



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

**ΕΦΑΡΜΟΓΗ ΑΛΓΟΡΙΘΜΩΝ ΚΑΤΑΤΜΗΣΗΣ ΕΙΚΟΝΑΣ  
ΓΡΑΜΜΑΤΩΝ, ΩΣ ΚΑΙ ΤΑΥΤΟΠΟΙΗΣΗΣ ΓΡΑΦΕΩΝ  
ΣΥΓΧΡΟΝΩΝ ΧΕΙΡΟΓΡΑΦΩΝ**

**ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ**

**ΑΝΤΩΝΙΟΣ Ι. ΑΝΤΩΝΙΑΔΗΣ**

**ΝΙΚΟΛΑΟΣ Ι. ΣΑΡΡΗΣ**

**Επιβλέπων:** Κωνσταντίνος Παπαοδυσσεύς  
Καθηγητής ΕΜΠ

Αθήνα, Αύγουστος 2013





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

**ΕΦΑΡΜΟΓΗ ΑΛΓΟΡΙΘΜΩΝ ΚΑΤΑΤΜΙΣΗΣ ΕΙΚΟΝΑΣ  
ΓΡΑΜΜΑΤΩΝ, ΩΣ ΚΑΙ ΤΑΥΤΟΠΟΙΗΣΗΣ ΓΡΑΦΕΩΝ  
ΣΥΓΧΡΟΝΩΝ ΧΕΙΡΟΓΡΑΦΩΝ**

**ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ**

**ΑΝΤΩΝΙΟΣ Ι. ΑΝΤΩΝΙΑΔΗΣ**

**ΝΙΚΟΛΑΟΣ Ι. ΣΑΡΡΗΣ**

**Επιβλέπων:** Κωνσταντίνος Παπαοδυσσεύς  
Καθηγητής ΕΜΠ

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 26<sup>η</sup> Αυγούστου 2013.

.....

.....

.....

Κωνσταντίνος Παπαοδυσσεύς  
Καθηγητής Ε.Μ.Π.

Ηλίας Κουκούτσης  
Επ.Καθηγητής Ε.Μ.Π.

Καμπουράκης Γεώργιος  
Αν.Καθηγητής Ε.Μ.

Αθήνα, Αύγουστος 2013



.....

Αντώνιος Ι. Αντωνιάδης

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π

.....

Νικόλαος Ι. Σαρρής

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π

Copyright Αντώνιος Ι. Αντωνιάδης, 2013

Copyright Νικόλαος Ι. Σαρρής, 2013

Copyright Κωνσταντίνος Παπαποδυσσεύς, 2013

Με επιφύλαξη παντός δικαιώματος. 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 τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς το συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευτούν ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσοβίου Πολυτεχνείου.



## **ΕΥΧΑΡΙΣΤΙΕΣ**

Φτάνοντας στο τέλος αυτής της διπλωματικής εργασίας αισθανόμαστε την ανάγκη να εκφράσουμε τις ευχαριστίες μας σε όλους όσους στήριξαν την προσπάθεια μας όλο αυτό τον καιρό.

Πρώτον, ευχαριστούμε τον καθηγητή κ. Παπαοδυσσέα για την επίβλεψη και την καθοδήγηση της διπλωματικής μας εργασίας καθώς επίσης για τη σπουδαία ακαδημαϊκή προσφορά του στη σχολή καθ' όλη τη διάρκεια των προπτυχιακών μας χρόνων.

Επίσης θα θέλαμε να ευχαριστήσουμε τους υποψήφιους διδάκτορες Δημήτρη Αραμπατζή και Φώτιο Γιαννόπουλο για την άψογη συνεργασία και υποστήριξη που μας παρείχαν όλο αυτό τον καιρό.

Τέλος θέλουμε να ευχαριστήσουμε τις οικογένειες μας και το περιβάλλον μας για τη βοήθεια που μας παρείχαν όλο αυτό καιρό σε όλες μας τις προσπάθειες.





## ΠΕΡΙΛΗΨΗ

Το ακαδημαϊκό ενδιαφέρον της παρούσας διπλωματικής εργασίας εστιάζεται στην αυτόματη αναγνώριση γραφέα αρχαίων και σύγχρονων κειμένων από τα γραπτά του και κατ' επέκταση του χρονικού προσδιορισμού αρχαίων επιγραφών . Αρχικά επιχειρείται να παρουσιαστεί στον αναγνώστη μια περιγραφή του προβλήματος που αντιμετωπίστηκε, της σπουδαιότητας της μελέτης καθώς και των δυσκολιών που κρύβονται πίσω από την προσπάθεια επίλυσης αυτού του προβλήματος.

Στη συνέχεια και αφού γίνεται μια σύντομη περιγραφή της μεθόδου, γίνεται μια ενδελεχής παρουσίαση των κριτηρίων και του μαθηματικού υποβάθρου που χρησιμοποιήθηκαν για να επιτευχτεί ο στόχος που τέθηκε. Παρ' όλο που η βάση επίλυσης είναι ίδια για αρχαίους και σύγχρονους γραφείς , ήταν απαραίτητο να διαχωριστεί για τις δυο περιπτώσεις ώστε να παρουσιαστούν πλήρως όλες οι λεπτομέρειες της μεθόδου για κάθε περίπτωση.

Τέλος, παρουσιάζουμε την εφαρμογή της μεθόδου και τα πλούσια και ενδιαφέροντα αποτελέσματα που αυτή προσέφερε. Τα αποτελέσματα αυτά επικύρωσαν το θεωρητικό μοντέλο που αναπτύχθηκε και προσέφερε εξαιρετικά συμπεράσματα γύρω από το αντικείμενο.



## **ABSTRACT**

The academic interest of the present thesis focuses on the automatic identification of writer of ancient and modern texts from his manuscripts and furthermore on dating ancient inscriptions.

At first, a description of the problem faced is presented to the reader, along with the importance and difficulties lying under the solving procedure of it.

Subsequently, a brief description of the method and a thorough presentation of the criteria and mathematical background used for goal achievement are given. Despite the fact that the base of the solution is the same for ancient and modern writers, it was necessary to discriminate the two cases in order to present all the details of each case.

Finally, we present the application of the method as well as the rich and interesting results we gathered. These results verified the theoretical model developed and offered excellent conclusions for the object studied.



# ΠΕΡΙΕΧΟΜΕΝΑ

<b>1.</b>	<b>Σκοπός της παρούσας διπλωματικής</b>	<b>7</b>
<b>2.</b>	<b>Εισαγωγή. Το πρόβλημα που αντιμετωπίστηκε</b>	<b>8</b>
2.1	Η σπουδαιότητα της αυτόματης αναγνώρισης γραφέα κειμένων	8
2.2	Υπάρχουσες μέθοδοι ταυτοποίησης γραφέα (state-of-the-art)	10
2.3	Η ειδική περίπτωση της αυτόματης ταυτοποίησης γραφέων αρχαίων κειμένων	11
2.3.1	Η σπουδαιότητα	11
2.3.2	Οι δυσκολίες	13
<b>3.</b>	<b>Μια νέα μέθοδος ταυτοποίησης γραφέα από τα κείμενα του.</b>	
	<b>Συνοπτική περιγραφή</b>	<b>15</b>
<b>4.</b>	<b>Κριτήρια ομοιότητας υλοποιήσεων γραμμάτων στη θέση βέλτιστης προσαρμογής</b>	<b>17</b>
4.1	Τα κριτήρια ομοιότητας αναφορικά με αρχαίες επιγραφές	17
4.1.1	Απεικονίσεις συμβολογραμμάτων αποτελούμενα από τέσσερα «κτυπήματα»	18
4.1.2	Απεικονίσεις συμβολογραμμάτων αποτελούμενα από τρία «κτυπήματα»	21
4.1.3	Απεικονίσεις συμβολογραμμάτων αποτελούμενα από δύο κτυπήματα	23
4.1.4	Απεικονίσεις συμβολογραμμάτων που παρουσιάζουν ιδιαίτερα χαρακτηριστικά	25
4.1.5	Ορισμός ενός κριτηρίου ομοιότητας στην θέση βέλτιστης προσαρμογής	26
4.2	Κριτήρια ομοιότητας για σύγχρονους γραφείς	34
<b>5.</b>	<b>Εφαρμογή των κριτηρίων στην εκτίμηση ομοιότητας γραμμάτων</b>	<b>43</b>
5.1	Εξαγωγή περιγραμμάτων των γραμμάτων	43
5.2	Η Διαδικασία εκτέλεσης	50
5.3	Τα αποτελέσματα	52
5.3.1	Από αρχαίους γραφείς	52
5.3.2	Από σύγχρονους γραφείς	59
<b>6.</b>	<b>Συμπεράσματα</b>	<b>65</b>
	<b>Βιβλιογραφία</b>	<b>66</b>



# **1. ΣΚΟΠΟΣ ΤΗΣ ΠΑΡΟΥΣΗΣ ΔΙΠΛΩΜΑΤΙΚΗΣ**

Ένα από τα πολύ μεγάλα ανοικτά προβλήματα της Αναγνώρισης Προτύπων, και της Ψηφιακής Επεξεργασίας Σημάτων και Εικόνων και κατ' επέκταση του κλάδου της Επιστήμης των Υπολογιστών, είναι η αυτόματη ταυτοποίηση του γραφέα από τα κείμενα του. Η επίλυση ενός τέτοιου προβλήματος, έχει μεγάλες πρακτικές εφαρμογές, όπως η επιβεβαίωση της γνησιότητας σημαντικών εγγράφων, της υπογραφής κάποιων υποκειμένων, της επίλυσης προβλημάτων που άπτονται του ποινικού και αστικού δικαίου, κλπ. Ειδικότερα, η αυτόματη αναγνώριση του γραφέα αρχαίων κειμένων, έχει τις εξής εξαιρετικά σημαντικές για την Αρχαιολογία και την Ιστορία δυνατότητες εφαρμογής: κατά την αρχαιότητα, αλλά και κατά τους Βυζαντινούς/Μεσαιωνικούς αλλά και μεταβυζαντινούς χρόνους, οι γραφείς κατά κανόνα δεν υπέγραφαν τα κείμενα τους ούτε έβαζαν ημερομηνία ή κάποια σχετικά προσδιοριστικά στοιχεία. Αυτό έχει σαν συνέπεια την γένεση αμφιβολιών, επιστημονικών διαφωνιών ενδεχομένως και αντεγκλήσεων αναφορικά με την ακριβή χρονολόγηση ενός αρχαίου κειμένου. Συνεπώς, η αυτόματη ταυτοποίηση του γραφέα ενός αρχαίου κειμένου με χρήση σύγχρονων μαθηματικών αλγορίθμων που υλοποιούνται στον υπολογιστή, μπορεί να οδηγήσει με ασφάλεια στη χρονολόγηση ενός αρχαίου κειμένου, άρα και στην επίλυση των σχετικών σημαντικών επιστημονικών προβλημάτων.

Ο σκοπός της παρούσης διπλωματικής εργασίας είναι η ενδελεχής κατανόηση μιας νέων μεθόδων ταυτοποίησης γραφέα ([1], [2], [3], [4], [5]) οι εφαρμογή αυτών στην αυτόματη ταυτοποίηση αρχαίων αλλά και σύγχρονων γραφέων ως και η στατιστική αξιολόγηση των σχετικών αποτελεσμάτων. Ειδικότερα, οι νέοι αυτοί αλγόριθμοι θα δοκιμαστούν στην αυτόματη ταυτοποίηση γραφέων βυζαντινών παπύρων, αλλά κυρίως και στην αυτόματη ταυτοποίηση σύγχρονων γραφέων από κείμενα τους που βρίσκονται σε διεθνή βάση δεδομένων αναφοράς.

## **2 ΕΙΣΑΓΩΓΗ.**

### **ΤΟ ΠΡΟΒΛΗΜΑ ΠΟΥ ΑΝΤΙΜΕΤΩΠΙΣΤΗΚΕ**

#### **2.1 Η σπουδαιότητα της αυτόματης αναγνώρισης γραφέα κειμένων**

Η αναγνώριση χειρογράφου μελετάται εδώ και αρκετές δεκαετίες από την ακαδημαϊκή κοινότητα με ιδιαίτερη προσήλωση και ζοηρότητα. Η βιομετρική αναγνώριση χειρογράφου, ή αλλιώς η αναγνώριση γραφέα, δεν πρέπει να συγχέεται με την οπτική αναγνώριση χαρακτήρων. Στην πρώτη περίπτωση ο σκοπός είναι να γίνει αναγνώριση και ταυτοποίηση του γραφέα ενός δοσμένου χειρόγραφου κειμένου, ανεξάρτητα από το περιεχόμενο, ενώ στην δεύτερη ο σκοπός είναι η αναγνώριση του κειμένου και ουσιαστικέ ή ψηφιοποίηση των χαρακτήρων, ανεξάρτητα του γραφέα.

Κάθε γραφέας μπορεί να χαρακτηριστεί από το δικό του τρόπο γραφής, από την αναπαραγωγή των λεπτομερών και του τρόπου που γράφει ασυναίσθητα. Για το λόγο αυτό, σε ορισμένες περιπτώσεις εμπειρογνωμοσύνης και ταυτοποίησης, τα δείγματα γραφής έχουν την ίδια αξία με εκείνη των δακτυλικών αποτυπωμάτων. Το πρόβλημα της αναγνώρισης γραφέα τίθεται συχνά στο Δικαστήριο των Ευρωπαϊκών Κοινοτήτων, όπου πρέπει κανείς να καταλήξει σε ένα συμπέρασμα σχετικά με τη γνησιότητα ενός εγγράφου (π.χ. διαθήκη). Προκύπτει, επίσης, σε τράπεζες για την επαλήθευση της υπογραφής, ή σε κάποια ιδρύματα που αναλύουν κείμενα παλαιότερων συγγραφέων, και ενδιαφέρονται για τη φύση των κειμένων αυτών, όπως για παράδειγμα τον προσδιορισμό των διαφόρων συγγραφέων που συμμετείχαν στην εκπόνηση ενός χειρογράφου ή ποιος έκανε διορθώσεις πάνω στο κείμενο.

Η Αυτόματη Αναγνώριση Γραφέα μπορεί να χωριστεί σε δύο βασικές κατηγορίες. Την στατική και την δυναμική. Στην στατική, γνωστή και ως “off-line” το χειρόγραφο κείμενο αρχικά ψηφιοποιείται μέσω ενός scanner ή κάμερας και το σύστημα που θέτουμε σε εφαρμογή αναγνωρίζει το κείμενο και ταυτοποιεί τον γραφέα μέσω των μετρήσεων που παίρνει από τους χαρακτήρες και των συγκρίσεων που κάνει. Στην δυναμική κατηγορία, γνωστή και ως “on-line”, γίνεται χρήση κάποιου ηλεκτρονικού tablet ή γραφίδας με λειτουργία PDA και το κείμενο συλλέγεται και επεξεργάζεται σε πραγματικό χρόνο.

Η Αυτόματη Αναγνώριση Γραφέα μέσω ενός ηλεκτρονικού υπολογιστικού συστήματος δίνει πολλές λύσεις στις μέρες μας και σε καμία περίπτωση δεν έρχεται να αντικαταστήσει τη δουλεία



ενός γραφολόγου, αλλά αντιθέτως είναι ένα χρήσιμο εργαλείο στα χέρια του που ελαχιστοποιεί τόσο την απόσταση όσο και το χρόνο. Τα παραδείγματα που δόθηκαν παραπάνω, η αναγνώριση μιας διαθήκης ή μιας υπογραφής, μπορούν να λυθούν πιο γρήγορα ή ακόμα να επιβεβαιωθούν με την βοήθεια ενός ηλεκτρονικού υπολογιστικού συστήματος. Επιπλέον κάποιες τραπεζικές ή ταχυδρομικές συναλλαγές που απαιτούν το γνήσιο της υπογραφής, μπορούν να γίνουν ευκολότερα με ένα δυναμικό σύστημα που αναφέραμε παραπάνω.

Τέλος υπάρχει τεράστιος όγκος τόσο αρχαίων, όσο και σύγχρονων κειμένων που δεν έχουν προλάβει να μελετήσουν και να ταυτοποιήσουν οι γραφολόγοι. Ένα αξιόπιστο σύστημα αυτόματης αναγνώρισης γραφέα θα ήταν ένα σημαντικό εργαλείο τόσο στην ταυτοποίηση του συγγραφέα και της γνησιότητας του κειμένου, αλλά παράλληλα και στην χρονολόγηση του χειρόγραφου.

## 2.2 Υπάρχουσες μέθοδοι ταυτοποίησης γραφέα (state-of-the-art).

Τα τελευταία χρόνια η έρευνα γύρω από του αντικείμενο της αυτόματης αναγνώρισης και ταυτοποίησης γραφέα γνωρίζει ιδιαίτερη άνθηση κυρίως όσον αφορά τα χειρόγραφα έγγραφα. Η on-line αναγνώριση γραφέα είναι ένας τομέας που αναπτύχθηκε σχετικά πρόσφατα και γι αυτό το λόγο υπάρχουν λίγες προσεγγίσεις πάνω σε αυτή. Μια μελέτη πάνω στην on-line αναγνώριση γραφέα είναι η [7] η οποία προτείνει μια μέθοδο δυναμικής αναγνώρισης γραφέα και χρησιμοποιεί τη σχέση μεταξύ στατικής και δυναμικής πληροφορίας ενός χειρόγραφου εγγράφου. Αντίθετα, το έργο που έχει παραχθεί πάνω στην off-line αναγνώριση γραφέα είναι πολύ περισσότερο. Συγκεκριμένα, αρκετοί ερευνητές εφαρμόζουν μορφολογική προσέγγιση [8], φίλτρα Gabor και προσέγγιση grayscale co-occurrence [3] ή texture identification. Στο έργο των [9],[10],[11] είναι βασισμένο στην εξαγωγή χαρακτηριστικών, ενώ στα έργα [12],[13],[14] εφαρμόζονται Hidden Markov Models. Στο έργο [15] παρουσιάζεται μια κυματομορφή βασισμένη στη γενικευμένη γκαουσιανή πυκνότητα (Generalized Gaussian Density (GGD)).

Από την άλλη πλευρά, τα έργα [16],[17] αντιμετωπίζουν τις διεργασίες αναγνώρισης και ταυτοποίησης γραφέα χρησιμοποιώντας τοπικά χαρακτηριστικά. Στο [18] προτείνεται μια προσέγγιση με μετασχηματισμό Fourier ενώ στο [19] εφαρμόζεται ένας μετασχηματισμός διχοτόμησης. Το [20] μετράει τη μοναδικότητα των χειρόγραφων χαρακτήρων μέσω μοντέλων αναγνώρισης και ταυτοποίησης. Η προσέγγιση textural και allograph πρωτοτύπου περιγράφεται στα [21],[22]. Ικανοποιητικά αποτελέσματα παρουσιάζουν δύο ακόμα μελέτες που χρησιμοποιούν ανάσυρση πληροφορίας με προσεγγίσεις ταιριάσματος allograph πρωτοτύπου η πρώτη [23] ή με διανομή διακριτού χαρακτήρα πρωτοτύπου η δεύτερη [24].

Πιο πρόσφατα, ερευνητές [25] χρησιμοποιούν προσέγγιση διανομής πρωτοτύπου συνεχούς χαρακτήρα με fuzzy c-means αλγόριθμους για να εκτιμήσουν την πιθανότητα ένας χαρακτήρας να έχει παραχθεί από ένα πρωτότυπο. Άλλοι ακαδημαϊκοί [26] χρησιμοποιούν ένα συνδυασμό από τοπικούς descriptors και learning τεχνικές ή από κατευθυντικά μορφολογικά χαρακτηριστικά [27]. Τέλος, στο [28] ειδική μέριμνα γίνεται σε ορισμένα χαρακτηριστικά γεωγραφικού τύπου, όπως skew, slant, pressure, thinning area κλπ για να κατατάξουν καλλιγραφικά χειρόγραφα έγγραφα ανάλογα με το γραφέα τους.

## **2.3 Η ειδική περίπτωση της αυτόματης ταυτοποίησης γραφέα αρχαίων κειμένων**

### **2.3.1 Η σπουδαιότητα της ταυτοποίησης γραφέα αρχαίων κειμένων και χειρογράφων**

Η έρευνα γύρω από την ταυτοποίηση γραφέα αρχαίων κειμένων αποτελεί ένα πολύ σημαντικό κεφάλαιο στην επιστήμη της ιστορίας και της αρχαιολογίας. Ιστορικές πηγές όπως διατηρημένα γραπτά έγγραφα σε λίθινες επιγραφές ή χειρόγραφα, προσφέρουν στους ιστορικούς ένα από τα πιο σημαντικά εργαλεία της εργασίας τους. Η μελέτη όμως μιας πηγής δεν μπορεί να γίνει σε κενό χρόνο. Είναι απολύτως απαραίτητο για τον ιστορικό επιστήμονα να γνωρίζει την ημερομηνία δημιουργίας της πηγής που μελετά καθώς αν αυτή δεν ερμηνευτεί μέσα στο ιστορικό πλαίσιο που γράφτηκε μπορεί είτε να οδηγήσει σε λάθος είτε σε ελλιπείς πληροφορίες. Όπως άλλωστε έγραψε πρόσφατα και ο μεγάλος ιστορικός Christian Habicht “σωστή χρήση των επιγραφών μπορεί να γίνει μόνο αν αυτές μπορούν να προσδιοριστούν χρονικά”. Ωστόσο, οι γραφείς δεν υπέγραφαν ποτέ τα κείμενα τους και δεν έβαζαν ημερομηνία.

Επίσης, το μεσαίωνα και μέχρι την εφεύρεση της τυπογραφίας, η λογοτεχνία του αρχαίου κόσμου διατηρούνταν και μεταφέρονταν μέσω παπύρων και περγαμινών. Η πρακτική αυτή παράγει παρόμοια προβλήματα με τις αρχαίες επιγραφές. Παραδείγματος χάριν, το έπος Ιλιάδα του Ομήρου διασώθηκε μέσα στους αιώνες σε μορφή χειρόγραφων τόμων. Αυτοί οι τόμοι, γράφτηκαν στην Κωνσταντινούπολη αλλά σε διάφορες χρονικές περιόδους χρειάστηκε να αλλάξουν τοποθεσία και να ταξιδέψουν σε βιβλιοθήκες της Ευρώπης. Κατά τη διάρκεια όλων αυτών των χρόνων, στο έργο προστέθηκαν διάφορες σημειώσεις και σχόλια στα περιθώρια του κειμένου καθώς και μεταξύ των γραμμών.

Η σημαντική αυτή έρευνα, που επιχειρούμε να μελετήσουμε και να αναπτύξουμε εδώ, δηλαδή του χρονικού προσδιορισμού αρχαίων επιγραφών μέσω της ταυτοποίησης γραφέα μέχρι και μερικά χρόνια πριν θα ήταν εκ των πραγμάτων ανέφικτη. Αυτό γιατί τα απαιτούμενα τεχνολογικά μέσα δεν είχαν αναπτυχθεί. Εργαλεία όπως η ψηφιακή φωτογραφία, φτηνές και μεγάλης χωρητικότητας αποθηκευτικές συσκευές καθώς επίσης δίκτυα υψηλής ταχύτητας είναι επιτεύγματα που αναπτύχθηκαν πρόσφατα και χωρίς αυτά η ψηφιακή σύγκριση κειμένων ήταν απαγορευτική.

Αντίθετα, στις μέρες μας η έρευνα αυτή, εκμεταλλευόμενη τα επιτεύγματα αυτά μπορεί να αναπτυχθεί και να παράξει πολύ σημαντικά αποτελέσματα όσον αφορά την ποσοτική και στατιστική ανάλυση σχετικά με τους διαφορετικούς γραφείς αρχαίων επιγραφών και τη φύση των κειμένων-πηγών.

Η μέθοδος που αναπτύσσεται εδώ, επιτυγχάνει έμμεσα το χρονικό προσδιορισμό των αρχαίων επιγραφών, εκμεταλλευόμενη την επαγωγική ιδιότητα. Συγκεκριμένα, αναφορικά με τις αρχαίες επιγραφές μπορούμε να βασιστούμε σε μια πολύ σημαντική ιδιότητα της περιόδου για τη μελέτη μας. Συγκεκριμένα οι γραφείς που σμίλευαν τις πλάκες στην αρχαιότητα ήταν επάγγελμα. Επομένως ο αριθμός των διαφορετικών “χεριών” είναι ιδιαίτερα περιορισμένος, τόσο ώστε να μπορούμε να υποθέσουμε ότι οι δεκάδες χιλιάδες πλάκες αποδίδονται σε μερικές δεκάδες “χέρια” . Αν λοιπόν είμαστε βέβαιοι για τη χρονολογία μιας πλάκας και μπορέσουμε να αποδώσουμε με σιγουριά μια άλλη πλάκα στον ίδιο χαρακτή, τότε μπορούμε να προσδιορίσουμε χρονικά κατά προσέγγιση τη δεύτερη πλάκα σύμφωνα με τη χρονική περίοδο επαγγελματικής δράσης του χαρακτή. Αξίζει να σημειωθεί εδώ, ότι η επαγγελματική περίοδος ενός χαρακτή κάλυπτε περίπου είκοσι με είκοσι πέντε χρόνια. Όπως γίνεται εύκολα αντιληπτό, η πιθανότητα να επιτύχει η μέθοδος αυξάνει όσο αυξάνει ο αριθμός των διαφορετικών επιγραφών που αποδίδονται σε αυτό το χέρι. Αυτό συμβαίνει γιατί αυξάνει και η πιθανότητα κάποια από αυτές τις επιγραφές να μπορέσει να προσδιοριστεί χρονικά.

Τα παραπάνω, ισχύουν και στην περίπτωση χειρογράφων. Όπως αναφέρθηκε ανωτέρω, το έργο του Ομήρου ταξίδεψε στην Ευρώπη και στους αιώνες και υπέστη πολλές προσθήκες σχολίων και σημειώσεων στο περιθώριο των πολυτελών αρχικών εκδόσεων. Συνεπώς, αν μπορέσουμε να προσδιορίσουμε με αυτόματα μέσα τα διαφορετικά “χέρια” που έγραψαν κείμενο , μπορούμε να αποκτήσουμε μια χρονολογική ταξινόμηση καθώς και επιπρόσθετες πολύ σημαντικές πληροφορίες.

Σύμφωνα με τα παραπάνω η αξία της ανάπτυξης ενός πακέτου μεθόδων και ενός σχετικού συστήματος πληροφορίας που πραγματοποιεί σωστή ταυτοποίηση γραφέα είναι απόλυτα απαραίτητη. Την αναγκαιότητα αυτή φιλοδοξεί να καλύψει η παρούσα μελέτη χρησιμοποιώντας μεθόδους από τα πεδία των Μαθηματικών, Αναγνώρισης προτύπου , Επεξεργασίας Εικόνας καθώς και Επιστήμης Υπολογιστών εν γένει.

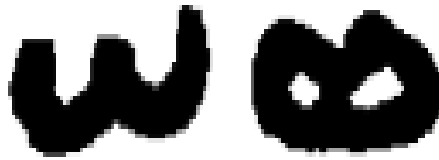
### 2.3.2 Οι δυσκολίες

Όπως σε κάθε καινοτόμα μεθοδολογία, έτσι και εδώ οι δυσκολίες δεν έλειψαν. Ο βαθμός δυσκολίας μάλιστα στη δικιά μας περίπτωση ήταν αυξημένες καθώς, η υποφαινόμενη μέθοδος αυτόματης ταυτοποίησης γραφέα παρουσιάζεται πρώτη φορά στην επιστημονική κοινότητα. Αυτό έχει ως αποτέλεσμα να μην υπάρχει σχετική βιβλιογραφία, να μην υπάρχουν κείμενα που μπορούν να χρησιμοποιηθούν ως πρότυπα και να μην υπάρχει μια βάση δεδομένων για συγκριτική και υποβοηθητική χρήση. Επίσης, εφόσον η σκοπιά από την οποία εξετάζουμε το αντικείμενο είναι αυτή το μηχανικού υπολογιστών και του μαθηματικού, είναι φυσιολογικό να απουσιάζει η ιστορική γνώση για τον αριθμό των διαφορετικών χειρών που είχαν γράψει τα σχετικά χειρόγραφα, μια επικουρική πληροφορία που θα ήταν πολύτιμη.

Εκτός από τις παραπάνω δυσκολίες, αντιμετωπίστηκαν και προβλήματα δομικής υφής. Για παράδειγμα, πολλές επιγραφές και βυζαντινοί κώδικες όπως είναι φυσιολογικό έχουν υποστεί σοβαρή φθορά στο πέρασμα των χρόνων. Επιπρόσθετα, μια ιδιομορφία που έχει να κάνει με τον ανθρώπινο γραφικό χαρακτήρα πρόσθεσε δυσκολίες στην εργασία μας. Πολλές φορές, δύο αλφαβητικά σύμβολα που έχουν γραφεί από τον ίδιο γραφέα παρουσιάζουν μικρότερη ομοιότητα απ' ό,τι δύο αλφαβητικά σύμβολα διαφορετικών γραφέων. Αυτό έχει ως αποτέλεσμα να εισάγεται ένας παράγοντας αβεβαιότητας στη μεθοδολογία ο οποίος αν και εξαλείφεται από τη στατιστική, χρήζει ιδιαίτερης προσοχής.

Ένα άλλο σημείο που έπρεπε να αντιμετωπιστεί, είναι ότι για πολλούς γραφείς υπάρχει ουσιώδης μεταβλητότητα στον τρόπο με τον οποίο ένας γραφέας σχηματίζει ένα αλφαβητικό σύμβολο σε ένα κείμενο. Για παράδειγμα, στο ίδιο έγγραφο και για τον ίδιο γραφέα ένα πλήθος υλοποιήσεων του “ω” μπορεί να είναι κλειστό, ενώ ένα άλλο πλήθος μπορεί να είναι ανοικτό όπως φαίνεται στην εικόνα 2.1. Η ίδια παρατήρηση ισχύει και για το γράμμα “ε” και άλλα συμβολογράμματα. Ένα χαρακτηριστικό παράδειγμα είναι το αλφαβητικό σύμβολο κάπα, το οποίο σε άλλες υλοποιήσεις μοιάζει με το λατινικό “u” ενώ σε άλλες μοιάζει με το κλασικό “κ”. Σε πολλές περιπτώσεις η μορφή του συμβολογράμματος που επιλέγεται από το γραφέα έχει να κάνει με τη θέση που εμφανίζεται αυτό μέσα σε μια λέξη (μέσα στη λέξη ή στο τέλος), ποιο γράμμα προηγείται ή ακολουθεί καθώς επίσης την κούραση του γραφέα τη στιγμή της υλοποίησης. Ένα άλλο χαρακτηριστικό παράδειγμα είναι το κεφαλαίο άλφα στις αρχαίες ελληνικές επιγραφές.

Πολλοί γραφείς δε συνδέουν τη μεσαία οριζόντια γραμμή του “Α” με το αριστερό πόδι , άλλοι με το δεξί πόδι και άλλες φορές δεν τη συνδέουν με κανένα.



Εικόνα 2.1

Τέλος, ακόμα και αν η μορφή του συμβολογράμματος είναι σταθερή, πολύ σημαντικό ρόλο παίζει ο θόρυβος του περιγράμματος. Το φαινόμενο του θορύβου, που εξαρτάται σε μεγάλο βαθμό από την κούραση, τη διάθεση και την ηλικία του γραφέα, τα μέσα που αυτός χρησιμοποιεί όπως την ποιότητα του μαρμάρου ή του υποθέματος γραφής, τη συγκεκριμένη μορφή του οργάνου γραφής (μολύβι, σμίλη κλπ) και από τη φθορά του κειμένου, εμφανίζεται πολύ συχνότερα στις αρχαίες επιγραφές (Εικόνα 2.2) χωρίς αυτό να απουσιάζει για παράδειγμα από βυζαντινούς κώδικες όπως φαίνεται.



Εικόνα 2.2

### **3. ΜΙΑ ΝΕΑ ΜΕΘΟΔΟΣ ΤΑΥΤΟΠΟΙΗΣΗΣ ΓΡΑΦΕΑ**

#### **ΑΠΟ ΤΑ ΚΕΙΜΕΝΑ ΤΟΥ. ΣΥΝΟΠΤΙΚΗ ΠΕΡΙΓΡΑΦΗ**

1. Σε κάθε έγγραφο χωριστά, έχουν εφαρμοστεί αυτόματες μέθοδοι τμηματοποίησης των εικόνων, έτσι ώστε να εξαχθεί η υλοποίηση κάθε συμβόλου της αλφαβήτου από το φόντο της εικόνας. Κάθε τέτοια υλοποίηση γράμματος που εξάγεται ενσωματώνεται σε ένα ορθογώνιο πλαίσιο. Το σώμα της υλοποίησης, κατά κανόνα, είναι αρκετά θορυβώδες. Συνεπώς, εφαρμόζονται κατάλληλα μορφολογικά φίλτρα καθώς και φίλτρα που μειώνουν το θόρυβο, προκειμένου να καταστήσουν την υλοποίηση του συμβόλου της αλφαβήτου ομοιογενές σώμα. Έτσι, τελικά, κάθε υλοποίηση είναι ενσωματωμένη σε ένα ορθογώνιο πλαίσιο, με το σώμα του να είναι απλό μαύρο και το φόντο του καθαρό λευκό.

2. Προκειμένου η επακόλουθη ανάλυση να είναι εφαρμόσιμη, το περίγραμμα του κάθε σώματος μιας αυθαίρετης υλοποίησης ενός γράμματος, πρέπει να έχει μια σαφή και καλά καθορισμένη μορφή. Για αυτό το σκοπό έχουν σχεδιαστεί καινοτόμοι αλγόριθμοι, οι οποίοι εκτελούν τις κατάλληλες ενέργειες φιλτραρίσματος.

3. Σκεφτείτε ένα περίγραμμα  $e$  από μια αυθαίρετη υλοποίηση ενός συμβόλου της αλφαβήτου, που εμφανίζεται σε ένα συγκεκριμένο έγγραφο. Αρχικά, χωρίζουμε το  $e$  σε συνεχόμενες αλυσίδες από pixels, ελαφρώς επικαλυπτόμενες. Στη συνέχεια, προσεγγίζουμε κάθε τέτοια αλυσίδα των pixels με κατάλληλες πολωνυμικές συναρτήσεις για τις  $x$  και  $y$  συντεταγμένες χωριστά, με ανεξάρτητη μεταβλητή το τόξο της αλυσίδας μήκους  $S$ . Έχουμε υποθέσει και επιβεβαιώσαμε ότι η καμπυλότητα αυτών των πολωνυμικών λειτουργιών, υπολογίζεται σε σημεία κοντά στο κέντρο κάθε αλυσίδας, προσφέροντας μια ιδιαίτερα καλή εκτίμηση της τάσης της καμπυλότητας του αρχικού περιγράμματος του γράμματος. Για να επιτευχθεί μια καλή εκτίμηση ανάλογης ποιότητας αυτής της τάσης καμπυλότητας, κοντά στο τέλος των αλυσίδων, έχουμε εφαρμόσει κατάλληλο ψαλίδισμα των ορίων της αλυσίδας, καθώς επίσης και έναν κατάλληλα σταθμισμένο υπολογισμό μέσης τιμής των τιμών καμπυλότητας των τελικών σημείων. Μέχρι το τέλος αυτής της διαδικασίας, κάθε περίγραμμα  $e$  μιας αυθαίρετης υλοποίησης συμβόλου της αλφαβήτου, συνδέεται αποκλειστικά με μία ακολουθία τιμών καμπυλότητας, η οποία αντιπροσωπεύει την τάση της καμπυλότητας σε κάθε pixel του  $e$ , με έναν πολύ ικανοποιητικό τρόπο.

4. Ας υποθέσουμε, τώρα, ότι θεωρούμε δύο περιγράμματα  $C1$ ,  $C2$  δύο διαφορετικών υλοποιήσεων του ίδιου συμβόλου της αλφαβήτου, το οποίο προκαλεί δύο διακριτές αλληλουχίες των καμπυλοτήτων  $K1$ ,  $K2$ , όπως αυτά που περιγράφονται στο ανωτέρω Βήμα 3. Έχουμε αναπτύξει και εφαρμόσει μια μαθηματική προσέγγιση και ένα αντίστοιχο αλγόριθμο, τα οποία επιτυγχάνουν άμεσο βέλτιστο ταίριασμα των  $K1$  και  $K2$ . Στη συνέχεια, τα περιγράμματα  $C1$  και  $C2$  ταιριάζουν, επίσης, βέλτιστα με τη βοήθεια ενός νέου λήμματος.
5. Η εφαρμογή των προηγούμενων βημάτων προσφέρει μια ιδιαίτερα καλή θέση που ταιριάζουν οι καμπύλες  $C1$  και  $C2$ . Παρόλα αυτά, σε αυτό το στάδιο, εισάγουμε ένα ακόμα ουσιαστικά διαφορετικό κριτήριο, το οποίο είναι ιδιαίτερα κατάλληλο για στατιστικές εκτιμήσεις. Μπορεί κανείς να αναφέρει το κριτήριο αυτό με έναν γενικό τρόπο, ως την περιοχή της τομής των δύο αντιστοιχισμένων γραμμών ανά τομέα της ένωσής τους.
6. Χρησιμοποιώντας τα κριτήρια που ορίζονται στα προηγούμενα βήματα, έχουμε δηλώσει και δοκιμάσει στατιστικές υποθέσεις, οι οποίες έδειξαν τελικά τον διαφορετικό αριθμό των συγγραφέων που έγραψαν όλα τα διαθέσιμα έγγραφα. Την ίδια στιγμή, ένα "αντιπροσωπευτικό" έγγραφο συνδέεται αποκλειστικά με κάθε συγγραφέα.
7. Τα υπόλοιπα έγγραφα που έχουν ταξινομηθεί στους αντίστοιχους γραφείς με βάση μια καινοτόμα εκτιμήτρια μέγιστης πιθανοφάνειας.



## **4. ΚΡΙΤΗΡΙΑ ΟΜΟΙΟΤΗΤΑΣ ΥΛΟΠΟΙΗΣΕΩΝ**

### **ΓΡΑΜΜΑΤΩΝ ΣΤΗ ΘΕΣΗ ΒΕΛΤΙΣΤΗΣ**

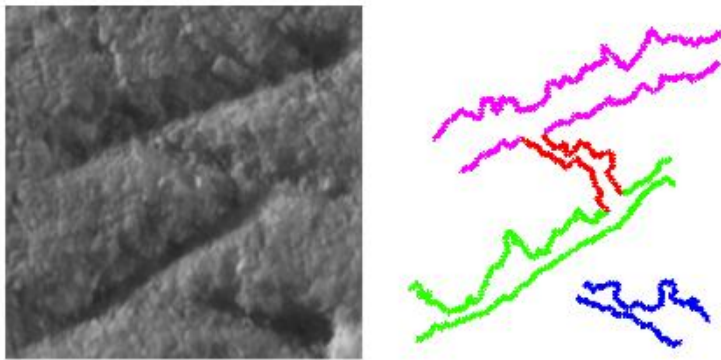
### **ΠΡΟΣΑΡΜΟΓΗΣ.**

#### **4.1 Τα κριτήρια ομοιότητας αναφορικά με αρχαίες επιγραφές .**

Στην παρούσα διπλωματική μας έχουν δοθεί όλα τα προς σύγκριση ζεύγη γραμμάτων στη θέση βέλτιστης προσαρμογής όσον αφορά είτε την καμπυλότητα των περιγραμμάτων είτε την απόσταση τους στον ευκλείδειο χώρο. Ο σκοπός της παρούσας εργασίας ήταν επομένως να ελεγχθεί η ομοιότητα ή μη των υλοποιήσεων των διαφόρων συμβόλων της αλφαβήτου που εμφανίζονται τόσο σε αρχαία όσο και σύγχρονα κείμενα . Εν συνεχεία , στα αποτελέσματα αυτών των συγκρίσεων εφαρμόζεται η μεθοδολογία ταυτοποίησης γραφέα που έχει αναπτυχθεί στα διδακτορικό Π.Ρ.) [1]. Η εφαρμογή όμως αυτή ξέφευγε εξ' αρχής του σκοπού αυτής της διπλωματικής εργασίας.

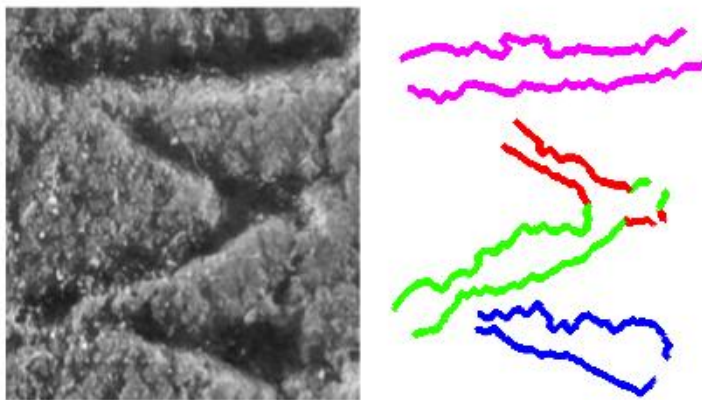
Για να εφαρμόσουμε κριτήρια ομοιότητας σε αρχαίες επιγραφές αρχικά θα κάνουμε κάποιες παραδοχές αναφορικά με τη μορφή των συμβολογραμμάτων. Καταρχήν, θεωρούμε ότι κάθε σύμβολο της αλφαβήτου αποτελείται κατά κανόνα από συγκεκριμένο αριθμό “κτυπημάτων”. Με τον όρο “κτύπημα” περιγράφουμε κάθε σκέλος ενός οποιουδήποτε γράμματος το οποίο κατά κανόνα δημιουργήθηκε με ένα κτύπημα της σμίλης επί του μαρμάρου. Για παράδειγμα, θεωρούμε ότι τα συμβολογράμματα Σ,Μ,Ε αποτελούνται από τέσσερα κτυπήματα , τα συμβολογράμματα Α,Ν από τρία και τα Τ,Λ από δύο. Για τα γράμματα που περιέχουν μη ευθύγραμμο τμήματα όπως τα Ρ,Ο,Β,Ω,Φ λαμβάνουμε ειδική μέριμνα όσον αφορά το πλήθος των “κτυπημάτων” που τα δημιούργησαν. Για παράδειγμα θεωρούμε ότι το γράμμα Ρ αποτελείται από δύο κτυπήματα, ένα ευθύγραμμο σκέλος και μια καπύλη. Η παραδοχή που κάνουμε για τα καμπύλα γράμματα, εμπεριέχει μιας μορφής κατάχρηση καθώς ο λιθοξόος πιθανώς να χάραξε το καμπύλο τμήμα με περισσότερα από ένα “κτυπήματα”. Στη συνέχεια παραθέτουμε μερικές εικόνες υλοποιήσεων συμβολογραμμάτων με τα περιγράμματα τους όπου για λόγους καλής εποπτείας χρησιμοποιούμε διαφορετικά χρώματα για τα αντίστοιχα “κτυπήματα”.

### 4.1.1 Απεικονίσεις συμβολογραμμάτων αποτελούμενα από τέσσερα «κτυπήματα».



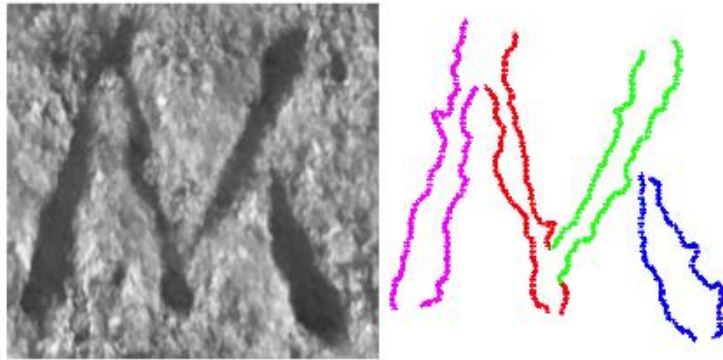
Εικόνα 4.1

Ψηφιακή εικόνα ενός γράμματος Σ με την τελική μορφή του περιγράμματος αυτού στο x-y επίπεδο. Το πρώτο “κτύπημα” παρουσιάζεται με χρώμα μωβ, το δεύτερο με χρώμα κόκκινο, το τρίτο με χρώμα πράσινο και το τέταρτο με χρώμα μπλε.



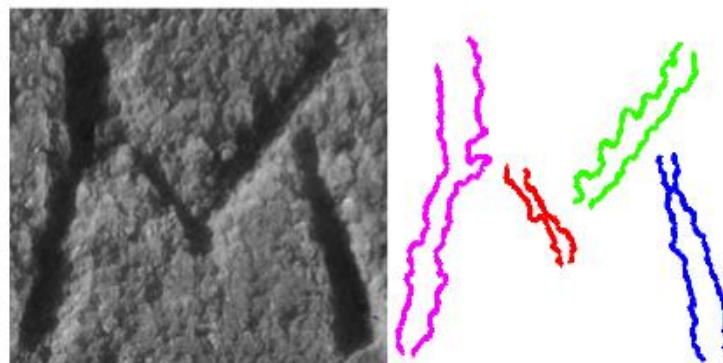
Εικόνα 4.2

Ομοίως με την εικόνα 4.1 παρουσιάζεται ένα δεύτερο παράδειγμα ψηφιακής εικόνας του γράμματος Σ με την τελική μορφή του περιγράμματος αυτού στο x-y επίπεδο ενώ τηρείται η αντιστοιχία των χρωμάτων.



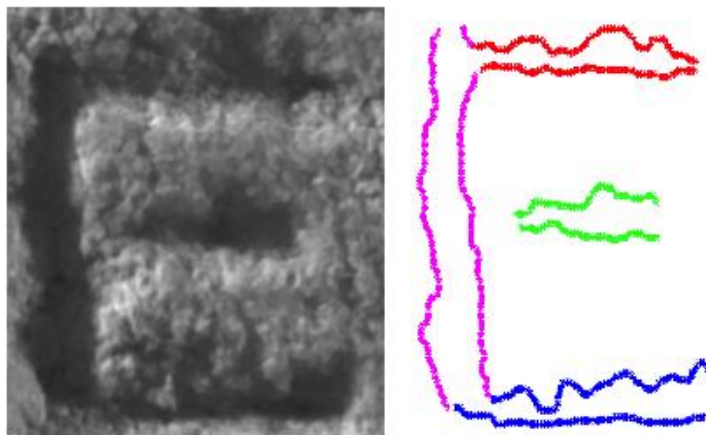
Εικόνα 4.3

Ψηφιακή εικόνα ενός γράμματος Μ με την τελική μορφή του περιγράμματος αυτού στο x-y επίπεδο. Το πρώτο “κτύπημα” παρουσιάζεται με χρώμα μωβ, το δεύτερο με χρώμα κόκκινο, το τρίτο με χρώμα πράσινο και το τέταρτο με χρώμα μπλε.



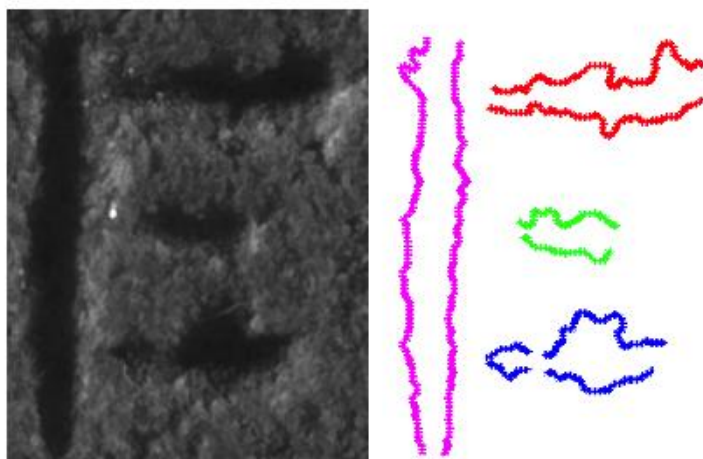
Εικόνα 4.4

Ομοίως με την εικόνα 4.3 παρουσιάζεται ένα δεύτερο παράδειγμα ψηφιακής εικόνας του γράμματος Μ με την τελική μορφή του περιγράμματος αυτού στο x-y επίπεδο ενώ τηρείται η αντιστοιχία των χρωμάτων.



Εικόνα 4.5

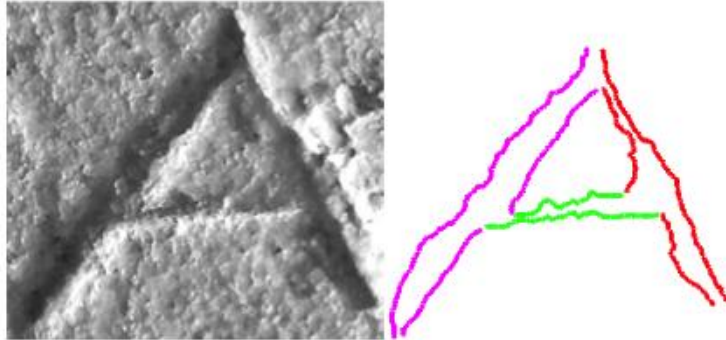
Ψηφιακή εικόνα ενός γράμματος Ε με την τελική μορφή του περιγράμματος αυτού στο x-y επίπεδο. Το πρώτο “κτύπημα” παρουσιάζεται με χρώμα μωβ, το δεύτερο με χρώμα κόκκινο, το τρίτο με χρώμα πράσινο και το τέταρτο με χρώμα μπλε.



Εικόνα 4.6

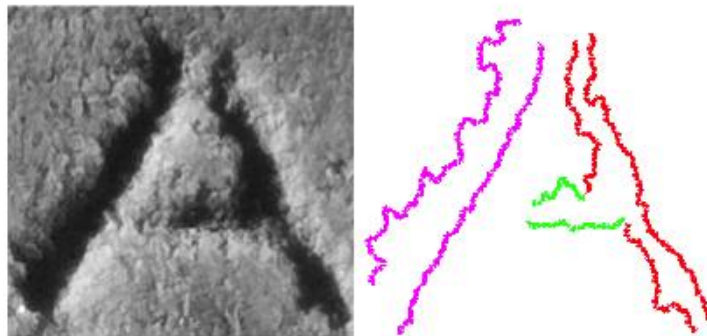
Ομοίως με την εικόνα 4.5 παρουσιάζεται ένα δεύτερο παράδειγμα ψηφιακής εικόνας του γράμματος Ε με την τελική μορφή του περιγράμματος αυτού στο x-y επίπεδο ενώ τηρείται η αντιστοιχία των χρωμάτων.

#### 4.1.2 Απεικονίσεις συμβολογραμμάτων αποτελούμενα από τρία «κτυπήματα».



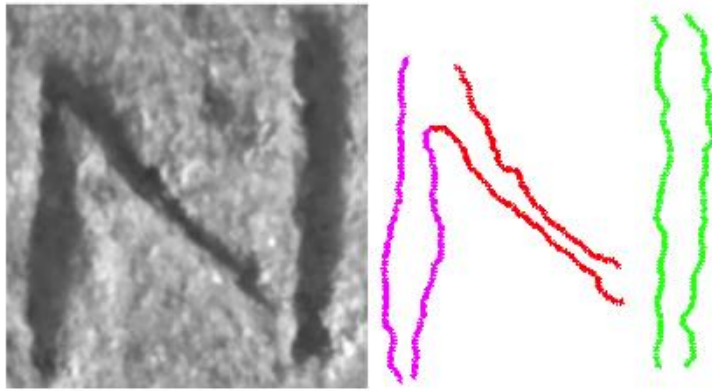
Εικόνα 4.7

Ψηφιακή εικόνα ενός γράμματος Α με την τελική μορφή του περιγράμματος αυτού στο x-y επίπεδο. Το πρώτο “κτύπημα” παρουσιάζεται με χρώμα μωβ, το δεύτερο με χρώμα κόκκινο και το τρίτο με χρώμα πράσινο.



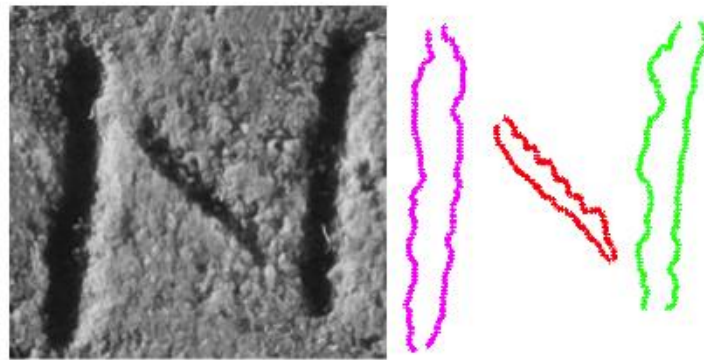
Εικόνα 4.8

Ομοίως με την εικόνα 4.7 παρουσιάζεται ένα δεύτερο παράδειγμα ψηφιακής εικόνας του γράμματος Α με την τελική μορφή του περιγράμματος αυτού στο x-y επίπεδο ενώ τηρείται η αντιστοιχία των χρωμάτων.



Εικόνα 4.9

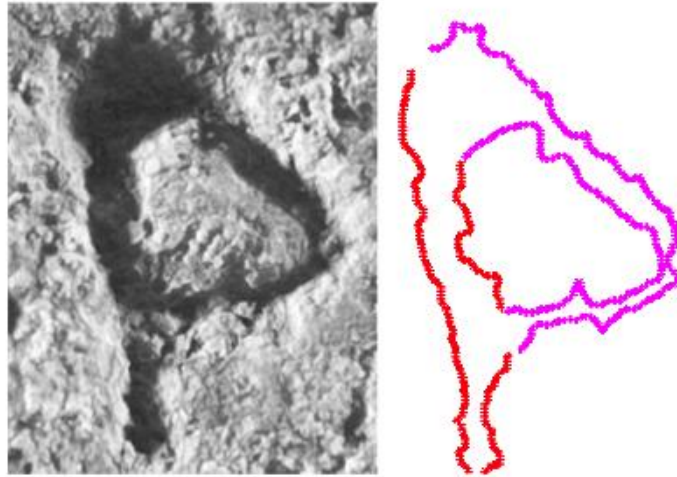
Ψηφιακή εικόνα ενός γράμματος N με την τελική μορφή του περιγράμματος αυτού στο x-y επίπεδο. Το πρώτο “κτύπημα” παρουσιάζεται με χρώμα μωβ, το δεύτερο με χρώμα κόκκινο και το τρίτο με χρώμα πράσινο.



Εικόνα 4.10

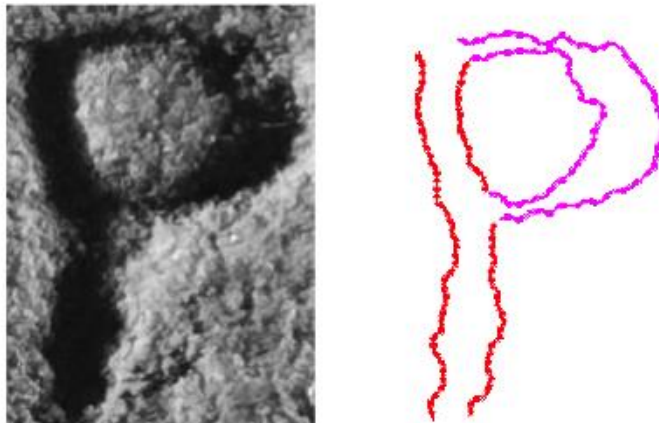
Ομοίως με την εικόνα 4.9 παρουσιάζεται ένα δεύτερο παράδειγμα ψηφιακής εικόνας του γράμματος N με την τελική μορφή του περιγράμματος αυτού στο x-y επίπεδο ενώ τηρείται η αντιστοιχία των χρωμάτων.

### 4.1.3 Απεικονίσεις συμβολογραμμάτων που αποτελούνται από δύο «κτυπήματα».



Εικόνα 4.11

Ψηφιακή εικόνα ενός γράμματος P με την τελική μορφή του περιγράμματος αυτού στο x-y επίπεδο. Το πρώτο “κτύπημα” παρουσιάζεται με χρώμα μωβ ενώ το δεύτερο με χρώμα κόκκινο

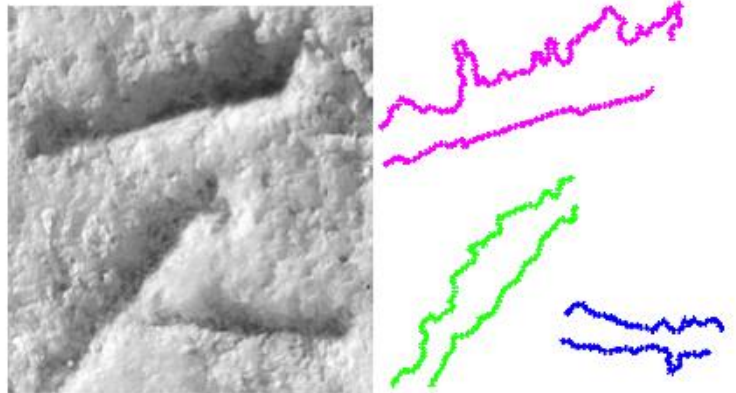


Εικόνα 4.12

Ομοίως με την εικόνα 4.9 παρουσιάζεται ένα δεύτερο παράδειγμα ψηφιακής εικόνας του γράμματος P με την τελική μορφή του περιγράμματος αυτού στο x-y επίπεδο ενώ τηρείται η αντιστοιχία των χρωμάτων.

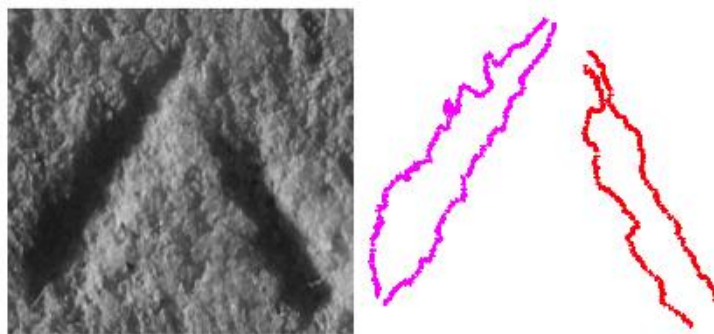


#### 4.1.4 Απεικονίσεις υλοποιήσεων συμβολογραμμάτων που παρουσιάζουν ιδιαίτερα χαρακτηριστικά.



Εικόνα 4.13

Ψηφιακή εικόνα ενός γράμματος  $\Sigma$  της “χειρός τρία” με την τελική μορφή του περιγράμματος αυτού στο x-y επίπεδο. Το πρώτο “κτύπημα” παρουσιάζεται με χρώμα μωβ, το δεύτερο “κτύπημα” δεν έχει υλοποιηθεί, το τρίτο με χρώμα πράσινο και το τέταρτο με χρώμα μπλε.



Εικόνα 4.14

Ψηφιακή εικόνα ενός γράμματος  $\text{A}$  της “χειρός τέσσερα” με την τελική μορφή του περιγράμματος αυτού στο x-y επίπεδο. Το πρώτο “κτύπημα” παρουσιάζεται με χρώμα μωβ, το δεύτερο “κτύπημα” με χρώμα κόκκινο και το τρίτο χρώμα δεν έχει υλοποιηθεί.



Ένα συχνό φαινόμενο που παρατηρήσαμε στους αρχαίους γραφείς είναι ότι πολλές φορές παρέλειπαν κάποιο σκέλος ενός συμβολογράμματος. Στις ανωτέρω εικόνες παρουσιάζονται δύο υλοποιήσεις συμβολογραμμάτων που δείχνουν ακριβώς αυτή τη συμπεριφορά. Ο λιθοξόος στον οποίο θα αναφερόμαστε ως “χέρι τρία” έχει την τάση να παραλείπει συχνά το δεύτερο σκέλος του συμβολογράμματος  $\Sigma$ , ενώ ο λιθοξόος “χέρι τέσσερα” συχνά δε χτυπούσε τη μεσαία κεραία του  $\Lambda$ . Η διαφοροποίηση του  $\Lambda$  από το  $\Lambda$  γίνεται από τα συμφραζόμενα.

## 4.1.5 Ορισμός ενός κριτηρίου ομοιότητας στη θέση βέλτιστης προσαρμογής.

Σύμφωνα με τις παραδοχές που κάναμε ανωτέρω όσον αφορά τη μορφή των συμβολογραμμάτων μπορούμε τώρα να μοντελοποιήσουμε μαθηματικά και να παράγουμε ένα κριτήριο ομοιότητας των συμβολογραμμάτων στη θέση βέλτιστης προσαρμογής. Έστω  $Y^{\Pi}$  και  $Y^T$  δύο υλοποιήσεις του ίδιου συμβολογράμματος που εμφανίζονται σε δύο διαφορετικές επιγραφές. Ας υποθεθεί ότι κάθε μια υλοποίηση αποτελείται από τέσσερα “κτυπήματα” και συγκεκριμένα η  $Y^{\Pi}$  από τα  $S_1^{\Pi}, S_2^{\Pi}, S_3^{\Pi}, S_4^{\Pi}$  ενώ η  $Y^T$  από τα  $S_1^T, S_2^T, S_3^T, S_4^T$ . Όσον αφορά την ειδική περίπτωση που εξετάσαμε στην παράγραφο 4.1.4 στο εξής θα θεωρούμε από αυστηρά μαθηματικής απόψεως ότι αν κάποιο “κτύπημα”  $S_i$  δεν υλοποιείται τότε αυτό είναι το κενό σύνολο.

Ας θεωρήσουμε τώρα ότι οι υλοποιήσεις  $Y^{\Pi}$  και  $Y^T$  είναι βέλτιστα προσαρμοσμένες μεταξύ τους με την  $Y^{\Pi}$  να παίζει το ρόλο της προτύπου υλοποίησης και η και  $Y^T$  της τρεχούσης<sup>1</sup>. Στη συνέχεια θεωρούμε δύο αντίστοιχα “κτυπήματα”  $S_1^{\Pi}$  και  $S_1^T$  σε αυτή τη θέση βέλτιστης προσαρμογής.

Επιπλέον, θεωρούμε τέσσερα κρίσιμα σημεία του υπό εξέταση “κτυπήματος”,  $A_1, A_2, B_1, B_2$  τα οποία δημιουργούν μία περιορισμένη εκδοχή αυτού. Τα  $A_1, A_2$  ορίζουν τη μια πλευρά του σκέλους και τα  $B_1, B_2$  την άλλη πλευρά αυτού. Ενώνοντας τα  $A_1, B_1$  αρχικά και τα  $A_2, B_2$  έπειτα σχηματίζουμε την κλειστή καμπύλη  $A_1 A_2 B_2 B_1$  όπως φαίνεται στην παρακάτω εικόνα (4.15).

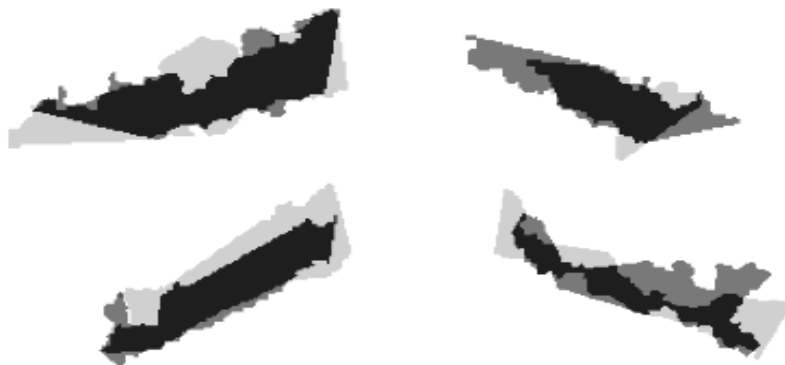
---

<sup>1</sup> Για τον προσδιορισμό της θέσης βέλτιστης προσαρμογής χρησιμοποιείται η ακόλουθη διαδικασία. Αρχικά τα περιγράμματα των δύο υλοποιήσεων τοποθετούνται σε κοινό πλαίσιο (καμβά) με τα κέντρα βάρους τους να συμπίπτουν. Το περίγραμμα της μιας υλοποίησης διατηρείται σταθερό και το θεωρούμε καμπύλη αναφοράς ενώ το περίγραμμα της έταιρης υλοποίησης περιστρέφεται, υφίσταται ομοιοθεσία και μετατοπίζεται παράλληλα. Στη συνέχεια ορίζεται το μέτρο της απόστασης μεταξύ των δύο περιγραμμάτων (του περιγράμματος αναφοράς και του προσαρμοσμένου) και επιλέγεται εκείνος ο μετασχηματισμός που προσφέρει τη μικρότερη απόσταση μεταξύ των δύο καμπυλών ή ισοδύναμα το μικρότερο λάθος προσαρμογής. Τέλος, τα δύο περιγράμματα συγκρίνονται και με μια επιπλέον μέθοδο δηλαδή με χρήση ενός ακόμα κατάλληλα επιλεγμένου κριτηρίου όσον αφορά τη μεταξύ τους προσαρμογή. Όπως αναφέρθηκε στην εισαγωγή αυτού του κεφαλαίου για το σκοπό αυτής της διπλωματικής εργασίας όλα τα προς σύγκριση ζεύγη γραμμάτων μας έχουν δοθεί στη θέση βέλτιστης προσαρμογής. Για το λόγο αυτό, σε αυτό το σημείο δε θα προχωρήσουμε στη μαθηματική παρουσίαση της διαδικασίας προσδιορισμού της θέσης βέλτιστης προσαρμογής, η οποία περιγράφεται αναλυτικά στο [1].



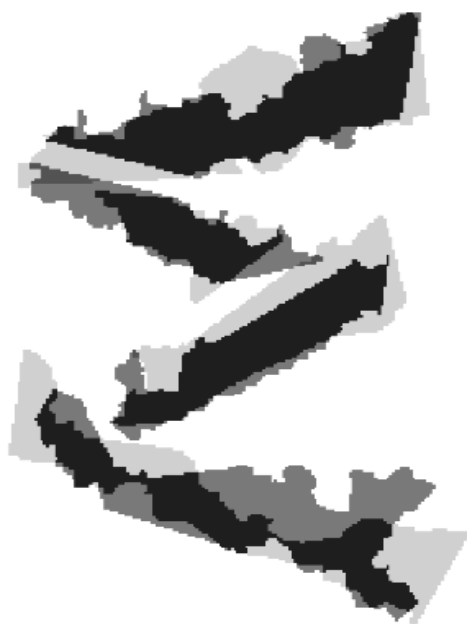
Εικόνα 4.15

Ειδική αντιμετώπιση χρήζει η περίπτωση όπου έχουμε ένα σκέλος το οποίο αποτελείται από νησίδες. Τότε εκτός των  $A_1 A_2$  και  $B_1 B_2$  διασυνδέουμε με ευθύγραμμα τμήματα διαδοχικά κρίσιμα σημεία που γεφυρώνουν δύο διαδοχικές αλυσίδες επί της ίδιας πλευράς του σκέλους έτσι ώστε να δημιουργηθεί μία κλειστή καμπύλη που περιβάλλει το σκέλος. Με χρήση της διαδικασίας που περιγράψαμε στα “κτυπήματα”  $S_1^H$  και  $S_1^T$  και λαμβάνουμε τις βέλτιστα προσαρμοσμένες κλειστές καμπύλες  $A_1^H A_2^H B_2^H B_1^H$  και  $A_1^T A_2^T B_2^T B_1^T$  όπως αυτές φαίνονται στην εικόνα 4.16.



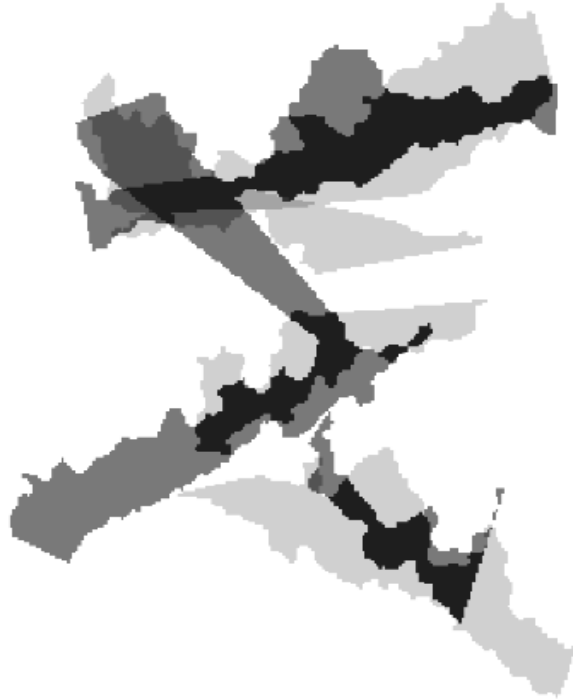
Εικόνα 4.16

Βέλτιστα προσαρμοσμένα “κτυπήματα” δύο υλοποιήσεων του συμβολογράμματος  $\Sigma$  της ίδιας χειρός. Η περιοχή που είναι χρωματισμένη με τον ανοικτότερο τόνο είναι της τρέχουσας υλοποίησης, αυτή με τον αμέσως πιο σκούρο της πρότυπης και η πιο σκούρα όλων η περιοχή της τομής των δύο.



Εικόνα 4.17

Συνολική παρουσίαση των βέλτιστα ταιριασμένων “κτυπημάτων” της εικόνας 4.16, των δύο υλοποιήσεων του συμβολογράμματος  $\Sigma$  που προέρχονται από το **ίδιο χέρι**.



Εικόνα 4.18

Συνολική παρουσίαση των βέλτιστα ταιριασμένων “κτυπημάτων” της εικόνας 4.16, των δύο υλοποιήσεων του συμβολογράμματος  $\Sigma$  που προέρχονται από **διαφορετικό χέρι**. Η περιοχή που είναι χρωματισμένη με τον ανοικτότερο τόνο είναι της τρέχουσας υλοποίησης, αυτή με τον αμέσως πιο σκούρο της πρότυπης και η πιο σκούρα όλων η περιοχή της τομής των δύο.

Κάθε κλειστή καμπύλη χωρίζει κατά τα γνωστά το επίπεδο σε δύο χωρία, ένα εσωτερικό  $E_1^{\Pi}$  ή  $E_1^T$  και ένα εξωτερικό  $X_1^{\Pi}$  ή  $X_1^T$ . Σύμφωνα με αυτή την παρατήρηση μπορούμε να ορίσουμε μία ποσότητα  $\delta E_1$  ίση με το πηλίκο δύο εμβαδών:

$$\delta E = \frac{\text{εμβαδόντης } E_1^{\Pi} \cap E_1^T}{\text{εμβαδόντης } E_1^{\Pi} \cup E_1^T} = \frac{EMBE_1^{\Pi} \cap E_1^T}{EMBE_1^{\Pi} \cup E_1^T}$$

Με τον ίδιο τρόπο ορίζουμε  $\delta E_k$  και για τα άλλα “κτύπηματα” :

$$\delta E_k = \frac{EMBE_k^{\Pi} \cap E_k^T}{EMBE_k^{\Pi} \cup E_k^T} \quad k = 2,3,4 \quad (4.1)$$

Θεωρούμε ότι όταν κάποιο “κτύπημα”  $k = 1,2,3,4$  απουσιάζει η αντίστοιχη ποσότητα  $E_k$  είναι ίση με μηδέν.

Εύκολα συμπεραίνουμε ότι η τιμή της ποσότητας  $\delta E_k$  είναι άμεσα συνδεδεμένη με την ομοιότητα των “κτυπημάτων”. Ιδανικά “κτυπήματα” που προέρχονται από τον ίδιο γραφέα και αυτός είχε την ικανότητα να αναπαραγάγει τέλεια το συμβολόγραμμα που είχε υπόψην, δίνουν τιμή για την ποσότητα  $\delta E_k$  ίση με ένα. Βέβαια, όπως είναι προφανές οι παράγοντες που επηρεάζουν ένα χαρακτή στη χάραξη του ίδιου συμβολογράμματος είναι πάρα πολλοί γεγονός που οδηγεί σε ποσότητες  $\delta E_k$  μικρότερες από τη μονάδα τόσο όσο λιγότερο καλλιγράφος είναι ο χαρακτήρας. Επίσης, όπως είναι φυσικό όταν πρόκειται για διαφορετικούς χαρακτήρες η ποσότητα  $\delta E_k$  είναι στατιστικά πολύ μικρότερη από την περίπτωση σύγκρισης συμβολογραμμάτων που έχουν γραφεί από το “ίδιο χέρι”.

### 1. Υπολογισμός των ποσοτήτων $\delta E_k$ για τις υλοποιήσεις ενός συμβολογράμματος σε μία επιγραφή.

Έστω μια επιγραφή, ας την ονομάσουμε  $I$ , με  $N_I$  υλοποιήσεις ενός συμβολογράμματος έστω του  $M$ . Η μεθοδολογία που χρησιμοποιούμε είναι να θεωρούμε διαδοχικά, ξεκινώντας από την πρώτη υλοποίηση, ως πρότυπο και να προσαρμόζουμε βέλτιστα όλες τις άλλες υλοποιήσεις σε αυτήν με τη διαδικασία που περιγράφηκε στην υποσημείωση 1. Στη συνέχεια υπολογίζουμε τις τέσσερις ποσότητες  $\delta E_k$   $k=1,2,3,4$  για τα τέσσερα “κτυπήματα” του συμβολογράμματος  $M$  σύμφωνα με την (4.1) και προσθέτουμε δύο δείκτες. Στον καινούριο συμβολισμό  $\delta E_{k,i,j}$  ο δείκτης  $i$  προσδιορίζει τον αύξοντα αριθμό της υλοποίησης που θεωρούμε ως πρότυπο και άρα παίρνει τιμές  $i= 1,2,\dots,N_I$  και ο δείκτης  $j$  προσδιορίζει όλες τις υπόλοιπες υλοποιήσεις του συμβολογράμματος  $M$  στην επιγραφή που προσαρμόζονται βέλτιστα στην  $i$  και άρα παίρνει τιμές  $j= 1,2,\dots,N_I$ . Μετά από των υπολογισμό των  $\delta E_{k,i,j}$ , υπολογίζουμε το μέσο όρο και τη δειγματική διασπορά των ποσοτήτων για κάθε “κτύπημα”  $k$ :

$$\delta E_k^{M,I} = \frac{1}{N_I - 1} \sum_{i=1}^{N_I - 1} \frac{1}{N_I - i} \sum_{j=i+1}^{N_I} \delta E_{k,i,j} \quad k = 1,2,3,4$$

$$SE_k^{M,I} = \sqrt{\frac{1}{N_I - 1} \sum_{i=1}^{N_I - 1} \frac{1}{N_I - i} \sum_{j=i+1}^{N_I} \delta E_{k,i,j} - (\delta E_k^{\Sigma I})^2} \quad k = 1,2,3,4$$

όπου το πρώτο άθροισμα αφορά τις πρότυπες υλοποιήσεις ενώ το δεύτερο στις τρέχουσες και ο πρώτος άνω δείκτης το εξεταζόμενο συμβολόγραμμα εδώ το M, ο δεύτερος άνω δείκτης τον αριθμό της επιγραφής και ο κάτω δείκτης τον αριθμό του αντίστοιχου “κτυπήματος”.

## 2. Υπολογισμός των ποσοτήτων $\delta E_k$ για τις υλοποιήσεις ενός συμβολογράμματος σε δύο διαφορετικές επιγραφές.

Έστω μια επιγραφή  $\alpha$ ς την ονομάσουμε I, με  $N_I$  υλοποιήσεις ενός συμβολογράμματος έστω του M, και μια δεύτερη επιγραφή  $\beta$ ς την ονομάσουμε J, με  $N_J$  υλοποιήσεις του ίδιου συμβολογράμματος. Ως πρότυπο υλοποίηση θεωρούμε την πρώτη εμφάνιση του συμβολογράμματος M στην I επιγραφή και προσαρμόζουμε βέλτιστα όλες τις υλοποιήσεις της J με τη διαδικασία που περιγράφηκε στην υποσημείωση 1. Για κάθε μια βέλτιστη προσαρμογή της J υπολογίζουμε τις ποσότητες  $\delta E_k$   $k=1,2,3,4$  σύμφωνα με την 4.1 τις  $\delta E_{k,1,j}$   $j=1,\dots,N_j$ . Στη συνέχεια θεωρούμε τη δεύτερη υλοποίηση του συμβολογράμματος M της επιγραφής I ως πρότυπο και υπολογίζουμε τις ποσότητες  $\delta E_{k,2,j}$   $j=1,\dots,N_j$  κ.ο.κ. για όλες τις  $N_I$  υλοποιήσεις του M στην επιγραφή I. Όπως και στο βήμα ένα υπολογίζουμε το μέσο όρο και τη δειγματική διασπορά των ποσοτήτων αυτών:

$$\delta E_k = \frac{S_{I,J}}{N_I} \sum_{i=1}^{N_I} \frac{1}{N_J} \sum_{j=1}^{N_J} \delta E_{k,i,j} \quad k = 1,2,3,4 \quad (4.2)$$

$$SE_k^{M,I,J} = \sqrt{\frac{1}{N_I} \sum_{i=1}^{N_I} \frac{1}{N_J} \sum_{j=1}^{N_J} (\delta E_{k,i,j} - \delta E_k^{M,I,J})^2} \quad k = 1,2,3,4 \quad (4.3)$$

Όπου ο πρώτος άνω δείκτης συμβολίζει το εξεταζόμενο συμβολόγραμμα εδώ το Μ, ο δεύτερος άνω δείκτης τον αριθμό της προτύπου επιγραφής, ο τρίτος την τρέχουσα , ενώ ο κάτω δείκτης συμβολίζει τον αριθμό του “κτυπήματος”.

**3. Ένα μέτρο για την εκτίμηση του εάν δύο διαφορετικές επιγραφές προέρχονται από τον ίδιο χαρακτή.**

Όπως παρατηρήσαμε και σε προηγούμενο σημείο οι τιμές των ποσοτήτων  $\delta E_k^{M,I}$  και  $\delta E_k^{M,I,J}$  έχουν άμεση συνάρτηση με τον αν οι επιγραφές I και J προέρχονται από τον ίδιο χαρακτή ή όχι. Στην πραγματικότητα οι ποσότητες αυτές αναμένεται να έχουν πολύ περισσότερο παραπλήσιες τιμές για όλα τα “κτυπήματα” όταν πρόκειται για επιγραφές που έχουν χαραχτεί από το ίδιο χέρι απ' όταν οι I και J προέρχονται από διαφορετικό χέρι. Η ορθότητα αυτού του ισχυρισμού ενισχύεται από το γεγονός ότι εφόσον κάθε ποσότητα  $\delta E_k$  είναι στην ουσία ποσοστό (λαμβάνουν τιμές πάντοτε στο  $[0,1]$ ) για  $N_I \geq 5$  και  $N_J \geq 5$  ακολουθεί κατά προσέγγιση κανονική κατανομή. Επίσης, εάν οι επιγραφές I και J προέρχονται από τον ίδιο χαρακτή, τότε οι πληθυσμιακές μέσες τιμές των ποσοτήτων  $\delta E_k^{M,I}$  και  $\delta E_k^{M,I,J}$  είναι ίσες. Άρα, ως μέτρο για το αν δύο επιγραφές προέρχονται από τον ίδιο χαρακτή επιλέγουμε το σύνολο των ποσοτήτων :

$$\delta_k^{M,I,J} = \frac{\delta E_k^{M,I} - \delta E_k^{M,I,J}}{\sqrt{\frac{(SE_k^{M,I})^2 + (SE_k^{M,I,J})^2}{N_I \frac{(N_I-1)}{2} + (N_I N_J)}}$$



Όσο πιο κοντά στο μηδέν είναι η ποσότητα  $\delta_k^{M,I,J}$  τόσο περισσότερο αναμένεται οι επιγραφές I και J να προέρχονται από τον ίδιο χαρακτή. Αντίστροφα, ένα οι I και J προέρχονται από διαφορετικό χέρι, αναμένεται να υπάρχει ικανός αριθμός συμβολογραμμάτων καθώς και “κτυπημάτων” των υλοποιήσεων τους για τα οποία οι ποσότητες  $\delta E_k^{M,I,J}$  διαφέρουν στατιστικά σημαντικά. Εφόσον, όπως προαναφέρθηκε οι ποσότητα του αριθμητή ακολουθεί κανονική κατανομή, η κάθε ποσότητα  $\delta_k^{M,I,J}$   $k = 1,2,3,4$  ακολουθεί κατανομή Student ή για  $N_I \geq 5$  και  $N_J \geq 5$  ακολουθεί με πολύ ικανοποιητική προσέγγιση τυπική κανονική κατανομή. Στην περίπτωση της κατανομής Student και εάν οι διασπορές  $(SE_k^{M,I})^2$  και  $(SE_k^{M,I,J})^2$  είναι ίσες τότε η κατανομή έχει βαθμούς ελευθερίας :

$$d_k^{M,I,J} = N_I N_J + \frac{N_I(N_I - 1)}{2} - 2$$

ενώ αν οι διασπορές είναι άνισες και θέσουμε  $N_{II} = N_I$  και  $N_{IJ} = N_I N_J$  τότε έχουμε :

$$d_k^{M,I,J} = \frac{((SE_k^{M,I})^2 / N_{II} + (SE_k^{M,I,J})^2 / N_{IJ})^2}{\frac{((SE_k^{M,I})^2 / N_{II})^2}{N_{II} - 1} + \frac{((SE_k^{M,I,J})^2 / N_{IJ})^2}{N_{IJ} - 1}}$$

Τέλος εφόσον δεν γνωρίζουμε από την αρχή αν οι πληθυσμιακές διασπορές των ποσοτήτων  $\delta_k^{M,I,J}$  είναι ίσες ή άνισες επιλέγουμε την τελευταία σχέση για τους βαθμούς ελευθερίας της Student διότι καλύπτει και τις δύο περιπτώσεις.

## 4.2 Τα κριτήρια ομοιότητας για σύγχρονους γραφείς.

Στο παρόν κεφάλαιο εισάγουμε ένα κριτήριο ομοιότητας αλφαβητικών συμβόλων σύγχρονων γραφών ακολουθώντας τη λογική που χρησιμοποιήσαμε αναφορικά με τις αρχαίες επιγραφές και περγαμηνές. Αρχικά, θεωρούμε δύο διαφορετικές υλοποιήσεις του ίδιου συμβόλου, είτε αυτές προέρχονται από το ίδιο έγγραφο είτε από διαφορετικό. Ονομάζουμε αυτές τις υλοποιήσεις  $Y_1$  και  $Y_2$  και τα περιγράμματα αυτών  $\Pi_1$  και  $\Pi_2$ . Τα περιγράμματα  $\Pi_1$  και  $\Pi_2$  θεωρούμε ως απλά πολύγωνα<sup>2</sup>.

Για να προχωρήσουμε στην εξαγωγή του κριτηρίου ομοιότητας των δύο συμβολογραμμάτων, θα πρέπει να προσδιορίσουμε τη θέση βέλτιστης προσαρμογής των δύο περιγραμμάτων σύμφωνα με τη μεθοδολογία που αναπτύχθηκε συνοπτικά στην υποσημείωση 1 του κεφαλαίου 4.1. Τουτέστιν, θεωρούμε το περίγραμμα  $\Pi_1$  ως πρότυπο και εφαρμόζουμε στο  $\Pi_2$  περιστροφή, ομοιοθεσία και παράλληλη μετατόπιση πολύ κοντά στη θέση ισορροπίας. Μετά από αυτή την προεργασία, είμαστε έτοιμοι να προχωρήσουμε στην εξαγωγή του κριτηρίου ομοιότητας.

Σε αυτή τη φάση και αφού έχουμε προσδιορίσει τις σχετικές παραμέτρους μετασχηματισμού  $\chi_\delta, \gamma_\delta, \varphi, \lambda$  της προαναφερθείσας διαδικασίας ορίζουμε ένα τετραγωνικό κύβο γύρο από αυτό το σημείο και μία τέλεια διαμέριση αυτού. Σε κάθε σημείο αυτής της διαμέρισης, αντιστοιχεί ένας κατάλληλος μετασχηματισμός του περιγράμματος  $\Pi_2$ . Για κάθε μετασχηματισμένη εκδοχή του περιγράμματος  $\Pi_2$ , θεωρούμε ότι το μετασχηματισμένο εσωτερικό της υλοποίησης  $Y_2$  και ορίζουμε το κριτήριο :

$$dE_n = \frac{\text{περιοχή της εσωτερικής } Y_1 \text{ και της μετασχηματισμένης } Y_2}{\text{περιοχή της ένωσης των εσωτερικών } Y_1 \text{ και της μετασχηματισμένης } Y_2} \quad (4.2.1)$$

όπου το  $n$  διατρέχει όλα τα σημεία διαμέρισης του τετραγωνικού κύβου. Στη συνέχεια, αφού υπολογίσουμε τη μέγιστη τιμή όλων των  $dE_n$  προκρίνουμε το ακόλουθο κριτήριο: “Η κατάλληλη

---

<sup>2</sup>Σε αυτό το σημείο πρέπει να σημειώσουμε το εξής. Αν και τα περιγράμματα των υλοποιήσεων των συμβόλων μπορεί να περικλύουν πολυγωνικές κυλότητες, σε όλες τις περιπτώσεις το λήμμα εφαρμόζεται μόνο στις εξωτερικές κυλότητες των περιγραμμάτων αυτών.

σχετική θέση του περιγράμματος  $\Pi_1$  μαζί με το κατάλληλα μετασχηματισμένο περίγραμμα  $\Pi_2$ , δίνουν την τιμή του  $\delta E_n$  η οποία βέλτιστα εκφράζουν την ομοιότητα των δύο υλοποιήσεων  $Y_1$  και  $Y_2$ . Σημειώνουμε ότι, στην ιδανική περίπτωση όπου το  $\Pi_1$  και το μετασχηματισμένο  $\Pi_2$  συμπίπτουν, τότε  $\delta E^0 = 1$ , ενώ όταν οι δύο υλοποιήσεις των συμβολογραμμάτων είναι ριζικά διαφορετικά, τότε το κριτήριο  $\delta E^0$  γίνεται χαρακτηριστικά μικρό.

Σε αυτό το σημείο και για λόγους πληρότητας, πρέπει να επεκτείνουμε το κριτήριο για την περίπτωση γραμμάτων που απαρτίζονται από περισσότερες από μια απλά συνδεδεμένη περιοχή. Για τις απλά συνδεδεμένες περιοχές των υλοποιήσεων συμβολογραμμάτων θα χρησιμοποιήσουμε τον όρο “χτύπημα”. Ο όρος αυτός, όπως γίνεται εμφανές και από την παράγραφο 4.1 έχει τη βάση του στους αρχαίους γραφείς – χαρακτες και στα εργαλεία που χρησιμοποιούσαν για να χαράξουν τα συμβολογράμματα πάνω στις πλάκες. Παρ' ότι το φαινόμενο των πολλαπλά συνδεδεμένων συμβολογραμμάτων συναντώνται περισσότερο στις αρχαίες επιγραφές, εμφανίζονται επίσης σε βυζαντινούς κώδικες καθώς επίσης και σε σύγχρονα κείμενα.

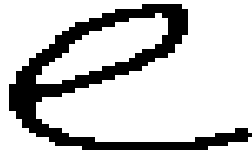
Έστω λοιπόν δύο υλοποιήσεις  $X_1$  και  $X_2$  οι οποίες αποτελούνται από  $N_1^S$  και  $N_2^S$  διακριτά “χτυπήματα αντίστοιχα, όπου όπως προείπαμε κάθε “χτύπημα” είναι μια απλά συνδεδεμένη περιοχή. Έπειτα, σύμφωνα και με τη μεθοδολογία που ακολουθήσαμε στην παράγραφο 5.1, αυθαίρετα ορίζουμε μια υλοποίηση ως σταθερή και εφαρμόζουμε στην άλλη υλοποίηση όλους τους γνωστούς μετασχηματισμούς δηλαδή περιστροφή, ομοιοθεσία και παράλληλη μετατόπιση. Κατά τα γνωστά, οι παράμετροι των μετασχηματισμών βρίσκονται μέσα σε ένα μικρό τετραγωνικό κύβο γύρω από το σημείο ισορροπίας. Αυτό που αλλάζει στην περίπτωση αυτή είναι η συνάρτηση σφάλματος η οποία είναι διαφορετική από αυτή της προηγούμενης παραγράφου. Στην πραγματικότητα αυτό που αλλάζει είναι ότι τώρα στο πηλίκο της συνάρτησης θα χρησιμοποιήσουμε το συνολικό εμβαδό των απλά συνδεδεμένων περιοχών. Το συνολικό εμβαδό για τη  $X_1$  είναι η ένωση των  $N_1^S$  “χτυπημάτων” έστω  $U_1$  και αντίστοιχα για τη  $X_2$  είναι η ένωση των  $N_2^S$  “χτυπημάτων” έστω  $U_2$ . Τώρα μπορούμε να προχωρήσουμε στον ορισμό της συνάρτησης σφάλματος :

$$\delta E_n^S = \frac{\text{περιοχή της στομήτων } U_1 \text{ και } U_2}{\text{περιοχή της ένωσης των } U_1 \text{ και } U_2} \quad (4.2.2)$$

Όσον αφορά το  $n$ , αυτό διατρέχει όλα τα σημεία της διαμέρισης του τετραγωνικού κύβου που έχουμε ορίσει γύρω από το σημείο που επιλέξαμε με βάση το Λήμμα.

Για να υπολογίσουμε το βέλτιστο ταίριασμα της υλοποίησης  $X_1$  με τη μετασχηματισμένη  $X_2$ , βρίσκουμε τη μέγιστη τιμή του  $\delta E_n^S$  πάνω στο  $n$ , η οποία μας δίνει τις βέλτιστες τιμές των παραμέτρων. Σε αυτό το σημείο πρέπει να σημειωθεί ότι ο παράγοντας που καθορίζει την ακρίβεια του βέλτιστου ταιριάσματος του  $X_1$  με το μετασχηματισμένο  $X_2$  είναι το βήμα της διαμέρισης του τετραγωνικού κύβου. Προφανώς, για το κριτήριο αυτό ισχύει η ίδια λογική που ισχύει και στο 4.2.1. Η τιμή της ποσότητας  $\delta E_n^S$  στη θέση βέλτιστης προσαρμογής την οποία και ονομάζουμε  $\delta E^0$  είναι το κριτήριο του βαθμού ομοιότητας των δύο υλοποιήσεων που συγκρίνονται. Όσο η τιμή του  $\delta E^0$  πλησιάζει τη μονάδα τόσο μεγαλύτερη είναι η ομοιότητα των δύο υλοποιήσεων. Αντίθετα, όσο μικρότερη είναι η τιμή του  $\delta E^0$  τόσο περισσότερο διαφέρουν οι δύο υλοποιήσεις.

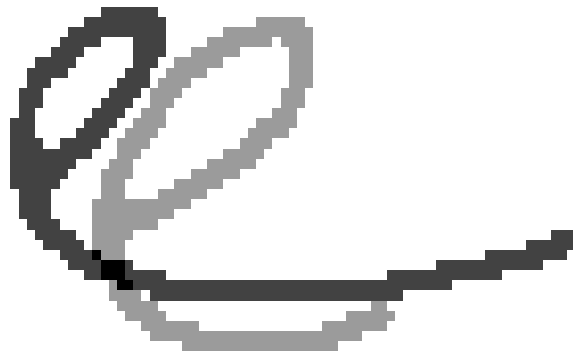
Παρακάτω παρατίθενται φωτογραφίες από υλοποιήσεις γραμμμάτων στην εξής μορφή: αρχικά παρουσιάζονται οι υλοποιήσεις των γραμμμάτων όπως χωρίς καμία επεξεργασία. Στη συνέχεια θεωρούμε τη μια ως πρότυπη και εφαρμόζουμε κατά τα γνωστά ομοιοθεσία, παράλληλη μετατόπιση και περιστροφή. Τέλος εμφανίζουμε το ταίριασμα των δυο υλοποιήσεων πριν την επεξεργασία και την τελική μορφή μετά την προσαρμογή. Να τονίσουμε εδώ, ότι η προσαρμογή που διενεργείται στη δεύτερη υλοποίηση είναι τέτοια ώστε να μην αλλοιώνεται η αρχική μορφή του γράμματος. Αν κάναμε μεγαλύτερη επεξεργασία στις υλοποιήσεις ελλόχευε ο κίνδυνος να αλλάξουμε το γραφικό χαρακτήρα του γραφέα και να αμφισβητηθούν έτσι τα αποτελέσματα μας. Ανάμεσα στις υλοποιήσεις που έχουμε διαλέξει υπάρχουν περιπτώσεις όπου μετά την επεξεργασία έχει βελτιωθεί κατά πολύ η σχετική θέση των δύο γραμμμάτων ενώ σε άλλες είτε αυτό δεν ήταν δυνατόν είτε δε χρειάστηκε. Στις απεικονίσεις όπου παρουσιάζονται τα ταιριάσματα των υλοποιήσεων με έντονο μαύρο χρώμα σημειώνεται η υλοποίηση που θεωρείται ως πρότυπη και με γκρι η υλοποίηση που υφίσταται ομοιοθεσία, παράλληλη μετατόπιση και περιστροφή.



Εικόνα 4.19 Η υλοποίηση του  $e$  που θα θεωρηθεί ως **πρότυπη** και δε θα υποστεί προσαρμογή



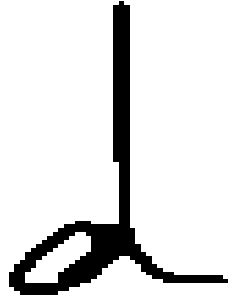
Εικόνα 4.20 Η υλοποίηση του  $e$  που θα υποστεί ομοιοθεσία, παράλληλη μετατόπιση και περιστροφή



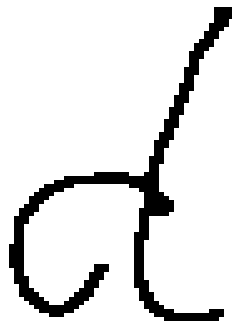
Εικόνα 4.21 Ένα πρώτο ταίριασμα των δύο υλοποιήσεων **πριν** την επεξεργασία.



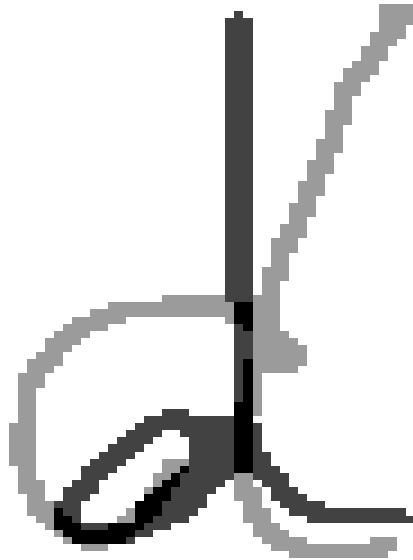
Εικόνα 4.22 Το ταίριασμα των δυο υλοποιήσεων **μετά** την προσαρμογή



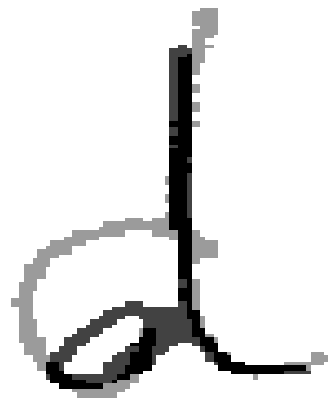
Εικόνα 4.23 Η υλοποίηση του  $\delta$  που θα θεωρηθεί ως **πρότυπη** και δε θα υποστεί προσαρμογή



Εικόνα 4.24 Η υλοποίηση του  $\delta$  που θα υποστεί ομοιοθεσία, παράλληλη μετατόπιση και περιστροφή

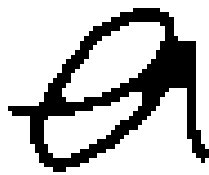


Εικόνα 4.25 Ένα πρώτο ταίριασμα των δύο υλοποιήσεων **πριν** την επεξεργασία.

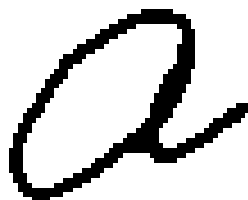


Εικόνα 4.26 Το ταίριασμα των δυο υλοποιήσεων **μετά** την προσαρμογή





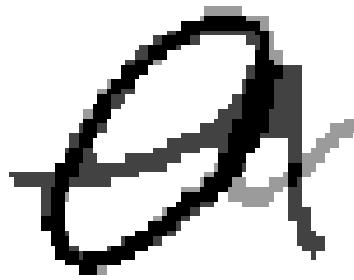
Εικόνα 4.27 Η υλοποίηση του  $\alpha$  που θα θεωρηθεί ως **πρότυπη** και δε θα υποστεί προσαρμογή



Εικόνα 4.28 Η υλοποίηση του  $\alpha$  που θα υποστεί ομοιοθεσία, παράλληλη μετατόπιση και περιστροφή



Εικόνα 4.29 Ένα πρώτο ταίριασμα των δύο υλοποιήσεων **πριν** την επεξεργασία.



Εικόνα 4.30 Το ταίριασμα των δυο υλοποιήσεων **μετά** την προσαρμογή

## 5. ΕΦΑΡΜΟΓΗ ΤΩΝ ΚΡΙΤΗΡΙΩΝ ΣΤΗΝ

## ΕΚΤΙΜΗΣΗ ΟΜΟΙΟΤΗΤΑΣ ΓΡΑΜΜΑΤΩΝ

### 5.1 Εξαγωγή των περιγραμμάτων των γραμμάτων

Για την εξαγωγή των περιγραμμάτων των γραμμάτων χρησιμοποιήσαμε προσέγγιση πολυωνύμων ως προς το μήκος. Η μέθοδος αυτή συνίσταται στον προσδιορισμό των κατάλληλων πολυωνύμων 3ου βαθμού τα οποία προσεγγίζει βέλτιστα μια ακολουθία εικονοστοιχείων του περιγράμματος το καθένα.

Αρχικά, επιλέγουμε ένα υποσύνολο εικονοστοιχείων του περιγράμματος πλήθους  $L_S$  έστω τα  $P_1, P_2, \dots, P_{L_S}$ . Η απόσταση μεταξύ δύο διαδοχικών εικονοστοιχείων δίνεται ως γνωστόν από τη σχέση:

$$d(P_i, P_{i+1}) = \sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2}$$

όπου  $(x_i, y_{i+1})$  είναι οι συντεταγμένες του  $P_i$  και  $(x_{i+1}, y_{i+1})$  οι συντεταγμένες του  $P_{i+1}$ . Στο επόμενο βήμα ορίζουμε τα πολυώνυμα 3ου βαθμού τα οποία προσεγγίζουν βέλτιστα τα υποσύνολα εικονοστοιχείων. Τα πολυώνυμα αυτά έχουν ως παράμετρο το μήκος του εκάστοτε αντιστοίχου προς προσέγγιση τμήματος. Η παράμετρος αυτή δίνεται από την παρακάτω ακολουθία:

$$\xi_1 = 0 \text{ και } \xi_i + 1 = \sum_{k=1}^i d(P_k, P_{(k+1)}) \quad i = 1, 2, \dots, L_S - 1. \quad (5.1.1)$$

Ο τυχαίος όρος  $\xi_i$  της ακολουθίας αντιστοιχεί στο μήκος του τμήματος που ξεκινάει από το εικονοστοιχείο  $P_1$  και καταλήγει στο  $P_i$ . Άρα, οι συντεταγμένες κάθε εικονοστοιχείου προσεγγίζονται με τα πολυώνυμα :

$$\hat{x}_i = a_0 \xi_i^3 + a_1 \xi_i^2 + a_2 \xi_i + a_3 i = 1, 2, \dots, L_S$$

και

$$\hat{y}_i = b_0 \xi_i^3 + b_1 \xi_i^2 + b_2 \xi_i + b_3 i = 1, 2, \dots, L_S$$

Οι συντελεστές  $a_0, a_1, a_2, a_3, b_0, b_1, b_2, b_3$  προσδιορίζονται όταν ελαχιστοποιήσουμε τη συνάρτηση σφάλματος προσέγγισης σύμφωνα με τη λογική των ελαχίστων τετραγώνων. Συγκεκριμένα η συνάρτηση αυτή είναι η ακόλουθη:

$$E = \sum_{i=1}^{L_S} \{(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2\}$$

Όπως παρατηρούμε, η συνάρτηση αυτή είναι συνεχώς παραγωγίσιμη ως προς τους συντελεστές  $a_0, a_1, a_2, a_3, b_0, b_1, b_2, b_3$  και είναι και κάτω φραγμένη από το 0. Άρα, έχει τουλάχιστον ένα απόλυτο ελάχιστο, στο οποίο όλες οι μερικές παράγωγοι είναι ίσες με το μηδέν. Ικανοποιώντας αυτή τη συνθήκη καταλήγουμε σε ένα γραμμικό σύστημα το οποίο μας δίνει τους συντελεστές που ελαχιστοποιούν την ποσότητα σφάλματος E. Συμβολίζουμε με  $Sx_1$  και  $Sy_1$  τα πολυώνυμα που προσδιορίσαμε και αντιστοιχούν στο σημειοσύνολο που είχαμε αρχικά επιλέξει  $P_1, P_2, \dots, P_{L_S}$ .

Η διαδικασία που περιγράψαμε παραπάνω για να προσδιοριστούν τα  $Sx_1$  και  $Sy_1$  δεν αρκεί για την επεξεργασία όλου του περιγράμματος ενός συμβολογράμματος. Αυτό που πρέπει να γίνει τώρα είναι να επαναληφθεί η διαδικασία τόσες φορές όσες χρειάζεται για να προσεγγιστεί όλο το περίγραμμα. Συνεπώς, σε δεύτερη φάση, ξεκινάμε από ένα τυχαίο pixel  $x_j, y_j$  του περιγράμματος, και προσδιορίζουμε κατά τα γνωστά το ζεύγος πολυωνύμων  $Sx_j, Sy_j$  το οποίο βέλτιστα προσεγγίζει τα pixels  $P_j, P_{j+1}, \dots, P_{j+L_S-1}$ . Το μόνο που μένει τώρα είναι να προσδιορίσουμε τα εικονοστοιχεία

εκκίνησης έτσι ώστε να καλύπτεται όλο το περίγραμμα του τυχαίου γράμματος με επικαλυπτόμενα ζεύγη πολυωνύμων 3ου βαθμού,

$$Sx_j, Sy_j, j = 1, 2, \dots, N_S$$

Η διαδικασία επιλογής των εικονοστοιχείων εκκίνησης είναι η ακόλουθη. Αρχικά επιλέγουμε τυχαία ένα εικονοστοιχείο του περιγράμματος το οποίο ονομάζουμε  $P_1$  και δημιουργούμε την αλυσίδα  $A_1P_1, P_2, \dots, P_{L_S-1}$  και εξάγουμε κατά τα γνωστά τις ακολουθίες  $Sx_1$  και  $Sy_1$ . Στη συνέχεια, προσδιορίζουμε την ποσότητα  $AM$  η οποία αντιστοιχεί στο 2.5% του συνολικού αριθμού των εικονοστοιχείων του περιγράμματος δηλαδή  $AM = [0.025L_p]$ . Ακολούθως, επαναλαμβάνοντας τη μεθοδολογία, δημιουργούμε την αλυσίδα  $A_2P_{1+AM}, \dots, P_{1+L_S+AM-1}$  για την οποία προσδιορίζουμε τις ακολουθίες  $Sx_2$  και  $Sy_2$ , κατόπιν την αλυσίδα  $A_2P_{1+2AM}, \dots, P_{1+L_S+2AM-1}$  και τα πολυώνυμα  $Sx_3$  και  $Sy_3$ , μέχρι να ικανοποιηθεί η συνθήκη  $vAM \leq L_p$  και  $(v+1)AM > L_p$ . Η προτεινόμενη μεθοδολογία, προσφέρει ένα πλήθος πολυωνύμων  $Sx_i, Sy_i$  για κάθε εικονοστοιχείο του περιγράμματος. Παίρνοντας το μέσο όρο αυτών των πολυωνύμων, επιτυγχάνουμε μια πολύ ομαλή κάλυψη του περιγράμματος του συμβολογράμματος, η οποία είναι και πρακτικά ανεξάρτητη του σημείου εκκίνησης. Ένα επιπλέον κέρδος που αποκομίζουμε χρησιμοποιώντας αυτή τη μέθοδο, είναι ότι καλύπτουμε τα πιθανά μικρά κενά που υπάρχουν σε πολλές περιπτώσεις λόγω φθοράς των επιγραφών με τα αντίστοιχα σημεία των πολυωνύμων προσέγγισης.

Τέλος, για κάθε σημείο της ομαλοποιημένης προσέγγισης του περιγράμματος υπολογίζουμε την καμπυλότητα, το μέσο όρο της και τη διασπορά της. Από τις τιμές αυτές επιλέγουμε αυτές που παρουσιάζουν τοπικό μέγιστο και έχουν τιμή τουλάχιστον τριπλάσια της διασποράς τους. Τα σημεία του περιγράμματος που ικανοποιούν τις παραπάνω συνθήκες τα θεωρούμε σημεία καμπής και τα ονομάζουμε κρίσιμα σημεία. Εξαίρεση θεωρούμε τα συμβολογράμματα  $O, \Phi, \Omega$  για τα οποία δεν ορίζουμε σημεία καμπής εφόσον αυτά είναι απομονωμένα. Προφανώς η ίδια διαδικασία εφαρμόζεται και σε περίπτωση που ένα συμβολόγραμμα αποτελείται από νησίδες, επαναλαμβάνοντας τη μέθοδο για κάθε νησίδα ξεχωριστά και ορίζοντας ως  $L_p$  το πλήθος των εικονοστοιχείων κάθε νησίδας.

Τέλος αφήνουμε κατάλληλα ζεύγη κρίσιμων σημείων καμπής να σχηματίζουν τις πλευρές του εκάστοτε συμβόλου της αλφαβήτου, σύμφωνα με την ανθρώπινη αίσθηση.



Εικόνα 5.1

Ψηφιακή αναπαράσταση του συμβολογράμματος Α. Με μπλε χρώμα μπορούμε να διακρίνουμε το περίγραμμα και με κίτρινο τις διαδοχικές προσεγγίσεις του περιγράμματος. Με κόκκινο χρώμα σημειώνουμε το σημείο εκκίνησης ενώ πρέπει να σημειωθεί ότι το κάθε “χτύπημα” προσεγγίζεται αυτόνομα.



Εικόνα 5.1

Ψηφιακή αναπαράσταση του συμβολογράμματος Σ. Με μπλε χρώμα μπορούμε να διακρίνουμε το περίγραμμα και με κίτρινο τις διαδοχικές προσεγγίσεις του περιγράμματος. Με κόκκινο χρώμα σημειώνουμε το σημείο εκκίνησης ενώ πρέπει να σημειωθεί ότι το κάθε “χτύπημα” προσεγγίζεται αυτόνομα.



Εικόνα 5.3

Ψηφιακή αναπαράσταση του συμβολογράμματος Α με πολλαπλές προσεγγίσεις . Εδώ παρουσιάζονται τέσσερις πολλαπλές προσεγγίσεις με διαφορετικό χρώμα η κάθε μια . Με πράσινο χρώμα σημειώνουμε την τελευταία προσέγγιση η οποία είναι η τελευταία χρονολογικά ενώ με μπλε χρώμα παρουσιάζεται το αρχικό περίγραμμα.





Εικόνα 5.4

Ψηφιακή αναπαράσταση του συμβολογράμματος  $\Sigma$  με πολλαπλές προσεγγίσεις . Εδώ παρουσιάζονται τέσσερις πολλαπλές προσεγγίσεις με διαφορετικό χρώμα η κάθε μια . Με πράσινο χρώμα σημειώνουμε την τελευταία προσέγγιση η οποία είναι η τελευταία χρονολογικά ενώ με μπλε χρώμα παρουσιάζεται το αρχικό περίγραμμα.

## 5.2 Η διαδικασία εκτέλεσης.

Μας δόθηκε μια βάση δεδομένων (συγκεκριμένα η IM-UNIPEN) που περιείχε φωτογραφίες από χειρόγραφα κείμενα στα Αγγλικά. Συγκεκριμένα, πρόκειται για απομονωμένες λέξεις με μαύρα γράμματα σε λευκό φόντο. Το πρώτο βήμα πριν τη διαδικασία τρεξίματος ήταν να “κόψουμε”, δηλαδή να απομονώσουμε τα συμβολογράμματα από τις λέξεις. Έτσι, δημιουργήσαμε μια βάση δεδομένων από εικόνες στην οποία κάθε καινούρια εικόνα περιείχε ένα μόνο σύμβολο. Κατά τη διαδικασία αυτή, τα σύμβολα επιλέχθηκαν με κριτήριο την πολυπλοκότητα του σχήματος τους ώστε να είναι περισσότερο διακριτές οι διαφορές ανάμεσα στους γραφείς. Για παράδειγμα προτιμήθηκαν συμβολογράμματα όπως τα “w” και “a”. Φροντίσαμε το δείγμα να είναι πολύ μεγάλο της τάξης μερικών χιλιάδων συμβολογραμμάτων για κάθε γράμμα και για μερικές δεκάδες γραφείων έτσι ώστε τα στατιστικά αποτελέσματα να προσφέρουν ένα ασφαλές συμπέρασμα για την εργασία μας.

Η διαδικασία της απομόνωσης συμβολογραμμάτων, ήταν μια επίπονη διαδικασία η οποία παρά την απλότητα που μπορεί να εμφανίζει με μια πρώτη ματιά, απαιτούσε ιδιαίτερη προσοχή και λεπτομερή εργασία. Για το σκοπό αυτό χρησιμοποιήθηκε εξειδικευμένο πρόγραμμα επεξεργασίας εικόνων. Για παράδειγμα κάποιοι γραφείς δε σηκώνουν τη γραφίδα όταν γράφουν, με αποτέλεσμα τα γράμματα να είναι συνεχόμενα ή ακόμα και να επικαλύπτονται. Έτσι, συχνά έπρεπε με μεγάλη ακρίβεια να απομονώσουμε το γράμμα που θέλαμε από τα υπόλοιπα από τα υπόλοιπα pixel που δημιουργούσαν ουσιαστικά “οπτικό θόρυβο”. Στη συνέχεια, μετά την επεξεργασία και απομόνωση των συμβολογραμμάτων ιδιαίτερη προσοχή απαιτούσε η ονοματοδοσία κάθε εικόνας. Χρειάστηκε να χρησιμοποιήσουμε συγκεκριμένο μοντέλο ονοματοδοσίας το οποίο μας ήταν απαραίτητο για την επόμενη φάση. Συγκεκριμένα κάθε εικόνα είχε όνομα της μορφής : “XXXXX\_XX\_XX\_YYY\_A\_ZZZ” όπου “XXXXX\_XX\_XX” αποτελούσε τον αριθμό που προσδιόριζε το κείμενο και το γραφέα, “YYY” ο αριθμός της λέξης του κειμένου “A” το γράμμα και “ZZZ” ο αύξοντας αριθμός της υλοποίησης του γράμματος στο κείμενο. Οι υλοποιήσεις των συμβολογραμμάτων αποθηκεύτηκαν ταξινομημένοι σε φακέλους ανά γράμμα για το κάθε κείμενο.

Σε δεύτερη φάση, για να προχωρήσουμε στη σύγκριση των συμβολογραμμάτων χρειάστηκε η εξαγωγή των περιγραμμάτων των υλοποιήσεων τους. Αυτό έγινε με χρήση του προγράμματος Matlab (σ. Έκδοση 2009 B) και κατάλληλου κώδικα που μας δόθηκε. Τα περιγράμματα για κάθε συμβολογράμμα αποθηκεύτηκαν σε φακέλους με την αντίστοιχη δομή και ονοματολογία με τους παραπάνω. Το όνομα δε κάθε περιγράμματος ήταν το όνομα του αντίστοιχου γράμματος προσθέτοντας τη λέξη “\_PERIGRAMMA” στο τέλος. Τα αρχεία που προέκυψαν από αυτή τη φάση

δεν ήταν πλέον εικόνες αλλά αρχεία .mat .

Στο τελευταίο βήμα το οποίο ήταν το σημαντικότερο και πιο χρονοβόρο έγινε η σύγκριση των περιγραμμάτων που είχαμε εξάγει προηγουμένως. Η σύγκριση έγινε και σε αυτό το στάδιο με χρήση του προγράμματος Matlab και χρήση κατάλληλου κώδικα. Καταρχήν, όπως είναι προφανές συγκρίναμε μόνο υλοποιήσεις ίδιων συμβολογραμμάτων. Δεν υπήρχε κανένα επιστημονικό ενδιαφέρον να συγκρίνουμε ένα “a” με ένα “w” . Η μέθοδος συγκρίσεων που ακολουθήθηκε μπορεί να ονομαστεί σχηματικά “όλα με όλα”. Σε αυτό το σημείο πρέπει να κάνουμε σαφές ότι οι συγκρίσεις έγιναν κείμενο προς κείμενο στη βάση ενός συγκεκριμένου γράμματος και για όλα τα γράμματα που είχαμε διαθέσιμα σε κάθε κείμενο. Συνεπώς, συγκρίναμε κάθε κείμενο με τον εαυτό του ώστε να έχουμε ένα βασικό κριτήριο που δείχνει πόσο μοιάζουν τα γράμματα του ίδιου γραφέα μεταξύ τους. Σε δεύτερη φάση συγκρίναμε διαφορετικά κείμενα μεταξύ τους βγάλαμε στατιστικά αποτελέσματα για την ομοιότητα των γραμμάτων. Η σύγκριση αυτών των αποτελεσμάτων με τα αρχικά αποτελέσματα ομοιότητας κειμένων με τον εαυτό τους μας βοήθησε να βγάλουμε ασφαλή συμπεράσματα για την ταυτοποίηση του κάθε γραφέα.

Οι παραπάνω κώδικες είχαν αρχικά σχεδιαστεί για να δουλέψουν πάνω σε αρχαίες επιγραφές και μια διαφορετική βάση δεδομένων. Έτσι για την ορθή λειτουργία και για τη συγκεκριμένη βάση δεδομένων και ονοματολογία χρειάστηκε να γίνουν κάποιες μικρές προσαρμογές.

## 5.3 Αποτελέσματα

### 5.3.1 Από αρχαίους γραφείς.

Σε αυτό το σημείο της εργασίας μας πραγματοποιήσαμε έλεγχο αληθείας το κριτηρίου σε δοσμένη βάση δεδομένων αρχαίων επιγραφών. Η βάση δεδομένων αρχαίων επιγραφών έχουν επιλεγεί από τον καθηγητή κ. Steven Tracy, καθηγητή στο center of advanced research του Princeton University και τ. Διευθυντής της Αμερικανικής Σχολής Κλασικών Σπουδών. Θεωρείται ένας από τους μεγαλύτερους επιγραφολόγους / αρχαιολόγους παγκοσμίως και έχει αφιερώσει πάνω από τέσσερις δεκαετίες της ζωής του στη μελέτη των Αρχαίων Ελληνικών επιγραφών. Η βάση δεδομένων που μας δόθηκε περιείχε 32 επιγραφές διαφόρων χειρών για τις οποίες δεν είχαμε καμία πληροφορία σχετικά με αυτές , ούτε για το σημείο στο οποίο ανεσκάφησαν ούτε για το που αναφαιρέτο το περιεχόμενο τους ούτε για την ευρύτερη εποχή κατά την οποία σμιλεύτηκαν, παρά μόνο ότι αυτές ανήκουν σε εννέα διαφορετικούς γραφείς. Επίσης, για να εκτελέσουμε το πειραματικό στάδιο μας δόθηκε η πληροφορία για το γραφέα τεσσάρων αρχαίων επιγραφών και η αντιστοιχία τους με τέσσερις διαφορετικούς γραφείς. Στόχος της μελέτης μας ήταν με βάση το κριτήριο μέγιστης πιθανοφάνειας να αποφανθούμε για τις υπόλοιπες “άγνωστες” επιγραφές αν ανήκουν σε κάποιο από τους παραπάνω τέσσερις γραφείς. Για το σκοπό αυτό ελέγξαμε το κριτήριο μέγιστης πιθανοφάνειας για τα γράμματα Α,Σ,Μ και Ο καθώς και τη συνολική πιθανοφάνεια.

Παρακάτω ακολουθεί ο πίνακας με τις αρχαίες επιγραφές όπως αυτές μας δόθηκαν με τον κωδικό της Αμερικανικής Σχολής Κλασικών Σπουδών. Για αυτές όπως αναφέρθηκε και παραπάνω γνωρίζαμε μόνο ότι οι επιγραφές 7400, 7519, 7190 και 0247 ανήκαν σε διαφορετικούς χαρακτες τους οποίους ονομάσαμε “Χέρι 1” , “Χέρι 2” , “Χέρι 3” και “Χέρι 4” αντίστοιχα. Τέλος ακολουθούν οι πίνακες με τα αποτελέσματα που συλλέξαμε από τη διαδικασία εκτέλεσης καθώς και ένας συγκεντρωτικός πίνακας με τα συμπεράσματα που προέκυψαν. Να σημειωθεί εδώ, ότι ο έλεγχος ήταν εξαντλητικός, δηλαδή για κάθε χέρι και για κάθε γράμμα Α,Σ,Μ και Ο θεωρήσαμε αυτά ως πρότυπα και προσαρμόσαμε τα αντίστοιχα γράμματα των επιγραφών στις υλοποιήσεις γραμμάτων των γνωστών χειρών. Το μοναδικό κριτήριο εφαρμογής ήταν ο αριθμός των

υλοποιήσεων να μην είναι μικρότερος των τεσσάρων σε κάθε επιγραφή.

Κατάλογος Διαθεσίμων Επιγραφών Προς Ταυτοποίηση Γραφώς				
$E_1$	0068		$E_{17}$	7237
$E_2$	0247		$E_{18}$	7245
$E_3$	0286		$E_{19}$	7254
$E_4$	1640		$E_{20}$	7335
$E_5$	4033		$E_{21}$	7400
$E_6$	4266		$E_{22}$	7405
$E_7$	4424		$E_{23}$	7446
$E_8$	4917		$E_{24}$	7457
$E_9$	5297		$E_{25}$	7481
$E_{10}$	6006		$E_{26}$	7482
$E_{11}$	6053		$E_{27}$	7542
$E_{12}$	6124		$E_{28}$	7587
$E_{13}$	6295		$E_{29}$	7519
$E_{14}$	6671		$E_{30}$	7566
$E_{15}$	7041		$E_{31}$	7567
$E_{16}$	7190		$E_{32}$	7723

**Πρότυπο 7400**  
(Απεδόθη στο λιθοξόο 1 )

α/ α	Πιθαν/α Α	Κατ/ξη Α	Πιθαν/α Σ	Κατ/ξη Σ	Πιθαν/α Μ	Κατ/ξη Μ	Πιθαν/α Ο	Κατ/ξη Ο	Συνολική Πιθαν/α	Συνολική Κατ/ξη
1	1.550e-002	7400	1.321e-002	7400	1.293e-002	7400	3.603e-002	7400	9.533e-008	7400
2	1.362e-002	7405	1.010e-002	7405	1.018e-002	7190	3.286e-002	7405	4.107e-008	7405
3	1.175e-002	7245	9.532e-003	7190	9.628e-003	7542	1.751e-002	7542	1.787e-008	7542
4	1.171e-002	6671	9.523e-003	6124	9.332e-003	6124	1.705e-002	6124	1.725e-008	6124
5	1.142e-002	7542	9.293e-003	7542	9.092e-003	7405	1.670e-002	7245	3.495e-012	7041
6	1.140e-002	6124	9.017e-003	7723	8.596e-003	7237	1.305e-003	7587	3.325e-012	6671
7	1.115e-002	7041	8.888e-003	7041	2.140e-003	7723	1.185e-003	7041	6.735e-013	7245
8	1.099e-002	4266	8.695e-003	7457	2.104e-003	6671	1.157e-003	7237	5.251e-013	4266
9	6.873e-004	6006	8.257e-003	0247	1.983e-003	7482	1.041e-003	6671	2.120e-013	7723
10	6.659e-004	0247	1.644e-003	4266	1.948e-003	5297	1.035e-003	4266	1.065e-013	7190
11	4.190e-005	7587	1.298e-004	6671	1.513e-004	0247	8.931e-004	7446	2.036e-014	0247
12	4.106e-005	7567	1.268e-004	7245	1.500e-004	4917	8.830e-004	7723	3.863e-015	7237
13	4.105e-005	7190	1.242e-004	0286	2.977e-005	7041	4.423e-004	6006	2.221e-016	6006
14	4.057e-005	4917	1.232e-004	7446	2.974e-005	7587	4.312e-004	0286	7.242e-017	7457
15	1.245e-005	7723	3.855e-005	4424	2.893e-005	7457	2.677e-005	7190	6.142e-017	7587
16	1.235e-005	7237	3.814e-005	4033	2.841e-005	7254	2.672e-005	7567	2.985e-017	7446
17	1.227e-005	4424	3.796e-005	0068	2.810e-005	4266	2.638e-005	4424	1.621e-017	0286
18	1.221e-005	7254	3.779e-005	7587	2.709e-005	7245	2.521e-005	4917	1.334e-017	7482
19	1.153e-005	7457	3.762e-005	4917	2.702e-005	0286	2.517e-005	7254	1.093e-017	5297
20	1.149e-005	7446	3.613e-005	1640	2.661e-005	4033	2.501e-005	6295	5.770e-018	4917
21	1.121e-005	0286	3.564e-005	7254	2.606e-005	4424	2.499e-005	7457	8.606e-019	7567
22	1.090e-005	6053	3.321e-005	7482	2.554e-005	6295	2.448e-005	0247	3.250e-019	4424
23	1.065e-005	7335	3.299e-005	5297	2.521e-005	7335	2.242e-005	7335	3.111e-019	7254
24	1.050e-005	7519	3.262e-005	6295	2.511e-005	7481	2.228e-005	6053	2.073e-019	6295
25	1.024e-005	0068	3.252e-005	6053	2.497e-005	1640	2.075e-005	7482	1.857e-019	0068
26	1.007e-005	7566	3.216e-005	6006	2.480e-005	7567	2.042e-005	7566	1.663e-019	4033
27	9.951e-006	6295	3.164e-005	7567	2.432e-005	7566	1.984e-005	7519	1.662e-019	7335
28	9.766e-006	7482	3.146e-005	7237	2.417e-005	0068	1.977e-005	0068	1.637e-019	1640
29	9.588e-006	7481	3.063e-005	7566	2.363e-005	7446	1.901e-005	7481	1.586e-019	6053
30	9.566e-006	1640	2.979e-005	7481	2.273e-005	6006	1.898e-005	1640	1.531e-019	7566
31	9.537e-006	5297	2.807e-005	7519	2.009e-005	6053	1.805e-005	4033	1.363e-019	7481
32	9.083e-006	4033	2.762e-005	7335	1.921e-005	7519	1.785e-005	5297	1.123e-019	7519

**Πρότυπο 7519**  
(Απεδόθη στο λιθοξόο 2 )

α/α	Πιθαν/α Α	Κατ/ξη Α	Πιθαν/α Σ	Κατ/ξη Σ	Πιθαν/α Μ	Κατ/ξη Μ	Πιθαν/α Ο	Κατ/ξη Ο	Συνολική Πιθαν/α	Συνολική Κατ/ξη
1	1.317e-002	7519	9.900e-003	7519	1.159e-002	7519	2.825e-002	7519	4.269e-008	7519
2	1.138e-002	6006	9.861e-003	7587	5.972e-003	5297	2.480e-002	6006	5.668e-009	6053
3	1.097e-002	7567	9.090e-003	6053	5.827e-003	6053	2.214e-002	7567	5.294e-009	6295
4	9.708e-003	5297	8.904e-003	4424	5.582e-003	6295	1.763e-002	5297	3.191e-009	5297
5	8.813e-003	7566	8.643e-003	7566	5.478e-003	7587	1.486e-002	7566	3.132e-009	6006
6	8.532e-003	0247	8.635e-003	6295	5.370e-003	7482	1.426e-002	6295	2.661e-009	7567
7	8.508e-003	6053	8.568e-003	6006	1.311e-003	7567	1.258e-002	6053	1.458e-009	7566
8	8.480e-003	0068	8.361e-003	7567	1.296e-003	6006	1.170e-002	0247	1.519e-011	6671
9	8.158e-003	6671	7.961e-003	7446	1.289e-003	7566	3.109e-003	6671	3.123e-013	7587
10	7.704e-003	6295	3.124e-003	5297	1.279e-003	6671	2.955e-003	7405	3.357e-014	0247
11	7.522e-003	7400	4.776e-004	7041	1.002e-004	4424	2.933e-003	7400	1.812e-014	7482
12	7.458e-003	7405	4.758e-004	4917	9.887e-005	0286	3.672e-004	0068	7.344e-015	7400
13	1.474e-003	7542	4.747e-004	7482	1.967e-005	7542	3.615e-004	1640	7.124e-015	7405
14	1.455e-003	0286	4.687e-004	6671	1.950e-005	6124	3.529e-004	0286	5.438e-015	4424
15	1.425e-003	7245	2.308e-005	7723	1.935e-005	7245	2.209e-005	7482	1.164e-015	0068
16	1.406e-003	4917	2.216e-005	7190	1.930e-005	4917	2.209e-005	7542	9.563e-016	0286
17	1.365e-003	6124	2.206e-005	7237	1.906e-005	7335	2.172e-005	6124	2.616e-016	4917
18	1.325e-003	4266	2.174e-005	7457	1.889e-005	7400	2.060e-005	7245	4.809e-017	7446
19	3.271e-004	7723	2.090e-005	0068	1.884e-005	7237	2.028e-005	4917	4.529e-017	7041
20	3.247e-004	1640	2.053e-005	7542	1.846e-005	0247	1.884e-005	7723	3.393e-017	1640
21	3.241e-004	4424	1.948e-005	4033	1.790e-005	4266	1.882e-005	4424	1.314e-017	7542
22	3.220e-004	7482	1.896e-005	4266	1.789e-005	0068	1.814e-005	7587	1.072e-017	7245
23	3.189e-004	7587	1.894e-005	7405	1.777e-005	7446	1.798e-005	4266	9.956e-018	6124
24	3.149e-004	7041	1.889e-005	7245	1.764e-005	7190	1.765e-005	7446	8.083e-018	4266
25	3.139e-004	7457	1.884e-005	0286	1.738e-005	7041	1.762e-005	7457	2.436e-018	7723
26	1.927e-005	7446	1.823e-005	0247	1.713e-005	7723	1.737e-005	7481	1.910e-018	7457
27	1.900e-005	7190	1.764e-005	7400	1.713e-005	7481	1.734e-005	7041	1.334e-019	7237
28	1.860e-005	7237	1.753e-005	7254	1.709e-005	7254	1.727e-005	7237	1.257e-019	7190
29	1.784e-005	7481	1.723e-005	6124	1.708e-005	7405	1.693e-005	7190	9.624e-020	4033
30	1.726e-005	4033	1.714e-005	1640	1.705e-005	4033	1.679e-005	4033	7.738e-020	7481
31	1.655e-005	7254	1.602e-005	7335	1.687e-005	1640	1.368e-005	7254	6.781e-020	7254
32	1.476e-005	7335	1.458e-005	7481	1.589e-005	7457	1.300e-005	7335	5.856e-020	7335

**Πρότυπο 7190**  
(Απεδότη στο λιθοξόο 3 )

α/ α	Πιθαν/α Α	Κατ/ξη Α	Πιθαν/α Σ	Κατ/ξη Σ	Πιθαν/α Μ	Κατ/ξη Μ	Πιθαν/α Ο	Κατ/ξη Ο	Συνολική Πιθαν/α	Συνολική Κατ/ξη
1	1.447e-002	7190	2.019e-002	7190	1.678e-002	7190	1.625e-002	7190	7.966e-008	7190
2	1.072e-002	7723	1.361e-002	7041	9.910e-003	7400	1.578e-002	7041	1.772e-008	7041
3	1.014e-002	7041	1.179e-002	7567	9.732e-003	7237	1.162e-002	4033	4.934e-010	7587
4	9.742e-003	7457	1.139e-002	7400	8.987e-003	7542	9.727e-003	0286	1.644e-010	7723
5	9.683e-003	7587	1.138e-002	7723	8.581e-003	7723	9.451e-003	0068	2.905e-011	7400
6	9.473e-003	7245	1.112e-002	7587	8.139e-003	7041	2.291e-003	7587	1.069e-011	6671
7	9.461e-003	7446	1.105e-002	7405	2.002e-003	7587	2.215e-003	7245	1.608e-012	7245
8	5.854e-003	4266	1.092e-002	4424	1.995e-003	7457	2.171e-003	1640	7.913e-013	7457
9	5.710e-003	4424	1.072e-002	6671	1.946e-003	6671	1.368e-003	7457	5.257e-013	4424
10	3.563e-003	7254	6.752e-003	1640	1.884e-003	7482	1.334e-003	7254	3.381e-013	4266
11	3.381e-003	6671	4.055e-003	7446	4.649e-004	0247	1.301e-003	4266	2.980e-013	7446
12	3.374e-003	7542	4.027e-003	4033	4.614e-004	5297	5.167e-004	7567	1.936e-013	7254
13	1.689e-003	7400	1.271e-003	0247	7.201e-005	0286	5.129e-004	5297	1.335e-013	7542
14	8.357e-004	7237	1.264e-003	7245	7.181e-005	4917	5.125e-004	6053	1.269e-013	7567
15	8.192e-004	7335	6.278e-004	4266	7.073e-005	4266	5.108e-004	0247	1.158e-013	0247
16	8.133e-004	7405	6.179e-004	7254	6.777e-005	7405	1.572e-004	7723	5.467e-014	4033
17	7.825e-004	4917	6.155e-004	0068	6.680e-005	6124	1.545e-004	6006	3.213e-014	7237
18	7.795e-004	6124	1.212e-004	6006	6.597e-005	7254	1.532e-004	6124	2.282e-014	7405
19	3.839e-004	0247	1.209e-004	7542	6.559e-005	1640	1.525e-004	7400	1.701e-014	1640
20	3.268e-004	0286	1.205e-004	6124	6.532e-005	7567	1.516e-004	6671	6.827e-015	0286
21	3.192e-004	7567	3.000e-005	6295	6.442e-005	4033	1.510e-004	6295	6.516e-015	0068
22	3.119e-004	6006	2.983e-005	0286	6.274e-005	7335	1.404e-004	7446	9.610e-016	6124
23	3.038e-004	6053	2.977e-005	7457	6.274e-005	0068	1.361e-004	4424	3.365e-016	6006
24	2.993e-004	7566	2.974e-005	6053	6.196e-005	4424	1.341e-004	7237	2.317e-016	6053
25	2.984e-004	6295	2.952e-005	5297	6.181e-005	7481	4.175e-005	7482	1.154e-016	5297
26	1.814e-005	4033	2.947e-005	7237	6.068e-005	7245	3.880e-005	4917	7.728e-017	6295
27	1.787e-005	0068	2.882e-005	4917	6.031e-005	7566	3.780e-005	7519	6.281e-017	4917
28	1.771e-005	1640	2.673e-005	7482	5.764e-005	6006	3.749e-005	7405	3.548e-017	7482
29	1.750e-005	7519	2.622e-005	7566	5.718e-005	6295	3.645e-005	7542	3.104e-017	7335
30	1.739e-005	7481	2.462e-005	7519	5.535e-005	7446	3.313e-005	7566	1.567e-017	7566
31	1.688e-005	7482	2.354e-005	7481	5.005e-005	6053	3.189e-005	7481	8.064e-019	7481
32	1.653e-005	5297	2.148e-005	7335	4.050e-005	7519	2.815e-005	7335	6.595e-019	7519



**Πρότυπο 4266**  
(Απεδότη στο λιθοξόο 4 )

α/ α	Πιθαν/α Α	Κατ/ξη Α	Πιθαν/α Σ	Κατ/ξη Σ	Πιθαν/α Μ	Κατ/ξη Μ	Πιθαν/α Ο	Κατ/ξη Ο	Συνολική Πιθαν/α	Συνολική Κατ/ξη
1	1.626e-002	4266	1.386e-002	4266	1.088e-002	4266	2.954e-002	4266	7.243e-008	4266
2	1.259e-002	7587	9.813e-003	4424	9.868e-003	7237	2.605e-002	4424	2.817e-008	7587
3	1.220e-002	4424	9.496e-003	7587	9.523e-003	7587	2.521e-002	7542	2.002e-008	4424
4	1.120e-002	7446	9.389e-003	7190	8.632e-003	7542	2.475e-002	7587	2.835e-010	7542
5	1.115e-002	6671	9.372e-003	1640	8.374e-003	7457	1.091e-002	7446	2.428e-010	7457
6	8.742e-003	7041	7.368e-003	0286	8.335e-003	7567	6.213e-003	7245	2.261e-010	7041
7	8.686e-003	7245	7.267e-003	7245	6.460e-003	6006	4.760e-003	7041	1.841e-010	7245
8	6.727e-003	7723	7.111e-003	7041	6.419e-003	4424	3.730e-003	7723	2.254e-011	7237
9	6.691e-003	4917	6.992e-003	4033	4.044e-003	4917	2.862e-003	7254	1.545e-011	7190
10	4.154e-003	7237	5.500e-003	7457	4.006e-003	7190	2.670e-003	7457	1.815e-012	7723
11	4.037e-003	7542	3.467e-003	0068	1.998e-003	7405	8.430e-004	7237	5.620e-013	7254
12	3.974e-003	6124	3.362e-003	7254	7.903e-004	7482	7.367e-004	6124	2.751e-013	7446
13	1.975e-003	7457	1.061e-003	0247	7.841e-004	1640	7.127e-004	7400	2.648e-013	7400
14	9.889e-004	7254	1.045e-003	6053	7.696e-004	6124	7.038e-004	7405	2.346e-013	4917
15	9.733e-004	7405	8.244e-004	7400	7.645e-004	7041	5.541e-004	4917	1.922e-013	0286
16	9.597e-004	7190	6.526e-004	7237	4.737e-004	7400	5.435e-004	7335	1.318e-013	7405
17	9.518e-004	7400	5.142e-004	7481	4.723e-004	0068	4.281e-004	7190	8.088e-014	6124
18	9.364e-004	7335	3.228e-004	7542	4.697e-004	7245	2.140e-004	0286	5.035e-014	6671
19	4.344e-004	0286	1.557e-004	7723	4.649e-004	7723	1.651e-004	6671	1.772e-014	1640
20	4.308e-004	6053	9.639e-005	7405	4.640e-004	6671	9.460e-005	0247	9.115e-015	6053
21	4.277e-004	0247	6.026e-005	5297	2.808e-004	0286	7.375e-005	6053	5.946e-015	6006
22	4.242e-004	6295	5.897e-005	6671	2.781e-004	7481	7.247e-005	6295	3.963e-015	0068
23	2.436e-004	7566	3.709e-005	7482	2.759e-004	6295	7.066e-005	7566	3.070e-015	4033
24	9.699e-004	6006	3.639e-005	7446	2.747e-004	6053	6.943e-005	7567	2.866e-015	0247
25	1.532e-004	7567	3.595e-005	6124	2.719e-004	4033	5.486e-005	6006	2.481e-015	7567
26	4.498e-005	0068	2.799e-005	7567	6.680e-005	0247	5.406e-005	1640	8.755e-016	7335
27	4.462e-005	1640	2.795e-005	6295	6.553e-005	7566	5.383e-005	0068	3.044e-016	7481
28	4.283e-005	7481	2.753e-005	7335	6.294e-005	5297	4.972e-005	7481	2.369e-016	6295
29	4.104e-005	4033	2.182e-005	7566	6.254e-005	7335	3.937e-005	4033	4.391e-017	7482
30	4.068e-005	7519	1.732e-005	6006	6.191e-005	7446	3.822e-005	7482	2.460e-017	7566
31	3.924e-005	5297	1.566e-005	4917	5.910e-005	7254	3.381e-005	7519	3.869e-018	5297
32	3.921e-005	7482	9.687e-006	7519	5.114e-005	7519	2.600e-005	5297	6.813e-019	7519

**Πίνακας Συγκεντρωτικών Αποτελεσμάτων**

<b>Χέρι 1</b>	<b>Χέρι 2</b>	<b>Χέρι 3</b>	<b>Χέρι 4</b>
<b>7400</b>	<b>7519</b>	<b>7190</b>	<b>4266</b>
7405	6053	7041	7587
7542	6295		4424
6124	5297		
	6006		
	7567		
	7566		

### 5.3.2 Από σύγχρονους γραφείς.

Στους πίνακες που ακολουθούν παρουσιάζονται αποτελέσματα των συγκρίσεων των επιγραφών που βρίσκονται στην πρώτη στήλη του πίνακα (πρότυπες επιγραφές) με αυτές που βρίσκονται στην πρώτη γραμμή του πίνακα για το γράμμα “d”.

Η διαγώνιος αντιπροσωπεύει το βαθμό ομοιότητας των υλοποιήσεων του γράμματος “d” μέσα σε κάθε επιγραφή. Μας δείχνει πόσο «συνεπής» είναι ο γραφέας κατά την υλοποίηση του συγκεκριμένου γράμματος της αλφαβήτου. Όσο μεγαλύτερος ο αριθμός (μέσος όρος των αποτελεσμάτων τομή/ένωση για όλες τις υλοποιήσεις του “d” που συγκρίθηκαν), τόσο πιο συνεπής ο γραφέας (ή τόσο πιο καλλιγράφος!). Το ζητούμενο εδώ είναι ο μέσος όρος του κάθε γραφέα να απέχει όσο το δυνατόν περισσότερο από το μέσο όρο που προκύπτει από τη σύγκριση των υλοποιήσεων του γράμματος “d” της επιγραφής του γραφέα που μας ενδιαφέρει με τις επιγραφές των υπολοίπων γραφέων (δηλαδή η τιμή που προκύπτει από τη σύγκριση π.χ. της 99013\_01\_01 με την 99015\_01\_01 να βρίσκεται όσο το δυνατόν πιο «δεξιά» ή «αριστερά» στην κατανομή που έχει σαν μέση τιμή την τιμή που προκύπτει από τη σύγκριση της κάθε επιγραφής με τον εαυτό της).

Πρότυπη/ Συγκρινόμενη	99013_01_01	99014_01_00	99014_01_01	99015_01_01	99019_01_00
99013_01_01	0,3197	0,3441	0,3485	0,2811	0,2941
99014_01_00	0,3638	0,5041	0,5664	0,3230	0,2830
99014_01_01	0,3701	0,5470	0,5989	0,3252	0,2716
99015_01_01	0,2761	0,3281	0,3112	0,3519	0,2387
99019_01_00	0,3067	0,2761	0,2783	0,2424	0,3176

Από τον πίνακα αυτό παρατηρούμε ότι δεν μπορούμε να εξάγουμε ασφαλές συμπέρασμα για το αν οι επιγραφές 99013\_01\_01 και οι 99014\_01\_00, 99014\_01\_01, 99015\_01\_01, 99019\_01\_00 δεν έχουν γραφτεί από το ίδιο χέρι. Οι τιμές που προκύπτουν από τις συγκρίσεις υποδεικνύουν ότι οι μέσοι όροι που προκύπτουν από τη σύγκριση της 99013\_01\_01 με τις υπόλοιπες επιγραφές βρίσκονται κοντά στο μέσο όρο της κατανομής (που εδώ είναι ο 0,3197). Επομένως, δεν μπορούμε να βασιστούμε στο γράμμα “d” για να πούμε με βεβαιότητα ότι οι επιγραφές αυτές δεν έχουν γραφτεί από το ίδιο χέρι (αυτό διότι ζητάμε απορριπτικό κριτήριο!). Για τον ίδιο λόγο δεν μπορούμε να πούμε με βεβαιότητα ότι οι 99019\_01\_00 και 99013\_01\_01 δεν έχουν γραφτεί από το ίδιο χέρι. Αντιθέτως, όλοι οι υπόλοιποι γραφείς διαχωρίζονται με βάση το γράμμα “d”.

Στον ακόλουθο πίνακα παρουσιάζονται ανάλογα αποτελέσματα για τις ίδιες επιγραφές, αλλά για το γράμμα “e”. Εδώ παρατηρούμε ότι οι γραφείς που δεν διαχωρίζονταν προηγουμένως με βάση το γράμμα “d”, με βάση το γράμμα “e” διαχωρίζονται πλήρως. Αντιθέτως, η επιγραφή 99015\_01\_01 δεν διαχωρίζεται με βεβαιότητα από τις υπόλοιπες επιγραφές που παρουσιάζονται, όσον αφορά το γράμμα “e”, ενώ, ο διαχωρισμός της από τις υπόλοιπες ήταν ξεκάθαρος για το γράμμα “d”, όπως συμπεραίνουμε από τον ανωτέρω πίνακα.

Πρότυπη/ Συγκρινόμενη	99013_01_01	99014_01_00	99014_01_01	99015_01_01	99019_01_00	99019_01_01
99013_01_01	<b>0,4404</b>	<b>0,3208</b>	<b>0,326</b>	<b>0,3367</b>	<b>0,3726</b>	<b>0,3719</b>
99014_01_00	<b>0,3397</b>	<b>0,5831</b>	<b>0,6378</b>	<b>0,3847</b>	<b>0,3502</b>	<b>0,323</b>
99014_01_01	<b>0,3435</b>	<b>0,6348</b>	<b>0,6582</b>	<b>0,4092</b>	<b>0,3421</b>	<b>0,3103</b>
99015_01_01	<b>0,3483</b>	<b>0,3445</b>	<b>0,358</b>	<b>0,3919</b>	<b>0,3651</b>	<b>0,3624</b>
99019_01_00	<b>0,3794</b>	<b>0,3384</b>	<b>0,3253</b>	<b>0,3685</b>	<b>0,4243</b>	<b>0,4368</b>
99019_01_01	<b>0,3746</b>	<b>0,313</b>	<b>0,2963</b>	<b>0,3621</b>	<b>0,4349</b>	<b>0,428</b>

Παρακάτω παρουσιάζονται πίνακες που δείχνουν τη μέση τιμή του συνολικού ταιριάσματος (μέση τιμή των τιμών των εμβადών τομής προς ένωση για όλες τις υλοποιήσεις ενός γράμματος για κάθε ζεύγος επιγραφών που συγκρίνονται).

Εδώ παρουσιάζονται αποτελέσματα της σύγκρισης των επιγραφών 99014\_01\_00 και 99014\_01\_01 (πρότυπες) με όλες τις υπόλοιπες επιγραφές του δοσμένου σετ για το γράμμα “e”.

Παρατηρούμε ότι ο γραφέας 99014 και στις δύο του επιγραφές διατηρεί υψηλό βαθμό συνέπειας στην υλοποίηση του γράμματος “e”. Θα λέγαμε ότι αυτό τον καθιστά εξαιρετικά «καλλιγράφο» ή πολύ συνεπή στις υλοποιήσεις του για το συγκεκριμένο γράμμα. Για το λόγο αυτό, η αντίστοιχη μέση τιμή είναι πολύ υψηλή όταν γίνεται σύγκριση των υλοποιήσεων του συγκεκριμένου γράμματος στις επιγραφές που έχει γράψει ο ίδιος, ενώ παρατηρούνται πολύ μικρές αντίστοιχες τιμές όταν συγκρίνονται οι υλοποιήσεις των “e” στις επιγραφές του 99014 και οι υλοποιήσεις του ίδιου γράμματος σε επιγραφές άλλων γραφέων.

	9906_ 01_00	9906_ 01_01	99012_ 01_00	99013_ 01_00	99013_ 01_01	99014_ 01_00	99014_ 01_01	99015_ 01_00	99015_ 01_01	99019_ 01_00	99019_ 01_01
99014_ 01_00	0,3573	0,3644	0,3615	0,3161	0,3397	0,5831	0,6378	0,3435	0,3847	0,3502	0,3230
99014_ 01_01	0,3563	0,3794	0,3688	0,3084	0,3435	0,6348	0,6582	0,3577	0,4092	0,3421	0,3103

	99021_ 01_00	99021_ 01_01	99022_ 01_00	99022_ 01_01	99023_ 01_00	99023_ 01_01	99025_ 01_00	99025_ 01_01	99027_ 01_00	99027_ 01_01	99030_ 01_00
99014_ 01_00	0,3743	0,4138	0,3736	0,3789	0,3537	0,3433	0,3433	0,3415	0,3696	0,3418	0,3674
99014_ 01_01	0,3678	0,3777	0,3951	0,4038	0,3508	0,3312	0,3491	0,3561	0,3702	0,3479	0,3573

01_01											
-------	--	--	--	--	--	--	--	--	--	--	--

	99030_ 01_01	99031_ 01_00	99031_ 01_01	99032_ 01_00	99032_ 01_01	99033_ 01_00	99033_ 01_01	99035_ 01_00	99035_ 01_01	99039_ 01_00	99039_ 01_01
99014_ 01_00	0,3850	0,3786	0,3532	0,2578	0,2575	0,5401	0,5381	0,3297	0,2852	0,3555	0,3291
99014_ 01_01	0,3892	0,3778	0,3610	0,2623	0,2617	0,5449	0,5333	0,3455	0,2935	0,3431	0,3241

	99040_ 01_00	99040_ 01_01	99045_ 01_00	99045_ 01_01	99046_ 01_00	99046_ 01_01	99047_ 01_00	99047_ 01_01	99065_ 01_00	99065_ 01_01
99014_ 01_00	0,3389	0,3449	0,3063	0,3094	0,4173	0,4417	0,3667	0,3641	0,3796	0,3677
99014_ 01_01	0,3438	0,3509	0,2976	0,2981	0,4091	0,4403	0,3754	0,3742	0,3753	0,3661

Αντίστοιχα συμπεράσματα προκύπτουν από τη μελέτη των κατωτέρω πινάκων για το γράμμα d και για τις επιγραφές 99033\_01\_00 και 99065\_01\_01 (πρότυπες επιγραφές).

	9906_ 01_00	9906_ 01_01	99012_ 01_00	99012_ 01_01	99013_ 01_01	99014_ 01_00	99014_ 01_01	99015_ 01_01	99019_ 01_00	99022_ 01_01	99023_ 01_00
99033_ 01_00	0,2890	0,3187	0,3962	0,4104	0,3430	0,5510	0,5813	0,3414	0,2863	0,3670	0,3718
99065_ 01_01	0,2738	0,2887	0,4215	0,3898	0,3743	0,422	0,4154	0,3809	0,3224	0,3296	0,4166

	99023_ 01_01	99025_ 01_00	99025_ 01_01	99027_ 01_00	99027_ 01_01	99030_ 01_00	99030_ 01_01	99031_ 01_00	99031_ 01_01	99032_ 01_01	99033_ 01_00
99033_ 01_00	0,3969	0,3668	0,3642	0,3802	0,4120	0,3979	0,3800	0,4867	0,3488	0,2616	0,6746
99065_ 01_01	0,3905	0,3397	0,3625	0,3663	0,4005	0,3688	0,398	0,4013	0,3238	0,3279	0,4841

	99033_ 01_01	99035_ 01_00	99035_ 01_01	99039_ 01_00	99039_ 01_01	99040_ 01_00	99040_ 01_01	99045_ 01_00	99046_ 01_00	99046_ 01_01	99047_ 01_00
99033_ 01_00	0,7231	0,4155	0,3043	0,3350	0,3096	0,3410	0,3564	0,2487	0,3457	0,3986	0,3585
99065_ 01_01	0,501	0,493	0,3834	0,3366	0,3239	0,3542	0,3853	0,275	0,3724	0,4189	0,3834

	<b>99047_01_01</b>	<b>99065_01_00</b>	<b>99065_01_01</b>
<b>99033_01_00</b>	<b>0,3844</b>	<b>0,5149</b>	<b>0,4889</b>
<b>99065_01_01</b>	<b>0,3921</b>	<b>0,5742</b>	<b>0,5899</b>



## 6. ΣΥΜΠΕΡΑΣΜΑΤΑ

Στην παρούσα διπλωματική εφαρμόστηκε η νέα κλάση αλγορίθμων ταυτοποίησης γραφέων , η οποία έχει προδημοσιευτεί στα [1] [2] [3] [4] [5] . Δόθηκε ιδιαίτερη έμφαση στην αυτόματη αναγνώριση του γραφέα σύγχρονων αγγλικών κειμένων που βρίσκονται στη σχετική διεθνή βάση δεδομένων IM-UNIPEN. Συγκεκριμένα, μας δόθηκε ένα υποσύνολο του περιεχομένου της βάσης που περιείχε φωτογραφίες από χειρόγραφα κείμενα στα Αγγλικά και ειδικότερα φωτογραφίες από απομονωμένες λέξεις με μαύρα γράμματα σε λευκό φόντο. Φροντίσαμε το δείγμα να είναι πολύ μεγάλο της τάξης μερικών χιλιάδων συμβολογραμμάτων για κάθε γράμμα και για μερικές δεκάδες γραφέων έτσι ώστε τα στατιστικά αποτελέσματα να προσφέρουν ένα ασφαλές συμπέρασμα για την εργασία μας.

Η εφαρμογή της μεθόδου οδήγησε σε αυτόματη ταυτοποίηση του σύγχρονου γραφέα πρακτικά με απόλυτη ασφάλεια.

## **ΒΙΒΛΙΟΓΡΑΦΙΑ**

- [1] “Ανάπτυξη προηγμένων μεθόδων αναγνώρισης προτύπων και μαθηματική θεμελιώση για τον εντοπισμό του τρόπου δημιουργίας και της ταυτότητας του δημιουργού σημαντικών αρχαιολογικών ευρημάτων”, Διδακτορική διατριβή, Ρουσόπουλος Παναγιώτης, 2010
- [2] “Automatic writer identification of ancient Greek inscriptions”, IEEE Transactions on Pattern Analysis and Machine Intelligence, M. Panagopoulos, C. Papaodysseus, P. Rousopoulos, D. Dafi and S. Tracy , vol. 31, no. 8, pp. 1404-1414, July 2009.
- [3] “Image and Pattern Analysis of 1850 B.C. Wall Paintings and Recognition”, Papaodysseus C., Exarhos M., Panagopoulos M. , Rousopoulos P. , Triantafillou C. , Panagopoulos T., IEEE Transactions on Systems, Man and Cybernetics, 2008
- [4] 17th International Conference on Digital Signal Processing, Corfu, Greece
- [5] 18th International Conference on Digital Signal Processing, Santorini, Greece
- [6] Archaeometry
- [7] S. Tracy and C. Papaodysseus, “The study of hands on Greek inscriptions: The need for a digital approach”, American Journal of Archaeology, pp. 99-102, Jan. 2009.
- [8] S.V.Tracy, C. Papaodysseus, P. Rousopoulos, M. Panagopoulos, D. Fragoulis, D. Dafi and Th. Panagopoulos “Identifying hands on ancient Athenian inscriptions: First steps towards a digital approach”, Archaeometry vol. 49, no. 4, pp. 749–764, 2007.
- [9] M. Panagopoulos, C. Papaodysseus, P. Rousopoulos, D. Dafi and S. Tracy, “Automatic writer identification of ancient Greek inscriptions”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, no. 8, pp. 1404-1414, July 2009.
- [9] G. E. Kokolakis, “Bayesian classification and classification performance for independent distributions”, IEEE Transactions on Information Theory, vol. 27, pp. 419–21, 1981.
- [10] 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.
- [11] E. N. Zois and V. Anastassopoulos, “Morphological waveform coding for writer identification”, Pattern Recognition, vol. 33, pp. 385-398, 2000.
- [12] H. Said, T. Tan and K. Baker, “Personal identification based on handwriting”, Pattern Recognition, vol. 33, no. 1, pp. 149-160, Jan. 2000.

- [13] Y. Zhu, T. Tan and Y. Wang, "Biometric personal identification based on handwriting", in Proceedings of the 15th International Conference on Pattern Recognition, IEEE, vol. 2, 2000, pp. 797-800.
- [14] U.-V. Marti, R. Messerli, and H. Bunke, "Writer identification using text line based features", in Proceedings of the 6th International Conference on Document Analysis and Recognition, Sept. 2001, pp. 101-105.
- [15] A. Bensefia, T. Paquet, and L. Heutte, "Handwriting analysis for writer verification", in Proceedings of the 9th International Workshop Frontiers in Handwriting Recognition, Oct. 2004, pp. 196-201.
- [16] G. Leedham and S. Chachra, "Writer identification using innovative binarised features of handwritten numerals", in Proceedings of the 7th International Conference on Document Analysis and Recognition, 2003.
- [17] A. Schlapbach and H. Bunke, "A writer identification and verification system using HMM based recognizers", Pattern Analysis and Applications, vol. 10, no. 1, Feb. 2007.
- [18] Z. He, X. You, Y. Yan Tang, "Writer identification of Chinese handwriting documents using hidden Markov tree model", Pattern Recognition, vol. 41, no. 4, pp. 1295-1307, 2008.
- [19] A. Schlapbach and H. Bunke, "Off-line handwriting identification using HMM based recognizers", in ICPR Proceedings of the 17th International Conference on Pattern Recognition, vol. 2, 2004, pp. 654-658.
- [20] Z. He, B. Fang, J. Du, Y. Yan Tang, and X. You, "A novel method for off-line handwriting based writer identification", in Proceedings of the 8th International Conference on Document Analysis and Recognition, 2005.
- [21] A. Bensefia, T. Paquet, and L. Heutte, "A writer identification and verification system", Pattern Recognition Letters, vol. 26, pp. 2080-2092, 2005.
- [22] V. Pervouchine and G. Leedham, "Extraction and analysis of forensic document examiner features used for writer identification", Pattern Recognition, vol. 40, 2007.
- [23] T. Pitak, T. Matsuura, "On-line writer recognition for Thai based on velocity of barycenter of pen-point movement", in International Conference on Image Processing, vol. 2, 2004, pp. 889-892.
- [24] S.-H. Cha and S. N. Srihari, "Multiple feature integration for writer verification", in Proceedings on 7th International Workshop Frontiers in Handwriting Recognition, 2000, pp. 333-342.
- [25] B. Zhang, S. N. Srihari and S. Lee, "Individuality of handwritten characters", in Proceedings of the 7th International Conference on Document Analysis and Recognition, Aug. 2003.

- [26] M. Bulacu and L. Schomaker, “Text-independent writer identification and verification using textural and allographic features”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, no. 4, Apr. 2007. AUTHOR: TITLE 35
- [27] L. Schomaker and M. Bulacu, “Automatic writer identification using connected-component contours and edge-based features of uppercase western script”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 26, no. 6, pp. 787-798, June 2004.
- [28] R. Niels, F. Gootjen and L. Vuurpijl, “Writer identification through information retrieval: The allograph weight vector”, in International Conference on Frontiers in Handwriting Recognition, 2008, pp. 481–486.