

## 2<sup>η</sup> ΕΡΓΑΣΙΑ

Θέμα: “Τμηματοποίηση εικόνας κειμένου σε λέξεις”

### Α' Μέρος - Μέθοδος

#### Περίληψη

Να γραφτεί πρόγραμμα σε Octave/Matlab το οποίο να λαμβάνει σαν είσοδο μια εικόνα κειμένου η οποία θα μετατρέπεται σε δυαδική. Στην συνέχεια, να εντοπίζει με (ημι)αυτόματο τρόπο τις λέξεις του κειμένου χρησιμοποιώντας διάφορες τεχνικές μορφολογικής επεξεργασίας εικόνων. Η μορφολογική επεξεργασία θα έχει σαν σκοπό:

- i) Την αφαίρεση τυχόν θορύβου, δηλαδή διάσπαρτων pixels που δεν ανήκουν σε κάποιο χαρακτήρα ή σημείο στίξης όπως επίσης και μεγάλων τμημάτων της εικόνας που δεν αντιστοιχούν σε λέξεις.
- ii) Την κατάτμηση του κειμένου σε λέξεις (word segmentation).

#### Αναλυτικά

Σας δίνεται ένα αρχείο **ergasia2\_main\_start.m** που περιέχει την δομή του κυρίως προγράμματος. Ορισμένα βήματα είναι ήδη **έτοιμα** και σε όλα υπάρχουν σχόλια. Μετονομάστε το αρχείο **ergasia2\_main\_start.m** σε **ergasia2\_main.m** και συνεχίστε με τα παρακάτω βήματα.

Βήμα A1) Διαβάστε και εμφανίστε την εικόνα κειμένου που σας δίνεται, όπως στην Εικόνα 1.

Βήμα A2) Εάν η εικόνα είναι έγχρωμη, τότε μετατρέψτε την σε grayscale και εμφανίστε την σε ένα καινούργιο γράφημα, όπως στην Εικόνα 2 (δεν διαφέρει σημαντικά από την εικόνα 1). Για την μετατροπή χρησιμοποιείτε την συνάρτηση **any\_image\_to\_grayscale\_func.m** που σας δίνεται.

Βήμα A3) Μετατρέψτε την grayscale εικόνα σε δυαδική χρησιμοποιώντας την εντολή **graythresh** του Octave/Matlab η οποία βασίζεται στην μέθοδο του **Otsu**. Εμφανίστε την δυαδική εικόνα σε ένα καινούργιο γράφημα, όπως στην Εικόνα 3.

ΠΑΡΑΤΗΡΗΣΗ: για την απεικόνιση μιας δυαδικής εικόνας κειμένου BW είναι σκόπιμο να απεικονίζεται η **άρνηση** της (~BW) ώστε τα γράμματα να εμφανίζονται μαύρα και το φόντο άσπρο.

Βήμα A4) Προτείνετε μια ακολουθία από μορφολογικές πράξεις ώστε να απαλείψετε:

- τον **θόρυβο** στην δυαδική εικόνα, δηλαδή διάσπαρτα pixel που δεν ανήκουν σε κάποιο χαρακτήρα ή σημείο στίξης.
- **μεγάλα συνδεδεμένα στοιχεία** της εικόνας που δεν αντιστοιχούν σε λέξεις (όπως π.χ. η σφραγίδα που βρίσκεται στο άνω δεξιό τμήμα της εικόνας).

Για τον σκοπό αυτό μπορείτε να χρησιμοποιήσετε διάφορες εντολές του Octave/Matlab όπως τις

**imdilate**

imerode  
imopen  
imclose  
regionprops  
bwlabel  
bwmorph  
scatterhist  
imclearborder

Μπορείτε επίσης να χρησιμοποιήσετε τις συνήθεις δυαδικές πράξεις **and**, **or**, **xor** του Octave/Matlab μεταξύ δύο δυαδικών εικόνων. Ανατρέξτε στην βοήθεια του Octave/Matlab για περισσότερες πληροφορίες σχετικά με τις παραπάνω συναρτήσεις. Για παράδειγμα, το αποτέλεσμα της διαδικασίας αφαίρεσης της σφραγίδας από την εικόνα δίνει μια δυαδική εικόνα όπως αυτή στην Εικόνα 4.

ΣΗΜΕΙΩΣΗ: Το κατώφλι μεγέθους σχετικά με το ποια στοιχεία της εικόνας είναι «πολύ μεγάλα» μπορείτε να το προσδιορίσετε βασιζόμενοι σε **στατιστικές μετρήσεις** π.χ. μέση τιμή και τυπική απόκλιση του εμβαδού των στοιχείων (connected components) της εικόνας.

**ΕΡΩΤΗΣΗ Α1:** Με ποιο σκεπτικό και ποια σειρά επιλέξατε τις διάφορες μορφολογικές και λογικές πράξεις στο βήμα Α4 με σκοπό τον καθαρισμό της εικόνας και την αφαίρεση της σφραγίδας; Αιτιολογήστε αναλυτικά την απάντησή σας.

Βήμα Α5) Προτείνετε μια ακολουθία από μορφολογικές πράξεις (όπως στο βήμα Α4 παραπάνω) ώστε να συνενωθούν τα συνδεδεμένα στοιχεία σε μεγαλύτερα με σκοπό κάθε συνδεδεμένο στοιχείο να αντιστοιχεί (όσο αυτό είναι δυνατό) σε κάποια **λέξη** του κειμένου. Έπειτα, εφαρμόστε τις **bwlabel** και **label2rgb** ώστε να εμφανίσετε με διαφορετικό χρώμα το κάθε συνδεδεμένο στοιχείο, όπως στην Εικόνα 5. Για τον σκοπό αυτό χρησιμοποιήστε την εντολή **bwlabel** του OCTAVE/MATLAB με συνδεσιμότητα 4 σημείων (4-connectivity).

ΣΗΜΕΙΩΣΗ: Ένα απλό παράδειγμα χρήσης των συναρτήσεων **bwlabel** και **label2rgb** δίνεται στο αρχείο **bwlabel\_demo.m**. Παρατηρήστε ότι η εντολή **bwlabel** επιστρέφει έναν πίνακα ιδίων διαστάσεων με την δυαδική εικόνα. Τα pixel του  $L$ -οστού στοιχείου (component) έχουν τιμή  $L$ , με  $1 \leq L \leq Count$ , όπου  $Count$  το πλήθος των συνδεδεμένων στοιχείων της εικόνας.

**ΕΡΩΤΗΣΗ Α2:** Με ποιο σκεπτικό και ποια σειρά επιλέξατε τις διάφορες μορφολογικές και λογικές πράξεις στο βήμα 5 με σκοπό τον προσδιορισμό των λέξεων; Αιτιολογήστε την απάντησή σας.

**ΕΡΩΤΗΣΗ Α3:** Πόσα συνολικά στοιχεία (components) βρέθηκαν στην εικόνα; Αν χρησιμοποιήσετε συνδεσιμότητα 8 σημείων (8-connectivity) σαν δεύτερη παράμετρο στην **bwlabel** πόσα συνολικά components δημιουργούνται; Πού οφείλεται η διαφορά στις δύο αυτές τιμές; Ποιος τύπος συνδεσιμότητας είναι καλύτερος κατά την γνώμη σας και γιατί;

Βήμα Α6) Για κάθε συνδεδεμένο στοιχείο που αντιστοιχεί σε κάποια λέξη (ή σε κάποιο σημείο στίξης) μπορούμε να ορίσουμε ένα **περιγεγραμμένο ορθογώνιο** με βάση το εύρος συντεταγμένων που καταλαμβάνει αυτό το στοιχείο στην  $x$ -διάσταση [ $xmin$  έως  $xmax$ ] όσο και στην  $y$ -διάσταση [ $ymin$  έως  $ymax$ ]

ΣΗΜΕΙΩΣΗ: Δείτε πώς υπολογίζονται τα [ $xmin$   $ymin$   $xmax$   $ymax$ ] και πώς εμφανίζεται το περιγεγραμμένο ορθογώνιο στο συνοδευτικό αρχείο **bwlabel\_demo.m**.

Σαν τελικό αποτέλεσμα εμφανίστε την αρχική εικόνα του κειμένου (Εικόνα 1) και πάνω της σχεδιάστε τα ορθογώνια των λέξεων που εντοπίσατε, όπως φαίνεται στην Εικόνα 6.

**ΕΡΩΤΗΣΗ 4:** Προτείνετε μια παραλλαγή στην ακολουθία που χρησιμοποιείτε στο βήμα 6 ώστε το ορθογώνιο που περιγράφει κάθε λέξη να συμπεριλαμβάνει (κατά το δυνατόν) και τα **σημεία στίξης** της εκάστοτε λέξης (τόνο, περισπωμένη κλπ). Τι αλλαγή κάνατε;

Βήμα A7) Τα ορθογώνια που αντιστοιχούν στις λέξεις που έχετε εντοπίσει αποθηκεύστε τα σε ένα αρχείο κειμένου «**results.txt**». Το αρχείο θα έχει τόσες γραμμές όσες είναι οι λέξεις που εντοπίστηκαν και σε κάθε γραμμή θα έχει τις συντεταγμένες του αντίστοιχου ορθογωνίου [ $x_{min}$   $y_{min}$   $x_{max}$   $y_{max}$ ]. Συνεπώς, εάν έχουν εντοπισθεί  $N$  λέξεις, τότε στο αρχείο θα έχει αποθηκευτεί ένας πίνακας τιμών  $N \times 4$ .

ΣΗΜΕΙΩΣΗ: Συμβουλευτείτε την εντολή **dlmwrite** για την εγγραφή πίνακα σε αρχείο κειμένου.

## Β' Μέρος - Αποτίμηση και συγκριτικές δοκιμές

### Περίληψη

Να γίνει αποτίμηση της μεθόδου εντοπισμού λέξεων με χρήση του Intersection Over Union (IOU) με βάση το οποίο θα συγκριθούν οι εντοπισμένες λέξεις με τις πραγματικές λέξεις του κειμένου.

Η αποτίμηση να πραγματοποιηθεί για διαφορετικές ζεύγη αρχικών συνθήκών που αφορούν την βελτίωση της grayscale εικόνας και τις τιμές κατωφλίου για την IOU.

ΠΑΡΑΤΗΡΗΣΗ: Στο Β' μέρος θεωρήστε ότι τα βήματα A4 και A5 του Α' μέρους παραμένουν αμετάβλητα. Δηλαδή, οι μορφολογικές πράξεις με βάση τις οποίες έγινε ο εντοπισμός των λέξεων δεν θα τροποποιηθούν άλλο.

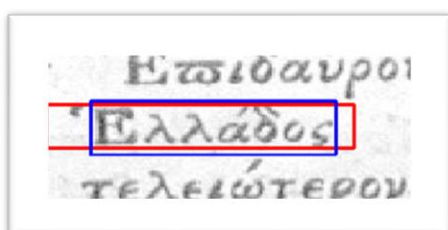
### Intersection Over Union

Το μέτρο Intersection Over Union μεταξύ δύο ορθογωνίων  $A$  και  $B$  όπου το πρώτο αντιστοιχεί στην εντοπισμένη θέση μια λέξης ενώ το δεύτερο στην πραγματική θέση της, υπολογίζεται ως:

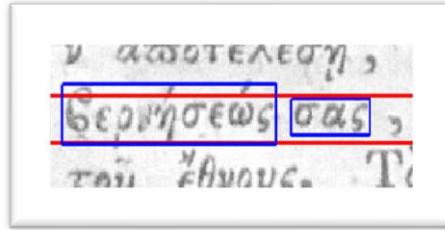
$$IOU = \frac{A \cap B}{A \cup B}$$

Οι τιμές που λαμβάνει το μέτρο είναι μεταξύ 0 (καθόλου επικάλυψη) μέχρι 1 (πλήρης ταύτιση).

Στο παρακάτω παράδειγμα το κόκκινο ορθογώνιο αντιστοιχεί στην εντοπισμένη θέση (method's bounding box) ενώ το μπλε ορθογώνιο στην πραγματική θέση της λέξης (ground truth bounding box). Το  $IOU$  είναι 0.726.



Υπάρχει περίπτωση κάποια προσδιορισμένη θέση να τέμνεται με περισσότερες από μια λέξεις του ground truth, όπως φαίνεται στο παρακάτω παράδειγμα.



Σε αυτή την περίπτωση η μέθοδός μας έδωσε μια εντοπισμένη λέξη (κόκκινο ορθογώνιο) που σχετίζεται με δύο πραγματικές λέξεις (μπλε ορθογώνια). Τα *IOU* είναι 0.532 και 0.172, αντίστοιχα.

ΣΗΜΕΙΩΣΗ: Περισσότερες πληροφορίες σχετικά με το *IOU* και την συσχέτισή του με την αποτίμηση της αποτελεσματικότητας ανάκλησης μπορείτε να βρείτε στο συνοδευτικό αρχείο **Αποτίμηση αποτελεσματικότητας.pdf** και να δείτε και τον σχετικό κώδικα στο **iou\_demo.m**.

## Αποτίμηση

Σας δίνεται ένα αρχείο **ergasia2\_evaluation\_start.m** που περιέχει την δομή του προγράμματος αποτίμησης. Ορισμένα βήματα είναι ήδη **έτοιμα** και σε όλα υπάρχουν σχόλια. Μετονομάστε το αρχείο **ergasia2\_evaluation\_start.m** σε **ergasia2\_evaluation.m** και συνεχίστε με τα παρακάτω βήματα.

Θέλουμε να αποτιμήσουμε το πώς επηρεάζει την συνολική απόδοση ο συνδυασμός του **gamma correction** της αρχικής grayscale εικόνας και του κατωφλίου  **$IOU_{Thres}$**  που θα εφαρμόσουμε στον πίνακα *IOU*. Για τον σκοπό αυτό δίνεται το αρχείο **Troizina 1827\_ground\_truth.txt** το οποίο περιέχει 87 ground truth λέξεις. Κάθε λέξη καθορίζεται από ένα ορθογώνιο που περιγράφεται με ένα διάνυσμα 4 τιμών [*gxmin*, *gymin*, *gxmax*, *gymax*] που αντιστοιχούν στην άνω αριστερή και την κάτω δεξιά γωνία του.

Βήμα B1) Διαβάστε το αρχείο των αποτελεσμάτων της μεθόδου **results.txt** που έχει δημιουργήσει το πρόγραμμα σας στο Α' μέρος βήμα A7 και υπολογίστε την πίνακα *IOU* που περιέχει την σύγκριση των λέξεων που βρέθηκαν με τις ground truth λέξεις.

ΣΗΜΕΙΩΣΗ: Ο υπολογισμός του πίνακα όλων των *IOU* γίνεται από την συνάρτηση **calcIOU.m** που σας δίνεται. Συμβουλευτείτε το παράδειγμα στο **iou\_demo.m** για περισσότερες πληροφορίες.

Βήμα B2) Συμπληρώστε τον παρακάτω πίνακα με τις τιμές αποτίμησης **F-Measure** για τα αντίστοιχα ζεύγη (gamma value – *IOU* threshold).

Πίνακας αποτίμησης <b>F-Measure</b>			
gamma value	IOU threshold		
	0.3	0.5	0.7
0.6			
0.8			
1.0			
1.2			
1.4			

Συγκεκριμένα, i) η τιμή gamma value καθορίζεται στην αρχή του αρχείου **ergasia2\_main\_start.m** που σας δίνεται (γραμμή 38) και ii) η τιμή IOU threshold καθορίζεται στο αρχείο **ergasia2\_evaluation\_start.m** που σας δίνεται (γραμμή 34).

**ΕΡΩΤΗΣΗ B1:** Παρουσιάστε και σχολιάστε τα αποτελέσματα που πίνακα. Ποιος συνδυασμός (gamma value – IOU threshold) δίνει το καλύτερο ποσοστό αποτίμησης **F-Measure**; Ποια από τις τρεις τιμές IOU threshold επηρεάζεται περισσότερο (έχει μεγαλύτερη διασπορά τιμών) από τις διάφορες τιμές gamma value; Αιτιολογείστε τις απαντήσεις σας.

ΣΗΜΕΙΩΣΗ: Ενδεικτικά, για τα bounding boxes που φαίνονται στην Εικόνα 6 και για gamma value = 1 και IOU threshold = 0.5 προκύπτει τιμή F-Measure περίπου 0.7. Αυτή η τιμή βέβαια θα διαφέρει στα δικά σας αποτελέσματα.

## Οδηγίες

- Το συνοδευτικό υλικό που αφορά την Εργασία (εικόνα, βοηθητικά αρχεία, ground truth κλπ.) μπορείτε να τα βρείτε στο φάκελο της εργασίας στο e-class.
- Τα αρχεία του Octave/Matlab θα πρέπει ΟΠΩΣΔΗΠΟΤΕ να περιέχουν και **σχόλια** για να είναι ευανάγνωστα.

## Παράδοση εργασίας

Ένα συμπιεσμένο αρχείο **ergasia.zip** (ή **.rar**) που να περιέχει

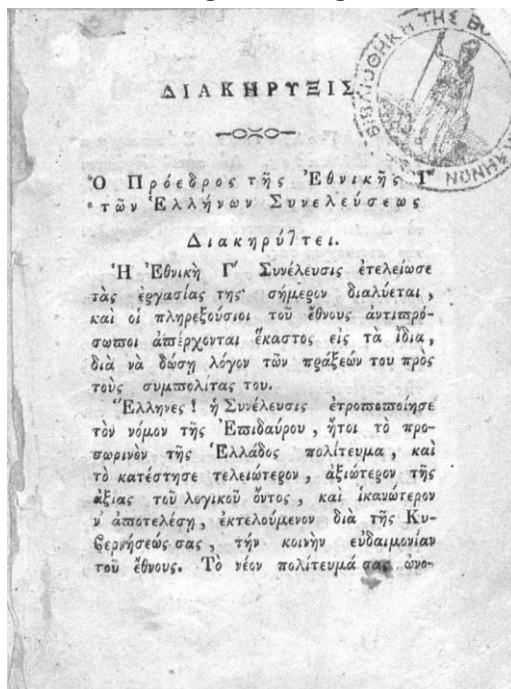
- τα αρχεία **ergasia2\_main.m** και **ergasia2\_evaluation.m** του Octave/Matlab (κώδικας)
- ένα αρχείο του WORD (παρουσίαση εργασίας + απαντήσεις στις ερωτήσεις).

Το συμπιεσμένο αρχείο να αποσταλεί ΜΟΝΟ μέσω eclass στην ενότητα Εργασίες μέχρι και την **Κυριακή 29/1/2023** στις 12:00 το βράδυ.

Για οποιαδήποτε διευκρίνιση μπορείτε να επικοινωνήσετε μαζί μου στο **akesidis@uniwa.gr** ή να αποστείλετε μήνυμα στο eclass.

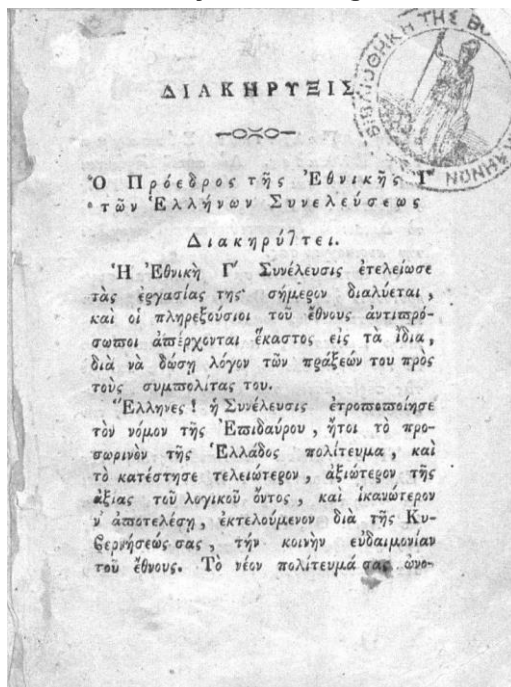
Τάσος Κεσίδης

Original image



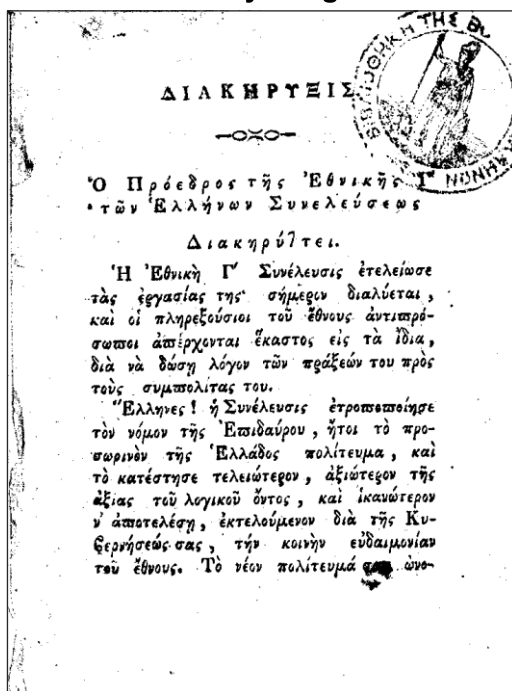
Εικόνα 1

Grayscale image



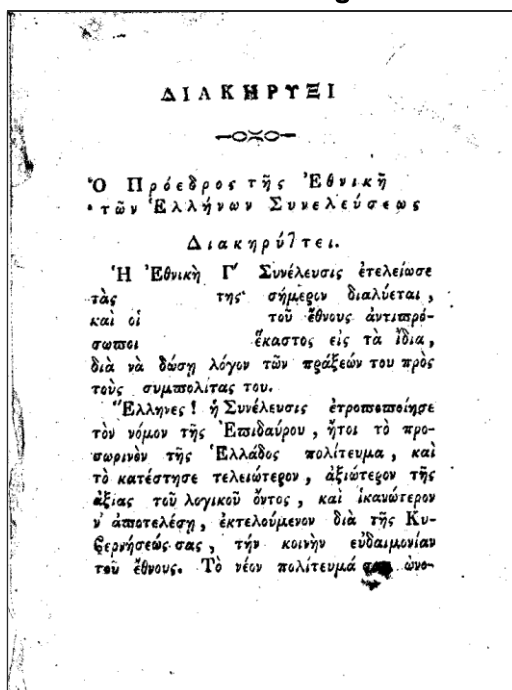
Εικόνα 2

Binary image



Εικόνα 3

Cleaned image



Εικόνα 4

[illegible][illegible]