2η Εργασία Επεξεργασίας και Ανάλυσης Εικόνας

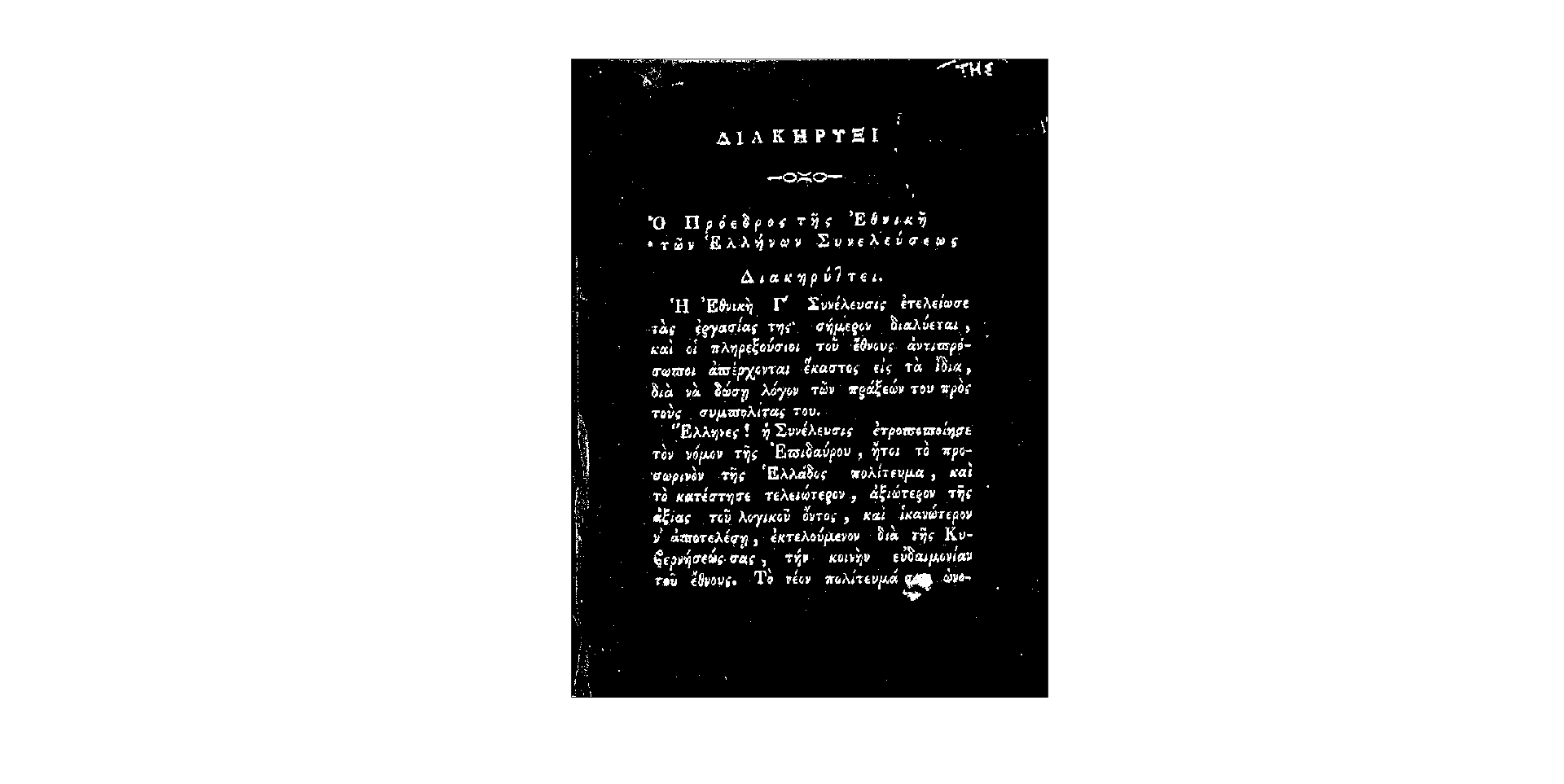
## Περιγραφή

Η 2η εργασία για το μάθημα της επεξεργασίας και ανάλυσης εικόνας αφορά την υλοποίηση ενός προγράμματος σε Matlab το οποίο θα λαμβάνει μια εικόνα κειμένου η οποία θα μετατρέπεται σε δυαδική. Ύστερα θα αφαιρούνται τυχόν θόρυβος και θα εντοπίζονται με ημιαυτόματο τρόπο τις λέξεις του κειμένου χρησιμοποιώντας διάφορες τεχνικές μορφολογικής επεξεργασίας εικόνων. Τέλος θα γίνει αποτίμηση της μεθόδου αυτής με χρήση ου Intersection Over Union όπου θα γίνει σύγκριση των εντοπισμένων λέξεων με τις πραγματικές λέξεις του κειμένου.

## Α’ Μέρος

Στο 1ο μέρος της εργασίας αυτής διαβάζεται η εικόνα που έχει δοθεί και μετατρέπεται σε grayscale χρησιμοποιώντας την συνάρτηση any\_image\_to\_grayscale\_func.m που μας δίνεται. Με την μέθοδο του Otsu μετατρέπουμε την εικόνα σε δυαδική. Στην συνέχεια ακολουθούν μια σειρά από μορφολογικές πράξεις ώστε να γίνει απαλοιφή μεγάλων συνδεδεμένων στοιχείων, όπως η σφραγίδα, και του θορύβου που έχει η εικόνα.

Αρχικά γίνεται διαστολή ώστε να ενωθούν τα γράμματα που βρίσκονται πιο κοντά μεταξύ τους και να σχηματίσουν την περιοχή της λέξης που ανήκουν με την συνάρτηση imdilate() και μετά βρίσκουμε τις στατιστικές πληροφορίες από τις περιοχές αυτές. Με την συνάρτηση labeled\_img() δίνουμε ένα label σε κάθε περιοχή και με την regionprops() παίρνουμε χρήσιμα χαρακτηριστικά όπως το εμβαδόν, το κέντρο και την περίμετρό του. Υπολογίζεται το μέσο εμβαδόν των περιοχών αυτών και η τυπική απόκλιση και βρίσκουμε τις περιοχές που θεωρούνται θόρυβος. Αυτές τις περιοχές τους δίνουμε την τιμή 0. Με αυτή την διαδικασία αφαιρείται μεγάλο κομμάτι της σφραγίδας καθώς η σφραγίδα σχηματίζει μια πιο μεγάλη περιοχή συγκριτικά με τις υπόλοιπες.



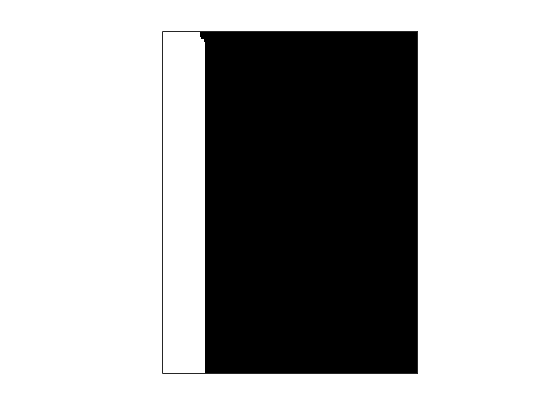
Εικόνα 1 Gamma=1

Για τον υπόλοιπο θόρυβο που υπάρχει ακολουθούνται δύο μέθοδοι. Πρώτα με την διαστολή των γραμμάτων οριζοντίως και ύστερα με την συστολή, επιτυγχάνεται να μείνουν μόνο ένα κομμάτι θορύβου στα αριστερά της εικόνας.



Εικόνα 2 Περιοχές όπου βρίσκεται θόρυβος

Επειδή στη αριστερή πλευρά δεν έχουμε δεδομένα που χρειαζόμαστε, γεμίζουμε την περιοχή μέχρι το κείμενο με την βοήθεια της διαστολής έτσι ώστε να καλύψουμε όλη την περιοχή που υπάρχει θόρυβος.



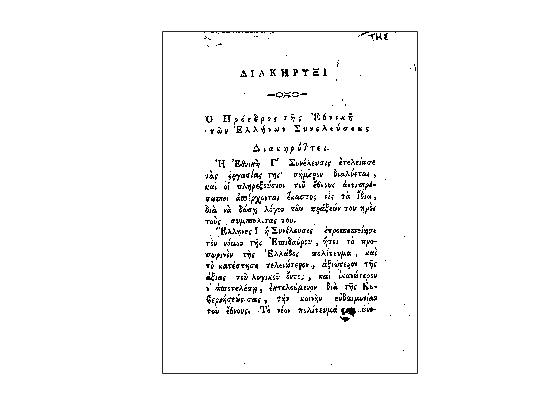
Εικόνα 3 Περιοχή θορύβου

Για να γίνει η αφαίρεση του θορύβου αρχικά εντοπίζονται όλα τα σημεία που υπάρχει με την λογική πράξη and μεταξύ της αρχικής εικόνας και της περιοχής που έχουμε υπολογίσει. Η πράξη αυτή επιστρέφει σημεία που έιναι και στις δυο εικόνες άσπρα (1).

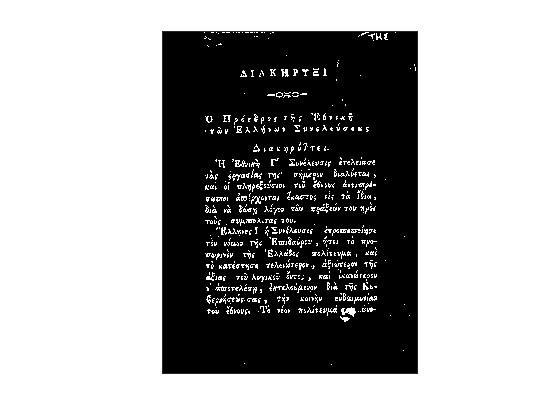


Εικόνα 4 Θόρυβος στην αριστερή πλευρά

Τέλος με την πράξη xor μεταξύ της αρχικής εικόνας και την άρνηση της εικόνας με τον θόρυβο, επιστρέφονται Pixel με τιμή 1 όταν έχει ,είτε στην μια εικόνα είτε στην άλλη, 1 ως τιμή, αλλιώς επιστρέφει 0.

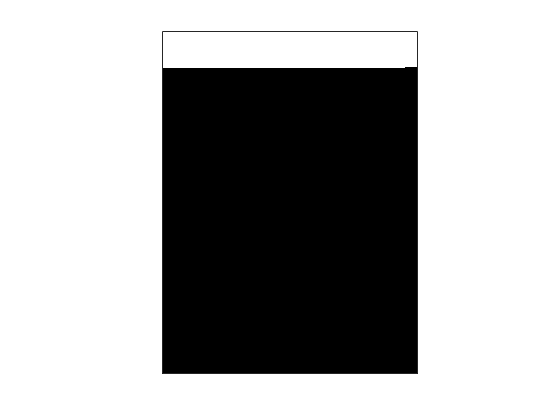


Εικόνα 5 Αποτέλεσμα xor

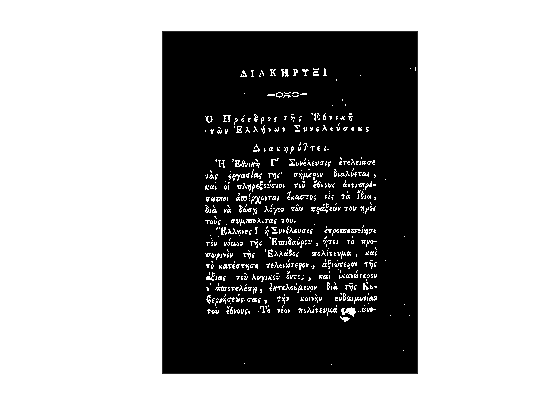


Εικόνα 6 Άρνηση εικόνας

Με τον ίδιο τρόπο εντοπίζεται η περιοχή του θορύβου στο πάνω μέρος της εικόνας, διατηρώντας μόνο τον θόρυβο στην εικόνα με τη συστολή και μετά με διαστολή των pixels που έχουν μείνει και με την λογική πράξη and της εικόνας με την άρνηση της εικόνας με τον θόρυβο αφαιρείται ο θόρυβος.



Εικόνα 7 Περιοχή θορύβου στο πάνω μέρος της εικόνας



Εικόνα 8 Αποτέλεσμα εικόνας με την and

Τέλος, για να αφαιρεθεί και ο θόρυβος στην υπόλοιπη εικόνα, πρώτα εφαρμόζουμε συστολή και μετά διαστολή για να μείνει μόνο το κείμενο.



Εικόνα 9 Διατήρηση περιοχής κειμένου



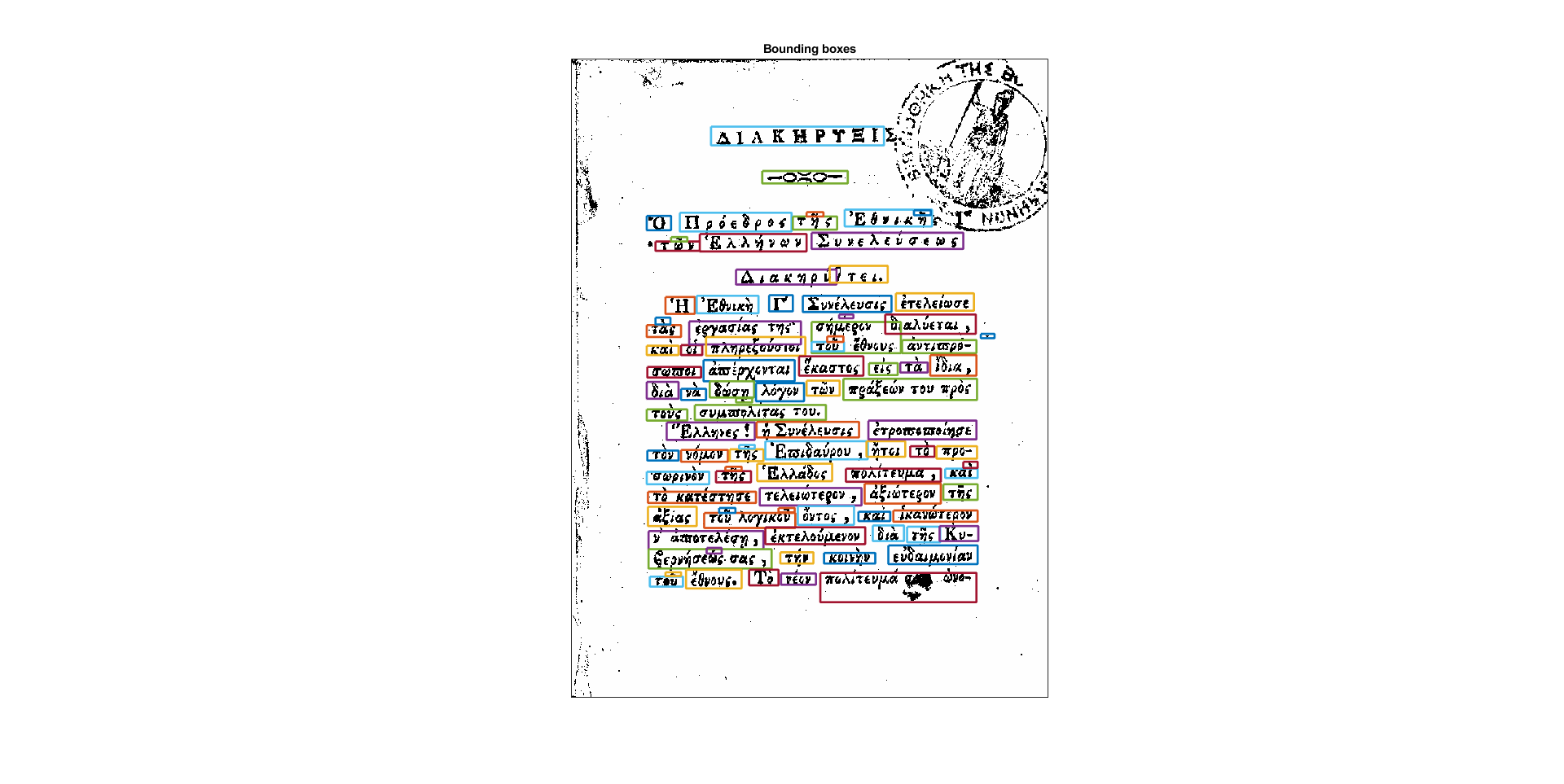
Εικόνα 10 Τελική εικόνα χωρίς θόρυβο και σφραγίδα

Αφότου έχει καθαριστεί η εικόνα από θόρυβο, με την διαστολή των γραμμάτων σχηματίζουμε πάλι περιοχές που σχηματίζουν τις περιοχές που βρίσκεται κάθε λέξη και με την bwlabel() τους δίνουμε μια ετικέτα με συνδεσιμότητα 4 σημείων. Η διαστολή δεν θα πρέπει να είναι μεγάλη διότι οι περιοχές των λέξεων θα καλύπτουν κι άλλες λέξεις με αποτέλεσμα να μην έχουμε τον διαχωρισμό που αναζητάμε αργότερα. Οι περιοχές αφότου αποκτήσουν Label, εμφανίζονται όπως στην παρακάτω εικόνα.



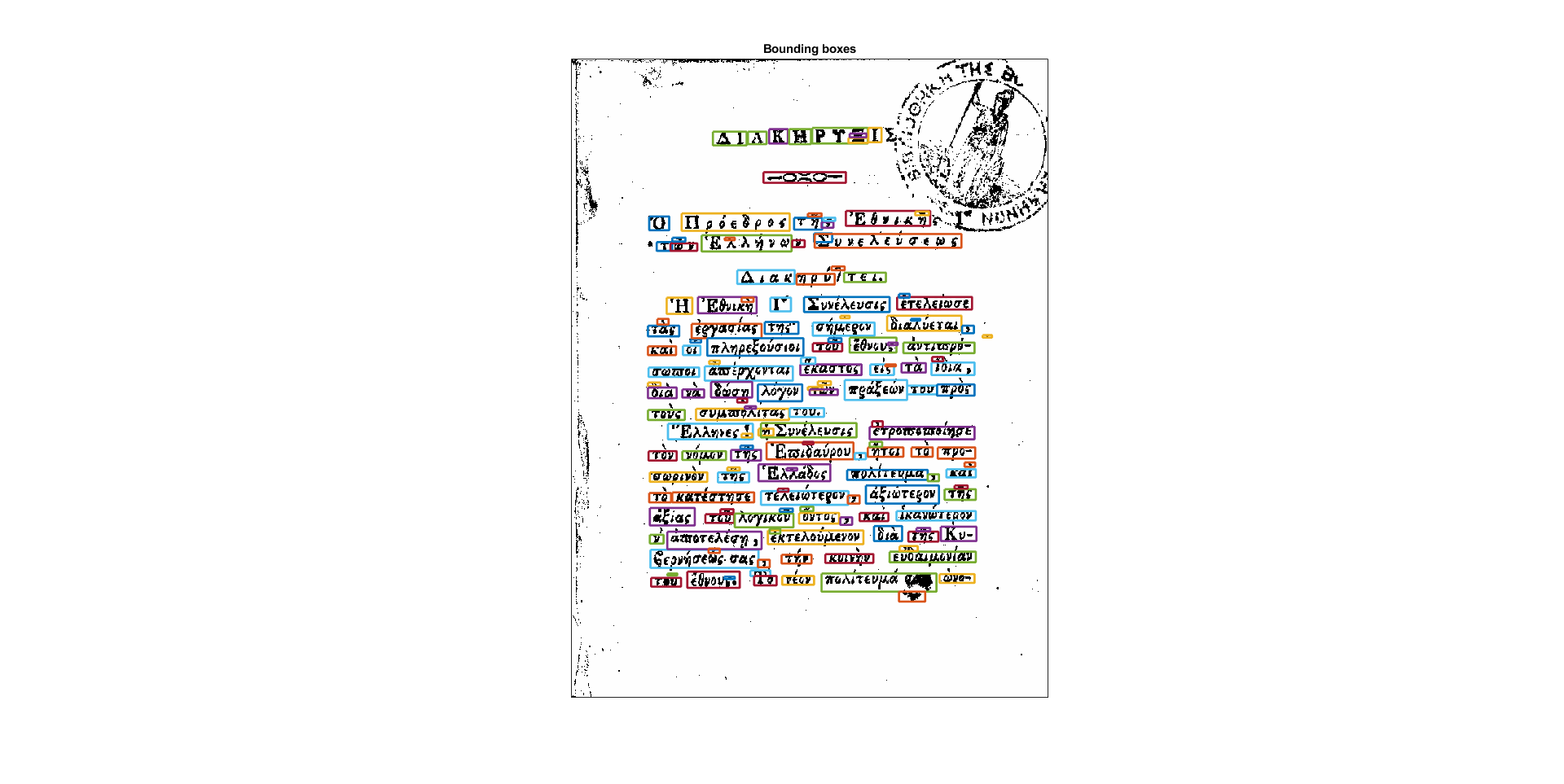
Εικόνα 11 Αποτέλεσμα με 92 περιοχές ξεχωριστά

Στην περίπτωση της συνδεσιμότητας 4 σημείων εντοπίζονται 92 στοιχεία ενώ αν δοκιμάσουμε με συνδεσιμότητα 8 σημείων εντοπίζονται 89 στοιχεία. Η διαφορά αυτή υπάρχει καθώς στην συνδεσιμότητας με 8 σημεία εξετάζονται περισσότερα Pixel με αποτέλεσμα τα στοιχεία που βρίσκονται πιο κοντά μεταξύ τους να δημιουργούν ένα μεγαλύτερο στοιχείο που τα καλύπτει.



Εικόνα 12 Τελικό αποτέλεσμα

Εάν θελήσουμε να βρούμε και τα σημεία στίξης πέρα από τις λέξεις θα πρέπει να αλλάξουμε την διαστολή με άνοιγμα πρώτα και μετά διαστολή. Με αυτόν τον τρόπο διαχωρίζονται περισσότερο ου σύνδεσμοι των γραμμάτων και διαγράφονται λεπτές προεξοχές. Με τα σημεία στίξης να απέχουν πλέον περισσότερο από πριν από τα γράμματα, μπορούμε με την διαστολή να βρούμε τις λέξεις οριζοντίως αλλά κάποια σημεία στίξης δημιουργούν ξεχωριστές περιοχές.



Εικόνα 13 Εντοπισμός σημείων στίξης

Τα όρια από τα κουτιά των λέξεων αποθηκεύονται στο αρχείο results.txt ώστε να συγκριθούν με το ground truth.

## Β’ Μέρος

Στο δεύτερο μέρος της εργασίας γίνεται η αποτίμηση της μεθόδου εντοπισμού των λέξεων με την χρήση του Intersection Over Union με βάση το οποίο θα συγκριθούν οι εντοπισμένες λέξεις με τις πραγματικές λέξεις του κειμένου. Η αποτίμηση θα πραγματοποιηθεί για διάφορους συνδυασμούς των τιμών της βελτίωσης της grayscale εικόνας και τις τιμές κατωφλίου για την IOU.

Η μέθοδος IOU υπολογίζεται με την τομή δύο ορθογωνίων προς την ένωσή τους και μετράει αν τα δύο αυτά ορθογώνια ταυτίζονται πλήρως (τιμή 1) ή δεν υπάρχει καθόλου επικάλυψη (τιμή 0). Τα ορθογώνια αυτά είναι τα πλαίσια των λέξεων που έχουν εντοπιστεί με τις μορφολογικές πράξεις που έγιναν στο Α’ μέρος και τα πλαίσια των λέξεων που χρησιμοποιούμαι ως ground truth.

Έχοντας βρει τον πίνακα με τις τιμές IOU, εφαρμόζουμε ένα κατώφλι ώστε να κρατήσουμε τιμές που βρίσκονται πάνω από αυτή την τιμή και δημιουργούμε έναν πίνακα ίδιων διαστάσεων με τιμές 0-1. Κάθε στήλη είναι οι λέξεις που έχουμε ως ground truth και κάθε γραμμή είναι οι λέξεις που έχουμε εντοπίσει προηγουμένως.

Με βάση αυτόν τον πίνακα (πίνακας σύγχυσης) υπολογίζουμε τις τιμές tp και fn (true positive, false negative) εξετάζοντάς τον κάθετα. Εάν στην στήλη συναντάμε «1», η διαδικασία εξέτασης εκείνης την στήλης σταματάει και η τιμή tp αυξάνεται κατά ένα. Στο τέλος, από το σύνολο των στηλών αφαιρούμε το πλήθος των στηλών που περιέχουν έστω μια φορά «1» και βρίσκουμε την τιμή fn.

Με την ίδια διαδικασία εξετάζουμε τον πίνακα οριζοντίως για να υπολογίσουμε την τιμή fp (false positive). Βρίσκουμε τις γραμμές που περιέχουν έστω ένα «1» και αφαιρούμε του πλήθος τους από τις συνολικές γραμμές για να βρούμε την fp τιμή.

Έχοντας αυτές τις τιμές υπολογισμένες, μπορούμε να υπολογίσουμε το Recall () και το Precision (). H τιμή F-measure υπολογίζεται ως εξής:

Δοκιμάζοντας για διάφορες τιμές βελτίωσης της grayscale εικόνας (gamma value) και διαφορετικές τιμές του IOU Thresholdοδηγούμαστε στα αποτελέσματα του παρακάτω πίνακα.

|  |  |  |  |
| --- | --- | --- | --- |
| Πίνακας αποτίμησης F-Measure | | | |
|  | IOU threshold | | |
| Gamma value | 0.3 | 0.5 | 0.7 |
| 0.6 | 0.84 | 0.71 | 0.31 |
| 0.8 | 0.83 | 0.73 | 0.31 |
| 1.0 | 0.82 | 0.72 | 0.30 |
| 1.2 | 0.76 | 0.63 | 0.28 |
| 1.4 | 0.71 | 0.55 | 0.22 |

Παρατηρώντας τα αποτελέσματα, οδηγούμαστε στο συμπέρασμα ότι από τις αλλαγές της τιμής gamma, η IOU threshold=0.5 επηρεάζεται πολύ περισσότερο (διαφορά 0.16) από τις 0.3 και 0.7. Το καλύτερο ποσοστό υπολογίζεται με IOU threshold= 0.3. Οι τιμές συνολικά που έχει ο πίνακας IOU δεν ξεπερνάνε το 0.7 σαν threshold σε αρκετά σημεία με αποτέλεσμα οι τιμές του F-measure να είναι μικρές και να μην διαφέρουν όσο και να μεταβάλλουμε το gamma.

Αυξάνοντας το gamma της εικόνας χάνονται τμήματα αυτής κατά την διάρκεια της μορφολογικής επεξεργασίας και χάνεται πληροφορία. Η πληροφορία που χάνεται (αν αυξάνουμε το gamma) ή αποκτάται (αν μειώνουμε το gamma), έχει σημαντικό ρόλο στα αποτελέσματα του IOU.