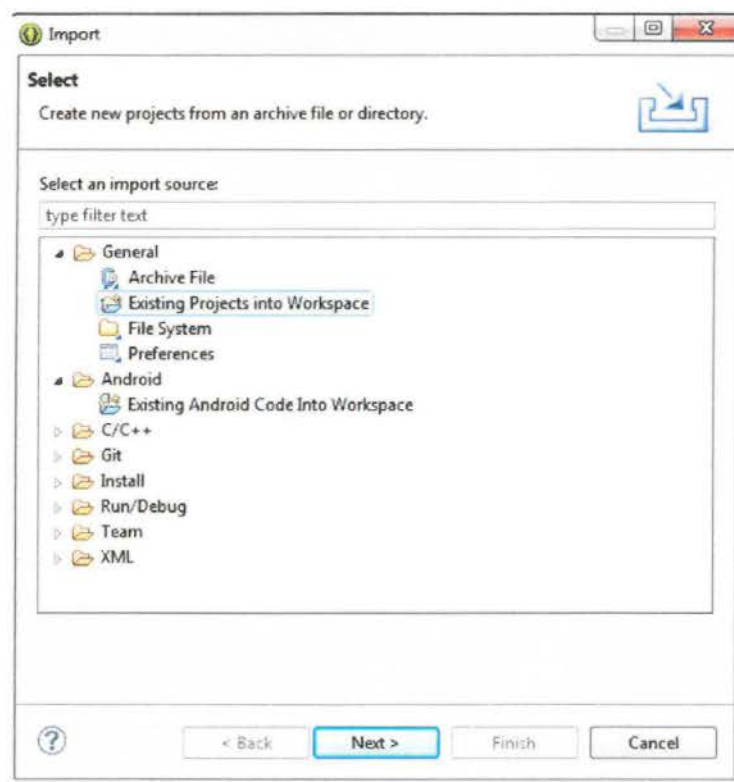


ην συνέχεια πρέπει να προσθέσουμε το tess-two στο Project μας σαν βιβλιοθήκη. Πινακίδα δεξιά κλικ πάνω στο Project μας, μετά πάμε *Properties* και στο παράθυρο που οίγει επιλέγουμε *Android*. Πατάμε το κουμπί *Add...* και επιλέγουμε το tess-two, μετά πάμε *Apply* και *OK*.

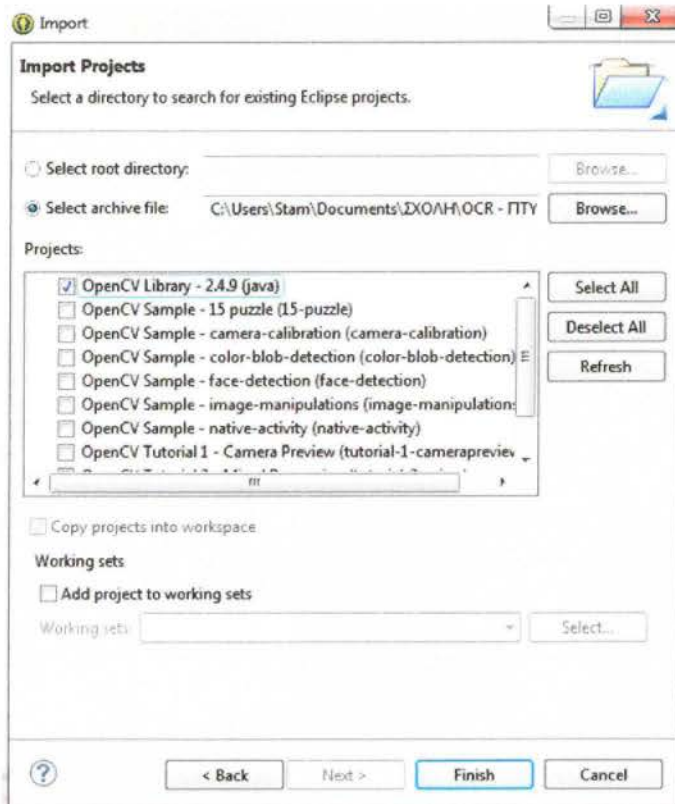


3.1.2 – OpenCV

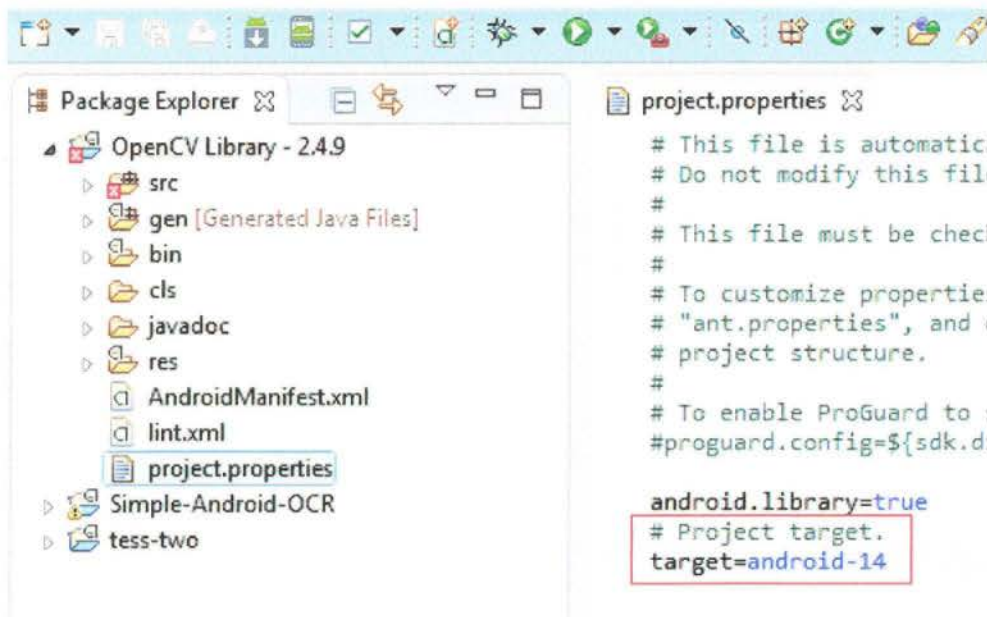
Για να εισάγουμε την βιβλιοθήκη OpenCV στο Eclipse , πάμε *File > Import...* μετα επιλέγουμε το *Existing Project into Workspace*.



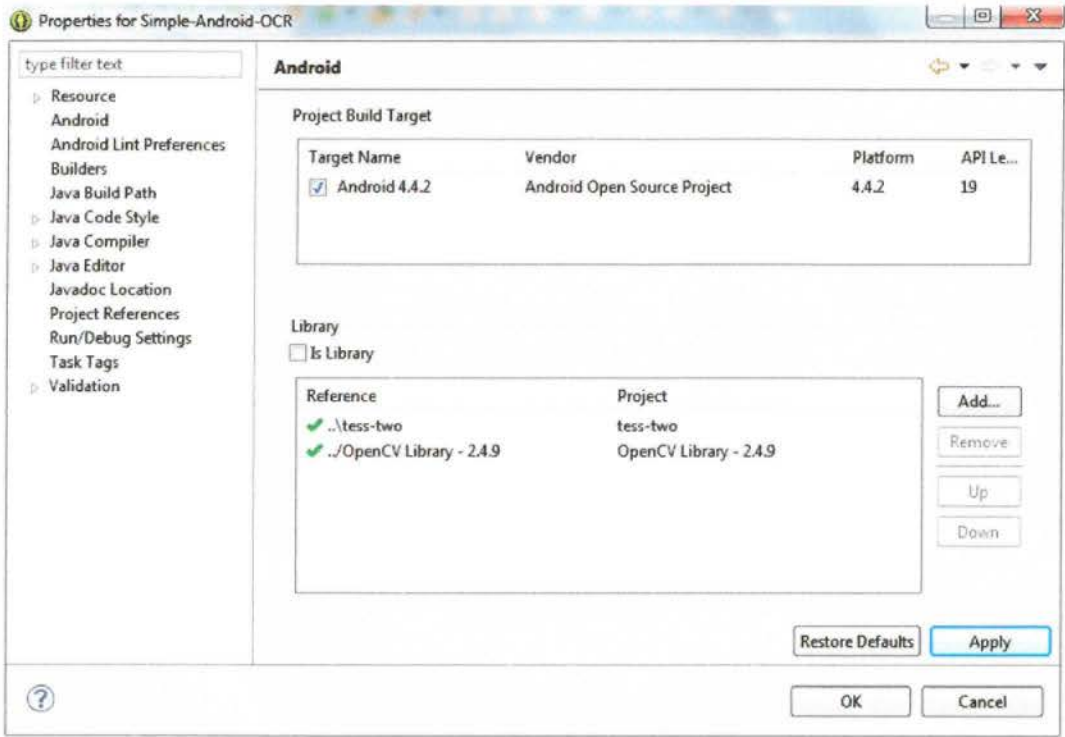
Μετά επιλέγουμε το *Select archive file:* και πατάμε το *Browse...* και βρίσκουμε το *OpenCV-2.4.9-android-sdk.zip* αρχείο. Έπειτα πατάμε το *Deselect All* και επιλέγουμε μόνο το *OpenCV Library* και πατάμε *Finish*.



έπουμε όμως ότι υπάρχουν κάποια λάθη (errors), για να τα διορώσουμε πρέπει να πτήσουμε πάνω στο OpenCV Library, να ανοίξουμε το αρχείο *project.properties* και να λάξουμε το *Project target* από *target = android-14* σε *target = android-19*.



Τέλος πρέπει να προσθέσουμε το OpenCV στο Project μας σαν βιβλιοθήκη. Κάνουμε δεξιά κλικ πάνω στο Project μας, μετά πάμε *Properties* και στο παράθυρο που ανοίγει επιλέγουμε *Android*. Πατάμε το κουμπί *Add...* και επιλέγουμε το OpenCV Library, μετά πατάμε *Apply* και *OK*.



Για να μπορέσει να λειτουργήσει το OpenCV στο κινητό ή στον Emulator πρέπει να εγκατασταθεί ο OpenCV Manager. Στο κινητό είναι πολύ απλό, είτε το καταβάζουμε απο το Google Play είτε θα ζητηθεί να κατέβει απο την ίδια την εφαρμογή. Για να το εγκαταστήσουμε στον Emulator πρέπει πρώτα να δούμε τι επεξεργαστή προσομοιώνουμε και ύστερα να κατεβάσουμε το κατάλληλο πακέτο.

Hardware Platform	Android ver.	Package name
armeabi-v7a (ARMv7-A + NEON)	>= 2.3	OpenCV_2.4.9_Manager_2.18_armv7a-neon.apk
armeabi-v7a (ARMv7-A + NEON)	= 2.2	OpenCV_2.4.9_Manager_2.18_armv7a-neon-android8.apk
armeabi (ARMv5, ARMv6)	>= 2.3	OpenCV_2.4.9_Manager_2.18_armeabi.apk
Intel x86	>= 2.3	OpenCV_2.4.9_Manager_2.18_x86.apk
MIPS	>= 2.3	OpenCV_2.4.9_Manager_2.18_mips.apk

ρού έχουμε κατεβάσει το σωστό πακέτο, στην περίπτωση μας είναι το πρώτο, το ταφέρουμε στον φάκελο όπου βρίσκετε το adb το οποίο θα μας βοηθήσει να κάνουμε κατάσταση το πακέτο στον Emulator. Ανοίγουμε τον φάκελο του ADT , μετά ανοίγουμε ν φάκελο sdk και τέλος στον φάκελο platform-tools βρίσκετε το adb, εκεί μετακινούμε το κέτο που καταβάσαμε.



ην συνέχεια ανοίγουμε τον Emulator και περιμένουμε μέχρι να φορτώσει, έπειτα οίγουμε το CMD και μετακινούμαστε στον φάκελο platform-tools με την εντολή cd (change Directory) και το path.

```
\Users\Stam>
```

```
Documents\adt-bundle-windows-x86_64-20131030\sdk\platform-tools
```

τελούμε την παρακάτω εντολή για να γίνει η εγκατάσταση.

```
\Users\Stam\Documents\adt-bundle-windows-x86_64-20131030\sdk\platform-tools>
```

```
adb install OpenCV_2.4.9_Manager_2.18_armv7a-neon.apk
```

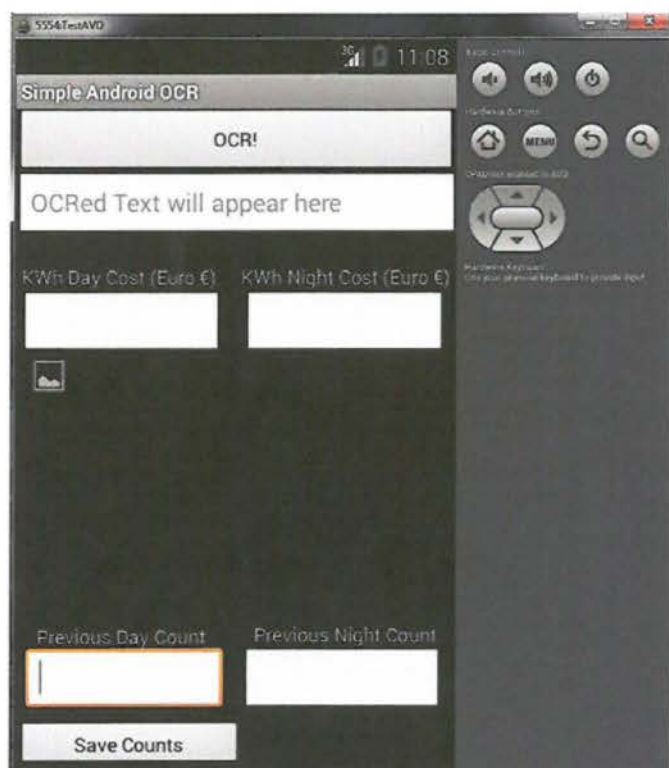


3.2 Παρουσίαση της Εφαρμογής

Η εφαρμογή τρέχει σε AVD (Android Virtual Device) που προσομοιώνει τον Nexus One, το οποίο έχει οθόνη 3.7", επεξεργαστή ARM (armeabi-v7a), RAM 512 MB και λειτουργικό Android 4.4.2 – API Level 19.

Όταν τρέξουμε για πρώτη φορά την εφαρμογή πρέπει να εισάγουμε τις προηγούμενες μετρήσεις του ρολογιού οι οποίες αναγράφονται πάνω στον λογαριασμό της ΔΕΗ. Όλα τα σύγχρονα ρολόγια περιλαμβάνουν και το ημερήσιο και το νυχτερινό μετρητή σε ένα ρολοι, οπότε η εφαρμογή υπολογίζει το ποσό που θα πρέπει να πληρωθεί με βάση και των δύο μετρήσεων.

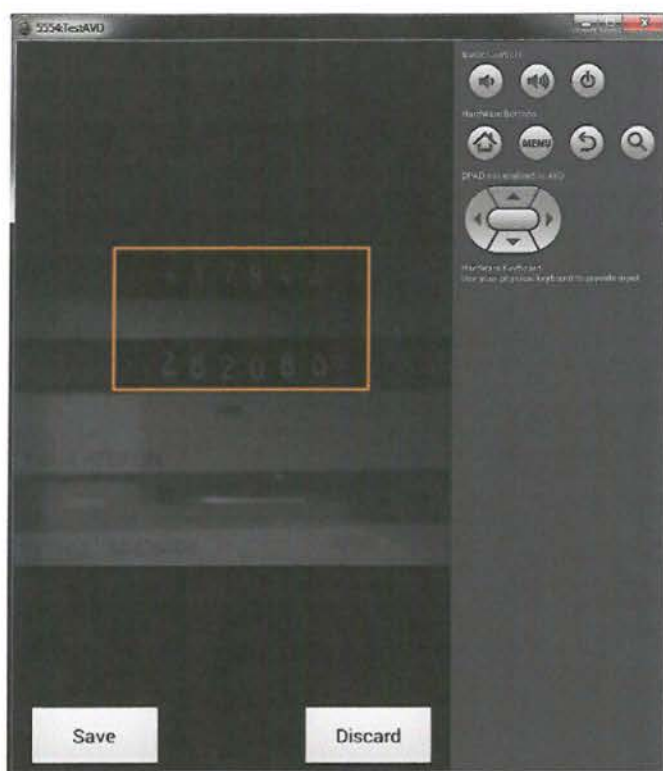
Για να εισάγουμε την προηγούμενη ημερήσια μέτρηση την γράφουμε στο πεδίο Previous Day Count και για να εισάγουμε την προηγούμενη νυχτερινή μέτρηση την γράφουμε στο πεδίο Previous Night Count και πατάμε το κουμπί Save Counts για να αποθηκεύσουμε τις μετρήσεις στην μνήμη της συσκευής.



ρού αποθηκεύσουμε τις μετρήσεις, πατώντας το κουμπί OCR! θα ανοίξει η βιβλιοθήκη page Gallery) με όλες τις εικόνες που είναι αποθηκευμένες στην συσκευή. Έχουμε ήδη γιοιοες φωτογραφίες περασμένες οι οποίες έχουν τραβηχτεί με ένα παλίο κινητό (Nokia 2) το οποίο διαθέτει 5 Mpixel κάμερα με μέτρια ποιότητα.



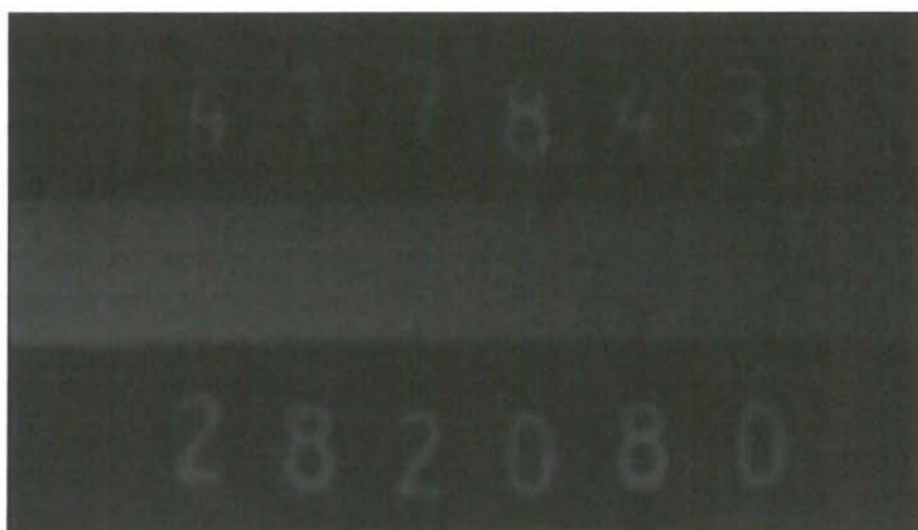
πιλέγουμε την φωτογραφία που μας ενδιαφέρει, στην περίπτωση μας είναι η πρώτη η οία είναι η πραγματική που τραβήχτηκε απο το κινητό, οι υπόλοιπες είναι εξεργασμένες στο Photoshop για να ελέγξουμε την εφαρμογή μια διάφορες ταστάσεις, δηλαδή υψηλή αντίθεση, έντονο φωτισμό, κτλ. Αφού την επιλέξουμε μας γείται να κόψουμε (crop) την φωτογραφία μετακινώντας το παραλληλόγραμμο πλαίσιο γύρω απο τα ψηφία. Αυτό γίνεται για δύο λογους, πρώτων για να μπορέσουμε να ιράγουμε ακριβέστερα αποτελέσματα απομονώνοντας μόνο το σημείο που μας διαφέρει και δεύτερον για να γίνονται οι υπολογισμοί γρήγορα και σε λιγότερο δυνατές σκευές και επίσης να αποφεύγετε να γεμίζει η μνήμη RAM (Out of Memory) λογο των λαστιων δεδομένων που πρέπει να επεξεργαστούν.



Πατώντας το Save ξεκινάει η επεξεργασία της φωτογραφίας, την οποία θα την δούμε παρακάτω. Εάν θέλουμε να επιλέξουμε διαφορετική φωτογραφία πατάμε το Discard και πηγαίνουμε πίσω να επιλέξουμε κάποια άλλη .

3.2.1 Διαδικασία Επεξεργασίας

Παρακάτω θα δούμε τα στάδια επεξεργασίας της φωτογραφίας αφού την έχουμε κόψει. Στην εικόνα 3.2.1 βλέπουμε την αρχική μας φωτογραφία μετά το κόψιμο.



Εικόνας 3.2.1 – Αρχική Εικόνα

την συνέχεια πρέπει να την μετατρέψουμε απο έγχρωμη (RGB) σε εικόνα φωτεινότητας (grayscale) , έτσι ώστε στην συνέχεια να την μετατρέψουμε σε δυαδική (Binary).



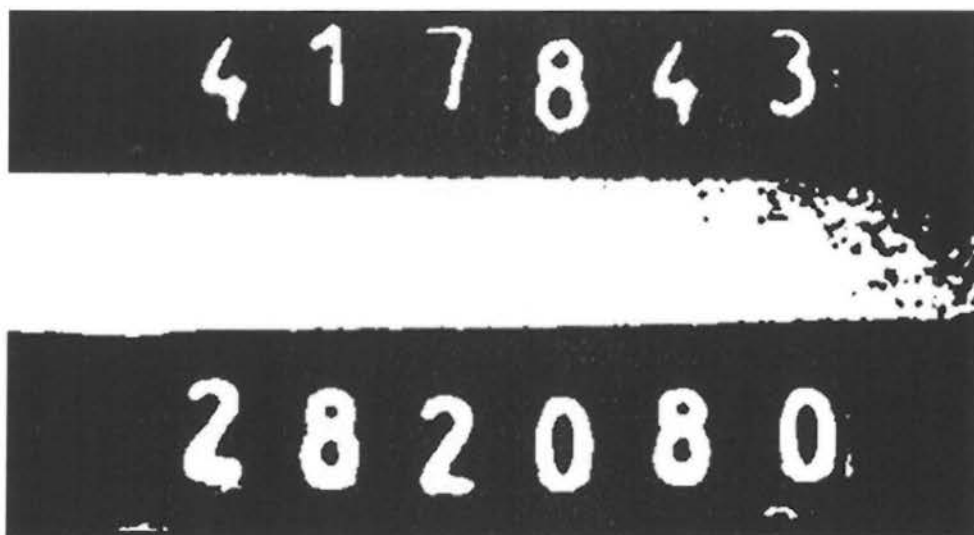
Εικόνα 3.2.2 – Grayscale

Χαρατηρώντας την εικόνα βλέπουμε οτι έχει αρκετό θόρυβο, ο οποίος θα επηρεάσει καταλυτά τα αποτελέσματα μας. Γι' αυτό το λόγο χρησιμοποιούμε ένα αμφίπλευρο φίλτρο (bilateral Filter) για την μείωσή του. Αυτή η διαδικασία θα θολώσει την εικόνα μας και θα αφαιρέσει τις ακμές ανέπαφες.



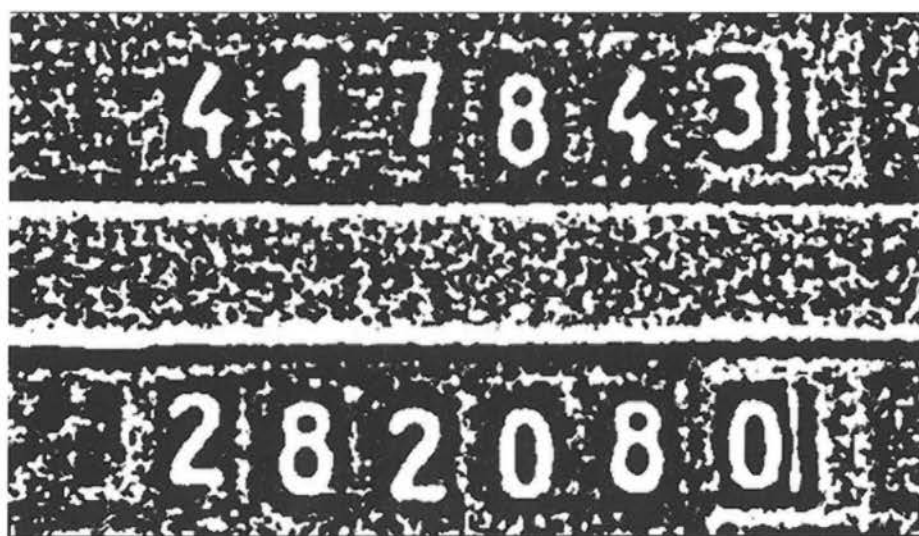
Εικόνα 3.2.3 – Bilateral Filter

Έπειτα πρέπει να μετατρέψουμε την εικόνα μας σε δυαδική (Binary) με την διαδικασία του κατωφλίου (Threshold), θα μπορούσαμε να κάνουμε την μέθοδο κατωφλίου του Otsu (Otsu's Method Threshold) αλλά η πυκνότητα των ψηφίων διαφέρει το οποίο θα μας δημιουργήσει πρόβλημα αργότερα.



Εικόνα 3.2.4 – Global Threshold (Otsu's Method)

Γι' αυτό χρησιμοποιούμε προσαρμοζόμενο κατώφλι (Adaptive Threshold) για ομοιόμορφο αποτέλεσμα, η μέθοδος που μας προσφέρει το OpenCV για Adaptive Threshold υλοποιεί παρόμοια μέθοδο δυαδικοποίησης με αυτή του Niblack. Βλέπουμε ότι η πυκνότητα όλων των ψηφίων είναι ίδια.



Εικόνα 3.2.5 – Adaptive Threshold (Niblack's Method)

Με την μέθοδο της συρρίκνωσης/διάβρωσης (Erosion) ελαχιστοποιούμε κατα πολύ τα περιττά σχήματα (Contours) για να είναι πιο εύκολο να τα αφαιρέσουμε εντελώς στην συνέχεια και να κρατήσουμε μόνο τους αριθμούς για ακριβέστερα αποτελέσματα.



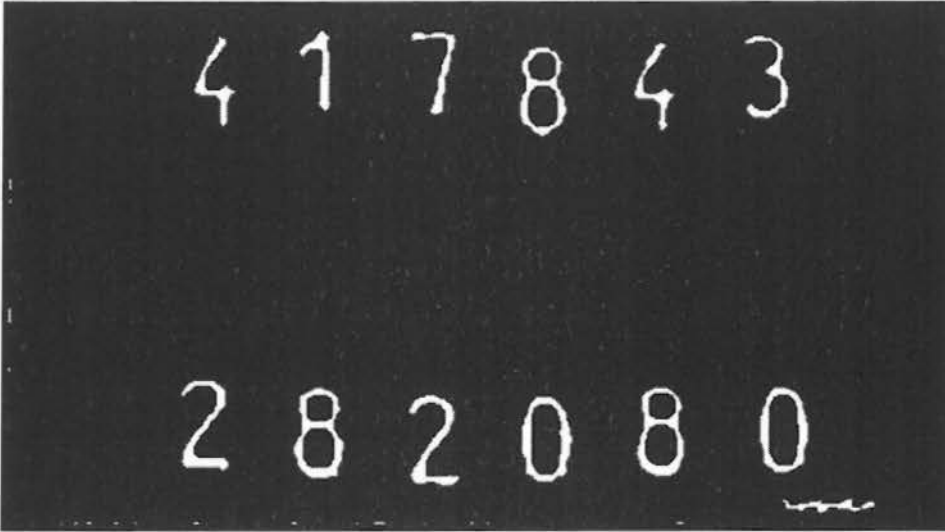
Εικόνα 3.2.6 - Erosion

Στην συνέχεια πρέπει να αφαιρέσουμε τα περιττά σχήματα (Contours) τα οποία θα επηρεάσουν τα αποτελέσματα μας.



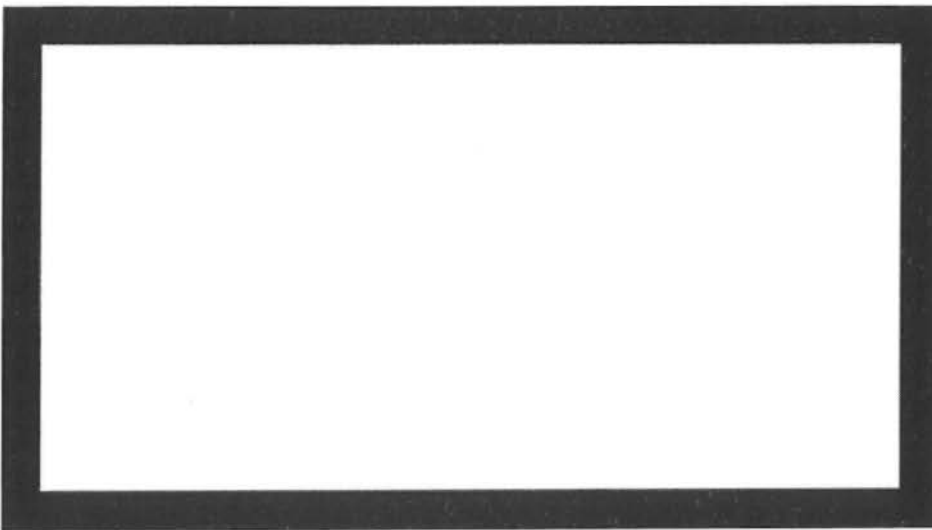
Εικόνα 3.2.7 – Αφαίρεση των Μεγάλων Σχημάτων

Αφαιρέσαμε τα μεγάλα σχήματα διατηρώντας τα ψηφία. Αν τα ψηφία μας είχαν διαφορετική πυκνότητα υπήρχε περίπτωση κάποια απο αυτά να διαγραφούν, διότι υπήρχε περίπτωση να έχουν τον ίδιο αριθμό pixels με κάποιο απο τα μεγάλα περιττά σχήματα ή ακόμη και με κάποιο απο τα μικρά.



Εικόνα 3.2.8 – Αφαίρεση των Μικρών Σχημάτων

Παρατηρούμε ότι στο περίγραμμα της εικόνας μας έχουν παραμείνει κάποια περιττά σχήματα , για να τα αφαιρέσουμε εντελώς δημιουργούμε μία νέα εικόνα (Εικόνα 3.2.9) την οποία θα την κάνουμε λογικό AND με την εικόνα μας ετσι ώστε να κρατήσουμε μόνο τα ψηφία.



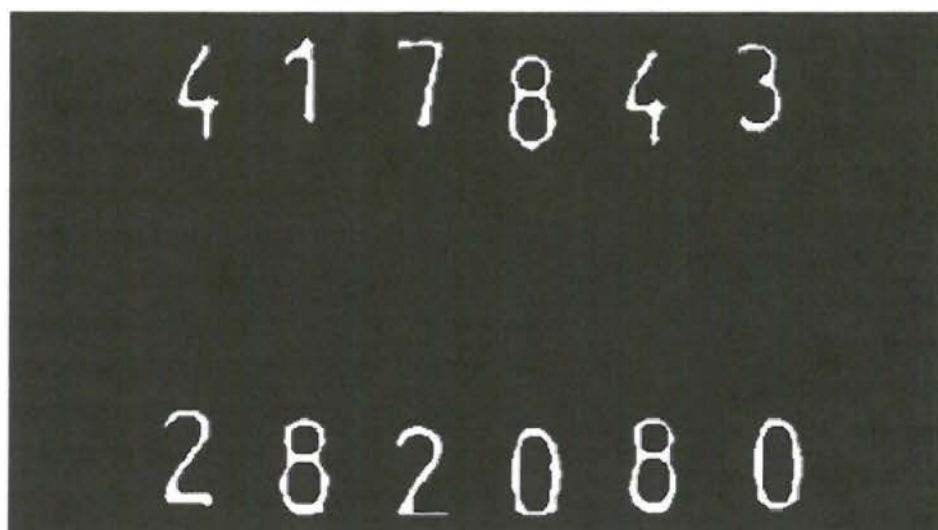
Εικόνα 3.2.9 – Εικόνα που θα γίνει λογικό AND με την φωτογραφία.

ο περίγραμμα της εικόνας που φτιάξαμε είναι μάυρο (0), οπότε ότι υπάρχει στο περίγραμμα της φωτογραφίας μας το οποίο είναι άσπρο (1), δηλαδή τα περιττα ζυματάκια, θα γίνει μάυρο (0) ενώ ότι ήταν μάυρο (0) θα παραμείνει μαυρο (0). Το ίδιο θα γίνει και στο εσωτερικό της εικόνας μας , δηλαδή θα διατηρήσουμε τα ψηφία και το φόντο (ackground).

INPUT		OUTPUT
A	B	A AND B
0	0	0
0	1	0
1	0	0
1	1	1

Εικόνα 3.2.10 – Λογική της πράξης AND

η εικόνα μας έχει καθαριστεί εντελώς και έχουν παραμείνει μόνο τα ψηφία που θέλουμε να διατηρήσουμε.



Εικόνα 3.2.11 – Τελική Εικόνα

Αφού ολοκληρωθεί η διαδικασία της αναγνώρισης των χαρακτηριστων, σειρά έχει ο υπολογισμός του ποσού που πρέπει να πληρώσουμε στην ΔΕΗ. Ο υπολογισμός γίνεται με βάση τον παρακάτω πίνακα αφού πρώτα υπολογίσουμε την διαφορά της τωρινής μέτρησης, δηλαδή των αναγνωρισμένων ψηφίων, και των προηγούμενων μετρήσεων που έχουμε αποθηκεύσει στην συσκευή.

Κλιμάκιο Κατανάλωσης σε Κιλοβατώρες	Χρέωση ανταγωνιστική €	Χρέωση μονοπωλιακή €	Συνολική χρέωση €/kWh
0 – 800	0.054	0.03016 + 0.0002	0.08436
801 – 1000	0.0686	0.03016 + 0.00528	0.10404
1001 – 1200	0.071	0.03016 + 0.00528	0.10644
1201 – 1600	0.0727	0.03016 + 0.00528	0.10814
1601 – 2000	0.0727	0.03016 + 0.01137	0.11423
2001 – 3000	0.08174	0.03016 + 0.03157	0.14347
3001 +	0.08174	0.03016 + 0.03608	0.14798

Νυχτερινό τιμολόγιο = 0,05591 €

Παράδειγμα

Προηγούμενη Ημερήσια Μέτρηση = 30455

Τωρινή Ημερήσια Μέτρηση = 32400

Προηγούμενη Νυχτερινή Μέτρηση = 12677

Τωρινή Νυχτερινή Μέτρηση = 13600

32400 – 30455 = 1945

Η χρέωση γι' αυτό είναι **0.11423** οπότε $1945 \cdot 0.11423 = 222.18 \text{ €}$

13600 – 12677 = 923

Η χρέωση γι' αυτό είναι **0,05591** οπότε $923 \cdot 0.05591 = 51.6 \text{ €}$

Όλως γίνονται οι υπολογισμοί η εφαρμογή εμφανίζει τα ψηφία που αναγνώρισε και την ομογραφία με την τελική της επεξεργασία για να συγκρίνουμε τα αποτελέσματα και να γουρευτούμε ότι δεν υπάρχει κάποιο λάθος. Τέλος μας εμφανίζει ξεχωριστά στο πίστοιχο πεδίο το ποσό που πρέπει να πληρωθεί για το ημερήσιο και το νυχτερινό ύμα.



Εικόνα 3.2.12 – Τελικό Αποτέλεσμα της Εφαρμογής

ΒΙΒΛΙΟΓΡΑΦΙΑ

Διαδύκτιο

- [1] Camera Sensors At Work: how your digital camera turns light into an image
<http://www.digitalcameraworld.com/2012/08/31/cheat-sheet-friday-how-your-digital-camera-turns-light-into-an-image/>
- [2] Digital Camera Sensors
<http://www.cambridgeincolour.com/tutorials/camera-sensors.htm>
- [3] Color Space and Color Profiles
<http://dpbestflow.org/color/color-space-and-color-profiles>
- [4] Color Models
http://dba.med.sc.edu/price/irf/Adobe_tg/models/main.html
- [5] Image Processing Toolbox – Displaying Image Information
<http://www.cs.uregina.ca/Links/class-info/425/Lab2/>
- [6] Bit-Depth
<http://www.shinno.biz/?q=node/3>
- [7] Τύποι Αρχείων
http://graficnotes.blogspot.gr/2014/04/blog-post_26.html
- [8] Digital Camera Image Noise
<http://www.cambridgeincolour.com/tutorials/image-noise.htm>
- [9] University of Basel – Image Restoration
[http://miac.unibas.ch/SIP/06-Restoration.html#\(1\)](http://miac.unibas.ch/SIP/06-Restoration.html#(1))
- [10] Bilateral Filtering
<http://www.mathworks.gr/matlabcentral/fileexchange/12191-bilateral-filtering>
- [11] Dilation and Erosion
<http://www-rohan.sdsu.edu/doc/matlab/toolbox/images/morph4.html#24720>

Βιβλία

- [12] I.N.Έλληνας (2010), Ψηφιακή Επεξεργασία Εικόνας & Βίντεο
- [13] J.Sauvola & M.Pietikainen (2000), Adaptive Document Image Binarization
- [14] W.Niblack (1986), An Introduction to Digital Image Processing
- [15] N.Otsu (1979), A thresholding selection method form grayscale histogram
- [16] O.Losson,L.Macaire & Y.Yang, Color demosaicing methods