



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ  
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΕΠΙΚΟΙΝΩΝΙΩΝ, ΗΛΕΚΤΡΟΝΙΚΗΣ ΚΑΙ ΣΥΣΤΗΜΑΤΩΝ  
ΠΛΗΡΟΦΟΡΙΚΗΣ

**ΜΕΛΕΤΗ ΚΑΙ ΕΦΑΡΜΟΓΗ ΝΕΩΝ ΜΕΘΟΔΩΝ  
ΤΑΥΤΟΠΟΙΗΣΗΣ ΓΡΑΦΕΑ ΠΑΛΑΙΩΝ ΚΕΙΜΕΝΩΝ**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Θεόδωρος Ε. Λαπίδης

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

Αθήνα, Απρίλιος 2016





ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ  
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΕΠΙΚΟΙΝΩΝΙΩΝ, ΗΛΕΚΤΡΟΝΙΚΗΣ ΚΑΙ ΣΥΣΤΗΜΑΤΩΝ  
ΠΛΗΡΟΦΟΡΙΚΗΣ

## ΜΕΛΕΤΗ ΚΑΙ ΕΦΑΡΜΟΓΗ ΝΕΩΝ ΜΕΘΟΔΩΝ ΤΑΥΤΟΠΟΙΗΣΗΣ ΓΡΑΦΕΑ ΠΑΛΑΙΩΝ ΚΕΙΜΕΝΩΝ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Θεόδωρος Ε. Λαπίδης

**Επιβλέπων :** Κωνσταντίνος Παπαοδυσσεύς

Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 4<sup>η</sup> Απριλίου 2016.

.....  
Κωνσταντίνος Παπαοδυσσεύς  
Καθηγητής Ε.Μ.Π.

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

.....  
Μιχαήλ Θεολόγου  
Καθηγητής Ε.Μ.Π.



.....  
Θεόδωρος Ε. Λαπίδης

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

Copyright © Θεόδωρος Λαπίδης, 2016

Copyright © Κωνσταντίνος Παπαοδυσσεύς, 2016

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 τον συγγραφέα.

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



## Περίληψη

Ένα πολύ σημαντικό αντικείμενο της παλαιογραφίας είναι η εύρεση του γραφέα που έχει γράψει ένα κείμενο. Ένας τρόπος να γίνει αυτό είναι μέσω της εξέτασης του γραφικού χαρακτήρα του γραφέα και της σύγκρισης του με γραφικούς χαρακτήρες ήδη γνωστών γραφέων ούτως ώστε να ταυτισθεί με κάποιον από αυτούς. Αυτό γινόταν μέχρι πρόσφατα με το παραδοσιακό τρόπο, δηλαδή έλεγχος του γραφικού χαρακτήρα με το μάτι και σύγκριση αυτού με τους υπόλοιπους.

Η τεχνολογία εξέλιξη μπορεί να μας βοηθήσει σε ένα μεγάλο βαθμό στο να γίνεται αυτή η αναγνώριση πολύ πιο γρήγορα και εύστοχα. Μπορούν να συγκριθούν πολύ μεγαλύτερος αριθμός γραφικών στυλ σε πολύ μικρότερο χρόνο. Στη συνέχεια μπορεί να γίνει και οπτική αναγνώριση για επιβεβαίωση του αποτελέσματος. Η τεχνική αυτή μπορεί να χρησιμοποιηθεί και για άλλους σκοπούς, όπως πχ. δικαστική έρευνα.

Υπάρχουν διάφορες μεθοδολογίες για την επίτευξη της παραπάνω τεχνικής. Μία από τις πιο σύγχρονες και ταυτόχρονα πιο ακριβείς είναι η σύγκριση των περιγραμμάτων του κάθε χαρακτήρα ξεχωριστά. Αυτό γίνεται εξάγοντας το περίγραμμα από όλες τις εμφανίσεις ενός συγκεκριμένου γράμματος. Στη συνέχεια γίνεται σύγκριση όλων αυτών των περιγραμμάτων, με όλα τα περιγράμματα από γνωστά σύνολα, ένα σύνολο για κάθε γραφέα, συγκρίνονται τα αποτελέσματα από τις συγκρίσεις και επιλέγεται το σύνολο το οποίο έχει τη μεγαλύτερη ομοιότητα με το εξεταζόμενο σύνολο. Τέλος, επιλέγεται ο γραφέας του συνόλου αυτού ως γραφέας του άγνωστου κειμένου. Η τεχνική αυτή ονομάζεται σύγκριση “όλα με όλα” και αναλύεται επακριβώς στο τελευταίο κεφάλαιο της εργασίας αυτής.

## Λέξεις-Κλειδιά

ταυτοποίηση, γραφείς, αναγνώριση, προτύπων, επεξεργασία, εικόνας, συγγραφείς, παλαιογραφία, περιγράμματα, χαρακτήρες





## **Abstract**

One of the most important parts of paleography is finding the writer of a text. This can be achieved through examining the handwriting of the writer, and then comparing it with the handwritings of already known writers. This was done till recently with the traditional way, examining each one of the handwritings with the vision and then comparing it with the others.

Technological evolution can help us, so that, this procedure can be done quicker and more accurate. It is possible to compare a greater number of handwritings in smaller time period. Later, the result that was produced, can be verified through visual examination. This technique can be used for other reason, like forensic reasons.

There are many different ways for do it. One of the most moderns and more precise of them is the comparison of character outlines one by one. It can be done through getting the outlines of all specific character in a text. After that all the outlines are compared with all the outlines from known groups of characters (each group is assigned to one known writer). Then the results are compared, and it is chosen the group which was the greater similarity with the group being examining. Finally, the writer of that group is chosen as the writer of the examined text. The technique is called "Comparison All-By-All" and it is analyzed in the last chapter of this thesis.

## **Keywords**

identification, scribe, pattern, recognition, image, processing, writer, paleography, outline, characters



## **Ευχαριστίες**

Με την εργασία αυτή, ολοκληρώνονται οι σπουδές μου στο Εθνικό Μετσόβιο Πολυτεχνείο. Θα ήθελα να ευχαριστήσω τον καθηγητή μου κ. Παπαοδυσσέα για όλα την ανάθεση της διπλωματικής και τις συμβουλές που μου παρείχε, όπως επίσης τον διδάκτωρ κ. Αραμπατζή Δημήτριο, ο οποίος με βοήθησε καθ' όλη την διάρκεια της εκπόνησης της διπλωματικής μου εργασίας.

Τέλος, θα ήθελα να ευχαριστήσω την οικογένεια μου, που τόσα χρόνια μου παρείχαν όλη την στήριξη και την βοήθεια που χρειαζόμουν ούτως ώστε να ολοκληρώσω τις σπουδές μου στο Εθνικό Μετσόβιο Πολυτεχνείο



## Περιεχόμενα

1 Εισαγωγή .....	13
2 Βιβλιογραφική Ανασκόπηση .....	16
2.1 Η αναγνώριση Γραφένων .....	18
2.2 Online Αναγνώριση Γραφέα .....	22
2.3 Η Βιομετρική .....	25
2.4 Βιομετρικά συστήματα .....	28
2.5 Γραφηματική Μέθοδος κωδικών .....	31
2.6 Παράγοντες που καθορίζουν την ατομικότητα του χειρογράφου .....	34
3 Ανάλυση Άρθρων .....	377
3.1 <i>New Mathematical and Algorithmic Schemes for Pattern Classification with Application to the Identification of Writer of Important Ancient Documents</i> .....	377
3.1.1 Θεωρητικό Υπόβαθρο .....	377
3.1.2 Σκοπός της Μελέτης .....	399
3.1.3 Περιγραφή της Μεθοδολογίας .....	39
3.1.4 Δυσκολίες στην Αναγνώριση .....	43
3.1.5 Συμπεράσματα της Μελέτης .....	44
3.2 <i>Identifying the Writer of Ancient Inscriptions and Byzantine Codices. A Novel Approach</i> .....	466
3.2.1 Θεωρητικό Υπόβαθρο της Μελέτης .....	488
3.2.2 Η προτεινόμενη Μεθοδολογία .....	52

3.2.3 Συμπεράσματα της Μελέτης.....	577
4 Έρευνα.....	666
5 Συμπεράσματα-Επίλογος .....	68
Βιβλιογραφία .....	70

## 1 Εισαγωγή

Το χειρόγραφο είναι ένα βασικό μέσο επικοινωνίας στον πολιτισμό μας, το οποίο έχει αναπτυχθεί και εξελιχθεί ιδιαίτερα με την πάροδο του χρόνου. Στο σχολείο, όλοι μαθαίνουν να γράφουν, σύμφωνα με ένα πρότυπο στυλ γραφής. Το γραπτό κείμενο όμως ποικίλλει ανάλογα με τη γεωγραφική θέση, τις χρονικές συνθήκες και τα πολιτιστικά και ιστορικά υπόβαθρα. Με το πέρασμα του χρόνου, οι άνθρωποι αναπτύσσουν ατομικά χαρακτηριστικά γραφής και ο γραφικός χαρακτήρας μπορεί να αρχίσει να αποκλίνει από το μαθέν στυλ. Αυτά τα μοναδικά χαρακτηριστικά χρησιμεύουν στη διάκριση του γραψίματος ενός ατόμου από ένα άλλο, ακόμη και αν οι δύο τύποι γραφής μοιάζουν, καθιστώντας δυνατή την αναγνώριση του γραφέα για τον οποίο κάποιος έχει ήδη δει ένα γραπτό κείμενο. Αυτή η αναγνώριση γραφέων χρησιμεύει ως μια πολύτιμη λύση για τους εξεταστές εγγράφων, τους παλαιογράφους, τους γραφολόγους και τους δικαστικούς γραφολόγους.

Σε αντίθεση με το ηλεκτρονικό ή το έντυπο κείμενο, το χειρόγραφο κείμενο φέρει πρόσθετες πληροφορίες σχετικά με την προσωπικότητα του ατόμου που το έχει γράψει. Ο Κινέζος φιλόσοφος Jo-Hau δήλωσε: «το χειρόγραφο μας δείχνει αλάνθαστα αν αυτό προέρχεται από ένα χυδαίο ή ένα ευγενές πνεύμα – πρόσωπο» [Olganova, 1960]. Στην πραγματικότητα, μεταξύ των εκφραστικών συμπεριφορών των ανθρώπων, το χειρόγραφο προσφέρει την πλουσιότερη πληροφόρηση σχετικά με τις φυσικές, διανοητικές και συναισθηματικές καταστάσεις του γραφέα. Κάθε γραπτή κίνηση ή νοητική επιλογή αποκαλύπτει ένα συγκεκριμένο χαρακτηριστικό της προσωπικότητας, ενώ και σε πρακτικό επίπεδο οι νευρο-μυϊκές τάσεις συσχετίζονται

με συγκεκριμένα παρατηρήσιμα χαρακτηριστικά της προσωπικότητας [Baggett, 2004]. Δύο προσωπικότητες μπορεί να μοιράζονται κοινά χαρακτηριστικά, αλλά ποτέ δεν μπορεί να είναι ακριβώς τα ίδια και αυτό ακριβώς συμβαίνει και με τη γραφή. Αυτό εξηγεί την σταθερότητα στο στυλ γραφής ενός ατόμου και τη μεταβλητότητα μεταξύ των γραπτών των διαφόρων γραφέων.

Η παλαιογραφία είναι η μελέτη των ιστορικών εγγράφων και γραμμάτων. Το μεγαλύτερο μέρος της εργασίας τους έγκειται στην μεταγραφή, την ερμηνεία και τη χρονολόγηση των χειρόγραφων, αλλά ένα σημαντικό στοιχείο είναι η αναγνώριση του χειρόγραφου ενός γραφέα ή συγγραφέα. Η αναγνώριση ενός αντιγράφου του γραφέα μπορεί να έχει σημαντικές επιπτώσεις στην επιβεβαίωση της γνησιότητάς του. Για παράδειγμα, τα χειρόγραφα που σχετίζονται με την προσωπική γραφή, είναι πιθανό να έχουν εποπτευθεί από τον γραφέα προσωπικά, επιτρέποντας στους ιστορικούς ένα βαθμό εμπιστοσύνης στην πιστότητά τους (Mooney, 2006). Η αναγνώριση μπορεί επίσης να αποτελέσει ένα στοιχείο εντοπισμού χειρόγραφου που να καταδεικνύει την έκταση της μετάδοσης ενός συγκεκριμένου κειμένου.

Το πρόβλημα της αναγνώρισης γραφέων σχετίζεται με το ζήτημα της αναγνώρισης χειρόγραφων [Vinciarelli, 2002]. Η αναγνώριση χειρόγραφου στοχεύει στην εξάλειψη των μεταβλητών που αλλάζουν λόγω άλλων διακυμάνσεων στα γραπτά κείμενα και έτσι τον εντοπισμό των μεμονωμένων χαρακτήρων και λέξεων. Η αναγνώριση γραφέων από την άλλη πλευρά βασίζεται σε αυτές τις συγκεκριμένες παραλλαγές μεταξύ των σχημάτων του χαρακτήρα του γραφέα που επιτρέπουν να χαρακτηριστεί ο γραφέας ενός συγκεκριμένου. Παρά την αντίφαση μεταξύ των δύο προσεγγίσεων,



η αναγνώριση γραφένων μπορεί να είναι βολική στην αναγνώριση χειρόγραφου, αξιοποιώντας την αρχή της προσαρμογής του συστήματος με τον τύπο του γραφέα [Nosary et al., 1999].

Με δεδομένα αυτά τα στοιχεία, η παρούσα εργασία έχει ως σκοπό να αναλύσει το ζήτημα της αναγνώρισης γραφένων, εξετάζοντας τη σημασία της, τα επιμέρους στοιχεία που αφορούν την αναγνώριση γραφένων και χειρόγραφων και κυρίως τις τεχνικές και μεθόδους αναγνώρισης που χρησιμοποιούνται σήμερα. Για το σκοπό αυτό θα γίνει μια βιβλιογραφική ανασκόπηση που στοχεύει να συγκεντρώσει τις πιο σημαντικές και βαρύνουσες απόψεις στο ζήτημα, και στη συνέχεια θα γίνει κριτική προσέγγιση και ανάλυση δυο συγκεκριμένων άρθρων που απαντούν στο θέμα της παρούσας εργασίας.

## 2 Βιβλιογραφική Ανασκόπηση

Αυτό το κεφάλαιο εξετάζει διεξοδικά τις διαθέσιμες τεχνικές υπολογιστικής αναγνώρισης γραφών. Παρά το γεγονός ότι το σύνολο των δεδομένων για το έργο αυτό έχει συνταχθεί σε παλαιογραφικό φόντο, οι σχετικές τεχνικές αφορούν την αυτοματοποιημένη επεξεργασία εγγράφου και τα βιομετρικά στοιχεία, και είναι αυτοί οι τομείς που θα πρέπει να ερευνηθούν.

Η αναγνώριση γραφέα βασίζεται σε βιομετρικά στοιχεία συμπεριφοράς: μια εμπεδωμένη δραστηριότητα από την οποία μπορούν να εξαχθούν προσωπικά στοιχεία ταυτότητας. Ως εκ τούτου, η διαδικασία ταυτοποίησης πέφτει στο θεωρητικό πλαίσιο της Αναγνώρισης Προτύπων. Ως περιοχή της ανάλυσης του εγγράφου, επίσης συνδέεται με τους τομείς της έρευνας χειρόγραφου, όπως την οπτική αναγνώριση χαρακτήρων (OCR).

Μέσα στην αυτοματοποιημένη αναγνώριση γραφέα, οι μέθοδοι χωρίζονται ανάλογα με το είδος των πληροφοριών που απασχολούν: την κίνηση και την πίεση κατά τη στιγμή της γραφής (on-line πληροφορίες), ή τα δεδομένα της τελικής εικόνας (offline πληροφορίες). Κατά την offline αναγνώριση, γίνεται μια περαιτέρω κατηγοριοποίηση ανάλογα με το επίπεδο της γραφής- με ειδικές γνώσεις/πληροφορίες που απαιτούνται για την αναγνώριση των στοιχείων. Αυτό κυμαίνεται από καμία πληροφορία, όπου οι εικόνες αντιμετωπίζονται μόνο ως pixel (0 ή 1), μέχρι και λεπτομερείς πληροφορίες σχετικά με το χαρακτήρα που έχει το κείμενο και του περιεχομένου του κάθε δείγματος.

Αυτό το κεφάλαιο περιγράφει τους συναφείς τομείς εργασίας της υπάρχουσας βιβλιογραφίας σε λεπτομέρειες, συμπεριλαμβανομένης της τρέχουσας κατάστασης της αυτοματοποιημένης αναγνώρισης γραφών για την αρχαιότητα.

## 2.1 Η αναγνώριση Γραφών

Η αναγνώριση είναι μια χρονοβόρα χειροκίνητη διαδικασία, που απαιτεί τεχνογνωσία όσον αφορά στο προσωπικό στυλ χειρόγραφου από εκατοντάδες γραφείς που εργάστηκαν και δημιούργησαν σε μια δεδομένη χρονική περίοδο. Οι μικρές λεπτομέρειες του στυλ των μεμονωμένων χαρακτήρων και ολόκληρων των μολυβιών το χαράζει ο γραφέας, εξετάστηκαν και συγκρίθηκαν μεταξύ των χειρόγραφων για να επιβεβαιωθεί εάν είχαν γραφτεί από τον ίδιο γραφέα. Αυτή η διαδικασία συνήθως περιπλέκεται από τη δυσκολία που έχουν οι σχετικοί μελετητές στο να βρουν αρκετά πολύτιμα χειρόγραφα από διάφορες συλλογές και να τα φέρουν όλα μαζί για σύγκριση (Davis, 2007).

Ένα χειρόγραφο κείμενο μπορεί είναι ενδιαφέρον για ποικίλους λόγους. Τα αρχαία χειρόγραφα, για παράδειγμα, χρησιμεύουν στους ιστορικούς και στους γλωσσολόγους στο να μελετήσουν την εξέλιξη του στυλ και της μορφής της γραφής με την πάροδο του χρόνου, που με τη σειρά του αντικατοπτρίζει τις ιστορικές και πολιτισμικές αλλαγές της κοινωνίας. Η γνώση σχετικά με μεμονωμένα γράμματα, τα δίψηφα, τα σημεία στίξης και τις συντμήσεις με τον τρόπο που έχουν εξελιχθεί επιτρέπει στους ιστορικούς και την επιστήμη της παλαιογραφίας να ταυτίζουν τις περιόδους κατά τις οποίες γράφτηκε ένα χειρόγραφο. Η ποσότητα αυτών των αρχαίων χειρόγραφων που αποθηκεύεται σε αρχεία, βιβλιοθήκες και ιδιωτικές

συλλογές είναι τεράστια και αυτό θα είναι χρήσιμο για την ανάπτυξη συστημάτων πληροφορικής που θα μπορούσαν να βοηθήσουν τους παλαιογράφους σε χειρόγραφα που χρονολογούνται, ταξινομούνται και πιστοποιούνται.

Οι πρακτικές λεπτομέρειες αυτής της κατάστασης είναι σε αντίθεση με την άλλη μεγάλη σύγχρονη εφαρμογή της αναγνώρισης γραφών. Η ιατροδικαστική εξέταση χειρόγραφων διενεργείται με την παροχή αποδεικτικών στοιχείων σε ποινικές έρευνες ή αστικές νομικές υποθέσεις. Εδώ οι ειδικοί έχουν συνήθως εύκολη πρόσβαση στα εν λόγω έγγραφα, και αυξανόμενη πρόσβαση στα ηλεκτρονικά συστήματα αναγνώρισης γραφών και τις σχετικές βάσεις δεδομένων για αυτούς (Tariador και Siguenza, 2004, Franke et al. 2003). Γι' αυτό υπάρχει ένας σημαντικός όγκος εργασιών διαθέσιμος στην αναγνώριση γραφών, ο οποίος προέρχεται από την εγκληματολογία, αντίθετα με τις εργασίες που προέρχονται από τον τομέα που συστηματικά απευθύνεται σε ιστορικά δεδομένα.

Αυτό είναι σημαντικό καθώς η προσέγγιση και ο σκοπός της γραφής μπορεί να είναι πολύ διαφορετικός ανά την εποχή. Αυτό επηρεάζει τους γραφείς και την τεχνική τους, και κατά συνέπεια, το στυλ γραφής που παράγουν [Nickell & Fischer, 1999].

Η αναγνώριση γραφών χωρίζεται γενικά στην αναγνώριση και την επαλήθευση των γραφών. Η αναγνώριση περιλαμβάνει μια μεγάλη αναζήτηση όπου δίνεται ένα χειρόγραφο δείγμα αγνώστου πατρότητας και μια βάση δεδομένων με τα δείγματα ενός αριθμού από γνωστούς γραφείς, με στόχο να βρεθεί ο γραφέας (ή μια πιθανή λίστα των γραφών) στη βάση δεδομένων. Η επαλήθευση γραφέα, από την άλλη πλευρά, περιλαμβάνει τη σύγκριση ανάμεσα σε δυο χειρόγραφα δείγματα, και ο

στόχος είναι να καθοριστεί αν και τα δύο έχουν γραφτεί από το ίδιο πρόσωπο ή όχι. Η ταυτοποίηση Γραφέα πραγματοποιείται γενικά με τον υπολογισμό ενός δείκτη ομοιότητας μεταξύ των υπό έλεγχο κειμένων και όλων των γραπτών των γνωστών γραφέων και τη διαλογή των ανακτημένων αποτελεσμάτων σε μια λίστα που δημιουργείται με την αύξηση της απόστασης από το ερώτημα. Η επιλογή των κατάλληλων ορίων αποδοχής σε αυτούς τους δείκτες ομοιότητας μπορεί στη συνέχεια να χρησιμοποιηθεί για να οδηγήσει τους μελετητές στην επαλήθευση του γραφέα.

Όταν έχουμε να κάνουμε με μεγάλα σύνολα δεδομένων, η αναγνώριση γραφέων μπορεί επίσης να χρησιμοποιηθεί ως ένα βήμα πριν από το φιλτράρισμα της εξακρίβωσης [Bensefia et al., 2005b]. Το στάδιο της αναγνώρισης θα μπορούσε να εξαγάγει ένα υποσύνολο πιθανών υποψήφιων γραπτών από τη βάση δεδομένων καθένα από τα οποία μπορεί στη συνέχεια να συγκριθεί με το υπό αμφισβήτηση έγγραφο είτε με το σύστημα ελέγχου είτε από έναν ειδικό. Ο στόχος απόδοσης του συστήματος αναγνώρισης γραφέων, όπως αναφέρεται στους Schomaker & Bulacu, (2004), αποσκοπεί στην επίτευξη ενός ποσοστού αναγνώρισης σχεδόν 100 τοις εκατό σε μια λίστα με 100 γραφείς, που υπολογίζεται από μια βάση δεδομένων που μπορεί να περιέχει έως και 10.000 δείγματα. Η επίδοση αυτή, ωστόσο, παραμένει μακριά από το να επιτευχθεί προς το παρόν. Στην παγκόσμια λογοτεχνία της παλαιογραφίας υπάρχουν μελέτες και έρευνες που έχουν προσπαθήσει να συμβάλουν στην επίτευξη αυτού του στόχου με την ανάπτυξη ενός συστήματος για την αναγνώριση γραφέων από σαρωμένη εικόνα των χειρόγραφων εγγράφων.

Παρά την ανάπτυξη των ηλεκτρονικών εγγράφων και τις προβλέψεις ενός κόσμου χωρίς χαρτί, η σημασία των χειρόγραφων εγγράφων έχει διατηρήσει τη θέση της και τα προβλήματα της αναγνώρισης και ταυτοποίησης των γραφέων είναι μια ενεργή περιοχή έρευνας τα τελευταία χρόνια. Έχει προταθεί μια ευρεία ποικιλία συστημάτων που βασίζονται στη χρήση των τεχνικών επεξεργασίας εικόνας του υπολογιστή και την αναγνώριση προτύπων, για την επίλυση των προβλημάτων που συναντώνται στην αυτόματη ανάλυση της γραφής και την αναγνώριση του γραφέα ενός υπό ερώτηση εγγράφου.

Στην παρούσα εργασία υπάρχει μεγαλύτερο ενδιαφέρον για την καταγραφή των προσεγγίσεων που αναπτύχθηκαν τα τελευταία χρόνια, χάρη στην ανανέωση του ενδιαφέροντος της ανάλυσης των εγγράφων για αυτόν τον τομέα. Οι τεχνικές για την αναγνώριση γραφέων και την ταξινόμηση χειρόγραφων παραδοσιακά ταξινομούνται σε δύο ευρείες κατηγορίες: τις κείμενο-εξαρτώμενες και τις κείμενο-ανεξάρτητες μεθόδους. Στις κείμενο-εξαρτώμενες μεθόδους τα δείγματα γραφής προς σύγκριση απαιτείται να περιέχουν το ίδιο σταθερό κείμενο. Ο έλεγχος της υπογραφής, για παράδειγμα, μπορεί να θεωρηθεί ότι εμπίπτει σε αυτή την κατηγορία. Αυτές οι μέθοδοι συνήθως χρησιμοποιούν τη σύγκριση μεταξύ των επιμέρους χαρακτήρων ή λέξεων γνωστής μεταγραφής και συνεπώς απαιτούν από ολόκληρο το κείμενο που πρόκειται να αναγνωρισθεί ή κατά διαστήματα να περιέχει χαρακτήρες ή λέξεις που αναγνωρίστηκαν νωρίτερα σε μια διαδικασία αναγνώρισης γραφέα [Schlarbach, 2007].

Οι κείμενο-ανεξάρτητες μέθοδοι από την άλλη πλευρά προσδιορίζουν το γραφέα ενός ανεξάρτητου έγγραφου από το σημασιολογικό περιεχόμενο του. Αυτές οι μέθοδοι χρησιμοποιούν τα χαρακτηριστικά που εξάγονται από ολόκληρη την εικόνα του κειμένου ή από μία περιοχή ενδιαφέροντος. Μια ελάχιστη ποσότητα χειρόγραφου είναι αναγκαία προκειμένου να αντλήσει κανείς τα σταθερά χαρακτηριστικά μηδενικής ευαισθησίας στο περιεχόμενο κειμένου των δειγμάτων [Bulacu & Schomaker, 2007].

Σε γενικές γραμμές, το κείμενο στις κειμενο-εξαρτώμενες μεθόδους λειτουργεί στο χαρακτήρα ή στο επίπεδο της λέξης, ενώ οι κείμενο-ανεξάρτητες μέθοδοι εργάζονται στα επίπεδα επί της γραμμής ή της παραγράφου. Προφανώς, οι κείμενο-ανεξάρτητες μέθοδοι είναι λιγότερο περιορισμένες και οι πιο χρήσιμες σε πρακτικές εφαρμογές, όπου η ανθρώπινη παρέμβαση τείνει να ελαχιστοποιηθεί. Θα πρέπει ωστόσο να σημειωθεί ότι οι κειμενο-εξαρτώμενες μέθοδοι ανάλυσης έχουν πολύ μεγάλη σημασία στην ιατροδικαστική εξέταση έγγραφων όπου οι ιατροδικαστές προσπαθούν να συγκρίνουν ορισμένα χαρακτηριστικά των επιμέρους σχημάτων του χαρακτήρα. Επιπλέον, η διακριτική δύναμη διαφορετικών χαρακτήρων μπορεί να διαφέρει όπως δηλώνουν οι Zhang et al., (2003] και Jain et al. (2008). Μια ιδανική λύση θα ήταν συνεπώς η χρήση αυτόματων συστημάτων για να βρεθεί μια σειρά από μεγάλα σύνολα δεδομένων και στη συνέχεια να δημιουργηθεί ο πιο περιορισμένος κατάλογος με διαχειρίσιμο μέγεθος στον εξεταστή έγγραφου.

## 2.2 Online Αναγνώριση Γραφέα

Η Αναγνώριση χωρίζεται σε δύο βασικούς κλάδους: την online και την offline αναγνώριση γραφέα. Η online αναγνώριση αναλύει δείγματα γραφής που έχουν παρθεί κατά τη διάρκεια της παραγωγής, κάνοντας χρήση της δυναμικής κίνησης, της κατεύθυνσης, του χρονοδιαγράμματος, τις πληροφορίες για διαθέσιμη πίεση (Li and Tan, 2009, Li et al., 2007), καθώς και τα δεδομένα σχήματα στα οποία βασίζεται (Namboodiri και Gupta, 2006, Blankers et al, 2007), ή ένα συνδυασμό προσεγγίσεων (Schlarbach και Bunke, 2007a). Σε αντίθεση με αυτό, ο offline εντοπισμός γραφέα χρησιμοποιεί μόνο μετρήσεις που προέρχονται από την στατική εικόνα του ολοκληρωμένου ίχνος μελανιού.

Η online αναγνώριση γραφέα έχει υψηλή ακρίβεια αναγνώρισης, καθώς και οι συμπληρωματικές πληροφορίες που είναι διαθέσιμες καθιστούν γενικά ανώτερη από αυτή την άποψη σε σχέση με τις αμιγώς offline τεχνικές (Charpan, 2006).

Το μεγαλύτερο μέρος των εργασιών στον προσδιορισμό γραφέων αφορά στον εντοπισμό και την ανάπτυξη νέων δυνατοτήτων. Η ενότητα αυτή θα περιγράψει τα χαρακτηριστικά που περιγράφουν το φάσμα των διαθέσιμων συναρτήσεων, και θα παρέχει ποιοτική επισκόπηση των βασικών μεθόδων που έχουν προταθεί.

Υπάρχουν τρεις τεχνικές διαθέσιμες για την αναγνώριση και την ταξινόμηση των χαρακτηριστικών ενός γραφέα. Τα τοπικά, τα γενικά, και τα χαρακτηριστικά υφής ενός γραφέα. Αυτά μπορούν να ομαδοποιηθούν ανάλογα με το βαθμό που είναι σχετικά με την γραφή. Τα χαρακτηριστικά υφής δεν χρησιμοποιούν πληροφορίες που



σχετίζονται με την γραφή: είναι συνήθως χαρακτηριστικά από την τελική εικόνα (Terzija και Geisselhardt, 2004, Daugman, 2005, Shahabi και Rahmati, 2009, Antonini κ.ά., 1992). Στο άλλο άκρο, τα τοπικά χαρακτηριστικά συνήθως μετρούνται απευθείας από συγκεκριμένα, αναγνωρισμένα τμήματα του ίχνους του μελανιού κατά τη στιγμή της γραφής, όπως για παράδειγμα την κίνηση μίας συνεχόμενης μολυβιάς ή οι χαρακτήρες (Pechwitz et al., 2002). Στην μέση των δύο αυτών τεχνικών, τα γενικά χαρακτηριστικά κάνουν χρήση της διάκρισης μεταξύ του ίχνους του μελανιού και του φόντου, αλλά δεν απαιτούν καμία μεταγραφή του περιεχομένου γραφής, την στιγμή του ελέγχου. Μπορούν να εξαχθούν ή να μετρηθούν από οποιοδήποτε κείμενο. Οι ονομασίες των χαρακτηριστικών είναι κατά προσέγγιση, και η ορολογία πολλές φορές ποικίλλει: τα τοπικά και γενικά χαρακτηριστικά περιγράφονται μερικές φορές ως δομικά και στατιστικά χαρακτηριστικά αντίστοιχα.

Όπως γίνεται σαφές, έχουν προταθεί διάφορες προσεγγίσεις για τον εντοπισμό γραφέων. Μια επιστημονική επικύρωση της ατομικότητας του χειρόγραφου γίνεται από τους Srihari et al. [2002]. Σε αυτή την μελέτη οι ερευνητές πήραν δείγματα χειρόγραφου 1500 ατόμων, αντιπροσωπευτικών του πληθυσμού των ΗΠΑ σε σχέση με το φύλο, την ηλικία, τις εθνικές ομάδες, κλπ. Ο γραφέας μπορεί να προσδιοριστεί 98% με βάση τα Macro (online) και τα Micro (offline) χαρακτηριστικά που εξάγονται από τα χειρόγραφα έγγραφα. Οι Said et al. [2005] πρότειναν μια γενική προσέγγιση που βασίζεται στο πολυκαναλικό φιλτράρισμα Gabor, όπου κάθε γραφέας χειρόγραφου θεωρείται ότι διαθέτει μια διαφορετική υφή. Οι Bensefia et al. [2004] που χρησιμοποίησαν τοπικά χαρακτηριστικά βασίστηκαν σε γραφήματα από κατάτμηση των καλλιγραφικών χειρόγραφων και μετά η αναγνώριση γραφέα

εκτελέστηκε από ένα μοντέλο ανάκτησης πληροφοριών βασισμένο στο κείμενο που περιεχόταν σε αυτές. Οι Schomaker et al. [2003] παρουσίασαν μια νέα προσέγγιση, χρησιμοποιώντας συνδεδεμένα περιγράμματα των μολυβιών και συναρτήσεις της πυκνότητας πιθανότητας τους. Οι Schlarbach et al [2004] προτείνουν μια προσέγγιση που βασίζεται στο κρυφό μοντέλο Μάρκοφ(HMM) για την αναγνώριση και την επαλήθευση γραφέα. Οι Marti et al. [2000] εξάγουν ένα σύνολο χαρακτηριστικών από τις γραμμές του χειρόγραφου κειμένου. Τα χαρακτηριστικά γνωρίσματα που εξάγονται αντιστοιχούν στα ορατά χαρακτηριστικά της γραφής, όπως για παράδειγμα, το πλάτος, την κλίση και το ύψος των τριών κύριων ζωνών γραφής.

Θεωρείται, ότι η γραφή είναι ένα παράδειγμα ενός βιομετρικού, προσωπικού χαρακτηριστικού που μπορεί να χρησιμοποιηθεί για την ταυτοποίηση ενός ατόμου. Η θεωρητική βάση για αυτό προέρχεται από το χώρο της ταξινόμησης και της αναγνώρισης προτύπων, η οποία ασχολείται με την απόδοση ποικίλων μετρήσεων για τη σωστή ταυτότητα και ταυτοποίηση. Το κομμάτι αυτό θα δώσει μια σύντομη επισκόπηση του πεδίου των βιομετρικών στοιχείων και τη διαδικασία αναγνώρισης προτύπων που εφαρμόζεται, και θα καταλήξει με μια σύνοψη των πιο στενών σχέσεων των ερευνητικών τομέων.

Η offline αναγνώριση γραφέα και οι προσεγγίσεις επαλήθευσης εμπίπτουν σε δύο γενικές κατηγορίες: τις κείμενο-εξαρτώμενες μεθόδους και τις κείμενο-ανεξάρτητες (Plamondon και Lorette 1989). Οι μέθοδοι που εξαρτώνται από το κείμενο είναι πολύ παρόμοιες με τις τεχνικές επαλήθευσης υπογραφής και χρησιμοποιούν την σύγκριση μεταξύ των επιμέρους χαρακτήρων ή λέξεων γνωστού σημασιολογικού

περιεχόμενου(πχ. ASCII). Συνεπώς, αυτές οι μέθοδοι απαιτούν τον εκ των προτέρων εντοπισμό και την κατάτμηση των σχετικών πληροφοριών. Αυτό γίνεται συνήθως αλληλεπιδραστικά από ένα ανθρώπινο χρήστη.

Οι κείμενο-ανεξάρτητες μέθοδοι αναγνώρισης και επαλήθευσης χρησιμοποιούν τα στατιστικά χαρακτηριστικά γνωρίσματα που εξάγονται από ολόκληρη την εικόνα ενός μπλοκ κειμένου. Μια ελάχιστη ποσότητα του χειρόγραφου (π.χ. μια παράγραφο που περιέχει λίγες γραμμές κειμένου) είναι απαραίτητη προκειμένου να αντλήσουν τα σταθερά χαρακτηριστικά που δεν επηρεάζονται από το περιεχόμενο των δειγμάτων. Το συγκριτικό πλεονέκτημα αυτής της μεθόδου είναι ότι η ανθρώπινη παρέμβαση ελαχιστοποιείται.

Τυπικά για τις προσεγγίσεις που είναι κείμενο ανεξάρτητες, τα χαρακτηριστικά που χρησιμοποιούνται για την ταυτοποίηση γραφέα παρέχουν μια συγκεντρωτική περιγραφή ολόκληρης της περιοχής που περιέχει το χειρόγραφο με την αφαίρεση των τοπικών πληροφοριών. Για το λόγο αυτό, είναι αμφίβολο να χρησιμοποιούνται κείμενο - ανεξάρτητες μέθοδοι ακόμα και στις περιπτώσεις όπου η περιεκτικότητα του κειμένου των δειγμάτων είναι σταθερή και γνωστή.

### **2.3 Η Βιομετρική**

Η "Βιομετρική είναι η επιστήμη της αναγνώρισης της ταυτότητας του προσώπου που βασίζεται στις φυσικές ή συμπεριφορικές ιδιότητες του ατόμου ...» (Jain et al., 2008).

Οι φυσικές ιδιότητες είναι οι μετρήσιμες βιολογικές ιδιότητες που κατέχει ένα πρόσωπο, όπως τα δακτυλικά αποτυπώματα, η ίριδα, ή η αναγνώριση του

προσώπου. Τα συμπεριφορικά χαρακτηριστικά είναι χαρακτηριστικά γνωρίσματα και αφορούν στον τρόπο που ένα πρόσωπο εκτελεί κάποια ενέργεια, π.χ. το βάδισμα, την ομιλία ή τη γραφή. Οι κύριες τρέχουσες χρήσεις των βιομετρικών στοιχείων αφορούν τον ασφαλή έλεγχο της ταυτότητας (επίσης γνωστός ως επαλήθευση), για παράδειγμα των δακτυλικών αποτυπωμάτων ενός ατόμου με τη σάρωση στο πρόγραμμα ταυτοποίησης, και την αναγνώριση της ίριδας. Η μακροχρόνια και ευρεία χρήση των προσωπικών υπογραφών για την πιστοποίηση των συναλλαγών ανήκει επίσης σε αυτή την κατηγορία. Το χειρόγραφο έχει αναγνωριστεί ως έγκυρο βιομετρικό στοιχείο, κυρίως όσον αφορά στην απόδειξη εγκληματολογικής εξέτασης ενός εγγράφου (FDES / QDEs), ώστε να κριθεί νομικά αποδεκτό (Srihari et al, 2002. Davis, 2007, Jain, 2002). Χρησιμοποιείται κυρίως στη μαρτυρία εμπειρογνομόνων σχετικά με τη γνησιότητα ενός συγκεκριμένου κειμένου, καθώς και για τον καθορισμό του συντάκτη του ιστορικού κειμένου, όπου αυτό είναι άγνωστο ή σε αμφισβήτηση.

Σύμφωνα με τους Jain et al (2005), επτά χαρακτηριστικά χρησιμοποιούνται για να κρίνουμε αν ένα πιθανό βιομετρικό στοιχείο είναι όντως βιομετρικό στοιχείο. Οι Dunstone και Yager (2009) απαριθμούν δεκατρία. Ιδιαίτερα όσων αφορά το χειρόγραφο (Schomaker, 2007) αυτά είναι κυρίως παράγοντες όπως η καθολικότητα, η μοναδικότητα, η μονιμότητα, και η δυνατότητα μέτρησης. Δηλαδή η βιομετρική πρέπει να υπάρχει σε όλα τα στοιχεία του σχετικού πληθυσμού, θα πρέπει να είναι μεμονωμένα διακριτά, θα πρέπει να παραμένουν ξεχωριστά, και θα πρέπει να είναι δυνατή η συλλογή των δεδομένων σε κατάλληλη μορφή. Περαιτέρω σκέψεις περιλαμβάνουν τα κοινωνικά, τα πρακτικά, και τα θέματα σχεδιασμού συστήματος. Οι Jain et al (2005) και οι Impedovo και Pirlo (2008) σημειώνουν ότι κανένα

βιομετρικό στοιχείο δε θα πληροί όλα τα κριτήρια, αλλά πρέπει να είναι κατάλληλο για την εφαρμογή σε συγκεκριμένες διαδικασίες αναγνώρισης.

Το χειρόγραφο γενικά πληροί τα βασικά κριτήρια για να είναι χρήσιμο βιομετρικά. Είναι καθολικό, καθώς ο πληθυσμός πρακτικά περιλαμβάνει όσους έχουν ήδη προσκομίσει γραπτά στοιχεία, η μοναδικότητα του έχει αποδειχθεί (π.χ. Srihari et al. (2002)) και η δυνατότητα μέτρησης του βοηθάει στην μελέτη εξαγωγής και επιλογής χαρακτηριστικών γνωρισμάτων, που καταλαμβάνουν μεγάλο μέρος της προσπάθειας αναγνώρισης της ταυτότητας των γραφένων. Το ζήτημα της μονιμότητας είναι πιο ενδιαφέρον. Είναι γνωστό ότι υπάρχουν πολλοί παράγοντες που επηρεάζουν το στυλ γραφής, τόσο συνήθεις (μιμητικοί) όσο και φυσιολογικοί (γενετικοί) (Schomaker και Bulacu, 2004). Η γήρανση, και οι συναφείς φυσικές συνθήκες που μπορούν να αναπτυχθούν, έχουν σαφώς κάποια επίδραση, αλλά δεν έχουν βρεθεί να αποτρέπουν την αναγνώριση (Walton, 1997) εκτός αν υπάρχει κάποια ασθένεια.

Η διαδικασία της αναγνώρισης γραφέα πραγματοποιείται μέσω της εφαρμογής ενός βιομετρικού συστήματος ταξινόμησης. Η ειδική τεχνική αναγνώρισης που έχει επιλεγεί για αυτόν τον τομέα είναι η γραφηματική διαδικασία κωδίκων. Οι περιγραφές και τα μοντέλα των συστημάτων αναγνώρισης δίνονται στις παρακάτω ενότητες.

## 2.4 Βιομετρικά συστήματα

Οι παρακάτω ενότητες παρέχουν μια σύντομη περιγραφή των διαδικασιών και της λειτουργίας ενός βιομετρικού συστήματος κατά τους Bolle et al. (2003) Jain et al. (2008) ή Jain et al (2004).

Ένα βιομετρικό σύστημα πρέπει να είναι σε θέση να συλλέγει, να επεξεργάζεται, να αποθηκεύει και να συγκρίνει τα βιομετρικά δείγματα, και να εξάγει μια απόφαση ως προς το κατά πόσον τα δύο δείγματα συλλέχθηκαν από το ίδιο άτομο (Dunstone και Yager, 2009). Αυτά τα συστατικά υποστηρίζουν τις διαδικασίες εγγραφής, καθώς και την ταυτοποίηση ή την επαλήθευση.

Η πρώτη προβλεπόμενη φάση του βιομετρικού συστήματος είναι η εγγραφή όταν το σύστημα αρχικοποιείται με βιομετρικά δείγματα από τα επαληθευμένα άτομα τα οποία θα πρέπει να είναι σε θέση να αναγνωρίσει. Τα στοιχεία του συστήματος που απαιτούνται σε αυτό το στάδιο είναι η συλλογή, η επεξεργασία και η αποθήκευση. Η συλλογή περιλαμβάνει ένα φυσικό αισθητήρα ικανό να δεχθεί τις βιομετρικές εισόδους, η επεξεργασία περιλαμβάνει οποιαδήποτε αυτόματη και χειροκίνητη επεξεργασία της εισόδου που απαιτείται για να παραχθεί μια καλή αναπαράσταση του βιομετρικού στοιχείου, και μετά ένα ή περισσότερα επαληθευμένα βιομετρικά δείγματα μπορούν να αποθηκευτούν για το κάθε άτομο. Αυτά τα δείγματα αποτελούν τα στοιχεία στα οποία το σύστημα θα στηριχθεί κατά τη λήψη αποφάσεων ταυτοποίησης, και υποτίθεται ότι είναι ακριβή.

Μετά την εγγραφή, το σύστημα είναι σε θέση να συγκρίνει τα εισερχόμενα δείγματα στη βάση δεδομένων των εγγεγραμμένων χρηστών και να αποφασίσει κατά πόσον έχει βρεθεί ταύτιση ή όχι. Υπάρχουν δύο κύριοι τρόποι λειτουργίας: η εξακρίβωση των στοιχείων και η επαλήθευση.

Στη λειτουργία της επαλήθευσης (που ονομάζεται επίσης έλεγχος ταυτότητας), το σύστημα ελέγχει αν ένα εισερχόμενο δείγμα ταιριάζει με την ταυτότητα που αξιώνει κάποιο άτομο. Αυτό απαιτεί, πρώτον, ότι η δηλωθείσα ταυτότητα είναι στη βάση δεδομένων του συστήματος, και ότι η επεξεργασία του δείγματος ταιριάζει στα αρχεία του συστήματος για το συγκεκριμένο άτομο αρκετά στενά, ώστε να μπορεί να κρίνει κάποιος ότι έχουν παραχθεί από το ίδιο πρόσωπο. Το σύστημα μπορεί να προσθέσει το προσφάτως συλλεχθέν δείγμα στη βάση δεδομένων του. Τα εξαρτήματα του συστήματος που εμπλέκονται σε αυτή τη διαδικασία είναι αρχικά η σύγκριση (για να προσδιοριστεί πόσο στενά ταιριάζει το νέο δείγμα στα υπάρχοντα δεδομένα για το άτομο αυτό), και η απόφαση (να προσδιοριστεί αν το αποτέλεσμα της σύγκρισης είναι αρκετά κοντά) και στη συνέχεια αν αυτό ταιριάζει η συλλογή και επεξεργασία (για το νέο δείγμα πλέον) και η αποθήκευση.

Στη λειτουργία αναγνώρισης της ταυτότητας (επίσης γνωστή ως ταυτοποίηση), το άτομο δεν χρειάζεται να διεκδικήσει μια ταυτότητα: το νέο βιομετρικό δείγμα συγκρίνεται με όλα τα άτομα που είναι γνωστά στο σύστημα. Αυτό απαιτεί ένα πολύ μεγαλύτερο αριθμό συγκρίσεων από τον έλεγχο. Επιπλέον, απαιτούνται δύο κατώτατα όρια σύγκρισης πριν εξαχθεί η απόφαση: η πρώτη, όπως και πριν, καθορίζει εάν ένα δείγμα είναι αρκετά κοντά σε ένα συγκεκριμένο άτομο για να

ταιριάζει, αλλά η δεύτερη πρέπει να προσδιορίσει αν το δείγμα είναι σε αρκετά στενή αντιστοιχία για κάθε άτομο που πρωτοκολλήθηκε στο σύστημα.

Τα συστατικά που εμπλέκονται σε αυτή τη διαδικασία είναι και πάλι η σύγκριση (για να καθορίσει το πόσο κοντά το νέο δείγμα ταιριάζει με τα υπάρχοντα δεδομένα για όλα τα άτομα) και η απόφαση (να προσδιοριστεί αν το αποτέλεσμα της σύγκρισης είναι αρκετά κοντά για να ταιριάζει με ένα μόνο άτομο). Αν ταιριάζει απόλυτα μόνο με ένα άτομο τότε γίνεται επεξεργασία του υπό εξέταση δεδομένο και αποθηκεύεται στην βάση.

Σε αντίθεση με τα περισσότερα συστήματα πιστοποίησης της αυθεντικότητας, τα οποία απαιτούν μια ακριβή αντιστοιχία μεταξύ κωδικών (όπως οι κωδικοί πρόσβασης), οι βιομετρικοί κωδικοί θα έχουν φυσικά κάποια μεταβλητότητα. Αυτό σημαίνει ότι η αντιστοίχιση του συστήματος πρέπει επίσης να ανεχθεί τη μεταβολή. Θεωρείται αυτονόητο επίσης ότι θα πρέπει να βρεθεί μια τέλεια αντιστοιχία, γιατί είναι πιθανό να αναφερθεί ότι ένα αντιγραμμένο δείγμα έχει χρησιμοποιηθεί σε μια προσπάθεια να ξεγελάσει το σύστημα. Το πολύ υψηλό ή πολύ χαμηλό όριο θα παράγει τόσο λάθη ταυτότητας, όσο και τα διάφορα μέτρα που χρησιμοποιήθηκαν για να περιγράψουν το πόσο καλά επιτυγχάνει αυτόν τον στόχο ένα σύστημα. Η ακρίβεια ενός βιομετρικού συστήματος συνήθως διατυπώνεται σε σχέση με τα ποσοστά σφάλματος του.

Πολλά βιομετρικά συστήματα διενεργούν εξακρίβωση, για την οποία υπάρχουν δύο πρότυποι τύποι λάθους: η Λάθος Τιμή Αποδοχής (FAR), η πιθανότητα δηλαδή να αποδοθεί ένα δείγμα σε λάθος άτομο, και η Λάθος Τιμή Απόρριψης (FDR), η



πιθανότητα να μην μπορέσουν οι μελετητές να ταιριάξουν ένα δείγμα που παρουσιάστηκε στο σωστό άτομο. Αυτά ονομάζονται σφάλματα τύπου I και II αντίστοιχα, και είναι επίσης γνωστά ως ψευδή ταύτιση / μη ταύτιση (ειδικά κατά την εξέταση προσδιορισμού, και όχι στα προβλήματα επαλήθευσης). Το ποσοστό αυτό μπορεί κοινότοπα να μειωθεί στο μηδέν με την απόρριψη ή την αποδοχή όλων των δειγμάτων που παρουσιάζονται σε αυτό. Θα πρέπει, συνεπώς, να χρησιμοποιούνται σε συνδυασμό για τον προσδιορισμό της σχέσης και επιρροής μεταξύ τους, δίνοντας τα χαρακτηριστικά σφάλματος ενός συγκεκριμένου συστήματος.

## 2.5 Γραφηματική Μέθοδος κωδικών

Αν και είναι σε χρήση μια ευρεία ποικιλία τεχνικών για την αναγνώριση γραφέν, υπάρχει σημαντικό μέρος της λογοτεχνίας που εστιάζει στην γραφηματική μέθοδο κωδικών (όπως περιγράφεται από τον Bulacu (2007)) ως βάση για την ανάλυση του συνόλου δεδομένων. Επιλέχθηκε για την απόδοση της υψηλής αναγνώρισης, τις σχετικά χαμηλές απαιτήσεις επεξεργασίας και των μεταδεδομένων, και την ομοιότητα στις χειροκίνητες διαδικασίες αναγνώρισης γραφέν. Τα στοιχεία αυτά ήταν σημαντικά σε αυτό το έργο, καθώς τα αρχαία ή τα μεσαιωνικά κείμενα δεν έχουν μεταγραφές διαθέσιμες, και οι μεθοδολογικές ομοιότητες προσφέρουν ένα καλό σημείο αναφοράς για τους σχετικούς μελετητές.

Η μέθοδος κωδικών βασίζεται γύρω από την κατάτμηση του κειμένου σε γραφήματα. Πρόκειται για αποσπάσματα του κειμένου σε χαρακτήρες που σχηματίζεται με τη διαίρεση καλλιγραφικών ιχνών μελανιού ευρηματικά. Αν και ο στόχος είναι να δώσει περίπου την κατάτμηση του κειμένου με βάση τους χαρακτήρες, το περιεχόμενο του

χαρακτήρα ή η έννοια του γραφήματος δεν είναι ξεκάθαρες και μερικές φορές και οι σε συγχώνευση χαρακτήρες είναι εξίσου έγκυροι. Ένας καλός εμπειρικός κανόνας θα δώσει μια συνεπή κατάτμηση, δηλαδή προτείνεται ότι παρόμοιο ίχνος μελανιού, θα παράγει παρόμοια γραφήματα εξόδου. Αυτό επιτρέπει τη σύγκριση της κατανομής των σχημάτων ίχνους μελάνης που σχηματίζεται από έναν γραφέα.

Το επόμενο βήμα για ένα μελετητή είναι να μετρήσει τα χαρακτηριστικά και τον υπολογισμό του διανύσματος για κάθε εικόνα στο σύνολο δεδομένων. Κάθε γράφημα σε μια εικόνα συγκρίνεται με όλα τα γραφήματα στο βιβλίο κωδίκων, και αντιστοιχεί με εκείνο που ταιριάζει περισσότερο. Αφού αυτό έχει γίνει για όλα τα γραφήματα σε μια εικόνα η αντιστοιχία είναι κανονικοποιημένη να συνοψιστεί σε μία, δίνοντας μια κατανομή πιθανοτήτων. Αυτή είναι η εικόνα χαρακτηριστικού διανύσματος σε σχέση με το επιλεγμένο κωδικό. Όταν όλα τα διανύσματα χαρακτηριστικών της εικόνας έχουν υπολογιστεί, μπορεί να λάβει χώρα η προαιρετική δυνατότητα επιλογής ή το στάδιο εκχύλισης. Το τελικό βήμα είναι η κατάταξη αναγνώρισης με τη σύγκριση των χαρακτηριστικών γνωρισμάτων των διανυσμάτων για την κατάρτιση αυτών που υπολογίζονται από ανώνυμες εικόνες για να καθορίσουν τον πιθανότερο γραφέα κάθε μία από αυτών.

Η ταυτοποίηση ενός γραφέα που βασίζεται σε εικόνες χειρόγραφου απαιτεί τρεις κύριες επιχειρησιακές φάσεις μετά την προεπεξεργασία της εικόνας:

- εξαγωγή χαρακτηριστικών
- συνδυασμός χαρακτηριστικών που ταιριάζουν

- αναγνώριση γραφέα και επαλήθευση

Ένα σύστημα αναγνώρισης γραφέα εκτελεί μια αναζήτηση ένα-προς-πολλά σε μια μεγάλη βάση δεδομένων με χειρόγραφα δείγματα γνωστής πατρότητας και επιστρέφει μια πιθανή λίστα των υποψηφίων (βλέπε Σχ. 1.1). Ο κατάλογος αυτός εξετάζεται περαιτέρω από κάποιον επαγγελματία ο οποίος λαμβάνει την τελική απόφαση σχετικά με την ταυτότητα του συντάκτη του δείγματος που ήταν υπό αμφισβήτηση. Η επαλήθευση γραφέα περιλαμβάνει μια σύγκριση ένα-προς-ένα για να αποφασιστεί το κατά πόσον ή όχι τα δύο δείγματα γράφτηκαν από το ίδιο πρόσωπο. Η διαδικασία επίλυσης αυτού του προβλήματος δίνει πληροφορίες για τη φύση της ατομικότητας του χειρόγραφου.

Σε αναζητήσεις αναγνώριση γραφέα, όλα τα δείγματα στο σύνολο δεδομένων διατάσσονται με την αύξηση της ανομοιότητας (ή εξ αποστάσεως) από το δείγμα - ερώτημα. Αυτό αντιπροσωπεύει μια ειδική περίπτωση ανάκτησης εικόνων, όπου η διαδικασία ανάκτησης βασίζεται σε χαρακτηριστικά σύλληψης της ατομικότητας του χειρόγραφου. Σε δοκιμές επαλήθευσης γραφέα, αν η απόσταση μεταξύ δύο δειγμάτων που επιλέχθηκαν είναι μικρότερη από ένα προκαθορισμένο όριο, τα δείγματα θεωρούνται ότι έχουν γραφτεί από το ίδιο πρόσωπο. Σε αντίθετη περίπτωση, τα δείγματα θεωρούνται ότι έχουν γραφτεί από διαφορετικούς γραφείς. Η επαλήθευση γραφέα έχει δυνατότητα εφαρμογής σε ένα σενάριο στο οποίο ένας συγκεκριμένος γραφέας, θα πρέπει να ανιχνεύεται αυτόματα σε μια ποσότητα χειρόγραφων εγγράφων.

## 2.6 Παράγοντες που καθορίζουν την ατομικότητα του χειρόγραφου

Καθώς ο γραφέας ωριμάζει, φεύγει από το στυλ τετράδιο που μαθαίνει στην τάξη και σταδιακά ενσωματώνετε στη δική του γραφή βάση της δικής του ατομικότητας. Ιδιαίτερα στις μέρες μας, όπου υπάρχει λιγότερη έμφαση στην καλλιγραφία στο σχολείο.

Υπάρχουν δύο βασικοί παράγοντες που συμβάλλουν στην ατομικότητα του γραπτού: οι γενετικοί (βιολογικοί) και μιμητικοί (πολιτιστικοί) παράγοντες. Ο πρώτος θεμελιώδης παράγοντας είναι η γενετική σύσταση του γραφέα. Οι γενετικοί παράγοντες είναι γνωστό ή μπορεί να υποτεθούν ότι συμβάλλουν στην ατομικότητα του χειρόγραφου στυλ.

- Η βιομηχανική δομή του χεριού, δηλαδή, τα σχετικά μεγέθη των οστών του καρπού του καρπού και των δακτύλων και την επιρροή τους επί της λαβής του στυλό
- Το αριστερό ή το δεξί χέρι (Francks et al 2003)
- Η μυϊκή δύναμη, η κόπωση, οι περιφερειακές κινητικές διαταραχές (Gulcher κ.ά. 1997)
- το κεντρικό νευρικό σύστημα (ΚΝΣ) δηλαδή, η ικανότητα για έλεγχο της κίνησης και η σταθερότητα του ΚΝΣ σε εκτέλεση εργασιών κίνησης (Van Galen et al. 1993).

Ο δεύτερος παράγοντας που αποτελείται από μιμητικά ή πολιτισμικά στοιχεία μεταφέρει επιρροές (Moritz 1990) σχετικά με το στυλ γραφής και τα σχήματα των χαρακτήρων (μοσχεύματα), που έχουν διαμορφωθεί κατά τη διάρκεια της

εκπαίδευσης ή έχουν γίνει κτήμα του γράφοντα από την παρατήρηση των κειμένων των άλλων προσώπων. Η κατανομή των αλληλογραφών σε ένα πληθυσμό γραφέων επηρεάζεται σε μεγάλο βαθμό από τις μεθόδους που διδάσκονται στο σχολείο, η οποία με τη σειρά της εξαρτάται από παράγοντες όπως η γεωγραφική κατανομή, η θρησκεία και οι τύποι σχολείων.

Μαζί, οι γενετικοί και μιμητικοί παράγοντες καθορίζουν μια συνήθη διαδικασία γραφής, με αναγνωρίσιμα στοιχεία στο επίπεδο του ίχνους του μελανιού, στο επίπεδο του χαρακτήρα στο σύνολό του και στο επίπεδο του χαρακτήρα στην τοποθέτηση και διάταξη σελίδας.

Το χειρόγραφο μπορεί να περιγραφεί ως μια ιεραρχική ψυχοκινητική διαδικασία: αρχικά, μια αφηρημένη κίνηση ανακτάται από τη μακροπρόθεσμη μνήμη. Στη συνέχεια καθορίζονται λεπτομερώς οι παράμετροι για την κίνηση, όπως το μέγεθος, το σχήμα, το χρονοδιάγραμμα. Τέλος, δημιουργούνται οι εντολές για τα βιοφυσικά συστήματα μυών (Maarse 1987).

Η εργασία αυτή παρουσιάζει μια νέα προσέγγιση για την αυτόματη αναγνώριση γραφέα. Η μέθοδος αυτή εφαρμόζεται για τον προσδιορισμό του γραφέα των αρχαίων ελληνικών επιγραφών που με τη σειρά του μπορεί να προσφέρει ακριβή και αντικειμενική χρονολόγηση του περιεχομένου των επιγραφών. Μια τέτοια χρονολόγηση είναι ζωτικής σημασίας για την ορθή γραφή της ιστορίας. Η μεθοδολογία αυτή βασίζεται στην ιδέα της δημιουργίας ενός ιδανικού εκπρόσωπου για κάθε σύμβολο στο αλφάβητο σε κάθε επιγραφή, μετά από κατάλληλη

τοποθέτηση όλων των επιτευγμάτων του συγκεκριμένου συμβόλου στην παρούσα επιγραφή.

### 3 Ανάλυση Άρθρων

#### 3.1 New Mathematical and Algorithmic Schemes for Pattern Classification with Application to the Identification of Writer of Important Ancient Documents

Στο άρθρο *New Mathematical and Algorithmic Schemes for Pattern Classification with Application to the Identification of Writer of Important Ancient Documents* των Αραμπατζή et al, τονίζεται η σημασία της αναγνώρισης γραφών και χειρόγραφων.

##### 3.1.1 Θεωρητικό Υπόβαθρο

Το σύνολο των σωζόμενων χειρόγραφων έγγραφων είναι μία από τις κύριες πηγές για την επιστήμη της Ιστορίας. Πιο συγκεκριμένα, λαξευμένα σε πέτρα κείμενα όπως και αρχαίες επιγραφές είναι σημαντικά μέσα για τη μελέτη της Αρχαιότητας. Ομοίως, χειρόγραφα σε παπύρους και περγαμηνές συνέβαλαν στη μετάδοση του αρχαίου κόσμου της λογοτεχνίας μέσα από το Μεσαίωνα, οδηγώντας τελικά στην Αναγέννηση και το Διαφωτισμό. Για παράδειγμα, οι γραφείς δηλώνουν ότι ακόμα και η ομηρική Ιλιάδα επιβιώνει κυρίως μέσω κάποιων μεγάλων χειρόγραφων βιβλίων, που παρήχθησαν στην Κωνσταντινούπολη κατά τη διάρκεια του 10ου ή του 11ου αιώνα και σήμερα βρίσκονται διάσπαρτα σε διάφορες βιβλιοθήκες σε όλη την Ευρώπη: τη Βενετία, το El Escorial στην Ισπανία, το Λονδίνο, τη Γενεύη, τη Φλωρεντία και τη Ρώμη. Οι τόμοι αυτοί περιέχουν τα ποιήματα του Ομήρου, καθώς και μια σειρά από διαφορετικά κείμενα σχολιασμού και σύντομες σημειώσεις στο περιθώριο του χειρόγραφου και μεταξύ των γραμμών. Όπως εύκολα μπορεί κανείς να υποθέσει από τα παραπάνω, η χρονολόγηση του περιεχομένου αυτών των επιγραφών και

χειρόγραφων είναι ιδιαίτερα υψηλής σημασίας και για τους δύο κλάδους, της Ιστορίας και της Αρχαιολογίας.

Ωστόσο, οι γραφείς των αρχαίων επιγραφών και χειρόγραφων σπάνια υπέγραφαν ή έβαζαν ημερομηνίες στα έγγρατά τους, καθιστώντας τη διαδικασία της χρονολόγησης τους πραγματικά δύσκολη. Το γεγονός αυτό συχνά προκαλεί διαμάχες και διαφωνίες μεταξύ των επιστημόνων. Ένας βασικός στόχος της παρούσας εργασίας είναι να αναλύσει ποσοτικά το περιεχόμενο ενός συγκεκριμένου συνόλου αρχαίων επιγραφών και βυζαντινών κωδίκων, έτσι ώστε να προσδιοριστούν οι σχετικές ημερομηνίες παραγωγής τους, καθώς και η σχέση μεταξύ τους.

Στο άρθρο αυτό εισάγεται μια νέα προσέγγιση για την ταξινόμηση καμπυλών σε σωστές λίστες και οικογένειες, ανάλογα με την ομοιότητα. Οι γραφείς εισάγουν μια μαθηματική ποσότητα που ονομάζουν επίπεδο καμπυλότητας και όπως επίσης και μια σειρά από προτάσεις, οι οποίες δηλώνονται και στην συνέχεια αποδεικνύονται. Εισάγονται κάποια κατάλληλα μέτρα για ομοιότητας δύο καμπυλών και στην συνέχεια εφαρμόζεται μια στατιστική ανάλυση. Η αποτελεσματικότητα της διαδικασίας προσαρμογής καμπύλης έχει δοκιμαστεί αρχικά σε 2 σχήματα συνόλων δεδομένων αναφοράς. Στη συνέχεια η μεθοδολογία έχει εφαρμοστεί στο πολύ σημαντικό πρόβλημα της ταξινόμησης 23 βυζαντινών κωδίκων και 46 αρχαίων επιγραφών στους γραφείς τους, επιτυγχάνοντας έτσι τη σωστή χρονολόγηση του περιεχομένου τους.



### 3.1.2 Σκοπός της Μελέτης

Η εργασία των Αραμπατζή et al επιχειρεί να αντιμετωπίσει τρία προβλήματα: 1) Να αναπτύξει και να παρουσιάσει μια νέα μέθοδο για το βέλτιστο ταίριασμα δύο καμπύλων, μια από τις οποίες μπορεί να υπόκειται σε δύο ανεξάρτητους μετασχηματισμούς κλίμακας (είτε X ή Y-άξονα), περιστροφής και παράλληλης μετάφρασης. 2) Να καθιερωθούν σωστά, νεωτεριστικά στατιστικά κριτήρια, προκειμένου να χαρακτηριστεί ένα συγκεκριμένο σύνολο των εν λόγω καμπύλων σε σωστές ομάδες / συστάδες, σύμφωνα με την ομοιότητά τους, με βάση την εισαγωγή νέων μέτρων ομοιότητας. 3) Να χαρακτηριστεί μια σειρά από σημαντικές αρχαίες επιγραφές και βυζαντινοί κώδικες στον σωστό γραφέα, έτσι ώστε τα έγγραφα αυτά να μπορούν αναμφίβολα να χρονολογηθούν ορθά.

Οι γραφείς υποθέτουν ότι, όταν ένας συγκεκριμένος γραφέας δημιουργεί μια υλοποίηση ενός συμβόλου του αλφάβητου σε ένα έγγραφο, τότε ο γραφέας μπορεί να αλλάξει τον προσανατολισμό, τη θέση και το μέγεθος του παραγόμενου γράμματος, μάλλον αυθαίρετα. Ωστόσο, εξακολουθεί να υπάρχει ένας πυρήνας που δημιουργείται στην πραγματοποίηση, η οποία παραμένει αμετάβλητη υπό τους προαναφερθέντες μετασχηματισμούς και, επιπλέον, είναι χαρακτηριστικό προς τον ίδιο τον γραφέα.

### 3.1.3 Περιγραφή της Μεθοδολογίας

Η σύγκριση σχημάτων αντιμετωπίζεται ως χαρακτηριστικά που ταιριάζουν στον προσανατολισμό και τα γεωμετρικά χαρακτηριστικά που αξιολογήθηκαν με διαφορεικές ιδιότητες της συνάρτησης. Στη συνέχεια, η ομαδοποίηση των σχημάτων

γίνεται σε δένδρα αποφάσεων - με βάση το ιεραρχικό πλαίσιο ομαδοποίησης. Έχει διατυπωθεί μια χαρακτηριστική προσέγγιση για το σχήμα κατηγοριοποίησης με ένα στατιστικό τρόπο. Τα χαρακτηριστικά των σχημάτων που σχηματίζουν μια ομάδα χαρτογραφήθηκαν επί της γραμμικής βάσης της βαθμολογίας Gaussian – δηλαδή μια μήτρα συνδιακύμανσης η οποία επιλέγεται να είναι η αντιπροσωπευτική ομάδα σχημάτων. Έπειτα μελετάται η επίπεδη εγγραφή σχημάτων κάτω από συσχετισμένους μετασχηματισμούς χρησιμοποιώντας την Ευκλείδεια αναπαράσταση της απόστασης των επίπεδων σχημάτων. Το πρόβλημα αντιστοίχισης στη συνέχεια διατυπώνεται μέσω της μεγιστοποίησης της αμοιβαίας πληροφορίας στο χώρο των παραμέτρων του συσχετισμένου μετασχηματισμού. Η μεγιστοποίηση επιτυγχάνεται μέσω μιας μεθόδου απότομης καθόδου.

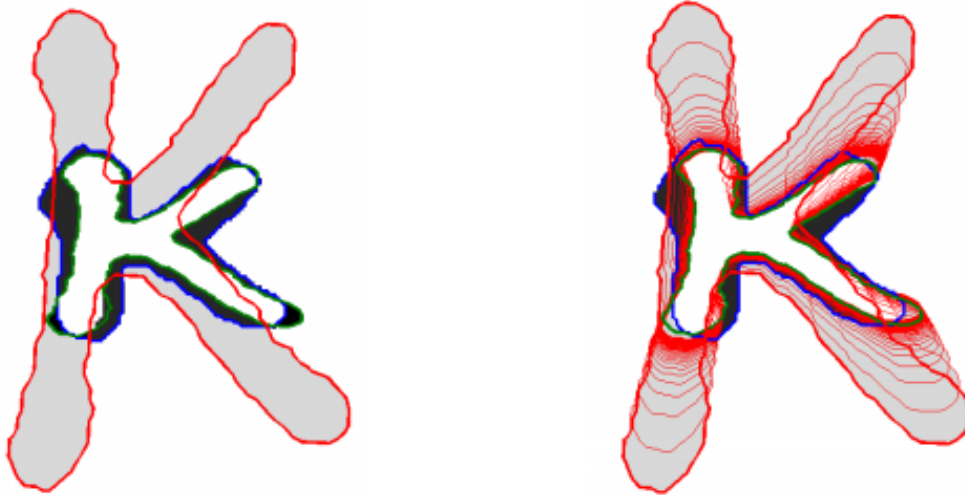
Ο προσανατολισμός της αμετάβλητης σύγκρισης των επίπεδων σχημάτων αντιμετωπίζεται σε όρους μιας "ορθής" σύγκριση μεταξύ των αλληλουχιών των αξιών καμπυλότητας τους. Η σύγκριση προσδιορίζεται ως η νέα παραμετροποίηση των καμπυλών που ως επί το πλείστον ωφελεί την καμπυλότητα, βελτιστοποιημένη για κάθε δυνατή απόκλιση από σημείο σε σημείο αντιστοιχίας, χρησιμοποιώντας τον αλγόριθμο του Dijkstra.

Η προσέγγιση που εισάγεται αφορά την εγγραφή καμπύλης N-διαστάσεων και τη σύγκριση με αξιολόγηση της ευθυγράμμισης της εφαιπτομενικής κατεύθυνση της καμπύλης με εφαιπτόμενες άλλης σταθερής καμπύλης στο πλαίσιο των συζυγών. Η σύγκριση των σχημάτων σε κλειστό επίπεδο βασίζεται στην κατασκευή πρωτοτύπων και διαμορφώνει αποκλίσεις σαν να είχαν προκληθεί από μια Νευτώνεια ροή φορέα

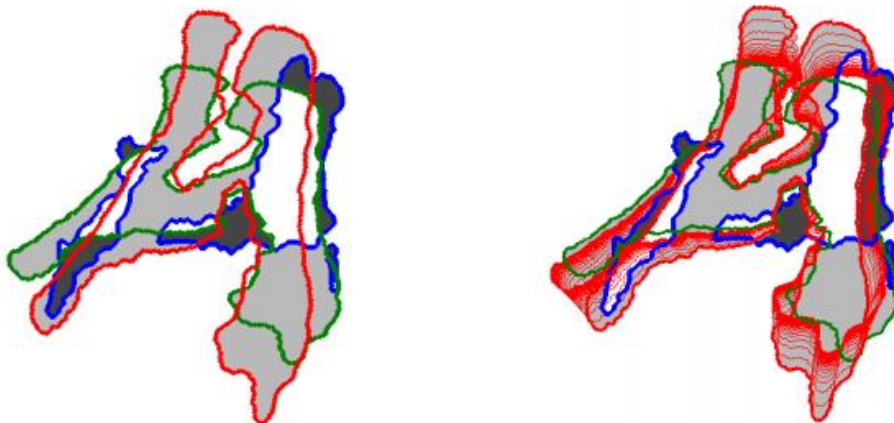
που ενεργεί στο εσωτερικό του κάθε σχήματος. Στη συνέχεια, η αντιστοιχία ένα-προς-ένα σχήμα προσδιορίζεται ως γεωδαιτικής διαδρομή που ελαχιστοποιεί την παραμόρφωση στελεχών. Η ροή του φορέα παραμόρφωσης διαμορφώνεται ως ένα άπειρο διαστάσεων (Hilbert) πολλαπλής πάνω σε συνάρτηση περιγραφικών των κλειστών καμπυλών κλιμάκωσης και περιστροφής. Στη συνέχεια, οι καμπύλες σφάλματος τοποθέτησης επιλέγονται έτσι ώστε να διατηρούν το εσωτερικό του αντικειμένου που ορίζεται στην πολλαπλότητα των εσωτερικών περιορισμών. Αυτό επιτρέπει την προβολή και τη μετάφραση της αναπαράστασης των σχημάτων κατά μήκος διαδρομών που ελαχιστοποιούν το σφάλμα.

Στη συνέχεια, κατασκευάζεται μια χαρτογράφηση, ώστε να μετατρέψει τον επιλεγμένο μετρικό χώρο σε έναν Ευκλείδειο και έτσι να "γραμμικοποιήσει" το γεωδαιτικό πρόβλημα. Οι γραφείς της μελέτης μετατρέπουν την αναπαράσταση ενός σχήματος από την μορφή ενός γραμμικού συνδυασμού συναρτήσεων 1 διάστασης σε συνάρτηση 2 διαστάσεων, χρησιμοποιώντας DC συναρτήσεις. Αυτό επιτυγχάνεται με την επίλυση ενός αντίστοιχου προβλήματος βελτιστοποίησης και Ευκλείδειους μετασχηματισμούς

Ένας πρώτος μείζων στόχος της παρούσας εργασίας είναι να επιτευχθεί η βέλτιστη προσαρμογή για ένα σύνολο καμπύλων. Προς αυτή την κατεύθυνση, οι μελετητές όρισαν πρώτα έναν νέο όρο που ονόμασαν "καμπυλότητα". Στην πράξη, όλες οι καμπύλες θα είναι ψηφιακές και θα βρίσκονται σε μία ψηφιακή εικόνα.



Εικόνα 1: Απεικόνιση της διαδικασίας αντιστοίχισης που εφαρμόζεται σε δύο περιγράμματα γραμμάτων που ανήκουν σε διαφορετικούς κώδικες.



Εικόνα 2 Απεικόνιση της διαδικασίας αντιστοίχισης που εφαρμόζεται σε δύο περιγράμματα γραμμάτων που ανήκουν σε διαφορετικές επιγραφές

### 3.1.4 Δυσκολίες στην Αναγνώριση

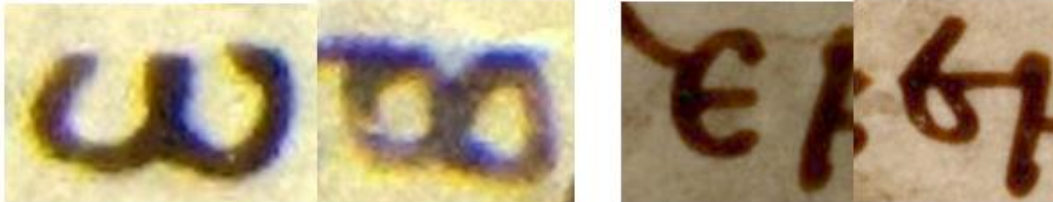
Μια σημαντική δυσκολία που αντιμετώπισαν οι γραφείς ήταν ότι δεν υπάρχει ένα σετ εκπαίδευσης διαθέσιμο, δηλαδή δεν υπάρχουν χειρόγραφα τα οποία μπορούν να χρησιμοποιηθούν ως αναφορά, σε αντίθεση με τις περισσότερες προσεγγίσεις που εμφανίζονται στη βιβλιογραφία μέχρι σήμερα. Αντίστοιχα, κανένα κείμενο γραμμένο από τον προς αναζήτηση γραφέα δε χρησιμοποιήθηκε ως έγγραφο αναφοράς και καμία προ υπάρχουσα βάση δεδομένων και πληροφοριών σχετικά με τα κείμενα ή γραφείς τους δεν έχει χρησιμοποιηθεί.

Μια άλλη σημαντική δυσκολία που αντιμετώπισαν οι γραφείς ήταν ότι δεν είχαν καμία προηγούμενη γνώση του αριθμού των διακριτών γραφέων που είχαν γράψει τα εξεταζόμενα χειρόγραφα. Επιπλέον, ορισμένα από τα πιο σημαντικά προβλήματα που προκύπτουν κατά την προσπάθεια να εντοπίσουν αυτόματα το γραφέα των αρχαίων έγγραφων είναι:

- Η προφανής μεταβλητότητα του τρόπου με τον οποίο ο γραφέας γράφει τα σύμβολα του αλφάβητου. Για παράδειγμα, τα σύμβολα Θ or Ω or E, ακόμα και αν είναι γραμμένα από το ίδιο χέρι, ακόμη και αν συναντώνται στο ίδιο χειρόγραφο, μπορεί να είναι είτε κλειστά είτε ανοικτά. Οι παρακάτω φωτογραφίες είναι ενδεικτικές της αναφοράς
- Επιπλέον, σε μια σειρά από έγγραφα, το σύμβολο του αλφάβητου κάπα μοιάζει με το λατινικό «u», ενώ κάπου αλλού στο ίδιο έγγραφο, μοιάζει με κ
- Ομοίως, σε πολλές αρχαίες επιγραφές, στο γράμμα A ο ίδιος γραφέας συνδέει τη μεσαία τραβέρσα του πολλές φορές με το αριστερό πόδι μόνο,

άλλες φορές μόνο με το δεξί πόδι, ενώ μερικές φορές δεν το συνδέει με οποιοδήποτε πόδι καθόλου, κ.λπ.

- Συχνά παρατηρείται ότι η ομοιότητα μεταξύ δύο συγκεκριμένων δειγμάτων από τα σύμβολα της αλφάβητου του ίδιου γραφέα είναι μικρότερη από την ομοιότητα μεταξύ άλλων ζευγών δειγμάτων από διαφορετικούς γραφείς.
- Τα υπό εξέταση χειρόγραφα μπορεί να έχουν υποστεί σοβαρή φθορά
- Με βάση τις προαναφερθείσες παρατηρήσεις, πρέπει να προσδιοριστεί ένας πυρήνας σε κάθε υλοποίηση των σύμβολων της αλφάβητου που παραμένει αναλλοίωτο στα διάφορα έγγραφα γραμμένα από τον ίδιο γραφέα. Επιπλέον, πρέπει να βρεθεί και να εκφραστεί μια μαθηματική περιγραφή αυτού του πυρήνα



Εικόνα 3: Εκδήλωση του σχήματος μεταβλητότητας που ανέκυψε μεταξύ της παραγωγής του ίδιου σύμβολου του αλφάβητου που παράγεται από τον ίδιο γραφέα

### 3.1.5 Συμπεράσματα της Μελέτης

Στην εργασία, παρουσιάστηκε μια καινοτόμος μεθοδολογία η οποία κατατάσσει τις καμπύλες σε σωστές οικογένειες σύμφωνα με την ομοιότητά τους. Οι μελετητές υποθέτουν ότι τα σύμβολα που εντάσσονται σε κάθε τέτοια οικογένεια παραμένουν

σε αυτή, κάτω από κάθε περιστροφή, παράλληλη μετάφραση και αλλαγή μεγέθους κατά μήκος των αξόνων  $\chi$  και  $\psi$  ανεξάρτητα.

Αυτή η μέθοδος έχει εφαρμοστεί στο πρόβλημα της αυτόματης ταξινόμησης των αρχαίων επιγραφών και βυζαντινών κωδίκων, ανάλογα με τους γραφείς τους. Ένας τέτοιος χαρακτηρισμός μπορεί να γίνει ένα πραγματικά ισχυρό, αποδοτικό και χρήσιμο εργαλείο για την ακριβή χρονολόγηση των αρχαίων χειρόγραφων. Η προσέγγιση που εισάγουν οι μελετητές εξάγει πρώτα το περίγραμμα των συμβόλων του αλφαβήτου που εμφανίζονται στην κάθε επιγραφή ή κώδικα. Στη συνέχεια, εισάγεται μια μαθηματική οντότητα, η οποία καλείται καμπυλότητα και αποδεικνύεται με μια σειρά από θεμελιώδεις προτάσεις. Με αυτόν τον τρόπο, αποδεικνύεται ότι, όταν εισάγεται ένα κατάλληλο, νέο μέτρο ομοιότητας του επιπέδου καμπυλότητας, τότε υπάρχει μια σωστή διαδικασία αλλαγής μεγέθους που ελαχιστοποιεί με τον καλύτερο δυνατό τρόπο το αντίστοιχο λάθος τοποθέτησης ανεξάρτητα από την περιστροφή και την παράλληλη μετάφραση. Έτσι, εφαρμόζεται ένας αλγόριθμος ελαχιστοποίησης, ο οποίος προσφέρει τη βέλτιστη προσαρμογή οποιονδήποτε δύο καμπύλων, όσον αφορά την καμπυλότητα. Στη συνέχεια, χρησιμοποιείται ένα σφάλμα τοποθέτησης Ευκλείδειου τύπου, το οποίο λαμβάνει υπόψη την περιστροφή και παράλληλη μετάφραση. Επιπλέον, οι μελετητές ορίζουν τα νέα μέτρα ομοιότητας των δύο καμπύλων με βάση τα προηγούμενα αποτελέσματα. Η στατιστική επεξεργασία της ομοιότητας μετρά τις τιμές, μέσω μιας νέας προσέγγισης, και μας δίνει τόσο τον αριθμό των διακριτών χειρών που έχουν γράψει αυτά τα έγγραφα, όσο και το σύνολο των επιγραφών ή κωδίκων που ανήκουν σε κάθε ένα από αυτά.

Η εισαχθείσα προσέγγιση εφαρμόστηκε σε 23 βυζαντινούς κώδικες και 46 αρχαία επιγραφές της Κλασικής και της Ελληνιστικής εποχής. Αξίζει να αναφερθεί στους περιορισμούς και στα ειδικά χαρακτηριστικά της μελέτης ότι η ομάδα των γραφέων της μελέτης αυτής, η οποία αποτελείται από Μαθηματικούς και ειδικούς Μηχανικούς, δεν χρησιμοποίησαν κανένα χειρόγραφο αναφοράς, δεν είχαν καμία ιδέα σχετικά με τον αριθμό των διακριτών χεριών που είχαν γράψει τα εξεταζόμενα χειρόγραφα και δεν είχαν οποιαδήποτε σχετική πληροφορία παντός είδους στη διάθεσή τους. Το σύστημα που αναπτύχθηκε κατέταξε τους 23 βυζαντινούς κώδικες σε 4 διαφορετικούς γραφείς και τις 46 επιγραφές σε 10 διαφορετικά χέρια. Διακεκριμένοι ειδικοί στην επιγραφολογία, Αρχαιολογία και Κλασικές Σπουδές επιβεβαίωσαν πλήρως τον χαρακτηρισμό που προσφέρεται από το σύστημα.

### **3.2 Identifying the Writer of Ancient Inscriptions and Byzantine Codices. A Novel Approach**

Στο άρθρο *Identifying the Writer of Ancient Inscriptions and Byzantine Codices. A Novel Approach* των Παπαοδυσσεύς, κ.α. προτείνεται μια νέα μεθοδολογία που παρουσιάζεται με σκοπό την αυτόματη αναγνώριση του γραφέα των αρχαίων επιγραφών και βυζαντινών κωδίκων. Αυτή η αναγνώριση μπορεί να προσφέρει σαφή χρονολόγηση αυτών των αρχαίων χειρόγραφων. Η εισαχθείσα μεθοδολογία εφαρμόζεται εκτός από μεμονωμένα γράμματα και σε περιγράμματα συμπλοκών γραμμάτων ή οποιασδήποτε κατηγορίας παρόμοιες καμπύλες.

Η μέθοδος που παρουσιάζεται στο άρθρο υπολογίζει αρχικά την κανονικοποιημένη καμπυλότητα σε κάθε pixel περιγράμματος ενός γράμματος. Στη συνέχεια, εκτελεί



κατά ζεύγη συγκρίσεις των ακολουθιών καμπυλοτήτων που αντιστοιχούν σε δύο υλοποιήσεις του ίδιου συμβόλου του αλφάβητου. Έπειτα, εισάγετε μια νέα πρόταση που, με βάση τα προηγούμενα αποτελέσματα, προσφέρει μια λύση στο πρόβλημα του ταιριάσματος δύο αντίστοιχων ψηφιακών περιγραμμάτων. Στη συνέχεια, χρησιμοποιείται ένα κριτήριο ποσοτικοποίησης της ομοιότητας δύο υλοποιήσεων του ίδιου συμβόλου του αλφάβητου. Τέλος, μια σειρά από στατιστικά κριτήρια χρησιμοποιούνται για την αυτόματη αναγνώριση του γραφέα αρχαίων χειρόγραφων. Η μέθοδος που εισήχθη δεν έχει κανένα χειρόγραφο αναφοράς, ούτε τον αριθμό των διακριτών χειρών που είχαν γράψει το υπό εξέταση σύνολο των χειρόγραφων ούτε απολύτως οποιαδήποτε σχετική πληροφορία, όπως ακριβώς και στο πρώτο υπό εξέταση άρθρο. Επίσης χρησιμοποιείται αρκετά αποτελεσματικά ακόμη και αν έχουμε ένα μικρό αριθμό ορισμένων συμβόλων του αλφαβήτου (λιγότερο από 6) σε ένα υπό εξέταση κείμενο. Η μόνη εκ των προτέρων γνώση είναι το αλφάβητο της γλώσσας υπό εξέταση. Οι μελετητές τονίζουν ότι σε διαφορετική περίπτωση η μέθοδος δεν εξαρτάται καθόλου από την ίδια τη γλώσσα. Δηλαδή δεν λαμβάνει υπόψη αν το αλφάβητο είναι λατινικό, ελληνικό, ετρουσκικό, κλπ. Η μεθοδολογία που αναπτύχθηκε, έχει εφαρμοστεί, όπως και η προηγούμενη, σε 46 αρχαίες επιγραφές της κλασικής και ελληνιστικής εποχής και 23 βυζαντινοί κώδικες, προσφέροντας 100% ακριβή αποτελέσματα, με την έννοια ότι τα λαμβανόμενα αποτελέσματα είναι σε πλήρη συμφωνία με διακεκριμένους επιστήμονες στον τομέα της Αρχαιολογίας, της Ιστορίας και των Κλασικών Σπουδών.

### 3.2.1 Θεωρητικό Υπόβαθρο της Μελέτης

Η ιδέα προέκυψε στους γραφείς του άρθρου, για να επιτευχθεί χρονολόγηση των αρχαίων εγγράφων μέσω της αυτόματης αναγνώρισης των γραφέα τους. Στην πραγματικότητα, όσον αφορά τις αρχαίες επιγραφές, οι γραφείς σκάλιζαν τις πέτρες, και υπήρχαν άνθρωποι που το έκαναν αυτό ως επάγγελμα κατά την αρχαιότητα. Αυτό έχει τις ακόλουθες συνέπειες: α) Ο αριθμός των διαφορετικών χεριών που επεξεργάζονται τις πέτρες για να γράψουν ένα κείμενο ήταν ιδιαίτερα περιορισμένος. Έτσι, μπορεί κανείς να αναμένει ότι οι δεκάδες χιλιάδες επιγραφές που ανακαλύφθηκαν στην Αττική μπορεί να ανήκουν σε μερικές εκατοντάδες διαφορετικούς γραφείς. β) Εάν κάποιος καταφέρει να αποδώσει μια σειρά από επιγραφές σε έναν γραφέα, τότε είναι πολύ εύκολο να καθοριστεί η περίοδος της καριέρας του, δεδομένου ότι, κατά κανόνα, το περιεχόμενο τουλάχιστον μίας επιγραφής θα το αποκαλύψει. Προφανώς, όσο μεγαλύτερος είναι ο αριθμός των επιγραφών που σωστά αποδίδεται σε αυτό το χέρι, τόσο μεγαλύτερη είναι η πιθανότητα να χρονολογηθεί η καριέρα του. γ) Εάν βρεθεί μια νέα επιγραφή το περιεχόμενο της οποίας δεν μπορούν να χρονολογηθεί, και η προτεινόμενη μεθοδολογία αποδώσει αυτήν την επιγραφή σε κάποιο γραφέα, τότε, το περιεχόμενο της επιγραφής κερδίζει μια ημερομηνία αμέσως, η οποία είναι η χρονική περίοδος κατά την οποία ο γραφέας ήταν ενεργός επαγγελματικά. Σημειώνουμε ότι η εργατική σταδιοδρομία των περισσότερων αρχαίων γραφέων κάλυπταν περίπου 20 έως 25 χρόνια.

Παρόμοιες παρατηρήσεις ισχύουν και για όσα διατηρημένα αρχαία κείμενα έγιναν αντιγραφή με το χέρι. Για παράδειγμα, όπως έχει ήδη επισημανθεί, λίγα χέρια

διατήρησαν τα ποιήματα του Ομήρου αντιγράφοντας τα στην Κωνσταντινούπολη κατά τη διάρκεια του 10ου ή του 11ου αιώνα, ενώ μερικοί άλλοι γραφείς προσέθεσαν τα σχόλιά τους στα χειρόγραφα. Αν κάποιος μπορεί με αυτόματα μέσα να προσδιορίσει τα διαφορετικά χέρια που είχαν γράψει κείμενο στα έγγραφα αυτά, αποκτά αμέσως μια χρονολογική κατάταξη, καθώς και επιπλέον πολύτιμες πληροφορίες.

Ως εκ τούτου, η σημασία της ανάπτυξης ενός συνόλου μεθόδων και ενός σχετικού πληροφοριακού συστήματος που εκτελεί την ορθή ταυτοποίηση του γραφέα είναι απολύτως απαραίτητη. Η ερευνητική ομάδα των γραφέων στην μελέτη έχει ήδη κάνει μια πρώτη προσπάθεια προς αυτή την κατεύθυνση με πολύ καρποφόρα αποτελέσματα, σε σχέση με αρχαίες επιγραφές, που παρουσιάστηκε στο προηγούμενο άρθρο της μελέτης.

Υπάρχει, ωστόσο, σημαντική καινοτομία στην παρούσα εργασία σε σχέση με την προηγούμενη και συγκεκριμένα: 1) Η μέθοδος που παρουσιάζεται εδώ είναι ουσιαστικά διαφορετική από τις μεθόδους που εισάγεται στο προηγούμενο άρθρο. Οι μελετητές τόνισαν ότι όταν αναπτύσσετε μια νέα ανεξάρτητη στατιστική μέθοδος για την επίλυση ενός προβλήματος, συμπεριλαμβανομένων των στοχαστικών διαδικασιών, τότε μπορεί κανείς να είναι ασυμπτωτικά βέβαιος για τα αποτελέσματα που του συγκρίνοντάς τα με τις ήδη υπάρχουσες μεθόδους. Με άλλα λόγια, κάθε νέα προσέγγιση η οποία επιβεβαιώνει τα ήδη υπάρχοντα αποτελέσματα που προσφέρονται από άλλες στατιστικές μεθόδους αυξάνει το βαθμό εμπιστοσύνης ότι αυτά τα αποτελέσματα είναι σωστά. 2) Οι μέθοδοι που αναπτύχθηκαν από τους

γραφείς παρουσιάζονται στη μελέτη αναλυτικά καθώς υπάρχει ένας επαρκής αριθμός πεπραγμένων για κάθε σύμβολο του αλφάβητου σε μια δοκιμασμένη επιγραφή.

Η σύγκριση δύο εγγράφων, προκειμένου να ελεγχθεί εάν έχουν γραφτεί από τον ίδιο γραφέα, είναι εξαντλητική. Με άλλα λόγια, επεξεργάζονται όλα τα ζεύγη υλοποίησης του ίδιου συμβόλου του αλφάβητου στα δύο έγγραφα που συγκρίνονται. Έτσι, για παράδειγμα, ας υποθέσουμε ότι κάποιος πρέπει να αποφασίσει αν δύο χειρόγραφα έχουν γραφτεί από τον ίδιο γραφέα ή όχι. Ας υποθέσουμε, επίσης, ότι υπάρχουν τέσσερις πραγματώσεις ενός ορισμένου συμβόλου της αλφαβήτου στο πρώτο χειρόγραφο και πέντε στο δεύτερο. Στη συνέχεια, η μέθοδος που εισάγεται εδώ, παράγει  $5 \times 5 = 25$  και  $4 \times 5 = 20$  συγκρίσεις, που μπορεί να προσφέρουν μια αρκετά αξιόπιστη σχετική απάντηση, μετά την προτεινόμενη στατιστική επεξεργασία.

Το προηγούμενο έργο των δημιουργών ασχολείται με αρχαίες επιγραφές. Η παρούσα εργασία ασχολείται τόσο με το πρόβλημα των αρχαίων επιγραφών, καθώς και με το πρόβλημα των βυζαντινών κωδίκων. Η ουσιαστική διαφορά μεταξύ αυτών των δύο περιπτώσεων γίνεται εμφανής στο κυρίως κείμενο του χειρόγραφου. Η μέθοδος που προτείνεται εδώ είναι εξαντλητική, σε σχέση με όλες τις υλοποιήσεις που εμφανίζονται σε ένα έγγραφο.

Σε κάθε περίπτωση, οι γραφείς ασχολούνται με το ζήτημα του προσδιορισμού του γραφέα των εγγράφων, για τα οποία δεν απαιτείται προηγούμενη πληροφορία, και καμία πληροφορία δεν είναι γνωστή. Με άλλα λόγια, δεν υπάρχουν καθόλου έγγραφα αναφοράς. Ισοδύναμα, δεν υπάρχει οποιαδήποτε βάση δεδομένων

υποστήριξης. Η μέθοδος που παρουσιάζεται στο άρθρο αυτό προσπαθεί να επιτύχει την αναγνώριση γραφέα με βάση τις ατομικές υλοποιήσεις γραμμάτων και / ή επαναλαμβανόμενων συμπλεγμάτων γραμμάτων. Πιο συγκεκριμένα η εισαχθείσα ανάλυση βασίζεται στους ατομικούς χαρακτήρες ή τμήματα χαρακτήρων.

Πιο αντικειμενικά, η εισαχθείσα προσέγγιση δεν είναι κείμενο-ανεξάρτητη. Ωστόσο, η μόνη προϋπόθεση για την εφαρμογή του συνόλου των ενεργειών για την αναγνώριση γραφέα, που παρουσιάζεται εδώ, είναι η γνώση του συνόλου των θεμελιωδών συμβόλων. Ισοδύναμα, προκειμένου να εφαρμόσουν την εισαχθείσα μέθοδο επαρκώς, αρκεί να είναι σε θέση να διακρίνουν τα διαφορετικά σύμβολα ενός αλφαβήτου, δηλαδή της ελληνικής και λατινικής γλώσσας ή προγενέστερου τέτοιου συνόλου συμβόλων, όπως η Βαβυλωνιακή, η ετρουσκική κ.λπ. Στην πραγματικότητα,, ο μόνος λόγος που χρειάζεται να είναι γνωστή η γλώσσα είναι η ανάγκη να γίνουν συγκρίσεις μόνο μεταξύ των υλοποιήσεων από τα ίδια σύμβολα ή συμπλέγματα συμβόλων.

Από εκεί και πέρα, καμία κείμενο-εξάρτηση δεν υπάρχει στην εισαχθείσα μέθοδο. Όσον αφορά την προσέγγιση που υιοθετήθηκε από τους μελετητές θα πρέπει να γίνουν δύο παρατηρήσεις: α) Θεωρείται ότι το κύριο κείμενο που περιέχεται σε μία πέτρα ή σε ένα πάπυρο έχει γραφτεί από το ίδιο χέρι. β) Κάθε πάπυρος αποτελείται από το κυρίως κείμενο, καθώς και μια σειρά από σχόλια γραμμένα στα πλαίσια της κάθε σελίδας είτε από το πρόσωπο που αντέγραψε το αρχικό κείμενο ή από τους μελετητές / ερευνητές στη συνέχεια. Στην παρούσα εργασία, οι μελετητές αγνοούν

εντελώς τα σχόλια. Ωστόσο, οι γραφείς σκοπεύουν να επιχειρήσουν την ταξινόμηση των παρατηρήσεων στο εγγύς μέλλον.

### 3.2.2 Η προτεινόμενη Μεθοδολογία

Υπάρχουν πάρα πολλές μελέτες και μεθοδολογίες , που έχουν αναπτυχθεί κατά καιρούς, και αφορούν την αναγνώριση γραφών.

Η μεθοδολογία που παρουσιάζετε εδώ μπορεί επίσης να εφαρμοστεί και στην περίπτωση της βέλτιστης αντιστοιχίας και της ταξινόμησης των αυθαίρετων σχημάτων, υπό την προϋπόθεση ότι περιλαμβάνουν μόνο Ευκλείδειους μετασχηματισμούς και κλιμάκωση.

Πράγματι, όσον αφορά την ταξινόμηση με βάση την καμπυλότητα των επίπεδων καμπυλών, αποδεικνύεται ότι η καμπυλότητα της καμπύλης όταν εκφράζεται ως συνάρτηση του μήκους καμπύλης, καθορίζει πλήρως τους διαφορικούς συσχετισμούς αυτής της καμπύλης. Επιπλέον, μια νέα καμπύλη με συντεταγμένες την καμπυλότητα της αρχικής καμπύλης και των παραγώγων της σε σχέση με το μήκος της καμπύλης, είναι αναλλοίωτη σε τοπικούς Ευκλείδειους μετασχηματισμούς. Σε ένα ευρύτερο πλαίσιο, τα αποτελέσματα της προηγούμενης μελέτης, επεκτάθηκαν για να συμπεριλάβουν οποιαδήποτε ομάδα των μετασχηματισμών, όχι απαραίτητα Ευκλείδειους. Με βάση τα αποτελέσματα αυτά, η καμπυλότητα των σημείων ελέγχου μιας καμπύλης (ή, απλώς, των κορυφών του) χρησιμοποιείται για να προσδιοριστούν τοπικά οι Ευκλείδειες σχέσεις ισοδυναμίας μεταξύ ενός ζεύγους καμπύλων. Στη

συνέχεια, η ομοιότητα ενός συνόλου καμπυλών προσδιορίζεται με τη βοήθεια της ομοιότητας των αντίστοιχων σχέσεων ισοδυναμίας.

Στην παρούσα μελέτη, δύο από τους γραφείς (Tracy και Blackwell) επέλεξαν μια σειρά από σημαντικές αρχαίες ελληνικές επιγραφές και βυζαντινούς κώδικες, αντίστοιχα. Στη μελέτη υπάρχουν εικόνες υψηλής ανάλυσης από κάθε τέτοιο έγγραφο. Στη συνέχεια, εξήχθησαν τα μεμονωμένα γράμματα από κάθε χειρόγραφο. Η εξαγωγή επιτεύχθηκε με τη βοήθεια μιας ημι-αυτόματης μεθόδου. Κάθε υλοποίηση γράμματος που εξάγεται από το έγγραφο ενσωματώνεται αμέσως σε ένα ορθογώνιο πλαίσιο. Για να μπορέσει η εισαχθείσα μέθοδος να εφαρμοστεί, το περίγραμμα του κάθε σώματος μίας αυθαίρετης υλοποίησης πρέπει να έχει μια σαφή, καλά καθορισμένη μορφή.

Στη συνέχεια, έγιναν οι ακόλουθες διαδικασίες:

1. Έστω  $C$  το περίγραμμα μιας αυθαίρετης υλοποίησης σύμβολου του αλφάβητου, που εμφανίζεται σε ένα συγκεκριμένο έγγραφο. Χωρίζουμε το  $C$  σε παρακείμενες αλυσίδες pixels, με ελαφρά επικάλυψη. Στη συνέχεια προσεγγίζουν οι μελετητές κάθε τέτοια αλυσίδα των pixels με συνάρτηση κατάλληλου πολυώνυμου των  $x$  και  $y$  συντεταγμένων χωριστά, με ανεξάρτητης μεταβλητής τόξο αλυσίδας μήκους,  $S$ . Μέχρι το τέλος αυτής της διαδικασίας, κάθε περίγραμμα  $C$  ενός συμβόλου του αλφάβητου σχετίζεται μοναδικά με μια ακολουθία τιμών καμπυλότητας, η οποία αντιπροσωπεύει την τάση της καμπυλότητας σε κάθε pixel της  $C$ , με ένα πολύ ικανοποιητικό τρόπο. Οι μελετητές έλαβαν ιδιαίτερη μέριμνα για την εξομάλυνση

των τιμών της ακολουθίας καμπυλότητας, πολλαπλασιάζοντας το κάθε ένα από αυτά με το μήκος του αντίστοιχου περιγράμματος.

2. Έστω τώρα, ότι υπάρχουν δύο καμπύλες  $C_1$ ,  $C_2$  από δύο διαφορετικές πραγματοποιήσεις του ίδιου συμβόλου του αλφάβητου, οι οποίες δημιουργούν δύο διακριτές ακολουθίες κανονικοποιημένων καμπυλοτήτων  $K_1$ ,  $K_2$ . Έχουν αναπτύξει και εφαρμόσει μια μαθηματική προσέγγιση και ένα αντίστοιχο αλγόριθμο, ο οποίος επιτυγχάνει την άμεση βέλτιστη αντιστοίχιση των  $K_1$  και  $K_2$  μέσω της ελαχιστοποίησης από μία σωστά επιλεγμένης συνάρτησης σφάλματος.

3. Σε αυτό το στάδιο, οι μελετητές έχουν εισαγάγει ένα άλλο ουσιαστικά διαφορετικό κριτήριο, το οποίο είναι ιδιαίτερα κατάλληλο για στατιστικούς λόγους. Πιο συγκεκριμένα, θεωρούν δύο υλοποιήσεις του ίδιου συμβόλου του αλφάβητου, που προσπαθούν να το ταιριάξουν άριστα και στην καλύτερη θέση υπολογίζουν την αναλογία της περιοχής της τομής τους με την περιοχή της ένωσής τους. Αυτή η αναλογία είναι τόσο ένα ισχυρό σφάλμα, όσο και μια ποσότητα αρκετά κατάλληλη για την επακόλουθη στατιστική ανάλυση.

4. Χρησιμοποιώντας το κριτήριο που ορίζεται στο προηγούμενο βήμα 3, οι μελετητές έχουν δηλώσει και δοκιμάσει τις στατιστικές υποθέσεις, οι οποίες τελικά έδειξαν τον αριθμό των διαφορετικών γραφών που έγραψαν όλα τα διαθέσιμα έγγραφα. Την ίδια στιγμή, ένα "αντιπροσωπευτικό" έγγραφο σχετίζεται μοναδικά με κάθε γραφέα.

5. Κάθε ένα από τα υπόλοιπα έγγραφα ταξινομείτε στο πιο κατάλληλο χέρι, βάση των νέων αποτελεσμάτων.



Οι μελετητές τόνισαν και σε αυτό το σημείο ότι δεν χρειάζεται εκ των προτέρων καμία γνώση σχετικά με τα έγγραφα για την εφαρμογή αυτής της προσέγγισης. Η μόνη απαίτηση είναι να αναζητηθούν οι συγκρίσεις του περιγράμματος ζευγαριών  $C1$  και  $C2$  που ανήκουν σε υλοποιήσεις του ίδιου συμβόλου αλφάβητο. Πέρα από αυτό, η προτεινόμενη μέθοδος είναι ουσιαστικά περιεχόμενο/κείμενο-ανεξάρτητη. Επιπλέον, η μέθοδος είναι εφαρμόσιμη σε περίγραμμα επαναλαμβανόμενων συμπλεγμάτων γραμμάτων ή σε οποιαδήποτε ανάλογη κατηγορία παρόμοιων δισδιάστατων καμπύλων.

Έστω ότι έχουμε δυο κώδικες ή επιγραφές, έστω  $D1$  και  $D2$ , και ένα αλφαβητικό σύμβολο, έστω  $L$ , του οποίου η υλοποίηση εμφανίζεται και στα δυο κείμενα. Αν επιλέξουμε  $L_i^1$  του  $L$  στο  $D1$  και  $L_j^2$  του  $L$  στο  $D2$ , όπου οι εκθέτες 1,2 συμβολίζουν την επιγραφή ή τον κώδικα και οι δείκτες συμβολίζουν τον απόλυτο αριθμό της υλοποίησης στο συγκεκριμένο κείμενο. Τμηματοποιούμε τις εικόνες των  $L_i^1$  και  $L_j^2$  με την εφαρμογή της μεθόδου που αναλύθηκε σε προηγούμενο κεφάλαιο.

Έπειτα, γίνεται η εξαγωγή των περιγραμμάτων των  $L_i^1$  και  $L_j^2$ , έτσι ώστε να είναι πάντα στη σωστή μορφή: α. Κάθε εικονοστοιχείο του περιγράμματος έχει 2 ακριβώς γειτονικά εικονοστοιχεία. β. Δεν υπάρχουν ορθές γωνίες μεταξύ διαδοχικών εικονοστοιχείων. γ. Δεν υπάρχουν μεμονωμένα εικονοστοιχεία. Για να επιτευχθεί αυτός ο σκοπός εφαρμόζονται κάποια μορφολογικά φίλτρα πάνω στις εξαγμένες εικόνες αυτών των γραμμάτων για να επιτύχουμε τις τρεις αυτές προϋποθέσεις.

Έστω ότι,  $I$  είναι η εικόνα που έχει ενσωματωθεί το γράμμα και  $C$  είναι το περίγραμμα του. Εφαρμόζονται οι ακόλουθες μορφολογικές μετατροπές.

α. Εφαρμογή ρομβοειδούς σχήματος στοιχείου με διαγώνιο μήκος πέντε εικονοστοιχείων, όπου δημιουργείτε η εικόνα  $I^E$ .

β. Εφαρμογή διαστολής στην  $I^E$  μέσω τετραγωνικού σχήματος στοιχείου με πλευρά δυο εικονοστοιχεία, όπου δημιουργείτε η εικόνα  $I^D$ .

γ. Εφαρμογή διαστολής στην  $I^D$  μέσω ενός σταυροειδούς σχήματος στοιχείου με 5 εικονοστοιχεία, σε ύψος και πλάτος, όπου δημιουργείτε η εικόνα  $I^{DC}$ .

δ. Τέλος, παίρνουμε το περίγραμμα του γράμματος αφαιρώντας τα  $I^{DC}$  και  $I^D$ .

Έστω ότι το  $C_i^1$  είναι εξωτερικό περίγραμμα του  $L_i^1$  και το  $C_j^2$  του  $L_j^2$ . Υποθέτετε ότι το καθένα από τα δύο περιγράμματα είναι πολύγωνα Jordan. Σε αρκετές περιπτώσεις το κάθε περίγραμμα είναι μια μονή κλειστή καμπύλη. Σε πολλές άλλες περιπτώσεις το κάθε περίγραμμα συνίσταται από δυο ή περισσότερες καμπύλες

Οι γραφείς που ειδικεύονται στους τομείς των Μηχανικών Ηλεκτρονικών Υπολογιστών και Μαθηματικών έχουν εφαρμόσει τα βήματα της μεθοδολογίας που συζητήθηκαν προηγουμένως. Συγκεκριμένα, έκαναν εξαντλητικές συγκρίσεις ζευγών Bonferroni όλων των επιγραφών, φθάνοντας έτσι στο συμπέρασμα ότι όλες οι 46 επιγραφές που δοκιμάστηκαν στη μελέτη τους έχουν γραφτεί από δέκα (10) διακεκριμένες γραφείς, όπως και όλες οι προηγούμενες μελέτες.

Working ID	Original ID	Working ID	Original ID	Working ID	Original ID	Working ID	Original ID	Working ID	Original ID	Working ID	Original ID
I1	0247	I9	4033	I17	6006	I25	7188	I33	7400	I41	7542
I2	0286	I10	4266	I18	6053	I26	7190	I34	7405	I42	7566
I3	1024	I11	4330	I19	6124	I27	7220	I35	7446	I43	7567
I4	1640	I12	4424	I20	6295	I28	7237	I36	7457	I44	7587
I5	2054	I13	4462	I21	6422	I29	7245	I37	7478	I45	7723
I6	2361	I14	4917	I22	6671	I30	7254	I38	7481	I46	10068
I7	3717	I15	5039	I23	7041	I31	7335	I39	7482		
I8	3855	I16	5297	I24	7156	I32	7398	I40	7519		

### 3.1.3 Συμπεράσματα της Μελέτης

Μια τέτοια ταξινόμηση όπως αυτή που χρησιμοποιήθηκε στο άρθρο μπορεί να γίνει ένα πραγματικά ισχυρό εργαλείο για την ακριβή χρονολόγηση των αντίστοιχων χειρόγραφων. Η μέθοδος που εισήχθη απασχολεί μεμονωμένες παραγωγές σύμβολων του αλφάβητου ή/και συνεχόμενα μπλοκ συμβόλων/γραμμάτων. Η μέθοδος που εισήχθη εξάγει πρώτα το περίγραμμα των συμβόλων του αλφαβήτου που εμφανίζονται στο κάθε χειρόγραφο. Εκτελεί ζεύγη συγκρίσεων των περιγραμμάτων του κάθε σύμβολου που εμφανίζεται σε κάθε θεωρημένο χειρόγραφο με εκτιμήσεις καμπυλότητας. Εισάγει μια νέα πρόταση που προσφέρει μία λύση στο πρόβλημα της βέλτιστης αντιστοίχισης δύο περιγραμμάτων που αποτελούνται από ίσο αριθμό πίξελ. Σε αυτήν την θέση ταιριάσματος, η μέθοδος πραγματοποιεί μια μικρορύθμιση και εισάγει ένα κριτήριο ομοιότητας μεταξύ δύο παραγωγών του ίδιου σύμβολου του αλφάβητου.

Τέλος, οι μελετητές εισήγαγαν έναν αριθμό νέων στατιστικών κριτηρίων προκειμένου να επιτύχουν την αναγνώριση γραφέν. Οι μελετητές τόνισαν ότι η προαναφερθείσα

προσέγγιση αποδίδει αρκετά καλά, ακόμη και αν ένας μικρός αριθμός επιτευγμάτων (λιγότερο από 6) ορισμένων συμβόλων αλφαβήτου εμφανίζονται σε ένα έγγραφο ταξινομημένα. Η μέθοδος έχει εφαρμοστεί σε 46 αρχαίες επιγραφές της κλασικής και ελληνιστικής εποχής και 23 βυζαντινούς κώδικες. Η μέθοδος που εισήχθη και το σχετικό σύστημα που αναπτύχθηκε δεν έχει κανένα χειρόγραφο αναφοράς. Ούτε ο αριθμός των διακριτών χειρών που έχουν γράψει τα χειρόγραφα θεωρείται γνωστός, όπως επίσης και καμία άλλη σχετική πληροφορία, εκτός από την ομαδοποίηση των επιτευγμάτων του ίδιου συμβόλου ή της ακολουθίας συμβόλων. Η μέθοδος ταξινομεί τις 46 επιγραφές σε 10 διαφορετικά χέρια και τα 23 βυζαντινούς κώδικες σε 4 διαφορετικούς γραφείς. Οι διακεκριμένοι μελετητές στην Επιγραφολογία, την αρχαιολογία και τις Κλασικές Σπουδές συμφωνούν πλήρως με την κατάταξη που προσφέρονται από το σύστημα. Οι γραφείς σκοπεύουν να επεκτείνουν τα αποτελέσματα από την εφαρμογή της μεθόδου σε ακόμη μεγαλύτερο αριθμό επιγραφών και βυζαντινούς κώδικες και να αντιμετωπίσουν διαφορετικά θέματα που σχετίζονται με άλλες απόψεις και διαφωνίες.

## 4 Έρευνα

Χρησιμοποιώντας τις συναρτήσεις που εξηγούνται στην διπλωματική της Ευδοκίας Κάλφα(2012) περί κατάτμησης και επεξεργασίας εικόνας σε αρχαίες περιγραφές, εξάγονται τα περιγράμματα γραμμάτων από 5 διαφορετικούς γραφείς. Στη συνέχεια γίνεται σύγκριση των όλων περιγραμμάτων ενός γραφέα με τα περιγράμματα του ίδιου γράμματος που έχουν εξαχθεί από έναν άλλο γραφέα.

Αρχικά για κάθε γράμμα αντικαθίστανται τα μεμονωμένα pixels που υπάρχουν είτε γύρω από το γράμμα (μπορεί να υπάρχουν μαύρα pixels, τα οποία βρίσκονται εκεί κατα λάθος και αντικαθίστανται με άσπρα) είτε μέσα στο γράμμα (άσπρα pixels τα οποία θα έπρεπε να ήταν μαύρα). Στη συνέχεια εντοπίζεται το περίγραμμα με τον παρακάτω τρόπο. Πρώτον, για κάθε άσπρο pixel ελέγχεται αν υπάρχει μαύρο pixel γύρω του (πάνω, κάτω, αριστερά, δεξιά) και αν υπάρχει τότε αυτό αποτελεί μέρος του περιγράμματος. Τέλος, εντοπίζονται και αφαιρούνται οι ορθές γωνίες από το παραπάνω περίγραμμα. Εντοπίζονται δηλαδή αν υπάρχουν τρία συνεχόμενα pixels τα οποία σχηματίζουν γάμα μεταξύ τους και διαγράφεται το pixel που βρίσκεται στην κορυφή της γωνίας (μεσαίο).

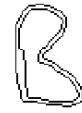
Παρακάτω φαίνονται τα περιγράμματα από κάποια από τα γράμματα που επεξεργάστηκαν.



arxiko gramma



stadio 1



stadio 2



arxiko gramma



stadio 1



stadio 2



arxiko gramma



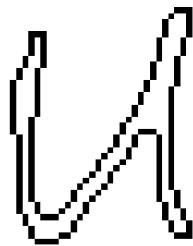
stadio 1



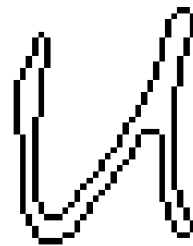
stadio 2



arxiko gramma



stadio 1



stadio 2



arxiko gramma



stadio 1



stadio 2



arxiko gramma



stadio 1



stadio 2



arxiko gramma



stadio 1



stadio 2



arxiko gramma



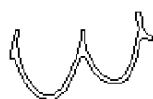
stadio 1



stadio 2



arxiko gramma



stadio 1

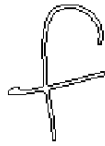


stadio 2





arxiko gramma



stadio 1



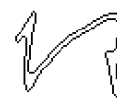
stadio 2



arxiko gramma



stadio 1



stadio 2



arxiko gramma



stadio 1



stadio 2



arxiko gramma



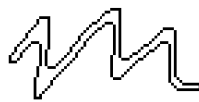
stadio 1



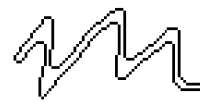
stadio 2



arxiko gramma



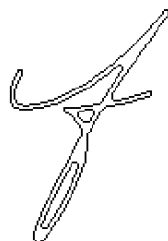
stadio 1



stadio 2



arxiko gramma



stadio 1



stadio 2



arxiko gramma



stadio 1



stadio 2



arxiko gramma



stadio 1



stadio 2



arxiko gramma



stadio 1



stadio 2

Στη συνέχεια, για κάθε γράμμα, συγκρίνουμε όλα τα δείγματα που έχουμε από έναν γραφέα, με όλα τα δείγματα από κάθε έναν διαφορετικό γραφέα άλλα και τα δείγματα ένας γραφέα μεταξύ τους. Η σύγκριση των περιγραμμάτων γίνεται με τον παρακάτω τρόπο. Αρχικά γίνεται μετακίνηση ούτως ώστε να έχουν το ίδιο κέντρο βάρους. Στη συνέχεια γίνεται περιστροφή στο ένα από τα δύο δείγματα ούτως ώστε να έχουν όσο το δυνατόν περισσότερα κοινά σημεία. Τέλος μεγεθύνεται ή σμικρύνεται το ένα δείγμα ώστε να έχουν κοινά όσο περισσότερα σημεία είναι δυνατόν. Η διαδικασία αυτή επαναλαμβάνεται έως ότου δεν υπάρχει καλύτερο ταίριασμα μεταξύ των δύο δειγμάτων-γραμμάτων (όπου ταίριασμα είναι το κλάσμα με αριθμητή την τομή των περιγραμμάτων των 2 γραμμάτων και παρονομαστή την ένωση αυτών). Μετά την ολοκλήρωση της παραπάνω διαδικασίας για όλες τις συγκρίσεις δημιουργείτε ένας δυσδιάστατος πίνακας με τα αποτελέσματα των συγκρίσεων.

Παρακάτω φαίνονται δύο πίνακες από την σύγκριση. Ο πρώτος δείχνει μία σύγκριση μεταξύ δύο διαφορετικών γραφών και ο δεύτερος δείχνει τις συγκρίσεις μεταξύ των γραμμάτων ίδιου γραφέα. Είναι εμφανές η ομοιότητα μεταξύ γραμμάτων ίδιο γραφέα αντίθετα από γράμματα που έχουν παραχθεί από διαφορετικά χέρια. Τα αποτελέσματα των συγκρίσεων έχουν στρογγυλοποιηθεί στο τρίτο δεκαδικό ψηφίο.

0,207	0,152	0,119	0,206	0,291	0,258	0,331	0,199	0,275	0,333	0,331	0,186
0,339	0,237	0,264	0,348	0,385	0,333	0,262	0,304	0,404	0,431	0,369	0,262
0,333	0,385	0,222	0,314	0,422	0,473	0,281	0,276	0,570	0,575	0,378	0,234
0,281	0,198	0,187	0,279	0,344	0,332	0,271	0,316	0,405	0,379	0,390	0,234
0,278	0,195	0,195	0,234	0,322	0,318	0,243	0,292	0,389	0,357	0,383	0,238
0,272	0,312	0,278	0,249	0,296	0,280	0,229	0,276	0,315	0,325	0,331	0,267
0,249	0,182	0,169	0,252	0,333	0,280	0,368	0,307	0,379	0,428	0,434	0,219
0,183	0,165	0,142	0,212	0,260	0,217	0,208	0,183	0,209	0,307	0,387	0,197
0,443	0,346	0,272	0,481	0,605	0,417	0,267	0,449	0,565	0,421	0,456	0,325
0,197	0,133	0,141	0,242	0,331	0,291	0,289	0,239	0,373	0,435	0,366	0,255

1,000	0,326	0,296	0,390	0,358	0,334	0,495	0,384	0,249	0,459
0,306	1,000	0,310	0,539	0,570	0,503	0,457	0,550	0,492	0,541
0,307	0,311	1,000	0,321	0,319	0,317	0,313	0,283	0,455	0,367
0,385	0,551	0,327	1,000	0,665	0,430	0,576	0,639	0,415	0,661
0,363	0,556	0,327	0,678	1,000	0,437	0,467	0,682	0,409	0,594
0,324	0,455	0,307	0,422	0,429	1,000	0,455	0,420	0,352	0,472
0,491	0,430	0,299	0,529	0,461	0,373	1,000	0,554	0,410	0,588
0,370	0,569	0,306	0,565	0,721	0,397	0,589	1,000	0,396	0,547
0,256	0,493	0,467	0,417	0,414	0,370	0,382	0,404	1,000	0,438
0,475	0,540	0,383	0,618	0,515	0,481	0,591	0,546	0,384	1,000

## 5 Συμπεράσματα-Επίλογος

Παρά το γεγονός ότι υπάρχει μια τάση για σύγχρονο εξοπλισμό, όπως τα τάνπλετ, μαζί με την έννοια ενός κόσμου χωρίς χαρτί, τα χειρόγραφα έγγραφα εξακολουθούν να παραμένουν σημαντικά, ιδιαίτερα σχετικά με νομικά και ιστορικά θέματα. Λόγω αυτού του γεγονότος, η εξακρίβωση της γνησιότητας των χειρόγραφων κείμενων και υπογραφών έχει πράγματι υπάρξει μια ανοιχτή περιοχή της έρευνας. Οι ερευνητές έχουν προσπαθήσει να προτείνουν ισχυρά συστήματα πιστοποίησης της ταυτότητας γραφέων για τα χειρόγραφα μέρη ενός εγγράφου. Σε αντίθεση με το τυπωμένο κείμενο, το χειρόγραφο κείμενο μας παρέχει επιπλέον πληροφορίες, όπως μια βαθύτερη κατανόηση για την προσωπικότητα του γραφέα. Το ατομικό στυλ γραφής είναι συνήθως σταθερό και συνεπές για μεγάλο χρονικό διάστημα. Αναλύοντας τους χαρακτήρες μας δίνετε η ευκαιρία να εντοπιστεί ο συντάκτης ενός κειμένου. Η έρευνα βρίσκεται σε εξέλιξη στον τομέα αυτό για πάνω από μια δεκαετία τώρα, αλλά τα επιθυμητά αποτελέσματα δεν έχουν ακόμη επιτευχθεί. Ένα άλλο ζήτημα που προκύπτει είναι ότι οι διαφορετικές γλώσσες, έχουν διαφορετικά εγγενή χαρακτηριστικά. Κατά συνέπεια, είναι πολύ δύσκολο να προτείνει κανείς ένα σύστημα που λειτουργεί για όλες τις γλώσσες. Σε αυτό το σημείο η έρευνα των Paraodysseus et al, (2014), είναι εξαιρετικά σημαντική, γιατί διατυπώνει την υπόθεση ότι αυτή η μέθοδος μπορεί να χρησιμοποιηθεί ανεξάρτητα από τη γλώσσα.

Από τα αποτελέσματα της δικής μου εργασίας προκύπτει ότι η ηλεκτρονική ταυτοποίηση γραφέα με την μέθοδο της εξαγωγής περιγραμμάτων για κάθε γράμμα

ξεχωριστά και στην συνέχεια τις σύγκρισης των περιγραμμάτων που έχουν εξαχθεί από ένα κείμενο με ενός άλλου κείμενου, προκύπτει ότι υπάρχει ένα αρκετά μεγάλο ποσοστό επιτυχίας το οποίο μπορεί να μας οδηγήσει σχεδόν με σιγουριά στην ταυτοποίηση του γραφέα ενός κειμένου. Αυτό μπορεί να επαληθευτεί συγκρίνοντας τα αποτελέσματα από αυτή την μέθοδο με αποτελέσματα που έχουν εξαχθεί από άλλες μεθόδους.

Επειδή η μέθοδος εξαρτάται μόνο από τους ξεχωριστούς χαρακτήρες είναι δυνατό να εφαρμοστεί σε όλους τις περιπτώσεις που υπάρχουν χειρόγραφα έγγραφα, από την αναγνώριση αρχαίων γραφών και γραφών μέχρι την εγκληματολογία και την ιατροδικαστική. Ακόμα χωρίς καμία αλλαγή μπορεί να χρησιμοποιηθεί για την ταυτοποίηση χειρόγραφων εγγράφων που έχουν γραφεί ηλεκτρονικά (πχ. ψηφιακό στυλό)

## Βιβλιογραφία

Antonini, M. Barlaud, P. Mathieu, and I. Daubechies. Image coding using wavelet transform. Image Processing, IEEE Transactions on, 1(2):205–220, 1992.

Arabadjis D, F. Giannopoulos , C. Papaodysseus , S. Zannos<sup>1</sup> , P. Rousopoulos , M. Panagopoulos and C. Blackwell (2013) New Mathematical and Algorithmic Schemes for Pattern Classification with Application to the Identification of Writers of Important Ancient Documents, Pattern Recognition 46, 2278-2296.

Ευδοκία Κάλφα. Συγκριτική μελέτη μεθόδων κατάτμησης και επεξεργασίας εικόνας με εφαρμογή σε αρχαίες επιγραφές(2012)

Arazi. Handwriting identification by means of run-length measurements. IEEE Trans. Syst., Man and Cybernetics, 7(12):878–881, 1977.

Bulacu and L. Schomaker. Combining Multiple Features for Text-Independent Writer Identification and Verification. pages 281–286, 2006.

Bulacu and L. Schomaker. Writer style from oriented edge fragments. In Proc. of the 10th Int. Conference on Computer Analysis of Images and Patterns, pages 460–469, 2003. M.

Gupta, S. (2008). Automatic Person Identification and Verification using Online Handwriting. Master Thesis. International Institute of Information Technology Hyderabad, India



Hertel, C., Bunke, H. (2003) —A Set of Novel Features for Writer Identification,|| Proc. Fourth Int'l Conf. Audio and Video-Based Biometric Person Authentication, pp. 679-687

Leclerc, F., Plamondon, R. (1994) Automatic signature verification: The state of the art 1989-1993. In Progress in Automatic Signature Verification edited by R. Plamondon, World Scientific Publ. Co., pp. 13-19.

Marti, U.V., Messerli, R., Bunke, H. (2001) —Writer Identification Using Text Line Based Features,|| Proc. Sixth Int'l Conf. Document Analysis and Recognition (ICDAR), pp. 101-105.

Papaodysseus, C., Rousopoulos, P., Giannopoulos, F., Zannos, S., Arabadjis, D., Panagopoulos, M., Kalfa, E., Blackwell, C., Tracy, S.: Identifying the writer of ancient inscriptions and Byzantine codices. A novel approach. Computer Vision and Image Understanding 121 (2014) 57–73

Plamondon, R., Lorette, G. (1989) —Automatic Signature Verification and Writer Identification—The State of the Art,|| Pattern Recognition, vol. 22, no. 2, pp. 107-131.  
[6]

Said, G.S. Peake, T.N. Tan and K.D. Baker, (1998) Writer identification from non-uniformly skewed handwriting images . Proceedings of the Ninth British Machine Vision Conference (BMVC 98), Southampton, England,, vol.2, pp. 478–487,

Schlapbach, A., Bunke, H. (2005) Writer identification using an HMM-based handwriting recognition system: to normalize the input or not? In: 12th Conference of the International Graphonomics Society, Salerno, Italy, June 26–29, pp.138–142.

Schlapbach, A., Bunke, H. (2007) A writer identification and verification system using HMM based recognizers, *Pattern Analysis Application* (Springer)10,33–43, doi:10.1007/s10044-006-0047-5.

Shahabi and M. Rahmati. A New Method for Writer Identification of Handwritten Farsi Documents. In 10th International Conference on Document Analysis and Recognition, pages 426–430, 2009. doi: 10.1109/ICDAR.2009.290.

Srihari, C. I. Tomai, Bin Zhang, and Sangjik Lee. Individuality of numerals. In 7th International Conference on Document Analysis and Recognition, pages 1096–1100, 2003.

Srihari, S. H. Cha, H. Arora, and S. Lee. Individuality of handwriting. *Journal of Forensic Sciences*, pages 1–17, 2002.

Terzija and Walter Geisselhardt. Digital image watermarking using complex wavelet transform. In *Proceedings of the 2004 Workshop on Multimedia and Security*, pages 193–198, New York, NY, USA, 2004. ACM.