



**ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ  
ΠΛΗΡΟΦΟΡΙΚΗ ΜΕ ΕΦΑΡΜΟΓΕΣ ΣΤΗ ΒΙΟΙΑΤΡΙΚΗ**

**Μορφομετρικός – μορφολογικός  
χαρακτηρισμός εικόνων με εφαρμογές  
στην Ιατρική και στη Βιολογία**

**ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ**

Φοιτητής : Λημναίος Αδαμάντιος

Αριθμός Μητρώου: ΠΒ0086

Επιβλέπων: Ιακωβίδης Δημήτριος

Λαμία, Οκτώβριος 2013



**ΠΑΝΕΠΙΣΤΗΜΙΟ  
ΘΕΣΣΑΛΙΑΣ**

ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

ΠΛΗΡΟΦΟΡΙΚΗ ΜΕ ΕΦΑΡΜΟΓΕΣ ΣΤΗ ΒΙΟΙΑΤΡΙΚΗ

**Μορφομετρικός – μορφολογικός  
χαρακτηρισμός εικόνων με εφαρμογές  
στην Ιατρική και στη Βιολογία**

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

Φοιτητής : Λημναίος Αδαμάντιος

Αριθμός Μητρώου: ΠΒ0086

Επιβλέπων: Ιακωβίδης Δημήτριος

Λαμία , Οκτώβριος 2013

*στην οικογένεια μου...*

© Λημναίος Αδαμάντιος

Διεύθυνση: Άνθιμου Γαζή 115 , Κερατσίνι , ΤΚ:18758

Τηλέφωνο: 6984788213

Email: adamantios13@hotmail.com

# Πίνακας Περιεχομένων

<b>ΠΕΡΙΛΗΨΗ.....</b>	
<b>ABSTRACT.....</b>	
<b>ΕΙΣΑΓΩΓΗ .....</b>	
<b>I. ΕΥΧΑΡΙΣΤΙΕΣ .....</b>	
<b>ΚΕΦΑΛΑΙΟ 1: ΕΙΣΑΓΩΓΗ.....</b>	
1.1 ΨΗΦΙΑΚΗ ΕΠΕΞΕΡΓΑΣΙΑ ΕΙΚΟΝΑΣ.....	
1.2 ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΕΙΚΟΝΑΣ .....	
1.3 ΕΝΔΟΣΚΟΠΙΚΗ ΚΑΨΟΥΛΑ .....	
1.4 ΕΠΕΞΕΡΓΑΣΙΑ ΚΑΙ ΑΝΑΛΥΣΗ ΕΝΔΟΣΚΟΠΙΚΩΝ ΔΕΔΟΜΕΝΩΝ.....	
1.5 ΑΝΤΙΚΕΙΜΕΝΙΚΟΙ ΣΤΟΧΟΙ.....	
1.6 ΔΟΜΗ ΕΡΓΑΣΙΑΣ .....	
<b>ΚΕΦΑΛΑΙΟ 2: ΑΝΑΛΥΣΗ ΕΙΚΟΝΑΣ ΚΑΙ ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΤΥΠΩΝ .....</b>	
2.1 ΕΙΣΑΓΩΓΗ ΣΤΗΝ ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΤΥΠΩΝ .....	
2.1.1 ΒΑΣΙΚΗ ΔΙΑΔΙΚΑΣΙΑ.....	
2.1.2 ΠΡΟΕΠΕΞΕΡΓΑΣΙΑ ΕΙΚΟΝΑΣ.....	
2.1.3 ΕΞΑΓΩΓΗ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ.....	
2.1.3.1 ΕΞΑΓΩΓΗ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ ΑΠΟ ΕΙΚΟΝΕΣ.....	
2.1.3.2 ΓΕΝΙΚΕΣ ΜΕΘΟΔΟΙ ΕΞΑΓΩΓΗΣ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ.....	
2.1.4 ΕΠΙΛΟΓΗ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ.....	
2.1.5 ΤΑΞΙΝΟΜΗΣΗ ΣΕ ΚΛΑΣΕΙΣ.....	
2.2 ΑΝΑΚΛΗΣΗ ΕΙΚΟΝΩΝ ΒΑΣΕΙ ΠΕΡΙΕΧΟΜΕΝΟΥ.....	
2.3 ΒΙΒΛΙΟΓΡΑΦΙΚΗ ΑΝΑΣΚΟΠΗΣΗ ΜΕΘΟΔΩΝ ΕΞΑΓΩΓΗΣ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ.....	
2.4 ΕΠΙΛΟΓΗ ΠΕΡΙΓΡΑΦΕΩΝ FOURIER.....	

**ΚΕΦΑΛΑΙΟ 3: FOURIER DESCRIPTORS .....**

3.1 ΑΝΑΠΤΥΓΜΑ ΣΥΝΑΡΤΗΣΗΣ ΣΕ ΣΕΙΡΑ FOURIER.....

3.1.1 ΕΚΘΕΤΙΚΗ ΣΕΙΡΑ FOURIER.....

3.1.2 ΤΡΙΓΩΝΟΜΕΤΡΙΚΗ ΣΕΙΡΑ FOURIER.....

3.2 ELLIPTIC FOURIER DESCRIPTORS (EFD).....

3.2.1 ΟΡΙΣΜΟΣ ΤΩΝ EFD.....

3.2.2 ΑΜΕΤΑΒΛΗΤΗ ΣΥΜΠΕΡΙΦΟΡΑ ΤΩΝ EFD ΣΕ ΓΕΩΜΕΤΡΙΚΕΣ ΑΛΛΑΓΕΣ.....

3.3 ΤΑΞΙΝΟΜΗΣΗ.....

**ΚΕΦΑΛΑΙΟ 4: ΠΕΙΡΑΜΑΤΙΚΟ ΜΕΡΟΣ.....**

4.1 ΕΠΙΛΟΓΗ ΒΑΣΕΩΝ ΔΕΔΟΜΕΝΩΝ.....

4.1.1 ΒΑΣΕΙΣ ΔΕΔΟΜΕΝΩ ΕΙΚΟΝΩΝ MPEG7 CE SHAPE-1 PART B.....

4.1.2 ΒΑΣΕΙΣ ΕΙΚΟΝΩΝ SSID -1.....

4.1.3 ΒΑΣΕΙΣ ΕΙΚΟΝΩΝ SSID -2.....

4.1.4 ΕΦΑΡΜΟΓΗ ΤΩΝ EFD ΣΕ ΣΥΓΚΕΚΡΙΜΕΝΑ ΓΕΩΜΕΤΡΙΚΑ ΣΧΗΜΑΤΑ.....

4.1.5 ΧΡΗΣΗ ΤΗΣ ΤΕΧΝΙΚΗΣ SAMMON'S MAPPING

4.1.6 ΒΑΣΗ ΕΙΚΟΝΩΝ ΜΕ ΓΕΩΜΕΤΡΙΚΑ ΣΧΗΜΑΤΑ.....

4.1.7 ΔΗΜΙΟΥΡΓΙΑ ΒΑΣΗΣ ΕΙΚΟΝΩΝ ΜΕ ΙΑΤΡΙΚΟ ΠΕΡΙΕΧΟΜΕΝΟ.....

4.2 ΠΡΟΕΠΕΞΕΡΓΑΣΙΑ ΕΙΚΟΝΩΝ.....

4.3 ΕΞΑΓΩΓΗ ΔΙΑΝΥΣΜΑΤΟΣ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ.....

4.4 ΜΕΘΟΔΟΣ ΕΠΙΚΥΡΩΣΗΣ.....

4.5 ΤΑΞΙΝΟΜΗΣΗ.....

**ΚΕΦΑΛΑΙΟ 5: ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΠΡΟΟΠΤΙΚΕΣ.....**

**ΒΙΒΛΙΟΓΡΑΦΙΑ.....**

## ΠΕΡΙΛΗΨΗ

Σκοπός της παρούσας πτυχιακής εργασίας είναι η ανάπτυξη μεθοδολογίας ανάλυσης ιατρικών εικόνων αξιοποιώντας χαρακτηριστικά σχήματος και η εφαρμογή της για την επίλυση πραγματικού ιατρικού προβλήματος και την υποβοήθηση της ιατρικής διάγνωσης. Η μεθοδολογία εφαρμόστηκε με επιτυχία για το χαρακτηρισμό ελκών του λεπτού εντέρου από βίντεο ενδοσκοπικής κάψουλας. Η πρωτοτυπία της εργασίας αυτής έγκειται στη εφαρμογή των ελλειπτικών περιγραφέων σχήματος Fourier (Elliptic Fourier Descriptors, EFD) και μεθόδων μηχανικής μάθησης για την ταξινόμηση των ελκών σε κατηγορίες. Τα βίντεο που χρησιμοποιήθηκαν προέρχονται από δημόσια βάση δεδομένων και επισημειώθηκαν από ειδικούς. Για την υλοποίηση αξιοποιήθηκαν αλγόριθμοι γραμμένοι σε Matlab και Java. Πραγματοποιήθηκε εκτενής πειραματική μελέτη τόσο σε ιατρικά όσο και σε τεχνητά δεδομένα για τη μελέτη και επιλογή των κατάλληλων αλγορίθμων. Η επιτυχία με την οποία διαχωρίζονται τα έλκη κυμαίνεται από έως

## ABSTRACT

The scope of the present thesis is the development of an analytic methodology for medical imagery by utilizing schematic characteristics and implementing it in order to solve actual medical issues and assisting medical diagnosis. This methodology was successfully applied to characterize small intestine ulcers using endoscopic capsule video. The originality of this work lies in the application of Elliptic Fourier Descriptors (EFD) and methods of artificial learning to allow taxonomy of ulcers into categories. Videos used, are taken from a public database, and annotated by specialists. Algorithms written in Matlab and Java were used for the materialization of the project. Extensive experimental study was carried out in medical as well as artificial data in order to study and choose the proper algorithms. Success in categorizing ulcers ranges from ..... to .....

## ΚΕΦΑΛΑΙΟ 1: ΕΙΣΑΓΩΓΗ

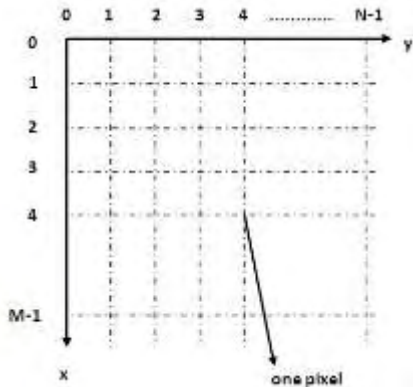
### 1.1 Ψηφιακή επεξεργασία εικόνας

Ξεκινώντας, θα θέλαμε να ορίσουμε την έννοια της ψηφιακής εικόνας (digital image) και αντίστοιχα τον τρόπο με τον οποίο την επεξεργαζόμαστε. Η εικόνα, λοιπόν, αποτελεί ένα πλούτο πληροφοριών, όπως σωστά αναφέρεται και σε μία κινέζικη παροιμία, η οποία χαρακτηριστικά λέει «Μία εικόνα αξίζει όσο χίλιες λέξεις». Με τη βοήθεια της όρασης, δηλαδή με την καταγραφή και ερμηνεία των εικόνων που βλέπουμε στο άμεσό μας περιβάλλον, είμαστε σε θέση να κινούμαστε, να προγραμματίζουμε τις ενέργειές μας, ακόμη και να αναπτύσσουμε τη νοημοσύνη μας. Σε κοινωνικό επίπεδο, η εικόνα διαδραματίζει σημαντικό ρόλο, καθώς σε αυτή στηρίζονται τα περισσότερα από τα σημερινά μέσα μαζικής επικοινωνίας και ενημέρωσης.

Ο τεράστιος αυτός όγκος πληροφοριών και η ανάγκη αποθήκευσης και επεξεργασίας τους οδήγησε τους επιστήμονες και τεχνικούς στην εξεύρεση μέσων ψηφιακής αποθήκευσης της εικόνας και στη συνέχεια επεξεργασίας της με τη βοήθεια ηλεκτρονικών υπολογιστών. Με τον τρόπο αυτό, αναπτύχθηκε ένας εντελώς νέος κλάδος της Πληροφορικής, που ονομάζεται ψηφιακή επεξεργασία εικόνας (digital image processing) [Πήτα01].

Προτού, όμως, αναλύσουμε την έννοια της ψηφιακής εικόνας, θα πρέπει να ορίσουμε την αναλογική εικόνα. Η εικόνα, λοιπόν, στην αναλογική της μορφή, είναι ένα δισδιάστατο σήμα  $x_a(t_1, t_2)$ , όπου  $t_1, t_2$  είναι οι δύο ορθογώνιες συντεταγμένες του επιπέδου και συνήθως αναφέρονται ως  $x, y$ . Το σήμα αυτό λαμβάνει τιμή για κάθε διαφορετικό ζευγάρι των  $(x, y)$ , ενώ η τιμή αυτή είναι ανάλογη της φωτεινότητας της εικόνας στο συγκεκριμένο σημείο και ονομάζεται ένταση στο επίπεδο του γκρι. Ως παραδείγματα αναφέρουμε τη φωτεινότητα ενός φιλμ ή το ηλεκτρικό ρεύμα σε μία ηλεκτρονική κάμερα. Πραγματοποιώντας, στο σημείο αυτό, τη διαδικασία ψηφιοποίησης του δισδιάστατου αυτού σήματος, προκύπτει η ψηφιακή εικόνα [Petr10].

Στο παρακάτω σχήμα (εικόνα 1) αναπαριστάται ο πιο απλός τρόπος ψηφιοποίησης του δισδιάστατου σήματος, που είναι η ομοιόμορφη δειγματοληψία του κατά μήκος των δύο ορθογώνιων αξόνων  $t_1, t_2$ . Ορίζοντας  $x_a(t_1, t_2)$  το αναλογικό σήμα και  $T_1, T_2$  τα διαστήματα δειγματοληψίας κατά μήκος των δύο αυτών αξόνων, το διακριτό σήμα  $x(n_1, n_2)$  δίνεται από τη σχέση  $x(n_1, n_2) = x_a(n_1 T_1, n_2 T_2)$ .



**Εικόνα 1**

Σε κάθε ψηφιακή εικόνα διακρίνουμε τα επιμέρους εικονοστοιχεία (pixels), τα οποία αποτελούν ένα πεπερασμένο αριθμό στοιχείων, με συγκεκριμένη τοποθεσία και τιμή. Θα πρέπει να αναφέρουμε ότι τα δύο βασικά χαρακτηριστικά της ψηφιακής εικόνας είναι:

- τα διαφορετικά επίπεδα φωτεινότητας (gray levels) που αυτή απεικονίζει
- η χωρική διακριτική ικανότητα της εικόνας (spatial resolution)

Τα διαφορετικά επίπεδα φωτεινότητας προκύπτουν από το πλήθος των διαφορετικών τιμών που μπορεί να πάρει κάθε εικονοστοιχείο. Το χαρακτηριστικό αυτό είναι εμφανές στο ανθρώπινο μάτι και η ποιότητα της εικόνας χαλάει αρκετά, όταν περιοριστούμε σε μικρό πλήθος δεπιπέδων φωτεινότητας, για παράδειγμα λιγότερο από 15. Στο παρακάτω (εικόνα 2) σχήμα απεικονίζεται μία εικόνα 256 επιπέδων γκρι και οι αντίστοιχες εικόνες με λιγότερα επίπεδα φωτεινότητας. Η διαφορά στον εντοπισμό λεπτομερειών της εικόνας είναι προφανής.



**Εικόνα 2**

Η χωρική διακριτική ικανότητα της εικόνας αντιστοιχίζεται στο μήκος της πλευρά κάθε εικονοστοιχείου. Προφανώς, όσο μικρότερο το μήκος της πλευράς κάθε εικονοστοιχείου, τόσο καλύτερης ποιότητας θεωρείται η εικόνα, εφόσον μπορούμε να εντοπίσουμε περισσότερες λεπτομέρειες σε αυτήν. Στο



παρακάτω σχήμα (εικόνα 3) παρατηρούμε μία αρχική εικόνα και τη μετατροπή της σε εικόνα υποτετραπλάσιας χωρικής διακριτικής ικανότητας. Η διαφορά είναι εμφανής.



Εικόνα 3

Μέχρι στιγμής έχουμε ασχοληθεί με τις μονοχρωματικές εικόνες, όπου η φωτεινότητα κάθε εικονοστοιχείου λαμβάνει μία τιμή και αντιστοιχεί στο επίπεδο του γκρι. Στις έγχρωμες, όμως, εικόνες, η τιμή κάθε εικονοστοιχείου είναι ένα ξεχωριστό διάνυσμα, που περιέχει τις τιμές κάθε μίας από τις συνιστώσες, που συνθέτουν το εκάστοτε χρωματικό μοντέλο [Gonz07]. Ένα από τα πλέον διαδεδομένα χρωματικά μοντέλα είναι το RGB, το οποίο απαρτίζεται από 3 επιμέρους συνιστώσες, κάθε μία από τις οποίες αντιστοιχεί στη φωτεινότητα του κόκκινου (Red), του πράσινου (Green) και του μπλε (blue) καναλιού. Συνθέτοντας και τις 3 αυτές τιμές λαμβάνουμε την τελική (έγχρωμη) τιμή του εικονοστοιχείου. Στο παρακάτω σχήμα (εικόνα 4) παρατηρούμε τον τρόπο, με τον οποίο συνθέτουμε διάφορα χρώματα χρησιμοποιώντας τα 3 αυτά χρώματα του RGB μοντέλου. Άλλα γνωστά χρωματικά μοντέλα είναι τα CMY και HSV. Αναλυτικότερα για τα χρωματικά μοντέλα θα μιλήσουμε στην επόμενη ενότητα.



Εικόνα 4

Οι εικόνες ως μέσο μετάδοσης πληροφορίας παρουσιάστηκαν για πρώτη φορά στις αρχές του 20<sup>ου</sup> αιώνα και πιο συγκεκριμένα τη δεκαετία του 1920. Τότε, υπήρξε η ανάγκη μετάδοσης ψηφιακών εικόνων μεταξύ Νέας Υόρκης και Λονδίνου, με σκοπό τη δημοσίευσή τους στις τότε εφημερίδες. Προφανώς, η γρήγορη μετάδοση της πληροφορίας μέσω του υποβρύχιου καλωδίου υπερτερούσε σε σχέση με τις άλλες μορφές της εποχής. Στο παρακάτω σχήμα (εικόνα 5) απεικονίζουμε μερικές από τις εικόνες που μεταδόθηκαν με τον τρόπο αυτό εκείνη την εποχή [McFa72].



**Εικόνα 5**

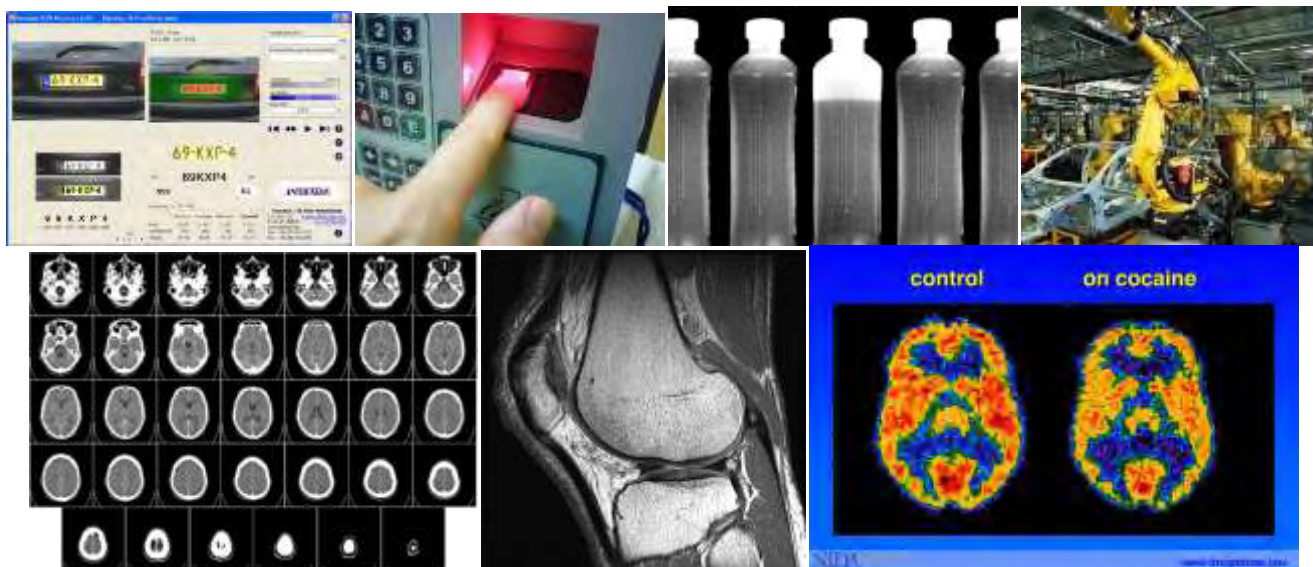
Στη συνέχεια, και για να περάσουμε στην έννοια της ψηφιακής επεξεργασίας της εικόνας, θα πρέπει να αναφερθούμε σε κάποιου είδους επεξεργασία, η οποία εφαρμόζεται στις ψηφιακές εικόνες, με τη βοήθεια ενός υπολογιστικού συστήματος. Σήμερα, στον όρο αυτό συμπεριλαμβάνουμε την επεξεργασία της εικόνας με σκοπό τη μείωση του θορύβου (de-noise), τη βελτίωση της αντίθεσης (contrast enhancement) και την τόνιση των ακμών (edge detection and image sharpen). Όμως, ακόμη και πιο πολύπλοκες διεργασίες, όπως εύρεση επιφανειών αντικειμένων (object contour) ή ταυτότητα αντικειμένων (object identity) ανήκουν στην περιοχή της ψηφιακής επεξεργασίας της εικόνας.

Από τα πρώτα πεδία εφαρμογής της ψηφιακής επεξεργασίας εικόνων αποτέλεσαν οι αποστολές στο διάστημα. Πιο αναλυτικά, τη δεκαετία του 1960, το διαστημικό σκάφος Ranger 7 απέστειλε από το διάστημα εικόνες του δορυφόρου της Γης, δηλαδή της Σελήνης [Cerr03]. Οι εικόνες αυτές είχαν επεξεργαστεί κατάλληλα, με σκοπό να μειωθεί ο θόρυβος της συγκεκριμένης φωτογραφικής μηχανής που χρησιμοποιήθηκε. Μία από τις εικόνες αυτές παρουσιάζονται στο παρακάτω σχήμα.



**Εικόνα 6**

Άμεσα, τα πλεονεκτήματα αντίστοιχων μεθοδολογιών έγιναν κατανοητά και έγιναν προσπάθειες οι τεχνικές αυτές να ενσωματωθούν σε πολλούς τομείς της σύγχρονης ζωής. Σήμερα, οι αλγόριθμοι της ψηφιακής επεξεργασίας εικόνας συνδέονται και συνεργάζονται με την ανάλυση εικόνας (image analysis) και την υπολογιστική όραση (computer vision), με σκοπό την κατανόηση μίας σκηνής, όπως θα την αντιλαμβανόταν ένας φυσιολογικός άνθρωπος. Ως χαρακτηριστικό παράδειγμα της καθημερινότητας αναφέρουμε τα συστήματα αυτόματης αναγνώρισης του αριθμού κυκλοφορίας των οχημάτων, τα οποία και συναντώνται σε αρκετούς –πλήρως αυτοματοποιημένους- χώρους στάθμευσης. Τα συστήματα αυτά μπορούν να χρησιμοποιηθούν και από τις αστυνομικές δυνάμεις, με σκοπό την αυτόματη αναγνώριση συγκεκριμένων οχημάτων. Η αστυνομία σίγουρα χρησιμοποιεί συστήματα αυτόματης αναγνώρισης και ταυτοποίησης δακτυλικών αποτυπωμάτων (fingerprint), καθώς πρέπει να είναι σε θέση να εντοπίζει με γρήγορο τρόπο τους δράστες παράνομων ενεργειών [Malt09]. Και στο βιομηχανικό, όμως, τομέα, τα συστήματα της ψηφιακής επεξεργασίας εικόνας έχουν εφαρμογή. Στη γραμμή παραγωγής ενός προϊόντος γίνεται χρήση ειδικών καμερών και στη συνέχεια τεχνικών επεξεργασίας των αντίστοιχων εικόνων, ώστε να εντοπισθούν περιπτώσεις ελαττωματικών αντικειμένων [Goln07]. Άλλη εφαρμογή συναντάται σε περιπτώσεις ρομποτικής, όπου το ρομπότ της γραμμής παραγωγής οφείλει να γνωρίζει για παράδειγμα την ακριβή θέση μίας οπής, για να προσαρτήσει στο σημείο αυτό μία βίδα. Κατεχοχήν, όμως, τομέας της ψηφιακής επεξεργασίας εικόνας αποτελεί η ιατρική, με τα διάφορα συστήματα που χρησιμοποιούν οι ειδικευμένοι επιστήμονες. Ενδεικτικά αναφέρουμε τον αξονικό τομογράφο (CT) [Kak88], το μαγνητικό τομογράφο (MRI) [Wehr88] και τον τομογράφο εκπομπής ποζιτρονίου (PET) [Phe185]. Στο ακόλουθο σχήμα (εικόνα 7) παραθέτουμε ορισμένα από τα παραπάνω συστήματα εφαρμογών του επιστημονικού πεδίου της ψηφιακής επεξεργασίας εικόνας που αναφέραμε.



Εικόνα 7

## 1.2 Χαρακτηριστικά εικόνας

Τα χαρακτηριστικά της εικόνας χρησιμοποιούνται για να περιγράψουν και να διαχωρίσουν τα αντικείμενά της. Στο ανθρώπινο μάτι πολλές φορές μοιάζει εύκολος διαχωρισμός και η εξαγωγή χαρακτηριστικών από μια εικόνα. Τα χρώματα και τα σχήματα των αντικειμένων γίνονται άμεσα αντιληπτά και ευδιάκριτα χωρίς πιθανότητα λάθους. Τι συμβαίνει όμως όταν πρέπει να αναθέσουμε στον ηλεκτρονικό υπολογιστή την διάκριση των εκάστοτε χαρακτηριστικών μιας εικόνας ; Ο υπολογιστής δεν έχει την ικανότητα να διαχωρίσει το άσπρο από το πράσινο , το μεγάλο από το μικρό ή ακόμα τη θέση ενός αντικειμένου στην εικόνα με οπτικό μέσο, όπως έχουμε εμείς τα μάτια. Γι αυτό πρέπει να γίνει η μετατροπή του γενικού συνόλου των εικονοστοιχείων (pixels) σε αριθμητικές τιμές οι οποίες και είναι μετρήσιμες από το σύστημα. Τις αριθμητικές αυτές τιμές ο υπολογιστής τις αντιλαμβάνεται ως χαρακτηριστικά της εικόνας. Βέβαια όταν τα δεδομένα εισόδου προς επεξεργασία σε έναν αλγόριθμο έχουν μεγάλο όγκο, και επιπλέον υπάρχει η υποψία ότι εμπεριέχουν περιττή πληροφορία, τότε αυτά θα πρέπει να μετατραπούν σε μια μειωμένη παρουσίαση ενός συνόλου των βασικών τους χαρακτηριστικών (γνωστό και ως διάνυσμα χαρακτηριστικών). Τα χαρακτηριστικά που μπορούν να εξαχθούν από μια εικόνα είναι πάρα πολλά και μπορούν να εξαχθούν με πολλούς και διάφορους τρόπους. Αν τα εξαγόμενα χαρακτηριστικά επιλεγθούν προσεκτικά, αναμένεται να μας δώσουν την άμεσα σχετική και επιθυμητή πληροφορία των δεδομένων εισόδου προκειμένου να εκτελέσουμε σε επόμενο βήμα την οποιαδήποτε διαδικασία με αισθητά μειωμένο όγκο πληροφορίας. Παρακάτω θα δούμε τους διάφορους τρόπους με τους οποίους μπορούμε να αναπαραστήσουμε το περιεχόμενο, τη πληροφορία μιας εικόνας:

- Ολικά χαρακτηριστικά:

Περιλαμβάνουν χαρακτηριστικά που περιγράφουν συνολικά ολόκληρη την εικόνα. Ενδεικτικά θα μπορούσαν να δοθούν ως παράδειγμα ο μέσος όρος, η τυπική απόκλιση των εικονοστοιχείων της εικόνας και το ιστογράμματα (συναρτήσεις πυκνότητας πιθανότητας) των αποχρώσεών της.

- Τοπικά χαρακτηριστικά:

Είναι τα χαρακτηριστικά τα οποία περιγράφουν μια περιοχή της εικόνας. Τα χαρακτηριστικά αυτά μπορούν να χρησιμοποιηθούν για το ταίριασμα προτύπων εντός της εικόνας. Τα τοπικά χαρακτηριστικά μπορεί να συνοδεύονται από την πληροφορία θέσης από την οποία εξαγονται μέσα στην εικόνα.

- Χαρακτηριστικά αντικειμένων:

Είναι τα χαρακτηριστικά τα οποία μπορούν να υπολογισθούν για κάθε αντικείμενο ξεχωριστά μέσα σε μια εικόνα. Τα αντικείμενα μπορούν να καταταμηθούν είτε χειροκίνητα είτε από τους διάφορους ημιαυτόματους και αυτόματους τρόπους. Η κατάτμηση αποσκοπεί στη διαίρεση της

εικόνας σε περιοχές οι οποίες έχουν ομογενείς ιδιότητες κατά κάποια έννοια (π.χ. περιοχές χρώματος) και μπορεί να επιτευχθεί με πολλούς τρόπους. Ένα παράδειγμα αλγόριθμου κατάτμησης είναι ο Expectation – Maximization όπου κατατμείται η εικόνα με βάση τις κλίσεις της και δεν τεμαχίζεται σε τμήματα με βάση την χρωματική κατανομή σε κάθε περιοχή. Δύο βασικές τεχνικές κατάτμησης είναι οι ακόλουθες:

- *Απλή κατωφλίωση*, αν πρόκειται για αντικείμενα τα οποία μπορούν ευκολα να διαχωριστούν, όπως για παράδειγμα ένας άνθρωπος με φόντο τη θάλασσα.
- *Τεχνικές πλημμύρας*, στις οποίες ο χρήστης αρχικοποιεί τη μέθοδο δίνοντας ένα εικονοστοιχείο μέσα στο αντικείμενο κάνοντας χρήση της πληροφορίας των γειτονικών εικονοστοιχείων, π.χ. παρόμοιο χρώμα.

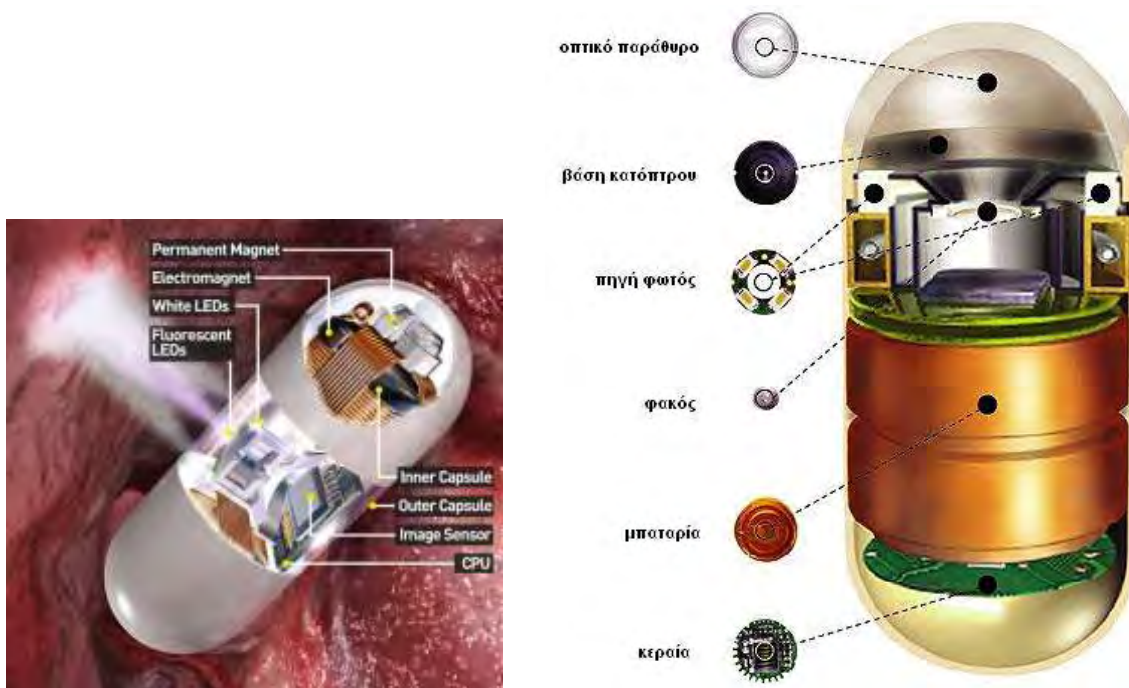
### **1.3 Ενδοσκοπική κάψουλα**

Το 1980 κυκλοφόρησε το βιβλίο του Isaac Asimov με τίτλο «Φανταστικό ταξίδι». Ο συγγραφέας διηγείται ότι επιστήμονες της NASA κατασκεύασαν ένα μικροσκοπικό όχημα που διερευνούσε το ανθρώπινο σώμα και επισκεύαζε βλάβες, ακόμα και σε κυτταρικό επίπεδο. Πολύ σύντομα το βιβλίο έγινε κινηματογραφική ταινία.

Είκοσι χρόνια αργότερα, στις 24 Μαΐου 2000, στις ειδήσεις του BBC παρουσιάστηκε η νέα τεχνολογία της ενδοσκοπικής κάψουλας ή βιντεοκάψουλας του λεπτού εντέρου, που επέτρεπε στους επιστήμονες «ένα φανταστικό ταξίδι» στο λεπτό έντερο του ανθρώπου.

Η ενδοσκοπική κάψουλα είναι πλαστική κάψουλα διαστάσεων 11x26 χιλιοστά και βάρους 3,7 γραμμάρια. Παρά το πολύ μικρό μέγεθός της περιέχει πηγή ψυχρού φωτισμού, μπαταρία διάρκειας οκτώ ωρών και μικροτσίπ για τη συνεχή λήψη ψηφιακών εικόνων, με ρυθμό 2-8 ανά δευτερόλεπτο. Καθώς η κάψουλα προωθείται από τις περισταλτικές κινήσεις του εντέρου, οι εικόνες που λαμβάνει μεταβιβάζονται συνεχώς και αποθηκεύονται σε δέκτη καταγραφής που φοράει στη μέση του ασθενούς. Μετά από την ολοκλήρωση της εξέτασης, ο δέκτης συνδέεται με ηλεκτρονικό υπολογιστή και γίνεται ανάγνωση του βίντεο από ειδικά εκπαιδευόμενο γαστρεντερολόγο.





**Εικόνα 8, δομή ενδοσκοπικής κάψουλας**

Τα **πλεονεκτήματα** της εξέτασης με βιντεοκάψουλα είναι ότι παρέχει άριστης ποιότητας εικόνες του βλεννογόνου και δεν απαιτεί νοσηλεία του ασθενούς ή καταστολή όπως στις κλασικές ενδοσκοπήσεις (γαστροσκόπηση, κολonosκόπηση). Επίσης ο ασθενής δεν υποβάλλεται σε ακτινοβολία, όπως με τις κλασικές ακτινογραφίες στομάχου ή εντέρων με βάριο, ή την αξονική τομογραφία.

Στα **μειονεκτήματα** της εξέτασης περιλαμβάνονται κυρίως ατελείς εξετάσεις. Στον οισοφάγο η εξέταση μπορεί να είναι ατελής λόγω ταχείας διόδου της κάψουλας στο στόμαχο. Πρόσφατα δοκιμάζεται η μέθοδος «κάψουλα δεμένη με νήμα», που επιτρέπει την λεπτομερή εξέταση του οισοφάγου.

Στο λεπτό έντερο η εξέταση είναι πλήρης στο 80-85% των ασθενών. Στους υπόλοιπους ασθενείς είτε ο τελικός ειλεός είναι ακάθαρτος και οι εικόνες ακατάλληλες είτε η μπαταρία εξαντλείται πριν να φθάσει η κάψουλα στην ειλεοτυφλική βαλβίδα. Στην εξέταση του παχέος εντέρου η επιτυχία της μεθόδου είναι 60-75%. Στους υπόλοιπους ασθενείς είτε το παχύ έντερο είναι ακάθαρτο είτε η μπαταρία εξαντλείται πριν η κάψουλα φθάσει στον πρωκτό.

Η τεχνική της ενδοσκοπικής κάψουλας, είναι μια κατ' εξοχήν ασφαλής και μη επώδυνη μέθοδος ελέγχου του λεπτού εντέρου, η οποία στα χέρια ενός έμπειρου ενδοσκόπου - γαστρεντερολόγου, μπορεί να έχει σημαντική διαγνωστική αξία. Το 2001 η βιντεοκάψουλα του λεπτού εντέρου έλαβε άδεια κλινικής χρήσεως από το Food and Drug Administration (FDA) των ΗΠΑ και υπολογίζεται ότι μέχρι σήμερα

έχει χρησιμοποιηθεί για τη διερεύνηση 1.000.000 ασθενών σε όλο τον κόσμο. Το 2004 κατασκευάστηκε και τέθηκε σε χρήση η βιντεοκάψουλα του οισοφάγου και το 2006 η βιντεοκάψουλα του παχέος εντέρου.



**Εικόνα 9**

Στη παραπάνω φωτογραφία (εικόνα 9) από βιντεοσκοπική κάψουλα φαίνεται καθαρά το έλκος στο μέσου του λεπτού εντέρου αλλά και θρόμβος αίματος που υποδηλώνει πρόσφατη αιμορραγία .

#### **1.4 Επεξεργασία και ανάλυση ενδοσκοπικών δεδομένων**

Τα συστήματα που χρησιμοποιούνται στην επεξεργασία και ανάλυση ενδοσκοπικών εικόνων διέπονται από τη τυπική δομή συστημάτων υποβοήθησης της ιατρικής διάγνωσης (Computer Aided Detection, CAD), και συμπεριλαμβάνουν τα παρακάτω στάδια:

- Προ-επεξεργασία: στάδιο αφαίρεσης θορύβου και βελτίωσης της ποιότητας της εικόνας
- Κατάτμηση: εύρεση διάφορων δομών της εικόνας
- Εξαγωγή χαρακτηριστικών: εξαγωγή τιμών των διάφορων επιλεγμένων χαρακτηριστικών
- Ταξινόμηση: κατηγοριοποίηση της εικόνας ως προς συγκεκριμένο χαρακτηριστικό γνώρισμα

Η επιστημονική και ακαδημαϊκή κοινότητα έχει προτείνει πλήθος CAD συστημάτων με εξειδίκευση στην επεξεργασία και ανάλυση ενδοσκοπικών εικόνων, κάθε ένα από τα οποία διαθέτει συγκεκριμένα πλεονεκτήματα και μειονεκτήματα, κάνοντάς το κατάλληλο για διαφορετικές εφαρμογές και περιπτώσεις. Ενδεικτικά αναφέρουμε μερικές εργασίες της βιβλιογραφίας, όπως την [Kark03], όπου γίνεται χρήση κυματιδίων, με σκοπό τον εντοπισμό όγκων σε ενδοσκοπικές εικόνες και την [WangP01], όπου τα χαρακτηριστικά υφής που εξαγονται από τις εικόνες ταξινομούνται με βάση νευρωνικό δίκτυο. Χαρακτηριστικά υφής χρησιμοποιούνται και στην [Galo99], ενώ στην [Hafn10] εξαγονται

χαρακτηριστικά με κριτήριο τις εντοπισμένες ακμές της εικόνας και ταξινομούνται με βάση τον ταξινομητή  $k$  πλησιέστερων γειτόνων.



**Εικόνα10,σύγχρονο CAD σύστημα επεξεργασίας ενδοσκοπικών εικόνων**

### **1.5 Αντικειμενικοί στόχοι**

- Η διερεύνηση εξαγωγής χαρακτηριστικών σχήματος από αντικείμενα που αναπαριστώνται σε ψηφιακές εικόνες.
- Η διερεύνηση συγκεκριμένων χαρακτηριστικών για την αναπαράσταση σχημάτων.
- Η εφαρμογή τους και αξιολόγηση της αποτελεσματικότητάς τους σε δημοσίως διαθέσιμες βάσεις εικόνων
- Η εφαρμογή τους για την επίλυση πραγματικού ιατρικού προβλήματος με στόχο την υποβοήθηση της ιατρικής διάγνωσης.
- Να αναπτυχθούν βάσεις εικόνων, που μπορούν να χρησιμοποιηθούν μελλοντικά από άλλες ερευνητικές ομάδες (π.χ. περίγραμμα που αποδόσαμε χειροκίνητα σε ιατρικές ενδοσκοπικές εικόνες)



## 1.6 Δομή εργασίας

Στο κεφάλαιο 2 (Ανάλυση Εικόνας και Αναγνώριση Προτύπων) θα κάνουμε μια βασική εισαγωγή στην αναγνώριση προτύπων, θα δούμε επιγραμματικά τα βήματα της βασικής της διαδικασίας και στη συνέχεια θα αναλύσουμε στάδια όπως η προεπεξεργασία της εικόνας, η επιλογή και η εξαγωγή των χαρακτηριστικών. Αφού αναφερθούμε στη ταξινόμηση σε κλάσεις και σε μερικούς από τους βασικότερους ταξινομητές, θα περάσουμε σε μια σύντομη ανασκόπηση των CBIR συστημάτων και στην ανάκληση εικόνων βάσει περιεχομένου και τέλος χρησιμοποιώντας την υπάρχουσα βιβλιογραφία θα πραγματοποιήσουμε ένα ταξίδι στο επιστημονικό παρελθόν κάνοντας μια ανασκόπηση στις διάφορες μεθόδους εξαγωγής χαρακτηριστικών και τις εφαρμογές τους. Το κεφάλαιο κλείνει κάνοντας μια εισαγωγή στους Fourier Descriptors εξηγώντας για ποιο λόγο επιλέχθηκαν για την εκπόνηση της συγκεκριμένης εργασίας.

Στο κεφάλαιο 3 θα περάσουμε στο βασικό κομμάτι της πτυχιακής εργασίας, αρχικά δίνοντας κάποιες βασικές πληροφορίες για τους περιγραφείς Fourier ( Fourier descriptors ) όπως ανάπτυξη συνάρτησης σε σειρά Fourier, εκθετική και τριγωνομετρική σειρά Fourier και εν συνεχεία θα εξηγήσουμε αναλυτικά τι είναι οι Elliptic Fourier Descriptors και πως αυτοί παραμένουν αμετάβλητοι σε γεωμετρικές αλλαγές. Το κεφάλαιο θα κλείσει κάνοντας μια αναφορά σε δυο βασικούς ταξινομητές διανυσμάτων, αυτόν των K-κοντινότερων γειτόνων και τον Naïve – Bayes ταξινομητή.

Το κεφάλαιο 4 αποτελείται από την αναλυτική παρουσίαση όλων των βάσεων εικόνων που χρησιμοποιήθηκαν, όπως επίσης και από την λεπτομερή περιγραφή των πειραμάτων που υλοποιήθηκαν. Θα δούμε και θα ερμηνεύσουμε την εφαρμογή των EFD στα βασικά γεωμετρικά σχήματα (παράλληλα θα αναφερθούμε στη χρήση της τεχνικής Sumtton και στον τρόπο λειτουργίας της), αλλά και σε ιατρικές εικόνες που είναι και η βασική μελέτη της πτυχιακής εργασίας. Ακόμη, παρουσιάζονται και ερμηνεύονται τα αποτελέσματα και πραγματοποιείται μία σύγκριση μεταξύ τους. Τέλος γίνεται ταξινόμηση στις εικόνες της κάθε βάσης σε κατηγορίες, ανάλογα με τις τιμές των EFD συντελεστών χρησιμοποιώντας το εργαλείο Weka.

Τέλος, στο κεφάλαιο 5 καταγράφουμε συνοπτικά τα συμπεράσματα της παρούσης εργασίας και προτείνουμε μελλοντικές προοπτικές της έρευνας στη συγκεκριμένη θεματολογία.

## ΚΕΦΑΛΑΙΟ 2 : Ανάλυση Εικόνας και Αναγνώριση Προτύπων

### 2.1 Εισαγωγή στην Αναγνώριση Προτύπων

Αναγνώριση Προτύπων (Pattern Recognition) είναι η επιστήμη που προσπαθεί να αναγνωρίσει με αυτόματο τρόπο χρήσιμες κανονικότητες σε θορυβώδη και πολύπλοκα περιβάλλοντα [Tsak06]

Πρότυπο είναι μια ασαφώς ορισμένη οντότητα στην οποία μπορεί να αποδοθεί κάποιο όνομα όπως ένα αποτύπωμα, ένα πρόσωπο ή ένας όγκος σε μια ιατρική εικόνα [wata85].

Μία μηχανή, όπως ένας ηλεκτρονικός υπολογιστής, πρέπει να εκπαιδευθεί κατάλληλα ώστε να αναγνωρίζει πρότυπα (patterns) και να τα κατηγοριοποιεί αυτόματα σε κατηγορίες. Ανάλογα με την εφαρμογή γίνεται κατάταξη των αντικειμένων σε κλάσεις με τη βοήθεια αλγορίθμων ταξινόμησης. Το ερευνητικό ενδιαφέρον για αυτά τα ζητήματα ξεκίνησε από τη δεκαετία του 1960, κατά την πρώτη περίοδο της ανάπτυξης της επιστήμης των υπολογιστών. Βασιζόμενη στο θεωρητικό υπόβαθρο που παρείχε η επιστήμη της Στατιστικής, η πρώιμη έρευνα επικεντρώθηκε στην ανάπτυξη θεωρητικών μεθόδων. Η αναγνώριση προτύπων περιλαμβάνει μεθοδολογίες για την περιγραφή, την ταξινόμηση (classification) ή την ομαδοποίηση (clustering) των προτύπων σε κλάσεις

Δεδομένης μίας υπάρχουσας συλλογής αντικειμένων για τα οποία είναι γνωστή κλάση (ή κατηγορία) στόχος είναι να βρεθεί μία συνάρτηση μεταβλητών που θα περιγράψει το μοντέλο της κλάσης. Η επιτυχία πρόβλεψης αξιολογείται με ένα νέο σύνολο δεδομένων.

#### 2.1.1 Βασική Διαδικασία

1. Καθορισμός τύπου δεδομένων.
2. Δημιουργία σώματος εκπαίδευσης.  
Το σώμα εκπαίδευσης πρέπει να είναι αντιπροσωπευτικό της πραγματικής φύσης του προβλήματος. Από μετρήσεις ή εξωτερική ανθρώπινη παρέμβαση αναθέτουμε στα δεδομένα του σώματος τις ετικέτες.
3. Εξαγωγή κατάλληλων χαρακτηριστικών.  
Μετατρέπουμε κάθε νέο αντικείμενο σε διάνυσμα χαρακτηριστικών (input vector) προκειμένου να ακολουθήσει η ταξινόμησή του. Είναι σημαντικό να επιλεγούν τα γνωρίσματα εκείνα που περιέχουν την απαραίτητη πληροφορία για την σωστή ανάθεση κλάσης ενώ παράλληλα ο αριθμός τους πρέπει να είναι διαχειρίσιμος υπολογιστικά. Επίσης πρέπει να γνωρίζουμε τον τύπο τιμής κάθε ιδιότητας (ονομαστικής (nominal), τακτικής (ordinal), αναλογικής (ratio)).
4. Επιλογή αλγορίθμου εκπαίδευσης.  
Υπάρχει πληθώρα αλγορίθμων και η επιλογή πρέπει να γίνει ανάλογα με το ποιος κρίνεται πιο αποδοτικός για τη συγκεκριμένη κατηγορία προβλημάτων.

5. Αξιολόγηση της διαδικασίας.

Μετά την ολοκλήρωση της εκπαίδευσης η ακρίβεια της παραγόμενης σχέσης θα πρέπει να αξιολογηθεί με ένα σύνολο αξιολόγησης (evaluation set).

### **2.1.2 Προεπεξεργασία εικόνας**

Ο στόχος της προεπεξεργασίας είναι η βελτίωση των αρχικών εικόνων μειώνοντας ανεπιθύμητες στρεβλώσεις ή αντίθετα ενισχύοντας κάποια χαρακτηριστικά τους που είναι σημαντικά για περετέρω επεξεργασία. Μερικές απ τις βασικές προεπεξεργασίες εικόνας είναι οι εξής:

- Μετασχηματισμός φωτεινότητας
- Γεωμετρικοί μετασχηματισμοί
- Παλινόρθωση εικόνας (Image Reconstruction), η οποία απαιτεί γνώση της συνολικής εικόνας

Ένα άλλο στάδιο προεπεξεργασίας των εικόνων είναι το γέμισμα των κενών περιοχών εσωτερικά των αντικειμένων. Θα πρέπει οι περιοχές αυτές να «γεμίσουν» με λευκά pixels, ώστε να εντοπισθεί σωστά το περίγραμμα του αντικειμένου. Ο αλγόριθμος που επιτελεί αυτή τη λειτουργία ονομάζεται flood fill, και θα τον δουμε αναλυτικότερα στο τέταρτο κεφάλαιο.

### **2.1.3 Εξαγωγή χαρακτηριστικών**

Στην αναγνώριση προτύπων και στην επεξεργασία εικόνας, η εξαγωγή χαρακτηριστικών παίζει πού σημαντικό ρόλο. Θα μπορούσαμε να πούμε πως η εξαγωγή χαρακτηριστικών είναι ένας γενικός όρος που περιλαμβάνει της μεθόδους με τις οποίες πετυχαίνουμε σμίκρυνση των δεδομένων εισόδου ενώ παράλληλα τα περιγράφουμε με επαρκή ακρίβεια. Εξαγωγή χαρακτηριστικών από την εικόνα γίνεται με χρήση αλγόριθμων και μετατροπή τους σε μαθηματικά δεδομένα για να είναι μετρήσιμα από το σύστημα και να γίνεται ανίχνευση ομοιότητας μεταξύ εικόνων με τη σύγκριση των χαρακτηριστικών αυτών. Όμως κάθε εφαρμογή έχει τις δικές τις ανάγκες για αυτό και ο κάθε τρόπος εξαγωγής τους είναι προσαρμοσμένος στο πρόβλημα που εκκείται να λύσει. Στόχος της εξαγωγής χαρακτηριστικών είναι η σμίκρυνση του αρχικού όγκου δεδομένων εισόδου, χρησιμοποιώντας μόνο εκείνα τα χαρακτηριστικά που είναι επαρκή, σαφή και ικανά να μας δώσουν τις ζητούμενες πληροφορίες και τα αποτελέσματα για το εκάστοτε πρόβλημα. Για παράδειγμα, έστω ότι χρειαζόμαστε χαρακτηριστικά ώστε να κατηγοριοποιήσουμε μια τάξη μαθητών σε καλούς, μέτριους και κακούς και τα διαθέσιμα χαρακτηριστικά μας είναι οι βαθμοί τους, το ύψος τους, το φύλλο και το, δείκτη IQ. Ποια από αυτά θα πρέπει να επιλεγούν υπόψη στο πρόβλημα μας; Θα μπορούσαμε να επιλέξουμε τους βαθμούς και τον δείκτη IQ ως τα πιο επαρκή από αυτά ή έναν συνδιασμό αυτών των δυο σε ένα νέο χαρακτηριστικό, το οποίο εξάγεται με κάποια μαθηματική μέθοδο. Βασικός στόχος είναι η μετατροπή του αρχικού χώρου X σε έναν νέον χώρο Y για να καταλήξουμε σε νέα χαρακτηριστικά που θα λειτουργήσουν καλύτερα για εμάς. Με αυτόν τον τρόπο επιτυγχάνεται συμπίεση της αρχικής πληροφορίας.

### **2.1.3.1 Εξαγωγή χαρακτηριστικών από εικόνες**

Στα περισσότερα συστήματα τα χαρακτηριστικά κάθε εικόνας υπολογίζονται και αποθηκεύονται στο παρασκήνιο (πριν τη λειτουργία του συστήματος) έτσι ώστε να ελαττώνεται ο υπολογιστικός φόρτος του. Βέβαια αν η βάση δεδομένων αποτελείται από κάποιες εκατοντάδες χιλιάδες εικόνες, το προεπεξεργαστικό αυτό βήμα χρειάζεται αρκετό χρόνο

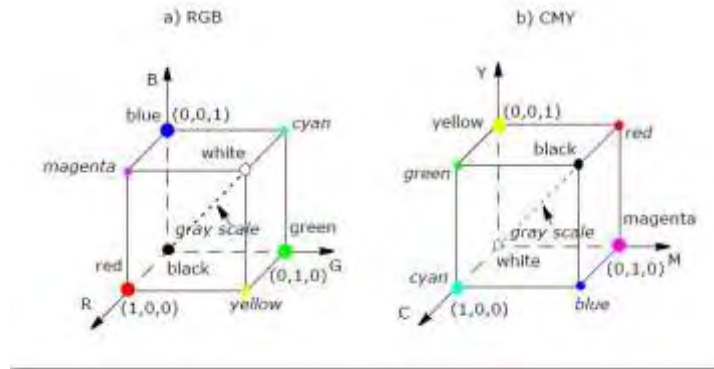
Τις περισσότερες φορές τα χαρακτηριστικά των εικόνων διαιρούνται σε χαμηλού και υψηλότερου επιπέδου. Συχνά τα υψηλού επιπέδου υπολογίζονται απο αυτά του χαμηλού επιπέδου. Θα αναλύσουμε το χρώμα, την υφή, το σχήμα και τη θέση. Πιο αναλυτικά έχουμε,

#### ***Χαμηλού επιπέδου χαρακτηριστικά:***

- **Χρώμα:**

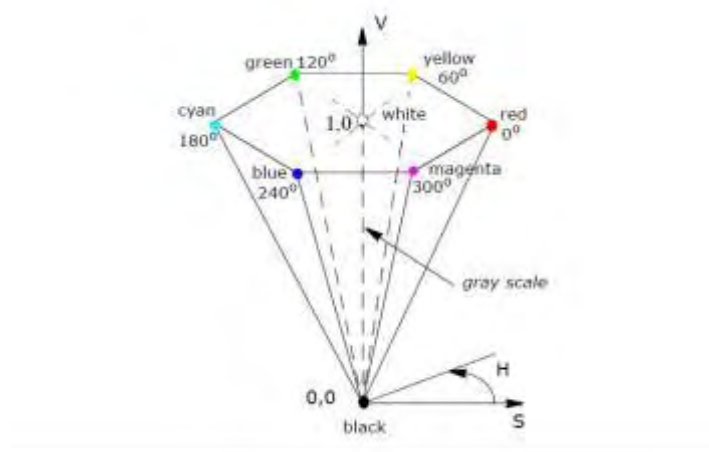
Περιλαμβάνει τη χρωματική πληροφορία τόσο στις εικόνες των τόνων του γκρι, όσο και στις έγχρωμες εικόνες. Είναι το πιο συχνά χρησιμοποιούμενο χαρακτηριστικό για ανάκληση εικόνων. Πριν όμως χρησιμοποιηθεί η χρωματική πληροφορία πρέπει να οριστεί ο χρωματικός χώρος. Δεν υπάρχει δεδομένη συμφωνία ανάμεσα στους επιστήμονες για το ποιός είναι ο καλύτερος, συμφωνούν όμως στο γεγονός ότι για να θεωρείται καλός πρέπει να είναι ομοιόμορφος.

- Ο γνωστός σε όλους RGB (Red, Green, Blue) χώρος συνήθως χρησιμοποιείται όπου δεν υπάρχουν μεταβολές φωτεινότητας όπως για παράδειγμα οι πίνακες ζωγραφικής, η σύνθεση εικόνων απο φωτογραφίες ή εμπορικά σήματα. Παρόλο που είναι αρκετά απλός και κατανοητός χώρος, δεν αποτελεί πρωταρχική επιλογή στα CBIR (Content-Based Image Retrieval) συστήματα, καθώς εξαρτάται έντονα από τις εκάστοτε συσκευές λήψης.
- Ο CMY (Cyan, Magenta, Yellow) είναι ένας εξίσου σημαντικός χρωματικός χώρος, χρησιμοποιείται στην απεικόνιση και εκτύπωση έγχρωμων εικόνων, όπως και το RGB, μοντέλα που διαφαίνονται στο παρακάτω σχήμα. Ο RGB και ο CMY δεν είναι ομοιόμορφοι χρωματικοί χώροι.



Εικόνα 11, σχηματική αναπαράσταση RGB και CMY χώρων

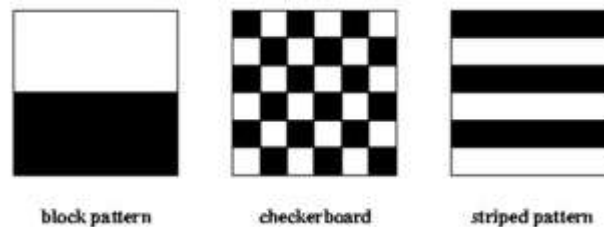
- Οι YIQ και YUV χρησιμοποιούνται από τις εταιρίες τηλεόρασης τόσο στην Ευρώπη (PAL) όσο και στην Αμερική (NTSC) και λαμβάνονται απο τον χρωματικό χώρο RGB με έναν απλό γραμμικό μετασχηματισμο. Η Y συνιστώσα περιέχει την πληροφορία φωτεινότητας της εικόνας και είναι αυτή που εμφανίζεται στις ασπρόμαυρες τηλεοράσεις. Η κωδικοποίηση της γίνεται στις άλλες δυο συνιστώσες.
- Ο HSV (Hue, Saturation, Value or Brightness) χρησιμοποιείται εκτεταμένα από τα γραφικά υπολογιστών και είναι περισσότερο διαισθητικοί στην περιγραφή ενός χρώματος. Επίσης αποτελείται απο τρεις συνιστώσες, όπου η hue παραμένει αμετάβλητη στις αλλαγές της φωτεινότητας κι έτσι θεωρείται η πιο αξιόπιστη. Το χρωματικό μοντέλο HSV προκύπτει από το RGB [Fole00] και θεωρείται ότι προσομοιώνει την ανθρώπινη αντίληψη, τον τρόπο δηλαδή που αντιλαμβάνεται ο ανθρώπινος οργανισμός το χρώμα.



Εικόνα 12, σχηματική αναπαράσταση HSV χώρου

- Υφή

Η υφή μίας εικόνας, δεν αποδίδεται με ένα συγκεκριμένο ορισμό [Cogg82], καθώς ο καθένας αποδίδει την έννοια αυτή με διαφορετικό τρόπο, ανάλογα με εκάστοτε εφαρμογή. Προσπαθώντας και εμείς με τη σειρά μας, θα λέγαμε ότι η υφή μίας εικόνας είναι ο τρόπος που κατανέμονται τα pixels με διαφορετικά επίπεδα του γκρι (spatial variation in pixel intensities), ενώ είναι ο πρώτος και βασικός τρόπος, με τον οποίο ο άνθρωπος διαχωρίζει μία μονοχρωματική εικόνα σε επιμέρους υποσύνολα. Ένα παράδειγμα εικόνων με διαφορετική υφή απεικονίζεται στο παρακάτω σχήμα.



**Εικόνα 13**

Η υφή μπορεί να χρησιμοποιήσει διάφορα μοντέλα:

- Χαρακτηριστικά Tamura,[Tamu78]:

Τα χαρακτηριστικά Tamura περιλαμβάνουν την ταχύτητα, την αντίθεση, τη κατεύθυνση την ομαλότητα και την σκληρότητα και είναι σχεδιασμένα με ψυχολογικές μελέτες πάνω στην ανθρώπινη αντίληψη της υφής. Τα πρώτα τρία χαρακτηριστικά έχουν χρησιμοποιηθεί σε κάποια πολύ γνωστά συστήματα ανάκλησης εικόνων, όπως είναι το Photobook και το QBIC.

- Ταυτόχρονα αυτό-παλινδρόμμενα μοντέλα (Simultaneous Auto- Regressive Models, SAR):

Τα SAR είναι μια περίπτωση των τυχαίων Μαρκοβιανών μοντέλων, τα οποία είχαν μεγάλη επιτυχία στην περιγραφή της υφής την προηγούμενη δεκαετία. Χαρακτηριστικό τους είναι πως έχουν καλύτερη απόδοση χρησιμοποιώντας λιγότερες παραμέτρους.

### Υψηλότερου επιπέδου χαρακτηριστικά:

- Σχήμα

Οι περιοχές και τα αντικείμενα μιας εικόνας μπορούν να αναπαρασταθούν με τα περιγράμματά τους (κώδικες αλυσίδας). Βέβαια η εικόνα θα πρέπει πιο πριν να έχει υποστεί κατάτμηση. Υπάρχει ένας μεγάλος αριθμός συστημάτων που έχουν αναπτυχθεί χρησιμοποιούν τη πληροφορία σχήματος. Ως περιγραφείς σχήματος μπορούν να θεωρηθούν οι εξής:

- Περιγραφείς Fourier (Fourier Descriptors). Απλοί χειρισμοί των συντελεστών Fourier μπορούν να εξαλείψουν τη σύνδεση του αντικειμένου με τη θέση, το μέγεθος και την κατεύθυνση, για παράδειγμα η περιγραφή του αντικειμένου με FD μπορεί να γίνει αμετάβλητη στην περιστροφή χρησιμοποιώντας μόνο το πλάτος των συντελεστών. Η εργασία αυτή βασίζεται πάνω στους περιγραφείς Fourier και συγκεκριμένα στους Elliptic Fourier Descriptors, που θα δούμε αναλυτικά στο τρίτο κεφάλαιο.
- Ροπές σχήματος, όπου ιδιέταιρα προτιμώνται οι αμετάβλητες. [Liao96], [Teh88], [Gonz07]
- Φάσμα του σχήματος της εικόνας (Ιστόγραμμα καμπυλότητας)
- Παραμορφώσιμα μοντέλα περιγράμματος, τα οποία χρησιμοποιούνται προκειμένου να μοντελοποιηθούν τα περιγράμματα των αντικειμένων της εικόνας υπό εξέταση.

- Θέση

Η χωρική θέση που έχουν τα χαρακτηριστικά, οι περιοχές ή τα αντικείμενα. Για παράδειγμα ο ουρανός ή ο ωκεανός μπορεί να υπάρχουν σε διάφορες εικόνες, αλλά να κατέχουν διαφορετική θέση σε αυτές. Οπότε θα πρέπει να ληφθεί υπόψη και η θέση που κατέχουν τα αντικείμενα στις διάφορες εικόνες. Μια πολύ διαδεδομένη μέθοδος καθορισμού θέσεως αντικειμένων είναι τα 2D strings όπου κατασκευάζονται προβάλλοντας την εικόνα στους  $x$  και  $y$  άξονες. Οι προβολές αυτές καθορίζουν δυο σύνολα, το  $A$  και το  $V$ . Κάθε σύμβολο στο  $A$  αναπαριστά ένα αντικείμενο ενώ στο  $V$  αναπαριστά τη χωρική σχέση μεταξύ των αντικειμένων.

### 2.1.3.2 Γενικές μέθοδοι εξαγωγής χαρακτηριστικών

Για τον σκοπό αυτό υπάρχουν διάφορες μέθοδοι που έχουν προταθεί (στην ενότητα 2.3 θα δούμε αναλυτικότερα και σε πράξη τις διάφορες μεθόδους εξαγωγής χαρακτηριστικών μέσα από τη βιβλιογραφική ανασκόπηση), μερικές από τις οποίες θα δούμε παρακάτω:

Η μέθοδος Principal Component Analysis (PCA) είναι η πλέον διαδεδομένη μέθοδος εξαγωγής χαρακτηριστικών. Είναι μια μαθηματική διαδικασία που χρησιμοποιεί ένα ορθογώνιο - γραμμικό μετασχηματισμό για να μετατρέψει ένα σύνολο παρατηρήσεων από πιθανώς συσχετιζόμενες μεταβλητές σε ένα σύνολο τιμών από γραμμικά μη συσχετιζόμενες μεταβλητές που ονομάζονται κύριες συνιστώσες (principal components) [Anil99]. Ο αριθμός των κυρίων συνιστωσών είναι μικρότερο ή το πολύ ίσος με αυτόν των αρχικών μεταβλητών. Ο μετασχηματισμός αυτός γίνεται με τέτοιο τρόπο ώστε η πρώτη κύρια συνιστώσα να έχει τη μεγαλύτερη δυνατή διακύμανση (να αντιπροσωπεύει όσο τον δυνατόν περισσότερη τη μεταβλητότητα των δεδομένων), και κάθε διαδοχική συνιστώσα να έχει με τη σειρά της την αμέσως επόμενη υψηλότερη διακύμανση με τον περιορισμό ότι θα είναι ορθογώνιες μεταξύ τους. Οι κύριες συνιστώσες εγγυώνται την ανεξαρτησία τους μόνο όταν το σύνολο των δεδομένων εισόδου κατανέμεται κανονικά [Jollif02].

Multidimensional scaling (MDS) είναι ένα σύνολο από σχετικές στατιστικές τεχνικές που χρησιμοποιούνται συχνά στην απεικόνιση πληροφοριών για την εξερεύνηση ομοιοτήτων ή διαφορών στα δεδομένα. Ο αλγόριθμος MDS ξεκινά με έναν πίνακα με τις ιδιότητες του κάθε στοιχείου, στη συνέχεια εκχωρεί μια θέση σε κάθε στοιχείο N-διάστατο χώρο, όπου το N καθορίζεται εκ των προτέρων. Όταν το N παίρνει μικρές τιμές, οι τελικές θέσεις μπορούν να παρασταθούν σε ένα γράφημα ή σε δυσδιάστατες τεχνικές απεικόνισης (scatterplots) [Anil99].

Στην κατηγορία αυτή ανήκουν και οι χάρτες του Sammon (Sammon's Mapping ή Sammon Projection) [Sam69], στους οποίους θα αφιερώσουμε περισσότερο χρόνο μιας και είναι μια τεχνική την οποία υιοθετούμε και χρησιμοποιούμε για τους σκοπούς της εργασίας αυτής.

Η μέθοδος παρουσιάστηκε από τον Sammon το 1969 με σκοπό να βοηθήσει στην αναπαράσταση n-διάστατων διανυσμάτων από τον n-διάστατο χώρο σε χώρο χαμηλότερης διάστασης, έτσι ώστε να διατηρηθεί η δομή των δεδομένων και να εντοπιστούν οι γεωμετρικές σχέσεις σε υποσύνολα διανυσμάτων, όπως ομαδοποιήσεις ή γραμμικές σχέσεις. Η διατήρηση της δομής των δεδομένων επιτυγχάνεται μέσω ενός αλγορίθμου ο οποίος σχεδιάζει σημεία στον χώρο χαμηλότερης διάστασης έτσι ώστε οι αποστάσεις μεταξύ τους να προσεγγίζουν τις αποστάσεις που υπάρχουν στον n-διάστατο χώρο.

Η διαδικασία έχει ως εξής:

Έστω ότι έχουμε  $k$  διανύσματα στον χώρο  $X_i$  με  $i=1,2,3,\dots,k$  στον  $n$ -διάστατο χώρο και σε αντιστοιχία προς αυτά ορίζουμε  $k$  διανύσματα  $Y_i$  με  $i=1,2,3,\dots,k$  στον  $d$ -διάστατο χώρο, όπου  $d < n$  και συνήθως  $d=2$ , καθώς μας ενδιαφέρει περισσότερο η αναπαράσταση στο επίπεδο. Ορίζουμε την απόσταση μεταξύ των διανυσμάτων  $X_i$  και  $X_j$  να είναι  $d_{ij}$  και την απόσταση μεταξύ των διανυσμάτων  $Y_i$  και  $Y_j$  να είναι  $d_{ij}$ . Ως μέτρο απόστασης χρησιμοποιείται συνήθως η Ευκλείδεια



απόσταση [Deza09] και στη συνέχεια θέτουμε τα διανύσματα  $Y$  στον  $d$ -διάστατο χώρο έτσι ώστε:

$$Y_1 = \begin{bmatrix} y_{11} \\ \vdots \\ y_{1d} \end{bmatrix}, Y_2 = \begin{bmatrix} y_{21} \\ \vdots \\ y_{2d} \end{bmatrix}, \dots, Y_k = \begin{bmatrix} y_{k1} \\ \vdots \\ y_{kd} \end{bmatrix}$$

Έπειτα υπολογίζουμε όλες τις αποστάσεις  $d_{ij}$  τις οποίες χρησιμοποιούμε για να ορίσουμε ένα σφάλμα  $E$  το οποίο δίνει μια εικόνα του πόσο καλά ταιριάζει ο σχηματισμός των  $k$  σημείων στον  $d$ -διάστατο χώρο με τα  $k$  σημεία στον  $n$ -διάστατο χώρο. Το σφάλμα αυτό είναι η συνάρτηση των  $d^*k$  μεταβλητών των  $grq$  όπου  $p=1,2,3,\dots,k$  και  $q=1,2,3,\dots,k$  και υπολογίζεται ως εξής:

$$E = \frac{1}{\sum_{i < j} [d_{ij}^*]} \sum_{i < j} \frac{[d_{ij}^* - d_{ij}]^2}{d_{ij}^*}$$

Στο τελευταίο βήμα του αλγορίθμου προσαρμόζουμε τις μεταβλητές  $grq$  ή ισοδύναμα αλλάζουμε τον σχηματισμό των διανυσμάτων  $Y$  στον  $d$ -διάστατο χώρο, ώστε να μειώσουμε το σφάλμα  $E$  μέχρι να πετύχουμε την ελαχιστοποίησή του. Το διάστημα τιμών του σφάλματος είναι  $[0,1]$  με την τιμή 0 να υποδηλώνει την τέλεια αναπαράσταση των δεδομένων σε χώρο χαμηλότερης διάστασης (χωρίς καμία απώλεια πληροφορίας).

Οι χάρτες του Sammon μπορούν να κατασκευαστούν με τη βοήθεια του Matlab, όπως θα δείξουμε στη πράξη σε επόμενη ενότητα.

Άλλες μέθοδοι εξαγωγής χαρακτηριστικών είναι:

- Semidefinite embedding [Saul04]
- Multifactor dimensionality reduction [Hahn01]
- Multilinear subspace learning [Plat11]
- Nonlinear dimensionality reduction [Verl07]
- Auto-Encoder [Hint08]

#### **2.1.4 Επιλογή χαρακτηριστικών**

Η επιλογή χαρακτηριστικών αποσκοπεί στην εύρεση των καταλληλότερων χαρακτηριστικών που διαφοροποιούν καλύτερα ένα αντικείμενο από άλλα, δηλαδή στην επιλογή εκείνων που περισσότερο διαχωρίσιμες κλάσεις.

Τρεις βασικές κατηγορίες μεθόδων είναι:

- Φίλτρουν (filter)
- Περιβάλλουσες (wrapper)
- Ενσωματωμένες (embedded)

Οι δυο πρώτες βασίζονται κυρίως σε στατιστικό κριτήριο ενώ η τρίτη στον ταξινομητή.

Ενδεικτικό παράδειγμα μεθόδου φίλτρου αποτελεί η μέθοδος της βαθμονόμησης βάσει του στατιστικού κριτηρίου t-test. Σύμφωνα με το Welch's t-test για κάθε χαρακτηριστικό  $f$  στον πίνακα των δεδομένων μας υπολογίζεται η μέση τιμή  $m$  και η τυπική απόκλιση  $s$  για κάθε κλάση. Η ανάλυση συσχέτισης επίσης μπορεί να χρησιμοποιηθεί για την επιλογή  $N$  χαρακτηριστικών. Παραδείγματα περιβάλλουσων μεθόδων είναι η SFFS (Sequential Forward Floating Selection) και η SBS (Sequential Backwards Selection). Χαρακτηρίζονται από ένα δυναμικά μεταβαλλόμενο αριθμό των χαρακτηριστικών και δίνουν αρκετά καλά αποτελέσματα. Μειονεκτήματα των στατιστικών μεθόδων, όπως η filter method, είναι ότι απαιτούν μεγάλο πλήθος διανυσμάτων.

#### **2.1.5 Ταξινόμηση σε Κλάσεις**

Η κατηγοριοποίηση (classification) [Vazi03] είναι μία τεχνική της εξόρυξης δεδομένων, κατά την οποία ένα στοιχείο ανατίθεται σε ένα προκαθορισμένο σύνολο κατηγοριών. Ο όρος κατηγοριοποίηση συναντάται στην βιβλιογραφία και ως *ταξινόμηση*. Γενικότερα, ο στόχος της διαδικασίας αυτής είναι η ανάπτυξη ενός μοντέλου, το οποίο αργότερα θα μπορεί να χρησιμοποιηθεί για την κατηγοριοποίηση μελλοντικών δεδομένων. Τέτοια παραδείγματα είναι ο διαχωρισμός των emails με βάση την επικεφαλίδα τους ή το περιεχόμενό τους, η πρόβλεψη καρκινικών κυττάρων χαρακτηρίζοντας τα ως καλοήθη ή κακοήθη, η κατηγοριοποίηση πελατών μιας τράπεζας ανάλογα με την πιστωτική τους ικανότητα κ.α. Στη παρούσα εργασία, σκοπός μας είναι να κατηγοριοποιήσουμε – ταξινομήσουμε τα χαρακτηριστικά διανύσματα που προκύπτουν από την εξαγωγή χαρακτηριστικών που αναφέραμε παραπάνω.

Η ταξινόμηση μπορεί να περιγραφεί ως μία διαδικασία δύο βημάτων [Fish36]:

- Εκμάθηση (Learning) : Στο πρώτο βήμα της διαδικασίας δημιουργείται/προσδιορίζεται το μοντέλο με βάση ένα σύνολο προκατηγοριοποιημένων παραδειγμάτων, που ονομάζεται δεδομένα εκπαίδευσης (training data). Τα δεδομένα εκπαίδευσης αναλύονται από ένα

αλγόριθμο κατηγοριοποίησης, προκειμένου να σχηματιστεί το μοντέλο. Λόγω του ότι τα δεδομένα εκπαίδευσης ανήκουν σε μία προκαθορισμένη κατηγορία, η οποία είναι γνωστή, η κατηγοριοποίηση αποτελεί μέθοδος εποπτευομένης μάθησης (supervised learning). Το μοντέλο, που λέγεται και αλλιώς κατηγοριοποιητής (classifier), αναπαρίσταται με τη μορφή κανόνων κατηγοριοποίησης(classification rules), δέντρων απόφασης(decision trees) ή μαθηματικών τύπων.

- Ταξινόμηση (Classification): Μετά την δημιουργία του μοντέλου, το επόμενο βήμα είναι η αξιολόγησή του. Για να επιτευχθεί αυτό, χρησιμοποιούμε τα δοκιμαστικά δεδομένα (test data) για να υπολογίσουν την ακρίβεια του μοντέλου. Το μοντέλο κατηγοριοποιεί τα δοκιμαστικά δεδομένα. Έπειτα, η κατηγορία που σχηματίστηκε με βάση τα δοκιμαστικά δεδομένα συγκρίνεται με την πρόβλεψη που έγινε για τα δεδομένα εκπαίδευσης, τα οποία είναι ανεξάρτητα από αυτά της δοκιμής. Η ακρίβεια του μοντέλου υπολογίζεται από το ποσοστό των δειγμάτων δοκιμής που κατηγοριοποιήθηκαν σωστά σε σχέση με το υπό εκπαίδευση μοντέλο.

Δεδομένης μιας διαχωριστικής επιφάνειας τα σημεία που αντιστοιχούν στα διανύσματα χαρακτηριστικών ταξινομούνται σε κλάσεις. Πολλά χαρακτηριστικά κάνουν το πρόβλημα της ταξινόμησης πιο πολύπλοκο, γι' αυτό είναι επιθυμητό το μικρότερο δυνατό αλλά επαρκές σύνολο χαρακτηριστικών. Τα χαρακτηριστικά μας πρέπει να παρουσιάζουν μεταξύ τους όσο το δυνατόν μικρότερη συσχέτιση.

Ενδεικτικά θα αναφέρουμε μερικές μεθόδους και αλγορίθμους ταξινόμησης:

- Ταξινομητής μέγιστης εντροπίας: Σημειώστε ότι η λογιστική παλινδρόμηση είναι ένας αλγόριθμος για ταξινόμηση παρά την ονομασία του. (Το όνομα προέρχεται από το γεγονός ότι η λογιστική παλινδρόμηση χρησιμοποιεί μια επέκταση ενός γραμμικού μοντέλου παλινδρόμησης για να μοντελοποιήσει την πιθανότητα μιας εισόδου να ανήκει σε μια συγκεκριμένη κλάση.) [Stan00]
- Ταξινομητής Naive Bayes [Zhan09]
- Διάνυσμα υποστήριξης μηχανής (Support Vector Machines - SVM) [Cort95]
- Αλγόριθμοι εκτίμησης πυρήνα και K-κοντινότερου γείτονα [Coom82]
- Perceptron [Rose57]
- Νευρωνικά δίκτυα (perceptron πολλών επιπέδων)

## **2.2 Ανάκληση εικόνων βάσει περιεχομένου**

Σήμερα, η ψηφιακή επεξεργασία εικόνας χρησιμοποιείται ευρέως στα συστήματα ανάκλησης εικόνων με βάση το περιεχόμενο (CBIR – Content based Image Retrieval systems). Τέτοια συστήματα έκαναν την εμφάνισή τους κυρίως στις αρχές της δεκαετίας του 1990 [Smeu00].

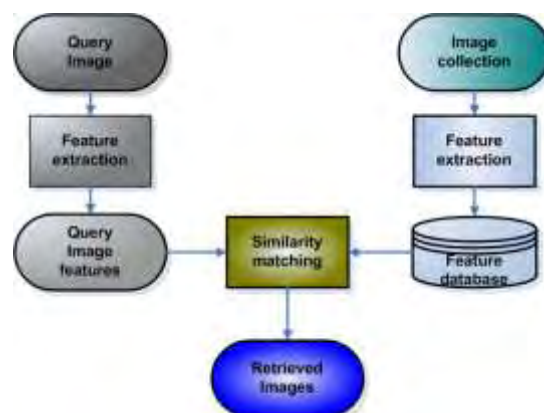
Ο λόγος δημιουργίας τέτοιων συστημάτων όμως εντοπίζεται σε παλαιότερες χρονικές περιόδους. Η αυξανόμενη τάση των ανθρώπων για χρήση του Διαδικτύου, σε συνδυασμό με την ανάπτυξη της

τεχνολογίας, έφερε στη διάθεση ενός μέσου χρήστη υπολογιστικών συστημάτων συλλογές μεγάλου πλήθους ψηφιακών εικόνων. Ως επακόλουθο υπήρξε η ανάγκη ομαδοποίησης των εικόνων σε επιμέρους κατηγορίες, με σκοπό την εύρεση συγκεκριμένων από αυτές, όταν ήταν αυτό απαραίτητο.

Αρχικά προτάθηκε η σύνδεση συγκεκριμένων μεταδεδομένων (metadata) -δηλαδή ορισμένες λέξεις κλειδιά, ετικέτες, ακόμη και περιγραφή με τη μορφή κειμένου- με κάθε εικόνα και ολοκλήρωση της εκάστοτε αναζήτησης με κριτήριο τα δεδομένα αυτά. Είναι σαφές όμως, ότι η μεθοδολογία αυτή διαθέτει αρκετά μειονεκτήματα. Εμφανέστερο πρόβλημα της χειροκίνητης απόδοσης μεταδεδομένων από ανθρώπους σε μεγάλο πλήθος εικόνων είναι μη αποδοτικότητα, το υψηλό κόστος της διαδικασίας αυτής, όπως επίσης και η μη απόδοση όλων των νοημάτων που απαντώνται σε μία εικόνα.

Βελτίωση της ανωτέρω μεθοδολογίας είναι η αναζήτηση με βάση τα πραγματικά δεδομένα κάθε εικόνας, δηλαδή το οπτικό της περιεχόμενο. Τα συστήματα αυτά είναι τα CBIR συστήματα και σε αντίθεση με τα προγενέστερά τους, είναι ικανά να επιστρέφουν πιο αξιόπιστα αποτελέσματα στο χρήστη τους.

Στο παρακάτω σχήμα (εικόνα 14) απεικονίζεται η λειτουργία ενός τυπικού CBIR συστήματος. Παρατηρώντας την είσοδο και την έξοδο του συστήματος, θα λέγαμε ότι ως είσοδο έχουμε μία εικόνα επερώτηση (query image) που δίνει ο χρήστης, ενώ ως έξοδο επιστρέφονται οι εικόνες με όμοιο περιεχόμενο (retrieved images). Το σύστημα αρχικά εξάγει κάποιες τιμές σε χαρακτηριστικά από την εικόνα επερώτηση (Feature extraction), ενώ στη συνέχεια εξάγονται οι τιμές των ίδιων χαρακτηριστικών σε όλες τις εικόνες της δεδομένης βάσης (Image collection). Έπειτα, πραγματοποιείται μία τεχνική σύγκρισης (similarity matching) μεταξύ των τιμών των χαρακτηριστικών της εικόνας επερώτησης και των τιμών των χαρακτηριστικών όλων των εικόνων της βάσης και επιστρέφονται στο χρήστη εκείνες οι εικόνες της βάσης (Retrieved images), που θεωρείται ότι «ταιριάζουν» περισσότερο με την εικόνα επερώτησης.



Εικόνα 14

Πραγματοποιώντας μία σύντομη ιστορική αναδρομή στα CBIR συστήματα [Velt00], παρατηρούμε ότι τα πρώτα τέτοια συστήματα υλοποιήθηκαν τη δεκαετία του 1980 [Chan80]. Όμως, το πρώτο εμπορικό CBIR σύστημα, που καθιερώθηκε, ήταν το QBIC (Query by Image Content) από την IBM το 1993 [Nibl93]. Συνέχεια δόθηκε με τα CBVQ (Query by Visual Query) [Smit95], WebSEEK [Smit1997] και VisualSEEK [Smit97]. Έπειτα, η ακαδημαϊκή κοινότητα αρχίζει την ανάπτυξη αντίστοιχων συστημάτων και πολλά CBIR συστήματα έρχονται στο προσκήνιο, όπως τα MARS [Orte97], MetaSEEK [Beni98], PhotoBook [Pent96], Picasso [Bimb97], Wise [Wang97] και Surfimage [Nast98]. Τα πιο σύγχρονα, που κατασκευάζονται μετά το 2000 (Blobworld [Cars99], PicHinter [Cox02], PicSOM [Laak00], PicToSeek [Geve00] και SIMPLiCity [Wang01]), έχουν υψηλότερη απόδοση και μεγαλύτερα ποσοστά επιτυχίας. Την τωρινή εποχή συνεχίζει και αναπτύσσεται ο επιστημονικός αυτός τομέας, ενώ υπάρχουν και online αρκετά CBIR συστήματα, δημόσια προσβάσιμα σε χρήστες. Αναφέρουμε τα Anaktisi [Chat10], FIRE, imgSEEK και SIMBA.

Στο παρακάτω σχήμα (εικόνα 15) απεικονίζουμε το γραφικό περιβάλλον (interface) μερικών CBIR συστημάτων και πιο συγκεκριμένα των Anaktisis και BRISC.



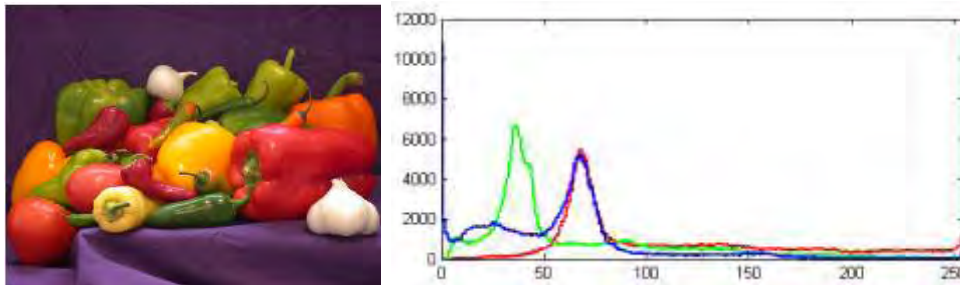
Εικόνα 15

Στην παραπάνω παράγραφο αναφερθήκαμε στην έννοια της εξαγωγής τιμών σε συγκεκριμένα χαρακτηριστικά (feature extraction). Πρόκειται για χρήση συγκεκριμένων μαθηματικών μοντέλων, με σκοπό να προκύψουν τιμές, που χαρακτηρίζουν την εικόνα. Προφανώς, όμοιες εικόνες θα πρέπει να έχουν και όμοιες τιμές σε κάθε χαρακτηριστικό, στην περίπτωση που τα επιλεγμένα χαρακτηριστικά είναι αποδοτικά και περιγράφουν σωστά τις εκάστοτε εικόνες.

### 2.3 Βιβλιογραφική ανασκόπηση μεθόδων εξαγωγής χαρακτηριστικών

Όπως είδαμε στη παράγραφο 1.2 κάποιες απ τις βασικές κατηγορίες χαρακτηριστικών σχήματος είναι το χρώμα , η υφή και το σχήμα. Σε αυτή την ενότητα θα αναφερθούμε στις μεθόδους εξαγωγής χαρακτηριστικών και πως αυτές έχουν βρει εφαρμογή μέσα απο διάφορες εργασίες που έχουν εκπονηθεί πάνω στο συγκεκριμένο αντικείμενο.

Ξεκινώντας με τα *χρωματικά ιστογράμματα* (color histograms), πρόκειται για αναπαράσταση της κατανομής της χρωματικής πληροφορίας εντός μίας εικόνας [Wang09] και χρησιμοποιείται για την ανάλυση κάθε μίας από τις χρωματικές συνιστώσες. Στην πράξη γίνεται χρήση χρωματικών ιστογραμμάτων για τα χρωματικά μοντέλα RGB και HSV, όπως απεικονίζεται και στο ακόλουθο σχήμα(εικόνα 16).



**Εικόνα 16, χρωματικό ιστόγραμμα**

Σε σχέση με τις *χρωματικές ροπές*, ένα χρωματικό ιστόγραμμα είναι αμετάβλητο στην περιστροφή και στη μετατόπιση του αντικειμένου, ενώ μπορεί να υπολογιστεί και σε υποσύνολο της εικόνας. Βασικό, όμως, μειονέκτημα είναι η εξάρτησή του μόνο από το χρώμα και όχι από το σχήμα ή την υφή τους προς εξέταση αντικειμένου. Είναι πιθανό, δύο διαφορετικές εικόνες να έχουν ίδια χρωματικά ιστογράμματα, επειδή τυγχάνει να έχουν αντικείμενα με ίδια χρωματική πληροφορία.

Προσπαθώντας την επίτευξη βελτίωσης της παραπάνω μεθοδολογίας, έχουν προταθεί διάφορες τεχνικές, όπως τα χρωματικά διανύσματα συνοχής (color coherence vectors) [Pass99], όπου εφαρμόζεται κατάτμηση της εικόνας σε δύο επιμέρους περιοχές και μετέπειτα υπολογισμός των ιστογραμμάτων σε κάθε μία από τις περιοχές αυτές. Πιο πολύπλοκες τεχνικές, όπως το χρωματικό διάγραμμα συσχέτισης (color correlogram) [Huan97] λαμβάνουν υπόψη και τη θέση του εκάστοτε pixel στην εικόνα (spatial information), απαιτούν όμως περισσότερο χρόνο για την εκτέλεση πιο περίπλοκων αλγορίθμων.

Όλα τα παραπάνω αποτελούν πρωταρχικές μεθοδολογίες, με σκοπό την εκμετάλλευση της χρωματικής πληροφορίας της εικόνας. Σήμερα, έχουν προταθεί πλήθος πολυπλοκότερων, αλλά αποδοτικότερων

αλγορίθμων, που ξεφεύγουν όμως από τους σκοπούς της παρούσης εισαγωγής. Για το λόγο αυτό και παραθέτουμε αντίστοιχη βιβλιογραφία, με σκοπό την περαιτέρω αναζήτηση [Geve04], [Smeu00], [Datt05].

Ως χαρακτηριστικά, που περιγράφουν την υφή μίας μονοχρωματικής εικόνας, ορίζονται οι *στατιστικές ροπές* (statistical moments) ως προς τη μέση τιμή [Shar01] διαφόρων τάξεων του ιστογράμματος της εικόνας, με βάση τον ακόλουθο τύπο.

$$\mu_n(u) = \sum_{i=0}^{L-1} (u_i - \langle u \rangle)^n p(u_i)$$

όπου  $n$  είναι η τάξη της ροπής που θέλουμε να υπολογίζουμε,  $u_i$  τα  $L$  διαφορετικά επίπεδα του γκρι της εικόνας,  $p(u_i)$  η πιθανότητα εμφάνισης του κάθε επιπέδου στην εικόνα –υπολογισμένο από το ιστογράμμα της- και  $\langle u \rangle$  η στατιστική μέση τιμή του επιπέδου γκρι, δηλαδή

$$\langle u \rangle = \sum_{i=0}^{L-1} u_i p(u_i)$$

Οι παραπάνω ροπές των τεσσάρων πρώτων τάξεων ονομάζονται αντίστοιχα μέση τιμή (mean value), διακύμανση (variance), στρέβλωση (skewness) και κύρτωση (kurtosis) και χαρακτηρίζουν τη μορφή του ιστογράμματος. Ειδικά, αναφέρουμε ότι η διακύμανση χρησιμοποιείται ευρέως ως ένα μέγεθος που προσδιορίζει την αντίθεση (contrast) της εικόνας. Στον παρακάτω πίνακα αναφέρουμε συνοπτικά και κάποια άλλα μεγέθη που υπολογίζονται από το ιστογράμμα της εικόνας.

<b>Χαρακτηριστικά που υπολογίζονται από ιστογράμμα της εικόνας</b>	
$p_{\max} = \max_k \{p(u_k)\}$	Μέγιστη πιθανότητα (maximum probability)
$u_{\max} = \arg \max_k \{p(u_k)\}$	Επικρατούσα τιμή (mode)
$u_{\text{med}} = \text{median}\{p(u_k)\}$	Διάμεσος (median)
$m_1 = \langle u \rangle = E[u] = \sum_{k=0}^{L-1} u_k p(u_k)$	Στατιστικός μέσος όρος (statistical average gray level)
$m_n = E[u^n] = \sum_{k=0}^{L-1} u_k^n p(u_k)$	Ροπή τάξης $n$ (moments of order $n$ )
$\hat{m}_n = E[ u ^n] = \sum_{k=0}^{L-1}  u_k ^n p(u_k)$	Απόλυτη ροπή τάξης $n$ (absolute moments of order $n$ )
$\mu_n = E[(u - E[u])^n] = \sum_{k=0}^{L-1} (u_k - m_1)^n p(u_k)$	Κεντρικές ροπές τάξης $n$ (central moments of order $n$ )
$\sigma^2 = \mu_2 = \sum_{k=0}^{L-1} (u_k - m_1)^2 p(u_k)$	Διακύμανση (variance)



$\hat{\mu}_n = E[ u - E[u] ^n] = \sum_{k=0}^{L-1}  u_k - m_1 ^n p(u_k)$	Απόλυτη κεντρική ροπή τάξης n (absolute central moments)
$m_{-n} = E[u^{-n}] = \sum_{k=0}^{L-1} \frac{p(u_k)}{u_k^n}, u_k \neq 0$	Αντίστροφη ροπή τάξης n (inverse moment of order n)
$E(u) = E[p] = \sum_{k=0}^{L-1} p^2(u_k)$	Ομοιομορφία ή ενέργεια (uniformity or energy)
$H = E[-\log_2 p] = -\sum_{k=0}^{L-1} p(u_k) \log_2 p(u_k)$	Εντροπία (entropy)
$H_a = \frac{1}{1-a} \log_2 \left( \sum_{k=0}^{L-1} p(u_k)^a \right)$	Εντροπία Renyi τάξης a (Reni entropy of order a)

Προσπαθώντας να ορίσουμε ροπές αναλλοίωτες ως προς τη μεταφορά, την περιστροφή και την αλλαγή κλίμακας, ξεκινάμε από τις κεντρικές ροπές  $\mu_{pq}$ , που υπολογίζονται όπως παρακάτω και είναι αμετάβλητες ως προς τη μεταφορά, καθώς υπολογίζονται με κριτήριο την απόσταση κάθε σημείου από το κεντροειδές  $(\bar{x}, \bar{y})$ , όπου  $\bar{x} = \frac{m_{10}}{m_{00}}$  και  $\bar{y} = \frac{m_{01}}{m_{00}}$ .

$$\mu_{pq} = \sum_x \sum_y (x - \bar{x})^p (y - \bar{y})^q f(x, y)$$

Έπειτα, ορίζουμε τις κανονικοποιημένες κεντρικές ροπές  $n_{pq}$ , οι οποίες είναι αμετάβλητες και ως προς την αλλαγή κλίμακα ακολούθως:

$$n_{pq} = \frac{\mu_{pq}}{\mu_{00}^\gamma}$$

Όπου  $\gamma = 1 + \frac{p+q}{2}$ .

Οι γεωμετρικές ροπές είναι ευρέως διαδεδομένες και έχουν χρησιμοποιηθεί σε πλήθος εφαρμογών ανάλυσης εικόνων [Liao96], [Teh88], [Gonz07], [Prok92], [Meht97], [Sajj97]. Παρόλα αυτά, δεν θα πρέπει να παραλείψουμε και τα μειονεκτήματά τους [Cele05]. Το βασικότερο από αυτά είναι ότι η χρήση ροπών χαμηλών τάξεων δεν αρκεί για να περιγράψει πλήρως το περίγραμμα ενός σχήματος. Αναγκαία είναι, λοιπόν, η χρήση ροπών αρκετά υψηλών τάξεων, οι οποίες όμως είναι αλγοριθμικά πολύπλοκο να υπολογισθούν και επίσης εμφανίζουν ευαισθησία στην παρουσία θορύβου.

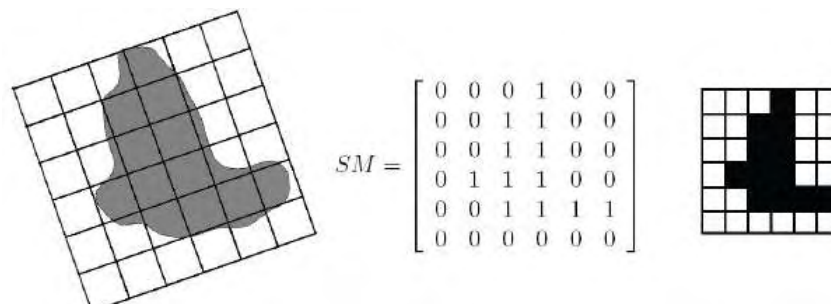
Επόμενο χαρακτηριστικό της κατηγορίας αυτής αποτελούν οι αλγεβρικές ροπές [Taub91], οι οποίες μάλιστα έχουν χρησιμοποιηθεί στο σύστημα ανάκλησης εικόνων QBIC [Nibl93] που αναφέρθηκε παραπάνω. Σημείο εκκίνησης για τον υπολογισμό των αλγεβρικών ροπών αποτελούν οι κεντρικές ροπές. Με βάση τις πρώτες m κεντρικές ροπές κατασκευάζονται ορισμένοι πίνακες  $M_{[j,k]}$ . Από τις



ιδιοτιμές των πινάκων αυτών εξάγονται οι τιμές των αλγεβρικών ροπών. Σε αντίθεση με τις ροπές  $H_u$ , οι αλγεβρικές ροπές είναι αμετάβλητες στους συγγενείς μετασχηματισμούς (affine invariance). Όμως δεν επιτυγχάνουν σταθερή -γενικά- απόδοση στις διαφορετικές εφαρμογές που έχουν χρησιμοποιηθεί. Όπως χαρακτηριστικά αναφέρεται και στη σχετική βιβλιογραφία [Scas94], οι αλγεβρικές ροπές είναι ικανές να αποδώσουν είτε πολύ καλά είτε πολύ άσχημα αποτελέσματα, χαρακτηριστικό που δεν τις καθιστά αξιόπιστες, ώστε να εφαρμοστούν σε πλήθος συστημάτων.

Στην [Teh88] πραγματοποιήθηκε μία εκτενής μελέτη και σύγκριση μεταξύ των διάφορων ροπών, όπως ορθογώνιες ροπές (Zernike, Legendre και pseudo-Zernike) και μη ορθογώνιες ροπές (γεωμετρικές ροπές, γενικευμένες σύνθετες ροπές (generalized complex moments) [Abo88]). Από τα συμπεράσματα της εργασίας αυτής αντιλαμβανόμαστε ότι οι γεωμετρικές ροπές, οι σύνθετες ροπές και οι pseudo-Zernike ροπές είναι οι πλέον ανεκτικές στο θόρυβο, σε αντίθεση με τις ροπές Legendre που μεταβάλλονται αρκετά παρουσία θορύβου. Η μελέτη καταλήγει στις Zernike και pseudo-Zernike ροπές ως τις βέλτιστες -σε σχέση με τις υπόλοιπες της σύγκρισης- για την περιγραφή του σχήματος αντικειμένων. Οι ροπές Zernike παραμένουν αναλλοίωτες στην περιστροφή του σχήματος, είναι ανεκτικές στο θόρυβο και σε μικρές αλλαγές του περιγράμματος του σχήματος και επιπλέον δεν περιέχουν περιττή πληροφορία εξαιτίας της ορθογωνιότητας της βάσης από την οποία σχηματίζονται. Λόγω των ανωτέρω, έχουν επιλεγεί και ως χαρακτηριστικό region-based από το πρότυπο MPEG-7 [Jean00].

Στη συνέχεια αναφέρουμε και άλλα είδη ροπών που έχουν χρησιμοποιηθεί για την περιγραφή του σχήματος ενός αντικειμένου, όπως τις *ροπές Zernike* [Khot90], [Shut02], *οι ροπές Legendre* [Chon04] ή *οι ακτινικές Chebyshev ροπές* (radial chebyshev moments) [Muku04] καθώς και μεθοδολογίες που βασίζονται σε τεχνικές πλέγματος για την εξαγωγή χαρακτηριστικών τιμών. Ονομάζονται grid based descriptors [Lu99], [Flus92] και εξάγονται υπολογίζοντας ολόκληρη την περιοχή του -προς ταξινόμηση- αντικειμένου. Στην τεχνική αυτή, αρχικά τοποθετείται ένα πλέγμα συγκεκριμένου πλήθους κελιών επάνω στην εικόνα. Το πλέγμα αυτό σαρώνεται από αριστερά προς τα δεξιά, και από το επάνω προς τα κάτω μέρος της εικόνας. Το αποτέλεσμα είναι μία δυαδική εικόνα, με τιμή 1 στα κελιά που αντιστοιχούν σε σημείο εντός του αντικειμένου και 0 εκτός αυτού, όπως διαφαίνεται από το ακόλουθο σχήμα (εικόνα 17).



Εικόνα 17

Άρα, ολόκληρο το αντικείμενο αναπαριστάται από ένα διάνυσμα με δυαδικές τιμές. Η σύγκριση δύο διαφορετικών αντικειμένων μετατρέπεται -ουσιαστικά- στη σύγκριση των αντίστοιχων διανυσμάτων, όπου και χρησιμοποιούνται ειδικές μετρικές, όπως η απόσταση Hamming ή η απόσταση city block. Ακόμη, προηγμένες τεχνικές έχουν προταθεί, με σκοπό την αναλλοίωτη συμπεριφορά των μετρικών στην μετακίνηση, την περιστροφή ή την αλλαγή κλίμακας του αντικειμένου. Οι τεχνικές πλέγματος, που παρουσιάστηκαν στην παράγραφο αυτή, θεωρούνται απλοϊκοί στην αναπαράσταση και την υλοποίηση, είναι όμως απαιτητικοί στους υπολογιστικούς πόρους που χρειάζονται για να εξαχθούν οι τιμές τους. Παρόλα αυτά, έχουν χρησιμοποιηθεί σε πραγματικά συστήματα ανάκλησης εικόνων, όπως το σύστημα MARS [Orte97].

Η επιστημονική κοινότητα, προσπαθώντας να βελτιώσει τις παραπάνω τεχνικές πλέγματος και να αποφύγει την επιπλέον επεξεργασία για την αναλλοίωτη συμπεριφορά των μετρικών αυτών, κατέληξε στις shape matrix τεχνικές βασισμένες σε μοντέλο με πολικές συντατεγμένες (polar model) [Gosh85], [Taza89] και όχι σε τετραγωνικό μοντέλο (square model), όπως το παραπάνω. Στην περίπτωση αυτή εντοπίζεται το κέντρο βάρους του αντικειμένου και χτίζεται ένα πλέγμα γύρω από αυτό. Το πλέγμα όμως βασίζεται σε ομόκεντρους κύκλους και το διαχωρισμό τους σε επιμέρους τόξα, όπως φαίνεται στο παρακάτω σχήμα (εικόνα 18). Ο πίνακας θα προκύψει από τα σημεία τομής των κύκλων και των ευθειών. Αν στα σημεία αυτά βρισκόμαστε εντός του αντικειμένου, η αντίστοιχη θέση του πίνακα θα λάβει τιμή 1, αλλιώς θα λάβει τιμή 0.

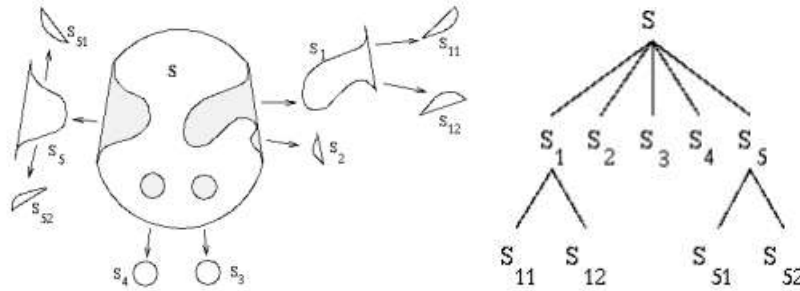


**Εικόνα 18, σχηματική αναπαράσταση τεχνικής shape matrix**

Διαφορετικές προσεγγίσεις, με σκοπό την εξαγωγή region-based χαρακτηριστικών αναλύουμε στη συνέχεια. Στις μεθοδολογίες που θα ακολουθήσουν, το αρχικό σχήμα αναλύεται και διαχωρίζεται σε επιμέρους υπο-σχήματα, τα οποία μετέπειτα χρησιμοποιούνται, ώστε να περιγράψουν και να αναλύσουν ολόκληρη την περιοχή ενδιαφέροντος. Οι τεχνικές αυτές χαρακτηρίζονται ως structural τεχνικές και αποτελούν ξεχωριστή υποκατηγορία των region-based χαρακτηριστικών σχήματος.

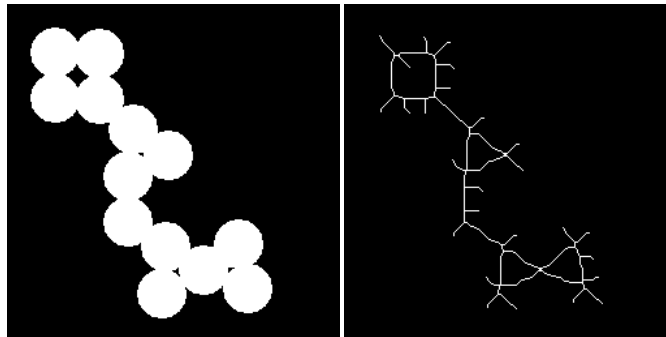
Στις εργασίες [Hlav93], [Gonz07] και [Davi04] χρησιμοποιείται η έννοια της κυρτής περιοχής (convex region). Μία περιοχή θεωρείται κυρτή, αν για κάθε ζεύγος σημείων της  $x_1, x_2$ , ολόκληρο το ευθύγραμμο τμήμα  $x_1x_2$  βρίσκεται εντός της περιοχής αυτής. Ως convex hull μίας περιοχής R, ορίζουμε την ελάχιστη κυρτή περιοχή H, για την οποία ισχύει  $R \subset H$ , ενώ η διαφορά H-R καλείται convex deficiency D της περιοχής R.

Πριν τη διαδικασία εξαγωγής του convex hull μίας περιοχής προηγούνται συνήθως τεχνικές που εξομαλύνουν το περίγραμμα του αντικειμένου, με σκοπό την ομαλοποίηση του ορίου και την αφαίρεση του αντίστοιχου θορύβου. Η σύγκριση δύο διαφορετικών περιοχών  $R_1, R_2$  αντιστοιχεί σε μία σύγκριση των δενδροειδών δομών τους, χρησιμοποιώντας για παράδειγμα τεχνικές σύγκρισης γράφων. Ένα παράδειγμα της μεθοδολογίας, που περιγράψαμε, διαφαίνεται στο παρακάτω σχήμα(εικόνα 19) [Hlav93].



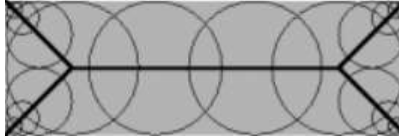
Εικόνα 19

Όμοια με την κυρτότητα μίας –προς ταξινόμηση- περιοχής θα μπορούσε να χρησιμοποιηθεί και ο σκελετός της περιοχής αυτής. Ως region skeleton ορίζεται ένα σύνολο από ευθείες (εντός του αντικειμένου), οι οποίες ενώνουν τα άκρα του σχήματος, όπως απεικονίζεται παρακάτω (εικόνα 20).



Εικόνα 20

Στην [Blum67] καταγράφεται μία προσπάθεια απεικόνισης των χαρακτηριστικών μίας περιοχής μέσω του σκελετού της περιοχής αυτής και προέκυψε ο μετασχηματισμός medial axis (medial axis transform - MAT). Ο μετασχηματισμός αυτός ορίζεται ως ο γεωμετρικός τόπος των σημείων που αντιστοιχούν στα κέντρα των κυκλικών δίσκων με την μεγαλύτερη δυνατή επιφάνεια, ώστε να μην ξεφεύγουν από τα όρια της περιοχής ενδιαφέροντος, όπως φαίνεται στο ακόλουθο σχήμα (εικόνα 21) [Mors94].

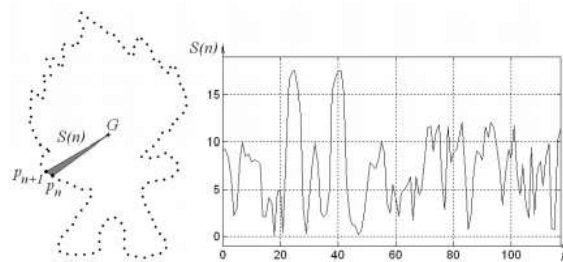


Εικόνα 21

Και στην περίπτωση αυτή, ο σκελετός διασπάται σε επιμέρους υπο-σκελετούς και καταλήγουμε σε ένα γράφο, ο οποίος προσδιορίζει το αντικείμενο. Στη συνέχεια, αντίστοιχες τεχνικές σύγκρισης γράφων λαμβάνουν χώρα, με σκοπό την εύρεση ομοιότητας δύο διαφορετικών εικόνων.

Συνοπτικά, σχετικά με τις structural μεθοδολογίες που παρουσιάσαμε, παρατηρούμε ότι δεν χρησιμοποιούνται ευρέως σε πραγματικές εφαρμογές. Βασικά τους μειονεκτήματα είναι η πολυπλοκότητα των αντίστοιχων αλγορίθμων και η υψηλή υπολογιστική ισχύς που απαιτούν. Ακόμη, βασίζονται σε τεχνικές σύγκρισης γράφων, ένα πρόβλημα που ακόμη και σήμερα θεωρείται δύσκολο προς επίλυση.

Άλλη τεχνική αποτελεί η *συνάρτηση περιοχής* (area function), που υπολογίζει το εμβαδόν του τριγώνου που σχηματίζουν δύο διαδοχικά σημεία του περιγράμματος και το κεντροειδές. Με τον τρόπο αυτό προκύπτει μία νέα συνάρτηση, όπως φαίνεται και στο παρακάτω σχήμα (εικόνα 22).



Εικόνα 22

Οι ροπές του περιγράμματος (boundary moments) υπολογίζονται από τις συναρτήσεις υπογραφής σχήματος [Hlav93]. Αν υποθέσουμε ότι έχουμε καταλήξει σε μία συνάρτηση  $z(i)$  υπογραφής σχήματος, τότε ορίζονται οι  $m_r$  (ροπές τάξης  $r$ ) και  $\mu_r$  (κεντρικές ροπές τάξης  $r$ ) ως ακολούθως:

$$m_r = \frac{1}{N} \sum_{i=1}^N [z(i)]^r$$

$$\mu_r = \frac{1}{N} \sum_{i=1}^N [z(i) - m_1]^r$$

Επιπλέον, οι κανονικοποιημένες ροπές  $\bar{m}_r = \frac{m_r}{\mu_2^{r/2}}$  και  $\bar{\mu}_r = \frac{\mu_r}{\mu_2^{r/2}}$  είναι αμετάβλητες ως προς την περιστροφή, τη μεταφορά και την αλλαγή κλίμακας, ενώ κι άλλες μετρικές με περισσότερη ανοχή στο θόρυβο μπορούν να οριστούν ως  $F_1 = \frac{\sqrt{\mu_2}}{m_1}$ ,  $F_2 = \frac{\mu_3}{\mu_2^{3/2}}$  και  $F_3 = \frac{\mu_4}{\mu_2^2}$ .

Άλλος τρόπος που παρουσιάζεται στη βιβλιογραφία [Gonz07], είναι να υπολογίσουμε το ιστόγραμμα  $p(v_i)$  της συνάρτησης  $z(i)$  και να παράγουμε τη ροπή  $r$  τάξης ως

$$\mu_r = \sum_{i=1}^K (v_i - m)^r p(v_i), \text{ όπου } m = \sum_{i=1}^K v_i p(v_i).$$

Στην περίπτωση που γίνει κανονικοποίηση του εμβαδού της  $z(i)$  στη μονάδα, τότε οι παραπάνω αναλλοίωτες ροπές υπολογίζονται ως

$$\mu_r = \sum_{i=1}^N (i - m)^r z(i), \text{ όπου } m = \sum_{i=1}^N iz(i).$$

Οι ροπές, που μόλις αναφέραμε, υπολογίζονται εύκολα και χωρίς μεγάλο υπολογιστικό κόστος. Παρόλα αυτά οι ροπες υψηλών τάξεων δεν είναι ικανές να συσχετισθούν με την οπτική αναπαράσταση ενός αντικειμένου, γεγονός που επιτρέπει τη χρήση μόνο των ροπών χαμηλών τάξεων.

Άλλη μεθοδολογία της κατηγορίας αυτής είναι οι τεχνικές *elastic matching*. Στην περίπτωση αυτή, ένα παραμορφώσιμο μοντέλο  $\phi(s)$  ορίζεται ως το άθροισμα ενός αυθεντικού μοντέλου  $\tau(s)$  και μίας παραμόρφωσης  $\theta(s)$ , ενώ η ομοιότητα ενός σχήματος με ένα αυθεντικό μοντέλο ορίζεται με χρήση συγκεκριμένης συνάρτησης [Bim97a]. Οι τεχνικές αυτές, αν και αποδίδουν υψηλά ποσοστά επιτυχίας, δεν χρησιμοποιούνται στην πράξη, καθώς απαιτούν υψηλή υπολογιστική ισχύ για τη σύγκριση δύο αντικειμένων.

Χρησιμοποιώντας ως σημείο εκκίνησης μία συνάρτηση υπογραφής σχήματος  $f$ , όπως αυτές που παρουσιάστηκαν παραπάνω, αναπτύσσονται και οι τεχνικές εύρεσης περιγραφέν σχήματος μέσω στοχαστικών μεθοδολογιών [Kash81], [Dubo86], [Das90], [Seki92]. Πιο συγκεκριμένα, πραγματοποιείται μία προσπάθεια μοντελοποίησης (με Auto-regressive μεθοδολογία) της συνάρτησης  $f$  και εύρεσης της τωρινής τιμής  $f_t$  ως ένα γραμμικό συνδυασμό παλαιότερων  $M$  τιμών της με βάση τον ακόλουθο τύπο.

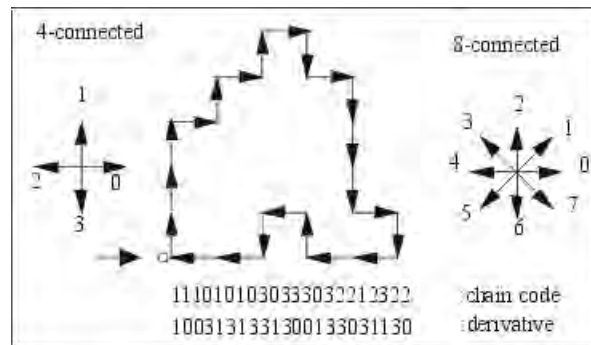
$$f_t = a + \sum_{j=1}^m \theta_j f_{t-j} + \sqrt{\beta} \omega_t$$

όπου  $\theta_j$  είναι οι συντελεστές του μοντέλου,  $m$  είναι η τάξη του μοντέλου,  $\sqrt{\beta} \omega_t$  είναι το σφάλμα στο βήμα  $t$  και αντιστοιχεί στην ακρίβεια της πρόβλεψης και το  $a$  είναι ανάλογο της μέσης τιμής της  $f$ . Οι τιμές των παραμέτρων, δηλαδή των  $a, \theta_j, \beta$ , υπολογίζονται με χρήση του κριτηρίου ελαχίστων τετραγώνων (least square). Οι μεθοδολογίες αυτές μειονεκτούν ως προς την κατάλληλη επιλογή των

παραμέτρων του προβλήματος. Σε περίπτωση περίπλοκου περιγράμματος απαιτείται ικανό πλήθος παραμέτρων, γεγονός που οδηγεί όμως στην ανάγκη υψηλής υπολογιστικής ισχύς. Γενικά, η επιλογή του  $m$  γίνεται εμπειρικά, καθώς οι παράμετροι  $\theta_j$  δεν έχουν καμία φυσική σημασία.

Υποκατηγορία των contour-based χαρακτηριστικών αποτελούν αυτά που διαχωρίζουν ολόκληρο το περίγραμμα σε επιμέρους τμήματα (structural shape representation), όπως οι structural τεχνικές των region-based χαρακτηριστικών, που αναφέρθηκαν παραπάνω. Κάθε επιμέρους τμήμα επεξεργάζεται και το τελικό αποτέλεσμα προκύπτει από τις τιμές αυτές.

Αρχικά, αναφέρουμε τις τεχνικές αναπαράστασης του περιγράμματος με κώδικα αλύσσου (chain code) [Free61], [Hu62]. Στην περίπτωση αυτή, ένα πλέγμα (grid) δύο διαστάσεων τοποθετείται επάνω στην εικόνα και σημειώνονται τα τετράγωνα που αντιστοιχούν σε περιοχή του περιγράμματος. Το αποτέλεσμα κωδικοποιείται με κώδικα αλυσίδας, χρησιμοποιώντας 8 ή 4 κατευθύνσεις, όπως απεικονίζονται στο παρακάτω σχήμα (εικόνα 23). Άρα, εκκινώντας τη διαδικασία από ένα συγκεκριμένο σημείο, προκύπτει τελικά μία συμβολοσειρά, που χαρακτηρίζει πλήρως το εκάστοτε περίγραμμα.



Εικόνα 23, κώδικας αλυσίδας

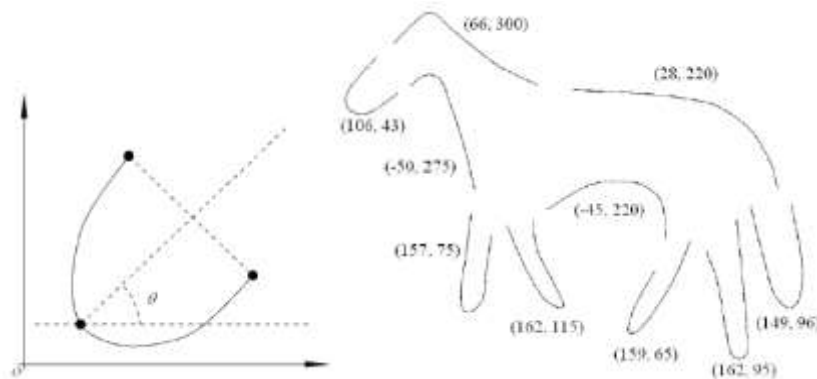
Με σκοπό την εύρεση ίδιου κώδικα, ανεξάρτητα με το σημείο εκκίνησης προτάθηκε ο διαφορικός κώδικας αλύσσου (differential chain code), όπου σε κάθε σημείο κωδικοποιείται η διαφορά της νέας τιμής με την αμέσως προηγούμενη και εφαρμόζεται στο αποτέλεσμα η πράξη modulo. Άλλες τεχνικές καταλήγουν στον κανονικοποιημένο κώδικα αλύσσου, ο οποίος δεν μεταβάλλεται με την περιστροφή του περιγράμματος. Οι συγκεκριμένες τεχνικές, όμως, εμφανίζουν μειονεκτήματα, όπως για παράδειγμα η υψηλή ευαισθησία στην παρουσία θορύβου. Παρόλα αυτά, χρησιμοποιούνται και από άλλες μεθοδολογίες, όπως οι μεθοδολογίες προσέγγισης μέσω πολυγώνου.

Οι μεθοδολογίες προσέγγισης μέσω πολυγώνου (polygon approximation) αναλύουν στο σχήμα του περιγράμματος σε ευθύγραμμα τμήματα [Gros90] [Gros92]. Το σχήμα προσεγγίζεται με ένα πολύγωνο, οι κορυφές των οποίων αποτελούν τα χρήσιμα δεδομένα, όπως διαφαίνεται από το παρακάτω σχήμα. Για κάθε κορυφή αποθηκεύονται τέσσερις τιμές (η γωνία, η απόσταση από την επόμενη κορυφή και οι συντεταγμένες), ενώ το σύνολο των τιμών για όλες τις κορυφές αποτελεί το χαρακτηριστικό διάνυσμα

του σχήματος. Η σύγκριση μεταξύ δύο διαφορετικών σχημάτων πραγματοποιείται με συγκεκριμένες αποστάσεις (edit distances).

Πιο προηγμένες τεχνικές, που βασίζονται στην προσέγγιση του περιγράμματος από ένα πολύγωνο, έχουν προταθεί [Meh95], με σκοπό την υλοποίηση πιο αξιόπιστων χαρακτηριστικών, που να χαρακτηρίζονται ως αναλλοίωτα στη μεταφορά, την περιστροφή και την αλλαγή κλίμακας. Γενικά, οι μεθοδολογίες αυτές καταλήγουν στην απλοποίηση του περιγράμματος, χωρίς να παρουσιάζεται θολό το αποτέλεσμα και επιτυγχάνουν να εξαλείψουν τυχόν θόρυβο, ενώ συχνά χρησιμοποιούνται σε στάδια προεπεξεργασία της εικόνας.

Το μοντέλο, που προσεγγίζει το περίγραμμα με ένα πολύγωνο, μπορεί να επεκταθεί σε πλήθος σχημάτων, όπως προτείνεται στην [Ber90]. Το αποτέλεσμα ονομάζεται ανάλυση σε λείες καμπύλες (smooth curve decomposition) και απεικονίζεται στο παρακάτω σχήμα (εικόνα 24).



Εικόνα 24

Παρατηρούμε ότι αρχικά εφαρμόζεται ένα φίλτρο μέσου όρου (average filter) στο περίγραμμα, με σκοπό να κατήξουμε σε μία λεία καμπύλη. Έπειτα, εντοπίζονται τα zero-crossing σημεία, ενώ τα τμήματα του περιγράμματος μεταξύ δύο τέτοιων διαδοχικών σημείων αποτελούν την αναπαράστασή του. Για κάθε τέτοιο τμήμα της περιμέτρου του αντικειμένου αποθηκεύονται η μέγιστη καμπυλότητά του και ο προσανατολισμός του και για την ομοιότητα δύο διαφορετικών τμημάτων χρησιμοποιείται το κριτήριο της ευκλείδειας απόστασης.

Scale-space μεθοδολογίες αναφέρθηκαν παραπάνω, εντοπίζονται όμως και στην περίπτωση structural τεχνικών, όπου το αρχικό περίγραμμα διαχωρίζεται σε επιμέρους περιοχές, οι οποίες και επεξεργάζονται περαιτέρω. Στην [Dude97] αφαιρείται από το αρχικό περίγραμμα τυχόν θόρυβος με χρήση φίλτρων μέσου όρου και στη συνέχεια πραγματοποιείται κατάτμησή του σε υπο-περιοχές. Για κάθε υπο-περιοχή υπολογίζονται οι περιγραφείς του ( $s_i^A$ ), δηλαδή το μήκος, η θέση (ordinal position) και η τιμή συντονισμού της καμπυλότητας (curvature tuning). Η σύγκριση δύο διαφορετικών σχημάτων αντιστοιχεί στη σύγκριση των αντίστοιχων χαρακτηριστικών διανυσμάτων τους  $A = (s_1^A, s_2^A, \dots, s_N^A)$  και  $B = (s_1^B, s_2^B, \dots, s_N^B)$ , που υλοποιείται με τεχνικές δυναμικού προγραμματισμού. Η χρήση των



μεθοδολογιών αυτών σε πραγματικά συστήματα ανάκλησης εικόνων είναι περιορισμένη, καθώς η επιλογή των περιγραφέων εξαρτάται και μεταβάλλεται από την εκάστοτε εφαρμογή.

Τέλος, αναφέρουμε συνοπτικά την κατηγορία των shape invariants χαρακτηριστικών, ως structural μεθοδολογίες. Οι τεχνικές αυτές προτάθηκαν, με σκοπό την αμετάβλητη συμπεριφορά των αντίστοιχων χαρακτηριστικών, όχι μόνο ως προς τη μεταφορά, την περιστροφή και την αλλαγή κλίμακας, αλλά και ως προς πλήθος άλλων μετασχηματισμών. Για παράδειγμα, οι περισσότερες μεθοδολογίες εξαρτώνται από το σημείο παρατήρησης, γεγονός το οποίο προσπαθούν να ξεπεράσουν τα shape invariants χαρακτηριστικά.

Στην κατηγορία αυτή μπορούμε να εντοπίσουμε:

- Γεωμετρικά χαρακτηριστικά (geometric invariants), όπως cross-ratio, length-ratio, λόγος απόστασης (distance ratio), γωνία, εμβαδόν, τρίγωνο ή χαρακτηριστικά από συνεπίπεδα σημεία (invariants from coplanar points) [Li98], [Huan98], [Hlav93].
- Αλγεβρικά χαρακτηριστικά (algebraic invariants), όπως ορίζουσα (determinant), ιδιοτιμές (eigenvalues), ίχνος (trace) [Squi00], [Hlav93].
- Διαφορικά χαρακτηριστικά (differential invariants), όπως καμπυλότητα (curvature), ροπή (torsion) ή γκαουσιανή καμπυλότητα (Gaussian curvature).

## **2.4 Επιλογή περιγραφέων Fourier (Fourier Descriptors)**

Διαφορετική προσέγγιση ακολουθείται στην περίπτωση των περιγραφέων Fourier (Fourier descriptors) και των περιγραφέων κυματιδίων (Wavelet descriptors). Αφετηρία αποτελεί μία συνάρτηση υπογραφής σχήματος, οπότε και το περίγραμμα απεικονίζεται πλήρως από μία μονοδιάστατη συνάρτηση  $f$ . Στη συνέχεια, οι μετρικές της κατηγορίας αυτής προκύπτουν από το φασματικό περιεχόμενο της συνάρτησης  $f$ , υπολογίζεται δηλαδή ο μετασχηματισμός της στο χώρο των συχνοτήτων.

Προσπαθώντας να αναλύσουμε περαιτέρω την περίπτωση της συνάρτησης υπογραφής των σύνθετων συντεταγμένων (complex coordinates), αναφέρουμε ότι δειγματοληπτείται το περίγραμμα του αντικειμένου, άρα προκύπτει μία ακολουθία σημείων  $(x_k, y_k)$  και συνεπώς μία νέα μιγαδική μεταβλητή  $c_k = x_k + jy_k$ . Υπολογίζοντας του συντελεστές του Διακριτού Μετασχηματισμού Fourier (DFT), καταλήγουμε στους Fourier Descriptors. Άξιο αναφοράς είναι το γεγονός, ότι οι συντελεστές αυτοί δεν είναι αμετάβλητοι στους τρεις βασικούς μετασχηματισμούς που αναφέραμε παραπάνω, υπάρχουν όμως τεχνικές ώστε να προκύψουν νέες μετρικές –με εκκίνηση τους συντελεστές Fourier- που χαρακτηρίζονται αναλλοίωτοι στη μεταφορά, στην περιστροφή και στην αλλαγή κλίμακας [Crim82], [Arbt90], [Gran72]. Σχετικά με τις τεχνικές αυτές θα αναφερθούμε εκτενέστερα σε επόμενο κεφάλαιο.



Οι περισσότερες εργασίες της βιβλιογραφίας καταλήγουν στην πολύ καλή απόδοση των περιγραφών Fourier. Ταυτόχρονα, βασίζονται στην άριστα μελετημένη και θεμελιωμένη θεωρία Fourier και είναι εύκολοι στην υλοποίηση, δίχως να απαιτούν υψηλή υπολογιστική ισχύ. Ένα ακόμη βασικό χαρακτηριστικό τους είναι η φυσική σημασία των συντελεστών, γεγονός που βοηθάει στην ερμηνεία και κατανόηση των αποτελεσμάτων. Για αυτό το λόγο και έχουν επιλεγεί –μαζί με άλλα χαρακτηριστικά- από το πρότυπο MPEG-7, για την αναπαράσταση των αντικειμένων, όπως φαίνεται και από τον παρακάτω πίνακα [Zhan07], όπου καταγράφονται όλα τα επιλεγμένα χαρακτηριστικά.

Contour-based		Region-based
Structural	Conventional	
Chain code	Perimeter	Area
Polygon	Compactness	Compactness
Gaussian smoothing	Eccentricity	Euler number
Invariants	Shape signature	Geometric moments
	Fourier descriptor	Zernike moments
	Wavelet descriptor	Pseudo-zernike moments
	Scale space	Legendre moments
	Autoregressive	Grid method
	Elastic matching	

Τα πλεονεκτήματα των περιγραφών Fourier είναι ποικίλα. Αρχικά, οι περιγραφείς Fourier είναι θεμελιωμένοι και επιστημονικώς τεκμηριωμένοι, καθώς βασίζονται και προκύπτουν από τη θεωρία του μετασχηματισμού Fourier. Επίσης, είναι απλοί στον υπολογισμό και δεν απαιτούν υψηλή υπολογιστική ισχύ, εφόσον μπορεί να γίνει χρήση του FFT (Fast Fourier Transform) [Gonz07]. Ταυτόχρονα, κάθε συντελεστής έχει μία φυσική σημασία, ενώ με μία κανονικοποίηση χαρακτηρίζονται και ως αναλλοίωτοι ως προς τη μεταφορά, την αλλαγή κλίμακας και την περιστροφή, εφόσον η ίδια συνάρτηση απόστασης από το κεντροειδές δεν μεταβάλλεται με την περιστροφή ή τη μετατόπιση του περιγράμματος. Επιπλέον, είναι ανεκτικοί ως προς την παρουσία θορύβου και επιτυγχάνουν υψηλές αποδόσεις, ως προς την ταξινόμηση του σχήματος του περιγράμματος ενός αντικειμένου.

Οι περιγραφείς Fourier αναγνωρίζονται από τη βιβλιογραφία ως οι πιο ευρέως διαδεδομένοι, εφόσον πλήθος εργασιών τους υιοθετούν [Bril68], [Gran72], [Zahn72], [Rich74], [Pers77], [Chel84], [Lin87], [Lin87a], [Krzy89], [Arbt89], [Arbt90], [Otto91], [Raub94], [Kaup95], [Meht97], [Rui97]. Οι συναρτήσεις υπογραφής σχήματος ποικίλουν, με αυτούς των σύνθετων συντεταγμένων (complex coordinates), της cumulative συνάρτησης εφαπτομένης (cumulative angle  $\phi_n$ ) και της συνάρτησης απόστασης από το κεντροειδές να είναι οι πιο συχνά χρησιμοποιούμενοι. Προβλήματα, που παρουσιάστηκαν στις πρώτες εργασίες, γρήγορα ξεπεράστηκαν, όπως για παράδειγμα η περιγραφή μη κλειστών ή επικαλυπτόμενων καμπύλων [Lin87], [Mitc84].

Εξαιτίας των παραπάνω, στην παρούσα εργασία επιλέχθηκε να παρουσιαστούν οι περιγραφείς Fourier και πιο συγκεκριμένα οι Elliptic Fourier Descriptors, όπως αυτοί αναλύονται σε επόμενο κεφάλαιο. Επιπλέον πληροφορίες για τις άλλες μεθοδολογίες είναι διαθέσιμες σε εργασίες που καταγράφηκαν στις ανωτέρω παραγράφους.

## ΚΕΦΑΛΑΙΟ 3 : Fourier Descriptors

### 3.1: Ανάπτυγμα συνάρτησης σε σειρά Fourier

Ξεκινώντας, θα θέλαμε να καλύψουμε τη βασική θεωρία γύρω από την ανάπτυξη μίας συνάρτησης του χρόνου ως σειρά Fourier.

#### 3.1.1: Εκθετική σειρά Fourier

Ένα περιοδικό, λοιπόν, σήμα  $x(t)$  γράφεται ως άθροισμα άπειρων εκθετικών όρων [Ορρε83] και πιο συγκεκριμένα

$$x(t) = \sum_{k=-\infty}^{+\infty} c_k e^{jk\omega t},$$

όπου  $\omega = \frac{2\pi}{T}$  με  $T$  την περίοδο του σήματος και

$$c_k = \frac{1}{T} \int_0^T x(t) e^{-jk\omega t} dt \text{ οι συντελεστές της εκθετικής σειράς Fourier του σήματος.}$$

Η παραπάνω αναπαράσταση είναι η ανάπτυξη ενός σήματος σε εκθετική σειρά Fourier, ή αλλιώς το ανάπτυγμα Fourier του σήματος.

Οι μιγαδικοί συντελεστές  $c_k$  κατέχουν έναν ιδιαίτερο ρόλο στον επιστημονικό κλάδο της επεξεργασίας σήματος και στην ανάλυση κατά Fourier. Οι  $c_k$  ονομάζονται συντελεστές Fourier ή φασματικές

γραμμές και ορίζουν το φάσμα του σήματος, ενώ για  $k=0$  παίρνουμε τον συντελεστή  $c_0 = \frac{1}{T} \int_0^T x(t) dt$

που είναι η συνεχής ή σταθερά συνιστώσα του φάσματος (DC συνιστώσα). Οι συντελεστές αυτοί έχουν και συγκεκριμένη φυσική σημασία, καθώς κάθε ένας από αυτούς δηλώνει το φασματικό περιεχόμενο του σήματος  $x(t)$  στη συχνότητα  $k\omega$ . Μας δείχνει δηλαδή κατά πόσο συμμετέχει η συχνότητα αυτή στον καθορισμό και την παραγωγή του σήματος  $x(t)$ . Κάθε συντελεστής  $c_k$  ονομάζεται και  $k^{\text{στή}}$  αρμονική συνιστώσα.

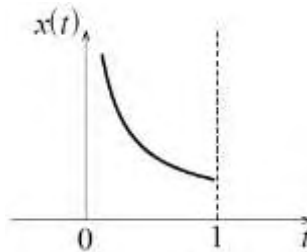
Παραπάνω αναφέραμε ότι για κάθε περιοδικό σήμα  $x(t)$ , μπορούμε να ορίσουμε την αντίστοιχη ανάπτυξή του ως σειρά Fourier. Αυτό δεν είναι αληθές, διότι αποκρύψαμε τις τρεις βασικές συνθήκες

που πρέπει να ισχύουν, για να υπάρχει η ανάπτυξη κατά σειρά Fourier ενός σήματος. Αυτές ονομάζονται συνθήκες Dirichlet και είναι οι εξής [Hsu95]:

1. Σε κάθε περίοδο θα πρέπει το σήμα  $x(t)$  να είναι απολύτως ολοκληρώσιμο, δηλαδή

$$\int_{\langle T \rangle} |x(t)| dt < +\infty$$

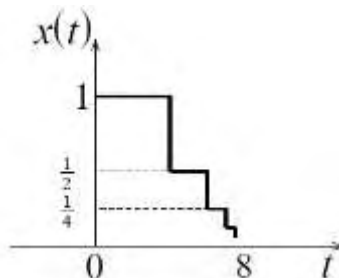
Ένα παράδειγμα ενός σήματος, για το οποίο δεν ισχύει η συνθήκη αυτή είναι το περιοδικό σήμα  $x(t) = \frac{1}{t}$ , για  $0 < t \leq 1$  με περίοδο  $T = 1$ . Η γραφική παράστασή του είναι αυτή του παρακάτω σχήματος.



2. Το σήμα  $x(t)$  θα πρέπει σε κάθε πεπερασμένο χρονικό διάστημα να είναι συνεχές ή να περιέχει πεπερασμένο αριθμό ασυνεχειών, κάθε μία από τις οποίες να είναι πεπερασμένου ύψους.

Ένα παράδειγμα περιοδικού σήματος, που περιέχει άπειρο αριθμό ασυνεχειών και η αντίστοιχη γραφική του παράσταση παρουσιάζονται παρακάτω.

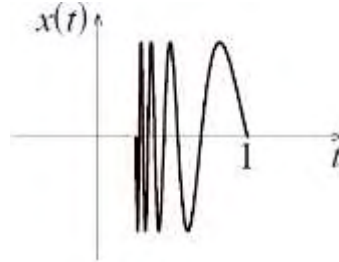
$$x(t) = \begin{cases} 1, & 0 \leq t < 4 \\ \frac{1}{2}, & 4 \leq t < 6 \\ \frac{1}{4}, & 6 \leq t < 7 \\ \vdots & \vdots \end{cases} \quad \text{με περίοδο } T=8$$



3. Το σήμα  $x(t)$  θα πρέπει να είναι φραγμένης κύμανσης σε κάθε πεπερασμένο χρονικό διάστημα. Δηλαδή θα πρέπει να υπάρχει πεπερασμένος αριθμός μεγίστων και ελαχίστων στο διάστημα.

Ως παράδειγμα περιοδικού σήματος, που δεν ισχύει η συνθήκη αυτή αναφέρουμε την παρακάτω συνάρτηση και απεικονίζουμε και τη γραφική του παράσταση.

$$x(t) = \sin\left(\frac{2\pi}{t}\right), \text{ για } 0 < t \leq 1, \text{ με περίοδο } T=1$$



Αν θελήσουμε να παρουσιάσουμε την εκθετική σειρά Fourier από την άποψη των μαθηματικών, θα υποστηρίζαμε ότι τα εκθετικά σήματα  $e^{jk\omega t}$ , με  $k = 0, \pm 1, \pm 2, \dots$  σχηματίζουν ένα ορθογώνιο σύνολο. Πραγματικά, το εσωτερικό γινόμενο τους είναι το ακόλουθο:

$$\langle e^{jk\omega t}, e^{jm\omega t} \rangle = \int_0^T e^{jk\omega t} e^{-jm\omega t} dt = T\delta(k-m) = \begin{cases} T, & k = m \\ 0, & k \neq m \end{cases}$$

Άρα, τα εκθετικά αυτά σήματα όντως σχηματίζουν ορθογώνιο σύνολο σε οποιοδήποτε πεπερασμένο χρονικό διάστημα  $[t_0, t_0 + T]$  και καλούνται και αρμονικά συσχετιζόμενα εκθετικά σήματα.

### **3.1.2: Τριγωνομετρική σειρά Fourier**

Εκτός από την εκθετική σειρά Fourier υπάρχει και η τριγωνομετρική σειρά Fourier. Γράψουμε την παραπάνω σχέση λίγο διαφορετικά, δηλαδή  $x(t) = c_0 + \sum_{k=1}^{+\infty} (c_k e^{jk\omega t} + c_{-k} e^{-jk\omega t})$

Παρατηρούμε ότι οι τιμές  $e^{jk\omega t}$  και  $e^{-jk\omega t}$  ορίζουν ζεύγος μιγαδικών συζυγών, άρα και οι αντίστοιχοι συντελεστές  $c_k$  και  $c_{-k}$  είναι μεταξύ τους συζυγείς, οπότε και γράφονται ως  $c_k = c_{k,1} - jc_{k,2}$ ,  $c_{-k} = c_{k,1} + jc_{k,2}$ .

$$\text{Άρα, } x(t) = c_0 + 2 \sum_{k=1}^{+\infty} \left( c_{k,1} \frac{e^{jk\omega t} + e^{-jk\omega t}}{2} + c_{k,2} \frac{-e^{jk\omega t} + e^{-jk\omega t}}{2} \right)$$

$$\text{Δηλαδή } x(t) = c_0 + 2 \sum_{k=1}^{+\infty} (c_{k,1} \cos(k\omega t) + c_{k,2} \sin(k\omega t)).$$

Αν τώρα ορίσουμε νέους συντελεστές  $a_k = 2c_{k,1}$  και  $b_k = 2c_{k,2}$ , τότε λαμβάνουμε την ανάπτυξη του σήματος σε τριγωνομετρική σειρά Fourier, δηλαδή

$$x(t) = c_0 + \sum_{k=1}^{+\infty} (a_k \cos(k\omega t) + b_k \sin(k\omega t)),$$

με τους νέους συντελεστές  $a_k, b_k$  να ονομάζονται και να είναι γνωστοί ως Fourier Descriptors. Η φυσική σημασία των συντελεστών αυτών είναι αρκετά μεγάλη, καθώς αντιστοιχούν στο «πόσο» συμμετέχει κάθε συχνότητα που εκπροσωπούν στην κατασκευή του τελικού σήματος, και πιο συγκεκριμένα στην κατασκευή της καμπύλης του περιγράμματος. Αναλυτικότερα, μία καμπύλη με έντονες μεταβολές (και άρα υψηλές συχνότητες) θα έχει μεγαλύτερες τιμές στις τιμές των  $a_k, b_k$  για μεγάλες τιμές των  $k$ , σε σύγκριση με μία άλλη καμπύλη, που εμφανίζει πολύ μικρότερη κυμάτωση (χαμηλές συχνότητες).

Οι συντελεστές  $a_k, b_k$  υπολογίζονται από τις παρακάτω εξισώσεις:

$$a_k = \frac{2}{T} \int_0^T x(t) \cos(k\omega t) dt, \quad b_k = \frac{2}{T} \int_0^T x(t) \sin(k\omega t) dt$$

και  $c_0$  είναι η μέση τιμή του σήματος.

Ενώ παρατηρήστε ότι συσχετίζονται με τους συντελεστές  $c_k$  της εκθετικής σειράς Fourier ως εξής:

$$c_k = \frac{a_k - jb_k}{2}, \quad c_{-k} = \frac{a_k + jb_k}{2}$$

Αν θελήσουμε να αναλύσουμε την τριγωνομετρική σειρά Fourier από τη μαθηματική της σκοπιά, μπορούμε να υποστηρίξουμε ότι τα σήματα  $\sin(k\omega t)$  και  $\cos(k\omega t)$  σε οποιοδήποτε πεπερασμένο χρονικό διάστημα  $[t_0, t_0 + T]$  καλούνται αρμονικά συσχετιζόμενα σήματα και σχηματίζουν ένα ορθογώνιο σύνολο. Το γεγονός αυτό ισχύει, εφόσον αποδεικνύονται οι παρακάτω σχέσεις.

$$\begin{aligned} \langle \sin(k\omega t), \sin(m\omega t) \rangle &= \int_0^T \sin(k\omega t) \sin(m\omega t) dt = \frac{1}{2} \int_0^T \cos((k-m)\omega t) dt - \frac{1}{2} \int_0^T \cos((k+m)\omega t) dt = \frac{T}{2} \delta(k-m) \\ \langle \cos(k\omega t), \cos(m\omega t) \rangle &= \int_0^T \cos(k\omega t) \cos(m\omega t) dt = \frac{1}{2} \int_0^T \cos((k-m)\omega t) dt + \frac{1}{2} \int_0^T \cos((k+m)\omega t) dt = \frac{T}{2} \delta(k-m) \\ \langle \sin(k\omega t), \cos(m\omega t) \rangle &= \int_0^T \sin(k\omega t) \cos(m\omega t) dt = \frac{1}{2} \int_0^T \sin((k-m)\omega t) dt + \frac{1}{2} \int_0^T \sin((k+m)\omega t) dt = 0 \end{aligned}$$

Προσπαθώντας να εκφράσουμε την τριγωνομετρική σειρά Fourier με ένα διαφορετικό τρόπο, μπορούμε να χρησιμοποιήσουμε την παρακάτω γνωστή τριγωνομετρική ταυτότητα.

$$b \cos \varphi + c \sin \varphi = A \cos(\varphi + \theta),$$

$$\text{Όπου } A = \sqrt{b^2 + c^2} \text{ και } \theta = -\tan^{-1} \frac{c}{b}.$$

Τότε

$$\begin{aligned} x(t) &= c_0 + \sum_{k=1}^{\infty} b_k \cos(k\omega t) + \sum_{k=1}^{\infty} c_k \sin(k\omega t) = \\ &= c_0 + b_1 \cos(\omega t) + c_1 \sin(\omega t) + b_2 \cos(2\omega t) + c_2 \sin(2\omega t) + \dots = \\ &= c_0 + A_1 \cos(\omega t + \theta_1) + A_2 \cos(\omega t + \theta_2) + \dots \end{aligned}$$

$$\text{Όπου } A_1 = \sqrt{b_1^2 + c_1^2}, \theta_1 = -\tan^{-1} \frac{c_1}{b_1}$$

$$A_2 = \sqrt{b_2^2 + c_2^2}, \theta_2 = -\tan^{-1} \frac{c_2}{b_2}, \text{ κτλ.}$$

Άρα, γενικά

$$x(t) = A_0 + \sum_{k=1}^{\infty} A_k \cos(k\omega t + \theta_k), \text{ με}$$

$$A_0 = c_0, A_k = \sqrt{b_k^2 + c_k^2}, \theta_k = -\tan^{-1} \frac{c_k}{b_k}$$

### **3.2 Elliptic Fourier Descriptors (EFD)**

#### **3.2.1 Ορισμός των EFD**

Στη θεματική αυτή ενότητα θα αναλύσουμε τη θεωρία των Elliptic Fourier Descriptors και θα εξηγήσουμε πως μπορούν αυτοί να βοηθήσουν στην αναπαράσταση μίας καμπύλης με επιμέρους τιμές.

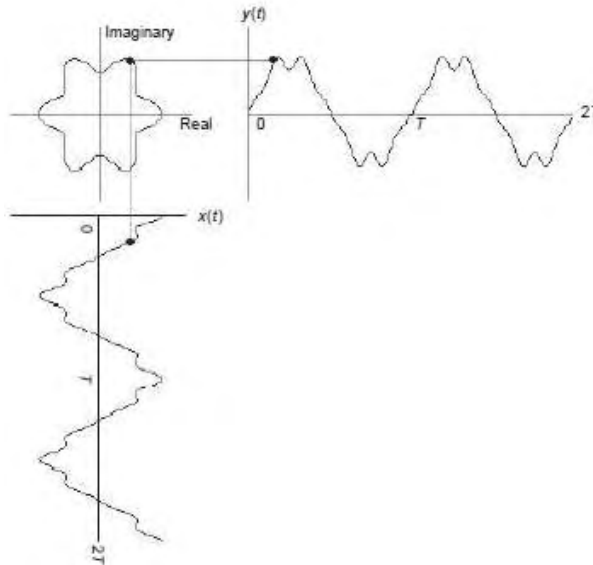
Στην αρχή αναφέρουμε ότι έχουμε στη διάθεσή μας μία καμπύλη από το περίγραμμα ενός αντικειμένου στην εικόνα. Έχουμε δηλαδή μία σειρά από pixel της εικόνας, που αντιστοιχούν στο περίγραμμα. Θεωρούμε ως  $x(t)$  τον οριζόντιο άξονα (γραμμές της εικόνας) και  $y(t)$  τον κάθετο άξονα

(στήλες της εικόνας). Αποθηκεύουμε δηλαδή τα pixels του περιγράμματος σε δύο συναρτήσεις, σε κάθε μία από τις οποίες δίνονται αντίστοιχα οι  $x$  και  $y$  συντεταγμένες.

Με τον τρόπο αυτό μπορούμε να ορίσουμε μία νέα μιγαδική συνάρτηση, την

$$c(t) = x(t) + jy(t),$$

η οποία μας περιγράφει πλήρως την καμπύλη του περιγράμματος. Στο παρακάτω σχήμα φαίνεται μία τέτοια απεικόνιση.



Αναλύοντας την  $c(t)$  ως εκθετική σειρά Fourier καταλήγουμε στην εξίσωση

$$c_k = c_{xk} + jc_{yk}$$

Όπου  $c_{xk}$  και  $c_{yk}$  οι συντελεστές της εκθετικής σειράς Fourier των  $x(t)$  και  $y(t)$  αντίστοιχα, δηλαδή

$$c_{xk} = \frac{1}{T} \int_0^T x(t) e^{-jk\omega t} dt \text{ και } c_{yk} = \frac{1}{T} \int_0^T y(t) e^{-jk\omega t} dt.$$

Σχετίζοντας τους παραπάνω συτελεστές της εκθετικής σειράς Fourier με τους αντίστοιχους της τριγωνομετρικής σειράς Fourier, προκύπτουν τα ακόλουθα:

$$c_{xk} = \frac{a_{xk} - jb_{xk}}{2}, \quad c_{yk} = \frac{a_{yk} - jb_{yk}}{2}$$

$$c_{x-k} = \frac{a_{xk} + jb_{xk}}{2}, \quad c_{y-k} = \frac{a_{yk} + jb_{yk}}{2}$$



Ενώ οι συντελεστές της τριγωνομετρικής σειράς βρίσκονται ως:

$$a_{xk} = \frac{2}{T} \int_0^T x(t) \cos(k\omega t) dt \quad \text{και} \quad b_{xk} = \frac{2}{T} \int_0^T x(t) \sin(k\omega t) dt$$

$$a_{yk} = \frac{2}{T} \int_0^T y(t) \cos(k\omega t) dt \quad \text{και} \quad b_{yk} = \frac{2}{T} \int_0^T y(t) \sin(k\omega t) dt$$

Από τα παραπάνω συμπεραίνουμε ότι οι συντελεστές  $c_k$  της εκθετικής σειράς Fourier του σήματος  $c(t)$  γράφονται ως:

$$c_k = c_{xk} + jc_{yk} = \frac{a_{xk} - jb_{xk}}{2} + j \frac{a_{yk} - jb_{yk}}{2} = \frac{a_{xk} + ja_{yk}}{2} - j \frac{b_{xk} + jb_{yk}}{2}$$

$$c_{-k} = c_{x-k} + jc_{y-k} = \frac{a_{xk} + jb_{xk}}{2} + j \frac{a_{yk} + jb_{yk}}{2} = \frac{a_{xk} + ja_{yk}}{2} + j \frac{b_{xk} + jb_{yk}}{2}$$

Δηλαδή,

$$c_k = A_k - jB_k \quad \text{και} \quad c_{-k} = A_k + jB_k$$

$$\text{Όπου} \quad A_k = \frac{a_{xk} + ja_{yk}}{2} \quad \text{και} \quad B_k = \frac{b_{xk} + jb_{yk}}{2}$$

$$\text{Επίσης,} \quad c(t) = c_0 + \sum_{k=1}^{+\infty} (A_k - jB_k) e^{jk\omega t} + \sum_{k=-\infty}^{-1} (A_k + jB_k) e^{jk\omega t}$$

Ή αλλιώς σε τριγωνομετρική μορφή

$$c(t) = \frac{a_{x0}}{2} + \sum_{k=1}^{+\infty} (a_{xk} \cos(k\omega t) + b_{xk} \sin(k\omega t)) + j \left( \frac{a_{y0}}{2} + \sum_{k=1}^{+\infty} (a_{yk} \cos(k\omega t) + b_{yk} \sin(k\omega t)) \right)$$

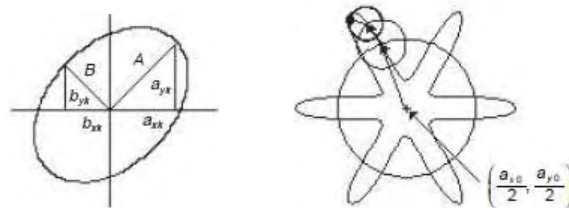
$$\text{Ή ισοδύναμα με μορφή πινάκων} \quad \begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = \frac{1}{2} \begin{bmatrix} a_{x0} \\ a_{y0} \end{bmatrix} + \sum_{k=1}^{\infty} \begin{bmatrix} a_{xk} & b_{xk} \\ a_{yk} & b_{yk} \end{bmatrix} \begin{bmatrix} \cos(k\omega t) \\ \sin(k\omega t) \end{bmatrix}.$$

Η φυσική σημασία των παραπάνω συντελεστών αναλύεται παρακάτω. Κάθε όρος του αθροίσματος αντιστοιχεί σε μία έλλειψη στο μιγαδικό επίπεδο. Οι άξονες της έλλειψης (major axes) καθορίζονται

από τις τιμές των συντελεστών  $A_k$  και  $B_k$ , ή ομοίως από τις τιμές των συντελεστών  $a_{xk}$ ,  $a_{yk}$ ,  $b_{xk}$  και  $b_{yk}$ , όπως φαίνεται από το παρακάτω σχήμα (α). Στο σχήμα Β παρατηρούμε τη διαδικασία εύρεσης σημείου της περιμέτρου με χρήση του παραπάνω αθροίσματος, για άθροισμα όμως τριών όρων.

Αρχικά ξεκινάμε από το σημείο  $\left(\frac{a_{x0}}{2}, \frac{a_{y0}}{2}\right)$  και ορίζουμε μία έλλειψη. Όσο αλλάζει η τιμή του  $t$

μετακινούμαστε πάνω στην έλλειψη αυτή (στην ουσία ο κεντρικός κύκλος). Όταν σταθεροποιήσουμε το  $t$  και πάρουμε και τον επόμενο όρο του αθροίσματος ξεκινάμε νέα έλλειψη με κέντρο το σημείο πάνω στον πρώτο κύκλο όπου έχουμε σταματήσει και η διαδικασία επαναλαμβάνεται. Προφανώς όσο περισσότερους συντελεστές χρησιμοποιήσουμε (δηλαδή περισσότερους όρους του αθροίσματος), οι καμπύλες που μπορούμε να περιγράψουμε με μεγάλη ακρίβεια πληθαίνουν. Στο συγκεκριμένο σχήμα η καμπύλη είναι αρκετά συμμετρική και εύκολη να περιγραφεί, γεγονός που είναι σπάνιο σε πραγματικές εφαρμογές.



### 3.2.2 Αμετάβλητη συμπεριφορά των EFD σε γεωμετρικούς μετασχηματισμούς

Με εκκίνηση την περιγραφή ενός σχήματος με χρήση των EFD μπορούμε να ορίσουμε νέους συντελεστές, ώστε αυτοί να είναι αμετάβλητη σε συγκεκριμένους γεωμετρικούς μετασχηματισμούς. Οι γεωμετρικοί μετασχηματισμοί που εξετάζουμε είναι μετατόπιση (translation), αλλαγή κλίμακας (scaling) και περιστροφή (rotation).

Ξεκινούμε με το μετασχηματισμό της μετατόπισης. Αν θεωρήσουμε δηλαδή ότι μετακινούμε το σχήμα πάνω στην εικόνα, θα ελέγξουμε την αλλαγή που θα έχουμε στις τιμές των συντελεστών.

Έχουμε ως αρχικό σημείο αναφοράς την εξίσωση που καταλήξαμε στην προηγούμενη παράγραφο.

$$\begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = \frac{1}{2} \begin{bmatrix} a_{x0} \\ a_{y0} \end{bmatrix} + \sum_{k=1}^{\infty} \begin{bmatrix} a_{xk} & b_{xk} \\ a_{yk} & b_{yk} \end{bmatrix} \begin{bmatrix} \cos(k\omega t) \\ \sin(k\omega t) \end{bmatrix}$$

Αν θεωρήσουμε ένα καινούργιο περίγραμμα ενός σχήματος

$$c'(t) = x'(t) + jy'(t)$$

Τότε η αντίστοιχη εξίσωση θα είναι η ακόλουθη.

$$\begin{bmatrix} x'(t) \\ y'(t) \end{bmatrix} = \frac{1}{2} \begin{bmatrix} a'_{x0} \\ a'_{y0} \end{bmatrix} + \sum_{k=1}^{\infty} \begin{bmatrix} a'_{xk} & b'_{xk} \\ a'_{yk} & b'_{yk} \end{bmatrix} \begin{bmatrix} \cos(k\omega t) \\ \sin(k\omega t) \end{bmatrix}$$

Θεωρούμε ότι η καμπύλη  $c'(t)$  είναι η αρχική καμπύλη  $c(t)$ , αλλά μετατοπισμένη κατά  $t_x$  και  $t_y$  στον πραγματικό και φανταστικό άξονα αντίστοιχα. Οπότε εξετάζουμε την αλλαγή στις τιμές των συντελεστών σε περίπτωση μετατόπισης της καμπύλης.

Τότε, η παραπάνω εξίσωση γράφεται και ως

$$\begin{bmatrix} x'(t) \\ y'(t) \end{bmatrix} = \frac{1}{2} \begin{bmatrix} a_{x0} \\ a_{y0} \end{bmatrix} + \sum_{k=1}^{\infty} \begin{bmatrix} a_{xk} & b_{xk} \\ a_{yk} & b_{yk} \end{bmatrix} \begin{bmatrix} \cos(k\omega t) \\ \sin(k\omega t) \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix}$$

Δηλαδή η μετατοπισμένη καμπύλη  $c'(t)$  αντιστοιχεί την εξίσωση

$$\begin{bmatrix} x'(t) \\ y'(t) \end{bmatrix} = \frac{1}{2} \begin{bmatrix} a_{x0} + 2t_x \\ a_{y0} + 2t_y \end{bmatrix} + \sum_{k=1}^{\infty} \begin{bmatrix} a_{xk} & b_{xk} \\ a_{yk} & b_{yk} \end{bmatrix} \begin{bmatrix} \cos(k\omega t) \\ \sin(k\omega t) \end{bmatrix}$$

Ενώ η αρχική  $c(t)$  στην εξίσωση

$$\begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = \frac{1}{2} \begin{bmatrix} a_{x0} \\ a_{y0} \end{bmatrix} + \sum_{k=1}^{\infty} \begin{bmatrix} a_{xk} & b_{xk} \\ a_{yk} & b_{yk} \end{bmatrix} \begin{bmatrix} \cos(k\omega t) \\ \sin(k\omega t) \end{bmatrix}$$

Συγκρίνοντας τις παραπάνω εξισώσεις μπορούμε να κατανοήσουμε την αλλαγή που επέρχεται στις τιμές των συντελεστών EFD, στην περίπτωση που μετακινήσουμε το περίγραμμα πάνω στην εικόνα. Από αυτές τις εξισώσεις συμπεραίνουμε ότι:

$$a'_{xk} = a_{xk}, b'_{xk} = b_{xk}, a'_{yk} = a_{yk}, b'_{yk} = b_{yk} \text{ για } k \neq 0$$

$$a'_{x0} = a_{x0} + 2t_x, a'_{y0} = a_{y0} + 2t_y$$

Άρα, όλοι οι συντελεστές παραμένουν σταθεροί, εκτός από τους  $a_{x0}$  και  $a_{y0}$ . Η παρατήρηση αυτή συμβαδίζει και με τη φυσική τους σημασία, εφόσον αυτοί αντιστοιχούν στο κέντρο βάρους του σχήματος, το οποίο είναι το μοναδικό που μετακινείται, εφόσον μετακινηθεί όλο το περίγραμμα σε άλλο σημείο της εικόνας.

Επόμενος γεωμετρικός μετασχηματισμός είναι ο μετασχηματισμός αλλαγής κλίμακας. Στην περίπτωση αυτή θεωρούμε ότι όλο το σχήμα αλλάζει κλίμακα (δηλαδή μεγαλώνει ή μικραίνει), αλλά συμμετρικά ως προς το κέντρο βάρους του.

Τότε, η αλλαγή που θα πραγματοποιηθεί μπορεί να μοντελοποιηθεί από μία σταθερά  $s$ , οπότε η νέα καμπύλη που θα προκύψει θα περιγράφεται από την παρακάτω εξίσωση.

$$\begin{bmatrix} x'(t) \\ y'(t) \end{bmatrix} = \frac{1}{2} \begin{bmatrix} a_{x0} \\ a_{y0} \end{bmatrix} + s \sum_{k=1}^{\infty} \begin{bmatrix} a_{xk} & b_{xk} \\ a_{yk} & b_{yk} \end{bmatrix} \begin{bmatrix} \cos(k\omega t) \\ \sin(k\omega t) \end{bmatrix}$$

Αν συγκρίνουμε πάλι με την αρχική εξίσωση

$$\begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = \frac{1}{2} \begin{bmatrix} a_{x0} \\ a_{y0} \end{bmatrix} + \sum_{k=1}^{\infty} \begin{bmatrix} a_{xk} & b_{xk} \\ a_{yk} & b_{yk} \end{bmatrix} \begin{bmatrix} \cos(k\omega t) \\ \sin(k\omega t) \end{bmatrix},$$

συμπεραίνουμε ότι οι τιμές των νέων συντελεστών σχετίζονται με τους αρχικούς ως εξής:

$$a'_{xk} = sa_{xk}, \quad b'_{xk} = sb_{xk}, \quad a'_{yk} = sa_{yk}, \quad b'_{yk} = sb_{yk} \quad \text{για } k \neq 0$$

$$a'_{x0} = a_{x0}, \quad a'_{y0} = a_{y0}$$

Δηλαδή, όλοι οι συντελεστές πολλαπλασιάζονται με τη σταθερά  $s$ , εκτός από τους  $a_{x0}$  και  $a_{y0}$  που παραμένουν αμετάβλητοι.

Τέλος, ασχολούμαστε με το γεωμετρικό μετασχηματισμό της περιστροφής του σχήματος κατά συγκεκριμένη γωνία  $\rho$ .

Τότε, η νέα καμπύλη  $c'(t)$  θα δίνεται από την παρακάτω εξίσωση.

$$\begin{bmatrix} x'(t) \\ y'(t) \end{bmatrix} = \frac{1}{2} \begin{bmatrix} a_{x0} \\ a_{y0} \end{bmatrix} + \begin{bmatrix} \cos(\rho) & \sin(\rho) \\ -\sin(\rho) & \cos(\rho) \end{bmatrix} \sum_{k=1}^{\infty} \begin{bmatrix} a_{xk} & b_{xk} \\ a_{yk} & b_{yk} \end{bmatrix} \begin{bmatrix} \cos(k\omega t) \\ \sin(k\omega t) \end{bmatrix}$$

Ομοίως με ανωτέρω, παρατηρούμε την εξίσωση αυτή σε σχέση με την αρχική

$$\begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = \frac{1}{2} \begin{bmatrix} a_{x0} \\ a_{y0} \end{bmatrix} + \sum_{k=1}^{\infty} \begin{bmatrix} a_{xk} & b_{xk} \\ a_{yk} & b_{yk} \end{bmatrix} \begin{bmatrix} \cos(k\omega t) \\ \sin(k\omega t) \end{bmatrix}$$

Και καταλήγουμε ότι οι νέοι συντελεστές θα σχετίζονται με τους αρχικούς σύμφωνα με τις παρακάτω σχέσεις.

$$a'_{xk} = a_{xk} \cos(\rho) + a_{yk} \sin(\rho), \quad b'_{xk} = b_{xk} \cos(\rho) + b_{yk} \sin(\rho),$$

$$a'_{yk} = -a_{xk} \sin(\rho) + a_{yk} \cos(\rho), \quad b'_{yk} = -b_{xk} \sin(\rho) + b_{yk} \cos(\rho) \text{ για } k \neq 0$$

$$a'_{x0} = a_{x0}, \quad a'_{y0} = a_{y0}$$

Δηλαδή, οι νέοι συντελεστές αλλάζουν και δίνονται από ένα γραμμικό συνδυασμό των παλαιών. Ο γραμμικός συνδυασμός εξαρτάται από την τιμή της γωνίας περιστροφής  $\rho$ .

Άρα, αποδείξαμε τις αλλαγές που πραγματοποιούνται στις τιμές των συντελεστών EFD, σε περίπτωση μετατόπισης, αλλαγής κλίμακας ή περιστροφής της αρχικής καμπύλης. Αν τώρα συνδυάσουμε όλους τους παραπάνω γεωμετρικούς μετασχηματισμούς, καταλήγουμε ότι

$$a'_{xk} = s(a_{xk} \cos(\rho) + a_{yk} \sin(\rho)), \quad b'_{xk} = s(b_{xk} \cos(\rho) + b_{yk} \sin(\rho))$$

$$a'_{yk} = s(-a_{xk} \sin(\rho) + a_{yk} \cos(\rho)), \quad b'_{yk} = s(-b_{xk} \sin(\rho) + b_{yk} \cos(\rho)) \text{ για } k \neq 0$$

$$a'_{x0} = a_{x0} + 2t_x, \quad a'_{y0} = a_{y0} + 2t_y$$

Με αφετηρία τις εξισώσεις αυτές προσπαθούμε να ορίσουμε άλλους συντελεστές, οι οποίοι να μην μεταβάλλονται όταν εφαρμόζουμε τους γεωμετρικούς αυτούς μετασχηματισμούς. Ως πρώτο βήμα, δεν χρησιμοποιούμε τους συντελεστές  $a_{x0}$  και  $a_{y0}$  που αντιστοιχούν στο κέντρο βάρους του σχήματος. Στη συνέχεια, ορίζουμε ως νέους αμετάβλητους συντελεστές τους ακόλουθους [Nixο08], ενώ μερικές εργασίες της βιβλιογραφίας ακολουθούν διαφορετικές προσεγγίσεις [Gran72].

$$\frac{|A_k|}{|A_1|} + \frac{|B_k|}{|B_1|}$$

Ενώ υπενθυμίζουμε ότι σε προηγούμενη παράγραφο είχαμε ορίσει

$$A_k = \frac{a_{xk} + ja_{yk}}{2} \text{ και } B_k = \frac{b_{xk} + jb_{yk}}{2}$$

Άρα, συμπεραίνουμε ότι

$$\frac{|A'_k|}{|A'_1|} = \frac{\sqrt{a_{xk}^2 + a_{yk}^2}}{\sqrt{a_{x1}^2 + a_{y1}^2}}$$

και αντίστοιχα

$$\frac{|B'_k|}{|B'_1|} = \frac{\sqrt{b_{xk}^2 + b_{yk}^2}}{\sqrt{b_{x1}^2 + b_{y1}^2}}$$

Παρατηρούμε δηλαδή ότι οι παραπάνω εξισώσεις δεν περιέχουν τις παραμέτρους  $(t_x, t_y)$ ,  $s$  και  $\rho$  των γεωμετρικών μετασχηματισμών (μετακίνησης, αλλαγής κλίμακας και περιστροφής), παραμένουν δηλαδή αναλλοίωτοι σε αυτούς.

Όλη η παραπάνω ανάλυση και τα αποτελέσματα, στα οποία καταλήξαμε, απεικονίζονται στον παρακάτω πίνακα.

<b>Μετατόπιση κατά <math>(t_x, t_y)</math></b>	$a'_{xk} = a_{xk}, b'_{xk} = b_{xk}, a'_{yk} = a_{yk}, b'_{yk} = b_{yk}$ για $k \neq 0$ $a'_{x0} = a_{x0} + 2t_x, a'_{y0} = a_{y0} + 2t_y$
<b>Αλλαγή κλίμακας κατά <math>s</math></b>	$a'_{xk} = sa_{xk}, b'_{xk} = sb_{xk}, a'_{yk} = sa_{yk}, b'_{yk} = sb_{yk}$ για $k \neq 0$ $a'_{x0} = a_{x0}, a'_{y0} = a_{y0}$
<b>Περιστροφή κατά γωνία <math>\rho</math></b>	$a'_{xk} = a_{xk} \cos(\rho) + a_{yk} \sin(\rho), b'_{xk} = b_{xk} \cos(\rho) + b_{yk} \sin(\rho),$ $a'_{yk} = -a_{xk} \sin(\rho) + a_{yk} \cos(\rho), b'_{yk} = -b_{xk} \sin(\rho) + b_{yk} \cos(\rho)$ για $k \neq 0$ $a'_{x0} = a_{x0}, a'_{y0} = a_{y0}$
<b>Μετατόπιση, αλλαγή κλίμακας και περιστροφή</b>	$a'_{xk} = s(a_{xk} \cos(\rho) + a_{yk} \sin(\rho)), b'_{xk} = s(b_{xk} \cos(\rho) + b_{yk} \sin(\rho))$ $a'_{yk} = s(-a_{xk} \sin(\rho) + a_{yk} \cos(\rho)), b'_{yk} = s(-b_{xk} \sin(\rho) + b_{yk} \cos(\rho))$ για $k \neq 0$ $a'_{x0} = a_{x0} + 2t_x, a'_{y0} = a_{y0} + 2t_y$
<b>Αναλλοίωτοι συντελεστές</b>	$\frac{ A_k }{ A_1 } + \frac{ B_k }{ B_1 },$ εφόσον $\frac{ A'_k }{ A'_1 } = \frac{\sqrt{a_{xk}^2 + a_{yk}^2}}{\sqrt{a_{x1}^2 + a_{y1}^2}}$ και $\frac{ B'_k }{ B'_1 } = \frac{\sqrt{b_{xk}^2 + b_{yk}^2}}{\sqrt{b_{x1}^2 + b_{y1}^2}}$

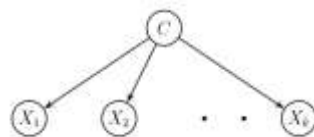
### 3.3: Ταξινόμηση

Για την κατηγοριοποίηση ή αλλιώς ταξινόμηση των διανυσμάτων, που προέκυψαν, από το στάδιο της εξαγωγής των τιμών των EFD συντελεστών, χρησιμοποιήθηκε σε πρώτη φάση ο ταξινομητής  $k$  πλησιέστερων γειτόνων ( $knn$ ) με ευκλείδεια απόσταση.

Ο συγκεκριμένος ταξινομητής λειτουργεί ως εξής: το άγνωστο πρότυπο αντιστοιχίζεται στην κλάση εκείνη, στην οποία ανήκει η πλειοψηφία από τους  $k$  πιο κοντινούς του γείτονες. Στην πράξη, δηλαδή, υπολογίζουμε τους  $k$  πλησιέστερους γείτονες του αγνώστου προτύπου και από αυτούς βρίσκουμε σε ποια κλάση ανήκουν οι περισσότεροι. Η απόσταση δύο προτύπων  $x, y$  σε διάσταση χώρου  $N$ , δίνεται από την κλασσική ευκλείδεια απόσταση, με βάση την ακόλουθη εξίσωση:

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_N - y_N)^2}$$

Ο συγκεκριμένος ταξινομητής επιλύει το πρόβλημα της ταξινόμησης σε τοπική κλίμακα, καθώς για κάθε άγνωστο πρότυπο εξετάζει την κλάση των γειτονικών του, ενώ ταυτόχρονα είναι πολύ απλός, δίνοντας τη δυνατότητα της εξαντλητικής αναζήτησης για την εύρεση της βέλτιστης λύσης. Στη συνέχεια ελέγχθηκε η απόδοση απλών Bayesian ταξινομητών. Πρόκειται για στατιστικούς ταξινομητές που μπορούν να υπολογίσουν τη πιθανότητα ένα δοθέν στιγμιότυπο κάποιου προβλήματος να ανήκει σε μια από τις προκαθορισμένες κλάσεις του προβλήματος [ERNT03]. Θεωρητικά οι Bayesian ταξινομητές έχουν τον μικρότερο ρυθμό λάθους συγκρινόμενοι με τους υπόλυτους ταξινομητές. Στην πράξη όμως δεν ισχύει πάντα, λόγω των σφαλμάτων που γίνεται στις υποθέσεις όπως για παράδειγμα στην υπόθεση για την ανεξαρτησία ως προς την κατανομή των κλάσεων. Βασικό πλεονέκτημά τους είναι ότι προσφέρουν μια θεωρητική αιτιολόγηση για άλλους ταξινομητές λόγω του θεωρήματος Bayes ενώ παρουσιάζουν υψηλή απόδοση σε ακρίβεια και ταχύτητα όταν εφαρμόζονται σε μεγάλες βάσεις δεδομένων. Ένας πολύ απλός Bayesian ταξινομητής είναι ο Naïve – Bayes, η απόδοση του οποίου συγκρίνεται με την απόδοση ταξινομητών όπως στα δέντρα απόφασης και τα νευρωνικά δίκτυα. Ένα παράδειγμα ενός Bayes δικτύου θα μπορούσε να απεικονίζει τις πιθανοτικές συσχετίσεις μεταξύ ασθενειών και συμπτωμάτων. Εφόσον το δίκτυο εκπαιδευτεί με τα κατάλληλα δεδομένα, τροφοδοτείται με τα συμπτώματα και παράγει ως έξοδο τις πιθανότητες ύπαρξης των διάφορων ασθενειών. Στην περίπτωση του Naïve Bayes ταξινομητή, το δίκτυο απλοποιείται περαιτέρω, χρησιμοποιώντας όμως κάποιες υποθέσεις, ως προς τη συσχέτιση των μεταβλητών που αυτό απεικονίζει. Πιο συγκεκριμένα, θεωρείται ότι οι προβλεπόμενες τιμές  $X_1, X_2, \dots, X_k$  είναι πλήρως ανεξάρτητες μεταξύ τους, παίρνοντας ως δεδομένη την τιμή της κλάσης  $C$ , όπως αυτό αναπαρίσταται στο παρακάτω σχήμα (εικόνα 27)[Geor95], [Bunt94], [Duda73].



Εικόνα 25

## ΚΕΦΑΛΑΙΟ 4: ΠΕΙΡΑΜΑΤΙΚΟ ΜΕΡΟΣ

### 4.1 Επιλογή βάσεων δεδομένων

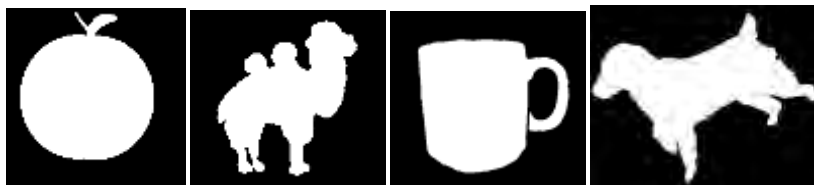
Για την επικύρωση των παραπάνω θεωρητικών μεθοδολογιών, που παρουσιάστηκαν και αναλύθηκαν, χρησιμοποιήθηκαν συγκεκριμένες βάσεις εικόνων, οι οποίες διατίθενται ελεύθερα από την επιστημονική και ακαδημαϊκή κοινότητα, με σκοπό την αξιολόγηση συγκεκριμένων αλγορίθμων και τεχνικών. Στα πλαίσια της παρούσης εργασίας επιλέχθηκαν μερικές από τις πιο γνωστές και ευρέως διαδεδομένες βάσεις εικόνων. Οι βάσεις αυτές, όπως επίσης και τα τεχνικά τους χαρακτηριστικά, παρουσιάζονται στο εδάφιο αυτό.

#### 4.1.1 Βάση εικόνων MPEG7 CE Shape-1 Part B

Η συγκεκριμένη βάση εικόνων εντοπίστηκε μετά από αναζήτηση στο Διαδίκτυο. Πιο αναλυτικά βρέθηκε η περιγραφή και η αναφορά της σε ιστοσελίδα σχετική με την επεξεργασία εικόνας [Imag]. Ο ιστοχώρος αυτός διαθέτει μία λίστα με βάσεις εικόνων, οι οποίες είναι δημόσια διαθέσιμες στην ακαδημαϊκή κοινότητα για ερευνητικούς σκοπούς.

Η βάση που επιλέξαμε δημιουργήθηκε ως αρωγή για την εξέλιξη του προτύπου MPEG-7 [Mpeg]. Το πρότυπο αυτό αναπτύχθηκε από την σχετική ομάδα MPEG (Moving Picture Experts Group) και στοχεύει στην περιγραφή πολυμεσικού περιεχομένου πληροφορίας. Την πληροφορία αυτή μπορεί να ερμηνεύσει, να προωθήσει και να χρησιμοποιήσει οποιαδήποτε υπολογιστική συσκευή (ηλεκτρονικός υπολογιστής, τηλεόραση, κινητό τηλέφωνο κτλ).

Με σκοπό, λοιπόν, την περιγραφή των αντικειμένων μίας εικόνας, δίνονται, από την ειδικευμένη ομάδα που αναφέραμε παραπάνω, πλήθος δυαδικών εικόνων, κάθε μία από τις οποίες περιγράφει ένα αντικείμενο. Παραδείγματα από εικόνες της βάσης αυτής βλέπουμε και στο παρακάτω σχήμα (εικόνα 26), όπου αναπαριστώνται ένα μήλο, μία καμήλα, μία κούπα και ένας σκύλος.



Εικόνα 26



Η βάση αυτή αποτελείται από συνολικά 1400 δυαδικές εικόνες, κάθε μία από τις οποίες ανήκει σε μία ξεχωριστή κλάση. Σε κάθε μία κλάση ανήκουν 20 διαφορετικές εικόνες, ενώ ο αριθμός των κλάσεων της βάσης είναι 70. Στον παρακάτω πίνακα αναφέρουμε επιγραμματικά τις κλάσεις που υπάρχουν στη βάση αυτή.

Κλάσεις της βάσης εικόνων MPEG-7						
Apple	Car	Cup	Device8	Frog	Key	Sea snake
Bat	Carriage	Deer	Device9	Glas	Lizzard	Shoe
Beetle	Cattle	Device0	Dog	Guitar	Lmfish	Spoon
Bell	Cellular phone	Device1	Elephant	Hammer	Misk	Spring
Bird	Chicken	Device2	Face	Hat	Octopus	Stef
Bone	Children	Device3	Fish	HCircle	Pencil	Teddy
Bottle	Chopper	Device4	Flatfish	Heart	Personal car	Tree
Brick	Classic	Device5	Fly	Horse	Pocket	Truck
Butterfly	Comma	Device6	Fork	Horseshoe	Rat	Turtle
Camel	Crown	Device7	Fountain	Jar	Ray	Watch

Η βάση αυτή είναι ευρέως διαδεδομένη [Late00], ενώ το πιο σημαντικό χαρακτηριστικό της είναι το μεγάλο πλήθος και η ποικιλομορφία των κλάσεων, στις οποίες είναι ομαδοποιημένες οι εικόνες που δίδονται. Επίσης, η βάση είναι διαθέσιμες σε δυαδικές εικόνες (bitmap), εφόσον η διαδικασία της εύρεσης του αντικειμένου έχει ήδη ολοκληρωθεί. Το αντικείμενο (object) έχει άσπρο φόντο, ενώ το υπόβαθρο (background) έχει μαύρο. Τέλος, θα θέλαμε να αναφέρουμε ότι ο τύπος αρχείου, που διατίθενται οι εικόνες, είναι .gif.

#### **4.1.2 Βάση εικόνων SSID-1**

Η συγκεκριμένη βάση εικόνων ονομάστηκε SSID από τα αρχικά Shape Indexing of Image Databases [Ssid]. Η βάση αυτή κατασκευάστηκε στα πλαίσια ενός προγράμματος, σκοπός του οποίου ήταν η υλοποίηση ενός συστήματος ανάκλησης εικόνων (image retrieval system), που να χρησιμοποιεί ως πρωτεύον κριτήριο το σχήμα των αντικειμένων της εικόνας [Seba01].

Οι εικόνες της βάσης περιέχουν συγκεκριμένα αντικείμενα, όπως φαίνεται και από το παρακάτω σχήμα (εικόνα 27), όπου απεικονίζονται μερικές εικόνες. Το πλήθος των εικόνων είναι 99, ενώ χαρακτηριστικό γνώρισμα της βάσης αυτής είναι το εξής: κάποιες από τις εικόνες της είναι επεξεργασμένες με σημείο εκκίνησης κάποια άλλη εικόνα της βάσης. Στο παρακάτω σχήμα, δηλαδή, παρατηρήστε τις δύο τελευταίες εικόνες. Η δεύτερη είναι όμοια με την πρώτη, με εξαίρεση το πάνω μέρος της, το οποίο και έχει αποκοπεί. Με τον τρόπο αυτό γίνεται μία προσπάθεια εκτίμησης των σφαλμάτων, που εισάγουν στη διαδικασία ανάκλησης των εικόνων πιθανές τροποποιήσεις.



Εικόνα 27

Στο σημείο αυτό θα θέλαμε να τονίσουμε ότι η κατηγοριοποίηση των εικόνων σε επιμέρους κλάσεις δεν είναι προφανής, με την έννοια ότι η ταξινόμηση δεν περιγράφεται από την ονομασία των εικόνων της βάσης, όπως γίνεται για παράδειγμα στη βάση MPEG7. Επειδή, όμως, θέλαμε η ταξινόμηση να είναι απολύτως ευδιάκριτη και σαφής για τη διεξαγωγή των πειραμάτων, κατατάξαμε τις εικόνες της βάσης αυτής σε διακριτές κλάσεις, ανάλογα με την κοινή λογική. Η ταξινόμηση αυτή πραγματοποιήθηκε σε 9 κλάσεις με ίδιο αριθμό εικόνων σε κάθε κλάση (11) , ενώ ο τρόπος με τον οποίο έλαβε χώρα η κατηγοριοποίηση αυτή, παραδίδεται και γνωστοποιείται μαζί με τα βοηθητικά αρχεία της εργασίας.

Στις εικόνες της βάσης αυτής το αντικείμενο απεικονίζεται με μαύρα pixel, ενώ το υπόβαθρο με λευκά, αντίθετα δηλαδή με την MPEG-7 βάση εικόνων. Οι εικόνες της SSID-1 δίνονται στο format .pgm (portable graymap). Αυτός ο τύπος αρχείων διαβάζεται εύκολα από το image processing toolbox του matlab, οπότε και δεν μας δημιουργεί κανένα πρόβλημα στην επεξεργασία του.

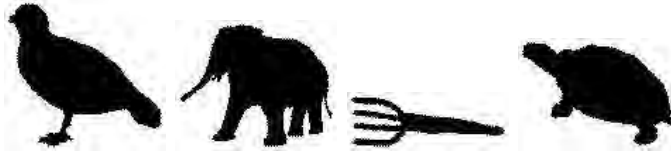
#### **4.1.3 Βάση εικόνων SSID-2**

Οι εικόνες της βάσης SSID-2 προέρχονται από ακριβώς την ίδια ομάδα εργασίας με αυτές τις SSID-1. Έχουν τα ίδια τεχνικά χαρακτηριστικά, αλλάζει όμως το πλήθος τους σε 216. Να τονίσουμε ότι στη βάση αυτή δεν παρατηρείται το φαινόμενο της επεξεργασίας εικόνων (π.χ. κόψιμο περιοχών κτλ), ενώ οι εικόνες προέρχονται από την πρώτη βάση MPEG-7, πρόκειται δηλαδή για ένα μικρό υποσύνολο 216 εικόνων από την MPEG-7.

Ο αριθμός των διακριτών κλάσεων, στις οποίες ανήκουν οι 216 εικόνες είναι 18. Επίσης, σε κάθε μία κλάση ανήκουν 12 εικόνες. Οι κλάσεις αναφέρονται στον παρακάτω πίνακα.

Κλάσεις της βάσης εικόνων SSID-2		
Bird	Classic	Hammer
Bone	Elephant	Heart
Brick	Face	Key
Camel	Fork	Misk
Car	Fountain	Ray
Children	Glas	Turtle

Στο παρακάτω σχήμα (εικόνα 28) δίνουμε μερικές εικόνες της βάσης αυτής.



Εικόνα 28

#### 4.1.4 Εφαρμογή των EFD σε συγκεκριμένα γεωμετρικά σχήματα

Ανοίγοντας μια παρένθεση στις βάσεις δεδομένων μας θα προσπαθήσουμε να ερμηνεύσουμε τις θεωρητικές τεκμηριώσεις του τρίτου κεφαλαίου με τη χρήση συγκεκριμένου προγράμματος. Σε πρώτη φάση δημιουργήσαμε εικόνες με αντικείμενα τρία γεωμετρικά σχήματα (κύκλος, τετράγωνο, τρίγωνο). Για το κάθε ένα γεωμετρικό σχήμα καταλήξαμε σε πολλές εικόνες, προσπαθώντας να έχουμε μετασχηματισμούς του σχήματος, δηλαδή μετατόπιση, αλλαγή κλίμακας και περιστροφή. Στη συνέχεια εξάγουμε τις τιμές των αναλλοίωτων συντελεστών Elliptic Fourier Descriptor, όπως αυτοί ορίστηκαν στην παραπάνω παράγραφο. Τότε, προσπαθούμε να ερμηνεύσουμε τις τιμές τους, ώστε αυτές να συνάδουν με τη θεωρητική ανάλυση που παρουσιάσαμε ανωτέρω. Τα αποτελέσματα της μεθοδολογίας παρουσιάζονται στην παρούσα ενότητα, ταξινομημένα ανά γεωμετρικό σχήμα (κύκλος, τετράγωνο, τρίγωνο).

Πριν ξεκινήσουμε την παρακάτω ανάλυση θα θέλαμε να αναφερθούμε στις τιμές που λαμβάνουν οι δύο πρώτοι αναλλοίωτοι συντελεστές EDF. Να υπενθυμίσουμε ότι στην ουσία υπολογίζουμε την τιμή

της παράστασης  $\frac{|A_k|}{|A_1|} + \frac{|B_k|}{|B_1|}$ , όπου  $A_k = \frac{a_{xk} + ja_{yk}}{2}$  και  $B_k = \frac{b_{xk} + jb_{yk}}{2}$ .

Από τα παραπάνω συμπεραίνουμε ότι ο δεύτερος συντελεστής, που αντιστοιχεί στην τιμή  $k = 1$  θα

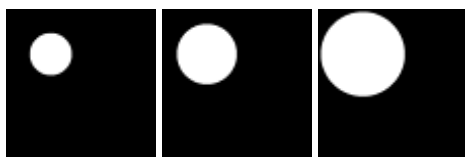
είναι πάντα 2, εφόσον  $\frac{|A_1|}{|A_1|} + \frac{|B_1|}{|B_1|} = 2$ .

Αντίστοιχα ο πρώτος συντελεστής θα δίνεται για  $k = 0$ , θα εξαρτάται δηλαδή από τα  $a_{x0}$  και  $a_{y0}$ , τα οποία δίνουν το κέντρο μάζας του σχήματος. Άρα και ο πρώτος συντελεστής δεν χρησιμεύει στην περιγραφή του περιγράμματος και δεν είναι λογικό να συμπεριληφθεί στην εξαγωγή του προς ταξινόμηση διανύσματος.

## Γεωμετρικό Σήμα: Κύκλος

Δημιουργούμε σε εικόνα μαύρου φόντου κύκλο χρώματος λευκού. Δημιουργούμε επίσης και διαφορετικές εκδοχές του κύκλου προσπαθώντας να επιτύχουμε τους γεωμετρικούς μετασχηματισμούς μετατόπισης και αλλαγής κλίμακας. Το γεωμετρικό μετασχηματισμό της περιστροφής δεν έχει νόημα να τον υλοποιήσουμε, εφόσον εξαιτίας του ορισμού του κύκλου δεν αλλάζει τίποτα σε σχέση με την αρχική εικόνα.

Στις παρακάτω τρεις εικόνες απεικονίζουμε έναν κύκλο μεταβλητής και μάλιστα αυξανόμενης τιμής διαμέτρου.



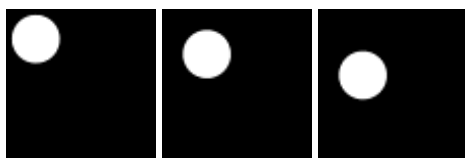
Εικόνα 29

Οι τιμές των αντίστοιχων αναλλοίωτων EFD συντελεστών είναι οι παρακάτω:

circle1.gif :	1258.9	2.0	1.12	1.06	0.93	0.91	0.87	0.87	0.84	0.82
circle2.gif :	1256.9	2.0	1.12	1.06	0.93	0.92	0.87	0.86	0.84	0.83
circle3.gif :	1257.1	2.0	1.12	1.06	0.93	0.92	0.87	0.87	0.84	0.83

Παρατηρούμε ότι όντως οι τιμές των αναλλοίωτων EFD συντελεστών (για  $k > 1$ ) όλων των εικόνων είναι ίδιες μεταξύ τους. Άρα επικυρώνουμε και τη θεωρητική ανάλυση, που ισχυρίζεται την αναλλοίωτη συμπεριφορά των συντελεστών στο μετασχηματισμό της μετακίνησης αλλαγής του σχήματος –πάντα για το αντικείμενο του κύκλου-.

Στη συνέχεια ασχολούμαστε με το γεωμετρικό μετασχηματισμό της μετακίνησης. Πιο συγκεκριμένα δημιουργούμε μία εικόνα με έναν κύκλο και νέες εικόνες, στις οποίες ο κύκλος έχει μετατοπιστεί. Το παρακάτω σχήμα απεικονίζει τους κύκλους αυτούς.



Εικόνα 30

Τα αποτελέσματα είναι τα παρακάτω.

```

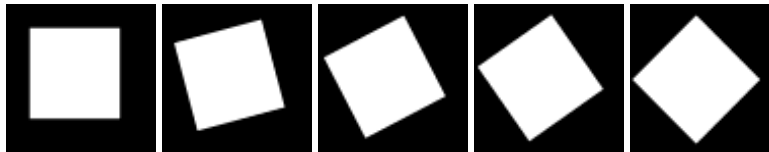
circle4.gif :      835.9 2.0  1.12  1.05  0.93  0.92  0.87  0.86  0.84  0.83
circle5.gif :     1258.1 2.0  1.12  1.05  0.93  0.92  0.87  0.86  0.84  0.83
circle6.gif :     1581.5 2.0  1.12  1.05  0.93  0.92  0.87  0.86  0.84  0.83

```

Παρατηρούμε ότι το αποτέλεσμα συμβαδίζει με τη θεωρητική ανάλυση.

### Γεωμετρικό Σήμα: Τετράγωνο

Πλέον ασχολούμαστε με το τετράγωνο και παράγουμε εικόνες ενός τετραγώνου, το οποίο έχει υποστεί περιστροφή.



Εικόνα 31

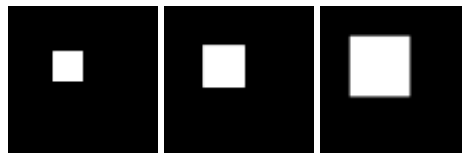
Τα αποτελέσματα που αποκομίζουμε είναι τα ακόλουθα.

```

rect1.gif :      1561.5 2.0  0.90  1.32  1.68  0.97  0.70  1.04  1.52  0.91
rect2.gif :     1176.3 2.0  0.74  1.35  0.76  0.61  1.12  0.71  0.63  0.99
rect3.gif :     2320.0 2.0  1.60  1.24  1.11  1.22  1.08  1.03  1.10  0.98
rect4.gif :     2357.7 2.0  1.47  1.28  1.12  1.19  1.07  1.14  1.03  1.06
rect5.gif :     2663.8 2.0  1.50  1.33  1.25  1.20  1.17  1.14  1.12  1.11

```

Ο επόμενος γεωμετρικός μετασχηματισμός που εξετάζουμε είναι της αλλαγής κλίμακας.



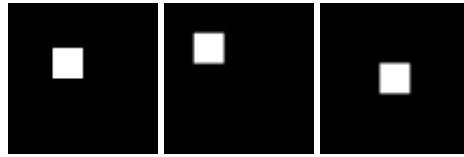
Εικόνα 32

```

rect6.gif :     1357.1 2.0  0.90  1.33  1.68  0.97  0.70  1.06  1.52  0.89
rect7.gif :     1357.1 2.0  0.90  1.33  1.68  0.97  0.70  1.05  1.52  0.90
rect8.gif :     1357.0 2.0  0.90  1.32  1.68  0.97  0.70  1.04  1.52  0.91

```

Στη συνέχεια, ασχολούμαστε με το μετασχηματισμό της μετακίνησης.

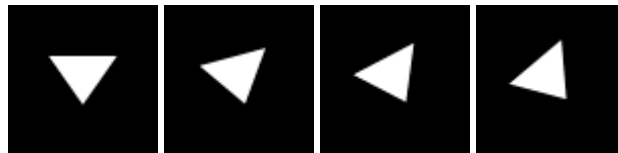


**Εικόνα 33**

```
rect9.gif :      1357.1 2.0  0.90  1.33  1.68  0.97  0.70  1.06  1.52  0.89
rect10.gif :     1016.2 2.0  0.90  1.33  1.68  0.97  0.70  1.06  1.52  0.89
rect11.gif :     1698.1 2.0  0.90  1.33  1.68  0.97  0.70  1.06  1.52  0.89
```

### Γεωμετρικό Σήμα: Τρίγωνο

Το τελευταίο γεωμετρικό σχήμα, με το οποίο ασχολούμαστε, είναι το τρίγωνο. Παράγουμε, λοιπόν, περιστροφές ενός αρχικού τριγώνου.



**Εικόνα 34**

```
triangle1.gif :   32.4 2.0  1.75  0.91  0.59  0.25  0.52  0.27  0.23  0.29
triangle2.gif :   13.7 2.0  0.63  0.39  0.35  0.24  0.23  0.17  0.14  0.13
triangle3.gif :   12.4 2.0  0.49  0.42  0.26  0.22  0.20  0.15  0.14  0.12
triangle4.gif :   14.4 2.0  0.71  0.36  0.34  0.19  0.18  0.15  0.13  0.13
```

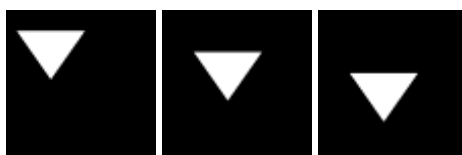
Έπειτα δημιουργούμε εικόνες με ένα τρίγωνο, το οποίο υφίσταται το γεωμετρικό μετασχηματισμό της αλλαγής κλίμακας.



Εικόνα 35

triangle5.gif :	41.6	2.0	1.75	0.91	0.58	0.27	0.52	0.27	0.22	0.29
triangle6.gif :	25.6	2.0	1.75	0.91	0.59	0.25	0.52	0.27	0.23	0.29
triangle7.gif :	20.2	2.0	1.75	0.91	0.59	0.25	0.52	0.26	0.23	0.28
triangle8.gif :	16.5	2.0	1.75	0.91	0.59	0.25	0.52	0.27	0.23	0.28

Και τέλος εφαρμόζουμε μετατόπιση σε ένα τρίγωνο και παρατηρούμε πιθανές αλλαγές στις τιμές των EFD συντελεστών.

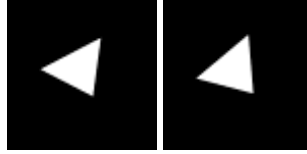


Εικόνα 36

triangle9.gif :	18.8	2.0	1.75	0.91	0.59	0.25	0.52	0.27	0.23	0.29
triangle10.gif:	28.3	2.0	1.75	0.91	0.59	0.25	0.52	0.27	0.23	0.29
triangle11.gif:	33.2	2.0	1.75	0.91	0.59	0.25	0.52	0.27	0.23	0.29

Για την επιβεβαίωση των σωστών αποτελεσμάτων ταξινομήσαμε όλες τις παραπάνω τιμές με τη βοήθεια της Weka. Χρησιμοποιήσαμε ταξινομητή NaiveBayes και k-fold cross-validation μεθοδολογία επικύρωσης (με k=10), ενώ το ποσοστό επιτυχίας είναι 100%, αποδुकνειόντας τη σωστή λειτουργία της παραπάνω μεθοδολογίας.

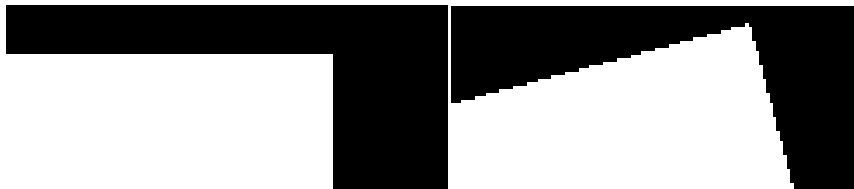
**Παρατήρηση:** Πιθανότατα ο αναγνώστης να εντόπισε ότι στην περίπτωση της περιστροφής οι τιμές των EFD συντελεστών δεν είναι ακριβώς ίδιοι μεταξύ τους, όπως συμβαίνει στους άλλους δύο γεωμετρικούς μετασχηματισμούς (μετατόπιση και αλλαγή κλίμακας). Αναφέρουμε ως παράδειγμα τα δύο παρακάτω σχήματα και τις αντίστοιχες τιμές που προκύπτουν.



**Εικόνα 37**

triangle3.gif :	12.4	2.0	0.49	0.42	0.26	0.22	0.20	0.15	0.14	0.12
triangle4.gif :	14.4	2.0	0.71	0.36	0.34	0.19	0.18	0.15	0.13	0.13

Εδώ θα θέλαμε να επισημάνουμε ότι στην περίπτωση της περιστροφής δεν παράγεται ακριβώς το ίδιο περίγραμμα, καθώς εξαιτίας της στροφής ενός σχήματος, μία πλήρως οριζόντια π.χ. γραμμή μπορεί να μετατραπεί σε γραμμή που σχηματίζει γωνία π.χ. 20 μοιρών. Τότε παρατηρείται το φαινόμενο να εντοπίζονται «σκαλοπατάκια» στην νέα γραμμή, και άρα αρκετά διαφορετικό περίγραμμα από την αρχική εικόνα. Το γεγονός αυτό απεικονίζεται και στο παρακάτω σχήμα.



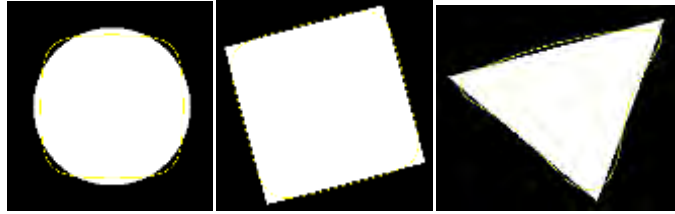
**Εικόνα 38**

Η παραπάνω παρατήρηση αλλάζει αρκετά το περίγραμμα (από πλήρως ευθεία γραμμή σε γραμμή με έντονες διακυμάνσεις), γεγονός που επιφέρει και αλλαγές στις τιμές των EFD συντελεστών. Εν γένει όμως, οι αλλαγές αυτές δεν είναι έντονες και σημαντικές, καθώς και τα σχήματα αυτά ταξινομούνται σωστά, όπως αποδείχθηκε με τη χρήση της Weka.

Κλείνοντας την παράγραφο αυτή, θα θέλαμε να αναφερθούμε σε μία μικρή σύγκριση των τιμών των EFD για τα διαφορετικά γεωμετρικά σχήματα, που χρησιμοποιήσαμε παραπάνω. Θα λέγαμε, ότι όσον αφορά το γεωμετρικό σχήμα του κύκλου, οι EFD συντελεστές δίνουν μία τιμή κοντά στη μονάδα, ενώ για το τετράγωνο παρατηρούμε μεγαλύτερες τιμές για τους συντελεστές 4 και 5. Τέλος, για το τρίγωνο παρατηρούμε τον συντελεστή 3 να παίρνει μεγάλη τιμή, σε αντίθεση με τους συντελεστές 5 και 6, που λαμβάνουν τιμές αρκετά μικρότερες της μονάδας.

Η συμπεριφορά αυτή σχετίζεται με το εκάστοτε γεωμετρικό σχήμα. Γενικά, όσο αυξάνουμε το πλήθος των συντελεστών, τόσο μεταβαίνουμε σε μεγαλύτερο επίπεδο λεπτομέρειας. Στο παρακάτω σχήμα παρατηρούμε την ανακατασκευή των γεωμετρικών σχημάτων, κάνοντας χρήση των τεσσάρων πρώτων συντελεστών.





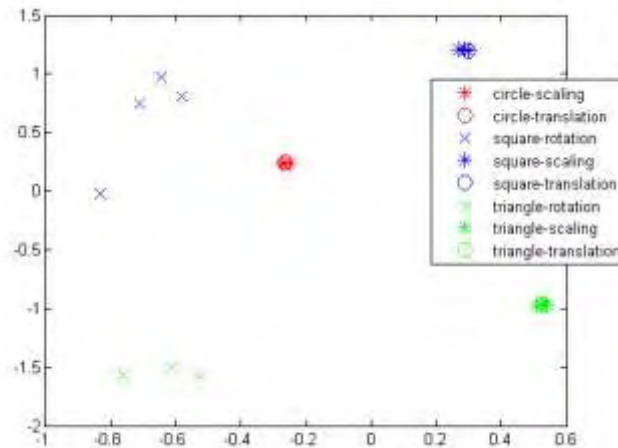
Εικόνα 39

#### **4.1.5 Χρήση της τεχνικής Sammon's Mapping**

Για την επικύρωση των παραπάνω χρησιμοποιήθηκε και η τεχνική του Sammon (Sammon projection ή Sammon's mapping) [Samm69]. Όπως είδαμε στην ενότητα 2.1.3 πρόκειται για μία μεθοδολογία, που μειώνει τη διάσταση του διανύσματος χαρακτηριστικών, με σκοπό τη δυνατότητα εκτύπωσης των δεδομένων και της καλύτερης κατανόησης του προβλήματος.

Ο αλγόριθμος, που προτάθηκε, μετασχηματίζει έναν χώρο υψηλής διάστασης σε έναν χώρο χαμηλότερης. Το κριτήριο του μετασχηματισμού είναι η διατήρηση της δομής των αποστάσεων μεταξύ των σημείων του χώρου υψηλής διάστασης και στο χώρο της χαμηλότερης. Αν, δηλαδή, δύο σημεία απέχουν μεταξύ τους κάποια δεδομένη απόσταση στο χώρο υψηλής διάστασης, δεν θα θέλαμε τα σημεία αυτά να συμπέσουν στο χώρο της χαμηλής διάστασης. Με τον τρόπο αυτό, μπορούμε να απεικονίζουμε στο χώρο των δύο ή και των τριών διαστάσεων τα δεδομένα του προβλήματος και να παρατηρούμε εκτενέστερα τη μορφή τους, κατανοώντας τα αποτελέσματα και την ταξινόμηση των αλγορίθμων που εφαρμόζουμε.

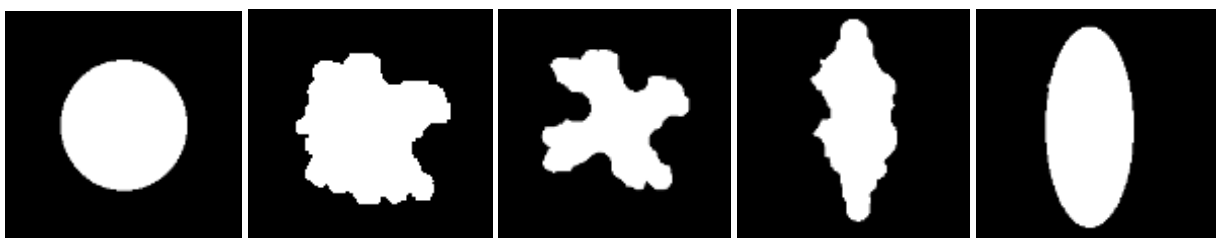
Εφαρμόσαμε τη συγκεκριμένη τεχνική με τη βοήθεια του matlab. Δώσαμε στο πρόγραμμα ως δεδομένα τις τιμές της παραπάνω παραγράφου. Οι τιμές αυτές αντιστοιχούν στα γεωμετρικά σχήματα του κύκλου, του τετραγώνου και του τριγώνου. Στα σχήματα αυτά εφαρμόσαμε τους γεωμετρικούς μετασχηματισμούς της αλλαγής κλίμακας (scaling), της μετατόπισης (translation) και της περιστροφής (rotation). Δίνοντας τα δεδομένα αυτά στο πρόγραμμα και μειώνοντας το διάνυσμα χαρακτηριστικών σε χώρο διάστασης δύο (2), προκύπτουν νέα διανύσματα χαρακτηριστικών, τα οποία και εκτυπώνουμε. Το αποτέλεσμα δίνεται στο παρακάτω σχήμα (εικόνα 40). Στο σχήμα αυτό παρατηρούμε το κάθε ένα διάνυσμα χαρακτηριστικών, το οποίο και εκτυπώνουμε με διαφορετικό σχήμα και χρώμα, ανάλογα με το γεωμετρικό σχήμα που αντιστοιχεί και το γεωμετρικό μετασχηματισμό που εφαρμόστηκε.



**Εικόνα 40**

Ως προς το σχολιασμό του συγκεκριμένου σχήματος, μπορούμε να αναφέρουμε τα ακόλουθα, όπως αυτά διαπιστώθηκαν και παραπάνω. Οι γεωμετρικοί μετασχηματισμοί της αλλαγής κλίμακας (scaling) και της μετατόπισης (translation) δεν μεταβάλλουν τις τιμές των συντελεστών EFD. Αυτό παρατηρείται από τους κύκλους και τους αστερίσκους του σχήματος. Όταν δηλαδή το γεωμετρικό σχήμα είναι κύκλος ή τετράγωνο ή τρίγωνο, όποιον από τους δύο αυτούς μετασχηματισμούς εφαρμόσουμε, οι νέες τιμές των συντελεστών EFD θα είναι πολύ κοντινές με τις αρχικές. Αλλαγές όμως έχουμε στην περίπτωση της περιστροφής (rotation) που αντιστοιχεί με τα σημεία 'x' στο σχήμα. Οι νέες τιμές των συντελεστών δεν ταυτίζονται με τις αρχικές, ομαδοποιούνται όμως σε συγκεκριμένα σημεία και άρα είναι δυνατή η αναγνώρισή τους από τους αλγορίθμους που χρησιμοποιούμε.

Με τη βοήθεια της τεχνικής του Sammon's Mapping πραγματοποιήσαμε ένα ακόμα πείραμα με τις παρακάτω εικόνες όπου αναπαραστήσαμε τα διανύσματα της κάθε μιας στον δισδιάστατο χώρο



**Εικόνα 1**

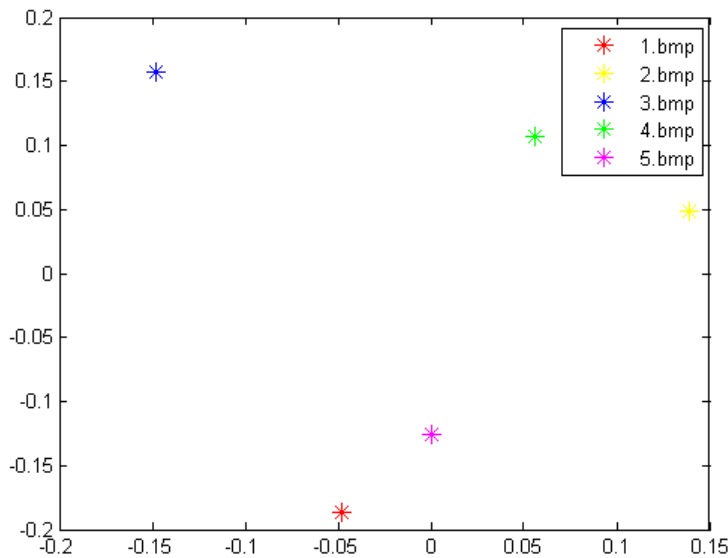
**Εικόνα 2**

**Εικόνα 3**

**Εικόνα 4**

**Εικόνα 5**

Σκοπός του πειράματος είναι να καταλάβουμε κατά πόσο οι Fourier Descriptors είναι ικανοί να διαφοροποιήσουν ομαλό από ανώμαλο περίγραμμα, κάτι που είναι συχνό στη διαφοροποίηση καλοθών από κακοθών όγκων.



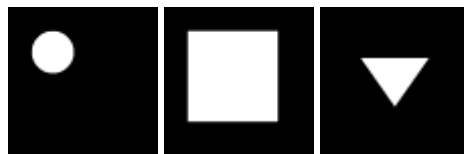
**Εικόνα 41, Γραφική αναπαράσταση Sammon's mapping**

Όπως φαίνεται και απ το παραπάνω αποτέλεσμα (εικόνα 41), βλέπουμε τον επιτυχή διαχωρισμό που θα θέλαμε.

#### **4.1.6 Βάση εικόνων με γεωμετρικά σχήματα**

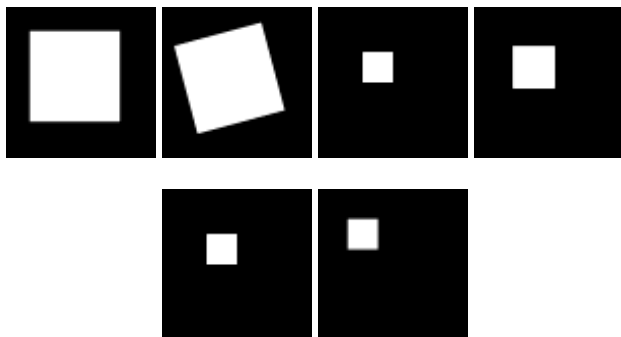
Στα πλαίσια της παρούσης εργασίας χρησιμοποιήσαμε και μία βάση εικόνων με συγκεκριμένα γεωμετρικά σχήματα. Ο σκοπός τους είναι να παρατηρήσουμε τη μεταβολή των αναλλοίωτων EFD συντελεστών, ώστε να αποδειχθεί και με πειραματικά αποτελέσματα η αμετάβλητη συμπεριφορά τους στους διάφορους γεωμετρικούς μετασχηματισμούς.

Πιο συγκεκριμένα, επιλέχθηκαν τα γεωμετρικά σχήματα του κύκλου, του τετραγώνου και του τριγώνου και δημιουργήθηκαν σχετικές εικόνες, με χρώμα λευκό του αντικειμένου και μαύρο του υποβάθρου. Τα σχήματα αυτά απεικονίζονται και στο παρακάτω σχήμα, ενώ αποθηκεύτηκαν ως αρχεία εικόνων με τη βοήθεια του αντίστοιχου προγράμματος σε πηγαίο κώδικα matlab.



**Εικόνα 42**

Στη συνέχεια εφαρμόσαμε τους γεωμετρικούς μετασχηματισμούς της μετατόπισης (translation), της αλλαγής κλίμακας (scaling) και της περιστροφής (rotation) σε κάθε ένα γεωμετρικό σχήμα, με εξαίρεση τον κύκλο, όπου η περιστροφή δεν έχει νόημα. Στο παρακάτω σχήμα (εικόνα 43) απεικονίζουμε μερικά παραδείγματα περιστροφής, αλλαγής κλίμακας και μετατόπισης του τετραγώνου. Αντίστοιχα σχήματα υπάρχουν και για τα άλλα δύο γεωμετρικά σχήματα.



Εικόνα 43

Στον παρακάτω πίνακα παρουσιάζουμε συνοπτικά τις εικόνες της βάσης αυτής, όπως επίσης και πως κάθε μία από αυτές αντιστοιχίζεται στα γεωμετρικά σχήματα και στους γεωμετρικούς μετασχηματισμούς που αναφέραμε παραπάνω.

Εικόνες της βάσης εικόνων με γεωμετρικά σχήματα			
	Περιστροφή	Αλλαγή κλίμακας	Μετατόπιση
Κύκλος		circle1	circle4
		circle2	circle5
		circle3	circle6
Τετράγωνο	rect1	rect6	rect9
	rect2	rect7	rect10
	rect3	rect8	rect11
	rect4		
	rect5		
Τρίγωνο	triangle1	triangle5	triangle9
	triangle2	triangle6	triangle10
	triangle3	triangle7	triangle11
	triangle4	triangle8	

#### **4.1.7 Δημιουργία βάσης εικόνων με ιατρικό περιεχόμενο**

Στο βήμα αυτό, θα θέλαμε να εφαρμόσουμε την τεχνική ανάκλησης εικόνων που υλοποιήσαμε, σε ιατρικά δεδομένα, που αναπτύξαμε μόνοι μας. Για το σκοπό αυτό, χρησιμοποιήσαμε τα δεδομένα, που παρατίθενται δημόσια και ελεύθερα -για ερευνητικούς σκοπούς- από σχετική ιστοσελίδα [caps]. Εκεί, διατίθενται πλήθος εικόνων από διάφορα ιατρικά περιστατικά (cases). Μάλιστα, υπάρχει και η σχετική επισήμανση (annotation) του κάθε περιστατικού από τους αντίστοιχους ειδικούς ιατρούς, αυτή όμως περιορίζεται κάθε φορά στον χαρακτηρισμό του περιστατικού και όχι στην επισήμανση της συγκεκριμένης περιοχής, στην οποία εντοπίζεται η εκάστοτε ανωμαλία. Επίσης, θα θέλαμε να τονίσουμε ότι ο συγκεκριμένος ιστότοπος διαθέτει για κάθε περιστατικό μία εικόνα και ένα video (wmv αρχείο) όπου απεικονίζεται η ανωμαλία.

Στην παρούσα εργασία περιοριστήκαμε στη μελέτη περιπτώσεων έλκους (Ulcer). Συνδεθήκαμε στην ιστοσελίδα και εντοπίσαμε τις περιπτώσεις έλκους, που χαρακτηρίζεται από τους ειδικούς ιατρούς σε μία από τις ακόλουθες κλάσεις:

- Quiescent
- Active moderate
- Active severe

Συνολικά εντοπίσαμε 9 περιπτώσεις (cases), 3 για κάθε μία από τις παραπάνω κλάσεις. Στις επιμέρους περιπτώσεις παρέχονται:

- Η περιγραφή του περιστατικού από ειδικούς ιατρούς με τη μορφή κειμένου (description.txt αρχείο), όπως για παράδειγμα δίνεται παρακάτω

```
Image Title: Ulcer  
Organ Type: Small Bowel  
Diagnosis Type: Quiescent
```

- Εικόνα που απεικονίζει την ανωμαλία που εντοπίστηκε στο συγκεκριμένο περιστατικό (αρχείο image.png)



**Εικόνα 44**

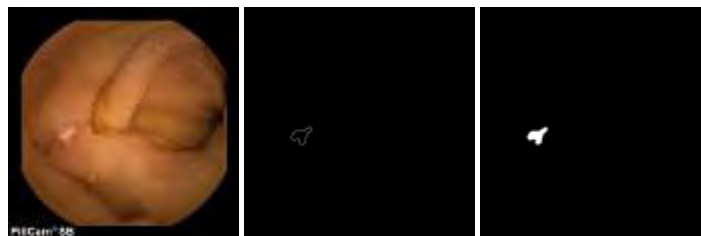
- Βίντεο αρχείο (.wmv) με συνολικά 100 frames, με την παραπάνω εικόνα να αντιστοιχεί στο frame 51.

Με σκοπό να εμπλουτίσουμε τη βάση δεδομένων με επιπλέον εικόνες, ώστε να προσεγγίσουμε έναν ικανοποιητικό αριθμό περιστατικών, χρησιμοποιήσαμε εικόνες από το δοθέν αρχείο βίντεο. Πιο συγκεκριμένα, υλοποιήθηκε κώδικας σε matlab, όπου με τη βοήθεια σχετικών συναρτήσεων (mmreader), έγινε εξαγωγή των frames από το αρχείο βίντεο. Κάθε ένα frame αποθηκεύτηκε στο σχετικό φάκελο, ενώ η ανάλυσή του παρέμεινε όμοια με του αρχικού αρχείου.

Εδώ θα θέλαμε να αναφέρουμε ότι η εικόνα που διατίθεται για κάθε περιστατικό έχει καλύτερη ποιότητα από το αντίστοιχο βίντεο, γεγονός που πιθανότατα δημιούργησε προβλήματα στα επόμενα βήματα της διαδικασίας, μειώνοντας το τελικό ποσοστό επιτυχίας που επιτεύχθηκε.

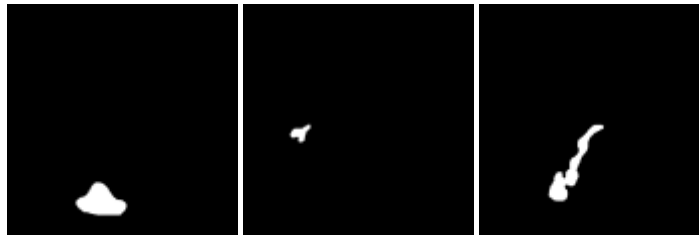
Στη συνέχεια και εφόσον έχουμε αποθηκεύσει όλα τα frames του βίντεο κάθε περιστατικού, απομονώνουμε εκείνα τα οποία αναπαριστούν την ανωμαλία που απεικονίζεται στην εικόνα του περιστατικού. Πλέον για κάθε περιστατικό δεν έχουμε μονάχα μία εικόνας, αλλά περισσότερες (συνήθως 3).

Επόμενο βήμα της διαδικασίας είναι η επισήμανση της ανωμαλίας που διαφαίνεται σε κάθε εικόνα. Στο σημείο αυτό έγινε χρήση του εργαλείου ratsnake [Iako11], με τη βοήθεια του οποίου δόθηκε το περίγραμμα της ανωμαλίας σε κάθε διαθέσιμη εικόνα. Το αποτέλεσμα αποθηκεύεται ως δυαδική εικόνα με την κατάλληλη ονομασία. Επιπρόσθετα, για κάθε ανωμαλία εκτός από το περίγραμμά της αποθηκεύουμε και ολόκληρη την περιοχή της. Όλα τα παραπάνω αναπαριστούνται στο ακόλουθο σχήμα (εικόνα 45).



**Εικόνα 45**

Εν κατακλείδι, διαθέτουμε συνολικά 30 εικόνες, σε κάθε μία από τις οποίες έχει επισημανθεί με τη μορφή μάσκας – δυαδικής εικόνας μία περιοχή που περικλείει την εντοπισμένη ανωμαλία του περιστατικού. Οι 30 αυτές περιπτώσεις είναι κατηγοριοποιημένες από ειδικευμένους ιατρούς σε 3 διακριτές κλάσεις. Παρακάτω (εικόνα 46) παραθέτουμε τρεις δυαδικές μάσκες που ανήκουν στις κλάσεις αυτές.



Εικόνα 46

#### 4.2 Προεπεξεργασία εικόνων

Εφόσον παρουσιάσαμε τις βάσεις που χρησιμοποιήσαμε στα πλαίσια της εργασίας αυτής, αναλύουμε μερικές τεχνικές προ-επεξεργασίας, που εφαρμόσαμε στις εικόνες αυτές. Οι τεχνικές αυτές κρίθηκαν απαραίτητες, ώστε υπολογιστούν σωστά οι τιμές των αναλλοίωτων συντελεστών Elliptic Fourier Descriptors.

Ξεκινώντας την παράγραφο αυτή, θα θέλαμε να αναφερθούμε σε ένα συγκεκριμένο πρόβλημα, που παρατηρήθηκε σε μερικές από τις εικόνες της βάσης MPEG-7. Πιο συγκεκριμένα, όπως είδαμε και παραπάνω, στις εικόνες της βάσης αυτής, το αντικείμενο απεικονίζεται με λευκά pixels, ενώ υπόβαθρο με μαύρα. Εφόσον χρησιμοποιούμε αριθμούς των 8 bit, τότε το λευκό θα απεικονίζεται με την τιμή 255, ενώ το μαύρο με την τιμή 0. Κάποιες, όμως, από τις εικόνες της βάσης αυτής έχουν τιμές διαφορετικές των 0 και 255, για παράδειγμα 250. Για την σωστή εκτέλεση των επόμενων βημάτων της εργασίας, χρησιμοποιήσαμε το matlab, ώστε να διαβάσουμε τις εικόνες και να διορθώσουμε τυγχόν τέτοιες περιπτώσεις. Οι εικόνες της βάσης MPEG-7 με το πρόβλημα που περιγράψαμε αποτυπώνονται στον παρακάτω πίνακα.

Εικόνες της βάσης MPEG-7 με λάθος τιμές pixels					
bell-1.gif	fork-10.gif	fork-4.gif	lizzard-6.gif	spring-6.gif	spring-15.gif
bell-10.gif	fork-14.gif	fork-5.gif	lizzard-9.gif	spring-7.gif	spring-16.gif
bird-1.gif	fork-16.gif	fork-6.gif	lmfish-1.gif	spring-8.gif	spring-17.gif
chopper-16.gif	fork-17.gif	fork-7.gif	sea_snake-12.gif	spring-9.gif	spring-18.gif
crown-1.gif	fork-18.gif	fork-8.gif	spring-1.gif	spring-10.gif	spring-19.gif
dog-1.gif	fork-19.gif	fork-9.gif	spring-2.gif	spring-11.gif	spring-20.gif
elephant-1.gif	fork-2.gif	guitar-1.gif	spring-3.gif	spring-12.gif	tree-8.gif
fish-1.gif	fork-20.gif	hammer-1.gif	spring-4.gif	spring-13.gif	
fork-1.gif	fork-3.gif	lizzard-19.gif	spring-5.gif	spring-14.gif	

Ένα άλλο στάδιο προεπεξεργασίας των εικόνων είναι το γέμισμα των κενών περιοχών εσωτερικά των αντικειμένων. Στο παρακάτω σχήμα παρατηρούμε μερικές εικόνες της βάσης MPEG-7, στις οποίες

εντοπίζονται μαύρες περιοχές εντός του αντικειμένου. Θα πρέπει οι περιοχές αυτές να «γεμίσουν» με λευκά ριξελς, ώστε να εντοπισθεί σωστά το περίγραμμα του αντικειμένου, πάνω στο οποίο θα υπολογιστούν και οι συντελεστές EFD [Zhang03].



**Εικόνα 47**

Ο αλγόριθμος ονομάζεται flood fill, είναι πολύ απλός και υλοποιείται από σχετική συνάρτηση του image processing toolbox του matlab [Soil99]. Οι αντίστοιχες -στις παραπάνω- εικόνες που προκύπτουν μετά την εφαρμογή του αλγορίθμου αυτού φαίνονται στο παρακάτω σχήμα.



**Εικόνα 48**

Πλέον, ακολουθεί ο αλγόριθμος εντοπισμού του περιγράμματος του αντικειμένου. Ο αλγόριθμος είναι υλοποιημένος σε java. Η σχετική συνάρτηση που χρησιμοποιούμε, είναι μία παραλλαγή πηγαίου κώδικα, ο οποίος εντοπίστηκε στο Διαδίκτυο, μετά την αναζήτηση που πραγματοποιήσαμε. Έγιναν όμως αλλαγές, με σκοπό να μπορεί να εφαρμοστεί στα δεδομένα του δικού μας προβλήματος.

Στο παρακάτω σχήμα απεικονίζουμε το περίγραμμα των παραπάνω εικόνων. Από αυτό το περίγραμμα θα εξαχθούν οι αναλλοίωτοι συντελεστές EFD, οι οποίοι είναι ικανοί να περιγράψουν το αντικείμενο που υποδηλώνει το εκάστοτε περίγραμμα.



**Εικόνα 49**

Σε περίπτωση που δεν είχε εκτελεστεί ο αλγόριθμος του γεμίσματος των κενών περιοχών, το περίγραμμα που θα εντόπιζε το βήμα αυτό θα ήταν λανθασμένο. Παρακάτω παραθέτουμε τις αντίστοιχες –με ανωτέρω- εικόνες, χωρίς όμως την εκτέλεση του αλγορίθμου flood fill.





Εικόνα 50

### 4.3 Εξαγωγή διανύσματος χαρακτηριστικών

Το διάνυσμα χαρακτηριστικών που εξάγεται από κάθε εικόνα είναι οι συντελεστές των αναλλοίωτων EFD, όπως αυτοί παρουσιάστηκαν και αναλύθηκαν στην παράγραφο 2.2. Να τονίσουμε ότι οι συντελεστές αυτοί υπολογίζονται πάνω στο περίγραμμα του αντικειμένου, που περιέχει κάθε εικόνα, όπως αυτό προέκυψε από τον αλγόριθμο της παραγράφου 3.2. Το πλήθος των συντελεστών επιλέχθηκε να είναι 10, καθώς αυτό είναι αρκετό για να υποστηρίξουμε ικανοποιητικά ποσοστά επιτυχίας στην ανάκληση εικόνων σύμφωνα με τη βιβλιογραφία [Yang08,Zhang03].

Επιλέχθηκε να χρησιμοποιηθεί πηγαίος κώδικας σε Java, ο οποίος και εντοπίστηκε στο Διαδίκτυο μετά τη σχετική αναζήτηση που πραγματοποιήθηκε [EFDJAVA]. Το συγκεκριμένο πρόγραμμα υλοποιεί και υπολογίζει τους συντελεστές των αναλλοίωτων Elliptic Fourier Descriptors της παραγράφου 2.2 [Nixο08].

Ο πηγαίος κώδικας εμπλουτίστηκε και προσαρμόστηκε στα δεδομένα του προβλήματος που εξετάζουμε. Στη συνέχεια εκτελέστηκε για όλες τις εικόνες των βάσεων που αναφέραμε παραπάνω. Για κάθε μία από τις βάσεις εικόνων παράγεται και ένα σχετικό αρχείο κειμένου. Κάθε σειρά του αρχείου αυτού είναι ως εξής:

```
apple-1.gif: 23.264649 2.0 0.772370 0.297397 0.269375 0.177995 0.239090 0.196705 0.125533 0.127073
```

περιέχει δηλαδή το όνομα του αρχείου εικόνας και 10 τιμές, που αντιστοιχούν στους πρώτους δέκα EFD συντελεστές.

Στο σημείο αυτό θα θέλαμε να θυμίσουμε ότι οι τιμές των δύο πρώτων συντελεστών δεν περιέχουν κάποιου είδους πληροφορία για τη μορφή του σχήματος. Αυτό ισχύει, διότι ο πρώτος συντελεστής αντιστοιχεί στο κέντρο μάζας του σχήματος και ο δεύτερος έχει τιμή πάντα 2, γεγονός που παρατηρείται και από τα αρχεία κειμένου με τις τιμές των συντελεστών.

#### **4.4 Μέθοδος επικύρωσης**

Για την επικύρωση των αποτελεσμάτων ταξινόμησης υλοποιήθηκε η τεχνική k-fold cross validation, με τιμή της μεταβλητής  $k=10$  [Theo09]. Η συγκεκριμένη κατηγορία αλγορίθμων έχει ως σκοπό τη βελτίωση της ικανότητας γενίκευσης (generalization) και τη ρεαλιστική αξιολόγηση της απόδοσης του συστήματος.

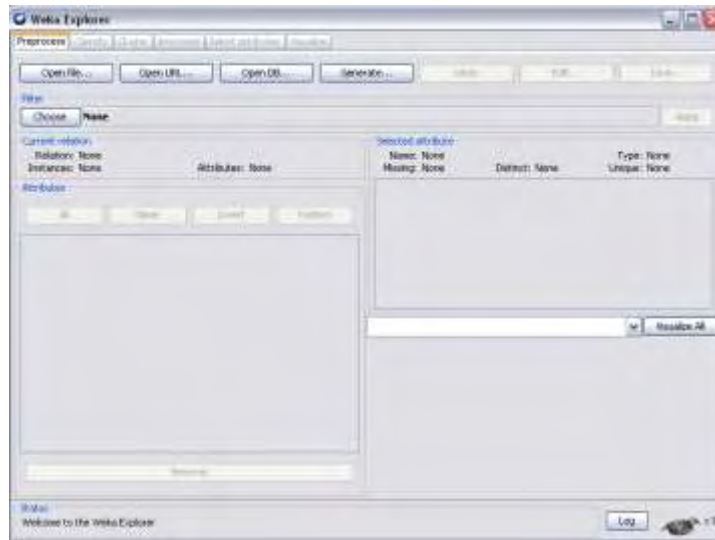
Ο αλγόριθμος είναι ο εξής: το διαθέσιμο σύνολο δεδομένων διαμερίζεται σε  $k$  (συνήθως ίσα) υποσύνολα. Ένα υποσύνολο χρησιμοποιείται για τον έλεγχο του συστήματος (testing ή validation data) ελέγχοντας την ορθότητα της ταξινόμησης, ενώ τα υπόλοιπα  $k-1$  υποσύνολα χρησιμοποιούνται ως δεδομένα εκπαίδευσης του ταξινομητή (training data). Η παραπάνω διαδικασία επαναλαμβάνεται  $k$  φορές, με κάθε ένα από τα  $k$  υποσύνολα να χρησιμοποιείται μόνο μία φορά ως δεδομένα ελέγχου (testing data). Ο μέσος όρος των αποτελεσμάτων από τις  $k$  επαναλήψεις είναι η τελική απόδοση του συστήματος.

Το βασικό πλεονέκτημα της μεθοδολογίας αυτής είναι ότι όλα τα διανύσματα χαρακτηριστικών χρησιμοποιούνται και ως δεδομένα εκπαίδευσης και ως δεδομένα ελέγχου, με κάθε ένα διάνυσμα να χρησιμοποιείται μόνο μία φορά ως δεδομένο ελέγχου.

#### **4.5 Ταξινόμηση**

Στο επόμενο βήμα της εργασίας θέλουμε να ταξινομήσουμε τις εικόνες της κάθε βάσης σε μία κατηγορία, ανάλογα με τις τιμές των EFD συντελεστών. Οφείλουμε να χωρίσουμε τα δεδομένα κάθε βάσης σε δύο υποσύνολα, στο υποσύνολο εκπαίδευσης (training set), με βάση το οποίο εκπαιδεύεται ο εκάστοτε ταξινομητής και στο υποσύνολο ελέγχου (test set), με βάση το οποίο ελέγχεται η απόδοση του κάθε ταξινομητή.

Για τα παραπάνω χρησιμοποιήθηκε το εργαλείο Weka [Weka]. Πρόκειται για μία συλλογή από αλγορίθμους machine learning, υλοποιημένους σε java. Βασικό χαρακτηριστικό του εργαλείου αυτού είναι ότι προσφέρει συγκεκριμένες συναρτήσεις, που μπορούν να καλεσθούν από οποιοδήποτε πρόγραμμα java. Πιο εύχρηστο όμως είναι το σχετικό interface που είναι διαθέσιμο και περιέχει πλήθος αλγορίθμων για προ-επεξεργασία, ταξινόμηση κτλ. Το interface είναι πολύ απλό και ο χρήστης μπορεί να επιλέξει παραμέτρους, όσον αφορά για παράδειγμα το μέγεθος των training set και test set και τον ταξινομητή που θέλει να χρησιμοποιηθεί. Στο παρακάτω σχήμα βλέπουμε ένα print screen από το αντίστοιχο interface που έχει η Weka.



**Εικόνα 51**

Στη Weka δίνουμε με τη μορφή ενός αρχείου κειμένου τα δεδομένα του προβλήματος, επιλέγουμε με τη βοήθεια του γραφικού interface τις παραμέτρους και το εργαλείο αυτό αναλαμβάνει να εκτελέσει τους αντίστοιχους αλγορίθμους και να μας επιστρέψει τα αντίστοιχα αποτελέσματα.

Ο τρόπος που δέχεται ως είσοδο η Weka τα δεδομένα του προβλήματος είναι με τη βοήθεια ενός αρχείου .arff. Διαβάζοντας το εγχειρίδιο χρήσης της Weka αντιλαμβανόμαστε ότι τα αρχεία .arff έχουν ένα συγκεκριμένο format. Στην αρχή έχουμε το header του αρχείου, που είναι της παρακάτω μορφής.

```
% 1. Title: Iris Plants Database
%
% 2. Sources:
% (a) Creator: R.A. Fisher
% (b) Donor: Michael Marshall (MARSHALL%PLU@io.arc.nasa.gov)
% (c) Date: July, 1988
%
@RELATION iris
@ATTRIBUTE sepallength NUMERIC
@ATTRIBUTE sepalwidth NUMERIC
@ATTRIBUTE petallength NUMERIC
@ATTRIBUTE petalwidth NUMERIC
@ATTRIBUTE class {Iris-setosa,Iris-versicolor,Iris-virginica}
```

Περιγράφει δηλαδή τον τύπο του κάθε χαρακτηριστικού του διανύσματος χαρακτηριστικών, ενώ αναφέρει και τις κλάσεις, στις οποίες ανήκουν τα δεδομένα του προβλήματος.

Στη συνέχεια ακολουθούν τα δεδομένα, που δίνονται στην ακόλουθη μορφή.

```
@DATA
5.1,3.5,1.4,0.2,Iris-setosa
4.9,3.0,1.4,0.2,Iris-setosa
```

4.7,3.2,1.3,0.2,Iris-setosa  
 4.6,3.1,1.5,0.2,Iris-setosa  
 5.0,3.6,1.4,0.2,Iris-setosa  
 5.4,3.9,1.7,0.4,Iris-setosa  
 4.6,3.4,1.4,0.3,Iris-setosa  
 5.0,3.4,1.5,0.2,Iris-setosa  
 4.4,2.9,1.4,0.2,Iris-setosa  
 4.9,3.1,1.5,0.1,Iris-setosa

Παρατηρούμε ότι τα δεδομένα χωρίζονται με το κόμμα, ενώ στο τέλος κάθε γραμμής ακολουθεί η κλάση, στην οποία ανήκει το χαρακτηριστικό διάνυσμα που πλοηγήθηκε.

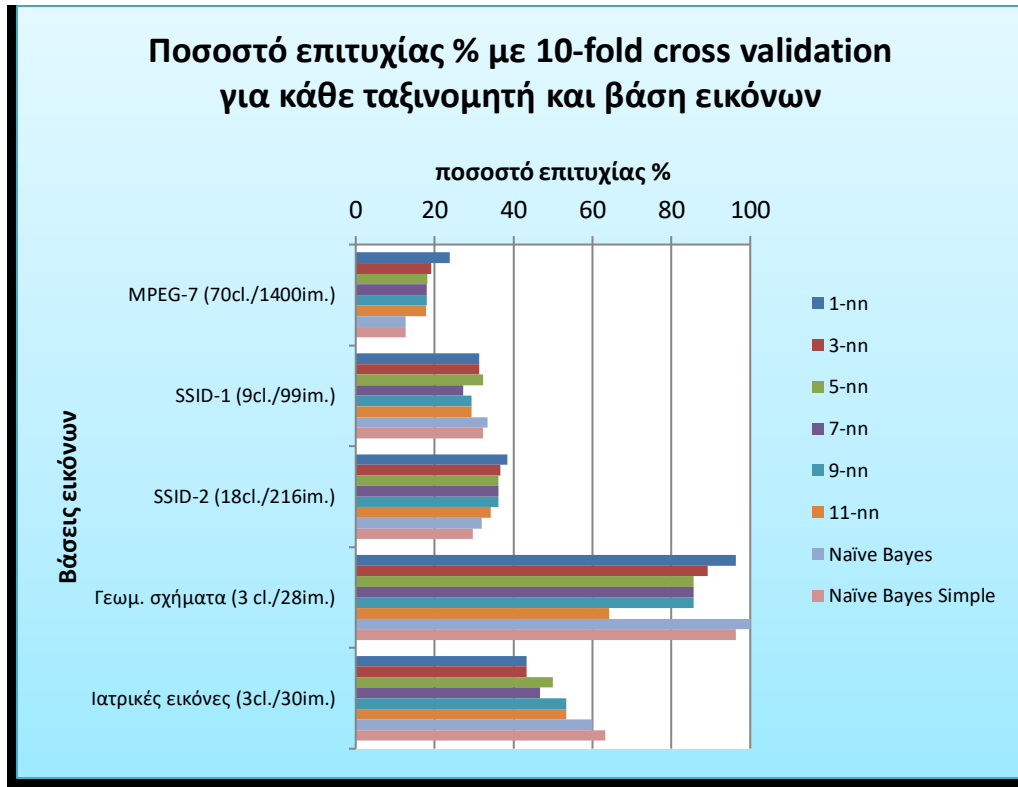
Τα δεδομένα του προβλήματός μας δεν βρίσκονται στη μορφή αυτή. Οπότε, κατασκευάζουμε σχετικό κώδικα σε matlab, ο οποίος διαβάζει τα αρχεία κειμένου με τις τιμές των EFD κάθε εικόνας των βάσεων και τα μετατρέπει στο αντίστοιχο .arff αρχείο κειμένου, ώστε να είναι δυνατή η επεξεργασία του από τη Weka και η εξαγωγή των αποτελεσμάτων της ταξινόμησης.

Στο σύστημα, που σχεδιάσαμε και υλοποιήσαμε, έγινε χρήση του ταξινομητή k πλησιέστερων γειτόνων (k-nearest neighbor classifier - knn), με κλασική ευκλείδεια μετρική απόστασης [Theo09]. Η Weka έχει υλοποιημένη σχετική συνάρτηση, οπότε και ελέγξαμε την απόδοση του συγκεκριμένου ταξινομητή, μεταβάλλοντας την τιμή της παραμέτρου k, από 1 έως 11 με βήμα 2. Τα αποτελέσματα, που εξάγαμε για όλες τις βάσεις εικόνων, παρουσιάζονται στον παρακάτω πίνακα.

Αποτελέσματα ταξινόμησης με 10-fold cross validation για τις βάσεις εικόνων					
	MPEG7	SSID-1	SSID-2	Γεωμετρικά σχήματα	Ιατρικές εικόνες
<b>k-nn (k=1)</b>	23.7857 %	31.3131 %	38.4259 %	96.4286 %	43.3333 %
<b>k-nn (k=3)</b>	19.0714 %	31.3131 %	36.5741 %	89.2857 %	43.3333 %
<b>k-nn (k=5)</b>	18.1429 %	32.3232 %	36.1111 %	85.7143 %	50 %
<b>k-nn (k=7)</b>	17.9286 %	27.2727 %	36.1111 %	85.7143 %	46.6667 %
<b>k-nn (k=9)</b>	17.9286 %	29.2929 %	36.1111 %	85.7143 %	53.3333 %
<b>k-nn (k=11)</b>	17.7857 %	29.2929 %	34.2593 %	64.2857 %	53.3333 %
<b>Naïve Bayes</b>	12.5714 %	33.3333 %	31.9444 %	100 %	60 %
<b>Naïve Bayes Simple</b>	12.6429 %	32.3232 %	29.6296 %	96.4286 %	63.3333 %
<b>Αριθμός κλάσεων</b>	70	9	18	3	3
<b>Αριθμός προτύπων ανά κλάση</b>	20	11	12		10
<b>Συνολικός αριθμός προτύπων</b>	1400	99	216	28	30

Εικόνα 52

Με σκοπό την οπτική επισκόπηση των παραπάνω αποτελεσμάτων παρουσιάζουμε και το παρακάτω διάγραμμα.



**Εικόνα 53**

Από τα αποτελέσματα του πίνακα παρατηρούμε ότι το υψηλότερο ποσοστό επιτυχίας, 100%, επιτυγχάνεται για τη βάση εικόνων με τα γεωμετρικά σχήματα. Το γεγονός αυτό είναι απολύτως λογικό, μιας και τη βάση αυτή θα τη χαρακτηρίζαμε ως την πιο 'εύκολη', καθώς περιέχει λίγες εικόνες που διαχωρίζονται πολύ εύκολα μεταξύ τους. Έπειτα, εντοπίζεται η βάση με τις ιατρικές εικόνες, που επιτυγχάνει ποσοστό επιτυχίας 63.3%. Στη συνέχεια ακολουθεί η βάση SSID-2 με 38.4%, που περιέχει ένα σχετικά μικρό υποσύνολο των εικόνων της βάσης MPEG7, η οποία και κατατάσσεται στην τελευταία θέση, όσον αφορά το ποσοστό επιτυχίας με 23.8%. Ενδιάμεσα εντοπίζεται η βάση SSID-1. Οι εικόνες της βάσης αυτής έχουν παραμορφωθεί, με σκοπό να εξαχθούν συμπεράσματα για την ικανότητα του συστήματος με παρουσία θορύβου στις εικόνες. Για το λόγο αυτό και παρόλο που περιέχει λιγότερες κλάσεις από τη βάση SSID-2, δεν επιτυγχάνει υψηλότερα ποσοστά επιτυχίας και αρκείται στο 32.3%.

Τα ποσοστά επιτυχίας που αναφέρουμε πιθανόν να φαίνονται πολύ μικρά στον αναγνώστη, να τονίσουμε όμως ότι το πρόβλημα που καλούμαστε να επιλύσουμε είναι δύσκολο. Πιο συγκεκριμένα θα πρέπει να αναφέρουμε ότι στη βάση MPEG7, για παράδειγμα, έχουμε 70 διακριτές κλάσεις με 20

πρότυπα ανά κλάση. Υψηλά ποσοστά επιτυγχάνονται στη βάση με τα γεωμετρικά σχήματα, καθώς εκεί έχουμε μονάχα 3.

Από τους διαφορετικούς ταξινομητές που δοκιμάσαμε και ελέγξαμε την απόδοσή τους, ικανοποιητικότερα αποτελέσματα φαίνεται να δίνει ο k-nn με τιμή k=1, όπως επίσης και οι Bayesian ταξινομητές [Theo09].

Πραγματοποιώντας μία προσπάθεια, να ερμηνεύσουμε με καλύτερο τρόπο τα αποτελέσματα, που λαμβάνουμε, από την εφαρμογή της μεθοδολογίας στην βάση με τις ιατρικές εικόνες, καταλήγουμε στον παρακάτω πίνακα.

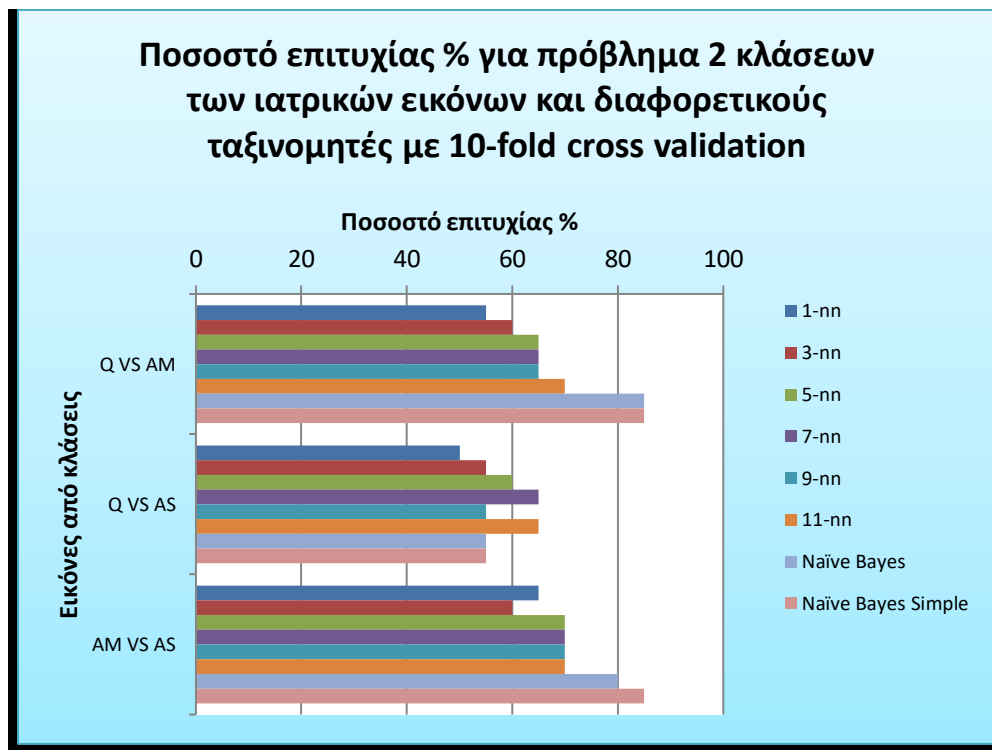
Στον πίνακα αυτό απεικονίζονται 3 ξεχωριστά πειράματα, τα δεδομένα των οποίων εξήχθησαν ακολούθως:

- Στη βάση με τις ιατρικές εικόνες διαθέτουμε 3 διακριτές κλάσεις (Quiescent, Active Moderate και Active Severe)
- Κατασκευάζουμε δεδομένα για το πείραμα 1, που αποτελούνται μόνο από τις Quiescent και Active Moderate περιπτώσεις
- Αντίστοιχα και για τα πειράματα 2 και 3, με Quiescent VS Active Severe και Active Moderate VS Active Severe.

Αποτελέσματα ταξινόμησης με 10-fold cross validation για τη βάση ιατρικών εικόνων			
	Quiescent VS Active Moderate	Quiescent VS Active Severe	Active Moderate VS Active Severe
k-nn (k=1)	55 %	50 %	65 %
k-nn (k=3)	60 %	55 %	60 %
k-nn (k=5)	65 %	60 %	70 %
k-nn (k=7)	65 %	<u>65 %</u>	70 %
k-nn (k=9)	65 %	55 %	70 %
k-nn (k=11)	70 %	65 %	70 %
Naïve Bayes	85 %	55 %	80 %
Naïve Bayes Simple	<u>85 %</u>	55 %	<u>85 %</u>
Αριθμός κλάσεων	2	2	2
Αριθμός προτύπων ανά κλάση	10	10	10
Συνολικός αριθμός προτύπων	20	20	20

Εικόνα 54

Προσπαθώντας και εδώ να επιτύχουμε μία οπτική αναπαράσταση των παραπάνω αποτελεσμάτων, καταλήγουμε στο ακόλουθο διάγραμμα.



Εικόνα 55

Παρατηρούμε ότι το πρόβλημα, που είναι δύσκολο να επιλυθεί και προκαλεί χαμηλά ποσοστά επιτυχίας, είναι το Quiescent VS Active Severe. Συνεπώς, καταλήγουμε ότι τα συγκεκριμένα χαρακτηριστικά που επιλέχθηκαν, δεν είναι ικανά να διαχωρίσουν ικανοποιητικά τα δεδομένα των δύο αυτών κλάσεων.

Σε μία προσπάθεια σύγκρισης της μεθοδολογίας που υλοποιήσαμε με αντίστοιχες εργασίες της βιβλιογραφίας, αναφέρουμε το άρθρο [Late00]. Εκεί χρησιμοποιήθηκε ολόκληρη η βάση εικόνων MPEG7, ενώ το διάγραμμα χαρακτηριστικών, που κατασκευάστηκε, έκανε χρήση πιο εξειδικευμένων και πολύπλοκων τεχνικών, όπως ροπές Zernike και αναπαράσταση περιγράμματος από wavelet. Μάλιστα, σημειώνεται ότι έχει αποδειχθεί, πως οι τεχνικές wavelet δίνουν καλύτερα αποτελέσματα σε σχέση με τους Fourier Descriptors [Chua96].

Παρόλα αυτά μία άμεση σύγκριση των αποτελεσμάτων της παρούσης εργασίας δεν είναι δυνατή, καθώς η [Late00] χρησιμοποιεί διαφορετικό τρόπο αξιολόγησης των αποτελεσμάτων. Εκεί, κάθε μία από τις 1400 εικόνες θεωρείται ως εικόνα-επερώτηση (query image). Η εικόνα αυτή δίνεται στο εκπαιδευμένο σύστημα και αυτό επιστρέφει στο χρήστη τις υπόλοιπες εικόνες της βάσης,

ταξινομημένες με κριτήριο την ομοιότητά τους ως προς την εικόνα-επερώτηση. Το σύστημα δηλαδή επιστρέφει στο χρήστη ως εικόνα νούμερο 1, αυτήν που εκείνο θεωρεί ότι είναι η πλέον όμοια με την εικόνα-επερώτηση, και συνεχίζει ανάλογα. Το σύστημα αξιολόγησης της εργασίας [Late00] μετράει τον αριθμό των εικόνων της ίδιας κλάσης με αυτήν της εικόνας-επερώτησης, αλλά στις πρώτες 40 εικόνες που επιστρέφει το σύστημα. Αν δηλαδή το σύστημα επιστρέψει και τις 20 εικόνες της κλάσης (σε κάθε κλάση της βάσης MPEG7 έχουμε 20 εικόνες) μέσα στις 40 πρώτες, τότε αυτό θεωρείται ως 100% επιτυχία.

Προφανώς ο τρόπος αυτός αξιολόγησης διαφέρει σημαντικά από το δικό μας. Σε συνδυασμό και με τα διαφορετικά και πιθανότατα πιο αποτελεσματικά χαρακτηριστικά που χρησιμοποιούνται στην [Late00], κάθε προσπάθεια άμεσης σύγκρισης δεν έχει νόημα. Όμως, μία αναφορά των αποτελεσμάτων θα είναι σίγουρα χρήσιμη. Παρατηρούμε, λοιπόν, ότι στην [Late00] το ποσοστό επιτυχίας κυμαίνεται μεταξύ 60% και 75%, υπολογισμένο όπως αναφέραμε παραπάνω. Το ποσοστό αυτό φαίνεται πολύ μεγαλύτερο από αυτό που επιτυγχάνει η παρούσα εργασία, να τονίσουμε όμως τη χρήση εντελώς διαφορετικού τρόπου αξιολόγησης των αποτελεσμάτων, με αυτόν την παρούσης εργασίας να είναι πολύ πιο 'αυστηρός'.

Μία άλλη εργασία, που χρησιμοποιεί ίδια βάση εικόνων με την παρούσα είναι η [Seba01]. Εκεί, ελέγχεται η απόδοση του προτεινόμενου αλγορίθμου με κριτήριο τη βάση εικόνων SSID-1. Ο αλγόριθμος, που αναπτύσσεται και υλοποιείται, στηρίζεται στην ανάλυση και αναπαράσταση των σχημάτων (shape deformation) σε πιο απλά σχήματα και στον ορισμό μία μετρικής απόστασης (edit distance) μεταξύ δύο σχημάτων ως επέκταση της ανάλυσης αυτής.

Πρόκειται για αρκετά ενδιαφέρουσα και μάλιστα πολύ αποδοτική τεχνική. Η μεθοδολογία, όμως, αξιολόγησης -και στην περίπτωση αυτή- διαφέρει από αυτήν της παρούσης εργασίας. Πιο συγκεκριμένα, στην [Seba01] γίνεται χρήση της μετρικής edit distance ( $dist(i,j)$ ), που αναφέραμε παραπάνω. Για κάθε μία εικόνα  $i$  από τις 99 της βάσης SSID-1 υπολογίζεται η τιμή της μετρικής αυτής ως προς όλες τις άλλες εικόνες  $j$ , δηλαδή βρίσκονται όλες οι  $dist(i,j)$  και ταξινομούνται με αυξανόμενη σειρά. Ως φυσικό επακόλουθο, εικόνες από την ίδια κλάση θα έχουν μικρή τιμή  $dist(i,j)$  και θα πρέπει να παρουσιάζονται στην αρχή της ταξινομημένης λίστας. Στο τέλος της επαναλαμβανόμενης διαδικασίας καταλήγει σε ένα διάνυσμα 10 τιμών. Στην πρώτη θέση του διανύσματος είναι ο μέσος όρος των εικόνων (για τις 99 επαναλήψεις) που βρίσκονται στην πρώτη θέση της ταξινομημένης λίστας και ανήκουν στην ίδια κλάση με την εικόνα-επερώτηση. Εφόσον έχουμε 11 εικόνες στην ίδια κλάση, οι όμοιες εικόνες με την εικόνα επερώτηση θα είναι 10. Άρα, παραθέτουμε την τιμή που αφορά τη 10<sup>η</sup> θέση του διανύσματος και είναι 87%. Δηλαδή, αν θεωρήσουμε ως άγνωστο πρότυπο μία εικόνα της βάσης, το 87% των πρώτων 10 -επιστρεφόμενων από το σύστημα- εικόνων θα ανήκει στην ίδια κλάση.

Αυτή η μεθοδολογία επικύρωσης ονομάζεται leave-one-out και είναι διαφορετική -και λιγότερο αξιόπιστη- από τη διαδικασία 10-fold cross validation, που χρησιμοποιήσαμε στην παρούσα εργασία. Το γεγονός αυτό, σε συνδυασμό με τη χρήση στην [Seba01] πιο πολύπλοκων αλλά και αποδοτικών τεχνικών για την περιγραφή ενός σχήματος, εξηγούν και τη διαφορά στα ποσοστά των δύο εργασιών.



Τέλος, θα θέλαμε να αναφέρουμε ότι στην παρούσα εργασία ελέγχθηκε μόνο η απόδοση του ταξινομητή πλησιέστερων γειτόνων ( $k$ nn) και των απλών Bayesian ταξινομητών. Οι συγκεκριμένοι ταξινομητές αποδίδουν συνήθως καλά αποτελέσματα, όχι όμως τα μέγιστα δυνατά, καθώς δεν χρησιμοποιούν πιο αποδοτικά μαθηματικά μοντέλα, όπως τα νευρωνικά δίκτυα (neural networks) [Hakn97] ή οι μηχανές διανυσματικής μάθησης (Support Vector Machine – SVM) [Varn06]. Οι τελευταίοι όμως ταξινομητές είναι πολύπλοκοι αλγοριθμικά, με αποτέλεσμα να είναι απαραίτητο πολύ μεγαλύτερο χρονικό διάστημα, για την εκπαίδευσή τους. Επίσης, οι παράμετροί τους είναι πολύ περισσότεροι από την μία παράμετρο -την μεταβλητή  $k$ - του  $k$ nn ταξινομητή, γεγονός που δημιουργεί την ανάγκη για εκτεταμένα πειράματα, με σκοπό την εύρεση των βέλτιστων τιμών των παραμέτρων. Εξαιτίας των παραπάνω λόγων δεν εξετάστηκε η χρήση άλλων ταξινομητών στην παρούσα εργασία, αλλά περιοριστήκαμε σε αυτούς που αναφέραμε παραπάνω.

## ΚΕΦΑΛΑΙΟ 5: ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΠΡΟΟΠΤΙΚΕΣ

Στην παρούσα εργασία έγινε ανάλυση και υλοποίηση των Elliptic Fourier Descriptors (EFD). Οι περιγραφείς αυτοί χρησιμοποιούνται για την ταξινόμηση του σχήματος ενός αντικειμένου, ανήκουν, λοιπόν, στην ευρύτερη κατηγορία των χαρακτηριστικών σχήματος (shape descriptors). Εξάγονται από ολόκληρο το περίγραμμα του αντικειμένου, εφόσον έχουν προηγουμένως εφαρμοστεί τεχνικές κατάτμησης σε μία εικόνα. Οι μετρικές EFD αποτελούν ειδική υποκατηγορία των περιγραφέων Fourier, οι οποίοι αναγνωρίζονται από τη βιβλιογραφία ως οι πλέον διαδεδομένοι ανάμεσα στα χαρακτηριστικά σχήματος.

Οι περιγραφείς Fourier διακρίνονται για την υψηλή τους απόδοση και αξιοπιστία, όσον αφορά τα ποσοστά επιτυχίας που δίνουν σε εφαρμογές ανάκλησης εικόνων με βάση το περιεχόμενο. Ταυτόχρονα, διέπονται από φυσική σημασία, καθώς έχουν θεμελιωθεί με βάση το –επιστημονικά τεκμηριωμένο και μελετημένο– μετασχηματισμό Fourier. Επίσης, η υλοποίησή τους είναι απλή και δεν απαιτούν αυξημένη υπολογιστική ισχύ. Λόγω των παραπάνω έχουν επιλεγεί μαζί με άλλα χαρακτηριστικά– από το πρότυπο MPEG-7 για την περιγραφή εικόνων [Zhan07].

Η απόδοση των μετρικών EFD ελέγχθηκε σε πλήθος εικόνων από βάσεις της βιβλιογραφίας, αλλά και από βάσεις εικόνων που κατασκευάσαμε στα πλαίσια της παρούσης εργασίας. Για το πρόβλημα της ταξινόμησης έγινε χρήση του ταξινομητή ελάχιστης ευκλείδειας απόστασης αλλά και απλών *bayesian ταξινομητών*. Όσον αφορά το πρόβλημα της αξιολόγησης, πραγματοποιήθηκε με κριτήριο το μοντέλο *10-fold cross validation*, με βάση το οποίο το 1/10 της διαθέσιμης βάσης εικόνων επιλέγεται ως σύνολο ελέγχου και το υπόλοιπο ως σύνολο εκπαίδευσης, με τη διαδικασία να επαναλαμβάνεται δέκα φορές.

Η πρώτη βάση, που χρησιμοποιήσαμε, είναι διαθέσιμη στο Διαδίκτυο [Imag] και δημιουργήθηκε με σκοπό την εξέλιξη του προτύπου MPEG-7 [Mpeg]. Αποτελείται συνολικά από 1400 εικόνες, που ταξινομούνται σε 70 κλάσεις με ισάριθμο αριθμό εικόνων (20) κάθε μία. Εξαιτίας του μεγάλου αριθμού κλάσεων, το πρόβλημα της ταξινόμησης των εικόνων της βάσης αυτής χαρακτηρίζεται ιδιαίτερα δύσκολο και απαιτητικό. Παρόλα αυτά, η μεθοδολογία που υλοποιήθηκε στα πλαίσια της παρούσης εργασίας απέδωσε ποσοστό επιτυχίας 23.8%. Ο συγκεκριμένος αριθμός μπορεί να φαίνεται χαμηλός, θα πρέπει όμως να αναλογιστούμε ότι το πρόβλημα της ταξινόμησης περιλαμβάνει 70 κλάσεις, γεγονός που δυσκολεύει σε μεγάλο βαθμό την επίλυσή του. Από τις εργασίες της βιβλιογραφίας, η [Late00] χρησιμοποιεί την ίδια βάση εικόνων και αποδίδει ποσοστό επιτυχίας μεταξύ 60% και 75%. Όμως, η μεθοδολογία αξιολόγησης είναι εντελώς διαφορετική, με αυτήν της παρούσης εργασίας να κρίνεται πολύ πιο ‘αυστηρή’. Επίσης, στη [Late00] υλοποιούνται πιο πολύπλοκες και εξειδικευμένες τεχνικές –σε σχέση με τους περιγραφείς EFD– για την παραγωγή του διανύσματος χαρακτηριστικών, όπως οι ροπές Zernike και οι περιγραφείς κυματιδίων. Αυτό οδηγεί σε καλύτερα αποτελέσματα με κόστος την ανάγκη μεγαλύτερης υπολογιστικής ισχύς.

Δεύτερη βάση εικόνων είναι η SSID [Ssid] και είναι διαθέσιμη στο ερευνητικό –και όχι μόνο– κοινό μέσω του Διαδικτύου. Απαρτίζεται από 99 εικόνες, που διαχωρίζονται ισόποσα σε εννέα διακριτές κλάσεις. Θα πρέπει να τονίσουμε ότι πρόκειται για εικόνες, που έχουν υποστεί τεχνητή αλλοίωση

(εισαγωγή θορύβου, αποκοπή περιοχών), με σκοπό να δυσκολέψουν το έργο των εφαρμογών ανάκλησης και να προσομοιώσουν καταστάσεις που αντιμετωπίζονται στον πραγματικό κόσμο. Η προτεινόμενη μεθοδολογία απέδωσε 33.3% ποσοστό επιτυχίας, τιμή που συγκριτικά με την αντίστοιχη της βιβλιογραφίας (87% στην [Seba01]) φαίνεται σημαντικά μικρότερη. Αναφέροντας όμως τη διαφορετική τεχνική αξιολόγησης (leave-one-out στην [Seba01] και 10-fold cross validation στην παρούσα), όπως επίσης και τη χρήση πιο πολύπλοκων τεχνικών για την περιγραφή ενός σχήματος, καταλαβαίνουμε ότι δεν μπορούμε να συγκρίνουμε τις δύο τιμές μεταξύ τους.

Η επόμενη βάση εικόνων αποτελείται από ένα υποσύνολο της βάσης MPEG που αναφέρθηκε παραπάνω. Πιο αναλυτικά, η τρίτη βάση διαθέτει 216 εικόνες, δηλαδή 18 κλάσεις με 12 εικόνες η κάθε μία. Το ποσοστό επιτυχίας προσεγγίζεται στο 38.4%, υψηλότερο από ότι ανωτέρω, παρόλο που το πρόβλημα είναι δυσκολότερο εξαιτίας των περισσότερων κλάσεων. Αυτό συμβαίνει, καθώς στις εικόνες της τρίτης βάσης δεν εντοπίζεται θόρυβος.

Η τέταρτη βάση αναπτύχθηκε και χρησιμοποιήθηκε για την επικύρωση των αποτελεσμάτων της παρούσης εργασίας. Διαθέτει 38 εικόνες, που ανήκουν σε τρεις διακριτές κλάσεις. Κάθε κλάση αντιστοιχεί σε ένα γεωμετρικό σχήμα εκ των κύκλου, τριγώνου και τετραγώνου. Οι εικόνες κάθε κλάσης προκύπτουν από μία αρχική εικόνα που έχει υποστεί τους τρεις βασικούς μετασχηματισμούς: μετατόπιση, περιστροφή και αλλαγή κλίμακας. Σκοπός της βάσης αυτής ήταν να αποδείξει ότι οι συντελεστές EFD είναι αναλλοίωτοι στους τρεις αυτούς μετασχηματισμούς, γεγονός που επαληθεύτηκε, εφόσον το ποσοστό επιτυχίας είναι 100%.

Η τελευταία βάση εικόνων δημιουργήθηκε στα πλαίσια της εργασίας αυτής και αφορά εικόνες από συγκεκριμένη εφαρμογή του ιατρικού τομέα. Πιο συγκεκριμένα, συγκεντρώθηκε υλικό από σχετική ιστοσελίδα [cars], όπου παρουσιάζονται ιατρικές εικόνες περιστατικών. Επιλέξαμε το πρόβλημα του έλκους και εντοπίσαμε εικόνες που παρουσιάζουν έλκος. Η αντίστοιχη παθολογία είναι ταξινομημένη σε τρεις διακριτές κλάσεις από ειδικούς ιατρούς, ενώ με χρήση ειδικού εργαλείου [Iako11] πραγματοποιήθηκε η χειροκίνητη κατάτμηση της περιοχής ενδιαφέροντος. Καταλήξαμε σε ένα πλήθος 30 εικόνων που κατηγοριοποιούνται σε τρεις διακριτές κλάσεις (Quiescent, Active Moderate και Active Severe). Το ποσοστό επιτυχίας της μεθοδολογίας μας είναι 63.3%, ενώ προσπαθώντας να κατανοήσουμε το αποτέλεσμα αυτό, προχωρούμε παρακάτω και ταξινομούμε τα δεδομένα ως προς τις δύο από τις τρεις κλάσεις κάθε φορά. Προκύπτουν τρία διαφορετικά πειράματα Quiescent VS Active Moderate, Quiescent VS Active Severe και Active Moderate VS Active Severe, με τα αποτελέσματα να κυμαίνονται στα ποσοστά 85%, 65% και 85% αντίστοιχα. Άρα, συμπεραίνουμε ότι οι περιγραφείς EFD αδυνατούν να περιγράψουν τη διαφορά μεταξύ Quiescent και Active Severe περιστατικών, ενώ στα άλλα δύο πειράματα επιτυγχάνουν ικανοποιητικά αποτελέσματα.

Όσον αφορά την αξιολόγηση των ταξινομητών, διακρίνεται μία ελαφριά υπεροχή των τεχνικών Bayes έναντι του ταξινομητή των  $k$  πλησιέστερων γειτόνων.

Συμπερασματικά, αναφέρουμε ότι επιλέχθηκαν:

- Απλοί περιγραφείς EFD, με σκοπό την κατανόηση της φυσικής τους σημασίας, αλλά και την χαμηλή υπολογιστική πολυπλοκότητα

- Αξιόπιστη μεθοδολογία 10-fold cross validation επικύρωσης των αποτελεσμάτων
- Ταξινομητής k πλησιέστερων γειτόνων και μεθοδολογίες Bayes.

Ως μελλοντική έρευνα προτείνουμε τον εμπλουτισμό των χαρακτηριστικών με νέες μετρικές, ικανές να εντοπίσουν και να διακρίνουν τις περιπτώσεις που οι EFD περιγραφείς αδυνατούν να ανταπεξέλθουν. Σημαντική όμως προσοχή θα πρέπει να δοθεί στην υπολογιστική πολυπλοκότητα των νέων αυτών χαρακτηριστικών, καθώς θέλουμε η συγκεκριμένη μεθοδολογία να εφαρμοστεί σε συστήματα ανάκλησης εικόνων πραγματικού χρόνου που οφείλουν να χρησιμοποιούν χαμηλής πολυπλοκότητας αλγόριθμους. Επιπλέον, βελτιωμένα αποτελέσματα αποδίδουν και πιο προηγμένες τεχνικές ταξινόμησης, όπως τα μοντέλα διανυσματικής μάθησης SVM ή τα νευρωνικά δίκτυα. Άρα, θα μπορούσαν να υλοποιηθούν αντίστοιχοι ταξινομητές, με σκοπό την επίτευξη υψηλότερων ποσοστών επιτυχίας.

## **Βιβλιογραφία**

- [Abba00] Abbasi, S.; Mokhtarian, F. & Kittler, J., Enhancing CSS-based shape retrieval for objects with shallow concavities, *Image and Vision Computing*, Elsevier, 2000, 18, 199-211
- [Abo88] Abo-Zaid, A.; Hinton, O. & Horne, E., About moment normalization and complex moment descriptors, *Pattern Recognition*, Springer, 1988, 399-409
- [Alaj06] Alajlan, N.; Kamel, M. S. & Freeman, G., Multi-object image retrieval based on shape and topology, *Signal Processing: Image Communication*, Elsevier, 2006, 21, 904-918
- [Alaj07] Alajlan, N.; El Rube, I.; Kamel, M. S. & Freeman, G., Shape retrieval using triangle-area representation and dynamic space warping, *Pattern Recognition*, Elsevier, 2007, 40, 1911-1920
- [Ang95] Ang, Y.; Li, Z. & Ong, S., Image retrieval based on multidimensional feature properties, *Proceedings of SPIE*, 1995, 2420, 4
- [Anil99] Anil K. Jain & Robert P.W., Jianchang Mao, *Statistical Pattern Recognition: A Review*, 1999
- [Arbt89] Arbter, K., *Affine-invariant Fourier descriptors, From Pixels to Features*, Elsevier Science Publishers, 1989, 153-164
- [Arbt90] Arbter, K.; Snyder, W.; Burkhardt, H. & Hirzinger, G., Application of Affine-Invariant Fourier Descriptors to Recognition of 3-D Objects, *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, Published by the IEEE Computer Society, 1990, 640-647
- [Beni98] Benitez, A.; Beigi, M. & Chang, S., Using relevance feedback in content-based image metasearch, *IEEE Internet Computing*, IEEE Educational Activities Department, 1998, 2, 59-69.
- [Berr00] Berretti, S.; Del Bimbo, A. & Pala, P., Retrieval by shape similarity with perceptual distance and effective indexing, *Multimedia*, IEEE Transactions on, IEEE, 2000, 2, 225-239
- [Bha11] Bhagat, A. P. & Atique, M., Medical image retrieval, indexing and enhancement techniques: a survey, *Proceedings of the 2011 International Conference on Communication, Computing & Security*, 2011, 387-390
- [Bimb97] Bimbo, A.; Mugnaini, M.; Pala, P. & Turco, F., Picasso: Visual querying by color perceptive regions, *Proceedings of the 2nd International Conference on Visual Information Systems*, San Diego, 1997, 125-131.
- [Bimb97a] Del Bimbo, A. & Pala, P., Visual image retrieval by elastic matching of user sketches, *Pattern Analysis and Machine Intelligence*, IEEE Transactions on, IEEE, 1997, 19, 121-132
- [Blum67] Blum, H. & others, A transformation for extracting new descriptors of shape, *Models for the perception of speech and visual form*, 1967, 19, 362-380
- [Bril68] Brill, E. L., Character recognition via Fourier descriptors, *WESCON*, session, 1968, 25, 1-10
- [Bui99] Bui, T. & Chen, G., Invariant Fourier-wavelet descriptor for pattern recognition, *Pattern recognition*, Citeseer, 1999, 32, 1083-1088
- [Bunt94] Buntine, W., Operations for learning with graphical models, *arXiv preprint cs/9412102*, 199

[Cars99] Carson, C.; Thomas, M.; Belongie, S.; Hellerstein, J. & Malik, J., Blobworld: A system for region-based image indexing and retrieval, Visual Information and Information Systems, 1999, 660-660.

[caps] <http://www.capsuleendoscopy.org/Image-Atlas/Pages/default.aspx>

[Cele05] Celebi, M. & Aslandogan, Y., A comparative study of three moment-based shape descriptors, Information Technology: Coding and Computing, 2005. ITCC 2005. International Conference on, 2005, 1, 788-793

[Cerr03] Ceruzzi, P., A history of modern computing, The MIT press, 2003.

[Chat10] S. A. Chatzichristofis, K. Zagoris, Y. S. Boutalis and Nikos Papamarkos, "ACCURATE IMAGE RETRIEVAL BASED ON COMPACT COMPOSITE DESCRIPTORS AND RELEVANCE FEEDBACK INFORMATION", «International Journal of Pattern Recognition and Artificial Intelligence (IJPRAI) », Vol. 24, No. 2, 207 – 244, 2010.

[Chan80] Chang, N. & Fu, K., Query-by-pictorial-example, IEEE Transactions on Software Engineering, 1980, 6, 519-524.

[Chel84] Chellappa, R. & Bagdazian, R., Fourier coding of image boundaries, Pattern Analysis and Machine Intelligence, IEEE Transactions on, IEEE, 1984, 102-105

[Chen12] Chen, Y. & Lee, J., A Review of Machine-Vision-Based Analysis of Wireless Capsule Endoscopy Video, Diagnostic and Therapeutic Endoscopy, Hindawi Publishing Corporation, 2012, 2012

[Chon04] Chong, C.-W.; Raveendran, P. & Mukundan, R., Translation and scale invariants of Legendre moments, Pattern Recognition, Elsevier, 2004, 37, 119-129

[Chua96] Chuang, G. & Kuo, C., Wavelet descriptor of planar curves: Theory and applications, Image Processing, IEEE Transactions on, IEEE, 1996, 5, 56-70 .

[Cogg82] Coggins, J. M., "A Framework for Texture Analysis Based on Spatial Filtering," Ph.D. Thesis, Computer Science Department, Michigan State University, East Lansing, Michigan, 1982

[Coom82] D. Coomans; D.L. Massart (1982). "Alternative k-nearest neighbour rules in supervised pattern recognition : Part 1. k-Nearest neighbour classification by using alternative voting rules".

[Cort95] Cortes, Corinna; and Vapnik, Vladimir N.; "Support-Vector Networks", Machine Learning, 20, 1995. <http://www.springerlink.com/content/k238jx04hm87j80g/>

[Cox02] Cox, I.; Miller, M.; Minka, T.; Papathomas, T. & Yianilos, P., The Bayesian image retrieval system, PicHunter: theory, implementation, and psychophysical experiments, Image Processing, IEEE Transactions on, IEEE, 2002, 9, 20-37.

[Crim82] Crimmins, T., A complete set of Fourier descriptors for 2-D shapes, IEEE Trans. Syst. Man Cybern, 1982, 12, 170-179

[Das90] Das, M.; Paulik, M. J. & Loh, N., A bivariate autoregressive technique for analysis and classification of planar shapes, Pattern Analysis and Machine Intelligence, IEEE Transactions on, IEEE, 1990, 12, 97-103

- [Datt05] Datta, R.; Li, J. & Wang, J., Content-based image retrieval: approaches and trends of the new age, Proceedings of the 7th ACM SIGMM international workshop on Multimedia information retrieval, November, 2005, 10-11
- [Daub90] Daubechies, I., The wavelet transform, time-frequency localization and signal analysis, IEEE Trans. Inf. Theory, 1990, 36, 961-1005
- [Daug98] Daugman, J., Complete discrete 2-D Gabor transforms by neural networks for image analysis and compression, Acoustics, Speech and Signal Processing, IEEE Transactions on, IEEE, 1998, 36, 1169-1179
- [Davi04] Davies, E. R., Machine vision: theory, algorithms, practicalities, Morgan Kaufmann, 2004
- [Deza09] Elena Deza & Michel Marie Deza (2009) Encyclopedia of Distances, page 94, Springer
- [Doi07] Doi, K., Computer-aided diagnosis in medical imaging: historical review, current status and future potential, Computerized medical imaging and graphics: the official journal of the Computerized Medical Imaging Society, NIH Public Access, 2007, 31, 198
- [Dubo86] Dubois, S. R. & Glanz, F. H., An autoregressive model approach to two-dimensional shape classification, Pattern Analysis and Machine Intelligence, IEEE Transactions on, IEEE, 1986, 55-66
- [Duda73] Richard Duda, Peter Hart (1973). Pattern Classification and Scene Analysis. Wiley, New York.
- [Dude97] Dudek, G. & Tsotsos, J. K., Shape representation and recognition from multiscale curvature, Computer Vision and Image Understanding, Elsevier, 1997, 68, 170-189
- [EFDJAVA]<https://java2s.com/Open-Source/Java/Image/objectj/ij/plugin/filter/EllipticFD.java.htm>
- [EMP1] <http://academics.epu.ntua.gr/LinkClick.aspx?fileticket=TEutWP3krj0%3D&...>
- [ERNT03] <http://infolab.cs.unipi.gr/people/ntoutsis/papers/master.pdf>
- [Engl92] Engle Jr, R., Attempts to use computers as diagnostic aids in medical decision making: a thirty-year experience., Perspectives in biology and medicine, 1992, 35, 207
- [Fish36] Fisher R.A. (1936) " The use of multiple measurements in taxonomic problems", Annals of Eugenics, 7, 179–188
- [Flus92] Flusser, J., Invariant shape description and measure of object similarity, Image Processing and its Applications, 1992., International Conference on, 1992, 139-142.
- [Fole00] Foley, J., Computer graphics: principles and practice, Addison-Wesley Professional, 1990
- [Free61] Freeman, H., On the encoding of arbitrary geometric configurations, Electronic Computers, IRE Transactions on, IEEE, 1961, 260-268
- [Fu76] Fu, K.-S. & Aizerman, M. A., Syntactic methods in pattern recognition, Systems, Man and Cybernetics, IEEE Transactions on, IEEE, 1976, 590-591
- [Galo99] Galousi, K.; Karkanis, S. & Maroulis, D., Classification of Endoscopic Images Based on Texture Spectrum., Proceedings of Workshop on Machine Learning in Medical Applications, Advance Course in Artificial Intelligence-ACAI99, Chania, Greece, 1999, 63-69

- [Geor95] George H. John, Pat Langley: Estimating Continuous Distributions in Bayesian Classifiers. In: Eleventh Conference on Uncertainty in Artificial Intelligence, San Mateo, 338-345, 1995.
- [Geve00] Gevers, T. & Smeulders, A., Pictoseek: Combining color and shape invariant features for image retrieval, *Image Processing, IEEE Transactions on*, IEEE, 2000, 9, 102-119.
- [Geve04] Gevers, T. & Smeulders, A., *Image search engines: An overview*, Emerging Topics in Computer Vision. Prentice Hall, Citeseer, 2004.
- [Goln07] Golnabi, H. & Asadpour, A., Design and application of industrial machine vision systems, *Robotics and Computer-Integrated Manufacturing*, Elsevier, 2007, 23, 630-637.
- [Gonz07] Gonzalez, R. & Woods, R., *Digital Image Processing*, Prentice Hall, 2007.
- [Gosh85] Goshtasby, A., Description and discrimination of planar shapes using shape matrices, *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, IEEE, 1985, 738-743.
- [Gran72] Granlund, G., Fourier preprocessing for hand print character recognition, *Computers, IEEE Transactions on*, IEEE, 1972, 100, 195-201
- [Gros90] Grosky, W. I. & Mehrotra, R., Index-based object recognition in pictorial data management, *Computer Vision, Graphics, and Image Processing*, Academic Press, 1990, 52, 416-436
- [Gros92] Grosky, W. I.; Neo, P. & Mehrotra, R., A pictorial index mechanism for model-based matching, *Data & knowledge engineering*, Elsevier, 1992, 8, 309-327
- [Hahn01] Ritchie MD, Hahn LW, Roodi N, Bailey LR, Dupont WD, Parl FF, Moore JH. Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. *Am J Hum Genet.* 2001 Jul;69(1):138-47
- [Hafn10] Hafner, M.; Gangl, A.; Liedlgruber, M.; Uhl, A.; Vécsei, A. & Wrba, F., Endoscopic image classification using edge-based features, *Pattern Recognition (ICPR), 2010 20th International Conference on*, 2010, 2724-2727
- [Hank97] Hanka, R. & Harte, T., Curse of dimensionality: Classifying large multi-dimensional images with neural networks, *Computer-Intensive Methods in Control and Signal Processing*, K. Warwick and M. K arn y (Ed. s), Birkhauser Boston, New York, 1997, 249-260
- [Hara73] Haralick, R.; Shanmugam, K. & Dinstein, I., Textural features for image classification, *IEEE Transactions on systems, man and cybernetics*, 1973, 3, 610-621
- [Hara79] Haralick, R., Statistical and structural approaches to texture, *Proceedings of the IEEE*, IEEE, 1979, 67, 786-804
- [Hint08] Transformation Auto-encoders, Department of Computer Science, University of Toronto, 2008
- [Hlav93] Hlavac, V.; Sonka, M. & Boyle, R., *Image processing, analysis and machine vision*, Brooks/Cole Thompson, 1993, 2
- [Hsu95] Hsu, H., *Theory and Problems of Signals and Systems*, McGraw-Hill, 1995



- [Hu62] Hu, M., Visual pattern recognition by moment invariants, *Information Theory, IRE Transactions on*, IEEE, 1962, 8, 179-187
- [Huan97] Huang, J.; Kumar, S.; Mitra, M.; Zhu, W. & Zabih, R., Image Indexing Using Color Correlograms, *Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition (CVPR'97)*, 1997, 762
- [Huan98] Huang, C.-L. & Huang, D.-H., A content-based image retrieval system, *Image and Vision Computing, Elsevier*, 1998, 16, 149-163
- [Iako11] D.K. Iakovidis, and C.V Smailis, "Efficient Semantically-Aware Annotation of Images", in *Proc. IEEE International Conference of Imaging Systems and Techniques (IST)*, pp. 146-149, Penang, Malaysia, May 2011, doi: 10.1109/IST.2011.5962190
- [Jain91] Jain, A. & Farrokhnia, F., Unsupervised texture segmentation using Gabor filters, *Pattern recognition, Elsevier*, 1991, 24, 1167-1186
- [Jain98] Jain, A. & Vailaya, A., Shape-based retrieval: A case study with trademark image databases, *Pattern recognition, Citeseer*, 1998, 31, 1369-1390
- [Jalb06] Jalba, A. C.; Wilkinson, M. H. & Roerdink, J. B., Shape representation and recognition through morphological curvature scale spaces, *Image Processing, IEEE Transactions on*, IEEE, 2006, 15, 331-341
- [Jean00] Jeannin, S., ISO/IEC JTC1/SC29/WG11/N3321: MPEG-7 visual part of experimentation model version 5.0, Nordwijkerhout, March, 2000
- [Joll02] Jolliffe I.T. *Principal Component Analysis*, Series: Springer Series in Statistics, 2nd ed., Springer, NY, 2002, XXIX, 487 p. 28 illus. ISBN 978-0-387-95442-4
- [Imag] [http://www.imageprocessingplace.com/root\\_files\\_V3/image\\_databases.htm](http://www.imageprocessingplace.com/root_files_V3/image_databases.htm)
- [Jain89] Jain, A., *Fundamentals of digital image processing*, Prentice-Hall, Inc. Upper Saddle River, NJ, USA, 1989
- [Kak88] Kak, A. C. & Slaney, M., *Principles of computerized tomographic imaging*, IEEE Press, 1988
- [Kark03] Karkanis, S. A.; Iakovidis, D. K.; Maroulis, D. E.; Karras, D. A. & Tzivras, M., Computer-aided tumor detection in endoscopic video using color wavelet features, *Information Technology in Biomedicine, IEEE Transactions on*, IEEE, 2003, 7, 141-152
- [Kash81] Kashyap, R. & Chellappa, R., Stochastic models for closed boundary analysis: Representation and reconstruction, *Information Theory, IEEE Transactions on*, IEEE, 1981, 27, 627-637
- [Kaup95] Kauppinen, H.; Seppanen, T. & Pietikainen, M., An experimental comparison of autoregressive and Fourier-based descriptors in 2D shape classification, *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, IEEE, 1995, 17, 201-207
- [Khal01] Khalil, M. I. & Bayoumi, M. M., A dyadic wavelet affine invariant function for 2D shape recognition, *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, IEEE, 2001, 23, 1152-1164
- [Khot90] Khotanzad, A. & Hong, Y., *Invariant image recognition by Zernike moments*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Published by the IEEE Computer Society, 1990, 489-497

- [Krzy89] Krzyzak, A.; Leung, S. & Suen, C., Reconstruction of two-dimensional patterns from Fourier descriptors, *Machine vision and applications*, Springer, 1989, 2, 123-140
- [Laak00] Laaksonen, J.; Koskela, M.; Laakso, S. & Oja, E., PicSOM-content-based image retrieval with self-organizing maps, *Pattern Recognition Letters*, Elsevier, 2000, 21, 1199-1207.
- [Late00] L. J. Latecki, R. Lakaemper, and U. Eckhardt: Shape Descriptors for Non-rigid Shapes with a Single Closed Contour. Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Hilton Head Island, South Carolina, pp. 424-429, June 2000
- [Li98] Li, S. Z., Shape matching based on invariants, 1998
- [Liao96] Liao, S. X. & Pawlak, M., On image analysis by moments, *Pattern analysis and machine intelligence*, IEEE Transactions on, IEEE, 1996, 18, 254-266.
- [Lied11] Liedlgruber, M. & Uhl, A., Computer-aided decision support systems for endoscopy in the gastrointestinal tract: a review, *Biomedical Engineering, IEEE Reviews in*, IEEE, 2011, 4, 73-88
- [Lin87] Lin, C. & Chellappa, R., Classification of partial 2-D shapes using Fourier descriptors, *Pattern Analysis and Machine Intelligence*, IEEE Transactions on, IEEE, 1987, 686-690
- [Lin87a] Lin, C.-S. & Hwang, C.-L., New forms of shape invariants from elliptic Fourier descriptors, *Pattern recognition*, Elsevier, 1987, 20, 535-545
- [Liu96] Liu, F. & Picard, R., Periodicity, directionality, and randomness: Wold features for image modeling and retrieval, *Pattern Analysis and Machine Intelligence*, IEEE Transactions on, IEEE, 1996, 18, 722-733
- [Lodw63] Lodwick, G.; Haun, C.; Smith, W.; Keller, R. & Robertson, E., Computer diagnosis of primary bone tumors: a preliminary report, *Radiology*, 1963, 80, 273-275
- [Lonc98] Loncaric, S., A survey of shape analysis techniques, *Pattern recognition*, Elsevier, 1998, 31, 983-1001
- [Lu02] Lu, K.-J. & Kota, S., Compliant mechanism synthesis for shape-change applications: Preliminary results, *Proceedings of SPIE Modeling, Signal Processing, and Control Conference*, 2002, 4693, 161-172
- [Lu99] Lu, G. & Sajjanhar, A., Region-based shape representation and similarity measure suitable for content-based image retrieval, *Multimedia Systems*, Springer, 1999, 7, 165-174.
- [Mall89] Mallat, S., A theory for multiresolution signal decomposition: The wavelet representation, *IEEE transactions on pattern analysis and machine intelligence*, 1989, 11, 674-693
- [Malt09] Maltoni, D.; Maio, D.; Jain, A. & Prabhakar, S., *Handbook of fingerprint recognition*, Springer-Verlag New York Inc, 2009.
- [Math98] Mathias, E. & Conci, A., Comparing the influence of color spaces and metrics in content-based image retrieval, *Computer Graphics, Image Processing, and Vision*, 1998. Proceedings. SIBGRAP'98. International Symposium on, 1998, 371-378
- [McFa72] McFarlane, M., Digital Pictures 50 Years ago, *Proc. IEEE*, 1972, 60, 768-770.

- [Mehr95] Mehrotra, R. & Gary, J. E., Similar-shape retrieval in shape data management, *Computer, IEEE*, 1995, 28, 57-62
- [Meht97] Mehtre, B. M.; Kankanhalli, M. S. & Lee, W. F., Shape measures for content based image retrieval: a comparison, *Information Processing & Management, Elsevier*, 1997, 33, 319-337.
- [Mitic84] Mitchell, O. & Grogan, T. A., Global and partial shape discrimination for computer vision, *Optical Engineering, International Society for Optics and Photonics*, 1984, 23, 235484-235484
- [Mokh88] Mokhtarian, F., A theory of multiscale, torsion-based shape representation for space curves, *Computer Vision and Image Understanding, Elsevier*, 1997, 68, 1-17
- [Mokh92] Mokhtarian, F. & Mackworth, A. K., A theory of multiscale, curvature-based shape representation for planar curves, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1992, 14, 789-805
- [Mors94] Morse, B., Computation of object cores from greyscale images, Ph. D. dissertation, Dept. of Computer Science, Univ. of North Carolina, 1994
- [Mpeg] <http://mpeg.chiariglione.org/standards/mpeg-7/mpeg-7.htm>
- [Muku04] Mukundan, R., A new class of rotational invariants using discrete orthogonal moments, University of Canterbury. *Computer Science and Software Engineering.*, 2004
- [Mull04] Müller, H.; Michoux, N.; Bandon, D. & Geissbuhler, A., A review of content-based image retrieval systems in medical applications--clinical benefits and future directions, *International journal of medical informatics, Elsevier*, 2004, 73, 1-23
- [Nast98] Nastar, C.; Mitschke, M.; Meilhac, C. & Boujemaa, N., Surfimage: a flexible content-based image retrieval system, *Proceedings of the sixth ACM international conference on Multimedia*, 1998, 339-344.
- [Nibl93] Niblack, C.; Barber, R.; Equitz, W.; Flickner, M.; Glasman, E.; Petkovic, D.; Yanker, P.; Faloutsos, C. & Taubin, G., The QBIC project: querying images by content, using color, texture, and shape, *Proceedings of SPIE*, 1993
- [Nixo08] Nixon, M. & Aguado, A., *Feature extraction and image processing*, Academic Press, 2008
- [Oppe83] Oppenheim, A.; Willsky, A. & Nawab, S., *Signals and systems*, Prentice-Hall Englewood Cliffs, NJ, 1983, 2
- [Orte97] Ortega, M.; Rui, Y.; Chakrabarti, K.; Mehrotra, S. & Huang, T., Supporting similarity queries in MARS, *Proceedings of the fifth ACM international conference on Multimedia*, 1997, 403-413.
- [Otto91] Van Otterloo, P. J., *A contour-oriented approach to shape analysis*, Prentice Hall International (UK) Ltd., 1991
- [Pass99] Pass, G. & Zabih, R., Comparing images using joint histograms, *Multimedia Systems, Springer*, 1999, 7, 234-240

- [Pear88] J. Pearl, Probabilistic Reasoning in Intelligent Systems, Morgan Kaufmann, San Francisco, California, 1988
- [Peng06] Peng, J.; Yang, W. & Cao, Z., A symbolic representation for shape retrieval in curvature scale space, Computational Intelligence for Modelling, Control and Automation, 2006 and International Conference on Intelligent Agents, Web Technologies and Internet Commerce, International Conference on, 2006, 45-245
- [Pent96] Pentland, A.; Picard, R. & Sclaroff, S., Photobook: Content-based manipulation of image databases, International Journal of Computer Vision, Springer, 1996, 18, 233-254.
- [Pers77] Persoon, E. & Fu, K.-S., Shape discrimination using Fourier descriptors, Systems, Man and Cybernetics, IEEE Transactions on, IEEE, 1977, 7, 170-179
- [Petr10] Petrou, M. & Petrou, C., Image processing: the fundamentals, Wiley, 2010.
- [Peur97] Peura, M. & Iivari, J., Efficiency of simple shape descriptors, Proceedings of the third international workshop on visual form, 1997, 443-451
- [Phel85] Phelps, M.; Mazziotta, J. & Schelbert, H., Positron emission tomography and autoradiography: principles and applications for the brain and heart, Raven Press Pub., New York, NY, 1985
- [Plat11] Plataniotis, K.N.; Venetsanopoulos, A.N. (2011). "A Survey of Multilinear Subspace Learning for Tensor Data"
- [Prok92] Prokop, R. J. & Reeves, A. P., A survey of moment-based techniques for unoccluded object representation and recognition, CVGIP: Graphical Models and Image Processing, Elsevier, 1992, 54, 438-460
- [Raub94] Rauber, T. W., Two-dimensional shape description, Technical Report: GR UNINOVA-RT-10-94, Universidade Nova de Lisboa, 1994
- [Rich74] Richard, C. W. & Hemami, H., Identification of three-dimensional objects using Fourier descriptors of the boundary curve, Systems, Man and Cybernetics, IEEE Transactions on, IEEE, 1974, 371-378
- [Rose57] Rosenblatt, Frank (1957), The Perceptron--a perceiving and recognizing automaton. Report 85-460-1, Cornell Aeronautical Laboratory.
- [Rui97] Rui, Y.; She, A. C. & Huang, T. S., A modified Fourier descriptor for shape matching in MARS, SERIES ON SOFTWARE ENGINEERING AND KNOWLEDGE ENGINEERING, WORLD SCIENTIFIC PUBLISHING, 1997, 8, 165-180
- [Sajj97] Sajjanhar, A., A technique for similarity retrieval of shapes, Monash University, 1997
- [Samm69] Sammon, J.W. Jr., "A nonlinear mapping for data structure analysis", IEEE Transactions on Computers, vol. C-18, no. 5, 1969, pp. 401-409
- [Saul04] Unsupervised learning of image manifolds by semidefinite programming K. Q. Weinberger and L. K. Saul (2004). In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR-04), Washington D.C.

- [Scas94] Scassellati, B. M.; Alexopoulos, S. & Flickner, M. D., Retrieving images by 2D shape: a comparison of computation methods with human perceptual judgments, IS&T/SPIE 1994 International Symposium on Electronic Imaging: Science and Technology, 1994, 2-14.
- [Scla97] Sclaroff, S.; Taycher, L. & La Cascia, M., Imagerover: A content-based image browser for the world wide web, Content-Based Access of Image and Video Libraries, 1997. Proceedings. IEEE Workshop on, 1997, 2-9
- [Seba01] Sebastian, T. B.; Klein, P. N. & Kimia, B. B., Recognition of Shapes by Editing Shock Graphs, ICCV, 2001
- [Seki92] Sekita, I.; Kurita, T. & Otsu, N., Complex autoregressive model for shape recognition, Pattern Analysis and Machine Intelligence, IEEE Transactions on, IEEE, 1992, 14, 489-496
- [Shap01] Shapiro, L. & Stockman, G., Computer Vision., Prentice Hall, 2001
- [Shut02] J. Shutler, 2002, "Complex Zernike Moments", Statistical Moments, Department of Electronics and Computer Science, University of Southampton, United Kingdom
- [Smeu00] Smeulders, A.; Worring, M.; Santini, S.; Gupta, A. & Jain, R., Content-based image retrieval at the end of the early years, IEEE Transactions on pattern analysis and machine intelligence, Citeseer, 2000, 22, 1349-1380
- [Smit95] Smith, J. & Chang, S., Automated image retrieval using color and texture., Columbia University, July 1995.
- [Smit97] Smith, J. & Chang, S., VisualSEEk: a fully automated content-based image query system, Proceedings of the fourth ACM international conference on Multimedia, 1997, 87-98.
- [Smit1997] Smith, J. & Chang, S., WebSEEk: a content-based image and video search engine for the world-wide web, IEEE Multimedia, 1997, 4, 12-20.
- [Soff97] Soffer, A. & Samet, H., Negative shape features for image databases consisting of geographic symbols, Proc. 3rd International Workshop on Visual Form, 1997
- [Soil99] Soille, P., Morphological Image Analysis: Principles and Applications, Springer-Verlag, 1999, pp. 173-174.
- [Squi00] Squire, D. M. & Caelli, T. M., Invariance Signatures: Characterizing contours by their departures from invariance, Computer Vision and Image Understanding, Elsevier, 2000, 77, 284-316
- [Ssid] <http://www.lems.brown.edu/vision/researchAreas/SIID/>
- [Stan00] [http://www.stanford.edu/class/cs124/lec/Maximum\\_Entropy\\_Classifiers.pdf](http://www.stanford.edu/class/cs124/lec/Maximum_Entropy_Classifiers.pdf)
- [Stri95] Stricker, M. & Orengo, M., Similarity of color images, Proc. SPIE Storage and Retrieval for Image and Video Databases, 1995, 2420, 381-392

- [Tamu78] Tamura, H.; Mori, S. & Yamawaki, T., Textural features corresponding to visual perception., IEEE Transactions on Systems, Man, & Cybernetics, Institute of Electrical & Electronics Engineers Inc, 1978
- [Taub91] Taubin, G. & Cooper, D., Recognition and positioning of rigid objects using algebraic moment invariants, San Diego,'91, San Diego, CA, 1991, 175-186
- [Taza89] Taza, A. & Suen, C. Y., Discrimination of planar shapes using shape matrices, Systems, Man and Cybernetics, IEEE Transactions on, IEEE, 1989, 19, 1281-1289.
- [Teag80] Teague, M. R., Image analysis via the general theory of moments, J. Opt. Soc. Am, 1980, 70, 920-930.
- [Teh88] Teh, C.-H. & Chin, R. T., On image analysis by the methods of moments, Pattern Analysis and Machine Intelligence, IEEE Transactions on, IEEE, 1988, 10, 496-513
- [Theo09] S. Theodoridis and K. Koutroumbas. Pattern Recognition. Academic Press, 4<sup>th</sup> edition, 2009.
- [Tien97] Tieng, Q. M. & Boles, W., Recognition of 2D object contours using the wavelet transform zero-crossing representation, Pattern Analysis and Machine Intelligence, IEEE Transactions on, IEEE, 1997, 19, 910-916
- [Tori73] Toriwaki, J.; Suenaga, Y.; Negoro, T. & Fukumura, T., Pattern recognition of chest X-ray images, Computer Graphics and Image Processing, Elsevier, 1973, 2, 252-271
- [Tsak06] P. Tsakalidis, university of Crete, "Pattern Recognition", 2006
- [Vapn06] Vapnik, V., Estimation of dependences based on empirical data, Springer-Verlag New York Inc, 2006
- [Vazi03] Μιχάλης Βαζιργιάννης, Μαρία Χαλκίδη, Εξόρυξη Γνώσης από Βάσεις Δεδομένων και τον Παγκόσμιο Ιστό, Εκδ. Gutenberg, 2003
- [Verl07] John A. Lee, Michel Verleysen, Nonlinear Dimensionality Reduction, Springer, 2007.
- [Velt00] Veltkamp, R. & Tanase, M., Content-based image retrieval systems: A survey, 2000
- [Velt99] Veltkamp, R. & Hagedoorn, M., State-of-the-art in shape matching, UU-CS, Utrecht University: Information and Computing Sciences, 1999
- [Wang01] Wang, J.; Li, J. & Wiederhold, G., SIMPLicity: Semantics-sensitive integrated matching for picture libraries, IEEE Transactions on pattern analysis and machine intelligence, Published by the IEEE Computer Society, 2001, 947-963.
- [WangP01] Wang, P.; Krishnan, S. M.; Kugean, C. & Tjoa, M., Classification of endoscopic images based on texture and neural network, Engineering in Medicine and Biology Society, 2001. Proceedings of the 23rd Annual International Conference of the IEEE, 2001, 4, 3691-3695

[Wang09] Xiang-Yang Wang, Jun-Feng Wu<sup>1</sup> and Hong-Ying Yang "Robust image retrieval based on color histogram of local feature regions" Springer Netherlands

[Wang97] Wang, J.; Wiederhold, G.; Firschein, O. & Wei, S., Wavelet-based image indexing techniques with partial sketch retrieval capability, Research and Technology Advances in Digital Libraries, 1997. ADL'97. Proceedings., IEEE International Forum on, 1997, 13-24.

[Wang99] Wang, Y.-P.; Lee, S. & Toraichi, K., Multiscale curvature-based shape representation using B-spline wavelets, Image Processing, IEEE Transactions on, IEEE, 1999, 8, 1586-1592

[Wehr88] Wehrli, F.; Shaw, D. & Kneeland, J., Biomedical magnetic resonance imaging, New York, NY (USA); VCH Publishers Inc., 1988

[Weka] <http://www.cs.waikato.ac.nz/ml/weka/>

[Wins67] Winsberg, F.; Elkin, M.; Macy, J.; Bordaz, V. & Weymouth, W., Detection of radiographic abnormalities in mammograms by means of optical scanning and computer analysis, Radiology, 1967, 89, 211-215

[Yaco09] Yacob, Y. M.; Sakim, H. A. M.; Baharudin, N.; Yeh, L. Y. & Isa, N. A. M., A survey on medical digital imaging of endoscopic gastritis, TENCON 2009-2009 IEEE Region 10 Conference, 2009, 1-6

[Yang07] Yang, M.; Kpalma, K. & Ronsin, J., Scale-controlled area difference shape descriptor, Electronic Imaging 2007, 2007, 650003-650003

[Yang08] Yang, M.; Kpalma, K.; Ronsin, J. & others, A survey of shape feature extraction techniques, 2008

[Yang98] Yang, H. S.; Lee, S. U. & Lee, K. M., Recognition of 2D object contours using starting-point-independent wavelet coefficient matching, Journal of Visual Communication and Image Representation, Elsevier, 1998, 9, 171-181

[Zahn72] Zahn, C. T. & Roskies, R. Z., Fourier descriptors for plane closed curves, Computers, IEEE Transactions on, IEEE, 1972, 100, 269-281

[Zhan01] Zhang, D. & Lu, G., A comparative study on shape retrieval using Fourier descriptors with different shape signatures, Proc. of international conference on intelligent multimedia and distance education (ICIMADE01), 2001, 1-9

[Zhan02] Zhang, D.; Lu, G. & others, A comparative study of Fourier descriptors for shape representation and retrieval, Proceedings of the Fifth Asian Conf. on Computer Vision, 2002, 646-651

[Zhang03] D. Zhang and G. Lu, A comparative study of curvature scale space and fourier descriptors for shape-based image retrieval, Visual Communication and Image Representation, vol. 14(1), 2003.

[Zhan04] Zhang, D. & Lu, G., Review of shape representation and description techniques, Pattern recognition, Elsevier, 2004, 37, 1-19

[Zhan07] Zhang, D. & Lu, G., Evaluation of MPEG-7 shape descriptors against other shape descriptors, Multimedia Systems, Springer, 2003, 9, 15-30

[Zhan09] Zhang, Harry. "The Optimality of Naive Bayes",2009

[Πήτα01] Πήτας, Ι., Ψηφιακή Επεξεργασία Εικόνας, 2001 .



