



ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

**ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ**

ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

Οπτική Επεξεργασία και Ανάλυση Ιστορικών Εγγράφων

Νικόλαος Σ. Σταματόπουλος

ΑΘΗΝΑ

ΣΕΠΤΕΜΒΡΙΟΣ 2011

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

Οπτική Επεξεργασία και Ανάλυση Ιστορικών Εγγράφων

Νικόλαος Σ. Σταματόπουλος

ΕΠΙΒΛΕΠΩΝ ΚΑΘΗΓΗΤΗΣ: Σέργιος Θεοδωρίδης, Καθηγητής ΕΚΠΑ

ΤΡΙΜΕΛΗΣ ΕΠΙΤΡΟΠΗ ΠΑΡΑΚΟΛΟΥΘΗΣΗΣ:

Σέργιος Θεοδωρίδης, Καθηγητής ΕΚΠΑ

Βασίλειος Γάτος, Ερευνητής Β' ΕΚΕΦΕ «Δημόκριτος»

Νικόλαος Παπαμάρκος, Καθηγητής ΔΠΘ

ΕΠΤΑΜΕΛΗΣ ΕΞΕΤΑΣΤΙΚΗ ΕΠΙΤΡΟΠΗ

**Σέργιος Θεοδωρίδης
Καθηγητής ΕΚΠΑ**

**Βασίλειος Γάτος
Ερευνητής Β' ΕΚΕΦΕ «Δημόκριτος»**

**Νικόλαος Παπαμάρκος
Καθηγητής ΔΠΘ**

**Σταύρος Περαντώνης
Ερευνητής Α' ΕΚΕΦΕ «Δημόκριτος»**

**Ιωάννης Πρατικάκης
Επίκουρος Καθηγητής ΔΠΘ**

**Γεώργιος Κουρουπέτρογλου
Αναπληρωτής Καθηγητής ΕΚΠΑ**

**Άγγελος Πικράκης
Λέκτορας ΠΑΠΕΙ**

Ημερομηνία εξέτασης 29/09/2011

ΠΕΡΙΛΗΨΗ

Η έρευνα ιστορικών βιβλίων αποτελεί μία σημαντική πηγή πληροφοριών, τόσο σε ότι αφορά την ιστορία των προηγούμενων περιόδων όσο και την ανάπτυξη, διαχρονικά, της ίδιας της πολιτιστικής τεκμηρίωσης. Μέχρι σήμερα έχουν γίνει αρκετές προσπάθειες ψηφιοποίησης και ηλεκτρονικής πλοήγησης, εντούτοις όμως δεν υπάρχει κατάλληλο περιβάλλον οπτικής επεξεργασίας και ανάλυσης του περιεχομένου των συλλογών αυτών. Κατά συνέπεια μεγάλος αριθμός ιστορικών βιβλίων δεν έχει μελετηθεί ακόμα και παραμένει ανεκμετάλλευτος. Η παρούσα διδακτορική διατριβή επικεντρώνεται στα στάδια επεξεργασίας των ιστορικών εγγράφων πριν την αναγνώριση και ειδικότερα στα στάδια βελτίωσης της ποιότητας και κατάτμησης του εγγράφου. Τα παραπάνω στάδια χρήζουν ιδιαίτερης προσοχής αφού επηρεάζουν άμεσα το αποτέλεσμα της αναγνώρισης. Στο στάδιο της βελτίωσης ποιότητας δόθηκε έμφαση στον εντοπισμό και την αφαίρεση του περιθωρίου καθώς και στη διόρθωση της τοπικής καμπυλότητας, τα οποία αποτελούν τα κατεξοχήν προβλήματα των ιστορικών εγγράφων. Πιο συγκεκριμένα, για τον εντοπισμό και την αφαίρεση του περιθωρίου αναπτύχθηκαν δύο μέθοδοι. Η πρώτη μέθοδος αφαιρεί συγχρόνως με το περιθώριο και τμήματα κειμένου γειτονικών σελίδων, ενώ η δεύτερη μέθοδος διαχωρίζει τα έγγραφα που περιέχουν δύο σελίδες καθώς η πλειοψηφία των τεχνικών επεξεργασίας εγγράφων θεωρεί ότι επεξεργάζονται μία σελίδα. Τα πειραματικά αποτελέσματα καταδεικνύουν την αποτελεσματικότητα των μεθόδων εφόσον καταφέρνουν να αφαιρέσουν το περιθώριο χωρίς να χάνεται χρήσιμη πληροφορία. Για τη διόρθωση της τοπικής καμπυλότητας αναπτύχθηκε μία μέθοδος η οποία εφαρμόζεται σε δύο φάσεις επιτυγχάνοντας να διορθώσει όλες τις παραμορφώσεις σε επίπεδο λέξης. Η αποτίμηση της μεθόδου βασίζεται, εκτός από τους έμμεσους τρόπους που συναντάμε στη βιβλιογραφία, και σε μία πρωτότυπη ημιαυτόματη μέθοδο η οποία δίνει τη δυνατότητα να συγκρίνονται άμεσα μέθοδοι διόρθωσης της τοπικής καμπυλότητας. Στο στάδιο της κατάτμησης αναπτύχθηκε μία πρωτότυπη μεθοδολογία συνδυασμού αποτελεσμάτων κατάτμησης γραμμών κειμένου, η οποία προσπαθεί να εκμεταλλευτεί συμπληρωματικές τεχνικές ώστε να αντιμετωπιστούν τα πολλά και σύνθετα προβλήματα που συναντάμε στα ιστορικά έγγραφα. Τα πειραματικά αποτελέσματα είναι ενθαρρυντικά για την εφαρμογή της μεθοδολογίας και σε άλλα επίπεδα κατάτμησης. Επιπρόσθετα, δημιουργήθηκε μία μέθοδος κατάτμησης χαρακτήρων σε ιστορικά έγγραφα. Τα πειραματικά αποτελέσματα σε σημαντικό αριθμό ιστορικών εγγράφων από διαφορετικές γλώσσες αλλά και χρονολογικές περιόδους επιβεβαιώνουν την αποτελεσματικότητα της. Τέλος, αναπτύχθηκε μία μέθοδος με στόχο να μειωθεί σημαντικά ο χρόνος που χρειάζεται για τη δημιουργία ground-truth αξιολόγησης μεθόδων κατάτμησης.

ΘΕΜΑΤΙΚΗ ΠΕΡΙΟΧΗ: Επεξεργασία Ιστορικών Εγγράφων

ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ: βελτίωση ποιότητας εγγράφου, αφαίρεση περιθωρίου, διόρθωση τοπικής καμπυλότητας, κατάτμηση εγγράφου, συνδυασμός μεθόδων κατάτμησης

ABSTRACT

The collections of historical books are an important source of information, both for the history of previous periods and for the development of the cultural documentation itself. Although to date, there have been made several attempts of digitalization and electronic navigation, there is not an appropriate frame of optical process and analysis of the content of these collections, consequently a large number of historical books have not been studied yet and remain unexploited. In this thesis, we studied the preprocessing stages which are performed before the recognition process and we focused on the enhancement and segmentation of historical documents. Preprocessing stages play an important role in document image processing since they affect the performance of subsequent processing, such as optical character recognition. At the enhancement stage, we focused on the border removal as well as on the dewarping of document images, which are common problems associated with historical documents. Two methodologies that detect and remove black borders as well as noisy text regions are proposed. Furthermore, optimal page frames of double page document images are detected since the majority of approaches are able to process only single page document images. The experimental results on several historical documents demonstrate the effectiveness of the proposed techniques since they remove the noisy borders without missing text information. Concerning the warping problem, a coarse-to-fine rectification methodology to compensate for undesirable document image distortions is proposed. To verify the validity of the proposed methodology, experiments have been carried out using indirect evaluation techniques as well as a novel semi-automatic evaluation methodology. Using the proposed evaluation methodology we can obtain objective evaluations and quantitative comparisons among the different dewarping techniques. At the document image segmentation stage we proposed a novel combination method of complementary text line segmentation techniques, where each technique can solve different difficult problems of historical problems. The experimental results are promising in order to apply the combination methodology in different segmentation tasks, such as word and character segmentation. Furthermore, a methodology for character segmentation in historical documents is suggested. Comparative experiments using several historical documents from different languages and time periods prove the efficiency of the proposed technique. Finally, in order to ease the construction of document image segmentation ground-truth that includes text-image alignment we presented an efficient technique.

SUBJECT AREA: Historical Document Processing

KEYWORDS: document image enhancement, border removal, document image dewarping, document image segmentation, combined segmentation techniques

ΕΥΧΑΡΙΣΤΙΕΣ

Θα ήθελα να ευχαριστήσω όλους εκείνους που βοήθησαν στην ολοκλήρωση της παρούσας διδακτορικής διατριβής. Ιδιαίτέρως, θα ήθελα να ευχαριστήσω τον ερευνητή του Ε.Κ.Ε.Φ.Ε. «Δημόκριτος» Β. Γάτο για την αμέριστη συμπαράσταση και βοήθεια του καθ' όλη τη διάρκεια της εκπόνησης της διατριβής αυτής. Επίσης, θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή Σ. Θεοδωρίδη καθώς και τον καθηγητή Ν. Παπαμάρκο, τον διευθυντή του Εργαστηρίου Υπολογιστικής Ευφυΐας του Ε.Κ.Ε.Φ.Ε. «Δημόκριτος» Σ. Περαντώνη και τον Επίκουρο Καθηγητή Ι. Πρατικάκη για τη συστηματική καθοδήγησή τους και την πολύτιμη βοήθειά τους όλα αυτά τα χρόνια.

ΠΕΡΙΕΧΟΜΕΝΑ

1. ΕΙΣΑΓΩΓΗ	17
1.1 Εισαγωγή στην Επεξεργασία και Ανάλυση Ιστορικών Εγγράφων	17
1.2 Στόχοι της Διδακτορικής Διατριβής.....	20
1.3 Συνεισφορά της Διδακτορικής Διατριβής	20
2. ΤΡΕΧΟΥΣΑ ΤΕΧΝΟΛΟΓΙΚΗ ΣΤΑΘΜΗ ΣΤΗΝ ΕΠΕΞΕΡΓΑΣΙΑ ΙΣΤΟΡΙΚΩΝ ΕΓΓΡΑΦΩΝ	25
2.1 Βελτίωση Ποιότητας	25
2.1.1 Δυαδική Μετατροπή.....	25
2.1.2 Αφαίρεση Θορύβου	28
2.1.3 Εντοπισμός και Αφαίρεση Περιθωρίου	31
2.1.4 Διόρθωση Κλίσης.....	35
2.1.5 Διόρθωση Τοπικής Καμπυλότητας	37
2.2 Ανάλυση Δομής και Κατάτμηση	46
2.2.1 Ανάλυση και Κατανόηση της Δομής του Εγγράφου	47
2.2.2 Κατάτμηση Γραμμών Κειμένου.....	49
2.2.3 Κατάτμηση Λέξεων	52
2.2.4 Κατάτμηση Χαρακτήρων.....	54
2.2.5 Αντιστοίχιση Εικόνας και Αρχείου Κειμένου	63
3. ΑΦΑΙΡΕΣΗ ΠΕΡΙΘΩΡΙΟΥ ΚΑΙ ΔΙΑΧΩΡΙΣΜΟΣ ΔΙΠΛΩΝ ΣΕΛΙΔΩΝ	71
3.1 Εισαγωγή.....	71
3.2 Εντοπισμός και Αφαίρεση Περιθωρίου καθώς και Τμημάτων Γειτονικής Σελίδας.....	71
3.3 Διαχωρισμός Διπλών Σελίδων	79
3.4 Πειραματικά Αποτελέσματα	84
4. ΔΙΟΡΘΩΣΗ ΤΟΠΙΚΗΣ ΚΑΜΠΥΛΟΤΗΤΑΣ	93
4.1 Εισαγωγή.....	93
4.2 Διόρθωση της Τοπικής Καμπυλότητας με Χρήση Κατάτμησης.....	94
4.2.1 Κατάτμηση Λέξεων και Γραμμών Κειμένου	94
4.2.2 Προσδιορισμός Γραμμών Αναφοράς των Λέξεων	96
4.2.3 Διόρθωση Παραμορφώσεων	97
4.3 Διόρθωση της Τοπικής Καμπυλότητας σε Δύο Φάσεις	99
4.3.1 Αρχική Προσέγγιση	101
4.3.2 Βελτιωμένη Προσέγγιση	106
4.4 Ημιαυτόματη Μέθοδος Αξιολόγησης.....	110
4.4.1 Μαρκάρισμα Εικόνας από τον Χρήστη	111
4.4.2 Αντιστοίχιση Σημείων.....	112
4.4.3 Μετρικές Αξιολόγησης	114
4.5 Πειραματικά Αποτελέσματα	116
5. ΚΑΤΑΤΜΗΣΗ ΓΡΑΜΜΩΝ ΚΕΙΜΕΝΟΥ	125

5.1	Εισαγωγή	125
5.2	Μεθοδολογία Συνδυασμού Αποτελεσμάτων Κατάτμησης Γραμμών Κειμένου	127
5.2.1	Γενική Περιγραφή	127
5.2.2	Ορισμοί	130
5.2.3	Μεθοδολογία	133
5.3	Πειραματικά Αποτελέσματα	140
6.	ΚΑΤΑΤΜΗΣΗ ΧΑΡΑΚΤΗΡΩΝ	145
6.1	Εισαγωγή	145
6.2	Αλγόριθμος Κατάτμησης Χαρακτήρων	145
6.3	Πειραματικά Αποτελέσματα	149
7.	ΜΕΘΟΔΟΣ ΓΡΗΓΟΡΗΣ ΔΗΜΙΟΥΡΓΙΑΣ GROUND-TRUTH	153
7.1	Εισαγωγή	153
7.2	Μέθοδος Δημιουργίας Ground-truth	153
7.2.1	Ανάλυση Αρχείου Κειμένου	154
7.2.2	Κατάτμηση Γραμμών Κειμένου	155
7.2.3	Κατάτμηση Λέξεων	157
7.2.4	Κατάτμηση Χαρακτήρων	159
7.2.5	Διόρθωση Αποτελεσμάτων	160
7.3	Πειραματικά Αποτελέσματα	160
8.	ΣΥΜΠΕΡΑΣΜΑΤΑ	165
	ΔΗΜΟΣΙΕΥΣΕΙΣ	169
	ΕΤΕΡΟΑΝΑΦΟΡΕΣ ΣΤΙΣ ΔΗΜΟΣΙΕΥΣΕΙΣ	173
	ΠΙΝΑΚΑΣ ΟΡΟΛΟΓΙΑΣ	179
	ΣΥΝΤΜΗΣΕΙΣ – ΑΡΚΤΙΚΟΛΕΞΑ – ΑΚΡΩΝΥΜΙΑ	183
	ΑΝΑΦΟΡΕΣ	185

ΠΡΟΛΟΓΟΣ

Ο Νικόλαος Σταματόπουλος αποφοίτησε από το Τμήμα Πληροφορικής και Τηλεπικοινωνιών του Εθνικού Καποδιστριακού Πανεπιστημίου Αθηνών τον Ιούνιο του 2006, ενώ το Νοέμβριο του ίδιου έτους ανακηρύχτηκε υποψήφιος διδάκτορας στο εν λόγω τμήμα. Η διδακτορική του διατριβή εκπονήθηκε σε συνεργασία με το Ινστιτούτο Πληροφορικής και Τηλεπικοινωνιών του Ε.Κ.Ε.Φ.Ε. «Δημόκριτος». Παράλληλα με την ερευνητική του δραστηριότητα στον τομέα της Οπτικής Επεξεργασίας και Ανάλυσης Ιστορικών Εγγράφων, συμμετείχε και σε δυο ερευνητικά έργα χρηματοδοτούμενα από τη Γενική Γραμματεία Έρευνας και Τεχνολογίας (έργο «Πολύτιμο»)*¹ και από την Ευρωπαϊκή Ένωση (έργο «IMPACT»)*², αντίστοιχα.

(*¹)Υπουργείο Ανάπτυξης, Γενική Γραμματεία Έρευνας και Τεχνολογίας (Γ.Γ.Ε.Τ.), Επιχειρησιακό Πρόγραμμα Επεξεργασίας Εικόνων, Ήχου & Γλώσσας, Σύστημα Επεξεργασίας, Διαχείρισης και Παροχής Πρόσβασης στο Περιεχόμενο Πολύτιμων Βιβλίων & Χειρογράφων («Πολύτιμο»).

(*²)IMProving ACcess to Text (IMPACT), FP7-ICT-2007-1, Digital libraries and technology-enhanced learning, Large-scale integrating project – IP. - <http://www.impact-project.eu/> .

1. ΕΙΣΑΓΩΓΗ

1.1 Εισαγωγή στην Επεξεργασία και Ανάλυση Ιστορικών Εγγράφων

Η έρευνα και η μελέτη ιστορικών βιβλίων αποτελούν την κύρια πηγή πληροφόρησης και τεκμηρίωσης τόσο για την επιστήμη της ιστορίας και τις ανθρωπιστικές επιστήμες, όσο και για τον απλό αναγνώστη. Η λεπτομερής εξέταση και παρουσίαση ενός εγγράφου είναι προϊόν μίας εξαιρετικά χρονοβόρας και πολύπλοκης διαδικασίας για την οποία πολύ συχνά απαιτείται η μετακίνηση του ερευνητή στον τόπο φύλαξης του υλικού. Είναι επόμενο λοιπόν η πλειονότητα των σημερινών αναγνωστών να αναζητά ευκολότερους τρόπους πρόσβασης στην πληροφορία, όπως το διαδίκτυο.

Διεθνώς αναγνωρισμένες βιβλιοθήκες επιδιώκουν την ψηφιοποίηση ιστορικών βιβλίων. Από αυτή τη διαδικασία δύναται ένας ικανοποιητικός αριθμός αναγνωστών να έχει πρόσβαση σε ιστορικά βιβλία που μέχρι σήμερα ήταν ανεκμετάλλευτα ενώ συγχρόνως συμβάλει στην προστασία και τη διαφύλαξη του. Ωστόσο, η πρόσβαση στην εικόνα ενός ψηφιοποιημένου εγγράφου και μόνο δεν είναι ικανοποιητική. Η μετατροπή των ιστορικών εγγράφων σε ηλεκτρονικά κείμενα δίνει τη δυνατότητα στον χρήστη να εφαρμόσει μία πληθώρα εργασιών όπως αναζήτηση, τοποθέτηση ετικέτας (tagging), επεξεργασία (αντιγραφή και επικόλληση), δημιουργία ευρετηρίου (indexing), ανάγνωση σε φορητές συσκευές κ.α. έτσι ώστε να διευκολύνεται η μελέτη και η πρόσβαση στο πολύτιμο ιστορικό υλικό.

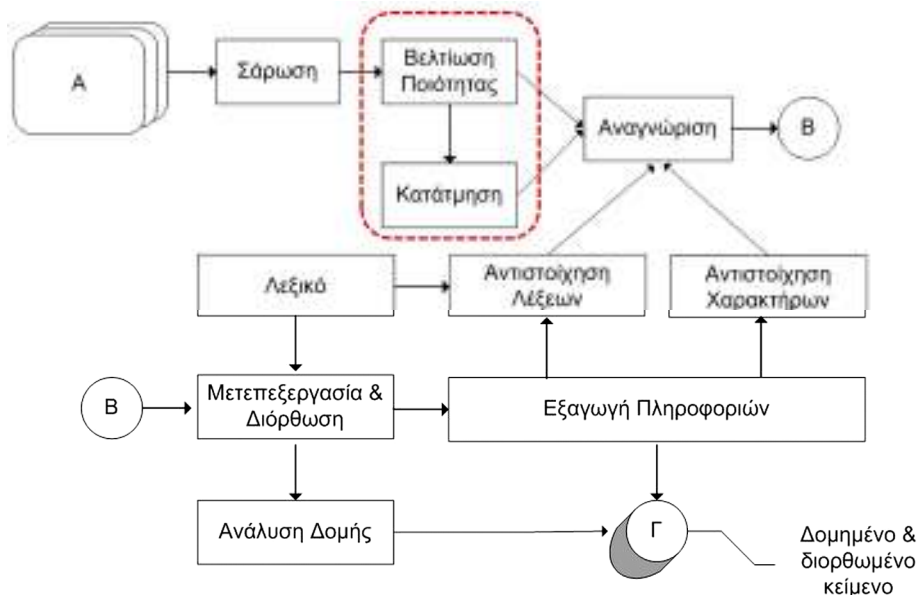
Ένα σύστημα Οπτικής Επεξεργασίας και Αναγνώρισης Εγγράφων καταστά εφικτή την επεξεργασία και αναγνώριση σύγχρονων εγγράφων με μεγάλη επιτυχία, όμως, στην περίπτωση των ιστορικών εγγράφων τα ποσοστά επιτυχίας δεν είναι το ίδιο ικανοποιητικά. Παρά την εκτεταμένη και ενδελεχή έρευνα που έχει πραγματοποιηθεί τα τελευταία χρόνια για την επίλυση του παραπάνω προβλήματος παραμένει ακόμα «ανοικτό» στην επιστημονική κοινότητα. Τα ειδικά χαρακτηριστικά των ιστορικών εγγράφων δυσκολεύουν την αυτόματη επεξεργασία και αναγνώρισή τους (Σχήμα 1.1). Για παράδειγμα, υπάρχουν διάφορα προβλήματα στην ποιότητα της εικόνας λόγω χαμηλής ποιότητας του αρχικού εγγράφου ή κακής ψηφιοποίησης ή παλαιότητας του εγγράφου όπως σκιές, μη ομοιόμορφη φωτεινότητα υπόβαθρου, έντονος θόρυβος, εμφάνιση περιθωρίου, παραμόρφωση, δυσδιάκριτες περιοχές κειμένου κ.α. Επιπλέον, εκτός από τα προβλήματα στην ποιότητα της εικόνας υπάρχουν και αλλά

ιδιαίτερα χαρακτηριστικά που επηρεάζουν την αυτόματη επεξεργασία και αναγνώριση όπως ασυνήθιστες γραμματοσειρές και πολύπλοκες δομές.



Σχήμα 1.1: Παραδείγματα εικόνων ιστορικών εγγράφων.

Ανεξάρτητα από τη μορφή του προβλήματος που έχουμε να αντιμετωπίσουμε, ιστορικά ή σύγχρονα έγγραφα, η γενική αρχιτεκτονική μιας ολοκληρωμένης διαδικασίας ψηφιοποίησης, η οποία ξεκινάει με τη σάρωση εγγράφων και καταλήγει σε δομημένα ηλεκτρονικά κείμενα, παρουσιάζεται στο Σχήμα 1.2.



Σχήμα 1.2: Η γενική αρχιτεκτονική μιας ολοκληρωμένης διαδικασίας ψηφιοποίησης εγγράφων.

Η **σάρωση** (scanning) των εγγράφων με τη βοήθεια μηχανημάτων οπτικής σάρωσης (scanner) αποτελεί ένα σημαντικό στάδιο καθώς καθορίζει σε μεγάλο βαθμό την ποιότητα της παραγόμενης εικόνας. Παράγοντες που πρέπει να

ληφθούν υπόψη στην επιλογή του σαρωτή είναι η οπτική ανάλυση, η περιοχή σάρωσης καθώς και ο χρόνος σάρωσης.

Η **βελτίωση της ποιότητας** του εγγράφου αποτελεί ένα ιδιαίτερα σημαντικό στάδιο, διότι βοηθάει στην επιτυχία των επόμενων σταδίων επεξεργασίας του, όπως της κατάτμησης και της αναγνώρισης. Το στάδιο αυτό περιλαμβάνει τη δυαδική μετατροπή και την ανάδειξη των περιοχών κειμένου (μετατροπή της εικόνας διαβάθμισης γκρι (grayscale) σε δυαδική), τον εντοπισμό και την αφαίρεση του περιθωρίου καθώς και τμημάτων γειτονικής σελίδας, τη διόρθωση της κλίσης και της τοπικής καμπυλότητας, την εξάλειψη θορύβου και την ανάδειξη μη ορατών περιοχών κειμένου (παλίμψηστα έγγραφα).

Στο στάδιο της **κατάτμησης** γίνεται ο εντοπισμός των βασικών συστατικών του εγγράφου (εικόνες, γραφικά, τμήματα κειμένου, παράγραφοι, γραμμές κειμένου, λέξεις, χαρακτήρες).

Το στάδιο της **αναγνώρισης** περιλαμβάνει τη μετατροπή των συστατικών του εγγράφου σε διανύσματα χαρακτηριστικών και τη δημιουργία ενός ταξινομητή για την κατάταξη κάθε συστατικού σε μία από τις γνωστές κλάσεις. Τα διανύσματα αυτά προκύπτουν από την ανάλυση των τοπολογικών, των γεωμετρικών, των στατιστικών ή των δομικών στοιχείων των συστατικών του εγγράφου.

Το στάδιο της **μετεπεξεργασίας** και της **διόρθωσης** περιλαμβάνει τον έλεγχο και τη διόρθωση του αποτελέσματος που προκύπτει από την αναγνώριση. Αυτό επιτυγχάνεται με τη χρήση **λεξικών** και την **εξαγωγή πληροφοριών** όπως για παράδειγμα το είδος της γραφής, την ανάλυση των συμπλεγμάτων, το γραφικό χαρακτήρα και την κλίση των γραμμάτων. Συνεπώς, η αναγνώριση βασίζεται στην **αντιστοίχιση λέξεων και χαρακτήρων** με προκαθορισμένα σύνολα λέξεων (που αναπαρίστανται με λεξικά) και χαρακτήρων.

Τέλος, η **ανάλυση της δομής** του εγγράφου μας βοηθάει στην εξαγωγή σημασιολογικής πληροφορίας και στην αυτόματη κατηγοριοποίηση των εγγράφων. Για παράδειγμα, ο αριθμός των στηλών του κειμένου, η ύπαρξη εικόνων σε συγκεκριμένη θέση, η εμφάνιση σημειώσεων ή παραπομπών, η ύπαρξη σφραγίδας κ.λπ. είναι στοιχεία τα οποία μας βοηθούν στην εξαγωγή σημασιολογικής πληροφορίας.

1.2 Στόχοι της Διδακτορικής Διατριβής

Η παρούσα διδακτορική διατριβή επικεντρώνεται στα στάδια επεξεργασίας των ιστορικών εγγράφων πριν την αναγνώριση και συγκεκριμένα στα στάδια βελτίωσης της ποιότητας και της κατάτμησης (διακεκομμένο πλαίσιο στο Σχήμα 1.2). Τα στάδια αυτά είναι πολύ σημαντικά αφού επηρεάζουν άμεσα το αποτέλεσμα της αναγνώρισης. Στα ιστορικά έγγραφα υπάρχουν προβλήματα στην ποιότητα της εικόνας τα οποία πρέπει να αντιμετωπιστούν για να έχουμε καλύτερα αποτελέσματα τόσο στο στάδιο της κατάτμησης όσο και στο στάδιο της αναγνώρισης. Από την άλλη, το αποτέλεσμα της κατάτμησης του εγγράφου στα βασικά του χαρακτηριστικά (γραμμές κειμένου, λέξεις και χαρακτήρες) επηρεάζει επίσης σε μεγάλο βαθμό το αποτέλεσμα της αναγνώρισης. Για παράδειγμα, η ύπαρξη χαρακτήρων που διαιρούνται σε δύο κομμάτια ή η συνένωση γειτονικών χαρακτήρων μπορεί να οδηγήσει σε σοβαρά λάθη αναγνώρισης.

Οι στόχοι που προσπαθήσαμε να επιτύχουμε είναι:

- Ο εντοπισμός και η αφαίρεση του περιθωρίου καθώς και τμημάτων γειτονικής σελίδας.
- Ο κατάλληλος διαχωρισμός των εγγράφων που περιέχουν δύο σελίδες ώστε να είναι δυνατή η εφαρμογή των επόμενων σταδίων επεξεργασίας.
- Η διόρθωση της τοπικής καμπυλότητας.
- Η δημιουργία κατάλληλου περιβάλλοντος αξιολόγησης των τεχνικών διόρθωσης της τοπικής καμπυλότητας.
- Η κατάτμηση των γραμμών κειμένου και των χαρακτήρων.
- Η γρήγορη και εύκολη δημιουργία ground-truth για την αξιολόγηση μεθόδων κατάτμησης.

1.3 Συνεισφορά της Διδακτορικής Διατριβής

Για την επίτευξη των παραπάνω στόχων αναπτύχθηκαν και προτείνονται καινοτόμες μέθοδοι η αποτελεσματικότητα των οποίων καταδεικνύεται από τα πειραματικά αποτελέσματα πάνω σε γνωστές βάσεις, αλλά και πάνω σε βάσεις που δημιουργήθηκαν στα πλαίσια αυτής της διατριβής. Συγκεκριμένα προτείνονται:

- Δύο μέθοδοι για τον εντοπισμό και την αφαίρεση του περιθωρίου από ψηφιοποιημένα έγγραφα. Η πρώτη μέθοδος εντοπίζει και αφαιρεί το περιθώριο καθώς και τμήματα κειμένου γειτονικών σελίδων. Η μέθοδος βασίζεται στις οριζόντιες και κατακόρυφες προβολές, στην ανάλυση των συνδεδεμένων συστατικών της εικόνας και στη συνάρτηση διασυσχέτισης του σήματος (signal cross-correlation). Η δεύτερη μέθοδος εφαρμόζεται σε έγγραφα που περιέχουν δύο σελίδες, επομένως ο στόχος της, εκτός από την αφαίρεση του περιθωρίου, είναι ο διαχωρισμός του εγγράφου σε δύο ξεχωριστές σελίδες. Η μέθοδος αυτή στηρίζεται σε μία παραλλαγή των κλασικών προβολών η οποία δίνει έμφαση στα μεγάλα συνεχόμενα τμήματα του υπόβαθρου της εικόνας.
- Μία νέα μέθοδος για τη διόρθωση της τοπικής καμπυλότητας η οποία βασίζεται μόνο στη διδιάστατη επεξεργασία του εγγράφου. Η προτεινόμενη μέθοδος εφαρμόζεται σε δύο φάσεις. Η πρώτη φάση έχει ως στόχο να διορθώσει τις έντονες παραμορφώσεις του εγγράφου και να παράγει ένα πρώτο αποτέλεσμα διόρθωσης. Στο στάδιο αυτό εφαρμόζεται ένας χαμηλής υπολογιστικής πολυπλοκότητας μετασχηματισμός, ο οποίος απεικονίζει την προβολή της κυρτής επιφάνειας του εγγράφου σε μία διδιάστατη ορθογώνια περιοχή. Η δεύτερη φάση βασίζεται στον εντοπισμό της γραμμής αναφοράς των λέξεων και διορθώνει τις τοπικές παραμορφώσεις του εγγράφου σε επίπεδο λέξης παράγοντας το τελικό διορθωμένο αποτέλεσμα απαλλαγμένο απ' όλες τις παραμορφώσεις.
- Μία πρωτότυπη ημιαυτόματη μέθοδος για την άμεση αξιολόγηση των τεχνικών διόρθωσης της τοπικής καμπυλότητας. Μέχρι σήμερα η αξιολόγηση των τεχνικών αυτών πραγματοποιείται είτε έμμεσα, με τη χρήση συστημάτων οπτικής αναγνώρισης χαρακτήρων (OCR), είτε άμεσα με τη βοήθεια οπτικών ελέγχων από χρήστες. Η προτεινόμενη μετρική αξιολόγησης λαμβάνει υπόψη την απόκλιση των γραμμών κειμένου στη διορθωμένη εικόνα από μία ευθεία, οριζόντια γραμμή αναφοράς η οποία θεωρείται το ιδανικό αποτέλεσμα. Επίσης, προτείνεται μία διαδικασία αντιστοίχισης με στόχο τη μείωση της προσπάθειας του χρήστη να προσδιορίσει τις γραμμές κειμένου στη διορθωμένη εικόνα. Η διαδικασία αυτή αντιστοιχίζει τα σημεία που έχει μαρκάρει ο χρήστης στην αρχική εικόνα με τα αντίστοιχα σημεία στη διορθωμένη εικόνα, δίνοντας έτσι τη δυνατότητα να αξιολογήσουμε

απεριόριστες μεθόδους χωρίς να χρειάζεται ο χρήστης να επαναλαμβάνει τη διαδικασία του μαρκαρίσματος.

- Μία πρωτότυπη μεθοδολογία συνδυασμού αποτελεσμάτων κατάτμησης των εγγράφων σε γραμμές κειμένου. Η μεθοδολογία προσπαθεί να συνδυάσει συμπληρωματικές τεχνικές κατάτμησης υπό την έννοια ότι διαφορετικές προσεγγίσεις μπορούν να αντιμετωπίσουν αποτελεσματικότερα διαφορετικές δυσκολίες που συναντάμε στα ιστορικά έγγραφα. Συνεπώς, ο στόχος της μεθοδολογίας είναι να εκμεταλλευτεί διαφορετικά αποτελέσματα κατάτμησης και συγκεκριμένα χαρακτηριστικά του αρχικού εγγράφου προκειμένου να παράγει ένα βελτιωμένο αποτέλεσμα κατάτμησης γραμμών κειμένου.
- Μία νέα μέθοδος κατάτμησης χαρακτήρων. Η μέθοδος αυτή βασίζεται στη χρήση διαδρομών κατάτμησης οι οποίες συνδέουν τα χαρακτηριστικά σημεία του σκελετού της λέξης με τα χαρακτηριστικά σημεία του σκελετού του υπόβαθρου. Η επιλογή των διαδρομών που παράγουν το καλύτερο αποτέλεσμα κατάτμησης πραγματοποιείται με τη βοήθεια διάφορων κριτηρίων όπως η απόκλιση του πλάτους του παραγόμενου χαρακτήρα από το αναμενόμενο πλάτος κ.α.
- Μία νέα μέθοδος για τη γρήγορη και εύκολη δημιουργία ground-truth για την αξιολόγηση μεθόδων κατάτμησης. Η μέθοδος χρησιμοποιεί πληροφορία από το αρχείο κειμένου, όπως ο αριθμός των γραμμών κειμένου και ο αριθμός των λέξεων ανά γραμμή κειμένου, με σκοπό να οδηγήσει συγκεκριμένες τεχνικές κατάτμησης σε καλύτερα αποτελέσματα και να μειωθεί σημαντικά ο χρόνος που χρειάζεται ο χρήστης για να δημιουργήσει το ground-truth. Τέλος, προκύπτει η αντιστοίχιση των γραμμών κειμένου, των λέξεων και των χαρακτήρων της εικόνας με το αρχείο κειμένου.

Η δομή της υπόλοιπης διατριβής έχει ως εξής: στο Κεφάλαιο 2 δίνεται μία αναλυτική περιγραφή των σταδίων επεξεργασίας με σκοπό τη βελτίωση της ποιότητας και την κατάτμηση των ιστορικών εγγράφων παραθέτοντας πληθώρα μεθόδων της τρέχουσας τεχνολογικής στάθμης. Στο Κεφάλαιο 3 παρουσιάζονται οι δύο μέθοδοι για τον εντοπισμό και την αφαίρεση του περιθωρίου καθώς και τον διαχωρισμό των εγγράφων σε δύο ξεχωριστές σελίδες. Στο Κεφάλαιο 4 περιγράφεται η μέθοδος διόρθωσης της τοπικής καμπυλότητας καθώς και η νέα ημιαυτόματη μέθοδος αξιολόγησης. Στο Κεφάλαιο 5 αναλύεται διεξοδικά η μεθοδολογία συνδυασμού αποτελεσμάτων κατάτμησης των εγγράφων σε

γραμμές κειμένου και στο Κεφάλαιο 6 περιγράφεται η τεχνική κατάτμησης χαρακτήρων. Στο Κεφάλαιο 7 προτείνεται μία τεχνική για τη γρήγορη και εύκολη δημιουργία ground-truth αξιολόγησης μεθόδων κατάτμησης. Τέλος, συζήτηση για τα συμπεράσματα γίνεται στο Κεφάλαιο 8.

2. ΤΡΕΧΟΥΣΑ ΤΕΧΝΟΛΟΓΙΚΗ ΣΤΑΘΜΗ ΣΤΗΝ ΕΠΕΞΕΡΓΑΣΙΑ ΙΣΤΟΡΙΚΩΝ ΕΓΓΡΑΦΩΝ

Σε αυτό το κεφάλαιο θα αναφερθούμε στα στάδια επεξεργασίας των ιστορικών εγγράφων που αφορούν τη βελτίωσή της ποιότητας και την κατάτμηση των εγγράφων παραθέτοντας τις σημαντικότερες υπάρχουσες τεχνικές.

2.1 Βελτίωση Ποιότητας

Τα δεδομένα, πριν υποστούν οποιαδήποτε επεξεργασία, ανάλογα με τον τρόπο που ψηφιοποιήθηκαν υποβάλλονται σε διάφορα στάδια επεξεργασίας για τη βελτίωσή της ποιότητας τους. Η επεξεργασία αυτή στοχεύει στο να παράγει εικόνες που θα είναι πιο εύκολο να χρησιμοποιηθούν στα επόμενα στάδια της κατάτμησης και της αναγνώρισης.

Οι κύριοι στόχοι αυτής της επεξεργασίας είναι οι εξής:

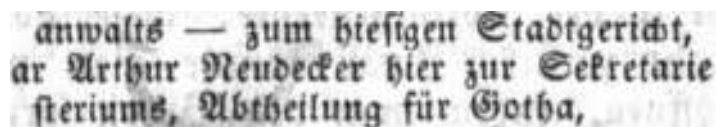
1. Δυαδική Μετατροπή (Binarization)
2. Αφαίρεση Θορύβου (Noise Removal)
3. Εντοπισμός και Αφαίρεση Περιθωρίου (Border Removal)
4. Διόρθωση Στροφής (Skew Correction)
5. Διόρθωση Τοπικής Καμπυλότητας (Dewarping)

2.1.1 Δυαδική Μετατροπή

Η δυαδική μετατροπή των εικόνων είναι το πρώτο βήμα των περισσότερων συστημάτων ανάλυσης και επεξεργασίας εγγράφων και αναφέρεται στη μετατροπή των εικόνων διαβάθμισης γκρι σε δυαδικές. Η δυαδική μετατροπή αποτελεί βασικό στοιχείο των συστημάτων επεξεργασίας εγγράφων διότι η βέλτιστη δυαδική μετατροπή αποτελεί τη βάση για τη συμπίεση περιεχομένου, τη βελτίωσή της εμφάνισης λόγω της αφαίρεσης θορύβου και τη διευκόλυνση και επιτάχυνση των αλγορίθμων κατάτμησης και αναγνώρισης. Οι τεχνικές δυαδικής μετατροπής χωρίζονται σε δυο βασικές κατηγορίες: **καθολικής κατωφλίωσης** και **τοπικής κατωφλίωσης**.

Στις μεθόδους καθολικής κατωφλίωσης επιλέγεται μία τιμή ως κατώφλι και στη συνέχεια όλες οι στάθμες του γκρι κάτω από το κατώφλι μετατρέπονται σε κείμενο και εκείνες που είναι πάνω από το κατώφλι μετατρέπονται σε υπόβαθρο.

Η σωστή επιλογή του κατωφλίου είναι σημαντική για την ποιότητα της τελικής εικόνας. Μία από τις κυριότερες τεχνικές αυτής της κατηγορίας είναι η τεχνική Otsu [1] η οποία έχει χαρακτηριστεί σαν μία από τις καλύτερες τεχνικές εύρεσης καθολικού κατωφλιού. Η μέθοδος αυτή βασίζεται στην επεξεργασία του ιστογράμματος της εικόνας και στον προσδιορισμό του κατωφλιού βάσει του κριτηρίου της μεγιστοποίησης της διαχωρισιμότητας μεταξύ των περιοχών κειμένου και υπόβαθρου. Στο Σχήμα 2.1 παρουσιάζεται το αποτέλεσμα εφαρμογής της μεθόδου Otsu σε ένα ιστορικό έγγραφο. Στο [2] η επιλογή του κατωφλιού γίνεται με χρήση σημείων ακμών. Τα σημεία ακμών είναι σημεία ανάμεσα στο αντικείμενο και το υπόβαθρο της εικόνας ή μεταξύ δύο αντικειμένων. Το ιστόγραμμα της εικόνας που προκύπτει μόνο από τα σημεία αυτά μας δίνει καλύτερη διαχωρισιμότητα κειμένου-υπόβαθρου σε σχέση με το ιστόγραμμα της συνολικής εικόνας. Στην εργασία [3], οι Ridler και Calvard προτείνουν μία επαναληπτική μέθοδο η οποία ξεκινά από μία αρχική εκτίμηση για τη τιμή του κατωφλιού και στη συνέχεια προσαρμόζεται λαμβάνοντας υπόψη τις περιοχές κειμένου και υπόβαθρου που προκύπτουν σε κάθε επανάληψη.



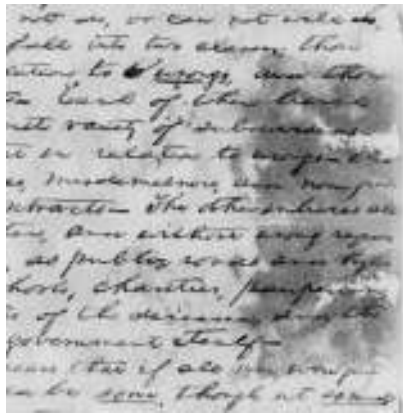
(α)

**anwalts — zum hiesigen Stadtgericht,
ar Arthur Neudecker hier zur Sekretarie
steriums, Abtheilung für Gotha,**

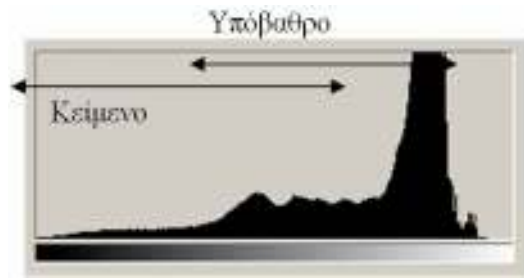
(β)

Σχήμα 2.1: Δυαδική μετατροπή με τη χρήση της μεθόδου Otsu [1]: (α) Εικόνα ιστορικού εγγράφου διαβάθμισης γκρι, (β) δυαδική εικόνα εγγράφου.

Οι μέθοδοι καθολικής κατωφλίωσης μπορούν να εφαρμοστούν μόνο στις περιπτώσεις όπου υπάρχει σαφής διάκριση μεταξύ των περιοχών κειμένου και υπόβαθρου. Αυτό όμως δε συμβαίνει στις περισσότερες περιπτώσεις των εγγράφων, ιδίως στα ιστορικά έγγραφα που έχουμε χαμηλή ποιότητα χαρακτήρων, σκιές, μη ομοιόμορφη φωτεινότητα υπόβαθρου, έντονο θόρυβο κ.λπ. (Σχήμα 2.2(α)). Στις περιπτώσεις αυτές παρατηρούμε ότι στο ιστόγραμμά οι τιμές του γκρι που αντιστοιχούν στο κείμενο και στο υπόβαθρο επικαλύπτονται (Σχήμα 2.2(β)).



(α)



(β)

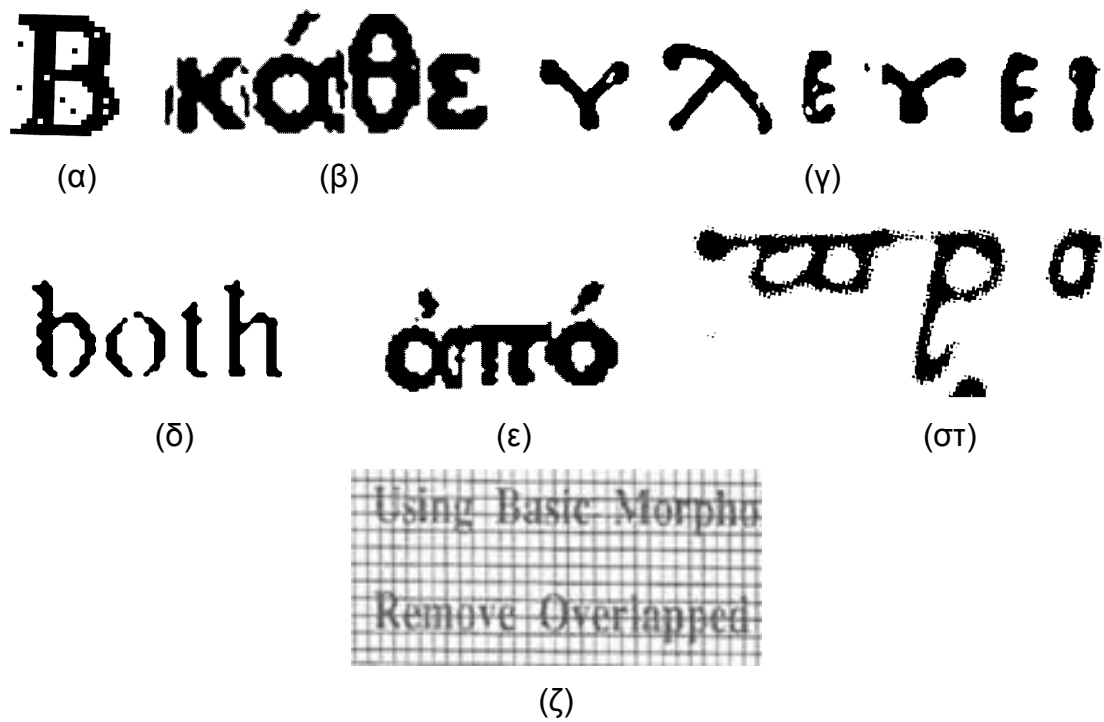
Σχήμα 2.2: (α) Ιστορικό έγγραφο και (β) το ιστόγραμμα του στο οποίο οι τιμές του γκρι που αντιστοιχούν στο κείμενο και στο υπόβαθρο επικαλύπτονται.

Για τους παραπάνω λόγους αναπτύχθηκαν οι μέθοδοι τοπικής κατωφλίωσης στις οποίες χρησιμοποιούνται πολλές τιμές κατωφλιών ανάλογα με την τοπική πληροφορία της εικόνας. Έτσι επιτυγχάνεται η βέλτιστη δυαδική μετατροπή ακόμα και στις περιπτώσεις που δεν υπάρχει σαφής διάκριση των περιοχών του κειμένου από το υπόβαθρο. Η τεχνική του Niblack [4] είναι μία τεχνική τοπικής κατωφλίωσης η οποία υπολογίζει ένα κατώφλι για κάθε εικονοστοιχείο (pixel) κάνοντας χρήση της μέσης τιμής και της τυπικής απόκλισης της φωτεινότητας σε ένα παράθυρο το οποίο μετακινείται σε όλη την εικόνα. Η μέθοδος αυτή μπορεί να ξεχωρίσει το κείμενο από το υπόβαθρο ακόμα και όταν οι φωτεινότητές τους δεν διαφέρουν αρκετά αλλά αφήνει αρκετή ποσότητα θορύβου στο υπόβαθρο. Ο αλγόριθμος του Sauvola [5] προσπαθεί να δώσει λύση σε αυτό το πρόβλημα εισάγοντας την υπόθεση ότι τα pixels κειμένου έχουν στάθμες του γκρι κοντά στο 0 ενώ τα pixels του υπόβαθρου έχουν στάθμες του γκρι κοντά στο 255. Μία τεχνική τοπικής κατωφλίωσης που εφαρμόζεται σε έγγραφα υποβαθμισμένης ποιότητας παρουσιάζεται στο [6] και χρησιμοποιεί πληροφορία από το υπόβαθρο. Επιπλέον, έχουν προταθεί τεχνικές δυαδικής μετατροπής οι οποίες είναι αποκλειστικά προσανατολισμένες σε ιστορικά έγγραφα. Στην εργασία [7], οι Chen και Leedham χρησιμοποιούν διανύσματα χαρακτηριστικών και βάρη με στόχο να επιλέξουν το καταλληλότερο τοπικό κατώφλι, ενώ στο [8] εφαρμόζουν πρώτα ένα ολικό κατώφλι και στη συνέχεια εντοπίζουν τις περιοχές που πιθανόν να περιέχουν ακόμα θόρυβο. Στα [9] και [10] υπάρχει εκτενής έρευνα και σύγκριση τεχνικών δυαδικής μετατροπής οι οποίες είναι προσανατολισμένες σε ιστορικά έγγραφα. Ιδιαίτερο ενδιαφέρον παρουσιάζουν και οι τεχνικές στις οποίες λαμβάνεται υπόψη το πλάτος του χαρακτήρα και έχουν αναπτυχθεί ειδικά για έγγραφα κειμένου όπως η τεχνική του Kamel [11] και του Yang [12].

2.1.2 Αφαίρεση Θορύβου

Ο θόρυβος είναι αποτέλεσμα της κακής ψηφιοποίησης, της χαμηλής ποιότητας του αρχικού εγγράφου, της παλαιότητας του εγγράφου κ.α. Μερικά από τα προβλήματα που προκαλεί ο θόρυβος στην ποιότητα των δεδομένων και πρέπει να αντιμετωπιστούν είναι τα ακόλουθα:

- **Θόρυβος μορφής αλατοπίπερου (salt and pepper noise):** Είναι η παρουσία μεμονωμένων pixels κειμένου σε περιοχές υπόβαθρου ή το αντίστροφο. Ο θόρυβος αυτός μπορεί να προέλθει λόγω ψηφιοποίησης ή δυαδικής μετατροπής (Σχήμα 2.3(α)).
- **Θόρυβος αποτελούμενος από περιοχές μικρού μεγέθους:** Περιοχές από pixels κειμένου με μικρό εμβαδόν οι οποίες έχουν προκύψει λόγω θορύβου ή κακής ψηφιοποίησης (Σχήμα 2.3(β)).
- **Οπές στο σώμα των χαρακτήρων:** Το σώμα των χαρακτήρων αντί να είναι συμπαγές παρουσιάζει οπές. Αυτό μπορεί να δημιουργήσει ιδιαίτερο πρόβλημα στην αναπαράσταση και αναγνώριση των χαρακτήρων (Σχήμα 2.3(γ)).
- **Ασυνέχειες στο σώμα των χαρακτήρων:** Δεν υπάρχει συνεκτικότητα στους χαρακτήρες καθώς εμφανίζονται σπασίματα ή κενά. Ιδιαίτερα σημαντικό πρόβλημα για το στάδιο της κατάτμησης (Σχήμα 2.3(δ)).
- **Ενωμένοι χαρακτήρες:** Γειτονικοί χαρακτήρες ενώνονται μεταξύ τους λόγω κακής ψηφιοποίησης ή ύπαρξης θορύβου ανάμεσα στους χαρακτήρες (Σχήμα 2.3(ε)).
- **Μη ομαλή περίμετρος των χαρακτήρων:** Η μη ομαλή περίμετρος των χαρακτήρων είναι πρόβλημα που εισάγεται κατά την ψηφιοποίηση και μπορεί να έχει επιπτώσεις στο στάδιο της αναγνώρισης χαρακτήρων (Σχήμα 2.3(στ)).
- **Επικάλυψη κειμένου από την υφή υπόβαθρου:** Υφή του υπόβαθρου (γραμμές ή περιοδικά εμφανιζόμενα σύμβολα) που επικαλύπτει τις περιοχές κειμένου (Σχήμα 2.3(ζ)).



Σχήμα 2.3: Παραδείγματα θορύβου: (α) μορφής αλατοπίπερου, (β) περιοχές μικρού μεγέθους, (γ) σπές, (δ) ασυνέχειες στο σώμα των χαρακτήρων, (ε) ενωμένοι χαρακτήρες, (στ) μη ομαλή περίμετρος των χαρακτήρων, (ζ) επικάλυψη κειμένου από την υφή υπόβαθρου.

Όλα τα παραπάνω προβλήματα θορύβου είναι απαραίτητα να εξαλειφθούν πριν την περαιτέρω επεξεργασία των εγγράφων. Οι κυριότεροι μέθοδοι που χρησιμοποιούνται για την αντιμετώπιση των παραπάνω προβλημάτων μπορούν να ταξινομηθούν σε γραμμικά ή μη γραμμικά φίλτρα.

Η χρήση φίλτρων στοχεύει στο να αφαιρεθεί ο θόρυβος, να βελτιωθεί η ποιότητα των περιοχών κειμένου καθώς και η συνεκτικότητα των χαρακτήρων αφαιρώντας απομονωμένα pixels και γεμίζοντας τυχόν σπασίματα ή κενά στο σώμα των χαρακτήρων. Γι' αυτό τον σκοπό έχει σχεδιαστεί ένα πλήθος φίλτρων τόσο στο πεδίο του χώρου όσο και στο πεδίο των συχνοτήτων. Η βασική ιδέα αποτελείται από μία προκαθορισμένη μάσκα (παράθυρο) η οποία διατρέχει την εικόνα και η τιμή κάθε pixel καθορίζεται συναρτήσει των τιμών των γειτονικών του pixels [13], [14]. Έχουν αναπτυχθεί φίλτρα για την εξομάλυνση της εικόνας (smoothing), την αύξηση της οξύτητας (sharpening), την κατωφλίωση (thresholding), την αφαίρεση αχνών σημείων κειμένου ή υπόβαθρου και για τη βελτίωσή της αντίθεσης της φωτεινότητας (contrast) [15].

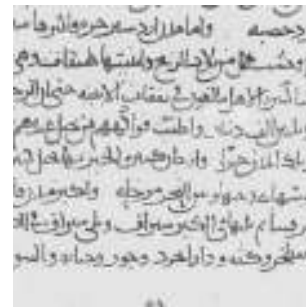
Η μαθηματική μορφολογία αναφέρεται στη μελέτη της γεωμετρίας, της τοπολογίας και της μορφής των αντικειμένων [16]. Αφορά συγκεκριμένες πράξεις κατά τις οποίες μία εικόνα αλληλεπιδρά με ένα δομικό στοιχείο και μετατρέπεται σε μία απλοποιημένη και λειτουργική μορφή διατηρώντας τα βασικά χαρακτηριστικά της.

Στοχεύει στη βελτίωση της δομής των αντικειμένων (φιλτράρισμα θορύβου, βελτίωσή της ποιότητας και απλοποίηση των αντικειμένων, σκελετός, λέπτυνση, πάχυνση της εικόνας, λείανση των περιγραμμάτων) και στην ποσοτική περιγραφή των αντικειμένων (χαρακτηριστικά εμβαδού, περιμέτρου, προβολών). Οι περισσότερες μορφολογικές πράξεις μπορούν να οριστούν χρησιμοποιώντας τις δύο βασικές μορφολογικές πράξεις οι οποίες είναι η διάβρωση (erosion) και η διαστολή (dilation). Επιπλέον, η μαθηματική μορφολογία μπορεί να χρησιμοποιηθεί για την απομάκρυνση συμβόλων του υπόβαθρου τα οποία επικαλύπτουν τους χαρακτήρες [17]. Τα σύμβολα αυτά μπορεί να είναι περιοδικά επαναλαμβανόμενα σχήματα ή ευθύγραμμα τμήματα (Σχήμα 2.3(ζ)).

Επιπλέον, έχουν προταθεί τεχνικές αφαίρεσης θορύβου οι οποίες αφορούν εικόνες διαβάθμισης γκρι και είναι προσανατολισμένες σε ιστορικά έγγραφα και έγγραφα χαμηλής ποιότητας. Στις εργασίες [18] και [19] παρουσιάζονται δύο μέθοδοι οι οποίες προσπαθούν να μειώσουν τον θόρυβο από το υπόβαθρο της εικόνας και να αναδείξουν καλύτερα το περιεχόμενό της. Τα Σχήματα 2.4 και 2.5 παρουσιάζουν παραδείγματα εφαρμογής των παραπάνω μεθόδων σε ιστορικά έγγραφα.

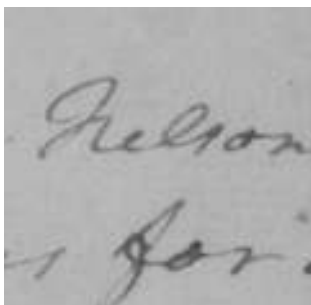


(α)

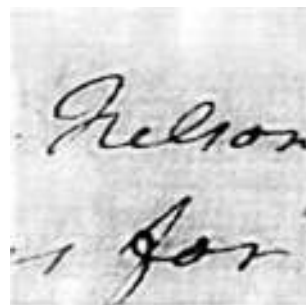


(β)

Σχήμα 2.4: Βελτίωση ποιότητας ιστορικού εγγράφου με τη μέθοδο [18]: (α) αρχική εικόνα, (β) εικόνα μετά την επεξεργασία.



(α)

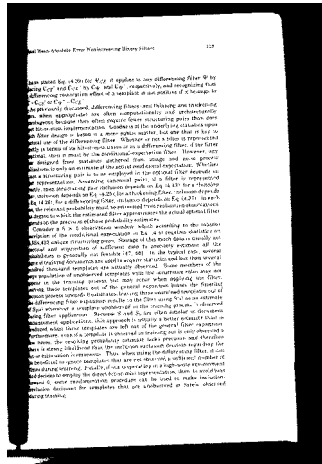


(β)

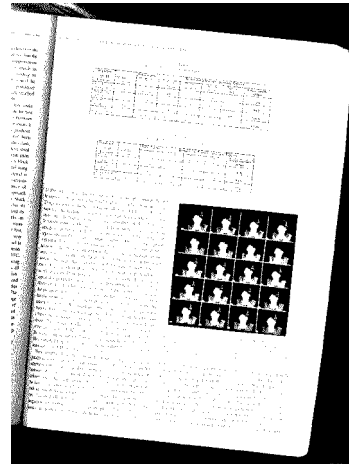
Σχήμα 2.5: Βελτίωση ποιότητας ιστορικού εγγράφου με τη μέθοδο [19]: (α) αρχική εικόνα, (β) εικόνα μετά την επεξεργασία.

2.1.3 Εντοπισμός και Αφαίρεση Περιθωρίου

Λόγω κακής ψηφιοποίησης του εγγράφου (π.χ. υπερφωτισμός στα όρια του εγγράφου) ή λόγω κακής δυαδικής μετατροπής είναι αρκετά συνηθισμένο να εμφανίζεται ένα περιθώριο στα όρια της εικόνας του εγγράφου (Σχήμα 2.6(α)), καθώς επίσης και τμήματα γειτονικών σελίδων (Σχήμα 2.6(β)). Αυτό το φαινόμενο αλλοιώνει την όψη του εγγράφου, δυσκολεύει τη μετέπειτα επεξεργασία του, ενώ συγχρόνως αυξάνει το μέγεθός του.



(α)

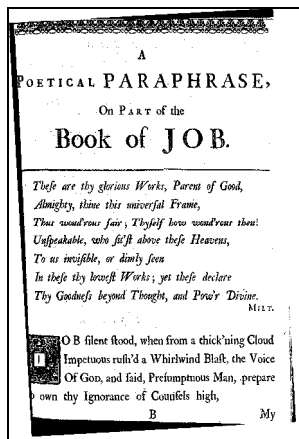


(β)

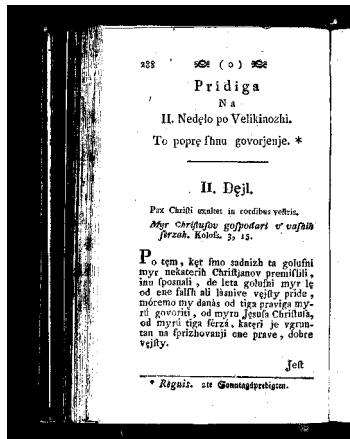
Σχήμα 2.6: Εμφάνιση (α) περιθωρίου και (β) τμήματος γειτονικής σελίδας στα όρια του εγγράφου.

Μερικά βασικά χαρακτηριστικά του περιθωρίου που δυσκολεύουν τον εντοπισμό και την αφαίρεσή του είναι τα ακόλουθα:

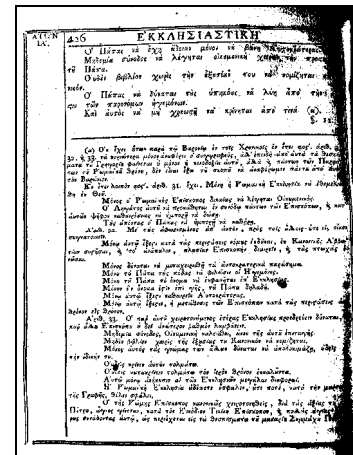
- Δε βρίσκεται πάντα στα όρια της εικόνας (Σχήμα 2.7(α)).
- Δεν είναι πάντα συμπαγές (Σχήμα 2.7(β)).
- Μπορεί να εισχωρεί στο έγγραφο με αποτέλεσμα να επικαλύπτει περιοχές κειμένου (Σχήμα 2.7(γ)).



(α)



(β)

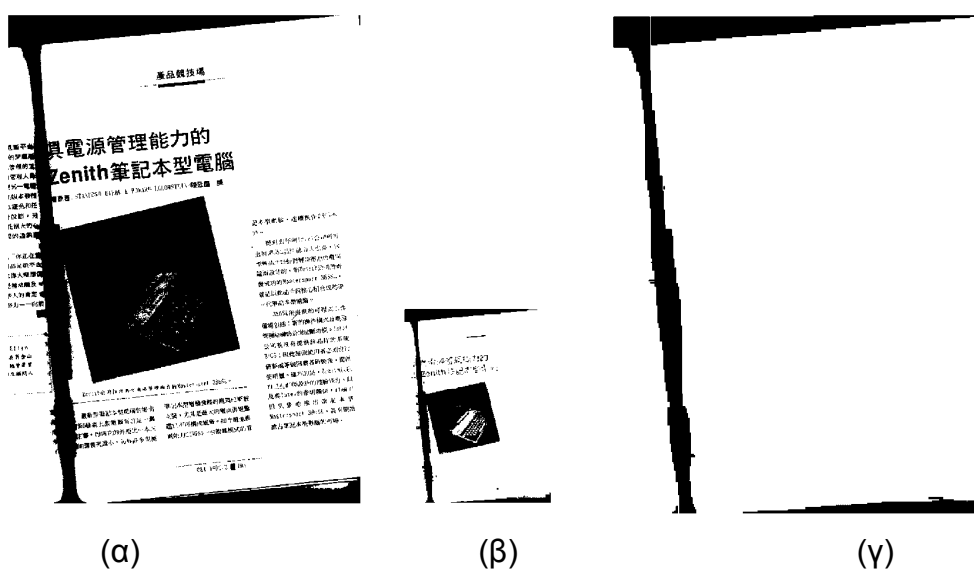


(γ)

Σχήμα 2.7: Παραδείγματα εμφάνισης περιθωρίου σε ιστορικά έγγραφα.

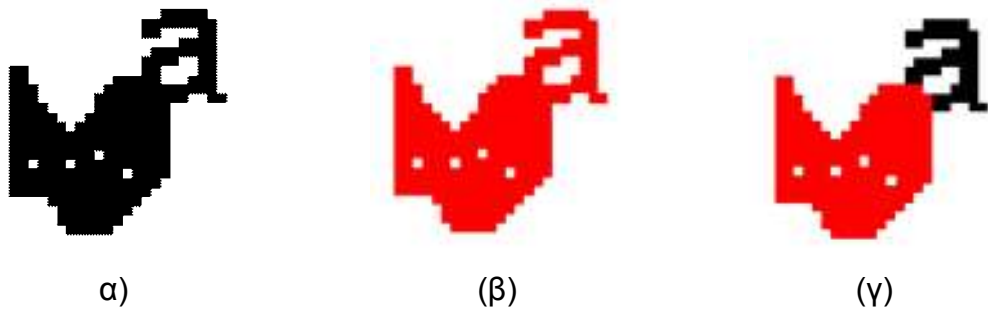
Στη βιβλιογραφία έχουν προταθεί τεχνικές που έχουν ως στόχο τον εντοπισμό της περιοχής κειμένου και την αφαίρεση τόσο του περιθωρίου όσο και των τμημάτων κειμένου γειτονικών σελίδων. Οι Le et al. [20] προτείνουν μία αυτοματοποιημένη μέθοδο εντοπισμού των ορίων του εγγράφου η οποία βασίζεται στην ταξινόμηση των γραμμών και των στηλών σε τρεις κατηγορίες (λευκές (blank), κειμένου (textual) και μη κειμένου (non-textual)), την κατάτμηση του εγγράφου και την ανάλυση των προβολών. Η μέθοδος στηρίζεται σε αρκετές παραδοχές και παραμέτρους με αποτέλεσμα να είναι εφαρμόσιμη σε περιορισμένες περιπτώσεις.

Στην εργασία [21], οι Fan et al. αρχικά μειώνουν την ανάλυση του εγγράφου ώστε να διατηρηθούν οι περιοχές του περιθωρίου και να απομακρυνθούν οι περιοχές κειμένου. Ο καθορισμός του κατάλληλου επιπέδου μείωσης της ανάλυσης του εγγράφου είναι σημαντικός για το τελικό αποτέλεσμα. Μία μεγάλη μείωση της ανάλυσης θα διατηρήσει τις μεγάλες περιοχές του περιθωρίου αλλά θα εξαφανίσει τις μικρότερες, ενώ μία μικρή μείωση θα διατηρήσει και περιοχές κειμένου. Στη συνέχεια, οι περιοχές που έχουν παραμείνει ταξινομούνται σε περιοχές περιθωρίου και μη χρησιμοποιώντας διάφορα στατιστικά στοιχεία όπως το μήκος, τη θέση και τις γειτονικές περιοχές. Το επόμενο στάδιο, μετά την ταξινόμηση των περιοχών, είναι η απομάκρυνση των περιοχών περιθωρίου. Η διαδικασία απομάκρυνσης του περιθωρίου υλοποιείται στο αρχικό έγγραφο και όχι στο έγγραφο μειωμένης ανάλυσης και αφορά μόνο το περιθώριο και όχι τμήματα από γειτονικές σελίδες. (Σχήμα 2.8).



Σχήμα 2.8: Παράδειγμα εφαρμογής της μεθόδου [21] για την αφαίρεση του περιθωρίου: (α) αρχικό έγγραφο, (β) μετά τη μείωση της ανάλυσης και (γ) τελικός προσδιορισμός του περιθωρίου.

Οι Anila και Lins [22],[23] προτείνουν τρεις αλγόριθμους για την αφαίρεση του περιθωρίου αλλά όχι τμημάτων γειτονικών σελίδων. Στο [22] προτείνουν τους αλγόριθμους εισβολής (invading) και μη εισβολής (non-invading). Ο αλγόριθμος εισβολής προϋποθέτει ότι το περιθώριο δεν επικαλύπτει περιοχές κειμένου ενώ ο αλγόριθμος μη εισβολής εξασφαλίζει ότι αν το περιθώριο επικαλύπτει περιοχές κειμένου τότε αυτές παραμένουν στην εικόνα και δεν αφαιρούνται (Σχήμα 2.9). Και οι δύο αλγόριθμοι χρησιμοποιούν την τεχνική της πλημμύρας (flood-fill algorithm) για την αφαίρεση του περιθωρίου και εφαρμόζουν στατιστική ανάλυση σε χαρακτηριστικά του εγγράφου για να υπολογίσουν τις παραμέτρους που θα σταματήσουν την τεχνική της πλημμύρας. Στην εργασία [23] προτείνουν έναν παρόμοιο αλγόριθμο ο οποίος βασίζεται στην τεχνική της πλημμύρας, στην ετικετοποίηση των συστατικών του εγγράφου (component labelling) και στους γράφους περιοχών γεινίασης (region adjacency graphs) με στόχο την ταξινόμηση τους σε περιοχές περιθωρίου και μη με τη χρήση φίλτρων.



Σχήμα 2.9: (α) Αρχικό έγγραφο όπου παρατηρείτε επικάλυψη του κειμένου από το θόρυβο περιθωρίου. Αποτέλεσμα αλγορίθμου (β) εισβολής και (γ) μη εισβολής όπου με ανοικτό χρώμα σημειώνονται τα pixels που αφαιρούνται ως θόρυβος περιθωρίου.

Η μέθοδος των Shafait et al. [24], σε αντίθεση με τις προαναφερθείσες μεθόδους, δεν προσπαθεί να εντοπίσει και να αφαιρέσει τις περιοχές που ανήκουν στο περιθώριο και τη γειτονική σελίδα, αλλά εντοπίζει το πλαίσιο της περιοχής κειμένου (page frame) και αφαιρεί ότι δεν ανήκει σε αυτό. Η μέθοδος λαμβάνει υπόψη τη δομή των τυπωμένων εγγράφων και προσπαθεί να χτίσει ένα γεωμετρικό μοντέλο πάνω στο πλαίσιο της περιοχής κειμένου. Έπειτα, με τη χρήση μεθόδων γεωμετρικής αντιστοίχισης και συναρτήσεων κόστους καταλήγει στο ιδανικό πλαίσιο της περιοχής κειμένου. Στο Σχήμα 2.10 φαίνεται το αποτέλεσμα της μεθόδου [24] σε έγγραφο όπου το περιθώριο επικαλύπτει μέρος της περιοχής κειμένου.



Σχήμα 2.10: Παράδειγμα εφαρμογής της μεθόδου [24] για την αφαίρεση του περιθωρίου: (α) Αρχικό έγγραφο όπου παρατηρείτε επικάλυψη του κειμένου από το περιθωρίου, (β) εντοπισμός του πλαισίου της περιοχής κειμένου, (γ) τελικό αποτέλεσμα μετά την αφαίρεση του περιθωρίου.

Τέλος, στην εργασία [25], οι Yacoub et al. περιγράφουν ένα σύστημα για την ψηφιοποίηση εγγράφων με σύνθετη δομή. Υποθέτουν ότι ένα έγγραφο περιέχει πάντα δύο σελίδες (Σχήμα 2.11), οπότε ο στόχος τους είναι, εκτός από την αφαίρεση του περιθωρίου, και ο διαχωρισμός του εγγράφου σε δύο ξεχωριστές σελίδες. Αρχικά, αφαιρούν το περιθώριο χρησιμοποιώντας οριζόντιες και κατακόρυφες προβολές για τον εντοπισμό ευθύγραμμων ακμών και στη συνέχεια, διαχωρίζουν το έγγραφο σε δύο σελίδες εντοπίζοντας απλά το κέντρο του.



Σχήμα 2.11: Παράδειγμα εγγράφου που περιέχει δύο σελίδες.

Στον Πίνακα 2.1 παρουσιάζονται συνοπτικά οι μέθοδοι που έχουν προταθεί για τον εντοπισμό της περιοχής κειμένου και την αφαίρεση τόσο του περιθωρίου όσο και των τμημάτων κειμένου γειτονικών σελίδων.

Πίνακας 2.1: Μέθοδοι Εντοπισμού και Αφαίρεσης Περιθωρίου

	Μέθοδος	Αφαίρεση περιθωρίου	Αφαίρεση γειτονικής σελίδας	Διαχωρισμός διπλής σελίδας
Le et al. 1996 [20]	Ανάλυση προβολών	✓	✓	-
Fan et al. 2002 [21]	Μείωση ανάλυσης εγγράφου	✓	-	-
Avila and Lins 2004 [22]	Τεχνική της πλημμύρας	✓	-	-
Avila and Lins 2004 [23]	Γράφοι περιοχών γειτνίασης	✓	-	-
Shafait et al. 2008 [24]	Γεωμετρικά μοντέλα	✓	✓	-
Yacoub et al 2005 [25]	Ανάλυση προβολών	✓	-	✓

2.1.4 Διόρθωση Κλίσης

Ένα από τα βασικά βήματα για τη βελτίωσή της ποιότητας των εγγράφων είναι ο εντοπισμός και η διόρθωση της κλίσης. Η κλίση του εγγράφου προέρχεται συνήθως από τη μη ευθυγραμμισμένη τοποθέτησή του στον ψηφιακό μετατροπέα. Το πρόβλημα αυτό είναι πολύ συνηθισμένο στις περιπτώσεις των ιστορικών βιβλίων και η διόρθωση του είναι απαραίτητη πριν προχωρήσουμε στα επόμενα στάδια (Σχήμα 2.12).



(α)



(β)

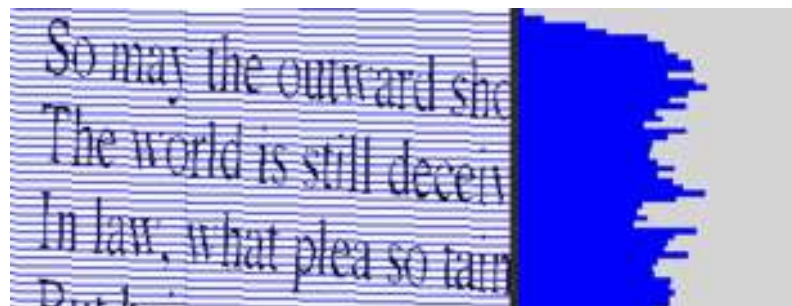
Σχήμα 2.12: (α) Αρχική εικόνα, (β) εικόνα μετά τη διόρθωση της κλίσης.

Πριν προχωρήσουμε στην αναζήτηση και τη διόρθωση της κλίσης των εγγράφων είναι απαραίτητο να εντοπίσουμε τον προσανατολισμό τους. Υπάρχουν δύο δυνατοί προσανατολισμοί, ο οριζόντιος (landscape) και ο κατακόρυφος (portrait). Μία απλή και αποτελεσματική μέθοδος βασίζεται στην οριζόντια και κάθετη εξομάλυνση της εικόνας [26]. Οι τεχνικές εξομάλυνσης της εικόνας στοχεύουν στο γέμισμα των περιοχών του υπόβαθρου ώστε να ενωθούν τμήματα της εικόνας που ανήκουν στην ίδια κατηγορία. Η πιο γνωστή τεχνική εξομάλυνσης είναι ο αλγόριθμος RLSA (Run Length Smoothing Algorithm – Αλγόριθμος εξομάλυνσης βασισμένος στο διατρέχον μήκος) [27].

Οι μέθοδοι για τη διόρθωση της κλίσης βασίζονται κυρίως σε τεχνικές όπως ο πλησιέστερος γείτονας (nearest neighbor – NN) [28], [29], η συνάρτηση συσχέτισης (correlation) [30] και η επεξεργασία των προβολών [31], [32]. Οι τεχνικές που βασίζονται στην επεξεργασία των προβολών αναλύουν τις προβολές του εγγράφου ή τμημάτων αυτού σε διαφορετικές διευθύνσεις. Αν η διεύθυνση προβολής ταυτίζεται με τη διεύθυνση της κλίσης του εγγράφου τότε το ιστόγραμμα της προβολής της εικόνας παρουσιάζει τοπικά μέγιστα τα οποία αντιστοιχούν στις γραμμές κειμένου (Σχήμα 2.13(α)). Σε αντίθετη περίπτωση το ιστόγραμμα της προβολής της εικόνας δεν παρουσιάζει περιοδικότητα (Σχήμα 2.13(β)).



(α)



(β)

Σχήμα 2.13: Το ιστόγραμμα της προβολής της εικόνας όταν η διεύθυνση προβολής (α) ταυτίζεται με την κλίση του εγγράφου ή (β) είναι τυχαία.

Επίσης, μία διαδεδομένη μέθοδος διόρθωσης της κλίσης είναι ο μετασχηματισμός Hough [33],[34]. Δεδομένου ενός συνόλου pixels κειμένου ο μετασχηματισμός Hough μπορεί να εντοπίσει τη γραμμή που διασχίζει τον μεγαλύτερο αριθμό από αυτά τα pixels. Συνεπώς, σε μία δυαδική εικόνα κειμένου, οι γραμμές που εντοπίζει ο μετασχηματισμός Hough περνούν συνήθως από τη μέση των γραμμών κειμένου λόγω της φυσικής διεύθυνσης της γραφής. Συνεπώς, θα έχουν την ίδια κλίση με τη κλίση ολόκληρης της σελίδας.

2.1.5 Διόρθωση Τοπικής Καμπυλότητας

Η ψηφιοποίηση εγγράφων με τη χρήση σαρωτών (επιτραπέζιων ή χειρός) ή ψηφιακών φωτογραφικών μηχανών οδηγούν συχνά σε έγγραφα που πάσχουν από παραμορφώσεις (Σχήμα 2.14) εξαιτίας της μορφής του υλικού προς ψηφιοποίηση (π.χ. βιβλιοδεσία πολυσέλιδων εντύπων), της διάταξης του εγγράφου κατά την ψηφιοποίηση (π.χ. μη ευθυγραμμισμένη τοποθέτηση του εγγράφου στον ψηφιακό μετατροπέα) ή λόγω περιβαλλοντικών συνθηκών (π.χ. παραμόρφωση εγγράφου λόγω υγρασίας). Τέτοιου είδους παραμορφώσεις, όπως οι καμπύλες γραμμές κειμένου, δε μειώνουν μόνο την αναγνωσιμότητα του εγγράφου αλλά επηρεάζουν και την αποτελεσματικότητα των επόμενων σταδίων επεξεργασίας όπως της κατάτμησης και της αναγνώρισης του εγγράφου.



(α)



(β)

Σχήμα 2.14: Ψηφιοποιημένα έγγραφα με τη χρήση (α) επιτραπέζιου σαρωτή και (β) ψηφιακής φωτογραφικής μηχανής.

Οι μέθοδοι που έχουν προταθεί για τη διόρθωση της τοπικής καμπυλότητας ταξινομούνται σε δύο κύριες κατηγορίες: (α) τεχνικές που βασίζονται στην **τριδιάστατη ανακατασκευή του εγγράφου** [35]-[42] και (β) τεχνικές που βασίζονται μόνο στη **διδιάστατη επεξεργασία του εγγράφου** [43]-[55].

- **Τεχνικές που βασίζονται στην τριδιάστατη ανακατασκευή του εγγράφου**

Οι μέθοδοι που ανήκουν σε αυτή την κατηγορία βασίζονται στην τριδιάστατη ανακατασκευή του εγγράφου για να διορθώσουν τις παραμορφώσεις και διαχωρίζονται σε δύο υποκατηγορίες. Οι τεχνικές της πρώτης υποκατηγορίας ανακτούν την τριδιάστατη δομή του εγγράφου βασιζόμενες σε επιπρόσθετη πληροφορία με τη χρήση ειδικού εξοπλισμού όπως σαρωτή λέιζερ (laser scanner) [35], στερεοσκοπική φωτογραφική μηχανή (stereo camera) [36], [37], ή τριδιάστατο σαρωτή δομημένου φωτός (structured-light 3D scanner) [38] (Σχήμα 2.15). Η χρήση ειδικού εξοπλισμού περιορίζει όπως είναι φυσικό την εφαρμογή των συγκεκριμένων τεχνικών.



(α)



(β)



(γ)

Σχήμα 2.15: Ειδικός εξοπλισμός για την ανάκτηση της τριδιάστατης δομής του εγγράφου: (α) σαρωτής λέιζερ, (β) στερεοσκοπική φωτογραφική μηχανή και (γ) τριδιάστατος σαρωτής δομημένου φωτός.

Οι τεχνικές που ανήκουν στη δεύτερη υποκατηγορία ανακτούν την τριδιάστατη δομή του εγγράφου βασιζόμενες μόνο σε πληροφορία που υπάρχει στο ψηφιοποιημένο έγγραφο. Οι Cao et al. [39] για να διορθώσουν τις παραμορφώσεις του εγγράφου προτείνουν την αναπαράστασή του με ένα κυλινδρικό μοντέλο. Οι παράμετροι του μοντέλου βασίζονται στο σκελετό των οριζόντιων γραμμών κειμένου του εγγράφου. Εκτός του περιορισμού του προτεινόμενου μοντέλου, αφού δεν είναι πάντα εφικτή η μοντελοποίηση του εγγράφου με κύλινδρο, η μέθοδος θέτει και άλλους περιορισμούς σχετικά με την τοποθέτηση του εγγράφου κατά την ψηφιοποίηση. Στην εργασία [40], οι Liang et al. υπολογίζουν την τριδιάστατη δομή του εγγράφου απευθείας από το πεδίο ροής του κειμένου. Η μέθοδος θέτει επίσης περιορισμούς σχετικά με την

τοποθέτηση του εγγράφου κατά την ψηφιοποίηση. Τέλος, οι Tan et al. [41] και οι Zhang et al. [42] βασίζονται στη σκίαση που δημιουργείται πάνω στα παραμορφωμένα έγγραφα για να ανακτήσουν την τριδιάστατη δομή τους (Σχήμα 2.16). Η πληροφορία για τον φωτισμό των εγγράφων δεν είναι διαθέσιμη τις περισσότερες φορές και συνεπώς η εφαρμογή τους είναι περιορισμένη.



Σχήμα 2.16: Παραμορφωμένο έγγραφο μαζί με τη σκίαση που δημιουργείται κατά την ψηφιοποίησή του.

- **Τεχνικές που βασίζονται στη διδιάστατη επεξεργασία του εγγράφου**

Στην κατηγορία αυτή ανήκουν οι μέθοδοι που βασίζονται μόνο στην πληροφορία που υπάρχει στα ψηφιοποιημένα έγγραφα και δουλεύουν αποκλειστικά στο διδιάστατο χώρο για τη διόρθωση των παραμορφώσεων. Η μεγάλη πλειοψηφία των μεθόδων αυτής της κατηγορίας στηρίζεται στον εντοπισμό των καμπύλων γραμμών κειμένου στα παραμορφωμένα έγγραφα, μία διαδικασία που είναι αρκετά δύσκολη και μπορεί να οδηγήσει συχνά σε λανθασμένα αποτελέσματα. Αρκετές τεχνικές προσεγγίζουν κάθε γραμμή κειμένου ξεχωριστά και προσπαθούν να την ευθυγραμμίσουν ώστε να διορθώσουν την τοπική καμπυλότητα. Οι Lavialle et al. [43] μοντελοποιούν κάθε γραμμή κειμένου με κυβικές B-splines, οι οποίες έχουν αποδειχθεί πιο ακριβείς από τις καμπύλες Bezier (Σχήμα 2.17). Προτείνουν μία αυτόματη διαδικασία για τον εντοπισμό των γραμμών κειμένων η αρχική κατάσταση της οποίας πρέπει να βρίσκεται κοντά στο επιθυμητό αποτέλεσμα. Επιπλέον, η τεχνική αυτή δεν είναι αποτελεσματική όταν στο έγγραφο υπάρχει ανομοιογένεια στα κενά διαστήματα ανάμεσα στις γραμμές κειμένου.

...felvesynto vs by our fenecs, and had f
...the condition of our birth, they shoul
...that we have learn't to become judges.
...mans reason is the generall controuller
...which embraceth al, and can do all, by i
...This anfwere were good among the C
...fo much as knowing the name of Phifil
...life. This anfwere might happily avail

(α)

felvesynto vs by our fenecs, and had f
the condition of our birth, they shoul
that we have learn't to become judges.
mans reason is the generall controuller
which embraceth al, and can do all, by i
This anfwere were good among the C
fo much as knowing the name of Phifil
life. This anfwere might happily avail

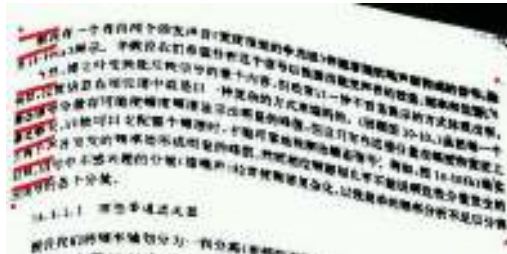
(β)

felvesynto vs by our fenecs, and had f
the condition of our birth, they shoul
that we have learn't to become judges.
mans reason is the generall controuller
which embraceth al, and can do all, by i
This anfwere were good among the C
fo much as knowing the name of Phifil
life. This anfwere might happily avail

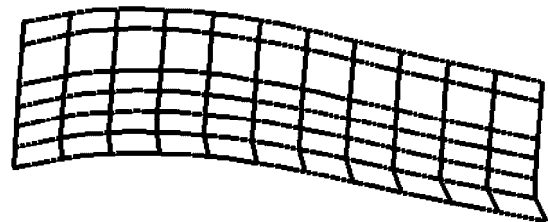
(γ)

Σχήμα 2.17: (α) Αρχικό έγγραφο, (β)-(γ) διορθωμένο έγγραφο με την χρήση καμπυλών Bezier και κυβικών B-splines, αντίστοιχα, σύμφωνα με τη μέθοδο [43].

Στην εργασία [44], οι Wu και Agam δημιουργούν ένα πλέγμα στο παραμορφωμένο έγγραφο προσεγγίζοντας τις γραμμές κειμένου με μη γραμμικές καμπύλες (Σχήμα 2.18) και προσπαθούν να το αντιστοιχίσουν σε ένα μη παραμορφωμένο πλέγμα ώστε να ανακτήσουν το διορθωμένο έγγραφο. Οι μη γραμμικές καμπύλες υπολογίζονται από τον εντοπισμό των πλαισίων των χαρακτήρων κάθε γραμμής κειμένου. Η τεχνική αυτή βασίζεται σε πολλούς ευρετικούς κανόνες και απαιτεί από τον χρήστη να καθορίσει τις τέσσερις γωνίες του πλέγματος στην παραμορφωμένη εικόνα, κάτι που την καθιστά πρακτικά μη εφαρμόσιμη. Επιπλέον, η μέθοδος δε μπορεί να χειριστεί πλέγματα με ανομοιομορφες στήλες.



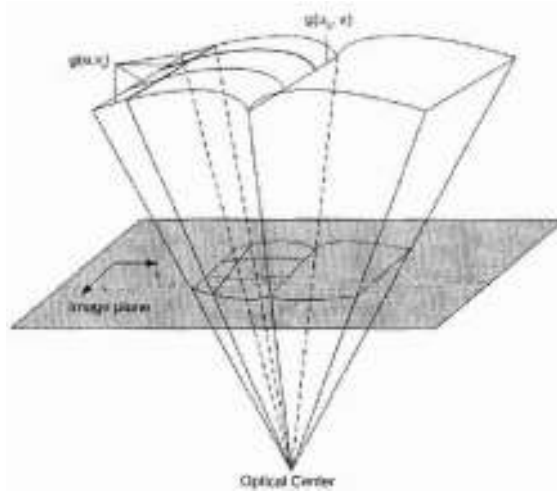
(α)



(β)

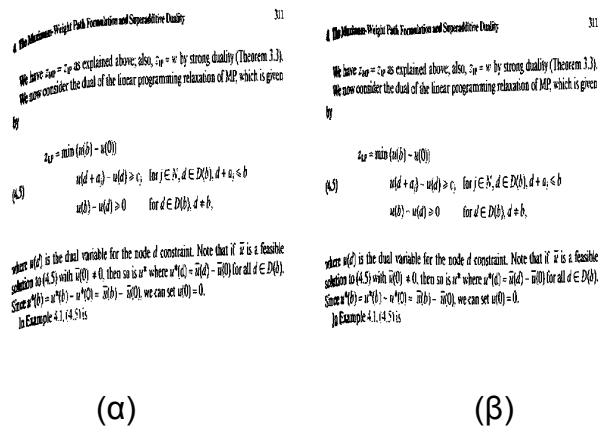
Σχήμα 2.18: (α) Αρχικό έγγραφο στο οποίο προσδιορίζονται οι γραμμές κειμένου (β) το πλέγμα που χτίζεται σύμφωνα με τη μέθοδο [44].

Οι Zhang και Tan [45] βασίζονται στο μοντέλο Gordon [56], το οποίο εξάγεται από τις γραμμές κειμένου, για τη μοντελοποίηση της επιφάνειας του εγγράφου (Σχήμα 2.19). Οι γραμμές κειμένου προσεγγίζονται με κυβικές καμπύλες χρησιμοποιώντας σημεία που προέρχονται από την ανάλυση των συνδεδεμένων συστατικών του εγγράφου. Το μοντέλο Gordon θέτει περιορισμούς σχετικά με τη θέση του εγγράφου.



Σχήμα 2.19: Το μοντέλο Gordon για τη μοντελοποίηση επιφάνειας βιβλίου.

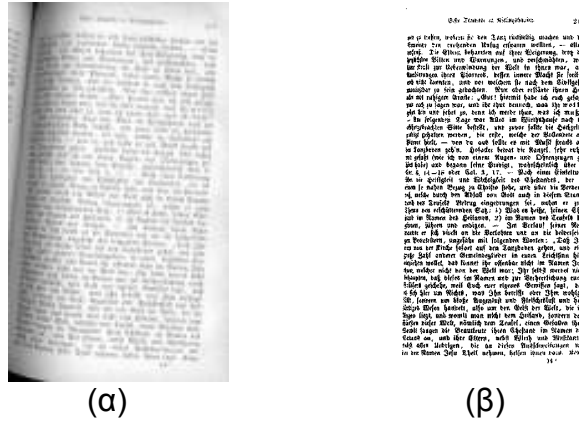
Στο [46], οι Ezaki et al. χρησιμοποιούν επίσης κυβικές καμπύλες, όχι όμως για να προσεγγίσουν μόνο τις γραμμές κειμένου αλλά και τα κενά διαστήματα στο υπόβαθρο ανάμεσα σε αυτές. Για να επιτύχουν πιο ακριβή αποτελέσματα διαχωρίζουν κάθετα το έγγραφο σε επιμέρους μικρότερα τμήματα. Το προτεινόμενο μοντέλο για την αναπαράσταση του εγγράφου δεν έχει ικανοποιητικά αποτελέσματα. Το Σχήμα 2.20 απεικονίζει ένα παράδειγμα εφαρμογής της μεθόδου [46] στο οποίο το αποτέλεσμα δεν είναι ικανοποιητικό.



Σχήμα 2.20: (α) Παραμορφωμένο έγγραφο (β) διορθωμένο έγγραφο μετά την εφαρμογή της μεθόδου [46].

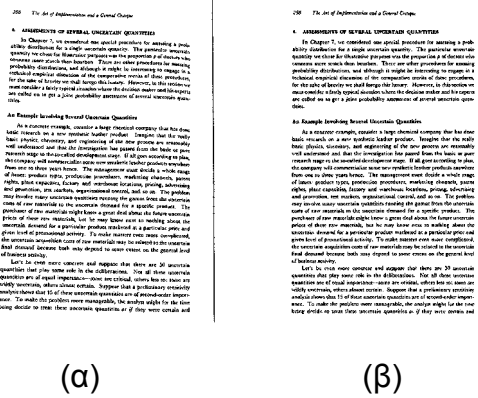
Επίσης, οι Mischke και Luther [47] δημιουργούν ένα πυκνό πλέγμα στο παραμορφωμένο έγγραφο προσεγγίζοντας τις γραμμές κειμένου με πολυώνυμα. Έπειτα, με τη βοήθεια ενός διγραμμικού μετασχηματισμού διορθώνει το παραμορφωμένο έγγραφο. Ο εντοπισμός των γραμμών κειμένου επιτυγχάνεται με τη χρήση του κέντρου μάζας των συνδεδεμένων συστατικών του εγγράφου. Η μέθοδος απαιτεί να έχει εφαρμοστεί ένα στάδιο προ-επεξεργασίας για τη διόρθωση της στροφής του εγγράφου και περιορίζεται σε συγκεκριμένου τύπου

παραμορφώσεις. Το Σχήμα 2.21 απεικονίζει ένα παράδειγμα εφαρμογής της μεθόδου [47].



Σχήμα 2.21: (α) Παραμορφωμένο έγγραφο (β) διορθωμένο έγγραφο μετά από τη δυαδική μετατροπή, τη διόρθωση στροφής και την εφαρμογή της μεθόδου [47].

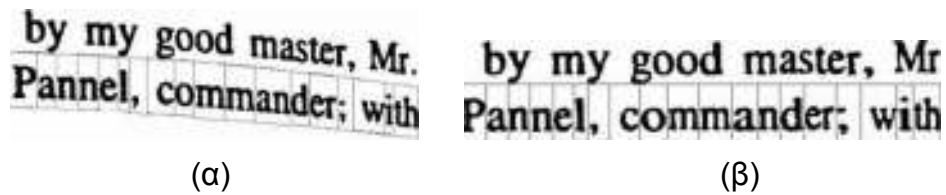
Στην εργασία [48], οι Fu et al. προτείνουν τη μοντελοποίηση του παραμορφωμένου εγγράφου με ένα κυλινδρικό μοντέλο και τη διόρθωση των παραμορφώσεων με τη χρήση ενός μετασχηματισμού. Οι παράμετροι του κυλινδρικού μοντέλου εξαγονται από τις γραμμές κειμένου. Βασικά μειονεκτήματα της μεθόδου είναι ότι βασίζεται σε πολύπλοκους και χρονοβόρους υπολογισμούς, και επιπλέον η μοντελοποίηση του εγγράφου με κύλινδρο δεν είναι πάντα εφικτή. Το Σχήμα 2.22 απεικονίζει ένα παράδειγμα εφαρμογής της μεθόδου [48].



Σχήμα 2.22: (α) Παραμορφωμένο έγγραφο (β) διορθωμένο έγγραφο μετά την εφαρμογή της μεθόδου [48].

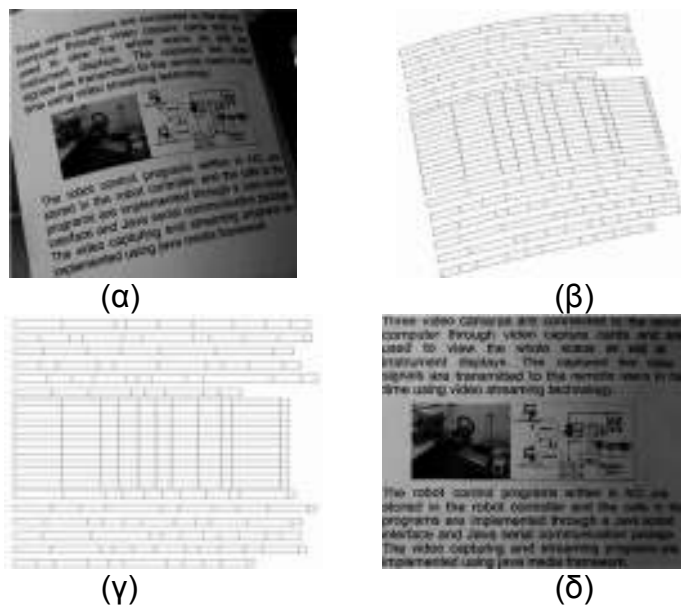
Πολλές τεχνικές δε στηρίζονται μόνο στον εντοπισμό των καμπύλων γραμμών κειμένου στα παραμορφωμένα έγγραφα αλλά δίνουν έμφαση και στον εντοπισμό της γραμμής αναφοράς (baseline). Οι Ulges et al. [49] προτείνουν μία τεχνική η οποία στηρίζεται στην υπόθεση ότι το έγγραφο περιέχει μόνο ισαπέχουσες γραμμές κειμένου ίδιας γραμματσειράς, καθώς και ότι τα κενά διαστήματα ανάμεσα στις λέξεις δεν είναι μεγάλα. Αρχικά, εντοπίζονται ένα τετράπλευρο κελί για κάθε χαρακτήρα με τη βοήθεια της γραμμής αναφοράς. Στη συνέχεια,

αντιστοιχίζουν κάθε κελί σε ένα ορθογώνιο παραλληλόγραμμο το οποίο έχει το σωστό μέγεθος και τη σωστή θέση στο τελικό διορθωμένο έγγραφο (Σχήμα 2.23).



Σχήμα 2.23: Αλυσίδα από κελιά κατά μήκος μίας γραμμής κειμένου (α) στο αρχικό έγγραφο και (β) στο διορθωμένο έγγραφο μετά την εφαρμογή της μεθόδου [49].

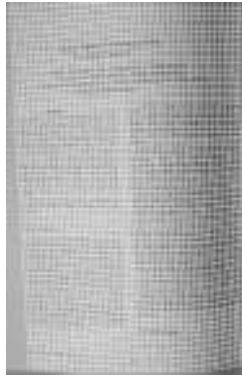
Με παρόμοιο τρόπο, στην εργασία [50], οι Lu et al. χωρίζουν το αρχικό έγγραφο σε πολλές τετράπλευρες περιοχές τις οποίες αντιστοιχίζουν στη συνέχεια στο διορθωμένο έγγραφο με βάση τον αριθμό των χαρακτήρων που βρίσκονται μέσα σε κάθε περιοχή. Το Σχήμα 2.24 απεικονίζει ένα χαρακτηριστικό παράδειγμα κατάτμησης εγγράφου και δημιουργίας του διορθωμένου εγγράφου. Η κατάτμηση του αρχικού εγγράφου βασίζεται στα κάθετα όρια των χαρακτήρων (vertical stroke boundaries – VSBs) καθώς και στις επάνω και κάτω γραμμές αναφοράς των γραμμών κειμένου. Η τεχνική βασίζεται σε πολλούς ευρετικούς κανόνες.



Σχήμα 2.24: Παράδειγμα εφαρμογής της μεθόδου [50] (α) Αρχικό έγγραφο, (β) οι περιοχές κατάτμησης, (γ) οι αντίστοιχες περιοχές στο διορθωμένο έγγραφο, (δ) τελικό αποτέλεσμα.

Στο [51], οι Schneider et al. εξάγουν τοπικά χαρακτηριστικά προσανατολισμού (orientation features) από τις γραμμές αναφοράς με στόχο να δημιουργήσουν ένα πλέγμα στην παραμορφωμένη εικόνα. Έπειτα, κατασκευάζουν το διορθωμένο έγγραφο προσεγγίζοντας τις μη γραμμικές παραμορφώσεις με πολλαπλές γραμμικές προβολές (Σχήμα 2.25). Το βασικό μειονέκτημα της μεθόδου είναι η δυσκολία εύρεσης των χαρακτηριστικών σημείων μετάβασης που απαιτούνται για

τον εντοπισμό των γραμμών αναφοράς, με αποτέλεσμα να οδηγείται συχνά σε λανθασμένα αποτελέσματα.



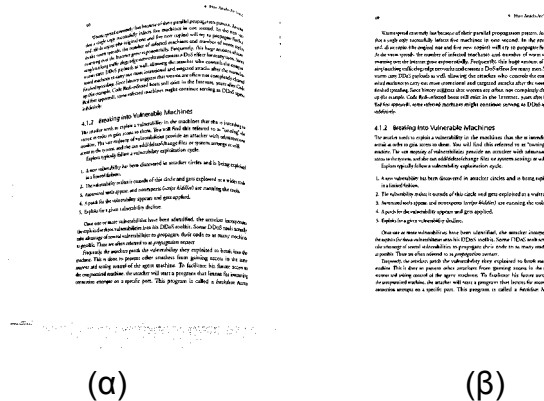
(α)



(β)

Σχήμα 2.25: (α) Παραμορφωμένο έγγραφο μαζί με το πλέγμα που δημιουργήσε η μέθοδος [51], (β) διορθωμένο έγγραφο.

Οι Bukhari et al. [52] εντοπίζουν αρχικά τις καμπύλες γραμμές κειμένου καθώς και το αντίστοιχο ζευγάρι γραμμών αναφοράς, πάνω και κάτω. Στη συνέχεια, αντιστοιχίζουν κάθε χαρακτήρα που περικλείεται από τις δύο καμπύλες γραμμές αναφοράς στο αντίστοιχο κελί από ευθείες, οριζόντιες γραμμές αναφοράς. Η τεχνική είναι πιθανό να αποτύχει όταν οι παραμορφώσεις ποικίλουν μέσα στο έγγραφο και ειδικά κατά μήκος της ίδιας γραμμής κειμένου. Στο Σχήμα 2.26 βλέπουμε ένα παράδειγμα εφαρμογής της μεθόδου των Bukhari et al.



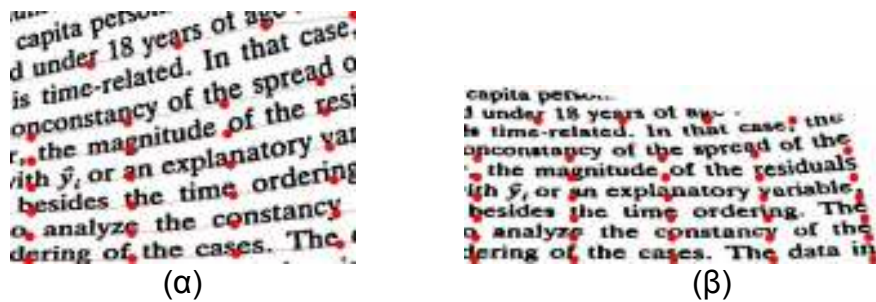
(α)

(β)

Σχήμα 2.26: (α) Παραμορφωμένο έγγραφο (β) διορθωμένο έγγραφο μετά την εφαρμογή της μεθόδου [52].

Στο [53], οι Zhang et al. εντοπίζουν διάφορα σημεία κλειδιά πάνω στο παραμορφωμένο έγγραφο με τη βοήθεια των γραμμών αναφοράς. Στη συνέχεια, χρησιμοποιώντας αυτά τα σημεία εφαρμόζουν Thin-Plane Splines (TPS) παρεμβολή για την ανάκτηση του διορθωμένου εγγράφου. Η προτεινόμενη μέθοδος δε στηρίζεται μόνο στον εντοπισμό των γραμμών κειμένου αλλά και των

χαρακτήρων. Στο παράδειγμα του Σχήματος 2.27 προσδιορίζονται τα σημεία κλειδιά στο αρχικό και στο διορθωμένο έγγραφο.



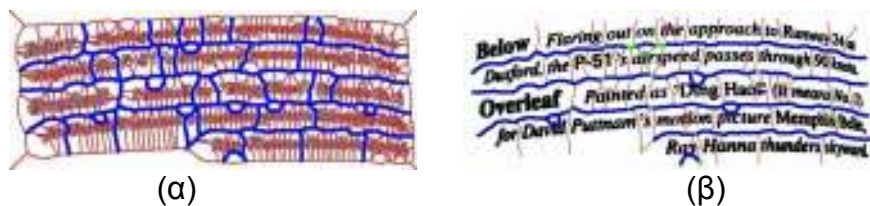
Σχήμα 2.27: (α) Παραμορφωμένο έγγραφο (β) διορθωμένο έγγραφο μετά την εφαρμογή της μεθόδου [53], όπου τα σημεία κλειδιά προσδιορίζονται με κύκλους.

Τέλος, έχουν προταθεί μέθοδοι για τη διόρθωση της τοπικής καμπυλότητας οι οποίες δε στηρίζονται στον εντοπισμό των γραμμών κειμένου. Οι Brown και Tsoi [54] προτείνουν μία μέθοδο για τη διόρθωση της τοπικής καμπυλότητας και της φωτεινότητας σε εικόνες έργων τέχνης (Σχήμα 2.28). Εντοπίζουν τα όρια του εγγράφου και εφαρμόζοντας γραμμική παρεμβολή Coons [57] ανακτούν το διορθωμένο έγγραφο. Για τον εντοπισμό των ορίων του εγγράφου η μέθοδος βασίζεται σε συγκεκριμένο πρότυπο με αποτέλεσμα η εφαρμογή της να είναι περιορισμένη.



Σχήμα 2.28: (α) Παραμορφωμένη εικόνα (β) διορθωμένη εικόνα μετά την εφαρμογή της μεθόδου [54] η οποία διορθώνει και τα προβλήματα της φωτεινότητας.

Οι Masalovitch και Mestetskiy [55] αρχικά δημιουργούν τον σκελετό του εγγράφου (Σχήμα 2.29(α)). Έπειτα, προσεγγίζουν τα κενά διαστήματα ανάμεσα στις γραμμές κειμένου με κυβικές καμπύλες Bezier (Σχήμα 2.29(β)) και χτίζουν το πλέγμα του εγγράφου βάσει του οποίου ανακτάται το διορθωμένο έγγραφο. Η μέθοδος είναι ευαίσθητη στον εντοπισμό των κάθετων ορίων του πλέγματος και έτσι μειώνεται η αποτελεσματικότητά της.



Σχήμα 2.29: (α) Σκελετός εγγράφου, (β) προσέγγιση κενών διαστημάτων με καμπύλες Bezier.

Μία γενική επισκόπηση των μεθόδων διόρθωσης της τοπικής καμπυλότητας που βασίζονται στη διδιάστατη επεξεργασία του εγγράφου παρουσιάζεται στον Πίνακα 2.2. Παρατηρούμε ότι η μεγάλη πλειοψηφία των μεθόδων στηρίζεται στον εντοπισμό των γραμμών κειμένου στα παραμορφωμένα έγγραφα, μία διαδικασία που είναι αρκετά δύσκολη και οδηγεί συχνά σε λανθασμένα αποτελέσματα.

Πίνακας 2.2: Τεχνικές διόρθωσης της τοπικής καμπυλότητας που βασίζονται στη διδιάστατη επεξεργασία του εγγράφου.

	Μέθοδος	Εντοπισμός Γραμμών Κειμένου
Lavialle et al. 2001 [43]	Μοντελοποίηση γραμμών κειμένου με κυβικές B-splines	✓
Wu and Agam 2002 [44]	Δημιουργία πλέγματος με μη-γραμμικές καμπύλες	✓
Zhang and Tan 2005 [45]	Μοντελοποίηση εγγράφου με το μοντέλο Gordon [56]	✓
Ezaki et al. 2005 [46]	Μοντελοποίηση γραμμών κειμένου και υπόβαθρου με κυβικές καμπύλες	✓
Mischke and Luther 2005 [47]	Δημιουργία πλέγματος με πολυώνυμα και χρήση διγραμμικού μετασχηματισμού	✓
Fu et al. 2007 [48]	Μοντελοποίηση εγγράφου με κύλινδρο	✓
Ulges et al. 2005 [49]	Δημιουργία κελιών με τη χρήση γραμμών αναφοράς	✓
Lu et al. 2006 [50]	Δημιουργία πλέγματος με τη βοήθεια των κάθετων ορίων των χαρακτήρων και των γραμμών αναφοράς	✓
Schneider et al. 2007 [51]	Δημιουργία πλέγματος με τοπικά χαρακτηριστικά προσανατολισμού	✓
Bukhari et al. 2009 [52]	Μοντελοποίηση γραμμών κειμένου με ζευγάρια γραμμών αναφοράς	✓
Zhang et al. 2008 [53]	Εφαρμογή Thin-Plane Splines (TPS) παρεμβολής χρησιμοποιώντας σημεία κλειδιά	✓
Brown and Tsoi 2006 [54]	Εντοπισμός των ορίων του εγγράφου και εφαρμογή γραμμικής παρεμβολής Coons [57]	
Masalovitch and Mestetskiy 2007 [55]	Δημιουργία πλέγματος από τον σκελετό του εγγράφου χρησιμοποιώντας κυβικές καμπύλες Bezier	

2.2 Ανάλυση Δομής και Κατάτμηση

Η κατάτμηση των ψηφιακών εγγράφων αφορά τη διαδικασία του εντοπισμού των δομικών στοιχείων (structure elements) με σκοπό την περιγραφή και την ανάλυση της δομής τους. Η διαδικασία της κατάτμησης αποτελεί ένα σημαντικό στάδιο της επεξεργασίας των εγγράφων.

Οι δομές των εγγράφων χωρίζονται σε δύο κατηγορίες, τις φυσικές (κείμενο, εικόνες, γραφήματα, πίνακες, γραμμές κ.λπ.) και τις λογικές (τίτλος, παράγραφος, επικεφαλίδες, λεζάντες κ.λπ.). Όταν μία εφαρμογή εξετάζει τις φυσικές δομές του εγγράφου ανήκει στις μεθόδους **ανάλυσης φυσικής δομής** (Physical/Geometric Document Structure Analysis), ενώ όταν αναλύει τις λογικές δομές του εγγράφου κατατάσσεται στις μεθόδους **ανάλυσης λογικής δομής** (Logical/Functional Document Structure Analysis). Για να είναι δυνατή η λογική ανάλυση του εγγράφου είναι απαραίτητο να προηγηθεί η ανάλυση της φυσικής του δομής. Η βασική κατηγοριοποίηση των τεχνικών κατάτμησης γίνεται με κριτήριο τη σειρά ομαδοποίησης των δομικών στοιχείων του εγγράφου. Συνεπώς, προκύπτουν οι παρακάτω κατηγορίες:

- Οι τεχνικές ανάλυσης «από πάνω προς τα κάτω» (top-down) [58], [59]. Ξεκινούν την ανάλυση ανιχνεύοντας μεγάλης κλίμακας δομές (π.χ. στήλες κειμένου) και με διαδοχικές διαιρέσεις καταλήγουν σε μικρής κλίμακας δομές (π.χ. λέξεις ή μεμονωμένοι χαρακτήρες).
- Οι τεχνικές ανάλυσης «από κάτω προς τα πάνω» (bottom-up) [60], [61]. Ενώνονται διαδοχικά συστατικά μικρού μεγέθους και προοδευτικά σχηματίζονται δομές μεγαλύτερης κλίμακας, π.χ. ομαδοποίηση χαρακτήρων για να τη δημιουργία λέξεων, λέξεων για τη δημιουργία γραμμών κειμένου κ.λπ.
- Οι υβριδικές τεχνικές στις οποίες η φιλοσοφία των παραπάνω κατηγοριών συνδυάζεται για να βελτιωθεί το αποτέλεσμα [62].

2.2.1 Ανάλυση και Κατανόηση της Δομής του Εγγράφου

Η ανάλυση και η κατανόηση της δομής του εγγράφου (Page Layout Analysis) αποτελεί τη διαδικασία της αναγνώρισης και της περιγραφής των δομών ενός εγγράφου. Διαχωρίζεται σε δύο είδη εργασιών, την κατάτμηση της σελίδας και την ταξινόμηση των εξαγόμενων περιοχών. Η κατάτμηση αφορά την ομαδοποίηση των δομικών στοιχείων σε ομοιογενή τμήματα ενώ η ταξινόμηση κατατάσσει το κάθε τμήμα στην κατάλληλη κατηγορία (π.χ. κείμενο, εικόνα, γράφημα κ.λπ.).

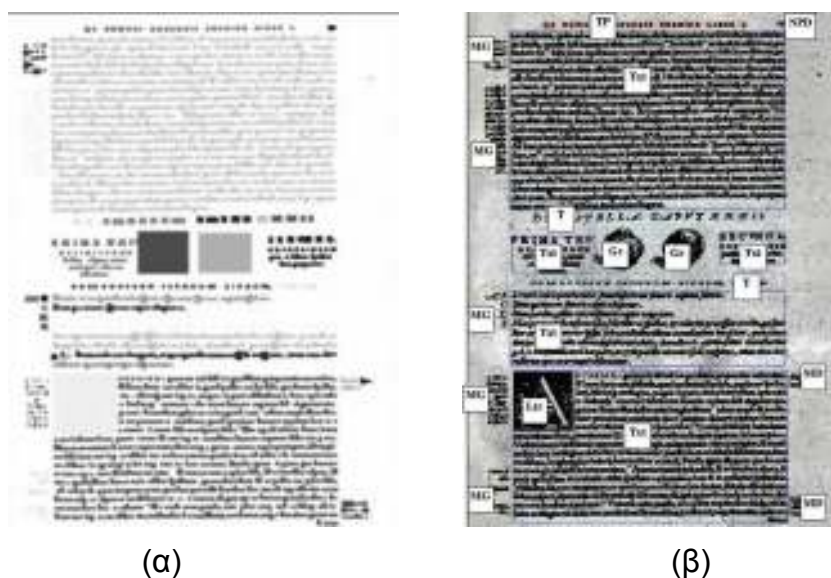
Πολλές τεχνικές έχουν προταθεί για την ανάλυση και την κατανόηση της δομής στα σύγχρονα τυπωμένα έγγραφα. Στην εργασία [61], ακολουθώντας την «από κάτω προς τα πάνω» προσέγγιση, η δομή του εγγράφου αναπαρίσταται ιεραρχικά με τη μορφή δέντρου έτσι ώστε να είναι δυνατός ο ορισμός της σειράς

ανάγνωσης. Δηλαδή, ξεκινώντας από τα συνδεδεμένα συστατικά της εικόνας, δημιουργούνται διαδοχικά οι λέξεις, οι γραμμές κειμένου, οι παράγραφοι κ.λπ. Η μέθοδος θεωρεί τα συνδεδεμένα συστατικά ως τις κορυφές ενός δέντρου και τις αποστάσεις μεταξύ τους ως τις πιθανές συνδέσεις των κορυφών, αποδίδοντας συντελεστές κόστους σε κάθε μία. Το πρόβλημα της συνένωσης των συστατικών σε λέξεις, γραμμές κειμένου κ.λπ. μπορεί να αντιμετωπιστεί βρίσκοντας δέντρα ελαχίστου κόστους. Αντίθετα, η εργασία [59] ακολουθεί την «από πάνω προς τα κάτω» μεθοδολογία και η κατάτμηση του εγγράφου υλοποιείται μέσω μίας διαδικασίας περιγραφής της εικόνας σε διαφορετικά επίπεδα ανάλυσης υπό τη μορφή πυραμίδας (pyramidal quadtree structure). Επίσης, η μέθοδος [63] βασίζεται στην κατάτμηση του εγγράφου μέσω της ανάλυσης του υπόβαθρου και της αναπαράστασής του με λευκές (χρώμα του υπόβαθρου) ορθογώνιες περιοχές (white tiles).

Οι περισσότερες τεχνικές που προτείνονται για τα τυπωμένα κείμενα δε μπορούν να βρουν εφαρμογή στα χειρόγραφα λόγω της μεταβλητότητας του χώρου. Τεχνικές που αναπτύχθηκαν για χειρόγραφα βρίσκουν εφαρμογή μόνο σε συγκεκριμένες κατηγορίες εγγράφων. Αυτό ήταν και το ερέθισμα για τη δημιουργία μίας γενικευμένης τεχνικής, η οποία θα προσαρμόζεται ανάλογα με τον τύπο του εγγράφου χρησιμοποιώντας κάθε φορά συναφείς πληροφορίες και χαρακτηριστικά. Η εργασία [64] αφορά την κατάτμηση και την ανάλυση της δομής περίπλοκων χειρόγραφων κειμένων και βασίζεται στα Μοντέλα Τυχαίων Μαρκοβιανών Πεδίων (Markov Random Field Models). Η κατάτμηση ανάγεται σε μία διαδικασία όπου τοποθετούνται ετικέτες μέσα στο έγγραφο, οι οποίες υποδηλώνουν τη μορφή του, και το πρόβλημα λύνεται με τεχνικές βελτιστοποίησης μίας συνάρτησης κατανομών.

Επιπλέον, έχουν προταθεί τεχνικές για την ανάλυση και κατανόηση της δομής ειδικά για τα ιστορικά έγγραφα. Η εργασία [65] αφορά ιστορικά τυπωμένα έγγραφα. Βασικός της στόχος είναι να διαχωρίσει τα έγγραφα σε περιοχές που περιέχουν είτε κείμενο είτε γράφημα. Το έγγραφο χωρίζεται σε ομοιογενείς περιοχές οι οποίες εξετάζονται ως προς τον προσανατολισμό τους με τη χρήση της συνάρτησης αυτοσυσχέτισης και κατηγοριοποιούνται ως κείμενο ή γράφημα. Οι Bulacu et al. [66], προτείνουν μία μέθοδο που αποτελείται από δύο στάδια. Στο πρώτο στάδιο γίνεται ένας πρώτος, πρόχειρος, εντοπισμός των περιοχών του εγγράφου χρησιμοποιώντας μόνο ευθείες γραμμές. Στη συνέχεια, έχοντας για

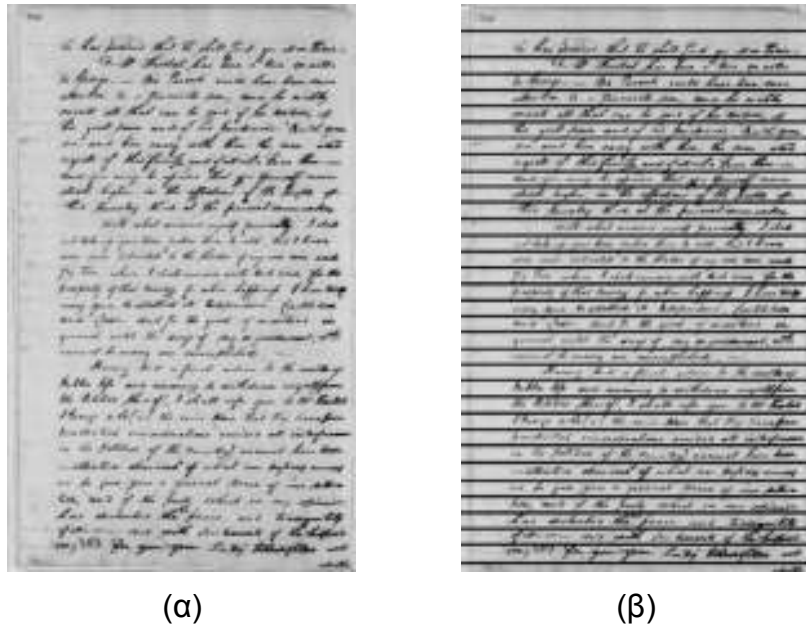
οδηγούς τις παραπάνω ευθείες, οριοθετούν τις περιοχές με καμπυλόγραμμους οδηγούς με τη χρήση των περιγραμμάτων των αντικειμένων. Στο [67] περιγράφεται μία υβριδική μεθοδολογία η οποία συνδυάζει τεχνικές ανάλυσης «από πάνω προς τα κάτω» και «από κάτω προς τα πάνω» και συγχρόνως ο χρήστης επεμβαίνει για να καθορίσει το τελικό αποτέλεσμα. Η μέθοδος αναλύει τα συνδεδεμένα συστατικά και το υπόβαθρο του εγγράφου για να πετύχει την καλύτερη κατάτμηση του εγγράφου. Έπειτα, ταξινομεί τις περιοχές με τη βοήθεια διάφορων σεναρίων που παράγει ο χρήστης. Στο Σχήμα 2.30 παρουσιάζεται ένα παράδειγμα εφαρμογής της παραπάνω μεθόδου. Τέλος, στο [68], οι Baechler και Ingold, αναπαριστούν τις περιοχές του εγγράφου με τη βοήθεια πολυγώνων και τις ταξινομούν σε τέσσερις κατηγορίες: (i) τις κατεστραμμένες περιοχές οι οποίες δεν αναγνωρίζονται από τον άνθρωπο και είναι συνηθισμένες στα ιστορικά έγγραφα, (ii) τις περιοχές κειμένου, (iii) τις περιοχές σχολίων που αφορά χειρόγραφο κείμενο το οποίο προστέθηκε αργότερα δίπλα στο αρχικό κείμενο και (iv) τις διακοσμητικές περιοχές.



Σχήμα 2.30: Παράδειγμα εφαρμογής της μεθόδου [67]: (α) Αποτέλεσμα κατάτμησης του εγγράφου (κάθε περιοχή επισημαίνεται με διαφορετικό επίπεδο του γκρι), (β) αποτέλεσμα ταξινόμησης των περιοχών με τη βοήθεια του χρήστη (MG-Αριστερό περιθώριο, MD-Δεξιό περιθώριο, Ltr-Διακοσμητικό γράμμα, NPD-Δεξιά αρίθμηση σελίδας, TP-Πρωτεύων τίτλο, T-Τίτλος).

2.2.2 Κατάτμηση Γραμμών Κειμένου

Το επόμενο βήμα, αφού εντοπιστούν οι περιοχές κειμένου, είναι η κατάτμηση των γραμμών κειμένου (Σχήμα 2.31).



Σχήμα 2.31: Κατάτμηση γραμμών κειμένου.

Η πιο γνωστή τεχνική για την κατάτμηση γραμμών κειμένου είναι οι οριζόντιες προβολές. Τα τοπικά ελάχιστα των προβολών είναι υποψήφια σημεία κατάτμησης γιατί συνήθως αντιπροσωπεύουν το κενό διάστημα μεταξύ των γραμμών κειμένου. Στην περίπτωση όμως των ιστορικών εγγράφων είναι δύσκολο να βρεθεί ένα «καθαρό» και ευθυγραμμισμένο κείμενο ώστε να χρησιμοποιηθούν με επιτυχία οι οριζόντιες προβολές. Για τον σκοπό αυτό έχουν προταθεί αλγόριθμοι που προσπαθούν να αντιμετωπίσουν τις δύσκολες αυτές περιπτώσεις.

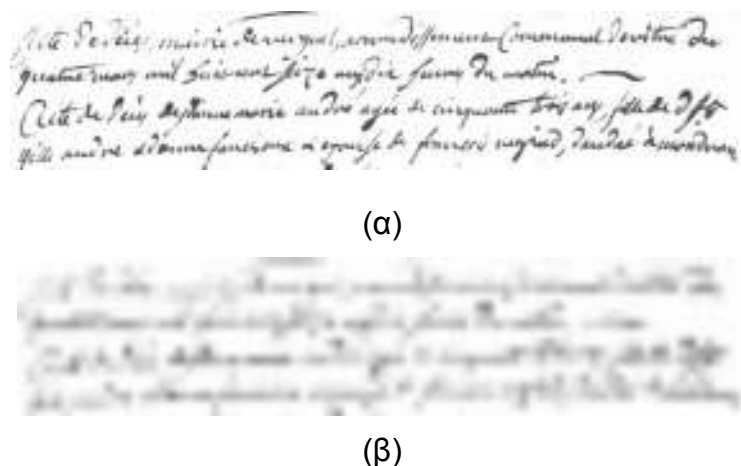
Στο [69] η κατάτμηση γραμμών κειμένου μοντελοποιείται ως ένα γενικότερο πρόβλημα κατάτμησης εικόνας. Η εικόνα εξομαλύνεται με ένα Gaussian φίλτρο και μετατρέπεται από δυαδική εικόνα σε εικόνα διαβάθμισης γκρι. Έπειτα, γίνεται μία αρχική εκτίμηση των γραμμών κειμένου επαναφέροντας την εικόνα σε δυαδική και χρησιμοποιείται μία συνάρτηση μερικών παραγώγων για την εξαγωγή του περιγράμματος των γραμμών κειμένου. Τέλος, εξάγονται οι γραμμές κειμένου και διαχωρίζονται από το θόρυβο. Η μέθοδος επιτυγχάνει υψηλά ποσοστά στην περίπτωση των χειρόγραφων εγγράφων. Οι Bruzzone και Coffetti [70] προτείνουν μία τεχνική απομόνωσης των γραμμών κειμένου χωρίς απώλεια της πληροφορίας. Ο προτεινόμενος αλγόριθμος βασίζεται στο ιστόγραμμα συνεχόμενων pixels κειμένου και στην ταξινόμηση των συνδεδεμένων συστατικών της εικόνας. Το πλεονέκτημα της μεθόδου έγκειται στην προστασία και στο σωστό διαχωρισμό των χαρακτήρων που έχουν τμήματά τους έξω από τη βασική περιοχή της γραμμής, όπως οι χαρακτήρες b, d, f, g, k, l κ.λπ. Η εργασία

[71] έχει ως στόχο να εντοπίσει τις γραμμές κειμένου σε ιστορικά χειρόγραφα έγγραφα. Χρησιμοποιεί ένα στάδιο προ-επεξεργασίας, και στη συνέχεια, χρησιμοποιεί έναν πίνακα μεταβάσεων (κείμενο – υπόβαθρο και αντίστροφα) για να εντοπίσει τις γραμμές κειμένου. Έπειτα, εφαρμόζει τον αλγόριθμο min cut/max flow για να διαχωρίσει τις ενωμένες γραμμές κειμένου. Τέλος, μικρά αντικείμενα που δεν ανήκουν σε γραμμές κειμένου ενσωματώνονται σε κάποιες από αυτές ή αφαιρούνται ως θόρυβος. Στο Σχήμα 2.32 απεικονίζεται το αποτέλεσμα μετά την εφαρμογή του πίνακα μεταβάσεων και πριν την τελική επεξεργασία.



Σχήμα 2.32: (α) Ιστορικό χειρόγραφο έγγραφο, (β) αποτέλεσμα μετά την εφαρμογή του πίνακα μεταβάσεων σύμφωνα με τη μέθοδο [71].

Στο [72] οι συγγραφείς ασχολούνται επίσης με την κατάτμηση γραμμών κειμένου σε ιστορικά χειρόγραφα έγγραφα. Η μέθοδος βασίζεται στην παρατήρηση ότι σε ένα έγγραφο με μικρή ανάλυση οι γραμμές κειμένου ενώνονται σε ευθύγραμμα τμήματα (Σχήμα 2.33). Ο εντοπισμός των γραμμών στηρίζεται στα φίλτρα Kalman.



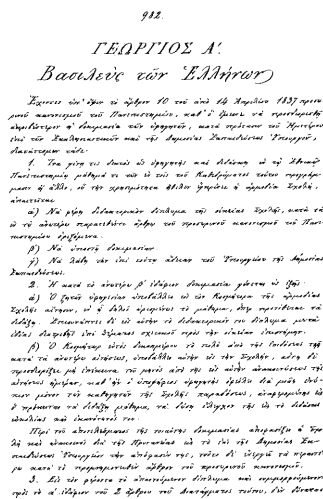
Σχήμα 2.33: (α) Αρχική εικόνα, (β) εικόνα μειωμένης ανάλυσης.

Στο [73] παρουσιάζεται μία ευέλικτη μέθοδος για την κατάτμηση ιστορικών εγγράφων εφαρμόζοντας οριζόντιες προβολές σε εικόνα διαβάθμισης γκρι. Επίσης, οι Feldbach και Tonpiet [74] προτείνουν μία μέθοδο κατάτμησης γραμμών κειμένου σε ιστορικά εκκλησιαστικά έγγραφα. Ο προτεινόμενος αλγόριθμος διαχωρίζει τις γραμμές κειμένου στις περιπτώσεις που είναι πολύ κοντά η μία στην άλλη ή ενωμένες ή όταν οι γραμμές δεν είναι απόλυτα ευθείες. Στο [75] υπάρχει εκτενής αναφορά σχετικά με τις μεθόδους κατάτμησης γραμμών κειμένου σε ιστορικά έγγραφα.

Επίσης, μία διαδοδομένη τεχνική για την κατάτμηση των γραμμών κειμένου είναι ο μετασχηματισμός Hough. Στο [76] εξάγονται τα συνδεδεμένα συστατικά και υπολογίζεται το μέσο ύψος του γράμματος. Στη συνέχεια, ένας block-based μετασχηματισμός Hough χρησιμοποιείται για τον εντοπισμό πιθανών γραμμών κειμένου. Τέλος, ενώνονται οι κατακερματισμένες γραμμές και εντοπίζονται γραμμές κειμένου που δε βρέθηκαν στο προηγούμενο βήμα.

2.2.3 Κατάτμηση Λέξεων

Μετά την κατάτμηση των γραμμών κειμένου ακολουθεί η κατάτμηση λέξεων (Σχήμα 2.34). Οι πιο συνηθισμένες μέθοδοι είναι οι κατακόρυφες προβολές των γραμμών κειμένου και η ανάλυση των συνδεδεμένων συστατικών (connected component analysis) [77].



(α)



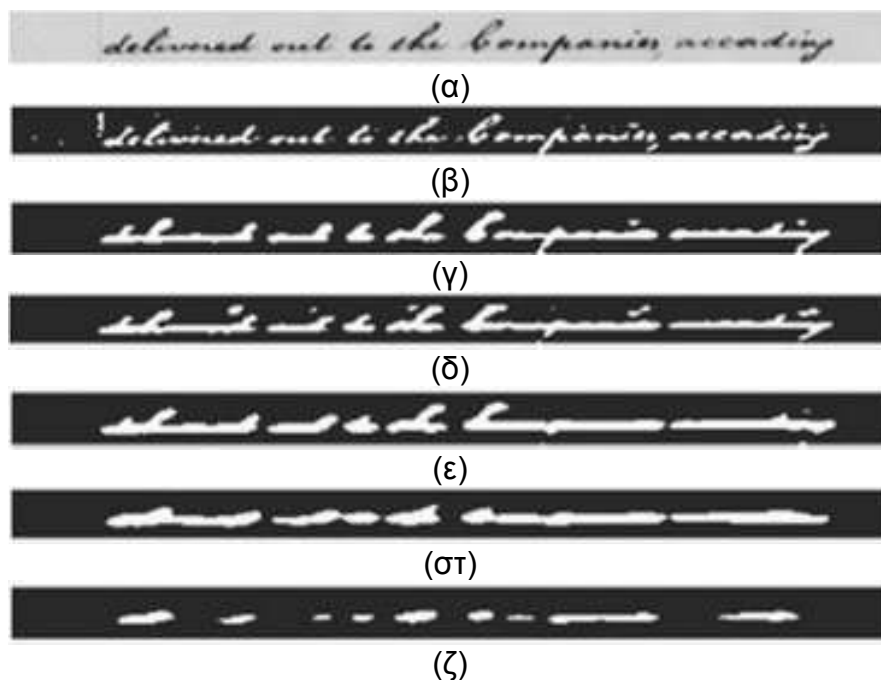
(β)

Σχήμα 2.34: Κατάτμηση λέξεων.

Σε αντίθεση με τα τυπωμένα έγγραφα, στα χειρόγραφα δεν υπάρχουν σταθερά κενά διαστήματα ανάμεσα στις λέξεις και τους χαρακτήρες. Η εργασία [78] εκμεταλλεύεται τα κενά διαστήματα των λέξεων μέσα στις γραμμές κειμένου. Η

καινοτομία της μεθόδου οφείλεται στο γεγονός ότι η απόφαση για το που βρίσκονται τα κενά μέσα σε μία γραμμή κειμένου λαμβάνεται σε σχέση με τον τρόπο γραφής του κειμένου. Οι παράμετροι του αλγορίθμου, οι οποίοι καθορίζουν εάν το κενό διάστημα είναι ανάμεσα σε λέξεις ή χαρακτήρες, δε χρειάζονται μεγάλο δείγμα εκπαίδευσης. Αντιθέτως, προσαρμόζονται κάθε φορά στην εξεταζόμενη γραμμή κειμένου. Προς αυτή την κατεύθυνση κινείται και η εργασία που προτείνεται στο [79] για τυπωμένα έγγραφα. Ανιχνεύει τα κενά διαστήματα και τα σύμβολα (σημεία στίξης) του εγγράφου ως αριστερά και δεξιά όρια των λέξεων. Η ανίχνευση των κενών διαστημάτων πραγματοποιείται με την ταξινόμηση όλων των κενών διαστημάτων της γραμμής κειμένου σε κατηγορίες. Η μέθοδος προορίζεται για την αγγλική και την κορεάτικη γραφή, ενώ περιλαμβάνει και μία διαδικασία αναγνώρισης ειδικών συμβόλων.

Η εργασία [80] έχει ως στόχο την κατάτμηση λέξεων σε ιστορικά χειρόγραφα έγγραφα. Η μέθοδος δέχεται ως είσοδο εικόνες διαβάθμισης γκρι και πραγματοποιείται η κατάτμηση γραμμών κειμένου με τη χρήση οριζόντιων προβολών. Έπειτα, κάθε γραμμή κειμένου φιλτράρεται με ένα Laplacian φίλτρο σε διαφορετικές κλίμακες. Η διαδικασία αυτή οδηγεί στην παραγωγή κηλίδων που σε μικρή κλίμακα μπορεί να αντιστοιχούν σε χαρακτήρες και σε μεγαλύτερη κλίμακα σε λέξεις (Σχήμα 2.35). Είναι σημαντικό για τη διαδικασία να επιλεγεί η κατάλληλη κλίμακα στην οποία οι κηλίδες θα αντιστοιχούν σε λέξεις.

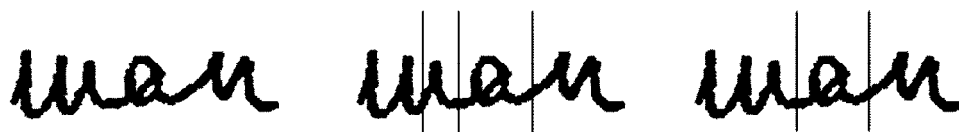


Σχήμα 2.35: Τεχνική δημιουργίας κηλίδων για την κατάτμηση λέξεων με τη χρήση Laplacian φίλτρου. (α) Γραμμή κειμένου, (β)-(ζ) αποτελέσματα για διαφορετικές κλίμακες.

Τέλος, η κατάτμηση των γραμμών κειμένου δεν είναι πάντα απαραίτητη για την κατάτμηση των λέξεων. Η εργασία [81] διαπραγματεύεται το θέμα της ανάκτησης ιστορικών εγγράφων με βάση το περιεχόμενό τους (content-based image retrieval). Η προσέγγιση του προβλήματος εστιάζεται στον εντοπισμό των λέξεων του εγγράφου οι οποίες δίνονται υπό μορφή ερωτήματος στο σύστημα. Ως λέξεις ορίζονται τα συνδεδεμένα συστατικά που προκύπτουν μετά την εφαρμογή του αλγορίθμου εξομάλυνσης RLSA και έχουν τις κατάλληλες διαστάσεις σύμφωνα με παραμέτρους που βασίζονται στο μέσο ύψος του γράμματος.

2.2.4 Κατάτμηση Χαρακτήρων

Έχοντας εντοπίσει τις λέξεις στις γραμμές κειμένου, στο επόμενο στάδιο προχωράμε στην κατάτμηση των χαρακτήρων. Ανεξάρτητα από τη μορφή του προβλήματος που έχουμε να αντιμετωπίσουμε το στάδιο κατάτμησης των χαρακτήρων επηρεάζει σημαντικά την απόδοση ενός συστήματος οπτικής αναγνώρισης καθώς η ύπαρξη χαρακτήρων που διαιρούνται σε δύο κομμάτια ή η συνένωση γειτονικών χαρακτήρων μπορεί να οδηγήσει σε σοβαρά σφάλματα στο στάδιο της αναγνώρισης (Σχήμα 2.36).

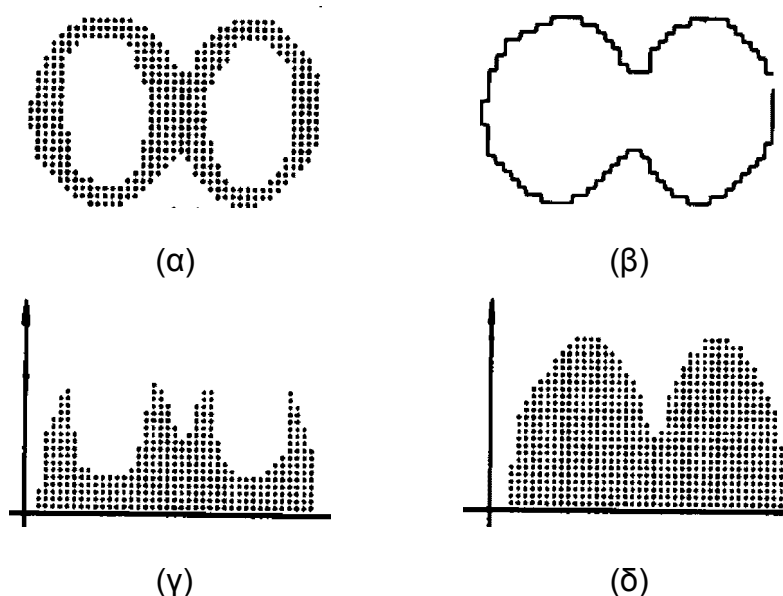


Σχήμα 2.36: Εντοπισμός υποψήφιων σημείων κατάτμησης χαρακτήρων.

Οι μέθοδοι κατάτμησης χαρακτήρων μπορούν να ταξινομηθούν σε δύο βασικές κατηγορίες ανάλογα με την προσέγγιση που ακολουθούν. Στην πρώτη κατηγορία ανήκουν οι μέθοδοι στις οποίες η κατάτμηση είναι ανεξάρτητη από την αναγνώριση (explicit segmentation) ενώ αντιθέτως, στη δεύτερη κατηγορία η κατάτμηση βασίζεται στην αναγνώριση (implicit segmentation). Σύμφωνα με την πρώτη προσέγγιση η λέξη χωρίζεται ρητώς σε χαρακτήρες οι οποίοι στη συνέχεια αναγνωρίζονται ο καθένας ξεχωριστά. Αντιθέτως, σύμφωνα με τη δεύτερη προσέγγιση, η κατάτμηση και η αναγνώριση λαμβάνουν χώρα ταυτόχρονα. Συνεπώς, το στάδιο της κατάτμησης δέχεται επιπλέον πληροφορία και προσαρμόζεται ανάλογα με το αποτέλεσμα της αναγνώρισης. Επιπρόσθετα, στη δεύτερη κατηγορία ανήκουν και οι τεχνικές στις οποίες οι λέξεις αναγνωρίζονται εξ' ολοκλήρου χωρίς κατάτμηση σε χαρακτήρες και χρησιμοποιούνται κυρίως σε περιπτώσεις στις οποίες το σύνολο των δεδομένων είναι μικρό και γνωστό εκ των

προτέρων, όπως τραπεζικές επιταγές και ταχυδρομικές διευθύνσεις. Στη βιβλιογραφία συναντώνται πολλές τεχνικές για την κατάτμηση τυπωμένων και χειρόγραφων εγγράφων καθώς και ψηφίων. Στη συνέχεια περιγράφονται οι βασικότερες μέθοδοι κατάτμησης χαρακτήρων.

Οι Liang et al. [82] προτείνουν ένα επαναληπτικό αλγόριθμο κατάτμησης τυπωμένων χαρακτήρων. Αρχικά, υπολογίζει υποψήφια σημεία τομής (break points) και στη συνέχεια προχωρά επαναληπτικά σε περαιτέρω κατάτμηση ή συνένωση ελέγχοντας την ορθογραφία των λέξεων. Συνεπώς, ελαχιστοποιεί τις πιθανότητες λανθασμένης κατάτμησης. Ο αλγόριθμος χρησιμοποιεί δύο είδη προβολών, τις κατακόρυφες προβολές των pixels κειμένου των χαρακτήρων καθώς και τις προβολές των pixels υπόβαθρου που περικλείονται από το περίγραμμα των χαρακτήρων (Σχήμα 2.37).



Σχήμα 2.37: (α) Συνδεδεμένοι χαρακτήρες “oo”, (β) περίγραμμα χαρακτήρων, (γ) κατακόρυφες προβολές, (δ) προβολές υπόβαθρου που περικλείεται από το περίγραμμα.

Οι Chang και Chen [83] προτείνουν μία μέθοδο που επικεντρώνεται στην κατάτμηση των ενωμένων χαρακτήρων καθώς προσαρμόζουν τις παραμέτρους της δυαδικής μετατροπής ώστε να αποφεύγονται οι σπασμένοι χαρακτήρες. Οι ενωμένοι χαρακτήρες εντοπίζονται χρησιμοποιώντας τοπολογικές ιδιότητες καθώς και τυπογραφικούς κανόνες. Πιο συγκεκριμένα, οι τοπολογικές ιδιότητες εξάγονται από τη χωρική διάταξη του κυρτού κελύφους (convex hull) που περικλείει τους χαρακτήρες. Στη συνέχεια, ένας αλγόριθμος εύρεσης του μικρότερου μονοπατιού μαζί με την πληροφορία από το κυρτό κέλυφος χρησιμοποιούνται για να διαχωρίσουν τους χαρακτήρες. Τα χαρακτηριστικά που εξάγονται από το κυρτό κέλυφος δεν επηρεάζονται από τον τύπο ή το μέγεθος

της γραμματοσειράς, συνεπώς η μέθοδος μπορεί να χειριστεί και δύσκολες περιπτώσεις ενωμένων ή επικαλυπτόμενων χαρακτήρων χωρίς να προηγηθεί διόρθωση της στροφής. Στο Σχήμα 2.38 παρουσιάζονται μερικά παραδείγματα χαρακτήρων περικλειόμενα από το κυρτό κέλυφος.



Σχήμα 2.38: Χαρακτήρες περικλειόμενοι από το κυρτό κέλυφος. Τα pixels που απεικονίζονται με μαύρο, ανοιχτό γκρι και σκούρο γκρι χρώμα αντιπροσωπεύουν τους χαρακτήρες, το κυρτό σχήμα και την περιοχή ανάμεσα στο κυρτό σχήμα και του πλαισίου του χαρακτήρα, αντίστοιχα.

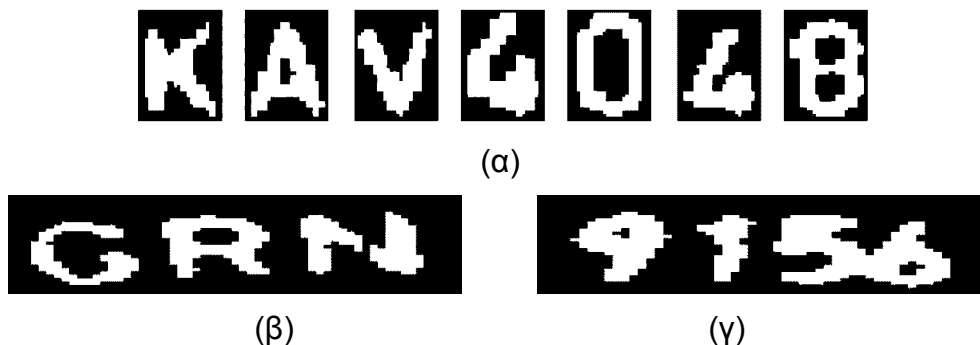
Στο [73], οι Antonacopoulos και Karatzas χρησιμοποιούν τις κατακόρυφες προβολές της λέξης για τον διαχωρισμό των χαρακτήρων σε ιστορικά τυπωμένα έγγραφα (Σχήμα 2.39). Το πρώτο σημείο διαχωρισμού εντοπίζεται σύμφωνα με το αναμενόμενο πλάτος χαρακτήρα και υπολογίζεται με ακρίβεια χρησιμοποιώντας πληροφορία από τα τοπικά ελάχιστα των προβολών. Το επόμενο σημείο διαχωρισμού εντοπίζεται με τον ίδιο τρόπο χρησιμοποιώντας επιπλέον πληροφορία από το προηγούμενο σημείο διαχωρισμού. Η μέθοδος δε μπορεί να χειριστεί περιπτώσεις επικαλυπτόμενων χαρακτήρων.



Σχήμα 2.39: Κατάτμηση χαρακτήρων με χρήση προβολών σύμφωνα με τη μέθοδο [73].

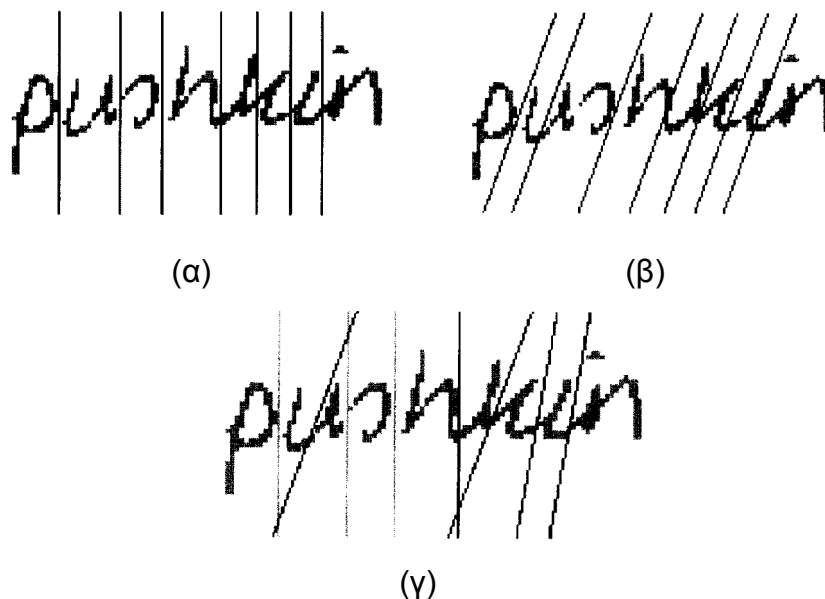
Στην εργασία [84], οι Nomura et al. προτείνουν έναν αλγόριθμο που βασίζεται στα ιστογράμματα. Ανιχνεύει σημεία τομής για τους κατακερματισμένους χαρακτήρες και διαχωρίζει χαρακτήρες που παρουσιάζουν επικάλυψη. Βασικό πεδίο εφαρμογής της εργασίας είναι οι πινακίδες αυτοκινήτων, συνεπώς θεωρεί ότι κάθε εικόνα αποτελείται από ένα σύνολο τριών χαρακτήρων που φτάνει μέχρι τη μέση και ένα σύνολο τεσσάρων ψηφίων. Αν δεν υπάρχουν επικαλυπτόμενοι ή κατακερματισμένοι χαρακτήρες ο διαχωρισμός επιτυγχάνεται εύκολα με κατακόρυφες προβολές. Αν υπάρχει επικάλυψη, εντοπίζει τα σημεία κατάτμησης χρησιμοποιώντας μία συνάρτηση κόστους γνωρίζοντας το συνολικό αριθμό των χαρακτήρων. Τέλος, αν υπάρχουν κατακερματισμένοι χαρακτήρες τους ενώνει με βάσει τις αποστάσεις των χαρακτήρων, το αναμενόμενο πλάτος και τα κενά

διαστήματα μεταξύ τους. Παραδείγματα των παραπάνω περιπτώσεων παρουσιάζονται στο Σχήμα 2.40.



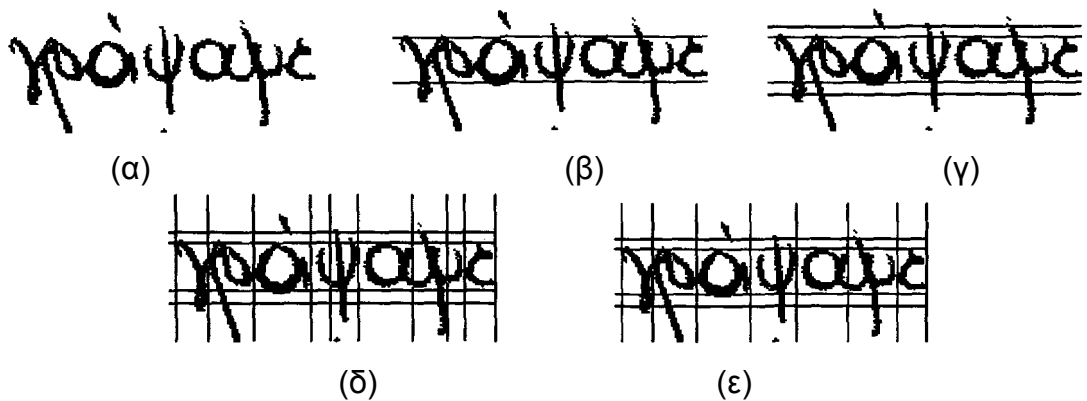
Σχήμα 2.40: Διαφορετικά προβλήματα στις εικόνες πινακίδων. (α) Ευδιάκριτοι, (β) κατακερματισμένοι και (γ) ενωμένοι χαρακτήρες.

Οι Yanikoglu και Sandon [85] προτείνουν έναν αλγόριθμο κατάτμησης χειρόγραφων χαρακτήρων λαμβάνοντας υπόψη κάποια γενικά χαρακτηριστικά που συναντώνται σε χειρόγραφα έγγραφα. Πιο συγκεκριμένα, στην αρχή γίνεται μία πρώτη κατάτμηση της λέξης με γραμμές διαφορετικών κλίσεων κοντά στην κλίση του κειμένου. Τα διαφορετικά αποτελέσματα κατάτμησης ενώνονται και προκύπτει το αποτέλεσμα που φαίνεται στο Σχήμα 2.41(γ). Τέλος, γίνεται η οριστική κατάτμηση των χαρακτήρων με μία συνάρτηση κόστους. Η συνάρτηση κόστους εξαρτάται τόσο από χαρακτηριστικά του κειμένου όπως το πάχος της πέννας γραφής, το πλάτος των χαρακτήρων και το ύψος της γραμμής κειμένου, όσο και από την απόσταση μεταξύ των σημείων κατάτμησης του πρώτου σταδίου και της κλίσης των διαχωριστικών γραμμών.



Σχήμα 2.41: (α)-(β) Κατάτμηση με διαχωριστικές γραμμές διαφορετικών κλίσεων, (γ) αποτέλεσμα που ενσωματώνει όλα τα αποτελέσματα κατάτμησης μαζί με το τελικό αποτελέσματα.

Οι Kavallieratou et al. [86] προτείνουν μία μέθοδο για την κατάτμηση χειρόγραφου κειμένου με κλίση (cursive handwritten text). Αρχικά, το έγγραφο υφίσταται κατάτμηση σε επίπεδο λέξεων και διορθώνεται η κλίση τους. Έπειτα, εντοπίζονται τα υποψήφια σημεία κατάτμησης και το τελικό αποτέλεσμα προκύπτει μέσα από μία διαδικασία μετασχηματισμού βάσει κανόνων μάθησης (transformation-based learning). Τα υποψήφια σημεία κατάτμησης εντοπίζονται στα τοπικά ελάχιστα της κατακόρυφης προβολής λαμβάνοντας παράλληλα υπόψη τυπογραφικούς κανόνες σε σχέση με το πλάτος που πρέπει να έχει κάθε χαρακτήρας (Σχήμα 2.42).



Σχήμα 2.42: Τα στάδια της μεθόδου [86]. (α) Αρχική εικόνα λέξης, (β) το κυρίως τμήμα της λέξης, (γ) το κυρίως τμήμα της λέξης αυξημένο κατά $\frac{1}{4}$ του κύριου ύψους, (δ) η λέξη μετά τον εντοπισμό των υποψήφιων σημείων κατάτμησης, (ε) το τελικό αποτέλεσμα της κατάτμησης.

Στο [87], οι Xiao και Leedham παρουσιάζουν μία μέθοδο κατάτμησης αγγλικών εγγράφων σε χαρακτήρες. Αρχικά, πραγματοποιεί μία ανάλυση συνδεδεμένων συστατικών η οποία τα διαχωρίζει σε κατάλληλα σημεία με βάση την καμπυλότητά τους. Στο Σχήμα 2.43 απεικονίζεται μία λέξη όπου επισημάνονται οι κυρτές, οι κοίλες και οι κλειστές περιοχές που σχηματίζονται γύρω από τους χαρακτήρες. Στη συνέχεια, ενώνει τα τμήματα εκείνα τα οποία, βάσει των χαρακτηριστικών τους, αποτελούν τμήματα του ίδιου χαρακτήρα. Η μέθοδος αυτή έχει το πλεονέκτημα της ανεξαρτησίας από το είδος της γραφής του κειμένου, δηλαδή το μέγεθος των λέξεων, τη κλίση και το πάχος της πένας. Επιπλέον, ο τρόπος κατάτμησης των αντικειμένων αποτρέπει τη δημιουργία τμημάτων χαρακτήρων όπως συμβαίνει σε άλλες μεθόδους.



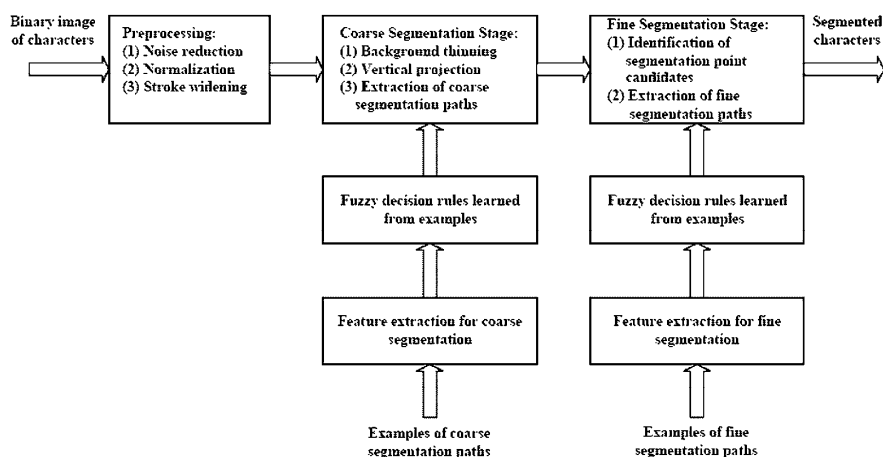
Σχήμα 2.43: Παράδειγμα λέξης όπου επισημάνονται οι κυρτές (D), οι κοίλες (U) και οι κλειστές (L) περιοχές γύρω από τους χαρακτήρες.

Οι Nicchiotti et al. [88] προτείνουν μία μέθοδο για την κατάτμηση χειρόγραφων χαρακτήρων. Αρχικά, εντοπίζουν όλα τα υποψήφια σημεία κατάτμησης με στόχο να βρίσκονται ανάμεσά τους εκείνα τα σημεία που θα οδηγήσουν στη σωστή κατάτμηση των χαρακτήρων (Σχήμα 2.44). Η αρχική κατάτμηση βασίζεται σε διάφορα χαρακτηριστικά όπως το πάχος των χαρακτήρων, τη γραμμή αναφοράς, τα τοπικά ελάχιστα του περιγράμματος και τις κλειστές περιοχές. Στη συνέχεια, απορρίπτονται σημεία κατάτμησης βάσει του αναμενόμενου πλάτους και της θέση των χαρακτήρων που προκύπτουν σε σχέση με τη γραμμή αναφοράς.



Σχήμα 2.44: Κατάτμηση της λέξης σε όλα τα υποψήφια σημεία σύμφωνα με τη μέθοδο [88].

Στην εργασία [89], ο Zhao et al. προτείνουν μία μέθοδο κατάτμησης χειρόγραφων κινέζικων χαρακτήρων η οποία αποτελείται από δύο στάδια. Το διάγραμμα ροής της μεθόδου παρουσιάζεται στο Σχήμα 2.45. Αρχικά, εφαρμόζονται τεχνικές βελτίωσης της ποιότητας όπως αφαίρεση θορύβου, κανονικοποίηση και πάχυνση. Στο πρώτο στάδιο, εφαρμόζεται κατάτμηση της λέξης κατά προσέγγιση σύμφωνα με τον εξωτερικό σκελετό και την κατακόρυφη προβολή. Έπειτα, αξιολογούνται όλα τα υποψήφια μονοπάτια κατάτμησης με τη βοήθεια ασαφών κανόνων απόφασης χρησιμοποιώντας γεωμετρικά χαρακτηριστικά. Συνεπώς, όλες οι λανθασμένες διαδρομές κατάτμησης απορρίπτονται. Στο δεύτερο στάδιο επαναλαμβάνεται η ίδια διαδικασία σε κάθε τμήμα που προέκυψε από το πρώτο στάδιο.



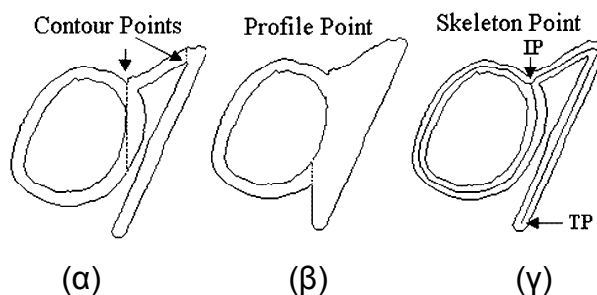
Σχήμα 2.45: Διάγραμμα ροής της μεθόδου [89] για την κατάτμηση χειρόγραφων κινέζικων χαρακτήρων.

Η βασική ιδέα της μεθόδου των Timar et al. [90] είναι ότι μπορούμε να εξαγάγουμε χρήσιμες πληροφορίες για τη διάρθρωση της λέξης από τα χαρακτηριστικά σημεία του σκελετού. Αρχικά, διορθώνουν τη κλίση της λέξης με τη βοήθεια των γραμμών αναφοράς και του ιστογράμματος έτσι ώστε κάθε χαρακτήρας να είναι κάθετα τοποθετημένος μέσα στη λέξη. Αφού έχει εξαχθεί ο σκελετός της λέξης, εντοπίζεται ο σκελετός του υπόβαθρου με τη μέθοδο της πλημμύρας (Σχήμα 2.46). Τέλος, εντοπίζονται οι διαδρομές κατάτμησης.



Σχήμα 2.46: Εξαγωγή σκελετού σύμφωνα με τη μέθοδο [90]: (α) Αρχική λέξη, (β) σκελετός της λέξης (μαύρος) και σκελετός του υπόβαθρου στο πάνω (κόκκινος) και κάτω (πράσινος) τμήμα της λέξης.

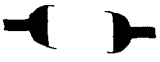




Οι Oliveira et al. [91] προτείνουν μία μέθοδο κατάτμησης χειρόγραφων ψηφίων βασισμένη σε δομικά χαρακτηριστικά. Η μέθοδος χρησιμοποιεί το συνδυασμό δύο τύπων δομικών χαρακτηριστικών έτσι ώστε να εντοπιστεί η καλύτερη διαδρομή κατάτμησης. Αρχικά, εντοπίζονται όλα τα υποψήφια σημεία κατάτμησης και παράγεται ένα σύνολο διαδρομών κατάτμησης βάσει εμπειρικών κανόνων. Ως σημεία κατάτμησης εντοπίζονται τα σημεία του περιγράμματος (contour points), τα σημεία προφίλ (profile points), τα σημεία διακλάδωσης (intersection points) καθώς και τα τερματικά σημεία (terminal points) (Σχήμα 2.47). Τέλος, σύμφωνα με μία διαδικασία βασισμένη στους γράφους, επιλέγονται τα σωστά μονοπάτια κατάτμησης.



Σχήμα 2.47: Σημεία κατάτμησης σύμφωνα με τη μέθοδο [91]: (α) σημεία περιγράμματος, (β) σημεία προφίλ, (γ) σημεία διακλάδωσης και τερματικά σημεία στον σκελετό.

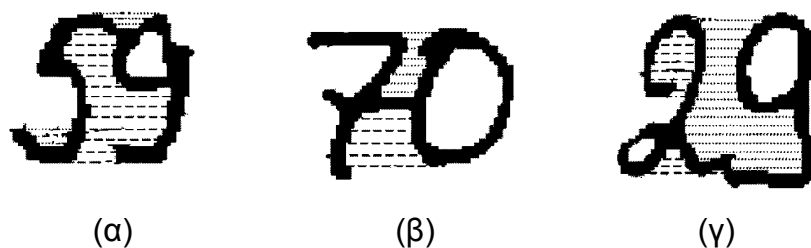
Οι Chen και Wang [92] σε μία παρόμοια μέθοδο προσπαθούν να εντοπίσουν τη βέλτιστη διαδρομή κατάτμησης για την κατάτμηση χειρόγραφων ψηφίων. Η μέθοδος στηρίζεται στην παραδοχή ότι υπάρχουν μόνο δύο ψηφία τα οποία συνδέονται σε ένα ή περισσότερα σημεία. Στο Σχήμα 2.48 παρουσιάζονται όλοι οι τύπου σύνδεσης δύο ψηφίων που μπορεί να διαχειριστεί με επιτυχία η μέθοδος.

Αρχικά, εφαρμόζεται ο αλγόριθμος εκλέπτυνσης στο σύμπλεγμα ψηφίων και στο υπόβαθρο ώστε να προκύψει ο εσωτερικός και εξωτερικός σκελετός, αντίστοιχα. Έπειτα, ο σκελετός κατηγοριοποιείται σε τέσσερις κατηγορίες και εξάγονται κάποια χαρακτηριστικά σημεία. Τέλος, σχηματίζονται όλα τα υποψήφια μονοπάτια που διασχίζουν τα χαρακτηριστικά σημεία και με τη βοήθεια μίας Gaussian συνάρτησης πιθανότητας και γεωμετρικών παραμέτρων επιλέγεται το βέλτιστο μονοπάτι κατάτμησης.

Category	Type	Style of touching	Examples
Single-touching	1		59 33
	2		24 02
	3		23 52
	4		40 00
Multiple-touching	5		78 38

Σχήμα 2.48: Τύποι σύνδεσης δύο ψηφίων που μπορεί να διαχειριστεί με επιτυχία η μέθοδος [92].

Τέλος, στο [93], οι Pal et al. προτείνουν επίσης ένα αλγόριθμο για την κατάτμηση χειρόγραφων ψηφίων ο οποίος βασίζεται στην επεξεργασία των περιοχών υπόβαθρου (reservoir) ανάμεσα στα ενωμένα ψηφία. Οι περιοχές αυτές εντοπίζονται με τη μέθοδο της πλημμύρας από το πάνω και κάτω μέρος της εικόνας. Έπειτα, αποφασίζεται ο τύπος σύνδεσης των ψηφίων σύμφωνα με τη θέση και το μέγεθος των περιοχών. Στο Σχήμα 2.49 εμφανίζονται οι τρεις τύποι σύνδεσης: πάνω, στη μέση και κάτω σύνδεση. Στη συνέχεια, προσδιορίζεται το καλύτερο σημείο κατάτμησης αναλύοντας τα σύνορα των περιοχών, τον τύπο σύνδεσης και διάφορα τοπολογικά χαρακτηριστικά. Τέλος, με τη βοήθεια μορφολογικών χαρακτηριστικών σχεδιάζεται η διαδρομή κατάτμησης.



Σχήμα 2.49: Τύποι σύνδεσης ψηφίων σύμφωνα με τη μέθοδο [93]: (α) πάνω (β) στη μέση και (γ) κάτω σύνδεση. Με διακεκομμένες γραμμές επισημάνονται οι περιοχές υπόβαθρου ανάμεσα στα ψηφία.

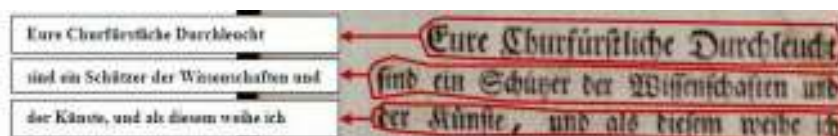
Στον Πίνακα 2.3 παρουσιάζονται συνοπτικά τα βασικά χαρακτηριστικά των μεθόδων που έχουν προταθεί για την κατάτμηση χαρακτήρων.

Πίνακας 2.3: Μέθοδοι Κατάτμησης Χαρακτήρων

	Εφαρμογή	Μέθοδος
Liang et al. 1994 [82]	Τυπωμένους χαρακτήρες	Επαναληπτικός αλγόριθμος κατάτμησης με τη χρήση προβολών και συνάρτησης κόστους.
Chang and Chen 1999 [83]	Τυπωμένους χαρακτήρες	Αλγόριθμος εύρεσης του μικρότερου μονοπατιού χρησιμοποιώντας την πληροφορία από το κυρτό κέλυφος (convex hull).
Antonacopoulos and Karatzas 2005 [73]	Τυπωμένους ιστορικούς χαρακτήρες	Χρησιμοποιεί τις κατακόρυφες προβολές και το αναμενόμενο πλάτος των χαρακτήρων για τον εντοπισμός των σημείων κατάτμησης.
Nomura et al. 2005 [84]	Τυπωμένους χαρακτήρες – πινακίδες	Αλγόριθμος κατάτμησης με βάση τα ιστογράμματα, μία συνάρτηση κόστους, τις αποστάσεις των χαρακτήρων και το αναμενόμενο πλάτος τους.
Yanikoglu and Sandon 1998 [85]	Χειρόγραφους χαρακτήρες	Εντοπίζει τα διαδοχικά σημεία κατάτμησης εκτιμώντας μία συνάρτηση κόστους σε κάθε σημείο κατά μήκος της γραμμής αναφοράς.
Kavallieratou et al. 2000 [86]	Χειρόγραφους χαρακτήρες	Χρησιμοποιεί μία διαδικασία μετασχηματισμού βάσει κανόνων μάθησης για την εξαγωγή κανόνων που προσδιορίζουν τα σημεία κατάτμησης.
Xiao and Leedham 2000 [87]	Χειρόγραφους αγγλικούς χαρακτήρες	Χρησιμοποιεί τη γνώση σχετικά με τη δομή των αγγλικών χαρακτήρων. Διαχωρίζει τις λέξεις σε κατάλληλα σημεία βάσει την καμπυλότητα τους, τα οποία στη συνέχεια ενώνει για να σχηματίσει τους χαρακτήρες.
Nicchiotti et al. 2000 [88]	Χειρόγραφους χαρακτήρες	Κατάτμηση λέξης σε όλα τα υποψήφια σημεία και εύρεση των τελικών σημείων βάσει το αναμενόμενο πλάτος των χαρακτήρων και τη θέση τους σε σχέση με τη γραμμή αναφοράς.
Zhao et al. 2003 [89]	Χειρόγραφους κινέζικους χαρακτήρες	Χρήση ασαφών κανόνων απόφασης, οι οποίοι εκπαιδεύονται από παραδείγματα, για την αξιολόγηση των υποψήφιων μονοπατιών κατάτμησης.
Timar et al. 2003 [90]	Χειρόγραφους χαρακτήρες	Χρησιμοποιεί χαρακτηριστικά σημεία του σκελετού και λογικούς κανόνες για να εντοπίσει τις διαδρομές κατάτμησης.
Oliveira et al. 2000 [91]	Χειρόγραφα ψηφία	Χρησιμοποιείται ο συνδυασμός δύο τύπων δομικών χαρακτηριστικών έτσι ώστε να εξασφαλιστεί η καλύτερη διαδρομή κατάτμησης με τη χρήση γράφων.
Chen and Wang 2000 [92]	Χειρόγραφα ψηφία	Συσχετίζει τα χαρακτηριστικά σημεία του σκελετού ώστε να εντοπίσει όλα τα υποψήφια μονοπάτια κατάτμησης και με τη βοήθεια μίας Gaussian συνάρτησης πιθανότητας εντοπίζει το βέλτιστο.
Pal et al. 2003 [93]	Χειρόγραφα ψηφία	Το καλύτερο σημείο κατάτμησης προσδιορίζεται αναλύοντας τα σύνορα των περιοχών υπόβαθρου ανάμεσα στα ψηφία, τον τύπο σύνδεσης και διάφορα τοπολογικά χαρακτηριστικά.

2.2.5 Αντιστοίχιση Εικόνας και Αρχείου Κειμένου

Πολλές ηλεκτρονικές βιβλιοθήκες παρέχουν πρόσβαση σε μεγάλες βάσεις ψηφιοποιημένων ιστορικών βιβλίων, περιοδικών και εφημερίδων. Σε πολλές περιπτώσεις, πέραν της πρόσβασης στην εικόνα εγγράφου, είναι διαθέσιμο και το αρχείο κειμένου με την ASCII κωδικοποίηση του εγγράφου (transcription). Τα αρχεία κειμένου έχουν προκύψει είτε από την εφαρμογή μίας μεθοδολογίας οπτικής αναγνώρισης των εικόνων και τη μετέπειτα διόρθωση τους από τον χρήστη είτε από την εξολοκλήρου πληκτρολόγηση τους από τον χρήστη. Το γεγονός αυτό έχει οδηγήσει στην ανάπτυξη πολλών μεθόδων που προσπαθούν να αντιστοιχήσουν την πληροφορία των εικόνων με την πληροφορία στα αρχεία κειμένου (transcript mapping) (Σχήμα 2.50). Η αντιστοίχιση αυτή δίνει τη δυνατότητα στους χρήστες να εντοπίζουν εύκολα τη θέση μίας πρότασης, μίας λέξης ή ακόμα και ενός χαρακτήρα στην εικόνα του εγγράφου ενώ διαβάζουν το αρχείο κειμένου. Για παράδειγμα, θα μπορεί ο χρήστης να βλέπει συγχρόνως την εικόνα και το αρχείο κειμένου και όταν επιλέγει μία λέξη στην εικόνα να μαρκάρετε αυτόματα η αντίστοιχη λέξη στο αρχείο κειμένου ή το αντίστροφο (Σχήμα 2.51).



(α)



(β)



(γ)

Σχήμα 2.50: Παράδειγμα αντιστοίχισης της εικόνας με το αρχείο κειμένου σε επίπεδο (α) γραμμής κειμένου, (β) λέξης και (γ) χαρακτήρα.



Σχήμα 2.51: Παράδειγμα εφαρμογής η οποία δίνει τη δυνατότητα στον χρήστη να εντοπίζει εύκολα τη θέση μίας λέξης στην εικόνα του εγγράφου ενώ διαβάζει το αρχείο κειμένου.

Μία αυτοματοποιημένη τεχνική αντιστοίχισης των εικόνων με τα αρχεία κειμένου έχει να αντιμετωπίσει πολλές δυσκολίες. Θα πρέπει πρώτα να εντοπίσει τις περιοχές κειμένου, τις γραμμές κειμένου, τις λέξεις ή/και τους χαρακτήρες, στην εικόνα και έπειτα να προχωρήσει στην αντιστοίχιση με το αρχείο κειμένου. Εάν δεν υπάρχουν λάθη στο στάδιο της κατάτμησης ή/και στο αρχείο κειμένου τότε μία απλή αντιστοίχιση ένα προς ένα είναι επαρκής. Για παράδειγμα, θα αντιστοιχούσε την πρώτη λέξη στην εικόνα με την πρώτη λέξη στο αρχείο κειμένου, τη δεύτερη με τη δεύτερη και ου το καθ' εξής. Εντούτοις, πιθανά λάθη στη κατάτμηση των περιοχών κειμένου ή στο αρχείο κειμένου δεν επιτρέπουν σε μια αντιστοίχιση ένα προς ένα να έχει επιτυχή αποτελέσματα. Επίσης, οι λέξεις που σπάνε στο τέλος της πρότασης και συνεχίζουν στην επόμενη γραμμή κειμένου είναι άλλη μία αιτία για πιθανά λάθη κατά τη διαδικασία της αντιστοίχισης. Γι' αυτό τον σκοπό έχουν προταθεί αρκετές μέθοδοι αντιστοίχισης [94-103] οι βασικότερες των οποίων παρουσιάζονται στη συνέχεια.

Οι μέθοδοι αντιστοίχισης των εικόνων με τα αρχεία κειμένου βασίζονται κυρίως είτε στα Hidden Markov Models (HMMs) [94-96] είτε σε τεχνικές Dynamic Time Warping (DTW) [97-99]. Επίσης, η μεγάλη πλειοψηφία των μεθόδων επικεντρώνεται στην αντιστοίχιση χειρόγραφων ιστορικών εγγράφων με τα αντίστοιχα αρχεία κειμένου σε επίπεδο λέξης.

Οι Zimmermann και Bunke [94] προτείνουν μία αυτοματοποιημένη μέθοδο κατάτμησης του χειρόγραφου εγγράφου σε λέξεις χρησιμοποιώντας πληροφορία από το αρχείο κειμένου. Η τεχνική κατάτμησης βασίζεται στα HMMs και αποτελείται από δύο στάδια. Στο πρώτο στάδιο, χρησιμοποιείται ο

αποκωδικοποιητής Viterbi ο οποίος δέχεται μία ολόκληρη γραμμή κειμένου από το αρχείο κειμένου και παράγει τη βέλτιστη κατάτμηση σε λέξεις. Η διαδικασία αυτή οδηγεί σε ικανοποιητικά αποτελέσματα αρκεί να προηγηθεί η κατάλληλη εκπαίδευση των HMMs καθώς επίσης να έχουν κανονικοποιηθεί οι γραμμές κειμένου της εικόνας για να μειωθεί έτσι η διακύμανση των διαφορετικών ειδών γραφής. Στο δεύτερο στάδιο, γίνεται η αντιστοίχιση του αποτελέσματος της κατάτμησης της κανονικοποιημένης εικόνας με την πραγματική εικόνα χρησιμοποιώντας μία ενδιάμεση αναπαράσταση της εικόνας. Η μέθοδος εφαρμόστηκε σε έγγραφα της βάσης IAM [104]. Το Σχήμα 2.52 παρουσιάζει ένα αποτέλεσμα αντιστοίχισης λέξεων για μία γραμμή κειμένου.



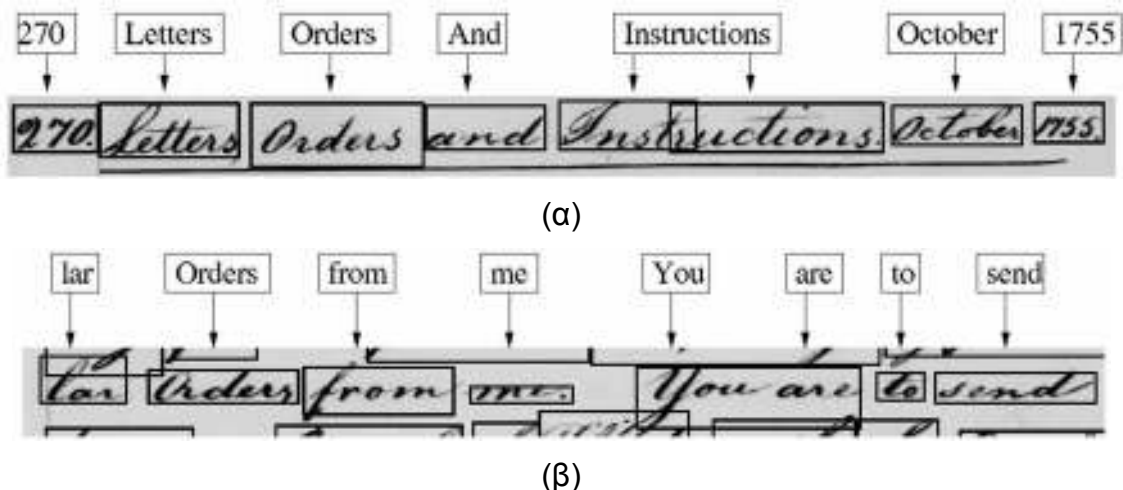
Σχήμα 2.52: Αποτέλεσμα αντιστοίχισης λέξεων για μία γραμμή κειμένου σύμφωνα με τη μέθοδο [94].

Με παρόμοιο τρόπο, στην εργασία [95], οι Toselli et al. χρησιμοποιούν τον αποκωδικοποιητή Viterbi των HMMs για την αντιστοίχιση λέξεων σε χειρόγραφα ιστορικά έγγραφα. Η μέθοδος αποτελείται από τέσσερα στάδια. Στο πρώτο στάδιο λαμβάνει χώρα η προ-επεξεργασία του εγγράφου και περιλαμβάνει: τη διόρθωση κλίσης, την αφαίρεση θορύβου και τέλος την κατάτμηση και την κανονικοποίηση των γραμμών κειμένου. Στο δεύτερο στάδιο, κάθε γραμμή κειμένου αναπαριστάται με ένα διάνυσμα χαρακτηριστικών και έπειτα ακολουθεί το στάδιο εκπαίδευσης των HMMs. Στο τελευταίο στάδιο, ολοκληρώνεται η αντιστοίχιση ανά γραμμή κειμένου χρησιμοποιώντας τον αποκωδικοποιητή Viterbi (Σχήμα 2.53). Για την αξιολόγηση της μεθόδου χρησιμοποιήθηκαν εικόνες από ένα ιστορικό βιβλίο της ηλεκτρονικής βιβλιοθήκης της Βαλένθια (BIVALDI) [105].



Σχήμα 2.53: Αποτέλεσμα αντιστοίχισης λέξεων σύμφωνα με τη μέθοδο [95].

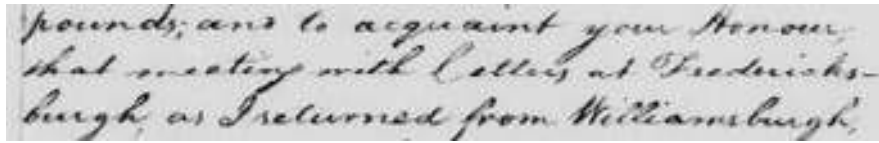
Στο [96], οι Rothfeder et al. βασίζονται επίσης στα HMMs για την αντιστοίχιση λέξεων σε χειρόγραφα ιστορικά έγγραφα χωρίς όμως να πραγματοποιούν αναγνώριση κάθε λέξης της εικόνας του εγγράφου. Αρχικά, εντοπίζονται όλες οι λέξεις του εγγράφου χρησιμοποιώντας μία τεχνική κατάτμησης βασισμένη στην ταξινόμηση αποστάσεων. Ο στόχος της μεθόδου είναι να αντιστοιχήσει μία ή περισσότερες λέξεις του αρχείου κειμένου σε κάθε λέξη που εντοπίστηκε αυτόματα στο στάδιο της κατάτμησης. Στην περίπτωση που κάποια λέξη της εικόνας έχει σπάσει σε δύο λέξεις (oversegmentation) θα πρέπει η αντίστοιχη λέξη του αρχείου κειμένου να αντιστοιχιστεί και με τις δύο λέξεις (Σχήμα 2.54(α)). Αντιθέτως, αν δύο λέξεις της εικόνας έχουν συνενωθεί (undersegmentation) θα πρέπει και οι δύο λέξεις του αρχείου κειμένου να αντιστοιχηθούν με την ίδια λέξη (Σχήμα 2.54(β)). Η μέθοδος εφαρμόστηκε σε χειρόγραφα ιστορικά κείμενα του George Washington [106].



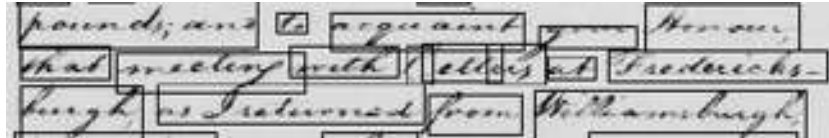
Σχήμα 2.54: Αντιστοίχιση λέξεων σύμφωνα με τη μέθοδο [96] στην περίπτωση (α) της διάσπασης μια λέξης και (β) της συνένωσης δύο λέξεων.

Η δεύτερη κύρια κατηγορία περιλαμβάνει τις μεθόδους που βασίζονται σε τεχνικές Dynamic Time Warping (DTW) οι οποίες στοιχίζουν δύο χρονικές σειρές ελαχιστοποιώντας την απόσταση μεταξύ τους. Οι Kornfield et al. [97] βασίζονται στην τεχνική DTW χρησιμοποιώντας ως σειρές το αποτέλεσμα της κατάτμησης του εγγράφου σε λέξεις και τις λέξεις του αρχείου κειμένου. Ως χαρακτηριστικά των λέξεων, τόσο της εικόνας όσο και του αρχείου κειμένου, χρησιμοποιούνται ο λόγος διαστάσεων, το πλάτος, η θέση των χαρακτήρων καθώς και αριθμός των ascenders και descenders των χαρακτήρων. Το στάδιο της κατάτμησης είναι αρκετά δύσκολο, ειδικά στα ιστορικά χειρόγραφα έγγραφα, και συνεπώς γίνονται πολλά λάθη που επηρεάζουν τη μετέπειτα διαδικασία. Γι' αυτό προτείνουν τη

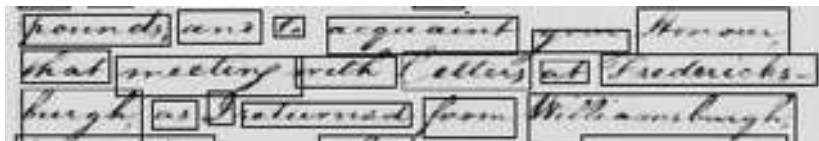
διόρθωση του αποτελέσματος της κατάτμησης από τον χρήστη ώστε να μην επηρεάσει την αντιστοίχιση (Σχήμα 2.55).



(α)



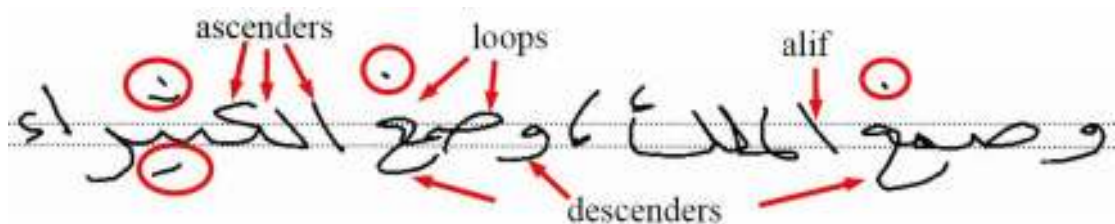
(β)



(γ)

Σχήμα 2.55: Κατάτμηση λέξεων σύμφωνα με τη μέθοδο [97], (α) αρχικό έγγραφο, (β) αποτέλεσμα αυτόματης κατάτμησης και (γ) αποτέλεσμα μετά τη διόρθωση από τον χρήστη.

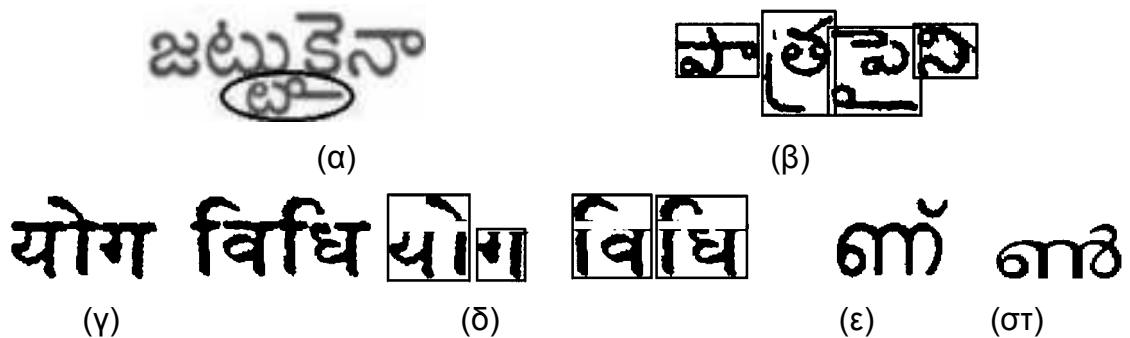
Οι Lorigo και Govindaraju [98] προτείνουν μία μέθοδο αντιστοίχισης εικόνας και αρχείου κειμένου για χειρόγραφα αραβικά έγγραφα. Λαμβάνουν υπόψη τους τις ιδιαιτερότητες της αραβικής γραφής και χρησιμοποιούν ως μονάδες τμήματα λέξεων και όχι λέξεις καθώς τα κενά διαστήματα ανάμεσα στους χαρακτήρες μπορεί να έχουν το ίδιο πλάτος με εκείνα ανάμεσα στις λέξεις. Προτείνουν μία παραλλαγή του κλασικού DTW, τον True Distance DTW, ο οποίος χρησιμοποιεί δύο αποστάσεις για την αντιστοίχιση πολλαπλών μονάδων από τη μία σειρά σε μία μόνο μονάδα στη δεύτερη σειρά. Χρησιμοποιούν έξι τύπους χαρακτηριστικών, τον αριθμό των βρόχων, των ascenders και descenders των χαρακτήρων, των σημαδιών πάνω και κάτω από τη γραμμή αναφοράς και τέλος ένα δυαδικό χαρακτηριστικό που δείχνει αν η λέξη είναι ο ειδικός χαρακτήρας “alif” (Σχήμα 2.56).



Σχήμα 2.56: Χαρακτηριστικά της μεθόδου [98].

Η μέθοδος των Jawahar και Kumar [99] δημιουργεί αντιστοίχιση σε επίπεδο γραμμής, λέξης και χαρακτήρα και εφαρμόζεται σε τυπωμένα Ινδικά έγγραφα.

Προτείνουν μία ιεραρχική προσέγγιση η οποία βασίζεται στη τεχνική του DTW και σε πολλές ημιαυτόματες διαδικασίες. Η μέθοδος προσπαθεί να αντιμετωπίσει τις δυσκολίες που εμφανίζονται γενικά στα τυπωμένα έγγραφα, όπως η συνένωση και η επικάλυψη χαρακτήρων αλλά και συγκεκριμένα προβλήματα που παρουσιάζονται στα Ινδικά έγγραφα, όπως η ύπαρξη μία γραμμής αναφοράς (Shirorekha) σε συγκεκριμένες διαλέκτους και η έλλειψη κωδικοποίησης κάποιων χαρακτήρων (Σχήμα 2.57).



Σχήμα 2.57: Προβλήματα που παρουσιάζουν τα τυπωμένα Ινδικά έγγραφα: (α) συνένωση και (β) επικάλυψη χαρακτήρων, (γ) ύπαρξη γραμμής αναφοράς (Shirorekha) σε συγκεκριμένες διαλέκτους και (δ) δημιουργία σπασμένων χαρακτήρων μετά την αφαίρεση της, (ε)-(στ) έλλειψη κωδικοποίησης χαρακτήρων.

Οι Tomai et al. [100] προτείνουν μία μέθοδο για να μειώσουν το μέγεθος του λεξικού που χρησιμοποιείται κατά τη διαδικασία αναγνώρισης των λέξεων με τη βοήθεια των αρχείων κειμένων. Δημιουργούν πολλές διαφορετικές υποθέσεις κατάτμησης μίας γραμμής κειμένου σε λέξεις και στη συνέχεια, για κάθε λέξη παράγεται μία ταξινομημένη λίστα από υποψήφιες λέξεις σύμφωνα με το λεξικό. Έπειτα, κάθε λέξη αντιστοιχίζεται με μία μόνο λέξη του λεξικού με τη βοήθεια δυναμικού προγραμματισμού. Εάν η αντιστοίχιση δεν πραγματοποιηθεί με ένα ικανοποιητικό βαθμό εμπιστοσύνης τότε η λέξη απορρίπτεται.

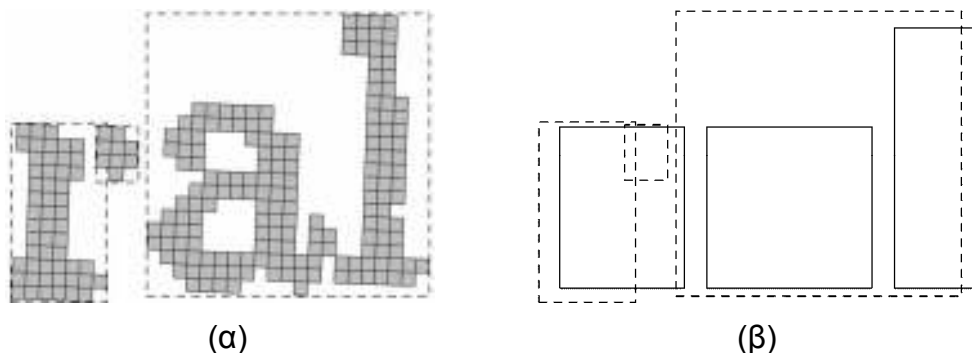
Στην εργασία [101], οι Huang και Srihari προτείνουν μία μέθοδο αντιστοίχισης για χειρόγραφα έγγραφα που βασίζεται στην αναγνώριση των λέξεων. Χρησιμοποιείται μία μηχανή αναγνώρισης λέξεων (WMR) [107] η οποία δέχεται ως είσοδο μία εικόνα λέξης και ένα λεξικό και παράγει τα πιθανά αποτελέσματα αναγνώρισης με το πρώτο αποτέλεσμα να είναι το βέλτιστο. Στη συνέχεια, με τη βοήθεια δυναμικού προγραμματισμού γίνεται η αντιστοίχιση της λέξης που παρήγαγε η μηχανή αναγνώρισης και της λέξης που βρίσκεται στο αρχείο κειμένου. Για να διορθώσουν πιθανά λάθη που γίνονται στη διαδικασία των λέξεων εφαρμόζουν ένα στάδιο μετεπεξεργασίας. Η μέθοδος δοκιμάστηκε σε έγγραφα της βάσης CEDAR [108].

Οι Zinger et al. [102] αρχικά εντοπίζουν αυτόματα τις γραμμές κειμένου σε ιστορικά έγγραφα χρησιμοποιώντας οπτικά χαρακτηριστικά και εισάγουν χειροκίνητα την αντίστοιχη κωδικοποίηση κάθε γραμμής κειμένου. Στο Σχήμα 2.58 απεικονίζεται η παραπάνω εφαρμογή. Κάθε γραμμή κειμένου χωρίζεται σε τόσες περιοχές όσες είναι οι λέξεις με βάση τα μεγαλύτερα κενά διαστήματα και η αντιστοίχιση υλοποιείται με τη βοήθεια μίας συνάρτησης κόστους. Τα χαρακτηριστικά που χρησιμοποιούνται βασίζονται στο μήκος της λέξης τόσο στην εικόνα του εγγράφου όσο και στο αρχείο κειμένου.



Σχήμα 2.58: Εφαρμογή της μεθόδου [102] για τη χειροκίνητη εισαγωγή της κωδικοποίησης των γραμμών κειμένου.

Τέλος, μία διαφορετική τεχνική παρουσιάζεται από τον Hobby [103] ο οποίος προσπαθεί να αντιστοιχήσει μία εικόνα τυπωμένου εγγράφου με το αντίστοιχο αρχείο κειμένου το οποίο όμως έχει την ίδια μορφοποίηση με την εικόνα. Πιο συγκεκριμένα, προσπαθεί να αντιστοιχήσει τους χαρακτήρες μίας εικόνας με τους χαρακτήρες του ηλεκτρονικού εγγράφου από το οποίο προήλθε η εικόνα. Εξάγει χαρακτηριστικά από τα πλαίσια των συνδεδεμένων συστατικών της εικόνας (Σχήμα 2.59(α)) και των χαρακτήρων στο αρχείο κειμένου (Σχήμα 2.59(β)) και με τη χρήση γεωμετρικού μετασχηματισμού πραγματοποιεί την αντιστοίχιση.



Σχήμα 2.59: (α) Τα πλαίσια των συνδεδεμένων συστατικών της εικόνας (διακεκομμένη γραμμή) και (β) τα πλαίσια των χαρακτήρων στο αρχείο κειμένου (συνεχόμενη γραμμή) από τα οποία εξάγει τα χαρακτηριστικά η μέθοδος [103].

Στον Πίνακα 2.4 παρουσιάζονται συνοπτικά τα βασικά χαρακτηριστικά των παραπάνω μεθόδων, όπως η τεχνική αντιστοίχισης, το επίπεδο αντιστοίχισης και ο τύπος των εγγράφων που εφαρμόζονται.

Πίνακας 2.4: Μέθοδοι αντιστοίχισης εικόνων με αρχεία κειμένου.

	Τεχνική αντιστοίχισης	Επίπεδο αντιστοίχισης	Εφαρμογή
Zimmermann and Bunke 2002 [94]	HMM	Λέξη	Χειρόγραφα Έγγραφα
Toselli et al. 2007 [95]	HMM	Λέξη	Χειρόγραφα Ιστορικά Έγγραφα
Rothfeder et al. 2006 [96]	HMM	Λέξη	Χειρόγραφα Ιστορικά Έγγραφα
Kornfield et al. 2004 [97]	DTW	Λέξη	Χειρόγραφα Ιστορικά Έγγραφα
Lorigo and Govindaraju 2007 [98]	DTW	Λέξη	Χειρόγραφα Αραβικά Έγγραφα
Jawahar and Kumar 2007 [99]	DTW	Γραμμή κειμένου, λέξη & χαρακτήρα	Τυπωμένα Ινδικά Έγγραφα
Tomai et al. 2002 [100]	Δυναμικό Προγραμματισμό	Λέξη	Χειρόγραφα Ιστορικά Έγγραφα
Huang and Srihari 2006 [101]	Δυναμικό Προγραμματισμό	Λέξη	Χειρόγραφα Έγγραφα
Zinger et al. 2009 [102]	Συνάρτηση κόστους	Λέξη	Χειρόγραφα Ιστορικά Έγγραφα
Hobby 1998 [103]	Γεωμετρικό μετασχηματισμό	Χαρακτήρα	Τυπωμένα Έγγραφα

3. ΑΦΑΙΡΕΣΗ ΠΕΡΙΘΩΡΙΟΥ ΚΑΙ ΔΙΑΧΩΡΙΣΜΟΣ ΔΙΠΛΩΝ ΣΕΛΙΔΩΝ

3.1 Εισαγωγή

Το περιθώριο στα όρια της εικόνας του εγγράφου, όπως αναφέραμε και στην παράγραφο 2.1.3, δημιουργεί πολλά προβλήματα στη μετέπειτα επεξεργασία του, συνεπώς είναι σημαντικό να εντοπιστεί και να αφαιρεθεί. Είναι αρκετά συνηθισμένο, κατά την ψηφιοποίηση βιβλίων ή περιοδικών, μαζί με το περιθώριο να εμφανίζονται και τμήματα γειτονικών σελίδων που επίσης δημιουργούν προβλήματα και πρέπει να αφαιρεθούν. Επιπλέον, μία συνηθισμένη πρακτική είναι η ψηφιοποίηση δύο σελίδων ταυτόχρονα με σκοπό την επιτάχυνση της διαδικασίας ψηφιοποίησης. Παρόλα αυτά, η πλειοψηφία των τεχνικών επεξεργασίας εγγράφων θεωρεί ότι επεξεργάζεται μία σελίδα, συνεπώς, για να είναι δυνατή η μετέπειτα επεξεργασία τέτοιων εγγράφων θα πρέπει να διαχωριστούν σε δύο ξεχωριστές εικόνες.

Σε αυτό το κεφάλαιο παρουσιάζονται δύο νέες μέθοδοι για τον εντοπισμό και την αφαίρεση του περιθωρίου. Η πρώτη μέθοδος, η οποία έχει δημοσιευθεί στο [109], βασίζεται στις οριζόντιες και κατακόρυφες προβολές, στην ανάλυση των συνδεδεμένων συστατικών της εικόνας και στη συνάρτηση διασυσχέτισης του σήματος. Στόχος μας είναι ο εντοπισμός και η αφαίρεση τόσο του περιθωρίου όσο και τμημάτων κειμένου γειτονικών σελίδων. Η δεύτερη μέθοδος, η οποία έχει δημοσιευθεί στο [110], εφαρμόζεται σε έγγραφα που περιέχουν δύο σελίδες συνεπώς ο στόχος είναι, εκτός από την αφαίρεση του περιθωρίου, και ο διαχωρισμός του εγγράφου σε δύο ξεχωριστές σελίδες. Η μέθοδος στηρίζεται σε μία παραλλαγή των κλασικών προβολών, η οποία δίνει έμφαση στα μεγάλα συνεχόμενα τμήματα του υπόβαθρου της εικόνας. Τα πειραματικά αποτελέσματα πάνω σε πληθώρα διαφορετικών ιστορικών βιβλίων που έχουμε στη διάθεσή μας τόσο από το έργο “ΠΟΛΥΤΙΜΟ” [111] όσο και από το έργο “IMPACT” [112] καταδεικνύουν την αποτελεσματικότητα των μεθόδων που αναπτύχθηκαν.

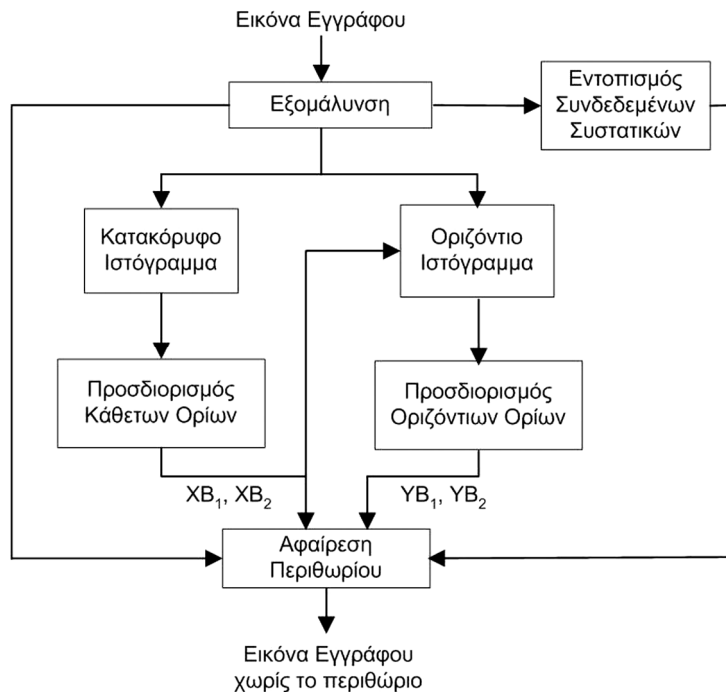
3.2 Εντοπισμός και Αφαίρεση Περιθωρίου καθώς και Τμημάτων Γειτονικής Σελίδας

Η πρώτη προτεινόμενη μέθοδος αποτελείται από δύο φάσεις. Αρχικά, εντοπίζει και αφαιρεί το περιθώριο και στη συνέχεια προσπαθεί να αφαιρέσει, εάν υπάρχουν, τμήματα κειμένου γειτονικής σελίδας. Η μέθοδος βασίζεται στην ανάλυση των προβολών και των συνδεδεμένων συστατικών της εικόνας για την

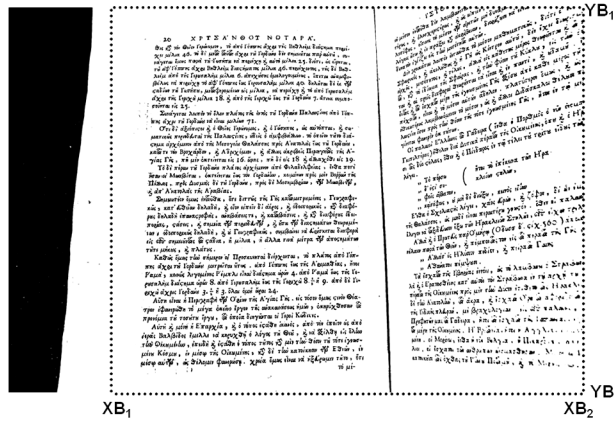
αφαίρεση του περιθωρίου καθώς και στη συνάρτηση διασυσχέτισης του σήματος (signal cross-correlation) για την αφαίρεση τμημάτων κειμένου γειτονικής σελίδας. Εάν η εικόνα είναι διαβάθμισης γκρι προηγείται η δυαδική μετατροπή. Η τεχνική δυαδικής μετατροπής που χρησιμοποιήσαμε περιγράφεται στο [6] και αποτελείται από πέντε στάδια: 1) εφαρμογή ενός χαμηλοπερατού φίλτρου Wiener, 2) μία πρώτη εκτίμηση των *rixels* κειμένου με βάση το [4], 3) τον υπολογισμό του υπόβαθρου, 4) κατωφλίωση που συνδυάζει το υπόβαθρο με την αρχική εικόνα και τέλος 5) ένα στάδιο μετεπεξεργασίας που βελτιώνει την ποιότητα της εικόνας.

i) Εντοπισμός και Αφαίρεση Περιθωρίου

Το διάγραμμα ροής του πρώτου σταδίου της μεθόδου δίνεται στο Σχήμα 3.1. Στόχος μας είναι να υπολογίσουμε τα όρια XB_1 , XB_2 , YB_1 και YB_2 του περιθωρίου όπως παρουσιάζονται στο Σχήμα 3.2. Για να το πετύχουμε αυτό, εφαρμόζουμε μία εξομάλυνση στην εικόνα και έπειτα υπολογίζουμε τα όρια των περιοχών του περιθωρίου και του κειμένου βάσει των ιστογραμμάτων. Τέλος, αφαιρούμε το περιθώριο χρησιμοποιώντας τα συνδεδεμένα συστατικά της εικόνας. Τα κυριότερα βήματα της προτεινόμενης μεθόδου περιγράφονται αναλυτικά στη συνέχεια.



Σχήμα 3.1: Διάγραμμα ροής για τον εντοπισμό και την αφαίρεση του περιθωρίου.



Σχήμα 3.2: Τα όρια του περιθωρίου XB_1 , XB_2 , YB_1 και YB_2 .

Εξομάλυνση: Εφαρμόζουμε οριζόντια και κάθετη εξομάλυνση της εικόνας με τον αλγόριθμο RLSA [27]. Η εικόνα διατρέχεται οριζόντια και κάθετα και τα διαδοχικά pixels υπόβαθρου με μήκος μικρότερο από ένα κατώφλι μετατρέπονται σε pixels κειμένου. Στόχος μας είναι οι περιοχές κειμένου και περιθωρίου να γίνουν πιο συμπαγείς και συνεπώς ευκολότερα διαχωρίσιμες.

Εύρεση Συνδεδεμένων Συστατικών: Υπολογίζουμε τα συνδεδεμένα συστατικά της εικόνας που προέκυψε μετά την εξομάλυνση σύμφωνα με την προσέγγιση που περιγράφεται στο [113]. Εάν η εικόνα αποτελείται από CS συνδεδεμένα συστατικά C_i η τελική εικόνα $I_l(x, y)$ δίνεται από τον τύπο:

$$I_l = \begin{cases} i & \text{εάν } (x, y) \in C_i, 0 < i \leq CS \\ 0 & \text{διαφορετικά} \end{cases}$$

Κατακόρυφο Ιστογράμμο: Υπολογίζουμε το κατακόρυφο ιστογράμμο H_y το οποίο προκύπτει από το άθροισμα των pixels κειμένου σε κάθε στήλη της εικόνας:

$$H_y(x) = \sum_{y=0}^{I_y-1} I(x, y) \quad , \quad 0 \leq x < I_x$$

όπου $I(x, y)$ η εικόνα που προέκυψε μετά την εξομάλυνση και I_x , I_y το πλάτος και το ύψος της εικόνας, αντίστοιχα.

Προσδιορισμός Κάθετων Ορίων: Προσδιορίζουμε τα κάθετα όρια του περιθωρίου XB_1 και XB_2 (Σχήμα 3.2). Αρχικά προσδιορίζουμε το αριστερό κάθετο όριο XB_1 εντοπίζοντας την αρχή (x_0) και το τέλος (x_1) του περιθωρίου με τη βοήθεια του κατακόρυφου ιστογράμματος (Σχήμα 3.3). Το x_0 υπολογίζεται ως εξής:

$$x_0 = \min(x) : H_v(x) > L_1 \text{ ή } H_v(x) < L_2, \quad 0 \leq x < I_x / 5$$

Η πρώτη συνθήκη ($H_v(x) > L_1$) ικανοποιείται όταν το περιθώριο ξεκινάει απευθείας από το αριστερό όριο της εικόνας, κάτι το οποίο είναι και το πιο συνηθισμένο (Σχήμα 3.2). Η δεύτερη συνθήκη ($H_v(x) < L_2$) ικανοποιείται όταν υπάρχει μία κενή περιοχή πριν το περιθώριο. Στην περίπτωση που δε βρεθεί x_0 που να ικανοποιεί αυτές τις συνθήκες θέτουμε $x_0 = -1$ και $x_1 = -1$. Διαφορετικά, υπολογίζουμε το x_1 ως εξής:

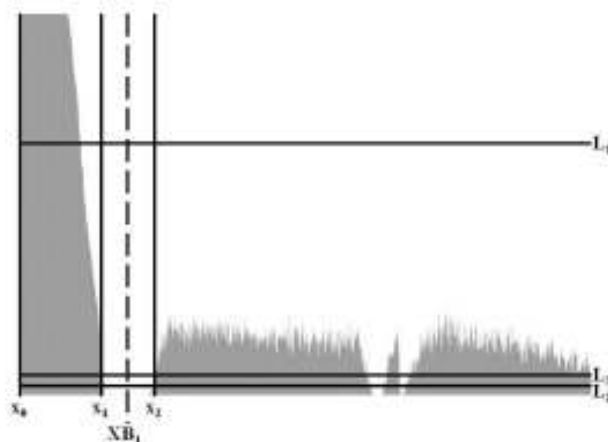
$$x_1 = \begin{cases} \min(x) : H_v(x) < L_2, & \text{εάν } H_v(x_0) > L_1 \\ \min(x) : H_v(x) > L_1, & \text{διαφορετικά} \end{cases}, \quad x_0 < x < I_x / 2$$

Στην περίπτωση που δε βρεθεί x_1 που να ικανοποιεί αυτές τις συνθήκες θέτουμε $x_0 = -1$ και $x_1 = -1$. Αφού έχουμε προσδιορίσει το όρια του περιθωρίου υπολογίζουμε την αρχή (x_2) της περιοχής κειμένου (Σχήμα 3.3) ως εξής:

$$x_2 = \min(x) : H_v(x) < L_1 \text{ και } H_v(x) > L_3, \quad x_1 < x < I_x / 2$$

Εάν δεν υπάρχει x_2 που να ικανοποιεί την παραπάνω συνθήκη θέτουμε $x_2 = -1$. Τέλος, υπολογίζουμε το αριστερό κάθετο όριο XB_1 ως εξής:

$$XB_1 = \begin{cases} 0 & \text{εάν } x_0 = -1 \\ x_0 + (x_1 - x_0) / 2 & \text{εάν } x_2 = -1 \\ x_1 + (x_2 - x_1) / 2 & \text{εάν } x_2 \neq -1 \end{cases}$$



Σχήμα 3.3: Το κατακόρυφο ιστόγραμμα της εικόνας του Σχήματος 3.2 και ο προσδιορισμός του αριστερού κάθετου ορίου XB_1 .

Αντίστοιχα, προσδιορίζεται και το δεξιό κάθετο όριο XB_2 του περιθωρίου. Πειραματικές τιμές των L_1 , L_2 και L_3 είναι οι εξής: $L_1 = (2/3) * I_y$, $L_2 = (1/50) * I_y$ και $L_3 = (1/20) * I_y$.

Αφού έχουν εντοπιστεί τα κάθετα όρια του περιθωρίου, εφαρμόζουμε μία παρόμοια διαδικασία για τον προσδιορισμό των οριζόντιων ορίων YB_1 , YB_2 (Σχήμα 3.2). Η βασική διαφοροποίηση έγκειται στο γεγονός ότι χρησιμοποιούμε το οριζόντιο ιστόγραμμα της περιοχής που οριοθετείται από τα κάθετα όρια του περιθωρίου XB_1 και XB_2 και όχι ολόκληρης της εικόνας.

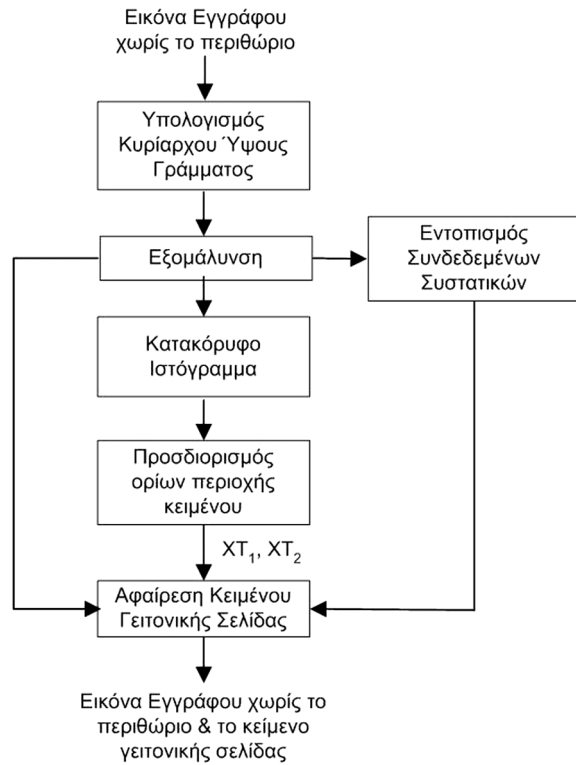
Αφαίρεση περιθωρίου: Έχοντας προσδιορίσει τα όρια XB_1 , XB_2 , YB_1 και YB_2 αφαιρούμε το περιθώριο και προκύπτει η τελική εικόνα $I_c(x, y)$ ως εξής:

$$I_c(x, y) = \begin{cases} 0 & \text{εάν } I_i(x, y) = i \text{ και} \\ & \exists (x_1, y_1) : I_i(x_1, y_1) = i \text{ και } (x_1 \leq XB_1 \text{ ή } x_1 \geq XB_2 \text{ ή } y_1 \leq YB_1 \text{ ή } y_1 \geq YB_2) \\ I(x, y) & \text{διαφορετικά} \end{cases}$$

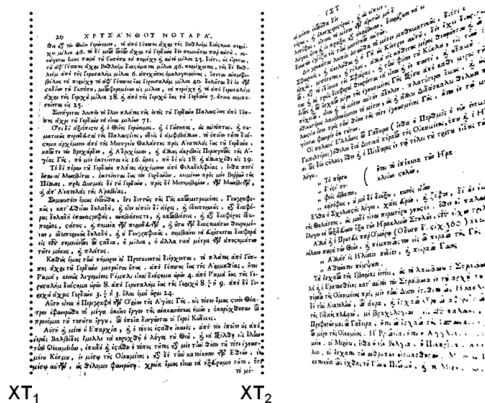
Δηλαδή, όλα τα pixels κειμένου που ανήκουν σε ένα συνδεδεμένο συστατικό C_i το οποίο έχει τουλάχιστον ένα pixel κειμένου εκτός των ορίων μετατρέπονται σε pixels υπόβαθρου. Μετά την αφαίρεση του περιθωρίου προχωράμε στον εντοπισμό και την αφαίρεση τμημάτων κειμένου γειτονικής σελίδας.

ii) Εντοπισμός και Αφαίρεση Τμημάτων Κειμένου Γειτονικής Σελίδας

Στο επόμενο στάδιο εντοπίζουμε και αφαιρούμε το κείμενο της γειτονικής σελίδας από την εικόνα που προέκυψε στο προηγούμενο στάδιο. Το διάγραμμα ροής του δεύτερου σταδίου δίνεται στο Σχήμα 3.4. Στόχος μας είναι να υπολογίσουμε τα όρια της περιοχής κειμένου XT_1 και XT_2 όπως φαίνονται στο Σχήμα 3.5. Αρχικά, εφαρμόζουμε μία εξομάλυνση στην εικόνα με σκοπό να ενώσουμε όλα τα pixels κειμένου που ανήκουν στην ίδια γραμμή κειμένου. Έπειτα, υπολογίζουμε το κατακόρυφο ιστόγραμμα της εικόνας ώστε να εντοπίσουμε τις ζώνες κειμένου. Τέλος, εντοπίζουμε το κείμενο της διπλανής σελίδας με τη βοήθεια της συνάρτησης διασυσχέτισης του σήματος. Τα κυριότερα βήματα της μεθόδου περιγράφονται στη συνέχεια.



Σχήμα 3.4: Διάγραμμα ροής για τον εντοπισμό και την αφαίρεση τμημάτων κειμένου γειτονικής σελίδας.



Σχήμα 3.5: Τα όρια της περιοχής κειμένου XT_1 και XT_2 .

Υπολογισμός Κυρίαρχου Ύψους Γράμματος: Το κυρίαρχο ύψος γράμματος υπολογίζεται σύμφωνα με τη μέθοδο που περιγράφεται στο [80]. Φτιάχνουμε το ιστογράμμα των υψών όλων των πλαισίων που περικλείουν τα συνδεδεμένα συστατικά της εικόνας και το κυρίαρχο ύψος γράμματος εκφράζεται από τη μέγιστη τιμή του ιστογράμματος.

Εξομάλυνση: Εφαρμόζουμε οριζόντια και κάθετη εξομάλυνση της εικόνας με τον αλγόριθμο RLSA [27]. Οι παράμετροι του αλγορίθμου εκφράζονται συναρτήσει του κυρίαρχου ύψους γράμματος. Στόχος μας είναι να ενώσουμε όλα τα pixels κειμένου που βρίσκονται στην ίδια γραμμή κειμένου. Η εμπειρική τιμή για το

κατώφλι στην οριζόντια εξομάλυνση είναι ίση με το κυρίαρχο ύψος γράμματος, ενώ στην κάθετη εξομάλυνση με το μισό του κυρίαρχου ύψους γράμματος.

Προσδιορισμός ορίων περιοχής κειμένου: Υπολογίζουμε τις ζώνες κειμένου που ικανοποιούν την παρακάτω συνθήκη:

$$\exists x_1, x_2 : H_v(x) > L_2 \quad \forall x \in [x_1, x_2] \text{ και } (x_2 - x_1) \geq (XB_2 - XB_1) / 3$$

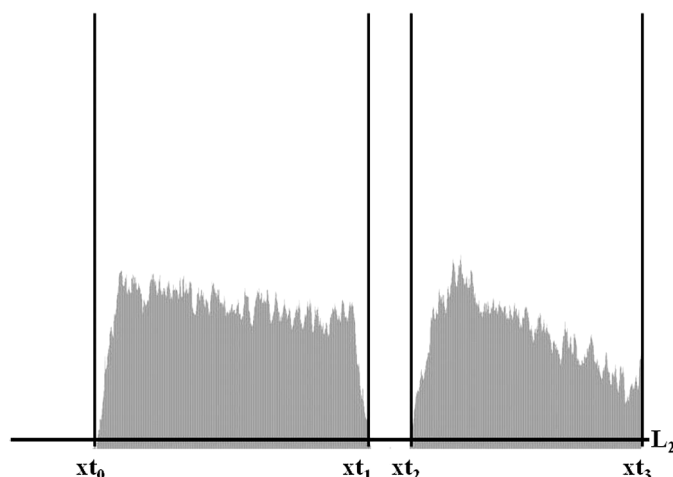
Με άλλα λόγια, ψάχνουμε για περιοχές κειμένου οι οποίες προέκυψαν μετά την εξομάλυνση και έχουν πλάτος μεγαλύτερο από το 1/3 του συνολικού πλάτους της εικόνας. Υπάρχουν τρεις διαφορετικές περιπτώσεις που έχουμε να αντιμετωπίσουμε ανάλογα με τη δομή του εγγράφου: να εντοπίσουμε (i) δύο περιοχές, (ii) μία περιοχή ή (iii) καμία περιοχή που να ικανοποιούν την παραπάνω συνθήκη. Στη συνέχεια εξετάζουμε όλες τις παραπάνω περιπτώσεις:

- Στην πρώτη περίπτωση εξετάζουμε αν μία από τις δύο περιοχές ανήκει σε γειτονική σελίδα και άρα πρέπει να αφαιρεθεί. Θέτουμε xt_0 , xt_1 και xt_2 , xt_3 τα όρια των περιοχών (Σχήμα 3.6) και υπολογίζουμε τη διασυσχέτιση του σήματος για κάθε περιοχή SC^1, SC^2 [114]. Για να υπολογίζουμε τη διασυσχέτιση του σήματος σε ολόκληρη την περιοχή υπολογίζουμε αρχικά τη διασυσχέτιση του σήματος SC_y^i για κάθε γραμμή της περιοχής i :

$$SC_y^i(a, y) = 1 - \frac{2}{M} \sum_{k=0}^M (I(k, y) XOR I(k, y+a))$$

όπου M είναι το πλάτος της περιοχής και a η απόσταση μεταξύ δύο γραμμών. Έπειτα, η συνολική διασυσχέτιση του σήματος SC^i της περιοχής i προκύπτει από τη μέση τιμή όλων των SC_y^i . Η διασυσχέτιση παίρνει τιμές από -1 (οι δύο γραμμές είναι εντελώς διαφορετικές) μέχρι 1 (οι δύο γραμμές είναι ίδιες). Θεωρούμε ότι σε μία περιοχή κειμένου γειτονικής σελίδας, όπου οι γραμμές κειμένου εμφανίζουν συνήθως μία παραμόρφωση, η διασυσχέτιση θα έχει μικρές τιμές (< -0.5). Συνεπώς, υπολογίζουμε τα όρια XT_1 και XT_2 ως εξής:

$$\begin{aligned} \text{εάν } (SC^1 > -0.5 \text{ και } SC^2 > -0.5) \text{ τότε } (XT_1 = xt_0 \text{ και } XT_2 = xt_3) \\ \text{αλλιώς εάν } (SC^1 > SC^2) \text{ τότε } (XT_1 = xt_0 \text{ και } XT_2 = xt_1) \\ \text{αλλιώς } (XT_1 = xt_2 \text{ και } XT_2 = xt_3) \end{aligned}$$



Σχήμα 3.6: Το κατακόρυφο ιστόγραμμα της εικόνας του Σχήματος 3.5 και ο προσδιορισμός των ορίων των δύο περιοχών κειμένου.

- Στη δεύτερη περίπτωση εξετάζουμε εάν το κείμενο της γειτονικής σελίδας και η περιοχή κειμένου βρίσκονται πολύ κοντά. Θέτουμε x_{t_0} , x_{t_1} τα όρια της περιοχής. Εάν το πλάτος της περιοχής είναι μικρότερο από το $0.7*(XB_2 - XB_1)$ τότε θεωρούμε ότι δεν υπάρχει κείμενο γειτονικής σελίδας. Διαφορετικά, διαιρούμε την περιοχή σε οκτώ υποπεριοχές και υπολογίζουμε τη διασυσχέτιση του σήματος για κάθε υποπεριοχή SC^1, \dots, SC^8 . Τέλος, υπολογίζουμε τα όρια XT_1 και XT_2 ως εξής:

Εξετάζουμε την πρώτη και την τελευταία υποπεριοχή. Εάν $SC^1 > -0.5$ και $SC^8 > -0.5$ θεωρούμε ότι δεν υπάρχει κείμενο γειτονικής σελίδας και άρα θέτουμε τα όρια $XT_1 = x_{t_0}$ και $XT_2 = x_{t_1}$.

Εάν $SC^1 < -0.5$ τότε είναι πιθανόν να υπάρχει κείμενο γειτονικής σελίδας στο αριστερό όριο του εγγράφου. Ψάχνουμε για την πρώτη διαδοχική υποπεριοχή i στην οποία αλλάζει κατάσταση η διασυσχέτιση του σήματος, δηλαδή $SC^i \geq -0.5$. Έπειτα, βρίσκουμε τη θέση x' για την οποία το κατακόρυφο ιστόγραμμα έχει την ελάχιστη τιμή μέσα στις πρώτες $i-1$ υποπεριοχές. Θεωρούμε τη θέση x' ως το όριο της περιοχής κειμένου, $XT_1 = x'$ & $XT_2 = x_{t_1}$, εάν το πλάτος της περιοχής που προκύπτει είναι επιτρεπτό, δηλαδή $(x_{t_1} - x') \geq (XB_2 - XB_1)/3$. Διαφορετικά θεωρούμε $XT_1 = x_{t_0}$ και $XT_2 = x_{t_1}$. Εφαρμόζουμε παρόμοια διαδικασία εάν $SC^8 < -0.5$, που σημαίνει ότι είναι πιθανόν να υπάρχει κείμενο γειτονικής σελίδας στο δεξιό όριο του εγγράφου.

- Στην τελευταία περίπτωση, όπου δεν προκύπτει καμία περιοχή, το έγγραφο αποτελείται μάλλον από δύο ή περισσότερες στήλες και προσπαθούμε να τις εντοπίσουμε και να τις διαχωρίσουμε από τη γειτονική σελίδα. Ελέγχουμε εάν υπάρχουν περιοχές με πλάτος μεγαλύτερο από $(XB_2 - XB_1)/4$. Εάν βρεθούν δύο ή λιγότερες περιοχές θεωρούμε ότι δεν υπάρχει κείμενο γειτονικής σελίδας. Εάν βρεθούν τρεις ή περισσότερες περιοχές τότε επιλεγούμε την πρώτη και την τελευταία περιοχή και εφαρμόζουμε τον αλγόριθμο της πρώτης περίπτωσης που περιγράψαμε προηγουμένως.

Αφαίρεση Κειμένου Γειτονικής Σελίδας: Έχοντας προσδιορίσει τα όρια XT_1 και XT_2 αφαιρούμε τη γειτονική σελίδα και προκύπτει η τελική εικόνα $I_f(x, y)$ ως εξής:

$$I_f(x, y) = \begin{cases} I_c(x, y) & \text{εάν } I_l(x, y) = i \text{ και } \exists (x_1, y_1): I_l(x_1, y_1) = i \text{ και } (x_1 \geq XT_1 \text{ ή } x_1 \leq XT_2) \\ 0 & \text{διαφορετικά} \end{cases}$$

Δηλαδή, όλα τα pixels κειμένου ενός συνδεδεμένου συστατικού C_i , το οποίο δεν έχει τουλάχιστον ένα pixel εντός των ορίων XT_1 και XT_2 , μετατρέπονται σε pixels υπόβαθρου.

3.3 Διαχωρισμός Διπλών Σελίδων

Η προτεινόμενη μέθοδος εφαρμόζεται σε εικόνες εγγράφων που περιέχουν δύο σελίδες. Στο Σχήμα 3.7 παρουσιάζεται το διάγραμμα ροής της μεθόδου το οποίο αποτελείται από τρία στάδια: ένα στάδιο προ-επεξεργασίας, την εύρεση των κάθετων ορίων των σελίδων και τέλος την εύρεση των οριζώντιων ορίων. Στόχος μας είναι ο εντοπισμός και η αφαίρεση του περιθωρίου καθώς και ο διαχωρισμός των σελίδων σε δύο ξεχωριστές εικόνες. Η μέθοδος στηρίζεται σε μία παραλλαγή των κλασικών προβολών η οποία δίνει έμφαση στα μεγάλα συνεχόμενα τμήματα του υπόβαθρου της εικόνας και προσπαθεί να εντοπίσει τα πλαίσια που περικλείουν τις δύο σελίδες αφήνοντας εκτός το περιθώριο (Σχήμα 3.8). Στη συνέχεια ακολουθεί η αναλυτική περιγραφή όλων των σταδίων.

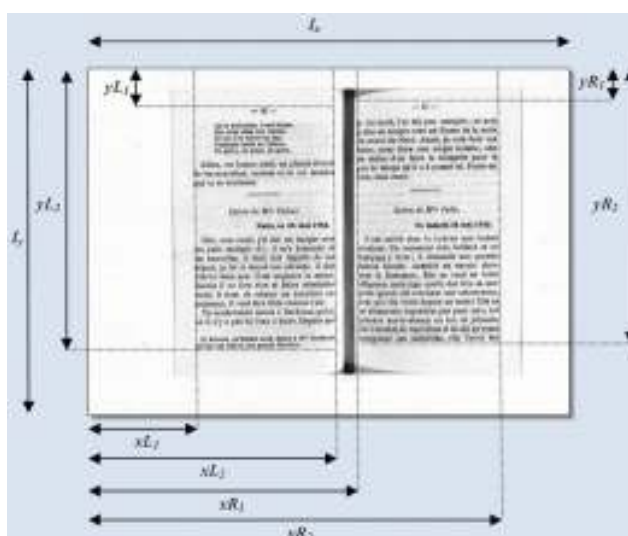
i) Προ-επεξεργασία

Στο στάδιο της προ-επεξεργασίας μετατρέπουμε την εικόνα σε δυαδική, εάν είναι διαβάθμισης γκρι, χρησιμοποιώντας τη μέθοδο που περιγράφεται στο [6]. Έπειτα, υπολογίζουμε το κυρίαρχο ύψος γράμματος φτιάχνοντας το ιστόγραμμα των υψών όλων των πλαισίων που περικλείουν τα συνδεδεμένα συστατικά της

εικόνας [80]. Στη συνέχεια απομακρύνουμε το θόρυβο αφαιρώντας τα συνδεδεμένα συστατικά που έχουν ύψος ή πλάτος μικρότερο από το ένα δέκατο του κυρίαρχου ύψους γράμματος. Τέλος, εφαρμόζουμε μία διαδικασία εξομάλυνσης της εικόνας με τη βοήθεια της τεχνικής ARLSA (Adaptive Run Length Smoothing Algorithm) [115]. Χρησιμοποιώντας αυτή την τεχνική πετυχαίνουμε να ενώσουμε τα συνδεδεμένα συστατικά που ανήκουν στην ίδια γραμμή κειμένου χωρίς να συνδέονται με το θόρυβο περιθωρίου.



Σχήμα 3.7: Διάγραμμα ροής προτεινόμενης μεθόδου για τον εντοπισμό και την αφαίρεση του περιθωρίου καθώς και τον διαχωρισμό των διπλών σελίδων.



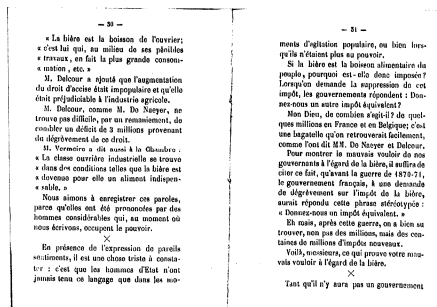
Σχήμα 3.8: Εικόνα εγγράφου που περιέχει δύο σελίδες μαζί με τα όρια των σελίδων όπως υπολογίζονται από την προτεινόμενη μέθοδο.

ii) Εύρεση κάθετων ορίων

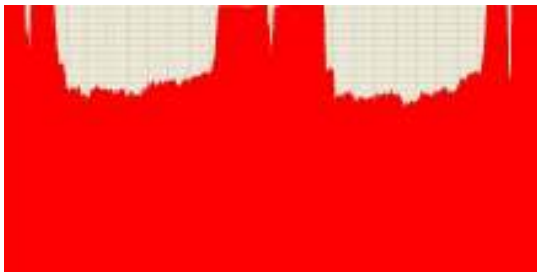
Στόχος μας είναι να εντοπίσουμε τα κάθετα όρια που περικλείουν τις δύο σελίδες, δηλαδή τα όρια xL_1, xL_2 για την αριστερή σελίδα και xR_1, xR_2 για τη δεξιά σελίδα (Σχήμα 3.8). Βασιζόμαστε σε μία παραλλαγή των κλασικών προβολών η οποία δίνει έμφαση στα μεγάλα συνεχόμενα τμήματα του υπόβαθρου της εικόνας. Οι προβολές αυτές αποδείχθηκαν πιο αποτελεσματικές για τον εντοπισμό περιοχών κειμένου και ορίζονται ως εξής:

$$HV(x) = \frac{\sum_{j=1}^{w_y} (y_{j2} - y_{j1})^2}{(1 - 2 * a) * I_y}$$

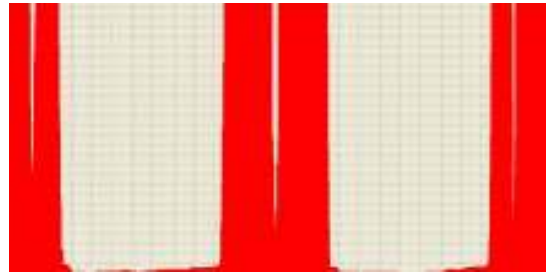
όπου w_y είναι το πλήθος των τμημάτων του υπόβαθρου $(i, y_{j1}) - (i, y_{j2})$ για τη στήλη $x = i$ μέσα στο διάστημα $y = a * I_y, \dots, (1 - a) * I_y$, με $a = 1/8$. Περιορίζουμε το διάστημα των προβολών για να μην επηρεαστούν από το θόρυβο που βρίσκεται στα όρια του εγγράφου. Συνεπώς $HV(x) \in [0, \dots, (1 - 2 * a) * I_y]$. Ένα παράδειγμα των προτεινόμενων προβολών σε σύγκριση με τις κλασικές προβολές παρουσιάζεται στο Σχήμα 3.9. Όπως φαίνεται, με τις προτεινόμενες προβολές οι περιοχές κειμένου είναι πιο εύκολα διαχωρίσιμες.



(α)



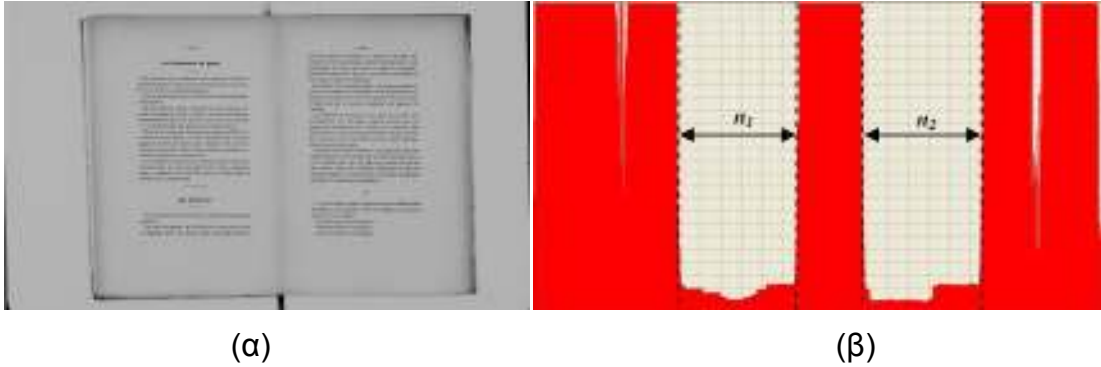
(β)



(γ)

Σχήμα 3.9: (α) Δυαδική εικόνα, (β) κλασικές προβολές pixels υπόβαθρου, (γ) προβολές συνεχόμενων τμημάτων του υπόβαθρου.

Μπορούμε να υποθέσουμε με ασφάλεια ότι εάν ισχύει η συνθήκη $HV(x) < (1 - 2 * a) I_y / 2$ τότε η στήλη x ανήκει σε περιοχή κειμένου. Συνεπώς, εάν έχουμε n συνεχόμενες στήλες, με $n > I_x / 6$, που να ικανοποιούν τη συνθήκη τότε θεωρούμε ότι έχουμε εντοπίσει μία περιοχή κειμένου. Στο Σχήμα 3.10 απεικονίζεται ένα παράδειγμα στο οποίο έχουν εντοπιστεί δύο περιοχές κειμένου με n_1 και n_2 συνεχόμενες στήλες κειμένου, αντίστοιχα.



Σχήμα 3.10: (α) Αρχική εικόνα, (β) προβολές συνεχόμενων τμημάτων του υπόβαθρου όπου επισημαίνονται οι δύο περιοχές κειμένου.

Αν και ο εντοπισμός δύο περιοχών που να ικανοποιούν την παραπάνω συνθήκη είναι το πιο συνηθισμένο, υπάρχουν και περιπτώσεις που εντοπίζουμε παραπάνω από δύο, μόνο μία ή και καμιά περιοχή κειμένου. Στη συνέχεια εξετάζουμε όλες τις παραπάνω περιπτώσεις:

- Εάν εντοπίσουμε δύο περιοχές κειμένου αναθέτουμε τα όρια της πρώτης περιοχής στα όρια xL_1 , xL_2 της αριστερής σελίδας και αντίστοιχα της δεύτερης περιοχής στα όρια xR_1 , xR_2 της δεξιάς σελίδας.
- Εντοπίζεται μόνο μία περιοχή κειμένου όταν η μία από τις δύο σελίδες είναι κενή ή περιέχει πολύ μικρό όγκο πληροφοριών. Σε αυτή την περίπτωση προσεγγίζουμε τα όρια της βασιζόμενοι στα όρια της περιοχής που εντοπίστηκε. Εάν θεωρήσουμε x_1 και x_2 τα όρια της περιοχής που έχουμε εντοπίσει τότε τα όρια των δύο σελίδων υπολογίζονται ως εξής:

$$xL_1 = \begin{cases} x_1, & \text{εάν } x_1 < \frac{I_x}{2} \\ I_x - x_2, & \text{διαφορετικά} \end{cases}, \quad xL_2 = \begin{cases} x_2, & \text{εάν } x_1 < \frac{I_x}{2} \\ I_x - x_1, & \text{διαφορετικά} \end{cases}$$

$$xR_1 = \begin{cases} x_1, & \text{εάν } x_1 > \frac{I_x}{2} \\ I_x - x_2, & \text{διαφορετικά} \end{cases}, \quad xR_2 = \begin{cases} x_2, & \text{εάν } x_1 > \frac{I_x}{2} \\ I_x - x_1, & \text{διαφορετικά} \end{cases}$$

- Στην περίπτωση που έχουμε έγγραφο με πολλές στήλες ή πολύπλοκη δομή (π.χ. περιοδικά) υπάρχει το ενδεχόμενο να εντοπίσουμε περισσότερες από δύο περιοχές κειμένου ή ακόμα και καμιά. Σε αυτή την περίπτωση ψάχνουμε για την κατακόρυφη διαχωριστική γραμμή SL_x ανάμεσα στις δύο σελίδες. Η SL_x είναι εκείνη η στήλη της εικόνας που περιέχει τα περισσότερα pixels υπόβαθρου και βρίσκεται πιο κοντά στο μέσο της εικόνας:

$$SL_x = \arg \max_{x \in [\frac{I_x}{2} - c * I_x, \frac{I_x}{2} + c * I_x]} HV(x)$$

όπου η παράμετρος c ορίζει την απόσταση από το μέσο της εικόνας. Αφού έχουμε εντοπίσει την κατακόρυφη διαχωριστική γραμμή SL_x τα όρια των δύο σελίδων υπολογίζονται ως εξής:

$$xL_1 = \max_{x \in [0, d * I_x]} (\arg \max HV(x)), \quad xL_2 = \min_{x \in [SL_x - e * I_x, SL_x]} (\arg \max HV(x))$$

$$xR_1 = \max_{x \in [SL_x, SL_x + e * I_x]} (\arg \max HV(x)), \quad xR_2 = \min_{x \in [I_x - d * I_x, I_x]} (\arg \max HV(x))$$

όπου η παράμετρος d ορίζει τη μέγιστη απόσταση που εκτιμούμε ότι μπορεί να ξεκινάει μία σελίδα από τα όρια της εικόνας και η παράμετρος e ορίζει τη μέγιστη απόσταση από το μέσο της εικόνας ($d = e = 1/6$).

Στις δύο πρώτες περιπτώσεις προχωράμε σε ένα επαναπροσδιορισμό των ορίων με στόχο να μην κοπούν μικρές περιοχές κειμένου που προεξέχουν από την κύρια σελίδα και επιτυγχάνεται με τις παρακάτω εξισώσεις:

$$xL_1 = \max_{x \in [xL_1 - d * I_x, xL_1]} (\arg \max HV(x)), \quad xL_2 = \min_{x \in [xL_2, xL_2 + e * I_x]} (\arg \max HV(x))$$

$$xR_1 = \max_{x \in [xR_1 - d * I_x, xR_1]} (\arg \max HV(x)), \quad xR_2 = \min_{x \in [xR_2, xR_2 + e * I_x]} (\arg \max HV(x))$$

Στο Σχήμα 3.11 παρουσιάζεται ένα παράδειγμα επαναπροσδιορισμού του δεξιού ορίου (xR_2) της σελίδας.

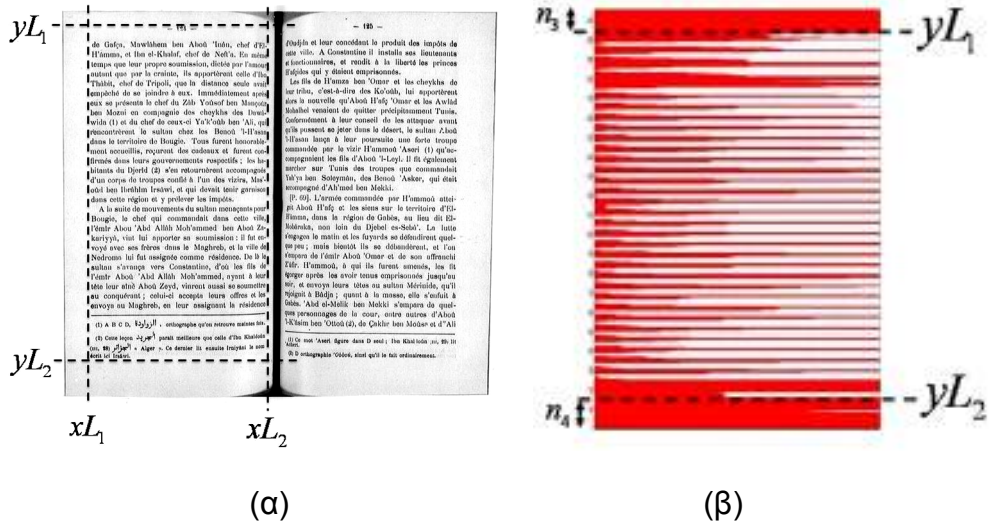


Σχήμα 3.11: Παράδειγμα με την αρχική εκτίμηση του ορίου xR_2 (κάθετη γραμμή "...") και μετά τον επαναπροσδιορισμό του (κάθετη γραμμή "---").

iii) Εύρεση οριζόντιων ορίων

Στο στάδιο αυτό επεξεργαζόμαστε κάθε σελίδα ξεχωριστά μέσα στα κάθετα όρια (xL_1 , xL_2 και xR_1 , xR_2) που υπολογίσαμε στο προηγούμενο στάδιο. Στόχος μας είναι να εντοπίσουμε τα οριζόντια όρια που περικλείουν τις δύο σελίδες, δηλαδή τα όρια yL_1 , yL_2 για την αριστερή σελίδα και yR_1 , yR_2 για τη δεξιά σελίδα (Σχήμα 3.8).

Αρχικά υπολογίζουμε τις οριζόντιες προβολές συνεχόμενων τμημάτων του υπόβαθρου $HH(y)$ με τον ίδιο τρόπο που υπολογίζονται και οι κατακόρυφες προβολές. Μπορούμε να υποθέσουμε ότι εάν ισχύει η συνθήκη $HH(y) > f^*(xL_2 - xL_1)$ (ή $HH(y) > f^*(xR_2 - xR_1)$) τότε η γραμμή y ανήκει στο περιθώριο. Συνεπώς, εάν έχουμε n συνεχόμενες γραμμές, με $n > I_y / 25$, που να ικανοποιούν τη συνθήκη τότε θεωρούμε ότι έχουμε εντοπίσει το επάνω ή το κάτω όριο της σελίδας. Για το επάνω όριο ψάχνουμε στο διάστημα $[0, g^* I_y]$ και αντίστοιχα στο $[(1-g)^* I_y, I_y]$ για το κάτω όριο ($f = 2/3, g = 1/4$). Αν δεν ικανοποιηθεί η συνθήκη τότε θεωρούμε ότι δεν υπάρχει το αντίστοιχο περιθώριο. Στο Σχήμα 3.12 απεικονίζεται ένα παράδειγμα στο οποίο έχουν εντοπιστεί τα όρια για την αριστερή σελίδα.



Σχήμα 3.12: Εντοπισμός οριζόντιων ορίων: (α) Αρχική εικόνα με τα κάθετα και οριζόντια όρια της αριστερής σελίδας, (β) προβολές συνεχόμενων τμημάτων του υπόβαθρου της αριστερής σελίδας με n_3 και n_4 συνεχόμενες γραμμές περιθωρίου.

3.4 Πειραματικά Αποτελέσματα

Για την αξιολόγηση των προτεινόμενων μεθόδων δημιουργήσαμε χειροκίνητα το ground-truth και χρησιμοποιήσαμε μία αυτοματοποιημένη διαδικασία αξιολόγησης. Η διαδικασία αυτή βασίζεται στον υπολογισμό του πλήθους των pixels κειμένου στην περιοχή κειμένου, βάσει του ground-truth, και των pixels κείμενου στο αποτέλεσμα που προκύπτει μετά την εφαρμογή της μεθόδου.

Υπολογίζουμε δύο μετρικές, την ακρίβεια (*Precision*) και τον αριθμό εύστοχων αποκρίσεων (*Recall*) που εκφράζουν την ακρίβεια της μεθόδου και το ποσοστό της περιοχής κειμένου που παραμένει στην τελική εικόνα, αντίστοιχα [116]. Έστω G το σύνολο των pixels κειμένου στην περιοχή κειμένου, βάσει του ground-truth,

R το σύνολο των pixels κειμένου στο αποτέλεσμα και $T(s)$ μία συνάρτηση που υπολογίζει τα στοιχεία του συνόλου s . Οι μετρικές υπολογίζονται ως εξής:

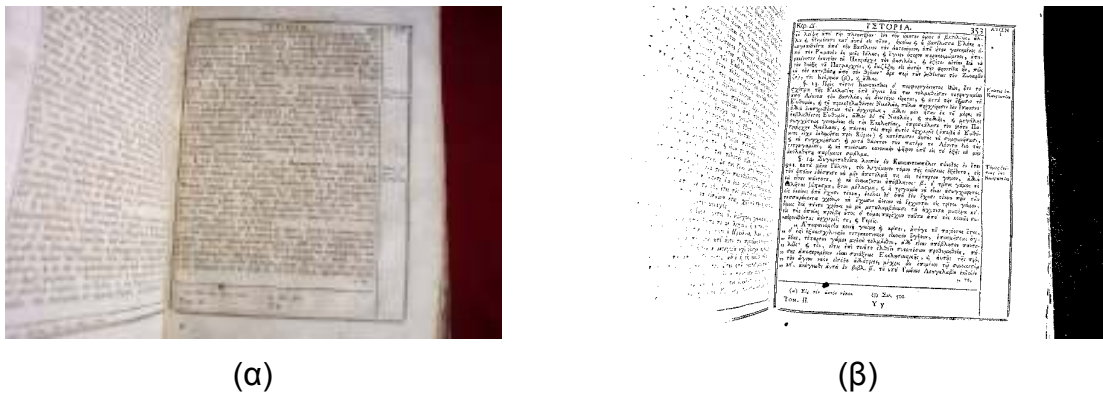
$$Precision = \frac{T(G \cap R)}{T(R)}, \quad Recall = \frac{T(G \cap R)}{T(G)}$$

Η μετρική F -Measure (FM), συνδυάζοντας το $Precision$ και $Recall$, υπολογίζεται ως εξής:

$$FM = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Αποτίμηση μεθόδου για τον εντοπισμό και την αφαίρεση του περιθωρίου καθώς και της γειτονικής σελίδας

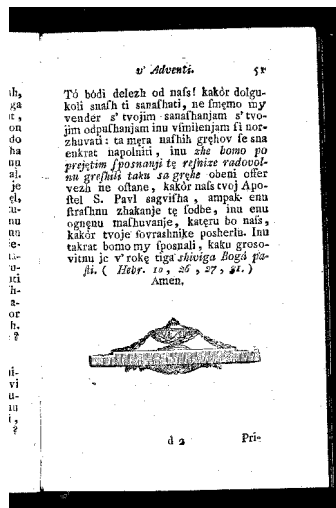
Η προτεινόμενη μέθοδος για τον εντοπισμό και την αφαίρεση του περιθωρίου καθώς και της γειτονικής σελίδας δοκιμάστηκε σε δύο σύνολα ιστορικών εγγράφων. Το πρώτο σύνολο αποτελείται από ιστορικά έγγραφα του έργου “ΠΟΛΥΤΙΜΟ” [111] το οποίο περιέχει 370 εικόνες εγγράφων από 6 διαφορετικά βιβλία. Όλες οι εικόνες του συνόλου έχουν περιθώριο και τμήματα κειμένου γειτονικής σελίδας που πρέπει να αφαιρεθούν. Το δεύτερο σύνολο αποτελείται από ιστορικά έγγραφα του έργου “IMPACT” [112] το οποίο περιέχει 22383 εικόνες εγγράφων από 7 διαφορετικές βιβλιοθήκες και περισσότερα από 150 ιστορικά βιβλία, εφημερίδες και περιοδικά. Χαρακτηριστικά παραδείγματα εικόνων από τα δύο σύνολα παρουσιάζονται στα Σχήματα 3.13 έως 3.15. Επίσης, για λόγους σύγκρισης εφαρμόσαμε στα δύο σύνολα τη μέθοδο [20] των Le et al. (§ 2.1.3), η οποία είναι η μοναδική που αφαιρεί το περιθώριο και τη γειτονική σελίδα, καθώς και τρία εμπορικά προγράμματα, το BookRestorer [117], το WiseBook [118] και το ScanFix [119]. Τα πειραματικά αποτελέσματα για τα δύο σύνολα παρουσιάζονται στους Πίνακες 3.1 και 3.2, αντίστοιχα.



Σχήμα 3.13: Ιστορικό έγγραφο από το σύνολο εικόνων “ΠΟΛΥΤΙΜΟ”: (α) Αρχική εικόνα διαβάθμισης γκρι, (β) δυαδική εικόνα μετά την εφαρμογή της τεχνικής [6].



(α)



(β)

Σχήμα 3.14: Ιστορικό έγγραφο από το σύνολο εικόνων “IMPACT” (NUK): (α) Αρχική εικόνα διαβάθμισης γκρι, (β) δυαδική εικόνα μετά την εφαρμογή της τεχνικής [6].



(α)



(β)

Σχήμα 3.15: Ιστορικό έγγραφο από το σύνολο εικόνων “IMPACT” (BL): (α) Αρχική εικόνα διαβάθμισης γκρι, (β) δυαδική εικόνα μετά την εφαρμογή της τεχνικής [6].

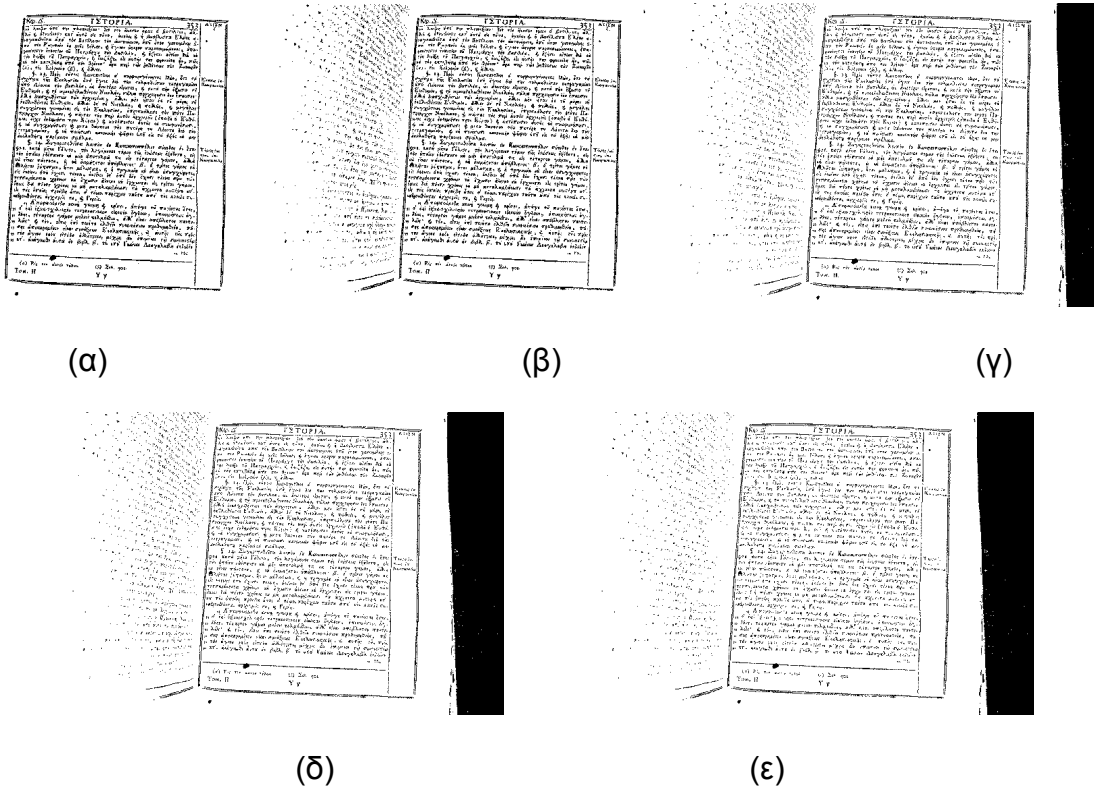
Πίνακας 3.1: Πειραματικά αποτελέσματα για το σύνολο εικόνων “ΠΟΛΥΤΙΜΟ”.

Μέθοδος	Precision (%)	Recall (%)	F-Measure (%)
Προτεινόμενη Μέθοδος	91.11	96.95	93.94
Le et al. [20]	70.90	99.33	82.74
BookRestorer [117]	74.33	95.47	83.58
WiseBook [118]	53.00	99.02	69.05
ScanFix [119]	51.49	87.96	64.95

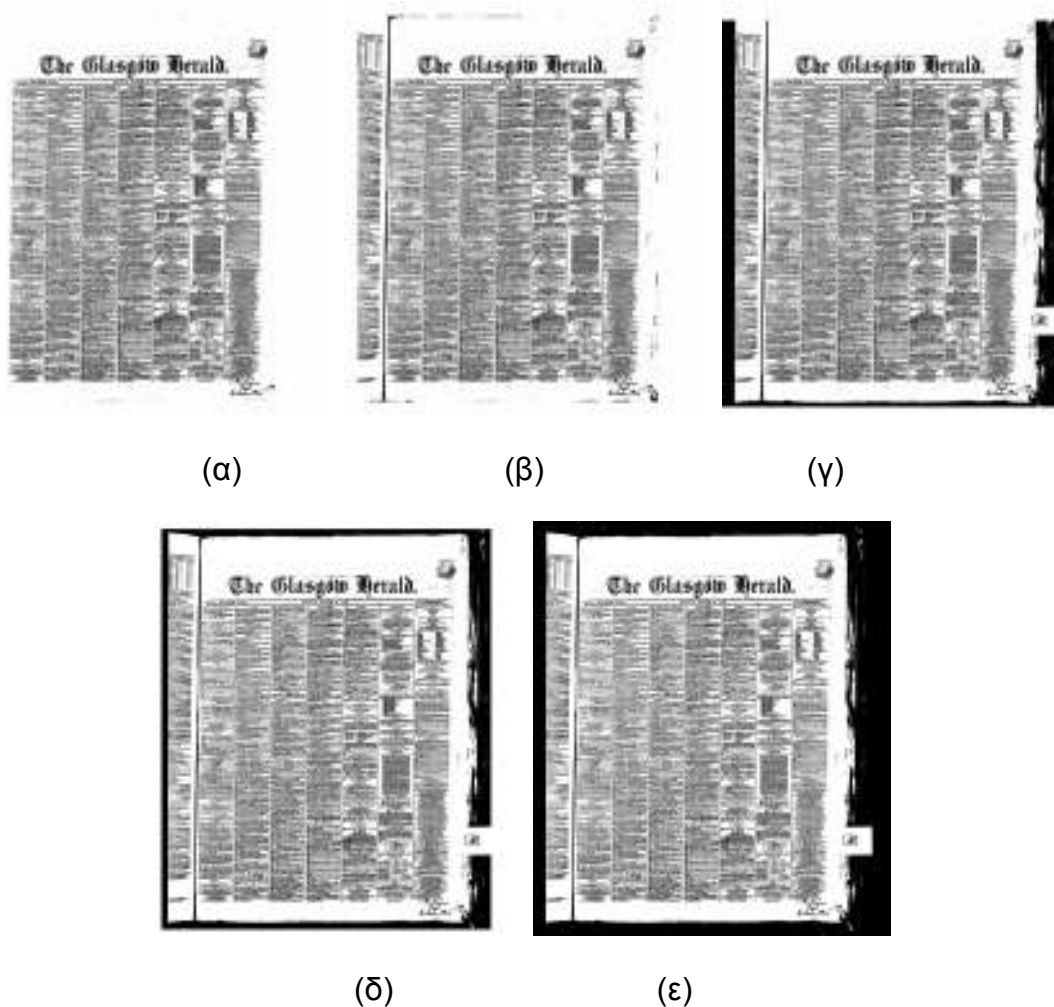
Πίνακας 3.2: Πειραματικά αποτελέσματα για το σύνολο εικόνων “IMPACT”.

Βιβλιοθήκη	Αριθμός Εικόνων	Μέθοδος	Precision (%)	Recall (%)	F-Measure (%)
BL [120]	1631	Προτεινόμενη Μέθοδος	98.94	98.18	98.56
		Le et al. [20]	88.89	99.05	93.70
		BookRestorer [117]	80.30	99.36	88.82
		WiseBook [118]	70.98	99.38	82.81
		ScanFix [119]	59.23	95.42	73.09
BNE [121]	7543	Προτεινόμενη Μέθοδος	99.88	99.27	99.57
		Le et al. [20]	99.58	86.64	92.66
		BookRestorer [117]	95.46	93.02	94.22
		WiseBook [118]	83.24	99.49	90.64
		ScanFix [119]	89.57	98.78	93.95
BNF [122]	7677	Προτεινόμενη Μέθοδος	98.29	99.26	98.77
		Le et al. [20]	97.98	98.86	98.42
		BookRestorer [117]	97.00	99.68	98.32
		WiseBook [118]	86.10	99.58	92.35
		ScanFix [119]	86.23	99.03	92.19
BSB[123]	2417	Προτεινόμενη Μέθοδος	96.86	93.14	94.96
		Le et al. [20]	96.08	91.53	93.75
		BookRestorer [117]	95.27	95.22	95.24
		WiseBook [118]	92.24	95.19	93.69
		ScanFix [119]	91.96	96.54	94.19
NUK [124]	1416	Προτεινόμενη Μέθοδος	98.01	99.15	98.58
		Le et al. [20]	93.20	99.09	96.05
		BookRestorer [117]	84.26	99.56	91.27
		WiseBook [118]	72.44	98.36	83.43
		ScanFix [119]	73.38	98.55	84.12
NLB [125]	315	Προτεινόμενη Μέθοδος	99.98	99.87	99.92
		Le et al. [20]	99.85	99.97	99.91
		BookRestorer [117]	99.83	99.96	99.89
		WiseBook [118]	99.18	98.61	98.89
		ScanFix [119]	99.04	80.61	88.88
ONB [126]	1384	Προτεινόμενη Μέθοδος	96.63	98.24	97.43
		Le et al. [20]	96.48	99.06	97.75
		BookRestorer [117]	93.76	99.62	96.60
		WiseBook [118]	74.77	99.09	85.23
		ScanFix [119]	80.14	97.67	88.04
Συνολικά	22383	Προτεινόμενη Μέθοδος	98.62	98.46	98.54
		Le et al. [20]	97.28	94.01	95.62
		BookRestorer [117]	94.11	96.92	95.50
		WiseBook [118]	83.32	98.94	90.46
		ScanFix [119]	85.00	98.04	91.05

Όπως φαίνεται από τους πίνακες, η προτεινόμενη μέθοδος επιτυγχάνει τα καλύτερα αποτελέσματα σε σχέση με την τεχνική των Le et al. [20] αλλά και τα τρία εμπορικά προγράμματα με συνολικό $FM=93.94\%$ για το σύνολο εικόνων του “ΠΟΛΥΤΙΜΟ” και $FM=98.54\%$ για το σύνολο εικόνων του “IMPACT”. Οι δύο μετρικές, *Precision* και *Recall*, έχουν υψηλές τιμές και στα δυο σύνολα εικόνων που δείχνει ότι η προτεινόμενη μέθοδος εντοπίζει και αφαιρεί το περιθώριο και τη γειτονική σελίδα χωρίς συγχρόνως να χάνεται χρήσιμη πληροφορία από το έγγραφο. Η τεχνική των Le et al. [20] και το BookRestorer [117] από τα εμπορικά προγράμματα επιτυγχάνουν τα καλύτερα αποτελέσματα από τις συγκρινόμενες τεχνικές με μικρές διαφορές μεταξύ τους. Στα Σχήματα 3.16 και 3.17 απεικονίζονται παραδείγματα εφαρμογής της προτεινόμενης μεθόδου και των συγκρινόμενων τεχνικών.



Σχήμα 3.16: Αποτέλεσμα εφαρμογής (α) προτεινόμενης μεθόδου, (β) Le et al. [20], (γ) BookRestorer [117], (δ) WiseBook [118] και (ε) ScanFix [119] στο έγγραφο του Σχήματος 3.13(β).



Σχήμα 3.17: Αποτέλεσμα εφαρμογής (α) προτεινόμενης μεθόδου, (β) Le et al. [20], (γ) BookRestorer [117], (δ) WiseBook [118] και (ε) ScanFix [119] στο έγγραφο του Σχήματος 3.15(β).

Αποτίμηση μεθόδου για τον διαχωρισμό διπλών σελίδων

Για την αξιολόγηση της προτεινόμενης μεθόδου η οποία εντοπίζει και αφαιρεί το περιθώριο αλλά συγχρόνως διαχωρίζει και τις διπλές σελίδες, χρησιμοποιήσαμε ένα σύνολο εικόνων που αποτελείται από ιστορικά έγγραφα του έργου “IMPACT” [112] και περιέχει 3467 διπλές σελίδες, από 4 διαφορετικές βιβλιοθήκες και περισσότερα από 50 ιστορικά βιβλία. Χαρακτηριστικά παραδείγματα εικόνων παρουσιάζονται στο Σχήμα 3.18. Για να υπολογίσουμε τις μετρικές σε μία εικόνα, υπολογίζουμε το *Precision* και το *Recall* για την αριστερή και τη δεξιά σελίδα και στη συνέχεια υπολογίζουμε το μέσο όρο. Η πλειοψηφία των τεχνικών που διαχωρίζουν τις διπλές σελίδες δεν εντοπίζουν το πλαίσιο που περικλείει κάθε σελίδα ώστε να αφαιρέσουν και το περιθώριο αλλά εντοπίζουν μόνο τη διαχωριστική γραμμή ανάμεσα στις δύο σελίδες, Συνεπώς, δεν είναι δυνατή η άμεση σύγκρισή τους. Παρόλα αυτά, εφαρμόσαμε στο σύνολο των εικόνων το

εμπορικό πρόγραμμα ABBYY FineReader Engine 10 [127], το οποίο προσπαθεί να εντοπίσει τη διαχωριστική γραμμή ανάμεσα στις δύο σελίδες. Τα πειραματικά αποτελέσματα παρουσιάζονται στον Πίνακα 3.3.

Πίνακας 3.3: Πειραματικά αποτελέσματα μεθόδων για τον διαχωρισμό των διπλών σελίδων.

Βιβλιοθήκη	Αριθμός Εικόνων	Προτεινόμενη μέθοδος			ABBYY FineReader Engine 10 [127]		
		<i>Precision</i> (%)	<i>Recall</i> (%)	<i>F-Measure</i> (%)	<i>Precision</i> (%)	<i>Recall</i> (%)	<i>F-Measure</i> (%)
BL [120]	2171	92.11	98.15	95.03	66.23	90.56	76.51
BNF [122]	458	98.97	98.99	98.98	78.03	98.22	86.97
BSB [123]	305	94.36	99.57	96.89	49.62	89.01	63.72
NUK [124]	533	84.47	97.90	90.69	42.38	91.15	57.86
Συνολικά	3467	92.04	98.35	95.09	62.66	91.53	74.39

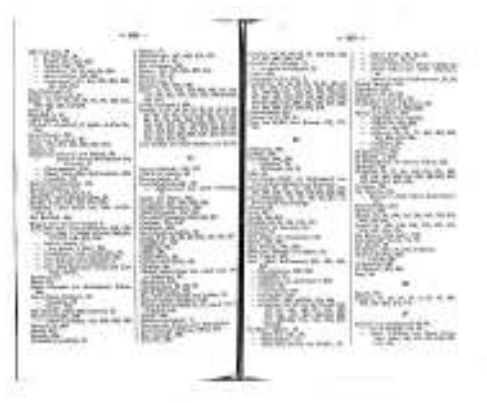
Η προτεινόμενη μέθοδος επιτυγχάνει συνολικό **FM=95.09%**. Το *Recall* είναι αρκετά υψηλό, **98.35%**, που δείχνει ότι διαχωρίζει πολύ καλά τις διπλές σελίδες χωρίς να χάνεται χρήσιμη πληροφορία, ενώ το *Precision* βρίσκεται στο **92.04%** που σημαίνει ότι υπάρχουν περιπτώσεις στις οποίες δεν αφαιρείται όλος ο θόρυβος περιθωρίου. Το εμπορικό πρόγραμμα ABBYY FineReader Engine 10 έχει πολύ χαμηλό *Precision*, 62.66%, αφού δεν αφαιρεί το θόρυβο περιθωρίου ενώ από το *Recall* φαίνεται ότι σε πολλές περιπτώσεις δε μπορεί να διαχωρίσει σωστά τις διπλές σελίδες. Στο Σχήμα 3.19 παρουσιάζονται παραδείγματα εφαρμογής της προτεινόμενης μεθόδου όσο και του προγράμματος ABBYY FineReader Engine 10.



(α)



(β)



(γ)



(δ)

Σχήμα 3.18: Ιστορικά έγγραφα που περιέχουν δύο σελίδες από το σύνολο εικόνων “IMPACT”: (α)-(β) Αρχικές εικόνες εγγράφων διαβάθμισης γκρι, (γ)-(δ) δυαδικές εικόνες εγγράφων μετά την εφαρμογή της τεχνικής [6].



(α)



(β)

Σχήμα 3.19: Οι δύο σελίδες που προκύπτουν μετά την εφαρμογή (α) της προτεινόμενης μεθόδου και (β) του προγράμματος ABBYY FineReader Engine 10 [127] στο έγγραφο του Σχήματος 3.18(α).

4. ΔΙΟΡΘΩΣΗ ΤΟΠΙΚΗΣ ΚΑΜΠΥΛΟΤΗΤΑΣ

4.1 Εισαγωγή

Οι τεχνικές διόρθωσης της τοπικής καμπυλότητας ταξινομούνται σε δύο βασικές κατηγορίες: (i) τεχνικές που βασίζονται στην τριδιάστατη ανακατασκευή του εγγράφου και (ii) τεχνικές που βασίζονται μόνο στη διδιάστατη επεξεργασία του εγγράφου όπως έχει ήδη συζητηθεί στην παράγραφο 2.1.5. Η πλειοψηφία των τεχνικών της δεύτερης κατηγορίας βασίζεται στην κατάτμηση των γραμμών κειμένου (Πίνακας 2.2). Η κατάτμηση όμως των γραμμών κειμένου στα παραμορφωμένα έγγραφα είναι μία δύσκολη διαδικασία που οδηγεί συχνά σε λάθη, με αποτέλεσμα οι τεχνικές διόρθωσης της τοπικής καμπυλότητας να αποτυγχάνουν.

Σε αυτό το κεφάλαιο παρουσιάζεται μία νέα μέθοδος διόρθωσης της τοπικής καμπυλότητας, η οποία έχει δημοσιευθεί στο [128]. Στόχος μας είναι να μπορέσουμε να εφαρμόσουμε με μεγαλύτερη επιτυχία το στάδιο της κατάτμησης των γραμμών κειμένου στα παραμορφωμένα έγγραφα. Η νέα μέθοδος βασίζεται στη διδιάστατη επεξεργασία του εγγράφου και εφαρμόζεται σε δύο φάσεις. Η πρώτη φάση έχει ως στόχο να διορθώσει τις έντονες παραμορφώσεις του εγγράφου και να παράγει ένα πρώτο αποτέλεσμα διόρθωσης. Το στάδιο αυτό εφαρμόζει ένα χαμηλής υπολογιστικής πολυπλοκότητας μετασχηματισμό ο οποίος απεικονίζει την προβολή της κυρτής επιφάνειας του εγγράφου σε μία διδιάστατη ορθογώνια περιοχή. Η δεύτερη φάση διορθώνει τις τοπικές παραμορφώσεις του εγγράφου σε επίπεδο λέξης παράγοντας το τελικό διορθωμένο αποτέλεσμα απαλλαγμένο απ' όλες τις παραμορφώσεις. Η δεύτερη φάση εφαρμόζεται μετά την αρχική διόρθωση του εγγράφου, συνεπώς οι έντονες παραμορφώσεις έχουν διορθωθεί και άρα είναι πιο εύκολο να εφαρμοστεί η κατάτμηση των γραμμών κειμένου με επιτυχία. Επιπλέον, στο κεφάλαιο αυτό παρουσιάζεται μία βελτίωση της προτεινόμενης μεθόδου ενισχύοντας τη σημαντικά, η οποία έχει δημοσιευθεί στο [129].

Η πλειοψηφία των τεχνικών διόρθωσης της τοπικής καμπυλότητας χρησιμοποιεί οπτικούς ελέγχους από χρήστες για την αποτίμηση του αποτελέσματος, με συνέπεια η αξιολόγησή τους να βασίζεται σε υποκειμενικούς κανόνες χωρίς να είναι δυνατή η αντικειμενική και ποσοτική σύγκριση διαφορετικών τεχνικών. Επιπλέον, η χρήση συστημάτων OCR, συγκρίνοντας το αποτέλεσμα της αναγνώρισης του αρχικού και του διορθωμένου εγγράφου, για έμμεση

αξιολόγηση των τεχνικών είναι ευρέως διαδεδομένη. Σε πολλές περιπτώσεις, όπως τα ιστορικά έγγραφα, δεν είναι δυνατή η χρήση ενός συστήματος OCR. Για την αξιολόγηση της προτεινόμενης μεθόδου, εκτός από την χρήση συστημάτων OCR, χρησιμοποιούμε επίσης και μία πρωτότυπη ημιαυτόματη μέθοδο η οποία παρουσιάζεται σε αυτό το κεφάλαιο και έχει δημοσιευθεί στο [130]. Η προτεινόμενη μετρική αξιολόγησης λαμβάνει υπόψη την απόκλιση των γραμμών κειμένου στη διορθωμένη εικόνα από μία ευθεία, οριζόντια γραμμή αναφοράς, η οποία θεωρείται το ιδανικό αποτέλεσμα. Επίσης, προτείνεται μία διαδικασία αντιστοίχισης σημείο προς σημείο με στόχο τη μείωση της προσπάθειας του χρήστη να προσδιορίσει τις γραμμές κειμένου στη διορθωμένη εικόνα. Η διαδικασία αυτή αντιστοιχίζει τα σημεία που έχει μαρκάρει ο χρήστης στην αρχική εικόνα με τα αντίστοιχα σημεία στη διορθωμένη εικόνα, δίνοντας έτσι τη δυνατότητα να αξιολογήσουμε απεριόριστους μεθόδους χωρίς να χρειάζεται ο χρήστης να επαναλαμβάνει τη διαδικασία του μαρκαρίσματος.

Αρχικά, πριν την περιγραφή των προτεινόμενων μεθόδων, περιγράφεται η μέθοδος [131] για τη διόρθωση της τοπικής καμπυλότητας η οποία βασίζεται στην κατάτμηση του εγγράφου σε γραμμές κειμένου και λέξεις. Περιγράφουμε αναλυτικά την μέθοδο [131] διότι η δεύτερη φάση της προτεινόμενης μεθόδου βασίζεται σε αυτή έχοντας πραγματοποιήσει σημαντικές βελτιώσεις στα πλαίσια αυτής της διατριβής.

4.2 Διόρθωση της Τοπικής Καμπυλότητας με Χρήση Κατάτμησης

Η μέθοδος [131], η οποία από εδώ και στο εξής θα αναφέρεται ως SBD (Segmentation Based Dewarping), εντοπίζει τις γραμμές κειμένου και τις λέξεις και προσπαθεί να διορθώσει τις παραμορφώσεις του εγγράφου περιστρέφοντας και μετατοπίζοντας τις λέξεις με τη βοήθεια των γραμμών αναφορών. Εάν η εικόνα είναι διαβάθμισης γκρι, το πρώτο βήμα είναι η δυαδική μετατροπή χρησιμοποιώντας την τεχνική που περιγράφεται στο [6].

4.2.1 Κατάτμηση Λέξεων και Γραμμών Κειμένου

Αρχικά, υπολογίζεται το κυρίαρχο ύψος γράμματος (AH) φτιάχνοντας το ιστόγραμμα των υψών όλων των πλαισίων που περικλείουν τα συνδεδεμένα συστατικά της εικόνας [80] με σκοπό να απομακρυνθούν τα μικρά και μεγάλα αντικείμενα που δεν ανήκουν στο κείμενο, όπως εικόνες, γράφοι, θόρυβος κ.λπ. Συνεπώς, απομακρύνονται όλα εκείνα τα συνδεδεμένα συστατικά της εικόνας που

ικανοποιούν την εξής συνθήκη:

$$h > 3 * AH \quad \text{ή} \quad h < AH / 4 \quad \text{ή} \quad w < AH / 4$$

όπου h και w προσδιορίζουν το ύψος και το πλάτος του συνδεδεμένου συστατικού, αντίστοιχα.

Για την κατάτμηση των λέξεων εφαρμόζεται εξομάλυνση της εικόνας, με τη βοήθεια του αλγορίθμου RLSA [27] και κατώφλι AH (Σχήμα 4.1(β)), ακολουθούμενη από την εύρεση των συνδεδεμένων συστατικών (Σχήμα 4.1(γ)). Αφού έχουν εντοπιστεί οι λέξεις, σαρώνεται η εικόνα από αριστερά προς τα δεξιά και από πάνω προς τα κάτω και η πρώτη λέξη που εντοπίζεται ορίζεται ως η πρώτη λέξη της πρώτης γραμμής κειμένου. Έπειτα, η λέξη αυτή συνδέεται με τις γειτονικές λέξεις στην οριζόντια διεύθυνση, αριστερά και δεξιά, με στόχο να εντοπιστούν όλες οι λέξεις που ανήκουν στην ίδια γραμμή κειμένου. Τέλος, συνεχίζεται η σάρωση και επαναλαμβάνεται η ίδια διαδικασία για τον εντοπισμό των επόμενων γραμμών κειμένου μέχρι όλες οι λέξεις της εικόνας να ανατεθούν σε κάποια γραμμή κειμένου (Σχήμα 4.1(δ)).

Η διαδικασία σύνδεσης δύο γειτονικών λέξεων είναι η ακόλουθη: Έστω ότι (x_1, y_1) , (x_2, y_2) προσδιορίζουν τις συντεταγμένες του πλαισίου της πρώτης λέξης που έχει εντοπιστεί και (x'_1, y'_1) , (x'_2, y'_2) τις συντεταγμένες του πλαισίου της υποψήφιας προς σύνδεσης γειτονικής λέξης. Απ' όλες τις λέξεις που βρίσκονται δεξιά από τη λέξη που έχει εντοπιστεί και ικανοποιούν τη συνθήκη $[y_1, y_2] \cap [y'_1, y'_2] \neq \emptyset$ (δηλαδή έχουν οριζόντια επικάλυψη), επιλέγεται εκείνη η λέξη με τη μικρότερη απόσταση $D = x'_1 - x_2$, μόνο εάν $0 < D < 5 * AH$. Από τη στιγμή που πολλές λέξεις μπορούν να ικανοποιούν τη συνθήκη της οριζόντιας επικάλυψης, επιλέγοντας εκείνη με τη μικρότερη απόσταση εξασφαλίζουμε ότι θα είναι ο άμεσος γείτονας που ανήκει στην ίδια γραμμή κειμένου. Στη συνέχεια, η λέξη αυτή ορίζεται ως λέξη της γραμμής κειμένου και συνεχίζουμε την αναζήτηση στη δεξιά κατεύθυνση μέχρι να μην υπάρχει άλλη υποψήφια προς σύνδεση γειτονική λέξη. Παρόμοια διαδικασία ακολουθείται και για τη σύνδεση των λέξεων στην αριστερή διεύθυνση.

Figure 6(a) shows the word **Laboratory** and the positions of the upward concavities. The a's show two upward concavities near the baseline, while the b shows one concavity near the baseline and one well above the baseline. Figure 6(b) shows the positions of upward concavities in a single Kanji character.

(α)

Figure 6(a) shows the word **Laboratory** and the positions of the upward concavities. The a's show two upward concavities near the baseline, while the b shows one concavity near the baseline and one well above the baseline. Figure 6(b) shows the positions of upward concavities in a single Kanji character.

(β)

Figure 6(a) shows the word **Laboratory** and the positions of the upward concavities. The a's show two upward concavities near the baseline, while the b shows one concavity near the baseline and one well above the baseline. Figure 6(b) shows the positions of upward concavities in a single Kanji character.

(γ)

Figure 6(a) shows the word **Laboratory** and the positions of the upward concavities. The a's show two upward concavities near the baseline, while the b shows one concavity near the baseline and one well above the baseline. Figure 6(b) shows the positions of upward concavities in a single Kanji character.

(δ)

Σχήμα 4.1: Κατάτμηση λέξεων και γραμμών κειμένου. (α) Αρχικό παραμορφωμένο έγγραφο, (β) εξομάλυνση της εικόνας, (γ) οι λέξεις και (δ) οι γραμμές κειμένου που εντοπίστηκαν. Η πρώτη λέξη κάθε γραμμής κειμένου που εντοπίζεται κατά την σάρωση προσδιορίζεται με χρωματισμένο πλαίσιο και τα βέλη δείχνουν τις γειτονικές λέξεις που ανήκουν στην ίδια γραμμή κειμένου.

4.2.2 Προσδιορισμός Γραμμών Αναφοράς των Λέξεων

Στο στάδιο αυτό προσδιορίζονται η κάτω και η επάνω γραμμή αναφοράς των λέξεων, οι οποίες ορίζουν το κυρίως σώμα της λέξης. Ξεκινώντας από την εικόνα που προκύπτει μετά την εξομάλυνση (Σχήμα 4.1(β)) χρησιμοποιούμε τη μέθοδο που περιγράφεται στο [132]. Εφαρμόζει γραμμική παλινδρόμηση στο σύνολο των pixels κειμένου της λέξης που βρίσκονται πιο χαμηλά, ή αντίστοιχα πιο ψηλά, σε κάθε στήλη της λέξης (Σχήμα 4.2). Μετά από αυτή τη διαδικασία, η επάνω και η κάτω γραμμή αναφοράς της λέξης W_i ορίζονται ως εξής:

$$y = a_i^u x + b_i^u \quad \& \quad y = a_i^l x + b_i^l$$



Σχήμα 4.2: Προσδιορισμός της επάνω και της κάτω γραμμής αναφοράς.

Επιπλέον, υπολογίζουμε τη γωνία κλίσης της επάνω (θ_i^u) και της κάτω (θ_i^l) γραμμής αναφοράς ως εξής:

$$\theta_i^u = \arctan(a_i^u) \quad \& \quad \theta_i^l = \arctan(a_i^l)$$

Κάθε λέξη W_i θα πρέπει να προσδιορίζεται από μία μόνο γραμμή αναφοράς (b_i, θ_i) . Η μικρότερη γωνία κλίσης είναι συνήθως και η πιο αντιπροσωπευτική, συνεπώς, για τη λέξη που βρίσκεται αριστερά σε κάθε γραμμή κειμένου η γραμμή αναφοράς ορίζεται ως εξής:

$$(b_i, \theta_i) = \begin{cases} (b_i^u, \theta_i^u), & \text{εάν } |\theta_i^u| < |\theta_i^l| \\ (b_i^l, \theta_i^l), & \text{διαφορετικά} \end{cases}$$

ενώ για τις υπόλοιπες λέξεις κάθε γραμμής κειμένου, η γραμμή αναφοράς επιλέγεται σε σχέση με την προηγούμενη λέξη ώστε να έχουν τη μικρότερη διαφορά ως προς την κλίση ως εξής:

$$(b_i, \theta_i) = \begin{cases} (b_i^u, \theta_i^u), & \text{εάν } |\theta_i^u - \theta_{i-1}^u| < |\theta_i^l - \theta_{i-1}^l| \\ (b_i^l, \theta_i^l), & \text{διαφορετικά} \end{cases}$$

4.2.3 Διόρθωση Παραμορφώσεων

Αυτό είναι το τελευταίο βήμα της μεθόδου στο οποίο κάθε λέξη $W_i(x, y)$ περιστρέφεται και μετατοπίζεται σύμφωνα με τη γραμμή αναφοράς ώστε να διορθωθούν οι παραμορφώσεις της εικόνας. Αρχικά, περιστρέφονται όλες οι λέξεις σύμφωνα με τη γωνία κλίσης τους (θ_i):

$$(x^r, y^r) = (x, (x - x_i^{\min}) * \sin(-\theta_i) + y * \cos(\theta_i))$$

όπου $W_i^r(x^r, y^r)$ είναι η λέξη που προκύπτει μετά την περιστροφή και x_i^{\min} είναι το αριστερότερο σημείο της λέξης $W_i(x, y)$. Μετά την περιστροφή, όλες οι λέξεις είναι παράλληλες ως προς την οριζόντια διεύθυνση αλλά δεν είναι ευθυγραμμισμένες ως προς τη γραμμή κειμένου (Σχήμα 4.3(β)). Συνεπώς πρέπει να μετατοπιστούν ως προς την κατακόρυφη διεύθυνση. Η κάθετη μετατόπιση γίνεται σε σχέση με την αριστερή λέξη κάθε γραμμής κειμένου, η οποία δε μετατοπίζεται σε αυτό το στάδιο. Κάθε λέξη μετατοπίζεται με στόχο η γραμμή αναφοράς της (b_i, θ_i) να είναι ευθυγραμμισμένη με τη γραμμή αναφοράς της αριστερής λέξης που ανήκει στην ίδια γραμμή κειμένου (Σχήμα 4.3(γ)). Η κατακόρυφη μετατόπιση d_i της λέξης

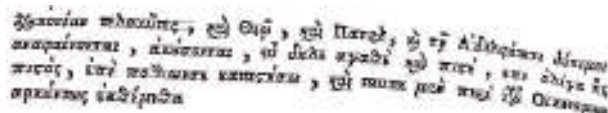
$W_i^r(x^r, y^r)$ υπολογίζεται ως εξής:

$$d_i = \begin{cases} y_0^{ur} - y_i^r, & \text{εάν } (b_i, \theta_i) \equiv (b_i^u, \theta_i^u) \\ y_0^{lr} - y_i^r, & \text{διαφορετικά} \end{cases}$$

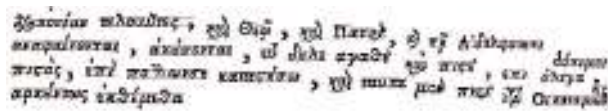
όπου οι παράμετροι $y_0^{ur} = (a_0^u x_0^{\min} + b_0^u) * \cos(\theta_0^u)$, $y_0^{lr} = (a_0^l x_0^{\min} + b_0^l) * \cos(\theta_0^l)$ αντιπροσωπεύουν την πάνω και κάτω γραμμή αναφοράς της αριστερής λέξης ως προς την οποία γίνεται η ευθυγράμμιση, και $y_i^r = (a_i x_i^{\min} + b_i) * \cos(\theta_i)$ αντιπροσωπεύει τη λέξη $W_i^r(x^r, y^r)$ που μετατοπίζεται. Για την αριστερή λέξη χρησιμοποιούμε και τις δύο γραμμές αναφοράς καθώς κάθε λέξη που μετατοπίζεται μπορεί να χρησιμοποιεί για ευθυγράμμιση είτε την πάνω είτε την κάτω γραμμή αναφοράς. Συνεπώς, μετά τον υπολογισμό της περιστροφής και της κάθετης μετατόπισης της λέξης $W_i^r(x^r, y^r)$ η τελική αποκατάσταση γίνεται ως εξής:

$$(x^{r'}, y^{r'}) = (x, y^r + d_i)$$

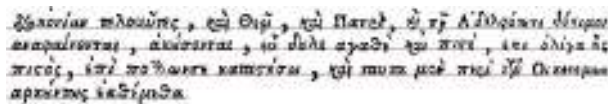
Τέλος, προσθέτουμε στην εικόνα τα συνδεδεμένα συστατικά που δεν ανήκουν στο κείμενο και αφαιρέθηκαν στην αρχή της μεθόδου (§ 4.2.1). Κάθε pixel κειμένου του συνδεδεμένου συστατικού κληρονομεί τους συντελεστές μεταφοράς $(x^{r'}, y^{r'})$ από το πλησιέστερο pixel, ως προς τον οριζόντιο άξονα, που συμμετείχε στη διαδικασία.



(α)



(β)



(γ)

Σχήμα 4.3: Διόρθωση των παραμορφώσεων: (α) αρχικό παραμορφωμένο έγγραφο στο οποίο επισημαίνονται οι γραμμές αναφοράς των λέξεων και το αποτέλεσμα που προκύπτει μετά (β) την περιστροφή και (γ) την κατακόρυφη μετατόπιση των λέξεων.

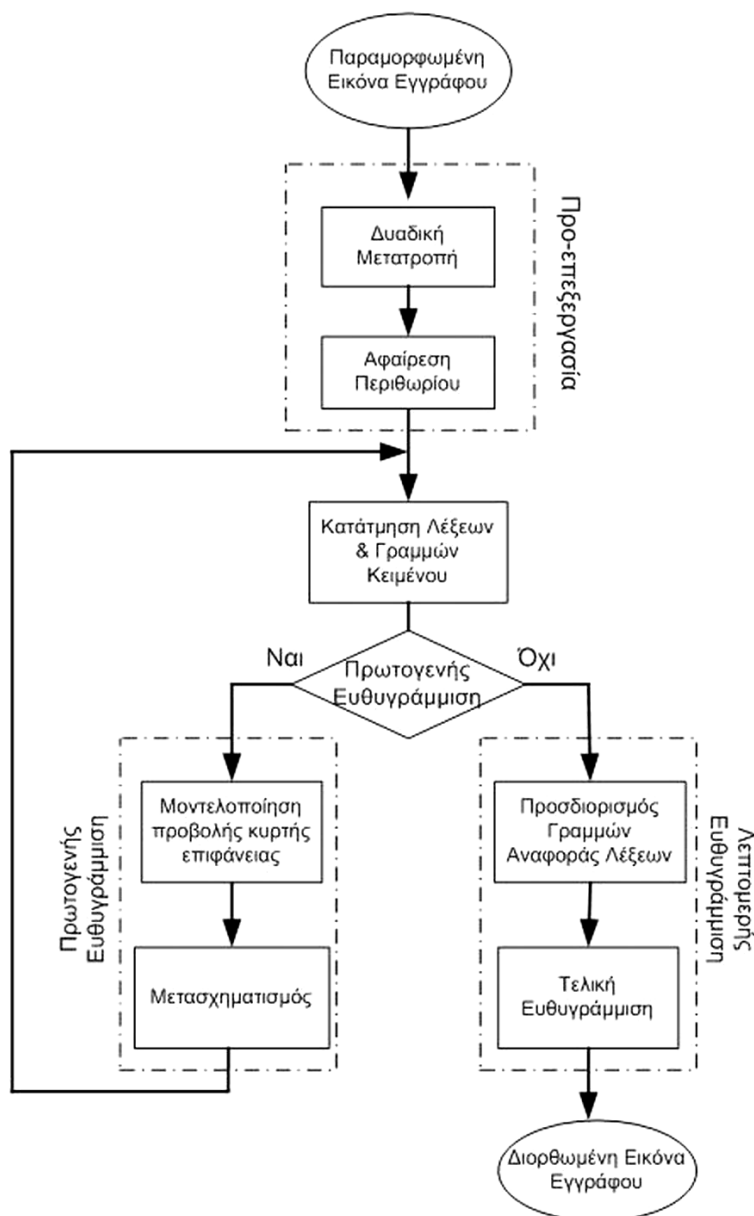
4.3 Διόρθωση της Τοπικής Καμπυλότητας σε Δύο Φάσεις

Σε αυτό το κεφάλαιο παρουσιάζεται η προτεινόμενη μέθοδος, η οποία από εδώ και στο εξής θα αναφέρεται ως TSD.v1 (Two-Step Dewarping). Επιπλέον, παρουσιάζεται και η βελτιωμένη προσέγγιση η οποία από εδώ και στο εξής θα αναφέρεται ως TSD.v2.

Η πρώτη φάση (πρωτογενής ευθυγράμμιση - coarse rectification) έχει ως στόχο να διορθώσει τις έντονες παραμορφώσεις του εγγράφου και να παράγει ένα πρώτο αποτέλεσμα διόρθωσης. Στη συνέχεια, η δεύτερη φάση (λεπτομερείς ευθυγράμμιση - fine rectification) βασίζεται στη μέθοδο [131] που περιγράψαμε προηγουμένως και διορθώνει τις τοπικές παραμορφώσεις του εγγράφου σε επίπεδο λέξης παράγοντας το τελικό διορθωμένο αποτέλεσμα απαλλαγμένο από όλες τις παραμορφώσεις. Σε σύγκριση με τις τεχνικές τρέχουσας τεχνολογικής στάθμης [45], [46], [47], [48] και [54], οι οποίες βασίζονται σε συγκεκριμένα πρότυπα και παραμέτρους που τις δυσκολεύουν να γενικευθούν ή χρησιμοποιούν πολύπλοκους υπολογισμούς, η προτεινόμενη μέθοδος χρησιμοποιεί ένα χαμηλής πολυπλοκότητας μετασχηματισμό, ο οποίος δε βασίζεται σε συγκεκριμένα μοντέλα. Αν και η μέθοδος προϋποθέτει το κείμενο να είναι στοιχισμένο και να μην περιέχει συνεχόμενη γραφή, ώστε να μπορούν να εντοπιστούν οι λέξεις, μπορεί να αντιμετωπίσει με επιτυχία έγγραφα που περιέχουν ανομοιογενές διαστήματα ανάμεσα στις γραμμές κειμένου ή περιοχές που δεν ανήκουν στο κείμενο. Πρέπει να σημειωθεί ότι η προτεινόμενη μέθοδος εφαρμόζεται σε μονόστηλα έγγραφα. Συνεπώς, σε περιπτώσεις που το έγγραφο περιέχει δύο ή περισσότερες στήλες θεωρούμε ότι έχει εφαρμοστεί προηγουμένως μία μέθοδος που εντοπίζει και ξεχωρίζει τις στήλες, όπως η τεχνική στο [133]. Το διάγραμμα ροής της μεθόδου παρουσιάζεται στο Σχήμα 4.4.

Όπως παρατηρούμε και στο διάγραμμα ροής, η κατάτμηση των γραμμών κειμένου και των λέξεων εφαρμόζεται πριν από κάθε φάση χρησιμοποιώντας τον αλγόριθμο που περιγράψαμε στην παράγραφο 4.2.1. Όμως, κάθε φάση έχει διαφορετική ανεκτικότητα στα λάθη. Η πρώτη φάση εφαρμόζει τον αλγόριθμο κατάτμησης στην αρχική παραμορφωμένη εικόνα, συνεπώς υπάρχει μεγάλη πιθανότητα να υπάρχουν λάθη, ειδικά όταν οι παραμορφώσεις είναι πολύ έντονες. Παρόλα αυτά, το συγκεκριμένο στάδιο δε βασίζεται στην ακρίβεια του αποτελέσματος κατάτμησης των γραμμών κειμένου και των λέξεων, καθώς χρειάζεται μόνο συγκεκριμένα σημεία για να μοντελοποιήσει την προβολή της

κυρτής επιφάνειας του εγγράφου. Για λόγους σαφήνειας, θα αναφέρουμε πιθανά λάθη στη διαδικασία κατάτμησης των γραμμών κειμένου και των λέξεων και πως αυτά επηρεάζουν την πρώτη φάση της μεθόδου. Στην περίπτωση που η μέθοδος σπάσει μία λέξη ή ενώσει δύο λέξεις που ανήκουν στην ίδια γραμμή κειμένου δε θα επηρεάσει τη συνέχεια της μεθόδου κατάτμησης καθώς οι λέξεις θα συνδεθούν κανονικά ως γειτονικές λέξεις (§ 4.2.1).



Σχήμα 4.4: Διάγραμμα ροής της προτεινόμενης μεθόδου για τη διόρθωση της τοπικής καμπυλότητας σε δύο φάσεις.

Από την άλλη, εάν ενωθούν δύο λέξεις που ανήκουν σε διαφορετικές γραμμές αυτό θα οδηγήσει σε λάθος κατάτμηση των γραμμών κειμένου. Όμως, η πρώτη φάση της μεθόδου δε θα επηρεαστεί καθώς χρησιμοποιεί μόνο τα αρχικά και τελικά σημεία κάθε γραμμής, επίσης, οι μικρές γραμμές κειμένου που πιθανόν να

προκύψουν από ένα τέτοιο λάθος αγνοούνται από τη μέθοδο (§ 4.3.1). Τέλος, εάν μία λέξη δεν εντοπιστεί καθόλου είναι πιθανόν να οδηγήσει σε σπάσιμο μίας γραμμής κειμένου σε δύο μικρότερες, οι οποίες, όπως είπαμε και προηγουμένως, θα αγνοηθούν και δε θα επηρεάσουν τη μέθοδο. Η δεύτερη φάση εφαρμόζεται μετά την αρχική διόρθωση του εγγράφου, συνεπώς οι έντονες παραμορφώσεις έχουν διορθωθεί και η κατάτμηση του εγγράφου σε γραμμές και λέξεις εφαρμόζεται με μεγαλύτερη επιτυχία.

Στο στάδιο της προ-επεξεργασίας εφαρμόζουμε, τη δυαδική μετατροπή [6], εάν η εικόνα είναι διαβάθμισης γκρι, και αφαίρεση του περιθωρίου χρησιμοποιώντας την τεχνική που αναπτύξαμε και περιγράψαμε στο προηγούμενο κεφάλαιο (§ 3.2).

4.3.1 Αρχική Προσέγγιση

(1) Πρωτογενής Ευθυγράμμιση (Coarse Rectification)

Στην πρώτη φάση εφαρμόζεται ένας χαμηλής υπολογιστικής πολυπλοκότητας μετασχηματισμός ο οποίος απεικονίζει την προβολή της κυρτής επιφάνειας του εγγράφου σε μία διδιάστατη ορθογώνια περιοχή με στόχο να παράγει ένα πρώτο αποτέλεσμα διόρθωσης απαλλαγμένο από τις έντονες παραμορφώσεις. Σε σύγκριση με τις τεχνικές τρέχουσας τεχνολογικής στάθμης [45], [48] και [54] οι οποίες προσπαθούν επίσης να μοντελοποιήσουν την επιφάνεια του παραμορφωμένου εγγράφου, η προτεινόμενη μέθοδος δε βασίζεται σε κανένα φυσικό πρότυπο ή γενικό μοντέλο (π.χ. κυλινδρικό μοντέλο). Οι τεχνικές [45] και [54] χρησιμοποιούν συγκεκριμένους τύπους παρεμβολής βασισμένες στο μοντέλο Gordon [56] και τη γραμμική παρεμβολή Coons [57], αντίστοιχα. Στην προτεινόμενη μέθοδο, δημιουργούμε μία αντιστοίχιση μεταξύ δύο καμπύλων γραμμών βάσει της οποίας γίνεται η απεικόνιση της προβολής της κυρτής επιφάνειας στη διδιάστατη ορθογώνια περιοχή. Στη συνέχεια περιγράφονται αναλυτικά όλα τα βήματα της μεθόδου.

Μοντελοποίηση της προβολής της κυρτής επιφάνειας

Θεωρούμε ότι η προβολή της κυρτής επιφάνειας οριοθετείται από δύο καμπύλες γραμμές, οι οποίες προσδιορίζονται από δύο γραμμές κειμένου στο ανώτερο και κατώτερο μέρος του εγγράφου, καθώς και δύο ευθείες γραμμές, οι οποίες προσδιορίζονται από το δεξιό και αριστερό όριο του εγγράφου. Στη συνέχεια, παραθέτουμε αναλυτικά τον αλγόριθμο για τον υπολογισμό όλων των γραμμών

που οριοθετούν την προβολή της κυρτής επιφάνειας. Ας θεωρήσουμε ότι τα σημεία $A(x_1, y_1)$, $B(x_2, y_2)$, $C(x_3, y_3)$ και $D(x_4, y_4)$ ορίζουν τις κορυφές της προβολής της κυρτής επιφάνειας (Σχήμα 4.5).

(α) Υπολογισμός Αριστερής & Δεξιάς Ευθείας Γραμμής

Στόχος μας είναι να υπολογίσουμε τις ευθείες γραμμές AD και BC οι οποίες προσδιορίζονται από το δεξιό και αριστερό όριο του εγγράφου.

Βήμα 1: Εντοπίζουμε τα αρχικά (x_{s_i}, y_{s_i}) και τελικά (x_{e_i}, y_{e_i}) σημεία κάθε γραμμής κειμένου, όπου $0 < i < NL$ και NL το πλήθος των γραμμών κειμένου.

Βήμα 2: Υπολογίζουμε τις μέσες τιμές X_s και X_e για τις τετμημένες όλων των αρχικών και τελικών σημείων, αντίστοιχα. Τα σημεία που ικανοποιούν τη συνθήκη $|x_{s_i} - X_s| > 2 \cdot AH$, ή $|x_{e_i} - X_e| > 2 \cdot AH$ αντίστοιχα, εξαιρούνται από τη διαδικασία, με στόχο να αποκλειστούν οι μικρές γραμμές που δεν αρχίζουν ή τελειώνουν στα όρια του εγγράφου (AH - κυρίαρχο ύψος γράμματος - § 4.2.1).

Βήμα 3: Χρησιμοποιώντας τα σημεία που έχουν απομείνει από το προηγούμενο βήμα υπολογίζουμε τις ευθείες AD και BC με τη μέθοδο ελαχίστων τετραγώνων (Least Square Estimation – LSE).

Συνεπώς, οι δύο ευθείες που προσδιορίζονται από το δεξιό και αριστερό όριο του εγγράφου ορίζονται ως εξής:

$$AD : y = a_l x + b_l \quad \& \quad BC : y = a_r x + b_r$$

(β) Υπολογισμός Ανώτερη & Κατώτερης Καμπύλης Γραμμής

Στόχος μας είναι να υπολογίσουμε τις καμπύλες γραμμές AB και DC οι οποίες προσδιορίζονται από δύο γραμμές κειμένου στο ανώτερο και κατώτερο μέρος του εγγράφου (Σχήμα 4.5).

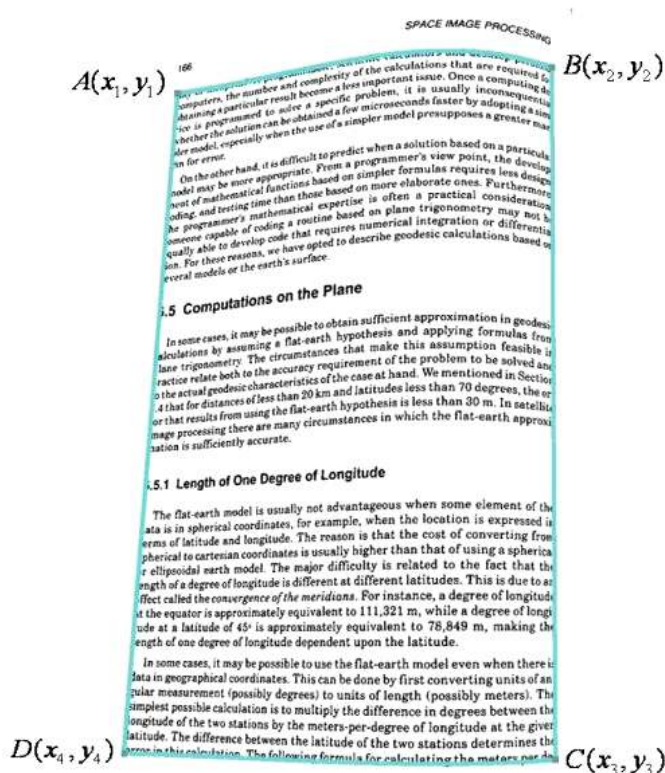
Βήμα 1: Επιλέγουμε τις πρώτες γραμμές κειμένου από το ανώτερο και κατώτερο μέρος του εγγράφου που ικανοποιούν τη συνθήκη: $(Dl + Dr) / 2 < 2 \cdot AH$, όπου Dl και Dr οι αποστάσεις του αρχικού και τελικού σημείου κάθε γραμμής κειμένου από τις ευθείες AD και BC , αντίστοιχα. Με αυτή τη συνθήκη επιτυγχάνουμε να επιλέξουμε αντιπροσωπευτικές γραμμές του εγγράφου που βρίσκονται κοντά στα όρια του και όχι μικρές γραμμές, όπως τίτλους. Αν καμιά γραμμή μέχρι το μέσο του εγγράφου δεν ικανοποιεί τη συνθήκη επιλέγουμε τη γραμμή που ελαχιστοποιεί το άθροισμα $(Dl + Dr)$.

Βήμα 2: Εντοπίζουμε τα ανώτερα σημεία (x_{u_i}, y_{u_i}) της επάνω γραμμής κειμένου και τα κατώτερα σημεία (x_{b_i}, y_{b_i}) της κάτω γραμμής κειμένου.

Βήμα 3: Χρησιμοποιώντας τα σημεία από το προηγούμενο βήμα υπολογίζουμε τις καμπύλες 3^{ου} βαθμού με τη μέθοδο ελαχίστων τετραγώνων.

Συνεπώς, οι δύο καμπύλες γραμμές που προσδιορίζονται από δύο γραμμές κειμένου στο ανώτερο και κατώτερο μέρος του εγγράφου ορίζονται ως εξής:

$$AB : y = a_{u1}x^3 + a_{u2}x^2 + a_{u3}x + a_{u4} \quad \& \quad DC : y = a_{b1}x^3 + a_{b2}x^2 + a_{b3}x + a_{b4}$$



Σχήμα 4.5: Μοντελοποίηση της προβολής της κυρτής επιφάνειας.

Μοντέλο Μετασχηματισμού

Αφού έχει μοντελοποιηθεί η προβολή της κυρτής επιφάνειας του εγγράφου στόχος μας είναι να παράγουμε ένα μετασχηματισμό ο οποίος θα την απεικονίζει σε μία διδιάστατη ορθογώνια περιοχή. Ας θεωρήσουμε ότι τα σημεία $A'(x'_1, y'_1)$, $B'(x'_2, y'_2)$, $C'(x'_3, y'_3)$ και $D'(x'_4, y'_4)$ ορίζουν τις κορυφές της ορθογώνιας περιοχής (Σχήμα 4.6). Επίσης, $|\widehat{AB}|$ και $|AB|$ συμβολίζουν το μήκος του τόξου \widehat{AB} και την Ευκλείδεια απόσταση μεταξύ των σημείων A και B , αντίστοιχα.

Βήμα 1: Προσδιορίζουμε την ορθογώνια περιοχή. Η μία κορυφή της συμπίπτει με την κορυφή της προβολής της κυρτής επιφάνειας ($A'(x'_1, y'_1) \equiv A(x_1, y_1)$) και οι υπόλοιπες κορυφές υπολογίζονται σύμφωνα με το πλάτος W και το ύψος H της ορθογώνιας περιοχής:

$$W = \min(|\widehat{AB}|, |\widehat{DC}|) \quad \& \quad H = \min(|AD|, |BC|)$$

Βήμα 2: Στόχος μας είναι κάθε σημείο της προβολής της κυρτής επιφάνειας $O(x, y)$ να απεικονιστεί στο αντίστοιχο σημείο $O'(x', y')$ της ορθογώνιας περιοχής. Κάθε σημείο $O(x, y)$ προσδιορίζεται από δύο σημεία που βρίσκονται πάνω στις δύο καμπύλες γραμμές AB και DC και ικανοποιούν την συνθήκη:

$$\frac{|\widehat{AE}|}{|\widehat{AB}|} = \frac{|\widehat{DG}|}{|\widehat{DC}|}$$

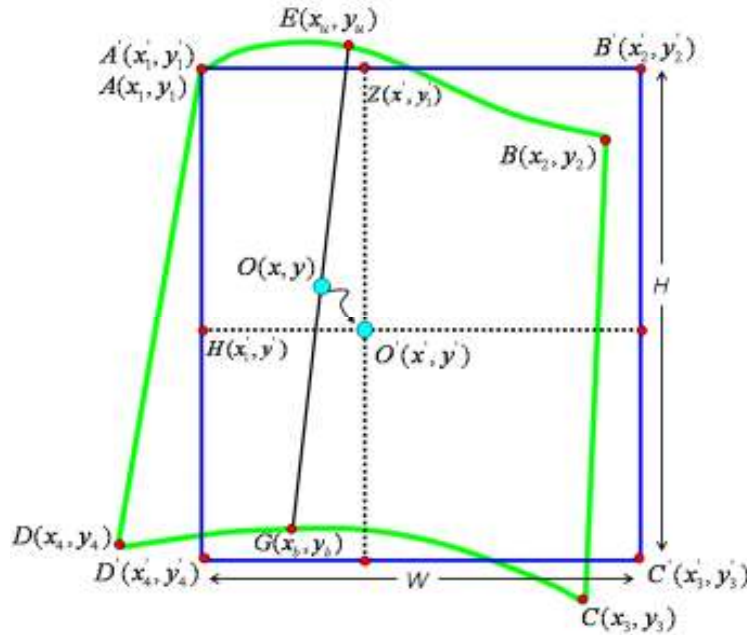
Το αντίστοιχο σημείο $O'(x', y')$ υπολογίζεται διατηρώντας το λόγο ανάμεσα στην προβολή της κυρτής επιφάνειας και την ορθογώνια περιοχή τόσο στην κάθετη διεύθυνση όσο και στην οριζόντια. Βασιζόμενοι σε αυτό τον στόχο, τα σημεία $Z(x'_1, y'_1)$ και $H(x'_1, y'_1)$, που ουσιαστικά προσδιορίζουν το σημείο $O'(x', y')$ (Σχήμα 4.6) υπολογίζονται ως εξής:

$$\frac{|\widehat{AB}|}{|\widehat{AE}|} = \frac{W}{|A'Z|} \Rightarrow |A'Z| = \frac{W}{|\widehat{AB}|} |\widehat{AE}|$$

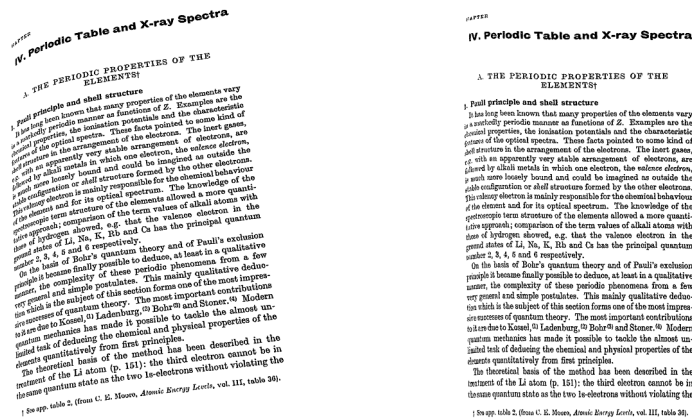
$$\frac{|EG|}{|EO|} = \frac{H}{|A'H|} \Rightarrow |A'H| = \frac{H}{|EG|} |EO|$$

Βήμα 3: Όσα σημεία της εικόνας δεν περικλείονται στην προβολή της κυρτής επιφάνειας κληρονομούν το μετασχηματισμό από το πλησιέστερο σημείο.

Ένα παράδειγμα εφαρμογής της πρώτης φάσης της μεθόδου παρουσιάζεται στο Σχήμα 4.7. Παρατηρούμε ότι οι έντονες παραμορφώσεις του εγγράφου έχουν διορθωθεί και έχει επιτευχθεί ένα πρώτο, βελτιωμένο αποτέλεσμα στο οποίο είναι ευκολότερο να εφαρμοστεί η κατάτμηση των γραμμών κειμένου. Επιπλέον, η κάθετη ευθυγράμμιση του εγγράφου έχει αποκατασταθεί.



Σχήμα 4.6: Μοντέλο μετασχηματισμού.



(α) (β)

Σχήμα 4.7 (α) Αρχικό παραμορφωμένο έγγραφο, (β) Έγγραφο μετά την εφαρμογή της πρώτης φάσης (πρωτογενής ευθυγράμμιση).

(2) Λεπτομερείς Ευθυγράμμιση (Fine Rectification)

Η δεύτερη φάση βασίζεται στη μέθοδο SBD [131] (§ 4.2) και διορθώνει τις τοπικές παραμορφώσεις του εγγράφου σε επίπεδο λέξης παράγοντας το τελικό διορθωμένο αποτέλεσμα απαλλαγμένο απ' όλες τις παραμορφώσεις. Η βασική διαφορά με τη μέθοδο SBD έγκειται στο γεγονός ότι τώρα εφαρμόζεται στο αποτέλεσμα του πρώτου σταδίου (Σχήμα 4.7), συνεπώς έχει μεγαλύτερες πιθανότητες να έχει καλύτερο αποτέλεσμα στο στάδιο της κατάμησης των γραμμών κειμένου.

Επιπλέον, σε σύγκριση με τη μέθοδο SBD, όταν προσθέτουμε στην εικόνα τα συνδεδεμένα συστατικά που δεν ανήκουν στο κείμενο (§ 4.2.1) κάθε ρixel κειμένου ενός συνδεδεμένου συστατικού δεν κληρονομεί τους συντελεστές μεταφοράς από το πλησιέστερο ρixel. Εφαρμόζεται ένα ενιαίος συντελεστής σε ολόκληρο το συνδεδεμένο συστατικό ο οποίος προκύπτει από το μέσο όρο των συντελεστών κάθε ρixel. Έτσι, τα συνδεδεμένα συστατικά που προστίθενται είναι ενιαία χωρίς σπασίματα (Σχήμα 4.8).



Σχήμα 4.8: Παράδειγμα προσθήκης συνδεδεμένου συστατικού στο τελικό αποτέλεσμα με χρήση της μεθόδου (α) SBD και (β) TSD.v1. Οι κύκλοι δείχνουν τα σπασίματα που προκύπτουν από τη μέθοδο SBD.

4.3.2 Βελτιωμένη Προσέγγιση

(1) Πρωτογενής Ευθυγράμμιση (Coarse Rectification)

Πειραματικά αποτελέσματα έδειξαν ότι όταν το έγγραφο, εκτός από τις παραμορφώσεις λόγω τοπικής καμπυλότητας, περιέχει και στροφή η μοντελοποίηση της προβολής της κυρτής επιφάνειας μπορεί να αποτύχει λόγω λανθασμένου υπολογισμού των ευθειών AD και BC . Αυτό οφείλεται στο γεγονός ότι ο υπολογισμός τους βασίζεται στο μέσο όρο των τετμημένων των αρχικών και τελικών σημείων κάθε γραμμής κειμένου (Σχήμα 4.9(α)). Συνεπώς, (i) προσαρμόσαμε τον υπολογισμό των ευθειών χρησιμοποιώντας τον μήκος των γραμμών κειμένου έτσι ώστε να μην επηρεάζεται από τη στροφή του εγγράφου. Επιπλέον, (ii) προσθέσαμε μία επαναληπτική διαδικασία για ακόμα πιο ακριβή αποτελέσματα (Σχήμα 4.9(β)). Η νέα διαδικασία υπολογισμού των ευθειών AD και BC περιγράφεται στη συνέχεια:

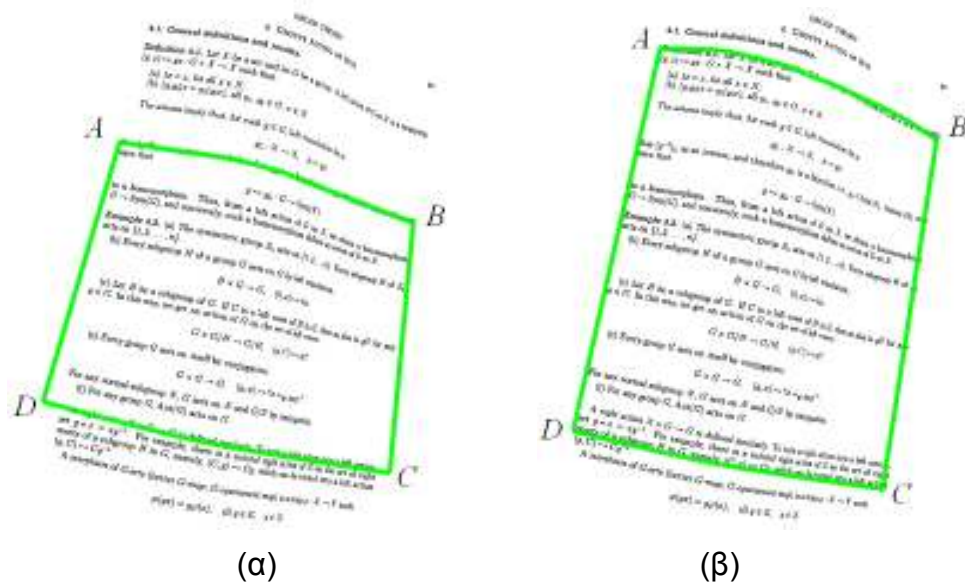
Βήμα 1: Εντοπίζουμε τα αρχικά (x_{s_i}, y_{s_i}) και τελικά (x_{e_i}, y_{e_i}) σημεία κάθε γραμμής κειμένου, όπου $0 < i < NL$ και NL το πλήθος των γραμμών κειμένου.

Βήμα 2: Υπολογίζουμε το μέσο μήκος των γραμμών κειμένου AL και κάθε γραμμή με μήκος μικρότερο ή ίσο του $0.8 * AL$ αποκλείεται από τη συνέχεια της διαδικασίας. Με αυτό τον τρόπο εξαιρούνται όλες οι μικρές γραμμές κειμένου όπως τίτλοι, υπότιτλοι, κείμενο περιθωρίου, μαθηματικές εκφράσεις κ.λπ.

Βήμα 3: Χρησιμοποιώντας τα σημεία που έχουν απομείνει από το προηγούμενο βήμα υπολογίζουμε τις ευθείες AD και BC με τη μέθοδο ελαχίστων τετραγώνων.

Βήμα 4: Μετά το δεύτερο βήμα, αν και οι μικρές γραμμές κειμένου έχουν αποκλειστεί, μπορεί να έχουν παραμείνει γραμμές κειμένου που δεν αρχίζουν ή τελειώνουν στα όρια του εγγράφου, όπως για παράδειγμα η πρώτη γραμμή κάθε παραγράφου. Γι' αυτό το λόγο εφαρμόζουμε μία επαναληπτική διαδικασία. Εάν η απόκλιση της ευθείας (AD ή BC) από τα σημεία των γραμμών (αρχικά ή τελικά) είναι μεγαλύτερη από το κυρίαρχο ύψος γράμματος αποκλείουμε το σημείο εκείνο που απέχει περισσότερο και υπολογίζουμε ξανά την ευθεία. Η επαναληπτική διαδικασία σταματάει όταν ικανοποιηθεί η παραπάνω συνθήκη ή μείνουν μόνο δύο σημεία. Από τη στιγμή που η επαναληπτική διαδικασία χρησιμοποιεί γραμμές κειμένου με μεγάλο μήκος η πιθανή απόκλιση της ευθείας δεν θα είναι μεγάλη. Συνεπώς, ακόμα και στη χειρότερη περίπτωση, μπορεί η μοντελοποίηση να μην είναι ακριβής ώστε να παραχθεί το καλύτερο δυνατό αποτέλεσμα στην πρώτη φάση, αλλά τουλάχιστον η μέθοδος δε θα αποτύχει.

Με τη νέα τεχνική υπολογισμού των ευθειών AD και BC οι γραμμές κειμένου που τελικά συμμετέχουν στη διαδικασία υπολογισμού είναι αντιπροσωπευτικές του εγγράφου. Συνεπώς, για τον υπολογισμό των καμπύλων γραμμών AB και DC επιλέγουμε την πρώτη και την τελευταία γραμμή κειμένου που συμμετέχουν στη διαδικασία (Σχήμα 4.9).



Σχήμα 4.9: Μοντελοποίηση της προβολής της κυρτής επιφάνειας σε έγγραφο με στροφή χρησιμοποιώντας τη μέθοδο (α) TSD.v1 και (β) TSD.v2.

(2) Λεπτομερείς Ευθυγράμμιση (Fine Rectification)

Όσο αναφορά τη δεύτερη φάση, παρατηρήθηκε ότι δεν επαρκεί ο κανόνας σύμφωνα με τον οποίον η επιλογή ανάμεσα στην πάνω και κάτω γραμμή αναφοράς της λέξης εξαρτάται από τη μικρότερη γωνία κλίσης ή τη σχέση της με την προηγούμενη λέξη καθώς μπορεί να έχουμε έντονες παραμορφώσεις μέσα στην ίδια γραμμή κειμένου (§ 4.2.2).

Με αφορμή αυτή την παρατήρηση, για κάθε γραμμή αναφοράς, πάνω και κάτω, εκτός από τη γωνία κλίσης, υπολογίζουμε και την απόκλιση S_i^u και S_i^l ως εξής:

$$S_i^u = \sum_{j=1}^n |y_j^u - (a_i^u x_j + b_i^u)| \quad \& \quad S_i^l = \sum_{j=1}^n |y_j^l - (a_i^l x_j + b_i^l)|$$

όπου n είναι ο αριθμός των στηλών της λέξης W_i , (x_j, y_j^u) και (x_j, y_j^l) ορίζουν τη θέση του ανώτερου και κατώτερου, αντίστοιχα, pixel κειμένου στη j στήλη της λέξης W_i . Η γραμμή αναφοράς με τη μικρότερη γωνία κλίσης και τη μικρότερη απόκλιση είναι συνήθως η πιο αντιπροσωπευτική. Συνεπώς, η γραμμή αναφοράς για κάθε λέξη W_i ορίζεται ως εξής:

$$(b_i, \theta_i, S_i) = \begin{cases} (b_i^u, \theta_i^u, S_i^u), & \text{εάν } (|\theta_i^u| < |\theta_i^l|) \text{ και } (S_i^u < S_i^l) \\ (b_i^l, \theta_i^l, S_i^l), & \text{διαφορετικά} \end{cases}$$

Παραδείγματα προσδιορισμού της γραμμής αναφοράς με τη νέα μέθοδο απεικονίζονται στο Σχήμα 4.10.

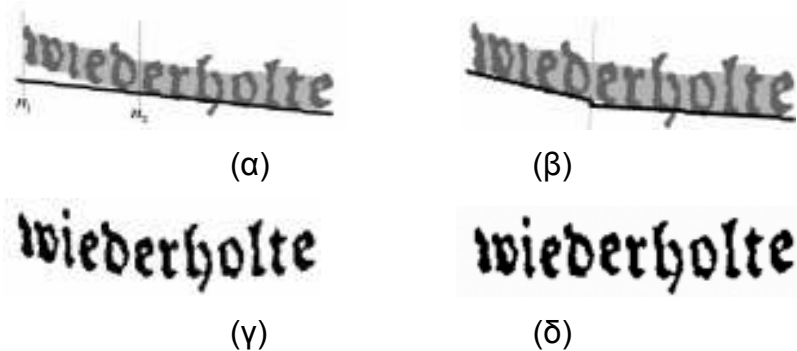


Σχήμα 4.10: Προσδιορισμός της γραμμής αναφοράς σε λέξεις με τη μέθοδο TSD.v2. Η γραμμή αναφοράς που επιλέγεται απεικονίζεται με μαύρο χρώμα.

Σε πολλές περιπτώσεις, η γραμμή αναφοράς δε μπορεί να προσδιοριστεί σωστά λόγω στρεβλώσεων που υπάρχουν μέσα στην ίδια τη λέξη (Σχήμα 4.11(α)). Συνεπώς, μόνο η περιστροφή και η κατακόρυφη μετατόπιση της λέξης δεν επαρκούν για την πλήρη αποκατάσταση των παραμορφώσεων (Σχήμα 4.11(γ)). Για αυτό το λόγο, διαιρούμε τη λέξη και επεξεργαζόμαστε κάθε τμήμα της ξεχωριστά. Για να αποφασίσουμε αν θα διαιρέσουμε μία λέξη W_i ελέγχουμε αν ικανοποιεί την εξής συνθήκη:

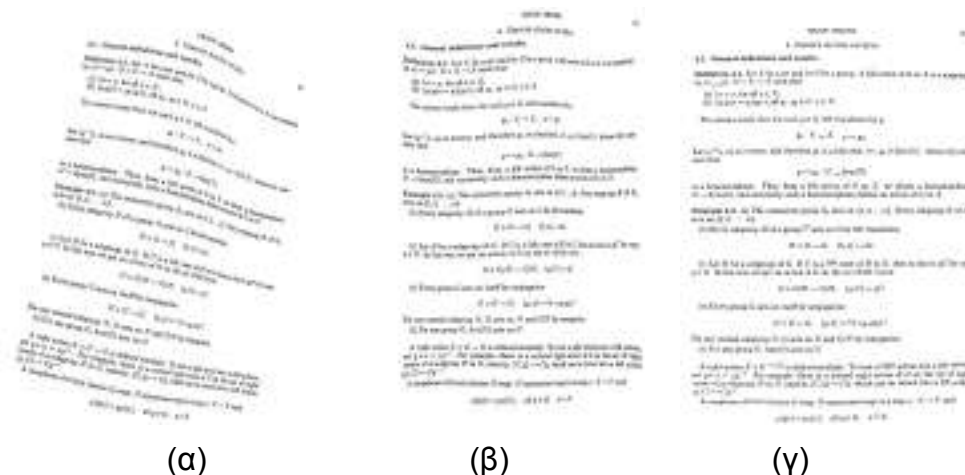
$$\exists n_1, n_2 : \prod_{j=n_1}^{n_2} |y_j - (a_i x_j + b_i)| \neq 0 \text{ και } (n_2 - n_1) \geq 2 * AH$$

όπου (x_j, y_j) ορίζει τη θέση του ανώτερου (ή κατώτερου) pixel κειμένου στη j στήλη της λέξης W_i . Σύμφωνα με το κριτήριο αυτό, διαιρούμε μία λέξη όταν η γραμμή αναφοράς αποκλίνει από τη λέξη για διάστημα μεγαλύτερο από το διπλάσιο του κυρίαρχου ύψους του γράμματος (δηλαδή για περισσότερο από δύο χαρακτήρες) (Σχήμα 4.11(α)). Στη συνέχεια, διαιρούμε τη λέξη στο σημείο από το οποίο η γραμμή αναφοράς αρχίζει να αποκλίνει και υπολογίζουμε τη νέα γραμμή αναφοράς για κάθε τμήμα ξεχωριστά. Στο επόμενο βήμα (§ 4.2.3), κάθε τμήμα θα περιστραφεί και θα μετατοπιστεί ανεξάρτητα πετυχαίνοντας με αυτό τον τρόπο την πλήρη αποκατάσταση των παραμορφώσεων (Σχήμα 4.11(δ)).



Σχήμα 4.11: Προσδιορισμός της γραμμής αναφοράς σε λέξη με στρεβλώσεις. (α) αρχική γραμμή αναφοράς όπου n_1 και n_2 καθορίζουν το διάστημα μέσα στο οποίο αποκλίνει από τη λέξη, (β) γραμμές αναφοράς μετά τον διαχωρισμό της λέξης, (γ)-(δ) περιστροφή της λέξης σύμφωνα με τη γραμμή αναφοράς στο (α) και (β), αντίστοιχα.

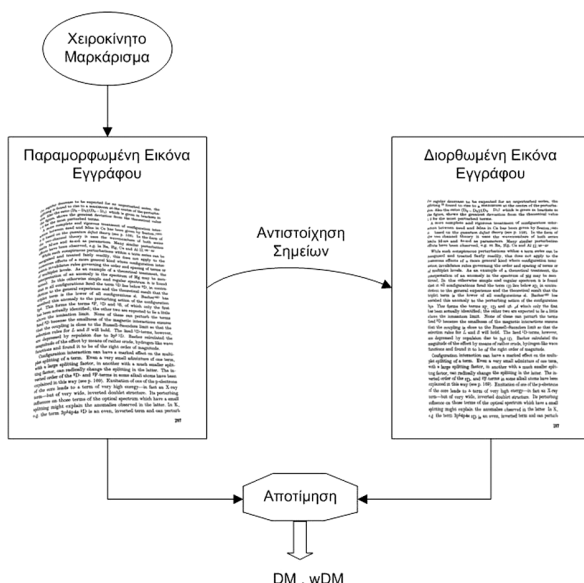
Τέλος, σύμφωνα με τη μέθοδο TSD.v1, στο τελικό στάδιο αποκατάστασης των παραμορφώσεων όλες οι λέξεις ευθυγραμμίζονται με την αντίστοιχη αριστερή λέξη της γραμμής κειμένου. Σε πολλές περιπτώσεις, ειδικά όταν το έγγραφο προέρχεται μετά από ψηφιοποίηση βιβλίου, οι παραμορφώσεις είναι ιδιαίτερα έντονες από τη μία μεριά του εγγράφου εξαιτίας της βιβλιοδεσίας. Αυτό έχει ως συνέπεια να μην πετυχαίνουμε το καλύτερο δυνατό αποτέλεσμα. Για τον σκοπό αυτό υπολογίζουμε τη μέση γωνία κλίσης για όλες τις λέξεις στην αριστερή και τη δεξιά μεριά του εγγράφου και επιλέγουμε να αποκαταστήσουμε την ευθυγράμμιση με τη μεριά εκείνη που έχει την μικρότερη τιμή, καθώς αυτή θα έχει και τις λιγότερες παραμορφώσεις. Το Σχήμα 4.12 παρουσιάζει ένα παράδειγμα εφαρμογής της μεθόδου TSD.v2.



Σχήμα 4.12: Παράδειγμα εφαρμογής της μεθόδου TSD.v2. (α) Αρχικό έγγραφο και το αποτέλεσμα μετά την εφαρμογή (β) της πρωτογενούς ευθυγράμμισης και (γ) της λεπτομερούς ευθυγράμμισης.

4.4 Ημιαυτόματη Μέθοδος Αξιολόγησης

Η προτεινόμενη μέθοδος αξιολόγησης λαμβάνει υπόψη την απόκλιση των γραμμών κειμένου στη διορθωμένη εικόνα από μία ευθεία, οριζόντια γραμμή αναφοράς η οποία θεωρείται το ιδανικό αποτέλεσμα. Αρχικά, ο χρήστης μαρκάρει κάποιες γραμμές κειμένου στην παραμορφωμένη εικόνα. Έπειτα, τα σημεία αυτά αντιστοιχίζονται με τα αντίστοιχα σημεία στη διορθωμένη εικόνα. Η αντιστοίχιση πραγματοποιείται με τη χρήση του μετασχηματισμού SIFT (Scale Invariant Feature Transform) [134]. Τέλος, με τη βοήθεια πολυωνύμων τρίτου βαθμού, υπολογίζεται η μετρική αξιολόγησης καμπυλότητας η οποία θεωρεί ότι το ιδανικό αποτέλεσμα θα πρέπει να αποτελείται μόνο από ευθείες, οριζόντιες γραμμές κειμένου. Το διάγραμμα ροής της μεθόδου παρουσιάζεται στο Σχήμα 4.13. Στη συνέχεια, περιγράφονται αναλυτικά όλα τα στάδια της μεθόδου.



Σχήμα 4.13: Διάγραμμα ροής της προτεινόμενης μεθόδου αξιολόγησης.

4.4.1 Μαρκάρισμα Εικόνας από τον Χρήστη

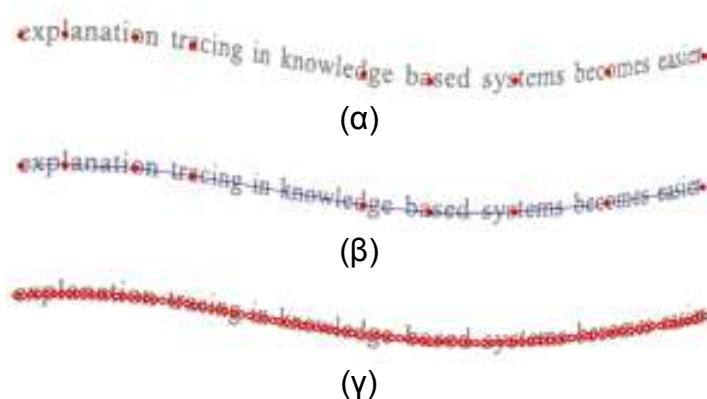
Στο πρώτο στάδιο της μεθόδου ο χρήστης επιλέγει και μαρκάρει N γραμμές κειμένου ορίζοντας διάφορα σημεία σε κάθε γραμμή. Το ιδανικό θα ήταν να μαρκάρει ο χρήστης όλες τις γραμμές κειμένου που περιέχει το έγγραφο, όμως κάτι τέτοιο δεν είναι πρακτικό. Συνεπώς, επιλέγονται N αντιπροσωπευτικές γραμμές κειμένου του παραμορφωμένου εγγράφου. Στο Σχήμα 4.14 απεικονίζεται ένα έγγραφο στο οποίο έχουν επιλεγεί έξι ($N = 6$) αντιπροσωπευτικές γραμμές κειμένου.



Σχήμα 4.14: Παράδειγμα επιλογής έξι αντιπροσωπευτικών γραμμών κειμένου σε ένα παραμορφωμένο έγγραφο.

Η διαδικασία μαρκαρίσματος είναι εύκολη και γρήγορη. Ο χρήστης, για να μαρκάρει μία γραμμή αρκεί να ορίσει μερικά μόνο σημεία εκεί όπου αλλάζει η καμπυλότητα της γραμμής. Πρέπει να σημειωθεί ότι τα σημεία ορίζονται στη μέση του κυρίως σώματος των λέξεων (Σχήμα 4.15(α)). Έπειτα, υπολογίζονται τα ευθύγραμμα τμήματα ανάμεσα στα ζευγάρια των σημείων που όρισε ο χρήστης (Σχήμα 4.15(β)) και σε κάθε ευθύγραμμο τμήμα εφαρμόζεται δειγματοληψία με βήμα πέντε pixels. Έτσι προκύπτει ένα σύνολο σημείων που αντιπροσωπεύει την επιλεγμένη γραμμή κειμένου στο παραμορφωμένο έγγραφο (Σχήμα 4.15(γ)).

Το στάδιο αυτό είναι το μοναδικό το οποίο χρειάζεται την επέμβαση του χρήστη. Όμως, αυτή η διαδικασία θα εφαρμοστεί μόνο μία φορά σε κάθε παραμορφωμένο έγγραφο. Στη συνέχεια, υπάρχει η δυνατότητα να αξιολογούνται διαφορετικές τεχνικές διόρθωσης της τοπικής καμπυλότητας χρησιμοποιώντας το ίδιο μαρκαρισμένο έγγραφο, με αποτέλεσμα να μπορούμε να έχουμε αξιόπιστες συγκρίσεις μεταξύ των διαφορετικών τεχνικών.



Σχήμα 4.15: Διαδικασία μαρκαρίσματος μίας γραμμής κειμένου: (α) σημεία που ορίζονται από τον χρήστη, (β) ευθύγραμμα τμήματα ανάμεσα στα ζευγάρια των σημείων και (γ) το τελικό σύνολο σημείων που αντιπροσωπεύει τη γραμμή.

4.4.2 Αντιστοίχιση Σημείων

Τα σημεία που έχουν προκύψει από το προηγούμενο βήμα αντιστοιχίζονται με τα αντίστοιχα σημεία στη διορθωμένη εικόνα, η οποία έχει προκύψει μετά από την εφαρμογή κάποιας τεχνικής διόρθωσης της τοπικής καμπυλότητας. Η αντιστοίχιση πραγματοποιείται με τη χρήση του μετασχηματισμού SIFT [134].

Ο μετασχηματισμός SIFT χρησιμοποιείται κυρίως στο εντοπισμό και την περιγραφή τοπικών χαρακτηριστικών σε εικόνες. Έχει χρησιμοποιηθεί σε πολλές εφαρμογές, όπως στην αναγνώριση αντικειμένων και χειρονομιών, στη ρομποτική πλοήγηση, στη συρραφή εικόνων, στην αναπαράσταση τριδιάστατων μοντέλων και στην αντιστοίχιση video. Μετασχηματίζει μία εικόνα σε μία μεγάλη συλλογή από τοπικά διανύσματα χαρακτηριστικών τα οποία είναι αναλλοίωτα στη μεταφορά, στην αλλαγή κλίμακας και στην περιστροφή καθώς και μερικώς αναλλοίωτα σε αλλαγές του φωτισμού της εικόνας. Αρχικά, εντοπίζει χαρακτηρίστηκα σημεία (key points) ψάχνοντας σε διαφορετικές κλίμακες της εικόνας για τοπικά ελάχιστα και μέγιστα της διαφοράς των εικόνων που έχει εφαρμοστεί το φίλτρο Gauss (difference-of-Gauss). Από κάθε key point εξάγεται ένα διάνυσμα χαρακτηριστικών το οποίο περιγράφει τοπικά την εικόνα. Τα διανύσματα χαρακτηριστικών που εξάγονται ονομάζονται SIFT keys.

Στην προτεινόμενη μέθοδο, αρχικά εξάγονται τα SIFT keys από την παραμορφωμένη και τη διορθωμένη εικόνα. Έπειτα, προσδιορίζεται η αντιστοιχία ανάμεσα στα SIFT keys των δύο εικόνων και με τη βοήθεια αυτής της πληροφορία αντιστοιχίζουμε τα μαρκαρισμένα σημεία με τα αντίστοιχα σημεία στη διορθωμένη εικόνα.

Έστω $M(x_m, y_m)$ ένα μαρκαρισμένο σημείο στην παραμορφωμένη εικόνα το οποίο θέλουμε να αντιστοιχίσουμε με το αντίστοιχο σημείο $M'(x'_m, y'_m)$ στη διορθωμένη εικόνα (Σχήμα 4.16). Εντοπίζουμε τα δύο πλησιέστερα key points $K_1(x_{k1}, y_{k1})$ και $K_2(x_{k2}, y_{k2})$ στην παραμορφωμένη εικόνα σύμφωνα με την Ευκλείδεια απόσταση. Έπειτα, με τη βοήθεια των αντίστοιχων key points στη διορθωμένη εικόνα, $K'_1(x'_{k1}, y'_{k1})$ και $K'_2(x'_{k2}, y'_{k2})$, προσδιορίζουμε το σημείο $M'(x'_m, y'_m)$ με γραμμική παρεμβολή:

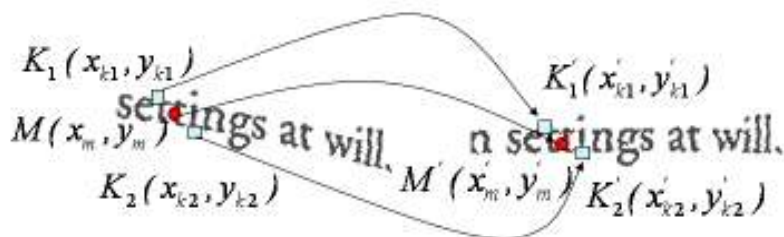
$$\left. \begin{aligned} x'_m &= x_m * a_x + b_x \\ y'_m &= y_m * a_y + b_y \end{aligned} \right\}$$

όπου οι συντελεστές a_x, b_x, a_y και b_y υπολογίζονται ως εξής:

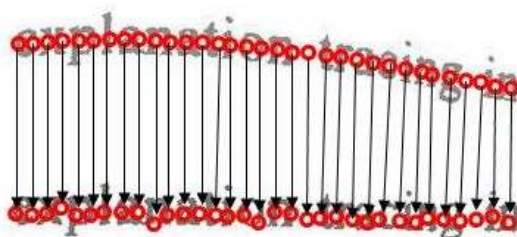
$$a_x = \begin{cases} \frac{x'_{k2} - x'_{k1}}{x_{k2} - x_{k1}} & \text{εάν } x_{k1} \neq x_{k2} \\ 1 & \text{διαφορετικά} \end{cases} \quad \& \quad b_x = x'_{k1} - x_{k1} * a_x$$

$$a_y = \begin{cases} \frac{y'_{k2} - y'_{k1}}{y_{k2} - y_{k1}} & \text{εάν } y_{k1} \neq y_{k2} \\ 1 & \text{διαφορετικά} \end{cases} \quad \& \quad b_y = y'_{k1} - y_{k1} * a_y$$

Το Σχήμα 4.17 απεικονίζει ένα παράδειγμα αντιστοίχισης σύμφωνα με την παραπάνω διαδικασία.



Σχήμα 4.16: Αντιστοίχιση σημείων με τη βοήθεια του μετασχηματισμού SIFT.



Σχήμα 4.17: Παράδειγμα αντιστοίχισης σημείων.

4.4.3 Μετρικές Αξιολόγησης

Στο τελικό στάδιο της μεθόδου υπολογίζεται η μετρική αξιολόγησης της καμπυλότητας η οποία αποδίδει την αποτελεσματικότητα μίας μεθόδου διόρθωσης της τοπικής καμπυλότητας. Λαμβάνει υπόψη την απόκλιση των γραμμών κειμένου στη διορθωμένη εικόνα από μία ευθεία, οριζόντια γραμμή αναφοράς η οποία θεωρείτε το ιδανικό αποτέλεσμα. Κάθε επιλεγμένη γραμμή στην παραμορφωμένη και τη διορθωμένη εικόνα προσεγγίζεται με πολυώνυμα τρίτου βαθμού. Η προσέγγιση γίνεται με μεγάλη ακρίβεια, ακόμα και αν η γραμμή πάσχει από πολλές και έντονες στρεβλώσεις, καθώς χρησιμοποιούνται πολλά διαδοχικά πολυώνυμα. Η μετρική αξιολόγησης της καμπυλότητας εξάγεται με τον υπολογισμό του ολοκληρώματος πολυωνύμου. Τα αναλυτικά βήματα για τον υπολογισμό της περιγράφονται στη συνέχεια.

Βήμα 1: Τα σημεία που αντιπροσωπεύουν τη j^{th} γραμμή κειμένου ($1 \leq j \leq N$) στην παραμορφωμένη και τη διορθωμένη εικόνα διαιρούνται σε k διαδοχικά σύνολα. Η τιμή της παραμέτρου k ισούται με τον αριθμό των σημείων που όρισε ο χρήστης κατά τη διαδικασία μαρκαρίσματος της γραμμής (§ 4.4.1).

Βήμα 2: Κάθε σύνολο σημείων προσεγγίζεται με ένα πολυώνυμο τρίτου βαθμού με τη μέθοδο ελαχίστων τετραγώνων. Τα πολυώνυμα του m^{th} συνόλου σημείων, για την παραμορφωμένη και τη διορθωμένη εικόνα, αντίστοιχα, δίνονται από τις εξισώσεις:

$${}^m y = {}^m a_3 x^3 + {}^m a_2 x^2 + {}^m a_1 x + {}^m a_0, \quad 1 \leq m \leq k$$

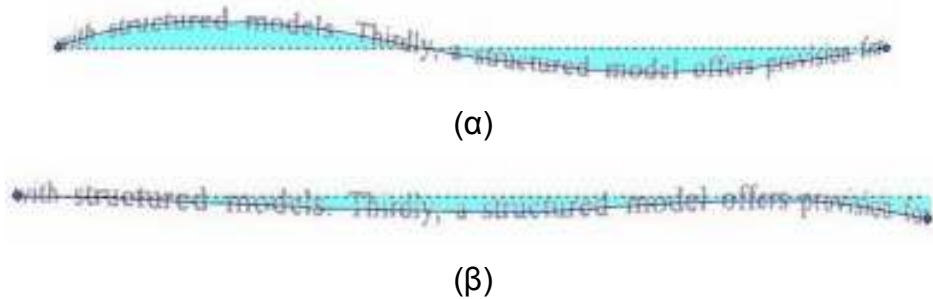
$${}^m y' = {}^m a'_3 x^3 + {}^m a'_2 x^2 + {}^m a'_1 x + {}^m a'_0, \quad 1 \leq m \leq k$$

Βήμα 3: Υπολογίζουμε το ολοκλήρωμα της j^{th} γραμμής κειμένου στην παραμορφωμένη (S_j) και τη διορθωμένη (S'_j) εικόνα προσθέτοντας τα ολοκληρώματα για κάθε πολυώνυμο (Σχήμα 4.18):

$$S_j = \sum_{m=1}^k \int_{x_s^m}^{x_e^m} ({}^m a_3 x^3 + {}^m a_2 x^2 + {}^m a_1 x) dx$$

$$S'_j = \sum_{m=1}^k \int_{x'_s^m}^{x'_e^m} ({}^m a'_3 x^3 + {}^m a'_2 x^2 + {}^m a'_1 x) dx$$

όπου ${}^m x_s, {}^m x_e$ και ${}^m x'_s, {}^m x'_e$ αντιπροσωπεύουν τα αρχικά και τελικά σημεία του m^{th} συνόλου σημείων στην παραμορφωμένη και τη διορθωμένη εικόνα, αντίστοιχα. Είναι προφανές ότι το S'_j ισούται με μηδέν όταν η j^{th} γραμμή κειμένου στη διορθωμένη εικόνα είναι μία οριζόντια ευθεία γραμμή.



Σχήμα 4.18: Η περιοχή ολοκληρώματος για μία γραμμή κειμένου (α) στην παραμορφωμένη και (β) στη διορθωμένη εικόνα.

Βήμα 4: Εξάγουμε τη μετρική αξιολόγησης της καμπυλότητας DM_j , η οποία αντιπροσωπεύει την απόδοση της τεχνικής βάσει της j^{th} γραμμής κειμένου, ως εξής:

$$DM_j = \begin{cases} 1 - \frac{S'_j}{S_j}, & \text{εάν } \frac{S'_j}{S_j} < 1 \\ 0, & \text{διαφορετικά} \end{cases}$$

Είναι προφανές ότι η μετρική DM_j κυμαίνεται στο διάστημα $[0, \dots, 1]$. Λαμβάνει τη μέγιστη τιμή ($DM_j = 1$) όταν η j^{th} γραμμή κειμένου στη διορθωμένη εικόνα είναι μία οριζόντια, ευθεία γραμμή, όπως είναι το αναμενόμενο ιδανικό αποτέλεσμα. Συνεπώς, συμπεραίνεται ότι η τεχνική διόρθωσης της τοπικής καμπυλότητας παρήγαγε το καλύτερο δυνατό αποτέλεσμα. Από την άλλη μεριά, λαμβάνει την ελάχιστη τιμή ($DM_j = 0$) όταν η j^{th} γραμμή κειμένου στη διορθωμένη εικόνα είναι το ίδιο παραμορφωμένη, ή σε χειρότερο βαθμό, με την αντίστοιχη γραμμή στην αρχική εικόνα ($S'_j \geq S_j$). Με άλλα λόγια, το ολοκλήρωμα S_j της αρχικής εικόνας χρησιμοποιείται σαν το κάτω όριο για τη j^{th} γραμμή κειμένου. Συνεπώς, συμπεραίνεται ότι η τεχνική διόρθωσης της τοπικής δε διόρθωσε πλήρως την εικόνα, ή ακόμα χειρότερα, ότι πρόσθεσε επιπλέον παραμόρφωση.

Βήμα 5: Επαναλαμβάνουμε όλα τα προηγούμενα βήματα για τις N επιλεγμένες γραμμές κειμένου με σκοπό να εξάγουμε τη συνολική μετρική αξιολόγησης της

καμπυλότητας που θα εκφράζει ολόκληρο το έγγραφο. Έπειτα εκφράζουμε τη μετρική DM ως το μέσο όρο όλων των επιμέρους μετρικών DM_j :

$$DM = \frac{\sum_{j=1}^N DM_j}{N} \times 100$$

όπου η μετρική DM κυμαίνεται στο διάστημα $[0, \dots, 100]$ και όσο μεγαλύτερη τιμή έχει τόσο καλύτερη είναι η απόδοση της τεχνικής διόρθωσης της τοπικής καμπυλότητας.

Είναι προφανές ότι η μετρική DM εξαρτάται από τις N γραμμές που έχουν επιλεγεί από τον χρήστη χωρίς να λαμβάνει υπόψη της το βαθμό παραμορφώσεων που έχει κάθε γραμμή ξεχωριστά. Για αυτό το λόγο, προτείνουμε και μία δεύτερη μετρική αξιολόγησης της καμπυλότητας (wDM) η οποία δίνει μεγαλύτερη βαρύτητα στις γραμμές κειμένου οι οποίες έχουν εντονότερες παραμορφώσεις. Ο βαθμός παραμόρφωσης κάθε γραμμής κειμένου εκφράζεται από το αντίστοιχο ολοκλήρωμα S_j . Συνεπώς, η μετρική wDM υπολογίζεται ως εξής:

$$wDM = \sum_{j=1}^N w_j DM_j \times 100, \text{ όπου } w_j = \frac{S_j}{\sum_{l=1}^N S_l}, \quad 1 \leq j \leq N$$

Είναι φανερό ότι όσο μεγαλύτερο είναι το ολοκλήρωμα S_j τόσο μεγαλύτερο είναι και το βάρος w_j της αντίστοιχης γραμμής κειμένου.

4.5 Πειραματικά Αποτελέσματα

Για την αξιολόγηση της προτεινόμενης μεθόδου διόρθωσης της τοπικής καμπυλότητας βασιστήκαμε τόσο στη νέα ημιαυτόματη μέθοδο αξιολόγησης (§ 4.4), όσο και στη χρήση ενός συστήματος οπτικής αναγνώρισης (OCR) για έμμεση αξιολόγηση η οποία είναι ευρέως διαδεδομένη [45], [47], [49-53], [55].

Η αξιολόγηση των τεχνικών διόρθωσης της τοπικής καμπυλότητας με τη χρήση OCR πραγματοποιείται υπολογίζοντας το ποσοστό αναγνώρισης των χαρακτήρων (character accuracy) [135] στο αρχικό και στο διορθωμένο έγγραφο μετά την εφαρμογή της τεχνικής. Το ποσοστό αναγνώρισης των χαρακτήρων ορίζεται ως ο λόγος του πλήθους των χαρακτήρων που αναγνωρίστηκαν σωστά (δηλαδή το πλήθος των χαρακτήρων στο ground-truth μείον τους χαρακτήρες

που αναγνωρίστηκαν λάθος) προς το συνολικό πλήθος των χαρακτήρων στο ground-truth:

$$\text{Character Accuracy} = \frac{\#characters - \#errors}{\#characters}$$

Για να ορίσουμε τα λάθη στην αναγνώριση χαρακτήρων υπολογίζουμε τον ελάχιστο αριθμό ενεργειών που πρέπει να γίνουν (εισαγωγή, διαγραφή ή αντικατάσταση) για να παραχθεί η σωστή συμβολοσειρά (Edit Distance) [136]. Επιπλέον, η πλειοψηφία των συστημάτων οπτικής αναγνώρισης όταν έχουν χαμηλή εμπιστοσύνη στην απόφαση τους σημαδεύουν του χαρακτήρες ως “ύποπτους” (“suspect” characters). Συνεπώς, όσο καλύτερη είναι η ποιότητα του εγγράφου που αναγνωρίζει ένα σύστημα οπτικής αναγνώρισης, τόσο μεγαλύτερη εμπιστοσύνη θα έχει και άρα θα παράγει λιγότερους “ύποπτους” χαρακτήρες.

Για την αξιολόγηση των προτεινόμενων μεθόδων χρησιμοποιήθηκε μία βάση από 100 παραμορφωμένα έγγραφα η οποία περιέχει έγγραφα στην Αγγλική και Γερμανική γλώσσα. Στα έγγραφα υπάρχει μεγάλη ποικιλία στο μέγεθος της γραμματοσειράς αλλά και στο είδος των παραμορφώσεων. Για λόγους σύγκρισης, εφαρμόσαμε και τη μέθοδο [131] στην ίδια βάση καθώς και το εμπορικό πρόγραμμα BookRestorer [117]. Η οπτική αναγνώριση πραγματοποιήθηκε από το πρόγραμμα ABBYY FineReader Engine 8.1 [127]. Στον Πίνακα 4.1 παρουσιάζονται τα συνολικά αποτελέσματα αναγνώρισης για όλες τις μεθόδους.

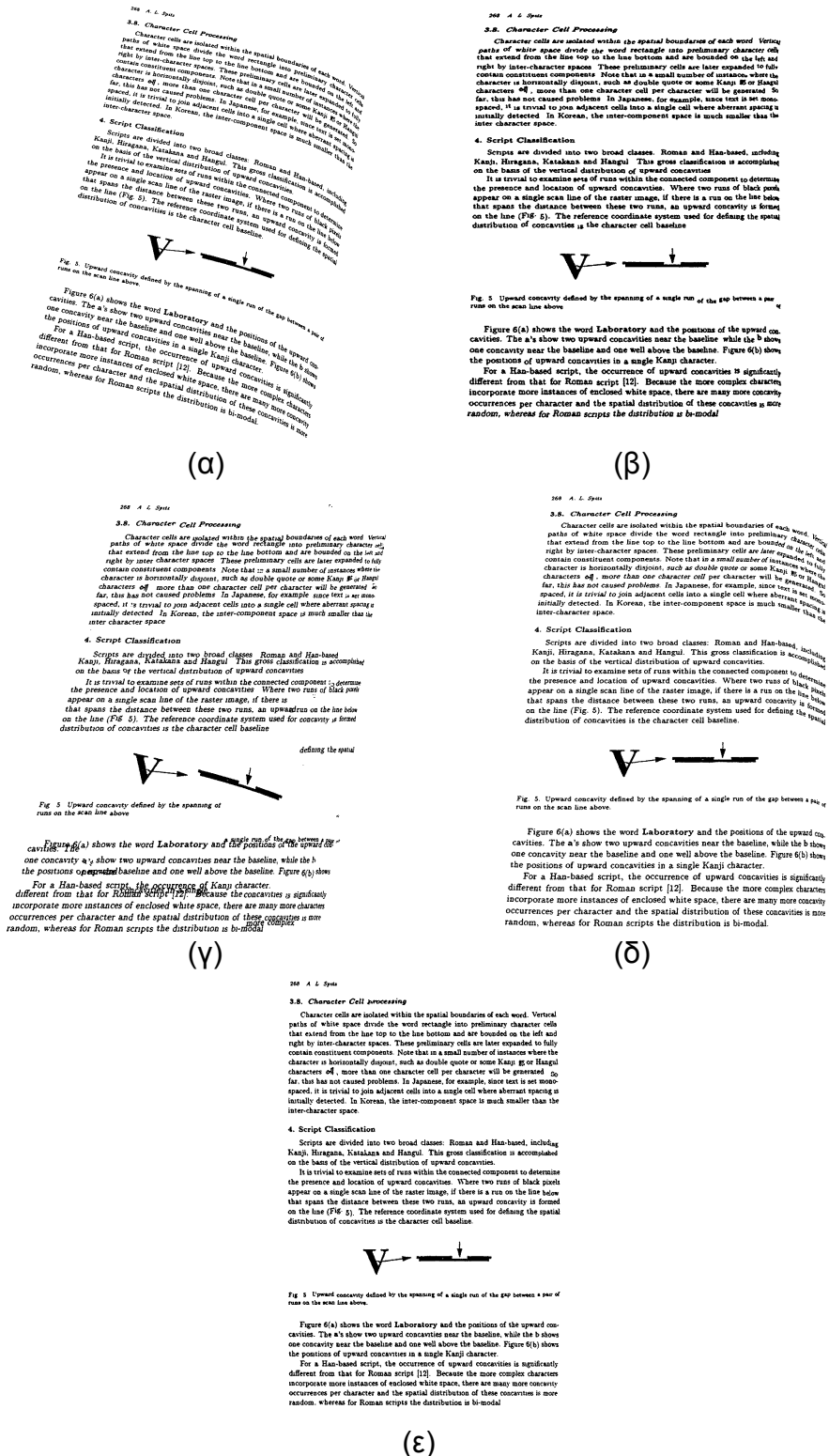
Όπως καταδεικνύουν και τα αποτελέσματα, η βελτιωμένη έκδοση της προτεινόμενης μεθόδου (TSD.v2) πετυχαίνει τα καλύτερα ποσοστά αναγνώρισης. Μετά την εφαρμογή της μεθόδου TSD.v2 το ποσοστό αναγνώρισης αυξάνεται κατά 37 ποσοστιαίες μονάδες, καταδεικνύοντας έτσι την αποτελεσματικότητα της προτεινόμενης μεθόδου. Επίσης, η βελτιωμένη έκδοση πετυχαίνει καλύτερα αποτελέσματα από την αρχική έκδοση (TSD.v1). Επιπλέον, και οι δύο εκδόσεις της προτεινόμενης μεθόδου είναι καλύτερες από τη μέθοδο SBD [131], συνεπώς, η εφαρμογή των δύο φάσεων βελτιώνει το αποτέλεσμα και βοηθάει στην καλύτερη κατάτμηση των γραμμών κειμένου. Τέλος, παρατηρούμε ότι η μέθοδος TSD.v2 παράγει τους λιγότερους “ύποπτους” χαρακτήρες. Άρα, η TSD.v2 παράγει έγγραφα καλύτερης ποιότητας και αυξάνει την εμπιστοσύνη στο σύστημα αναγνώρισης. Ο χαμηλός αριθμός “ύποπτων” χαρακτήρων είναι μία ένδειξη ότι

εάν χρησιμοποιήσουμε περισσότερα έγγραφα η μέθοδος TSD.v2 έχει μεγαλύτερη πιθανότητα να παράγει καλύτερα αποτελέσματα από τις άλλες μεθόδους.

Πίνακας 4.1: Πειραματικά αποτελέσματα χρησιμοποιώντας το ποσοστό αναγνώρισης χαρακτήρων.

Μέθοδος	Χαρακτήρες	Λάθη	“ύποπτοι” χαρακτήρες	Ποσοστό Αναγνώρισης Χαρακτήρων
Χωρίς Διόρθωση	170726	74191	2739	56,54%
SBD [131]	170726	31553	3896	81,51%
Προτεινόμενη μέθοδος TSD.v1	170726	24637	3467	85,56%
BookRestorer [117]	170726	16170	2627	90,52%
Προτεινόμενη μέθοδος TSD.v2	170726	10549	879	93,82%

Το Σχήμα 4.19 παρουσιάζει ένα χαρακτηριστικό αποτέλεσμα όλων των μεθόδων. Η προτεινόμενη μέθοδος TSD.v2 διορθώνει όλες τις παραμορφώσεις, τη στροφή καθώς και την κάθετη ευθυγράμμιση του εγγράφου. Επίσης, παρατηρούμε ότι μπορεί να χειριστεί και έγγραφα που περιέχουν εικόνες. Τέλος, παρατηρούμε και κάποια λάθη, κυρίως στην ευθυγράμμιση των γραμμών κειμένου όπως στη λεζάντα της εικόνα, λόγω λανθασμένου προσδιορισμού της γραμμής αναφοράς των λέξεων. Όμως, αυτά τα λάθη δεν επηρεάζουν την αναγνώριση. Η μέθοδος TSD.v1 προκαλεί αρκετά λάθη λόγω λανθασμένου προσδιορισμού της γραμμής αναφοράς των λέξεων καθώς επίσης παρατηρούμε και πολλές παραμορφωμένες λέξεις στη δεξιά μεριά του εγγράφου. Η μέθοδος SBD προκαλεί επικάλυψη των γραμμών κειμένου λόγω λανθασμένης κατάτμησης και δε διορθώνει την κάθετη ευθυγράμμιση του εγγράφου. Τέλος, το εμπορικό πρόγραμμα BookRestorer δε μπορεί να χειριστεί όλες τις παραμορφώσεις και γι’ αυτό πολλές γραμμές κειμένου δεν είναι ευθείες.



Σχήμα 4.19: Αποκατάσταση παραμορφωμένου εγγράφου; (α) αρχικό έγγραφο και το αποτέλεσμα μετά την εφαρμογή της μεθόδου (β) TSD.v1, (γ) SBD [131], (δ) BookRestorer [117] και (ε) TSD.v2.

Η κατάμηση των λέξεων είναι ένα σημαντικό στάδιο της προτεινόμενης μεθόδου (§ 4.2.1). Για αυτό τον σκοπό αξιολογήσαμε τις παραπάνω μεθόδους υπολογίζοντας επιπλέον και το ποσοστό αναγνώρισης των λέξεων (word accuracy) [135] τόσο στο αρχικό όσο και στο διορθωμένο έγγραφο. Το ποσοστό

αναγνώρισης των λέξεων ορίζεται ως ο λόγος του πλήθους των λέξεων που αναγνωρίστηκαν σωστά (δηλαδή το πλήθος των λέξεων στο ground-truth μείον τις λέξεις που αναγνωρίστηκαν λάθος) προς το συνολικό πλήθος των λέξεων στο ground-truth:

$$\text{Word Accuracy} = \frac{\# \text{ words} - \# \text{ misrecognized } _ \text{ words}}{\# \text{ words}}$$

Μία λέξη θεωρείται ότι αναγνωρίστηκε σωστά όταν όλοι οι χαρακτήρες της έχουν αναγνωριστεί σωστά. Στον Πίνακα 4.2 παρουσιάζονται τα συνολικά αποτελέσματα. Όπως και στο προηγούμενο πείραμα, όπου χρησιμοποιήθηκε το ποσοστό αναγνώρισης χαρακτήρων, η βελτιωμένη έκδοση της προτεινόμενης μεθόδου TSD.v2 πετυχαίνει τα καλύτερα ποσοστά αναγνώρισης τα οποία βελτιώνονται κατά 39 ποσοστιαίες μονάδες σε σχέση με τα αρχικά έγγραφα. Επιπλέον, η σειρά κατάταξης όλων των μεθόδων παραμένει η ίδια.

Πίνακας 4.2: Πειραματικά αποτελέσματα χρησιμοποιώντας το ποσοστό αναγνώρισης λέξεων.

Μέθοδος	Λέξεις	Λάθη	Ποσοστό Αναγνώρισης Λέξεων
Χωρίς Διόρθωση	27012	14916	44,78%
SBD [131]	27012	10072	62,71%
Προτεινόμενη μέθοδος TSD.v1	27012	9169	66,06%
BookRestorer [117]	27012	5714	78,85%
Προτεινόμενη μέθοδος TSD.v2	27012	4303	84,07%

Για να αξιολογήσουμε το στάδιο κατάτμησης των λέξεων, το αντικαταστήσαμε με δύο γνωστές τεχνικές της τρέχουσας τεχνολογικής στάθμης, τις τεχνικές κατάτμησης λέξεων που χρησιμοποιούν η ABBYY FineReader Engine 8.1 [127] και το OCROpus [137] και επαναλάβουμε τα ίδια πειράματα. Στον Πίνακα 4.3 παρουσιάζονται τα τελικά αποτελέσματα. Όπως είναι φανερό, από τα ποσοστά αναγνώρισης χαρακτήρων και λέξεων, η προτεινόμενη μέθοδος κατάτμησης λέξεων συμπεριφέρεται καλύτερα στα παραμορφωμένα έγγραφα.

Πίνακας 4.3: Πειραματικά αποτελέσματα χρησιμοποιώντας διαφορετικές τεχνικές κατάτμησης λέξεων.

Τεχνική Κατάτμησης Λέξεων	Χαρακτήρες	Λάθη Χαρακτήρων	“ύποπτοι” χαρακτήρες	Ποσοστό Αναγνώρισης Χαρακτήρων	Λέξεις	Λάθη Λέξεων	Ποσοστό Αναγνώρισης Λέξεων
OCROpus	170726	13279	2109	92,22%	27012	5442	79,85%
FineReader	170726	49018	2162	71,28%	27012	10886	59,70%
Προτεινόμενη μέθοδος TSD.v2	170726	10549	879	93,82%	27012	4303	84,07%

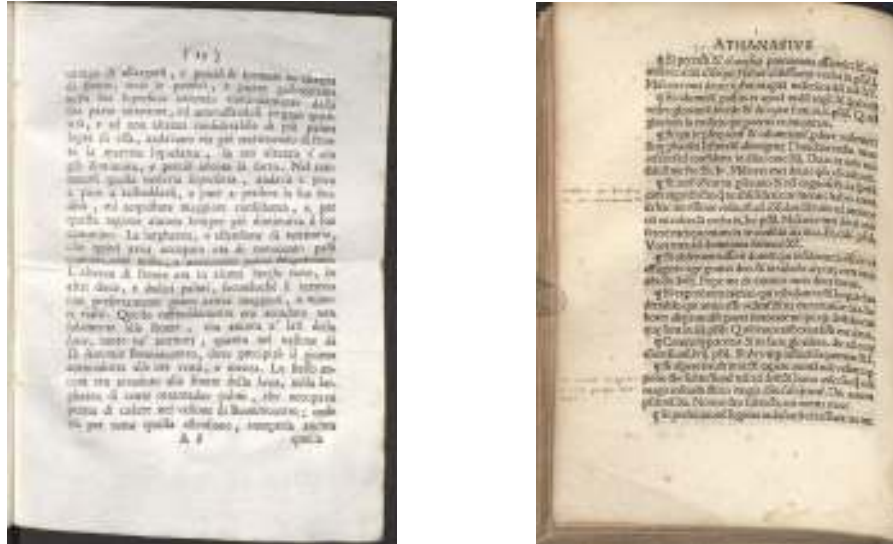
Στη συνέχεια, αξιολογήσαμε όλες τις μεθόδους διόρθωσης της τοπικής καμπυλότητας με την προτεινόμενη ημιαυτόματη μέθοδο. Στον Πίνακα 4.4 παρουσιάζονται τα συνολικά αποτελέσματα και για τις δύο προτεινόμενες μετρικές, DM και wDM , έχοντας μαρκάρει έξι αντιπροσωπευτικές γραμμές κειμένου σε κάθε έγγραφο ($N=6$). Παρατηρούμε ότι η σειρά κατάταξης είναι η ίδια με τους Πίνακες 4.1 και 4.2. Συνεπώς, οι δύο προτεινόμενες μετρικές αξιολόγησης είναι σε συμφωνία με το ποσοστό αναγνώρισης χαρακτήρων και λέξεων. Η προτεινόμενη τεχνική TSD.v2 έχει την καλύτερη απόδοση. Η μετρική wDM , η οποία δίνει μεγαλύτερη βαρύτητα στις γραμμές κειμένου που έχουν εντονότερες παραμορφώσεις, παρουσιάζει ελαφρώς χαμηλότερα ποσοστά για όλες τις μεθόδους σε σύγκριση με τη μετρική DM .

Πίνακας 4.4: Πειραματικά αποτελέσματα χρησιμοποιώντας την προτεινόμενη ημιαυτόματη μέθοδο αξιολόγησης.

Μέθοδος	DM	wDM
SBD [131]	79.35%	78.84%
Προτεινόμενη μέθοδος TSD.v1	82.53%	81.97%
Προτεινόμενη μέθοδος TSD.v2	91.90%	91.64%
BookRestorer [117]	84.12%	83.56%

Η ημιαυτόματη μέθοδος αξιολόγησης δοκιμάστηκε επιπλέον και σε ένα δεύτερο σύνολο εγγράφων (HSET) το οποίο αποτελείται από 50 ιστορικά έγγραφα για τα οποία δεν είναι δυνατόν να εφαρμόσουμε με επιτυχία οπτική αναγνώριση

χαρακτήρων. Το Σχήμα 4.20 παρουσιάζει δύο αντιπροσωπευτικά ιστορικά έγγραφα από το δεύτερο σύνολο, ενώ στον Πίνακα 4.5 παρουσιάζονται τα συνολικά αποτελέσματα και για τις δύο προτεινόμενες μετρικές έχοντας μαρκάρει επίσης έξι αντιπροσωπευτικές γραμμές κειμένου σε κάθε έγγραφο ($N = 6$).



Σχήμα 4.20: Παραδείγματα ιστορικών εγγράφων από το σύνολο HSET.

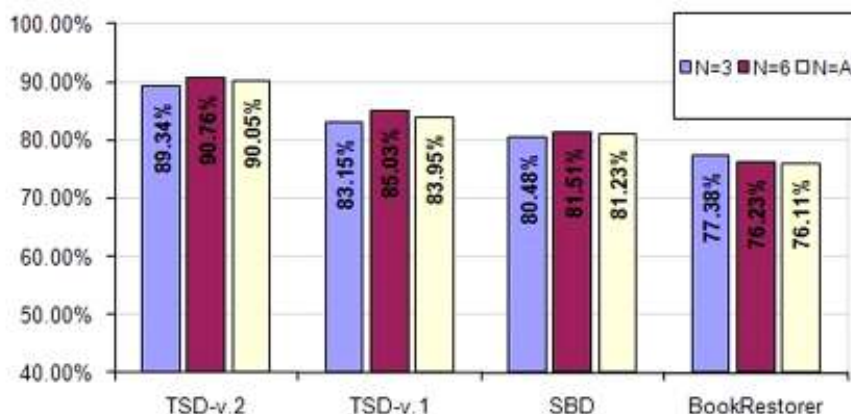
Πίνακας 4.5: Πειραματικά αποτελέσματα για το σύνολο HSET.

Μέθοδος	<i>DM</i>	<i>wDM</i>
SBD [131]	81.51%	80.92%
Προτεινόμενη μέθοδος TSD.v1	85.03%	84.22%
Προτεινόμενη μέθοδος TSD.v2	90.76%	89.64%
BookRestorer [117]	76.23%	75.67%

Στον Πίνακα 4.5 παρατηρούμε ότι η γενική σειρά κατάταξης έχει αλλάξει σε σύγκριση με το προηγούμενο σύνολο εγγράφων. Η προτεινόμενη τεχνική TSD.v2 παραμένει πρώτη, αλλά το εμπορικό πρόγραμμα BookRestorer έχει το χαμηλότερο ποσοστό. Αυτό συμβαίνει διότι το BookRestorer δε μπορεί να χειριστεί πάντα με επιτυχία ιστορικά έγγραφα με πολύπλοκη δομή ή με μεγάλη ποικιλία παραμορφώσεων. Δηλαδή, μπορεί να διορθώσει με επιτυχία τις παραμορφώσεις όταν είναι ομοιόμορφες σε ολόκληρο το έγγραφο, όπως στην

περίπτωση εγγράφου από βιβλιοδετημένο τόμο, αλλά όχι όταν κάθε γραμμή παρουσιάζει διαφορετική παραμόρφωση, όπως στο Σχήμα 4.20.

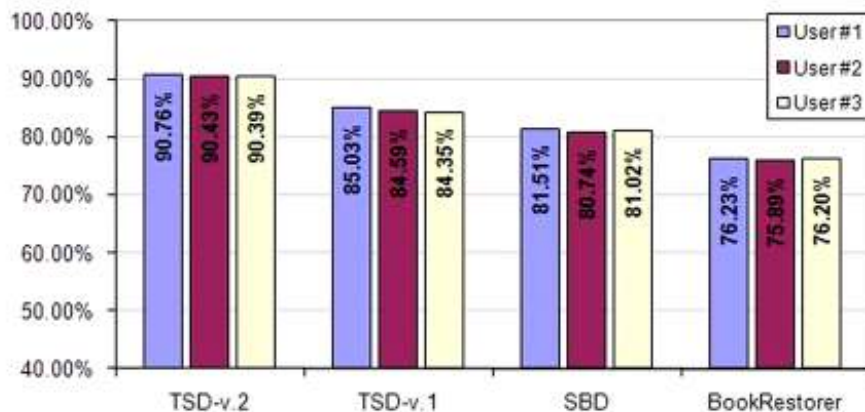
Σε όλα τα παραπάνω πειράματα, για να υπολογίσουμε τις δύο προτεινόμενες μετρικές επιλέξαμε και μαρκάραμε σε κάθε έγγραφο έξι αντιπροσωπευτικές γραμμές κειμένου ($N=6$). Για να εξετάσουμε την επιρροή που έχει το πλήθος των επιλεγμένων γραμμών κειμένου πραγματοποιήσαμε το ίδιο πείραμα στο σύνολο HSET άλλες δύο φορές, έχοντας μαρκάρει τη μία φορά τρεις γραμμές κειμένου ($N=3$) και την άλλη όλες τις γραμμές κειμένου κάθε εγγράφου ($N=All$). Στο γράφημα του Σχήματος 4.21 παρουσιάζονται τα πειραματικά αποτελέσματα χρησιμοποιώντας μόνο τη μετρική *DM*. Σε όλες τις περιπτώσεις η σειρά κατάταξης των μεθόδων δεν αλλάζει παρατηρώντας, όπως είναι φυσικό, μόνο πολύ μικρές μεταβολές στα ποσοστά κάθε μεθόδου. Συνεπώς, συμπεραίνουμε ότι για να έχουμε αξιόπιστα αποτελέσματα δε χρειάζεται να μαρκάρουμε όλες τις γραμμές κειμένου ενός εγγράφου αλλά μόνο ένα ικανοποιητικό αριθμό αντιπροσωπευτικών γραμμών κειμένου.



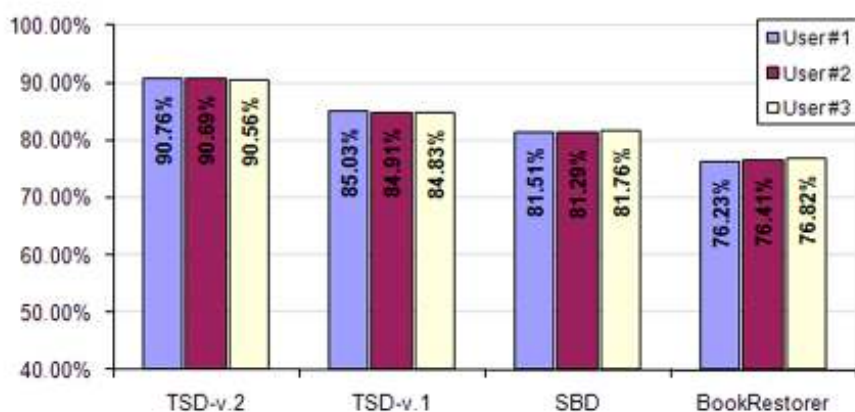
Σχήμα 4.21: Συγκριτικά αποτελέσματα για το σύνολο HSET σύμφωνα με τη μετρική *DM* μαρκάροντας (i) τρεις, (ii) έξι και (iii) όλες τις γραμμές κειμένου των εγγράφων.

Για να εξετάσουμε το βαθμό που επηρεάζει ο ανθρώπινος παράγοντας την προτεινόμενη μέθοδο πραγματοποιήσαμε δύο επιπλέον πειράματα στο σύνολο HSET. Στο πρώτο πείραμα χρησιμοποιήσαμε τρεις διαφορετικούς χρήστες για να επιλέξουν και να μαρκάρουν έξι γραμμές κειμένου σε κάθε έγγραφο. Στόχος μας είναι να εξετάσουμε πόσο επηρεάζει τα αποτελέσματα η επιλογή των γραμμών κειμένου. Στο δεύτερο πείραμα χρησιμοποιήσαμε επίσης τρεις διαφορετικούς χρήστες για να μαρκάρουν τις ίδιες γραμμές κειμένου σε κάθε έγγραφο. Στόχος μας είναι να εξετάσουμε πόσο επηρεάζει τα αποτελέσματα η διαδικασία μαρκαρίσματος. Τα συνολικά αποτελέσματα για τα δυο πειράματα εμφανίζονται

στα Σχήματα 4.22 και 4.23 χρησιμοποιώντας τη μετρική *DM*. Από τα αποτελέσματα συμπεραίνουμε ότι η διαδικασία μαρκαρίσματος δεν επηρεάζει σημαντικά τη μέθοδο καθώς παρατηρούνται πάρα πολύ μικρές μεταβολές στα ποσοστά κάθε μεθόδου.



Σχήμα 4.22: Συγκριτικά αποτελέσματα για το σύνολο HSET σύμφωνα με τη μετρική *DM* όπου τρεις διαφορετικοί χρήστες επέλεξαν και μάρκαραν έξι γραμμές κειμένου.



Σχήμα 4.23: Συγκριτικά αποτελέσματα για το σύνολο HSET σύμφωνα με τη μετρική *DM* όπου τρεις διαφορετικοί χρήστες μάρκαραν τις ίδιες έξι γραμμές κειμένου.

5. ΚΑΤΑΤΜΗΣΗ ΓΡΑΜΜΩΝ ΚΕΙΜΕΝΟΥ

5.1 Εισαγωγή

Ο συνδυασμός ταξινομητών για τη βελτιστοποίηση των αποτελεσμάτων είναι συνήθως τακτική σε πολλούς τομείς της αναγνώρισης προτύπων [138], όπως για παράδειγμα στην αναγνώριση λέξεων [139], στην επεξεργασία ομιλίας [140] και σε εφαρμογές βιομετρικής [141]. Στις περιπτώσεις αυτές χρησιμοποιούνται κάποιοι κανόνες για να συνδυάσουν τις εξόδους πολλών ταξινομητών με στόχο να αξιοποιήσουν συμπληρωματικές πληροφορίες από κάθε ταξινομητή και να δημιουργήσουν ένα υβριδικό ταξινομητή που θα έχει καλύτερη απόδοση από τον καθένα ξεχωριστά. Με αλλά λόγια, το συνδυασμό ταξινομητών μπορούμε να το δούμε σαν μία συλλογή αδύναμων ταξινομητών, όπου ο καθένας μπορεί να λύσει διαφορετικά δύσκολα προβλήματα. Μερικές από τις πιο συνηθισμένες μεθόδους συνδυασμών ταξινομητών που χρησιμοποιούνται είναι η πλειοψηφική επιλογή (voting) καθώς και η γραμμική και λογιστική παλινδρόμηση (linear and logistic regression).

Στον τομέα της επεξεργασίας και ανάλυσης εγγράφων έχουν προταθεί διάφορες συνδυαστικές τεχνικές για τη βελτιστοποίηση των αποτελεσμάτων οπτικής αναγνώρισης [142], [143], [144]. Οι τεχνικές συνδυάζουν τις συμβολοσειρές που προκύπτουν από διαφορετικές μηχανές οπτικής αναγνώρισης με στόχο τη βελτίωση του τελικού αποτελέσματος. Για να εφαρμοστούν οι τεχνικές αυτές με επιτυχία θα πρέπει να υπάρχει αντιστοίχιση μεταξύ των διαφορετικών αποτελεσμάτων καθώς επίσης και τα λάθη να είναι ασυσχέιστα μεταξύ τους. Επιπλέον, στο [145] προτείνεται μία μεθοδολογία για τη βελτιστοποίηση των αποτελεσμάτων κατάτμησης του εγγράφου σε ομογενείς περιοχές (page segmentation) συνδυάζοντας διαφορετικά αποτελέσματα. Ένα βασικό μειονέκτημα της μεθοδολογίας είναι ότι οι κανόνες συνδυασμού σχεδιάστηκαν μετά την ανάλυση και τον οπτικό έλεγχο των αποτελεσμάτων συγκεκριμένων τεχνικών κατάτμησης, συνεπώς είναι καθαρά προσανατολισμένη σε αυτές τις τεχνικές κατάτμησης και δε μπορεί να γενικευτεί.

Στο κεφάλαιο αυτό προτείνουμε μία πρωτότυπη μεθοδολογία συνδυασμού αποτελεσμάτων κατάτμησης των εγγράφων σε γραμμές κειμένου η οποία έχει δημοσιευθεί στο [147]. Η μεθοδολογία προσπαθεί να συνδυάσει συμπληρωματικές τεχνικές κατάτμησης υπό την έννοια ότι διαφορετικές

προσεγγίσεις μπορούν να αντιμετωπίσουν αποτελεσματικότερα διαφορετικές δυσκολίες που συναντάμε στα ιστορικά έγγραφα.

Στη βιβλιογραφία υπάρχουν πολλές μέθοδοι για την κατάτμηση του εγγράφου, τις βασικότερες των οποίων αναφέρουμε στο δεύτερο κεφάλαιο (§ 2.2). Οι περισσότερες τεχνικές κατάτμησης, ανάλογα με την προσέγγιση που ακολουθούν, μπορούν τα ταξινομηθούν σε τέσσερις βασικές κατηγορίες: (1) ανάλυση προβολών, (2) εξομάλυνση, (3) μετασχηματισμό Hough και (4) φάσμα εγγράφου. Τεχνικές από κάθε κατηγορία μπορούν να αντιμετωπίσουν αποτελεσματικότερα διαφορετικά προβλήματα που συναντάμε στην κατάτμηση ιστορικών εγγράφων, όπως συνδεδεμένα ή επικαλυπτόμενα συστατικά, χαμηλή ποιότητα, ποικιλομορφία στις γωνίες κλίσης και τον προσανατολισμό, θόρυβο κ.α.

Οι τεχνικές κατάτμησης που βασίζονται στην ανάλυση των προβολών είναι οι πιο διαδεδομένες [77], [81]. Χρησιμοποιούνται συνήθως στα τυπωμένα έγγραφα αλλά μπορούν να προσαρμοστούν και σε χειρόγραφα. Οι τεχνικές που ανήκουν σε αυτή την κατηγορία μπορούν να χειριστούν έγγραφα με ποικιλομορφία στα κενά διαστήματα ανάμεσα στις λέξεις και τους χαρακτήρες. Δεν μπορούν όμως να χειριστούν έγγραφα που παρουσιάζουν επικαλύψεις μεταξύ των συστατικών τους και κλίση. Για παράδειγμα, βασικά προβλήματα στην κατάτμηση των γραμμών κειμένου είναι η ποικιλομορφία στις γωνίες κλίσης και τον προσανατολισμό ακόμα και μέσα στην ίδια γραμμή κειμένου καθώς και η παρουσία συνδεδεμένων ή επικαλυπτόμενων συστατικών που ανήκουν σε διαφορετικές γραμμές κειμένου.

Οι τεχνικές εξομάλυνσης, όπως η fuzzy RLSA [148] και η adaptive RLSA [115], βασίζονται στη συνένωση τμημάτων της εικόνας που ανήκουν στην ίδια κατηγορία. Οι τεχνικές αυτής της κατηγορίας, με τη βοήθεια τοπικών χαρακτηριστικών, μπορούν να αντιμετωπίσουν προβλήματα όπως τα επικαλυπτόμενα ή συνδεδεμένα συστατικά. Επιπλέον, εφαρμόζονται με επιτυχία σε έγγραφα με διαφορετικά μεγέθη συμβολοσειρών. Παρόλα αυτά, δε μπορούν να χειριστούν έγγραφα που περιέχουν διαφορετικές γωνίες κλίσης αλλά και κενά διαστήματα ανάμεσα στις λέξεις και τους χαρακτήρες.

Ο μετασχηματισμός Hough είναι ένα από το πιο διαδεδομένα εργαλεία στην επεξεργασία εγγράφων. Έχει χρησιμοποιηθεί για τον εντοπισμό της κλίσης, των περιοχών κειμένου καθώς και των γραμμών κειμένου [34],[76]. Βασίζεται στην πλειοψηφική τεχνική (voting process) και η επιτυχία του έγκειται στη γενικότητά του καθώς δεν απαιτεί κάποια γνώση εκ των προτέρων. Το πλεονέκτημα των

τεχνικών που ανήκουν σε αυτή την κατηγορία είναι ότι οι γραμμές κειμένου μπορούν να βρίσκονται σε οποιαδήποτε γωνία και προσανατολισμό, ακόμα και εντός πινάκων ή σχημάτων. Όμως, δυσκολεύονται να χειριστούν προβλήματα όπου η γωνία και ο προσανατολισμός αλλάζουν μέσα στην ίδια γραμμή κειμένου.

Τέλος, το φάσμα εγγράφου (Document Spectrum) [60] είναι μία αναπαράσταση του εγγράφου η οποία περιγράφει τα καθολικά δομικά χαρακτηριστικά του και μπορεί να χρησιμοποιηθεί για την κατάτμησή του. Η χρήση του φάσματος του εγγράφου είναι μία προσέγγιση «από κάτω προς τα πάνω» η οποία βασίζεται στην ομαδοποίηση των αντικειμένων του εγγράφου [74]. Οι κατάλληλοι κανόνες ομαδοποίησης μπορούν να λύσουν πολλά προβλήματα όπως η ποικιλομορφία στις γωνίες κλίσης και τον προσανατολισμό ακόμα και μέσα στην ίδια γραμμή κειμένου. Όμως, πιθανά λάθη κατά την αρχικοποίηση οδηγούν σε αποτυχία όπως στις περιπτώσεις όπου οι γραμμές κειμένου, οι λέξεις ή οι χαρακτήρες βρίσκονται πολύ κοντά.

Συνεπώς, ο στόχος της προτεινόμενης μεθοδολογίας είναι να εκμεταλλευτεί αποτελέσματα κατάτμησης τεχνικών που ανήκουν σε διαφορετικές κατηγορίες καθώς και συγκεκριμένα χαρακτηριστικά του αρχικού εγγράφου ώστε να παραχθεί ένα βελτιωμένο αποτέλεσμα κατάτμησης. Αν και η προτεινόμενη μεθοδολογία αφορά την κατάτμηση σε γραμμές κειμένου, μπορεί να επεκταθεί και στην κατάτμηση λέξεων και χαρακτήρων.

5.2 Μεθοδολογία Συνδυασμού Αποτελεσμάτων Κατάτμησης Γραμμών Κειμένου

5.2.1 Γενική Περιγραφή

Θεωρούμε ότι έχουμε διαφορετικά αποτελέσματα κατάτμησης για μία εικόνα και τα συνδυάζουμε με στόχο να αυξήσουμε την ακρίβεια του αποτελέσματος κατάτμησης. Ως υποπεριοχές κατάτμησης ορίζουμε τις περιοχές εκείνες που προκύπτουν από την τομή των διαφορετικών αποτελεσμάτων. Η προτεινόμενη μεθοδολογία αποτελείται από πέντε βήματα:

Βήμα 1: Εξαγωγή χαρακτηριστικών από υποπεριοχές.

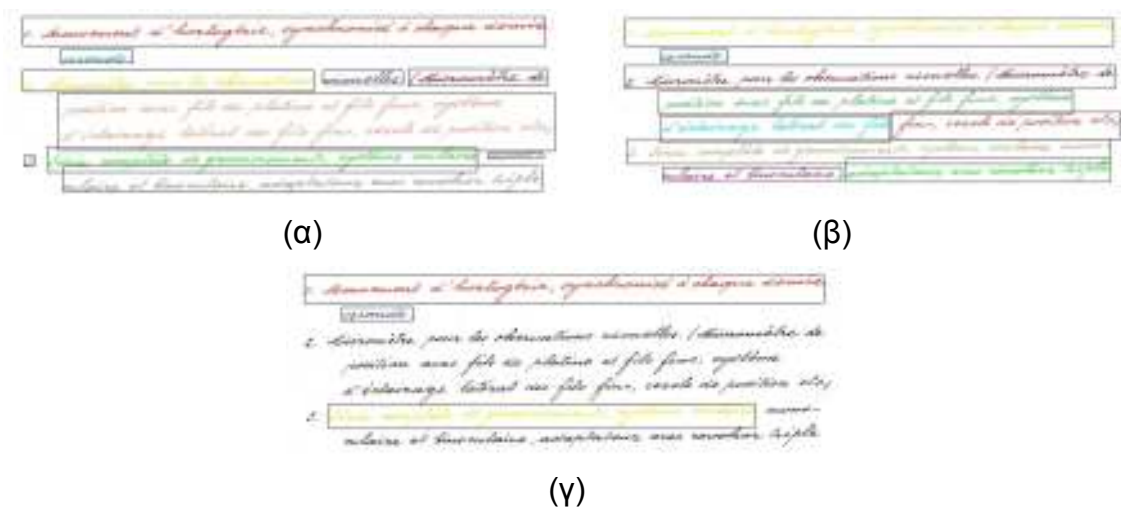
Βήμα 2: Εντοπισμός υποπεριοχών κατάτμησης που έχουν εντοπιστεί σωστά.

Βήμα 3: Ομαδοποίηση υποπεριοχών.

Βήμα 4: Επεξεργασία ομάδων και δημιουργία νέων περιοχών κατάτμησης.

Βήμα 5: Τελική επεξεργασία αποτελέσματος.

Στο “Βήμα 1” εντοπίζουμε τις υποπεριοχές κατάτμησης για τις οποίες συμφωνούν όλα τα αποτελέσματα πάνω από ένα κατώφλι (υψηλό ποσοστό επικάλυψης) και εξάγουμε χαρακτηριστικά από κάθε υποπεριοχή. Έπειτα, υπολογίζουμε τη μέση τιμή κάθε χαρακτηριστικού. Στόχος μας είναι οι μέσες τιμές των χαρακτηριστικών που υπολογίσαμε να προσεγγίζουν τις μέσες τιμές των χαρακτηριστικών των σωστών περιοχών κατάτμησης. Για παράδειγμα, στην κατάτμηση γραμμών κειμένου περιμένουμε τα χαρακτηριστικά των υποπεριοχών να προσεγγίζουν τα χαρακτηριστικά των γραμμών κειμένου, όπως το μήκος ή το ύψος της γραμμής κειμένου (Σχήμα 5.1).

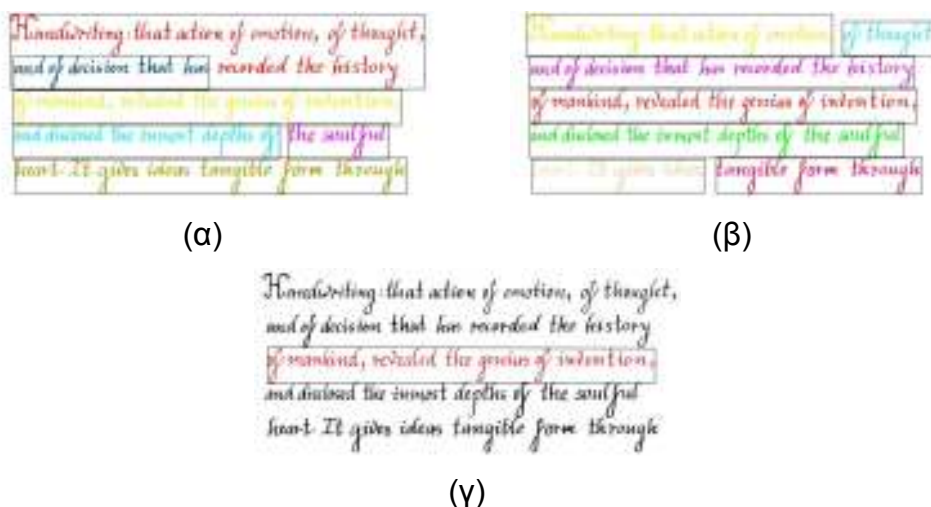


Σχήμα 5.1: “Βήμα 1” της προτεινόμενης μεθοδολογίας: (α), (β) αποτελέσματα κατάτμησης, (γ) υποπεριοχές από τις οποίες εξάγουμε χαρακτηριστικά.

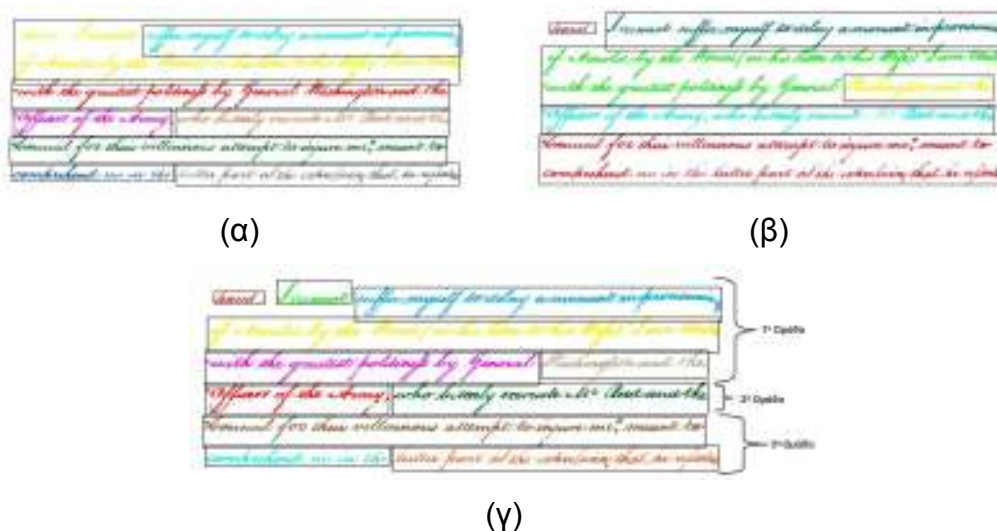
Με παρόμοιο τρόπο εργαζόμαστε και στο “Βήμα 2” εντοπίζοντας τις υποπεριοχές κατάτμησης για τις οποίες συμφωνούν όλα τα αποτελέσματα πάνω από ένα κατώφλι (πολύ υψηλό ποσοστό επικάλυψης). Θεωρούμε ότι αυτές οι υποπεριοχές έχουν εντοπιστεί σωστά. Συνεπώς, τις προσθέτουμε στο καινούριο αποτέλεσμα κατάτμησης και δε συμμετέχουν στα επόμενα βήματα (Σχήμα 5.2).

Στο “Βήμα 3” ομαδοποιούμε τις υπόλοιπες υποπεριοχές. Ενσωματώνουμε σε μία ομάδα όλες τις υποπεριοχές που συσχετίζονται σύμφωνα με τα αρχικά αποτελέσματα κατάτμησης. Στόχος μας είναι όλες οι υποπεριοχές που μπορούν να σχηματίσουν μία σωστή περιοχή κατάτμησης να είναι μέλη της ίδιας ομάδας. Συνεπώς, σε μία ομάδα συμμετέχουν υποπεριοχές που μπορούν να σχηματίσουν μία ή περισσότερες σωστές περιοχές κατάτμησης. Δηλαδή, όταν επεξεργαζόμαστε αποτελέσματα κατάτμησης γραμμών κειμένου οι ομάδες

περιέχουν υποπεριοχές που σχηματίζουν μία ή περισσότερες γραμμές κειμένου (Σχήμα 5.3).



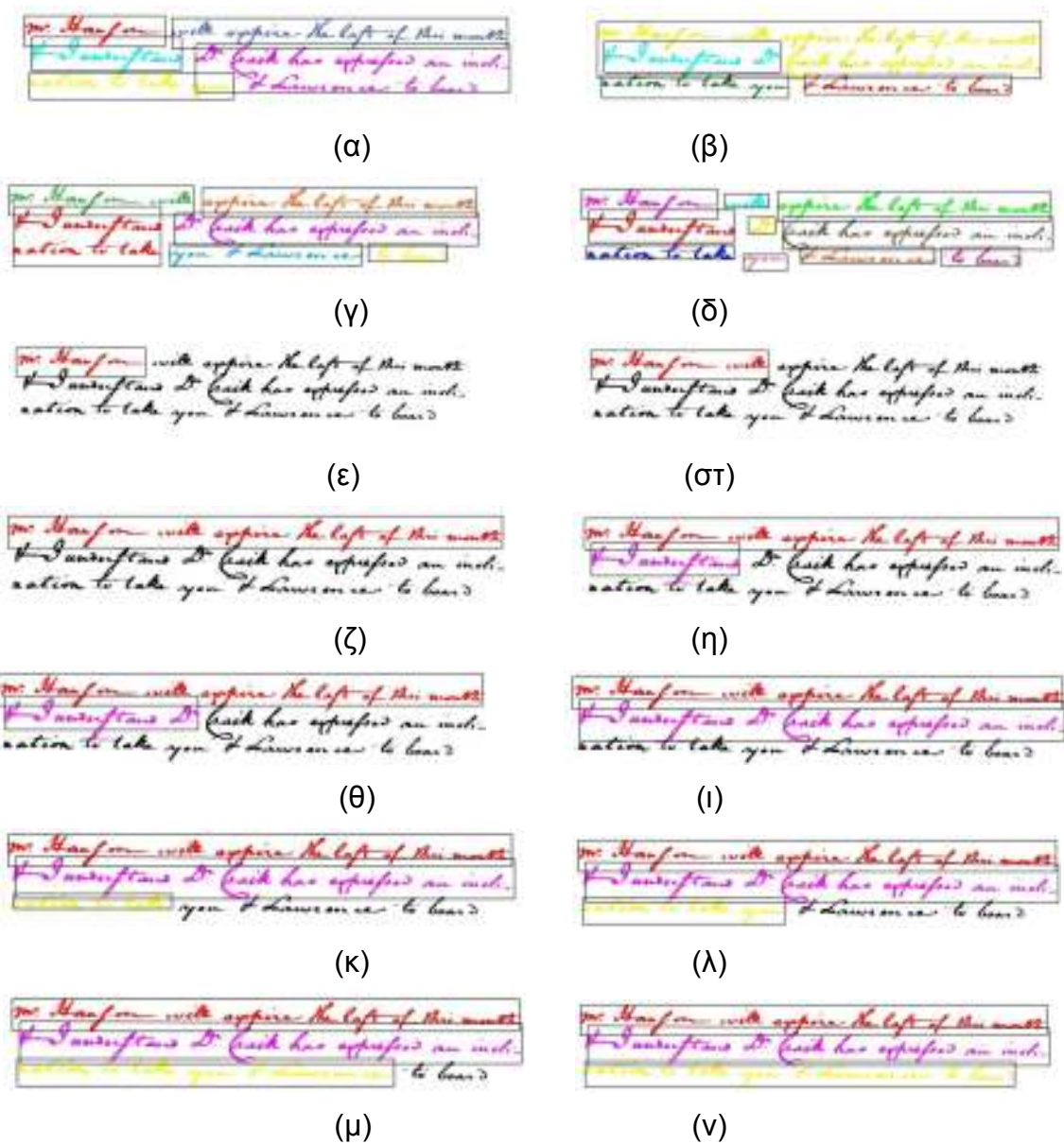
Σχήμα 5.2: “Βήμα 2” της προτεινόμενης μεθοδολογίας: (α), (β) αποτελέσματα κατάτμησης, (γ) υποπεριοχές που έχουν εντοπιστεί σωστά και προστίθενται στο καινούριο αποτέλεσμα κατάτμησης.



Σχήμα 5.3: “Βήμα 3” της προτεινόμενης μεθοδολογίας: (α), (β) αποτελέσματα κατάτμησης, (γ) ομαδοποίηση υποπεριοχών.

Αφού έχουμε ομαδοποιήσει τις υποπεριοχές, στο “Βήμα 4” επεξεργαζόμαστε κάθε ομάδα ξεχωριστά με σκοπό να αποφασίσουμε ποιες υποπεριοχές θα συγχωνευτούν για να προκύψουν οι νέες περιοχές κατάτμησης (Σχήμα 5.4). Ξεκινάμε από την υποπεριοχή με το μεγαλύτερο ποσοστό επικάλυψης σύμφωνα με τα αρχικά αποτελέσματα. Στη συνέχεια, εξετάζουμε ποιες από τις υποπεριοχές της ομάδας πρέπει να συγχωνευτούν μαζί της ώστε να προκύψει μία νέα περιοχή με χαρακτηριστικά κοντά στις μέσες τιμές που υπολογίσαμε στο “Βήμα 1”. Η νέα περιοχή που προκύπτει προστίθεται στο καινούριο αποτέλεσμα κατάτμησης και επαναλαμβάνουμε τη διαδικασία μέχρι να προστεθούν όλες οι υποπεριοχές της ομάδας.

Τέλος, στο “Βήμα 5”, εντοπίζουμε όλες τις υποπεριοχές που δεν ανήκουν σε καμιά περιοχή κατάτμησης στο καινούριο αποτέλεσμα και τις συγχωνεύουμε με την πλησιέστερη περιοχή.



Σχήμα 5.4: “Βήμα 4” της προτεινόμενης μεθοδολογίας: (α)-(γ) αποτελέσματα κατάτμησης, (δ) υποπεριοχές που ανήκουν στην ίδια ομάδα, διαδοχικά βήμα για τη δημιουργία της (ε)-(ζ) πρώτης, (η)-(ι) δεύτερης και (κ)-(ν) τρίτης γραμμής κειμένου.

5.2.2 Ορισμοί

Σε αυτή την παράγραφο θα αναφέρουμε τους βασικούς ορισμούς που θα χρησιμοποιήσουμε στην επόμενη ενότητα όπου περιγράφεται αναλυτικά η προτεινόμενη μεθοδολογία.

Ορισμός 1: Αρχική Εικόνα

Θεωρούμε ότι έχουμε μία δυαδική εικόνα:

$$I(x, y) = \begin{cases} 1 \\ 0 \end{cases}, \text{ όπου } 1 \leq x \leq I_x, 1 \leq y \leq I_y$$

όπου 1 και 0 αντιστοιχούν στην εικόνα και το υπόβαθρο, αντίστοιχα.

Ορισμός 2: Αποτελέσματα Κατάτμησης

Έστω $R_1(x, y), R_2(x, y), \dots, R_N(x, y)$ τα αποτελέσματα N διαφορετικών τεχνικών κατάτμησης, οι οποίες έχουν εφαρμοστεί στην εικόνα $I(x, y)$:

$$R_j(x, y) \in A_j, \quad A_j = \{1, \dots, n_j\}$$

Κάθε τιμή r_j στο σύνολο A_j προσδιορίζει ότι το pixel (x, y) ανήκει στην r_j^{th} περιοχή κατάτμησης σύμφωνα με τη j^{th} μέθοδο κατάτμησης. Στο Σχήμα 5.5 παρουσιάζεται ένα παράδειγμα με τρεις ($N = 3$) μεθόδους κατάτμησης.

Στόχος μας είναι να δημιουργήσουμε ένα καινούριο αποτέλεσμα κατάτμησης $R(x, y)$:

$$R(x, y) \in A, \quad A = \{1, \dots, n\}$$

Ορισμός 3: Τομή Αποτελεσμάτων Κατάτμησης

Ορίζουμε τη δυαδική εικόνα $C(x, y)_{(r_1, r_2, \dots, r_N)}$ ως εξής:

$$C(x, y)_{(r_1, r_2, \dots, r_N)} = \begin{cases} 1 & \text{εάν } (R_1(x, y) = r_1 \text{ και } \dots \text{ και } R_N(x, y) = r_N) \\ 0 & \text{διαφορετικά} \end{cases}$$

όπου $1 \leq r_j \leq n_j, j = 1, \dots, N$, η οποία αντιπροσωπεύει την τομή των αποτελεσμάτων κατάτμησης $R_1(x, y), R_2(x, y), \dots, R_N(x, y)$ για τις περιοχές κατάτμησης με τιμές (r_1, r_2, \dots, r_N) (Σχήμα 5.5(ε)).

Επίσης, ορίζουμε την εικόνα $D_j(x, y)_{(r_j)}$:

$$D_j(x, y)_{(r_j)} = \begin{cases} 1 & \text{εάν } R_j(x, y) = r_j \\ 0 & \text{διαφορετικά} \end{cases}$$

όπου $1 \leq r_j \leq n_j, j = 1, \dots, N$, η οποία αντιπροσωπεύει την τομή της αρχικής εικόνας $I(x, y)$ με την r_j^{th} περιοχή κατάτμησης σύμφωνα με τη j^{th} μέθοδο κατάτμησης (Σχήμα 5.5(στ)).

Ορισμός 4: Επικάλυψη Αποτελεσμάτων Κατάτμησης

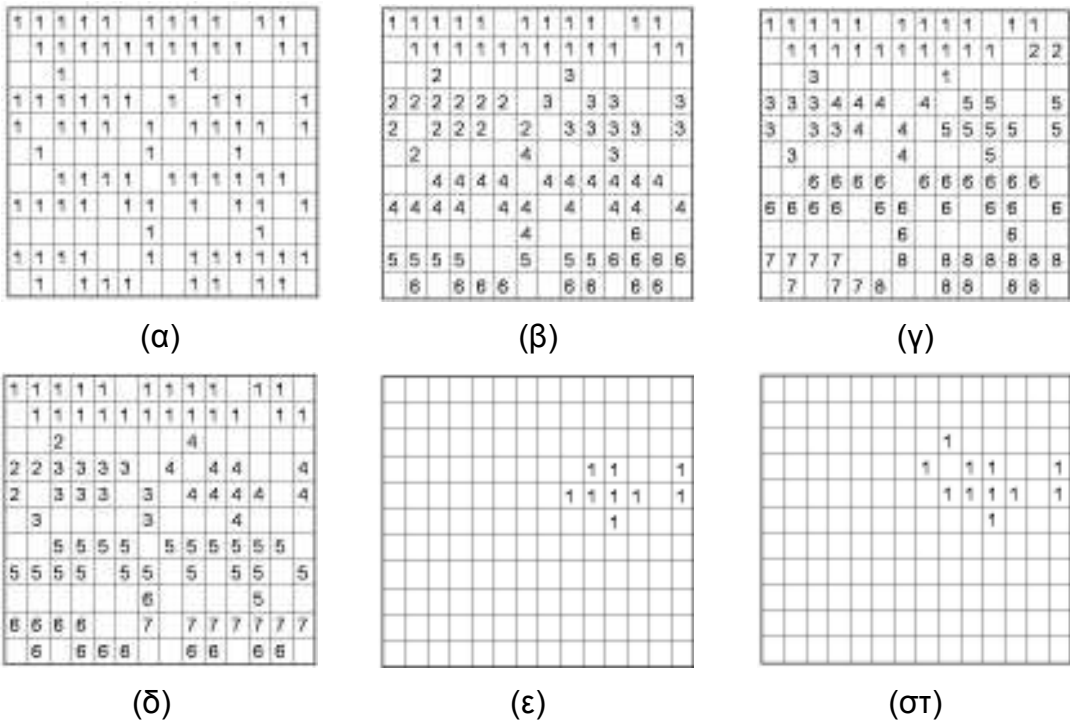
Για να προσδιορίσουμε την επικάλυψη των αποτελεσμάτων κατάτμησης ορίζουμε την ακόλουθη συνάρτηση:

$$f_j(r_1, r_2, \dots, r_N) = \begin{cases} \frac{\sum_{x,y} C(x, y)_{(r_1, r_2, \dots, r_N)}}{\sum_{x,y} D_j(x, y)_{(r_j)}} & \text{εάν } \sum_{x,y} D_j(x, y)_{(r_j)} \neq 0 \\ 0 & \text{διαφορετικά} \end{cases}$$

όπου $1 \leq r_j \leq n_j, j = 1, \dots, N$.

Για παράδειγμα, στο Σχήμα 5.5 έχουμε: $f_1(3, 5, 4) = \frac{\sum_{x,y} C(x, y)_{(3,5,4)}}{\sum_{x,y} D_1(x, y)_{(3)}} = \frac{9}{11} = 0.81$, που

σημαίνει ότι το 81% των pixels που έχουν τιμή 3 στο αποτέλεσμα $R_1(x, y)$, έχουν επίσης τιμή 5 στο αποτέλεσμα $R_2(x, y)$ και τιμή 4 στο αποτέλεσμα $R_3(x, y)$.



Σχήμα 5.5: Παράδειγμα με τρία αποτελέσματα κατάτμησης, όπου τα κενά pixels αντιπροσωπεύουν τα pixels υπόβαθρου. (α) Αρχική εικόνα $I(x, y)$ (β)-(δ) αποτελέσματα κατάτμησης $R_1(x, y)$, $R_2(x, y)$ και $R_3(x, y)$, (ε) $C(x, y)_{(3,5,4)}$ και (στ) $D_1(x, y)_{(3)}$.

Ορισμός 5: Ποσοστά Επικάλυψης

Με τη βοήθεια της συνάρτησης $f_j(r_1, r_2, \dots, r_N)$ ορίζουμε τα σύνολα υποπεριοχών HO και VHO . Το σύνολο HO περιέχει υποπεριοχές με τιμές (r_1, r_2, \dots, r_N) για τις οποίες όλα τα αποτελέσματα έχουν υψηλό ποσοστό επικάλυψης (70%) και ορίζεται ως εξής:

$$(r_1, r_2, \dots, r_N) \in HO \text{ εάν } f_j(r_1, r_2, \dots, r_N) \geq 0.7 \quad \forall j = 1, \dots, N$$

Παρόμοια ορίζουμε το σύνολο VHO το οποίο περιέχει υποπεριοχές με τιμές (r_1, r_2, \dots, r_N) για τις οποίες όλα τα αποτελέσματα έχουν πολύ υψηλό ποσοστό επικάλυψης (90%) και ορίζεται ως εξής:

$$(r_1, r_2, \dots, r_N) \in VHO \text{ εάν } f_j(r_1, r_2, \dots, r_N) \geq 0.9 \quad \forall j = 1, \dots, N$$

Ορισμός 6: Εξαγωγή Χαρακτηριστικών

Ορίζουμε την ακόλουθη συνάρτηση:

$$V[Q(x, y)] = FeatureExtraction [Q(x, y)]$$

η οποία δέχεται ως είσοδο μία δυαδική εικόνα $Q(x, y)$ και επιστρέφει ένα διάνυσμα χαρακτηριστικών $V[Q(x, y)] = \{v_1, v_2, \dots, v_p\}$, όπου $0 \leq v_i \leq 1, i = 1, \dots, p$. Επιλέγουμε τα χαρακτηριστικά ανάλογα με το πρόβλημα κατάτμησης που αντιμετωπίζουμε. Στην παράγραφο 5.3 παρουσιάζονται συγκεκριμένα χαρακτηριστικά για το συνδυασμό αποτελεσμάτων κατάτμησης γραμμών κειμένου.

5.2.3 Μεθοδολογία

Η μεθοδολογία αποτελείται από πέντε βήματα, τα οποία περιγράψαμε εν συντομία στην παράγραφο 5.2.1. Στην ενότητα αυτή θα περιγράψουμε αναλυτικά όλα τα βήματα της μεθοδολογίας.

Βήμα 1: Εξαγωγή χαρακτηριστικών από υποπεριοχές

Στο βήμα αυτό εξάγουμε χαρακτηριστικά από τις υποπεριοχές που ανήκουν στο σύνολο HO (§ 5.2.2) και υπολογίζουμε τη μέση τιμή $AV = \{av_1, av_2, \dots, av_p\}$. Στόχος μας είναι το διάνυσμα AV να προσεγγίζει τις μέσες τιμές των χαρακτηριστικών των σωστών περιοχών κατάτμησης. Ο αλγόριθμος που υλοποιεί το πρώτο βήμα είναι ο εξής:

Αλγόριθμος 1 - alg1 []: Υπολογισμός διανύσματος χαρακτηριστικών AV

{

(i) Εξαγωγή χαρακτηριστικών από κάθε υποπεριοχή του συνόλου HO

$$\forall (r_1, r_2, \dots, r_N) \in HO \quad V[C(x, y)_{(r_1, r_2, \dots, r_N)}] = FeatureExtraction[C(x, y)_{(r_1, r_2, \dots, r_N)}]$$

(ii) Υπολογισμός μέσης τιμής

$$AV = \frac{\sum_{i=1}^m V[C(x, y)_{(r_1, r_2, \dots, r_N)}]}{m}$$

}

όπου $m = |HO|$, δηλαδή το πλήθος των υποπεριοχών που ανήκουν στο σύνολο HO .

Βήμα 2: Εντοπισμός υποπεριοχών κατάτμησης που έχουν εντοπιστεί σωστά

Προσθέτουμε στο νέο αποτέλεσμα κατάτμησης $R(x, y)$ τις υποπεριοχές εκείνες που ανήκουν στο σύνολο VHO . Θεωρούμε ότι αυτές οι υποπεριοχές έχουν εντοπιστεί σωστά και συνεπώς δε συμμετέχουν στα επόμενα βήματα. Ο αλγόριθμος που υλοποιεί αυτό το βήμα είναι ο εξής:

Αλγόριθμος 2 - alg2 []: Εντοπισμός σωστών περιοχών κατάτμησης

{

(i) Αρχικοποίηση νέου αποτελέσματος

$$R(x, y) = 0 \quad \forall x, y$$

(ii) Προσθήκη υποπεριοχών που ανήκουν στο σύνολο VHO στην $R(x, y)$

$$k = 0$$

$$\forall (r_1, r_2, \dots, r_N) \in VHO$$

{

$$k = k + 1$$

$$R(x, y) = k * C(x, y)_{(r_1, r_2, \dots, r_N)} \quad \forall x, y$$

}

}

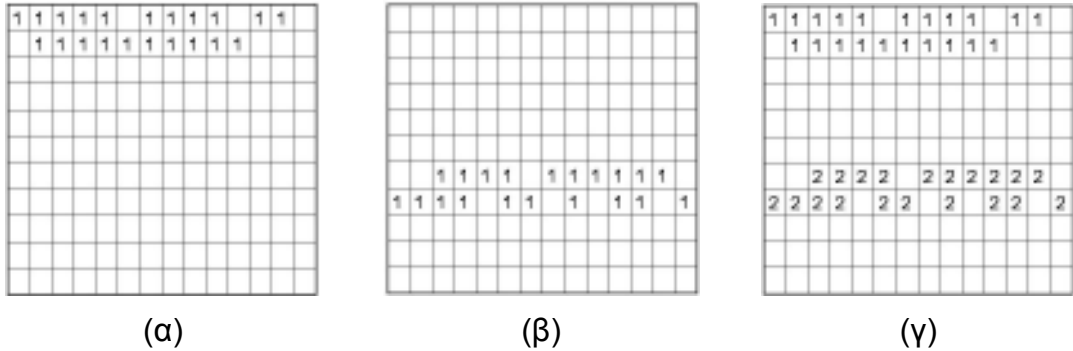
όπου $k = 1, \dots, z$ και $z = |VHO|$, δηλαδή το πλήθος των υποπεριοχών που ανήκουν στο σύνολο VHO .

Στο παράδειγμα του Σχήματος 5.5 παρατηρούμε δύο υποπεριοχές που ανήκουν στο σύνολο VHO , τις (1,1,1) και (4,6,5), καθώς ισχύουν τα ακόλουθα:

$$f_1(1,1,1) = \frac{21}{23} = 0.91, \quad f_2(1,1,1) = \frac{21}{22} = 0.95, \quad f_3(1,1,1) = \frac{21}{23} = 0.91 \text{ και}$$

$$f_1(4, 6, 5) = \frac{20}{22} = 0.90, \quad f_2(4, 6, 5) = \frac{20}{22} = 0.90, \quad f_3(4, 6, 5) = \frac{20}{21} = 0.95$$

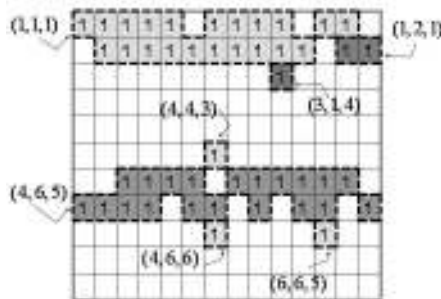
Στο Σχήμα 5.6 απεικονίζονται οι εικόνες που προκύπτουν από την τομή των αποτελεσμάτων για τις συγκεκριμένες υποπεριοχές καθώς και το νέο αποτέλεσμα που προκύπτει μετά την εφαρμογή του δεύτερου βήματος.



Σχήμα 5.6: Εφαρμογή του δεύτερου βήματος της μεθοδολογίας στο παράδειγμα του Σχήματος 5.5 (α) $C(x, y)_{(1,1)}$, (β) $C(x, y)_{(4,6,5)}$ και (γ) το νέο αποτέλεσμα $R(x, y)$.

Βήμα 3: Ομαδοποίηση υποπεριοχών

Οι υποπεριοχές με τιμές (r_1, r_2, \dots, r_N) που συμμετείχαν στο προηγούμενο βήμα, ή έχουν τουλάχιστον μία κοινή τιμή r_1, r_2, \dots, r_N με αυτές, δε συμμετέχουν σ' αυτό το βήμα. Οι υποπεριοχές από το παράδειγμά μας (Σχήμα 5.6) που δε συμμετέχουν απεικονίζονται στο Σχήμα 5.7.



Σχήμα 5.7: Υποπεριοχές του Σχήματος 5.5 που δε συμμετέχουν στο τρίτο βήμα της μεθοδολογίας.

Διαιρούμε τις υπόλοιπες υποπεριοχές σε ομάδες $(G_i, i = 1, \dots, l)$ ενσωματώνοντας σε μία ομάδα όλες τις υποπεριοχές (r_1, r_2, \dots, r_N) που συσχετίζονται σύμφωνα με τα αρχικά αποτελέσματα κατάτμησης, δηλαδή $\sum_{x,y} C(x, y)_{(r_1, r_2, \dots, r_N)} \neq 0$. Στόχος μας είναι

όλες οι υποπεριοχές που μπορούν να σχηματίσουν μία σωστή περιοχή κατάτμησης να είναι μέλη της ίδιας ομάδας. Με άλλα λόγια, σε μία ομάδα συμμετέχουν υποπεριοχές που μπορούν να σχηματίσουν μία ή περισσότερες γραμμές κειμένου. Επομένως, στο επόμενο βήμα θα εξετάσουμε ποιες

υποπεριοχές μίας ομάδας θα ενωθούν για να σχηματίσουν μία καινούρια γραμμή κειμένου. Ο αλγόριθμος που διαιρεί τις υποπεριοχές σε ομάδες είναι ο εξής:

Αλγόριθμος 3 – alg3 [] : Ομαδοποίηση υποπεριοχών

```

{
  i = 0
  Μέχρι όλες οι υποπεριοχές  $(r_1, r_2, \dots, r_N) : \sum_{x,y} C(x,y)_{(r_1, r_2, \dots, r_N)} \neq 0$  να έχουν
  ενσωματωθεί σε μία ομάδα:
  {
    (i) Δημιουργία νέας ομάδας  $G_i$ 
     $i = i + 1$ 
     $G_i = \emptyset$ 
    (ii) Εντοπισμός υποπεριοχής που δεν ανήκει σε καμία ομάδα
    Find  $(r_1, r_2, \dots, r_N) : [ \sum_{x,y} C(x,y)_{(r_1, r_2, \dots, r_N)} \neq 0 \text{ AND } (r_1, r_2, \dots, r_N) \notin G_j, j = 1, \dots, i-1 ]$ 
    (iii) Προσθήκη στην ομάδα
    alg4 [  $G_i, (r_1, r_2, \dots, r_N)$  ]
  }
}

```

όπου η αναδρομική συνάρτηση **alg4** [$G_i, (r_1, r_2, \dots, r_N)$], με είσοδο μία κενή ομάδα και μία υποπεριοχή, ορίζεται ως εξής:

Αλγόριθμος 4 - alg4 [$G_i, (r_1, r_2, \dots, r_N)$] : Προσθήκη στην ομάδα

```

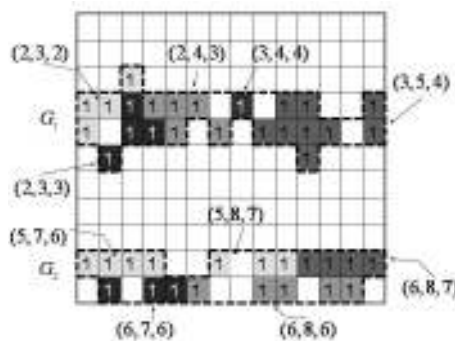
{
  (i) Προσθέτουμε την υποπεριοχή στην ομάδα
   $G_i = G_i \cup \{(r_1, r_2, \dots, r_N)\}$ 
  (ii) Εντοπίζουμε τις υποπεριοχές που έχουν τουλάχιστον μία κοινή τιμή με την
  αρχική υποπεριοχή και δεν ανήκουν σε καμία ομάδα
   $\forall (r'_1, r'_2, \dots, r'_N) : [ \sum_{x,y} C(x,y)_{(r'_1, r'_2, \dots, r'_N)} \neq 0 \text{ AND } (r'_1, r'_2, \dots, r'_N) \notin G_j, j = 1, \dots, i ]$ 
  if  $(r_1 = r'_1 \text{ OR } \dots \text{ OR } r_N = r'_N)$ 
  alg4 [  $G_i, (r'_1, r'_2, \dots, r'_N)$  ]
}

```


Στο παράδειγμά μας δημιουργούνται μόνο δύο ομάδες, G_1 και G_2 (Σχήμα 5.8):

$$G_1 = \{(2,3,2), (2,3,3), (2,4,3), (3,4,4), (3,5,4)\}$$

$$G_2 = \{(5,7,6), (5,8,7), (6,7,6), (6,8,6), (6,8,7)\}$$



Σχήμα 5.8: Οι ομάδες G_1 και G_2 που δημιουργούνται στο παράδειγμα του Σχήματος 5.5.

Βήμα 4: Επεξεργασία ομάδων και δημιουργία νέων περιοχών κατάτμησης

Επεξεργαζόμαστε κάθε ομάδα G_i , $i=1, \dots, l$ ξεχωριστά. Ο αλγόριθμος 5 εντοπίζει μία υποπεριοχή με τιμές (r_1, r_2, \dots, r_N) για την οποία τα αποτελέσματα έχουν το μεγαλύτερο βαθμό επικάλυψης, σύμφωνα με ένα από τα αποτελέσματα, και επίσης δεν έχει προστεθεί στο τελικό αποτέλεσμα. Στη συνέχεια, ο αλγόριθμος 6 εξετάζει ποιες από τις υπόλοιπες υποπεριοχές της ομάδας πρέπει να συγχωνευτούν μαζί της ώστε τα χαρακτηριστικά της περιοχής που προκύπτει να είναι πιο κοντά στις μέσες τιμές AV που υπολογίστηκαν στο πρώτο βήμα. Συνεπώς, η νέα περιοχή που προκύπτει προστίθεται στο τελικό αποτέλεσμα $R(x, y)$ και επαναλαμβάνεται η διαδικασία μέχρι να προστεθούν όλες οι υποπεριοχές της ομάδας. Οι αλγόριθμοι 5 και 6 δίνονται στη συνέχεια.

Αλγόριθμος 5 - alg5 [G_i]: Επεξεργασία ομάδας

{

(i) Αρχικοποίηση συνάρτησης g και των παραμέτρων k και END

$$\forall (r_1, r_2, \dots, r_N) \in G_i \quad g((r_1, r_2, \dots, r_N)) = 0$$

$$k = z$$

$$END = false$$

(ii) Μέχρι κάθε υποπεριοχή της ομάδας G_i να προστεθεί στο $R(x, y)$

Until ($END = true$)

{

$$k = k + 1$$

(iii) Εντοπίζουμε την υποπεριοχή με τη μεγαλύτερη επικάλυψη

$$MI = \max_{\substack{(r_1, r_2, \dots, r_N) \in G_i \\ g((r_1, r_2, \dots, r_N))=0}} \{ f_1(r_1, r_2, \dots, r_N), \dots, f_N(r_1, r_2, \dots, r_N) \}$$

$$Find (r_1, r_2, \dots, r_N) \in G_i : [g((r_1, r_2, \dots, r_N)) = 0 \text{ AND } (f_1(r_1, r_2, \dots, r_N) = MI \text{ OR } \dots \text{ OR } f_N(r_1, r_2, \dots, r_N) = MI)]$$

Εάν $MI = 0$ τότε $END = true$ διαφορετικά

{

(iv) Προσθέτουμε την περιοχή στο $R(x, y)$

$$R(x, y) = k * C(x, y)_{(r_1, r_2, \dots, r_N)} \quad \forall x, y$$

$$g((r_1, r_2, \dots, r_N)) = 1$$

(v) Υπολογίζουμε τη διαφορά ανάμεσα στα χαρακτηριστικά της περιοχής και το διάνυσμα AV

$$V[C(x, y)_{(r_1, r_2, \dots, r_N)}] = FeatureExtraction[C(x, y)_{(r_1, r_2, \dots, r_N)}]$$

$$Dif = | AV - V[C(x, y)_{(r_1, r_2, \dots, r_N)}] |$$

$$M(x, y) = C(x, y)_{(r_1, r_2, \dots, r_N)} \quad \forall x, y$$

(vi) Ελέγχουμε αν θα συγχωνευτεί άλλη υποπεριοχή της ομάδας

$$alg6 [G_i, Dif, M(x, y), k]$$

}

}

}

όπου η συνάρτηση g δείχνει εάν μία υποπεριοχή έχει προστεθεί ή όχι στο καινούριο αποτέλεσμα $R(x, y)$ και ο Αλγόριθμος 6 εξετάζει αν πρέπει να συγχωνευτούν και άλλες υποπεριοχές της ομάδας στην περιοχή $M(x, y)$ που δέχεται σαν είσοδο. Επιπλέον, παίρνει σαν είσοδο το σύνολο G_i , τη διαφορά των χαρακτηριστικών της περιοχής $M(x, y)$ από τις μέσες τιμές AV καθώς και την τιμή k που ορίζει την καινούρια περιοχή στο τελικό αποτέλεσμα $R(x, y)$.

Αλγόριθμος 6 - $alg6 [G_i, Dif, M(x, y), k]$: Συνένωση υποπεριοχών

{

(i) Αρχικοποίηση παραμέτρου END

$$END = false$$

(ii) Σταματάει μέχρι να ελέγξει όλες τις υποπεριοχές της ομάδας G_i

Until ($END = true$)

{

(iii) Εντοπίζουμε την υποπεριοχή με τη μεγαλύτερη επικάλυψη η οποία δεν έχει ελεγχθεί

$$MI = \max_{\substack{(r_1, r_2, \dots, r_N) \in G_i \\ g((r_1, r_2, \dots, r_N))=0}} \{f_1(r_1, r_2, \dots, r_N), \dots, f_N(r_1, r_2, \dots, r_N)\}$$

$$Find (r_1, r_2, \dots, r_N) \in G_i : [g((r_1, r_2, \dots, r_N)) = 0 \text{ AND } (f_1(r_1, r_2, \dots, r_N) = MI \text{ OR } \dots \text{ OR } f_N(r_1, r_2, \dots, r_N) = MI)]$$

Εάν $MI = 0$ τότε $END = true$ διαφορετικά

{

(iv) Συνένωση της υποπεριοχής με την περιοχή $M(x, y)$

$$N(x, y) = M(x, y) + C(x, y)_{(r_1, r_2, \dots, r_N)} \quad \forall x, y$$

$$g((r_1, r_2, \dots, r_N)) = 1$$

(v) Υπολογίζουμε τη διαφορά ανάμεσα στα χαρακτηριστικά της περιοχής και το διάνυσμα AV

$$V[N(x, y)] = FeatureExtraction[N(x, y)]$$

$$Dif_1 = |AV - V[N(x, y)]|$$

Εάν $Dif_1 \geq Dif$ τότε $g((r_1, r_2, \dots, r_N)) = 0$ διαφορετικά

{

(vi) Προσθέτουμε την περιοχή στο $R(x, y)$

$$R(x, y) = k * C(x, y)_{(r_1, r_2, \dots, r_N)} \quad \forall x, y$$

$$M(x, y) = N(x, y) \quad \forall x, y$$

}

}

}

}

Το Σχήμα 5.9 παρουσιάζει το αποτέλεσμα που προκύπτει μετά την εφαρμογή του τέταρτου βήματος στο παράδειγμα του Σχήματος 5.5. Παρατηρούμε ότι όλες οι υποπεριοχές της ομάδας G_1 συγχωνεύονται και δημιουργούν μία περιοχή στο τελικό αποτέλεσμα $R(x, y)$ με τιμή 5, ενώ από την ομάδα G_2 προκύπτουν δύο περιοχές. Οι υποπεριοχές $\{(5, 7, 6), (5, 8, 7), (6, 8, 7)\}$ σχηματίζουν την περιοχή με τιμή 3 και οι υπόλοιπες, $\{(6, 7, 6), (6, 8, 6)\}$, την περιοχή με τιμή 4.

1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1
	5									
5	5	5	5	5	5	5	5	5		5
5	5	5	5	5	5	5	5	5	5	5
	5							5		
	2	2	2	2	2	2	2	2	2	2
2	2	2	2	2	2	2	2	2	2	2
3	3	3	3		3	3	3	3	3	3
4	4	4	4		4	4	4	4	4	4

Σχήμα 5.9: Αποτέλεσμα μετά την εφαρμογή του τέταρτου βήματος στο παράδειγμα του Σχήματος 5.5.

Βήμα 5: Τελική επεξεργασία αποτελέσματος

Στο πέμπτο και τελευταίο βήμα, εντοπίζουμε όλες τις υποπεριοχές που δεν ανήκουν σε καμιά περιοχή στο νέο αποτέλεσμα κατάτμησης $R(x,y)$ (Σχήμα 5.7) και παίρνουν την τιμή της πλησιέστερης περιοχής (Σχήμα 5.10).

1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1
	5				1					
5	5	5	5	5	5	5	5	5		5
5	5	5	5	5	5	5	5	5	5	5
	5				2			5		
	2	2	2	2	2	2	2	2	2	2
2	2	2	2	2	2	2	2	2	2	2
					2				3	
3	3	3	3		3	3	3	3	3	3
4	4	4	4		4	4	4	4	4	4

Σχήμα 5.10: Αποτέλεσμα μετά την εφαρμογή του πέμπτου βήματος στο παράδειγμα του Σχήματος 5.9.

5.3 Πειραματικά Αποτελέσματα

Για να εξετάσουμε την αποτελεσματικότητα της προτεινόμενης μεθοδολογίας χρησιμοποιήσαμε δύο διαφορετικές μεθόδους κατάτμησης γραμμών κειμένου. Η πρώτη βασίζεται στην ανάλυση των προβολών [77]. Εντοπίζει τα ελάχιστα της συνάρτησης προβολών και τα θεωρεί ως υποψήφια σημεία κατάτμησης. Τα σημεία αυτά ταξινομούνται σε σχέση με τις αποστάσεις τους από τα γειτονικά σημεία. Τέλος, αναλύοντας όλες τις αποστάσεις της εικόνας αποφασίζει τα τελικά σημεία κατάτμησης. Η δεύτερη μέθοδος βασίζεται στην εξομάλυνση της εικόνας [115]. Προτείνουν τον αλγόριθμο adaptive RLSA, μία παραλλαγή του κλασσικού RLSA, στον οποίο οι παράμετροι εξομάλυνσης επιλέγονται με βάση τα γεωμετρικά χαρακτηριστικά των συνδεδεμένων συστατικών της εικόνας. Χρησιμοποιώντας αυτή την τεχνική πετυχαίνουν να ενώσουν τα συνδεδεμένα συστατικά που ανήκουν στην ίδια γραμμή κειμένου αποφεύγοντας τον θόρυβο.

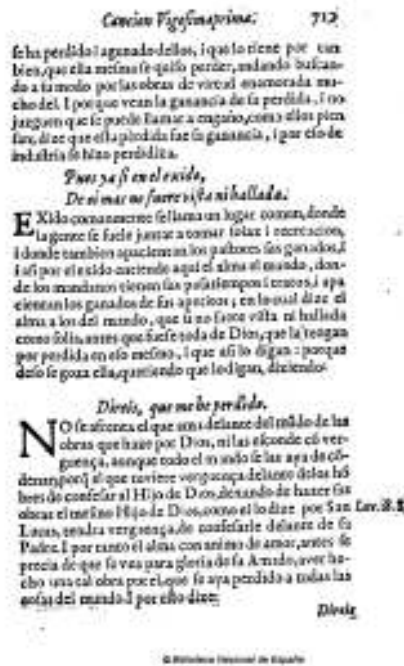
Εφαρμόσαμε τις παραπάνω μεθόδους κατάτμησης σε ένα σύνολο 50 ιστορικών εγγράφων (Σχήμα 5.11(α)) τα οποία περιέχουν συνολικά 1633 γραμμές κειμένου. Έπειτα, χρησιμοποιώντας τα δύο αποτελέσματα κατάτμησης για κάθε εικόνα

δημιουργήσαμε ένα νέο αποτέλεσμα βάσει της προτεινόμενης μεθοδολογίας. Επίσης, χρησιμοποιήσαμε και ένα δεύτερο σύνολο 50 σύγχρονων χειρόγραφων εγγράφων (Σχήμα 5.11(β)) τα οποία περιέχουν συνολικά 1144 γραμμές κειμένου. Ως χαρακτηριστικά των περιοχών χρησιμοποιήσαμε τα ακόλουθα:

$$v_1 = \frac{\text{Length of the bounding box}}{\text{Length of image}}, v_2 = \frac{\text{Height of the bounding box}}{\text{Height of image}}, v_3 = \frac{v_2}{v_1}$$

$$v_4 = \frac{\text{Foreground pixels}}{\text{Total pixels}}, v_5 = \frac{x \text{ co-ordinate of middle point}}{\text{Length of image}} \text{ και}$$

$$v_6 = \frac{\text{Median stroke width in the row with the maximum number of black-white transitions}}{\text{Length of image}}$$



(α)



(β)

Σχήμα 5.11: Παράδειγμα (α) ιστορικού και (β) χειρόγραφου εγγράφου που χρησιμοποιήθηκαν για την αξιολόγηση της προτεινόμενης μεθοδολογίας.

Για την αποτίμηση της κατάτμησης δημιουργήσαμε χειροκίνητα το ground-truth και χρησιμοποιήσαμε μία αυτοματοποιημένη διαδικασία. Η διαδικασία αυτή βασίζεται στον υπολογισμό του πλήθους των γραμμών κειμένου που εντοπίζονται σωστά με βάση το ground-truth [149]. Έστω I το σύνολο των σημείων της εικόνας, G_j το σύνολο των σημείων μέσα στην j περιοχή του ground-truth, R_i το σύνολο των σημείων μέσα στην i περιοχή του αποτελέσματος και $T(s)$ μία συνάρτηση που υπολογίζει τα στοιχεία του συνόλου s . Ο πίνακας

$MatchScore(i, j)$ απεικονίζει το αποτέλεσμα της αντιστοίχισης της περιοχής j του ground-truth με την περιοχή i του αποτελέσματος:

$$MatchScore(i, j) = \frac{T(G_j \cap R_i \cap I)}{T((G_j \cup R_i) \cap I)}$$

Θεωρούμε ένα ζευγάρι περιοχών (i, j) σαν one-to-one match όταν η τιμή του πίνακα $MatchScore$ είναι πάνω από ένα επιλεγμένο κατώφλι T_a . Αν N είναι το πλήθος των περιοχών του ground-truth και M το πλήθος των περιοχών του αποτελέσματος υπολογίζουμε το ποσοστό εντοπισμού (Detection Rate – DR) και την ακρίβεια αναγνώρισης (Recognition Accuracy – RA) με βάση τα παρακάτω:

$$DR = \frac{o2o}{N} \text{ και } RA = \frac{o2o}{M}$$

Η συνολική μετρική F -Measure (FM) υπολογίζεται ως εξής:

$$FM = \frac{2 \cdot DR \cdot RA}{DR + RA}$$

Αξιολογήσαμε την απόδοση των δύο τεχνικών κατάτμησης και τον συνδυασμό τους για τις εικόνες κάθε συνόλου χρησιμοποιώντας κατώφλι $T_a = 95\%$. Όπως παρατηρούμε στους Πίνακες 5.1 και 5.2 το τελικό αποτέλεσμα κατάτμησης που προκύπτει μετά από τον συνδυασμό τους έχει καλύτερη απόδοση από τις δύο αρχικές τεχνικές αυξάνοντας σημαντικά το F -Measure. Μετά τον συνδυασμό των αποτελεσμάτων εντοπίζονται σωστά σχεδόν 170 γραμμές κειμένου περισσότερες σε σχέση με τις αρχικές τεχνικές κατάτμησης στο σύνολο των ιστορικών εγγράφων και 200 γραμμές κειμένου στο σύνολο των χειρόγραφων εγγράφων. Στα Σχήματα 5.12 και 5.13 παρουσιάζονται δύο παραδείγματα εφαρμογής της προτεινόμενης μεθοδολογίας. Μετά τον συνδυασμό των αποτελεσμάτων παρατηρούμε ότι όλες οι γραμμές κειμένου έχουν εντοπιστεί σωστά, ακόμα και στην περίπτωση όπου η γραμμή κειμένου έχει εντοπιστεί λάθος και από τις δύο τεχνικές κατάτμησης (Σχήμα 5.13).

Πίνακας 5.1: Πειραματικά αποτελέσματα για το σύνολο των ιστορικών εγγράφων.

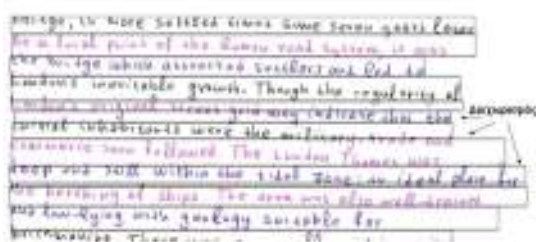
Μέθοδος Κατάτμησης	N	M	ο2ο	DR(%)	RA(%)	FM(%)
Ανάλυση Προβολών [77]	1633	1577	1327	81.26	84.15	82.68
Adaptive RLSA [115]	1633	1594	1358	83.16	85.19	84.16
Συνδυασμός αποτελεσμάτων	1633	1605	1529	93.63	95.26	94.44

Πίνακας 5.2: Πειραματικά αποτελέσματα για το σύνολο των χειρόγραφων εγγράφων.

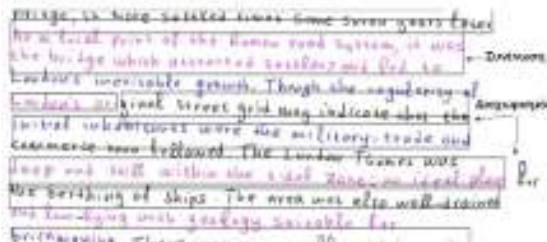
Μέθοδος Κατάτμησης	N	M	ο2ο	DR(%)	RA(%)	FM(%)
Ανάλυση Προβολών [77]	1144	1248	841	73.51	67.39	70.32
Adaptive RLSA [115]	1144	1314	860	75.17	65.45	69.98
Συνδυασμός αποτελεσμάτων	1144	1152	1071	93.62	92.97	93.29



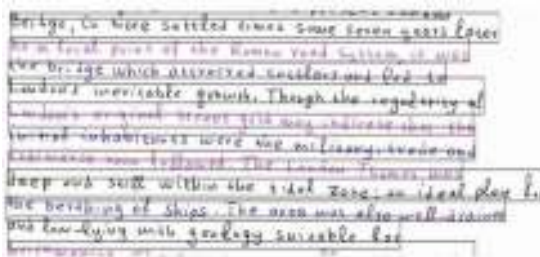
Σχήμα 5.12: Αποτελέσματα κατάτμησης ιστορικής εικόνας εγγράφου με τη χρήση της μεθόδου (α) ανάλυσης προβολών [77], (β) adaptive RLSA [115] και (γ) μετά τον συνδυασμό τους.



(α)



(β)



(γ)

Σχήμα 5.13: Αποτελέσματα κατάτμησης χειρόγραφης εικόνας εγγράφου με τη χρήση της μεθόδου (α) ανάλυσης προβολών [77], (β) adaptive RLSA [115] και (γ) μετά τον συνδυασμό τους.

6. ΚΑΤΑΤΜΗΣΗ ΧΑΡΑΚΤΗΡΩΝ

6.1 Εισαγωγή

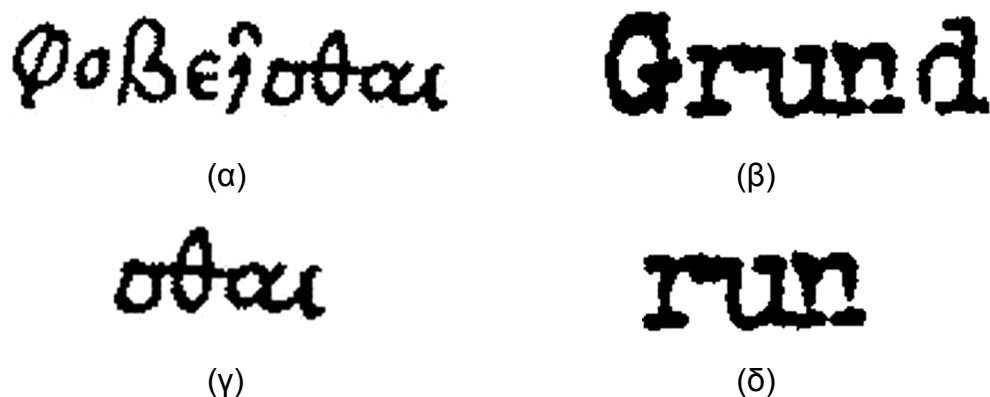
Το πιο σημαντικό στάδιο, και συγχρόνως το ποιοό δύσκολο, στην κατάτμηση των ιστορικών εγγράφων είναι η κατάτμηση των χαρακτήρων καθώς επηρεάζει άμεσα την απόδοση ενός συστήματος οπτικής αναγνώρισης. Στην παράγραφο 2.2.4 αναφέρουμε τις κυριότερες μεθόδους κατάτμησης χαρακτήρων που συναντάμε στη βιβλιογραφία, τα βασικά χαρακτηριστικά των οποίων παρουσιάζονται συνοπτικά στον Πίνακα 2.3. Στο κεφάλαιο αυτό προτείνουμε μία νέα μέθοδο κατάτμησης χαρακτήρων σε ιστορικά έγγραφα η οποία έχει δημοσιευθεί στο [115]. Στην εργασία [115] περιγράφονται όλα τα στάδια κατάτμησης ιστορικών εγγράφων, από γραμμές κειμένου και λέξεις μέχρι και χαρακτήρες. Στο πλαίσιο αυτής της διατριβής επικεντρωθήκαμε μόνο στο στάδιο της κατάτμησης χαρακτήρων, το οποίο περιγράφουμε σε αυτό το κεφάλαιο. Η προτεινόμενη μέθοδος βασίζεται στη σχεδίαση διαδρομών κατάτμησης οι οποίες συνδέουν χαρακτηριστικά σημεία του σκελετού της λέξης με χαρακτηριστικά σημεία του σκελετού του υπόβαθρου. Η επιλογή των διαδρομών που παράγουν το καλύτερο αποτέλεσμα κατάτμησης πραγματοποιείται με τη βοήθεια κριτηρίων όπως η απόκλιση του πλάτους του παραγόμενου χαρακτήρα από το αναμενόμενο πλάτος κ.α.

6.2 Αλγόριθμος Κατάτμησης Χαρακτήρων

Η βασική ιδέα είναι η σχεδίαση πιθανών διαδρομών κατάτμησης (segmentation paths) οι οποίες συνδέουν τα χαρακτηριστικά σημεία του σκελετού της λέξης και του υπόβαθρου. Βασιστήκαμε στη μέθοδο [92] η οποία έχει προταθεί για τον διαχωρισμό δύο χειρόγραφων ψηφίων που συνδέονται με συγκεκριμένους τύπους σύνδεσης.

Αρχικά, υπολογίζεται το κυρίαρχο ύψος γράμματος (AH) με το ιστόγραμμα των υψών όλων των πλαισίων που περικλείουν τα συνδεδεμένα συστατικά της εικόνας, όπως περιγράφεται στην εργασία [80]. Θεωρούμε ότι το πλάτος κάθε χαρακτήρα δε μπορεί να είναι μικρότερο από $MinCharWidth = 0.5 * AH$ και μεγαλύτερο από $MaxCharWidth = 1.5 * AH$. Έπειτα, εφαρμόζονται τα παρακάτω βήματα στα συνδεδεμένα συστατικά των οποίων ο λόγος ύψος προς πλάτος είναι μικρότερος ή ίσος του 0.5, καθώς τα συνδεδεμένα συστατικά που ικανοποιούν

αυτή τη συνθήκη είναι πιθανόν να αποτελούνται από περισσότερους από έναν χαρακτήρες και συνεπώς είναι υποψήφια προς διάσπαση (Σχήμα 6.1).



Σχήμα 6.1: (α)-(β) Εικόνες λέξεων, (γ)-(δ) συνδεδεμένα συστατικά υποψήφια προς διάσπαση.

Βήμα 1: Εξάγεται ο σκελετός των αντικειμένων και του υπόβαθρου (Σχήμα 6.2). Για τον υπολογισμό του σκελετού γίνεται χρήση της επαναληπτικής μεθόδου που αναλύεται στην εργασία [150].



Σχήμα 6.2: Σκελετός αντικειμένου (πράσινο) και υπόβαθρου (μπλε) για τα Σχήματα 6.1(γ) και 6.1(δ) αντίστοιχα.

Βήμα 2: Ο σκελετός ταξινομείται στα ακόλουθα τμήματα (Σχήμα 6.3):

(i) **Πάνω τμήμα:** Το τμήμα του σκελετού του υπόβαθρου που βρίσκεται πάνω από το σκελετό του αντικειμένου.

(ii) **Κάτω τμήμα:** Το τμήμα του σκελετού του υπόβαθρου που βρίσκεται κάτω από το σκελετό του αντικειμένου.

(iii) **Τμήμα χαρακτήρων:** Ο σκελετός του αντικειμένου.

(iv) **Τμήμα οπών:** Το τμήμα του σκελετού του υπόβαθρου που περικλείεται από το σκελετό του αντικειμένου.



Σχήμα 6.3: Ταξινόμηση των τμημάτων του σκελετού: (i) πάνω τμήμα (γαλάζιο), (ii) κάτω τμήμα (ροζ), (iii) τμήμα χαρακτήρων (πράσινο) και (iv) τμήμα οπών (μπλε).

Βήμα 3: Εντοπίζονται τα χαρακτηριστικά σημεία του σκελετού (Σχήμα 6.4):

(i) **Σημείο διακλάδωσης:** Το σημείο στο οποίο διασταυρώνονται περισσότερα από δυο τμήματα του σκελετού.

(ii) **Τελικό σημείο:** Το σημείο στο οποίο καταλήγει ένα τμήμα του σκελετού. Δηλαδή το σημείο που έχει μόνο ένα γείτονα. Τα σημεία κατάληξης ανήκουν μόνο στο σκελετό του υπόβαθρου και όχι στο σκελετό του αντικειμένου.

(iii) **Σημείο γωνίας:** Το σημείο σε ένα τμήμα του σκελετού στο οποίο η καμπυλότητα αλλάζει απότομα.



(α)



(β)

Σχήμα 6.4: Χαρακτηριστικά σημεία στο σκελετό: (i) σημεία διασταύρωσης (μωβ), (ii) σημεία κατάληξης (καφέ) και (iii) σημεία γωνίας (μπλε).

Βήμα 4: Σε αυτό το βήμα σχεδιάζονται όλες οι υποψήφιες διαδρομές κατάτμησης (Σχήμα 6.5). Εφαρμόζουμε δύο μεθόδους αναζήτησης, “από πάνω προς τα κάτω” και “από κάτω προς τα πάνω”. Στην “από πάνω προς τα κάτω” αναζήτηση, κάθε διαδρομή ξεκινάει από ένα χαρακτηριστικό σημείο το οποίο ανήκει στο πάνω τμήμα του σκελετού, διασχίζει ένα ή δύο χαρακτηριστικά σημεία του σκελετού των χαρακτήρων και καταλήγει σε ένα χαρακτηριστικό σημείο στο κάτω τμήμα του σκελετού. Αν δεν υπάρχει χαρακτηριστικό σημείο στο σκελετό των χαρακτήρων ή στο κάτω τμήμα του σκελετού τότε δημιουργείται μία κάθετη διαδρομή που τερματίζει στο κάτω τμήμα του σκελετού. Αντίστοιχη διαδικασία χρησιμοποιείται και στην “από κάτω προς τα πάνω” αναζήτηση. Μία διαδρομή για να είναι υποψήφια διαδρομή κατάτμησης πρέπει να ικανοποιεί τους παρακάτω περιορισμούς:

(i) Το μήκος της να είναι μικρότερο από AH

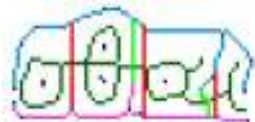
(ii) Το πλάτος της να είναι μικρότερο από $AH / 3$.

(iii) Ο λόγος του αριθμού των pixels κειμένου προς τον αριθμό των pixels υπόβαθρου να είναι μικρότερος από τρία.

(iv) Η απόσταση ανάμεσα σε δύο χαρακτηριστικά σημεία που διασχίζει η διαδρομή να είναι μικρότερη από $0.8 \cdot AH$.

(v) Δεν πρέπει να υπάρχει τμήμα του σκελετού του χαρακτήρα ανάμεσα σε δύο χαρακτηριστικά σημεία.

(vi) Αν η διαδρομή είναι κάθετη θα πρέπει να κόβει το τμήμα του σκελετού του χαρακτήρα σε ένα μόνο σημείο.



(α)



(β)

Σχήμα 6.5: Υποψήφιες διαδρομές κατάτμησης.

Βήμα 5: Αφού έχουν σχεδιαστεί όλες οι υποψήφιες διαδρομές κατάτμησης, στο πέμπτο και τελευταίο βήμα της μεθόδου επιλέγονται οι καταλληλότερες για την κατάτμηση του αντικειμένου (Σχήμα 6.6). Ξεκινώντας από την αρχή του αντικειμένου (ή από την τελευταία διαδρομή που επιλέχθηκε) εντοπίζουμε τις διαδρομές εκείνες που οδηγούν σε χαρακτήρες με πλάτος μέσα στα επιτρεπτά όρια [*MinCharWidth*, *MaxCharWidth*]. Ανάμεσα από αυτές τις υποψήφιες διαδρομές κατάτμησης θα επιλεγθεί εκείνη που ελαχιστοποιεί τα ακόλουθα κριτήρια :

(i) Την απόκλιση του πλάτους του χαρακτήρα που προκύπτει από το αναμενόμενο πλάτος AH .

(ii) Την απόκλιση του ύψους του χαρακτήρα που προκύπτει από το αναμενόμενο ύψος AH .

(iii) Το μήκος της υποψήφιας διαδρομής κατάτμησης.

(iv) Το πλάτος της υποψήφιας διαδρομής κατάτμησης.

Η διαδικασία αυτή επαναλαμβάνεται έως ότου το αντικείμενο δε μπορεί να κατατμηθεί σε περισσότερους χαρακτήρες ή δεν υπάρχουν άλλες υποψήφιες διαδρομές κατάτμησης.



(α)



(β)



(γ)



(δ)

Σχήμα 6.6: Επιλογή τελικών διαδρομών κατάτμησης για το Σχήμα 6.5.

Στον Πίνακα 6.1 παρουσιάζονται τα συνολικά αποτελέσματα για όλες τις μεθόδους με $N = 71818$ και κατώφλι $T_a = 90\%$. Παρατηρούμε ότι η προτεινόμενη μέθοδος ξεπερνάει όλες τις τεχνικές της τρέχουσας τεχνολογικής στάθμης και τα εμπορικά προγράμματα, με ποσοστό $FM=84.49\%$. Εντοπίζει σωστά 62425 χαρακτήρες από τους συνολικά 71818. Ακολουθεί η μέθοδος που βασίζεται στο RLSA [151] με ποσοστό $FM=80.01\%$, εντοπίζοντας σωστά σχεδόν 6000 χαρακτήρες λιγότερους. Έπειτα ακολουθούν τα προγράμματα ABBYY FineReader Engine 8.1 [127] και OCROpus [137] και τελευταία η μέθοδος που βασίζεται στην ανάλυση των προβολών [73]. Στο Σχήμα 6.8 παρουσιάζονται αντιπροσωπευτικά αποτελέσματα για όλες τις μεθόδους.

Πίνακας 6.1: Πειραματικά αποτελέσματα κατάτμησης χαρακτήρων.

Μέθοδος Κατάτμησης	N	M	ο2ο	DR (%)	RA (%)	FM (%)
Ανάλυση Προβολών [73]	71818	71948	49449	68.85	68.73	68.79
RLSA [151]	71818	69065	56361	78.48	81.61	80.01
ABBYY FineReader Engine 8.1 [127]	71818	74721	52782	73.49	70.64	72.04
OCROpus [137]	71818	79575	53648	74.70	67.42	70.87
Προτεινόμενη Μέθοδος	71818	75955	62425	86.92	82.19	84.49

glauben gehalten. Hinfurt ist mir bey gelegt die
Kron der Berechtigkeit / welche mir der Herr
an jenem tag / der Berechte Richter / geben
wird. Nicht mir aber allein / sondern auch alle
men die seine Erscheinung lieb haben.

(α)

glauben gehalten. Hinfurt ist mir bey gelegt die
Kron der Berechtigkeit / welche mir der Herr
an jenem tag / der Berechte Richter / geben
wird. Nicht mir aber allein / sondern auch alle
men die seine Erscheinung lieb haben.

(β)

glauben gehalten. Hinfurt ist mir bey gelegt die
Kron der Berechtigkeit / welche mir der Herr
an jenem tag / der Berechte Richter / geben
wird. Nicht mir aber allein / sondern auch alle
men die seine Erscheinung lieb haben.

(γ)

glauben gehalten. Hinfurt ist mir bey gelegt die
Kron der Berechtigkeit / welche mir der Herr
an jenem tag / der Berechte Richter / geben
wird. Nicht mir aber allein / sondern auch alle
men die seine Erscheinung lieb haben.

(δ)

glauben gehalten. Hinfurt ist mir bey gelegt die
Kron der Berechtigkeit / welche mir der Herr
an jenem tag / der Berechte Richter / geben
wird. Nicht mir aber allein / sondern auch alle
men die seine Erscheinung lieb haben.

(ε)

glauben gehalten. Hinfurt ist mir bey gelegt die
Kron der Berechtigkeit / welche mir der Herr
an jenem tag / der Berechte Richter / geben
wird. Nicht mir aber allein / sondern auch alle
men die seine Erscheinung lieb haben.

(στ)

Σχήμα 6.8: (α) Αρχικό έγγραφο, αποτελέσματα κατάτμησης χαρακτήρων χρησιμοποιώντας: (β) ανάλυση προβολών [73], (γ) RLSA [151], (δ) ABBYY FineReader Engine 8.1 [127], (ε) OCROpus [137] και (στ) την προτεινόμενη μέθοδο.

7. ΜΕΘΟΔΟΣ ΓΡΗΓΟΡΗΣ ΔΗΜΙΟΥΡΓΙΑΣ GROUND-TRUTH

7.1 Εισαγωγή

Σε αυτό το κεφάλαιο παρουσιάζουμε μία νέα μέθοδο για τη δημιουργία ground-truth με σκοπό την αξιολόγηση μεθόδων κατάτμησης η οποία έχει δημοσιευθεί στο [152]. Η μέθοδος χρησιμοποιεί την πληροφορία από το αρχείο κειμένου με σκοπό να κατευθύνει συγκεκριμένες τεχνικές κατάτμησης ώστε να παράγουν καλύτερα αποτελέσματα και να μειωθεί σημαντικά ο χρόνος που χρειάζεται ο χρήστης για να δημιουργήσει το ground-truth. Απαραίτητη προϋπόθεση για την εφαρμογή της προτεινόμενης μεθοδολογίας είναι το αρχείο κειμένου να περιέχει την πληροφορία για τον διαχωρισμό των γραμμών κειμένου. Τέλος, με μια απλή αντιστοίχιση ένα προς ένα προκύπτει η αντιστοίχιση των γραμμών κειμένου, των λέξεων και των χαρακτήρων της εικόνας με το αρχείο κειμένου. Επίσης, η προτεινόμενη μέθοδος, εκτός από τα ιστορικά έγγραφα, δοκιμάστηκε και σε σύγχρονα χειρόγραφα έγγραφα τα αποτελέσματα της οποίας έχουν δημοσιευθεί στο [153].

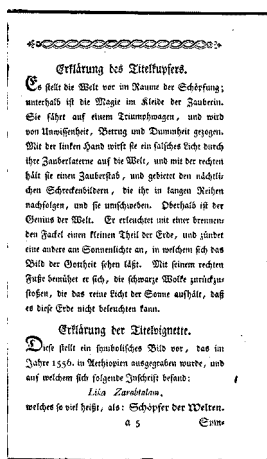
7.2 Μέθοδος Δημιουργίας Ground-truth

Αρχικά, εφαρμόζουμε ένα στάδιο προ-επεξεργασίας με στόχο να παράγουμε μία δυαδική εικόνα όπου θα διατηρούνται μόνο οι περιοχές κειμένου. Περιλαμβάνει τη δυαδική μετατροπή της εικόνας, εάν χρειάζεται, με τη χρήση της μεθόδου [6] καθώς και την αφαίρεση του περιθωρίου χρησιμοποιώντας την τεχνική που περιγράψαμε στο Κεφάλαιο 3 (Σχήμα 7.1). Με μία απλή ανάλυση του αρχείου κειμένου υπολογίζουμε τον αριθμό των γραμμών κειμένου της εικόνας καθώς και τον αριθμό των λέξεων και των χαρακτήρων για κάθε γραμμή κειμένου. Με τη βοήθεια αυτής της πληροφορίας τεχνικές της τρέχουσας τεχνολογικής στάθμης για την κατάτμηση γραμμών κειμένου, λέξεων και χαρακτήρων καθοδηγούνται με σκοπό την παραγωγή καλύτερων αποτελεσμάτων. Για την κατάτμηση των γραμμών κειμένου βασιζόμαστε σε μία τεχνική που χρησιμοποιεί τον μετασχηματισμό Hough, ενώ για την κατάτμηση των λέξεων και των χαρακτήρων χρησιμοποιούμε τεχνικές ταξινόμησης των αποστάσεων. Τέλος, ο χρήστης παρεμβαίνει για τη διόρθωση των αποτελεσμάτων και τη δημιουργία του τελικού ground-truth. Το διάγραμμα ροής της προτεινόμενης μεθόδου παρουσιάζεται στο Σχήμα 7.2. Όπως αποδεικνύουν τα πειραματικά αποτελέσματα μόνο ένας μικρός αριθμός λαθών χρειάζονται διόρθωση, συνεπώς η δημιουργία του ground-truth

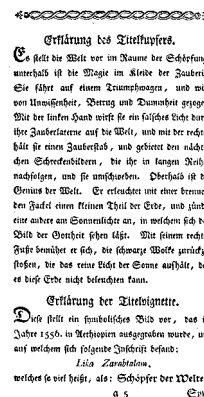
γίνεται πολύ πιο εύκολα και γρήγορα. Στη συνέχεια περιγράφονται αναλυτικά όλα τα στάδια της προτεινόμενης μεθόδου.



(α)

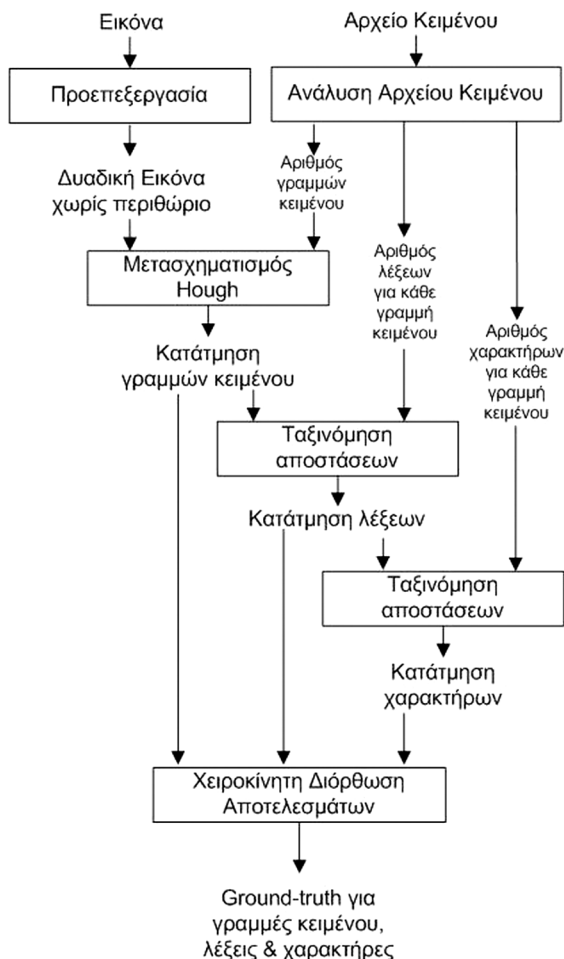


(β)



(γ)

Σχήμα 7.1: (α) Εικόνα ιστορικού εγγράφου, αποτέλεσμα μετά (β) τη δυαδική μετατροπή και (γ) την αφαίρεση του περιθωρίου.



Σχήμα 7.2: Διάγραμμα ροής της προτεινόμενης μεθόδου.

7.2.1 Ανάλυση Αρχείου Κειμένου

Το αρχείο κειμένου περιέχει χρήσιμη πληροφορία για την κατάτμηση της εικόνας σε γραμμές κειμένου, λέξεις και χαρακτήρες. Θεωρούμε ότι διαθέτουμε ένα αρχείο

κειμένου για κάθε εικόνα εγγράφου το οποίο περιέχει την πληροφορία για τον διαχωρισμό των γραμμών κειμένου. Συνεπώς, με μία γρήγορη και εύκολη ανάλυση του αρχείου κειμένου υπολογίζουμε τον αριθμό των γραμμών κειμένου NL που εμφανίζονται στην εικόνα, καθώς και τον αριθμό των λέξεων NW_i και των χαρακτήρων NC_i για κάθε γραμμή κειμένου i .

7.2.2 Κατάτμηση Γραμμών Κειμένου

Για την κατάτμηση των γραμμών κειμένου χρησιμοποιούμε μία παραλλαγή της μεθόδου που παρουσιάζεται στο [154] η οποία λαμβάνει υπόψη της τον αριθμό των γραμμών κειμένου NL που εμφανίζονται στην εικόνα. Περιλαμβάνει τρία στάδια: (i) προ-επεξεργασία, (ii) μετασχηματισμός Hough και (iii) μετεπεξεργασία.

(i) Προ-επεξεργασία

Στο στάδιο της προ-επεξεργασίας εφαρμόζεται ανάλυση συνδεδεμένων συστατικών με στόχο την απομάκρυνση αντικειμένων που δεν ανήκουν στο κείμενο. Αρχικά, υπολογίζεται το κυρίαρχο ύψος γράμματος (AH) φτιάχνοντας το ιστόγραμμα των υψών όλων των πλαισίων που περικλείουν τα συνδεδεμένα συστατικά της εικόνας [80] και έπειτα επιτυγχάνεται εξομάλυνση της εικόνα με τον αλγόριθμο RLSA και κατώφλι $Th = 0.4 \cdot AH$. Τέλος, αφαιρούνται τα συνδεδεμένα συστατικά που ικανοποιούν μία από τις παρακάτω συνθήκες:

$$(CS_w > 0.6I_w) \text{ ή } (CS_h > 5AH) \ \& \ (\frac{n_b}{CI_w CI_h} > 0.5) \text{ και } (\frac{CI_w}{CI_h} > 3) \text{ και } (CI_w > 3AH)$$

όπου I_w το πλάτος της εικόνας, CS_w, CS_h το πλάτος και το ύψος του αντικειμένου μετά την εξομάλυνση, n_b, CI_w, CI_h το πλήθος των pixels κειμένου, το πλάτος και το ύψος του αντικειμένου πριν την εξομάλυνση. Με την εφαρμογή της πρώτης συνθήκης αφαιρούνται τα μεγάλα αντικείμενα, όπως σχέδια και εικόνες (Σχήμα 7.3(α)), ενώ με τη δεύτερη συνθήκη αφαιρούνται μικρές σχεδιαστικές γραμμές που εμφανίζονται συχνά σε ιστορικά έγγραφα (Σχήμα 7.3(β)).

Inhalt.

Bermischte, ältere und neue elektrische Experimete.	Seite 425
Neuerfundene magnetische Experimente.	431
Weitere Beschreibung magnetischer Verische, die bekannt sind.	437
Bermischte, theils mechanische, theils chemische Verische.	446
Die Zauberkunst durch Wunderkräfte und Erfindungen, die durch Raubkunst, optische Täuschung, und Charlatanerie werden.	471
Uebersicht der Geschwindigkeit und Beschleunigung der Luft.	472
Wie man die Muffate verfertigen soll.	474
Einige Grundzüge zu Kartens-Kunststücken.	479



1. Seite

Symbolische Erklärung.

Das Haupt der Verische sind den Schicksal der Welt vor -- den Ereignissen, die alles schuf -- die Einheit. Die drei Klammern, die sich dazu umgeben, sind die symbolische Zahl der Vollkommenheit. Die verischen Klammern, welche und sich in sich selbst der Körperlichkeit, den Gedanken, den Geist der Seele in der Einheit. Was, Zahl und Ordnung in sich selbst der Körperlichkeit, und in sich selbst der Seele, Ordnung, Ordnung und Welt. Die neuen Verische sind die symbolische Zeichen der ersten Zahl, die aus der Einheit entsteht -- die Zahl der Schöpfung -- das Symbol der Produktion. Welt und Seele liegen auf der Erde, und sind mit einander verbunden. Die Klammern, die diese Verische ausfüllen, ist das Symbol der Harmonie. Es ruht auf 7 Werten, die die Klammern der 7 Harmonien der Natur sind, und welche 7 Töne verischen. Die neue Klammern des Instrumentes sind das Symbol des Tetrachordoms, durch welches das erste Musik-Instrument gefunden wird, und wozu die Klammern alle Klammern der Harmonie in der Zahl 4 verischen. Es ist auch das Symbol der Mäßigkeit der Dinge: das mathematische Punkt, der Linie, des Plans und der Erde. Die ganze Natur drückt dieses Hieroglyph aus, nämlich: Die Weisheit, die Weisheit, die Weisheit, die Weisheit und Weisheit der Dinge.

21

(α)

(β)

Σχήμα 7.3: Ιστορικά έγγραφα που περιέχουν (α) εικόνες και (β) σχεδιαστικές γραμμές.

(ii) Μετασχηματισμός Hough

Το βασικό στάδιο της διαδικασίας κατάτμησης των γραμμών είναι η εφαρμογή του μετασχηματισμού Hough όπως περιγράφεται στην εργασία [154]. Ο μετασχηματισμός Hough αντιστοιχίζει τα σημεία μίας ευθείας του επιπέδου x-y σε ένα σημείο (ρ, θ) του επιπέδου των πολικών συντεταγμένων. Μία ευθεία του επιπέδου x-y περιγράφεται από τη σχέση:

$$x \cos(\theta) + y \sin(\theta) = \rho$$

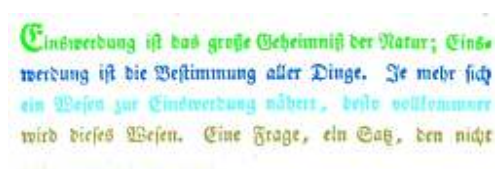
όπου ρ η κάθετη απόσταση της αρχής των αξόνων από την ευθεία και θ η γωνία που σχηματίζει η κάθετη στην ευθεία με τον άξονα x. Για όλα τα σημεία της εικόνας υπολογίζεται ο μετασχηματισμός Hough όλων των ευθειών που περνούν από αυτά και τα εξαγόμενα ζευγάρια τιμών (ρ, θ) αποθηκεύονται σε ένα πίνακα συσσώρευσης ο οποίος δημιουργείται με κατάλληλο βήμα κβάντισης των ρ, θ . Για το ρ θέτουμε βήμα $0.2 \cdot \Delta H$, ενώ για το θ θέτουμε βήμα 1 μοίρα και παίρνει τιμές μεταξύ 85 και 95 μοιρών.

Μετά τον υπολογισμό του πίνακα συσσώρευσης ακολουθεί η εξής διαδικασία: Εντοπίζουμε τα κελιά (ρ_i, θ_i) με τη μέγιστη συμβολή και αντιστοιχίζουμε στη γραμμή (ρ_i, θ_i) όλα τα σημεία που συνεισφέρουν στην περιοχή $(\rho_i - 5, \theta_i) \dots (\rho_i + 5, \theta_i)$. Στην εργασία [154], ο αριθμός των κελιών (ρ_i, θ_i) με τη μέγιστη συμβολή προσδιορίζεται σύμφωνα με ένα κριτήριο διακοπής. Στην προτεινόμενη μέθοδο, το κριτήριο διακοπής είναι ο αριθμός των κελιών (ρ_i, θ_i) με τη μέγιστη συμβολή να μη ξεπερνάει τον αριθμό των γραμμών κειμένου NL που

εμφανίζονται στην εικόνα. Συνεπώς, καθοδηγούμε τον μετασχηματισμό Hough να εντοπίσει NL γραμμές κειμένου.

(iii) Μετεπεξεργασία

Το τελικό στάδιο για την κατάτμηση των γραμμών κειμένου αποτελείται από δύο βήματα. Στο πρώτο βήμα λαμβάνουν χώρα συγχωνεύσεις μεταξύ γραμμών που εντοπίστηκαν από τον μετασχηματισμό Hough για να διορθωθούν μερικά λάθη όπως περιγράφεται στην εργασία [154]. Αυτό, όπως είναι φυσικό, οδηγεί σε μείωση των γραμμών που εντοπίστηκαν και άρα σε μικρότερο αριθμό από τον επιθυμητό αριθμό γραμμών NL . Συνεπώς, στο δεύτερο βήμα, συνδεδεμένα συστατικά που δεν αντιστοιχούν σε καμία γραμμή ελέγχονται μήπως μπορούν να δημιουργήσουν μία καινούρια γραμμή την οποία δε μπόρεσε να εντοπίσει ο μετασχηματισμός Hough. Καθοδηγούμε λοιπόν τον αλγόριθμο να δημιουργήσει τόσες γραμμές όσες χρειάζονται για να φτάσουμε τον επιθυμητό αριθμό NL . Μετά τη δημιουργία των νέων γραμμών κειμένου, όλα τα αταξινόμητα συνδεδεμένα συστατικά ενσωματώνονται στην πλησιέστερη γραμμή. Αν και ο αλγόριθμος καθοδηγείται να εντοπίσει NL γραμμές, υπάρχουν περιπτώσεις στις οποίες εντοπίζει λιγότερες καθώς δεν υφίστανται αταξινόμητα συνδεδεμένα συστατικά για τη δημιουργία νέας γραμμής. Στο Σχήμα 7.4 παρουσιάζονται αποτελέσματα κατάτμησης γραμμών κειμένου σε ένα ιστορικό έγγραφο καθώς και σε ένα χειρόγραφο έγγραφο.



(α)

gültiges Fehlurteil gelassen umj. bis zur Hinrichtung durch den Schiedsrichter beschäftigt. Ich und die zu Besuch im Gefängnis weilenden Freunde und Schüler philosophische Fragen.

(β)

Σχήμα 7.4: Κατάτμηση γραμμών κειμένου σε (α) τυπωμένο ιστορικό και (β) χειρόγραφο έγγραφο σύμφωνα με την προτεινόμενη μέθοδο.

7.2.3 Κατάτμηση Λέξεων

Η μέθοδος κατάτμησης των λέξεων βασίζεται επίσης στην τεχνική που περιγράφεται στο [154] και αποτελείται από δύο στάδια. Στο πρώτο στάδιο υπολογίζονται οι αποστάσεις μεταξύ των γειτονικών συστατικών μίας γραμμής κειμένου, ενώ στο δεύτερο στάδιο οι αποστάσεις ταξινομούνται σε αποστάσεις μεταξύ λέξεων ή αποστάσεις μεταξύ χαρακτήρων. Η μέθοδος καθοδηγείται ώστε να εντοπίσει NW_i λέξεις σε κάθε i γραμμή κειμένου.

(i) Υπολογισμός Αποστάσεων

Αρχικά, διορθώνουμε την κλίση της γραμμής και στη συνέχεια υπολογίζουμε τις αποστάσεις ανάμεσα στα επικαλυπτόμενα συστατικά. Ως επικαλυπτόμενα συστατικά θεωρούμε τα συνδεδεμένα συστατικά που έχουν κάθετη επικάλυψη ώστε να αποφύγουμε τις αποστάσεις μεταξύ χαρακτήρων και τόνων ή τμημάτων σπασμένων χαρακτήρων.

Για τον υπολογισμό της απόστασης ανάμεσα σε δύο επικαλυπτόμενα συστατικά υπολογίζουμε την Ευκλείδεια απόσταση ανάμεσα σε όλα τα ζευγάρια των pixels κειμένου και κρατάμε τη μικρότερη απόσταση. Για να επιταχύνουμε τον υπολογισμό της απόστασης δε χρησιμοποιούμε όλα τα pixels κειμένου κάθε συστατικού αλλά ένα υποσύνολο που αποτελούν το αριστερό και δεξιό προφίλ κάθε συστατικού, αντίστοιχα.

(ii) Ταξινόμηση αποστάσεων

Για την ταξινόμηση των αποστάσεων χρησιμοποιούμε ένα ξεχωριστό κατώφλι για κάθε γραμμή κειμένου. Όλες οι αποστάσεις που είναι μεγαλύτερες από το κατώφλι θεωρούνται αποστάσεις μεταξύ λέξεων ενώ οι υπόλοιπες θεωρούνται αποστάσεις μεταξύ χαρακτήρων. Για τον υπολογισμό του κατωφλίου στη γραμμή i βασιζόμαστε στον αναμενόμενο αριθμό λέξεων NW_i της γραμμής.

Έστω L το πλήθος των επικαλυπτόμενων συστατικών της γραμμής i . Συνεπώς, ο συνολικός αριθμός των αποστάσεων είναι $L-1$, οι οποίες ορίζονται ως $d_j, j=1, \dots, L-1$ και τις ταξινομούμε με φθίνουσα σειρά. Θεωρούμε την πρώτη απόσταση, δηλαδή τη μεγαλύτερη, ως το κατώφλι για τη γραμμή i και εντοπίζουμε τις λέξεις που προκύπτουν σύμφωνα με αυτό. Εάν ο αριθμός των λέξεων είναι ίσος ή μεγαλύτερος από τον αναμενόμενο αριθμό NW_i σταματάμε και ορίζουμε αυτή την απόσταση ως κατώφλι της γραμμής. Διαφορετικά, θεωρούμε την επόμενη απόσταση ως κατώφλι και επαναλαμβάνουμε τη διαδικασία μέχρι να ικανοποιηθεί η συνθήκη. Στο Σχήμα 7.5 παρουσιάζεται ένα παράδειγμα κατάτμησης λέξεων χρησιμοποιώντας διαφορετικά κατώφλια. Επιπλέον, στο Σχήμα 7.6 παρουσιάζονται αποτελέσματα κατάτμησης λέξεων σε ένα ιστορικό έγγραφο καθώς και σε ένα χειρόγραφο έγγραφο.

to d_1 arrive d_2 from d_3 ambiguity

(α)

to arrive from ambiguity

(β)

to arrive from ambiguity

(γ)

to arrive from ambiguity

(δ)

Σχήμα 7.5: Παράδειγμα κατάτμησης λέξεων χρησιμοποιώντας διαφορετικά κατώφλια. (α) αρχική εικόνα με τις τρεις μεγαλύτερες αποστάσεις (d_1, d_2, d_3). Αποτέλεσμα που προκύπτει χρησιμοποιώντας ως κατώφλι την απόσταση (β) d_1 , (γ) d_2 και (δ) d_3 . Το κατώφλι d_3 οδηγεί στο σωστό αποτέλεσμα καθώς ο αναμενόμενος αριθμός λέξεων NW_i είναι 4.

Einsteins ist das große Geheimnis. Einsteins ist die Bestimmung aller Dinge. Er ist ein Wesen zur Einwirkung. Er ist die vollkommene Welt dieses Wesens. Eine Frage, ein Satz, den nicht alle Wesen verstehen.

(α)

gültiges Fehlerteil. Einsteins ist die Bestimmung aller Dinge. Er ist ein Wesen zur Einwirkung. Er ist die vollkommene Welt dieses Wesens. Eine Frage, ein Satz, den nicht alle Wesen verstehen.

(β)

Σχήμα 7.6: Κατάτμηση λέξεων σε (α) τυπωμένο ιστορικό και (β) χειρόγραφο έγγραφο σύμφωνα με την προτεινόμενη μέθοδο.

7.2.4 Κατάτμηση Χαρακτήρων

Για την κατάτμηση των χαρακτήρων χρησιμοποιούμε την ίδια διαδικασία την οποία χρησιμοποιήσαμε και για τις λέξεις. Δύο είναι οι βασικές διαφορές: (i) οι αποστάσεις υπολογίζονται ανάμεσα στα συνδεδεμένα συστατικά και όχι τα επικαλυπτόμενα συστατικά της γραμμής κειμένου, και (ii) για τον υπολογισμό του κατωφλίου στη γραμμή i βασιζόμαστε στον αναμενόμενο αριθμό χαρακτήρων NC_i της γραμμής αντί για NW_i . Η διαδικασία κατάτμησης χαρακτήρων εφαρμόζεται μόνο σε τυπωμένα έγγραφα καθώς για να είναι δυνατή η εύρεση των αποστάσεων μεταξύ των χαρακτήρων δεν πρέπει η γραφή να είναι συνεχόμενη και άρα οι χαρακτήρες ενωμένοι. Στο Σχήμα 7.7 παρουσιάζεται ένα αποτέλεσμα κατάτμησης χαρακτήρων σε ένα ιστορικό έγγραφο.



Σχήμα 7.7: Κατάτμηση χαρακτήρων σε τυπωμένο ιστορικό έγγραφο σύμφωνα με την προτεινόμενη μέθοδο.

7.2.5 Διόρθωση Αποτελεσμάτων

Για τη δημιουργία του τελικού ground-truth ο χρήστης διορθώνει τα αποτελέσματα που προέκυψαν από την κατάτμηση των γραμμών κειμένου, των λέξεων και των χαρακτήρων. Η διόρθωση των αποτελεσμάτων γίνεται με τη βοήθεια ειδικού εργαλείου το οποίο επιτρέπει στον χρήστη να επεξεργαστεί, να εισάγει ή να διαγράψει περιοχές (Σχήμα 7.8). Όπως προκύπτει από τα πειραματικά αποτελέσματα οι διορθώσεις που χρειάζονται είναι ελάχιστες, καθώς η προτεινόμενη μέθοδος έχει αποδειχτεί αποτελεσματική, συνεπώς η δημιουργία του ground-truth γίνεται γρήγορα και εύκολα.



Σχήμα 7.8: Εργαλείο διόρθωσης αποτελεσμάτων: (α) αποτέλεσμα κατάτμησης λέξεων (β) ground-truth μετά τη διόρθωση από τον χρήστη.

7.3 Πειραματικά Αποτελέσματα

Για την αξιολόγηση της προτεινόμενης μεθόδου χρησιμοποιήσαμε δύο σύνολα εικόνων. Το πρώτο σύνολο περιλαμβάνει ιστορικά έγγραφα από το βιβλίο Eckartshausen που δημοσιεύτηκε το 1788 και ανήκει στη βιβλιοθήκη Bavarian State Library [123], ενώ το δεύτερο περιλαμβάνει χειρόγραφες εικόνες από τον διαγωνισμό “ICDAR2009 Handwriting Segmentation Contest” [155]. Για την αξιολόγηση της τεχνικής κατάτμησης γραμμών κειμένου και λέξεων χρησιμοποιήσαμε 94 εικόνες εγγράφων από το βιβλίο Eckartshausen, στις οποίες μαρκάραμε χειροκίνητα 2647 γραμμές κειμένου και 18575 λέξεις, και τις 200 εικόνες εγγράφων του διαγωνισμού που αποτελούνται από 4034 γραμμές κειμένου και 29717 λέξεις. Για την αξιολόγηση της τεχνικής κατάτμησης χαρακτήρων χρησιμοποιήσαμε 18 αντιπροσωπευτικές εικόνες εγγράφων από το βιβλίο Eckartshausen στις οποίες μαρκάραμε χειροκίνητα 14667 χαρακτήρες. Η

διαδικασία αξιολόγησης βασίζεται στον υπολογισμό του Detection Rate (DR), Recognition Accuracy (RA) και F-Measure (FM) όπως περιγράφεται στην παράγραφο 5.3 με κατώφλι $T_a = 95\%$ για τις γραμμές κειμένου και $T_a = 90\%$ για τις λέξεις και τους χαρακτήρες.

Αξιολογήσαμε την απόδοση της προτεινόμενης μεθόδου πριν την επέμβαση του χρήστη για τη διόρθωση των λαθών και τη συγκρίναμε με τεχνικές της τρέχουσας τεχνολογικής στάθμης. Στο σύνολο των ιστορικών τυπωμένων εγγράφων εφαρμόσαμε δύο μεθόδους, η πρώτη βασίζεται στον RLSA [151] και η δεύτερη στην ανάλυση των προβολών [73], καθώς και τα πρόγραμμα ABBYY FineReader Engine 8.1 [127] και OCROpus [137]. Στο σύνολο των χειρόγραφων εικόνων εφαρμόσαμε επίσης τις δύο μεθόδους που βασίζονται στο RLSA [151] και στην ανάλυση των προβολών [73] καθώς και τις πέντε καλύτερες τεχνικές του διαγωνισμού “ICDAR2009 Handwriting Segmentation Contest” [155] (ILSP-LWSeg-09, PAIS, CMM, CUBS και CASIA-MSTSeg). Επιπλέον, εφαρμόσαμε και στα δύο σύνολο την τεχνική [154] στην οποία βασιστήκαμε χωρίς όμως να λαμβάνει υπόψη της την πληροφορία από το αρχείο κειμένου, με σκοπό να εξετάσουμε πόσο βελτιώνεται.

Αφού αξιολογήσαμε την απόδοση της προτεινόμενης μεθόδου πριν την επέμβαση του χρήστη, στη συνέχεια, χρονομετρήσαμε το συνολικό χρόνο που απαιτείται για τη δημιουργία του ground-truth για τις εξής περιπτώσεις: (i) την εξολοκλήρου χειροκίνητη δημιουργία, (ii) αφού εφαρμόσουμε πρώτα μία τεχνική της τρέχουσας τεχνολογικής στάθμης και μετά διορθώσουμε τα λάθη και (iii) αφού εφαρμόσουμε πρώτα την προτεινόμενη μέθοδο και μετά διορθώσουμε τα λάθη. Για τη χειροκίνητη δημιουργία του ground-truth για τις γραμμές, τις λέξεις και τους χαρακτήρες στο πρώτο σύνολο, καταγράψαμε ότι ο χρήστης χρειάζεται για μία εικόνα 420, 900 και 5400 δευτερόλεπτα, αντίστοιχα. Ανάλογα, για τη χειροκίνητη δημιουργία του ground-truth για τις γραμμές και τις λέξεις στο δεύτερο σύνολο, καταγράψαμε ότι ο χρήστης χρειάζεται για μία εικόνα 280 και 600 δευτερόλεπτα, αντίστοιχα. Επίσης, υπολογίσαμε ότι ο μέσος χρόνος που χρειάζεται για να διορθώσει ένα λάθος στο αποτέλεσμα των λέξεων και των χαρακτήρων είναι 5 δευτερόλεπτα, ενώ για να διορθώσει ένα λάθος στο αποτέλεσμα των γραμμών είναι 10 δευτερόλεπτα, χρησιμοποιώντας το εργαλείο που περιγράψαμε στην παράγραφο 7.2.5. Επιπλέον, λαμβάνουμε υπόψη ότι ο μέσος χρόνος για τον οπτικό έλεγχο ενός αποτελέσματος κατάτμησης γραμμών κειμένου, λέξεων και

χαρακτήρων είναι 20, 40 και 80 δευτερόλεπτα ανά εικόνα για το πρώτο σύνολο και 13 και 27 δευτερόλεπτα ανά εικόνα για το δεύτερο σύνολο. Τέλος, θεωρούμε ότι έχοντας το σωστό αποτέλεσμα κατάτμησης έχουμε αυτόματα και την αντιστοίχιση της εικόνας με το αρχείου κειμένου.

Στους Πίνακες 7.1 και 7.2 παρουσιάζονται αναλυτικά όλα τα αποτελέσματα από την αποτίμηση της κατάτμησης της προτεινόμενης μεθόδου και των τεχνικών της τρέχουσας τεχνολογικής στάθμης καθώς και ο χρόνος που χρειάζεται ο χρήστης για να δημιουργήσει το ground-truth ξεκινώντας κάθε φορά από το αντίστοιχο αποτέλεσμα. Όπως παρατηρούμε, ο χρόνο για τη δημιουργία του ground-truth μειώνεται κατά 93% για τα ιστορικά τυπωμένα έγγραφα και κατά 92% για τα χειρόγραφα σε σχέση με την εξολοκλήρου χειροκίνητη δημιουργία. Εάν συγκρίνουμε την προτεινόμενη μέθοδο με την καλύτερη τεχνική της τρέχουσας τεχνολογικής στάθμης σε κάθε σύνολο, ο χρόνο μειώνεται κατά 29% και 12% για το πρώτο και δεύτερο σύνολο, αντίστοιχα. Επίσης, παρατηρούμε ότι η τεχνική [154] βελτιώνεται αρκετά με τη βοήθεια της επιπλέον πληροφορίας από το αρχείο κειμένου.

Πίνακας 7.1: Συγκριτικά αποτελέσματα για το σύνολο των ιστορικών εγγράφων.

	Κατάτμηση Γραμμών				Κατάτμηση Λέξεων				Κατάτμηση Χαρακτήρων				Χρόνος Δημιουργίας Ground-truth
	ο2ο	DR (%)	RA (%)	FM (%)	ο2ο	DR (%)	RA (%)	FM (%)	ο2ο	DR (%)	RA (%)	FM (%)	
Χειροκίνητη δημιουργία	-	-	-	-	-	-	-	-	-	-	-	-	61.5 ώρες
RLSA [151]	2516	95.0	92.2	93.6	13230	71.2	86.5	78.1	12686	86.5	84.3	85.4	12.5 ώρες
Ανάλυση προβολών [73]	2382	90.0	92.5	91.2	16175	87.1	91.8	89.4	11697	79.7	75.4	77.5	10.2 ώρες
OCROpus [137]	2448	92.5	93.2	92.8	17972	96.7	87.1	91.7	11851	80.8	71.3	75.7	7.2 ώρες
ABBYY FineReader [127]	2593	97.9	98.6	98.2	17842	96.0	94.6	95.2	12470	85.0	78.1	81.4	6.2 ώρες
Hough - Gap Classification [154]	2351	89.8	94.7	92.2	18058	97.2	89.5	93.1	-	-	-	-	-
Προτεινόμενη Μέθοδος	2618	98.9	98.9	98.9	18248	98.2	98.2	98.2	13291	90.6	90.1	90.3	4.4 ώρες

Πίνακας 7.2: Συγκριτικά αποτελέσματα για το σύνολο των χειρόγραφων εγγράφων.

	Κατάτμηση Γραμμών				Κατάτμηση Λέξεων				Χρόνος Δημιουργίας Ground-truth
	ο2ο	DR (%)	RA (%)	FM (%)	ο2ο	DR (%)	RA (%)	FM (%)	
Χειροκίνητη δημιουργία	-	-	-	-	-	-	-	-	48.9 ώρες
Ανάλυση προβολών [73]	2538	62.9	57.8	60.2	20143	67.8	52.5	59.2	19.7 ώρες
RLSA [151]	1615	40.0	38.9	39.4	24006	80.8	77.7	79.2	16.9 ώρες
CASIA-MSTSeg [155]	3867	95.9	95.5	95.7	25938	87.3	82.6	84.8	7.9 ώρες
CUBS [155]	4016	99.6	99.5	99.5	26631	89.6	84.4	87.0	6.6 ώρες
CMM [155]	3975	98.5	98.3	98.4	27078	91.1	86.8	88.9	6.0 ώρες
PAIS [155]	3973	98.5	98.6	98.5	27288	91.8	89.3	90.5	5.8 ώρες
ILSP-LWSeg-09 [155]	4000	99.2	98.9	99.0	28279	95.1	94.4	94.8	4.3 ώρες
Hough - Gap Classification [154]	3898	96.6	96.5	95.6	28138	94.7	93.7	94.2	4.8 ώρες
Προτεινόμενη Μέθοδος	3913	97.0	97.1	97.1	28845	97.1	97.2	97.1	3.8 ώρες

8. ΣΥΜΠΕΡΑΣΜΑΤΑ

Στις μέρες μας η επεξεργασία και η ανάλυση σύγχρονων εγγράφων γίνεται με μεγάλη ακρίβεια, γεγονός που πιστοποιείται και από την πληθώρα εμπορικών προϊόντων που υπάρχουν. Αντιθέτως, στην περίπτωση των ιστορικών εγγράφων τα ποσοστά επιτυχίας δεν είναι το ίδιο ικανοποιητικά. Υπάρχουν προβλήματα που δυσκολεύουν την αυτόματη επεξεργασία και αναγνώριση ιστορικών εγγράφων. Εκτός από τα ιδιαίτερα χαρακτηριστικά, όπως ασυνήθιστες γραμματοσειρές και πολύπλοκες δομές εγγράφων, συναντάμε προβλήματα και στην ποιότητα της εικόνας λόγω παλαιότητας ή χαμηλής ποιότητας του εγγράφου ή λόγω κακής ψηφιοποίησης, αφού πολλές φορές δεν είναι εφαρμόσιμοι οι συνηθισμένοι τρόποι ψηφιοποίησης για λόγους προστασίας του αυθεντικού υλικού.

Η παρούσα διδακτορική διατριβή επιχειρεί να πάει ένα βήμα μπροστά την έρευνα σε αυτόν τον τομέα και επικεντρώνεται στα στάδια βελτίωσης της ποιότητας και κατάτμησης του εγγράφου. Ειδικότερα:

- Προτείνει μεθόδους για τον εντοπισμό και την αφαίρεση του περιθωρίου, ένα πρόβλημα που εμφανίζεται συχνά στα ιστορικά έγγραφα και επηρεάζει τα επόμενα στάδια επεξεργασίας. Ταυτόχρονα με το περιθώριο, τμήματα κειμένου γειτονικών σελίδων αφαιρούνται ή έγγραφα που περιέχουν δύο σελίδες διαχωρίζονται καθώς η πλειοψηφία των τεχνικών επεξεργασίας εγγράφων θεωρούν ότι επεξεργάζονται μία σελίδα κάθε φορά. Τα πειραματικά αποτελέσματα σε ποικίλα ιστορικά βιβλία επιβεβαιώνουν την αποτελεσματικότητα των μεθόδων καθώς καταφέρνουν να αφαιρέσουν το περιθώριο ενώ συγχρόνως δε χάνεται χρήσιμη πληροφορία.
- Προτείνει μία νέα μέθοδο για τη διόρθωση της τοπικής καμπυλότητας και των παραμορφώσεων που εμφανίζονται στα ιστορικά έγγραφα εξαιτίας της μορφής του υλικού προς ψηφιοποίηση, της διάταξης του εγγράφου κατά την ψηφιοποίηση ή λόγω περιβαλλοντικών συνθηκών και παλαιότητας του εγγράφου. Η αποτίμηση της μεθόδου βασίζεται, εκτός από τους έμμεσους τρόπους που συναντάμε στη βιβλιογραφία, και σε μία πρωτότυπη ημιαυτόματη μέθοδο που προτείνουμε η οποία δίνει τη δυνατότητα να αξιολογήσουμε και να συγκρίνουμε απεριόριστους μεθόδους με ελάχιστο κόπο από τον χρήστη. Η προτεινόμενη μέθοδος αξιολόγησης είναι ένα πρώτο βήμα για την απεξάρτηση από τους έμμεσους τρόπους αξιολόγησης και τις μηχανές οπτικής αναγνώρισης. Ωστόσο, χρειάζεται περαιτέρω εξέλιξη με

στόχο να λαμβάνει υπόψη της, πέρα από τη διόρθωση της καμπυλότητας των γραμμών κειμένου, και όλες τις άλλες παραμέτρους όπως είναι η διόρθωση της κλίμακας των χαρακτήρων, η κατακόρυφη ευθυγράμμιση κ.α. Επιπροσθέτως, αν και την τελευταία δεκαετία έχουν προταθεί πολλές μέθοδοι για τη διόρθωση της τοπικής καμπυλότητας, δεν υπάρχει η δυνατότητα άμεσης σύγκρισης των μεθοδολογιών καθώς δεν είναι διαθέσιμη μία κοινή βάση ιστορικών ή σύγχρονων εγγράφων. Η δημιουργία μίας βάσης με έγγραφα που παρουσιάζουν το πρόβλημα αυτό θα διευκολύνει την περαιτέρω εξέλιξη των μεθόδων.

- Παρουσιάζει μία πρωτότυπη μεθοδολογία συνδυασμού αποτελεσμάτων κατάτμησης των εγγράφων σε γραμμές κειμένου. Στη βιβλιογραφία υπάρχει πληθώρα μεθόδων κατάτμησης γραμμών κειμένου, πολλές εκ των οποίων είναι συμπληρωματικές υπό την έννοια ότι κάθε μέθοδος μπορεί να αντιμετωπίσει αποτελεσματικότερα διαφορετικές δυσκολίες. Η προτεινόμενη μεθοδολογία προσπαθεί να συνδυάσει συμπληρωματικές τεχνικές κατάτμησης ώστε να αντιμετωπιστούν τα πολλά και σύνθετα προβλήματα που συναντάμε στα ιστορικά έγγραφα. Τα πειραματικά αποτελέσματα είναι ενθαρρυντικά και θα πρέπει η μεθοδολογία να εξελιχθεί και να εφαρμοστεί και σε άλλα επίπεδα κατάτμησης, όπως οι λέξεις και οι χαρακτήρες, καθώς τα αποτελέσματα στα ιστορικά έγγραφα για κάθε μέθοδο ξεχωριστά είναι χαμηλά.
- Προτείνει μία μέθοδο κατάτμησης χαρακτήρων σε ιστορικά έγγραφα. Τα πειραματικά αποτελέσματα σε σημαντικό αριθμό ιστορικών εγγράφων, από διαφορετικές γλώσσες αλλά και χρονολογικές περιόδους επιβεβαιώνουν την αποτελεσματικότητα της μεθόδου σε σύγκριση με τεχνικές της τρέχουσας τεχνολογικής στάθμης καθώς επίσης και εμπορικά προγράμματα.
- Προτείνει μία μέθοδο με στόχο τη γρήγορη και εύκολη δημιουργία ground-truth για την αξιολόγηση μεθόδων κατάτμησης. Η δημιουργία του ground-truth είναι μία χρονοβόρα και επίπονη διαδικασία, και κυρίως για τα ιστορικά έγγραφα δεν υπάρχει διαθέσιμη καμία βάση. Από τα αποτελέσματα, σε ιστορικά τυπωμένα και σύγχρονα χειρόγραφα έγγραφα, φαίνεται ότι ο χρόνος που χρειάζεται ο χρήστης για να δημιουργήσει το ground-truth μειώνεται σημαντικά. Αν και παρουσιάζει εντυπωσιακά αποτελέσματα σε σχέση με τη μείωση του χρόνου που απαιτείται για τη δημιουργία του ground-truth, έχει τη

δυνατότητα να παράγει ακόμα καλύτερα αποτελέσματα χρησιμοποιώντας επιπρόσθετες πληροφορίες από το αρχείο κειμένου όπως για παράδειγμα το μήκος κάθε λέξης κ.α.

ΔΗΜΟΣΙΕΥΣΕΙΣ

A. ΠΕΡΙΟΔΙΚΑ

- A1. N. Stamatopoulos, B. Gatos, I. Pratikakis and S.J. Perantonis, “Goal-oriented Rectification of Camera-Based Document Images”, IEEE Transactions on Image Processing, vol. 20, no. 4, 2011, pp. 910-920.
- A2. N. Stamatopoulos, B. Gatos and S.J. Perantonis, “A Method for Combining Complementary Techniques for Document Image Segmentation”, Pattern Recognition, vol. 42, no. 12, 2009, pp. 3158-3168.
- A3. B. Gatos, N. Stamatopoulos and G. Louloudis, “ICDAR2009 Handwriting Segmentation Contest”, International Journal on Document Analysis and Recognition, vol. 14, no. 1, 2011, pp. 25-33.
- A4. N. Nikolaou, M. Makridis, B. Gatos, N. Stamatopoulos and N. Papamarkos, “Segmentation of historical machine-printed documents using Adaptive Run Length Smoothing and skeleton segmentation paths”, Image and Vision Computing, vol. 28, no. 4, 2010, pp. 590-604.

B. ΠΡΑΚΤΙΚΑ ΣΥΝΕΔΡΙΩΝ

- B1. B. Gatos, G. Louloudis and N. Stamatopoulos, “Greek Polytonic OCR based on Efficient Character Class Number Reduction”, 11th International Conference on Document Analysis and Recognition, Beijing, China, 2011.
- B2. G. Louloudis, N. Stamatopoulos and B. Gatos, “ICDAR 2011 - Writer Identification Contest”, 11th International Conference on Document Analysis and Recognition, Beijing, China, 2011.
- B3. N. Stamatopoulos, G. Louloudis and B. Gatos, “Efficient Transcript Mapping to Ease the Creation of Document Image Segmentation Ground Truth with Text-Image Alignment”, 12th International Conference on Frontiers in Handwriting Recognition, Kolkata, India, 2010, pp. 226-231.
- B4. B. Gatos, N. Stamatopoulos and G. Louloudis, “ICFHR 2010 Handwriting Segmentation Contest”, 12th International Conference on Frontiers in Handwriting Recognition, Kolkata, India, 2010, pp. 737-742.
- B5. N. Stamatopoulos, B. Gatos and T. Georgiou, “Page Frame Detection for Double Page Document Images”, 9th International Workshop on Document Analysis Systems, Boston, MA, USA, 2010, pp. 401-408.
- B6. G. Vamvakas, N. Stamatopoulos, B. Gatos and S.J. Perantonis, “Automatic Unsupervised Parameter Selection for Character Segmentation”, 9th

- International Workshop on Document Analysis Systems, Boston, MA, USA, 2010, pp. 409-415.
- B7. N. Stamatopoulos, B. Gatos and I. Pratikakis, "A Methodology for Document Image Dewarping Techniques Performance Evaluation", 10th International Conference on Document Analysis and Recognition, Barcelona, Spain, 2009, pp. 956-960.
- B8. G. Louloudis, N. Stamatopoulos and B. Gatos, "A Novel Two Stage Evaluation Methodology for Word Segmentation Techniques", 10th International Conference on Document Analysis and Recognition, Barcelona, Spain, 2009, pp. 686-690.
- B9. B. Gatos, N. Stamatopoulos and G. Louloudis, "ICDAR2009 Handwriting Segmentation Contest", 10th International Conference on Document Analysis and Recognition, Barcelona, Spain, 2009, pp. 1393-1397.
- B10. N. Stamatopoulos, G. Louloudis and B. Gatos, "A Comprehensive Evaluation Methodology for Noisy Historical Document Recognition Techniques", 3rd Workshop on Analytics for Noisy Unstructured Text Data, Barcelona, Spain, 2009, pp. 47-54.
- B11. N. Stamatopoulos, B. Gatos, I. Pratikakis and S.J. Perantonis, "A Two-Step Dewarping of Camera Document Images", 8th International Workshop on Document Analysis Systems, Nara, Japan, 2008, pp. 209-216.
- B12. G. Vamvakas, B. Gatos, N. Stamatopoulos and S.J. Perantonis, "A Complete Optical Character Recognition Methodology for Historical Documents", 8th International Workshop on Document Analysis Systems, Nara, Japan, 2008, pp. 525-532.
- B13. N. Stamatopoulos, B. Gatos and S.J. Perantonis, "A Method for Combining Complementary Techniques for Document Image Segmentation", 11th International Conference on Frontiers in Handwriting Recognition, Montreal, Canada, 2008, pp. 235-240.
- B14. N. Stamatopoulos, B. Gatos and A. Kesidis, "Automatic Borders Detection of Camera Document Images", 2nd International Workshop on Camera-Based Document Analysis and Recognition, Curitiba, Brazil, 2007, pp.71-78.
- B15. B. Gatos, A. Antonacopoulos and N. Stamatopoulos, "ICDAR2007 Handwriting Segmentation Contest", 9th International Conference on Document Analysis and Recognition, Curitiba, Brazil, 2007, pp. 1284-1288.

- B16. G. Vamvakas, B. Gatos, S. Petridis and N. Stamatopoulos, "An Efficient Feature Extraction and Dimensionality Reduction Scheme for Isolated Greek Handwritten Character Recognition", 9th International Conference on Document Analysis and Recognition, Curitiba, Brazil, 2007, pp. 1073-1077.
- B17. G. Vamvakas, N. Stamatopoulos, B. Gatos, I. Pratikakis and S.J. Perantonis, "Greek Handwritten Character Recognition", 11th Panhellenic Conference on Informatics, Patra, Greece, 2007, pp. 343-352.
- B18. G. Vamvakas, B. Gatos, I. Pratikakis, N. Stamatopoulos, A. Roniotis and S.J. Perantonis, "Hybrid Off-Line OCR for Isolated Handwritten Greek Characters", 4th IASTED International Conference on Signal Processing, Pattern Recognition, and Applications, Innsbruck, Austria, 2007, pp. 197-202.

ΕΤΕΡΟΑΝΑΦΟΡΕΣ ΣΤΙΣ ΔΗΜΟΣΙΕΥΣΕΙΣ

- M. Zhao, S. Li and J. Kwok, “Text detection in images using sparse representation with discriminative dictionaries”, *Image and Vision Computing*, vol. 28, no. 12, 2010, pp. 1590-1599.
- Y. Zhang and L. Wu, “Fast document image binarization based on an improved adaptive Otsu's method and destination word accumulation”, *Journal of Computational Information Systems*, vol. 7, no. 6, 2011, pp. 1886-1892.
- S. Pletschacher and A. Antonacopoulos, “The PAGE (Page Analysis and Ground-Truth Elements) Format Framework”, 20th International Conference on Pattern Recognition, Istanbul, Turkey, 2010, pp. 257-260.
- J. Kumar, L. Kang, D. Doermann and W. Abd-Almageed, “Segmentation of Handwritten Textlines in Presence of Touching Components”, 11th International Conference on Document Analysis and Recognition, Beijing, China, 2011.
- A. Alaei, P. Nagabhushan and U. Pal, “Piece-wise painting technique for line segmentation of unconstrained handwritten text: a specific study with Persian text documents”, *Pattern Analysis & Applications* DOI: 10.1007/s10044-011-0226-x, 2011.
- R. Sarkar, N. Das, S. Basu, M. Kundu, M. Nasipuri and D.K. Basu, “CMATERdb1: a database of unconstrained handwritten Bangla and Bangla-English mixed script document image”, *International Journal on Document Analysis and Recognition*, DOI: 10.1007/s10032-011-0148-6, 2011.
- E. Kavallieratou and F. Daskas, “Text Line Detection and Segmentation: Uneven Skew Angles and Hill-and-Dale Writing”, *Journal of Universal Computer Science*, vol. 17, no. 1, 2011, pp. 16-29.
- A. Lemaitre, J. Camillerapp and B. Coüason, “A perceptive method for handwritten text segmentation”, *Document recognition and retrieval XVIII - Electronic Imaging*, San Francisco, United States, No. 78740C, 2011.

- A. Alaei, U. Pal and P. Nagabhushan, "A new scheme for unconstrained handwritten text-line segmentation", *Pattern Recognition* vol. 44, no. 4, 2011, pp. 917-928.
- P. Nagabhushan and A. Alaei, "Tracing and Straightening the Baseline in Handwritten Persian/Arabic Text-line: A New Approach Based on Painting-technique", *International Journal on Computer Science and Engineering*, vol. 2, no. 4, 2010, pp. 907-916.
- V. Papavassiliou, V. Katsouros and G. Carayannis, "A Morphological Approach for Text-Line Segmentation in Handwritten Documents", 12th International Conference on Frontiers in Handwriting Recognition, Kolkata, India, 2010, pp. 19-24.
- D.M. Oliveira, R.D. Lins, G. Torreão, J. Fan and M. Thielo, "A new algorithm for segmenting warped text-lines in document images", *ACM Symposium on Applied Computing*, Taichung, Taiwan, 2011, pp. 259-265.
- S.S. Bukhari, F. Shafait and T.M. Breuel, "Performance Evaluation of Curled Textlines Segmentation Algorithms", 9th International Workshop on Document Analysis Systems, (short paper), Boston, MA, USA, 2010, pp. 555-558.
- R.D. Lins, D.M. Oliveira, G. Torreao, J. Fan and M. Thielo, "Correcting Book Binding Distortion in Scanned Documents", 7th International Conference on Image Analysis and Recognition, Portugal, 2010, pp. 355-365.
- D.M. Oliveira, R.D. Lins, G. Torreao, J. Fan and M. Thielo, "A New Method for Text-Line Segmentation for Warped Documents", 7th International Conference on Image Analysis and Recognition, Portugal, 2010, pp. 398-408.
- H.I. Koo and N.I. Cho, "State Estimation in a Document Image and Its Application in Text Block Identification and Text Line Extraction", 11th European Conference on Computer Vision, Crete, Greece, 2010, pp. 421-434.
- S.S. Bukhari, T.M. Breuel and F. Shafait, "Textline information extraction from grayscale camera-captured document images", 16th International Conference on Image Processing, Cairo, 2009, pp. 2013-2016.
- S.S. Bukhari, F. Shafait and T.M. Breuel, "Ridges based Curled Textline Region Detection from Grayscale Camera-Captured Document Images", 13th

International Conference on Computer Analysis of Images and Patterns, Münster, Germany, 2009, pp. 173-180.

- S.S. Bukhari, F. Shafait and T.M. Breuel, “Coupled Snakelet Model for Curled Textline Segmentation of Camera-Captured Document Images”, 10th International Conference on Document Analysis and Recognition, Barcelona, Spain, 2009, pp. 61-65.
- S.S. Bukhari, F. Shafait and T.M. Breuel, “Dewarping of document images using coupled-snakes”, International Workshop on Camera-Based Document Analysis and Recognition, Barcelona, Spain, 2009, pp. 34-41.
- M. Diem and R. Sablatnig, “Are Characters Objects?”, 12th International Conference on Frontiers in Handwriting Recognition, Kolkata, India, 2010, pp. 565-570.
- C. Colutto, “Introducing a new image dissimilarity measure with an application to character image clustering in degraded historical documents”, 9th International Workshop on Document Analysis Systems, Boston, USA, 2010, pp. 325-332.
- D.R. Leeandl and S. Oh, “Minimum-Cost Path Algorithm for Separating Touching Characters”, 7th IASTED International Conference on Signal Processing, Pattern Recognition, and Applications, Austria, 2010, pp. 164-168.
- M. Diem and R. Sablatnig, “Recognizing characters of ancient manuscripts”, Proceedings of SPIE - The International Society for Optical Engineering, vol. 7531, article number 753106, 2010.
- F. Shafait and T.M. Breuel, “The Effect of Border Noise on the Performance of Projection-Based Page Segmentation Methods”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 33, no. 4, 2011, pp. 846-851.
- M.M. Haji, T.D. Bui and C.Y. Suen, “Simultaneous Document Margin Removal and Skew Correction Based on Corner Detection in Projection Profiles”, 15th International Conference on Image Analysis and Processing, Italy, 2009, pp. 1025-1034.

- F. Shafait, D. Keysers and T.M. Breuel, "Response to "Projection Methods Require Black Border Removal", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, no. 4, 2009, pp.763-764.
- F. Shafait, J. Beusekom, D. Keysers and T.M. Breuel, "Document cleanup using page frame detection", International Journal on Document Analysis and Recognition, vol. 11, no. 2, 2008, pp. 81-96.
- A. Rehman and T. Saba, "Off-line cursive script recognition: current advances, comparisons and remaining problems", Artificial Intelligence Review, DOI 10.1007/s10462-011-9229-7, 2011.
- A. Sánchez, C.A.B. Mello, P.D. Suárez and A. Lopes, "Automatic line and word segmentation applied to densely line-skewed historical handwritten document images", Integrated Computer-Aided Engineering, vol. 18, no. 2, pp. 125-142, 2011.
- A. Rehman and T. Saba, "Performance analysis of character segmentation approach for cursive script recognition on benchmark database", Digital Signal Processing, vol. 21, no. 3, 2011, pp. 486-490.
- E. Kavallieratou, "Text line detection and segmentation: uneven skew angles and hill-and-dale writing", ACM Symposium on Applied Computing, Sierre, Switzerland, 2010, pp. 59-60.
- R. Doumat, E.E. Zsigmond and J.M. Pinon, "User Trace-Based Recommendation System for a Digital Archive", 8th International Conference on Case-Based Reasoning, Alessandria, Italy, 2010, pp. 360-374.
- N. Ouwayed, A. Belaïd and F. Auger, "General text line extraction approach based on locally orientation estimation", 17th Document Recognition and Retrieval Conference, San Jose, CA, United States, 2010, pp. 1-10.
- V. Papavassiliou, T. Stafylakis, V. Katsouros and G. Carayiannis, "Handwritten document image segmentation into text lines and words", Pattern Recognition, vol. 43, no. 1, 2010, pp. 369-377.
- F. Kurniawan and D. Mohamad, "Performance Comparison between Contour-Based and Enhanced Heuristic-Based for Character Segmentation", 5th International Conference on Signal-Image Technology & Internet-Based Systems, Marrakesh, Morocco, 2009, pp. 112-117.

- A. Khandelwal, P. Choudhury, R. Sarkar, S. Basu, M. Nasipuri and N. Das, “Text Line Segmentation for Unconstrained Handwritten Document Images Using Neighborhood Connected Component Analysis”, 3rd International Conference on Pattern Recognition and Machine Intelligence, New Delhi, India, 2009, pp. 369-374.
- S.S. Bukhari, F. Shafait and T.M. Breuel, “Script-Independent Handwritten Textlines Segmentation using Active Contours”, 10th International Conference on Document Analysis and Recognition, Barcelona, Spain, 2009, pp. 446-450.
- E. Saund, J. Lin and P. Sarkar, “PixLabeler: User Interface for Pixel-Level Labeling of Elements in Document Images”, 10th International Conference on Document Analysis and Recognition, Barcelona, Spain, 2009, pp. 646-650.
- R.P. Santos, G.S. Clemente, T.I. Ren and G.D.C. Cavalcanti, “Text Line Segmentation Based on Morphology and Histogram Projection”, 10th International Conference on Document Analysis and Recognition, Barcelona, Spain, 2009, pp. 651-655.
- A. Rehman, D. Mohamad, F. Kurniawan and M. Ilays, “Performance analysis of segmentation approach for cursive handwriting on benchmark database”, International Conference on Computer Systems and Applications, Rabat, Morocco, 2009, pp. 265-270.
- R. Doumat, E. Egyed-Zsigmond and J.M. Pinon, “Digitized ancient documents ... What's next?”, Document Numerique, vol. 12, no. 1, 2009, pp. 31-51.
- T. Stafylakis, V. Papavassiliou, V. Katsouros and G. Carayiannis, “Robust text-line and word segmentation for handwritten documents images”, International Conference on Acoustics, Speech and Signal Processing, Las Vegas, NV, USA, 2008, pp. 3393–3396.
- S.S. Bukhari, F. Shafait and T.M. Breuel, “Segmentation of Curled Textlines Using Active Contours”, 8th International Workshop on Document Analysis Systems, Nara, Japan, 2008, pp. 270-277.

- R. Doumat, E.E. Zsigmond, J.M. Pinon and E. Csiszar, “Online ancient documents: Armarius”, 8th ACM symposium on Document Engineering, Sao Paulo, Brazil, 2008, pp. 127-130.
- K.S. Siddharth, M. Jangid, R. Dhir and R. Rani, “Handwritten Gurmukhi Character Recognition Using Statistical and Background Directional Distribution”, International Journal on Computer Science and Engineering, vol. 3, no. 6, 2011, pp. 2332-2345.
- S. Dabra, S. Agrawal and R.K. Challa, “A novel feature set for recognition of similar shaped handwritten Hindi characters using machine learning”, 1st International Conference on Computer Science, Engineering and Applications, Chennai, India, 2011, pp. 25-35.
- H. Hamdi and M. Khemakhem, “Distributing Arabic Handwriting Recognition System Based on the Combination of Grid Meta-Scheduling and P2P Technologies (Omnivore)”, Universal Journal of Computer Science and Engineering Technology, vol. 1, no 1, 2010, pp. 31-35.
- P.A. Phuong, N.Q. Tao and L.C. Mai, “An Efficient Model for Isolated Vietnamese Handwritten Recognition”, International Conference on Intelligent Information Hiding and Multimedia Signal Processing, Harbin, China, 2008, pp. 358-361.

ΠΙΝΑΚΑΣ ΟΡΟΛΟΓΙΑΣ

Ξενόγλωσσος όρος	Ελληνικός Όρος
Baseline	Γραμμή Αναφοράς
Binarization	Δυαδική Μετατροπή
Border Removal	Αφαίρεση Περιθωρίου
Bottom-up technique	Τεχνική “από κάτω προς τα πάνω”
Break Points	Σημεία Κατάτμησης
Classification	Ταξινόμηση
Connected Components	Συνδεδεμένα Συστατικά
Connected Components Analysis	Ανάλυση Συνδεδεμένων Συστατικών
Content-based image retrieval	Ανάκτηση Εικόνας με βάση το Περιεχόμενο
Contour Points	Σημεία Περιγράμματος
Contrast	Αντίθεση φωτεινότητας
Crossing Points	Σημεία τομής
Detection Rate	Ποσοστό Εντοπισμού
Dewarping	Διόρθωση Τοπικής Καμπυλότητας
Dilation	Διαστολή
Distances	Αποστάσεις
Document Boundaries	Όρια Εγγράφου
Document Image Understanding	Κατανόηση Εικόνας Εγγράφου
Document Page Segmentation	Κατάτμηση Σελίδας Εγγράφου
Document Spectrum	Φάσμα Εγγράφου
Document Structure Analysis	Ανάλυση Δομής του Εγγράφου
Enhancement	Βελτίωση Ποιότητας
Erosion	Διάβρωση
Feature Extraction	Εξαγωγή Χαρακτηριστικών
Flood-fill algorithm	Αλγόριθμος πλημμύρας
Horizontal Histogram	Ιστόγραμμα Οριζόντιων Προβολών
Indexing	Δημιουργία Ευρετηρίου
Intersection Points	Σημεία Διακλάδωσης
Invading Algorithm	Αλγόριθμος Εισβολής
Portrait	Κατακόρυφος Προσανατολισμός
Laser Scanner	Σαρωτής Λείζερ
Linear Regression	Γραμμική Παλινδρόμηση
Logistic Regression	Λογιστική Παλινδρόμηση

Markov Random Field Models	Μοντέλα Τυχαίων Μαρκοβιανών Πεδίων
Nearest Neighbor Algorithm	Αλγόριθμος Πλησιέστερου Γείτονα
Non-Invading Algorithm	Αλγόριθμος Μη Εισβολής
Optical Character Recognition	Οπτική Αναγνώριση Χαρακτήρων
Orientation Features	Χαρακτηριστικά Προσανατολισμού
Page frame	Πλαίσιο Περιοχής Κειμένου
Pattern Recognition	Αναγνώριση Προτύπων
Pixel	Εικονοστοιχείο
Landscape	Οριζόντιος προσανατολισμός
Post-Processing	Μετεπεξεργασία
Pre-processing	Προ-επεξεργασία
Precision	Ακρίβεια
Process	Επεξεργασία
Profile points	Σημεία προφίλ
Projections	Προβολές
Recall	Αριθμός εύστοχων αποκρίσεων
Recognition Accuracy	Ακρίβεια Αναγνώρισης
Reference Points	Σημεία Αναφοράς
Region Adjacency Graph	Γράφος Περιοχών Γειτνίασης
Run Length Smoothing Algorithm	Αλγόριθμος Εξομάλυνσης Βασισμένος στο Διατρέχον Μήκος
Salt and pepper noise	Θόρυβος μορφής αλατοπίπερου
Scanner	Σαρωτής
Scanning	Σάρωση
Segmentation	Κατάτμηση
Signal cross-correlation	Διασυσχέτιση του σήματος
Skeletonization	Εξαγωγή Σκελετού
Skew Correction	Διόρθωση Στροφής
Smoothing	Εξομάλυνση Εικόνας
Stereo Camera	Στερεοσκοπική Φωτογραφική Μηχανή
Structured Elements	Δομικά Στοιχεία
Structured light 3D scanner	Τριδιάστατος Σαρωτής Δομημένου Φωτός
Tagging	Τοποθέτηση Ετικέτας
Terminal Points	Τερματικά σημεία
Text Line	Γραμμή Κειμένου
Threshold	Κατώφλι

Thresholding	Κατωφλίωση
Top-down technique	Τεχνική “από πάνω προς τα κάτω”
Transcript mapping	Αντιστοίχιση Εικόνας και Αρχείου Κειμένου
Transcription	ASCII Κωδικοποίηση της Εικόνας
Vertical Histogram	Ιστόγραμμα Κατακόρυφων Προβολών
Vertical Stroke Boundaries	Κάθετα όρια χαρακτήρων

ΣΥΝΤΜΗΣΕΙΣ – ΑΡΚΤΙΚΟΛΕΞΑ – ΑΚΡΩΝΥΜΙΑ

ARLSA	Adaptive Run Length Smoothing Algorithm
ASCII	American Standard Code for Information Interchange
CCL	Connected Components Labelling
CEDAR	Center of Excellence for Document Analysis and Recognition
CIL	Computational Intelligence Laboratory
DIU	Document Image Understanding
DTW	Dynamic Time Warping
Docsprum	Document Spectrum
DR	Detection Rate
FM	F-Measure
HMM	Hidden Markov Models
IAM	Informatics and Applied Mathematics
LSE	Least Square Estimation
NN	Nearest Neighbor
OCR	Optical Character Recognition
RA	Recognition Accuracy
RLSA	Run Length Smoothing Algorithm
SBD	Segmentation Based Dewarping
SIFT	Scale Invariant Feature Transform
SVM	Support Vector Machines
TPS	Thin-Plane Splines
TSD	Two-Step Dewarping
VSBs	Vertical Stroke Boundaries
ΕΚΕΦΕ	Εθνικό Κέντρο Ερευνών και Φυσικών Επιστημών
ΕΚΠΑ	Εθνικό Καποδιστριακό Πανεπιστήμιο Αθηνών
ΙΠΤ	Ινστιτούτο Πληροφορικής και Τηλεπικοινωνιών

.

ΑΝΑΦΟΡΕΣ

- [1] N. Otsu, "A Threshold Selection Method from Gray-level Histograms", IEEE Transactions on Systems, Man and Cybernetics, vol. 9, no. 1, 1979, pp. 62-66.
- [2] J.S. Weszka, R.N. Nagel and A. Rosenfeld, "A threshold selection technique", IEEE Transactions on Computers, vol. C-23, no. 12, 1974, pp. 1322-1326.
- [3] T.W. Ridler and S. Calvard, "Picture Thresholding Using an Iterative Selection Method", IEEE Transactions On Systems, Man and Cybernetics, vol. 8, no. 8, 1978, pp. 630-632.
- [4] W. Niblack, "An Introduction to Digital Image Processing", Prentice Hall, Englewood Cliffs, 1986, pp. 115-116.
- [5] J. Sauvola and M. Pietikainen, "Adaptive document image binarization", Pattern Recognition, vol. 33, no. 2, 2000, pp. 225-236.
- [6] B. Gatos, I. Pratikakis and S.J. Perantonis, "Adaptive degraded document image binarization", Pattern Recognition vol. 39, no. 3, 2006, pp. 317-327.
- [7] Y. Chen and G. Leedham, "Decompose algorithm for thresholding degraded historical document images", Vision Image and Signal Processing, vol. 152, no. 6, 2005, pp. 702-714.
- [8] E. Kavallieratou and S. Stathis, "Adaptive Binarization of Historical Document Images", 18th International Conference on Pattern Recognition, Hong Kong, 2006, pp. 742-745.
- [9] J. He, Q.D.M. Do, A.C. Downton and J.H. Kim, "A comparison of binarization methods for historical archive documents", 8th International Conference on Document Analysis and Recognition, Seoul, Korea, 2005, pp. 538-542.
- [10] P. Stathis, E. Kavallieratou and N. Papamarkos, "An evaluation survey of binarization algorithms on historical documents", 19th International Conference on Pattern Recognition, Tampa, Florida, USA, 2008, pp. 1-4.
- [11] M. Kamel and A. Zhao, "Extraction of Binary Character/Graphics Images from Grayscale Document Images", CVGIP: Graphical Models and Image Processing, vol. 55, no. 3, 1993, pp. 203-217.
- [12] Y. Yang and H. Yan, "An Adaptive Logical Method for Binarization of Degraded Document Images", Pattern Recognition, vol. 33, no. 5, 2000, pp. 787-807.
- [13] C.Y. Suen, C. Nadal, R. Legault, T.A. Mai and L. Lam, "Computer recognition of unconstrained handwritten numerals", Proceedings of the IEEE, vol. 80, no. 7, 1992, pp. 1162-1180.
- [14] J. Serra, "Morphological filtering: An overview", Signal Process, vol. 38, no. 1, 1994, pp. 3-11.
- [15] R.C. Gonzalez and R.E. Woods, "Digital Image Processing", Second Edition, Prentice Hall, 2002.
- [16] F. Matheron, "Random Sets and Integral Geometry", John Wiley & Sons Inc., New York, 1975.
- [17] S. Liang, M. Ahmadi and M. Shridhar, "A morphological approach to text string extraction from regular periodic overlapping text/background images", CVGIP: Graphical Models and Image Processing, vol. 56, no. 5, 1994, pp. 402-413.

- [18] R.F. Moghaddam and M. Cheriet, "RSLDI: Restoration of single-sided low-quality document images", *Pattern Recognition*, vol. 42, no. 12, 2009, pp. 3355-3364.
- [19] R.F. Moghaddam and M. Cheriet, "Beyond pixels and regions: A non-local patch means (NLPM) method for content-level restoration, enhancement, and reconstruction of degraded document images", *Pattern Recognition*, vol. 44, no. 2, 2011, pp. 363-374.
- [20] D.X. Le, G.R. Thoma and H. Wechsler, "Automated borders detection and adaptive segmentation for binary document images", *International Conference on Pattern Recognition*, Vienna, Austria, 1996, pp. 737-741.
- [21] K.C. Fan, Y.K. Wang and T.R. Lay, "Marginal noise removal of document images", *Pattern Recognition*, vol. 35, no. 11, 2002, pp. 2593-2611.
- [22] B.T. Avila and R.D. Lins, "A New Algorithm for Removing Noisy Borders from Monochromatic Documents", *ACM Symposium on Applied Computing*, Nicosia, Cyprus, 2004, pp. 1219-1225.
- [23] B.T. Avila and R.D. Lins, "Efficient Removal of Noisy Borders from Monochromatic Documents", *International Conference on Image Analysis and Recognition*, Porto, Portugal, 2004, pp. 249-256.
- [24] F. Shafait, J.v. Beusekom, D. Keysers and T.M. Breuel, "Document cleanup using page frame detection", *International Journal on Document Analysis and Recognition*, vol. 11, no. 2, 2008, pp. 81-96.
- [25] S. Yacoub, J. Burns, P. Faraboschi, D. Ortega, J.A. Peiro and V. Saxena, "Document Digitization Lifecycle for Complex Magazine Collection", *ACM Symposium on Document Engineering*, Bristol, United Kingdom, 2005, pp. 197-206.
- [26] P.Y. Yin, "Skew detection and block classification of printed documents", *Image and Vision Computing*, vol. 19, no. 8, 2001, pp. 567-579.
- [27] F.M. Wahl, K.Y. Wong and R.G. Casey, "Block Segmentation and Text Extraction in Mixed Text/Image Documents", *Computer Graphics and Image Processing*, vol. 20, no. 4, 1982, pp. 375-390.
- [28] A. Hashizume, P.S. Yeh and A. Rosenfeld, "A method of detecting the orientation of aligned components", *Pattern Recognition Letters*, vol. 4, no. 2, 1986, pp. 125-132.
- [29] X. Jiang, H. Bunke and D. Widmer-Kljajo, "Skew Detection of Document Images by Focused Nearest-Neighbor Clustering", *5th International Conference on Document Analysis and Recognition*, Bangalore, India, 1999, pp. 629-632.
- [30] B. Gatos, N. Papamarkos and C. Chamzas, "Skew detection and text line position determination in digitized documents", *Pattern Recognition*, vol. 30, no. 9, 1997, pp. 1505-1519.
- [31] H.S. Baird, "The Skew Angle of Printed Documents", *Society of Photographic Scientists and Engineers*, Rochester, N.Y., 1987, pp. 21-24.
- [32] T. Steinherz, N. Intrator and E. Rivlin, "Skew detection via principal components analysis", *5th International Conference on Document Analysis and Recognition*, Bangalore, India, 1999, pp. 153-156.

- [33] S.C. Hinds, J.L. Fisher and D.P. D'Amato, "A document skew detection method using run-length encoding and the Hough transform", 10th International Conference on Pattern Recognition, Atlantic City, NJ, 1990, pp. 464-468.
- [34] A. Amin and S. Fischer, "A Document Skew Detection Method Using the Hough Transform", Pattern Analysis & Applications, vol. 3, no. 3, 2000, pp. 243-253.
- [35] L. Zhang, Y. Zhang and C.L. Tan, "An Improved Physically-Based Method for Geometric Restoration of Distorted Document Images", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 30, no. 4, 2008, pp. 728-734.
- [36] A. Ulges, C.H. Lampert and T. Breuel, "Document capture using stereo vision", ACM Symposium on Document Engineering, Milwaukee, Wisconsin, USA, 2004, pp. 198-200.
- [37] A. Yamashita, A. Kawarago, T. Kaneko and K.T. Miura, "Shape reconstruction and image restoration for non-flat surfaces of document with a stereo vision system", 17th International Conference on Pattern Recognition, Cambridge, UK, 2004, pp. 482-485.
- [38] M.S. Brown and W.B. Seales, "Image restoration of arbitrarily warped documents", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 26, no. 10, 2004, pp. 1295-1306.
- [39] H. Cao, X. Ding and C. Liu, "Rectifying the bound document image captured by the camera: A model based approach", 7th International Conference on Document Analysis and Recognition, Scotland, 2003, pp. 71-75.
- [40] J. Liang, D. DeMenthon and D. Doermann, "Geometric rectification of camera-captured document images", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 30, no. 4, 2008, pp. 591-605.
- [41] C.L. Tan, L. Zhang, Z. Zhang and T. Xia, "Restoring warped document images through 3D shape modeling", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 2, 2006, pp. 195-208.
- [42] L. Zhang, A.M. Yip, M.S. Brown and C.L. Tan, "A unified framework for document restoration using inpainting and shape-from-shading", Pattern Recognition, vol. 42, no. 11, 2009, pp. 2961-2978.
- [43] O. Lavielle, X. Molines, F. Angella and P. Baylou, "Active Contours Network to Straighten Distorted Text Lines", International Conference on Image Processing, Thessaloniki, Greece, 2001, pp. 748-751.
- [44] C. Wu and G. Agam, "Document image De-warping for Text/Graphics recognition", IAPR International Workshop on Structural, Syntactic and Statistical Pattern Recognition, Windsor, Canada, 2002, pp. 348-357.
- [45] L. Zhang and C.L. Tan, "Warped image restoration with applications to digital libraries", 8th International Conference on Document Analysis and Recognition, Seoul, Korea, 2005, pp. 192-196.
- [46] H. Ezaki, S. Uchida, A. Asano and H. Sakoe, "Dewarping of document image by global optimization", 8th International Conference on Document Analysis and Recognition, Seoul, Korea, 2005, pp. 302-306.

- [47] L. Mischke and W. Luther, "Document Image De-warping Based on Detection of Distorted Text Lines", International Conference on Image Analysis and Processing, Cagliari, Italy, 2005, pp. 1068-1075.
- [48] B. Fu, M.Wu, R. Li, W. Li, Z. Xu and C. Yang, "A model-based book dewarping method using text line detection", International Workshop on Camera-Based Document Analysis and Recognition, Curitiba, Brazil, 2007, pp. 63-70.
- [49] A. Ulges, C.H. Lampert and T.M. Breuel, "Document image dewarping using robust estimation of curled text lines", 8th International Conference on Document Analysis and Recognition, Seoul, Korea, 2005, pp. 1001-1005.
- [50] S.J. Lu, B.M. Chen and C.C. Ko, "A partition approach for the restoration of camera images of planar and curled document", Image and Vision Computing, vol. 24, no. 8, 2006, pp. 837-848.
- [51] D.C. Schneider, M. Block and R. Rojas, "Robust Document Warping with Interpolated Vector Fields", 9th International Conference on Document Analysis and Recognition, Curitiba, Brazil, 2007, pp. 113-117.
- [52] S.S. Bukhari, F. Shafait and T.M. Breuel, "Dewarping of document images using coupled-snakes", International Workshop on Camera-Based Document Analysis and Recognition, Barcelona, Spain, 2009, pp. 34-41.
- [53] Y. Zhang, C. Liu, X. Ding and Y. Zou, "Arbitrary warped document image restoration based on segmentation and Thin-Plate Splines", 19th International Conference on Pattern Recognition, Florida, USA, 2008, pp. 1-4.
- [54] M.S. Brown and Y.C. Tsai, "Geometric and shading correction for images of printed materials using boundary", IEEE Transactions on Image Processing, vol. 15, no. 6, 2006, pp. 1544-1554.
- [55] A. Masalovitch and L. Mestetskiy, "Usage of continuous skeletal image representation for document images de-warping", International Workshop on Camera-Based Document Analysis and Recognition, Curitiba, Brazil, 2007, pp. 45-53.
- [56] G. Farin, "Curves and Surfaces for Computer Aided Geometric Design: A practical guide", 4th edition, Academic Press, 1996.
- [57] S. Coons, "Surfaces for Computer Aided Design", Technical report, Mass. Inst. Technol., Cambridge, 1968.
- [58] G. Nagy and S. Seth, "Hierarchical representation of optically scanned documents", 7th International Conference on Pattern Recognition, Montreal, Canada, 1984, pp. 347-349.
- [59] S.W. Lee and D.S. Ryu, "Parameter-Free Geometric Document Layout Analysis", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 11, 2001, pp. 1240-1256.
- [60] L. O'Gorman, "The Document Spectrum for Page Layout Analysis", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 15, no. 11, 1993, pp. 1162-1173.

- [61] A. Simon, J-C Pret and A.P. Johnson, "A fast algorithm for bottom-up document layout analysis", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 3, 1997, pp. 273-277.
- [62] S.L. Taylor, D.A. Dahl, M. Lipshutz, C. Weir, L.M. Norton, R.W. Nilson and M.C. Linebarger, "Integrating Natural Language Understanding with Document Structure Analysis", *Artificial Intelligence Review*, vol. 8, no. 2-3, 1994, pp. 255-276.
- [63] A. Antonacopoulos, "Page Segmentation Using the Description of the Background", *Computer Vision and Image Understanding*, vol. 70, no. 3, 1998, pp. 350-369.
- [64] S. Nicolas, T. Paquet and L. Heutte, "Complex handwritten page segmentation using contextual models", 2nd International Conference on Document Image Analysis for Libraries, Lyon, France, 2006, pp. 46-59.
- [65] N. Journet, R. Mullot, J.Y. Ramel and V. Eglin, "Ancient Printed Documents Indexation: A New Approach", 3rd International Conference on Advances in Pattern Recognition, Bath, United Kingdom, 2005, pp. 580-589.
- [66] M. Bulacu, R. van Koert, L. Schomaker and T. van der Zant, "Layout Analysis of Handwritten Historical Documents for Searching the Archive of the Cabinet of the Dutch Queen", 9th International Conference on Document Analysis and Recognition, Curitiba, Brazil, 2007, pp. 357-361.
- [67] J. Y. Ramel, S. Leriche, M. L. Demonet and S. Busson, "User-driven page layout analysis of historical printed books", *International Journal on Document Analysis and Recognition*, vol. 9, no. 2-4, 2007, pp. 243-261.
- [68] M. Baechler and R. Ingold, "Medieval manuscript layout model", *ACM Symposium on Document Engineering*, Manchester, United Kingdom, 2010, pp. 275-278.
- [69] Y. Li, Y. Zheng and D. Doermann, "Detecting Text Lines in Handwritten Documents", 18th International Conference on Pattern Recognition, Hong Kong, 2006, pp. 1030-1033.
- [70] E. Bruzzone and M.C. Coffetti, "An algorithm for extracting cursive text lines", 5th International Conference on Document Analysis and Recognition, Bangalore, India, 1999, pp. 749-752.
- [71] D.J. Kennard and W.A. Barrett, "Separating Lines of Text in Free-Form Handwritten Historical Documents", 2nd International Conference on Document Image Analysis for Libraries, Lyon, France, 2006, pp. 12-23.
- [72] A. Lemaitre and J. Camillerapp, "Text Line Extraction in Handwritten Document with Kalman Filter Applied on Low Resolution Image", 2nd International Conference on Document Image Analysis for Libraries, Lyon, France, 2006, pp. 38-45.
- [73] A. Antonacopoulos and D. Karatzas, "Semantics-based content extraction in typewritten historical documents", 8th International Conference on Document Analysis and Recognition, Seoul, Korea, 2005, pp. 48-53.
- [74] M. Feldbach and K.D. Tonnies, "Line Detection and Segmentation in Historical Church Registers", 6th International Conference on Document Analysis and Recognition, Seattle, WA, USA, 2001, pp. 743-747.

- [75] L. Linkforman-Sumem, A. Zahour and B. Taconet, "Text line segmentation of historical documents: a survey", *International Journal on Document Analysis and Recognition*, vol. 9, no. 2, 2006, pp. 123–138.
- [76] G. Louloudis, B. Gatos, I. Pratikakis and C. Halatsis, "Text line detection in handwritten documents", *Pattern Recognition*, vol. 41, no. 12, 2008, pp. 3758-3772.
- [77] A. Antonacopoulos and D. Karatzas, "Document Image analysis for World War II personal records", 1st International Workshop on Document Image Analysis for Libraries, Palo Alto, USA, 2004, pp. 336-341.
- [78] J. Park and V. Govindaraju, "Use of adaptive segmentation in handwritten phrase recognition", *Pattern Recognition*, vol. 35, no. 1, 2002, pp. 245-252.
- [79] S.H. Kim, C.B. Jeong, H.K. Kwag and C.Y. Suen, "Word segmentation of printed text lines based on gap clustering and special symbol detection", 16th International Conference on Pattern Recognition, Quebec, Canada, 2002, pp. 320-323.
- [80] R. Manmatha and J.L. Rothfeder, "A Scale Space Approach for Automatically Segmenting Words from Historical Handwritten Documents", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, 2005, pp. 1212-1225.
- [81] B. Gatos, T. Konidakis, K.Ntzios, I. Pratikakis and S.J. Perantonis, "A Segmentation-free Approach for Keyword Search in Historical Typewritten Documents", 8th International Conference on Document Analysis and Recognition, Seoul, Korea, 2005, pp. 54-58.
- [82] S. Liang, M. Shridhar and M. Ahmadi, "Segmentation of touching characters in printed document recognition", *Pattern Recognition*, vol. 27, no. 6, 1994, pp. 825-840.
- [83] T.C. Chang and S.Y. Chen, "Character Segmentation Using Convex-Hull Techniques", *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 13, no. 6, 1999, pp. 833-858.
- [84] S. Nomura, K. Yamanaka, O. Katai, H. Kawakami and T. Shiose, "A novel adaptive morphological approach for degraded character image segmentation", *Pattern Recognition*, vol. 38, no. 11, 2005, pp. 1961-1975.
- [85] B.A. Yanikoglu and P.A. Sandon, "Segmentation of off-line cursive handwriting using linear programming", *Pattern Recognition*, vol. 31, no. 12, 1998, pp. 1825-1833.
- [86] E. Kavallieratou, E. Stamatatos, N. Fakotakis and G. Kokkinakis, "Handwritten Character Segmentation using Transformation-based Learning", 15th International Conference on Pattern Recognition, Barcelona, Spain, 2000, pp. 634-637.
- [87] X. Xiao and G. Leedham, "Knowledge-based English cursive script segmentation", *Pattern Recognition Letters*, vol. 21, no. 10, 2000, pp. 945-954.
- [88] G. Nicchiotti, C. Scagliola and S. Rimassa, "A Simple and Effective Cursive Word Segmentation Method", 7th International Workshop on Frontiers in Handwriting Recognition, Amsterdam, Netherlands, 2000, pp. 499-504.
- [89] S. Zhao, Z. Chi, P. Shi and H. Yan, "Two-stage segmentation of unconstrained handwritten Chinese character", *Pattern Recognition*, vol. 36, no. 1, 2003, pp. 145-156.

- [90] G. Timar, K. Karacs and C. Rekeczky, "Analogic Preprocessing and Segmentation Algorithms for Off-line Handwriting Recognition", *Journal of Circuits, Systems, and Computers*, vol. 12, no. 6, 2003, pp. 783-804.
- [91] L.S. Oliveira, E. Lethelier, F. Bortolozzi and R. Sabourin, "Handwritten Digits Segmentation based on Structural Approach", *13th Brazilian Symposium on Computer Graphics and Image Processing*, Florianopolis, Brazil, 2000, pp. 67-73
- [92] Y.K. Chen and J.F. Wang, "Segmentation of Single- or Multiple-Touching Handwritten Numeral String Using Background and Foreground Analysis", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, 2000, pp. 1304-1317.
- [93] U. Pal, A. Belaid and C. Choisy, "Touching numeral segmentation using water reservoir concept", *Pattern Recognition Letters*, vol. 24, no. 1-3, 2003, pp. 261-272.
- [94] M. Zimmermann and H. Bunke, "Automatic Segmentation of the IAM Off-line Database for Handwritten English Text", *16th International Conference on Pattern Recognition*, Quebec, Canada, 2002, pp. 35-39.
- [95] A.H. Toselli, V. Romero and E. Vidal, "Viterbi based alignment between text images and their transcripts", *Workshop on Language Technology for Cultural Heritage Data*, Prague, Czech, 2007, pp. 9-16.
- [96] J. Rothfeder, R. Manmatha and T.M. Rath, "Aligning Transcripts to Automatically Segmented Handwritten Manuscripts", *7th Workshop on Document Analysis Systems*, Nelson, New Zealand, 2006, pp. 84-95.
- [97] E.M. Kornfield, R. Manmatha and J. Allan, "Text Alignment with Handwritten Documents", *1st Workshop on Document Image Analysis for Libraries*, Palo Alto, USA, 2004, pp. 195-211.
- [98] L.M. Lorigo and V. Govindaraju, "Transcript Mapping for Handwritten Arabic Documents", *14th Conference on Document Recognition and Retrieval*, San Jose, USA, 2007, pp. 277-284.
- [99] C.V. Jawahar and A. Kumar, "Content-level Annotation of Large Collection of Printed Document Images", *9th International Conference on Document Analysis and Recognition*, Curitiba, Brazil, 2007, pp.799-803.
- [100] C.I. Tomai, B. Zhang and V. Govindaraju, "Transcript Mapping for Historic Handwritten Document Images", *8th International Workshop on Frontiers in Handwriting Recognition*, Ontario, Canada, 2002, pp. 413-418.
- [101] C. Huang and S.N. Srihari, "Mapping transcripts to handwritten text", *10th International Workshop on Frontiers in Handwriting Recognition*, La Baule, France, 2006, pp. 15-20.
- [102] S. Zinger, J. Nerbonne and L. Schomaker, "Text-image alignment for historical handwritten documents", *Document Recognition and Retrieval XVI*, San Jose, USA, 2009, pp. 1-8.
- [103] J.D. Hobby, "Matching document images with ground truth", *International Journal on Document Analysis and Recognition*, vol. 1, no. 1, 1998, pp. 52-61.
- [104] IAM Handwritten Database: <http://www.iam.unibe.ch/fki/databases/iam-handwriting-database>

- [105] Biblioteca Valenciana Digital (BIVALDI): <http://bv2.gva.es>
- [106] George Washington Papers collection from the Manuscript Division at the Library of Congress: <http://frontiers.loc.gov/ammem/gwhtml/gwhome.html>
- [107] G. Kim and V. Govindaraju, "A lexicon driven approach to handwritten word recognition for real-time applications", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 4, 1997, pp. 366-379.
- [108] S.N. Srihari, S. Cha, H. Arora and S. Lee, "Individuality of handwriting", Journal of Forensic Sciences, vol. 47, no. 4, 2002, pp. 1-17.
- [109] N. Stamatopoulos, B. Gatos and A. Kesidis, "Automatic Borders Detection of Camera Document Images", 2nd International Workshop on Camera-Based Document Analysis and Recognition, Curitiba, Brazil, 2007, pp. 71-78.
- [110] N. Stamatopoulos, B. Gatos and T. Georgiou, "Page Frame Detection for Double Page Document Images", 9th International Workshop on Document Analysis Systems, Boston, MA, USA, 2010, pp. 401-408.
- [111] POLYTIMO project, <http://iit.demokritos.gr/cil/POLYTIMO>.
- [112] IMPACT project, European Community's Seventh Framework Programme under grant agreement N° 215064, <http://www.impact-project.eu/>.
- [113] F. Chang, C.J. Chen and C.J. Lu, "A linear-time component-labeling algorithm using contour tracing technique", Computer Vision and Image Understanding, vol. 93, no. 2, 2004, pp. 206-220.
- [114] J. Sauvola and M. Pietikainen, "Page segmentation and classification using fast feature extraction and connectivity analysis", 3rd International Conference on Document Analysis and Recognition, Montreal, Canada, 1995, pp. 1127-1131.
- [115] N. Nikolaou, M. Makridis, B. Gatos, N. Stamatopoulos and N. Papamarkos, "Segmentation of historical machine-printed documents using Adaptive Run Length Smoothing and skeleton segmentation paths", Image and Vision Computing, vol. 28, no. 4, 2010, pp. 590-604.
- [116] M. Junker, R. Hoch and A. Dengel, "On the evaluation of document analysis components by recall, precision, and accuracy", 5th International Conference on Document Analysis and Recognition, Bangalore, India, 1999, pp. 713-716.
- [117] BookRestorer: <http://www.i2s-bookscanner.com/>
- [118] WiseBook: <http://www.cadcam.org/wise-book.php>
- [119] ScanFix Xpress: <http://www.accusoft.com/scanfix.htm>
- [120] The British Library: <http://www.bl.uk/>
- [121] Biblioteca Nacional de España: <http://www.bne.es/es/Inicio/index.html>
- [122] Bibliothèque Nationale de France: <http://www.bnf.fr>
- [123] Bayerische Staatsbibliothek (Bavarian State Library): <http://www.bsb-muenchen.de>
- [124] Narodna in univerzitetna knjižnica (National and University Library of Slovenia): <http://www.nuk.uni-lj.si/>

- [125] St. Cyril and Methodius National Library (National Library of the Republic of Bulgaria): http://search.nationallibrary.bg/cyrisweb/catalog_en.html
- [126] Österreichische Nationalbibliothek (Austrian National Library): <http://www.onb.ac.at/>
- [127] ABBYY FineReader: <http://finereader.abbyy.com/>
- [128] N. Stamatopoulos, B. Gatos, I. Pratikakis and S.J. Perantonis, “A Two-Step Dewarping of Camera Document Images”, 8th International Workshop on Document Analysis Systems Nara, Japan, 2008, pp. 209-216.
- [129] N. Stamatopoulos, B. Gatos, I. Pratikakis and S.J. Perantonis, “Goal-oriented Rectification of Camera-Based Document Images”, IEEE Transactions on Image Processing, vol. 20, no. 4, 2011, pp. 910-920.
- [130] N. Stamatopoulos, B. Gatos and I. Pratikakis, “A Methodology for Document Image Dewarping Techniques Performance Evaluation”, 10th International Conference on Document Analysis and Recognition, Barcelona, Spain, 2009, pp. 956-960.
- [131] B. Gatos, I. Pratikakis and K. Ntirogiannis, “Segmentation Based Recovery of Arbitrarily Warped Document Images”, 9th International Conference on Document Analysis and Recognition, Curitiba, Brazil, 2007, pp. 989-993.
- [132] U.V. Marti and H. Bunke, “Using a statistical language model to improve the performance of an HMM-based cursive handwriting recognition system”, International Journal of Pattern Recognition and Artificial intelligence, vol. 15, no. 1, 2001, pp. 65–90.
- [133] T.M. Breuel, “Two Geometric Algorithms for Layout Analysis”, International Workshop on Document Analysis Systems, Princeton, USA, 2002, pp. 188-199.
- [134] D.G. Lowe, “Distinctive image features from scale-invariant keypoints”, International Journal of Computer Vision, vol. 60, no. 2, 2004, pp. 91-110.
- [135] J. Kanai, T.A. Nartker, S. Rice and G. Nagy, “Performance metrics for document understanding systems”, 2nd International Conference on Document Analysis and Recognition, Japan, 1993, pp. 424-427.
- [136] V.I. Levenshtein, “Binary codes capable of correcting deletions, insertions, and reversals”, Soviet Physics Doklady, vol. 10, no. 8, 1966, pp. 707–710.
- [137] The OCROpus open source document analysis system: <http://code.google.com/p/ocropus/>
- [138] J. Kittler, M. Hatef, R. Duin and J. Matas, “On combining classifiers”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20, no. 3, 1998, pp. 226–239.
- [139] A.L. Koerich, R. Sabourin and C.Y. Suen, “Recognition and verification of unconstrained handwritten words”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 10, 2005, pp. 1509–1522.
- [140] M. Magimai-Doss, D. Hakkani-Tur, O. Cetin, E. Shriberg, J. Fung and N. Mirghafori, “Entropy Based Classifier Combination for Sentence Segmentation”, IEEE International Conference on Acoustics, Speech and Signal Processing, Honolulu, Hawaii, 2007, pp. 189-192.

- [141] S. Tulyakov and V. Govindaraju, "Classifier Combination Types for Biometric Applications", Conference on Computer Vision and Pattern Recognition Workshop, New York, USA, 2006, pp. 58-65.
- [142] J.C. Handley, "Improving OCR accuracy through combination: a survey", International Conference on Systems, Man, and Cybernetics, California, USA, 1998, pp. 4330-4333.
- [143] J.C. Handley and T.B. Hickey, "Merging three optical character recognition outputs for improved precision using a minimum edit distance function", Online Computer Library, US Patent No. 5459739, 1995.
- [144] S.V. Rice, J. Kanai and T.A. Nartker, "An Algorithm for Matching OCR-Generated Text Strings", International Journal of Pattern Recognition and Artificial Intelligence, vol. 8, no. 5, 1994, pp. 1259-1268.
- [145] M. Ferrer and E. Valveny, "Combination of OCR Engines for Page Segmentation Based on Performance Evaluation", 9th International Conference on Document Analysis and Recognition, Curitiba, Brazil, 2007, pp. 784-788.
- [146] N. Stamatopoulos, B. Gatos and S.J. Perantonis, "A Method for Combining Complementary Techniques for Document Image Segmentation", 11th International Conference on Frontiers in Handwriting Recognition, Montreal, Canada, 2008, pp. 235-240.
- [147] N. Stamatopoulos, B. Gatos and S.J. Perantonis, "A Method for Combining Complementary Techniques for Document Image Segmentation", Pattern Recognition, vol. 42, no. 12, 2009, pp. 3158-3168.
- [148] Z. Shi and V. Govindaraju, "Line Separation for Complex Document Images Using Fuzzy Runlength", 1st International Workshop on Document Image Analysis for Libraries, 2004, pp. 306-312.
- [149] I. Phillips and A. Chhabra, "Empirical Performance Evaluation of Graphics Recognition Systems", IEEE Transactions Pattern Analysis and Machine Intelligence, vol. 21, no. 9, 1999, pp. 849-870.
- [150] H.J. Lee and B. Chen, "Recognition of handwritten Chinese characters via short line segments", Pattern Recognition, vol. 25, no. 5, 1992, pp. 543-552.
- [151] K. Khurshidt, "Analysis and Retrieval of Historical Document Images", PhD Thesis, Université Paris Descartes, 2009.
- [152] N. Stamatopoulos, G. Louloudis and B. Gatos, "A Comprehensive Evaluation Methodology for Noisy Historical Document Recognition Techniques", 3rd Workshop on Analytics for Noisy Unstructured Text Data, Barcelona, Spain, 2009, pp. 47-54.
- [153] N. Stamatopoulos, G. Louloudis and B. Gatos, "Efficient Transcript Mapping to Ease the Creation of Document Image Segmentation Ground Truth with Text-Image Alignment", 12th International Conference on Frontiers in Handwriting Recognition, Kolkata, India, 2010, pp. 226-231.
- [154] G. Louloudis, B. Gatos, I. Pratikakis and C. Halatsis, "Text line and word segmentation of handwritten documents", Pattern Recognition, vol. 42, no. 12, 2009, pp. 3169-3183.

- [155] B. Gatos, N. Stamatopoulos and G. Louloudis, "ICDAR2009 Handwriting Segmentation Contest", *International Journal on Document Analysis and Recognition*, vol. 14, no. 1, 2011, pp. 25-33.