ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Αναγνώριση Γλώσσας Χειρόγραφων Εγγράφων με Χρήση LBP και SIFT Χαρακτηριστικών

Βλάσιος Σ. Κωτσοβίλης

Επιβλέπων: **Γάτος Βασίλειος,** Κύριος Ερευνητής, ΕΚΕΦΕ «Δημόκριτος»

ΑΘΗΝΑ

ΜΑΡΤΙΟΣ 2020

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Αναγνώριση Γλώσσας Χειρόγραφων Εγγράφων με Χρήση LBP και SIFT Χαρακτηριστικών

Βλάσιος Σ. Κωτσοβίλης

Α.Μ.: M1573

ΕΠΙΒΛΕΠΩΝ: **Γάτος Βασίλειος,** Κύριος Ερευνητής, ΕΚΕΦΕ «Δημόκριτος»

ΕΞΕΤΑΣΤΙΚΗ ΕΠΙΤΡΟΠΗ: **Παναγάκης Ιωάννης,** Αναπληρωτής Καθηγητής ΕΚΠΑ

ΜΑΡΤΙΟΣ 2020

ΠΕΡΙΛΗΨΗ

Η αναγνώριση γλώσσας εικόνων χειρόγραφων εγγράφων είναι ένα πρόβλημα ανάλυσης εγγράφων στο οποίο οι γλώσσες αντιστοιχούν σε ένα σύνολο γραφικών αναπαραστάσεων που χρησιμοποιούνται για να εκφράσουν ένα συγκεκριμένο σύστημα γραφής. Κάθε γλώσσα έχει τα δικά της χαρακτηριστικά όχι μόνο αναφορικά με τη φυσική της μορφή, αλλά και με το στυλ γραφής της. Η υφή μιας εικόνας είναι ένα μοναδικό χαρακτηριστικό, το οποίο μπορεί να χρησιμοποιηθεί για την αναγνώριση της γλώσσας μιας εικόνας εγγράφου και μπορεί να οριστεί ως ένα επαναλαμβανόμενο μοτίβο από εικονοστοιχεία (pixels) με δομημένο τρόπο. Προκειμένου να εξαχθούν χαρακτηριστικά βασισμένα στην υφή, χρησιμοποιούνται τα Τοπικά Δυαδικά Πρότυπα (Local Binary Patterns - LBP), που είναι απλά στην εφαρμογή τους και δεν επηρεάζονται από αλλαγές στις τιμές έντασης των εικονοστοιχείων μιας εικόνας. Τα LBP χαρακτηρίζουν τμήματα της εικόνας χρησιμοποιώντας δυαδικούς κώδικες, οι οποίοι κωδικοποιούν τη σχέση μεταξύ του κεντρικού εικονοστοιχείου και των γειτόνων του. Από την άλλη πλευρά, τα χαρακτηριστικά που είναι βασισμένα στην κλίση, όπως είναι οι περιγραφείς του Μετασχηματισμού Χαρακτηριστικών Αμετάβλητης Κλίμακας (Scale Invariant Feature Transform - SIFT), περιγράφουν οπτικά χαρακτηριστικά σε τοπικές περιοχές των χειρογράφων χωρίς να απαιτείται τμηματοποίηση. Συγκεκριμένα, ο SIFT είναι ένας αλγόριθμος ανίχνευσης σημείων κλειδιών (keypoints) που εντοπίζει τοπικές αλλαγές στην ένταση των εικονοστοιχείων των εικόνων. Παρέχει επίσης έναν επαρκή αριθμό σημείων κλειδιών για λεπτομερή χρήση. Στην παρούσα εργασία παρουσιάζουμε ένα σύστημα αυτόματης αναγνώρισης γλώσσας σε εικόνες χειρόγραφων εγγράφων χωρίς την εφαρμογή τμηματοποίησης. Η αναγνώριση της γλώσσας μπορεί να θεωρηθεί ως ένα πρόβλημα ταξινόμησης στο οποίο κάθε γλώσσα αντιπροσωπεύει μια κλάση. Κωδικοποιούμε τις δομές του κειμένου χρησιμοποιώντας περιγραφείς υφής ή περιγραφείς αμετάβλητης κλίμακας και περιστροφής, που προέρχονται από τα LBP και τα SIFT χαρακτηριστικά αντίστοιχα. Τα LBP και SIFT χαρακτηριστικά χρησιμοποιούνται ανεξάρτητα σε πειράματα, ώστε να εξαχθούν τα χαρακτηριστικά από τις εικόνες των εγγράφων. Η αναγνώριση της γλώσσας επιτυγχάνεται με τη χρήση των ταξινομητών Κ Πλησιέστερου Γείτονα (K Nearest Neighbour - KNN), Naive Bayes Nearest Neighbour (NBNN) και Local NBNN. Η ταξινόμηση των άγνωστων εγγράφων σε μια συγκεκριμένη γλώσσα βασίζεται στη σύγκριση με τα χαρακτηριστικά των εγγράφων του συνόλου αναφοράς. Τα πειράματα για την αξιολόγηση του συστήματος εκτελούνται σε εικόνες χειρόγραφων εγγράφων γραμμένες στη γαλλική, γερμανική, ελληνική και αγγλική γλώσσα οι οποίες είναι μέρος μιας δημόσιας βάσης δεδομένων που περιέχει 208 έγγραφα από 26 γραφείς και έχουν χρησιμοποιηθεί στη βιβλιογραφία σε διαγωνισμούς εντοπισμού γραφέα. Η εργασία αυτή περιλαμβάνει λεπτομερή αποτελέσματα για όλες τις παραπάνω μεθόδους τα οποία σε κάποιες περιπτώσεις ξεπερνούν το 85% (ποσοστό ορθής ταξινόμησης στη γλώσσα του χειρογράφου).

ΘΕΜΑΤΙΚΗ ΠΕΡΙΟΧΗ: Αναγνώριση Γλώσσας

ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ: LBP, SIFT, χειρόγραφα έγγραφα, αναγνώριση γλώσσας, local NBNN

ABSTRACT

Language identification for handwritten document images is a document analysis problem in which languages correspond to a set of graphical representations used to express a particular system of writing. Each language corresponds to unique features not only concerning the physical form, but also the writing style. Texture of an image is a unique feature that can be used to identify the language of a document image and can be defined as a repeating pattern of pixels in a structured way. In order to extract texture-based features, Local Binary Patterns (LBP) are used, which are simple in implementation and provide robustness to changes in the intensity values of image's pixels. LBP characterizes image patches using binary codes which encode the relationship between the central pixel and its neighbours. On the other hand, gradient-based features, such as Scale Invariant Feature Transform (SIFT) descriptors, describe visual features on local regions of handwritings without the need for segmentation. Particularly, SIFT is a keypoints detection algorithm which detects local changes in the intensity of pixels in images. It also provides a sufficient number of keypoints for an in-depth use. In this thesis we present a system for automatic language identification in handwritten document images, without applying any segmentation step. Language identification can be viewed as a problem of classification in which each language represents a class. We encode text structures using texture, scale and rotation invariant descriptors derived from LBP and SIFT features respectively. LBP and SIFT features are used in experiments independently in order to extract the features from document images. Identification of language is accomplished by using K Nearest Neighbour (KNN), Naive Bayes Nearest-Neighbour (NBNN) and Local NBNN classifiers. Classification of test documents is based on the distance from features of training documents. The experiments for the evaluation of the system are performed on handwritten document images written in French, German, Greek and English languages and are part of a public dataset which contains 208 documents from 26 writers and has been used in several writer identification competitions. This thesis includes detailed results of all the above methods and it is demonstrated that language classification accuracy can reach a percentage of over 85%.

SUBJECT AREA: Language Identification

KEYWORDS: LBP, SIFT, handwritten documents, language identification, local NBNN

Η διπλωματική εργασία αυτή αφιερώνεται σε όσους συνέβαλαν στην ολοκλήρωσή της.

ΠΕΡΙΕΧΟΜΕΝΑ

ΚΑΤΑΛΟΓΟΣ ΕΙΚΟΝΩΝ

ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

ΠΡΟΛΟΓΟΣ

Η παρούσα Διπλωματική Εργασία εκπονήθηκε στην Αθήνα από τον Νοέμβριο του 2018 μέχρι και τον Μάρτιο του 2020. Αποτελεί αναπόσπαστο κομμάτι για την απόκτηση του πτυχίου και διεξήχθη κατά το τελευταίο έτος της φοίτησής μου ως μεταπτυχιακός φοιτητής στο Τμήμα Πληροφορικής και Τηλεπικοινωνιών του Εθνικού και Καποδιστριακού Πανεπιστημίου Αθηνών. Για τη διεκπεραίωσή της, θα ήθελα να ευχαριστήσω τον επιβλέποντα, κύριο ερευνητή, κ. Βασίλειο Γάτο και τον επιστημονικό ερευνητή, κ. Παναγιώτη Καδδά, για τη συνεργασία και την πολύτιμη συμβολή τους στην ολοκλήρωσή της.

1. ΕΙΣΑΓΩΓΗ

1.1 Αναγνώριση της γλώσσας χειρόγραφων εγγράφων

Η αναγνώριση της γλώσσας χειρόγραφων εγγράφων είναι η διαδικασία ανάθεσης μιας γλώσσας σε άγνωστα δείγματα χειρογράφων χρησιμοποιώντας γνωστά δείγματα τα οποία θεωρούνται ως σύνολο αναφοράς. Με άλλα λόγια ελέγχεται η ομοιότητα μεταξύ των δειγμάτων. Πρόκειται για ένα πρόβλημα ταξινόμησης εικόνων, όπου οι εικόνες είναι τα δείγματα των χειρόγραφων εγγράφων και όλα τα δείγματα της ίδιας γλώσσας αναπαριστούν μια κλάση. Είναι μια αρκετά δύσκολη διαδικασία καθώς η ποικιλία των γραφικών χαρακτήρων είναι μεγάλη ακόμη κι αν πρόκειται για την ίδια γλώσσα γραφής.

Ο έλεγχος της ομοιότητας των δειγμάτων πραγματοποιείται συγκρίνοντας ορισμένα χαρακτηριστικά τους που εξάγονται από τις εικόνες με συγκεκριμένες τεχνικές επεξεργασίας εικόνας. Πρόκειται για χαμηλού επιπέδου (low-level) τεχνικές οι οποίες ελέγχουν κάθε εικονοστοιχείο (pixel) της εικόνας για τον εντοπισμό του χαρακτηριστικού. Πολλές φορές όμως το υπολογιστικό τους κόστος είναι μεγάλο και ενδεχομένως να απαιτούν πολύ χρόνο ώστε να ολοκληρωθούν. Για το λόγο αυτό υπάρχουν εναλλακτικές τεχνικές υψηλότερου επιπέδου (high-level), οι οποίες ελέγχουν συγκεκριμένα κομμάτια της εικόνας για τον εντοπισμό των χαρακτηριστικών και όχι εικονοστοιχείο προς εικονοστοιχείο.

Υπάρχει πληθώρα χαρακτηριστικών που μπορούν να χρησιμοποιηθούν για τον σκοπό αυτό και η επιλογή των κατάλληλων εξαρτάται κυρίως από το εκάστοτε πρόβλημα και το είδος της εφαρμογής.

1.2 Οπτικά Χαρακτηριστικά

Τα οπτικά χαρακτηριστικά (visual features) χρησιμοποιούνται για την επίλυση υπολογιστικών προβλημάτων στα επιστημονικά πεδία της μηχανικής όρασης (computer vision) και της επεξεργασίας εικόνας (image processing). Τα χαρακτηριστικά μπορεί να είναι συγκεκριμένες δομές στην εικόνα όπως σημεία, άκρες (edges) ή αντικείμενα, αλλά και αποτέλεσμα αλγορίθμων ανίχνευσης χαρακτηριστικών (feature detection).

Η ιδέα τους είναι αρκετά γενική και η κατάλληλη επιλογή τους εξαρτάται σε πολύ μεγάλο βαθμό από το είδος του προβλήματος.

1.3 Οπτικοί Περιγραφείς

Οι οπτικοί περιγραφείς (visual descriptors) είναι περιγραφές των οπτικών χαρακτηριστικών του περιεχομένου των εικόνων. Περιγράφουν στοιχειώδη χαρακτηριστικά όπως το σχήμα (shape), το χρώμα (color) και η υφή (texture).

Το **σχήμα** περιέχει σημασιολογικές πληροφορίες λόγω της ικανότητας του ανθρώπου να αναγνωρίζει αντικείμενα μέσω αυτού. Αυτές οι πληροφορίες μπορούν να εξαχθούν μόνο μέσω της κατάτμησης που είναι μια διαδικασία παρόμοια με αυτή που εφαρμόζει το ανθρώπινο οπτικό σύστημα ώστε να ξεχωρίσει αντικείμενα. Οπτικοί περιγραφείς που περιγράφουν το σχήμα είναι, για παράδειγμα, ο Περιγραφέας Σχήματος με βάση την Περιοχή (Region-based Shape Descriptor), ο Περιγραφέας Σχήματος με βάση το Περίγραμμα

(Contour-based Descriptor) και ο Περιγραφέας 3-D Σχήματος (3-Dimensional Shape Descriptor).

Το **χρώμα** είναι το βασικότερο στοιχείο ενός οπτικού περιεχομένου και περιγράφεται από περιγραφείς χρώματος, όπως είναι ο Περιγραφέας Δομής Χρώματος (Color Structure Descriptor) και ο Περιγραφέας Διάταξης Χρώματος (Color Layout Descriptor).

Η **υφή** αποτελεί εξίσου σημαντικό στοιχείο περιγραφής μιας εικόνας. Οι οπτικοί περιγραφείς υφής χαρακτηρίζουν υφές ή περιοχές της εικόνας. Παρατηρούν την ομοιογένεια της περιοχής και τα ιστογράμματα (histograms) των συνόρων των περιοχών αυτών. Τέτοιοι περιγραφείς είναι, για παράδειγμα, ο Περιγραφέας Περιήγησης Υφής (Texture Browsing Descriptor) και ο Περιγραφέας Ιστογράμματος Άκρων (Edge Histogram Descriptor).

1.4 Ανίχνευση Χαρακτηριστικών

Η ανίχνευση χαρακτηριστικών (feature detection) περιλαμβάνει μεθόδους για τον υπολογισμό συγκεκριμένων πληροφοριών των εικόνων και τη λήψη αποφάσεων σε κάθε σημείο της εικόνας για το αν υπάρχει ένα χαρακτηριστικό συγκεκριμένου τύπου το οποίο αναζητούν ή όχι. Τα χαρακτηριστικά που προκύπτουν από την ανίχνευση τους είναι υποσύνολα του περιεχομένου των εικόνων, συχνά με τη μορφή απομονωμένων σημείων (isolated points), συνεχών καμπυλών (continuous curves) ή συνδεδεμένων περιοχών (connected regions). Τα χαρακτηριστικά αυτά μπορεί να είναι άκρες, γωνίες (corners) ή σημεία ενδιαφέροντος (interest points) και περιοχές ενδιαφέροντος (binary large objects - blobs).

Οι **άκρες** είναι σημεία που οριοθετούν δύο περιοχές της εικόνας. Μία άκρη μπορεί να έχει οποιοδήποτε σχήμα ακόμη και να περιέχει διασταυρώσεις περιοχών. Πρακτικά, οι άκρες προσδιορίζονται σαν ένα σύνολο σημείων της εικόνας. Για παράδειγμα, μέθοδοι που εντοπίζουν άκρες είναι οι: Καμπυλότητα Καμπύλης Επιπέδου (Level Curve Curvature), Shi & Tomasi και Laplacian of Gaussian.

Οι **γωνίες**, ή αλλιώς σημεία ενδιαφέροντος, αναφέρονται σε χαρακτηριστικά που μοιάζουν με σημεία σε μία εικόνα και έχουν μια τοπική δομή δύο διαστάσεων. Γωνίες μπορούν να θεωρηθούν και τμήματα της εικόνας που δεν είναι γωνίες με την παραδοσιακή έννοια, όπως για παράδειγμα ένα φωτεινό σημείο σε σκοτεινό υπόβαθρο. Αυτά είναι τα σημεία ενδιαφέροντος, αλλά έχει επικρατήσει ο όρος γωνία για την ονομασία τους. Οι μέθοδοι Canny, Sobel και Smallest Univalue Segment Assimilating Nucleus (SUSAN) αποτελούν παραδείγματα ανιχνευτών γωνιών.

Οι **περιοχές ενδιαφέροντος (blobs)** περιγράφουν τις δομές της εικόνας σε επίπεδο περιοχών οι οποίες αποτελούνται από πολλά σημεία. Οι περιγραφείς περιοχών ενδιαφέροντος έχουν τη δυνατότητα να περιέχουν και μόνο ένα σημείο, όχι δηλαδή απαραίτητα μόνο περιοχές. Έτσι υπερτερούν έναντι των ανιχνευτών γωνιών καθώς μπορούν και ανιχνεύουν περιοχές της εικόνας που είναι πολύ ομαλές και ενδεχομένως δεν μπορούν να ανιχνευτούν από τους ανιχνευτές γωνιών. Αυτό είναι και το χαρακτηριστικό που ουσιαστικά τους διαφοροποιεί. Παραδείγματα μεθόδων ανίχνευσης περιοχών ενδιαφέροντος είναι οι Features from Accelerated Segment Test (FAST), Maximally Stable External Regions (MSER) και Difference of Gaussians (DoG).

Αφού εντοπιστούν τα χαρακτηριστικά, αυτά ποσοτικοποιούνται στους περιγραφείς χαρακτηριστικών (feature descriptors) οι οποίοι ονομάζονται και διανύσματα χαρακτηριστικών (feature vectors).

1.5 Αντικείμενο και στόχοι της διπλωματικής εργασίας

Η εργασία στοχεύει στη μελέτη και την ανάπτυξη διαφορετικών μεθόδων εξαγωγής χαρακτηριστικών από χειρόγραφες ψηφιοποιημένες εικόνες και στην παρουσίαση μεθόδων αναγνώρισης της γλώσσας, δηλαδή ταξινόμησης, αναλόγως τα χαρακτηριστικά αυτά. Τα χαρακτηριστικά που εξάγονται από τις εικόνες είναι τα LBP [1] και ταξινομούνται με έναν αλγόριθμο ο οποίος είναι βασισμένος στη λογική NN, και τα SIFT [2] χαρακτηριστικά που η ταξινόμηση των εικόνων βασίζεται στον NBNN [3] και στον Local NBNN [4] αλγόριθμο.

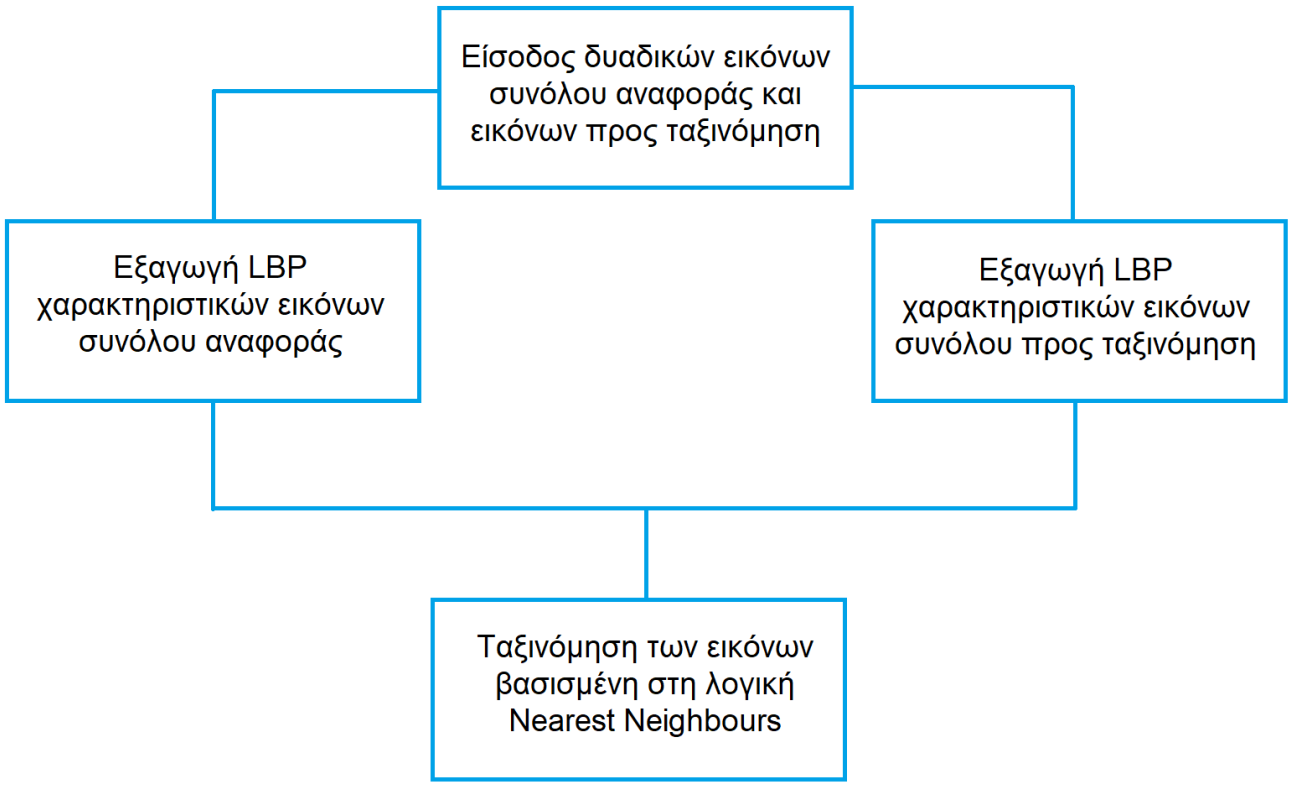
Όσον αφορά τη μέθοδο εξαγωγής χαρακτηριστικών LBP, μελετάται η εφαρμογή της πάνω σε δυαδικές εικόνες και παρουσιάζεται μια επέκταση της. Αυτή έχει βασιστεί στην εργασία [5] αλλά διαφοροποιείται ως προς τον τρόπο επιλογής των γειτονικών pixels και παρέχει βελτιωμένα αποτελέσματα σε σχέση με την αρχική μορφή των LBP. Η μέθοδος εξαγωγής των χαρακτηριστικών SIFT χρησιμοποιείται στην αρχική της μορφή.

Σχετικά με τους ταξινομητές, παρουσιάζεται ένας ταξινομητής βασισμένος στη λογική NN για τα LBP χαρακτηριστικά και γίνεται μελέτη πάνω στον τρόπο επιλογής των πλησιέστερων γειτόνων κατά την διαδικασία απόφασης της κλάσης. Για τους ταξινομητές NBNN και Local NBNN, που εφαρμόζονται με τα SIFT χαρακτηριστικά, παρουσιάζονται πρακτικές βελτίωσης τους ως προς την ταχύτητα αλλά και τα ποσοστά της επιτυχούς ταξινόμησής των εικόνων [6].

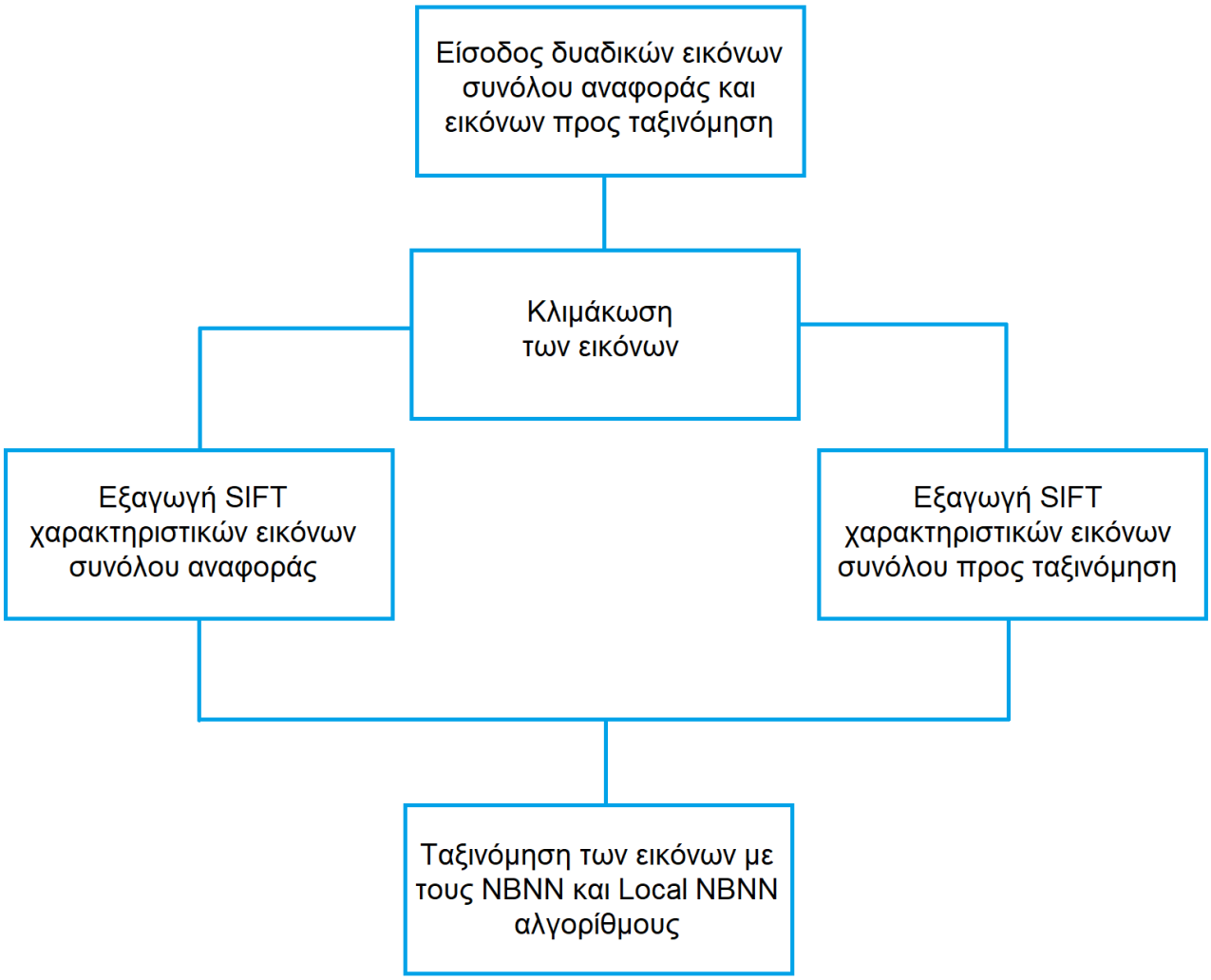
Το σύστημα που αναπτύσσεται απαιτεί την είσοδο εικόνων που είναι σε δυαδική μορφή καθώς οι μέθοδοι εξαγωγής των χαρακτηριστικών έχουν επικεντρωθεί στην εφαρμογή τους σε δυαδικές εικόνες. Επίσης, σκοπός της εργασίας είναι η αξιοποίηση αυτών των μεθόδων χωρίς προ-επεξεργασία των εικόνων, γεγονός που μειώνει την υπολογιστική πολυπλοκότητα του συστήματος αλλά και τους πόρους που απαιτούνται από αυτό.

1.6 Περιληπτική παρουσίαση των μεθόδων

Αρχικά εισάγονται στο σύστημα οι εικόνες που αποτελούν το σύνολο αναφοράς και το άγνωστο σύνολο, δηλαδή τις εικόνες προς ταξινόμηση. Έπειτα, υπολογίζονται τα LBP και SIFT χαρακτηριστικά των εικόνων που ανήκουν στο σύνολο αναφοράς και αποθηκεύονται σε συγκεκριμένες δομές. Στη συνέχεια, επαναλαμβάνεται η διαδικασία εξαγωγής των χαρακτηριστικών για τα LBP των άγνωστων εικόνων και οι εικόνες ταξινομούνται στις αντίστοιχες κλάσεις με τη μέθοδο ταξινόμηση KNN. Τέλος, εξάγονται τα SIFT χαρακτηριστικά του άγνωστου συνόλου και πραγματοποιείται ταξινόμησή τους με τον NBNN ή Local NBNN αλγόριθμο. Η βασική διαφορά των μεθόδων είναι ότι στη μέθοδο εξαγωγής των χαρακτηριστικών SIFT παρέχεται η δυνατότητα κλιμάκωσης των εικόνων εισόδου. Στις εικόνες 1 και 2 παρουσιάζονται τα βασικά στάδια κάθε τεχνικής.



Εικόνα 1: Βασικά στάδια μεθόδου ταξινόμησης με χρήση των LBP χαρακτηριστικών



Εικόνα 2: Βασικά στάδια μεθόδου ταξινόμησης με χρήση των SIFT χαρακτηριστικών

Η εργασία δομείται ως εξής: στο κεφάλαιο 2 αναφέρεται η τρέχουσα τεχνολογική στάθμη για την αναγνώριση γραφέα και γλώσσας σε έγγραφα. Στο κεφάλαιο 3 παρουσιάζονται τα Τοπικά Δυαδικά Πρότυπα (LBP χαρακτηριστικά), η υλοποίησή τους και η επέκταση τους ως προς την επιλογή των γειτονικών εικονοστοιχείων. Στο κεφάλαιο 4 παρουσιάζονται τα χαρακτηριστικά που προκύπτουν από τον αλγόριθμο SIFT και αναλύονται τα βασικά βήματα λειτουργίας του. Στο κεφάλαιο 5 περιγράφεται ο ταξινομητής των LBP χαρακτηριστικών που βασίζεται στη λογική NN και αναλύονται οι διαφορετικοί τρόποι επιλογής των πλησιέστερων γειτόνων κατά την εφαρμογή της μεθόδου. Επίσης, περιγράφονται οι ταξινομητές των SIFT χαρακτηριστικών και παρουσιάζονται δύο τεχνικές βελτίωσης των αποτελεσμάτων τους. Στο κεφάλαιο 6 περιγράφεται το Γραφικό Περιβάλλον, οι δυνατότητες της εφαρμογής που αναπτύχθηκε και παρουσιάζονται ορισμένα σενάρια λειτουργίας της. Στο κεφάλαιο 7 αναλύονται τα πειραματικά αποτελέσματα που προκύπτουν από τον έλεγχο της εφαρμογής στο σύνολο του δείγματος των εικόνων και συγκρίνονται όλα μεταξύ τους. Τέλος, στο κεφάλαιο 8 γίνεται αξιολόγηση των παραπάνω και παρατίθενται τα συμπεράσματα που προκύπτουν.

2. ΤΡΕΧΟΥΣΑ ΤΕΧΝΟΛΟΓΙΚΗ ΣΤΑΘΜΗ ΓΙΑ ΤΗΝ ΑΝΑΓΝΩΡΙΣΗ ΓΡΑΦΕΑ ΚΑΙ ΓΛΩΣΣΑΣ

2.1 Ταυτοποίηση γραφέα

Η ταυτοποίηση γραφέα (writer identification) είναι μια μέθοδος η οποία χρησιμοποιεί τεχνικές που εφαρμόζονται σε χειρόγραφα έγγραφα. Αποτελεί μέθοδο αναγνώρισης χειρογράφων και πραγματοποιείται με την αντιστοίχιση άγνωστων χειρογράφων με χειρόγραφα μιας βάσης δεδομένων, η οποία αποτελείται από δείγματα γνωστής προέλευσης γραφέα. Θεωρείται ένα πολύ σημαντικό θέμα έρευνας κι επομένως έχει μελετηθεί πολύ έχοντας μια ευρεία ποικιλία εφαρμογών, όπως είναι η ασφάλεια, η οικονομική δραστηριότητα και ο έλεγχος πρόσβασης. Ιδιαίτερα η ανάλυση χειρογράφων έχει μεγάλη σημασία για τα συστήματα ποινικής δικαιοσύνης.

Μέχρι στιγμής έχει χρησιμοποιηθεί μια ευρεία ποικιλία χαρακτηριστικών για το έργο της αναγνώρισης του γραφέα, όπως τα χαρακτηριστικά Quill στην εργασία [7], χαρακτηριστικά με βάση το διατρέχον μήκος (run-length) στην εργασία [8], χαρακτηριστικά βάσει περιγράμματος (contour) στην εργασία [9], καθώς και χαρακτηριστικά με βάση την υφή και την κλίση (gradient) στην εργασία [10]. Στη συνέχεια, ακολουθεί σύντομη παρουσίαση των μεθόδων στις οποίες χρησιμοποιήθηκαν τα παραπάνω χαρακτηριστικά για το πρόβλημα της αναγνώρισης του γραφέα.

Στην εργασία [7], χρησιμοποιούνται τα χαρακτηριστικά Quill που στην ουσία είναι μια κατανομή πιθανότητας (probability distribution) της σχέσης μεταξύ της κατεύθυνσης (direction) και του πλάτους (width) του μελανιού (ink) πάνω σε χειρόγραφα έγγραφα. Τα σύνολα δεδομένων εικόνων που χρησιμοποιούνται για την αξιολόγηση της απόδοσης της μεθόδου που προτείνουν, είναι δύο σύνολα τα οποία αποτελούνται από μεσαιωνικά χειρόγραφα έγγραφα και άλλα δύο που περιέχουν πιο σύγχρονα έγγραφα.

Πριν το βήμα της ταξινόμησης, πραγματοποιείται προ-επεξεργασία (pre-processing) των εικόνων χρησιμοποιώντας τεχνικές περικοπής περιοχών του κειμένου και μεθόδους κατωφλίωσης (thresholding). Τα μεσαιωνικά έγγραφα απαιτούν περεταίρω προ-επεξεργασία κι εφαρμόζονται σε αυτά οι εξής τεχνικές: επιλογή μικρών περιοχών κειμένου, διόρθωση της προοπτικής (perspective), αυτόματη κλιμάκωση (scaling) και φιλτράρισμα υψηλής διέλευσης (high pass filtering).

Όσον αφορά το στάδιο της αναγνώρισης, η ταξινόμηση βασίζεται στον ταξινομητή των πλησιέστερων γειτόνων και δεν περιέχει το βήμα της εκμάθησης (learning). Πριν την ταξινόμηση όμως εφαρμόζονται ορισμένες μέθοδοι για τον υπολογισμό του βέλτιστου συνδυασμού των τιμών των παραμέτρων που περιέχουν τα χαρακτηριστικά Quill. Τα ποσοστά ορθής αναγνώρισης γραφέα κυμαίνονται μεταξύ των τιμών 63% και 95%.

Αποδεικνύουν ότι ο συνδυασμός των μετρήσεων της κατεύθυνσης και του πλάτους του ίχνους ενός γραφέα είναι μοναδικά. Επίσης, η εργασία ενισχύει τη θεμελίωση ότι η γωνία και το πλάτος του ίχνους του μελανιού, χαρακτηριστικά που παρουσιάζονται κατά κόρον στην παλαιογραφία, αποτελούν ισχυρή πηγή πληροφόρησης για τον εντοπισμό γραφέα. Θεμελίωσαν λοιπόν την αξία των μετρήσεων του πλάτους του ίχνους στην αναγνώριση γραφέα, πέρα από την αξία της κατεύθυνσης του που ήταν ήδη θεμελιωμένη από το παρελθόν.

Στην εργασία [8], παρουσιάζεται το χαρακτηριστικό General Pattern Run-Length Transform (GPRLT) που είναι βασισμένο στο κείμενο (textural-based) και στην ουσία είναι το ιστόγραμμα του διατρέχοντος μήκους των πολύπλοκων μοτίβων. Το διατρέχον μήκος είναι ένα άθροισμα από διαδοχικά εικονοστοιχεία, τα οποία έχουν την ίδια τιμή. Το χαρακτηριστικό αυτό έχει δύο διαφορετικές μορφές, εκ των οποίων η πρώτη αναφέρεται σε δυαδικές εικόνες και η δεύτερη σε εικόνες γκρίζας κλίμακας. Χρησιμοποιείται χωρίς την εφαρμογή τμηματοποίησης στις εικόνες.

Τα χαρακτηριστικά GPRLT αποτελούνται από ορισμένες παραμέτρους και πριν το στάδιο της ταξινόμησης εφαρμόζουν μεθόδους ώστε να εντοπίσουν τον βέλτιστο συνδυασμό των τιμών τους. Για την ταξινόμηση των εγγράφων χρησιμοποιούν τον ταξινομητή πλησιέστερων γειτόνων. Αφού υπολογιστούν οι αποστάσεις μεταξύ των εικόνων προς ταξινόμηση και των εικόνων του συνόλου αναφοράς, αποδίδεται σε αυτά ο γραφέας στον οποίο ανήκουν.

Τα σύνολα δεδομένων που χρησιμοποιούν περιέχουν έγγραφα γραμμένα στην κινέζικη γλώσσα, έγγραφα γραμμένα στην αγγλική γλώσσα και έγραφα που περιέχουν και κινέζικους και αγγλικούς χαρακτήρες. Επίσης, εφαρμόζουν τη μέθοδό τους και σε ένα σύνολο εικόνων που περιέχει έγγραφα στην ελληνική και την αγγλική γλώσσα.

Στα πειράματά τους δείχνουν ότι το χαρακτηριστικό GPRLT αποδίδει πολύ καλύτερα αποτελέσματα στο πρόβλημα αναγνώρισης γραφέα για τις εικόνες γκρίζας κλίμακας παρά για τις δυαδικές εικόνες. Επίσης, πετυχαίνουν καλύτερη διάκριση των γραφέων σε σχέση με τις παραδοσιακές τεχνικές τρέχοντος μήκους. Το ποσοστό επιτυχούς ταξινόμησης είναι ίσο με 75.2% για τον διαχωρισμό των δυαδικών εικόνων και 91.4% για τον διαχωρισμό των εικόνων γκρίζας κλίμακας.

Στην εργασία [9] προτείνουν μία μέθοδο ανίχνευσης διασταυρώσεων (junction detection) σε χειρόγραφες εικόνες. Η μέθοδος αυτή χρησιμοποιεί την κατανομή του μήκους διαδρομής σταθερού πλάτους (stroke-length) προς όλες τις κατευθύνσεις γύρω από ένα σημείο αναφοράς, το οποίο βρίσκεται εντός του μελανιού στο κείμενο του εγγράφου. Η τεχνική αυτή παρέχει ένα χαρακτηριστικό που στην ουσία αποτελεί έναν περιγραφέα τοπικής πληροφορίας.

Για την αξιολόγηση της απόδοσης της μεθόδου που προτείνουν, παρουσιάζουν ένα νέο σύνολο δεδομένων από εικόνες το οποίο αποτελείται από έγγραφα γραμμένα στην αγγλική και την κινέζικη γλώσσα. Τα έγγραφα περιέχουν διαφορετικούς τρόπους γραφής που έχουν δημιουργηθεί από τους ίδιους γραφείς. Για το πρόβλημα της αναγνώρισης γραφέα εφαρμόζουν τη μέθοδο ανίχνευσης διασταυρώσεων στο σύνολο αυτό, το οποίο στην ουσία αποτελεί αναπαράσταση ενός βιβλίου κωδίκων (codebook-based) που έχει εκπαιδευτεί από τις ανιχνευθείσες διασταυρώσεις.

Επίσης, χρησιμοποιούν σύνολα δεδομένων από ιστορικά και πιο σύγχρονα χειρόγραφα έγγραφα για την αξιολόγηση του περιγραφέα που δημιουργείται με την τεχνικής τους. Τα ποσοστά επιτυχούς αναγνώρισης είναι διαφορετικά για κάθε σύνολο δεδομένων εικόνων και κυμαίνονται μεταξύ των τιμών 80% και 98.5%.

Τα πειραματικά αποτελέσματα δείχνουν ότι η προτεινόμενη μέθοδος ανίχνευσης παραμένει αμετάβλητη σε αλλαγές περιστροφής και κλίμακας. Η επιτυχής αναγνώριση γραφέα αποδεικνύει ότι οι διασταυρώσεις αποτελούν σημαντικό στοιχείο χαρακτηρισμού των διαφορετικών τρόπων γραφής.

Στην εργασία [10] παρουσιάζεται μια τεχνική αναγνώρισης γραφέα χρησιμοποιώντας τα SIFT χαρακτηριστικά και το χαρακτηριστικό κατεύθυνσης περιγράμματος (contour directional feature - CDF). Η προτεινόμενη μέθοδος περιλαμβάνει δύο στάδια.

Στο πρώτο στάδιο κατασκευάζεται ένα βιβλίο κωδίκων τοπικών προτύπων υφής, το οποίο δημιουργείται μέσω της συσταδοποίησης (clustering) ενός συνόλου από SIFT περιγραφείς που εξάγονται από τις εικόνες. Χρησιμοποιώντας αυτό το βιβλίο κωδίκων, υπολογίζονται οι ομοιότητες μεταξύ των εικόνων με τη χρήση ιστογραμμάτων εμφάνισης (occurrence histograms). Το αποτέλεσμα για κάθε εικόνα προς ταξινόμηση είναι μια λίστα από υποψήφιες εικόνες που ανήκουν στο σύνολο αναφοράς.

Το επόμενο στάδιο είναι η βελτίωση της λίστας των υποψήφιων εικόνων και επιτυγχάνεται με τη χρήση των χαρακτηριστικών κατεύθυνσης περιγράμματος και του SIFT περιγραφέα. Η μέθοδος εφαρμόζεται σε δύο σύνολα δεδομένων εικόνων και τα πειραματικά αποτελέσματα δείχνουν ότι η προτεινόμενη τεχνική υπερβαίνει κορυφαίους αλγορίθμους που έχουν προταθεί στο παρελθόν για το πρόβλημα της αναγνώρισης γραφέα.

2.2 Ταυτοποίηση γλώσσας

Η ταυτοποίηση της γλώσσας (language identification) είναι μια αντίστοιχη μέθοδος με την ταυτοποίηση γραφέα που βασίζεται όμως στην αναγνώριση και την αντιστοίχιση της γλώσσας των δειγμάτων. Εφαρμόζεται σε συστήματα αναγνώρισης εγγράφων όπως είναι οι ψηφιακές βιβλιοθήκες, δεδομένου ότι τέτοιου είδους μηχανισμοί αναγνώρισης απαιτούν την ενσωμάτωση γλωσσικών μοντέλων ώστε να αυξηθεί η απόδοσή τους.

Τα χαρακτηριστικά που έχουν χρησιμοποιηθεί μέχρι στιγμής για την αναγνώριση της γλώσσας, είναι τα χαρακτηριστικά με βάση το σχήμα στην εργασία [11] και τα χαρακτηριστικά υφής αμετάβλητης περιστροφής (rotation invariant texture features) στην εργασία [12]. Επίσης, προτάθηκε ένα σχήμα κωδικοποίησης λέξης (word shape coding) στην εργασία [13] με σκοπό την ταυτοποίηση της γλώσσας, το οποίο και επεκτάθηκε στην εργασία [14]. Στη συνέχεια, ακολουθεί σύντομη παρουσίαση των μεθόδων στις οποίες χρησιμοποιήθηκαν τα παραπάνω χαρακτηριστικά για το πρόβλημα της αναγνώρισης γλώσσας.

Στην εργασία [11] παρουσιάζεται ένα σύστημα αναγνώρισης του είδους γραφής (script) και της γλώσσας σε εικόνες χειρόγραφων εγγράφων. Αναπτύχθηκε χρησιμοποιώντας ένα σύνολο δεδομένων εγγράφων που αποτελείται από 6 διαφορετικά είδη γραφής, 8 γλώσσες και 281 γραφείς.

Σύμφωνα με την τεχνική που προτείνουν, τα έγγραφα χαρακτηρίζονται από τη μέση τιμή, την τυπική απόκλιση και τη στροφή (skew) πέντε ορισμένων χαρακτηριστικών συνδεδεμένων συστατικών (connected component features). Ένα συνδεδεμένο συστατικό είναι το βασικότερο αντικείμενο ομαδοποίησης των εικονοστοιχείων ενδιαφέροντος. Τα χαρακτηριστικά των συνδεδεμένων συστατικών που χρησιμοποιήθηκαν είναι το σχετικό κέντρο Y (relative Y centroid), το σχετικό κέντρο X, ο αριθμός των λευκών τρυπών (white holes) που περιέχονται στους χαρακτήρες, η σφαιρικότητα (sphericity) και η αναλογία απεικόνισης (aspect ratio). Δημιουργείται έτσι ένα διάνυσμα χαρακτηριστικών 15 στοιχείων για κάθε έγγραφο.

Πριν το στάδιο της αναγνώρισης, πραγματοποιείται προ-επεξεργασία σε κάθε έγγραφο ώστε να αφαιρεθούν στίγματα, διαχωριστικές γραμμές και στοιχεία των εγγράφων που είναι πολύ

μεγάλα σε μέγεθος. Στη συνέχεια, εξάγονται τα πέντε παραπάνω χαρακτηριστικά και υπολογίζεται η μέση τιμή, η τυπική απόκλιση και η στροφή για κάθε ένα από αυτά.

Για την αναγνώριση του είδους γραφής και της γλώσσας, εκπαιδεύεται και χρησιμοποιείται μια γραμμική ανάλυση διακρίσεων (linear discriminant analysis) η οποία δοκιμάστηκε χρησιμοποιώντας διασταυρούμενη επικύρωση (cross-validation). Το ποσοστό ορθής ταξινόμησης των 6 διαφορετικών ειδών γραφής ισούται με 88% και το βέλτιστο ποσοστό αναγνώρισης της γλώσσας είναι κατά μέσο όρο ίσο με 85% και επιτεύχθηκε στον διαχωρισμό εγγράφων της αγγλικής και της γερμανικής γλώσσας.

Στην εργασία [12] η τεχνική που προτείνεται βασίζεται στην εξαγωγή χαρακτηριστικών υφής που είναι αμετάβλητα ως προς την περιστροφή με σκοπό την αναγνώριση του είδους γραφής εικόνων χειρογράφων. Ο υπολογισμός των χαρακτηριστικών αυτών βασίζεται σε μια επέκταση της δημοφιλούς πολυκαναλικής τεχνικής φιλτραρίσματος Gabor (multi-channel Gabor filtering technique) και η αποτελεσματικότητά της χρήσης τους ελέγχεται με 300 τυχαία περιστρεφόμενα δείγματα 15 υφών.

Τα πειράματα για την αξιολόγηση της απόδοσης της προτεινόμενης μεθόδου, πραγματοποιούνται σε 6 γλώσσες (Κινέζικα, Αγγλικά, Ελληνικά, Ρωσικά, Περσικά και Μαλαγιάλαμ) και τα αποτελέσματα αποδεικνύουν τις δυνατότητες που έχει η συγκεκριμένη προσέγγιση με βάση την υφή στην επιτυχή ταυτοποίηση του είδους γραφής.

Στην εργασία [13] παρουσιάζεται μια τεχνική αναγνώρισης γλώσσας εικόνων εγγράφων που είναι γραμμένα σε γλώσσες βασισμένες στη λατινική. Η προτεινόμενη μέθοδος προϋποθέτει την εφαρμογή τεχνικών τμηματοποίησης της εικόνας σε γραμμές και λέξεις και αναγνωρίζει τη γλώσσα μέσω ενός σχήματος κωδικοποίησης γλώσσας. Σύμφωνα με αυτό, κάθε εικόνα λέξης μετατρέπεται σε ένα σχήμα κωδικοποίησης λέξης και κατά συνέπεια μετατρέπει κάθε εικόνα εγγράφου σε ένα διάνυσμα χαρακτηριστικών.

Για κάθε γλώσσα που μελετάται, δημιουργείται ένα πρότυπο μέσω ενός συστήματος εκμάθησης και η γλώσσα του εγγράφου προς ταξινόμηση καθορίζεται σύμφωνα με την ομοιότητα μεταξύ του διανύσματος του άγνωστου εγγράφου και των πολλαπλών προτύπων γλωσσών που δημιουργήθηκαν. Η προτεινόμενη μέθοδος αναγνώρισης γλώσσας είναι γρήγορη, ακριβής και δεν επηρεάζεται από πιθανά σφάλματα κατάτμησης του κειμένου που ενδέχεται να συμβούν λόγω θορύβου (noise) ή άλλων τύπων υποβάθμισης (degradation) των εγγράφων.

Τα πειραματικά αποτελέσματα πραγματοποιούνται σε ένα σύνολο δεδομένων εικόνων εγγράφων που δημιουργήθηκε για τους σκοπούς της εργασίας και αποτελείται από 80 έγγραφα, γραμμένα σε 4 διαφορετικές γλώσσες που περιέχουν τουλάχιστον 15 γραμμές κειμένου το καθένα. Επίσης, δοκιμάστηκαν και διαφορετικές γραμματοσειρές. Η αναγνώριση της γλώσσας έχει μέσο ποσοστό επιτυχίας ίσο με 97.8%.

Στην εργασία [14] αναφέρεται μια τεχνική αναγνώρισης βασισμένη στη γνώση (knowledge-based) που διαφοροποιεί είδη γραφής και γλώσσες σε εικόνες εγγράφων, ακόμη κι αν αυτές περιέχουν θόρυβο ή γενικότερα είναι υποβαθμισμένη η ποιότητά τους. Σύμφωνα με τη μέθοδο που προτείνουν, η αναγνώριση των ειδών γραφής και των γλωσσών πραγματοποιείται μέσω της διαδικασίας μετατροπής των εγγράφων σε διανύσματα (document vectorization). Αυτό έχει σαν αποτέλεσμα τη μετατροπή κάθε εικόνας εγγράφου σε ένα διάνυσμα, το οποίο χαρακτηρίζει το σχήμα και τη συχνότητα των χαρακτήρων ή λέξεων που περιέχονται στα έγγραφα.

Η μέθοδός τους απαιτεί προ-επεξεργασία των εικόνων, ώστε να αντισταθμιστεί η υποβάθμιση του εγγράφου και να παραχθούν καθαρά (clean) έγγραφα κειμένου σε δυαδική μορφή, προτού προχωρήσουν στην εξαγωγή των χαρακτηριστικών. Τα βήματα προ-επεξεργασίας που ακολουθούν είναι η κατάτμηση του εγγράφου με σκοπό τον διαχωρισμό του κειμένου από το φόντο, η επισήμανση των δυαδικών στοιχείων μέσω της ανάλυσης συνδεδεμένων συστατικών (connected component analysis) και η αφαίρεση του θορύβου με τη χρήση διάφορων μεθόδων φιλτραρίσματος, αναλόγως το είδος του θορύβου.

Έπειτα, οι εικόνες των εγγράφων μετατρέπονται σε διανύσματα με τη χρήση των ακραίων σημείων (extremum points) των χαρακτήρων και του αριθμού των χαρακτήρων. Αυτά τα δύο χαρακτηριστικά έχουν τη δυνατότητα να παραμένουν αμετάβλητα κατά την παρουσία θορύβου στο έγγραφο και τη χρήση διαφορετικών γραμματοσειρών.

Για κάθε είδος γραφής ή γλώσσα που εξετάζεται, δημιουργείται ένα πρότυπο μέσω της διαδικασίας της εκμάθησης, το οποίο αποτελείται από ένα σύνολο εικόνων αναφοράς. Τα είδη γραφής και οι γλώσσες καθορίζονται στη συνέχεια ανάλογα με τις αποστάσεις μεταξύ των διανυσμάτων των άγνωστων εικόνων εγγράφων και των πολλαπλών εκμαθημένων προτύπων. Τα πειράματα πραγματοποιούνται πάνω σε ένα σύνολο εγγράφων που περιέχει 6 διαφορετικά είδη γραφής και 8 γλώσσες που βασίζονται στη λατινική.

Τα αποτελέσματά δείχνουν ότι η μέθοδός τους είναι γρήγορη και ακριβής ακόμη κι αν τα έγγραφα είναι υποβαθμισμένα ως προς την ποιότητά τους και περιέχουν διάφορες μορφές θορύβου. Παρουσιάζουν πειραματικά αποτελέσματα για όλες τις διαφορετικές μορφές θορύβου που μελετούν και τα ποσοστά επιτυχούς αναγνώρισης κυμαίνονται μεταξύ 93.48% και 96.88% για τα είδη γραφής και μεταξύ 96.34% και 98.87% για τις διαφορετικές γλώσσες.

3. ΤΟΠΙΚΑ ΔΥΑΔΙΚΑ ΠΡΟΤΥΠΑ

3.1 Γενική έννοια των Τοπικών Δυαδικών Προτύπων

Τα Τοπικά Δυαδικά Πρότυπα (Local Binary Patterns - LBP) [1] είναι ένας απλός τύπος οπτικών περιγραφέων υφής που μπορούν να περιγράψουν την τοπική δομή των εικόνων. Εφαρμόζονται επιτυχώς σε πολλά προβλήματα μηχανικής όρασης καθώς και σε προβλήματα Ανάλυσης Εικόνων Εγγράφων (Document Image Analysis), συμπεριλαμβανομένου της οπτικής αναγνώρισης γραμματοσειρών και της αναγνώρισης γραφέων.

Τα Τοπικά Δυαδικά Πρότυπα επισημαίνουν τα εικονοστοιχεία μιας εικόνας ελέγχοντας μια γειτονία από εικονοστοιχεία ως προς ένα όριο (threshold) και δίνουν σαν αποτέλεσμα έναν δυαδικό αριθμό. Η πιο σημαντική ιδιότητά τους είναι η αντοχή τους στις μονοτονικές μεταβολές γκρίζας κλίμακας (gray-scale) που προκαλούνται, για παράδειγμα, από αλλαγές στον φωτισμό. Ένα άλλο εξίσου σημαντικό χαρακτηριστικό τους είναι η υπολογιστική απλότητά τους η οποία καθιστά δυνατή την ανάλυση εικόνων ακόμη και σε προκλητικές ρυθμίσεις που συμβαίνουν σε πραγματικό χρόνο.

Η βασική ιδέα ανάπτυξης των LBP βασίστηκε στο γεγονός ότι οι δισδιάστατες επιφανειακές υφές μπορούν να περιγραφούν με δύο συμπληρωματικά μέτρα: τα τοπικά χωρικά μοτίβα (local spatial patterns) και την αντίθεση της γκρίζα κλίμακας. Ο βασικός χειριστής LBP συγκρίνει την τιμή ενός κεντρικού εικονοστοιχείου με τις τιμές των εικονοστοιχείων της 3x3 γειτονίας του και δίνει ετικέτες στα γειτονικά. Το αποτέλεσμα είναι ένας δυαδικός αριθμός ο οποίος στη συνέχεια μετατρέπεται σε δεκαδικό και μπορεί να πάρει τιμές εντός του εύρους {0, …, 256} (2^8=256). Το ιστόγραμμα της συχνότητας εμφάνισης των τιμών αυτών χρησιμοποιείται σαν ένας περιγραφέας υφής (texture descriptor). Ακολουθεί ο τρόπος υλοποίησής τους.

3.2 Υλοποίηση των Τοπικών Δυαδικών Προτύπων

Το διάνυσμα περιγραφέα LBP στην απλούστερη του μορφή υπολογίζεται ως εξής:

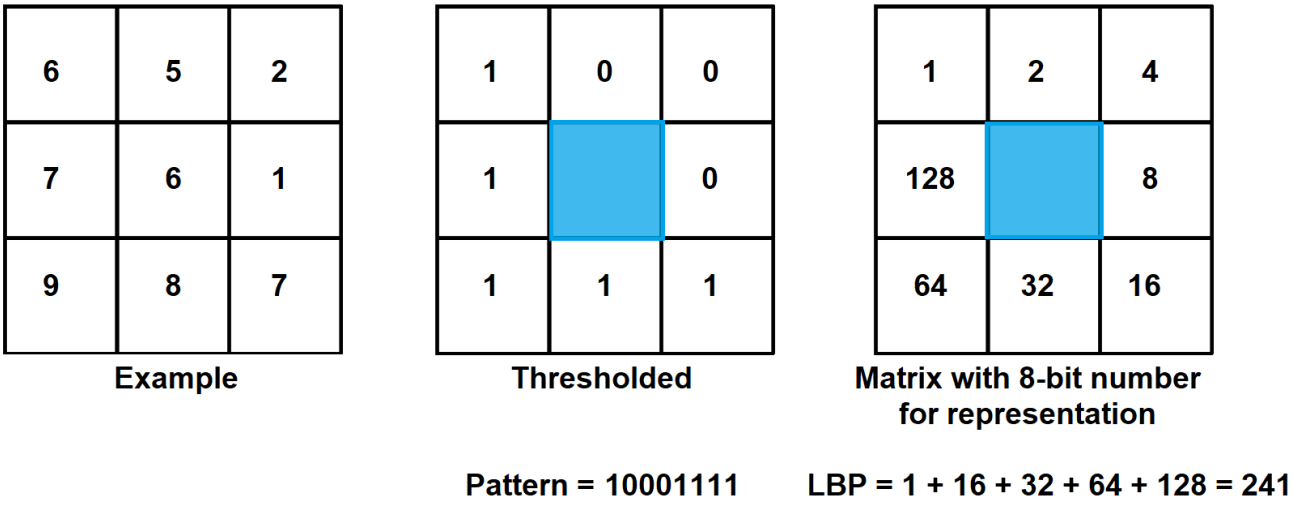
1. Υποδιαίρεση της εικόνας σε κελιά από συγκεκριμένο αριθμό εικονοστοιχείων.
2. Συγκρίνεται κάθε εικονοστοιχείο του κελιού με τα οχτώ γειτονικά του ακολουθώντας δεξιόστροφη κυκλική φορά.

   2.1. Αν η τιμή του κεντρικού εικονοστοιχείου είναι μεγαλύτερη από την τιμή του γειτονικού του, τότε το γειτονικό παίρνει την 0, αλλιώς την τιμή 1.

   2.2. Το τελικό αποτέλεσμα είναι ένας οκταψήφιος δυαδικός αριθμός, ο οποίος μετατρέπεται σε δεκαδικό.
3. Υπολογίζεται το ιστόγραμμα της συχνότητας κάθε αριθμού που προκύπτει για το συγκεκριμένο κελί και τελικά το αποτέλεσμα είναι ένα ιστόγραμμα 256 διαστάσεων που αποτελεί και το διάνυσμα χαρακτηριστικών.
4. Πραγματοποιείται προαιρετικά κανονικοποίηση του ιστογράμματος.
5. Ενώνονται τα ιστογράμματα όλων των κελιών σε ένα και προκύπτει το τελικό ιστόγραμμα του παραθύρου.

Το διάνυσμα χαρακτηριστικών μπορεί πλέον να χρησιμοποιηθεί από αλγορίθμους ταξινόμησης.

Στην εικόνα 3, παρουσιάζεται ένα παράδειγμα υπολογισμού του Τοπικού Δυαδικού Προτύπου ενός εικονοστοιχείου. Στο πρώτο σχήμα της εικόνας, φαίνονται οι τιμές του κεντρικού εικονοστοιχείου και των γειτονικών του. Στο μεσαίο σχήμα, παρουσιάζονται οι ετικέτες των γειτονικών εικονοστοιχείων που προκύπτουν μετά από τον έλεγχο των τιμών τους με την κεντρική τιμή. Στη συνέχεια, υπολογίζεται ο οκταψήφιος δυαδικός αριθμός λαμβάνοντας τις ετικέτες με δεξιόστροφη κυκλική φορά, ξεκινώντας από το εικονοστοιχείο που βρίσκεται στη γωνία πάνω-αριστερά. Στο τελευταίο σχήμα της εικόνας, παρουσιάζεται ο τρόπος μετατροπής του δυαδικού αριθμού σε δεκαδικό. Σύμφωνα με αυτόν, κάθε τιμή των εικονοστοιχείων πολλαπλασιάζεται με τον αντίστοιχο αριθμό του σχήματος και προκύπτει τελικά ο δεκαδικός αριθμός του Τοπικού Δυαδικού Προτύπου αθροίζοντας τα επιμέρους αποτελέσματα.

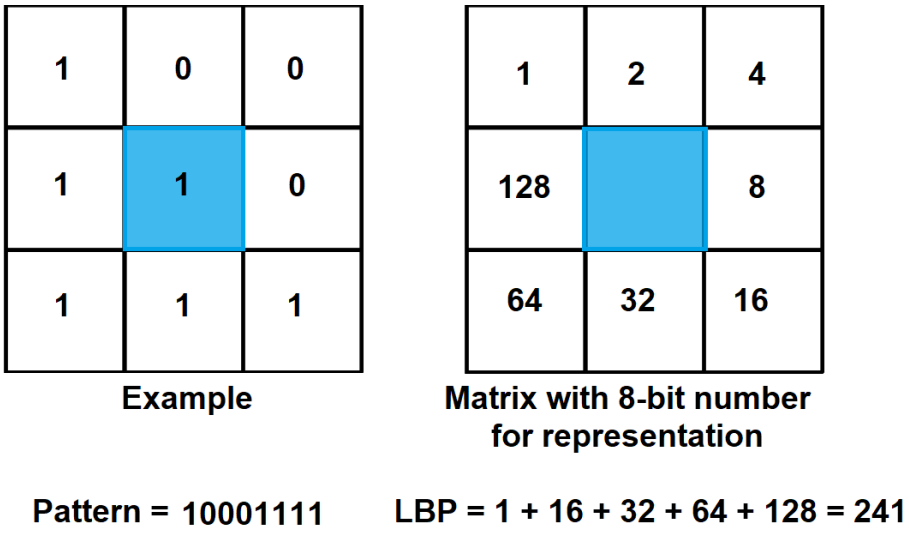


Εικόνα 3: Υπολογισμός LBP

3.3 Επέκταση των Τοπικών Δυαδικών Προτύπων

Ο αρχικός χειριστής LBP σχεδιάστηκε για εικόνες γκρίζου επιπέδου (gray-level). Η εφαρμογή του στις δυαδικές εικόνες μειώνει τον αριθμό των εικονοστοιχείων που αποτελούν σημεία ενδιαφέροντος καθώς λαμβάνονται υπόψη μόνο αυτά που έχουν την τιμή 1. Επίσης, απλοποιείται ο υπολογισμός των δυαδικών αριθμών των γειτονιών γιατί δεν πραγματοποιείται έλεγχος μεταξύ των τιμών τους και του κεντρικού, απλώς λαμβάνονται υπόψη μόνο τα pixels που έχουν την τιμή 1.

Στην εικόνα 4, παρουσιάζεται ένα παράδειγμα υπολογισμού του Τοπικού Δυαδικού Προτύπου σε δυαδική εικόνα. Στο πρώτο σχήμα της εικόνας φαίνονται οι τιμές του κεντρικού εικονοστοιχείου και των γειτονικών του. Ο οκταψήφιος δυαδικός αριθμός προκύπτει από τις τιμές τους με δεξιόστροφη κυκλική φορά, ξεκινώντας από το εικονοστοιχείο που βρίσκεται στη γωνία πάνω-αριστερά. Στο δεύτερο σχήμα παρουσιάζεται ο τρόπος μετατροπής του δυαδικού αριθμού σε δεκαδικό. Σύμφωνα με αυτόν, κάθε τιμή πολλαπλασιάζεται με τον αντίστοιχο αριθμό και στη συνέχεια αθροίζονται και προκύπτει ο δεκαδικός αριθμός του Τοπικού Δυαδικού Προτύπου.



Εικόνα 4: Υπολογισμός LBP σε δυαδική εικόνα

Η επέκταση του υπολογισμού των LBP χαρακτηριστικών [5] αφορά την αύξηση της ακτίνας (μέγιστη ακτίνα) απόστασης των γειτονικών εικονοστοιχείων και τον υπολογισμό των ιστογραμμάτων για όλες τις πιθανές τιμές της, μεταξύ της τιμής 1 και της μέγιστης τιμής. Προκύπτουν έτσι τόσα ιστογράμματα όσες είναι οι πιθανές τιμές της ακτίνας και τελικά ενώνονται όλα σε ένα ιστόγραμμα. Για παράδειγμα, έστω ότι η μέγιστη ακτίνα παίρνει την τιμή 5. Υπολογίζονται τα ιστογράμματα για το σύνολο των τιμών {1, 2, 3, 4, 5} και τελικά ενώνονται σε ένα ιστόγραμμα το οποίο έχει διαστάσεις 256*radius, όπου radius είναι η τιμή της μέγιστης ακτίνας, δηλαδή 5.

Τα βήματα υπολογισμού του νέου διανύσματος περιγραφέα είναι τα εξής:

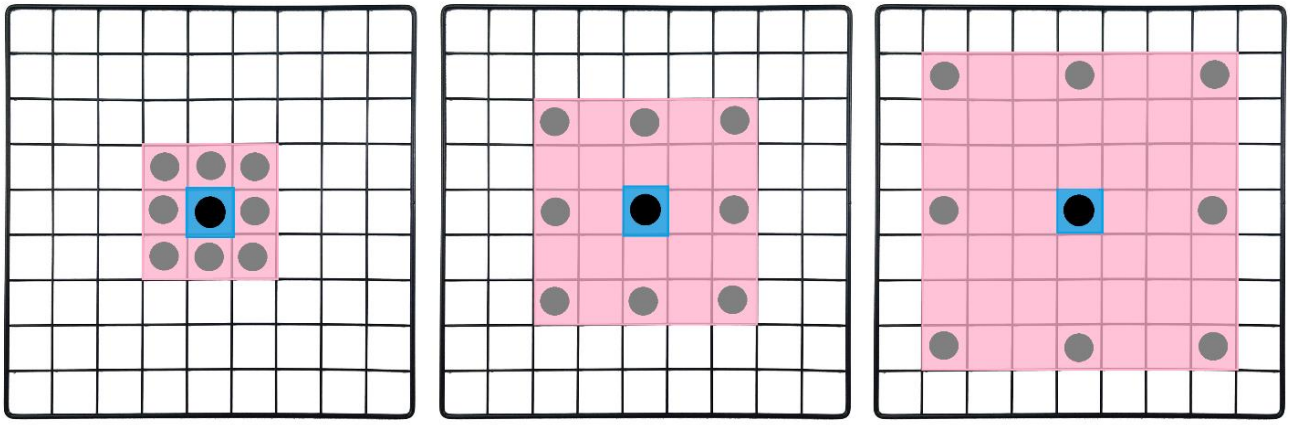
1. Υποδιαίρεση της εικόνας σε μικρότερα παράθυρα (windows) ίσου μεγέθους.
2. Έστω ότι η τιμή της ακτίνας ισούται με 1 και η μέγιστη τιμή της είναι radius=5.
3. Για κάθε pixel του παραθύρου που αποτελεί σημείο ενδιαφέροντος, έχει δηλαδή την τιμή 1, λαμβάνονται υπόψη τα οχτώ γειτονικά του που απέχουν από αυτό τόσο όσο η ακτίνα, ακολουθώντας δεξιόστροφη κυκλική φορά.

3.1 Υπολογίζεται ο οκταψήφιος δυαδικός αριθμός και μετατρέπεται σε δεκαδικό.

4. Υπολογίζεται το ιστόγραμμα της συχνότητας κάθε αριθμού που προκύπτει για το συγκεκριμένο παράθυρο και τελικά το αποτέλεσμα είναι ένα ιστόγραμμα 256 διαστάσεων.
5. Πραγματοποιείται προαιρετικά κανονικοποίηση του ιστογράμματος.
6. Επαναλαμβάνονται τα βήματα 3 ως 5 για όλες τις τιμές της ακτίνας με μέγιστη τιμή την τιμή radius.
7. Ενώνονται τα ιστογράμματα που προκύπτουν από όλες τις ακτίνες σε ένα ιστόγραμμα και το αποτέλεσμα είναι ένα διάνυσμα χαρακτηριστικών διάστασης 256*radius το οποίο αποτελεί και το τελικό διάνυσμα χαρακτηριστικών του συγκεκριμένου παραθύρου.

Στην εικόνα 5, παρουσιάζεται ο τρόπος αύξησης της ακτίνας κατά τον υπολογισμό των Τοπικών Δυαδικών Προτύπων. Στο πρώτο σχήμα η ακτίνα είναι ίση με 1, όπως ακριβώς και

στην απλή περίπτωση. Στο μεσαίο σχήμα η ακτίνα είναι ίση με 2, δηλαδή κάθε εικονοστοιχείο έχει απόσταση από το κεντρικό ίση με δύο εικονοστοιχεία. Στο τελευταίο σχήμα, η ακτίνα είναι ίση με την τιμή 3. Η τιμή της ακτίνας μπορεί να λάβει οποιαδήποτε τιμή.



Εικόνα 5: Επέκταση LBP

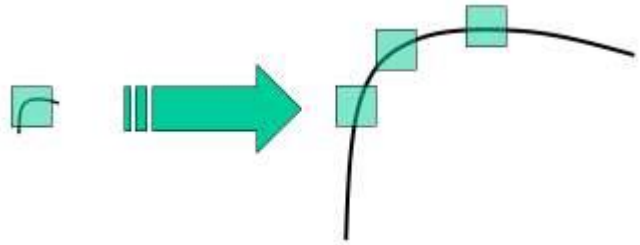
Τελικά, το διάνυσμα χαρακτηριστικών προκύπτει από την ένωση όλων των ιστογραμμάτων που προέρχονται από τις διαφορετικές τιμές της ακτίνας σε ένα ιστόγραμμα. Έτσι, το διάνυσμα αποκτά μεγαλύτερη ακρίβεια περιγραφής των χαρακτηριστικών της εικόνας και μπορεί να επιφέρει καλύτερα αποτελέσματα ταξινόμησης.

4. ΜΕΤΑΣΧΗΜΑΤΙΣΜΟΣ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ ΑΜΕΤΑΒΛΗΤΗΣ ΚΛΙΜΑΚΑΣ

4.1 Γενική έννοια του SIFT αλγορίθμου

Ο Μετασχηματισμός Χαρακτηριστικών Αμετάβλητης Κλίμακας (Scale Invariant Feature Transform - SIFT) είναι ένας αλγόριθμος ανίχνευσης χαρακτηριστικών που χρησιμοποιείται στη μηχανική όραση για ανίχνευση και περιγραφή των τοπικών χαρακτηριστικών των εικόνων. Έχει προταθεί από τον David Lowe [2] και με τη χρήση του εντοπίζονται σημεία ενδιαφέροντος (keypoints) στις εικόνες και υπολογίζονται οι περιγραφείς χαρακτηριστικών που αντιστοιχούν σε αυτά.

Τα SIFT χαρακτηριστικά έχουν την ιδιότητα να παραμένουν αμετάβλητα σε αλλαγές ως προς την κλίμακα, την περιστροφή και την τοποθεσία των αντικειμένων πάνω στην εικόνα. Αυτό έχει σαν αποτέλεσμα να μπορούν να ανιχνευτούν επιτυχώς ακόμη και στην περίπτωση που προκληθούν τέτοιου είδους αλλαγές. Για παράδειγμα, ο αλγόριθμος SIFT μπορεί και εντοπίζει γωνίες που έχει αλλάξει η κλίμακά τους, κάτι που άλλοι αλγόριθμοι δεν μπορούν να πετύχουν. Στην εικόνα 6, φαίνεται ένα παράδειγμα γωνίας και πως πρακτικά μεταβάλλεται αν αλλάξει η κλίμακα της.



Εικόνα 6: Αλλαγή κλίμακας γωνίας

4.2 Βασικά βήματα του SIFT αλγορίθμου

Τα βασικά βήματα του SIFT αλγορίθμου είναι τέσσερα και αναλύονται παρακάτω.

4.2.1 Ανίχνευση ακραίων χώρων κλίμακας

Από την εικόνα 6, είναι προφανές ότι δεν είναι δυνατόν να συσχετιστούν σημεία ενδιαφέροντος με διαφορετική κλίμακα. Πραγματοποιείται λοιπόν μια διαδικασία ώστε να ξεπεραστεί το πρόβλημα αυτό.

Αρχικά, υπολογίζεται ο χώρος κλίμακας (scale-space) της εικόνας που ορίζεται ως η συνάρτηση $L(x, y, \sigma)$ (τύπος 3.1) και ισούται με τη συνέλιξη της αρχικής εικόνας I με την Gaussian θόλωση (blur) $G(x, y, \sigma)$ (τύπος 3.2). Αυτό γίνεται για διάφορες τιμές του σ, το οποίο λειτουργεί σαν παράμετρος θόλωσης, που σημαίνει ότι όσο μεγαλύτερο είναι, τόσο περισσότερη είναι και η θόλωση της εικόνας.

$$L(x, y, \sigma) = G\,(\,x\,,y\,,\sigma\,) \times I(x, y) \qquad (3.1)$$

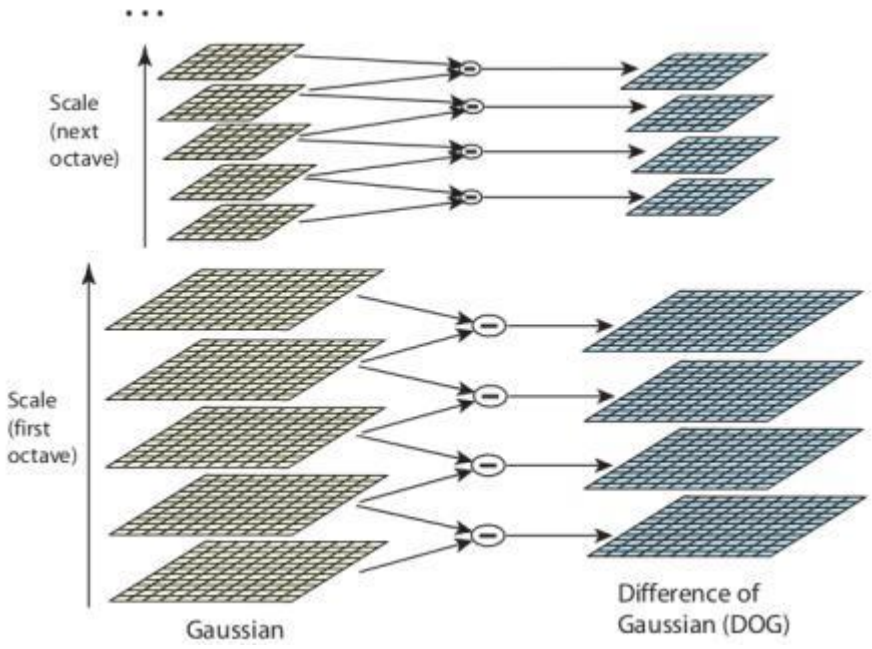
$$G(x, y, \sigma) = \frac{1}{2 \times \pi \times \sigma^2} e^{\frac{-(x^2+y^2)}{2\times\sigma^2}} \qquad (3.2)$$

Στη συνέχεια, για τον εντοπισμό των ακραίων χώρων κλίμακας υπολογίζεται η διαφορά των Gaussians (Difference of Gaussians) $D(x, y, \sigma)$ (τύπος 3.3) για δυο διαφορετικές τιμές του σ, έστω σ και $k * \sigma$. Το k είναι ένας σταθερός παράγοντας που στην ουσία διαχωρίζει τις διαδοχικές κλίμακες θόλωσης της εικόνας.

$$D(x, y, \sigma) = L\,(\,x\,, y\,, k * \sigma\,) - L\,(\,x\,, y\,, \sigma\,) \quad (3.3)$$

Η παραπάνω διαδικασία επαναλαμβάνεται για διάφορες οκτάβες (octaves), με τις οποίες εκφράζεται η μείωση του μεγέθους της εικόνας. Η γενική ιδέα είναι η προοδευτική θόλωση της εικόνας, η σμίκρυνσή της, η προοδευτική θόλωση της νέας εικόνας, η σμίκρυνσή της και ούτω καθεξής. Έπειτα, υπολογίζονται οι διαφορές των Gaussians.

Στην εικόνα 7, παρουσιάζεται μια απεικόνιση της παραπάνω διαδικασίας.



Εικόνα 7: Διαφορά Gaussians [2]

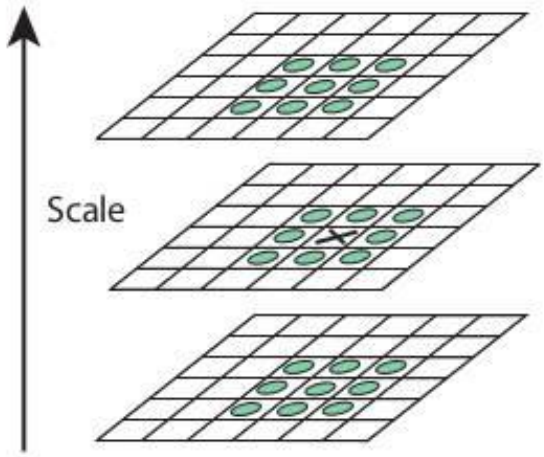
Όσον αφορά τις τιμές των παραμέτρων, σύμφωνα με την εργασία [2], έχουν προκύψει εμπειρικά και προτείνονται οι εξής: αριθμός από οκτάβες ίσος με 4, αριθμός των διαφορετικών επιπέδων θόλωσης ίσος με 5, αρχική τιμή του σ ίση με 1.6 και k ίσο με $\sqrt{2}$.

4.2.2 Εντοπισμός τοποθεσίας των keypoints

Σε αυτό το βήμα, αναζητούνται τα τοπικά μέγιστα (maxima) και ελάχιστα (minima) στους χώρους κλίμακας της διαφοράς των Gaussians $D(x, y, \sigma)$. Ο εντοπισμός τους επιτυγχάνεται συγκρίνοντας τα εικονοστοιχεία μεταξύ τους, κατά μήκος της κλίμακας θόλωσης των DoG, όπως παρουσιάζεται στην εικόνα 8.

Κάθε εικονοστοιχείο συγκρίνεται με τους οχτώ γείτονές του, καθώς επίσης και με τα εννιά αντίστοιχα εικονοστοιχεία της ίδιας εικόνας στην προηγούμενη κλίμακα, και με τα εννιά στην

επόμενη. Αν είναι τοπικό μέγιστο ή ελάχιστο, τότε θεωρείται υποψήφιο keypoint στην κλίμακα που εντοπίστηκε.



Εικόνα 8: Σύγκριση των pixels κατά μήκος της κλίμακας [2]

Οι τοποθεσίες των keypoints που εντοπίστηκαν είναι κατά προσέγγιση, καθώς τα τοπικά μέγιστα και ελάχιστα δεν βρίσκονται ποτέ ακριβώς πάνω σε εικονοστοιχεία. Βρίσκονται κάπου ανάμεσά τους, όπου είναι αδύνατη η πρόσβασή τους. Επομένως, πρέπει να εντοπιστούν με μαθηματικό τρόπο οι θέσεις των υποεικονοστοιχείων (subpixels) τους.

Σύμφωνα με την εργασία [15], δημιουργούνται οι τιμές των υποεικονοστοιχείων με την Taylor επέκταση της συνάρτησης χώρου κλίμακας $D(x, y, \sigma)$ που ανήκει το υποψήφιο keypoint (τύπος 3.4).

$$D\,(x) = D + \frac{\partial D^T}{\partial x} x + \frac{1}{2} x^T \frac{\partial^2 D}{\partial x^2} x \quad (3.4)$$

Το D και τα παράγωγά του υπολογίζονται στο υποψήφιο keypoint και $x = (x, y, \sigma)^T$ είναι η αντιστάθμιση (offset) από το keypoint. Η τοποθεσία του ακραίου σημείου (extremum) $\hat{x}$, υπολογίζεται παίρνοντας την παράγωγο της συνάρτηση 3.4 ως προς x και θέτοντάς την ίση με το 0 (τύπος 3.5).

$$\hat{x} = -\frac{\partial^2 D}{\partial x^2} \frac{\partial D}{\partial x}^{-1} \qquad (3.5)$$

Έτσι, εντοπίζονται τα ακραία σημεία, ή αλλιώς τα μέγιστα και τα ελάχιστα, στις συναρτήσεις χώρου κλίμακας $D(x, y, \sigma)$, τα οποία αποτελούν υποψήφια keypoints. Στη συνέχεια, κάποια από αυτά θα απορριφθούν, είτε γιατί δεν έχουν αρκετή αντίθεση (contrast), είτε γιατί βρίσκονται πάνω σε άκρες.

Η συνάρτηση τιμής $D(\hat{x})$ στο ακραίο σημείο (τύπος 3.6), είναι χρήσιμη για την απόρριψη των ασταθών (unstable) ακραίων σημείων που παρουσιάζουν χαμηλή αντίθεση. Αυτή προκύπτει αντικαθιστώντας την εξίσωση 3.5 στην 3.4.

$$D(\hat{x}) = D + \frac{1}{2} \frac{\partial D^T}{\partial x} \hat{x} \qquad (3.6)$$

Όλα τα ακραία σημεία με μέγεθος έντασης $|D(\hat{x})|$ μικρότερο από την τιμή κατώφλι 0.03 [2], απορρίπτονται. Το κατώφλι αυτό ονομάζεται κατώφλι αντίθεσης (contrast threshold).

Η διαφορά των Gaussians παρουσιάζει υψηλή απόκριση σε άκρες, οι οποίες πρέπει να αφαιρεθούν. Αυτό επιτυγχάνεται με την χρήση ενός $2x2$ Hessian πίνακα ώστε να υπολογιστούν οι κύριες καμπυλότητες (principal curvatures) στην τοποθεσία και την κλίμακα του keypoint (τύπος 3.7).

$$H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix} \qquad (3.7)$$

Σύμφωνα με την εργασία [15], ο πίνακας H και τα παράγωγα του D υπολογίζονται λαμβάνοντας υπόψη τις διαφορές των γειτονικών σημείων του keypoint.

Οι ιδιοτιμές του H είναι ανάλογες των κύριων καμπυλοτήτων του D και σύμφωνα με την εργασία [16], δε χρειάζεται να υπολογιστούν. Σημασία έχει μόνο ο λόγος της ιδιοτιμής με το μεγαλύτερο μέγεθος και της ιδιοτιμής με το μικρότερο. Έστω α η ιδιοτιμή με το μεγαλύτερο μέγεθος και β με το μικρότερο, τότε ο λόγος τους ισούται με $r = \alpha/\beta$.

Στη συνέχεια, υπολογίζεται το άθροισμα των ιδιοτιμών από το ίχνος (trace) του πίνακα H (τύπος 3.8) και το γινόμενό τους από την ορίζουσα του πίνακα (τύπος 3.9).

$$Tr(H) = D_{xx} + D_{yy} = \alpha + \beta \qquad (3.8)$$

$$Det(H) = D_{xx}D_{yy} - (D_{yy})^2 = \alpha\beta \qquad (3.9)$$

Αποδεικνύεται ότι ο λόγος $R = Tr(H)^2/Det(H)$, είναι ίσος με τον λόγο $(r+1)^2/r$ (τύπος 3.10).

$$R = \frac{Tr(H)^2}{Det(H)} = \frac{(\alpha+\beta)^2}{\alpha\beta} = \frac{(r\beta+\beta)^2}{r\beta^2} = \frac{(r+1)^2}{r} \qquad (3.10)$$

Τελικά, αν ο λόγος R των κύριων καμπυλοτήτων, για ένα υποψήφιο keypoint, είναι μεγαλύτερος από τον λόγο $(r_{th}+1)^2/r_{th}$, όπου r_{th} είναι ένα κατώφλι άκρων (edge threshold), τότε το συγκεκριμένο keypoint απορρίπτεται. Η κατάλληλη τιμή του κατωφλίου είναι η τιμή 10 [2].

Επομένως, εξαλείφονται όλα τα keypoints χαμηλής αντίθεσης και όσα βρίσκονται σε άκρες.

4.2.3 Ανάθεση προσανατολισμού

Σε αυτό το βήμα, εκχωρείται σε κάθε keypoint ένας ή περισσότεροι προσανατολισμοί. Πρόκειται για το βασικό στάδιο του αλγορίθμου SIFT, με το οποίο επιτυγχάνεται αμεταβλητότητα ως προς την περιστροφή.

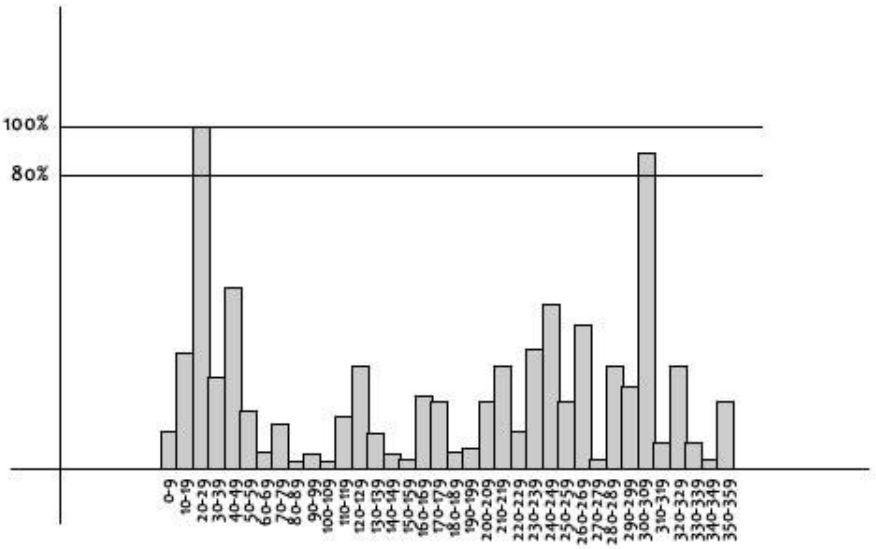
Αρχικά, λαμβάνεται υπόψη η Gaussian εξομαλυνθείσα εικόνα $L(x, y, \sigma)$ στην κλίμακα σ του keypoint, έτσι ώστε να εκτελεστούν όλοι οι υπολογισμοί με αμετάβλητο ως προς την κλίμακα τρόπο. Για όλα τα σημεία μιας επιλεγμένης περιοχής γύρω από το keypoint, υπολογίζεται το μέγεθος κλίσης (gradient magnitude) $m(x, y)$ και ο προσανατολισμός $\theta(x, y)$, σύμφωνα με τους τύπους 3.11 και 3.12 αντίστοιχα.

$$m(x,y) = \sqrt{(L(x+1,y) - L(x-1,y))^2 + (L(x,y+1) - L(x,y-1))^2} \qquad (3.11)$$

$$\theta(x,y) = tan^{-1}((L(x,y+1) - L(x,y-1))/(L(x+1,y) - L(x-1,y))) \qquad (3.12)$$

Προκύπτει λοιπόν ένα ιστόγραμμα προσανατολισμών που αποτελείται από 36 κάδους των 10 μοιρών, καλύπτοντας συνολικά 360 μοίρες. Για παράδειγμα, έστω ότι η κατεύθυνση της κλίσης για ένα σημείο είναι 18.75 μοίρες, τότε θα τοποθετηθεί στον κάδο των 10-19 μοιρών. Η ποσότητα που προστίθεται στους κάδους είναι ανάλογη του μεγέθους κλίσης των σημείων και ενός κυκλικού παραθύρου σταθμισμένου ως προς την Gaussian συνάρτηση, με σ που είναι 1.5 φορές μεγαλύτερο από ότι η κλίμακα σ στην οποία ανήκει το keypoint.

Στην εικόνα 9, παρουσιάζεται ένα παράδειγμα ιστογράμματος προσανατολισμών μετά την ολοκλήρωση της παραπάνω διαδικασίας. Η υψηλότερη κορυφή του βρίσκεται στον τρίτο κάδο, δηλαδή των 20-29 μοιρών, και η αμέσως επόμενη στον κάδο 31, δηλαδή των 300-309 μοιρών. Παρατηρείται ότι η δεύτερη κορυφή υπερβαίνει το 80% της τιμής της πρώτης, ενώ όλες οι υπόλοιπες βρίσκονται κάτω από αυτό.



Εικόνα 9: Ιστόγραμμα SIFT προσανατολισμών

Τελικά, το keypoint παίρνει τον προσανατολισμό της υψηλότερης κορυφής. Επίσης, όσες κορυφές είναι πάνω από το 80% της τιμής της υψηλότερης, οδηγούν στη δημιουργία ενός νέου keypoint. Το keypoint αυτό έχει την ίδια τοποθεσία και κλίμακα με το προηγούμενο αλλά διαφορετικό προσανατολισμό, ίσο με αυτόν της αντίστοιχης κορυφής.

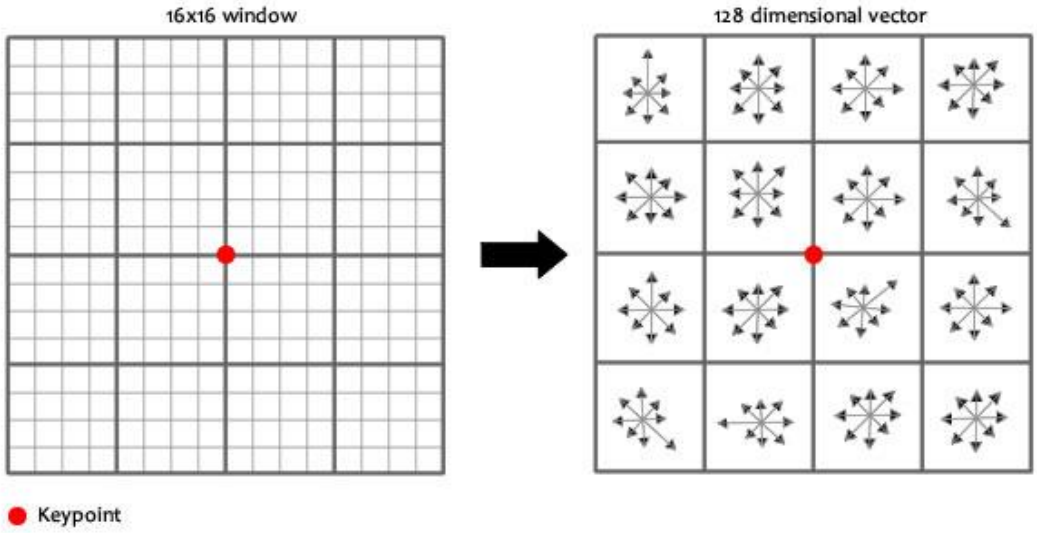
4.2.4 Περιγραφέας των keypoints

Σε αυτό το βήμα, υπολογίζεται το διάνυσμα περιγραφέα του keypoint.

Αρχικά, επιλέγεται μια περιοχή μεγέθους 16x16 γύρω από το keypoint και χωρίζεται σε 16 υπό-τετράγωνα μεγέθους 4x4. Για κάθε υπό-τετράγωνο, δημιουργείται ένα ιστόγραμμα προσανατολισμών 8 κάδων. Σε κάθε παράθυρο 4x4 υπολογίζονται τα μεγέθη κλίσης και οι προσανατολισμοί των εικονοστοιχείων και προστίθενται στο ιστόγραμμα. Η διαδικασία αυτή παρουσιάζεται στην εικόνα 10.

Για παράδειγμα, οι προσανατολισμοί που ανήκουν στην περιοχή 0-44 μοίρες, προστίθενται στον πρώτο κάδο. Οι προσανατολισμοί της περιοχής 45-89 στο επόμενο και ούτω καθεξής. Η ποσότητα που προστίθεται στους κάδους εξαρτάται από το μέγεθος της κλίσης και από

την απόσταση από το keypoint. Όσο μεγαλύτερη είναι η απόσταση, τόσο μικρότερη είναι και η ποσότητα.



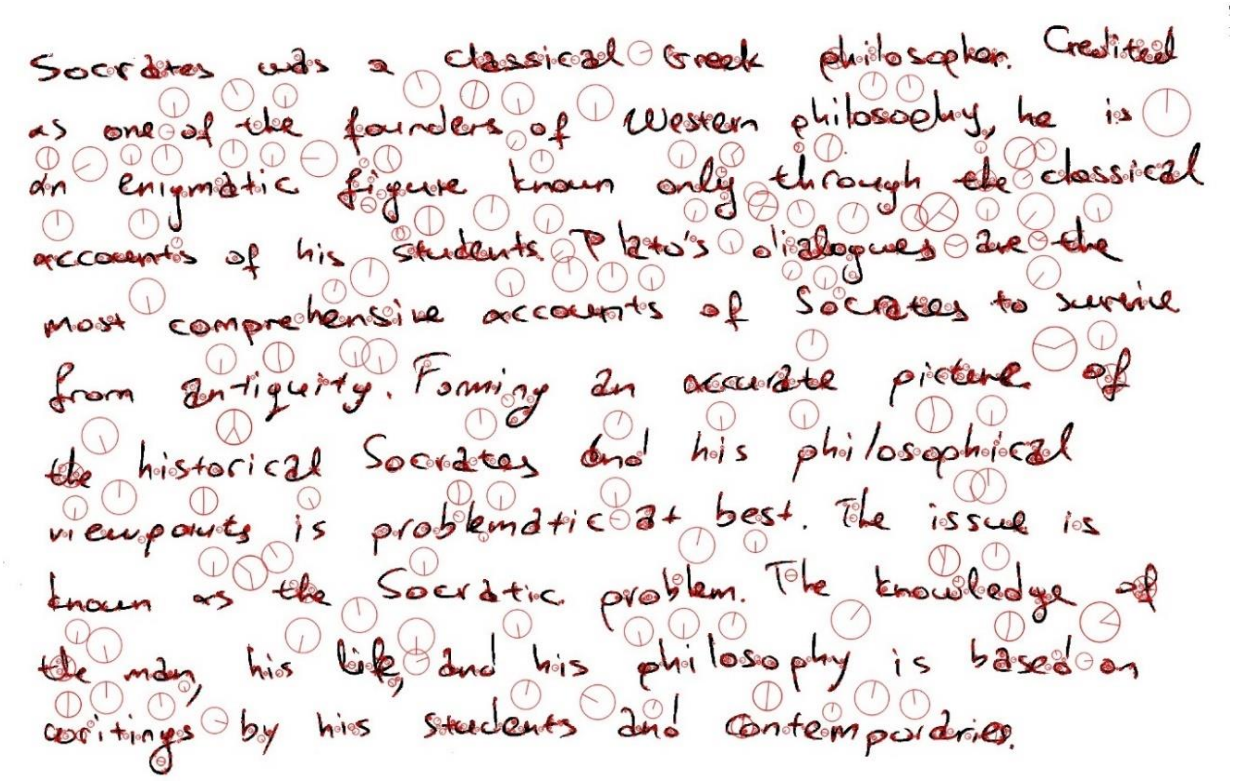
Εικόνα 10: Προσανατολισμοί στην περιοχή γύρω από το keypoint

Η παραπάνω διαδικασία πραγματοποιείται και για τις 16 περιοχές μεγέθους 4x4, καταλήγοντας έτσι σε 128 (16x8) αριθμούς. Προκύπτει έτσι το διάνυσμα χαρακτηριστικών του keypoint.

Στη συνέχεια, το διάνυσμα τροποποιείται ώστε να μειωθούν οι επιδράσεις της αλλαγής περιστροφής. Αυτό επιτυγχάνεται με την αφαίρεση της περιστροφής του keypoint από κάθε προσανατολισμό με αποτέλεσμα να σχετίζεται πλέον κάθε ένας με τον προσανατολισμό του keypoint.

Τέλος, τροποποιείται για να μειωθούν οι επιδράσεις της αλλαγής φωτισμού, μέσω της κανονικοποίησης του ως προς τη μονάδα. Έπειτα, όποιος αριθμός του είναι μεγαλύτερος από την τιμή 0.2 [2], παίρνει την τιμή αυτή και ξαναγίνεται κανονικοποίηση ως προς τη μονάδα. Το διάνυσμα πλέον έχει την τελική του μορφή.

Στην εικόνα 11, φαίνεται ένα παράδειγμα απεικόνισης των keypoints σε εικόνα χειρόγραφου εγγράφου. Τα keypoints που εντοπίστηκαν παρουσιάζονται με κύκλους με διαφορετικά κέντρα και διαφορετικές ακτίνες ως προς το μέγεθός και την κλίση τους. Τα κέντρα των κύκλων είναι οι συντεταγμένες των τοποθεσιών των keypoints, τα μήκη των ακτίνων τους αντικατοπτρίζουν το μέγεθος της κλίσης τους και οι κλίσεις των ακτίνων εκφράζουν τον προσανατολισμό τους.



Εικόνα 11: Απεικόνιση SIFT keypoints σε εικόνα

5. ΤΑΞΙΝΟΜΗΤΕΣ

Η αναγνώριση της γλώσσας σε ένα χειρόγραφο έγγραφο μπορεί να θεωρηθεί ως ένα πρόβλημα ταξινόμησης εικόνων όπου οι εικόνες αποτελούν τα δείγματα ψηφιοποιημένων εγγράφων και τα δείγματα που ανήκουν στην ίδια γλώσσα εκπροσωπούν μια κλάση. Υπάρχουν δύο ήδη ταξινομητών που μπορούν να εφαρμοστούν στο πρόβλημα αναγνώρισης της γλώσσας. Οι παραμετρικοί ταξινομητές που προϋποθέτουν την εκμάθησή τους και οι μη-παραμετρικοί που δεν απαιτούν το βήμα της εκμάθησης.

Οι παραμετρικοί ταξινομητές δημιουργούν ένα μοντέλο από το σύνολο αναφοράς των δειγμάτων (train set) και προσπαθούν να εκτιμήσουν παραμέτρους για το μοντέλο αυτό. Οι μη-παραμετρικοί συγκρίνουν κατευθείαν το σύνολο αναφοράς με το άγνωστο σύνολο των δειγμάτων (test set). Επιπλέον, οι παραμετρικοί ταξινομητές απαιτούν περισσότερους πόρους από τους μη παραμετρικούς και καθυστερούν την ταξινόμηση, καθώς πρέπει να προηγηθεί το βήμα της εκμάθησης. Αντίθετα, οι μη-παραμετρικοί ταξινομητές είναι σαφώς πιο γρήγοροι.

Οι ταξινομητές που χρησιμοποιούνται για την επίλυση του προβλήματος της αναγνώρισης γλώσσας στην παρούσα εργασία ανήκουν στην κατηγορία των μη παραμετρικών.

5.1 Ταξινομητής για τα LBP χαρακτηριστικά

Για την ταξινόμηση των εικόνων με τη χρήση των LBP χαρακτηριστικών χρησιμοποιείται ένας ταξινομητής βασισμένος στην εκτίμηση της απόστασης κοντινότερων γειτόνων (K Nearest Neighbour - KNN). Ειδικότερα, η λογική ταξινόμησης βασίζεται στις αποστάσεις των παραθύρων στα οποία διαχωρίζονται οι εικόνες και πιο συγκεκριμένα στις αποστάσεις των LBP χαρακτηριστικών των παραθύρων.

Αρχικά, διαχωρίζονται όλες οι εικόνες σε ίσο αριθμό μικρότερων τμημάτων (παράθυρα) κι έπειτα εξάγονται τα LBP χαρακτηριστικά από κάθε παράθυρο ξεχωριστά. Αυτό συμβαίνει και στο σύνολο αναφοράς των εικόνων και στο άγνωστο σύνολο. Έτσι, δημιουργείται μια βάση από LBP χαρακτηριστικά προς ταξινόμηση και μία βάση που αποτελείται από τα χαρακτηριστικά αναφοράς, είναι γνωστές δηλαδή οι κλάσεις στις οποίες ανήκουν.

Ο τρόπος ταξινόμησης μιας άγνωστης εικόνας είναι ο εξής:

- Για κάθε παράθυρο, υπολογίζεται η απόστασή του από όλα τα παράθυρα όλων των εικόνων του συνόλου αναφοράς. Η απόσταση ενός παραθύρου από ένα άλλο υπολογίζεται ως το άθροισμα των απόλυτων αποστάσεων κάθε μίας τιμής των LBP χαρακτηριστικών. Το σύνολο των τιμών των χαρακτηριστικών είναι 256 επί την ποσότητα radius, όπως παρουσιάστηκε στο υποκεφάλαιο 3.3.
- Γίνεται ταξινόμηση των αποστάσεων κατά φθίνουσα σειρά και επιλέγονται οι πλησιέστεροι γείτονες.
- Δίνονται οι αντίστοιχες ψήφοι για κάθε γλώσσα που ανήκουν οι πλησιέστεροι γείτονες.

Τα παραπάνω βήματα επαναλαμβάνονται για όλα τα παράθυρα της εικόνας προς ταξινόμηση και τελικά η εικόνα ταξινομείται στη γλώσσα με τις περισσότερες ψήφους.

5.1.1 Επιλογή των πλησιέστερων γειτόνων

Η επιλογή των πλησιέστερων γειτόνων των παραθύρων μπορεί να πραγματοποιηθεί με τέσσερις διαφορετικούς τρόπους, οι οποίοι παρουσιάζονται στην εικόνα 12 και επεξηγούνται παρακάτω.

Ο πρώτος τρόπος (0m) είναι η επιλογή του ενός κοντινότερου γείτονα κι έτσι δίνεται μία ψήφος ίση με τη μονάδα στη γλώσσα που ανήκει. Ο δεύτερος τρόπος (1m) είναι η επιλογή τόσων κοντινότερων γειτόνων όσων έχουν ίση ελάχιστη απόσταση με τον πρώτο κοντινότερο και αποδίδονται ψήφοι, ίσες με τη μονάδα, σε κάθε γλώσσα που ανήκουν. Ο τρίτος τρόπος (2m) είναι η επιλογή του συνόλου των κοντινότερων γειτόνων με τις 3 κοντινότερες αποστάσεις και η απόδοση ψήφων ίσων με τη μονάδα σε κάθε γλώσσα που ανήκουν. Ο τέταρτος και τελευταίος τρόπος (3m) είναι ίδιος με τον προηγούμενο, με τη διαφορά ότι οι ψήφοι δεν είναι ίσες μεταξύ τους, αλλά έχουν το αντίστοιχο βάρος σύμφωνα με την κατάταξή τους. Οι ψήφοι των πρώτων κοντινότερων γειτόνων ισούνται με τον αριθμό 3, άρα προστίθεται στις αντίστοιχες ψήφους ο αριθμός 3, οι ψήφοι των αμέσως επόμενων ισούνται με 2 και οι τελευταίες είναι ίσες με τη μονάδα.

Classifiers							
0m *(Only first)*		**1m** *(1-NN)*		**2m** *(k-NN)*		**3m** *(k-NN with Weight)*	
Αγγλικά	10	Αγγλικά	10	Αγγλικά	**10**	Αγγλικά	**10 (+k)**
Αγγλικά	10	Αγγλικά	10	Αγγλικά	10	Αγγλικά	10 (+k)
Γαλλικά	10	Γαλλικά	10	Γαλλικά	10	Γαλλικά	10 (+k-1)
Γερμανικά	11	Γερμανικά	11	Γερμανικά	**11**	Γερμανικά	**11 (+k-1)**
Ελληνικά	11	Ελληνικά	11	Ελληνικά	11	Ελληνικά	11 (+k-1)
Αγγλικά	12	Αγγλικά	12	Αγγλικά	**12**	Αγγλικά	**12 (+k-2)**
Αγγλικά	12	Αγγλικά	12	Αγγλικά	12	Αγγλικά	12 (+k-2)
Αγγλικά	12	Αγγλικά	12	Αγγλικά	12	Αγγλικά	12 (+k-2)
Γερμανικά	25	Γερμανικά	25	Γερμανικά	25	Γερμανικά	25
Γερμανικά	48	Γερμανικά	48	Γερμανικά	48	Γερμανικά	48

Εικόνα 12: Τρόποι επιλογής πλησιέστερων γειτόνων

5.2 Ταξινομητές για τα SIFT χαρακτηριστικά

Στην ταξινόμηση των εικόνων με τη χρήση των SIFT χαρακτηριστικών, οι ταξινομητές που βασίζονται στην εκτίμηση της απόστασης του κοντινότερου γείτονα, και ειδικότερα ο ταξινομητής Naive Bayes Nearest-Neighbour (NBNN) [3], μπορούν να δώσουν βελτιωμένα αποτελέσματα, αν αποφευχθούν δύο πρακτικές που οδηγούν στην υποβάθμιση των μεθόδων ταξινόμησης.

Η πρώτη πρακτική που πρέπει να αποφευχθεί, είναι η μείωση του αριθμού των περιγραφέων που περιγράφουν την εικόνα. Η μείωση του αριθμού τους έχει σκοπό να παραμείνουν οι πιο αντιπροσωπευτικοί, κάτι το οποίο μπορεί να προκαλέσει μεγάλη απώλεια πληροφορίας για τους μη παραμετρικούς αλγορίθμους, οι οποίοι δεν έχουν το στάδιο της εκμάθησης για να μπορέσουν να ανταποκριθούν στην απώλεια αυτή.

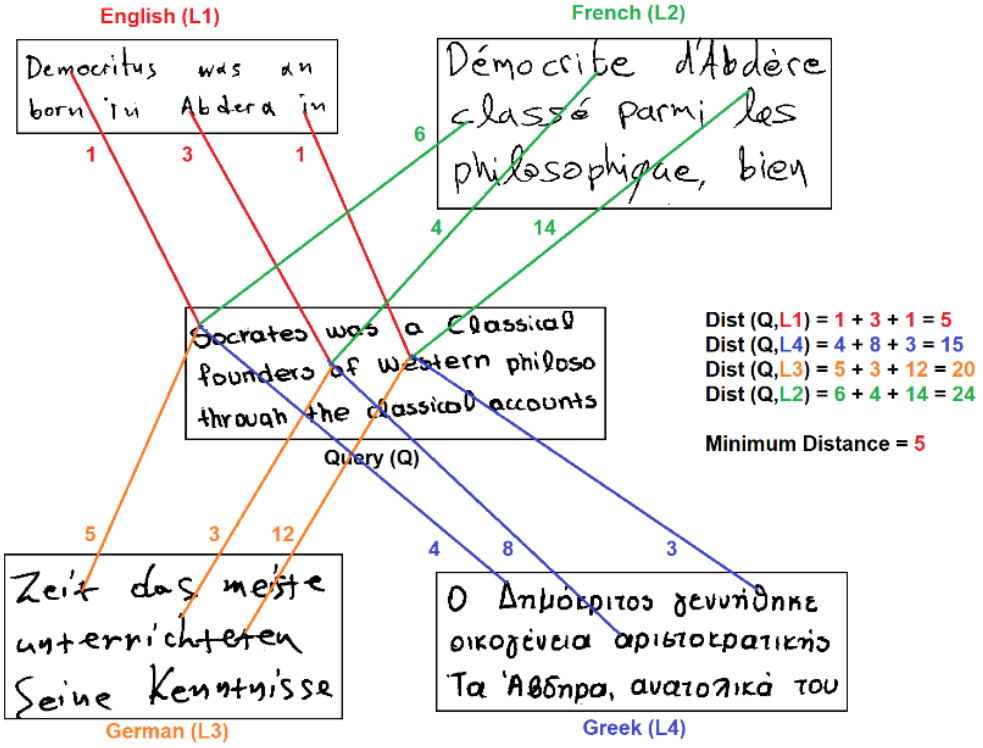
Η δεύτερη πρακτική αφορά την απόσταση των εικόνων, όπου στην ουσία υπολογίζονται οι αποστάσεις μεταξύ τους. Σε αντίθεση με αυτή την πρακτική, ο υπολογισμός της απόστασης των εικόνων από τις κλάσεις (οι κλάσεις είναι οι γλώσσες) γενικεύει την αναζήτηση του κοντινότερου γείτονα σε αντιστοίχιση κλάσης αντί για αντιστοίχιση εικόνας. Οι ταξινομητές που δεν έχουν το στάδιο της εκμάθησης ανταποκρίνονται πολύ καλύτερα με αυτόν τον τρόπο σε διακυμάνσεις που συμβαίνουν εντός των κλάσεων.

5.2.1 Ταξινομητής Naive Bayes Nearest Neighbour

Ο ταξινομητής NBNN [3] έχει προταθεί για το πρόβλημα ταξινόμησης φυσικών σκηνών. Αποδείχθηκε ότι οι υπό συνθήκη πιθανότητες των κλάσεων μπορούν να προσεγγιστούν επιτυχώς από την τετράγωνη Ευκλείδεια απόσταση του περιγραφέα προς ταξινόμηση από το πλησιέστερο διάνυσμα χαρακτηριστικών που ανήκει στη σωστή κλάση (τύπος 5.1). Με άλλα λόγια, αρκεί να βρεθεί η κλάση με την ελάχιστη Ευκλείδεια απόσταση των χαρακτηριστικών της από εκείνα της άγνωστης εικόνας.

$$c = \ argmin_C \left[\sum_{i=1}^{n} \left\| d_i - NN_C(d_i) \right\|^2 \right] \quad (5.1)$$

Μια απεικόνιση του NBNN ταξινομητή φαίνεται στην εικόνα 13, χρησιμοποιώντας εικόνες από ψηφιοποιημένα έγγραφα. Ο NBNN αλγόριθμος αναζητεί τον κοντινότερο γείτονα κάθε περιγραφέα της εικόνας προς ταξινόμηση Q (στην εικόνα φαίνονται μόνο 3 περιγραφείς) σε όλες τις κλάσεις (γλώσσες) οι οποίες είναι οι $L1, L2, L3, L4$. Στη συνέχεια ο αλγόριθμος υπολογίζει τις αποστάσεις από κάθε κλάση ξεχωριστά και τελικά η εικόνα Q ταξινομείται στην κλάση με τη μικρότερη συνολική απόσταση, που είναι η κλάση $L1$ στο παράδειγμα της παρακάτω εικόνας.



Εικόνα 13: Απεικόνιση NBNN

Ο NBNN αλγόριθμος μπορεί να συνοψιστεί ως εξής:

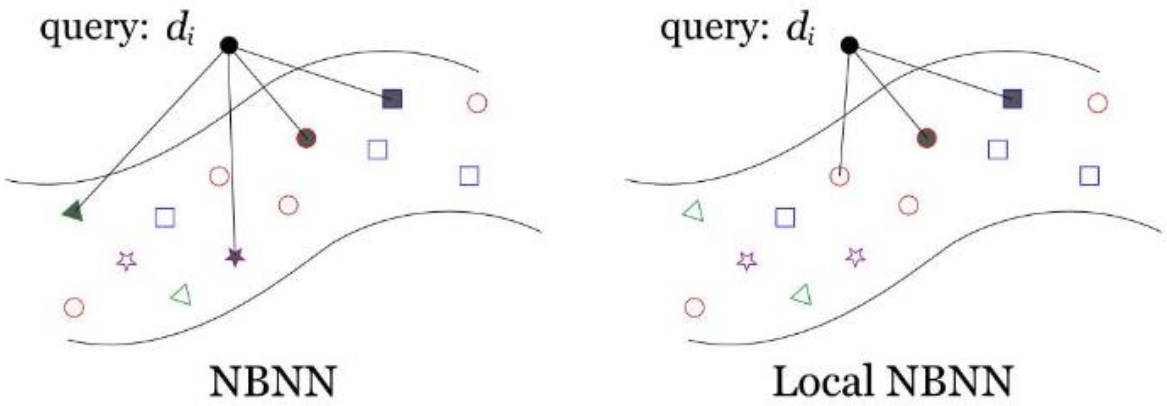
- Υπολογισμός των n περιγραφέων $d_1, \ldots, d_n$ της άγνωστης εικόνας Q.
- $\forall d_i \, \forall c \in C$ υπολογίζεται ο κοντινότερος γείτονας NN του d_i στην κλάση c: $NN_C(d_i)$, όπου NN_C είναι ο κοντινότερος γείτονας στην κλάση c και C είναι το σύνολο των κλάσεων.
- $\hat{C} = argmin_C \left[\sum_{i=1}^{n} \left\| d_i - NN_C(d_i) \right\|^2\right]$, όπου $\hat{C}$ είναι η κλάση με την ελάχιστη συνολική απόσταση.

Για τον εντοπισμό των πλησιέστερων γειτόνων δημιουργείται ένα ευρετήριο αναζήτησης για όλες τις κλάσεις χρησιμοποιώντας την εφαρμογή των k-dimensional (k-d) δέντρων που παρέχονται από τη βιβλιοθήκη Fat Library for Approximate Nearest Neighbours (FLANN) [17], ώστε η αναζήτηση να είναι αποτελεσματική.

5.2.2 Ταξινομητής Local Naïve Bayes Nearest Neighbour

Ο Local Naive Bayes Nearest Neighbour [4] είναι ο βελτιωμένος NBNN αλγόριθμος, ο οποίος όχι μόνο παρέχει υψηλότερη ακρίβεια ταξινόμησης αλλά είναι και ταχύτερος για εικόνες φυσικών σκηνών. Ως εκ τούτου μπορεί να εφαρμοστεί και να δώσει βελτιωμένα αποτελέσματα σε πολλούς διαφορετικούς κλάδους.

Η βασική του ιδέα είναι η εξάλειψη της ανάγκης για αναζήτηση του πιο κοντινού γείτονα σε όλες τις κλάσεις. Αντίθετα, λαμβάνονται υπόψη μόνο οι κλάσεις που βρίσκονται εντός μιας συγκεκριμένης γειτονίας του άγνωστου περιγραφέα. Στην εικόνα 14, παρουσιάζεται η βασική του διαφορά με τον NBNN, ο οποίος αναζητεί τον κοντινότερο γείτονα του άγνωστου περιγραφέα σε όλες τις κλάσεις. Ο Local NBNN συγκρίνει τον περιγραφέα προς ταξινόμηση μόνο με τους περιγραφείς που βρίσκονται εντός μια συγκεκριμένης γειτονίας, ανεξάρτητα από την κλάση στην οποία ανήκουν.



Εικόνα 14: Βασική διαφορά NBNN και Local NBNN

Η προτεινόμενη μέθοδος βασίζεται στον Local NBNN αλγόριθμο [4] και στην εργασία [6] προτείνονται οι παρακάτω τύποι:

$$Dist_{local}^{c} = \sum_{i=1}^{n} [(\|d_i - \varphi(NN_c(d_i))\|^2 - \|d_i - N_{k+1}(d_i)\|^2)] \qquad (5.2)$$

$$\hat{C} = argmin_C(Dist_{local}^{c}) \qquad (5.3)$$

$$\varphi(NN_c(d_i)) = \begin{cases} NN_c(d_i), & \alpha\nu\ NN_c(d_i) \le NN_{k+1}(d_i) \\ N_{k+1}(d_i), & \alpha\nu\ NN_c(d_i) > NN_{k+1}(d_i) \end{cases} \qquad (5.4)$$

όπου $NN_{k+1}(d_i)$ είναι ο γείτονας $(k+1)$ του περιγραφέα d_i.

Για τον εντοπισμό των πλησιέστερων γειτόνων, χρησιμοποιείται πάλι ένα ευρετήριο αναζήτησης για όλες τις κλάσεις με k-d δέντρα ώστε να είναι αποτελεσματική η αναζήτηση.

Στη συνέχεια, υπολογίζονται οι 10 πλησιέστεροι γείτονες για κάθε περιγραφέα της χειρόγραφης εικόνας προς ταξινόμηση (ο αριθμός 10 έχει αποδειχθεί πειραματικά [4] όπου επιβεβαιώνεται ότι είναι ο βέλτιστος από όλα τα πειράματα που έγιναν σε χειρόγραφες εικόνες). Σύμφωνα με την εργασία [4], χρησιμοποιείται η απόσταση του k+1 κοντινότερου γείτονα ώστε να υπολογιστούν οι “background” αποστάσεις από τις κλάσεις που δεν εντοπίστηκαν στους k πλησιέστερους γείτονες.

Ο Local NBNN μπορεί να συνοψιστεί ως εξής:

1. Για όλους τους περιγραφείς $d_i \in Q$, όπου Q είναι η εικόνα προς ταξινόμηση
2. $\{p_1, p_2, \ldots, p_{k+1}\} \leftarrow NN(d_i, k+1)$
3. $dist_B \leftarrow \|d_i - p_{k+1}\|^2$
4. Για όλες τις κατηγορίες C που βρέθηκαν στους k πλησιέστερους γείτονες
5. $dist_C = min_{\{p_j | Class(p_j)=C\}} \|d_i - p_j\|^2$
6. $totals[C] \leftarrow totals[C] + dist_C - dist_B$
7. Τέλος επανάληψης
8. Τέλος επανάληψης
9. επέστρεψε $argmin_C totals[C]$

Η βασική διαφορά του Local NBNN από τον NBNN είναι η αφαίρεση της απόστασης του k+1 πλησιέστερου γείτονα κάθε φορά που προστίθενται οι αποστάσεις των k πλησιέστερων γειτόνων στις συνολικές αποστάσεις των κλάσεων που ανήκουν. Η διαδικασία επαναλαμβάνεται για όλους τους περιγραφείς της εικόνας λαμβάνοντας υπόψη τους αντίστοιχους κοντινότερους γείτονες τους. Τελικά η κάθε εικόνα ταξινομείται στη γλώσσα με το ελάχιστο άθροισμα αποστάσεων.

5.2.3 Κατώφλι προσανατολισμού

Τα μοτίβα των γλωσσών αποδίδουν πολλά keypoints με παρόμοια χαρακτηριστικά που όμως έχουν διαφορετικούς προσανατολισμούς. Δεδομένου ότι ο προσανατολισμός των χαρακτηριστικών μπορεί να αποτελέσει χαρακτηριστικό γνώρισμα των γλωσσών, παρέχει μια ξεχωριστή ιδιότητα στα χαρακτηριστικά. Για τον λόγο αυτόν πρέπει να αποφευχθεί η αντιστοίχιση των όμοιων περιγραφέων που έχουν διαφορετικό προσανατολισμό. Προτείνεται λοιπόν η παρακάτω συνθήκη [6]:

$$\left|Ort_{kpt1} - Ort_{kpt2}\right| \le T_r \qquad (5.5)$$

όπου Ort_{kpt1} και Ort_{kpt2} είναι οι προσανατολισμοί των δύο keypoints (σε μοίρες) που ελέγχονται τα χαρακτηριστικά τους ώστε να ταιριάξουν και T_r είναι το κατώφλι διαφοράς των προσανατολισμών.

Με άλλα λόγια, τα χαρακτηριστικά με διαφορά προσανατολισμών μεγαλύτερη από το κατώφλι δε θεωρούνται κατάλληλα ώστε να ταιριάξουν. Σύμφωνα με την εργασία [6], η πιο κατάλληλη τιμή του κατωφλίου είναι η τιμή 10. Ο έλεγχος διαφοράς προσανατολισμού απευθύνεται στα χαρακτηριστικά που ανιχνεύονται μέσω του SIFT αλγορίθμου κι επομένως εφαρμόζεται και στον NBNN και στον Local NBNN.

5.2.4 Κανονικοποίηση αποστάσεων των κλάσεων

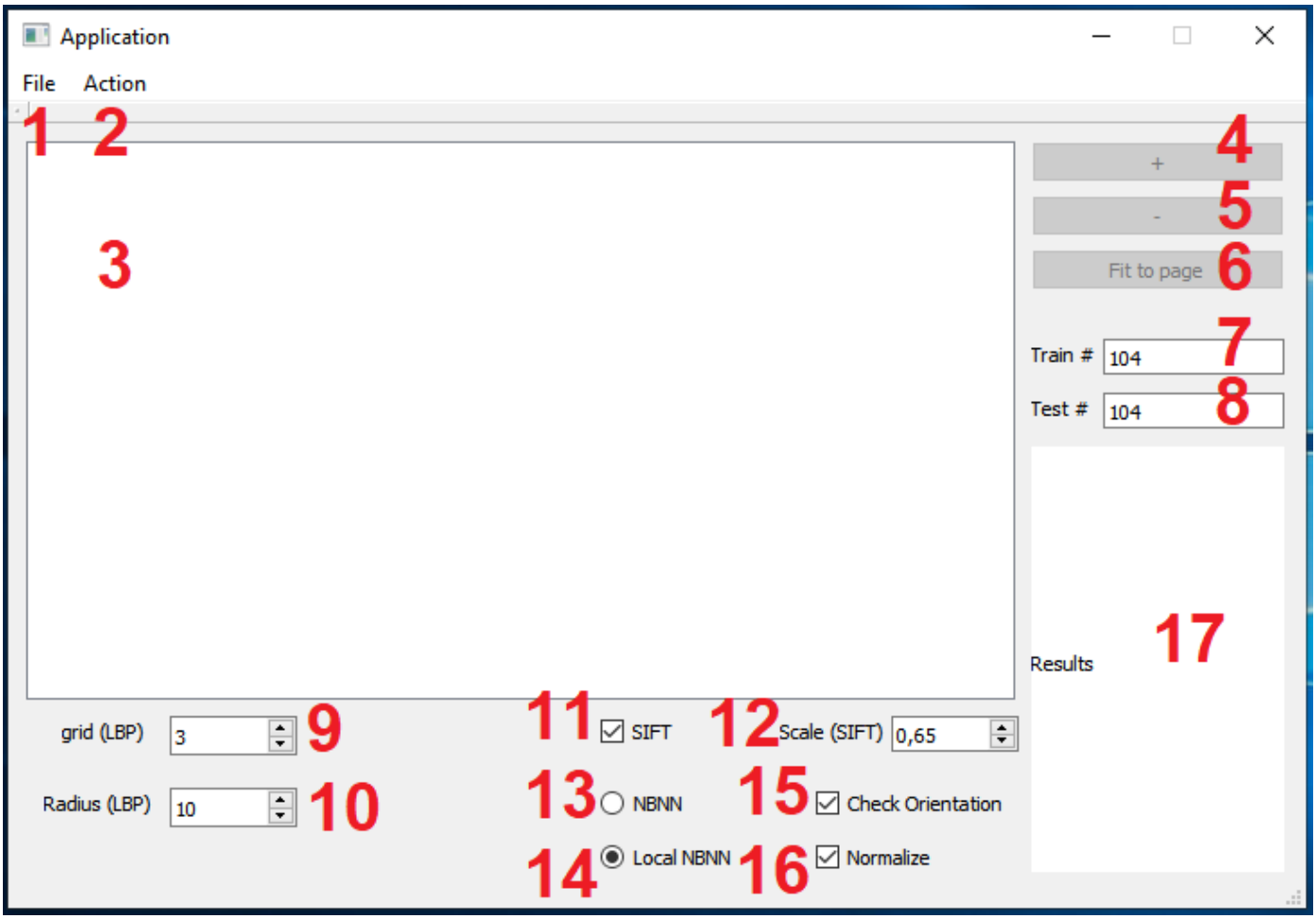
Πολλές φορές τα δείγματα των διαφορετικών γλωσσών μπορεί να χαρακτηριστούν ανισόρροπα όταν τουλάχιστον μία από τις κλάσεις αποτελείται από μικρό αριθμό δειγμάτων. Αυτό μπορεί να επηρεάσει το αποτέλεσμα της ταξινόμησης και να μειώσει κατά πολύ το ποσοστό επιτυχίας της. Για τον λόγο αυτόν, πραγματοποιείται κανονικοποίηση των τελικών αποστάσεων $Dist_{local}^{C}$ από κάθε κλάση με τον αριθμό των keypoints της κάθε κλάσης αντίστοιχα [6] (τύπος 5.6).

$$\hat{C} = argmin_C\left(\frac{Dist_{local}^{C}}{K_C}\right) \qquad (5.6)$$

Ο αριθμός των keypoints K_C κάθε κλάσης υπολογίζεται ως το πλήθος των χαρακτηριστικών που πέρασαν επιτυχώς τον έλεγχο προσανατολισμού.

6. ΠΑΡΟΥΣΙΑΣΗ ΕΦΑΡΜΟΓΗΣ ΤΑΞΙΝΟΜΗΣΗΣ ΕΙΚΟΝΩΝ

Για την υλοποίηση της ταξινόμησης των εικόνων αναπτύχθηκε μια εφαρμογή βασισμένη στη γλώσσα προγραμματισμού C++ [18]. Το ανεξάρτητο πλατφόρμας (cross-platform) περιβάλλον ανάπτυξης λογισμικού (software development environment) που χρησιμοποιήθηκε είναι το «Qt» [19]. Παράλληλα έγινε χρήση της βιβλιοθήκης «OpenCV» [20] για την εφαρμογή του αλγορίθμου SIFT. Στην εικόνα 15, διακρίνεται το Γραφικό Περιβάλλον Αλληλεπίδρασης Χρήστη-Υπολογιστή (Graphical User Interface - GUI) και στη συνέχεια σχολιάζονται οι επιλογές που δίνονται από αυτό.

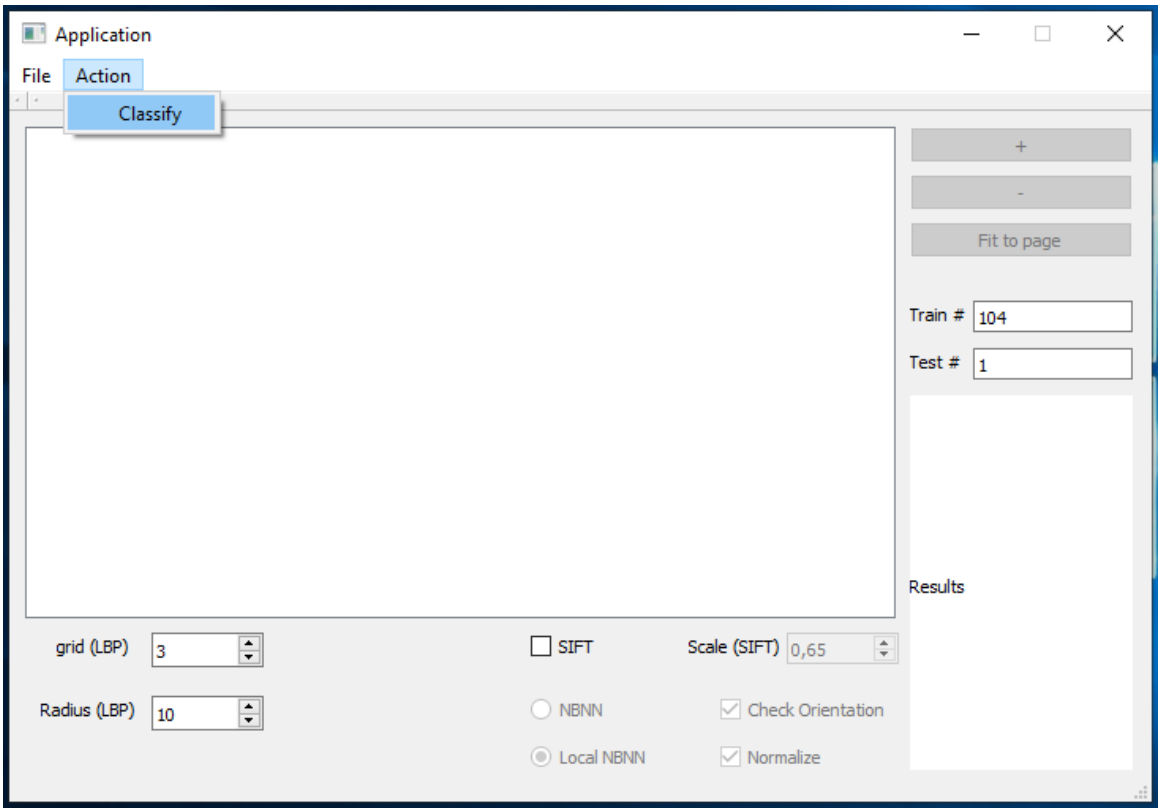


Εικόνα 15: Γραφικό Περιβάλλον Αλληλεπίδρασης Χρήστη-Υπολογιστή

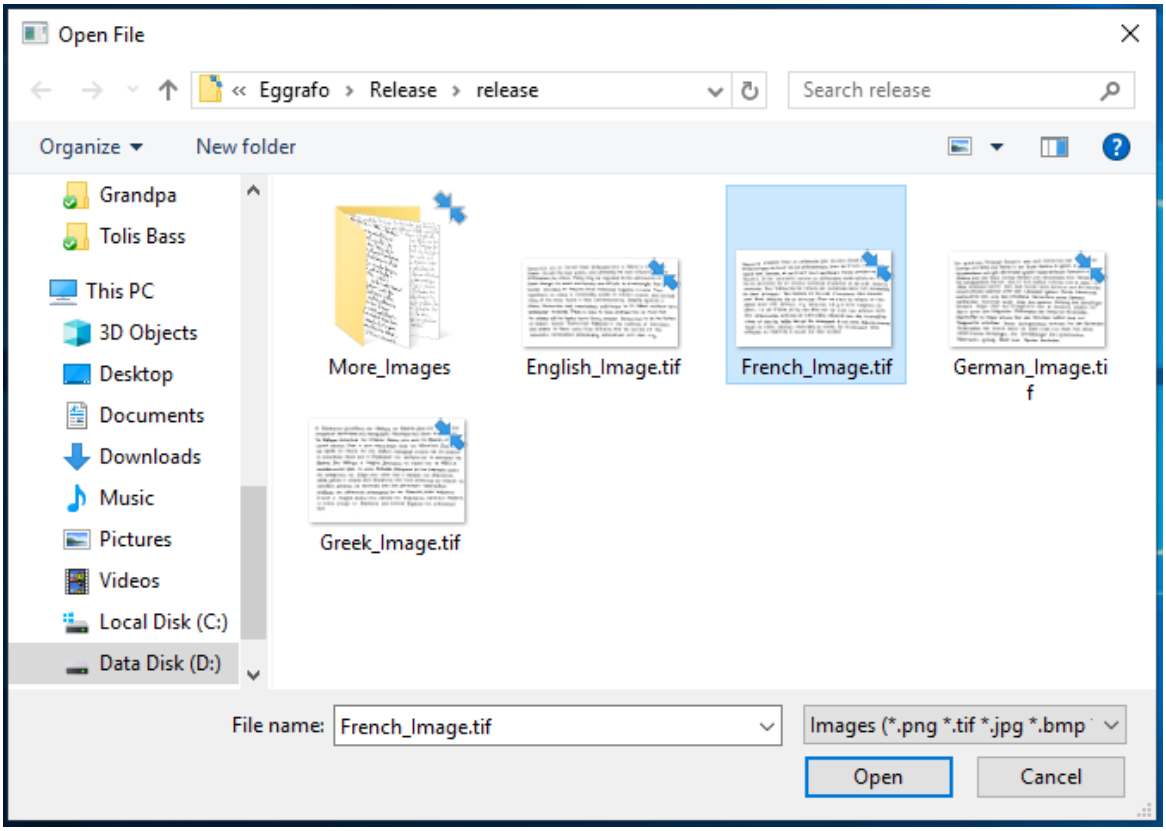
Οι εικόνες που θα χρησιμοποιηθούν σαν test και train σύνολα και απαιτούνται για την ομαλή εκτέλεση, πρέπει να βρίσκονται στους υπό-φακέλους του φακέλου «Images» ο οποίος βρίσκεται εντός του προγράμματος και να είναι χωρισμένες αναλόγως τη γλώσσα και την κατηγορία συνόλου αναφοράς ή συνόλου προς ταξινόμηση. Οι επιλογές που δίνονται από το Γραφικό Περιβάλλον είναι οι ακόλουθες:

1. **Κλείσιμο/επανεκκίνηση εφαρμογής:** Ο χρήστης μπορεί να επιλέξει «File-Restart» για την επανεκκίνηση της εφαρμογής ή «File-Exit» για να την κλείσει.
2. **Εκτέλεση εφαρμογής:** Πραγματοποιείται εκτέλεση της εφαρμογής επιλέγοντας «Action-Classify».
3. **Παρουσίαση εικόνας:** Στην περίπτωση εκτέλεσης της εφαρμογής για μία test εικόνα, η εικόνα εμφανίζεται στο συγκεκριμένο παράθυρο.
4. **Μεγέθυνση εικόνας:** Μεγέθυνση της εικόνας του παραθύρου
5. **Προσαρμογή εικόνας:** Προσαρμογή εικόνας στις διαστάσεις του παραθύρου
6. **Σμίκρυνση εικόνας:** Σμίκρυνση της εικόνας του παραθύρου
7. **Αριθμός train εικόνων:** Επιλογή του αριθμού των εικόνων που θα αποτελέσουν το σύνολο των εικόνων αναφοράς.
8. **Αριθμός test εικόνων:** Επιλογή του αριθμού των εικόνων που θα αποτελέσουν το σύνολο των εικόνων προς ταξινόμηση. Στην περίπτωση που ο χρήστης επιθυμεί να εκτελέσει το πρόγραμμα για μία άγνωστη εικόνα, αφού βάλει τις εισόδους στο Γραφικό Περιβάλλον και εκτελέσει το πρόγραμμα, του εμφανίζεται ένα νέο παράθυρο ώστε να επιλέξει την αντίστοιχη εικόνα.
9. **Αριθμός grid:** Καθορίζεται το σε πόσα κομμάτια θα μοιραστεί το ύψος και το πλάτος των εικόνων (grid×grid) που θα αποτελούν τα νέα μεγέθη των παραθύρων των εικόνων ώστε να εφαρμοστεί η μέθοδος ταξινόμησης με τη χρήση των LBP χαρακτηριστικών, όπως περιγράφεται στο υποκεφάλαιο 3.3.
10. **Αριθμός radius:** Επιλογή της μέγιστης ακτίνας (radius) για τον υπολογισμό των LBP με τιμές ακτίνας από 1 μέχρι την τιμή radius και ένωση των ιστογραμμάτων των LBP χαρακτηριστικών όλων των ακτίνων σε ένα ιστόγραμμα, όπως περιγράφεται στο υποκεφάλαιο 3.3.
11. **Επιλογή SIFT:** Επιλογή υπολογισμού των SIFT χαρακτηριστικών και ταξινόμησής τους
12. **Επιλογή κλιμάκωσης εικόνας:** Επιλογή της σμίκρυνσης των εικόνων εισόδου μόνο για τη μέθοδο με τα SIFT χαρακτηριστικά. Επιλέγοντας μία τιμή από το σύνολο, πολλαπλασιάζεται η κάθε πλευρά των εικόνων με το αντίστοιχο νούμερο με αποτέλεσμα την κλιμάκωση τους. Στην περίπτωση που επιλεγεί η τιμή 1, δεν υπάρχει καμία αλλαγή στο μέγεθός τους.
13. **Επιλογή ταξινομητή NBNN:** Επιλέγεται ο ταξινομητής NBNN για την ταξινόμηση των εικόνων με τη χρήση των SIFT χαρακτηριστικών.
14. **Επιλογή ταξινομητή Local NBNN:** Επιλέγεται ο ταξινομητής Local NBNN για την ταξινόμηση των εικόνων με τη χρήση των SIFT χαρακτηριστικών.
15. **Επιλογή ελέγχου προσανατολισμού:** Επιλογή ελέγχου της διαφοράς του προσανατολισμού των keypoints, όπως περιγράφεται στο υποκεφάλαιο 5.2.
16. **Επιλογή κανονικοποίησης:** Επιλογή κανονικοποίησης των τελικών αποστάσεων από κάθε κλάση με το πλήθος των keypoints κάθε κλάσης, όπως περιγράφεται στο υποκεφάλαιο 5.2.
17. **Πίνακας αποτελεσμάτων:** Παρουσιάζονται στο χρήστη τα αποτελέσματα της ταξινόμησης των εικόνων εγγράφων

Ακολουθούν ενδεικτικά σενάρια και εικόνες των παραπάνω λειτουργιών της εφαρμογής.



Εικόνα 16: Στιγμιότυπο της εφαρμογής πριν την εκκίνηση της

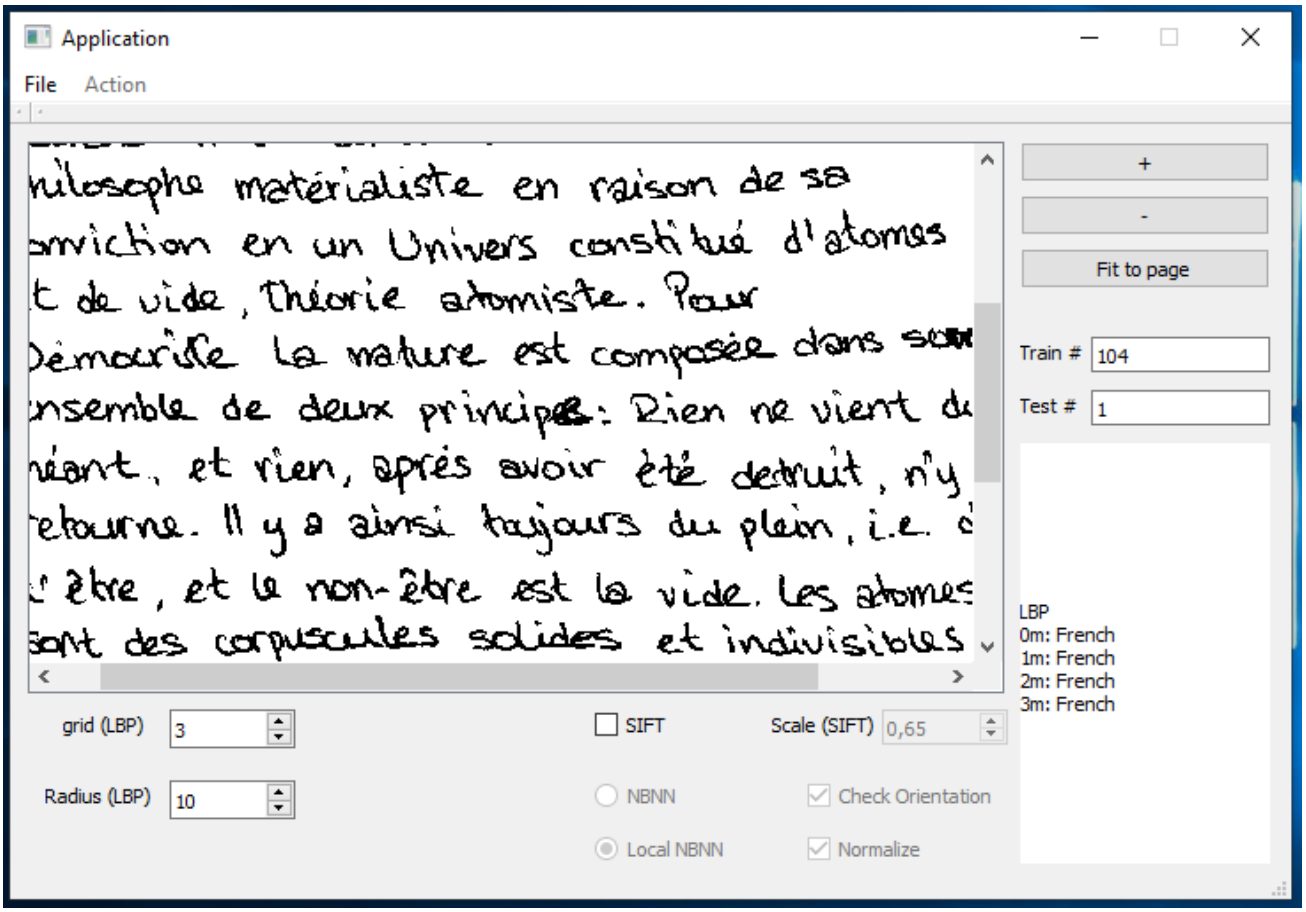


Εικόνα 17: Στιγμιότυπο του παραθύρου που εμφανίζεται στον χρήστη για την επιλογή μίας εικόνας προς ταξινόμηση

Στην εικόνα 16, παρουσιάζεται η εφαρμογή με τις τιμές των παραμέτρων που εισάγει ο χρήστης με σκοπό την ταξινόμηση μίας εικόνας με τη χρήση μόνο των LBP χαρακτηριστικών. Έχουν επιλεχθεί 104 εικόνες για το σύνολο αναφοράς με το οποίο θα συγκριθεί η εικόνα και οι τιμές των παραμέτρων «grid» και «Radius» είναι 3 και 10 αντίστοιχα. Ο αριθμός των εικόνων που εισάγεται στην εφαρμογή, διαιρείται με το 4 και το αποτέλεσμα είναι το πλήθος των εικόνων για κάθε μία από τις 4 γλώσσες των Αγγλικών, Γαλλικών, Γερμανικών και Ελληνικών. Έτσι, για παράδειγμα, για την γλώσσα των Αγγλικών οι εικόνες του συνόλου αναφοράς είναι 26 (104/4). Το ίδιο συμβαίνει και σε όλες τις υπόλοιπες γλώσσες.

Επιλέγοντας «Action-Classify», εμφανίζεται ένα παράθυρο στον χρήστη όπου καλείται να επιλέξει την εικόνα προς ταξινόμηση, όπως παρουσιάζεται στην εικόνα 17. Ο χρήστης επιλέγει μια εικόνα εγγράφου στη γαλλική γλώσσα και πατώντας το κουμπί «Open», εκκινεί την εφαρμογή.

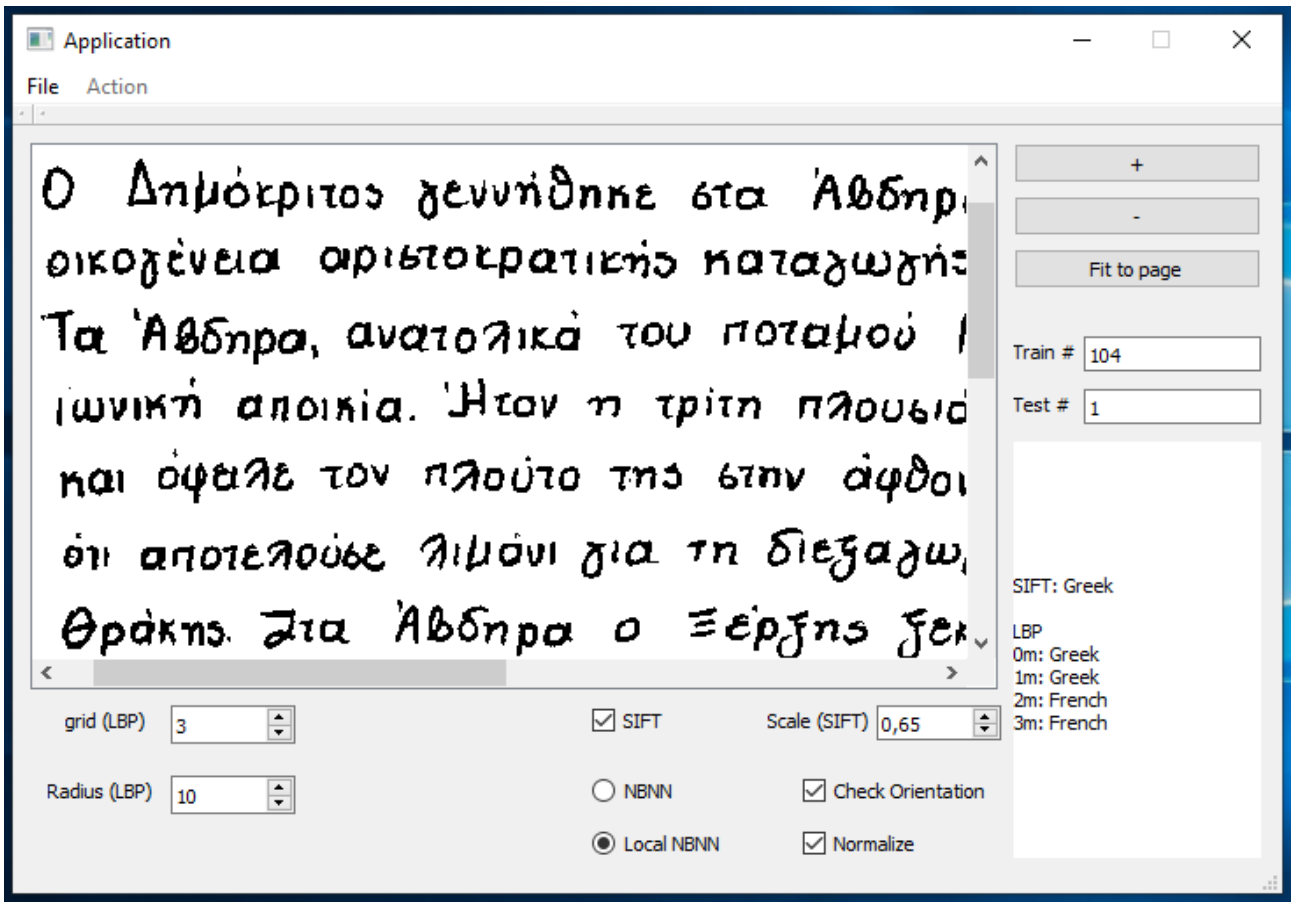
Στην εικόνα 18, έχει ολοκληρωθεί η διαδικασία της ταξινόμησης και παρουσιάζονται τα αποτελέσματα του παραπάνω σεναρίου. Στον πίνακα της εφαρμογής με τα αποτελέσματα φαίνεται ότι η εικόνα ταξινομήθηκε σωστά με όλους τους διαφορετικούς τρόπους επιλογής των πλησιέστερων γειτόνων, όπως περιγράφονται στο υποκεφάλαιο 5.1.1. Στο μεγάλο παράθυρο της εφαρμογής παρουσιάζεται η εικόνα αυτή.



Εικόνα 18: Στιγμιότυπο της παρουσίασης των αποτελεσμάτων ταξινόμησης για μία εικόνα προς ταξινόμηση με χρήση των LBP χαρακτηριστικών

Στην εικόνα 19, παρουσιάζεται ένα αντίστοιχο σενάριο με το προηγούμενο όπου ο χρήστης όμως έχει επιλέξει να πραγματοποιηθεί η ταξινόμηση μιας εικόνας με τη χρήση των LBP χαρακτηριστικών και των SIFT χαρακτηριστικών. Έχουν επιλεχθεί 104 εικόνες για το σύνολο αναφοράς με το οποίο θα συγκριθεί η εικόνα και οι τιμές των παραμέτρων «grid» και «Radius» είναι 3 και 10 αντίστοιχα. Οι τιμές των παραμέτρων που αφορούν τα SIFT χαρακτηριστικά είναι 0.65 για την κλιμάκωση των εικόνων και έχει επιλεχτεί ο ταξινομητής Local NBNN με ενεργοποιημένες τις επιλογές για τον έλεγχο των προσανατολισμών των keypoints και την κανονικοποίηση των τελικών αποστάσεων από τις κλάσεις, όπως περιγράφονται στο υποκεφάλαιο 5.2. Ο χρήστης στο συγκεκριμένο σενάριο επιλέγει μία εικόνα εγγράφου στην ελληνική γλώσσα.

Αφού ολοκληρωθεί η διαδικασία της ταξινόμησης, παρουσιάζονται τα αποτελέσματα του παραπάνω σεναρίου. Στον πίνακα της εφαρμογής με τα αποτελέσματα φαίνεται ότι η εικόνα ταξινομήθηκε σωστά με τη χρήση των LBP χαρακτηριστικών μόνο με τους δύο πρώτους τρόπους επιλογής των πλησιέστερων γειτόνων, όπως περιγράφονται στο υποκεφάλαιο 5.1.1. Με την χρήση των χαρακτηριστικών SIFT, η αναγνώριση γλώσσας ήταν επιτυχής.



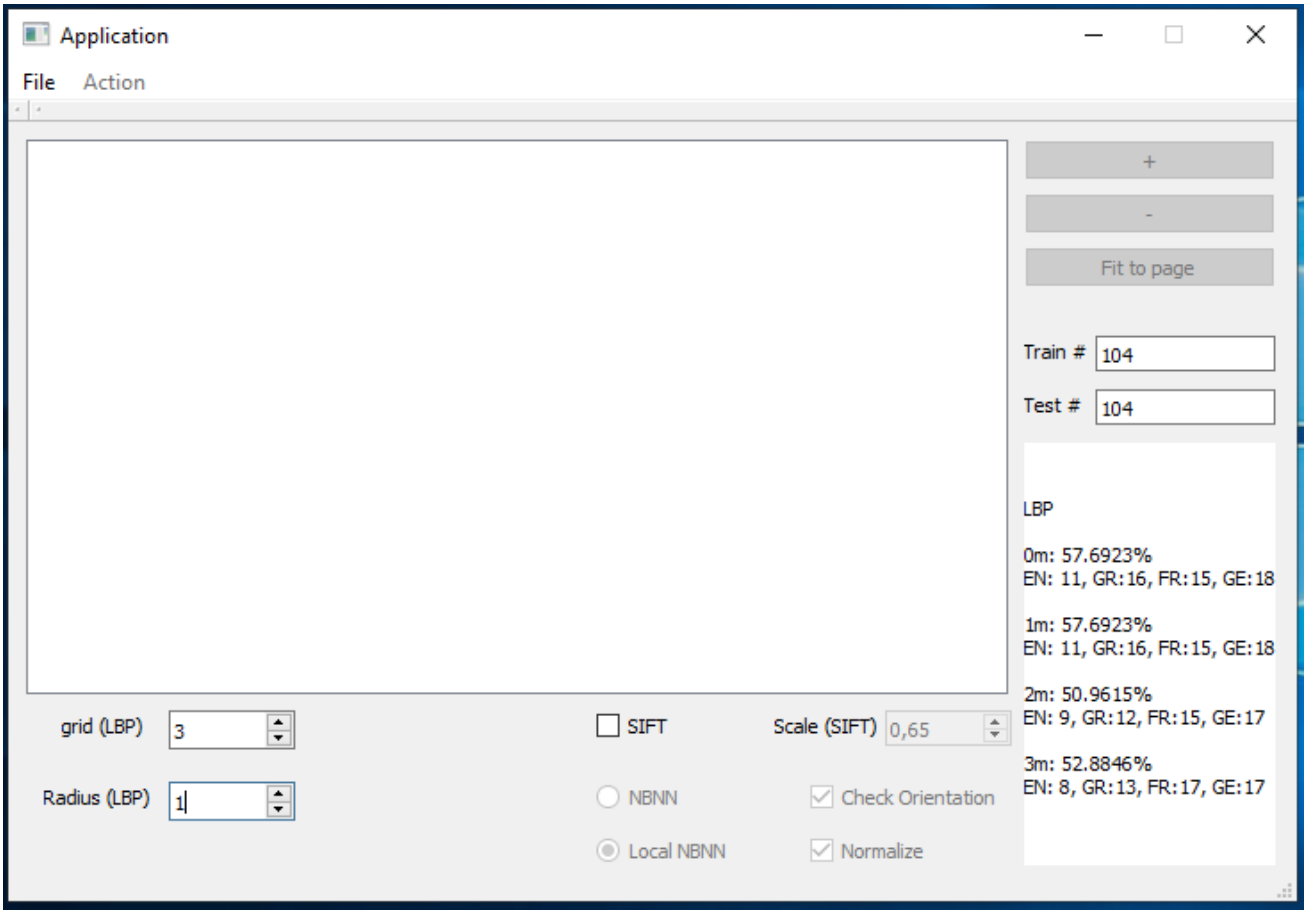
Εικόνα 19: Στιγμιότυπο της παρουσίασης των αποτελεσμάτων ταξινόμησης για μία εικόνα προς ταξινόμηση με χρήση των LBP και SIFT χαρακτηριστικών

Στην εικόνα 20, παρουσιάζεται ένα σενάριο όπου ο χρήστης έχει επιλέξει να πραγματοποιηθεί ταξινόμηση πολλών εικόνων μόνο με τη χρήση των LBP χαρακτηριστικών. Έχουν επιλεχθεί 104 εικόνες για το σύνολο αναφοράς και 104 εικόνες για το σύνολο εικόνων

προς ταξινόμηση. Ο αριθμός των εικόνων που εισάγεται στην εφαρμογή, διαιρείται με το 4 και το αποτέλεσμα είναι το πλήθος των εικόνων για κάθε μία από τις 4 γλώσσες των Αγγλικών, Γαλλικών, Γερμανικών και Ελληνικών. Αυτό συμβαίνει και για το σύνολο αναφοράς και για το σύνολο προς ταξινόμηση. Έτσι, για παράδειγμα, για τη γλώσσα των Αγγλικών οι εικόνες του συνόλου αναφοράς είναι 26 (104/4) και του συνόλου προς ταξινόμηση επίσης 26 (104/4). Το ίδιο συμβαίνει και σε όλες τις υπόλοιπες γλώσσες. Οι τιμές των παραμέτρων «grid» και «Radius» είναι 3 και 1 αντίστοιχα.

Επιλέγοντας «Action-Classify», ο χρήστης εκκινεί την εφαρμογή.

Αφού ολοκληρωθεί η διαδικασία της ταξινόμησης, παρουσιάζονται τα αποτελέσματα του παραπάνω σεναρίου. Στον πίνακα της εφαρμογής φαίνονται τα ποσοστά επιτυχούς ταξινόμησης για κάθε έναν από τους τρόπους επιλογής των πλησιέστερων γειτόνων, όπως περιγράφονται στο υποκεφάλαιο 5.1.1. Ο πρώτος τρόπος έχει ποσοστό επιτυχούς αναγνώρισης γλώσσας ίσο με 57.69%, ο δεύτερος ίσο με 57.69%, ο τρίτος ίσο με 50.96% και ο τέταρτος ίσο με 52.88%. Κάτω από τα αντίστοιχα ποσοστά, παρουσιάζεται το ακριβές νούμερο των εικόνων για κάθε γλώσσα που ταξινομήθηκαν σωστά.

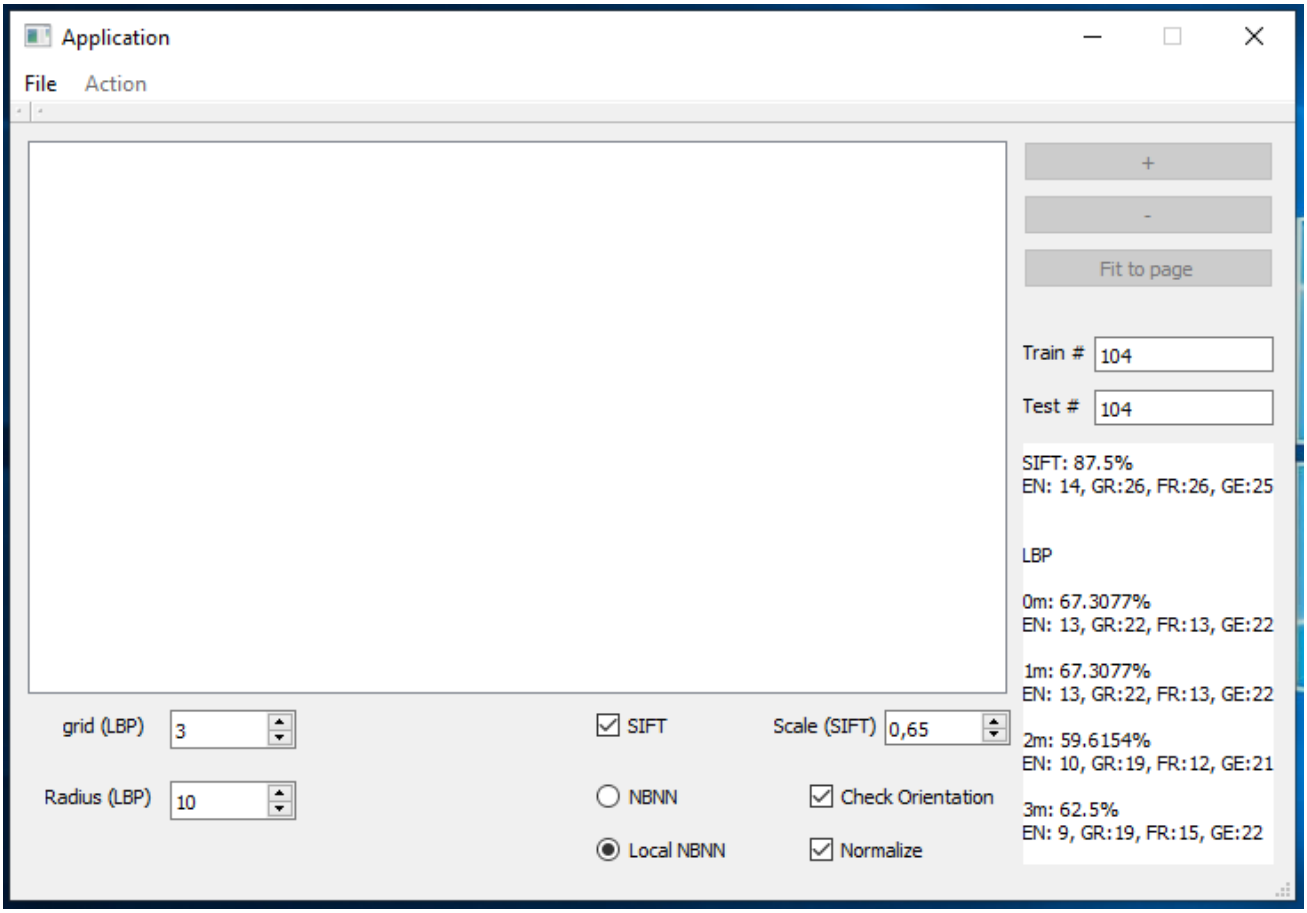


Εικόνα 20: Στιγμιότυπο της παρουσίασης των αποτελεσμάτων ταξινόμησης για πολλές εικόνες προς ταξινόμηση με χρήση των LBP χαρακτηριστικών

Στην εικόνα 21, παρουσιάζεται ένα αντίστοιχο σενάριο με το προηγούμενο όπου ο χρήστης όμως έχει επιλέξει να πραγματοποιηθεί η ταξινόμηση πολλών εικόνων με τη χρήση των LBP

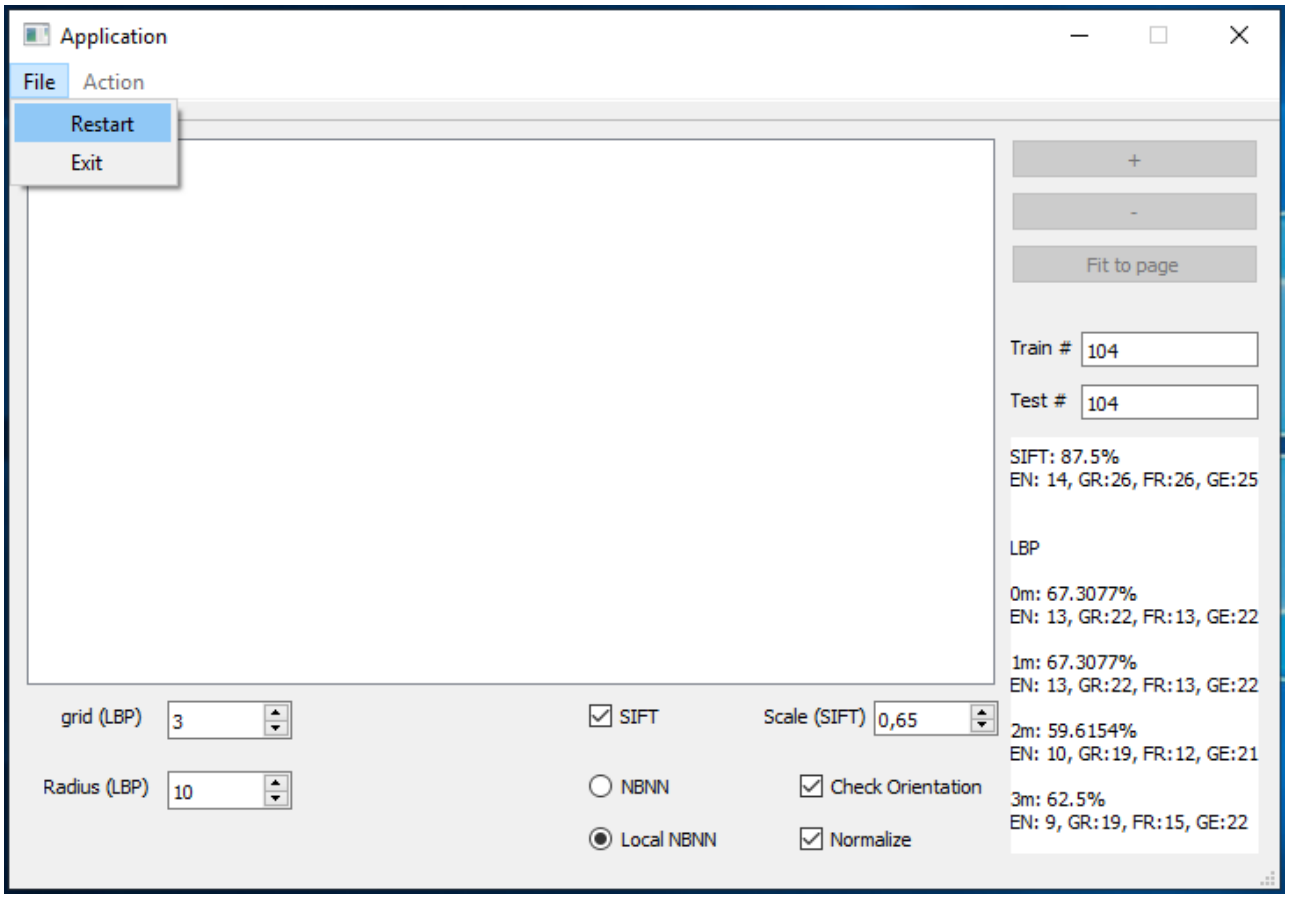
χαρακτηριστικών και των SIFT χαρακτηριστικών. Έχουν επιλεχθεί 104 εικόνες για το σύνολο αναφοράς και 104 εικόνες για το σύνολο εικόνων προς ταξινόμηση. Οι τιμές των παραμέτρων «grid» και «Radius» είναι 3 και 10 αντίστοιχα. Οι τιμές των παραμέτρων που αφορούν τα SIFT χαρακτηριστικά είναι 0.65 για την κλιμάκωση των εικόνων και έχει επιλεχτεί ο ταξινομητής Local NBNN με ενεργοποιημένες τις επιλογές για τον έλεγχο των προσανατολισμών των keypoints και την κανονικοποίηση των τελικών αποστάσεων από τις κλάσεις, όπως περιγράφονται στο υποκεφάλαιο 5.2.

Αφού ολοκληρωθεί η διαδικασία της ταξινόμησης, παρουσιάζονται τα αποτελέσματα του παραπάνω σεναρίου. Στον πίνακα της εφαρμογής φαίνονται τα ποσοστά επιτυχούς ταξινόμησης με τα LBP χαρακτηριστικά για κάθε έναν από τους τρόπους επιλογής των πλησιέστερων γειτόνων, όπως περιγράφονται στο υποκεφάλαιο 5.1.1. Ο πρώτος τρόπος έχει ποσοστό επιτυχούς αναγνώρισης γλώσσας ίσο με 67.30%, ο δεύτερος ίσο με 67.30%, ο τρίτος ίσο με 59.61% και ο τέταρτος ίσο με 62.5%. Το ποσοστό επιτυχούς αναγνώρισης της γλώσσας με τα SIFT χαρακτηριστικά ισούται με 87.5%..



Εικόνα 21: Στιγμιότυπο της παρουσίασης των αποτελεσμάτων ταξινόμησης για πολλές εικόνες προς ταξινόμηση με χρήση των LBP και SIFT χαρακτηριστικών

Στην εικόνα 22, παρουσιάζονται οι επιλογές που δίνονται στον χρήστη σε περίπτωση που θέλει να κλείσει την εφαρμογή ή να κάνει επανεκκίνησή της, αφού ολοκληρώσει τα πειράματά του.



Εικόνα 22: Στιγμιότυπο της εφαρμογής πριν την έξοδο ή επανεκκίνησή της

7. ΠΕΙΡΑΜΑΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ

Για τον έλεγχο των μεθόδων ταξινόμησης, τα πειράματα πραγματοποιήθηκαν στο σύνολο εικόνων benchmarking dataset ICDAR 2011 Writer Identification Contest [21] [22], ένα σύνολο χειρόγραφων εικόνων που αποτελείται από 208 εικόνες. Οι εικόνες αυτές προέρχονται από 26 διαφορετικούς γραφείς εκ των οποίων κάθε ένας έχει γράψει 2 εικόνες για κάθε μία από τις γλώσσες Αγγλικά, Γαλλικά, Γερμανικά και Ελληνικά. Ο τρόπος που χωρίστηκε το σύνολο των εικόνων στο σύνολο αναφοράς και στο σύνολο προς ταξινόμηση είναι 1 εικόνα του κάθε γραφέα για την κάθε γλώσσα στο σύνολο αναφοράς και μία στο άγνωστο σύνολο. Έτσι, και τα δύο σύνολα αποτελούνται από 104 εικόνες, 26 εικόνες για κάθε γλώσσα αντίστοιχα.

Όλες οι εικόνες είναι σε δυαδική μορφή και χρησιμοποιούνται χωρίς περαιτέρω μετατροπές σαν είσοδοι στις μεθόδους ταξινόμησης που αφορούν τα LBP χαρακτηριστικά.

7.1 Πειραματικά αποτελέσματα LBP χαρακτηριστικών

Για τον έλεγχο της απόδοσης του ταξινομητή των LPB χαρακτηριστικών πραγματοποιήθηκαν πειράματα ως προς τον αριθμό διαχωρισμού των αρχικών εικόνων σε παράθυρα και ως προς την ακτίνα απόστασης των γειτονικών εικονοστοιχείων. Για την απλή μορφή των LBP η ακτίνα είναι ίση με 1 σε όλα τα πειράματα, ενώ για την επέκταση της αρχικής μορφής η ακτίνα λαμβάνει τιμές από την τιμή 1 μέχρι και την μέγιστη τιμή της, όπως παρουσιάζεται στο υποκεφάλαιο 3.3.

Στον πίνακα 1, παρουσιάζονται τα αποτελέσματα με ακτίνα 1 για πλήθος παραθύρων των εικόνων με τιμές από 2x2 μέχρι 10x10, σε φθίνουσα σειρά ως προς το ποσοστό επιτυχίας της ταξινόμησης. Η πρώτη στήλη «Classifier» περιέχει τους τέσσερις διαφορετικούς τρόπους επιλογής των πλησιέστερων γειτόνων, που επεξηγούνται στην εικόνα 12. Το πλήθος των παραθύρων που διαχωρίζονται οι αρχικές εικόνες, όπως περιγράφεται στο υποκεφάλαιο 3.3, αντιστοιχεί στη στήλη «Windows».

Παρατηρείται ότι για μεγάλο πλήθος παραθύρων, από 7x7 μέχρι 10x10, το ποσοστό ορθής ταξινόμησης των εικόνων είναι κάτω από 50%. Ο αριθμός των παραθύρων των εικόνων με τον οποίον η ταξινόμηση έχει τα καλύτερα αποτελέσματα είναι 2x2 και 3x3, δηλαδή στις περιπτώσεις που τα παράθυρα έχουν το μεγαλύτερο επιτρεπτό μέγεθος μετά τη διαίρεση των αρχικών εικόνων.

Όσον αφορά την ταξινόμηση των διαφορετικών γλωσσών, τα καλύτερα αποτελέσματα εμφανίζονται στο διαχωρισμό των Γερμανικών και των Ελληνικών. Αντίθετα, η ταξινόμηση των Αγγλικών και Γαλλικών έχει μικρότερο ποσοστό ορθής ταξινόμησης. Αυτό παρατηρείται για όλους τους συνδυασμούς του πλήθους των παραθύρων που χωρίζονται οι εικόνες αλλά και για όλους τους τρόπους επιλογής των πλησιέστερων γειτόνων.

Σχετικά με τους διαφορετικούς τρόπους επιλογής των πλησιέστερων γειτόνων, παρατηρείται ότι τα καλύτερα αποτελέσματα προέρχονται από την επιλογή του πρώτου πλησιέστερου γείτονα και την επιλογή όλων των πλησιέστερων γειτόνων που έχουν ίση ελάχιστη απόσταση με τον πρώτο πλησιέστερο, εφόσον υπάρχουν.

Πίνακας 1: Αποτελέσματα ταξινόμησης LBP

Αποτελέσματα ταξινόμησης LBP με πλήθος παραθύρων από 2x2 μέχρι 10x10 (σε φθίνουσα σειρά ως προς το «Rate»)						
Classifier	Windows	Rate	German	Greek	English	French
0m	3x3	57.69%	18	16	11	15
1m	3x3	57.69%	18	16	11	15
0m	2x2	54.81%	13	15	14	15
1m	2x2	54.81%	13	15	14	15
3m	2x2	53.85%	14	16	8	18
3m	3x3	52.88%	17	13	8	17
2m	5x5	52.88%	18	15	9	13
3m	6x6	52.88%	23	17	8	7
2m	2x2	51.92%	15	14	8	17
2m	3x3	50.96%	17	12	9	15
0m	4x4	49.04%	17	13	8	13
1m	4x4	49.04%	17	13	8	13
2m	4x4	49.04%	16	12	11	12
3m	4x4	49.04%	16	12	10	13
3m	5x5	49.04%	17	16	7	11
0m	6x6	49.04%	21	11	9	10
1m	6x6	49.04%	15	18	9	9
2m	6x6	48.08%	22	14	7	7
0m	7x7	46.15%	22	12	6	8
1m	5x5	45.19%	14	15	4	14
0m	5x5	44.23%	13	14	5	14
0m	9x9	44.23%	20	8	12	6
0m	8x8	43.27%	17	13	8	7
2m	7x7	41.35%	14	20	6	3
3m	7x7	41.35%	14	22	4	3
1m	7x7	40.38%	14	21	4	3
2m	8x8	37.50%	13	24	1	1
3m	8x8	34.62%	10	23	2	1
0m	10x10	34.62%	17	6	10	3
1m	8x8	33.65%	8	23	2	2
2m	9x9	33.65%	9	24	2	0
3m	9x9	31.73%	8	24	1	0
1m	9x9	30.77%	6	24	1	1
2m	10x10	29.81%	5	26	0	0
1m	10x10	28.85%	3	26	1	0
3m	10x10	27.88%	3	26	0	0

Όσον αφορά την επέκταση της ακτίνας του υπολογισμού των LBP χαρακτηριστικών για όλες τις πιθανές τιμές από την τιμή 1 μέχρι την μέγιστη τιμή της, πραγματοποιήθηκαν πειράματα για πλήθος παραθύρων με τιμές από 2x2 μέχρι 10x10. Τα ακριβή ποσοστά ταξινόμησης παρουσιάζονται στον πίνακα 2 κατά φθίνουσα σειρά.

Η στήλη «Classifier» περιέχει τους τέσσερις διαφορετικούς τρόπους επιλογής των πλησιέστερων γειτόνων που επεξηγούνται στην εικόνα 12. Η στήλη «Windows» είναι το πλήθος των παραθύρων που διαχωρίζονται οι αρχικές εικόνες και η στήλη «Radius» αναφέρεται στο μέγεθος της μέγιστης ακτίνας κατά τη διαδικασία δημιουργίας των LBP χαρακτηριστικών, όπως παρουσιάζονται στο υποκεφάλαιο 3.3.

Τα αποτελέσματα είναι βέλτιστα για τα μικρά πλήθη παραθύρων διαχωρισμού των αρχικών εικόνων, δηλαδή για τα μεγαλύτερα παράθυρα ως προς το μέγεθός τους σε σχέση με τις υπόλοιπες περιπτώσεις. Πιο συγκεκριμένα, για τιμές από 2x2 μέχρι 5x5 η ορθή ταξινόμηση ξεπερνά το 60%. Για τις μεγαλύτερες τιμές του πλήθους των παραθύρων, τα αποτελέσματα δεν ήταν το ίδιο επιτυχή, όπως ακριβώς συνέβη και στην απλή περίπτωση των LBP χαρακτηριστικών. Όσο μεγαλώνει το πλήθος των παραθύρων, όσο δηλαδή μικραίνει το μέγεθός τους, τόσο μειώνεται και η επιτυχία της ταξινόμησης των εικόνων.

Ο τρόπος επιλογής των πλησιέστερων γειτόνων που εμφάνισαν την ορθότερη ταξινόμηση είναι πάλι οι δύο πιο απλές περιπτώσεις. Η επιλογή του πρώτου πλησιέστερου γείτονα δίνει το καλύτερο αποτέλεσμα από όλα τα πειράματα, το οποίο ισούται με 67.31%. Η δεύτερη περίπτωση είναι η επιλογή όλων των πλησιέστερων γειτόνων που έχουν ελάχιστες αποστάσεις ίσες με την ελάχιστη απόσταση του πρώτου και παρουσιάζουν ακριβώς το ίδιο ποσοστό επιτυχούς ταξινόμησης με την περίπτωση επιλογής μόνο του πρώτου πλησιέστερου.

Οι κλάσεις που εμφανίζουν τα βέλτιστα αποτελέσματα κατά την ταξινόμησή τους είναι και σε αυτή την περίπτωση οι γλώσσες των Γερμανικών και των Ελληνικών. Τα ποσοστά είναι περίπου ίσα ή ξεπερνούν το 75%, με ελάχιστες εξαιρέσεις. Αντίθετα, η ταξινόμηση των γλωσσών των Αγγλικών και των Γαλλικών είναι περίπου στο 50% και παρατηρείται σε όλους τους πιθανούς συνδυασμούς των παραμέτρων. Αυτό έχει σαν αποτέλεσμα τη μείωση της ολικής απόδοσης του ταξινομητή.

Στον πίνακα 2 που ακολουθεί, παρουσιάζονται τα αποτελέσματα των πειραμάτων με τις αντίστοιχες τιμές των παραμέτρων για ποσοστά ορθής ταξινόμησης που ξεπερνούν το 60%. Τα βέλτιστα αποτελέσματα προέρχονται από πλήθος παραθύρων 3x3 και από μέγιστη ακτίνα ίση με την τιμή 10. Τα επόμενα καλύτερα αποτελέσματα προέρχονται από τον υπολογισμό των LBP με τιμή μέγιστης ακτίνας 6 και 7 και με πλήθος παραθύρων 3x3, όπως δηλαδή και το πλήθος των παραθύρων που επιφέρουν τα βέλτιστα αποτελέσματα.

Πίνακας 2: Αποτελέσματα ταξινόμησης επέκτασης LBP σε εύρος μέγιστων ακτίνων

Αποτελέσματα ταξινόμησης επέκτασης LBP με μέγιστες τιμές ακτίνας από 3 μέχρι 10 και πλήθος παραθύρων από 2x2 μέχρι 5x5 (σε φθίνουσα σειρά ως προς το «Rate»)							
Classifier	Windows	Radius	Rate	German	Greek	English	French
0m	3x3	10	67.31%	22	22	13	13
1m	3x3	10	67.31%	22	22	13	13
0m	3x3	6	64.42%	22	20	10	15

1m	3x3	6	64.42%	22	20	10	15
0m	3x3	7	64.42%	22	20	10	15
1m	3x3	7	64.42%	22	20	10	15
0m	4x4	10	64.42%	23	20	12	12
1m	4x4	10	64.42%	23	20	12	12
0m	5x5	10	64.42%	23	19	13	12
0m	3x3	9	63.46%	22	19	12	13
1m	3x3	9	63.46%	22	19	12	13
0m	5x5	8	63.46%	22	19	12	13
1m	5x5	8	63.46%	22	19	12	13
2m	5x5	9	63.46%	23	20	11	12
1m	5x5	10	63.46%	23	19	13	11
3m	5x5	10	63.46%	23	19	12	12
0m	2x2	4	62.50%	20	16	13	16
1m	2x2	4	62.50%	20	16	13	16
0m	3x3	5	62.50%	21	18	10	16
1m	3x3	5	62.50%	21	18	10	16
0m	3x3	8	62.50%	22	20	10	13
1m	3x3	8	62.50%	22	20	10	13
3m	3x3	8	62.50%	22	19	10	14
3m	3x3	9	62.50%	21	20	10	14
3m	3x3	10	62.50%	22	19	9	15
0m	4x4	7	62.50%	23	18	12	12
1m	4x4	7	62.50%	23	18	12	12
3m	5x5	9	62.50%	23	20	11	11
3m	2x2	4	61.54%	21	15	12	16
0m	2x2	9	61.54%	19	20	12	13
1m	2x2	9	61.54%	19	20	12	13
0m	2x2	10	61.54%	19	20	12	13
1m	2x2	10	61.54%	19	20	12	13
0m	4x4	3	61.54%	20	14	13	17
1m	4x4	3	61.54%	20	14	13	17
0m	4x4	5	61.54%	20	18	11	15
1m	4x4	5	61.54%	20	18	11	15
0m	4x4	8	61.54%	22	18	11	13
1m	4x4	8	61.54%	22	18	11	13
0m	4x4	9	61.54%	22	19	11	12
1m	4x4	9	61.54%	22	19	11	12
1m	5x5	6	61.54%	21	18	12	13
3m	5x5	8	61.54%	23	19	10	12
2m	5x5	10	61.54%	23	19	10	12
1m	6x6	4	61.54%	20	19	11	14
3m	6x6	4	61.54%	23	19	10	12

Εφόσον παρατηρήθηκε ότι το ποσοστό επιτυχούς ταξινόμησης των εικόνων είναι βέλτιστο για τιμή μέγιστης ακτίνας ίση με 10, έγινε δοκιμή και με ακόμη μεγαλύτερες τιμές. Για την ακρίβεια πραγματοποιήθηκαν πειράματα για μέγιστη ακτίνα με τιμή από 11 μέχρι και 20, αλλά δεν υπήρξε περαιτέρω βελτίωση. Οπότε σαν βέλτιστη τιμή παρουσιάζεται η τιμή 10, η οποία εξασφαλίζει και αξιοποίηση λιγότερων υπολογιστικών πόρων σε σύγκριση με τις μεγαλύτερες τιμές.

Στον πίνακα 3 παρουσιάζονται τα αποτελέσματα για όλες τις πιθανές τιμές του πλήθους των παραθύρων και μέγιστη ακτίνα ίση με 10, η οποία εμφάνισε το μεγαλύτερο ποσοστό επιτυχίας. Στόχος είναι η μελέτη της επίδρασης του πλήθους των παραθύρων στην ταξινόμηση των εικόνων, αλλά και των διαφορετικών τρόπων επιλογής των πλησιέστερων γειτόνων.

Η στήλη «Classifier» αναφέρεται στους τέσσερις διαφορετικούς τρόπους επιλογής των πλησιέστερων γειτόνων που επεξηγούνται στην εικόνα 12. Η στήλη «Windows» είναι το πλήθος των παραθύρων που διαχωρίζονται οι αρχικές εικόνες και η στήλη «Radius» είναι η μέγιστη ακτίνα κατά τη διαδικασία δημιουργίας των LBP χαρακτηριστικών, όπως περιγράφονται στο υποκεφάλαιο 3.3. Η μέγιστη ακτίνα έχει την τιμή 10, η οποία αποδείχτηκε προηγουμένως ότι επιφέρει τα βέλτιστα αποτελέσματα.

Παρατηρείται ότι καθώς μεγαλώνει το πλήθος των παραθύρων, το ποσοστό ορθής ταξινόμησης μειώνεται όλο και περισσότερο. Ειδικά για τις μεγαλύτερες τιμές του πειράματος, 8x8 μέχρι 10x10, το ποσοστό επιτυχίας είναι κάτω από το 50%. Αυτό συμβαίνει για όλους τους διαφορετικούς τρόπους επιλογής των πλησιέστερων γειτόνων, εκτός από λίγες εξαιρέσεις που κυμαίνονται μεταξύ των ποσοστών 50% και 60%.

Σχετικά με την ταξινόμηση των γλωσσών, οι κλάσεις που ταξινομούνται με μεγαλύτερη επιτυχία είναι τα Γερμανικά και τα Ελληνικά. Αυτό παρατηρείται για την πλειοψηφία των πιθανών τιμών του πλήθους των παραθύρων που διαχωρίζονται οι αρχικές εικόνες αλλά και για όλους τους πιθανούς τρόπους επιλογής των πλησιέστερων γειτόνων.

Πίνακας 3: Αποτελέσματα ταξινόμησης επέκτασης LBP με μέγιστη ακτίνα 10

Αποτελέσματα ταξινόμησης επέκτασης LBP με μέγιστη ακτίνα 10 και πλήθος παραθύρων από 2x2 μέχρι 10x10 (σε φθίνουσα σειρά ως προς το «Rate»)							
Classifier	Windows	Radius	Rate	German	Greek	English	French
0m	3x3	10	67.31%	22	22	13	13
1m	3x3	10	67.31%	22	22	13	13
0m	4x4	10	64.42%	23	20	12	12
1m	4x4	10	64.42%	23	20	12	12
0m	5x5	10	64.42%	23	19	13	12
1m	5x5	10	63.46%	23	19	13	11
3m	5x5	10	63.46%	23	19	12	12
3m	3x3	10	62.50%	22	19	9	15
0m	2x2	10	61.54%	19	20	12	13
1m	2x2	10	61.54%	19	20	12	13
2m	5x5	10	61.54%	23	19	10	12
2m	3x3	10	59.62%	21	19	10	12
0m	7x7	10	59.62%	24	16	7	15

2m	4x4	10	58.65%	22	19	11	9
3m	4x4	10	58.65%	23	19	11	8
0m	6x6	10	57.69%	21	20	8	11
1m	6x6	10	57.69%	19	21	9	11
2m	6x6	10	57.69%	22	20	8	10
3m	6x6	10	57.69%	22	20	8	10
0m	8x8	10	56.73%	23	18	10	8
0m	9x9	10	56.73%	23	16	6	14
2m	2x2	10	54.81%	19	17	9	12
3m	2x2	10	54.81%	18	18	8	13
2m	7x7	10	54.81%	20	22	10	5
0m	10x10	10	53.85%	23	14	6	13
3m	7x7	10	52.88%	17	23	10	5
1m	7x7	10	50.00%	16	23	7	6
2m	8x8	10	39.42%	11	25	3	2
3m	8x8	10	38.46%	10	25	3	2
2m	9x9	10	38.46%	11	25	3	1
3m	9x9	10	38.46%	8	26	4	2
1m	8x8	10	35.58%	8	25	3	1
1m	9x9	10	34.62%	6	26	3	1
2m	10x10	10	33.65%	6	26	2	1
3m	10x10	10	32.69%	6	26	2	0
1m	10x10	10	28.85%	3	26	1	0

Οι παρατηρήσεις σε αυτό το πείραμα συμφωνούν απόλυτα με αυτές των προηγούμενων πειραμάτων, με αποτέλεσμα τα συμπεράσματα που αφορούν την επιλογή των πλησιέστερων γειτόνων, την επιλογή του πλήθους των παραθύρων αλλά και τον επιτυχή διαχωρισμό των γλωσσών να ενισχύονται.

Συνοπτικά λοιπόν, η τιμή 3x3 για το πλήθος των παραθύρων, η επιλογή του πρώτου πλησιέστερου γείτονα και των τόσων πλησιέστερων γειτόνων που έχουν ίση ελάχιστη απόσταση με τον πρώτο, αλλά και η τιμή 10 για μέγιστη ακτίνα υπολογισμού των LBP, παρουσιάζουν ποσοστό ορθής ταξινόμησης ίσο με 67.31% που είναι και το βέλτιστο.

7.2 Πειραματικά αποτελέσματα SIFT χαρακτηριστικών

Για τον έλεγχο των μεθόδων ταξινόμησης των εικόνων με τη χρήση του SIFT αλγορίθμου, τα πειράματα πραγματοποιήθηκαν στο ίδιο σύνολο εικόνων [21] με τα LBP χαρακτηριστικά. Τα αποτελέσματα παρουσιάζονται στον πίνακα 4, όπου η στήλη «Classifier» αναφέρεται στους διαφορετικούς ταξινομείς των εικόνων, που παρουσιάστηκαν στο υποκεφάλαιο 5.2.

Όλες οι εικόνες είναι σε δυαδική μορφή και πραγματοποιείται κλιμάκωση τους πριν εισαχθούν στη μέθοδο. Πιο συγκεκριμένα, σμικρύνονται πολλαπλασιάζοντας την κάθε τους πλευρά με τον αριθμό που επιλέγει ο χρήστης ώστε να μειωθεί το μέγεθός τους για αξιοποίηση λιγότερων υπολογιστικών πόρων. Τα αποτελέσματα που παρουσιάζονται στον

πίνακα 4, προέρχονται από κλιμάκωση των εικόνων εισόδου, οι οποίες πολλαπλασιάστηκαν με τον αριθμό 0.65.

Παρατηρώντας τα ποσοστά ορθής ταξινόμησης, φαίνεται ξεκάθαρα η υπεροχή του ταξινομητή Local NBNN έναντι του NBNN στην εφαρμογή του σε χειρόγραφες εικόνες, από τις οποίες εξάγονται τα χαρακτηριστικά μέσω του SIFT αλγορίθμου. Ο έλεγχος της διαφοράς των προσανατολισμών των keypoints των SIFT χαρακτηριστικών και η κανονικοποίηση των τελικών αποστάσεων από κάθε κλάση με το πλήθος των keypoints κάθε κλάσης εφαρμόζονται και στους δύο.

Όσον αφορά την ταξινόμηση των διαφορετικών γλωσσών, οι κλάσεις που παρουσιάζουν τα βέλτιστα αποτελέσματα κατά τον διαχωρισμό τους είναι οι γλώσσες των Γερμανικών και των Ελληνικών και στους δύο αλγορίθμους ταξινόμησης. Το ποσοστό των Γερμανικών είναι περίπου ίσο με 96% και στους δύο ταξινομητές, ενώ των Ελληνικών είναι περίπου 65% στον NBNN, αλλά 100% στον Local NBNN. Παρατηρείται λοιπόν μεγάλη αύξηση του ποσοστού επιτυχίας διαχωρισμού των Ελληνικών στον Local NBNN, καταφέρνοντας την ορθή ταξινόμηση όλων των εικόνων.

Η κλάση των Γαλλικών παρουσιάζει ποσοστό ορθής ταξινόμησης κάτω από το 50% στον NBNN, αλλά στον Local NBNN ταξινομούνται όλα τα δείγματα σωστά. Αντίθετα, η ταξινόμηση των Αγγλικών δεν είναι επιτυχής σε κανέναν από τους δύο αλγορίθμους, καθώς διαχωρίζονται σωστά μόνο οι μισές εικόνες στον Local NBNN και στον NBNN το ποσοστό τους πέφτει κάτω από το 50%.

Η ανεπιτυχής λοιπόν ταξινόμηση των Αγγλικών στους δύο ταξινομητές και των Γαλλικών και Ελληνικών μόνο στον ταξινομητή NBNN, είναι ο λόγος που επηρεάζεται η ολική απόδοση τους.

Πίνακας 4: Αποτελέσματα ταξινόμησης SIFT

Ταξινομητές SIFT					
Classifier	**Rate**	**German**	**Greek**	**English**	**French**
Κανονικοποιημένος Local NBNN με κατώφλι διαφοράς προσανατολισμού	87.50%	25	26	14	26
Κανονικοποιημένος NBNN με κατώφλι διαφοράς προσανατολισμού	63.46%	25	17	12	12

7.3 Σύγκριση βέλτιστων αποτελεσμάτων

Στον πίνακα 5, παρουσιάζονται συγκεντρωμένα τα αποτελέσματα ταξινόμησης όλων των μεθόδων.

Όσον αφορά τα αποτελέσματα ταξινόμησης των SIFT χαρακτηριστικών του πίνακα, η στήλη «Classifier-SIFT» αναφέρεται στους διαφορετικούς ταξινομητές των εικόνων, όπως επεξηγήθηκαν στο υποκεφάλαιο 5.2. Παρατηρείται ότι το μέγιστο ποσοστό ορθής ταξινόμησης προκύπτει από τον Local NBNN αλγόριθμο. Με εξαίρεση την κλάση των Αγγλικών, διαχωρίζει όλες τις κατηγορίες σχεδόν με απόλυτη επιτυχία.

Η υπεροχή του οφείλεται στο γεγονός ότι αναζητά τους πιο κοντινούς γείτονες των χαρακτηριστικών μόνο στις κλάσεις που βρίσκονται εντός μιας συγκεκριμένης γειτονίας και

όχι στην αναζήτηση του κοντινότερου γείτονα σε όλες τις κλάσεις, όπως λειτουργεί ο NBNN. Επίσης, ο έλεγχος του προσανατολισμού των keypoints και η κανονικοποίηση των ολικών αποστάσεων των κλάσεων με το πλήθος των keypoints κάθε κλάσης, βελτιώνουν αισθητά την ολική του απόδοση.

Σχετικά με τα αποτελέσματα ταξινόμησης των LBP χαρακτηριστικών που παρουσιάζονται στον πίνακα, η στήλη «Classifier-LBP» αναφέρεται στους τέσσερις διαφορετικούς τρόπους επιλογής των πλησιέστερων γειτόνων, όπως επεξηγούνται στην εικόνα 12. Η στήλη «Windows» είναι το πλήθος των παραθύρων που διαχωρίζονται οι αρχικές εικόνες και η στήλη «Radius» αναφέρεται στο μέγεθος της μέγιστης ακτίνας κατά τη διαδικασία δημιουργίας των LBP χαρακτηριστικών, όπως παρουσιάζονται στο υποκεφάλαιο 3.3.

Το βέλτιστο αποτέλεσμα προέρχεται από την επέκταση του κλασικού αλγορίθμου υπολογισμού των χαρακτηριστικών. Αυξάνοντας την ακτίνα και λαμβάνοντας υπόψη όλες τις πιθανές τιμές εντός του εύρους με ελάχιστη τιμή τη μονάδα και μέγιστη την τιμή 10, τα αποτελέσματα βελτιώνονται αισθητά σε σύγκριση με την απλή μορφή υπολογισμού των χαρακτηριστικών.

Επίσης, από τα πειράματα αποδεικνύεται ότι το πλήθος παραθύρων που διαχωρίζονται οι αρχικές εικόνες πρέπει να είναι σχετικά μικρό, δηλαδή το μέγεθός τους να είναι μεγάλο. Είναι λογικό καθώς η πληροφορία που εξάγεται από αυτά πριν ταξινομηθούν στις κλάσεις που ανήκουν είναι μεγαλύτερη άρα και πιο ακριβής.

Σχετικά με τους τρόπους επιλογής των κοντινότερων γειτόνων, οι καλύτερες επιλογές είναι του κοντινότερου γείτονα με την ελάχιστη απόσταση και των τόσων κοντινότερων γειτόνων όσων έχουν ίση ελάχιστη απόσταση με τον πρώτο.

Πίνακας 5: Σύγκριση βέλτιστων αποτελεσμάτων

Ταξινομητές SIFT							
Classifier-SIFT			**Rate**	**German**	**Greek**	**English**	**French**
Κανονικοποιημένος Local NBNN με κατώφλι διαφοράς προσανατολισμού			87.50%	25	26	14	26
Κανονικοποιημένος NBNN με κατώφλι διαφοράς προσανατολισμού			63.46%	25	17	12	12
Ταξινομητές LBP							
Classifier-LBP	**Windows**	**Radius**	**Rate**	**German**	**Greek**	**English**	**French**
0m	3x3	10	67.31%	22	22	13	13
1m	3x3	10	67.31%	22	22	13	13
0m	3x3	1	57.69%	18	16	11	15
1m	3x3	1	57.69%	18	16	11	15

8. ΣΥΜΠΕΡΑΣΜΑΤΑ

Η ταξινόμηση εικόνων είναι μια διαδικασία κατά την οποία διαχωρίζονται οι εικόνες σύμφωνα με το οπτικό τους περιεχόμενο. Ενώ για τον άνθρωπο πρόκειται για μια τετριμμένη διαδικασία, για τον τομέα της μηχανικής μάθησης δεν είναι τόσο απλή. Η αποτελεσματική ταξινόμηση αποτελεί πρόκληση σε όλες τις εφαρμογές.

Το αντικείμενο της διπλωματικής αυτής είναι ο εντοπισμός της γλώσσας που είναι γραμμένο ένα χειρόγραφο έγγραφο. Πρόκειται για ένα πολύπλοκο πρόβλημα καθώς χαρακτηρίζεται από πολλούς παράγοντες και απαιτεί την υποδιαίρεση του σε επιμέρους μικρότερα προβλήματα. Η επιλογή των κατάλληλων χαρακτηριστικών που θα εξαχθούν από τις εικόνες των χειρόγραφων εγγράφων ώστε να τις εκπροσωπήσουν καθώς και των ταξινομητών που θα διαχωρίσουν επιτυχώς τα δείγματα, είναι δύο από αυτά.

Υπάρχει πληθώρα χαρακτηριστικών και η ορθή επιλογή τους είναι βασική προϋπόθεση για την επιτυχή ταξινόμηση. Η χρήση των SIFT αποδείχθηκε ότι έχει πολύ καλά αποτελέσματα σε δυαδικές εικόνες ψηφιοποιημένων εγγράφων αν επιλεχθεί και ο κατάλληλος αλγόριθμος ταξινόμησης, που στη συγκεκριμένη περίπτωση είναι ο Local NBNN. Αξιοσημείωτο για την επιτυχία του αλγορίθμου είναι το γεγονός ότι εφαρμόστηκε έλεγχος στους προσανατολισμούς των keypoints των SIFT χαρακτηριστικών πριν ταιριάξουν μεταξύ τους. Τα μοτίβα των γλωσσών αποδίδουν πολλά keypoints με παρόμοια χαρακτηριστικά αλλά διαφορετικούς προσανατολισμούς. Δεδομένου ότι ο προσανατολισμός παρέχει μια ξεχωριστή ιδιότητα, έπρεπε να αποφευχθεί η αντιστοίχιση των όμοιων περιγραφέων με διαφορετικό προσανατολισμό. Επίσης, η κανονικοποίηση των ολικών αποστάσεων των κλάσεων με το πλήθος των keypoints κάθε κλάσης, πριν ληφθεί η απόφαση ταξινόμησης, βελτιώνει αισθητά την απόδοση του ταξινομητή.

Αντίθετα, η εφαρμογή των LBP χαρακτηριστικών στην απλή τους μορφή δεν ήταν επιτυχής σε συνδυασμό με τον ταξινομητή των πλησιέστερων γειτόνων. Παρόλα αυτά, αποδείχθηκε ότι η επέκταση του LBP αλγορίθμου, που παρουσιάστηκε, μπορεί να βελτιώσει την ταξινόμηση κατά σημαντικό βαθμό. Αυτό οφείλεται κυρίως στο γεγονός ότι αυξάνεται η πληροφορία που εξάγεται από τις εικόνες και τις εκπροσωπεί με αποτέλεσμα να λειτουργούν καλύτερα και οι ταξινομητές με τους οποίους διαχωρίζονται τα δείγματα.

Το παραπάνω αποδεικνύεται και από το γεγονός ότι η ταξινόμηση είναι βέλτιστη για μικρό πλήθος παραθύρων στα οποία διαχωρίζονται οι εικόνες, δηλαδή για παράθυρα που έχουν μεγάλο μέγεθος, άρα και περισσότερη πληροφορία. Αντίθετα, όσο μεγαλώνει το πλήθος των παραθύρων, άρα το μέγεθος του καθενός γίνεται όλο και πιο μικρό, τόσο μειώνεται και η επιτυχία της ταξινόμησης των εικόνων.

Η απόδοση των προτεινόμενων μεθόδων εφαρμόστηκε και αξιολογήθηκε στο σύνολο εικόνων benchmarking dataset ICDAR 2011 Writer Identification Contest [21] [22]. Η αξιολόγηση της απόδοσης έδειξε τη διακριτική ισχύ των προτεινόμενων μεθόδων σε σχέση με διαφορετικά είδη γραφής και γλώσσες, καθώς το σύνολο εικόνων δημιουργήθηκε με τη βοήθεια 26 διαφορετικών γραφέων και αποτελείται από 4 διαφορετικές γλώσσες (Αγγλικά, Γαλλικά, Γερμανικά, Ελληνικά). Επομένως τα αποτελέσματα έδειξαν τη γενικότητα των μεθόδων και τη δυνατότητα κλιμάκωσης σε μεγάλο αριθμό κατηγοριών.

ΠΙΝΑΚΑΣ ΟΡΟΛΟΓΙΑΣ

Ξενόγλωσσος Όρος	**Ελληνικός Όρος**
Aspect ratio	Αναλογία απεικόνισης
Assignment	Ανάθεση
Binary	Δυαδικό
Blur	Θόλωμα
Browsing	Περιήγηση
Classifier	Ταξινομητής
Clean	Καθαρός
Clustering	Συσταδοποίηση
Codebook-based	Βιβλίο κωδίκων
Coding	Κωδικοποίηση
Color	Χρώμα
Computer vision	Μηχανική όραση
Connected component analysis	Ανάλυση συνδεδεμένων συστατικών
Connected component features	Χαρακτηριστικά συνδεδεμένων συστατικών
Connected regions	Συνδεδεμένες περιοχές
Continuous curves	Συνεχείς καμπύλες
Contour	Περίγραμμα
Contour directional feature	Χαρακτηριστικό κατεύθυνσης περιγράμματος
Contrast threshold	Κατώφλι αντίθεσης
Corner	Γωνία
Cross-platform	Ανεξάρτητο πλατφόρμας
Cross-validation	Διασταυρούμενη επικύρωση
Curvature	Καμπυλότητα
Curve	Καμπύλη
Degradation	Υποβάθμιση
Descriptor	Περιγραφέας
Detection	Ανίχνευση
Direction	Κατεύθυνση
Document	Έγγραφο
Edge threshold	Κατώφλι άκρων
Edges	Άκρες
Extremum	Ακραίο σημείο
Feature	Χαρακτηριστικό
Gradient	Κλίση
Gray-level	Γκρίζου επιπέδου
Gray-scale	Γκρίζας κλίμακας
High-level	Υψηλού επιπέδου
High-pass filtering	Φιλτράρισμα υψηλής διέλευσης
Histogram	Ιστόγραμμα
Identification	Ταυτοποίηση
Image	Εικόνα
Ink	Μελάνι

Interest points	Σημεία ενδιαφέροντος
Invariant	Αμετάβλητο
Isolated points	Απομονωμένα σημεία
Junction	Διασταύρωση
Keypoint	Σημείο ενδιαφέροντος
Knowledge-based	Βασισμένη στη γνώση
Language	Γλώσσα
Layout	Διάταξη
Learning	Εκμάθηση
Level	Επίπεδο
Linear discriminant analysis	Γραμμική ανάλυση διακρίσεων
Local	Τοπικά
Localization	Εντοπισμός τοποθεσίας
Low-level	Χαμηλού επιπέδου
Magnitude	Μέγεθος
Maxima	Μέγιστο
Minima	Ελάχιστο
Multi-channel Gabor filtering technique	Πολυκαναλικής τεχνικής φιλτραρίσματος Gabor
Nearest Neighbour	Κοντινότερος γείτονας
Noise	Θόρυβος
Occurrence histograms	Ιστογράμματα εμφάνισης
Octaves	Οκτάβες
Offset	Αντιστάθμιση
Orientation	Προσανατολισμός
Pattern	Πρότυπο
Perspective	Προοπτική
Pixel	Εικονοστοιχείο
Probability distribution	Κατανομή πιθανότητας
Pre-processing	Προ-επεξεργασία
Principal curvatures	Κύριες καμπυλότητες
Processing	Επεξεργασία
Pyramid	Πυραμίδα
Rate	Ποσοστό
Region	Περιοχή
Relative centroid	Σχετικό κέντρο
Rotation	Περιστροφή
Rotation invariant texture features	Χαρακτηριστικά υφής αμετάβλητης περιστροφής
Run-length	Διατρέχον μήκος
Scale	Κλίμακα
Scale Invariant Feature Transform	Μετασχηματισμός χαρακτηριστικών αμετάβλητης κλίμακας
Scale-space	Χώρος κλίμακας
Scale-space filtering	Φιλτράρισμα στον χώρο κλίμακας
Scaling	Κλιμάκωση
Script	Είδος γραφής

Shape	Σχήμα
Skew	Στροφή
Software development environment	Περιβάλλον ανάπτυξης λογισμικού
Space	Χώρος
Sphericity	Σφαιρικότητα
Stroke-length	Μήκος διαδρομής σταθερού πλάτους
Structure	Δομή
Subpixels	Υποεικονοστοιχείο
Textural-based	Βασισμένο στο κείμενο
Texture	Υφή
Test set	Άγνωστο σύνολο δειγμάτων
Threshold	Όριο
Thresholding	Κατωφλίωση
Trace	Ίχνος
Train set	Σύνολο αναφοράς δειγμάτων
Unstable	Ασταθής
Vector	Διάνυσμα
Vectorization	Μετατροπή σε διάνυσμα
White holes	Λευκές τρύπες
Width	Πλάτος
Visual	Οπτικά
Word shape coding	Σχήμα κωδικοποίησης λέξης
Writer	Γραφέας

ΣΥΝΤΜΗΣΕΙΣ – ΑΡΚΤΙΚΟΛΕΞΑ – ΑΚΡΩΝΥΜΙΑ

3-D	3-dimensional
Blob	Binary large object
CDF	Contour Directional Feature
DoG	Difference of Gaussians
FAST	Features from Accelerated Segment Test
FLANN	Fast Library for Approximate Nearest Neighbours
GPRLT	General Pattern Run-Length Transform
GUI	Graphical User Interface
SIFT	Scale-space Extrema Detection
K-d	K-dimensional
LBP	Local Binary Patterns
LoG	Laplacian of Gaussian
MSER	Maximally Stable Extremal Regions
NN	Nearest Neighbour
NBNN	Naïve Bayes Nearest-Neighbour
SUSAN	Smallest Univalue Segment Assimilating Nucleus

ΑΝΑΦΟΡΕΣ

[1] T. Ojala, M. Pietikainen and T. Maenpaa, “Multiresolution gray-scale and rotation invariant texture classification with local binary patterns,” in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 7, pp. 971-987, July 2002.

[2] Lowe, David G. “Distinctive Image Features from Scale-Invariant Keypoints,” International Journal of Computer Vision, 2004, pp 91-110.

[3] O. Boiman, E. Shechtman and M. Irani, “In defense of Nearest-Neighbour based image classification,” 2008 IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, 2008, pp. 1-8.

[4] S. McCann and D. G. Lowe, “Local Naive Bayes Nearest Neighbour for image classification,” 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, 2012, pp. 3650-3656.

[5] A. Nicolaou, A. D. Bagdanov, M. Liwicki and D. Karatzas, “Sparse radial sampling LBP for writer identification,” 2015 13th International Conference on Document Analysis and Recognition (ICDAR), Tunis, 2015, pp. 716-720.

[6] H. Mohammed, V. Mäergner, T. Konidaris and H. S. Stiehl, “Normalised Local Naïve Bayes Nearest-Neighbour Classifier for Offline Writer Identification,” 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), Kyoto, 2017, pp. 1013-1018.

[7] Brink, J. Smit, M. Bulacu, and L. Schomaker, “Writer identification using directional ink-trace width measurements,” Pattern Recognition, vol. 45, no. 1, pp. 162–171, 2012.

[8] S. He and L. Schomaker, “General Pattern Run-Length Transform for Writer Identification,” 2016 12th IAPR Workshop on Document Analysis Systems (DAS), Santorini, 2016, pp. 60-65.

[9] S. He, M. Wiering, and L. Schomaker, “Junction detection in hand-written documents and its application to writer identification,” Pattern Recognition, vol. 48, no. 12, pp. 4036–4048, 2015.

[10] Y. Xiong, Y. Wen, P. S. P. Wang and Y. Lu, “Text-independent writer identification using SIFT descriptor and contour-directional feature,” 2015 13th International Conference on Document Analysis and Recognition (ICDAR), Tunis, 2015, pp. 91-95.

[11] J. Hochberg, K. Bowers, M. Cannon, and P. Kelly, “‘Script and language identification for handwritten document images,” IJDAR 45–52 (1999).

[12] T. N. Tan, “Rotation invariant texture features and their use in automatic script identification,” in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20, no. 7, pp. 751-756, July 1998.

[13] S. J. Lu, L. Li and C. L. Tan, “Identification of Latin-Based Languages through Character Stroke Categorization,” Ninth International Conference on Document Analysis and Recognition (ICDAR 2007), Parana, 2007, pp. 352-356.

[14] L. Shijian and C. Lim Tan, “Script and Language Identification in Noisy and Degraded Document Images,” in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 30, no. 1, pp. 14-24, Jan. 2008.

[15] Brown, Matthew, and David G. Lowe. “Invariant features from interest point groups,” BMVC. Vol. 4. 2002.

[16] C. Harris and M. Stephens, “A combined corner and edge detector.,” Alvey vision conference, vol. 15, pp. 10–5244, 1988.

[17] M. Muja and D. G. Lowe, “Fast approximate nearest neighbours with automatic algorithm configuration.,” in VISAPP (1), pp. 331–340, 2009.

[18] C++ Programming Language; http://www.cplusplus.com/ [Προσπελάστηκε 28/3/2020]

[19] Qt Software Development Environment; https://www.qt.io/ [Προσπελάστηκε 28/3/2020]

[20] OpenCV Library of Programming Functions; http://opencv.org/ [Προσπελάστηκε 28/3/2020]

[21] http://www.iit.demokritos.gr/~louloud/Writer_Identification_Contest/Benchmarking_Dataset.rar [Προσπελάστηκε 28/3/2020]

[22] G. Louloudis, N. Stamatopoulos and B. Gatos, “ICDAR 2011 Writer Identification Contest,” 2011 International Conference on Document Analysis and Recognition, Beijing, 2011, pp. 1475-1479.