ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΣΗΜΑΤΩΝ ΕΛΕΓΧΟΥ ΚΑΙ ΡΟΜΠΟΤΙΚΗΣ

Εντοπισμός, διαχωρισμός, κατάτμηση: Διεργασίες επεξεργασίας χειρόγραφων και πολυμεσικών δεδομένων εν όψει εφαρμογών Αναγνώρισης, Αρχειοθέτησης και Δεικτοδότησης

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

Βασίλειος Παπαβασιλείου

ΜΑΪΟΣ 2010

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΣΗΜΑΤΩΝ ΕΛΕΓΧΟΥ ΚΑΙ ΡΟΜΠΟΤΙΚΗΣ

Εντοπισμός, διαχωρισμός, κατάτμηση: Διεργασίες επεξεργασίας χειρόγραφων και πολυμεσικών δεδομένων εν όψει εφαρμογών Αναγνώρισης, Αρχειοθέτησης και Δεικτοδότησης

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

Βασίλειος Παπαβασιλείου

Τριμελής Συμβουλευτική Επιτροπή

Γ. Καραγιάννης (Επιβλέπων)	Π. Μαραγκός	Σ. Κόλλιας

Επταμελής Επιτροπή

Γ. Καραγιάννης καθηγητής ΕΜΠ	Π. Μαραγκός καθηγητής ΕΜΠ	Σ. Κόλλιας καθηγητής ΕΜΠ
Α. Σταφυλοπάτης καθηγητής ΕΜΠ	Γ. Καμπουράκης καθηγητής ΕΜΠ	Β. Μέρτζιος καθηγητής ΔΠΘ

Β. Κατσούρος
Ερευνητής Β' ΙΕΛ

Πρόλογος

Η καθημερινή δημιουργία και ανάγνωση ηλεκτρονικών, έντυπων και χειρόγραφων κειμένων καθιστούν το γραπτό λόγο ως έναν από τους βασικούς φορείς πληροφορίας. Επομένως, είναι χρήσιμη η ανάπτυξη μεθόδων για τη διαχείριση της πληροφορίας αυτής, τη μετάδοσή της, την επεξεργασία της, τη διατήρησή της και τη δεικτοδότησή της. Οι προσπάθειες που γίνονται σε αυτή την κατεύθυνση έχουν οδηγήσει στην παραγωγή αξιόλογων και αποτελεσματικών προϊόντων λογισμικού για την επεξεργασία εικόνων έντυπων κειμένων. Οι ερευνητικές προσπάθειες επεκτείνονται και στην επεξεργασία εικόνων χειρόγραφων κειμένων. Η ποικιλομορφία των χειρόγραφων γραφημάτων και της διάταξης των χειρόγραφων εγγράφων είναι τα προβλήματα που καλείται να επιλύσει κάθε τέτοια προσπάθεια.

Η διατριβή αυτή επικεντρώνεται στην επεξεργασία ψηφιακών δυαδικών εικόνων χειρόγραφων κειμένων και ιδιαίτερα στον εντοπισμό των φυσικών στοιχείων του κειμένου, όπως οι γραμμές και οι λέξεις. Στην εργασία περιγράφονται δύο νέες τεχνικές για την οριοθέτηση των γραμμών κειμένου. Η πρώτη υιοθετεί τον αλγόριθμο Viterbi για την εύρεση της βέλτιστης ακολουθίας περιοχών κειμένου και κενών σε κατακόρυφες ζώνες του κειμένου και αποσκοπεί στη βελτίωση της γνωστής τεχνικής των επιμέρους προβολών. Η μέθοδος υποβλήθηκε προς αξιολόγηση σε δύο σχετικούς διεθνείς διαγωνισμούς (ICDAR2007 και ICDAR2009 Handwriting Segmentation Contests) και κατέλαβε την πρώτη και δεύτερη θέση αντίστοιχα. Η δεύτερη μέθοδος στοχεύει στη βελτίωση των αποτελεσμάτων των τεχνικών διάχυσης και εξέλιξης, προτείνοντας την ενσωμάτωση ενός σταδίου ελέγχου με την εφαρμογή μορφολογικών τελεστών δυαδικών εικόνων σε κάθε επανάληψη. Για την κατάτμηση του κειμένου σε λέξεις, προτείνεται μια νέα προσέγγιση που βασίζεται στην ποσοτικοποίηση των κενών μεταξύ διαδοχικών χαρακτήρων μέσω της αντικειμενικής συνάρτησης ενός γραμμικού ταξινομητή διανυσμάτων υποστήριξης χαλαρού περιθωρίου που τους διαχωρίζει. Η προτεινόμενη μέθοδος αξιολογήθηκε στα πλαίσια των προαναφερόμενων διαγωνισμών και παρουσίασε τα καλύτερα αποτελέσματα. Ως επέκταση της ανάλυσης εικόνων κειμένου, στην εργασία περιγράφεται μια τεχνική εντοπισμού πρόσθετου κειμένου σε πλαίσια βίντεο, η οποία ενσωματώνει ένα νέο στάδιο επαλήθευσης, στο οποίο οι εντοπισμένες περιοχές κατηγοριοποιούνται σε κειμενικές ή μη, με τη βοήθεια μιγμάτων γκαουσιανών κατανομών. Η προτεινόμενη τεχνική αποτελεί βαθμίδα του συστήματος δεικτοδότησης του ειδησεογραφικού τηλεοπτικού προγράμματος που λειτουργεί στο Εθνικό Συμβούλιο Ραδιοτηλεόρασης.

Στο πλαίσιο της διδακτορικής μου διατριβής συνεργάστηκα με ανθρώπους που με βοήθησαν να ανταποκριθώ στις προκλήσεις αυτής της έρευνας. Θέλω να ευχαριστήσω θερμά τον επιβλέποντά μου, καθηγητή Γιώργο Καραγιάννη για την αμέριστη συμπαράστασή του και την εμπιστοσύνη του σε κάθε εκπαιδευτική και επαγγελματική προσπάθειά μου. Ένας άνθρωπος με τον οποίο συνεργάστηκα στενά είναι ο κ. Βασίλης Κατσούρος, Ερευνητής Β' του Ινστιτούτου Επεξεργασίας του Λόγου (ΙΕΛ). Είναι δύσκολο με λίγες λέξεις να εκφράσω τις ευχαριστίες μου για την αδιάλειπτη βοήθειά του. Οι γνώσεις και η εμπειρία του συνέτειναν στο να μπορέσω να ξεπεράσω δυσκολίες που πολλές φορές φαίνονταν ως ανυπέρβλητα εμπόδια.

Επίσης, θα ήθελα να ευχαριστήσω τον Θέμο Σταφυλάκη για την καθημερινή συνεργασία μας και τη βοήθειά του, καθώς και τον κ. Γρηγόρη Σταϊνχάουερ και τον κ. Ιωάννη Δολόγλου, Ερευνητές Α΄ του ΙΕΛ. Θέλω ακόμα να ευχαριστήσω τον κ. Χριστόφορο Νίκου, Επίκουρο καθηγητή στο Τμήμα Πληροφορικής του Πανεπιστημίου Ιωαννίνων, για τα χρήσιμα σχόλια και τις υποδείξεις του. Στο σημείο αυτό θέλω να αφιερώσω την προσπάθεια μου στην Τερίνα και να την ευχαριστήσω για την φροντίδα της στο Θάνο και στο Ντίνο.

Περίληψη

Η ανάλυση εικόνων κειμένου έχει ως στόχο τη μετατροπή των έντυπων και χειρόγραφων κειμένων στα αντίστοιχα ηλεκτρονικά έγγραφα. Πρόκειται για μια σύνθετη διαδικασία που υλοποιείται σε επιμέρους στάδια επεξεργασίας, όπως η ψηφιοποίηση του πρωτοτύπου, ο εντοπισμός των περιοχών κειμένου, η κατάτμησή τους σε βασικά τμήματα του γραπτού λόγου (π.χ. γραμμές κειμένου, λέξεις και παραγράφους), η κατανόηση του ρόλου κάθε τμήματος, η αναγνώριση των χαρακτήρων και η δημιουργία του αντίστοιχου ηλεκτρονικού εγγράφου. Αν και έχουν αναπτυχθεί αποδοτικά εμπορικά προϊόντα για την επεξεργασία εντύπων, δεν έχει σημειωθεί η αντίστοιχη πρόοδος για τα χειρόγραφα. Η συγκεκριμένη εργασία επικεντρώνεται στην επεξεργασία ψηφιακών δυαδικών εικόνων χειρόγραφων κειμένων που περιέχουν μόνο κειμενικά στοιχεία και εστιάζει στα στάδια κατάτμησής τους σε γραμμές κειμένου και σε λέξεις.

Στην πρώτη ενότητα περιγράφονται δύο τεχνικές για την οριοθέτηση των γραμμών κειμένου. Η πρώτη τεχνική στοχεύει στη βελτίωση της υπάρχουσας μεθοδολογίας των επιμέρους προβολών, προτείνοντας τη μοντελοποίηση των κατακόρυφων ζωνών ανάλυσης ως ακολουθίες παρατηρήσεων που προκύπτουν από ένα κρυφό Μαρκοβιανό μοντέλο. Η προτεινόμενη τεχνική υποβλήθηκε προς αξιολόγηση σε δύο διεθνείς διαγωνισμούς κατάτμησης χειρόγραφου κειμένου σε γραμμές και παρουσίασε καλύτερα αποτελέσματα και από τις αντίστοιχες (προβολές) και από υλοποιήσεις άλλων μεθόδων. Η δεύτερη τεχνική βασίζεται στην εφαρμογή τελεστών δυαδικής μορφολογίας. Η διαφοροποίησή της έγκειται στην εισαγωγή ενός σταδίου ελέγχου μετά από κάθε επανάληψη για τον εντοπισμό προτύπων τα οποία δηλώνουν ότι τμήματα γειτονικών γραμμών τείνουν να ενωθούν ή έχουν ήδη ενωθεί. Η συγκριτική αξιολόγησή της με παρόμοιες τεχνικές, έδειξε ότι η ενσωμάτωση του σταδίου ελέγχου συμβάλει στη βελτίωση της επίδοσης.

Στη δεύτερη ενότητα εξετάζεται το πρόβλημα κατάτμησης του χειρόγραφου κειμένου σε λέξεις. Αν θεωρηθεί ότι τα εικονοστοιχεία δύο διαδοχικών γραφημάτων ανήκουν σε δύο τάξεις, τότε μπορεί να υπολογιστεί ο γραμμικός ταξινομητής διανυσμάτων υποστήριξης που τις διαχωρίζει. Για την εκτίμηση της απόστασης μεταξύ των γραφημάτων προτείνεται μια τιμή ανάλογη του περιθωρίου ταξινόμησης. Η κατηγοριοποίηση των αποστάσεων σε κενά μεταξύ λέξεων και σε κενά μεταξύ γραμμάτων της ίδιας λέξης, γίνεται με τη χρήση κατωφλίου που υπολογίζεται από τη συνάρτηση πυκνότητας πιθανότητας των αποστάσεων. Η αξιολόγηση της προτεινόμενης μεθόδου μέσω της συμμετοχής της σε δύο διεθνείς διαγωνισμούς, την ανέδειξε ως την αποτελεσματικότερη.

Ως επέκταση της ανάλυσης εικόνων που περιέχουν μόνο κειμενικά στοιχεία, στην τρίτη ενότητα περιγράφεται μια τεχνική εντοπισμού πρόσθετου κειμένου σε πλαίσια βίντεο, η οποία ενσωματώνει ένα στάδιο επαλήθευσης, στο οποίο οι εντοπισμένες περιοχές κατηγοριοποιούνται σε κειμενικές ή μη, με τη βοήθεια μιγμάτων γκαουσιανών κατανομών.

Abstract

The thesis focuses on handwritten document image analysis, so as to study and propose methods for two critical preprocessing stages in the workflow of an optical character recognition application, such as text-line and word segmentation. The shortcomings of the existing methods are discussed and two novel techniques for text-lines segmentation and one for locating words are introduced.

The first text-line segmentation algorithm is based on locating the optimal succession of text and gap areas within vertical zones by applying Viterbi algorithm on a Hidden Markov Model with parameters drawn from statistics of each type of area from the whole document image. Then, a text-line separator drawing technique is applied and finally the connected components are assigned to text lines according to simple geometrical constraints that conclude if a connected component can be directly assigned or it should be split because it lies across successive text lines. The algorithm participated in the ICDAR07 and ICDAR09 handwriting segmentation contests and took the first and second place respectively.

The second method is based on binary morphology. The basic steps of the approach are: a) texture reduction (by combining dilation and sub-sampling) to produce a low resolution image, in which the underlying texture of text lines is apparent while preventing aliasing and b) application of dilations and (p,q)-th generalized foreground rank openings successively to join close and horizontally overlapping regions while preventing a merge in the vertical direction. These operations evolve the candidate text lines and distinguish special patterns, which imply that text lines have come very close or have been merged. Finally, each connected component of the initial document image is assigned to the text line that intersects, whereas if it intersects more than one text lines, we cut it using the local ridges produced with the application of the watershed algorithm.

Word segmentation can be seen as a problem which requires the formulation of a metric of the gap between successive components and the clustering of the gaps in "inter" or "intra" word classes. To measure the gap metric, we use the negative logarithm of the objective function of a soft-margin linear Support Vector Machine. We employ a nonparametric approach to estimate the probability density function of the gap metrics and have observed that the “inter” words gaps are accumulated to the most right lobe of the probability density function while the “intra” word gaps are gathered to the left lobe. The classification threshold is chosen to be equal to the minimum between the two main lobes. The algorithm tested on the benchmarking datasets of ICDAR07 and ICDAR09 handwriting segmentation contests and outperformed the participating algorithms.

Furthermore, the thesis studies the problem of locating artificial text in video frames. A new method for verifying text areas detected in video streams is proposed. The algorithm explores the spectral properties of the horizontal projection of candidate text regions in order to reduce the high amount of false alarms that most text detection algorithms suffer from. The algorithm has been tested on newscast video sequences and we conclude that the addition of the

verification module increased the precision rate significantly while keeping the recall rate almost unaffected.

Περιεχόμενα

Εισαγωγή

Η ανάγνωση έντυπων ή χειρόγραφων κειμένων είναι ένας καθημερινός τρόπος λήψης της πληροφορίας. Όμως, η διαχείριση της καταγεγραμμένης σε χαρτί πληροφορίας δεν είναι εύκολη. Η διατήρηση και η αποθήκευσή της, η γρήγορη αναζήτηση συγκεκριμένου τμήματός της, η δεικτοδότησή της και φυσικά η διόρθωση-επεξεργασία της είναι μερικά από βασικά προβλήματα που προκύπτουν. Για τους λόγους αυτούς έχουν αναπτυχθεί καινοτόμες διαδικασίες για την ψηφιοποίηση των εγγράφων και την επεξεργασία τους, ώστε να προκύψουν τα αντίστοιχα ηλεκτρονικά έγγραφα. Ο θεματικός τομέας στον οποίο εντάσσονται αυτές οι ερευνητικές και τεχνολογικές προσπάθειες ονομάζεται ανάλυση εικόνων κειμένου (Document Image Analysis, DIA). Στον τομέα αυτό βρίσκουν εφαρμογή αλγόριθμοι και τεχνικές που προέρχονται από τις περιοχές της ψηφιακής επεξεργασίας εικόνας, της όρασης υπολογιστών και της μηχανικής μάθησης. Είναι ένας σύγχρονος επιστημονικός τομέας που προσελκύει το ενδιαφέρον όλο και περισσότερων ερευνητικών ομάδων και τεχνολογικών φορέων, όπως αποδεικνύει το αυξανόμενο πλήθος των εξειδικευμένων συνεδρίων που διοργανώνονται, των στοχευμένων επιστημονικών περιοδικών που εκδίδονται και των καινοτόμων εμπορικών εφαρμογών που αναπτύσσονται [1].

Τα πιο γνωστά συστήματα ανάλυσης εικόνων κειμένου είναι τα συστήματα οπτικής αναγνώρισης χαρακτήρων (Optical Character Recognition). Ο στόχος τους είναι η μετατροπή της ψηφιακής εικόνας κειμένου σε ένα μορφότυπο αφενός κατάλληλο για την προβολή και την επεξεργασία από τους διαθέσιμους ηλεκτρονικούς κειμενογράφους και αφετέρου όμοιο ή ακόμα καλύτερα ευκρινέστερο από το πρότυπο έγγραφο. Επομένως, θα πρέπει να περιλαμβάνει διαδικασίες προεπεξεργασίας όπως η απομάκρυνση των περιττών στοιχείων (π.χ. αστοχίες των μέσων ψηφιοποίησης), η κατάτμηση της εικόνας σε περιοχές με διαφορετικές αναπαραστάσεις της πληροφορίας (π.χ. περιοχές κείμενου, εικόνες, πίνακες κ.λπ.), η κατάτμηση των κειμενικών περιοχών σε γραμμές κειμένου, σε λέξεις και σε χαρακτήρες. Το επόμενο στάδιο επεξεργασίας αφορά στην αναγνώριση των χαρακτήρων και το τελικό στάδιο στην ανασύνθεση των κειμενικών και μη στοιχείων, ώστε να προκύψει το επιθυμητό ηλεκτρονικό έγγραφο. Η αναγνώριση χαρακτήρων και ιδιαίτερα των χειρόγραφων εξακολουθεί να αποτελεί μια από τις μεγαλύτερες προκλήσεις. Είναι φυσικά, άμεσα συνδεδεμένη με τη γλώσσα που χρησιμοποιείται στο κείμενο και προϋποθέτει την κατάλληλη εκπαίδευση της μηχανής αναγνώρισης. Εξαιρουμένης της αναγνώρισης, τα υπόλοιπα στάδια μπορούν να θεωρηθούν ανεξάρτητα γλώσσας και να τύχουν ενιαίας αντιμετώπισης.

Αυτές οι διαδικασίες αφορούν στην ανάλυση της διάταξης του εγγράφου (document layout analysis) και διακρίνονται σε εκείνες που α) εντοπίζουν τα φυσικά στοιχεία της εικόνας (document physical layout representation) και β) καθορίζουν το ρόλο των στοιχείων (document logical layout representation) [2]. Τα φυσικά στοιχεία της εικόνας είναι οι λέξεις, οι γραμμές κειμένου, οι παράγραφοι, οι πίνακες, οι οδηγοί (διαχωριστικά) και οι εικόνες. Οι πιθανοί ρόλοι τους σχετίζονται άμεσα με τον τύπο του εξεταζόμενου ψηφιοποιημένου εγγράφου. Ως

παράδειγμα αναφέρεται ότι στην περίπτωση που εξετάζεται η πρώτη σελίδα ενός επιστημονικού άρθρου, οι βασικοί ρόλοι είναι ο τίτλος του περιοδικού, ο τίτλος της εργασίας, τα ονόματα των συγγραφέων, η περίληψη και οι λέξεις-κλειδιά, ενώ αν πρόκειται για τη σελίδα μιας εφημερίδας τότε οι βασικοί ρόλοι είναι οι τίτλοι των θεμάτων και οι οδηγοί που τα διαχωρίζουν. Αν εντοπιστούν και ιεραρχηθούν οι περιοχές ενδιαφέροντος, τότε σε συνδυασμό με τα αποτελέσματα ενός συστήματος οπτικής αναγνώρισης χαρακτήρων, παρέχεται η δυνατότητα αυτόματης δημιουργίας μεταδεδομένων για τη δεικτοδότηση των αρχείων.

Η διάταξη των εγγράφων ποικίλει ανάλογα με το είδος τους. Πράγματι, ενώ η διάταξη των στοιχείων σε επιστημονικά άρθρα μπορεί να θεωρηθεί παρόμοια, διαφέρει σημαντικά από αυτή των εφημερίδων ή των διαφημιστικών φυλλαδίων. Αυτός είναι και ο λόγος που πολλές από τις προτεινόμενες τεχνικές εστιάζονται στην ανάλυση εγγράφων συγκεκριμένου είδους. Βέβαια, έχουν προταθεί και εύρωστες μέθοδοι των οποίων η αποτελεσματικότητα παραμένει υψηλή για την επεξεργασία έντυπων εγγράφων διαφορετικών ειδών. Τα αποτελέσματα των διαγωνισμών κατάτμησης εικόνων έντυπων εγγράφων (Page Segmentation Competition ICDAR2001-2009 [3]) που πραγματοποιούνται από το 2001 με τη συμμετοχή τεχνικών προτεινόμενων από ερευνητικές ομάδες και γνωστών εμπορικών εφαρμογών αναδεικνύουν ότι το συγκεκριμένο πρόβλημα παραμένει ανοιχτό.

Η διαφοροποίηση της διάταξης των εγγράφων γίνεται ιδιαίτερα έντονη στην περίπτωση των χειρόγραφων κειμένων, μια και η διάταξη αποτελεί επιλογή του εκάστοτε γραφέα. Είναι επομένως λογικό, στην ανάλυση εικόνων χειρόγραφων κειμένων να μην έχει υπάρξει η αντίστοιχη πρόοδος. Όπως αναφέρεται στην εκτενή μελέτη για την επεξεργασία χειρογράφων [4], ακόμα και σε χειρόγραφα κείμενα που περιέχουν μόνο κειμενικά στοιχεία, ο εντοπισμός των γραμμών του κειμένου και των λέξεων παραμένει ένα ανοιχτό ερευνητικό θέμα.

Η συγκεκριμένη εργασία επικεντρώνεται στην ανάλυση εικόνων χειρόγραφων κειμένων και ιδιαίτερα στην πρώτη βαθμίδα επεξεργασίας για την εξαγωγή των φυσικών στοιχείων της εικόνας. Στο πρώτο κεφάλαιο παρουσιάζονται οι κυριότερες τεχνικές και προτείνονται δύο νέες μέθοδοι για τον διαχωρισμό του χειρόγραφου κειμένου σε γραμμές κειμένου. Η πρώτη βασίζεται στη χρήση των επιμέρους προβολών και η δεύτερη στην εφαρμογή τελεστών της δυαδικής μαθηματικής μορφολογίας. Στο δεύτερο κεφάλαιο εξετάζεται το πρόβλημα διαχωρισμού του κειμένου σε λέξεις και προτείνεται ένας νέος τρόπος ποσοτικοποίησης των κενών μεταξύ των λέξεων με την υιοθέτηση των μηχανών διανυσμάτων υποστήριξης χαλαρών περιθωρίων. Ως προέκταση της ανάλυσης εικόνων κειμένου, στο τρίτο κεφάλαιο περιγράφεται το πρόβλημα εντοπισμού των πιθανών περιοχών κειμένου σε πλαίσια βίντεο και προτείνεται ένα πρόσθετο στάδιο επεξεργασίας για την επαλήθευση των περιοχών.

Η μεγάλη πλειοψηφία των χειρόγραφων κειμένων χαρακτηρίζονται εκ φύσεως από υψηλή ευκρίνεια αφού κατά τη δημιουργία τους, το μέσο γραφής και το χαρτί επιλέγονται με τέτοιο τρόπο ώστε να είναι έντονη η χρωματική αντίθεση των ιχνών του μέσου και του χαρτιού. Αυτός είναι και ο λόγος που κατά την ψηφιοποίηση των εγγράφων επιλέγεται σχεδόν αποκλειστικά η παραγωγή ψηφιακών δυαδικών εικόνων. Στις εικόνες αυτές, τα ίχνη του μέσου

γραφής αναπαρίστανται με εικονοστοιχεία (pixels) που έχουν τιμή 1 και συνιστούν το κείμενο. Τα υπόλοιπα εικονοστοιχεία έχουν την τιμή 0 και αντιστοιχούν στο φόντο. Η σύμβαση για την απόδοση των τιμών 1 και 0 στα εμφανιζόμενα δισδιάστατα αντικείμενα (σχήματα) και στο φόντο αντίστοιχα είναι η πλέον συνήθης στις μελέτες που αφορούν στις δυαδικές εικόνες.

Ένας από τους όρους που χρησιμοποιείται ευρέως στην ανάλυση ψηφιακών δυαδικών εικόνων είναι αυτός των συνεκτικών αντικειμένων (Connected Components, CCs). Συχνά αναφέρονται και ως αντικείμενα, σχήματα, ή υποσύνολα της εικόνας. Η συνεκτικότητα ορίζεται από τον τρόπο γειτνίασης των pixels και την τιμή που έχουν (για τις δυαδικές αν είναι 1 ή 0). Υποθέτοντας ότι η ψηφιακή εικόνα αναπαρίσταται σε ένα τετραγωνικό πλέγμα, τότε κάθε pixel, έστω *p*, έχει 8 γειτονικά pixels (2 κατά τον οριζόντιο άξονα, 2 κατά τον κατακόρυφο και 4 διαγώνια). Αν καθορίσουμε τον τρόπο γειτνίασης ως 8-n, τότε και τα 8 pixels είναι γείτονες του *p*. Αν επιλέξουμε γειτνίαση τύπου 4-n, τότε μόνο 4 pixels (2 κατά τον οριζόντιο και 2 κατά τον κατακόρυφο άξονα) θα είναι γείτονες του *p*. Αν τώρα οι γείτονες έχουν την ίδια τιμή με το *p*, τότε ορίζουν μια συνεκτική περιοχή της δυαδικής εικόνας. Γενικότερα, τα συνεκτικά αντικείμενα μιας δυαδικής εικόνας ορίζονται ως τα υποσύνολα με το μέγιστο πλήθος γειτονικών και ομότιμων pixels (maximal connected subsets) [κεφ. 7 στο 5]. Ο εντοπισμός και η δεικτοδότηση των συνεκτικών αντικειμένων επιτυγχάνεται με τον αλγόριθμο που περιγράφεται στο [6].

Κεφάλαιο 1. Κατάτμηση χειρόγραφου κειμένου σε γραμμές

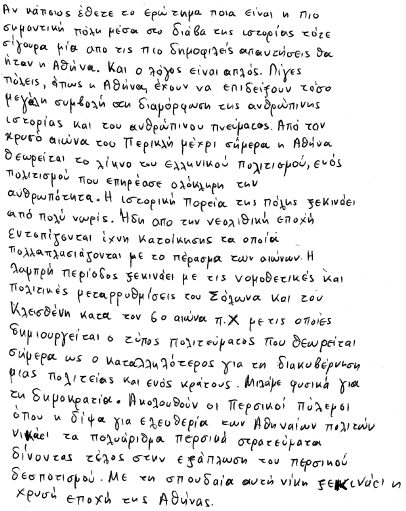
Ο διαχωρισμός ενός αυθόρμητα γραμμένου (χωρίς περιορισμούς) χειρόγραφου κειμένου σε γραμμές κειμένου αποτελεί ένα από τα σημαντικά στάδια ανάλυσης της εικόνας κειμένου. Πράγματι, αν οι γραμμές κειμένου εντοπιστούν, τότε είναι δυνατή είτε η ομαδοποίησή τους σε παραγράφους, είτε ο διαχωρισμός τους σε λέξεις και φυσικά η κατηγοριοποίησή τους ως επικεφαλίδες, υποσημειώσεις κ.λπ.

Τα κυριότερα προβλήματα που καλείται να επιλύσει κάθε προτεινόμενη τεχνική είναι:
α) η ακανόνιστη μορφοποίηση του κειμένου (non-Manhattan layout),
β) η μεταβλητή κλίση των γραμμών κειμένου (skew angle variation),
γ) η ύπαρξη ενωμένων γραφημάτων που ανήκουν σε διαφορετικές γραμμές (touching lines),
δ) η μεταβλητότητα του μεγέθους των χαρακτήρων (character size variation) και ιδιαίτερα αυτών που εκ φύσεως εκτείνονται προς τα πάνω (ascenders) όπως δ, ζ, λ, ξ, f, h, κ.ά. ή προς τα κάτω (descenders) όπως γ, ρ, g, y, κ.ά. προκαλώντας την οριζόντια επικάλυψη τμημάτων των διαδοχικών γραμμών (overlapping components) και
ε) η ποικιλομορφία των γραφικών χαρακτήρων (writing styles) και των γλωσσών (scripts).
Μερικά παραδείγματα χειρόγραφων κειμένων παρουσιάζονται στο σχ. 1.1.

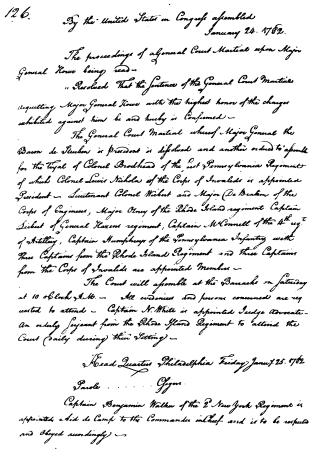
Οι αρχικές προσεγγίσεις του προβλήματος υιοθέτησαν γνωστές μεθόδους που ήδη είχαν εφαρμοστεί στην επεξεργασία εικόνων έντυπων κειμένων για την εκτίμηση της κλίσης του κειμένου [8]. Η βασική υπόθεση ήταν ότι η κλίση του κειμένου (document's skew) είναι σταθερή και οφείλεται κυρίως σε πιθανές αστοχίες είτε κατά την παραγωγή (π.χ. εκτύπωση ή φωτοτύπηση) της σελίδας του κειμένου είτε κατά την ψηφιοποίησή της. Τα βασικά «εργαλεία» για την εκτίμηση της κλίσης είναι οι προβολές, ο μετασχηματισμός Hough και η εφαρμογή δυαδικών μορφολογικών τελεστών [9-11]. Στην ακόλουθη ενότητα παρουσιάζεται ο τρόπος εφαρμογής τους στις δυαδικές εικόνες χειρόγραφων κειμένων.

Η διαφοροποίηση των χειρόγραφων από τα έντυπα κείμενα αποτυπώνεται στο σχήμα 1.2. Είναι αναμενόμενο η πολύ αποτελεσματική μέθοδος Docstrum [12] για την επεξεργασία εντύπων, να αντιμετωπίζει πολλές δυσκολίες στον εντοπισμό των γραμμών κειμένου στο χειρόγραφο. Αξίζει να αναφερθεί ότι η μέθοδος Docstrum έχει εφαρμοστεί σε εκατοντάδες σελίδες επιστημονικών περιοδικών διαφορετικών εκδοτικών οίκων και έχει παρουσιάσει εξαιρετικά αποτελέσματα. Πρόκειται για μια κλασσική bottom-up μέθοδο αφού εξετάζει ένα προς ένα τα συνεκτικά αντικείμενα της εικόνας με στόχο την ομαδοποίησή τους σε μεγαλύτερες οντότητες του εγγράφου (π.χ. γραμμές κειμένου, παραγράφους, κ.λπ.). Στο πρώτο στάδιο της μεθόδου, για κάθε CC της εικόνας εντοπίζονται τα πέντε κοντινότερά του CCs με βάση την ευκλείδεια απόσταση των κέντρων μάζας τους. Επίσης, υπολογίζονται οι γωνίες που σχηματίζουν τα ευθύγραμμα τμήματα, τα οποία συνδέουν τα κέντρα μάζας τους. Από το ιστόγραμμα των γωνιών επιλέγεται η μέγιστη τιμή ως η κλίση του εγγράφου και η αντίστοιχη

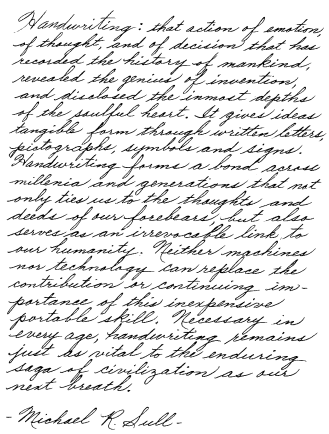
απόσταση ως η απόσταση μεταξύ των CCs που ανήκουν στην ίδια γραμμής κειμένου. Τέλος, τα CCs ομαδοποιούνται συγκρίνοντας τις αποστάσεις τους με την υπολογισμένη απόσταση. Η μέθοδος Docstrum επεκτείνεται και στην ομαδοποίηση των γραμμών κειμένου σε παραγράφους ή στήλες, αλλά αυτό που χρήζει ιδιαίτερης προσοχής είναι ότι ήδη έχουν γίνει δύο βασικές υποθέσεις που δεν μπορούν να υιοθετηθούν στην επεξεργασία χειρόγραφων κειμένων. Η πρώτη είναι πως κάθε CC εκτείνεται σε μια και μοναδική γραμμή κειμένου και η δεύτερη ότι η απόσταση μεταξύ διαδοχικών γραμμών κειμένου είναι μεγαλύτερη από αυτή μεταξύ των CCs της ίδιας γραμμής.



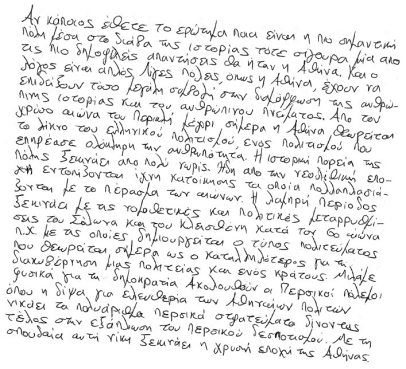
024.tif ICDAR07



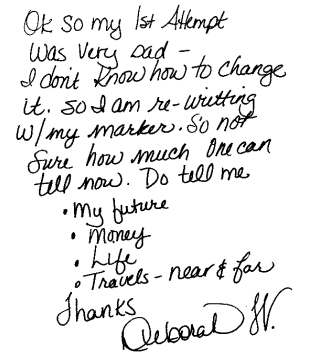
016.tif ICDAR07



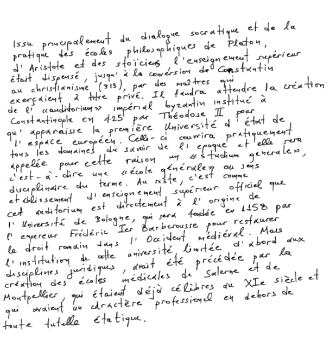
044.tif ICDAR07



010.tif ICDAR07

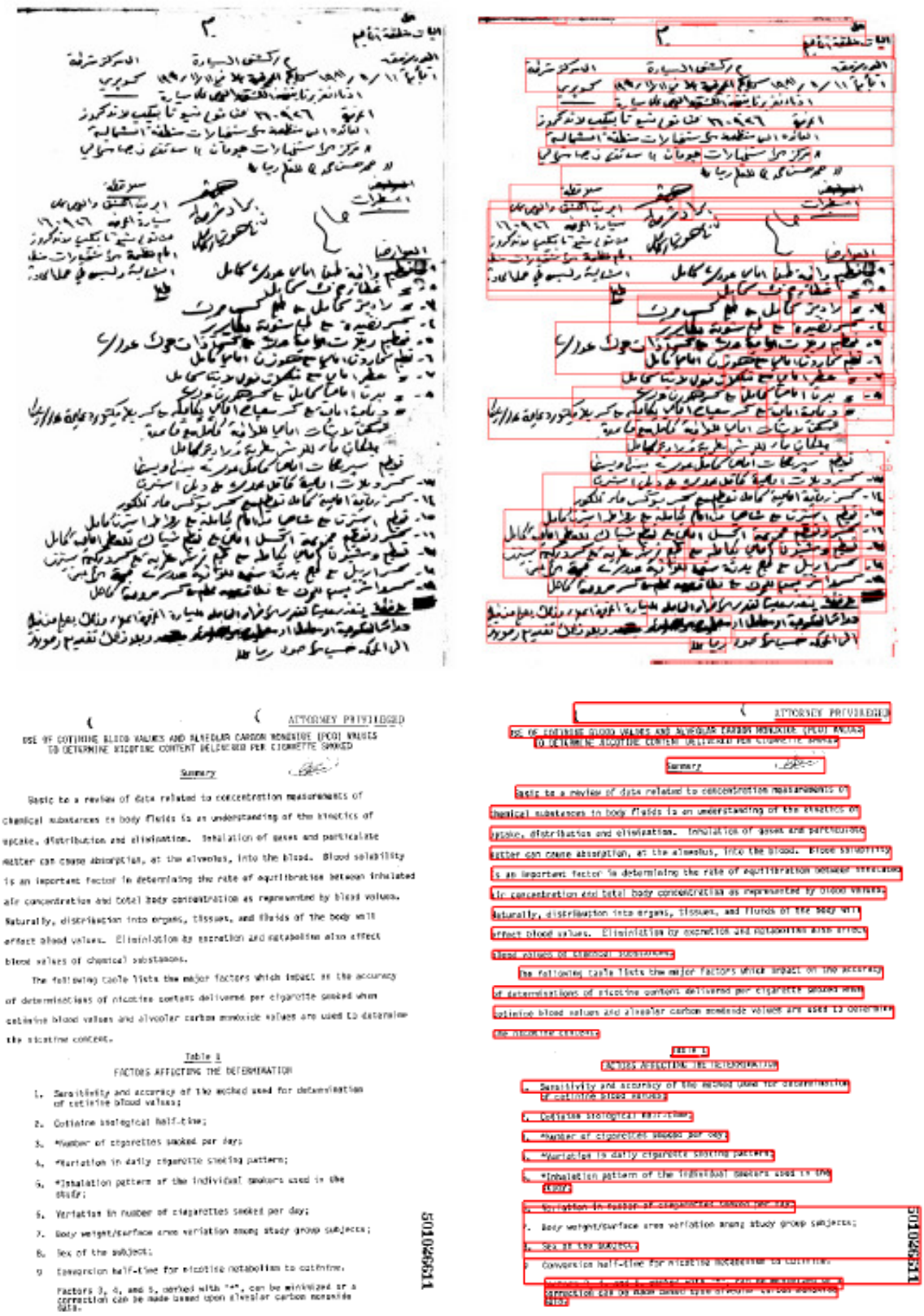


068.tif ICDAR07



026.tif ICDAR07

Σχήμα 1.1 Παραδείγματα χειρόγραφων κειμένων από τις συλλογές που χρησιμοποιήθηκαν στο διαγωνισμό ICDAR2007 Handwriting Segmentation [7]



ATTORNEY PRIVILEGED

USE OF COTININE BLOOD VALUES AND ALVEOLAR CARBON MONOXIDE (PCO) VALUES TO DETERMINE NICOTINE CONTENT DELIVERED PER CIGARETTE SMOKED

Summary

Basic to a review of data related to concentration measurements of chemical substances in body fluids is an understanding of the kinetics of uptake, distribution and elimination. Inhalation of gases and particulate matter can cause absorption, at the alveolus, into the blood. Blood solubility is an important factor in determining the rate of equilibration between inhaled air concentration and total body concentration as represented by blood values. Naturally, distribution into organs, tissues, and fluids of the body will affect blood values. Elimination by excretion and metabolism also affect blood values of chemical substances.

The following table lists the major factors which impact on the accuracy of determinations of nicotine content delivered per cigarette smoked when cotinine blood values and alveolar carbon monoxide values are used to determine the nicotine content.

Table 1

FACTORS AFFECTING THE DETERMINATION

1. Sensitivity and accuracy of the method used for determination of cotinine blood values;
2. Cotinine biological half-time;
3. *Number of cigarettes smoked per day;
4. *Variation in daily cigarette smoking pattern;
5. *Inhalation pattern of the individual smokers used in the study;
6. Variation in number of cigarettes smoked per day;
7. Body weight/surface area variation among study group subjects;
8. Sex of the subjects;
9. Conversion half-time for nicotine metabolism to cotinine.

Factors 3, 4, and 5, marked with "*", can be minimized or a correction can be made based upon alveolar carbon monoxide data.

5010926514

Σχήμα 1.2 Χρήση του αλγόριθμου Docstrum για την κατάτμηση σε γραμμές ενός χειρόγραφου και ενός έντυπου κειμένου [26].

1.1. Προσεγγίσεις

Οι τεχνικές που έχουν εφαρμοστεί στο διαχωρισμό κειμένου σε γραμμές κατηγοριοποιούνται σε τέσσερις μεγάλες κατηγορίες με βάση τη μεθοδολογία που ακολουθούν [13, 14]. Συγκεκριμένα, διακρίνονται σε τεχνικές που αναλύουν τις προβολές (projections ή projection profiles), που εφαρμόζουν το μετασχηματισμό Hough, που ομαδοποιούν τα CCs με βάση τις σχετικές θέσεις τους (CCs analysis) και σε αυτές που υιοθετούν τεχνικές διάχυσης και εξέλιξης. Στη συνέχεια του κεφαλαίου παρουσιάζονται συνοπτικά τεχνικές που είτε έχουν κριθεί ως πολύ αποτελεσματικές στα αντίστοιχα σύνολα εξέτασης, είτε αποτελούν τμήματα ολοκληρωμένων συστημάτων για την επεξεργασία ειδικών χειρόγραφων κειμένων (π.χ. ταχυδρομικών διευθύνσεων, εκκλησιαστικών κειμένων κ.ά.).

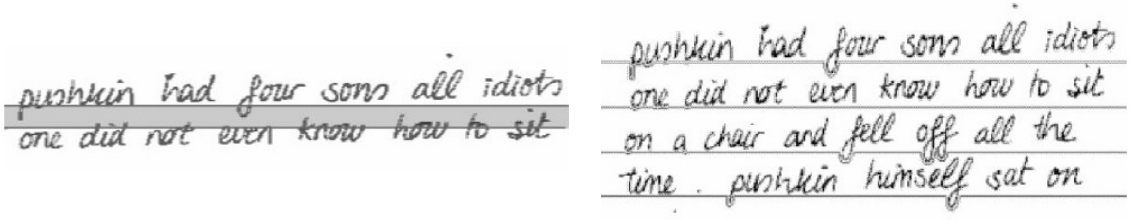
Στην πρώτη κατηγορία των σχετικών αλγορίθμων ανήκουν αυτοί που βασίζονται στις προβολές. Η προβολή μιας εικόνας A διαστάσεων $M \times N$, ορίζεται με την ακόλουθη σχέση, ως ένα σήμα διάρκειας M (κατακόρυφη διάσταση εικόνας) με κάθε τιμή του να ισούται με το άθροισμα των τιμών των pixels που έχουν την ίδια τετμημένη:

$$PR(i) = \sum_{j=1}^{N} A(i, j), \quad i = 1, ..., M \qquad \text{(Εξ. 1.1)}$$

Είναι προφανές ότι οι προβολές παρουσιάζουν σημαντικούς λοβούς στις περιοχές της εικόνας που υπάρχει κείμενο και εξίσου σημαντικές κοιλάδες μεταξύ των γραμμών κειμένου (πιθανές θέσεις διαχωριστικών). Όμως, αν το κείμενο παρουσιάζει κλίση, τότε η άμεση χρήση των προβολών δεν μπορεί να οδηγήσει σε ασφαλή συμπεράσματα για τα όρια των γραμμών κειμένου.

Οι Yanikoglu και Sandon [15] υποθέτουν ότι το κείμενο έχει σταθερή κλίση και αρχικός στόχος τους είναι ο υπολογισμός της και φυσικά η διόρθωσή της. Για το λόγο αυτό, η προς επεξεργασία εικόνα περιστρέφεται διαδοχικά από $-\varphi^o$ ως φ^o (με βήμα 0.5^o ή 1^o ανάλογα την επιθυμητή ακρίβεια) και οι προβολές κατά τον οριζόντιο άξονα υπολογίζονται κάθε φορά. Η προβολή που παρουσιάζει τη μέγιστη τιμή με βάση ένα επιλεγμένο κριτήριο (π.χ. το άθροισμα των τετραγώνων των διαφορών δύο διαδοχικών τιμών της προβολής) αντιστοιχεί στη γωνία κλίσης. Η περιστροφή της εικόνας κατά τη γωνία αυτή έχει ως αποτέλεσμα την εξάλειψη της κλίσης και επομένως τώρα οι λοβοί της προβολής θα αναδείξουν τις περιοχές κειμένου. Αντίστοιχα, τα τοπικά ελάχιστα (μέγιστα) της πρώτης παράγωγου της προβολής θα οριοθετούν την έναρξη (τέλος) περιοχής του κενού μεταξύ δύο διαδοχικών περιοχών κειμένου. Η διαφοροποίησή τους από τις τεχνικές που χρησιμοποιούνται για τα έντυπα κείμενα, έγκειται στον τρόπο χάραξης των διαχωριστικών μεταξύ των γραμμών κειμένου. Οι προβολές δε χρησιμοποιούνται για την άμεση χάραξη των διαχωριστικών (π.χ. στα αντίστοιχα τοπικά ελάχιστα), αλλά για να οριοθετήσουν μια περιοχή μεταξύ δύο γραμμών κειμένου στην οποία θα πρέπει να χαραχθεί το διαχωριστικό (βλ. Εν. 1.2). Ξεκινώντας από το αριστερό όριο της εικόνας και μέσα στην οριοθετημένη περιοχή του κενού (σχ. 1.3α), έστω από το pixel με τεταγμένη *y*, το

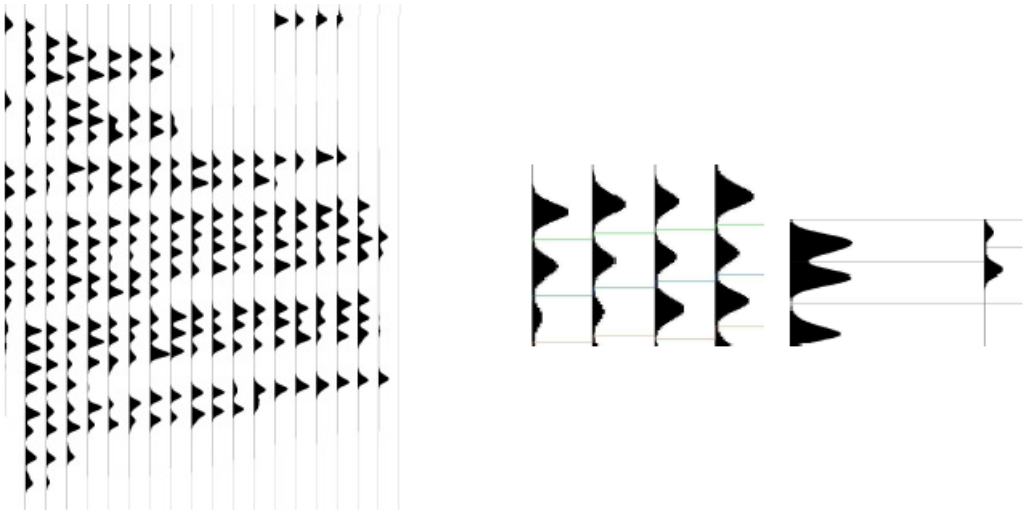
διαχωριστικό επεκτείνεται προς τα δεξιά μέχρι να συναντήσει κάποιο CC. Τότε ακολουθεί την περιβάλλουσα του CC (σχ. 1.3β), μέχρι να βρεθεί ξανά στην τεταγμένη y, κ.ο.κ. Αν ένα CC εκτείνεται σε όλο το μέτωπο της οριοθετημένης περιοχής, τότε το διαχωριστικό ξεπερνά το CC, τέμνοντάς το στη μεσαία γραμμή (row) της οριοθετημένης περιοχής. Μια βασική αδυναμία της συγκεκριμένης μεθόδου είναι προφανώς ότι λόγω της χρήσης των ολικών προβολών, δεν αντιμετωπίζει επιτυχώς τις περιπτώσεις κειμένων στα οποία η κλίση ποικίλει από γραμμή σε γραμμή.



(α) (β)

Σχήμα 1.3. Κατάτμηση χειρόγραφου κειμένου σε γραμμές [15]. (α) Εντοπισμός κρίσιμης περιοχής. (β) Χάραξη διαχωριστικών.

Αυτή η αδυναμία των ολικών προβολών είχε ως αποτέλεσμα την υιοθέτηση των επιμέρους προβολών (piece-wise projection profiles) [16]. Η βασική ιδέα είναι να χωριστεί η εικόνα σε κατακόρυφες ζώνες όπου η κλίση κάθε γραμμής του κειμένου μπορεί να θεωρηθεί αμελητέα. Επομένως, τα τοπικά ελάχιστα κάθε επιμέρους προβολής δηλώνουν τις θέσεις των αντίστοιχων διαχωριστικών ανά ζώνη (σχ. 1.4α). Στη συνέχεια, κάθε διαχωριστικό συνδυάζεται με το κοντινότερο διαχωριστικό της επόμενης ζώνης για να προκύψουν τα όρια των γραμμών κειμένου (σχ. 1.4β). Κάθε CC που περιέχεται εξ ολοκλήρου στην περιοχή μεταξύ δύο διαχωριστικών, ανατίθεται στην αντίστοιχη γραμμή κειμένου. Για την ανάθεση των CCs που εκτείνονται σε δύο ή περισσότερες περιοχές εξετάστηκαν δύο τεχνικές. Η πρώτη αφορά στη μοντελοποίηση κάθε γραμμής κειμένου με δισδιάσταση κανονική κατανομή, της οποίας οι παράμετροι $\boldsymbol{\mu}$ και $\boldsymbol{\Sigma}$ υπολογίζονται από τις τετμημένες και τεταγμένες των pixels των CCs που της έχουν ήδη ανατεθεί. Η απόφαση για το εξεταζόμενο CC προκύπτει από τη σύγκριση των πιθανοτήτων να περιγράφεται από τις κατανομές των γραμμών στις οποίες εκτείνεται. Η δεύτερη τεχνική υιοθετεί απλούς γεωμετρικούς κανόνες και αναθέτει το εξεταζόμενο CC στη γραμμή με τα pixels της οποίας παρουσιάζει τη μεγαλύτερη επικάλυψη. Η προσέγγιση αυτή εφαρμόστηκε σε 720 χειρόγραφα κείμενα που περιέχουν 11581 γραμμές κειμένου και το 97.31% από αυτές εντοπίστηκαν σωστά [17-18]. Από την ανάλυση των αποτελεσμάτων συμπεραίνει κανείς πως ο συνδυασμός των διαχωριστικών από ζώνη σε ζώνη δεν είναι πάντα προφανής. Πράγματι, όταν το κείμενο είναι αραιογραμμένο (π.χ. όταν το κενό μεταξύ δύο διαδοχικών λέξεων είναι μεγαλύτερο από το πλάτος της ζώνης, δεν θα προκύψει το αντίστοιχο διαχωριστικό) ή το μέγεθος των χαρακτήρων μια γραμμής κειμένου ποικίλει σημαντικά, δημιουργούνται αμφισημίες (βλ. Εν. 1.2).

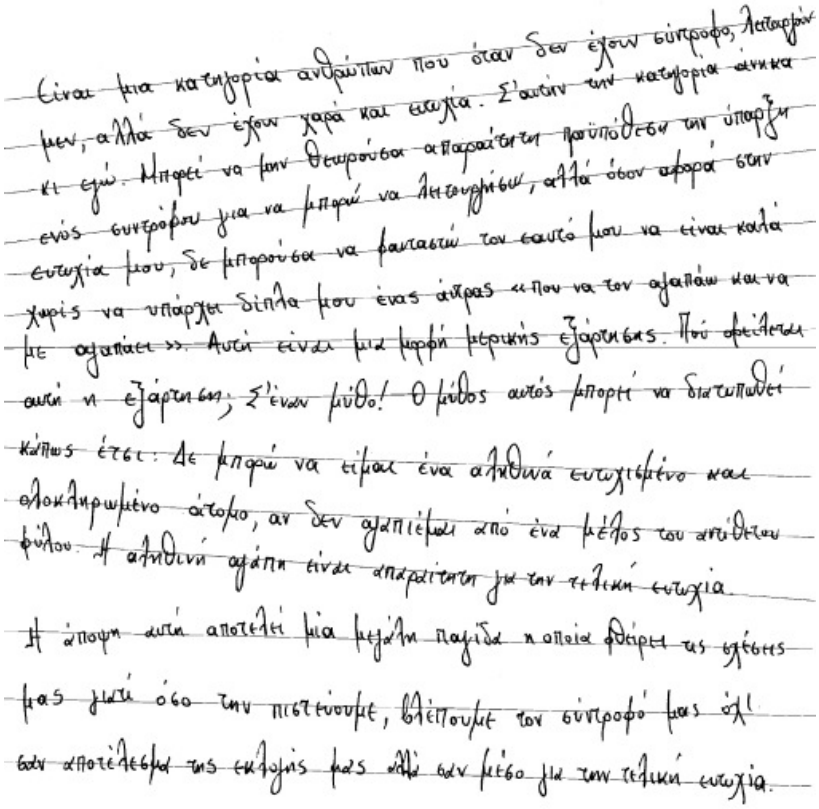


(α) (β)

Σχήμα 1.4. Κατάτμηση με τη χρήση επιμέρους προβολών [16]. (α) Οι επιμέρους προβολές των κατακόρυφων ζωνών. (β) Χάραξη των διαχωριστικών

Η δεύτερη κατηγορία περιλαμβάνει τις τεχνικές που βασίζονται στο μετασχηματισμό Hough [19] για την εύρεση συνευθειακών σημείων της εικόνας. Η υιοθέτησή του για την κατάτμηση κειμένου σε γραμμές, προϋποθέτει την επιλογή των κατάλληλων σημείων που αντιπροσωπεύουν τα CCs (π.χ. κέντρα μάζας των CCs) [20]. Σε μια τέτοια τεχνική [21], αρχικά υπολογίζονται το πλάτος και το ύψος για κάθε CC και με την εφαρμογή απλών κανόνων επιλέγονται αυτά με μεγάλο σχετικά πλάτος και μέτριο ύψος. Όπως αναφέρεται, τα αντικείμενα αυτά, αφενός δεν εκτείνονται σε περισσότερες από μία γραμμές (μέτριο ύψος) και αφετέρου περιέχουν κρίσιμη πληροφορία για την κλίση της γραμμής (μεγάλο πλάτος). Τα κέντρα μάζας των επιλεγμένων CCs τροφοδοτούν το μετασχηματισμό. Κάθε ευθεία που διέρχεται από ένα σημείο (x, y) αναπαρίσταται με τη σχέση $\rho = x\cos\theta + y\sin\theta$ όπου $\rho \in [-\sqrt{2}D, \sqrt{2}D]$ η απόσταση της ευθείας από το σημείο αναφοράς (D το μήκος της κύριας διαγωνίου της εικόνας κειμένου) και $\theta \in [-90^o, 90^o]$ η αντίστοιχη γωνία. Για τις παραμέτρους (ρ, θ) επιλέγονται οι επιθυμητές διακριτές τιμές τους, με βάση την επιθυμητή ακρίβεια. Έστω μ και ν τα πλήθη των επιλεγμένων διακριτών τιμών των παραμέτρων θ και ρ αντίστοιχα και $\mathbf{H}_{\mu\times\nu}$ ο πίνακας «συνάθροισης». Για κάθε προς επεξεργασία σημείο της εικόνας και για κάθε επιλεγμένη τιμή $\theta_i, i = 1,...,\mu$ υπολογίζονται οι αντίστοιχες τιμές της παραμέτρου $\rho_i, i = 1,...,\mu$ και κάθε μια αντιστοιχίζεται στην κοντινότερη επιλεγμένη τιμή $\rho_j, j = 1,...,\nu$. Με τον τρόπο αυτό, ορίζονται κάποιες ευθείες, οι οποίες διέρχονται από το εξεταζόμενο σημείο. Ακολούθως, οι τιμές των κελιών *(i , j)* του πίνακα «συνάθροισης», αυξάνονται κατά ένα. Τα σημεία που αντιστοιχούν σε κάθε κελί του πίνακα $\mathbf{H}$, είναι σημεία της εικόνας που «ανήκουν» στην ίδια ευθεία. Από την επαναληπτική εξέταση του πίνακα συνάθροισης, προκύπτει σε κάθε βήμα η ομάδα των σημείων που βρίσκονται κοντά στην ίδια ευθεία και συνεπώς τα CCs που ανήκουν στην αντίστοιχη

γραμμή κειμένου. Είναι προφανές, ότι η βασική υπόθεση που γίνεται, είναι πως η κλίση της γραμμής κειμένου παραμένει σταθερή. Αυτό έχει ως αποτέλεσμα την αδυναμία εντοπισμού των γραμμών κειμένου με μεταβλητή κατά μήκος τους κλίση, όπως άλλωστε αναφέρεται και στην ανάλυση των αποτελεσμάτων (σχ. 1.5).



Σχήμα 1.5. Η υιοθέτηση του μετασχηματισμού Hough για την κατάτμηση του χειρόγραφου κειμένου σε γραμμές είναι αποτελεσματική όταν η κλίση κάθε γραμμής είναι σταθερή [21]

Η τρίτη κατηγορία περιλαμβάνει μεθόδους που μελετούν τις σχετικές θέσεις και τις διαστάσεις των CCs του κειμένου. Οι συγκεκριμένες τεχνικές είναι ιδιαίτερα αποτελεσματικές στις περιπτώσεις που οι αποστάσεις μεταξύ διαδοχικών γραμμών κειμένου είναι μεγαλύτερες από αυτές μεταξύ των CCs εντός της ίδιας γραμμής. Όμως, αποτυγχάνουν στις περιπτώσεις που κάποια CCs εκτείνονται σε περισσότερες από μία γραμμές κειμένου (overlapping CCs). Για το λόγο αυτό, δε χρησιμοποιούνται στην ανάλυση αυθόρμητα γραμμένων χειρόγραφων κειμένων, αλλά αποτελούν τμήματα συστημάτων που έχουν σχεδιαστεί για ειδικές εφαρμογές. Στο [22] περιγράφεται μια τεχνική για την επεξεργασία εκκλησιαστικών κειμένων του 17^{ου} αιώνα, που βασίζεται στις γεωμετρικές αποστάσεις μεταξύ των CCs για τον υπολογισμό των βασικών γραμμών των περιοχών κειμένου. Η μέθοδος απαιτεί τον καθορισμό παραμέτρων που σχετίζονται με τον γραφικό χαρακτήρα του εκάστοτε γραφέα και παρουσιάζει 97% ποσοστό ορθού εντοπισμού γραμμών. Αν όμως, οι παράμετροι γενικευτούν για όλους τους γραφείς, η ακρίβεια περιορίζεται στο 90%. Μια άλλη bottom-up προσέγγιση [23] αποτελεί τη βαθμίδα για το διαχωρισμό του χειρόγραφου κειμένου σε γραμμές που ενσωματώνει το σύστημα αυτόματης

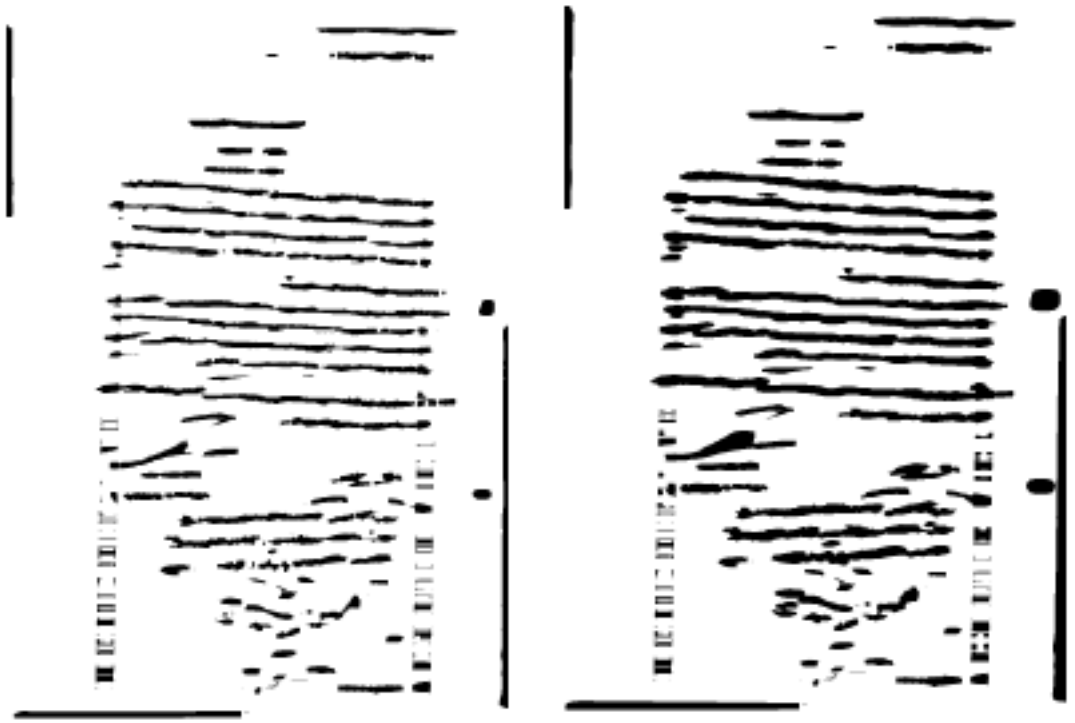
διαλογής των ταχυδρομικών φακέλων (Handwritten Address Interpretation System). Η συγκεκριμένη τεχνική αναθέτει τα CCs της εικόνας της χειρόγραφης ταχυδρομικής διεύθυνσης στην κατάλληλη γραμμή κειμένου με βάση την οριζόντια επικάλυψή τους.

Η τέταρτη κατηγορία περιλαμβάνει τις τεχνικές διάχυσης (smearing) και εξέλιξης (evolution). Η βασική ιδέα είναι ο αρχικός εντοπισμός των κρίσιμων περιοχών (π.χ. τμήματα των γραμμών κειμένου) και η εξέλιξή τους υπό όρους ώστε να οριοθετηθούν οι επιθυμητές περιοχές. Για την ανάδειξη των κρίσιμων περιοχών προτείνεται στο [24], η δημιουργία μιας νέας εικόνας στην οποία η τιμή του pixel αντιστοιχεί στο πλήθος των pixels του φόντου που μπορεί να συναντήσει κανείς, ξεκινώντας από το pixel αυτό και μετακινούμενος προς τα δεξιά ή αριστερά του (horizontal background run-length). Πειραματικά προσδιορίστηκε ότι αν κατά τη μετακίνηση αγνοηθεί η συνάντηση με συγκεκριμένο πλήθος pixels κειμένου (π.χ. 10), οι τιμές των pixels στη νέα εικόνα (horizontal background fuzzy runlength), αναδεικνύουν με μεγαλύτερη επιτυχία τις περιοχές κειμένου (σχ. 1.6). Προφανώς, στη νέα εικόνα οι μικρές τιμές αντιστοιχούν σε περιοχές κειμένου και οι μεγαλύτερες στο φόντο. Με τη δυαδικοποίηση της εικόνας μέσω της μεθόδου ολικής κατωφλίωσης του Otsu [25], αναδεικνύονται οι κρίσιμες περιοχές κειμένου ως συνεκτικά αντικείμενα της δυαδικής εικόνας. Στη συνέχεια τα αντικείμενα ομαδοποιούνται με βάση την απόστασή τους και δημιουργούνται οι γραμμές κειμένου.

Σχήμα 1.6. Αρχική εκτίμηση των γραμμών κειμένου [24].

Μια άλλη πρόταση [26] για τον αρχικό εντοπισμό των γραμμών κειμένου είναι η χρήση της μεθόδου παραθύρωσης Parzen για τον υπολογισμό της αντίστοιχης «πιθανοτικής εικόνας» Το βέλτιστο μέγεθος του παραθύρου υπολογίστηκε πειραματικά ως 30×120. Στη συνέχεια, εφαρμόζεται η μέθοδος τοπικής κατωφλίωσης του Niblack [27] για τη δυαδικοποίηση της νέας εικόνας και τη δημιουργία συνεκτικών αντικειμένων που αντιστοιχούν σε τμήματα των γραμμών κειμένου (σχ. 1.7α). Η μέθοδος Niblack υπολογίζει ένα κατώφλι T για κάθε pixel της εικόνας, μέσω της σχέσης $\mathrm{T}=\mu-0.2\sigma$, όπου μ και σ η μέση τιμή και η τυπική απόκλιση των τιμών των pixels στο παράθυρο ανάλυσης με κέντρο το εκάστοτε σημείο. Οι περιβάλλουσες των CCs εξελίσσονται επαναληπτικά με τη μέθοδο των επιπεδοσυνόλων (level sets) [28]. Η ταχύτητα

εξέλιξης των περιβαλλουσών ελέγχεται από τις τιμές της «πιθανοτικής εικόνας». Επομένως, η περιβάλλουσα εξελίσσεται γοργά όταν βρίσκεται σε περιοχές κειμένου. Όπως αναφέρεται, μετά από 10 επαναλήψεις οι περιβάλλουσες έχουν καλύψει μεγάλα τμήματα των γραμμών κειμένου και η απλή εξέταση των αποστάσεων των τμημάτων αυτών, έχει ως αποτέλεσμα την ομαδοποίησή τους σε γραμμές κειμένου (σχ. 1.7β). Μετά από κάθε επανάληψη, ελέγχεται η επικάλυψη των ορθογωνίων που περικλείουν κάθε εξελισσόμενη περιοχή (zero level set). Αν η επικάλυψη είναι σημαντική (μεγαλύτερη από ένα επιλεγμένο κατώφλι) το βήμα εξέλιξης «ακυρώνεται», η περιοχή επικάλυψης της πιθανοτικής εικόνας μηδενίζεται (ώστε να αποτραπεί η εξέλιξη μέσα σε αυτή) και συνεχίζεται η διαδικασία εξέλιξης. Η μέθοδος εφαρμόστηκε σε 400 χειρόγραφα κείμενα Αραβικών, Κορεατικών, Ινδικών και Κινεζικών και παρουσίασε ποσοστά ορθού εντοπισμού 85.6%, 92%, 95% και 96% αντίστοιχα. Στην ανάλυση των αποτελεσμάτων, σημειώνεται από τους συγγραφείς, ότι η μέθοδος αδυνατεί να διακρίνει διαδοχικές γραμμές κειμένου όταν αρκετοί χαρακτήρες τους επικαλύπτονται.

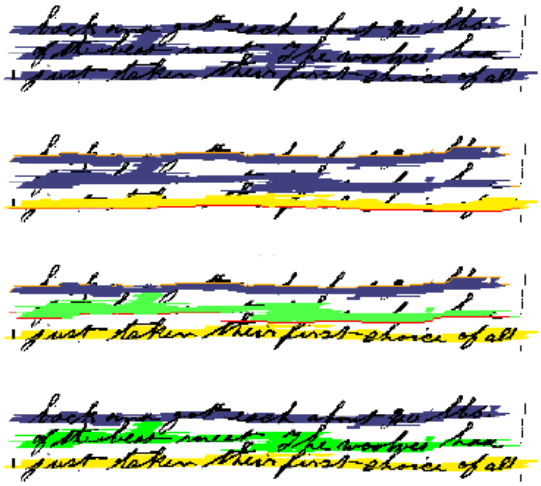


(α) (β)

Σχήμα 1.7. Οριοθέτηση των γραμμών κειμένου με τη χρήση επιπεδοσυνόλων [26]. (α) Αρχική εκτίμηση γραμμών κειμένου. (β) Οι γραμμές κειμένου μετά από 10 επαναλήψεις.

Σε μια άλλη προσέγγιση [29], που βασίζεται σε διαχωρισμό γράφων, δημιουργείται μια νέα εικόνα (πίνακας μεταβάσεων) στην οποία κάθε pixel έχει τιμή ίση με το πλήθος των μεταβάσεων από κείμενο σε φόντο που γίνονται σε ένα παράθυρο ανάλυσης με κέντρο το pixel αυτό. Με τη δυαδικοποίηση της νέας εικόνας προκύπτουν νέα CCs που αντιστοιχούν σε μία ή περισσότερες γραμμές κειμένου (σχ. 1.8). Η διάκρισή τους γίνεται με βάση το πλήθος των εσωτερικών σημείων που περιέχονται στο ορθογώνιο το οποίο περικλείει κάθε CC. Εσωτερικό

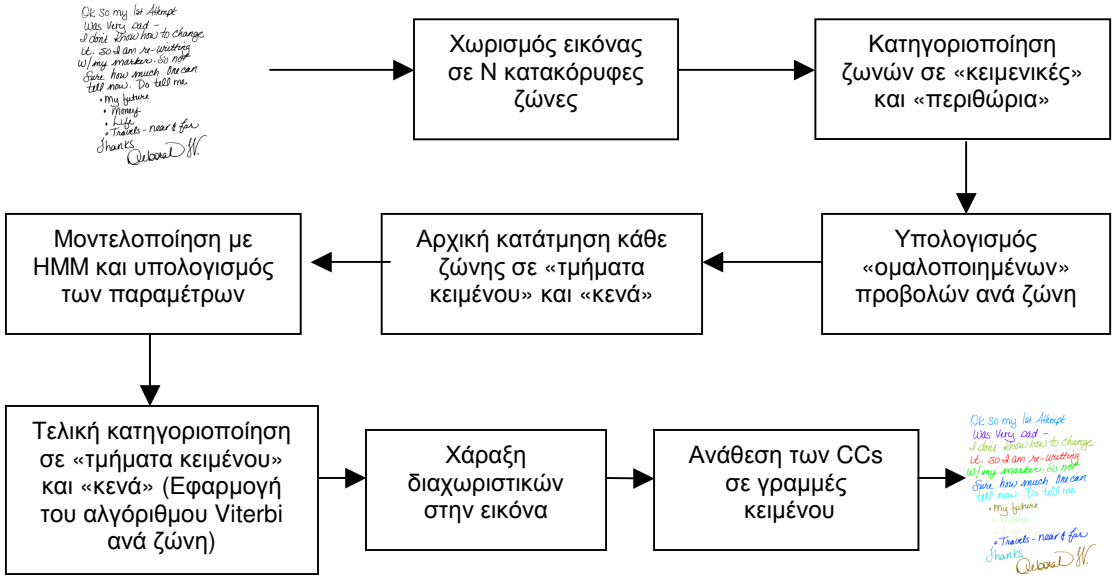
σημείο ονομάζεται κάθε pixel του φόντου που έχει εκατέρωθεν δύο τουλάχιστον pixels του CC με την ίδια τεταγμένη. Αν ο λόγος του πλήθους των εσωτερικών σημείων ενός CC προς το εμβαδό του ορθογωνίου είναι μεγαλύτερος από ένα προκαθορισμένο κατώφλι, τότε δημιουργείται ο αντίστοιχος γράφος με αρχικό (source) και τελικό σημείο (sink) το ανώτερο (κατώτερο) σημείο του CC και κόστη ανάλογα της σχέσης γειτνίασης των pixels (π.χ. 1 για γείτονες κατά τον κατακόρυφο άξονα, 400 για γείτονες κατά τον οριζόντιο άξονα και 2 για διαγώνιους γείτονες). Με τον τρόπο αυτό, ενισχύεται ο διαχωρισμός του γράφου κατά τον οριζόντιο άξονα [30]. Η διαδικασία επαναλαμβάνεται μέχρι τα νέα CCs να μην περιέχουν σημαντικό πλήθος εσωτερικών σημείων (σχ. 1.8).



Σχήμα 1.8. Κατάτμηση των αρχικών CCs σε μικρότερα που αντιστοιχούν σε μία μόνο γραμμή κειμένου με διαχωρισμό γράφων [29].

1.2. Προτεινόμενη μέθοδος με την εφαρμογή των επιμέρους προβολών

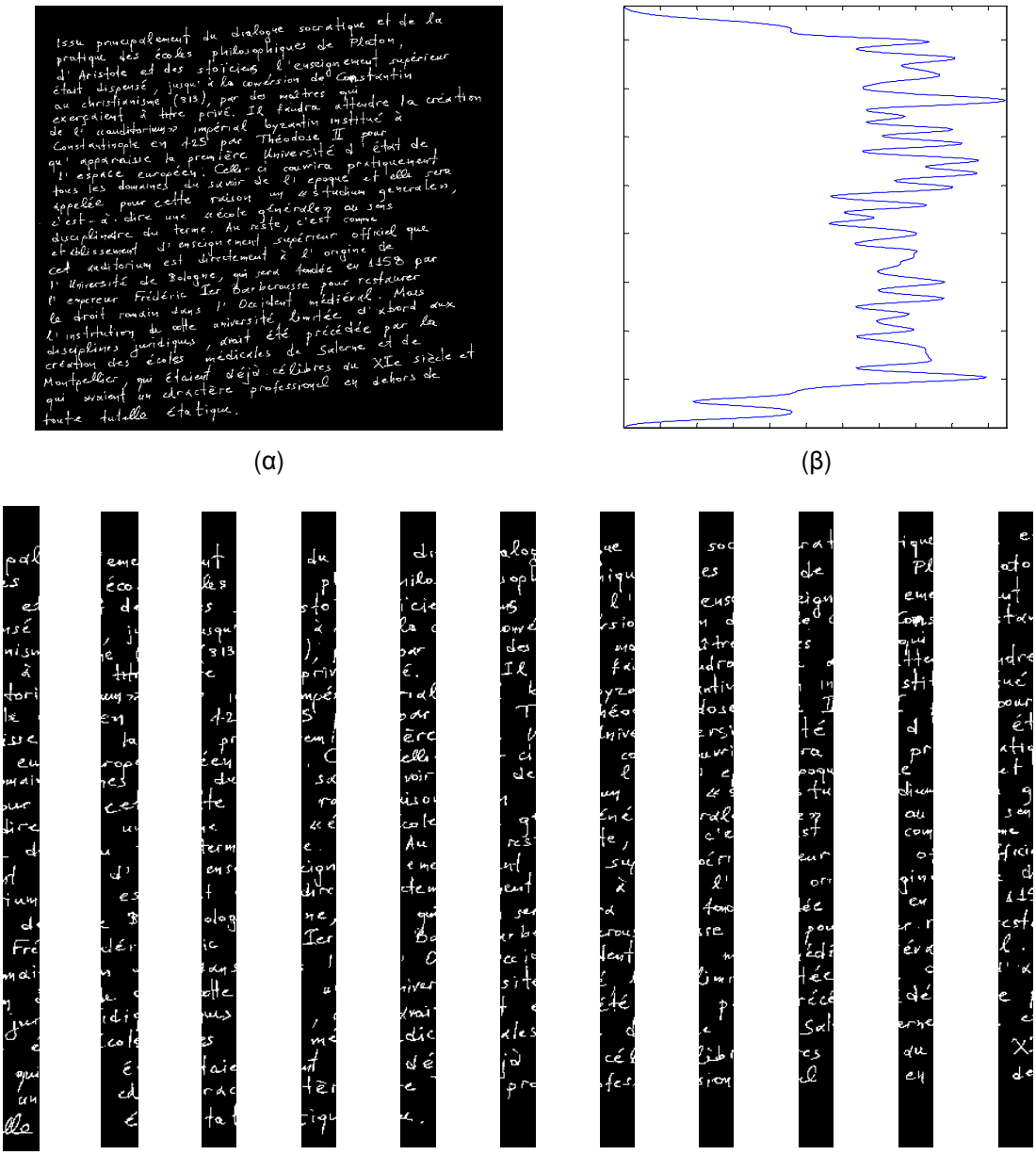
Στη συγκεκριμένη ενότητα παρουσιάζεται μια μέθοδος για την κατάτμηση της εικόνας χειρόγραφου κειμένου σε γραμμές κειμένου. Τα στάδια επεξεργασίας της προτεινόμενης μεθόδου [31] παρουσιάζονται στο σχ. 1.9.



Σχήμα 1.9. Τα στάδια της προτεινόμενης μεθόδου για την κατάτμηση χειρόγραφου κειμένου σε γραμμές, με τη χρήση επιμέρους προβολών.

1.2.1. Χωρισμός εικόνας σε κατακόρυφες ζώνες

Αρχικά, η εικόνα κειμένου A χωρίζεται σε N κατακόρυφες ζώνες A_i, $i = 1,...,N$ ίσου πλάτους. Το πλάτος κάθε ζώνης θα πρέπει να είναι αρκετά μικρό, ώστε να είναι ρεαλιστική η υπόθεση ότι η τοπική κλίση ανά γραμμή είναι αμελητέα. Ταυτόχρονα, το πλάτος της κάθε ζώνης θα πρέπει να είναι αρκετά μεγάλο, ώστε να περιέχει αρκετή πληροφορία (pixels κειμένου) για να είναι αξιόπιστες οι μετρήσεις που πρόκειται να γίνουν ανά ζώνη. Μια λογική επιλογή που ικανοποιεί τις απαιτήσεις αυτές, είναι το πλάτος κάθε ζώνης να ισούται με το 5% του συνολικού πλάτους της εικόνας κειμένου (δηλ. $N = 20$). Σημειώνεται ότι περαιτέρω ανάλυση για την επιλογή της τιμής αυτής και την επίδρασή της στην αποτελεσματικότητα της μεθόδου περιγράφεται στην ενότητα 1.2.7.



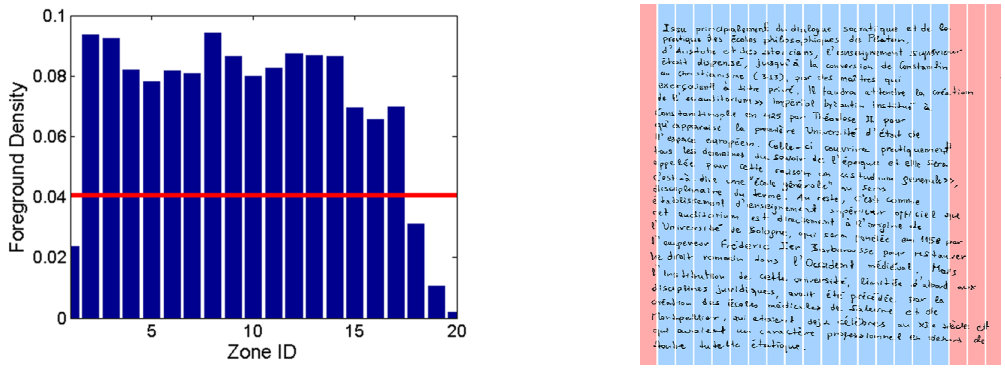
Σχήμα 1.10. (α) Η εικόνα 026.tif από ICDAR07. (β) Η ολική προβολή της. (γ) Οι ζώνες 5 ως 15.

Ένα παράδειγμα για τη χρησιμότητα αυτής της διαδικασίας παρουσιάζεται στο σχ. 1.10. Η χρήση της ολικής προβολής (σχ. 1.10β) για τον υπολογισμό των διαχωριστικών μεταξύ των γραμμών (στα σημεία που η προβολή παρουσιάζει τοπικά ελάχιστα) αποτυγχάνει, λόγω της κλίσης του κειμένου (σχ. 1.10α). Αντίθετα, μετά το διαχωρισμό της εικόνας σε κατακόρυφες ζώνες, προκύπτουν τμήματα κειμένου με αμελητέα κλίση (σχ. 1.10γ).

1.2.2. Κατηγοριοποίηση κατακόρυφων ζωνών

Με την κατάτμηση του κειμένου σε κατακόρυφες ζώνες, είναι πολύ πιθανό να δημιουργηθούν ζώνες, κυρίως κοντά στα όρια του κειμένου, οι οποίες περιέχουν ελάχιστα pixels κειμένου, λόγω

της ακανόνιστης μορφοποίησης των χειρόγραφων κειμένων. Οι συγκεκριμένες ζώνες δεν περιέχουν σημαντική πληροφορία και επομένως θα αποκλειστούν από τα ακόλουθα βήματα επεξεργασίας.



(α) (β)

Σχήμα 1.11. Κατηγοριοποίηση των ζωνών για την εικόνα κειμένου 012.tif από ICDAR07. (α) Οι πυκνότητες κειμένου για τις κατακόρυφες ζώνες (μπλε ράβδοι) και το υπολογιζόμενο κατώφλι (κόκκινη γραμμή). (β) Οι ζώνες με πυκνότητα κειμένου μικρότερη του κατωφλίου, αντιστοιχούν σε περιθώρια (ροζ).

Προκειμένου να εντοπιστούν αυτές οι ζώνες, υπολογίζεται η πυκνότητα κειμένου (foreground density) ανά ζώνη και συγκρίνεται με ένα κατώφλι (*th*) που υπολογίζεται για κάθε εικόνα. Η τιμή του κατωφλίου τίθεται ίση με το μισό της ενδιάμεσης τιμής των πυκνοτήτων των ζωνών. Η ενδιάμεση τιμή επιλέγεται αφενός γιατί το πλήθος των ζωνών είναι μικρό (*N*=20) και αφετέρου γιατί κάποιες τιμές των πυκνοτήτων θα είναι πιθανότατα ακραίες. Κάθε ζώνη με πυκνότητα μεγαλύτερη ή μικρότερη του κατωφλίου χαρακτηρίζεται ως «ζώνη κειμένου» ή «ζώνη περιθωρίου» αντίστοιχα. Η κατηγοριοποίηση των ζωνών γίνεται με την ακόλουθη σχέση:

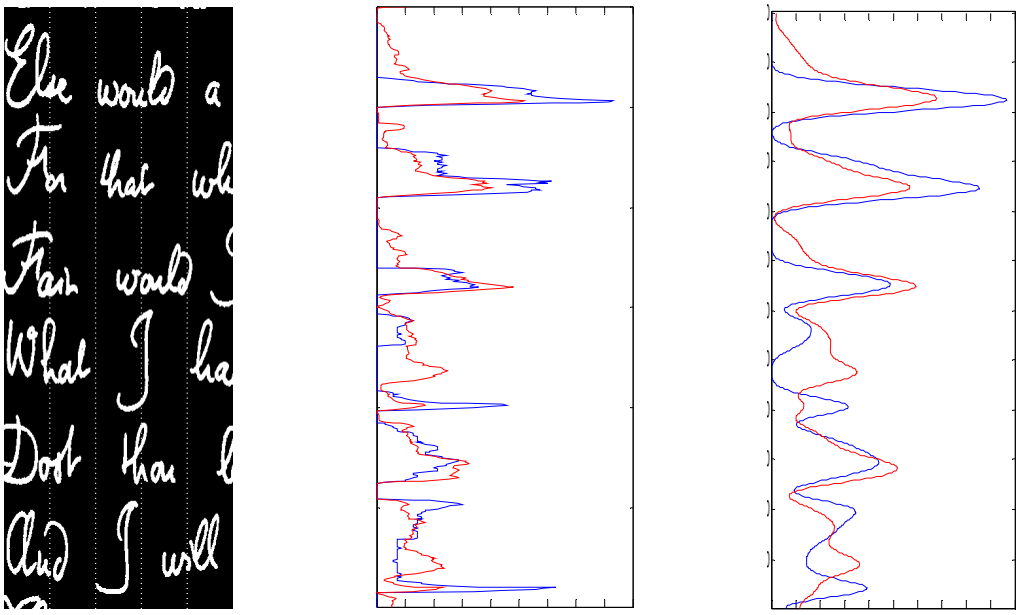
$$\delta_i = [1 + T(d_i - th)]/2,\ i \in \{1,2,...,N\} \qquad \text{(Εξ. 1.2)}$$

όπου $T(x) = \begin{cases} 1 & x \geq 0 \\ -1 & x < 0 \end{cases}$, $d_i = (1/M_i) \cdot \sum_{j=1}^{M_i} A_i(j)$ η πυκνότητα κειμένου της *i-οστης* ζώνης, M_i το πλήθος των pixels της *i-οστης* ζώνης και $th = 0.5 \cdot median\{d_1,...,d_i,...,d_N\}$. Στο σχ. 1.11 παρουσιάζεται η διάκριση των ζωνών σε «κειμενικές» ($\delta_i = 1$) και «περιθώριο» ($\delta_i = 0$).

1.2.3. Υπολογισμός «ομαλοποιημένων» προβολών

Όπως έχει αναφερθεί, οι επιμέρους προβολές είναι ιδιαίτερα ευαίσθητες σε περιπτώσεις που το κείμενο είτε είναι αραιογραμμένο, είτε παρουσιάζει ιδιαίτερη μεταβλητότητα στο μέγεθος των χαρακτήρων κατά μήκος της ίδιας γραμμής. Μπορεί κανείς να διακρίνει το φαινόμενο αυτό

παρατηρώντας τους λοβούς που εμφανίζει η προβολή PR_3 (μπλε χρώμα στο σχ. 1.12β) της τρίτης ζώνης του κειμένου 013.tif από ICDAR07 (σχ. 1.12α) στις θέσεις που πραγματώνεται το γράμμα «I» ως  και  . Για να περιοριστεί η επιρροή τέτοιων πραγματώσεων στις προβολές, «ομαλοποιούμε» την προβολή κάθε ζώνης λαμβάνοντας υπόψη και τις προβολές των *M* γειτονικών ζωνών (εκατέρωθεν).



(α) (β) (γ)

Σχήμα 1.12. (α) Τμήματα των ζωνών 1 ως 5 της εικόνας 013.tif από ICDAR07. (β) Η προβολή της ζώνης 3 (μπλε) και η αντίστοιχη ομαλοποιημένη (κόκκινη) προβολή. (γ) Οι προβολές μετά την εφαρμογή βαθυπερατού φίλτρου (Gaussian) τάξης ίσης με το μέσο ύψος των CCs του κειμένου.

Είναι όμως απαραίτητο, η συνεισφορά κάθε γειτονικής ζώνης να μην είναι ισότιμη με αυτή της τρέχουσας ζώνης, ώστε το αποτέλεσμα να μην επηρεάζεται σημαντικά από την ενδεχόμενη κλίση των γραμμών στις γειτονικές ζώνες. Για το λόγο αυτό, στον υπολογισμό της συνεισφοράς των γειτονικών προβολών, χρησιμοποιούνται κανονικοποιημένοι συντελεστές, που φθίνουν εκθετικά ως προς την απόσταση της «τρέχουσας ζώνης» από τις γειτονικές. Τα κανονικοποιημένα βάρη ορίζονται με την ακόλουθη σχέση:

$$w_j = \exp(-c\cdot|\,j\,|/(2M+1)) \,/\, \sum_{k=-M}^{M} \exp(-c\cdot|\,k\,|/(2M+1)), \quad j\in\{-M,...,0,...,M\} \quad \text{(Εξ. 1.3)}$$

Λαμβάνοντας υπόψη τις 4 γειτονικές ζώνες (2 εκατέρωθεν, *M=2*) και θέτοντας *c=3*, προκύπτουν οι ακόλουθες τιμές των βαρών: $w_{-2} = w_2 = 0.1116$, $w_{-1} = w_1 = 0.2033$, $w_0 = 0.3704$.

Επομένως, η ομαλοποιημένη προβολή (ΟΠ) SPR_i της *i-οστης* «ζώνης κειμένου» ορίζεται ως εξής:

$$SPR_i = \sum_{j=-2}^{2} \delta_{i+j} \cdot w_j \cdot PR_{i+j} \qquad \text{(Εξ. 1.4)}$$

Στο σχ. 1.12β (κόκκινο χρώμα) παρουσιάζεται η SPR_3 του κειμένου της εικόνας 013.tif από ICDAR07. Η διαφοροποίηση των προβολών φαίνεται παραστατικά στο σχ. 1.12γ, όπου οι θέσεις των τμημάτων κειμένου κάθε γραμμής κειμένου, αναδεικνύονται με τους αντίστοιχους λοβούς της «ομαλοποιημένης» προβολής (κόκκινο χρώμα). Η διαφορά είναι εμφανής και στις προβολές του σχ. 1.12γ.

1.2.4. Κατάτμηση κάθε ζώνης σε «τμήματα κειμένου» και «κενά»

Οι θέσεις στις οποίες η ΟΠ παρουσιάζει τοπικά ελάχιστα, αντιστοιχούν στις θέσεις των υποψήφιων διαχωριστικών των γραμμών ανά ζώνη. Ο εντοπισμός των «πραγματικών» διαχωριστικών, δηλαδή των σημαντικών τοπικών ελάχιστων, δεν είναι πάντοτε μια απλή διαδικασία. Συνήθως, εξετάζεται το ύψος και το πλάτος κάθε λοβού και με τη χρήση κάποιου κατωφλίου αποφασίζεται η αποδοχή ή απόρριψη του τοπικού ελάχιστου ως θέση διαχωριστικού μεταξύ των γραμμών.

Η προτεινόμενη μέθοδος αποφεύγει την επισφαλή διαδικασία της άμεσης εκτίμησης των διαχωριστικών ανά ζώνη και υιοθετεί τη μοντελοποίηση των ζωνών ως ακολουθίες παρατηρήσεων που παράγονται από ένα Κρυφό Μαρκοβιανό Μοντέλο (Hidden Markov Model). Τα στάδια της διαδικασίας είναι:

α) σε κάθε ζώνη ορίζονται περιοχές δύο κατηγοριών: «τμήματα κειμένου» και «κενά»,

β) υπολογισμός στατιστικών τιμών για κάθε τάξη και υιοθέτησή τους για την παραμετροποίηση του μοντέλου,

γ) χρήση του HMM ανά ζώνη για την επίτευξη νέας κατηγοριοποίησης.

α) Αρχικός διαχωρισμός

Οι θέσεις των άνω και κάτω ορίων των περιοχών κειμένου ανά ζώνη προσεγγίζονται αντίστοιχα, από τα τοπικά μέγιστα και ελάχιστα της πρώτης παραγώγου της εκάστοτε ΟΠ. Προφανώς, η θέση στην οποία η ΟΠ «αυξάνεται απότομα», δηλαδή σε τοπικό μέγιστο της πρώτης παραγώγου (σχ. 1.13β με κόκκινο), είναι η γραμμή από την οποία ξεκινά μια περιοχή με πολλά pixels κειμένου και επομένως αποτελεί πιθανό άνω όριο μιας γραμμής κειμένου. Ομοίως, η θέση στην οποία η ΟΠ «ελαττώνεται απότομα», δηλαδή σε τοπικό ελάχιστο της πρώτης παραγώγου, είναι η γραμμή από την οποία ξεκινά μια περιοχή με λίγα pixels κειμένου και επομένως αποτελεί πιθανό κάτω όριο μιας γραμμής κειμένου.

Η τιμή της πρώτης παραγώγου σε ένα σημείο j της ΟΠ είναι η κλίση της εφαπτομένης της στο σημείο αυτό. Επομένως, αν επιλεγούν l σημεία εκατέρωθεν του j (άρα $2l+1$ σημεία

συνολικά), η κλίση της ευθείας που τα προσεγγίζει, είναι η εκτίμηση της πρώτης παραγώγου. Η πρώτη παράγωγος ΔSPR_i της ΟΠ της ζώνης i υπολογίζεται ως εξής:

$$\Delta SPR_i(j) = \sum_{k=-l}^{l} k \cdot SPR_i(j+k) \quad / \quad \sum_{k=-l}^{l} k^2 \qquad \text{(Εξ. 1.5)}$$

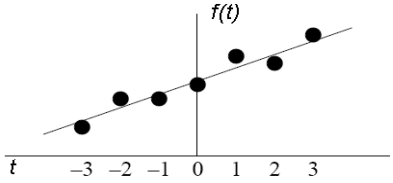
Η τιμή l είναι ιδιαίτερα κρίσιμη γιατί ορίζει το βαθμό ομαλοποίησης της παραγώγου. Αν είναι πολύ μικρή, τότε στο αποτέλεσμα θα εμφανίζονται οι μικρές και μη σημαντικές αλλαγές της ΟΠ, ενώ αν είναι πολύ μεγάλη είναι πιθανό να μην εμφανίζονται κρίσιμες αλλαγές της ΟΠ. Μια λογική επιλογή είναι το μέσο ύψος των CCs που περιέχονται στην αρχική δυαδική εικόνα, υποθέτοντας ότι με τον τρόπο αυτό προσεγγίζεται το μέσο ύψος των γραμμών κειμένου. Σημειώνεται ότι για τον εντοπισμό των CCs εφαρμόζεται ο αλγόριθμος που περιγράφεται στο [6].

Ακολούθως, παρουσιάζεται η διαδικασία που εξηγεί τον υπολογισμό της παραγώγου μέσω της Εξ. 1.5. Έστω, ένα σύνολο σημείων (x_i, y_i), $i = 1,...,N$, και ότι είναι επιθυμητή η προσέγγισή του από ένα πολυώνυμο τάξης P. Τότε, για τα αντίστοιχα σημεία $(x_i, \hat{y}_i)$ ισχύει ότι $\hat{y}_i = \sum_{k=0}^{P} a_k x_i^k$, όπου a_k οι συντελεστές του πολυωνύμου. Επομένως, το σφάλμα προσέγγισης σε κάθε σημείο είναι: $e_i = y_i - \hat{y}_i$. Ως εκτίμηση του συνολικού σφάλματος E επιλέγεται το άθροισμα των τετραγώνων των σφαλμάτων σε κάθε σημείο $E = \sum_{i=1}^{N} e_i^2$.

Το πολυώνυμο που προσεγγίζει καλύτερα το σύνολο των σημείων, είναι αυτό το οποίο ελαχιστοποιεί το ολικό σφάλμα. Άρα, για τους συντελεστές του βέλτιστου πολυωνύμου ισχύει ότι: $\frac{\partial E}{\partial a_m} = 0 \Leftrightarrow \sum_{i=1}^{N}((y_i - \sum_{k=0}^{P} a_k x_i^k)\, x_i^m) = 0 \Leftrightarrow \sum_{i=1}^{N} y_i x_i^m = \sum_{k=0}^{P}(a_k \sum_{i=1}^{N} x_i^{k+m}),\ m = 0,...,P$

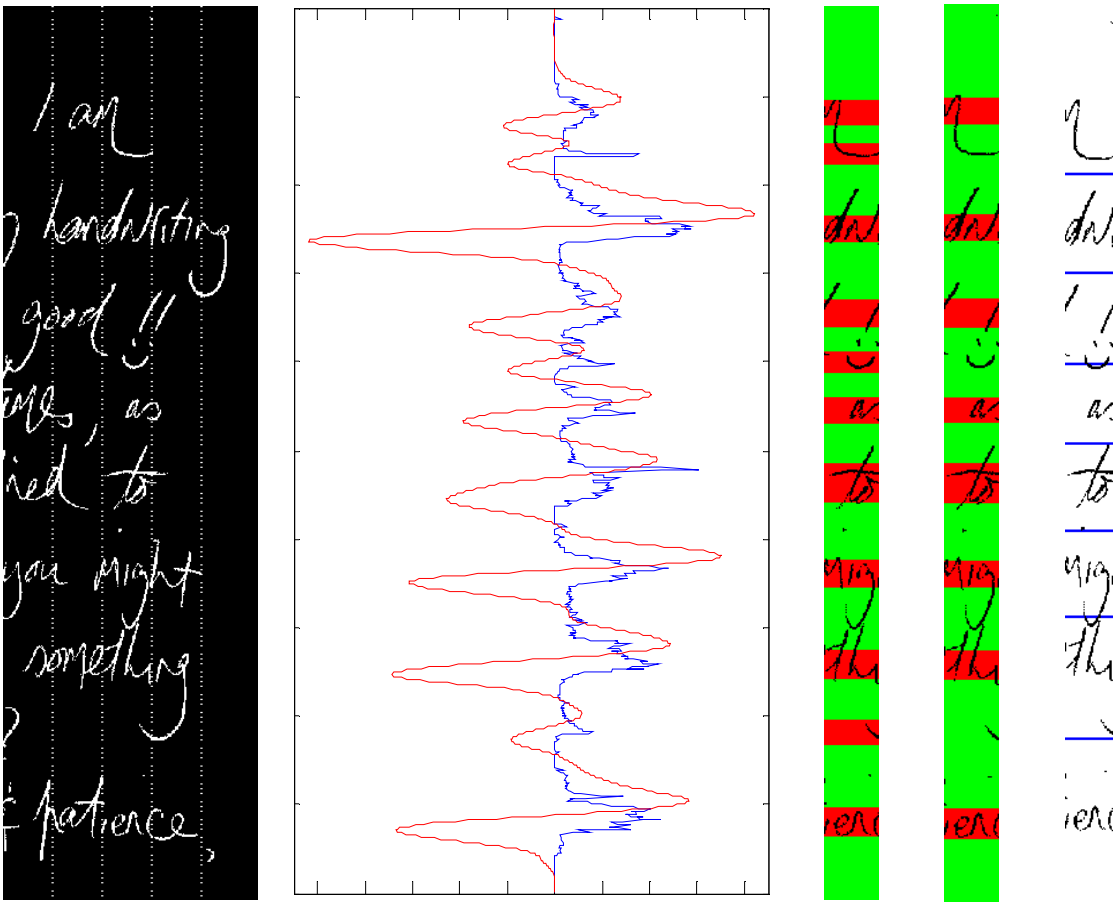
ή αλλιώς $\begin{bmatrix} \sum x^0 & \sum x^1 & \sum x^2 & \dots \\ \sum x^1 & \sum x^2 & \sum x^3 & \\ \sum x^2 & \sum x^3 & \sum x^4 & \\ \vdots & & & \ddots \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \end{bmatrix} = \begin{bmatrix} \sum yx^0 \\ \sum yx^1 \\ \sum yx^2 \\ \vdots \end{bmatrix}$.

Έστω τώρα η συνάρτηση f του διπλανού σχήματος.



Αν για τον υπολογισμό της πρώτης παραγώγου της στο t_i=0 επιλεγούν l σημεία εκατέρωθεν ($2l+1$ συνολικά), τότε οι παράμετροι της βέλτιστης ευθείας που τα προσεγγίζει, υπολογίζονται ως εξής:

$\begin{bmatrix} 2l+1 & 0 \\ 0 & \sum t^2 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} \sum f(t)t^0 \\ \sum f(t)t^1 \end{bmatrix}$ και επομένως, $a_1 = \frac{\sum f(t)t}{\sum t^2}$. Υιοθετώντας αυτή την προσέγγιση για τον υπολογισμό της πρώτης παραγώγου σε κάθε σημείο της ΟΠ, προκύπτει η (Εξ. 1.5).



(α) (β) (γ) (δ) (ε)

Σχήμα 1.13. (α) Τμήματα των ζωνών 16 ως 20 της εικόνας 067.tif από ICDAR07. (β) Η ΟΠ της ζώνης 18 (μπλε) και η πρώτη παράγωγός της (κόκκινη). (γ) Αρχικός διαχωρισμός της ζώνης 18 σε τμήματα κειμένου (κόκκινο) και κενά (πράσινα). (δ) Τελική κατηγοριοποίηση των τμημάτων σε τμήματα κειμένου και κενά. (ε) Τα διαχωριστικά των γραμμών κειμένου για τη ζώνη 18.

Η επιλογή των τοπικών ακροτάτων της ΔSPR_i και η οριοθέτηση των περιοχών κειμένου και κενών, αποτελεί ένα στάδιο υπερ-κατάτμησης κάθε ζώνης σε εναλλασσόμενες περιοχές (σχ. 1.13γ). Επίσης, μπορεί να θεωρηθεί ως μια αρχική κατηγοριοποίηση των περιοχών της εικόνας σε δύο τάξεις.

β) Μοντελοποίηση με χρήση HMM

Για να επιτευχθεί η νέα κατηγοριοποίηση που θα αναδείξει ως περιοχές κειμένου μόνο αυτές που αντιστοιχούν στο κύριο τμήμα κάθε γραμμής κειμένου ανά ζώνη, ορίστηκε ένα απλό

Κρυφό Μαρκοβιανό Μοντέλο (Hidden Markov Model, HMM) [32] με δύο καταστάσεις, μία για τις περιοχές κειμένου και η άλλη για τα κενά, που δηλώνονται με c_1 και c_2 αντίστοιχα. Για τον υπολογισμό των παραμέτρων του προτεινόμενου μοντέλου χρησιμοποιούνται τα στατιστικά στοιχεία που εξάγονται από τις ήδη κατηγοριοποιημένες περιοχές σε όλη τη σελίδα του κειμένου. Δεδομένου του μοντέλου και της ακολουθίας παρατηρήσεων ανά ζώνη, εφαρμόζεται σε κάθε ζώνη ο αλγόριθμος Viterbi για να υπολογιστεί η πιο πιθανή ακολουθία καταστάσεων.

Τα ενδεχόμενα να βρεθούμε στη μια ή στην άλλη κατάσταση στην αρχή του πειράματος είναι ισοπίθανα και επομένως οι αρχικές πιθανότητες ορίζονται ως $\pi_1 = \pi_2 = 0.5$.

Για τη μοντελοποίηση των πιθανοτήτων μετάβασης προτείνεται η εκθετική κατανομή με παράμετρο τη μέση τιμή των υψών των περιοχών σε κάθε κλάση m_j, $j \in \{1,2\}$:

$$a_{jj}(i) = \mathrm{P}\left(s_{[h,h+H_i]} = c_j \,\middle|\, s_{[h-H_{i-1},h]} = c_j\right) = \exp\left(-H_i / m_j\right), \quad j \in \{1,2\} \qquad \text{(Εξ. 1.6)}$$

όπου με H_i δηλώνεται το ύψος της *i-οστής* περιοχής και με $s_{[h,h+H_i]}$ η κατάσταση της περιοχής που εκτείνεται από τη θέση h ως τη θέση $h+H_i$. Για της πιθανότητες μετάβασης $a_{12}(i)$ και $a_{21}(i)$ ισχύει ότι: $a_{12}(i) = 1 - a_{11}(i)$ και $a_{21}(i) = 1 - a_{22}$. Η χρήση της εκθετικής κατανομής για τον υπολογισμό των πιθανοτήτων μετάβασης δηλώνει ότι η πιθανότητα αλλαγής κατάστασης αυξάνεται καθώς το ύψος της εξεταζόμενης περιοχής προσεγγίζει ή ξεπερνά τη μέση τιμή του ύψους στην προηγούμενη κατάσταση.

Οι πιθανότητες παρατήρησης για κάθε κατάσταση (emission probabilities) θα υπολογιστούν με βάση τις πυκνότητες σε pixels κειμένου των αρχικά κατηγοριοποιημένων τμημάτων. Για τη μοντελοποίηση των πιθανοτήτων παρατήρησης χρησιμοποιήθηκε η κανονική κατανομή, ως εξής:

$$b_j(i) = \mathrm{p}\left(x_i \,\middle|\, s_{[h,h+H_i]} = c_j\right) \sim \boldsymbol{N}\left(\mu_j, \sigma_j^2\right), \quad j \in \{1,2\} \qquad \text{(Εξ. 1.7)}$$

όπου η τυχαία μεταβλητή x_i δηλώνει τη λογαριθμική πυκνότητα της *i-οστής* περιοχής $s_{[h,h+H_i]}$ που οριοθετείται από τα ύψη h και $h+H_i$. Η μέση τιμή και η μεταβλητότητα των λογαριθμικών πυκνοτήτων των περιοχών που ανήκουν στην κατάσταση-κλάση c_j δηλώνονται με μ_j και σ_j^2 αντίστοιχα.

γ) Τελική κατηγοριοποίηση

Κάθε ζώνη του κειμένου μπορεί να θεωρηθεί ως μια ακολουθία παρατηρήσεων $O = \{O_1, O_2, ..., O_T\}$, όπου O_t, $t = 1,2,...,T$ είναι τα διανύσματα που περιέχουν την λογαριθμική πυκνότητα καθενός από τα T τμήματα, στα οποία έχει χωριστεί κάθε κατακόρυφη

ζώνη. Δεδομένου του μοντέλου $\lambda = (A, B, \pi)$, θα πρέπει να βρεθεί η ακολουθία καταστάσεων $Q = \{q_1, q_2, ..., q_T\}$ που παράγει με τη μεγαλύτερη πιθανότητα τη συγκεκριμένη ακολουθία παρατηρήσεων. Για τον προσδιορισμό της βέλτιστης ακολουθίας καταστάσεων, εφαρμόζουμε τον αλγόριθμο Viterbi [33] σε κάθε ζώνη. Ένα αριθμητικό παράδειγμα για την εφαρμογή του Viterbi παρουσιάζεται στο Παράρτημα Α. Όμως, η ακολουθία των καταστάσεων δηλώνει επί της ουσίας τη νέα κατηγοριοποίηση των εξεταζόμενων περιοχών σε τμήματα «κειμένου» και «κενού» (σχ. 1.13δ). Τέλος, τα διαχωριστικά των γραμμών κειμένου ανά ζώνη, χαράσσονται στο μέσο των τμημάτων κενού (σχ. 1.13ε).

1.2.5. Χάραξη των διαχωριστικών στην εικόνα κειμένου

Με τα προηγούμενα βήματα έχει επιτευχθεί ο εντοπισμός των διαχωριστικών σε κάθε κειμενική ζώνη. Ορίζουμε με $\left\{\psi_i^j, j = 1, ..., S_i\right\}$ τις θέσεις των διαχωριστικών της *i-οστής* ζώνης όπου S_i το πλήθος των διαχωριστικών στη ζώνη αυτή. Σημειώνεται ότι με τον όρο θέση δηλώνουμε τον αύξοντα αριθμό της γραμμής της εικόνας κειμένου στην οποία βρίσκεται το διαχωριστικό. Η χάραξη των διαχωριστικών σε όλη την εικόνα κειμένου γίνεται με το συνδυασμό των διαχωριστικών που προέκυψαν από το προηγούμενο στάδιο, ξεκινώντας από την πιο αριστερά ζώνη κειμένου και κινούμενοι προς τα δεξιά κατά το πλάτος της εικόνας.

Αρχικά, κάθε διαχωριστικό ψ_i^j συσχετίζεται με το κοντινότερό του (με βάση την κατακόρυφη απόσταση) από την *(i+1)-οστή* (π.χ. ψ_{i+1}^k) ζώνη. Από τη συσχέτιση αυτή, θα προκύψει κάποια από τις ακόλουθες τρεις περιπτώσεις:

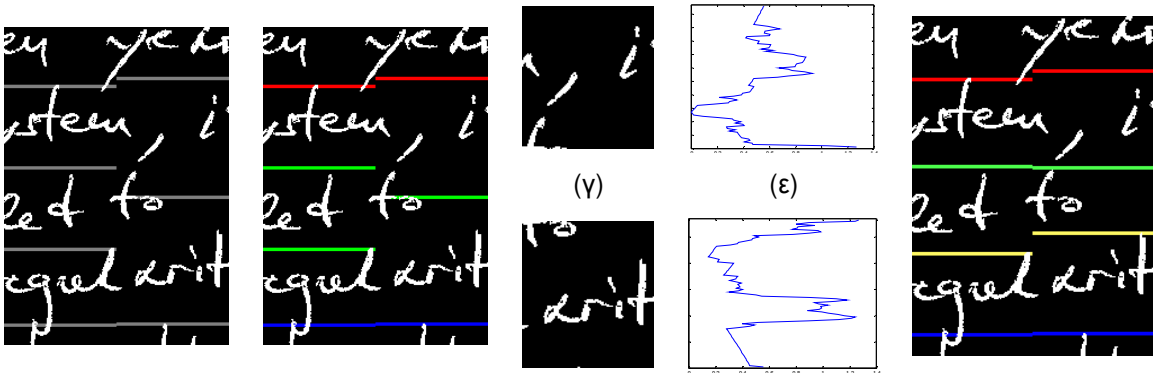
α) το ψ_{i+1}^k σχετίζεται με ένα μόνο διαχωριστικό από την *i-οστή* ζώνη (π.χ. τα κόκκινα και μπλε διαχωριστικά του σχήματος σχ. 1.14β). Τότε η αντιστοίχηση είναι «1 προς 1» και δεν απαιτείται περαιτέρω επεξεργασία.

β) το ψ_{i+1}^k σχετίζεται με *r* διαχωριστικά της *i-οστής* ζώνης, έστω $\left\{\psi_i^\ell, \ell = j, ..., j + r - 1\right\}$ (πράσινα διαχωριστικά στο σχ.1.14β). Αυτό συμβαίνει είτε γιατί κάποια γραμμή κειμένου είναι πιο σύντομη από τις άλλες, είτε γιατί υπάρχουν ιδιαίτερα μεγάλα κενά μεταξύ των λέξεων. Τότε, για κάθε ψ_i^ℓ ορίζεται η περιοχή *L* της *(i+1)-οστής* ζώνης, η οποία οριοθετείται από τα διαχωριστικά της *(i+1)-οστής* ζώνης που βρίσκονται κοντινότερα στο ψ_i^ℓ (σχ. 1.14γ,δ). Σε κάθε περιοχή *L* θα επιλεγεί ένα κατάλληλο διαχωριστικό που θα συσχετιστεί με το αντίστοιχο ψ_i^ℓ. Τα νέα διαχωριστικά στην *(i+1)-οστή* ζώνη θα πρέπει να εκτείνονται κοντά στα αντίστοιχα διαχωριστικά της *i-οστής* ζώνης και να συναντούν όσο το δυνατό λιγότερα pixels κειμένου.

Αυτές οι απαιτήσεις μπορούν να εκφραστούν ως ένα πρόβλημα υπολογισμού της γραμμής (row) m της περιοχής *L* που ελαχιστοποιεί την ακόλουθη συνάρτηση:

$$Q_m = (d_m + c_1) \cdot (P_m + c_2) \ , \qquad \text{(Εξ. 1.8)}$$

όπου d_m είναι η κανονικοποιημένη απόσταση της m-οστής γραμμής από το ψ_i^{ℓ}, P_m είναι η κανονικοποιημένη τιμή της προβολής της περιοχής *L* στην m-οστή γραμμή και τα c_1 και c_2 είναι επιλεγμένες σταθερές που τέθηκαν ίσες με τη μονάδα.



Σχήμα 1.14. Παράδειγμα του αλγόριθμου χάραξης των διαχωριστικών για την 15η ζώνη της εικόνας 025.tif ICDAR07. (α) Τα διαχωριστικά των ζωνών 14 και 15. (b) Τα διαχωριστικά με το ίδιο χρώμα συσχετίζονται αρχικά. (γ, δ) Οι περιοχές στις οποίες θα αναζητηθούν τα νέα διαχωριστικά. (ε, στ) Τα γραφήματα της μετρικής για τις αντίστοιχες περιοχές. (ζ). Η τελική συσχέτιση των διαχωριστικών.

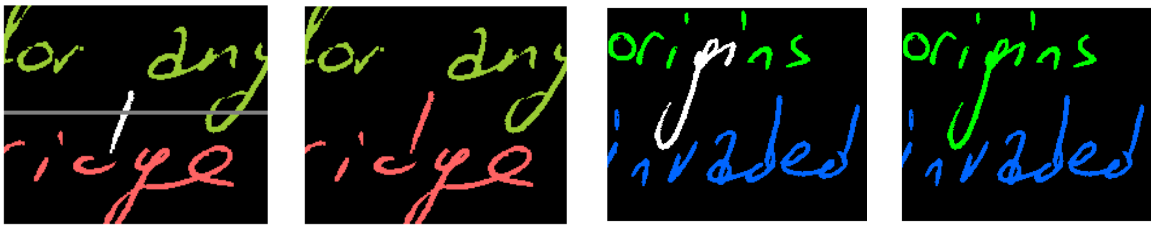
Ο πρώτος παράγοντας της (Εξ. 1.8) δηλώνει την απόσταση του νέου διαχωριστικού στην *(i+1)-οστή* ζώνη από το εξεταζόμενο διαχωριστικό ψ_i^{ℓ}. Ο δεύτερος όρος αντιστοιχεί στο πλήθος των pixels κειμένου που συναντά το νέο διαχωριστικό. Τα σχ. 1.14ε,στ δείχνουν τη γραφική παράσταση της συνάρτησης Q_m για τις περιοχές των σχ. 1.14γ,δ αντίστοιχα. Τα δύο ζεύγη των συσχετιζόμενων διαχωριστικών που προκύπτουν παρουσιάζονται στο σχ. 1.14ζ (πράσινο και κίτρινο). Από το σχ. 1.14β, μπορεί κανείς να παρατηρήσει ότι αν η σταθερά c_1 ήταν ίση με το 0, κάθε νέο διαχωριστικό θα προέκυπτε ως προέκταση του αντίστοιχου ψ_i^{ℓ} και επομένως θα έτεμνε τον τόνο και το χαρακτήρα «t» της λέξης «regularity». Η παράλειψη της σταθεράς c_2, θα οδηγούσε στην επιλογή μιας γραμμής που δεν έχει pixels κειμένου, χωρίς να λαμβάνεται υπόψη η απόσταση από το διαχωριστικό ψ_i^{ℓ}, και επομένως θα ήταν πολύ πιθανό να μην ακολουθείται η τοπική κλίση της γραμμής κειμένου.

γ) το ψ_{i+1}^{k} δε σχετίζεται με κανένα διαχωριστικό της i-οστής γραμμής. Αυτό σημαίνει ότι υπάρχει μια νέα γραμμή κειμένου, που ξεκινά από αυτή τη ζώνη και πέρα. Η χάραξη των σχετικών διαχωριστικών γίνεται με την εφαρμογή της τεχνικής που περιγράφεται στο (β) με κατεύθυνση από δεξιά προς αριστερά.

1.2.6 Ανάθεση των CCs σε γραμμές κειμένου

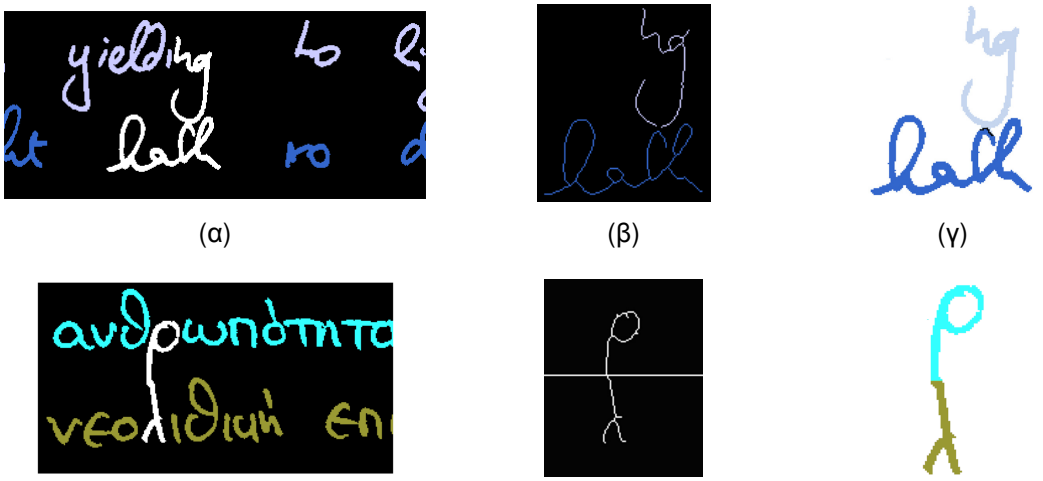
Ο αντικειμενικός στόχος ενός αλγορίθμου κατάτμησης κειμένου σε γραμμές κειμένου είναι η ανάθεση των CCs του κειμένου σε αυτές. Στο τελικό στάδιο επεξεργασίας, η ταξινόμηση των CCs του κειμένου στις κατάλληλες γραμμές γίνεται με τη χρήση ευρετικών κανόνων. Προφανώς, κάθε γραμμή οριοθετείται από δύο διαδοχικά (κατά τον κατακόρυφο άξονα) ανά ζώνη διαχωριστικά. Κάθε CC εντάσσεται σε μια γραμμή κειμένου αν η επικάλυψή του με την περιοχή που οριοθετούν αυτά τα διαδοχικά διαχωριστικά, είναι μεγαλύτερη από ένα συγκεκριμένο κατώφλι *R* π.χ.(75% του ύψους του εξεταζόμενου CC). Σημειώνεται ότι περαιτέρω ανάλυση για την επιλογή της τιμής αυτής και την επίδρασή της στην αποτελεσματικότητα της μεθόδου, παρουσιάζεται στην ενότητα 1.2.7.

Για τα CCs που δεν επαληθεύουν τον παραπάνω κανόνα, θα εξεταστεί αν θα διαχωριστούν μια και περιέχουν χαρακτήρες από δύο ή περισσότερες γραμμές, ή όχι μια και απλά έχουν μεγάλο ύψος. Έστω ένα τέτοιο CC που εκτείνεται στην i–οστή ζώνη και στις j και (j+1) γραμμές κειμένου. Η i-οστή ζώνη επεκτείνεται οριζόντια και από τις δύο πλευρές μέχρι να περιλάβει ένα εύλογο πλήθος από pixels κειμένου, που έχουν ήδη ανατεθεί σε κάθε μία από τις γραμμές j και j+1, όπως φαίνεται στα σχ. 1.15α,γ και 1.16α,δ. Έστω τώρα, N_{j}^{b} και N_{j+1}^{b} τα πλήθοι των pixels της νέας περιοχής που έχουν ήδη ανατεθεί στις γραμμές j και j+1 αντίστοιχα, και N_{j}^{a} και N_{j+1}^{a} τα πλήθη των pixels που έχουν ίδια τετμημένη με τα pixels του εξεταζόμενου CC. Οι λόγοι $r_k = N_k^a / N_k^b$ με *k=j, (j+1)* μπορούν να θεωρηθούν ως δείκτες της έλξης που ασκεί κάθε γραμμή κειμένου στο εξεταζόμενο CC. Αν και οι δύο λόγοι είναι μικρότεροι από ένα κατώφλι, που πειραματικά ορίστηκε ίσο με 0.4, τότε το εξεταζόμενο CC εντάσσεται στη γραμμή που περιέχει το μεγαλύτερο κατά ύψος τμήμα του (σχ. 1.15α,β). Με τον τρόπο αυτό καλύπτονται περιπτώσεις που το CC είναι ένα σημείο στίξης, ή τόνος, ή ένα τμήμα ενός «σπασμένου» χαρακτήρα. Αν μόνο ένας λόγος είναι μεγαλύτερος του κατωφλίου, τότε το CC εντάσσεται στην αντίστοιχη γραμμή, μια και πρόκειται συνήθως για CC με μεγάλο ύψος, πιθανότατα γιατί περιέχει ascenders και/ή descenders (σχ. 1.15γ,δ).



(α) (β) (γ) (δ)

Σχήμα 1.15. Παραδείγματα ένταξης δύο CCs της εικόνας 014.tif από ICDAR07. (α, β) Οριζόντια προέκταση της ζώνης ($6^{η}$) εκατέρωθεν. Οι λόγοι για το CC με id 257 (άσπρο) είναι $r_{πράσινο}$ = 0.2072 και $r_{κόκκινο}$=0.1248. Η κόκκινη γραμμή κειμένου περιέχει το 66% του CC και επομένως το CC ανατίθεται σε αυτή. (γ, δ) Οριζόντια προέκταση της ζώνης ($2^{η}$) εκατέρωθεν. Οι λόγοι για το CC με id id 29 (άσπρο) είναι $r_{πράσινο}$= 0.8552 και $r_{μπλε}$=0.3175 και επομένως το εξεταζόμενο CC ανατίθεται στη μπλε γραμμή κειμένου.



(α) (β) (γ)

(δ) (ε) (στ)

Σχήμα 1.16. Κατάτμηση CCs που εκτείνονται σε περισσότερες από μία γραμμές κειμένου. (α) Επέκταση της ζώνης για το CC με id 210 της εικόνας 013.tif από ICDAR07. (β) Η εκλεπτυσμένη μορφή χωρίζεται σε τρία τμήματα. (γ) Ανάθεση των αντίστοιχων τμημάτων του CC σε γραμμές κειμένου. (δ) Επέκταση της ζώνης για το CC με id 161 της εικόνας 003.tif από ICDAR07. (ε) Η εκλεπτυσμένη μορφή και το διαχωριστικό των γραμμών. Δεν υπάρχει κατάλληλο σημείο διαχωρισμού. (στ) Το CC διαχωρίζεται στα κοινά σημεία του με το διαχωριστικό των γραμμών.

Τέλος, αν και οι δύο λόγοι είναι μεγαλύτεροι από το κατώφλι, τότε το CC θα πρέπει να «σπάσει» σε δύο τμήματα γιατί προφανώς εκτείνεται και στις δύο γραμμές (σχ. 1.16α, δ). Για το διαχωρισμό του CC χρησιμοποιείται η πιο λεπτή εκδοχή της εικόνας του CC [34, 35]. Τα πιθανά σημεία για το «σπάσιμο» του CC είναι τα pixels της λεπτής μορφής που έχουν περισσότερους από 2 γείτονες (δεδομένου τύπου γειτνίασης 8-η) και βρίσκονται κοντά στο διαχωριστικό της γραμμής κειμένου. Ξεκινώντας από το pixel που βρίσκεται κοντινότερα στο διαχωριστικό, το

εξεταζόμενο pixel και τα γειτονικά του σε σημεία του φόντου, ώστε η λεπτή εκδοχή του CC να χωριστεί σε 2 τουλάχιστον νέα τμήματα (σχ. 1.16β). Έπειτα, για κάθε τμήμα υπολογίζονται οι λόγοι r_k και εξετάζονται οι κανόνες που αναφέρθηκαν παραπάνω. Αν είναι δυνατή η ένταξη όλων των τμημάτων στις γραμμές κειμένου, τότε το σημείο επιλέγεται ως το καταλληλότερο (σχ. 1.16γ). Αλλιώς, η διαδικασία επαναλαμβάνεται για το επόμενο υποψήφιο σημείο διαχωρισμού. Στην περίπτωση που δεν επιτευχθεί ο εντοπισμός σημείου διαχωρισμού (σχ. 1.16ε), το εξεταζόμενο CC χωρίζεται απευθείας στη γραμμή (row) που βρίσκεται το διαχωριστικό (σχ. 1.16στ). Μια παρόμοια διαδικασία για το διαχωρισμό ενός CC, περιγράφεται στο [36].

1.2.7. Αξιολόγηση της προτεινόμενης τεχνικής

Η προτεινόμενη τεχνική (ILSP-LWSeg) υποβλήθηκε προς αξιολόγηση στους διαγωνισμούς κατάτμησης χειρόγραφου κειμένου (Handwriting Segmentation Contests) που διεξήχθησαν στα πλαίσια των International Conferences on Document Analysis and Recognition 2007 και 2009. Τα σετ εκμάθησης και εξέτασης του πρώτου διαγωνισμού αποτελούνταν από 20 και 80 δυαδικές εικόνες αντίστοιχα. Στο δεύτερο διαγωνισμό, ως σετ εκμάθησης χρησιμοποιήθηκαν οι 100 εικόνες του ICDAR07 και το σετ εξέτασης αποτελούταν 200 δυαδικές εικόνες χειρόγραφου κειμένου. Το συνολικό πλήθος των γραμμών κειμένου ανά σετ εξέτασης ήταν 1771 και 4034. Σύμφωνα με τους διοργανωτές του διαγωνισμού, τα δύο σετ περιέχουν εικόνες κειμένου, οι οποίες καλύπτουν ευρύ φάσμα των ιδιαιτεροτήτων που απαντώνται σε χειρόγραφα κείμενα. Τα κείμενα είτε προέρχονται από διαφορετικούς γραφείς που τους ζητήθηκε να γράψουν το ίδιο κείμενο, είτε είναι ιστορικά έγγραφα, είτε αντλήθηκαν από το διαδίκτυο. Τα κείμενα είναι γραμμένα σε διάφορες γλώσσες (π.χ. Αγγλικά, Γερμανικά, Γαλλικά και Ελληνικά) και δεν περιέχουν μη κειμενικά στοιχεία όπως γραμμές, σχήματα, σκίτσα κ.λπ. Το μέγεθός τους ποικίλει από 650x825 σε 2500x3500 pixels.

Η μέθοδος αξιολόγησης βασίζεται στη μέτρηση του πλήθους των ορθών αντιστοιχιών μεταξύ των αποτελεσμάτων του αλγορίθμου και των πραγματικών δεδομένων [37]. Έστω I το σύνολο που αντιστοιχεί στην εξεταζόμενη εικόνα, G_j το σύνολο για την j γραμμή κειμένου της ορθά επισημειωμένης εικόνας (ground truth) και R_i το σύνολο για την i γραμμή κειμένου της εικόνας που προέκυψε από τον προτεινόμενη τεχνική. Ο πίνακας αντιστοιχιών $MatchScore(i, j)$ συμπληρώνεται ως εξής:

$$MatchScore(i, j) = \frac{|G_j \cap R_i \cap I|}{|(G_j \cup R_i) \cap I|} \quad \text{(Εξ. 1.9)}$$

Κάθε στοιχείο του πίνακα με τιμή μεγαλύτερη ή ίση από 0.95 δηλώνει ότι υπάρχει ορθή αντιστοίχηση μεταξύ της εντοπισμένης και της επισημειωμένης γραμμής κειμένου. Το πλήθος των ορθών αντιστοιχιών δηλώνεται ως $o_g 2 o_d$, όπου ο δείκτης g χρησιμοποιείται για τις πραγματικές γραμμές κειμένου και ο δείκτης d για τις γραμμές που έχει εντοπίσει η

αξιολογούμενη μέθοδος. Κάθε άλλη τιμή δηλώνει είτε το ταίριασμα μιας γραμμής τύπου g με πολλές γραμμές τύπου d ($o_g 2m_d$), είτε πολλών g με μία τύπου d ($m_g 2o_d$), είτε μιας d με πολλές g ($o_d 2m_g$), είτε πολλών d με μια g ($m_d 2o_g$). Αν G και F είναι τα πλήθη των πραγματικών και των εντοπισμένων περιοχών αντίστοιχα, τότε οι βαθμοί εντοπισμού DR (detection rate) και ακρίβειας RA (recognition accuracy) και ο αρμονικός μέσος FM υπολογίζονται ως εξής:

$$DR = (w_1 \cdot o_g 2o_d + w_2 \cdot o_g 2m_d + w_3 \cdot m_g 2o_d) / G \quad \text{(Εξ. 1.10)}$$

$$RA = (w_1 \cdot o_d 2o_g + w_2 \cdot o_d 2m_g + w_3 \cdot m_d 2o_g) / F \quad \text{(Εξ. 1.11)}$$

$$FM = 2 \cdot DR \cdot RA / (DR + RA) \quad \text{(Εξ.1.12)}$$

όπου w_1, w_2 και w_3 είναι προκαθορισμένα βάρη ίσα με 1, 0.25 και 0.25. Ως σχόλιο για τον τρόπο αξιολόγησης αναφέρεται ότι η χρήση των βαρών w_2 και w_3 για τον υπολογισμό της απόδοσης των αλγορίθμων δεν έχει ιδιαίτερη προσφορά. Στην ενότητα 3.1.3 (πίνακας 1.6) παρουσιάζονται τα αποτελέσματα χωρίς τη συμμετοχή των βαρών αυτών. Άλλωστε, στην αξιολόγηση των αλγορίθμων για το διαγωνισμό του 2009, τα βάρη αυτά τέθηκαν ίσα με 0.

Τα αποτελέσματα της προτεινόμενης τεχνικής στα δύο σετ εξέτασης και τα συγκριτικά αποτελέσματα των αλγορίθμων που συμμετείχαν, παρουσιάζονται στους πίνακες 1.1, 1.2. και 1.3. Αναλυτικές πληροφορίες για την οργάνωση των διαγωνισμών καθώς και σύντομες περιγραφές των αλγορίθμων που συμμετείχαν, περιέχονται στα [7, 38].

Πίνακας 1.1. Αναλυτικά αποτελέσματα προτεινόμενης μεθόδου (ICDAR07) [7]

G	F	$o_g 2o_d$	$o_g 2m_d$	$m_g 2o_d$	$o_d 2m_g$	$m_d 2o_g$	DR (%)	RA (%)
1771	1773	1713	5	34	17	10	97.3	97.0

Πίνακας 1.2. Συγκριτικά αποτελέσματα (ICDAR07) [7]

	DR (%)	RA (%)	FM %
BESUS [39]	86.6	79.7	83.0
DUTH-ARLSA	73.9	70.2	72.0
ILSP-LWSeg [31]	***97.3***	***97.0***	***97.1***
PARC	92.2	93.0	92.6
UoA-HT [21]	95.5	95.4	95.4
PROJECTIONS	68.8	63.2	65.9

Πίνακας 1.3. Συγκριτικά αποτελέσματα (ICDAR09) [38]

	M	$o_g 2o_d$	DR (%)	RA (%)	FM (%)
CASIA-MSTSeg [40]	4049	3867	95.86	95.51	95.68
CMM	4044	3975	98.54	98.29	98.42
CUBS [43]	**4036**	**4016**	**99.55**	**99.50**	**99.53**
ETS	4033	3496	86.66	86.68	86.67
ILSP-LWSeg-09 [31]	***4043***	***4000***	***99.16***	***98.94***	***99.05***
Jadavpur Univ	4075	3541	87.78	86.90	87.34
LRDE [65]	4423	3901	96.70	88.20	92.25
PAIS	4031	3973	98.49	98.56	98.52
AegeanUniv [41]	4054	3130	77.59	77.21	77.40
PortoUniv [42]	4028	3811	94.47	94.61	94.54
PPSL	4084	3792	94.00	92.85	93.42
REGIM	4563	1629	40.38	35.70	37.90

Επιγραμματικά, αναφέρεται ότι έξι μέθοδοι (DUTH-ARLSA, Jadavpur Univ, CASIA-MSTSeg, CMM, PPSL και REGIM) εξετάζουν τις σχετικές θέσεις και τις διαστάσεις των CCs της εικόνας κειμένου και με την εφαρμογή κανόνων επιτυγχάνεται η ομαδοποίησή τους σε γραμμές κειμένου. Εξετάζοντας την απόδοσή τους και λαμβάνοντας υπόψη ότι για τις συγκεκριμένες τεχνικές απαιτείται η χρήση διαφόρων παραμέτρων (κατωφλίων) συμπεραίνει κανείς την ευαισθησία τους στις επιλεγμένες παραμέτρους. Οι τεχνικές PAIS, ILSP και AegeanUniv υιοθέτησαν τις επιμέρους προβολές και παρουσίασαν υψηλά αποτελέσματα, εκτός της AegeanUniv στην οποία το κείμενο χωρίζεται σε τρεις μόλις κατακόρυφες ζώνες και επομένως δεν εντοπίζονται σωστά γραμμές κειμένου με μεταβλητή κλίση. Η μεταβλητή κλίση των γραμμών είναι και ο λόγος που οι ολικές προβολές (PROJECTIONS) παρουσίασαν χαμηλά ποσοστά επιτυχίας. Οι μέθοδοι BESUS, ETS και PARC αρχικά εφαρμόζουν δυαδικούς μορφολογικούς μετασχηματισμούς (dilation, closing) για την ενοποίηση κοντινών pixels κειμένου και στη συνέχεια τα νέα CCs ομαδοποιούνται σε γραμμές κειμένου με βάση τις αποστάσεις τους (βλ. Εν. 1.3.3). Μία τεχνική (UoA-HT) εφαρμόζει τον αλγόριθμο Hough. Η μέθοδος CUBS, η οποία αξιολογήθηκε ως η πιο αποτελεσματική στο ICDAR09, είναι παρόμοια με αυτή που περιγράφεται στην ενότητα 1.1 (horizontal background fuzzy runlength). Η σημαντική διαφορά είναι ότι ο αλγόριθμος RLSA δεν εφαρμόζεται μόνο κατά τον κατακόρυφο άξονα αλλά σε πέντε επιλεγμένες διευθύνσεις (-20^o ,-10^o , 0^o, 10^o και 20^o) [43]. Η μέθοδος PortoUniv βασίζεται στο

διαχωρισμό γράφων. Τα στάδια επεξεργασίας που προτείνει η LRDE είναι: η συνέλιξη της εικόνας με ανισοτροπικό γκαουσιανό φίλτρο για την ανάδειξη των περιοχών κειμένου, η εφαρμογή του μορφολογικού τελεστή κλεισίματος για την ομογενοποίηση των περιοχών αυτών, η χρήση του μετασχηματισμού watershed για την κατάτμηση της εικόνας σε περιοχές που περιέχουν μεγάλα τμήματα κάθε γραμμής κειμένου και τέλος η ενοποίηση των τμημάτων που παρουσιάζουν σημαντική επικάλυψη κατά τον οριζόντιο άξονα. Πρόκειται για αποτελεσματικό και ταχύ αλγόριθμο και θα μπορούσε να είναι πιο αποδοτικός αν συμπεριληφθεί ένα στάδιο ελέγχου για τον εντοπισμό και το διαχωρισμό ενοποιημένων γραμμών, πριν την εφαρμογή του μετασχηματισμού watershed.

Μετά τη λήξη των συνεδρίων τα σετ εξέτασης ήταν διαθέσιμα στους συμμετέχοντες. Παρατηρώντας προσεκτικά τα αποτελέσματα σε κάθε εικόνα κειμένου, προκύπτει ότι η προτεινόμενη τεχνική αντιμετωπίζει επιτυχώς τη μεταβλητότητα στην κλίση του κειμένου και στο μέγεθος των χαρακτήρων, την ακανόνιστη μορφή των περιθωρίων καθώς και τους «ενωμένες» γραμμές κειμένου. Όμως, η αποτελεσματικότητά της μειώνεται στις περιπτώσεις που η πλειοψηφία των χαρακτήρων είναι τεμαχισμένοι σε πολλά τμήματα.

Σχήμα 1.17. Τμήμα της εικόνας 037.tif από ICDAR07. Το πάνω τμήμα του χαρακτήρα “f”, ως ανεξάρτητο CC, ανατίθεται στην μπλε γραμμή κειμένου και κάποια τμήματα του “n” στην μωβ.

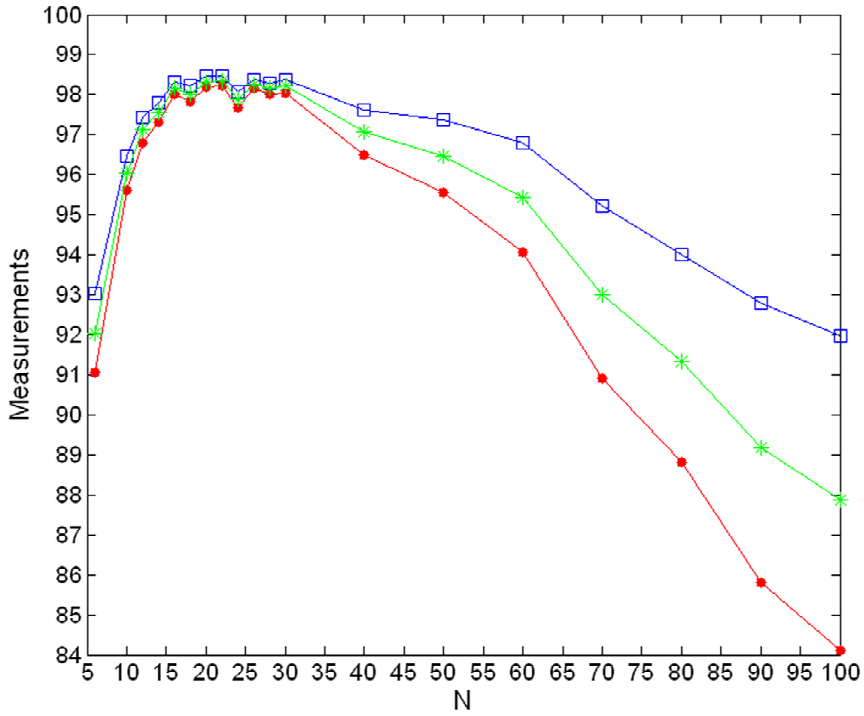
Η ύπαρξη «σπασμένων» χαρακτήρων οφείλεται κυρίως στην ποιότητα του οργάνου γραφής. Πράγματι, αν έχει χρησιμοποιηθεί ένα στυλό που εμφανίζει συχνές διακοπές στο μελάνι, από τη δυαδικοποίηση της ψηφιακής εικόνας κειμένου θα προκύψουν πολλά μικρά τμήματα που ανήκουν στον ίδιο χαρακτήρα. Η αποτυχία της προτεινόμενης μεθόδου οφείλεται στον τρόπο ανάθεσης των CCs σε γραμμές (σχ. 1.17) που περιγράφεται στην ενότητα 1.2.6.

Πέρα από την αποτελεσματικότητά της, η τεχνική εξετάστηκε στα σετ του ICDAR07 (100 εικόνες) και ως προς την ευστάθειά της κατά τη μεταβολή των παραμέτρων *N* και *R*. Στο πίνακα 1.4 παρουσιάζονται τα αποτελέσματα για διάφορες τιμές της παραμέτρου *N* (το πλήθος των κατακόρυφων ζωνών στις οποίες τεμαχίζεται αρχικά η εικόνα του χειρόγραφου κειμένου) και το σχ. 1.18 αναπαριστά την αντίστοιχη μεταβολή της απόδοσης του αλγορίθμου. Για τιμές μεταξύ 16 και 30, η μετρική *FM* ποικίλει από 97.88% σε 98.33%. Αντίθετα, για μικρότερες τιμές του *N*, η τεχνική γίνεται λιγότερο αποτελεσματική. Αυτό συμβαίνει γιατί οι μικρές τιμές του *N* αντιστοιχούν σε ζώνες μεγάλου πλάτους στις οποίες η κλίση του κειμένου δεν μπορεί να θεωρηθεί σταθερή.

Επομένως, για τέτοιες τιμές, η τεχνική δεν μπορεί να αντιμετωπίσει την μεταβλητή κλίση των γραμμών των χειρογράφων. Αντίστοιχα, για μεγάλες τιμές της ποσότητας *N*, το πλάτος των ζωνών είναι ιδιαίτερα μικρό και πολλές ζώνες δεν περιέχουν αρκετή πληροφορία για τον υπολογισμό αξιόπιστων στατιστικών τιμών.

Πίνακας 1.4. Αποτελέσματα αξιολόγησης για διάφορες τιμές της παραμέτρου *N*.

N	DR%	RA%	FM%	N	DR%	RA%	FM%	N	DR%	RA%	FM%
6	93.04	91.06	92.04	**20**	**98.46**	**98.2**	**98.33**	**50**	97.37	95.55	96.45
8	94.15	93.88	94.01	**22**	98.48	98.21	98.34	**60**	96.8	94.07	95.42
10	96.47	95.62	96.04	**24**	98.07	97.69	97.88	**70**	95.23	90.9	93.01
12	97.43	96.81	97.12	**26**	98.38	98.15	98.26	**80**	94	88.8	91.33
14	97.8	97.3	97.55	**28**	98.29	98	98.15	**90**	92.8	85.81	89.17
16	98.31	98.01	98.16	**30**	98.38	98.04	98.21	**100**	91.98	84.12	87.87
18	98.21	97.83	98.02	**40**	97.63	96.5	97.06				

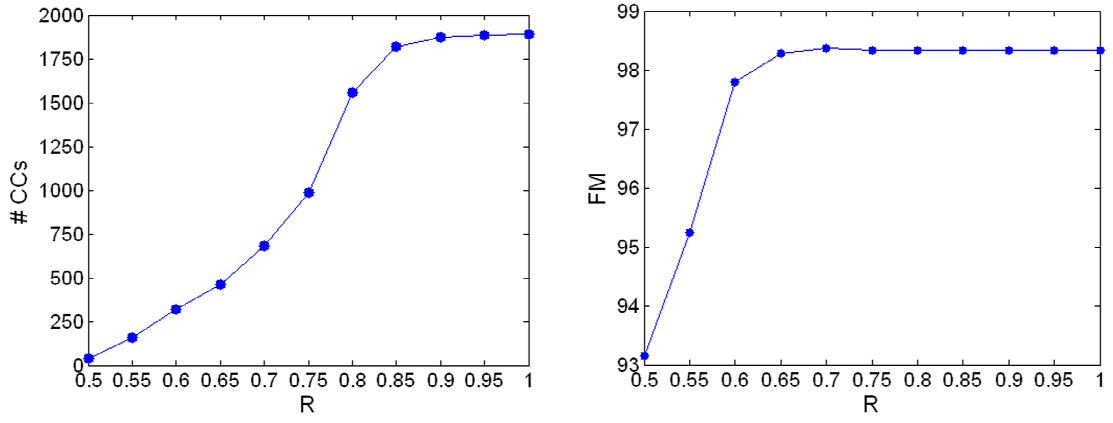


Σχήμα 1.18. Απόδοση του αλγορίθμου κατάτμησης χειρόγραφου σε γραμμές για διάφορες τιμές της *N*. Οι ποσότητες *DR*, *RA* και *FM* εμφανίζονται με μπλε, κόκκινο και πράσινο χρώμα αντίστοιχα.

Μια άλλη προκαθορισμένη ποσότητα που χρησιμοποιείται στην προτεινόμενη τεχνική είναι το κατώφλι *R* (βλέπε ενότητα 1.2.6). Η αποτελεσματικότητα της τεχνικής και το πλήθος των CCs που δεν εντάσσονται άμεσα σε γραμμές κειμένου, εξετάστηκαν για διάφορες τιμές του *R* (πίνακας 1.5). Όπως αναμενόταν, το πλήθος των CCs που δεν επαληθεύουν τον κανόνα ανάθεσης, αυξάνεται καθώς αυξάνεται το *R* (ο κανόνας γίνεται πιο αυστηρός), όπως φαίνεται στο σχ. 1.19α. Επομένως, καθώς αυξάνεται το *R,* περισσότερα CCs προωθούνται στα επόμενα στάδια επεξεργασίας για να ανατεθούν σε γραμμές κειμένου. Όμως, παρατηρώντας το σχ. 1.19β, προκύπτει το συμπέρασμα η αποτελεσματικότητα της τεχνικής παραμένει σταθερή. Σημειώνεται ότι για τιμές του *R* μικρότερες από 0.65, ο κανόνας είναι ιδιαίτερα χαλαρός και δεν ενδείκνυται για τα χειρόγραφα κείμενα που η πραγμάτωση χαρακτήρων με πολλές προεκτάσεις προς τα πάνω ή προς τα κάτω είναι πολύ συνήθης.

Πίνακας 1.5. Αποτελέσματα αξιολόγησης για διάφορες τιμές του *R.*

R	#CCs	FM %	R	#CCs	FM %	R	#CCs	FM %
0.5	-	93.48	0.7	684	98.61	0.9	1872	98.33
0.55	157	95.46	0.75	984	98.33	0.95	1886	98.33
0.6	318	97.84	0.8	1554	98.33	1	1892	98.33
0.65	462	98.34	0.85	1821	98.33			



(α) (β)

Σχήμα 1.19. α) Πλήθος CCs που δεν ανατέθηκαν άμεσα σε γραμμές κειμένου για διάφορες τιμές της παραμέτρου R. β) Μεταβολή της αποτελεσματικότητας (FM) για διάφορες τιμές της R.

1.2.8 Συμπεράσματα

Η διαφορετικότητα των γραφικών χαρακτήρων και των γλωσσών έχουν ως αποτέλεσμα τη μεγάλη ποικιλομορφία των χειρόγραφων κειμένων. Η προτεινόμενη μέθοδος είναι

ανεξάρτητη από τη γλώσσα γραφής και προσαρμόζεται στις ιδιαιτερότητες του εξεταζόμενου κειμένου. Επίσης, αντιμετωπίζει επιτυχώς τις κυριότερες δυσκολίες που παρουσιάζονται κατά την κατάτμηση του χειρόγραφου κειμένου σε γραμμές.

Συγκεκριμένα, υιοθετεί τη χρήση των επιμέρους προβολών ώστε να είναι δυνατός ο χειρισμός κειμένων με μεταβλητή κλίση μεταξύ των γραμμών αλλά και κατά μήκος της ίδιας γραμμής. Επομένως, υπερέχει των τεχνικών που βασίζονται στις ολικές προβολές, αφού αυτές προϋποθέτουν ότι η κλίση όλων των γραμμών κειμένου είναι σταθερή. Επίσης, υπερτερεί των μεθόδων που χρησιμοποιούν το μετασχηματισμό Hough, μια και αυτές προϋποθέτουν ότι η κλίση κατά μήκος μιας γραμμής είναι σχεδόν σταθερή.

Βέβαια, οι επιμέρους προβολές, όπως και οι τεχνικές διάχυσης, είναι ιδιαίτερα ευαίσθητες στη μεταβλητότητα του μεγέθους των χαρακτήρων στην ίδια γραμμή κειμένου και στην εμφάνιση μεγάλων κενών μεταξύ διαδοχικών λέξεων. Για την αντιμετώπιση αυτού του προβλήματος, υπολογίζονται οι ομαλοποιημένες προβολές λαμβάνοντας υπόψη και τα δεδομένα των γειτονικών ζωνών. Ο αρχικός υπερ-τεμαχισμός κάθε ζώνης σε τμήματα κειμένου και κενών με τη χρήση των ομαλοποιημένων προβολών και η εφαρμογή ενός πιθανοτικού μοντέλου που αναδεικνύει μόνο τα κύρια τμήματα των γραμμών κειμένου ανά ζώνη, έχουν ως αποτέλεσμα τη χάραξη των κατάλληλων διαχωριστικών των γραμμών κειμένου ανά ζώνη. Ο συνδυασμός των διαχωριστικών και η χρήση κατάλληλης μετρικής οδηγεί στην οριοθέτηση των γραμμών κειμένου. Τέλος, με τη χρήση απλών γεωμετρικών κανόνων τα CCs του κειμένου ανατίθενται πλήρως ή μερικώς στις αντίστοιχες γραμμές κειμένου.

Συμπερασματικά, η προτεινόμενη τεχνική είναι ιδιαίτερα αποτελεσματική για την εξέταση εικόνων χειρόγραφων κειμένων που περιέχουν μόνο κειμενικά στοιχεία. Επομένως, μπορεί να αποτελέσει βαθμίδα ενός συστήματος επεξεργασίας εικόνων χειρόγραφων κειμένου που περιλαμβάνει λειτουργίες όπως ο εντοπισμός των περιοχών με σκίτσα και η αναγνώριση χειρόγραφων χαρακτήρων με τελικό στόχο τη δημιουργία των αντίστοιχων ηλεκτρονικών κειμένων.

1.3. Προτεινόμενη μέθοδος με την εφαρμογή τελεστών δυαδικών εικόνων

Στη συγκεκριμένη ενότητα προτείνεται μια μέθοδος για την κατάτμηση δυαδικών εικόνων χειρόγραφου κειμένου σε γραμμές με τη χρήση μορφολογικών τελεστών. Η ανάλυση δυαδικών εικόνων έντυπων κειμένων με την εφαρμογή τελεστών μαθηματικής μορφολογίας έχει αποδειχθεί ιδιαίτερα αποτελεσματική. Πράγματι, στο [44] περιέχονται αρκετές αποτελεσματικές και ταχείες τεχνικές για ειδικές εφαρμογές, όπως ο εντοπισμός των πλάγιων (*italic*) και έντονων (*bold*) χαρακτήρων του κειμένου, των εικόνων ή άλλων μη κειμενικών στοιχείων και ο υπολογισμός της κλίσης του κειμένου, σε έντυπα. Το βασικό κίνητρο για την ανάπτυξη της προτεινόμενης μεθόδου είναι το γεγονός ότι παρά την αποδεδειγμένη ισχύ των τελεστών στην ανάλυση δυαδικών εικόνων έντυπων κειμένων, οι αντίστοιχες μέθοδοι που έχουν προταθεί για τα χειρόγραφα κείμενα δεν έχουν την ίδια απόδοση.

1.3.1. Βασικές έννοιες

Η μαθηματική μορφολογία προτάθηκε από τους Marathon και Serra στα μέσα της δεκαετίας του 1960 και αποτελεί ένα ισχυρό μέσο για την επεξεργασία ψηφιακών εικόνων με πολλές εφαρμογές [45], χωρίς βέβαια να περιορίζεται μόνο στην ανάλυση σημάτων δύο διαστάσεων. Ειδικότερα, για την επεξεργασία δυαδικών εικόνων, όπως οι εικόνες κειμένου που εξετάζονται στη συγκεκριμένη εργασία, έχει καθιερωθεί ο όρος της δυαδικής μορφολογίας. Ο βασικός φορμαλισμός είναι η αναπαράσταση των εικόνων ως σύνολα και επομένως και οι «μετασχηματισμοί» των εικόνων ως πράξεις μεταξύ συνόλων. Μια προφανής αναπαράσταση μιας ψηφιακής δυαδικής εικόνας μπορεί να γίνει με τη συνάρτηση $f(\mathbf{x}): A \to \{0,1\}, \mathbf{x} \in A \subseteq \mathbb{Z}^2$. Θεωρώντας ότι τα εικονοστοιχεία με τιμή 1 συνιστούν τα αντικείμενα-σχήματα της εικόνας, τότε αυτή μπορεί να περιγραφεί από το σύνολο $F = \{\mathbf{x} \in A : f(\mathbf{x}) = 1\}$ [5]. Κατ' αναλογία, μια δυαδική εικόνα κειμένου είναι το σύνολο των pixels του κειμένου (foreground), ενώ τα υπόλοιπα pixels αντιστοιχούν στο φόντο (background) και συνιστούν το συμπλήρωμα F^c του F.

Έστω, $X \subseteq \mathbb{Z}^2$ το σύνολο των pixels με τιμή 1 μιας δυαδικής εικόνας. Έστω επίσης, το σύνολο $B \subseteq \mathbb{Z}^2$. Το σύνολο B συνήθως ονομάζεται δομικό στοιχείο (structuring element, SE) και στις εφαρμογές έχει μικρότερο μέγεθος και απλούστερο σχήμα από το X. Ακολούθως ορίζονται οι τελεστές δυαδικών εικόνων που εφαρμόζονται στην προτεινόμενη μέθοδο, όπως προτείνονται στο [46]. Μια αναλυτική παρουσίαση των ορισμών τους, όπως έχουν προταθεί στην αντίστοιχη βιβλιογραφία περιλαμβάνεται στο [47].

Ο τελεστής **erosion** (συστολή) του X από το B ορίζεται ως:

$$X \ominus B = \{z : B + z \subseteq X\} = \bigcap_{z \in B} X - z \qquad \text{(Εξ. 1.13)}$$

όπου $X - z = \{x - z : x \in X\}$ η μετατόπιση του συνόλου X κατά το διάνυσμα $-z$. Από την (Εξ. 1.13) προκύπτει ότι το αποτέλεσμα της συστολής είναι το σύνολο των σημείων z που όταν αποτελούν το κέντρο του B (δηλαδή όταν το B έχει μετατοπιστεί κατά z), τότε το B περιέχεται στο X. Επομένως, ο συγκεκριμένος τελεστής μπορεί να χρησιμοποιηθεί για τον εντοπισμό του σχήματος του δομικού στοιχείου στη δυαδική εικόνα.

Ο τελεστής **dilation** (διαστολή) ορίζεται ως:

$$X \oplus B = \{z : \overset{\vee}{B} + z \cap X \neq \varnothing\} = \bigcup_{z \in B} X + z \qquad \text{(Εξ. 1.14)}$$

όπου $\overset{\vee}{B} = \{-b : b \in B\}$ το συμμετρικό, ως προς το σημείο αναφοράς (origin), σύνολο του B. Από την (Εξ. 1.14) προκύπτει ότι το αποτέλεσμα της διαστολής είναι το σύνολο των σημείων z που όταν αποτελούν το κέντρο του $\overset{\vee}{B}$, τότε το $\overset{\vee}{B}$ και το X έχουν ένα τουλάχιστον κοινό

στοιχείο. Επομένως, ο συγκεκριμένος τελεστής μπορεί να χρησιμοποιηθεί για την «επέκταση» των αντικειμένων-σχημάτων της εικόνας σύμφωνα με το σχήμα του δομικού στοιχείου.
Αν οι τελεστές erosion και dilation εφαρμοστούν διαδοχικά με το ίδιο SE, τότε προκύπτει ο τελεστής **opening** που ορίζεται ως:

$$X \odot B = (X \ominus B) \oplus B \qquad \text{(Εξ. 1.15)}$$

Αν τώρα εφαρμοστούν με την αντίστροφη σειρά προκύπτει ο τελεστής **closing** που ορίζεται ως:

$$X \bullet B = (X \oplus B) \ominus B \qquad \text{(Εξ. 1.16)}$$

Έστω δύο σύνολα B_1, B_2 ξένα μεταξύ τους ($B_1 \cap B_2 = \varnothing$). Ο τελεστής **hit-miss** ορίζεται ως:

$$X \otimes (B_1, B_2) = (X \ominus B_1) \cap (X^c \ominus B_2) \qquad \text{(Εξ. 1.17)}$$

Επομένως, ο πρώτος τελεστής συστολής εντοπίζει τις πραγματώσεις του σχήματος του B_1 από τα pixels της δυαδικής εικόνας με τιμή 1 (και το B_1 ονομάζεται “hit” SE), ενώ ο δεύτερος αναδεικνύει τις θέσεις εμφάνισης του σχήματος του B_2 στο φόντο της εικόνας (και το B_2 ονομάζεται “miss” SE). Τα δομικά στοιχεία B_1, B_2 έχουν το ίδιο σημείο αναφοράς και συνήθως χρησιμοποιείται ένα δομικό στοιχείο, έστω $\mathbf{B}$, στο οποίο φυσικά δηλώνονται τα hit και miss pixels, για να τα αναπαραστήσει. Το αποτέλεσμα του τελεστή είναι μια νέα δυαδική εικόνα με τιμή 1 στα pixels «γύρω» από τα οποία εμφανίζεται το πρότυπο που ορίζει το $\mathbf{B}$.
Έστω ένα δομικό στοιχείο W. Το **binary r-th rank order filter** (δυαδικό φίλτρο διάταξης τάξης r), $r = 1, 2, ..., |W|$, του X από το W ορίζεται (σαν τελεστής συνόλων) ως:

$$X \boxdot_r W = \{z : | X \cap (W + z) | \geq r\} \qquad \text{(Εξ. 1.18)}$$

δηλώνοντας ότι για το δυαδικό φίλτρο διάταξης δεν είναι απαραίτητη η διάταξη των τιμών των pixels κατά αύξουσα σειρά και η επιλογή της τιμής στην r-ιοστή θέση, αλλά αρκεί η απλή καταμέτρηση των κοινών στοιχείων και η σύγκριση του πλήθους τους με το κατώφλι (τάξη) r [48]. Επομένως, το δυαδικό φίλτρο διάταξης μπορεί να χαρακτηριστεί ως μια διαδικασία εντοπισμού του σχήματος W στην εικόνα X με βαθμό «χαλαρότητας» r. Αν $r = |W|$ τότε αναζητούμε το ακριβές ταίριασμα του W στο X και προφανώς η διαδικασία είναι ισοδύναμη με αυτή του τελεστή συστολής. Στην ειδική περίπτωση που $|W|$ είναι περιττός αριθμός και $r = (|W|+1)/2$ η (Εξ. 1.18) γράφεται $med_W(X)$, δηλώνοντας (για τις δυαδικές εικόνες) ότι το αποτέλεσμα στη θέση z θα είναι μονάδα αν το πλήθος των κοινών στοιχειών των X και $W + z$ πλειοψηφεί.

Στο [46] προτείνεται ένας μετασχηματισμός που συνδυάζει αυτούς που ορίζονται στις (Εξ. 1.17) και (Εξ. 1.18). Πρόκειται για τον (p,q) **-th rank hit-miss** που ορίζεται ως:

$$X \otimes_{p,q} (B_1, B_2) = (X \boxdot_p B_1) \cap (X^c \boxdot_q B_2) \qquad \text{(Εξ. 1.19)}$$

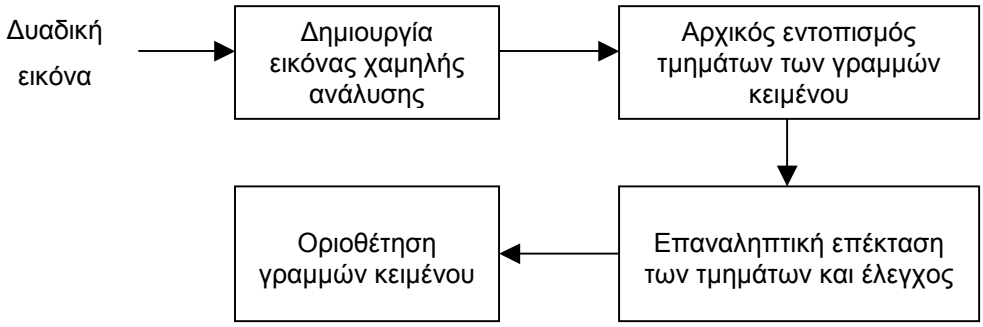
όπου $p = 1,2,...|B_1|$ και $q = 1,2,...|B_2|$. Επομένως, το ταίριασμα του σχήματος B_1 στην εικόνα X και του B_2 στο φόντο της (X^c) ελέγχεται από τα κατώφλια p και q αντίστοιχα. Το βασικό πλεονέκτημα αυτού του μετασχηματισμού είναι ότι επιτρέπει τον εντοπισμό και των δύο σχημάτων με μικρές διαφοροποιήσεις από αυτά των δομικών στοιχείων.

Προφανώς, η ανακατασκευή του προτύπου που ορίζουν τα B_1, B_2 μπορεί να επιτευχθεί με την εφαρμογή του τελεστή διαστολής με το B_1. Έτσι ορίζεται ο ακόλουθος μετασχηματισμός (p,q) **-th generalized foreground rank opening** ως γενίκευση του τελεστή ανοίγματος (opening) ως εξής:

$$\Psi_{p,q}(X; B_1, B_2) = [(X \boxdot_p B_1) \cap (X^c \boxdot_q B_2)] \oplus B_1 \qquad \text{(Εξ. 1.20)}$$

1.3.2. Προτεινόμενη μέθοδος

Τα στάδια επεξεργασίας της προτεινόμενης μεθόδου παρουσιάζονται στο ακόλουθο διάγραμμα:

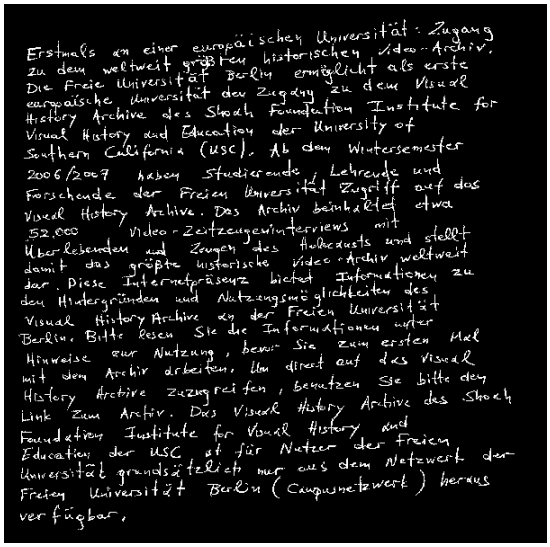


Σχήμα 1.20. Τα στάδια της προτεινόμενης μεθόδου για την κατάτμηση χειρόγραφου κειμένου σε γραμμές, με τη τελεστών δυαδικών εικόνων.

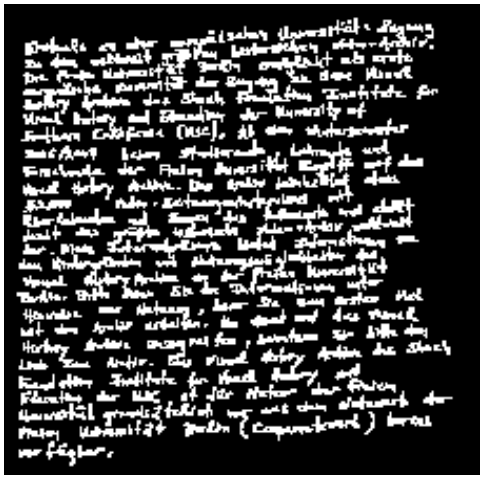
α) Δημιουργία εικόνας χαμηλής ανάλυσης

Οι ψηφιακές δυαδικές εικόνες κειμένου περιέχουν μεγάλο πλήθος εικονοστοιχείων αφού ψηφιοποιούνται συνήθως σε ανάλυση 300ppi, ώστε να είναι κατάλληλες για επεξεργασία από τα συστήματα οπτικής αναγνώρισης χαρακτήρων. Όμως, η ανάδειξη περιοχών με ιδιαίτερη υφή, όπως οι γραμμές κειμένου ή οι λέξεις, μπορεί να επιτευχθεί σε μικρότερες αναλύσεις. Υιοθετώντας την πρόταση που περιγράφεται στο [49], θα δημιουργηθεί μια εικόνα μικρότερης ανάλυσης (40ppi). Η άμεση υποδειγματοληψία (downsampling) της αρχικής εικόνας με παράγοντα 8 στην οριζόντια και στην κατακόρυφη κατεύθυνση, είναι πιθανό να προκαλέσει την εξαφάνιση των CCs της εικόνας που έχουν μικρή διάσταση και ίσως την αλλοίωση των

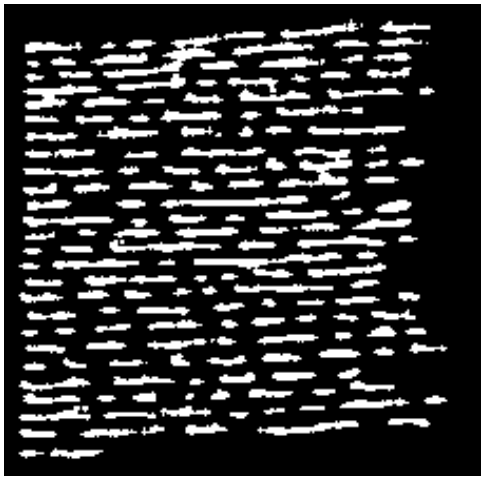
ιδιοτήτων υφής κάποιων περιοχών της αρχικής εικόνας. Στο [50] αποδεικνύεται ότι αν η υποδειγματοληψία πρόκειται να γίνει κρατώντας ένα pixel ανά $N \times N$ pixels της αρχικής εικόνας, τότε η εφαρμογή των μορφολογικών τελεστών διαστολή, ή συστολή, ή συνδυασμών τους με συμπαγή SE (όλα τα στοιχεία τους έχουν τιμή 1) διάστασης ίσης με $1 \times N$ και δειγματοληψία $\downarrow N$ ανά άξονα, θα οδηγήσει στη δημιουργία εικόνων χαμηλής ανάλυσης (χωρίς αναδίπλωση-aliasing στις δύο κατευθύνσεις) που σε κάθε μια θα αναδεικνύονται περιοχές με συγκεκριμένη υφή. Πράγματι, εφαρμόζοντας διαστολή με συμπαγές δομικό στοιχείο 1×8, υποδειγματοληψία με παράγοντα 8 κατά την οριζόντια διεύθυνση, διαστολή με το συμπαγές δομικό στοιχείο 8×1 και υποδειγματοληψία $\downarrow 8$ κατά τον κατακόρυφο άξονα, προκύπτει η εικόνα χαμηλής ανάλυσης (σχ. 1.21β), όπου οι περιοχές με κοντινά pixels κειμένου (π.χ. λέξεις), έχουν συγκροτήσει συνεκτικά αντικείμενα.



(α)



(β)



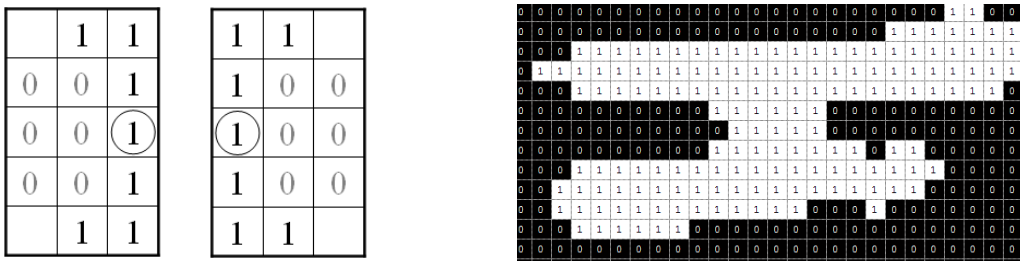
(γ)

Σχήμα 1.21. (α) Η αρχική εικόνα (028.tif ICDAR-07). (β) Η εικόνα χαμηλής ανάλυσης (X). γ) Η εικόνα $Y = med_W(X)$.

β) Αρχικός εντοπισμός τμημάτων των γραμμών κειμένου

Ένα από τα βασικά χαρακτηριστικά των χειρογράφων κειμένων είναι η μεταβλητότητα των αποστάσεων μεταξύ των γραμμών κειμένου και η ύπαρξη γραφημάτων που εκτείνονται και σε γειτονικές γραμμές. Το γεγονός αυτό εξηγεί την εμφάνιση αντικειμένων στην εικόνα χαμηλής ανάλυσης που επίσης εκτείνονται σε γειτονικές γραμμές. Έχοντας κατά νου ότι οι γραμμές κειμένου είναι επί τους ουσίας περιοχές με πληθώρα εικονοστοιχείων κειμένου (κυρίως κατά τον οριζόντιο άξονα) και με δεδομένη την προβολή των ιδιοτήτων της αρχικής εικόνας στην εικόνα χαμηλής ανάλυσης, συμπεραίνει κανείς πως μια εικόνα που θα δήλωνε σαφέστερα τα βασικά τμήματα των γραμμών κειμένου θα ήταν αυτή στην οποία δε θα εμφανίζονταν τα τμήματα που αντιστοιχούν στις προεκτάσεις (ascenders και descenders) των χαρακτήρων. Πειραματικά προσδιορίστηκε ότι η εφαρμογή του δυαδικού φίλτρου διάταξης $med_W(X)$, όπου X η εικόνα χαμηλής ανάλυσης και W το δομικό στοιχείο μεγέθους *3x7* συμβάλει στην απομάκρυνση αυτών των τμημάτων και στην ενοποίηση των κοντινών και οριζόντια διατεταγμένων περιοχών (σχ. 1.21γ). Η επιλογή του μεγέθους του W εξηγείται από το γεγονός ότι αφενός θα πρέπει να είναι μικρού ύψους ώστε να εφαρμόζεται κάθε φορά σε περιοχές εντός γραμμής κειμένου ή μεταξύ γραμμών κειμένου και αφετέρου να είναι μεγαλύτερου πλάτους ώστε να «ακολουθεί» τον οριζόντιο προσανατολισμό των γραμμών κειμένου.

Στη νέα εικόνα, έστω $Y = med_W(X)$, έχουν αναδειχθεί τα βασικά τμήματα των γραμμών κειμένου, αλλά παραμένουν κάποια CCs που εκτείνονται σε περισσότερες γραμμές κειμένου. Παρατηρώντας τα σχήματα που δημιουργούνται σε τέτοιες περιπτώσεις, μπορεί κανείς να συμπεράνει ότι τα στενά τμήματά τους είναι αυτά που αντιστοιχούν στην τοπική επικάλυψη των χαρακτήρων γειτονικών γραμμών. Η απομάκρυνση αυτών των τμημάτων θα μπορούσε να επιτευχθεί με τον τελεστή opening με ένα κατάλληλα επιλεγμένο δομικό στοιχείο. Όμως τότε θα υπήρχε ο κίνδυνος για την ταυτόχρονη απομάκρυνση και άλλων αντικειμένων της εικόνας με μέγεθος μικρότερο του SE.

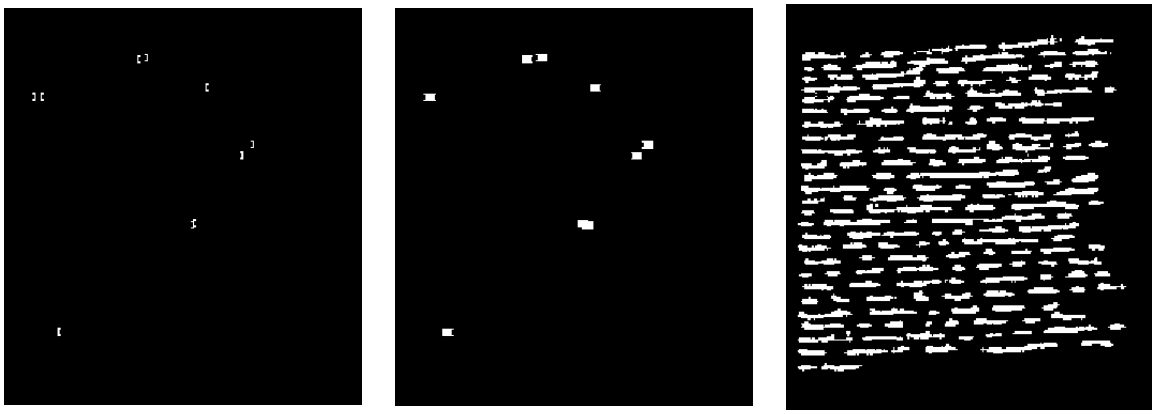


(α) (β)

Σχήμα 1.22. (α) Τα δομικά στοιχεία χρησιμοποιούνται για τον εντοπισμό των προτύπων που εμφανίζονται στις περιοχές επικάλυψης των γραμμών κειμένου. (β) Το στενό τμήμα του CC αντιστοιχεί στην περιοχή επικάλυψης.

Μια πιο ασφαλής επιλογή είναι η εφαρμογή των μετασχηματισμών $\Psi_{7,3}(Y;B_1,B_2)$ και $\Psi_{7,3}(Y;B_3,B_4)$ για τον εντοπισμό αυτών των σχημάτων, όπου τα B_1 και B_3 (B_2 και B_4) περιλαμβάνουν τα hits (miss) των SE του σχήματος 1.22α. Σημειώνεται ότι το πρώτο τμήμα του μετασχηματισμού που αφορά στα pixels κειμένου είναι συστολή (κάθε δομικό στοιχείο έχει 7 hits) και το δεύτερο είναι δυαδικό φίλτρο διάταξης τάξης 3, ώστε να εμφανίζει ανοχή στην ύπαρξη πρόσθετων pixels κειμένου (σχ. 1.22β).

Έστω $Z_1 = \Psi_{7,3}(Y;B_1,B_2)$ και $Z_2 = \Psi_{7,3}(Y;B_3,B_4)$ οι εικόνες που περιέχουν τα πρότυπα-εντοπισμένα σχήματα (σχ. 1.23α). Αν επεκτείνουμε κατάλληλα τα σχήματα των Z_1 και Z_2 προς τα δεξιά και αριστερά αντίστοιχα (σχ. 1.23β), τότε η διαφορά της ένωσής τους από την εικόνα Y, θα είναι μια νέα εικόνα, έστω I, η οποία δε θα περιέχει τα τμήματα που αντιστοιχούν στην επικάλυψη των χαρακτήρων (σχ. 1.23γ). Επομένως, $I = Y \setminus [(Z_1 \oplus B_5) \cup (Z_2 \oplus B_6)]$, όπου τα $B_5 = [0\ 0\ 0\ 0\ 1\ 1\ 1\ 1\ 1\ 1]$ και $B_6 = \overset{\vee}{B}_5$ ορίζουν την κατεύθυνση και το μέγεθος της επέκτασης.



(α) (β) (γ)

Σχήμα 1.23. (α) $Z_1 \cup Z_2$. (β) $(Z_1 \oplus B_5) \cup (Z_2 \oplus B_6)$. (γ) $I = Y \setminus [(Z_1 \oplus B_5) \cup (Z_2 \oplus B_6)]$

γ) Σχηματισμός των γραμμών κειμένου

Η εικόνα I περιέχει τα σημαντικά τμήματα (π.χ. λέξεις) των γραμμών κειμένου. Υποθέτοντας ότι για κάθε γραμμή κειμένου δύο τέτοια διαδοχικά τμήματα επικαλύπτονται κατά τον οριζόντιο άξονα, είναι δυνατή η ενοποίησή τους (ώστε να σχηματιστεί το σχήμα της γραμμής κειμένου) με την χρήση του τελεστή διαστολής και ένα απλό SE με σχήμα οριζόντιας γραμμής. Η επιλογή ενός τέτοιου SE μεγάλης διάστασης και η άμεση εφαρμογή της διαστολής πιθανότατα θα είχε ως αποτέλεσμα την ένωση αντικειμένων που αντιστοιχούν σε διαφορετικές γραμμές κειμένου μια και συνήθως αυτές παρουσιάζουν μεταβαλλόμενη κλίση. Επομένως, η διαδικασία αυτή θα πρέπει να γίνει σταδιακά ώστε να ελέγχονται κάθε φορά τα σχήματα που προκύπτουν. Κάθε βήμα της επαναληπτικής διαδικασίας περιλαμβάνει την επέκταση των αντικειμένων, τον

εντοπισμό των προτύπων που αντιστοιχούν σε ανεπιθύμητες ενώσεις τμημάτων και την απομάκρυνσή τους. Ακολούθως παρουσιάζεται η επαναληπτική διαδικασία:

$I_1 = I$

$\text{for } j = 1 : rep$

$K = I_j \oplus B_7$

$J_1 = (K \otimes_{3,4} \mathbf{B}_8) \oplus B_9$

$J_2 = (K \otimes_{3,4} \mathbf{B}_{10}) \oplus B_{11}$

$I_{j+1} = K \setminus (J_1 \cup J_2)$

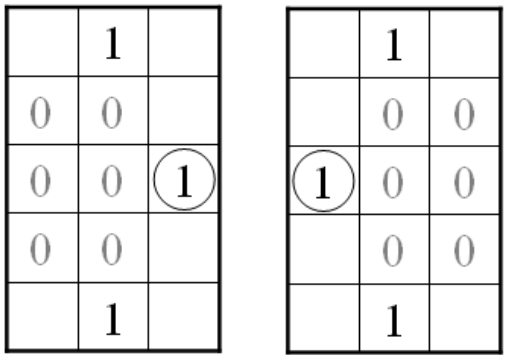
end

Τα δομικά στοιχεία που χρησιμοποιούνται έχουν επιλεγεί ως εξής:

$B_7 = [1\ 1\ 1\ 1\ 1]$ το SE για τη διαστολή,

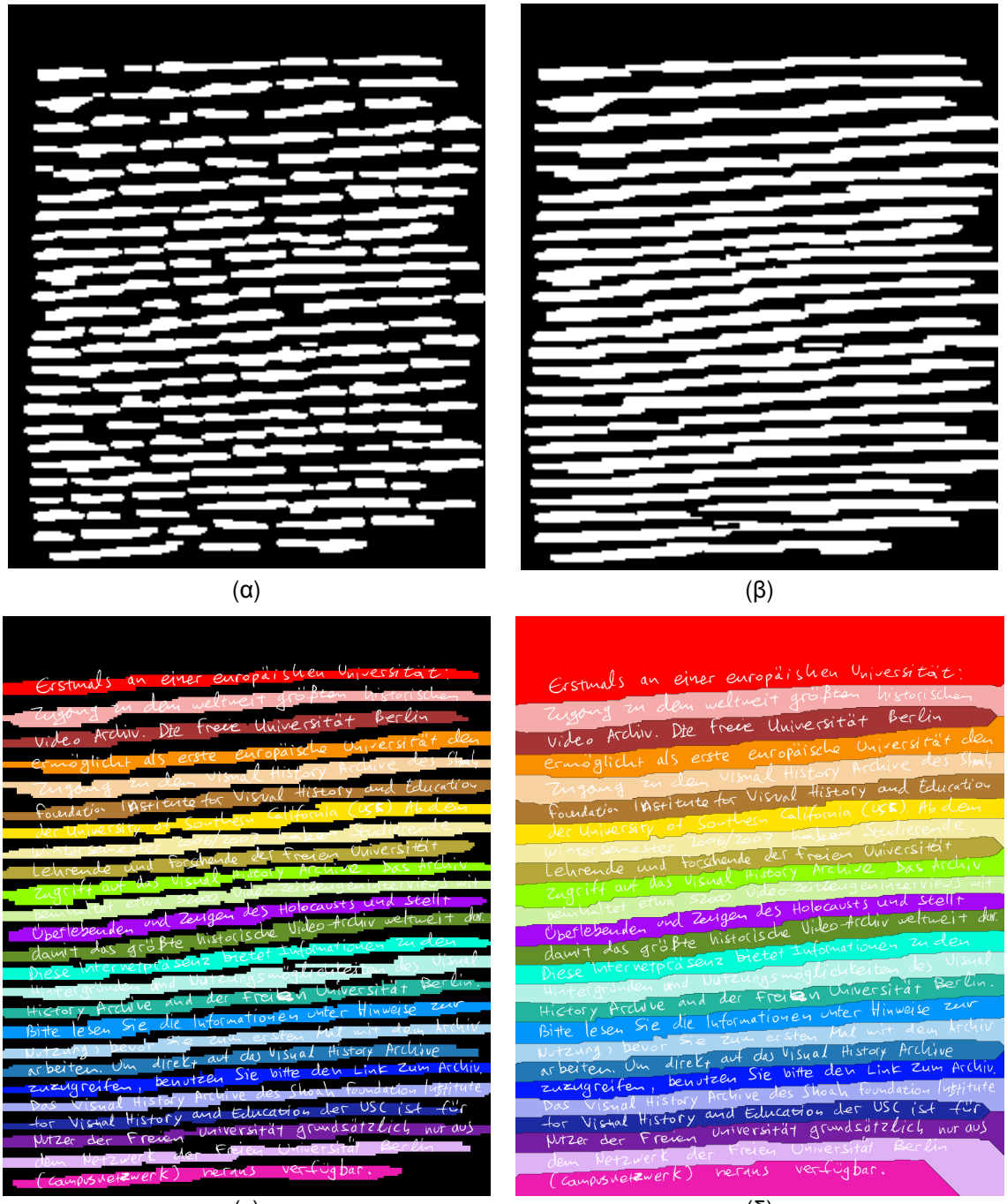
$\mathbf{B}_8$ και $\mathbf{B}_{10}$ τα ζεύγη των SE (hit και miss) του σχ. 1.24. Τα πρότυπα που εντοπίζονται με αυτά τα SE δηλώνουν ότι είτε δύο διαδοχικές γραμμές κειμένου τείνουν να ενωθούν, είτε έχουν ήδη ενωθεί.

$B_9 = \begin{bmatrix} 0\ 0\ 0\ 0\ 1\ 1\ 1\ 1\ 1 \\ 0\ 0\ 0\ 0\ 1\ 1\ 1\ 1\ 1 \\ 0\ 0\ 0\ 0\ 1\ 1\ 1\ 1\ 1 \end{bmatrix}$ και $B_{11} = \overset{\vee}{B}_9$ δηλώνουν την επέκταση των προτύπων.



Σχήμα 1.24. Τα ζεύγη των SE που χρησιμοποιούνται για τον εντοπισμό προτύπων και δηλώνουν ότι δύο γραμμές κειμένου τείνουν να ενωθούν ή έχουν ενωθεί.

Αν και ο αριθμός των επαναλήψεων που απαιτείται για το σχηματισμό των αντικειμένων που καλύπτουν τις περιοχές των γραμμών κειμένου ποικίλει από εικόνα σε εικόνα, προσδιορίστηκε πειραματικά ότι *rep=10* επαναλήψεις είναι αρκετές. Στα σχήματα (1.25α-β) παρουσιάζονται δύο φάσεις από την εξελικτική διαδικασία για την κατάτμηση της εικόνας 050.tif (ICDAR07).



Σχήμα 1.25. Κατάτμηση της εικόνας κειμένου σε γραμμές (050.tif – ICDAR07). (α) I_4. (β) I_7. (γ) Η εικόνα F. Τα CCs έχουν χρωματιστεί για να παρέχεται μεγαλύτερη ευκρίνεια. Επίσης, έχουν συμπεριληφθεί και τα pixels του κειμένου για να υπάρχει άμεση αξιολόγηση του αποτελέσματος. (δ) Η κατάτμηση της εικόνας με την εφαρμογή του αλγόριθμου watershed.

Κάθε CC της εικόνας χαμηλής ανάλυσης I_{10}, δηλώνει την ύπαρξη μιας γραμμής κειμένου στην αρχική εικόνα κειμένου A. Αν F, η εικόνα που προκύπτει από την «επέκταση»

της I_{10} στην αρχική ανάλυση, τότε κάθε CC που περιέχεται σε αυτή «καλύπτει» και την αντίστοιχη γραμμή κειμένου (σχ. 1.25γ).

Το τελευταίο στάδιο επεξεργασίας περιλαμβάνει τα ακόλουθα βήματα:

i) Την εξαγωγή των CCs που περιέχει η εικόνα F [6].

ii) Την απομάκρυνση των CCs που περιέχουν ελάχιστη κειμενική πληροφορία. Έστω $C \subseteq F$ το σύνολο που αντιπροσωπεύει ένα αντικείμενο της F και A η αρχική εικόνα κειμένου. Αν $|C \cap A| < thr$, τότε το αντικείμενο απορρίπτεται ως περιττό. Σημειώνεται ότι η τιμή του κατωφλίου thr επιλέχθηκε ίση με τριπλάσιο της ενδιάμεσης τιμής του πλήθους των pixels των αντικειμένων της αρχικής εικόνας, υποθέτοντας ότι κάθε γραμμή κειμένου περιέχει τουλάχιστον τρία αντικείμενα (π.χ. χαρακτήρες).

iii) Εφαρμογή του μετασχηματισμού watershed [51] στο συμπλήρωμα της εικόνας για τη χάραξη των διαχωριστικών μεταξύ των γραμμών κειμένου (σχ. 1.25δ).

iv) Δεικτοδότηση των pixels κειμένου. Έστω R_i μία από τις περιοχές που οριοθετήθηκαν στο προηγούμενο βήμα. Τότε η αντίστοιχη γραμμή κειμένου T_i προκύπτει ως $T_i = R_i \cap A$.

1.3.3. Αξιολόγηση

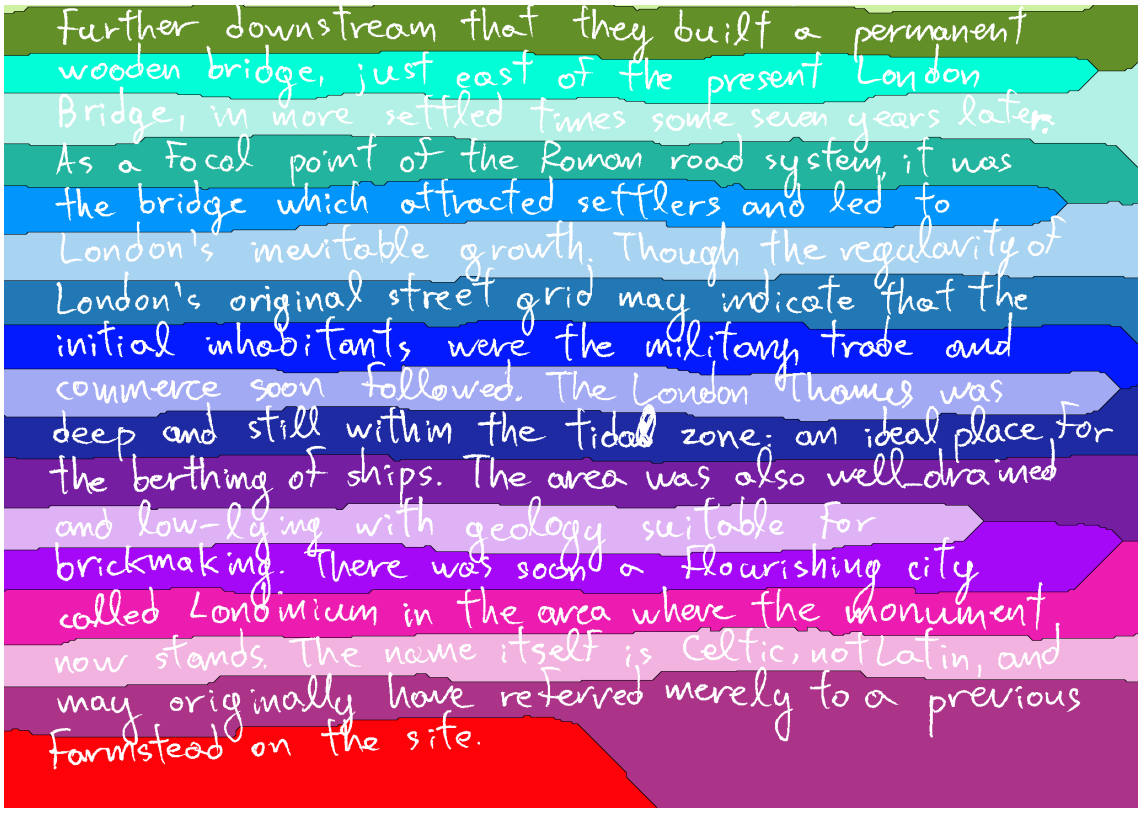
Η αξιολόγηση της προτεινόμενης μεθόδου έγινε στις 80 εικόνες (1771 γραμμές κειμένου) του σετ εξέτασης του ICDAR07 και τα αποτελέσματα παρουσιάζονται στον πίνακα 1.6. Στον ίδιο πίνακα συμπεριλαμβάνονται και τα αποτελέσματα των αλγορίθμων που συμμετείχαν στο διαγωνισμό. Σημειώνεται όμως, πως ο υπολογισμός της απόδοσης των αλγορίθμων έγινε λαμβάνοντας υπόψη μόνο τις 1-1 αντιστοιχίες των εντοπισμένων γραμμών κειμένου με τις επισημειωμένες.

Πίνακας 1.6. Συγκριτικά αποτελέσματα (ICDAR07)

	M	$o_g 2o_d$	DR (%)	RA (%)	FM (%)
BESUS [39]	1904	1494	84.36	78.47	81.31
DUTH-ARLSA	1894	1214	68.55	64.10	66.25
ILSP-LWSeg [31]	1773	1713	96.73	96.62	96.67
PARC	1756	1604	90.57	91.34	90.95
UoA-HT [21]	1770	1674	94.52	94.58	94.55
Προτεινόμενη μέθοδος	1783	1660	93.73	93.10	93.42

Από την παρατήρηση των αποτελεσμάτων για κάθε εξεταζόμενη εικόνα, προέκυψε το συμπέρασμα ότι η βασική αδυναμία της μεθόδου έγκειται στον τρόπο ανάθεσης των CCs στην αντίστοιχη γραμμή κειμένου. Ένα χαρακτηριστικό παράδειγμα παρουσιάζεται στο σχ. 1.26,

όπου εμφανίζεται ένα τμήμα της εικόνας 074.tif, μετά την κατάτμησή της. Πολλοί χαρακτήρες εκτείνονται σε περισσότερες από μια περιοχές και επομένως θα «σπάσουν» και τα τμήματά τους θα ανατεθούν σε αυτές. Άρα, κατά την αξιολόγηση της μεθόδου με βάση το πλήθος των pixels που έχουν ανατεθεί στη σωστή γραμμή, θα προκύψει πως κάποιες γραμμές δεν έχουν οριοθετηθεί με την επιθυμητή ακρίβεια. Η ενσωμάτωση μιας διαδικασίας σαν αυτή που περιγράφεται στην ενότητα 1.2.6, θα μπορούσε να ελέγξει το «σπάσιμο» των χαρακτήρων και να συμβάλει στη βελτίωση της προτεινόμενης τεχνικής.



Σχήμα 1.26. Αποτέλεσμα κατάτμησης της εικόνας 074.tif

Κεφάλαιο 2. Κατάτμηση χειρόγραφου κειμένου σε λέξεις

Ο διαχωρισμός κειμένου σε λέξεις είναι η κατάτμηση της εικόνας κειμένου σε μικρότερες εικόνες που καθεμία από αυτές περιέχει μια λέξη. Οι νέες αυτές εικόνες αποτελούν την είσοδο των συστημάτων αναγνώρισης είτε για την άμεση αναγνώριση των λέξεων, είτε για την περαιτέρω επεξεργασία τους (π.χ. διαχωρισμός σε χαρακτήρες), είτε για τη δεικτοδότηση του κειμένου εφόσον κάποιες εντοπισμένες λέξεις έχουν κρίσιμο ρόλο στο κείμενο. Ενδεικτική είναι η εργασία [52], στην οποία περιγράφεται ένας ταχύς και αποτελεσματικός τρόπος για τον εντοπισμό των λέξεων που είναι γραμμένες με πλάγια γραφή. Η διαδικασία περιλαμβάνει αφενός την κατάτμηση του κειμένου σε λέξεις και αφετέρου τον εντοπισμό ειδικών προτύπων-σχημάτων που χαρακτηρίζουν την πλάγια γραφή. Χρησιμοποιώντας ως αρχική εικόνα την εικόνα των προτύπων (seeds) και ως τελική την εικόνα των λέξεων (mask), αναδεικνύονται μόνο οι πλάγιες λέξεις.

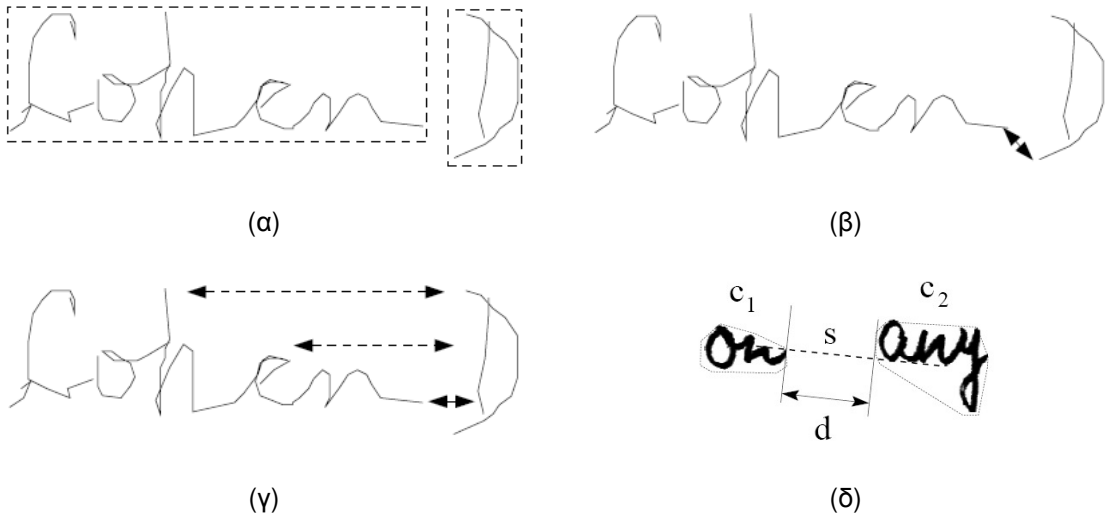
Αν και η κατάτμηση έντυπων κειμένων σε λέξεις μπορεί να επιτευχθεί με πολύ μεγάλη ακρίβεια, η αντίστοιχη εργασία για τα χειρόγραφα κείμενα παραμένει ανοιχτό ερευνητικό θέμα. Το συγκεκριμένο στάδιο επεξεργασίας, ως βαθμίδα ενός συστήματος ανάλυσης εικόνων χειρόγραφου κειμένου, συνήθως τοποθετείται αμέσως μετά το στάδιο κατάτμησης του κειμένου σε γραμμές.

2.1. Σχετικές εργασίες

Η βασική υπόθεση που υιοθετούν οι περισσότερες γνωστές τεχνικές διαχωρισμού μιας γραμμής κειμένου σε λέξεις είναι ότι δεν υπάρχουν ενωμένες λέξεις, δηλαδή δεν υπάρχει CC που να περιέχει τμήματα από δύο γειτονικές λέξεις. Επομένως, κάθε κενό μεταξύ δύο γειτονικών CCs είναι υποψήφια θέση για την τοποθέτηση ενός διαχωριστικού μεταξύ λέξεων. Η κυρίαρχη τάση για την κατάτμηση μιας γραμμής κειμένου σε λέξεις περιλαμβάνει δύο στάδια επεξεργασίας: α) την υιοθέτηση κάποιας ποσότητας για την εκτίμηση του κενού (της απόστασης μεταξύ διαδοχικών CCs) και β) τη διάκριση των κενών σε αυτά που βρίσκονται μεταξύ λέξεων (ΜΛ) και σε αυτά που βρίσκονται εντός λέξεων (ΕΛ).

Για την εκτίμηση του μεγέθους των κενών (gap measures) υιοθετήθηκαν διάφορες ποσότητες, όπως αυτές που εμφανίζονται στο σχήμα 2.1. Η πρώτη προσέγγιση αφορά στην οριζόντια απόσταση των ορθογωνίων που περικλείουν τα εξεταζόμενα CCs (Bounding Boxes Distance, BBD), όπως παρουσιάζεται στο σχ. 2.1α. Στα έντυπα κείμενα, οι αποστάσεις μεταξύ διαδοχικών λέξεων είναι αρκετά μεγαλύτερες από αυτές μεταξύ των χαρακτήρων της ίδιας λέξης και επομένως η χρήση αυτού του μεγέθους για τη ποσοτικοποίηση των κενών είναι ενδεδειγμένη [54]. Όμως, η χρήση της BBD στα χειρόγραφα είναι αναποτελεσματική, γιατί αφενός δεν εμφανίζεται η συγκεκριμένη κανονικότητα και αφετέρου είναι πιθανή η κατακόρυφη

επικάλυψη χαρακτήρων γειτονικών λέξεων (τότε η απόσταση θεωρείται μηδενική). Για το λόγο αυτό προτάθηκαν άλλες ποσότητες όπως η ελάχιστη Ευκλείδεια απόσταση (Minimum Euclidian Distance, ED) των CCs (σχ. 2.1β) και η ελάχιστη οριζόντια απόσταση (σχ. 2.1γ) μεταξύ των pixels κειμένου με την ίδια τεταγμένη (Minimum Run-Length Distance, RLD).



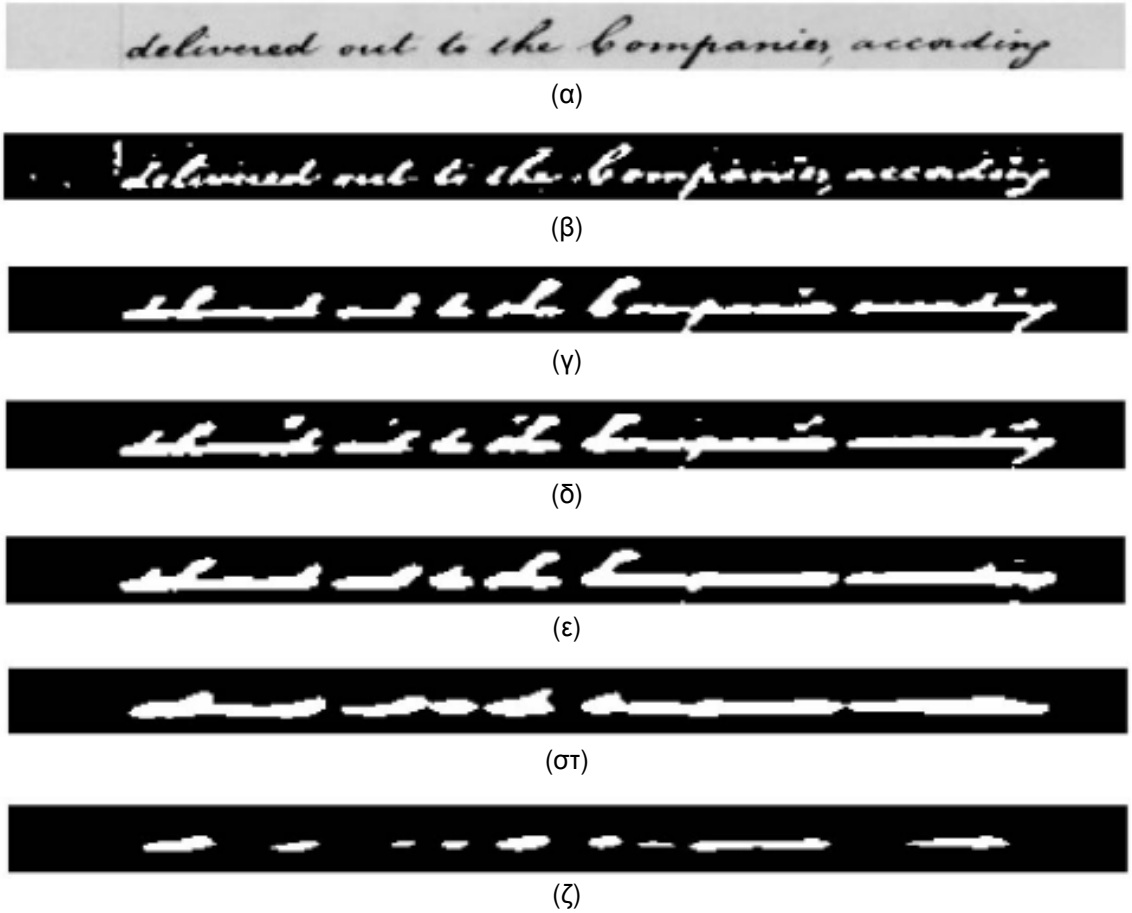
Σχήμα 2.1 Συνήθεις ποσότητες για την εκτίμηση των κενών μεταξύ διαδοχικών CCs, όπως παρουσιάζονται στο [53]). (α) BBD. (β) ED. (γ) RLD. (δ) Απόσταση κυρτών πολυγώνων.

Βέβαια, το γεγονός ότι η κατακόρυφη επικάλυψη των εξεταζόμενων χαρακτήρων δε λαμβάνεται υπόψη κατά τον υπολογισμό αυτών των ποσοτήτων, είναι πιθανό σε μερικές περιπτώσεις να μην αποδίδει κατάλληλα τη «φυσική απόσταση» των χαρακτήρων. Στο [53] παρουσιάζεται η χρήση αυτών των ποσοτήτων για την κατάτμηση χειρόγραφων ταχυδρομικών διευθύνσεων και αναδεικνύεται η ανάγκη επιλογής της πιο ταιριαστής απόστασης κατά περίπτωση. Μια άλλη ποσότητα που αντιπροσωπεύει με επιτυχία τα κενά μεταξύ των διαδοχικών χαρακτήρων προτείνεται στο [55]. Αρχικά, υπολογίζονται τα ελάχιστα κυρτά πολύγωνα (convex hulls) των CCs και τα αντίστοιχα κέντρα μάζας των πολυγώνων. Η προτεινόμενη ποσότητα ισούται με την ευκλείδεια απόσταση των pixels των περιβαλλουσών των πολυγώνων που βρίσκονται στην ευθεία που συνδέει τα κέντρα μάζας (d στο σχ. 2.1δ). Η επιλογή της συγκεκριμένης απόστασης αποδείχθηκε αποτελεσματικότερη κατά την εξέτασή της σε τμήμα της συλλογής χειρόγραφων κειμένου IAM [56]. Σημειώνεται ότι η αδυναμία των προτεινόμενων ποσοτήτων να περιγράψουν τη μεγάλη ποικιλομορφία των κενών που εμφανίζονται στα χειρόγραφα κείμενα, κατέστησε αναγκαία την ενσωμάτωση πρόσθετης *a priori* γνώσης. Η πιο ενδεδειγμένη διαδικασία αφορά στον εντοπισμό ειδικών συμβόλων «.», «,», «(», «)», «[» και «]», που οριοθετούν λέξεις στο γραπτό λόγο.

Το επόμενο στάδιο επεξεργασίας περιλαμβάνει την ταξινόμηση των «αποστάσεων» σε αυτές που περιγράφουν κενά ΜΛ και ΕΛ. Προφανώς, η χρήση προκαθορισμένων ταξινομητών (π.χ. κατώφλια, εκπαιδευμένα μοντέλα) μπορεί να εφαρμοστεί αποτελεσματικά για την επίλυση του προβλήματος. Είναι όμως σημαντικό να αναφερθεί ότι η ποικιλομορφία των χειρόγραφων

κειμένων είναι τόσο μεγάλη, που ίσως περιορίζει σημαντικά τη δυνατότητα γενίκευσης των εκπαιδευμένων ταξινομητών. Για το λόγο αυτό η προτεινόμενη μέθοδος υιοθετηθεί τη χρήση μιας μεθόδου υπολογισμού του κατωφλίου ταξινόμησης των «αποστάσεων» ανά χειρόγραφο κείμενο.

Οι μέθοδοι που βασίζονται στις εκτιμήσεις των κενών μεταξύ των CCs, εμφανίζουν μεγάλη πολυπλοκότητα και αρκετές φορές εσφαλμένα αποτελέσματα όταν οι χαρακτήρες του κείμενου στην προς επεξεργασία εικόνα παρουσιάζονται «σπασμένοι» (broken characters). Η ύπαρξη τέτοιων χαρακτήρων οφείλεται κυρίως σε αστοχίες του μέσου γραφής (π.χ. στυλό που εμφανίζει συχνές διακοπές στο μελάνι) και είναι πολύ συχνή σε ιστορικά χειρόγραφα. Στο [57] προτείνεται μια μέθοδος για την κατάτμηση χειρόγραφων κειμένων του George Washington σε λέξεις, που βασίζεται στην επεξεργασία της εικόνας κάθε γραμμής κειμένου με LoG (Laplacian of Gaussian) φίλτρα διαφορετικής διακύμανσης.



(α)

(β)

(γ)

(δ)

(ε)

(στ)

(ζ)

Σχήμα 2.2. Ανάδειξη των λέξεων μέσω της ανάλυσης της εικόνας της γραμμής κειμένου με LoG φίλτρα διαφορετικής διακύμανσης [57]. (α) Εικόνα γραμμής κειμένου. (β) $\sigma_x = 1$, $\sigma_y = 2$. (γ) $\sigma_x = 2$, $\sigma_y = 4$. (δ) $\sigma_x = 2$, $\sigma_y = 8$. (ε) $\sigma_x = 2.55$, $\sigma_y = 10.2$. (στ) $\sigma_x = 4$, $\sigma_y = 16$. (ζ) $\sigma_x = 5$, $\sigma_y = 20$

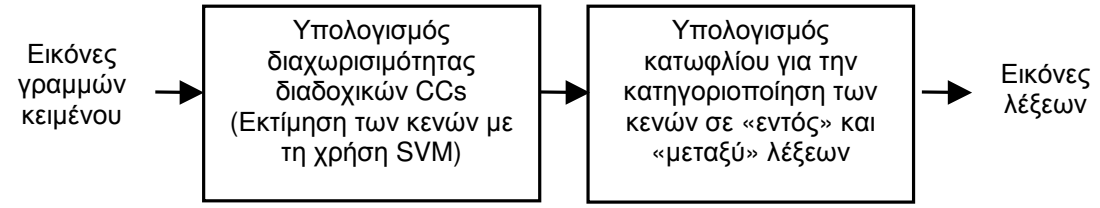
Συγκεκριμένα, για κάθε gray-scale εικόνα γραμμής κειμένου $f(x, y)$ δημιουργείται μια οικογένεια εικόνων

$$I(x, y; \sigma_x, \sigma_y) = L(x, y; \sigma_x, \sigma_y) * f(x, y)$$

συνελίσσοντας την αρχική εικόνα με τη δεύτερη παράγωγο $L(x, y; \sigma_x, \sigma_y)$ (Laplacian) της γκαουσιανής συνάρτησης $G(x, y; \sigma_x, \sigma_y) = (1/2\pi\sigma_x\sigma_y)\exp(-((x^2/2\sigma_x) + (y^2/2\sigma_y)))$, όπου σ_x και σ_y οι τυπικές αποκλίσεις κατά τον οριζόντιο και κατακόρυφο άξονα. Εφαρμόζοντας τον τελεστή L σε αυξανόμενες κλίμακες προκύπτει ότι σε μικρές κλίμακες αναδεικνύονται περιοχές (blobs) που αντιστοιχούν στους χαρακτήρες (σχ. 2.2), σε μεγαλύτερες κλίμακες «εμφανίζονται» οι περιοχές των λέξεων και σε ακόμη μεγαλύτερες είτε ομάδες λέξεων είτε τμήματά τους (λόγω του έντονου θολώματος). Πειραματικά προσδιορίστηκε ότι για $\sigma_x / \sigma_y = 4$ και σ_y ίσο με το 10% του ύψους της εξεταζόμενης γραμμής κειμένου, στην εικόνα αυτής της κλίμακας οι εμφανιζόμενες περιοχές αντιστοιχούν στις λέξεις του κειμένου [58].

2.2. Κατάτμηση χειρόγραφου κειμένου σε λέξεις με τη χρήση SVM

Η προτεινόμενη μέθοδος για την κατάτμηση χειρόγραφου κειμένου σε λέξεις προϋποθέτει, όπως και οι περισσότερες γνωστές τεχνικές, το χωρισμό του κειμένου σε γραμμές. Επίσης, χρησιμοποιεί την υπόθεση ότι κάθε CC ανήκει σε μία μόνο λέξη, δηλαδή διαδοχικές λέξεις δεν είναι ενωμένες. Επομένως, τα υποψήφια διαχωριστικά των λέξεων βρίσκονται στα κενά μεταξύ διαδοχικών CCs. Όπως έχει αναφερθεί, το πρόβλημα εντοπισμού των λέξεων ανάγεται σε ένα πρόβλημα ποσοτικοποίησης των κενών και κατηγοριοποίησής τους σε αυτά που διαχωρίζουν λέξεις (ΜΛ, μεταξύ λέξεων) και σε αυτά που διαχωρίζουν CCs της ίδιας λέξης (ΕΛ, εντός λέξεων). Αυτά τα βήματα ανάλυσης ακολουθεί και η προτεινόμενη μέθοδος (σχ. 2.3).



Σχήμα 2.3 Διάγραμμα προτεινόμενης μεθόδου.

Για την εκτίμηση του κενού μεταξύ διαδοχικών CCs, προτείνεται μια νέα ποσότητα που βασίζεται στη «διαχωρισιμότητα» των δύο τάξεων που τα περιέχουν. Αν θεωρήσουμε ότι τα εικονοστοιχεία κάθε CC συνιστούν μια κλάση, τότε το περιθώριο ταξινόμησης που αντιστοιχεί στο βέλτιστο γραμμικό ταξινομητή χαλαρών περιθωρίων (linear soft-margin Support Vector Machine), εκφράζει τη «διαχωρισιμότητα» των τάξεων και μπορεί να θεωρηθεί ως ένας καλός

εκτιμητής της «φυσικής απόστασης» διαδοχικών χαρακτήρων. Υπολογίζοντας την ποσότητα αυτή για κάθε ζεύγος διαδοχικών CCs ανά γραμμή κειμένου, προκύπτει μια συνολική εκτίμηση για τις «αποστάσεις» των χαρακτήρων του κειμένου. Υποθέτοντας ότι οι αποστάσεις μεταξύ των λέξεων είναι μεγαλύτερες από αυτές μεταξύ των χαρακτήρων της ίδιας λέξης, ο εντοπισμός των ΜΛ κενών, μπορεί να επιτευχθεί με τη διάκριση των ποσοτήτων σε «μικρές» και «μεγάλες». Για τον υπολογισμό του κατωφλίου που διακρίνει τις ποσότητες, υπολογίζεται η συνάρτηση πυκνότητας πιθανότητας των ποσοτήτων (σε όλο το κείμενο). Ως κατάλληλο κατώφλι επιλέγεται η τιμή που αντιστοιχεί στο δεξιότερο τοπικό ελάχιστο της συνάρτησης πυκνότητας πιθανότητας.

Η προτεινόμενη μέθοδος αξιολογήθηκε στο πλαίσιο των διαγωνισμών κατάτμησης χειρόγραφων κειμένων (ICDAR07 και ICDAR2009 handwriting segmentation contests) και κρίθηκε ως η πιο αποτελεσματική. Τα αναλυτικά αποτελέσματα των διαγωνισμών με τις επιμέρους συγκρίσεις και τις συνοπτικές περιγραφές των διαγωνιζόμενων αλγορίθμων παρουσιάζονται στα [7, 38].

2.2.1. Εκτίμηση κενών

Το πρώτο στάδιο επεξεργασίας είναι η ταξινόμηση κατά αύξουσα σειρά των CCs μιας γραμμής κειμένου, με βάση την τετμημένη του κέντρου μάζας τους. Έτσι, κάθε CC της γραμμής δεικτοδοτείται με το δείκτη $k = 1, 2, ..., K_l$, όπου K_l το πλήθος των CCs της l-οστής γραμμής κειμένου. Όπως έχει αναφερθεί, κάθε κενό μεταξύ διαδοχικών CC αποτελεί υποψήφιο κενό μεταξύ λέξεων. Ένα κατάλληλο μέγεθος για την εκτίμηση του κενού g_k^l, μεταξύ του C_k και του C_{k+1}, θα πρέπει να ενσωματώνει τη γεωμετρική απόσταση μεταξύ των δύο CCs σε όλο το μέτωπο της γραμμής κειμένου και την επικάλυψή τους κατά τον κατακόρυφο άξονα.

Σχήμα 2.4 Εξέταση του κενού μεταξύ των χαρακτήρων «l» και «t» της 10ης γραμμής κειμένου (022.tif, ICDAR-07).

Α) Γραμμικά διαχωρίσιμες τάξεις

Έστω ότι εξετάζεται η απόσταση g_k^l για το κενό μεταξύ των χαρακτήρων «l» και «t» (αντικείμενα C_k και C_{k+1} αντίστοιχα) της γραμμής κειμένου στο σχ. 2.4. Θεωρώντας ότι τα αντικείμενα αποτελούνται από M pixels συνολικά, αναπαριστούμε με $\mathbf{x}_i \in \mathbb{Z}^2$, $i = 1, 2, ..., M$ τα διανύσματα συντεταγμένων των M pixels και τα δεικτοδοτούμε με $y_i = -1$ ή $y_i = 1$ αν ανήκουν στο αντικείμενο C_k (τάξη ω_1) ή C_{k+1} (τάξη ω_2) αντίστοιχα. Προφανώς, η εκτίμηση της «διαχωρισιμότητας» των δύο τάξεων είναι ένας τρόπος ποσοτικοποίησης της απόστασης μεταξύ των χαρακτήρων.

Αν οι τάξεις ω_1 και ω_2 είναι γραμμικά διαχωρίσιμες, τότε υπάρχει μία τουλάχιστον γραμμική συνάρτηση $f(\mathbf{x}_i) = \mathbf{w}^T\mathbf{x}_i + w_0$: $\begin{Bmatrix} f(\mathbf{x}_i) > 0,\ \mathbf{x}_i \in \omega_1 \\ f(\mathbf{x}_i) < 0,\ \mathbf{x}_i \in \omega_2 \end{Bmatrix}$, όπου $\mathbf{w}$ και w_0 οι παράμετροι και το κατώφλι που προσδιορίζονται από το σετ εκπαίδευσης $Z_k = (X_k, Y_k)$, $X_k = \{\mathbf{x}_i\}$ και $Y_k = \{y_i\}$. Ο γραμμικός αυτός ταξινομητής ορίζει την ευθεία $f(\mathbf{x}) = 0$ ως την ευθεία διαχωρισμού των δύο τάξεων. Κάθε ανισότητα επιβάλλει έναν περιορισμό στη θέση της ευθείας διαχωρισμού. Επομένως, το σύνολο των περιορισμών ορίζει έναν «τόπο λύσεων», έναν τόπο αποδεκτών θέσεων της ευθείας διαχωρισμού. Ο βέλτιστος ταξινομητής, δηλαδή αυτός με την μεγαλύτερη δυνατότητα γενίκευσης, είναι αυτός που βρίσκεται «βαθύτερα» στον τόπο λύσεων με την έννοια ότι με αυτόν είναι μεγαλύτερη η πιθανότητα να ταξινομηθεί σωστά ένα νέο άγνωστο πρότυπο. Υιοθετώντας τη μεθοδολογία των μηχανών διανυσμάτων υποστήριξης [κεφ. 7 στο 59] για τον υπολογισμό των βαρών $\mathbf{w}$ και του κατωφλίου w_0 του βέλτιστου ταξινομητή, θα επιλεγεί η ευθεία διαχωρισμού που:

α) ισαπέχει από τα πλησιέστερα σε αυτή σημεία των δύο κλάσεων, (κόκκινη και διακεκομμένη ευθεία του σχ. 2.5α)

β) μεγιστοποιεί την απόσταση των σημείων αυτών από την ευθεία διαχωρισμού (κόκκινη ευθεία του σχ. 2.5α).

Τα κοντινότερα εκατέρωθεν σημεία στην ευθεία διαχωρισμού ονομάζονται διανύσματα υποστήριξης και το διπλάσιο της μέγιστης απόστασης είναι το περιθώριο ταξινόμησης (margin). Σύμφωνα με τα προηγούμενα, η ευθεία διαχωρισμού με το μέγιστο περιθώριο ορίζει το βέλτιστο ταξινομητή για το πρόβλημα. Επομένως, το περιθώριο ταξινόμησης μπορεί να αποτελέσει μια κατάλληλη ποσότητα για την εκτίμηση της «απόστασης» μεταξύ των δύο κλάσεων, ή αντίστοιχα της «διαχωρισιμότητας» των δύο κλάσεων. Άρα, στο συγκεκριμένο πρόβλημα μπορεί να χρησιμοποιηθεί για την απόδοση του μεγέθους του κενού μεταξύ δύο διαδοχικών χαρακτήρων.

Έστω δύο γραμμικά διαχωρίσιμες τάξεις (ω_1 και ω_2), $f(\mathbf{x}) = 0$ η ευθεία διαχωρισμού και $\mathbf{x}_1 \in \omega_1, \mathbf{x}_2 \in \omega_2$ τα κοντινότερα σε αυτή σημεία από κάθε τάξη (σχ. 2.5β). Αν d_1 και d_2 οι αποστάσεις των σημείων αυτών από την ευθεία διαχωρισμού και $\mathbf{x}_{p1}, \mathbf{x}_{p2}$ οι προβολές τους σε αυτή τότε: $\mathbf{x}_1 = \mathbf{x}_{p1} + d_1\frac{\mathbf{w}}{\|\mathbf{w}\|}$ και $\mathbf{x}_2 = \mathbf{x}_{p2} - d_2\frac{\mathbf{w}}{\|\mathbf{w}\|}$.

Επίσης,

$$f(\mathbf{x}_1) = \mathbf{w}^T\left(\mathbf{x}_{p1} + d_1\frac{\mathbf{w}}{\|\mathbf{w}\|}\right) + w_0 = \mathbf{w}^T\mathbf{x}_{p1} + w_0 + d_1\mathbf{w}^T\frac{\mathbf{w}}{\|\mathbf{w}\|} = f(\mathbf{x}_{p1}) + d_1\frac{\|\mathbf{w}\|^2}{\|\mathbf{w}\|} = d_1\|\mathbf{w}\| \qquad \text{(Εξ. 2.1)}$$

και

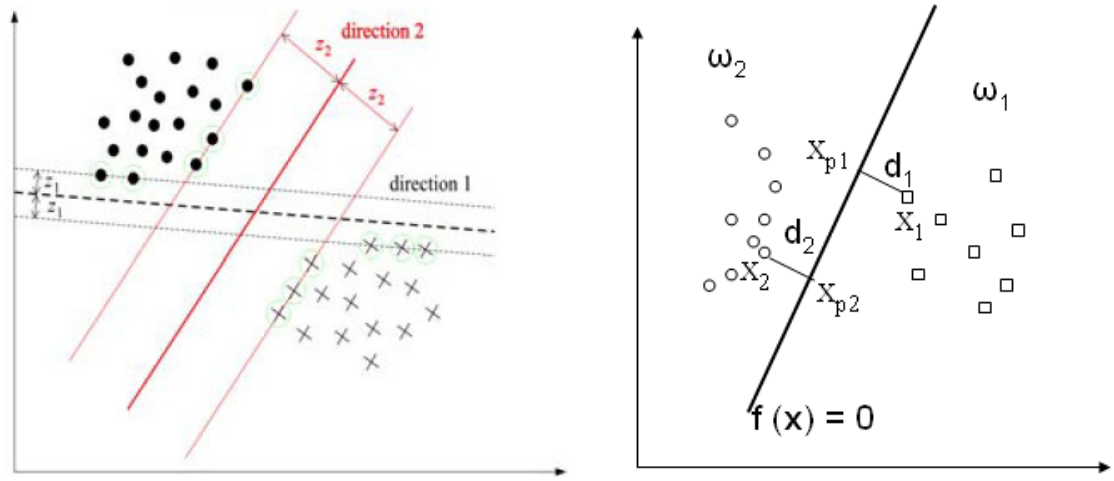
$$f(\mathbf{x}_2) = \mathbf{w}^T\left(\mathbf{x}_{p2} - d_2\frac{\mathbf{w}}{\|\mathbf{w}\|}\right) + w_0 = \mathbf{w}^T\mathbf{x}_{p2} + w_0 - d_2\mathbf{w}^T\frac{\mathbf{w}}{\|\mathbf{w}\|} = f(\mathbf{x}_{p2}) - d_2\frac{\|\mathbf{w}\|^2}{\|\mathbf{w}\|} = d_2\|\mathbf{w}\| \qquad \text{(Εξ.2.2)}$$

Από τις (Εξ. 2.1) και (Εξ. 2.2) προκύπτει ότι $d_1 = f(\mathbf{x}_1)/\|\mathbf{w}\|$ και $d_2 = -f(\mathbf{x}_2)/\|\mathbf{w}\|$.

Λαμβάνοντας υπόψη ότι τα κοντινότερα σημεία πρέπει να ισαπέχουν από την ευθεία διαχωρισμού (d_1=d_2) και τροποποιώντας τις παραμέτρους $\mathbf{w}$ και w_0 έτσι ώστε $f(\mathbf{x}_1) = 1$ και $f(\mathbf{x}_2) = -1$, προκύπτει ότι το περιθώριο ταξινόμησης γ ισούται με $2/\|\mathbf{w}\|$.

$$\gamma = 2/\|\mathbf{w}\| \qquad \text{(Εξ. 2.3)}$$

Φυσικά, για κάθε άλλο σημείο $\mathbf{x}_i$ θα πρέπει να ισχύει ότι $|f(\mathbf{x}_i)| - 1 > 0$.



(α) (β)

Σχήμα 2.5 α) Επιλογή της ευθείας διαχωρισμού με τη μέθοδο των μηχανών διανυσμάτων υποστήριξης. β) Υπολογισμός απόστασης των διανυσμάτων υποστήριξης από την ευθεία διαχωρισμού.

Το δεύτερο στάδιο της μεθοδολογίας των SVM είναι η μεγιστοποίηση του περιθωρίου, δηλαδή η ελαχιστοποίηση της ποσότητας $\|\mathbf{w}\|$, χωρίς φυσικά να γίνονται λάθη ταξινόμησης. Επομένως, το πρόβλημα μπορεί να διατυπωθεί ως εξής:

«ελαχιστοποίηση της $J(\mathbf{w}, w_0) = \frac{1}{2}\|\mathbf{w}\|^2$ με $y_i\left(\mathbf{w}^T\mathbf{x}_i + w_0\right) - 1 \geq 0,\ i = 1,...,M$ »

ή ισοδύναμα με την εισαγωγή των συντελεστών Lagrange:

«ελαχιστοποίηση της $L(\mathbf{w}, w_0, \boldsymbol{\lambda}) = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{N}\lambda_i\left(y_i\left(\mathbf{w}^T\mathbf{x}_i + w_0\right) - 1\right)$ με $\lambda_i \geq 0$ »

Σημειώνεται ότι $\min L(\mathbf{w}, w_0, \boldsymbol{\lambda}) \leq \min J$ αφού $\sum_{i=1}^{M}\lambda_i\left(y_i\left(\mathbf{w}^T\mathbf{x}_i + w_0\right) - 1\right) \geq 0$

Οι συναρτήσεις $J(\mathbf{w}, w_0)$ και $L(\mathbf{w}, w_0, \boldsymbol{\lambda})$ είναι κυρτές και υπόκεινται σε γραμμικούς περιορισμούς, οπότε οι συνθήκες Karush–Kuhn–Tucker (KKT) είναι ικανές και αναγκαίες για την

ύπαρξη ολικού ελάχιστου στα ανωτέρω προβλήματα [60]. Εφαρμόζοντας τις συνθήκες ΚΚΤ έχουμε:

$$\frac{\partial}{\partial \mathbf{w}} L(\mathbf{w}^*, w_0^*, \boldsymbol{\lambda}) = \mathbf{0} \qquad \Rightarrow \ \mathbf{w}^* = \sum_{i=1}^{M} \lambda_i \, y_i \mathbf{x}_i \tag{Εξ. 2.4}$$

$$\frac{\partial}{\partial w_0} L(\mathbf{w}^*, w_0^*, \boldsymbol{\lambda}) = 0 \qquad \Rightarrow \ \sum_{i=1}^{M} \lambda_i \, y_i = 0 \tag{Εξ. 2.5}$$

$$\lambda_i \geq 0, \ i = 1, ..., M \tag{Εξ. 2.6}$$

$$\lambda_i \left(y_i \left(\mathbf{w}^T \mathbf{x}_i + w_0 \right) - 1 \right) = 0, \ i = 1, ..., M \tag{Εξ. 2.7}$$

Αντικαθιστώντας την (Εξ. 2.4) στη συνάρτηση κόστους προκύπτει ότι:

$$\min L(\mathbf{w}^*, w_0^*, \boldsymbol{\lambda}) = \min\left(\frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M} \lambda_i \lambda_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j - \sum_{i=1}^{M}\sum_{j=1}^{M} \lambda_i \lambda_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j - \sum_{i=1}^{M} \lambda_i y_i w_0 + \sum_{i=1}^{M} \lambda_i\right) =$$

$$= \min\left(\sum_{i=1}^{M} \lambda_i - \frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M} \lambda_i \lambda_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j\right) \leq \min J$$

Επομένως, το πρόβλημα περιγράφεται ως εξής:

$$\max_{\boldsymbol{\lambda}} L(\mathbf{w}^*, w_0^*, \boldsymbol{\lambda}) \text{ με τους περιορισμούς } \sum_{i=1}^{M} \lambda_i y_i = 0 \text{ και } \lambda_i \geq 0$$

ή ισοδύναμα (λόγω της κυρτής συνάρτησης κόστους, ορίζουμε το δυικό πρόβλημα):

$$\min_{\boldsymbol{\lambda}}\left(\frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M} \lambda_i \lambda_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j - \sum_{i=1}^{M} \lambda_i\right) \text{ με τους περιορισμούς } \sum_{i=1}^{M} \lambda_i y_i = 0 \text{ και } \lambda_i \geq 0$$

Από την (Εξ. 2.7) συμπεραίνουμε πως οι τιμές λ_i μπορούν να είναι μη μηδενικές μόνο για τα διανύσματα $\mathbf{x}_i$ που $| f(\mathbf{x}_i) | = 1$, δηλαδή τα κοντινότερα στην ευθεία διαχωρισμού. Επομένως, από την (Εξ. 2.4) προκύπτει ότι τα βέλτιστα βάρη $\mathbf{w}^*$ εξαρτώνται μόνο από αυτά τα διανύσματα, τα οποία ονομάζονται διανύσματα υποστήριξης.

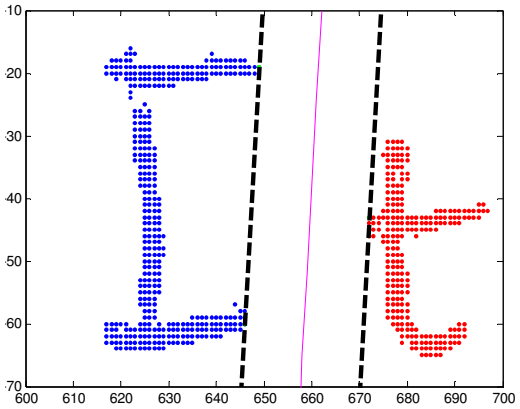
Από την επίλυση του προβλήματος προκύπτει άμεσα η τιμή της αντικειμενικής συνάρτησης, L^*, η οποία ως άμεσα συνδεδεμένη με το περιθώριο ταξινόμησης, είναι κατάλληλη για την εκτίμηση της «διαχωρισιμότητας» των δύο κλάσεων, άρα και για την απόδοση του κενού μεταξύ των χαρακτήρων.

Β) Παραδείγματα

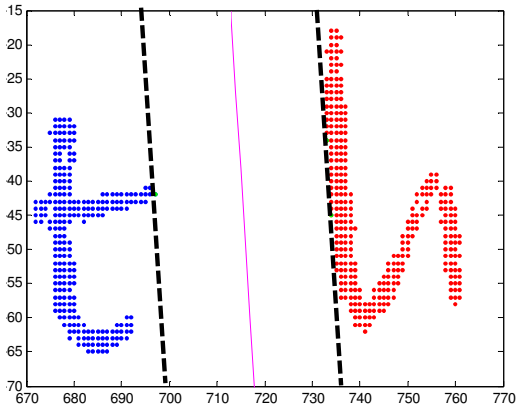
Στο σχ. 2.6-2.8 παρουσιάζονται μερικά παραδείγματα εκτίμησης των κενών με βάση τα ανωτέρω. Οι ευθείες διαχωρισμού εμφανίζονται με μωβ χρώμα ενώ οι αντίστοιχες παράλληλες ευθείες που διέρχονται από τα διανύσματα υποστήριξης εμφανίζονται διακεκομμένες. Για το

κενό που εξετάζεται στο σχ. 2.6β, οι τιμές του περιθωρίου ταξινόμησης και της αντικειμενικής συνάρτησης είναι 24.6901 και 0.0033 αντίστοιχα. Ομοίως, για το κενό στο σχ. 2.6γ, οι τιμές είναι 36.5764 και 0.0015. Από τη σύγκριση των τιμών των περιθωρίων ταξινόμησης για κάθε πρόβλημα και την παρατήρηση των «φυσικών αποστάσεων» κάθε ζεύγους χαρακτήρων (σχ. 2.5α), προκύπτει ότι το μεγαλύτερο περιθώριο ταξινόμησης αντιστοιχεί στο μεγαλύτερο κενό. Το αντίστροφο ισχύει για τις τιμές των αντικειμενικών συναρτήσεων, όπως άλλωστε είναι φυσικό, λόγω του ορισμού τους. Σημειώνεται επίσης, ότι για τα συγκεκριμένα προβλήματα, τα περιθώρια ταξινόμησης έχουν τιμές που είναι πολύ κοντινές στις αντίστοιχες BBD, μια και η ευθεία διαχωρισμού έχει πολύ μικρή κλίση σε σχέση με τον κατακόρυφο άξονα.

(α)



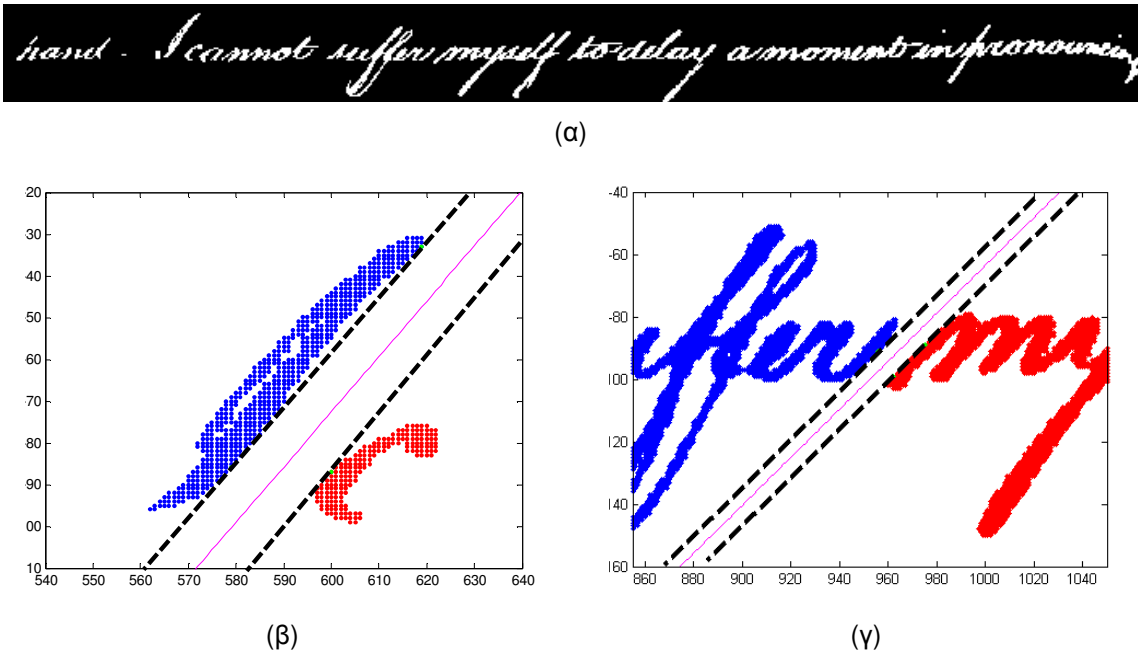
(β)



(γ)

Σχήμα 2.6 α) Εικόνα της 10ης γραμμής κειμένου (022.tif, ICDAR-07). β) Εξέταση του κενού μεταξύ «l» και «t». γ) Εξέταση του κενού μεταξύ «t» και «h».

Αντίθετα, για τα κενά των σχ. 2.7β και 2.7γ, όπου οι χαρακτήρες παρουσιάζουν επικάλυψη κατά τον κατακόρυφο άξονα λόγω του τρόπου γραφής που ακολουθεί ο συγγραφέας, οι BBD τιμές θα λαμβάνονταν ως μηδενικές. Με την προτεινόμενη μέθοδο οι αντίστοιχες τιμές είναι 17.3644 (0.0066) και 9.9993 (0.02) που συμφωνούν με τα «φυσικά μεγέθη» των κενών (σχ. 2.7α). Παρόμοια εκτίμηση των κενών επιτυγχάνεται με τη χρήση της ελάχιστης Ευκλείδειας απόστασης (17.4929 και 10 αντίστοιχα).



(α)

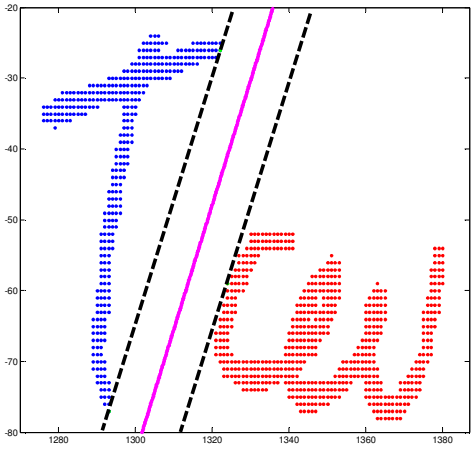
(β) (γ)

Σχήμα 2.7 α) Εικόνα της 2ης γραμμής κειμένου (009.tif ICDAR-07). β) Εξέταση του κενού μεταξύ «I» και «c». γ) Εξέταση του κενού μεταξύ των «suffer» και «myself».

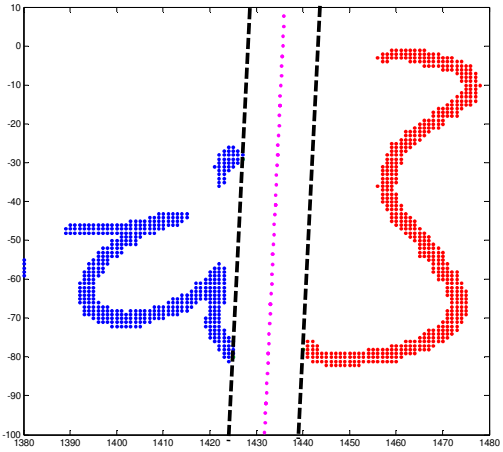
Η εκτίμηση των κενών των σχ. 2.8β και γ, προσεγγίζεται καλύτερα με την προτεινόμενη ποσότητα. Πράγματι, οι τιμές των περιθωρίων ταξινόμησης (αντικειμενικών συναρτήσεων) είναι 18.0506 (0.0061) και 14.9101 (0.009) αντίστοιχα. Οι τιμές των ελάχιστων ευκλείδειων αποστάσεων και των ελάχιστων οριζόντιων αποστάσεων για το σχ. 2.8β είναι 26.5707 και 29, ενώ για το σχ. 2.8γ είναι 15 και 15. Επομένως, η προτεινόμενη ποσότητα έχει αποδώσει στα κενά «κοντινότερες τιμές», γεγονός που διευκολύνει την ταξινόμησή τους στην ίδια κατηγορία (π.χ. ΜΛ) στο επόμενο στάδιο επεξεργασίας.

Από τα προηγούμενα παραδείγματα, γίνεται αντιληπτό ότι η προτεινόμενη μέθοδος χρησιμοποιεί τις συντεταγμένες των pixels για να ορίσει την κατάλληλη διεύθυνση (κάθετη στην ευθεία διαχωρισμού) για τη «μέτρηση» της απόστασης. Πράγματι, όταν το κείμενο δεν εμφανίζει κλίση, η ευθεία διαχωρισμού τείνει να γίνει κατακόρυφη, ενώ όταν υπάρχει πλάγια γραφή, ακολουθεί την κλίση της γραφής. Επίσης, η ευθεία διαχωρισμού προσαρμόζεται κατάλληλα όταν υπάρχει κατακόρυφη επικάλυψη των εξεταζόμενων CCs.

(α)



(β)

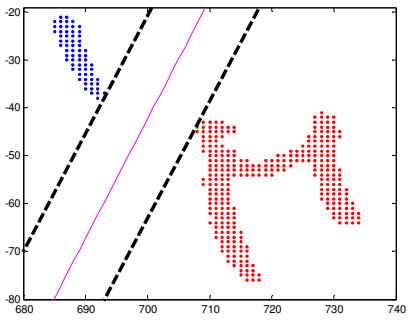


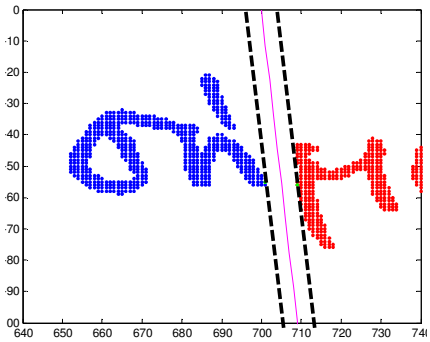
(γ)

Σχήμα 2.8 α) Εικόνα τμήματος της 1ης γραμμής κειμένου (117.tif, ICDAR-09). β) Εξέταση του κενού μεταξύ των «τ» και «αυ». γ) Εξέταση του κενού μεταξύ των «τί» και «ζ».

Γ) Επιλογή σημείων

Όπως, έχει αναφερθεί, ως πιθανές θέσεις των διαχωριστικών θεωρούνται τα κενά μεταξύ διαδοχικών CCs. Όμως, κάποιες περιπτώσεις όπως αυτή του σχ. 2.9 θα πρέπει να αποκλειστούν. Για το λόγο αυτό κατά την εξέταση του κενού g_k^l μεταξύ των χαρακτήρων C_k και του C_{k+1}, χρησιμοποιούνται και υπόλοιπα CCs της γραμμής κειμένου l. Συγκεκριμένα, διακρίνονται σε δύο ομάδες με την πρώτη να αποτελείται από τα pixels των C_1, C_2,..., C_k και η δεύτερη να περιλαμβάνει τα pixels των C_{k+1}, C_{k+2}, ... , C_{end}, με βάση την αρχική ταξινόμησή τους. Με τον τρόπο αυτό, αποφεύγονται παραπλανητικές εκτιμήσεις, όπως αυτή του σχ. 2.9β για το 23° κενό, και αντικαθίστανται από ρεαλιστικές όπως αυτή του σχ. 2.9γ.

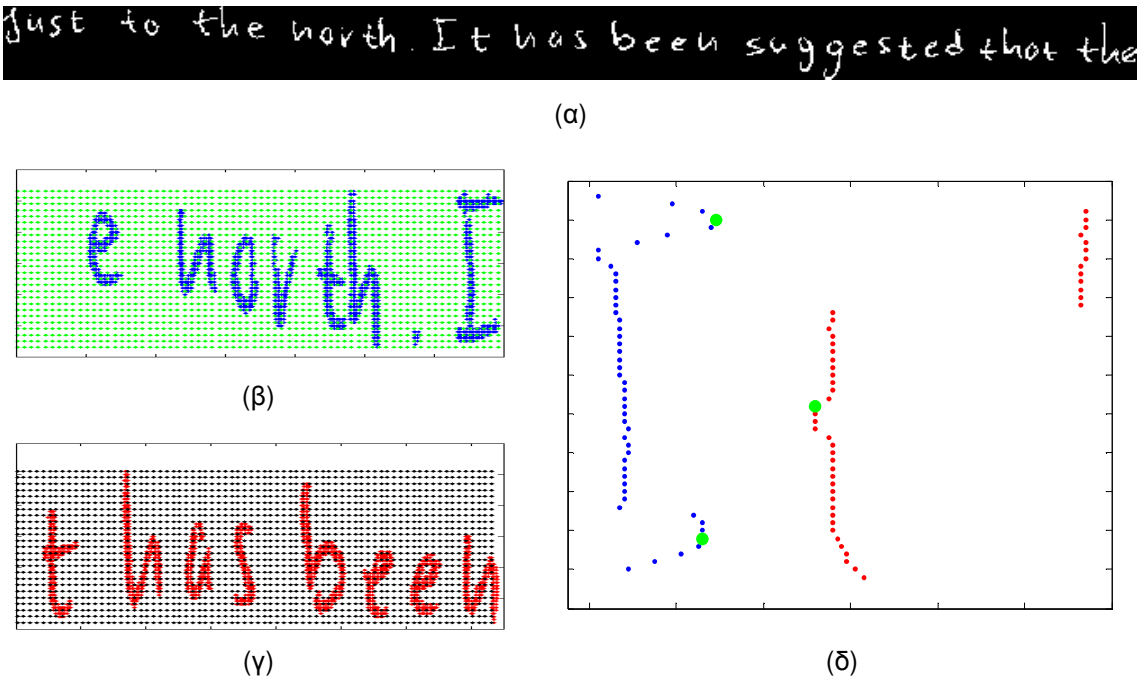




(α) (β) (γ)

Σχήμα 2.9 α) Διατεταγμένα CCs. β) Παραπλανητική εκτίμηση κενού. γ) Ρεαλιστική εκτίμηση κενού.

Όπως είναι φυσικό, το υπολογιστικό φορτίο για την εκτίμηση κάθε κενού θα ήταν ιδιαίτερα μεγάλο αν οι τάξεις συμπεριελάμβαναν όλα τα pixels κειμένου. Όμως, σύμφωνα με τη μεθοδολογία των SVM (βλ. Εξ. 2.3, 2.4), η τιμή του περιθωρίου ταξινόμησης εξαρτάται μόνο από τα κοντινότερα διανύσματα στην ευθεία διαχωρισμού. Επομένως, για την ποσοτικοποίηση της διαχωρισιμότητας των δύο ομάδων είναι απαραίτητη η χρήση μόνο ορισμένων από τα pixels που ανήκουν σε αυτές. Είναι προφανές ότι θα πρέπει να επιλεγούν τα pixels με την μεγαλύτερη (μικρότερη) τετμημένη για την αριστερή (δεξιά) ομάδα. Επίσης, θα πρέπει να εκτείνονται σε όλο το μέτωπο (ύψος) κάθε ομάδας. Για το λόγο αυτό επιλέγουμε το μέγιστο πλήθος επιλεγμένων σημείων ανά ομάδα $N_{max}= 200$ (υποθέτοντας ότι η τιμή αυτή είναι ένα λογικό ανώτατο όριο για το ύψος των χαρακτήρων). Αν h το ύψος μιας ομάδας, τότε επιλέγουμε (N_{max}/h) pixels από κάθε γραμμή (σχ. 2.10). Με τον τρόπο αυτό επιταχύνεται η διαδικασία βελτιστοποίησης που ακολουθεί, χωρίς να επηρεάζεται η διαχωρισιμότητα των κλάσεων.



(α) (β) (γ) (δ)

Σχήμα 2.10 α) Η 10η γραμμή κειμένου της εικόνας 022.tif από ICDAR07. (β-γ) Οι ομάδες για την εκτίμηση του κενού μεταξύ «l» και «t». (δ) Τα επιλεγμένα pixels από κάθε ομάδα (τα διανύσματα υποστήριξης δηλώνονται με πράσινο χρώμα).

Δ) Μη γραμμικά διαχωρίσιμες τάξεις

Η βασική υπόθεση που έχει ως τώρα υιοθετηθεί, είναι ότι τα pixels των διαδοχικών χαρακτήρων (ή των ομάδων) συνιστούν γραμμικά διαχωρίσιμες τάξεις. Όμως, πολύ συχνά αυτό δεν επαληθεύεται. Σε τέτοιες περιπτώσεις, η μεθοδολογία των SVM προτείνει τη χρήση των μηχανών διανυσμάτων υποστήριξης χαλαρών περιθωρίων [61]. Η βασική διαφοροποίηση από την αρχική θεώρηση είναι ότι πλέον επιτρέπεται η ύπαρξη διανυσμάτων εντός του περιθωρίου ταξινόμησης. Επομένως, οι περιορισμοί που είχαμε στην περίπτωση των γραμμικά

διαχωρίσιμων τάξεων, γίνονται τώρα πιο χαλαροί με την εισαγωγή των «χαλαρών μεταβλητών» (slack variables) ξ_i ως εξής:

$$y_i\left(\mathbf{w}^T\mathbf{x}_i + w_0\right) \geq 1 - \xi_i,\ i = 1,...,M$$
$$\xi_i \geq 0 \qquad \text{(Εξ. 2.8)}$$

Από την (Εξ. 2.8) συμπεραίνουμε ότι κάθε διάνυσμα $\mathbf{x}_i$ στο οποίο αντιστοιχεί $0 < \xi_i \leq 1$, είναι σωστά ταξινομημένο, αλλά βρίσκεται εντός του περιθωρίου ταξινόμησης. Αντίθετα, αν $\xi_i > 1$, το αντίστοιχο διάνυσμα δεν έχει ανατεθεί στη σωστή τάξη. Ο στόχος είναι η εύρεση ενός γραμμικού ταξινομητή που αφενός θα έχει μεγάλο περιθώριο ταξινόμησης (δυνατότητα γενίκευσης) και αφετέρου θα κάνει λίγα λάθη. Συνεπώς, το πρόβλημα παίρνει την ακόλουθη μορφή:

« ελαχιστοποίηση της $J(\mathbf{w}, w_0, \boldsymbol{\xi}) = \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{M}\xi_i$ με τους ανωτέρω περιορισμούς»,

όπου ο πρώτος όρος αντιστοιχεί στη μεγιστοποίηση του περιθωρίου και ο δεύτερος στην ελαχιστοποίηση των πιθανών σφαλμάτων ταξινόμησης. Η παράμετρος $C > 0$ δηλώνει την ποινή που επιθυμούμε να επιβληθεί για κάθε σφάλμα. Με την εισαγωγή των συντελεστών Lagrange, το πρόβλημα γίνεται:

ελαχιστοποίηση της $L(\mathbf{w}, w_0, \boldsymbol{\xi}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{M}\xi_i - \sum_{i=1}^{M}\lambda_i\left(y_i\left(\mathbf{w}^T\mathbf{x}_i + w_0\right) - 1 + \xi_i\right) - \sum_{i=1}^{M}\mu_i\xi_i$

$$\text{με } \lambda_i, \mu_i \geq 0 \qquad \text{(Εξ.2.9)}$$

Εφαρμόζοντας τις συνθήκες ΚΚΤ για τη βέλτιστη λύση έχουμε:

$$\frac{\partial}{\partial \mathbf{w}} L(\mathbf{w}^*, w_0^*, \boldsymbol{\xi}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = 0 \quad \Rightarrow \mathbf{w}^* = \sum_{i=1}^{M}\lambda_i y_i \mathbf{x}_i \qquad \text{(Εξ. 2.10)}$$

$$\frac{\partial}{\partial w_0} L(\mathbf{w}^*, w_0^*, \boldsymbol{\xi}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = 0 \quad \Rightarrow \sum_{i=1}^{M}\lambda_i y_i = 0 \qquad \text{(Εξ. 2.11)}$$

$$\frac{\partial}{\partial \boldsymbol{\xi}} L(\mathbf{w}^*, w_0^*, \boldsymbol{\xi}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = 0 \quad \Rightarrow C - \lambda_i - \mu_i = 0 \qquad \text{(Εξ. 2.12)}$$

$$\lambda_i\left(y_i\left(\mathbf{w}^T\mathbf{x}_i + w_0\right) - 1 + \xi_i\right) = 0, \quad i = 1,...,M \qquad \text{(Εξ. 2.13)}$$

$$\lambda_i \geq 0, \quad i = 1,...,M \qquad \text{(Εξ. 2.14)}$$

$$\mu_i \geq 0, \quad i = 1,...,M \qquad \text{(Εξ. 2.15)}$$

$$\mu_i\xi_i = 0, \quad i = 1,...,M \qquad \text{(Εξ. 2.16)}$$

Από τις (Εξ. 2.9) και (Εξ. 2.10) προκύπτει ότι:

$$L(\mathbf{w}^*, w_0^*, \boldsymbol{\xi}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M}\lambda_i\lambda_j y_i y_j \mathbf{x}_i^T\mathbf{x}_j + C\sum_{i=1}^{M}\xi_i - \sum_{i=1}^{M}\sum_{j=1}^{M}\lambda_i\lambda_j y_i y_j \mathbf{x}_i^T\mathbf{x}_j - \sum_{i=1}^{M}\lambda_i y_i w_0$$
$$+\sum_{i=1}^{M}\lambda_i - \sum_{i=1}^{M}\lambda_i\xi_i - \sum_{i=1}^{M}\mu_i\xi_i =$$
$$= \sum_{i=1}^{M}\lambda_i - \frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M}\lambda_i\lambda_j y_i y_j \mathbf{x}_i^T\mathbf{x}_j + \sum_{i=1}^{M}(C-\lambda_i-\mu_i)\xi_i$$

και αντικαθιστώντας από τη (Εξ. 2.12) έχουμε:

$$L(\mathbf{w}^*, w_0^*, \boldsymbol{\xi}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \sum_{i=1}^{M}\lambda_i - \frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M}\lambda_i\lambda_j y_i y_j \mathbf{x}_i^T\mathbf{x}_j \qquad \text{(Εξ. 2.17)}$$

Από τις (Εξ. 2.12), (Εξ. 2.14) και (Εξ. 2.15) προκύπτει ότι: $0 \le \lambda_i \le C$. Επομένως, καταλήγουμε στην επίλυση [62] του ακόλουθου προβλήματος βελτιστοποίησης:

$$\min_{\boldsymbol{\lambda}}\left(\frac{1}{2}\sum_{i=1}^{N_s}\sum_{j=1}^{N_s}\lambda_i\lambda_j y_i y_j \mathbf{x}_i^T\mathbf{x}_j - \sum_{i=1}^{N_s}\lambda_i\right) \text{ με τους περιορισμούς } \sum_{i=1}^{N_s}\lambda_i y_i = 0 \text{ και } 0 \le \lambda_i \le C)$$

Η συνάρτηση κόστους είναι ίδια με την περίπτωση των γραμμικά διαχωρίσιμων τάξεων, αλλά τώρα οι τιμές των συντελεστών Lagrange ελέγχονται και από την επιλεγμένη σταθερά C.

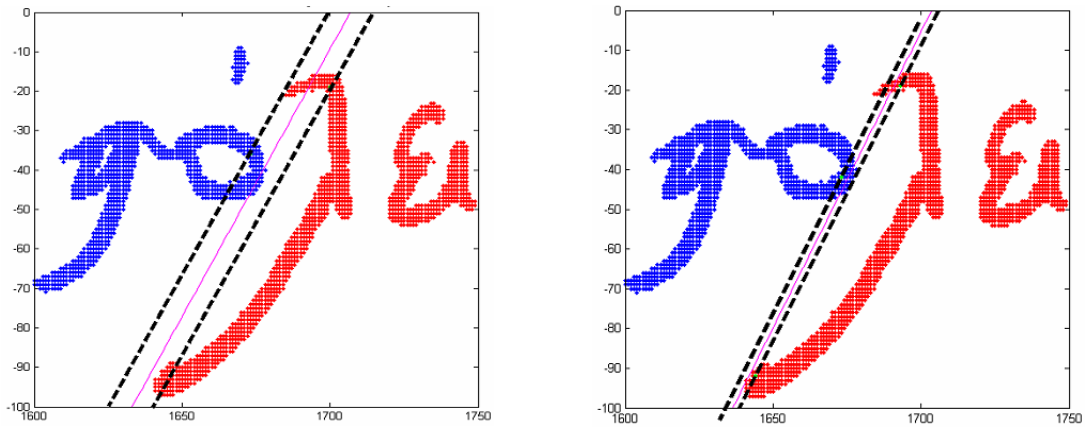
Η συγκεκριμένη σταθερά λειτουργεί αντισταθμιστικά στην επέκταση του περιθωρίου, όταν αυτή συνεπάγεται την εσφαλμένη ταξινόμηση κάποιου σημείου. Επομένως, οι μικρές τιμές του C επιτρέπουν την επέκταση του περιθωρίου μια και δεν επιβάλλουν αυστηρή ποινή στις εσφαλμένες ταξινομήσεις. Αν λοιπόν, για το συγκεκριμένο πρόβλημα ποσοτικοποίησης των κενών μεταξύ διαδοχικών χαρακτήρων, επιλέξουμε τέτοιες τιμές, τότε η εκτίμηση της απόστασης δε θα ανταποκρίνεται στην πραγματικότητα (σχ. 2.11α). Αντίστροφα, αν επιλεγούν πολύ μεγάλες τιμές για το C, τότε επί της ουσίας θα έχουν αποκλειστεί τα κενά μεταξύ μη γραμμικά διαχωρίσιμων χαρακτήρων, μια και η τιμή του περιθωρίου ταξινόμησης θα είναι πολύ μικρή. Από την πειραματική μελέτη προέκυψε ότι είναι ελάχιστες οι περιπτώσεις που τέτοιοι χαρακτήρες ανήκουν σε διαφορετικές λέξεις και επομένως η συμπεριφορά της τεχνικής για τιμές του $C \ge 10$ αποδείχθηκε σταθερή.

Ο αντικειμενικός στόχος δεν είναι η εύρεση του βέλτιστου ταξινομητή ($\mathbf{w}^*, w_0^*$) για κάθε ζεύγος διαδοχικών CCs, αλλά η εκτίμηση της διαχωρισιμότητά τους, δηλαδή οι αντίστοιχες τιμές είτε για το περιθώριο ταξινόμησης ($2/\|\mathbf{w}^*\|$) είτε για τη συνάρτηση κόστους (L^*). Στο σχ. 2.12 παρουσιάζονται βέλτιστες τιμές των περιθωρίων γ_i και οι «τροποποιημένες» τιμές των συναρτήσεων κόστους $-\log(L_i)$ για τα κενά g_i, $i = 1,...,N$, (N το πλήθος των κενών) που

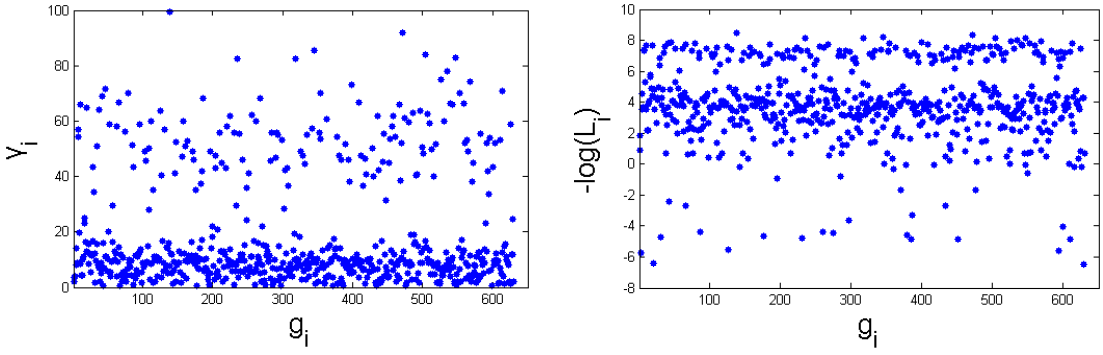
ποσοτικοποιήθηκαν στην εικόνα 004.tif (ICDAR-07). Η εισαγωγή του λογάριθμου γίνεται για τη μείωση του εύρους των τιμών των συναρτήσεων κόστους και η χρήση του αρνητικού πρόσημου, ώστε οι μεγαλύτερες τιμές να αντιστοιχούν σε μεγαλύτερες τιμές περιθωρίων και αντίστροφα. Η προτεινόμενη ποσότητα για την ποσοτικοποίηση των κενών ορίστηκε ως:

$$g_i = -\log\{L_i\} \qquad \text{(Εξ. 2.18)}$$

Η επιλογή της οφείλεται κυρίως στο γεγονός ότι συντελεί στη συγκρότηση μιας συμπαγούς τάξης για τις ποσότητες (μεγαλύτερες τιμές) που πιθανότατα αντιστοιχούν σε κενά ΜΛ, όπως θα εξηγηθεί στην ενότητα 2.2.2. Ένας πρόσθετος λόγος είναι ότι η τιμή της L_i προκύπτει άμεσα από την επίλυση του προβλήματος βελτιστοποίησης.



Σχήμα 2.11. Η εκτίμηση του κενού για C=0.01 και C=10 αντίστοιχα.



Σχήμα 2.12. Οι τιμές των περιθωρίων γ_i και των συναρτήσεων κόστους $-\log(L_i)$ για τα αντίστοιχα κενά g_i (004.tif-ICDAR07).

2.2.2. Κατηγοριοποίηση κενών

Ο υπολογισμός της προτεινόμενης ποσότητας γίνεται για κάθε ζεύγος διαδοχικών CCs σε κάθε γραμμή κειμένου. Με τον τρόπο αυτό έχουν ποσοτικοποιηθεί όλα τα κενά του κειμένου και απομένει η ταξινόμησή τους σε κενά ΜΛ και ΕΛ. Υποθέτοντας ότι οι αποστάσεις μεταξύ των λέξεων αντιστοιχούν σε μεγαλύτερες ποσότητες από αυτές μεταξύ των χαρακτήρων της ίδιας λέξης, η κατηγοριοποίηση των κενών μπορεί να επιτευχθεί με τη χρήση κάποιου κατωφλίου. Η ποικιλομορφία των χειρόγραφων κειμένων, καθιστά την επιλογή ενός καθολικού κατωφλίου για τη χρήση του σε κάθε χειρόγραφο πολύ δύσκολη και πιθανότατα αναποτελεσματική.

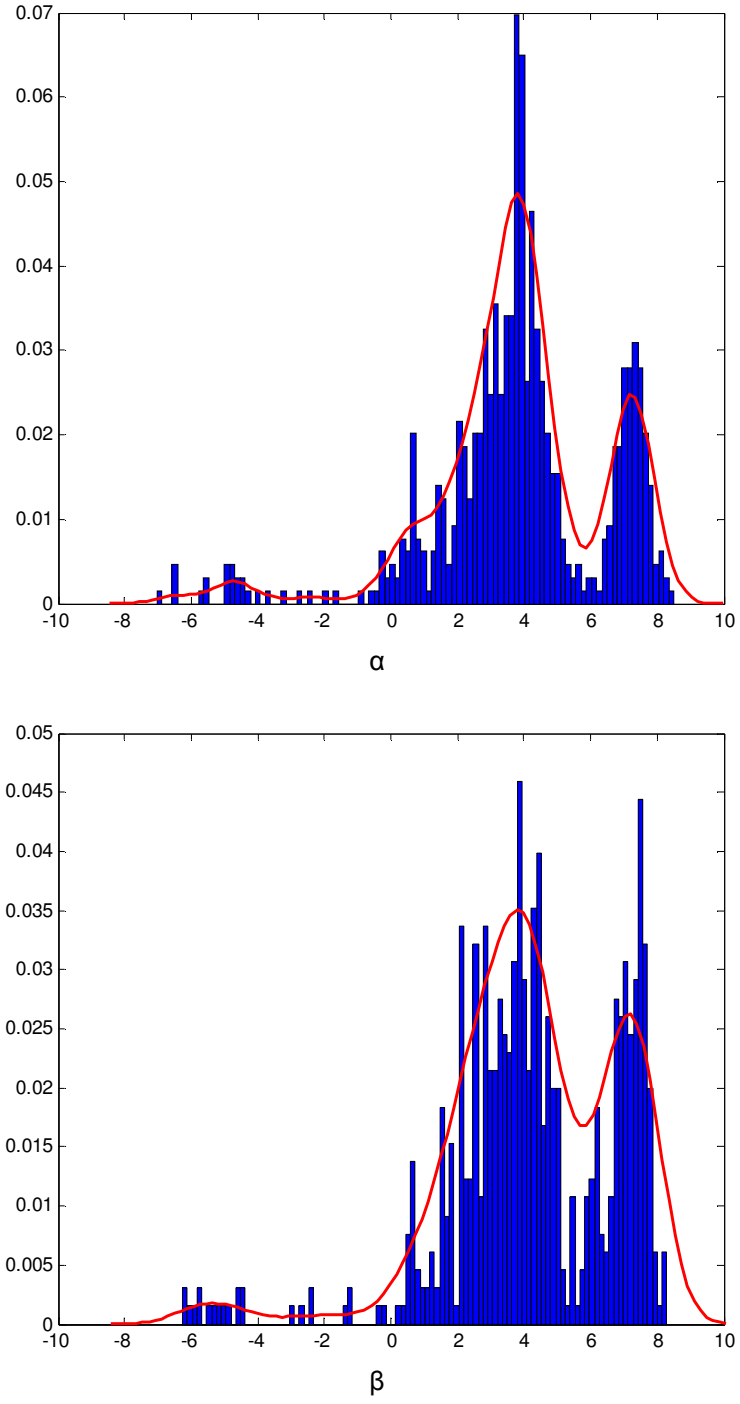
Για το λόγο αυτό προτείνεται ο υπολογισμός του κατωφλίου να γίνεται εκ νέου για κάθε κείμενο με βάση τα δεδομένα που συλλέγονται κατά την εκτίμηση των κενών. Έστω, $G = \{g_k\}_{n=1}^{N}$ το σύνολο των N ποσοτήτων για όλη την εικόνα κειμένου. Στο ιστόγραμμα (100 bins) του G παρατηρούμε ότι εμφανίζεται μια σχετικά συμπαγής ομάδα στις υψηλές τιμές η οποία πιθανότατα αντιστοιχεί στα ΜΛ κενά (σχ.2.13). Από την αντίστοιχη συνάρτηση πυκνότητας πιθανότητας, συμπεραίνει κανείς ότι παρουσιάζει ένα στενό λοβό στις μεγάλες τιμές και αρκετούς για τις μικρότερες τιμές. Αυτό είναι λογικό μια και κατά τη γραφή οι «αποστάσεις» μεταξύ των χαρακτήρων μπορεί να διαφοροποιούνται σημαντικά, ενώ αυτές μεταξύ των λέξεων δεν εμφανίζουν μεγάλη διασπορά. Επομένως, η επιλογή ενός κατωφλίου που θα διαχωρίζει τις υψηλές τιμές από τις υπόλοιπες, μπορεί να αποτελέσει το κριτήριο για την ταξινόμηση των αντίστοιχων κενών. Από το σχ. 2.12β προκύπτει ότι μια κατάλληλη τιμή για το κατώφλι διαχωρισμού θα ήταν η τιμή 6. Από τη συνάρτηση πυκνότητας πιθανότητας (σχ.2.13α) συμπεραίνουμε ότι η ίδια τιμή αντιστοιχεί στο δεξιότερο τοπικό ελάχιστο.

Για την εκτίμηση μιας ομαλής συνάρτησης πυκνότητας πιθανότητας (σ.π.π.) χρησιμοποιείται η μέθοδος παραθύρωσης Parzen [63] και περιγράφεται από την ακόλουθη εξίσωση:

$$\hat{p}(x) = \frac{1}{Nh}\sum_{t=1}^{N} K\left(\frac{x - x^t}{h}\right) \qquad \text{(Εξ. 2.19)}$$

όπου η τυχαία μεταβλητή x ορίζει τη θέση στην οποία θα υπολογιστεί η $\hat{p}(x)$, N το πλήθος των δειγμάτων (ποσοτήτων για τα κενά), x^t οι τιμές των δειγμάτων στην περιοχή πλάτους h γύρω από την τυχαία μεταβλητή x και $K(u) = (1/\sqrt{2\pi})\exp(-u^2/2)$ η κανονική συνάρτηση πυρήνα που καθορίζει την επιρροή των ποσοτήτων x^t στην τιμή της $\hat{p}(x)$. Για τον υπολογισμό της βέλτιστης τιμής του πλάτους του παραθύρου h (width of smoothing window) υιοθετήθηκε η σχέση που περιγράφεται στο [64], όπου αποδεικνύεται ότι για τον υπολογισμό της σ.π.π. με τη χρήση κανονικής συνάρτησης πυρήνα η βέλτιστη τιμή του είναι $h^* = \{median[abs(DATA) - median(DATA)]/0.6745\} \cdot [4/(3N)]^{1/5}$, με $DATA$ να συμβολίζει τις τιμές των N συγκεντρωμένων δειγμάτων. Επίσης, αναφέρεται ότι ως πρώτο

(τελευταίο) σημείο υπολογισμού της σ.π.π. επιλέγεται το $\min(DATA) - 3 \cdot h^*$ ($\max(DATA) + 3 \cdot h^*$) και τα υπόλοιπα επιλέγονται σε ισαπέχουσες θέσεις.



Σχήμα 2.13. Τα ιστογράμματα και οι συναρτήσεις πυκνότητας πιθανότητας για τις εικόνες (α) 004.tif και (β) 007.tif (ICDAR-07).

2.2.3. Αξιολόγηση

Η προτεινόμενη μέθοδος (ILSP-LWSeg) υποβλήθηκε προς αξιολόγηση στους διαγωνισμούς κατάτμησης χειρόγραφου κειμένου (Handwriting Segmentation Contests) που διεξήχθησαν στα πλαίσια των International Conference on Document Analysis and Recognition 2007 και 2009. Τα σετ εξέτασης των διαγωνισμών αποτελούνται από 80 και 200 δυαδικές εικόνες χειρόγραφου κειμένου και το συνολικό πλήθος των λέξεων είναι 13 307 και 29 717 για κάθε σετ αντίστοιχα. Τα χαρακτηριστικά των σετ και ο τρόπος αξιολόγησης των αλγορίθμων που συμμετείχαν, περιγράφονται στην ενότητα 1.2.7. Τα συγκριτικά αποτελέσματα παρουσιάζονται στους πίνακες 2.1 και 2.2.

Στο [7] περιέχονται αναλυτικές πληροφορίες για την οργάνωση του διαγωνισμού καθώς και σύντομες περιγραφές των αλγορίθμων που συμμετείχαν. Επιγραμματικά, αναφέρεται ότι όλες οι μέθοδοι προϋποθέτουν την κατάτμηση του κειμένου σε γραμμές κειμένου. Σε όλες τις μεθόδους, εκτός της προτεινόμενης, η εκτίμηση των κενών μεταξύ διαδοχικών CCs, γίνεται με τη χρήση είτε της Ευκλείδειας απόστασης είτε της ελάχιστης οριζόντιας απόστασης. Η ταξινόμηση των κενών σε διαχωριστικά λέξεων και μη, γίνεται με τη χρήση κατωφλίου, που υπολογίζεται από τα δεδομένα της εξεταζόμενης εικόνας κειμένου (π.χ. μέση τιμή στην DUTH-ARLSA, άθροισμα μέσης τιμής και τυπικής απόκλισης στην BESUS). Στη μέθοδο PARC, η κατηγοριοποίηση γίνεται από έναν εκπαιδευμένο ταξινομητή (δυστυχώς, δεν αναφέρονται περισσότερες πληροφορίες). Από τη σύγκριση των αποτελεσμάτων προκύπτει ότι η προτεινόμενη μέθοδος ανταποκρίνεται με σχετική επιτυχία στο πρόβλημα. Βέβαια, πρέπει να ληφθεί υπόψη ότι οι πιθανές αστοχίες στην κατάτμηση του κειμένου σε γραμμές ($1^{ο}$ στάδιο διαγωνισμού, βλ. Ενότητα 1.2.7) επηρεάζουν την απόδοση των τεχνικών για τον εντοπισμό των λέξεων.

Πίνακας 2.1. Συγκριτικά αποτελέσματα (ICDAR 2007)

	M	**o2o**	**g_o2m**	**g_m2o**	**d_o2m**	**d_o2m**	**DR (%)**	**RA (%)**	**FM (%)**
BESUS	19 091	9 114	327	6 172	2 449	823	80.7	52.0	63.3
DUTH-ARLSA	16 620	9 100	394	5 896	2 440	954	80.2	61.3	69.5
ILSP-LWSeg	**13 027**	**11 732**	**303**	**834**	**378**	**819**	**90.3**	**92.4**	**91.3**
PARC	14 965	10 246	422	3 482	1 524	1 088	84.3	72.8	78.1
UoA-HT	13 824	11 794	263	1 418	668	602	91.7	87.6	89.6

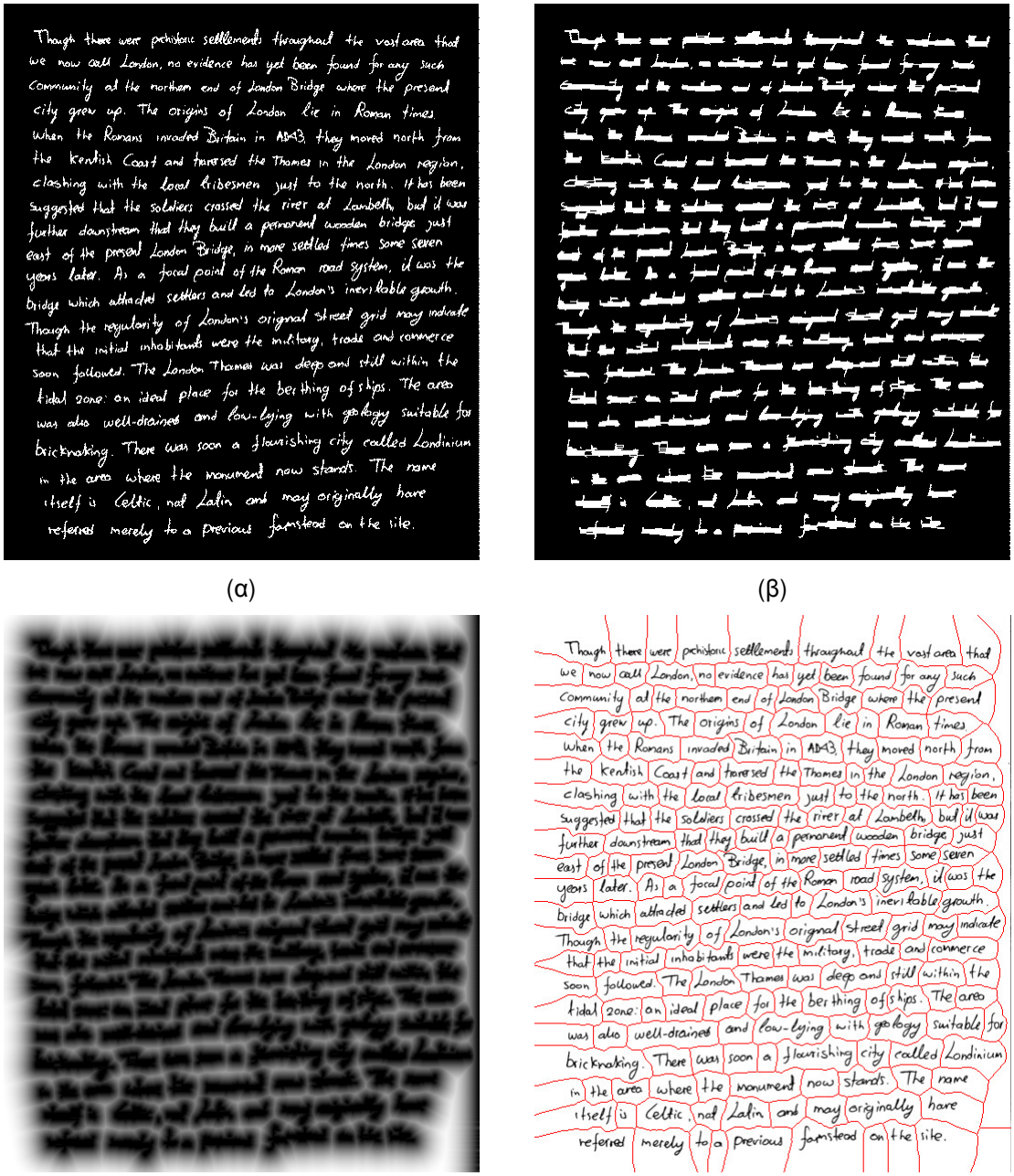
Πίνακας 2.2. Συγκριτικά αποτελέσματα (ICDAR 2009) [38]

	M	**o_g2o_d**	**DR(%)**	**RA(%)**	**FM(%)**
CASIA-MSTSeg	31 421	25 938	87.28	82.55	84.55
CMM	31 197	27 078	91.12	86.80	88.91
CUBS	31 533	26 631	89.62	84.45	86.96
ETS	30 848	26 720	86.55	83.38	84.93
ILSP-LWSeg-09	**29 962**	**28 279**	**95.16**	**94.38**	**94.77**
Jadavpur Univ	27 596	23 710	79.79	85.92	82.74
LRDE	33 006	26 318	88.56	79.74	83.92
PAIS	30 560	27 288	91.83	89.29	90.54

Στο [38] περιέχονται αναλυτικές πληροφορίες για την οργάνωση του διαγωνισμού καθώς και σύντομες περιγραφές των αλγορίθμων που συμμετείχαν. Επιγραμματικά, αναφέρεται ότι η CASIA εξάγει 11 γεωμετρικά χαρακτηριστικά για κάθε ζεύγος διαδοχικών CCs (οριζόντια και κατακόρυφη απόσταση των κέντρων μάζας, οριζόντια και κατακόρυφη επικάλυψη, κ.λπ.) σε κάθε γραμμή κειμένου και κατηγοριοποιεί τα κενά με βάση την απόφαση ενός ταξινομητή SVM (χωρίς να δίνονται περισσότερες λεπτομέρειες). Στην CMM χρησιμοποιείται η απόσταση BBD για την εκτίμηση των κενών και ένα προκαθορισμένο κατώφλι για την κατηγοριοποίησή τους. Η CUBS αρχικά δημιουργεί μια εικόνα *f* (buffer image) και αποθέτει σε κάθε pixel *(x,y)* της εικόνας τιμή *f(x,y)* ίση με την οριζόντια απόσταση των κοντινότερων pixel κειμένου που βρίσκονται εκατέρωθεν του pixel *(x,y)* (horizontal background run-length). Στη συνέχεια, επιλέγει τα pixels με τιμή μεγαλύτερη από ένα προκαθορισμένο κατώφλι ως πιθανές θέσεις διαχωριστικών μεταξύ των λέξεων (άρα και τα ζεύγη των CCs). Τέλος, υπολογίζει την απόσταση των ελάχιστων κυρτών πολυγώνων μεταξύ των CCs και τις ταξινομεί ως ΜΛ και ΕΛ χρησιμοποιώντας ως κατώφλι τη μέση τιμή των αποστάσεων αυξημένη κατά την τυπική απόκλισή τους Η ETS είναι παρόμοια με τη μέθοδο που περιγράφεται στην ενότητα 2.1 [57]. Η Jadavpur εξετάζει τις διαστάσεις κάθε CC και αποφαίνεται για το είδος της γραφής που υιοθετεί ο γραφέας. Ανάλογα με το είδος γραφής χρησιμοποιείται το προκαθορισμένο κατώφλι για την κατηγοριοποίηση των BBD αποστάσεων. Η PAIS αποθέτει σε κάθε pixel μιας νέας εικόνας το πλήθος των

μεταβάσεων από φόντο σε κείμενο και αντίστροφα που πραγματώνονται γύρω από αυτό (σε ένα ορθογώνιο παράθυρο ανάλυσης). Με ολική κατωφλίωση [25] της νέας εικόνας προκύπτουν τα CCs που αντιστοιχούν στις λέξεις της εξεταζόμενης γραμμής κειμένου.

Η LRDE [65] είναι η μοναδική από τις συμμετέχουσες μεθόδους που δεν προϋποθέτει την κατάτμηση της εικόνας κειμένου σε γραμμές. Αρχικά, η δυαδική εικόνα υποδειγματοληπτείται (1:4) και υπολογίζεται το συμπλήρωμά της (σχ. 2.14α). Στη συνέχεια, εφαρμόζεται μορφολογικό κλείσιμο (closing) με ορθογώνιο στοιχείο (SE) για να ενοποιηθούν τα κοντινά αντικείμενα και να συγκροτηθούν μεγαλύτερα που αντιστοιχούν σε λέξεις. Ακολούθως, υπολογίζεται ο μετασχηματισμός απόστασης ώστε να προκύψει μια νέα εικόνα σε κάθε pixel της οποίας αποδίδεται τιμή ίση με την απόστασή του από το κοντινότερο pixel με μη μηδενική τιμή (εμφανίζονται με άσπρο χρώμα στο σχ. 2.14β). Στη νέα εικόνα (σχ. 2.14γ) εντοπίζονται οι σκούρες συνεκτικές περιοχές (regional minima) που αποτελούνται από λιγότερα από *λ* pixels και απομακρύνονται (area closing, [36]). Τέλος, εφαρμόζεται ο μετασχηματισμός watershed [51] που ορίζει τα διαχωριστικά μεταξύ των λέξεων (σχ. 2.14δ). Σημειώνεται ότι αν κατά το δεύτερο βήμα επεξεργασίας (δυαδικός τελεστής κλεισίματος) δύο λέξεις ενοποιηθούν, τότε το σφάλμα θα διατηρηθεί.



(α) (β)

(γ) (δ)

Σχήμα 2.14. Κατάτμηση χειρόγραφου κειμένου σε λέξεις [65]. (α) Η αρχική εικόνα. (β) Η εικόνα μετά την εφαρμογή του τελεστή κλεισίματος. (γ) Το αποτέλεσμα του μετασχηματισμού απόστασης. (δ) Το αποτέλεσμα του μετασχηματισμού watershed.

2.2.4. Συμπεράσματα

Από την ανάλυση των αποτελεσμάτων της προτεινόμενης μεθόδου προκύπτουν δύο βασικά συμπεράσματα:

Α) Η αποτελεσματικότητα της μεθόδου επηρεάζεται άμεσα, όπως ήταν φυσικό, από την απόδοση της μεθόδου διαχωρισμού του κειμένου σε γραμμές. Ένα παράδειγμα που αποτυπώνει αυτή τη σχέση, παρουσιάζεται στο σχ. 2.15. Ένα μεγάλο τμήμα του «of» και τα δύο

ανώτερα τμήματα του χαρακτήρα «n» ανατέθηκαν στην ίδια γραμμή και τελικά στην ίδια λέξη, μια και τα μεταξύ τους κενά κατηγοριοποιήθηκαν ως ΕΛ. Αντίθετα, τα δύο κατακόρυφα τμήματα του χαρακτήρα «n» ανατέθηκαν στην ίδια γραμμή κειμένου αλλά σε διαφορετικές λέξεις, αφού το μεταξύ τους κενό ταξινομήθηκε ως ΜΛ.

Β) Όπως έχει αναφερθεί η επιλογή του κατωφλίου κατηγοριοποίησης των κενών γίνεται με την επεξεργασία του συνόλου των ποσοτήτων που έχουν υπολογιστεί κατά την εκτίμηση των κενών. Με τον τρόπο αυτό καθίσταται διαθέσιμη η μέγιστη πληροφορία που έχει αντληθεί από το κείμενο. Επομένως, είναι λογικό να υποθέσει κανείς πως η συνάρτηση πυκνότητας πιθανότητας των ποσοτήτων «περιγράφει» τις επιλογές του γραφέα σχετικά με τις αποστάσεις των χαρακτήρων κατά τη σύνταξη του κειμένου. Με την υπόθεση αυτή υπονοείται πως ο γραφέας ακολουθεί μια συγκεκριμένη τακτική γραφής σε όλη την έκταση του κειμένου. Συχνά όμως οι γραφείς διαφοροποιούν τις επιλογές τους για να αποδώσουν ιδιαίτερη σημασία σε κάποιο τμήμα του κειμένου (π.χ. ο τίτλος είναι συνήθως αραιογραμμένος). Σε αυτές τις περιπτώσεις η προτεινόμενη τεχνική αδυνατεί να εντοπίσει τη διαφοροποίηση και συνήθως κατακερματίζει τον τίτλο (σχ.2.16).

Σχήμα 2.15. α) Τμήμα της εικόνας 037.tif από ICDAR07, όπως έχει χωριστεί σε γραμμές. β) Το ίδιο τμήμα της εικόνας, όπως έχει χωριστεί σε λέξεις.

Σχήμα 2.16. Εσφαλμένη κατάτμηση κειμένου σε λέξεις (τμήμα της εικόνας 049.tif ICDAR-07)

Κεφάλαιο 3. Εντοπισμός κειμένου σε βίντεο

Στα προηγούμενα κεφάλαια παρουσιάστηκαν τεχνικές για την κατάτμηση δυαδικών εικόνων χειρόγραφου κειμένου σε γραμμές κειμένου και λέξεις. Η βασική υπόθεση που συνοδεύει την προηγούμενη ανάλυση είναι ότι η εικόνα περιέχει μόνο κειμενικά στοιχεία. Υπάρχουν όμως και πολλά παραγόμενα ψηφιακά αρχεία που περιλαμβάνουν και μη κειμενικά στοιχεία όπως γραμμές, σχήματα, σκίτσα κ.λπ. Η συνύπαρξη κειμενικών και μη στοιχείων, δημιουργεί την ανάγκη για ανάπτυξη μεθόδων εντοπισμού και εξαγωγής του κειμένου [66].

Η εφαρμογή των τεχνικών εντοπισμού κειμένου έχει επεκταθεί τα τελευταία χρόνια και στην επεξεργασία πολυμεσικών αρχείων, όπου το κείμενο δεν είναι το κυρίαρχο δομικό στοιχείο, αν και αποτελεί βασικό φορέα πληροφορίας. Πράγματι, η δεικτοδότηση της αυξανόμενης ποσότητας παραγόμενων πολυμεσικών αρχείων, είναι δυνατό να επιτευχθεί με την εξαγωγή και αναγνώριση του κειμένου που πιθανώς εμφανίζεται σε αυτά. Ένα παράδειγμα είναι η επεξεργασία του βίντεο ενός δελτίου ειδήσεων και η εξαγωγή της θεματολογίας του με βάση το κείμενο (π.χ. τίτλοι θεμάτων) που εμφανίζονται κατά τη μετάδοσή του [67].

Στο συγκεκριμένο κεφάλαιο παρουσιάζεται ένας τρόπος εντοπισμού κειμένου σε πλαίσια βίντεο. Η διαδικασία εντοπισμού συνοδεύεται από ένα στάδιο επαλήθευσης των εντοπισμένων περιοχών ώστε να μειωθούν οι πιθανές αστοχίες. Τα δύο αυτά στάδια συνιστούν μια βαθμίδα του ΠΑΝΟΠΤΗ [68], ενός συστήματος που λειτουργεί στο Εθνικό Συμβούλιο Ραδιοτηλεόρασης για την καταγραφή και τη δεικτοδότηση της μεταδιδόμενης πληροφορίας. Ο ΠΑΝΟΠΤΗΣ καταγράφει τις ειδησεογραφικές και ενημερωτικές εκπομπές λόγου που μεταδίδονται από 10 επιλεγμένους ελληνικούς τηλεοπτικούς σταθμούς και ενσωματώνει σύγχρονες γλωσσικές τεχνολογίες όπως η αναγνώριση φωνής, η αναγνώριση ομιλητή και ο εντοπισμός και η οπτική αναγνώριση χαρακτήρων [69].

3.1. Γενικές αρχές

Οι κειμενικές πληροφορίες που εμφανίζονται σε ένα τηλεοπτικό πρόγραμμα, διακρίνονται σε αυτές που έχουν προστεθεί στο βίντεο (overlay ή artificial text) και σε αυτές που αποτελούν τμήμα της προβαλλόμενης σκηνής (scene text). Οι συνήθεις πραγματώσεις των πρόσθετων κειμένων είναι οι τίτλοι των ειδησεογραφικών θεμάτων, τα ονόματα και οι ιδιότητες των ομιλητών και οι επωνυμίες των εταιριών και των προϊόντων τους σε διαφημίσεις. Επομένως, το πρόσθετο κείμενο είναι σημαντικός φορέας πληροφορίας και η εμφάνισή του είναι προσεκτικά σχεδιασμένη ώστε να το καθιστά ευανάγνωστο (σχ. 3.1). Οι βασικοί παράγοντες που καθιστούν το κείμενο εύληπτο, είναι το μέγεθος, ο χρωματισμός, ο προσανατολισμός και η διάρκεια εμφάνισης [70].

Τα κυριότερα χαρακτηριστικά του πρόσθετου κειμένου που εμφανίζεται σε εκπομπές ειδησεογραφικού περιεχομένου είναι:

α) το ύψος των χαρακτήρων περιορίζεται μεταξύ συγκεκριμένων τιμών,

β) οι χαρακτήρες κάθε μιας κειμενικής πραγμάτωσης είναι σχεδόν «μονόχρωμοι»,

γ) ο προσανατολισμός είναι συνήθως οριζόντιος,

δ) η φωτεινότητα των χαρακτήρων βρίσκεται σε έντονη αντίθεση με το φόντο και

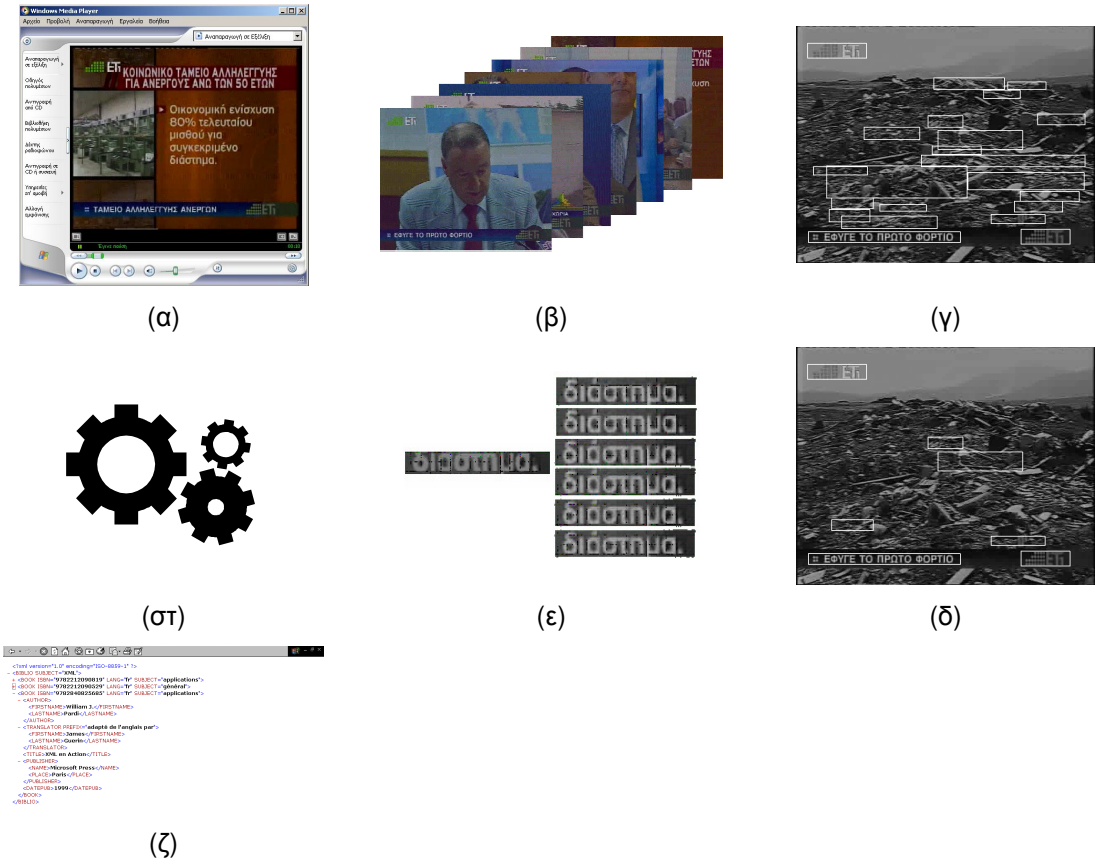
ε) η εμφάνισή του διαρκεί για εύλογο χρονικό διάστημα (π.χ. τουλάχιστον 1 sec) και επομένως περιέχεται σε αρκετά διαδοχικά πλαίσια του βίντεο.

Σχήμα 3.1. Παραδείγματα πλαισίων βίντεο που περιέχουν κείμενο. Τα εμφανιζόμενα κείμενα διαφοροποιούνται ως προς το μέγεθος, το χρώμα και τον προσανατολισμό.

Η συνήθης δομή των συστημάτων εντοπισμού και αναγνώρισης κειμένου σε βίντεο παρουσιάζεται στο σχ. 3.2. Θεωρώντας το βίντεο ως μια ακολουθία εικόνων, το πρώτο στάδιο είναι η επιλογή των πλαισίων που αντιπροσωπεύουν το βίντεο, χωρίς την απώλεια κρίσιμης πληροφορίας (σχ. 3.2β). Η επεξεργασία κάθε πλαισίου παρέχει τη δυνατότητα της πλέον διεξοδικής ανάλυσης του βίντεο, αλλά ο όγκος της εισερχόμενης πληροφορίας είναι τόσο μεγάλος που την καθιστά εξαιρετικά χρονοβόρα. Τα κριτήρια της επιλογής είναι το είδος κωδικοποίησης του βίντεο και η ελάχιστη χρονική διάρκεια εμφάνισης του κειμένου. Για παράδειγμα, αν η ελάχιστη διάρκεια εμφάνισης ενός σημαντικού κειμένου έχει οριστεί ίση με 0.8 sec και η είσοδος είναι ένα αρχείο με 25fps, τότε ο ενδεδειγμένος ρυθμός δειγματοληψίας είναι 1:10, ώστε να εξασφαλίζεται η εμφάνιση κάθε «κρίσιμου» πρόσθετου κειμένου σε τουλάχιστον δύο επιλεγμένα πλαίσια.

Το επόμενο στάδιο είναι ο εντοπισμός των περιοχών κειμένου σε κάθε ένα από τα επιλεγμένα πλαίσια (text localization). Σε αυτή τη διαδικασία τα πλαίσια θεωρούνται ανεξάρτητες

εικόνες και επομένως η απόκρισή της μπορεί να χαρακτηριστεί ως η «στατική» πληροφορία του βίντεο. Οι αστοχίες της προηγούμενης διαδικασίας θα μειωθούν στο στάδιο επαλήθευσης των εντοπισμένων περιοχών (text verification). Ακολούθως, οι υποψήφιες περιοχές κειμένου που εντοπίστηκαν σε διαδοχικά πλαίσια συγκρίνονται ως προς τη θέση και το περιεχόμενό τους, με στόχο αφενός τη δημιουργία ενοποιημένων εικόνων μεγαλύτερης ευκρίνειας για κάθε εντοπισμένο κείμενο και αφετέρου την απομάκρυνση περιοχών που εμφανίστηκαν μόνο σε ένα πλαίσιο και πιθανότατα δεν περιέχουν κρίσιμη πληροφορία (temporal redundancy). Το στάδιο αυτό αναδεικνύει τη «δυναμική» πληροφορία του βίντεο. Τέλος, οι εικόνες κειμένου οδηγούνται σε μια μηχανή οπτικής αναγνώρισης χαρακτήρων. Το αναγνωρισμένο κείμενο και τα στοιχεία των εντοπισμένων εικόνων κειμένου (χρονικές στιγμές και θέσεις εμφάνισης) συνδυάζονται για να δημιουργηθεί το κατάλληλο αρχείο (XML μορφότυπο) που δεικτοδοτεί το βίντεο.



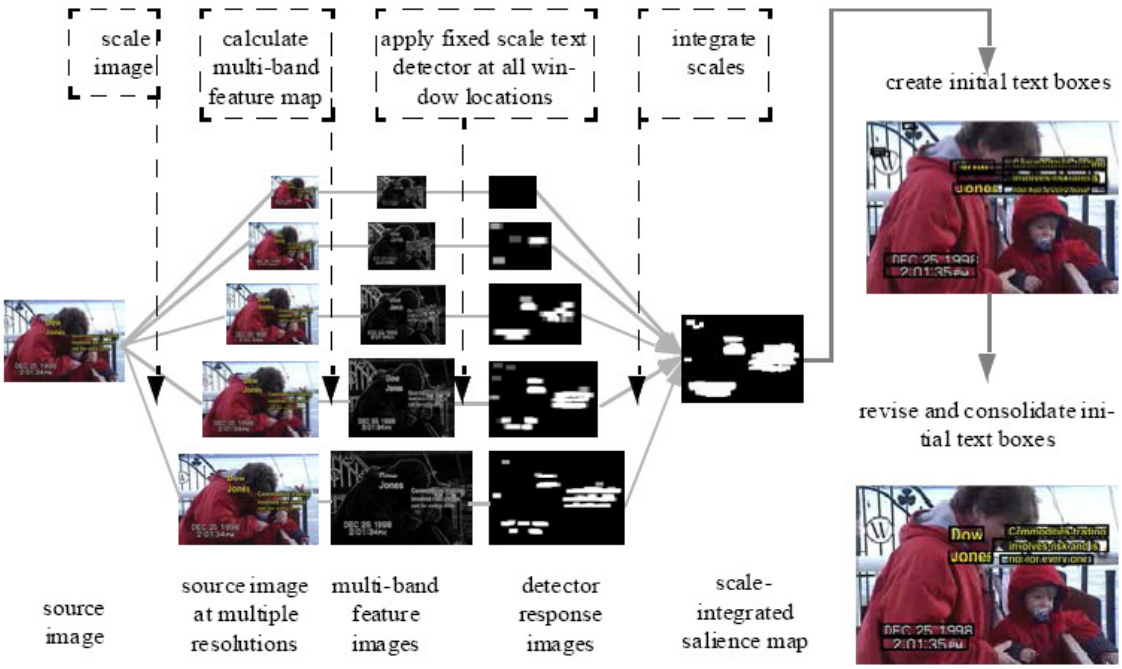
Σχήμα 3.2. Η δομή ενός συστήματος δεικτοδότησης βίντεο με βάση το εμφανιζόμενο κείμενο. (α) Το βίντεο εισόδου. (β) Τα επιλεγμένα πλαίσια. (γ) Το αποτέλεσμα του σταδίου εντοπισμού κειμένου. (δ) Το αποτέλεσμα του σταδίου επαλήθευσης. (ε) «Δυναμική πληροφορία». (στ) Μηχανή οπτικής αναγνώρισης χαρακτήρων. (ζ) Μορφότυπο XML.

3.2. Εντοπισμός κειμένου

Για τον εντοπισμό του κειμένου σε πλαίσια βίντεο έχουν προταθεί διάφορες τεχνικές που βασίζονται σε κάποιες από τις ιδιότητες του κειμένου. Υποθέτοντας ότι τα εικονοστοιχεία των

χαρακτήρων του κειμένου έχουν κοντινές τιμές φωτεινότητας, σε κάποιες τεχνικές εντοπίζονται αρχικά τα αντίστοιχα συνεκτικά αντικείμενα [71]. Στη συνέχεια εξετάζονται τα γεωμετρικά χαρακτηριστικά τους και είτε απορρίπτονται ως τμήματα του φόντου είτε θεωρούνται υποψήφια τμήματα περιοχών κειμένου. Τέλος, με κριτήρια τη διάταξη και την απόστασή τους αποφασίζεται η ενοποίησή τους για τη συγκρότηση περιοχών κειμένου. Μια άλλη ιδιότητα του κειμένου είναι η χρωματική αντίθεσή του με το φόντο. Σε πολλές μεθόδους [72] που βασίζονται στην ιδιότητα αυτή, εντοπίζονται οι ακμές (π.χ. με τη χρήση των βασικών μορφολογικών τελεστών ή με τον αλγόριθμο Canny [73]) και εφαρμόζονται μορφολογικοί τελεστές (closing και opening) με κατάλληλα δομικά στοιχεία για την ενοποίηση κοντινών ακμών ή την απομάκρυνση των απομονωμένων.

Οι περιοχές με πρόσθετο κείμενο χαρακτηρίζονται από ιδιαίτερη υφή λόγω της διάταξης των χαρακτήρων και της έντονης αντίθεσής τους με το φόντο. Για την «περιγραφή» της υφής έχουν προταθεί διάφορα χαρακτηριστικά όπως η μεταβλητότητα των τιμών των εικονοστοιχείων, η πυκνότητα των ακμών [74] και κυματίδια (wavelets) [75]. Οι τιμές τέτοιων χαρακτηριστικών εξαρτώνται από το μέγεθος των χαρακτήρων του κειμένου. Η κοινή πρακτική για να επιτευχθεί η ανεξαρτησία από το μέγεθος των χαρακτήρων, είναι η επεξεργασία της εικόνας σε διάφορες αναλύσεις (σχ. 3.3). Με τον τρόπο αυτό, εντοπίζονται αρχικά οι περιοχές κειμένου συγκεκριμένων διαστάσεων σε κάθε εικόνα διαφορετικής κλίμακας και τελικά προβάλλονται στην αρχική εικόνα.



Σχήμα 3.3. Εντοπισμός περιοχών πρόσθετου κειμένου [76].

Πιο αναλυτικά, σε κάθε εικόνα διαφορετικής κλίμακας εφαρμόζεται η κατάλληλη διαδικασία (π.χ. εντοπισμός ακμών) για την ανάδειξη των επιλεγμένων χαρακτηριστικών ($3^η$ στήλη στο σχ. 3.3). Στη συνέχεια εξετάζεται κάθε περιοχή της κάθε εικόνας (με τη χρήση ενός

ολισθένοντος παραθύρου του οποίου οι διαστάσεις σχετίζονται άμεσα με το μέγεθος του αναζητούμενου κειμένου) και αξιολογείται ως πιθανή περιοχή κειμένου ($4^{η}$ στήλη στο σχ. 3.3). Για την αξιολόγηση της περιοχής χρησιμοποιούνται κυρίως νευρωνικά δίκτυα [75, 76] και μηχανές διανυσμάτων υποστήριξης [77]. Οι αποκρίσεις των εκπαιδευμένων ταξινομητών είτε είναι δυαδικές (κείμενο ή όχι) και αποδίδονται μόνο στο κεντρικό pixel του τρέχοντος παραθύρου ανάλυσης [77], είτε εκφράζουν την πιθανότητα να πρόκειται για κειμενική περιοχή και αποδίδονται σε όλα τα pixels της [76]. Στο τελευταίο στάδιο, οι υποψήφιες περιοχές από κάθε κλίμακα προβάλλονται στην αρχική εικόνα, εξετάζονται η επικάλυψή τους και τα αποτελέσματα της αξιολόγησής τους και επιλέγονται οι πιο πιθανές περιοχές κειμένου.

3.2.1. Προτεινόμενη μέθοδος

Η μέθοδος που περιγράφεται ακολούθως στοχεύει στον εντοπισμό του πρόσθετου κειμένου που εμφανίζεται σε τηλεοπτικές εκπομπές ειδησεογραφικού περιεχομένου. Από την παρατήρηση των εμφανιζόμενων κειμένων προκύπτουν τα ακόλουθα συμπεράσματα: α) το μέγεθος του κειμένου ποικίλει από 10 ως 50 pixels (σε πλαίσια βίντεο διαστάσεων 720×576), β) ο προσανατολισμός του είναι οριζόντιος και γ) η φωτεινότητα των χαρακτήρων διαφέρει σημαντικά από αυτή του φόντου. Οι ιδιότητες αυτές αξιοποιούνται στα ακόλουθα βήματα επεξεργασίας:

α) Το εύρος του μεγέθους του εμφανιζόμενου κειμένου δεν είναι μεγάλο και επομένως δεν απαιτείται η αναζήτησή του σε διάφορες κλίμακες της εικόνας. Στην προτεινόμενη μέθοδο επιλέχθηκε η σμίκρυνση της αρχικής εικόνας κατά δύο φορές ($\downarrow 2$), ώστε να μειωθεί η πολυπλοκότητα των ακόλουθων βημάτων (σχ. 3.4α).

β) Στα όρια των χαρακτήρων εμφανίζονται έντονες μεταβολές της φωτεινότητας και επομένως οι περιοχές κειμένου περιέχουν πληθώρα τέτοιων μεταβολών κυρίως κατά τον οριζόντιο άξονα. Για την ανάδειξη των μεταβολών χρησιμοποιείται η οριζόντια εκδοχή του φίλτρου Prewitt (Εξ. 3.1), το οποίο εκτιμά την παράγωγο κατά τον οριζόντιο άξονα (σχ. 3.4β).

$$f_{\mathrm{Pr}ewitt} = \begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix} \qquad \text{(Εξ. 3.1)}$$

Βέβαια, η περιοχή κειμένου περιέχει και τα εικονοστοιχεία που βρίσκονται μεταξύ των εντοπισμένων ακμών. Για την ανάδειξη και αυτών των pixels χρησιμοποιείται η ακόλουθη σχέση:

$$A(x,y) = \left[\sum_{i=-\lceil S/2 \rceil}^{\lceil S/2 \rceil} \left(\frac{\partial I}{\partial x}(x+i,y) \right)^2 \right]^{1/2} \qquad \text{(Εξ. 3.2)}$$

όπου A η εικόνα των συσσωρευμένων συντελεστών της παραγώγου και I η εικόνα στην επιλεγμένη κλίμακα. Με την (Εξ. 3.2) η τιμή του κάθε pixel προκύπτει ως η συσσωρευμένη παράγωγος (accumulated gradients) S γειτονικών pixels. Με τον τρόπο αυτό δημιουργούνται

περιοχές που πιθανότατα αντιστοιχούν σε περιοχές κειμένου (σχ. 3.4γ). Η παράμετρος S εξαρτάται από την απόσταση μεταξύ των ακμών, δηλαδή από το μέγεθος των χαρακτήρων, και τέθηκε ίση με 5 μια και στη συγκεκριμένη ανάλυση της αρχικής εικόνας το ελάχιστο μέγεθος των χαρακτήρων είναι 5 pixels.

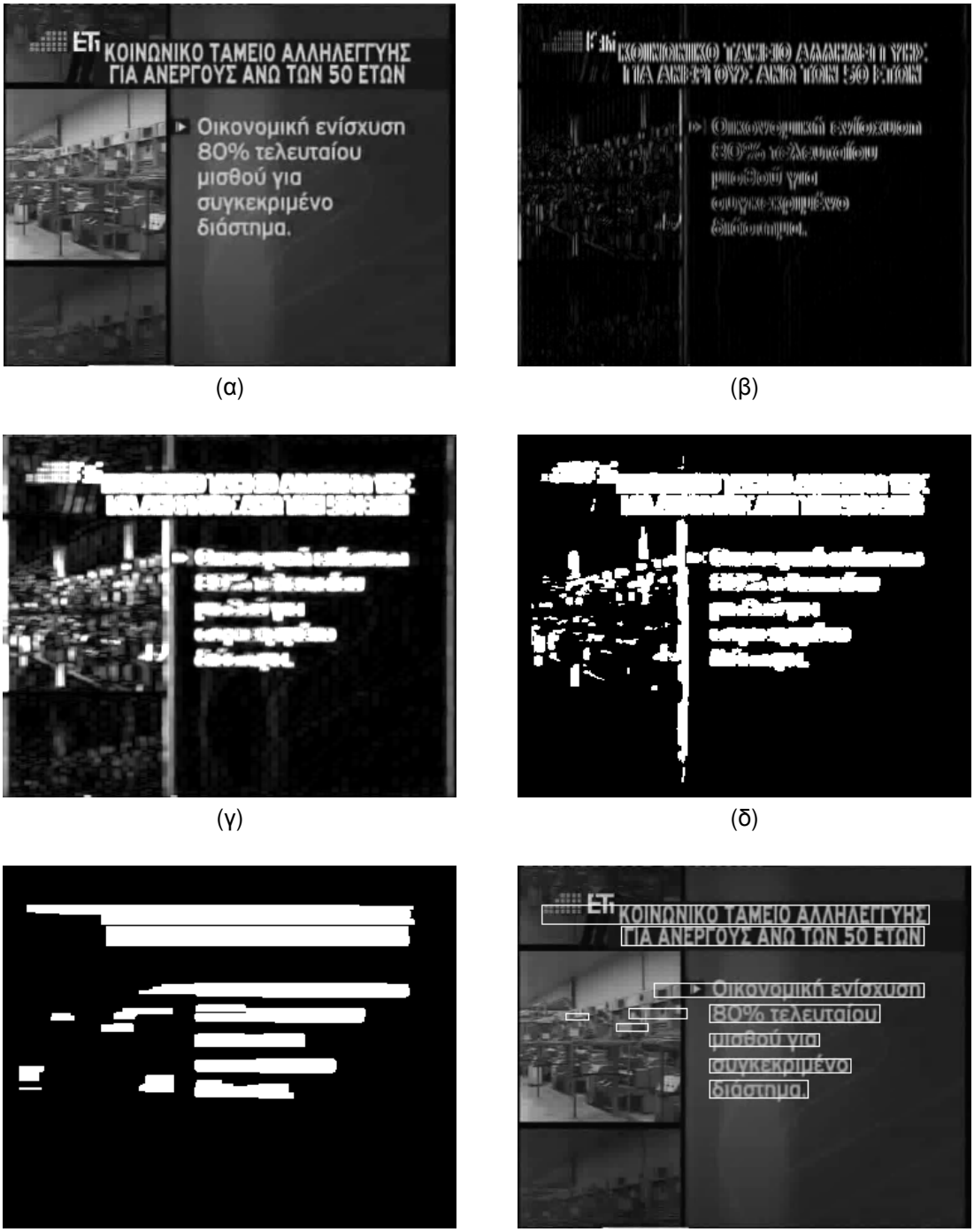
γ) Η τιμή που έχει αποδοθεί σε κάθε pixel της A δηλώνει την πιθανότητα να αντιστοιχεί το υπό εξέταση pixel σε περιοχή κειμένου. Η κατηγοριοποίηση των pixels σε pixels περιοχών κειμένου και μη, αντιστοιχεί στη μετατροπή της γκρίζας πιθανοτικής εικόνας σε δυαδική (σχ. 3.4δ) και πραγματοποιείται με την Εξ. 3.3.

$$B(x,y)=\begin{Bmatrix}1 \ , & A(x,y)\geq th\\ 0 \ , & A(x,y)<th\end{Bmatrix} \qquad \text{(Εξ. 3.3)}$$

όπου *th* είναι το κατώφλι που υπολογίζεται με τη μέθοδο ολικής κατωφλίωσης του Otsu [25].

δ) Η δυαδική εικόνα περιέχει πληθώρα αντικειμένων, τα οποία εκτείνονται σε πιθανές περιοχές κειμένου. Για την δημιουργία αντικειμένων με σχήμα αντίστοιχο με αυτό των γραμμών κειμένου (σχ. 3.4ε) εφαρμόζονται απλοί δυαδικοί μορφολογικοί μετασχηματισμοί και κανόνες :

- Εφαρμογή του τελεστή ανοίγματος με συμπαγές δομικό στοιχείο 4×4 για την απομάκρυνση αντικειμένων μικρού μεγέθους που μπορούν να θεωρηθούν μη χρήσιμα (το ελάχιστο ύψος των χαρακτήρων είναι 5 pixels)
- Εφαρμογή του τελεστή κλεισίματος με το δομικό στοιχείο $\left[\boxed{1}\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\right]$ για την επέκταση των αντικειμένων κατά 10 pixels προς τα δεξιά. Με τον τρόπο αυτό ενοποιούνται αντικείμενα που αντιστοιχούν σε λέξεις της ίδιας γραμμής κειμένου.
- Εφαρμογή του τελεστή ανοίγματος με συμπαγές δομικό στοιχείο 1×15 για την απομάκρυνση των στενών κατακόρυφων τμημάτων των αντικειμένων.
- Απομάκρυνση των τμημάτων των περιοχών που έχουν τοπικό πλάτος μικρότερο του ¼ πλάτους της περιοχής. Με τον τρόπο αυτό οι περιοχές κειμένου που περιέχουν περισσότερες από μία γραμμές κειμένου θα διαχωριστούν ώστε κάθε περιοχή να περιέχει μία γραμμή κείμενου (text line).
- Σχηματισμός ορθογωνίων πλαισίων (bounding boxes) που εγκιβωτίζουν τα εναπομείναντα CCs (σχ. 3.4στ). Κάθε ορθογώνιο με ύψος μικρότερο από 5 pixels ή με λόγο πλάτους-ύψους μικρότερο από 2 απορρίπτεται, ως περιοχή που κατά κανόνα δεν περιέχει σημαντική κειμενική πληροφορία. Επομένως, κειμενικές πραγματώσεις, όπως θερμοκρασίες σε δελτία καιρού δεν προτείνονται ως υποψήφιες περιοχές κειμένου.



Σχήμα 3.4. Εντοπισμός πρόσθετου κειμένου σε πλαίσιο βίντεο. (α) Η αρχική εικόνα. (β) Η παράγωγος κατά τον οριζόντιο άξονα. (γ) Η συσσωρευμένη παράγωγος. (δ) Η δυαδική εικόνα. (ε) Η εικόνα μετά την εφαρμογή μορφολογικών τελεστών. (στ) Οι υποψήφιες περιοχές κειμένου.

Στο σχ. 3.5 παρουσιάζεται ένα συγκριτικό παράδειγμα των αποτελεσμάτων της προτεινόμενης μεθόδου (σχ.3.5α) και της συναφούς μεθόδου [78] (3.5β), όπως προέκυψε από τη διεύθυνση http://liris.cnrs.fr/christian.wolf/demos/index.html. Από τη σύγκριση των αποτελεσμάτων συμπεραίνει κανείς πως πολλές μη κειμενικές περιοχές προτείνονται ως

πιθανές περιοχές κειμένου. Επομένως, θα πρέπει να ακολουθήσει ένα στάδιο επαλήθευσης των εντοπισμένων περιοχών ώστε να μειωθεί το πλήθος των αστοχιών (βλ. Εν. 3.3). Μια σημαντική διαφορά είναι ότι η προτεινόμενη μέθοδος οριοθετεί ξεχωριστά κάθε γραμμή κειμένου. Με τον τρόπο αυτό αφενός μειώνεται το πλήθος των pixels του φόντου και αφετέρου διευκολύνεται η διαδικασία απομάκρυνσης των αστοχιών. Σημειώνεται ότι ο μη εντοπισμός του λογότυπου οφείλεται στο γεγονός πως ο λόγος πλάτους-ύψους του αντίστοιχου ορθογωνίου είναι μικρότερος από το επιλεγμένο κατώφλι.

(α) (β)

Σχήμα 3.5. α) Αποτέλεσμα προτεινόμενης μεθόδου. β) Αποτέλεσμα της [78].

Στο σχ. 3.6 παρουσιάζονται μερικά ενδεικτικά αποτελέσματα της προτεινόμενης μεθόδου σε πλαίσια βίντεο από ελληνικά τηλεοπτικά κανάλια και σε εικόνες που ήταν διαθέσιμες μέσω του διαδικτύου από τις θέσεις http://liris.cnrs.fr/christian.wolf/demos/index.html (σχ. 3.5ε) και http://www.informatik.unimannheim.de/informatik/pi4/projects/MOCA (σχ. 3.5στ). Από την παρατήρηση των αποτελεσμάτων συμπεραίνει κανείς πως η πλειοψηφία των πρόσθετων κειμένων εντοπίζεται με επιτυχία, αλλά συχνά είτε προτείνονται και μη κειμενικές περιοχές, είτε οι περιοχές κειμένου δεν οριοθετούνται με την επιθυμητή ακρίβεια. Αξίζει να σημειωθεί ότι η προτεινόμενη μέθοδος εντοπίζει και κειμενικές πραγματώσεις που αποτελούν μέρος της προβαλλόμενης σκηνής (scene text) όταν αυτές κατά την εμφάνισή τους έχουν ιδιότητες παρόμοιες με αυτές του πρόσθετου κειμένου (σχ. 3.5γ, δ).

Η αποτελεσματικότητα της μεθόδου εξετάστηκε σε 3700 πλαίσια βίντεο από εμπορικό και ειδησεογραφικό πρόγραμμα δέκα ελληνικών τηλεοπτικών σταθμών (ET1, NET, ET3, MEGA, ANT1, STAR, ALTER, ALPHA, ΣΚΑΙ και EXTRA). Επίσης, συμπεριλήφθηκαν και 59 πλαίσια από το σετ εξέτασης της [78] που ήταν διαθέσιμα. Τα αποτελέσματα δίνονται στον πίνακα Π.3.1.

Στα 3759 πλαίσια του σετ εξέτασης περιέχονται 11952 πραγματώσεις κειμένου (Ground Truth, GT) με ποικίλο ύψος από 10 ως 50 pixels. Το πλήθος πραγματικών περιοχών κειμένου που εντοπίστηκαν (Localized Text, LT) είναι 11497. Το πλήθος των εντοπισμένων περιοχών που δεν αντιστοιχεί σε κείμενο (Localized Non-Text, LNT) είναι 37589. Ορίζοντας το ποσοστό

ορθού εντοπισμού $Recall = LT / GT$ και ως ποσοστό ακρίβειας $Precision = LT / (LT + LNT)$, προκύπτει το συμπέρασμα ότι η μέθοδος είναι μεν αποτελεσματική (recall~96%), αλλά δεν είναι ακριβής (~23%).

Πίνακας 3.1. Αξιολόγηση μεθόδου εντοπισμού κειμένου.

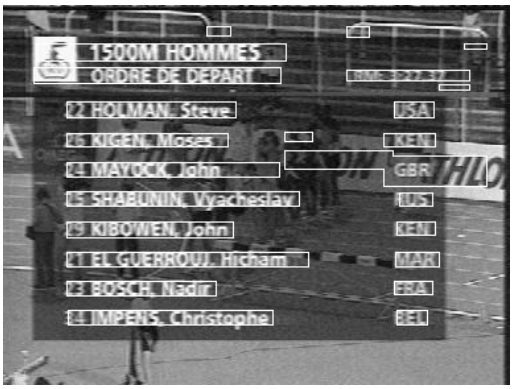
GT	LT	LNT	Recall	Precision
11952	11497	37589	0.96	0.23

(α)

(β)

(γ)

(δ)



(ε)

(στ)

Σχήμα 3.6. Παραδείγματα εντοπισμού περιοχών κειμένου.

3.3. Επαλήθευση περιοχών κειμένου

Οι αλγόριθμοι εντοπισμού κειμένου σε εικόνα και ιδιαίτερα σε πλαίσια (frames) βίντεο έχουν υψηλά ποσοστά ορθού εντοπισμού περιοχών κειμένου (real text areas, RT), αλλά σχετικά χαμηλά έως πολύ χαμηλά ποσοστά ακρίβειας, αφού προτείνουν ως υποψήφιες περιοχές κειμένου πολλά τμήματα της εικόνας που δεν περιέχουν κείμενο (non-text areas, NT). Επομένως, η «στατική πληροφορία» που θα μεταδοθεί στα επόμενα στάδια περιέχει και πολλά περιττά στοιχεία. Για τη μείωση του πλήθους των αστοχιών (false alarms), προτείνεται η εισαγωγή ενός πρόσθετου σταδίου επεξεργασίας, αυτού της επαλήθευσης των περιοχών κειμένου. Το στάδιο επαλήθευσης περιλαμβάνει συνήθως την εφαρμογή ευρετικών κανόνων για την κατηγοριοποίηση των υποψήφιων περιοχών κειμένου.

Η προσέγγιση που έχει υιοθετηθεί για τον εντοπισμό του κειμένου καθορίζει και τον τρόπο με τον οποίο θα αξιολογηθεί κάθε προτεινόμενη περιοχή στο στάδιο της επαλήθευσης. Αν εντοπίζονται περιοχές που αντιστοιχούν σε γραμμές κειμένου τότε εξετάζονται απλά γεωμετρικά χαρακτηριστικά των εγκιβωτισμένων υποψήφιων περιοχών, όπως το ύψος και ο λόγος πλάτους ύψους. Αν οι προτεινόμενες περιοχές αντιστοιχούν σε περισσότερες από μία γραμμές κειμένου, τότε οι περιοχές χωρίζονται σε μικρότερες με βάση τα τοπικά ελάχιστα της οριζόντιας προβολής της εικόνας [74, 76]. Με τον τρόπο αυτό οριοθετούνται οι γραμμές κειμένου στις «πραγματικές» περιοχές κειμένου, ενώ οι μη κειμενικές περιοχές απορρίπτονται μια και χωρίζονται σε πολλά τμήματα μικρού ύψους. Ένα άλλο κριτήριο που έχει προταθεί είναι ο λόγος του πλήθους των pixels, που έχουν ταξινομηθεί ως pixels κειμένου κατά το στάδιο εντοπισμού και συνιστούν ένα CC, προς το εμβαδό του ορθογωνίου που εγκιβωτίζει το CC [77]. Αξιοποιώντας τη διαφορά φωτεινότητας μεταξύ του κειμένου και του φόντου, (μέθοδος HWDavid) στο [79], εξετάζεται αν το χρωματικό ιστόγραμμα της υποψήφιας περιοχής περιέχει τουλάχιστον μια βαθειά κοιλάδα.

Στο [80] παρουσιάζεται αναλυτικά μια μέθοδος κατηγοριοποίησης των υποψήφιων περιοχών κειμένου σε RT ή NT. Αρχικά, οι εικόνες κανονικοποιούνται ώστε να έχουν ύψος 16 pixels, εντοπίζονται οι ακμές (Canny) και υπολογίζονται οι τιμές των παραγώγων στον οριζόντιο και κατακόρυφο άξονα (Sobel). Για την παραμετροποίηση της εικόνας εφαρμόζεται ολισθαίνον παράθυρο (16x16 με βήμα 4 pixels) και για κάθε θέση του υπολογίζονται οι τιμές των παραγώγων σε κάθε pixel και η ευκλείδεια απόσταση κάθε pixel από το κοντινότερο pixel που ανήκει σε ακμή. Επομένως, κάθε εικόνα αντιπροσωπεύεται από την αντίστοιχη ακολουθία διανυσμάτων (παραθύρων). Για την ταξινόμηση των διανυσμάτων χρησιμοποιείται μηχανή διανυσμάτων υποστήριξης (με πυρήνα RBF [59]) και η τελική απόφαση για την εικόνα προκύπτει από το συνδυασμό των αποκρίσεων του SVM.

3.3.1. Προτεινόμενη μέθοδος

Η προτεινόμενη μέθοδος στοχεύει στη μείωση του πλήθους των υποψήφιων περιοχών που δεν περιέχουν κείμενο. Η βασική ιδέα είναι να αξιοποιηθεί το γεγονός ότι οι περιοχές κειμένου διέπονται από ένα είδος κανονικότητας με την παρουσία κενού μετά από κάθε

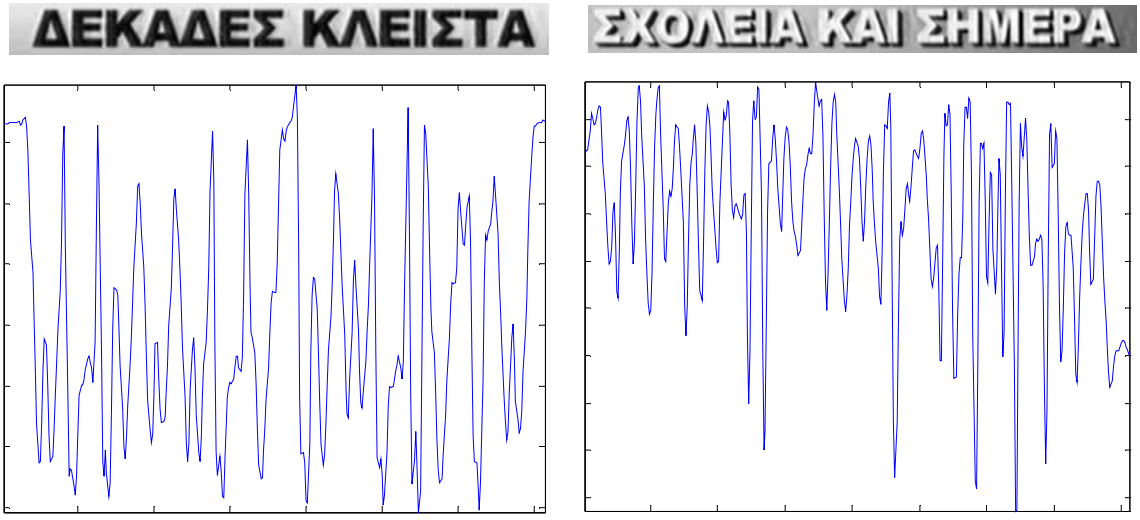
χαρακτήρα [81]. Αυτή η κανονικότητα στην εμφάνιση των κειμένων μπορεί να αποδοθεί μέσω της προβολής των εικόνων. Η προβολή P της εικόνας $\mathbf{A}_{M\times N}$ υπολογίζεται με την ακόλουθη σχέση:

$$P(i) = (1/M)\sum_{j=1}^{M}\mathbf{A}(j,i) - \bar{\mathbf{A}} \quad , \quad i = 1,...,N \qquad \text{(Εξ. 3.4)}$$

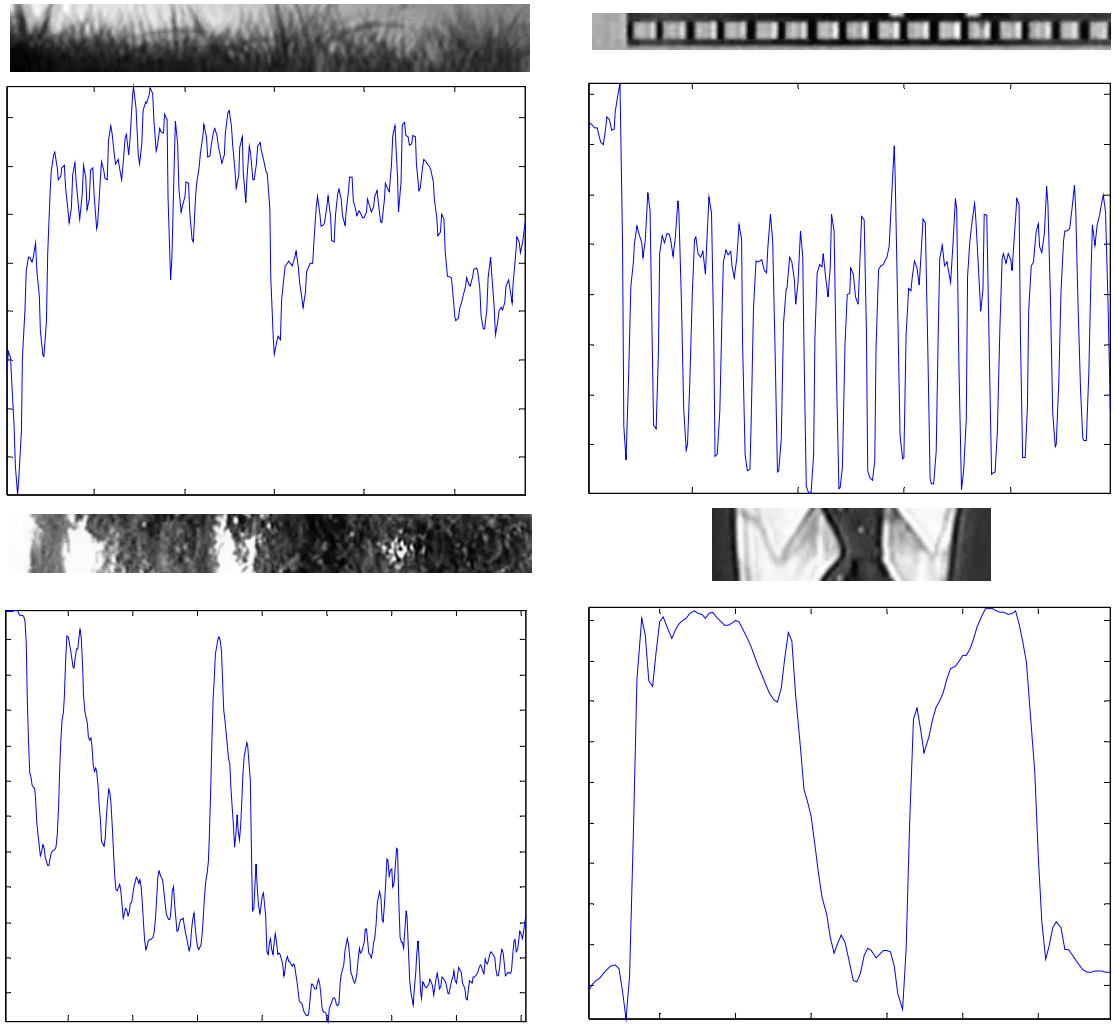
όπου $\bar{\mathbf{A}}$ η μέση τιμή των τιμών των pixels της εικόνας $\mathbf{A}$. Στην περίπτωση που η εικόνα αποτελείται από φωτεινά γράμματα και σκούρο φόντο, οι κοιλάδες της προβολής αντιστοιχούν στα κενά μεταξύ των χαρακτήρων. Το αντίστροφο ισχύει όταν εμφανίζονται σκουρόχρωμοι χαρακτήρες σε ανοιχτό φόντο (σχ. 3.7). Όπως είναι φυσικό, για τις μη κειμενικές περιοχές δεν μπορεί να δοθεί μια αντίστοιχη ερμηνεία (σχ. 3.8).

Η απόσταση μεταξύ των κενών εξαρτάται άμεσα από το μέγεθος των χαρακτήρων, άρα και από το ύψος της περιοχής κειμένου. Για να περιοριστεί αυτή η μεταβλητότητα, οι υποψήφιες περιοχές κειμένου κανονικοποιούνται ώστε να αποκτήσουν ύψος ίσο με 30 pixels. Η συγκεκριμένη τιμή επιλέχθηκε γιατί αποτελεί το πιο σύνηθες ύψος των εντοπισμένων κειμενικών περιοχών, όπως παρατηρήθηκε πειραματικά.

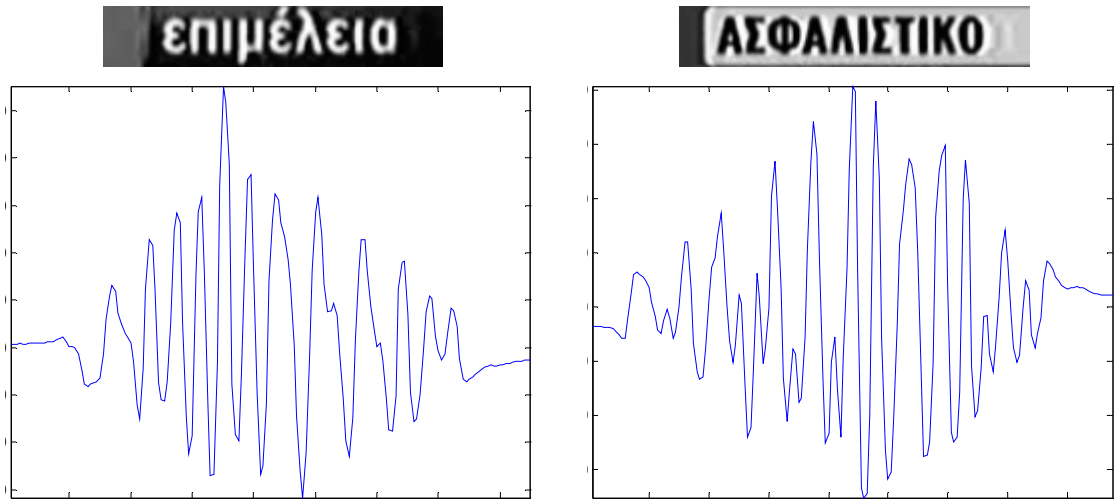
Μια άλλη παράμετρος που επηρεάζει τη μορφή της προβολής είναι η μη ακριβής οριοθέτηση του κειμένου. Συχνά, οι εντοπισμένες περιοχές κειμένου περιέχουν περιττά τμήματα που ανήκουν στο φόντο (σχ. 3.9) και συνεπώς στα αντίστοιχα τμήματα της προβολής δεν αποτυπώνεται η παρουσία κειμένου. Για τη μείωση της συμβολής των pixels που βρίσκονται κοντά στα δεξιά και αριστερά όρια της εικόνας, εφαρμόζεται παράθυρο Hamming οριζόντιου προσανατολισμού.



Σχήμα 3.7. Περιοχές κειμένου και οι αντίστοιχες προβολές τους.



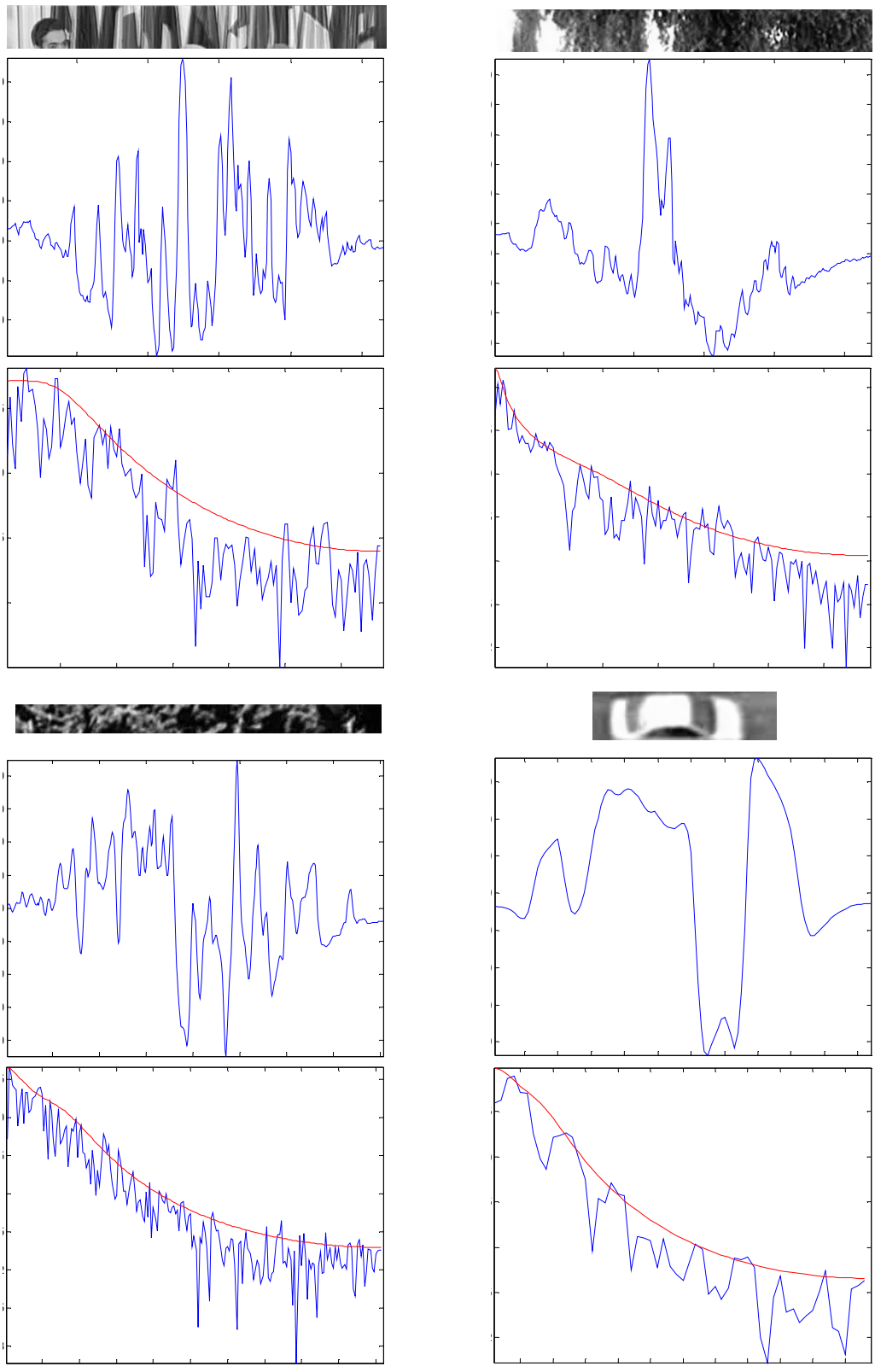
Σχήμα 3.8. Μη κειμενικές περιοχές και οι αντίστοιχες προβολές τους.



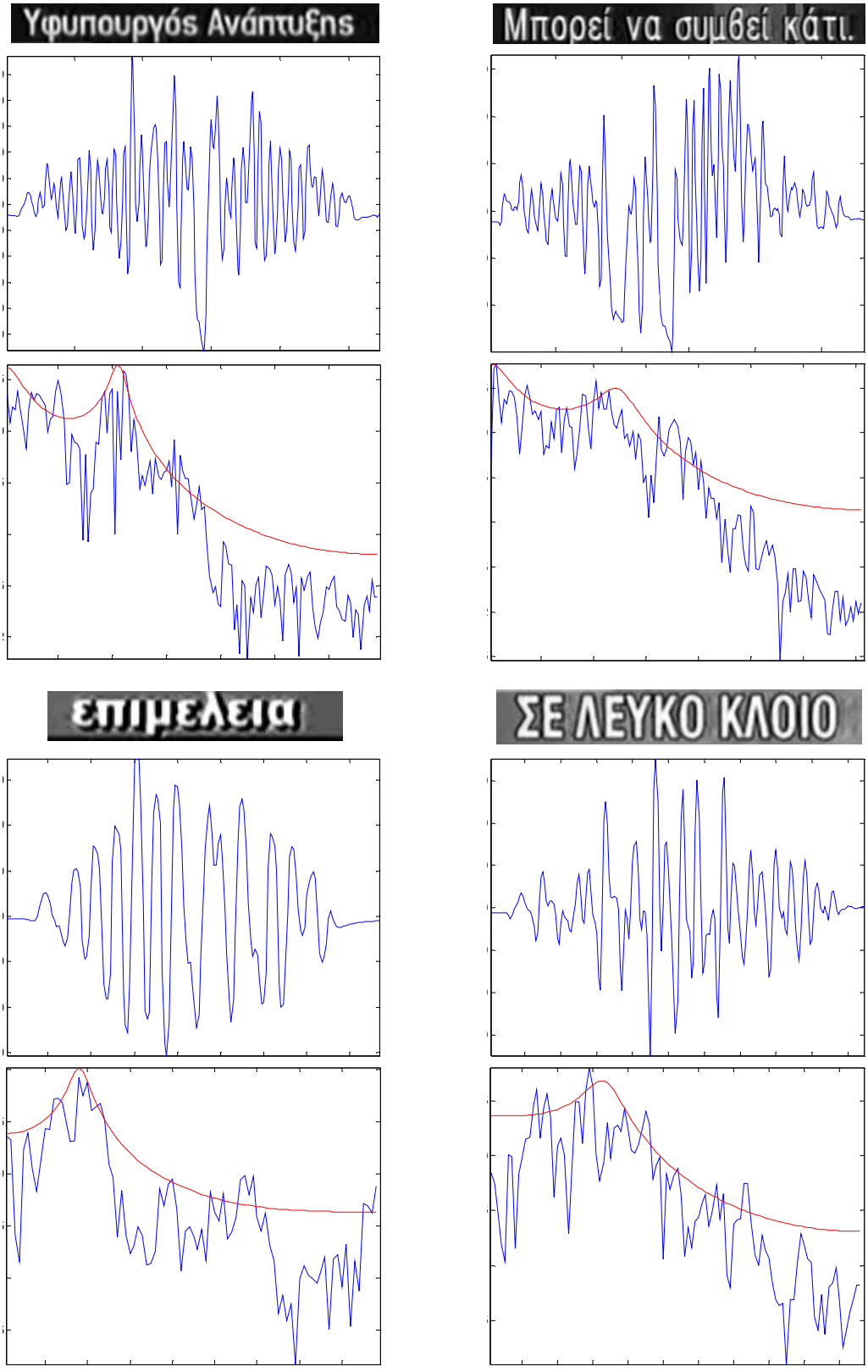
Σχήμα 3.9. Οι προβολές των εικόνων κειμένου μετά την εφαρμογή του παραθύρου Hamming.

Στη συνέχεια εξετάζεται αν αυτή η περιοδική εμφάνιση των κενών που αποτυπώνεται στις προβολές, μπορεί να αποτελέσει διακριτικό χαρακτηριστικό μεταξύ των κειμενικών και μη περιοχών. Για το λόγο αυτό συγκεντρώθηκαν 2400 RT και 8900 NT, όπως αυτά προέκυψαν από την εφαρμογή του αλγόριθμου εντοπισμού κειμένου, και υπολογίστηκαν τα αντίστοιχα φάσματα (σχ. 3.10-3.11). Από την παρατήρηση των φασμάτων προέκυψε ότι στα φάσματα των RT υπάρχει μια περιοχή συχνοτήτων που συγκεντρώνει σημαντική ενέργεια. Αντίθετα, στα φάσματα πολλών NT δεν παρατηρήθηκε κάτι ανάλογο. Για την εξαγωγή χαρακτηριστικών που εκφράζουν αυτή τη διαφορετικότητα επιλέχθηκε η παραμετρική μοντελοποίηση της προβολής μέσω ενός τρίτης τάξης μοντέλου γραμμικής πρόβλεψης. Σημειώνεται ότι ο αντικειμενικός στόχος δεν είναι η προσέγγιση του φάσματος, αλλά η εξαγωγή χρήσιμων χαρακτηριστικών για την κατηγοριοποίηση των υποψήφιων περιοχών κειμένου σε NT και RT. Με την εφαρμογή αυτού του μοντέλου αναμένεται να εξαχθεί μια εκτίμηση για τη θέση της κρίσιμης περιοχής συχνοτήτων μέσω της γωνίας του πόλου και της δυναμικής της μέσω του πλάτους του πόλου. Αναμένεται στις κειμενικές περιοχές να αντιστοιχούν υψηλές τιμές πλάτους και οι τιμές των γωνιών να περιορίζονται σε συγκεκριμένο εύρος. Στην τρίτη γραμμή των σχ. 3.10 και 3.11 παρουσιάζεται με κόκκινο χρώμα ο λογάριθμος του φάσματος της κρουστικής απόκρισης των αντίστοιχων σε κάθε προβολή μοντέλων. Η εμφάνιση στενού λοβού υποδηλώνει μεγάλη τιμή πλάτους για τον πόλο.

Ως τρίτο χαρακτηριστικό επιλέχθηκε το κέντρο βάρους του φάσματος με σκοπό την περιγραφή του σχήματος του φάσματος [82]. Θεωρώντας το φάσμα ως κατανομή των συχνοτήτων x με πιθανότητες να παρατηρηθούν αυτές, τις κανονικοποιημένες τιμές του φάσματος το κέντρο βάρους $\mu = \int x\, p(x) dx$. Αναμένεται η τιμή του κέντρου βάρους για τις περιοχές κειμένου να είναι μεγαλύτερη από αυτή πολλών περιοχών NT.

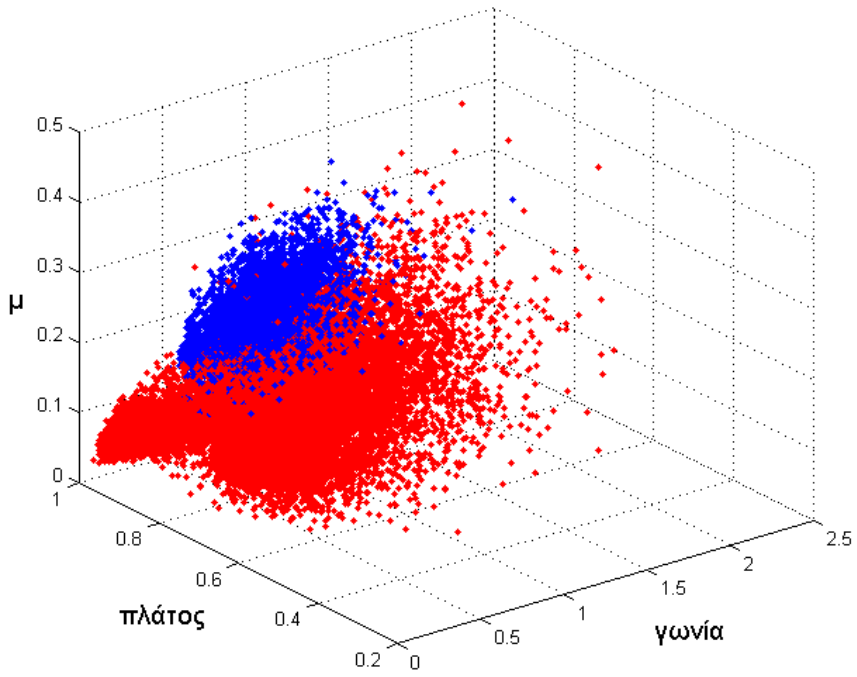


Σχήμα 3.10. Παραδείγματα μη κειμενικών περιοχών (1η και 4η γραμμή), οι προβολές τους (2η και 5η γραμμή), ο λογάριθμος κάθε φάσματος (μπλε) και ο λογάριθμος του φάσματος της κρουστικής απόκρισης του αντίστοιχου μοντέλου (3η και 6η γραμμή).



Σχήμα 3.11. Παραδείγματα κειμενικών περιοχών (1η και 3η γραμμή), οι προβολές τους (2η και 4η γραμμή), ο λογάριθμος κάθε φάσματος (μπλε) και ο λογάριθμος του φάσματος της κρουστικής απόκρισης του αντίστοιχου μοντέλου (3η και 6η γραμμή).

Οι θέσεις των διανυσμάτων των χαρακτηριστικών για τις περιοχές κειμένου (μπλε) και τις μη κειμενικές περιοχές (κόκκινο) παρουσιάζονται στο σχ. 3.12. Είναι προφανές ότι αν και οι τάξεις δεν είναι διαχωρίσιμες, μπορεί να επιτευχθεί σημαντική μείωση των αστοχιών της διαδικασίας εντοπισμού.



Σχήμα 3.12. Η διάταξη των διανυσμάτων χαρακτηριστικών RT (μπλε) και NT (κόκκινο).

Όπως έχει αναφερθεί, ο αντικειμενικός στόχος της διαδικασίας επαλήθευσης είναι η απόρριψη υποψήφιων περιοχών κειμένου που εσφαλμένα έχουν προταθεί από το στάδιο εντοπισμού, χωρίς φυσικά να μειώνεται το πλήθος των πραγματικών εντοπισμένων περιοχών κειμένου. Για την επίτευξη του στόχου, επιλέχθηκε η υιοθέτηση στατιστικών μοντέλων που εκτιμούν τις πιθανότητες κάθε εξεταζόμενη περιοχή να ανήκει στη μια ή την άλλη τάξη. Η απόφαση για την ταξινόμηση της περιοχής γίνεται με την εξέταση του λόγου των πιθανοτήτων. Για να ελεγχθεί η επιρροή του σταδίου στις κειμενικές περιοχές, ο λόγος των πιθανοφανειών (likelihood ratio) δεν συγκρίνεται με τη μονάδα, αλλά με ένα κατώφλι $thres < 1$.

Για την περιγραφή της δομής κάθε τάξης επιλέχθηκε το αντίστοιχο μίγμα γκαουσιανών (Gaussian Mixture Model, GMM) [59]. Έστω GMM_RT και GMM_NT τα μοντέλα για τις τάξεις RT και NT αντίστοιχα. Το μίγμα γκαουσιανών για την τάξη RT περιγράφεται ως εξής:

$$p(\mathbf{x} \mid GMM\_RT) = \sum_{j=1}^{ncRT} \pi_j^{RT} \mathcal{N}(\mathbf{x} \mid \mathbf{m}_j^{RT}, \mathbf{\Sigma}_j^{RT}), \quad j = 1,...,ncRT$$

με $ncRT$ το πλήθος των συνιστωσών του μίγματος και π_j^{RT}, $\mathbf{m}_j^{RT}$ και $\mathbf{\Sigma}_j^{RT}$ ο συντελεστής μίξης, η μέση τιμή και ο πίνακας συμμεταβλητότητας της j-οστής συνιστώσας.

Ομοίως, το μίγμα γκαουσιανών για την τάξη NT περιγράφεται ως:

$$p(x \mid GMM\_NT) = \sum_{k=1}^{ncNT} \pi_k^{NT} N(x \mid \mathbf{m}_k^{NT}, \mathbf{\Sigma}_k^{NT}),\ k = 1,...,ncNT$$

με $ncNT$ το πλήθος των συνιστωσών του μίγματος και π_j^{NT}, $\mathbf{m}_k^{NT}$ και $\mathbf{\Sigma}_k^{NT}$ ο συντελεστής μίξης, η μέση τιμή και ο πίνακας συμμεταβλητότητας της k -οστής συνιστώσας.

Μια υποψήφια εικόνα κειμένου $\mathbf{x}$ ταξινομείται ως RT αν ο λόγος των πιθανοφανειών είναι μεγαλύτερος ή ίσος από το κατώφλι $thres$

$$\frac{p(x \mid GMM\_RT)}{p(x \mid GMM\_NT)} \geq thres \Rightarrow x \in RT \qquad \text{(Εξ. 3.5)}$$

και ως NT αν είναι μικρότερος:

$$\frac{p(x \mid GMM\_RT)}{p(x \mid GMM\_NT)} < thres \Rightarrow x \in NT \qquad \text{(Εξ. 3.6)}$$

Για τον προσδιορισμό του πλήθους των συνιστωσών κάθε μίγματος ($ncRT$ και $ncNT$) και του κατωφλίου $thres$ εφαρμόστηκε η μέθοδος 10-fold cross-validation και για τον υπολογισμό των βέλτιστων παραμέτρων (π_j^{RT}, $\mathbf{m}_j^{RT}$, $\mathbf{\Sigma}_j^{RT}$, π_k^{NT}, $\mathbf{m}_k^{NT}$ και $\mathbf{\Sigma}_k^{NT}$) ο αλγόριθμος Expectation-Maximization (EM) [83]. Η συγκεκριμένη διαδικασία περιγράφεται ακολούθως:

Έστω $A = \{RT_i\},\ i = 1,...,2400$ και $B = \{NT_i\},\ i = 1,...,8900$ τα σύνολα των RT και NT που χρησιμοποιούνται για την εκπαίδευση των μοντέλων. Κάθε σύνολο χωρίζεται σε 10 υποσύνολά του A_m και B_m, $m = 1,...,10$, έτσι ώστε να είναι ξένα μεταξύ τους και $A = A_1 \cup A_2 ... \cup A_{10}$ και $B = B_1 \cup B_2 ... \cup B_{10}$. Σε κάθε επανάληψη, δύο υποσύνολα A_m και B_m συνιστούν το τρέχον σετ εξέτασης και η ένωση των υπολοίπων $A - A_m$ και $B - B_m$ τα τρέχοντα σετ εκμάθησης. Για κάθε δυνατό συνδυασμό του πλήθους των συνιστωσών των δύο μιγμάτων υπολογίζονται με την εφαρμογή του αλγορίθμου EM, οι παράμετροι των μιγμάτων. Για την αρχικοποίηση του EM εφαρμόζεται ο αλγόριθμος k-means [κεφ. 9 στο 59]. Με δεδομένα τα πλήθη των συνιστωσών των μιγμάτων και των αντίστοιχων παραμέτρων τους, υπολογίζονται οι πιθανοφάνειες για τον τρέχον σετ εξέτασης. Για τιμές του κατωφλίου από 0.1 ως 1 με βήμα 0.1, υπολογίζονται ο συντελεστής ορθής ταξινόμησης των κειμενικών περιοχών

$$R = \frac{\text{πλήθος σωστά ταξινομημένων κειμενικών περιοχών από το τρέχον σετ εξέτασης}}{\text{πλήθος κειμενικών περιοχών στο τρέχον σετ εξέτασης}}$$

και ο συντελεστής ακρίβειας

$$PR = \frac{\text{πλήθος σωστά ταξινομημένων κειμενικών περιοχών από το τρέχον σετ εξέτασης}}{\text{πλήθος περιοχών ταξινομημένων ως κειμενικές από το τρέχον σετ εξέτασης}}$$

Η περιγραφή της διαδικασίας με τη μορφή ψευδοκώδικα παρουσιάζεται ακολούθως:

for $m = 1:10$

$Tr\_RT_m = A - A_m$ /* τρέχον σετ εκπαίδευσης για το GMM_RT */

$Tr\_NT_m = B - B_m$ /* τρέχον σετ εκπαίδευσης για το GMM_NT */

for $j = 2:30$ /* πλήθος συνιστωσών για το GMM_RT */

Υπολογισμός των παραμέτρων π_j^{RT}, $\mathbf{m}_j^{RT}$ και $\mathbf{\Sigma}_j^{RT}$ υποθέτοντας j συνιστώσες και σετ δεδομένων το Tr_RT_m, έστω $GMM\_RT_m^j$ (αλγόριθμος EM)

for $k = 2:30$ /* πλήθος συνιστωσών για το GMM_RT */

Υπολογισμός των παραμέτρων π_k^{NT}, $\mathbf{m}_k^{NT}$ και $\mathbf{\Sigma}_k^{NT}$ υποθέτοντας k συνιστώσες και σετ δεδομένων το Tr_NT_m, $GMM\_NT_m^k$ (αλγόριθμος EM)

$PTA = p(A_m \mid GMM\_RT_m^j);$

$PNTA = p(A_m \mid GMM\_NT_m^k);$

$PTB = p(B_m \mid GMM\_RT_m^j);$

$PNTB = p(B_m \mid GMM\_NT_m^k);$

$T\_as\_T = sum[(PTA./PNTA) \geq thres]$

$NT\_as\_T = sum[(PTB./PNTB) \geq thres]$

$R(m,j,k) = T\_as\_T \ / \ |A_m|$ /* recall για τα τρέχοντα σετ*/

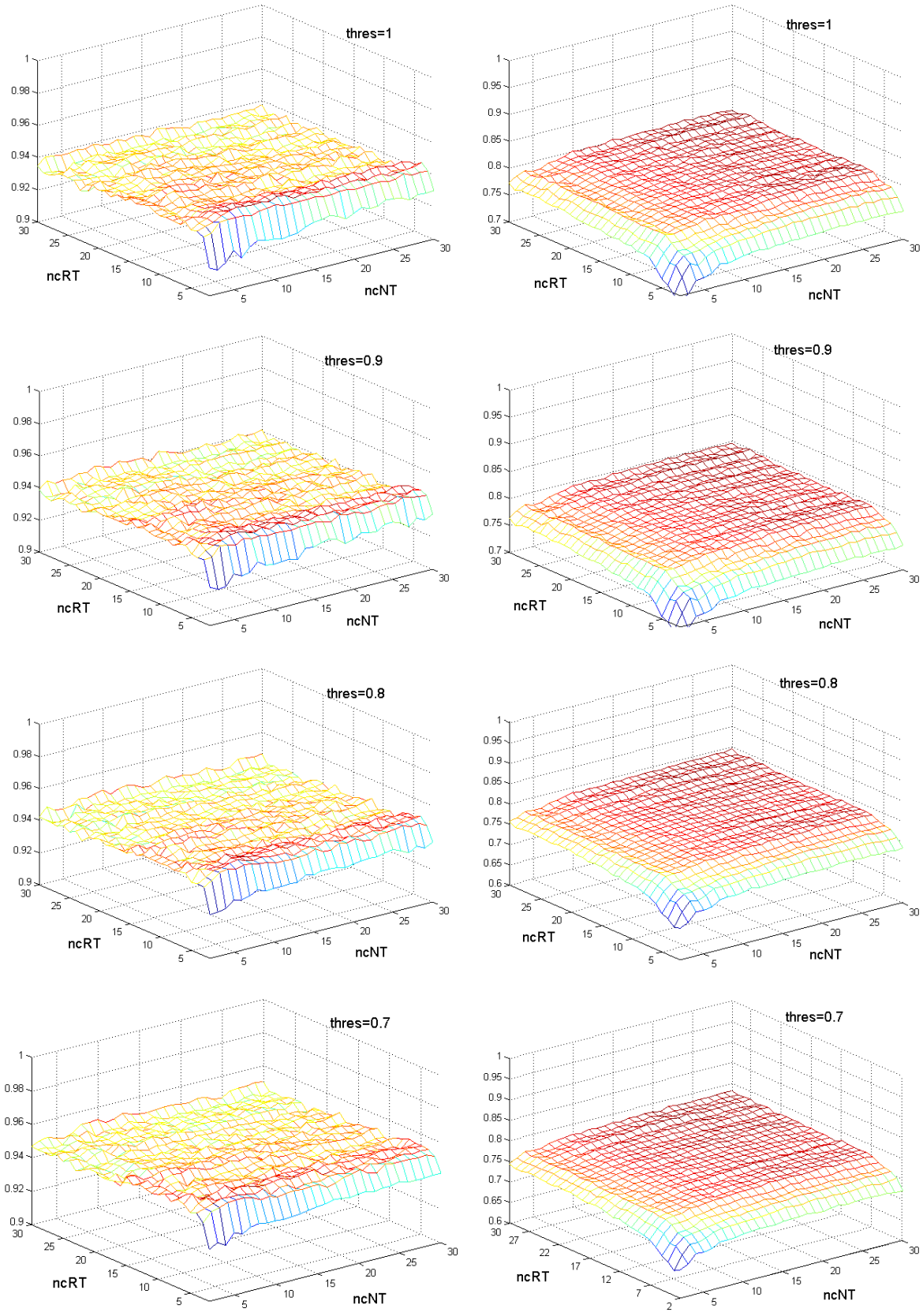
$PR(m,j,k) = T\_as\_T \ / \ (T\_as\_T + NT\_as\_T)$ /*precision*/

end

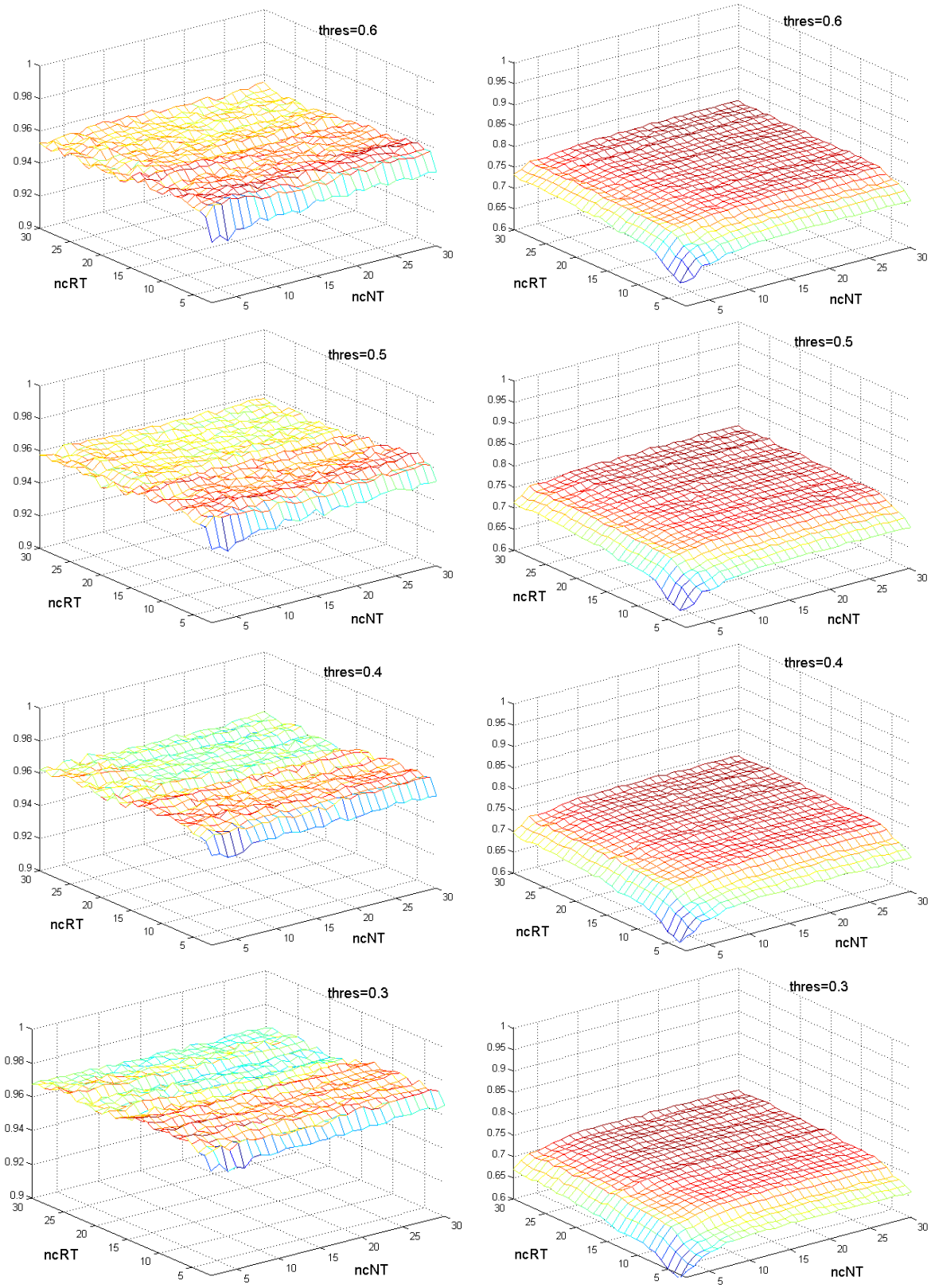
end

end

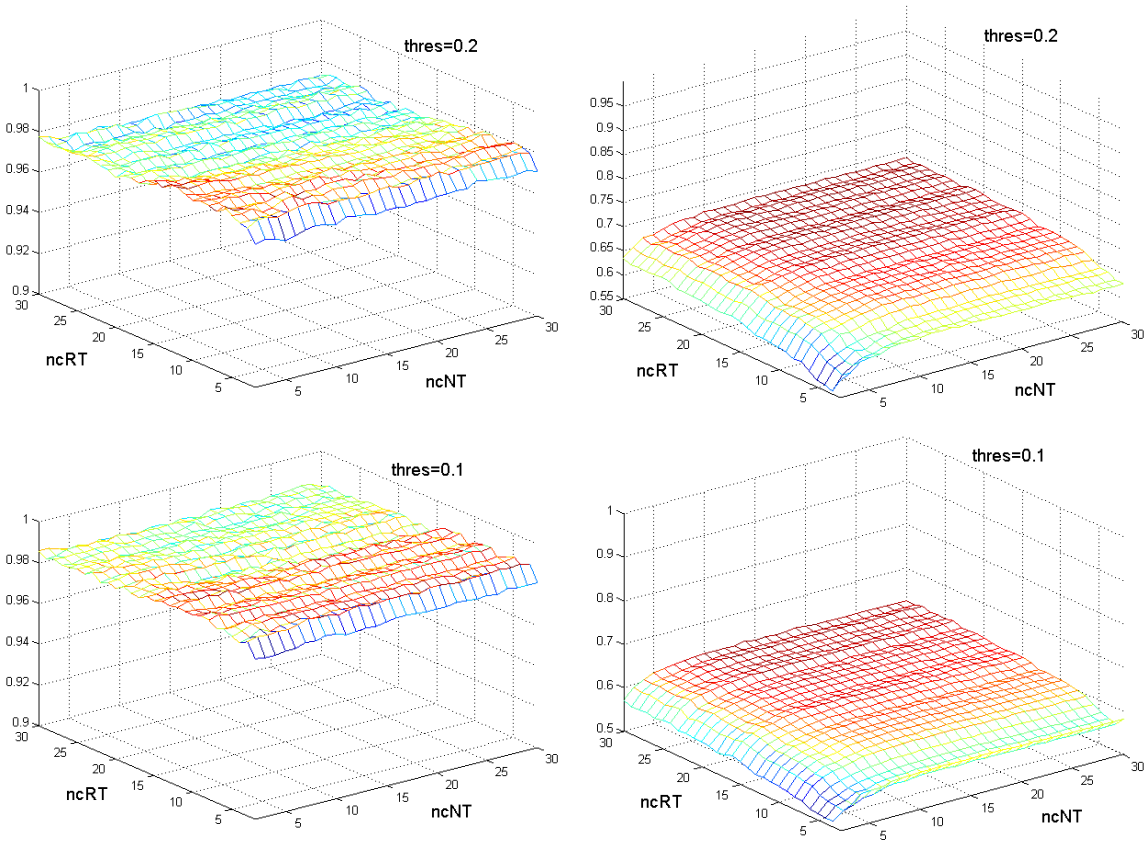
Με τη χρήση του κατωφλίου ελέγχεται η επιρροή του αλγόριθμου επαλήθευσης στην απόδοση του συστήματος. Με μικρές τιμές του κατωφλίου, πριμοδοτούνται οι κειμενικές περιοχές, έτσι ώστε να μην μειώνεται η απόδοση του σταδίου εντοπισμού. Βέβαια, τότε η μείωση των αστοχιών του σταδίου εντοπισμού είναι μικρότερη. Αντίθετα, για μεγάλες τιμές του κατωφλίου, πολλές μη κειμενικές περιοχές απομακρύνονται επιτυχώς, αλλά ταυτόχρονα και πολλές κειμενικές περιοχές απορρίπτονται Τα αποτελέσματα για την ορθή ταξινόμηση των περιοχών κειμένου και της ακρίβειας για διάφορες τιμές του κατωφλίου $thres$ και των συνιστωσών $ncRT$ και $ncNT$ παρουσιάζονται στα σχ.3.13α, β, γ.



Σχήμα 3.13α. Ορθή ταξινόμηση (αριστερά) και ακρίβεια (δεξιά) για τις τιμές $thres$ 1, 0.9, 0.8 και 0.7 και των συνιστωσών $ncNT$ και $ncRT$.



Σχήμα 3.13β. Ορθή ταξινόμηση (αριστερά) και ακρίβεια (δεξιά) για τις τιμές του $thres$ 0.6, 0.5, 0.4 και 0.3 και των συνιστωσών $ncNT$ και $ncRT$.



Σχήμα 3.13γ. Ορθή ταξινόμηση (αριστερά) και ακρίβεια (δεξιά) για τις τιμές του $thres$ 0.2. και 0.1 και των συνιστωσών $ncNT$ και $ncRT$.

Στον πίνακα 3.2 παρουσιάζονται οι τιμές των συνιστωσών που αντιστοιχούν στη μέγιστη ορθή ταξινόμηση των κειμενικών περιοχών για κάθε τιμή του κατωφλίου. Αντίστοιχα, στον πίνακα 3.3 παρουσιάζονται οι τιμές των συνιστωσών που αντιστοιχούν στη μέγιστη ακρίβεια. Παρατηρώντας τα διαγράμματα και τα ενδεικτικά στοιχεία των πινάκων, μπορεί να κανείς επιλέξει το κατάλληλο κατώφλι ώστε το στάδιο επαλήθευσης να προσαρμοστεί στις ανάγκες της εκάστοτε εφαρμογής και να είναι είτε χαλαρό (π.χ. μικρό κατώφλι), είτε αυστηρό. Αν κανείς επιλέξει την ελάχιστη αποδεκτή τιμή της ορθής ταξινόμησης ίση με 96%, τότε είναι προφανές ότι το επιλεγμένο κατώφλι θα είναι ≤ 0.5 . Για να επιτευχθεί και η μέγιστη μείωση των NT, θα πρέπει να επιλεγεί η μεγαλύτερη επιτρεπτή τιμή για το κατώφλι. Επομένως, η προφανής επιλογή είναι $thresh = 0.5$. Για την επιλογή του πλήθους των συνιστωσών, επιλέγονται αρχικά τα ζεύγη που παρουσιάζουν ορθή ταξινόμηση ≥ 0.96 και μεταξύ αυτών, επιλέγεται το ζεύγος με τη μεγαλύτερη ακρίβεια. Από το αντίστοιχο διάγραμμα προκύπτει ότι για το επιλεγμένο κατώφλι, οι κατάλληλες παράμετροι είναι $ncRT = 10$ και $ncNT = 28$ με R = 96.28% και PR = 74.71%. Δεδομένων των παραμέτρων, τα μοντέλα εκπαιδεύονται εκ νέου χρησιμοποιώντας τα σύνολα A και B .

Πίνακας 3.2. Πλήθη συνιστωσών με τη μέγιστη ορθή ταξινόμηση ανά τιμή κατωφλίου.

Κατώφλι	**Μέγιστη ορθή ταξινόμηση**			
	$ncRT$	$ncNT$	R **(%)**	PR **(%)**
1	4	12	94.6	77.78
0.9	4	7	94.85	75.62
0.8	6	7	95.23	75.84
0.7	6	7	95.65	74.85
0.6	5	29	95.94	74.14
0.5	5	29	96.53	72.73
0.4	6	7	97.03	70.19
0.3	6	2	97.57	61.40
0.2	6	2	98.24	57.59
0.1	12	7	98.95	60.12

Πίνακας 3.3. Πλήθη συνιστωσών με τη μέγιστη ακρίβεια ανά τιμή κατωφλίου.

Κατώφλι	**Μέγιστη ακρίβεια**			
	$ncRT$	$ncNT$	R **(%)**	PR **(%)**
1	27	22	93.22	80.84
0.9	27	22	93.68	80.08
0.8	30	30	94.64	79.28
0.7	27	22	94.64	78.22
0.6	25	18	95.02	76.88
0.5	25	18	95.61	75.64
0.4	25	24	96.23	74.00
0.3	30	19	96.69	71.79
0.2	21	22	97.32	68.63
0.1	30	14	98.33	63.09

Η προτεινόμενη μέθοδος εξετάστηκε αρχικά στο σετ δεδομένων που χρησιμοποιήθηκε για το διαγωνισμό ICDAR 2003 Robust Word Recognition Contest. Το σετ αποτελείται από 2437 εικόνες εντοπισμένων περιοχών κειμένου σε φωτογραφίες. Οι περιοχές με λόγο πλάτους ύψους μικρότερο του 2 δε συμμετείχαν στο πείραμα μια και δεν ενδιαφέρουν τον αλγόριθμο επαλήθευσης (τέτοιες περιοχές έχουν απορριφθεί από το στάδιο εντοπισμού). Από τις υπόλοιπες 1794 εικόνες, οι 1483 κατηγοριοποιήθηκαν σωστά (σχ. 3.14α) και οι υπόλοιπες απορρίφθηκαν (σχ. 3.14β). Το ποσοστό ορθής ταξινόμησης είναι 82.66% και κρίνεται

ικανοποιητικό αν ληφθεί υπόψη ότι το σετ εξέτασης περιλαμβάνει εικόνες κειμένου που αποτελούν τμήμα της φωτογραφιζόμενης σκηνής και όχι εικόνες πρόσθετου κειμένου.

Σχήμα 3.14. Εικόνες από το [79] που είτε επαληθεύτηκαν ως κειμενικές περιοχές (αριστερή στήλη) είτε απορρίφθηκαν (μεσαία και δεξιά).

Επίσης, ο αλγόριθμος επαλήθευσης εξετάστηκε στο ίδιο σετ δεδομένων με τον αλγόριθμο εντοπισμού κειμένου (βλ. Εν. 3.2) και τα συγκριτικά αποτελέσματα παρουσιάζονται στον Πίνακα 3.4. Από τις 11497 πραγματικές περιοχές κειμένου που εντοπίστηκαν (LT), οι 11319 κατηγοριοποιήθηκαν σωστά ως κειμενικές (RT). Επομένως, ο συνολικός συντελεστής εντοπισμού και επαλήθευσης (*recall)* είναι 0.947. Πρέπει να σημειωθεί ότι η συγκεκριμένη τιμή δεν αποτελεί το συνολικό δείκτη απόδοσης του συστήματος μια και εξετάζει κάθε πλαίσιο του βίντεο ανεξάρτητα και όχι σαν μια ακολουθία εικόνων. Με άλλα λόγια, η απόρριψη μιας περιοχής κειμένου που εντοπίστηκε σε κάποιο πλαίσιο δεν συνεπάγεται και την οριστική απόρριψή της γιατί είναι πιθανό να επαληθευτεί σε κάποια άλλη πραγμάτωσή της σε κάποιο κοντινό χρονικά πλαίσιο. Πράγματι, όπως θα εξηγηθεί στην επόμενη ενότητα, μια περιοχή κειμένου που πραγματώνεται σε Ν διαδοχικά από τα επιλεγμένα πλαίσια του βίντεο, θα χαθεί αν δεν εντοπιστεί ή δεν επαληθευτεί σε Ν-1 πλαίσια. Παράλληλα, το πλήθος των αστοχιών μειώνεται από 37589 (LNT) σε 3473 (RNT) και συνεπώς, η ακρίβεια του αλγορίθμου βελτιώθηκε σημαντικά από 0.23 σε 0.765.

Πίνακας 3.4. Αξιολόγηση μεθόδου εντοπισμού κειμένου.

GT	LT	LNT	Recall	Precision	RT	RNT	Recall	Precision
11952	11497	37589	0.96	0.23	11319	3473	0.947	0.765

Δύο χαρακτηριστικά παραδείγματα παρουσιάζονται στο ακόλουθο σχήμα. Στα σχ. 3.15α, γ παρουσιάζονται οι υποψήφιες περιοχές κειμένου, ως αποτέλεσμα του σταδίου εντοπισμού. Στα σχ. 3.15β, δ εμφανίζονται οι περιοχές που έχουν ταξινομηθεί ως περιοχές κειμένου από το στάδιο επαλήθευσης. Είναι προφανές ότι ενώ η μείωση των αστοχιών είναι ιδιαίτερα σημαντική,

μόλις μια πραγματική περιοχή κειμένου απορρίπτεται.

(α) (β)

(γ) (δ)

Σχήμα 3.15. Παραδείγματα μείωσης των αστοχιών.

3.4. Εξαγωγή «δυναμικής πληροφορίας»

Η μετάδοση πληροφορίας μέσω του εμφανιζόμενου κειμένου καθιστά αναγκαίες κάποιες ιδιότητες του κειμένου. Στα προηγούμενα στάδια έχουν εξεταστεί και χρησιμοποιηθεί ιδιότητες όπως η έντονη αντίθεση των χαρακτήρων με το φόντο, η οριζόντια διάταξή τους και η περιοδική τοποθέτησή τους. Μια άλλη ιδιότητα που δεν έχει συμπεριληφθεί ως τώρα είναι η διάρκεια εμφάνισής του. Είναι προφανές ότι κάθε κείμενο θα πρέπει να παραμένει εμφανές για εύλογο χρονικό διάστημα (π.χ. 1 sec), δηλαδή να εμφανίζεται σε διαδοχικά πλαίσια του βίντεο, ώστε να είναι εύληπτο. Σύμφωνα με τη συχνότητα δειγματοληψίας (βλ. Εν. 3.1), η επιλεγμένη ακολουθία πλαισίων είναι τέτοια ώστε κάθε κείμενο να εμφανίζεται σε δύο τουλάχιστον διαδοχικά πλαίσια. Η συσχέτιση της μεταδιδόμενης μέσω διαδοχικών πλαισίων πληροφορίας αναδεικνύει τη «δυναμική» πληροφορία του βίντεο. Η διαδικασία εξαγωγής της περιγράφεται ακολούθως:

α) Για κάθε υποψήφια περιοχή κειμένου αποθηκεύονται σε κατάλληλο αρχείο η θέση, οι διαστάσεις και οι αύξοντες αριθμοί των πλαισίων αρχικής και τελικής εμφάνισής της (δείκτες

έναρξης / τέλους). Σημειώνεται ότι αρχικά οι δείκτες εμφάνισης είναι ίσοι.

β) Δεδομένου ότι κάθε περιοχή κειμένου εμφανίζεται σε δύο τουλάχιστον γειτονικά πλαίσια, κάθε υποψήφια περιοχή που βρέθηκε στο πλαίσιο f_i (δείκτης έναρξης i) συγκρίνεται με τις περιοχές που βρέθηκαν στα πλαίσια i-10 ως i-1. Αν η επικάλυψη δύο περιοχών είναι σημαντική (π.χ. >70%), τότε υπολογίζεται ο δισδιάστατος συντελεστής συσχέτισης *r* (2-D correlation coefficient) των επικαλυπτόμενων τμημάτων.

$$r = \sum_x \sum_y \left(A_{xy} - \overline{A} \right) \left(B_{xy} - \overline{B} \right) \Big/ \sqrt{\left(\sum_x \sum_y \left(A_{xy} - \overline{A} \right)^2 \right) \left(\sum_x \sum_y \left(B_{xy} - \overline{B} \right)^2 \right)} \qquad \text{(Εξ. 3.7)}$$

όπου $\overline{A} = \frac{1}{nm} \sum_{x=1}^{m} \sum_{y=1}^{n} A(x, y)$, $\overline{B} = \frac{1}{nm} \sum_{x=1}^{m} \sum_{y=1}^{n} B(x, y)$

Αν ο συντελεστής (εξ. 3.6) είναι υψηλός (π.χ. >0.8), δηλαδή οι εξεταζόμενες περιοχές μοιάζουν σημαντικά, τότε επιλέγεται η μεγαλύτερη σε έκταση περιοχή, αντικαθιστώντας το κοινό τμήμα της, με τη μέση τιμή των επικαλυπτόμενων τμημάτων (averaging). Σημειώνεται ότι ως συντελεστής ομοιότητας έχει επιλεγεί ο Pearson product moment συντελεστής συσχέτισης που εκφράζεται από το λόγο της συμμεταβλητότητας των εικόνων προς το γινόμενο των τυπικών αποκλίσεών τους. Ως νέοι δείκτες εμφάνισης της επιλεγμένης περιοχής, ορίζονται ο μικρότερος και ο μεγαλύτερος από τους δείκτες έναρξης και τέλους αντίστοιχα. Η μικρότερη περιοχή σημειώνεται ως περιττή αφού η «πληροφορία» της έχει πλέον «ενσωματωθεί» στην επιλεγμένη περιοχή.

γ) Τέλος, κάθε περιοχή, η οποία είναι σημειωμένη ως περιττή, ή έχει ίσους δείκτες εμφάνισης απορρίπτεται.

Με τον τρόπο αυτό απομακρύνονται οι περιοχές που εμφανίστηκαν σε ένα μόνο πλαίσιο του βίντεο και επομένως είτε δεν είναι σημαντικοί φορείς πληροφορίας είτε δεν είναι πραγματικές περιοχές κειμένου. Οι εικόνες αυτές οδηγούνται στην εμπορική μηχανή οπτικής αναγνώρισης χαρακτήρων ABBYY Finereader 8.1 και κάθε αναγνωρισμένο κείμενο συνδυάζεται με τις αντίστοιχες πληροφορίες (έναρξη, τέλος και θέση εμφάνισης), ώστε να δημιουργηθεί το αρχείο δεικτοδότησης του βίντεο σε μορφότυπο XML.

3.5. Αξιολόγηση συστήματος

Όπως έχει αναφερθεί το σύστημα αποτελεί βαθμίδα του Πανόπτη και συμβάλει στη δεικτοδότηση της μεταδιδόμενης πληροφορίας. Σε τακτικούς ελέγχους που πραγματοποιεί η αρμόδια ομάδα από το Εθνικό Συμβούλιο Ραδιοτηλεόρασης δεν έχουν αναφερθεί σημαντικές αστοχίες. Πέρα από την εξέταση της αποτελεσματικότητας του σταδίου εντοπισμού και της διαδικασίας επαλήθευσης, έγινε και αξιολόγηση του συνολικού συστήματος. Συγκεκριμένα, χρησιμοποιήθηκαν 40 ώρες τηλεοπτικού προγράμματος ειδησεογραφικού περιεχομένου από δέκα γνωστούς τηλεοπτικούς σταθμούς και τα ποσοστά ορθού εντοπισμού και ακρίβειας είναι 94.08% και 78.93% αντίστοιχα. Σημειώνεται ότι στη μείωση του συντελεστή ακρίβειας συμβάλει

και το γεγονός ότι περίπου το 30% των τελικών εικόνων περιέχει πραγματώσεις κειμένου που αποτελούν μέρος της σκηνής, όπως αυτές που φαίνονται στο ακόλουθο σχήμα.

Σχήμα 3.16. Παραδείγματα κειμένων που αποτελούν μέρος της προβαλλόμενης σκηνής.

Βέβαια, μια σημαντική παράμετρος για την αξιολόγηση του συστήματος είναι η «ποιότητα» του ηλεκτρονικού κειμένου (ASCII) που προκύπτει από την οπτική αναγνώριση του εντοπισμένου κειμένου. Όπως, έχει αναφερθεί, για τη συγκεκριμένη εργασία χρησιμοποιείται μια εμπορική εφαρμογή οπτικής αναγνώρισης χαρακτήρων. Η συγκεκριμένη εφαρμογή είναι σχεδιασμένη για την επεξεργασία εικόνων κειμένου υψηλής ανάλυσης και συνεπώς αντιμετωπίζει δυσκολίες στην αναγνώριση των εικόνων χαμηλής ανάλυσης όπως αυτές που προκύπτουν από τα πλαίσια του βίντεο. Για το λόγο αυτό ένα επόμενο στάδιο εξέλιξης του συστήματος θα πρέπει να στοχεύει στη δημιουργία εικόνων κατάλληλης ευκρίνειας. Αυτό μπορεί να επιτευχθεί είτε με τον κατάλληλο συνδυασμό των εικόνων που περιέχουν το ίδιο κείμενο, είτε με την ανάπτυξη κατάλληλης τεχνικής δυαδικοποίησης της εικόνας κειμένου.

Ένα πρόσθετο κριτήριο για την αξιολόγηση κάθε τέτοιου συστήματος είναι ο χρόνος απόκρισής του. Από την παρατήρηση της καθημερινής λειτουργίας του συστήματος στις εγκαταστάσεις του Εθνικού Συμβουλίου Ραδιοτηλεόρασης έχει προκύψει το συμπέρασμα ότι ο απαιτούμενος χρόνος για την επεξεργασία ενός ωριαίου τηλεοπτικού προγράμματος (3600 πλαίσια, ανάλυσης 720×576) ποικίλει από 15 ως 20 λεπτά της ώρας.

Κεφάλαιο 4. Συμπεράσματα

Η ανάλυση εικόνων κειμένου είναι ένας θεματικός τομέας, στον οποίο δραστηριοποιούνται πολλοί τεχνολογικοί φορείς με στόχο την ανάπτυξη σύγχρονων και αποτελεσματικών προϊόντων. Ειδικά για την επεξεργασία εικόνων εντύπων, ήδη είναι διαθέσιμα μερικά πολύ γνωστά προϊόντα λογισμικού, όπως τα συστήματα οπτικής αναγνώρισης χαρακτήρων, που ενσωματώνουν ταχείς και αποτελεσματικές διαδικασίες για τον εντοπισμό των κυρίαρχων στοιχείων της εικόνας (π.χ. κείμενο, εικόνες, πίνακες, κ.λπ.). Αντίθετα, στην επεξεργασία εικόνων χειρόγραφων κειμένων δεν έχει συντελεστεί η ανάλογη πρόοδος, προφανώς λόγω της μεγάλης ποικιλομορφίας που παρουσιάζουν. Άλλωστε, ακόμα και για χειρόγραφα που περιέχουν μόνο κειμενικά στοιχεία, τα προβλήματα κατάτμησής τους σε γραμμές κειμένου και λέξεις δεν έχουν βρει τις αντίστοιχες λύσεις. Στη συγκεκριμένη εργασία εξετάστηκαν τα δύο αυτά προβλήματα και προτάθηκαν νέες τεχνικές για την επίλυσή τους.

Αν και οι προσεγγίσεις που έχουν υιοθετηθεί για την κατάτμηση σε γραμμές κειμένου διαφοροποιούνται με βάση την μέθοδο στην οποία βασίζονται, τελικά δύο είναι οι μεγάλες κατηγορίες. Η πρώτη περιλαμβάνει αυτές που εξετάζουν συνολικά την εικόνα κειμένου με στόχο τον εντοπισμό των περιοχών που αντιστοιχούν σε γραμμές κειμένου (μέθοδοι top-down). Μια από τις πιο γνωστές top-down μεθόδους είναι αυτή που βασίζεται στη χρήση των επιμέρους προβολών (χωρισμός της εικόνας σε κατακόρυφες ζώνες και εξέταση κάθε ζώνης ξεχωριστά). Όπως είναι φυσικό, λόγω της τμηματικής εξέτασης της εικόνας κειμένου, οι επιμέρους προβολές είναι αποτελεσματικές στις περιπτώσεις κειμένων με μεταβλητή κλίση μεταξύ των γραμμών αλλά και κατά μήκος της ίδιας γραμμής. Επομένως, η προτεινόμενη μέθοδος υπερέχει των τεχνικών που βασίζονται στις ολικές προβολές, αφού αυτές προϋποθέτουν ότι η κλίση όλων των γραμμών κειμένου είναι σταθερή. Επίσης, υπερτερεί των μεθόδων που χρησιμοποιούν το μετασχηματισμό Hough, μια και αυτές προϋποθέτουν ότι η κλίση κατά μήκος μιας γραμμής είναι σχεδόν σταθερή.

Βέβαια, η τμηματική εξέταση της εικόνας κειμένου, καθιστά τις επιμέρους προβολές ιδιαίτερα ευαίσθητες σε τοπικά φαινόμενα που απαντώνται στα χειρόγραφα, όπως είναι η μεταβλητότητα του μεγέθους των χαρακτήρων και η ύπαρξη μεγάλων κενών, που δυσχεραίνουν τον εντοπισμό των σημαντικών τοπικών ελαχίστων των προβολών, δηλαδή τις θέσεις των διαχωριστικών μεταξύ διαδοχικών γραμμών κειμένου. Για την αντιμετώπιση αυτών των φαινομένων προτείνονται ο αρχικός διαχωρισμός κάθε ζώνης σε τμήματα κειμένου και κενού, και η δημιουργία ενός Κρυφού Μαρκοβιανού Μοντέλου για τη μοντελοποίηση των ζωνών ως ακολουθίες παρατηρήσεων που παράγονται από αυτό. Με τον τρόπο αυτό, αποφεύγεται η επισφαλής διαδικασία της άμεσης οριοθέτησης των γραμμών κειμένου ανά ζώνη και προτείνεται η ανάδειξή τους ως απόκριση ενός μοντέλου που ενσωματώνει πληροφορία από όλη την εικόνα κειμένου.

Μια πρόσθετη δυσκολία που πρέπει να αντιμετωπιστεί κατά την εφαρμογή των επιμέρους προβολών είναι η επίλυση των αμφισημιών που δημιουργούνται κατά το ταίριασμα των

διαχωριστικών μεταξύ διαδοχικών ζωνών. Η προτεινόμενη μέθοδος υιοθετεί τη χρήση μιας συνάρτησης κόστους που περιλαμβάνει την κατακόρυφη απόσταση των διαχωριστικών (όπως όλες οι σχετικές μέθοδοι) και εξετάζει την πυκνότητα κειμένου στην περιοχή της αμφισημίας.

Η ενσωμάτωση των δύο αυτών σταδίων επεξεργασίας καθιστά τις επιμέρους προβολές πιο αποδοτικές στην κατάτμηση του χειρογράφου σε γραμμές κειμένου. Αυτό προκύπτει και από την αξιολόγηση της προτεινόμενης μεθόδου μέσω της συμμετοχής της στους διαγωνισμούς κατάτμησης χειρόγραφου κειμένου [7, 38].

Στη δεύτερη κατηγορία μεθόδων για την κατάτμηση χειρόγραφου σε γραμμές ανήκουν αυτές που εξετάζουν τις σχέσεις μεταξύ των CCs με στόχο την προοδευτική δημιουργία μεγαλύτερων τμημάτων ή ομάδων που αντιστοιχούν σε γραμμές κειμένου. Στην παρούσα εργασία προτείνεται μια τεχνική βασισμένη στους τελεστές δυαδικής μορφολογίας. Το πρώτο στάδιο είναι η δημιουργία μιας εικόνας χαμηλής ανάλυσης στην οποία αναδεικνύονται τα αρχικά τμήματα κάθε γραμμής κειμένου. Στο επόμενο στάδιο τα αρχικά τμήματα επεκτείνονται επαναληπτικά. Η βασική διαφορά της προτεινόμενης τεχνικής από παρόμοιες μεθόδους είναι ότι κάθε στάδιο επέκτασης των αρχικών τμημάτων ακολουθείται από το αντίστοιχο στάδιο ελέγχου για το σχηματισμό ειδικών προτύπων-σχημάτων που δηλώνουν ότι τμήματα διαδοχικών γραμμών κειμένου τείνουν να ενωθούν. Αν εντοπιστούν τέτοια πρότυπα, τότε τα αντίστοιχα pixels μετατρέπονται σε pixels του φόντου και συνεχίζεται η διαδικασία επέκτασης. Με τον τρόπο αυτό, η προτεινόμενη τεχνική συμβάλει στην πρόληψη του ανεπιθύμητου φαινόμενου της ενοποίησης διαδοχικών γραμμών από το οποίο υποφέρουν παρόμοιες τεχνικές.

Στη συγκεκριμένη εργασία εξετάστηκε και το πρόβλημα της κατάτμησης χειρόγραφων κειμένων σε λέξεις. Η κοινή πρακτική που ακολουθείται είναι η επιλογή μιας ποσότητας για την εκτίμηση των αποστάσεων μεταξύ διαδοχικών CCs του κειμένου και η κατηγοριοποίηση των ποσοτήτων σε αυτές που αντιστοιχούν σε κενά μεταξύ λέξεων ή εντός λέξεων. Οι ποσότητες που έχουν προταθεί στη βιβλιογραφία δεν μπορούν να περιγράψουν κατάλληλα τις αποστάσεις μεταξύ των CCs και συνήθως είτε χρησιμοποιούνται συνδυασμοί τους είτε γίνονται κάποιες υποθέσεις που περιορίζουν σημαντικά τη δυνατότητα γενίκευσης. Στην παρούσα εργασία προτείνεται μια νέα ποσότητα, για τον υπολογισμό της οποίας λαμβάνεται υπόψη το σχήμα και ο προσανατολισμός των εξεταζόμενων CCs. Αν θεωρηθεί ότι τα pixels των δύο CCs συνιστούν δύο τάξεις, τότε μια ποσότητα ανάλογη του περιθωρίου ταξινόμησης του βέλτιστου γραμμικού ταξινομητή που τις διαχωρίζει, είναι μια κατάλληλη ποσότητα για την εκτίμηση του κενού μεταξύ των CCs. Με τον τρόπο αυτό, ποσοτικοποιούνται κατάλληλα οι αποστάσεις μεταξύ επικαλυπτόμενων ή μη, και πλάγιων ή μη χαρακτήρων, μια και η ευθεία διαχωρισμού θα τοποθετείται κατάλληλα μέσω της επίλυσης του προβλήματος βελτιστοποίησης.

Μια μελλοντική εφαρμογή του προτεινόμενου τρόπου ποσοτικοποίησης των κενών μπορεί να είναι η εκτίμηση της κλίσης της γραφής. Υπολογίζοντας την κλίση του βέλτιστου γραμμικού ταξινομητή για κάθε ζεύγος διαδοχικών CCs, είναι εφικτή η εκτίμηση της κλίσης των γραμμάτων που υιοθετεί ο γραφέας του κειμένου. Η πληροφορία αυτή μπορεί να αξιοποιηθεί σε επόμενα στάδια επεξεργασίας και κυρίως στην αναγνώριση του χειρόγραφου κειμένου.

Για την κατηγοριοποίηση των ποσοτήτων γίνονται συνήθως οι δύο ακόλουθες υποθέσεις: α) οι μεγάλες αποστάσεις αντιστοιχούν σε κενά μεταξύ λέξεων ενώ οι μικρότερες σε κενά μεταξύ των χαρακτήρων της ίδιας λέξης και β) οι ποσότητες που αντιστοιχούν στα κενά μεταξύ λέξεων είναι κοντινές. Υιοθετώντας τις υποθέσεις αυτές για την κατηγοριοποίηση των ποσοτήτων, προτείνεται η εκτίμηση της συνάρτησης πυκνότητας πιθανότητας και ως κατώφλι επιλέγεται η τιμή που αντιστοιχεί στο δεξιότερο τοπικό ελάχιστό της.

Η προτεινόμενη μέθοδος για την κατάτμηση του χειρόγραφου κειμένου σε λέξεις, περιλαμβάνει έναν νέο τρόπο ποσοτικοποίησης των κενών μεταξύ διαδοχικών χαρακτήρων που προσαρμόζεται στις ιδιαιτερότητες των κειμένων. Επίσης, για την κατηγοριοποίηση των κενών δε χρησιμοποιεί προκαθορισμένες παραμέτρους. Επομένως, μπορεί να είναι αποτελέσει μια αποτελεσματική τεχνική για την επεξεργασία ποικίλων χειρόγραφων κειμένων. Η αξιολόγησή της μέσω της υποβολής της στους διαγωνισμούς κατάτμησης χειρόγραφου κειμένου [7, 38] ανέδειξε την αποτελεσματικότητά της.

Ως επέκταση της επεξεργασίας εικόνων που περιέχουν μόνο κειμενικά στοιχεία, στην εργασία περιγράφεται μια γρήγορη τεχνική για τον εντοπισμό πρόσθετου κειμένου οριζόντιου προσανατολισμού σε πλαίσια βίντεο. Αν και με την προτεινόμενη τεχνική εντοπίζονται τέτοιες κειμενικές πραγματώσεις, προτείνονται ως υποψήφιες περιοχές κειμένου και πολλά τμήματα του πλαισίου που δεν περιέχουν κείμενο. Για τη μείωση του πλήθους των αστοχιών προτείνεται ένα πρόσθετο στάδιο επεξεργασίας για την επαλήθευση των υποψήφιων περιοχών κειμένου. Η μέθοδος στοχεύει στην παραμετροποίηση των εικόνων και στην ταξινόμησή τους σε κειμενικές και μη. Για την παραμετροποίηση των εικόνων επιλέχθηκαν τρία χαρακτηριστικά για την περιγραφή του φάσματος της προβολής των εικόνων. Δύο μίγματα γκαουσιανών υπολογίστηκαν για την περιγραφή της δομής κάθε τάξης και η ταξινόμηση προκύπτει από τη σύγκριση του λόγου των πιθανοτήτων με ένα κατάλληλα επιλεγμένο κατώφλι. Η χρήση του κατωφλίου υιοθετήθηκε για να είναι εφικτός ο έλεγχος της επιρροής του αλγορίθμου στο σύστημα. Από την αξιολόγηση της μεθόδου προέκυψε ότι συμβάλει στη μείωση των αστοχιών χωρίς να μειώνει σημαντικά το πλήθος των πραγματικών εντοπισμένων περιοχών κειμένου. Επομένως, μπορεί να ενσωματωθεί σε ανάλογα συστήματα και να συμβάλει στη μείωση της περιττής πληροφορίας που μεταδίδεται στα επόμενα στάδια επεξεργασίας. Βέβαια, υπάρχουν περιθώρια βελτίωσής της και ως μελλοντική εργασία ορίζεται η εξαγωγή και χρήση κι άλλων χαρακτηριστικών που περιγράφουν τη μορφή του φάσματος, όπως η ασυμμετρία και η κύρτωση.

Οι προτεινόμενες μέθοδοι εντοπισμού και επαλήθευσης αποτελούν τμήμα του συστήματος Πανόπτης που χρησιμοποιείται από το Εθνικό Συμβούλιο Ραδιοτηλεόρασης για τη δεικτοδότηση του μεταδιδόμενου τηλεοπτικού προγράμματος με βάση το εμφανιζόμενο κείμενο. Η συγκεκριμένη βαθμίδα επεξεργασίας θα μπορούσε να φανεί χρήσιμη και σε συστήματα αναγνώρισης του ομιλητή. Ο εντοπισμός και η αναγνώριση της κειμενικής πληροφορίας που περιέχει τα στοιχεία του ομιλητή θα μπορούσε να αξιοποιηθεί από τέτοια συστήματα και να

βελτιώσει την αποτελεσματικότητά τους. Ο συγκερασμός δύο τέτοιων συστημάτων είναι ένας από τους μελλοντικούς στόχους.

Βιβλιογραφία

[1] G. Nagy, Twenty Years of Document Image Analysis, IEEE Transactions on Pattern Analysis and Machine Intelligence, V. 22, No. 1, pp. 38–61, 2000.

[2] S. Mao, A. Rosenfeld & T. Kanungo, Document Structure Analysis Algorithms: A Literature Survey, in Proc. of Document Recognition and Retrieval X, V. 5010, pp. 197-207, 2003.

[3] A. Antonacopoulos, P. Pletschacher, D. Bridson & C. Papadopoulos, ICDAR2009 Page Segmentation Competition, in Proc. of Int'l Conf. on Document Analysis and Recognition, pp. 1370-1374, 2009.

[4] R. Plamondon & S.N. Srihari, On-Line and Off-Line Handwriting Recognition: A Comprehensive Survey, IEEE Transactions on Pattern Analysis Machine Intelligence, V. 22, No. 1, pp. 63–84, 2000.

[5] Π. Μαραγκός, Όραση Υπολογιστών, ΕΜΠ 2005.

[6] R. M. Haralick & L. G. Shapiro, Computer and Robot Vision, V. I, Addison-Wesley, 1992.

[7] B. Gatos, A. Antonacopoulos & N. Stamatopoulos, ICDAR2007 handwriting segmentation contest, in Proc. of Int'l Conf. on Document Analysis and Recognition, pp. 1284-1288, 2007.

[8] J. J. Hull, Document Image Skew Detection: Survey and Annotated Bibliography, in: J. J. Hull & S. L. Taylor (Eds.), Document Analysis Systems II, World Scientific, pp. 40-64, 1998.

[9] S. Chen & R.M. Haralick, An Automatic Algorithm for Text Skew Estimation in Document Images Using Recursive Morphological Transforms, in Proc. of Int'l Conf. on Image Processing, pp. 139-143,1994.

[10] L. Najman, Using mathematical morphology for document skew estimation, in Proc. of SPIE Document Recognition and Retrieval XI, V. 5296, pp. 182-191, 2004.

[11] S. Chen & R.M. Haralick, Recursive Erosion, Dilation, Opening, and Closing Transforms, IEEE Transactions on Image Processing, V. 4, No. 3, 1995.

[12] L. O'Gorman, The document spectrum for page layout analysis, IEEE Transactions on Pattern Analysis and Machine Intelligence, V. 15, No. 11, pp. 1162–1173, 1993.

[13] Z. Razak, K. Zulkiflee, et al., Off-line Handwriting Text Line Segmentation: A Review, International Journal of Computer Science and Network Security, V.8, No.7, pp. 12-20, 2008.

[14] L. Likforman-Sulem, A. Zahour & B. Taconet, Text Line Segmentation of Historical Documents: A Survey, International Journal on Document Analysis and Recognition, V. 9, Is. 22, pp.123-138, 2007.

[15] B. Yanikoglu & P.A. Sandon, Segmentation of Off-line Cursive Handwriting Using Linear Programming, Pattern Recognition, V. 31, No. 12, pp. 1825-1833, 1998.

[16] E. Bruzzone & M.C. Coffeti, An Algorithm for Extracting Cursive Text Lines, in Proc. of Int'l Conf. on Document Analysis and Recognition, pp. 749, 1999.

[17] M. Arivazhagan, H. Srinivasan & S. Srihari, A statistical approach to line segmentation in handwritten documents, in Proc. of Document Recognition and Retrieval XIV, V. 6500, No.1, 2007.

[18] K. Kuzhinjedathu, H. Srinivansan & S. Srihari, Robust Line Segmentation for Handwritten Documents, in Proc. of Document Recogntion and Retrieval XV, SPIE V. 6815, 2008.

[19] R. O. Duda & P. E. Hart, Use of the Hough Transformation to Detect Lines and Curves in Pictures, Communications of the ACM, V. 15, No. 1, pp. 11-15, 1972.

[20] L. Likforman-Sulem, A. Hanimyan & C. Faure, A Hough Based Algorithm for Extracting Text Lines in Handwritten Documents, in Proc. of Int'l Conf. on Document Analysis and Recognition, pp. 774-777, 1995.

[21] G. Louloudis, B. Gatos & C. Halatsis, Text line detection in handwritten documents, Pattern Recognition, 41, Is. 12, pp. 3758-3772, 2008.

[22] M. Feldbach & K. D. Tonnies, Line Detection and Segmentation in Historical Church Registers, in Proc. of Int'l Conf. on Document Analysis and Recognition, pp. 743-747, 2001.

[23] E. Cohen, J.J. Hull & S.N. Srihari, Control Structure for Interpreting Handwritten Addresses, IEEE Transactions on Pattern Analysis and Machine Intelligence, V. 16, No. 10, pp.1049-1055, 1994.

[24] Z. Shi & V. Covindaraju, Line Separation for Complex Document Images Using Fuzzy Runlength, in Proc. of International Workshop on Document Analysis for Libraries, pp. 306-312, 2004.

[25] N. Otsu, A Threshold Selection Method From Gray-Level Histograms, IEEE Transactions on Systems. Man and Cybernetics, Vol. 9, No. 1, pp. 62-66, 1979.

[26] Y. Li, Y. Zheng, D. Doermann & S. Jaeger, Script-Independent Text Line Segmentation in Freestyle Handwritten Documents, IEEE Transactions on Pattern Analysis and Machine Intelligence, V. 30, No. 8, pp. 1313-1329, 2008.

[27] W. Niblack, An Introduction to Digital Image Processing, Englewood Cliffs, Prentice Hall, pp. 115-116, 1986

[28] S. Osher & R. Fedkiw, Level Set Methods and Dynamic Implicit Surfaces, Springer 2003.

[29] D. J. Kennar, W. A. Barrett, Separating Lines of Text in Free-Form Handwritten Historical Documents, in Proc. Int'l Workshop Document Image Analysis for Libraries 2006, pp. 12-23.

[30] Y. Boykov & V. Kolmogorov, An Experimental Comparison of Min-Cut/Max-Flow Algorithms for Energy Minimization in Vision, IEEE Transactions on Pattern Analysis and Machine Intelligence, V. 26, No. 9, pp. 1124-1137, 2004.

[31] V. Papavassiliou, T. Stafylakis, V. Katsouros, G. Carayannis, Handwritten Document Image Segmentation into Text Lines and Words, Pattern Recognition, V. 43, No 1, 2010.

[32] R. O. Duda, P.E. Hart & D.G. Stork, Pattern Classification, (2nd ed.), 2001.

[33] Rakesh Dugad, U. B. Desai, A Tutorial on Hidden Markov Models, Technical Report No: SPANN-96.1, Signal Processing and Artificial Neural Networks Laboratory, Department of Electrical Engineering, Indian Institute of Technology – Bombay, India, 1996.

[34] C. Gonzalez & R. E. Woods, Digital Image Processing, Prentice-Hall, 2002

[35] T. Pavlidis, Algorithms for Graphics and Image Processing, pp. 195-214, Computer Science Press, Rockville USA, 1982.

[36] P. Soille, Morphological Image Analysis Principles and Applications, (2nd ed.), Springer 2004.

[37] I. Phillips, A. Chhabra, Empirical Performance Evaluation of Graphics Recognition Systems, IEEE Transactions on Pattern Analysis and Machine Intelligence, 21, No. 9, (1999), pp. 849-870.

[38] B. Gatos, N. Stamatopoulos & G. Louloudis, ICDAR2009 Handwriting Segmentation Contest, in Proc. of Int'l Conf. on Document Analysis and Recognition, pp. 1393-1397, 2009.

[39] A.K. Das, A. Gupta & B. Chanda, A Fast Algorithm for Text Line & Word Extraction from Handwritten Documents, Image Processing & Communications, V. 3, No. 1-2, pp. 85-94, 1997

[40] F. Yin & C.L. Liu, Handwritten Text Line Segmentation by Clustering with Distance Metric Learning, in Proc. of Int'l Conf. on Frontiers in Handwriting Recognition, pp.229-234, 2008.

[41] E. Kavallieratou, N. Dromazou, N. Fakotakis & G. Kokkinakis, An Intergrated System for Handwritten Document Image Processing, International Journal of Pattern Recongition and Artificial Intelligence, V. 17, N. 4, pp. 101-120, 2003.

[42] J.S. Cardoso, A. Capela, A. Rebelo & C. Guedes, A Connected Path Approach for Staff Detection on a Music Score, in Proc. of Int'l Conf. on Image Processing, pp.1005-1008, 2008.

[43] Z. Shi, S. Seltur, V. Govindaraju, A Steerable Directional Local Profile Technique for Extraction of Arabic Text Lines, in Proc. of Int'l Conf. on Document Analysis and Recognition, pp. 176-180, 2009.

[44] www.leptonica.com

[45] J. Serra, Image Analysis and Mathematical Morphology, Academic Press Inc. 1982.

[46] D.S. Bloomberg & P. Maragos, Generalized Hit-Miss Operations, in Proc. of SPIE Conf. Image Analysis and Morphological Image Processing, pp. 116-128, 1990.

[47] P. Soille, On morphological operators based on rank filters, Pattern Recognition, V. 35, Is. 2, pp. 527-535, 2002

[48] P. Maragos & R. W. Schafer, Morphological Filters-Part II, Their Relations to Median, Order-Statistic and Stack Filters, IEEE Transactions on Acoustics, Speech, and Signal Processing, V. ASSP-35, No. 8, 1987.

[49] D. S. Bloomberg, Image analysis using threshold reduction, in Proc. of SPIE Conf. Image Algebra and Morphological Image Processing II Conference, pp. 38-52, 1991.

[50] D. S. Bloomberg, Textured Reductions for Document Image Analysis, in Proc. of IS&T/SPIE EI '96, Conference 2660: Document Recognition III, pp. 160-174,1996.

[51] L. Vincent & P. Soille, Watershed in Digital Spaces: An Efficient Algorithm Based on Immersion Simulations, IEEE Transactions on Pattern Analysis and Machine Intelligence, V. 13, No. 6, pp. 583-598, 1991.

[52] D.S. Bloomberg, Multiresolution Morphological Analysis of Document Images, in Proc. of SPIE Visual Communications and Image Processing, V. 1818, pp. 648-662, 1992.

[53] G. Seni & E. Cohen, External word segmentation of off-line handwritten text lines, Pattern Recognition, 27, (1994), pp. 41-52.

[54] S. H. Kim, C. B. Joeng, H. K. Kwag & C. Y. Chen, Word Segmentation of Printed Text Lines on Gap Clustering and Special Symbol Detection, in Proc. of Int'l Conf. on Pattern Recognition, Vol. 2, pp. 320-323, 2002.

[55] U.V. Mart & H. Bunke, Text Line Segmentation and Word Recognition in a System for General Writer Independent Handwriting Recognition, in Proc. of Int'l Conf. on Document Analysis and Recognition, pp. 159-163, 2001.

[56] U.V. Marti & H. Bunke, The IAM-Database: an English sentence database for off-line handwriting recognition, Int'l Journal on Document Analysis and Recognition, V.5, No. 1, pp. 39-46, 2002.

[57] R. Manmatha & J. L. Rothfeder, A Scale Space Approach for Automatically Segmenting Words from Historical Handwritten Documents, IEEE Transactions on Pattern Analysis and Machine Intelligence, V. 27, No.8, pp. 1212-1225, 2005.

[58] T. Lindeberg & J.O. Eklundh, Scale-Space Primal Sketch: construction and experiments, Image and Vision Computing, V. 10, No.1, pp. 3–18, 1992.

[59] C.M. Bishop, Pattern Recognition and Machine Learning, Springer 2006.

[60] S. S. Rao, Engineering Optimization, Wiley, 1996.

[61] C.J.C. Burges, A Tutorial on Support Vector Machines for Pattern Recognition, Data Mining and Knowledge Discovery, Is. 2, pp. 121-167, 1998.

[62] J. Nocedal & S.J. Wright, Numerical Optimization, Springer, 2006.

[63] E. Alpaydin, Introduction to Machine Learning, The MIT Press, Cambridge USA, 2004.

[64] A.W. Bowman & A. Azzalini, Applied Smoothing Techniques for Data Analysis, Oxford University Press, pp. 25-46, 1997.

[65] http://www.lrde.epita.fr/cgi-bin/twiki/view/Olena/ModuleIcdar

[66] K. Jung, K. I. Kim & A. K. Jain, Text information extraction in images and video: a survey, Pattern Recognition, V. 35, Is. 5, pp. 977-997, 2004.

[67] T. Sato, T. Kanade, E. Hughes, M. Smith & S. Satoh, Video OCR: Indexing Digital News Libraries by Recognition of Superimposed Captions, Multimedia Systems, V. 7, No. 5, pp. 385-395, 1999.

[68] I. Demiros, G Carayannis, V. Antonopoulos, G. Kambourakis, V. Katsouros, P. Kolevris, M. Nottas, H. Papageorgiou, V. Papavasiliou, S. Raptis, F. Simistira & T. Stafylakis, PANOPTIS: A System for Intelligent Monitoring of the Hellenic Broadcast Sector, in Proc. of Int'l Conf. on Database and Expert Systems Application, pp. 605-609, 2008.

[69] C. Snoek & M. Worring, Multimodal Video Indexing: A Review of the state-of-the-art, Multimedia Tools and Applications, V.25, No. 1, pp. 5-35, 2005.

[70] R. Lienhart, Video OCR: A Survey and Practitioner's Guide in Video Mining, Kluwer Academic Publisher, pp. 155-184, 2003.

[71] J.C. Shim, C. Dorai & R. Bolle, Automatic Text Extraction from Video for Content-based Annotation and Retrieval, in Proc. of Int'l Conference on Pattern Recognition, V. 1, pp. 618-620, 1998.

[72] Y. Hassan & L. Karam, Morphological Text Extraction from Images, IEEE Transactions on Image Processing, V. 9, Is. 11, pp. 1978-1983, 2000.

[73] J. Canny, A Computational Approach for Edge Detection, IEEE Transactions on Pattern Analysis and Machine Intelligence, V. 8, No. 6, pp. 679-698, 1986

[74] V. Wu, R. Manmatha, E.M. Riseman, Textfinder: An Automatic System to Detect and Recognize Text in Images, IEEE Transactions on Pattern Analysis and Machine Intelligence, V. 21, Is. 11, pp. 1224-1229, 1999.

[75] H. Li, D. Doerman & O. Kia, Automatic Text Detection and Tracking in Digital Video, IEEE Transactions on Image Processing, V. 9, No.1, pp. 147-156, 2000.

[76] A. Wernicke, R. Lienhart, On the Segmentation of Text in Videos, in Proc. of IEEE Int'l Conference on Multimedia and Expo (ICME2000), Vol. 3, pp. 1511-1514, 2000.

[77] C.W. Lee, K. Jung, H.J. Kim, Automatic Text Detection and Removal in Video Sequences, Pattern Recognition Letters, V. 24, pp. 2607-2623, 2003.

[78] C. Wolf, J.M. Jolion, Extraction and Recognition of Artificial Text in Multimedia Documents, Technical Report RFV-RR-2002.01, Laboratoire Reconnaissance de Formes et Vision, INSA de Lyon, France, February 2002.

[79] S.M. Lucas et al., ICDAR 2003 Robust Reading Competitions: Entries, Results and Future Directions, Int'l Journal on Document Analysis and Recognition, V. 7, No. 2-3, pp. 105-122, 2005.

[80] D. Chen, J.M. Odobez & J.P. Thiran, A localization/verification scheme for finding text in images and video frames based on contrast independent features and machine learning methods, Signal Processing: Image Communication, ELSEVIER, V. 19, pp. 205-217, 2004.

[81] V. Papavassiliou, T. Stafylakis, V. Katsouros, G. Carayannis, A Parametric Spectral-Based Method for Verification of Text in Videos, in Proc. of Inte'l Conf. on Document Analysis and Recognition, pp. 879-883, 2007.

[82] G. Peeters, A Large Set of Audio Features for Sound Description, Technical report published by IRCAM, 2003.

[83] J.A. Bilmes, A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models, Technical Report ICSI-TR-97-02, University of Berkeley, 1997.

Παράρτημα Α

Παράδειγμα υπολογισμού των νέων διαχωριστικών

Μετά τον αρχικό διαχωρισμό της εικόνας κειμένου σε περιοχές κειμένου και κενών, η διόρθωση των πιθανών αστοχιών θα γίνει με την εφαρμογή του προτεινόμενου HMM, όπως περιγράφεται στο κεφάλαιο 1.2.4. Οι στατιστικές τιμές για τις δύο τάξεις είναι:

$$\mu_1 = -3.9395,\ \sigma_1 = 0.71339,\ \mu_2 = -5.1631,\ \sigma_2 = 0.43015,\ m_1 = 31.985,\ m_2 = 60.888$$

όπου μ_1, μ_2, σ_1 και σ_2 οι μέσες τιμές και οι τυπικές αποκλίσεις των πυκνοτήτων των περιοχών κειμένου και κενού αντίστοιχα και m_1 και m_2 οι μέσες τιμές των υψών. Θα εφαρμόσουμε τον αλγόριθμο Viterbi για τη 18η ζώνη της εικόνας 067.tif από ICDAR07. Οι αρχικές πιθανότητες ορίζονται ίσες με 0.5. Οι περιοχές-παρατηρήσεις και οι αντίστοιχες τιμές των πιθανοτήτων παρατήρησης κατάστασης και μετάβασης σε λογαριθμική κλίμακα παρουσιάζονται στον ακόλουθο πίνακα:

Περιοχή	Πυκν. (ln)*	ύψος	ln(b_1(i))	ln(b_2(i))	ln(α_{11}(i))	ln(α_{12}(i))	ln(α_{22}(i))	ln(α_{21}(i))	i
	-4.4628	30	-0.8502	-1.4006	-0.93793	-0.49666	-0.49271	-0.94409	**1**
	-5.2462	19	-2.2587	-0.093979	-0.59402	-0.80319	-0.31205	-1.3166	**2**
	-4.9424	24	-1.5694	-0.2069	-0.75034	-0.63905	-0.39417	-1.1216	**3**
	-5.058	56	-1.8102	-0.10517	-1.7508	-0.19072	-0.91973	-0.50854	**4**
	-3.2388	31	-1.0636	-10.081	-0.96919	-0.47705	-0.50913	-0.91883	**5**
	-5.0856	63	-1.8716	-0.091562	-1.9697	-0.15025	-1.0347	-0.43903	**6**
	-4.016	33	-0.5869	-3.6308	-1.0317	-0.44067	-0.54198	-0.8713	**7**
	-4.9514	26	-1.5872	-0.1964	-0.81287	-0.58624	-0.42702	-1.0569	**8**
	-4.6334	24	-1.0543	-0.83347	-0.75034	-0.63905	-0.39417	-1.1216	**9**
	-5.5652	27	-3.1778	-0.51233	-0.84414	-0.56199	-0.44344	-1.0267	**10**
	-4.0953	30	-0.6050	-3.1566	-0.93793	-0.49666	-0.49271	-0.94409	**11**
	-5.1621	43	-2.0498	-0.075319	-1.3444	-0.30206	-0.70622	-0.68024	**12**
	-3.9186	46	-0.5816	-4.2603	-1.4382	-0.27098	-0.75549	-0.63446	**13**
	-5.8081	63	-4.0116	-1.1996	-1.9697	-0.15025	-1.0347	-0.43903	**14**
	-3.6148	31	-0.6847	-6.5529	-0.96919	-0.47705	-0.50913	-0.91883	**15**
	-4.6442	69	-1.0691	-0.80288	-2.1572	-0.1229	-1.1332	-0.38859	**16**
	-3.5799	34	-0.7082	-6.8484	-1.063	-0.42377	-0.55841	-0.84891	**17**
	-5.4639	45	-2.8643	-0.31991	-1.4069	-0.28091	-0.73907	-0.64925	**18**
	-5.0651	29	-1.8265	-0.10124	-0.90666	-0.51729	-0.47629	-0.97044	**19**
	-5.5534	68	-3.1401	-0.48696	-2.126	-0.12706	-1.1168	-0.39649	**20**
	-3.6091	34	-0.688	-6.6013					**21**

Κατά την εφαρμογή του αλγόριθμου Viterbi, εισάγεται η ποσότητα $\delta_i(j)$ που δηλώνει το

υπολογιζόμενο «κόστος» να «βρισκόμαστε» στην κατάσταση j κατά την i-οστή παρατήρηση. Επίσης, εισάγεται η ποσότητα $\psi_i(j)$ που δηλώνει την κατάσταση για την (i-1)-ιοστή παρατήρηση που έχει το μικρότερο κόστος για να «μεταβούμε» στην κατάσταση j για την i-ιοστή παρατήρηση. Εφαρμόζοντας τα δύο πρώτα βήματα του αλγόριθμου Viterbi:

1. $\delta_1(1) = \pi_1 \cdot b_1(1)$, $\delta_1(2) = \pi_2 \cdot b_2(1)$ ή αλλιώς $\delta_1(1) = \ln(\pi_1) + \ln(b_1(1))$, $\delta_1(2) = \ln(\pi_2) + \ln(b_2(1))$.

Επίσης, $\psi_1(1) = \psi_1(2) = 0$.

2. $\delta_i(j) = \max_k \delta_{i-1}(k) a_{kj}(i) \cdot b_j(i)$ $\psi_i(j) = \arg\max_k \delta_{i-1}(k) \cdot a_{kj}(i)$ $2 \le i \le 21,\ j,k = 1,2$

προκύπτουν οι ακόλουθες τιμές:

i	$\ln(\delta_{i-1}(1)) + \ln(a_{11}(i))$	$\ln(\delta_{i-1}(2)) + \ln(a_{21}(i))$	$\ln(\delta_{i-1}(1)) + \ln(a_{12}(i))$	$\ln(\delta_{i-1}(2)) + \ln(a_{22}(i))$	$\delta_i(1)$	$\delta_i(2)$	$\psi_i(1)$	$\psi_i(2)$	q_i^*
1					-1.5434	-2.0937	0	0	1
2	-2.4813	-3.0378	-2.04	-2.5864	-4.74	-2.134	1	1	2
3	-5.3340	-3.4506	-5.5431	-2.4461	-5.0199	-2.6530	2	2	2
4	-5.7703	-3.7746	-5.659	-3.0471	-5.5848	-3,1523	2	2	2
5	-7.3356	-3.6608	-5.7755	-4.072	-4,7244	-14,153	2	2	1
6	-5.6936	-15.0720	-5.2014	-14.6623	-7,5652	-5,293	1	1	2
7	-9.5348	-5.7320	-7.7154	-6.3277	-6,319	-9,9585	2	2	1
8	-7.3507	-10.8298	-6.7597	-10.5005	-8,9379	-6,9561	1	1	2
9	-9.7508	-8.0129	-9.5241	-7.3831	-9,0672	-8,2165	2	2	2
10	-9.8175	-9.3381	-9.7062	-8.6107	-12,516	-9,123	2	2	2
11	-13.3601	-10.1498	-13.0779	-9.5665	-10,755	-12,723	2	2	1
12	-11.6928	-13.6672	-11.2515	-13.2158	-13,743	-11,327	1	1	2
13	-15.0869	-12.0070	-14.0446	-12.0330	-12,589	-16,293	2	2	1
14	-14.0268	-16.9278	-12.8597	-17.0488	-18,038	-14,059	1	1	2
15	-20.0080	-14.4983	-18.1886	-15.0939	-15,183	-21,647	2	2	1
16	-16.1522	-22.5656	-15.6601	-22.1559	-17,221	-16,463	1	1	2
17	-19.3786	-16.8516	-17.3442	-17.5962	-17,56	-24,193	2	1	1
18	-18.6228	-25.0415	-17.9836	-24.7510	-21,487	-18,303	1	1	2
19	-22.8940	-18.9527	-21.7680	-19.0425	-20,779	-19,144	2	2	2
20	-21.6854	-20.1142	-21.2960	-19.6201	-23,254	-20,107	2	2	2
21	-25.3803	-20.5035	-23.3814	-21.2239	-21,192	-27,825	2	2	1

Από τα δύο επόμενα βήματα:

3. $p^* = \max_j \delta_{21}(j)$ $q_{21}^* = \arg\max_j \delta_{21}(j),\ \ j = 1,2$ έχουμε p^* =-21,192 και q_{21}^*=1

4. $q_i^* = \psi_{i+1}(q_{i+1}^*),\ \ i = 20,19,...,1$

προκύπτει ως πιο πιθανή ακολουθία καταστάσεων για την ακολουθία παρατηρήσεων (ζώνη 18) η 12221212221212121221.

Επομένως, η νέα κατηγοριοποίηση είναι: TGGGTGTGGGTGTGTGTGGGT

Παράρτημα Β

Δημοσιεύσεις – Διακρίσεις

V. Papavassiliou, T. Stafylakis, V. Katsouros, G. Carayannis, Handwritten Document Image Segmentation into Text Lines and Words, Pattern Recognition, V. 43, No 1, 2010.

V. Papavassiliou, V. Katsouros, G. Carayannis, A Morphological Approach for Text-Line Segmentation in Handwritten Documents, in Proc. of Int'l Conf. of Frontiers on Handwriting, 2010.

I. Demiros, G Carayannis, V. Antonopoulos, G. Kambourakis, V. Katsouros, P. Kolevris, M. Nottas, H. Papageorgiou, V. Papavasiliou, S. Raptis, F. Simistira & T. Stafylakis, PANOPTIS: A System for Intelligent Monitoring of the Hellenic Broadcast Sector, in Proc. of Int'l Conf. on Database and Expert Systems Application, pp. 605-609, 2008.

T. Stafylakis, V. Papavassiliou, V. Katsouros, G. Carayannis, Robust text-line and word segmentation for handwritten documents images, in Proc. of IEEE Int'l Conf. on Acoustics, Speech and Signal Processing ICASSP 2008, pp. 3393-3396.

V. Papavassiliou, T. Stafylakis, V. Katsouros, G. Carayannis, A Parametric Spectral-Based Method for Verification of Text in Videos, in Proc. of Int'l Conf. on Document Analysis and Recognition, pp. 879-883, 2007.

Β. Παπαβασιλείου, Θ. Σταφυλάκης, Β. Κατσούρος, Γ. Καραγιάννης, Ανίχνευση Κειμένου σε Βίντεο, Πρακτικά 1ου Πανελλήνιου Συνέδριου Φοιτητών Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών ΣΦΗΜΜΥ 2007, Αθήνα, 27-28 Μαίου 2007. (ISBN: 978-960-89028-6-2) σελ. 178-184.

Β. Παπαβασιλείου, Θ. Σταφυλάκης, Β. Κατσούρος, Γ. Καραγιάννης, Κατάτμηση χειρόγραφου κειμένου σε γραμμές και λέξεις, Πρακτικά 1ου Πανελλήνιου Συνέδριου Φοιτητών Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών ΣΦΗΜΜΥ 2007, Αθήνα, 27-28 Μαίου 2007. (ISBN: 978-960-89028-6-2) σελ. 185-189.

Μέλος της ερευνητικής ομάδας του ΙΕΛ (Β. Παπαβασιλείου, Θ. Σταφυλάκης, Β. Κατσούρος, Γ. Καραγιάννης) που συμμετείχε στο διεθνή διαγωνισμό ICDAR2007 handwriting segmentation contest ($1^{η}$ θέση για την κατάτμηση χειρόγραφου κειμένου σε γραμμές και λέξεις). Σύντομη παρουσίαση των αλγορίθμων και των αποτελεσμάτων δημοσιεύονται στην εργασία B. Gatos, A. Antonacopoulos & N. Stamatopoulos, ICDAR2007 handwriting segmentation contest, in Proc. of Int'l Conf. on Document Analysis and Recognition, pp. 1284-1288, 2007.

Μέλος της ερευνητικής ομάδας του ΙΕΛ (Β. Παπαβασιλείου, Θ. Σταφυλάκης, Β. Κατσούρος, Γ. Καραγιάννης) που συμμετείχε στο διεθνή διαγωνισμό ICDAR2009 handwriting segmentation contest ($2^{η}$ θέση για την κατάτμηση χειρόγραφου κειμένου σε γραμμές και $1^{η}$ θέση για την

κατάτμηση χειρόγραφου κειμένου λέξεις). Σύντομη παρουσίαση των αλγορίθμων και των αποτελεσμάτων δημοσιεύονται στην εργασία B. Gatos, N. Stamatopoulos & G. Louloudis, ICDAR2009 Handwriting Segmentation Contest, in Proc. of Int'l Conf. on Document Analysis and Recognition, pp. 1393-1397, 2009.