

Word Spotting και Clustering εικόνων χειρόγραφων λέξεων

Η Διπλωματική Εργασία
παρουσιάστηκε ενώπιον
του Διδακτικού Προσωπικού του
Πανεπιστημίου Αιγαίου

Σε Μερική Εκπλήρωση
των Απαιτήσεων για το Δίπλωμα του
Μηχανικού Πληροφοριακών και Επικοινωνιακών Συστημάτων

του
Πολυχρόνη Μάριου και
Ευαγόρου Ανδρέα
ΕΑΡΙΝΟ ΕΞΑΜΗΝΟ 2011

Η ΤΡΙΜΕΛΗΣ ΕΠΙΤΡΟΠΗ ΔΙΔΑΣΚΟΝΤΩΝ ΕΠΙΚΥΡΩΝΕΙ
ΤΗ ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ
ΤΟΥ ΠΟΛΥΧΡΟΝΗ ΜΑΡΙΟΥ ΚΑΙ
ΕΥΑΓΟΡΟΥ ΑΝΔΡΕΑ:

ΚΑΒΑΛΛΙΕΡΑΤΟΥ ΕΡΓΙΝΑ , Επιβλέπων 30 ΙΟΥΝΙΟΥ 2011
Τμήμα Μηχανικών Πληροφοριακών και
Επικοινωνιακών Συστημάτων

ΣΤΑΜΑΤΑΤΟΣ ΕΥΣΤΑΘΙΟΣ, Μέλος
Τμήμα Μηχανικών Πληροφοριακών και
Επικοινωνιακών Συστημάτων

ΜΑΡΑΓΚΟΥΔΑΚΗΣ ΜΑΝΩΛΗΣ, Μέλος
Τμήμα Μηχανικών Πληροφοριακών και
Επικοινωνιακών Συστημάτων

ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ
ΕΑΡΙΝΟ ΕΞΑΜΗΝΟ 2011

ΠΕΡΙΛΗΨΗ

Η συγκεκριμένη διπλωματική εργασία, ασχολείται με την τεχνική του Word Spotting καθώς και την ομαδοποίηση (Clustering) εικόνων χειρόγραφων λέξεων. Σκοπός της εργασίας ήταν η τροποποίηση και βελτίωση κάποιων ήδη υπάρχον μεθόδων, για την ομαδοποίηση χειρόγραφων λέξεων, γραμμένες από πλήθος διαφορετικών ατόμων. Χρησιμοποιήθηκαν συνολικά πέντε μέθοδοι, εκ των οποίων οι δύο πάρθηκαν από τη βιβλιογραφία ως πρότυπο και με βάση αυτές, δημιουργήθηκαν και υλοποιήθηκαν τρεις νέες μέθοδοι με σκοπό να φέρουν καλύτερα αποτελέσματα από τις δύο ήδη υπάρχουσες. Για το σκοπό της σύγκρισης των δύο υπάρχον και των τριών νέων μεθόδων, δημιουργήθηκαν τρία διαφορετικά σύνολα δεδομένων, τα οποία αξιολογήθηκαν με βάση τέσσερα μέτρα αξιολόγησης (Precision, Recall, F-measure, Purity). Αξίζει να σημειωθεί, ότι μέσα από τη σύγκριση των εν λόγω μεθόδων, προέκυψαν καλύτερα αποτελέσματα ομαδοποίησης στις τρεις νέες μεθόδους που υλοποιήθηκαν από τις δύο ήδη υπάρχουσες, επιτυγχάνοντας έτσι το σκοπό της εργασίας.

Συγκεκριμένα στο κεφάλαιο 1 δίνεται μία μικρή εισαγωγή της εργασίας αυτής. Στο κεφάλαιο 2, παρουσιάζονται κάποια γενικά στοιχεία του Word Spotting, αρχίζοντας με τον ορισμό της τεχνικής του, και συνεχίζοντας με τα προβλήματα τα οποία καλείται να αντιμετωπίσει, την βασική του ιδέα, τα στάδια από τα οποία αποτελείται, καθώς και κάποιες εργασίες που έχουν γίνει γύρω από το θέμα αυτό. Επίσης, γίνεται μια μικρή αναφορά σχετικά με την αποτυχία του Optical Character Recognition (OCR), στις περιπτώσεις χειρόγραφων εγγράφων. Στο κεφάλαιο 3, παρουσιάζονται βασικά στάδια υλοποίησης, που περιλαμβάνονται μέσα στο Word Spotting, τα οποία είναι απαραίτητα τόσο για την επιτυχία του Word Spotting, όσο για τη σωστή ομαδοποίηση των χειρόγραφων εικόνων, καθώς επίσης δίνεται και μια περιγραφή δύο ειδών μετασχηματισμού, οι οποίοι είναι απαραίτητοι για την υλοποίηση των μεθόδων της εργασίας. Στο κεφάλαιο 4, αρχικά δίνεται μια μικρή εισαγωγή του κεφαλαίου, την οποία ακολουθεί μια σύντομη περιγραφή των 5 μεθόδων που υλοποιήθηκαν για τους σκοπούς της εργασίας. Στη συνέχεια, παρουσιάζονται τα σύνολα δεδομένων που χρησιμοποιούνται στα πειράματα της εργασίας, καθώς και η μέθοδος αξιολόγησης των πειραμάτων αυτών. Τέλος, δίνεται αναλυτική περιγραφή της υλοποίησης των πέντε σεναρίων, δηλαδή, των δύο ήδη υπάρχον και των τριών νέων προτεινόμενων, τα οποία έχουν στόχο την ομαδοποίηση εικόνων χειρόγραφων λέξεων διάφορων συγγραφέων. Στο κεφάλαιο 5, παρουσιάζονται τα συγκριτικά αποτελέσματα που έδωσαν οι 5 μέθοδοι, και δίνονται τα συμπεράσματα που εξάχθηκαν από τα αποτελέσματα αυτά.

© 2011

του

ΠΟΛΥΧΡΟΝΗ ΜΑΡΙΟΥ ΚΑΙ ΕΥΑΓΓΟΡΟΥ ΑΝΔΡΕΑ

Τμήμα Μηχανικών Πληροφοριακών και Επικοινωνιακών Συστημάτων

ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ

ABSTRACT

In this dissertation, some stages of the Word Spotting technique are applied, and more particularly, the stages that are needed, so that the Clustering of handwritten word images can be performed successfully. The main goal of this project was to produce the implementations of 2 methods that have been already proposed in an older paper, and based on those implementations, apply some modifications, so that some new, better performed, methods can be created. A significant difference between the initial methods and the ones proposed here, is that the initial methods were created with the target of applying the Word Spotting technique on handwritten document images of one (or a small group of) writer(s), while, on the other hand, the proposed methods of this dissertation, were created with the target of applying the Clustering process of handwritten word images of many writers. Although not all the stages of the Word Spotting technique were applied in the presented methods, it is our belief that the hard work has been done, and that adding the rest of the stages is not of huge importance.

More specifically, the first of the initial methods made use of the Dynamic Time Warping (DTW) algorithm in the Matching stage, while the second one made use of the Discrete Fourier Transform (DFT), to transform the extracted features of the word images. While our implementation of those two methods might not be an exact match of the implementations of the original paper, we believe that it came very close to the originals, as we followed almost all the stages presented in that paper. The method that we implemented and was purporting the original first method was method 1 of this project, while the method that we implemented and was purporting the original second method was method 3. One of the new methods created in this project (method 2), was a modification of method 1, where the modifications that were implemented were mostly in the pre-processing stage, the feature extraction stage and the Clustering stage. In more detail, the pre-processing was chosen so that it would work best for our data sets, some new features were added in the list of extracted features from each word image, and a different Clustering algorithm was used, with a different distance metric. Another one of the new methods created during this dissertation (method 4), was a modification of method 3, where the modifications that were implemented were mostly in the pre-processing stage, and the feature extraction stage. In more detail, the pre-processing again was chosen so that it would work best for our data sets, and some new features were added in the list of extracted features from each word image. The last one of the methods proposed (method 5), is an alternate approach of method 4, where instead of the DFT, the Discrete Cosine Transform (DCT) was used to transform the features extracted from each image. Three data sets were used in our experiments, where two of them were of good quality (one of few images (200), and one of many images (1008)), and one of them (this set was also consisted of 200 images), contained randomly taken images from a larger set that contained both good and bad quality's word images. For the evaluation of the performance of the Clustering process, four performance indicators were used: precision, recall, F-measure, and purity.

For a small description of the dissertation, in chapter 1, a small introduction is given, while in chapter 2, a general approach of the Word Spotting technique is given. Specifically, in chapter 2 a definition of the technique, and some of the problems that it has to face are provided, the reason why Optical Character Recognition (OCR) fails in the case of handwritten documents is given, the basic idea of Word Spotting is presented, followed by a description of its stages. Finally, the chapter ends with a brief discussion of some related work that has been done on this field. In chapter 3, the basic stages of the Word Spotting technique which are implemented in this dissertation are presented, while a small description of the two transformations that are used in some of the implemented methods is also provided. Proceeding to chapter 4, a small introduction

of the chapter is found, followed by a brief description of the methods implemented in this dissertation. Then, a presentation of the data sets is given, and the evaluation method of our experiments is explained, an extensive description of the implementations of the methods is also given, and finally, we present a Graphical User Interface (GUI), which was created so that the use of our methods can be performed with ease. Chapter 5 contains the final results of all 5 methods, and the conclusions extracted from those results.

By viewing the final results, it is our opinion that we have accomplished our goal, by managing to modify the original methods and got them to give better results, and by making them able to be applied successfully in handwritten word images written by many different writers.

© 2011

POLYCHRONIS MARIOS AND EVAGOROU ANDREAS

Department of Information and Communication Systems Engineering

UNIVERSITY OF THE AEGEAN

ΕΥΧΑΡΙΣΤΙΕΣ - ΑΦΙΕΡΩΣΕΙΣ

Στο σημείο αυτό θα θέλαμε αρχικά, να ευχαριστήσουμε ιδιαίτερος την επιβλέπουσα λέκτορα της διπλωματικής μας εργασίας κα. Εργίνα Καβαλλιεράτου, για την ουσιαστική βοήθεια που προσέφερε με την άψογη συνεργασία, τις πολύτιμες συμβουλές και την καθοδήγησή της σε όλη την διάρκεια της εκπόνησης αυτής της διπλωματικής εργασίας.

Επίσης, ευχαριστούμε την επιτροπή της διπλωματικής μας εργασίας, και συγκεκριμένα τον επίκουρο καθηγητή κ. Ευστάθιο Σταματάτο, για τις παρατηρήσεις και τις υποδείξεις του στη βελτίωση της εργασίας αυτής, καθώς επίσης και τον λέκτορα κ. Εμμανουήλ Μαραγκουδάκη, για τις βασικές γνώσεις που μας είχε δώσει μέσα από τα μαθήματά του, τα οποία ήταν αναγκαία σε πολλά σημεία στην εργασία αυτή.

ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

ΠΕΡΙΛΗΨΗ	iii
ABSTRACT	iv
ΕΥΧΑΡΙΣΤΙΕΣ - ΑΦΙΕΡΩΣΕΙΣ	vi
ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ	vii
ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ	ix
ΚΑΤΑΛΟΓΟΣ ΣΧΗΜΑΤΩΝ	xi
ΚΕΦΑΛΑΙΟ 1 - ΕΙΣΑΓΩΓΗ	1
ΚΕΦΑΛΑΙΟ 2 - ΓΕΝΙΚΑ ΓΙΑ ΤΟ WORD SPOTTING.....	2
2.1 Ορισμός του Word Spotting	2
2.2 Προβλήματα που καλείται να αντιμετωπίσει η τεχνική Word Spotting	2
2.3 Αποτυχία του OCR στην αναγνώριση χειρόγραφων λέξεων	2
2.4 Βασική ιδέα του Word Spotting	3
2.5 Στάδια του Word Spotting.....	4
2.6 Στάδια του Word Spotting που μελετώνται.....	4
2.7 Εργασίες που έγιναν στο Word Spotting.....	5
2.7.1 Εργασία με θέμα « Word Spotting for Historical Documents ».....	5
2.7.2 Εργασία με θέμα « Features for Word Spotting in Historical Manuscripts ».....	6
2.7.3 Εργασία με θέμα « Word Spotting: Indexing Handwritten Archives ».....	6
2.7.4 Εργασία με θέμα « Nearest Neighbor based Collection OCR ».....	7
2.7.5 Εργασία με θέμα « A probabilistic method for keyword retrieval in handwritten document images »	8
2.7.6 Εργασία με θέμα « Scale Space technique for word segmentation in handwritten manuscripts ».....	9
2.7.7 Εργασία με θέμα « Indexing for a Digital Library of George Washington’s manuscripts – A study of word Matching techniques».....	10
2.7.8 Εργασία με θέμα « Word image Matching using dynamic time warping»	11
2.7.9 Εργασία με θέμα « Application of multi-level classifiers and Clustering for automatic Word Spotting in historical document images»	12
2.7.10 Εργασία με θέμα « Keyword-guided Word Spotting in historical printed documents using synthetic data and user feedback».....	12
ΚΕΦΑΛΑΙΟ 3 - ΣΤΑΔΙΑ ΥΛΟΠΟΙΗΣΗΣ	14
3.1 Προ-Επεξεργασία (Preprocessing)	14
3.1.1 Εφαρμογές προ-επεξεργασίας (Preprocessing Applications).....	14
3.2 Εξαγωγή Χαρακτηριστικών (Feature Extraction)	19
3.2.1 Χαρακτηριστικά (Features)	19

3.2.2 Παρεμβολή Χαρακτηριστικών (Interpolation).....	26
3.2.3 Κανονικοποίηση Χαρακτηριστικών (Feature Normalization)	27
3.2.4 Κανονικοποίηση Μεγέθους (Standardization)	28
3.3 Matching.....	29
3.3.1 Dynamic Time Warping	30
3.4 Διακριτοί Μετασχηματισμοί	32
3.4.1 Διακριτός Μετασχηματισμός Fourier (Discrete Fourier Transform - DFT)	32
3.4.2 Διακριτός Μετασχηματισμός Συνημίτονου (Discrete Cosine Transform – DCT).....	35
3.5 Ομαδοποίηση (Clustering)	37
3.5.1 Ιεραρχικοί αλγόριθμοι ομαδοποίησης (Hierarchical Clustering Algorithms).....	38
3.5.2 Ομαδοποίηση μεθόδων υλοποίησης.....	41
ΚΕΦΑΛΑΙΟ 4 - ΣΕΝΑΡΙΑ – ΜΕΘΟΔΟΙ ΥΛΟΠΟΙΗΣΗΣ	43
4.1 Εισαγωγή.....	43
4.2 Περιγραφή Μεθόδων.....	44
4.3 Σύνολα Δεδομένων.....	45
4.3.1 Small Set of Good Data.....	45
4.3.2 Small Set of Random Data	45
4.3.3 Large Set of Good Data.....	45
4.4 Μέθοδος Αξιολόγησης (Evaluation Method).....	46
4.4.1 Precision	47
4.4.2 Recall.....	47
4.4.3 F-Measure.....	48
4.4.4 Purity	48
4.5 Περιγραφή Μεθόδων.....	50
4.5.1 Μέθοδος 1	50
4.5.2 Μέθοδος 2	55
4.5.3 Μέθοδος 3	64
4.5.4 Μέθοδος 4	69
4.5.5 Μέθοδος 5	77
4.6 Διασύνδεση Μεθόδων σε Graphical User Interface (GUI)	84
ΚΕΦΑΛΑΙΟ 5 - ΑΠΟΤΕΛΕΣΜΑΤΑ ΚΑΙ ΣΥΜΠΕΡΑΣΜΑΤΑ	85
5.1 Συγκριτικά Αποτελέσματα Μεθόδων.....	85
5.2 Παρατηρήσεις και συμπεράσματα	87
ΒΙΒΛΙΟΓΡΑΦΙΑ.....	88
ΣΥΝΤΟΜΟ ΒΙΟΓΡΑΦΙΚΟ ΣΗΜΕΙΩΜΑ	91

ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

Πίνακας 2-1: Average precision για ένα σύνολο δεδομένων του [3] των αλγορίθμων: XOR, Sum of Squared Differences (SSD), Euclidean Distance Mapping (EDM), και ενός αλγορίθμου των Scott and Longuet-Higgins (SLH).....	11
Πίνακας 2-2: Συγκριτικοί χρόνοι εκτέλεσης των αλγορίθμων: XOR, Sum of Squared Differences (SSD), ενός αλγορίθμου των Scott and Longuet-Higgins (SLH), Shape Context (SC), Euclidean Distance Mapping (EDM), και Dynamic Time Warping (DTW).	11
Πίνακας 3-1: Παράδειγμα λειτουργίας της γραμμικής παρεμβολής για UWP. 1 ^η γραμμή: Αριθμός θέσης στο διάνυσμα, 2 ^η γραμμή: Αρχικές τιμές του UWP, 3 ^η γραμμή: Υπολογισμός των τιμών του παρεμβλημένου UWP, 4 ^η γραμμή: Οι νέες τιμές του UWP.	26
Πίνακας 4-1: Παράδειγμα πίνακα αποστάσεων για σύνολο δεδομένων που αποτελείται από 5 εικόνες λέξεων.....	53
Πίνακας 4-2: Αποτελέσματα μεθόδου 1 για τα τρία σύνολα δεδομένων. Στον χρόνο περιλαμβάνεται ο χρόνος όλων των λειτουργιών που εκτελούνται στη μέθοδο. Οι χρόνοι υπολογίστηκαν κατά την εκτέλεση της μεθόδου στη Matlab σε υπολογιστή με 2-πύρηνο επεξεργαστή 2,27 GHz.	54
Πίνακας 4-3: Αποτελέσματα μεθόδου 2 για το Small Set of Good Data. Για κάθε ένα από τους 7 αλγόριθμους Linkage, δίνεται το ποσοστό των δεικτών αξιολόγησης: Precision, Recall, F-Measure και Purity, ενώ η μετρική απόσταση που χρησιμοποιήθηκε εδώ ήταν η απόσταση συνημίτονων (Cosine Distance).	61
Πίνακας 4-4: Αποτελέσματα μεθόδου 2 για το Small Set of Good Data. Για κάθε μια από τις 11 μετρικές αποστάσεων, δίνεται το ποσοστό των δεικτών αξιολόγησης: Precision, Recall, F-Measure και Purity, ενώ ο αλγόριθμος ομαδοποίησης που χρησιμοποιήθηκε εδώ ήταν ο Ward Linkage.	62
Πίνακας 4-5: Αποτελέσματα μεθόδου 2 για τα τρία σύνολα δεδομένων. Στον χρόνο περιλαμβάνεται ο χρόνος όλων των λειτουργιών που εκτελούνται στη μέθοδο. Οι χρόνοι υπολογίστηκαν κατά την εκτέλεση της μεθόδου στη Matlab σε υπολογιστή με 2-πύρηνο επεξεργαστή 2,27 GHz.	63
Πίνακας 4-6: Αποτελέσματα μεθόδου 3 για τα τρία σύνολα δεδομένων. Στον χρόνο περιλαμβάνεται ο χρόνος όλων των λειτουργιών που εκτελούνται στη μέθοδο. Οι χρόνοι υπολογίστηκαν κατά την εκτέλεση της μεθόδου στη Matlab σε υπολογιστή με 2-πύρηνο επεξεργαστή 2,27 GHz.	68
Πίνακας 4-7: Αποτελέσματα μεθόδου 4 για το Small Set of Good Data. Για κάθε ένα από τους 7 αλγόριθμους Linkage, δίνεται το ποσοστό των δεικτών αξιολόγησης: Precision, Recall, F-Measure και Purity, ενώ η μετρική απόσταση που χρησιμοποιήθηκε εδώ ήταν η ευκλείδεια απόσταση.....	74

Πίνακας 4-8: Αποτελέσματα μεθόδου 4 για το Small Set of Good Data. Για κάθε μια από τις 10 μετρικές αποστάσεων, δίνεται το ποσοστό των δεικτών αξιολόγησης: Precision, Recall, F-Measure και Purity, ενώ ο αλγόριθμος ομαδοποίησης που χρησιμοποιήθηκε εδώ ήταν ο Ward Linkage.	75
Πίνακας 4-9: Αποτελέσματα μεθόδου 4 για τα τρία σύνολα δεδομένων. Στον χρόνο περιλαμβάνεται ο χρόνος όλων των λειτουργιών που εκτελούνται στη μέθοδο. Οι χρόνοι υπολογίστηκαν κατά την εκτέλεση της μεθόδου στη Matlab σε υπολογιστή με 2-πύρηνο επεξεργαστή 2,27 GHz.	76
Πίνακας 4-10: Αποτελέσματα μεθόδου 5 για τα τρία σύνολα δεδομένων. Στον χρόνο περιλαμβάνεται ο χρόνος όλων των λειτουργιών που εκτελούνται στη μέθοδο. Οι χρόνοι υπολογίστηκαν κατά την εκτέλεση της μεθόδου στη Matlab σε υπολογιστή με 2-πύρηνο επεξεργαστή 2,27 GHz.	82
Πίνακας 5-1: Συγκριτικά αποτελέσματα Μεθόδων 1,2,3,4 και 5 για το Small Set of Good Data.	85
Πίνακας 5-2: Συγκριτικά αποτελέσματα Μεθόδων 1,2,3,4 και 5 για το Small Set of Random Data.	86
Πίνακας 5-3: Συγκριτικά αποτελέσματα Μεθόδων 1,2,3,4 και 5 για το Large Set of Good Data.	86

ΚΑΤΑΛΟΓΟΣ ΣΧΗΜΑΤΩΝ

Σχήμα 2-1: [1] Αποτυχία του OCR στην αναγνώριση αρχαίων ιστορικών χαρακτήρων γραμμένων στα ελληνικά. Το σχήμα (a) είναι η αρχική εικόνα προς αναγνώριση και το σχήμα (b) δείχνει τα αποτελέσματα που δίνει το OCR.	3
Σχήμα 2-2: Βασική ιδέα του Word Spotting με τα επιμέρους στάδια που το αποτελούν.	3
Σχήμα 2-3: [9] Τα αποτελέσματα της αναγνώρισης λέξεων αναπαριστώνται με πιθανοτικό μοντέλο.	9
Σχήμα 3-1: Παράδειγμα αφαίρεσης θορύβου της λέξης “ΒΑΡΕΤΟΣ”.....	15
Σχήμα 3-2: Παράδειγμα σύνδεσης σπασμένων κομματιών.	15
Σχήμα 3-3: Παράδειγμα γεμίματος γωνιάς.	15
Σχήμα 3-4: Παράδειγμα σύνδεσης σπασμένων κομματιών σε λέξη.	16
Σχήμα 3-5: Παράδειγμα φαινομένου skew στη λέξη “δημοκρατία”.....	16
Σχήμα 3-6: Παράδειγμα διόρθωσης της γωνιάς skew στη λέξη “δημοκρατία”.....	16
Σχήμα 3-7: Παράδειγμα φαινομένου slant στη λέξη “βρέχω”.....	17
Σχήμα 3-8: Τροποποίηση της γωνιάς (shear angle) στη λέξη “βρέχω” και κάθετα ιστογράμματα προβολής των τροποποιημένων εικόνων για γωνιές -90, 0 και 90 μοιρών.	17
Σχήμα 3-9: Παράδειγμα περιορισμού της λέξης “ΒΑΡΕΤΟΣ” στο πλαίσιο οριοθέτησης.....	18
Σχήμα 3-10: Παράδειγμα κανονικοποίησης του font στη λέξη “δορυφόρος”.....	18
Σχήμα 3-11: Παράδειγμα Upper-Lower-Left-Right Word Profile της λέξης “δημοκρατία”.....	21
Σχήμα 3-12: Παράδειγμα Lower Minus Upper Word Profile της λέξης “δημοκρατία”.....	21
Σχήμα 3-13: Παράδειγμα Vertical-Horizontal Projection Profile της λέξης “δημοκρατία”.....	22
Σχήμα 3-14: Παράδειγμα Vertical-Horizontal Background to Ink Transitions Profile της λέξης “δημοκρατία”.....	23
Σχήμα 3-15: Παράδειγμα υπολογισμού των End Points από την αρχική εικόνα. Τα τελικά σημεία (End Points) είναι ο συνολικός αριθμός των 1 που εμφανίζονται στην δεξιά εικόνα.	24
Σχήμα 3-16: Παράδειγμα των τελικών σημείων (End Points) της λέξης “ΜΑΝΙΑ”. Στην μεσαία εικόνα τα pixel που αναπαριστούν End Points σημειώνονται με πράσινο χρώμα ενώ η κάτω εικόνα δείχνει μόνο τα End Points (άσπρα pixel).	24
Σχήμα 3-17: Παράδειγμα των περιμετρικών σημείων (Perimeter Points) της λέξης “ΜΑΝΙΑ”. Στην μεσαία εικόνα τα pixel που αναπαριστούν Perimeter Points σημειώνονται με πράσινο χρώμα ενώ η κάτω εικόνα δείχνει μόνο τα Perimeter Points (άσπρα pixel).....	25
Σχήμα 3-18: Παράδειγμα στις τρύπες που ανιχνεύονται μέσα στη λέξη “ΜΑΝΙΑ”.....	25
Σχήμα 3-19: Παράδειγμα εφαρμογής της παρεμβολής σε UWP μιας εικόνας χειρόγραφης λέξης.	26

Σχήμα 3-20: Παράδειγμα κανονικοποίησης μεγέθους για κάθε εικόνα μεταβαλλόμενου διανύσματος 1XN (κάθετο προφίλ) σε σταθερό διάνυσμα μεγέθους 1X175. Ομοίως ισχύει για μεταβαλλόμενο διάνυσμα 1XM (οριζόντιο προφίλ).	28
Σχήμα 3-21: Παράδειγμα αντιστοίχισης δύο σημάτων με τον Dynamic Time Warping.	30
Σχήμα 3-22: Ο περιορισμός Sakoe-Chiba band.	31
Σχήμα 3-23: Παράδειγμα δια-ομαδικών και ενδο-ομαδικών αποστάσεων τριών ομάδων.	37
Σχήμα 3-24: [28] Παράδειγμα οπτικοποίησης 5 ομάδων ως ένα δενδρόγραμμα.	38
Σχήμα 3-25: Μέτρο ομοιότητας του ιεραρχικού αλγόριθμου Single Linkage.	39
Σχήμα 3-26: Μέτρο ομοιότητας του ιεραρχικού αλγόριθμου Complete Linkage.	39
Σχήμα 3-27: Μέτρο ομοιότητας του ιεραρχικού αλγόριθμου Average Linkage.	40
Σχήμα 3-28: Μέτρο ομοιότητας του ιεραρχικού αλγόριθμου Centroid Linkage.	40
Σχήμα 3-29: Παράδειγμα δενδρογράμματος με 12 διαφορετικές εικόνες, αρχίζοντας με 12 διαφορετικές ομάδες (φύλλα) καταλήγοντας στη ρίζα του δέντρου (με όλες τις εικόνες).	42
Σχήμα 4-1: Ο τρόπος με τον οποίο χωρίζονται τα έγγραφα στην ανάκτηση πληροφορίας ώστε να κατανοηθεί ευκολότερα η χρήση των μέτρων αξιολόγησης Precision και Recall.	46
Σχήμα 4-2: Παράδειγμα υπολογισμού του Precision για 2 ομάδες (Clusters).	47
Σχήμα 4-3: Παράδειγμα υπολογισμού του Recall για 2 ομάδες (Clusters).	48
Σχήμα 4-4: Παράδειγμα υπολογισμού του Purity για 3 ομάδες (Clusters).	49
Σχήμα 4-5: Διάγραμμα ροής μεθοδολογίας για την μέθοδο 1.	51
Σχήμα 4-6: Διάγραμμα ροής προ-επεξεργασίας μεθόδου 1.	52
Σχήμα 4-7: Διάγραμμα ροής μεθοδολογίας για την μέθοδο 2.	55
Σχήμα 4-8: Διάγραμμα ροής προ-επεξεργασίας μεθόδου 2.	56
Σχήμα 4-9: Παράδειγμα τεμαχισμού αρχικής εικόνας χειρόγραφης λέξης σε 9 μικρότερες υπό-εικόνες.	57
Σχήμα 4-10: Παράδειγμα δημιουργίας πινάκων χαρακτηριστικών.	59
Σχήμα 4-11: Παράδειγμα δημιουργίας πινάκων χαρακτηριστικών.	60
Σχήμα 4-12: Διάγραμμα ροής μεθοδολογίας για την μέθοδο 3.	64
Σχήμα 4-13: Διάγραμμα ροής προ-επεξεργασίας μεθόδου 3.	65
Σχήμα 4-14: Παράδειγμα μετασχηματισμού των τεσσάρων χαρακτηριστικών σε DFT της λέξης “δημοκρατία”.	66
Σχήμα 4-15: Σχηματική αναπαράσταση της όλης διαδικασίας της μεθόδου 3, στην οποία η επεξεργασμένη λέξη “δημοκρατία” αναπαριστάται ως ένα διάνυσμα σε σειρά με DFT συντελεστές.	67
Σχήμα 4-16: Παράδειγμα ενός πίνακα δεδομένων με 800 λέξεις της μεθόδου 3.	68
Σχήμα 4-17: Διάγραμμα ροής μεθοδολογίας για την μέθοδο 4.	69
Σχήμα 4-18: Διάγραμμα ροής προ-επεξεργασίας μεθόδου 4.	70

Σχήμα 4-19: Παράδειγμα μετασχηματισμού των έντεκα χαρακτηριστικών σε DFT της λέξης “δημοκρατία” .	72
Σχήμα 4-20: Σχηματική αναπαράσταση της όλης διαδικασίας της μεθόδου 4, στην οποία η επεξεργασμένη λέξη “δημοκρατία” αναπαριστάται ως ένα διάνυσμα σε σειρά με DFT συντελεστές.	73
Σχήμα 4-21: Παράδειγμα ενός πίνακα δεδομένων με 800 λέξεις της μεθόδου 4.	74
Σχήμα 4-22: Διάγραμμα ροής μεθοδολογίας για την μέθοδο 5.	77
Σχήμα 4-23: Διάγραμμα ροής προ-επεξεργασίας μεθόδου 5.	78
Σχήμα 4-24: Παράδειγμα μετασχηματισμού των έντεκα χαρακτηριστικών σε DCT της λέξης “δημοκρατία” .	80
Σχήμα 4-25: Σχηματική αναπαράσταση της όλης διαδικασίας της μεθόδου 5, στην οποία η επεξεργασμένη λέξη “δημοκρατία” αναπαριστάται ως ένα διάνυσμα σε σειρά με DCT συντελεστές.	81
Σχήμα 4-26: Παράδειγμα ενός πίνακα δεδομένων με 800 λέξεις της μεθόδου 5.	82
Σχήμα 4-27: Graphical User Interface σε εφαρμογή.	84

ΚΕΦΑΛΑΙΟ 1 - ΕΙΣΑΓΩΓΗ

Κάθε άτομο έχει το δικό του γραφικό χαρακτήρα, με αποτέλεσμα μερικές φορές να είναι δύσκολο ακόμα και για ένα άνθρωπο να διαβάσει και να επεξεργαστεί λέξεις κάποιου άλλου ατόμου. Στο παρελθόν, το πιο πάνω έκανε αδύνατο για έναν Ηλεκτρονικό Υπολογιστή να αναγνώσει και να επεξεργαστεί ένα χειρόγραφο κείμενο. Ωστόσο οι ανάγκες της σύγχρονης κοινωνίας, οδήγησαν πολλές εταιρείες και οργανισμούς στη δική τους ανάγκη να κάνουν διαθέσιμο ένα τεράστιο ποσό χειρόγραφων εγγράφων σε ηλεκτρονική μορφή, στο διαδίκτυο ή σε ψηφιακά μέσα. Ως αποτέλεσμα, η αναζήτηση λέξεων μέσα από χειρόγραφα έγγραφα και η επεξεργασία τους από Ηλεκτρονικό Υπολογιστή, εξακολουθεί μέχρι και σήμερα να είναι ένα από τα σοβαρότατα προβλήματα, μιας και πρέπει αρχικά να σκαναριστούν. Το σκανάρισμα δημιουργεί και άλλου είδους προβλήματα, όπως είναι η εισαγωγή επιπλέον θορύβου στην εικόνα, ενώ άλλες υποβαθμίσεις που παρουσιάζονται συνήθως από διάφορες αιτίες είναι το ξεθώριασμα στο μελάνι, διαφορές στο font, υποβαθμίσεις στις λέξεις, όπως σπασμένα γράμματα ή ελλιπείς γράμματα μέσα στη λέξη. Λόγω του μεγάλου όγκου της μεταβλητότητας στην γραφή, και τα υψηλά επίπεδα θορύβου στα χειρόγραφα έγγραφα, η επεξεργασία για την αναγνώρισή των λέξεων που βρίσκονται μέσα σε αυτά είναι μία εξαιρετικά δύσκολη διαδικασία.

Χρειάζεται λοιπόν, ένας μηχανισμός ο οποίος θα χειρίζεται τη διαδικασία αυτή, με σκοπό να γίνεται πιο εύκολος ο χειρισμός, η ανάγνωση και η επεξεργασία των χειρόγραφων εγγράφων μέσα από Ηλεκτρονικό Υπολογιστή. Η παραδοσιακή τεχνική του OCR, αναγνωρίζει τις λέξεις χαρακτήρα με χαρακτήρα, όμως η εφαρμογή της στην αναγνώριση χειρόγραφων λέξεων αποτυγχάνει σε πολύ μεγάλο βαθμό. Αναζητάμε λοιπόν, μία νέα μέθοδο η οποία θα μπορεί να εφαρμοστεί σε χειρόγραφες λέξεις, όπου πλέον δεν θα μελετώνται οι χαρακτήρες της λέξης, αλλά θα μελετάται ολόκληρη η λέξη.

Το Word Spotting είναι η τεχνική, με την οποία μπορούμε να επιτύχουμε αναγνώριση χειρόγραφων λέξεων μέσα από διάφορα χειρόγραφα έγγραφα εικόνων, και με την οποία ασχολείται η παρούσα εργασία. Για να εκπληρωθεί ο στόχος της συγκεκριμένης εργασίας, χρησιμοποιήθηκαν δύο υπάρχουσες μέθοδοι ως πρότυπο για τη δημιουργία, υλοποίηση και αξιολόγηση τριών νέων μεθόδων, που δημιουργήθηκαν για τις ανάγκες της εργασίας. Σκοπός των μεθόδων είναι η ομαδοποίηση (Clustering) εικόνων χειρόγραφων λέξεων γραμμένες από διαφορετικά άτομα. Κυρίως στόχος τους, είναι να τοποθετήσουν σωστά τη λέξη στη σωστή ομάδα. Αξιοσημείωτο είναι το γεγονός ότι οι νέες μέθοδοι που χρησιμοποιήθηκαν έφεραν καλύτερα αποτελέσματα από τις δύο ήδη υπάρχουσες. Συγκεκριμένα, έγινε μια προσπάθεια υλοποίησης δύο ήδη υπάρχον μεθόδων, οι οποίες είχαν ως σκοπό την εφαρμογή της τεχνικής του Word Spotting σε εικόνες χειρόγραφων κειμένων ενός (ή μιας μικρής ομάδας) συγγραφέων. Οι δικές μας προσεγγίσεις των υλοποιήσεων αυτών, δηλαδή οι 3 νέες μέθοδοι που υλοποιήθηκαν στα πλαίσια της εργασίας, είχαν ως σκοπό να εφαρμόσουν κάποια στάδια του Word Spotting, με στόχο την επιτυχή αναγνώριση εικόνων χειρόγραφων λέξεων διάφορων συγγραφέων.

ΚΕΦΑΛΑΙΟ 2 - ΓΕΝΙΚΑ ΓΙΑ ΤΟ WORD SPOTTING

2.1 Ορισμός του Word Spotting

Το Word Spotting, είναι μία τεχνική-μεθοδολογία που χρησιμοποιείται για την αναγνώριση χειρόγραφων λέξεων, μέσα από ένα σύνολο χειρόγραφων εγγράφων. Στόχος της τεχνικής του Word Spotting, είναι να εφαρμοστεί σε χειρόγραφα έγγραφα, με σκοπό να αναγνωρίσει πρότυπα λέξεων, τα οποία με το μάτι ίσως να μην είναι εφικτό να αναγνωριστούν. Η τεχνική καλείται να αντιμετωπίσει βασικά προβλήματα που αντιμετωπίζουμε σήμερα. Συνηθισμένη προσέγγιση του Word Spotting, είναι η είσοδος μιας εικόνας λέξης query, και η επιστροφή από το σύστημα όλων των παρόμοιων λέξεων του συνόλου δεδομένων. Εναλλακτική προσέγγιση, είναι η είσοδος ενός συνόλου δεδομένων, και η αυτόματη ταξινόμηση των λέξεων του συνόλου από το σύστημα σε διάφορες ομάδες παρόμοιων λέξεων.

2.2 Προβλήματα που καλείται να αντιμετωπίσει η τεχνική Word Spotting

Η τεχνική του Word Spotting καλείται κυρίως να αντιμετωπίσει τα παρακάτω προβλήματα:

- Πρόβλημα αναγνώρισης λέξεων διαφορετικών ατόμων, μιας και κάθε άνθρωπος έχει μοναδικό γραφικό χαρακτήρα.
- Μεγάλος όγκος δεδομένων από χειρόγραφα έγγραφα χρειάζεται να γίνει διαθέσιμος σε ηλεκτρονική μορφή, στο διαδίκτυο ή σε ψηφιακά μέσα, σε σύντομο χρονικό διάστημα.
- Πρόβλημα αναζήτησης λέξεων μέσα από πλήθος χειρόγραφων εγγράφων, και η επεξεργασία τους από Ηλεκτρονικό Υπολογιστή.
- Προβλήματα όπως θόρυβος, ξεθώριασμα στο μελάνι, διαφορές στο font και άλλου είδους υποβαθμίσεις.

2.3 Αποτυχία του OCR στην αναγνώριση χειρόγραφων λέξεων

Οι παραδοσιακοί μηχανισμοί OCR χρησιμοποιούνται ευρέως σε ποικίλες εφαρμογές, και μπορούμε να πούμε ότι έχουν επιτυχία σε σύγχρονα έγγραφα και σε έγγραφα υψηλής ποιότητας εκτύπωσης. Στις περιπτώσεις όμως που έχουμε χειρόγραφα έγγραφα ή ιστορικά αρχαία έγγραφα, όπως για παράδειγμα έγγραφα που περιλαμβάνουν λέξεις στα αρχαία ελληνικά, υπάρχουν πολλοί παράγοντες που επηρεάζουν την τελική απόδοση, όπως η χαμηλή ποιότητα του χαρτιού, διάφορες παραλλαγές στην τοποθέτηση του χαρτιού, όπως η κλίση του, ατέλειες στοιχειοθεσίας κτλ. Αφού λοιπόν το OCR κάνει τμηματοποίηση χαρακτήρων με σκοπό την αναγνώριση αυτών των χαρακτήρων, αποτυγχάνει σε μεγάλο βαθμό να τμηματοποιήσει χειρόγραφα ή ιστορικά έγγραφα σε μεμονωμένους χαρακτήρες. Λόγω ατελειών στοιχειοθέτησης καθώς και της χαμηλής ποιότητας χειρόγραφων ή ιστορικών εγγράφων, δεν επιτράπηκε σε κανένα OCR να κάνει επιτυχημένη αναγνώριση.[1]

Στο σχήμα 2-1 βλέπουμε τα αποτελέσματα ενός OCR στην προσπάθεια του να αναγνωρίσει μία εικόνα με ιστορικούς χαρακτήρες γραμμένους στα αρχαία ελληνικά.

σαῖδρα τύπον, τὴν ὕδατος ποιότητα ἔχε· τὸ δὲ τῷ κύβῳ σχῆμα ἀπεδίδα
 τῇ γῆ, σερωτάτη ὑπαρχέση, ἐ δραιοτάτη· τῷ δὲ σχήματι τῷ δωδεκαέδρῳ
 πρὸς τὸ πᾶν κατεχρήσατο. Πάντων δὲ τούτων ἀρχικωτέρα φύσις ἢ τῶν ἐπιπέδων·
 τῶν γὰρ σερεῶν προηγείται τὰ ἐπίπεδα· τῆς δὲ ἐπιπέδα φύσεως ὡσπέρινα

(a)

τοιβτιητι» ἐ'λί· τβ ίε' τ» χι!^8 <Γ%?μ* «Τίδίιιι *9νί· «{«ωτάτῷ
 υίτ«(χιίί·ί|),\$ «'ακιτάτΗ · τψ }ε «·χκ'μ*τ< τ9 ίιοίεχίκ'ί^ Τ{β< τβ τί»
 χατεχ^νατο. Πάντων ίέ τ«των β(χι)<οτε{α φύικ Η των |«·ικ'?'ν· των γ*}
 ς>(ί«ν »(ίκγ«7τ<ιι τα £*«·(»·«\$«· τίξ ίκ ίιηίβ φίβ'λαι αρηστα*

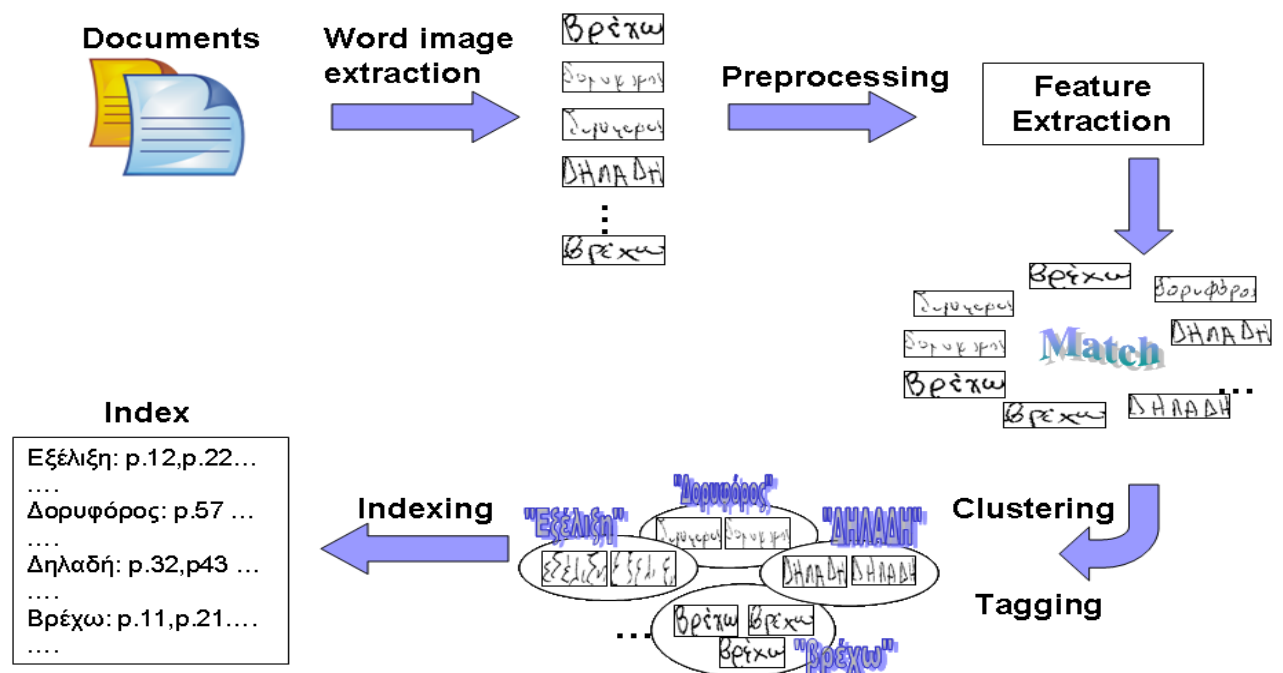
(b)

Σχήμα 2-1: [1] Αποτυχία του OCR στην αναγνώριση αρχαίων ιστορικών χαρακτήρων γραμμένων στα ελληνικά. Το σχήμα (a) είναι η αρχική εικόνα προς αναγνώριση και το σχήμα (b) δείχνει τα αποτελέσματα που δίνει το OCR.

Από τα αποτελέσματα του σχήματος 2-1 (b), βλέπουμε το OCR να αποτυγχάνει σε πολύ μεγάλο βαθμό στην προσπάθεια του να αναγνωρίσει την εικόνα γραμμένη στα αρχαία ελληνικά. Οι ατέλειες στην στοιχειοθέτηση της αρχικής εικόνας, η πολυτονική γραφή (περισπωμένη, δασεία κτλ.), η γραμματοσειρά, οι ατέλειες στο μέγεθος, καθώς και η χαμηλή ποιότητα της εικόνας, δεν επέτρεψαν στο OCR να αναγνωρίσει το ιστορικό έγγραφο.

2.4 Βασική ιδέα του Word Spotting

Στο σχήμα 2-2 δείχνουμε σχηματικά τη βασική ιδέα του Word Spotting και τα επιμέρους στάδια που το αποτελούν:



Σχήμα 2-2: Βασική ιδέα του Word Spotting με τα επιμέρους στάδια που το αποτελούν.

2.5 Στάδια του Word Spotting

Το Word Spotting αποτελείται από 8 στάδια, τα οποία φαίνονται και στο σχήμα 2-2 στην σχηματική εικόνα με τα επιμέρους στάδια που το αποτελούν.

Παρακάτω αναφέρουμε τα 8 στάδια του Word Spotting:

1. Αρχικά λαμβάνεται μία σκαναρισμένη grey-level εικόνα εγγράφου.
2. Στάδιο τμηματοποίησης (Segmentation): Η σελίδα χωρίζεται σε υπό-μονάδες (λέξεις), με σκοπό να εξαχθεί κάθε λέξη από το έγγραφο ξεχωριστά. Το στάδιο αυτό περιλαμβάνει την τμηματοποίηση γραμμής (Line Segmentation), καθώς και την τμηματοποίηση λέξης (Word Segmentation). Μία βασική τεχνική που εφαρμόζεται στο στάδιο αυτό είναι η Scale-Space η οποία μπορεί να βρεθεί στο [2]. Το στάδιο αυτό, σε ορισμένες περιπτώσεις μπορεί να παραληφθεί από τη μέθοδο του Word Spotting.
3. Στάδιο εξαγωγής των εικόνων λέξεων (Word Image Extraction): Κάθε λέξη που έχει τμηματοποιηθεί στο έγγραφο, εξάγεται μέσα σε ένα πλαίσιο οριοθέτησης.
4. Στάδιο προ-επεξεργασίας (Preprocessing): Στο στάδιο αυτό, σε κάθε λέξη που εξάγεται από το έγγραφο, εφαρμόζεται προ-επεξεργασία, με σκοπό να διορθώσει διάφορα προβλήματα, όπως αφαίρεση του θορύβου, διόρθωση της κλίσης, διόρθωση του font, εισαγωγή καλύτερου πλαισίου οριοθέτησης κτλ.
5. Στάδιο εξαγωγής χαρακτηριστικών (Feature Extraction): Στο στάδιο αυτό, για κάθε εικόνα λέξης που είναι πλέον επεξεργασμένη, εξάγονται κάποια χαρακτηριστικά της λέξης, τα οποία είναι απαραίτητα σε επόμενα στάδια.
6. Στάδιο ταιριάσματος (Matching): Στο στάδιο αυτό, υπολογίζονται οι ανά ζεύγος αποστάσεις μεταξύ όλων των εικόνων που υπάρχουν στο σύνολο, με βάση τα χαρακτηριστικά που έχουν εξαχθεί από κάθε εικόνα. Για το σκοπό αυτό χρησιμοποιούνται διάφοροι αλγόριθμοι ταιριάσματος. Ο υπολογισμός των αποστάσεων όμως για όλες τις λέξεις σε μεγάλα σύνολα δεδομένων είναι πολύ δαπανηρός, οπότε σε τέτοιες περιπτώσεις χρησιμοποιούνται τεχνικές κλαδέματος (pruning), για να μειώσουν το κόστος, όπως γίνεται για παράδειγμα στο [3]. Να σημειώσουμε ότι η χρήση του σταδίου αυτού είναι προαιρετική, μιας και το Word Spotting μπορεί να περάσει κατευθείαν στο στάδιο της ομαδοποίησης.
7. Στάδιο ομαδοποίησης (Clustering): Στο στάδιο αυτό, όλες οι εικόνες ομαδοποιούνται σε ομάδες-κλάσεις με βάση την ομοιότητα τους, κάνοντας χρήση κάποιου αλγόριθμου ομαδοποίησης.
8. Στάδιο ευρετηρίασης (Indexing): Στο στάδιο αυτό, σημειώνεται σε κάθε κλάση το ισοδύναμο της σε ASCII χαρακτήρες, και φτιάχνεται ένα ευρετήριο από όπου θα ανακτώνται οι λέξεις.

2.6 Στάδια του Word Spotting που μελετώνται

Τα στάδια με τα οποία ασχολείται η παρούσα εργασία, είναι το στάδιο της προ-επεξεργασίας, το στάδιο της εξαγωγής χαρακτηριστικών, το στάδιο του ταιριάσματος (Matching) και το στάδιο της ομαδοποίησης (Clustering), που είναι και ο τελικός στόχος. Θεωρούμε ότι τα αρχικά στάδια έχουν πραγματοποιηθεί, δηλαδή έχουν εισαχθεί και σκαναριστεί τα χειρόγραφα έγγραφα, έχει γίνει τμηματοποίηση λέξεων για κάθε ένα χειρόγραφο έγγραφο, και έχουν εξαχθεί όλες οι λέξεις από όλα τα έγγραφα. Έχοντας δηλαδή στη διάθεση μας όλες τις εικόνες λέξεων που έχουν εξαχθεί, πραγματοποιούνται όλα τα στάδια μέχρι το στάδιο της ομαδοποίησης των εικόνων.

2.7 Εργασίες που έγιναν στο Word Spotting

Παρακάτω θα μιλήσουμε για διάφορες εργασίες που έχουν πραγματοποιηθεί στο Word Spotting, καθώς και εργασίες που κάνουν χρήση του Word Spotting ή παραλλαγές του.

2.7.1 Εργασία με θέμα « Word Spotting for Historical Documents »

Το άρθρο [4], είναι ένα από τα πιο δημοφιλή στο χώρο του Word Spotting, των T.M.Rath και R.Manmatha. Στο άρθρο αυτό, γίνεται μία σειρά από συγκρίσεις αλγορίθμων ομαδοποίησης (Clustering) και αλγορίθμων ταιριάσματος (Matching), με σκοπό να αναδειχτεί ο καλύτερος αλγόριθμος, μέσα από μία συλλογή χειρόγραφων λέξεων, σκαναρισμένων εγγράφων της συλλογής του George Washington. Έχουν πραγματοποιηθεί δύο διαφορετικά πειράματα, στα οποία αποδείχτηκε στο πρώτο πείραμα καλύτερος αλγόριθμος ομαδοποίησης να είναι ο Average Linkage, με καλύτερο Word Error Rate (WER), ενώ στο δεύτερο πείραμα αναδείχτηκε ο Ward Linkage. Τα συγκεκριμένα πειράματα που εφαρμόστηκαν ακολουθούν τις δύο πιο κάτω μεθοδολογίες:

- a) Το πρώτο πείραμα αποτελείται από ένα σύνολο 4860 χειρόγραφων λέξεων, το οποίο περιλαμβάνει ένα τετραγωνικό πίνακα μεγέθους 4860 X 4860, ο οποίος περιέχει τις αποστάσεις ανά ζεύγος μεταξύ των εικόνων των λέξεων, που υπολογίζονται κάνοντας χρήση του Dynamic Time Warping (DTW) Matching αλγόριθμου. Με βάση τον πίνακα αυτό εκτελείται η ομαδοποίηση των λέξεων. Λόγω του ότι ο DTW υπολογίζει ανά ζεύγος αποστάσεις, ο υπολογισμός του πίνακα των αποστάσεων απαιτεί περίπου μια εβδομάδα σε μηχανή 4^{ov} CPUs 500MHz. [4]
- b) Το δεύτερο πείραμα αποτελείται και αυτό από ένα σύνολο 4860 χειρόγραφων λέξεων, στο οποίο κάθε λέξη αναπαριστάται αποκλειστικά από Discrete Fourier Transform (DFT) συντελεστές, οι οποίοι εξάγονται από το Upper Word Profile, το Lower Word Profile, Vertical Projection Profile και Vertical Background to Ink Transitions Profile. Στο πείραμα χρειάστηκαν συγκεκριμένα 33 συντελεστές DFT για την αναπαράσταση κάθε λέξης, όπου ο αριθμός 33 είναι ο βέλτιστος αριθμός που καθορίστηκε για αυτό το σύνολο δεδομένων. [4]

Μέσα από ένα πλήθος αλγορίθμων ομαδοποίησης, όπως ο K-Means, ο Single Linkage, ο Complete Linkage, ο Average Linkage, ο Weighted Linkage και ο Ward Linkage, αναδείχτηκαν ως καλύτεροι αλγόριθμοι ομαδοποίησης, με βάση το χαμηλότερο Word Error Rate (WER), για το πρώτο πείραμα ο Average Linkage ενώ για το δεύτερο πείραμα ο Ward Linkage.

Με αφορμή τα δύο αυτά πειράματα, και τα αποτελέσματα που έχουν βρεθεί, υλοποιήσαμε τις δύο ήδη υπάρχουσες μεθόδους (1 και 3) που θα δούμε αργότερα στο κεφάλαιο 4, για την ομαδοποίηση εικόνων χειρόγραφων λέξεων.

Να σημειωθεί επίσης ότι τα τέσσερα χαρακτηριστικά που εξάγονται, κανονικοποιούνται στο διάστημα [0,1], ενώ κατά τον DTW χρησιμοποιείται ο περιορισμός Sakoe-Chiba με μέτρο ομοιότητας την Ευκλείδεια απόσταση. Επίσης πριν γίνει η ομαδοποίηση, αρχικά εκτιμάται ο αριθμός των ομάδων που πρόκειται να δημιουργηθούν, με βάση τον εμπειρικό κανόνα Hear's Law.

2.7.2 Εργασία με θέμα « Features for Word Spotting in Historical Manuscripts »

Στο άρθρο [5], γίνεται μία σύγκριση διάφορων συνδυασμών χαρακτηριστικών (features), που εξάγονται από κάθε χειρόγραφο λέξη, με σκοπό να βρεθεί ο καλύτερος συνδυασμός για το στάδιο του Matching, χρησιμοποιώντας τον DTW αλγόριθμο, ο οποίος συγκρίνει σύνολα χαρακτηριστικών που εξάγονται από δύο εικόνες (δηλαδή δύο feature vectors). Για κάθε συνδυασμό χαρακτηριστικών, έχει μετρηθεί η απόδοση του μετά από μία σειρά δοκιμών. Έχει βρεθεί μία μέση ακρίβεια (Average Precision) γύρω στο 72% κάνοντας συνδυασμό χαρακτηριστικών, ο οποίος υπερτερεί από άλλες τεχνικές, τόσο σε ταχύτητα όσο και σε ακρίβεια. Λόγω του υψηλού θορύβου που παρουσιάζουν τα χειρόγραφα έγγραφα, έχει τονιστεί η ανάγκη της σωστής επιλογής των χαρακτηριστικών και του σωστού συνδυασμού. Λόγω σημαντικής υποβάθμισης στα χειρόγραφα έγγραφα, όπως ξεθωριασμένο μελάνι, λεκέδες κτλ. οι παραδοσιακοί χειρόγραφοι αναγνωριστές (handwritten recognizers) βασισμένοι στο OCR δεν αποδίδουν καθόλου καλά σε αυτό το έργο. Η ιδέα του Word Spotting που προτάθηκε, είναι η εναλλακτική λύση στο πρόβλημα, το οποίο εφαρμόζεται με επιτυχία κυρίως για συλλογές λέξεων συγκεκριμένου ατόμου. Η προσέγγιση περιγράφεται και από το σχήμα 2.2, με πρώτο στάδιο την τμηματοποίηση λέξεων από διάφορα έγγραφα, και συνεχίζει με την αναπαράσταση των λέξεων ως εικόνες, την εξαγωγή χαρακτηριστικών από κάθε εικόνα, το στάδιο του Matching, το στάδιο της ομαδοποίησης των λέξεων και τη δημιουργία του ευρετηρίου, κάνοντας χρήση δεικτών από τη συλλογή. Όπως αναφέρθηκε, στο [5] γίνεται μία έρευνα σε χαρακτηριστικά (features) που μπορούν να χρησιμοποιηθούν για ένα επιτυχή ταίριασμα (Matching) με βάση τον DTW. Για να υπάρχει συμβατικότητα μεταξύ τιμών στα χαρακτηριστικά διάφορων λέξεων, γίνεται η κανονικοποίηση των χαρακτηριστικών στο διάστημα μεταξύ [0,1], και τα χαρακτηριστικά εισάγονται σε ένα διάνυσμα σε σειρά. Η αρχική εικόνα κάθε λέξης εξομαλύνεται (smoothed) με ένα ιστροπικό Gaussian φίλτρο (Isotropic Gaussian Kernel) και μετατρέπεται το ύψος σε μια σταθερή τιμή, π.χ. σε ύψος 15 γραμμών. Τα χαρακτηριστικά έχουν αξιολογηθεί στη συλλογή του George Washington. Για να μειωθεί ένας μεγάλος αριθμός συγκρίσεων έχει εφαρμοστεί η τεχνική του pruning. Μετά από πειράματα, οι συγγραφείς έδειξαν ότι τα χαρακτηριστικά που αποδίδουν καλύτερα είναι τα: Vertical Projection Profile, Upper Word Profile, Lower Word Profile και Vertical Background to Ink Transitions Profile, τα οποία θα δούμε και στην ενότητα 3.2 της παρούσας εργασίας, με ποσοστό επιτυχίας του DTW γύρω στο 72.56% καθώς επίσης αποδείχτηκε και πιο γρήγορος στο χρόνο απόκρισης σε σχέση με το δεύτερο καλύτερο, τον EDM (Euclidean Distance Mapping), με ποσοστό 67.67%.

2.7.3 Εργασία με θέμα « Word Spotting: Indexing Handwritten Archives »

Στο άρθρο [6], η προσέγγιση που χρησιμοποιείται είναι η εξής: Αρχικά λαμβάνεται μία σκαναρισμένη χειρόγραφο grey level εικόνα εγγράφου. Η εικόνα μειώνεται στο μισό με Gaussian φιλτράρισμα και δειγματοληψία. Στη συνέχεια η εικόνα γίνεται δυαδική (binarized) δηλ. με pixels 0 και 1 με χρήση ενός κατωφλίου, δηλ. ποιά θα είναι μαύρα και ποιά άσπρα pixels. Η δυαδική εικόνα στη συνέχεια τμηματοποιείται σε εικόνες λέξεων. Αυτό γίνεται με διεργασία εξομάλυνσης (smoothing) και κατωφλίου (thresholding). Στη συνέχεια η εικόνα, χρησιμοποιείται ως πρότυπο (template) και συγκρίνεται έναντι όλων των υπόλοιπων. Αυτό, επαναλαμβάνεται για κάθε εικόνα του εγγράφου. Το στάδιο του ταιριάσματος (Matching) εκτελείται σε δύο φάσεις. Αρχικά ο αριθμός των εικόνων που πρόκειται να συγκριθούν μειώνεται με τεχνική pruning (τεχνική κλαδέματος), χρησιμοποιώντας τις περιοχές (areas) και τις αναλογίες εικόνας (aspect ratio) των εικόνων (δηλαδή η εικόνα η οποία πρόκειται να συγκριθεί δεν μπορεί να έχει περιοχή ή αναλογία εικόνας η οποία να είναι πολύ διαφορετική με το πρότυπο). Στη συνέχεια επιτυγχάνεται το ταίριασμα με βάση κάποιο Matching αλγόριθμο. Το στάδιο του ταιριάσματος διαιρεί τις εικόνες σε κλάσεις (ομάδες) ισοδυναμίας, όπου κάθε κλάση περιέχει άλλες εικόνες της ίδιας

λέξης. Εν τέλει γίνεται η δεικτοδότηση (Indexing). Για κάθε κλάση ισοδυναμίας, μετρίεται ο αριθμός των εικόνων που υπάρχουν σε αυτή την κλάση. Καθορίζονται οι top n κλάσεις ισοδυναμίας με βάση τον αριθμό των μελών κάθε κλάσης. Δηλαδή αν οι κλάσεις έχουν μεγάλο πλήθος εικόνων, το πιο πιθανόν να είναι λέξεις που χρησιμοποιούνται συχνά π.χ. η λέξη «και», οι οποίες απορρίπτονται, είναι δηλαδή απαγορευτικές λέξεις (stop-words). Από τις κλάσεις που παραμένουν, ο χρήστης εμφανίζει ένα μόνο μέλος από κάθε κλάση και το αντιστοιχεί στο ισοδύναμο του ASCII. Οι λέξεις αυτές μπορούν πλέον να δεικτοδοτηθούν οπουδήποτε εμφανίζονται μέσα στο έγγραφο.

Με βάση το άρθρο [6], η συνάρτηση ταιριάσματος πρέπει να τηρεί δύο βασικά κριτήρια. Πρώτο, θα πρέπει να παράγει χαμηλό σφάλμα ταιριάσματος (low match error) για τις λέξεις που είναι παρόμοιες με το πρότυπο. Και δεύτερο, θα πρέπει να παράγει μεγάλο σφάλμα ταιριάσματος (high match error) για τις λέξεις που είναι διαφορετικές με το πρότυπο.

Στα πειράματα χρησιμοποιήθηκαν οι αλγόριθμοι ταιριάσματος EDM και SLH (Scot and Longuet-Higgins [7]), όπου αναδείχτηκε καλύτερος ο SLH.

2.7.4 Εργασία με θέμα « Nearest Neighbor based Collection OCR »

Στην εργασία που περιγράφεται στο [8], παρουσιάζεται ένα OCR συλλογής (Collection OCR), στο οποίο πολλά παραδείγματα ίδιας λέξης (συνήθως σε ίδιο font), επέρχονται σε ένα έγγραφο ή σε μία συλλογή εγγράφων. Ως πρώτο στάδιο της εργασίας στο [8], είναι η χρήση ενός OCR, στο οποίο δημιουργείται ένα μερικό σύνολο αναγνωρισμένων λέξεων, το οποίο εισάγεται στο σύνολο εκπαίδευσης (training set). Το σύστημα στη συνέχεια τμηματοποιεί από διάφορες σελίδες εγγράφων τις λέξεις, και υπολογίζει από κάθε λέξη τα χαρακτηριστικά βασισμένα σε προφίλ (profile features). Ως δεύτερο στάδιο, ένας αλγόριθμος πλησιέστερου γείτονα (Nearest Neighbor Algorithm) συγκρίνει τις υπόλοιπες εικόνες λέξεων που βρίσκονται στο σύνολο δοκιμής (test set) με αυτές που έχουν ήδη αναγνωριστεί και δίνονται κάποιες ετικέτες (labels) από τους πλησιέστερους γείτονες. Αποδεικνύεται ότι με τη χρήση ενός κατά προσέγγιση γρήγορου πλησιέστερου γείτονα βασισμένο σε ιεραρχικό K-Means αλγόριθμο (Hierarchical K-Means), μπορεί να γίνει αυτό με μεγαλύτερη ακρίβεια και αποτελεσματικότητα. Αποδεικνύεται επίσης, ότι τα χαρακτηριστικά βασισμένα σε προφίλ (profile features), λειτουργούν καλύτερα από τα σχηματικά χαρακτηριστικά, δηλαδή τα SIFT and Pyramid Histogram of Gradient (PHOG). Το παραπάνω αναδεικνύει τα χαρακτηριστικά βασισμένα σε προφίλ, ως πιο ανθεκτικά ακόμα και σε υποβαθμισμένες λέξεις, σε σχέση με τα σχηματικά χαρακτηριστικά. Η προσέγγιση αυτή εφαρμόζεται σε μία συλλογή του Telugu books. Αποδεικνύεται ότι από μία επιλογή από 33 Telugu books, κάνοντας OCR στο 30% της συλλογής με αντίστοιχα labels, μπορεί να αναγνωριστεί το υπόλοιπο 70% των λέξεων στη συλλογή, με 70% ακρίβεια, με βάση αυτή την προσέγγιση. Η προσέγγιση αυτή δεν υποθέτει συγκεκριμένη γλώσσα, κάτι το οποίο την κάνει ικανή να μπορεί να εφαρμοστεί σε μεγάλο πλήθος γλωσσών. Στο συγκεκριμένο άρθρο, δηλαδή στο [8], χρησιμοποιείται σε σύνολο δεδομένων της ινδικής γλώσσας.

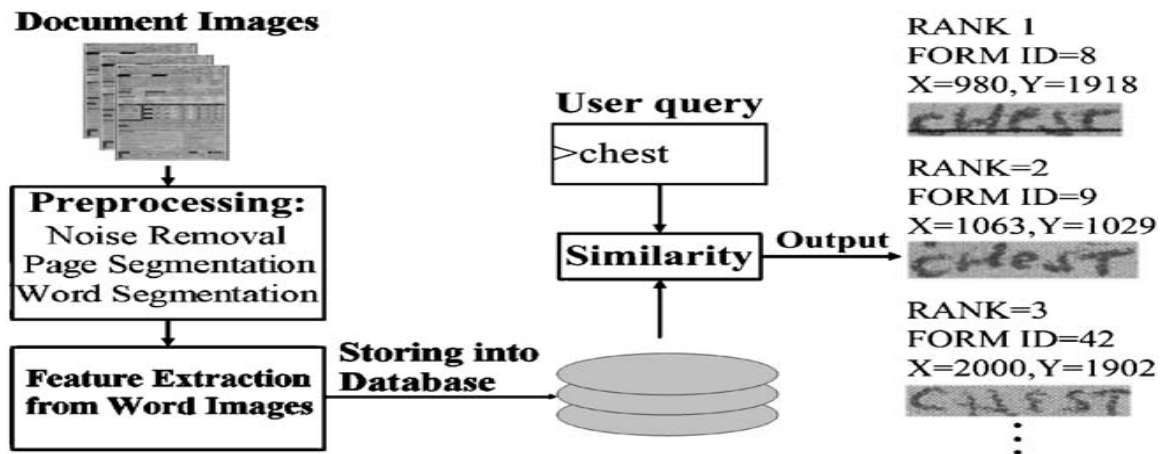
2.7.5 Εργασία με θέμα « A probabilistic method for keyword retrieval in handwritten document images »

Στην εργασία [9], περιγράφεται μια βελτίωση υπάρχον μεθόδων ανάκτησης στο Word Spotting, μοντελοποιώντας την ατελή τμηματοποίηση των λέξεων ως πιθανότητες και ενσωματώνοντας αυτές τις πιθανότητες στον αλγόριθμο του Word Spotting. Οι βαθμολογίες που επιστρέφει ο αναγνωριστής λέξεων (word recognizer), μετατρέπονται και αυτές επίσης σε πιθανότητες και ενσωματώνονται στο πιθανοτικό μοντέλο του Word Spotting.

Υπάρχουν δύο κοινές προσεγγίσεις για την ανάκτηση λέξεων από χειρόγραφα έγγραφα. Η μία προσέγγιση είναι η χρησιμοποίηση της ομοιότητας μεταξύ λέξεων (image to image Matching), ενώ η άλλη προσέγγιση είναι βασισμένη στην ομοιότητα bit με bit (bitwise Matching). Στην πρώτη προσέγγιση κατά την ανάκτηση, κάθε λέξη μετατρέπεται σε μία εικόνα με σκοπό να επιτευχθεί η ομοιότητα. Αυτό φυσικά γίνεται δημιουργώντας ένα μικρό σύνολο από εικόνες που σχετίζονται με τον συγκεκριμένο χρήστη. Όταν ο χρήστης εισάγει στο query μία λέξη, υπολογίζεται η ομοιότητα μεταξύ του query και κάθε εικόνας που υπάρχει στη βάση δεδομένων. Όλες οι λέξεις που επιστρέφονται, είναι σε φθίνουσα σειρά, με βάση την ομοιότητα όλων των εικόνων της βάσης δεδομένων με το query. Η ομοιότητα μεταξύ δύο εικόνων μετράται ως η απόσταση μεταξύ των δύο διανυσμάτων χαρακτηριστικών (feature vectors) που εξάγονται από κάθε εικόνα. Για τον υπολογισμό της ομοιότητας, μεταξύ των δύο διανυσμάτων χαρακτηριστικών των δύο εικόνων, εφαρμόζεται ο DTW ταιριάζοντας τα χαρακτηριστικά βασισμένα στα προφίλ (profile features). Στην δεύτερη προσέγγιση, υπολογίζεται η ομοιότητα bit με bit μεταξύ δύο διανυσμάτων χαρακτηριστικών δύο εικόνων για GSC (Gradient Structural Concavity [10],[11]) χαρακτηριστικά.

Η τεχνική του Word Spotting, απαιτεί και On-line ταίριασμα μεταξύ λέξεων, κάτι το οποίο είναι πολύ χρονοβόρο. Η ακρίβεια και η ταχύτητα παίζουν μεγάλο παράγοντα για το ταίριασμα σε μεγάλες βάσεις δεδομένων. Οι προσεγγίσεις όμως που υπάρχουν για γρήγορο ταίριασμα είναι περιορισμένες στην επιλογή των χαρακτηριστικών και στην πολυπλοκότητα του ταιριάσματος καθώς και των μεθόδων εκπαίδευσης. Το γεγονός αυτό περιορίζει το πεδίο εφαρμογής τους σε εφαρμογές που ασχολούνται με ένα χρήστη, δηλαδή εκπαίδευση εικόνων βάση ενός χρήστη. Σε αντίθεση τα αποτελέσματα ευρετηρίασης του OCR, δεν αντιμετωπίζουν το πρόβλημα της ταχύτητας. Σε αυτές τις μεθόδους, οι δείκτες χτίζονται με βάση τα αποτελέσματα του OCR, ή με διάνυσμα χαρακτηριστικών πυκνότητας πιθανότητας που προκύπτει από τις αποστάσεις που επιστρέφονται από τον αναγνωριστή λέξεων. Αφού εκτελεστεί η αναγνώριση λέξεων, ακολουθείται το στάδιο της ευρετηρίασης, με σκοπό να παρακολουθήσει κάποιες βασικές πληροφορίες, όπως θέσεις και αποτελέσματα αναγνώρισης (recognition scores), που συνήθως είναι πυκνότητες πιθανότητας, πιθανότητες ή κάποια άλλα μέτρα απόστασης.

Στην εργασία [9], έχουν βελτιώσει τα αποτελέσματα, με βάση τη μέθοδο ευρετηρίασης του OCR, ενσωματώνοντας πιθανότητες της τμηματοποίησης των λέξεων στο μέτρο ομοιότητας της ανάκτησης. Η απόδοση του Word Spotting μπορεί να αυξηθεί, μοντελοποιώντας τις πιθανότητες της τμηματοποίησης των λέξεων. Περιγράφεται ένα πιθανοτικό μοντέλο του Word Spotting το οποίο ενσωματώνει πιθανότητες τμηματοποίησης λέξεων (word segmentation probabilities) και πιθανότητες αναγνώρισης λέξεων (word recognition probabilities). Οι πιθανότητες τμηματοποίησης λέξεων, λαμβάνονται μοντελοποιώντας τη δεσμευμένη κατανομή της απόστασης χαρακτηριστικών των κενών λέξεων. Τα αποτελέσματα της αναγνώρισης λέξεων αναπαριστώνται επίσης και αυτά με ένα πιθανοτικό μοντέλο. Η μοντελοποίηση των πιθανοτήτων των λέξεων αναγνώρισης, λαμβάνονται από τις αποστάσεις που επιστρέφονται από τον αναγνωριστή λέξεων (word recognizer) όπως φαίνεται στο σχήμα 2-3.



Σχήμα 2-3: [9] Τα αποτελέσματα της αναγνώρισης λέξεων αναπαριστώνται με πιθανοτικό μοντέλο.

Για την πρώτη προσέγγιση (image to image Matching), έχοντας τα διανύσματα χαρακτηριστικών, υπολογίζεται η ομοιότητα μεταξύ λέξεων με βάση τον DTW. Ο DTW, αξιολογήθηκε σε χειρόγραφα έγγραφα του George Washington, με μέτρο αξιολόγησης την μέση ακρίβεια (Mean Average Precision – MAP). Τα αποτελέσματα δίνουν, ότι σε 2372 λέξεις καλής ποιότητας το MAP βρέθηκε 40.98% και σε 3262 λέξεις κακής ποιότητας το MAP βρέθηκε 16.50%.

Για την δεύτερη προσέγγιση (bitwise Matching), κάθε εικόνα αναπαριστάται με 1024 bits από χαρακτηριστικά GSC. Η λέξη υποδιαιρείται σε 32 περιφέρειες και εξάγονται 16 δυαδικά χαρακτηριστικά από κάθε περιφέρεια. Η ομοιότητα μεταξύ λέξεων, μετράται από το ταίριασμα bit με bit των αντίστοιχων διανυσμάτων GSC χαρακτηριστικών των δύο εικόνων.

Μετά από πειράματα μεταξύ τεσσάρων μεθόδων ανάκτησης σε μία συλλογή από έντυπα του New York Pre-Hospital Care Report (PCR), όπου η πρώτη μέθοδος είναι η πρώτη προσέγγιση (image to image Matching) δηλαδή η χρήση του DTW, η δεύτερη μέθοδος είναι η δεύτερη προσέγγιση (bitwise Matching), δηλαδή η χρήση των GSC χαρακτηριστικών, και δύο μέθοδοι που ενσωματώνουν πιθανότητες στο μέτρο ομοιότητας τους, αναδείχθηκε ότι οι δύο τελευταίες μέθοδοι υπερέχουν των δύο πρώτων.

2.7.6 Εργασία με θέμα « Scale Space technique for word segmentation in handwritten manuscripts »

Καθώς το στάδιο εξαγωγής των εικόνων λέξεων από το αρχικό χειρόγραφο έγγραφο, είναι ένα από τα πρώτα στάδια της τεχνικής του Word Spotting, όπως φαίνεται και από το σχήμα 2-2, θεωρήσαμε σκόπιμο να αναφέρουμε και μια τεχνική που εφαρμόζει τμηματοποίηση γραμμών και λέξεων. Η τεχνική αυτή παρουσιάζεται στο [2] και είναι μια από τις πιο ευρέως χρησιμοποιούμενες τεχνικές για τον σκοπό της τμηματοποίησης χειρόγραφων εγγράφων. Οι συγγραφείς του [2], ξεκινούν εξηγώντας αρχικά την γενική ιδέα της θεωρίας scale space, και στη συνέχεια περιγράφουν τα στάδια της τεχνικής τους. Πρώτο στάδιο είναι η εφαρμογή προεπεξεργασίας στο έγγραφο, κατά την οποία, αφαιρείται ο θόρυβος με τον εξής τρόπο: Με την βοήθεια ενός Laplacian of Gaussian (LOG) φίλτρου και την κατωφλίωση του Laplacian σε θετικές τιμές, λαμβάνεται μια δυαδικής μορφής εικόνα του εγγράφου. Επειδή ο θόρυβος που παρουσιάζεται στην περίπτωση τους είναι κάθετες και οριζόντιες γραμμές που εμφανίζονται στην περιφέρεια της σελίδας, εφαρμόζεται μια ανάλυση συνδεδεμένων κομματιών (connected component analysis), με τη βοήθεια ενός συρτού παραθύρου, αφαιρώντας γραμμές ανάλογα με το

λόγο διαστάσεων (aspect ratio), το ύψος και το πλάτος των γραμμών. Το δεύτερο στάδιο είναι η τμηματοποίηση γραμμών, η οποία πραγματοποιείται με την χρήση του projection profile. Εν συντομία, παίρνοντας το horizontal projection profile της εικόνας, και εφαρμόζοντας σε αυτό ένα low pass Gaussian φίλτρο, παίρνουμε εξομαλυσμένα τοπικά ελάχιστα και μέγιστα, όπου τα ελάχιστα είναι οι γραμμές που περιέχουν μελάνι, επομένως είναι εύκολο να γίνει η τμηματοποίηση από εδώ και πέρα. Στο τρίτο στάδιο, κάθε γραμμή περνάει από ένα ανισοτροπικό Gaussian φίλτρο για να μπορούν να διαχειριστούν οι κλίμακες ύψους και πλάτους της γραμμής, έτσι ώστε οι χαρακτήρες των λέξεων τελικά να ενωθούν και να μπορεί να πραγματοποιηθεί η τμηματοποίηση των λέξεων. Το δύσκολο κομμάτι του σταδίου αυτού είναι η σωστή επιλογή των κλιμάκων αυτών. Στο τελευταίο στάδιο εφαρμόζεται εξαγωγή blob και μετά-επεξεργασία, όπου blob είναι τα κομμάτια που αποτελούν τις τροποποιημένες λέξεις και εξάχθηκαν στο προηγούμενο στάδιο. Συγκεκριμένα, κάθε κομμάτι blob, περιορίζεται μέσα σε ένα πλαίσιο οριοθέτησης και αντιστοιχίζεται το ίδιο πλαίσιο οριοθέτησης στο αρχικό έγγραφο, το οποίο λογικά θα περιλαμβάνει την αντίστοιχη λέξη. Η επεξεργασία που εφαρμόζεται σε αυτό το στάδιο είναι η επέκταση του πλαισίου οριοθέτησης, έτσι ώστε να περιλαμβάνει ascenders (για παράδειγμα το ελληνικό γράμμα «δ» είναι ascender, παρουσιάζει ανέβασμα) και descenders (για παράδειγμα το ελληνικό γράμμα «γ» είναι descender, παρουσιάζει κατέβασμα) της λέξης που αποκόπηκαν κατά το τρίτο στάδιο, ενώ επίσης εφαρμόζεται και ένα φίλτρο για να καθαρίσει την εικόνα. Τα αποτελέσματα που δίνονται στο [2] δίνουν μέσο όρο ακρίβειας για την τεχνική αυτή 87.6% στα δεδομένα τους, ενώ η τεχνική χρειάστηκε περίπου 120 δευτερόλεπτα για να εφαρμόσει τμηματοποίηση σε ένα έγγραφο 800x600 pixels, σε ηλεκτρονικό υπολογιστή με επεξεργαστή Pentium 200 MHz.

2.7.7 Εργασία με θέμα « Indexing for a Digital Library of George Washington's manuscripts – A study of word Matching techniques »

Στο [3], γίνεται μια σύγκριση αλγόριθμων Matching, σε 3 σύνολα δεδομένων, 10 σελίδων το καθένα, τα οποία πάρθηκαν από την συλλογή χειρόγραφων κειμένων του George Washington. Πιο συγκεκριμένα, στα σύνολα αυτά, εφαρμόζονται οι τεχνικές της τμηματοποίησης γραμμών και λέξεων, του pruning, της κανονικοποίησης, και τελικά του Matching. Η τεχνική τμηματοποίησης γραμμών και λέξεων που εφαρμόστηκε ήταν η τεχνική που περιγράφεται στο [2]. Κατά την εφαρμογή της τεχνικής του pruning, έγιναν κάποιες συγκρίσεις μεταξύ διάφορων συνδυασμών κριτηρίων κλαδέματος, όπου καλύτερος συνδυασμός αποδείχτηκε να είναι για την περίπτωση τους, η περιοχή της εικόνας (image area), ο λόγος διαστάσεων (aspect ratio) και το πλήθος των descenders της λέξης. Για την κανονικοποίηση, εφαρμόσαν δυαδικοποίηση της εικόνας, και αφαίρεσαν τον θόρυβο, προχωρώντας στη συνέχεια με κανονικοποίηση της κλίμακας, διόρθωση της οριζόντιας κλίσης της λέξης (slant), και ευθυγράμμιση των λέξεων. Μια περιγραφή του φαινομένου slant δίνεται στο κεφάλαιο 3. Κατά την εφαρμογή του Matching στα 3 σύνολα δεδομένων, έγινε σύγκριση ανάμεσα στα αποτελέσματα που έδωσαν οι ακόλουθοι αλγόριθμοι: XOR, EDM, Sum of Squared Differences (SSD), SLH, και Shape Context(SC). Επειδή τα 2 από τα 3 σύνολα δεδομένων ήταν κακής ποιότητας σύμφωνα με τους συγγραφείς, δόθηκαν μόνο τα αποτελέσματα ενός συνόλου. Τα αποτελέσματα αυτού του πειράματος, έφεραν την ακόλουθη κατάταξη σε 4 αλγόριθμους ξεκινώντας από αυτόν που έφερε τα καλύτερα αποτελέσματα: EDM, XOR, SSD, SLH, όπως φαίνεται και από τον πίνακα 2-1. Ο SC δεν κατάφερε να χρησιμοποιηθεί αποδοτικά και γι' αυτό δεν εμφανίστηκαν τα αποτελέσματα για αυτόν τον αλγόριθμο. Από τα αποτελέσματα, επίσης παρατηρούμε ότι ο SLH χρειάζεται πολύ περισσότερο χρόνο εκτέλεσης σε σύγκριση με τους EDM και XOR, ενώ τονίζεται επίσης και η σημασία της κανονικοποίησης στην τεχνική του Word Spotting.

	XOR	SSD	EDM	SLH
Average Precision	0.5414	0.5266	0.7338	0.4243

Πίνακας 2-1: Average precision για ένα σύνολο δεδομένων του [3] των αλγορίθμων: XOR, Sum of Squared Differences (SSD), Euclidean Distance Mapping (EDM), και ενός αλγόριθμου των Scott and Longuet-Higgins (SLH).

2.7.8 Εργασία με θέμα « Word image Matching using dynamic time warping»

Στο [12], και πάλι γίνεται μια σύγκριση διάφορων αλγορίθμων στο στάδιο του Matching, σε 2 σύνολα δεδομένων, 10 σελίδων το καθένα, που πάρθηκαν από την συλλογή χειρόγραφων κειμένων του George Washington, και από τα οποία το πρώτο ήταν αποδεκτής ποιότητας ενώ το δεύτερο ήταν αρκετά υποβαθμισμένης ποιότητας. Οι συγγραφείς, αρχικά παρουσιάζουν τον αλγόριθμο DTW και τα πλεονεκτήματα που αυτός προσφέρει. Και σε αυτή την περίπτωση χρησιμοποιήθηκε η τεχνική που περιγράφεται στο [2] για την τμηματοποίηση των λέξεων, ενώ σαν τεχνική του pruning, χρησιμοποιήθηκε η τεχνική που περιγράφεται στο [13]. Κατά την κανονικοποίηση, διορθώνονται οι κατακόρυφες (skew) και οριζόντιες κλίσεις (slant) που υπάρχουν στις λέξεις, ενώ επίσης πραγματοποιείται και μια ευθυγράμμιση των χαρακτήρων των λέξεων. Τα χαρακτηριστικά που εξάγονται από κάθε εικόνα λέξης είναι τα εξής: vertical projection profile, upper και lower word profile και vertical background to ink transitions profile, τα οποία κανονικοποιούνται στο διάστημα τιμών [0,1]. Μια περιγραφή των χαρακτηριστικών καθώς και των φαινομένων skew και slant δίνεται στο κεφάλαιο 3. Για να προχωρήσουν στο στάδιο του Matching, οι συγγραφείς κατασκεύασαν 4 σύνολα δεδομένων από τα 2 αρχικά σύνολα, από τα οποία το κάθε ένα περιείχε:

A: 15 εικόνες από το 1ο σύνολο δεδομένων.

B: Όλες τις εικόνες του 1ου συνόλου δεδομένων. 2381 εικόνες (9 από τις οποίες δεν περιείχαν λέξεις).

C: 32 εικόνες του 2ου συνόλου δεδομένων.

D: Όλες τις εικόνες του 2ου συνόλου δεδομένων. 3370 εικόνες (108 από τις οποίες δεν περιείχαν λέξεις).

Οι αλγόριθμοι που συγκρίθηκαν στο στάδιο του Matching, φαίνονται στον πίνακα 2-2, που περιέχει τους χρόνους εκτέλεσης τους σε ηλεκτρονικό υπολογιστή με επεξεργαστή 400 MHz. Από τους 6 αυτούς αλγόριθμους, προκύπτει από τα τελικά αποτελέσματα που δίνονται στο [12], ότι για τα σύνολα δεδομένων τους, καλύτερο ποσοστό average precision δίνει ο DTW, με ποσοστό περίπου 65%, ενώ δεύτερος έρχεται ο EDM.

Αλγόριθμος	XOR	SSD	SLH	SC	EDM	DTW
Time (sec)	13	72	121	~50	14	~2

Πίνακας 2-2: Συγκριτικοί χρόνοι εκτέλεσης των αλγορίθμων: XOR, Sum of Squared Differences (SSD), ενός αλγόριθμου των Scott and Longuet-Higgins (SLH), Shape Context (SC), Euclidean Distance Mapping (EDM), και Dynamic Time Warping (DTW).

2.7.9 Εργασία με θέμα « Application of multi-level classifiers and Clustering for automatic Word Spotting in historical document images»

Στο [14], παρουσιάζεται ένας εναλλακτικός τρόπος εφαρμογής του Word Spotting. Αυτή η τεχνική δεν απαιτεί τμηματοποίηση γραμμών ή λέξεων, και είναι μια εύρωστη τεχνική που παρουσιάζει ανεξαρτησία τρόπου γραφής και γλώσσας κειμένου, όπως ισχυρίζονται οι συγγραφείς. Υπάρχουν δύο στάδια σε αυτή την τεχνική, το στάδιο της προ-επεξεργασίας, και το στάδιο του Word Spotting, στο οποίο χρησιμοποιούνται η Euclidean Distance Technique (EDT) και ο DTW. Το στάδιο της προ-επεξεργασίας, ως συνήθως εφαρμόζεται για να διαχωριστεί το background από το foreground της εικόνας, κάτι που επιτυγχάνεται με την χρησιμοποίηση δύο content level classifiers, όπως χαρακτηρίζονται στο κείμενο, σαν χαρακτηριστικά ομαδοποίησης. Οι δύο αυτοί ταξινομητές είναι: ο Estimated Background (EB), ο οποίος παίρνει την τιμή ενός pixel που δεν είναι stroke, και ο Stroke Gray Level (SGL), ο οποίος παριστάνει το μέτρο της τοπικής έντασης stroke. Παρόλο που δεν διευκρινίζεται πούθενά στο κείμενο τι ακριβώς εννοείται με τον όρο stroke, υποθέτουμε ότι σαν stroke λαμβάνεται ένα pixel που έχει τιμή μικρότερη από ένα συγκεκριμένο κατώφλι αν η εικόνα είναι gray scale. Ακόμα ένας ταξινομητής, ο Edge Profile (EP), χρησιμοποιείται για να βοηθήσει στον υπολογισμό του EB. Προχωρώντας τώρα στο στάδιο του Word Spotting, χρησιμοποιείται μια τεχνική βασισμένη στην σύγκριση σχημάτων. Τα ενωμένα κομμάτια (Connected Components, CC), λαμβάνονται από μια εικόνα, και δημιουργείται ένα βασικό πρότυπο, βασισμένο στις αμοιβαίες αποστάσεις των κομματιών αυτών. Όλα τα βασικά πρότυπα εισάγονται σε μια βιβλιοθήκη, και η εικόνα query συγκρίνεται μόνο με αυτά τα βασικά στοιχεία. Η απόσταση που χρησιμοποιείται για την σύγκριση είναι η EDT, με το πλάτος του stroke να χρησιμοποιείται σαν εκ των προτέρων πληροφορία, ενώ στη συνέχεια, εφαρμόζεται ο DTW ξεχωριστά κατά μήκος της οριζόντιας και κάθετης διεύθυνσης, έτσι ώστε να αντισταθμιστούν οι χαμηλής έντασης περιοχές των strokes. Επίσης, κάθε φορά χρησιμοποιείται μια τιμή κατωφλίου για την επιλογή του βασικού προτύπου. Κατά την επεξεργασία της ομαδοποίησης, χρησιμοποιούνται 6 χαρακτηριστικά: ο λόγος διαστάσεων (aspect ratio), η οριζόντια συχνότητα (horizontal frequency), το κατά κλίμακα κάθετο κέντρο μάζας (scaled vertical center of mass), το πλήθος των branch points, ο λόγος ύψους προς ύψος γραμμής (height ratio to line height), και η παρουσία τρυπών (presence of holes). Για την ομαδοποίηση χρησιμοποιήθηκε το δίκτυο Kohonen. Στα πειράματα που έγιναν, χρησιμοποιήθηκε ένα σύνολο δεδομένων 20 σελίδων, από τα δεδομένα του Juma Al Majid Center, στο οποίο η ποιότητα κάποιων από τις σελίδες ήταν υποβαθμισμένη. Χρησιμοποιώντας και τον δείκτη αξιολόγησης F-Measure, που περιγράφεται στο 4.4.3, και βλέποντας τα αποτελέσματα της τεχνικής, εξάγεται το συμπέρασμα ότι τα αποτελέσματα είναι αρκετά ικανοποιητικά, ενώ δίνεται ότι το υπολογιστικό κόστος της προ-επεξεργασίας είναι περίπου 1 λεπτό ανά Megapixel, και ότι η προ-επεξεργασία και η ομαδοποίηση των CC μπορεί να εκτελεστεί μόνο μια φορά. Ο χρόνος αναζήτησης στην συγκεκριμένη υλοποίηση είναι περίπου 0.06 sec/Basic CC.

2.7.10 Εργασία με θέμα « Keyword-guided Word Spotting in historical printed documents using synthetic data and user feedback»

Ακόμα μια διαφορετική προσέγγιση της τεχνικής του Word Spotting, εφαρμόζεται στο [1], όπου παρουσιάζεται μια νέα μέθοδος Word Spotting για ιστορικά χειρόγραφα έγγραφα. Η μέθοδος αυτή αποτελείται από 4 στάδια: Δημιουργία συνθετικών εικόνων λέξεων, τμηματοποίηση λέξεων με την χρήση δυναμικών παραμέτρων, εξαγωγή χαρακτηριστικών από κάθε εικόνα λέξης, και μια διαδικασία ανάκτησης με βελτιστοποίηση που επιτυγχάνεται από την τροφοδότηση στοιχείων από τον χρήστη. Το σύνολο δεδομένων που χρησιμοποιήθηκε ήταν ιστορικά τυπωμένα ελληνικά κείμενα της περιόδου 1471-1821. Αρχικά, η εικόνα που λαμβάνεται δυαδικοποιείται και

τροποποιείται, όπως περιγράφεται στα [15] και [16], ενώ στη συνέχεια, γίνεται μια εκτίμηση του μέσου ύψους χαρακτήρων, κάτι που θα χρησιμεύσει σε επόμενα στάδια. Τελικά, η προεπεξεργασία τελειώνει με την εφαρμογή της αφαίρεσης πλαισίου, όπως περιγράφεται στο [17]. Κατά την διαδικασία τμηματοποίησης, χρησιμοποιείται το μέσο ύψος χαρακτήρων που υπολογίστηκε προηγουμένως, στον αλγόριθμο Run Length Smoothing Algorithm (RLSA) [18,19], για να επιτευχθεί τμηματοποίηση γραμμών και λέξεων. Ο χρήστης επιλέγει ένα πρότυπο εικόνας για κάθε χαρακτήρα, για να δημιουργήσει την λέξη query, και στη συνέχεια, ευθυγραμμίζει το κάτω μέρος των χαρακτήρων της, για να ελαχιστοποιηθούν προβλήματα ευθυγράμμισης. Για να καθοριστούν οι αποστάσεις μεταξύ των χαρακτήρων της λέξης, χρησιμοποιείται και πάλι το μέσο ύψος χαρακτήρων. Προχωρώντας στο στάδιο της εξαγωγής χαρακτηριστικών, πρώτα εφαρμόζεται κανονικοποίηση. Ο χρήστης σε αυτό το σημείο, ορίζει ένα πλαίσιο οριοθέτησης που θα περιέχει κάθε λέξη, και κάθε λέξη θα βρίσκεται στο κέντρο του πλαισίου οριοθέτησης της. Μετά την κανονικοποίηση, έρχεται το υβριδικό σχέδιο χαρακτηριστικών ή hybrid feature scheme, όπως το αναφέρουν οι συγγραφείς. Τα χαρακτηριστικά που επιλέγονται να εξαχθούν από κάθε εικόνα λέξης είναι τα εξής: 1) Η εικόνα χωρίζεται σε ζώνες, και η υπολογισμένη ένταση των pixel κάθε χαρακτήρα εμφανίζεται σε κάθε ζώνη. 2) Η εικόνα χωρίζεται σε ζώνες, στις οποίες βρίσκονται τα υπολογισμένα upper και lower word profiles της λέξης. Το τμήμα του [1], με τίτλο word image retrieval, ξεκινά με την διαδικασία σύγκρισης των εικόνων λέξεων, όπου τα χαρακτηριστικά της λέξης query, συγκρίνονται με τα χαρακτηριστικά όλων των λέξεων του συνόλου δεδομένων. Στη συνέχεια, όσες εικόνες επιστρέφονται από την μέθοδο αυτή, εμφανίζονται μέσω ενός Graphical User Interface (GUI) στον χρήστη, ο οποίος επιλέγει τις εικόνες λέξεων που επιθυμεί να παραμείνουν στη λίστα. Όταν οι άσχετες εικόνες απομακρυνθούν μετά την παρέμβαση του χρήστη, η νέα λίστα δεδομένων εμφανίζεται και πάλι στον χρήστη. Σύμφωνα με τα αποτελέσματα η μέθοδος μπορεί να είναι ικανοποιητικής απόδοσης, ενώ προκύπτει το συμπέρασμα ότι η παρέμβαση του χρήστη επηρεάζει αρκετά την απόδοση της μεθόδου.

ΚΕΦΑΛΑΙΟ 3 - ΣΤΑΔΙΑ ΥΛΟΠΟΙΗΣΗΣ

3.1 Προ-Επεξεργασία (Preprocessing)

Το στάδιο της προ-επεξεργασίας, όπως και κάθε στάδιο του Word Spotting είναι ζωτικής σημασίας για την εφαρμογή των επόμενων σταδίων, καθώς και την εξαγωγή σωστών αποτελεσμάτων των σταδίων αυτών, αφού η αναποτελεσματική εφαρμογή προ-επεξεργασίας είναι πολύ πιθανό να οδηγήσει στην αποτυχία ολόκληρο το μοντέλο του Word Spotting.

Σκοπός του σταδίου αυτού είναι η αφαίρεση του θορύβου, η μετατροπή των εικόνων λέξεων σε μια κάπως ομοιότυπη μορφή, καθώς και η βελτίωση των εικόνων σε περιπτώσεις υποβαθμισμένων λέξεων.

Στη συνέχεια θα παρουσιάσουμε κάποιες εφαρμογές που περιλαμβάνονται στο στάδιο της προ-επεξεργασίας, σχεδόν όλες από τις οποίες εφαρμόζονται σε μεθόδους που θα περιγράψουμε σε επόμενο κεφάλαιο.

3.1.1 Εφαρμογές προ-επεξεργασίας (Preprocessing Applications)

Κάποιες από τις πιο γνωστές και ευρέως χρησιμοποιούμενες εφαρμογές προ-επεξεργασίας είναι οι πιο κάτω:

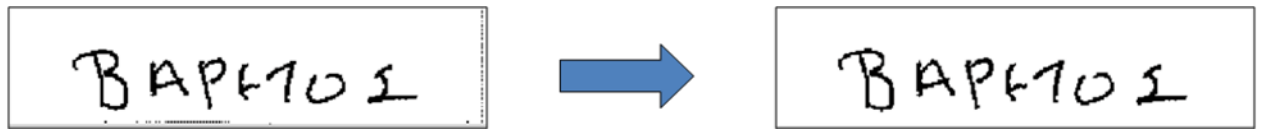
- 1) Αφαίρεση θορύβου (Noise removal).
- 2) Ένωση σπασμένων κομματιών της λέξης (Connect broken parts of the word image).
- 3) Εντοπισμός και αφαίρεση του φαινομένου skew (Skew detection & correction).
- 4) Εντοπισμός και αφαίρεση του φαινομένου slant (Slant detection & correction).
- 5) Περιορισμός της λέξης μέσα σε πλαίσιο οριοθέτησης (Contain the word in it's bounding box).
- 6) Κανονικοποίηση του font (Font normalization).

3.1.1.1 Noise Removal

Ο θόρυβος εμφανίζεται πολύ συχνά στις περιπτώσεις εικόνων που παράχθηκαν μέσω σάρωσης (scan) σαν οριζόντιες ή και κάθετες γραμμές μέσα στην εικόνα. Η παρουσία του θορύβου στην εικόνα, ακόμα και πριν την σάρωση είναι φυσικά πολύ πιθανή π.χ. με τη μορφή λεκέδων, ενώ η ύπαρξη ή η εμφάνιση διαφόρων άλλων μορφών θορύβου είναι πάντα πιθανή.

Στη περίπτωση μας, η πιο συχνή εμφάνιση θορύβου που υπήρχε στα δεδομένα μας ήταν της μορφής μικρών συνεχόμενων και σπασμένων οριζόντιων και κάθετων κομματιών στις άκρες της εικόνας, που πιθανόν να παρέμειναν από την τμηματοποίηση γραμμών και λέξεων του αρχικού εγγράφου, ή και να εμφανίστηκαν κατά τη σάρωση του εγγράφου. Ο θόρυβος αντιμετωπίστηκε με την αφαίρεση των κομματιών που υπήρχαν μέσα στην εικόνα και είχαν μέγεθος μικρότερο από ένα συγκεκριμένο αριθμό pixel, ο οποίος στην περίπτωση μας ήταν ο αριθμός 10.

Η εφαρμογή αυτή χρησιμοποιήθηκε επειδή τα δεδομένα μας ήταν σχετικά καλής ποιότητας, αφού η εφαρμογή αυτής της μεθόδου σε δεδομένα υποβαθμισμένης ποιότητας είναι πολύ πιθανό να τα υποβαθμίσει ακόμα περισσότερο. Ένα παράδειγμα αφαίρεσης θορύβου φαίνεται στο σχήμα 3-1.

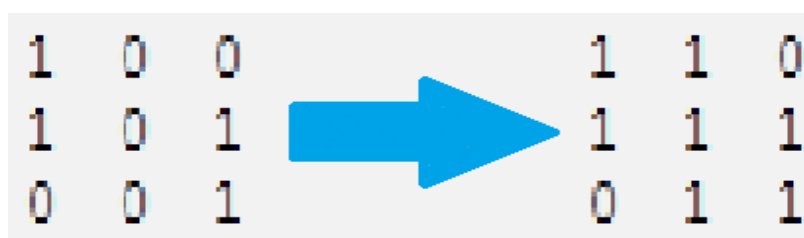


Σχήμα 3-1: Παράδειγμα αφαίρεσης θορύβου της λέξης “ΒΑΡΕΤΟΣ”.

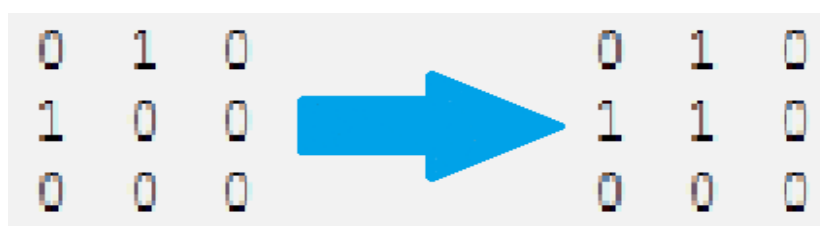
3.1.1.2 Connect Broken Parts of the Word Image

Αφού ο θόρυβος έχει πλέον αφαιρεθεί από τις εικόνες, το επόμενο βήμα είναι η προσπάθεια να βελτιωθεί ως ένα βαθμό η ποιότητα της εικόνας. Μια μέθοδος που θα μπορούσε να βοηθήσει σε αυτό είναι η ένωση σπασμένων κομματιών μέσα στη λέξη. Καθώς αυτή είναι η μόνη εφαρμογή που εκτελείται στην υλοποίηση μας με σκοπό τη βελτίωση της εικόνας, δεν μπορούμε να περιμένουμε και μεγάλη αλλαγή στη λέξη, αλλά έχει βρεθεί ότι σε αρκετές περιπτώσεις επιφέρει μια σημαντική αλλαγή. Φυσικά με την ύπαρξη πολύ υποβαθμισμένης ποιότητας λέξεων δεν μπορεί να επιφέρει σημαντική διόρθωση των εικόνων η χρήση μιας τέτοιας εφαρμογής.

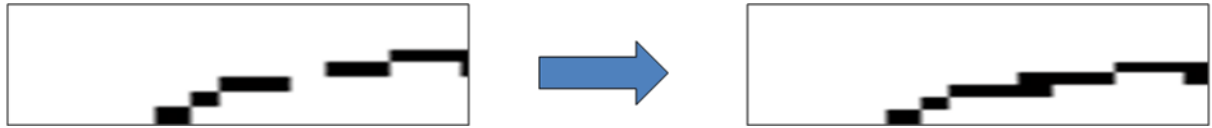
Αναλυτικότερα, ο τρόπος που λειτουργεί η εφαρμογή αυτή είναι ο εξής: αν υπάρχει κενό (π.χ. pixel με τιμή 0) ανάμεσα σε δύο κοντινά κομμάτια (π.χ. pixel με τιμή 1), τότε αυτό το κενό γεμίζει με pixel τιμής 1, όπως φαίνεται στο σχήμα 3-2. Αυτές οι συμπληρώσεις pixel πραγματοποιείται μόνο αν το κενό ανάμεσα στα κομμάτια είναι πλάτους ενός pixel. Επίσης, αν υπάρχει κάπου γωνία στην οποία για παράδειγμα υπάρχει pixel με τιμή 0, ενώ τα υπόλοιπα στοιχεία του κομματιού στο οποίο περιέχεται η γωνία είναι pixel τιμής 1, τότε το pixel στη γωνιά παίρνει και αυτό τιμή 1, όπως ακριβώς φαίνεται στο σχήμα 3-3. Στο σχήμα 3-4 φαίνεται η χρησιμοποίηση της εφαρμογής που περιγράφηκε πιο πάνω σε σπασμένα κομμάτια ενός γράμματος μιας λέξης.



Σχήμα 3-2: Παράδειγμα σύνδεσης σπασμένων κομματιών.



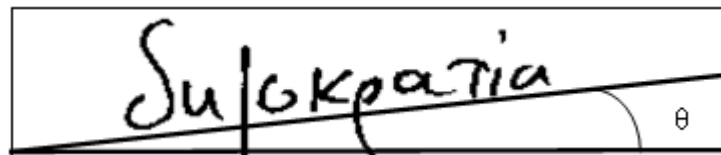
Σχήμα 3-3: Παράδειγμα γεμίσματος γωνιάς.



Σχήμα 3-4: Παράδειγμα σύνδεσης σπασμένων κομματιών σε λέξη.

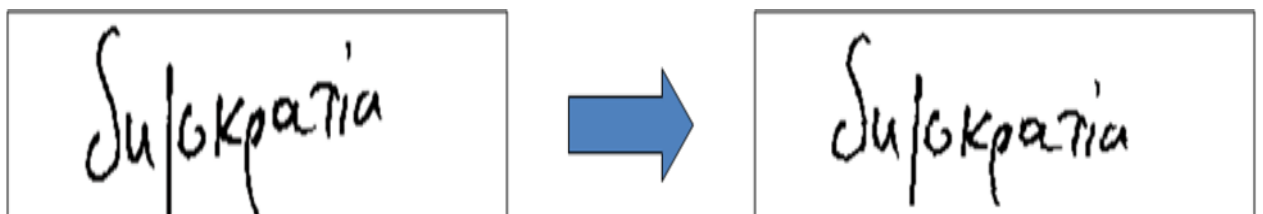
3.1.1.3 Skew Detection & Correction

Στα χειρόγραφα έγγραφα, λόγω των πολλών διαφορετικών τρόπων γραφής που υπάρχουν, καθώς ο κάθε άνθρωπος έχει το δικό του τρόπο γραφής, ίδιες λέξεις μπορούν να εμφανιστούν με πολύ διαφορετικούς τρόπους. Για να αντιμετωπιστεί αυτό το πρόβλημα, έστω και μερικώς, πρέπει να γίνει μια προσπάθεια, ούτως ώστε οι λέξεις να μετατραπούν σε μια κάπως ομοιότυπη μορφή για να μπορεί στα επόμενα στάδια οι λέξεις αυτές να είναι συγκρίσιμες. Στις εικόνες λέξεων που εξάγονται από χειρόγραφα έγγραφα, πολλές φορές εμφανίζεται η ύπαρξη του φαινομένου skew, το οποίο ορίζεται σαν η ύπαρξη γωνίας (έστω θ) ανάμεσα στην ευθεία πάνω στην οποία είναι η λέξη και μιας οριζόντιας ευθείας, όπως φαίνεται για παράδειγμα στο σχήμα 3-5.



Σχήμα 3-5: Παράδειγμα φαινομένου skew στη λέξη “δημοκρατία”.

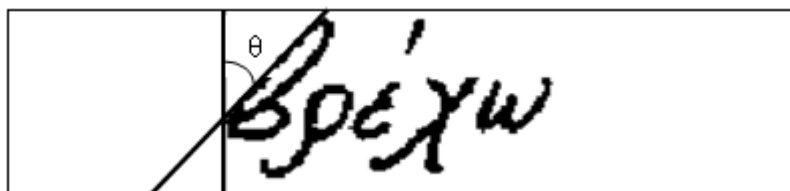
Ενώ στην περίπτωση των μεθόδων που υλοποιήθηκαν στην παρούσα εργασία, η εφαρμογή του εντοπισμού και αφαίρεσης του φαινομένου skew δεν χρησιμοποιείται, εντούτοις είναι μια από τις πιο διαδεδομένες εφαρμογές που πραγματοποιούνται στο στάδιο της προ-επεξεργασίας. Στο σχήμα 3-6 φαίνεται η διόρθωση του φαινομένου αυτού, ενώ δίνονται τα [20],[21] και [22] σαν απλή αναφορά σχετική με το θέμα αυτό.



Σχήμα 3-6: Παράδειγμα διόρθωσης της γωνιάς skew στη λέξη “δημοκρατία”.

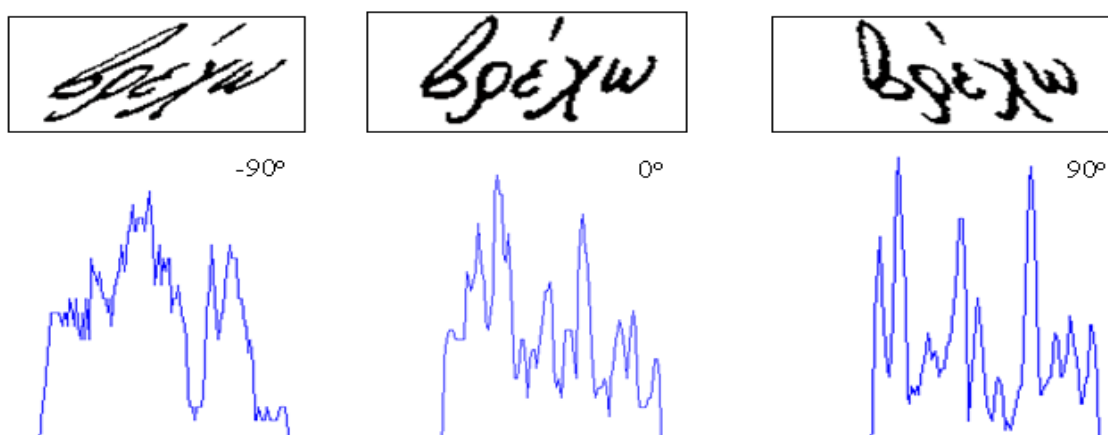
3.1.1.4 Slant Detection & Correction

Το φαινόμενο slant είναι αντίστοιχο με το φαινόμενο skew, μόνο που αυτή τη φορά μιλάμε για την κάθετη περίπτωση ύπαρξης γωνίας αντί της οριζόντιας. Δηλαδή, το φαινόμενο αυτό ορίζεται σαν η ύπαρξη γωνίας (έστω θ) ανάμεσα στην ευθεία πάνω στην οποία ορίζονται καθέτως τα γράμματα της λέξης και μιας κατακόρυφης ευθείας, όπως φαίνεται για παράδειγμα στο σχήμα 3-7.



Σχήμα 3-7: Παράδειγμα φαινομένου slant στη λέξη “βρέχω”.

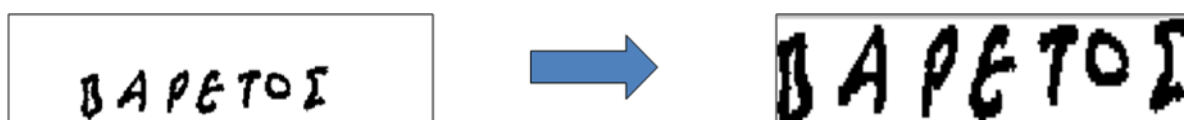
Στην περίπτωση της υλοποίησης της εφαρμογής εντοπισμού και αφαίρεσης του φαινομένου slant, στις μεθόδους αυτής της εργασίας που θα μελετηθούν στο κεφάλαιο 4, χρησιμοποιήθηκε μια παρόμοια μέθοδος με αυτή που περιγράφεται στο [23], σύμφωνα με την οποία, ο εντοπισμός και η διόρθωση της γωνίας slant, επιτυγχάνεται μέσω της παρατήρησης του κάθετου ιστογράμματος προβολής (Vertical projection histogram) της εικόνας. Συγκεκριμένα, η εικόνα μετασχηματίζεται τροποποιώντας την γωνιά (shear angle) της για γωνίες από -45 μοίρες ως 45 μοίρες, κάθε φορά εξάγοντας το κάθετο ιστόγραμμα προβολής της τροποποιημένης εικόνας. Η τροποποιημένη εικόνα της οποίας το κάθετο ιστόγραμμα προβολής έχει τις υψηλότερες κορυφές είναι συνήθως και η διορθωμένη από slant εικόνα, σύμφωνα με την περιγραφή που δίνεται στο [23]. Στην δική μας περίπτωση οι γωνίες που παίρνουμε είναι από -90 μοίρες ως 90 μοίρες. Αν για παράδειγμα έχουμε την εικόνα του σχήματος 3-7, βλέπουμε πως αυτή καθώς και το κάθετο ιστόγραμμα προβολής της τροποποιούνται κατά την αλλαγή της shear angle στο σχήμα 3-8. Στην περίπτωση αυτής της εικόνας, η γωνία (shear angle) που τελικά επιλέχθηκε ήταν αυτή των 80 μοιρών.



Σχήμα 3-8: Τροποποίηση της γωνιάς (shear angle) στη λέξη “βρέχω” και κάθετα ιστογράμματα προβολής των τροποποιημένων εικόνων για γωνίες -90 , 0 και 90 μοιρών.

3.1.1.5 Contain the Word in Its Bounding Box.

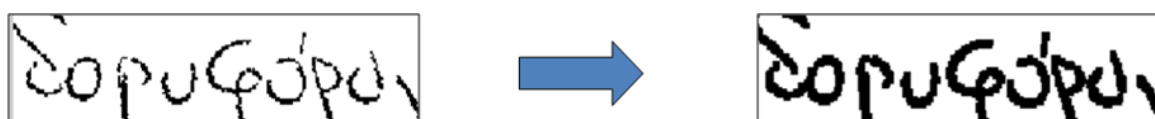
Για τον υπολογισμό των χαρακτηριστικών που ακολουθεί σε επόμενο στάδιο, και τη σωστή σύγκριση αυτών μεταξύ των εικόνων λέξεων που πραγματοποιείται αργότερα, καθώς και για την εφαρμογή της κανονικοποίησης του font που ακολουθεί, είναι σημαντικό η λέξη να περιοριστεί μέσα σε ένα πλαίσιο οριοθέτησης. Το πλαίσιο αυτό ορίζεται σαν το πλαίσιο το οποίο περιέχει τα στοιχεία της λέξης από την πρώτη γραμμή που περιέχει pixel με μελάνι μέχρι την τελευταία γραμμή που περιέχει pixel με μελάνι, και αντίστοιχα, από την πρώτη στήλη που περιέχει pixel με μελάνι μέχρι την τελευταία στήλη που περιέχει pixel με μελάνι. Με άλλα λόγια, απλά αφαιρεί τα κενά που βρίσκονται γύρω από την λέξη. Ένα παράδειγμα δίνεται στο σχήμα 3-9.



Σχήμα 3-9: Παράδειγμα περιορισμού της λέξης “ΒΑΡΕΤΟΣ” στο πλαίσιο οριοθέτησης.

3.1.1.6 Font Normalization

Συνεχίζοντας την προσπάθεια μετατροπής των εικόνων λέξεων σε ομοιότυπη μορφή, εφαρμόζουμε την κανονικοποίηση του font κάθε λέξης με τον εξής τρόπο: αφού έχουμε περιορίσει τη λέξη μέσα στο πλαίσιο οριοθέτησης της, παίρνουμε το λόγο (έστω R) του πλήθους των pixel που περιέχουν μελάνι στην εικόνα προς το πλήθος όλων των pixel της εικόνας, και τροποποιώντας την εικόνα με την πρόσθεση pixel μαύρου χρώματος γύρω από τα γράμματα της λέξης, μετατρέπουμε αυτό το λόγο σε μια τιμή λίγο μεγαλύτερη από 0.18. Η τιμή 0.18 σαν τιμή κατωφλίου για αυτό το σκοπό εφαρμόστηκε, αφού λίγες ήταν οι λέξεις στα δεδομένα μας που παρουσίαζαν τιμές σημαντικά μεγαλύτερες από αυτή την τιμή σαν υπολογισμένο λόγο R. Βρίσκουμε ότι η εφαρμογή της κανονικοποίησης του font, εκτός από το ότι βοηθά στην μετατροπή των λέξεων σε μια κάπως ομοιότυπη μορφή, σε αρκετές περιπτώσεις που το μελάνι της λέξης σε μια εικόνα είναι εξασθενημένο, βοηθά στην μερική αναβάθμιση της ποιότητας της λέξης. Ένα παράδειγμα εφαρμογής της κανονικοποίησης font φαίνεται στο σχήμα 3-10.



Σχήμα 3-10: Παράδειγμα κανονικοποίησης του font στη λέξη “δορυφόρος”.

3.2 Εξαγωγή Χαρακτηριστικών (Feature Extraction)

Η εξαγωγή χαρακτηριστικών και κυρίως η σωστή επιλογή τους, αποτελεί ένα από τους βασικότερους παράγοντες για να πραγματοποιηθεί με επιτυχία το στάδιο της ομαδοποίησης (Clustering).

Η εξαγωγή χαρακτηριστικών είναι ο μετασχηματισμός κάθε εικόνας σε ένα σύνολο από χαρακτηριστικά. Λόγω του ότι τα δεδομένα εισόδου σε ένα αλγόριθμο πολλές φορές είναι μεγάλα σύνολα δεδομένων, είναι δύσκολο να υποβληθούν σε επεξεργασία και έτσι χρειάζεται να υπόκεινται σε μία μειωμένη αναπαράσταση για να επιτελείται πιο εύκολα η επεξεργασία τους. Τα δεδομένα εισόδου, στην περίπτωση μας οι εικόνες, μετατρέπονται σε μία μειωμένη αναπαράσταση ως διανύσματα, τα οποία ονομάζονται διανύσματα χαρακτηριστικών (Features Vectors). Κάθε εικόνα λοιπόν θα μπορεί να αναπαρασταθεί με μειωμένη αναπαράσταση, δηλ. ως ένα διάνυσμα χαρακτηριστικών αντί ως μία εικόνα πλήρους μεγέθους.

Τα χαρακτηριστικά που εξάγονται πρέπει να είναι καλά επιλεγμένα, έτσι ώστε να εξάγουν όσο το δυνατόν περισσότερη πληροφορία. Όση περισσότερη πληροφορία έχουμε από κάθε εικόνα, τόσο το δυνατόν καλύτερα αποτελέσματα θα πάρουμε. Κάθε χαρακτηριστικό που εξάγεται από την εικόνα θα πρέπει να διαθέτει τη μέγιστη πληροφορία, έτσι ώστε τα επιλεγμένα χαρακτηριστικά να προσδιορίζουν μοναδικά την εικόνα ξεχωριστά. Η εξαγωγή χαρακτηριστικών με χαμηλή πληροφορία, δεν έχει καθόλου νόημα, μιας και δεν μπορεί να αντιπροσωπεύσει σωστά μια εικόνα, κάτι το οποίο μπορεί να οδηγήσει σε προβλήματα στις επόμενες διεργασίες φέρνοντας λανθασμένα αποτελέσματα.

3.2.1 Χαρακτηριστικά (Features)

Τα κύρια χαρακτηριστικά που εξάγονται και προσφέρουν την περισσότερη πληροφορία στην περίπτωση μας είναι τα παρακάτω προφίλ (profiles):

- 1) Upper Word Profile (UWP)
- 2) Lower Word Profile (LWP)
- 3) Left Word Profile (LeWP)
- 4) Right Word Profile (RWP)
- 5) Lower Minus Upper Word Profile (LMUWP)
- 6) Vertical Projection Profile (VPP)
- 7) Horizontal Projection Profile (HPP)
- 8) Vertical Background to Ink Transitions Profile (VBTITP)
- 9) Horizontal Background to Ink Transitions Profile (HBTITP)

Γενικά τα προφίλ μας δίνουν μία κατανομή της λέξης. Μας δίνουν δηλαδή μία αίσθηση στο πώς κατανέμεται το μελάνι (μαύρα pixels) μέσα στη λέξη.

Μερικά δευτερεύοντα χαρακτηριστικά, τα οποία προσφέρουν επίσης αρκετή πληροφορία είναι τα παρακάτω:

- 10) End Points
- 11) Perimeters Points
- 12) Holes Detection

Γενικά τα δευτερεύοντα χαρακτηριστικά, και συγκεκριμένα ο αριθμός των τελικών σημείων (End Points) και ο αριθμός των περιμετρικών σημείων (Perimeter Points), μας δίνουν πληροφορία σχετικά με τον όγκο της λέξης, δηλαδή πόσο μεγάλη ή μικρή είναι μια λέξη. Τα Holes Detection μας δίνουν πληροφορία σχετικά με το πού υπάρχουν τρύπες μέσα στη λέξη.

3.2.1.1 Upper Word Profile

Το Upper Word Profile είναι χαρακτηριστικό το οποίο περιγράφει το επάνω μέρος του σχήματος της λέξης. Ο υπολογισμός του χαρακτηριστικού αυτού είναι ως εξής: Σε κάθε στήλη της εικόνας, υπολογίζεται η απόσταση από την κορυφή του πλαισίου οριοθέτησης (Bounding Box), μέχρι το πλησιέστερο pixel με μελάνι (μαύρο pixel).

3.2.1.2 Lower Word Profile

Το Lower Word Profile είναι χαρακτηριστικό το οποίο περιγράφει το κάτω μέρος του σχήματος της λέξης. Ο υπολογισμός του χαρακτηριστικού αυτού είναι ως εξής: Σε κάθε στήλη της εικόνας, υπολογίζεται η απόσταση από το κάτω μέρος του πλαισίου οριοθέτησης (Bounding Box), μέχρι το πλησιέστερο pixel με μελάνι (μαύρο pixel).

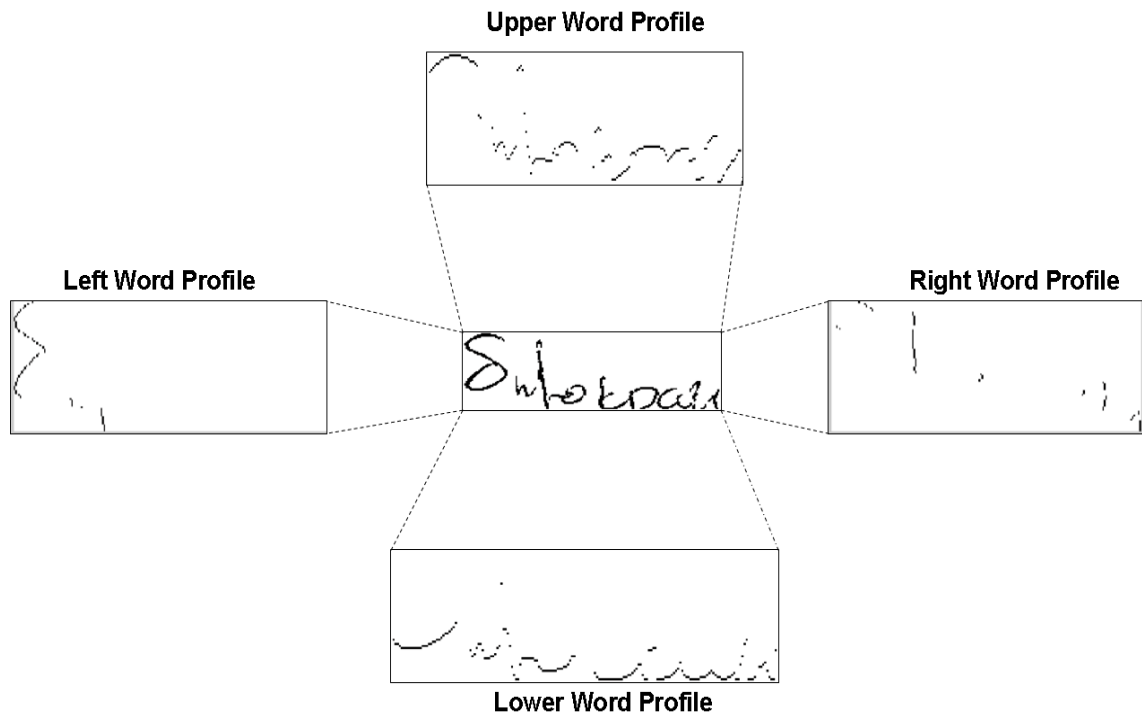
3.2.1.3 Left Word Profile

Το Left Word Profile είναι χαρακτηριστικό το οποίο περιγράφει το αριστερό μέρος του σχήματος της λέξης. Ο υπολογισμός του χαρακτηριστικού αυτού είναι ως εξής: Σε κάθε γραμμή της εικόνας, υπολογίζεται η απόσταση από το αριστερό μέρος του πλαισίου οριοθέτησης (Bounding Box), μέχρι το πλησιέστερο pixel με μελάνι (μαύρο pixel).

3.2.1.4 Right Word Profile

Το Right Word Profile είναι χαρακτηριστικό το οποίο περιγράφει το δεξί μέρος του σχήματος της λέξης. Ο υπολογισμός του χαρακτηριστικού αυτού είναι ως εξής : Σε κάθε γραμμή της εικόνας, υπολογίζεται η απόσταση από το δεξί μέρος του πλαισίου οριοθέτησης (Bounding Box), μέχρι το πλησιέστερο pixel με μελάνι (μαύρο pixel).

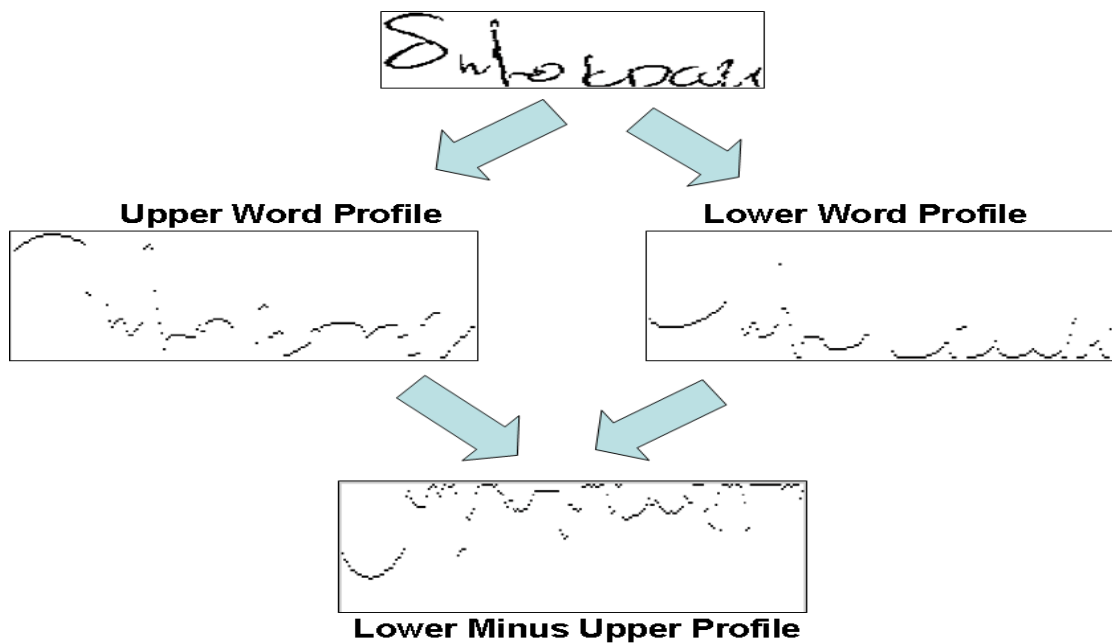
Ένα παράδειγμα των 4^{ov} χαρακτηριστικών που περιγράφονται πιο πάνω, φαίνεται στο σχήμα 3-11.



Σχήμα 3-11: Παράδειγμα Upper-Lower-Left-Right Word Profile της λέξης “δημοκρατία”.

3.2.1.5 Lower Minus Upper Word Profile

Το Lower Minus Upper Word Profile είναι χαρακτηριστικό το οποίο περιγράφει το εσωτερικό μέρος του σχήματος της λέξης. Ο υπολογισμός του χαρακτηριστικού αυτού είναι η διαφορά μεταξύ του Lower και του Upper Word Profile της λέξης που περιγράφηκαν προηγουμένως.



Σχήμα 3-12: Παράδειγμα Lower Minus Upper Word Profile της λέξης “δημοκρατία”.

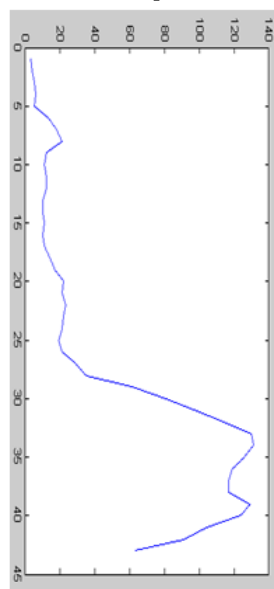
3.2.1.6 Vertical Projection Profile

Το Vertical Projection Profile είναι χαρακτηριστικό το οποίο περιγράφει την κατανομή του μελανιού στις στήλες της λέξης. Ο υπολογισμός του χαρακτηριστικού αυτού είναι ως εξής: Σε κάθε στήλη της εικόνας, υπολογίζεται το άθροισμα των pixel στα οποία υπάρχει μελάνι (μαύρα pixel).

3.2.1.7 Horizontal Projection Profile

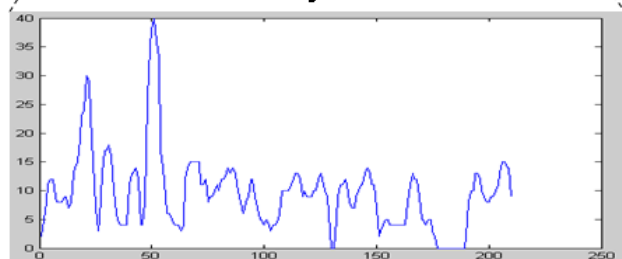
Το Horizontal Projection Profile είναι χαρακτηριστικό το οποίο περιγράφει την κατανομή του μελανιού στις γραμμές της λέξης. Ο υπολογισμός του χαρακτηριστικού αυτού είναι ως εξής: Σε κάθε γραμμή της εικόνας, υπολογίζεται το άθροισμα των pixel στα οποία υπάρχει μελάνι (μαύρα pixel).

Horizontal Projection Profile



δημοκρατία

Vertical Projection Profile



Σχήμα 3-13: Παράδειγμα Vertical-Horizontal Projection Profile της λέξης “δημοκρατία”.

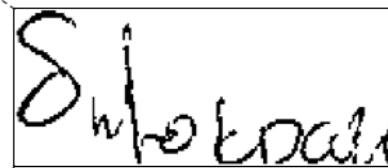
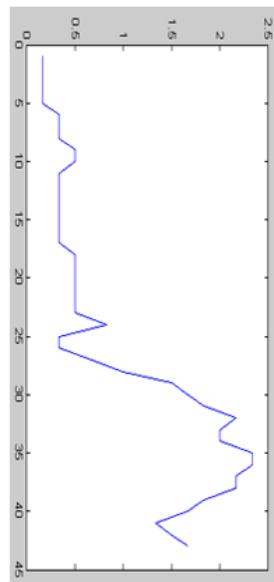
3.2.1.8 Vertical Background to Ink Transitions Profile

Το Vertical Background to Ink Transitions Profile είναι χαρακτηριστικό το οποίο περιγράφει την εσωτερική δομή της λέξης. Ο υπολογισμός του χαρακτηριστικού αυτού είναι ως εξής: Σε κάθε στήλη της εικόνας, υπολογίζεται ο αριθμός των μεταβολών από pixel του background σε pixel με μελάνι.

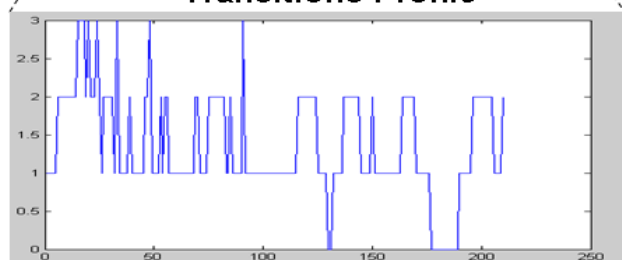
3.2.1.9 Horizontal Background to Ink Transitions Profile

Το Horizontal Background to Ink Transitions Profile είναι χαρακτηριστικό το οποίο περιγράφει την εσωτερική δομή της λέξης. Ο υπολογισμός του χαρακτηριστικού αυτού είναι ως εξής: Σε κάθε γραμμή της εικόνας υπολογίζεται ο αριθμός των μεταβολών από pixel του background σε pixel με μελάνι.

Horizontal Background to Ink Transitions Profile



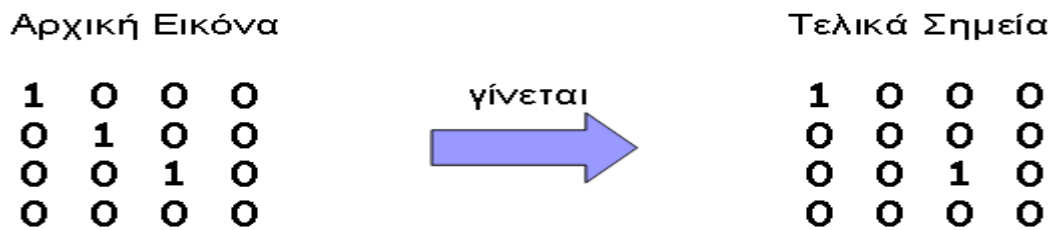
Vertical Background to Ink Transitions Profile



Σχήμα 3-14: Παράδειγμα Vertical-Horizontal Background to Ink Transitions Profile της λέξης “δημοκρατία”.

3.2.1.10 End Points

Ο αριθμός των pixel που αναφέρονται ως τελικά σημεία στη λέξη (End Points) είναι χαρακτηριστικό το οποίο περιγράφει τον όγκο της λέξης. Ο υπολογισμός του χαρακτηριστικού γίνεται ως εξής:



Σχήμα 3-15: Παράδειγμα υπολογισμού των End Points από την αρχική εικόνα. Τα τελικά σημεία (End Points) είναι ο συνολικός αριθμός των 1 που εμφανίζονται στην δεξιά εικόνα.

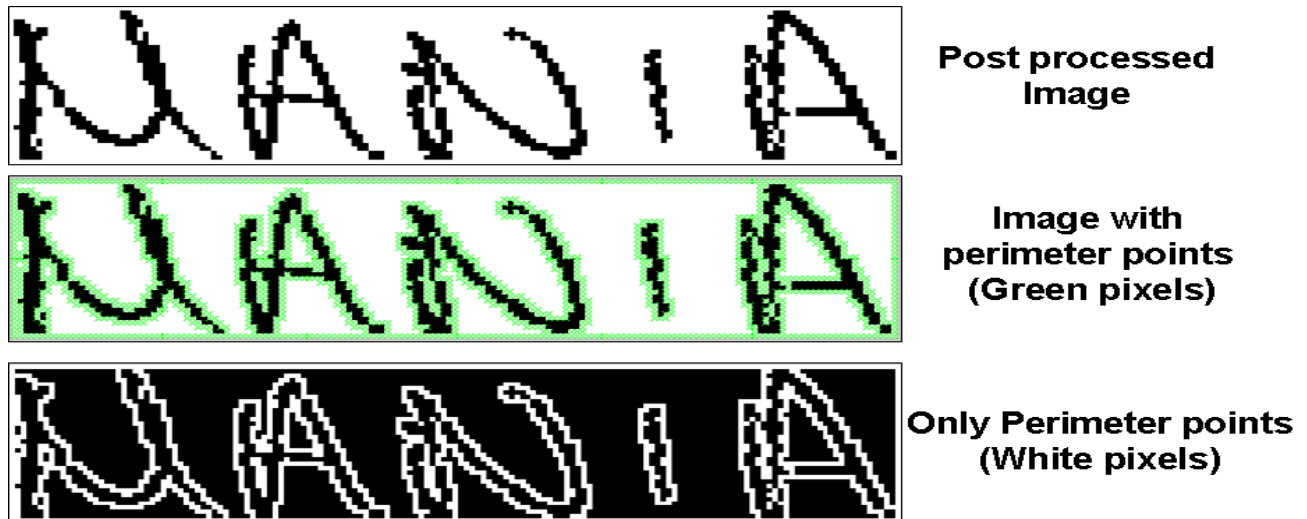
Στο σχήμα 3-16 βλέπουμε τα End Points της λέξης “MANIA”.



Σχήμα 3-16: Παράδειγμα των τελικών σημείων (End Points) της λέξης “MANIA”. Στην μεσαία εικόνα τα pixel που αναπαριστούν End Points σημειώνονται με πράσινο χρώμα ενώ η κάτω εικόνα δείχνει μόνο τα End Points (άσπρα pixel).

3.2.1.11 Perimeter Points

Ο αριθμός των pixel που αναφέρονται ως περιμετρικά σημεία στη λέξη (Perimeter Points) είναι χαρακτηριστικό το οποίο επίσης περιγράφει τον όγκο της λέξης. Ο υπολογισμός του χαρακτηριστικού αυτού γίνεται ανιχνεύοντας όλα τα περιμετρικά pixel της λέξης. Στο σχήμα 3-17 βλέπουμε τα Perimeter Points της λέξης “MANIA”.



Σχήμα 3-17: Παράδειγμα των περιμετρικών σημείων (Perimeter Points) της λέξης “MANIA”.

Στην μεσαία εικόνα τα pixel που αναπαριστούν Perimeter Points σημειώνονται με πράσινο χρώμα ενώ η κάτω εικόνα δείχνει μόνο τα Perimeter Points (άσπρα pixel).

3.2.1.12 Holes Detection

Οι τρύπες που εντοπίζονται μέσα στη λέξη (Holes Detection) και συγκεκριμένα τα pixel που αποτελούν το κέντρο κάθε τρύπας, καθώς επίσης το μέγεθος της τρύπας είναι χαρακτηριστικά τα οποία περιγράφουν κλειστά σημεία της λέξης. Ο υπολογισμός του χαρακτηριστικού αυτού γίνεται ανιχνεύοντας αρχικά όλες τις τρύπες της λέξης και για κάθε τρύπα υπολογίζεται το σημείο του κέντρου και το μέγεθος της τρύπας. Σε αυτό το σημείο η χρήση κατωφλίου (threshold) είναι απαραίτητη για την απόρριψη τρυπών πολύ μικρού μεγέθους ως χαρακτηριστικό. Στο σχήμα 3-18 βλέπουμε τις τρύπες που εμφανίζονται μέσα στη λέξη “MANIA”.



Σχήμα 3-18: Παράδειγμα στις τρύπες που ανιχνεύονται μέσα στη λέξη “MANIA”.

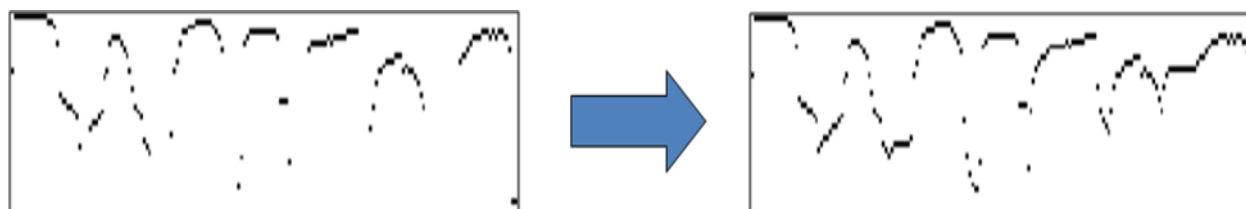
3.2.2 Παρεμβολή Χαρακτηριστικών (Interpolation)

Ακολουθώντας τη μέθοδο που εφαρμόστηκε στο [4], τα χαρακτηριστικά UWP, LWP, RWP και LeWP, χρειάστηκε να περάσουν από τη διαδικασία της παρεμβολής με την έννοια του γεμίσματος των κενών σημείων τους. Αυτό γίνεται γιατί σε πολλές περιπτώσεις, η εξασθένηση του μελανιού στις εικόνες, που προκαλείται από διάφορους παράγοντες, αφήνει κενές στήλες μέσα στη λέξη οι οποίες θα έπρεπε να περιέχουν μελάνι.

Για να διορθωθεί το πιο πάνω πρόβλημα, χρησιμοποιείται μια μέθοδος γραμμικής παρεμβολής. Δηλαδή, ακολουθείται μια διαδικασία, κατά την οποία εξετάζεται κάθε τιμή των διανυσμάτων των χαρακτηριστικών αυτών, και αν βρεθεί μια τιμή 0 σε κάποια θέση (έστω στη θέση i) του διανύσματος που περιέχει το χαρακτηριστικό, τότε παίρνουμε τις τιμές των δυο κοντινότερων γειτονικών του σημείων στο διάνυσμα που δεν περιέχουν μηδενική τιμή, και βάζουμε την τιμή που προκύπτει από τον μέσο όρο των γειτονικών αυτών τιμών σαν τιμή της θέσης i . Στην περίπτωση που δεν βρεθεί κάποιο από τα δύο γειτονικά στοιχεία που να μην περιέχει μηδενική τιμή, τότε η τιμή του i γίνεται ίδια με την τιμή του μόνου γειτονικού στοιχείου που βρέθηκε. Ένα απλό παράδειγμα της λειτουργίας της πιο πάνω διαδικασίας φαίνεται στον πίνακα 3-1, και ένα παράδειγμα της εφαρμογής της ίδιας διαδικασίας στο UWP μιας εικόνας χειρόγραφης λέξης φαίνεται στο σχήμα 3-19.

1	2	3	4	5	6	7	8
0	55	0	0	12	33	0	0
UWP(1) = UWP(2)	UWP(2)	UWP(3)= round[(UWP(2) +UWP(5))/2]	UWP(4)= round[(UWP(3) +UWP(5))/2]	UWP(5)	UWP(6)	UWP(7) = UWP(6)	UWP(8) = UWP(7)
55	55	34	23	12	33	33	33

Πίνακας 3-1: Παράδειγμα λειτουργίας της γραμμικής παρεμβολής για UWP. 1^η γραμμή: Αριθμός θέσης στο διάνυσμα, 2^η γραμμή: Αρχικές τιμές του UWP, 3^η γραμμή: Υπολογισμός των τιμών του παρεμβλημένου UWP, 4^η γραμμή: Οι νέες τιμές του UWP.



Σχήμα 3-19: Παράδειγμα εφαρμογής της παρεμβολής σε UWP μιας εικόνας χειρόγραφης λέξης.

3.2.3 Κανονικοποίηση Χαρακτηριστικών (Feature Normalization)

Λόγω μεταβολών στην ποιότητα των εικόνων όπως το ξεθώριασμα, οι διαφορές στο font, στη γραφή, οι διαφορές στο ύψος και το πλάτος κάθε εικόνας, τα χαρακτηριστικά που εξάγονται δεν κυμαίνονται στο ίδιο εύρος τιμών, παρουσιάζουν δηλαδή διαφορές στην κλίμακα, με αποτέλεσμα να δημιουργείται το πρόβλημα της σύγκρισης μεταξύ χαρακτηριστικών. Για να εκτελεστεί με επιτυχία η σύγκριση των εικόνων, όλα τα χαρακτηριστικά πρέπει να κανονικοποιηθούν σε ένα καθορισμένο εύρος τιμών. Πρέπει να υπάρχει δηλαδή μία κοινή κλίμακα για όλα τα χαρακτηριστικά κατά την σύγκριση. Το στάδιο της κανονικοποίησης καλείται να λύσει το πρόβλημα διαφορετικών κλιμάκων. Τις περισσότερες φορές, η κανονικοποίηση αναφέρεται ως η διαίρεση διαφόρων συνόλων δεδομένων με ένα κοινό συντελεστή, με σκοπό να φέρει όλα τα σύνολα δεδομένων σε μία κοινή κλίμακα, επιτρέποντας έτσι την σύγκρισή τους. Για το σκοπό αυτό, η κοινή κλίμακα που επιλέχθηκε για όλα τα χαρακτηριστικά είναι το διάστημα [0,1].

Στη συνέχεια θα μιλήσουμε για το πώς γίνεται η κανονικοποίηση για κάθε ένα από τα χαρακτηριστικά ξεχωριστά.

- Τα χαρακτηριστικά τα οποία ασχολούνται με τις στήλες της λέξης, δηλαδή τα κάθετα προφίλ τα οποία είναι:
 - Upper Word Profile
 - Lower Word Profile
 - Lower Minus Upper Word Profile
 - Vertical Projection Profile
 - Vertical Background to Ink Transitions Profile

κανονικοποιούνται διαιρώντας το κάθε διάνυσμα που τα περιέχει με το ύψος της λέξης, εκτός το Vertical Background to Ink Transitions Profile το οποίο διαιρείται με το 6 για να μετατραπούν σε τιμές του διαστήματος [0,1].

Να σημειώσουμε ότι ο λόγος που το Vertical Background to Ink Transitions Profile διαιρείται με το 6 αντί με το ύψος της λέξης, είναι διότι όλες οι μεταβολές από background σε μελάνι για κάθε στήλη της λέξης τουλάχιστον για τα δικά μας σύνολα δεδομένων, δεν υπερβαίνουν τις 6 μεταβολές. Διαιρώντας τα με το ύψος, δεν θα είχαν σχεδόν καθόλου νόημα, μιας και θα είχαμε τιμές κοντά στο 0 όπως 0.03, 0.05, ενώ διαιρώντας τα με το 6 κυμαίνονται σε πιο μεγάλες τιμές όπως 0.3, 0.5 και έτσι αποκτούν περισσότερη αξία.

- Τα χαρακτηριστικά τα οποία ασχολούνται με τις γραμμές της λέξης, δηλαδή τα οριζόντια προφίλ τα οποία είναι:
 - Left Word Profile
 - Right Word Profile
 - Horizontal Projection Profile
 - Horizontal Background to Ink Transitions Profile

κανονικοποιούνται διαιρώντας το κάθε διάνυσμα που τα περιέχει με το πλάτος της λέξης, εκτός το Horizontal Background to Ink Transitions Profile το οποίο διαιρείται με το 30, για να μετατραπούν σε τιμές του διαστήματος [0,1].

Να σημειώσουμε ότι ο λόγος που το Horizontal Background to Ink Transitions Profile διαιρείται με το 30 αντί με το πλάτος της λέξης, είναι ακριβώς ο ίδιος όπως αναφέραμε για το Vertical Background to Ink Transitions Profile.

➤ Τα χαρακτηριστικά τα οποία ασχολούνται με τον αριθμό των pixel δηλαδή τα :

- End Points
- Perimeter Points

κανονικοποιούνται διαιρώντας τα με το πλήθος των pixel της λέξης, όπου το πλήθος των pixel είναι το ύψος επί το πλάτος της λέξης για να μετατραπούν σε τιμές του διαστήματος [0,1].

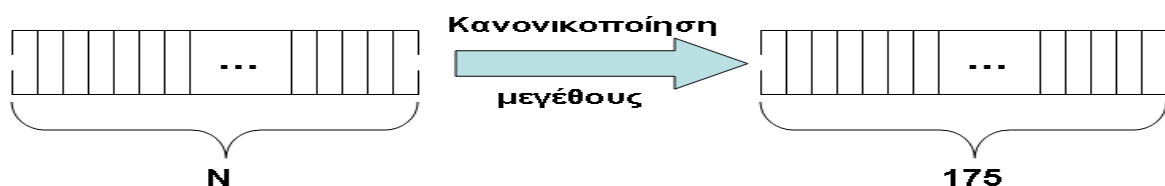
➤ Το χαρακτηριστικό το οποίο ασχολείται με κλειστά σημεία μέσα στη λέξη δηλαδή το:

- Holes Detection

η γραμμή του σημείου του κέντρου των τρυπών κανονικοποιείται διαιρώντας το με το πλήθος των γραμμών της λέξης, ενώ η στήλη του σημείου του κέντρου των τρυπών κανονικοποιείται διαιρώντας το με το πλήθος των στηλών της λέξης, ενώ το μέγεθος κάθε τρύπας κανονικοποιείται διαιρώντας το με το πλήθος των pixel της λέξης για να μετατραπούν σε τιμές του διαστήματος [0,1].

3.2.4 Κανονικοποίηση Μεγέθους (Standardization)

Λόγω του ότι όλες οι εικόνες έχουν διαφορετικό ύψος και πλάτος, δεν μπορούμε να τις διαχειριστούμε χωρίς να υπάρχει ένα σταθερό μέγεθος. Δεν μπορεί να πραγματοποιηθεί η σύγκρισή μεταξύ εικόνων, κατά συνέπεια να μην μπορεί να γίνει η ομαδοποίηση. Η εξαγωγή χαρακτηριστικών που αφορούν κάθετα και οριζόντια προφίλ, αναπαρίστανται ως ένα διάνυσμα μίας γραμμής και N στηλών και μίας γραμμή και M στηλών αντίστοιχα. Για κάθε διαφορετική όμως εικόνα, ο αριθμός N ο οποίος προσδιορίζει το πλάτος της λέξης για τα κάθετα προφίλ και ο αριθμός M ο οποίος προσδιορίζει το ύψος της λέξης για τα οριζόντια προφίλ μεταβάλλεται, κάτι το οποίο εμποδίζει τον υπολογισμό των αποστάσεων (στήλη ανά στήλη) μεταξύ δύο εικόνων κατά την διάρκεια της ομαδοποίησης (Clustering). Σκοπός λοιπόν της κανονικοποίησης μεγέθους είναι να μετατρέψει ένα μεταβαλλόμενο διάνυσμα γραμμής μεγέθους $1 \times N$ (κάθετο προφίλ) ή $1 \times M$ (οριζόντιο προφίλ) σε ένα σταθερό διάνυσμα γραμμής μεγέθους $1 \times K$, όπου K ένα σταθερό μέγεθος στηλών για όλες τις εικόνες. Ως αριθμό κανονικοποίησης μεγέθους K επιλέχθηκε το 175. Έτσι για όλες τις εικόνες, κάθε μεταβαλλόμενο διάνυσμα μεγέθους $1 \times N$ ή $1 \times M$ κανονικοποιείται σε ένα σταθερό μέγεθος 1×175 . Ο τρόπος με τον οποίο γίνεται η κανονικοποίηση μεγέθους έχει ληφθεί από το [24].



Σχήμα 3-20: Παράδειγμα κανονικοποίησης μεγέθους για κάθε εικόνα μεταβαλλόμενου διανύσματος $1 \times N$ (κάθετο προφίλ) σε σταθερό διάνυσμα μεγέθους 1×175 . Ομοίως ισχύει για μεταβαλλόμενο διάνυσμα $1 \times M$ (οριζόντιο προφίλ).

3.3 Matching

Το στάδιο του Matching είναι το στάδιο κατά το οποίο τα χαρακτηριστικά που εξάχθηκαν από κάθε εικόνα λέξης, συγκρίνονται με τα χαρακτηριστικά όλων των υπόλοιπων λέξεων, έτσι ώστε να βρεθούν κάποιες αποστάσεις μεταξύ αυτών. Με τη χρήση αυτών των αποστάσεων θα πραγματοποιηθεί αργότερα το στάδιο της ομαδοποίησης. Η υλοποίηση του σταδίου αυτού, δεν είναι ένα απαραίτητο βήμα στη μέθοδο του Word Spotting, καθώς τα χαρακτηριστικά μπορούν να εισαχθούν απευθείας στον αλγόριθμο ομαδοποίησης. Αν λάβουμε υπόψη ότι το στάδιο του Matching δεν είναι απαραίτητο για την εκτέλεση ενός συστήματος Word Spotting, και ότι το στάδιο αυτό όταν υλοποιείται, υλοποιείται μόνο για σκοπούς ομαδοποίησης, μπορούμε να πούμε ότι αυτό το στάδιο δεν είναι από μόνο του ένα στάδιο του Word Spotting, αλλά ένα προαιρετικό υπό-στάδιο του σταδίου της ομαδοποίησης. Παρόλο που το στάδιο του Matching είναι προαιρετικό, η σημασία του δεν πρέπει να υποτιμάται, καθώς η υλοποίηση ενός τέτοιου σταδίου μπορεί να επιφέρει σημαντικά καλύτερα αποτελέσματα από την μη χρησιμοποίησή του. Φυσικά αυτό εξαρτάται και από την περίπτωση του εκάστοτε προβλήματος.

Είναι σημαντικό να σημειωθεί ότι η σύγκριση μεταξύ όλων των λέξεων δεν είναι πάντα απαραίτητη, καθώς η υλοποίηση του Matching μπορεί να εφαρμοστεί με κάποιο διαφορετικό τρόπο, όπως γίνεται για παράδειγμα στο [14]. Εκεί, μια λέξη συγκρίνεται μόνο με κάποιες επιλεγμένες λέξεις, οι οποίες αντιπροσωπεύουν όλες τις υπόλοιπες λέξεις του συνόλου δεδομένων που κατέχουν παρόμοια χαρακτηριστικά. Επίσης, σε πολλές περιπτώσεις εφαρμόζεται μια διαδικασία κατά την οποία λέξεις που δεν είναι υψηλού ενδιαφέροντος και εμφανίζονται πολλές φορές μέσα στο αρχικό κείμενο (stop words), όπως για παράδειγμα οι λέξεις: “και”, “για”, “το”, δεν συμπεριλαμβάνονται στην εφαρμογή εύρεσης απόστασης από τις υπόλοιπες λέξεις, καθώς αυτό μπορεί να μειώσει σημαντικά τον χρόνο ενός τέτοιου υπολογισμού. Η διαδικασία αυτή καλείται pruning.

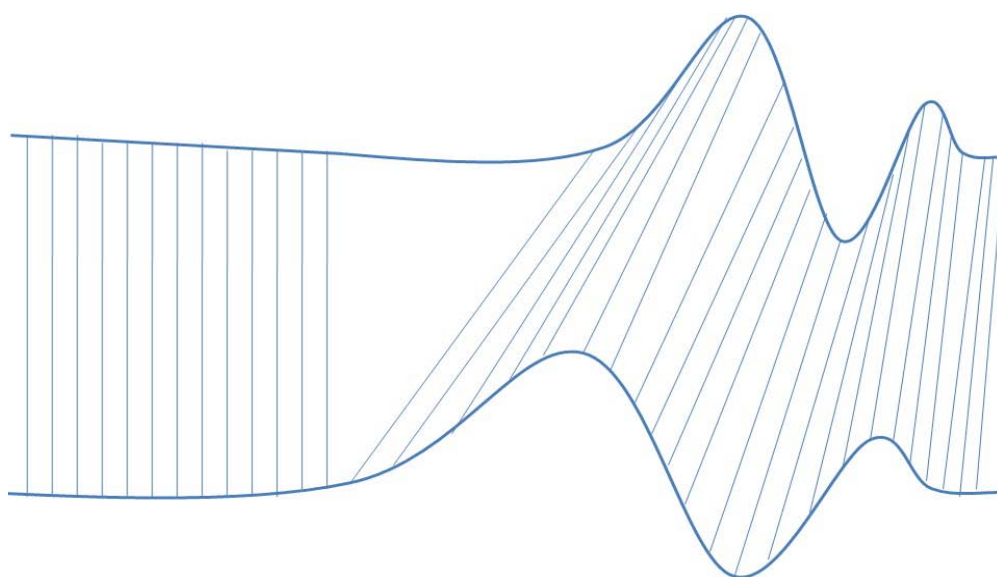
Υπάρχουν πολλοί αλγόριθμοι που μπορούν να χρησιμοποιηθούν στην υλοποίηση του σταδίου Matching υπολογίζοντας συγκεκριμένες αποστάσεις μεταξύ των λέξεων, αναφορικά δίνονται μερικοί από τους πιο γνωστούς:

- 1) XOR: Επιστρέφει το πλήθος των pixel που απομένουν με την αφαίρεση μεταξύ δύο εικόνων.
- 2) Sum of square distances (SSD): Επιστρέφει το ελάχιστο κόστος που υπολογίζεται με βάση το άθροισμα των τετραγωνικών διαφορών των εντάσεων των pixel μεταξύ δύο εικόνων.
- 3) Euclidean distance mapping (EDM): Λειτουργεί με παρόμοιο τρόπο με τη μέθοδο XOR, μόνο που αυτή τη φορά χρησιμοποιούνται κάποια βάρη στον υπολογισμό της απόστασης.
- 4) Corner point correlation (CORR): Η ανίχνευση συγκεκριμένου πλήθους γωνιών στις δύο εικόνες, και η σύγκριση του πλήθους αυτού για κάθε εικόνα καθώς και της περιοχής στην οποία βρέθηκε κάθε γωνιά, χρησιμοποιούνται για τον υπολογισμό της απόστασης μεταξύ δύο εικόνων.
- 5) Dynamic time warping (DTW): Ο αλγόριθμος αυτός θα περιγραφεί αναλυτικά πιο κάτω.

Μια σύγκριση αυτών των πέντε αλγορίθμων, καθώς και δύο άλλων μπορεί να βρεθεί στο [4].

3.3.1 Dynamic Time Warping

Ο αλγόριθμος DTW χρησιμοποιείται ευρέως για τον υπολογισμό της διαφοράς μεταξύ δυο ακολουθιών. Το σημαντικό πλεονέκτημα που προσφέρει αυτός ο αλγόριθμος είναι το γεγονός ότι μπορεί να ανιχνεύσει ομοιότητες μεταξύ δύο χρονοσειρών, οι οποίες έχουν διαφορά φάσης, συχνότητας ή και πλάτους μεταξύ τους, όπως είναι για παράδειγμα η περίπτωση δύο σημάτων φωνής. Ένα παράδειγμα αντιστοίχισης σημάτων όπως θα αναμενόταν να πραγματοποιηθεί με τη χρήση του DTW φαίνεται στο σχήμα 3-21. Ο αλγόριθμος αυτός χρησιμοποιείται σε πάρα πολλές εφαρμογές, όπως είναι για παράδειγμα, η αναγνώριση φωνής, ο υπολογισμός της ομοιότητας μεταξύ δύο βίντεο κ.α. Το γεγονός ότι ο DTW επιτρέπει μια μη γραμμική αντιστοιχία μεταξύ δύο χρονοσειρών, τον κάνει ιδανικό για την περίπτωση της σύγκρισης εικόνων χειρόγραφων λέξεων, αφού είναι σπάνιο το φαινόμενο δύο εικόνες χειρόγραφων λέξεων να είναι πλήρως ευθυγραμμισμένες.



Σχήμα 3-21: Παράδειγμα αντιστοίχισης δύο σημάτων με τον Dynamic Time Warping.

Στη γενική περίπτωση, ο αλγόριθμος DTW λαμβάνει δύο ακολουθίες σαν είσοδο και επιστρέφει τη διαφορά των δύο αυτών ακολουθιών. Ο τρόπος με τον οποίο επιτυγχάνεται αυτός ο υπολογισμός της διαφοράς είναι ο εξής: Αν $X = (x_1, x_2, \dots, x_n)$ είναι ένα διάνυσμα και $Y = (y_1, y_2, \dots, y_m)$ ένα δεύτερο διάνυσμα, τα οποία εισάγονται στον DTW με σκοπό τον υπολογισμό της διαφοράς τους, τότε αρχικά κατασκευάζεται ένας πίνακας D , διαστάσεων $n \times m$, στον οποίο κάθε στοιχείο του $D(i, j)$, όπου $i = 2, \dots, n$ και $j = 2, \dots, m$ περιέχει το άθροισμα μιας διαφοράς $d(x_i, y_j)$ συν την ελάχιστη τιμή μεταξύ των $[D(i-1, j-1), D(i-1, j), D(i, j-1)]$. Στο πρώτο στοιχείο του D , δηλαδή το $D(1, 1)$ περιέχεται μόνο η τιμή της διαφοράς $d(x_1, y_1)$ ενώ κάθε στοιχείο της πρώτης γραμμής υπολογίζεται από τον τύπο:

$$\triangleright D(i, 1) = D(i - 1, 1) + d(x_i, y_1), \text{ όπου } i = 2, \dots, n$$

Και κάθε στοιχείο της πρώτης στήλης υπολογίζεται από τον τύπο:

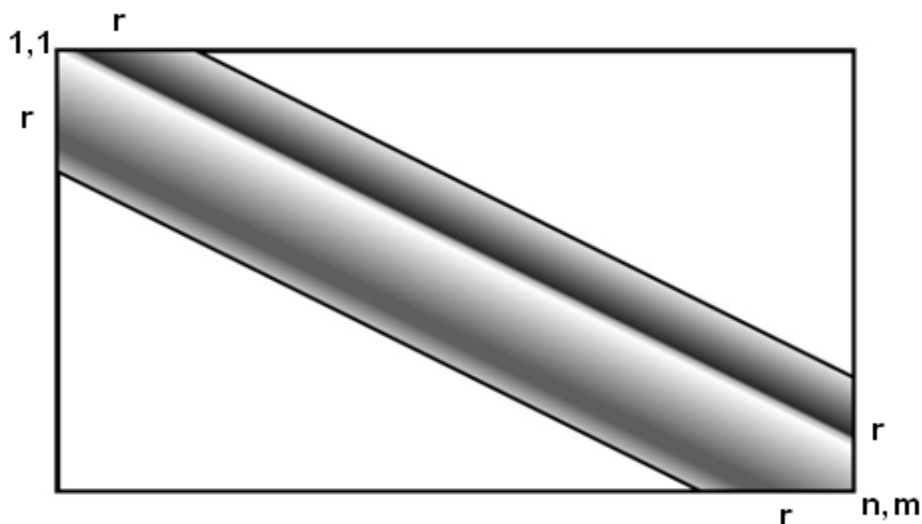
$$\triangleright D(1, j) = D(1, j - 1) + d(x1, yj), \text{ όπου } j = 2, \dots, m$$

Αφού ο πίνακας D έχει κατασκευαστεί, το επόμενο βήμα είναι ο εντοπισμός του μονοπατιού (warping path), το οποίο συνήθως υπολογίζεται ξεκινώντας από το σημείο D(n,m) και καταλήγοντας στο D(1,1), προχωρώντας κάθε φορά που βρισκόμαστε στο D(i,j) στο $\min[D(i-1, j-1), D(i-1, j), D(i, j-1)]$, όπου $i = 1, \dots, n$ και $j = 1, \dots, m$. Η τελική απόσταση που θα επιστραφεί από τον αλγόριθμο είναι η διαίρεση του στοιχείου D(n,m) με το πλήθος των βημάτων του μονοπατιού.

Η διαφορά d μπορεί να είναι οποιαδήποτε διαφορά όπως π.χ. η ευκλείδεια απόσταση ή η απόσταση συνημίτονων, αναλόγως με το πρόβλημα που καλούμαστε να αντιμετωπίσουμε. Στην περίπτωση που η είσοδος είναι δύο πίνακες αντί δύο διανύσματα, τότε η απόσταση d μπορεί να δημιουργηθεί έτσι ώστε να υπολογίζει και να επιστρέφει μια απόσταση μεταξύ των δύο συγκεκριμένων στηλών i και j των δύο πινάκων X και Y για όλες τις γραμμές των πινάκων αυτών, όπως φαίνεται για παράδειγμα από τον πιο κάτω τύπο:

$$\triangleright d(xi, yj) = \sum_{k=1}^L (x_{k,i} - y_{k,j})^2, \text{ όπου } L \text{ είναι το πλήθος των γραμμών των δύο πινάκων.}$$

Ο περιορισμός Sakoe-Chiba band [25], μπορεί να εφαρμοστεί στην υλοποίηση του αλγόριθμου DTW με τον εξής τρόπο: κατά τον υπολογισμό του μονοπατιού, τα στοιχεία του D που περιλαμβάνονται στο μονοπάτι πρέπει να βρίσκονται μέσα στην σκιασμένη περιοχή που φαίνεται στο σχήμα 3-22. Το r μπορεί να πάρει οποιαδήποτε ακέραια τιμή, η οποία παριστάνει την απόσταση από την κύρια διαγώνιο του D. Η επιλογή της τιμής αυτής πρέπει να γίνει μετά από διάφορα πειράματα έτσι ώστε να επιλεγθεί αυτή που δίνει τα καλύτερα αποτελέσματα. Στη περίπτωση μας το r είναι ίσο με 10. Η χρήση ενός τέτοιου περιορισμού έχει αποδειχθεί ότι σε πολλές περιπτώσεις επιφέρει καλύτερα αποτελέσματα αλλά και μειώνει σημαντικά το χρόνο υπολογισμού της απόστασης από τον αλγόριθμο DTW.



Σχήμα 3-22: Ο περιορισμός Sakoe-Chiba band.

3.4 Διακριτοί Μετασχηματισμοί

3.4.1 Διακριτός Μετασχηματισμός Fourier (Discrete Fourier Transform - DFT)

Ο διακριτός μετασχηματισμός Fourier είναι είδος διακριτού μετασχηματισμού, ο οποίος ασχολείται με διακριτά πεδία και έχει εφαρμογή στην ανάλυση Fourier. Σκοπός του είναι να μετασχηματίζει μία συνάρτηση από το πεδίο του χρόνου στο πεδίο της συχνότητας. Ο διακριτός μετασχηματισμός Fourier δέχεται ως είσοδο μία συνάρτηση η οποία είναι διακριτή, δηλαδή περιέχει μη μηδενικές τιμές που έχουν περιορισμένη διάρκεια, υπάρχει δηλαδή τέλος σε αυτή την διάρκεια. Τέτοιου είδους συναρτήσεις δημιουργούνται συνήθως με δειγματοληψία μίας συνεχούς συνάρτησης. Η είσοδος του διακριτού μετασχηματισμού Fourier είναι μία πεπερασμένη ακολουθία, δηλαδή υπάρχει τέλος σε αυτή την ακολουθία πραγματικών ή μιγαδικών αριθμών. Στην περίπτωση μας ο διακριτός μετασχηματισμός Fourier δέχεται ως είσοδο ένα κανονικοποιημένο διάνυμα γραμμής για κάθε οριζόντιο και κάθετο προφίλ με διακριτό μέγεθος μίας γραμμής και 175 στηλών, ενώ για τον αριθμό των τελικών σημείων (End Points) και περιμετρικών σημείων (Perimeter Points) δέχεται ένα κανονικοποιημένο διάνυμα μίας γραμμής και μίας στήλης. Ο διακριτός μετασχηματισμός Fourier (DFT) μπορεί να υπολογιστεί αποτελεσματικά κάνοντας χρήση ενός αλγόριθμου ταχύ μετασχηματισμού Fourier (Fast Fourier Transform – FFT).

Ο υπολογισμός μίας ακολουθίας N αριθμών x_0, x_1, \dots, x_{N-1} μετασχηματίζεται στο πεδίο της συχνότητας σε N μιγαδικούς αριθμούς X_0, X_1, \dots, X_{N-1} από την παρακάτω εξίσωση DFT :

$$X_k = \sum_{n=0}^{N-1} x_n e^{-\frac{2\pi i}{N}kn} \quad k = 0, \dots, N-1 \quad [26]$$

ή ως εξής:

$$X_k = \sum_{n=0}^{N-1} x_n \left[\cos\left(\frac{-2\pi kn}{N}\right) + i \sin\left(\frac{-2\pi kn}{N}\right) \right] \quad k = 0, \dots, N-1$$

3.4.1.1 Παράδειγμα Διακριτού Μετασχηματισμού Fourier

Δίνουμε ένα παράδειγμα στη συνέχεια για να καταλάβουμε καλύτερα την λειτουργία του μετασχηματισμού Fourier στο πεδίο της συχνότητας.

Έχουμε για παράδειγμα την ακολουθία $[x_0 = 1, x_1 = 2, x_2 = 3]$ η οποία είναι πεπερασμένη και αποτελείται από 3 αριθμούς δηλαδή έχουμε $N=3$.

Για να υπολογίσουμε τους συντελεστές X_0, X_1, X_2 με βάση DFT στο πεδίο της συχνότητας έχουμε ως εξής:

$$\begin{aligned}
\Gamma_{\alpha} k = 0 : X_0 &= \sum_{n=0}^2 x_n \left[\cos\left(\frac{-2\pi 0n}{3}\right) + i \sin\left(\frac{-2\pi 0n}{3}\right) \right] = \\
&= x_0 \left[\cos\left(\frac{-2\pi * 0 * 0}{3}\right) + i \sin\left(\frac{-2\pi * 0 * 0}{3}\right) \right] + x_1 \left[\cos\left(\frac{-2\pi * 0 * 1}{3}\right) + i \sin\left(\frac{-2\pi * 0 * 1}{3}\right) \right] + \\
&x_2 \left[\cos\left(\frac{-2\pi * 0 * 2}{3}\right) + i \sin\left(\frac{-2\pi * 0 * 2}{3}\right) \right] = \\
&= 1 * \left[\cos\left(\frac{-2\pi * 0 * 0}{3}\right) + i \sin\left(\frac{-2\pi * 0 * 0}{3}\right) \right] + 2 * \left[\cos\left(\frac{-2\pi * 0 * 1}{3}\right) + i \sin\left(\frac{-2\pi * 0 * 1}{3}\right) \right] + \\
&3 * \left[\cos\left(\frac{-2\pi * 0 * 2}{3}\right) + i \sin\left(\frac{-2\pi * 0 * 2}{3}\right) \right] = \\
&= 1 * [\cos(0) + i \sin(0)] + 2 * [\cos(0) + i \sin(0)] + 3 * [\cos(0) + i \sin(0)] = \\
&= 1 * [1 + 0] + 2 * [1 + 0] + 3 * [1 + 0] = 1 + 2 + 3 = 6 \\
\Rightarrow X_0 &= 6 + 0i
\end{aligned}$$

$$\begin{aligned}
\Gamma_{\alpha} k = 1 : X_1 &= \sum_{n=0}^2 x_n \left[\cos\left(\frac{-2\pi 1n}{3}\right) + i \sin\left(\frac{-2\pi 1n}{3}\right) \right] = \\
&= x_0 \left[\cos\left(\frac{-2\pi * 1 * 0}{3}\right) + i \sin\left(\frac{-2\pi * 1 * 0}{3}\right) \right] + x_1 \left[\cos\left(\frac{-2\pi * 1 * 1}{3}\right) + i \sin\left(\frac{-2\pi * 1 * 1}{3}\right) \right] + \\
&x_2 \left[\cos\left(\frac{-2\pi * 1 * 2}{3}\right) + i \sin\left(\frac{-2\pi * 1 * 2}{3}\right) \right] = \\
&= 1 * \left[\cos\left(\frac{-2\pi * 1 * 0}{3}\right) + i \sin\left(\frac{-2\pi * 1 * 0}{3}\right) \right] + 2 * \left[\cos\left(\frac{-2\pi * 1 * 1}{3}\right) + i \sin\left(\frac{-2\pi * 1 * 1}{3}\right) \right] + \\
&3 * \left[\cos\left(\frac{-2\pi * 1 * 2}{3}\right) + i \sin\left(\frac{-2\pi * 1 * 2}{3}\right) \right] = \\
&= 1 * [\cos(0) + i \sin(0)] + 2 * \left[\cos\left(\frac{-2\pi}{3}\right) + i \sin\left(\frac{-2\pi}{3}\right) \right] + 3 * \left[\cos\left(\frac{-4\pi}{3}\right) + i \sin\left(\frac{-4\pi}{3}\right) \right] = \\
&= 1 * [1 + 0] + 2 * [-0.5 - 0.866i] + 3 * [-0.5 + 0.866i] = 1 - 1 - 1.732i - 1.5 + 2.6i = \\
&= -1.5 + 0.868i \\
\Rightarrow X_1 &= -1.5 + 0.868i
\end{aligned}$$

$$\begin{aligned}
\text{Για } k = 2 : X_2 &= \sum_{n=0}^2 x_n \left[\cos\left(\frac{-2\pi 2n}{3}\right) + i \sin\left(\frac{-2\pi 2n}{3}\right) \right] = \\
&= x_0 \left[\cos\left(\frac{-2\pi * 2 * 0}{3}\right) + i \sin\left(\frac{-2\pi * 2 * 0}{3}\right) \right] + x_1 \left[\cos\left(\frac{-2\pi * 2 * 1}{3}\right) + i \sin\left(\frac{-2\pi * 2 * 1}{3}\right) \right] + \\
&x_2 \left[\cos\left(\frac{-2\pi * 2 * 2}{3}\right) + i \sin\left(\frac{-2\pi * 2 * 2}{3}\right) \right] = \\
&= 1 * \left[\cos\left(\frac{-2\pi * 2 * 0}{3}\right) + i \sin\left(\frac{-2\pi * 2 * 0}{3}\right) \right] + 2 * \left[\cos\left(\frac{-2\pi * 2 * 1}{3}\right) + i \sin\left(\frac{-2\pi * 2 * 1}{3}\right) \right] + \\
&3 * \left[\cos\left(\frac{-2\pi * 2 * 2}{3}\right) + i \sin\left(\frac{-2\pi * 2 * 2}{3}\right) \right] = \\
&= 1 * \left[\cos(0) + i \sin(0) \right] + 2 * \left[\cos\left(\frac{-4\pi}{3}\right) + i \sin\left(\frac{-4\pi}{3}\right) \right] + 3 * \left[\cos\left(\frac{-8\pi}{3}\right) + i \sin\left(\frac{-8\pi}{3}\right) \right] = \\
&= 1 * [1 + 0] + 2 * [-0.5 + 0.866i] + 3 * [-0.5 - 0.866i] = 1 - 1 + 1.732i - 1.5 - 2.6i = \\
&= -1.5 - 0.868i \\
\Rightarrow X_2 &= -1.5 - 0.868i
\end{aligned}$$

Αρα όπως βλέπουμε και από το παράδειγμα, τα αποτελέσματα που δίνει ο διακριτός μετασχηματισμός Fourier στο πεδίο της συχνότητας είναι μιγαδικοί αριθμοί, οι οποίοι έχουν την μορφή $X = A + Bi$ με πραγματικό μέρος το A και φανταστικό μέρος το B .

Ο μετασχηματισμός Fourier εφαρμόζεται για κάθε ένα από τα χαρακτηριστικά που έχουν εξαχθεί και κανονικοποιηθεί. Κατά το στάδιο όμως της ομαδοποίησης πρέπει να απαλείφουν πρώτα τα μιγαδικά για να εκτελεστούν με επιτυχία οι συγκρίσεις μεταξύ λέξεων. Για το σκοπό αυτό, από τα μετασχηματισμένα αποτελέσματα που παίρνουμε, υπολογίζεται η απόλυτη τιμή τους και το αποτέλεσμα της απόλυτης τιμής αυτής χρησιμοποιείται ως είσοδος στο στάδιο της ομαδοποίησης. Η απόλυτη τιμή κάθε μετασχηματισμένου στοιχείου X του προφίλ (1x175) για τα οριζόντια και κάθετα προφίλ, και κάθε μετασχηματισμένου διανύσματος (1x1) για τα περιμετρικά και τελικά σημεία υπολογίζεται ως εξής :

$$|X_k| = \sqrt{\text{Re}(X_k)^2 + \text{Im}(X_k)^2} \quad [26]$$

Όπου για το παράδειγμα μας πιο πάνω θα έχουμε:

$$X_0 = 6 + 0i \Rightarrow |X_0| = \sqrt{6^2 + 0^2} = 6$$

$$X_1 = -1.5 + 0.868i \Rightarrow |X_1| = \sqrt{(-1.5)^2 + 0.868^2} = 1.733$$

$$X_2 = -1.5 - 0.868i \Rightarrow |X_2| = \sqrt{(-1.5)^2 + (-0.868)^2} = 1.733$$

3.4.2 Διακριτός Μετασχηματισμός Συνημίτονου (Discrete Cosine Transform – DCT)

Ο διακριτός μετασχηματισμός συνημίτονου (DCT) είναι συναφής μετασχηματισμός με τον διακριτό μετασχηματισμό Fourier (DFT) και χρησιμοποιείται μόνο για πραγματικούς αριθμούς. Διατυπώνει μία πεπερασμένη ακολουθία πολλών σημείων μέσω ενός αθροίσματος συναρτήσεων από συνημίτονα, ταλαντεύοντας τα σε διαφορετικές συχνότητες. Ο διακριτός μετασχηματισμός συνημίτονου έχει πολλές εφαρμογές στον κλάδο της επιστήμης και της μηχανικής, όπως είναι η συμπίεση ήχου και εικόνας καθώς επίσης, χρησιμοποιείται για την επίλυση μερικών διαφορικών εξισώσεων. Η χρήση του συνημίτονου (Cosine) αντί του ημιτόνου (Sine), είναι καθοριστικής σημασίας, όπως για παράδειγμα στη συμπίεση, αποδεικνύεται ότι οι συναρτήσεις συνημίτονων είναι πολύ πιο αποτελεσματικές μιας και χρειάζονται πιο λίγες για την προσέγγιση ενός τυπικού σήματος, αφού οι μικρές υψηλές συχνότητες μπορούν εύκολα να απορρίπτονται, ενώ για τις διαφορικές εξισώσεις τα συνημίτονα εκφράζουν συγκεκριμένη επιλογή των οριακών τιμών. Ο διακριτός μετασχηματισμός συνημίτονου (DCT) είναι ισοδύναμος με τον διακριτό μετασχηματισμό Fourier (DFT) με περίπου διπλάσιο μήκος, που λειτουργεί σε πραγματικά δεδομένα με άρτια συμμετρία. Από την άλλη, ο διακριτός μετασχηματισμός ημιτόνου (DST), είναι και αυτός ισοδύναμος με τον διακριτό μετασχηματισμό Fourier (DFT) για πραγματικές και περιττές συναρτήσεις. Η προφανής διαφορά μεταξύ DCT και DFT, είναι ότι η DCT χρησιμοποιεί μόνο συναρτήσεις συνημίτονου ενώ η DFT χρησιμοποιεί συναρτήσεις και ημιτόνων και συνημίτονων, οι οποίες αναπαρίστανται σε μορφή μιγαδικού εκθετικού όπως έχουμε δει. Από αυτή την εμφανές διαφορά φαίνεται εύκολα ότι ο DCT υποδηλώνει διαφορετικές οριακές τιμές σε σχέση με τον DFT και άλλων συναφών μετασχηματισμών. Ο DFT όπως για παράδειγμα η σειρά Fourier, υποδηλώνει μία περιοδική επέκταση της αρχικής συνάρτησης, ενώ ο DCT υποδηλώνει μία άρτια επέκταση της αρχικής συνάρτησης. Υπάρχουν οχτώ παραλλαγές DCT, εκ των οποίων οι τέσσερις είναι κοινές. Η πιο κοινή παραλλαγή διακριτού μετασχηματισμού συνημίτονου (DCT) είναι η τύπου II-DCT η οποία απλά ονομάζεται “DCT” ενώ η τύπου III-DCT είναι η αντιστροφή (inverse) της DCT, η οποία ονομάζεται απλά Inverse-DCT ή πιο απλά “IDCT”.

Στην περίπτωση μας ο διακριτός μετασχηματισμός συνημίτονου δέχεται ως είσοδο ένα κανονικοποιημένο διάνυσμα γραμμής για κάθε οριζόντιο και κάθετο προφίλ με διακριτό μέγεθος μίας γραμμής και 175 στηλών, ενώ για τον αριθμό των τελικών σημείων (End Points) και περιμετρικών σημείων (Perimeter Points), δέχεται ένα κανονικοποιημένο διάνυσμα μίας γραμμής και μίας στήλης.

Ο υπολογισμός μίας ακολουθίας N αριθμών x_0, x_1, \dots, x_{N-1} μετασχηματίζεται στο πεδίο της συχνότητας σε N πραγματικούς αριθμούς X_0, X_1, \dots, X_{N-1} από την παρακάτω εξίσωση DCT :

$$X_k = \sum_{n=0}^{N-1} x_n \cos\left[\frac{\pi}{N}\left(n + \frac{1}{2}\right)k\right] \quad k = 0, \dots, N-1 \quad [27]$$

3.4.2.1 Παράδειγμα Διακριτού Μετασχηματισμού Συνημίτονου

Δίνουμε ένα παράδειγμα στη συνέχεια για να καταλάβουμε καλύτερα την λειτουργία του μετασχηματισμού συνημίτονου στο πεδίο της συχνότητας.

Έχουμε για παράδειγμα την ακολουθία $[x_0 = 1, x_1 = 2, x_2 = 3]$ η οποία είναι πεπερασμένη και αποτελείται από 3 αριθμούς δηλαδή έχουμε $N=3$.

Για να υπολογίσουμε τους συντελεστές X_0, X_1, X_2 με βάση DCT στο πεδίο της συχνότητας έχουμε ως εξής:

$$\begin{aligned}\text{Για } k=0 : X_0 &= \sum_{n=0}^2 x_n \cos\left[\frac{\pi}{3}\left(n + \frac{1}{2}\right) * 0\right] = \\ &= x_0 \cos\left[\frac{\pi}{3}\left(0 + \frac{1}{2}\right) * 0\right] + x_1 \cos\left[\frac{\pi}{3}\left(1 + \frac{1}{2}\right) * 0\right] + x_2 \cos\left[\frac{\pi}{3}\left(2 + \frac{1}{2}\right) * 0\right] = \\ &= 1 * \cos\left[\frac{\pi}{3}\left(0 + \frac{1}{2}\right) * 0\right] + 2 * \cos\left[\frac{\pi}{3}\left(1 + \frac{1}{2}\right) * 0\right] + 3 * \cos\left[\frac{\pi}{3}\left(2 + \frac{1}{2}\right) * 0\right] = \\ &= 1 * \cos[0] + 2 * \cos[0] + 3 * \cos[0] = 1 + 2 + 3 = 6 \\ &\Rightarrow X_0 = 6\end{aligned}$$

$$\begin{aligned}\text{Για } k=1 : X_1 &= \sum_{n=0}^2 x_n \cos\left[\frac{\pi}{3}\left(n + \frac{1}{2}\right) * 1\right] = \\ &= x_0 \cos\left[\frac{\pi}{3}\left(0 + \frac{1}{2}\right) * 1\right] + x_1 \cos\left[\frac{\pi}{3}\left(1 + \frac{1}{2}\right) * 1\right] + x_2 \cos\left[\frac{\pi}{3}\left(2 + \frac{1}{2}\right) * 1\right] = \\ &= 1 * \cos\left[\frac{\pi}{3}\left(0 + \frac{1}{2}\right) * 1\right] + 2 * \cos\left[\frac{\pi}{3}\left(1 + \frac{1}{2}\right) * 1\right] + 3 * \cos\left[\frac{\pi}{3}\left(2 + \frac{1}{2}\right) * 1\right] = \\ &= 1 * \cos\left[\frac{\pi}{6}\right] + 2 * \cos\left[\frac{3\pi}{6}\right] + 3 * \cos\left[\frac{5\pi}{6}\right] = 0.866 + 0 - 2.6 = -1.734 \\ &\Rightarrow X_1 = -1.734\end{aligned}$$

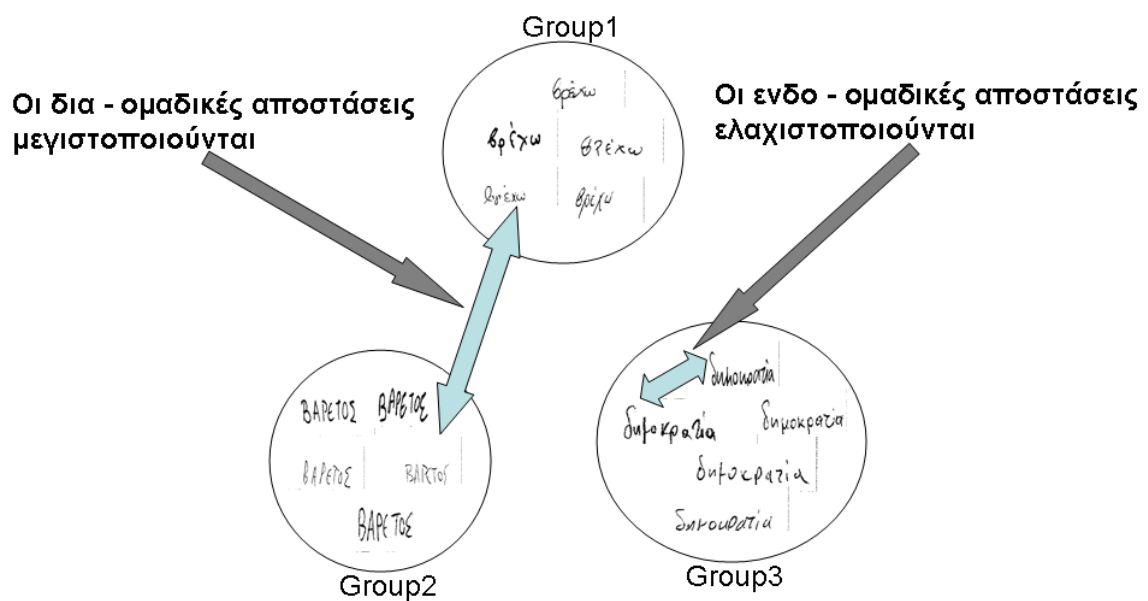
$$\begin{aligned}\text{Για } k=2 : X_2 &= \sum_{n=0}^2 x_n \cos\left[\frac{\pi}{3}\left(n + \frac{1}{2}\right) * 2\right] = \\ &= x_0 \cos\left[\frac{\pi}{3}\left(0 + \frac{1}{2}\right) * 2\right] + x_1 \cos\left[\frac{\pi}{3}\left(1 + \frac{1}{2}\right) * 2\right] + x_2 \cos\left[\frac{\pi}{3}\left(2 + \frac{1}{2}\right) * 2\right] =\end{aligned}$$

$$\begin{aligned}
&= 1 * \cos\left[\frac{\pi}{3}\left(0 + \frac{1}{2}\right) * 2\right] + 2 * \cos\left[\frac{\pi}{3}\left(1 + \frac{1}{2}\right) * 2\right] + 3 * \cos\left[\frac{\pi}{3}\left(2 + \frac{1}{2}\right) * 2\right] = \\
&= 1 * \cos\left[\frac{\pi}{3}\right] + 2 * \cos[\pi] + 3 * \cos\left[\frac{5\pi}{3}\right] = 0.5 - 2 + 1.5 = 0 \\
&\Rightarrow X_2 = 0
\end{aligned}$$

Ο διακριτός μετασχηματισμός συνημίτονου εφαρμόζεται για κάθε ένα από τα χαρακτηριστικά που έχουν εξαχθεί και κανονικοποιηθεί. Όπως βλέπο με και από το παράδειγμα, τα αποτελέσματα που δίνει ο διακριτός μετασχηματισμός συνημίτονου στο πεδίο της συχνότητας, είναι πραγματικοί αριθμοί, οι οποίοι χρησιμοποιούνται κατευθείαν ως έχουν στο στάδιο της ομαδοποίησης για την σύγκριση μεταξύ λέξεων.

3.5 Ομαδοποίηση (Clustering)

Το στάδιο της ομαδοποίησης αποτελεί την δημιουργία ομάδων από λέξεις, τέτοιων ώστε οι λέξεις μίας ομάδας να είναι όμοιες ή συσχετιζόμενες μεταξύ τους, και διαφορετικές ή μη συσχετιζόμενες από τις λέξεις άλλων ομάδων. Με λίγα λόγια οι ενδο-ομαδικές αποστάσεις να ελαχιστοποιούνται και οι δια-ομαδικές αποστάσεις να μεγιστοποιούνται όπως φαίνεται στο σχήμα 3-23.

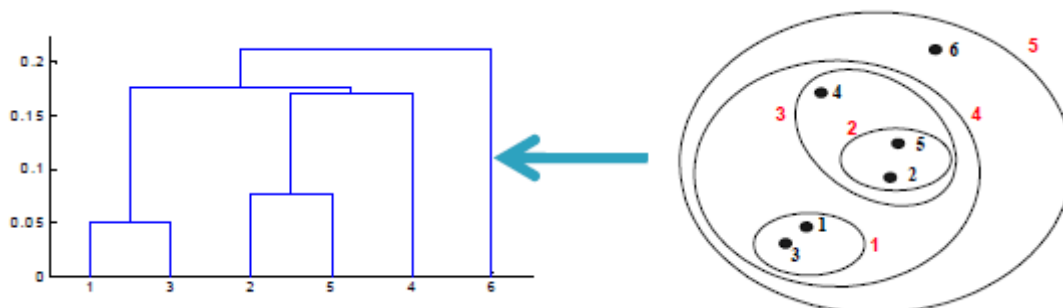


Σχήμα 3-23: Παράδειγμα δια-ομαδικών και ενδο-ομαδικών αποστάσεων τριών ομάδων.

Υπάρχουν δύο βασικές προσεγγίσεις ομαδοποίησης οι οποίες ακολουθούν ιεραρχική προσέγγιση ή διαχωριστική προσέγγιση. Η ιεραρχική ομαδοποίηση (Hierarchical Clustering) περιέχει ένα σύνολο από εμφωλευμένες ομάδες και επιτρέπει σε μία ομάδα να έχει υποομάδες οργανωμένες σε ένα ιεραρχικό δέντρο. Η διαχωριστική ομαδοποίηση (Partitional Clustering) περιέχει ένα

διαμερισμό των αντικειμένων σε μη επικαλυπτόμενα υποσύνολα, τέτοιο ώστε κάθε αντικείμενο να ανήκει ακριβώς σε ένα υποσύνολο. [28]

Εμείς θα ασχοληθούμε με την ιεραρχική ομαδοποίηση (Hierarchical Clustering). Η ιεραρχική ομαδοποίηση παράγει ένα σύνολο εμφωλιασμένων ομάδων που οργανώνονται σαν ένα ιεραρχικό δέντρο. Οπτικοποιούνται σαν ένα δενδρόγραμμα, δηλαδή ένα διάγραμμα που μοιάζει με δέντρο, και καταγράφει τις ακολουθίες από συγχωνεύσεις (merges) και διαχωρισμούς (splits). [28]



Σχήμα 3-24: [28] Παράδειγμα οπτικοποίησης 5 ομάδων ως ένα δενδρόγραμμα.

Η ιεραρχική ομαδοποίηση αποτελείται από δύο διαφορετικές προσεγγίσεις. Η μία είναι η συγκεντρωτική προσέγγιση (Agglomerative) και η άλλη είναι η Διαιρετική (Divisive). Η συγκεντρωτική προσέγγιση ξεκινά με κάθε σημείο (στην περίπτωσή μας λέξη) ως μία ξεχωριστή ομάδα (φύλλο του δέντρου), και σε κάθε βήμα συγχωνεύει το πιο κοντινό ζευγάρι ομάδων μέχρι να μείνει μόνο μία ομάδα (ρίζα του δέντρου με όλες τις λέξεις μέσα). Σε αντίθετη περίπτωση η διαιρετική προσέγγιση ξεκινά με μία περιεκτική ομάδα (ρίζα του δέντρου με όλες τις λέξεις μέσα) και σε κάθε βήμα διαιρείται η ομάδα εως ότου κάθε ομάδα να περιέχει ένα σημείο (μία λέξη σε κάθε ένα από τα φύλλα του δέντρου).

Τα πλεονεκτήματα της ιεραρχικής ομαδοποίησης είναι ότι δεν χρειάζεται να υποθέτουμε ένα συγκεκριμένο αριθμό από ομάδες, οποιοσδήποτε επιθυμητός αριθμός από ομάδες μπορεί να επιτευχθεί κόβοντας το δενδρόγραμμα στο κατάλληλο επίπεδο. Επίσης μπορούν να αντιστοιχούν σε εύλογες ταξινομήσεις. [28]

3.5.1 Ιεραρχικοί αλγόριθμοι ομαδοποίησης (Hierarchical Clustering Algorithms)

Αλγόριθμοι που χρησιμοποιούν ιεραρχική ομαδοποίηση και τους οποίους θα δούμε στη συνέχεια είναι οι:

- Single Linkage
- Complete Linkage
- Average Linkage
- Centroid Linkage
- Ward Linkage

Αναφορικά εκτός από τους πέντε πιο πάνω αλγόριθμους που χρησιμοποιούν ιεραρχική ομαδοποίηση τους οποίους θα δούμε και στη συνέχεια, υπάρχουν στο χώρο και οι Median Linkage και Weighted Linkage.

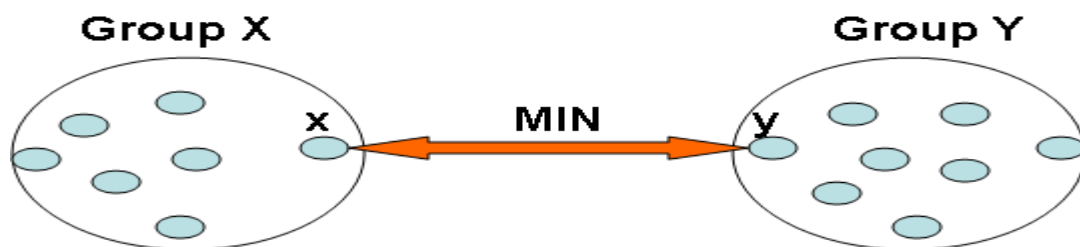
3.5.1.1 Single Linkage

Στον αλγόριθμο Single Linkage η απόσταση (ή αλλιώς μέτρο ομοιότητας) μεταξύ δύο ομάδων, υπολογίζεται ως η απόσταση μεταξύ των δύο πιο όμοιων (κοντινότερων) σημείων στις δύο αυτές ομάδες. Η απόσταση αυτή ονομάζεται και ως απόσταση πλησιέστερου γείτονα ή ελάχιστη απόσταση μεταξύ δύο ομάδων.

Μαθηματικά η απόσταση $D(X, Y)$ μεταξύ ομάδων X και Y ορίζεται ως εξής:

$$D(X, Y) = \min_{x \in X, y \in Y} d(x, y) \quad [29]$$

όπου X και Y είναι οποιεσδήποτε δύο ομάδες που περιέχουν τα σημεία x και y αντίστοιχα και $d(x, y)$ είναι η ελάχιστη απόσταση μεταξύ των δύο ομάδων.



Σχήμα 3-25: Μέτρο ομοιότητας του ιεραρχικού αλγόριθμου Single Linkage.

3.5.1.2 Complete Linkage

Στον αλγόριθμο Complete Linkage η απόσταση (ή αλλιώς μέτρο ομοιότητας) μεταξύ δύο ομάδων, υπολογίζεται ως η απόσταση μεταξύ των δύο λιγότερο όμοιων (μακρινότερων) σημείων στις δύο αυτές ομάδες. Η απόσταση αυτή ονομάζεται και ως απόσταση απομακρυσμένου γείτονα ή μέγιστη απόσταση μεταξύ δύο ομάδων.

Μαθηματικά η απόσταση $D(X, Y)$ μεταξύ ομάδων X και Y ορίζεται ως εξής:

$$D(X, Y) = \max_{x \in X, y \in Y} d(x, y) \quad [30]$$

όπου X και Y είναι οποιεσδήποτε δύο ομάδες που περιέχουν τα σημεία x και y αντίστοιχα και $d(x, y)$ είναι η μέγιστη απόσταση μεταξύ των δύο ομάδων.



Σχήμα 3-26: Μέτρο ομοιότητας του ιεραρχικού αλγόριθμου Complete Linkage.

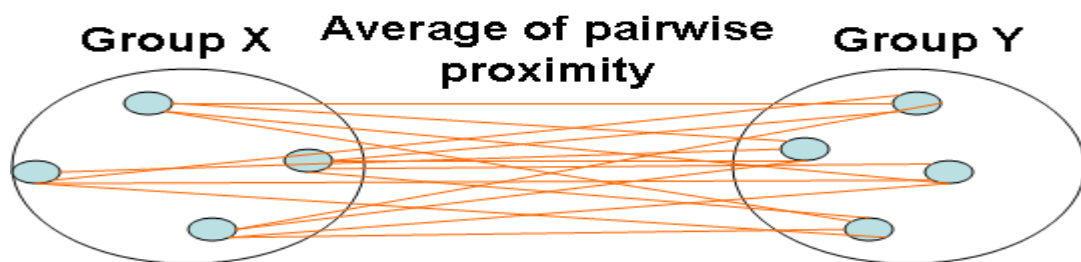
3.5.1.3 Average Linkage

Στον αλγόριθμο Average Linkage η απόσταση (ή αλλιώς μέτρο ομοιότητας) μεταξύ δύο ομάδων, είναι η μέση απόσταση μεταξύ όλων των σημείων ανά ζεύγος στις δύο αυτές ομάδες.

Μαθηματικά η απόσταση $D(X, Y)$ μεταξύ ομάδων X και Y ορίζεται ως εξής:

$$D(X, Y) = \frac{1}{N_X N_Y} \sum_{i=1}^{N_X} \sum_{j=1}^{N_Y} d(x_{Xi}, x_{Yj}) \quad [31]$$

όπου X και Y είναι οποιεσδήποτε δύο ομάδες, $d(x_{Xi}, x_{Yj})$ είναι η ανά ζεύγος απόσταση μεταξύ όλων των σημείων των δύο ομάδων, N_X είναι ο συνολικός αριθμός των σημείων της ομάδας X και N_Y είναι ο συνολικός αριθμός των σημείων της ομάδας Y .



Σχήμα 3-27: Μέτρο ομοιότητας του ιεραρχικού αλγόριθμου Average Linkage.

3.5.1.4 Centroid Linkage

Στον αλγόριθμο Centroid Linkage η απόσταση (ή αλλιώς μέτρο ομοιότητας) μεταξύ δύο ομάδων, είναι η απόσταση μεταξύ των κέντρων για τις δύο αυτές ομάδες. Χρησιμοποιεί την ευκλείδεια απόσταση μεταξύ των κέντρων των δύο ομάδων.

Μαθηματικά η απόσταση $D(X, Y)$ μεταξύ ομάδων X και Y ορίζεται ως εξής:

$$D(X, Y) = \|\bar{x}_X - \bar{x}_Y\|_2 \quad \text{όπου} \quad \bar{x}_X = \frac{1}{N_X} \sum_{i=1}^{N_X} x_{Xi} \quad \text{και} \quad \bar{x}_Y = \frac{1}{N_Y} \sum_{j=1}^{N_Y} x_{Yj} \quad [31]$$

όπου X και Y είναι οποιεσδήποτε δύο ομάδες, $\|\bar{x}_X - \bar{x}_Y\|_2$ είναι η ευκλείδεια απόσταση μεταξύ των κέντρων των δύο ομάδων, \bar{x}_X είναι το κέντρο της ομάδας X , \bar{x}_Y είναι το κέντρο της ομάδας Y , N_X είναι ο συνολικός αριθμός των σημείων της ομάδας X και N_Y είναι ο συνολικός αριθμός των σημείων της ομάδας Y .



Σχήμα 3-28: Μέτρο ομοιότητας του ιεραρχικού αλγόριθμου Centroid Linkage.

3.5.1.5 Ward Linkage

Στον αλγόριθμο Ward Linkage η απόσταση (ή αλλιώς μέτρο ομοιότητας) μεταξύ δύο ομάδων, είναι η απόσταση των τετραγωνικών λαθών για τις δύο αυτές ομάδες. Χρησιμοποιεί το στοιχειώδες άθροισμα των τετραγώνων. Το άθροισμα των τετραγώνων της εντός ομάδας, ορίζεται ως το άθροισμα των τετραγώνων των αποστάσεων μεταξύ όλων των σημείων στην ομάδα και το κέντρο της ομάδας.

Μαθηματικά η απόσταση $D(X, Y)$ μεταξύ ομάδων X και Y ορίζεται ως εξής:

$$D(X, Y) = \sqrt{\frac{2N_X N_Y}{(N_X + N_Y)}} \|\bar{x}_X - \bar{x}_Y\|_2 \quad [31]$$

$$\text{όπου } \bar{x}_X = \frac{1}{N_X} \sum_{i=1}^{N_X} x_{Xi} \quad \text{και} \quad \bar{x}_Y = \frac{1}{N_Y} \sum_{j=1}^{N_Y} x_{Yj}$$

όπου X και Y είναι οποιεσδήποτε δύο ομάδες, $\|\bar{x}_X - \bar{x}_Y\|_2$ είναι η ευκλείδεια απόσταση μεταξύ των κέντρων των δύο ομάδων, \bar{x}_X είναι το κέντρο της ομάδας X , \bar{x}_Y είναι το κέντρο της ομάδας Y , N_X είναι ο συνολικός αριθμός των σημείων της ομάδας X και N_Y είναι ο συνολικός αριθμός των σημείων της ομάδας Y .

3.5.2 Ομαδοποίηση μεθόδων υλοποίησης

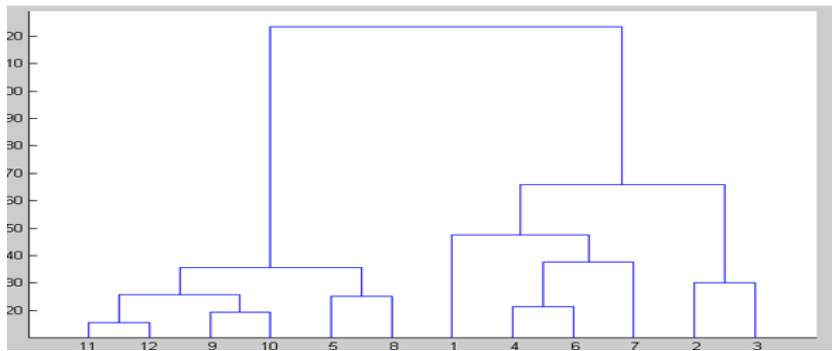
Σε όλα τα σενάρια - μέθοδοι υλοποίησης που θα δούμε σε επόμενο κεφάλαιο, γίνεται χρήση του Ward Linkage για την ομαδοποίηση των λέξεων, εκτός από μία μέθοδο η οποία κάνει χρήση του Average Linkage.

Ο Ward Linkage καθώς και ο Average Linkage είναι συγκεντρωτικοί (bottom-up) ιεραρχικοί αλγόριθμοι, όπως αναφέραμε και προηγουμένως, οι οποίοι φτιάχνουν ένα δέντρο από ομάδες. Αρχίζουν με κάθε λέξη ως ξεχωριστή ομάδα (φύλλα του δέντρου), και σε κάθε βήμα συγχωνεύουν το πιο κοντινό ζευγάρι ομάδων με βάση κάποιο μέτρο ομοιότητας, όπως για παράδειγμα την ευκλείδεια απόσταση, μέχρι να μείνει μία ομάδα (η ρίζα του δέντρου) με όλες τις εικόνες μέσα.

Ο Ward Linkage και ο Average Linkage δέχονται ως είσοδο τον πίνακα δεδομένων (προς ομαδοποίηση) καθώς και την μετρική απόστασης (πχ. Ευκλείδεια απόσταση), για να υπολογίσουν τις αποστάσεις μεταξύ γραμμών (εικόνων) του πίνακα δεδομένων, και επιστρέφουν ως έξοδο τα Cluster labels κάθε εικόνας. Για κάθε εικόνα επιστρέφεται μία ετικέτα (label) στην οποία αναγράφεται ο αριθμός της ομάδας στην οποία έχει εισαχθεί.

Στο σχήμα 3-29, βλέπουμε ένα παράδειγμα με 12 διαφορετικές εικόνες με τα αντίστοιχα label που επιστρέφει ο αλγόριθμος ομαδοποίησης.

Δενδρόγραμμα:



Labels:

1,2,3 : Βαρετός

4,5,6 : Βρέχω

7,8,9 : Για

10,11,12 : Και

Σχήμα 3-29: Παράδειγμα δενδρογράμματος με 12 διαφορετικές εικόνες, αρχίζοντας με 12 διαφορετικές ομάδες (φύλλα) καταλήγοντας στη ρίζα του δέντρου (με όλες τις εικόνες).

ΚΕΦΑΛΑΙΟ 4 - ΣΕΝΑΡΙΑ – ΜΕΘΟΔΟΙ ΥΛΟΠΟΙΗΣΗΣ

4.1 Εισαγωγή

Σε αυτό το κεφάλαιο, θα δούμε αρχικά μια σύντομη περιγραφή των μεθόδων που υλοποιήθηκαν για τους σκοπούς αυτής της εργασίας, στη συνέχεια θα παρουσιαστούν τα σύνολα δεδομένων που χρησιμοποιήθηκαν στα πειράματα της εργασίας, καθώς και ο τρόπος αξιολόγησης των πειραμάτων αυτών. Ακολουθεί μια εκτενής περιγραφή υλοποίησης κάθε μεθόδου, και το κεφάλαιο κλείνει με την σύντομη παρουσίαση μιας φιλικής διεπαφής χρήστη που κατασκευάστηκε για τους σκοπούς της εργασίας.

Οι μέθοδοι-σενάρια υλοποίησης που θα περιγράψουμε στη συνέχεια, αποτελούνται από πέντε διαφορετικές προσεγγίσεις, οι οποίες ακολουθούν μία συγκεκριμένη μεθοδολογία. Με αφορμή το άρθρο [4], ένα από τα πιο δημοφιλή στο χώρο του Word Spotting των T.M.Rath και R.Manmatha, το οποίο έχουμε δει στο κεφάλαιο 2.7.1, υλοποιήθηκαν οι δύο από τις πέντε μεθόδους που θα δούμε στο παρόν κεφάλαιο. Αξίζει να σημειωθεί ότι οι μέθοδοι που παρουσιάζονται στο [4], έχουν σκοπό την εφαρμογή του Word Spotting σε εικόνες χειρόγραφων λέξεων ενός συγγραφέα (ή μιας μικρής ομάδας συγγραφέων), ενώ οι δικές μας μέθοδοι αφορούν τροποποιήσεις των μεθόδων αυτών, σε τέτοιο βαθμό έτσι ώστε να μπορούν να εφαρμοστούν με επιτυχία σε εικόνες χειρόγραφων λέξεων γραμμένων από διάφορους συγγραφείς.

Με αφορμή τα αποτελέσματα που έδωσε το συγκεκριμένο άρθρο για τα δύο πειράματα που έχουμε δει στο κεφάλαιο 2.7.1, υλοποιήσαμε μια δική μας προέγγιση των δύο καλύτερων μεθόδων που παρουσιάστηκαν στο άρθρο αυτό, και τις εφαρμόσαμε στα δικά μας σύνολα δεδομένων, τα οποία θα δούμε αργότερα. Με διάφορους τρόπους, προσπαθήσαμε να εφαρμόσουμε επάνω στις ήδη υπάρχον μεθοδολογίες, τροποποιήσεις πραγματοποιώντας δικά μας πειράματα και σκέψεις με στόχο να πετύχουμε όσο το δυνατόν καλύτερα αποτελέσματα.

Συγκεκριμένα έχουμε υλοποιήσει πέντε διαφορετικές μεθόδους – σενάρια. Από τις πέντε αυτές μεθόδους τις οποίες θα περιγράψουμε στο παρόν κεφάλαιο, η μέθοδος 1 αντιπροσωπεύει την ήδη υπάρχον μεθοδολογία που έχει εφαρμοστεί στο πείραμα (a) του άρθρου [4], το οποίο έχουμε δει στο κεφάλαιο 2.7.1, αλλά για τα δικά μας σύνολα δεδομένων, ενώ η μέθοδος 3 αντιπροσωπεύει την ήδη υπάρχον μεθοδολογία που έχει εφαρμοστεί στο πείραμα (b) του άρθρου [4], το οποίο έχουμε δει στο κεφάλαιο 2.7.1, αλλά και πάλι για τα δικά μας σύνολα δεδομένων. Οι μέθοδοι 2, 4 και 5 αντιπροσωπεύουν εντελώς δικά μας πειράματα και τροποποιήσεις των ήδη υπάρχον μεθόδων 1 και 3. Συγκεκριμένα η μέθοδος 2 είναι μια παραλλαγή της μεθόδου 1, ενώ οι μέθοδοι 4 και 5 είναι παραλλαγές της μεθόδου 3.

4.2 Περιγραφή Μεθόδων

Παρακάτω περιγράφουμε εισαγωγικά τις πέντε μεθόδους με τις οποίες ασχολούμαστε και τις οποίες έχουμε υλοποιήσει:

- 1) Η μέθοδος 1 όπως έχουμε αναφέρει είναι η ήδη υπάρχον μέθοδος, προσπαθεί δηλαδή να ακολουθήσει την ίδια ακριβώς μεθοδολογία που εφαρμόζεται στο άρθρο [4] για το (α) πείραμα αλλά για τα δικά μας σύνολα δεδομένων. Εξάγει τέσσερα χαρακτηριστικά προφίλ (Profile features), από τα οποία υπολογίζεται ο πίνακας αποστάσεων, κάνοντας χρήση του Dynamic Time Warping Matching αλγορίθμου. Με βάση τον πίνακα αυτό, εκτελείται η ομαδοποίηση με αλγόριθμο ομαδοποίησης τον Average Linkage (αυτόν δηλαδή που αναδείχτηκε ως καλύτερος για αυτή τη μέθοδο στο άρθρο [4]) και χρήση της Ευκλείδειας απόστασης.
- 2) Η μέθοδος 2 αφορά αποκλειστικά δική μας παραλλαγή της μεθόδου 1, στην οποία εξάγονται έντεκα χαρακτηριστικά (features), ενώ αυτή τη φορά τα χαρακτηριστικά που εξάγονται, δεν εξάγονται για ολόκληρη την λέξη μόνο, αλλά και για κομμάτια της λέξης. Στη συνέχεια, υπολογίζεται ο πίνακας αποστάσεων κάνοντας χρήση του Dynamic Time Warping Matching αλγορίθμου, και με βάση τον πίνακα αυτό εκτελείται η ομαδοποίηση με αλγόριθμο ομαδοποίησης τον Ward Linkage και χρήση της απόστασης συνημίτονου (Cosine Distance).
- 3) Η μέθοδος 3 όπως έχουμε αναφέρει είναι η ήδη υπάρχον μέθοδος, προσπαθεί δηλαδή να ακολουθήσει την ίδια ακριβώς μεθοδολογία που εφαρμόζεται στο άρθρο [4] για το (b) πείραμα αλλά για τα δικά μας σύνολα δεδομένων. Η κάθε λέξη αναπαριστάται αποκλειστικά από Discrete Fourier Transform (DFT) συντελεστές, οι οποίοι εξάγονται από τα τέσσερα χαρακτηριστικά προφίλ (profile features). Υπολογίζεται ένας πίνακας για όλες τις λέξεις, με τον οποίο εκτελείται η ομαδοποίηση, με αλγόριθμο ομαδοποίησης τον Ward Linkage (αυτόν δηλαδή που αναδείχτηκε ως καλύτερος για αυτή τη μέθοδο στο άρθρο [4]) και χρήση της Ευκλείδειας απόστασης.
- 4) Η μέθοδος 4 αφορά αποκλειστικά δική μας παραλλαγή της μεθόδου 3, στην οποία η κάθε λέξη αναπαριστάται αποκλειστικά από Discrete Fourier Transform (DFT) συντελεστές, οι οποίοι εξάγονται από έντεκα διαφορετικά χαρακτηριστικά (features). Υπολογίζεται ένας πίνακας για όλες τις λέξεις, με τον οποίο εκτελείται η ομαδοποίηση με αλγόριθμο ομαδοποίησης τον Ward Linkage και χρήση της Ευκλείδειας απόστασης.
- 5) Η μέθοδος 5 αφορά επίσης αποκλειστικά δική μας παραλλαγή της μεθόδου 3, στην οποία η κάθε λέξη αυτή τη φορά αναπαριστάται αποκλειστικά αντί από Discrete Fourier Transform (DFT) συντελεστές από Discrete Cosine Transform (DCT) συντελεστές, οι οποίοι εξάγονται από έντεκα διαφορετικά χαρακτηριστικά (features). Υπολογίζεται ένας πίνακας για όλες τις λέξεις, με τον οποίο εκτελείται η ομαδοποίηση με αλγόριθμο ομαδοποίησης τον Ward Linkage και χρήση της Ευκλείδειας απόστασης.

4.3 Σύνολα Δεδομένων

Για να συγκρίνουμε την απόδοση των πέντε μεθόδων που αναφέρθηκαν, έχουμε δημιουργήσει τρία διαφορετικά σεντ δεδομένων, τα οποία είναι:

- 1) Small set of Good Data
- 2) Small set of Random Data
- 3) Large set of Good Data

4.3.1 Small Set of Good Data

Σε αυτό το σεντ δεδομένων έχουμε εισάγει συνολικά 200 διαφορετικές χειρόγραφες λέξεις οι οποίες αποτελούνται από 10 λέξεις με 20 στιγμιότυπα κάθε λέξης. Δηλαδή 10 λέξεις οι οποίες είναι γραμμένες από 20 διαφορετικά άτομα. Σε αυτό το σύνολο δεδομένων οι λέξεις που εισάγαμε ήταν γενικά καλής ποιότητας λέξεις, δηλαδή επιλεγμένες έτσι ώστε να περιέχουν λίγο θόρυβο. Λόγω του ότι οι περισσότερες εικόνες λέξεων που είχαμε στη διάθεσή μας ήταν πολύ υποβαθμισμένες, δηλαδή είχαν προβλήματα όπως ξεθώριασμα, σπασμένα γράμματα, ανολοκλήρωτες λέξεις και υπερβολικό θόρυβο, σε αυτό το σεντ, επιλέχθηκαν μόνο καλής ποιότητας λέξεις, δηλαδή λέξεις οι οποίες διαβάζονται κανονικά.

Στο σύνολο αυτό οι 10 λέξεις που εισήχθησαν από 20 διαφορετικά άτομα είναι οι λέξεις: “ΒΑΡΕΤΟΣ”, “ΔΗΛΑΔΗ”, “εξέλιξη”, “ΕΝΤΟΣ”, “για”, “ΓΙΑΤΙ”, “ΜΑΝΙΑ”, “ΟΧΙ”, “ΨΑΡΙ”, “ΖΗΤΩ”.

4.3.2 Small Set of Random Data

Σε αυτό το σεντ δεδομένων έχουμε εισάγει και πάλι συνολικά 200 διαφορετικές χειρόγραφες λέξεις οι οποίες αποτελούνται από 10 λέξεις με 20 στιγμιότυπα κάθε λέξης. Δηλαδή 10 λέξεις οι οποίες είναι γραμμένες από 20 διαφορετικά άτομα. Σε αυτό το σύνολο δεδομένων οι λέξεις που εισάγαμε ήταν γενικά τυχαίες λέξεις, δηλαδή επιλεγμένες τυχαία αλλά όχι σε σημείο που να είναι υπερβολικά υποβαθμισμένες λέξεις. Επιλέχθηκαν τυχαία, έτσι ώστε να διαβάζονται κανονικά ακόμη και με την παρωπία αρκετού θορύβου, αλλά όχι σε σημείο πολύ υποβαθμισμένης κατάστασης.

Στο σύνολο αυτό οι 10 λέξεις που εισήχθησαν από 20 διαφορετικά άτομα είναι οι λέξεις: “ΒΑΡΕΤΟΣ”, “βρέχω”, “δημοκρατία”, “ΔΗΛΑΔΗ”, “εξέλιξη”, “ΕΝΤΟΣ”, “ΕΥΤΥΧΩΣ”, “φαίνεται”, “ΓΙΑΤΙ”, “και”.

4.3.3 Large Set of Good Data

Σε αυτό το σεντ δεδομένων έχουμε εισάγει συνολικά 1008 διαφορετικές χειρόγραφες λέξεις οι οποίες αποτελούνται από 12 λέξεις με 84 στιγμιότυπα κάθε λέξης. Δηλαδή 12 λέξεις οι οποίες είναι γραμμένες από 84 διαφορετικά άτομα. Σε αυτό το σύνολο δεδομένων οι λέξεις που εισάγαμε ήταν γενικά καλής ποιότητας, δηλαδή επιλεγμένες έτσι ώστε να περιέχουν λίγο θόρυβο, και να διαβάζονται κανονικά.

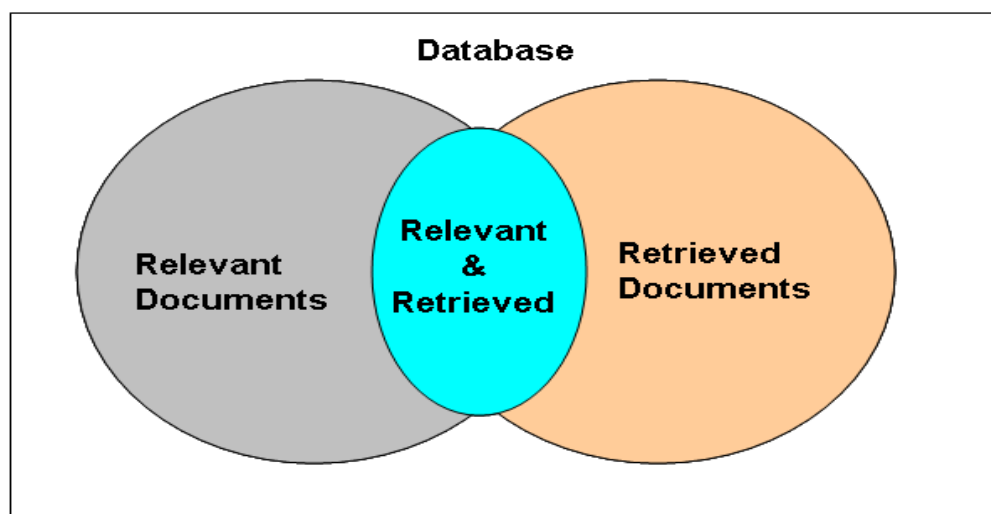
Στο σύνολο αυτό οι 12 λέξεις που εισήχθησαν από 84 διαφορετικά άτομα είναι οι λέξεις: “ΒΑΡΕΤΟΣ”, “βρέχω”, “δημοκρατία”, “ΔΗΛΑΔΗ”, “εξέλιξη”, “ΕΝΤΟΣ”, “για”, “ΓΙΑΤΙ”, “ΜΑΝΙΑ”, “ΟΧΙ”, “ΨΑΡΙ”, “ΖΗΤΩ”.

4.4 Μέθοδος Αξιολόγησης (Evaluation Method)

Λόγω του ότι τα αποτελέσματα που δίνει το στάδιο της ομαδοποίησης, δεν είναι εύκολο να αξιολογηθούν με ανθρώπινο μάτι, και κυρίως όταν έχουμε να ασχοληθούμε με μεγάλα σύνολα δεδομένων, στην περίπτωση μας εικόνες λέξεων, η διαδικασία της αξιολόγησης είναι χρονοβόρα και κουραστική για τον άνθρωπο. Χρειάζεται λοιπόν κάποιο στάδιο το οποίο θα μπορεί να αξιολογεί τα δεδομένα, να δίνει κάποια αποτελέσματα, ούτως ώστε να έχουμε τη δυνατότητα να πάρουμε κάποιες αποφάσεις για αυτά. Το στάδιο της αξιολόγησης αποτελεί βασικότατο παράγοντα για την αξιολόγηση των αποτελεσμάτων που προκύπτουν από τα σενάρια - μεθόδους υλοποίησης που θα μελετήσουμε στο παρόν κεφάλαιο. Στόχος είναι, να αξιολογεί τις ομάδες που δημιουργούνται από κάθε σενάριο, για κάθε ένα από τα τρία σύνολα δεδομένων που αναφέραμε προηγουμένως στο κεφάλαιο 4-3, κάνοντας χρήση κάποιων μέτρων αξιολόγησης. Θέλουμε δηλαδή να μάθουμε κατά πόσο μία λέξη έχει ενταχθεί στη σωστή ομάδα, και τελικά να δούμε κατά πόσο έχουμε επιτύχει ή αποτύχει συνολικά για όλες τις λέξεις. Τα μέτρα αξιολόγησης θα μας δώσουν ποσοστά επί τις εκατό, για το σκοπό αυτό.

Δύο ευρέως χρησιμοποιούμενα μέτρα αξιολόγησης των αποτελεσμάτων που χρησιμοποιούνται στην αναγνώριση προτύπων, είναι το Precision και το Recall. Είναι απλές μετρικές οι οποίες υπολογίζουν το ποσοστό των εμφανίσεων των σωστών αποτελεσμάτων.

Στην ανάκτηση πληροφορίας για τον υπολογισμό αυτών των μετρικών, μία ολόκληρη συλλογή δεδομένων χωρίζεται σε σχετικά έγγραφα (relevant documents) και σε ανακτημένα έγγραφα (retrieved documents) όπως βλέπουμε στο παρακάτω σχήμα:



Σχήμα 4-1: Ο τρόπος με τον οποίο χωρίζονται τα έγγραφα στην ανάκτηση πληροφορίας ώστε να κατανοηθεί ευκολότερα η χρήση των μέτρων αξιολόγησης Precision και Recall.

Οι μέθοδοι αξιολόγησης με τις οποίες θα ασχοληθούμε είναι:

1. Precision
2. Recall
3. F-Measure
4. Purity

4.4.1 Precision

Το Precision είναι σε ποσοστό επί της % ο αριθμός των ανακτημένων εγγράφων (retrieved documents) τα οποία είναι σχετικά (relevant).

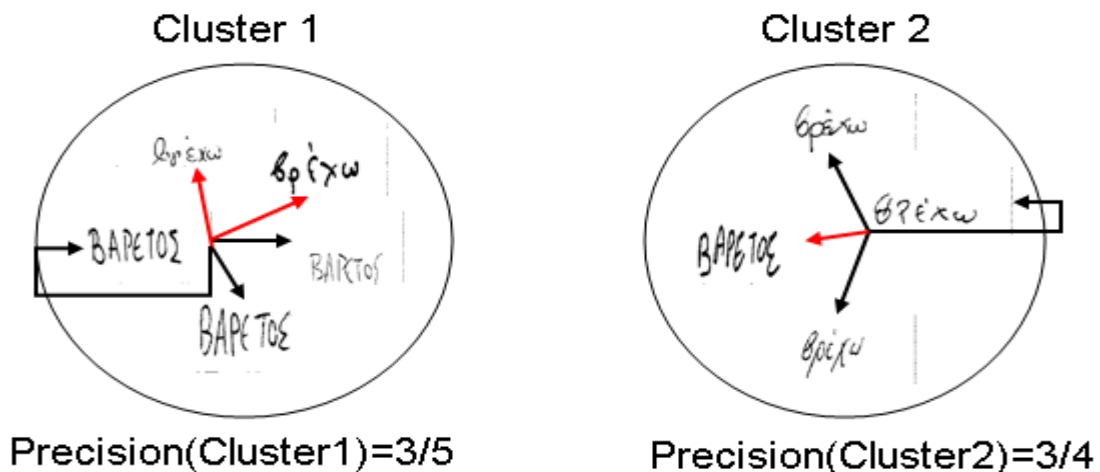
$$\text{Precision} = \frac{\# \text{Relevant} + \text{Retreived}}{\# \text{Retreived}} \quad [32]$$

Δηλαδή ο αριθμός των σχετικών (relevant) εγγράφων που έχει ανακτηθεί (retrieved) δια τον συνολικό αριθμό των εγγράφων που έχει ανακτηθεί (retrieved). Με άλλα λόγια το precision είναι ο αριθμός των σωστών αποτελεσμάτων διαιρούμενο με τον αριθμό όλων των αποτελεσμάτων που επιστράφηκαν.

Στην περίπτωση μας το Precision μίας ομάδας από λέξεις είναι ο αριθμός των λέξεων που έχουν εισαχθεί σωστά στην ομάδα δια το συνολικό αριθμό των λέξεων της ομάδας αυτής. Ομάδα μιας λέξης ορίζεται ως η ομάδα της λέξης που εμφανίζεται πιο συχνά. Για να βρούμε το συνολικό Precision όλων των ομάδων απλά υπολογίζουμε τον μέσο όρο τους.

Για παράδειγμα έχουμε μία ομάδα η οποία αποτελείται από 5 στιγμιότυπα της λέξης “Γιατί” και 2 στιγμιότυπα της λέξης “Όχι”, τότε έχουμε Precision= 5/7 =71.4286% για αυτή την ομάδα.

Στο σχήμα 4-2 δείχνουμε παράδειγμα υπολογισμού του precision για δύο ομάδες:



Σχήμα 4-2: Παράδειγμα υπολογισμού του Precision για 2 ομάδες (Clusters).

4.4.2 Recall

Το Recall είναι σε ποσοστό επί της % ο αριθμός όλων των σχετικών εγγράφων (relevant documents) τα οποία έχουν ανακτηθεί (retrieved).

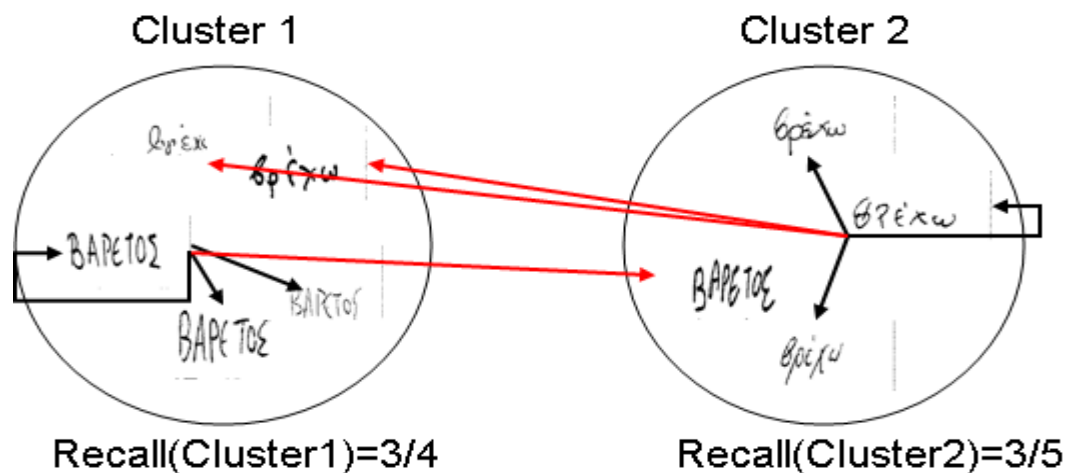
$$\text{Recall} = \frac{\# \text{Relevant} + \text{Retreived}}{\text{Total } \# \text{Relevant in Database}} \quad [32]$$

Δηλαδή ο αριθμός των σχετικών (relevant) εγγράφων που έχει ανακτηθεί (retrieved) δια τον συνολικό αριθμό των σχετικών (relevant) εγγράφων που υπάρχει στη βάση ή στη συλλογή. Με άλλα λόγια το recall είναι ο αριθμός των σωστών αποτελεσμάτων διαιρούμενο με τον αριθμό όλων των αποτελεσμάτων που θα έπρεπε να επιστραφούν.

Στην περίπτωση μας το Recall μίας ομάδας από λέξεις είναι ο αριθμός των λέξεων που έχουν εισαχθεί σωστά στην ομάδα (δηλαδή ο αριθμός με τα πιο πολλά στιγμιότυπα) δια το συνολικό αριθμό της λέξης που αντιπροσωπεύει την ομάδα που υπάρχει μέσα στη βάση ή στη συλλογή. Για να βρούμε το συνολικό Recall όλων των ομάδων απλά υπολογίζουμε τον μέσο όρο τους.

Για παράδειγμα έχουμε μία ομάδα η οποία αποτελείται από 5 στιγμιότυπα της λέξης “Γιατί” και 2 στιγμιότυπα της λέξης “Όχι”, ενώ τα συνολικά στιγμιότυπα της λέξης “Γιατί” είναι 10 (αυτά που υπάρχουν συνολικά στη βάση) τότε έχουμε $Recall = 5/10 = 50\%$ για αυτή την ομάδα.

Στο σχήμα 4-3 δείχνουμε παράδειγμα υπολογισμού του recall για δύο ομάδες:



Σχήμα 4-3: Παράδειγμα υπολογισμού του Recall για 2 ομάδες (Clusters).

4.4.3 F-Measure

Το F-Measure είναι ένα μέτρο αξιολόγησης το οποίο συνδυάζει το Precision και το Recall με την παρακάτω εξίσωση:

$$F - Measure = \frac{2 * Precision * Recall}{Precision + Recall} \quad [33]$$

Με βάση το προηγούμενο παράδειγμα στο οποίο βρήκαμε Precision = 71.4286% και Recall = 50%, εφαρμόζοντας τον τύπο, θα έχουμε F-Measure = 58.823% για την συγκεκριμένη ομάδα.

Για να βρούμε το συνολικό F-Measure όλων των ομάδων απλά υπολογίζουμε τον μέσο όρο τους για όλες τις ομάδες.

4.4.4 Purity

Το Purity είναι ένα μέτρο αξιολόγησης το οποίο επικεντρώνεται στη συχνότητα εμφάνισης της πιο κοινής κατηγορίας σε κάθε ομάδα. Δείχνει τον βαθμό ως προς το κατά πόσο μία ομάδα περιλαμβάνει έγγραφα μίας μόνο κατηγορίας.

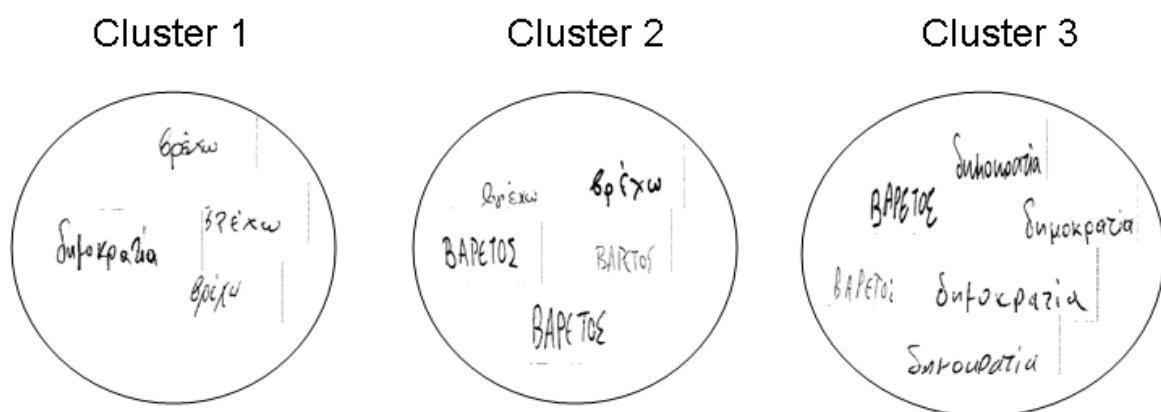
$$Purity(C_j) = \frac{1}{|C_j|} \max_i (|C_j|_{class=i}) \quad [34]$$

Όπου C_j είναι η ομάδα για την οποία υπολογίζεται το purity, $|C_j|$ είναι το μέγεθος της ομάδας C_j , $|C_j|_{class=i}$ δηλώνει τον αριθμό των αντικειμένων της κλάσης (ή κατηγορίας) i που ανατίθενται στην ομάδα j .

Το συνολικό Purity για ένα σύνολο δεδομένων D , k ομάδων εκφράζεται ως ένα άθροισμα των επιμέρους purities κάθε ομάδας ως εξής :

$$Purity(C_j) = \sum_{j=1}^k \frac{|C_j|}{|D|} purity(C_j) \quad [34]$$

Στο σχήμα 4-4 δείχνουμε παράδειγμα υπολογισμού του purity για τρεις ομάδες:



Σχήμα 4-4: Παράδειγμα υπολογισμού του Purity για 3 ομάδες (Clusters).

Από το σχήμα 4-4, έχουμε συνολικά στο σύνολο δεδομένων μας 15 λέξεις οι οποίες έχουν ενταχθεί σε 3 ομάδες. Ο αριθμός των μελών της κλάσης με την υψηλότερη πλειοψηφία για τις τρεις παραπάνω ομάδες είναι: για το Cluster 1 πλειοψηφεί η κλάση “βρέχω” με 3 στιγμιότυπα, για το Cluster 2 πλειοψηφεί η κλάση “βαρετός” με 3 στιγμιότυπα και για το Cluster 3 πλειοψηφεί η κλάση “δημοκρατία” με 4 στιγμιότυπα.

Συνεπώς το Purity υπολογίζεται ως εξής : $Purity = \frac{1}{15} * (3 + 3 + 4) = 66.6\%$

Από τα αποτελέσματα που παίρνουμε από τις μεθόδους που ακολουθούν στη συνέχεια στο κεφάλαιο 4.5, το purity δίνει τα ίδια ακριβώς αποτελέσματα με το recall, κάτι που είναι απόλυτα λογικό αν ληφθούν υπόψη οι τρόποι υπολογισμού των δύο αυτών δεικτών αξιολόγησης. Αυτό φυσικά συμβαίνει, γιατί επιλέγεται σε κάθε περίπτωση αριθμός ομάδων ίσος με τον αριθμό των διαφορετικών λέξεων που υπάρχουν σε κάθε σύνολο δεδομένων. Σε μια πιο ρεαλιστική προσέγγιση, όπου δεν αναμένεται κάθε λέξη να μπει στη δική της ομάδα, και επομένως επιλέγεται ελαφρώς μεγαλύτερος αριθμός ομάδων από τις διαφορετικές λέξεις του συνόλου δεδομένων, οι τιμές των δεικτών αυτών θα διαφέρουν. Εμείς απλά επιλέξαμε να αφήσουμε και τις δύο τιμές, αφού διαλέξαμε να παρουσιάσουμε και τους δύο αυτούς δείκτες.

4.5 Περιγραφή Μεθόδων

Σε αυτήν την ενότητα θα παρουσιαστούν λεπτομερώς οι μέθοδοι που υλοποιήθηκαν για τους σκοπούς της παρούσας εργασίας.

4.5.1 Μέθοδος 1

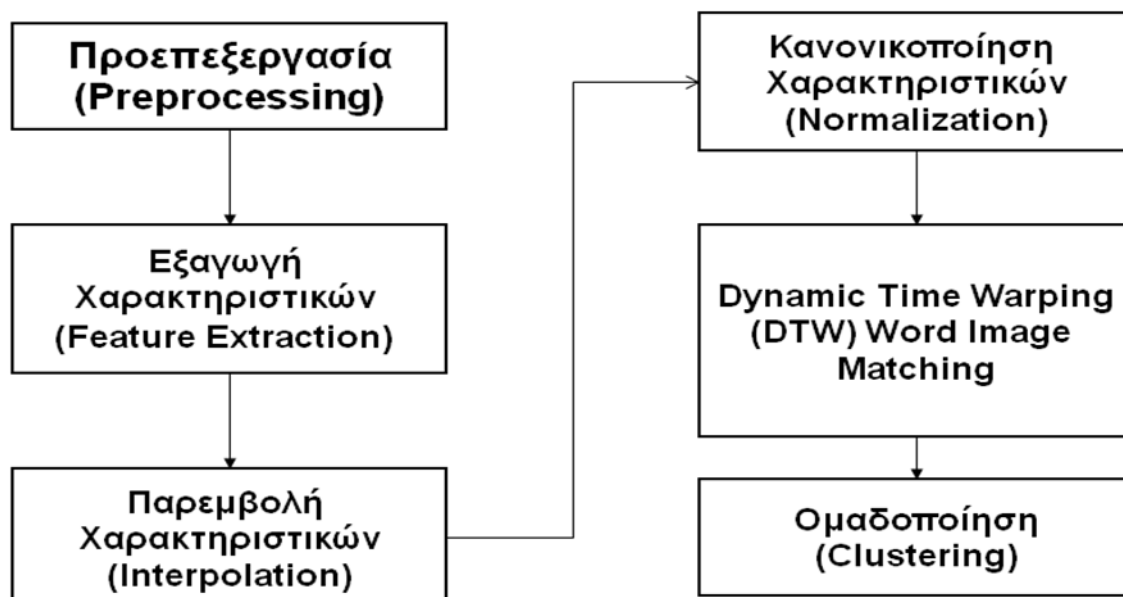
Σκοπός μας για την υλοποίηση της πρώτης μεθόδου ήταν η κατά προσέγγιση ανάπτυξη μιας μεθόδου που θα ακολουθούσε συγκεκριμένα βήματα υλοποίησης όπως περιγράφονται στο [4]. Φυσικά, δεν είναι δυνατόν η μέθοδος μας να είναι πλήρως πανομοιότυπη με την μέθοδο που αναπτύχθηκε στο [4], καθώς δεν μπορούμε να γνωρίζουμε λεπτομέρειες σχετικά με την υλοποίηση των συγγραφέων, ενώ σε μερικά σημεία η μέθοδος μας διαφέρει από την αρχική μέθοδο που υλοποιήθηκε στο συγκεκριμένο άρθρο. Στις επόμενες παραγράφους, θα περιγράψουμε συνοπτικά την αρχική μέθοδο και στην συνέχεια θα περιγράψουμε τη δική μας υλοποίηση για τη μέθοδο αυτή, αναφέροντας κατά τη διάρκεια της ενότητας τα σημεία στα οποία διαφέρει από την αρχική μέθοδο, και τελικά δίνοντας τα αποτελέσματα που προέκυψαν στα δεδομένα μας από την χρησιμοποίηση της μεθόδου 1.

4.5.1.1 Αρχική μέθοδος

Αναφέροντας την ακολουθία των εφαρμογών και υλοποιήσεων ακριβώς όπως αναφέρονται και στο [4], θα προσπαθήσουμε να δώσουμε μια σύντομη περιγραφή της υλοποίησης της αρχικής μεθόδου, η οποία είναι μια επιλεγμένη από εμάς μέθοδος, από τις διάφορες που δοκιμάζονται και συγκρίνονται στο [4]. Σύμφωνα με το [4], λαμβάνεται αρχικά μια gray scale εικόνα με τις τιμές των pixels της εικόνας να βρίσκονται στο διάστημα [0..255]. Αφού στο αρχικό έγγραφο εφαρμοστεί τμηματοποίηση γραμμών και λέξεων με την μέθοδο που περιγράφεται στο [35] και [2], το μόνο που αναφέρεται κατά την προ-επεξεργασία, είναι πως η εικόνα κάθε λέξης υπόκειται σε μεθόδους, που έχουν σκοπό να ανιχνεύσουν και να κανονικοποιήσουν διακυμάνσεις μέσα στη λέξη, όπως είναι οι γωνίες skew και slant. Τα χαρακτηριστικά που εξάγονται από κάθε εικόνα λέξης, είναι το Vertical Projection Profile, το Upper Word Profile, το Lower Word Profile και το Vertical Background to Ink Transitions Profile, τα οποία κανονικοποιούνται σε τιμές του διαστήματος [0,1], ενώ τα Upper Word Profile και Lower Word Profile, περνούν και από μια μέθοδο γραμμικής παρεμβολής, για να γεμίσουν τα κενά που υπάρχουν στα διανύσματα που τα περιέχουν. Ακολούθως, χρησιμοποιώντας μια συγκεκριμένη τιμή καταωφλίου οι συγγραφείς εφάρμοσαν την τεχνική του pruning στα δικά τους δεδομένα για να μειώσουν τις συγκρίσεις μεταξύ των λέξεων. Οι συγκρίσεις μεταξύ των λέξεων έγιναν με την χρήση του DTW, όπως ακριβώς περιγράφεται στο κεφάλαιο 3, χρησιμοποιώντας επίσης σαν απόσταση για τον σκοπό αυτό, αυτήν που περιγράφηκε στο ίδιο κεφάλαιο, δηλαδή το τετράγωνο της ευκλείδειας απόστασης. Έχοντας πλέον τις ανά ζεύγος αποστάσεις μεταξύ των εικόνων λέξεων αποθηκευμένες σε ένα πίνακα, τις έδωσαν σαν είσοδο σε 5 ιεραρχικούς αλγόριθμους με επιλεγμένη απόσταση την ευκλείδεια απόσταση. Από αυτούς, προέκυψε καλύτερος ο Average Linkage, ο οποίος έφερε το χαμηλότερο Word Error Rate, 34.12%, γι' αυτό και επιλέχθηκε σαν ο αλγόριθμος που θα εφαρμοστεί στην δική μας μέθοδο 1.

4.5.1.2 Διάγραμμα ροής μεθοδολογίας

Τα βήματα που ακολουθούνται στην μεθοδολογία υλοποίησης της μεθόδου 1, είναι ακριβώς όπως φαίνονται και από το διάγραμμα ροής μεθοδολογίας της μεθόδου στο σχήμα 4-5.



Σχήμα 4-5: Διάγραμμα ροής μεθοδολογίας για την μέθοδο 1.

4.5.1.3 Περιγραφή μεθόδου

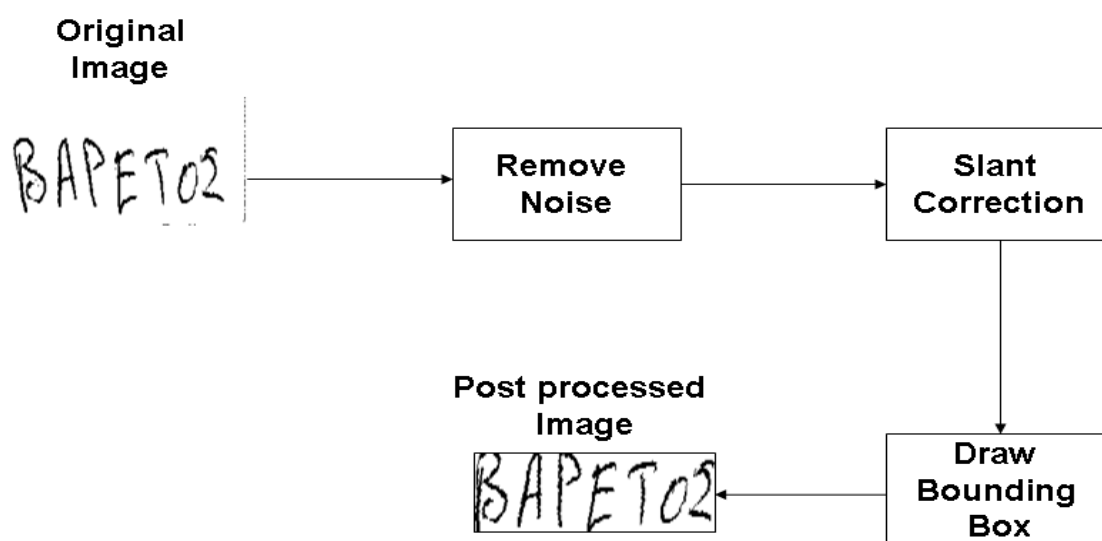
Προσπαθώντας να ακολουθήσουμε τα βήματα της αρχικής μεθόδου, υλοποιήσαμε τη μέθοδο 1. Σε αντίθεση όμως με την αρχική μέθοδο, τα δικά μας δεδομένα δεν ήταν ολόκληρο το κείμενο αλλά ήταν ήδη εικόνες λέξεων, οπότε το στάδιο της τμηματοποίησης δεν εφαρμόστηκε από εμάς, ενώ επίσης οι εικόνες ήταν δυαδικής μορφής, δηλαδή είχαν τιμές pixel 0 και 1.

4.5.1.3.1 Προ-επεξεργασία

Η προ-επεξεργασία των εικόνων έγινε χρησιμοποιώντας κάποιες από τις μεθόδους που περιγράφονται στο κεφάλαιο 3.1, και πιο συγκεκριμένα εφαρμόστηκαν οι πιο κάτω εφαρμογές κατά την προ-επεξεργασία κάθε εικόνας χειρόγραφης λέξης:

- 1) Αφαίρεση θορύβου (Noise removal).
- 2) Εντοπισμός και αφαίρεση του φαινομένου slant (Slant detection & correction).
- 3) Περιορισμός της λέξης μέσα σε πλαίσιο οριοθέτησης (Contain the word in it's bounding box).

Όπως πιθανόν έχει εφαρμοστεί αυτή η προ-επεξεργασία και στην περίπτωση της αρχικής μεθόδου, με τη διαφορά ό τ στην δική μας υλοποίηση δεν εφαρμόστηκε η ανίχνευση και διόρθωση του φαινομένου skew, καθώς οι πλείστες εικόνες μας δεν έπασχαν από σοβαρή ύπαρξη αυτού του φαινομένου. Στο σχήμα 4-6 φαίνεται το διάγραμμα ροής μεθοδολογίας για την εφαρμογή της προ-επεξεργασίας που εφαρμόστηκε στην μέθοδο 1.



Σχήμα 4-6: Διάγραμμα ροής προ-επεξεργασίας μεθόδου 1.

4.5.1.3.2 Εξαγωγή χαρακτηριστικών (Feature extraction)

Τα χαρακτηριστικά που εξάχθηκαν από κάθε εικόνα λέξης, ήταν ακριβώς τα ίδια με τα χαρακτηριστικά που εξάχθηκαν και στην αρχική μέθοδο, δηλαδή τα πιο κάτω:

- 1) Upper Word Profile
- 2) Lower Word Profile
- 3) Vertical Projection Profile
- 4) Vertical Background to Ink Transitions Profile

Είναι δηλαδή διανύσματα μιας γραμμής και n στηλών, όπου n είναι το πλήθος των στηλών της εικόνας της χειρόγραφης λέξης.

4.5.1.3.3 Παρεμβολή και κανονικοποίηση χαρακτηριστικών (Interpolation & normalization of the extracted features)

Από τα πιο πάνω χαρακτηριστικά, τα δύο πρώτα πέρασαν από γραμμική παρεμβολή όπως έγινε και στην περίπτωση της αρχικής μεθόδου. Ειδικότερα, η γραμμική παρεμβολή στην περίπτωση μας εφαρμόστηκε όπως ακριβώς περιγράφεται στο κεφάλαιο 3.2.2. Επίσης, όλα τα χαρακτηριστικά που εξάχθηκαν κατά την μέθοδο 1, κανονικοποιήθηκαν σε τιμές του διαστήματος $[0,1]$ όπως περιγράφεται στο κεφάλαιο 3.2.3.

4.5.1.3.4 Matching

Ακολουθώντας και πάλι τα βήματα που περιγράφουν την αρχική μέθοδο, στο στάδιο του Matching, εφαρμόζουμε ακριβώς τις μεθόδους που περιγράφονται στο [4], και τις έχουμε υλοποιήσει όπως ακριβώς περιγράφονται στο κεφάλαιο 3.3.

Αναλυτικότερα, για κάθε εικόνα χειρόγραφης λέξης, αφού εξαχθούν τα χαρακτηριστικά που αναφέραμε πιο πάνω, εισάγονται σε ένα πίνακα, ο οποίος περιέχει 4 γραμμές, όπου σε κάθε γραμμή βρίσκεται ένα διάνυσμα n στηλών που περιγράφει ένα χαρακτηριστικό, και όπου n είναι το πλήθος των στηλών της εικόνας. Οι πίνακες αυτοί, συγκρίνονται μεταξύ τους για κάθε συνδυασμό εικόνων λέξεων του συνόλου δεδομένων, με την χρήση του DTW, και οι τιμές που λαμβάνονται σαν αποστάσεις μεταξύ κάθε συνδυασμού πινάκων των εικόνων, εισάγονται μέσα σε ένα άλλο πίνακα, τον πίνακα αποστάσεων. Δηλαδή, ο τελευταίος πίνακας είναι ένας πίνακας διαστάσεων $M \times M$, όπου M είναι το πλήθος όλων των λέξεων του συνόλου δεδομένων, και περιέχει τις ανά ζεύγος αποστάσεις μεταξύ κάθε συνδυασμού λέξεων. Για παράδειγμα, στις θέσεις (1,3) και (3,1) του πίνακα, βρίσκεται η απόσταση που υπολογίστηκε με την χρήση του DTW, μεταξύ των πινάκων χαρακτηριστικών της πρώτης και τρίτης λέξης του συνόλου δεδομένων. Αφού κάθε εικόνα λέξης συγκρινόμενη με τον εαυτό της όπως είναι λογικό επιστρέφει μηδενική απόσταση, η κεντρική διαγώνιος του πίνακα αποστάσεων περιέχει μόνο μηδενικές τιμές. Για παράδειγμα αν έχουμε σύνολο δεδομένων που αποτελείται από 5 εικόνες λέξεων, ο πίνακας αποστάσεων θα είναι της μορφής του πίνακα που φαίνεται στον πίνακα 4-1.

	1	2	3	4	5
1	0	0.3313	0.5123	0.9123	2.3311
2	0.3313	0	1.1451	1.2309	3.1520
3	0.5123	1.1451	0	0.3145	2.1219
4	0.9123	1.2309	0.314	0	0.9991
5	2.3311	3.1520	2.1219	0.9991	0

Πίνακας 4-1: Παράδειγμα πίνακα αποστάσεων για σύνολο δεδομένων που αποτελείται από 5 εικόνες λέξεων.

4.5.1.3.5 Ομαδοποίηση (Clustering)

Αφού τώρα έχουμε τον πίνακα αποστάσεων, τον εισάγουμε σαν είσοδο στον αλγόριθμο ομαδοποίησης Average Linkage, ο οποίος περιγράφεται στο κεφάλαιο 3.5.1.3, με επιλεγμένη μετρική απόσταση την ευκλείδεια απόσταση, όπως γίνεται δηλαδή και στην περίπτωση της αρχικής μεθόδου.

4.5.1.4 Αποτελέσματα (Results)

Εφαρμόζοντας την μέθοδο 1 που περιγράψαμε πιο πάνω για τα τρία σύνολα δεδομένων που περιγράφονται στο κεφάλαιο 4.3, παίρνουμε τα αποτελέσματα που φαίνονται στον πίνακα 4-2, τα οποία αξιολογήθηκαν με βάση τους δείκτες Precision, Recall, F-Measure και Purity, που περιγράφονται στο κεφάλαιο 4.4.

Average Linkage	Small Set of Good Data (10 Clusters)	Small Set of Random Data (10 Clusters)	Large Set of Good Data (12 Clusters)
Precision (%)	79.9755	72.4071	79.7755
Recall (%)	37.5000	37.0000	16.5674
F-Measure (%)	51.0588	48.9742	27.4369
Purity (%)	37.5000	37.0000	16.5674
Time (Minutes)	19.1207	29.1244	539.8002 (8.9966 Hours)

Πίνακας 4-2: Αποτελέσματα μεθόδου 1 για τα τρία σύνολα δεδομένων. Στον χρόνο περιλαμβάνεται ο χρόνος όλων των λειτουργιών που εκτελούνται στη μέθοδο. Οι χρόνοι υπολογίστηκαν κατά την εκτέλεση της μεθόδου στη Matlab σε υπολογιστή με 2-πύρηνο επεξεργαστή 2,27 GHz.

Τα καλύτερα αποτελέσματα όπως φαίνεται στον πίνακα 4-2 δίνονται για το Small Set of Good Data. Και στις τρεις περιπτώσεις, βλέπουμε ότι το Precision είναι σε αποδεκτά επίπεδα, το Recall όμως δίνει πολύ χαμηλές τιμές, κάτι που υπεηρεάζει και την τιμή του F-Measure. Επίσης παρατηρούμε ότι το Purity και το Recall έχουν ακριβώς τις ίδιες τιμές, κάτι που προκύπτει από τον ορισμό των δύο δεικτών αξιολόγησης. Η εκτέλεση της μεθόδου για αριθμό ομάδων διαφορετικό από τον ακριβή αριθμό των διαφορετικών λέξεων, θα δώσει διαφορετικές τιμές στους δύο δείκτες. Επίσης, παρατηρούμε ότι η μέθοδος αυτή είναι αρκετά χρονοβόρα αφού χρειάζονται περίπου 9 ώρες για την εκτέλεση της σε ένα σύνολο δεδομένων 1008 εικόνων λέξεων (Large Set of Good Data), ενώ παρατηρούμε επιπλέον ότι το σύνολο δεδομένων που περιέχει και υποβαθμισμένης ποιότητας εικόνες λέξεων (Small Set of Random Data), χρειάζεται περισσότερο χρόνο σε εκτέλεση σε σύγκριση με το σύνολο δεδομένων που δεν περιέχει εικόνες λέξεων υποβαθμισμένης ποιότητας (Small Set of Good Data) και είναι του ίδιου πλήθους εικόνων.

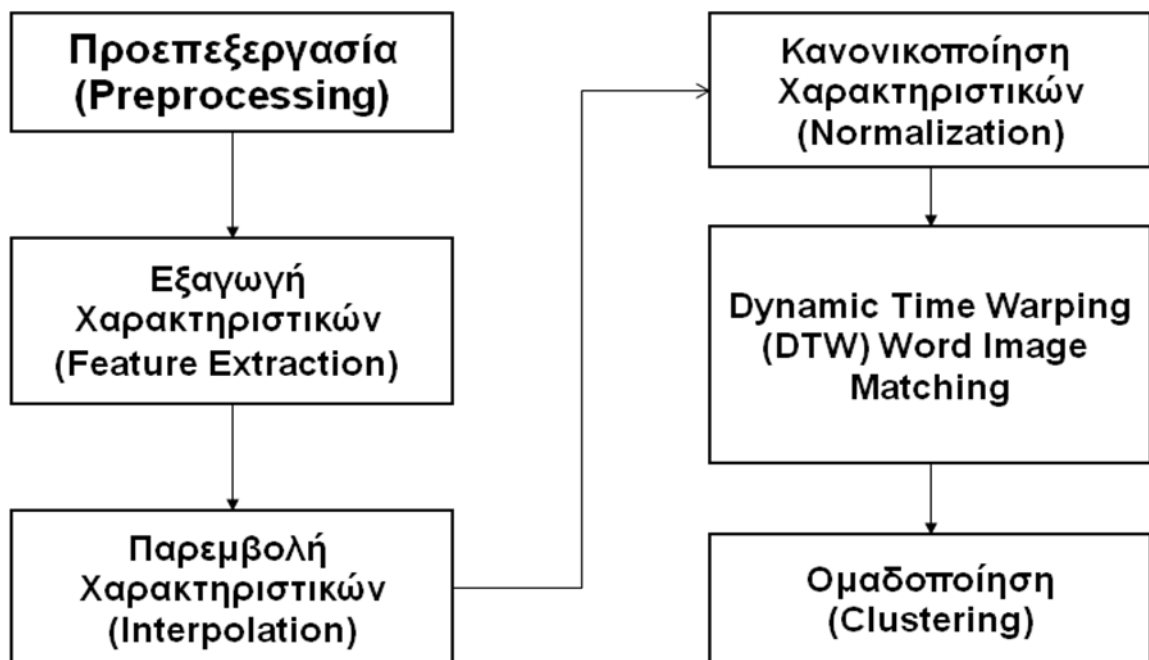
4.5.2 Μέθοδος 2

Η μέθοδος 2 είναι αποκλειστικά δική μας μέθοδος. Συγκεκριμένα είναι μια παραλλαγή της μεθόδου 1, με τροποποιήσεις στην προ-επεξεργασία στην οποία υπόκεινται οι εικόνες των λέξεων κάθε συνόλου δεδομένων, καθώς έχουν προστεθεί χαρακτηριστικά προς εξαγωγή για τις εικόνες των λέξεων, αλλά δεν έχει αλλάξει η μεθοδολογία της μεθόδου. Τα βήματα από τα οποία αποτελείται η μεθοδολογία που χρησιμοποιήθηκε για την ανάπτυξη της μεθόδου 2 είναι τα ίδια με τα βήματα της μεθόδου 1, δηλαδή:

- 1) Προ-επεξεργασία (Preprocessing)
- 2) Εξαγωγή χαρακτηριστικών (Feature Extraction)
- 3) Παρεμβολή Χαρακτηριστικών (Interpolation)
- 4) Κανονικοποίηση Χαρακτηριστικών (Normalization)
- 5) Matching με DTW
- 6) Ομαδοποίηση (Clustering)

4.5.2.1 Διάγραμμα ροής μεθοδολογίας

Τα βήματα που αναφέρονται πιο πάνω φαίνονται και στο σχήμα 4-5, το οποίο μεταφέρεται στο σχήμα 4-7 για ευκολία του αναγνώστη. Το σχήμα αυτό δείχνει το διάγραμμα ροής της μεθοδολογίας που εφαρμόστηκε στην περίπτωση της μεθόδου 1 και 2.



Σχήμα 4-7: Διάγραμμα ροής μεθοδολογίας για την μέθοδο 2.

4.5.2.2 Περιγραφή μεθόδου

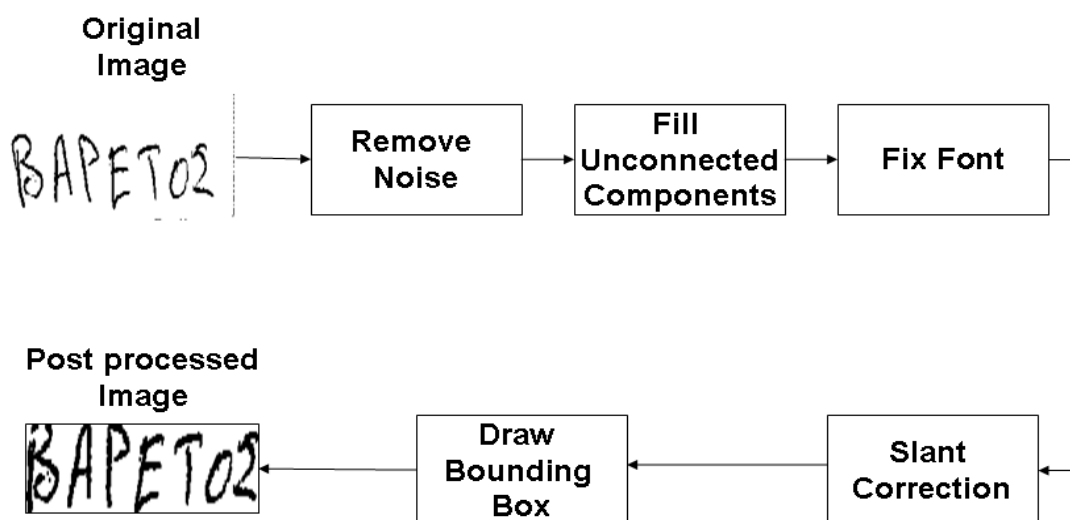
Εδώ θα αναφέρουμε τις διαφορές μεταξύ των υλοποιήσεων της μεθόδου 1 και της μεθόδου 2, τις οποίες θα εξηγήσουμε με περισσότερη λεπτομέρεια κατά την διάρκεια αυτής της ενότητας. Ενώ η μεθοδολογία των δύο μεθόδων σε γενικά επίπεδα φαίνεται ίδια, εντούτοις υπάρχουν αρκετές διαφορές ανάμεσα στις δύο υλοποιήσεις. Στο στάδιο της προ-επεξεργασίας που εφαρμόζεται στην μέθοδο 2, προστέθηκαν δύο ακόμα εφαρμογές στις ήδη υπάρχουσες εφαρμογές προ-επεξεργασίας της μεθόδου 1, οι οποίες έχει βρεθεί ότι συνολικά φέρνουν καλύτερα αποτελέσματα από την προ-επεξεργασία που εφαρμόστηκε στη μέθοδο 1. Επίσης, κατά την εξαγωγή χαρακτηριστικών, και κάποια νέα χαρακτηριστικά έχουν επιλεγεί να εξαχθούν από τις εικόνες των λέξεων εκτός από τα 4 που χρησιμοποιήθηκαν στην μέθοδο 1. Άλλες διαφορές μεταξύ των δύο μεθόδων βρίσκονται στο στάδιο του Matching, κατά τον υπολογισμό δηλαδή του πίνακα αποστάσεων, καθώς και στο στάδιο της ομαδοποίησης όπου στην περίπτωση της μεθόδου 2 χρησιμοποιείται διαφορετικός αλγόριθμος ομαδοποίησης καθώς και διαφορετική μετρική απόσταση για τον αλγόριθμο αυτό.

4.5.2.2.1 Προ-επεξεργασία

Όπως αναφέρεται και πιο πάνω, η προ-επεξεργασία της μεθόδου 2 αποτελείται από τα βήματα της προ-επεξεργασίας της μεθόδου 1, καθώς και από τα βήματα: Ένωση σπασμένων κομματιών της λέξης και κανονικοποίηση του font. Τα βήματα που ακολουθούν την προ-επεξεργασία της μεθόδου 2 είναι:

- 1) Αφαίρεση θορύβου (Noise removal).
- 2) Ένωση σπασμένων κομματιών της λέξης (Connect broken parts of the word image).
- 3) Εντοπισμός και αφαίρεση του φαινομένου slant (Slant detection & correction).
- 4) Περιορισμός της λέξης μέσα σε πλαίσιο οριοθέτησης (Contain the word in it's bounding box).
- 5) Κανονικοποίηση του font (Font normalization).

Η περιγραφή και η σημασία κάθε ενός από τα πιο πάνω βήματα δίνεται στο κεφάλαιο 3.1. Στο σχήμα 4-8 φαίνεται το διάγραμμα ροής της προ-επεξεργασίας που εφαρμόστηκε στην μέθοδο 2.



Σχήμα 4-8: Διάγραμμα ροής προ-επεξεργασίας μεθόδου 2.

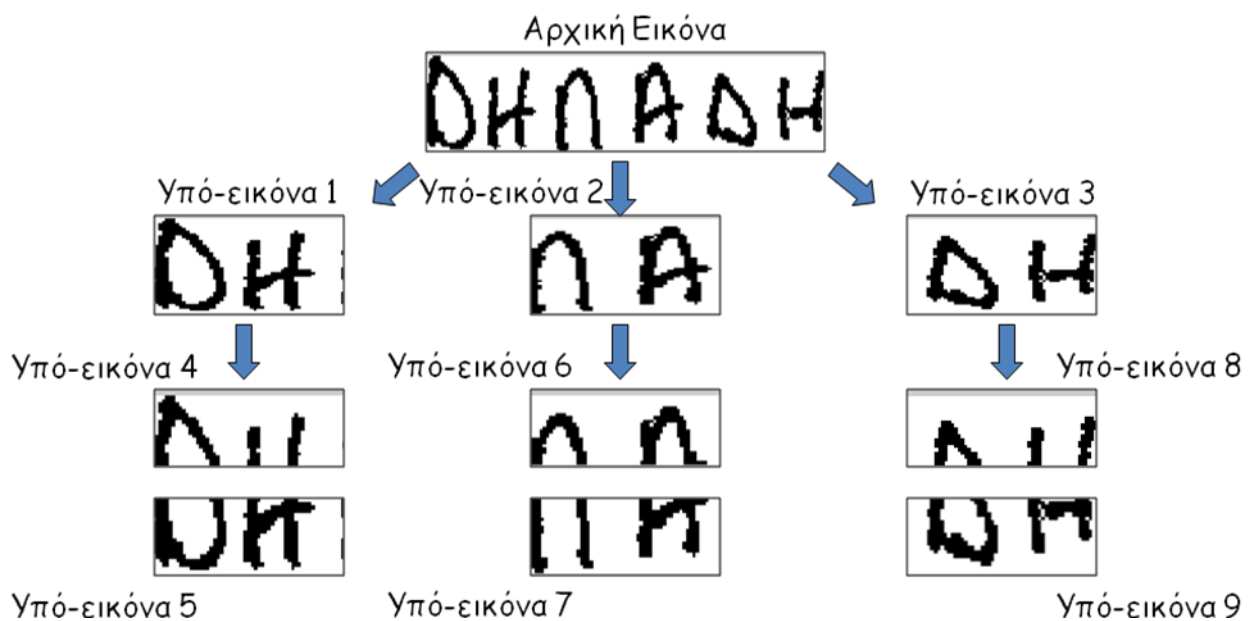
4.5.2.2.2 Εξαγωγή χαρακτηριστικών (Feature Extraction)

Τα χαρακτηριστικά που εξάγονται για κάθε εικόνα χειρόγραφης λέξης του συνόλου δεδομένων, είναι τα:

- 1) Upper Word Profile
- 2) Lower Word Profile
- 3) Left Word Profile
- 4) Right Word Profile
- 5) Lower Minus Upper Word Profile
- 6) Vertical Projection Profile
- 7) Horizontal Projection Profile
- 8) Vertical Background to Ink Transitions Profile
- 9) Horizontal Background to Ink Transitions Profile
- 10) End Points
- 11) Holes Detection

Για κάθε ένα από τα χαρακτηριστικά αυτά δίνεται λεπτομερής περιγραφή και επεξήγηση της σημασίας του στο κεφάλαιο 3.2.

Η σημασία του βήματος εξαγωγής χαρακτηριστικών αυτής της μεθόδου, δεν έγκειται μόνο στην επιλογή του συγκεκριμένου συνδυασμού χαρακτηριστικών, αλλά και στον τρόπο με τον οποίο τα χαρακτηριστικά αυτά εξάγονται από την εικόνα της λέξης. Συγκεκριμένα, στο βήμα αυτό, η κάθε εικόνα λέξης, τεμαχίζεται σε διάφορες μικρότερες υπό-εικόνες, όπου από την κάθε υπό-εικόνα εξάγονται διαφορετικά χαρακτηριστικά. Παράδειγμα ενός τέτοιου τεμαχισμού φαίνεται στο σχήμα 4-9.



Σχήμα 4-9: Παράδειγμα τεμαχισμού αρχικής εικόνας χειρόγραφης λέξης σε 9 μικρότερες υπό-εικόνες.

Συγκεκριμένα, από τις εικόνες που παρουσιάζονται στο σχήμα 4-9:

- Από την αρχική εικόνα, εξάγεται μόνο το χαρακτηριστικό:
 - Holes Detection

- Από τις υπό-εικόνες 1, 2 και 3, εξάγονται τα χαρακτηριστικά:
 - Upper Word Profile
 - Lower Word Profile
 - Lower Minus Upper Word Profile
 - Vertical Projection Profile
 - Vertical Background To Ink Transitions Profile
 - Holes Detection

- Από τις υπό-εικόνες 4 και 5, εξάγονται τα χαρακτηριστικά:
 - Left Word Profile
 - Horizontal Projection Profile
 - Horizontal Background To Ink Transitions Profile
 - End Points

- Από τις υπό-εικόνες 6 και 7, εξάγονται τα χαρακτηριστικά:
 - Horizontal Projection Profile
 - Horizontal Background To Ink Transitions Profile
 - End Points

- Από τις υπό-εικόνες 8 και 9, εξάγονται τα χαρακτηριστικά:
 - Right Word Profile
 - Horizontal Projection Profile
 - Horizontal Background To Ink Transitions Profile
 - End Points

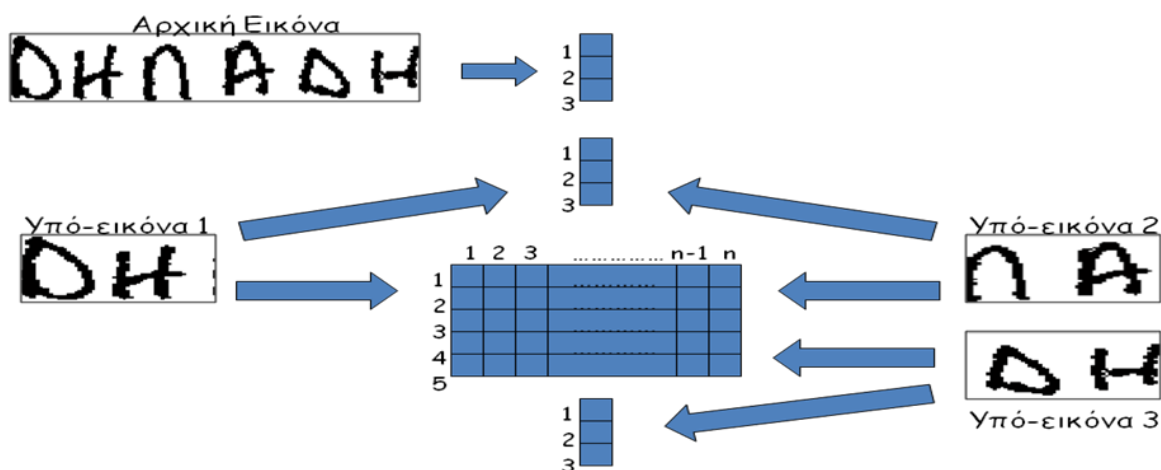
Είναι σημαντικό να αναφερθεί ότι η επιλογή του συγκεκριμένου συνδυασμού χαρακτηριστικών, ο τεμαχισμός κάθε εικόνας με αυτόν τον τρόπο, καθώς και η επιλογή των συγκεκριμένων χαρακτηριστικών για κάθε υπό-εικόνα, έγιναν μετά από πολλά πειράματα που αποτελούνταν από συνδυασμούς διάφορων χαρακτηριστικών, και διαφορετικών μεθόδων τεμαχισμού. Τελικά, αποφασίστηκε και επιλέχθηκε η χρησιμοποίηση του συγκεκριμένου συνδυασμού χαρακτηριστικών και μεθόδου τεμαχισμού που παρουσιάζεται πιο πάνω, αφού ήταν αυτή που έδινε τα καλύτερα αποτελέσματα από όλα τα πειράματα που επιχειρήσαμε.

4.5.2.2.3 Παρεμβολή και κανονικοποίηση χαρακτηριστικών (Interpolation & normalization of the extracted features)

Από τα πιο πάνω χαρακτηριστικά, τα upper, lower, left και right word profiles πέρασαν από γραμμική παρεμβολή όπως έγινε και στην περίπτωση της αρχικής μεθόδου. Ειδικότερα, η γραμμική παρεμβολή στην περίπτωση μας, εφαρμόστηκε όπως ακριβώς περιγράφεται στο κεφάλαιο 3.2.2. Επίσης, όλα τα χαρακτηριστικά που εξάχθηκαν κατά την μέθοδο 2, κανονικοποιήθηκαν σε τιμές του διαστήματος $[0,1]$ όπως περιγράφεται στο κεφάλαιο 3.2.3.

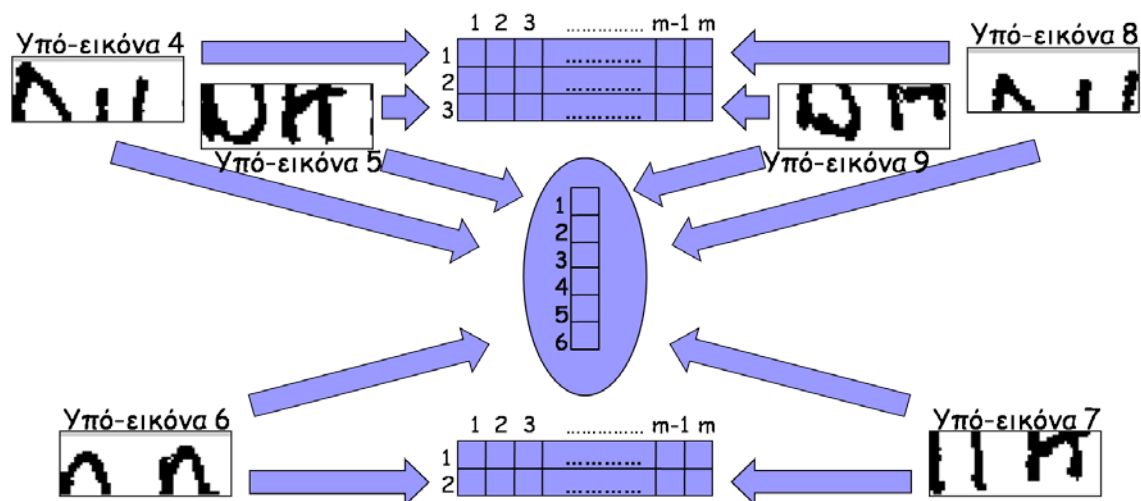
4.5.2.2.4 Matching

Παρατηρώντας τα χαρακτηριστικά που επιλέχθηκαν να εξαχθούν από κάθε εικόνα λέξης, και λαμβάνοντας υπόψη την επιλογή μας να τεμαχίσουμε την εικόνα σε μικρότερες υπό-εικόνες, είναι εύκολο να καταλήξουμε στο συμπέρασμα ότι τα χαρακτηριστικά που εξάγονται από κάθε εικόνα λέξης είναι διαφορετικών διαστάσεων. Αυτό δυσκολεύει το έργο κατασκευής πινάκων χαρακτηριστικών που να είναι κατάλληλοι για εισαγωγή στον DTW, έτσι ώστε με την εισαγωγή αυτή να επιστρέφεται η απόσταση μεταξύ των πινάκων αυτών για να κατασκευαστεί τελικά ο πίνακας αποστάσεων μεταξύ των εικόνων λέξεων. Για τον σκοπό της επίλυσης αυτού του προβλήματος, επιλέξαμε μια απλή λύση, κατά την οποία για κάθε εικόνα κατασκευάζονταν 14 πίνακες, από τους οποίους ο καθένας περιείχε συγκεκριμένα χαρακτηριστικά της εικόνας. Κάθε ένας από αυτούς τους πίνακες, συγκρινόταν με τους πίνακες που περιείχαν τα αντίστοιχα χαρακτηριστικά κάθε εικόνας του συνόλου δεδομένων. Έτσι τελικά είχαμε 14 πίνακες αποστάσεων, τους οποίους προσθέταμε για να δημιουργηθεί ο τελικός πίνακας αποστάσεων των εικόνων λέξεων. Τα σχήματα 4-10 και 4-11 παρουσιάζουν την δημιουργία των 14^{ων} πινάκων για την εικόνα που φαίνεται στο σχήμα 4-9.



Σχήμα 4-10: Παράδειγμα δημιουργίας πινάκων χαρακτηριστικών.

Όπως φαίνεται και από το σχήμα 4-10, για την αρχική εικόνα κατασκευάζεται ένας πίνακας 3^{ων} γραμμών και 1στήλης (Holes Detection). Για τις 3 υπό-εικόνες που φαίνονται στο σχήμα κατασκευάζονται για κάθε μια από αυτές, ένας πίνακας 3^{ων} γραμμών και 1 στήλης (Holes Detection) καθώς και ένας πίνακας 5 γραμμών και n στηλών (UWP, LWP, LMUWP, VPP, VBTTP), όπου n είναι το πλήθος των στηλών της κάθε υπό-εικόνας.



Σχήμα 4-11: Παράδειγμα δημιουργίας πινάκων χαρακτηριστικών.

Όπως φαίνεται και από το σχήμα 4-11, για τις υπό-εικόνες 4 και 5 κατασκευάζεται ένας πίνακας 3^{ov} γραμμών και m στηλών (LWP, HPP, HBTITP). Για τις υπό-εικόνες 6 και 7 κατασκευάζεται ένας πίνακας 2 γραμμών και m στηλών (HPP, HBTITP). Για τις υπό-εικόνες 8 και 9 κατασκευάζεται ένας πίνακας 3^{ov} γραμμών και m στηλών (RWP, HPP, HBTITP), όπου m είναι το πλήθος στηλών της κάθε υπό-εικόνας. Από όλες τις υπό-εικόνες του σχήματος δημιουργείται ακόμα ένας πίνακας 6 γραμμών και 1 στήλης, που περιέχει σε κάθε γραμμή τα End Points κάθε υπό-εικόνας.

Πρέπει επίσης να σημειώσουμε, ότι επιχειρήσαμε να αλλάξουμε τον τρόπο που υπολογίζεται η απόσταση από τον DTW, δηλαδή αντί για το τετράγωνο της ευκλείδειας απόστασης, χρησιμοποιήσαμε την απόσταση συνημίτονων, αλλά αυτό μας έδινε χειρότερα αποτελέσματα, οπότε κρατήσαμε τελικά την αρχική μέθοδο υπολογισμού απόστασης.

4.5.2.2.5 Ομαδοποίηση (Clustering)

Έχοντας πλέον τον πίνακα αποστάσεων, που αποτελεί την είσοδο του αλγόριθμου ομαδοποίησης, δοκιμάσαμε στο στάδιο αυτό 7 συγκεντρωτικούς ιεραρχικούς αλγόριθμους Linkage, δηλαδή τους:

- 1) Average Linkage
- 2) Centroid Linkage
- 3) Complete Linkage
- 4) Median Linkage
- 5) Single Linkage
- 6) Ward Linkage
- 7) Weighted Linkage

Στον πίνακα 4-3 φαίνονται τα αποτελέσματα που προέκυψαν κατά την εκτέλεση της μεθόδου 2 για τους 7 συγκεντρωτικούς ιεραρχικούς αλγόριθμους που αναφέραμε πιο πάνω. Η εκτέλεση

αυτή έγινε στο σύνολο δεδομένων Small Set of Good Data, με επιλεγμένη μετρική απόσταση την απόσταση συνημίτονων.

Small Set of Good Data	Precision (%)	Recall (%)	F – Measure (%)	Purity (%)
Average Linkage	87.4891	76.5000	81.6263	76.5000
Centroid Linkage	86.8589	75.5000	80.7821	75.5000
Complete Linkage	82.9285	81.0000	81.9529	81.0000
Median Linkage	79.8965	67.5000	73.1769	67.5000
Single Linkage	91.1764	25.0000	39.2405	25.0000
Ward Linkage	93.7518	93.0000	93.3744	93.0000
Weighted Linkage	83.6868	73.0000	77.9789	73.0000

Πίνακας 4-3: Αποτελέσματα μεθόδου 2 για το Small Set of Good Data. Για κάθε ένα από τους 7 αλγόριθμους Linkage, δίνεται το ποσοστό των δεικτών αξιολόγησης: Precision, Recall, F-Measure και Purity, ενώ η μετρική απόσταση που χρησιμοποιήθηκε εδώ ήταν η απόσταση συνημίτονων (Cosine Distance).

Από τους αλγόριθμους που συγκρίναμε προέκυψε ότι τα καλύτερα αποτελέσματα, έδινε ο Ward Linkage, όπως φαίνεται και στον πίνακα 4-3 για το Small Set of Good Data, γι' αυτό και αποφασίσαμε να χρησιμοποιηθεί αυτός σαν ο αλγόριθμος ομαδοποίησης για την μέθοδο 2.

Επίσης, δοκιμάσαμε και διάφορες άλλες μετρικές αποστάσεων κατά το στάδιο της ομαδοποίησης εκτός από την ευκλείδεια απόσταση, στις οποίες περιλαμβάνονται η seuclidean (standardized Euclidean Distance), η απόσταση συνημίτονων, και η απόσταση Jaccard. Στον πίνακα 4-4 φαίνονται τα αποτελέσματα που προέκυψαν κατά την εκτέλεση της μεθόδου 2 για το Small Set of Good Data, για 11 διαφορετικές μετρικές αποστάσεων κατά την ομαδοποίηση, με αλγόριθμο ομαδοποίησης τον Ward Linkage.

Small Set of Good Data	Precision (%)	Recall (%)	F – Measure (%)	Purity (%)
Euclidean Distance	69.3081	59.5000	64.0306	59.5000
Seuclidean Distance	78.0503	73.0000	75.4407	73.0000
Cityblock Distance	71.1197	62.0000	66.2474	62.0000
Minkowski Distance	69.3081	59.5000	64.0306	59.5000
Chebychev Distance	91.2931	89.5000	90.3876	89.5000
Mahalanobis Distance	34.7717	17.0000	22.8356	17.0000
Cosine Distance	93.7518	93.0000	93.3744	93.0000
Correlation Distance	84.1007	82.0000	83.0370	82.0000
Spearman Distance	82.8403	80.0000	81.3954	80.0000
Hamming Distance	91.0869	18.0000	30.0597	18.0000
Jaccard Distance	91.0869	18.0000	30.0597	18.0000

Πίνακας 4-4: Αποτελέσματα μεθόδου 2 για το Small Set of Good Data. Για κάθε μια από τις 11 μετρικές αποστάσεων, δίνεται το ποσοστό των δεικτών αξιολόγησης: Precision, Recall, F-Measure και Purity, ενώ ο αλγόριθμος ομαδοποίησης που χρησιμοποιήθηκε εδώ ήταν ο Ward Linkage.

Από τις μετρικές αποστάσεων που έχουν δοκιμαστεί, τα καλύτερα αποτελέσματα έφερε η απόσταση συνημίτονων (Cosine Distance), όπως φαίνεται και στον πίνακα 4-4 για το Small Set of Good Data, γι' αυτό και επιλέχθηκε να χρησιμοποιηθεί αυτή σαν η μετρική απόσταση για την ομαδοποίηση της μεθόδου 2. Μια συνοπτική περιγραφή της κάθε μετρικής απόστασης που εμφανίζεται στον πίνακα 4-4 μπορεί να βρεθεί στο [31].

4.5.2.3 Αποτελέσματα (Results)

Εφαρμόζοντας την μέθοδο 2 που περιγράψαμε πιο πάνω για τα τρία σύνολα δεδομένων που περιγράφονται στο κεφάλαιο 4.3, παίρνουμε τα αποτελέσματα που φαίνονται στον πίνακα 4-5, τα οποία αξιολογήθηκαν με βάση τους δείκτες Precision, Recall, F-Measure και Purity, που περιγράφονται στο κεφάλαιο 4.4.

Ward Linkage	Small Set of Good Data (10 Clusters)	Small Set of Random Data (10 Clusters)	Large Set of Good Data (12 Clusters)
Precision (%)	93.7518	75.3480	91.2443
Recall (%)	93.0000	71.5000	87.7976
F-Measure (%)	93.3744	73.3736	89.4878
Purity (%)	93.0000	71.5000	87.7976
Time (Minutes)	10.9238	15.5273	297.2059 (4.9534 Hours)

Πίνακας 4-5: Αποτελέσματα μεθόδου 2 για τα τρία σύνολα δεδομένων. Στον χρόνο περιλαμβάνεται ο χρόνος όλων των λειτουργιών που εκτελούνται στη μέθοδο. Οι χρόνοι υπολογίστηκαν κατά την εκτέλεση της μεθόδου στη Matlab σε υπολογιστή με 2-πύρηνο επεξεργαστή 2,27 GHz.

Από τα αποτελέσματα όπως είναι λογικό βλέπουμε αρκετά καλύτερα ποσοστά για το Small Set of Good Data από ότι για τα άλλα δύο. Στην περίπτωση του Small Set of Good Data και Large Set of Good Data έχουμε ικανοποιητικά ποσοστά, ενώ για το Small Set of Random Data βλέπουμε ότι είναι κάποιου βαθμού χαμηλότερα, κάτι που είναι απόλυτα λογικό και αναμενόμενο, αφού στο Small Set of Random Data περιέχονται και λίγο υποβαθμισμένης ποιότητας εικόνες λέξεων. Επίσης, παρατηρούμε ότι σε σύγκριση με την μέθοδο 1, τα αποτελέσματα έχουν αυξηθεί σημαντικά, ενώ παρόλο που και η μέθοδος 2 είναι χρονοβόρα, έχει μειωθεί περίπου στο μισό ο χρόνος εκτέλεσης σε σύγκριση με τον χρόνο που απαιτείται για την μέθοδο 1. Και σε αυτή την περίπτωση βλέπουμε ότι το Small Set of Random Data χρειάζεται περισσότερο χρόνο εκτέλεσης από το Small Set of Good Data, κάτι που είναι λογικό αφού το στάδιο της προ-επεξεργασίας χρειάζεται περισσότερο χρόνο για να διορθώσει εικόνες λέξεων που είναι λίγο υποβαθμισμένες.

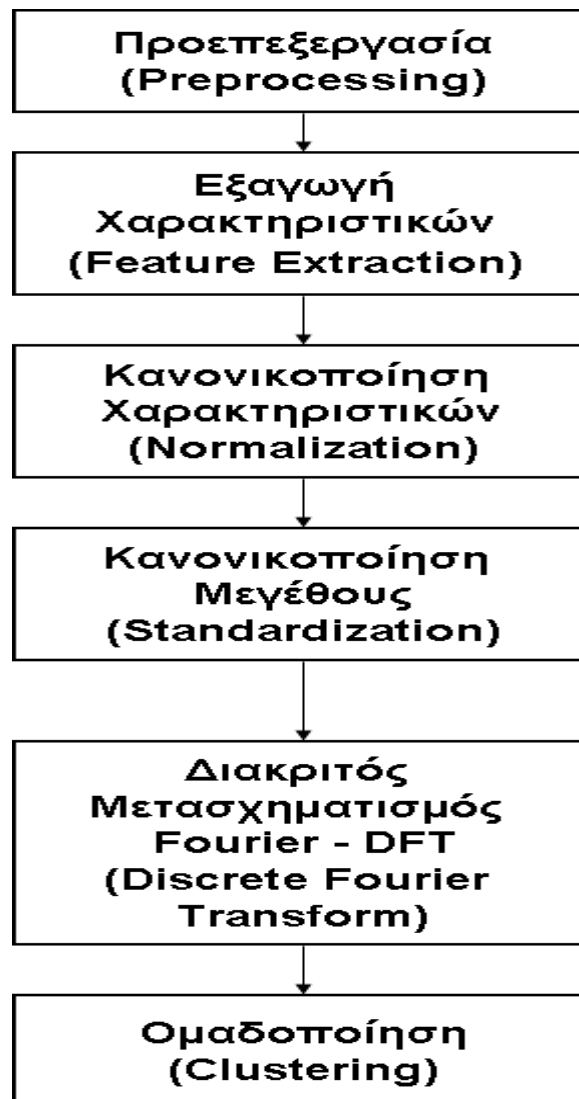
4.5.3 Μέθοδος 3

Η μέθοδος 3 όπως έχουμε αναφέρει, αφορά ήδη υπάρχον μέθοδο η οποία αποτελείται από τα παρακάτω βήματα:

- 1) Προ-επεξεργασία (Preprocessing)
- 2) Εξαγωγή χαρακτηριστικών (Feature Extraction)
- 3) Κανονικοποίηση Χαρακτηριστικών (Normalization)
- 4) Κανονικοποίηση Μεγέθους (Standardization)
- 5) Διακριτός Μετασχηματισμός Fourier (Discrete Fourier Transform)
- 6) Ομαδοποίηση (Clustering)

4.5.3.1 Διάγραμμα ροής μεθοδολογίας

Στο σχήμα 4-12 δείχνουμε το διάγραμμα ροής που ακολουθεί η μέθοδος 3.



Σχήμα 4-12: Διάγραμμα ροής μεθοδολογίας για την μέθοδο 3.

4.5.3.2 Περιγραφή Μεθόδου

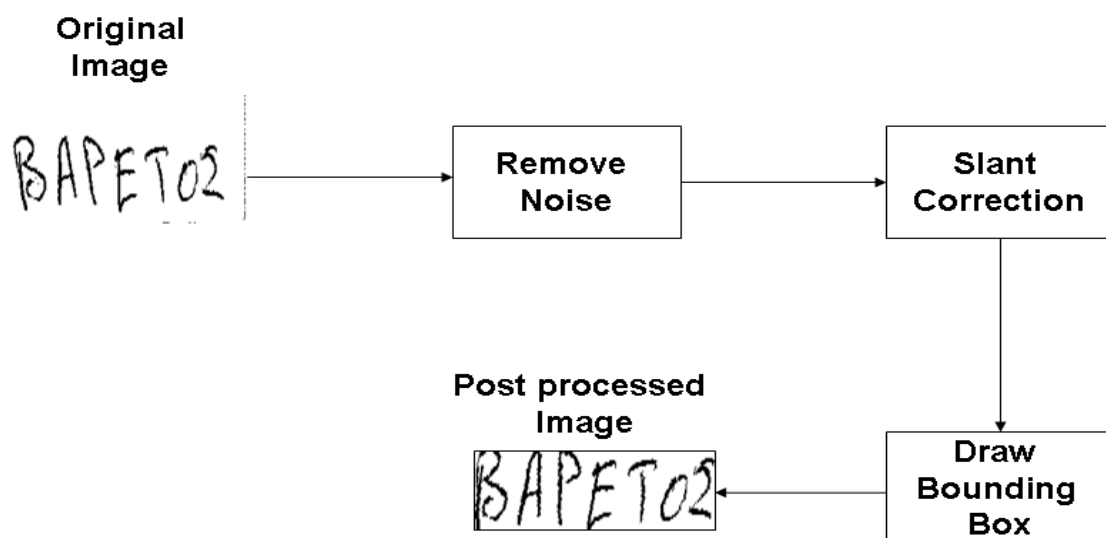
Η μέθοδος 3 όπως έχουμε πει είναι ήδη υπάρχον μέθοδος, η οποία ξεκινά με το στάδιο της προ-επεξεργασίας, εξάγει τέσσερα χαρακτηριστικά, τα κανονικοποιεί, εκτελεί κανονικοποίηση μεγέθους, μετασχηματίζει σε DFT συντελεστές κάθε διάνυσμα χαρακτηριστικών, φτιάχνει τον πίνακα δεδομένων και εν τέλει πραγματοποιεί το Clustering με βάση τον Ward Linkage.

4.5.3.2.1 Προ-επεξεργασία

Η μέθοδος 3, αρχίζει με πρώτο στάδιο την προ-επεξεργασία κάθε χειρόγραφης εικόνας. Λαμβάνεται αρχικά μία σκαναρισμένη δυαδική εικόνα χειρόγραφης λέξης (με pixels 0 και 1), η οποία υποφέρει από διάφορα προβλήματα όπως θόρυβος, ξεθώριασμα, διαφορές στο font, διαφορές στην κλίση κτλ., στην οποία χρειάζεται μία σωστή προ-επεξεργασία ώστε να διορθωθεί όσο το δυνατόν καλύτερα.

Η προ-επεξεργασία αρχίζει με πρώτο βήμα την αφαίρεση του θορύβου της λέξης, στη συνέχεια διορθώνεται η κλίση της και εν τέλει σχεδιάζεται το πλαίσιο οριοθέτησης (bounding box) της.

Στο σχήμα 4-13 δείχνουμε το διάγραμμα ροής που ακολουθεί η προ-επεξεργασία της μεθόδου 3:



Σχήμα 4-13: Διάγραμμα ροής προ-επεξεργασίας μεθόδου 3.

4.5.3.2.2 Εξαγωγή χαρακτηριστικών (Feature Extraction)

Αφού λοιπόν σε κάθε εικόνα έχει πραγματοποιηθεί το στάδιο της προ-επεξεργασίας, η λέξη είναι πλέον έτοιμη να περάσει στο στάδιο της εξαγωγής χαρακτηριστικών. Στο στάδιο της εξαγωγής χαρακτηριστικών εξάγονται από κάθε λέξη τα τέσσερα πιο κάτω χαρακτηριστικά:

- 1) Upper Word Profile
- 2) Lower Word Profile
- 3) Vertical Projection Profile
- 4) Vertical Background to Ink Transitions Profile

Το κάθε ένα από τα πιο πάνω χαρακτηριστικά αναπαριστάται ως ένα διάνυσμα μιας γραμμής και N στηλών, όπου N είναι το πλάτος της κάθε λέξης το οποίο φυσικά μεταβάλλεται για κάθε λέξη.

4.5.3.2.3 Κανονικοποίηση (Normalization) και Κανονικοποίηση Μεγέθους (Standardization)

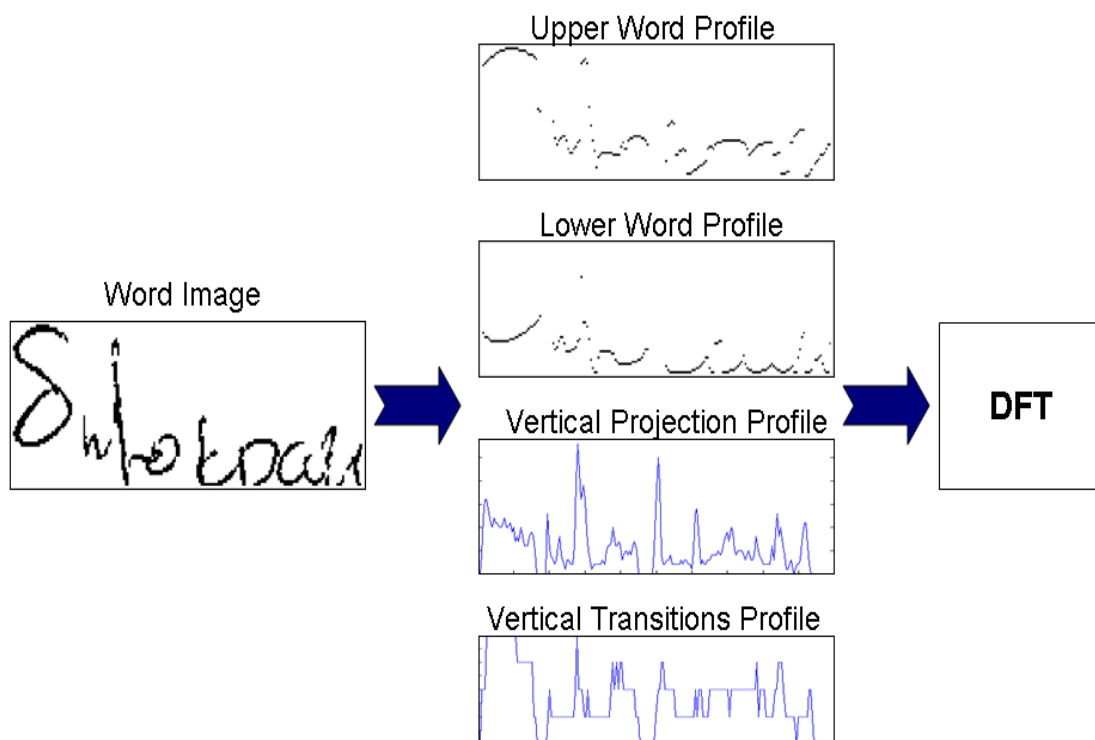
Στη συνέχεια, το κάθε ένα από τα τέσσερα διανύσματα γραμμής μεγέθους $1 \times N$, περνά στο στάδιο της κανονικοποίησης (normalization), όπου πραγματοποιείται η κανονικοποίηση του σε κλίμακα $[0,1]$, έτσι ώστε να μπορούν τα χαρακτηριστικά να συγκρίνονται μεταξύ τους. Η κανονικοποίηση των τεσσάρων αυτών χαρακτηριστικών επιτελείται ακριβώς όπως έχουμε αναφέρει στο κεφάλαιο 3 για τα συγκεκριμένα χαρακτηριστικά.

Αφού λοιπόν τα τέσσερα αυτά χαρακτηριστικά έχουν πλέον κανονικοποιηθεί, περνούν στο στάδιο της κανονικοποίησης μεγέθους (standardization). Εδώ όπως μιλήσαμε και στο κεφάλαιο 3, λόγω του ότι τα πλάτη κάθε εικόνας μεταβάλλονται δεν έχουμε σταθερά μεγέθη, έτσι κάθε ένα από τα τέσσερα μεταβαλλόμενα διανύσματα μεγέθους $1 \times N$ κανονικοποιείται σε ένα σταθερό μέγεθος, στην περίπτωσή μας σε 1×175 . Η κανονικοποίηση μεγέθους είναι πολύ σημαντική για να επιτευχθεί με επιτυχία η σύγκριση μεταξύ λέξεων.

4.5.3.2.4 Διακριτός Μετασχηματισμός Fourier (Discrete Fourier Transform – DFT)

Μετά την κανονικοποίηση μεγέθους, τα τέσσερα αυτά διανύσματα μεγέθους πλέον 1×175 , εισάγονται ως είσοδος στο στάδιο του διακριτού μετασχηματισμού Fourier (Discrete Fourier Transform - DFT). Εκεί, τα τέσσερα αυτά διανύσματα μετασχηματίζονται στο πεδίο της συχνότητας κάνοντας χρήση του μετασχηματισμού Fourier. Δηλαδή τα τέσσερα αυτά διανύσματα, αναπαριστώνται πλέον ως τέσσερα μετασχηματισμένα διανύσματα με συντελεστές Fourier.

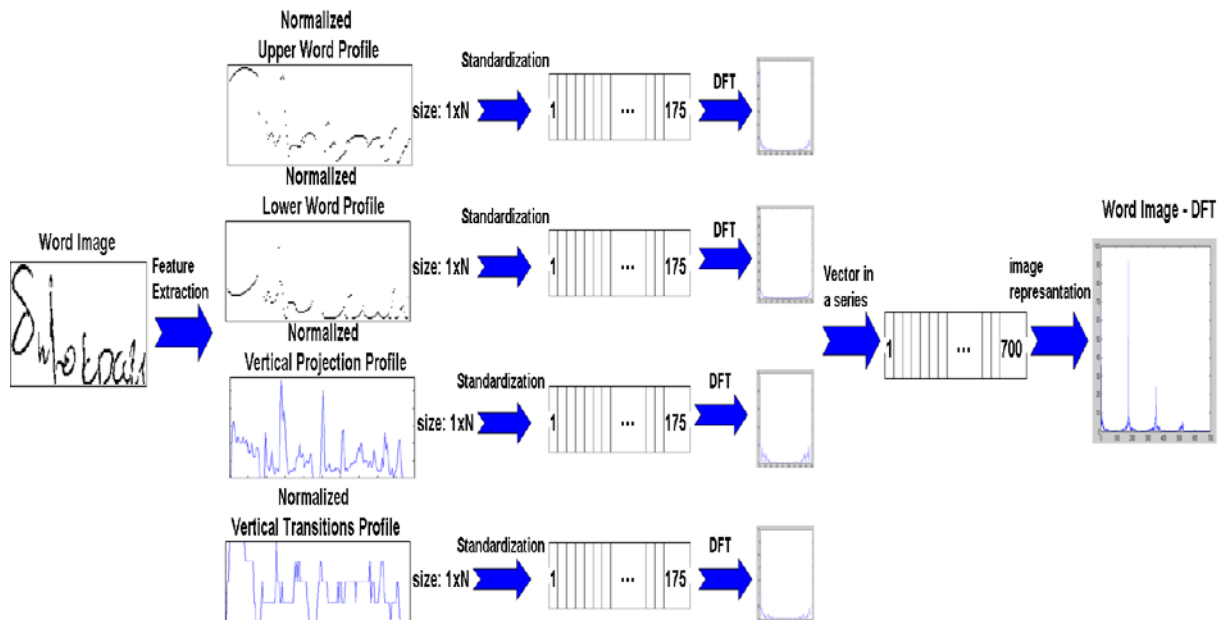
Παρακάτω δείχνουμε παράδειγμα μετασχηματισμού DFT των τεσσάρων χαρακτηριστικών:



Σχήμα 4-14: Παράδειγμα μετασχηματισμού των τεσσάρων χαρακτηριστικών σε DFT της λέξης “δημοκρατία”.

Εν τέλει τα τέσσερα αυτά μετασηματισμένα διανύσματα εισάγονται σε ένα διάνυσμα σε σειρά. Μπαίνουν δηλαδή σε ένα άλλο διάνυσμα, το οποίο πλέον αντιπροσωπεύει την εικόνα. Το διάνυσμα αυτό έχει μέγεθος 1X700 (175 X 4 μετασηματισμένα χαρακτηριστικά).

Παρακάτω δείχνουμε σχηματικά μία εικόνα η οποία έχει περάσει το στάδιο της προ-επεξεργασίας και πώς από εκεί και πέρα αναπαριστάται ως ένα διάνυσμα σε σειρά με συντελεστές Fourier:

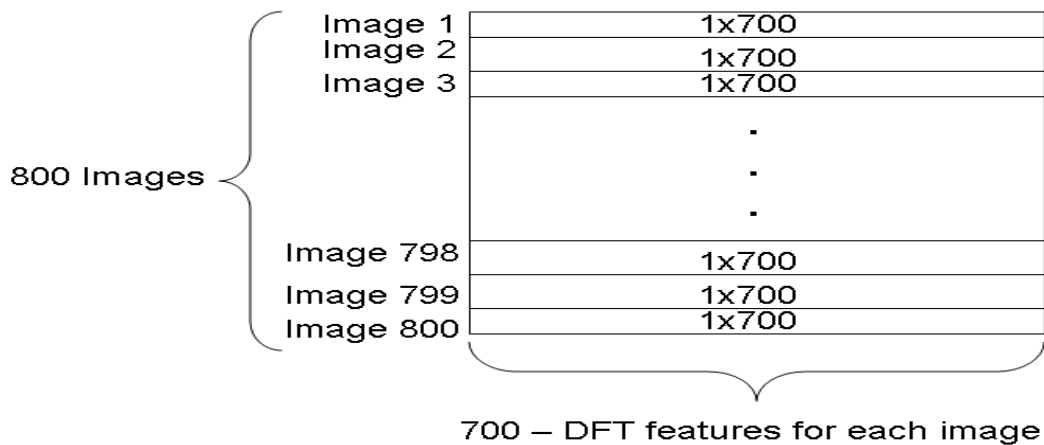


Σχήμα 4-15: Σχηματική αναπαράσταση της όλης διαδικασίας της μεθόδου 3, στην οποία η επεξεργασμένη λέξη “δημοκρατία” αναπαριστάται ως ένα διάνυσμα σε σειρά με DFT συντελεστές.

Το τελικό διάνυσμα το οποίο περιέχει σε σειρά και τα τέσσερα κανονικοποιημένα χαρακτηριστικά μεγέθους 1X700 με τους DFT συντελεστές, αναπαριστά πλήρως την εικόνα. Για κάθε διαφορετική εικόνα υπολογίζεται ένα τέτοιο διάνυσμα με DFT συντελεστές.

4.5.3.2.5 Ομαδοποίηση (Clustering)

Αρα πριν προχωρήσουμε στο στάδιο της ομαδοποίησης (Clustering), φτιάχνουμε τον πίνακα δεδομένων. Ο πίνακας δεδομένων λοιπόν, υλοποιείται εισάγοντας σε κάθε γραμμή ένα τέτοιο διάνυσμα μεγέθους 1X700 με DFT συντελεστές. Αν για παράδειγμα έχουμε 800 διαφορετικές λέξεις, ο πίνακας δεδομένων θα έχει μέγεθος 800X700, όπου οι γραμμές αναπαριστούν 800 διαφορετικές εικόνες ενώ οι 700 στήλες αναπαριστούν τα χαρακτηριστικά της συγκεκριμένης γραμμής (δηλαδή της συγκεκριμένης λέξης) με DFT συντελεστές.



Σχήμα 4-16: Παράδειγμα ενός πίνακα δεδομένων με 800 λέξεις της μεθόδου 3.

Αφού φτιαχτεί και ο πίνακας δεδομένων, μπορούμε πλέον να πραγματοποιήσουμε το στάδιο της ομαδοποίησης. Για την ομαδοποίηση των λέξεων χρησιμοποιείται ο Ward Linkage αλγόριθμος, τον οποίο έχουμε δει στο κεφάλαιο 3, ο οποίος δέχεται ως είσοδο τον πίνακα δεδομένων, τον αριθμό των ομάδων που πρόκειται να δημιουργηθούν και την μετρική απόσταση μεταξύ λέξεων, η οποία στην μέθοδο 3 είναι η ευκλείδεια απόσταση.

4.5.3.3 Αποτελέσματα (Results)

Εφαρμόζοντας την μέθοδο 3 στα τρία διαφορετικά σύνολα δεδομένων που αναφέραμε, δηλαδή τα Small set of good data, Small set of random data και Large set of good data, και αξιολογώντας τα αποτελέσματα που παίρνουμε με τα μέτρα αξιολόγησης που έχουμε δει στο κεφάλαιο 4.4, δηλαδή τα Precision, Recall, F-Measure και Purity βρίσκουμε τα παρακάτω αποτελέσματα:

Ward Linkage	Small Set of Good Data (10 Clusters)	Small Set of Random Data (10 Clusters)	Large Set of Good Data (12 Clusters)
Precision (%)	78.0267	48.4975	71.3454
Recall (%)	72.0000	45.5000	67.8571
F-Measure (%)	74.8923	46.9509	69.5575
Purity (%)	72.0000	45.5000	67.8571
Time (Minutes)	0.6323 (37.9380 Sec)	0.7499 (44.9968 Sec)	3.3565

Πίνακας 4-6: Αποτελέσματα μεθόδου 3 για τα τρία σύνολα δεδομένων. Στον χρόνο περιλαμβάνεται ο χρόνος όλων των λειτουργιών που εκτελούνται στη μέθοδο. Οι χρόνοι υπολογίστηκαν κατά την εκτέλεση της μεθόδου στη Matlab σε υπολογιστή με 2-πύρηνο επεξεργαστή 2,27 GHz.

Από τα αποτελέσματα του πίνακα 4-6, βλέπουμε γενικά καλύτερα ποσοστά για το Small Set of Good Data και Large Set of Good Data, μιας και η ποιότητα των λέξεων είναι αρκετά καλή σε σχέση με το Small Set of Random Data. Στην περίπτωση του Small Set of Random Data βλέπουμε ότι τα αποτελέσματα είναι πολύ χαμηλότερα, κάτι που είναι απόλυτα λογικό και αναμενόμενο, αφού στο Small Set of Random Data περιέχονται και λίγο υποβαθμισμένες ποιότητας εικόνες λέξεων. Όσο αφορά τώρα τους χρόνους εκτέλεσης, γενικά είναι πολύ γρήγοροι και στα τρία σύνολα δεδομένων. Συγκεκριμένα τα δύο μικρά σύνολα δεδομένων ολοκληρώνονται σε πιο λίγο από ένα λεπτό, ενώ το μεγάλο σύνολο δεδομένων ολοκληρώνεται σε τρία λεπτά.

4.5.4 Μέθοδος 4

Η μέθοδος 4 όπως έχουμε αναφέρει αφορά αποκλειστικά δική μας παραλλαγή της μεθόδου 3, η οποία αποτελείται από τα παρακάτω βήματα:

- 1) Προ-επεξεργασία (Preprocessing)
- 2) Εξαγωγή χαρακτηριστικών (Feature Extraction)
- 3) Κανονικοποίηση Χαρακτηριστικών (Normalization)
- 4) Κανονικοποίηση Μεγέθους (Standardization)
- 5) Διακριτός Μετασχηματισμός Fourier (Discrete Fourier Transform)
- 6) Ομαδοποίηση (Clustering)

4.5.4.1 Διάγραμμα ροής μεθοδολογίας

Στο σχήμα 4-17 δείχνουμε το διάγραμμα ροής που ακολουθεί η μέθοδος 4.



Σχήμα 4-17: Διάγραμμα ροής μεθοδολογίας για την μέθοδο 4.

4.5.4.2 Περιγραφή Μεθόδου

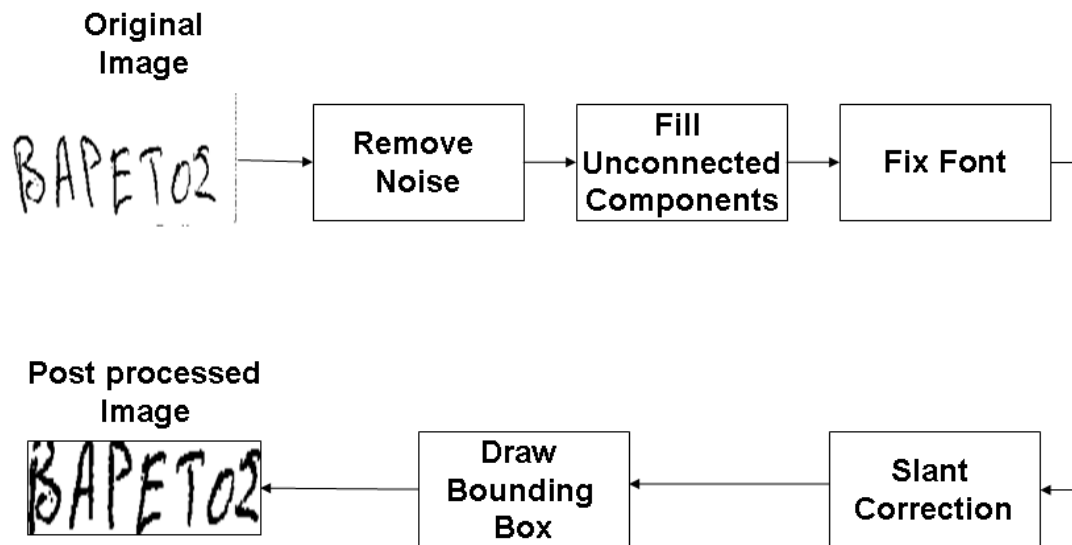
Η μέθοδος 4 όπως έχουμε πει είναι αποκλειστικά δική μας παραλλαγή της μεθόδου 3 με τροποποιήσεις στο στάδιο της προ-επεξεργασίας και της εξαγωγής χαρακτηριστικών. Ξεκινά με το στάδιο της προ-επεξεργασίας, εξαγει έντεκα χαρακτηριστικά, τα κανονικοποιεί, εκτελεί κανονικοποίηση μεγέθους, μετασχηματίζει σε DFT συντελεστές κάθε διάνυσμα χαρακτηριστικών, φτιάχνει τον πίνακα δεδομένων και εν τέλει πραγματοποιεί το Clustering με βάση τον Ward Linkage.

4.5.4.2.1 Προ-επεξεργασία

Η μέθοδος 4, αρχίζει με πρώτο στάδιο την προ-επεξεργασία κάθε χειρόγραφης εικόνας. Λαμβάνεται αρχικά μία σκαναρισμένη δυαδική χειρόγραφη λέξη (με pixels 0 και 1), η οποία υποφέρει από διάφορα προβλήματα όπως θόρυβο, ξεθώριασμα, διαφορές στο font, διαφορές στην κλίση κτλ., στην οποία χρειάζεται μία σωστή προ-επεξεργασία ώστε να διορθωθεί όσο το δυνατόν καλύτερα.

Η προ-επεξεργασία αρχίζει με πρώτο βήμα την αφαίρεση του θορύβου της λέξης, στη συνέχεια πραγματοποιείται η ένωση σπασμένων κομματιών μέσα στη λέξη, μετά διορθώνεται το font της λέξης, ακολουθεί η διόρθωση της κλίσης της και εν τέλει σχεδιάζεται το πλαίσιο οριοθέτησης (bounding box) της.

Στο σχήμα 4-18 δείχνουμε το διάγραμμα ροής που ακολουθεί η προ-επεξεργασία της μεθόδου 4:



Σχήμα 4-18: Διάγραμμα ροής προ-επεξεργασίας μεθόδου 4.

4.5.4.2.2 Εξαγωγή χαρακτηριστικών (Feature Extraction)

Αφού λοιπόν σε κάθε εικόνα έχει πραγματοποιηθεί το στάδιο της προ-επεξεργασίας, η λέξη είναι πλέον έτοιμη να περάσει στο στάδιο της εξαγωγής χαρακτηριστικών. Στο στάδιο της εξαγωγής χαρακτηριστικών εξάγονται από κάθε λέξη τα έντεκα πιο κάτω χαρακτηριστικά:

- 1) Upper Word Profile
- 2) Lower Word Profile
- 3) Left Word Profile
- 4) Right Word Profile
- 5) Lower Minus Upper Word Profile
- 6) Vertical Projection Profile
- 7) Horizontal Projection Profile
- 8) Vertical Background to Ink Transitions Profile
- 9) Horizontal Background to Ink Transitions Profile
- 10) End Points
- 11) Perimeter Points

Το κάθε ένα από τα πιο πάνω χαρακτηριστικά εκτός των End Points και Perimeter Points αναπαριστάται ως ένα διάνυσμα μιας γραμμής και N στηλών για τα κάθετα χαρακτηριστικά, όπου N είναι το πλάτος της κάθε λέξης και ως ένα διάνυσμα μιας γραμμής και M στηλών για τα οριζόντια χαρακτηριστικά, όπου M είναι το ύψος της κάθε λέξης. Τα N και M φυσικά μεταβάλλονται για κάθε λέξη. Τα End Points και Perimeter Points είναι σταθερά διανύσματα μεγέθους 1×1 μιας και περιέχουν τον αριθμό των End Points και Perimeter Points αντίστοιχα της κάθε λέξης.

4.5.4.2.3 Κανονικοποίηση (Normalization) και Κανονικοποίηση Μεγέθους (Standardization)

Στη συνέχεια, το κάθε ένα από τα πέντε κάθετα διανύσματα γραμμής μεγέθους $1 \times N$ (δηλαδή τα: Upper Word Profile, Lower Word Profile, Lower Minus Upper Word Profile, Vertical Projection Profile, Vertical Background to Ink Transitions Profile), το κάθε ένα από τα τέσσερα οριζόντια διανύσματα γραμμής μεγέθους $1 \times M$ (δηλαδή τα: Left Word Profile, Right Word Profile, Horizontal Projection Profile, Horizontal Background to Ink Transitions Profile) και τα δύο σταθερά διανύσματα μεγέθους 1×1 (δηλαδή τα: End Points, Perimeter Points), περνούν στο στάδιο της κανονικοποίησης (normalization) όπου πραγματοποιείται η κανονικοποίηση τους σε κλίμακα $[0,1]$, έτσι ώστε να μπορεί τα χαρακτηριστικά να συγκρίνονται μεταξύ τους. Η κανονικοποίηση των έντεκα αυτών χαρακτηριστικών επιτελείται ακριβώς όπως έχουμε αναφέρει στο κεφάλαιο 3 για τα συγκεκριμένα χαρακτηριστικά.

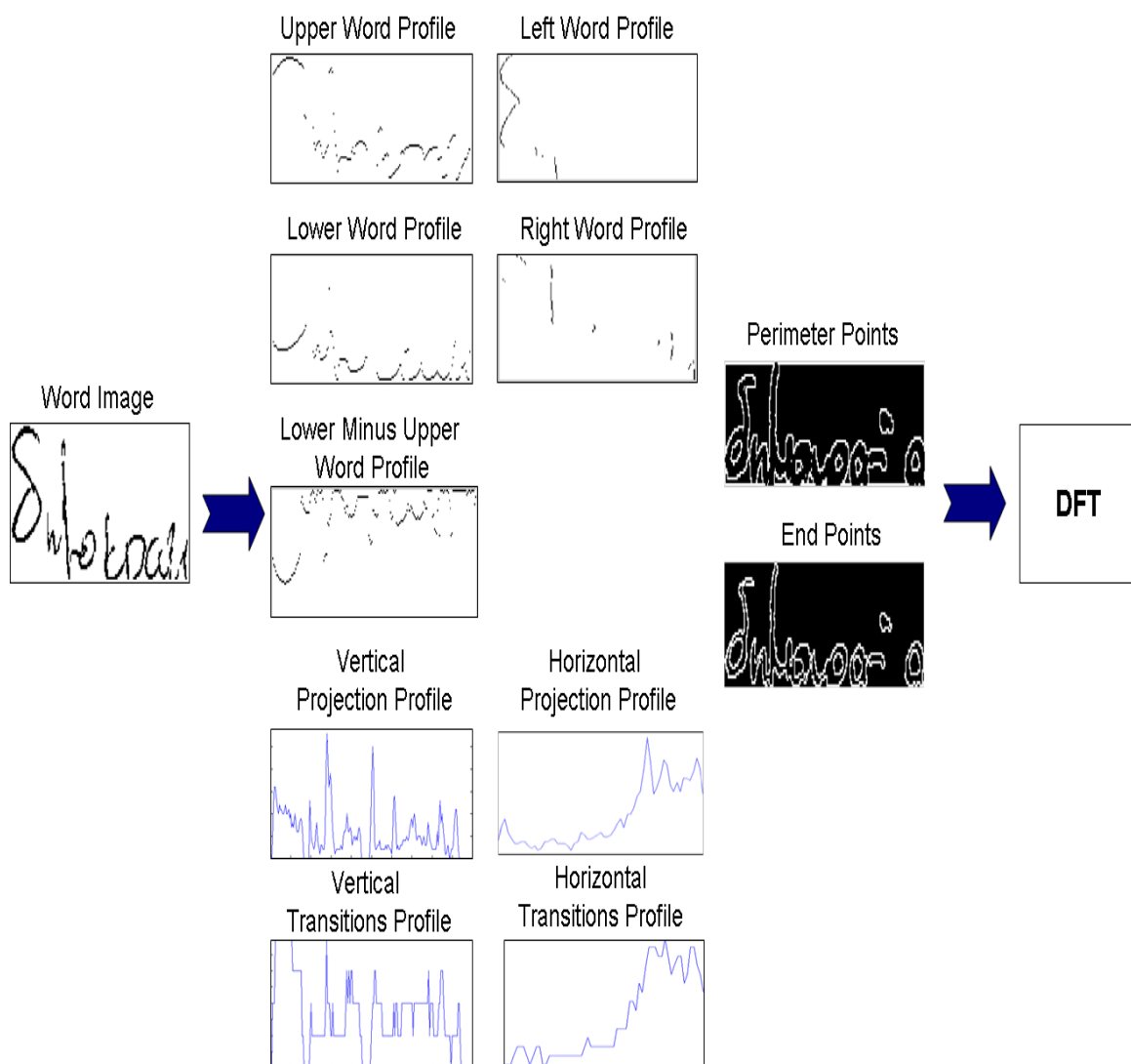
Αφού λοιπόν τα πέντε κάθετα και τα τέσσερα οριζόντια χαρακτηριστικά έχουν πλέον κανονικοποιηθεί, περνούν στο στάδιο της κανονικοποίησης μεγέθους (standardization). Εδώ όπως μιλήσαμε και στο κεφάλαιο 3, λόγω του ότι το πλάτος και το ύψος κάθε εικόνας μεταβάλλονται δεν έχουμε σταθερά μεγέθη, έτσι κάθε ένα από τα πέντε κάθετα μεταβαλλόμενα διανύσματα μεγέθους $1 \times N$ και κάθε ένα από τα τέσσερα οριζόντια μεταβαλλόμενα διανύσματα μεγέθους $1 \times M$, κανονικοποιούνται σε ένα σταθερό μέγεθος στην περίπτωση μας σε 1×175 . Τα End Points και Perimeter Points δεν χρειάζεται να περάσουν το στάδιο της κανονικοποίησης

μεγέθους μιας και έχουν από μόνα τους το σταθερό μέγεθος 1X1. Η κανονικοποίηση μεγέθους είναι πολύ σημαντική για να επιτευχθεί με επιτυχία η σύγκριση μεταξύ λέξεων.

4.5.4.2.4 Διακριτός Μετασχηματισμός Fourier (Discrete Fourier Transform – DFT)

Μετά την κανονικοποίηση μεγέθους, τα πέντε κάθετα και τα τέσσερα οριζόντια αυτά διανύσματα μεγέθους πλέον 1X175, και τα δύο σταθερού μεγέθους 1X1 διανύσματα, εισάγονται ως είσοδο στο στάδιο του διακριτού μετασχηματισμού Fourier (Discrete Fourier Transform - DFT). Εκεί τα έντεκα αυτά διανύσματα μετασχηματίζονται στο πεδίο της συχνότητας κάνοντας χρήση του μετασχηματισμού Fourier. Δηλαδή τα έντεκα αυτά διανύσματα, αναπαριστώνται πλέον ως έντεκα μετασχηματισμένα διανύσματα με συντελεστές Fourier.

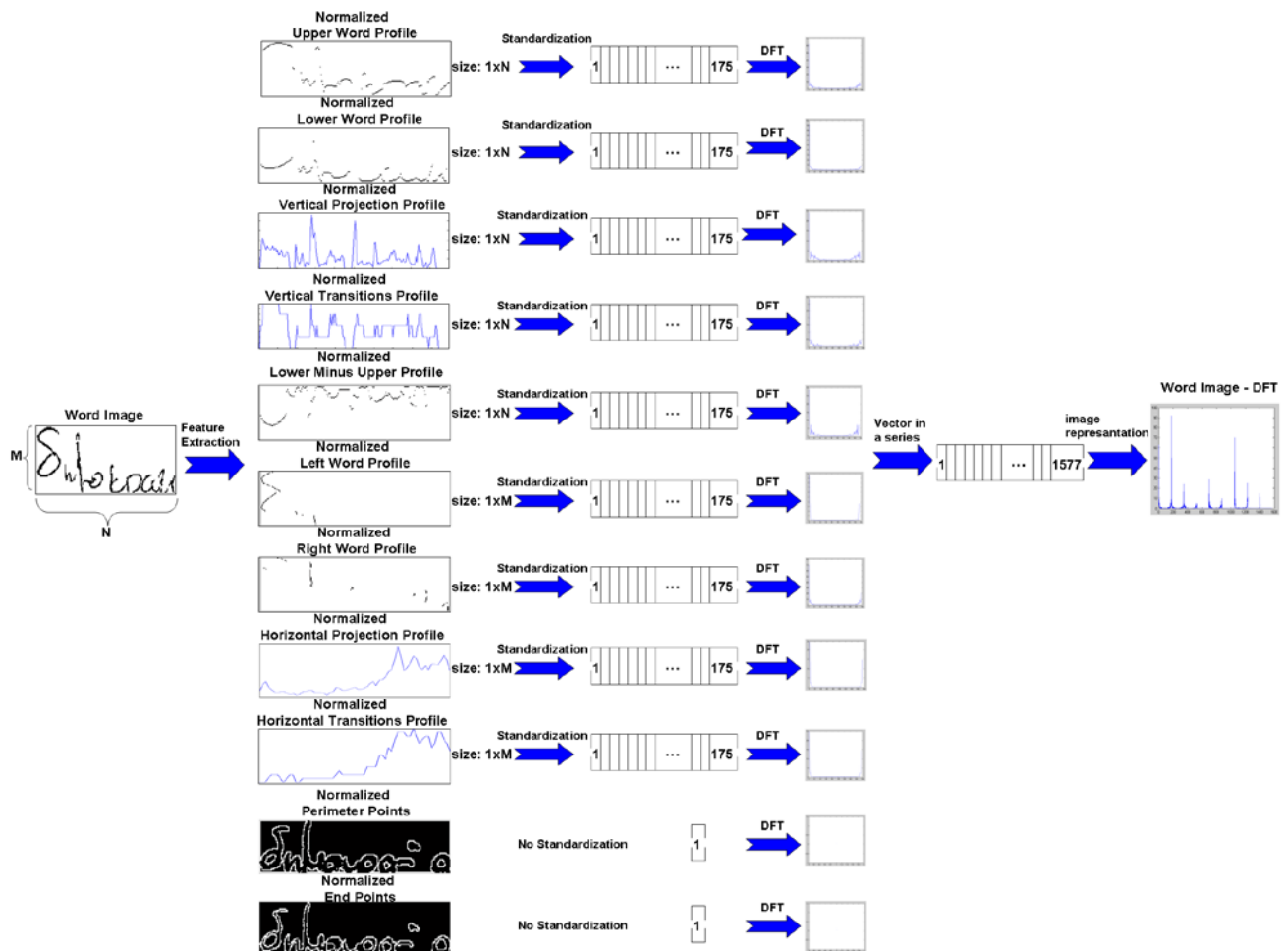
Παρακάτω δείχνουμε παράδειγμα μετασχηματισμού DFT των έντεκα χαρακτηριστικών:



Σχήμα 4-19: Παράδειγμα μετασχηματισμού των έντεκα χαρακτηριστικών σε DFT της λέξης “δημοκρατία”.

Εν τέλει τα έντεκα αυτά μετασχηματισμένα διανύσματα, εισάγονται σε ένα διάνυσμα σε σειρά. Μπαίνουν δηλαδή σε ένα άλλο διάνυσμα, το οποίο πλέον αντιπροσωπεύει την εικόνα. Το διάνυσμα αυτό έχει μέγεθος 1X1577 ((175 X 9) +2) μετασχηματισμένα χαρακτηριστικά).

Παρακάτω δείχνουμε σχηματικά μία εικόνα η οποία έχει περάσει το στάδιο της προ-επεξεργασίας και πώς από εκεί και πέρα αναπαριστάται ως ένα διάνυσμα σε σειρά με συντελεστές Fourier:

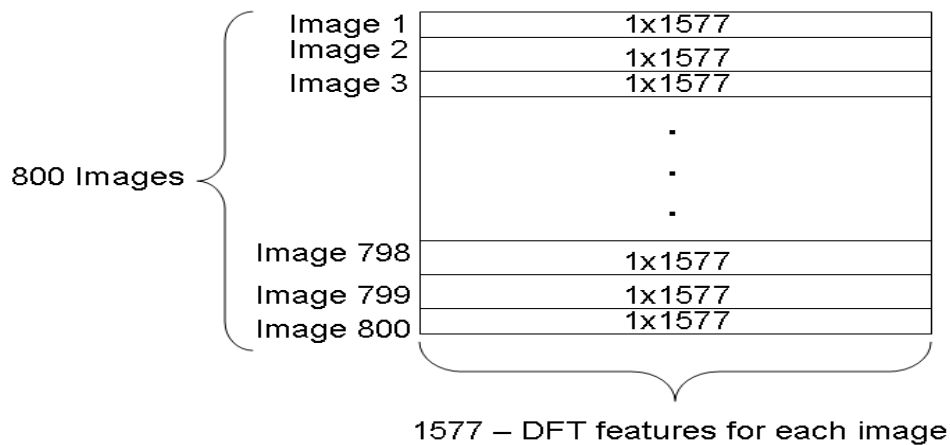


Σχήμα 4-20: Σχηματική αναπαράσταση της όλης διαδικασίας της μεθόδου 4, στην οποία η επεξεργασμένη λέξη “δημοκρατία” αναπαριστάται ως ένα διάνυσμα σε σειρά με DFT συντελεστές.

Το τελικό διάνυσμα το οποίο περιέχει σε σειρά και τα έντεκα κανονικοποιημένα χαρακτηριστικά μεγέθους 1×1577 με τους DFT συντελεστές, αναπαριστά πλήρως την εικόνα. Για κάθε διαφορετική εικόνα υπολογίζεται ένα τέτοιο διάνυσμα με DFT συντελεστές.

4.5.4.2.5 Ομαδοποίηση (Clustering)

Αρα πριν προχωρήσουμε στο στάδιο της ομαδοποίησης (Clustering), φτιάχνουμε τον πίνακα δεδομένων. Ο πίνακας δεδομένων λοιπόν υλοποιείται εισάγοντας σε κάθε γραμμή ένα τέτοιο διάνυσμα μεγέθους 1×1577 με DFT συντελεστές. Αν για παράδειγμα έχουμε 800 διαφορετικές λέξεις, ο πίνακας δεδομένων θα έχει μέγεθος 800×1577 , όπου οι γραμμές αναπαριστούν 800 διαφορετικές εικόνες ενώ οι 1577 στήλες αναπαριστούν τα χαρακτηριστικά της συγκεκριμένης γραμμής (δηλαδή της συγκεκριμένης λέξης) με DFT συντελεστές.



Σχήμα 4-21: Παράδειγμα ενός πίνακα δεδομένων με 800 λέξεις της μεθόδου 4.

Αφού φτιαχτεί και ο πίνακας δεδομένων, μπορούμε πλέον να πραγματοποιήσουμε το στάδιο της ομαδοποίησης. Για την ομαδοποίηση των λέξεων χρησιμοποιείται ο Ward Linkage αλγόριθμος, τον οποίο έχουμε δει στο κεφάλαιο 3, ο οποίος δέχεται ως είσοδο τον πίνακα δεδομένων, τον αριθμό των ομάδων που πρόκειται να δημιουργηθούν και την μετρική απόσταση μεταξύ λέξεων, η οποία στην μέθοδο 4 είναι η ευκλείδεια απόσταση.

Ο Ward Linkage επιλέχθηκε ως καλύτερος αλγόριθμος ομαδοποίησης μετά από μία σειρά συγκρίσεων που εφαρμόστηκαν για 7 αλγορίθμους ομαδοποίησης.

Στον πίνακα 4-7 φαίνονται τα αποτελέσματα που προέκυψαν κατά την εκτέλεση της μεθόδου 4, για τους 7 συγκεντρωτικούς ιεραρχικούς αλγόριθμους. Η εκτέλεση αυτή έγινε στο σύνολο δεδομένων Small Set of Good Data, με επιλεγμένη μετρική απόσταση την ευκλείδεια απόσταση.

Small Set of Good Data	Precision (%)	Recall (%)	F-Measure (%)	Purity (%)
Average Linkage	79.5974	31.0000	44.6216	31.0000
Centroid Linkage	91.1173	20.5000	33.4698	20.5000
Complete Linkage	76.0627	66.0000	70.6749	66.0000
Median Linkage	87.1562	24.5000	38.2482	24.5000
Single Linkage	91.0471	14.5000	25.0159	14.5000
Ward Linkage	84.7237	84.5000	84.6117	84.5000
Weighted Linkage	77.8831	49.5000	60.5294	49.5000

Πίνακας 4-7: Αποτελέσματα μεθόδου 4 για το Small Set of Good Data. Για κάθε ένα από τους 7 αλγόριθμους Linkage, δίνεται το ποσοστό των δεικτών αξιολόγησης: Precision, Recall, F-

Measure και Purity, ενώ η μετρική απόσταση που χρησιμοποιήθηκε εδώ ήταν η ευκλείδεια απόσταση.

Από τους αλγόριθμους που συγκρίναμε προέκυψε ότι τα καλύτερα αποτελέσματα έδινε ο Ward Linkage, όπως φαίνεται και στον πίνακα 4-7 για το Small Set of Good Data, γι' αυτό και αποφασίσαμε να χρησιμοποιηθεί αυτός σαν ο αλγόριθμος ομαδοποίησης για την μέθοδο 4.

Επίσης, δοκιμάσαμε και διάφορες άλλες μετρικές αποστάσεων κατά το στάδιο της ομαδοποίησης. Στον πίνακα 4-8 φαίνονται τα αποτελέσματα που προέκυψαν κατά την εκτέλεση της μεθόδου 4 για το Small Set of Good Data, για 10 διαφορετικές μετρικές αποστάσεων κατά την ομαδοποίηση, με αλγόριθμο ομαδοποίησης τον Ward Linkage.

Small Set of Good Data	Precision (%)	Recall (%)	F-Measure (%)	Purity (%)
Euclidean Distance	84.7237	84.5000	84.6117	84.5000
Seuclidean Distance	51.2881	44.0000	47.3653	44.0000
Cityblock Distance	86.6383	82.5000	84.5185	82.5000
Minkowski Distance	84.7237	84.5000	84.6117	84.5000
Chebychev Distance	61.6348	53.0000	56.9922	53.0000
Cosine Distance	81.9956	77.5000	79.6844	77.5000
Correlation Distance	83.2284	79.5000	81.3215	79.5000
Spearman Distance	50.2499	31.5000	38.7247	31.5000
Hamming Distance	91.0989	19.0000	31.4422	19.0000
Jaccard Distance	91.0989	19.0000	31.4422	19.0000

Πίνακας 4-8: Αποτελέσματα μεθόδου 4 για το Small Set of Good Data. Για κάθε μια από τις 10 μετρικές αποστάσεων, δίνεται το ποσοστό των δεικτών αξιολόγησης: Precision, Recall, F-Measure και Purity, ενώ ο αλγόριθμος ομαδοποίησης που χρησιμοποιήθηκε εδώ ήταν ο Ward Linkage.

Από τις μετρικές αποστάσεων που δοκιμάστηκαν, τα καλύτερα αποτελέσματα έφερε η ευκλείδεια απόσταση (Euclidean Distance) σε ισοβαθμία με την μετρική απόσταση Minkowski, όπως φαίνεται και στον πίνακα 4-8 για το Small Set of Good Data. Λόγω του ότι η ευκλείδεια απόσταση, έχει χρησιμοποιηθεί στην υπάρχουσα μέθοδο 3, επιλέχθηκε αυτή να χρησιμοποιηθεί σαν η μετρική απόσταση για την ομαδοποίηση της μεθόδου 4. Μια συνοπτική περιγραφή της κάθε μετρικής απόστασης που εμφανίζεται στον πίνακα 4-8 μπορεί να βρεθεί στο [31].

4.5.4.3 Αποτελέσματα (Results)

Εφαρμόζοντας την μέθοδο 4 στα τρία διαφορετικά σύνολα δεδομένων που αναφέραμε, δηλαδή τα Small set of good data, Small set of random data και Large set of good data, και αξιολογώντας τα αποτελέσματα που παίρνουμε με τα μέτρα αξιολόγησης που έχουμε δει στο κεφάλαιο 4.4, δηλαδή τα Precision, Recall, F-Measure και Purity βρίσκουμε τα παρακάτω αποτελέσματα:

Ward Linkage	Small Set of Good Data (10 Clusters)	Small Set of Random Data (10 Clusters)	Large Set of Good Data (12 Clusters)
Precision (%)	84.7237	58.5413	74.9404
Recall (%)	84.5000	53.5000	69.2460
F-Measure (%)	84.6117	55.9072	71.9807
Purity (%)	84.5000	53.5000	69.2460
Time (Minutes)	0.6462 (38.7767 Sec)	0.7830 (46.9809 Sec)	3.5211

Πίνακας 4-9: Αποτελέσματα μεθόδου 4 για τα τρία σύνολα δεδομένων. Στον χρόνο περιλαμβάνεται ο χρόνος όλων των λειτουργιών που εκτελούνται στη μέθοδο. Οι χρόνοι υπολογίστηκαν κατά την εκτέλεση της μεθόδου στη Matlab σε υπολογιστή με 2-πύρηνο επεξεργαστή 2,27 GHz.

Από τα αποτελέσματα του πίνακα 4-9, βλέπουμε γενικά και εδώ καλύτερα ποσοστά για το Small Set of Good Data και Large Set of Good Data μιας και η ποιότητα των λέξεων είναι αρκετά καλή σε σχέση με το Small Set of Random Data. Στην περίπτωση του Small Set of Random Data βλέπουμε ότι τα αποτελέσματα είναι χαμηλότερα, κάτι που είναι απόλυτα λογικό και αναμενόμενο, αφού στο Small Set of Random Data περιέχονται και λίγο υποβαθμισμένης ποιότητας εικόνες λέξεων. Όσο αφορά τώρα τους χρόνους εκτέλεσης, γενικά και εδώ είναι πολύ γρήγοροι και στα τρία σύνολα δεδομένων. Συγκεκριμένα, τα δύο μικρά σύνολα δεδομένων ολοκληρώνονται σε πιο λίγο από ένα λεπτό, ενώ το μεγάλο σύνολο δεδομένων ολοκληρώνεται σε 3.5 λεπτά. Να τονίσουμε στο σημείο αυτό ότι έχουμε βρει γενικά καλύτερα αποτελέσματα και στα τρία σύνολα δεδομένων σε σχέση με την ήδη υπάρχον μέθοδο 3.

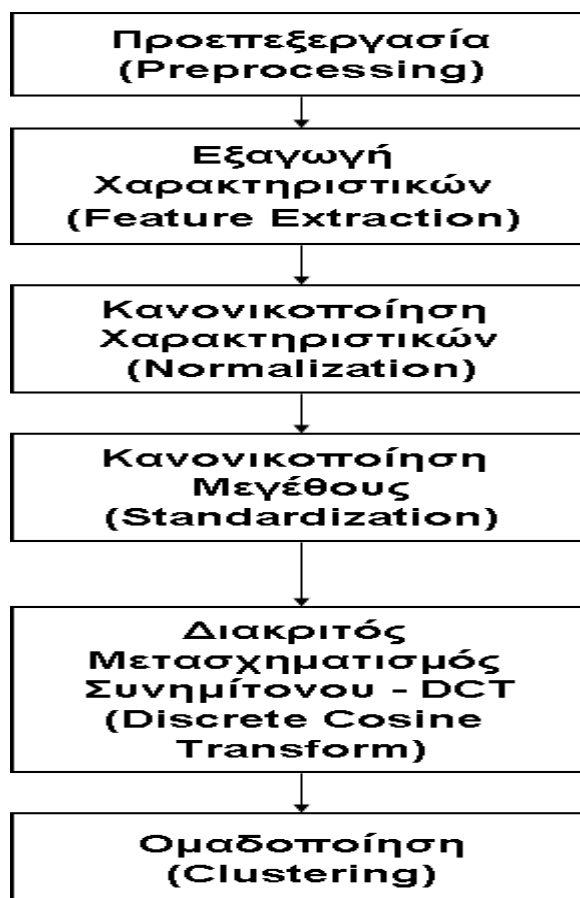
4.5.5 Μέθοδος 5

Η μέθοδος 5 όπως έχουμε αναφέρει αφορά επίσης αποκλειστικά δική μας παραλλαγή της μεθόδου 3, η οποία αποτελείται από τα παρακάτω βήματα:

- 1) Προ-επεξεργασία (Preprocessing)
- 2) Εξαγωγή χαρακτηριστικών (Feature Extraction)
- 3) Κανονικοποίηση Χαρακτηριστικών (Normalization)
- 4) Κανονικοποίηση Μεγέθους (Standardization)
- 5) Διακριτός Μετασχηματισμός Συνημίτονου (Discrete Cosine Transform)
- 6) Ομαδοποίηση (Clustering)

4.5.5.1 Διάγραμμα ροής μεθοδολογίας

Στο σχήμα 4-22 δείχνουμε το διάγραμμα ροής που ακολουθεί η μέθοδος 5.



Σχήμα 4-22: Διάγραμμα ροής μεθοδολογίας για την μέθοδο 5.

4.5.5.2 Περιγραφή Μεθόδου

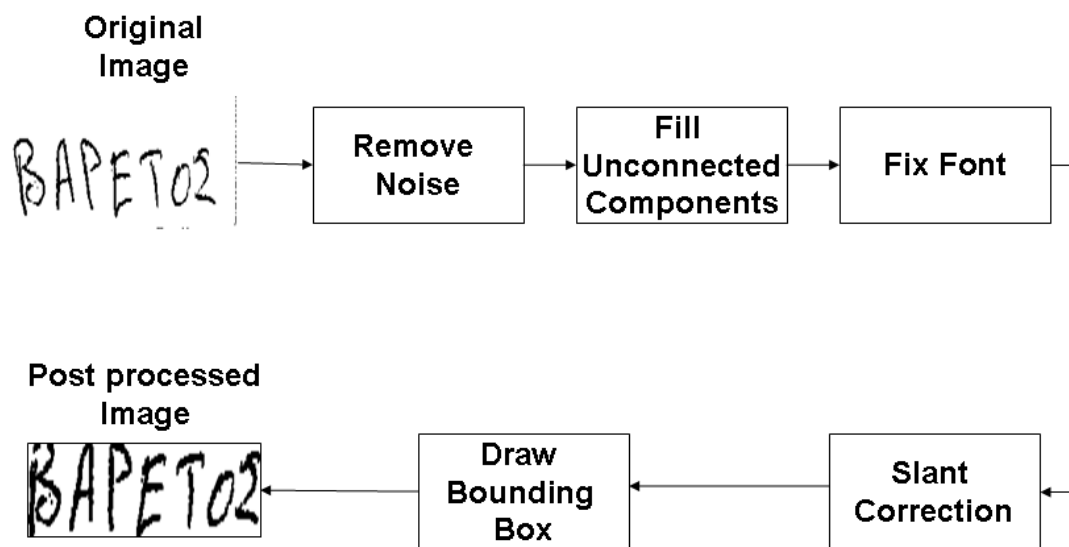
Η μέθοδος 5 όπως έχουμε πει είναι αποκλειστικά δική μας παραλλαγή της μεθόδου 3 με τροποποιήσεις στο στάδιο της προ-επεξεργασίας και της εξαγωγής χαρακτηριστικών, ακριβώς αυτές που έχουμε εφαρμόσει στη μέθοδο 4, μόνο που αυτή τη φορά στη θέση του διακριτού μετασχηματισμού Fourier, χρησιμοποιείται ο διακριτός μετασχηματισμός Συνημίτονου. Ξεκινά με το στάδιο της προ-επεξεργασίας, εξάγει έντεκα χαρακτηριστικά, τα κανονικοποιεί, εκτελεί κανονικοποίηση μεγέθους, μετασχηματίζει αυτή τη φορά σε DCT συντελεστές κάθε διάνυσμα χαρακτηριστικών, φτιάχνει τον πίνακα δεδομένων και εν τέλει πραγματοποιεί το Clustering με βάση τον Ward Linkage.

4.5.5.2.1 Προ-επεξεργασία

Η μέθοδος 5, αρχίζει με πρώτο στάδιο την προ-επεξεργασία κάθε χειρόγραφης εικόνας. Λαμβάνεται αρχικά μία σκαναρισμένη δυαδική χειρόγραφη λέξη (με pixels 0 και 1) η οποία υποφέρει από διάφορα προβλήματα όπως θόρυβος, ξεθώριασμα, διαφορές στο font, διαφορές στην κλίση κτλ., στην οποία χρειάζεται μία σωστή προ-επεξεργασία ώστε να διορθωθεί όσο το δυνατόν καλύτερα.

Η προ-επεξεργασία της μεθόδου 5 είναι ακριβώς ίδια με τη μέθοδο 4, αρχίζει δηλαδή με πρώτο βήμα την αφαίρεση του θορύβου της λέξης, στη συνέχεια πραγματοποιείται η ένωση σπασμένων κομματιών μέσα στη λέξη, μετά διορθώνεται το font της λέξης, ακολουθεί η διόρθωση της κλίσης της και εν τέλει σχεδιάζεται το πλαίσιο οριοθέτησης (bounding box) της.

Στο σχήμα 4-23 δείχνουμε το διάγραμμα ροής που ακολουθεί η προ-επεξεργασία της μεθόδου 5:



Σχήμα 4-23: Διάγραμμα ροής προ-επεξεργασίας μεθόδου 5.

4.5.5.2.2 Εξαγωγή χαρακτηριστικών (Feature Extraction)

Αφού λοιπόν σε κάθε εικόνα έχει πραγματοποιηθεί το στάδιο της προ-επεξεργασίας, η λέξη είναι πλέον έτοιμη να περάσει στο στάδιο της εξαγωγής χαρακτηριστικών. Στο στάδιο της εξαγωγής χαρακτηριστικών εξάγονται από κάθε λέξη τα έντεκα πιο κάτω χαρακτηριστικά τα οποία είναι αυτά που χρησιμοποιούνται και στη μέθοδο 4, δηλαδή τα:

- 1) Upper Word Profile
- 2) Lower Word Profile
- 3) Left Word Profile
- 4) Right Word Profile
- 5) Lower Minus Upper Word Profile
- 6) Vertical Projection Profile
- 7) Horizontal Projection Profile
- 8) Vertical Background to Ink Transitions Profile
- 9) Horizontal Background to Ink Transitions Profile
- 10) End Points
- 11) Perimeter Points

Το κάθε ένα από τα πιο πάνω χαρακτηριστικά εκτός των End Points και Perimeter Points αναπαριστάται ως ένα διάνυσμα μιας γραμμής και N στηλών για τα κάθετα χαρακτηριστικά, όπου N είναι το πλάτος της κάθε λέξης και ως ένα διάνυσμα μιας γραμμής και M στηλών για τα οριζόντια χαρακτηριστικά, όπου M είναι το ύψος της κάθε λέξης. Τα N και M φυσικά μεταβάλλονται για κάθε λέξη. Τα End Points και Perimeter Points είναι σταθερά διανύσματα μεγέθους 1×1 μιας και περιέχουν τον αριθμό των End Points και Perimeter Points αντίστοιχα της κάθε λέξης.

4.5.5.2.3 Κανονικοποίηση (Normalization) και Κανονικοποίηση Μεγέθους (Standardization)

Στη συνέχεια το κάθε ένα από τα πέντε κάθετα διανύσματα γραμμής μεγέθους $1 \times N$ (δηλαδή τα: Upper Word Profile, Lower Word Profile, Lower Minus Upper Word Profile, Vertical Projection Profile, Vertical Background to Ink Transitions Profile), το κάθε ένα από τα τέσσερα οριζόντια διανύσματα γραμμής μεγέθους $1 \times M$ (δηλαδή τα: Left Word Profile, Right Word Profile, Horizontal Projection Profile, Horizontal Background to Ink Transitions Profile) και τα δύο σταθερά διανύσματα μεγέθους 1×1 (δηλαδή τα: End Points, Perimeter Points) περνούν στο στάδιο της κανονικοποίησης (normalization) όπου πραγματοποιείται η κανονικοποίηση τους σε κλίμακα $[0,1]$, έτσι ώστε να μπορεί τα χαρακτηριστικά να συγκρίνονται με άλλες λέξεις. Η κανονικοποίηση των έντεκα αυτών χαρακτηριστικών επιτελείται ακριβώς όπως έχουμε αναφέρει στο κεφάλαιο 3 για τα συγκεκριμένα χαρακτηριστικά.

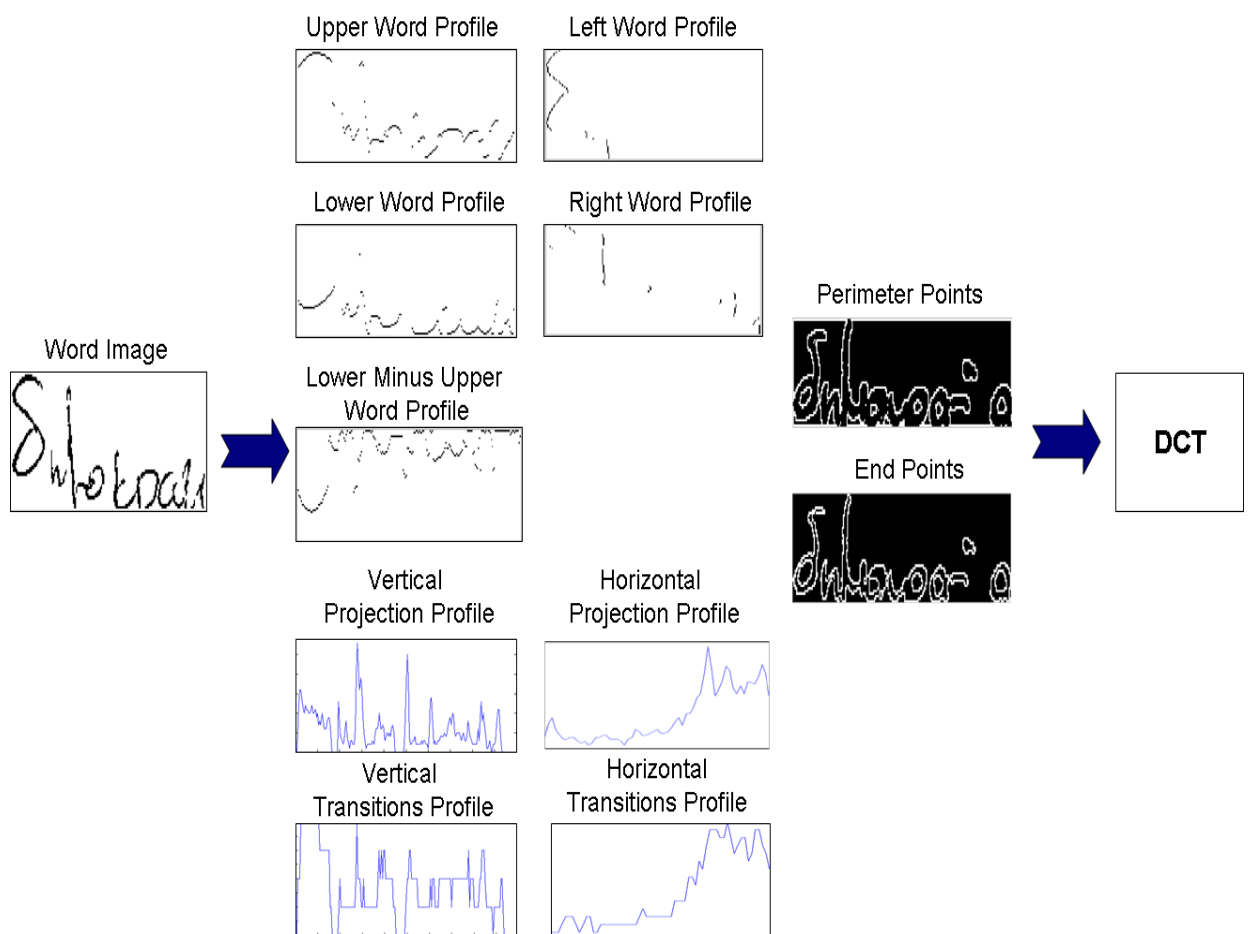
Αφού λοιπόν τα πέντε κάθετα και τα τέσσερα οριζόντια χαρακτηριστικά έχουν πλέον κανονικοποιηθεί, περνούν στο στάδιο της κανονικοποίησης μεγέθους (standardization). Εδώ όπως μιλήσαμε και στο κεφάλαιο 3, λόγω του ότι το πλάτος και το ύψος κάθε εικόνας μεταβάλλονται δεν έχουμε σταθερά μεγέθη, έτσι κάθε ένα από τα πέντε κάθετα μεταβαλλόμενα διανύσματα μεγέθους $1 \times N$ και κάθε ένα από τα τέσσερα οριζόντια μεταβαλλόμενα διανύσματα μεγέθους $1 \times M$, κανονικοποιούνται σε ένα σταθερό μέγεθος στην περίπτωση μας σε 1×175 . Τα

End Points και Perimeter Points δεν χρειάζεται να περάσουν το στάδιο της κανονικοποίησης μεγέθους, μιας και έχουν από μόνα τους το σταθερό μέγεθος 1X1. Η κανονικοποίηση μεγέθους είναι πολύ σημαντική για να επιτευχθεί με επιτυχία η σύγκριση μεταξύ λέξεων.

4.5.5.2.4 Διακριτός Μετασχηματισμός Συνημίτονου (Discrete Cosine Transform – DCT)

Μετά την κανονικοποίηση μεγέθους, τα πέντε κάθετα και τα τέσσερα οριζόντια αυτά διανύσματα μεγέθους πλέον 1X175, και τα δύο σταθερού μεγέθους 1X1 διανύσματα, εισάγονται ως είσοδο στο στάδιο του διακριτού μετασχηματισμού Συνημίτονου (Discrete Cosine Transform - DCT). Εκεί, τα έντεκα αυτά διανύσματα μετασχηματίζονται στο πεδίο της συχνότητας κάνοντας χρήση του μετασχηματισμού Συνημίτονου. Δηλαδή τα έντεκα αυτά διανύσματα, αναπαριστώνται πλέον ως έντεκα μετασχηματισμένα διανύσματα με συντελεστές Συνημίτονων.

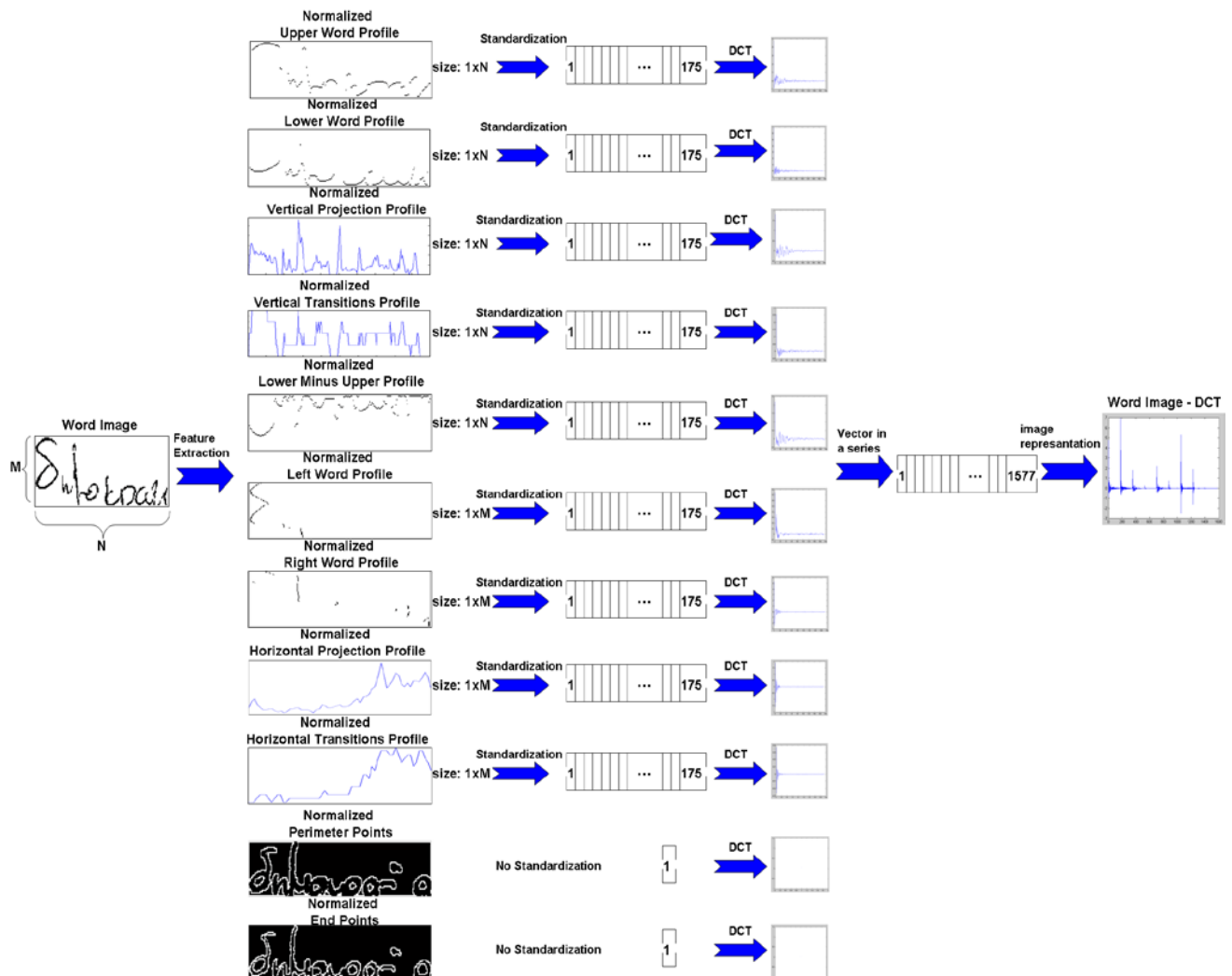
Παρακάτω δείχνουμε παράδειγμα μετασχηματισμού DCT των έντεκα χαρακτηριστικών:



Σχήμα 4-24: Παράδειγμα μετασχηματισμού των έντεκα χαρακτηριστικών σε DCT της λέξης “δημοκρατία”.

Εν τέλει τα έντεκα αυτά μετασχηματισμένα διανύσματα εισάγονται σε ένα διάνυσμα σε σειρά. Μπαίνουν δηλαδή σε ένα άλλο διάνυσμα, το οποίο πλέον αντιπροσωπεύει την εικόνα. Το διάνυσμα αυτό έχει μέγεθος 1X1577 ((175 X 9) + 2) μετασχηματισμένα χαρακτηριστικά).

Παρακάτω δείχνουμε σχηματικά μία εικόνα η οποία έχει περάσει το στάδιο της προ-επεξεργασίας και πώς από εκεί και πέρα αναπαριστάται ως ένα διάνυσμα σε σειρά με συντελεστές Συνημίτονων:

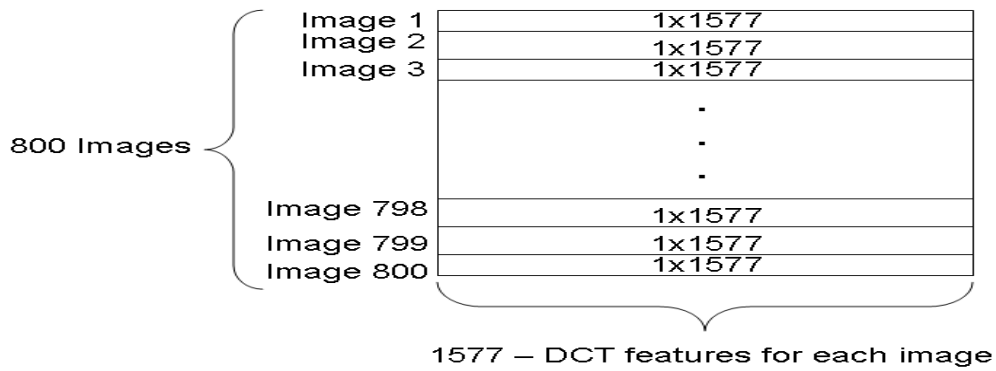


Σχήμα 4-25: Σχηματική αναπαράσταση της όλης διαδικασίας της μεθόδου 5, στην οποία η επεξεργασμένη λέξη “δημοκρατία” αναπαριστάται ως ένα διάνυσμα σε σειρά με DCT συντελεστές.

Το τελικό διάνυσμα το οποίο περιέχει σε σειρά και τα έντεκα κανονικοποιημένα χαρακτηριστικά μεγέθους 1×1577 με τους DCT συντελεστές, αναπαριστά πλήρως την εικόνα. Για κάθε διαφορετική εικόνα υπολογίζεται ένα τέτοιο διάνυσμα με DCT συντελεστές.

4.5.5.2.5 Ομαδοποίηση (Clustering)

Άρα πριν προχωρήσουμε στο στάδιο της ομαδοποίησης (Clustering), φτιάχνουμε τον πίνακα δεδομένων. Ο πίνακας δεδομένων λοιπόν υλοποιείται εισάγοντας σε κάθε γραμμή ένα τέτοιο διάνυσμα μεγέθους 1×1577 με DCT συντελεστές. Αν για παράδειγμα έχουμε 800 διαφορετικές λέξεις, ο πίνακας δεδομένων θα έχει μέγεθος 800×1577 , όπου οι γραμμές αναπαριστούν 800 διαφορετικές εικόνες ενώ οι 1577 στήλες αναπαριστούν τα χαρακτηριστικά της συγκεκριμένης γραμμής (δηλαδή της συγκεκριμένης λέξης) με DCT συντελεστές.



Σχήμα 4-26: Παράδειγμα ενός πίνακα δεδομένων με 800 λέξεις της μεθόδου 5.

Αφού φτιαχτεί και ο πίνακας δεδομένων μπορούμε πλέον να πραγματοποιήσουμε το στάδιο της ομαδοποίησης. Για την ομαδοποίηση των λέξεων χρησιμοποιείται ο Ward Linkage αλγόριθμος, τον οποίο έχουμε δει στο κεφάλαιο 3, ο οποίος δέχεται ως είσοδο τον πίνακα δεδομένων, τον αριθμό των ομάδων που πρόκειται να δημιουργηθούν και την μετρική απόσταση μεταξύ λέξεων, η οποία στην μέθοδο 5 είναι η ευκλείδεια απόσταση. Ο Ward Linkage και η μετρική απόσταση, ευκλείδεια απόσταση, επιλέχθηκαν ως οι καλύτεροι, με βάση τα αποτελέσματα που πήραμε στην μέθοδο 4 για όλους τους αλγορίθμους ομαδοποίησης και μετρικών αποστάσεων.

4.5.5.3 Αποτελέσματα (Results)

Εφαρμόζοντας την μέθοδο 5 στα τρία διαφορετικά σύνολα δεδομένων που αναφέραμε, δηλαδή τα Small set of good data, Small set of random data και Large set of good data, και αξιολογώντας τα αποτελέσματα που παίρνουμε με τα μέτρα αξιολόγησης που έχουμε δει στο κεφάλαιο 4.4, δηλαδή τα Precision, Recall, F-Measure και Purity βρίσκουμε τα παρακάτω αποτελέσματα:

Ward Linkage	Small Set of Good Data (10 Clusters)	Small Set of Random Data (10 Clusters)	Large Set of Good Data (12 Clusters)
Precision (%)	94.4203	63.3481	96.6526
Recall (%)	93.0000	53.0000	95.6349
F-Measure (%)	93.7047	57.7139	96.1410
Purity (%)	93.0000	53.0000	95.6349
Time (Minutes)	0.6458 (38.7494 Sec)	0.7714 (46.2851 Sec)	3.5419

Πίνακας 4-10: Αποτελέσματα μεθόδου 5 για τα τρία σύνολα δεδομένων. Στον χρόνο περιλαμβάνεται ο χρόνος όλων των λειτουργιών που εκτελούνται στη μέθοδο. Οι χρόνοι υπολογίστηκαν κατά την εκτέλεση της μεθόδου στη Matlab σε υπολογιστή με 2-πύρηνο επεξεργαστή 2,27 GHz.

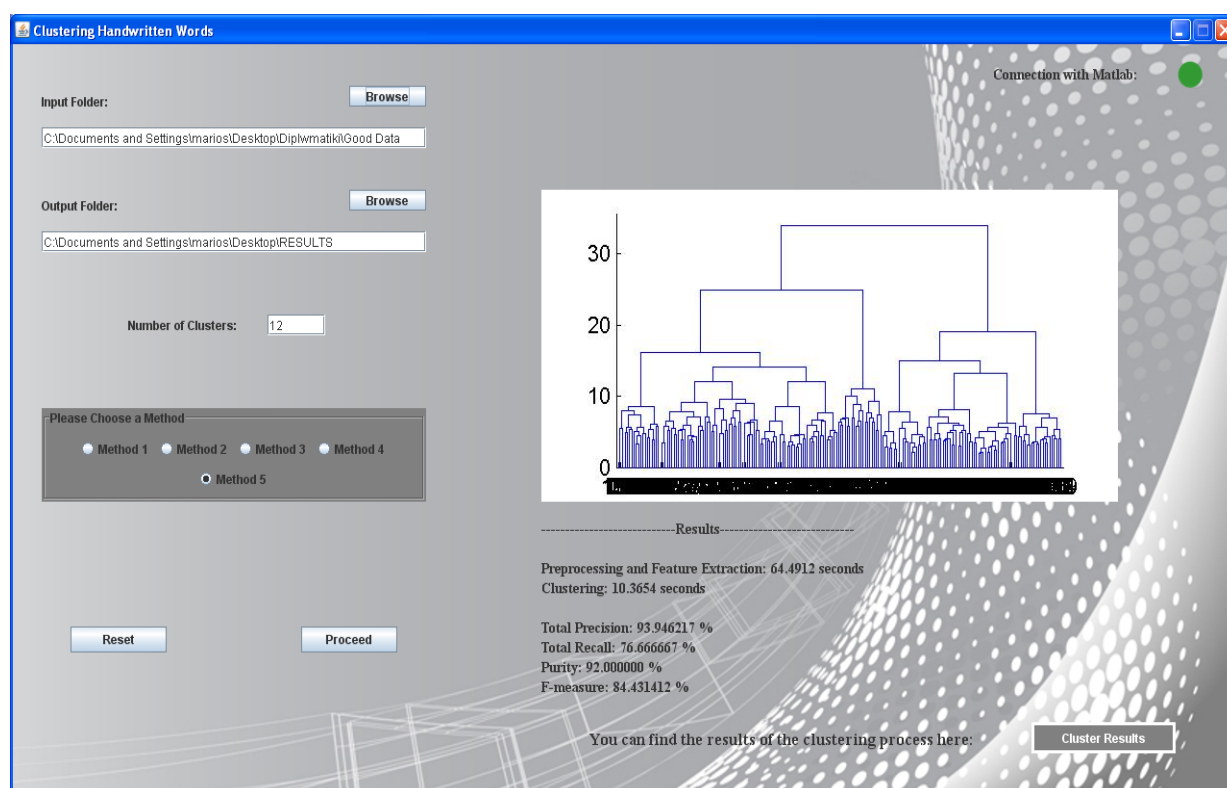
Από τα αποτελέσματα του πίνακα 4-10 βλέπουμε επίσης και εδώ καλύτερα ποσοστά για το Small Set of Good Data και Large Set of Good Data μιας και η ποιότητα των λέξεων είναι αρκετά καλή σε σχέση με το Small Set of Random Data. Στην περίπτωση του Small Set of Random Data, βλέπουμε ότι τα αποτελέσματα είναι λίγο χαμηλότερα, κάτι που είναι απόλυτα λογικό και αναμενόμενο, αφού στο Small Set of Random Data περιέχονται και λίγο υποβαθμισμένης ποιότητας εικόνες λέξεων. Όσο αφορά τώρα τους χρόνους εκτέλεσης, γενικά και εδώ είναι πολύ γρήγοροι και στα τρία σύνολα δεδομένων. Συγκεκριμένα τα δύο μικρά σύνολα δεδομένων ολοκληρώνονται σε πιο λίγο από ένα λεπτό, ενώ το μεγάλο σύνολο δεδομένων ολοκληρώνεται σε 3.5 λεπτά. Να τονίσουμε στο σημείο αυτό, ότι έχουμε βρει γενικά ακόμα καλύτερα αποτελέσματα σε αυτή τη μέθοδο και στα τρία σύνολα δεδομένων σε σχέση με την ήδη υπάρχον μέθοδο 3, καθώς και σε σχέση με τη νέα μέθοδο 4. Αυτό οφείλεται λόγω του ότι οι συντελεστές Συνημίτονων (DCT) έχουν την εξαιρετική ικανότητα να συγκεντρώνουν το μεγαλύτερο μέρος της πληροφορίας που μετασχηματίζουν, στους πρώτους όρους σε σχέση με τους συντελεστές Fourier (DFT). Επίσης να σημειώσουμε την τεράστια πρόοδο των αποτελεσμάτων της μεθόδου αυτής σε σχέση με τη αρχική μέθοδο 3.

4.6 Διασύνδεση Μεθόδων σε Graphical User Interface (GUI)

Λόγω του ότι οι πέντε μέθοδοι που έχουμε δει είναι υλοποιημένοι σε γλώσσα προγραμματισμού της Matlab, και λόγω του ότι υπάρχουν χρήστες οι οποίοι δεν είναι σχετικοί με το αντικείμενο της Matlab, έχουμε υλοποιήσει ένα Graphical User Interface (GUI), με τη χρήση της γλώσσας προγραμματισμού Java. Για ευκολία του χρήστη, έχουμε ενσωματώσει τις πέντε μεθόδους που υλοποιήσαμε, σε ένα φιλικό γραφικό περιβάλλον (Interface), από όπου ο χρήστης θα έχει τη δυνατότητα να δει τα αποτελέσματα των παραπάνω μεθόδων με τέσσερα απλά βήματα. Πρώτο βήμα να επιλέξει το μονοπάτι στο οποίο βρίσκεται το σύνολο δεδομένων, δεύτερο βήμα να επιλέξει το μονοπάτι στο οποίο πρόκειται να μπουν τα αποτελέσματα, τρίτο βήμα να επιλέξει τον αριθμό των ομάδων που πρόκειται να δημιουργηθούν και τέταρτο βήμα να επιλέξει μία από τις πέντε παραπάνω μεθόδους, στην οποία ενδιαφέρεται να χρησιμοποιήσει. Για ευκολία κατανόησης της κάθε μεθόδου, κατά την επιλογή της αντίστοιχης μεθόδου αναγράφονται δύο λόγια για το πώς λειτουργεί η επιλεγθείσα μέθοδος. Όσο αφορά τα αποτελέσματα των μέτρων αξιολόγησης όπως (Precision, Recall, F-Measure, Purity) εμφανίζονται στη διεπαφή του χρήστη κατά την ολοκλήρωση της αντίστοιχης μεθόδου, καθώς επίσης αναγράφονται και σε αρχείο το οποίο βρίσκεται μαζί με τις ομάδες στο μονοπάτι στο οποίο έχουν εισαχθεί τα αποτελέσματα.

Να σημειώσουμε επίσης ότι ο χρήστης έχει φυσικά το δικαίωμα εκτός των συνόλων δεδομένων που έχουμε δει (δηλαδή τα Small Set of Good Data, Small Set of Random Data, Large Set of Good Data) να εισάγει ως είσοδο το δικό του σύνολο δεδομένων και να δει τα δικά του αποτελέσματα ομαδοποίησης.

Στο σχήμα 4-27 δείχνουμε το Graphical User Interface το οποίο έχουμε υλοποιήσει:



Σχήμα 4-27: Graphical User Interface σε εφαρμογή.

ΚΕΦΑΛΑΙΟ 5 - ΑΠΟΤΕΛΕΣΜΑΤΑ ΚΑΙ ΣΥΜΠΕΡΑΣΜΑΤΑ

5.1 Συγκριτικά Αποτελέσματα Μεθόδων

Παρακάτω δείχνουμε συγκριτικά αποτελέσματα και των πέντε μεθόδων για τα τρία σύνολα δεδομένων μας. Συγκεκριμένα στον πίνακα 5-1 είναι τα συγκριτικά αποτελέσματα για το σύνολο Small set of good data, στον πίνακα 5-2 είναι τα συγκριτικά αποτελέσματα για το σύνολο Small set of random data και στον πίνακα 5-3 είναι τα συγκριτικά αποτελέσματα για το σύνολο Large set of good data.

SMALL SET OF GOOD DATA (200 GOOD IMAGES)					
Ward (10 Clusters)	Method 1	Method 2	Method 3	Method 4	Method 5
Precision (%)	79.9755	93.7518	78.0267	84.7237	94.4203
Recall (%)	37.5000	93.0000	72.0000	84.5000	93.0000
F-Measure (%)	51.0588	93.3744	74.8923	84.6117	93.7047
Purity (%)	37.5000	93.0000	72.0000	84.5000	93.0000
Time (Minutes)	19.1207	10.9238	0.6323	0.6462	0.6458

Πίνακας 5-1: Συγκριτικά αποτελέσματα Μεθόδων 1,2,3,4 και 5 για το Small Set of Good Data.

SMALL SET OF RANDOM DATA (200 RANDOM IMAGES)					
Ward (10 Clusters)	Method 1	Method 2	Method 3	Method 4	Method 5
Precision (%)	72.4071	75.3480	48.4975	58.5413	63.3481
Recall (%)	37.0000	71.5000	45.5000	53.5000	53.0000
F-Measure (%)	48.9742	73.3736	46.9509	55.9072	57.7139
Purity (%)	37.0000	71.5000	45.5000	53.5000	53.0000
Time (Minutes)	29.1244	15.5273	0.7499	0.7830	0.7714

Πίνακας 5-2: Συγκριτικά αποτελέσματα Μεθόδων 1,2,3,4 και 5 για το Small Set of Random Data.

LARGE SET OF GOOD DATA (1008 GOOD IMAGES)					
Ward (12 Clusters)	Method 1	Method 2	Method 3	Method 4	Method 5
Precision (%)	79.7755	91.2443	71.3454	74.9404	96.6526
Recall (%)	16.5674	87.7976	67.8571	69.2460	95.6349
F-Measure (%)	27.4369	89.4878	69.5575	71.9807	96.1410
Purity (%)	16.5674	87.7976	67.8571	69.2460	95.6349
Time (Minutes)	539.8002	297.2059	3.3565	3.5211	3.5419

Πίνακας 5-3: Συγκριτικά αποτελέσματα Μεθόδων 1,2,3,4 και 5 για το Large Set of Good Data.

5.2 Παρατηρήσεις και συμπεράσματα

Με βάση τα συγκριτικά αποτελέσματα και των πέντε μεθόδων συμπεραίνουμε τα παρακάτω:

1. Οι τρεις νέες μέθοδοι που δημιουργήθηκαν για το σκοπό της εργασίας, ανέδειξαν καλύτερα αποτεύματα από τις δύο ήδη υπάρχουσες μεθόδους σε όλα τα σύνολα δεδομένων που εφαρμόστηκαν. Συγκεκριμένα η νέα μέθοδος 2 ανέδειξε καλύτερα αποτελέσματα από την ήδη υπάρχουσα μέθοδο 1, ενώ οι νέες μέθοδοι 4 και 5 ανέδειξαν καλύτερα αποτελέσματα από την ήδη υπάρχουσα μέθοδο 3. Επίσης, η νέα μέθοδος 2 ρίχνει περίπου στο μισό το χρόνο εκτέλεσης σε σχέση με την μέθοδο 1. Επομένως, ο σκοπός της εργασίας επιτεύχθηκε.
2. Στα σύνολα δεδομένων καλής ποιότητας λέξεων όπως είναι τα Small Set of Good Data και Large Set of Good Data, η νέα μέθοδος 5 η οποία κάνει χρήση του DCT αποδίδει καλύτερα σε σύγκριση με τις υπόλοιπες μεθόδους. Από την άλλη, στο σύνολο δεδομένων τυχαίων λέξεων όπως είναι το Small Set of Random Data, η νέα μέθοδος 2 η οποία κάνει χρήση του DTW αποδίδει καλύτερα σε σύγκριση με τις υπόλοιπες μεθόδους. Αυτό μας οδηγεί στο συμπέρασμα ότι, σε περιπτώσεις που τα δεδομένα είναι καλής ποιότητας, προτιμάται η νέα μέθοδος 5 η οποία κάνει χρήση του DCT, καθώς είναι αρκετά γρήγορη και επιφέρει τα καλύτερα αποτελέσματα από όλες τις υπόλοιπες. Αν έχουμε υποβαθμισμένα δεδομένα ίσως η χρήση της μεθόδου 2 να είναι κάπως πιο χρήσιμη.
3. Γενικά, οι μέθοδοι 1 και 2 παρουσιάζουν πιο σταθερές τιμές στα τρία σύνολα δεδομένων, από τις μεθόδους 3, 4 και 5. Αυτή η παρατήρηση μας οδηγεί στο συμπέρασμα ότι η χρήση του DTW, οδηγεί σε πιο σταθερές αποδόσεις σε σχέση με την ποιότητα των δεδομένων από ότι η χρήση των DFT και DCT.
4. Ο DTW είναι γενικά χρονοβόρος, έτσι πολλές φορές είναι απαγορευτική η χρήση του σε μεγάλα σύνολα δεδομένων, ενώ ο DFT και ο DCT μπορούν εύκολα να εφαρμοστούν σε μεγάλα σύνολα δεδομένων αφού γενικά είναι αρκετά γρήγοροι.
5. Γενικά τα αποτελέσματα μπορούν να βελτιωθούν με την εφαρμογή κάποιας διαφορετικής προ-επεξεργασίας των λέξεων και συνδυασμό κάποιων άλλων χαρακτηριστικών. Επίσης, η χρήση μίας τεχνικής που θα απορρίπτει πολύ υποβαθμισμένες λέξεις (pruning) σε σύνολα δεδομένων που περιλαμβάνουν αρκετές υποβαθμισμένες λέξεις, πιθανόν να φέρει καλύτερα αποτελέσματα σε τέτοιου είδους σύνολα.
6. Βασικότατο σημείο που μπορεί να βελτιώσει γενικά την απόδοση και τα αποτελέσματα του συστήματος, είναι η χρήση συνόλου δεδομένων του οποίου οι λέξεις θα αποτελούνται από ένα ή λίγους συγγραφείς (θα είναι δηλαδή οι χειρόγραφες λέξεις προς ομαδοποίηση γραμμένες από ένα ή λίγα άτομα). Σε κάθε περίπτωση οι προτεινόμενες μέθοδοι δίνουν αρκετά καλά αποτελέσματα και για λέξεις πολλών συγγραφέων.

BIBΛΙΟΓΡΑΦΙΑ

- [1] T.Konidakis, B.Gatos, K.Ntzios, I.Pratikakis, S.Theodoridis, J.Perantonis, “Keyword-guided Word Spotting in historical printed documents using synthetic data and user feedback”. IJDAR (2007) 9:167-177.
- [2] R.Manmatha and N.Srimal, “Scale space technique for word segmentation in handwritten manuscripts”. In the Proc. of the Second International Conference on Scale-Space Theories in Computer Vision (Scale-Space’99), pp. 22-33, Sep. 1999.
- [3] T.M.Rath, S. Kane, A.Lehman, E.Partridge and R.Manmatha, “Indexing for a Digital Library of George Washington’s Manuscripts – A Study of Word Matching Techniques”, CIIR Technical Report MM-36, 2002.
- [4] T.M. Rath, and R. Manmatha, “Word Spotting for historical documents”. International Journal on Document Analysis and Recognition (IJ DAR), 9(2-4):139–152, 2007.
- [5] Toni M. Rath, and R. Manmatha, “Features for Word Spotting in historical manuscripts”, ICDAR, B03 conference, vol. 1, pp. 218-222.
- [6] Manmatha, R., Croft, W.B. Word Spotting: Indexing Handwritten Archives, In M. Maybury, editor, Intelligent Multimedia Information Retrieval Collection. AAAI/MIT Press, Menlo Park, CA, 1997.
- [7] G.L. Scot., and H.C Loguet-Higgins, “An algorithm for associating the features of two patterns”, Proc. Of the Royal Society of London B224 (1991), 21-26.
- [8] Pramod Sankar K., C.V. Jawahar, and R. Manmatha, “Nearest neighbor based collection OCR”, DAS’10, June 9-11, 2010, Boston, MA, USA.
- [9] Huaigu Cao, Anurag Bhardwaj, and Venu Govindaraju, “A probabilistic method for keyword retrieval in handwritten document images”, Pattern Recognition (2009).
- [10] B. Zhang, S.N. Srihari, and C. Huang, “Word image retrieval using binary features”, Document Recognition and Retrieval XI, vol. 5296, SPIE, Greenbelt, MD, 2004, pp. 45–53.
- [11] S.N. Srihari, and G.R. Ball, “Language independent Word Spotting in scanned documents” International Conference on Asian Digital Libraries, 2008, pp. 134–143.
- [12] Toni M. Rath, and R. Manmatha, “Word image Matching using dynamic time warping”, CVPR-03 conference, vol. 2, pp. 521-527.
- [13] Manmatha, R., Croft, W.B., “Word Spotting: Indexing Handwritten Archives”, In M. Maybury, editor, Intelligent Multimedia Information Retrieval Collection. AAAI/MIT Press, Menlo Park, CA, 1997.
- [14] Reza Farrahi Moghaddam and Mohamed Cheriet, “Application of multi-level classifiers and Clustering for automatic Word Spotting in historical document images”, 2009 10th International Conference on Document Analysis and Recognition, 2009.
- [15] Gatos, B., Pratikakis, I., Perantonis, S.J.:An adaptive binarisation technique for low quality historical documents In: IAPR Workshop on Document Analysis Systems (DAS2004), Lecture Notes in Computer Science (3163), September 2004, pp. 102–113 (2004).
- [16] Gatos, B., Pratikakis, I., Perantonis, S.J.: Adaptive degraded document image binarization. Pattern Recogn. **39**, 317–327 (2006).

- [17] Gatos, B., Danatsas, D., Pratikakis I., Perantonis, S.J.: Automatic table detection in document images. In: Proceedings of the Third International Conference on Advances in Pattern Recognition (ICAPR'05), Lecture Notes in Computer Science (3686), pp. 609–618, Path, UK (2005).
- [18] S. Theodoridis, and K. Koutoumbas, Pattern Recognition. Academic, New York (1997).
- [19] F.M. Wahl, K.Y. Wong, and R.G. Casey, “Block segmentation and text extraction in mixed text/image documents”, Comput. Graph. Image Process. 20, 375-390 (1982).
- [20] E. Kavallieratou, N. Fakotakis and G. Kokkinakis, “Skew angle estimation for printed and handwritten documents using the Wigner–Ville distribution”, Image and Vision Computing 20 (2002) 813–824, 2002.
- [21] Rajiv Kapoor, Deepak Bagai and T.S. Kamal, “A new algorithm for skew detection and correction”, Pattern Recognition Letters 25 (2004) 1215–1229, 2004.
- [22] Ibrahim S. I. Abuhaiba, “Skew Correction of Textural Documents”, J. King Saud Univ., Vol. 15, Comp. & Info. Sci., pp. 67-86 (A.H. 1423/2003), 2002.
- [23] Frank de Zeeuw, “Slant Correction using Histograms”, Bachelor’s Thesis in Artificial Intelligence, 2006.
- [24] Doulgeri Nikoleta, Kavalieratou Ergina, “Retrieval of Historical Documents by Word Spotting”, Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 7247, 724706, 2009 SPIE-IS&T – CCC code: 0277-786X/09/\$18 – doi: 10.1117/12.805602.
- [25] Sakoe, H., and Chiba, S., “Dynamic programming optimization for spoken word recognition”, IEEE Trans. on Acoustics, Speech and Signal Processing 26 (1980), 623-625, 1980.
- [26] Wikipedia, “Discrete Fourier Transform”, [Online] Available: http://en.wikipedia.org/wiki/Discrete_Fourier_transform, 13 May 2011 (Last Modified).
- [27] Wikipedia, “Discrete Cosine Transform”, [Online] Available: http://en.wikipedia.org/wiki/Discrete_cosine_transform, 27 May 2011 (Last Modified).
- [28] Maragkoudakis Manolis, “Data Mining – Chapter 7: Cluster Analysis - Lecture”, 2010-2011.
- [29] Wikipedia, “Single-Linkage Clustering”, [Online] Available: http://en.wikipedia.org/wiki/Single-Linkage_Clustering, 31 May 2011 (Last Modified).
- [30] Wikipedia, “Complete-Linkage Clustering”, [Online] Available: http://en.wikipedia.org/wiki/Complete-Linkage_Clustering, 28 February 2011 (Last Modified).
- [31] Mathworks, “Product Documentation - Linkage”, [Online] Available: <http://www.mathworks.com/help/toolbox/stats/Linkage.html>, 1984-2011 – The Mathworks Inc.
- [32] PowerShow, “Precision and Recall Presentation”, [Online] Available: http://www.powershow.com/view/b2691-Yzc2M/Precision_and_Recall_flash_ppt_presentation, 2011- CrystalGraphics Inc.
- [33] Wikipedia, “F1 score”, [Online] Available: http://en.wikipedia.org/wiki/F1_score, 25 May 2011 (Last Modified).
- [34] Delip Rao, “A note on cluster purity”, February 8, 2006.

[35] Manmatha R., and Rothfeder, J., “A scale space approach for automatically segmenting words from historical handwritten documents”, *IEEE Trans. On Pattern Analysis and Machine Intelligence*, 2005.

ΣΥΝΤΟΜΟ ΒΙΟΓΡΑΦΙΚΟ ΣΗΜΕΙΩΜΑ

ΠΟΛΥΧΡΟΝΗΣ ΜΑΡΙΟΣ

Γεννήθηκε στην πόλη της Λάρνακας της Κύπρου, στις 01-01-1987. Από το 2001 μέχρι το 2004, φοίτησε στο Παγκύπριο Λύκειο Λάρνακας στο κλάδο των Θετικών Επιστημών. Από το 2004 μέχρι το 2006, εκπλήρωσε την 25μήνη στρατιωτική του θητεία. Το 2006 εισήχθη στο Πανεπιστήμιο Αιγαίου, της Πολυτεχνικής Σχολής Θετικών Επιστημών, στο τμήμα Μηχανικών Πληροφοριακών και Επικοινωνιακών Συστημάτων από το οποίο και αποφοίτησε τον Ιούλιο του 2011. Το Σεπτέμβριο του 2011, θα συνεχίσει τις σπουδές του για μεταπτυχιακό τίτλο στο Πανεπιστήμιο Κύπρου, με θέμα «Επαγγελματικό Μάστερ σε Προηγμένες Τεχνολογίες Πληροφορικής», στο τμήμα Πληροφορικής της Σχολής Θετικών και Εφαρμοσμένων Επιστημών. Γνωρίζει άπταιστα την αγγλική γλώσσα. Είναι μέλος του οργανισμού IEEE (Institute of Electrical and Electronics Engineers), μέλος του οργανισμού IEEE Computer Society καθώς και του οργανισμού IEEE Communications Society.

ΕΥΑΓΟΡΟΥ ΑΝΤΡΕΑΣ

Γεννήθηκε στην πόλη της Λευκωσίας της Κύπρου, το 1986. Από το 1998 μέχρι το 2001 φοίτησε στο Γυμνάσιο Αγίου Δομετίου, και από το 2001 μέχρι το 2004, φοίτησε στο Ενιαίο Λύκειο Κύκκου Α της Λευκωσίας στον κλάδο των Θετικών Επιστημών. Από το 2004 μέχρι το 2006, εκπλήρωσε την 25μήνη στρατιωτική του θητεία. Το 2006 εισήχθη στο Πανεπιστήμιο Αιγαίου, της Πολυτεχνικής Σχολής Θετικών Επιστημών, στο τμήμα Μηχανικών Πληροφοριακών και Επικοινωνιακών Συστημάτων από το οποίο και αποφοιτά μέσα στο 2011. Γνωρίζει άπταιστα την αγγλική γλώσσα. Είναι μέλος του οργανισμού IEEE (Institute of Electrical and Electronics Engineers), μέλος του οργανισμού IEEE Computer Society καθώς και του οργανισμού IEEE Communications Society.