

## ΔΗΜΟΚΡΙΤΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΡΑΚΗΣ



ΔΗΜΟΚΡΙΤΕΙΟ  
ΠΑΝΕΠΙΣΤΗΜΙΟ  
ΘΡΑΚΗΣ ■ DEMOCRITUS  
UNIVERSITY  
OF THRACE

ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ  
ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΗΛΕΚΤΡΟΝΙΚΗΣ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ  
ΣΥΣΤΗΜΑΤΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ

# Επεξεργασία Ψηφιοποιημένων Εγγράφων σε Διαδικτυακό Περιβάλλον

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Αθανάσιος Μ. Μαναλικάδης

Επιβλέπων: Ιωάννης Πρατικάκης,

Αναπληρωτής Καθηγητής

Ξάνθη, Ιούλιος 2019

## Περιεχόμενα

<b>Περιεχόμενα .....</b>	<b>2</b>
<b>1. Εισαγωγή .....</b>	<b>4</b>
<b>1.1 Γενικά .....</b>	<b>4</b>
<b>1.2 Στόχος και περιγραφή της διπλωματικής .....</b>	<b>5</b>
<b>2. Μελέτη μεθοδολογιών Δυαδικοποίησης και Κατάτμησης .....</b>	<b>6</b>
<b>2.1 Δυαδική Μετατροπή .....</b>	<b>6</b>
<b>2.2 Μέθοδος Otsu .....</b>	<b>8</b>
<b>2.2.1 Γενική επισκόπηση .....</b>	<b>8</b>
<b>2.2.2 Μαθηματική ανάλυση .....</b>	<b>9</b>
<b>2.2.3 Τα βασικά βήματα εκτέλεσης .....</b>	<b>11</b>
<b>2.3 Μέθοδος Sauvola και Pietikainen .....</b>	<b>12</b>
<b>2.3.1 Δυαδικοποίηση αντικειμένων κειμένου .....</b>	<b>12</b>
<b>2.3.2 Επιλογή κατωφλιού με παρεμβολή .....</b>	<b>15</b>
<b>2.4 Μέθοδος GPP .....</b>	<b>16</b>
<b>2.4.1 Γενική επισκόπηση .....</b>	<b>16</b>
<b>2.4.2 Ανάλυση μεθοδολογίας .....</b>	<b>17</b>
<b>2.5 Σύγκριση τεχνικών δυαδικοποίησης .....</b>	<b>23</b>
<b>2.6 Περί κατάτμησης .....</b>	<b>29</b>
<b>2.7 Η μεθοδολογία κατάτμησης της διπλωματικής .....</b>	<b>31</b>
<b>2.7.1 Συνοπτική περιγραφή της μεθοδολογίας .....</b>	<b>31</b>
<b>2.7.2 Συνδεδεμένα αντικείμενα .....</b>	<b>32</b>
<b>2.7.3 Προσαρμοστικός αλγόριθμος εξομάλυνσης .....</b>	<b>34</b>

<b>2.7.4 Ανάλυση της Μεθοδολογίας Κατάτμησης λέξεων .....</b>	<b>38</b>
<b>2.7.5 Αξιολόγηση της τεχνικής.....</b>	<b>46</b>
<b>3. Ανάλυση της προγραμματιστικής υλοποίησης και παρουσίαση της εφαρμογής.....</b>	<b>51</b>
<b>3.1 Γλώσσες Προγραμματισμού που χρησιμοποιήθηκαν .....</b>	<b>51</b>
<b>3.2 Τεχνολογίες που χρησιμοποιήθηκαν .....</b>	<b>52</b>
<b>3.3 Το πλαίσιο λογισμικού Angular .....</b>	<b>53</b>
<b>3.3.1 Γενική περιγραφή.....</b>	<b>53</b>
<b>3.3.2 Η αρχιτεκτονική της Angular .....</b>	<b>53</b>
<b>3.4 Παρουσίαση της εφαρμογής .....</b>	<b>56</b>
<b>3.4.1 Γενική επισκόπηση των βασικών περιοχών της εφαρμογής.....</b>	<b>56</b>
<b>3.4.2 Η καρτέλα δυαδικής μετατροπής .....</b>	<b>57</b>
<b>3.4.3 Η καρτέλα κατάτμησης.....</b>	<b>60</b>
<b>3.4.4 Η καρτέλα μαζικής επεξεργασίας.....</b>	<b>65</b>
<b>3.4.5 Η καρτέλα εμφάνισης δεδομένων .....</b>	<b>69</b>
<b>3.4.6 Εμφάνιση της εφαρμογής σε κινητές συσκευές.....</b>	<b>71</b>
<b>4. Συμπεράσματα .....</b>	<b>73</b>
<b>Αναφορές.....</b>	<b>75</b>

# 1. Εισαγωγή

## 1.1 Γενικά

Η σύγχρονη εποχή χαρακτηρίζεται ως η εποχή της πληροφορίας ή ως εποχή των υπολογιστών. Αυτό ερμηνεύεται με το γεγονός, πως έχει αυξηθεί σε μεγάλο βαθμό, η δυνατότητα των ανθρώπων να ανταλλάσσουν και να μεταφέρουν πληροφορίες ελεύθερα, καθώς και να έχουν εύκολη και άμεση πρόσβαση σε αυτές. Όλες οι παραπάνω δυνατότητες έχουν επιτευχθεί μέσα από την ραγδαία ανάπτυξη του παγκόσμιου ιστού (internet) και την τεράστια εξέλιξη στον τομέα της ψηφιοποίησης και του αυτοματισμού. Η ψηφιοποίηση, δηλαδή η μετατροπή της πληροφορίας από αναλογική σε ψηφιακή μορφή, είναι μια διαδικασία που διαρκώς εξελίσσεται και βελτιώνεται. Παράλληλα, μέσα από τη χρήση του διαδικτύου, επιτρέπεται η μετάδοση και η διαχείριση κάθε είδους δεδομένων (ήχος, εικόνα, βίντεο) σε σχεδόν πραγματικό χρόνο. Ζούμε λοιπόν σε μια ψηφιακή εποχή ή διανύουμε μια ψηφιακή επανάσταση, όπου η επεξεργασία, ανάλυση, αποθήκευση και ανάκτηση κάθε είδους πληροφορίας (ήχος, εικόνα, βίντεο) κρίνεται πολύ σημαντική και αποτελεί ένα ακόμα στοίχημα που πρέπει να κερδηθεί.

Μεγάλο μερίδιο από το σύνολο του όγκου της πληροφορίας κατέχουν οι εικόνες και τα έγγραφα. Βασικό πλεονέκτημα ενός ψηφιοποιημένου εγγράφου είναι πως η κατάστασή του παραμένει αναλλοίωτη και είναι ανεξάρτητη της φυσικής φθοράς, που μπορεί να περιέλθει μια εικόνα αναλογικής μορφής. Το αντικείμενο της παρούσας διπλωματικής εντάσσεται στην ψηφιακή επεξεργασία εικόνων εγγράφου σε διαδικτυακό περιβάλλον. Γενικά οι ψηφιακές εικόνες κατηγοριοποιούνται σε τρείς γενικούς τύπους ανάλογα τη φωτεινότητά τους:

- Δυαδικές εικόνες (binary images): Ύπαρξη δύο επιπέδων φωτεινότητας, συνήθως το μαύρο και το άσπρο.
- Εικόνες αποχρώσεων του γκρι (greyscale images): Ύπαρξη 256 επιπέδων φωτεινότητας του γκρι.
- Έγχρωμες εικόνες (color images): Ύπαρξη τουλάχιστον 256 αποχρώσεων (8-bit εικόνα) έως 16,777,216 αποχρώσεων (24-bit εικόνα).

Η αποτελεσματική διαχείριση και ανάλυση της πληροφορίας εικόνων εγγράφου παραμένει ακόμα και σήμερα ένα δύσκολο πρόβλημα και ενδιαφέρων πεδίο έρευνας.

## 1.2 Στόχος και περιγραφή της διπλωματικής

Στόχος της παρούσας διπλωματικής είναι να μελετήσει ήδη υπάρχουσες τεχνικές αποκατάστασης και άντλησης της πληροφορίας από εικόνες εγγράφου και να τις υλοποιήσει με χρήση τεχνολογιών διαδικτύου.

Σύμφωνα με τη διάρθρωση της διπλωματικής, το πρώτο πρόβλημα με το οποίο ασχολείται, είναι η δυαδικοποίηση ψηφιακών εγγράφων. Εξετάζονται λοιπόν τρεις τεχνικές δυαδικοποίησης, οι οποίες εφαρμόζονται σε έγγραφα αποχρώσεων του γκρι και τα διαχωρίζουν σε δύο επίπεδα φωτεινότητας (υπόβαθρο και προσκήνιο), ώστε να αναδεικνύεται τοπικά η πληροφορία του κειμένου. Οι τεχνικές αυτές υποστηρίζουν τόσο σύγχρονα όσο και αλλοιωμένα έγγραφα. Σύγχρονα θεωρούνται τα έγγραφα όπου τα επίπεδα θορύβου είναι πολύ χαμηλά και η πληροφορία (περιοχές κειμένου) μπορεί να εξαχθεί πιο εύκολα. Αντίθετα με τον όρο αλλοιωμένα έγγραφα περιγράφονται παλιά έγγραφα που συνήθως έχουν προϋπάρξει σε έντυπη μορφή, έχουν φθαρεί είτε τεχνικά είτε φυσικά και κατά τη διάρκεια της ψηφιοποίησης τους έχουν υποστεί πιθανών επιπλέον αλλοιώσεις. Προέρχονται κυρίως από τη σάρωση τυπωμένων βιβλίων και χειρόγραφων κειμένων (γράμματα, επιστολές, βιβλία). Ιδιαίτερη βαρύτητα έχει δοθεί στα ιστορικά κείμενα λόγω της αξίας τους ως φορείς πολιτισμικής κληρονομιάς. Επίσης, τα κείμενα αυτά συνήθως παρουσιάζουν διάφορα προβλήματα ποιότητας (θόρυβος), καθιστώντας τη διαδικασία αποκατάστασής τους δύσκολη και επιστημονικά ενδιαφέρουσα. Στη συνέχεια, η διπλωματική εστιάζει στη διαδικασία της κατάτμησης και εξαγωγής περιοχών του κειμένου της εικόνας και συγκεκριμένα στην τμηματοποίηση λέξεων.

Οι παραπάνω μεθοδολογίες αναπτύχθηκαν και ενσωματώθηκαν σε μία εφαρμογή διαδικτύου (web app) ώστε να επιτευχθεί:

- Άμεση πρόσβαση και χρήση της εφαρμογής κατευθείαν από όλες τις συσκευές που υποστηρίζουν προγράμματα περιήγησης στο διαδίκτυο (web browsers) με τη γλώσσα προγραμματισμού JavaScript. Δεν απαιτείται εγκατάσταση κάποιου λογισμικού καθώς οι αλγόριθμοι είναι έτοιμοι για χρήση και δοκιμή μέσω διαδικτύου από οποιοδήποτε φοιτητή, ερευνητή ή ακαδημαϊκό ασχολείται με την επεξεργασία εικόνων εγγράφου.
- Εξασφάλιση της ιδιωτικότητας των εγγράφων προς ανάλυση και αυξημένη ταχύτητα εκτέλεσης. Το στάδιο της επεξεργασίας της εικόνας γίνεται τοπικά στον υπολογιστή του χρήστη, χωρίς να απαιτείται το ανέβασμα του εγγράφου σε κάποιο απομακρυσμένο διακομιστή (server). Έτσι, η διαδικασία εκτελείται γρηγορότερα, αφού γλιτώνουμε το χρόνο upload/download και η ταχύτητα εκτέλεσης εξαρτάται μόνο από την επεξεργαστική ισχύ του εκάστοτε μηχανήματος χρήστης.

Η εφαρμογή είναι διαθέσιμη για άμεση χρήση από το σύνδεσμο <https://athanasios.netlify.com>.

## 2. Μελέτη μεθοδολογιών Δυαδικοποίησης και Κατάτμησης

### 2.1 Δυαδική Μετατροπή

Η δυαδική μετατροπή, δυαδικοποίηση ή κατωφλίωση εικόνας αναφέρεται στην μετατροπή μιας εικόνας επιπέδων του γκρι χρώματος (greyscale) σε δυαδική, δηλαδή ασπρόμαυρη. Κύριος σκοπός της είναι ο διαχωρισμός των εικονοστοιχείων της εικόνας σε δύο κατηγορίες, στα αντικείμενα ενδιαφέροντος (κείμενο σε εικόνες εγγράφων) και στο υπόβαθρο.

Η δυαδικοποίηση στην απλούστερη της μορφή πραγματοποιείτε με την επιλογή ενός κατωφλιού τέτοιου ώστε όλα τα εικονοστοιχεία της εικόνας με τιμές μικρότερες (σκοτεινότερης απόχρωσης) από αυτό να αντιστοιχούν στο κείμενο, ενώ οι υπόλοιπες τιμές (φωτεινότερης απόχρωσης) να αντιπροσωπεύουν το υπόβαθρο. Η σωστή επιλογή του κατωφλιού είναι σημαντική τόσο για το περιορισμό του θορύβου όσο και για την ποιότητα και το πάχος των χαρακτήρων της τελικής εικόνας.

$$B(x, y) = \begin{cases} 1 & (\text{κείμενο}), \\ 0 & (\text{υπόβαθρο}), \end{cases} \quad \begin{cases} I(x, y) \leq T \\ I(x, y) > T \end{cases} \quad (2.1)$$

όπου

- $B(x, y)$  η δυαδική εικόνα (binary image).
- $I(x, y)$  η αρχική γκρι αποχρώσεων εικόνα (greyscale image).
- $T$  το κατώφλι (threshold).

uel comprehendit foris idem. Barbaus canus: caput alii ul eleuati  
ridens & lenia idem. Trilua: transuersus: diminutus: lenex: niger:  
& caput unum idem. Rileus & ruffus idem. Albus & candidus  
idem. Cicer & coniferius idem. Conunctio & cundinato idem.  
Puer mundus facie: maxilla munda idem. Puella imberbis: belliper  
& flauis idem. Caput draconis: limes supior: limes unctans ul li  
mes inuis idem. Cauda draconis: limes inferior: limes exiens: uel  
limes extra idem lunt. Et confitat q̄ figure predictar pluribus aliis  
nōibus nōānatur barbariae: arabiae: & grec. Sed de calibus nōibus  
nō est mihi cura q̄uis ipsar: nōā in illis idiomatibus ab antiquis  
dæmonibus plures inueniūsem.

(α)

uel comprehendit foris idem. Barbaus canus: caput alii ul eleuati  
ridens & lenia idem. Trilua: transuersus: diminutus: lenex: niger:  
& caput unum idem. Rileus & ruffus idem. Albus & candidus  
idem. Cicer & coniferius idem. Conunctio & cundinato idem.  
Puer mundus facie: maxilla munda idem. Puella imberbis: belliper  
& flauis idem. Caput draconis: limes supior: limes unctans ul li  
mes inuis idem. Cauda draconis: limes inferior: limes exiens: uel  
limes extra idem lunt. Et confitat q̄ figure predictar pluribus aliis  
nōibus nōānatur barbariae: arabiae: & grec. Sed de calibus nōibus  
nō est mihi cura q̄uis ipsar: nōā in illis idiomatibus ab antiquis  
dæmonibus plures inueniūsem.

(β)

**Εικόνα 2.1: Παράδειγμα κατωφλίωσης. (α) Αρχική εικόνα απόχρωσης του γκρι, (β) δυαδική μετατροπή με κατώφλι  $T = 137$ .**

Οι τεχνικές δυαδικής μετατροπής ταξινομούνται σε δύο μεγάλες κατηγορίες, στις τεχνικές **καθολικής κατωφλίωσης** και στις τεχνικές **τοπικής ή προσαρμοζόμενης κατωφλίωσης**.

- Στις μεθόδους καθολικής κατωφλίωσης χρησιμοποιείτε μια μοναδική τιμή κατωφλιού για το διαχωρισμό του κειμένου από το υπόβαθρο, με τη χρήση τεχνικών όπως η ανάλυση ιστογράμματος κ.α.
- Στις μεθόδους τοπικής κατωφλίωσης αναζητούνται πολλές τιμές κατωφλιών ανάλογα με την τοπική πληροφορία της εικόνας (μέση τιμή, τυπική απόκλιση). Υπολογίζεται δηλαδή διαφορετικό κατώφλι για κάθε εικονοστοιχείο ή ευρύτερη περιοχή.

Η δυαδική μετατροπή αποτελεί πρωταρχικό στάδιο στα περισσότερα συστήματα επεξεργασίας εικόνων εγγράφου. Παίζει σημαντικό ρόλο στην ανάλυση κειμένων, αφού το αποτέλεσμά της επηρεάζει και τα μετέπειτα στάδια της επεξεργασίας, όπως η κατάτμηση γραμμών, λέξεων, γραμμάτων αλλά και το τελικό στάδιο της αναγνώρισης. Με την εφαρμογή της, αυξάνεται η ποιότητα ανάγνωσης (μειώνεται ο θόρυβος στο υπόβαθρο, ενισχύεται η αντίθεση, τονίζονται οι αχνοί χαρακτήρες, απαλείφονται οι χαρακτήρες της πίσω σελίδας και οι λεκέδες κ.α.). Τα παραπάνω συνθέτουν είδη θορύβου που συναντώνται συνήθως σε εικόνες ιστορικών εγγράφων, επηρεάζουν την απόδοση των τεχνικών δυαδικοποίησης και την καθιστούν μια δύσκολη διαδικασία.

its salutary effect in  
of biliary affection,  
which it is liable to  
months....It appears

(α)

baker distribute the  
between them. Before

(β)

Oἱ Α'. τόμοι τοῦ ΒΡΕΤΤΑΝΙΚΟΥ ΑΣΤΕΡΟΣ, οὓς  
ὑπεσχέθημεν εἰς τοὺς ἐν Ἑλλάδι συνδρομητὰς ἡμῶν  
εἰσὶν ἡδὴ ἔτοιμοι, καὶ δυνάμεθα νὰ παραδώσωμεν αὐτὸν εἰς ὅποιονδήποτε ἐνταῦθα ἀνταποκριτὴν ἢ φίλον των ἥθελον ἐκεῖνοι μᾶς παραγγείλει. Τὸ μετρον  
τοῦτο λαμβάνομεν, ὅπως ἀποφύγωμεν τὴν ἐν Ἑλλάδι  
κατάσχεσιν τῶν τόμων.

(γ)

with reference to the entries in the Diary made by  
offices from the time the overseer left his headquarters  
described limit of cash to be conveyed by the overseer is

The S.P.M. will make suitable remarks in the  
to be sent to the account office on that date about  
detained because of limit of cash with the overseer  
exceeded.

Note.—In the case of towns where mail motor

(δ)

**Εικόνα 2.2: Παραδείγματα εικόνων κειμένου με διάφορα είδη θορύβου. (α) κηλίδες,  
(β) χαρακτήρες πίσω σελίδας, (γ) αχνοί χαρακτήρες, (δ) λεκέδες.**

Στη παρούσα διπλωματική υλοποιήθηκαν και εξετάστηκαν οι ακόλουθες τρεις τεχνικές δυαδικοποίησης:

- Otsu [1]
- Sauvola και Pietikainen [2]
- GPP (Gatos, Pratikakis, Perantonis) [3]

## 2.2 Μέθοδος Otsu

### 2.2.1 Γενική επισκόπηση

Η μέθοδος δυαδικοποίησης του Otsu [1] είναι μια εδραιωμένη τεχνική που χρησιμοποιείται για την αυτόματη κατωφλίωση εικόνων γενικώς. Αποτελεί μια τεχνική καθολικής κατωφλίωσης, η οποία βασίζεται στην επεξεργασία του ιστογράμματος της εικόνας για την εύρεση ενός βέλτιστου κατωφλιού  $T$  τέτοιου ώστε να μεγιστοποιείται η διαχωρισμότητα των δύο κλάσεων της εικόνας

(κείμενο και υπόβαθρο). Στο ιστόγραμμα αποτυπώνεται η συχνότητα εμφάνισης της κάθε φωτεινότητας (επίπεδο γκρι χρώματος), δηλαδή το πλήθος των εικονοστοιχείων με φωτεινότητα  $0,1,2\dots 255$  σε μια εικόνα των 8-bit. Επομένως, για κάθε υποψήφιο κατώφλι  $T$ , το ιστόγραμμα της εικόνας χωρίζεται σε δύο πληθυσμούς-κλάσεις και υπολογίζεται η τιμή του  $T$  για την οποία η διακύμανση των τιμών της εικόνας στο υπόβαθρο (background) και στο αντικείμενο ενδιαφέροντος-κείμενο (foreground) γίνεται ταυτόχρονα ελάχιστη (within-class variance). Ισοδύναμα το κατώφλι  $T$  μπορεί να προκύψει από τη μεγιστοποίηση της διακύμανσης των τιμών ανάμεσα στο υπόβαθρο και το κείμενο (between-class variance).

## 2.2.2 Μαθηματική ανάλυση

- Έστω μια εικόνα διαστάσεων  $M \times N$  με πίνακα ιστογράμματος  $H(i)$ ,  $i \in [0, 255]$
- Η πιθανότητα εμφάνισης της στάθμης  $i$  στην εικόνα:

$$p(i) = \frac{H(i)}{M \times N} \quad (2.2)$$

- Η πιθανότητα ένα εικονοστοιχείο να ανήκει στο υπόβαθρο:

$$W_b = \sum_{i=T+1}^{255} p(i) \quad (2.3)$$

- Η πιθανότητα ένα εικονοστοιχείο να ανήκει στο κείμενο:

$$W_f = \sum_{i=0}^T p(i) \quad (2.4)$$

ή ισοδύναμα

$$W_f = 1 - W_b \quad (2.5)$$

- Η μέση τιμή  $\mu_b$  και η διακύμανση  $\sigma_b^2$  των επιπέδων του γκρι χρώματος στο υπόβαθρο:

$$\mu_b = \frac{\sum_{i=T+1}^{255} i p(i)}{W_b} \quad (2.6)$$

$$\sigma_b^2 = \frac{\sum_{i=T+1}^{255} p(i)(i - \mu_b)^2}{W_b} \quad (2.7)$$

- Αντίστοιχα, η μέση τιμή  $\mu_f$  και η διακύμανση  $\sigma_f^2$  των επιπέδων του γκρι χρώματος στο κείμενο:

$$\mu_f = \frac{\sum_{i=0}^T i p(i)}{W_f} \quad (2.8)$$

$$\sigma_f^2 = \frac{\sum_{i=0}^T p(i)(i - \mu_f)^2}{W_f} \quad (2.9)$$

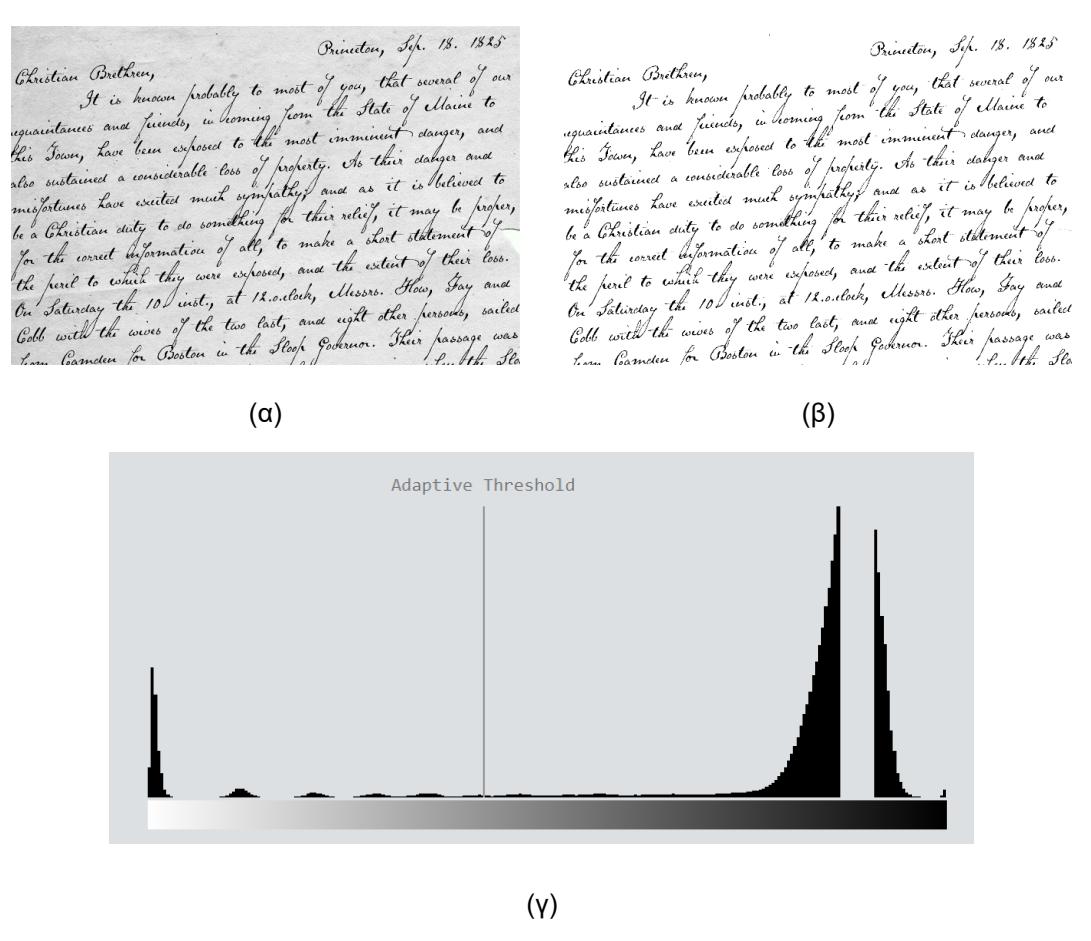
- Η εντός κλάσης διακύμανση (within-class variance) ορίζεται ως:

$$\sigma_w^2(T) = W_b + \sigma_b^2 + W_f + \sigma_f^2 \quad (2.10)$$

- Σύμφωνα με τη μέθοδο Otsu, η βέλτιστη τιμή κατωφλιού  $T$  είναι εκείνη για την οποία η  $\sigma_w^2(T)$  γίνεται ελάχιστη.
- Προκύπτει το ίδιο αποτέλεσμα εάν μεγιστοποιηθεί η μεταξύ κλάσεων διακύμανση (between-class variance), η οποία δίνεται από τη σχέση:

$$\sigma_B^2(T) = W_b W_f (\mu_b - \mu_f)^2 \quad (2.11)$$

- Ως κατώφλι επιλέγεται η τιμή του  $i$  που μεγιστοποιεί την ποσότητα  $\sigma_B^2(T)$ .



Εικόνα 2.3: Παράδειγμα καθολικής κατωφλίωσης της μεθόδου Otsu. (α) Αρχική εικόνα απόχρωσης γκρι, (β) δυαδική εικόνα Otsu, (γ) το ιστόγραμμα της αρχικής εικόνας μαζί με το υπολογισμένο κατώφλι ( $T = 126$ ).

### 2.2.3 Τα βασικά βήματα εκτέλεσης

Στην υλοποίηση του αλγορίθμου ακολουθήθηκαν τα παρακάτω βασικά βήματα:

1. Υπολογισμός Ιστογράμματος  $H(i)$ ,  $i \in [0, 255]$ .
2. Για όλες τις πιθανές τιμές του κατωφλιού  $T$ ,  $T \in [0, 255]$ :

**2.1** Ανανέωση του αριθμού των εικονοστοιχείων που ανήκουν στο υπόβαθρο και στο κείμενο.

$$\omega_b = \sum_{i=T+1}^{255} H(i) \quad (2.12)$$

$$\omega_f = \sum_{i=0}^{255} H(i) - \omega_b \quad (2.13)$$

**2.2** Ανανέωση της μέσης τιμής των επιπέδων του γκρι χρώματος στο υπόβαθρο και στο κείμενο.

$$\mu_b(T) = \frac{\sum_{i=T+1}^{255} i H(i)}{\omega_b} \quad (2.14)$$

$$\mu_f(T) = \frac{\sum_{i=0}^{255} i H(i) - \sum_{i=T+1}^{255} i H(i)}{\omega_f} \quad (2.15)$$

**2.3** Υπολογισμός της διακύμανσης μεταξύ των δύο κλάσεων.

$$\sigma_B^2(T) = \omega_b \omega_f (\mu_b - \mu_f)^2 \quad (2.16)$$

**2.4** Άνταξη  $\sigma_B^2(T) > \sigma_B^2(T-1)$  επιλέγεται το  $T$ .

**3.** Το τελικό κατώφλι  $T$  είναι αυτό για το οποίο μεγιστοποιείται το  $\sigma_B^2(T)$ .

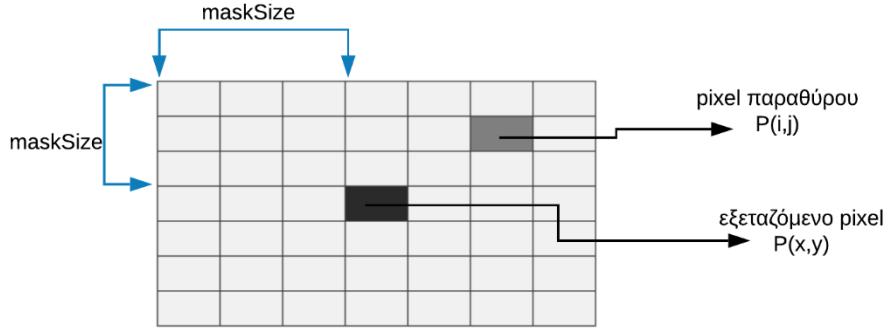
$$T = argmax(\sigma_B^2(T)) \quad (2.17)$$

## 2.3 Μέθοδος Sauvola και Pietikainen

Η μέθοδος Sauvola και Pietikainen [2] αποτελεί μια τεχνική τοπικής δυαδικής μετατροπής.

### 2.3.1 Δυαδικοποίηση αντικειμένων κειμένου

Στη προσαρμοζόμενη κατωφλίωση Sauvola [2], σε κάθε εικονοστοιχείο κεντράρεται ένα ορθογώνιο παράθυρο  $W$  και το αντίστοιχο κατώφλι υπολογίζεται με βάση την πληροφορία που υπάρχει εντός του παραθύρου.



Εικόνα 2.4: παράθυρο  $W$ .

$$m(x, y) = \frac{1}{Sum} \sum_{i=x-maskSize}^{x+maskSize} \sum_{j=y-maskSize}^{y+maskSize} P(i, j) \quad (2.18)$$

$$s(x, y) = \sqrt{\frac{1}{Sum} \sum_{i=x-maskSize}^{x+maskSize} \sum_{j=y-maskSize}^{y+maskSize} (P(i, j) - m(x, y))^2} \quad (2.19)$$

Όπου  $Sum = (2 \times maskSize + 1)^2$  ο συνολικός αριθμός εικονοστοιχείων του παραθύρου  $W$ .

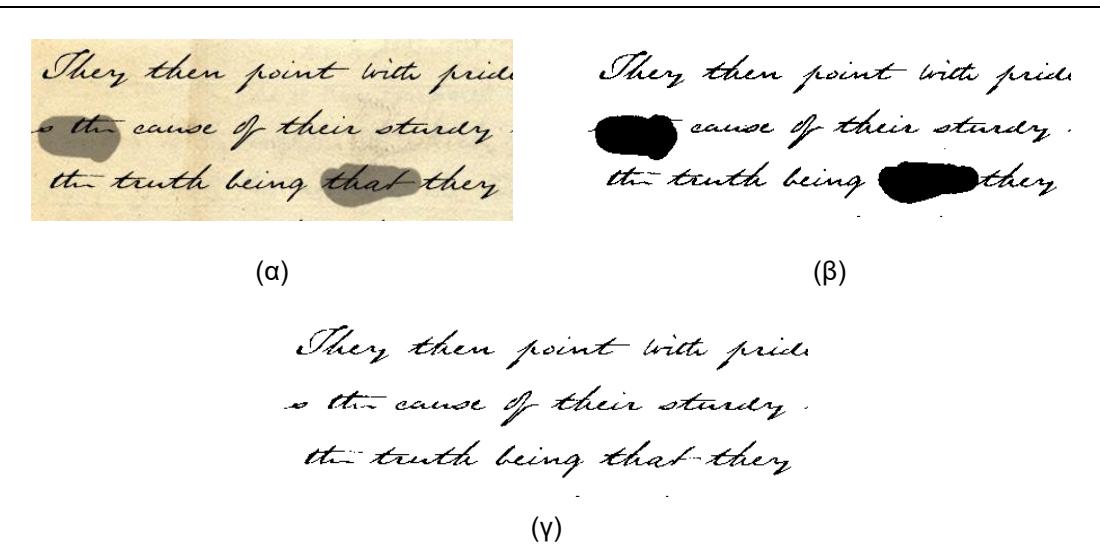
Για την εύρεση του κατωφλιού  $T$ , χρησιμοποιείτε η μέση τιμή  $m(x, y)$  και η τυπική απόκλιση  $s(x, y)$  των εικονοστοιχείων εντός του παραθύρου σύμφωνα με την παρακάτω εξίσωση:

$$T(x, y) = m(x, y) \left[ 1 + k \left( \frac{s(x, y)}{R} - 1 \right) \right] \quad (2.20)$$

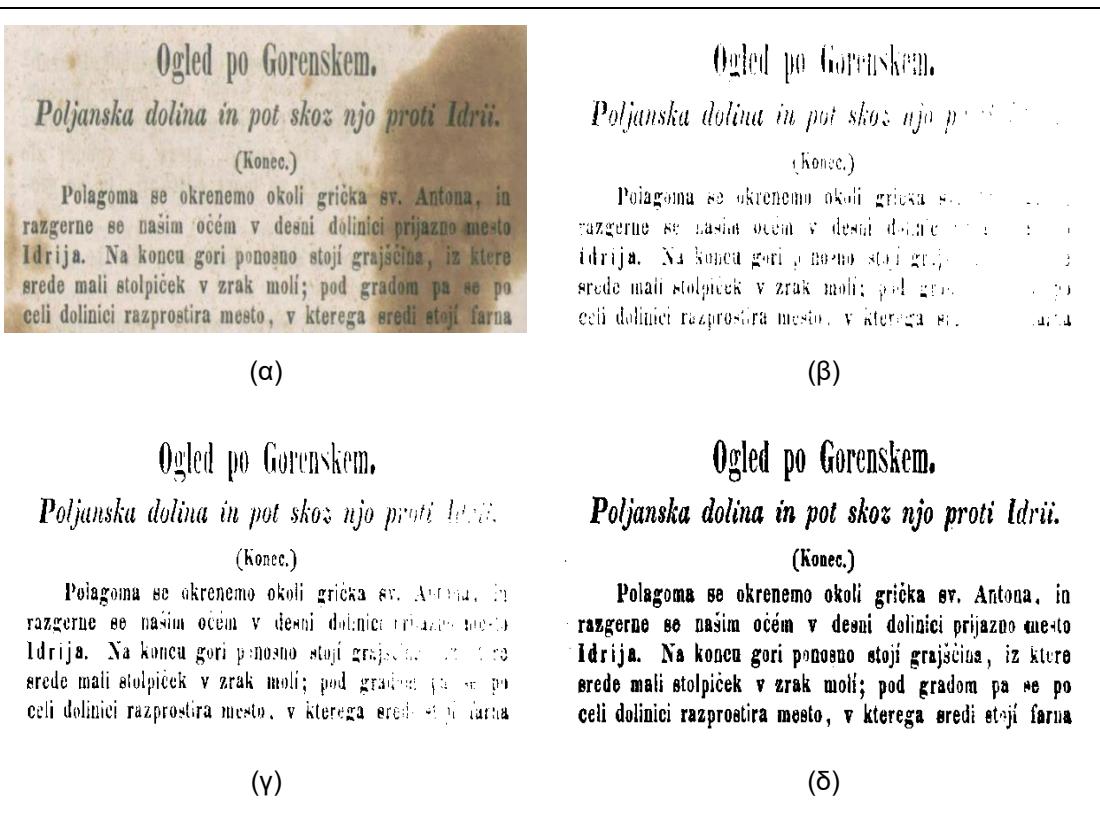
óπου

- $R$ : παράμετρος που αντιστοιχεί στο δυναμικό εύρος της τυπικής απόκλισης και
  - $k$ : παράμετρος που ρυθμίζει την ποσότητα του θορύβου που προθέτεται ή αφαιρείται.

Ο αλγόριθμος έχει μεγάλη ευαισθησία στη τιμή της παραμέτρου  $k$ , η οποία παίρνει θετικές τιμές. Για καλύτερα αποτελέσματα προτείνονται οι τιμές παραμέτρων  $R=128$  και  $k=0.5$ , που μειώνουν δραστικά το θόρυβο.

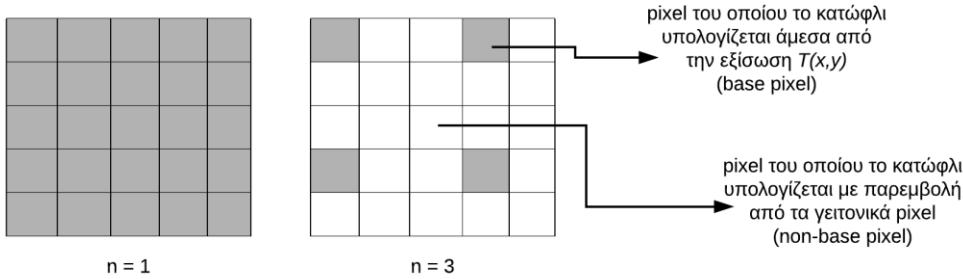


Εικόνα 2.5: Διαδικοποίηση της εικόνας (α) κατά (β) Otsu και κατά (γ) Sauvola και Pietikainen.



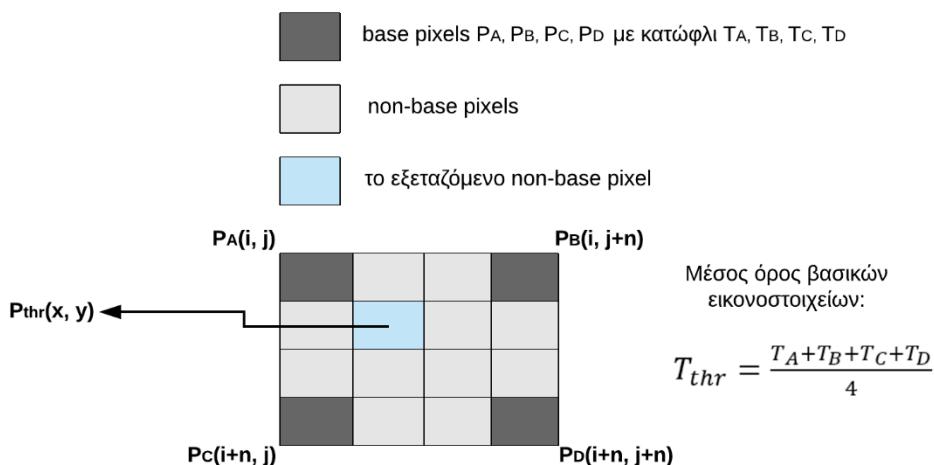
Εικόνα 2.6: Ευαισθησία του αλγορίθμου Sauvola στην τιμή της παραμέτρου  $k$ . (α) αρχική εικόνα, (β)  $k = 0.5$ , (γ)  $k = 0.4$ , (δ)  $k = 0.2$ .

### 2.3.2 Επιλογή κατωφλιού με παρεμβολή



Εικόνα 2.7: Παρεμβολή για επιλογή κατωφλιού.

Πρώτα υπολογίζεται το κατώφλι άμεσα από την εξίσωση  $T(x, y)$ , μετακινώντας το παράθυρο  $W$  σε κάθε  $n$ -οστό εικονοστοιχείο και στη συνέχεια (για  $n > 1$ ) τα υπόλοιπα, χρησιμοποιώντας παρεμβολή από τα γειτονικά εικονοστοιχεία.



Εικόνα 2.8: Παράθυρο παρεμβολής.

Για τον υπολογισμό του κατωφλιού στα μη-βασικά εικονοστοιχεία (non-base pixels) χρησιμοποιούμε τη μέθοδο του μέσου όρου. Αποτελεί κατάλληλη προσέγγιση για τα περισσότερα έγγραφα με τυχαίο θόρυβο και  $n > 5$ . Για κάθε εικονοστοιχείο ανάμεσα στα τέσσερα βασικά (base-pixels), υπολογίζεται το κατώφλι  $T_{thr}(x, y)$  ως ο μέσος όρος των κατωφλιών των base pixels:

$$T_{thr}(x, y) = \frac{T(i, j) + T(i, j + n) + T(i + n, j) + T(i + n, j + n)}{4} \quad (2.21)$$

Ένας άλλος τρόπος, είναι με τη χρήση διγραμμικής παρεμβολής υπολογίζοντας την απόσταση από τα βασικά εικονοστοιχεία:

$$W_A = |P_A(i, j) - P_{thr}(x, y)| \quad (2.22)$$

$$W_B = |P_B(i, j + n) - P_{thr}(x, y)| \quad (2.23)$$

$$W_C = |P_C(i + n, j) - P_{thr}(x, y)| \quad (2.24)$$

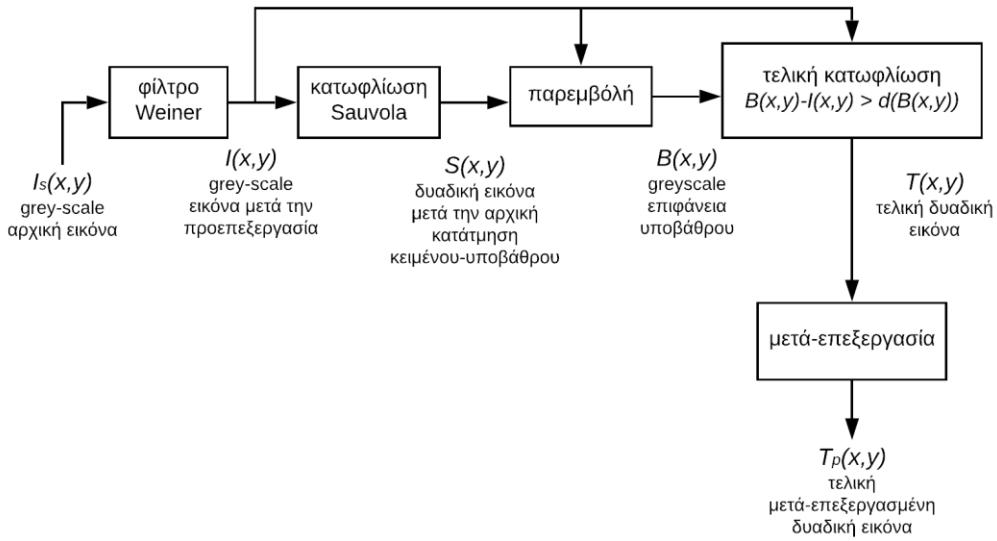
$$W_D = |P_D(i + n, j + n) - P_{thr}(x, y)| \quad (2.25)$$

$$T_{thr}(x, y) = \frac{W_A T(i, j) + W_B T(i, j + n) + W_C T(i + n, j) + W_D T(i + n, j + n)}{4} \quad (2.26)$$

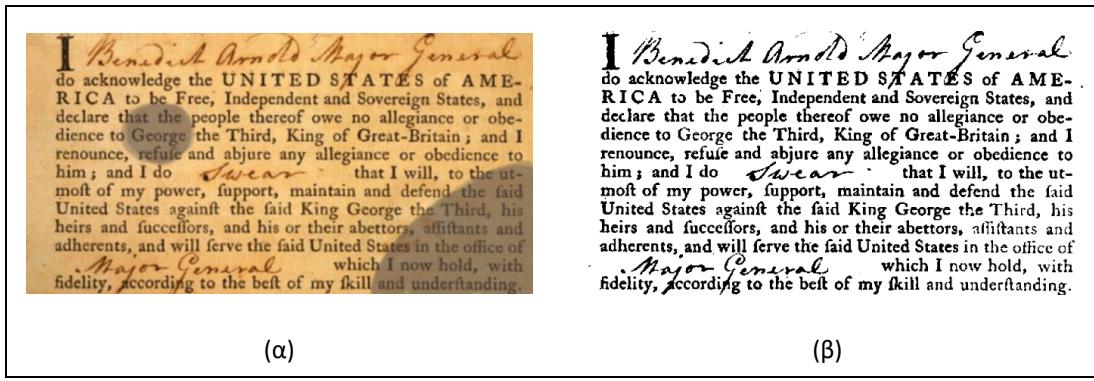
## 2.4 Μέθοδος GPP

### 2.4.1 Γενική επισκόπηση

Η μέθοδος GPP [3] αποτελεί μια καινοτόμα τεχνική τοπικής προσαρμοζόμενης κατωφλίωσης, η οποία εξειδικεύεται στην αυτόματη δυαδικοποίηση και βελτίωση χαμηλής ποιότητας εικόνων κειμένου ιστορικών εγγράφων, χωρίς τη ρύθμιση παραμέτρων από το χρήστη. Αρχικά πραγματοποιείτε ένα στάδιο προ-επεξεργασίας για την μείωση του θορύβου. Έπειτα, εφαρμόζεται η κατωφλίωση Sauvola [2] και υπολογίζεται η επιφάνεια υποβάθρου με παρεμβολή γειτονικών εικονοστοιχείων. Στη συνέχεια, ακολουθεί η τελική δυαδικοποίηση συνδυάζοντας πληροφορία από το εκτιμώμενο υπόβαθρο και την αρχική εικόνα. Το κείμενο προκύπτει όταν η διαφορά της αρχικής εικόνας από το υπόβαθρο είναι μεγαλύτερη από ένα κατώφλι. Το κατώφλι είναι δυναμικό και προσαρμόζεται στις γκρι αποχρώσεις του υποβάθρου. Τέλος, έπειται μια μετά-επεξεργασία, όπου βελτιώνεται η συνολική ποιότητα της εικόνας.



Εικόνα 2.9: Μπλοκ διάγραμμα μεθόδου GPP.



Εικόνα 2.10: Παράδειγμα δυαδικής μετατροπής με τη μέθοδο GPP. (α) Αρχική εικόνα, (β) δυαδικό αποτέλεσμα.

## 2.4.2 Ανάλυση μεθοδολογίας

Παρακάτω ακολουθούν τα βασικά βήματα εκτέλεσης της μεθόδου GPP [3].

### 1. Προ-επεξεργασία

Για αλλοιωμένης και χαμηλής ανάλυσης έγγραφα η προ-επεξεργασία της γκρι απόχρωσης αρχικής εικόνας είναι απαραίτητη για την μείωση των περιοχών θορύβου, την εξομάλυνση του φόντου στο υπόβαθρο και την βελτίωση της αντίθεσης μεταξύ κειμένου και υποβάθρου. Οπότε, στο στάδιο αυτό χρησιμοποιείται ένα προσαρμοσμένο χαμηλοπερατό φίλτρο Wiener με βάση τα στατιστικά στοιχεία της τοπικής περιοχής ανάμεσα σε κάθε εικονοστοιχείο. Η αρχική γκρι απόχρωσης (grey-scale) εικόνα  $I_s(x,y)$  “φιλτράρεται” σε μια grey-scale εικόνα  $I(x,y)$  σύμφωνα με την παρακάτω σχέση:

$$I(x, y) = \mu + (\sigma^2 - \nu^2)(I_S(x, y) - \mu)/\sigma^2 \quad (2.27)$$

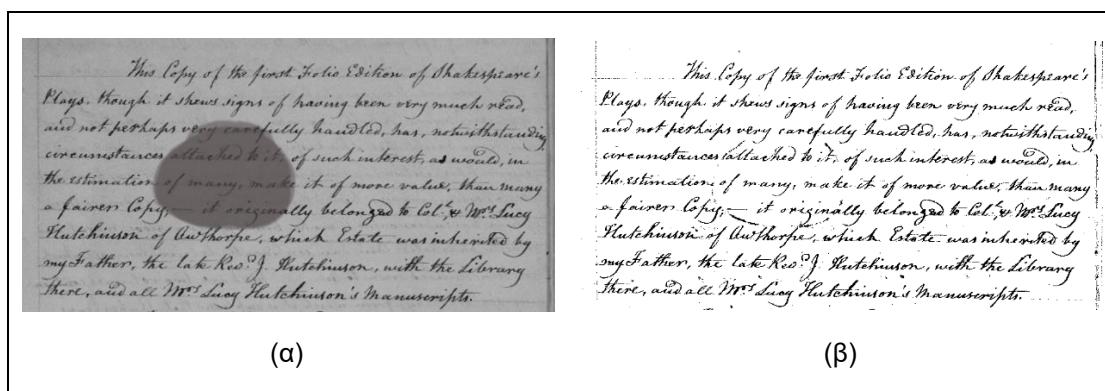
όπου  $\mu$ : η μέση τιμή,  $\sigma$ : η διακύμανση σε μία μια  $3 \times 3$  περιοχή (γειτονιά) ανάμεσα σε κάθε εικονοστοιχείο και  $\nu^2$ : η μέση τιμή των διακυμάνσεων που αντιστοιχούν σε κάθε εικονοστοιχείο της γειτονιάς.

## 2. Κατάτμηση περιοχών κειμένου και υποβάθρου

Σε αυτό το στάδιο διαχωρίζονται οι περιοχές υποβάθρου και κειμένου, χρησιμοποιώντας το δυαδικό αποτέλεσμα  $S(x, y)$  της μεθόδου Sauvola [2]. Συγκεκριμένα, από την εικόνα  $I(x, y)$  εξάγεται η εικόνα  $S(x, y)$ . Για την επιλογή του κατωφλιού γίνεται χρήση της κλασικής εξίσωσης Sauvola:

$$T_S(x, y) = \mu \left( 1 + k \left( \frac{\sigma}{R} - 1 \right) \right) \quad (2.28)$$

έχοντας ρυθμίσει χαμηλά την παράμετρο  $k = 0.2$ , ώστε να εντοπίζεται περισσότερο κείμενο, και επιλέγοντας ένα παράθυρο παρεμβολής δέκα εικονοστοιχείων.



**Εικόνα 2.11:** Τοπική κατωφλίωση της προσέγγισης Sauvola. (a) Αρχική εικόνα απόχρωσης του γκρι, (b) δυαδική εικόνα Sauvola.

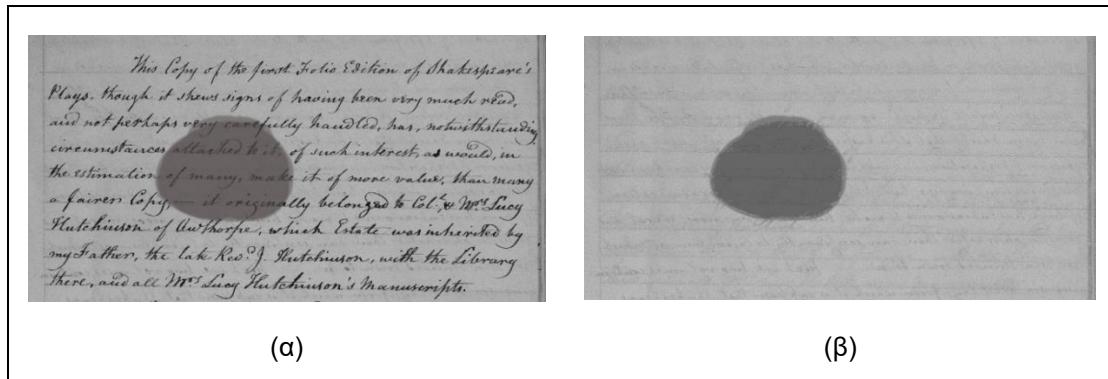
## 3. Εκτίμηση της επιφάνειας υποβάθρου

Σε αυτό το βήμα της υλοποίησης, υπολογίζεται κατά προσέγγιση η επιφάνεια υποβάθρου  $B(x, y)$  της εικόνας  $I(x, y)$  με παρεμβολή γειτονικών εικονοστοιχείων σύμφωνα με τον τύπο:

$$B(x, y) = \begin{cases} I(x, y), & S(x, y) = 0 \\ \frac{\sum_{ix=x-dx}^{x+dx} \sum_{iy=y-dy}^{y+dy} (I(ix, iy)(1 - S(ix, iy)))}{\sum_{ix=x-dx}^{x+dx} \sum_{iy=y-dy}^{y+dy} (1 - S(ix, iy))}, & S(x, y) = 1 \end{cases} \quad (2.29)$$

όπου  $dx \times dy$  οι διαστάσεις του παραθύρου παρεμβολής, το οποίο αρχικοποιείται ώστε να καλύπτει τουλάχιστον δύο χαρακτήρες κειμένου.

Στην υλοποίηση αυτής της διπλωματικής χρησιμοποιείτε ένα τετραγωνικό παράθυρο διάστασης  $dw = 20$  εικονοστοιχεία.



Εικόνα 2.12: Εκτίμηση υποβάθρου της μεθόδου GPP. (α) Αρχική εικόνα απόχρωσης του γκρι, (β) εκτιμώμενο υπόβαθρο.

Σε κάθε εικονοστοιχείο “κεντράρεται” το παράθυρο  $dx \times dy$  και εντός αυτού υπολογίζεται η μέση τιμή των εικονοστοιχείων της αρχικής εικόνας που αντιστοιχούν στα εικονοστοιχεία του υποβάθρου της μεθόδου Sauvola [2].

Πιο αναλυτικά, για την εκτίμηση του υποβάθρου  $B(x, y)$ , τα εικονοστοιχεία που αντιστοιχούν στο υπόβαθρο Sauvola ( $S(x, y) = 0$ ) παίρνουν τις αντίστοιχες τιμές των εικονοστοιχείων της αρχικής εικόνας  $I(x, y)$ . Επιπλέον, τα εικονοστοιχεία που αντιστοιχούν στο κείμενο Sauvola ( $S(x, y) = 1$ ) παίρνουν τη μέση τιμή των εικονοστοιχείων της αρχικής εικόνας  $I(ix, iy)$  που δεν ανήκουν σε αυτό ( $1 - S(ix, iy)$ ), χρησιμοποιώντας το παράθυρο παρεμβολής.

#### 4. Τελική κατωφλίωση

Σε αυτή τη φάση πραγματοποιείται η τελική κατωφλίωση έπειτα από σύγκριση της εικόνας υποβάθρου  $B(x, y)$  και της αρχικής εικόνας  $I(x, y)$ . Οι περιοχές κειμένου εντοπίζονται όταν σε κάθε σημείο, η διαφορά της αρχικής εικόνας  $I(x, y)$  από το εκτιμώμενο υπόβαθρο  $B(x, y)$  είναι μεγαλύτερη από ένα κατώφλι  $d$ . Προτείνεται η τιμή κατωφλιού  $d$  να αλλάζει δυναμικά ανάλογα με την αντίστοιχη τιμή της επιφάνειας  $B(x, y)$ , έτσι ώστε να διασώζεται η πληροφορία του κειμένου ακόμα και σε περιοχές με πολύ σκούρο υπόβαθρο (το κατώφλι  $d$  παίρνει μικρότερες τιμές σε πιο σκούρο περιοχές).

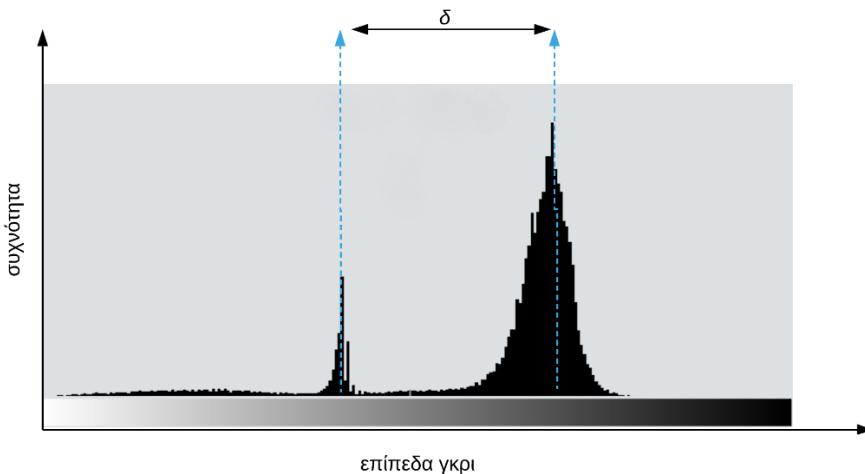
Η τελική δυαδική εικόνα  $T(x, y)$  δίνεται από το παρακάτω τύπο:

$$T(x, y) = \begin{cases} 1, & B(x, y) - I(x, y) > d(B(x, y)) \\ 0, & \text{διαφορετικά} \end{cases} \quad (2.30)$$

Σε ένα τυπικό ιστόγραμμα ενός εγγράφου αποτυπώνονται δύο κορυφές, μία για τις περιοχές κειμένου και μία για το υπόβαθρο. Η μέση απόσταση λοιπόν  $\delta$ , μεταξύ των κορυφών υποβάθρου και κειμένου, υπολογίζεται ως ο λόγος της συνολικής διαφοράς των εικονοστοιχείων υποβάθρου και αρχικής εικόνας προς τον συνολικό αριθμό εικονοστοιχείων του κειμένου Sauvola:

$$\delta = \frac{\sum_{x=0}^{I_x} \sum_{y=0}^{I_y} (B(x, y) - I(x, y))}{\sum_{x=0}^{I_x} \sum_{y=0}^{I_y} S(x, y)} \quad (2.31)$$

όπου  $I_x, I_y$  οι διαστάσεις της εικόνας.



**Εικόνα 2.13:** Ιστόγραμμα όπου φαίνεται η μέση απόσταση  $\delta$ , μεταξύ των κορυφών υποβάθρου και κειμένου.

Στην περίπτωση εικόνων εγγράφου με ομοιόμορφο φωτισμό, το ελάχιστο κατώφλι  $d$  μπορεί να προσδιοριστεί ως το γινόμενο  $q \cdot \delta$ , όπου  $q$  είναι μια παράμετρος βάρους. Στις χαμηλής αντίθεσης της εικόνας, απαιτείται μικρότερη τιμή κατωφλιού  $d$ . Για αυτό το λόγο, υπολογίζεται η μέση τιμή  $b$  του υποβάθρου  $B(x, y)$  που αντιστοιχεί στο κείμενο Sauvola σύμφωνα με τη παρακάτω εξίσωση:

$$b = \frac{\sum_{x=0}^{I_x} \sum_{y=0}^{I_y} B(x, y) S(x, y)}{\sum_{x=0}^{I_x} \sum_{y=0}^{I_y} S(x, y)} \quad (2.32)$$

Χρειάζεται να ικανοποιούνται και οι παρακάτω δύο περιπτώσεις:

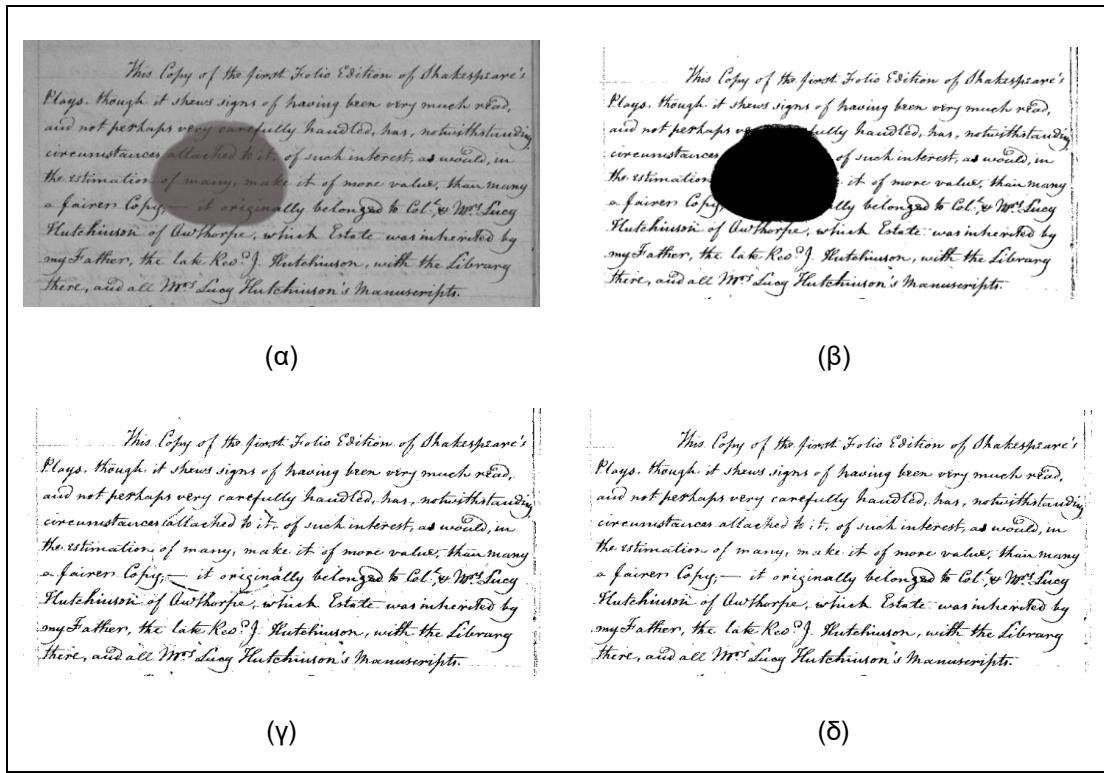
- Το κατώφλι να είναι ίσο με τη τιμή  $q \cdot \delta$ , όταν η τιμή του υποβάθρου είναι μεγαλύτερη από τη μέση τιμή  $b$ .
- Το κατώφλι να είναι ίσο με  $p_2 \cdot q \cdot \delta$ , όταν η τιμή του υποβάθρου είναι μικρότερη από το γινόμενο  $p_1 \cdot b$ .

όπου  $p_1, p_2 \in [0,1]$  και  $q$  είναι παράμετροι της συνάρτησης που ακολουθεί το δυναμικό κατώφλι  $d$ .

Για την επίτευξη όλων των παραπάνω προϋποθέσεων, χρησιμοποιείται μια λογαριθμική σιγμοειδής συνάρτηση:

$$d(B(x, y)) = q \cdot \delta \left( \frac{(1 - p_2)}{1 + \exp\left(\frac{-4B(x, y)}{b(1 - p_1)} + \frac{2(1 + p_1)}{(1 - p_1)}\right)} + p_2 \right) \quad (2.33)$$

Έπειτα από πειραματικές μελέτες, προτείνονται οι τιμές παραμέτρων  $p_1 = 0.5$ ,  $p_2 = 0.8$  και  $q = 0.6$ .



Εικόνα 2.14: Αποτέλεσμα μεθόδου GPP. (α) Αρχική εικόνα απόχρωσης του γκρι, (β) σύγκριση με την τεχνική Otsu, (β) προσέγγιση Sauvola, (γ) τελική δυαδική εικόνα GPP.

## 5. Δειγματοληψία

Για να επιτευχθεί καλύτερη δυαδική μετατροπή, εφαρμόζεται μια τεχνική δειγματοληψίας με δικυβική παρεμβολή (bicubic interpolation) και τη χρήση ενός πυρήνα  $4 \times 4 = 16$  εικονοστοιχείων. Η σχέση (2.30) που δίνει τη τελική δυαδική εικόνα στο προηγούμενο βήμα, μετασχηματίζεται στον επόμενο τύπο:

$$T(x', y') = \begin{cases} 1, & B(x, y) - I_u(x', y') > d(B(x, y)) \\ 0, & \text{διαφορετικά} \end{cases} \quad (2.34)$$

όπου  $T(x', y')$  η δυαδική εικόνα μεγέθους  $M$  φορές το μέγεθος της αρχικής γκρι απόχρωσης εικόνας,  $x = x'/M$ ,  $y = y'/M$  οι ακέραιοι δείκτες των εικονοστοιχείων στην δισδιάστατη εικόνα και  $I_u$  δίνεται από την ακόλουθη σχέση:

$$I_u(x', y') = -b(1-b)^2 F(x', y-1) + (1-2b^2+b^3)F(x', y) + b(1+b-b^2)F(x', y+1) - b^2(1-b)F(x', +2) \quad (2.35)$$

Η συνάρτηση  $F()$  υπολογίζεται όπως παρακάτω:

$$F(x', m) = -a(1-a)^2 I(x-1, m) + (1-2a^2+a^3)I(x, m) + a(1+a-a^2)I(x+1, m) - a^2(1-a)I(+2, m) \quad (2.36)$$

όπου  $a = (x'/M) - x$ ,  $b = (y'/M) - y$ .

Προτείνεται η τιμή  $M = 2$  (διπλάσια δυαδική εικόνα) για καλύτερα ποιοτικά αποτελέσματα.

## 6. Μετά-επεξεργασία

Στο τελικό στάδιο, εκτελείται μετά-επεξεργασία της δυαδικής εικόνας με σκοπό την εξάλειψη του θορύβου, τη βελτίωση της ποιότητας των περιοχών του κειμένου και τη τελική αφαίρεση μικρών κενών και οπών.

Αρχικά, με ανάλυση συνδεδεμένων αντικειμένων υπολογίζεται το μέσο ύψος χαρακτήρων  $l_n$  (σχηματίζεται ένα ιστόγραμμα ύψους των συνδεδεμένων αντικειμένων και επιλέγεται η μεγαλύτερη κορυφή ως το μέσο ύψος χαρακτήρων). Έπειτα από πειραματικές μελέτες, προτείνονται οι ακόλουθη παράμετροι:

$$n = 0.15l_n, k_{sh} = 0.9n^2, k_{sw} = 0.05n^2, dx = dy = 0.25n, k_{sw1} = 0.35n^2$$

Στη συνέχεια ακολουθούν τρία βήματα μετά-επεξεργασίας:

### Βήμα 1°

Χρησιμοποιείται ένα μορφολογικό φίλτρο συρρίκνωσης (shrink filter) για την τελική αφαίρεση του θορύβου από το υπόβαθρο. Εξετάζεται ολόκληρη η δυαδική εικόνα και σε κάθε εικονοστοιχείο κειμένου ( $T(x, y) = 1$ ) κεντράρεται ένα παράθυρο διαστάσεων  $n \times n$ . Αν ισχύει  $P_{sh} > K_{sh}$ , τότε το αντίστοιχο εικονοστοιχείο αλλάζει σε υπόβαθρο ( $T(x, y) = 0$ ), όπου  $P_{sh}$  ο αριθμός εικονοστοιχείων του υποβάθρου εντός του παραθύρου.

### Βήμα 2°

Εφαρμόζεται ένα swell filter για να γεμίσει τα πιθανά κενά, ασυνέχειες και οπές των περιοχών κειμένου. Εξετάζεται ολόκληρη η δυαδική εικόνα και σε κάθε εικονοστοιχείο υποβάθρου ( $T(x, y) = 0$ ) κεντράρεται ένα παράθυρο

διαστάσεων  $n \times n$ . Αν ισχύει  $P_{sw} > K_{sw}$  και  $|x - x_a| < dx$  και  $|y - y_a| < dy$ , τότε το αντίστοιχο εικονοστοιχείο αλλάζει σε κείμενο ( $T(x, y) = 1$ ), όπου  $P_{sw}$  ο αριθμός εικονοστοιχείων του κειμένου εντός του παραθύρου.

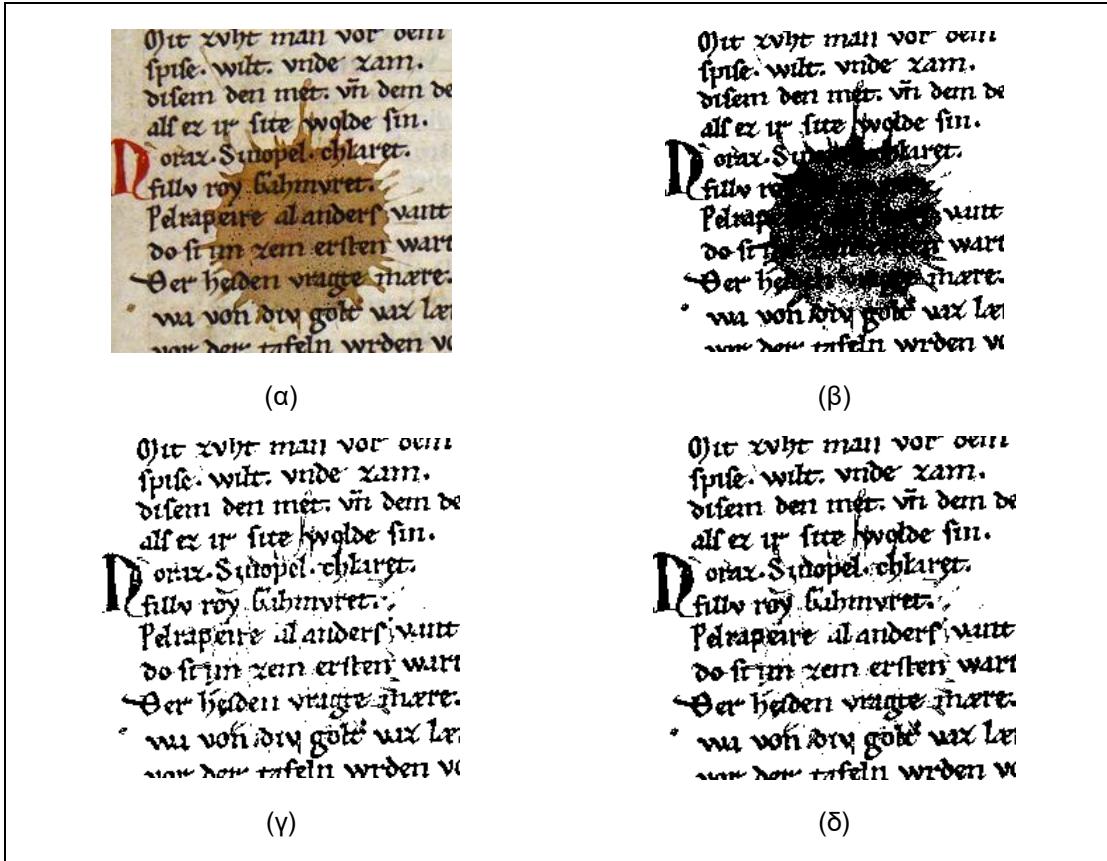
### Βήμα 3°

Τέλος, γίνεται χρήση και πάλι ενός swell filter για τη βελτίωση της ποιότητας αποτύπωσης των χαρακτήρων κειμένου. Εξετάζεται ολόκληρη η δυαδική εικόνα και σε κάθε εικονοστοιχείο υποβάθρου ( $T(x, y) = 0$ ) κεντράρεται ένα παράθυρο διαστάσεων  $n \times n$ . Αν ισχύει  $P_{sw1} > K_{sw1}$  τότε το αντίστοιχο εικονοστοιχείο αλλάζει σε κείμενο ( $T(x, y) = 1$ ), όπου  $P_{sw1}$  ο αριθμός εικονοστοιχείων του κειμένου εντός του παραθύρου.

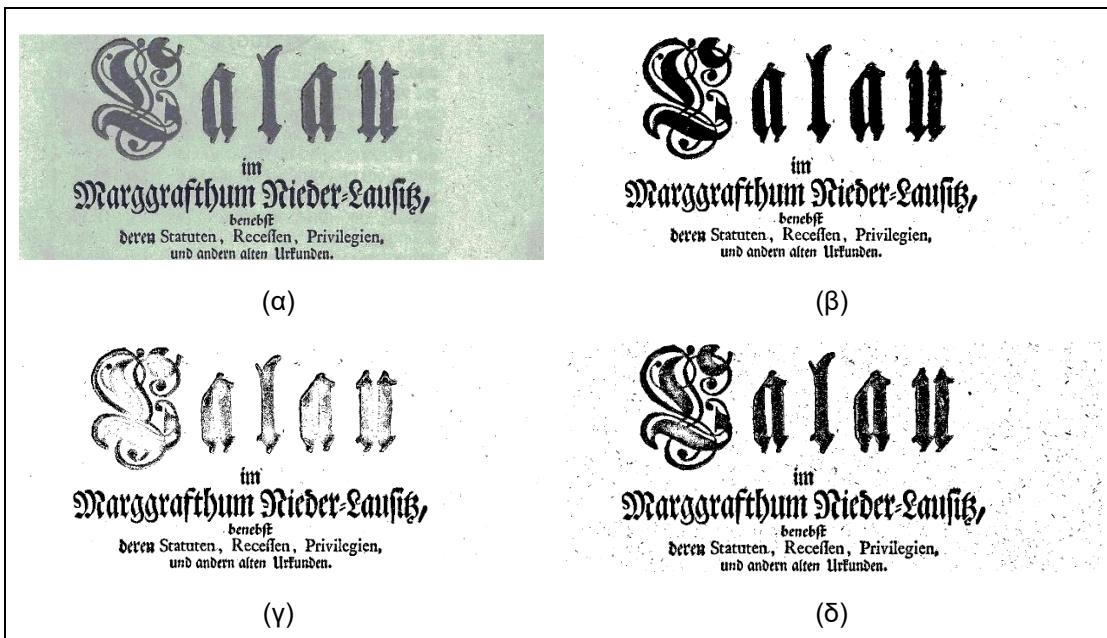
Έτσι, από τη δυαδική εικόνα  $T(x, y)$  εξάγεται η τελική μετά-επεξεργασμένη δυαδική εικόνα  $T_p(x, y)$ .

## 2.5 Σύγκριση τεχνικών δυαδικοποίησης

Η χρήση ενός μοναδικού καθολικού κατωφλιού παρέχει ικανοποιητικά αποτελέσματα για εικόνες, όπου η φωτεινότητα του υποβάθρου διαφέρει για όλες τις περιοχές της εικόνας από τη φωτεινότητα του κειμένου. Ωστόσο, υπάρχουν περιπτώσεις, όπου η φωτεινότητα του υποβάθρου μιας περιοχής είναι παρόμοια με τη φωτεινότητα του κειμένου. Η χρήση ενός καθολικού κατωφλιού κρίνεται ανεπαρκής για αυτές τις περιπτώσεις, διότι είναι αδύνατη η διάσωση όλου του κειμένου και η ταυτόχρονη αφαίρεση όλου του θορύβου από το υπόβαθρο. Εστιάζοντας σε τοπικό επίπεδο, υπάρχει αρκετή αντίθεση για τον διαχωρισμό του κειμένου από το υπόβαθρο και οι τεχνικές που εκμεταλλεύονται την τοπική πληροφορία, προσαρμόζοντας καταλλήλως το κατώφλι για κάθε εικονοστοιχείο ή ευρύτερη περιοχή, παράγουν καλύτερα αποτελέσματα.



Εικόνα 2.15: (α) Αρχική εικόνα με κηλίδα, δυαδική μετατροπή κατά (β) Otsu, (γ) Sauvola και (δ) GPP.



Εικόνα 2.16: (α) Αρχική εικόνα με θόρυβο στο υπόβαθρο και το τίτλο, δυαδική μετατροπή κατά (β) Otsu, (γ) Sauvola και (δ) GPP.

98 *The First Part of HENRY IV.*

K. HENRY.

Heav'n pardon thee. Yet let me wonder, Harry,  
 At thy affections, which do hold a wing  
 Quite from the flight of all thy ancestors.  
 Thy place in council thou haft rudely lost,

(α)

98 *The First Part of HENRY IV.*

K. HENRY.

Heav'n pardon thee. Yet let me wonder, Harry,  
 At thy affections, which do hold a wing  
 Quite from the flight of all thy ancestors.  
 Thy place in council thou haft rudely lost,

(γ)

98 *The First Part of HENRY IV.*

K. HENRY.

Heav'n pardon thee. Yet let me wonder, Harry,  
 At thy affections, which do hold a wing  
 Quite from the flight of all thy ancestors.  
 Thy place in council thou haft rudely lost,

(β)

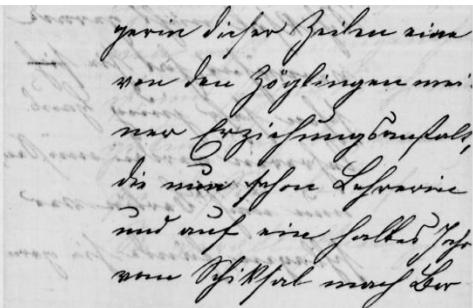
98 *The First Part of HENRY IV.*

K. HENRY.

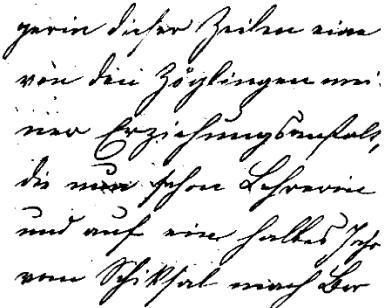
Heav'n pardon thee. Yet let me wonder, Harry,  
 At thy affections, which do hold a wing  
 Quite from the flight of all thy ancestors.  
 Thy place in council thou haft rudely lost,

(δ)

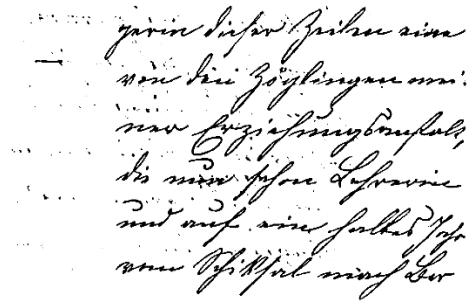
Εικόνα 2.17: (α) Αρχική εικόνα με αχνούς χαρακτήρες, δυαδική μετατροπή κατά (β) Otsu, (γ) Sauvola και (δ) GPP.



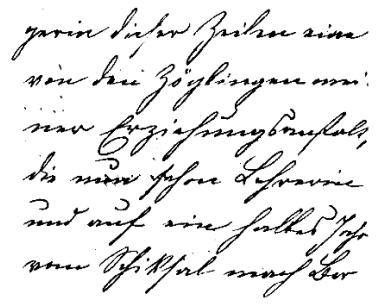
(α)



(β)

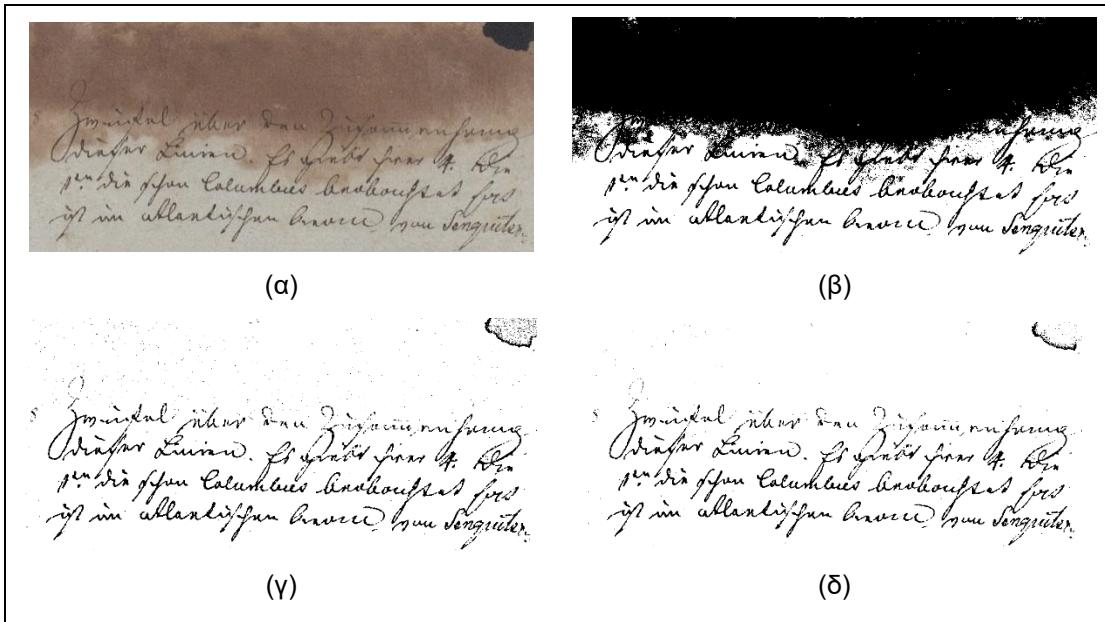


(γ)



(δ)

Εικόνα 2.18: (α) Αρχική εικόνα με χαρακτήρες πίσω σελίδας, δυαδική μετατροπή κατά (β) Otsu, (γ) Sauvola και (δ) GPP.



Εικόνα 2.19: (α) Αρχική εικόνα με μεταβαλλόμενη φωτεινότητα στο υπόβαθρο, δυαδική μετατροπή κατά (β) Otsu, (γ) Sauvola και (δ) GPP.

Με βάση τα παραπάνω, η καθολική τεχνική κατωφλίωσης του Otsu [1] κρίνεται μη αποτελεσματική για έγγραφα υποβαθμισμένης ποιότητας με μεταβαλλόμενη αντίθεση και διαφορετικά μοτίβα στο υπόβαθρο. Μοναδικό της πλεονέκτημα είναι η χαμηλή πολυπλοκότητα και η ταχύτητα εκτέλεσης του αλγορίθμου.

Οι τεχνικές προσαρμοζόμενης κατωφλίωσης Sauvola [2] και GPP [3] δίνουν καλύτερα αποτελέσματα, καθώς είναι ικανές να αφαιρέσουν το ανομοιόμορφο υπόβαθρο ακόμα και σε παλιά έγγραφα, όπου έχουμε χαμηλή ποιότητα χαρακτήρων, σκιές, μη ομοιόμορφη φωτεινότητα υποβάθρου κτλ. Πιο συγκεκριμένα η προσέγγιση Sauvola [2] εμφανίζει ικανοποιητική απόδοση αλλά σε συγκεκριμένες κατηγορίες εγγράφων, εισάγει ποσοστά θορύβου στο αποτέλεσμα και οι χαρακτήρες των λέξεων εμφανίζονται “λεπτοί” και “σπασμένοι”. Η μέθοδος δυαδικοποίησης GPP [3] επιτυγχάνει ανώτερη δυαδική μετατροπή από τις παραπάνω και αντεπεξέρχεται με μεγάλη επιτυχία στα διάφορα είδη θορύβου. Ωστόσο, και στους δύο αλγορίθμους, σε εικόνες με αρκετά έντονο θόρυβο, η αυτόματη κατωφλίωση με τις προτεινόμενες προκαθορισμένες παραμέτρους δεν καταλήγει πάντα σε πολύ αξιόλογα αποτελέσματα και χρειάζεται η χειροκίνητη ρύθμιση τους από το χρήστη για τη βέλτιστη δυαδικοποίηση.

Τα παραπάνω συμπεράσματα παρέχουν ποιοτική ανάλυση των αποτελεσμάτων, αφού είναι βασισμένα στον ανθρώπινο παράγοντα και συνεπώς αποτελούν υποκειμενική αξιολόγηση.

Παράλληλα, κρίνεται απαραίτητη και η ποσοτική αξιολόγηση με τη χρήση μιας υπολογιστικής μεθόδου, η οποία απαντάται με τον όρο αντικειμενική αξιολόγηση. Στην παρούσα διπλωματική λοιπόν, ακολουθήθηκε αξιολόγηση με βάση τα εικονοστοιχεία, η οποία πραγματοποιείται με τη μία προς μία

αντιστοίχιση των εικονοστοιχείων μεταξύ της δυαδικής Ground Truth εικόνας και της εικόνας της δυαδικής μετατροπής.

Επίσης, για την αποτίμηση του αποτελέσματος της δυαδικής μετατροπής χρησιμοποιήθηκαν μετρικές αξιολόγησης που βασίζονται στον αριθμό των σωστών και λανθασμένων εικονοστοιχείων. Συγκεκριμένα υπολογίστηκαν οι μετρικές Recall (Ανάκληση), Precision (Ακρίβεια) και F-Measure. Οι μετρικές αυτές στηρίζονται στον υπολογισμό των True Positive (TP), False Positive (FP) και False Negative (FN).

- Ένα εικονοστοιχείο  $P(x, y)$  ορίζεται ως True Positive (TP), όταν  $P(x, y) = 1$  (μαύρο) και στις δύο εικόνες του Ground Truth και του αποτελέσματος της δυαδικής μετατροπής.
- Ένα εικονοστοιχείο  $P(x, y)$  ορίζεται ως False Positive (FP), όταν  $P(x, y) = 1$  (μαύρο) μόνο στο αποτέλεσμα της δυαδικής μετατροπής.
- Ένα εικονοστοιχείο  $P(x, y)$  ορίζεται ως False Negative (FN), όταν  $P(x, y) = 1$  (μαύρο) μόνο στην Ground Truth εικόνα.

Αν  $C_{TP}$ ,  $C_{FP}$ ,  $C_{FN}$  ο αριθμός των TP, FP, FN εικονοστοιχείων αντίστοιχα, τότε οι μετρικές αξιολόγησης Recall ( $RC$ ) και Precision ( $PR$ ) δίνονται από τους παρακάτω τύπους:

$$RC = \frac{C_{TP}}{C_{FN} + C_{TP}} \quad (2.37)$$

$$PR = \frac{C_{TP}}{C_{FP} + C_{TP}} \quad (2.38)$$

Μια μετρική που συνδυάζει το Precision και το Recall είναι ο μέσος όρος τους F-Measure ( $FM$ ):

$$FM = \frac{2 \times RC \times PR}{RC + PR} \quad (2.39)$$

Για την αξιολόγηση χρησιμοποιήθηκαν εικόνες από τους διαγωνισμούς DIBCO 2017 [4] και H-DIBCO 2018 [5]. Ο DICBO 2017 [4] περιέχει 20 εικόνες εγγράφου, οι οποίες εμφανίζουν υψηλά ποσοστά θορύβου, κυρίως χαρακτήρων της πίσω σελίδας. Ωστόσο, περιέχει και εικόνες με άλλα είδη θορύβου όπως κηλίδες και αχνούς χαρακτήρες. Από την άλλη, οι 10 εικόνες του H-DIBCO 2018 [5] αποτελούν χειρόγραφα έγγραφα, τα οποία εμφανίζουν κυρίως διαφορετικά επίπεδα φωτεινότητας στο υπόβαθρο τους σε συνδυασμό με τα προαναφερόμενα είδη θορύβου. Στους παρακάτω πίνακες αποτυπώνονται τα αποτελέσματα της αξιολόγησης.

DIBCO 2017	Otsu			Sauvola			GPP		
εικόνα	RC	PR	FM	RC	PR	FM	RC	PR	FM
1	96	35	51	71	82	76	56	95	71
2	96	88	92	84	99	91	73	99	84
3	99	32	48	87	88	87	72	98	83
4	100	22	36	82	77	80	67	90	77
5	98	37	53	91	67	77	82	87	84
6	94	82	88	97	85	91	91	90	91
7	97	79	87	96	85	90	91	91	91
8	85	91	88	96	83	89	93	89	91
9	84	84	84	95	87	91	94	90	92
10	87	87	87	92	92	92	90	94	92
11	98	89	93	88	97	92	83	98	90
12	97	69	81	75	73	83	79	93	84
13	100	51	67	87	76	81	80	82	81
14	97	75	85	89	83	86	84	86	85
15	97	81	88	85	96	90	80	98	88
16	93	84	88	90	96	93	84	98	90
17	78	93	85	91	83	87	89	85	87
18	92	79	85	92	82	86	86	85	86
19	91	75	83	83	92	87	81	94	87
20	81	90	85	67	98	80	67	98	80
<b>Μέσος όρος</b>	93	71,2	77,7	86,9	86	86,5	81	92	85,7

Πίνακας 2.1: Αποτελέσματα δυαδικής μετατροπής εικόνων εγγράφου DIBCO 2017.

H-DIBCO 2018	Otsu			Sauvola			GPP		
εικόνα	RC	PR	FM	RC	PR	FM	RC	PR	FM
1	84	29	43	86	71	77	87	71	78
2	83	9	16	79	49	60	73	63	67
3	83	84	84	88	86	87	88	87	88
4	64	14	23	80	75	78	80	81	81
5	55	18	27	82	44	57	75	50	60
6	95	48	64	99	71	83	92	95	94
7	63	53	58	96	61	75	92	78	84
8	90	74	81	80	89	84	83	86	85
9	93	29	45	86	85	86	90	86	88
10	88	64	74	82	83	83	72	83	77
<b>Μέσος όρος</b>	79,8	42,2	51,5	85,8	71,4	77	83,2	78	80,2

Πίνακας 2.2: Αποτελέσματα δυαδικής μετατροπής εικόνων εγγράφου H-DIBCO 2018.

	Recall (RC)	Precision (PR)	F-Measure (FM)
Otsu	88,60	61,50	68,97
Sauvola	86,53	81,17	83,30
GPP	81,80	87,33	83,87

Πίνακας 2.3: Μέσος όρος των συνολικών αποτελεσμάτων της αξιολόγησης.

Όπως φαίνεται από τους παραπάνω πίνακες, η μέθοδος Otsu [1] πετυχαίνει υψηλά ποσοστά ανάκλησης (Recall), καθώς εντοπίζει τα περισσότερα εικονοστοιχεία του κειμένου. Ωστόσο, αυτό είναι πλασματικό, αφού συμβαίνει λόγω της μεγάλης εισαγωγής θορύβου στη δυαδική εικόνα, γεγονός που φαίνεται από τα χαμηλά ποσοστά ακρίβειας (Precision).

Οι μέθοδοι Sauvola [2] και GPP [3] αναμενόμενα παρουσιάζουν πολύ καλύτερη απόδοση. Συγκεκριμένα, η μέθοδος Sauvola επιτυγχάνει ελαφρώς μεγαλύτερα ποσοστά ανάκλησης (Recall) από τη μέθοδο GPP, η οποία αντίστροφα δίνει καλύτερα ποσοστά ακρίβειας (Precision). Επίσης, και στις δύο τεχνικές απαιτείται η χειροκίνητη ρύθμιση των παραμέτρων τους, για να επιτευχθεί το καλύτερο δυαδικό αποτέλεσμα. Κυρίως χρειάζεται η ρύθμιση των παραμέτρων  $k, R$  της εξίσωσης Sauvola 2.20, παράμετροι που αφορούν και την μέθοδο GPP για την εκτίμηση του κειμένου και του υποβάθρου (εξίσωση 2.28). Στα περισσότερα έγγραφα απαιτήθηκαν χαμηλές τιμές  $k \leq 0.3$  και  $R \leq 100$ , ώστε να εντοπιστεί περισσότερο κείμενο. Οι χαμηλές αυτές τιμές όμως συνεπάγονται με ενδεχόμενη εισαγωγή περισσότερου θορύβου. Έτσι, συνολικά οι τεχνικές Sauvola [2] και GPP [3] εμφανίζουν ικανοποιητικά αποτελέσματα και καταφέρνουν να βρουν σωστά το κείμενο, αλλά για να το πετύχουν σε ορισμένες περιπτώσεις εισάγουν ανεπιθύμητο θόρυβο στο τελικό δυαδικό αποτέλεσμα, περισσότερο δε η προσέγγιση Sauvola.

## 2.6 Περί κατάτμησης

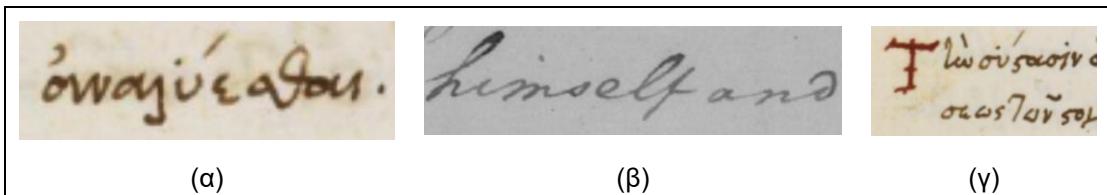
Η κατάτμηση εικόνας είναι μια πολύ σημαντική διαδικασία του τομέα της ψηφιακής επεξεργασίας. Συνήθως, αποτελεί ένα ενδιάμεσο βήμα, έπειτα του σταδίου προ-επεξεργασίας (αφαίρεση θορύβου, δυαδική μετατροπή κτλ.) και προηγείται του τελικού σταδίου αναγνώρισης (εξαγωγή χαρακτηριστικών, ταξινόμηση κτλ.). Συνθέτει ένα από τα πιο δύσκολα προβλήματα στον τομέα στον τομέα της όρασης υπολογιστών και ένα ανοιχτό πεδίο έρευνας.



**Εικόνα 2.20:** Τα κυριότερα στάδια της οπτικής ψηφιακής επεξεργασίας εικόνων εγγράφου.

Γενικά ως κατάτμηση εικόνας, ορίζεται η διαδικασία διαχωρισμού μιας εικόνας σε διακριτές, ομοιογενείς υπό-περιοχές, ασυνεχείς μεταξύ τους και μη αλληλεπικαλυπτόμενες. Οι περιοχές αυτές είναι ομοιογενείς ως προς κάποιο χαρακτηριστικό της εικόνας, όπως η φωτεινότητα, το χρώμα ή η υφή. Η κάθε περιοχή είναι αρκετά διαφορετική από τις γειτονικές περιοχές, και διαχωρίζεται με σαφήνεια από αυτές. Το αποτέλεσμα της κατάτμησης, εκτός από ένα σύνολο από ομοιογενείς περιοχές που καλύπτουν ολόκληρη την εικόνα, ενδέχεται να είναι και ένα σύνολο από περιγράμματα που έχουν εξαχθεί από αυτήν. Όσον αφορά, τις εικόνες εγγράφου που μελετά αυτή η διπλωματική, η κατάτμηση των ψηφιακών εγγράφων αφορά τη διαδικασία του εντοπισμού και της ομαδοποίησης των δομικών τους στοιχείων σε ομοιογενή τμήματα με σκοπό την περιγραφή και την ανάλυση της δομής τους. Οι δομές των εγγράφων ταξινομούνται σε δύο κατηγορίες, τις φυσικές (κείμενο, εικόνες, γραφήματα, πίνακες, γραμμές κ.τ.λ.) και τις λογικές (τίτλος, παράγραφος, επικεφαλίδες, λεζάντες κ.τ.λ.).

Η δυαδική μετατροπή που αναλύθηκε προηγουμένως στο κεφάλαιο αποτελεί και αυτή ένα είδος κατάτμησης, αφού διαχωρίζει τις περιοχές υποβάθρου και κειμένου σε μια εικόνα εγγράφου. Παρακάτω εξετάζεται το επόμενο βήμα της οπτικής επεξεργασίας, ανάλυσης και αναγνώρισης, το οποίο είναι το στάδιο της εξαγωγής τμημάτων γραμμών κειμένου, λέξεων και χαρακτήρων. Συγκεκριμένα, περιγράφεται η μεθοδολογία κατάτμησης λέξεων που υλοποιήθηκε σε αυτή τη διπλωματική και βασίζεται στον προσαρμοστικό αλγόριθμο εξομάλυνσης κατά μήκος διαδρομής-ARLSA (Adaptive Run Length Smoothing Algorithm). Επρόκειτο για τεχνική τμηματοποίησης (κατάτμησης) τυπωμένων ιστορικών ψηφιακών εγγράφων. Τα έγγραφα αυτά ενδέχεται να παρουσιάζουν ιδιαίτερα χαρακτηριστικά τα οποία καθιστούν την τμηματοποίησή τους αρκετά δύσκολη διαδικασία. Ορισμένα από αυτά τα χαρακτηριστικά, είναι η χαμηλή ποιότητα εκτύπωσης που εμφανίζεται ως θόρυβος, η τοπική παραμόρφωση κλίσης του κειμένου (λοξό ή στρεβλό κείμενο), η μεταβλητή απόσταση μεταξύ των γραμμών κειμένου, λέξεων και χαρακτήρων, η ύπαρξη κειμένου ή άλλων γραφικών στοιχείων στο περιθώριο της σελίδας, τα διαφορετικά μεγέθη και γραμματοσειρά στους χαρακτήρες κ.α. Συνεπάγεται λοιπόν, πως το αποτέλεσμα της κατάτμησης (βαθμός διαχωρισμού λέξεων, γραμμών ή χαρακτήρων) επηρεάζει καταλυτικά την αποδοτικότητα της τμηματοποίησης κειμένου.

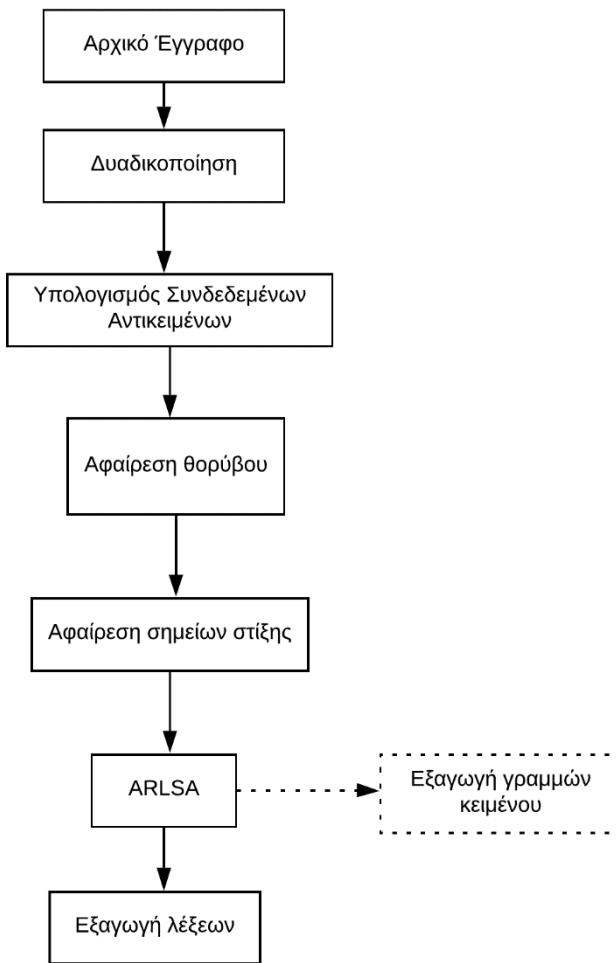


**Εικόνα 2.21: Παραδείγματα τυπωμένων ιστορικών εικόνων με ιδιαίτερα χαρακτηριστικά.**  
**(α) Διαφορετικό ύψος, πάχος και απόσταση χαρακτήρων, (β) λοξό κείμενο, (γ) γραφικό στοιχείο στο περιθώριο της σελίδας.**

## 2.7 Η μεθοδολογία κατάτμησης της διπλωματικής

### 2.7.1 Συνοπτική περιγραφή της μεθοδολογίας

Η μέθοδος κατάτμησης, που υλοποιήθηκε στην παρούσα διπλωματική, αφορά τη κατάτμηση λέξεων σε τυπωμένα ιστορικά έγγραφα και βασίζεται στην εργασία των Nikolaou, Makridis, Gatos, Stamatopoulos και Papamarkos [6], όπου γίνεται χρήση ενός προσαρμοστικού αλγορίθμου εξομάλυνσης κατά μήκος διαδρομής ARLSA (Adaptive Run Length Smoothing Algorithm). Επρόκειτο για τεχνική η οποία συνδυάζει ιδέες από την επεξεργασία συνδεδεμένων αντικειμένων και την ανάλυση υποβάθρου. Η αρχική γκρι απόχρωσης εικόνα μετατρέπεται σε δυαδική μέσω της προσαρμοστικής τεχνικής δυαδικοποίησης GPP [3] που εξετάστηκε προηγουμένως στο κεφάλαιο. Ο χρήστης βέβαια έχει τη δυνατότητα να επιλέξει ανάμεσα και στις άλλες τεχνικές κατωφλίωσης Otsu [1] και Sauvola [2] που υλοποιήθηκαν. Έπειτα, εξάγονται τα συνδεδεμένα αντικείμενα που περιέχονται στο έγγραφο. Η δυαδική εικόνα φιλτράρεται έτσι ώστε να απομακρυνθούν τα μικρά αντικείμενα θορύβου, τα οποία επηρεάζουν τις στατικές ιδιότητες του εγγράφου και δύναται να οδηγήσουν σε λανθασμένα αποτελέσματα. Επίσης, εντοπίζονται τα σημεία στίξης και αφαιρούνται από το έγγραφο. Στη συνέχεια, στην εικόνα που προκύπτει μετά το φιλτράρισμα του θορύβου, εφαρμόζεται ο προσαρμοστικός αλγόριθμος εξομάλυνσης ARLSA [6]. Τέλος, πραγματοποιείται η κατάτμηση λέξεων με τη χρήση ιστογράμματος των κάθετων προβολών των αντικειμένων που ομαδοποιήθηκαν μετά την εφαρμογή του ARLSA. Βάση του ιστογράμματος υπολογίζονται οι κατάλληλοι παράμετροι, οι οποίοι χρησιμοποιούνται για τη τελική τμηματοποίηση των λέξεων.



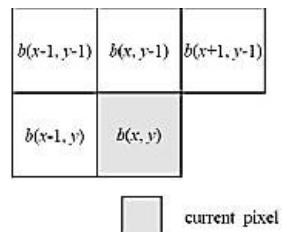
**Εικόνα 2.22: Διάγραμμα ροής της μεθοδολογίας κατάτμησης γραμμών κειμένου και λέξεων σε τυπωμένα ιστορικά έγγραφα.**

Παρόλο που η υλοποιημένη τεχνική αφορά κυρίως την εξαγωγή λέξεων, υπάρχει και η δυνατότητα εξαγωγής των γραμμών κειμένου εφαρμόζοντας τον ARLSA με διαφορετικές παραμέτρους από τις προκαθορισμένες.

## 2.7.2 Συνδεδεμένα αντικείμενα

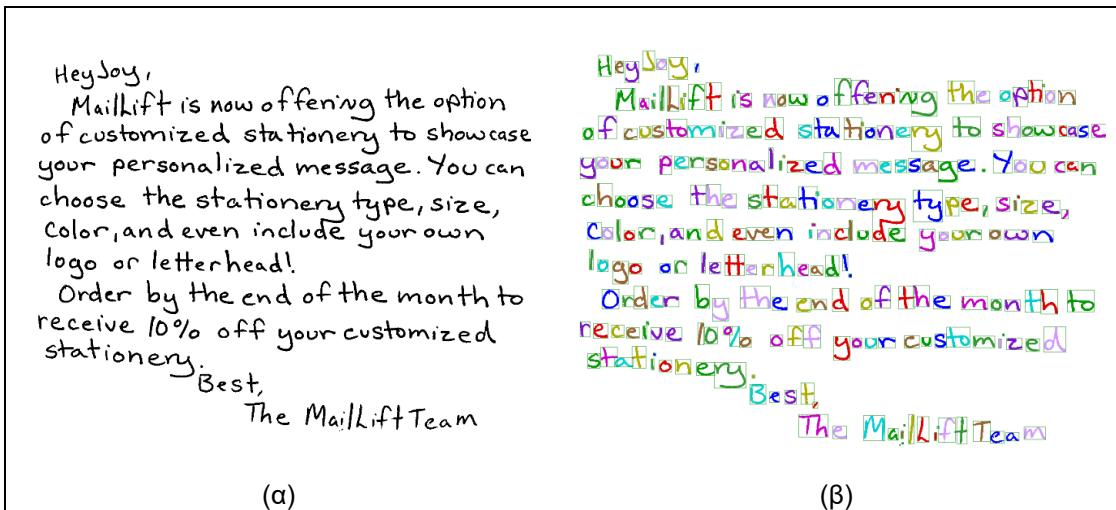
Με την τεχνική των συνδεδεμένων αντικειμένων (connected components analysis) επιτυγχάνεται ο εντοπισμός και η εξαγωγή όλων των αντικειμένων (κοινές περιοχές προσκηνίου) που περιέχονται σε μια δυαδική εικόνα εγγράφου. Για διαδικασία αυτή ακολουθούνται τα παρακάτω βήματα:

- Δημιουργία ενός μετρητή ετικετών περιοχών προσκηνίου.
- Σάρωση της εικόνας από αριστερά προς δεξιά και από πάνω προς τα κάτω.
  - Για κάθε pixel (εικονοστοιχείο) προσκηνίου εξετάζεται το αριστερό, το πάνω-αριστερό, το πάνω και το πάνω-δεξιό pixel (γειτονικά pixels).



**Εικόνα 2.23: Γειτονικά εικονοστοιχεία προς εξέταση.**

- Αν κανένα από τα γειτονικά pixels δεν ανήκει στο προσκήνιο, τότε στο εξεταζόμενο pixel θέτεται μια καινούργια ετικέτα και αυξάνεται ο μετρητής ετικετών κατά ένα.
- Αν μόνο ένα γειτονικό pixel ανήκει στο προσκήνιο, τότε στο εξεταζόμενο pixel θέτεται η ετικέτα του γειτονικού pixel.
- Αν περισσότερα γειτονικά pixels ανήκουν στο προσκήνιο και όλα έχουν την ίδια ετικέτα, τότε στο εξεταζόμενο pixel θέτεται η κοινή ετικέτα.
- Αν περισσότερα γειτονικά pixels ανήκουν στο προσκήνιο αλλά έχουν διαφορετικές ετικέτες, τότε στο εξεταζόμενο pixel θέτεται η ετικέτα του πάνω-δεξιού pixel και όλες οι γειτονικές ετικέτες σχετίζονται ισοδύναμες.
- Σαρώνεται η εικόνα, όλες οι σχετικά ισοδύναμες ετικέτες ομαδοποιούνται σε μια κοινή ετικέτα.
- Έτσι όλες οι κοινές περιοχές προσκηνίου ενώνονται και σχηματίζονται τα συνδεδεμένα αντικείμενα.
- Για κάθε αντικείμενο υπολογίζονται οι διαστάσεις των ορθογωνίων πλαισίων που το περικλείουν.

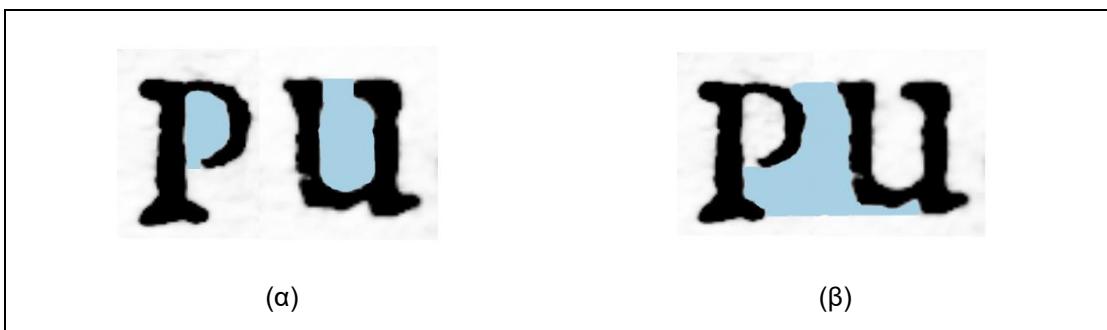


Εικόνα 2.24: Χρωματική απεικόνιση των συνδεδεμένων αντικειμένων. (α) Δυαδική εικόνα, (β) εξαγμένα συνδεδεμένα αντικείμενα.

### 2.7.3 Προσαρμοστικός αλγόριθμος εξομάλυνσης

Ένας από τους πιο γνωστούς αλγορίθμους για την επεξεργασία της δομής εικόνας εγγράφου είναι η εξομάλυνση κατά μήκος διαδρομής RLSA. Η τμηματοποίηση των λέξεων βασίζεται στην εφαρμογή εξομάλυνσης κατά την οριζόντια ή την κάθετη κατεύθυνση. Για κάθε κατεύθυνση, τα συνεχόμενα λευκά τμήματα του υποβάθρου, που έχουν μήκος μεγαλύτερο από ένα κατώφλι, μετατρέπονται σε μαύρα τμήματα του προσκηνίου, ώστε να σχηματιστούν οι λέξεις. Ο προσαρμοστικός Αλγόριθμος Εξομάλυνσης ARLSA [6] αποτελεί μια τροποποιημένη έκδοση του RLSA για τυπωμένα ιστορικά έγγραφα, με πλεονέκτημα την ορθότερη ομαδοποίηση όμοιων συνδεδεμένων αντικειμένων.

Οι ακολουθίες λευκών εικονοστοιχείων υποβάθρου, μπορούν να χωριστούν σε δύο κατηγορίες, η πρώτη περιλαμβάνει τις ακολουθίες που παρεμβάλλονται μεταξύ εικονοστοιχείων του ίδιου αντικειμένου, ενώ η δεύτερη αφορά ακολουθίες που παρεμβάλλονται μεταξύ εικονοστοιχείων δύο διαφορετικών αντικειμένων.



Εικόνα 2.25: Οι δύο τύποι ακολουθίων εικονοστοιχείων υποβάθρου. (α) Εικονοστοιχεία μεταξύ του ίδιου αντικειμένου, (β) εικονοστοιχεία ανάμεσα σε δύο διαφορετικά αντικείμενα.

Τα εικονοστοιχεία που ανήκουν στη πρώτη περίπτωση, μετατρέπονται σε μαύρα εικονοστοιχεία προσκηνίου. Στη δεύτερη όμως, η αντικατάσταση των εικονοστοιχείων υποβάθρου με εικονοστοιχεία προσκηνίου γίνεται μόνο όταν ικανοποιείται η παρακάτω σχέση:

$$(L(s) \leq T_l) \wedge (H_R(s) \leq T_h) \wedge (O_H(s) \geq T_o) \wedge (N(s) = 1) \quad (2.40)$$

Τα παραπάνω μεγέθη αναφέρονται σε:

- $L(s)$ : το μήκος μια ακολουθίας  $S(i, j)$  εικονοστοιχείων υποβάθρου μεταξύ δύο συνδεδεμένων αντικειμένων  $CC_i$  και  $CC_j$ .
- $H_R(s)$ : ο λόγος υψών των δύο αντικειμένων  $CC_i$  και  $CC_j$ , ο οποίος ορίζεται ως εξής:

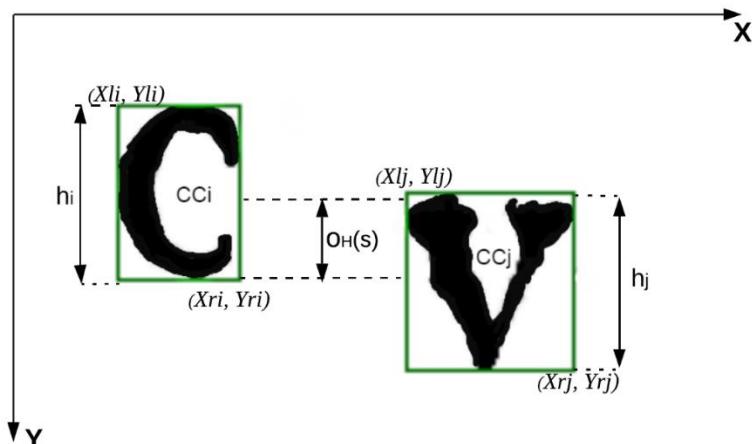
$$H_R(s) = \frac{\max(h_i, h_j)}{\min(h_i, h_j)} \quad (2.41)$$

όπου  $h_i, h_j$  τα ύψη των  $CC_i$  και  $CC_j$  αντίστοιχα.

- $O_H(s)$ : η επικάλυψη κατά την οριζόντια κατεύθυνση των ορθογωνίων πλαισίων που περικλείουν τα  $CC_i$  και  $CC_j$ .

$$O_H(s) = \max(Yl_i, Yl_j) - \min(Yr_i, Yr_j) \quad (2.42)$$

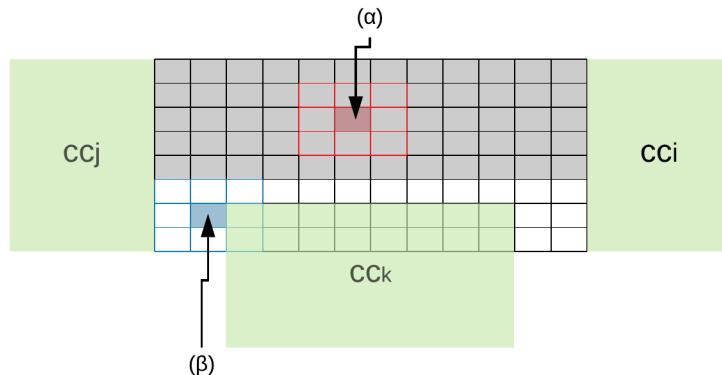
όπου  $(Xl_i, Yl_i)$  και  $(Xr_i, Yr_i)$  οι συντεταγμένες της πάνω αριστερά και κάτω δεξιά γωνίας του παραλληλογράμμου πλαισίου των συνδεδεμένων αντικειμένων.



Εικόνα 2.26: Οριζόντια επικάλυψη μεταξύ δύο αντικειμένων.

- $N(s)$ : συνάρτηση δυαδικής εξόδου, της οποίας η έξοδος λαμβάνει την τιμή 0, όταν στην  $3 \times 3$  γειτονιά ενός τουλάχιστον εικονοστοιχείου της ακολουθίας  $S$  υπάρχει έστω και ένα άλλο εικονοστοιχείο κάποιου άλλου αντικειμένου  $CC_k$  με  $k \neq i, j$ , αλλιώς παίρνει τη τιμή 1.

Σκοπός της συνάρτησης  $N(s)$  είναι να αποτρέψει τη λανθασμένη σύνδεση αντικειμένων και αποδεικνύεται πολύ χρήσιμη σε έγγραφα με πολύ στενό περιθώριο ανάμεσα στις γραμμές του κειμένου.



**Εικόνα 2.27:** Γραφική επεξίγηση της συνάρτησης  $N(s)$ . (α)  $N(s) = 1$ , στην  $3 \times 3$  γειτονιά των εικονοστοιχείων υποβάθρου (απεικόνιση με γρι χρώμα) δεν παρεμβάλλεται κανένα άλλο αντικείμενο εκτός των  $CC_i$  και  $CC_j$ , (β)  $N(s) = 0$ , στην  $3 \times 3$  γειτονιά των εικονοστοιχείων υποβάθρου (απεικόνιση με λευκό χρώμα) παρεμβάλλεται ένα τρίτο αντικείμενο  $CC_k$ .

- $T_l$ : κατώφλι με το οποίο ελέγχεται το μήκος της ακολουθίας  $S$  και σχετίζεται με τα ύψη  $h_i, h_j$  των αντικειμένων  $CC_i$  και  $CC_j$  αντίστοιχα.

$$T_l = \alpha \cdot \min(h_i, h_j) \quad (2.43)$$

όπου  $\alpha$  σταθερή τιμή.

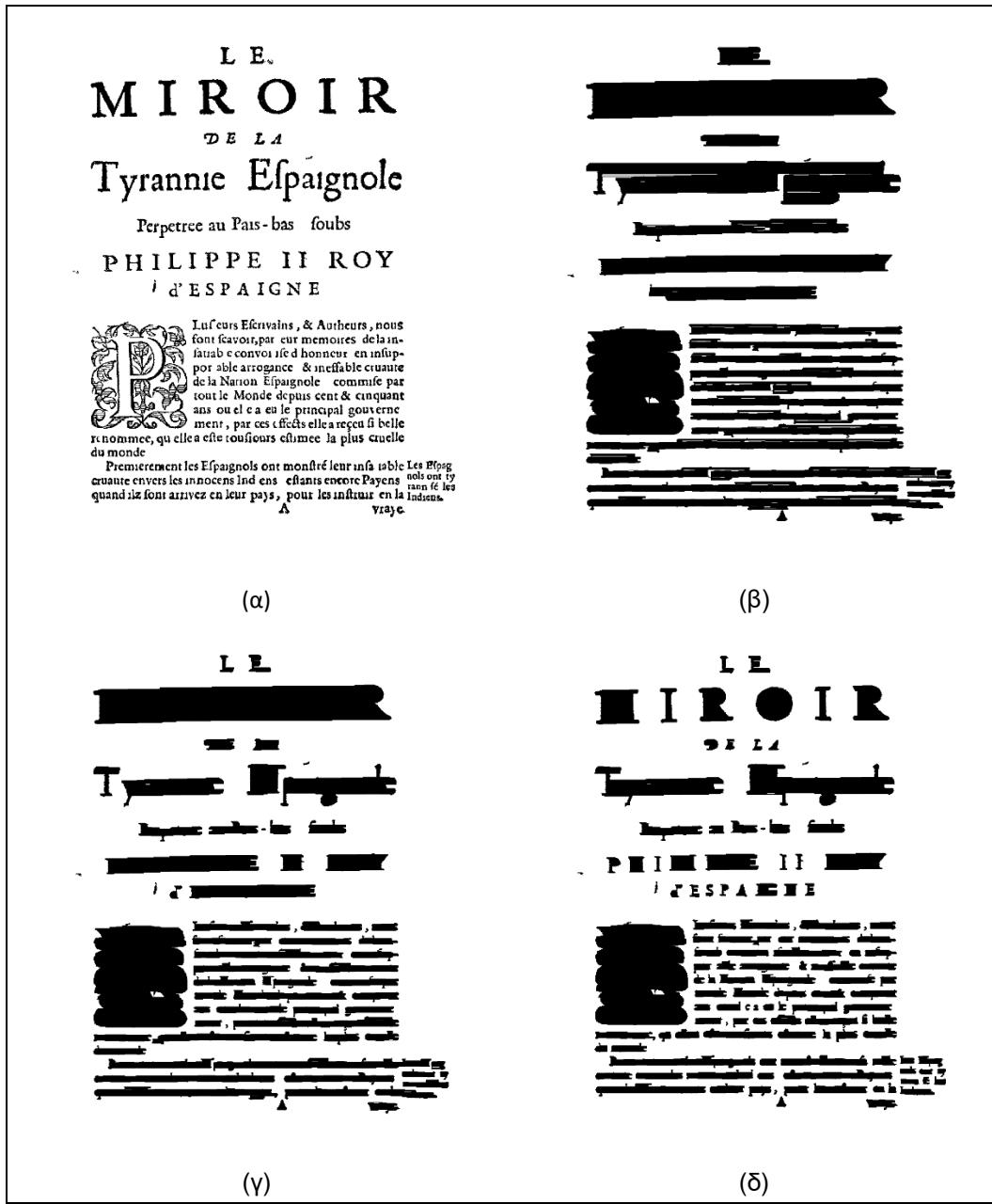
- $T_o$ : κατώφλι με το οποίο ρυθμίζεται το ποσοστό της οριζόντιας επικάλυψης που θα πρέπει να έχουν τα δύο αντικείμενα, εκφράζεται συνάρτηση του αντικειμένου με το μικρότερο ύψος.

$$T_o = c \cdot \min(h_i, h_j) \quad (2.44)$$

όπου  $c \in [0,1]$  συντελεστής που εκφράζει το ποσοστό επικάλυψης.

Στην εργασία του ARLSA [6] προτείνεται η τιμή 0.4. Η πρακτική σημασία αυτής της τιμής εξηγεί πως για να είναι δυνατή η σύνδεση των δύο αντικειμένων  $CC_i$  και  $CC_j$ , θα πρέπει το αντικείμενο με το μικρότερο ύψος, να επικαλύπτεται στην οριζόντια διεύθυνση κατά 40%.

- $T_h$ : κατώφλι το οποίο ορίζει τη μέγιστη ποσοστιαία διαφορά που μπορούν να έχουν τα ύψη των δύο αντικειμένων  $CC_i$  και  $CC_j$ . Για το είδος των εγγράφων που μελετώνται προτείνεται η τιμή 3.5. Με βάση αυτή τη τιμή, ένα αντικείμενο  $CC_i$  δεν μπορεί να έχει πάνω από 3.5 φορές το ύψος ενός γειτονικού αντικειμένου  $CC_j$  (ισχύει και το αντίστροφο).

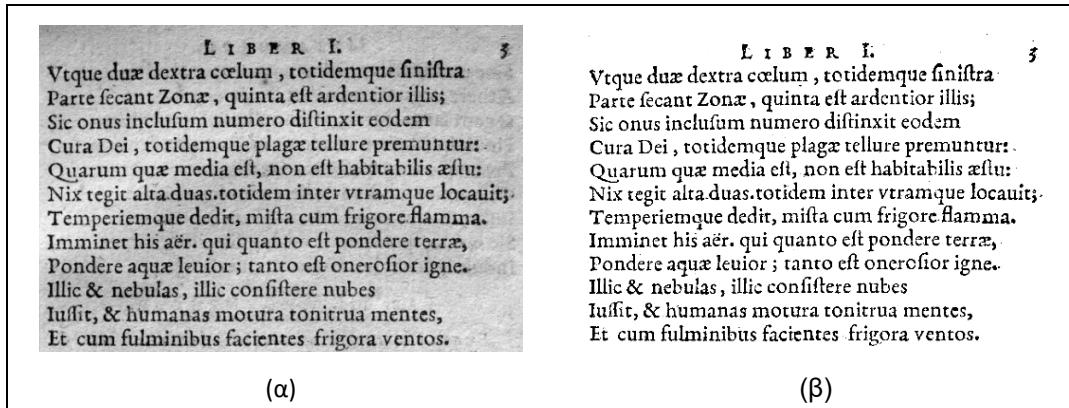


**Εικόνα 2.28:** Παράδειγμα χρήσης του προσαρμοστικού αλγορίθμου εξομάλυνσης. (α) Αρχική δυαδική εικόνα, (β) ARLSA με  $\alpha = 5$ ,  $T_h = 3.5$ ,  $c = 0.4$ , τα αντικείμενα συγχωνεύονται ώστε να σχηματιστούν οι γραμμές του κειμένου, (γ)-(δ) ARLSA με  $T_h = 3.5$ ,  $c = 0.4$ , τα αντικείμενα συγχωνεύονται ώστε να σχηματιστούν ομάδες λέξεων ( $\alpha = 1$ ) ή τμήματα λέξεων ( $\alpha = 0.5$ ).

## 2.7.4 Ανάλυση της Μεθοδολογίας Κατάτμησης λέξεων

Για την κατάτμηση λέξεων ακολουθούνται τα εξής βήματα:

- 1) Η αρχική γκρι απόχρωσης (greyscale) εικόνα δυαδικοποιείται με τη μέθοδο GPP [3] και προκύπτει η δυαδική εικόνα  $I_o$ .



Εικόνα 2.29: Δυαδική μετατροπή GPP. (α) Greyscale εικόνα, (β) αποτέλεσμα κατωφλίωσης.

- 2) Στη συνέχεια, εξάγονται τα συνδεδεμένα αντικείμενα της εικόνας  $I_o$  και υπολογίζονται οι διαστάσεις (πλάτος, ύψος) των ορθογωνίων πλαισίων που τα περικλείουν.

L I B E R I .

Vtque duæ dextra cœlum , totidemque finistra  
 Parte fecant Zonæ , quinta est ardentior illis;  
 Sic onus inclusum numero distinxit eodem  
 Cura Dei , totidemque plagæ tellure premuntur:  
 Quarum quæ media est, non est habitabilis æstu:  
 Nix tegit alta duas. totidem inter vtramque locauit;  
 Temperiemque dedit, mista cum frigore flamma.  
 Imminet his aér. qui quanto est pondere terræ,  
 Pondere aquæ leuior ; tanto est onerosior igne.  
 Illic & nebulas, illuc confistere nubes  
 Iussit, & humanas motura tonitrua mentes,  
 Et cum fulminibus facientes frigora ventos.

Εικόνα 2.30: Χρωματική απεικόνιση των συνδεδεμένων αντικειμένων μαζί με τα ορθογώνια παραλληλόγραμμα που τα περικλείουν.

- 3) Επόμενο βήμα της μεθόδου είναι η απομάκρυνση των αντικειμένων θορύβου, ώστε να βελτιωθεί το αποτέλεσμα της κατάτμησης. Για κάθε αντικείμενο  $CC_i$  εξάγονται δύο νέα χαρακτηριστικά:

- Η πυκνότητα:

$$\text{Density} = \frac{P_{total}}{H \cdot W} \quad (2.45)$$

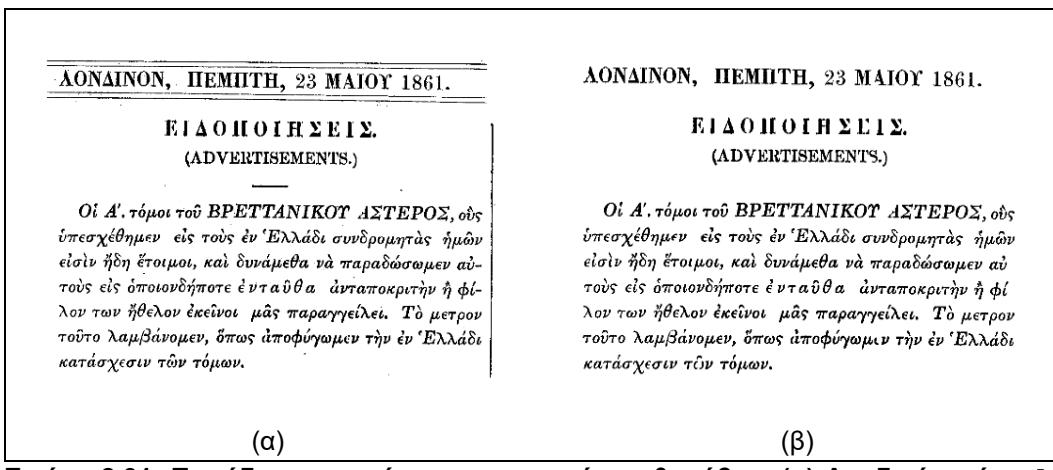
όπου

- $P_{total}$ : ο συνολικός αριθμός εικονοστοιχείων προσκηνίου του αντικειμένου  $CC_i \in I_o$ .
- $H \cdot W$ : το εμβαδόν του ορθογωνίου πλαισίου του αντικειμένου  $CC_i$  με  $H, W$  το ύψος και το πλάτος αντίστοιχα.

- Η επιμήκυνση:

$$\text{Elongation} = \frac{\min(H, W)}{\max(H, W)} \quad (2.46)$$

Όσα αντικείμενα έχουν  $\text{Density} \leq 0.05$  ή  $\text{Density} \geq 0.9$  ή  $\text{Elongation} \leq 0.08$  θεωρούνται θόρυβος και διαγράφονται. Οι περιοχές προσκηνίου που απορρίφθηκαν, μετατρέπονται σε υπόβαθρο και προκύπτει η εικόνα  $I_1$ .



Εικόνα 2.31: Παράδειγμα αφαίρεσης αντικειμένων θορύβου. (α) Δυαδική εικόνα  $I_o$ , (β) εικόνα  $I_1$  μετά το φιλτράρισμα του θορύβου.

- 4) Ακολουθεί ένα ακόμα στάδιο φιλτραρίσματος, όπου αφαιρούνται τα σημεία στίξης. Εκτελείται ο ARSLA στην εικόνα  $I_1$  με  $\alpha = 1.5, T_h = 3.5, c = 0.7$  και μετά την εφαρμογή του προκύπτει η εικόνα  $I_2$ , από την οποία εξάγονται τα συνδεδεμένα αντικείμενα. Η διαδικασία βασίζεται στη σύγκριση των αντικειμένων των εικόνων  $I_1$  και  $I_2$ .

Αν ισχύει η σχέση:

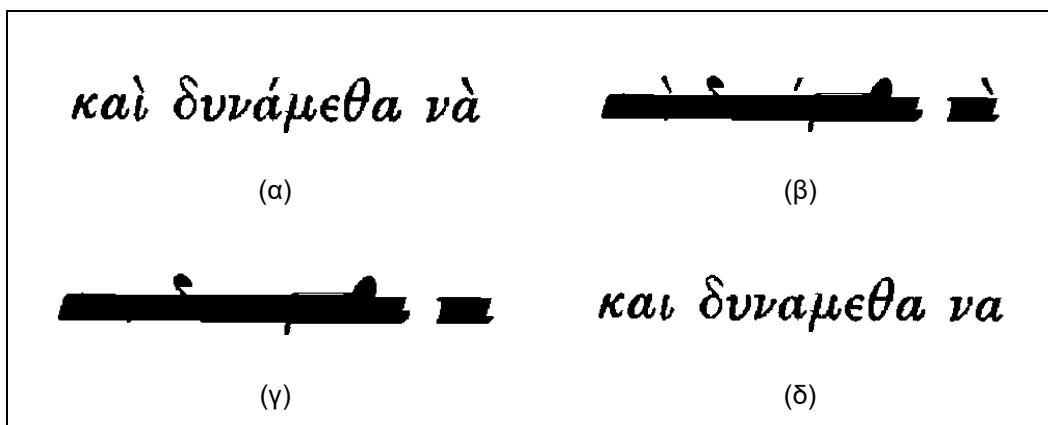
$$\frac{P_2}{P_1} \leq T_R \quad (2.47)$$

τότε διαγράφονται όσα αντικείμενα αντιστοιχούν σε λόγο μικρότερο από  $T_R$  και τα εικονοστοιχεία τους αλλάζουν σε υπόβαθρο, σχηματίζοντας την εικόνα  $I_3$ , όπου

- $P_2$ : το πλήθος των εικονοστοιχείων (προσκηνίου) του αντικειμένου  $CC_i$  της εικόνας  $I_2$ ,  $CC_i \in I_2$ .
- $P_1$ : το πλήθος των εικονοστοιχείων (προσκηνίου) της εικόνας  $I_1$  μέσα στη περιοχή που ορίζει το αντικείμενο  $CC_i \in I_2$ .
- $T_R$ : ένα σταθερό κατώφλι με τιμή 1.15.

Με εφαρμογή του τελεστή AND μεταξύ των εικόνων  $I_1$  και  $I_3$  προκύπτει η εικόνα  $I_4$ .

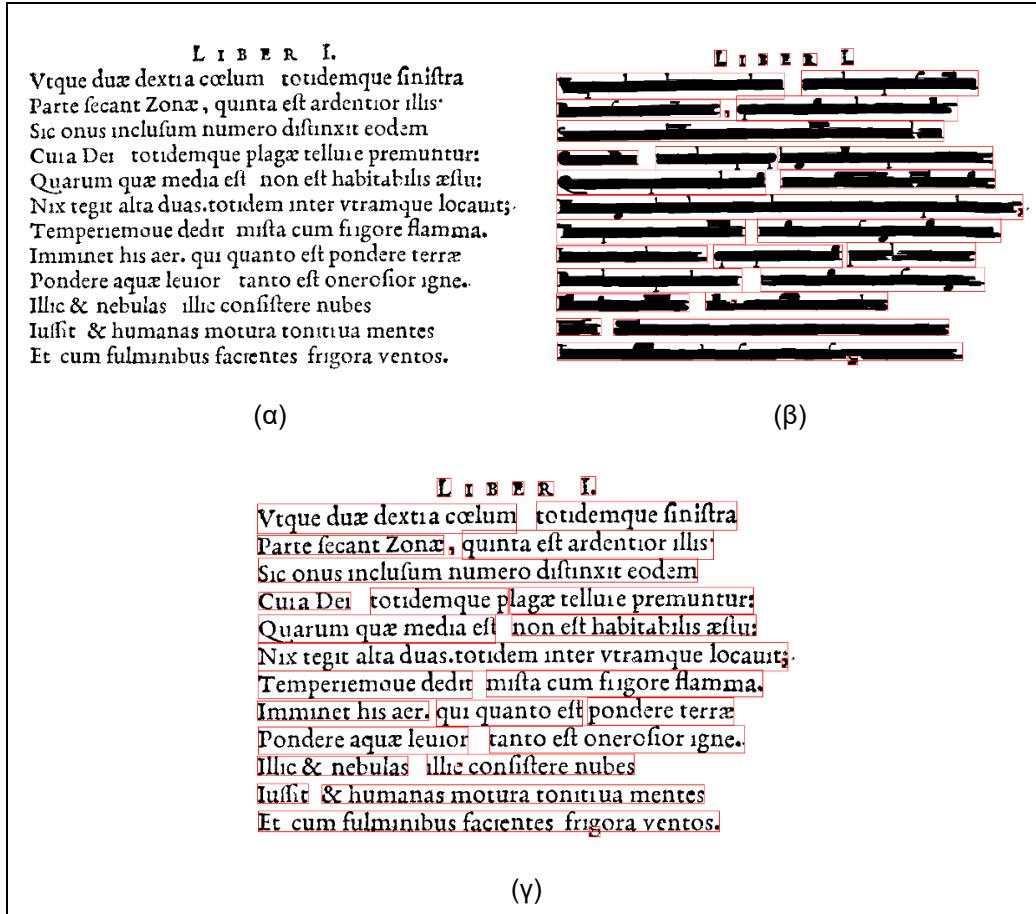
Η τιμή κατωφλιού  $T_R = 1.15$  εκφράζει τη πιθανότητα, το μέγεθος των αντικειμένων της εικόνας  $I_2$  (άρα και το πλήθος των εικονοστοιχείων  $P_2$ ) να είναι 15% μεγαλύτερο από το μέγεθος των αντικειμένων της εικόνας  $I_1$ . Έτσι, μετά την εφαρμογή του ARLSA, τα αντικείμενα κειμένου θα ομαδοποιηθούν μεταξύ τους, σχηματίζοντας αντικείμενα πολύ μεγαλύτερου μεγέθους σε σχέση με τα αρχικά ( $\frac{P_2}{P_1} \gg 1.15$ ). Τα μικρά απομονωμένα αντικείμενα όμως (σημεία στίξης συνήθως) ενδέχεται να παραμείνουν ίδια ή να μεγαλώσουν ελάχιστα. Συνεπώς θα ισχύει  $\frac{P_2}{P_1} \leq 1.15$  και θα αφαιρεθούν.



Εικόνα 2.32: Διαδικασία αφαίρεσης σημείων στίξης. (α) Εικόνα  $I_1$ , (β) εικόνα  $I_2$  μετά την εφαρμογή του ARLSA, (γ) εικόνα  $I_3$ , (δ) τελική εικόνα  $I_4$ .

- 5) Έπειτα, εκτελείται ο ARLSA ( $\alpha = 1$ ,  $T_h = 3.5$ ,  $c = 0.7$ ) στην εικόνα  $I_4$  και το αποτέλεσμά του δίνει την εικόνα  $I_5$ . Σκοπός της εφαρμογής του αλγορίθμου είναι να συνδεθούν αντικείμενα κειμένου μεταξύ τους και να

σχηματίσουν ομάδες λέξεων. Μετά εξάγονται τα συνδεδεμένα αντικείμενα των εικόνων  $I_4$  και  $I_5$  και τα αντικείμενα της εικόνας  $I_4$  ομαδοποιούνται βάση των διαστάσεων των ορθογωνίων πλαισίων που περικλείουν τα αντικείμενα της εικόνας  $I_5$ . Ακολούθως διαγράφονται τα αντικείμενα με πλάτος  $W \leq 10$ .



Εικόνα 2.33: Παράδειγμα εφαρμογής του ARLSA. (α) Εικόνα  $I_4$ , (β) εικόνα  $I_5$  - αποτέλεσμα ομαδοποίησης λέξεων με ARLSA ( $\alpha = 1$ ,  $T_h = 3.5$ ,  $c = 0.7$ ), (γ) ομαδοποιημένα αντικείμενα κειμένου της εικόνας  $I_4$ .

- 6) Τα τμήματα των λέξεων εξάγονται με την επεξεργασία κάθε ομαδοποιημένου αντικειμένου  $CC_{group} \in I_4$  που προέκυψε από την εφαρμογή του ARLSA στο προηγούμενο βήμα. Πρώτα ορίζεται η ποσότητα:

$$M_d = 15 \cdot \overline{AW} \quad (2.48)$$

όπου  $\overline{AW}$  το μέσο πλάτος των αντικειμένων της εικόνας  $I_o$  με πλάτος  $W \leq 20$ . Για κάθε  $CC_{group} \in I_4$  με πλάτος  $W < M_d$ , τα εικονοστοιχεία της εικόνας  $I_4$  αντικαθίστανται με τα αντίστοιχα εικονοστοιχεία της εικόνας  $I_5$  (αποτέλεσμα ARLSA). Για τα υπόλοιπα αντικείμενα με  $W \geq M_d$  υπολογίζεται η κάθετη προβολή των εικονοστοιχείων (αριθμός pixel κειμένου της κάθετης διεύθυνσης) και εξάγεται το ιστόγραμμα  $H_v(w)$  των

ακολουθιών λευκών εικονοστοιχείων (υποβάθρου) της κάθετης διεύθυνσης.

Με επεξεργασία του ιστογράμματος με τη μέθοδο Otsu [1] υπολογίζεται το κατώφλι:

$$T_{thr} = T_{otsu} + 1 \quad (2.49)$$

για τον εντοπισμό των αποστάσεων διαχωρισμού μεταξύ των λέξεων και των γραμμάτων, όπου  $T_{otsu}$  το κατώφλι που προέκυψε από τη τεχνική Otsu. Με βάση το κατώφλι  $T_{thr}$ , το ιστόγραμμα  $H_v(w)$  χωρίζεται σε δύο περιοχές:

$$H_v(w) = \begin{cases} H_{vl}(w), & 0 \leq w < T_{thr} \\ H_{vw}(w), & w \geq T_{thr} \end{cases} \quad (2.50)$$

Επίσης, ορίζονται δύο νέοι παράμετροι:

- Το περιθώριο μεταξύ γραμμάτων (letterspacing):

$$l_{sp} = argmax(H_{vl}(w)) \quad (2.51)$$

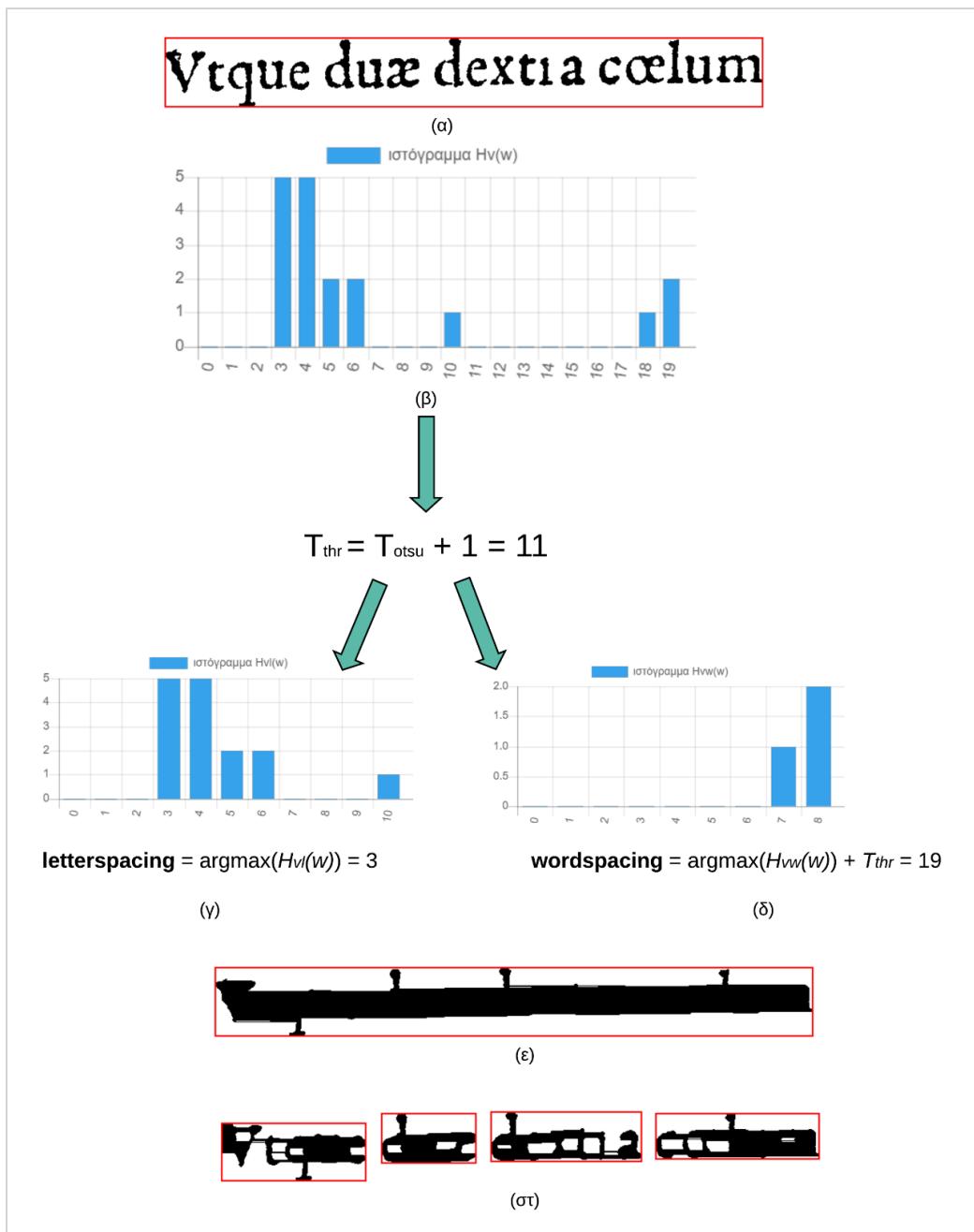
- Το περιθώριο μεταξύ λέξεων (wordspacing):

$$w_{sp} = argmax(H_{vw}(w)) + T_{thr} \quad (2.52)$$

Συνάρτηση των παραπάνω εκφράζεται το μέγεθος:

$$S_{sp} = \begin{cases} W, & (w_{sp} - l_{sp}) < l_{sp} \\ ceil\left(\frac{w_{sp} + T_{thr}}{2}\right), & διαφορετικά \end{cases} \quad (2.53)$$

Έπειτα εφαρμόζεται εξομάλυνση κατά την οριζόντια κατεύθυνση. Υπολογίζεται το οριζόντιο μήκος ακολουθιών εικονοστοιχείων υποβάθρου μεταξύ των εικονοστοιχείων προσκηνίου. Αν το μήκος είναι μικρότερο ή ίσο της παραμέτρου  $S_{sp}$  τότε τα εικονοστοιχεία της ακολουθίας αντικαθίστανται με μαύρα εικονοστοιχεία προσκηνίου. Έτσι, η εικόνα  $I_4$  μετατρέπεται στην εξομαλυμένη εικόνα  $I_6$ , από την οποία εξάγονται τα συνδεδεμένα αντικείμενα και απορρίπτονται όσα έχουν  $W \leq 5$  και  $H \leq 5$ .



**Εικόνα 2.34:** Παράδειγμα τελικής εξομάλυνσης. (α) Ομαδοποιημένο αντικείμενο της εικόνας  $I_4$ , (β) ιστόγραμμα  $H_v(w)$ , (γ) εξαγωγή απόστασης γραμμάτων από το ιστόγραμμα  $H_{vl}(w)$ , (δ) εξαγωγή απόστασης λέξεων από το ιστόγραμμα  $H_{vw}(w)$ , (ε) εικόνα  $I_5$ , (στ) εξομαλυμένη εικόνα  $I_6$ .

- 7) Τέλος, με βάση τις διαστάσεις των ορθογωνίων πλαισίων των τελικών αντικειμένων της εικόνας  $I_6$ , ομαδοποιούνται τα αντικείμενα της αρχικής εικόνας  $I_0$  και ολοκληρώνεται η διαδικασία κατάτμησης λέξεων του εγγράφου.

## LIBER I.

Vtque dux dextra cœlum , totidemque sinistra  
Parte secant Zonæ , quinta est ardenter illis;  
Sic onus inclusum numero distinxit eodem  
Cura Dei , totidemque plaga tellure premuntur;  
Quarum qua media est , non est habitabilis æstu:  
Nix tegit alta duas,totidem inter utramque locauit;  
Temperiemque dedit, mista cum frigore flamma.  
Imminet his aëri qui quanto est pondere terra,  
Pondere aquæ leuior ; tanto est onerosior igne.  
Ilic & nebulas , illuc confistere nubes  
Iusfit , & humanas motura tonitrua mentes,  
Et cum fulminibus facientes frigora ventos.

(α)

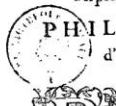
## LIBER II.

Vtque dux dextra cœlum , totidemque sinistra  
Parte secant Zonæ , quinta est ardenter illis;  
Sic onus inclusum numero distinxit eodem  
Cura Dei , totidemque plaga tellure premuntur;  
Quarum qua media est , non est habitabilis æstu:  
Nix tegit alta duas,totidem inter utramque locauit;  
Temperiemque dedit, mista cum frigore flamma.  
Imminet his aëri qui quanto est pondere terra,  
Pondere aquæ leuior ; tanto est onerosior igne.  
Ilic & nebulas , illuc confistere nubes  
Iusfit , & humanas motura tonitrua mentes,  
Et cum fulminibus facientes frigora ventos.

(β)

L E.  
**M I R O I R**  
DE LA  
Tyrannie Espagnole

Perpetrée au Pays-bas , soubs

PHILIPPE II. ROY  
d'ESPAGNE.

Lesseurs Escrivains , & Autheurs , nous font scavoir , par leur memoires , de la infatiable convoitise d'honneur , en infupable arrogance , & incetible cruauter de la Nation Espagnole , commise par tout le Monde depuis cent & cinquante ans , ou elle a eu le principal gouvernement , par ces effets elle a reçeu le bel et renommee qu'elle a été toujours estimée la plus cruelle du monde .

Premièrement les Espagnols ont montré leur infatiable Les Espag cruauter envers les innocens Indiens , etians encore Payens , tannite les quand ilz sont arriviez en leur pays , pour les instruire en la Indiens viaye .

(γ)

L E.  
**M I R O I R**

DE LA

Tyrannie Espagnole

Perpetrée au Pays-bas , soubs



Lesseurs Escrivains , & Autheurs , nous font scavoir , par leur memoires , de la infatiable convoitise d'honneur , en infupable arrogance , & incetible cruauter de la Nation Espagnole , commise par tout le Monde depuis cent & cinquante ans ou elle a eu le principal gouvernement , par ces effets elle a reçeu le bel et renommee qu'elle a été toujours estimée la plus cruelle du monde .

Premièrement les Espagnols ont montré leur infatiable Les Espag cruauter envers les innocens Indiens , etians encore Payens , quand ilz sont arriviez en leur pays , pour les instruire en la Indiens viaye .

(δ)

ΑΟΝΔΙΝΟΝ , ΠΕΜΠΤΗ , 23 ΜΑΪΟΥ 1861.

ΕΙΔΟΗ ΗΣΕΙΣ.  
(ADVERTISEMENTS.)

Oi A' , τόμοι τοῦ BΡΕΤΤΑΝΙΚΟΤ ΑΣΤΕΡΟΣ , οὓς ὑπερσχέθημεν εἰς τοὺς ἐν Ἑλλάδι συνδρομητὰς ἡμῶν εἰσὶν ἥδη ἔτοιμοι , καὶ δυνάμεθα παραδώσωμεν αὐτοὺς εἰς ὄποιονδήποτε εὐταῖθα ἀνταποκριτὴν ἢ φίλον τῶν ἥθελοι ἔκεινοι μᾶς παραγγείλει . Τὸ μετρον τοῦτο λαμβάνομεν , δπως ἀποφύγωμεν τὴν ἐν Ἑλλάδι κατάσχεσιν τῶν τόμων .

(ε)

## ΕΙΔΟΗ ΗΣΕΙΣ

(ADVERTISEMENTS.)

Οι A' , τόμοι τοῦ BΡΕΤΤΑΝΙΚΟΤ ΑΣΤΕΡΟΣ , οὓς ὑπερσχέθημεν εἰς τοὺς ἐν Ἑλλάδι συνδρομητὰς ἡμῶν εἰσὶν ἥδη ἔτοιμοι , καὶ δυνάμεθα παραδώσωμεν αὐτοὺς εἰς ὄποιονδήποτε εὐταῖθα ἀνταποκριτὴν ἢ φίλον τῶν ἥθελοι ἔκεινοι μᾶς παραγγείλει . Τὸ μετρον τοῦτο λαμβάνομεν , δπως ἀποφύγωμεν τὴν ἐν Ἑλλάδι κατάσχεσιν τῶν τόμων .

(στ)

» μυήσομ<sup>ε</sup> ἐταίρου. Ομήρω<sup>α</sup> μὲν ἐποιήθησαν οἱ σῆμαι<sup>ο</sup> δένοες αὐτ<sup>ο</sup>, οὐκ οἶδα εἰ Αχιλλ<sup>ο</sup>ς τοι<sup>ς</sup> Γαρύκλας μᾶλλον ἀξιος εἰρηθα<sup>ν</sup>, ή ἔμοι τοι<sup>ς</sup> σομ<sup>της</sup> φιλ τάτης τε καὶ διεργέτιδος κεφαλῆς. ὡς ἐγώ μάρτυρα ποιοῦμαγ<sup>η</sup> θεὸν, ὃν φιλοσοφία φερεσθέλ<sup>η</sup>, τῆς ιερᾶς σ<sup>ο</sup> <sup>της</sup> Γλυκείας Φυχῆς ὅπει μέσος καρδίας ἄγαλ μα τοι<sup>ς</sup> φέρω, καὶ ἐμβομβέδ<sup>η</sup> μων ταῦς ἀκοαῆς ή θαυμαση<sup>η</sup> σεν τῷ Κρανι<sup>α</sup> γένων ήχω. τῇ πατείδ<sup>η</sup> δε<sup>τη</sup> διδυμήσας στότης Αἰγυπτίου, <sup>της</sup> Δυοῖν στιαν-

(ζ)

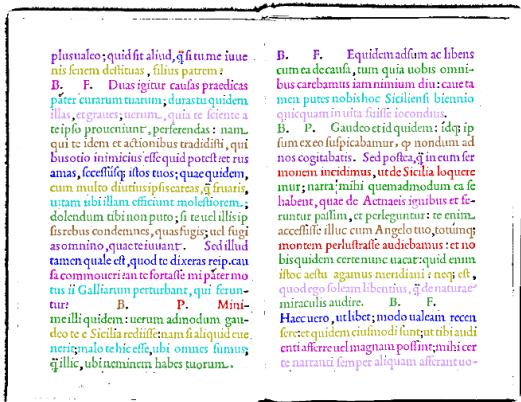
» μυήσομ<sup>ε</sup> ἐταίρου. Ομήρω<sup>α</sup> μὲν ἐποιήθησαν οἱ σῆμαι<sup>ο</sup> δένοες αὐτ<sup>ο</sup>, οὐκ οἶδα εἰ Αχιλλ<sup>ο</sup>ς τοι<sup>ς</sup> Γαρύκλας μᾶλλον ἀξιος εἰρηθα<sup>ν</sup>, ή ἔμοι τοι<sup>ς</sup> σομ<sup>της</sup> φιλ τάτης τε καὶ διεργέτιδος κεφαλῆς. ὡς ἐγώ μάρτυρα ποιοῦμαγ<sup>η</sup> θεὸν, ὃν φιλοσοφία φερεσθέλ<sup>η</sup>, τῆς ιερᾶς σ<sup>ο</sup> <sup>της</sup> Γλυκείας Φυχῆς ὅπει μέσος καρδίας ἄγαλ μα τοι<sup>ς</sup> φέρω, καὶ ἐμβομβέδ<sup>η</sup> μων ταῦς ἀκοαῆς ή θαυμαση<sup>η</sup> σεν τῷ Κρανι<sup>α</sup> γένων ήχω. τῇ πατείδ<sup>η</sup> δε<sup>τη</sup> διδυμήσας στότης Αἰγυπτίου, <sup>της</sup> Δυοῖν στιαν-

(η)

Εικόνα 2.35: Χρωματική απεικόνιση της τελικής τμηματοποίησης λέξεων. (α)-(γ)-(ε)-(ζ) Αρχικές εικόνες, (β)-(δ)-(στ)-(η) αποτέλεσμα κατάτμησης λέξεων.

Αξίζει να σημειωθεί πως το βήμα 6 αποτελεί επιπλέον στάδιο εξομάλυνσης της εικόνας και τμηματοποίησης των λέξεων μετά την εφαρμογή του ARLSA. Εναλλακτικά, στο βήμα 5 μπορεί κάλλιστα να εφαρμοστεί ο ARLSA με  $\alpha < 1$ , ώστε αντί για ομάδες λέξεων να ομαδοποιηθούν μόνο τα αντικείμενα που ανήκουν στην ίδια λέξη. Έτσι, προκύπτουν οι τελικές περιοχές των λέξεων και η διαδικασία της κατάτμησης ολοκληρώνεται με παρόμοιο τρόπο όπως στο βήμα 7.

Επίσης, μολονότι η μεθοδολογία που αναλύθηκε παραπάνω εστιάζει κυρίως στην κατάτμηση λέξεων, υπάρχει δυνατότητα εξαγωγής των γραμμών κειμένου εφαρμόζοντας τον ARLSA στην εικόνα  $I_4$ , η οποία προέκυψε από το στάδιο αφαίρεσης θορύβου και σημείων στίξης. Ο αλγόριθμος εκτελείται με τιμή της παραμέτρου  $\alpha$  ίσης ή μεγαλύτερης του 5, τιμή που είναι σχετικά μεγάλη, ώστε να ενωθούν γειτονικά αντικείμενα της ίδιας γραμμής κειμένου με μεγάλη απόσταση μεταξύ τους.



(α)

L I B E R . I  
Vtque duæ dextra ccelum, totidemque sinistra:  
Parte secant Zonæ, quinta est ardentior illis;  
Sic onus inclusum numero distinxit eodem  
Cura Dei, totidemque plagæ tellure premuntur:  
Quarum quæ media est, non est habitabilis æstu:  
Nix tegit alta duas. totidem inter vtramque locavit;  
Temperiemque dedit, mixta cum frigore flamma.  
Imminet his aér. qui quanto est pondere terræ,  
Pondere aquæ leuior; tanto est onerosior igne.  
Illic & nebulas, illuc confistere nubes  
Iussit, & humanas motura tonitrua mentes,  
Et cum fulminibus facientes frigora ventos.

(β)

## LA VRAYE ENARRATION De la destruction des INDES OCCIDENTALES:

*Le Chapitre Premier*

**Uand les INDES OCCIDENTALES estoient decouverts au l'an 1492.** l'ile Eg  
l'annee suivante foye venue la les contient en  
**Chrestiens pour instruire, former, & po  
puler les terres; a l'çavoir les Espagnols**  
ainsi sont passez 49 ans, que y sont ar  
rivez les Espagnols en grand nombre,  
& ils sont abbordez au l'ile Espagnio  
la grande, & fort heureuse, contenant en rondeur 600 lieux;  
il y a plusieurs autres grandes & riches, lesquels nous aper  
çumes de loing, éstaits fort habitez de ḡs naturels, Indes.  
On decouvre aujord'huy le pays ferme, & on a decouvert  
plus que dix milieux de terre, distante de la plus que 250  
lieux, pleine de gens, comme une ruche à miel, pleine de  
mouches à miel: Le decouvertement dernier a esté au l'an  
1531: il semble que Dieu a respandu par ce terres une bende,  
cu tas des hommes, des toutes les raçons innumerables.  
Dieu a crée ces hommes innumerables, fort simples, sans Les nature  
finesse, ou doubletelle, fort obediens, fidels au Princes naturels, des Ind  
& aux Chrestiens, lesquels ils servent fort humbles, patients,  
pauvres, & modestes, sans fanfer, troubler, & redire, non moi  
les, mais Dieu.

Princeton, Sept. 18. 1825

Christian Brothers,

It is known probably to most of you, that several of our  
acquaintances and friends, in coming from the State of Maine to  
their Town, have been exposed to the most imminent danger, and  
also sustained a considerable loss of property. As their dangers and  
misfortunes have excited much sympathy, and as it is believed to  
be a Christian duty to do something for their relief, it may be proper,  
for the general information of all, to make a short statement of  
the peril to which they were exposed, and the extent of their loss.  
On Saturday the 10th inst, at Rockport, Meas. How, Gay and  
Cobb with the wives of the two last, and eight other persons, sailed  
from Portland for Boston in the Stage Governor. Their passage was  
to end at Boston in the Stage.

(γ)

(δ)

Εικόνα 2.36: Χρωματική απεικόνιση τμηματοποίησης λέξεων κειμένου. (α)-(β)-(γ)  
Ικανοποιητική εξαγωγή γραμμών, (δ) η μικρή απόσταση μεταξύ των γραμμών οδηγεί σε  
λανθασμένη ομαδοποίηση τους.

### 2.7.5 Αξιολόγηση της τεχνικής

Για την αξιολόγηση της μεθόδου υπολογίστηκαν τα μετρικά μεγέθη **Recall** (Ανάκληση) και **Precision** (Ακρίβεια), τα οποία χρησιμοποιούνται ευρέως για την αξιολόγηση συστημάτων ανάκτησης της πληροφορίας. Ορίζονται ως:

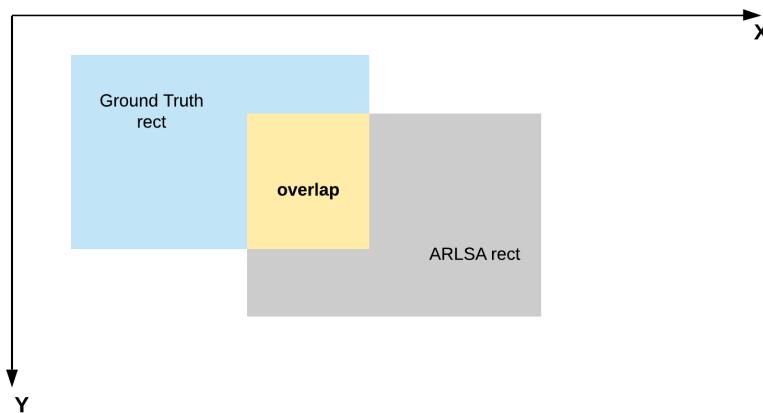
$$Recall = \frac{T_p}{N_{GT}} \quad (2.54)$$

$$Precision = \frac{T_p}{N_{ARLSA}} \quad (2.55)$$

όπου

- $T_p$ : τα σωστά overlaps.
- $N_{GT}$ : ο αριθμός των Rectangles (ορθογώνια πλαίσια γύρω από τις κατατμημένες λέξεις) του Ground Truth.
- $N_{ARLSA}$ : ο αριθμός των Rectangles που έχουν εξαχθεί από την υλοποιημένη μέθοδο ARLSA.

Το overlap μεταξύ δύο Rectangle ορίζεται ως το εμβαδόν της περιοχής που έχουν επικάλυψη.



**Εικόνα 2.37: Γραφική απεικόνιση επικάλυψης μεταξύ δύο Rectangles.**

Έστω  $A$  το εμβαδόν ενός Rectangle από το Ground Truth και  $B$  το εμβαδόν ενός Rectangle από την υλοποιημένη τεχνική. Το overlap θεωρείται σωστό όταν ξεπερνάει ένα κατώφλι:

$$T_{overlap} = \frac{intersection(overlap)}{union} = \frac{A \cap B}{A \cup B} \quad (2.56)$$

Στις περισσότερες περιπτώσεις θεωρούμε σωστό overlap όταν  $T_{overlap} \geq 0.5$ .

Όπως φαίνεται από τις παραπάνω εξισώσεις, η τέλεια ανάκληση (Recall) είναι ο αριθμός 1.0 και δείχνει ότι όλα τα σχετικά-συναφή Rectangles έχουν κατατμηθεί. Αυτό όμως δεν λέει πόσα λανθασμένα μη σχετικά Rectangles έχουν επίσης εξαχθεί. Αυτό φαίνεται από την ακρίβεια (Precision), η οποία στην τέλεια της κατάσταση έχει την τιμή 1.0 και δηλώνει ότι κάθε Rectangle που έχει βρεθεί είναι συναφής, δεν εκφράζει όμως πόσα συναφή Rectangles δεν έχουν

βρεθεί. Συνεπώς, τα δύο αυτά μεγέθη είναι αντιστρόφως ανάλογα αλλά χρειάζονται και τα δύο για να προκύψει μια ολοκληρωμένη αξιολόγηση των αποτελεσμάτων της τμηματοποίησης λέξεων.

Αρχικά, η υλοποιημένη μέθοδος κατάτμησης ARLSA [6] δοκιμάστηκε σε 20 σύγχρονες χειρόγραφες εικόνες εγγράφου του διαγωνισμού H-KWS 2014 [7]. Πρόκειται για έγγραφα με ελάχιστα ποσοστά θορύβου, τα οποία περιέχουν κείμενο διαφορετικού γραφικού χαρακτήρα ανά εικόνα. Για το σύνολο των εγγράφων ορίστηκε ένα μεγάλο κατώφλι  $T_{overlap} \geq 0.8$ . Τα αποτελέσματα φαίνονται στο παρακάτω πίνακα.

	$T_{overlap} \geq 0.8$	
εικόνα	Recall %	Precision %
1	98	98
2	98	93
3	82	84
4	95	96
5	96	96
6	98	99
7	94	93
8	82	83
9	85	85
10	94	94
11	88	84
12	84	83
13	76	72
14	80	84
15	90	88
16	92	92
17	81	78
18	82	80
19	66	67
20	76	81
μέσος όρος	86,85	86,5

Πίνακας 2.4: Αποτελέσματα αξιολόγησης σε σύγχρονα χειρόγραφα έγγραφα.

Παρατηρείται πως η μέθοδος αποδίδει πολύ καλά σε όλες τις εικόνες και καταφέρνει να εντοπίσει και να εξάγει τις λέξεις του κειμένου με μεγάλη ακρίβεια. Φυσικά, με μεγαλύτερη ανοχή επικάλυψης (μικρότερο  $T_{overlap}$ ) η τεχνική θα επιτυγχάνει ακόμα καλύτερα αποτελέσματα.

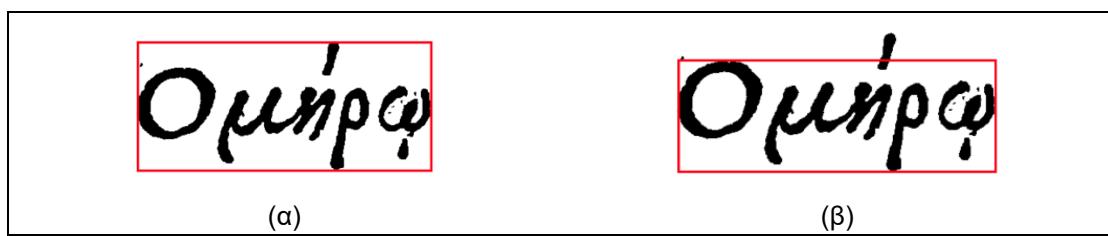
Έπειτα, η μέθοδος εξετάστηκε σε ένα σύνολο 10 ιστορικών εικόνων εγγράφου. Το σύνολο αυτό περιέχει τόσο τυπωμένα όσο και χειρόγραφα έγγραφα, τα οποία εμφανίζουν μεταβολές μεγέθους και απόστασης μεταξύ των γραμμών, λέξεων και χαρακτήρων του κειμένου, λοξό κείμενο, καλλιτεχνική γραφή κ.α.

Για τις εικόνες αυτές και τις προηγούμενες, δημιουργήθηκαν τα αντίστοιχα Ground Truth τους χειροκίνητα με τη χρήση γραφικού εργαλείου, το οποίο θα παρουσιαστεί στο επόμενο κεφάλαιο. Τα αποτελέσματα της αξιολόγησης αποτυπώνονται στον παρακάτω πίνακα.

	$T_{overlap} \geq 0.3$		$T_{overlap} \geq 0.5$		$T_{overlap} \geq 0.8$	
εικόνα	Recall	Precision	Recall	Precision	Recall	Precision
1	95	92	86	84	52	50
2	96	83	89	78	66	57
3	91	62	73	50	28	20
4	97	94	96	93	84	81
5	98	70	96	68	39	28
6	86	81	77	72	51	48
7	87	85	78	76	53	51
8	82	86	76	80	63	66
9	91	93	80	81	58	60
10	86	83	80	77	70	68
μέσος όρος	90,9	82,9	83,1	75,9	56,4	52,9

Πίνακας 2.5: Αποτελέσματα αξιολόγησης σε ιστορικά έγγραφα.

Παρατηρούνται αρκετά ικανοποιητικά αποτελέσματα για  $T_{overlap} \geq 0.5$ . Όπως αναμένεται όμως, με μεγαλύτερο κατώφλι  $T_{overlap} \geq 0.8$  τα ποσοστά Recall και Precision χειροτερεύουν. Στις περισσότερες εικόνες ο αλγόριθμος εντοπίζει της λέξεις αλλά, λόγω των σημείων στίξης που δεν φιλτραρίστηκαν, προκύπτουν παραπάνω αντικείμενα (Rectangles), τα οποία μειώνουν το Precision (Ακρίβεια). Επίσης, στην υλοποιημένη προσέγγιση ARLSA της διπλωματικής, τα σημεία τονισμού που αφαιρέθηκαν δεν συνυπολογίζονται στην τελική εξαγωγή των λέξεων. Έτσι, παρόλο που το σώμα των λέξεων (χωρίς τόνους) εντοπίζεται σωστά, προκύπτει λανθασμένη επικάλυψη με τη Ground Truth λέξη κατά την αξιολόγηση, σε περιπτώσεις μεγάλου κατωφλιού. Το πρόβλημα αυτό φαίνεται στη παρακάτω εικόνα.



Εικόνα 2.38: Αδυναμία σωστής κατάτμησης ολόκληρης της λέξης λόγω αφαίρεσης των σημείων τονισμού. (α) Ground Truth λέξη, (β) ARLSA λέξη.

Συγκεντρωτικά σε αυτό το σύνολο εγγράφων, συμπεραίνεται πως ο αλγόριθμος αποδίδει καλύτερα σε τυπωμένα (machine printed) ιστορικά έγγραφα παρά σε ιστορικές χειρόγραφες εικόνες κειμένου.

Τέλος, εξετάστηκαν 10 εικόνες εγγράφου από τη συλλογή “Konzilsprotokolle (1795)” [8]. Τα αποτελέσματα της αξιολόγησης παρουσιάζονται στον παρακάτω πίνακα.

εικόνα	$T_{overlap} \geq 0.3$		$T_{overlap} \geq 0.5$	
	Recall	Precision	Recall	Precision
1	79	72	55	54
2	71	71	53	53
3	73	70	53	55
4	73	75	46	48
5	65	68	47	49
6	77	70	53	48
7	80	76	58	55
8	72	72	48	48
9	64	73	46	53
10	76	75	49	49
μέσος όρος	73	72,2	50,8	51,2

Πίνακας 2.6: Αποτελέσματα αξιολόγησης σε ιστορικά έγγραφα της συλλογής “Konzilsprotokolle (1795)”.

Σε αυτή τη βάση εγγράφων παρατηρούνται μη αξιόλογα αποτελέσματα. Αυτό οφείλεται στο γεγονός ότι σε όλες τις εικόνες της συλλογής, οι λέξεις περιέχουν καλλιγραφικούς πλάγιους χαρακτήρες με μεγάλη κλίση, οι οποίοι εμφανίζουν μεταβλητή διαφορά ύψους μεταξύ τους. Έτσι, χαρακτήρες της ίδιας λέξεις δεν καταφέρνουν να ομαδοποιηθούν αποτελεσματικά μεταξύ τους και να σχηματίσουν την σωστή λέξη. Ακόμα, εξαιτίας του μεγάλου ύψους των χαρακτήρων, υπάρχει πολύ μικρό περιθώριο μεταξύ των γραμμών του κειμένου, με αποτέλεσμα σε ορισμένες περιπτώσεις οι λέξεις διαφορετικής γραμμής να μπλέκονται μεταξύ τους.

Συνολικά, η υλοποιημένη μέθοδος κατάτμησης βασιζόμενη στον ARSLA [6], ανάλογα με το είδος του εγγράφου, μπορεί να επιτύχει αξιόλογη εξαγωγή των λέξεων του κειμένου. Μεγάλο της πλεονέκτημα αποτελεί η σωστή σύνδεση όμοιων αντικειμένων, ώστε να σχηματιστούν οι λέξεις σε έγγραφα με μεταβλητή απόσταση μεταξύ των γραμμών κειμένου, λέξεων και χαρακτήρων. Βέβαια για επιτυχημένη εφαρμογή του αλγορίθμου, απαιτείται η χειροκίνητη ρύθμιση των παραμέτρων του ανάλογα με τις διαστάσεις των αντικειμένων.

### 3. Ανάλυση της προγραμματιστικής υλοποίησης και παρουσίαση της εφαρμογής

#### 3.1 Γλώσσες Προγραμματισμού που χρησιμοποιήθηκαν

Οι γλώσσες προγραμματισμού με τις οποίες υλοποιήθηκε η εφαρμογή διαδικτύου της διπλωματικής είναι οι παρακάτω:

- **HTML/CSS**

Η HTML (Hypertext Markup Language – Γλώσσα Σήμανσης Υπερκειμένου) και η CSS (Cascading Style Sheets – Διαδοχικά Φύλλα Ύφους) είναι δύο βασικές τεχνολογίες για την δημιουργία εφαρμογών διαδικτύου και ιστοσελίδων. Η HTML αποτελεί τη κύρια γλώσσα σήμανσης μέσω της οποίας περιγράφονται τα δομικά στοιχεία της διεπαφής χρήστη, όπως το κείμενο, οι εικόνες, οι φόρμες εισαγωγής και όλα τα άλλα γραφικά στοιχεία που περιέχονται σε μια ιστοσελίδα. Η CSS ανήκει στην κατηγορία των γλωσσών φύλλων ύφους (μορφοποίησης) και χρησιμοποιείται για τον έλεγχο της εμφάνισης και διάταξης των διάφορων δομικών στοιχείων στης HTML ιστοσελίδας.

- **JavaScript/TypeScript**

Η JavaScript(JS) είναι μια διερμηνευμένη γλώσσα προγραμματισμού για ηλεκτρονικούς υπολογιστές. Αποτελεί, μαζί με την HTML και την CSS, μία από τις βασικές τεχνολογίες του Διαδικτύου. Ενσωματώνεται στους περισσότερους φυλλομετρητές Ιστού (web Browsers) μέσω των οποίων εκτελείται τοπικά στον υπολογιστή του χρήστη (client side), επιτρέποντας την ασύγχρονη ανταλλαγή δεδομένων και την δημιουργία διαδραστικών εφαρμογών δυναμικού περιεχόμενου. Υποστηρίζει αντικειμενοστραφή, βασιζόμενο σε γεγονότα (event driven), προστακτικό και συναρτησιακό στυλ προγραμματισμού. Η TypeScript(TS), με την οποία υλοποιήθηκε το μεγαλύτερο κομμάτι της διπλωματικής, είναι μια γλώσσα ανοιχτού λογισμικού που δημιουργήθηκε και συντηρείται από τη Microsoft. Πρόκειται ουσιαστικά για ένα υπερσύνολο της JavaScript (ECMAScript 6 – ES6) με αυστηρή και στατική σύνταξη, το οποίο μεταγλωττίζεται σε JavaScript κώδικα. Προσθέτει επιπλέον χαρακτηριστικά όπως οι κλάσεις (classes), οι λειτουργικές μονάδες (modules), οι στατικοί τύποι (static typing), η εισαγωγή/εξαγωγή ενοτήτων (module import/export) κ.α.

## 3.2 Τεχνολογίες που χρησιμοποιήθηκαν

Οι βασικές τεχνολογίες που χρησιμοποιήθηκαν για την υλοποίηση της διαδικτυακής εφαρμογής ανάλυσης και επεξεργασίας εικόνων εγγράφου είναι οι ακόλουθες:

- **Angular** [9]  
Η Angular είναι ένα JavaScript Framework για τη δημιουργία διαδραστικών client side εφαρμογών που τρέχουν στον browser.
- **Canvas API** [10]  
Το Canvas API (Application Program Interface) παρέχει χαμηλού επιπέδου δυνατότητες σχεδιασμού γραφικών μέσω JavaScript και HTML, εστιάζοντας κυρίως στα δισδιάστατα γραφικά. Μπορεί να χρησιμοποιηθεί για σχεδιοκίνηση (animation), γραφικά παιχνιδιών, οπτικοποίηση δεδομένων (data visualization), επεξεργασία εικόνας και βίντεο κ.α.
- **FabricJS** [11]  
Το FabricJS είναι μια ανοιχτού λογισμικού (open source) βιβλιοθήκη JavaScript η οποία παρέχει δυνατότητα αλληλεπίδρασης του χρήστη με τα γραφικά αντικείμενα που εμφανίζονται στο Canvas (περιοχή της οθόνης όπου αποτυπώνονται τα γραφικά στοιχεία).
- **Web Workers API** [12]  
Με τη χρήση Web Workers επιτυγχάνεται η εκτέλεση κώδικα σε ξεχωριστό νήμα εκτέλεσης (thread) από το κύριο. Έτσι, διεργασίες μεγάλου υπολογιστικού κόστους και κατανάλωσης πόρων μπορούν να εκτελεστούν απομονωμένα ή παράλληλα, επιτρέποντας το κύριο νήμα (που αφορά κυρίως το γραφικό περιβάλλον) να τρέχει χωρίς καθυστέρηση.
- **Angular Material Library** [13]  
Το Angular Material είναι μια συλλογή από επαναχρησιμοποιήσιμα, διαδραστικά και προσαρμοζόμενα γραφικά στοιχεία (UI elements), θέματα (themes) και συστατικά (components) διεπαφής χρήστη σύμφωνα με τα σχεδιαστικά πρότυπα του Material Design.
- **FileReader API** [14]  
Μέσω του FileReader API της JavaScript επιτρέπεται το ασύγχρονο διάβασμα και φόρτωμα αρχείων από τον υπολογιστή του χρήστη στην διαδικτυακή εφαρμογή (web app).
- **Web Storage API** [15]  
Το Web Storage API της JavaScript δίνει δυνατότητα αποθήκευσης δεδομένων στο τοπικό χώρο αποθήκευσης (localStorage) του προγράμματος περιήγησης (browser) του χρήστη.

### 3.3 Το πλαίσιο λογισμικού Angular

#### 3.3.1 Γενική περιγραφή

Η Angular είναι ένα ολοκληρωμένο πλαίσιο λογισμικού (framework) για την δημιουργία εφαρμογών διαδικτύου με τη χρήση HTML/CSS και TypeScript. Αποτελεί έργο ανοιχτού λογισμικού, το οποίο αναπτύσσεται και συντηρείται κυρίως από ομάδες προγραμματιστών της Google. Υποστηρίζει τόσο τις κινητές συσκευές όσο και τους σταθερούς υπολογιστές. Το τελευταίο διάστημα έχει αποτελέσει ένα από τα δημοφιλέστερα και σημαντικότερα framework για την ανάπτυξη client side εφαρμογών πταγκοσμίως. Με το συγκεκριμένο framework αναπτύχθηκε η εφαρμογή της παρούσας διπλωματικής. Για αυτό το λόγω, παρακάτω στο κεφάλαιο αναλύεται η αρχιτεκτονική της Angular και περιγράφεται ο τρόπος χρήσης της υλοποιημένης εφαρμογής.

#### 3.3.2 Η αρχιτεκτονική της Angular

Το πλαίσιο λογισμικού Angular [9] είναι σχεδιασμένο ώστε να δημιουργούνται εφαρμογές οι οποίες χτίζονται με συστατικά λογισμικού (web components).

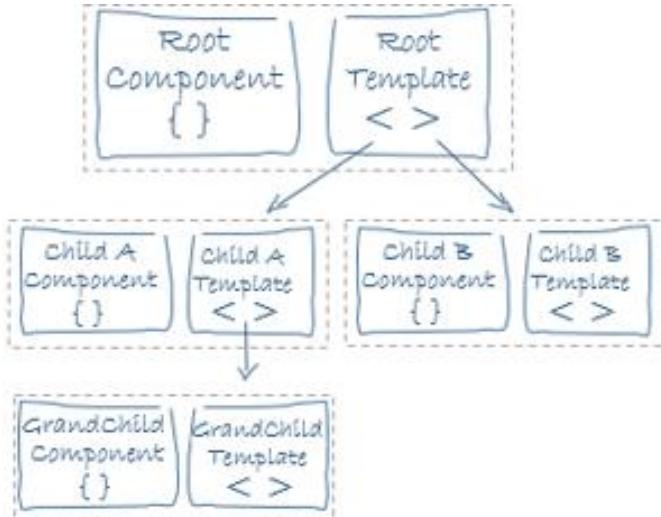
Τα δομικά στοιχεία μιας Angular εφαρμογής είναι τα ακόλουθα:

- **Modules (Λειτουργικές Μονάδες)**

Οι Λειτουργικές Μονάδες περιέχουν components (συστατικά), services (υπηρεσίες) και άλλα αρχεία κώδικα, τα οποία είναι αφιερωμένα σε ένα τομέα της εφαρμογής, μιας ροής εργασιών ή σε ένα πολύ συναφές σύνολο δυνατοτήτων. Μπορούν να χρησιμοποιηθούν σε συνδυασμό με άλλα Modules για την ολοκλήρωση της διαδικτυακής εφαρμογής. Η οργάνωση του κώδικα σε ξεχωριστές λειτουργικές ενότητες βοηθά στη διαχείριση της ανάπτυξης σύνθετων εφαρμογών με έμφαση την αξιοπιστία, την επαναχρησιμότητα του κώδικα και την επεκτασιμότητα.

- **Components (Συστατικά)**

Τα Συστατικά είναι το κύριο εργαλείο για τη δόμηση μια Angular εφαρμογής. Περιέχουν TypeScript κώδικα, ο οποίος είναι υπεύθυνος για τις μονομερείς και διμερείς διασυνδέσεις δεδομένων της σελίδας αλλά και για τη διαχείριση γεγονότων (events) όπως το πάτημα κουμπιών. Κάθε Component ορίζει μια κλάση, η οποία περιέχει τα δεδομένα της εφαρμογής, την προγραμματιστική λογική και σχετίζεται με ένα HTML Template. Κάθε συστατικό μπορεί να γίνει ανεξάρτητο και εύκολα επαναχρησιμοποιήσιμο σε διαφορετικά σημεία της εφαρμογής.

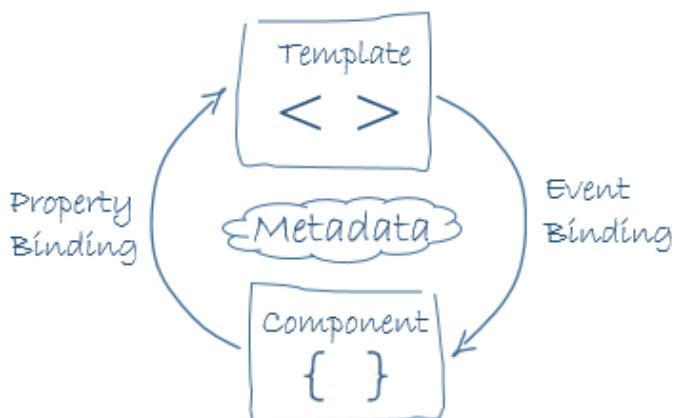


Εικόνα 3.1: Ιεραρχική δόμηση των components σε μια εφαρμογή Angular.

- **Templates (Πρότυπα)**

Τα πρότυπα είναι HTML κομμάτια κώδικα με συντακτικές προσθήκες της Angular που περιγράφουν το component view (το σύνολο των γραφικών στοιχείων του component, τα οποία εμφανίζονται στην οθόνη). Αυτές οι επιπλέον προσθήκες ή δυνατότητες αναφέρονται σε:

- **Data binding (σύνδεση δεδομένων)**: μεταφορά δεδομένων μεταξύ του component class και του template.
- **Pipes (διασωλήνωση)**: μορφοποίηση δεδομένων που εμφανίζονται.
- **Directives (οδηγίες)**: αλλαγή διάταξης, εμφάνισης ή συμπεριφοράς προσθέτοντας, αφαιρώντας και μεταβάλλοντας στοιχεία στην οθόνη.

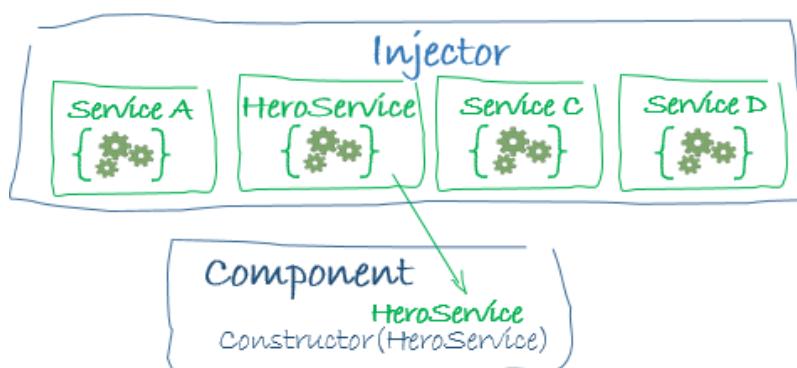


Εικόνα 3.2: Επικοινωνία μεταξύ component και template.

Ένα Template μπορεί να περιέχει ιεραρχικά άλλα εμφωλευμένα πρότυπα που ανήκουν σε διαφορετικά συστατικά.

- **Services (Υπηρεσίες)**

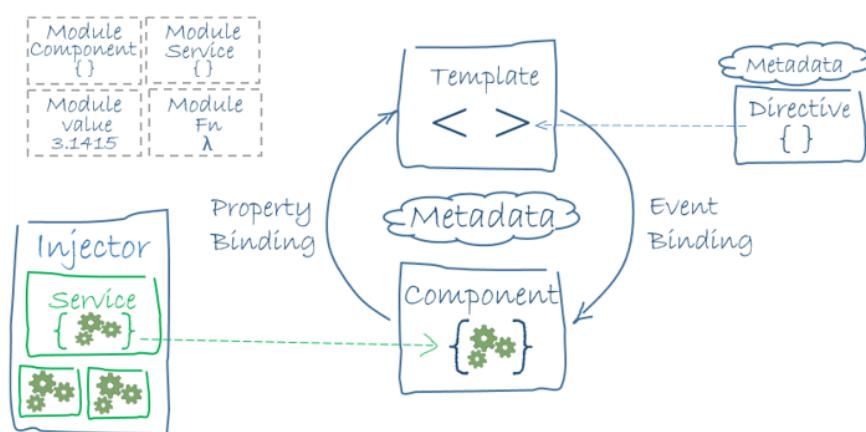
Οι Υπηρεσίες είναι τυπικά κλάσεις της TypeScript, οι οποίες εκτελούν μια συγκεκριμένη ενέργεια ή διαδικασία και παρέχουν ένα καλό ορισμένο αποτέλεσμα όπως τιμή, συνάρτηση ή χαρακτηριστικό που χρειάζεται η εφαρμογή. Χρησιμοποιούνται για τη καλύτερη δόμηση του κώδικα και την επαναχρησιμοποίηση κοινών λειτουργιών σε διαφορετικά σημεία της εφαρμογής. Τα Services παρέχονται στα Components μέσω Dependency Injection.



Εικόνα 3.3: Παροχή services στο component με τη χρήση Dependency Injection.

- **Routing (Δρομολόγηση)**

Η Δρομολόγηση κατά βάση αφορά την πλούγηση σε διαφορετικές καταστάσεις και ιεραρχικές όψεις της εφαρμογής, όπως μεταξύ διαφορετικών συστατικών. Περιλαμβάνει χαρτογραφήσεις, δηλαδή ανταποκρίσεις ανάμεσα στο συστατικό και ένα URL μονοπάτι.



Εικόνα 3.4: Αρχιτεκτονική της Angular.

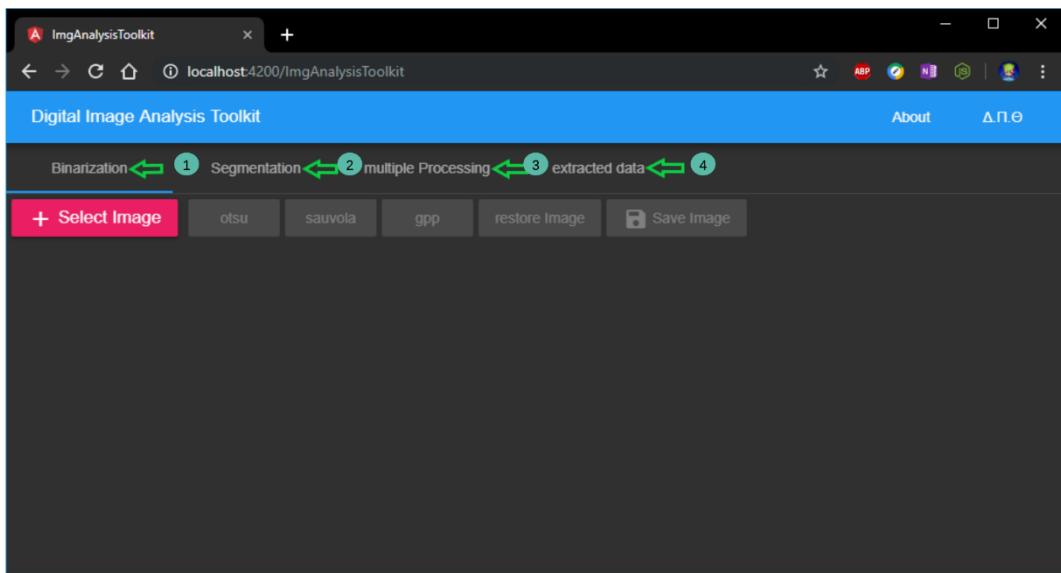
Η παραπάνω δομή και αρχές ανάπτυξης ακολουθήθηκαν πιστά στην υλοποίηση της εφαρμογής. Συνοπτικά έγινε:

- Δημιουργία υπηρεσιών για την επικοινωνία και ανταλλαγή δεδομένων μεταξύ των components, για την εισαγωγή εικόνων και Ground Truth εγγράφων, για την αποθήκευση και ανάκτηση δεδομένων από το χώρο αποθήκευσης του browser και για την διαχείριση των Web Workers, μέσω των οποίων εκτελούνται οι αλγόριθμοι δυαδικοποίησης, κατάτμησης και αξιολόγησης σε ξεχωριστό νήμα.
- Η εφαρμογή επικεντρώνεται σε ένα μοναδικό τομέα, αυτό της ανάλυσης εικόνων εγγράφου, οπότε χρησιμοποιήθηκε ένα βασικό module, στο οποίο δηλώθηκαν τα συστατικά και οι υπηρεσίες της εφαρμογής. Παράλληλα στο βασικό module προστέθηκε και η λειτουργική μονάδα που παρέχει η βιβλιοθήκη Angular Material [13], η οποία περιέχει έτοιμα συστατικά, όπως γραφικά στοιχεία πινάκων, φορμών εισαγωγής, καρτελών περιεχομένου, κουμπιών επιλογής κ.α.
- Πλοήγηση μεταξύ της κύριας εφαρμογής και της σελίδας πληροφοριών μέσω της υπηρεσίας δρομολόγησης της Angular.
- Ευρέα και καθολική χρήση των πρόσθετων χαρακτηριστικών της Angular, όπως οι δυνατότητες data binding, τα pipes και τα directives.
- Χωρισμός των διάφορων HTML στοιχείων, λειτουργιών και τις προγραμματιστικής λογικής πίσω από αυτά (διαδικασιών κατωφλίωσης, κατάτμησης, προβολής δεδομένων, εικόνων, αποτελεσμάτων, άλλων γραφικών στοιχείων αλληλεπίδρασης κ.α. ) σε ξεχωριστά components, ώστε να επιτευχθεί εύκολη συντήρηση και επέκταση στην εφαρμογή.

## 3.4 Παρουσίαση της εφαρμογής

### 3.4.1 Γενική επισκόπηση των βασικών περιοχών της εφαρμογής.

Όταν ο χρήστης επισκέπτεται με το πρόγραμμα περιήγησης την εφαρμογή, η πρώτη εικόνα της διεπαφής που θα συναντήσει είναι η παρακάτω:

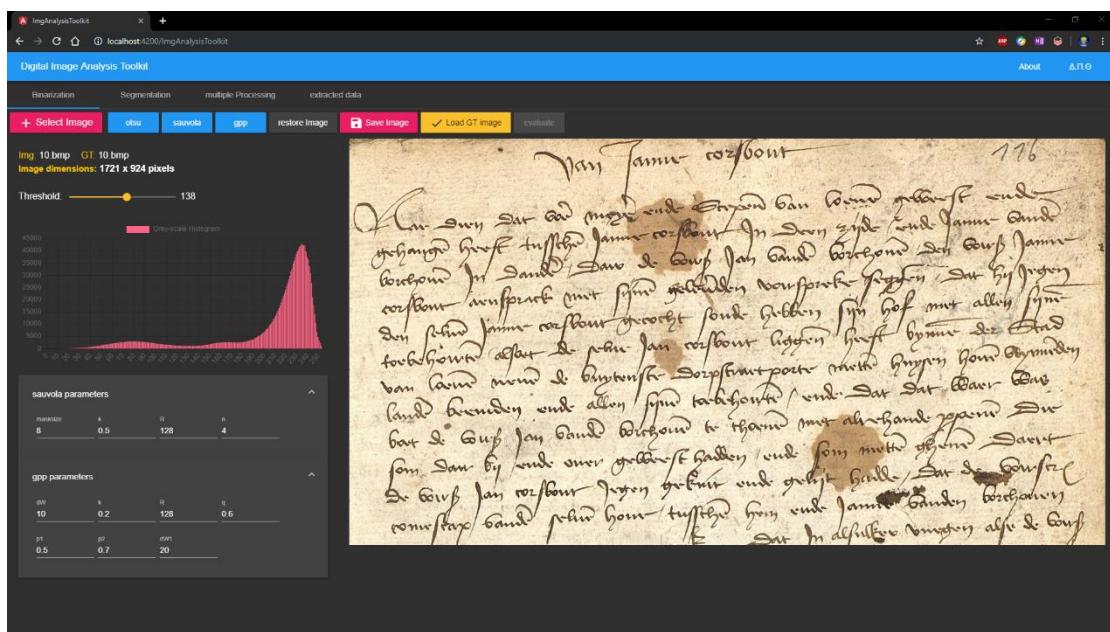


**Εικόνα 3.5: Οι τέσσερις περιοχές χρήσης της εφαρμογής.**

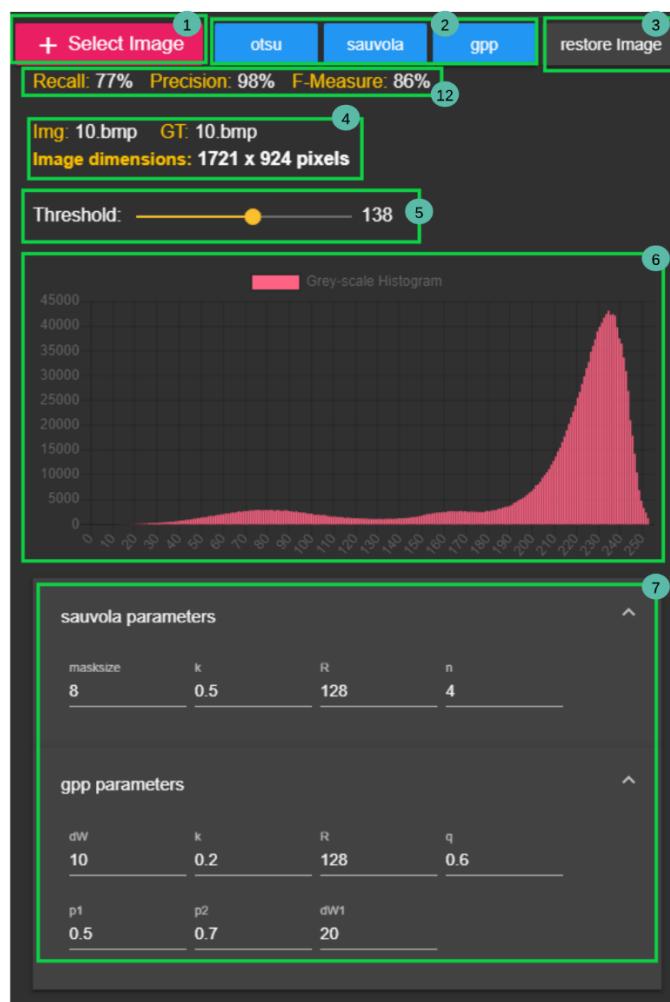
Όπως φαίνεται η εφαρμογή μέσω καρτελών (Tabs) χωρίζεται σε τέσσερεις περιοχές χρήσης. Οι πρώτες δύο αποτελούν βασικές ενότητες της εφαρμογής και αφορούν τη δυαδική μετατροπή “*Binarization*” (περιοχή 1) και τη κατάτμηση “*Segmentation*” (περιοχή 2) μιας εικόνας εγγράφου. Η τρίτη ενότητα “*multiple Processing*” (περιοχή 3) επιτρέπει τη μαζική επεξεργασία (κατωφλίωση και τμηματοποίηση) πολλών εικόνων εγγράφου μαζί. Τέλος, η τέταρτη καρτέλα “*extracted data*” (περιοχή 4) παρουσιάζει τα δεδομένα που εξάχθηκαν.

### 3.4.2 Η καρτέλα δυαδικής μετατροπής

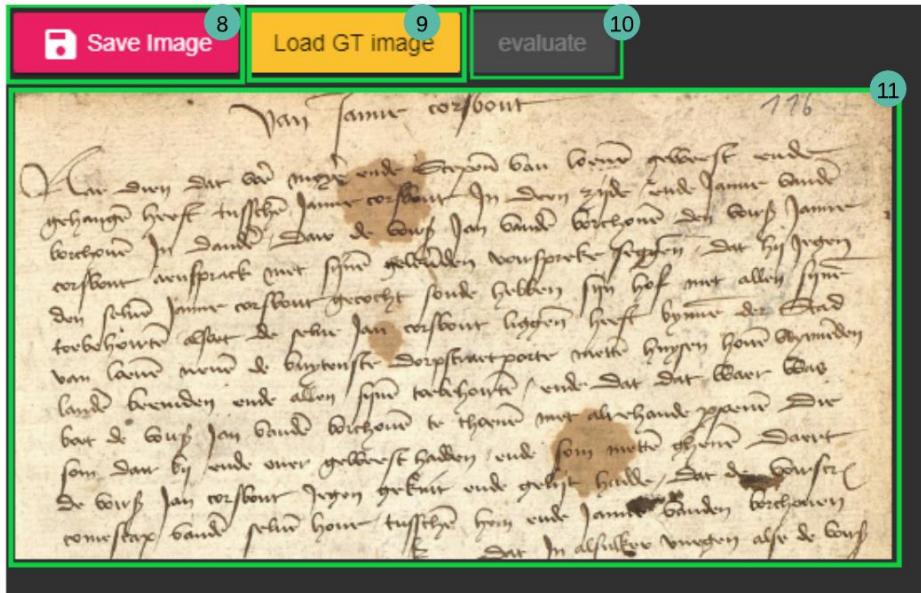
Η πρώτη βασική ενότητα της εφαρμογής αφορά τη δυαδική μετατροπή.



Εικόνα 3.6: Η καρτέλα δυαδικής μετατροπής.



Εικόνα 3.7: Γραφικά στοιχεία της διεπαφής χρήστη - καρτέλας “Binarization” (περιοχές 1-7, 12).

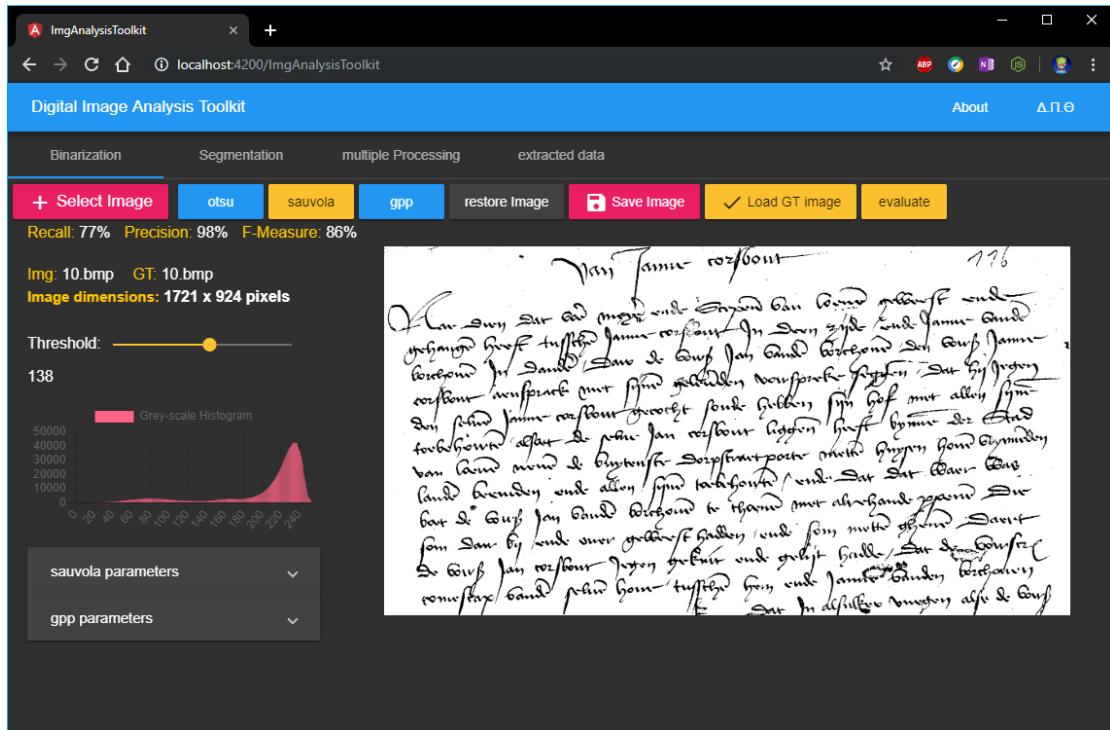


**Εικόνα 3.8:** Γραφικά στοιχεία της διεπαφής χρήστη - καρτέλας “*Binarization*” (περιοχές 8-10) και το *canvas* - το τμήμα της οθόνης όπου απεικονίζεται η εικόνα εγγράφου (περιοχή 11).

Αρχικά πρέπει να φορτωθεί η εικόνα εγγράφου στην εφαρμογή. Αυτό γίνεται από το κουμπί εισαγωγής εγγράφου “*Select Image*” (περιοχή 1). Αφού επιλεχτεί το αρχείο εικόνας, διαβάζεται και φορτώνεται από την εφαρμογή μέσω του FileReader API [14] της JavaScript. Έπειτα, με χρήση του Canvas API [10] εξάγονται τα δεδομένα της εικόνας εγγράφου, δηλαδή οι διαστάσεις και τα εικονοστοιχεία της, και η εικόνα εμφανίζεται στο *canvas* (περιοχή 11). Επίσης, δημιουργείται το ιστόγραμμα γκρι αποχρώσεων της εικόνας και αναπαρίσταται γραφικά (περιοχή 6). Ακόμη, εμφανίζονται οι διαστάσεις και το όνομα της εικόνας (περιοχή 4). Με τη χρήση της ροδέλας του ποντικιού (mouse wheel) παρέχεται δυνατότητα οπτικής μεγέθυνσης/σμίκρυνσης (zoom in/out) της εικόνας στην οθόνη.

Ο χρήστης μπορεί να επιλέξει μια εκ των τριών μεθόδων δυαδικοποίησης Otsu [1], Sauvola [2] και GPP [3] που υλοποιήθηκαν (περιοχή 2). Μόλις επιλέξει, ξεκινά η εφαρμογή της αντίστοιχης μεθόδου δυαδικής μετατροπής, διαδικασία η οποία εκτελείται ασύγχρονα σε ξεχωριστό νήμα από το κύριο νήμα (*main thread*) του GUI (Graphical User Interface), με τη χρήση Web Worker [12]. Όταν ολοκληρωθεί η διαδικασία, το αποτέλεσμα, δηλαδή η δυαδική εικόνα, επιστρέφεται στο *main thread* της JavaScript και αποτυπώνεται εκ νέου στην οθόνη μέσω του Canvas API (περιοχή 11). Για να αναιρεθεί η δυαδική μετατροπή, χρησιμοποιείται το κουμπί επαναφοράς “*restore Image*” (περιοχή 3). Μέσω των πεδίων επιλογής (περιοχή 7) ο χρήστης δύναται να αλλάξει τις προκαθορισμένες παραμέτρους των τεχνικών κατωφλίωσης Sauvola και GPP. Επιπλέον, με το slider επιλογής (περιοχή 5) ο χρήστης μπορεί να ρυθμίσει χειροκίνητα το κατώφλι, με βάση το οποίο ακολουθεί η καθολική κατωφλίωση

της εικόνας. Επιπρόσθετα, υπάρχει επιλογή αποθήκευσης της δυαδικής εικόνας εγγράφου στον υπολογιστή του χρήστη με το κουμπί αποθήκευσης “Save Image” (περιοχή 4).

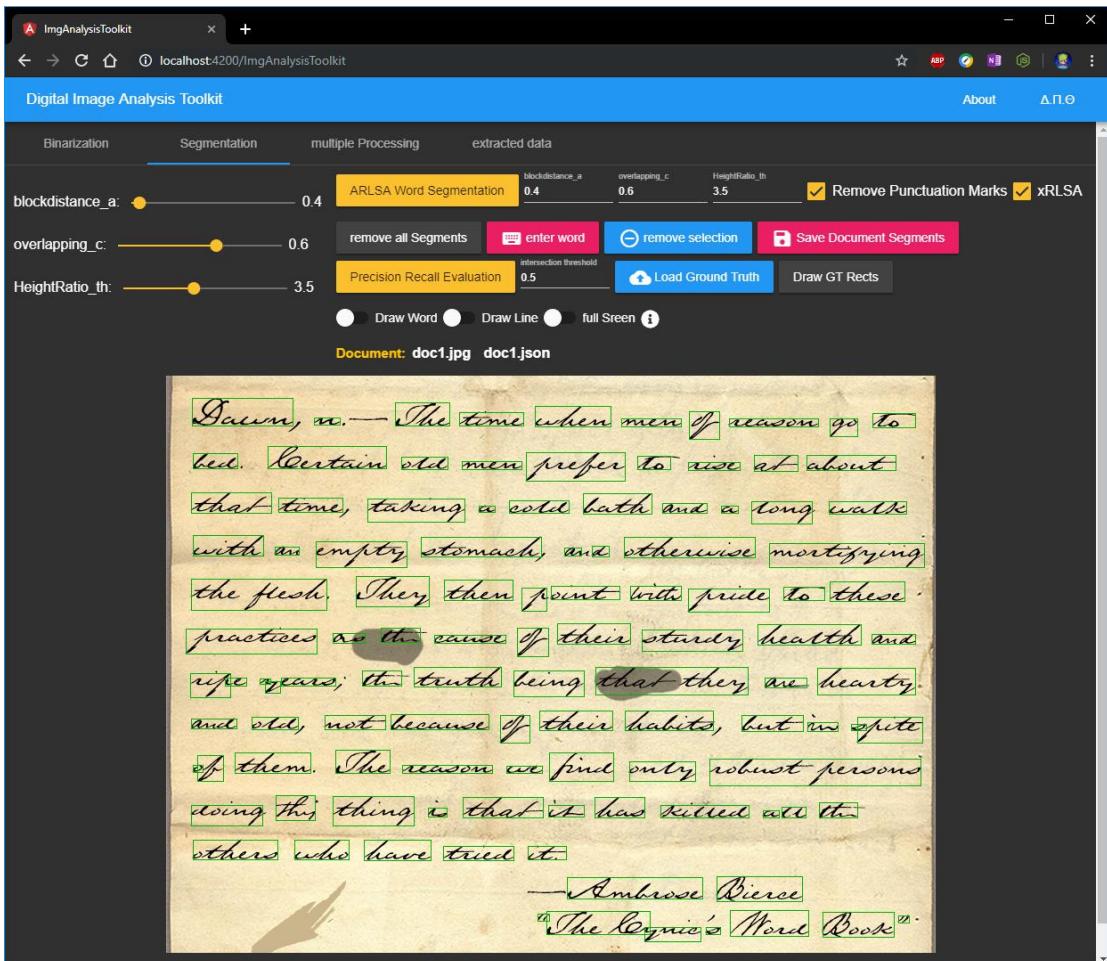


**Εικόνα 3.9:** Με την ολοκλήρωση της δυαδικής μετατροπής, το canvas ανανεώνεται, εμφανίζοντας τη νέα δυαδική εικόνα.

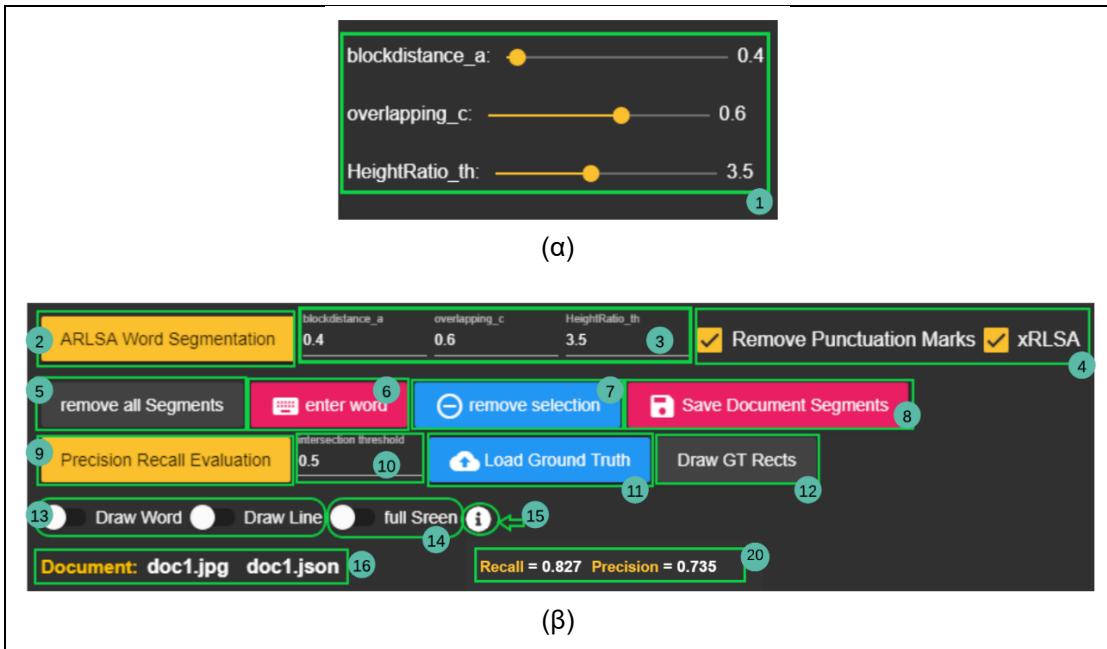
Τέλος, δίνεται δυνατότητα αξιολόγησης του δυαδικού αποτελέσματος. Από το κουμπί εισαγωγής “Load GT Image” (περιοχή 9) φορτώνεται η δυαδική Ground Truth εικόνα και με το διπλανό κουμπί “evaluate” (περιοχή 10) εφαρμόζεται η αξιολόγηση. Έτσι, εμφανίζονται την οθόνη τα ποσοστά Recall, Precision και F-Measure (περιοχή 12).

### 3.4.3 Η καρτέλα κατάτμησης

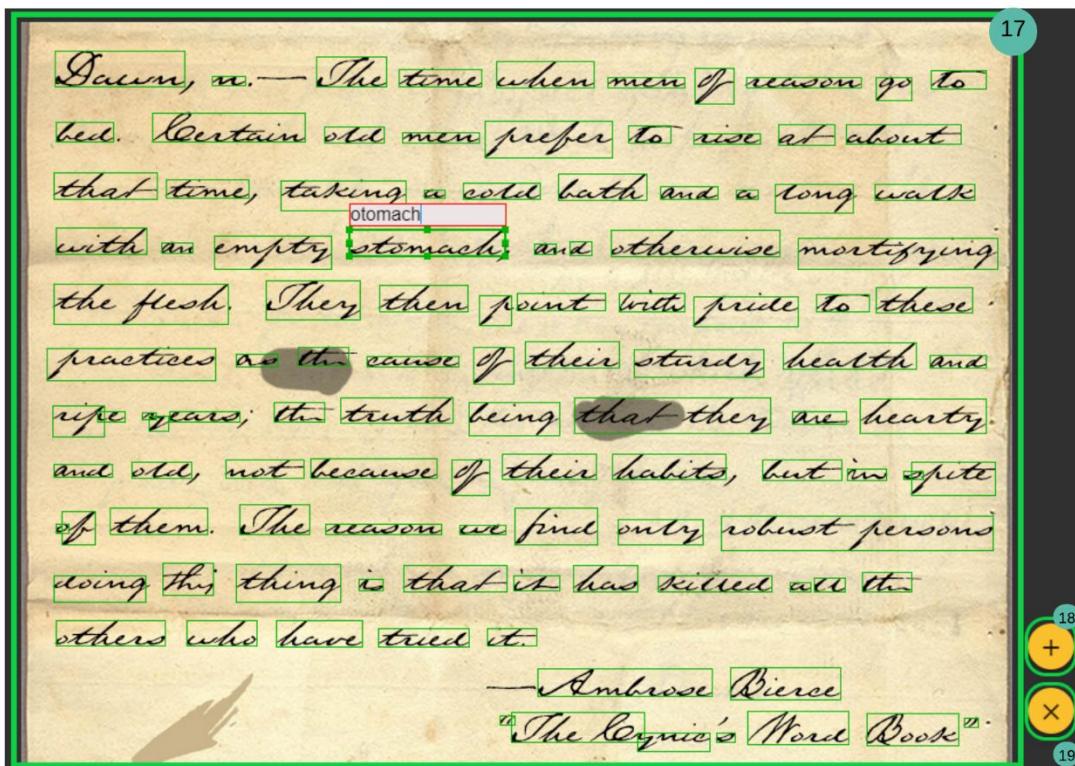
Αφού ο χρήστης φορτώσει στην εφαρμογή μια εικόνα εγγράφου και την μετατρέψει σε δυαδική, μπορεί να περιηγηθεί στην επόμενη βασική κατηγορία της εφαρμογής η οποία αφορά τη κατάτμηση περιοχών κειμένου.



Εικόνα 3.10: Η καρτέλα κατάτμησης.



Εικόνα 3.11: Γραφικά στοιχεία της διεπαφής χρήστη – καρτέλας “Segmentation”. (α) Περιοχή 1, (β) περιοχές 2-16, 20.



Εικόνα 3.12: Το βασικό canvas εμφάνισης και επεξεργασίας των αποτελεσμάτων της κατάτμησης (περιοχή 17) μαζί με γραφικά στοιχεία (περιοχές 18-19).

Με επιλογή του κουμπιού κατάτμησης “ARLSA Word Segmentation” (περιοχή 2) ξεκινά η διαδικασία τμηματοποίησης και εξαγωγής του κειμένου. Η διαδικασία αυτή εκτελείται ασύγχρονα σε ξεχωριστό νήμα με τη χρήση Web Workers και μόλις ολοκληρωθεί, σχηματίζονται στο canvas (περιοχή 17) τα ορθογώνια πλαίσια (Segments) τα οποία αντιστοιχούν στις λέξεις ή γραμμές που εντόπισε ο αλγόριθμος. Ο καθαρισμός του canvas, δηλαδή η διαγραφή των segments, γίνεται με το κουμπί διαγραφής “remove all Segments” (περιοχή 5). Αν ο χρήστης δεν είχε δυαδικοποιήσει την εικόνα στην προηγούμενη καρτέλα τότε, πριν ξεκινήσει η διαδικασία της κατάτμησης, γίνεται αυτόματη δυαδική μετατροπή στο έγγραφο με τη μέθοδο GPP [3].

Η αλλαγή των προκαθορισμένων παραμέτρων-κατωφλιών του ARLSA [6] μπορεί να γίνει είτε από τα slider επιλογής (περιοχή 1) είτε από τις φόρμες εισαγωγής (περιοχή 3). Επίσης, υπάρχει επιλογή απενεργοποίησης της αφαίρεσης σημείων στίξης (“Remove Punctuation Marks”) και της κατάτμησης λέξεων με βάση το ιστόγραμμα των κάθετων προβολών του υποβάθρου (“xRLSA”) (περιοχή 4).

Για να αξιολογηθεί η υλοποιημένη μέθοδος κατάτμησης, πρέπει πρώτα να εισαχθεί το κατάλληλο αρχείο τύπου JSON, το οποίο θα περιέχει τα Ground Truth Segments (σωστές περιοχές που βρίσκονται οι λέξεις). Αυτό γίνεται με το κουμπί εισαγωγής “Load Ground Truth” (περιοχή 11). Η εκτέλεση της

διαδικασίας αξιολόγησης ξεκινά με το κουμπί “Precision Recall Evaluation” (περιοχή 9) και μόλις ολοκληρωθεί, οι τιμές Precision και Recall εμφανίζονται στη οθόνη (περιοχή 20). Ο χρήστης μπορεί να καθορίσει το κατώφλι επικάλυψης για την αξιολόγηση μέσω της φόρμας εισαγωγής “intersection threshold” (περιοχή 10). Ακόμη, έχει τη δυνατότητα ταυτόχρονης απεικόνισης των αποτελεσμάτων της κατάτμησης ARLSA και των σωστόν περιοχών του Ground Truth, επιλέγοντας το κουμπί “Draw GT Segments” (περιοχή 12).

**Oι Α'. τόμοι του ΒΡΕΤΤΑΝΙΚΟΥ ΑΣΤΕΡΟΣ, οὓς ὑπεσχέθημεν εἰς τοὺς ἐν Ελλάδι συνδρομητὰς ἡμῶν εἰσὶν ἡδη ἔτοιμοι, καὶ δυνάμεθα νὰ παραδώσωμεν αὐτὸὺς εἰς ὅποιονδήποτε ἐνταῦθα ἀνταποκριτὴν ἢ φίλον των ἥθελον ἐκεῖνοι μᾶς παραγγείλει. Τὸ μετρον τοῦτο λαμβάνομεν, ὅπως ἀποφύγωμεν τὴν ἐν Ελλάδι κατάσχεσιν τῶν τόμων.**

Εικόνα 3.13: Ταυτόχρονη εμφάνιση στο canvas των ARLSA (πράσινα πλαίσια) και Ground Truth (μπλε πλαίσια) Segments.

Το όνομα της εικόνας εγγράφου μαζί με το όνομα του Ground Truth αρχείου εμφανίζονται στην οθόνη (περιοχή 16).

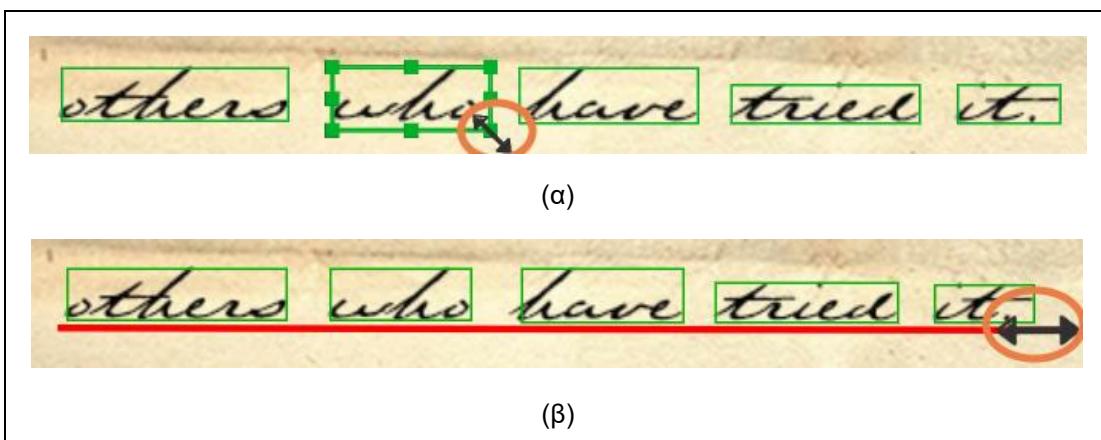
Επιπλέον, αν η εικόνα έχει πλάτος μικρότερο της οθόνης, ο χρήστης μπορεί να μεγαλώσει το εύρος του canvas με τη χρήση του slide-toggle διακόπτη “full screen” (περιοχή 14).

Επιπρόσθετα, στο σημείο αυτό η εφαρμογή περιέχει το εργαλείο που αναπτύχθηκε με τη βιβλιοθήκη FabricJS [11], η οποία δίνει τη δυνατότητα αλληλεπίδρασης του χρήστη με αντικείμενα που είναι σχεδιασμένα στο canvas. Σκοπός του εργαλείου είναι η οπτική αναπαράσταση των segments, η δημιουργία Ground Truth segments λέξεων και γραμμών κειμένου και η εισαγωγή λέξεων στα αντίστοιχα segments. Συγκεκριμένα αναπτύχθηκε κώδικας για να υλοποιηθούν τα παρακάτω χαρακτηριστικά:

- Responsive layout (ανταποκρινόμενη διάταξη) ώστε το canvas και κάθε αντικείμενο που περιέχει, να αποτυπώνονται σωστά σε κάθε μέγεθος οθόνης και να προσαρμόζονται δυναμικά όταν ο χρήστης μεταβάλει το μέγεθος του παραθύρου περιήγησης.
- Άλλαγή μεγέθους, θέσης και διάταξης των segments μέσα στο canvas.

- Δημιουργία εργαλείων για σχεδιασμό ορθογώνιων παραλληλογράμμων και γραμμών με τη χρήση του ποντικιού.
- Προσθήκη πεδίων εισαγωγής λέξεων πάνω από κάθε word segment.
- Δυνατότητα χειρισμού της διεπαφής με το πληκτρολόγιο.

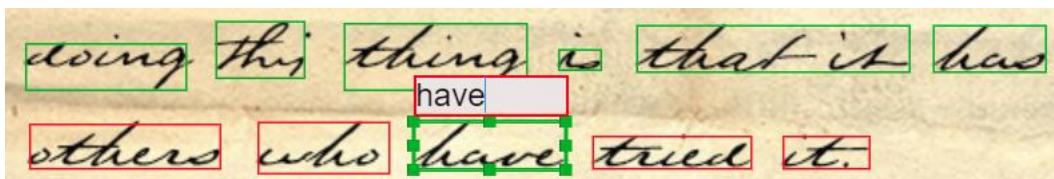
Από τους slide-toggle διακόπτες “Draw Word” και “Draw Line” (περιοχή 13) μπορούν να ενεργοποιηθούν τα αντίστοιχα εργαλεία σχεδιασμού τετραγώνων και γραμμών. Με τα εργαλεία αυτά ο χρήστης επιλέγει τις λέξεις και τις γραμμές του εγγράφου, σχεδιάζοντας με τον κέρσορα του ποντικιού τετράγωνα πλαίσια και γραμμές στις αντίστοιχες περιοχές του κειμένου. Με αυτό το τρόπο σχηματίζονται τα word και line segments πάνω στο canvas.



Εικόνα 3.14: Παράδειγμα επιλογής των λέξεων και γραμμών του κειμένου με την χρήση των εργαλείων (a) σχεδιασμού τετραγώνου και (b) σχεδιασμού γραμμής.

Όσον αφορά την σχεδίαση των ορθογωνίων πλαισίων γύρω από τις λέξεις, ο χρήστης δεν χρειάζεται να επιλέξει ακριβώς τα όρια της λέξης, αλλά το εργαλείο ανιχνεύει τις ακμές της και την επιλέγει αυτόματα.

Παράλληλα, προστέθηκε δυνατότητα εισαγωγής των λέξεων που περιέχει η εικόνα εγγράφου. Πάνω από κάθε segment ανοίγει μια φόρμα εισαγωγής κειμένου, όπου ο χρήστης πληκτρολογεί την αντίστοιχη λέξη.



Εικόνα 3.15: Παράδειγμα εισαγωγής λέξης.

Το χρώμα των τετραγωνικών word segments στα οποία έχουν προστεθεί λέξεις εμφανίζονται με κόκκινο χρώμα, ενώ τα υπόλοιπα με πράσινο χρώμα.

Με τοποθέτηση του ποντικιού πάνω στο εικονίδιο “i” (περιοχή 15) εμφανίζονται οι διαθέσιμες συντομεύσεις πληκτρολογίου και ποντικιού, με τις οποίες ο χρήστης αλληλεπιδρά με τα αντικείμενα που είναι σχεδιασμένα στο canvas (γραφικά στοιχεία word και line Segments). Οι συντομεύσεις αυτές είναι:

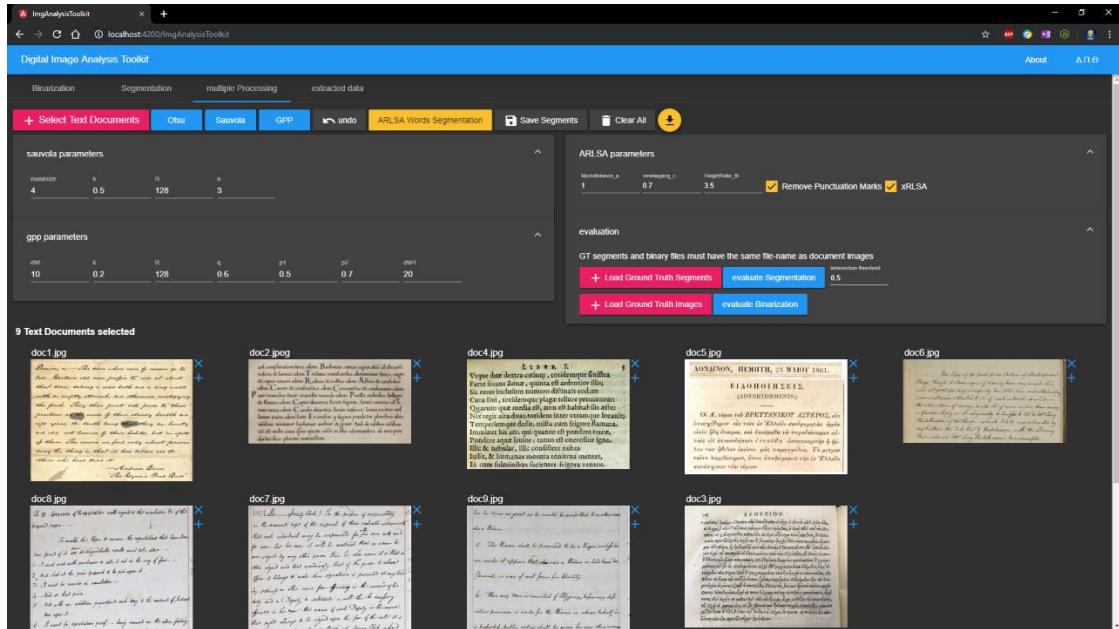
- 1) **Mouse wheel** (ροδέλα ποντικιού): μεγέθυνση οπτικά της εικόνας.
- 2) **ALT + mouse drag** (μετακίνηση ποντικιού): μετακίνηση της εικόνας μέσα στο canvas.
- 3) **CRTL + BACKSPACE**: διαγραφή από το canvas του επιλεγμένου αντικειμένου.
- 4) **CRTL + SPACE**: άνοιγμα της φόρμας εισαγωγής λέξης πάνω από το αντίστοιχο word segment.
- 5) **ENTER**: προσθήκη της πληκτρολογημένης λέξης και μετακίνηση της φόρμας εισαγωγής στο επόμενο word segment.

Οι ενέργειες των συντομεύσεων 3, 4 και 5 μπορούν να επιτευχθούν και με τα κουμπιά επιλογής “remove selection” (περιοχή 7), “enter word” (περιοχή 6) και του εικονιδίου “+” (περιοχή 18) αντίστοιχα, ενώ με το εικονίδιο “x” (περιοχή 19) κλείνει η φόρμα εισαγωγής λέξεων, αν είναι ανοιχτή.

Τέλος, με το κουμπί “Save Document Segments” (περιοχή 8), ο χρήστης έχει τη δυνατότητα αποθήκευσης (στο localStorage του browser) των word segments (διαστάσεις και θέση μέσα στην εικόνα) που είναι σχεδιασμένα στο canvas μαζί με τις αντίστοιχες εισαγμένες λέξεις, ώστε την επόμενη φορά να συνεχίσει την εργασία του από εκεί που την άφησε. Την επόμενη φορά λοιπόν, που θα επιλεχθεί η ίδια εικόνα, η εφαρμογή ψάχνει στο localStorage για αποθηκευμένα segments και λέξεις, και αν υπάρχουν, τα φορτώνει αυτόματα στο canvas.

#### 3.4.4 Η καρτέλα μαζικής επεξεργασίας

Η τρίτη ενότητα της εφαρμογής επιτρέπει τη μαζική επεξεργασία (κατωφλίωση και τμηματοποίηση) πολλών εικόνων εγγράφου ταυτόχρονα.



Εικόνα 3.16: Η καρτέλα μαζικής επεξεργασίας.

The screenshot shows the 'Digital Image Analysis Toolkit' interface with the 'multiple Processing' tab selected. It displays nine document thumbnails labeled doc1.jpg through doc9.jpg, each with a green plus sign icon. The documents include various historical and handwritten texts. A green circle with the number 18 is positioned above the doc4.jpg thumbnail, and a green circle with the number 19 is positioned above the doc1.jpg thumbnail. Below the thumbnails, there are two sets of performance metrics:

- doc1.jpg:** Recall = 0.895, Precision = 0.764
- doc2.jpg:** Recall = 0.827, Precision = 0.735
- doc4.jpg:** Recall = 0.963, Precision = 0.849

Εικόνα 3.17: Γραφικά στοιχεία της διεπαφής χρήστη – καρτέλας “multiple Processing” (περιοχές 17-21).



Εικόνα 3.18: Γραφικά στοιχεία της διεπαφής χρήστη – καρτέλας “multiple Processing”.  
 (α) Περιοχές 1-6, (β) περιοχές 7-16.

Αρχικά, με το κουμπί επιλογής “Select Text Documents” (περιοχή 1) διαλέγονται και φορτώνονται πολλές εικόνες μαζί, και απεικονίζονται στη οθόνη (περιοχή 13).

Στη συνέχεια, ο χρήστης δύναται να επιλέξει μέθοδο δυαδικής μετατροπής μεταξύ των τεχνικών Otsu [1], Sauvola [2] και GPP [3] (περιοχή 2). Μόλις εφαρμοστεί η μέθοδος, οι δυαδικές εικόνες που προκύπτουν εμφανίζονται στην οθόνη (περιοχή 18). Η αλλαγή των παραμέτρων της μεθόδου Sauvola γίνεται από τα πεδία εισαγωγής “sauvola parameters” (περιοχή 5), ενώ της GPP από τα πεδία εισαγωγής “gpp parameters” (περιοχή 6). Για αναίρεση της δυαδικοποίησης των εγγράφων χρησιμοποιείται το κουμπί αναίρεσης “undo” (περιοχή 3). Για την αξιολόγηση της δυαδικής μετατροπής πρώτα πρέπει να φορτωθούν οι σωστές Ground Truth εικόνες από το κουμπί εισαγωγής “Load Ground Truth Images” (περιοχή 15). Έπειτα, με το διπλανό κουμπί “evaluate

*Binarization*" (περιοχή 16) εκτελείτε η αξιολόγηση και εμφανίζονται οι τιμές Precision και Recall κάτω από κάθε εικόνα (περιοχή 21). Τα συνολικά αποτελέσματα όλων των εικόνων αποθηκεύονται σε ένα αρχείο κειμένου txt, το οποίο ο χρήστης μπορεί να κατεβάσει στον υπολογιστή του από το κουμπί download (περιοχή 9).

Η μαζική εξαγωγή των περιοχών κειμένου εκτελείται με το κουμπί κατάτμησης "ARLSA words Segmentation" (περιοχή 4). Αν δεν έχει προηγηθεί δυαδικοποίηση των εικόνων, τότε πριν την εφαρμογή της κατάτμησης, προηγείται αυτόματη δυαδική μετατροπή με τη τεχνική GPP [3]. Η ρύθμιση των παραμέτρων εκτέλεσης της τμηματοποίησης ARLSA [6] γίνεται από τα πεδία εισαγωγής "ARLSA parameters" (περιοχή 10).

Επίσης, επιλέγοντας το κουμπί "Load Ground Truth Segments" (περιοχή 12), ο χρήστης μπορεί να φορτώσει μαζικά τα αντίστοιχα JSON αρχεία που περιέχουν τα Ground Truth Segments. Για να αντιστοιχιστεί σωστά κάθε εικόνα με το σωστό Ground Truth αρχείο, πρέπει αμφότερα να έχουν το ίδιο όνομα. Αν ο χρήστης έχει φορτώσει τα αντίστοιχα Ground Truth αρχεία των εικόνων, τότε μόλις τελειώσει η κατάτμηση, τα αποτελέσματα της αξιολογούνται αυτόματα και εμφανίζονται οι τιμές Precision και Recall κάτω από κάθε εικόνα (περιοχή 21). Διαφορετικά, αν φορτωθούν τα Ground Truth αρχεία μετά την εφαρμογή της κατάτμησης, ο χρήστης μπορεί να αξιολογήσει τα αποτελέσματα της, επιλέγοντας χειροκίνητα το κουμπί "evaluate Segmentation" (περιοχή 13). Η ρύθμιση του κατωφλιού επικάλυψης για την αξιολόγηση επιτυγχάνεται μέσω της φόρμας εισαγωγής "*intersection threshold*" (περιοχή 14). Παράλληλα, τα συνολικά αποτελέσματα αξιολόγησης της κατάτμησης όλων των εικόνων αποθηκεύονται σε ένα αρχείο κειμένου txt, το οποίο ο χρήστης μπορεί να κατεβάσει στον υπολογιστή του από το κουμπί download (περιοχή 9).

Ακόμη, εμφανίζεται στην οθόνη ο συνολικός αριθμός εγγράφων που είναι ανοιχτός από την εφαρμογή (περιοχή 17). Επιπλέον, με το κουμπί διαγραφής "Clear All" (περιοχή 8) διαγράφονται όλες οι φορτωμένες εικόνες.

Επιπρόσθετα, δίπλα από κάθε εικόνα υπάρχουν δύο εικονίδια. Με επιλογή του πρώτου (περιοχή 19) η επιλεγμένη εικόνα διαγράφεται, ενώ με επιλογή του δεύτερου (περιοχή 20) η εφαρμογή μεταβαίνει στην καρτέλα της κατάτμησης για προβολή των segments και πιθανή επεξεργασίας τους ή μεμονωμένα κατάτμησης ξανά του εγγράφου. Τέλος, όπως και στη προηγούμενη ενότητα, υπάρχει κουμπί μαζικής αποθήκευσης "Save Segments" (περιοχή 7) των segments που εξάχθηκαν.

Οι διαδικασίες μαζικής δυαδικοποίησης και κατάτμησης των εικόνων απαιτούν αρκετούς υπολογιστικούς πόρους από το σύστημα, κυρίως μνήμης RAM. Οπότε, η χρήση πολλών Web Workers (έναν για κάθε εικόνα) που εκτελούνται όλοι μαζί παράλληλα σε ξεχωριστά νήματα, κρίνεται εφικτή μόνο για έναν ελάχιστο αριθμό εγγράφων μικρού μεγέθους. Η δυνατότητα αυτή υλοποιήθηκε

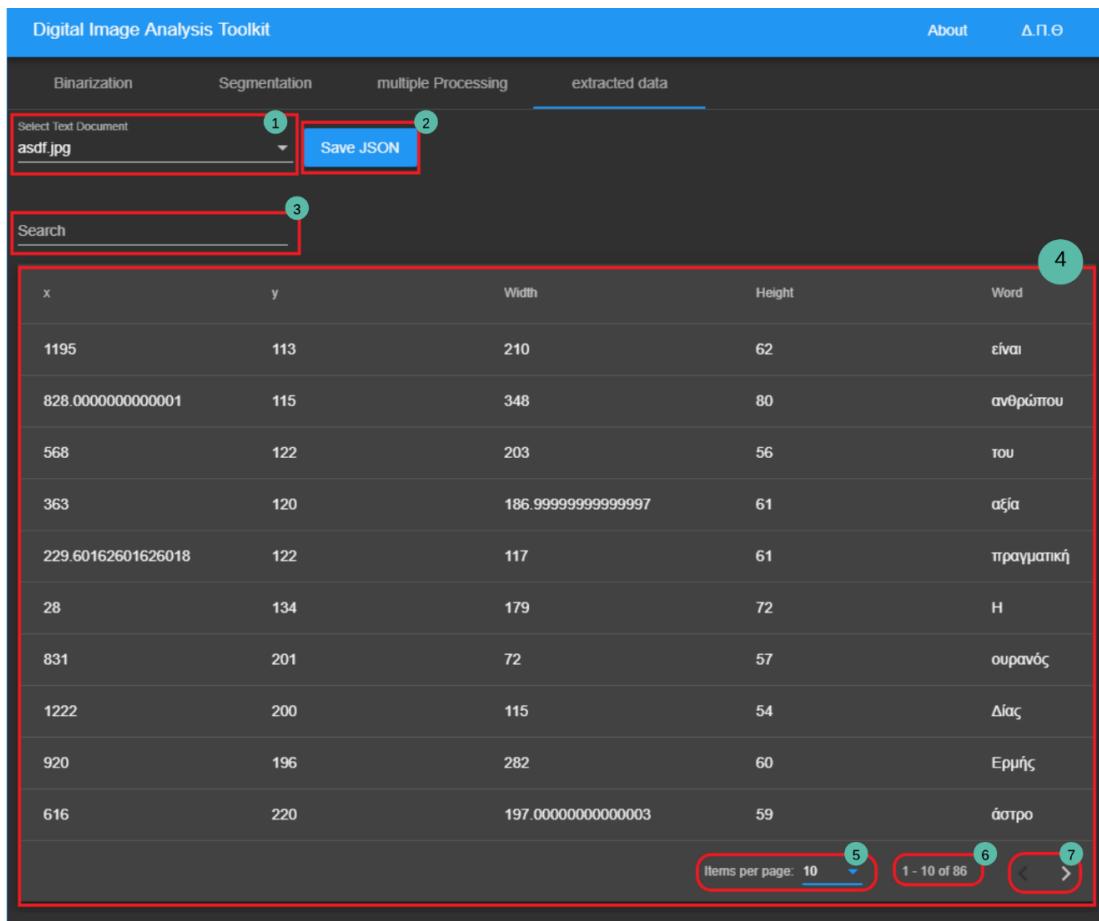
αρχικά, αλλά για να καταστεί δυνατή η χρήση της εφαρμογής σε συσκευές διαφορετικής επεξεργαστική ισχύς, τελικώς γίνεται χρήση ενός μόνου Web Worker, με τον οποίο εκτελείται ο αλγόριθμος μια φορά σειριακά για κάθε εικόνα.

### 3.4.5 Η καρτέλα εμφάνισης δεδομένων

Η τελευταία ενότητα της εφαρμογής παρουσιάζει τα δεδομένα που εξάχθηκαν και αποθηκεύτηκαν στο localStorage του browser.

x	y	Width	Height	Word
1195	113	210	62	είναι
828.000000000001	115	348	80	ανθρώπου
568	122	203	56	του
363	120	186.9999999999997	61	αξια
229.60162601626018	122	117	61	πραγματική
28	134	179	72	Η
831	201	72	57	ουρανός
1222	200	115	54	Δίας
920	196	282	60	Ερμής
616	220	197.0000000000003	59	άστρο

Εικόνα 3.19: Η καρτέλα εμφάνισης δεδομένων.



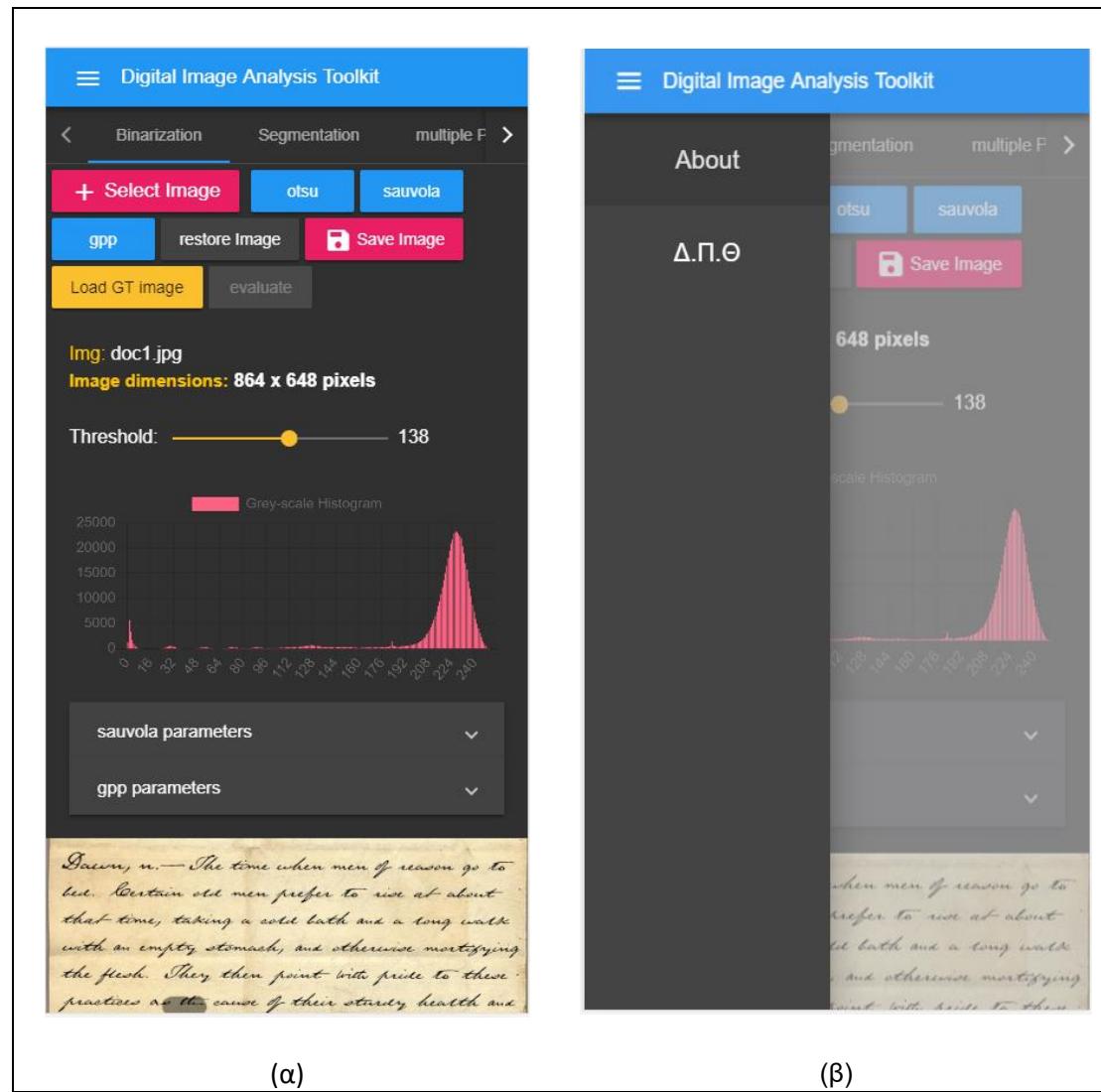
**Εικόνα 3.20: Γραφικά στοιχεία της διεπαφής χρήστη – καρτέλας “extracted data” (περιοχές 1-7).**

Αρχικά, ο χρήστης διαλέγει από τη λίστα επιλογής εγγράφου (περιοχή 1), το όνομα της εικόνας της οποίας τα αποθηκευμένα segments επιθυμεί να προβάλει. Μόλις επιλέξει λοιπόν, εμφανίζεται στη οθόνη ο πίνακας 4, ο οποίος περιέχει πληροφορίες για τα segments. Συγκεκριμένα, εμφανίζονται η θέση του κάθε segment μέσα στην εικόνα (αριθμός γραμμής και στήλης), οι διαστάσεις του (πλάτος και ύψος), καθώς και η αντίστοιχη λέξη που αντιπροσωπεύει.

Ακόμη, στο κάτω μέρος του πίνακα δίνονται επιλογές προσπέλασης του. Με τον επιλογέα (περιοχή 5) ρυθμίζεται πόσα segments θα εμφανιστούν σε μια σελίδα του πίνακα, ενώ δίπλα εμφανίζεται ο τρέχων αριθμός σελίδας (περιοχή 6). Με το βέλος (περιοχή 7) αλλάζει η σελίδα και εμφανίζονται τα επόμενα segments στον πίνακα. Επίσης, υπάρχει πεδίο αναζήτησης λέξεων των segments (περιοχή 3). Τέλος, ο χρήστης μπορεί να κατεβάσει τοπικά στον υπολογιστή του ένα αρχείο τύπου JSON, το οποίο θα περιέχει όλα τα δεδομένα των segments του εγγράφου. Αυτό γίνεται με το κουμπί “Save JSON” (περιοχή 2).

### 3.4.6 Εμφάνιση της εφαρμογής σε κινητές συσκευές

Η εφαρμογή μπορεί να εκτελεστεί από οποιαδήποτε πρόγραμμα περιήγησης που υποστηρίζει JavaScript. Εκτός από τους σταθερούς και φορητούς υπολογιστές, η δυνατότητα αυτή υπάρχει και στους browsers των σύγχρονων κινητών συσκευών. Έτσι, ιδιαίτερη έμφαση δόθηκε στην σωστή εμφάνιση του γραφικού περιβάλλοντος σε μικρότερες οθόνες υπολογιστών, όπως smartphones (“έξυπνων” τηλεφώνων) και tablets (υπολογιστών “ταμπλέτας”).



Εικόνα 3.21: (α) Εμφάνιση της καρτέλας δυαδικής μετατροπής σε κινητές συσκευές, (β) τοποθέτηση των κουμπιών του Toolbar σε side-panel (πλαϊνό πτυσσόμενο μενού).

**Digital Image Analysis Toolkit**

Segmentation

blockdistance\_a: 1  
overlapping\_c: 0.7  
HeightRatio\_th: 3.5

ARLSA Word Segmentation: 1  
overlapping\_c: 0.7  
HeightRatio\_th: 3.5

Remove Punctuation Marks  xRLSA

remove all Segments

Precision Recall Evaluation: 0.5

Draw Word  Draw Line  full Screen

**Document:** doc1.jpg

**Digital Image Analysis Toolkit**

multiple Processing

+ Select Text Documents Otsu

Sauvola GPP

ARLSA Words Segmentation

sauvola parameters

gpp parameters

ARLSA parameters

evaluation

**2 Text Documents selected**

doc1.jpg

doc2.jpeg

(α)

(β)

**Digital Image Analysis Toolkit**

multiple Processing extracted data

Select Text Document: fdgv.jpg

Search \_\_\_\_\_

x	y	Width	Height	Word
29	25	42	31	a
257	30	242	27	s
587	40	34	31	d
793	42	48	16	f
62	46	80	9	g

Items per page: 5   1 - 5 of 114

(γ)

Εικόνα 3.22: Εμφάνιση της εφαρμογής σε κινητές συσκευές. (α) καρτέλα κατάτμησης, (β) καρτέλα μαζικής επεξεργασίας, (γ) καρτέλα εμφάνισης δεδομένων.

## 4. Συμπεράσματα

Στη παρούσα διπλωματική μελετήθηκε η επεξεργασία ψηφιοποιημένων εγγράφων σε διαδικτυακό περιβάλλον. Συγκεκριμένα, αναπτύχθηκε μια client side εφαρμογή διαδικτύου (web app), η οποία παρέχει δυνατότητες δυαδικοποίησης και κατάτμησης εικόνων εγγράφου. Η εφαρμογή υποστηρίζει τρεις τεχνικές δυαδικής μετατροπής (Otsu [1], Sauvola [2], GPP [3]) καθώς και μία τεχνική τμηματοποίησης (εξαγωγής) λέξεων ή γραμμών του κειμένου (ARLSA [6]). Παράλληλα, επιτρέπει και την αξιολόγηση των παραπάνω μεθοδολογιών με την εισαγωγή Ground Truth αρχείων και τον υπολογισμό μετρικών μεγεθών αξιολόγησης.

Συνοπτικά, η μέθοδος Otsu [1], όντας καθολική μέθοδος κατωφλίωσης, πετυχαίνει καλή δυαδική μετατροπή μόνο σε εικόνες με όμοια αντίθεση στο υπόβαθρο. Αντίθετα, οι μέθοδοι Sauvola [2] και GPP [3] εκμεταλλεύονται τη τοπική πληροφορία του κειμένου και δίνουν αρκετά ικανοποιητικά αποτελέσματα, αλλά είναι δύσκολο να διαχειριστούν αυτόματα τη μεταβολή της φωτεινότητας τοπικά στο έγγραφο χωρίς τη χειροκίνητη ρύθμιση των παραμέτρων τους από το χρήστη. Επίσης, το αποτέλεσμα των τεχνικών δυαδικοποίησης επηρεάζει και την απόδοση των αλγορίθμων κατάτμησης. Όσον αφορά τη τεχνική τμηματοποίησης ARLSA [6], ανάλογα το είδος του εγγράφου, μπορεί να επιτύχει αποτελεσματική εξαγωγή των λέξεων ή γραμμών του κειμένου. Η επιτυχία της συγκεκριμένης τεχνικής βασίζεται στον αποκλεισμό της σύνδεσης των αντικειμένων, τα οποία δεν παρουσιάζουν ομοιότητες ως προς το ύψος, το πλάτος, και την επικάλυψη σε μια συγκεκριμένη διεύθυνση.

Ακόμα, συμπεραίνουμε πως χάρη στη ραγδαία ανάπτυξη και βελτίωση των σύγχρονων προγραμμάτων περιήγησης στο διαδίκτυο (web browsers) και της γλώσσας προγραμματισμού JavaScript μαζί με τα frameworks και τα APIs που την υποστηρίζουν, η επεξεργασία και ανάλυση εικόνων εγγράφου σε διαδικτυακό περιβάλλον κρίνεται απολύτως εφικτή και χρήσιμη διαδικασία, η οποία “απολαμβάνει” όλα τα πλεονεκτήματα μιας client side web εφαρμογής.

Ορισμένες μελλοντικές βελτιώσεις που μπορούν να γίνουν στην διαδικτυακή εφαρμογή είναι:

- Προσθήκη επιπρόσθετων τεχνικών δυαδικοποίησης και κατάτμησης της υπάρχουσας βιβλιογραφίας. Έτσι, ο χρήστης θα μπορεί να δοκιμάσει και να διαλέξει τη τεχνική που δίνει το βέλτιστο δυαδικό αποτέλεσμα ή τη βέλτιστη εξαγωγή του κειμένου. Επίσης οποιοσδήποτε ερευνητής έχει αναπτύξει τη δικιά του μέθοδο θα έχει τη δυνατότητα να τη συγκρίνει εύκολα με τις προσεγγίσεις της διεθνής βιβλιογραφίας που θα υλοποιηθούν.

- Προσθήκη περισσότερων μετρικών αξιολόγησης, ώστε να προκύψει πιο αντικειμενική αποτίμηση των αποτελεσμάτων της κατωφλίωσης και τμηματοποίησης περιοχών του κειμένου.
- Δημιουργία μιας διαδικτυακής βάσης δεδομένων στην εφαρμογή με εικόνες εγγράφου μαζί με τα αντίστοιχα Ground-Truth(GT) αρχεία τους (GT δυαδικές εικόνες και GT segments), ώστε να καταστεί ευκολότερη η αντικειμενική αξιολόγηση κάθε υλοποιημένης μεθοδολογίας.
- Βελτίωση της ταχύτητας και απόδοσης της εφαρμογής, υποστηρίζοντας πιο εξειδικευμένο υλικό (hardware) του υπολογιστή όπως πολυπύρηνου επεξεργαστή και κάρτας γραφικών, αν είναι διαθέσιμα.
- Μεγαλύτερη υποστήριξη κινητών συσκευών smartphones και tablets. Οι συσκευές αυτές, πέρα από την γρήγορη πρόσβαση στο διαδίκτυο, έχουν και το πλεονέκτημα της εύκολης σάρωσης και ψηφιοποίησης εικόνων, μέσω της κάμερας που διαθέτουν. Ωστόσο, στα περισσότερα κινητά η ανάλυση της φωτογραφικής τους μηχανής κρίνεται μέτρια. Συνεπώς, η άμεση βελτίωση της ποιότητας των εγγράφων μέσω δυαδικής μετατροπής μπορεί να φανεί χρήσιμη.
- Δημιουργία ξεχωριστού API της εφαρμογής, ώστε οι υλοποιημένες μεθοδολογίες να μπορούν να εισαχθούν ως βιβλιοθήκες και να χρησιμοποιηθούν από άλλες εφαρμογές διαδικτύου.

Τέλος, ο κώδικας της εφαρμογής μπορεί να γίνει open source ώστε τα μέλη τις ακαδημαϊκής κοινότητας ή ανεξάρτητοι ερευνητές να βελτιώσουν τις ήδη υπάρχουσες μεθοδολογίες, να προτείνουν καινούργιες, να υλοποιήσουν και να προσθέσουν επιπλέον δικίες τους τεχνικές και δυνατότητες. Με αυτό το τρόπο θα διευκολυνθεί η περεταίρω ανάπτυξη και χρησιμότητα της εφαρμογής ως ερευνητικό έργο.

## Αναφορές

- [1] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Systems, Man, and Cybernetics*, vol. 9, pp. 62-66, 1979.
- [2] J. Sauvola and M. Pietikainen, "Adaptive document image binarization," *Pattern Recognition*, vol. 33, no. 2, pp. 225-236, February 2000.
- [3] B. Gatos, I. Pratikakis and S. J. Perantonis, "Adaptive degraded document image binarization," *Pattern Recognition*, vol. 39, no. 3, pp. 317-327, March 2006.
- [4] I. Pratikakis, K. Zagoris, G. Barlas and B. Gatos, "ICDAR2017 Competition on Document Image Binarization (DIBCO 2017)," in *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, Kyoto, Japan, Nov. 2017.
- [5] I. Pratikakis, K. Zagoris, B. Gatos and P. Kaddas, "ICFHR 2018 Competition on Handwritten Document Image Binarization (H-DIBCO 2018)," in *2018 16th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, Niagara Falls, NY, USA, Aug. 2018.
- [6] 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, pp. 590-604, April 2010.
- [7] I. Pratikakis, K. Zagoris, B. Gatos, G. Louloudis and N. Stamatopoulos, "ICFHR 2014 Competition on Handwritten Keyword Spotting (H-KWS 2014)," in *2014 14th International Conference on Frontiers in Handwriting Recognition*, Heraklion, Greece, Sept. 2014.
- [8] Konzil, "Konzilsprotokolle 1795 - 1796," [Online]. Available: [http://www.digitale-bibliothek-mv.de/viewer/image/PPNUAG\\_0\\_1\\_St\\_647/1/#topDocAnchor](http://www.digitale-bibliothek-mv.de/viewer/image/PPNUAG_0_1_St_647/1/#topDocAnchor).
- [9] «Το πλαίσιο λογισμικού Angular,» [Ηλεκτρονικό]. Available: <https://angular.io/>.
- [10] «Το Canvas API της JavaScript,» [Ηλεκτρονικό]. Available: [https://developer.mozilla.org/en-US/docs/Web/API/Canvas\\_API](https://developer.mozilla.org/en-US/docs/Web/API/Canvas_API).
- [11] «Η βιβλιοθήκη FabricJS,» [Ηλεκτρονικό]. Available: <http://fabricjs.com/>.
- [12] «Το Web Workers API της JavaScript,» [Ηλεκτρονικό]. Available: [https://developer.mozilla.org/en-US/docs/Web/API/Web\\_Workers\\_API](https://developer.mozilla.org/en-US/docs/Web/API/Web_Workers_API).
- [13] «Η βιβλιοθήκη λογισμικού Angular Material,» [Ηλεκτρονικό]. Available: <https://material.angular.io/>.
- [14] «Το FileReader API της JavaScript,» [Ηλεκτρονικό]. Available: <https://developer.mozilla.org/en-US/docs/Web/API/FileReader>.

[15] «Το Web Storage API της JavaScript,» [Ηλεκτρονικό]. Available:  
[https://developer.mozilla.org/en-US/docs/Web/API/Web\\_Storage\\_API](https://developer.mozilla.org/en-US/docs/Web/API/Web_Storage_API).