



## ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ  
ΤΟΜΕΑΣ ΕΠΙΚΟΙΝΩΝΙΩΝ, ΗΛΕΚΤΡΟΝΙΚΗΣ ΚΑΙ ΣΥΣΤΗΜΑΤΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ

### **Αυτόματη Κατάτμηση Εικόνων Χειρογράφου με Σκοπό την Ταυτοποίηση του Γραφέα Τους τους με Εφαρμογή σε Κείμενα του Ρήγα Φεραίου**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

της

**ΚΑΜΠΟΥΡΗ ΠΑΝΑΓΙΩΤΑΣ-ΔΩΡΟΘΕΑΣ**

**Επιβλέπων :** Κωνσταντίνος Παπαοδυσσέας  
Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούλιος 2021





ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ  
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ  
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ  
ΤΟΜΕΑΣ ΕΠΙΚΟΙΝΩΝΙΩΝ, ΗΛΕΚΤΡΟΝΙΚΗΣ ΚΑΙ ΣΥΣΤΗΜΑΤΩΝ  
ΠΛΗΡΟΦΟΡΙΚΗΣ

## **Αυτόματη Κατάτμηση Εικόνων Χειρογράφου με Σκοπό την Ταυτοποίηση Του Γραφέα Τους τους με Εφαρμογή σε Κείμενα του Ρήγα Φεραίου**

### **ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ**

της

**ΚΑΜΠΟΥΡΗ ΠΑΝΑΓΙΩΤΑΣ-ΔΩΡΟΘΕΑΣ**

**Επιβλέπων :** Κωνσταντίνος Παπαοδυσσέας  
Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 14<sup>η</sup> Ιουλίου 2021.

.....  
Κωνσταντίνος Παπαοδυσσέας  
Καθηγητής Ε.Μ.Π.

.....  
Κουκούτσης Ηλίας  
Καθηγητής Ε.Μ.Π.

.....  
Παναγόπουλος Μιχαήλ  
Επίκουρος Καθηγητής Ιονίου  
Πανεπιστημίου

Αθήνα, Ιούλιος 2021

.....

**ΚΑΜΠΟΥΡΗ ΠΑΝΑΓΙΩΤΑ-ΔΩΡΟΘΕΑ**

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

Copyright © Καμπούρη Παναγιώτα-Δωροθέα, 2021

Copyright © Κωνσταντίνος Παπαοδυσσέας, 2021

All rights reserved. Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν στη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς το συγγραφέα. Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν το συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

## Περίληψη

Το πρόβλημα της κατάτμησης ψηφιακών εικόνων αποτελεί σημαντικό αντικείμενο μελέτης στους τομείς της Ψηφιακής Επεξεργασίας Εικόνας και της Όρασης Υπολογιστών. Πρόκειται για τη διαδικασία διαχωρισμού μιας ψηφιακής εικόνας σε πολλά τμήματα, με σκοπό την εξαγωγή των πληροφοριών που μας ενδιαφέρουν. Η κατάτμηση εικόνας αποτελεί αναπόσπαστο βήμα της αυτόματης ταυτοποίησης γραφέα, η οποία χρησιμοποιείται ευρέως σε παλαιογραφικές, αρχαιομετρικές και βιομετρικές εφαρμογές. Ειδικότερα τα τελευταία χρόνια η επιστημονική έρευνα έχει στραφεί και σε εφαρμογές στον τομέα της Ιστορίας και της Αρχαιολογίας. Σε αυτό το πλαίσιο, στην παρούσα διπλωματική εργασία μελετάται μια μέθοδος αυτόματης κατάτμησης εικόνων χειρογράφων με σκοπό την ταυτοποίηση του γραφέα τους και την εφαρμογή της σε ιστορικά κείμενα. Συγκεκριμένα, με την προτεινόμενη μέθοδο θα μελετήσουμε χειρόγραφα κείμενα του Ρήγα Φεραίου, τόσο κάποια που είναι ήδη γνωστά, όσο και κάποια πρόσφατα ανακαλυφθέντα στη Ρουμανία, με αφορμή τη διαφωνία της επιστημονικής κοινότητας αναφορικά με το αν αυτά γράφθηκαν δια χειρός Ρήγα ή όχι.

**Λέξεις Κλειδιά:** Κατάτμηση Εικόνας, Ψηφιακή Επεξεργασία Εικόνας, Αυτόματη Αναγνώριση Γραφέα, Ιστορικά Κείμενα, Χειρόγραφα, Αρχαιομετρία, Ταυτοποίηση Γραφέα, Παλαιογραφία



## **Abstract**

Image segmentation constitutes an important aspect of Digital Image Processing and Computer Vision. It is the process of dividing a digital image into several parts, in order to extract useful information to the user(s). Image segmentation is also an integral part of automatic writer identification, which is widely used in various fields, such as biometrics, archaeometry, palaeography etc. We would like to emphasize that in recent years, scientific research has turned to applications for automatic writer identification in the field of History and Archaeology. This dissertation presents a method of automatic segmentation of manuscript images in order to identify the hand who wrote them, and its application to historical texts. Specifically, the method studied in this dissertation will be applied to the manuscripts of Rhigas Feraios, both those known and the recently discovered ones in Romania, in order to attempt to resolve the disagreement of the scientific community as to whether these two documents were written by Rhigas or not.

**Keywords:** Image Segmentation, Digital Image Processing, Automatic Writer Identification, Historical Texts, Manuscripts, Archaeometry, Palaeography





## Πίνακας περιεχομένων

<b>1</b>	<b>Εισαγωγή</b> .....	<b>1</b>
1.1	Σκοπός διπλωματικής εργασίας .....	1
1.2	Σημασία και Δυσκολίες Της Αυτόματης Ταυτοποίησης Γραφέα .....	1
1.3	Οργάνωση του Κειμένου της Παρούσης Διπλωματικής και Συνοπτική Παρουσίαση των Ενοτήτων που θα Ακολουθήσουν.....	5
<b>2</b>	<b>Βιβλιογραφική Ανασκόπηση Μεθόδων Αυτόματης Αναγνώρισης Γραφέα</b> .....	<b>6</b>
2.1	Επί τόπου (on-line) αναγνώριση γραφέα .....	7
2.2	Εκ των υστέρων (off-line) αναγνώριση γραφέα .....	7
<b>3</b>	<b>Ιστορικά Στοιχεία και Η Σημασία της Ποσοτικής Απόδειξης ότι Ένα Χειρόγραφο Ανήκει του Ρήγα Φεραίου</b> .....	<b>10</b>
3.1	Βιογραφικά Στοιχεία Ρήγα Φεραίου .....	10
3.2	Το Συγγραφικό Έργο του Ρήγα Φεραίου .....	11
3.3	Τα ανακαλυφθέντα δυο χειρόγραφα στη Ρουμανία ανήκουν στο Ρήγα ή όχι; .....	14
<b>4</b>	<b>Μαθηματικό Υπόβαθρο</b> .....	<b>18</b>
4.1	Στατιστικές Κατανομές .....	18
4.1.1	Κανονική Κατανομή.....	18
4.1.2	Κατανομή $\chi^2$ .....	19
4.1.3	Κατανομή $t$ ή <i>Student</i> .....	20
4.2	Έλεγχος Στατιστικών Υποθέσεων.....	21
4.2.1	Υποθέσεις για τη μέση τιμή ενός πληθυσμού από ένα δείγμα του.....	21
4.2.2	Υποθέσεις που αφορούν τη διαφορά των μέσων δύο πληθυσμών.....	23
4.2.3	Έλεγχος Υποθέσεων Κατανομής με τεστ <i>Kolmogorov-Smirnoff</i> .....	25
<b>5</b>	<b>Περιγραφή Προτεινόμενης Μεθοδολογίας</b> .....	<b>26</b>
5.1	Κατάτμηση Εικόνας.....	26
5.1.1	Σημασία Κατάτμησης Εικόνας στην Παρούσα Εφαρμογή .....	26
5.1.1	Το Πρόβλημα της Κατάτμησης Εικόνας.....	26
5.1.2	Μέθοδοι Κατάτμησης Εικόνας με βάση Διαχωριστικές Τεχνικές .....	27

5.1.3	<i>Μέθοδοι Κατάτμησης Εικόνας με βάση Ιεραρχικές Τεχνικές</i> .....	27
5.1.4	<i>Μια Νέα Εναλλακτική Μέθοδος Κατάτμησης</i> .....	29
5.2	<i>Εξαγωγή Περιγραμμάτων των Υλοποιήσεων των Γραμμάτων της Αλφαβήτου</i> .....	32
5.3	<i>Υποβέλτιστη Προσαρμογή Περιγραμμάτων</i> .....	34
5.3.1	<i>Αυτοματοποιημένη Διαδικασία Υποβέλτιστης Προσαρμογής Περιγραμμάτων</i> .....	34
5.4	<i>Τελεσίδικο Κριτήριο για την βέλτιστη προσαρμογή δύο υλοποιήσεων</i> .....	36
5.4.1	<i>Ορισμός Κριτηρίου Για Δύο Απλές Συνεκτικές Υλοποιήσεις του Ίδιου Γράμματος</i> .....	36
5.4.2	<i>Ορισμός Κριτηρίου Για Δύο Μη Συνεκτικές Υλοποιήσεις του Ίδιου Γράμματος</i> .....	37
5.5	<i>Στατιστικός Έλεγχος Ομοιότητας Δύο Υλοποιήσεων</i> .....	38
5.5.1	<i>Πιθανοφάνεια στη θέση βέλτιστης Προσαρμογής Δύο Υλοποιήσεων Συμβόλων της Αλφαβήτου</i> .....	40
<b>6</b>	<b>Εφαρμογή Μεθοδολογίας στα Χειρόγραφα του Ρήγα Φεραίου</b> .....	<b>43</b>
6.1	<i>Εφαρμογή των Μεθόδων Κατάτμησης Εικόνας στα Κείμενα του Ρήγα Φεραίου</i> .....	43
6.1.1	<i>Αποτελέσματα Κατάτμησης</i> .....	43
6.1.2	<i>Διερεύνηση αποτελεσμάτων κατάτμησης ως προς τις τιμές των παραμέτρων του αλγορίθμου κατάτμησης</i> .....	50
6.2	<i>Εξαγωγή Περιγραμμάτων των Επιλεχθέντων Γραμμάτων των Κειμένων του Ρήγα Φεραίου</i> .....	57
6.3	<i>Αποτελέσματα Μεθόδου για τα Επιλεχθέντα Γράμματα</i> .....	59
6.3.1	<i>Σύγκριση Υλοποιήσεων επιλεγμένων γραμμάτων κάθε έργου</i> .....	59
6.3.2	<i>Σύγκριση Υλοποιήσεων επιλεγμένων γραμμάτων μεταξύ διαφορετικών έργων</i> .....	88
<b>7</b>	<b>Επίλογος</b> .....	<b>101</b>
7.1	<i>Σύνοψη και συμπεράσματα</i> .....	101
7.2	<i>Μελλοντικές επεκτάσεις</i> .....	101
<b>8</b>	<b>Βιβλιογραφία</b> .....	<b>102</b>

# 1

## *Εισαγωγή*

### ***1.1 Σκοπός διπλωματικής εργασίας***

Σκοπός της παρούσης διπλωματικής εργασίας είναι η παρουσίαση μιας μεθόδου αυτόματης κατάταξης εικόνων χειρογράφων με σκοπό την ταυτοποίηση του γραφέα τους και η εφαρμογή της σε κείμενα μεγάλης ιστορικής αξίας. Συγκεκριμένα, με την προτεινόμενη μέθοδο θα μελετήσουμε χειρόγραφα κείμενα του Ρήγα Φεραίου τόσο κάποια που είναι ήδη γνωστά, όσο και τα πρόσφατα ανακαλυφθέντα στη Ρουμανία, με αφορμή τη διαφωνία της επιστημονικής κοινότητας αναφορικά με το αν αυτά γράφθηκαν δια χειρός Ρήγα ή όχι. Τα δύο αυτά κείμενα είναι το «Σαγανάκι της Τρέλλας» και «Η Δοκιμασία της Φιλίας». Συγκεκριμένα, τα δύο αυτά κείμενα ανακαλύφθηκαν το 1998 από τη Ρουμάνια ελληνίστρια ιστορικό κυρία Lia Brad Chisacof . Η κυρία Chisacof παρουσίασε τα εν λόγω κείμενα και υποστήριξε ότι ανήκουν στο Ρήγα. Έκτοτε υπήρξε σημαντική επιστημονική διαφωνία και διάλογος επί της πατρότητας των κειμένων [1].

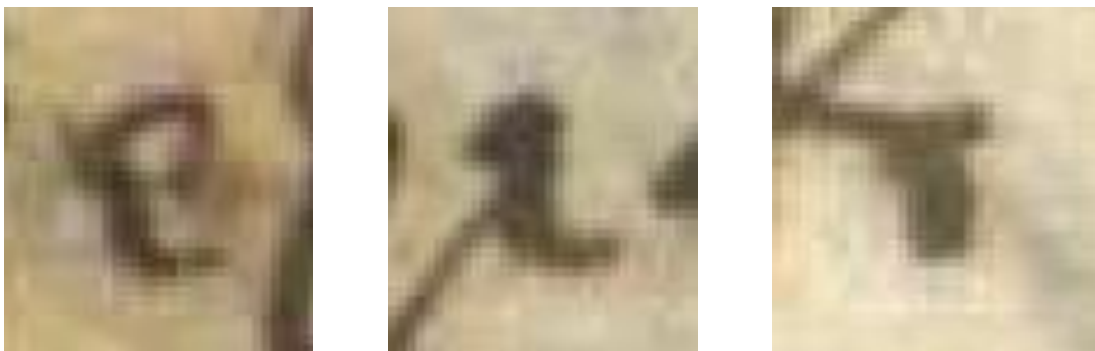
### ***1.2 Σημασία και Δυσκολίες Της Αυτόματης Ταυτοποίησης Γραφέα***

Η εφαρμογή της αυτόματης ταυτοποίησης γραφέα παρουσιάζει εξαιρετικό ενδιαφέρον, ειδικά στον τομέα της Ιστορίας και της Αρχαιολογίας και περισσότερο της Επιγραφολογίας. Όμως, είναι ιδιαίτερος σημαντικό ότι σε πολλά διασωθέντα κείμενα είτε αυτά είναι αρχαίες επιγραφές, είτε αρχαίοι πάπυροι, είτε περισσότερο σύγχρονα χειρόγραφα, ο γραφέας δεν υπέγραφε τα κείμενα του ούτε επισήμαινε την ημερομηνία, ούτε τον τόπο γραφής. Δεδομένου αυτού, προκύπτει το πρόβλημα της ορθής χρονολόγησης των κειμένων αυτών, αλλά και της ορθής απόδοσής τους στον πραγματικό γραφέα/συγγραφέα τους. Όσον αφορά τα αρχαιότερα

κείμενα, η χρονολόγηση αυτών είναι εξαιρετικής σημασίας, καθώς οι πληροφορίες που αυτά περιέχουν είναι πολύτιμες για τη μελέτη της εποχής στην οποία αυτά γράφτηκαν. Όσον αφορά τα νεότερα κείμενα, πάλι η ορθή χρονολόγησή τους και η απόδοση τους στον πραγματικό γραφέα είναι συχνά καθοριστικής σημασίας για την Λογοτεχνία, την Φιλοσοφία, την Ιστορία, καθώς και για την καλύτερη κατανόηση του αντίστοιχου πολιτιστικού και κοινωνικού περιβάλλοντος. Έως τη σημαντική ανάπτυξη, των Μαθηματικών και της Επιστήμης των Υπολογιστών, η μελέτη που αφορούσε τον γραφέα ενός κειμένου ήταν ιδιαίτερα δύσκολη και πραγματοποιείτο εμπειρικά από ειδικούς επιστήμονες.

Η μέθοδος που παρουσιάζεται εδώ χρησιμοποιεί μαθηματικά κριτήρια και αρκετά πολύπλοκους αλγόριθμους, οι οποίοι έχουν ως τώρα παρουσιαστεί στις δημοσιεύσεις [2,3,4]. Σημειώνεται ότι, έως τώρα, η προτεινόμενη εδώ μεθοδολογία προσφέρει εξαιρετικά ακριβή αποτελέσματα.

Οι δυσκολίες που παρουσιάζονται σε μια οποιαδήποτε αυτοματοποιημένη διαδικασία ταυτοποίησης γραφέα, οφείλονται σε πολλούς παράγοντες, όπως η ποιότητα των αρχικών φωτογραφιών του υπό μελέτη κειμένου και συγκεκριμένα η φωτεινότητα και η ανάλυσή τους, όπως φαίνεται στο Σχήμα 1.



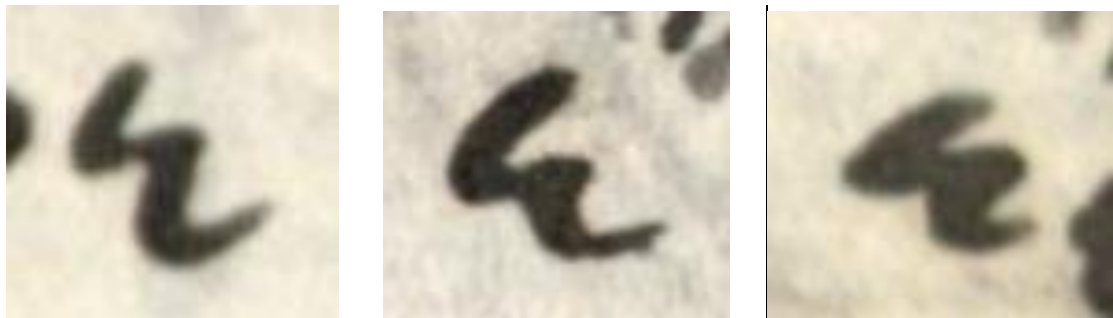
*Σχήμα 1: Παραδείγματα γραμμάτων που προέρχονται από φωτογραφίες χαμηλής ανάλυσης και χαμηλής φωτεινότητας*

Εξίσου σημαντικός παράγοντας είναι η παρουσία «ανωμαλιών» στο κείμενο, όπως σημάδια από μελάνι ή σημάδια φθοράς του χαρτιού ή γενικότερα της επιφάνειας γραφής, όπως φαίνεται στο Σχήμα 2. Δε θα πρέπει να παραλειφθεί και η διάβρωση που εμφανίζεται στην περίπτωση των επιγραφών και όχι μόνον και αλλοιώνει το σχήμα των υπό ανάλυση γραμμάτων, όπως και η κακή απόδοση του εργαλείου γραφής, είτε αυτό είναι σκαρπέλο, όπως συμβαίνει στις αρχαίες επιγραφές είτε είναι σύγχρονη πένα. Όλοι οι προαναφερθέντες παράγοντες αποτελούν συνολικά ένα είδος πρόσθετου «θορύβου» που καλούμαστε να αντιμετωπίσουμε με τη σωστή επεξεργασία των εικόνων του κειμένου.



Σχήμα 2: Παραδείγματα εικόνων με έντονη παρουσία «θορύβου»

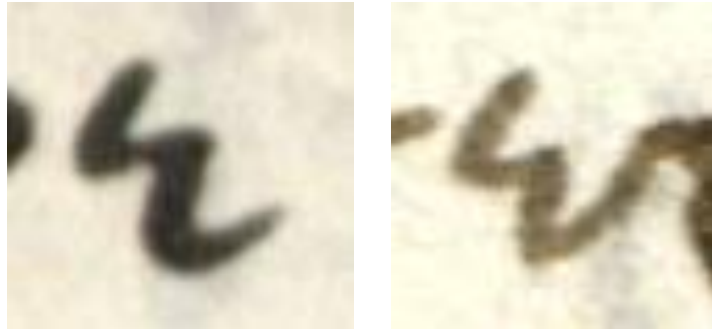
Εκτός από την ποιότητα του σαρωμένου ή φωτογραφισμένου κειμένου, καλούμαστε να αντιμετωπίσουμε τη μεταβλητότητα ως προς το σχήμα, το μέγεθος και τον προσανατολισμό των γραμμάτων του ίδιου γραφέα. Ενδεικτικά παραθέτουμε τρεις διαφορετικές υλοποιήσεις του ίδιου γράμματος από μία μόνο σελίδα των χειρογράφων που εξετάστηκαν:



Σχήμα 3: Τρεις διαφορετικές υλοποιήσεις του γράμματος της αλφαβήτου «ε», οι οποίες υπάρχουν σε μια συγκεκριμένη σελίδα του κειμένου «Το Σαγανάκι της Τρέλλας»

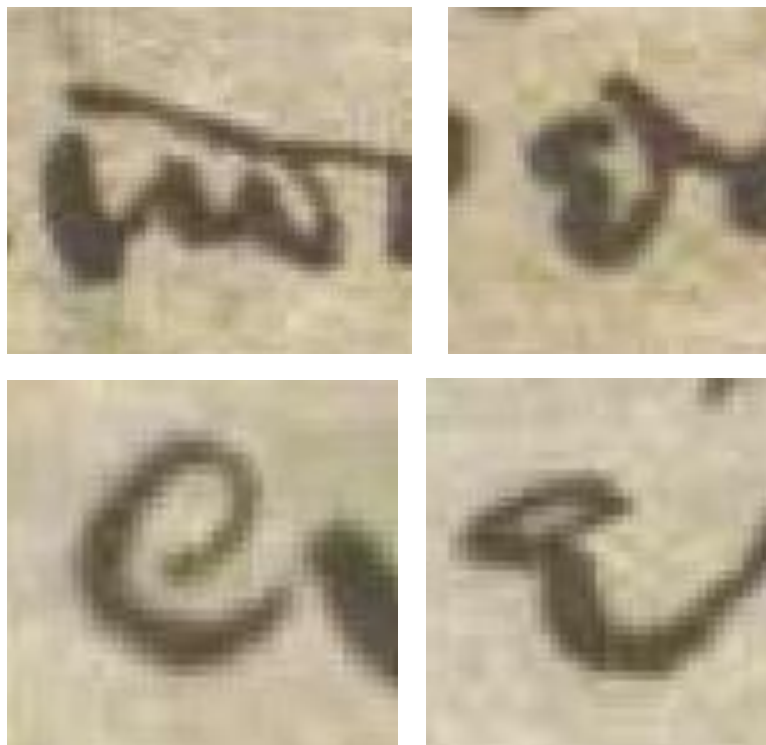
Άμεσο επακόλουθο αυτών των διαφοροποιήσεων μεταξύ των γραμμάτων του ίδιου γραφέα, είναι αυτές να είναι συγκρίσιμες με τις διαφοροποιήσεις μεταξύ διαφορετικών γραφέων. Ενδεικτικά παραθέτουμε δύο ζεύγη γραμμάτων, εκ των οποίων το πρώτο είναι του ίδιου γραφέα, ενώ τα στοιχεία του δεύτερου ζεύγους ανήκουν σε διαφορετικούς γραφείς:





*Σχήμα 4: Παράδειγμα στο οποίο ένα ζεύγος γραμμάτων που προέρχεται από διαφορετικούς γραφείς (πάνω ζεύγος) έχει περισσότερες ομοιότητες από ένα ζεύγος που προέρχεται από τον ίδιο γραφέα (πάνω ζεύγος)*

Τέλος, δε θα πρέπει να παραλειφθούν και οι διαφορετικές τεχνοτροπίες του ίδιου γράμματος που μπορεί να εμφανίζονται στο κείμενο του ίδιου γραφέα. Χαρακτηριστικό παράδειγμα αποτελούν τα ζεύγη γραμμάτων του Σχήματος 5, που εντοπίζονται στις ίδιες σελίδες ενός εκ των υπό μελέτη χειρογράφων:



*Σχήμα 5: Διαφορετικές τεχνοτροπίες των συμβόλων «π» και «ε» που βρίσκονται στην ίδια σελίδα*

### ***1.3 Οργάνωση του Κειμένου της Παρούσης Διπλωματικής και Συνοπτική Παρουσίαση των Ενοτήτων που θα Ακολουθήσουν***

Στο Κεφάλαιο 2 πραγματοποιείται μια σύντομη βιβλιογραφική ανασκόπηση των σύγχρονων μεθόδων αυτόματης αναγνώρισης γραφέα, για λόγους πληρότητας. Στο Κεφάλαιο 3 παραθέτουμε τις βασικές πληροφορίες για τη ζωή και τα έργα του Ρήγα Φεραίου, καθώς και για τη σημασία της μελέτης των ανακαλυφθέντων χειρογράφων. Στο Κεφάλαιο 4 αναλύονται οι απαραίτητες μαθηματικές έννοιες που θα χρησιμοποιηθούν στην προτεινόμενη μέθοδο της παρούσης εργασίας για την αναγνώριση γραφέα, της οποίας τα βήματα περιγράφονται στο Κεφάλαιο 5. Στο Κεφάλαιο 6 παρουσιάζονται αποτελέσματα της μεθόδου κατάτμησης των χειρογράφων του Ρήγα Φεραίου, για ένα συγκεκριμένο σύνολο γραμμάτων της αλφαβήτου. Στο Κεφάλαιο 7 παρουσιάζεται μια σύνοψη της εργασίας, καθώς και πιθανές μελλοντικές επεκτάσεις. Τέλος, στο Κεφάλαιο 8 παραθέτουμε τη βιβλιογραφία που χρησιμοποιήθηκε στο πλαίσιο της παρούσης εργασίας.

# 2

## *Βιβλιογραφική Ανασκόπηση Μεθόδων Αυτόματης Αναγνώρισης Γραφέα*

Τα τελευταία χρόνια το επιστημονικό ενδιαφέρον στον τομέα της αναγνώρισης γραφέα επικεντρώνεται κυρίως σε βιομετρικές εφαρμογές, προκειμένου να διασφαλίζεται η αυθεντικότητα των γραπτών κειμένων ή των υπογραφών. Συγκεκριμένα, ως βιομετρική ορίζουμε την αυτόματη αναγνώριση ενός ατόμου με βάση της συμπεριφοράς του, των κειμένων του, τα φυσικά του χαρακτηριστικά και λοιπά. Από τον ορισμό της βιομετρικής προκύπτουν δύο κατηγορίες των εν λόγω χαρακτηριστικών. Η πρώτη περιλαμβάνει τα φυσιολογικά βιομετρικά δεδομένα, τα οποία βασίζονται στην άμεση μέτρηση τμημάτων του ανθρώπινου σώματος, όπως της ίριδας, των δακτυλικών αποτυπωμάτων, του αμφιβληστροειδούς και των διαστάσεων του προσώπου [5]. Η δεύτερη κατηγορία είναι τα βιομετρικά συμπεριφοράς, τα οποία βασίζονται στη συμπεριφορά του ατόμου όπως το βάδισμα, η φωνή, ο τρόπος πληκτρολόγησης, η υπογραφή, και ο γραφικός χαρακτήρας[5]. Κατά συνέπεια, η ταυτοποίηση γραφέα ανήκει στην ομάδα των βιομετρικών συμπεριφοράς. Με την ταυτοποίηση γραφέα είμαστε σε θέση να προσδιορίσουμε το χέρι που έγραψε ένα κείμενο, από μια ομάδα γραφέων που είναι γνωστοί από το σύστημα [5]. Επιπλέον, δε θα πρέπει να παραλειφθεί η συμβολή της αναγνώρισης γραφέα στην ιατροδικαστική επιστήμη, καθώς είναι αποδεκτή σε δικαστήρια ως αποδεικτικό στοιχείο εδώ και περισσότερο από έναν αιώνα [5].

Η αναγνώριση του συγγραφέα μπορεί γενικά να πραγματοποιηθεί είτε εκ των υστέρων (off-line) είτε επί τόπου (on-line). Η πρώτη περίπτωση βασίζεται στην επεξεργασία μιας



σαρωμένης εικόνας του χειρόγραφου κειμένου. Στη δεύτερη περίπτωση, διατίθενται χωρικές και χρονικές πληροφορίες σχετικά με τον τρόπο γραφής του κειμένου, όπως δεδομένα της γωνίας της γραφίδας, της ταχύτητας γραφής, της πίεσης που το υποκείμενο ασκεί στη γραφίδα, κλπ [6]. Επίσης, πολλές προσεγγίσεις αναγνώρισης γραφέα μπορούν να κατηγοριοποιηθούν σε *ανεξάρτητες από το κείμενο (text independent)* και *εξαρτώμενες από το κείμενο μεθόδους (text dependent)* [6]. Στην πολύ σύγχρονη γραφολογία, στις ανεξάρτητες από κείμενο μεθόδους, ο γραφέας ζητείται από το γραφέα να γράψει οποιοδήποτε κείμενο στη δοκιμαστική φάση, ενώ στις μεθόδους που εξαρτώνται από το κείμενο, ο κάθε γραφέας θα πρέπει να γράψει στη δοκιμαστική φάση το ίδιο ακριβώς κείμενο με αυτό που θα χρησιμοποιηθεί στη διαδικασία εκπαίδευσης. Κατά συνέπεια, οι μέθοδοι που εξαρτώνται από το κείμενο δεν είναι εφαρμόσιμες στην πράξη, όπως για παράδειγμα στην ποινική δικαιοσύνη, όπου πρέπει να συγκριθεί το περιεχόμενο διαφόρων, ποικίλων εγγράφων αναφορικά με τον γραφέα τους [5].

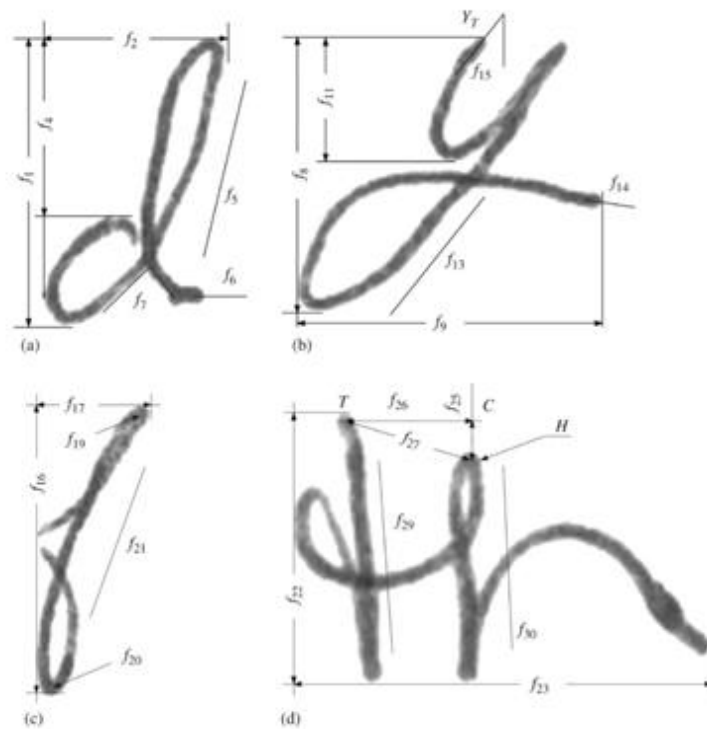
## **2.1 *Επί τόπου (on-line) αναγνώριση γραφέα***

Στην περίπτωση της επί τόπου αναγνώρισης γραφέα υπάρχει σχετικά περιορισμένος αριθμός μελετών σε σχέση με την εκ των υστέρων αναγνώριση, λόγω του σύνθετου χαρακτήρα που πρέπει να έχουν οι χρησιμοποιούμενοι μέθοδοι. Ο προσδιορισμός των τεχνοτροπιών γραφής πραγματοποιείται με διαφορετικούς τρόπους όπως η χρήση των σχέσεων μεταξύ στατικών και δυναμικών στοιχείων χειρόγραφων κειμένων [7], η χρήση ενός μοντέλου που ονομάζεται Beta – Elliptic, για τη μελέτη αραβικών κειμένων στο [5] και μέσω Random Hybrid Strokes, που περιγράφει τις κινήσεις του εργαλείου γραφής του εκάστοτε γραφέα στο [6]. Η τελική κατάταξη των γραφέων επιτυγχάνεται με τη χρήση νευρωνικών δικτύων και συγκεκριμένα στο [7] μέσω ενός Νευρωνικού Δικτύου Μίας Κατεύθυνσης (Feed Forward Neural Network) και στο [6] μέσω ενός Νευρωνικού Δικτύου με Ανάδραση (Recurrent Neural Network).

## **2.2 *Εκ των υστέρων (off-line) αναγνώριση γραφέα***

Στον τομέα της εκ των υστέρων αναγνώρισης και επαλήθευσης γραφέα έχει σημειωθεί μεγαλύτερη ερευνητική πρόοδος, η οποία επικεντρώνεται κυρίως σε τεχνικές επεξεργασίας εικόνας και αναγνώρισης προτύπων. Στο πλαίσιο αυτό έχουν προταθεί διάφορες τεχνικές όπως μορφολογικές προσεγγίσεις [8], επεξεργασία των εικόνων με φιλτράρισμα Gabor [9], αλλά και τεχνικές αναγνώρισης υφής, όπως η [10]. Τεχνικές που επιχειρούν να περιγράψουν οπτικά χαρακτηριστικά των γραμμάτων, όπως το ύψος, το πλάτος και η κλίση, όπως φαίνεται

στο Σχήμα 6, μπορούν επίσης να χρησιμοποιηθούν για την αναγνώριση και την επαλήθευση του γραφέα[11–14]. Επίσης, οι [15–17] χρησιμοποιούν μοντέλα Hidden Markov (HMM), δηλαδή προτείνουν τη χρήση πολλαπλών HMM συστημάτων αναγνώρισης γραφικού χαρακτήρα, το καθένα από τα οποία εκπαιδεύεται χρησιμοποιώντας δεδομένα από έναν γραφέα. Η επιλογή των HMM για αναγνώριση κειμένου έγκειται στην αντοχή τους στο θόρυβο, στην δυνατότητα διαχωρισμού της κάθε γραμμής σε λέξεις και χαρακτήρες και στην ύπαρξη διαθέσιμων σχετικών αλγορίθμων και τεχνικών. Οι ίδιοι συγγραφείς με το [17] πρότειναν στη συνέχεια τη χρήση Gaussian Mixture Model (GMM) για αναγνώριση γραφέα [18]. Αυτή η μέθοδος εκπαιδεύσε επίσης ένα μοντέλο για καθέναν από τους συγγραφείς και κάθε μοντέλο έγινε ειδικός στο χειρόγραφο ενός συγκεκριμένου γραφέα. Το προκύπτον σύστημα συγκρίθηκε με ένα σύστημα βασισμένο σε HMM, και αποδείχθηκε ότι ήταν απλούστερο και γρηγορότερο στην εκπαίδευση.



Σχήμα 6: Παράδειγμα εξαγωγής χαρακτηριστικών από την εργασία [14]

Μια ιδιαίτερη μέθοδος offline αυτόματης ταυτοποίησης γραφέα είναι η δημιουργία ενός συνόλου χαρακτηριστικών που ονομάζεται codebook. Η δημιουργία codebook είναι στην ουσία βασική συνιστώσα του μοντέλου bag-of-words[19]. Η χρήση αυτής της μεθοδολογίας είναι ιδιαίτερα δημοφιλής στη διεθνή βιβλιογραφία[19]. Για παράδειγμα, στην εργασία [20], το codebook δημιουργείται με πληροφορίες από το ιστόγραμμα συχνότητας εμφάνισης των διαφόρων σχημάτων, ενώ η πιο δημοφιλής τεχνική είναι με χρήση clustering  $k$ -μέσων [19].

Δε θα πρέπει επίσης να παραλειφθεί πως η χρήση των αλλογράφων είναι μια ευρέως χρησιμοποιούμενη μέθοδος για την ταυτοποίηση γραφέα[21,22]. Κατά την μελέτη χειρογράφων, οι ειδικοί συχνά δημιουργούν λίστες αλλογράφων που εμφανίζονται σε αυτά. Με βάση τη λίστα αυτή γίνεται η διάκριση μεταξύ κοινών αλλογράφων (δηλαδή αλλογράφων που εντοπίζονται στο γραφικό χαρακτήρα της πλειοψηφίας) και λιγότερο κοινών αλλογράφων, τα οποία είναι χαρακτηριστικά για περιορισμένο μόνο αριθμό γραφέων. Εάν το ίδιο ασυνήθιστο αλλόγραφο εμφανίζεται σε δύο έγγραφα, οι πιθανότητες είναι υψηλότερες αυτά να ανήκουν στον ίδιο γραφέα, από ό,τι όταν εντοπίζονται δύο κοινά αλλόγραφα σε δύο έγγραφα. Στη συνέχεια, οι λίστες αλλογράφων διαφορετικών εγγράφων μπορούν να συγκριθούν και να ελεγχθούν προσεκτικά για να καταλήξουμε στο πόσο καλά ταιριάζουν δύο χειρόγραφα[23]. Σε επίπεδο έρευνας, στην εργασία [23], οι συγγραφείς χρησιμοποιούν τη δημιουργία διανυσμάτων συχνότητας αλλογράφου για την ταυτοποίηση γραφέα.

Η εφαρμογή των παραπάνω μεθόδων σε ιστορικά κείμενα και σε κείμενα που περιέχουν σύμβολα αλφαβήτων εκτός του ελληνικού ή του λατινικού πραγματοποιήθηκε τα τελευταία χρόνια [24 – 31]. Μια ιδιαίτερη περίπτωση εφαρμογής των τεχνικών αναγνώρισης γραφέα και ομαδοποίησης κειμένων ανάλογα με το γραφέα είναι η αυτόματη ομαδοποίηση μουσικών χειρογράφων [32]. Τέλος, χαρακτηριστικά όπως η καμπυλότητα, η υφή και η κλίση της γραφής έχουν χρησιμοποιηθεί και για την ταυτοποίηση του φύλου του γραφέα [33].

# 3

## *Ιστορικά Στοιχεία και Η Σημασία της Ποσοτικής Απόδειξης ότι Ένα Χειρόγραφο Ανήκει του Ρήγα Φεραίου*

### **3.1 Βιογραφικά Στοιχεία Ρήγα Φεραίου**

Ο Ρήγας Βελεστινλής ή Ρήγας Φεραίος γεννήθηκε στο Βελεστίνο, τις αρχαίες Φερές, το 1757, από εύπορη οικογένεια, η οποία όμως βίωσε την τουρκική μανία. Είναι γνωστός για το επαναστατικό, πολιτικό και συγγραφικό του έργο.

Διδάχθηκε τα πρώτα του γράμματα από έναν ιερέα στο Βελεστίνο και κατόπιν στη Ζαγορά. Η δίψα του για μάθηση τον οδήγησε στα Αμπελάκια για περαιτέρω μόρφωση και όταν επέστρεψε, έγινε δάσκαλος στην κοινότητα Κισσού Πηλίου. Στην ηλικία των είκοσι ετών έφυγε από το Βελεστίνο λόγω ενός θερμού επεισοδίου με τους Τούρκους, και κατέφυγε αρχικά στο Λιτόχωρο του Ολύμπου, όπου κατατάχθηκε στο σώμα των αρματολών του θείου του, Σπύρου Ζήρα. Αργότερα μετέβη στο Άγιο Όρος στη μονή Βατοπεδίου, όπου του εδόθη ελεύθερη πρόσβαση στη βιβλιοθήκη της, στις οποίας τα έργα αφιέρωσε σοβαρό χρόνο μελέτης. Στη συνέχεια, ταξίδεψε στην Κωνσταντινούπολη, μετά από πρόσκληση του Πρέσβη της Ρωσίας για σπουδές, όπου διέυρνε τις γνώσεις του στη Γαλλική, την Ιταλική και τη Γερμανική γλώσσα.

Το 1786 ξεκίνησε την επαγγελματική του πορεία ως γραμματέας του ηγεμόνα της Βλαχίας Νικόλαου Μαυρογένη, αδερφό του παππού της Μαντώς Μαυρογένους και εγκαταστάθηκε στο Βουκουρέστι που ήταν η έδρα της ηγεμονίας. Ωστόσο, η έκβαση του

Ρωσοτουρκικού πολέμου το 1790 οδήγησε στον αποκεφαλισμό του Μαυρογένη και ανάγκασε το Ρήγα να καταφύγει στη Βιέννη, καθιστώντας την έδρα της επαναστατικής δράσης του. Εκεί συνοδοιπόροι στο έργο του ήταν κυρίως Έλληνες έμποροι ή φοιτητές, σημαντικότεροι εκ των οποίων ήταν οι τυπογράφοι αδελφοί Πούλιου. Στο τυπογραφείο τους τυπώθηκαν σημαντικά έργα του Ρήγα, και συγκεκριμένα ο Θούριος και η Χάρτα, το Σχολείον των Ντελικάτων Εραστών, το Φυσικής απάνθισμα, ο Ηθικός Τρίπους, Το Σύνταγμα της Ελληνικής Δημοκρατίας, Τα Δίκαια του ανθρώπου, καθώς και ο πρώτος τόμος του βιβλίου Νέου Αναχάρσις.

Παρέμεινε στη Βιέννη έως τον Δεκέμβριο του 1797 προετοιμάζοντας το επαναστατικό σχέδιο μέσω του συγγραφικού του έργου. Στη συνέχεια, μετέβη στην Τεργέστη με σκοπό να κατέβει στην Ελλάδα, αλλά επροδόθη από τον Έλληνα έμπορο Δημήτριο Οικονόμου και συνελήφθη από την Αυστριακή Αστυνομία [34].

Στις 10 Μαΐου 1798 παρεδόθη με συνοδεία των αυστριακών αρχών στους Τούρκους του Βελιγραδίου και φυλακίστηκε στον Πύργο Νεμπόισα μαζί με τους επτά ομοϊδεάτες συντρόφους του Ευστράτιο Αργέντη, Δημήτριο Νικολίδη, Αντώνιο Κορωνιό, Ιωάννη Καρατζά, Θεοχάρη Γεωργίου, Ιωάννη Εμμανουήλ και Παναγιώτη Εμμανουήλ. Στις 24 Ιουνίου του 1798, ύστερα από συνεχή βασανιστήρια, όλα τα μέλη αυτής της επαναστατικής ομάδας συμπεριλαμβανομένου του Ρήγα στραγγαλίστηκαν και τα σώματά τους ρίχτηκαν στον Δούναβη, σύμφωνα με τα επίσημα Αυστριακά έγγραφα. Διαδόθηκαν ψευδείς φήμες από τις Οθωμανικές Αρχές ότι οι οκτώ έγκλειστοι είχαν δραπετεύσει, εξαπολύθηκαν εικονικές έρευνες για την ανεύρεσή τους και τελικά δηλώθηκε ότι οι δραπέτες είχαν πνιγεί στον ποταμό.

### **3.2 Το Συγγραφικό Έργο του Ρήγα Φεραίου**

Δεδομένου ότι στην παρούσα εργασία πραγματοποιείται μελέτη συγκεκριμένων χειρογράφων του Ρήγα, κρίνεται σκόπιμη η συνοπτική παράθεση του συνολικού συγγραφικού του έργου για λόγους πληρότητας:

1. *Το Σαγανάκι της Τρέλλας*, 1786, στο οποίο ανήκουν τα υπό ανάλυση χειρόγραφα και η επιστημονική διαφωνία που θα αναλύσουμε στην επόμενη υποενότητα.
2. *Φυσικής Απάνθισμα*, Βιέννη 1790. Αποτελεί το πρώτο εθνικό διαφωτιστικό έργο του Ρήγα και την πρώτη εκλαϊκευμένη έκδοση φυσικής στην εποχή του Νεοελληνικού Διαφωτισμού. Βασίστηκε στην Γαλλική Εγκυκλοπαίδεια των Denis Diderot και Jean le Rond d'Alembert, για την καταπολέμηση της δεισιδαιμονίας και των προλήψεων στον ελληνικό χώρο.

3. *Σχολείον των ντελικάτων Εραστών*, Βιέννη 1790. Πρόκειται για διασκευή έξι ερωτικών διηγημάτων του Γάλλου Restif de la Bretonne [35]
4. *Ο Στρατηγός Κεβενχύλλερ ή Στρατιωτικόν Εγκόλπιον*. Πρόκειται για ένα εγχειρίδιο για τη στρατιωτική εκπαίδευση των Ελλήνων επαναστατών, καθώς ο Ρήγας οραματιζόταν να αρχίσει την επανάστασή του από τα οργανωμένα, τότε, στρατιωτικά μέρη της Μάνης και του Σουλίου και στη συνέχεια να προχωρούσε στην απελευθέρωση των υπολοίπων περιοχών της Ελλάδος και της Βαλκανικής [36].
5. *Χαλκογραφία του Μεγάλου Αλεξάνδρου*, Βιέννη, 1797. Πρόκειται για μια χαλκογραφημένη έντυπη εικόνα διαστάσεων 45×29 εκατοστών, που απεικονίζει την προσωπογραφία του Μεγάλου Αλεξάνδρου πλαισιωμένη από πολεμικές σκηνές και προσωπογραφίες των στρατηγών του.
6. *Νέα Χάρτα της Βλαχίας και μέρους της Τρανσυλβανίας*, Βιέννη, 1797. Είναι ένας από τους χάρτες των Παραδουνάβιων Ηγεμονιών που τύπωσε ο Ρήγας Βελεστινλής σε χάραξη Franz Müller.
7. *Γενική Χάρτα της Μολδαβίας*, Βιέννη 1797. Είναι ο δεύτερος χάρτης των Παραδουνάβιων Ηγεμονιών που τύπωσε ο Ρήγας Βελεστινλής στο τυπογραφείο του Ιακώβου Νιτς σε χάραξη Franz Müller.
8. *Χάρτα της Ελλάδος*, Βιέννη 1796-7. Είναι ένας μεγάλων διαστάσεων χάρτης που απεικονίζει τον ελλαδικό χώρο και την ευρύτερη περιοχή της Βαλκανικής χερσονήσου νότια του Δούναβη. Πρόκειται για ένα από τα πιο σημαντικά έργα του Νεοελληνικού Διαφωτισμού, έργο του Ρήγα Βελεστινλή, και το πιο σημαντικό δείγμα της ελληνικής χαρτογραφίας της προεπαναστατικής περιόδου.
9. *Επιπεδογραφία της Κωνσταντινουπόλεως*, Βιέννη, 1796. Αποτελεί το πρώτο τμήμα της Χάρτας και είχε χαλκογραφηθεί από τον Franz Müller.
10. *Ηθικός Τρίπους*, Βιέννη 1797. Εμπεριέχει τρία έργα από την ιταλική, γαλλική και γερμανική λογοτεχνία: τα δύο πρώτα «Τα Ολύμπια» του Μεταστάσιο και «Η Βοσκοπούλα των Άλπεων» του Μαρμοντέλ μεταφράστηκαν από τον Ρήγα Βελεστινλή, ενώ το τρίτο έργο «Ο Πρώτος Ναύτης» του Γκέσνερ, μεταφράστηκε από τον σύντροφο του Ρήγα, Αντώνιο Κορωνιό [37].
11. *Νέος Ανάχαρσις*, μετάφραση του Voyage du jeune Anacharsis en Grece, του J. Barthelemy, τόμος τέταρτος, μαζί με τον Γεώργιο Βεντότη, Βιέννη 1797. Ο γιατρός Γεώργιος Σακελάριος, μετέφρασε και εξέδωσε τον πρώτο τόμο, Βιέννη 1797.
12. *Νέα Πολιτική Διοικήσεις των κατοίκων της Ρούμελης, της Μικράς Ασίας, των Μεσογείων νήσων και της Βλαχομπογδανίας*, Βιέννη 1797. Πρόκειται για ένα κείμενο στο οποίο, δίχως δυσκολία, γίνονται φανερές οι επιδράσεις του Ευρωπαϊκού φιλελεύθερου Διαφωτισμού όπως αυτές αποτυπώνονται μέσα από την πατριωτική του προσήλωση και αγωνία για την Ελλάδα . Περιλάμβανε τέσσερα τμήματα:

- a. *Επαναστατική Προκήρυξις Υπέρ των νόμων και της πατρίδος*
- b. *Τα Δίκαια του ανθρώπου*, σε 35 άρθρα
- c. *Το Σύνταγμα της Ελληνικής Δημοκρατίας*, σε 124 άρθρα
- d. *Θούριος*.

13. «Αγαθάγγελος» ή «Οπτασία του μακαρίου ιερομονάχου Αγαθαγγέλου», περίπου 1790.

Δεν φέρει την υπογραφή του Ρήγα αλλά πολλοί ερευνητές πιστεύουν ότι αυτός είναι ο εκδότης.

Τονίζεται ότι για κάποια από αυτά τα έργα σώζονται και χειρόγραφες εκδοχές τους μέχρι σήμερα. Σε αυτή την κατηγορία υπάγονται «το Σύνταγμα της Ελλάδος», «ο Θούριος», το «το Φυσικής Απάνθισμα», **χωρίς αυτό να σημαίνει ότι αυτά τα κείμενα εγράφησαν δια χειρός του μεγάλου Ρήγα Φεραίου**. Επίσης, ένα μονάχα χειρόγραφο του «Σχολείου των ντελικάτων εραστών» βρέθηκε στη υπό την κατοχή του κυρίου Μ. Π. Κοσσονή στη Σμύρνη, αλλά είναι προ πολλού εξαφανισμένο [38]. **Το μόνο κείμενο από τα προαναφερθέντα που είναι ιστορικά αδιαμφισβήτητο ότι εγράφη από τον ίδιο το Ρήγα Φεραίο, είναι το «Φυσικής Απάνθισμα»**. Συνεπώς, αυτό το χειρόγραφο κείμενο θα χρησιμοποιήσουμε ως σημείο αναφοράς και σύγκρισης, για να θεμελιώσουμε ότι κάποιο άλλο κείμενο εγράφη δια χειρός του Ρήγα ή κάποιου άλλου.

### **3.3 Τα ανακαλυφθέντα δυο χειρόγραφα στη Ρουμανία**

#### ***ανήκουν στο Ρήγα ή όχι;***

Το 1998 ανακαλύφθηκαν στη Ρουμανία από τη ρουμανικής καταγωγής ελληνίστρια ιστορικό κυρία Lia Brad Chisacof δυο χειρόγραφα με ανέκδοτα κείμενα: το «Σαγανάκι της Τρέλλας», και η «Δοκιμασμένη Φιλία». Η συγκεκριμένη επιστήμονας σύντομα απεφάνθη στο επιλογικό κείμενό της [38] ότι τα χειρόγραφα αυτά ανήκαν στο Ρήγα Φεραίο. Εντούτοις, αρκετοί ειδικοί στον τομέα όπως ο Δρ. Δημήτριος Καραμπερόπουλος [1] εξέφρασαν σημαντικές επιστημονικές διαφωνίες, όπως θα παραθέσουμε αναλυτικότερα κατωτέρω.

Εκκινούμε με μια σύντομη περιγραφή των δυο ανεκδότητων χειρογράφων:

#### *Το σαγανάκι της Τρέλλας*

Υποστηρίχθηκε όπως προαναφέρθηκε ότι πρόκειται για ένα «χαμένο» θεατρικό έργο του Ρήγα που δεν ανέβηκε ποτέ στη σκηνή και παρέμεινε για χρόνια στο χειρόγραφο. Πρόκειται για μια κωμωδία-λίβελο κατά του αμφιλεγόμενου ηγεμόνα της Βλαχίας Νικόλαο Μαυρογένη (1735-1790), έναν άνθρωπο υπερβολικό και φιλοπόλεμο, «αμαθή, αγράμματο, κακοαναθρεμμένο», κοντά στον οποίο εργάστηκε ο Ρήγας ως γραμματέας, όπως αναφέρθηκε στην προηγούμενη ενότητα 3.1. Το έργο βασίζεται στο χιούμορ, και στην πολυγλωσσία, ενώ είναι εμφανής η επιρροή του συγγραφέα από τον Μολιέρο [39].

#### *Η δοκιμασμένη Φιλία*

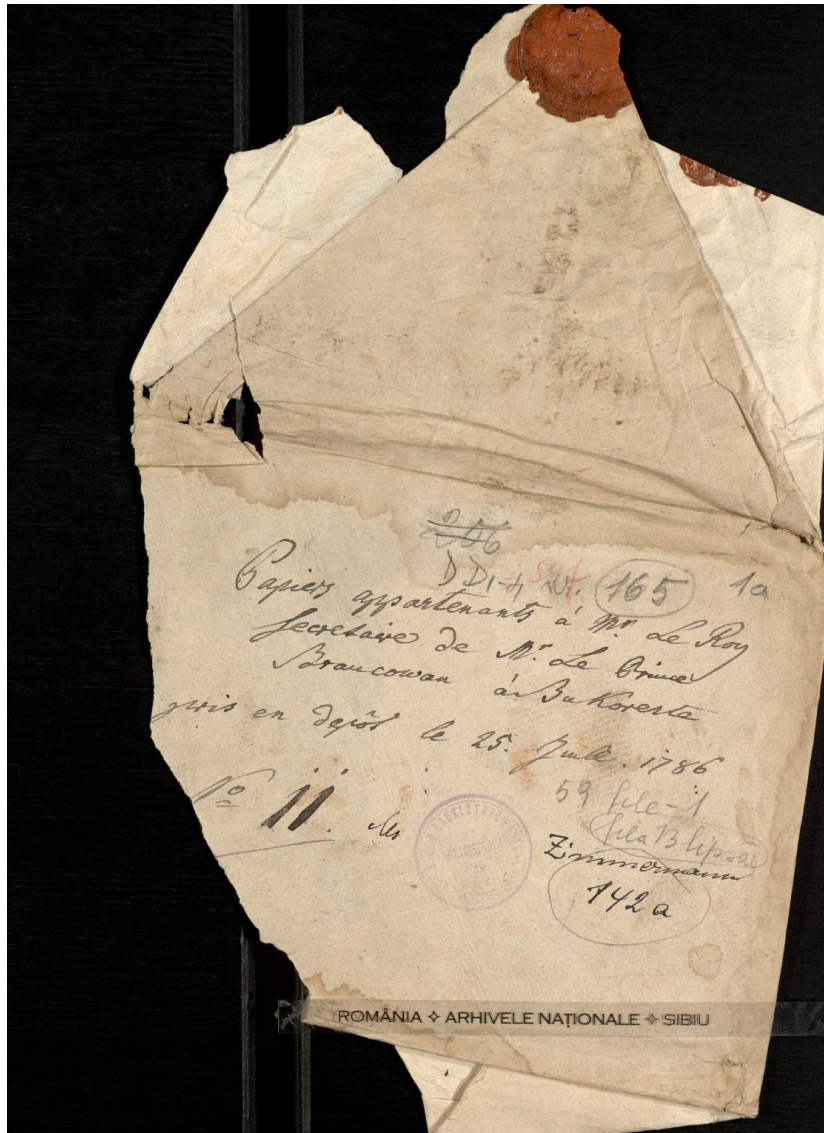
Το αφήγημά «*Η δοκιμασμένη φιλία*» (*L'amitié en épreuve*, 1761), που διασώζεται σε χειρόγραφο με χρονολογία 1785, ανήκει στα «ηθικά διηγήματα» του Jean-Francois Marmontel (1723-1799), ο οποίος ανήκε στο στενό περιβάλλον του Βολταίρου [39].

#### *Οι Επιστημονικές Διαφωνίες για την πιστότητα των κειμένων*

Τα επιχειρήματα για την πιστότητα του χειρογράφου του θεατρικού έργου «Το σαγανάκι της τρέλλας», επανεξετάζει ο Δρ. Καραμπερόπουλος στο άρθρο του “*Δεν είναι τελικά ο Ρήγας ο συγγραφέας των «Ανέκδοτων κειμένων» με την κωμωδία «Το σαγανάκι της Τρέλλας»*” [1]. Στο συγκεκριμένο άρθρο αντικρούονται οι απόψεις για την πατρότητα του χειρογράφου της κυρίας Chisacof με χρήση τεκμηρίων και στοιχείων. Αναλυτικότερα, στο εισαγωγικό κείμενό της η κυρία Chisacof έχει ειδικό κεφάλαιο με τίτλο «*Η πατρότητα του κειμένου*» (σελ. 31-35), στο οποίο εκθέτει τα επιχειρήματά της, τα οποία όμως αντικρούει ο Δρ. Καραμπερόπουλος στο δικό του άρθρο.



Το πρώτο στοιχείο που διατυπώνει η κυρία Chisacof είναι η σημείωση επάνω στο «φάκελο» των χειρογράφων, με το εξής περιεχόμενο: *Des papiers appartenant a monsieur Le Roy secre'taire de Me le Prince Brancowan a Bucoreste pris en de'po't de 25 Juillet 1786* (Έγγραφα που ανήκουν στο Κύριο Le Roy, γραμματέα του Πρίγκιπα του Μπρινκοβεάνου στο Βουκουρέστι, κατατέθηκαν στις 25 Ιουλίου 1786) (Σχήμα 7). Σημειώνεται ότι ο φάκελος φέρει τη βούλα του αυστριακού προξενείου του Βουκουρεστίου. Το όνομα Le Roy υποτέθηκε ότι αποτελεί μετάφραση του ονόματος του Ρήγα, όχι μόνον επειδή χρησιμοποιείται η γαλλική γλώσσα, αλλά και ίσως επειδή υπήρχε πρόθεση συγκάλυψης. Ωστόσο, σύμφωνα με τον Δρα. Καραμπερόπουλο, δεν υπάρχει άμεση σύνδεση μεταξύ του «φακέλου» και του χειρογράφου, εφόσον δεν πρόκειται καν για φάκελο αλλά για μία κατάλληλα διπλωμένη κόλλα χαρτί, όπως φαίνεται και στο Σχήμα 7. Επίσης, υποστηρίζει ότι δε συνάδουν οι ημερομηνίες, καθώς ο Μαυρογένης έφτασε στο Βουκουρέστι στις 17/28 Μαΐου 1786, οπότε το χρονικό διάστημα των δύο μηνών φαίνεται ότι είναι πολύ μικρό για την αναγνώριση τόσο της προσωπικότητας του Μαυρογένη όσο και την υποστηριζόμενη δυσaréσκεια του συγγραφέα Ρήγα. Τέλος, επισημαίνει ότι το όνομα Le Roy μπορεί να μην αναφέρεται στο πρόσωπο του Ρήγα, αλλά να αναφέρεται στο Γάλλο ναυπηγό με το όνομα Le Roy, ο οποίος το 1786 βρισκόνταν στην Πόλη και θα είχε και έναν προσωπικό λόγο να επιθυμεί τη διάδοση ενός αντιμαυρογενικού έργου, καθώς γνώριζε τον Μαυρογένη από παλιά [1].

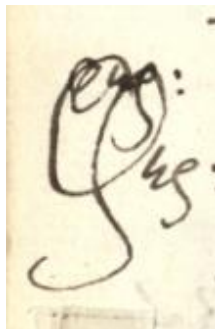


Σχήμα 7: Ο «φάκελος» των χειρογράφων στον οποίο βασίστηκε το πρώτο επιχείρημα της κυρίας Chisacof. Είναι εμφανές ότι όντως πρόκειται για μια σελίδα χαρτί, που έχει διπλωθεί κατάλληλα στη μορφή ενός φακέλου.

Στη συνέχεια, ένα σημαντικό επιχείρημα της κυρίας Chisacof ήταν ο γραφικός χαρακτήρας των χειρογράφων, ο οποίος είναι όμοιος με τον αντίστοιχο ενός χειρογράφου του Ρήγα, της Εθνικής Βιβλιοθήκης Ελλάδος 1788. Ωστόσο, σύμφωνα με τον Δρα. Καραμπερόπουλο εκείνη την εποχή δεν ήταν σπάνιο να μοιάζουν οι γραφικοί χαρακτήρες, οπότε δεν είναι δυνατόν η παλαιογραφική παράμετρος να παίζει μεγάλο ρόλο στην ταυτοποίηση του χειρογράφου[1].

Επίσης, δε θα πρέπει να παραλειφθεί το επιχείρημα της κυρίας Chisacof ως προς τον χαρακτήρα του Φεραρή, που θεωρεί ότι είναι το alter ego του συγγραφέα. Αρχικά, εστιάζει στην συντομογραφία του ονόματος Φεραρή και συγκεκριμένα παρατηρεί ότι στο φύλλο 49ν του χειρογράφου γράφθηκε με κεφαλαίο (Φερ) αφού προηγουμένως είχε γραφθεί με πεζά γράμματα (φερ) και σβήστηκε (Σχήμα 8). Το γεγονός αυτό το αποδίδει σε συγκαλυμμένες

προσπάθειες για την απόκρυψη της ταυτότητας του συγγραφέα. Όμως, σύμφωνα με τον Δρ. Καραμπερόπουλο, το όνομα «Φεραρής» δεν είναι το alter ego του συγγραφέα, όπως υποστηρίζεται από την κ. Chisacof, αλλά σχετίζεται με τον «Γιαννακάκη Τζανέτο», ο οποίος είχε καταφύγει στους Ρώσους (φιραρής=φυγάς). Στη συνέχεια, η κυρία Chisacof υποστηρίζει ότι το εν λόγω όνομα θυμίζει το τοπωνύμιο Φερραί, τη γενέτειρα του Ρήγα. Ωστόσο, σύμφωνα με τον Δρ. Καραμπελόπουλο ο Ρήγας ποτέ του δεν χρησιμοποίησε το αρχαίο όνομα «Φεραί» της γενέτειράς του Βελεστίνου, καθώς σε όλα τα έργα του και τα έγγραφα πάντοτε υπέγραφε ως «Ρήγας Βελεστινλής», και ποτέ ως «Φεραίος» [1].



*Σχήμα 8: Η διόρθωση του αρχικού γράμματος στην οποία βασίστηκε το τρίτο επιχείρημα της κυρίας Chisacof*

Τέλος, η κυρία Chisacof υπογραμμίζει τη σημασία της χρήσης συγκεκριμένων επιστημονικών όρων, όπως η λέξη «ηλεκτρισμός», καθώς ο όρος αυτός παραπέμπει στο έργο του Ρήγα «Φυσικής Απάνθισμα». Όμως, σύμφωνα με τον Δρα. Καραμπερόπουλο, οι συγκεκριμένοι όροι δε δύνανται να αποτελέσουν στοιχείο ταυτότητας του έργου, επειδή ο ηλεκτρισμός ήταν κάτι γνωστό αρκετά χρόνια πριν τη συγγραφή του έργου του Ρήγα «Φυσικής Απάνθισμα» [1].

Με αφορμή την παραπάνω επιστημονική αντιπαράθεση, εξετάστηκε στην παρούσα διπλωματική εργασία ποιο χέρι έγραψε τα εν λόγω χειρόγραφα. Αναλυτικότερα, έγινε αυτόματη κατάτμηση και εξαγωγή περιγραμμάτων των γραμμάτων «π» και «ε» από συγκεκριμένο δείγμα των σελίδων των χειρογράφων, με σκοπό αυτά να τύχουν επεξεργασίας, με πρωτότυπες μεθόδους, από την επιστημονική ομάδα του Αρχιμήδη, τα μέλη της οποίας αναγράφονται στην αντίστοιχη ιστοσελίδα<sup>1</sup>.

---

<sup>1</sup> <http://archimedes.ece.ntua.gr/index.php/members>

# 4

## Μαθηματικό Υπόβαθρο

Για τη βέλτιστη περιγραφή της χρησιμοποιούμενης μεθοδολογίας ταυτοποίησης γραφέα των χειρογράφων, κρίνεται σκόπιμη η παρουσίαση των βασικών μαθηματικών εννοιών που θα χρησιμοποιηθούν στα επόμενα Κεφάλαια.

### 4.1 Στατιστικές Κατανομές

#### 4.1.1 Κανονική Κατανομή

Μια τυχαία μεταβλητή  $X$  θεωρούμε ότι ακολουθεί κανονική κατανομή με μέση τιμή  $\mu$  και διασπορά  $\sigma^2$ , δηλαδή  $X \sim (\mu, \sigma^2)$ , όταν η συνάρτηση πυκνότητας πιθανότητας (σππ)  $f(x)$  είναι της μορφής:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, x \in \mathfrak{R}$$

Όπου, για την  $f(x)$  έχει αποδειχθεί ότι

$$\int_{-\infty}^{+\infty} f(x) dx = 1$$

#### Τυπική Κανονική Κατανομή

Μια τυχαία μεταβλητή που ακολουθεί την τυπική κανονική κατανομή, έχει μέση τιμή 0 ( $\mu=0$ ) και διασπορά 1 ( $\sigma^2=1$ ), συνήθως συμβολίζεται με  $Z$ , δηλαδή  $Z \sim N(0,1)$  και η συνάρτηση πυκνότητάς της με  $\varphi(z)$ ,  $-\infty < z < +\infty$

Με τον τύπο  $Z = \frac{X - \mu}{\sigma}$  μπορούν να μετατραπούν τα δεδομένα μιας μεταβλητής που ακολουθεί τυχούσα κανονική κατανομή σε τυπική και να υπολογιστούν οι σχετικές πιθανότητες για την  $X$  χρησιμοποιώντας τους πίνακες της τυπικής κανονικής κατανομής. Οι πίνακες αυτοί δίνουν τις αθροιστικές πιθανότητες της κατανομής της  $Z$ , δηλαδή το εμβαδόν της κατανομής από  $-\infty$  έως και μία συγκεκριμένη τιμή της  $Z$ . Κατ' αυτόν τον τρόπο, οι πίνακες δίνουν τις πιθανότητες  $P(Z \leq z)$  για όλα τα  $z$  από  $-3,99$  έως  $+3,99$  με βήμα  $0,01$ .

Επίσης, όπως προαναφέρθη, μέσω της τυπικής κατανομής μπορούμε να υπολογίσουμε τις πιθανότητες οποιασδήποτε κανονικής κατανομής, καθώς η γνώση των πιθανοτήτων μιας τυπικής κατανομής  $Z$  παρέχει αμέσως τη γνώση των πιθανοτήτων μιας κανονικής κατανομής  $X$  μέσω του τύπου:

$$P(X \leq t) = \int_{-\infty}^t f(x) dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^t e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \stackrel{z = \frac{x-\mu}{\sigma}}{=} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{t-\mu}{\sigma}} e^{-\frac{z^2}{2}} dz = P\left(Z \leq \frac{t-\mu}{\sigma}\right)$$

#### 4.1.2 Κατανομή $\chi^2$

Έστω  $n$  ανεξάρτητες τυπικές κανονικές τυχαίες μεταβλητές  $Z_1, Z_2, \dots, Z_n$ , δηλαδή,  $Z_i \sim N(0,1)$ , όπου  $i = 1, 2, \dots, n$ , τότε η κατανομή της τυχαίας μεταβλητής

$$X = Z_1^2 + Z_2^2 + \dots + Z_n^2$$

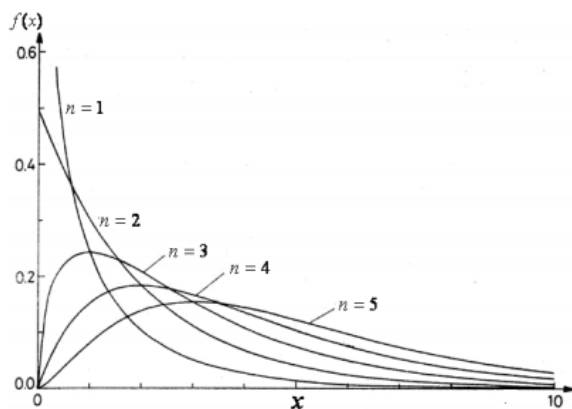
ονομάζεται *κατανομή  $\chi$ -τετράγωνο με  $n$  βαθμούς ελευθερίας* και συμβολίζεται με  $\chi_n^2$ .

Δεδομένου ότι πρόκειται για οικογένεια κατανομών, για κάθε τιμή του  $n$  παίρνουμε και μια άλλη κατανομή  $\chi^2$ . Είναι επίσης προφανές ότι τυχαία μεταβλητή  $X$  που ακολουθεί κατανομή  $\chi_n^2$  δεν παίρνει ποτέ αρνητικές τιμές. Η συνάρτηση πυκνότητας πιθανότητας της εν λόγω κατανομής ορίζεται μέσω της συνάρτησης *Γάμα*, δηλαδή

$$f_{\chi^2}(x) = \begin{cases} \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

$$\text{όπου } \Gamma(n) = \int_0^{\infty} x^{n-1} e^{-x} dx, (n > 0)$$

Στο Σχήμα 9 φαίνεται η γραφική παράσταση της συνάρτησης πυκνότητας πιθανότητας της  $X = Z_1^2 + Z_2^2 + \dots + Z_n^2$  για διάφορες τιμές του βαθμού ελευθερίας  $n$ .



Σχήμα 9

#### 4.1.3 Κατανομή $t$ ή Student

Έστω  $Z$  μια τυχαία μεταβλητή η οποία ακολουθεί τυπική κανονική κατανομή, δηλαδή  $Z \sim N(0,1)$ , και  $S_n$  μια τυχαία μεταβλητή ανεξάρτητη από την  $Z$  η οποία ακολουθεί την κατανομή  $\chi_n^2$ , δηλαδή  $S_n \sim \chi_n^2$ . Τότε, η κατανομή της τυχαίας μεταβλητής  $T$ ,

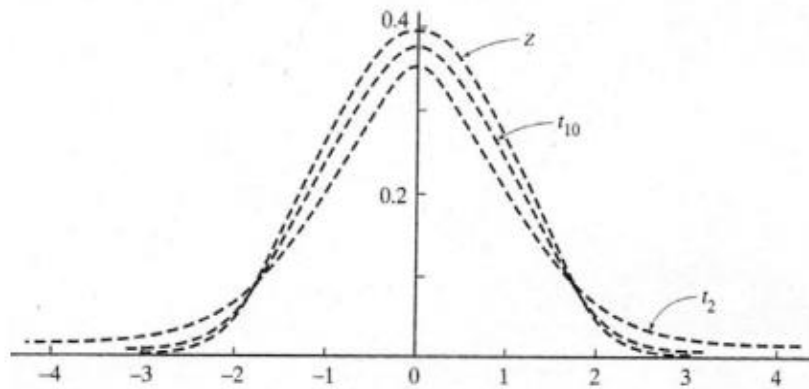
$$T = \frac{Z}{\sqrt{\frac{S_n}{n}}}$$

ονομάζεται *κατανομή  $t$  ή κατανομή Student με  $n$  βαθμούς ελευθερίας* και συμβολίζεται με  $t_n$ . Δεδομένου ότι πρόκειται για οικογένεια κατανομών, για κάθε τιμή του  $n$  παίρνουμε και μια άλλη κατανομή  $t$ . Είναι επίσης προφανές ότι μια τυχαία μεταβλητή  $T$  που ακολουθεί μια  $t_n$  κατανομή παίρνει τιμές από  $-\infty$  έως  $+\infty$  όπως η κανονική κατανομή. Η συνάρτηση πυκνότητας πιθανότητας της εν λόγω κατανομής ορίζεται μέσω της συνάρτησης *Γάμα*, δηλαδή

$$f_{t,n}(x) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{\pi n} \Gamma(\frac{n}{2})} \left(1 + \frac{x^2}{n}\right)^{-(n+1)/2} \quad (-\infty < x < \infty)$$

$$\text{όπου } \Gamma(n) = \int_0^{\infty} x^{n-1} e^{-x} dx, (n > 0)$$

Επίσης, στο Σχήμα 10 φαίνεται η γραφική παράσταση της συνάρτησης πυκνότητας της  $Z \sim N(0,1)$  και η γραφική παράσταση της συνάρτησης πυκνότητας της  $T \sim t_n$  για  $n = 2$  και για  $n = 10$ .



Σχήμα 10

## 4.2 Έλεγχος Στατιστικών Υποθέσεων

### 4.2.1 Υποθέσεις για τη μέση τιμή ενός πληθυσμού από ένα δείγμα του

#### 4.2.1.1 Όταν η πληθυσμιακή διασπορά είναι γνωστή

Έστω ότι έχουμε έναν κανονικό πληθυσμό με γνωστή διασπορά από τον οποίο δειγματοληπτούμε τυχαία  $N$  άτομα, οπότε η μέση τιμή δείγματος είναι  $\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$ .

Υποθέτουμε ότι η κανονική κατανομή από την οποία έγινε η δειγματοληψία έχει μέση πληθυσμιακή τιμή  $\mu$ . Ελέγχουμε την υπόθεση αυτή με το εξής κριτήριο ελέγχου:

#### Κριτήριο Ελέγχου Υπόθεσης 1

Δεδομένης μιας υπόθεσης  $H_0$  : η μέση τιμή του πληθυσμού είναι  $\mu$

και της εναλλακτικής της  $H_1$  : η μέση τιμή του πληθυσμού δεν είναι  $\mu$

Εξετάζουμε την ορθότητα της  $H_0$  υπολογίζοντας την ποσότητα:

$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{N}}}$$

και με βάση δύο προκαθορισμένες τιμές του πεδίου ορισμού της τυπικής κανονικής κατανομής  $-t_a$  και  $t_a$ , ελέγχουμε αν το  $Z$  ανήκει στο ανοικτό διάστημα  $(-t_a, t_a)$ . Στην

περίπτωση που ανήκει, η υπόθεσή μας θεωρείται *ορθή* με βαθμό εμπιστοσύνης

$$P = \int_{-t_a}^{t_a} f(z) dz, \text{ όπου } f(z) \text{ η σππ της τυπικής κανονικής κατανομής. Στην περίπτωση που το } Z$$

δεν ανήκει στο διάστημα  $(-t_a, t_a)$ , η  $H_0$  κρίνεται *απορριπτέα* και η  $H_1$  *ορθή* με τον ίδιο βαθμό εμπιστοσύνης. Η πιθανότητα να συμβεί η τελευταία περίπτωση ορίζεται ως επίπεδο σημαντικότητας  $\alpha$  και δίνεται από τον κάτωθι τύπο:

$$\alpha = 2 \int_{t_a}^{\infty} f(z) dz$$

Επομένως, όσο μικρότερο είναι το  $\alpha$ , τόσο πιθανότερο είναι η υπόθεση  $H_0$  να μην ισχύει. Μερικές τυπικές τιμές του  $\alpha$  είναι 0.05, 0.01, 0.001, και 0.0001.

#### 4.2.1.2 Όταν η πληθυσμιακή διασπορά δεν είναι γνωστή

Έστω ότι έχουμε έναν κανονικό πληθυσμό με άγνωστη διασπορά από τον οποίο δειγματοληπτούμε τυχαία  $N$  άτομα, οπότε η μέση τιμή δείγματος  $\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$ . Υποθέτουμε και σε αυτή την περίπτωση ότι η κανονική κατανομή από την οποία έγινε η δειγματοληψία έχει μέση πληθ τιμή  $\mu$ . Ελέγχουμε την υπόθεση αυτή με το εξής κριτήριο ελέγχου:

##### Κριτήριο Ελέγχου Υπόθεσης 2

Δεδομένης μιας υπόθεσης  $H_0$ : η μέση τιμή του πληθυσμού είναι  $\mu$

και της εναλλακτικής της  $H_1$ : η μέση τιμή του πληθυσμού δεν είναι  $\mu$

Εξετάζουμε την ορθότητα της  $H_0$ , υπολογίζοντας της τυπικής απόκλισης του δείγματος

$$S = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2} \text{ και κατόπιν την ποσότητα:}$$

$$t = \frac{\bar{x} - \mu}{\frac{S}{\sqrt{N}}}$$

Στη συνέχεια, με βάση δύο προκαθορισμένες τιμές  $-t_a$  και  $t_a$  του πεδίου ορισμού της κατανομής Student με  $N - 1$  βαθμούς ελευθερίας, ελέγχουμε αν το  $t$  ανήκει στο διάστημα  $(-t_a, t_a)$ . Στην περίπτωση που ανήκει, η υπόθεσή μας θεωρείται *ορθή*. Στην περίπτωση που το  $Z$  δεν ανήκει στο διάστημα  $(-t_a, t_a)$ , η  $H_0$  κρίνεται *απορριπτέα* και η  $H_1$  *ορθή*. Η πιθανότητα



να συμβεί η τελευταία περίπτωση είναι και πάλι ίση με  $a = 2 \int_{t_a}^{\infty} f(t) dt$ , όπου  $f(t)$  η σππ της κατανομής Student.

#### 4.2.2 Υποθέσεις που αφορούν τη διαφορά των μέσων δύο πληθυσμών

##### 4.2.2.1 Όταν η πληθυσμιακές διασπορές είναι γνωστές

Έστω ότι έχουμε δύο κανονικούς πληθυσμούς  $P_1 \sim N(\mu_1, \sigma_1^2)$  και  $P_2 \sim N(\mu_2, \sigma_2^2)$  με άγνωστες μέσες τιμές  $\mu_1$  και  $\mu_2$  αντίστοιχα και γνωστές τυπικές αποκλίσεις  $\sigma_1$  και  $\sigma_2$ . Από τους πληθυσμούς αυτούς δειγματοληπτούμε τυχαία  $N_1$  και  $N_2$  άτομα με μέση τιμή δειγμάτων  $\bar{x} = \frac{1}{N_1} \sum_{i=1}^{N_1} x_i$  και  $\bar{y} = \frac{1}{N_2} \sum_{i=1}^{N_2} y_i$ . Υποθέτουμε ότι οι δύο κανονικοί πληθυσμοί έχουν διαφορά μέσων τιμών ίση με μια γνωστή ποσότητα  $\delta\mu = \mu_1 - \mu_2$ . Για την υπόθεση αυτή υπάρχει ένα κριτήριο ελέγχου ορθότητας που παρατίθενται κατωτέρω:

##### Κριτήριο Ελέγχου Υπόθεσης Διαφοράς Μέσων Κανονικών Πληθυσμών

Δεδομένης μιας υπόθεσης  $H_0$ : οι δύο πληθυσμοί έχουν διαφορά μέσων τιμών  $\delta\mu$  και της εναλλακτικής της  $H_1$ : οι δύο πληθυσμοί έχουν διαφορά μέσων τιμών διαφορετική της  $\delta\mu$

Εξετάζουμε την ορθότητα της  $H_0$  υπολογίζοντας την ποσότητα:

$$Z = \frac{(\bar{x} - \bar{y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}}$$

αφού θέσουμε και με βάση δύο προκαθορισμένες τιμές του πεδίου ορισμού της τυπικής κανονικής κατανομής  $-t_a$  και  $t_a$ , ελέγχουμε αν το  $Z$  ανήκει στο διάστημα  $(-t_a, t_a)$ . Στην περίπτωση που ανήκει, η υπόθεσή μας θεωρείται *ορθή*. Στην περίπτωση που το  $Z$  δεν ανήκει στο διάστημα  $(-t_a, t_a)$ , η  $H_0$  κρίνεται *απορριπτέα* και η  $H_1$  *ορθή*. Η πιθανότητα να συμβεί η

τελευταία περίπτωση είναι πάλι  $a = 2 \int_{t_a}^{\infty} f(z) dz$ , όπου  $f(z)$  η σππ της τυπικής κανονικής κατανομής.

#### 4.2.2.2 Όταν η πληθυσμιακές διασπορές δεν είναι γνωστές

Έστω ότι έχουμε δύο κανονικούς πληθυσμούς  $P_1 \sim N(\mu_1, \sigma_1^2)$  και  $P_2 \sim N(\mu_2, \sigma_2^2)$  με άγνωστες μέσες τιμές  $\mu_1$  και  $\mu_2$ , και άγνωστες τυπικές αποκλίσεις  $\sigma_1$  και  $\sigma_2$ . Από τους πληθυσμούς αυτούς δειγματοληπτούμε τυχαία  $N_1$  και  $N_2$  άτομα με μέση τιμή δειγμάτων  $\bar{x} = \frac{1}{N_1} \sum_{i=1}^{N_1} x_i$  και  $\bar{y} = \frac{1}{N_2} \sum_{i=1}^{N_2} y_i$  αντίστοιχα. Υποθέτουμε ότι οι δύο κανονικοί πληθυσμοί έχουν διαφορά μέσων τιμών ίση με μια γνωστή ποσότητα  $\delta\mu = \mu_1 - \mu_2$ . Για την υπόθεση αυτή υπάρχει ένα κριτήριο ελέγχου ορθότητας που παρατίθενται κατωτέρω:

#### Κριτήριο Ελέγχου Υπόθεσης Διαφοράς Δύο Μέσων Κανονικών Πληθυσμών Με Άγνωστες Διασπορές

Δεδομένης μιας υπόθεσης  $H_0$ : οι δύο πληθυσμοί έχουν διαφορά μέσων τιμών  $\delta\mu$  και της εναλλακτικής της  $H_1$ : οι δύο πληθυσμοί έχουν διαφορά μέσων τιμών διαφορετική του  $\delta\mu$

Εξετάζουμε την ορθότητα της  $H_0$  υπολογίζοντας την ποσότητα:

$$t = \frac{(\bar{x} - \bar{y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{N_1} + \frac{S_2^2}{N_2}}}$$

όπου  $S_1 = \sqrt{\frac{1}{N_1 - 1} \sum_{i=1}^{N_1} (x_i - \bar{x})^2}$  και  $S_2 = \sqrt{\frac{1}{N_2 - 1} \sum_{i=1}^{N_2} (y_i - \bar{y})^2}$ , οι τυπικές αποκλίσεις των δύο δειγμάτων.

Η ποσότητα  $t$  ακολουθεί κατανομή Student με  $n$  βαθμούς ελευθερίας, που ισούται με το ακέραιο μέρος της ποσότητας:

$$\frac{\left(\frac{S_1^2}{N_1} + \frac{S_2^2}{N_2}\right)^2}{\frac{\left(\frac{S_1^2}{N_1}\right)^2}{N_1 - 1} + \frac{\left(\frac{S_2^2}{N_2}\right)^2}{N_2 - 1}}$$

Με βάση δύο προκαθορισμένες τιμές  $-t_a$  και  $t_a$  του πεδίου ορισμού της κατανομής Student με  $N - 1$  βαθμούς ελευθερίας, ελέγχουμε αν το  $t$  ανήκει στο διάστημα  $(-t_a, t_a)$ . Στην περίπτωση που ανήκει, η υπόθεσή μας θεωρείται *ορθή*. Στην περίπτωση που το  $Z$  δεν ανήκει στο

διάστημα  $(-t_a, t_a)$ , η  $H_0$  κρίνεται απορριπτέα και η  $H_1$  ορθή. Η πιθανότητα να συμβεί η τελευταία περίπτωση είναι πάλι  $\alpha = 2 \int_{t_a}^{\infty} f(t) dt$ , όπου  $f(t)$  η σππ της κατανομής Student.

### 4.2.3 Έλεγχος Υποθέσεων Κατανομής με τεστ Kolmogorov-Smirnoff

Έστω ένα σύνολο μετρήσεων μιας ιδιότητας ενός πληθυσμού, στο οποίο η μέτρηση  $\chi_1$  επιβεβαιώθηκε  $v_1$  φορές, η μέτρηση  $\chi_2$  επιβεβαιώθηκε  $v_2$  φορές, και η μέτρηση  $\chi_n$  επιβεβαιώθηκε  $v_n$  φορές. Θεωρώντας ότι τα  $\chi_1, \chi_2, \dots, \chi_n$  είναι διατεταγμένα υποθέτουμε ότι οι τιμές της συγκεκριμένης ιδιότητας του πληθυσμού ακολουθούν μια κατανομή με γνωστή συνάρτηση κατανομής  $\Phi_t$ . Για να ελέγξουμε την εν λόγω υπόθεση εφαρμόζουμε το Τεστ Kolmogorov-Smirnoff, σύμφωνα με το οποίο ορίζουμε τις σχετικές συχνότητες εμφάνισης των  $\chi_i$  ( $f_i$ ), και τις αθροιστικές σχετικές συχνότητες εμφάνισης  $F_i$  ως:

$$f_i = \frac{v_i}{\sum_{k=1}^n v_k} \text{ και } F_i = \sum_{j=1}^i f_j, i=1,2,\dots,n$$

Στη συνέχεια, υπολογίζουμε την τιμή της θεωρητικής συνάρτησης κατανομής  $\Phi(\chi_i)$  για κάθε τιμή μέτρησης  $\chi_i$  και κατόπιν την ακολουθία διαφορών

$$D_i = |F_i - \Phi(\chi_i)|, i=1,2,\dots,n$$

Ορίζουμε το  $D_0$  ως το ελάχιστο άνω φράγμα των υπολογισθέντων διαφορών και συγκεκριμένα

$$D_0 = \sup\{D_i, i=1,2,\dots,n\}$$

Δεδομένης μιας υπόθεσης  $H_0$ : Η ιδιότητα του συγκεκριμένου πληθυσμού έχει συνάρτηση κατανομής  $\Phi_t$

Ελέγχουμε αν  $D_0 < D_{n,\alpha}$ , όπου  $D_{n,\alpha}$  είναι σταθερές, των οποίων οι τιμές καθορίζονται από το πλήθος του δείγματος  $N = \sum_{j=1}^n v_j$  και από το επίπεδο σημαντικότητας  $\alpha$ .

Στην περίπτωση που ισχύει  $D_0 \geq D_{n,\alpha}$ , η  $H_0$  απορρίπτεται.

# 5

## *Περιγραφή Προτεινόμενης Μεθοδολογίας*

### **5.1 Κατάτμηση Εικόνας**

#### **5.1.1 Σημασία Κατάτμησης Εικόνας στην Παρούσα Εφαρμογή**

Όπως αναλύθηκε στο Κεφάλαιο 1, στο πλαίσιο της προ-επεξεργασίας των δεδομένων υπεισέρχεται πλήθος παραγόντων που αλλοιώνουν την ποιότητα των υπό μελέτη εικόνων, δεδομένου ότι αυτές αποτελούν ψηφιακές φωτογραφίες χειρόγραφου κειμένου. Προκειμένου να εξαγάγουμε την επιθυμητή πληροφορία από τις εν λόγω εικόνες, δηλαδή το σχήμα του εκάστοτε γράμματος απαιτείται η κατάλληλη επεξεργασία των φωτογραφιών. Απώτερος σκοπός της διαδικασίας αυτής είναι η απομόνωση του κάθε γράμματος από το υπόβαθρο, που στην περίπτωσή μας είναι η σελίδα ή κυρίως το πλαίσιο στο οποίο έχουμε θέσει το γράμμα. Αυτό επιτυγχάνεται μέσω της διαδικασίας της *κατάτμησης εικόνας*, η οποία πραγματοποιείται μέσω διαφόρων τεχνικών, εκ των οποίων θα αναλύσουμε τις κυριότερες για λόγους πληρότητας.

#### **5.1.1 Το Πρόβλημα της Κατάτμησης Εικόνας**

Το πρόβλημα της *κατάτμησης εικόνας* αποτελεί σημαντικό αντικείμενο μελέτης στους τομείς της Ψηφιακής Επεξεργασίας Εικόνας και στην Όραση Υπολογιστών. Πρόκειται για τη διαδικασία διαχωρισμού μιας ψηφιακής εικόνας σε πολλά τμήματα, δηλαδή σε σύνολα εικονοστοιχείων γνωστά και ως αντικείμενα της εικόνας ή *περιοχές ενδιαφέροντος*. Ο

διαχωρισμός αυτός επιτυγχάνεται ορίζοντας ένα σύνολο κριτηρίων ομοιότητας, με βάση τα οποία ομαδοποιούνται τα εικονοστοιχεία που παρουσιάζουν ισχυρό βαθμό συσχέτισης.

Ο στόχος της κατάτμησης είναι να απλοποιήσει ή και να αλλάξει την αναπαράσταση μιας εικόνας σε κάτι που είναι πιο ουσιαστικό και πιο εύκολο να αναλυθεί.

Βασική συνιστώσα στη διαδικασία της κατάτμησης αποτελεί η ομαδοποίηση των δεδομένων (data clustering), η οποία χρησιμοποιείται επίσης και σε πολλές άλλες συνθέστερες εφαρμογές, όπως η αναζήτηση σε μια βάση δεδομένων. Οι τεχνικές ομαδοποίησης διακρίνονται σε Διαχωριστικές Τεχνικές και Ιεραρχικές Τεχνικές, οι οποίες θα αναλυθούν στις επόμενες υπό-ενότητες.

### **5.1.2 Μέθοδοι Κατάτμησης Εικόνας με βάση Διαχωριστικές Τεχνικές**

Οι εν λόγω μέθοδοι διακρίνονται σε *Μεθόδους κέντρου βάρους (centroid based techniques)* και *Τεχνικές μεσαίας τιμής (medoid based techniques)*.

Στην πρώτη κατηγορία αποδίδονται σημεία σε μια υποπεριοχή, όταν ελαχιστοποιείται η απόστασή τους (χρωματική, ευκλείδεια κλπ.) από το κέντρο βάρους της υποπεριοχής αυτής. Βασικός αντιπρόσωπος της τεχνικής αυτής είναι η *μέθοδος K-means*, στην οποία επιλέγονται τυχαία K περιοχές και αποδίδονται σε κάθε μία αυτές σημεία της αρχικής περιοχής με βάση το προαναφερθέν κριτήριο, με δυναμική αναπροσαρμογή της θέσης του κέντρου βάρους, με βάση κάθε νεοεισελθέν εικονοστοιχείου. Δε θα πρέπει να παραλειφθεί και ο αλγόριθμος ISODATA, που ακολουθεί παρόμοια λογική.

Στη δεύτερη κατηγορία χρησιμοποιούνται σημεία αντιπρόσωποι (medoids) αντί για κέντρα βάρους (centroid), έτσι ώστε να ελαχιστοποιείται η μέση διαφορά τους από τα υπόλοιπα σημεία της ομάδας (cluster), όπου ως διαφορά εννοούμε τον βαθμό **μη** ομοιότητας μεταξύ των δεδομένων.

### **5.1.3 Μέθοδοι Κατάτμησης Εικόνας με βάση Ιεραρχικές Τεχνικές**

Σε αντίθεση με τις διαχωριστικές τεχνικές, οι Ιεραρχικές τεχνικές λειτουργούν με βάση τη συγχώνευση διαφορετικών υποπεριοχών. Αναλυτικότερα, αρχικά θεωρούν ότι το κάθε εικονοστοιχείο της εικόνας συνιστά μια ξεχωριστή υποπεριοχή και στη συνέχεια συγχωνεύουν περιοχές που βρίσκονται κοντινότερα με βάση κάποιο κριτήριο. Το κριτήριο αυτό ορίζεται ως ένα *μέτρο ομοιότητας* μεταξύ των σημείων, όπως η χρωματική τους απόσταση. Με βάση το επιλεγθέν μέτρο ομοιότητας κατασκευάζουμε έναν πίνακα

ομοιότητας  $m \times m$  όλων των εικονοστοιχείων μεταξύ τους. Από τον πίνακα αυτό παράγεται σε κάθε βήμα ένας νέος αραιός πίνακας, στον οποίο τα μηδενικά αντιστοιχούν στις ανόμοιες περιοχές και οι μονάδες στις όμοιες αντίστοιχα. Λόγω των αραιών πινάκων, προκύπτει άμεσα ότι η διαδικασία μπορεί να αναπαρασταθεί από αραιούς γράφους, στους οποίους οι κόμβοι είναι τα δεδομένα και οι ακμές είναι οι ομοιότητες μεταξύ τους.

Η προαναφερθείσα διαδικασία επαναλαμβάνεται μέχρις ότου να φτάσουμε σε ένα επιθυμητό αριθμό περιοχών είτε μέχρι να ικανοποιηθεί ένα κατάλληλο κατώφλι διαφοράς μεταξύ των τελικών ακμών που όπως προαναφέραμε εκφράζουν το βαθμό ομοιότητας. Σύμφωνα με όσα ελέγχθησαν ανωτέρω, το γεγονός αυτό κάθε ιεραρχική τεχνική συνιστά μια αναδρομική διαδικασία.

### Ειδικές ιεραρχικές τεχνικές

Ανάλογα με το κριτήριο ομοιότητας που χρησιμοποιείται κάθε φορά έχουμε τις εξής παραλλαγές των ιεραρχικών τεχνικών:

- Η μέθοδος στην οποία ορίζεται σε κάθε cluster ένας αντιπρόσωπος (π.χ. κέντρο βάρους). Αρχικά, ορίζουμε τόσους αντιπροσώπους clusters όσα και τα εικονοστοιχεία της εικόνας και κατόπιν μετράται η ομοιότητα μεταξύ των αντιπροσώπων κάθε cluster. Με βάση αυτό το κριτήριο και κάποιο συγκεκριμένο κατώφλι αποφασίζεται αν θα συγχωνευτούν δύο clusters[3].
- single link method: Κάθε cluster αντιπροσωπεύεται από όλα του τα εικονοστοιχεία και η ομοιότητα μεταξύ δύο clusters καθορίζεται από την ομοιότητα του πιο κοντινού ζεύγους στοιχείων που ανήκουν σε διαφορετικά clusters[3].
- CURE method: Η μέθοδος αυτή είναι ουσιαστικά συνδυασμός των δύο προηγούμενων μεθόδων. Εδώ ορίζονται αντιπρόσωποι, αλλά περισσότεροι από ένας για κάθε cluster. Η ομοιότητα μετράται μεταξύ των ζευγών των αντιπροσώπων των cluster και το κριτήριο συνένωσης δύο περιοχών εξαρτάται από τον βαθμό ομοιότητας των ζευγών αυτών συνολικά που ανήκουν σε διαφορετικά clusters. Σε κάθε συγχώνευση clusters, δυναμικά ορίζονται νέοι αντιπρόσωποι του νέου cluster που προέκυψε. Ακριβώς επειδή ο αλγόριθμος αυτός αποδίδει πολλούς αντιπροσώπους σε κάθε περιοχή, αλλά και γιατί τους συρρικνώνει προς το κέντρο βάρους, αντιμετωπίζει καταστάσεις διαφορετικών μεγεθών και σχημάτων clusters με μεγαλύτερη επιτυχία. Ο αριθμός των αντιπροσώπων εξαρτάται από το μέγεθος, το σχήμα και την παρουσία θορύβου σε κάθε cluster [3].
- group-average method: Η μέθοδος αυτή υποθέτει διασύνδεση όλων των δεδομένων των clusters και υπολογίζει όλες τις αποστάσεις μεταξύ τους. Κανονικοποιεί τις τιμές σχηματίζοντας ένα  $n \times m$ , πίνακα όπου  $n$  είναι ο αριθμός των δεδομένων στη μία

περιοχή και  $m$  ο αντίστοιχος στη δεύτερη, ώστε να αντιμετωπίσει το πρόβλημα που προκύπτει όταν τα clusters έχουν διαφορετικό μέγεθος [3].

- ROCK method: Ο αλγόριθμος αυτός μοιάζει με αυτόν της group-average μεθόδου, κανονικοποιεί όμως με βάση ένα ορισμένο από τον χρήστη κατώφλι, αντί να χρησιμοποιεί ένα πίνακα  $n \times m$  [3].

#### **5.1.4 Μια Νέα Εναλλακτική Μέθοδος Κατάτμησης**

Σε αυτό το κεφάλαιο περιγράφεται μια πρωτότυπη μέθοδος διαχωρισμού εικόνας σε *δύο θεματικές περιοχές*, που παρουσιάστηκε πρώτη φορά στην εργασία [3]. Συνήθως η πρώτη περιοχή περιλαμβάνει την πολύτιμη για μας πληροφορία, στην περίπτωσή μας το γράμμα, ενώ η δεύτερη περιοχή είναι το υπόβαθρο. Η μέθοδος αυτόματης κατάτμησης εικόνας που αναπτύχθηκε αποτελείται από τα βήματα που περιγράφονται κατωτέρω:

##### 1. Την εφαρμογή μορφολογικών φίλτρων για εξάλειψη θορύβου

Το παρόν βήμα ανήκει στη φάση της προ-επεξεργασίας των δεδομένων μας και ξεκινά με την μετατροπή των εικονοστοιχείων της εικόνας σε κλίμακα του γκρι και συγκεκριμένα σε κανονικοποιημένες τιμές φωτεινότητας στο πεδίο  $[0,1]$ . Στη συνέχεια, απαιτείται η εξομάλυνση της εικόνας, η οποία πραγματοποιείται με την εφαρμογή ενός *φίλτρου σμίκρυνσης* και ενός *φίλτρου διαστολής*.

Με την εφαρμογή του πρώτου φίλτρου λαμβάνουμε τη λεγόμενη 'ανοικτή' εκδοχή της εικόνας (open image), στην οποία η απόχρωση του γκρι κάθε εικονοστοιχείου είναι μικρότερη ή ίση της απόχρωσης της αρχικής εικόνας. Δηλαδή η 'ανοικτή' εικόνα είναι ένα κάτω φράγμα της αρχικής σε όλα τα σημεία της.

Σε επόμενο στάδιο, στην ανοικτή εκδοχή της εικόνας, εφαρμόζουμε τους εξής μετασχηματισμούς: αρχικά με κέντρο κάθε σημείο  $M$  της εικόνας σχηματίζουμε μια περιοχή με ακτίνα ένα εικονοστοιχείο και σε αυτή την περιοχή εφαρμόζουμε τη διαδικασία διαστολής, δηλαδή αποδίδουμε στο  $M$  τη μέγιστη απόχρωση στην περιοχή αυτή. Επαναλαμβάνουμε αυτή τη διαδικασία αυξάνοντας συνεχώς την ακτίνα της περιοχής έως ότου υπάρξει ένα τουλάχιστον σημείο και μια αντίστοιχη περιοχή η μέγιστη απόχρωση της οποίας είναι μεγαλύτερη από την απόχρωση γκρι του  $M$  στην αρχική εικόνα. Στο τέλος αυτής της διαδικασίας λαμβάνουμε μια ομαλοποιημένη εκδοχή της αρχικής εικόνας, με ένταση απόχρωσης του γκρι μικρότερη ή ίση αυτής.

## 2. Τη Δημιουργία Ιστογραμμμάτων γκριζών εικόνων

Χρησιμοποιώντας ένα πλαίσιο εικονοστοιχείων κατάλληλα επιλεγμένων διαστάσεων που ολισθαίνει οριζόντια κάθε γραμμή της ομαλοποιημένης εικόνας, δημιουργούμε το ιστόγραμμα διαβάθμισης του γκρι όλων των εικονοστοιχείων του πλαισίου. Στη συνέχεια, χωρίζουμε το διάστημα των τιμών φωτεινότητας του οριζόντιου άξονα σε ίσα υποδιαστήματα. Σε κάθε υποδιάστημα, αποδίδουμε το πλήθος των pixels που έχουν τιμή μεγαλύτερη ή ίση της αρχής του διαστήματος και μικρότερη του τέλους. Ο αριθμός των υποδιαστημάτων αυτών ονομάζεται **διαμέριση** και συνιστά σημαντική παράμετρο στη διαδικασία της κατάτμησης, όπως θα αναλύσουμε στο επόμενο κεφάλαιο.

## 3. Την Προσέγγιση Ομαλοποιημένου Ιστογράμματος

Σε αυτό το βήμα προσεγγίζουμε βέλτιστα με ένα πολυώνυμο, με την έννοια ελαχίστων τετραγώνων, το ιστόγραμμα αποχρώσεων του γκρι κάθε πλαισίου. Προς την κατεύθυνση αυτή επιτυγχάνεται ορίζοντας ένα κατάλληλα μεγάλου βαθμού πολυώνυμο, το οποίο έχει πεδίο ορισμού τις κεντρικές τιμές φωτεινότητας του ιστογράμματος και ως πεδίο τιμών τις κανονικοποιημένες συχνότητες των pixels με τη συγκεκριμένη φωτεινότητα. Ο **βαθμός** του πολυωνύμου που ταιριάζει βέλτιστα στα ζεύγη κανονικοποιημένων συχνότητων και κεντρικών τιμών φωτεινότητας, αποτελεί σημαντική παράμετρο του εν λόγω αλγορίθμου κατάτμησης.

## 4. Την Εύρεση των σημείων καμπής του πολυωνύμου

Παρατηρούμε ότι στα σημεία καμπής του ομαλού πολυωνύμου που προσεγγίζει το ιστόγραμμα της γκριζας εικόνας γίνεται μετάβαση από το σώμα του γράμματος στο υπόβαθρο της εικόνας, καθώς σε αυτά επισυμβαίνει μέγιστη μεταβολή της απόχρωσης του γκρι. Η εύρεση των σημείων καμπής ανάγεται στην εύρεση των σημείων αλλαγής προσήμου της δεύτερης παραγώγου του πολυωνύμου.

Στη συνέχεια, ορίζουμε διαστήματα τιμών φωτεινότητας γύρω από τα σημεία αλλαγής προσήμου της 2<sup>ης</sup> παραγώγου του πολυωνύμου που προσεγγίζει το ιστόγραμμα. Το **εύρος** των διαστημάτων αυτών αποτελεί επίσης μια πολύ σημαντική παράμετρο του αλγορίθμου κατάτμησης, όπως θα δούμε στα αποτελέσματα του αλγορίθμου.

Με βάση τα υπολογισθέντα σημεία καμπής, ελέγχουμε για την φωτεινότητα κάθε σημείου σε ποιο διάστημα τιμών φωτεινότητας αντιστοιχεί. Αν αντιστοιχεί στο πρώτο σημείο αλλαγής προσήμου, τότε αυτό έχει χαμηλή τιμή φωτεινότητας, οπότε μπορούμε να ισχυριστούμε ότι ανήκει στο σώμα του γράμματος. Αν ανήκει στο διάστημα τιμών φωτεινότητας που αντιστοιχεί στο δεύτερο σημείο αλλαγής



προσήμου, τότε το σημείο ανήκει στο όριο μεταξύ του εικονιζόμενου γράμματος και του υποβάθρου. Τέλος, αν ανήκει στο διάστημα τιμών φωτεινότητας που αντιστοιχεί στο τελευταίο σημείο αλλαγής προσήμου, τότε αυτό έχει υψηλή τιμή φωτεινότητας, οπότε μπορούμε να ισχυριστούμε ότι ανήκει στο υπόβαθρο της εικόνας.

5. Την αρχική εκτίμηση των θεματικών περιοχών με βάση τα σημεία καμψής

Ορίζουμε μια δυαδική εικόνα κατάτμησης, όπου στα pixels που ανήκουν στο διάστημα πριν το πρώτο σημείο αλλαγής προσήμου της  $2^{ns}$  παραγώγου αποδίδουμε την τιμή 1, οπότε δημιουργείται η εικόνα της αρχικής κατάτμησης της γκριζας εικόνας σε δύο σύνολα «γράμμα» και «υπόβαθρο».

6. Την εξάλειψη του θορύβου και ομογενοποίηση των θεματικών περιοχών

Με το πέρας του βήματος 5, έχουμε δημιουργήσει δύο περιοχές, από τις οποίες θα προσπαθήσουμε να εξαλείψουμε το 'θόρυβο', δηλαδή να τις ομογενοποιήσουμε με χρήση των κάτωθι μορφολογικών διαδικασιών. Καταρχήν, ορίζεται ως «θόρυβος λόγω φθοράς» ή «decay noise» η τυχαία φθορά χρώματος και υψής που με τον καιρό δημιουργήθηκε σε κάθε ένα από τα χειρόγραφα που μελετώνται. Αυτός ο θόρυβος μπορεί να εμφανίζεται είτε ως μαύρα εικονοστοιχεία μέσα στο σώμα του λευκού γράμματος, είτε ως λευκά εικονοστοιχεία στο μαύρο υπόβαθρο.

Για την εξάλειψη του μαύρου θορύβου ορίζουμε ένα κυλιόμενο πλαίσιο σε κάθε σημείο του συνόλου «γράμμα», δηλαδή των pixels με τιμή 1. Αν οι τιμές των pixels του ορίου του εν λόγω πλαισίου είναι ίσες με 1, θεωρούμε ότι το πλαίσιο ανήκει στο γράμμα, οπότε αποδίδουμε την τιμή 1 σε όποιο σημείο έχει τιμή 0.

Επιπρόσθετα, μπορούμε να εφαρμόσουμε τη συμπληρωματική διαδικασία ομογενοποίησης για την εξάλειψη του λευκού θορύβου. Μπορούμε δηλαδή να ορίσουμε ένα κυλιόμενο πλαίσιο με κέντρο κάθε σημείο υποβάθρου (με τιμή 0) και αν οι τιμές των σημείων του ορίου του εν λόγω πλαισίου είναι ίσες με 0, θεωρούμε ότι το πλαίσιο ανήκει στο υπόβαθρο, οπότε αποδίδουμε την τιμή 0 σε όποιο σημείο έχει τιμή 1.

Τελικά, μετά τη δημιουργία των ομογενοποιημένων περιοχών, που πραγματοποιήθηκε στο βήμα 7, λαμβάνουμε τις τελικές εκδόσεις των δύο περιοχών τόσο του σώματος του γράμματος όσο και του υποβάθρου.

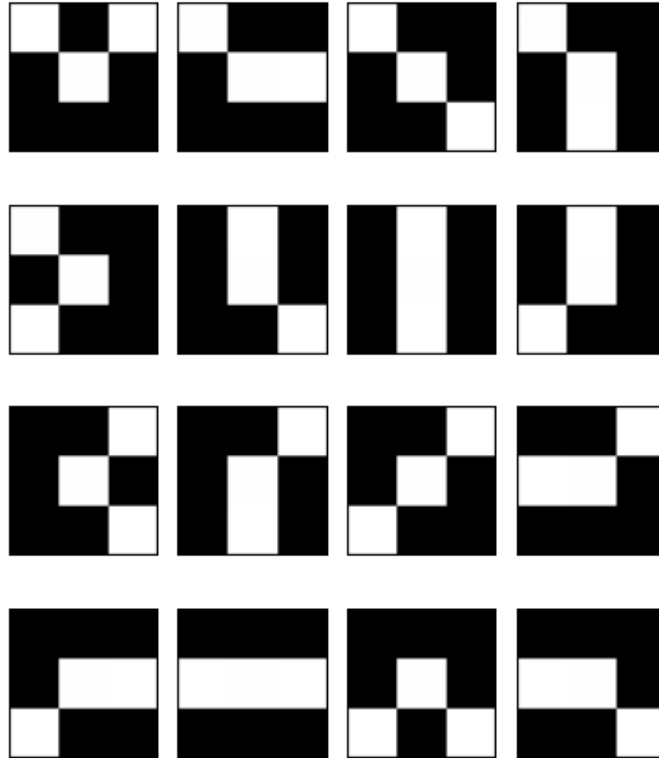
## 5.2 Εξαγωγή Περιγραμμάτων των Υλοποιήσεων των

### Γραμμάτων της Αλφαβήτου

Αφού έχουμε μετατρέψει την αρχική εικόνα του γράμματος από το χειρόγραφο σε δυαδική, δηλαδή έχουμε απομονώσει το σώμα του γράμματος από το υπόβαθρο της σελίδας, το επόμενο λογικό βήμα για την περαιτέρω επεξεργασία του γράμματος μέσω των αλγορίθμων που θα παρουσιαστούν σε επόμενες υποενότητες είναι να εξάγουμε το περίγραμμά του. Τα στάδια εξαγωγής του περιγράμματος είναι τα εξής:

1. Το πρώτο στάδιο εύρεσης περιγράμματος είναι η επιλογή των σημείων που έχουν επακριβώς δύο γειτονικά εικονοστοιχεία, όπου με τον όρο γειτονικά ενός σημείου εννοούμε τα εικονοστοιχεία που περιέχονται σε ένα πλαίσιο  $3 \times 3$  με κέντρο το εν λόγω pixel.
2. Το δεύτερο στάδιο εξαγωγής περιγράμματος είναι η βελτίωση του παρόντος περιγράμματος με την αφαίρεση των συνδυασμών που σχηματίζουν ορθή γωνία, καθώς επιτρέπονται μόνο συγκεκριμένοι σχηματισμοί των τριών εικονοστοιχείων που περιέχονται σε πλαίσιο  $3 \times 3$ , οι οποίοι παρουσιάζονται στο Σχήμα 11.
3. Στο τρίτο στάδιο ορίζουμε ένα κατώφλι κάτω από το οποίο απαγορεύονται οι σχηματισμοί μεμονωμένων συμπλεγμάτων εικονοστοιχείων. Συνήθως αυτό το κατώφλι ορίζεται ως το 5% του συνολικού αριθμού των εικονοστοιχείων του περιγράμματος.

Δε θα πρέπει να παραλειφθεί πως η αρίθμηση των εικονοστοιχείων του περιγράμματος πρέπει να ακολουθεί την ωρολογιακή φορά, δηλαδή η αρίθμηση να ξεκινά από το εικονοστοιχείο που βρίσκεται πιο ψηλά και αριστερά στην εικόνα και να συνεχίζει ωρολογιακά. Τονίζουμε ότι ένα τυχαίο εικονοστοιχείο  $n$  έχει ως γειτονικά του τα εικονοστοιχεία  $n - 1$  και  $n + 1$  σύμφωνα με τους περιορισμούς του Σχήματος 11.



*Σχήμα 11: Επιτρεπτοί συνδυασμοί εικονοστοιχείων. Το άσπρο χρώμα αντιστοιχεί στα εικονοστοιχεία του περιγράμματος*

Επειδή καμία τεχνική κατάτμησης εικόνας δεν μπορεί να προσφέρει τέτοια περιγράμματα, αναπτύχθηκε ειδικός κώδικας και έγινε χρήση πρωτότυπων μορφολογικών μετασχηματισμών, ώστε από την κατατμημένη εικόνα να εξάγεται αυτόματα το περίγραμμα στην επιθυμητή αυτή μορφή.

### 5.3 Υποβέλτιστη Προσαρμογή Περιγραμμάτων

Με βάση τα παραχθέντα περιγράμματα, το πρώτο βήμα για τον έλεγχο των υλοποιήσεων ενός ζεύγους γραμμάτων αναφορικά με την ταυτοποίηση γραφέα είναι η βέλτιστη προσαρμογή των περιγραμμάτων τους και η εκτίμηση του λάθους αυτής της προσαρμογής. Οι δυσκολίες που παρουσιάζονται έγκεινται στο διαφορετικό προσανατολισμό, στη διαφορετική θέση και στο διαφορετικό μέγεθος των υλοποιήσεων των γραμμάτων που παρατηρούνται μεταξύ διαφορετικών ή και του ίδιου χειρογράφου. Για την αντιμετώπιση των προαναφερθεισών δυσκολιών απαιτείται αρχικά να τοποθετήσουμε τα δύο υπό μελέτη περιγράμματα σε κοινό πλαίσιο, ώστε να συμπίπτουν τα κέντρα βάρους τους. Στη συνέχεια, απαιτείται η περιστροφή, η ομοιοθεσία και η παράλληλη μετατόπιση του ενός από τα δύο περιγράμματα, ώστε να βρεθεί στη βέλτιστη θέση προσαρμογής με το άλλο. Για τον εντοπισμό αυτής της θέσης ορίζεται ένα μέτρο απόστασης μεταξύ του περιγράμματος αναφοράς και του μετασχηματισμένου περιγράμματος και την εύρεση της μικρότερης δυνατής τιμής αυτής της απόστασης.

Προκειμένου να πραγματοποιηθεί η παραπάνω διαδικασία αυτόματα για κάθε ζεύγος περιγραμμάτων δύο υλοποιήσεων γραμμάτων, αναπτύχθηκε ειδικός αλγόριθμος που παρουσιάστηκε στην εργασία [1] και τον οποίο αναλύουμε στην παρακάτω Υποενότητα.

#### 5.3.1 Αυτοματοποιημένη Διαδικασία Υποβέλτιστης Προσαρμογής

##### Περιγραμμάτων

Θεωρούμε δύο ισοπληθή σύνολα σημείων  $(x_i, y_i)$  και  $(p_i, q_i)$   $i=1,2,\dots,N$ , που αντιπροσωπεύουν τα κέντρα των συνόλων των εικονοστοιχείων των περιγραμμάτων των δύο υλοποιήσεων των γραμμάτων. Ζητείται η εύρεση κατάλληλης γωνίας  $\varphi^o$ , η μετατόπιση στον οριζόντιο άξονα  $x_s^o$ , η μετατόπιση στον κατακόρυφο άξονα  $y_s^o$  και ο συντελεστής ομοιοθεσίας  $\lambda^o \in \mathbb{R}^+$ , τα οποία όταν δρουν στην καμπύλη  $(p_i, q_i)$  την προσαρμόζουν στην καμπύλη  $(x_i, y_i)$  βέλτιστα κατά την έννοια των ελαχίστων τετραγώνων. Θεωρούμε ότι έχουμε όπως αναλύθηκε στην εισαγωγή, ότι η διαδικασία αυτή συνίσταται στα στάδια της στροφής, της παράλληλης μετατόπισης και τέλος της ομοιοθεσίας.

Για τη στροφή και την παράλληλη μετατόπιση ορίζουμε ως κριτήριο βέλτιστης προσαρμογής την ελαχιστοποίηση της ποσότητας:

$$\varepsilon_{RT}(\varphi, x_s, y_s) = \sum_{i=1}^N \left\{ (x_i - p_i \cos \varphi + q_i \sin \varphi - x_s)^2 + (y_i - p_i \sin \varphi + q_i \cos \varphi - y_s)^2 \right\}$$

όπου  $\varphi$  η γωνία περιστροφής, τα  $x_\delta, y_\delta$  οι παράλληλες μετατοπίσεις κατά τον άξονα  $x$  και  $y$  αντίστοιχα και τα γινόμενα των  $(p_i, q_i)$  προκύπτουν από τον πολλαπλασιασμό με τη μήτρα

$$\text{μετασχηματισμού: } TR = \begin{bmatrix} \cos \varphi & -\sin \varphi \\ \sin \varphi & \cos \varphi \end{bmatrix}$$

Η ελαχιστοποίηση της παραπάνω ποσότητας οδηγεί στη λύση  $(\varphi^O, x_\delta^O, y_\delta^O)$  όπου

$$\tan \varphi^O = \frac{\sum_{i=1}^N \{(x_i - \bar{x})q_i + (y_i - \bar{y})p_i\}}{\sum_{i=1}^N \{(y_i - \bar{y})q_i + (x_i - \bar{x})p_i\}}$$

$$x_\delta^O = \bar{x} - \bar{p} \cos \varphi + \bar{q} \sin \varphi$$

$$y_\delta^O = \bar{y} - \bar{p} \sin \varphi - \bar{q} \cos \varphi$$

όπου τα  $\bar{x}, \bar{y}, \bar{p}, \bar{q}$  είναι τα βαρύκεντρα των αντίστοιχων ψηφιακών καμπυλών:

$$\bar{x} = \sum_{i=1}^N \frac{x_i}{N}, \bar{y} = \sum_{i=1}^N \frac{y_i}{N}, \bar{p} = \sum_{i=1}^N \frac{p_i}{N}, \bar{q} = \sum_{i=1}^N \frac{q_i}{N}$$

Για την εύρεση του βέλτιστου συντελεστή ομοιοθεσίας ελαχιστοποιούμε την ποσότητα:

$$\varepsilon_{RTO}(\lambda) = \sum_{i=1}^N \left\{ (x_i - \lambda \text{sign}(x_i, a - \bar{a})(a_i - \bar{a}) - \bar{a})^2 + (y_i - \lambda \text{sign}(y_i, \beta - \bar{\beta})(\beta_i - \bar{\beta}) - \bar{\beta})^2 \right\}$$

όπου

$$\begin{bmatrix} a_i \\ \beta_i \end{bmatrix} = \begin{bmatrix} \cos \varphi^O & -\sin \varphi^O \\ \sin \varphi^O & \cos \varphi^O \end{bmatrix} \begin{bmatrix} p_i \\ q_i \end{bmatrix} + \begin{bmatrix} x_\delta^O \\ y_\delta^O \end{bmatrix} \text{ και } \bar{a} = \sum_{i=1}^N \frac{a_i}{N}, \bar{\beta} = \sum_{i=1}^N \frac{\beta_i}{N}$$

Η ελαχιστοποίηση της  $\varepsilon_{RTO}(\lambda)$  μας δίνει λύση:

$$\lambda^O = \frac{\sum_{i=1}^N \left\{ |(x_i - \bar{a})(a_i - \bar{a})| + |(y_i - \bar{\beta})(\beta_i - \bar{\beta})| \right\}}{\sum_{i=1}^N \left\{ (a_i - \bar{a})^2 + (\beta_i - \bar{\beta})^2 \right\}}$$

## 5.4 Τελεσίδικο Κριτήριο για την βέλτιστη προσαρμογή δύο υλοποιήσεων

Στα νεότερα χειρόγραφα, ένα γράμμα μπορεί να αποτελείται από μια ή περισσότερες απλές συνεκτικές κλειστές καμπύλες, οι οποίες οφείλονται στην άρση και επανατοποθέτηση του εργαλείου γραφής στην επιφάνεια γραφής. Επομένως, θα ορίσουμε ξεχωριστά κριτήρια για καθεμία από τις προαναφερθείσες περιπτώσεις

### 5.4.1 Ορισμός Κριτηρίου Για Δύο Απλές Συνεκτικές Υλοποιήσεις του Ίδιου

#### Γράμματος

Έστωσαν δύο υλοποιήσεις  $L_1, L_2$  του ίδιου γράμματος που βρίσκονται στο ίδιο ή σε διαφορετικά κείμενα, και  $C_1, C_2$  τα αντίστοιχα εξωτερικά περιγράμματα των υλοποιήσεων αυτών. Στην περίπτωση που τα περιγράμματα  $C_1, C_2$  αποτελούνται από ίσο αριθμό εικονοστοιχείων εφαρμόζουμε τη μεθοδολογία που αναλύθηκε στην Υποενότητα 5.4.1 για μια αρχική καλή προσαρμογή του περιγράμματος  $C_2$  στο  $C_1$ . Εάν τα  $C_1, C_2$  δεν είναι ισοπληθή όσον αφορά τα pixels τους, αφαιρούμε ή συγχωνεύουμε pixels από το πολυπληθέστερο που απέχουν κατάλληλο πλήθος μεταξύ τους. Παραδείγματος χάριν, ας υποθεθεί ότι το  $C_1$  αποτελείται από 50 pixels ενώ το  $C_2$  από 55. Τότε, συγχωνεύοντας δύο ή αφαιρώντας ένα pixel ανά 10 pixels, λαμβάνουμε μια εκδοχή του  $C_2$  ισοπληθή με το  $C_1$ . Τονίζουμε ότι η ισοπληθής με το  $C_1$  εκδοχή του  $C_2$  χρησιμοποιείται **μόνον** για την αρχική εκτίμηση της υποβέλτιστης περιοχής. Επομένως, απαιτείται να ορίσουμε ένα κριτήριο ικανό να ταιριάζει περιγράμματα ανεξάρτητα από τον αριθμό των εικονοστοιχείων τους, αλλά και με ικανοποιητικά αποτελέσματα για τη στατιστική τους μελέτη.

Στο πλαίσιο αυτής της προσπάθειας, βρίσκουμε αρχικά την υποβέλτιστη σχετική θέση των  $C_1, C_2$  με βάση το κριτήριο ελαχίστων τετραγώνων, επιλέγοντας το  $C_1$  ως σταθερό και την ισοπληθή με αυτό εκδοχή  $C_2$  ως μετακινήσιμο περίγραμμα. Στη θέση ελαχιστοποίησης του κριτηρίου κόστους ελαχίστων τετραγώνων ορίζεται τετραδιάστατος κύβος γύρω από το σημείο  $(x_\delta, y_\delta, \varphi, \lambda)$  και μια λεπτή διαμέριση αυτού. Στη συνέχεια, με σημείο έναρξης την υποβέλτιστη προσαρμογή, γίνεται εξαντλητική αναζήτηση στον κύβο, για την εύρεση του σημείου που θα δίνει την καλύτερη σχετική θέση των δύο περιγραμμάτων με βάση το κάτωθι κριτήριο.

Συγκεκριμένα, για την εύρεση αυτού του σημείου θα χρησιμοποιήσουμε τα ζεύγη σημειοσυνόλων  $L_1, L'_2$  (του σώματος των γραμμμάτων) και των καμπυλών  $C_1, C'_2$  (του εξωτερικού περιγράμματος). Για κάθε μετασχηματισμένο σώμα γράμματος  $L'_2$  ορίζουμε το κριτήριο τομής δια ένωση για κάθε  $n$  σημείο της διαμέρισης του τετρασδιάστατου κύβου:

$$\delta E_n = \frac{\text{εμβαδόν}(L_1 \cap L'_2)}{\text{εμβαδόν}(L_1 \cup L'_2)} = \frac{\text{EMB}(L_1 \cap L'_2)}{\text{EMB}(L_1 \cup L'_2)}$$

Με βάση τις υπολογισθείσες τιμές εντοπίζουμε τη μέγιστη τιμή  $\delta E_B$ , καθώς η μεγιστοποίηση του κριτηρίου αυτού οδηγεί στην επιλογή της καλύτερης σχετικής θέσης των δύο υλοποιήσεων  $L_1, L'_2$ . Τέλος, ορίζουμε τις παραμέτρους  $(x_\delta^B, y_\delta^B, \varphi^B, \lambda^B)$  που αντιστοιχούν σε αυτή τη μέγιστη τιμή. Ιδανικά, για τα γράμματα που προέρχονται από τον ίδιο γραφέα, θα περιμέναμε η ποσότητα  $\delta E_n$  να είναι ίση με τη μονάδα. Ωστόσο, λόγω ενός πλήθους παραγόντων που αναλύθηκαν στο Κεφάλαιο 1 η ποσότητα αυτή πρακτικά πάντα αποκλίνει από τη μονάδα.

#### 5.4.2 Ορισμός Κριτηρίου Για Δύο Μη Συνεκτικές Υλοποιήσεις του Ίδιου

##### Γράμματος

Στην περίπτωση που κάποιο από τα υπό μελέτη γράμματα αποτελείται από παραπάνω από μια κλειστή καμπύλη, λόγω διαδοχικών μεμονωμένων μονοκονδυλιών ή γενικότερα λόγω επανατοποθέτησης του εργαλείου γραφής στην επιφάνεια γραφής επαναπροσδιορίζουμε το παραπάνω κριτήριο.

Αναλυτικότερα, έστωσαν δύο υλοποιήσεις  $L_1, L_2$  του ίδιου γράμματος που βρίσκονται στο ίδιο ή σε διαφορετικά κείμενα, που αποτελούνται  $N_1, N_2$  απλές συνεκτικές περιοχές αντίστοιχα. Θεωρούμε επίσης τα σύνολα  $S^1, S^2$  των απλών συνεκτικών περιοχών  $N_1, N_2$  αντίστοιχα. Προσαρμόζουμε το  $L_2$  στο  $L_1$  χρησιμοποιώντας τους μετασχηματισμούς στροφής, παράλληλης μετατόπισης και ομοιοθεσίας και για κάθε σημείο  $n$  της διαμέρισης της μετασχηματισμένης εκδοχής του  $L_2$  ορίζουμε το κριτήριο τομής δια ένωση:

$$\delta E_n = \frac{\text{EMB}\left(\left(\bigcup_{i=1}^{N_1} S_i^1\right) \cap \left(\bigcup_{j=1}^{N_2} S_j^2\right)\right)}{\text{EMB}\left(\left(\bigcup_{i=1}^{N_1} S_i^1\right) \cup \left(\bigcup_{j=1}^{N_2} S_j^2\right)\right)}$$

Όπως και στην προηγούμενη περίπτωση, εντοπίζουμε τη μέγιστη τιμή  $\delta E_B$ , καθώς και τις αντίστοιχες παραμέτρους  $(x_s^B, y_s^B, \varphi^B, \lambda^B)$  που παρήγαγαν το  $\delta E_B$ . Δε θα πρέπει να παραλειφθεί ότι η ακρίβεια της προκύπτουσας προσαρμογής εξαρτάται από την κατάλληλη επιλογή του βήματος στη διαμέριση του τετραδιάστατου κύβου που ορίζουν οι παράμετροι των μετασχηματισμών.

## 5.5 Στατιστικός Έλεγχος Ομοιότητας Δύο Υλοποιήσεων

Με βάση τη βέλτιστη προσαρμογή των υλοποιήσεων του ίδιου συμβόλου της αλφαβήτου που προκύπτει με μία εκ των μεθοδολογιών της Υποενότητας 5.4, παρατηρήθηκε ότι η στατιστική τους σύγκριση μπορεί να οδηγήσει στην ταυτοποίηση γραφέα, με πολύ ικανοποιητικό τρόπο, μέσω της κάτωθι διαδικασίας που παρουσιάστηκε πρώτη φορά στην εργασία [1].

1. Υπολογισμός των ποσοτήτων  $\delta E_B$  για όλες τις υλοποιήσεις ενός γράμματος που βρίσκονται στην ίδια σελίδα ή στο ίδιο τμήμα ενός χειρογράφου, έστω TX

Ας είναι  $I$  μια δοκιμαστική σελίδα ή ένα TX ενός οποιοδήποτε χειρογράφου, στην οποία εντοπίζονται  $N_I$  υλοποιήσεις ενός γράμματος, έστω το «π». Με βάση τη μεθοδολογία προσαρμογής που αναλύθηκε στην ενότητα 5.4, προσαρμόζουμε όλες τις  $N_I - 1$  υλοποιήσεις του γράμματος «π» στην πρώτη υλοποίηση, και υπολογίζουμε σε κάθε προσαρμογή τις ποσότητες  $\delta E_{1,j}$  με  $j = 2, \dots, N_I$ . Στη συνέχεια εκτελούμε την ίδια διαδικασία προσαρμόζοντας όλες τις  $N_I - 2$  υλοποιήσεις στη δεύτερη υλοποίηση και υπολογίζουμε τα  $\delta E_{2,j}$  με  $j = 3, \dots, N_I$  κ.ο.κ.

Για να εκμεταλλευτούμε τις υπολογισθείσες ποσότητες βρίσκουμε το μέσο όρο και τη δειγματική διασπορά τους, δηλαδή για την σελίδα ή το TX  $i$  για το γράμμα «π» έχουμε:

$$\delta E^{\pi,i} = \frac{1}{N_I - 1} \sum_{i=2}^{N_I} \left( \frac{1}{N_I - i} \sum_{j=i}^{N_I} \delta E_{i,j} \right)$$

$$SE^{\pi,i} = \sqrt{\frac{1}{N_I - 1} \sum_{i=2}^{N_I} \frac{1}{N_I - i} \sum_{j=i}^{N_I} (\delta E_{i,j} - \delta E^{\pi,i})^2}$$



2. Υπολογισμός των ποσοτήτων  $\delta E_B$  για όλες τις υλοποιήσεις ενός γράμματος που βρίσκονται σε διαφορετικές σελίδες ή TX του ιδίου ή ενός άλλου κειμένου

Θεωρούμε ότι έχουμε δύο διαφορετικές σελίδες ή TX  $I$  και  $J$ , χωρίς να γνωρίζουμε αν αυτές ανήκουν στο ίδιο χειρόγραφο ή στον ίδιο γραφέα. Έστω ότι στη σελίδα  $I$  έχουμε  $N_I$  υλοποιήσεις του υπό μελέτη γράμματος και στη σελίδα  $J$  έχουμε  $N_J$  υλοποιήσεις αντίστοιχα. Τότε με βάση τη μεθοδολογία προσαρμογής που αναλύθηκε στην ενότητα 5.4, προσαρμόζουμε όλες τις  $N_J$  υλοποιήσεις του γράμματος «π» της σελίδας  $J$  στην πρώτη υλοποίηση της σελίδας  $I$ , και υπολογίζουμε σε κάθε προσαρμογή τις ποσότητες  $\delta E_{1,j}$  με  $j=1, \dots, N_J$ . Στη συνέχεια εκτελούμε την ίδια διαδικασία προσαρμόζοντας όλες τις  $N_J$  υλοποιήσεις της σελίδας  $J$  στη δεύτερη υλοποίηση της σελίδας  $I$ , και υπολογίζουμε τα  $\delta E_{2,j}$  με  $j=1, \dots, N_J$ . Επαναλαμβάνουμε τη διαδικασία αυτή μέχρι να φτάσουμε στην τελευταία υλοποίηση της σελίδας  $I$ .

Για να εκμεταλλευτούμε τις υπολογισθείσες ποσότητες βρίσκουμε το μέσο όρο και τη δειγματική διασπορά τους, δηλαδή για την σελίδες ή τα TX  $I, J$  και το γράμμα «π» έχουμε:

$$\delta E^{\pi,I,J} = \frac{1}{N_I} \sum_{i=1}^{N_I} \frac{1}{N_J} \sum_{j=1}^{N_J} \delta E_{i,j}$$

$$SE^{\pi,I,J} = \sqrt{\frac{1}{N_I} \sum_{i=1}^{N_I} \frac{1}{N_J} \sum_{j=1}^{N_J} (\delta E_{i,j} - \delta E^{\pi,I,J})^2}$$

3. Υπολογισμός ενός μέτρου με το οποίο εκτιμάμε αν δύο διαφορετικές σελίδες ή TX προέρχονται από τον ίδιο γραφέα

Το κριτήριο εκτίμησης ταυτότητας γραφέα βασίζεται στη σύγκριση των ποσοτήτων  $\delta E^{\pi,I,J}$  και  $SE^{\pi,I,J}$ . Δεδομένου ότι μια ποσότητα  $\delta E^B$  λαμβάνει τιμές στο διάστημα  $[0,1]$ , αποτελεί ουσιαστικά ένα ποσοστό και επομένως για  $N_I \geq 5, N_J \geq 5$  ακολουθεί κατά προσέγγιση κανονική κατανομή. Επιβεβαιώνουμε την υπόθεση αυτή με τεστ Kolmogorov- Smirnov. Εν συνεχεία, υποθέτουμε ότι αν οι σελίδες  $I$  και  $J$  προέρχονται από το ίδιο «χέρι», τότε αναμένουμε να ισχύει  $\delta E^{\pi,I} = \delta E^{\pi,I,J}$ .

Δηλαδή, ελέγχουμε τη στατιστική υπόθεση  $H_0: \delta E^{\pi,I} = \delta E^{\pi,I,J}$ , η οποία αν ισχύει θεωρούμε ότι, όσον αφορά το γράμμα «π», τα τμήματα χειρογράφου  $I$  και  $J$  έχουν γραφθεί από το ίδιο χέρι. Προφανώς, ως δει, η  $H_0$  ελέγχεται έναντι της εναλλακτικής υποθέσεως  $H_1: \delta E^{\pi,I} \neq \delta E^{\pi,I,J}$ .

Η ποσοτικοποίηση του ελέγχου της υπόθεσης  $H_0$  μας οδηγεί στο κάτωθι μέτρο που περιγράφει αν δύο σελίδες ή τμήματα χειρογράφου προέρχονται από το ίδιο χέρι:

$$\delta^{\pi,I,J} = \frac{\delta E^{\pi,I} - \delta E^{\pi,I,J}}{\sqrt{\frac{(SE^{\pi,I})^2}{N_I(N_I-1)} + \frac{(SE^{\pi,I,J})^2}{N_I N_J}}}$$

Είναι εμφανές ότι όσο πιο κοντά στο μηδέν είναι η παραπάνω ποσότητα για ένα ή περισσότερα γράμματα τόσο περισσότερο αναμένεται τα τμήματα χειρογράφου  $I$  και  $J$  να προέρχονται από τον ίδιο γραφέα.

Δεδομένου ότι τα  $\delta E$  στον αριθμητή ακολουθούν κανονική κατανομή, οι ποσότητες  $\delta^{\pi,I,J}$  θα ακολουθούν κατανομή *Student* ή για συνολικό πλήθος συγκρίσεων μεγαλύτερο ίσο του 30 θα ακολουθούν με πολύ ικανοποιητική προσέγγιση τυπική κανονική κατανομή.

Άμεσο επακόλουθο αυτής της διαπίστωσης είναι ότι μπορούμε να ανάγουμε τη σύγκριση της ταυτότητας του γραφέα μεταξύ δύο τμημάτων χειρογράφων στη σύγκριση των ποσοτήτων  $\delta^{letter,I,J}$  για έναν οποιοδήποτε αριθμό συμβόλων της αλφαβήτου. Συγκεκριμένα, διατυπώνουμε τόσες μηδενικές υποθέσεις  $H_0^{letter}$  όσες και τα υπό μελέτη σύμβολα της αλφαβήτου (στην περίπτωσή μας τα σύμβολα «π» και «ε»). Ελέγχουμε τις υποθέσεις αυτές ως προς την ορθότητά τους για κάθε ζεύγος τμημάτων των χειρογράφων πραγματοποιώντας τόσα τεστ όσα και τα εξεταζόμενα γράμματα (στην περίπτωσή μας 2) με βαθμό εμπιστοσύνης  $\alpha \leq 10^{-4}$  [1].

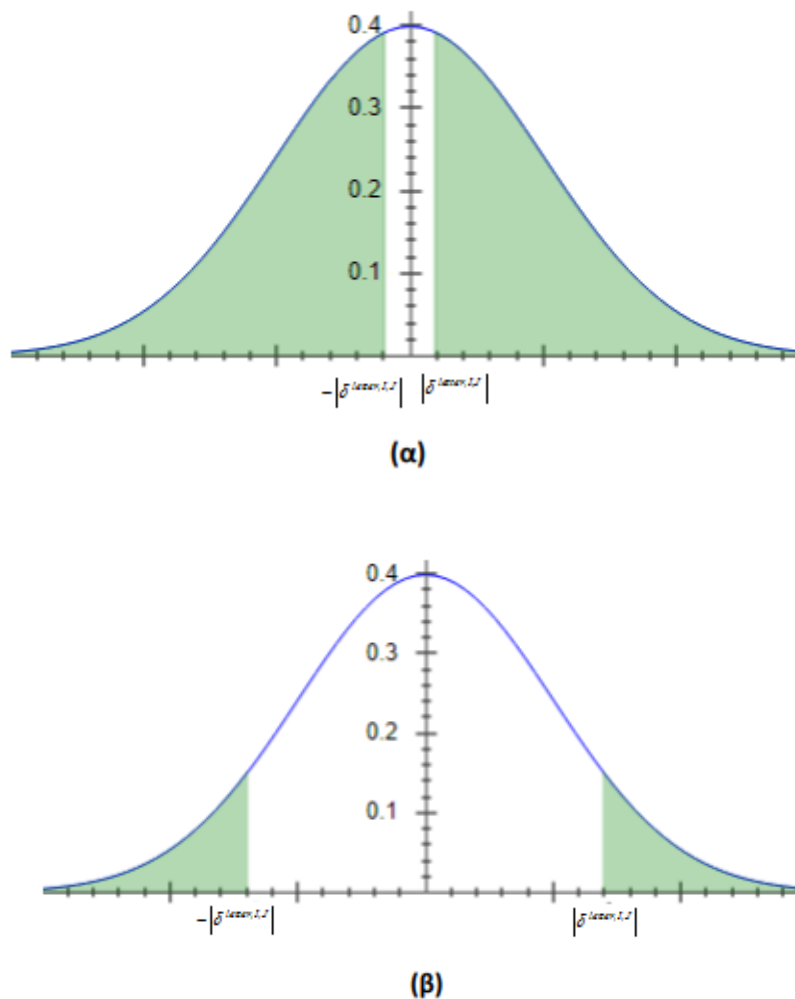
### 5.5.1 Πιθανοφάνεια στη θέση βέλτιστης Προσαρμογής Δύο Υλοποιήσεων

#### Συμβόλων της Αλφαβήτου

Συμπληρωματικά με το παραπάνω κριτήριο, μπορούμε να ορίσουμε ένα επιπλέον μέτρο σύγκρισης των δύο υλοποιήσεων των συμβόλων της αλφαβήτου που ανήκουν σε διαφορετικά κείμενα  $I, J$ , ως προς την ταυτότητα του γραφέα τους. Συγκεκριμένα, όπως διατυπώθηκε στην προηγούμενη Υποενότητα, οι ποσότητες  $\delta^{letter,I,J}$  ακολουθούν υπό ορισμένες συνθήκες τυπική κανονική κατανομή. Συνεπώς, η πιθανοφάνεια τα κείμενα  $I$  και  $J$  να ανήκουν στον ίδιο γραφέα, όσον αφορά ένα σύμβολο της αλφαβήτου είναι

$$\begin{aligned}
 P_{I,J}^{letter} &= P\left(-\infty < x < -|\delta^{letter,I,J}| \cup |\delta^{letter,I,J}| < x < +\infty\right) \\
 &= \int_{-\infty}^{-|\delta^{letter,I,J}|} \frac{1}{\sqrt{2 \cdot \pi}} \cdot e^{-\frac{x^2}{2}} dx + \int_{|\delta^{letter,I,J}|}^{\infty} \frac{1}{\sqrt{2 \cdot \pi}} \cdot e^{-\frac{x^2}{2}} dx = \sqrt{\frac{2}{\pi}} \cdot \int_{|\delta^{letter,I,J}|}^{\infty} e^{-\frac{x^2}{2}} dx
 \end{aligned}$$

Επομένως, στην περίπτωση που τα δύο κείμενα προέρχονται από τον ίδιο γραφέα, δηλαδή έχουμε μεγάλο βαθμό ομοιότητας των κειμένων ως προς τα υπό μελέτη γράμματα, αναμένουμε η ποσότητα  $\delta^{letter,I,J}$  για το κάθε εξεταζόμενο γράμμα να πλησιάζει στο μηδέν, και η απόλυτη τιμή της αντίστοιχης πιθανοφάνειας  $P_{I,J}^{letter}$  να είναι όσο το δυνατόν μεγαλύτερη, όπως φαίνεται στο παρακάτω Σχήμα . Αντίστοιχα, στην περίπτωση που τα δύο κείμενα προέρχονται από διαφορετικούς γραφείς, αναμένουμε η απόλυτη τιμή της ποσότητας  $\delta^{letter,I,J}$  να είναι αρκετά μεγάλη και η τιμή της πιθανοφάνειας  $P_{I,J}^{letter}$  να είναι σημαντικά μικρότερη, για ικανό αριθμό συμβόλων της αλφαβήτου, όπως φαίνεται στο Σχήμα 12 .



Σχήμα 12: Γραφική παράσταση της πιθανοφάνειας δύο κείμενα να έχουν γραφθεί από τον ίδιο γραφέα με βάση ένα σύμβολο “letter”. Στο (α) έχουμε μεγάλη πιθανοφάνεια, ενώ στο (β) μικρή.

Τελική πιθανοφάνεια που βασίζεται σε ένα σύνολο γραμμάτων

Με βάση την παραπάνω παρατήρηση ορίζουμε ένα σύνολο  $n$  συμβόλων της αλφαβήτου  $\Sigma^S$ , που περιέχει τα  $n$  γράμματα που θα εξεταστούν. Για να προβούμε στην ταυτοποίηση του «χεριού» για ένα ζεύγος κειμένων  $I, J$  ορίζουμε ως  $a_{1,2}$  την πιθανοφάνεια τα κείμενα αυτά να έχουν γραφεί από το ίδιο χέρι:

$$a_{1,2} = \sqrt[n]{\prod_{letter \in \Sigma^S} P_{I,J}^{letter}}$$

Όσο μικρότερη είναι η τιμή της παραπάνω πιθανοφάνειας, τόσο πιο απίθανο είναι τα δύο υπό μελέτη κείμενα να προέρχονται από τον ίδιο γραφέα. Αντιστρόφως, όσο μεγαλύτερη είναι η συγκεκριμένη πιθανοφάνεια, τόσο πιθανότερο είναι τα κείμενα να ανήκουν στον ίδιο γραφέα.

# 6

## *Εφαρμογή Μεθοδολογίας στα Χειρόγραφα του Ρήγα Φεραίου*

Με βάση τη μεθοδολογία που αναλύθηκε στο Κεφάλαιο 5, και με αφορμή τη διαφωνία για την πατρότητα των ανακαλυφθέντων στη Ρουμανία χειρογράφων, πραγματοποιήθηκε συγκριτική μελέτη των χειρογράφων του Ρήγα από τα έργα «Πολιτική Διοίκησης» (συγκεκριμένα από τα έργα «Συνταγμα» και «Θούριος»), το «Φυσικής Απάνθισμα», και τα ανέκδοτα έργα «Σαγανάκι της Τρέλλας» και «Η δοκιμασμένη Φιλία». Στο πλαίσιο της παρούσας μελέτης επιλέχθηκαν τα γράμματα «π» και «ε», και συγκεκριμένα 30 «π» και 30 «ε» από 4 σελίδες του χειρογράφου κάθε έργου, προκειμένου να συγκριθούν τα αποτελέσματα της μεθόδου για κάθε έργο.

### *6.1 Εφαρμογή των Μεθόδων Κατάτμησης Εικόνας στα Κείμενα του Ρήγα Φεραίου*

#### *6.1.1 Αποτελέσματα Κατάτμησης*

Το πρώτο βήμα της επεξεργασίας των γραμμάτων όπως περιγράψαμε στην Υποενότητα 5.1 είναι η περικοπή των γραμμάτων του από τη φωτογραφία του σαρωμένου κειμένου και η κατάλληλη κατάτμησή τους, ώστε να απομονωθεί το γράμμα από το υπόβαθρο της σελίδας. Για την επίδειξη των αποτελεσμάτων της διαδικασίας κατάτμησης επιλέχθηκαν δύο αντιπροσωπευτικά γράμματα από το σύνολο των δοκιμαστικών μας γραμμάτων και συγκεκριμένα από τις σελίδες 5 και 19 του χειρογράφου του έργου «Το Σαγανάκι της

Τρέλλας» (sag\_005\_f2\_e\_030.tif, sag\_019\_f1\_p\_007.tif). Για κάθε βήμα της Υποενότητας 5.1.4 παραθέτουμε τα αποτελέσματα της διαδικασίας:

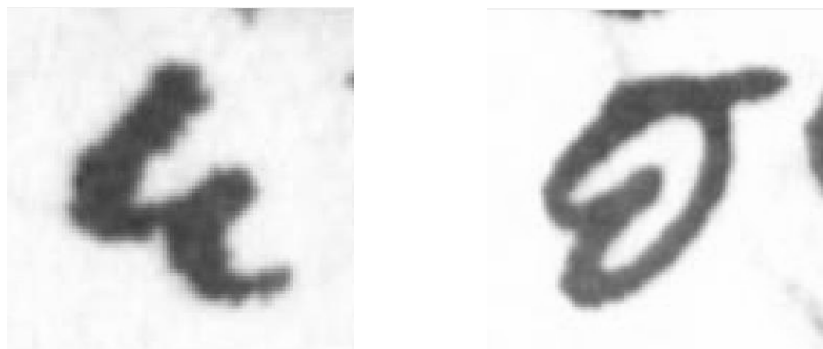
1. Εφαρμογή μορφολογικών φίλτρων για εξάλειψη θορύβου

Αρχικά, μετατρέπουμε την έγχρωμη εικόνα του γράμματος σε κανονικοποιημένη κλίμακα του γκρι. Τα αποτελέσματα της πρώτης φάσης της προ-επεξεργασίας παρουσιάζονται στο Σχήμα 13.



*Σχήμα 13: Αρχική έγχρωμη εικόνα των επιλεγθέντων γραμμάτων & Κανονικοποιημένη εικόνα των επιλεγθέντων γραμμάτων*

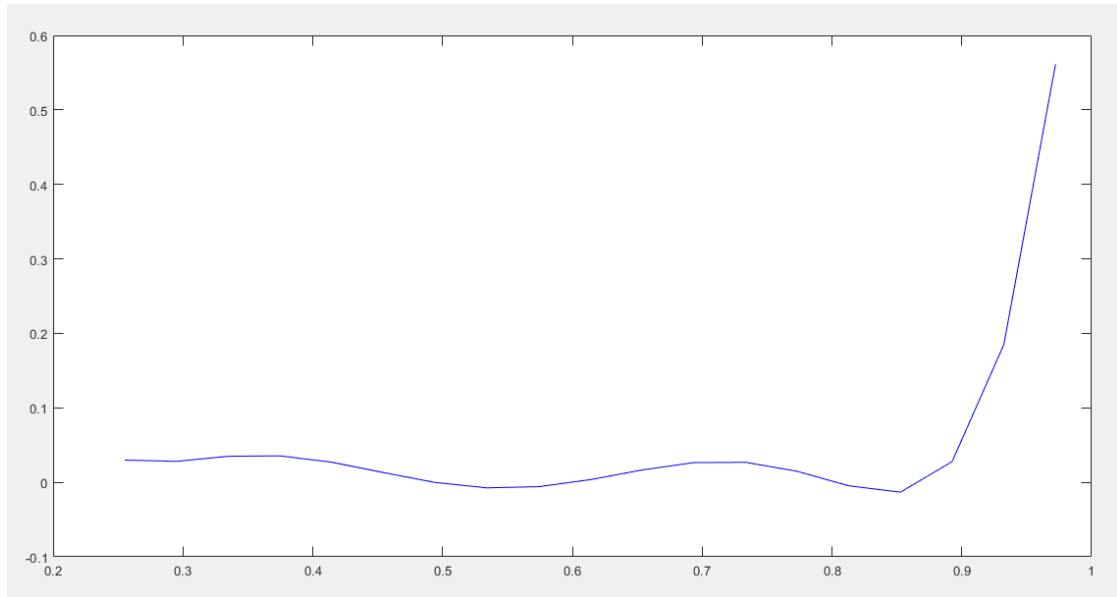
Στη συνέχεια, έχουμε την εφαρμογή των μορφολογικών φίλτρων για την εξομάλυνση. Συγκεκριμένα, παρακάτω παρουσιάζονται τα αποτελέσματα του φίλτρου διαστολής με δίσκο ακτίνας 1. Παρατηρούμε πως ενώ εξαλείφθηκε ο θόρυβος δομικού στοιχείου, αλλά «απλώθηκε» το σχήμα του γράμματος που είναι σε χαμηλές συχνότητες κοντά στο 0.



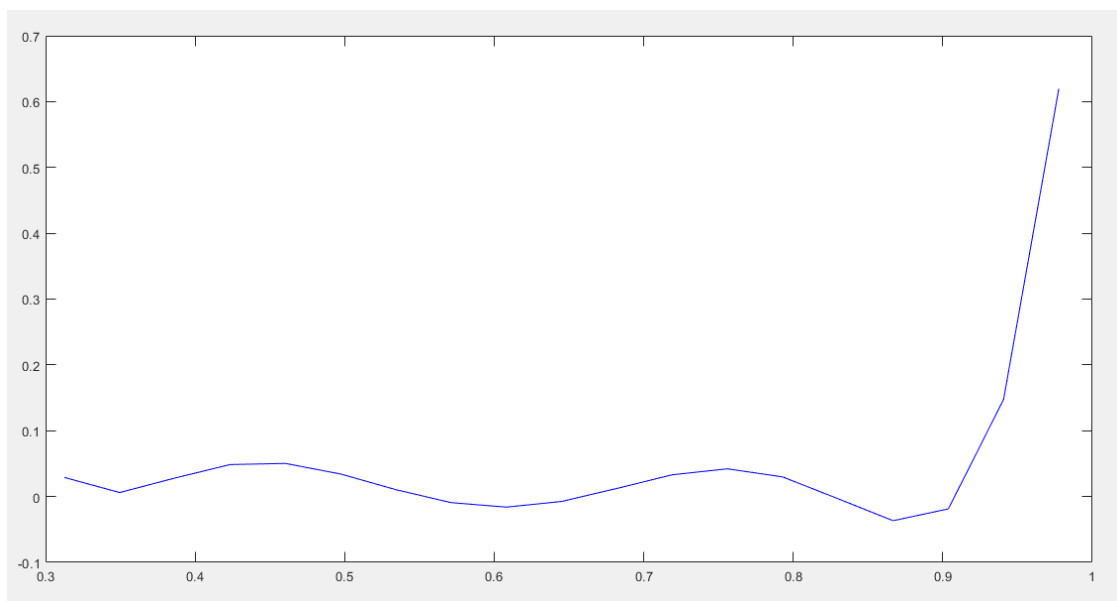
*Σχήμα 14: Εξομαλυμένη εικόνα των επιλεγθέντων γραμμάτων*

## 2. Δημιουργία Ιστογραμμάτων γκριζών εικόνων

Στο παρόν βήμα παράγονται τα ιστογράμματα των επιπέδων φωτεινότητας των δύο εξομαλυμένων εικόνων των γραμμάτων, εκ των οποίων το πρώτο ανήκει στο γράμμα «ε» και το δεύτερο στο γράμμα «π». Αξίζει να σημειωθεί ότι ορίζεται διαμέριση ίση με 19, δηλαδή έχουμε 19 διαστήματα τιμών στο ιστόγραμμα.



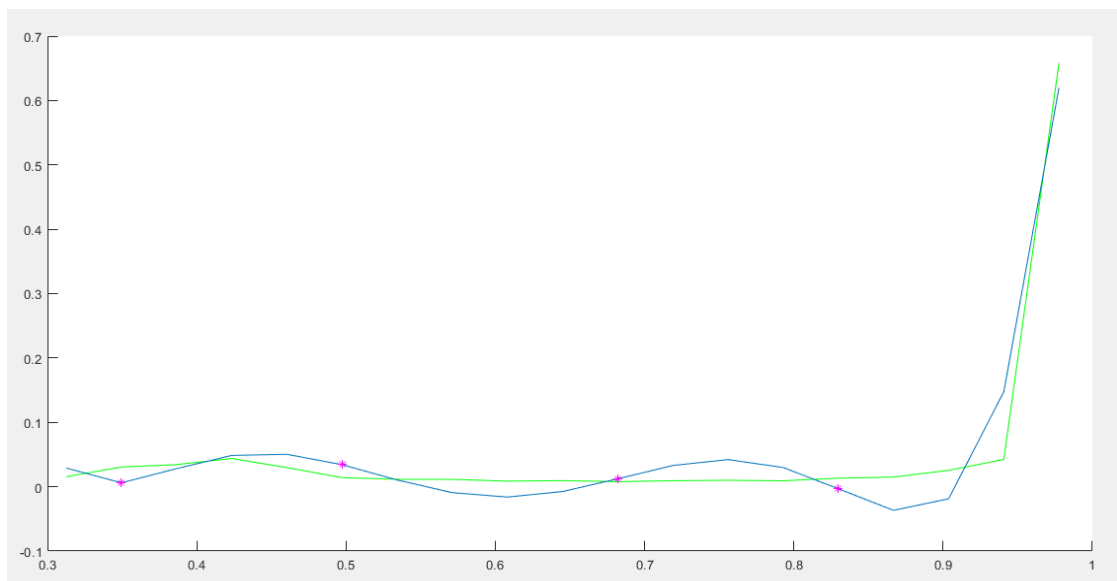
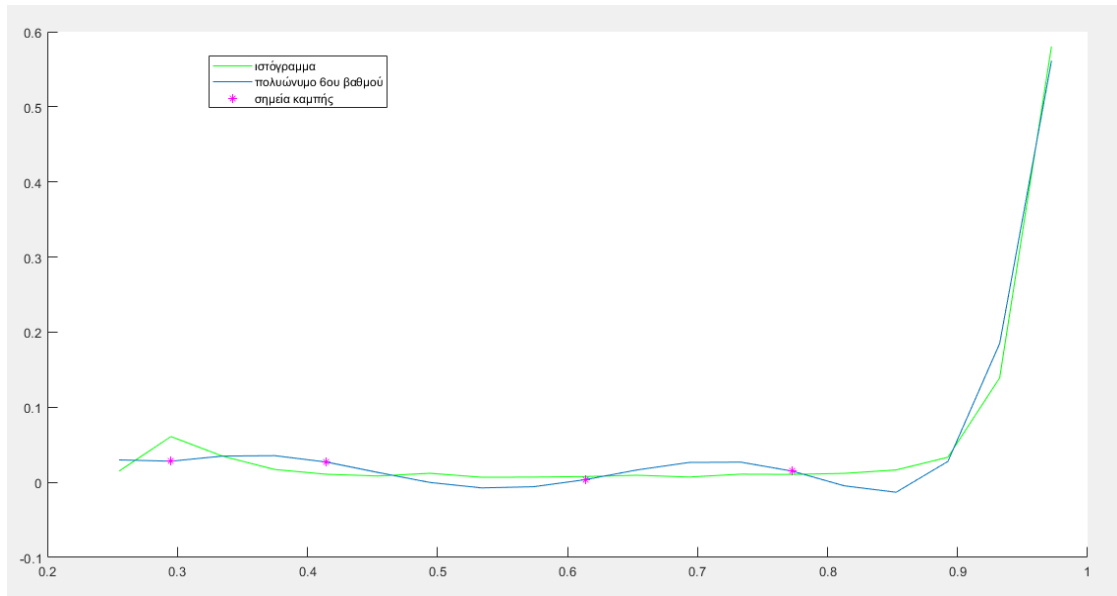
Σχήμα 15: Ιστόγραμμα Κανονικοποιημένων Τιμών Φωτεινότητας για το γράμμα «ε»



Σχήμα 16: Ιστόγραμμα Κανονικοποιημένων Τιμών Φωτεινότητας για το γράμμα «π»

### 3. Προσέγγιση Ομαλοποιημένου Ιστογράμματος

Παρατηρούμε στα παρακάτω διαγράμματα ότι το ομαλό πολυώνυμο που προσεγγίζει βέλτιστα το εκάστοτε κανονικοποιημένο ιστογράμμά μας με διαμέριση ίση με 19 είναι ένα πολυώνυμο 6<sup>ου</sup> βαθμού.

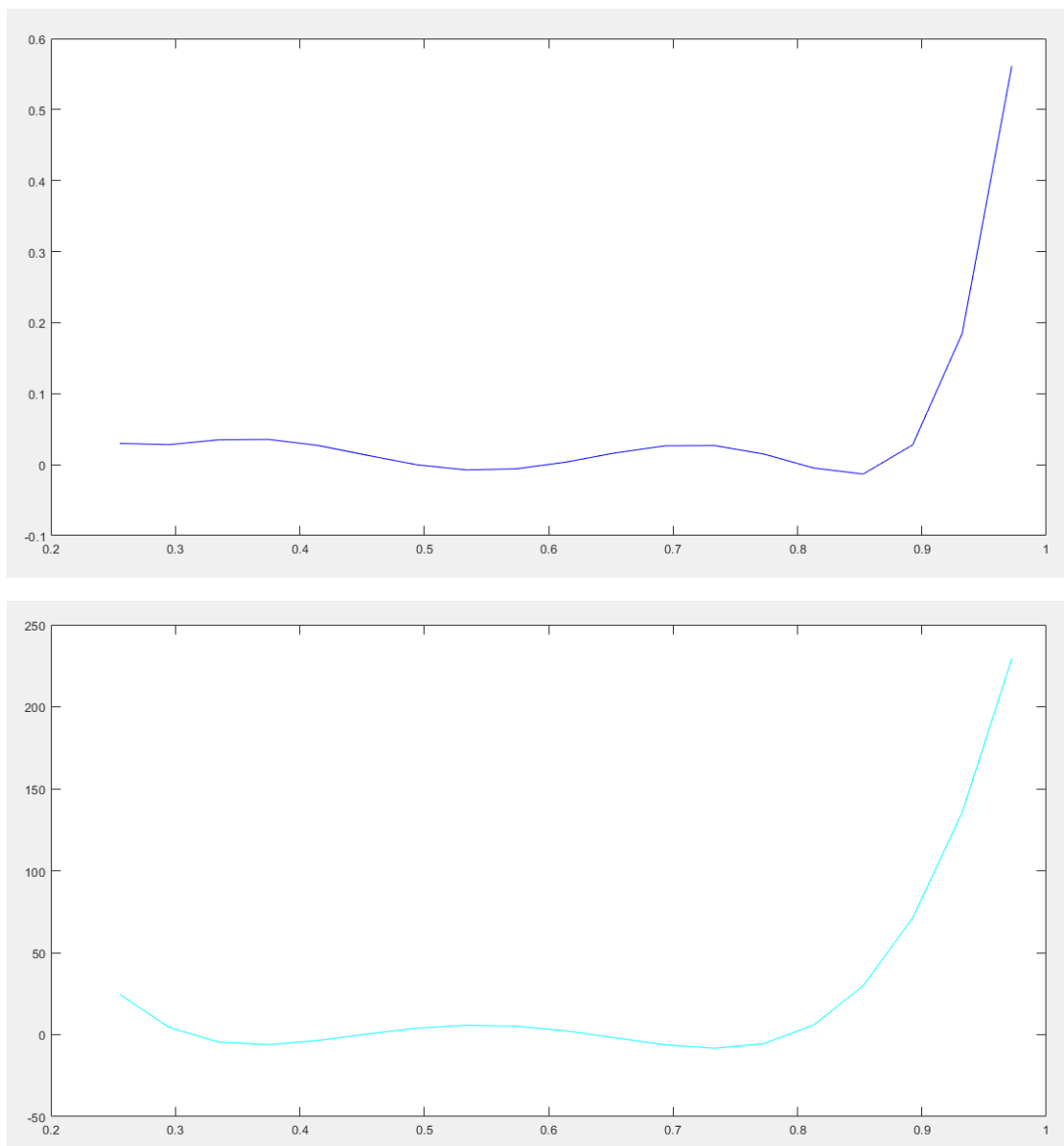


Σχήματα 17: Ιστόγραμμα κανονικοποιημένων τιμών φωτεινότητας (πράσινη καμπύλη), που προσεγγίζεται βέλτιστα από ομαλό πολυώνυμο 6<sup>ου</sup> βαθμού (μπλε καμπύλη), του οποίου τα σημεία καμπής σημειώνονται με ροζ



#### 4. Εύρεση των σημείων καμπής πολυωνύμου

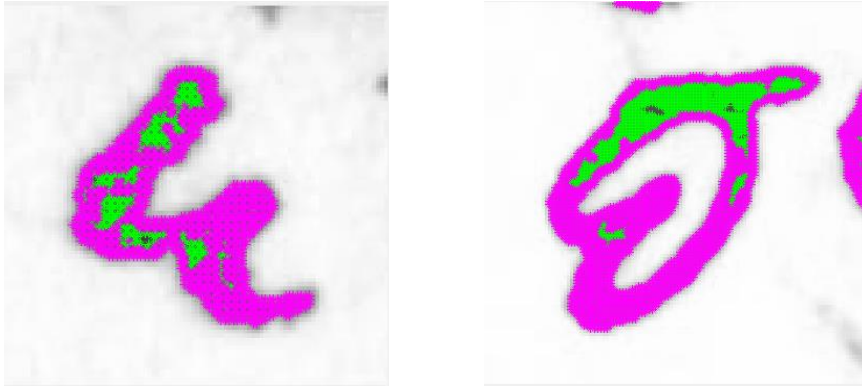
Με βάση το ομαλοποιημένο ιστόγραμμα φωτεινότητας ορίζουμε τα σημεία καμπής του. Συγκεκριμένα, παρατηρούμε στα παρακάτω σχήματα τη σύγκριση της καμπύλης ενός ομαλοποιημένου ιστογράμματος και της 2<sup>ης</sup> παραγώγου του για τον εντοπισμό των σημείων αλλαγής προσήμου του.



Σχήματα 18: Ομαλοποιημένο ιστόγραμμα φωτεινότητας (πάνω) και η δεύτερη παράγωγός του(κάτω). Από τη σύγκριση των δύο εν λόγω καμπυλών προκύπτουν τα σημεία καμπής

Με βάση τα σημεία καμπής δημιουργούνται διαστήματα τιμών φωτεινότητας. Τα σημεία του πρώτου διαστήματος, δηλαδή τα σημεία που έχουν χαμηλή τιμή φωτεινότητας, μπορούμε να ισχυριστούμε ότι ανήκουν στο σώμα του γράμματος και εμφανίζονται με πράσινο. Τα σημεία με ροζ χρώμα βρίσκονται στο διάστημα μεταξύ του πρώτου και δεύτερου σημείου καμπής και έχουν μεγαλύτερη φωτεινότητα από αυτά του πρώτου διαστήματος, αλλά εξακολουθούν να έχουν χαμηλή τιμή

φωτεινότητας, οπότε μπορούν επίσης να αντιστοιχιστούν στο σώμα του γράμματος. Δε θα πρέπει να παραλειφθεί η απουσία εσφαλμένων αντιστοιχίσεων, δηλαδή αντιστοίχισης σημείων του υποβάθρου στο σώμα του γράμματος. Το γεγονός αυτό οφείλεται στην καλή ποιότητα της αρχικής φωτογραφίας του χειρογράφου.



Σχήματα 19: Ένας πρώτος διαχωρισμός των εικονοστοιχείων του γράμματος από το υπόβαθρο της σελίδας με βάση τα διαστήματα που ορίζουν τα σημεία καμπής του ομαλοποιημένου ιστογράμματος.

5. Αρχική εκτίμηση των θεματικών περιοχών με βάση τα σημεία καμπής

Με βάση τον παραπάνω πρώτο διαχωρισμό σε «γράμμα» και «υπόβαθρο» της εικόνας μας, τη μετατρέπουμε σε δυαδική εικόνα, δηλαδή δημιουργούμε την πρώτη μορφή της κατατμημένης εικόνας, όπως φαίνεται στο Σχήμα 20.



Σχήμα 20: Αρχική κατάτμηση των υπό μελέτη γραμμάτων σε «γράμμα» και «υπόβαθρο»

6. Εξάλειψη του θορύβου και ομογενοποίηση των θεματικών περιοχών

Στο πλαίσιο του βήματος αυτού απομακρύνεται ο θόρυβος φθοράς, που συνίσταται στο λευκό και το μαύρο θόρυβο όπως ορίστηκε στο προηγούμενο Κεφάλαιο. Παρατηρούμε στα Σχήματα 21 και 22 ότι έχει απομακρυνθεί μεγάλο ποσοστό μαύρου θορύβου, κυρίως μεμονομένα εικονοστοιχεία θορύβου, ενώ μεγαλύτερες ομάδες εικονοστοιχείων παραμένουν ανέπαφες.

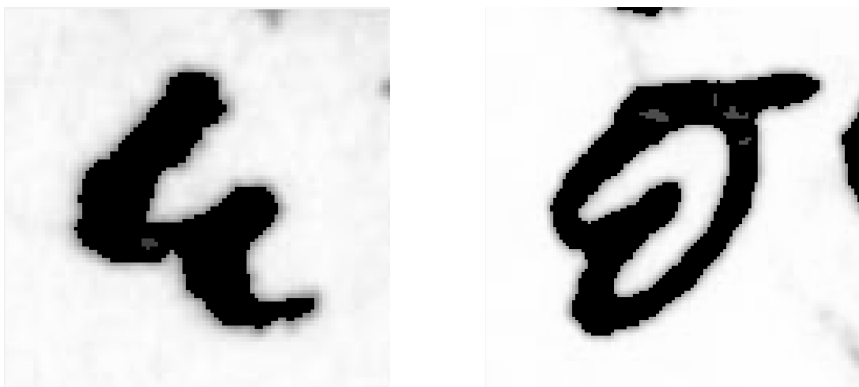


Σχήματα 21: Εικόνες Γραμμάτων μετά την απομάκρυνση του μαύρου θορύβου



Σχήματα 22: Εικόνες Γραμμάτων μετά την απομάκρυνση του λευκού θορύβου

7. Τελικός καθορισμός των 2 θεματικών περιοχών του σώματος του γράμματος  
Με βάση τα παραπάνω βήματα καταλήγουμε στις κατανεμημένες εικόνες των γραμμάτων που παρουσιάζονται στο Σχήμα 23:

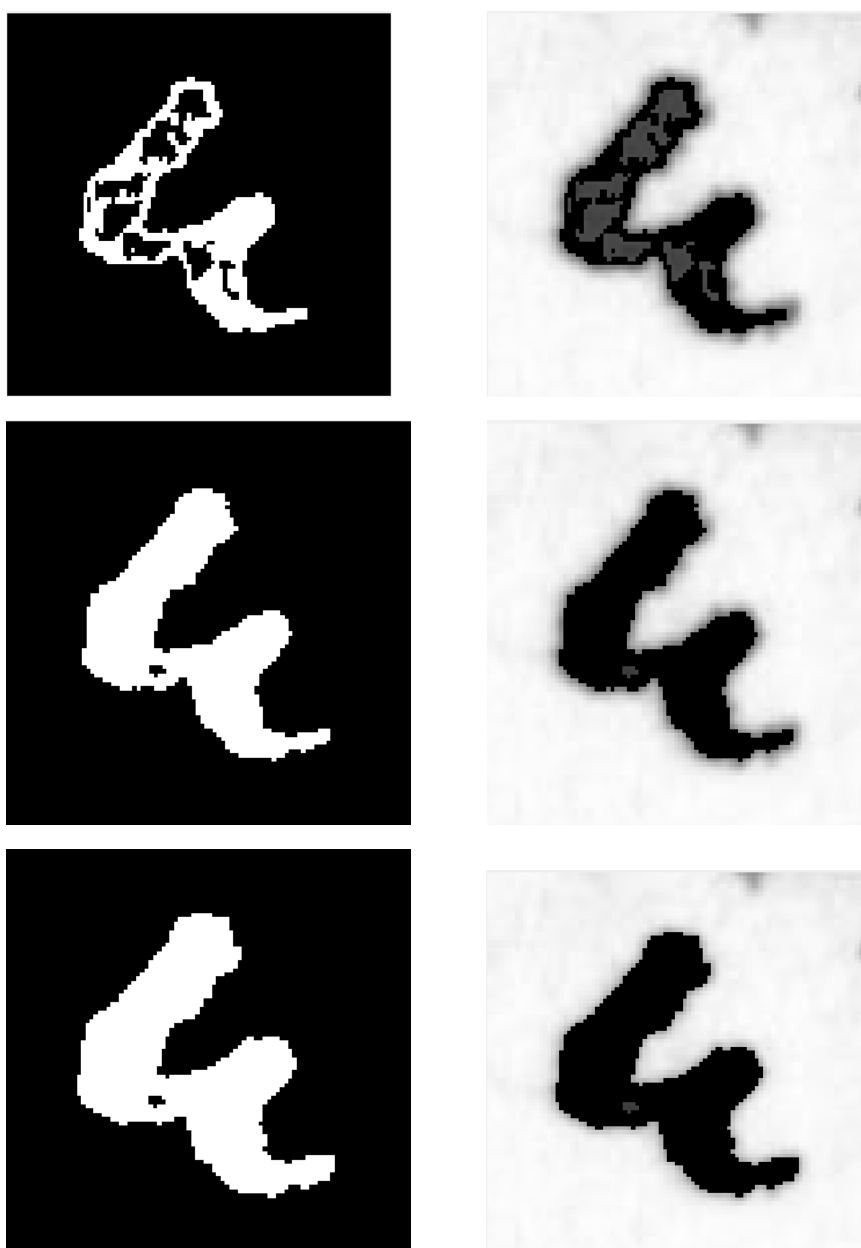


Σχήμα 23: Τελικά Αποτελέσματα Αφαίρεσης Κατάτμησης Από Την Γκρίζα Εικόνα

### 6.1.2 Διερεύνηση αποτελεσμάτων κατάτμησης ως προς τις τιμές των παραμέτρων του αλγορίθμου κατάτμησης

#### 1<sup>η</sup> Παράμετρος: Βαθμός Πολυωνύμου που Προσεγγίζει το Ιστόγραμμα Κανονικοποιημένων Τιμών Φωτεινότητας

Για τη διερεύνηση της επίδρασης της παραμέτρου του βαθμού του πολυωνύμου που καλείται να προσεγγίσει το Ιστόγραμμα του γράμματος, εκτελέσαμε την παραπάνω διαδικασία για σταθερή διαμέριση ίση με 19 και εύρος ίσο με 2 και βαθμό ίσο με 5, 6, 7, 8, και 9. Τα αποτελέσματα της κατατμημένης εικόνας και της αφαίρεσης αυτής από την γκριζα εικόνα για κάθε τιμή βαθμού παρατίθενται παρακάτω:





*Σχήμα 24: Τελική δυαδική κατάτμηση της εικόνας sag\_005\_f2\_e\_030.tif και αφαίρεση αυτής από την κανονικοποιημένη εικόνα για βαθμό πολωνύμου ίσο με 5,6,7,8,9*

Παρατηρούμε στα παραπάνω Σχήματα ότι το πολωνύμο 5<sup>ο</sup> βαθμού οριακά προσεγγίζει το ιστόγραμμα επιπέδων φωτεινότητας, επομένως η προκύπτουσα κατάτμηση περιέχει ελάχιστα ανιχνευμένα σώματα γράμματος. Αυξάνοντας τον βαθμό του πολωνύμου παρατηρούμε ότι το αποτέλεσμα κατάτμησης υπερκαλύπτει όλο και περισσότερο το σώμα του γράμματος, όπως φαίνεται στην περίπτωση του βαθμού ίσου με 9.

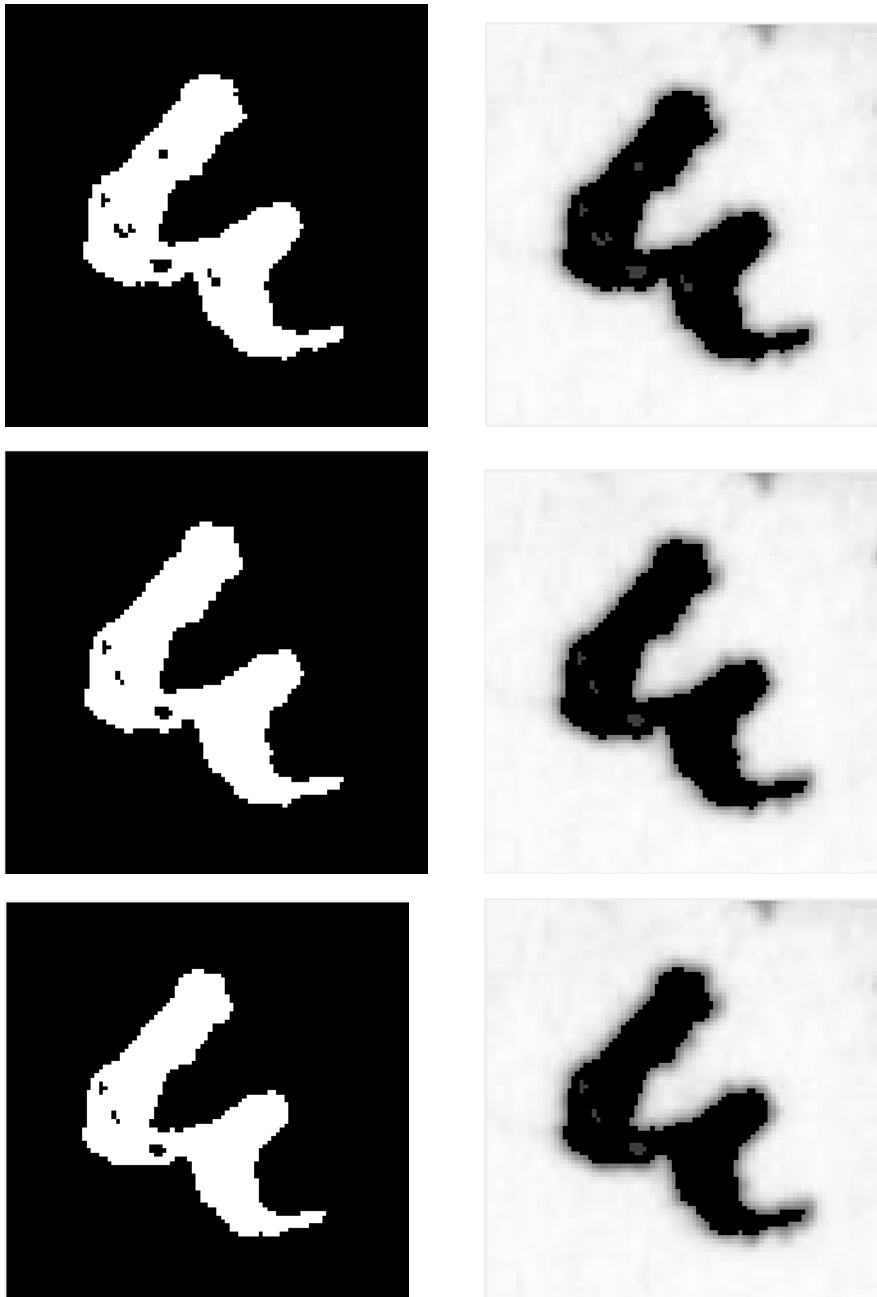
Αξίζει να σημειωθεί ότι σε εικόνες γραμμάτων με πολύ υψηλά επίπεδα θορύβου ή χαμηλή ανάλυση, ενδέχεται το πολωνύμο να μην ανιχνεύσει καθόλου σώμα γράμματος. Αυτό φαίνεται στο παρακάτω παράδειγμα από το χειρόγραφο του έργου «Σύνταγμα», στο οποίο για ένα τυχαία επιλεγμένο γράμμα, τα πολωνύμα 5<sup>ο</sup>, 6<sup>ο</sup> και 7<sup>ο</sup> βαθμού δε δύνανται να ανιχνεύσουν σώμα γράμματος:

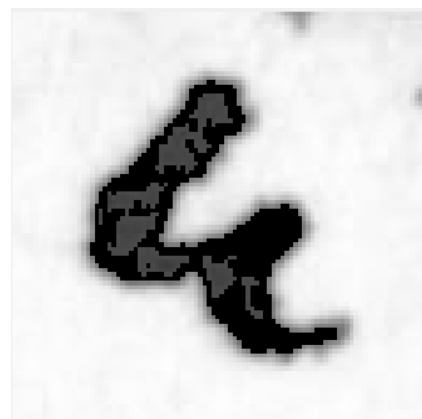
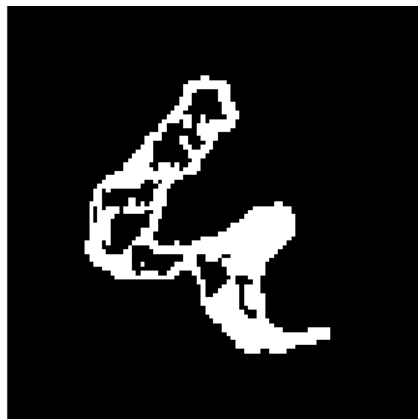
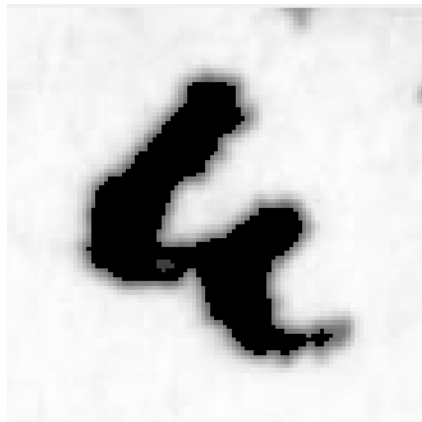
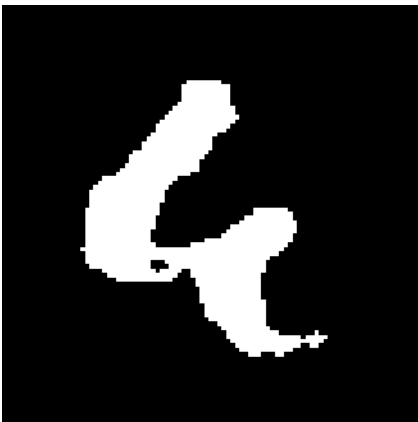
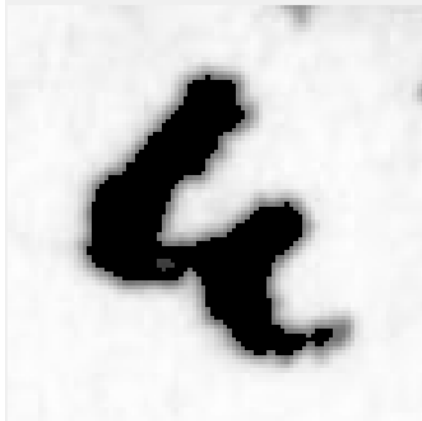
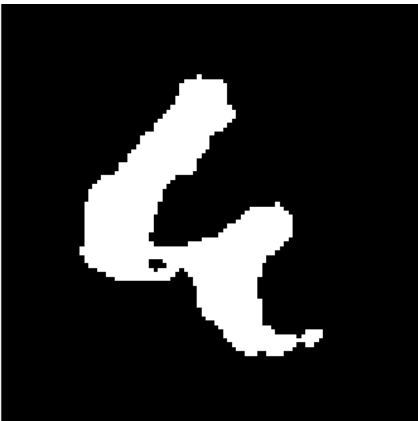
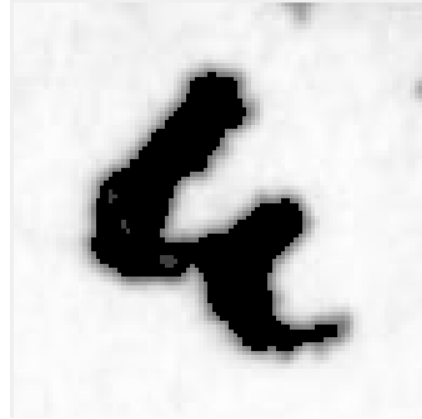
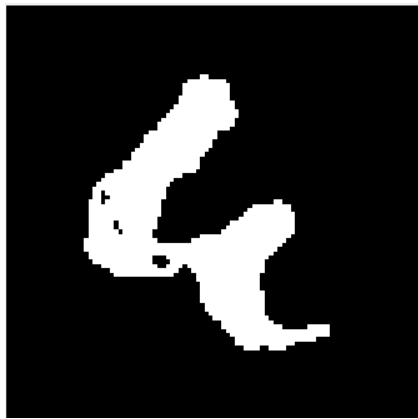


*Σχήμα 25: Αποτέλεσμα κατάτμησης για βαθμό πολωνύμου με τιμή από 5 έως και 8 σε θορυβώδη εικόνα. Παρατηρούμε ότι λόγω της χαμηλής ποιότητας της φωτογραφίας στις 3 πρώτες περιπτώσεις η μέθοδος δεν ανιχνεύει καθόλου το σώμα του γράμματος.*

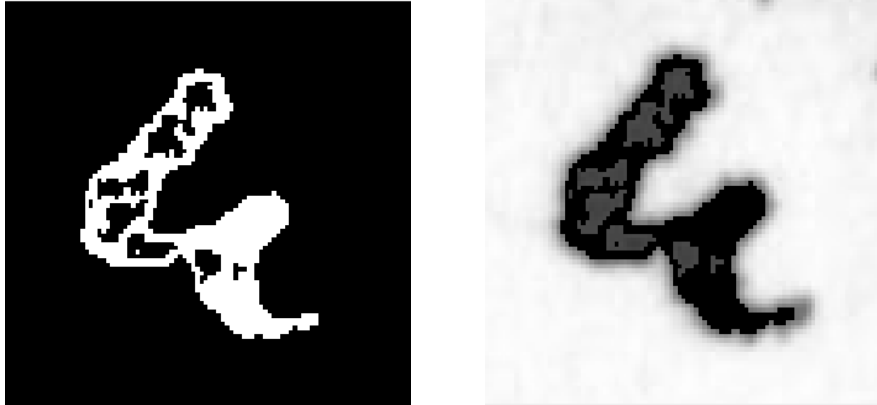
2<sup>η</sup> Παράμετρος: Αριθμός Διαστημάτων Διαμέρισης των Τιμών Φωτεινότητας

Για τη διερεύνηση της επίδρασης της παραμέτρου των αριθμών διαστημάτων διαμέρισης των τιμών φωτεινότητας του Ιστογράμματος, εκτελέσαμε την παραπάνω διαδικασία για σταθερό βαθμό πολωνύμου ίσο με 5, εύρος ίσο με 2 και διαμέριση με τιμή από 13 έως 20. Τα αποτελέσματα της κατατιμημένης εικόνας και της αφαίρεσης αυτής από την γκριζα εικόνα για κάθε τιμή βαθμού παρατίθενται παρακάτω. Συγκεκριμένα, το πρώτο ζεύγος έχει διαμέριση 13, το δεύτερο 14 κ.ο.κ.:







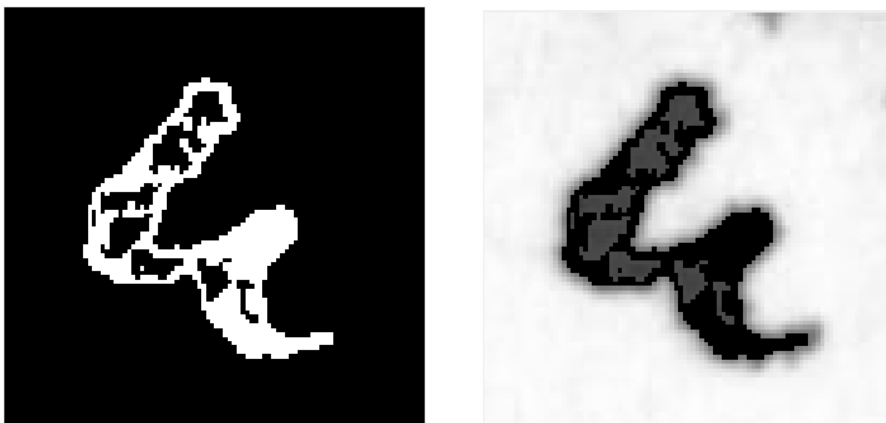


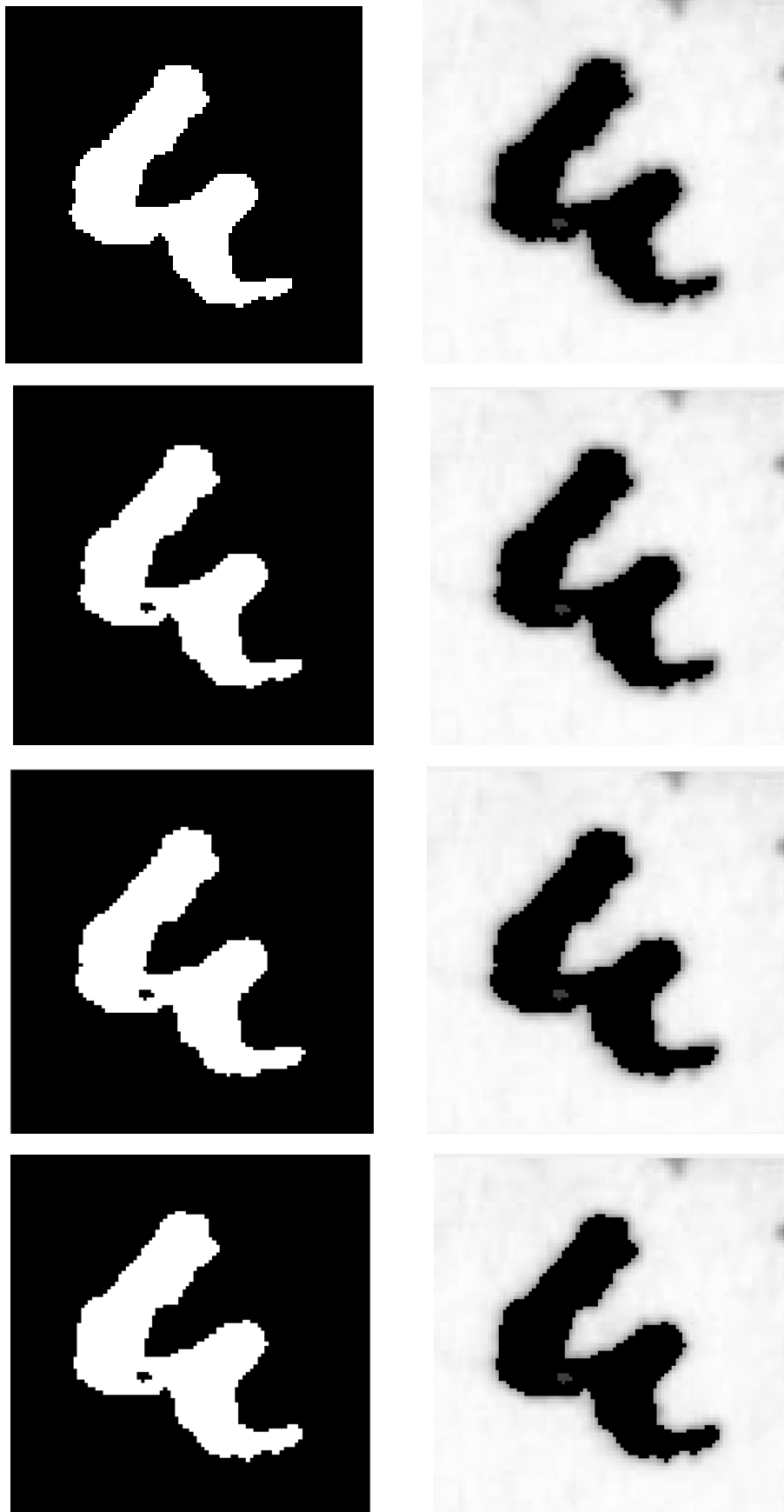
Σχήμα 26: Αποτελέσματα διερεύνησης της παραμέτρου του αριθμού των διαστημάτων διαμέρισης

Παρατηρούμε στα παραπάνω σχήματα ότι αυξάνοντας τον αριθμό των διαστημάτων διαμέρισης του ιστογράμματος της γκριζας εικόνας, μεταβάλλεται η ακρίβεια επιπέδων φωτεινότητας ιστογράμματος στην γκριζα εικόνα.

### 3<sup>η</sup> Παράμετρος: Εύρος Διαστημάτων Φωτεινότητας

Για τη διερεύνηση της επίδρασης της παραμέτρου του εύρους διαστημάτων των τιμών φωτεινότητας του Ιστογράμματος, εκτελέσαμε την παραπάνω διαδικασία για σταθερό βαθμό πολωνύμου ίσο με 5, διαμέριση με τιμή 19 και εύρος ίσο από 2 έως 6 με βήμα 1. Τα αποτελέσματα της κατατημένης εικόνας και της αφαίρεσης αυτής από την γκριζα εικόνα για κάθε τιμή βαθμού παρατίθενται παρακάτω:



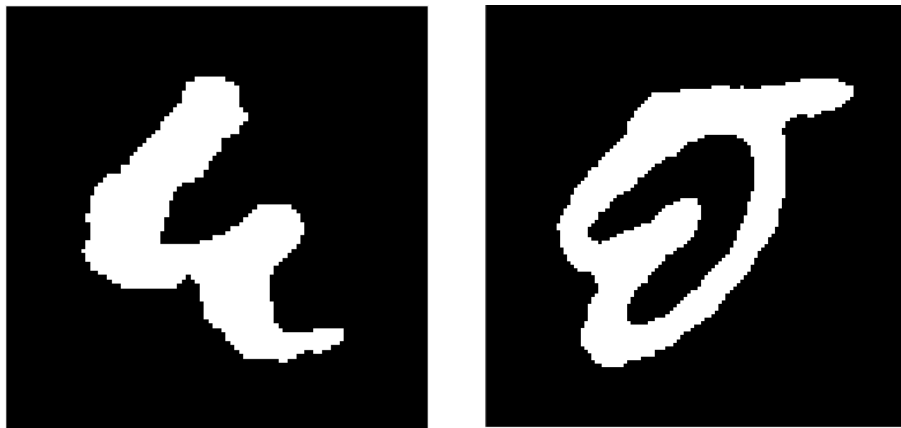


Σχήμα 27: Αποτελέσματα διερεύνησης της παραμέτρου του εύρους των διαστημάτων διαμέρισης

Παρατηρούμε ότι όσο μεγαλώνει η παράμετρος του εύρους το αποτέλεσμα κατάτμησης υπερκαλύπτει το σώμα του γράμματος.

## ***6.2 Εξαγωγή Περιγραμμάτων των Επιλεχθέντων Γραμμάτων των Κειμένων του Ρήγα Φεραίου***

Μετά την επιλογή των παραμέτρων που δίνουν τη βέλτιστη κατάτμηση των εικόνων γραμμάτων, ακολουθούν τα στάδια εξαγωγής του περιγράμματος για τη μετέπειτα μελέτη και σύγκριση των γραμμάτων. Παρακάτω παρουσιάζονται τα κατατμημένα γράμματα, τα αποτελέσματα από τα τρία στάδια εξαγωγής περιγράμματος και το τελικό περίγραμμα πάνω στο αρχικό μας γράμμα:



*Σχήμα 28: Κατατμημένα γράμματα*



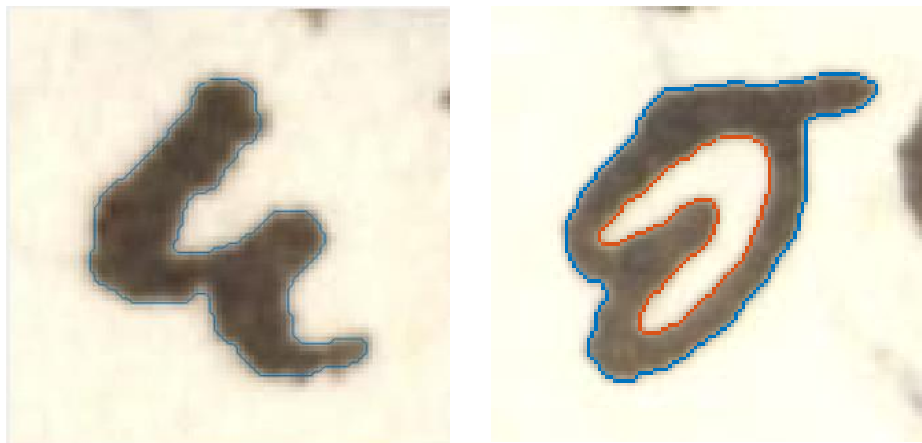
*Σχήμα 29: Αποτελέσματα περιγραμμάτων 1<sup>ου</sup> σταδίου*



Σχήμα 30: Αποτελέσματα περιγραμμάτων 2<sup>ου</sup> σταδίου



Σχήμα 31: Αποτελέσματα περιγραμμάτων 3<sup>ου</sup> σταδίου



Σχήμα 32: Τελικά αποτελέσματα περιγραμμάτων

## 6.3 Αποτελέσματα Μεθόδου για τα Επιλεγθέντα Γράμματα

### 6.3.1 Σύγκριση Υλοποιήσεων επιλεγμένων γραμμάτων κάθε έργου

Στο πλαίσιο της παρούσης Υποενότητας παρουσιάζουμε τα αποτελέσματα της αυτοματοποιημένης σύγκρισης για 2 ζεύγη γραμμάτων («ε» και «π» αντίστοιχα) για κάθε έργο ξεχωριστά.

Στα σχήματα που απεικονίζουν τα γράμματα στη θέση βέλτιστης προσαρμογής με μαύρο απεικονίζονται τα pixels που ανήκουν στην τομή των εμβαδών των σωμάτων των γραμμάτων, σε κλίμακα του γκρι απεικονίζονται τα pixels που ανήκουν στην ένωση των εμβαδών και με λευκό απεικονίζεται το υπόβαθρο της σελίδας.

*Η Δοκιμασία της Φιλίας*

---

Επιλέγουμε τα εξής γράμματα «ε» από το δείγμα των γραμμάτων που διαθέτουμε:



Σχήμα 33: Τα επιλεγθέντα γράμματα «ε», όπου η αρίθμηση θα χρησιμοποιηθεί στη σύγκριση των γραμμάτων

Με βάση τη μεθοδολογία που διατυπώθηκε στην Ενότητα 5.5, θεωρούμε την πρώτη υλοποίηση του γράμματος «ε» ως πρότυπο και προσαρμόζουμε σε αυτή τις υλοποιήσεις των υπόλοιπων 3 γραμμάτων. Επομένως, θα έχουμε σύγκριση πρώτου με δεύτερου, πρώτου με τρίτο και πρώτο με τέταρτο. Από τις συγκρίσεις αυτές προκύπτει η τιμή του κριτηρίου του

εμβαδού της τομής δια του εμβαδού της ένωσης των δύο υλοποιήσεων, δηλαδή οι ποσότητες  $\delta E_{1,2}, \delta E_{1,3}, \delta E_{1,4}, \delta E_{2,3}, \delta E_{2,4}, \delta E_{3,4}$ :



Σχήμα 34: Σύγκριση του πρώτου με το δεύτερο γράμμα. Το αποτέλεσμα του κριτηρίου τομής δια ένωση είναι  $\delta E_{1,2} = 0.6944$



Σχήμα 35: Σύγκριση του πρώτου με το δεύτερο γράμμα. Το αποτέλεσμα του κριτηρίου τομής δια ένωση είναι  $\delta E_{1,3} = 0.6347$



Σχήμα 36: Σύγκριση του πρώτου με το τέταρτο γράμμα. Το αποτέλεσμα του κριτηρίου τομής δια ένωση είναι  $\delta E_{1,4} = 0,7215$

Στη συνέχεια, θεωρούμε ως πρότυπο τη δεύτερη υλοποίηση του γράμματος «ε» και προσαρμόζουμε σε αυτή τις υπόλοιπες δύο υλοποιήσεις που απομένουν.



Σχήμα 37: Σύγκριση του δεύτερου με το τρίτο γράμμα. Το αποτέλεσμα του κριτηρίου τομής δια ένωση είναι  $\delta E_{2,3} = 0,6965$



Σχήμα 38: Σύγκριση του δεύτερου με το τέταρτο γράμμα. Το αποτέλεσμα του κριτηρίου τομής δια ένωση είναι  $\delta E_{2,4} = 0.8065$

Τέλος, θεωρούμε ως πρότυπο την τρίτη υλοποίηση του γράμματος «ε» και προσαρμόζουμε σε αυτή την τέταρτη και τελευταία υλοποίηση που απομένει.



Σχήμα 39: Σύγκριση του τρίτου με το τέταρτο γράμμα. Το αποτέλεσμα του κριτηρίου τομής δια ένωση είναι  $\delta E_{3,4} = 0,7164$



Ακολουθούμε ακριβώς την ίδια διαδικασία για τα γράμματα «π», οπότε αρχικά επιλέγουμε τα εξής γράμματα «π» από το δείγμα των γραμμάτων που διαθέτουμε:



Σχήμα 40: Τα επιλεγθέντα γράμματα «π», όπου η αρίθμηση θα χρησιμοποιηθεί στη σύγκριση των γραμμάτων

Και με βάση τη μεθοδολογία της Ενότητας 5.5, θεωρούμε τις εξής συγκρίσεις:



Σχήμα 41: Σύγκριση του πρώτου με το δεύτερο γράμμα. Το αποτέλεσμα του κριτηρίου τομής δια ένωση είναι  $\delta E_{1,2} = 0.6616$



Σχήμα 42: Σύγκριση του πρώτου με το τρίτο γράμμα. Το αποτέλεσμα του κριτηρίου τομής δια ένωση είναι  $\delta E_{1,3} = 0,7837$



Σχήμα 43: Σύγκριση του πρώτου με το τέταρτο γράμμα. Το αποτέλεσμα του κριτηρίου τομής δια ένωση είναι  $\delta E_{1,4} = 0,6221$

Στη συνέχεια, θεωρούμε ως πρότυπο τη δεύτερη υλοποίηση του γράμματος «ε» και προσαρμόζουμε σε αυτή τις υπόλοιπες δύο υλοποιήσεις που απομένουν.



Σχήμα 44: Σύγκριση του δεύτερου με το τρίτο γράμμα. Το αποτέλεσμα του κριτηρίου τομής δια ένωση είναι  $\delta E_{2,3} = 0,7554$



Σχήμα 45: Σύγκριση του δεύτερου με το τέταρτο γράμμα. Το αποτέλεσμα του κριτηρίου τομής δια ένωση είναι  $\delta E_{2,4} = 0,6120$

Τέλος, θεωρούμε ως πρότυπο την τρίτη υλοποίηση του γράμματος «ε» και προσαρμόζουμε σε αυτή την τέταρτη και τελευταία υλοποίηση που απομένει.



Σχήμα 46: Σύγκριση του τρίτου με το τέταρτο γράμμα. Το αποτέλεσμα του κριτηρίου τομής δια ένωση είναι  $\delta E_{3,4} = 0,6131$

Επιλέγουμε τα γράμματα:



Σχήμα 47: Τα τέσσερα υπό σύγκριση γράμματα, όπου η αρίθμηση θα χρησιμοποιηθεί στη σύγκριση των γραμμάτων.

Όπως και πριν, θεωρούμε την πρώτη υλοποίηση του γράμματος «ε» ως πρότυπο και προσαρμόζουμε σε αυτό τις υλοποιήσεις των υπόλοιπων 3 γραμμάτων:



Σχήμα 48: Σύγκριση του πρώτου με το δεύτερο γράμμα. Το αποτέλεσμα του κριτηρίου τομής δια ένωση είναι  $\delta E_{1,2} = 0,5891$



Σχήμα 49: Σύγκριση του πρώτου με το τρίτο γράμμα. Το αποτέλεσμα του κριτηρίου τομής δια ένωση είναι  $\delta E_{1,3} = 0,53223$



Σχήμα 50: Σύγκριση του πρώτου με το τέταρτο γράμμα. Το αποτέλεσμα του κριτηρίου τομής δια ένωση είναι  $\delta E_{1,4} = 0,5969$

Στη συνέχεια, θεωρούμε ως πρότυπο τη δεύτερη υλοποίηση του γράμματος «ε» και προσαρμόζουμε σε αυτή τις υπόλοιπες δύο υλοποιήσεις που απομένουν.

Σχήμα 51: Σύγκριση του δεύτερου με το τρίτο γράμμα. Το αποτέλεσμα του κριτηρίου τομής δια ένωση είναι  $\delta E_{2,3} = 0,7784$

Σχήμα 52: Σύγκριση του δεύτερου με το τέταρτο γράμμα. Το αποτέλεσμα του κριτηρίου τομής δια ένωση είναι  $\delta E_{2,4} = 0,8173$

Τέλος, θεωρούμε ως πρότυπο την τρίτη υλοποίηση του γράμματος «ε» και προσαρμόζουμε σε αυτή την τέταρτη και τελευταία υλοποίηση που απομένει.



Σχήμα 53: Σύγκριση του τρίτου με το τέταρτο γράμμα. Το αποτέλεσμα του κριτηρίου τομής δια ένωση είναι  $\delta E_{3,4} = 0,8020$

Ακολουθούμε ακριβώς την ίδια διαδικασία για τα γράμματα «π», οπότε αρχικά επιλέγουμε τα εξής γράμματα «π» από το δείγμα των γραμμάτων που διαθέτουμε:



Σχήμα 54: Τα επιλεγθέντα γράμματα «π», όπου η αρίθμηση θα χρησιμοποιηθεί στη σύγκριση των γραμμάτων

Και με βάση τη μεθοδολογία της Ενότητας 5.5, θεωρούμε τις εξής συγκρίσεις:





Σχήμα 55: Σύγκριση του πρώτου με το δεύτερο γράμμα. Το αποτέλεσμα του κριτηρίου τομής δια ένωση είναι  $\delta E_{1,2} = 0,6923$



Σχήμα 56: Σύγκριση του πρώτου με το δεύτερο γράμμα. Το αποτέλεσμα του κριτηρίου τομής δια ένωση είναι  $\delta E_{1,3} = 0,7519$



Σχήμα 57: Σύγκριση του πρώτου με το τέταρτο γράμμα. Το αποτέλεσμα του κριτηρίου τομής δια ένωση είναι  $\delta E_{1,4} = 0,6301$

Στη συνέχεια, θεωρούμε ως πρότυπο τη δεύτερη υλοποίηση του γράμματος «ε» και προσαρμόζουμε σε αυτή τις υπόλοιπες δύο υλοποιήσεις που απομένουν.



Σχήμα 58: Σύγκριση του δεύτερου με το τρίτο γράμμα. Το αποτέλεσμα του κριτηρίου τομής δια ένωση είναι  $\delta E_{2,3} = 0,6878$



Σχήμα 59: Σύγκριση του δεύτερου με το τέταρτο γράμμα. Το αποτέλεσμα του κριτηρίου τομής δια ένωση είναι  $\delta E_{2,4} = 0,6145$

Τέλος, θεωρούμε ως πρότυπο την τρίτη υλοποίηση του γράμματος «π» και προσαρμόζουμε σε αυτή την τέταρτη και τελευταία υλοποίηση που απομένει.



Σχήμα 60: Σύγκριση του τρίτου με το τέταρτο γράμμα. Το αποτέλεσμα του κριτηρίου τομής δια ένωση είναι  $\delta E_{3,4} = 0,6993$

Επιλέγουμε τα γράμματα «ε»:



Σχήμα 61: Τα τέσσερα υπό σύγκριση γράμματα, όπου η αρίθμηση θα χρησιμοποιηθεί για τη σύγκρισή τους

Όπως και πριν, θεωρούμε την πρώτη υλοποίηση του γράμματος «ε» ως πρότυπο και προσαρμόζουμε σε αυτό τις υλοποιήσεις των υπόλοιπων 3 γραμμάτων:



Σχήμα 62: Σύγκριση του πρώτου με το δεύτερο γράμμα. Το αποτέλεσμα του κριτηρίου τομής δια ένωση είναι  $\delta E_{1,2} = 0,8213$



Σχήμα 63: Σύγκριση του πρώτου με το τρίτο γράμμα. Το αποτέλεσμα του κριτηρίου τομής δια ένωση είναι  $\delta E_{1,3} = 0,6925$



Σχήμα 64: Σύγκριση του πρώτου με το τέταρτο γράμμα. Το αποτέλεσμα του κριτηρίου τομής δια ένωση είναι  $\delta E_{1,4} = 0,7919$

Στη συνέχεια, θεωρούμε όπως και πριν ως πρότυπο τη δεύτερη υλοποίηση του γράμματος «ε» και προσαρμόζουμε σε αυτή τις υπόλοιπες δύο υλοποιήσεις που απομένουν.



Σχήμα 65: Σύγκριση του δεύτερου με το τρίτο γράμμα. Το αποτέλεσμα του κριτηρίου τομής δια ένωση είναι  $\delta E_{2,3} = 0,6986$



Σχήμα 66: Σύγκριση του δεύτερου με το τέταρτο γράμμα. Το αποτέλεσμα του κριτηρίου τομής δια ένωση είναι  $\delta E_{2,4} = 0,7594$

Τέλος, θεωρούμε ως πρότυπο την τρίτη υλοποίηση του γράμματος «ε» και προσαρμόζουμε σε αυτή την τέταρτη και τελευταία υλοποίηση που απομένει.



Σχήμα 67: Σύγκριση του τρίτου με το τέταρτο γράμμα. Το αποτέλεσμα του κριτηρίου τομής δια ένωση είναι  $\delta E_{3,4} = 0,7877$

Ακολουθούμε ακριβώς την ίδια διαδικασία για τα γράμματα «π», οπότε αρχικά επιλέγουμε τα εξής γράμματα «π» από το δείγμα των γραμμάτων που διαθέτουμε:



Σχήμα 68: Τα επιλεγθέντα γράμματα «π», όπου η αρίθμηση θα χρησιμοποιηθεί στη σύγκριση των γραμμάτων

Και με βάση τη μεθοδολογία της Ενότητας 5.5, θεωρούμε τις εξής συγκρίσεις:



Σχήμα 69: Σύγκριση του πρώτου με το δεύτερο γράμμα. Το αποτέλεσμα του κριτηρίου τομής δια ένωση είναι  $\delta E_{1,2} = 0,6455$



Σχήμα 70: Σύγκριση του πρώτου με το δεύτερο γράμμα. Το αποτέλεσμα του κριτηρίου τομής δια ένωση είναι  $\delta E_{1,3} = 0,6417$





Σχήμα 71: Σύγκριση του πρώτου με το τέταρτο γράμμα. Το αποτέλεσμα του κριτηρίου τομής δια ένωση είναι  $\delta E_{1,4} = 0,6112$

Στη συνέχεια, θεωρούμε ως πρότυπο τη δεύτερη υλοποίηση του γράμματος «π» και προσαρμόζουμε σε αυτή τις υπόλοιπες δύο υλοποιήσεις που απομένουν.



Σχήμα 72: Σύγκριση του δεύτερου με το τρίτο γράμμα. Το αποτέλεσμα του κριτηρίου τομής δια ένωση είναι  $\delta E_{2,3} = 0,6821$



Σχήμα 73: Σύγκριση του δεύτερου με το τέταρτο γράμμα. Το αποτέλεσμα του κριτηρίου τομής δια ένωση είναι  $\delta E_{2,4} = 0,6158$

Τέλος, θεωρούμε ως πρότυπο την τρίτη υλοποίηση του γράμματος «π» και προσαρμόζουμε σε αυτή την τέταρτη και τελευταία υλοποίηση που απομένει.



Σχήμα 74: Σύγκριση του τρίτου με το τέταρτο γράμμα. Το αποτέλεσμα του κριτηρίου τομής δια ένωση είναι  $\delta E_{3,4} = 0,6196$

Επιλέγουμε τα γράμματα:



Σχήμα 75: Τα τέσσερα υπό σύγκριση γράμματα, όπου η αρίθμηση θα χρησιμοποιηθεί για τη σύγκρισή τους

Όπως και πριν, θεωρούμε την πρώτη υλοποίηση του γράμματος «ε» ως πρότυπο και προσαρμόζουμε σε αυτό τις υλοποιήσεις των υπόλοιπων 3 γραμμάτων:



Σχήμα 76: Σύγκριση του πρώτου με το δεύτερο γράμμα. Το αποτέλεσμα του κριτηρίου τομής δια ένωση είναι  $\delta E_{1,2} = 0,7310$



Σχήμα 77: Σύγκριση του πρώτου με το τρίτο γράμμα. Το αποτέλεσμα του κριτηρίου τομής δια ένωση είναι  $\delta E_{1,3} = 0,7219$



Σχήμα 78: Σύγκριση του πρώτου με το τέταρτο γράμμα. Το αποτέλεσμα του κριτηρίου τομής δια ένωση είναι  $\delta E_{1,4} = 0,6284$

Στη συνέχεια, θεωρούμε όπως και πριν ως πρότυπο τη δεύτερη υλοποίηση του γράμματος «π» και προσαρμόζουμε σε αυτή τις υπόλοιπες δύο υλοποιήσεις που απομένουν.



Σχήμα 79: Σύγκριση του δεύτερου με το τρίτο γράμμα. Το αποτέλεσμα του κριτηρίου τομής δια ένωση είναι  $\delta E_{2,3}=0,7329$



Σχήμα 80: Σύγκριση του δεύτερου με το τέταρτο γράμμα. Το αποτέλεσμα του κριτηρίου τομής δια ένωση είναι  $\delta E_{2,4} = 0,7274$

Τέλος, θεωρούμε ως πρότυπο την τρίτη υλοποίηση του γράμματος «ε» και προσαρμόζουμε σε αυτή την τέταρτη και τελευταία υλοποίηση που απομένει.



Σχήμα 81: Σύγκριση του τρίτου με το τέταρτο γράμμα. Το αποτέλεσμα του κριτηρίου τομής δια ένωση είναι  $\delta E_{3,4}=0,7930$

Ακολουθούμε ακριβώς την ίδια διαδικασία για τα γράμματα «π», οπότε αρχικά επιλέγουμε τα εξής γράμματα «π» από το δείγμα των γραμμάτων που διαθέτουμε:



Σχήμα 82: Τα επιλεγθέντα γράμματα «π», όπου η αρίθμηση θα χρησιμοποιηθεί στη σύγκριση των γραμμάτων

Και με βάση τη μεθοδολογία της Ενότητας 5.5, θεωρούμε τις εξής συγκρίσεις:



Σχήμα 83: Σύγκριση του πρώτου με το δεύτερο γράμμα. Το αποτέλεσμα του κριτηρίου τομής δια ένωση είναι  $\delta E_{1,2}=0,7675$



Σχήμα 84: Σύγκριση του πρώτου με το δεύτερο γράμμα. Το αποτέλεσμα του κριτηρίου τομής δια ένωση είναι  $\delta E_{1,3}=0,6924$



Σχήμα 85: Σύγκριση του πρώτου με το τέταρτο γράμμα. Το αποτέλεσμα του κριτηρίου τομής δια ένωση είναι  $\delta E_{1,4}=0,6478$

Στη συνέχεια, θεωρούμε ως πρότυπο τη δεύτερη υλοποίηση του γράμματος «ε» και προσαρμόζουμε σε αυτή τις υπόλοιπες δύο υλοποιήσεις που απομένουν.



Σχήμα 86: Σύγκριση του δεύτερου με το τρίτο γράμμα. Το αποτέλεσμα του κριτηρίου τομής δια ένωση είναι  $\delta E_{2,3}=0,6612$





Σχήμα 87: Σύγκριση του δεύτερου με το τέταρτο γράμμα. Το αποτέλεσμα του κριτηρίου τομής δια ένωση είναι  $\delta E_{2,4}=0,6145$

Τέλος, θεωρούμε ως πρότυπο την τρίτη υλοποίηση του γράμματος «ε» και προσαρμόζουμε σε αυτή την τέταρτη και τελευταία υλοποίηση που απομένει.



Σχήμα 88: Σύγκριση του τρίτου με το τέταρτο γράμμα. Το αποτέλεσμα του κριτηρίου τομής δια ένωση είναι  $\delta E_{3,4}=0,6188$

### 6.3.2 Σύγκριση Υλοποιήσεων επιλεγμένων γραμμάτων μεταξύ διαφορετικών έργων

Με βάση το δεύτερο βήμα της μεθοδολογίας που διατυπώθηκε στην Υποενότητα 5.5, προσαρμόζουμε όλες τις 4 υλοποιήσεις του εκάστοτε γράμματος ενός κειμένου J σε όλες τις υλοποιήσεις του ίδιου γράμματος του κειμένου I, θεωρώντας κάθε φορά ως πρότυπο ένα μόνο γράμμα του κειμένου I. Συγκεκριμένα, θα συγκρίνουμε τις υλοποιήσεις των υπό μελέτη κειμένων και θα παρουσιάσουμε ενδεικτικά τα αποτελέσματα των 4 συγκρίσεων.

*Σύγκριση της «Δοκιμασμένης Φιλίας» και του έργου «Φυσικής Απάνθισμα»*

---



Σχήμα 89: Σύγκριση του τρίτου γράμματος της «Δοκιμασμένης Φιλίας» με το τέταρτο γράμμα του έργου «Φυσικής Απάνθισμα». Το αποτέλεσμα του κριτηρίου τομής δια ένωση είναι  $\delta E_{3,4}=0,6932$



Σχήμα 90: Σύγκριση του τρίτου «π» της «Δοκιμασμένης Φιλίας» με το τρίτο «π» του έργου «Φυσικής Απάνθισμα».  
Το αποτέλεσμα του κριτηρίου τομής δια ένωση είναι  $\delta E_{3,3} =$



Σχήμα 91: Σύγκριση του τέταρτου γράμματος της «Δοκιμασμένης Φιλίας» με το τέταρτο γράμμα του έργου «Φυσικής Απάνθισμα». Το αποτέλεσμα του κριτηρίου τομής δια ένωση είναι  $\delta E_{4,4} = 0,7372$



Σχήμα 92: Σύγκριση του τέταρτου γράμματος του «Φυσικής Απάνθισμα» της με το τέταρτο «ε» της «Δοκιμασμένης Φιλίας». Το αποτέλεσμα του κριτηρίου τομής δια ένωση είναι  $\delta E_{4,4}=0,7371$

Σύγκριση της «Δοκιμασμένης Φιλίας» και του έργου «Σαγανάκι της Τρέλλας»

---



Σχήμα 93: Σύγκριση του τρίτου γράμματος της «Δοκιμασμένης Φιλίας» με το ένατο γράμμα του έργου «Σαγανάκι της Τρέλλας». Το αποτέλεσμα του κριτηρίου τομής δια ένωση είναι  $\delta E_{3,4}=0,7168$



Σχήμα 94: Σύγκριση του πρώτου «π» από το «Σαγανάκι της Τρέλλας» με το τρίτο «π» του έργου «Δοκιμασμένης Φιλίας». Το αποτέλεσμα του κριτηρίου τομής δια ένωση είναι  $\delta E_{3,3}=0,7187$



Σχήμα 95: Σύγκριση του τέταρτου γράμματος της «Δοκιμασμένης Φιλίας» με το πρώτο γράμμα του έργου «Σαγανάκι της Τρέλλας». Το αποτέλεσμα του κριτηρίου τομής δια ένωση είναι  $\delta E_{4,4}=0,7620$



Σχήμα 96: Σύγκριση του πρώτου γράμματος του «Σαγανάκι της Τρέλλας» της με το τέταρτο «ε» της «Δοκιμασμένης Φιλίας». Το αποτέλεσμα του κριτηρίου τομής δια ένωση είναι  $\delta E_{4,4}=0,7607$

Σύγκριση της «Δοκιμασμένης Φιλίας» και του έργου «Σύνταγμα της Ελλάδος»

---



Σχήμα 97: Σύγκριση του τρίτου γράμματος της «Δοκιμασμένης Φιλίας» με το δέκατο έκτο γράμμα του «Συντάγματος». Το αποτέλεσμα του κριτηρίου τομής δια ένωση είναι  $\delta E_{3,16}=0,5675$



Σχήμα 98: Σύγκριση του πρώτου «π» από το «Δοκιμασμένης Φιλίας» με το τέταρτο «π» του «Συντάγματος». Το αποτέλεσμα του κριτηρίου τομής δια ένωση είναι  $\delta E_{3,4}=0,6064$



Σχήμα 99: Σύγκριση του τέταρτου γράμματος της «Δοκιμασμένης Φιλίας» με το δεύτερο γράμμα του έργου «Σύνταγμα». Το αποτέλεσμα του κριτηρίου τομής δια ένωση είναι  $\delta E_{4,2}=0,6675$



Σχήμα 100: Σύγκριση του δεύτερου γράμματος του «Συντάγματος» με το τέταρτο «ε» της «Δοκιμασμένης Φιλίας».  
Το αποτέλεσμα του κριτηρίου τομής δια ένωση είναι  $\delta E_{2,4}=0,6665$

Σύγκριση του έργου «Φυσικής Απάνθισμα» με το έργο «Σαγανάκι της Τρέλλας»

---



Σχήμα 101: Σύγκριση του τρίτου γράμματος του «Φυσικής Απάνθισμα» με το πρώτο γράμμα του έργου «Σαγανάκι της Τρέλλας». Το αποτέλεσμα του κριτηρίου τομής δια ένωση είναι  $\delta E_{1,3}=0,6341$





Σχήμα 102: Σύγκριση του πρώτου γράμματος του «Σαγανάκι της Τρέλλας» με το τρίτο γράμμα του έργου «Φυσικής Απάνθισμα». Το αποτέλεσμα του κριτηρίου τομής δια ένωση είναι  $\delta E_{3,1}=0,6311$



Σχήμα 103: Σύγκριση του τέταρτου γράμματος του «Φυσικής Απάνθισμα» με το πρώτο γράμμα του έργου «Σαγανάκι της Τρέλλας». Το αποτέλεσμα του κριτηρίου τομής δια ένωση είναι  $\delta E_{4,1}=0,7655$



Σχήμα 104: Σύγκριση του πρώτου γράμματος από το «Σαγανάκι της Τρέλλας» με το τέταρτο «ε» του «Φυσικής Απάνθισμα». Το αποτέλεσμα του κριτηρίου τομής δια ένωση είναι  $\delta E_{1,4}=0,7565$

Σύγκριση του έργου «Φυσικής Απάνθισμα» με το έργο «Σύνταγμα της Ελλάδος»

---



Σχήμα 105: Σύγκριση του τρίτου γράμματος του «Φυσικής Απάνθισμα» με το τέταρτο γράμμα του «Συντάγματος». Το αποτέλεσμα του κριτηρίου τομής δια ένωση είναι  $\delta E_{3,4}=0,5499$



Σχήμα 106: Σύγκριση του τέταρτου γράμματος του «Συντάγματος» με το τρίτο γράμμα του έργου «Φυσικής Απάνθισμα». Το αποτέλεσμα του κριτηρίου τομής δια ένωση είναι  $\delta E_{4,3}=0,5590$



Σχήμα 107: Σύγκριση του τέταρτου γράμματος του «Φυσικής Απάνθισμα» με το δεύτερο γράμμα του έργου «Σύνταγμα». Το αποτέλεσμα του κριτηρίου τομής δια ένωση είναι  $\delta E_{4,2}=0,6692$



Σχήμα 108: Σύγκριση του δεύτερου γράμματος από το «Σύνταγμα» με το τέταρτο «ε» του «Φυσικής Απάνθισμα». Το αποτέλεσμα του κριτηρίου τομής δια ένωση είναι  $\delta E_{2,4} = 0,6650$

Σύγκριση του έργου «Σαγανάκι της Τρέλλας» με το έργο «Σύνταγμα της Ελλάδος»

---



Σχήμα 109: Σύγκριση του πρώτου γράμματος του «Σαγανάκι της Τρέλλας» με το τέταρτο γράμμα του «Συντάγματος». Το αποτέλεσμα του κριτηρίου τομής δια ένωση είναι  $\delta E_{1,4} = 0,6064$



Σχήμα 110: Σύγκριση του τέταρτου γράμματος του «Σαγανάκι της Τρέλλας» με το πρώτο γράμμα του «Συντάγματος». Το αποτέλεσμα του κριτηρίου τομής δια ένωση είναι  $\delta E_{4,1}=0,6115$



Σχήμα 111: Σύγκριση του πρώτου γράμματος από το «Σαγανάκι της Τρέλλας» με το δεύτερο γράμμα του «Συντάγματος». Το αποτέλεσμα του κριτηρίου τομής δια ένωση είναι  $\delta E_{1,2}=0,6707$



*Σχήμα 112: Σύγκριση του δεύτερου γράμματος από το «Σύνταγμα» με το πρώτο «ε» του «Σαγανάκι της Τρέλλας».*  
*Το αποτέλεσμα του κριτηρίου τομής δια ένωση είναι  $\delta E_{2,1}=0,6752$*

# 7

## *Επίλογος*

### **7.1** *Σύνοψη και συμπεράσματα*

Με βάση τα αποτελέσματα της παραπάνω διαδικασίας, φαίνεται πως τα γράμματα «ε» και «π» στα χειρόγραφα των έργων «Το Σαγανάκι της Τρέλλας», «Η Δοκιμασμένη Φιλία» και το «Φυσικής Απάνθισμα» προσφέρουν αρχικά σοβαρή πιθανότητα να είναι γραμμένα από το ίδιο χέρι. Αντίθετα, το «Σύνταγμα» φαίνεται να είναι γραμμένο από διαφορετικό χέρι, σύμφωνα με τα πρώτα στατιστικά αποτελέσματα. Είναι, επομένως, απολύτως απαραίτητο να εκτελεστούν οι σχετικοί αλγόριθμοι ταυτοποίησης γραφέα αναφορικά με περισσότερες υλοποιήσεις των συμβόλων της αλφαβήτου «π, ε».

Επιπλέον, για την επιβεβαίωση των εν λόγω συμπερασμάτων απαιτείται περαιτέρω ανάλυση των χειρογράφων και συγκεκριμένα η πραγματοποίηση της παραπάνω ανάλυσης για μεγαλύτερο σύνολο συμβόλων της αλφαβήτου (τόσο συμφώνων όσο και φωνηέντων) πέραν των «π, ε».

### **7.2** *Μελλοντικές επεκτάσεις*

Για την πληρέστερη μελέτη των χειρογράφων των έργων του Ρήγα Φεραίου, απαιτείται και η μελέτη των χειρογράφων του έργου «Σχολείο των Ντελικάτων Εραστών» και συγκεκριμένα του Συμυρναϊκού χειρογράφου. Το εν λόγω χειρόγραφο αν και αρχικά θεωρήθηκε αυτόγραφο, σύντομα αμφισβητήθηκε ως αντίγραφο του τυπωμένου βιβλίου, όπως αναφέρεται στο [38]. Επομένως, τίθεται το θέμα της επεξεργασίας του για τον εντοπισμό του χεριού που έγραψε.

# 8

## *Βιβλιογραφία*

- [1] Δ. Καραμπερόπουλος, “Τελικά δεν είναι ο Ρήγας ο συγγραφέας των «Ανέκδοτων κειμένων» με την κωμωδία «Σαγανάκι της τρέλλας»,” Έκδοση Επιστημονικής Εταιρείας Μελέτης Φερών-Βελεστίνου-Ρήγα, Αθήνα, Μάρτιος 2012. Διαθέσιμο στο σύνδεσμο: <http://karaberopoulos.gr/>.
- [2] Π. Ρουσόπουλος, "Ανάπτυξη Προηγμένων Μεθόδων Αναγνώρισης Προτύπων Και Μαθηματική Θεμελίωση Για Τον Εντοπισμό Του Τρόπου Δημιουργίας Και Της Ταυτότητας Του Δημιουργού Σημαντικών Αρχαιολογικών Ευρημάτων," Διδακτορική διατριβή, Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών, Εθνικό Μετσόβιο Πολυτεχνείο, Αθήνα, 2010.
- [3] Δ. Ντάφη, "Ψηφιακή Επεξεργασία, Προτυποποίηση (modeling) Και Εικονική Ανακατασκευή Σημαντικών Αρχαιολογικών Ευρημάτων", Διδακτορική Διατριβή, Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών, Εθνικό Μετσόβιο Πολυτεχνείο, Αθήνα, 2011.
- [4] Φ. Γιαννόπουλος, "Πρωτότυπες Μέθοδοι Αυτόματης Αναγνώρισης Γραφέα και Αντιμετώπισης Υπολογιστικών Προβλημάτων Υλοποίησης," Διδακτορική Διατριβή, Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών, Εθνικό Μετσόβιο Πολυτεχνείο, Αθήνα, 2018.
- [5] T. Dhieb, W. Ouarda, H. Boubaker and A. M. Alimi, “Beta-Elliptic Model for Writer Identification from Online Arabic Handwriting,” *Journal of Information Assurance & Security*, vol. 11, no. 5, pp. 263-272, 2016
- [6] X.Y. Zhang, G. S. Xie, C. L. Liu and Y. Bengio, “End-to-End Online Writer Identification With Recurrent Neural Network,” *IEEE Transactions on Human-Machine Systems*, vol. 47, no. 2, pp. 285-292, April 2017. doi: 10.1109/THMS.2016.2634921



- [7] J.Chapran, "Biometric Writer Identification: Feature Analysis and Classification," International Journal of Pattern Recognition and Artificial Intelligence, vol. 20, no. 4, pp. 483-503, 2006.
- [8] E.N. Zois and V. Anastassopoulos, "Morphological Waveform Coding for Writer Identification," Pattern Recognition, vol. 33, pp. 385-398, 2000.
- [9] H. Said, T. Tan and K. Baker, "Personal Identification Based on Handwriting," Pattern Recognition, vol. 33, no. 1, pp. 149-160, January 2000.
- [10] Y. Zhu, T. Tan and Y. Wang, "Biometric Personal Identification Based on Handwriting," Proceedings 15th International Conference on Pattern Recognition. ICPR-2000, Barcelona, 2000, vol. 2, pp. 797-800. doi: 10.1109/ICPR.2000.906196
- [11] U.V. Marti, R. Messerli and H. Bunke, "Writer Identification Using Text Line Based Features," Proceedings 6th International Conference on Document Analysis and Recognition. Seattle, WA, 2001, pp. 101-105. doi: 10.1109/ICDAR.2001.953763
- [12] A. Bensefia, T. Paquet and L. Heutte, "Handwriting Analysis for Writer Verification," Proceedings Ninth International Workshop on Frontiers in Handwriting Recognition, 2004, pp. 196-201. doi: 10.1109/IWFHR.2004.49
- [13] G. Leedham and S. Chachra, "Writer Identification Using Innovative Binarised Features of Handwritten Numerals," Proceedings Seventh International Conference on Document Analysis and Recognition, 2003, vol. 1, pp. 413-416. doi: 10.1109/ICDAR.2003.1227700
- [14] Pervouchine, V. and Leedham, G. "Extraction and analysis of forensic document examiner features used for writer identification", Pattern Recognition, vol 40, p. 1004 – 1013, 2007.
- [15] A. Schlapbach and H. Bunke, "A Writer Identification and Verification System Using HMM Based Recognizers," Pattern Analysis and Applications, vol. 10, no. 1, February 2007.
- [16] Z. He, X. You and Y. Yan Tang, "Writer Identification of Chinese Handwriting Documents Using Hidden Markov Tree Model," Pattern Recognition, 2008, vol. 41, no. 4, pp. 1295-1307.
- [17] A. Schlapbach and H. Bunke, "Off-Line Handwriting Identification Using HMM Based Recognizers," Proceedings 17th International Conference on Pattern Recognition, ICPR2004, 2004, vol. 2, pp. 654-658.
- [18] A. Schlapbach, H. Bunke, Off-line writer identification using gaussian mixture models, in: In Proc. the 18<sup>th</sup> International Conference on Pattern Recognition, 2006: pp. 992-995.

- [19] E. Khalifa, S. Al-maadeed, M. A. Tahir, A. Bouridane and A. Jamshed, "Off-Line Writer Identification Using an Ensemble of Grapheme Codebook Features," *Pattern Recognition Letters*, vol. 59, pp. 18-25, July 2015.
- [20] G. Ghiasi and R. Safabakhsh, "Offline Text-Independent Writer Identification Using Codebook and Efficient Code Extraction Methods," *Image and Vision Computing*, vol. 31, no. 5, pp. 379-391, May 2013.
- [21] Ralph Niels and Louis Vuurpijl. Generating copybooks from consistent handwriting styles. In *Proc. ICDAR 2007*, pages 1009–1013, Curitiba, Brazil, September 2007.
- [22] Ralph Niels, Louis Vuurpijl, and Lambert Schomaker, Automatic allograph matching in forensic writer identification. *International Journal of Pattern Recognition and Artificial Intelligence*, 21(1):61–81, February 2006.
- [23] R. Niels, F. Gootjen and L. Vuurpijl, "Writer Identification Through Information Retrieval: The Allograph Weight Vector," *Proceedings International Conference on Frontiers in Handwriting Recognition*, 2008, pp. 481-486.
- [24] J. Tan, J. H. Lai, C. D. Wang and M. S. Feng, "A Stroke Shape and Structure Based Approach for Off-Line Chinese Handwriting Identification," *International Journal of Intelligent Systems and Applications*, vol. 3, no. 2, pp. 1-8, March 2011.
- [25] E. Dalton and N. R. Howe, "Style-Based Retrieval for Ancient Syriac Manuscripts," *Proceedings 2011 Workshop on Historical Document Imaging and Processing, HIP-2011*, ACM, 2011, pp. 1-5.
- [26] M. Contreras Seitz, "Towards Chilean Spanish Language Diachronic Corpus [Hacia la Constitución de un Corpus Diacrónico del Español de Chile]," *Revista de Lingüística Teórica y Aplicada*, vol. 47, no. 2, pp. 11-134, 2009.
- [27] L. Schomaker, K. Franke and M. Bulacu, "Using Codebooks of Fragmented ConnectedComponent Contours in Forensic and Historic Writer Identification," *Pattern Recognition Letters*, vol. 28, no. 6, pp. 719-727, 2007.
- [28] I. Bar – Yosef, I. Beckman, K. Kedem and I. Dinstein, "Binarization, Character Extraction, and Writer Identification of Historical Hebrew Calligraphy Documents," *International Journal on Document Analysis and Recognition*, vol. 2, no. 9, pp. 89-99, 2007.
- [29] C. Adak and B. B. Chaudhuri, "Writer Identification from Offline Isolated Bangla Characters and Numerals," *Proceedings 13th International Conference on Document Analysis and Recognition, ICDAR-2015*, Tunis, 2015, pp. 486-490. doi: 10.1109/ICDAR.2015.7333809

- [30] R. Kumar, K. Kumar Ravulakollu and R. Bhat, “Fuzzy-Membership Based Writer Identification from Handwritten Devnagari Script,” *Journal of Information Processing Systems*, vol. 13, no. 4, pp. 893-913, 2017.
- [31] C. Halder, K. Thakur, S. Phadikar and K. Roy, “Writer Identification from Handwritten Devanagari Script,” *Proceedings Second International Conference on Information Systems Design and Intelligent Applications*, , India, 2015, vol. 2, pp. 497-505.
- [32] M. Niitsuma, L. Schomaker, J.-P. Van Oosten, Y. Tomita and D. Bell, “Musicologist-Driven Writer Identification in Early Music Manuscripts,” *Multimedia Tools and Applications*, vol. 75, no. 11, pp. 6463-6479, June 2016.
- [33] I. Siddigi, C. Djeddi, A. Raza and L. Souici-Meslati, “Automatic Analysis of Handwriting for Gender Classification,” *Pattern Analysis and Applications*, vol. 18, no. 4, pp. 887-899, November 2015.
- [34] Δ. Καραμπερόπουλος, “Η συνωμοτική δράση του Ρήγα Βελεστινλή. Το σχέδιο μιάς απραγματοποίητης επανάστασης”, *Ιστορικά Θέματα*, τεύχ. 41, σελ. 65-73, Ιούνιος 2005, Διαθέσιμο στο σύνδεσμο: <http://karaberopoulos.gr/>.
- [35] Δ. Καραμπερόπουλος, “Το Φυσικής απάνθισμα του Ρήγα Βελεστινλή, πρώτη απόπειρα σε «απλούν ύφος»”, Αθήνα 2010, Διαθέσιμο στο σύνδεσμο: <http://karaberopoulos.gr/>.
- [36] E. Legrand, “Ανέκδοτα Έγγραφα Περί Ρήγα Βελεστινλή Και Τω Συν Αυτώ Μαρτυρησαντων”, Έκδοση Επιστημονικής Εταιρείας Μελέτης Φερών-Βελεστίνου-Ρήγα, Αθήνα, Δεκέμβριος 2000
- [37] P. Βελεστινλής, “Ο Ηθικός Τρίπους, Βιέννη 1797-Επιμέλεια-Εισαγωγή-Ευρετήριο Δημ. Καραμπερόπουλος”, Έκδοση Επιστημονικής Εταιρείας Μελέτης Φερών-Βελεστίνου-Ρήγα, Αθήνα 2001
- [38] Λ. Βρανούσης, “Έκδόσεις και χειρόγραφα του «Σχολείου των ντελικάτων εραστών»”. *The Gleaner*, 8, σελ 14, 2016. doi: <https://doi.org/10.12681/er.9939>
- [39] P. Βελεστινλής, “Ρήγας: Ανέκδοτα Κείμενα”, Εκδόσεις Gutenberg, Δεκέμβριος 2011