



Τεχνολογικό Εκπαιδευτικό Ίδρυμα Κρήτης

Σχολή Τεχνολογικών Εφαρμογών

**Τμήμα Εφαρμοσμένης Πληροφορικής &
Πολυμέσων**



Πτυχιακή Εργασία

Τίτλος: **Κατηγοριοποίηση και αναγνώριση χαρακτήρων σε
αρχαίες ελληνικές επιγραφές**

Ιωάννης Χρ. Κυριαζίδης

AM: 1664

Ευχαριστίες

Ευχαριστώ πολύ τους γονείς μου Χρυσόστομο, Βιολέτα και τον αδελφό μου Αλέξανδρο για τη συμπαράστασή τους και τη βοήθεια που μου πρόσφεραν κατά τη διάρκεια των σπουδών μου.

Abstract

In this project, methods and techniques are presented in digital image processing which are used or can be used in systems of character classification and recognition in images and particularly in images of ancient inscriptions.

Once we understand some basic elements for the science of studying ancient inscriptions we will identify the points at which computer applications can provide solutions, what kind of applications there are, and how developing some new applications can help to further simplify and automate additional processes of this science.

After research, the most notable of them (according to literature) were gathered and categorized according to their intended use. Thus, some of these were employed in image preprocessing, segmentation, isolation of the characters, feature extraction and finally character recognition and classification.

Because of the special nature of the inscriptions' images—i.e. the large presence of noise due to their antiquity or due to their damaged condition, the size of images and other problems that will be reported in subsequent chapters—there were followed techniques different from the ones in a typical optical character recognition system (OCR-Optical Character Recognition). Also it was checked whether typical OCR implementations could be used and what changes-modifications should be made in order to get the best possible result.

After the configuration of these methods, we performed experiments in which some of the methods were rejected (for reasons which will be reported) and then, we proposed the most suitable of them, which were used to simulate the functions of a system of categorization and recognition of characters in an ancient inscription.

Σύνοψη

Στην παρούσα εργασία γίνεται αναφορά και παρουσίαση μεθόδων και τεχνικών του τομέα της ψηφιακής επεξεργασίας εικόνας, που χρησιμοποιούνται ή που μπορούν να χρησιμοποιηθούν σε συστήματα κατηγοριοποίησης και αναγνώρισης χαρακτήρων σε εικόνες και πιο συγκεκριμένα σε εικόνες αρχαίων επιγραφών.

Αφού κατανοήσουμε κάποια βασικά στοιχεία της θεωρητικής επιστήμης της μελέτης των αρχαίων επιγραφών θα γίνει ο εντοπισμός των σημείων στα οποία εφαρμογές της πληροφορικής δίνουν λύσεις και ποιες είναι αυτές, καθώς και πώς αναπτύσσοντας κάποιες νέες εφαρμογές μπορούμε να συμβάλουμε στη περαιτέρω απλούστευση ή αυτοματοποίηση επιπλέον διαδικασιών της επιστήμης αυτής.

Ύστερα από έρευνα, συγκεντρώθηκαν οι πιο αξιόλογες εξ αυτών (σύμφωνα με τη βιβλιογραφία) και κατηγοριοποιήθηκαν ανάλογα με τον τομέα στον οποίο προορίζεται να χρησιμοποιηθούν. Έτσι ορισμένες από αυτές χρησιμοποιήθηκαν στην προεπεξεργασία της εικόνας, την κατάτμηση της, την απομόνωση των χαρακτήρων, την εξαγωγή χαρακτηριστικών των χαρακτήρων και τέλος την αναγνώριση και κατηγοριοποίηση των χαρακτήρων.

Λόγω τις ιδιαίτερης φύσης των εικόνων των επιγραφών—δηλαδή της μεγάλης παρουσίας θορύβου λόγω παλαιότητας ή λόγω φθορών, του μεγέθους των εικόνων και άλλων προβλημάτων που θα αναφερθούν σε ακόλουθα κεφάλαια—οι τεχνικές που ακολουθήθηκαν διαφέρουν από αυτές που ακολουθούνται σε ένα συνηθισμένο σύστημα οπτικής αναγνώρισης χαρακτήρων (OCR-Optical Character Recognition). Έγινε έλεγχος αν και κατά πόσο μπορούν υπάρχουσες υλοποιήσεις (OCR) να χρησιμοποιηθούν και με ποιες αλλαγές-τροποποιήσεις θα ήταν εφικτό το καλύτερο δυνατό αποτέλεσμα.

Μετά από παραμετροποίηση των μεθόδων αυτών, στη συνέχεια εκτελέστηκαν πειράματα και αφού κάποιες από αυτές τις μεθόδους απορρίφθηκαν (για λόγους οι οποίοι θα αναφερθούν), προτάθηκαν οι καταλληλότερες, οι οποίες και χρησιμοποιήθηκαν για να προσομοιώσουν τις λειτουργίες ενός συστήματος κατηγοριοποίησης και αναγνώρισης χαρακτήρων σε μια αρχαία επιγραφή.

Πίνακας Περιεχομένων

1.	Εισαγωγή.....	1
1.1.	Σκοπός και Στόχοι Εργασίας.....	1
1.2.	Δομή εργασίας.....	2
2.	Ανάλυση προβλήματος.....	3
2.1.	Υπάρχουσες τεχνικές.....	3
2.2.	Εναλλακτικές εφαρμογές.....	5
2.3.	Προτεινόμενη λύση.....	6
2.4.	Ψηφιακή Επεξεργασία Εικόνας.....	6
2.5.	Matlab.....	7
2.5.1.	Γιατί Matlab;.....	8
2.5.2.	Matlab Image Processing Toolbox.....	8
2.5.3.	Matlab Neural Network Toolbox.....	8
3.	Μεθοδολογία Υλοποίησης.....	9
3.1.	Προεπεξεργασία.....	9
3.1.1.	Διαδική μετατροπή.....	10
3.1.2.	Βελτίωση Ποιότητας.....	26
3.2.	Κατάτμηση Χαρακτήρων.....	40
3.3.	Εξαγωγή Χαρακτηριστικών.....	44
3.3.1.	Γενικοί μετασχηματισμοί και ανάπτυξη σειρών.....	45
3.3.2.	Χαρακτηριστικά που προέρχονται από τη στατιστική κατανομή των σημείων.....	47
3.3.3.	Γεωμετρικά, τοπολογικά και μορφολογικά χαρακτηριστικά.....	49
3.4.	Αναγνώριση – Κατηγοριοποίηση.....	55
3.4.1.	Template matching.....	55
3.4.2.	Στατιστικές τεχνικές.....	56
3.4.3.	Δομικές τεχνικές.....	58
3.4.4.	Νευρωνικά δίκτυα.....	59
4.	Πειραματικό μέρος.....	63
4.1.	Προεπεξεργασία.....	63
4.2.	Κατάτμηση χαρακτήρων.....	65
4.3.	Εξαγωγή χαρακτηριστικών.....	67
4.4.	Κατηγοριοποίηση - Αναγνώριση.....	67

5. Συμπεράσματα και μελλοντικές εργασίες	72
6. Αναφορές – Βιβλιογραφία.....	74

Πίνακας εικόνων

Εικόνα 1. Φωτογραφία έκτυπου σε στυπόχαρτο.....	3
Εικόνα 2. Αριστερά τα έκτυπα και δεξιά οι 3D ανακατασκευές τους.....	5
Εικόνα 3. Το λογότυπο του MATLAB.....	7
Εικόνα 4. Δομή συστήματος.....	9
Εικόνα 5. Επιλογή κατώφλιου μέσω οπτικής ανάλυσης του ιστογράμματος της εικόνας.....	11
Εικόνα 6. Ασπρόμαυρες εικόνες που προκύπτουν από την εφαρμογή καθολικής κατωφλίωσης στην επιγραφή της Εικόνας 5 επιλέγοντας διαφορετικές τιμές ($T_\alpha < T_\beta < \dots < T_\zeta$) για το κατώφλι.....	12
Εικόνα 7. Εφαρμογή της μεθόδου του Otsu στο MATLAB.....	14
Εικόνα 8. Gray scale εικόνα στην οποία θα γίνει εφαρμογή της μεθόδου του Karur.....	18
Εικόνα 9. Τρεις εφαρμογές της μεθόδου του Karur, όπου διακρίνουμε τις διαφορετικές κλάσεις του γκρι.....	18
Εικόνα 10. Εύρεση ακμών με τη μέθοδο του 'Sobel' στο MatLab στην επιγραφή της εικόνας 8.....	19
Εικόνα 11. Εύρεση ακμών με τη μέθοδο του 'Canny' στο MatLab στην επιγραφή της εικόνας 8.....	20
Εικόνα 12. Εφαρμογή της μεθόδου του 'Canny' του MatLab με αρκετά καλά αποτελέσματα.....	20
Εικόνα 13.Επάνω:αρχική εικόνα με ανομοιομορφία στη φωτεινότητα. Κάτω αριστερά: καθολική κατωφλίωση. Κάτω δεξιά τοπική κατωφλίωση.....	21
Εικόνα 14. Η συνάρτηση $d(B(x,y))$	25
Εικόνα 15. Προβλήματα ποιότητας σε δυαδικές εικόνες. (α) Θόρυβος με μορφή αλατοπίπερου.....	28
Εικόνα 16.Χρήση μασκών για την αφαίρεση θορύβου μορφής αλατοπίπερου. (α) Βάση των 4 γειτονικών pixels. (β) Βάσει των 8 γειτονικών pixels.....	29
Εικόνα 17. Αποτέλεσμα της χρήσης μασκών για την αφαίρεση θορύβου μορφής αλατοπίπερου (salt & pepper noise). (α) Αρχική εικόνα. (β) Αφαίρεση θορύβου χρησιμοποιώντας τη μάσκα των 8 γειτονικών pixels. (γ) Αφαίρεση θορύβου χρησιμοποιώντας τη μάσκα των 4 γειτονικών pixels.....	29
Εικόνα 18. Εξομάλυνση της επιφάνειας και βελτίωση της ποιότητας των χαρακτήρων με χρήση μασκών. (α)-(δ) Μάσκες που χρησιμοποιούνται. (ε) Τελικό αποτέλεσμα.....	30
Εικόνα 19. Μάσκα φίλτρου μέσης τιμής.....	30
Εικόνα 20. Λειτουργία φίλτρου μέσης τιμής.....	31
Εικόνα 21. Σημεία στα άκρα της εικόνας.....	31
Εικόνα 22. Λειτουργία μέσου φίλτρου (median filter).....	32
Εικόνα 23. Επάνω: εφαρμογή του κανόνα για dilation σε δυαδική εικόνα. Κάτω: εφαρμογή του κανόνα dilation για grayscale εικόνα.....	34
Εικόνα 24. Παράδειγμα dilation και erosion σε μια δυαδική εικόνα. (α) Το δομικό στοιχείο.....	35

Εικόνα 25. Αναπαράσταση του opening και του closing της εικόνας. (α) Αναπαράσταση του opening. (β) Αναπαράσταση του closing.	37
Εικόνα 26. Εφαρμογή μορφολογικών πράξεων για την προεπεξεργασία εικόνας εγγράφου.	38
Εικόνα 27. Παράδειγμα διαδοχικής εφαρμογής φίλτρων συρρίκνωσης (shrink) και διόγκωσης (swell) σε δυαδική εικόνα εγγράφου. (α) Αρχική ασπρόμαυρη εικόνα. (β) Τελική εικόνα μετά την εφαρμογή των φίλτρων.	39
Εικόνα 28. Ζώνες κατάτμησης της λέξης. Παράδειγμα από γραπτό Gurmukhi. (a) Άνω ζώνη (b) Μέση ζώνη (c) Κάτω ζώνη.	40
Εικόνα 29. Παράθυρο με συντεταγμένες (a,b) και (c,d).	41
Εικόνα 30. Κατάτμηση χαρακτήρων χωρίς να προηγηθεί μετατροπή της εικόνας σε ασπρόμαυρη. ...	42
Εικόνα 31. Ομαδοποίηση γραμμών κειμένου. Έχουμε δύο ομάδες γραμμών (πρώτη γραμμή - δεύτερη και τρίτη γραμμή).	43
Εικόνα 32. Χαρακτηριστικά γραμμής κειμένου με χαρακτήρες ίδιου πλάτους: Διάστημα ανάμεσα στους χαρακτήρες, κενό ανάμεσα στους χαρακτήρες, πλάτος χαρακτήρων.	43
Εικόνα 33. Εντοπισμός των χαρακτήρων με χρήση των κάθετων προβολών σε ομάδα.	44
Εικόνα 34. Morlet Wavelet	47
Εικόνα 35. Παράδειγμα προβολής γράμματος.	48
Εικόνα 36. Παράδειγμα κανονικοποίησης χαρακτήρων.	50
Εικόνα 37. Υπολογισμός της πυκνότητας των pixels σε ζώνες στις οποίες έχουμε χωρίσει τον χαρακτήρα. (α) 16 ζώνες. (β) 9 ζώνες που επικαλύπτουν τις πρώτες.	51
Εικόνα 38. Χρήση βαρών για τον υπολογισμό της πυκνότητας των pixels σε κάθε παράθυρο. (α) Τα βάρη σε κάθε παράθυρο. (β) Οι τιμές των χαρακτηριστικών που προκύπτουν.	52
Εικόνα 39. Εξαγωγή χαρακτηριστικών εικόνας.	52
Εικόνα 40. Παράδειγμα διανύσματος καθέτου ιστογράμματος. (α) Κάθετο ιστόγραμμα. (β) Διάνυσμα ιστογράμματος.	53
Εικόνα 41. Πάνω και κάτω προφίλ ολόκληρης λέξης.	54
Εικόνα 42. Ο αλγόριθμος k-means.	57
Εικόνα 43. Παράδειγμα κατηγοριοποίησης με χρήση του k-means.	58
Εικόνα 44. Η δομική μονάδα ενός νευρωνικού δικτύου, ένας νευρώνας.	59
Εικόνα 45. Δυαδική μετατροπή και αφαίρεση θορύβου. (α) Αρχική εικόνα. (β) Η εικόνα μετά την κατωφλίωση. (γ) Η εικόνα μετά τη χρήση του φίλτρου μέσου.	64
Εικόνα 46. Η επιγραφή της εικ. 45 μετά την εφαρμογή των μορφολογικών λειτουργιών open & dilation.	65
Εικόνα 47. Χαρακτήρες της επιγραφής της εικ. 46 ύστερα από την κατάτμηση. Σε κόκκινο κύκλο βρίσκονται οι χαρακτήρες που δεν έχουν αποκοπεί ορθά.	66
Εικόνα 48. Παράδειγμα επιγραφής της οποίας οι χαρακτήρες αποκόπηκαν χωρίς να παρουσιαστούν σφάλματα.	66

Εικόνα 49. Σημεία των χαρακτήρων, τα οποία χρησιμοποιεί ο αλγόριθμος για να εξάγει τα χαρακτηριστικά για μια εικόνα-γράμμα. 67

Εικόνα 50. Οι ομάδες των χαρακτήρων με τους οποίους εκπαιδεύτηκε το δίκτυο. Με τα κόκκινα βέλη δείχνουμε τον αριθμό που θα αντιπροσωπεύει στο πείραμα την κάθε κλάση. 70

Εικόνα 51. Η σειρά των χαρακτήρων που χρησιμοποιήθηκε για τη δοκιμή. Σαν αποτέλεσμα περιμένουμε στη θέση του κάθε χαρακτήρα τον αριθμό της κλάσης στην οποία ανήκει. 70

Λίστα πινάκων

Πίνακας 1. Κανόνες για dilation και erosion.....	34
Πίνακας 2. Κανόνες γεμίματος εικόνων κατά τη διαδικασία επεξεργασίας των συνοριακών pixels. 36	
Πίνακας 3. Απόσπασμα του πίνακα P με τα χαρακτηριστικά του κάθε χαρακτήρα.....	68

1. Εισαγωγή

Ένας από τους βασικότερους τομείς της αρχαιολογίας είναι και η επιγραφική, ο κλάδος που ασχολείται με τις αρχαίες επιγραφές. Ο συγκεκριμένος τομέας έχει μεγάλο πεδίο εφαρμογής σε παγκόσμιο επίπεδο και είναι καθοριστικής σημασίας για την ορθή καταγραφή της ιστορίας και όχι μόνον. Αυτό δεν ισχύει μόνο για τις ελληνικές επιγραφές, αλλά και για αυτές που αφορούν και σε άλλους αρχαίους πολιτισμούς, με χαρακτηριστικότερα παραδείγματα τους πολιτισμούς που αναπτύχθηκαν στη Μεσοποταμία, την Αίγυπτο, την Ιταλία. Στην Ελλάδα ο πλούτος των επιγραφών που έχει ανακαλυφθεί σε όλους τους αρχαιολογικούς χώρους είναι τεράστιος, κάτι που καθιστά τον τομέα της επιγραφικής έναν από τους πιο σημαντικούς. Οι πληροφορίες από τη μελέτη των επιγραφών είναι πολύτιμες, όχι μόνο για την αρχαιολογία αυτή καθαυτή, αλλά και για την ιστορία, τη φιλολογία, τη γλωσσολογία, τη θρησκευολογία, την ιστορία του δικαίου και των πολιτευμάτων, καθώς και για άλλες ανθρωπιστικές επιστήμες.

Προκειμένου να επιτευχθεί η όσο το δυνατόν καλύτερη αξιοποίηση των πληροφοριών αυτών, γίνεται κατάταξη των επιγραφών με βάση τη χρονολογία χάραξής τους, καθώς επίσης και του ποιος χαρακτήρας έχει χαραχθεί κάθε μία εξ αυτών. Μέχρι σήμερα η κατάταξη αυτή γίνεται με βάση συγκριτικές χρονολογήσεις, το «ένστικτο» και την εν γένη εμπειρία των επιγραφολόγων. Οι τρόποι αυτοί προσφέρουν ενδεικτικά και μόνον αποτελέσματα, γιατί κανείς από αυτούς τους τρόπους δεν οδηγεί με συνέπεια σε αντικειμενικά και αδιαμφισβήτητα συμπεράσματα, αλλά και οι διαδικασίες που ακολουθούνται απασχολούν αρκετό αριθμό ατόμων και είναι χρονοβόρες. Επομένως είναι ιδιαίτερος σημαντική για την επιγραφική η εξεύρεση μαθηματικοποιημένων κριτηρίων με τελικό στόχο την κατάταξη και ομαδοποίηση επιγραφών ανά χρονολογική περίοδο καθώς και ανά χαρακτήρα με ένα ποιο αντικειμενικό αλλά και αυτοματοποιημένο και λιγότερο χρονοβόρο τρόπο. Τα μαθηματικοποιημένα αυτά κριτήρια, σε αντίθεση με τα όσα εφαρμόζονται μέχρι σήμερα, δίνουν ποσοτικά αποτελέσματα, τα οποία με τη σειρά τους επιτρέπουν το διαχωρισμό και την κατάταξη των επιγραφών ακόμη και την ταυτοποίηση το χαρακτήρα¹.

Τις επιγραφές, αποσπάσματα των οποίων χρησιμοποιήσαμε στην παρούσα εργασία, παρεχωρήθησαν σε εμάς από τον κ. Ιωάννη Τζιφόπουλο, αναπληρωτή Καθηγητή του τμήματος Φιλολογίας της Φιλοσοφικής Σχολής του Αριστοτελείου Πανεπιστημίου Θεσσαλονίκης, ο οποίος ασχολείται ερευνητικά με τον τομέα της επιγραφικής.

1.1. Σκοπός και Στόχοι Εργασίας

Η τεχνολογική εξέλιξη στον τομέα της πληροφορικής μπορεί να αποτελέσει αρωγό στην προσπάθεια εκσυγχρονισμού των μεθόδων που μέχρι στιγμής χρησιμοποιούνται στην αρχαιολογική έρευνα και ειδικά στον τομέα της επιγραφικής. Η συγκεκριμένη εργασία στοχεύει στο να παρουσιάσει τα σημεία εκείνα στα οποία η πληροφορική μπορεί να βοηθήσει αποτελεσματικά και κυρίως τον τρόπο που μπορεί αυτό να γίνει. Ακόμα σκοπός της είναι με την ολοκλήρωσή της να αποτελέσει βάση έρευνας για μελλοντικές εργασίες τόσο ενδιαφερόμενων σπουδαστών όσο και του ίδιου του συντάκτη φοιτητή, αλλά και να εξοικειώσει τον δεύτερο με τις έννοιες, τεχνολογίες και μεθόδους που θα χρησιμοποιηθούν, παρέχοντας του εμπειρία χρήσιμη για δράσεις στη μεταπτυχιακή του πορεία.

1.2. Δομή εργασίας

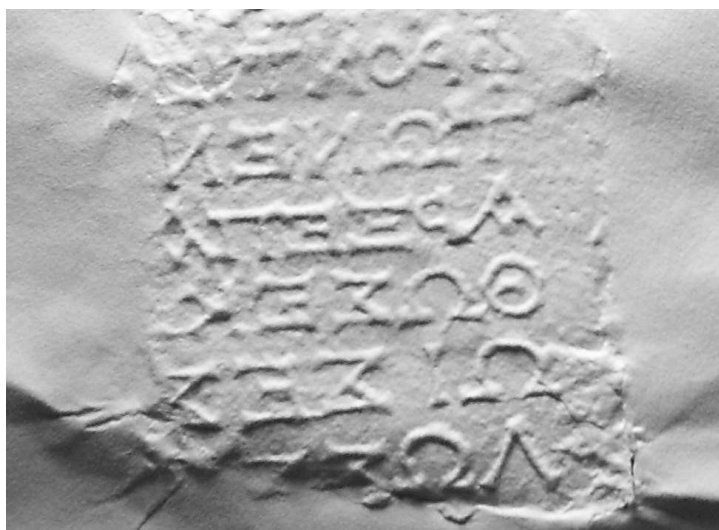
- A. Στο δεύτερο κεφάλαιο θα γίνει ανάλυση του προβλήματος, θα παρασχεθούν εξηγήσεις για το πού και πώς μπορεί ένα τέτοιο σύστημα να φανεί χρήσιμο και θα αναφερθούν τα εργαλεία που χρησιμοποιήθηκαν για την προσομοίωση.
- B. Στο τρίτο κεφάλαιο θα αναπτύξουμε το σχέδιο δράσης και θα γίνει λεπτομερέστερη περιγραφή του προβλήματος. Ακόμα, χωρίζοντας το συνολικό πρόβλημα σε υποπροβλήματα, θα αναλύσουμε τα βήματα που πρέπει να ακολουθηθούν για να φτάσουμε στο τελικό αποτέλεσμα. Στο θεωρητικό αυτό μέρος της εργασίας αναλύονται οι μέθοδοι και τεχνικές που βρήκαμε στη βιβλιογραφία.
- C. Στο τέταρτο κεφάλαιο θα παρουσιαστούν τα πειράματα τα οποία υλοποιούν τις τεχνικές που επιλέξαμε για την προσομοίωση του συστήματος και βασίζονται στα όσα θεωρητικά έχουν αναφερθεί στο τρίτο κεφάλαιο, καθώς και τα αποτελέσματά τους.
- D. Στο πέμπτο κεφάλαιο γίνεται ο σχολιασμός των αποτελεσμάτων και τα συμπεράσματα από τη θεωρητική και πειραματική εφαρμογή. Ακόμα θα εξεταστούν οι δυνατότητες περαιτέρω ανάπτυξης και εξέλιξης του συνόλου των τεχνικών αυτών και ο συνδυασμός τους με υπάρχουσες υλοποιήσεις, με σκοπό τη δημιουργία ολοκληρωμένων εφαρμογών.
- E. Στο έκτο κεφάλαιο παρατίθεται η βιβλιογραφία που χρησιμοποιήθηκε για την εκπόνηση της παρούσας εργασίας.

2. Ανάλυση προβλήματος

Κύριο αντικείμενο μελέτης μιας επιγραφής είναι φυσικά οι χαρακτήρες που είναι χαραγμένοι πάνω σε σκληρή επιφάνεια πέτρας, πηλού, μετάλλου κάθε είδους ή ξύλου. Από την ανάλυση κυρίως των χαρακτήρων μιας επιγραφής εξάγουμε τις σημαντικές για την επιγραφική πληροφορίες. Για τη μελέτη και ανάλυση των χαρακτήρων μιας επιγραφής, επειδή οι περισσότερες είναι στις αρχαιολογικές αποθήκες ή λόγω όγκου είναι τοποθετημένες στον χώρο όπου βρέθηκαν, ακολουθούνται κάποιες μέθοδοι δημιουργίας «αντιγράφων» με τα οποία ο επιγραφολόγος μπορεί να διεξαγάγει την έρευνα και μελέτη των επιγραφικών κειμένων.

2.1. Υπάρχουσες τεχνικές

Μέχρι στιγμής η μέθοδος για τη μελέτη των ενεπίγραφων θραυσμάτων που χρησιμοποιείται περισσότερο, επειδή δεν προκαλεί φθορές, είναι η κατασκευή των εκτύπων, η οποία έχει ως εξής: οι επιγραφολόγοι χρησιμοποιούν ένα ειδικό τύπο χαρτιού το οποίο βρεγμένο με νερό χτυπούν με ειδική βούρτσα έτσι ώστε να εφαρμοστεί στην επιφάνεια της επιγραφής. Όταν το χαρτί στεγνώσει, δημιουργείται ένα έκτυπο της επιγραφής σε ανάγλυφη τρισδιάστατη μορφή. Το είδος του χαρτιού που χρησιμοποιείται είναι ένα είδος χαρτιού-φίλτρου (στυπόχαρτο) του οποίου το πάχος ποικίλλει. Ένα λεπτότερο χαρτί χρειάζεται λιγότερη εργασία πάνω στην πέτρα, αλλά μπορεί να διαλυθεί πιο εύκολα². Άλλα μέσα τα οποία έχουν χρησιμοποιηθεί στο παρελθόν για το σκοπό αυτό, σύμφωνα με τη βιβλιογραφία, είναι η επάλειψη της ενεπίγραφης επιφάνειας με latex ή υγρό καουτσούκ, το οποίο αφού στεγνώσει παρέχει ένα στερεό αλλά δύσκολο στη μεταχείριση αντίγραφο, και το άπλωμα σκόνης κάρβουνου πάνω στην υγρή με νερό ενεπίγραφη επιφάνεια.



Εικόνα 1. Φωτογραφία έκτυπου σε στυπόχαρτο

Χρησιμοποιώντας την παραπάνω τεχνική το αποτέλεσμα είναι ότι έχουν δημιουργηθεί μεγάλες συλλογές εκτύπων από επιγραφές, οι οποίες εν τω μεταξύ έχουν σε πολλές περιπτώσεις αλλοιωθεί ή καταστραφεί τελείως. Τα έκτυπα αυτά φυλάσσονται σε διάφορες βιβλιοθήκες και ιδρύματα σε όλο τον κόσμο και διατίθενται προς μελέτη στους ερευνητές. Παρόλο που τα έκτυπα αποτελούν πιστά

αντίγραφα των επιγραφών, μπορούν να προκύψουν αρκετά προβλήματα κατά τη χρήση τους: η προσβασιμότητα σε αυτά είναι περιορισμένη και οι μελετητές πρέπει να ταξιδέψουν με σκοπό να εξετάσουν μια ποικιλία εκτύπων, από τον τρόπο φύλαξης και διατήρησης προκύπτει το πρόβλημα ότι, διπλώνοντας και ξεδιπλώνοντάς τα αρκετές φορές, μπορεί να προκληθεί η αλλοίωσή τους ή ακόμα και η καταστροφή τους.

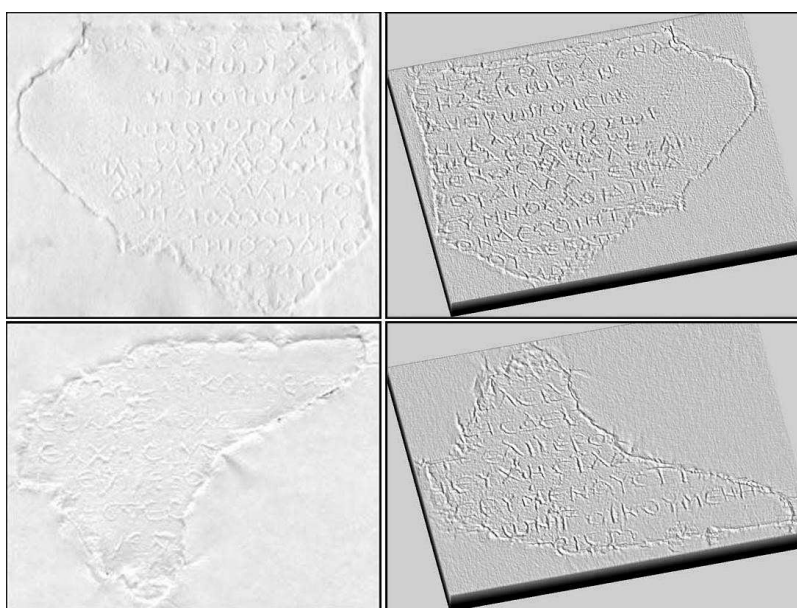
2.2. Εναλλακτικές εφαρμογές

Τα παραπάνω προβλήματα έχουν οδηγήσει τους επιστήμονες στο να αναζητήσουν νέες μεθόδους αποθήκευσης και διατήρησης των εκτύπων και γενικά των αντιγράφων των επιγραφών με σκοπό η πληροφορία εν τέλει να γίνει πιο εύκολα προσβάσιμη σε έναν όσο το δυνατόν μεγαλύτερο αριθμό ερευνητών.

Με την πάροδο του χρόνου και με την εξέλιξη της τεχνολογίας παρουσιάστηκαν νέες τεχνικές που θα μπορούσαν να δώσουν λύση στα πιο πάνω προβλήματα. Μια απλή προσέγγιση είναι η ψηφιακή δισδιάστατη (2D) φωτογράφιση των επιγραφών ή εναλλακτικά η σάρωση των εκτύπων ώστε μπορέσει να αποθηκευθεί η τρισδιάστατη πληροφορία τους σε ηλεκτρονική μορφή. Το πρόβλημα όμως με τις προαναφερθείσες μεθόδους είναι ότι μια δισδιάστατη (2D) εικόνα δεν μπορεί να αποτυπώσει πλήρως το τρισδιάστατο περιεχόμενο αυτών αντικειμένων και είναι επίσης ευαίσθητη στις συνθήκες φωτισμού και στη γωνία λήψης την φωτογραφικής μηχανής, ενώ και η σάρωση των εκτύπων απαιτεί εξελιγμένης μορφής και προς το παρόν κοστοβόρους σαρωτές.

Αρκετές από προκαθορισμένη θέση λήψης και κάτω από συγκεκριμένες συνθήκες φωτισμού φωτογραφίες μιας επιγραφής μπορούν να αποτυπώσουν με καλύτερο τρόπο τη χαραγμένη επιφάνεια. Σύμφωνα με την τεχνική αυτή, ένα ημισφαιρικό πλέγμα φωτεινών πηγών φωτίζει το αντικείμενο από μια σειρά συγκεκριμένων κατευθύνσεων. Στη συνέχεια η τοπική αμφίδρομη κατανομή φωτός μπορεί να προσεγγιστεί από ένα καθορισμένο σύστημα πολωνύμων, γνωστό ως Polynomial Texture Map. Μολονότι αυτή η μέθοδος λειτουργεί πολύ καλά στην εκμετάλλευση των χαρακτηριστικών των ανακλάσεων, η κατασκευή του ημισφαιρικού πλέγματος φωτεινών πηγών μπορεί να κοστίζει αρκετά, αν σκεφτεί κανείς ότι το μέγεθος του πλέγματος είναι ανάλογο του μεγέθους της επιγραφής, η οποία σε ορισμένες περιπτώσεις μπορεί να είναι αρκετά μεγάλη.

Μια video ακολουθία ή ένα σύνολο δισδιάστατων εικόνων μπορεί επίσης να χρησιμοποιηθεί για την ανακατασκευή ενός τρισδιάστατου σκηνικού. Αυτή η τεχνική είναι πολύ αποτελεσματική και απαιτεί μόνο μια κάμερα μη αυτόματης βαθμονόμησης. Τέτοιες μέθοδοι έχουν χρησιμοποιηθεί επιτυχημένα για την 3D ανακατασκευή αρχαιολογικών χώρων, συμπεριλαμβανομένων κτηρίων και περίπλοκων αγαλμάτων³.



Εικόνα 2. Αριστερά τα έκτυπα και δεξιά οι 3D ανακατασκευές τους.

Ωστόσο, η ακρίβεια και η λεπτομέρεια των ανακατασκευών που προκύπτουν δεν είναι αρκετά υψηλές για να καλύψουν λεπτομέρειες χαραγμένων επιφανειών όπως γράμματα και σύμβολα.

Επίσης, ένας πιο εξελιγμένος τρόπος είναι η χρήση laser scanner, συσκευή η οποία μπορεί να αποτυπώσει με έναν πολύ λεπτομερή τρόπο τη χαραγμένη επιφάνεια. Αυτή η μέθοδος έχει εφαρμοστεί σε ιερογλυφικά και δείχνει πολλά υποσχόμενη. Στα μειονεκτήματά της, ωστόσο, συγκαταλέγεται το γεγονός ότι η κατασκευή ενός laser σαρωτή μπορεί να αποδειχθεί εξαιρετικά δαπανηρή και η μεταφορά και χρήση του ενδέχεται να μην είναι πάντα εφικτή, αν αναλογιστεί κανείς την περιορισμένη πρόσβαση από πολλές ομάδες ερευνητών.

2.3. Προτεινόμενη λύση

Έναν άλλο, λιγότερο δαπανηρό και απλούστερο τρόπο που δε χρειάζεται τρισδιάστατη απεικόνιση θα εξετάσουμε στην παρούσα εργασία. Το σύστημα το οποίο θα επιχειρήσουμε να προσομοιώσουμε θα λαμβάνει ως πληροφορία μια ψηφιακή δισδιάστατη εικόνα επιγραφής ή έκτυπου και θα προσπαθεί να κατηγοριοποιήσει και ταυτοποιήσει με αυτόματο τρόπο τους χαρακτήρες με βάση την πληροφορία που εξάγεται από τις ψηφιακές απεικονίσεις των αρχαίων επιγραφών, χωρίς καμία αναφορά στο περιεχόμενο του κειμένου της επιγραφής, τη συντακτική και γραμματική του δομή. Αυτό μπορεί να επιτευχθεί εξάγοντας συγκεκριμένα χαρακτηριστικά των γραμμάτων με σκοπό τη δημιουργία ενός πίνακα-άτλαντα χαρακτήρων για κάθε επιγραφή, δηλαδή του ιδιαίτερου αλφαβήτου του χαρακτήρα της κάθε επιγραφής. Ο πίνακας αυτός μπορεί να φανεί εξαιρετικά χρήσιμος στην επιγραφική έρευνα, καθώς από αυτόν μπορούν να εξαχθούν πληροφορίες για τη χρονολόγηση της κάθε επιγραφής, την εξακρίβωση της προέλευσής της, τις ιδιομορφίες των γραμμάτων και το στυλ του χαρακτήρα. Στην ουσία, λαμβάνοντας υπ' όψιν τα προβλήματα που εμφάνισαν μέθοδοι που ήδη χρησιμοποιούνται, θα τροποποιήσουμε ένα σύστημα οπτικής αναγνώρισης χαρακτήρων (OCR), ώστε να είναι εφικτή η εφαρμογή του σε μια επιγραφή. Δυσκολίες που πιθανόν προκύψουν, θα εξεταστούν σε ακόλουθα κεφάλαια όπου θα προταθούν και ανάλογες λύσεις ώστε αυτές να ξεπεραστούν.

Όσον αφορά τα OCR συστήματα, δηλαδή τον εντοπισμό αλφαριθμητικών συμβόλων σε εικόνες και την αναγνώρισή τους, αυτά αποτελούν τομέα της έρευνας για αρκετές δεκαετίες, αφού η προσομοίωση της όρασης του ανθρώπου με αυτή των μηχανών είναι ένας τομέας με ανεξάντλητο ενδιαφέρον. Εφαρμογές αυτών των μεθόδων θα χρησιμοποιήσουμε στα πειράματά μας τα οποία θα γίνουν στη σουίτα MATLAB, η οποία παρέχει πληθώρα εργαλείων και βιβλιοθηκών, και στην οποία καθίσταται πιο κατανοητή η όλη διαδικασία.

2.4. Ψηφιακή Επεξεργασία Εικόνας

Ψηφιακή επεξεργασία εικόνας είναι η χρήση αλγορίθμων υπολογιστών για την εκτέλεση επεξεργασίας εικόνας σε ψηφιακές εικόνες. Ως υποκατηγορία ή τομέας της ψηφιακής επεξεργασίας σήματος, η ψηφιακή επεξεργασία εικόνας έχει πολλά πλεονεκτήματα σε σχέση με την αναλογική επεξεργασία. Κατ' αρχάς, επιτρέπει ένα πολύ μεγαλύτερο φάσμα αλγορίθμων, οι οποίοι εφαρμόζονται στα δεδομένα εισόδου και βοηθούν στο να αποφευχθούν προβλήματα όπως η συσσώρευση θορύβου και η παραμόρφωση του σήματος κατά την επεξεργασία. Επίσης οι αλγόριθμοι αυτοί που χρησιμοποιούνται είναι πολύ μεγαλύτερης πολυπλοκότητας και, ως εκ τούτου, μπορούν να προσφέρουν πολύ πιο εξελιγμένες επιδόσεις σε απλές εργασίες, αλλά και εφαρμογές μεθόδων που θα

ήταν αδύνατες με τον αναλογικό τρόπο. Ειδικότερα, η ψηφιακή επεξεργασία εικόνας βρίσκει πρακτικά εφαρμογή στην:

- Κατηγοριοποίηση
- Εξαγωγή χαρακτηριστικών
- Αναγνώριση προτύπων
- Προβολή-Σχεδίαση
- Multi-level ανάλυση σήματος,

χρησιμοποιώντας τεχνικές όπως η ψηφιοποίηση, η ανάλυση κύριων συνιστωσών (Principal Components Analysis), νευρωνικά δίκτυα και αρκετούς μετασχηματισμούς (wavelet, Fourier, Laplace κ.α).

2.5. Matlab

Το MATLAB είναι μια υψηλού επιπέδου γλώσσα προγραμματισμού και ένα διαδραστικό περιβάλλον για την ανάπτυξη αλγορίθμων, την οπτικοποίηση δεδομένων, την ανάλυση δεδομένων και την πραγματοποίηση αριθμητικών υπολογισμών. Το όνομα του προκύπτει από την συνένωση των αρχικών χαρακτήρων των λέξεων **Matrix Laboratory**. Όπως και το όνομά του προδίδει, είναι ειδικά σχεδιασμένο για την εκτέλεση υπολογισμών με πίνακες.



Χρησιμοποιώντας το προϊόν MATLAB μπορούν να λυθούν τεχνικά προβλήματα υπολογιστών ταχύτερα από ό,τι με τις παραδοσιακές γλώσσες προγραμματισμού, όπως C, C++ και Fortran.

Εικόνα 3. Το λογότυπο του MATLAB

Το MATLAB μπορεί να χρησιμοποιηθεί σε ένα ευρύ φάσμα εφαρμογών, συμπεριλαμβανομένης της ψηφιακής επεξεργασίας σήματος και εικόνας, σε εφαρμογές στον τομέα των επικοινωνιών, στην ανάλυση και σχεδιασμό οικονομικών μοντέλων στην υπολογιστική βιολογία κ.α. Οι πρόσθετες εργαλειοθήκες (συλλογές ειδικού σκοπού MATLAB συναρτήσεων, που παρέχονται ξεχωριστά⁴ επεκτείνουν το περιβάλλον του MATLAB για την επίλυση συγκεκριμένων κατηγοριών προβλημάτων σε αυτούς τους τομείς εφαρμογών. Χαρακτηριστικά που περιλαμβάνονται είναι:

- Υψηλού επιπέδου γλώσσα για τεχνικούς υπολογιστών
- Ανάπτυξη περιβάλλοντος για τη διαχείριση κώδικα, αρχεία και δεδομένα
- Διαδραστικά εργαλεία για την επαναληπτική έρευνα, σχεδιασμό και την επίλυση προβλημάτων
- Μαθηματικές συναρτήσεις για γραμμική άλγεβρα, στατιστικά στοιχεία, ανάλυση Fourier, το φιλτράρισμα, τη βελτιστοποίηση και αριθμητική ολοκλήρωση
- 2-D και 3-D λειτουργίες γραφικών για οπτικοποίηση δεδομένων
- Εργαλεία για τη δημιουργία προσαρμοσμένων γραφικών διεπαφών χρήστη
- Λειτουργίες για την ενσωμάτωση MATLAB αλγορίθμων με εξωτερικές εφαρμογές και γλώσσες, όπως η C, C++, Fortran, Java™, COM, και η Microsoft®, Excel®.

2.5.1. Γιατί Matlab;

Το περιβάλλον του MatLab προτιμήθηκε στην παρούσα εργασία, γιατί με τα εργαλεία και τις συναρτήσεις των εργαλειοθηκών (π.χ. Image Processing Toolbox, Neural Network Toolbox) που, όπως προαναφέρθηκε, παρέχει καθιστά απλούστερη τη διεξαγωγή των προσομοιώσεων-πειραμάτων και ευκολότερη την παρουσίαση των αποτελεσμάτων αυτών.

2.5.2. Matlab Image Processing Toolbox

Η εργαλειοθήκη επεξεργασίας εικόνας (Image Processing Toolbox) παρέχει μια πλήρη σειρά τυποποιημένων αλγορίθμων και γραφικών εργαλείων για την επεξεργασία εικόνας, την ανάλυση, την απεικόνιση, και την ανάπτυξη αλγορίθμων. Με τη χρήση του μπορούν να εκτελεστούν εργασίες για τη βελτίωση της εικόνας, την ανίχνευση χαρακτηριστικών, τη μείωση του θορύβου, την κατάτμηση εικόνας, για μετασχηματισμούς διαστημάτων, και την καταγραφή της εικόνας. Πολλές λειτουργίες στην εργαλειοθήκη είναι πολυνηματικές (multithreaded), με αποτέλεσμα να είναι εκμεταλλεύσιμες λειτουργίες υπολογιστών πολλαπλών πυρήνων και πολλαπλών επεξεργαστών.

(<http://www.mathworks.com/products/image/#thd1>)

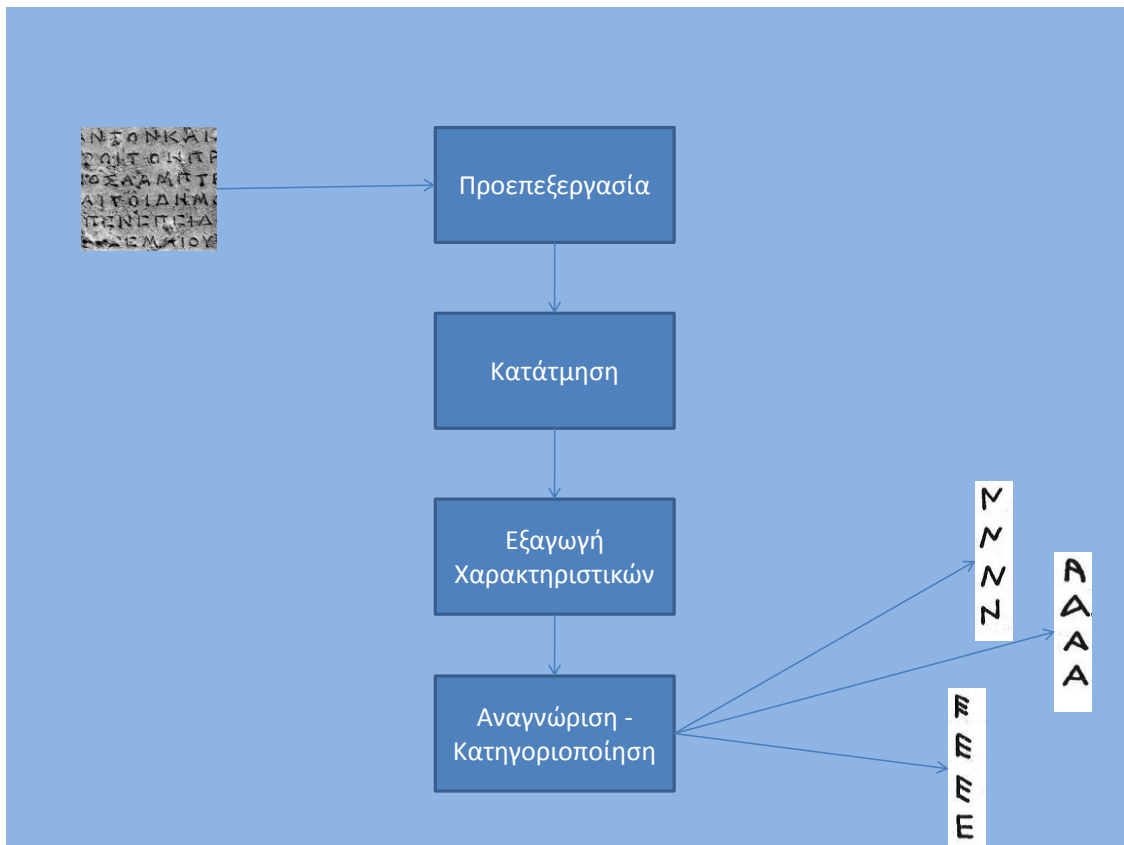
2.5.3. Matlab Neural Network Toolbox

Η εργαλειοθήκη νευρωνικών δικτύων (Neural Network Toolbox) παρέχει εργαλεία για τον σχεδιασμό, την εφαρμογή, την οπτικοποίηση και την προσομοίωση νευρωνικών δικτύων. Τα νευρωνικά δίκτυα χρησιμοποιούνται για τις εφαρμογές όπου κάποια απλή ανάλυση θα ήταν δύσκολο έως αδύνατο να παράσχει αποτελέσματα, όπως η αναγνώριση προτύπων και η μη-γραμμική αναγνώριση και έλεγχος συστήματος. Η εργαλειοθήκη υποστηρίζει feed forward δίκτυα, ακτινικής βάσης δίκτυα, δυναμικά δίκτυα, self-organizing χάρτες, και άλλα αποδεδειγμένα παραδείγματα δικτύου. (<http://www.mathworks.com/products/neuralnet/>)

3. Μεθοδολογία Υλοποίησης

Για την υλοποίηση του προτεινόμενου συστήματος χρειάζεται να διαιρέσουμε τις εργασίες που πρέπει να γίνουν σε κάποια βασικά στάδια, τα οποία είναι παρόμοια στα περισσότερα συστήματα οπτικής αναγνώρισης χαρακτήρων. Αναλύοντας το καθένα ξεχωριστά, θα εξετάσουμε τις τεχνικές που μπορούν να χρησιμοποιηθούν και ποιες είναι αυτές που μπορούν να μας δώσουν το καλύτερο δυνατό αποτέλεσμα. Τα βασικά αυτά στάδια είναι τα ακόλουθα:

- Προεπεξεργασία
- Κατάτμηση
- Εξαγωγή Χαρακτηριστικών
- Αναγνώριση - Κατηγοριοποίηση



Εικόνα 4. Δομή συστήματος.

3.1. Προεπεξεργασία

Περιλαμβάνει τη δυαδική μετατροπή (μετατροπή της gray scale εικόνας σε ασπρόμαυρης), τη βελτίωση της ποιότητας (εξάλειψη θορύβου, βελτίωση της ποιότητας του κειμένου) και τη διόρθωση της στροφής της εικόνας (διόρθωση στροφής της εικόνας η οποία έχει προκύψει λόγω μη ευθυγραμμισμένης φωτογραφικής μηχανής, ή μη ευθυγραμμισμένης τοποθέτησης του προς εξέταση αντικειμένου στον σαρωτή).

3.1.1. Δυαδική μετατροπή

Η δυαδική μετατροπή των εικόνων (binarization - thresholding) είναι το πρώτο από τα βήματα των περισσότερων συστημάτων ανάλυσης και επεξεργασίας εγγράφων και αναφέρεται στην μετατροπή των gray scale εικόνων σε ασπρόμαυρες. Η δυαδική μετατροπή αποτελεί βασικό στοιχείο των συστημάτων επεξεργασίας εγγράφων, γιατί μία βέλτιστη δυαδική μετατροπή αποτελεί τη βάση για σωστή κατάτμηση και αναγνώριση των εγγράφων. Οι βασικοί λόγοι για τους οποίους το στάδιο της δυαδικής μετατροπής είναι απαραίτητο είναι οι παρακάτω:

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

Στη βιβλιογραφία η δυαδική μετατροπή μπορεί να χρησιμοποιείται και ως τεχνική εντοπισμού του κειμένου, αφού συνήθως διακρίνει τις περιοχές του κειμένου από τις περιοχές του υποβάθρου της εικόνας. 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 κατωφλίου (threshold), σύμφωνα με το οποίο όλες οι γκρι στάθμες κάτω από αυτό μετατρέπονται σε 1 ενώ εκείνες που είναι πάνω από το κατώφλι μετατρέπονται σε 0 ή και το αντίστροφο:

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

Οι μέθοδοι, που χρησιμοποιούν μια τιμή κατωφλίου T για τη μετατροπή της εικόνας σε ασπρόμαυρη σύμφωνα με τον πιο πάνω τύπο, ονομάζονται *μέθοδοι καθολικής κατωφλίωσης* (global thresholding methods).

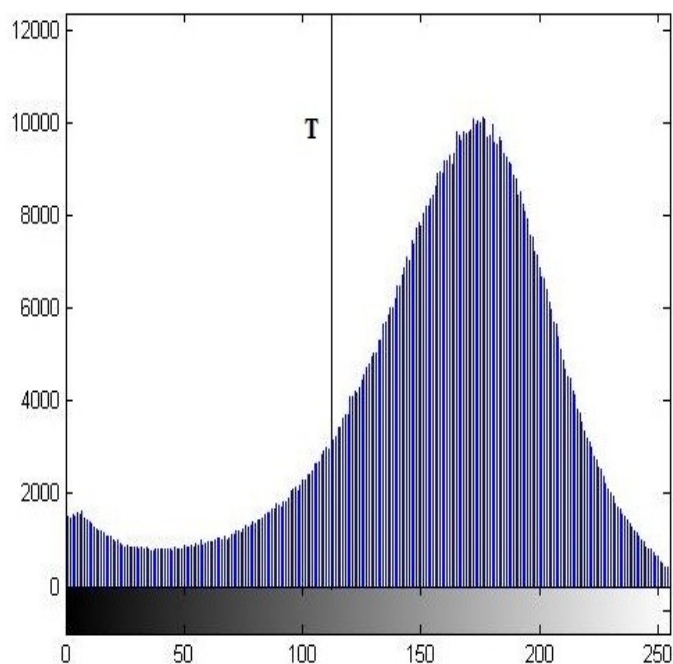
Εκτός από τις μεθόδους καθολικής κατωφλίωσης, στη δυαδική μετατροπή των εικόνων χρησιμοποιούνται και άλλες τεχνικές. Γενικά οι μέθοδοι κατωφλίωσης σύμφωνα με τους **Sezgin and Sankur**⁵ χωρίζονται στις έξι (6) ακόλουθες κατηγορίες, ανάλογα με το είδος της πληροφορίας που οι αλγόριθμοι τους διαχειρίζονται. Έτσι τις καταθέτουμε σε αυτές που λειτουργούν βάση:

- Ιστογράμματος (histogram shape-based methods), όπου για παράδειγμα οι κορυφές, οι κοιλάδες και οι καμπές σε ένα ιστόγραμμα αναλύονται.
- Ομαδοποίησης (clustering-based methods), όπου τα δείγματα επιπέδων του γκρι χωρίζονται σε δυο ομάδες σαν προσκήνιο και φόντο (foreground-background), ή εναλλακτικά μοντελοποιούνται σαν μίξη δυο Gaussian κατανομών.
- Εντροπίας (entropy-based methods), όπου οι αλγόριθμοί τους χρησιμοποιούν την εντροπία του προσκήνιου και του φόντου, αυτή μεταξύ αρχικής εικόνας και δυαδικής, κλπ.

- Χαρακτηριστικών των αντικειμένων, όπου ερευνούν το μέτρο ομοιότητας μεταξύ των γκριζαρισμένων (gray-level) και δυαδικών εικόνων, αντιστοίχιση - σύμπτωση άκρων, ασαφή ομοιότητα σχημάτων, σε έγχρωμες χρωματικές ομοιότητες κλπ.
- Χωρικών μεθόδων, όπου χρησιμοποιούν ανωτέρας τάξης κατανομές πιθανοτήτων και/ή συσχέτιση μεταξύ των pixels.
- Τοπικού κατωφλίου μέθοδοι, όπου προσαρμόζουν την τιμή του κατωφλίου σε κάθε pixel των τοπικών χαρακτηριστικών της εικόνας.

Εφαρμογές και παραδείγματα ορισμένων από τις παραπάνω μεθόδους και πως αυτές μπορούν ή όχι να βοηθήσουν στην εργασία θα δούμε στις παραγράφους που ακολουθούν, αρχίζοντας με μεθόδους καθολικής κατωφλίωσης.

Ένας απλός τρόπος για να επιλέξουμε ένα κατώφλι είναι μέσω οπτικής επιθεώρησης του ιστογράμματος της εικόνας. Το ιστόγραμμα της *Εικόνας 5* φαίνεται ότι έχει δύο ξεχωριστές κορυφές, με αποτέλεσμα να είναι εύκολο να επιλέξεις ένα κατώφλι ανάμεσα σε αυτές.



Εικόνα 5. Επιλογή κατωφλίου, μέσω οπτικής ανάλυσης του ιστογράμματος της εικόνας.

Ένας άλλος τρόπος να επιλεγεί ένα κατώφλι είναι με τη δοκιμή διαφορετικών τιμών, έως ότου βρεθεί εκείνη η τιμή που θα παράγει το καλύτερο αποτέλεσμα σύμφωνα με την κρίση του παρατηρητή (Εικόνα 6). Αυτή η μέθοδος ειδικά θα μπορούσε να είναι αρκετά αποτελεσματική σε ένα διαδραστικό περιβάλλον, το οποίο θα επιτρέπει στο χρήστη να μεταβάλλει το κατώφλι με τη χρήση κάποιου 'εργαλείου' (widget) όπως ένα slider, για να βλέπει άμεσα τα αποτελέσματα.



Εικόνα 6. Ασπρόμαυρες εικόνες που προκύπτουν από την εφαρμογή καθολικής κατωφλίωσης στην επιγραφή της Εικόνας 5 επιλέγοντας διαφορετικές τιμές ($T_a < T_b < \dots < T_z$) για το κατώφλι.

Για την αυτόματη επιλογή κατωφλίου, οι Gonzalez & Woods⁶ στο βιβλίο τους περιγράφουν την ακόλουθη επαναληπτική διαδικασία:

- Κάντε μια πρώτη εκτίμηση του T (εννοούμε το κατώφλι, 'eng: Threshold'). Μια προτεινόμενη αρχική εκτίμηση είναι το ενδιάμεσο σημείο μεταξύ των ελάχιστων και των μέγιστων τιμών έντασης της εικόνας.
- Κάντε την κατάτμηση της εικόνας χρησιμοποιώντας το T . Αυτή η διαδικασία θα παραγάγει δύο ομάδες από pixels: την G_1 , η οποία θα περιέχει όλα τα pixels με τιμές έντασης μεγαλύτερες ή ίσες του T ($\geq T$) και την G_2 , η οποία θα περιέχει τα pixels με τιμές μικρότερες του T ($< T$).
- Υπολογίστε τις μέσες τιμές έντασης μ_1 και μ_2 για τα pixels τις περιοχές G_1 και G_2 , αντίστοιχα.

- Υπολογίστε μια νέα τιμή κατωφλίου σύμφωνα με τον τύπο:

$$T = \frac{1}{2}(\mu_1 + \mu_2)$$

- Επαναλάβετε τα βήματα 2 έως 4 μέχρις ότου η διαφορά του T σε διαδοχικές επαναλήψεις να είναι μικρότερη από μια προκαθορισμένη παράμετρο T₀.

Η πιο πάνω διαδικασία η οποία ονομάζεται ‘Επιλογή μετά από επανάληψη’ διατυπώθηκε από τον **Ridler**⁷ και στην ουσία η μέθοδος αυτή ξεκινά από μία αρχική εκτίμηση για την τιμή του κατωφλίου, ενώ στη συνέχεια γίνεται διόρθωση αυτής της τιμής λαμβάνοντας υπόψη τις περιοχές κειμένου και υποβάθρου που προκύπτουν κάθε φορά. Η αρχική εκτίμηση του κατωφλίου 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 στη μεταβλητή T_b. Αντίστοιχα, υπολογίζεται και η μέση τιμή των gray scale τιμών των pixels πάνω από το αρχικό κατώφλι και αποθηκεύεται σε μια μεταβλητή. Στη συνέχεια υπολογίζεται μία νέα προσέγγιση του κατωφλίου και η διαδικασία επαναλαμβάνεται μέχρι τελικά να μην υπάρχει αλλαγή στο κατώφλι μεταξύ δύο διαδοχικών επαναλήψεων.

Από τις σημαντικότερες και περισσότερο χρησιμοποιημένες μεθόδους καθολικής κατωφλίωσης είναι η μέθοδος του **Otsu**⁸. Η εργαλειοθήκη του MATLAB που χρησιμοποιούμε παρέχει μια συνάρτηση, την επονομαζόμενη *graythresh*, η οποία για τον υπολογισμό του κατωφλίου χρησιμοποιεί την μέθοδο του Otsu. Η μέθοδος του Otsu βασίζεται στην επεξεργασία του ιστογράμματος της εικόνας και στον προσδιορισμό του κατωφλίου βάσει του κριτηρίου της μεγιστοποίησης της διαχωριστικότητας 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:

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

και N το συνολικό πλήθος των pixels της εικόνας.

Για κάθε υποψήφιο κατώφλι t το ιστόγραμμα της εικόνας χωρίζεται σε δυο περιοχές – κλάσεις. Μπορούμε να υπολογίσουμε 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 επί τη συνολική μέση τιμή όλων των pixels και υπολογίζεται από τον παρακάτω τύπο:

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

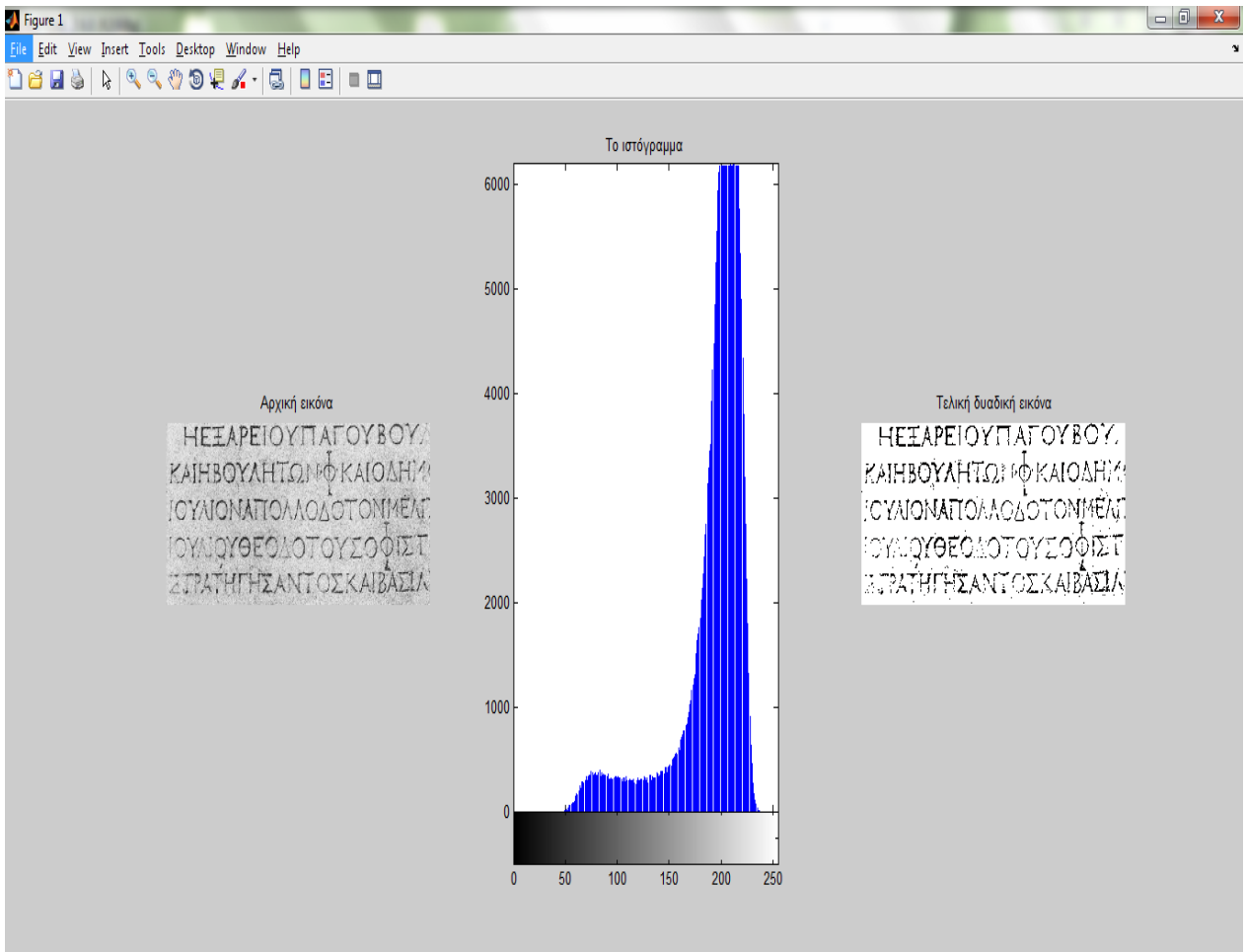
όπου τα ω_0 , ω_1 , μ_0 και μ_1 υπολογίζονται από τους τύπους:

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

Σύμφωνα με τον Otsu, η βέλτιστη τιμή για το κατώφλι επιτυγχάνεται όταν μεγιστοποιείται ο λόγος του τετραγώνου της τυπικής απόκλισης μεταξύ των δυο 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 μια από τις καλύτερες τεχνικές εύρεσης καθολικού κατωφλίου.



Εικόνα 7. Εφαρμογή της μεθόδου του Otsu στο MATLAB

Αρκετές αναφορές στη βιβλιογραφία των τεχνικών κατωφλίωσης έχουν και οι τεχνικές βασιζόμενες στην εντροπία, όπως αυτές των **Pun** και **Kapur**⁹.

Η εντροπία στη θεωρία της πληροφορίας είναι ένα «μέτρο αβεβαιότητας» που διακατέχει ένα σύστημα. Ο όρος *εντροπία* χρησιμοποιήθηκε αρχικά στη θερμοδυναμική. Στη θεωρία πληροφορίας εισήχθη από τον *Κλωντ Σάνον* (Claude Shannon) το 1948 και γι' αυτό ονομάζεται και *εντροπία του Σάνον*^{10 11}. Η εντροπία της θερμοδυναμικής μπορεί να ιδωθεί ως εφαρμογή της εντροπίας στη θεωρία πληροφορίας.

Η εντροπία είναι μια μετρική που σχετίζεται με το περιεχόμενο της πληροφορίας. Η μέθοδος κατωφλίωσης με χρήση της εντροπίας διαμορφώθηκε από τον Pun. Η εντροπία του ιστογράμματος μετά την κατωφλίωση επιλέγεται σαν κριτήριο. Αν έχουμε n πιθανά σύμβολα (π.χ. γράμματα ή ψηφία) και το σύμβολο i εμφανίζεται με πιθανότητα $p(i)$, η εντροπία που σχετίζεται με τα σύμβολα είναι σύμφωνα με τον ορισμό του Σάνον:

$$H = - \sum_{i=0}^n p_i \log(p_i).$$

Μετά από δύο επίπεδα κατωφλίωσης, η εικόνα έχει μόνο δύο κλάσεις: μαύρη (D^b) και άσπρη (D^w), όπου $D^b \cup D^w = D$ και $D^b \cap D^w = \emptyset$. Η εντροπία της εικόνας είναι:

$$H = -p_w \log(p_w) - p_b \log(p_b),$$

όπου $p_w = P(D^w)$ και $p_b = P(D^b)$. Για να ικανοποιηθεί η συνθήκη ότι η εντροπία είναι η μέγιστη, το κατώφλι πρέπει να επιλεγεί σαν $p_w = p_b = \frac{1}{2}$. Το κατώφλι δεν είναι προσαρμοστικό. Επίσης, μπορούμε να δούμε ότι ανεξαρτήτως του πόσα pixels έχει το αντικείμενο (η χρήσιμη πληροφορία) στην αρχική εικόνα, μετά την κατωφλίωση πρέπει πάντα να απομείνουν τα μισά από τον συνολικό αριθμό των pixels. Τα pixels του αντικειμένου σε μια εικόνα είναι πολύ πιθανό να μην είναι ακριβώς τα μισά των pixels ολόκληρης της εικόνας. Αυτό σημαίνει ότι πρέπει να διαγράψουμε κάποια Pixels του αντικειμένου αν είναι περισσότερα από τα μισά, ή, αν είναι λιγότερα, να θεωρήσουμε μερικά pixels από το φόντο σαν στοιχεία της πληροφορίας για να επέλθει η ισορροπία. Αυτό μπορεί να οδηγήσει σε κακή κατάσταση της εικόνας.

Ο Pun παρουσιάζει μια νέα προσέγγιση που προέρχεται από την προηγούμενη εντροπική κατωφλίωση. Ο *συντελεστής ανισοτροπίας* α εισάγεται και ορίζεται:

$$\alpha = \left(\sum_{i=0}^t p_i \cdot \log(p_i) \right) / \left(\sum_{i=0}^n p_i \cdot \log(p_i) \right),$$

όπου το t είναι το κατώφλι και το α χρησιμοποιείται για να περιγράψει το γεωμετρικό σχήμα του ιστογράμματος. Το κατώφλι t επιλέγεται:

$$\sum_{i=0}^t p_i = \frac{1}{2} + \left| \frac{1}{2} - a \right| = \begin{cases} 1 - a, & a \leq \frac{1}{2}, \\ a, & a > \frac{1}{2}. \end{cases}$$

Σαν επέκταση, στην k -level κατώφλωση, τα $k - 1$ κατώφλια t_1, t_2, \dots, t_{k-1} , πρέπει να επιλεγούν. Μπορούμε να βρούμε $k - 1$ επίπεδα s_1, s_2, \dots, s_{k-1} να διαιρούν το ιστόγραμμα σε k ίσα μέρη και οι $k - 1$ συντελεστές ανιστροποίας a_1, a_2, \dots, a_{k-1} ορίζονται:

$$a_1 = \left(\sum_{i=0}^{s_1} p_i \cdot \log(p_i) \right) / \left(\sum_{i=0}^n p_i \cdot \log(p_i) \right),$$

$$a_{k-1} = \left(\sum_{i=0}^{s_{k-1}} p_i \cdot \log(p_i) \right) / \left(\sum_{i=0}^n p_i \cdot \log(p_i) \right).$$

και τα $k - 1$ κατώφλια t_1, t_2, \dots, t_{k-1} , επιλέγονται ως:

$$\sum_{i=0}^{t_1} p_i = \sum_{i=0}^{s_1} p_i + \left| \sum_{i=0}^{s_1} p_i - a \right|,$$

$$\sum_{i=0}^{t_{k-1}} p_i = \sum_{i=0}^{s_{k-1}} p_i + \left| \sum_{i=s_{k-2}+1}^{s_{k-1}} p_i - a_{k-1} \right|.$$

Αυτή η νέα προσέγγιση μπορεί να προσαρμόσει το κατώφλι σύμφωνα με το σχήμα του ιστογράμματος. Συγκρινόμενη με την αρχική μέθοδο, η οποία περιορίζει τα κατώφλια που επιλέγονται να χωρίσουν το ιστόγραμμα σε ίσα μέρη, αυτή είναι προσαρμοζόμενη και πετυχαίνει καλά αποτελέσματα σε κάποιες εικόνες. Όμως καθώς προέρχεται από παραλλαγή της προηγούμενης μεθόδου (ίδια περίπου φιλοσοφία), η κατανομή των πιθανοτήτων είναι περίπου η ίδια. Στην πραγματικότητα γνωρίζουμε ότι κάθε 'κλάση' στην εικόνα μπορεί να καταλαμβάνει διάφορα τμήματά της. Επίσης η εισαγωγή του 'α' εισήγαγε νέα προβλήματα. Δηλαδή σε κάποιες εικόνες με ασυνήθιστο ιστόγραμμα, αυτή η μέθοδος αποτυγχάνει να παραγάγει αποδεκτά αποτελέσματα.

Ο **Kapur** επίσης χρησιμοποίησε τη γενική ιδέα του Σάνον για την εντροπία αλλά από διαφορετική οπτική γωνία. Δύο κατανομές πιθανοτήτων ολόκληρης της εικόνας αντί μίας ερευνώνται σε αυτή τη μέθοδο. Μία για το προσκίνητο (αντικείμενο - πληροφορία) και μια άλλη για το φόντο. Το άθροισμα από τις εντροπίες φόντου και πληροφορίας στη συνέχεια μεγιστοποιείται. Με άλλα λόγια, αυτό θα οδηγήσει σε ισοτίμων πιθανοτήτων γκρι επίπεδα σε κάθε περιοχή, μεγιστοποιώντας έτσι το

άθροισμα της ομοιογένειας στα επίπεδα του γκρι στο αντικείμενο και το φόντο, αφού τα επίπεδα του γκρι καθίστανται εξίσου πιθανά σε οποιαδήποτε περιοχή.

Η κατανομή πιθανοτήτων των επιπέδων του γκρι στα μαύρα μέρη της εικόνας είναι:

$$\frac{p_0}{P_B}, \frac{p_1}{P_B}, \dots, \frac{p_s}{P_B},$$

και για τα λευκά μέρη είναι:

$$\frac{p_{s+1}}{1 - P_B}, \frac{p_{s+2}}{1 - P_B}, \dots, \frac{p_{n-1}}{1 - P_B},$$

στα οποία, s είναι το κατώφλι, $p_i (i=0, 1, \dots, n-1)$ είναι η στατιστική πιθανότητα των pixels με επίπεδο γκρι i σε ολόκληρη την εικόνα, P_B είναι η πιθανότητα των pixels με επίπεδο γκρι μικρότερο ή ίσο του κατωφλίου s :

$$P_B = \sum_{i=0}^s p_i$$

Η εντροπία του μαύρου μέρους (πληροφορίας) της εικόνας είναι:

$$H_B^{(s)} = - \sum_{i=0}^s \frac{p_i}{P_B} \log_2 \left(\frac{p_i}{P_B} \right)$$

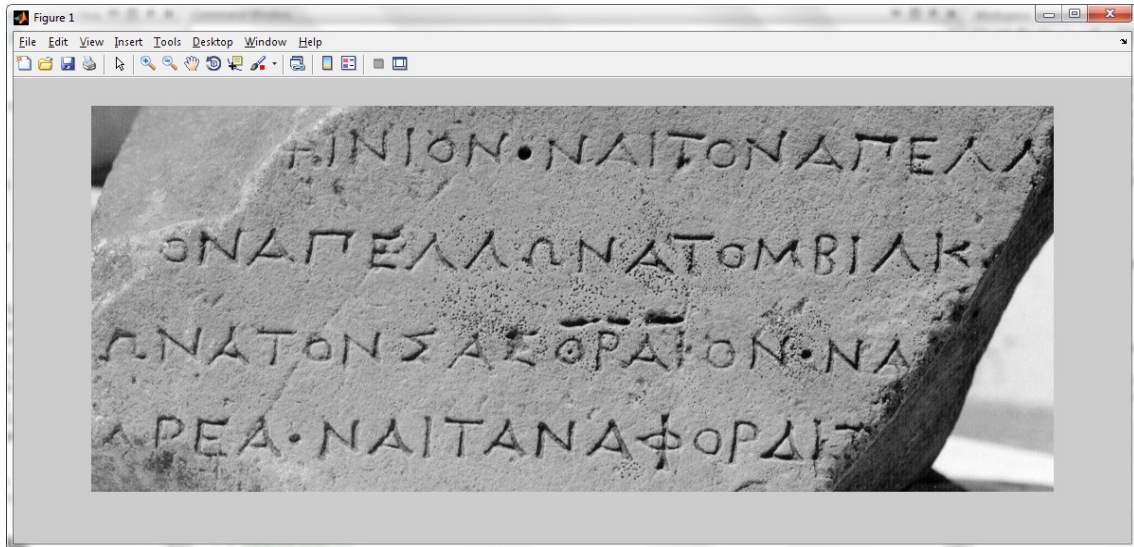
και αυτή του άσπρου είναι:

$$H_W^{(s)} = - \sum_{i=s+1}^{n-1} \frac{p_i}{1 - P_B} \log_2 \left(\frac{p_i}{1 - P_B} \right)$$

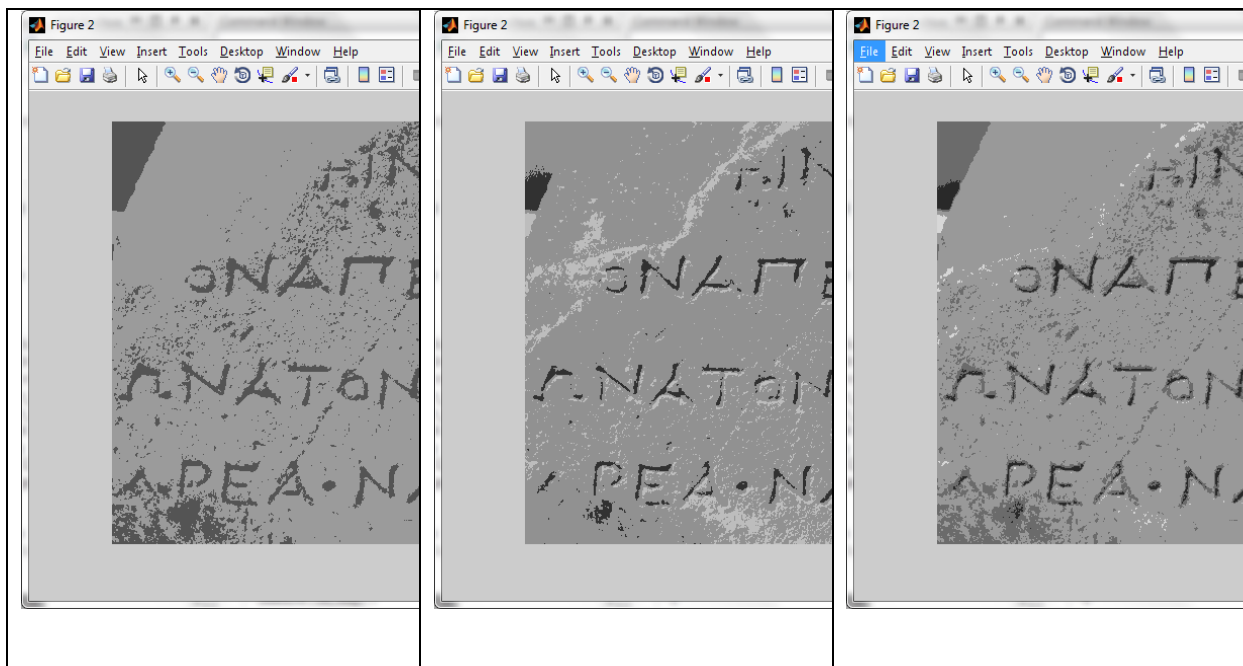
Η ολική εντροπία της εικόνας συνεπώς ορίζεται ως:

$$H_T^{(s)} = H_B^{(s)} + H_W^{(s)}.$$

Για κατώφλι s επιλέγεται αυτό το οποίο μεγιστοποιεί το $H_T^{(s)}$. Στις εικόνες που ακολουθούν βλέπουμε εφαρμογές της μεθόδου του Karur στο MATLAB. Πρώτη βλέπουμε την αρχική εικόνα και στη συνέχεια τρεις εφαρμογές με ένα, δύο και τρία κατώφλια που δίνουν δύο, τρεις και τέσσερις κλάσεις του γκρι αντίστοιχα.



Εικόνα 8. Gray scale εικόνα στην οποία θα γίνει εφαρμογή της μεθόδου του Karur.



Εικόνα 9. Τρεις εφαρμογές της μεθόδου του Karur, όπου διακρίνουμε τις διαφορετικές κλάσεις του γκρι.

Μια ακόμα μέθοδος καθολικής κατωφλίωσης που αξίζει να σημειωθεί είναι αυτή η οποία χρησιμοποιεί τα σημεία ακμών, τα οποία 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 προσφέρει καλύτερη διαχωριστικότητα κειμένου-υποβάθρου σε σχέση με το ιστόγραμμα της συνολικής εικόνας. Η ιδέα αυτή χρησιμοποιήθηκε από τον **Weszka**¹² 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:

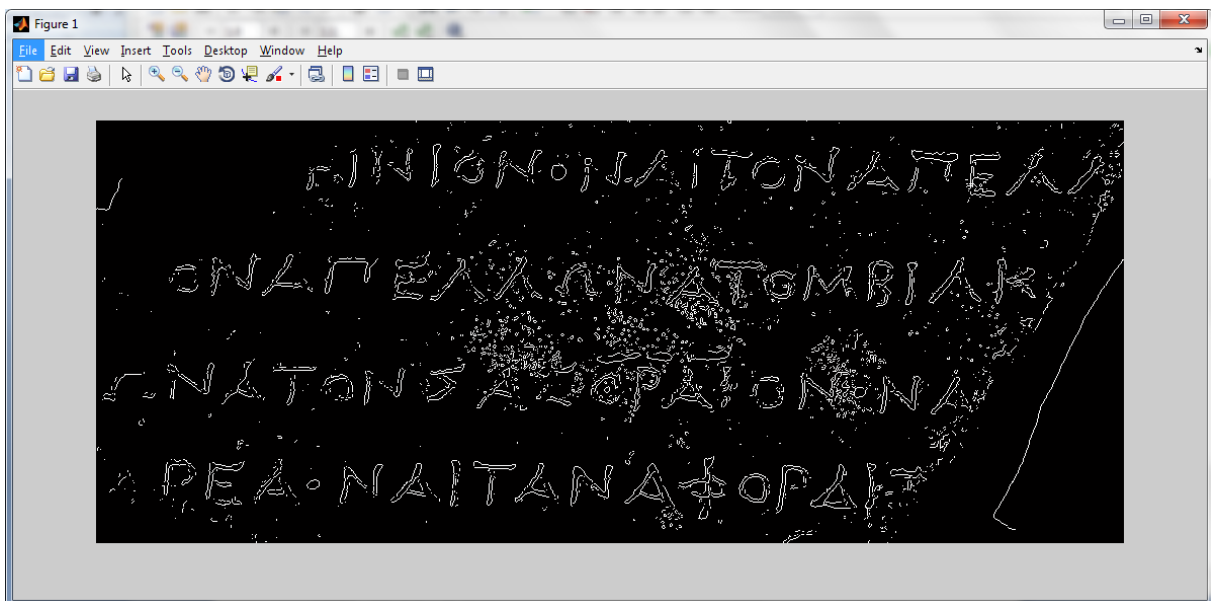
$$\begin{matrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{matrix}$$

Στη συνέχεια προσδιορίζεται το ιστόγραμμα της αρχικής εικόνας λαμβάνοντας υπόψη μόνο τα σημεία που έχουν υψηλές Laplacian τιμές. Για τον εντοπισμό των δύο κορυφών του ιστογράμματος οι δύο τιμές της φωτεινότητας i_1 και i_2 μπορούν να εντοπιστούν ως εξής:

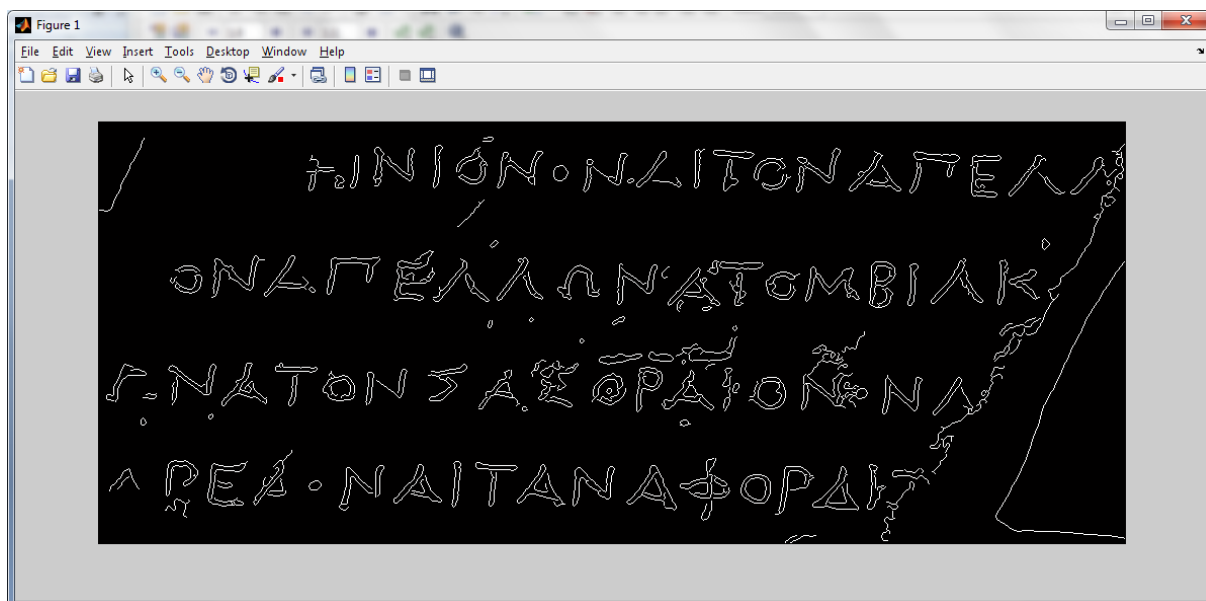
$$\begin{aligned} i_1: \text{Hist}(i) &= \max \text{ για } i = i_1 \\ i_2: \text{Hist}(i) * (i - i_1)^2 &= \max \text{ για } i = i_2 \end{aligned}$$

Με τον αλγόριθμο αυτό διασφαλίζεται ότι η δεύτερη κορυφή του ιστογράμματος απέχει από την πρώτη.

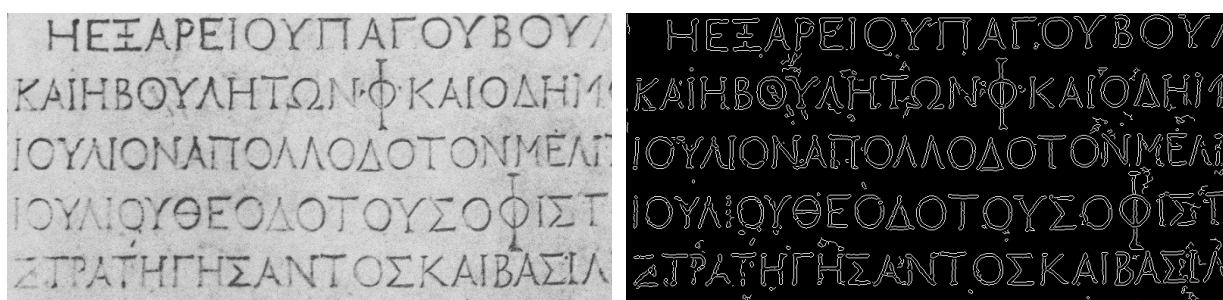
Στην εργαλειοθήκη που χρησιμοποιούμε στο Matlab υπάρχουν αρκετές υλοποιήσεις κατωφλίωσης-κατάτμησης της εικόνας με τον εντοπισμό των σημείων ακμών. Όπως θα δούμε και στα παραδείγματα που ακολουθούν, η τεχνική αυτή δεν είναι πάντα αποδοτική στις εικόνες επιγραφών. Παρόλο που σε κάποιες περιπτώσεις είναι αρκετά ικανοποιητικά τα αποτελέσματα, αυτό δε συμβαίνει κατά βάση. Λόγω των ιδιαιτεροτήτων που παρουσιάζουν οι εικόνες αυτές (ρωγμές στις επιγραφές, ιδιόμορφες επιφάνειες χάραξης, αλλοιωμένα γράμματα κ.ά.), καλά αποτελέσματα προκύπτουν μόνο από εικόνες με πολύ «φιλικά» χαρακτηριστικά, επομένως η συγκεκριμένη μέθοδος δεν συνίσταται.



Εικόνα 10. Έγρευση ακμών με τη μεθοδο του 'Sobel' στο MatLab στην επιγραφή της εικόνας 8.

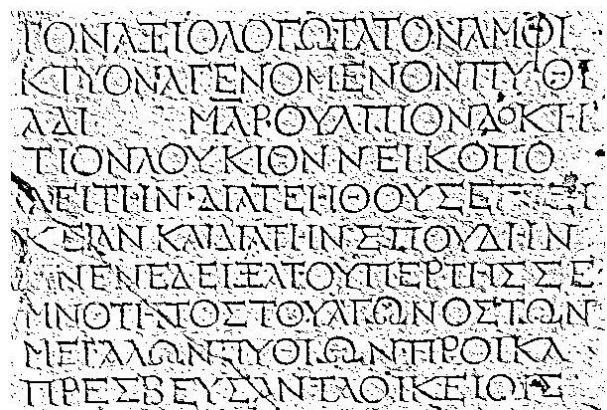
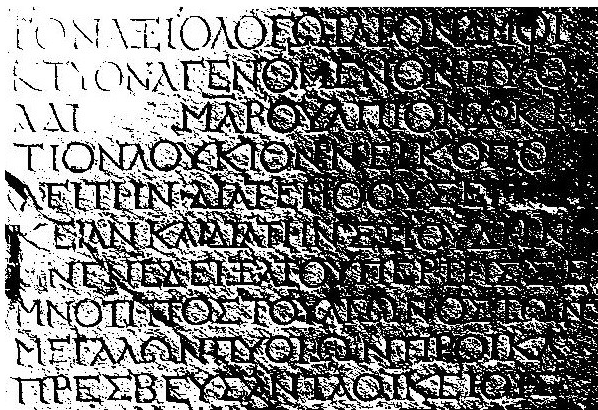
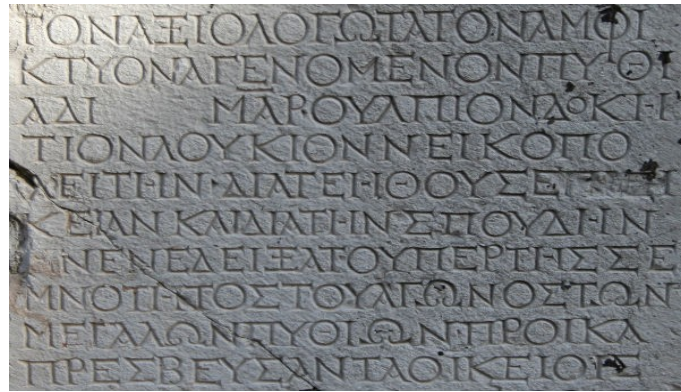


Εικόνα 11. Εύρεση ακμών με τη μέθοδο του 'Canny' στο MatLab στην επιγραφή της εικόνας 8.



Εικόνα 12. Εφαρμογή της μεθόδου του 'Canny' του MatLab με αρκετά καλά αποτελέσματα.

Οι μέθοδοι καθολικής κατωφλίωσης μπορούν να εφαρμοστούν μόνο στις περιπτώσεις όπου υπάρχει σαφής διάκριση των περιοχών του κειμένου από το υπόβαθρο. Αυτό όμως δεν συμβαίνει στις περισσότερες περιπτώσεις των επιγραφών. Ιδίως σε εικόνες επιγραφών που έχουν υποστεί μεγάλο βαθμό αλλοίωσης οι χαρακτήρες, σε εικόνες με ανομοιόμορφη φωτεινότητα υποβάθρου, σε εικόνες που υπάρχουν σκιές λόγω κακής λήψης κ.ά. Στα ιστογράμματα των εικόνων αυτών οι τιμές του γκρι που αντιστοιχούν στην πληροφορία-κείμενο και στο υπόβαθρο, επικαλύπτονται. Για τους λόγους αυτούς αναπτύχθηκαν οι μέθοδοι **προσαρμοσμένης ή τοπικής κατωφλίωσης** (adaptive-local 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 όπου δεν υπάρχει σαφής διάκριση των περιοχών του κειμένου από το υπόβαθρο.



Εικόνα 13. Επάνω: αρχική εικόνα με ανομοιομορφία στη φωτεινότητα. Κάτω αριστερά: καθολική κατωφλίωση. Κάτω δεξιά: τοπική κατωφλίωση.

Στην προηγούμενη εικόνα χρησιμοποιήθηκε αλγόριθμος στο MatLab διαμορφωμένος από τον *Guanglei Xiong*, ο οποίος είναι βασισμένος στη θεωρία των **Chow & Kaneko**¹³. Σύμφωνα με την μέθοδο αυτή, 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. Καταρχάς, 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.

Μέθοδος Niblack

Μια διαφορετική προσέγγιση ακολούθησε ο Niblack¹⁴ με τη μέθοδο του, η οποία προτάθηκε το 1986. Η κεντρική ιδέα είναι η εύρεση ενός τοπικού κατωφλιού για κάθε εικονοστοιχείο της εικόνας, ανάλογα με την τοπική μέση τιμή και την τοπική τυπική απόκλιση. Συγκεκριμένα, το κατώφλι στο εικονοστοιχείο (x, y) υπολογίζεται από τη σχέση:

$$T(x, y) = m(x, y) + k \cdot s(x, y)$$

όπου $m(x, y)$ η μέση τιμή και $s(x, y)$ η τυπική απόκλιση σε μια γειτονιά του εικονοστοιχείου (x, y) . Με την τιμή του συντελεστή k , ρυθμίζουμε το ποσοστό των συνολικών μαύρων εικονοστοιχείων κυρίως στα όρια των αντικειμένων. Έχει βρεθεί ότι η τιμή $k = -0.2$ δίνει αντικείμενα που διαχωρίζονται αρκετά καλά από το φόντο. Η διάσταση του παραθύρου της γειτονιάς πρέπει να είναι αρκετά μικρή ώστε να μη χάνονται τοπικές λεπτομέρειες αλλά ταυτόχρονα και αρκετά μεγάλη για να αφαιρείται τυχόν θόρυβος. Τα αποτελέσματα γενικά δεν είναι ιδιαίτερα ευαίσθητα στο μέγεθος του παραθύρου αρκεί το παράθυρο να περιλαμβάνει τουλάχιστον 1-2 χαρακτήρες.

Μέθοδος Sauvola

Τη λύση σε αυτό το πρόβλημα δίνει ο αλγόριθμος του Sauvola¹⁵ που, όπως και ο αλγόριθμος του Niblack, βασίζεται στην εκτίμηση της τοπικής μέσης τιμής και της τοπικής τυπικής απόκλισης, όμως εισάγει στην υπόθεση τα 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:

$$T(x, y) = m(x, y) \left[1 + k \left(1 - \frac{s(x, y)}{R} \right) \right]$$

όπου R είναι το δυναμικό εύρος της τυπικής απόκλισης, συνήθως ίσο με 128 και k είναι σταθερά με τιμή 0,5 (κατά άλλους $k = 0,1$).

Μέθοδος Gatos

Μια μέθοδος προσαρμοσμένης κατωφλίωσης που θα μπορούσε να δώσει πολύ καλά αποτελέσματα στην περίπτωση μας, είναι αυτή που βασίζεται στον εντοπισμό του υποβάθρου¹⁶ αναπτύχθηκε για τον εντοπισμό του κειμένου σε χαμηλής ποιότητας ιστορικά έγγραφα. Αρχικά, χρησιμοποιείται η προσέγγιση του Niblack ώστε να γίνει μία πρώτη προσέγγιση των pixels που ανήκουν στο κείμενο-φόντο. Συνήθως, τα πραγματικά pixels του κειμένου είναι υποσύνολο του αποτελέσματος του Niblack, αφού ο Niblack εισάγει θόρυβο εντοπίζοντας ταυτόχρονα όλες τις περιοχές κειμένου. Στη συνέχεια, υπολογίζεται η επιφάνεια του υποβάθρου της εικόνας με τη χρήση της παρεμβολής για τον εντοπισμό των φωτεινοτήτων των pixels που έχουν εντοπιστεί κατά το πρώτο στάδιο. Στο τελικό στάδιο, πραγματοποιείται η τελική κατωφλίωση συνδυάζοντας την εντοπισμένη επιφάνεια του υποβάθρου της εικόνας με την αρχική εικόνα. Οι περιοχές κειμένου εντοπίζονται αν η απόσταση της φωτεινότητας της αρχικής εικόνας από την αντίστοιχη φωτεινότητα της επιφάνειας του υποβάθρου ξεπερνάει κάποιο κατώφλι. Το κατώφλι αυτό προσαρμόζεται ανάλογα με τις τιμές της επιφάνειας του υποβάθρου ώστε να γίνεται εντοπισμός του κειμένου ακόμα και σε αρκετά σκοτεινές περιοχές. Αναλυτικά τα διάφορα στάδια έχουν ως εξής:

- **Αρχική εκτίμηση του κειμένου:** Επεξεργαζόμαστε την αρχική εικόνα $I(x,y)$ ώστε να παραγάγουμε τη δυαδική εικόνα $N(x,y)$ σύμφωνα με την προσέγγιση του Niblack.

- **Προσέγγιση της επιφάνειας του υποβάθρου:** Στο στάδιο αυτό υπολογίζουμε προσεγγιστικά την επιφάνεια του υποβάθρου $B(x,y)$ της εικόνας. Τα σημεία της αρχικής εικόνας $I(x,y)$ ανήκουν στην επιφάνεια του υποβάθρου $B(x,y)$ μόνον αν τα αντίστοιχα σημεία της εικόνας $N(x,y)$ της αρχικής εκτίμησης του κειμένου έχουν μηδενικές τιμές. Τα υπόλοιπα σημεία της επιφάνειας $B(x,y)$ υπολογίζονται με παρεμβολή από τα γειτονικά τους. Ο τύπος υπολογισμού της $B(x,y)$ έχει ως εξής:
- Το παράθυρο $dx \times dy$ που χρησιμοποιείται για την παρεμβολή ορίζεται ώστε να καλύπτει τουλάχιστον δύο χαρακτήρες του εγγράφου.

$$B(x, y) = \begin{cases} I(x, y) & \text{if } N(x, y) = 0 \\ \frac{\sum_{i=x-dx}^{x+dx} \sum_{j=y-dy}^{y+dy} (I(i, j)(1 - N(i, j)))}{\sum_{i=x-dx}^{x+dx} \sum_{j=y-dy}^{y+dy} (1 - N(i, j))} & \text{if } N(x, y) = 1 \end{cases}$$

- Το παράθυρο $dx \times dy$ που χρησιμοποιείται για την παρεμβολή ορίζεται ώστε να καλύπτει τουλάχιστον δύο χαρακτήρες του εγγράφου.
- **Τελική κατοφλίωση:** Στο στάδιο αυτό πραγματοποιούμε την τελική κατοφλίωση συνδυάζοντας την υπολογισμένη επιφάνεια B με την αρχική εικόνα I . Οι περιοχές κειμένου εντοπίζονται αν η απόσταση της αρχικής εικόνας από την επιφάνεια του υποβάθρου είναι μεγαλύτερη από ένα κατώφλι d . Προτείνεται η τιμή του κατωφλιού d να αλλάζει ανάλογα με την αντίστοιχη τιμή της επιφάνειας B , έτσι ώστε να διασώζεται η πληροφορία του κειμένου ακόμα και σε περιοχές με πολύ σκούρο υπόβαθρο. Για τον λόγο αυτό προτείνεται η τιμή του κατωφλιού d να παίρνει μικρότερες τιμές στις πιο σκούρες περιοχές. Η τελική δυαδική εικόνα T δίδεται από τον παρακάτω τύπο:

$$T(x, y) = \begin{cases} 1, & \text{αν } B(x, y) - I(x, y) > d \\ 0, & \text{αλλιού} \end{cases}$$

Η μέση απόσταση δ ανάμεσα στις περιοχές κειμένου και φόντου μπορεί να υπολογιστεί από τον παρακάτω τύπο:

$$\delta = \frac{\sum_x \sum_y (B(x, y) - I(x, y))}{\sum_x \sum_y N(x, y)}$$

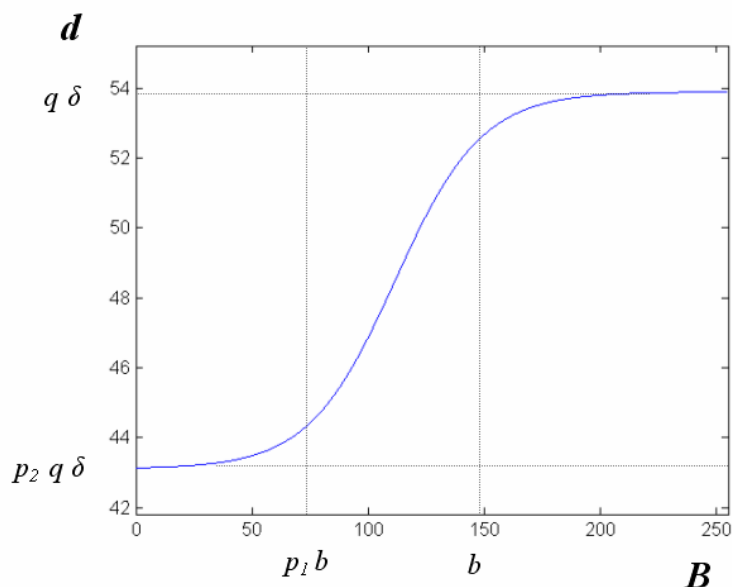
Για τις συνηθισμένες περιπτώσεις εικόνων εγγράφων με ομοιόμορφη φωτεινότητα υποβάθρου, το ελάχιστο κατώφλι d μεταξύ των pixels κειμένου και υποβάθρου μπορεί να οριστεί με επιτυχία σαν $q \cdot \delta$, όπου q μεταβλητή κοντά στο 0.8 η οποία βοηθάει τη διάσωση του συνολικού σώματος των χαρακτήρων ώστε να έχουμε βέλτιστα OCR αποτελέσματα. Στην περίπτωση των παλιών εγγράφων με χαμηλή ποιότητα και μη ομοιόμορφη κατανομή φωτεινότητας θέλουμε να έχουμε μικρότερη τιμή για το κατώφλι d για τις πιο σκούρες περιοχές υποβάθρου, αφού είναι συνηθισμένη περίπτωση σε τέτοιες εικόνες να έχουμε κείμενο σε σκούρες περιοχές υποβάθρου όπου η διαφορά στη φωτεινότητα μεταξύ κειμένου και υποβάθρου είναι μικρή. Για να το πετύχουμε αυτό, πρώτα υπολογίζουμε την μέση τιμή b των σημείων της επιφάνειας του υποβάθρου B που αντιστοιχούν στις περιοχές κειμένου της εικόνας N :

$$b = \frac{\sum_x \sum_y (B(x, y)(1 - N(x, y)))}{\sum_x \sum_y (1 - N(x, y))}$$

Αυτό που θέλουμε να πετύχουμε είναι το κατώφλι να είναι περίπου ίσο με $q \cdot \delta$, όταν η τιμή του υποβάθρου είναι μεγάλη (περίπου μεγαλύτερη από την μέση τιμή υποβάθρου b), και περίπου ίσο με $p_2 \cdot q \cdot \delta$, όταν η τιμή του υποβάθρου είναι μικρή (περίπου μικρότερη από μικρότερο από $p_1 \cdot b$) με $p_1, p_2 \in [0, 1]$. Για να εξομοιώσουμε την επιθυμητή συμπεριφορά του κατωφλιού, χρησιμοποιούμε την παρακάτω λογαριθμική σιγμοειδή συνάρτηση η οποία παρουσιάζει την επιθυμητή οριακή συμπεριφορά για μεγάλες και μικρές τιμές του υποβάθρου:

$$d(B(x, y)) = q \delta \left(\frac{(1 - p_2)}{1 + \exp\left(\frac{-4 B(x, y)}{b(1 - p_1)} + \frac{2(1 + p_1)}{(1 - p_1)}\right)} + p_2 \right)$$

Μετά από πειραματικά αποτελέσματα, για την περίπτωση χαμηλής ποιότητας ιστορικών εγγράφων, προτείνονται οι παρακάτω παράμετροι: $q = 0.6$, $p_1 = 0.5$, $p_2 = 0.8$.



Εικόνα 14. Η συνάρτηση $d(B(x,y))$.

Μέθοδος Εξαγωγής Προσκηνίου¹⁷.

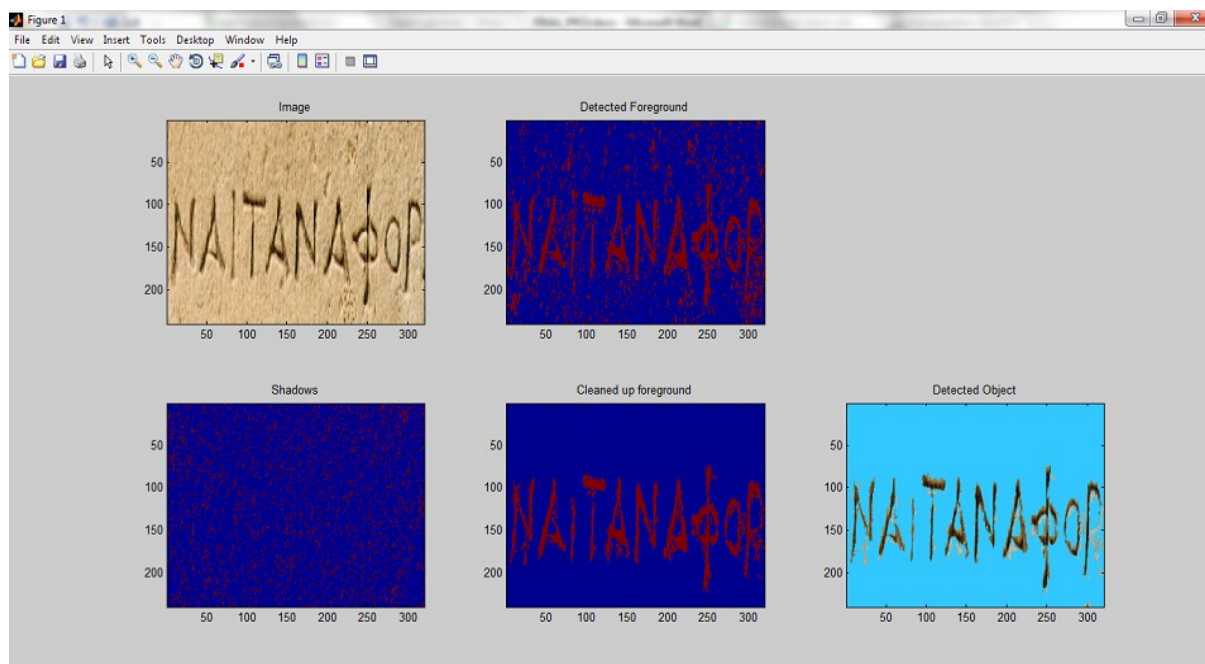
Αυτή η μέθοδος, η οποία χρησιμοποιείται αρκετά σε συστήματα ανίχνευσης κίνησης, μπορεί με κάποιες τροποποιήσεις να αποφέρει καλά αποτελέσματα, σε περιορισμένο όμως είδος εικόνων επιγραφών. Παρότι είναι λιγότερο πιθανό να χρησιμοποιηθεί σε σύστημα σαν και αυτό με το οποίο η παρούσα εργασία ασχολείται, λόγω της πολυπλοκότητάς του και του όγκου των δεδομένων που χρειάζεται, αξίζει να αναφερθεί διότι ως ιδέα προσφέρει αρκετές δυνατότητες τροποποίησης και εξέλιξης.

Συνοπτικά η λειτουργία της είναι η εξής: Ο αλγόριθμος λαμβάνει σαν δεδομένα εισαγωγής εικόνες του υποβάθρου με σκοπό τη μοντελοποίηση του. Το μοντέλο του κάθε pixel μπορεί να εκπροσωπείται από και μέχρι 5 χρώματα., τα οποία καθορίζονται στην περίπτωση μας από τις εικόνες που θα λάβει αρχικά ο αλγόριθμος. Κατά τη διαδικασία της εκτέλεσης της κατωφλίωσης, ένα pixel θεωρείται ότι είναι μέρος του φόντου (υποβάθρου), εάν είναι εντός T απόστασης από τα χρώματα του φόντου μέσα στο χρωματικό χώρο RGB. Με βάση πειραματικές παρατηρήσεις έχει επιλεγεί ως προεπιλεγμένη τιμή του $T=20$. Αφού όπως αναφέραμε και σε προηγούμενη παράγραφο, οι ανομοιομορφίες στον φωτισμό των εικόνων και οι σκιές είναι φαινόμενο που παρατηρείται αρκετές φορές, τα pixels που αρχικά έχουν εξαχθεί σαν pixels προσκηνίου υποβάλλονται σε περαιτέρω δοκιμή. Η διαφορά χρώματος/φωτεινότητας D ορίζεται ως εξής:

$$D = k \times D_{gb} + |\log(V_{BG}/V_{curr})|$$

Όπου D_{gb} είναι η Ευκλείδεια απόσταση μεταξύ του τρέχοντος pixel και του πιο κοινού pixel του υποβάθρου, χρωματικά κανονικοποιημένου, V_{BG} είναι η φωτεινότητα στο φόντο και V_{curr} είναι η φωτεινότητα του τρέχοντος pixel. Η σταθερά k καθορίζεται από τον δημιουργό του αλγορίθμου στην τιμή 18 πειραματικά. Για ένα συγκεκριμένο pixel του προσκηνίου, εάν $D < 0.5$, τότε αυτό το pixel επισημαίνεται σαν αλλαγή φωτισμού (σκιά/υψηλή φωτεινότητα) και απορρίπτεται. Στην εφαρμογή του αλγορίθμου στο MatLab που θα παρουσιάσουμε στην πιο κάτω εικόνα, ο δημιουργός της

χρησιμοποιεί κάποιες τεχνικές βελτίωσης ποιότητας, στις οποίες θα αναφερθούμε στην υποενότητα 3.1.2.



3.1.2. Βελτίωση Ποιότητας

Οι δυαδικές εικόνες των εγγράφων συνήθως χρειάζονται ένα στάδιο προεπεξεργασίας για τη βελτίωση της ποιότητάς τους. Στο στάδιο αυτό αντιμετωπίζονται διάφορα προβλήματα στην ποιότητα της εικόνας λόγω χαμηλής ποιότητας αρχικού εγγράφου ή λόγω κακής ψηφιοποίησης ή λόγω παλαιότητας του εγγράφου κλπ. Το στάδιο αυτό είναι ιδιαίτερα σημαντικό γιατί βοηθάει στην επιτυχία των επόμενων σταδίων επεξεργασίας του εγγράφου που αφορούν την κατάτμηση και αναγνώριση των χαρακτήρων. Τα βασικά προβλήματα στην ποιότητα της εικόνας που πρέπει να αντιμετωπιστούν είναι τα ακόλουθα:

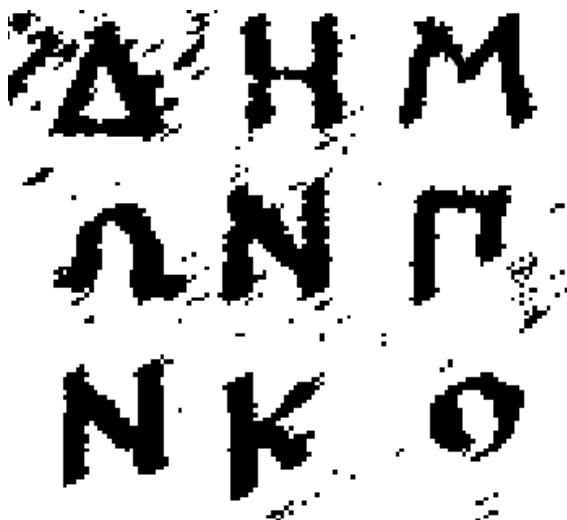
- **Θόρυβος μορφής αλατοπίπερου ή κρουστικός θόρυβος (salt and pepper noise or impulse noise):** Είναι η παρουσία μεμονωμένων μαύρων pixels σε λευκές περιοχές ή μεμονωμένων λευκών pixels σε μαύρες περιοχές. Ο θόρυβος αυτός μπορεί να προέλθει από νεκρά εικονοστοιχεία (dead pixels), από σφάλματα κατά την ψηφιοποίηση ή δυαδική μετατροπή, από bit errors κατά τη μετάδοση κ.ά. (Εικόνα 15α).
- **Θόρυβος αποτελούμενος από τμήματα εικόνας μικρού μεγέθους:** Τμήματα της εικόνας με μικρό εμβαδό τα οποία έχουν προκύψει λόγω θορύβου ή λόγω του επιπέδου περιγραφής του εγγράφου κατά την ψηφιοποίηση (Εικόνα 15β).
- **Οπές στο σώμα των χαρακτήρων:** Το σώμα των χαρακτήρων αντί να είναι συμπαγές, παρουσιάζει οπές. Αυτό μπορεί να δημιουργήσει ιδιαίτερο πρόβλημα στην αναπαράσταση και αναγνώριση των χαρακτήρων (Εικόνα 15γ).
- **Ασυνέχειες στο σώμα των χαρακτήρων:** Δεν υπάρχει συνεκτικότητα των χαρακτήρων. Σπασίματα ή κενά εμφανίζονται στο σώμα των χαρακτήρων. Ιδιαίτερα σημαντικό πρόβλημα για το στάδιο της κατάτμησης των χαρακτήρων (Εικόνα 15δ).

- **Ενωμένα γράμματα:** Λόγω κακής ψηφιοποίησης ή ύπαρξης θορύβου ανάμεσα στους χαρακτήρες, γειτονικοί χαρακτήρες ενώνονται μεταξύ τους (Εικόνα 15ε).
- **Μη ομαλή περίμετρος των χαρακτήρων:** Η μη ομαλή περίμετρος των χαρακτήρων είναι πρόβλημα που εισάγεται κατά την ψηφιοποίηση και μπορεί να έχει επίπτωση στο στάδιο της αναγνώρισης των χαρακτήρων (Εικόνα 15ζ).
- **Επικάλυψη συμβόλων του υποβάθρου με το κείμενο:** Σύμβολα του υποβάθρου (γραμμές ή περιοδικά εμφανιζόμενα σύμβολα) επικαλύπτουν τις περιοχές κειμένου.

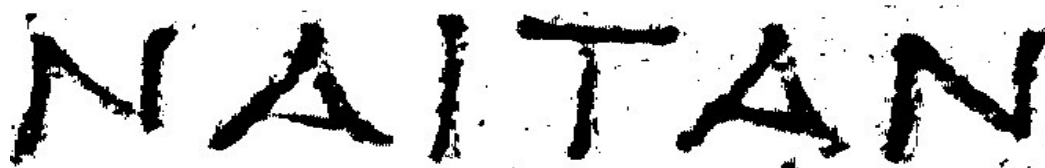
Για την αντιμετώπιση των παραπάνω προβλημάτων χρησιμοποιούνται μέθοδοι που βασίζονται κυρίως στη χρήση масκών, στη μαθηματική μορφολογία, σε φίλτρα συρρίκνωσης και διόγκωσης καθώς και σε ειδικούς αλγορίθμους που εντοπίζουν και αφαιρούν τα μαύρα περιθώρια στα όρια της εικόνας.



(α)



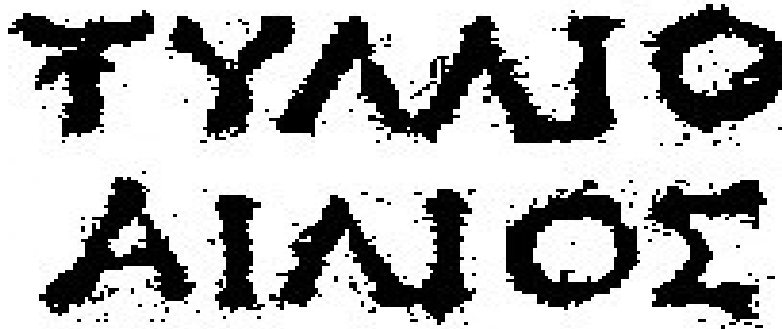
(β)



(γ)



(δ)



(ε)



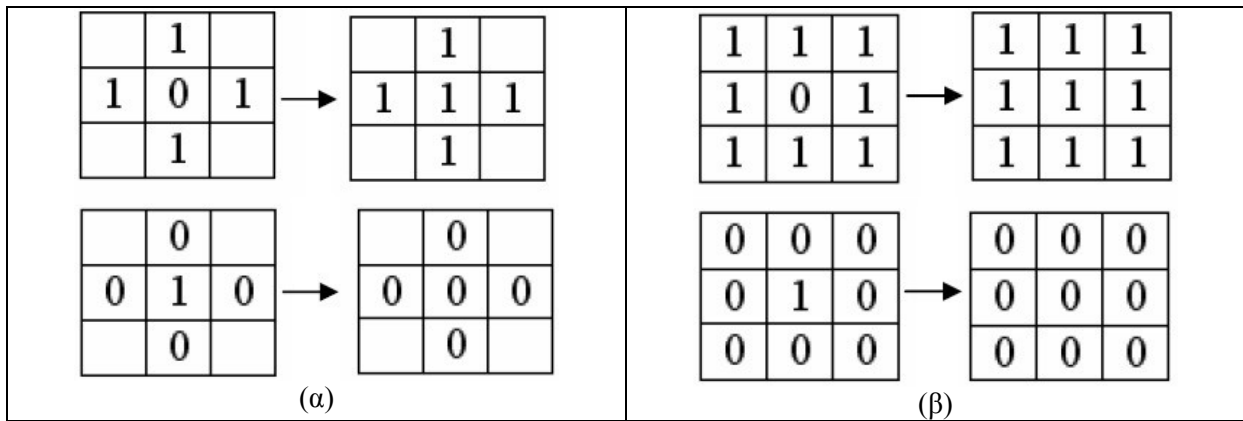
(ζ)

Εικόνα 15. Προβλήματα ποιότητας σε δυαδικές εικόνες. (α) Θόρυβος με μορφή αλατοπίπερου. (β) Θόρυβος αποτελούμενος από μικρά τμήματα εικόνας. (γ) Οπές στο σώμα των χαρακτήρων. (δ) Ασυνέχειες στο σώμα των χαρακτήρων. (ε) Ενωμένοι χαρακτήρες. (ζ) Μη ομαλή περίμετρος των χαρακτήρων.

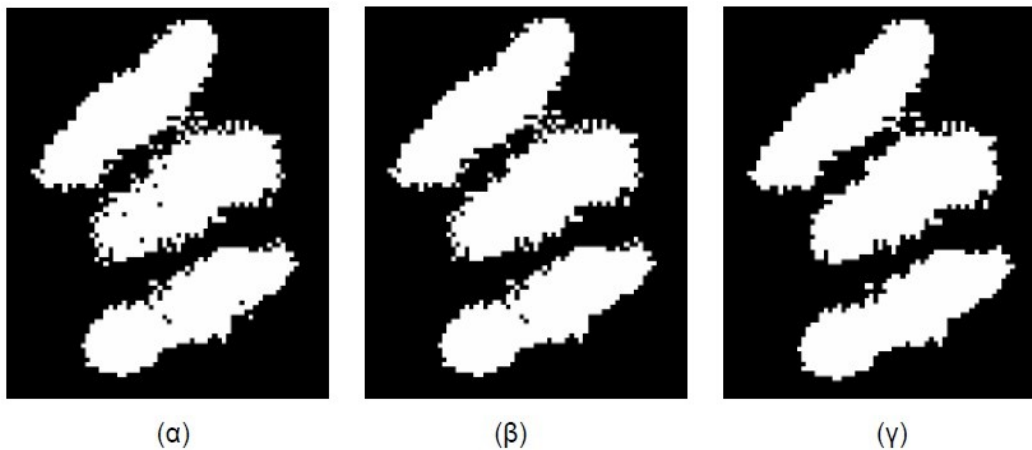
Αντιμετώπιση θορύβου μορφής αλατοπίπερου.

Χρήση Μασκών

Για την αφαίρεση τέτοιου είδους θορύβου, χρησιμοποιούνται διάφορες μάσκες, οι οποίες εφαρμόζονται σε κάθε σημείο της επιφάνειας της εικόνας. Οι πιο απλές από αυτές που χρησιμοποιούνται σε αυτή τη περίπτωση, είναι οι 3x3 μάσκες βάσει των 8 ή των 4 γειτονικών pixels (Εικόνα 16). Αν η μάσκα ταιριάζει στην εικόνα, τότε η εικόνα μετατρέπεται όπως στην πιο κάτω Εικόνα 17, όπου φαίνεται το αποτέλεσμα της χρήσης τέτοιων масκών.

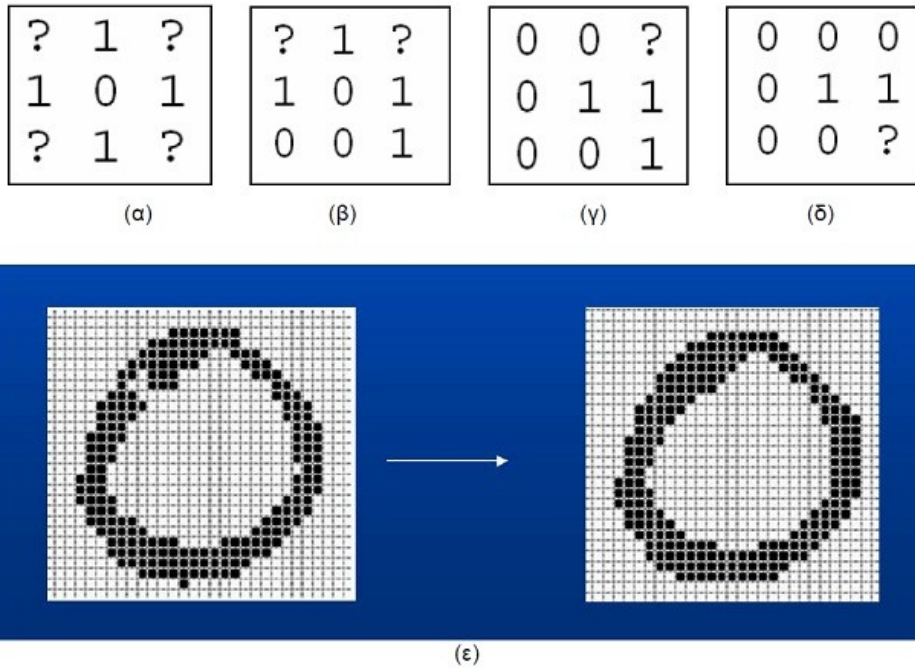


Εικόνα 16. Χρήση μασκών για την αφαίρεση θορύβου μορφής αλατοπίπερου. (α) Βάση των 4 γειτονικών pixels. (β) Βάσει των 8 γειτονικών pixels.



Εικόνα 17. Αποτέλεσμα της χρήσης μασκών για την αφαίρεση θορύβου μορφής αλατοπίπερου (salt & pepper noise). (α) Αρχική εικόνα. (β) Αφαίρεση θορύβου χρησιμοποιώντας τη μάσκα των 8 γειτονικών pixels. (γ) Αφαίρεση θορύβου χρησιμοποιώντας τη μάσκα των 4 γειτονικών pixels.

Οι μάσκες χρησιμοποιούνται και για την εξομάλυνση της επιφάνειας και βελτίωση της ποιότητας των χαρακτήρων. Σύμφωνα με την προσέγγιση του Suen¹⁸, αν η εικόνα ταιριάζει με τη μάσκα της Εικόνας 18α, τότε το κεντρικό pixel μετατρέπεται σε pixel εικόνας. Ίδια μετατροπή πραγματοποιείται όταν η εικόνα ταιριάζει με την μάσκα της Εικόνας 18β ή τις μάσκες που προκύπτουν με στροφή κατά 90°, 180° και 270°. Αντίστοιχα, διαγραφή του κεντρικού pixel πραγματοποιείται όταν ταιριάζει στην εικόνα η μάσκα της Εικόνας 18γ ή η μάσκα της Εικόνας 18δ ή οποιαδήποτε από τις έξι μάσκες προκύπτουν με στροφή κατά 90°, 180° και 270° των δύο αυτών μασκών. Στην Εικόνα 18ε δίνεται ένα παράδειγμα εξομάλυνσης της επιφάνειας και βελτίωσης της ποιότητας των χαρακτήρων με χρήση μασκών.



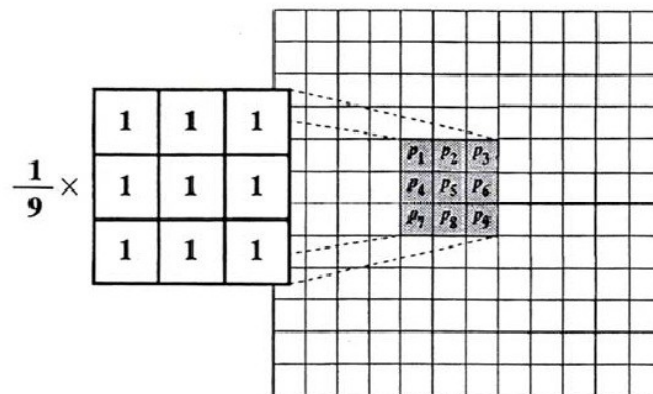
Εικόνα 18. Εξομάλυνση της επιφάνειας και βελτίωση της ποιότητας των χαρακτήρων με χρήση μασκών. (α)-(δ) Μάσκες που χρησιμοποιούνται. (ε) Τελικό αποτέλεσμα.

Δύο τεχνικές ακόμα, που μπορούν να βοηθήσουν στην απαλλαγή από τον θόρυβο της μορφής αλατοπίπερου, είναι η χρήση φίλτρου μέσου όρου (average) και φίλτρου μέσου (median).

Φίλτρο μέσου όρου ή μέσης τιμής

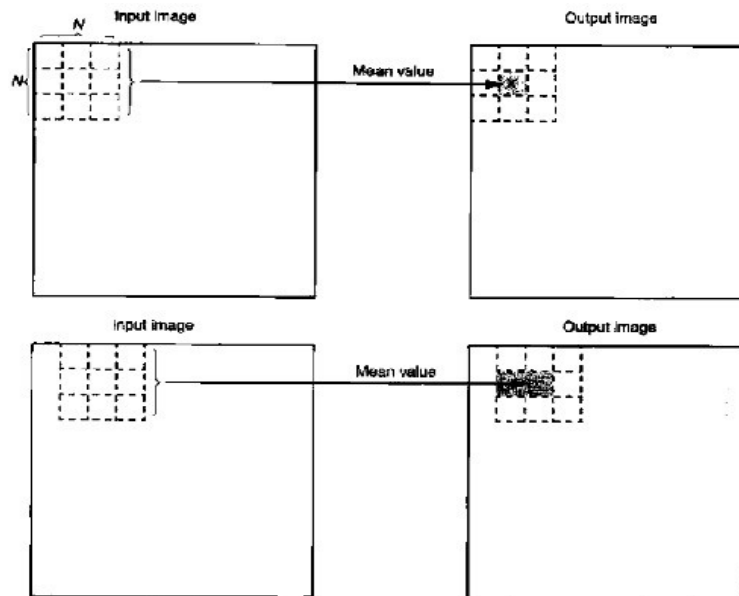
(<http://students.ceid.upatras.gr/~mprokala/techarticles/DigitalImageProcessing/imageRestoration.htm>)

Το φίλτρο της μέσης τιμής (*mean filter-average filter*) είναι μια απλή τεχνική εξάλειψης και εξομάλυνσης του θορύβου από ψηφιακές εικόνες αποχρώσεων του γκρι. Σύμφωνα με την τεχνική αυτή από την αρχική εικόνα παράγεται μία νέα εικόνα ίδιων διαστάσεων όπου κάθε εικονοστοιχείο της έχει φωτεινότητα τη μέση τιμή των τιμών φωτεινότητας μιας γειτονιάς του αντίστοιχου εικονοστοιχείου της αρχικής εικόνας. Είναι το απλούστερο δισδιάστατο χαμηλοπερατό (lowpass) FIR (Finite Impulse Response) φίλτρο.



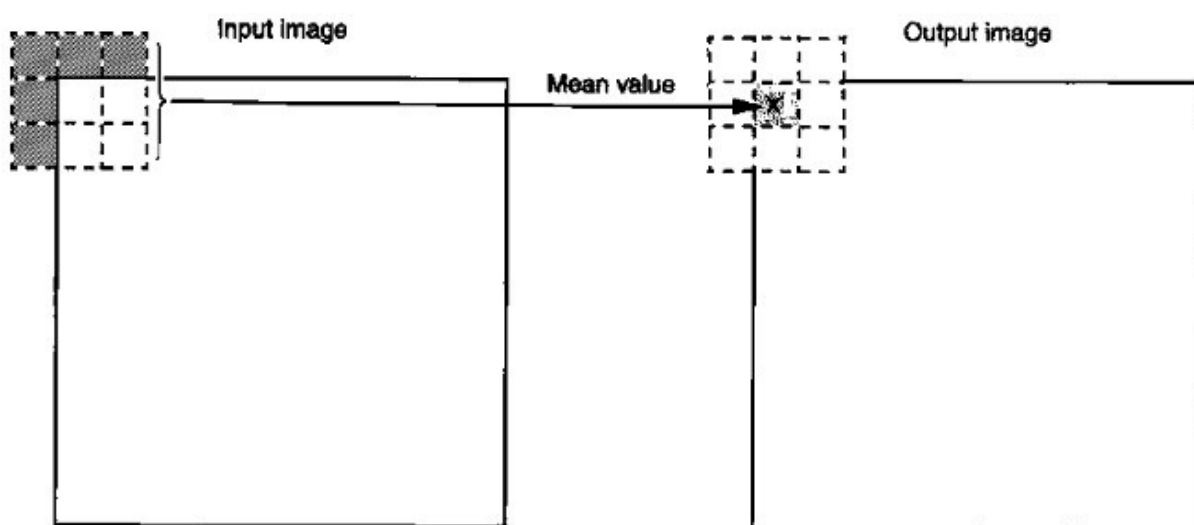
Εικόνα 19. Μάσκα φίλτρου μέσης τιμής

Η νέα εικόνα είναι το αποτέλεσμα του φιλτραρίσματος της αρχικής εικόνας με τη επάνω μάσκα. Είναι δυνατόν να χρησιμοποιηθούν και άλλες μάσκες για την εξομάλυνση μιας εικόνας. Μάσκες 5×5 ή μεγαλύτερες εξομαλύνουν ακόμα περισσότερο την εικόνα.



Εικόνα 20. Λειτουργία φίλτρου μέσης τιμής

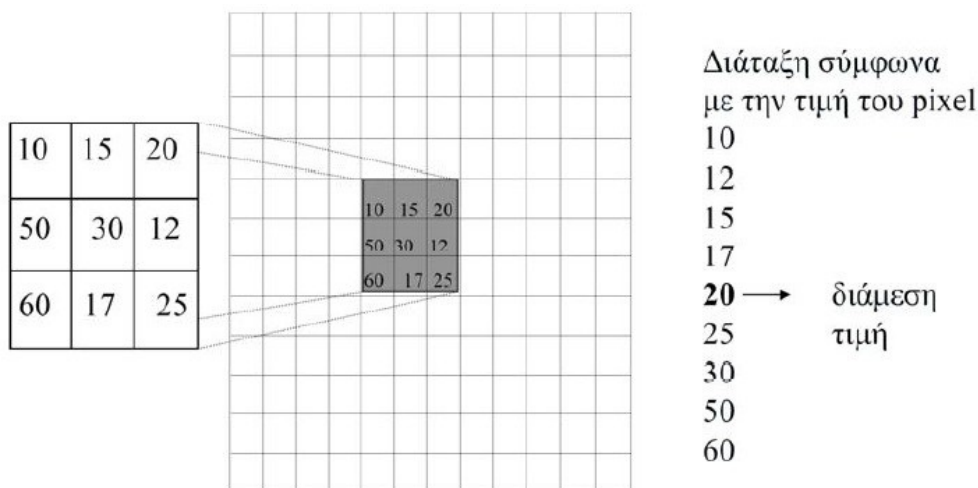
Για να υπολογιστούν οι τιμές των pixel που βρίσκονται στα άκρα της νέας εικόνας, η μάσκα τοποθετείται έτσι ώστε κάποια στοιχεία της να βρίσκονται έξω από τα όρια της αρχικής εικόνας. Σ' αυτή την περίπτωση θεωρείται αυθαίρετα ότι τα στοιχεία της αρχικής εικόνας που δεν υπάρχουν έχουν κάποια συγκεκριμένη τιμή (padding). Συνήθως η τιμή αυτή είναι το μηδέν (zero padding) ή η κοντινότερη τιμή της αρχικής εικόνας.



Εικόνα 21. Σημεία στα άκρα της εικόνας.

Φίλτρο ενδιάμεσης τιμής (Median filter)

Πέρα από το φίλτρο μέσου όρου, για την αποκατάσταση των εικόνων μπορεί να χρησιμοποιηθεί και κάποιο μη γραμμικό χαμηλοπερατό φίλτρο, το φίλτρο μεσαίου (median), με στόχο την εξάλειψη του θορύβου αλλά παράλληλα τη διατήρηση των λεπτομερειών και των περιγραμμάτων των εικόνων. Το φίλτρο μεσαίου βασίζεται στη διάταξη των δεδομένων. Αν έχουμε n παρατηρήσεις (n περιττός) και τις διατάξουμε σε αύξουσα σειρά από το μικρότερο στο μεγαλύτερο, τότε ως μεσαίος ορίζεται η παρατήρηση εκείνη που βρίσκεται στο μέσο της διάταξης, όχι η μεσαία τιμή ή ο μέσος όρος. Για παράδειγμα αν έχουμε τους αριθμούς 0,0,0,0,0,6,8,12,14 (είναι ήδη διαταγμένοι) τότε ο μεσαίος είναι το 0.



Εικόνα 22. Λειτουργία μέσου φίλτρου (median filter)

Πιο φορμαλιστικά, αν θεωρήσουμε n στοιχεία x_i , $i = 1, \dots, n$ (n περιττός) διατεταγμένα από το μικρότερο στο μεγαλύτερο, τότε το στοιχείο που βρίσκεται στη μεσαία θέση της διάταξης ονομάζεται μεσαίος (συμβολίζεται με $med(x_i)$) και είναι το στοιχείο εκείνο που ελαχιστοποιεί την L_1 νόρμα:

$$\sum_{i=1}^n |x_i - med| \rightarrow \min$$

Από την παραπάνω σχέση, ο μεσαίος είναι ο εκτιμητής της μέγιστης πιθανοφάνειας (maximum likelihood) της θέσης, αν τα δεδομένα ακολουθούν κάποια κατανομή όπως η Laplace. Ένα δισδιάστατο φίλτρο μεσαίου έχει τον ακόλουθο ορισμό:

$$y(i, j) = med\{x(i + r, j + s) : (r, s) \in A, (i, j) \in Z^2\}$$

όπου με $Z^2 = Z \times Z$ δηλώνεται το επίπεδο της εικόνας και το A (υποσύνολο του Z^2) είναι το παράθυρο του φίλτρου που χρησιμοποιείται.

Για την υλοποίηση του φίλτρου μεσαίου, θεωρείται όπως και στην προηγούμενη περίπτωση ένα παράθυρο μεγέθους $n \times m$ (n, m περιττοί αριθμοί) με κέντρο κάθε φορά ένα pixel της εικόνας. Υπολογίζεται ο μεσαίος των τιμών όλων των pixels του παραθύρου και ο μεσαίος που προκύπτει είναι η νέα τιμή του κεντρικού pixel του παραθύρου. Η διαδικασία αυτή πραγματοποιείται για όλα τα pixels

της εικόνας, εκτός από εκείνα στα άκρα της εικόνας για τα οποία δεν μπορεί να σχηματιστεί το παράθυρο με κέντρο αυτά.

Προφανώς η σχέση που παρουσιάστηκε παραπάνω δεν μπορεί να ισχύσει για τα ακραία pixels της εικόνας. Για να εφαρμοστεί το φίλτρο σε αυτά τα pixels υπάρχουν δύο προσεγγίσεις. Η μία είναι να αποκοπεί κατάλληλα το παράθυρο A του φίλτρου και η άλλη είναι να συμπληρωθεί η εικόνα με επαρκή αριθμό νέων δειγμάτων (pixels στα άκρα της εικόνας) ώστε να μπορεί να εφαρμοστεί το παράθυρο A .

Σε σύγκριση με το προηγούμενο φίλτρο (μέσης τιμής) τα αποτελέσματα είναι πιο ικανοποιητικά. Τούτο συμβαίνει γιατί τα φίλτρα μεσαίου έχουν ορισμένες ενδιαφέρουσες ιδιότητες. Όπως έχει αναφερθεί είναι χαμηλοπερατά φίλτρα άρα έχουν την ιδιότητα να αφαιρούν το θόρυβο που εμφανίζεται κυρίως στις υψηλές συχνότητες. Επίσης, επειδή ο μεσαίος είναι ένας ανθεκτικός εκτιμητής, ένα απλό σφάλμα (όπως π.χ. ένας μοναδιαίος παλμός) δεν επηρεάζει την επίδοση του φίλτρου, ακόμα και αν το μέγεθός του είναι πολύ μεγάλο ή πολύ μικρό. Αυτές οι ιδιότητες ανθεκτικότητας καθιστούν το φίλτρο μεσαίου κατάλληλο για την αφαίρεση κρουστικού θορύβου. Επιπλέον άλλο ένα χαρακτηριστικό του φίλτρου αυτού είναι ότι έχει την τάση να διατηρεί τα περιγράμματα στην εικόνα. Το γεγονός αυτό έχει ως αποτέλεσμα να εμφανίζεται η εικόνα καλύτερη στο μάτι σε σχέση με το φίλτρο μέσου όρου που τείνει να εξομαλύνει τα περιγράμματα.

Σαν αρνητικό αυτής της μεθόδου αξίζει να σημειωθεί ότι όσο το μέγεθος της μάσκας μεγαλώνει τόσο μεγαλύτερη απώλεια πληροφορίας έχουμε στο αποτέλεσμα που τείνει να μοιάσει τεχνητά ζωγραφισμένο.

Μορφολογικές Λειτουργίες

Η *μορφολογία* στην επεξεργασία εικόνας αποτελείται από μία ευρεία σειρά λειτουργιών επεξεργασίας εικόνας η οποία βασίζεται σε σχήματα. Αναφέρεται στη μελέτη της γεωμετρίας, της τοπολογίας και της μορφής των αντικειμένων. Οι μορφολογικές λειτουργίες εφαρμόζουν ένα δομικό στοιχείο (structuring element) σε μια εικόνα εισόδου, δημιουργώντας σαν έξοδο μια εικόνα ίδιου μεγέθους. Σε μια μορφολογική λειτουργία, η τιμή του κάθε pixel της εικόνας εξόδου βασίζεται σε μια σύγκριση του αντίστοιχου pixel της εικόνας εισόδου με τους «γείτονές» του. Επιλέγοντας το μέγεθος και το σχήμα της γειτονιάς, μπορούμε να κατασκευάσουμε μια μορφολογική λειτουργία η οποία είναι ευαίσθητη σε συγκεκριμένα σχήματα στην εικόνα εισόδου.

Οι μορφολογικές λειτουργίες-πράξεις χρησιμοποιούνται κυρίως για τους ακόλουθους σκοπούς:

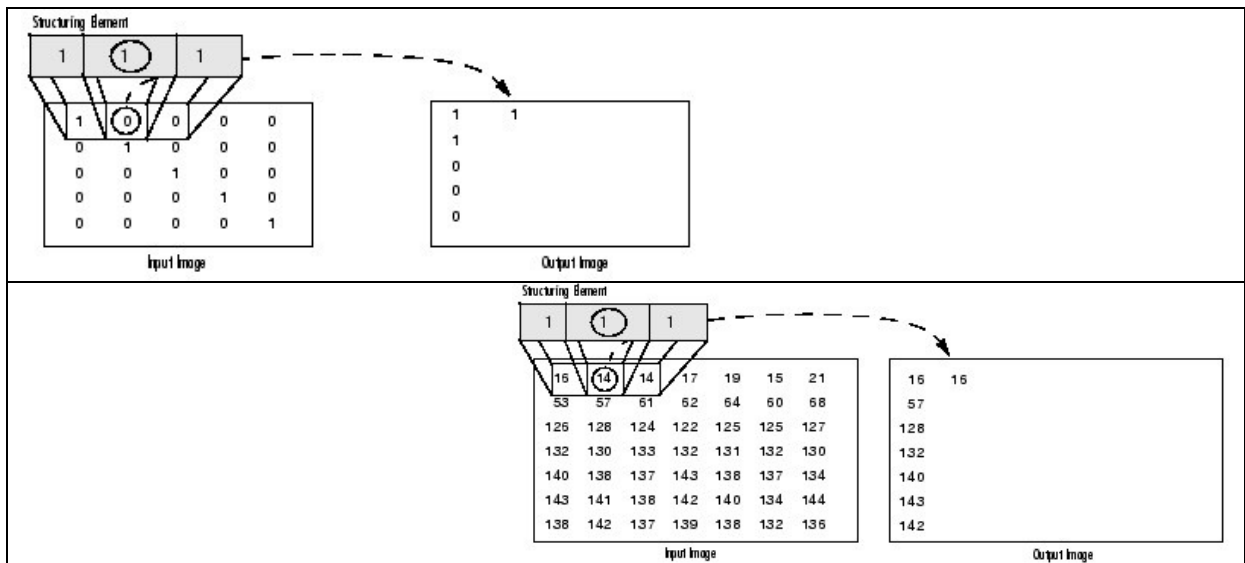
- Προεπεξεργασία της εικόνας και βελτίωση της δομής των αντικειμένων, ουσιαστικά για εργασίες όπως φιλτράρισμα θορύβου, βελτίωση της ποιότητας της εικόνας και απλοποίηση των αντικειμένων, δημιουργία σκελετού αντικειμένου, λέπτυνση – πάχυνση της εικόνας.
- Ποσοτική περιγραφή των αντικειμένων (χαρακτηριστικά εμβαδού, περιμέτρου, προβολών).

Οι πιο βασικές μορφολογικές λειτουργίες είναι το dilation (διαστολή) και το erosion (διάβρωση). Το dilation προσθέτει pixels στα όρια των αντικειμένων σε μια εικόνα, ενώ το erosion αφαιρεί. Ο αριθμός των pixels που προστίθενται ή αφαιρούνται από τα αντικείμενα σε μια εικόνα εξαρτάται από το μέγεθος και το σχήμα του *δομικού στοιχείου* που χρησιμοποιείται για την επεξεργασία της εικόνας. Στις μορφολογικές λειτουργίες διάβρωση και διαστολή, η κατάσταση κάθε pixel της εικόνας εξόδου καθορίζεται εφαρμόζοντας έναν κανόνα στο αντίστοιχο pixel της εικόνας εισόδου και στα γειτονικά σε αυτό pixels. Ο κανόνας που θα επιλεγεί να χρησιμοποιηθεί στην επεξεργασία των pixels ορίζει το

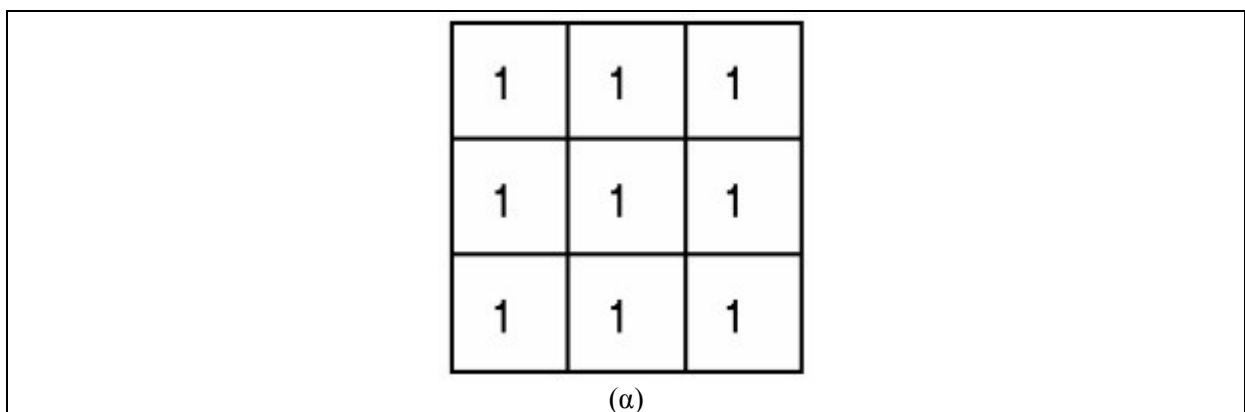
αν η λειτουργία θα είναι διαστολή ή διάβρωση. Ο παρακάτω πίνακας παρουσιάζει τους κανόνες αυτούς.

Λειτουργία	Κανόνας
Dilation	Η τιμή του pixel στην έξοδο είναι η <i>μέγιστη</i> τιμή όλων των pixels στη γειτονιά του pixel εισόδου. Σε μια ασπρόμαυρη εικόνα, αν κάποιο από αυτά τα pixels έχει την τιμή 1, η τιμή του pixel εξόδου γίνεται 1.
Erosion	Η τιμή του pixel στην έξοδο είναι η <i>ελάχιστη</i> τιμή όλων των pixels στη γειτονιά του pixel εισόδου. Σε μια ασπρόμαυρη εικόνα, αν κάποιο από αυτά τα pixels έχει την τιμή 0, η τιμή του pixel εξόδου γίνεται 0.

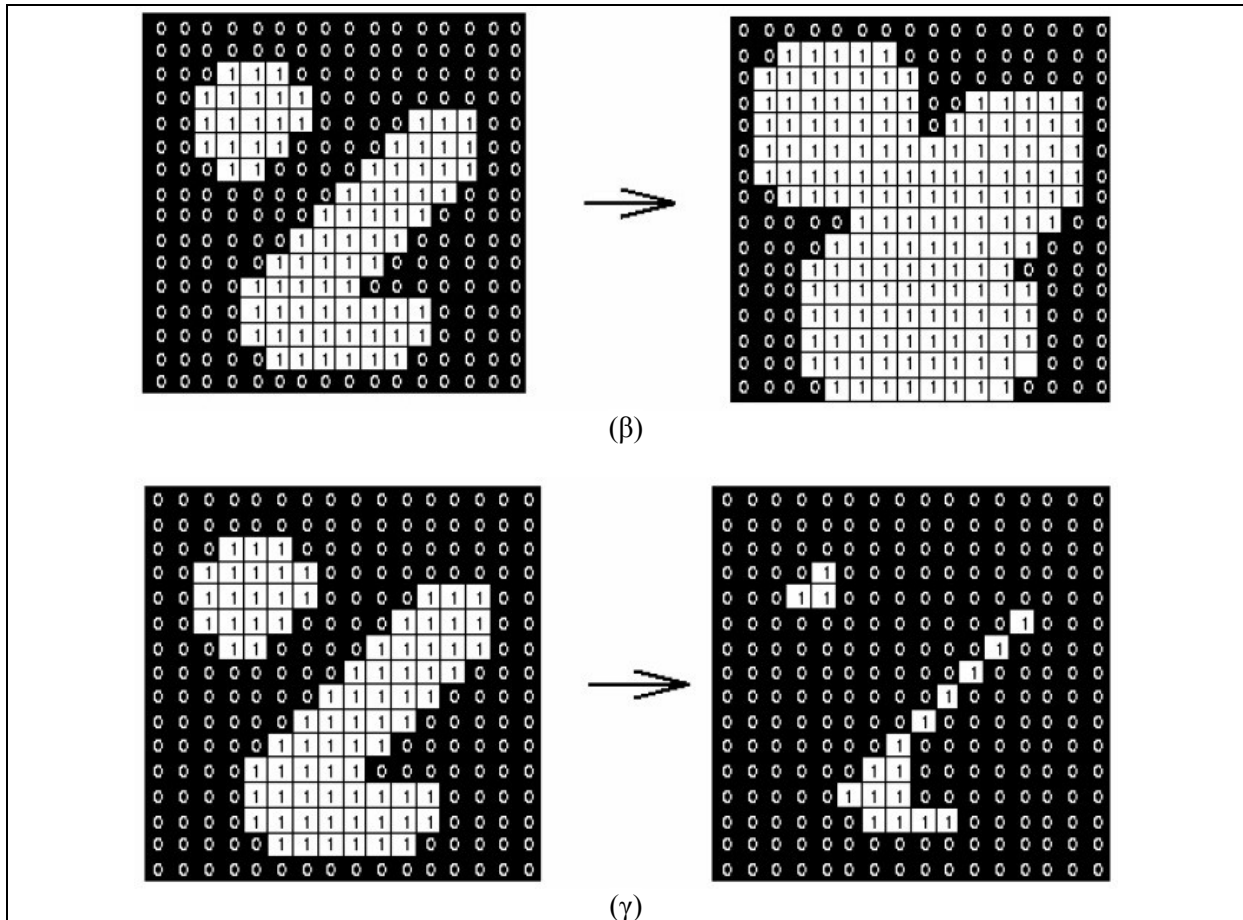
Πίνακας 1. Κανόνες για dilation και erosion¹



Εικόνα 23. Επάνω: εφαρμογή του κανόνα για dilation σε δυαδική εικόνα. Κάτω: εφαρμογή του κανόνα dilation για grayscale εικόνα.



¹ <http://www.mathworks.com/help/toolbox/images/f18-12508.html>



Εικόνα 24. Παράδειγμα dilation και erosion σε μια δυαδική εικόνα. (α) Το δομικό στοιχείο. (β) Dilation. (γ) Erosion.

Οι μορφολογικές αυτές λειτουργίες δρουν όπως είδαμε τοποθετώντας τη βάση του structuring element, δηλαδή το κεντρικό στοιχείο του, πάνω από το pixel που θα επεξεργαστεί στην εικόνα εισόδου. Όσον αφορά τα pixels στα σύνορα-άκρα της εικόνας, τμήματα της γειτνίασης που προσδιορίζονται από το δομικό στοιχείο μπορούν να επεκταθούν πέρα από το περίγραμμα της εικόνας.

Για να επεξεργαστούν τα pixel που βρίσκονται στα όρια της εικόνας, οι μορφολογικές λειτουργίες εκχωρούν μια τιμή σε αυτά τα απροσδιόριστα pixel και η διαδικασία συνεχίζεται σαν να έχει 'γεμίσει' η εικόνα με πρόσθετες γραμμές και στήλες. Οι τιμές αυτών των pixel γεμίσματος (padding pixels) διαφέρει στο dilation και στο erosion. Ο πίνακας που ακολουθεί περιγράφει τους κανόνες γεμίσματος για dilation και erosion για ασπρόμαυρες και γκριζαρισμένες εικόνες.

Λειτουργία	Κανόνες
Dilation	Σε pixels πέρα από τα σύνορα της εικόνας αποδίδεται η ελάχιστη τιμή που παρέχεται από τον τύπο των δεδομένων. Για ασπρόμαυρες (binary) εικόνες στα pixel αυτά η τιμή είναι ρυθμισμένη στο 0. Για γκριζαρισμένες εικόνες, η ελάχιστη τιμή για εικόνες τύπου uint8 είναι το 0.
Erosion	Σε pixels πέρα από τα σύνορα της εικόνας αποδίδεται η μέγιστη τιμή που παρέχεται από τον τύπο των δεδομένων. Για ασπρόμαυρες (binary) εικόνες στα pixel αυτά η τιμή είναι ρυθμισμένη στο 1. Για γκριζαρισμένες εικόνες, η μέγιστη τιμή για εικόνες τύπου

	uint8 είναι το 255.
--	---------------------

Πίνακας 2. Κανόνες γεμίματος εικόνων κατά τη διαδικασία επεξεργασίας των συνοριακών pixels.

Σημείωση: Χρησιμοποιώντας την ελάχιστη τιμή για τη λειτουργία dilation και τη μέγιστη για τη λειτουργία erosion, αποφεύγουμε το *border effect*, κατά το οποίο περιοχές κοντά στα σύνορα-άκρα της εικόνας εξόδου δεν εμφανίζονται να είναι ομοιογενείς με την υπόλοιπη εικόνα. Για παράδειγμα, αν η λειτουργία διάβρωσης (erosion) γεμίσει με την ελάχιστη τιμή (σε αντίθεση με τον κανόνα), θα έχουμε ως αποτέλεσμα ένα μαύρο σύνορο γύρω από τα άκρα της εικόνας εξόδου.

Συνδυασμός Dilation και Erosion

Οι λειτουργίες dilation και erosion συχνά χρησιμοποιούνται συνδυαζόμενες σε λειτουργίες επεξεργασίας εικόνας. Για παράδειγμα, ο ορισμός της μορφολογικής λειτουργίας *opening* μιας εικόνας είναι ένα erosion ακολουθούμενο από ένα dilation, χρησιμοποιώντας το ίδιο δομικό στοιχείο και στις δύο εφαρμογές. Η συγγενική μορφολογική λειτουργία *closing* σε μια εικόνα είναι το αντίστροφο: αποτελείται από ένα dilation ακολουθούμενο από erosion, με χρήση του ίδιου δομικού στοιχείου. Έστω το αντικείμενο A και το δομικό στοιχείο B εικόνες του δισδιάστατου χώρου. B_x είναι η μετατροπή του B ώστε η αρχή του είναι το x . Το erosion της εικόνας A με δομικό στοιχείο B ορίζεται σαν:

$$A \ominus B = \{x: B_x \subseteq A\}.$$

Αντίστοιχα το dilation της εικόνας A με δομικό στοιχείο B ορίζεται:

$$A \oplus B = \{x: \widehat{B}_x \cap A \neq \emptyset\}.$$

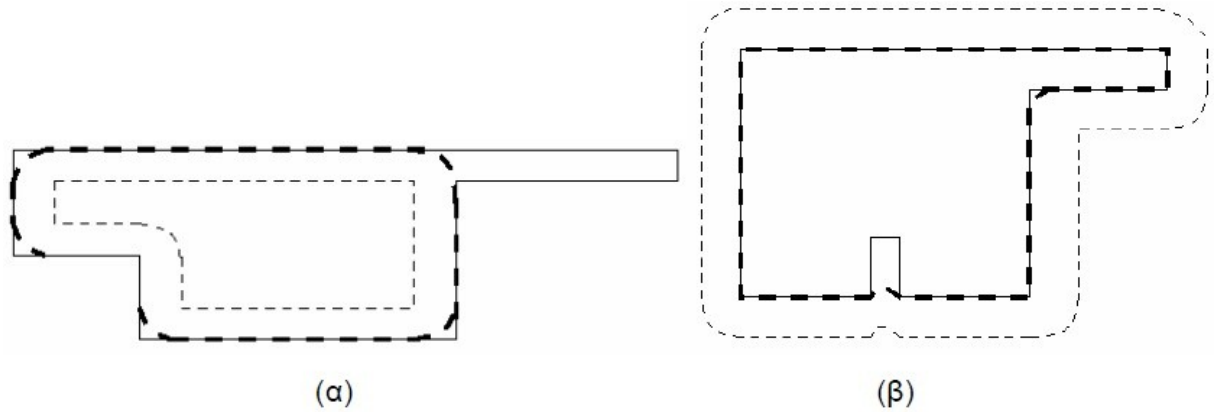
Το opening είναι ένα erosion το οποίο ακολουθείται από ένα dilation:

$$A \circ B = (A \ominus B) \oplus B,$$

ενώ το closing είναι ένα dilation το οποίο ακολουθείται από ένα erosion:

$$A \bullet B = (A \oplus B) \ominus B.$$

Το opening και το closing δεν έχουν κάποια επίδραση στο πάχος της εικόνας (διόγκωση ή συρρίκνωση) όμως το opening βοηθάει στην μείωση του θορύβου, στην εξομάλυνση της περιμέτρου των αντικειμένων, στην απομόνωση αντικειμένων που έχουν ενωθεί λόγω κακής ψηφιοποίησης ή θορύβου, ενώ το closing βοηθάει στην κάλυψη οπών ή ασυνεχειών στο σώμα των αντικειμένων καθώς και στην εξομάλυνση της περιμέτρου των αντικειμένων. Στην *Εικόνα 25* φαίνεται μία αναπαράσταση του opening και του closing. Η αρχική εικόνα έχει συνεχόμενη γραμμή, ενώ με έντονη διακεκομμένη γραμμή αναπαρίσταται η τελική εικόνα μετά τον μετασχηματισμό. Με λεπτή διακεκομμένη γραμμή αναπαρίσταται ο ενδιάμεσος μετασχηματισμός (erosion για το opening και dilation για το closing).



Εικόνα 26. Αναπαράσταση του opening και του closing της εικόνας. (α) Αναπαράσταση του opening.
(β) Αναπαράσταση του closing.

Στις εικόνες που ακολουθούν, δίδονται παραδείγματα εφαρμογής των βασικών αυτών μορφολογικών πράξεων για προεπεξεργασία δυαδικής εικόνας εγγράφου. Συγκεκριμένα, χρησιμοποιείται το opening για εξομάλυνση της περιμέτρου των χαρακτήρων, το erosion για να ξεχωρίσουν οι χαρακτήρες και το closing για την κάλυψη οπών ή ασυνεχειών στο σώμα των χαρακτήρων.

Dr. P.N. C
Mining Sur
Holroyd Ro
Reading,
Berks.



Dr. P.N. C
Mining Sur
Holroyd Ro
Reading,
Berks.

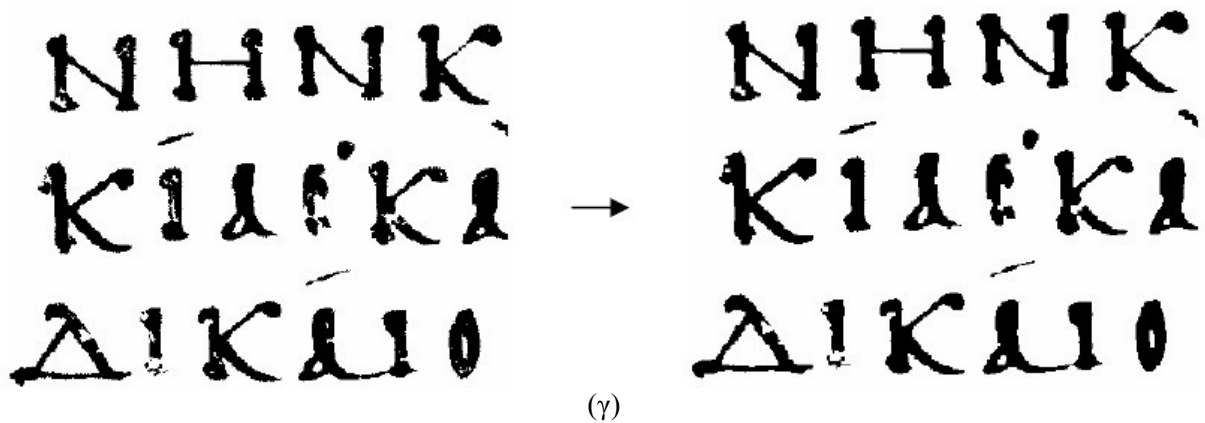
(α)

Permit
transmission



Permit
transmission

(β)

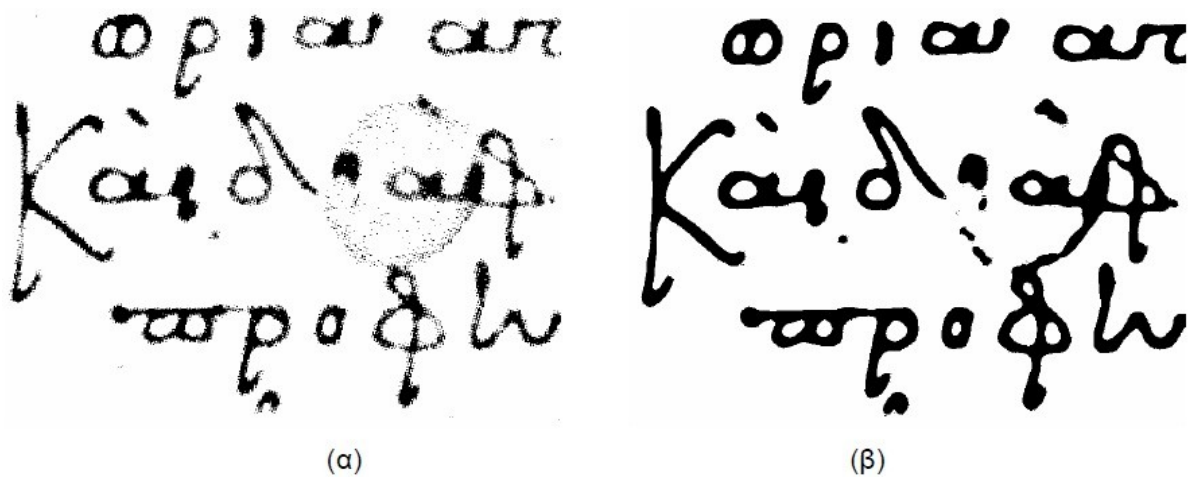


Εικόνα 27. Εφαρμογή μορφολογικών πράξεων για την προεπεξεργασία εικόνας εγγράφου. (α) Εφαρμογή opening για εξομάλυνση της περιμέτρου των χαρακτήρων. (β) Εφαρμογή erosion για απομόνωση των χαρακτήρων. (γ) Εφαρμογή closing για την κάλυψη οπών ή ασυνεχειών στο σώμα των χαρακτήρων.

Φίλτρα συρρίκνωσης και διόγκωσης

Δύο παραλλαγές των μορφολογικών λειτουργιών dilation και erosion είναι τα φίλτρα διόγκωσης και συρρίκνωσης (swell και shrink αντίστοιχα) (Schilling)¹⁹. Ένα $n \times n$ παράθυρο εφαρμόζεται σε κάθε σημείο εικόνας ή υποβάθρου και αφού υπολογίζεται το άθροισμα των σημείων υποβάθρου ή εικόνας μέσα στο παράθυρο αποφασίζεται η μετατροπή ή όχι του συγκεκριμένου σημείου. Τα φίλτρα συρρίκνωσης και διόγκωσης εφαρμόζονται στις δυαδικές εικόνες εγγράφων για να μειωθεί ο θόρυβος, να βελτιωθεί η ποιότητα των περιοχών κειμένου καθώς και η συνεκτικότητα των χαρακτήρων αφαιρώντας απομονωμένα pixels και γεμίζοντας τυχόν σπασίματα, κενά ή τρύπες στο σώμα των χαρακτήρων. Πιο συγκεκριμένα:

- Το φίλτρο συρρίκνωσης (shrink filter) χρησιμοποιείται για την αφαίρεση θορύβου από το υπόβαθρο της εικόνας. Σαρώνεται ολόκληρη η ασπρόμαυρη εικόνα και εξετάζεται κάθε σημείο κειμένου. Αν P_{sh} είναι ο αριθμός των pixels του υποβάθρου σε ένα κινούμενο παράθυρο $n \times n$, το οποίο έχει το σημείο του κειμένου σαν κεντρικό, τότε το σημείο μετατρέπεται σε σημείο υποβάθρου της εικόνας αν $P_{sh} > k_{sh}$ όπου k_{sh} ορίζεται πειραματικά.
- Το φίλτρο διόγκωσης (swell filter) χρησιμοποιείται για το γέμισμα πιθανών ασυνεχειών, κενών ή οπών στην επιφάνεια των χαρακτήρων. Σαρώνεται ολόκληρη η ασπρόμαυρη εικόνα και εξετάζεται κάθε σημείο υποβάθρου. Αν P_{sw} είναι ο αριθμός των pixels του κειμένου σε ένα κινούμενο παράθυρο $n \times n$, το οποίο έχει το σημείο (x,y) του υποβάθρου σαν κεντρικό, τότε το σημείο μετατρέπεται σε σημείο κειμένου της εικόνας αν $P_{sw} > k_{sw}$ όπου k_{sw} ορίζεται πειραματικά. Για να αποφύγουμε την πάχυνση των χαρακτήρων, μπορούμε εκτός από την συνθήκη $P_{sw} > k_{sw}$ να ελέγχουμε και τις συνθήκες $|x-x_a| < dx$ και $|y-y_a| < dy$ για τη μετατροπή ενός σημείου υποβάθρου σε κείμενο, όπου x_a, y_a οι μέσες τιμές των συντεταγμένων x, y στο παράθυρο $n \times n$. Με αυτό τον τρόπο ελέγχουμε μόνο τα σημεία υποβάθρου που βρίσκονται ανάμεσα σε ομοιόμορφα κατανεμημένα σημεία κειμένου. Η διαδοχική εφαρμογή των παραπάνω φίλτρων σε δυαδικές εικόνες εγγράφων έχει σημαντικά αποτελέσματα ως προς την βελτίωση της ποιότητας (Εικόνα 27).



Εικόνα 28. Παράδειγμα διαδοχικής εφαρμογής φίλτρων συρρίκνωσης (*shrink*) και διόγκωσης (*swell*) σε δυαδική εικόνα εγγράφου. (α) Αρχική ασπρόμαυρη εικόνα. (β) Τελική εικόνα μετά την εφαρμογή των φίλτρων.

Όσον αφορά άλλα προβλήματα όπως η κλίση των χαρακτήρων ή η κλίση ολόκληρης της εικόνας τις επιγραφής, ίσως σε κάποιες επιγραφές μόνο η κλίση των χαρακτήρων να σχολιασθεί, αφού αυτούς μεμονωμένα εξετάζουμε. Ωστόσο, οι περιπτώσεις που παρουσιάζεται κλίση των χαρακτήρων σε κείμενο αρχαίων επιγραφών, κυρίως αυτών που είναι χαραγμένες σε επιστόμια ή κυλίνδρους όπου ο βαθμός δυσκολίας της επεξεργασίας είναι πολύ υψηλότερος, δεν θα μας απασχολήσουν προς το παρόν, επειδή τίθενται διαφορετικά κριτήρια στην αντιμετώπιση από αυτά που στην παρούσα εργασία εξετάζουμε.

3.2. Κατάτμηση Χαρακτήρων

Το επόμενο βήμα, στη διαδρομή για την κατηγοριοποίηση και αναγνώριση των χαρακτήρων, είναι η κατάτμηση – απομόνωση της περιοχής που περιλαμβάνει τους χαρακτήρες. Το στάδιο αυτό είναι ένα από τα πιο σημαντικά στάδια της διαδικασίας αναγνώρισης κειμένου. Ο βαθμός επιτυχίας στο στάδιο αυτό εξαρτάται κατά πολύ από τα αποτελέσματα του σταδίου της προεπεξεργασίας (δυναδική μετατροπή, βελτίωση ποιότητας κλπ). Σε μια εικόνα χωρίς πολύ θόρυβο και με ευδιάκριτα τα χαρακτηριστικά των γραμμάτων μπορούν να εφαρμοστούν πολύ πιο απλές μέθοδοι κατάτμησης των χαρακτήρων, οι οποίες δεν απαιτούν μεγάλη επεξεργαστική ισχύ, ούτε μεγάλο χρόνο εκτέλεσης.

Στις υπάρχουσες τεχνικές συγκαταλέγονται αυτές που εφαρμόζονται σε δυαδικές εικόνες, τεχνικές οι οποίες προχωρούν στην κατάτμηση-εξαγωγή του χαρακτήρα από την εικόνα και στη συνέχεια μετατρέπουν την εικόνα του αποκομμένου χαρακτήρα σε δυαδική, και τεχνικές οι οποίες απλά προχωρούν στην κατάτμηση του χαρακτήρα και στη συνέχεια προχωρούν στο επόμενο στάδιο κατηγοριοποίησης. Η επιλογή της εκάστοτε μεθόδου γίνεται σύμφωνα με το αποτέλεσμα στο οποίο θέλουμε να καταλήξουμε, τους πόρους που διαθέτουμε, και το χρόνο στον οποίο θέλουμε να λάβουμε το αποτέλεσμα.

Ο απλούστερος τρόπος για την κατάτμηση των γραμμάτων είναι να χρησιμοποιήσουμε το κενό διάστημα μεταξύ τους σαν όριο της προς κατάτμηση περιοχής. Ωστόσο αυτή η τεχνική οδηγεί αρκετές φορές σε αποτυχία αν το κείμενο που έχουμε προς κατάτμηση περιέχει ενωμένους ή μη συνεχείς χαρακτήρες (κακή προεπεξεργασία όπως αναφέρθηκε και πιο πάνω).

Μέθοδος Rajiv Kumar και Amardeep Singh

Οι προαναφερθέντες ερευνητές χρησιμοποίησαν την μέθοδο αυτή για κατάτμηση χειρόγραφου κειμένου της Gurmukhi η οποία είναι από τους πιο συνηθισμένους τρόπους για να γράψεις Punjabi, μια δημοφιλή γλώσσα της βόρειας Ινδίας²⁰.

Στη μέθοδο αυτή, το γραπτό χωρίζεται σε τρεις οριζόντιες ζώνες ονομαζόμενες *άνω ζώνη*, *μέση ζώνη* και *κάτω ζώνη*. Οι χαρακτήρες εμφανίζονται κυρίως στη μέση ζώνη, ενώ στις άνω και κάτω ζώνες περιέχονται τόνοι, πνεύματα και αποσπάσματα σημείων στίξης. Χρήση των πληροφοριών των ζωνών αυτών γίνεται κατά τη διαδικασία κατάτμησης των γραμμάτων, όσο και στη διαδικασία αναγνώρισης-κατηγοριοποίησης των χαρακτήρων.



Εικόνα 29. Ζώνες κατάτμησης της λέξης. Παράδειγμα από γραπτό Gurmukhi. (a) Άνω ζώνη, (b) Μέση ζώνη, (c) Κάτω ζώνη.

Βασικό κομμάτι της τεχνικής των εν λόγω ερευνητών είναι το ‘Ελαστικό παράθυρο’ το οποίο θέτει τα όρια της λέξης πριν αυτή χωριστεί σε περαιτέρω ζώνες. Ελαστικό παράθυρο είναι εκείνο του οποίου οι διαστάσεις μπορούν να οριστούν σύμφωνα με τις απαιτήσεις της κάθε περίπτωσης.

Παρακάτω ακολουθεί περιγραφή του θέματος ενός παραθύρου το οποίο είναι ελαστικό ως προς το μέγεθος.

Από την αρχή της εικόνας (συνήθως είναι το σημείο 0,0), παίρνουμε τη θέση του πρώτου pixel που παρουσιάζεται στη γραμμή και κινούμαστε προς τα δεξιά μέχρι να συναντήσουμε το πιο δεξί pixel της συγκεκριμένης γραμμής. Έτσι έχουμε το πιο δεξί και το πιο αριστερό σημείο της γραμμής αυτής. Υποθέτουμε ότι υπάρχουν n γραμμές από pixels π.χ. $r1, r2, \dots, rn$ σε μια γραμμή κειμένου. Έτσι θα υπάρχουν n θέσεις pixel για τα πιο δεξιά και αριστερά άκρα επίσης. Αυτές είναι οι $Lx1, Lx2, \dots, Lxn$ για την αριστερή πλευρά (Left) και $Rx1, Rx2, \dots, Rxn$ για τη δεξιά πλευρά (Right). Οπότε για μέγιστο και ελάχιστο:

$$MIN X = gub(Lx1, Lx2, \dots, Lxn)$$

//gub – greatest upper bound

$$MAX X = llb(Rx1, Rx2, \dots, Rxn)$$

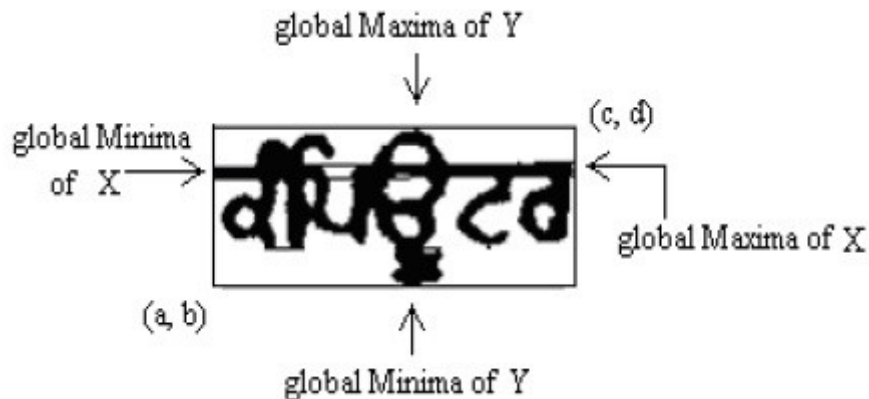
//llb = least lower bound

Για να βρούμε τώρα το οριζόντιο σύνορο του παραθύρου, χρησιμοποιούμε την ιδέα της ζώνης χωρίς pixel (no pixel zone) δηλαδή τη θεωρία ότι ανάμεσα από δύο ζώνες χωρίς pixel (πληροφορίας), υπάρχει μια γραμμή κειμένου. Υποθέτουμε ότι υπάρχουν n θέσεις pixel για την πιο επάνω πλευρά και n θέσεις για την πιο κάτω κοντά στις ζώνες χωρίς pixel. Έτσι $Ly1, Ly2, \dots, Lyn$ είναι οι n θέσεις pixel για την κάτω πλευρά και $Ry1, Ry2, \dots, Ryn$ για την άνω πλευρά. Οπότε για μέγιστο και ελάχιστο:

$$MIN Y = gub(Ly1, Ly2, \dots, Lyn)$$

$$MAX Y = llb(Ry1, Ry2, \dots, Ryn)$$

Κατ' αυτό τον τρόπο, ένα παράθυρο με συντεταγμένες $(MINX, MINY)$, $(MINX, MAXY)$, $(MAXX, MAXY)$ και $(MAXX, MINY)$ έχει σχηματιστεί, οι οποίες καταγράφονται σε ένα αρχείο. Έτσι, όλες οι γραμμές κειμένου έχουν ταυτοποιηθεί και οι θέσεις τους βρίσκονται γραμμένες σε αρχείο.



Εικόνα 30. Παράθυρο με συντεταγμένες (a, b) και (c, d) .

Με ανάλογο τρόπο ορίζονται παράθυρα για τα γράμματα της κάθε λέξης.

Μέθοδος Barmpoutis, Bozia, Wagman

Η συγκεκριμένη μέθοδος, η οποία περιγράφεται στο επιστημονικό άρθρο των Angelos Barmpoutis et al.²¹, δεν εφαρμόζεται σε εικόνα η οποία έχει γίνει εκ των προτέρων ασπρόμαυρη. Στο άρθρο αυτό γίνεται 3D ανακατασκευή επιγραφών μετά από επεξεργασία των εκτύπων τους, ενώ παρουσιάζεται ένα σύνολο τεχνικών για κατηγοριοποίηση χαρακτήρων αρχαίων επιγραφών.

Πιο αναλυτικά, η μέθοδος που ακολουθείται έχει ως εξής: η εικόνα της ανακατασκευασμένης επιγραφής έχει τον απαραίτητο προσανατολισμό ως προς τους x - y άξονες τέτοιο, ώστε κάθε γραμμή κειμένου να είναι παράλληλη με τον άξονα των x . Χρησιμοποιώντας την υπόθεση κάθε γράμμα μπορεί να περιλαμβάνεται σε μια ορθογώνια περιοχή προσανατολισμένη παράλληλα στους καρτεσιανούς άξονες.

Κάθε ορθογώνια περιοχή μπορεί να παραμετροποιηθεί χρησιμοποιώντας 4 πραγματικούς αριθμούς x_1, x_2, y_1, y_2 που θα δηλώνουν την πιο αριστερή και δεξιά συντεταγμένη στον άξονα x και την πιο πάνω και πιο κάτω συντεταγμένη στον άξονα y της περιοχής αυτής αντίστοιχα. Με τον καθορισμό αυτών των παραμέτρων ώστε η περιοχή που προκύπτει να είναι προκαθορισμένου μεγέθους και κεντραρισμένη σε συγκεκριμένο σημείο, μπορούμε να μεταβάλουμε ελαφρώς την ορθογώνια περιοχή με σκοπό να ελαττώσουμε το χρώρο που καταλαμβάνει.

Ο καθορισμός του πρώτου ορθογωνίου τμήματος γίνεται με μη αυτόματο τρόπο, μαρκάροντας τον επάνω αριστερά χαρακτήρα της επιγραφής χρησιμοποιώντας ένα γραφικό περιβάλλον χρήστη. Έχοντας υπολογίσει τις παραμέτρους του πρώτου ορθογωνίου, μπορούμε να καθορίσουμε αυτόματα την περιοχή που περιλαμβάνει το επόμενο γράμμα (μόνο το δεξί) κεντράροντάς το στο:

$$((x_2 - x_1)/2 + c, (y_1 + y_2)/2),$$

όπου c είναι προκαθορισμένη μεταβλητή. Παρόμοιος κανόνας μπορεί να εφαρμοστεί για τον καθορισμό την περιοχής του γράμματος και στην επόμενη σειρά γραμμάτων.



Εικόνα 31. Κατάτμηση χαρακτήρων χωρίς να προηγηθεί μετατροπή της εικόνας σε ασπρόμαυρη.

Μια μέθοδος η οποία μπορεί να χρησιμοποιηθεί και η οποία έχει πολύ καλά αποτελέσματα σε λέξεις με χαρακτήρες ίδιου πλάτους, είναι αυτή των κάθετων προβολών. Η κάθετη προβολή $V(x)$ είναι το ιστόγραμμα που προκύπτει αθροίζοντας όλα τα σημεία εικόνας σε κατακόρυφη διεύθυνση στην οριζόντια θέση x . Σε μία γραμμή κειμένου όπου οι χαρακτήρες χωρίζονται σωστά, η $V(x)$ έχει μηδενικά ανάμεσα στους χαρακτήρες. Επομένως, αν οι χαρακτήρες είναι σε άριστη κατάσταση, δηλαδή διαχωρίζονται τέλεια μεταξύ τους και δεν έχουν σπασίματα, ο διαχωρισμός των χαρακτήρων μπορεί να επιτευχθεί κατευθείαν από τις κάθετες προβολές σε κάθε γραμμή κειμένου. Αν και στις περισσότερες περιπτώσεις το κείμενο δεν είναι σε καλή κατάσταση, εντούτοις οι κάθετες προβολές

αποτελούν το βασικό εργαλείο για την κατάτμηση των χαρακτήρων. Σύμφωνα με τον Lu²², το πρώτο στάδιο αφορά την ομαδοποίηση των γραμμών ανάλογα με τη μεταξύ τους απόσταση και το ύψος του γράμματος. Για παράδειγμα, στην *Εικ.32* έχουμε δύο ομάδες γραμμών κειμένου, η πρώτη γραμμή κειμένου ανήκει στην πρώτη ομάδα, ενώ η δεύτερη και τρίτη γραμμή ανήκουν στην δεύτερη ομάδα.



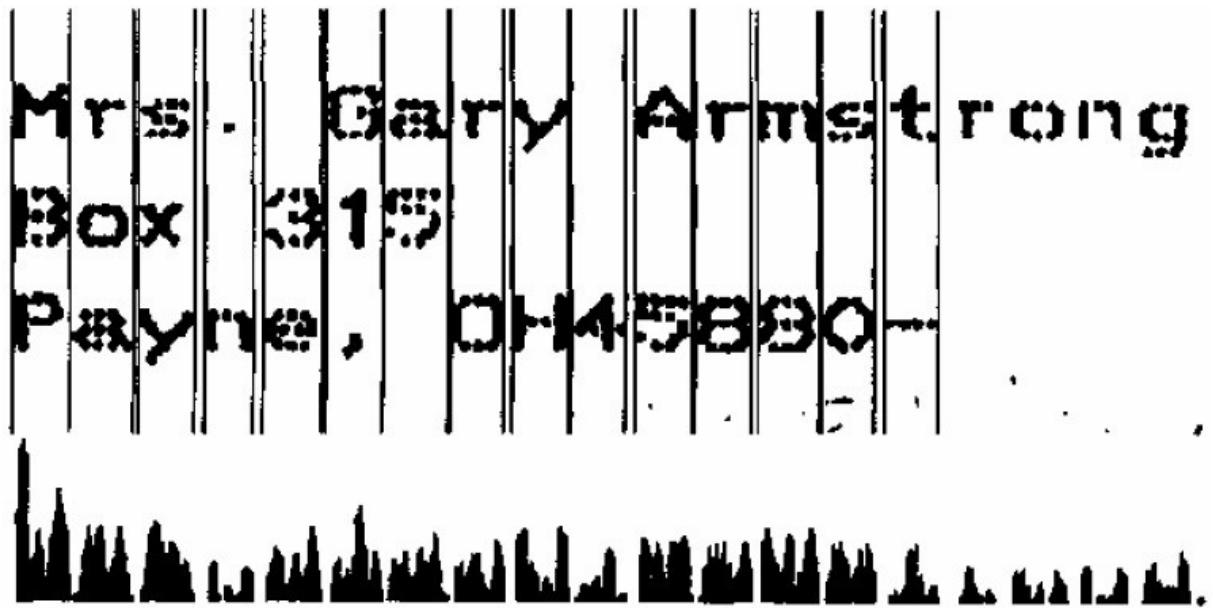
Εικόνα 32. Ομαδοποίηση γραμμών κειμένου. Έχουμε δύο ομάδες γραμμών (πρώτη γραμμή -δεύτερη και τρίτη γραμμή).

Στη συνέχεια υπολογίζονται οι κάθετες προβολές για κάθε ομάδα γραμμών. Τα κενά που μπορεί να εμφανίζονται σε χαρακτήρες με σπασίματα δεν επηρεάζουν την κάθετη προβολή της ομάδας των γραμμών κειμένου. Αντίθετα, ο πιθανός θόρυβος ανάμεσα στους χαρακτήρες μπορεί να εμποδίσει την εμφάνιση μηδενικών στην κάθετη προβολή. Για την αντιμετώπιση ανάλογων προβλημάτων, εξάγουμε στατιστικές τιμές για το πλάτος και την οριζόντια απόσταση των χαρακτήρων (*Εικόνα 33*).



Εικόνα 33. Χαρακτηριστικά γραμμής κειμένου με χαρακτήρες ίδιου πλάτους: Διάστημα ανάμεσα στους χαρακτήρες, κενό ανάμεσα στους χαρακτήρες, πλάτος χαρακτήρων.

Για να έχουμε ακριβή υπολογισμό, οι στατιστικές τιμές υπολογίζονται μόνο για περιοχές που αποτελούνται από περισσότερες της μίας γραμμής κειμένου. Για παράδειγμα, στην *Εικόνα 33* εξάγουμε στατιστικά υπολογίζοντας την κάθετη προβολή μέχρι και για το τέλος της δεύτερης μεγαλύτερης γραμμής.



Εικόνα 34. Εντοπισμός των χαρακτήρων με χρήση των κάθετων προβολών σε ομάδα.

Χρησιμοποιώντας τις στατιστικές τιμές για το πλάτος και την οριζόντια απόσταση των χαρακτήρων μπορούμε να διορθώσουμε το αποτέλεσμα ως εξής: (α) Αν από την κάθετη προβολή προκύπτουν γράμματα με πλάτος μεγαλύτερο από το μέσο πλάτος που έχουμε υπολογίσει προχωράμε σε διαίρεση της αντίστοιχης περιοχής. (β) Αν προκύπτουν γράμματα με μικρότερο πλάτος προχωράμε σε ενώσεις γειτονικών περιοχών μέχρι να προκύψει το μέσο πλάτος του γράμματος που έχουμε υπολογίσει. Αφού ολοκληρωθεί η παραπάνω διαδικασία, επαναλαμβάνεται ο στατιστικός υπολογισμός για το πλάτος και την οριζόντια απόσταση των χαρακτήρων. Αν υπάρχουν ακόμα γράμματα με διαφορετικό πλάτος, τότε κάθε γραμμή της ομάδας οδηγείται σε αντίστοιχη διαδικασία τμηματοποίησης που αφορά μόνο μία γραμμή κειμένου.

Εναλλακτική μέθοδος είναι η μέθοδος εντοπισμού της περιοδικότητας των κενών ανάμεσα στα γράμματα (GPD: gap periodicity detection algorithm). Βασίζεται στο γεγονός ότι τα κενά ανάμεσα στα γράμματα εμφανίζονται σε περιοδικά διαστήματα όταν το πλάτος των γραμμάτων είναι σταθερό. Για κάθε ομάδα γραμμών κειμένου, υπολογίζουμε τις αντίστροφες κάθετες προβολές (προβολές των σημείων υποβάθρου) και βρίσκουμε τον βέλτιστο συνδυασμό του μήκους των κενών ανάμεσα στους χαρακτήρες και της απόστασης ανάμεσα στα κενά που ταιριάζουν καλύτερα στις αντίστροφες κάθετες προβολές. Αν δεν υπάρχει κατάλληλος συνδυασμός μήκους των κενών και της απόστασης ανάμεσα στα κενά ώστε να δίνει αντίστοιχα διαδοχικά μέγιστα στις αντίστροφες κάθετες προβολές, τότε δεν έχει γίνει σωστή ομαδοποίηση των γραμμών κειμένου ή οι γραμμές κειμένου δεν αποτελούνται από χαρακτήρες ομοίου πλάτους.

3.3. Εξαγωγή Χαρακτηριστικών

Η εξαγωγή χαρακτηριστικών γνωρισμάτων αφορά την ανάκτηση των ιδιοτήτων καθορισμού. Είδη χαρακτηριστικών γνωρισμάτων που χρησιμοποιούνται είναι ο αριθμός τρυπών στο χαρακτήρα ή ο αριθμός κοιλωμάτων στο εξωτερικό περίγραμμά του, η σχετική προεξοχή των άκρων του χαρακτήρα, και τα τοπικά χαρακτηριστικά γνωρίσματα, όπως οι σχετικές θέσεις των διασταυρώσεων,

γραμμών και των γωνιών. Ανάλογα με τον τύπο της τεχνικής για εξαγωγή χαρακτηριστικών, ομαδοποιούμε²³ τις τεχνικές ανάλυσης χαρακτηριστικών ως:

1. Γενικοί μετασχηματισμοί και ανάπτυξη σειρών
2. Χαρακτηριστικά που προέρχονται από τη στατιστική κατανομή των σημείων
3. Γεωμετρικά, τοπολογικά και μορφολογικά χαρακτηριστικά

Ένα σύνολο ικανοποιητικών χαρακτηριστικών πρέπει να πληροί ορισμένες προϋποθέσεις όπως:

- Διακριτική ικανότητα. Τα χαρακτηριστικά πρέπει να έχουν αρκετά διαφορετικές τιμές για χαρακτήρες οι οποίοι ανήκουν σε διαφορετικές κλάσεις.
- Αξιοπιστία. Τα χαρακτηριστικά πρέπει να έχουν παραπλήσιες τιμές για χαρακτήρες οι οποίοι ανήκουν στην ίδια κλάση.
- Ανεξαρτησία. Τα χαρακτηριστικά δεν πρέπει να επικαλύπτονται μεταξύ τους.
- Μικρό μέγεθος. Ο αριθμός των χαρακτηριστικών πρέπει να είναι μικρός ώστε η αναγνώριση να γίνεται απλά και γρήγορα.
- Μικρό υπολογιστικό κόστος.
- Μικρή πολυπλοκότητα εξαγωγής.
-

Επίσης, τα χαρακτηριστικά θα πρέπει να μένουν αναλλοίωτα όταν ο χαρακτήρας εμφανίζει:

- Θόρυβο
- Μετακίνηση
- Αλλαγή στο μέγεθος
- Στροφή
- Κλίση
- Παραμόρφωση

3.3.1. Γενικοί μετασχηματισμοί και ανάπτυξη σειρών

Μετασχηματισμός ενός σήματος είναι η αναπαράσταση του με κάποιο διαφορετικό τρόπο χωρίς όμως να αλλάξουμε την πληροφορία που μεταφέρει. Η τεχνική γενικών μετασχηματισμών και ανάπτυξης σειρών βοηθάει στη μείωση των διαστάσεων του διανύσματος χαρακτηριστικών και δίνει χαρακτηριστικά ανεξάρτητα σε μερικές παραμορφώσεις. Ερευνητές έχουν χρησιμοποιήσει μετασχηματισμούς όπως του Fourier, Walsh, Hadamard και άλλους. Τα χαρακτηριστικά εξάγονται εύκολα με αυτές τις τεχνικές αλλά απαιτούν μεγάλη υπολογιστική ισχύ:

Μετασχηματισμοί Fourier

Η γενική διαδικασία είναι να επιλεγεί το φάσμα πλάτους του διανύσματος μέτρησης ως χαρακτηριστικό γνώρισμα σε ένα n -διάστατο ευκλείδειο χώρο. Μια από τις ελκυστικότερες ιδιότητες του μετασχηματισμού Φουριέ είναι η δυνατότητα να αναγνωριστεί η θέση μετατοπισμένων

χαρακτήρων, όταν παρατηρηθεί το φάσμα πλάτους και αγνοηθεί η φάση. Οι μετασχηματισμοί Fourier έχουν εφαρμοστεί στην αναγνώριση χαρακτήρα με πολλούς τρόπους.

Μετασχηματισμοί Wavelet

Ένα Wavelet είναι μια κυματοειδής ταλάντωση $\psi \in L^2(\mathbb{R})$ με ένα εύρος που αρχίζει από μηδέν, αυξάνει, και έπειτα μειώνεται πάλι σε μηδέν, δηλαδή ένα σήμα ενέργειας. Ο μετασχηματισμός Wavelet είναι μια τεχνική ανάπτυξης σειράς που μας επιτρέπει να απεικονίσουμε το σήμα σε διαφορετικά επίπεδα ευκρίνειας, δηλαδή να αναλύσουμε διαφορετικές συχνότητες με διαφορετικές κλίμακες. Αυτή η διαδικασία ονομάζεται πολυδιακριτική ανάλυση.

Πιο συγκεκριμένα: Η πολυδιακριτική ανάλυση ενός χώρου $L^2(\mathbb{R})$ αποτελείται από μία ακολουθία υποχώρων:

$$\{0\} \cdots \subset V_0 \subset V_1 \cdots \subset V_n \subset V_{n+1} \cdots \subset L^2(\mathbb{R}),$$

οι οποίοι ικανοποιούν τις εξής απαιτήσεις:

$$\forall (j, k) \in \mathbb{Z}^2, \quad f(t) \in V_j \Leftrightarrow f(t - 2^j k) \in V_j,$$

$$\forall j \in \mathbb{Z}, \quad V_{j+1} \subset V_j,$$

$$\forall j \in \mathbb{Z}, \quad f(t) \in V_j \Leftrightarrow f\left(\frac{t}{2}\right) \in V_{j+1},$$

$$\lim_{j \rightarrow +\infty} V_j = \bigcap_{j=-\infty}^{+\infty} V_j = \{0\},$$

$$\lim_{j \rightarrow -\infty} V_j = \text{Closure} \left(\bigcup_{j=-\infty}^{+\infty} V_j \right) = L^2(\mathbb{R}),$$

και υπάρχει θ τέτοιο ώστε $\{\theta(t-n)\}_{n \in \mathbb{Z}}$ να είναι βάση Riesz του V_0 . (Mallat). Δηλαδή:

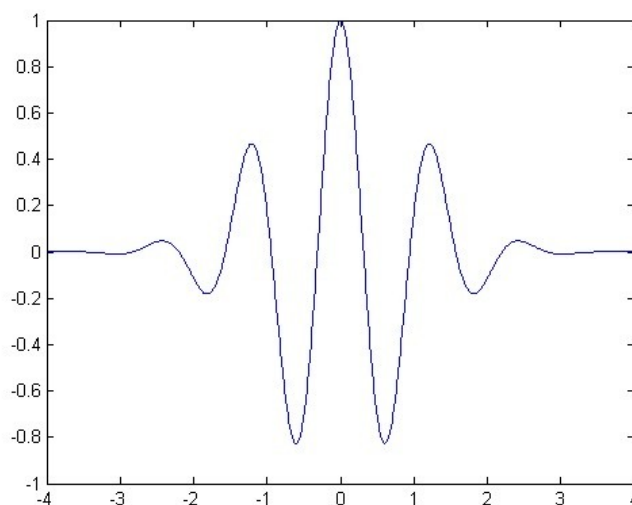
1. Κάθε υποχώρος παραμένει αναλλοίωτος όταν μετατεθεί κατά κάποιο πολλαπλάσιο του 2^{-k} .
2. Για υποχώρους $V_k \subset V_l$, $k < l$, ισχύει ότι ένας είναι κλίμακα του άλλου κατά ένα συντελεστή καθυστέρησης ίσο με 2^{k-l} .
3. Ο υποχώρος V_0 πρέπει να δημιουργείται ως το γραμμικό περίβλημα (linear hull) μίας ή περισσότερων συναρτήσεων που ονομάζονται συναρτήσεις κλίμακας.
4. Οι υποχώροι καλύπτουν ολόκληρο το χώρο L^2 .

Συνεπώς, μεταθέτοντας και καθυστερώντας το mother wavelet μπορούμε να κατασκευάσουμε την οικογένεια από wavelets ως:

$$\left\{ \psi_{j,n}(t) = \frac{1}{\sqrt{2^j}} \psi \left(\frac{t - 2^j n}{2^j} \right) \right\}_{(j,n) \in \mathbb{Z}^2},$$

η οποία αποτελεί ορθοκανονική βάση του L^2 .

Τα τμήματα της εικόνας κειμένου, που αντιστοιχούν σε γράμματα ή λέξεις, αντιπροσωπεύονται από τους συντελεστές Wavelet, που αντιστοιχούν στα διάφορα επίπεδα ευκρίνειας. Αυτοί οι συντελεστές προωθούνται έπειτα σε έναν ταξινομητή για την αναγνώριση. Η απεικόνιση στην ανάλυση multiresolution (MRA) με χαμηλή ευκρίνεια μπορεί να απορροφήσει καλύτερα παραλλαγές στη γραφή από ότι σε MRA με υψηλή ευκρίνεια. Εντούτοις, η απεικόνιση στη χαμηλή ευκρίνεια μπορεί να είναι τέτοια, ώστε σημαντικές λεπτομέρειες για το στάδιο αναγνώρισης να χαθούν.



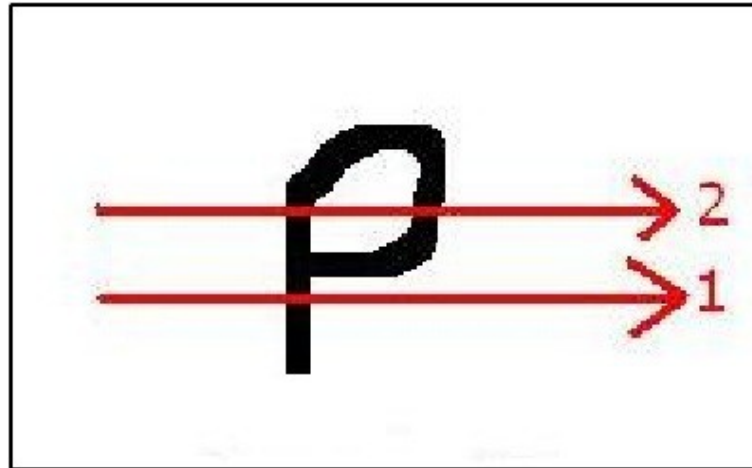
Εικόνα 35. Morlet Wavelet

3.3.2. Χαρακτηριστικά που προέρχονται από τη στατιστική κατανομή των σημείων

Χαρακτηριστικά που προέρχονται από τη στατιστική κατανομή των σημείων είναι ανθεκτικά σε παραμόρφωση και εφαρμόζονται σε αρκετές διαφορετικές γραμματοσειρές. Παρόλο που η εφαρμογή τους είναι σχετικά γρήγορη και απλή, η δημιουργία μάσκας είναι δύσκολη. Μερικές από τις χαρακτηριστικές τεχνικές μέσα σε αυτήν την περιοχή παρατίθενται πιο κάτω:

- **Χωρισμός (Zoning)**
Το ορθογώνιο που περιγράφει το χαρακτήρα διαιρείται σε αρκετές περιοχές που επικαλύπτονται ή όχι, και οι πυκνότητες των μαύρων σημείων μέσα σε αυτές τις περιοχές υπολογίζονται και χρησιμοποιούνται ως χαρακτηριστικά γνωρίσματα.
- **Διασταυρώσεις και αποστάσεις** Στην τεχνική διασταύρωσης τα χαρακτηριστικά γνωρίσματα βρίσκονται από τον αριθμό φορών που η μορφή χαρακτήρα διασχίζεται από τα διανύσματα

κατά μήκος ορισμένων κατευθύνσεων. Αυτή η τεχνική χρησιμοποιείται συχνά από τα εμπορικά συστήματα επειδή μπορεί να εκτελεσθεί με υψηλή ταχύτητα και απαιτεί χαμηλή πολυπλοκότητα. Κατά τη χρησιμοποίηση της τεχνικής απόστασης ορισμένα από τα μήκη μετρώνται κατά τη διάσχιση των διανυσμάτων από τη μορφή χαρακτήρα. Παραδείγματος χάριν το μήκος των διανυσμάτων μέσα στο όριο των χαρακτήρα.

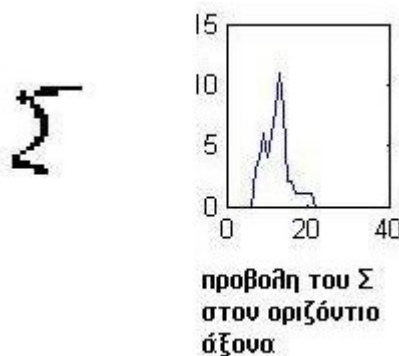


- **n-tuples**

Η ύπαρξη μαύρων και άσπρων σημείων (πρώτο πλάνο και υπόβαθρο) σε ορισμένες προσδιορισμένες διατάξεις, χρησιμοποιείται ως χαρακτηριστικό γνώρισμα.

- **Προβολές**

Η κατανομή σημείων χαρακτήρων προβεβλημένων πάνω σε κάποιο άξονα μπορεί να προσφέρει πληροφορίες για την ταυτότητα του εκάστοτε χαρακτήρα. Στην ουσία, μετατρέπουμε ένα δισδιάστατο σήμα σε μονοδιάστατο.



Εικόνα 36. Παράδειγμα προβολής γράμματος.

3.3.3. Γεωμετρικά, τοπολογικά και μορφολογικά χαρακτηριστικά

Η μέθοδος ανάλυσης με γεωμετρικά και τοπολογικά χαρακτηριστικά είναι η πιο διαδεδομένη στους ερευνητικούς κύκλους. Τα χαρακτηριστικά αυτά μπορούν να αντιπροσωπεύουν γενικές και ειδικές ιδιότητες των χαρακτήρων όπως ακμές, κόλπους, κύκλους κτλ. Έχουν επίσης μεγάλη αντοχή σε παραμορφώσεις και διαφορετικά στιλ χαρακτήρων. Με τη βοήθειά τους η επεξεργασία χαρακτήρων γίνεται αρκετά γρήγορη. Από την άλλη μεριά η διαδικασία εξαγωγής τους είναι επίπονη και η δημιουργία μασκών δύσκολη. Ομαδοποιώντας τις διαφορετικές μεθόδους ανάλυσης, μπορούμε να τις εντάξουμε στις εξής κατηγορίες:

- **Εξαγωγή τοπολογικών δομών**

Σε αυτή τη κατηγορία, αναζητούμε μια προκαθορισμένη δομή μέσα σε κάποιο χαρακτήρα ή λέξη. Ο αριθμός ή η σχετική με το χαρακτήρα θέση αυτών των δομών δημιουργεί μια περιγραφική απεικόνιση. Κοινές πρωτογενείς δομές είναι οι ακμές, οι καμπύλες, οι ευθείες γραμμές κτλ.

- **Μέτρηση και προσέγγιση των γεωμετρικών ιδιοτήτων**

Σε πολλές μελέτες, οι μετρήσεις γεωμετρικών ποσοτήτων απεικονίζουν τους χαρακτήρες. Τέτοιες είναι η αναλογία μεταξύ πλάτους και ύψους του χαρακτήρα, η σχετική οριζόντια και κάθετη απόσταση μεταξύ πρώτου και τελευταίου σημείου, η απόσταση μεταξύ δύο σημείων, το μήκος μιας λέξης. Ένα πολύ σημαντικό χαρακτηριστικό μέτρο είναι η καμπυλότητα ή η μεταβολή της καμπυλότητας.

- **Κωδικοποίηση**

Μία πολύ δημοφιλής μέθοδος κωδικοποίησης είναι η μέθοδος Freeman. Αυτή η κωδικοποίηση επιτυγχάνεται κυρίως με τη χαρτογράφηση ενός χαρακτήρα σε έναν δύο διαστάσεων παραμετρικό χώρο.

- **Διάφορα μορφολογικά χαρακτηριστικά**

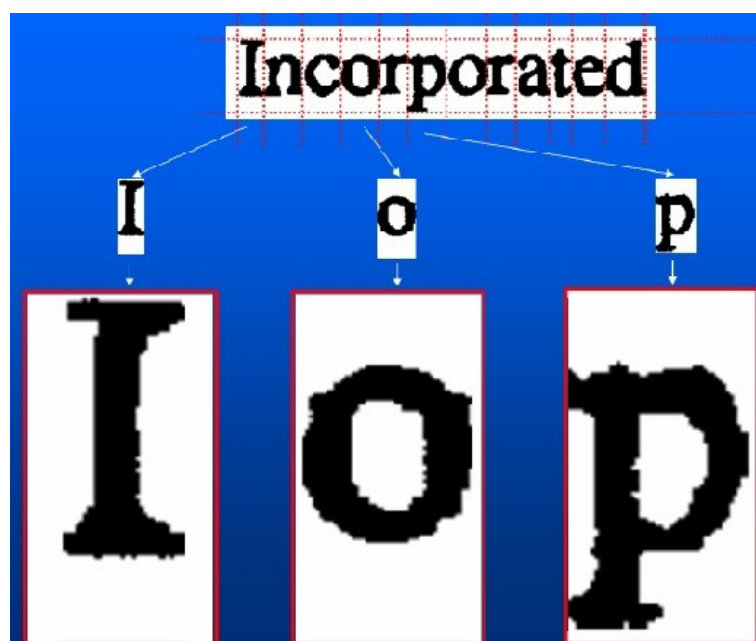
Οι περισσότερες και βασικότερες μέθοδοι εξαγωγής μορφολογικών χαρακτηριστικών, οι οποίες βασίζονται στην πληροφορία που υπάρχει στο συνολικό σώμα των χαρακτήρων, είναι οι μέθοδοι απευθείας σύγκρισης (template matching), ο χωρισμός σε ζώνες, οι γεωμετρικές ροπές και οι προβολές. Ακόμα από το σώμα των χαρακτήρων μπορούν να εξαχθούν χαρακτηριστικά που αφορούν το ποσοστό κάλυψης των μαύρων pixels, τον αριθμό των σημείων τομής με ευθείες (crossing points) καθώς και την ύπαρξη οπών.

Κανονικοποίηση των χαρακτήρων

Ένα βασικό στάδιο πριν την εξαγωγή των χαρακτηριστικών και την αναγνώριση ή κατηγοριοποίηση, είναι η κανονικοποίηση των χαρακτήρων που αποτελεί στάδιο προεπεξεργασίας. Το στάδιο αυτό είναι απαραίτητο γιατί θα πρέπει οι χαρακτήρες που θα συγκριθούν με τους χαρακτήρες της βάσης αναγνώρισης να έχουν το ίδιο μέγεθος. Συνήθως, γίνεται κανονικοποίηση των χαρακτήρων σε πίνακες των οποίων το ύψος είναι σχεδόν διπλάσιο από το πλάτος, π.χ. 24x48. Οι διαστάσεις του παραθύρου κανονικοποίησης εξαρτώνται και από τη μέθοδο εξαγωγής

χαρακτηριστικών που χρησιμοποιείται. Για παράδειγμα, αν χρησιμοποιήσουμε τον χωρισμό των χαρακτήρων σε παράθυρα διαστάσεων 5×10 , τότε το παράθυρο κανονικοποίησης θα πρέπει να έχει πλάτος πολλαπλάσιο του 5 και ύψος πολλαπλάσιο του 10, π.χ. 30×60 .

Για να προχωρήσουμε σε σωστή κανονικοποίηση κάθε χαρακτήρα θα πρέπει να χρησιμοποιήσουμε τις πληροφορίες τόσο για την οριοθέτηση του χαρακτήρα στην γραμμή του κειμένου όσο και για την οριοθέτηση της γραμμής του κειμένου στην συνολική εικόνα (Εικόνα 36). Το αριστερό και δεξί όριο του χαρακτήρα θεωρείται το όριο διαχωρισμού από το προηγούμενο και το επόμενο γράμμα. Το πάνω και κάτω όριο του χαρακτήρα θα πρέπει να είναι το πάνω και κάτω όριο της αντίστοιχης γραμμής κειμένου. Αυτό γίνεται για να διαχωριστούν τα κεφαλαία από τα μικρά γράμματα (π.χ. «ο» - «Ο») και οι χαρακτήρες που εξαρτώνται από την θέση τους στη γραμμή του κειμένου (π.χ. «.» - «'»). Εναλλακτικά, μπορούμε να κανονικοποιήσουμε κάθε χαρακτήρα στα όριά του, όμως θα πρέπει να αποθηκεύσουμε κάποια επιπλέον πληροφορία την οποία θα την χρησιμοποιήσουμε στην ταξινόμηση και αφορά το μέγεθος και τη θέση των χαρακτήρων ως προς τη γραμμή του κειμένου.



Εικόνα 37. Παράδειγμα κανονικοποίησης χαρακτήρων.

Απευθείας σύγκριση (template matching)

Η μεθοδολογία της απευθείας σύγκρισης (template matching) θεωρεί ότι τα pixels της εικόνας του χαρακτήρα είναι και το διάνυσμα των χαρακτηριστικών του²⁴. Αν θέλουμε να συγκρίνουμε δύο χαρακτήρες X, Y (X είναι το πρότυπο και Y ο χαρακτήρας που εξετάζουμε) οι οποίοι έχουν κανονικοποιηθεί στις ίδιες διαστάσεις και αποτελούνται από n σημεία, τότε ο αριθμός n_{ij} των σημείων, όπου ο χαρακτήρας X έχει τιμή i και ο χαρακτήρας Y έχει τιμή j , με $i, j \in \{0, 1\}$, δίδεται από τον τύπο:

$$n_{ij} = \sum_{m=1}^n \delta_m(i, j)$$

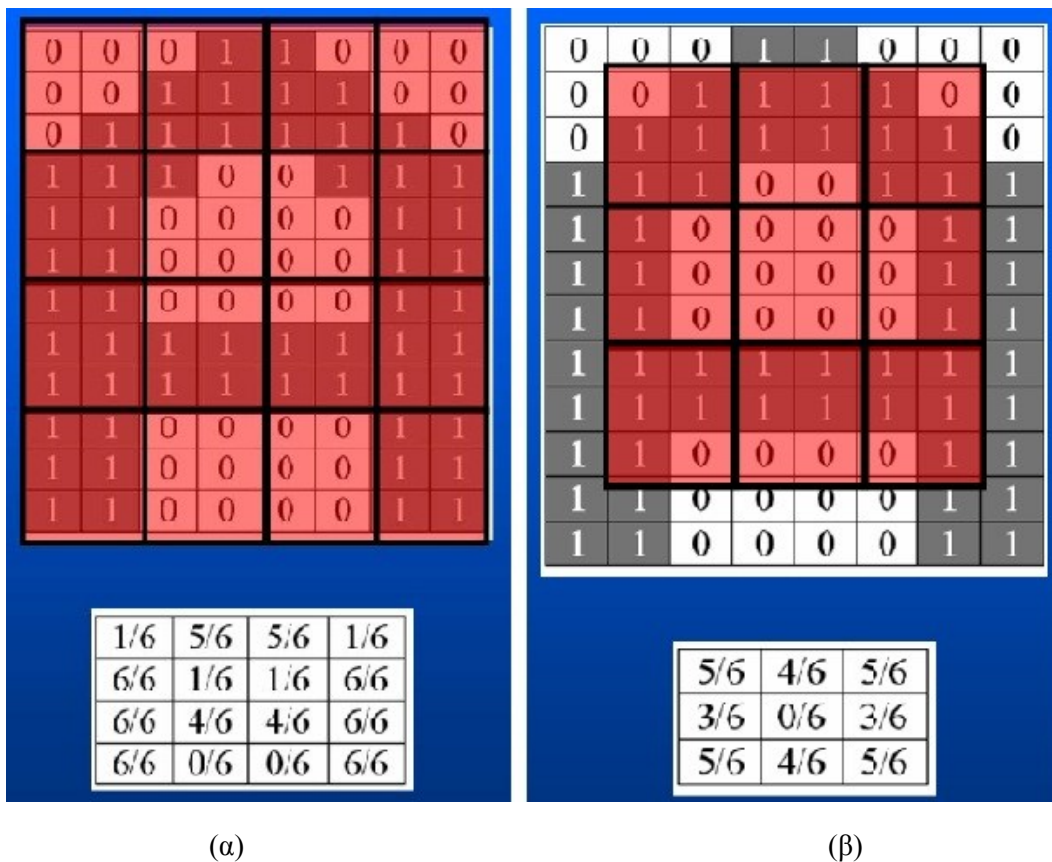
όπου:

$$\delta_m(i, j) = \begin{cases} 1, & \text{if } (x_m = i) \wedge (y_m = j) \\ 0, & \text{otherwise} \end{cases}$$

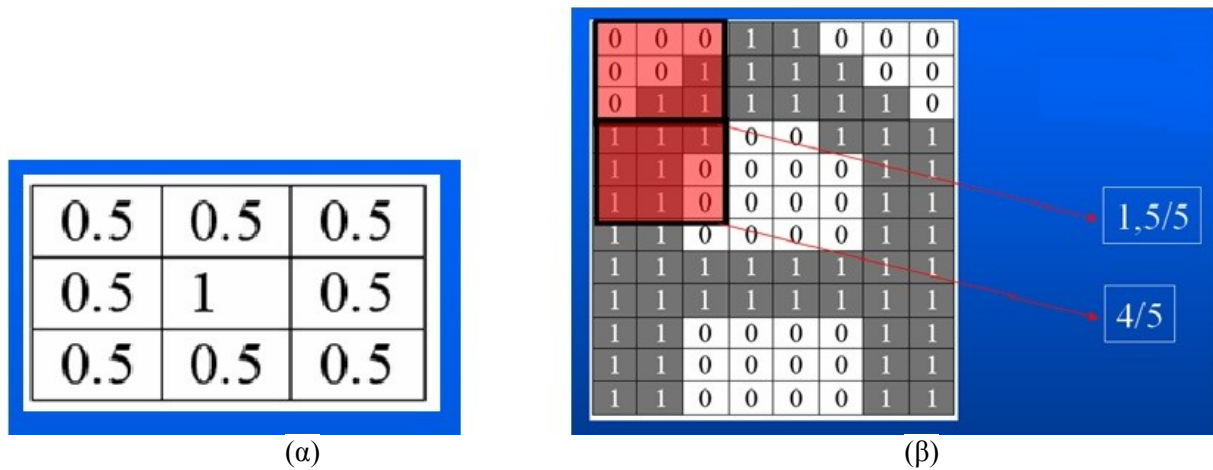
και x_m και y_m είναι τα m -ιστά σημεία των προς σύγκριση χαρακτήρων X και Y .

Χωρισμός σε ζώνες

Σύμφωνα με την προσέγγιση χωρισμού σε ζώνες, υπολογίζεται η κατανομή των μαύρων pixels σε διάφορες ζώνες στις οποίες χωρίζουμε τον χαρακτήρα²⁵. Για την εξαγωγή χαρακτηριστικών με χρήση των ζωνών του χαρακτήρα, η πιο απλή εφαρμογή είναι ο χωρισμός του χαρακτήρα σε $n \times m$ ζώνες και ο υπολογισμός της πυκνότητας των pixels σε κάθε ζώνη. Στο παράδειγμα της *Εικόνας 37* υπολογίζεται η πυκνότητα των pixels σε 16 ζώνες καθώς και σε 9 ζώνες που επικαλύπτουν τις πρώτες στις οποίες έχουμε χωρίσει τον χαρακτήρα. Εναλλακτικά, μπορούμε να δώσουμε βάρη στα pixels κάθε παραθύρου που έχουμε χωρίσει τον χαρακτήρα τονίζοντας περισσότερο τα pixels που βρίσκονται στο κέντρο του παραθύρου²⁶ (*Εικόνα 38*). Με αυτό τον τρόπο επιτυγχάνουμε μεγαλύτερη αντοχή στον θόρυβο.



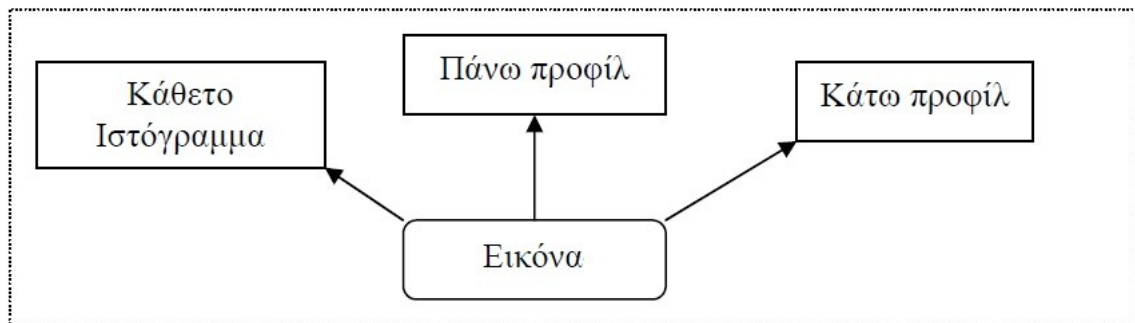
Εικόνα 38. Υπολογισμός της πυκνότητας των pixels σε ζώνες στις οποίες έχουμε χωρίσει τον χαρακτήρα. (α) 16 ζώνες. (β) 9 ζώνες που επικαλύπτουν τις πρώτες.



Εικόνα 39. Χρήση βαρών για τον υπολογισμό της πυκνότητας των pixels σε κάθε παράθυρο. (α) Τα βάρη σε κάθε παράθυρο. (β) Οι τιμές των χαρακτηριστικών που προκύπτουν.

Κάθετο ιστόγραμμα και άνω & κάτω προφίλ

Μια πολύ ενδιαφέρουσα προσέγγιση γίνεται στην εργασία της Ν. Δουλγέρη²⁷, η οποία για την οπτική αναγνώριση λέξης χρησιμοποιεί μόνο μορφολογικά χαρακτηριστικά, που μπορούν να λειτουργήσουν με απλούς αλγορίθμους κατηγοριοποίησης. Τα χαρακτηριστικά αυτά είναι το κάθετο ιστόγραμμα, το άνω προφίλ και το κάτω προφίλ.



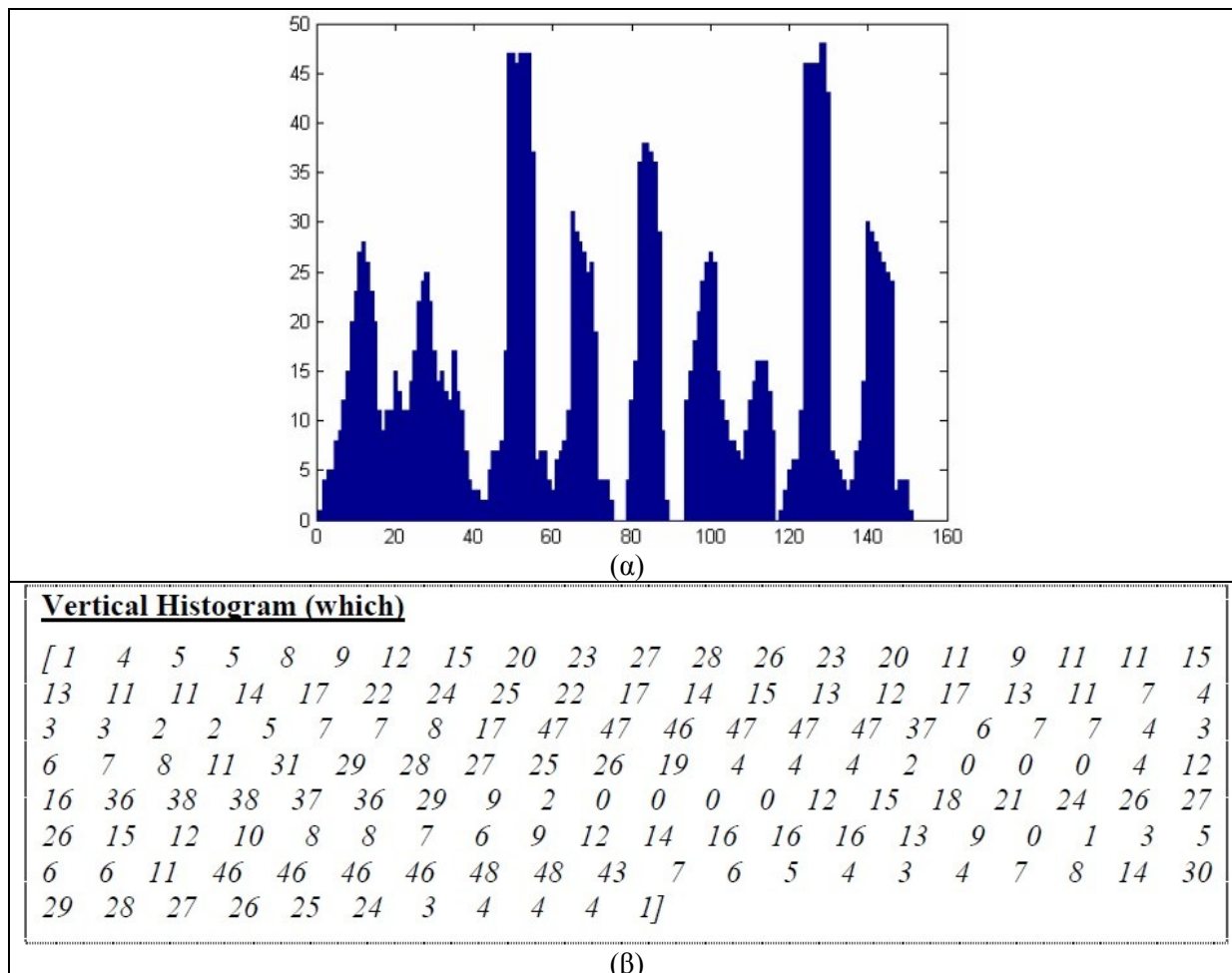
Εικόνα 40. Εξαγωγή χαρακτηριστικών εικόνας.

Κάθετο ιστόγραμμα

Αρχικά θεωρήθηκε αναγκαίο να υπάρχει μια ένδειξη του πλήθους των μαύρων pixels που υπάρχουν σε κάθε εικόνα-λέξη, που να τις διαφοροποιεί αρκετά μεταξύ τους. Αν θεωρήσουμε την εικόνα ως ένα πίνακα διαστάσεων $M \times N$, ως κάθετο ιστόγραμμα θεωρούμε το πλήθος των μαύρων pixels που συναντάμε σε κάθε στήλη της εικόνας, ενώ ως οριζόντιο ιστόγραμμα το πλήθος των μαύρων pixels που συναντάμε σε κάθε γραμμή της εικόνας.

Το κάθετο ιστόγραμμα προσφέρει ένα μέτρο ομοιότητας, αφού μας παρουσιάζει μια κατανομή του μελανιού σε στήλες. Με αυτό τον τρόπο έχουμε ένα δείγμα των ίδιων των γραμμμάτων στη λέξη, των ουρών πάνω και κάτω από το κύριο σώμα της λέξης, των τόνων που ίσως υπάρχουν και άλλων κοινών σημείων των λέξεων.

Από την εφαρμογή του κάθετου ιστογράμματος παίρνουμε ένα μονοδιάστατο διάνυσμα με μήκος (στήλες) ίσο με αυτό ολόκληρης της λέξης και τιμή για κάθε στήλη το πλήθος των μαύρων pixels της αντίστοιχης στήλης της εικόνας. Ένα παράδειγμα του παραγόμενου διανύσματος φαίνεται στην παρακάτω εικόνα.



Εικόνα 41. Παράδειγμα διανύσματος κάθετου ιστογράμματος. (α) Κάθετο ιστογράμμα. (β) Διάνυσμα ιστογράμματος.

Προφίλ

Τα προφίλ (*profiles*) είναι ένας βασικός τρόπος περιγραφής διαφόρων πραγμάτων γενικότερα. Για παράδειγμα το προφίλ ενός ανθρώπου μπορεί να περιλαμβάνει το όνομά του, την ηλικία του, μια περιγραφή της εμφάνισής του, όπως επίσης και κάποιων πνευματικών χαρακτηριστικών του. Έχοντας στην κατοχή μας τα προφίλ διαφόρων ανθρώπων μπορούμε να δούμε ποιοι μοιάζουν μεταξύ τους, ποιοι διαφέρουν, ακόμα κι αν μιλάμε για το ίδιο άτομο.

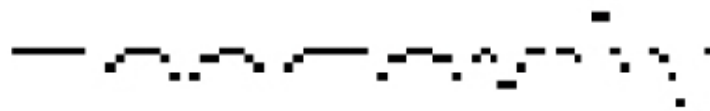
Μια μέθοδος οπτικής αναγνώρισης χαρακτήρων ή λέξεων εκμεταλλεύεται τις γενικές ιδιότητες των προφίλ τα οποία χρησιμοποιεί για την αναγνώρισή τους. Αποκτούμε έτσι μια γενική αναπαράσταση του σχήματος των χαρακτήρων ή των λέξεων, έτοιμη να συγκριθεί με άλλες. Βέβαια στην βιβλιογραφία συναντάμε διάφορα είδη προφίλ, τα οποία χρησιμοποιούνται ανάλογα με τις εκάστοτε ανάγκες.

Θεωρώντας όπως προηγουμένως ότι η εικόνα είναι ένας πίνακας διαστάσεων $M \times N$ μπορούμε να εξάγουμε τους παρακάτω ορισμούς:

- Πάνω προφίλ ορίζεται ως η θέση του πρώτου μαύρου pixel που συναντάμε σε κάθε στήλη της εικόνας, ξεκινώντας από την πρώτη γραμμή και διατρέχοντας την εικόνα προς τα κάτω.
- Κάτω προφίλ ορίζεται ως η θέση του πρώτου μαύρου pixel που συναντάμε σε κάθε στήλη της εικόνας, ξεκινώντας από την τελευταία γραμμή και διατρέχοντας την εικόνα προς τα πάνω.
- Αριστερό προφίλ ορίζεται ως η θέση του πρώτου μαύρου pixel που συναντάμε σε κάθε γραμμή της εικόνας, ξεκινώντας από την πρώτη στήλη και διατρέχοντας την εικόνα προς τα δεξιά.
- Δεξιό προφίλ ορίζεται ως η θέση του πρώτου μαύρου pixel που συναντάμε σε κάθε γραμμή της εικόνας, ξεκινώντας από την τελευταία στήλη και διατρέχοντας την εικόνα προς τα αριστερά.



(α) Δέξη



(β) Πάνω Προφίλ



(γ) Κάτω Προφίλ

Εικόνα 42. Πάνω και κάτω προφίλ ολόκληρης λέξης.

Τα προφίλ χρησιμοποιήθηκαν αρκετά στην αναγνώριση μεμονωμένων χαρακτήρων, όπως στους E.Kavallieratou et al.²⁸ και Heute²⁹ τα οποία χρησιμοποιήθηκαν και σε ολόκληρες λέξεις^{30 31}. Το σίγουρο πάντως είναι πως αποτελούν μια βασική περιγραφή του σχήματος είτε μιας ολόκληρης λέξης, είτε ενός χαρακτήρα γι' αυτό και επιλέχθηκαν ως βασικά μορφολογικά χαρακτηριστικά στο εν λόγω σύστημα.

3.4. Αναγνώριση – Κατηγοριοποίηση

Τελικά, η διαδικασία οπτικής αναγνώρισης χαρακτήρων κατατάσσει μια εικόνα χαρακτήρα σε μια κατηγορία με τη χρησιμοποίηση ενός αλγορίθμου ταξινόμησης βασισμένου στα χαρακτηριστικά γνωρίσματα που εξάγονται και τις σχέσεις μεταξύ των χαρακτηριστικών γνωρισμάτων. Δεδομένου ότι τα μέλη μιας κατηγορίας χαρακτήρα είναι ισοδύναμα ή παρόμοια από τη στιγμή που μοιράζονται καθοριστικές ιδιότητες, η μέτρηση της ομοιότητας είναι σημαντική σε οποιοδήποτε ταξινομητή.

Γενικά η κατηγοριοποίηση περιλαμβάνει τον διαχωρισμό ενός συνόλου σημείων δεδομένων (data points) σε μη επικαλυπτόμενες κατηγορίες ή κλάσεις (classes), όπου τα σημεία σε μια κλάση είναι «πιο κοντά» το ένα στο άλλο από τα σημεία σε άλλες κλάσεις. Με τον όρο «πιο κοντά» όταν αναφερόμαστε σε σημεία κλάσεων εννοούμε ότι βάση ενός μέτρου προσέγγισης είναι πιο κοντά. Όταν ένα σύνολο δεδομένων κατηγοριοποιείται, κάθε σημείο ανατίθεται σε μια κλάση και κάθε κλάση μπορεί να χαρακτηριστεί με ένα μοναδικό σημείο αναφοράς (συνήθως ο μέσος όρος των σημείων που περιέχονται στην κλάση), που λέγεται κεντρικό σημείο (centroid).

Το αν μια μέθοδος θεωρείται καλή εξαρτάται από τα χαρακτηριστικά της ίδιας της εφαρμογής στην οποία πρόκειται να χρησιμοποιηθεί. Στη βιβλιογραφία συναντάμε πολλές μεθόδους για την εύρεση κλάσεων (classes) βασισμένων σε διάφορα κριτήρια, είτε τυχαίες είτε συστηματικές. Έτσι συναντάμε προσεγγίσεις που στηρίζονται στις διαδικασίες του διαχωρισμού και της συνένωσης (ISODATA), κάποιες τυχαίες προσεγγίσεις (CLARA, CLARANS), μεθόδους που σχετίζονται με νευρωνικά δίκτυα καθώς και πολλές άλλες τεχνικές³² όπως τεχνικές template matching (αναφορά σε αυτό το είδος μεθόδων έγινε και στο προηγούμενο υποκεφάλαιο 3.3.3), τεχνικές κοντινότερου γείτονα, δυαδικά δέντρα απόφασης κ.ά.

Τι όμως είναι αυτό που κάνει μια κατηγοριοποίηση να θεωρείται «καλή»; Ας θεωρήσουμε μια μοναδική κλάση σημείων μαζί με το κεντρικό σημείο της ή μέσο. Αν τα υπόλοιπα σημεία βρίσκονται πολύ κοντά στο κεντρικό σημείο, τότε αυτό θα είναι και αντιπροσωπευτικό όλων των σημείων της κλάσης. Το βασικό μέτρο μέτρησης της διασποράς των σημείων σε μια ομάδα γύρω από το μέσο είναι συνήθως το άθροισμα των τετραγώνων της απόστασης ανάμεσα σε κάθε σημείο και το μέσο. Αν τα σημεία είναι κοντά στο μέσο, η διασπορά θα είναι μικρή. Μια γενίκευση αυτής της διασποράς, όπου το κεντρικό σημείο αντικαθίσταται από ένα σημείο αναφοράς, είτε είναι είτε όχι το κεντρικό, χρησιμοποιείται στην ανάλυση κλάσεων (classes analysis) και το άθροισμα όλων των διασπορών αποτελεί ένα μέτρο λάθους (error measure) το E , το οποίο ουσιαστικά προσφέρει μια αντικειμενική μέθοδο σύγκρισης των διαφόρων μεθόδων κατηγοριοποίησης³³.

3.4.1. Template matching

Απευθείας σύγκριση

Ένας χαρακτήρας εισόδου συγκρίνεται άμεσα με ένα πρότυπο σύνολο καταχωρημένων πρωτοτύπων και σύμφωνα με ένα μέτρο ομοιότητας γίνεται ταίριασμα πρωτοτύπων για την αναγνώριση. Οι τεχνικές αντιστοίχισης μπορούν να είναι απλές, όπως η ένα προς ένα σύγκριση, ή σύνθετες όπως η ανάλυση δέντρων απόφασης στην οποία μόνο τα επιλεγμένα εικονοκύτταρα εξετάζονται. Αν και η άμεση αντιστοίχιση έχει σταθερές μαθηματικές βάσεις, το ποσοστό αναγνώρισης αυτής της μεθόδου είναι πολύ ευαίσθητο στον θόρυβο.

Παραμορφώσιμα πρωτότυπα και ελαστική αντιστοίχιση

Μια εναλλακτική μέθοδος είναι η χρήση των παραμορφώσιμων πρωτοτύπων, όπου μια παραμόρφωση εικόνας χρησιμοποιείται για να ταιριάζει με μια άγνωστη εικόνα με μια βάση δεδομένων των γνωστών εικόνων. Ένα μέτρο ανομοιότητας προκύπτει από το ποσό παραμόρφωσης που απαιτείται.

Χαλαρή αντιστοίχιση

Είναι μια τεχνική αντιστοίχισης εικόνας συμβολικού επιπέδου που χρησιμοποιεί μία χαρακτηριστική περιγραφή για την εικόνα χαρακτήρα. Καταρχάς, προσδιορίζονται οι περιοχές αντιστοίχισης. Κατόπιν, τα στοιχεία εικόνας συγκρίνονται με το μοντέλο με βάση μερικές καθορισμένες εκτιμήσεις των αναθέσεων. Αυτή η διαδικασία απαιτεί μια τεχνική αναζήτησης σε ένα πολυδιάστατο διάστημα για την εύρεση του σφαιρικού μεγίστου μερικών λειτουργιών.

3.4.2. Στατιστικές τεχνικές

Στις προσεγγίσεις στατιστικής ταξινόμησης, τα σχέδια εικόνας χαρακτήρα αντιπροσωπεύονται από σημεία σε ένα πολυδιάστατο χώρο χαρακτηριστικών γνωρισμάτων. Κάθε συστατικό του χώρου χαρακτηριστικών γνωρισμάτων είναι μια αξία μέτρησης, η οποία λειτουργεί ως μια τυχαία μεταβλητή που απεικονίζει την έμφυτη μεταβλητότητα ανάμεσα στις κατηγορίες. Ένας ταξινομητής χωρίζει το χώρο χαρακτηριστικών γνωρισμάτων σε περιοχές που συνδέονται με κάθε κατηγορία, αντιστοιχίζοντας το σχέδιο που παρατηρούμε με κάποιο χαρακτήρα σύμφωνα με την περιοχή κατηγορίας στην οποία περιέρχεται³⁴. Οι σημαντικότεροι τρόποι προσέγγισης είναι οι ακόλουθοι:

Μη-παραμετρική αναγνώριση

Αυτή η μέθοδος χρησιμοποιείται για το διαχωρισμό, σε διαφορετικές κλάσεις προτύπων. Η πιο γνωστή μέθοδος non-parametric ταξινόμησης είναι ο κοντινότερος γείτονας (*NN*) και χρησιμοποιείται εκτενώς στην αναγνώριση χαρακτήρα. Δεν προϋποθέτει *a priori* πληροφορίες για τα στοιχεία. Ένα εισερχόμενο πρότυπο ταξινομείται χρησιμοποιώντας τον τομέα (*cluster*), του οποίου το κέντρο έχει την ελάχιστη απόσταση από το πρότυπο σε σχέση με όλους τους άλλους τομείς.

Παραμετρική αναγνώριση

Δεδομένου ότι οι *a priori* πληροφορίες είναι διαθέσιμες για τους χαρακτήρες στα δεδομένα εκπαίδευσης (*training data*), είναι δυνατό να ληφθεί ένα παραμετρικό μοντέλο για κάθε χαρακτήρα. Από τη στιγμή που οι παράμετροι του μοντέλου, που είναι βασισμένο σε μερικές πιθανότητες, λαμβάνονται, οι χαρακτήρες ταξινομούνται σύμφωνα με μερικούς κανόνες απόφασης όπως η μέγιστη πιθανότητα ή η μέθοδος Bayes (*Bayesian method*).

Ανάλυση Ομαδοποίησης (*clustering*) Οι τομείς των χαρακτηριστικών γνωρισμάτων χαρακτήρα, που αντιπροσωπεύουν τις ευδιάκριτες κλάσεις, αναλύονται με μεθόδους ομαδοποίησης. Η διαδικασία ομαδοποίησης είναι η εξής:

- Επιλογή του μέτρου ομοιότητας ή ανομοιότητας
- Επιλογή του κριτηρίου ομαδοποίησης
- Ορισμός του αλγορίθμου βελτιστοποίησης.

Η ομαδοποίηση μπορεί να πραγματοποιηθεί είτε με συσσωρευτικούς είτε διαχωριστικούς αλγορίθμους. Οι συσσωρευτικοί αλγόριθμοι λειτουργούν βαθμιαία και συγχωνεύουν τους μικρούς τομείς σε μεγαλύτερους με βάση το κριτήριο ομαδοποίησης., ενώ οι διαχωριστικές μέθοδοι χωρίζουν τις κλάσεις χαρακτήρων.

k-means Clustering

Μια από τις πιο διαδεδομένες μεθόδους κατηγοριοποίησης, που βασίζεται στην ελαχιστοποίηση μιας αντικειμενικής συνάρτησης είναι η k-Means κατηγοριοποίηση (k-Means classification). Προτάθηκε το 1967 από τον J. MacQueen ως μια μη-επιβλεπόμενη τεχνική κατηγοριοποίησης. Δοσμένου ενός συνόλου n σημείων δεδομένων σε έναν χώρο d -διαστάσεων, R^d , και ενός ακεραίου k , το πρόβλημα είναι να καθοριστεί ένα σύνολο k σημείων στο R^d , που καλούνται κεντρικά σημεία (centroids), τέτοια ώστε να ελαχιστοποιηθεί η μέση τετράγωνη απόσταση από κάθε δεδομένο στο κοντινότερό του κέντρο.

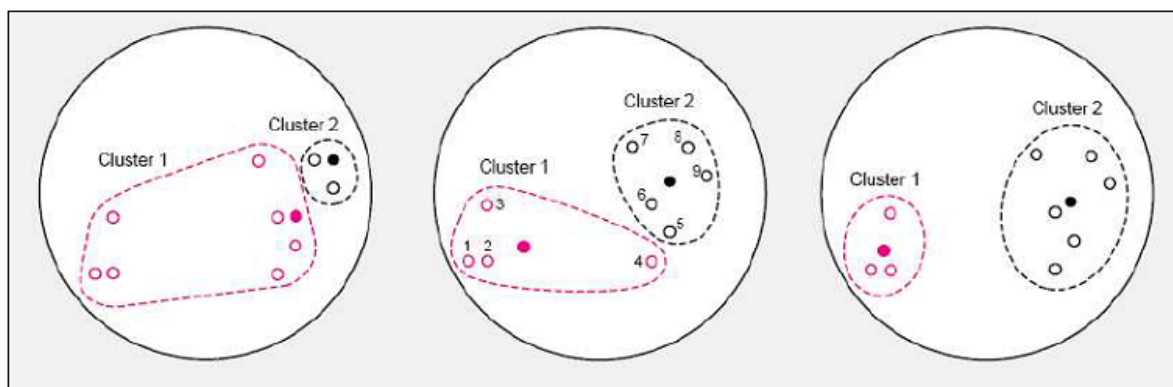
Για την επίλυση του k-mean προβλήματος υπάρχει ένας ευριστικός αλγόριθμος που βασίζεται σε ένα απλό επαναληπτικό σχήμα για την εύρεση μιας τοπικά ελάχιστης λύσης, που καλείται k-mean αλγόριθμος, με πλήθος παραλλαγών³⁵. Μια γενική περιγραφή του αλγορίθμου φαίνεται στην *Εικόνα 42* και ένα παράδειγμα κατηγοριοποίησης με τη χρήση του k-Means στην *Εικόνα 43*. Πάντως ο k-mean αλγόριθμος, ανεξάρτητα παραλλαγής επειδή συνεχώς ανανεώνει τις κλάσεις, δεν απαιτεί πολλές επαναλήψεις και γενικά θεωρείται αρκετά γρήγορος αλγόριθμος.

k-Means Αλγόριθμος

1. Όρισε το πλήθος των κλάσεων
2. Αρχικοποίησε τις κλάσεις με:
 - a. Τυχαία κατανομή παραδειγμάτων στις κλάσεις
 - Η
 - b. Τυχαία επιλογή κέντρων κλάσεων
3. Υπολόγισε το μέσο κάθε κλάσης
4. Απέδωσε κάθε δείγμα στο πλησιέστερο μέσο
5. Αν η κατανομή των δειγμάτων δεν άλλαξε τερμάτισε αλλιώς πήγαινε στο βήμα 3

Εικόνα 43. Ο αλγόριθμος k-means

<p>α) Αρχικοποίηση Σημείο Αναφοράς 1 (κόκκινος κύκλος με γέμισμα) και Σημείο Αναφοράς 2 (μαύρος κύκλος με γέμισμα) Κάθε δεδομένο συμμετέχει σε μια κλάση, με βάση την απόσταση από τα σημεία αναφοράς</p>	<p>β) Πρώτη Επανάληψη Σημεία Αναφοράς γίνεται τα κεντρικά σημεία κάθε κλάσης και τα υπόλοιπα δεδομένα ταξινομούνται στις δυο κλάσεις με βάση την απόστασή τους από τα νέα σημεία αναφοράς</p>	<p>γ) Δεύτερη Επανάληψη Το προηγούμενο βήμα εκτελείται ξανά και προκύπτει σταθερή κατηγοριοποίηση</p>
--	---	---



Εικόνα 44. Παράδειγμα κατηγοριοποίησης με χρήση του *k-means*.

3.4.3. Δομικές τεχνικές

Η επαναλαμβανόμενη περιγραφή ενός σύνθετου προτύπου, μέσω απλούστερων βασισμένων στη μορφή του αντικειμένου, ήταν η αρχική ιδέα για τη δημιουργία της δομικής αναγνώρισης προτύπων. Αυτά τα πρότυπα χρησιμοποιούνται για να περιγράψουν και να ταξινομήσουν τους χαρακτήρες στα συστήματα αναγνώρισης χαρακτήρα. Οι χαρακτήρες απεικονίζονται ως η ένωση των δομικών πρωτεύοντων στοιχείων. Τα πρωτεύοντα στοιχεία χαρακτήρα που εξάγονται από το γράμμα πρέπει είναι ποσοτικά προσδιορίσιμα και συσχετίσιμα. Οι ακόλουθες δομικές μέθοδοι εφαρμόζονται στα προβλήματα αναγνώρισης χαρακτήρα.

Γραμματικές μέθοδοι

Οι γραμματικές μέθοδοι δημιουργούν μερικούς κανόνες παραγωγής προκειμένου να διαμορφωθούν μέσω των επίσημων γραμματικών οι χαρακτήρες από ένα σύνολο πρωτόγονων. Αυτές οι μέθοδοι μπορούν να συνδυάσουν οποιοδήποτε τύπο τοπολογικών και στατιστικών χαρακτηριστικών γνωρισμάτων σύμφωνα με μερικούς συντακτικούς ή/και σημασιολογικούς κανόνες.

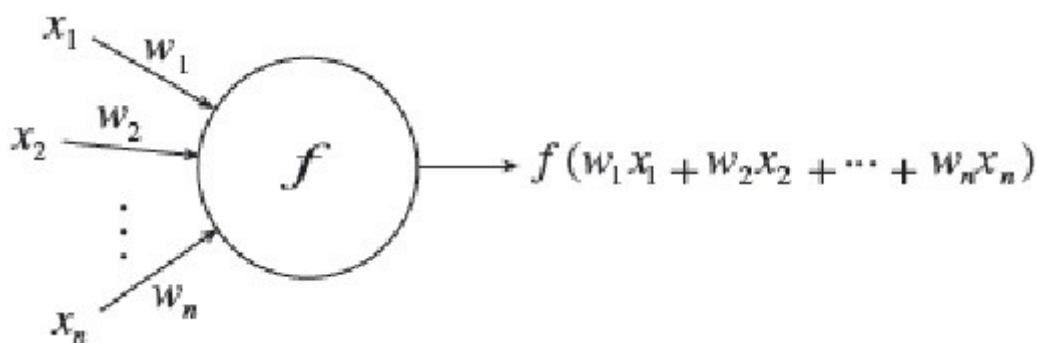
Γραφικές μέθοδοι

Οι γραφικές μονάδες αντιπροσωπεύονται από τα δέντρα, τους γράφους ή τις γραφικές παραστάσεις. Τα πρωτεύοντα στοιχεία χαρακτήρα (π.χ. καμπύλες) επιλέγονται από μια δομική

προσέγγιση, ανεξάρτητα από το πώς γίνεται η λήψη τελικής απόφασης στην αναγνώριση. Για κάθε κλάση, ένας γράφος ή ένα δέντρο διαμορφώνεται στο στάδιο κατάρτισης για να αντιπροσωπεύσει τις καμπύλες, τα γράμματα ή τις λέξεις. Το στάδιο αναγνώρισης κατατάσσει την άγνωστη γραφική παράσταση σε μια από τις κλάσεις με τη χρησιμοποίηση ενός μέτρου ομοιότητας.

3.4.4. Νευρωνικά δίκτυα

Ένα τεχνητό νευρωνικό δίκτυο (ΤΝΔ) είναι ένα μοντέλο επεξεργασίας πληροφοριών που εμπνέεται από τον τρόπο που επεξεργάζονται την πληροφορία τα βιολογικά νευρωνικά συστήματα, όπως ο εγκέφαλος. Το βασικό στοιχείο αυτού του μοντέλου είναι η δομή του συστήματος επεξεργασίας πληροφοριών, αποτελείται από τη μαζικά παράλληλη διασύνδεση των προσαρμοστικών "νευρωνικών" επεξεργαστών, νευρώνων (Εικόνα 44).



Εικόνα 45. Η δομική μονάδα ενός νευρωνικού δικτύου, ένας νευρώνας.

Κάθε είσοδος του νευρώνα είναι πολλαπλασιασμένη με ένα μέγεθος που ονομάζεται βάρος. Η f είναι μία προεπιλεγμένη συνάρτηση που μας υπολογίζει την έξοδο του νευρώνα. Λόγω της παράλληλης φύσης της δομής ενός ΤΝΔ, αυτό μπορεί να εκτελέσει τους υπολογισμούς ταχύτερα από τις κλασσικές τεχνικές. Ένα ΤΝΔ διαμορφώνεται για μια συγκεκριμένη εφαρμογή, όπως η αναγνώριση προτύπων ή η ταξινόμηση στοιχείων, μέσω μιας διαδικασίας εκμάθησης. Λόγω της προσαρμοστικής φύσης του, μπορεί να προσαρμοστεί στις αλλαγές των στοιχείων και να μάθει τα χαρακτηριστικά του σήματος εισόδου.

Ένα νευρωνικό δίκτυο περιέχει πολλούς κόμβους. Η έξοδος από έναν κόμβο γίνεται είσοδος σε έναν άλλο στο δίκτυο και η τελική απόφαση εξαρτάται από τη σύνθετη αλληλεπίδραση όλων των κόμβων.

Οι γενικές μαθηματικές σχέσεις που συμβολίζουν τη λειτουργία ενός νευρωνικού δικτύου είναι οι εξής:

$$Y=f(x)$$
$$f(x) = K \left(\sum_i w_i g_i(x) \right)$$

όπου:

- K η συνάρτηση ενεργοποίησης
- Y η έξοδος του νευρωνικού δικτύου
- g_i η έξοδος κάθε νευρώνα
- w_i τα βάρη των νευρώνων

Πλεονεκτήματα Νευρωνικών δικτύων

- Ένα νευρωνικό δίκτυο μπορεί να εκτελέσει εργασίες που ένα γραμμικό πρόγραμμα δεν μπορεί.
- Όταν ένα στοιχείο του νευρωνικού δικτύου αποτυγχάνει, η διαδικασία μπορεί να συνεχιστεί χωρίς πρόβλημα λόγω της παράλληλης διάταξης των νευρώνων.
- Ένα νευρωνικό δίκτυο μαθαίνει και δεν χρειάζεται να επαναπρογραμματιστεί.
- Μπορεί να εφαρμοστεί σε οποιαδήποτε εφαρμογή.

Μειονεκτήματα Νευρωνικών δικτύων

- Το δίκτυο είναι αναγκαίο να εκπαιδευτεί για να λειτουργήσει.
- Η αρχιτεκτονική ενός νευρωνικού δικτύου είναι διαφορετική από την αρχιτεκτονική των μικροεπεξεργαστών επομένως πρέπει να εξομοιωθεί.
- Απαιτείται υψηλός χρόνος επεξεργασίας για τα μεγάλα νευρωνικά δίκτυα.

Τοπολογία δικτύων

Οι νευρωνικές δικτυακές αρχιτεκτονικές μπορούν να ταξινομηθούν σε δύο σημαντικές ομάδες: feed-forward δίκτυα και ανατροφοδότησης.

- Στα feed-forward δίκτυα η ροή δεδομένων γίνεται αυστηρά μόνο με κατεύθυνση από την είσοδο προς την έξοδο.
- Στα δίκτυα ανατροφοδότησης η ροή μπορεί να γίνεται και αντίστροφα, με ανατροφοδότηση της εξόδου στην είσοδο στο ίδιο layer η και προηγούμενο.

Τα πιο κοινά νευρωνικά δίκτυα που χρησιμοποιούνται στα συστήματα αναγνώρισης χαρακτήρα είναι το πολυστρωματικό perceptron των feed-forward δικτύων και ο αυτοοργανωτικός χάρτης του Kohonen (SOM) των δικτύων ανατροφοδότησης.

Εκπαίδευση

Διάφορες προσεγγίσεις υπάρχουν για την εκπαίδευση των νευρωνικών δικτύων³⁶. Αυτές περιλαμβάνουν τη δοκιμή και σφάλμα, την πίσω διάδοση, τη μηχανή Boltzman, το δίκτυο Hebbian και την ανταγωνιστική εκμάθηση, οι οποίες καλύπτουν τη δυαδική και συνεχή εκτιμημένη είσοδο, καθώς επίσης και την εποπτευμένη και άνευ επίβλεψης εκμάθηση. Μπορούμε επίσης να προγραμματίσουμε ένα νευρωνικό δίκτυο ώστε να πραγματοποιήσει μια διεργασία δημιουργώντας χειροκίνητα την τοπολογία του και θέτοντας τα βάρη των νευρώνων. Όμως έτσι, παραμερίζουμε τη θεμελιώδη ιδιότητα των νευρωνικών δικτύων η οποία είναι η δυνατότητα να προγραμματίζονται μόνα τους.

Εκπαίδευση με επίβλεψη

Η φιλοσοφία αυτής της μεθόδου είναι πως για κάθε συγκεκριμένη είσοδο στο νευρωνικό δίκτυο εισάγουμε εμείς εκ των προτέρων μια επιθυμητή έξοδο. Με αυτό το τρόπο το νευρωνικό δίκτυο υπολογίζει το σφάλμα μεταξύ πραγματικής και επιθυμητής εξόδου και αλλάζοντας τα βάρη σύμφωνα με ένα καθορισμένο αλγόριθμο επιχειρεί να μειώσει αυτό το σφάλμα και να βρει τις τιμές βαρών που το καθιστούν ελάχιστο, με άλλα λόγια εκπαιδεύεται.

- **Δοκιμή και σφάλμα**

Η πιο βασική μέθοδος της εκπαίδευσης ενός νευρωνικού δικτύου είναι η δοκιμή και σφάλμα. Αν το δίκτυο δεν συμπεριφέρεται με τον τρόπο που πρέπει, αλλάζουμε το βάρος ενός τυχαίου συνδέσμου κατά ένα τυχαίο ποσό. Εάν η ακρίβεια του δικτύου μειώνεται, αναιρούμε τις αλλαγές και δοκιμάζουμε άλλες. Χρειάζεται χρόνος, αλλά η μέθοδος αυτή παράγει αποτελέσματα.

- **Πίσω-διάδοση**

Ο αλγόριθμος πίσω-διάδοσης συγκρίνει τα αποτελέσματα που λαμβάνονται με το αποτέλεσμα που αναμενόταν. Στη συνέχεια, χρησιμοποιεί αυτές τις πληροφορίες για να τροποποιήσει συστηματικά τα βάρη όλου του νευρωνικού δικτύου. Η εκπαίδευση αυτή διαρκεί μόνο ένα κλάσμα του χρόνου που απαιτεί η μέθοδος δοκιμή και σφάλμα. Μπορεί επίσης να χρησιμοποιηθεί αξιόπιστα για την κατάρτιση των δικτύων με ένα μόνο μέρος των δεδομένων. Τα δίκτυα που προκύπτουν είναι συνήθως σωστά και μπορούν να λύνουν προβλήματα για τα οποία δεν έχουν ποτέ εκπαιδευτεί ειδικά.

- **Εκπαίδευση χωρίς επίβλεψη**

Σε αυτή τη μέθοδο το σύστημα προσπαθεί μόνο του να οργανώσει και να ομαδοποιήσει τα δεδομένα αναγνωρίζοντας τις ιδιότητές τους. Συνήθως, πραγματοποιείται κάποιου είδους προεπεξεργασία όπως το data mining ώστε να προκύψουν οι ιδιότητες των δεδομένων. Παραδείγματα τέτοιων μεθόδων είναι η ομαδοποίηση και η μείωση διαστάσεων.

- **Ομαδοποίηση (clustering)**

Ομαδοποίηση ονομάζεται η ταξινόμηση των δεδομένων σε υποομάδες σύμφωνα με κάποιο στοιχείο ομοιότητας μεταξύ τους. Για την πραγματοποίησή της, το σύστημα πρέπει να εξαγάγει με κάποιο αλγόριθμο την ιδιότητα των δεδομένων προς σύγκριση. Έπειτα η ομοιότητα υπολογίζεται σύμφωνα με κάποια απόσταση την οποία έχουμε ορίσει.

- **Μείωση διαστάσεων(dimensionality reduction)**

Με αυτή τη μέθοδο μετασχηματίζουμε τα δεδομένα ώστε να περάσουμε από ένα χώρο με πολλές διαστάσεις σε κάποιον με λιγότερες. Σκοπός μας είναι να διατηρήσουμε όσο το δυνατόν πιο σταθερή τη διασπορά των δεδομένων ώστε να παραμείνουν σταθερές οι μεταξύ τους σχετικές αποστάσεις.

4. Πειραματικό μέρος

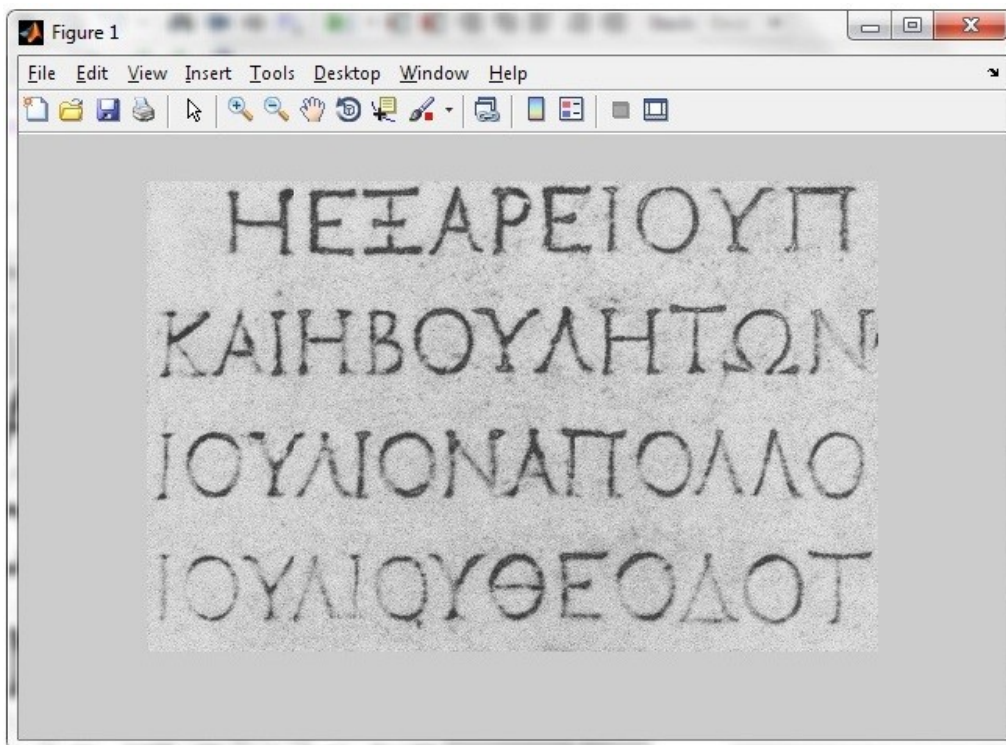
Στο σημείο αυτό θα παρουσιάσουμε τα πειράματα που έχουμε κάνει στη σουίτα MATLAB 7.6 σε σύστημα με επεξεργαστή quad core Intel i7 1, 6GHz, 4GB RAM και λειτουργικό Windows 7.

Τα βήματα που θα ακολουθήσουμε είναι αυτά που παρουσιάσαμε στο προηγούμενο κεφάλαιο, δηλαδή, προεπεξεργασία, κατάτμηση χαρακτήρων, εξαγωγή χαρακτηριστικών και τέλος κατηγοριοποίηση και αναγνώριση.

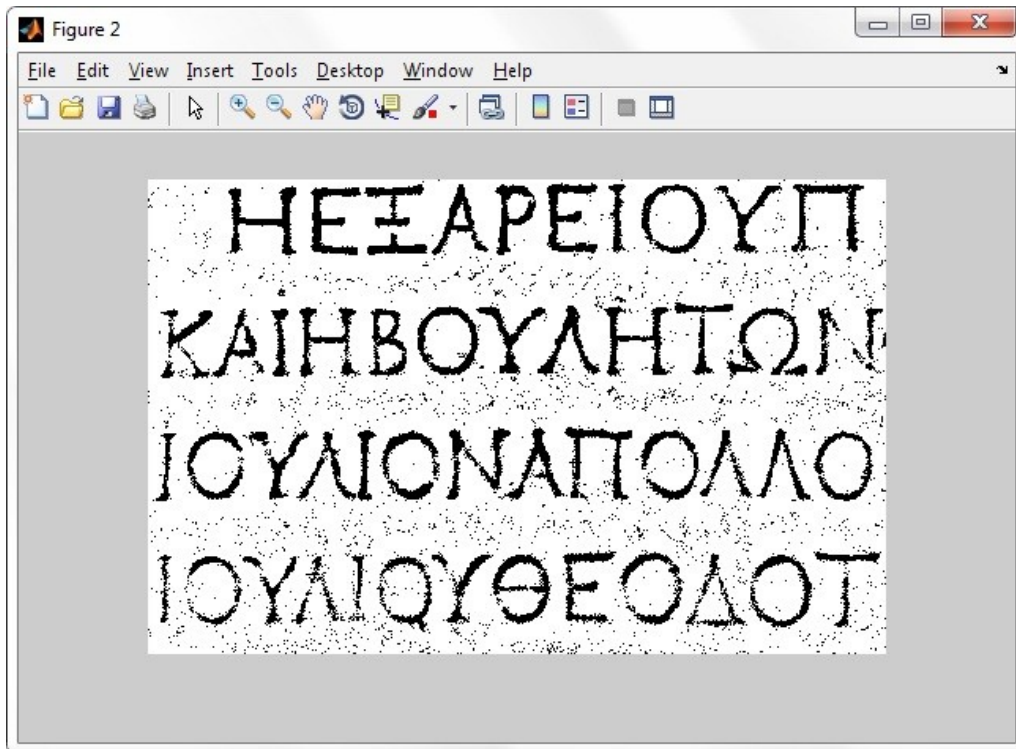
4.1. Προεπεξεργασία

Επιλέγουμε για το στάδιο της δυαδικής μετατροπής προσαρμοζόμενη, τοπική καταφλίωση (adaptive thresholding) με μέγεθος παραθύρου 30. Το κατώφλι επιλέχθηκε μη-αυτόματα μετά από δοκιμή για βελτιστοποίηση του αποτελέσματος.

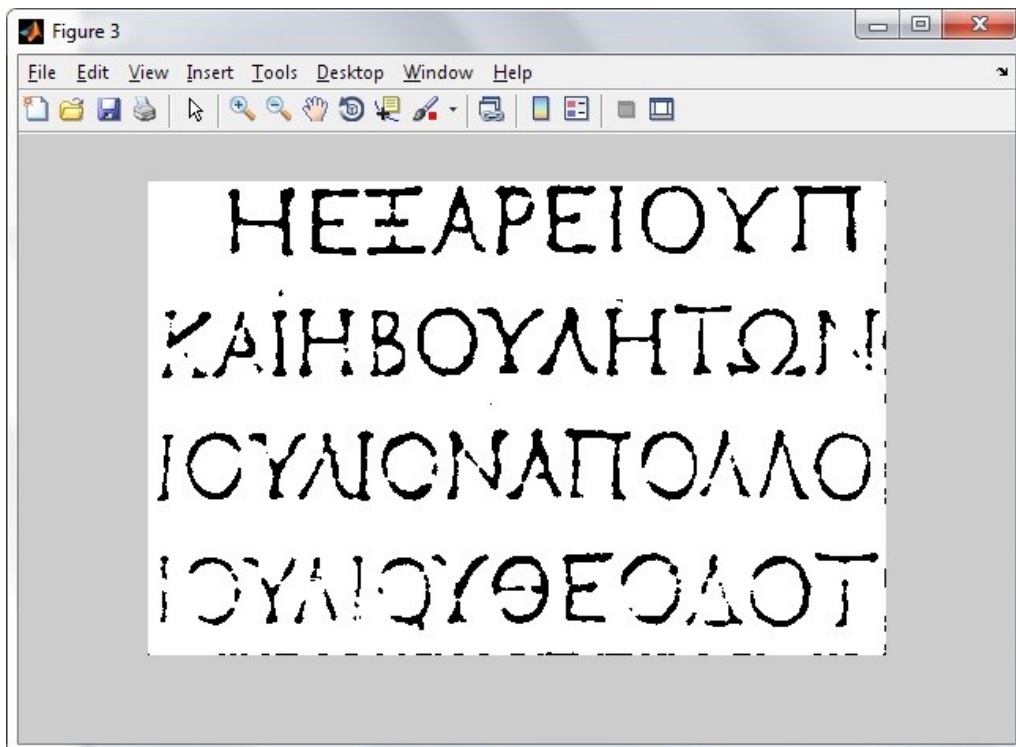
Στη συνέχεια χρησιμοποιήσαμε φίλτρο μέσου (median filter) για την απομάκρυνση του θορύβου μορφής αλατοπίπερου. Η υλοποίηση του φίλτρου έγινε με τη συνάρτηση του MATLAB 'medfilt2' και μέγεθος μάσκας 4x4.



(α)



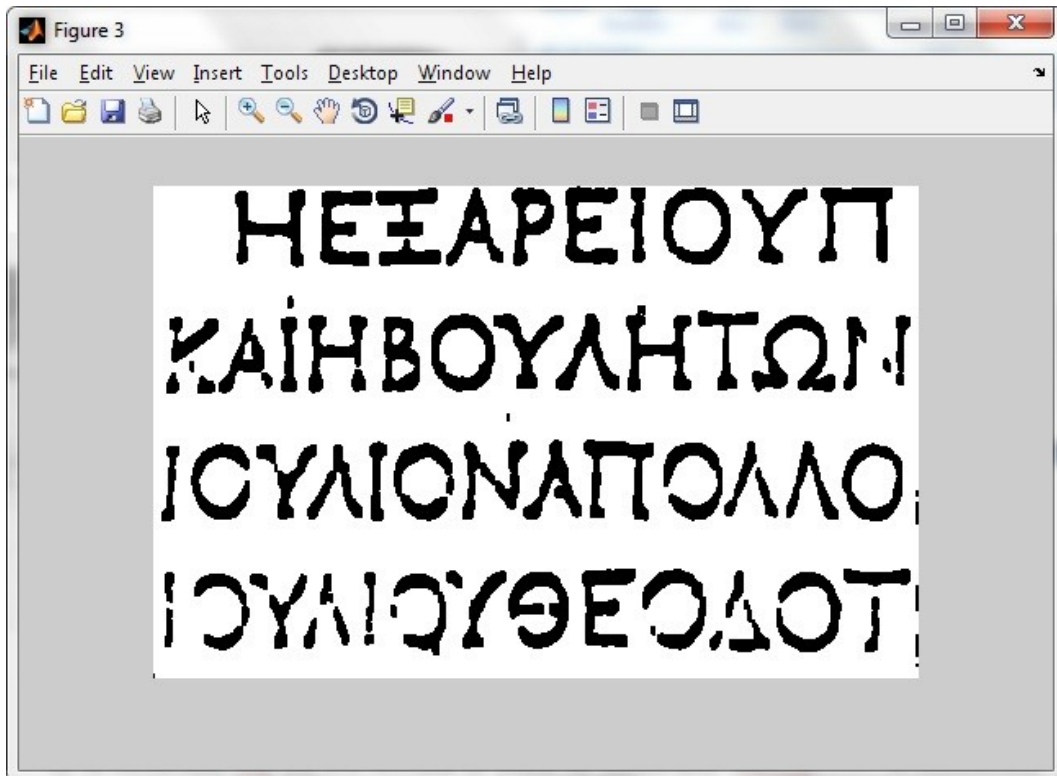
(β)



(γ)

Εικόνα 46. Διαδική μετατροπή και αφαίρεση θορύβου. (α) Αρχική εικόνα. (β) Η εικόνα μετά την κατοψλίωση. (γ) Η εικόνα μετά τη χρήση του φίλτρου μέσου.

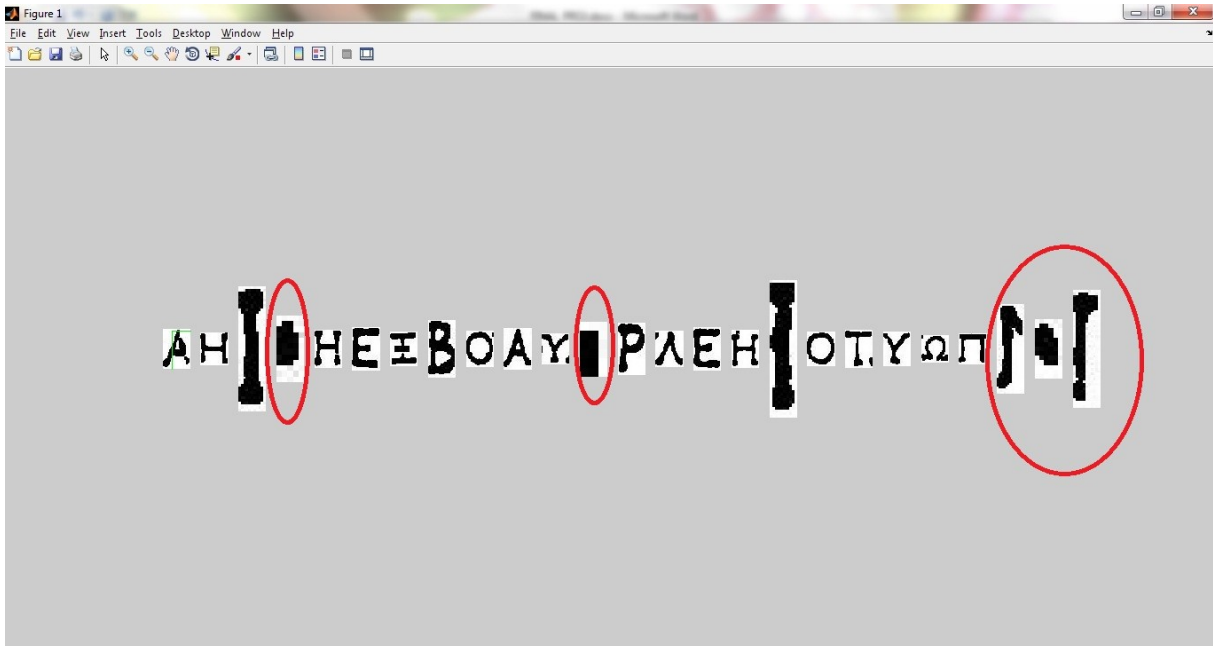
Στη συνέχεια επεξεργαζόμαστε την τελευταία εικόνα με κάποιες μορφολογικές εφαρμογές. Αρχικά κάνουμε χρήση της λειτουργίας ‘open’ με τη βοήθεια της συνάρτησης του MATLAB *bwmorph* και στη συνέχεια χρησιμοποιούμε τη μορφολογική διαστολή ‘dilation’. Για τη διαστολή δημιουργήσαμε πρώτα το δομικό στοιχείο (στην προκειμένη περίπτωση επιλέξαμε ορθογώνιο παραλληλόγραμμο 5x2) με τη συνάρτηση *strel* και στη συνέχεια με χρήση της συνάρτησης *imerode* το εφαρμόσαμε στην εικόνα.



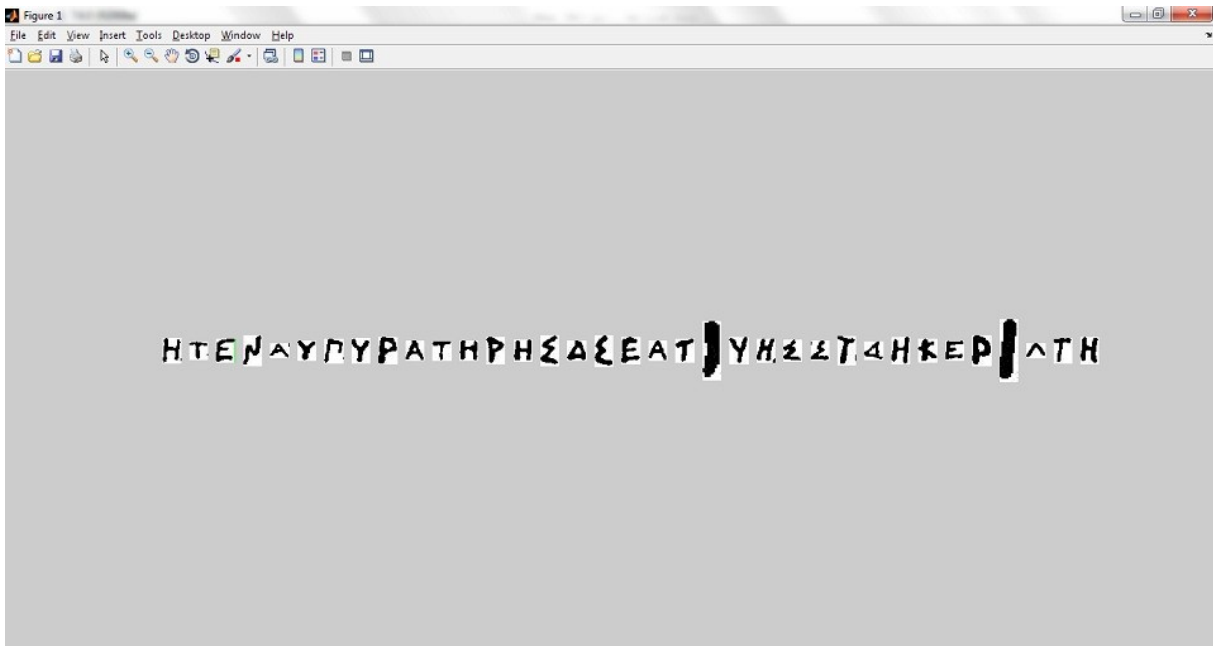
Εικόνα 47. Η επιγραφή της Εικόνας 45 μετά την εφαρμογή των μορφολογικών λειτουργιών *open* & *dilation*.

4.2. Κατάτμηση χαρακτήρων

Για την κατάτμηση των χαρακτήρων θα ακολουθήσουμε την εξής τεχνική στο MATLAB: Στην ήδη επεξεργασμένη εικόνα θα χρησιμοποιήσουμε μια συνάρτηση με την οποία θα προσπαθήσουμε να δημιουργήσουμε συμπαγείς χαρακτήρες γεμίζοντας τις οπές των γραμμών. Στη συνέχεια θα υπολογίσουμε το που βρίσκονται οι περιοχές αυτές και θα τις κόψουμε, θα αλλάξουμε το μέγεθός τους, θα τις εμφανίσουμε και στη συνέχεια θα τις αποθηκεύσουμε σε ξεχωριστά αρχεία εικόνας. Οι συναρτήσεις που θα χρησιμοποιήσαμε είναι : η *imdilate* για επιπλέον διόγκωση των γραμμών, την *imfill* για το γέμισμα των οπών των γραμμών, τις *bwlabel* & *regionprops* για τον εντοπισμό και την επιλογή των γραμμών και την *imcrop* για την αποκοπή των χαρακτήρων από την εικόνα. Όπως θα δούμε και στην εικόνα που ακολουθεί, προβλήματα παρουσιάζονται σε χαρακτήρες που μετά την προεπεξεργασία παρουσίασαν μεγάλα διαστήματα ασυνεχειών και σε αυτό το στάδιο ‘θεωρούνται’ ξεχωριστοί χαρακτήρες.



Εικόνα 48. Χαρακτήρες της επιγραφής της Εικόνας 47 ύστερα από την κατάτμηση. Σε κόκκινο κύκλο βρίσκονται οι χαρακτήρες που δεν έχουν αποκοπεί ορθά.



Εικόνα 49. Παράδειγμα επιγραφής της οποίας οι χαρακτήρες αποκόπηκαν χωρίς να παρουσιαστούν σφάλματα.

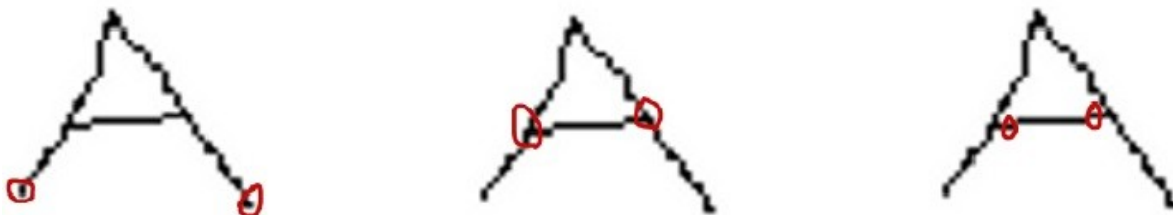
4.3. Εξαγωγή χαρακτηριστικών

Για την εξαγωγή των χαρακτηριστικών θα χρησιμοποιήσουμε μια τεχνική την οποία έχει υλοποιήσει στο MATLAB, ο **Dinesh Dileep**³⁷, με κάποιες αλλαγές. Ο αλγόριθμος του προαναφερθέντος υπολογίζει τα χαρακτηριστικά βασίζομενος στη γεωμετρία των χαρακτήρων. Τα χαρακτηριστικά υπολογίζονται από τις βασικές γραμμές που αποτελούν το σκελετό των χαρακτήρων. Το σύστημα αυτό επιστρέφει ένα διάνυσμα ως έξοδο. Τα χαρακτηριστικά αυτά μπορούν να χρησιμοποιηθούν για την εκπαίδευση κάποιου νευρωνικού δικτύου ή μπορούν να κατηγοριοποιηθούν με τη βοήθεια κάποιου αλγορίθμου κατηγοριοποίησης (εμείς θα χρησιμοποιήσουμε τον k-means).

Η βασική ιδέα του αλγορίθμου είναι η εξής:

- Χωρίζουμε την αρχική εικόνα σε τρεις υποεικόνες διαιρώντας την κάθετα σε τρία τμήματα (1x3). Έπειτα εξάγουμε τα χαρακτηριστικά για κάθε ξεχωριστή υποεικόνα.
- Χωρίζουμε την αρχική εικόνα σε τρεις υποεικόνες διαιρώντας την οριζόντια σε τρία τμήματα (3x1). Έπειτα εξάγουμε τα χαρακτηριστικά για κάθε ξεχωριστή υποεικόνα.
- Ενώνουμε τα αποτελέσματα των δύο εκτελέσεων.

Παραλλαγή της μεθόδου αυτής είναι ο χωρισμός της αρχικής εικόνας σε 9 τομείς (3x3), η εξαγωγή των χαρακτηριστικών για κάθε τομέα ξεχωριστά και τέλος η συνένωση των αποτελεσμάτων.



Εικόνα 50. Σημεία των χαρακτήρων, τα οποία χρησιμοποιεί ο αλγόριθμος για να εξαγάγει τα χαρακτηριστικά για μια εικόνα-γράμμα.

Ένα ακόμα χαρακτηριστικό που εξαγάγαμε είναι το centroid του κάθε πίνακα διανύσματος με τη χρήση της μεθόδου *k-means* του MATLAB. Αυτό μπορεί να χρησιμοποιηθεί και για την εκπαίδευση κάποιου νευρωνικού δικτύου στο στάδιο της αναγνώρισης, αλλά και ως δεδομένο κατηγοριοποίησης των χαρακτήρων, αν επιλεγεί για τις δοκιμές κάποιος αλγόριθμος clustering. Στη δεύτερη περίπτωση, δεν θα χρειαστεί να ‘γνωρίζουμε’ ποιος είναι ο κάθε χαρακτήρας, γιατί απλά θα κατηγοριοποιηθούν βάση της ομοιότητας των διανυσμάτων των χαρακτηριστικών.

4.4. Κατηγοριοποίηση - Αναγνώριση

Ολοκληρώνοντας τα προηγούμενα στάδια με επιτυχία και με κατά το δυνατό λιγότερα σφάλματα, μπορούμε να ελπίζουμε σε όσο το δυνατόν καλύτερα αποτελέσματα στο στάδιο της κατηγοριοποίησης-αναγνώρισης. Για την αναγνώριση των χαρακτήρων θα εκπαιδεύσουμε ένα νευρωνικό δίκτυο με δείγματα χαρακτήρων τα οποία θα χαρακτηρίζουν ένα γράμμα-κλάση, για παράδειγμα το Α ή Β και στη συνέχεια εισάγοντας τα πειραματικά δείγματα θα πάρουμε ως απάντηση

τον αριθμό της κλάσης στην οποία ανήκουν (π.χ. κλάση1=A, κλάση2=E ... κλπ.). Η εκπαίδευση θα γίνει δίνοντας τα στοιχεία για κάθε κλάση με μη αυτόματο τρόπο.

Το νευρωνικό δίκτυο, θα είναι ένα feed-forward (εμπροσθοτροφοδοτούμενο) δίκτυο, για τη δημιουργία του οποίου χρησιμοποιήθηκαν οι εξής βασικές συναρτήσεις:

- newff(): Είναι η συνάρτηση που δημιουργεί ένα νευρωνικό δίκτυο εμπρόσθιας τροφοδότησης και δέχεται σαν παραμέτρους τον πίνακα εκπαίδευσης, τον πίνακα των θέσεων των κλάσεων εκπαίδευσης, τον αριθμό των κλάσεων που θέλουμε να αναγνωριστούν, τη μέθοδο εκπαίδευσης, καθώς και αρκετά ακόμη στοιχεία ώστε να παραμετροποιηθεί πλήρως η συγκεκριμένη τεχνική.
- train(): Η συγκεκριμένη συνάρτηση δέχεται σαν όρισμα το προηγούμενο δίκτυο και αναλαμβάνει να το εκπαιδεύσει ανάλογα με τις παραμέτρους που έχουν προηγουμένως ορισθεί.
- sim(): Δέχεται σαν εισόδους το εκπαιδευόμενο δίκτυο και τον πίνακα δοκιμών και αποδίδει το αποτέλεσμα της αναγνώρισης.

Σαν πίνακα εκπαίδευσης θα πάρει στην είσοδο τον πίνακα P, ο οποίος σε κάθε στήλη του περιέχει τα χαρακτηριστικά του κάθε γράμματος και τον πίνακα T, με τον οποίο θα εκπαιδεύσουμε το δίκτυο, ορίζοντας τις θέσεις των κλάσεων.

Για τις ανάγκες του πειράματος χρησιμοποιήσαμε τέσσερις εκδοχές δέκα διαφορετικών χαρακτήρων και μια πέμπτη εκδοχή κάθε χαρακτήρα για το τεστ της αναγνώρισης.

	1	2	3	4	5	6	7	8	9	10	11	12	13
1	0	0.8400	0.0600	0.4700	0.4700	0	0.0600	0.0400	0.9000	0	0	0.5300	0.2800
2	0.2300	0.9400	0.5600	0.8000	0.4600	0.2900	0.8600	0.7300	0.3000	0	0.1400	0.9700	0.7200
3	0.9800	0.3500	0.8200	0.7900	0	0.9100	0.6600	1	0	0	0.9400	0.6100	0.9600
4	0.0700	0.0100	1	0.4300	0	0.4200	0.3000	0.8300	0.0800	0	0.1400	0.2400	0.6400
5	0	0	0.8200	0.2800	0.3100	0	0.0500	0.1400	0.5400	0.8600	0	0	0.0400
6	0	0.6900	0.6400	0.3900	0.4700	0.3500	0.1200	0.7600	0.7300	0.0600	0	0.7000	0.8000
7	0.5300	1	0.7300	0.4100	0.5000	0.9900	0.9000	0.7200	0.9400	0.7000	0.4800	0.9700	0
8	1	0.9200	0.2900	0.9700	0	0.6500	0.6000	0	0.2100	0.0500	1	0.9400	0.7600
9	0.3700	0.8900	0.2000	0.5700	0.1000	0.1200	0.7400	0.6200	0.8300	0	0.5000	0.9900	0.6800
10	0	0.4700	0.0700	0.8000	0.9000	0	0.9200	0.7600	0.6900	0.7000	0	0.3400	0.0800
11	0	0.5200	0.7300	0	0.5500	0.8100	0.2700	1	0.0800	0.2300	0	0.7000	0.8000
12	0.8000	0.7200	0.5200	0.1200	0.7200	0.9700	0.6000	0.2500	0.9300	1	0.8200	0.6000	0.5600
13	0.7400	0.0700	0.3000	0.7000	0.3900	0.2200	0	0	1	0.4800	0.2600	0.0900	0.0400
14	0.5400	0.8200	0.1000	0	0.3900	0	0.1000	0.1000	0.8200	0.3700	0.8400	0.8200	0.1600
15	0	0.8700	0	0	0.6600	0	0.7300	1	0.0300	0.7000	0.0200	0.8100	0
16	0.0900	0.3000	0.8000	0	0.6000	0.2400	0.4000	1	0	0.5000	0.2200	0.7000	0.8000
17	0.9700	0.9800	0.7200	0.2000	0.7500	1	0.5000	0.1000	0.5100	0.4700	0.9700	0.7200	0.8800
18	0.3500	0.9400	0.5800	0.7500	0.6000	0.9600	0	0	1	0.7400	0.7200	0.0600	0.8000
19	0.9000	0.7300	0.1100	0	0.8300	0.1000	0.2300	0.1000	0.3100	0.4800	1	0.7500	0.7200
20	0.0100	0.0700	0	0	0.4500	0	0.6400	1	0	0.6300	0.1600	1	0
21	0.4600	0.2500	0.9800	0	0.6800	0.8900	0.5900	1	0	0.6300	0.6200	0.7700	0.7600
22	0.9400	1	0.0300	0.2000	0.1000	0.8600	0.3100	0.2500	0.4300	0	0.8000	0.9900	0.4000
23	0.7000	0.4600	0	0.9000	0	0.2000	0	0	1	0.4400	0.4000	0.8800	0
24	0.9900	0	0	0	0.5000	0	0.0700	0.2700	0.1500	1	0.7700	0.9500	0
25	0.3800	0	0	0	0.3800	0	0.2000	0.7100	0	0.5000	0.4200	0.6300	0
26	0.8500	0.1900	1	0	0.8700	1	0.7800	0.7000	0	0.9700	0.9300	1	0.8400
27	0.3500	1	0.6600	0.1900	0.1700	0.9100	0.1700	0.7800	0.6800	0	0.0800	0.8700	0.9200
28	0	0.2300	0.7800	0.9000	0	0.4900	0	0.4100	1	0.0100	0	0.0600	0.8000

Πίνακας 3. Απόσπασμα του πίνακα P με τα χαρακτηριστικά του κάθε χαρακτήρα.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	1	0	0	0	0	0	0	0	0	0	1	0	0	0
2	0	1	0	0	0	0	0	0	0	0	0	1	0	0
3	0	0	1	0	0	0	0	0	0	0	0	0	1	0
4	0	0	0	1	0	0	0	0	0	0	0	0	0	1
5	0	0	0	0	1	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	1	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	1	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	1	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	1	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	1	0	0	0	0
11														
12														

Πίνακας 4. Απόσπασμα του πίνακα T των στόχων (targets), με τον οποίο ορίζουμε τις ομάδες των χαρακτήρων.

Neural Network Training (ntraintool)

Neural Network

Input → Layer (W, b) → Layer (W, b) → Output

Algorithms

Training: Gradient Descent Backpropagation with Adaptive Learning Rate. (trainingdx)
 Performance: Sum Squared Error (sse)

Progress

Epoch:	0	430 iterations	5000
Time:		0:00:02	
Performance:	99.7	0.0948	0.100
Gradient:	1.00	0.129	1.00e-10
Validation Checks:	0	0	6

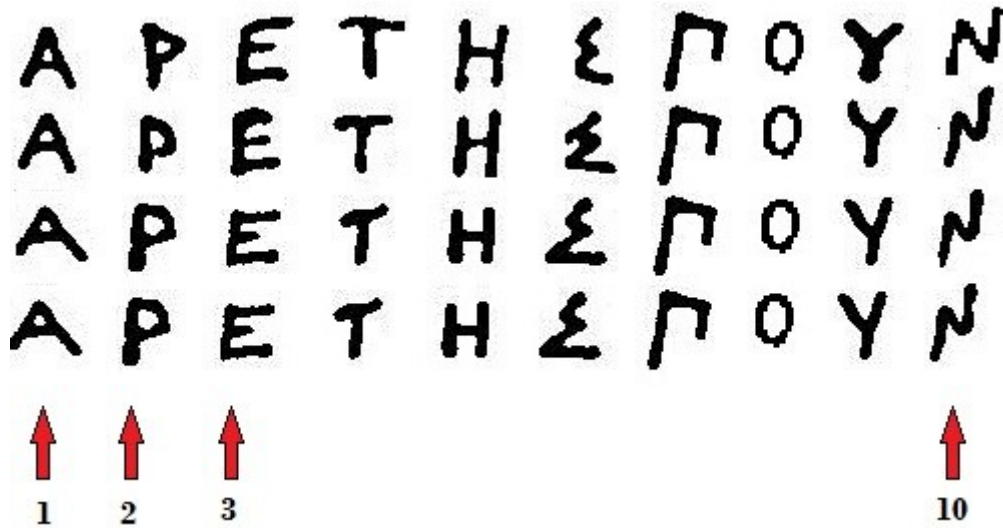
Plots

Performance (plotperform)
 Training State (plottrainstate)
 Regression (plotregression)

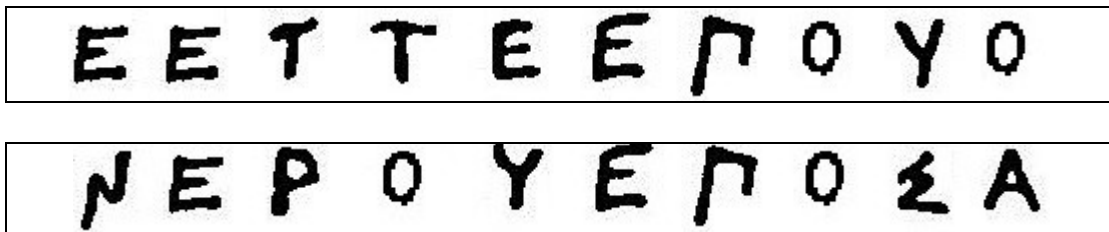
Plot Interval: 1 epochs

✓ Performance goal met.

Stop Training Cancel



Εικόνα 51. Οι ομάδες των χαρακτήρων με τους οποίους εκπαιδεύτηκε το δίκτυο. Με τα κόκκινα βέλη δείχνουμε τον αριθμό που θα αντιπροσωπεύει στο πείραμα την κάθε κλάση.



Εικόνα 52. Οι σειρές των χαρακτήρων που χρησιμοποιήθηκαν για τη δοκιμή. Σαν αποτέλεσμα περιμένουμε στη θέση του κάθε χαρακτήρα τον αριθμό της κλάσης στην οποία ανήκει.


```

Command Window
File Edit Debug Desktop Window Help
Warning: NEWFF used in an obsolete way.
> In nntobsu at 18
   In newff at 86
   In createnn at 9
   In imgpreprocess22 at 64
      See help for NEWFF to update calls to the new argument list.

      3   3   4   4   3   3   7   8   9   8

Warning: NEWFF used in an obsolete way.
> In nntobsu at 18
   In newff at 86
   In createnn at 9
   In imgpreprocess22 at 64
      See help for NEWFF to update calls to the new argument list.

      10   3   2   8   9   3   7   8   6   1

>>
OVR
    
```

Εικόνα 53. Τα επιτυχή αποτελέσματα για τις ακολουθίες χαρακτήρων της Εικόνας 51.

Τα αποτελέσματα αυτά τα πήραμε ύστερα από δοκιμή που έγινε μετά την πρώτη εκπαίδευση του δικτύου το οποίο αποτελούσαν δύο ‘κρυφά’ επίπεδα με δέκα (10) νευρώνες στο πρώτο και μεταβλητό αριθμό (ανάλογα με τους στόχους) στο δεύτερο.

Μεταβάλλοντας τον αριθμό των νευρώνων παρουσιάστηκαν λάθη στα αποτελέσματα, κυρίως στα γράμματα που παρουσιάζουν κάποια ομοιότητα στη μορφή τους (π.χ. E-Σ, Π-T, M-N).

Επίσης, από τα χαρακτηριστικά που εξαγάγαμε για τον κάθε χαρακτήρα, με χρήση της συνάρτησης του MATLAB *k-means*, κάναμε την εξαγωγή των centroid για κάθε διάνυσμα χαρακτηριστικών. Έτσι με χρήση διαφορετικής σύνταξης της προαναφερθείσας συνάρτησης επιχειρήσαμε να κάνουμε κατηγοριοποίηση των γραμμάτων. Η συνάρτηση δεχόταν ως ορίσματα έναν πίνακα ο οποίος περιείχε τα centroid κάθε διανύσματος χαρακτηριστικών και τον αριθμό των κλάσεων στις οποίες χωρίστηκαν τα δεδομένα. Τα αποτελέσματα ήταν θετικά μόνο για έναν πολύ μικρό αριθμό κλάσεων και κατ’ επέκταση δεν θεωρήσαμε αξιόλογη τη μέθοδο αυτή για να τη συνεχίσουμε.

Παρόλα αυτά, η ιδέα αυτή και το όλο πρόβλημα είναι αρκετά ενδιαφέροντα, και θα άξιζε να μελετηθεί και σε μελλοντική εργασία.

5. Συμπεράσματα και μελλοντικές εργασίες

Ολοκληρώνοντας τον κύκλο των πειραμάτων, και εξετάζοντας τα αποτελέσματα του συστήματος το οποίο προσομοιώσαμε, παρατηρούμε ότι αυτά είναι αρκετά ενθαρρυντικά, σε σχέση με τα δεδομένα τα οποία διαθέτουμε. Παρόλα αυτά όμως το σύστημα της προσομοίωσης επιδέχεται βελτιώσεων κατά κύριο λόγο σε δύο βασικούς τομείς.

Ο πρώτος τομέας έχει να κάνει με τον χρόνο εκτέλεσης, ο οποίος σε αρκετές περιπτώσεις είναι 'ασύμφορα' μεγάλος. Σε αυτό το γεγονός αρνητική συμβολή έχει το MATLAB το οποίο παρόλο που παρέχει πληθώρα ευκολιών και μεγάλη ευελιξία, υστερεί σε ταχύτητα σε σχέση με άλλες γλώσσες προγραμματισμού όπως η C ή C++. Αν η ανάπτυξη του ολοκληρωμένου πια συστήματος μπορούσε να γίνει με τη βοήθεια αυτών ή του ίδιου επιπέδου γλωσσών, θα είχαμε μια συμβολή στο να μειωθεί σημαντικά ο χρόνος εκτέλεσης και να γίνει η εφαρμογή πιο αποτελεσματική σε μεγαλύτερο αριθμό δεδομένων ή συνόλων δεδομένων πλησιάζοντας περισσότερο τις απαιτήσεις που θα υπήρχαν υπό πραγματικές συνθήκες.

Ο δεύτερος τομέας ο οποίος επιδέχεται βελτιώσεις για ταχύτερα αποτελέσματα, πέραν της βελτίωσης του τρόπου εκτέλεσης του κώδικα των πειραμάτων και κατ' επέκταση του συστήματος, θα μπορούσαν να γίνουν πρόσθετες αλλαγές και στην ίδια τη δομή του κώδικα, οι οποίες αφορούν τη διαδικασία εκπαίδευσης του συστήματος αλλά και το όλο στάδιο κατηγοριοποίησης-αναγνώρισης. Στα πειράματά μας η εκπαίδευση έγινε με μη αυτόματο τρόπο και τα δεδομένα αντλήθηκαν από άλλα τμήματα της ίδιας επιγραφής ή από τμήματα επιγραφών της ίδιας κατηγορίας. Αυτό έγινε για τον λόγο ότι δεν υπάρχουν ακόμα πλήρεις και λειτουργικές βάσεις δεδομένων με χαρακτήρες της αρχαίας ελληνικής γλώσσας, για να χρησιμοποιηθούν για την αυτόματη εισαγωγή δεδομένων εκπαίδευσης στην περίπτωση του νευρωνικού δικτύου. Σε αυτή την κατεύθυνση κινούνται κάποιοι ερευνητές οι οποίοι παρουσίασαν ένα ανάλογο σύστημα³⁸.

Με τη χρήση αυτών των δεδομένων και νέων πιο βελτιωμένων αλγορίθμων εκπαίδευσης νευρωνικών δικτύων ή γενικά τεχνικών μηχανικής μάθησης υπάρχει δυνατότητα βελτίωσης σε σημαντικό βαθμό της αποδοτικότητας ενός τέτοιου συστήματος.

Επιπλέον, μελέτες όπως του Παπαοδυσσέα Tracy et al.³⁹ μπορούν δώσουν νέα δεδομένα προς επεξεργασία, τα οποία να οδηγούν σε καλύτερα και πιο αξιόπιστα αποτελέσματα, όπως η ταυτοποίηση δύο διαφορετικών ενεπίγραφων αρχαιολογικών αντικειμένων με βάση το σχήμα των γραμμάτων τους, ότι δηλαδή ανήκουν στον ίδιο χαρακτή και συνεπώς πρόκειται για ένα αντικείμενο και την ίδια επιγραφή.

Μελλοντική εργασία, για την οποία βάση θα αποτελεί ένα σύστημα σαν και αυτό που αναλύουμε, είναι η δημιουργία λογισμικού με το οποίο, αφού γίνει η επεξεργασία και αναγνώριση-κατηγοριοποίηση των χαρακτήρων, δημιουργείται ένα αλφάβητο με αυτόματο τρόπο, για την εκάστοτε επιγραφή, πράγμα που θα εξυπηρετούσε ιδιαίτερα την επιγραφική έρευνα και όχι μόνον. Το αλφάβητο της κάθε επιγραφής είναι μια κατανομή των χαρακτήρων που χρησιμοποιήθηκαν κατά τη χάραξή της. Με τη μελέτη του, μπορούμε να εξαγάγουμε πληροφορίες όπως, π.χ.:

- τις ιδιομορφίες στη χάραξη κάθε γράμματος του χαρακτή (όλα τα Α είναι στο ίδιο σχήμα αλλά όλα δεν είναι απολύτως ίδια),
- το συγκεκριμένο αλφάβητο του χαρακτή της επιγραφής με τις ιδιομορφίες του και συνεπώς τον έλεγχο άλλων επιγραφών που τυχόν χάραξε ο ίδιος,
- με τη συνένωση των αλφαβητών της κάθε επιγραφής από μία συγκεκριμένη περιοχή (Αττική, Βοιωτία, Εύβοια, Κρήτη κλπ.) δημιουργείται και το τοπικό αλφάβητο της περιοχής από όπου προήλθαν οι επιγραφές με σκοπό την παρακολούθηση της εξέλιξης στη χάραξη των

γραμμάτων από την εμφάνιση του ελληνικού αλφαβήτου τον 8ο αιώνα π.Χ. κ.ε. (για μια ενδεικτική χαρτογράφηση των αρχαίων αλφαβήτων βλ. <http://poinikastas.csad.ox.ac.uk/>).

Ένα τέτοιο πρόγραμμα θα μπορούσε να είναι προσβάσιμο και μέσω του διαδικτύου έτσι ώστε ο εξουσιοδοτημένος ερευνητής να μπορεί από απόσταση να αποκτήσει online πληροφορίες για τα αλφάβητα μιας περιοχής και να συνδυάσει την επιγραφή που τον ενδιαφέρει, κυρίως αν αυτή είναι νέα ανακάλυψη, με τα επιγραφικά δεδομένα της περιοχής αυτής (εξέλιξη του αλφαβήτου της περιοχής, ιδιομορφίες του, χρονολόγηση επιγραφών, κ.ά.).

6. Αναφορές – Βιβλιογραφία

¹ Παναγόπουλος Μ.Κ., «Ταυτοποίηση χαρακτών αρχαίων επιγραφών και εντοπισμός τεχνοτροπίας αρχαιολογικών ευρημάτων με χρήση πρωτότυπων αλγορίθμων αναγνώρισης προτύπων και επεξεργασίας εικόνας», Διδακτορική Διατριβή, Εθνικό Μετσόβειο Πολυτεχνείο, Αθήνα 2008.

² Woodhead A. G., *The Study of Greek Inscriptions*. Cambridge University Press, 1967.

³ Barmoutis A., Bozia E., Wagman R.S., «A novel framework for 3D reconstruction and analysis of ancient inscriptions», *Journal of Machine Vision and Applications*, 2009

⁴ <http://www.mathworks.com/products/matlab/description1.html>

⁵ Sezgin M., Sankur B., Survey over image thresholding techniques and quantitative performance evaluation, *Journal of Electronic Imaging* 13(1), 146–165 January 2004.

⁶ Gonzalez R.C, Woods R.E , Eddins S.L., «Digital image processing using matlab» 405-408.

⁷ Ridler, T.W. and Calvard, S., «Picture Thresholding Using an Iterative Selection Method.» *IEEE Trans. On Systems, Man and Cybernetics*, Vol. SMC-8 1978, 630-632.

⁸ Otsu, N., «A Threshold Selection Method from Gray-level Histograms.» *IEEE Trans. On Systems, Man and Cybernetics*, Vol. 9, 1979 377-393

⁹ Pun, T., «Entropic Thresholding: A New Approach.» *CVGIP, Graphical Models and Image Processing* 16, (1981) 210—239.

Kapur, J. N., Sahoo, P. K. and Wong, A. K. C. «A New Method for Gray-Level Picture Thresholding Using the Entropy of the Histogram.» *Computer Vision, Graphics and Image Processing* 29 (1985) 273-285.

¹⁰ Shannon, C.E., «Prediction and entropy of printed English», *The Bell System Technical Journal*, 30:50-64, January 1951.

¹¹ Shannon C.E., «A Mathematical Theory of Communication», *Bell System Technical Journal*, vol. 27, pp. 379–423, 623-656, July, October, 1948.

¹² Weszka, J.S., Nagel, R.N., and Rosenfeld, A., A Threshold Selection Technique, *TC(23)*, 1974, pp. 1322-1326.

¹³ Chow, C.K. and Kaneko, T., «Boundary detection of radiographic images by a threshold method.» In *Proceedings IFIP Congress - 71, Booklet - TA-7*, (1972) 130-134

¹⁴ Niblack, W., «An Introduction to Digital Image Processing», Englewood Cliffs, Prentice Hall, pp. 115-116, 1986.

¹⁵ Sauvola, J., Pietikainen, «Adaptive document image binarization.» *Pattern Recognition* 33 (2000) 225-236
Sauvola, J., Seppanen, T., Haapakoski, S. and Pietikainen, M., «Adaptive Document Binarization.» *ICDAR 97*, pp. 147-152, Ulm, Germany, 1997.

¹⁶ Gatos, B., Pratikakis, I. and Perantonis, S. J., «Locating Text in Historical Collection Manuscripts.» *3rd Hellenic Conference on Artificial Intelligence*, Samos, Greece 2004.

¹⁷ Kelly, Philip and Ó Conaire, Ciarán and Monaghan, David and Kuklyte, Jogile and Connaghan, Damien and Pérez-Moneo Agapito, Juan Diego and Daras, Petros (2010) «Performance analysis and visualisation in tennis using a low-cost camera network.» In: *Multimedia Grand Challenge Track at ACM Multimedia 2010*, 25-29, Firenze, Italy October 2010.

- ¹⁸ Suen, C.Y., Nadal, C., Legault, R., Mai, T.A., Lam, L., «Computer recognition of unconstrained handwritten numerals.» Proc. IEEE, 80(7) (1992) 1162-1180.
- ¹⁹ Schilling, R. J., Fundamental of Robotics Analysis and Control. Prentice-Hall, Englewood Cliffs, NJ 1990.
- ²⁰ Rajiv K. Sharma, Amardeep Singh, «Detection and segmentation of Handwritten Text in Gurmukhi Script using Flexible Windowing», International Journal of Computer Theory and Engineering, Vol. 2, No. 3, June, 2010 1793-8201 329-332.
- ²¹ Ο.π. σημ. 03
- ²² Lu, Y., Haist, B., Harmon, L., Trenkle, J., Vogt, R., «An accurate and efficient system for segmenting machine-printed text.» Proc. of the U.S. Postal Service 5th Advanced Technology Conference, Vol. 3 (1992) 93-105.
- ²³ Παπαθανασίου Α.Σ, «Μελέτη,σχεδιασμός και υλοποίηση αλγορίθμων εντοπισμού θέσης και αναγνώρισης χαρακτήρων σε τυπωμένες εικόνες» Διπλωματική Εργασία, Τμήμα Ηλεκτρολόγων Μηχανικών και Τεχνολογίας Υπολογιστών, Πολυτεχνική Σχολή Πανεπιστημίου Πατρών, Πάτρα, 2010, 13-24.
- ²⁴ Pratt, W.K.: Digital Image Processing, 2nd edn. Wiley, New York 1991.
- ²⁵ Bokser, M.: Omnidocument technologies. Proc. IEEE 80 1992 1066-1078.
- ²⁶ Gatos, B., Karras, D., Perantonis, S., «Optical Character Recognition Using Novel Feature Extraction & Neural Network Classification Techniques.» Proc. of the Workshop on Neural Network Application and Tools, IEEE Computer Society Press (1993) 65-72.
- ²⁷ Δουλιγέρη Ν., «Οπτική Αναγνώριση Λέξης», Διπλωματική εργασία, Πανεπιστήμιο Αιγαίου, Καρλόβασι 2007.
- ²⁸ E.Kavallieratou, N.Fakotakis, and G.Kokkinakis, «Un Off-line Unconstrained Handwriting Recognition System», International Journal of Document Analysis and Recognition, no 4, pp. 226-242, 2002.
- ²⁹ Heute, Paquetet, Moreau, Lecourtier, Olivier, «A structural/statistical feature based vector for handwritten character recognition», Pattern Recognition Letters, vol 19, pp 629-641, 1998.
- ³⁰ Lavrenko V.,Rath T.M, Manmatha R., «Holistic Word Recognition for Handwritten Historical Documents», DIAL, pp 278-287,2004.
- ³¹ Zhang B., Shrihari S., Huang C., «Word Image Retrieval using Binary Features», SPIE, Document Recognition and Retrieval XI, San Jose, California, USA, 2004.
- ³² Kanungo T., Netanyahu N.S., A.Y.Wu, «An Efficient k-Means Clustering Algorithm: Analysis and Implementation», IEEE Trans on Pattern Analysis and Machine Intelligence, Vol24, No7, pp 881-892 , 2002.
- ³³ Faber V., «Clustering and the Continuous k-Means Algorithm», Los Alamos Science, No 22, pp138-144, 1994.

³⁴ Ο.π. σημ. 23

³⁵ Ο.π. σημ. 32

³⁶ Ο.π. σημ. 23

³⁷ Dileep D., «A feature extraction technique based on character geometry for character recognition», (www.mathworks.com) .

³⁸ Ο.π. σημ. 03

³⁹ Ο.π. σημ. 01