



ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ
ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΜΕ ΕΦΑΡΜΟΓΕΣ ΣΤΗ
ΒΙΟΙΑΤΡΙΚΗ

**ΑΝΑΚΛΗΣΗ ΕΙΚΟΝΩΝ ΒΑΣΕΙ ΠΕΡΙΕΧΟΜΕΝΟΥ, ΜΕ
ΕΦΑΡΜΟΓΗ ΣΤΗΝ ΥΠΟΛΟΓΙΣΤΙΚΑ ΥΠΟΒΟΗΘΟΥΜΕΝΗ
ΙΑΤΡΙΚΗ ΔΙΑΓΝΩΣΗ ΑΠΟ ΕΙΚΟΝΕΣ ΑΣΥΡΜΑΤΗΣ
ΕΝΔΟΣΚΟΠΙΚΗΣ ΚΑΨΟΥΛΑΣ**

Παναγιώτης Χρυσανθόπουλος

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

Υπεύθυνος

Ιακωβίδης Δημήτριος

Αναπληρωτής Καθηγητής

Λαμία, 25-05 έτος 2017



ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ
ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΜΕ ΕΦΑΡΜΟΓΕΣ ΣΤΗ ΒΙΟΙΑΤΡΙΚΗ

ΑΝΑΚΛΗΣΗ ΕΙΚΟΝΩΝ ΒΑΣΕΙ ΠΕΡΙΕΧΟΜΕΝΟΥ, ΜΕ ΕΦΑΡΜΟΓΗ ΣΤΗΝ
ΥΠΟΛΟΓΙΣΤΙΚΑ ΥΠΟΒΟΗΘΟΥΜΕΝΗ ΙΑΤΡΙΚΗ ΔΙΑΓΝΩΣΗ ΑΠΟ
ΕΙΚΟΝΕΣ ΑΣΥΡΜΑΤΗΣ ΕΝΔΟΣΚΟΠΙΚΗΣ ΚΑΨΟΥΛΑΣ

Παναγιώτης Χρυσανθόπουλος

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

Επιβλέπων

Ιακωβίδης Δημήτριος

Αναπληρωτής Καθηγητής

Λαμία, 25-05 έτος 2017

Με ατομική μου ευθύνη και γνωρίζοντας τις κυρώσεις ⁽¹⁾, που προβλέπονται από της διατάξεις της παρ. 6 του άρθρου 22 του Ν. 1599/1986, δηλώνω ότι:

1. Δεν παραθέτω κομμάτια βιβλίων ή άρθρων ή εργασιών άλλων αυτολεξεί **χωρίς να τα περικλείω σε εισαγωγικά** και χωρίς να αναφέρω το συγγραφέα, τη χρονολογία, τη σελίδα. Η αυτολεξεί παράθεση χωρίς εισαγωγικά χωρίς αναφορά στην πηγή, είναι λογοκλοπή. Πέραν της αυτολεξεί παράθεσης, λογοκλοπή θεωρείται και η παράφραση εδαφίων από έργα άλλων, συμπεριλαμβανομένων και έργων συμφοιτητών μου, καθώς και η παράθεση στοιχείων που άλλοι συνέλεξαν ή επεξεργάστηκαν, χωρίς αναφορά στην πηγή. Αναφέρω πάντοτε με πληρότητα την πηγή κάτω από τον πίνακα ή σχέδιο, όπως στα παραθέματα.
2. Δέχομαι ότι η αυτολεξεί **παράθεση χωρίς εισαγωγικά**, ακόμα κι αν συνοδεύεται από αναφορά στην πηγή σε κάποιο άλλο σημείο του κειμένου ή στο τέλος του, είναι αντιγραφή. Η αναφορά στην πηγή στο τέλος π.χ. μιας παραγράφου ή μιας σελίδας, δεν δικαιολογεί συρραφή εδαφίων έργου άλλου συγγραφέα, έστω και παραφρασμένων, και παρουσίασή τους ως δική μου εργασία.
3. Δέχομαι ότι υπάρχει επίσης περιορισμός στο μέγεθος και στη συχνότητα των παραθεμάτων που μπορώ να εντάξω στην εργασία μου εντός εισαγωγικών. Κάθε μεγάλο παράθεμα (π.χ. σε πίνακα ή πλαίσιο, κλπ), προϋποθέτει ειδικές ρυθμίσεις, και όταν δημοσιεύεται προϋποθέτει την άδεια του συγγραφέα ή του εκδότη. Το ίδιο και οι πίνακες και τα σχέδια
4. Δέχομαι όλες τις συνέπειες σε περίπτωση λογοκλοπής ή αντιγραφής.

Ημερομηνία:/...../20.....

Ο – Η Δηλ.

(Υπογραφή)

(1) «Όποιος εν γνώσει του δηλώνει ψευδή γεγονότα ή αρνείται ή αποκρύπτει τα αληθινά με έγγραφη υπεύθυνη δήλωση του άρθρου 8 παρ. 4 Ν. 1599/1986 τιμωρείται με φυλάκιση τουλάχιστον τριών μηνών. Εάν ο υπαίτιος αυτών των πράξεων σκόπευε να προσπορίσει στον εαυτόν του ή σε άλλον περιουσιακό όφελος βλάπτοντας τρίτον ή σκόπευε να βλάψει άλλον, τιμωρείται με κάθειρξη μέχρι 10 ετών.

**ΑΝΑΚΛΗΣΗ ΕΙΚΟΝΩΝ ΒΑΣΕΙ ΠΕΡΙΕΧΟΜΕΝΟΥ, ΜΕ ΕΦΑΡΜΟΓΗ ΣΤΗΝ
ΥΠΟΛΟΓΙΣΤΙΚΑ ΥΠΟΒΟΗΘΟΥΜΕΝΗ ΙΑΤΡΙΚΗ ΔΙΑΓΝΩΣΗ ΑΠΟ
ΕΙΚΟΝΕΣ ΑΣΥΡΜΑΤΗΣ ΕΝΔΟΣΚΟΠΙΚΗΣ ΚΑΨΟΥΛΑΣ**

Παναγιώτης Χρυσανθόπουλος

Τριμελής Επιτροπή:

Δημήτρης Ιακωβίδης, Αναπληρωτής Καθηγητής (επιβλέπων)

Βασίλης Πλαγιανάκος, Αναπληρωτής Καθηγητής

Κωνσταντίνος Δελήμπασης, Επίκουρος Καθηγητής

Περίληψη

Τα συστήματα ανάκλησης εικόνων με βάση το περιεχόμενο (Content Based Image Retrieval) έχουν ως στόχο τη διευκόλυνση της αναζήτησης εικόνων σε μεγάλες βάσεις δεδομένων, η αναζήτηση γίνεται μόνο με γνώμονα το οπτικό περιεχόμενο της εικόνας. Στον ιατρικό τομέα τα συστήματα ανάκλησης μπορούν να συνεισφέρουν στη διάγνωση, τη θεραπεία ακόμα και στην εκπαίδευση των ιατρών. Στη παρούσα πτυχιακή εργασία αναπτύχθηκε ένα σύστημα ανάκλησης εικόνων με βάση το περιεχόμενο με σκοπό την εφαρμογή του σε ενδοσκοπικές εικόνες, οι οποίες αποκτήθηκαν από ενδοσκοπική κάψουλα. Στα πλαίσια της εργασίας αναπτύχθηκε και παρουσιάζεται ένας νέος αλγόριθμος εξαγωγής τοπικών χαρακτηριστικών. Ο προτεινόμενος αλγόριθμος εξαγωγής χαρακτηριστικών εφαρμόζει τη μέθοδο τμηματοποίησης SLIC (Simple Linear Iterative Clustering) για το διαχωρισμό της εικόνας σε συστάδες εικονοστοιχείων (υπερεικονοστοιχεία). Μετά το διαχωρισμό σε συστάδες εφαρμόζεται η μέθοδος εξαγωγής σημείων ενδιαφέροντος SURF (Speed-Up Robust Features) για την εύρεση σημείων ενδιαφέροντος στην εικόνα. Τα τελικά εξαγόμενα οπτικά χαρακτηριστικά αναπαριστούνται με το μοντέλο Bag of Visual Words όπου οι οπτικές λέξεις προκύπτουν από τα υπερεικονοστοιχεία που περιέχουν σημεία ενδιαφέροντος SURF. Μετά τη διαδικασία της εξαγωγής χαρακτηριστικών, τα χαρακτηριστικά οργανώνονται σε ένα ευρετήριο για την ταχύτατη και αξιόπιστη ανάκληση με τη βοήθεια της μηχανής αναζήτησης Apache Lucene και της ανοικτού κώδικα βιβλιοθήκης LIRE. Η προτεινόμενη τεχνική εξαγωγής χαρακτηριστικών συγκρίνεται με ήδη υπάρχουσες τεχνικές με σκοπό την ανάδειξη της αποτελεσματικότητας της μεθόδου.

Περιεχόμενα

Κεφάλαιο 1	8
Εισαγωγή	8
1.1 Συστήματα ανάκλησης εικόνων με βάση το περιεχόμενο	8
1.2 Στόχοι της πτυχιακής εργασίας	9
1.3 Επιστημονική συνεισφορά	9
1.4 Δομή της πτυχιακής εργασίας	10
Κεφάλαιο 2	11
Αρχιτεκτονική ενός συστήματος CBIR	11
2.1 Εισαγωγή	11
2.2 Ανάκληση εικόνων βάση περιεχομένου	12
2.3 Εξαγωγή Χαρακτηριστικών	13
2.4 Σύγκριση Χαρακτηριστικών	14
2.5 Δημιουργία ευρετηρίου	14
2.6 Ανάδραση Συνάφειας	15
Κεφάλαιο 3	16
Ανασκόπηση συστημάτων ανάκλησης ιατρικών εικόνων	16
3.1 Γενικά συστήματα ανάκλησης εικόνων	16
3.2 Συστήματα ανάκλησης ιατρικών εικόνων	17
Κεφάλαιο 4	20
Μέθοδοι εξαγωγής χαρακτηριστικών εικόνας	20
4.1 Χρωματικοί χώροι	20
4.1.1 Χρωματικός χώρος RGB	20
4.1.2 Χρωματικός χώρος HSV και HSI	21
4.1.3 Χρωματικός χώρος XYZ	22
4.1.4 Χρωματικός χώρος CIE-L*a*b*	22
4.1.5 Ιστογράμματα	23
4.2 Χαρακτηριστικά υφής	23
4.2.1 Τοπικά δυαδικά πρότυπα	24
4.2.2 Μετασχηματισμός χαρακτηριστικών αμετάβλητης κλίμακας	24
4.2.3 Υπολογιστικά επιταχυμένα ανθεκτικά χαρακτηριστικά	27
4.3 Σάκος οπτικών λέξεων	28
Κεφάλαιο 5	31
Τμηματοποίηση εικόνας	31
5.1 Εισαγωγή	31
5.2 Μέθοδοι τμηματοποίησης εικόνας	31
5.3 Απλή Γραμμική Επαναληπτική Συσταδοποίηση	33
Κεφάλαιο 6	35
Ευρετήρια	35
6.1 Εισαγωγή	35

6.2 Σχισιακές Βάσεις Δεδομένων	35
6.3 Αντεστραμμένα Αρχεία (Inverted-File)	36
6.4 Δένδρα	38
6.4.1 Δυαδικά Δένδρα αναζήτησης	38
6.4.2 Β-δένδρα	39
6.4.3 Δένδρα Περιοχής	39
6.5 Μέθοδοι δημιουργίας ευρετηρίων σε συστήματα CBIR	40
Κεφάλαιο 7	41
Προτεινόμενο σύστημα ανάκλησης ενδοσκοπικών εικόνων βάση περιεχομένου	41
7.1 Εισαγωγή	41
7.2 Αρχιτεκτονική του συστήματος	42
7.3 Αλγόριθμος εξαγωγής χαρακτηριστικών	43
7.4 Δημιουργία Λεξιλογίου	46
7.5 Δημιουργία Ευρετηρίου	46
7.6 Σύγκριση Εικόνων	47
Κεφάλαιο 8	49
Πειράματα και Αποτελέσματα	49
8.1 Δεδομένα	49
8.2 Μετρικές απόδοσης	49
8.3 Πειράματα	50
8.3.1 Σύγκριση χαρακτηριστικών εξαγωγής με βάση το μέγεθος του λεξιλογίου	51
8.3.2 Καμπύλες ακρίβειας – ανάκλησης	56
8.3.3 Αξιολόγησης της ταχύτητας ανάκλησης εικόνων	57
Κεφάλαιο 9	61
Συμπεράσματα – Μελλοντική εργασία	61
9.1 Συμπεράσματα	61
9.2 Μελλοντική Εργασία	61
Αναφορές	63

Εισαγωγή

1.1 Συστήματα ανάκλησης εικόνων με βάση το περιεχόμενο

Η ανάπτυξη της ψηφιακής φωτογραφίας, του βίντεο, του διαδικτύου καθώς επίσης και η ραγδαία εξέλιξη των ιατρικών απεικονιστικών συστημάτων, έχει οδηγήσει στην αύξηση της διαθέσιμης οπτικής πληροφορίας. Επιπλέον, όλο και περισσότεροι οργανισμοί και φορείς όπως νοσοκομεία, μουσεία, υπηρεσίες δίωξης εγκλήματος ψηφιοποιούν τα δεδομένα τους και τα αποθηκεύουν σε μορφή εικόνων. Για την αξιοποίηση όλου του όγκου των δεδομένων απαιτούνται καινοτόμες τεχνικές αναζήτησης, οι μέθοδοι αναζήτησης βασισμένες σε λέξεις κλειδιά απαιτεί εκτενή περιγραφή των εικόνων που περιέχονται σε μια βάση δεδομένων, έργο ακριβό και χρονοβόρο, επιπροσθέτως μια εικόνα δεν μπορεί να περιγραφεί πλήρως από μια συλλογή από λέξεις κλειδιά, καθιστώντας τέτοιου είδους μηχανές αναζήτησης αναποτελεσματικές.

Η ανάκληση των εικόνων με βάση το περιεχόμενο είναι η διαδικασία αναζήτησης εικόνων σε μια βάση δεδομένων με βάση το οπτικό περιεχόμενο της εικόνας επερώτησης. Η αναζήτηση ξεκινά όταν ο χρήστης εισάγει στο σύστημα την εικόνα επερώτησης, έπειτα το σύστημα ανακτά από τη βάση δεδομένων παρόμοιες εικόνες.

Τα συστήματα ανάκλησης εικόνων με βάση το περιεχόμενο (Content Base Image Retrieval , CBIR) είναι η εφαρμογή της υπολογιστικής όρασης στο πρόβλημα της αναζήτησης ψηφιακών εικόνων σε μεγάλες βάσεις δεδομένων. Τα CBIR συστήματα είναι αντικείμενο εντατικής έρευνας για πάνω από μια δεκαετία, καθώς η μεγάλη ποικιλομορφία των εικόνων απαιτεί διαφορετικά συστήματα ανάκλησης εικόνων, η ανάγκη για διαφορετικά και πιο αποτελεσματικά CBIR συστήματα έχει οδηγήσει σε μια πληθώρα προκλήσεων. Οι προκλήσεις και το μεγάλο εύρος επιστημονικών πεδίων που μπορούν να χρησιμοποιήσουν συστήματα ανάκλησης με βάση το περιεχόμενο έχει προκαλέσει το ενδιαφέρον των επιστημόνων, με αποτέλεσμα στην αύξηση των επιστημονικών δημοσιεύσεων στο τομέα της ανάκλησης των εικόνων με βάση το περιεχόμενο.

Η ανάκληση ιατρικών εικόνων έχει σημαντικές εφαρμογές στη διάγνωση ιατρικών παθήσεων, στην εκπαίδευση ιατρών καθώς και στην έρευνα. Τα ιατρικά κέντρα, παράγουν και αποθηκεύουν ένα μεγάλο όγκο ιατρικών δεδομένων, για την βέλτιστη αξιοποίηση των δεδομένων είναι απαραίτητη η χρήση ενός συστήματος ανάκλησης. Ιατρικές εικόνες ασθενών που βρίσκονται αποθηκευμένες στις βάσεις δεδομένων των ιατρικών κέντρων, με ένα αξιόπιστο σύστημα ανάκλησης εικόνων, μπορούν να βοηθήσουν τον θεράποντα ιατρό στην καλύτερη διάγνωση, θεραπεία και συνταγογράφηση φαρμάκων.

1.2 Στόχοι της πτυχιακής εργασίας

Στα πλαίσια αυτής της ερευνητικής εργασίας είναι η διερεύνηση των συστημάτων ανάκλησης εικόνων με βάση το περιεχόμενο που έχουν προταθεί για την ανάκληση ιατρικών εικόνων με ιδιαίτερη έμφαση σε συστήματα που έχουν προταθεί για ανάκληση ενδοσκοπικών εικόνων. Κύριος στόχος της παρούσας πτυχιακής εργασίας είναι η κατασκευή ενός συστήματος ανάκλησης εικόνων με βάση το περιεχόμενο σε ενδοσκοπικά δεδομένα. Το σύστημα που δημιουργήθηκε αποσκοπεί στην αξιόπιστη ανάκληση ενδοσκοπικών εικόνων, για αυτό το λόγο εφαρμόζονται καινοτόμες τεχνικές εξαγωγής οπτικών χαρακτηριστικών και αξιόπιστες μεθοδολογίες για την δημιουργία του ευρετηρίου, με σκοπό την αποθήκευση, αναζήτηση και ανάκληση των εικόνων.

Για την εξαγωγή των οπτικών χαρακτηριστικών, που είναι ζωτικής σημασίας για την απόδοση του συστήματος, υιοθετούνται τεχνικές τμηματοποίησης εικόνας καθώς και μια αξιόπιστη τεχνική για την εξαγωγή οπτικών χαρακτηριστικών από τα τμηματοποιημένα μέρη της εικόνας, κατάλληλη για τις απαιτήσεις ενός συστήματος ανάκλησης ενδοσκοπικών εικόνων. Τα δεδομένα που χρησιμοποιήθηκαν για τη δημιουργία του συστήματος ανάκλησης έχουν αποκτηθεί από ενδοσκοπικές επεμβάσεις στις οποίες έχει χρησιμοποιηθεί ενδοσκοπική κάψουλα.

1.3 Επιστημονική συνεισφορά

Τα συστήματα ανάκλησης εικόνων με βάση το περιεχόμενο είναι υπεύθυνα για την αποθήκευση και αξιόπιστη αναζήτηση εικόνων σε μεγάλες βάσεις δεδομένων. Ένα σύστημα ανάκλησης δανείζεται τεχνικές από έναν μεγάλο αριθμό επιστημονικών πεδίων, διότι η κατασκευή ενός τέτοιου συστήματος απαιτεί αποδοτικές μεθόδους αποθήκευσης και αναζήτησης σε μεγάλες βάσεις δεδομένων αλλά και μεθόδους εξαγωγής οπτικών χαρακτηριστικών. Οι επιστημονικές συνεισφορές της εργασίας είναι οι εξής :

- Κατασκευή ενός ολοκληρωμένου συστήματος ανάκλησης ενδοσκοπικών εικόνων βάση περιεχομένου.
- Δημιουργία πρωτότυπης μεθοδολογίας για την εξαγωγή οπτικών χαρακτηριστικών.
- Διερεύνηση της πρωτότυπης μεθοδολογίας για την ανάκληση ενδοσκοπικών εικόνων.

Η κατασκευή του συστήματος, βασίστηκε σε προηγούμενες ερευνητικές εργασίες που έχουν πραγματοποιηθεί από την ερευνητική ομάδα του Πανεπιστημίου Θεσσαλίας. Οι εργασίες που συνέβαλαν στη δημιουργία του συστήματος είναι :

- Dimitris K. Iakovidis, Dimitris Chatzis, Panos Chrysanthopoulos, and Anastasios Koulaouzidis, "Blood Detection in Wireless Capsule Endoscope Images based on Salient Superpixels" in Proceedings of International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC, Milano, 2015, pp. 731-734
- D. Chatzis, P. Chrysanthopoulos, D. K. Iakovidis, A. Koulaouzidis "Color Based Segmentation Of Capsule Endoscopy Images For Automated Lesion Size

Measurements" in United European Gastroenterology Journal, Gut Vol. 3 Suppl. 5, A36, 2015

- Panos Chrysanthopoulos, Dimitris Chatzis, Anastasios Koulaouzidis, and **D.K.Iakovidis**, "Automated Color-Based Lesion Measurement for Wireless Capsule Endoscopy", in Proceedings of International Conference "Science in Technology", SCinTe, Athens, 2015, Vol 3, pp. 83-85,
- Spiros V. Georgakopoulos, Panos Chrysanthopoulos, Dimitris Chatzis, Anastasios Koulaouzidis, **Vasilis P. Plagianakos** and **Dimitris K. Iakovidis**, "Towards a Color Space for Automated Lesion Segmentation Robotic Capsule Endoscopy", in International Conference on Intelligent Robots and Systems, Hamburg, IROS, 2015

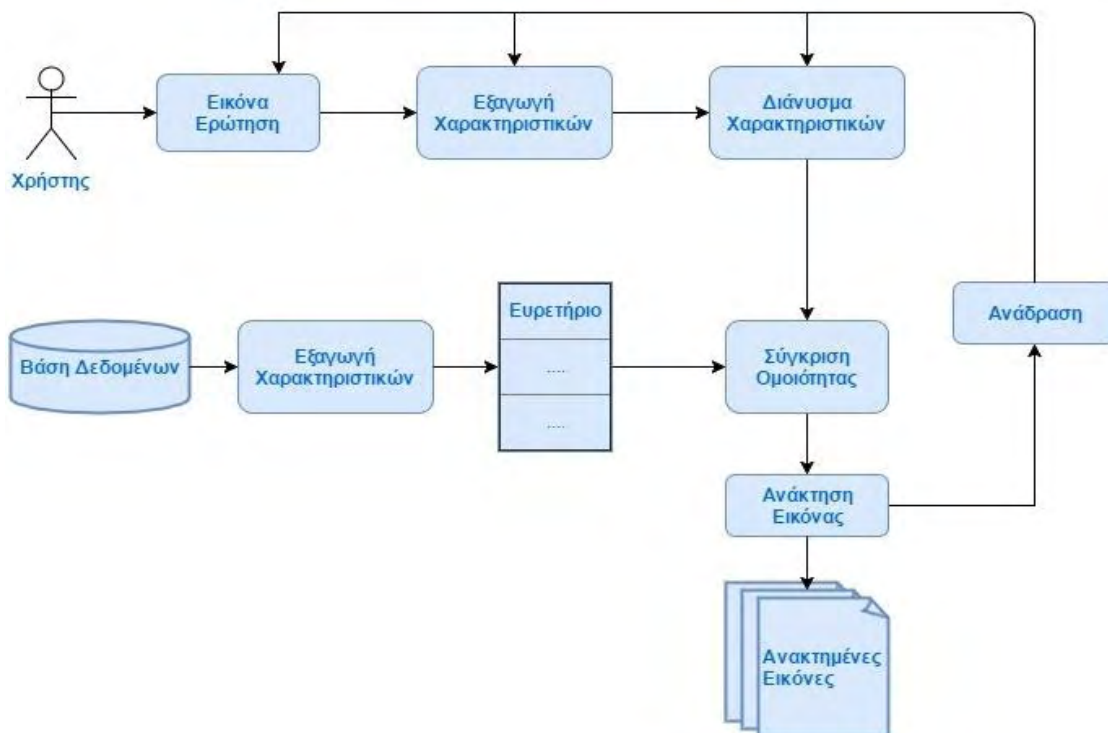
1.4 Δομή της πτυχιακής εργασίας

Η παρούσα ερευνητική εργασία αποτελείται από εννέα κεφάλαια. Το κεφάλαιο 2 περιγράφει την αρχιτεκτονική και τις αρχές λειτουργίας ενός συστήματος ανάκλησης εικόνων. Στο κεφάλαιο 3 γίνεται ανασκόπηση των συστημάτων CBIR, η ανασκόπηση αρχίζει με τα εμπορικά συστήματα αλλά επικεντρώνεται σε συστήματα ανάκλησης που έχουν προταθεί με στόχο να εφαρμοστούν σε ιατρικά δεδομένα. Στο κεφάλαιο 4 περιγράφονται οι χρωματικοί μετασχηματισμοί που χρησιμοποιήθηκαν και διαδικασίες εξαγωγής χαρακτηριστικών που εφαρμόζονται σε συστήματα ανάκλησης. Στο κεφάλαιο 5 παρουσιάζονται μέθοδοι τμηματοποίησης της εικόνας με σκοπό τον διαχωρισμό της σε περιοχές ενδιαφέροντος. Στο κεφάλαιο 6, γίνεται αναφορά στους τρόπους σύγκρισης ομοιότητας μεταξύ των εικόνων. Στην συνέχεια, στο κεφάλαιο 7 η εργασία επικεντρώνεται στις μεθόδους δημιουργίας ευρετηρίου που είναι ζωτικής σημασίας για ένα σύστημα ανάκλησης. Στο κεφάλαιο 8 αναλύεται η λειτουργία του προτεινόμενου συστήματος ανάκλησης εικόνων και στο κεφάλαιο 9 παρουσιάζονται τα πειράματα που διεξάχθηκαν και η σύγκριση των προτεινόμενων χαρακτηριστικών με ήδη υπάρχοντα χαρακτηριστικά που έχουν προταθεί.

Αρχιτεκτονική ενός συστήματος CBIR

2.1 Εισαγωγή

Στη διαδικασία της ανάκλησης όμοιων εικόνων είναι απαραίτητη η εξαγωγή οπτικών χαρακτηριστικών. Το παρακάτω σχήμα (Σχ. 2.1) περιγράφει συνοπτικά τη λειτουργία ενός CBIR συστήματος. Για κάθε εικόνα που υπάρχει στη βάση δεδομένων εξάγονται οπτικά χαρακτηριστικά σε μορφές διανυσμάτων, τα οποία οργανώνονται σε μορφές ευρετηρίου. Τα εξαγόμενα χαρακτηριστικά μπορεί να είναι χαμηλού επιπέδου όπως είναι το χρώμα και η υφή ή χαρακτηριστικά τα οποία περιέχουν σημασιολογικές πληροφορίες για το περιεχόμενο της εικόνας. Όταν ο χρήστης εισάγει την εικόνα επερώτηση (query image) στο σύστημα, αρχικά εξάγονται οπτικά χαρακτηριστικά, έπειτα με βάση αυτά τα χαρακτηριστικά, αναζητούνται παρόμοιες εικόνες. Το σύστημα ανάκλησης εικόνων επιστέφει τις εικόνες που βρίσκονται «πιο κοντά» (σύμφωνα με ένα μέτρο ομοιότητας) σε σχέση με την επερώτηση. Τα χαρακτηριστικά που εξάγονται από την εικόνα, που έχει εισαχθεί στο σύστημα πρέπει να είναι ίδια με τα χαρακτηριστικά που έχουν εξαχθεί από τις εικόνες της βάσης δεδομένων.



Σχ. 2.1 Σχηματική απεικόνιση ενός συστήματος CBIR

2.2 Ανάκληση εικόνων βάση περιεχομένου

Ένα σύστημα ανάκλησης εικόνων έχει τη δυνατότητα να δεχθεί ερωτήματα με διαφορετικές μορφές, η κυριότερη μορφή ερωτήματος σε συστήματα ανάκλησης είναι η επερώτηση με παράδειγμα, άλλες δημοφιλείς μορφές ερωτημάτων που συναντάμε σε συστήματα ανάκλησης είναι η επερώτηση με κείμενο και η επερώτηση με σκίτσο [34]

Επερώτηση με παράδειγμα : Στα περισσότερα συστήματα ανάκλησης δεδομένων με βάση το περιεχόμενο η αναζήτηση ξεκινάει όταν ο χρήστης εισάγει μια εικόνα στο σύστημα σαν επερώτηση (Εικ. 2.1)[34,35, 6, 9,10]. Η εικόνα που λειτουργεί ως επερώτηση έχει παρόμοια μορφή με τις εικόνες που είναι αποθηκευμένες στη βάση δεδομένων. Το σύστημα ξεκινά τη διαδικασία αναζήτησης εξάγοντας χαρακτηριστικά από την εικόνα και στη συνέχεια αναζητεί παρόμοιες στη βάση δεδομένων. Μειονέκτημα αυτής της τεχνικής είναι ότι ο χρήστης μπορεί αρχικά να μην διαθέτει κάποια παρόμοια εικόνα με αυτές που βρίσκονται ήδη αποθηκευμένες στη βάση (page zero problem). Για την μεγιστοποίηση της σαφήνειας των ερωτημάτων και τη βελτιστοποίηση της αποτελεσματικότητας των συστημάτων ανάκλησης έχουν προταθεί διάφορες τεχνικές όπως για παράδειγμα η δυνατότητα επιλογής χαρακτηριστικών από το χρήστη μέσα από μια ποικιλία χαρακτηριστικών, έτσι ο χρήστης μέσω μιας διαδικασίας πειραματισμού και λάθους θα καταλήξει στα χαρακτηριστικά που ταιριάζουν καλύτερα στις απαιτήσεις του.



Εικ. 2.1 Επερώτηση με παράδειγμα, αριστερά: η εικόνα επερώτηση, δεξιά : ανακτημένες εικόνες

Επερώτηση με σκίτσο : Ένα σκίτσο ή μία ζωγραφιά μπορεί να χρησιμοποιηθεί για να περιγράψει μια εικόνα (Εικ. 2.2), με βάση αυτό έχουν προταθεί συστήματα ανάκλησης εικόνων που χρησιμοποιούν σκίτσα σαν επερώτηση [28,29]. Η επερώτηση με σκίτσο έχει προταθεί και στα συστήματα ανάκλησης που έχουν δημιουργηθεί για ιατρικούς σκοπούς [11,26,27]. Λαμβάνοντας υπόψη ότι ένα σκίτσο που περιγράφει με σαφήνεια μια ιατρική εικόνα χρειάζεται χρόνο και ταλέντο στη ζωγραφική, η μέθοδος αυτή περιορίζεται σε μικρού εύρους εικόνων όπως σχήματα όγκων ή ακτινογραφίες ασθενών. Επιπλέον αυτή η τεχνική απαιτεί χρόνο και καλλιτεχνικό ταλέντο από το θεράποντα ιατρό καθιστώντας την αναποτελεσματική για την καθημερινότητα των ιατρών. Για τα συστήματα γενικού τύπου ανάκλησης εικόνων, πάλι ο παράγοντας του χρόνου παίζει καθοριστικό ρόλο καθώς η δημιουργία ενός σκίτσου που αναπαριστά περιγραφικά μια εικόνα είναι μια χρονοβόρα διαδικασία, επιπροσθέτως τα αποτελέσματα συχνά δεν ταιριάζουν στο σκίτσο.



Εικ. 2.2

Επερώτηση με κείμενο : Η τεχνική της επερώτησης με κείμενο [6,10,30,35] είναι πολύ δημοφιλής και χρησιμοποιείται από πολλές μηχανές αναζήτησης, διότι είναι πολύ γνώριμη και εύχρηστη για το χρήστη. Η διαδικασία ξεκινά με το χρήστη να εισάγει στο σύστημα ανάκλησης εικόνων λέξεις κλειδιά που περιγράφουν μια εικόνα. Η επερώτηση με κείμενο χρησιμοποιείται και σε σύστημα ανάκλησης ιατρικών εικόνων. Πολλά συστήματα ανάκλησης ιατρικών εικόνων χρησιμοποιούν κείμενο από το προσωπικό αρχείο του ασθενούς[17] , άλλα συστήματα ορίζουν δικές τους γραμματικές[36] ή χρησιμοποιούν προκαθορισμένο λεξιλόγιο[37], η χρήση του κειμένου για αναζήτηση είναι αδιαμφισβήτητα αξιόπιστη. Επιπλέον για καλύτερα αποτελέσματα μπορεί να χρησιμοποιηθεί ο συνδυασμός επερώτησης με παράδειγμα και επερώτησης με κείμενο.

Για την υιοθέτηση αυτής της τεχνικής απαιτούνται βάσεις δεδομένων που περιέχουν εικόνες μαζί με τον απαραίτητο σχολιασμό τους (image annotation). Ο σχολιασμός των εικόνων είναι μια χρονοβόρα και επίπονη διαδικασία καθώς και μια διαδικασία επιρρεπής σε λάθη διότι η περιγραφή της εικόνας επίκειται σε μερικές λέξεις κλειδιά. Επιπλέον η αναζήτηση και η ανάκληση της εικόνας δεν βασίζεται στο οπτικό περιεχόμενο της, αντίθετα βασίζεται σε λέξεις κλειδιά.



2.3 Εξαγωγή Χαρακτηριστικών

Μια ψηφιακή εικόνα αποτελείται από εικονοστοιχεία, η πληροφορία που εμπεριέχεται στα εικονοστοιχεία δεν επαρκεί για συγκρίσεις ομοιότητας εικόνων. Τα συστήματα ανάκλησης εικόνων ξεπερνούν αυτό το πρόβλημα εξαγοντας χαρακτηριστικά από τις εικόνες που ανταποκρίνονται στις ιδιότητες της εικόνας. Τα εξαγόμενα χαρακτηριστικά αναπαριστούνται με τη μορφή πολυδιάστατων διανυσμάτων που ονομάζονται διανύσματα χαρακτηριστικών. Ανάλογα την περιοχή από τη οποία εξαγονται τα χαρακτηριστικά, διαχωρίζονται στις εξής κατηγορίες:

- **Καθολικά χαρακτηριστικά:** στην εξαγωγή καθολικών χαρακτηριστικών κωδικοποιείται η πληροφορία όλης της εικόνας.

- **Τοπικά χαρακτηριστικά:** εξάγονται από μικρές περιοχές της εικόνας και κωδικοποιούν τις πληροφορίες της κάθε ξεχωριστής περιοχής που δημιουργείται. Οι περιοχές ενδιαφέροντος της εικόνας μπορεί να έχουν ορθογώνιο ή αυθαίρετο σχήμα [34,35].

Τα εξαγόμενα χαρακτηριστικά μπορούμε να τα διαχωρίσουμε ανάλογα με το επίπεδο της σημασιολογία τους σε χαρακτηριστικά υψηλού και χαμηλού επιπέδου:

- **Χαμηλού επιπέδου :** Ονομάζονται τα χαρακτηριστικά που περιγράφουν την εικόνα σε επίπεδο εικονοστοιχείων, όπως το χρώμα και η υφή.
- **Υψηλού επιπέδου:** Ονομάζονται τα χαρακτηριστικά που αναπαριστούν σημασιολογικές έννοιες. Οι άνθρωποι μπορούν να περιγράψουν και να ερμηνεύσουν το περιεχόμενο μια εικόνας χρησιμοποιώντας υψηλές έννοιες όπως για παράδειγμα η αναγνώριση των αντικειμένων της εικόνας ή η σύνδεση των εικόνων με ασαφείς έννοιες όπως τα συναισθήματα, αντίθετα η υπολογιστές αντιλαμβάνονται χαμηλού επιπέδου χαρακτηριστικά.

2.4 Σύγκριση Χαρακτηριστικών

Μια από τις μεθόδους για σύγκριση ομοιότητας εικόνων είναι ο υπολογισμός της απόστασης των διανυσμάτων των χαρακτηριστικών με βάση κάποια μετρική απόστασης [34,35]. Απλές μετρικές απόστασης που χρησιμοποιούνται σε συστήματα ανάκλησης είναι οι νόρμες L_p (αποστάσεις Minkowski), ειδικές περιπτώσεις των οποίων είναι και η Ευκλείδεια απόσταση, η απόσταση Μανχάταν και L_∞ . Υπάρχουν όμως περιπτώσεις από η εφαρμογή των παραπάνω μετρικών δεν δίνει καλά αποτελέσματα για συγκεκριμένα χαρακτηριστικά, για το λόγο αυτό έχουν αναπτυχθεί μετρικές που λαμβάνουν υπόψη την ιδιαιτερότητα και την κατανομή κάθε χαρακτηριστικού, τέτοιες μετρικές είναι χ^2 (δοκιμασία χ^2) η απόσταση EMD (Earth Mover's Distance [67]) κ.α. . Επιπροσθέτως υπάρχουν τεχνικές που εκτιμούν τη συνολική ομοιότητα των εικόνων, τέτοιες τεχνικές για παράδειγμα είναι το μοντέλο στρέβλωσης εικόνας (Image Distortion Model, IDM) [68] και ο αλγόριθμος SamMatch [69].

2.5 Δημιουργία ευρετηρίου

Τα συστήματα ανάκλησης εικόνων με βάση το περιεχόμενο εφαρμόζουν την υπόθεση ότι παρόμοιες εικόνες έχουν κοντινά διανύσματα χαρακτηριστικών με βάση κάποια μετρική απόστασης. Επομένως η αναζήτηση για παρόμοιες εικόνες μετατρέπεται σε μια αναζήτηση κοντινότερων γειτόνων στο χώρο των διανυσμάτων των χαρακτηριστικών. Η δυνατότητα αποδοτικής εκτέλεσης τέτοιων αναζητήσεων είναι κρίσιμη για την απόδοση του συστήματος. Για την επιτάχυνση των αναζητήσεων συχνά χρησιμοποιούνται πολυδιάστατα ευρετήρια όπως τα ΚΔ- δένδρα [70] (Κ-Διαστάσεων), τα δένδρα περιοχής [71] (R- trees) ή ακόμα και σχεσιακές βάσεις δεδομένων [72].

Οι δομές ευρετηρίων καθιστώνται λιγότερο αποδοτικές όταν αυξάνονται οι διαστάσεις των διανυσμάτων, πολλές φορές οι δομές ευρετηρίων έχουν χειρότερη απόδοση από την σειριακή αναζήτηση. Για το λόγο αυτό, σε μια προσπάθεια μείωσης του πλήθους των διαστάσεων, συχνά επιστρατεύονται τεχνικές προβολής σε χώρους μικρότερης διάστασης.

2.6 Ανάδραση Συνάφειας

Με τον όρο ανάδραση συνάφειας (relevance feedback) εννοούμε την διαδικασία αξιολόγησης των αποτελεσμάτων του συστήματος ανάκλησης με βάση το περιεχόμενο, από το χρήστη. Ο χρήστης αξιολογεί τα αποτελέσματα που επιστράφηκαν από το σύστημα και το σύστημα αποκρίνεται σύμφωνα με την αξιολόγηση του χρήστη και επιστέφει νέα προσαρμοσμένα αποτελέσματα. Ο κύκλος της αξιολόγησης του χρήστη και της απόκρισης του συστήματος μπορεί να επαναληφθεί πολλές φορές έως ότου τα αποτελέσματα καλύψουν τις ανάγκες του χρήστη.

Αρνητική ανάδραση ονομάζεται όταν ο χρήστης αξιολογεί τα αποτελέσματα του συστήματος ως μη συναφή με την αναζήτηση του. Αντίθετα, θετική ανάδραση ονομάζεται όταν ο χρήστης αξιολογήσει τις επιστρεφόμενες από το σύστημα εικόνες ως συναφή με την αναζήτηση του.

Ανασκόπηση συστημάτων ανάκλησης ιατρικών εικόνων

3.1 Γενικά συστήματα ανάκλησης εικόνων

Το πρώτο εμπορικό σύστημα ανάκλησης εικόνων με βάση το περιεχόμενο είναι το QBIC [1] το οποίο δημιουργήθηκε από την IBM το 1993, το QBIC υποστηρίζει ερωτήματα με παράδειγμα, με σκίτσο αλλά μπορούμε επίσης να εισαχθούν στο σύστημα λέξεις κλειδιά για την υποβοήθηση της αναζήτησης. Από τη πρώτη εμφάνιση του QBIC μέχρι σήμερα έχουν εμφανιστεί δεκάδες συστήματα ανάκλησης εικόνων που προορίζονται για εμπορική χρήση. Ενδεικτικά θα αναφέρουμε μερικά δημοφιλή προϊόντα όπως το Picalike το οποίο είναι ένα CBIR σύστημα το οποίο χρησιμοποιείται για διαφημιστικούς σκοπούς, το QUERBIE ένα CBIR σύστημα γενικής χρήσης, το Empora είναι ένα σύστημα ανάκλησης για σύγκριση προϊόντων, κλείνοντας με τα εμπορικά συστήματα θα αναφέρουμε τη δημοφιλέστερη μηχανή αναζήτησης εικόνων η οποία είναι το Google Image Search που χρησιμοποιείται από εκατομμύρια χρήστες.

Ξεχωριστά από τα εμπορικά συστήματα υπάρχουν και τα συστήματα ανοιχτού κώδικα που είναι ελεύθερα στη χρήση και παρέχουν τη δυνατότητα ανάπτυξης τους από προγραμματιστές. Το GNU Image Finding tool είναι ένα λογισμικό ανοιχτού κώδικα, το Photobook[2] είναι ένα σύστημα ανάκλησης εικόνων αναπτυγμένο από το εργαστήριο πολυμέσων του MIT, το ViRaL[3] (Visual Image Retrieval and Localization) το οποίο χρησιμοποιεί φωτογραφίες από τον ιστότοπο Flickr, εντοπίζει από ποια περιοχή έχει τραβηχτεί η φωτογραφία που έχει δοθεί στο σύστημα σαν επερώτηση, όλα αυτά τα συστήματα είναι μερικά από τα δεκάδες συστήματα ανοιχτού κώδικα που είναι διαθέσιμα. Για το σκοπό αυτής της ερευνητικής εργασίας θα επικεντρωθούμε σε συστήματα ανάκλησης εικόνων με βάση το περιεχόμενο που έχουν εφαρμογή σε ιατρικές εικόνες.

Το LIRE [51] (Lucene Image Retrieval) είναι μια ανοιχτού κώδικα βιβλιοθήκη για ανάκληση εικόνων με βάση το περιεχόμενο. Περιέχει μια πληθώρα χαμηλού επιπέδου καθολικών χαρακτηριστικών τα οποία μπορούν να χρησιμοποιηθούν ανάλογα με το πρόβλημα. Πέραν των καθολικών χαρακτηριστικών το LIRE παρέχει τη δυνατότητα να εξαγει χαρακτηριστικά με την καινοτόμα τεχνική *bag of visual words*. Επιπλέον το LIRE προσφέρει μεθόδους δημιουργίας ευρετηρίου και αναζήτησης εικόνων. Η ιδιαιτερότητα του LIRE επικείται στη δημιουργία του ευρετηρίου. Το LIRE χρησιμοποιεί την αξιόπιστη μηχανή αναζήτησης ανοιχτού κώδικα Apache Lucene για τη δημιουργία του ευρετηρίου. Η βιβλιοθήκη του LIRE είναι βασικά για γενικού τύπου εικόνες, αλλά λόγω της μεγάλης ποικιλίας των χαρακτηριστικών που παρέχει και του αξιόπιστου ευρετηρίου έχει χρησιμοποιηθεί για εργασίες που βασίζονται σε ιατρικά δεδομένα. Στα πλαίσια αυτής της ερευνητικής

εργασίας χρησιμοποιηθεί η βιβλιοθήκη του LIRE σαν πλατφόρμα ανάπτυξης νέων χαρακτηριστικών και αναζήτησης εικόνων στη βάση δεδομένων.

3.2 Συστήματα ανάκλησης ιατρικών εικόνων

Λόγω των μεγάλων απαιτήσεων και προκλήσεων που έχουν τα ιατρικά συστήματα ανάκλησης, έχουν προταθεί μια πληθώρα συστημάτων ανάκλησης εικόνων βάση περιεχόμενο. Σε αυτό το μέρος του κεφαλαίου αναφέρονται ενδεικτικά κάποια συστήματα ανάκλησης ιατρικών εικόνων. Λιγότερα είναι τα συστήματα ανάκλησης ενδοσκοπικών εικόνων.

Ανάκληση Ακτινογραφιών : Το μεγαλύτερο μέρος των συστημάτων ανάκλησης ιατρικών εικόνων αναφέρονται σε ακτινογραφίες [4,5,7]. Οι Muller et al [6] πρότεινε ένα σύστημα για την ταξινόμηση και ανάκληση ακτινογραφιών για την εξαγωγή χαρακτηριστικών χρησιμοποίησε το μοντέλο bag of visual words μαζί με χαρακτηριστικά υφής. Το MIRS [10] (Multimedia Information Retrieval System, MIRS) είναι ένα σύστημα για τη αποθήκευση και ανάκληση πληροφοριών που προέρχονται από πολυμέσα, για τη δημιουργία των διανυσμάτων των χαρακτηριστικών εφαρμόζει τη τεχνική CEDD (Color Edge Directivity Descriptor, CEDD) επιπλέον εξάγονται και τοπικοί περιγραφείς SIFT, το σύστημα MIRS παρέχει και τη δυνατότητα ανατροφοδότησης για καλύτερη απόδοση. Οι Qian et al [9] εφάρμοσε τεχνικές εξαγωγής σχήματος από ακτινογραφίες με σκοπό την ανάκληση τους. Το CBIR2 [11] είναι ένα υβριδικού τύπου σύστημα ανάκλησης εικόνων με βάση το περιεχόμενο υποστηρίζοντας ερωτήματα με παράδειγμα, ερωτήματα με κείμενο και ερωτήματα με σκίτσο, το σύστημα τους χωρίζεται σε δυο μέρη, το πρώτο μέρος είναι η δημιουργία του ευρετηρίου και το δεύτερο είναι το σύστημα ανάκλησης. Το μέρος του συστήματος που είναι υπεύθυνο για τη δημιουργία του ευρετηρίου, περιλαμβάνει μεθόδους τμηματοποίησης εικόνας, εξαγωγής χαρακτηριστικών, υπολογισμού διανυσμάτων χαρακτηριστικών και οργάνωσης δεδομένων κειμένου. Το μέρος της ανάκλησης είναι η διεπαφή του χρήστη με το σύστημα. Το IRMA [17] (Image Retrieval in Medical Application, IRMA) είναι μια κοινή ερευνητική εργασία μεταξύ τμημάτων του πανεπιστημίου του Άαχεν. Το IRMA είναι κατασκευασμένο πάνω σε ένα δίκτυο μεταξύ των συνεργαζόμενων τμημάτων του Πανεπιστημίου διευκολύνοντας την έρευνα και την ανάπτυξη του συστήματος μέσω του διαμοιρασμού των πόρων, δηλαδή υπάρχει διαμοιρασμός των εικόνων, των εξαγόμενων χαρακτηριστικών, των μεθόδων εξαγωγής ακόμα και της υπολογιστικής ισχύος. Το σύστημα παρέχει μια πληθώρα μεθόδων για την εξαγωγή χαρακτηριστικών. Για την εύρεση των κατάλληλων χαρακτηριστικών από το σύνολο των διαθέσιμων χαρακτηριστικών χρησιμοποιούνται οι ιατρικές διαγνώσεις. Χαρτογραφώντας τη διάγνωση σε κάθε εικόνα μπορούν να επιλεγθούν τα καταλληλά χαρακτηριστικά που περιγράφουν και διαχωρίζουν καλύτερα το οπτικό περιεχόμενο της εικόνας. Αυτό συμπεριλαμβάνει και την αναγνώριση του σημείου ενδιαφέροντος (ROI). Μετά από τοπική ανάλυση και τμηματοποίηση της εικόνας το σύστημα έχει αποκτήσει μια ιεραρχική απεικόνιση των άμορφων περιοχών που περιγράφουν τη δομή της εικόνας. Οι Iakovidis et al [8] πρότεινε ένα σύστημα το οποίο

βασίζεται σε χαρακτηριστικά υφής τα οποία έχουν υπολογιστεί μετά τον μετασχηματισμό της εικόνας.

Ανάκληση τομογραφιών: Ένα σημαντικό μέρος της έρευνας για τα συστήματα ανάκλησης αναφέρεται σε εικόνες που προέρχονται από τομογράφους. Napel et al [12] πρότεινε ένα σύστημα CBIR για αυτόματη ανάκληση εικόνων του ήπατος από αξονικό τομογράφο, για την αναπαράσταση των διανυσμάτων των χαρακτηριστικών εφαρμόστηκαν αλγόριθμοι για τον υπολογισμό της υφής και εύρεσης των περιγραμμάτων μαζί με υψηλού επιπέδου χαρακτηριστικά. Οι Derwin et al [13] χρησιμοποίησε χαρακτηριστικά υφής σε συνεργασία με τον αλγόριθμο Kanade-Lucas-Tomasi για την εξαγωγή χαρακτηριστικών με σκοπό την ανάκληση εικόνων του εγκεφάλου που έχουν προέλθει από μαγνητικό τομογράφο. Το COBRA[15] ή αλλιώς Content Base Retrieval Architecture είναι ένα σύστημα κατασκευασμένο να επεκτείνει της δυνατότητες διάγνωσης και έρευνας των PACS συστημάτων προσθέτοντας τους δυνατότητα ανάκλησης εικόνων με βάση το περιεχόμενο, η εξαγωγή χαρακτηριστικών βασίζεται σε χαρακτηριστικά χρώματος, υφής, σχήματος αλλά και σε ανωτέρου επιπέδου χαρακτηριστικά όπως είναι η αναγνώριση αντικειμένων στην εικόνα. Το MIMS [16] (Medical Image Management System, MIMS) είναι ένα σύστημα για την αποθήκευση και ανάκληση εικόνων, τα χαρακτηριστικά που χρησιμοποιούνται από το MIMS είναι χαρακτηριστικά που περιγράφουν το σχήμα των αντικειμένων στην εικόνα μαζί με σημασιολογικά χαρακτηριστικά υψηλού επιπέδου. Το ASSERT [14] είναι ένα σύστημα ανάκλησης εικόνων με βάση το περιεχόμενο το οποίο αναπτύχθηκε από το τμήμα ηλεκτρολόγων μηχανικών και μηχανικών υπολογιστών του Πανεπιστημίου Purdue, το σύστημα διαχειρίζεται υψηλής ανάλυσης τομογραφίες του πνεύμονα, το σύστημα εξάγει 26 χαρακτηριστικά εντός και εκτός μιας προσημασμένης περιοχής της εικόνας.

Ανάκληση μαστογραφιών: Η μαστογραφία είναι μια διαγνωστική εξέταση με στόχο την ανίχνευση του καρκίνου του μαστού, οι Jiang et al [18] πρότειναν ένα σύστημα για την υπολογιστική διάγνωση και ανάκληση εικόνων που έχουν προέρθει από μαστογραφίες, για την εξαγωγή των χαρακτηριστικών χρησιμοποιήθηκαν τοπικοί περιγραφείς SIFT για τον εντοπισμό σημείων ενδιαφέροντος (ROIs), στη συνέχεια το σύστημα αποθηκεύει τα σημεία ενδιαφέροντος σε ένα δέντρο για την καλύτερη απόδοση του συστήματος. Οι Quelled et al [19, 20] χρησιμοποίησε υπογραφές εικόνων μαζί με ένα νευρωνικό δίκτυο Bayes με σκοπό την ανάκληση μαστογραφιών [19], επίσης οι Quelled et al [20] επιστράτευσε δένδρα αποφάσεων σε συνδυασμό με υπογραφές εικόνων για την αποτελεσματική ανάκληση μαστογραφιών.

Ανάκληση ενδοσκοπικών εικόνων: Σε αυτό το τμήμα της ερευνητικής εργασίας θα αναλύσουμε μερικές τεχνικές που έχουν εφαρμοστεί για τη δημιουργία CBIR συστημάτων σε ενδοσκοπικά δεδομένα. Οι Beecks et al [21], πρότεινε ένα σύστημα ανάκλησης για ενδοσκοπικά βίντεο, για την εξαγωγή των διανυσμάτων των χαρακτηριστικών χρησιμοποίησε υπογραφές εικόνων μαζί με μια νέα τεχνική για τη σύγκριση των υπογραφών των εικόνων. Οι υπογραφές των εικόνων υπολογίστηκαν με βάση τοπικούς περιγραφείς, τη τιμή του χρώματος στο *CIE-Lab* χώρο χρώματος, και τη τιμή της τραχύτητας και της αντίθεσης, από ένα τυχαίο αριθμό εικονοστοιχείων από το σύνολο των εικονοστοιχείων της εικόνας, στη συνέχεια τα εξαγόμενα χαρακτηριστικά συσταδοποιούνται με τον αλγόριθμο *k-means*. Οι Manish et al [22] εφάρμοσαν

πολυκλιμακωτή γεωμετρική ανάλυση (Multiscale Geometric Analysis) του μετασχηματισμού Counterlet και το στατιστικό μοντέλο για τη εξαγωγή χαρακτηριστικών με σκοπό την ανάκληση ενδοσκοπικών εικόνων. Ashish Oberoi και Manpreet Singh [23] δημιούργησαν ένα CBIR σύστημα το οποίο δοκίμασαν σε ενδοσκοπικά δεδομένα, καθώς και σε οδοντιατρικά δεδομένα και ακτινογραφίες του κρανίου. Για την εξαγωγή χαρακτηριστικών χρησιμοποίησαν περιγραφείς Fourier και κυματομορφές Haar. Η σύγκριση ομοιότητας πραγματοποιήθηκε με το υπολογισμό της Ευκλείδειας απόστασης και της απόστασης Canberra μεταξύ των εξαγόμενων χαρακτηριστικών. Οι Shunren et al [25] δημιούργησαν ένα CBIR εξάγοντας χαμηλού επιπέδου χαρακτηριστικά από ενδοσκοπικές εικόνες όπως είναι η υφή και το χρώμα. Το συγκεκριμένο σύστημα CBIR επιτρέπει στο χρήστη να μεταβάλλει τα βάρη των αποστάσεων που χρησιμοποιούνται στη σύγκριση μεταξύ των διανυσμάτων των χαρακτηριστικών, εφαρμόζοντας την τεχνική της ανατροφοδότησης του χρήστη. Οι Vatamanu et al [24] πρότεινε ένα CBIR σύστημα το οποίο εξάγει χαρακτηριστικά υφής εφαρμόζοντας την τεχνική του LBP (Local Binary Pattern). Το αναφερόμενο σύστημα ακολουθεί τεχνικές εξόρυξης δεδομένων για την βέλτιστη επιλογή παραμέτρων στην εξαγωγή LBP ιστογραμμάτων που θα χρησιμοποιηθούν ως διανύσματα χαρακτηριστικών.

Πιν. 3.1 Ενδεικτικά σύστημα ανάκλησης ιατρικών εικόνων μέχρι το 2010.

Αναφορά	Χαρακτηριστικά	Αλγ. Εξαγωγής	Εικόνες	Ερωτήματα
Henning et al. 2016 [6]	Υφή	LBP, BOVW	ακτινογραφίες	Παράδειγμα
Beecks et al 2015[2]	Υπογραφές Εικόνων	Προτείνεται	ενδοσκοπικές	Παράδειγμα
Manish et al 2015[22].	Χρώμα, Υφή	Μετασχηματισμός counterlet	ενδοσκοπικές	Παράδειγμα
Vatamanu et al. 2015[24]	Υφή	LBP	ενδοσκοπικές	Παράδειγμα
Trupti S et al. 2014 [10]	Σχήμα, Τοπικοί περιγραφείς	CEDD, SIFT	ακτινογραφίες	Παράδειγμα, Κείμενο
Ashish et al. 2012 [23]	Περίγραμμα, Αναγνώριση αντικειμ.	Περιγραφείς Fourier, Κυματομορφές Haar	ενδοσκοπικές	Παράδειγμα
Menglin et al. 2011 [18]	Τοπικοί περιγραφείς	SIFT	μαστογραφίες	Παράδειγμα
Quelleg et al 2011 και 2010 [19,20]	Υπογραφές εικόνας	Μετασχηματισμός Εικόνων με κυματομορφές	μαστογραφίες	Παράδειγμα
Qian et al. 2010 [11]	Σχήμα		ακτινογραφίες	Παράδειγμα
Unay et al 2010 [13]	Τοπικοί περιγραφείς Υφή	LBP, Kanade-Lucas-Tomasi	τομογραφίες	Παράδειγμα
Napel et al. 2010 [12]	Υφή, περίγραμμα	φίλτρα Gabor, Ιστογράμματα	τομογραφίες	Παράδειγμα

Κεφάλαιο 4

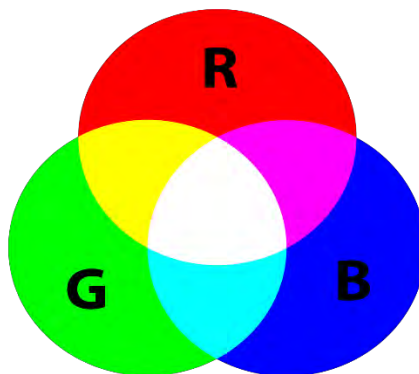
Μέθοδοι εξαγωγής χαρακτηριστικών εικόνας

4.1 Χρωματικοί χώροι

Η μαθηματική αναπαράσταση του χρώματος έχει τη μορφή ενός διανύσματος τριών ή τεσσάρων διαστάσεων. Οι χρωματικοί χωρίζονται σε γραμμικούς χώρους χρώματος (πχ. RGB, XYZ) και μη γραμμικούς χώρους (πχ. CIE-L*a*b*, HSV). Οι γραμμικοί χώροι χρώματος βασίζονται σε ένα σύνολο βασικών χρωμάτων και περιγράφουν το χρώμα ως ένα σταθμισμένο άθροισμα της έντασης των βασικών χρωμάτων που αποτελούν το χώρο. Οι μη γραμμικοί χώροι έχουν τη δυνατότητα να περιέχουν πιο πολλές ιδιότητες του χρώματος και μπορούν να – για να περιγράψουν καλύτερα το χρώμα[74].

4.1.1 Χρωματικός χώρος RGB

Είναι ο πιο διαδεδομένος χρωματικός χώρος, στο RGB χώρο το χρώμα εκφράζεται ως γραμμικός συνδυασμός των τριών βασικών χρωμάτων (Εικ. 4.1), κόκκινο (Red), πράσινο (Green) και μπλε (Blue). Ο χώρος RGB χρησιμοποιείται από πολλές συσκευές απεικόνισης και επεξεργασίας εικόνας. Λόγω της ευρείας διάδοσης του RGB χώρου, εφαρμόζεται και σε συστήματα ανάκλησης με βάση το περιεχόμενο, όμως υπάρχουν χρωματικοί χώροι που ανταποκρίνονται καλύτερα σε εφαρμογές της υπολογιστικής όρασης. Ο RGB χρωματικός χώρος έχει διάφορες παραλλαγές όπως είναι ο sRGB, Adobe RGB, Adobe Wide Gamut RGB, scRGB κ.α.



Εικ. 4.1 Απεικόνιση του RGB χρωματικού χώρου

4.1.2 Χρωματικός χώρος HSV και HSI

Ο HSV (Hue Saturation Value) είναι ένας κυλινδρικός χώρος χρώματος (Εικ. 4.2) που εκφράζει το χρώμα χρησιμοποιώντας μια τριάδα τιμών που αντιπροσωπεύουν την απόχρωση (Hue), το κορεσμό (Saturation) και την τιμή του χρώματος (Value). Η απόχρωση είναι το χρώμα που γίνεται αντιληπτό λόγω του μήκους κύματος, η καθαρότητα ή κορεσμός υποδηλώνει τη πρόσμιξη του λευκού στο χρώμα και η τιμή υποδηλώνει το βαθμό μίξης του καθαρού χρώματος με το μαύρο. Η απόχρωση είναι κυκλικό μέγεθος και παίρνει τιμές από 0° ως 360°. Τα κανάλια S και V παίρνουν τιμές στο διάστημα [0,1]. Τα συμπληρωματικά χρώματα στον HSV χώρο απέχουν 180°. Το πλεονέκτημα που παρουσιάζει ο χρωματικός χώρος είναι ότι η απόχρωση (H) και ο κορεσμός (S) είναι συσχετισμένα με το τρόπο αντίληψης του ανθρώπινου ματιού, καθιστώντας το χρωματικό χώρο HSV ιδανικό υποψήφιο για ανάπτυξη εφαρμογών ανάκλησης εικόνας. Η μετατροπή στο χρωματικό χώρο HSV από το RGB πραγματοποιείται με την βοήθεια των παρακάτω σχέσεων (Σχ. 4.1– 4.4).

Στο χρωματικό μοντέλο HSI, τα κανάλια H και S υπολογίζονται ομοίως όπως στο μοντέλο HSV. Η διαφορά των δυο μοντέλων βρίσκεται στον υπολογισμό της έντασης του χρώματος (Σχέση 4.5).

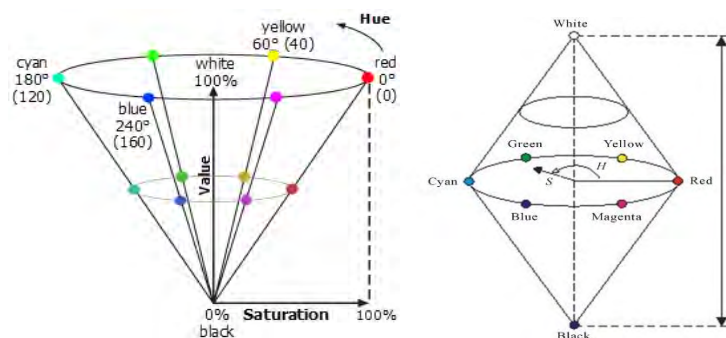
$$H_1 = \cos^{-1} \left\{ \frac{\frac{1}{2}[(R-G) + (R-B)]}{\sqrt{(R-G)^2 + (R-B)(G-R)}} \right\} \quad (\Sigma\chi. 4.1)$$

$$H = \begin{cases} H_1, & \text{αν } B \leq G \\ 360^\circ - H_1 & \text{αν } B > G \end{cases} \quad (\Sigma\chi. 4.2)$$

$$S = \frac{\max(R,G,B) - \min(R,G,B)}{\max(R,G,B)} \quad (\Sigma\chi. 4.3)$$

$$V = \frac{\max(R,G,B)}{255} \quad (\Sigma\chi. 4.4)$$

$$I = \frac{1}{3}(R + G + B) \quad (\Sigma\chi. 4.5)$$



Εικ. 4.2 Απεικόνιση του HSV και HSI χώρου

4.1.3 Χρωματικός χώρος XYZ

Ο χρωματικός χώρος XYZ είναι ένας θεμελιώδης χώρος που βασίζεται στα πρότυπα της CIE (Commission internationale de l'éclairage). Έπειτα από πειράματα η CIE κατέληξε σε ένα σύνολο μεγεθών που συμφωνούν με το τρόπο αντίδρασης του αμφιβληστροειδή στη διεργασία της όρασης. Τα X,Y,Z χρησιμοποιούνται για την περιγραφή ενός χρώματος ως αναλογίες λαμβάνοντας ποσοστά από 0% έως 100% και σε μερικές περιπτώσεις ελάχιστα πάνω από 100%. Οι σχέσεις (Σχέση 4.6) μετασχηματισμού από τον RGB χώρο στον XYZ είναι

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.412453 & 0.357580 & 0.180423 \\ 0.212671 & 0.715160 & 0.072169 \\ 0.019334 & 0.119193 & 0.950227 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (\text{Σχ. 4.6})$$

4.1.4 Χρωματικός χώρος CIE-L*a*b*

Ο χρωματικός χώρος CIE-L*a*b* αποτελείται από τρία κανάλια (Εικ. 4.4), το L* κανάλι το οποίο κωδικοποιεί τη φωτεινότητα και παίρνει τιμές συνήθως στο διάστημα [0,100], το a* κανάλι που εκφράζει την ανάμειξη κόκκινου – πράσινου και το b* κανάλι που κωδικοποιεί την ανάμειξη κίτρινου και μπλε. Τα κανάλια a*, b* παίρνουν συνήθως τιμές στο διάστημα [-127, 127]. Το πλεονέκτημα του χρωματικού χώρου CIE-L*a*b* είναι ότι η Ευκλείδεια απόσταση των L*a*b* συντεταγμένων αποδίδει προσεγγιστικά τη χρωματική απόσταση που αντιλαμβάνεται ο άνθρωπος. Η ιδιότητα αυτή είναι χρήσιμη σε συστήματα ανάκλισης εικόνων με βάση το περιεχόμενο καθώς οι μικρές χρωματικές αλλαγές γίνονται αμέσως αντιληπτές. Οι χρωματικοί χώροι που ικανοποιούν αυτή την ιδιότητα ονομάζονται αντιληπτά ομοιόμορφοι χρωματικοί χώροι (perceptual uniform color spaces). Για το μετασχηματισμό στο CIE-L*a*b* χρωματικό χώρο από το RGB, χρησιμοποιείται πρωτιά ο μετασχηματισμός RGB σε XYZ, και στη συνέχεια μετατρέπεται σε CIE-L*a*b*. Για το μετασχηματισμό χρησιμοποιείται ένα προκαθορισμένο σημείο λευκού (standard illuminant) στη συγκεκριμένη περίπτωση έχει χρησιμοποιηθεί το D65 που αντιπροσωπεύει περίπου το φως του μεσημεριού στην Δυτική και Βόρεια Ευρώπη. Οι σχέσεις για το μετασχηματισμό είναι :

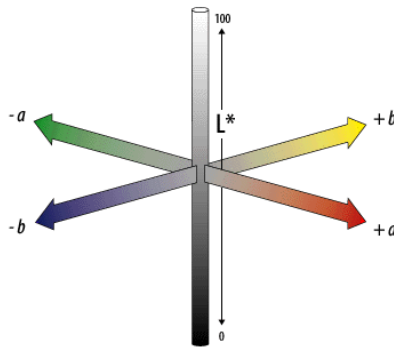
$$L = 116f(Y/Y_n) - 16 \quad (\text{Σχ. 4.7})$$

$$a^* = 500[f(X/X_n) - f(Y/Y_n)] \quad (\text{Σχ. 4.8})$$

$$b^* = 200[f(Y/Y_n) - f(Z/Z_n)] \quad (\text{Σχ. 4.9})$$

$$f(t) = \begin{cases} t^{1/3} & t > 0.008856 \\ 7.787 \cdot t + 16/116 & \text{αλλιώς} \end{cases} \quad (\text{Σχ. 4.10})$$

$$D65 = [X_n \ Y_n \ Z_n] = [0.95045 \ 1 \ 1.088754] \quad (\text{Σχ. 4.11})$$



Εικ. 4.4 Σχηματική απεικόνιση του CIE-Lab χρωματικού χώρου

4.1.5 Ιστογράμματα

Το ιστόγραμμα μια εικόνας εκφράζει τη κατανομή των αποχρώσεων σε μια εικόνα. Ιστογράμματα μπορούν να εξαχθούν από όλα τα κανάλια του χρωματικού χώρου που βρίσκεται η εξεταζόμενη εικόνα. Ένα ιστόγραμμα είναι ένα γράφημα που στο οριζόντιο άξονα έχει τιμές $[0, L-1]$ όπου L το εύρος του καναλιού του χρωματικού χώρου στον οποίο βρίσκεται η εικόνα, συνήθως είναι 0-255 και στο κατακόρυφο άξονα έχει το πλήθος των εικονοστοιχείων που έχουν κάθε τιμή του χρωματικού χώρου. Ο κατακόρυφος άξονας μπορεί να κανονικοποιηθεί με βάση τη μέγιστη τιμή του ιστογράμματος. Σύμφωνα με τα παραπάνω το ιστόγραμμα $h(k)$ μια εικόνας I , διαστάσεων $N \times M$ με αποχρώσεις ενός καναλιού στο διάστημα $[0, L-1]$ μπορεί να προσδιοριστεί με τη βοήθεια των σχέσεων:

$$h(k) = \begin{cases} h(k) + 1, & \text{αν } I(i, j) = k \\ h(k) & \text{διαφορετικά} \end{cases} \quad (\text{Σχ. 4.12})$$

$$\text{όπου } (i, j) \in \{0, 1, \dots, L-1\} \quad \sum_{k=0}^{L-1} h(k) = N \cdot M$$

4.2 Χαρακτηριστικά υφής

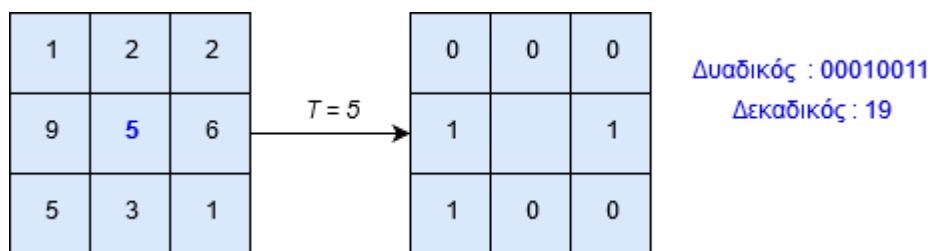
Στην υπολογιστική όραση δεν υπάρχει κάποιος συγκεκριμένος ορισμός για την έννοια της υφής (texture) σε μια εικόνα, καθώς έχουν προταθεί διάφοροι ορισμοί στο πλαίσιο διαφορετικών εφαρμογών. Η υφή είναι μια δομή της εικόνας που είναι ομοιογενής, τυχαία και επαναλαμβάνεται σε μια περιοχή της εικόνας πολύ μεγαλύτερη από την ίδια τη δομή [34]. Άλλος ένας ορισμός της υφής είναι, ότι απομένει στην εικόνα μετά την αφαίρεση των αντικειμένων που περιέχει [35]. Στη συνέχεια αναλύονται αντιπροσωπευτικές τεχνικές για

εξαγωγή χαρακτηριστικών υφής που χρησιμοποιούνται σε συστήματα ανάκλησης με βάση το περιεχόμενο.

4.2.1 Τοπικά δυαδικά πρότυπα

Τα χαρακτηριστικά των τοπικών δυαδικών προτύπων (Local Binary Patterns, LBP) έχουν εφαρμοστεί σε πολλά συστήματα ανάκλησης αλλά και σε πολλές εφαρμογές υπολογιστικής όρασης καθώς θεωρούνται ισχυρά χαρακτηριστικά υφής. Τα LBP προτάθηκαν αρχικά το 1996 από τους Ojala et al [53] έκτοτε έχουν τροποποιηθεί για την βέλτιστη απόδοσή τους.

Για την εξαγωγή LBP χαρακτηριστικών, σύμφωνα με την αρχική τους μορφή, χρησιμοποιείται ένα παράθυρο 3x3 το οποίο σαρώνει όλη την εικόνα με βήμα ένα εικονοστοιχείο (pixel) κάθε φορά. Η τιμή του κεντρικού εικονοστοιχείου λαμβάνεται ως τοπικό κατώφλι T , τα υπόλοιπα 8 εικονοστοιχεία της γειτονιάς λαμβάνουν τιμή 0 ή 1 ανάλογα αν η τιμή τους είναι μικρότερη ή μεγαλύτερη από το T αντίστοιχα. Μετά την κατωφλίωση παράγεται ένας δυαδικός αριθμός από τις τιμές των γειτονικών εικονοστοιχείων ο οποίος μετατρέπεται σε δεκαδικό αριθμό και εκφράζει τη τιμή του LBP για τη συγκεκριμένη γειτονιά (Εικ. 4.8). Στο τελευταίο στάδιο το αλγόριθμου υπολογίζεται ένα ιστόγραμμα από τις μετασχηματισμένες τιμές του LBP το οποίο χρησιμοποιείται ως διάνυσμα χαρακτηριστικών



Εικ. 4.8 Υπολογισμός LBP

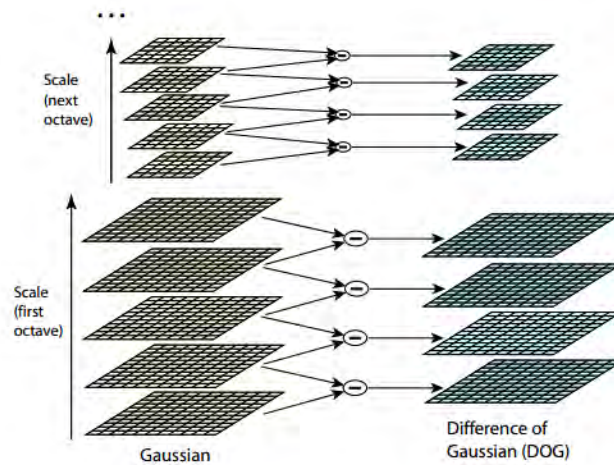
4.2.2 Μετασχηματισμός χαρακτηριστικών αμετάβλητης κλίμακας

Ο μετασχηματισμός χαρακτηριστικών αμετάβλητης κλίμακας (Scale Invariant Feature Transform, SIFT) είναι ένας αλγόριθμος για τον εντοπισμό και τη περιγραφή τοπικών χαρακτηριστικών στις εικόνες. Τα χαρακτηριστικά που ανιχνεύονται είναι σημεία πάνω στην εικόνα που φέρουν χρήσιμη πληροφορία η οποία δεν μπορεί να μεταβληθεί λόγω κλιμάκωσης, στροφής, μεταβολής της φωτεινότητας. Τα χαρακτηριστικά SIFT προτάθηκαν το 1999 από το David Lowe [54], από τότε το αλγόριθμος έχει αναπτυχθεί περαιτέρω. Τα κυρία βήματα για τον υπολογισμό των σημείων SIFT σύμφωνα με τον δημιουργό των χαρακτηριστικών είναι τα ακόλουθα:

- Δημιουργία κλιμακούμενων Γκαουσιανών Οκτάβων
- Προσδιορισμός τοπικών ακρότατων

- Ακριβής εντοπισμός σημείων
- Εκχώρηση προσανατολισμού
- Εύρεση του τοπικού περιγραφέα

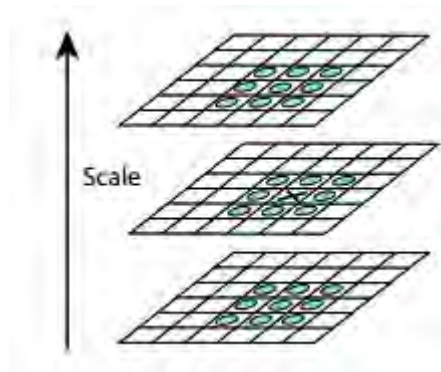
Στο πρώτο βήμα του αλγορίθμου δημιουργούνται οι κλιμακούμενες Γκαουσιανές οκτάβες, δηλαδή δημιουργούνται ομάδες εικόνων στην οποία κάθε εικόνα προκύπτει έπειτα από μια συνέλιξη της προηγούμενης με μια Γκαουσιανή. Η πρώτη εικόνα είναι αποτέλεσμα της συνέλιξης της αρχικής εικόνας με ένα φίλτρο Gauss με τυπική απόκλιση σ ενώ κάθε επόμενη εικόνας της οκτάβας προκύπτει ως η συνέλιξη της αρχικής εικόνας με ένα φίλτρο Gauss με τυπική απόκλιση πολλαπλάσια του προηγούμενου φίλτρου $k\sigma$ (Σχ. 4.5). Σύμφωνα με το Lowe το k πρέπει να είναι της μορφής $k = 2^{1/s}$ όπου s ένας ακέραιος αριθμός. Μετά από κάθε οκτάβα η εικόνα υποδειγματοληπτείται κατά παράγοντα 2. Οι διαδοχικές εικόνες αφαιρούνται και δημιουργούνται εικόνες DoG (Difference of Gaussians).



Σχ.4.5 Δημιουργία οκτάβων και εικόνων DoG

Μετά το πέρας της δημιουργίας των εικόνων DoG, μπορούν να εντοπιστούν χαρακτηριστικά σημεία στην εικόνα (Σχ.4.6), τα σημεία ανιχνεύονται αρχικά ως τοπικά ακρότατα των DoG εικόνων. Κατά τη διαδικασία αυτή ένα σημείο των DoG εικόνων παραμείνει ως πιθανό χαρακτηριστικό μόνο εάν είναι τοπικό μέγιστο ή ελάχιστο.

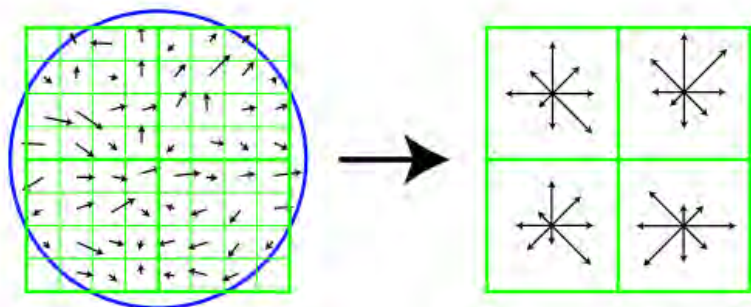
Τα πιθανά χαρακτηριστικά που επιλέχθηκαν στο προηγούμενο βήμα, είναι μεγάλα στον αριθμό και πρέπει να μειωθούν και να μείνουν τα πιο ευσταθή από αυτά. Στη διαδικασία του ακριβή εντοπισμού των σημείων επιδιώκουμε να απορρίψουμε τα σημεία που βρίσκονται πάνω σε ακμές ή σημεία που έχουν χαμηλή αντίθεση και να διατηρήσουμε αυτά που βρίσκονται σε γωνίες. Το φιλτράρισμα των σημείων γίνεται με μια τρισδιάστατη προσέγγιση των 26 γειτονικών σημείων της μεταξύ της οκτάβας που βρισκόμαστε και των διπλανών οκτάβων, με σκοπό την εύρεση της ακριβής θέσης του ακρότατου. Η εύρεση του ακρότατου γίνεται με τη βοήθεια των σειρών Taylor. Ένα επιπλέον φιλτράρισμα των σημείων γίνεται στη συνέχεια, χρησιμοποιώντας τις ιδιοτιμές του πίνακα Hessian κάθε πιθανού ακρότατου.



Σχ. 4.6 Εντοπισμός ακρότατων γύρω από το εικονοστοιχείο (X) σε

Η ανεξαρτητοποίηση των σημείων από περιστροφές της εικόνας επιτυγχάνεται με εκχώρηση του προσανατολισμού κάθε σημείου. Για την εκχώρηση του προσανατολισμού χρησιμοποιείται ένα ιστόγραμμα 36 θέσεων που καλύπτει το φάσμα των 360° (10 μοίρες ανά θέση), για κάθε εικονοστοιχείο της γειτονιάς του ακρότατου υπολογίζεται το μέτρο και η κλίση του. Οι κλίσεις των γειτονικών εικονοστοιχείων συμπληρώνουν το ιστόγραμμα των 36 θέσεων. Οι τιμές του ιστογράμματος σταθμίζονται με τη βοήθεια της μέσης τιμής των μέτρων των διανυσμάτων της γειτονιάς και με μια Gaussian με τυπική απόκλιση 1.5σ . Η υψηλότερη κορυφή που εντοπίζεται είναι και ο τελικός προσανατολισμός του σημείου. Οι κορυφές που αντιστοιχούν στο 80% της μέγιστης κορυφής διατηρούνται σαν ένα καινούργιο χαρακτηριστικό σημείο με την ίδια θέση.

Η εξαγωγή του τοπικού περιγραφέα γίνεται λαμβάνοντας υπόψη μια γειτονιά 16x16 εικονοστοιχείων γύρω από το χαρακτηριστικό σημείο. Η γειτονιά υποδιαιρείται σε υπογειτονίες 4x4, με αποτέλεσμα να δημιουργούνται 16 γειτονιές 4x4. Από κάθε υπογειτονιά υπολογίζονται 8 διανύσματα με τη διαδικασία που περιγράψαμε παραπάνω για τον υπολογισμό του προσανατολισμού, τελικά θα προκύψουν 128 διανύσματα ακμών (16x8) τα μέτρα των οποίων σχηματίζουν το τελικό περιγραφέα (Σχ. 4.7).



Σχ. 4.7 Εξαγωγή τοπικού περιγραφέα

4.2.3 Υπολογιστικά επιταχυμένα ανθεκτικά χαρακτηριστικά

Ο αλγόριθμος εξαγωγής υπολογιστικά επιταχυμένων ανθεκτικών χαρακτηριστικών (Speeded-Up Robust Features, SURF[52]) είναι εμπνευσμένος από τον αλγόριθμο SIFT. Ο αλγόριθμος SURF υπόσχεται μια δραματική αύξηση στη ταχύτητα του υπολογισμού των σημείων, το οποίο επιτυγχάνεται με τη χρήση των «εικόνων ολοκλήρωμα» ή integral images. Η εικόνα ολοκλήρωμα υπολογίζεται σύμφωνα με τη σχέση (Σχ. 4.13):

$$I_{int}(x, y) = \sum_{i=0}^x \sum_{j=0}^y I(i, j) \quad (\text{Σχ. 4.13})$$

Όπου I είναι η αρχική εικόνα και x, y οι συντεταγμένες του εικονοστοιχείου στην εικόνα.

Με τις integral images μπορούμε να υπολογίσουμε το άθροισμα των εικονοστοιχείων μιας ορθογώνιας περιοχής της εικόνας με μόλις τρεις προσθέσεις/αφαιρέσεις. Επιπλέον επιτρέπει το ταχύτατο υπολογισμό οποιουδήποτε φίλτρου που υλοποιείται από γραμμική συνέλιξη. Ο αλγόριθμος SURF για τη επιλογή της θέσης των χαρακτηριστικών χρησιμοποιεί τον πίνακα Hessian, δοθέντος ενός σημείου $P(x, y)$ ο πίνακας παίρνει τη μορφή:

$$H(P, \sigma) = \begin{bmatrix} L_{xx}(x, \sigma) & L_{xy}(x, \sigma) \\ L_{xy}(x, \sigma) & L_{yy}(x, \sigma) \end{bmatrix} \quad (\text{Σχ. 4.14})$$

L_{xx} είναι η συνέλιξη της παραγωγού δεύτερης τάξης ως προς x της συνάρτησης Gauss με την εικόνα στο σημείο P . Ομοίως ορίζονται και οι L_{xy} και L_{yy} . Οι παράγωγοι δευτέρου βαθμού προσεγγίζονται από φίλτρα μέσου όρου (box -filters). Η εφαρμογή των box- filters απαιτεί λιγότερη υπολογιστική ισχύ χρησιμοποιώντας της integral images.

Ο αλγόριθμος SURF δημιουργεί οκτάβες από εικόνες όπως και ο SIFT, με τη διαφορά όμως ότι στη για τη δημιουργία των οκτάβων αλλάζει το μέγεθος των φίλτρων μέσου όρου που χρησιμοποιούνται για τον υπολογισμό του Hessian. Επιπλέον λόγω της χρήσης integral images δεν απαιτείται η χρήση του ίδιου φίλτρου στην έξοδο του προηγούμενου στρώματος αλλά μπορούμε να εφαρμόσουμε τα φίλτρα οποιουδήποτε μεγέθους στην αρχική εικόνα. Με αυτή τη τεχνική κάθε οκτάβα αντιπροσωπεύει μια σειρά από φιλτραρίσματα, η κατασκευή του χώρου των οκτάβων ξεκινάει συνήθως με χρήση φίλτρου διαστάσεων 9×9 και σε κάθε οκτάβα το μέγεθος του παραθύρου αυξάνεται.

Ο εντοπισμός της ακριβής θέσης των χαρακτηριστικών επιτυγχάνεται με παρόμοιο τρόπο όπως στον αλγόριθμο SIFT εντοπίζοντας τα σημεία που μεγιστοποιείται η τιμή της ορίζουσας του Hessian.

Για τη εκχώρηση του προσανατολισμού αντί για την κατανομή των παραγώγων που χρησιμοποιεί ο αλγόριθμος SIFT, χρησιμοποιούνται οι κατανομές των κυματιδίων Haar πρώτης τάξης στην οριζόντια και κάθετη κατεύθυνση.

Ο τοπικός περιγραφέας του αλγορίθμου SURF προσδιορίζεται μια τετράγωνη περιοχή με κέντρο ένα σημείο SURF που έχει εξαχθεί. Η περιοχή χωρίζεται σε παράθυρα 4x4 τετράγωνες περιοχές, σε κάθε περιοχή υπολογίζονται οι αποκρίσεις των κυματιδίων Haar. Στη συνέχεια αθροίζονται οι αποκρίσεις των σωματιδίων, από κάθε υποπεριοχή k εξαγονται τα εξής χαρακτηριστικά :

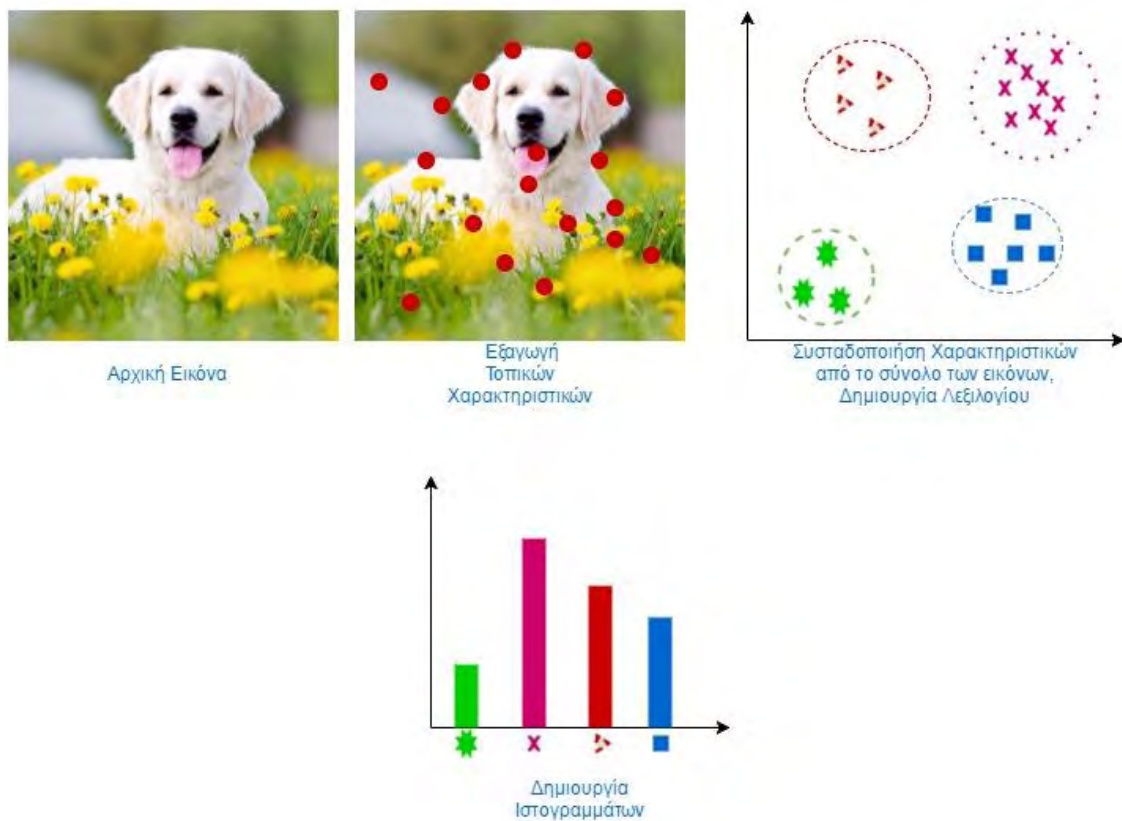
$$Fk = (\sum dx, \sum dy, \sum |dx|, \sum |dy|) \text{ (Σχ. 4.15)}$$

Επειδή από κάθε υποπεριοχή εξαγονται 4 χαρακτηριστικά, ο τοπικός περιγραφέας για την αρχική περιοχή έχει μέγεθος 64.

4.3 Σάκος οπτικών λέξεων

Το μοντέλο του Σάκου οπτικών λέξεων (Bag of Visual Words, BOVW) (Σχ. 4.10) είναι πολύ δημοφιλή στα συστήματα ανάκλησης με βάση το περιεχόμενο [31, 32, 33, 34]. Η τεχνική είναι εμπνευσμένη από το μοντέλο *bag of words* (BOW) που χρησιμοποιείται στην αναζήτηση εγγράφων κειμένου. Στο μοντέλο BOW, παράγεται ένα ιστόγραμμα με τις συχνότητες των λέξεων που περιέχει το έγγραφο, η λογική πίσω από το μοντέλο BOW είναι ότι έγγραφα με παρόμοιο περιεχόμενο θα έχουν ίδιες λέξεις με παρόμοιες συχνότητες στο περιεχόμενο τους, με αποτέλεσμα να είναι εφικτή η σύγκριση των εγγράφων με βάση τα ιστογράμματα που έχουν δημιουργηθεί. Στην υπολογιστική όραση για την εφαρμογή του μοντέλου είναι απαραίτητη η εξαγωγή οπτικών χαρακτηριστικών τα οποία θα έχουν το ρόλο των λέξεων, αυτό επιτυγχάνεται με την εξαγωγή τοπικών χαρακτηριστικών από σημεία ενδιαφέροντος της εικόνας. Συνεπώς μια εικόνα θα αποτελεί ένα έγγραφο που θα περιέχει μια σειρά από οπτικές λέξεις. Μετά την εξαγωγή τοπικών χαρακτηριστικών από τις εικόνες τα χαρακτηριστικά που υπολογίστηκαν συσταδοποιούνται και δημιουργούνται οι κωδικές λέξεις του λεξιλογίου. Στο επόμενο στάδιο υπολογίζεται ένα ιστόγραμμα εμφάνισης των κωδικών λέξεων σε κάθε εικόνα, με αποτέλεσμα κάθε εικόνα να εκφράζεται από ένα ιστόγραμμα. Ο αλγόριθμος συνοψίζεται στα εξής βήματα :

- Εξαγωγή τοπικών χαρακτηριστικών
- Συσταδοποίηση των εξαγόμενων τοπικών χαρακτηριστικών της εικόνας
- Δημιουργία λεξιλογίου με βάση τις συστάδες που δημιουργήθηκαν
- Δημιουργία των ιστογραμμάτων



Σχ. 4.10 Σχηματική Απεικόνιση του μοντέλου BOVW.

Το πρώτο βήμα για την δημιουργία BOVW χαρακτηριστικών είναι η εξαγωγή τοπικών χαρακτηριστικών από την εικόνα. Για την εξαγωγή τοπικών ιδιοτήτων από την εικόνα μπορεί να εφαρμοστούν αλγόριθμοι που εξάγουν τοπικούς περιγραφείς όπως για παράδειγμα ο αλγόριθμος SIFT ή SURF. Επιπροσθέτως, μπορεί να εφαρμοστεί και τμηματοποίηση της εικόνας και εξαγωγή χαρακτηριστικών από κάθε ξεχωριστό τμήμα που προκύπτει. Γενικότερα, στο παρόν στάδιο μπορεί να εφαρμοστεί οποιοσδήποτε αλγόριθμος για την εξαγωγή τοπικών χαρακτηριστικών. Μετά την εφαρμογή των αλγορίθμων εξαγωγής προκύπτουν διανύσματα χαρακτηριστικών από κάθε περιοχή της εικόνας.

Ύστερα από την εξαγωγή τοπικών χαρακτηριστικών (διανύσματα) από το σύνολο των εικόνων, τα διανύσματα χαρακτηριστικών συσταδοποιούνται με βάση τον αλγόριθμο *k-means*, η συσταδοποίηση είναι υπεύθυνη για την δημιουργία του λεξιλογίου. Ο αλγόριθμος *k-means* περιγράφεται στα εξής βήματα

1. Επιλογή τυχαίων αρχικών κέντρων.
2. Ανάθεση κάθε αντικειμένου στο πιο κοντινό κέντρο.
3. Υπολογισμός νέων κέντρων με βάση τη μέση τιμή των αντικειμένων που ανατέθηκαν σε κάθε κέντρο στο βήμα 2.
4. Επανάληψη βημάτων 2 και 3 μέχρι να μην υπάρχει αλλαγή στα κέντρα.

Η εφαρμογή του αλγορίθμου *k-means* συσταδοποιεί κάθε διάνυσμα χαρακτηριστικών σε ένα κέντρο, ο αριθμός των αρχικών κέντρων είναι επιλογή του χρήστη.

Τα κέντρα των διανυσμάτων που υπολογίστηκαν από τον αλγόριθμο *k-means*, αποτελούν τις κωδικές λέξεις (codewords) και με τη σειρά τους το σύνολο των κωδικών λέξεων αποτελούν το κωδικό λεξιλόγιο (codebook). Δηλαδή ο αριθμός των κέντρων υποδηλώνει το μέγεθος του λεξιλογίου.

Στο τελευταίο βήμα για τη δημιουργία χαρακτηριστικών BOVW, υπολογίζονται ιστογράμματα από την εικόνα που υποδηλώνουν τη συχνότητα εμφάνισης κάθε κωδικής λέξης στην εικόνα που εξετάζεται. Με αποτέλεσμα κάθε εικόνα να χαρακτηρίζεται από ένα ιστόγραμμα.

Η προσθήκη μια νέας εικόνας στο ήδη δημιουργημένο λεξιλόγιο ακολουθεί μια διαφορετική διαδικασία. Αρχικά εξάγονται τοπικοί περιγραφείς από την εικόνα, μετά την ολοκλήρωση της διαδικασίας εξαγωγής χαρακτηριστικών, υπολογίζεται με βάση ένα μέτρο απόστασης, η απόσταση μεταξύ των διανυσμάτων των χαρακτηριστικών από την καινούργια εικόνα που προστέθηκε στο λεξιλόγιο με τις ήδη υπάρχουσες κωδικές λέξεις του λεξιλογίου. Τα νέα χαρακτηριστικά προστίθενται στη κωδική λέξη που βρίσκεται πιο κοντά σύμφωνα με το μέτρο απόστασης.

Κεφάλαιο 5

Τμηματοποίηση εικόνας

5.1 Εισαγωγή

Στην υπολογιστική όραση η τμηματοποίηση μια εικόνας είναι ο χωρισμός της εικόνας σε περιοχές με παρόμοιες ιδιότητες [34] όπως χρώμα, υφή, φωτεινότητα κ.α. Εφαρμόζεται κυρίως για την ανίχνευση αντικειμένων και σχημάτων που περιέχονται σε μια εικόνα. Η τμηματοποίηση γεφυρώνει το σημασιολογικό κενό μεταξύ της επεξεργασίας χαμηλού επιπέδου και υψηλού επιπέδου. Οποιαδήποτε εφαρμογή που απαιτεί εύρεση, αναγνώριση και μέτρηση αντικειμένου χρησιμοποιεί κάποιου είδους τμηματοποίηση. Ένας αλγόριθμος τμηματοποίησης μπορεί να είναι αυτόματος, δηλαδή να μην χρειάζεται καμία παρεμβολή του χρήστη, καθώς και ημιαυτόματος όπου ο χρήστης πρέπει να υποδείξει μια περιοχή η το περίγραμμα μια περιοχής γύρω από την οποία επιθυμεί να πραγματοποιηθεί η τμηματοποίηση. Στις ιατρικές εικόνες η τμηματοποίηση εφαρμόζεται για να μελετήσουμε την ανατομία, διαφορές ανωμαλίες όπως κακώσεις και όγκους καθώς και τη μέτρηση διαστάσεων ενός ιστού για εντοπισμό όγκου.

Οι αλγόριθμοι για τμηματοποίηση εικόνας χωρίζονται στις εξής κατηγορίες [34]

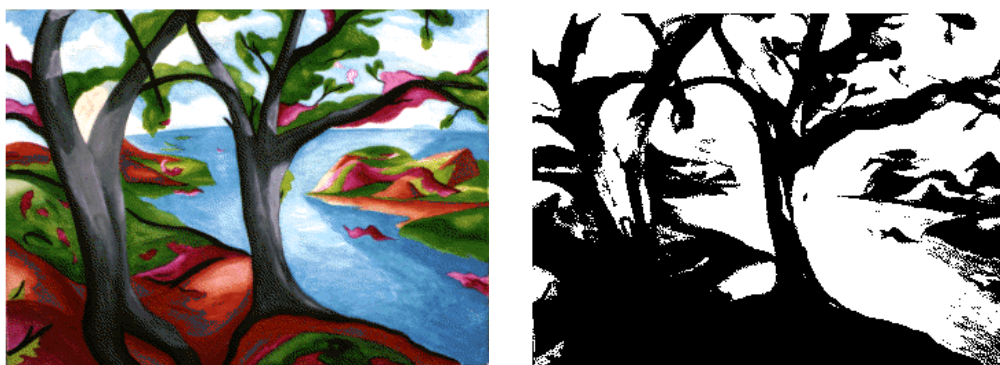
- Μέθοδοι βασισμένοι στο ιστογράμμα
- Μέθοδοι βασισμένοι στη κατωφλίωση
- Μέθοδοι βασισμένοι σε συστάδες

Οι παραπάνω κατηγορίες είναι οι κατηγορίες που βασίζονται η πλειοψηφία των αλγορίθμων τμηματοποίησης, υπάρχουν αλγόριθμοι που βασίζονται σε γράφους [57,58] σε μοντέλα με προκαθορισμένο σχήμα [59,60], σε συγχώνευση περιοχών [61,62], σε μοντέλα ενεργών περιγραμμάτων [63,64] κ.α.

5.2 Μέθοδοι τμηματοποίησης εικόνας

Οι μέθοδοι που βασίζονται στο ιστογράμμα είναι αποδοτικοί καθώς είναι υπολογιστικά φτηνοί επειδή χρειάζεται μόνο ένα πέρασμα της εικόνας. Η συγκεκριμένη τεχνική απαιτεί τον υπολογισμό του ιστογράμματος της εικόνας, για τον υπολογισμό μπορεί να χρησιμοποιηθεί η ένταση της φωτεινότητας ή το χρώμα. Η εικόνα χωρίζεται σε παρόμοιες περιοχές με βάση τις κορυφές και τις κοιλάδες που εμφανίζονται στο ιστογράμμα. Η τεχνική αυτή μπορεί να εφαρμοστεί επαναληπτικά και κάθε φορά η εικόνα να χωρίζεται σε μικρότερα τμήματα ώσπου να μην μπορεί να χωριστεί περαιτέρω. Ένα μειονέκτημα αυτής της μεθόδου είναι αρκετές φορές είναι δύσκολο να προσδιοριστούν οι κορυφές και οι κοιλάδες του ιστογράμματος.

Η κατωφλίωση [65] είναι η απλούστερη μέθοδος τμηματοποίησης εικόνων. Αρχικά επιλέγεται κατάλληλα μία τιμή κατωφλίωσης (μπορεί να είναι και περισσότερες από μία). Έπειτα κάθε εικονοστοιχείο της εικόνας κατανέμεται ανάλογα με την τιμή της έντασής του (συγκρίνεται με την τιμή κατωφλίωσης). Το κλειδί της επιτυχίας των αλγορίθμων που βασίζονται σε αυτή τη μέθοδο είναι η κατάλληλη επιλογή της τιμής κατωφλίωσης. Η τιμή της κατωφλίωσης μπορεί να βρεθεί με την χρήση διαφόρων μεθόδων όπως την μελετώντας τις κορυφές και τις κοιλάδες σε ένα ιστόγραμμα, με τη μέθοδο του Otsu [75].



Σχ 5.1 Εφαρμογή της μεθόδου κατωφλίωσης στην εικόνα.

Στις μεθόδους που βασίζονται σε συστάδες γίνεται ο διαχωρισμός των δεδομένων σε ένα σύνολο από κλάσεις. Ο πιο γνωστός αλγόριθμος για τη δημιουργία των συστάδων είναι ο *k-means* που αναφέραμε και παραπάνω. Για τον διαχωρισμό των εικονοστοιχείων (pixel) της εικόνας σε συστάδες, ο αλγόριθμος *k-means* αρχικοποιείται επιλεγώντας τυχαία κέντρα, τα εικονοστοιχεία της εικόνας τοποθετούνται στη συστάδα της οποίας το κέντρο απέχει λιγότερη απόσταση από τη τιμή του εικονοστοιχείου που εξετάζεται, στον αλγόριθμο *k-means* χρησιμοποιείται η Ευκλείδεια απόσταση. Στη συνέχεια, υπολογίζονται εκ νέου τα κέντρα των συστάδων και αλγόριθμος συνεχίζει επαναληπτικά ώσπου να μην έχουμε αλλαγές στα κέντρα. Ο αριθμός των αρχικών κέντρων μπορεί να επιλεγεί από το χρήστη ή τυχαία. Η επιτυχία της μεθόδου βασίζεται στην αρχική τοποθέτηση των κέντρων και στον αριθμό τους.



Σχ 5.2 Αποτέλεσμα του αλγορίθμου τμηματοποίησης *k-means*

5.3 Απλή Γραμμική Επαναληπτική Συσταδοποίηση

Ο αλγόριθμος της Απλής Γραμμικής Επαναληπτικής Συσταδοποίησης (Simple Linear Iterative Clustering, SLIC) [38] (Αλγ.) είναι μια παραλλαγή του αλγορίθμου k-means για συσταδοποίηση εικονοστοιχείων [40]. Ο συγκεκριμένος αλγόριθμος τμηματοποίησης είναι πιο αποδοτικός σε σύγκριση με άλλους αλγορίθμους τμηματοποίησης διότι έχει λιγότερες υπολογιστικές απαιτήσεις από συγχρόνους αλγορίθμους τμηματοποίησης και προσκολλάται καλύτερα στα περιγράμματα των αντικειμένων της εικόνας [39].

Σκοπός του αλγορίθμου είναι να δημιουργήσει παρόμοιες συστάδες εικονοστοιχείων σύμφωνα με το χρώμα και την θέση των εικονοστοιχείων στην εικόνα. Οι συστάδες των εικονοστοιχείων που δημιουργούνται ονομάζονται υπερεικονοστοιχεία (superpixels). Στον αλγόριθμο SLIC κάθε εικονοστοιχείο της εικόνας αναπαρίσταται με ένα διάνυσμα 5 χαρακτηριστικών (L_k, a_k, b_k, x, y) αυτά είναι : οι τιμές του εικονοστοιχείου k στα κανάλια L, a, b, του CIE-Lab χρωματικού χώρου και τις συντεταγμένες του εικονοστοιχείου στον οριζόντιο και κάθετο άξονα.

Ως παράμετρο ο αλγόριθμος SLIC δέχεται μόνο τον επιθυμητό αριθμό των υπερεικονοστοιχείων K που θα παραχθούν. Μια εικόνα με N εικονοστοιχεία θα περιέχει υπερεικονοστοιχεία μεγέθους N/K . Το μέγεθος των υπερεικονοστοιχείων θα είναι περίπου ίδιο, συνεπώς θα υπάρχει ένα κέντρο ενός υπερεικονοστοιχείου σε διάστημα περίπου $S = \sqrt{N/K}$. Ο αλγόριθμος ξεκινά αρχικοποιώντας τα K κέντρα των υπερεικονοστοιχείων τα οποία τοποθετούνται σε τυχαίες θέσεις. Έπειτα λαμβάνεται μια γειτονιά 3×3 γύρω από κάθε κέντρο και υπολογίζεται η κλίση (Σχέση 5.1) και τα κέντρα επανατοποθετούνται στην θέση με τη μικρότερη κλίση. Με τον υπολογισμό της κλίσης αποφεύγεται η τοποθέτηση των κέντρων σε ακμές της εικόνας και μειώνεται η πιθανότητα να επιλεγθεί ως κέντρο κάποιο εικονοστοιχείο που περιέχει θόρυβο. Η κλίση της εικόνας υπολογίζεται από τον τύπο:

$$G(x, y) = \|I(x + 1, y) - I(x - 1, y)\|^2 + \|I(x, y + 1) - I(x, y - 1)\|^2$$

Σχέση 5.1 Όπου $\|\cdot\|$ είναι η νορμα $L2$.

Επειδή η χωρική διάταξη των υπερεικονοστοιχείων θα είναι ένα υπερεικονοστοιχείο ανά περίπου S^2 , μπορούμε να συμπεράνουμε ότι τα εικονοστοιχεία που θα ανήκουν σε ένα υπερεικονοστοιχείο θα βρίσκονται σε μια περιοχή $2S \cdot 2S$ γύρω από το κεντρο του υπερεικονοστοιχείου που εξετάζεται. Μειώνοντας την περιοχή αναζήτησης ώστε να είναι ανάλογη του μεγέθους του υπερεικονοστοιχείου αλγόριθμος επιτυγχάνει να μείωση στο πλήθος των υπολογισμών που απαιτούνται για την ολοκλήρωση της τμηματοποίησης, καθώς δεν απαιτείται ο έλεγχος όλων των εικονοστοιχείων της εικόνας.

Στη συνέχεια ο αλγόριθμος συνεχίζει τοποθετώντας τα εικονοστοιχεία στον κατάλληλο πυρήνα με βάση τη μετρική απόστασης D_s . Η μετρική απόστασης D_s υπολογίζεται από τους τύπους:

$$D_{lab} = \sqrt{(L_k - L_i)^2 + (a_k - a_i)^2 + (b_k - b_i)^2} \quad (\Sigma\chi. 5.2)$$

$$Dxy = \sqrt{(X_k - X_i)^2 + (Y_k - Y_i)^2} \quad (\Sigma\chi. 5.3)$$

$$Ds = Dlab + \frac{m}{s} Dxy \quad (\Sigma\chi. 5.4)$$

Η μεταβλητή m ελέγχει το βαθμό που θα είναι συμπαγής κάθε υπερεικονοστοιχείο που θα δημιουργηθεί και παίρνει τιμές στο διάστημα $[1,20]$.

Η παραπάνω διαδικασία επαναλαμβάνεται για όλα τα εικονοστοιχεία της περιοχής $2S \cdot 2S$ γυρω από το κεντρο του υπερεικονοστοιχειου που εξετάζεται. Μόλις ανατεθούν όλα τα εικονοστοιχεία σε κάποιο κέντρο, υπολογίζεται εκ νέου τα κέντρα των υπερεικονοστοιχείων το οποίο είναι ο μέσος όρος των διανυσμάτων (L_k, a_k, b_k, x, y) που ανήκουν στο υπερεικονοστοιχείο, στην συνέχεια υπολογίζεται το σφάλμα (E) το οποίο είναι η απόσταση $L1$ μεταξύ του παλαιού και το νέου κέντρου, ο αλγόριθμος επαναλαμβάνετε μέχρι το σφάλμα (E) να είναι μικρότερο από ένα κατώφλι (T).

Algorithm 1: SLIC αλγόριθμος

Result: Τμηματοποιημένη Εικόνα

Αρχικοποίηση κέντρων με τη μικρότερη κλίση

while $E \leq T$ **do**

for για κάθε κέντρο **do**

 Ανάθεσε τα εικονοστοιχεία απο μια γειτονια $2S \times 2S$ σύμφωνα με την μετρική απόστασης

end

 Υπολόγισε τα νέα κέντρα το σφάλμα E (Απόσταση $L1$ μεταξύ νέου και παλιού κέντρου)

end



(A)

(B)

(Γ)

Σχήμα 5.3 Αρχική εικόνα (A), αποτέλεσμα αλγορίθμου SLIC (B), χρωματική απεικόνιση των υπερεικονοστοιχείων (Γ)

Κεφάλαιο 6

Ευρετήρια

6.1 Εισαγωγή

Μετά το στάδιο της εξαγωγής χαρακτηριστικών από την εικόνα, έχουν παραχθεί πολυδιάστατα δεδομένα τα οποία πρέπει να τα διαχειριστεί το σύστημα ανάκλησης εικόνων. Για την αποθήκευση των δεδομένων χρησιμοποιούνται ευρετήρια, ένα ιδανικό ευρετήριο πρέπει να έχει μικρό χρόνο απόκρισης, χαμηλή πολυπλοκότητα ακόμα και για μεγάλο αριθμό διαστάσεων, χαμηλές απαιτήσεις σε αποθηκευτικό χώρο, δυνατότητα παράλληλης αναζήτησης και δυνατότητα προσθήκης ή αφαίρεσης στοιχείων μετά τη δημιουργία του ευρετηρίου. Ένα ευρετήριο είναι μια δομή αρχείου που κάνει πιο αποδοτική την αναζήτηση μιας εγγραφής σε ένα αρχείο [34].

Η εξαγωγή ιδιοτήτων από την εικόνα και ειδικότερα τα χαρακτηριστικά υψηλού επιπέδου που μεταφέρουν και σημασιολογικές πληροφορίες οδηγούν στην δημιουργία πολυδιάστατων δεδομένων. Η «κατάρρα» των πολυδιάστατων δεδομένων (curse of dimensionality) είναι το φαινόμενο το οποίο προκαλείται όταν αυξάνεται η διάσταση των δεδομένων, ο όγκος των διαστάσεων μεγαλώνει με γρήγορους ρυθμούς και τα υπάρχοντα δεδομένα γίνονται αραιά μέσα στις διαστάσεις, για το λόγο αυτό συχνά απαιτούνται τεχνικές μείωσης των διαστάσεων τους. Στόχος της μείωσης των διαστάσεων των δεδομένων είναι να μεταμορφώσει τα πολυδιάστατα και αραιά δεδομένα, σε δεδομένα μικρότερης διάστασης. Τα πλεονέκτημα της μείωσης των διαστάσεων είναι ότι μειώνει το χρόνο και το χώρο αποθήκευσης, βελτιώνει την απόδοση των συστημάτων και βοηθάει στην οπτικοποίηση των δεδομένων.

Η διανυσματική περιγραφή σε μια συλλογή εικόνων προς ανάκληση αποθηκεύεται συνήθως σε σχεσιακές βάσεις δεδομένων στις οποίες η αναζήτηση διευκολύνεται με τη χρήση ευρετηρίων.

6.2 Σχεσιακές Βάσεις Δεδομένων

Το σχεσιακό μοντέλο δεδομένων αναπτύχθηκε από το Codd το 1970, η απλή και κατανοητή δομή είναι δυο από τα αίτια της τεράστιας απήχησης του μοντέλου για τη δημιουργία Βάσεων Δεδομένων. Οι βασικοί στόχοι του μοντέλου είναι η ανεξαρτησία των δεδομένων, η αποφυγή πλεονασμού των δεδομένων, διατήρηση της ακεραιότητας και η υποστήριξη ανάπτυξης γλωσσών χειρισμού δεδομένων οι οποίες στηρίζονται στη θεωρία των συνόλων και διευκολύνουν την ανάπτυξη ερωτημάτων προς το Σύστημα Διαχείρισης Βάσεων Δεδομένων. Σήμερα, η βασική γλώσσα που υποστηρίζουν τα συστήματα είναι η γλώσσα SQL (Structured Query Language).

Στο σχεσιακό μοντέλο τα δεδομένα αναπαριστούνται σε πίνακες ή σχέσεις. Κάθε πίνακας αποτελείται από ένα σύνολο γραμμών και στηλών. Μια εγγραφή στο πίνακα καταλαμβάνει μια γραμμή και ονομάζεται πλειάδα, ο αριθμός των πλειάδων ονομάζεται πληθυκότητα. Οι στήλες του πίνακα ορίζουν τις ιδιότητες της κάθε εγγραφής. Ένα χαρακτηριστικό ενός πίνακα ή ένα σύνολο χαρακτηριστικών ονομάζεται πρωτεύον κλειδί το οποίο έχει μοναδική τιμή και διακρίνει τις εγγραφές του πίνακα.

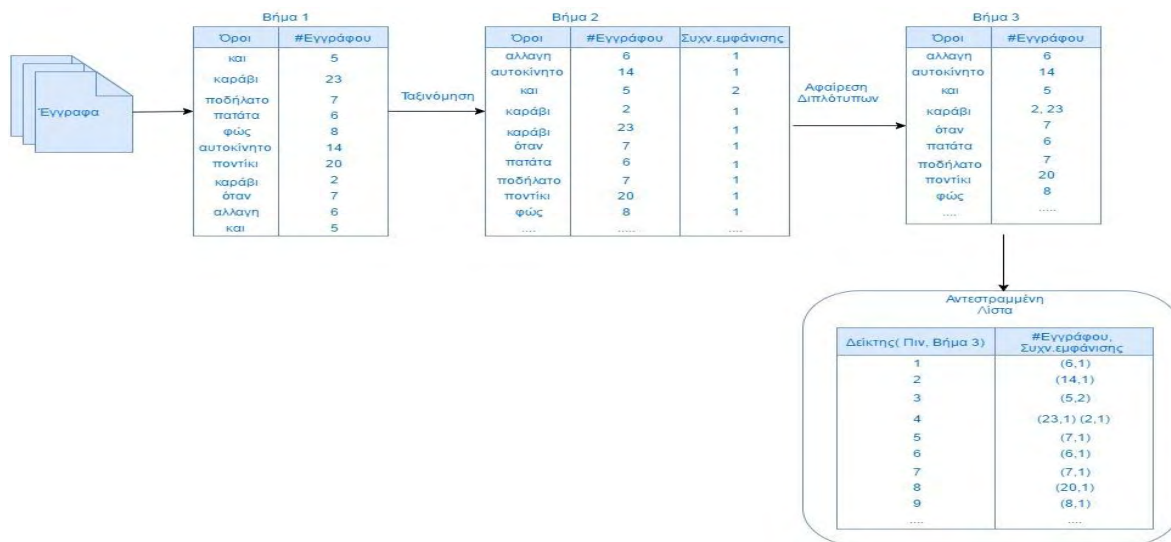
Σε συστήματα ανάκλησης εικόνων, παρόμοιες εικόνες μπορούν να αποτελέσουν ένα πίνακα στο σχεσιακό μοντέλο, για παράδειγμα εικόνες που περιέχουν αυτοκίνητα ως κύριο μέρος της εικόνας και τα εξαγόμενα χαρακτηριστικά να αναπαριστούν τις γραμμές του πίνακα.

Για την γρήγορη αναζήτηση και ανάκληση πληροφοριών από μια σχεσιακή βάση δεδομένων, έχουν προταθεί τεχνικές για τη δημιουργία ευρετηρίων. Ένα αρχείο ευρετηρίου περιέχει τα πρωτεύοντα κλειδιά των εγγράφων ενός πίνακα και τη θέση των αντίστοιχων εγγράφων στο πίνακα, επιτρέποντας έτσι τη γρήγορη αναζήτηση και ανάκληση των εγγράφων. Δημοφιλείς δομές για την δημιουργία ευρετηρίων είναι τα δένδρα και τα αντεστραμμένα αρχεία (inverted – files) οι οποίες αναφέρονται στη συνέχεια του κεφαλαίου.

6.3 Αντεστραμμένα Αρχεία (Inverted-File)

Η μέθοδος αυτή, όπως και άλλες τεχνικές που εφαρμόζονται στα συστήματα ανάκλησης εικόνων προέρχεται από την ανάκληση πληροφοριών (information retrieval). Η τεχνική των αντεστραμμένων αρχείων (Εικ. 7.1) χρησιμοποιείται συνήθως σε μηχανές αναζήτησης. Τα αντεστραμμένα αρχεία εφαρμόζονται συχνά για τη δημιουργία ευρετηρίου σε συστήματα ανάκλησης εικόνων όταν έχουμε εξαγωγή τοπικών χαρακτηριστικών ή χαρακτηριστικών *bag of visual words*.

Δοθέντος μιας συλλογής εγγράφων, αποδομούνται οι λέξεις που περιέχονται στα έγγραφα και τοποθετούνται σε μια λίστα μαζί με το ID του εγγράφου (Βήμα 1, Εικ.6.1). Στο επόμενο βήμα (Βήμα 2) της κατασκευής του αντεστραμμένου αρχείου οι όροι των εγγράφων ταξινομούνται και στη συνέχεια συνενώνονται οι εμφανίσεις του ίδιου όρου σε περισσότερα έγγραφα. Στόχος των αντεστραμμένων αρχείων είναι να δημιουργηθεί μια λίστα η οποία θα περιέχει όλους τους όρους που περιέχονται στην αρχική συλλογή των εγγράφων μαζί με το ID του εγγράφου καθώς και τη συχνότητα εμφάνισης του όρου σε κάθε έγγραφο.



Εικ. 6.1 Αντεστραμμένα Αρχεία

Η συχνότητα εμφάνισης (term frequency, *tf*) υπολογίζει τον αριθμό των εμφανίσεων της λέξης ή όρου (term), σε ένα έγγραφο (document), η πιο απλή μορφή της είναι το άθροισμα των εμφανίσεων της λέξεις ($f_{t,d}$), αλλά μπορεί να πάρει και άλλες μορφές μετά από κανονικοποίηση της τιμής της.

Η αντεστραμμένη συχνότητα εμφάνισης (inverse term frequency, *idf*) παρουσιάζει την εμφάνιση ενός όρου σε όλα τα έγγραφα, ουσιαστικά αντιπροσωπεύει ποσό συχνός η σπάνιος είναι ένας όρος σε όλα τα διαθέσιμα έγγραφα. Η αντεστραμμένη συχνότητα υπολογίζεται από τον τύπο (Σχ. 6.1):

$$idf(t, D) = \log \frac{N}{1 + |\{d \in D : t \in d\}|} \quad (\text{Σχ. 6.1})$$

Όπου N είναι ο αριθμός των εγγράφων που διαθέτουμε, και ο παρονομαστής συμβολίζει τον αριθμό των εγγράφων που εμφανίζετε όρος t

Σε πολλές μηχανές αναζήτησης, συνηθίζεται ο συνδυασμός των δυο αυτών συχνοτήτων και να υπολογίζεται η συχνότητα εμφάνισης- αντεστραμμένη συχνότητα εμφάνισης (*tf-idf*). Η συχνότητα υπολογίζεται από τη σχέση (Σχ. 6.2):

$$tfidf(t, d, D) = tf(t, d) * idf(t, D) \quad (\text{Σχ. 6.2})$$

Υψηλή τιμή τις *tf-idf* αποδεικνύει υψηλή συχνότητα εμφάνισης σε ένα δοθέν έγγραφο και χαμηλή συχνότητα εμφάνισης του όρου στη συλλογή των εγγράφων. Όταν ένας όρος εμφανίζεται σε περισσότερα έγγραφα της συλλογής η τιμή της *tf-idf* θα πλησιάζει το μηδέν, λόγω του λογάριθμου που εμπεριέχεται στον υπολογισμό της *idf*, η τιμή της *tf-idf* θα είναι μεγαλύτερη ίση του μηδενός.

6.4 Δένδρα

Αρκετές τεχνικές δημιουργίας ευρετηρίου απεικονίζουν τις διαστάσεις των δεδομένων σε μορφές ιεραρχικών δένδρων. Ένα δένδρο είναι συνεκτικό άκυκλο γράφημα το οποίο περιέχει κορυφές (κόμβους) και ακμές. Τα δένδρα είναι οργανωμένα με τέτοιο τρόπο ώστε οι διαδικασίες αναζήτησης και εισαγωγή κόμβων να είναι αποδοτικές. Μερικοί ορισμοί για τα δένδρα:

- **Φύλλο:** κόμβος χωρίς απογόνους
- **Πατέρας :** όταν υπάρχει ακμή από ένα κόμβο u σε ένα κόμβο n τότε ο u λέγεται **πατέρας** ή **πρόγονος** του n και ο n λέγεται **παιδί** του u .
- **Βαθμός κόμβου:** ο αριθμός των παιδιών του.
- **Βαθμός δένδρου :** ο μέγιστος από τους βαθμούς των κόμβων του.
- **Διαδρομή ή Μονοπάτι :** μια ακολουθία κόμβων v_1, v_2, \dots, v_k όπου κάθε v_i είναι ο πατέρας του v_{i+1} , **μήκος μονοπατιού** είναι ο αριθμός ακμών της διαδρομής που ακολουθούμε.
- **Ύψος ενός κόμβου:** το μήκος του μέγιστου μονοπατιού από τον κόμβο προς κάποιο φύλλο.
- **Ύψος δένδρου :** το ύψος της ρίζας.

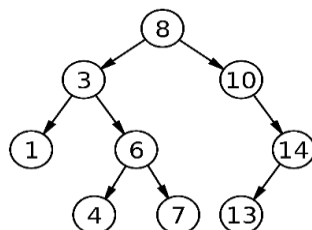
Τα δένδρα έχουν πολλές παραλλαγές όπως τα Δυαδικά δένδρα, R-δένδρα, B- δένδρα, Δένδρα-KΔ (K- διαστάσεων), μελάνυθρα δένδρα κ.α. παρακάτω αναφέρουμε ενδεικτικά μερικούς τύπους δένδρων.

6.4.1 Δυαδικά Δένδρα αναζήτησης

Ένα δένδρο ονομάζεται δυαδικό δένδρο αν όλοι οι κόμβοι του έχουν βαθμό ≤ 2 . Το ύψος ενός δυαδικού δένδρου με n κόμβους είναι $\log_2 n$. Δυαδικά Δένδρα Αναζήτησης (Εικ. 6.2) ονομάζονται τα δένδρα τα οποία κάθε κόμβος u πρέπει να ικανοποιεί τα εξής:

- Οι τιμές του αριστερού υποδένδρου του u να είναι μικρότερες από την τιμή του u
- Οι τιμές του δεξιού υποδένδρου του u να είναι μεγαλύτερες από την τιμή του u

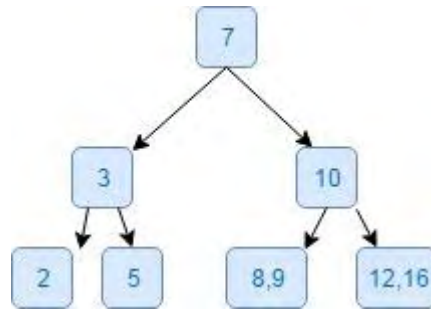
Τα Δυαδικά Δένδρα Αναζήτησης επιταχύνουν την αναζήτηση σε ένα σύνολο στοιχείων καθώς ανάλογα με την τιμή που αναζητούμε ανατρέχουμε στο δεξιό ή αριστερό υποδένδρο.



Εικ. 6.2 Δυαδικό δένδρο αναζήτησης

6.4.2 B-δένδρα

B-δένδρο (Εικ.6.3) ονομάζεται ένα ισοζυγισμένο δένδρο αναζήτησης στο οποίο κάθε κόμβος έχει προκαθορισμένο αριθμό παιδιών. Στην εισαγωγή και διαγραφή κάποιου κόμβου οι εναπομείναντες κόμβοι διαχωρίζονται ή συνενώνονται για να έχουν τον προκαθορισμένο αριθμό παιδιών. Μια παραλλαγή των B δένδρων είναι τα B⁺ δένδρα, σε αντίθεση με τα B-δένδρα τα δεδομένα στα B⁺ δένδρα αποθηκεύονται μόνο στις ρίζες του δένδρου και όχι στους εσωτερικούς κόμβους του δένδρου.

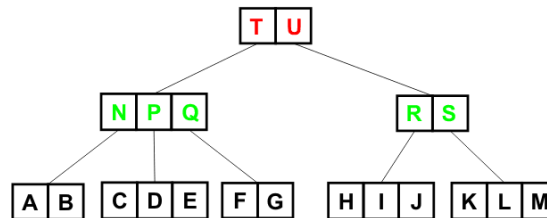
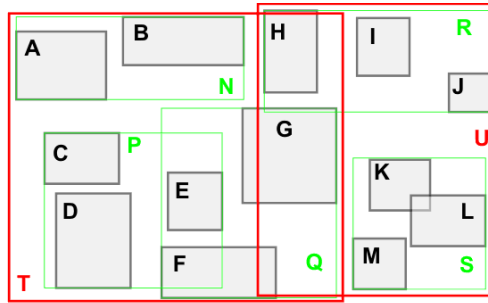


Εικ. 6.3 B- δένδρο μέγιστου βαθμού 3

6.4.3 Δένδρα Περιοχής

Η ιδέα των δένδρων περιοχής (R- tree) είναι να ομαδοποιήσουν γειτονικά δεδομένα με ένα και να αναπαρασταθούν με ένα παραλληλόγραμμο. Τα δένδρα R-Tree είναι μια παραλλαγή των B-δένδρων (B- Tree). Όπως και τα B-δένδρα, τα δένδρα περιοχής είναι ισοζυγισμένα δένδρα αναζήτησης. Τα φύλλα του δένδρου περιέχονται δείκτες προς τα πολυδιάστατα δεδομένα, κάθε κόμβος περιέχει το πολύ M εγγραφές από δυάδες $(I, \text{δείκτης})$. Το I είναι ένα n -διάστατο παραλληλόγραμμο που περικλείει ένα υποσύνολο του n -διάστατου χώρου και ο δείκτης δείχνει σε ένα κόμβο παιδί (Εικ. 6.4).

Τα δένδρα περιοχής είναι αποτελεσματικά στα περισσότερα πραγματικά δεδομένα, αλλά όσο αυξάνονται οι διαστάσεις η απόδοση υποβιβάζεται. Όταν τα δεδομένα οργανώνονται σε δένδρα περιοχής ο υπολογισμός των k -κοντινότερων γειτόνων μπορεί να γίνει αποδοτικά εφαρμόζοντας χωρικές συνενώσεις.



Εικ. 6.4 Δένδρο περιοχής

6.5 Μέθοδοι δημιουργίας ευρετηρίων σε συστήματα CBIR

Πέραν από τους δημοφιλείς τρόπους δημιουργίας ευρετηρίου που αναλύσαμε παραπάνω, έχουν αναπτυχθεί μέθοδοι για συγκεκριμένα συστήματα CBIR. Οι Zhihua et al. [41] πρότεινε ένα σύστημα ανάκλησης όπου το ευρετήριο αποθηκεύεται σε έναν εξυπηρετητή στο διαδίκτυο (cloud server). Η αποθήκευση του ευρετηρίου στον εξυπηρετητή του διαδικτύου παρέχει ευκολία στην πρόσβαση καθώς είναι διαθέσιμο οποιαδήποτε στιγμή και αξιόπιστο διότι τα δεδομένα δεν χάνονται σε περίπτωση δυσλειτουργίας του συστήματος. Για τη δημιουργία του ευρετηρίου και την διατύπωση των ερωτημάτων χρησιμοποιούνται αυξημένα επίπεδα ασφαλείας εφαρμόζοντας κρυπτογράφηση, καθώς οι εικόνες και ιδιαίτερα οι ιατρικές εικόνες περιέχουν ευαίσθητο περιεχόμενο.

Οι Meglin et al. [18] χρησιμοποιεί μια παραλλαγή των δένδρων σε συνεργασία με τη μέθοδο των αντεστραμμένων αρχείων για την δημιουργία του ευρετηρίου. Τα εξαγόμενα χαρακτηριστικά κβαντίζονται σε έναν αριθμό κόμβου (ID) το οποίο αντιπροσωπεύει το μονοπάτι από τη ρίζα του δένδρου στο φύλλο. Τα φύλλα των κόμβων περιέχουν αντεστραμμένα αρχεία που καταγράφουν τα IDs που περιέχονται σε ένα στιγμιότυπο του κόμβου.

Οι δομές ευρετηρίων που αναφέρθηκαν έχουν ως σκοπό την ελαχιστοποίηση επισκέψεων σε κόμβους της δομής και των δεδομένων. Όμως πολλές φορές, με την αύξηση των διαστάσεων του χώρου έχουν χειρότερη απόδοση από τη σειριακά αναζήτηση. Στη σειριακά αναζήτηση δεν υπάρχει κάποια δομή για την αποθήκευση τα χαρακτηριστικών, η αναζήτηση για παρόμοιες εικόνες πραγματοποιείται με τον έλεγχο όλων των αποθηκευμένων δεδομένων για εντοπισμό όμοιων χαρακτηριστικών και κατά συνέπεια εικόνων.

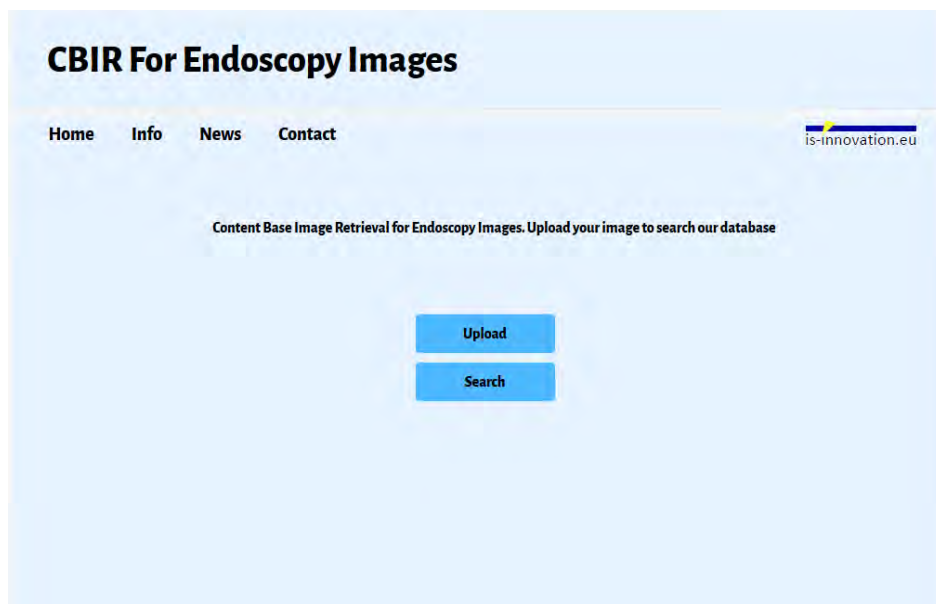
Προτεινόμενο σύστημα ανάκλησης ενδοσκοπικών εικόνων βάση περιεχομένου

7.1 Εισαγωγή

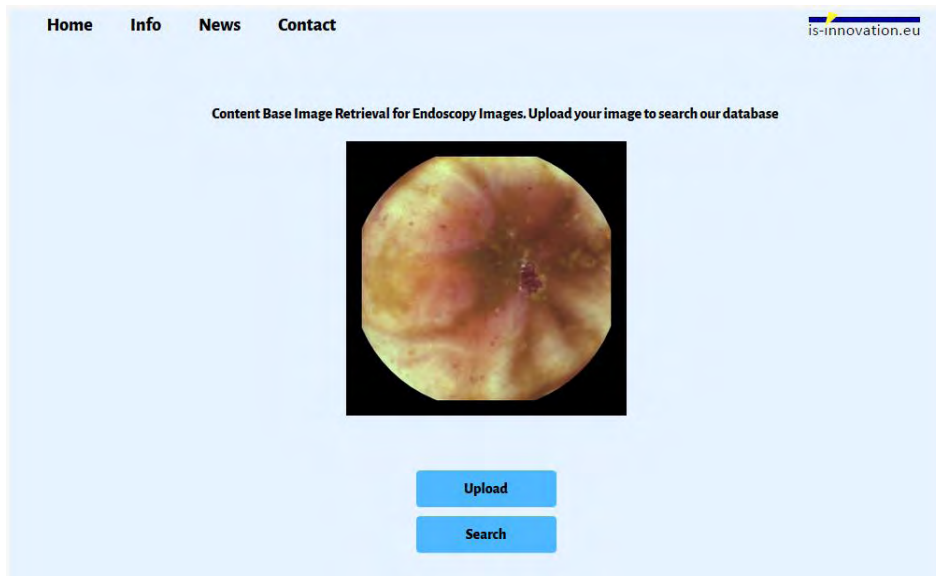
Στο κεφάλαιο αυτό παρουσιάζεται το σύστημα το οποίο κατασκευάστηκε στα πλαίσια αυτής της ερευνητικής εργασίας. Κύριος στόχος της εργασίας είναι η ανάπτυξη ενός συστήματος ανάκλησης με βάση το περιεχόμενο με κατάλληλα χαρακτηριστικά για τη βέλτιστη απόδοση σε εικόνες που έχουν εξαχθεί από ενδοσκοπική κάψουλα. Τα χαρακτηριστικά που αναπτύχθηκαν ακολουθούν το μοντέλο *bag of visual words* επιπροσθέτως έχει δοθεί έμφαση στην επεκτασιμότητα για περαιτέρω ανάπτυξη του συστήματος.

Η υλοποίηση του συστήματος πραγματοποιήθηκε με τη γλώσσα προγραμματισμού JAVA. Η γλώσσα προγραμματισμού JAVA επιλέχθηκε διότι προφέρει μεταφερτότητα του λογισμικού, έχει καλή απόδοση και είναι ευρέως διαδομένη, διευκολύνοντας έτσι την επέκταση του συστήματος. Το προτεινόμενο σύστημα έχει τη δυνατότητα να χρησιμοποιηθεί ως πλατφόρμα ανάπτυξης νέων μεθοδολογιών και αλγορίθμων εξαγωγής χαρακτηριστικών.

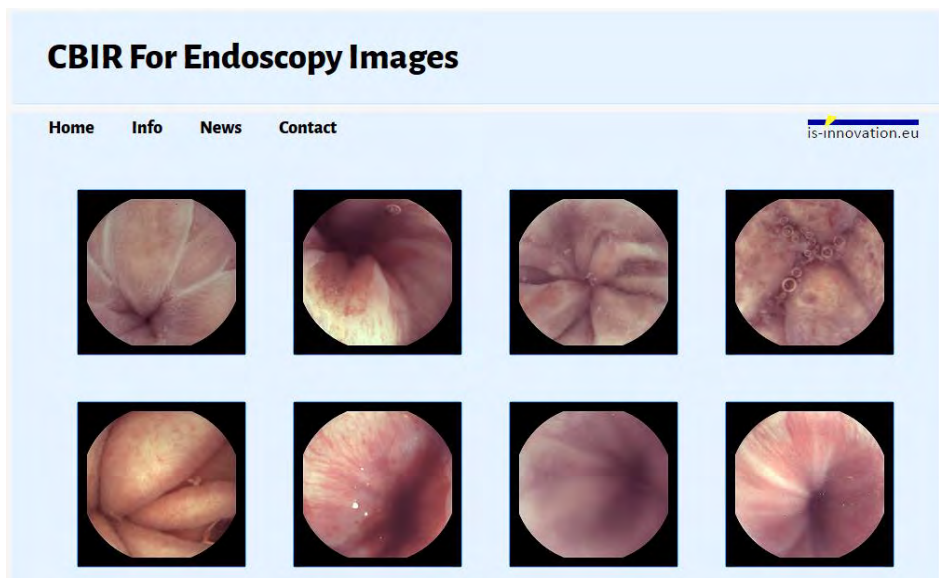
Η διεπαφή του χρήστη με το σύστημα ανάκλησης, γίνεται μέσω του διαδικτύου, όπου ο χρήστης εισάγει την εικόνα - επερώτηση και το σύστημα του επιστέφει παρόμοιες εικόνες. Η διαδικτυακή διεπαφή επιτρέπει την πρόσβαση των χρηστών στο σύστημα από όλο το κόσμο και σχεδιάστηκε με στόχο την ευχρηστία του (Εικ. 7.1 - 7.3).



Εικ. 7.1 Αρχική σελίδα διεπαφής



Εικ. 7.2 Ανέβασμα εικόνας επερώτησης



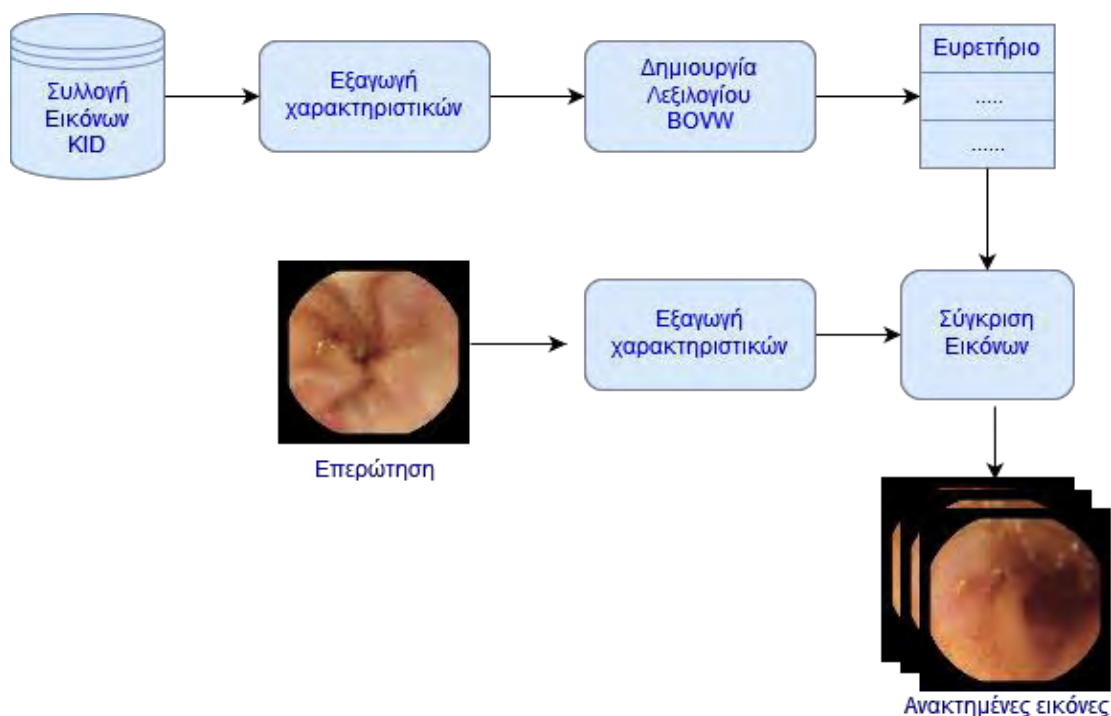
Εικ. 7.3 Ανακλημένες εικόνες

7.2 Αρχιτεκτονική του συστήματος

Το προτεινόμενο σύστημα με βάση το περιεχόμενο βασίζεται στο γενικό μοντέλο των συστημάτων CBIR. Το προτεινόμενο σύστημα (Διαγρ. 7.1), διατρέπει όλες τις εικόνες της βάσης δεδομένων και εξάγει χαρακτηριστικά με βάση το μοντέλο *bag of visual words* από τις αποθηκευμένες εικόνες που διατηρούμε στη βάση δεδομένων. Στο στάδιο της εξαγωγής ιδιοτήτων απαιτείται και η τμηματοποίηση της εικόνας. Η τμηματοποίηση της εικόνας γίνεται με τον αλγόριθμο SLIC τον οποίο αναλύσαμε παραπάνω. Ο αλγόριθμος εξαγωγής χαρακτηριστικών αναλύεται σε βάθος σε επόμενο τμήμα του κεφαλαίου. Στη συνέχεια γίνεται η δημιουργία του ευρετηρίου για την οργάνωση των εξαγόμενων χαρακτηριστικών.

Τα ερωτήματα που γίνονται στο σύστημα ακολουθούν το πρότυπο της επερώτησης με παράδειγμα. Ο χρήστης τοποθετεί την εικόνα αναζήτησης στην διεπαφή του συστήματος,

μετά την εισαγωγή της εικόνας το σύστημα προχωράει στην εξαγωγή χαρακτηριστικών. Τα χαρακτηριστικά που εξήχθησαν συγκρίνονται με αυτά που βρίσκονται στο ευρετήριο του συστήματος ανάκλησης. Η σύγκριση γίνεται μέσω μιας μετρικής απόστασης. Μετά τη σύγκριση των χαρακτηριστικών της εικόνας - επερώτησης και των χαρακτηριστικών του ευρετηρίου επιστρέφονται ταξινομημένες εικόνες. Η ταξινόμηση των εικόνων γίνεται με βάση την απόσταση των χαρακτηριστικών που έχουν εξαχθεί.



Διαγρ 7.1 Απεικόνιση του προτεινόμενου συστήματος CBIR

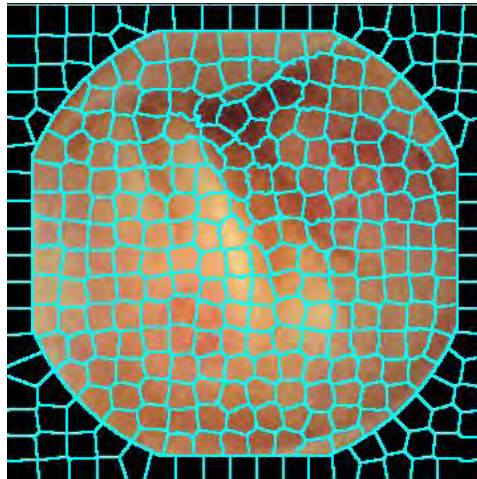
7.3 Αλγόριθμος εξαγωγής χαρακτηριστικών

Τα χαρακτηριστικά που δημιουργούνται στο στάδιο εξαγωγής χαρακτηριστικών έχουν τη μορφή μονοδιάστατων διανυσμάτων $F = (f_1, f_2, \dots, f_n)$. Τα f_i αναπαριστούνται με αριθμούς κινητής υποδιαστολής διπλής ακρίβειας.

Ο αλγόριθμος (Αλγ 7.1) που προτείνεται για την εξαγωγή χαρακτηριστικών με σκοπό την ανάκληση εικόνων με βάση το περιεχόμενο, ξεκινάει με τον μετασχηματισμό της εικόνας από τον *RGB* χώρο χρώματος στον *CIE-Lab* χώρο. Στην συνέχεια λαμβάνει χώρα η τμηματοποίηση της εικόνας. Για την διαδικασία της τμηματοποίησης εφαρμόζεται ο αλγόριθμος *SLIC* (Simple Linear Iterative Clustering) ο οποίος χωρίζει τη εικόνα σε ομάδες από εικονοστοιχεία τα οποία ονομάζονται υπερεικονοστοιχεία (superpixels).

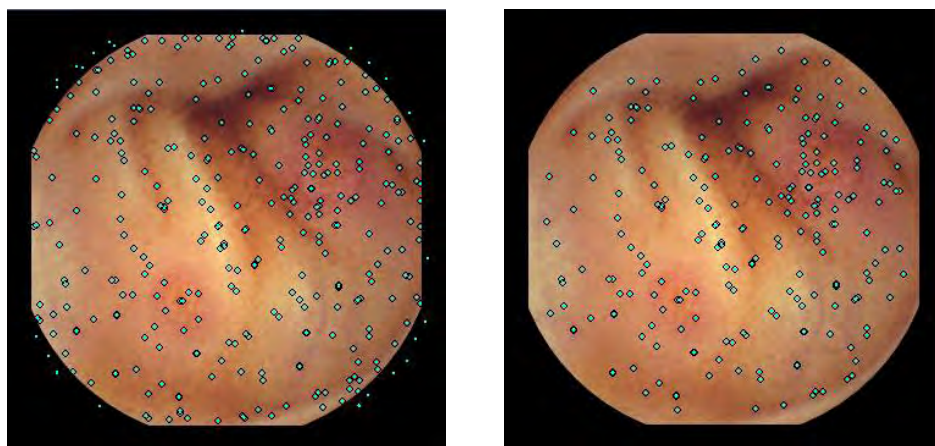
Μετά την τμηματοποίηση γίνεται εύρεση σημείων ενδιαφέροντος (salient points) στην εικόνα εφαρμόζοντας τον αλγόριθμο εξαγωγής τοπικών χαρακτηριστικών *SURF*. Τα σημεία *SURF* εξάγονται χρησιμοποιώντας το κανάλι a του χρωματικού χώρου *CIE-Lab*. Το κανάλι a

χρησιμοποιείται διότι στο χρωματικό χώρο *CIE-Lab* η χρωματική πληροφορία για το κόκκινο χρώμα βρίσκεται στο *a* κανάλι.



Σχήμα 7.1 Αποτέλεσμα αλγορίθμου *SURF* σε μια ενδοσκοπική εικόνα

Λόγω του χαρακτηριστικού μαύρου περιγράμματος των ενδοσκοπικών εικόνων πολλά σημεία ενδιαφέροντος έχουν «τραβηχτεί» προς το περίγραμμα της εικόνας. Τα σημεία αυτά δεν περιχέουν κάποια χρήσιμη πληροφορία και πρέπει να αποκοπούν από τον υπολογισμό των τελικών χαρακτηριστικών. Για το φιλτράρισμα των σημείων ενδιαφέροντος *SURF*, λαμβάνουμε ένα παραθύρο γύρω από κάθε σημείο *SURF*. Στα πλαίσια του παραθύρου αριθμείτε ο αριθμός των μαύρων εικονοστοιχείων, αν ο αριθμός των μαύρων εικονοστοιχείων είναι μεγαλύτερος από ένα ποσοστό ρ του συνολικού αριθμού εικονοστοιχείων του παραθύρου, τότε το σημείο *SURF* δεν το λαμβάνουμε υπόψη μας.



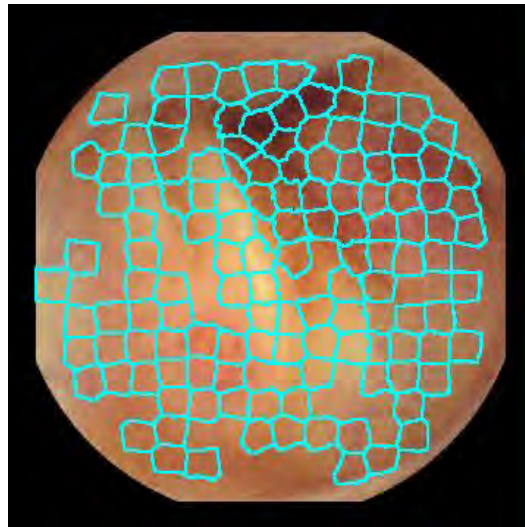
(A)

(B)

Σχήμα 7.2 Εύρεση σημείων *SURF* πριν το φιλτράρισμα (A), μετά το φιλτράρισμα (B)

Ο αλγόριθμος συνεχίζει με την εύρεση των υπερεικονοστοιχείων που περιχέουν μέσα τους τα φιλτραρισμένα σημεία ενδιαφέροντος *SURF* (salient superpixel) Σχ. 8.3. Μετά τον εντοπισμό των υπερεικονοστοιχείων ενδιαφέροντος εξάγονται, τα εξής χαρακτηριστικά: Salient L, Salient a, Salient b, min L, min a, min b, max L, max a, max b. Τα χαρακτηριστικά

Salient είναι οι τιμές των καναλιών L,a,b του χρωματικού χώρου *CIE-Lab* στη θέση του πιο ισχυρού *SURF* σημείου που βρίσκεται στο υπερεικονοστοιχείο που εξετάζουμε και *min*, *max* είναι η μέγιστη και η ελάχιστη τιμή του καναλιού που εξετάζουμε.



Σχήμα 7.3 Salient Superpixels

Ψευδοκώδικας

ΑΛΓΟΡΙΘΜΟΣ: Εξαγωγή Χαρακτηριστικών

ΑΡΧΗ

1. **ΜΕΤΑΒΛΗΤΕΣ** $F[]$, Τιμή1 , Τιμή2 , SL, Sa, Sb, minL ,mina , minb, maxL, maxa, maxb, I, $\Sigma[]$, ILab[], dist
2. $I \leftarrow$ Διάβασε Εικόνα
3. **ILab[L,a,b]** \leftarrow Μετατροπή από RGB σε Lab
4. Τμηματοποίηση I με τον αλγόριθμο SLIC
5. Εξαγωγή σημείων ενδιαφέροντος από την **ILab[a]** εφαρμόζοντας τον αλγόριθμο SURF
6. **Για κάθε** σημείο SURF
7. Λήψη παραθύρου γύρω από το σημείο
8. **Τιμη1** \leftarrow Άθροισμα μαύρων εικονοστοιχείων
9. **Αν** **Τιμη1** $>$ ρ * Σύνολο εικονοστοιχείων παραθύρου
10. **Τότε** Απόρριψη σημείου
11. **Τέλος Αν**
12. **Τέλος Επανάληψης**
13. **Για κάθε** υπερεικονοστοιχείο
14. **Αν** περιέχει σημεία SURF
15. Διατήρηση υπερεικονοστοιχείου
16. **Αλλιώς** απόρριψη υπερεικονοστοιχείου
17. **Τέλος Αν**
18. **Τέλος Επανάληψης**
19. **Για κάθε** διατηρημένο υπερεικονοστοιχείο
20. $(x,y) \leftarrow$ εύρεση συντεταμένων ισχυρότερου σημείου SURF

21. **SL** <- **ILab[L]** στην θέση (x,y)
22. **Sa** <- **ILab[a]** στην θέση (x,y)
23. **Sb** <- **ILab[b]** στην θέση (x,y)
24. **MinL** <- υπολογισμός μικρότερης τιμής του καναλιού L στο υπερεικονοστοιχείο
25. **Mina** <- υπολογισμός μικρότερης τιμής του καναλιού α στο υπερεικονοστοιχείο
26. **Minb** <- υπολογισμός μικρότερης τιμής του καναλιού b στο υπερεικονοστοιχείο
27. **MaxL** <- υπολογισμός μεγαλύτερης τιμής του καναλιού L στο υπερεικονοστοιχείο
28. **Maxa** <- υπολογισμός μεγαλύτερης τιμής του καναλιού α στο υπερεικονοστοιχείο
29. **Maxb** <- υπολογισμός μεγαλύτερης τιμής του καναλιού b στο υπερεικονοστοιχείο
30. **Fi** <- [**SL, Sa, Sb, minL, mina, minB, maxL, maxa, maxb**]
31. Τέλος Επανάληψης

Τέλος

7.4 Δημιουργία Λεξιλογίου

Μετά την ολοκλήρωση της εξαγωγής των χαρακτηριστικών, έχουν δημιουργηθεί διανύσματα χαρακτηριστικών για κάθε υπερεικονοστοιχείο της εικόνας. Με βάση το πρότυπο της τεχνικής bag of visual words αυτά τα διανύσματα θα δημιουργήσουν τις κωδικές λέξεις του λεξιλογίου.

Τα διανύσματα από κάθε υπερεικονοστοιχείο συσταδοποιούνται με βάση τον αλγόριθμο *k-means* σε *K* κέντρα. Ο αριθμός των κέντρων είναι μια παράμετρος που εξετάζεται στο πειραματικό στάδιο της εργασίας. Έπειτα της συσταδοποίησης έχει δημιουργηθεί το κωδικό λεξιλόγιο, η διαφορά της προτεινόμενης μεθόδου είναι ότι στη δημιουργία του λεξιλογίου χρησιμοποιούνται οι οπτικές πληροφορίες των υπερεικονοστοιχείων και όχι τοπικών περιγραφών του αλγορίθμου SURF ή SIFT. Χρησιμοποιώντας τα υπερεικονοστοιχεία, εξάγονται χαρακτηριστικά από άμορφες περιοχές της εικόνας που περιέχουν σημαντικές οπτικές πληροφορίες για τη δομή της εικόνας.

Στη συνέχεια, όπως και στην τεχνική των bag of visual word υπολογίζονται ιστογράμματα που αποδεικνύουν τη συχνότητα εμφάνισης κάθε κωδικής λέξης στην εξεταζόμενη εικόνα. Τα ιστογράμματα που υπολογίζονται θα χρησιμοποιηθούν για τη σύγκριση ομοιότητας μεταξύ των εικόνων.

7.5 Δημιουργία Ευρετηρίου

Τα εξαγόμενα χαρακτηριστικά πρέπει να οργανωθούν σε μια δομή ευρετηρίου για την γρήγορη και αποδοτική αναζήτηση των εικόνων. Η δημιουργία του ευρετηρίου του προτεινόμενου συστήματος επιτυγχάνεται με τη συμβολή του LIRE (Lucene Image Retrieval) που αναλύσαμε σε προηγούμενο κεφάλαιο. Όπως αναφέραμε το LIRE χρησιμοποιεί την επεκτάσιμη βιβλιοθήκη του Apache Lucene για τη δημιουργία του ευρετηρίου. Ο Apache Lucene είναι μια υψηλής απόδοσης αξιόπιστη μηχανή αναζήτησης, που προσφέρει

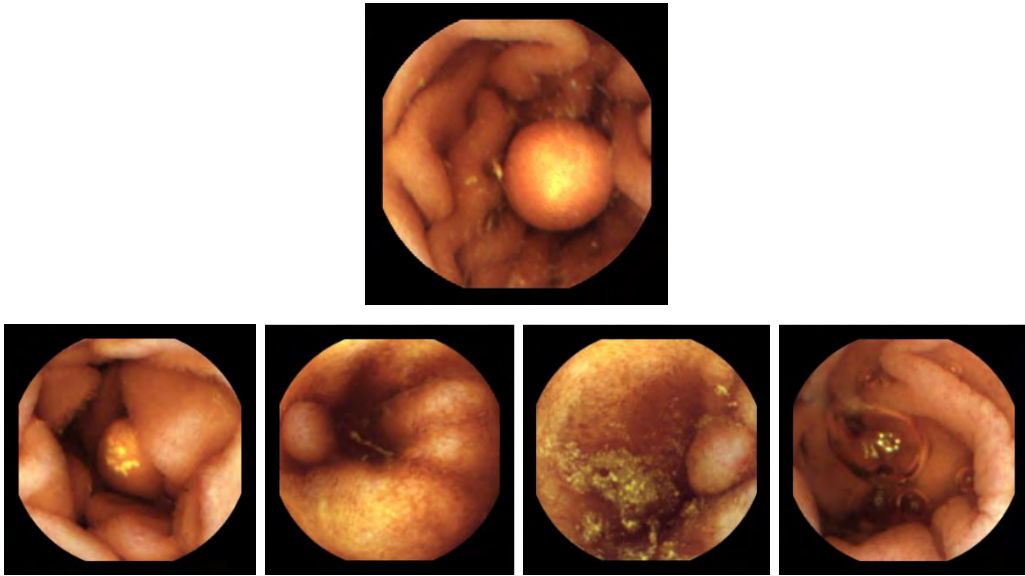
ισχυρούς αλγορίθμους αναζήτησης και δημιουργίας ευρετηρίου. Η μηχανή του Apache καλύπτει όλες τις πλατφόρμες ανάπτυξης, καθώς είναι ολοκληρωτικά γραμμένη χρησιμοποιώντας τη γλώσσα προγραμματισμού JAVA, επιπλέον προφέρει υψηλή απόδοση στη δημιουργία του ευρετηρίου με μικρές απαιτήσεις μνήμης και υψηλές ταχύτητες, πάνω από 150GB/ανά ώρα δεδομένα μπορούν να επεξεργαστούν για τη δημιουργία του ευρετηρίου. Για την δημιουργία του ευρετηρίου ακολουθείται την τεχνική των αντεστραμμένων αρχείων.

Το ευρετήριο που δημιουργείται περιέχει τα εξαγόμενα χαρακτηριστικά. Κάθε εικόνα που περιέχεται στη βάση των δεδομένων αντιπροσωπεύει ένα έγγραφο. Όπως αναφέραμε στην περιγραφή του αλγορίθμου, από κάθε εικόνα εξάγονται χαρακτηριστικά από κάθε υπερεικονοστοιχείο που περιέχει η εικόνα. Με βάση τα εξαγόμενα χαρακτηριστικά εφαρμόζεται το μοντέλο bag of visual words και δημιουργούνται οι κωδικές λέξεις και το κωδικό λεξιλόγιο. Το έγγραφο του ευρετηρίου περιέχει τις κωδικές λέξεις που εμπεριέχονται στο έγγραφο, δηλαδή τα εξαγόμενα χαρακτηριστικά από κάθε υπερεικονοστοιχείο της εικόνας, όπως επίσης εμπεριέχει το ιστόγραμμα των συχνοτήτων των κωδικών λέξεων με το οποίο θα γίνει η σύγκριση της ομοιότητας.

7.6 Σύγκριση Εικόνων

Μετά το πέρας της δημιουργίας του ευρετηρίου, το προτεινόμενο σύστημα ανάκλησης εικόνων είναι έτοιμο να δεχθεί ερωτήματα με σκοπό την ανάκληση όμοιων εικόνων. Αρχικά, δίνεται στο σύστημα μια εικόνα που θα λειτουργήσει σαν επερώτηση, από την εικόνα εξάγονται χαρακτηριστικά με τον αλγόριθμο εξαγωγής που αναφέραμε παραπάνω. Με την ολοκλήρωση της εξαγωγής έχουν δημιουργηθεί διανύσματα χαρακτηριστικών. Ουσιαστικά έχει δημιουργηθεί ένα καινούργιο έγγραφο που περιέχει τα εξαγόμενα χαρακτηριστικά της εικόνας που εισάχθηκε στο σύστημα, συνεπώς η σύγκριση γίνεται μεταξύ του εγγράφου της εικόνας επερώτησης και των εγγράφων που υπάρχουν στη βάση δεδομένων. Η σύγκριση των εγγράφων γίνεται με τη βοήθεια της μηχανής αναζήτησης Apache Lucene.

Η μηχανή αναζήτησης Apache Lucene, αναλαμβάνει να συγκρίνει το έγγραφο που δόθηκε στο σύστημα σαν επερώτηση, δηλαδή την εικόνα επερώτηση μετά την εφαρμογή του αλγορίθμου εξαγωγής χαρακτηριστικών που αναφέραμε, και των εγγράφων-εικόνων που βρίσκονται στη βάση δεδομένων. Η αναζήτηση γίνεται με το συνδυασμό των μοντέλων VSM (Vector Space Model) και το BIR (Boolean Model of Information Retrieval). Το μοντέλο VSM αναπαριστά ένα έγγραφο με ένα διάνυσμα. Η ιδέα πίσω από το μοντέλο του VSM είναι ότι όσο πιο συχνά εμφανίζεται ένας όρος στο έγγραφο της επερώτησης σε σχέση με το πόσο συχνά εμφανίζεται ο ίδιος όρος σε ένα έγγραφο από τη συλλογή των εγγράφων, τόσο πιο σχετικό είναι αυτό έγγραφο στο ερώτημα. Για τον υπολογισμό του αντιπροσωπευτικού διανύσματος του εγγράφου χρησιμοποιείται η συχνότητα εμφάνισης του όρου (tf) και η αντεστραμμένη συχνότητα εμφάνισης (idf) [66]. Το μοντέλο BIR εφαρμόζεται από τη μηχανή αναζήτησης ως ένα πρωτεύον φίλτρο για μείωση των εγγράφων στα οποία θα γίνει η τελική σύγκριση.



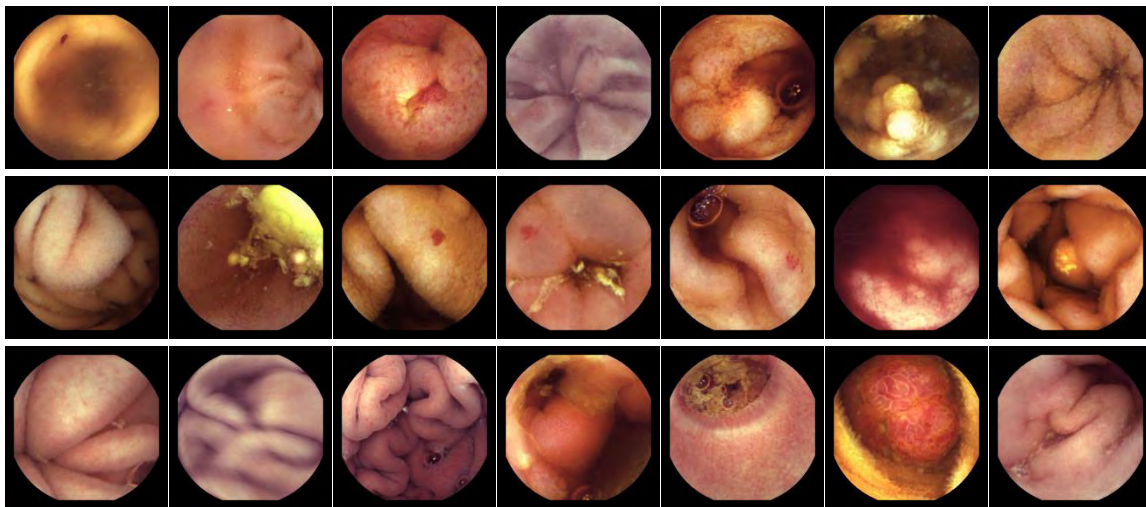
Εικ. 7.4 Πάνω : Εικόνα επερώτηση , Κάτω: Ανακτημένες εικόνες.

Κεφάλαιο 8

Πειράματα και Αποτελέσματα

8.1 Δεδομένα

Τα δεδομένα που χρησιμοποιήθηκαν για τα πειράματα αποτελούνται από εικόνες που περιέχουν παθήσεις του γαστρεντερικού συστήματος μαζί με φυσιολογικές εικόνες. Οι εικόνες αποτελούν ένα αντιπροσωπευτικό σύνολο από ενδοσκοπικές επεμβάσεις που πραγματοποιήθηκαν χρησιμοποιώντας την ενδοσκοπική κάψουλα MiroCam® (IntroMedic® Co., Seoul, South Korea) Εικ. 8.1. Τα δεδομένα είναι διαθέσιμα μέσω του διαδικτύου χρησιμοποιώντας την βάση δεδομένων για ενδοσκοπικές εικόνες KID [42,43]. Οι εικόνες συνοδεύονται με τις αντίστοιχες τους μάσκες, οι οποίες έχουν παραχθεί χειροκίνητα από ειδικούς χρησιμοποιώντας ένα εργαλείο σχολιασμού εικόνων[44].



Εικ. 8.1 Παραδείγματα εικόνων από τη συλλογή KID

8.2 Μετρικές απόδοσης

Η μέτρηση της απόδοσης ενός συστήματος ανάκλησης εικόνων με βάση το περιεχόμενο είναι μια ιδιαίτερα δύσκολη διαδικασία διότι δεν υπάρχει αντικειμενικός ορισμός της ομοιότητας των εικόνων. Για την υπερπήδηση του εμποδίου της ομοιότητας των εικόνων, συνήθως κατατάσσουμε τις εισαγόμενες εικόνες του συστήματος σε κλάσεις. Η κατάταξη των εικόνων σε κλάσεις βοηθάει στη διάκριση των επιστρεφόμενων εικόνων του συστήματος σε σχετικές ή μη σχετικές.

Για την σύγκριση μεταξύ συστημάτων ανάκληση είναι απαραίτητη η ύπαρξη τυποποιημένων μετρικών απόδοσης. Έχουν προταθεί πολλές διαφορετικές μετρικές [49] για την μέτρηση της απόδοσης ενός συστήματος ανάκλησης, στη συνέχεια θα αναφέρουμε τις πιο δημοφιλείς από αυτές .

$$\text{Ακρίβεια (Precision)} : P_k = \frac{\text{πλήθος σχετικών ανακτημένων εικονών}}{\text{πλήθος ανακτημένων εικονών}} \quad (\text{Σχ. 8.1})$$

Ο συμβολισμός P_k είναι η ακρίβεια του συστήματος όταν έχουν ανακτηθεί k εικόνες, η ακρίβεια παρουσιάζει το ποσοστό των σχετικών εικόνων που επιστράφηκαν από το σύστημα. Συχνά χρησιμοποιείται η P_{20} , P_{50} ή P_{100} δηλαδή η ακρίβεια του συστήματος όταν έχουν ανακτηθεί 20, 50, 100 εικόνες αντίστοιχα.

$$\text{Ανάκληση (Recall)} : R_k = \frac{\text{πλήθος σχετικών ανακτημένων εικόνων}}{\text{πλήθος σχετικών εικόνων}} \quad (\text{Σχ. 8.2})$$

Ο συμβολισμός R_k υποδηλώνει την ανάκληση του συστήματος όταν έχουν ανακτηθεί k εικόνες, συχνά χρησιμοποιείται R_{100} , R_{1000} ή $R(P = 0.5)$ δηλαδή η ανάκληση όταν η ακρίβεια είναι 0.5.

$$\text{Κατάταξη (Rank)} : rank = \frac{1}{NN_R} \left(\sum_{i=1}^{N_R} R_i - \frac{N_R(N_R-1)}{2} \right) \quad (\text{Σχ. 8.3})$$

Όπου N είναι το μέγεθος της συλλογής των εικόνων, N_R είναι οι σχετικές εικόνες της συλλογής και R_i είναι κατάταξη της i -οστής εικόνας. Η κατάταξη τείνει προς το 0 αν υπάρχει τέλεια απόδοση του συστήματος, η απόδοση μειώνεται καθώς η κατάταξη τείνει προς το 1, αν η κατάταξη είναι 0.5 τότε η ανάκληση είναι τυχαία.

8.3 Πειράματα

Για την καλύτερη αξιολόγηση της μεθόδου εξαγωγής χαρακτηριστικών του συστήματος ανάκλησης, η μέθοδος που προτείνεται συγκρίθηκε με τα χαρακτηριστικά SURF και με μια καινοτόμα μέθοδο εξαγωγής χαρακτηριστικών που προτάθηκε από τους Shuai et al. [45]. Στην εξαγωγή των χαρακτηριστικών σύγκρισης ακολουθήθηκε το μοντέλο bag of visual words και η δημιουργία του ευρετήριο έγινε με τη βοήθεια του Apache Lucene.

Η μέθοδος εξαγωγής χαρακτηριστικών SURF ακολουθεί τα βήματα που περιγράψαμε στο Κεφάλαιο 4, εξάγονται τοπικοί περιγραφείς από σημεία ενδιαφέροντος στην εικόνα, οι οποίοι θα αποτελέσουν τις κωδικές λέξεις και το κωδικό λεξιλόγιο, σύμφωνα με την τεχνική bag of visual words.

Η τεχνική που προτάθηκε από τους Shuai et al. ξεκινάει τμηματοποιώντας την εικόνα εφαρμόζοντας τον αλγόριθμο SLIC και στη συνέχεια από κάθε υπερεικονοστοιχείο της εικόνας που δημιουργείται εξάγονται ένα σύνολο ιστογραμμάτων. Πιο συγκεκριμένα ιστογράμματα του καναλιού I από τον HSI χώρο χρώματος, ιστογράμματα των τριών

καναλιών του χρωματικού χώρου *HSV*, κανονικοποιημένα ιστογράμματα των καναλιών του RGB χώρου, ιστόγραμμα από το χώρο RG- χρωματικότητας, επιπλέον εξαγονται χαρακτηριστικά υψής χρησιμοποιώντας την τεχνική LBP και ιστογράμματα από έναν χώρο χρώματος ειδικά σχεδιασμένο για αναγνώριση αντικειμένων σε μια εικόνα [46]. Τα εξαγόμενα ιστογράμματα δειγματοληπτούνται και το τελικό διάνυσμα των χαρακτηριστικών για κάθε υπερεικονοστοιχείο έχει μέγεθος 240. Με βάση τα εξαγόμενα διανύσματα από κάθε υπερεικονοστοιχείο δημιουργούνται οι κωδικές λέξεις και το λεξιλόγιο ακολουθώντας της τεχνική bag of visual words, το λεξιλόγιο που προτείνεται από τους Shuai et al. έχει μέγεθος 200 κωδικές λέξεις.

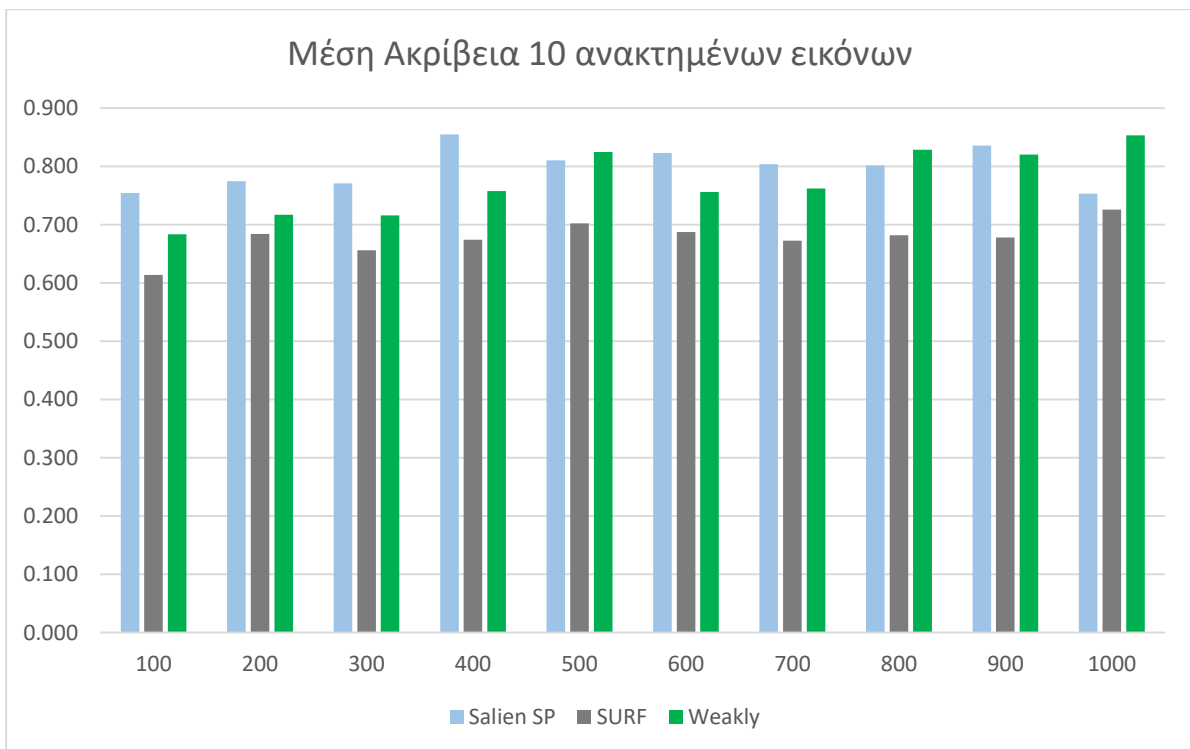
Οι παράμετροι της προτεινόμενης τεχνικής για την εξαγωγή οπτικών χαρακτηριστικών βασίστηκαν σε προηγούμενες εργασίες [47,48] που έχουν πραγματοποιηθεί από την ερευνητική ομάδα του Πανεπιστημίου Θεσσαλίας *is-innovation*. Πιο συγκεκριμένα ο αριθμός των υπερεικονοστοιχείων που χωρίζεται η εικόνα με βάση τον αλγόριθμο τμηματοποίησης SLIC είναι 320 και ο αριθμός των σημείων ενδιαφέροντος SURF που εξαγονται είναι υπολογισμένος ώστε να υπάρχει τουλάχιστον ένα σημείο SURF σε περιοχή ανωμαλίας της εικόνας.

8.3.1 Σύγκριση χαρακτηριστικών εξαγωγής με βάση το μέγεθος του λεξιλογίου

Για την καλύτερη αξιολόγηση της απόδοσης των τεχνικών εξαγωγής χαρακτηριστικών και της απόδοσης του συστήματος ανάκλησης, συγκρίθηκαν οι τεχνικές εξαγωγής μεταβάλλοντας το μέγεθος του λεξιλογίου που δημιουργείται. Η διαδικασία της αξιολόγησης, διεξάχθηκε πραγματοποιώντας τυχαία ερωτήματα στη βάση δεδομένων, ζητώντας από το σύστημα ανάκλησης να μας επιστρέψει αρχικά τις 10 κοντινότερες εικόνες και στη συνέχεια αυξήθηκε ο αριθμός των εικόνων που ζητάμε να μας επιστρέψει το σύστημα. Στα παρακάτω διαγράμματα φαίνεται η μέση ακρίβεια του συστήματος όταν έχει ζητηθεί η ανάκληση 10,20,50 και 100 εικόνων, μεταβάλλοντας κάθε φορά το μέγεθος του λεξιλογίου. Το μέγεθος του λεξιλογίου ξεκινάει από τις 100 λέξεις και συνέχεια αυξάνεται σταδιακά το μέγεθος του λεξιλόγιο κατά 100 λέξεις ως ο του το μέγεθος του λεξιλογίου φθάσει τις 1000 κωδικές λέξεις. Στα παρακάτω διαγράμματα η ετικέτα Weakly συμβολίζει την τεχνική που προτάθηκε από τους Shuai Wang et al. και η ετικέτα Salient SP (Superpixel) συμβολίζει την προτεινόμενη τεχνική εξαγωγής.

Πιν. 8.1 Μέση Ακρίβεια που επιτεύχθηκε με διαφορετικές μεθοδολογίες σε 10 ανακλημένες εικόνες

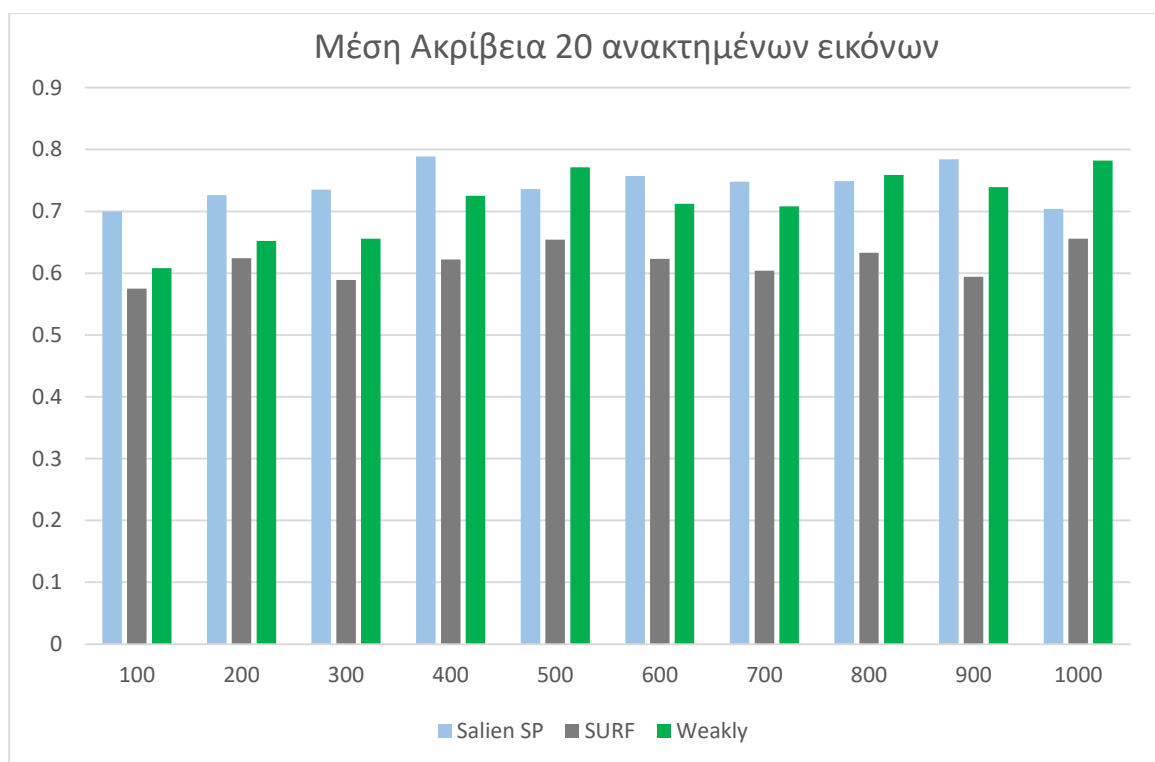
Μέγεθος Λεξιλογίου	Salient SP		SURF		Shuai et al. (Weakly)	
	Mean 10	stdev	Mean 10	stdev	Mean 10	stdev
100	0,754	0,088	0,614	0,140	0,684	0,133
200	0,774	0,083	0,684	0,122	0,717	0,120
300	0,771	0,084	0,656	0,142	0,716	0,125
400	0,855	0,062	0,674	0,122	0,758	0,090
500	0,810	0,078	0,702	0,131	0,824	0,085
600	0,823	0,074	0,687	0,150	0,756	0,091
700	0,804	0,077	0,673	0,132	0,762	0,099
800	0,802	0,082	0,682	0,126	0,828	0,112
900	0,836	0,063	0,678	0,142	0,820	0,088
1000	0,753	0,095	0,726	0,111	0,853	0,090



Διαγρ. 8.1 Μέση Ακρίβεια που επιτεύχθηκε με διαφορετικές μεθοδολογίες σε 10 ανακλημένες εικόνες

Πιν. 8.2 Μέση Ακρίβεια που επιτεύχθηκε με διαφορετικές μεθοδολογίες σε 20 ανακλημένες εικόνες

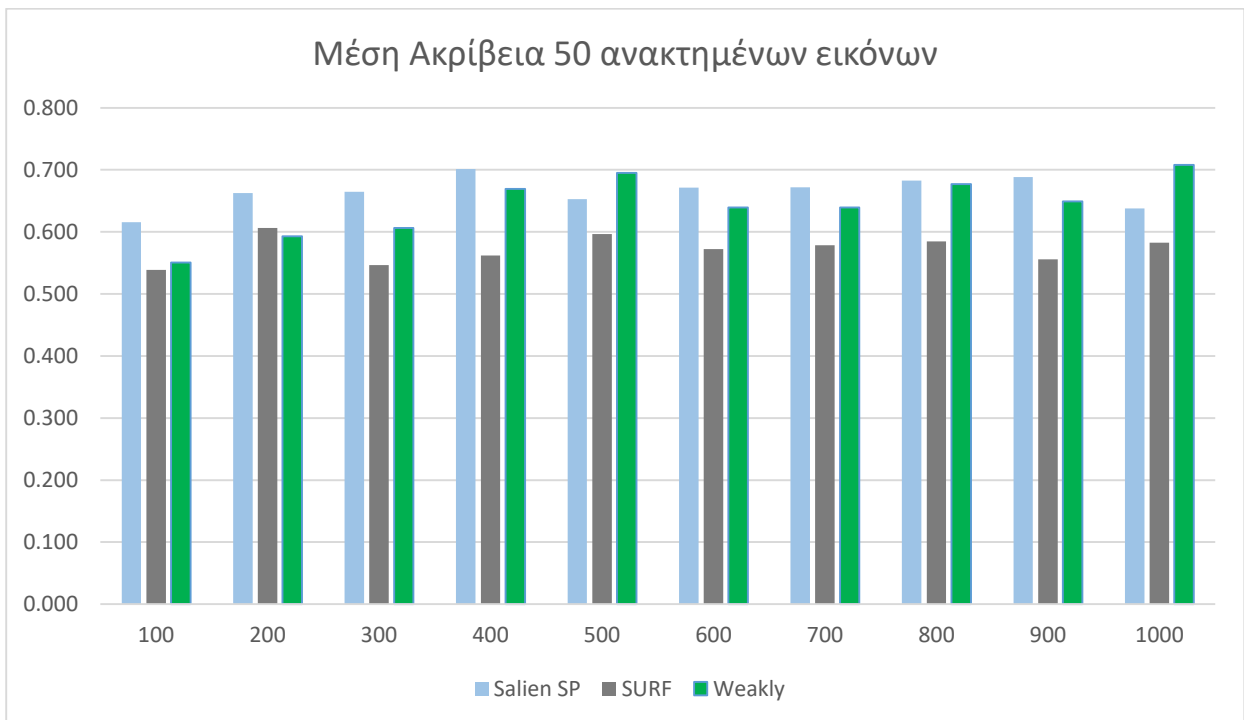
Μέγεθος Λεξιλογίου	Salient SP		SURF		Shuai et al. (Weakly)	
	Mean 20	stdev	Mean 20	stdev	Mean 20	stdev
100	0,7	0,08	0,575	0,105	0,608	0,12
200	0,726	0,075	0,624	0,104	0,652	0,107
300	0,735	0,069	0,589	0,119	0,656	0,106
400	0,789	0,082	0,622	0,100	0,725	0,071
500	0,736	0,097	0,654	0,103	0,771	0,079
600	0,757	0,085	0,623	0,122	0,712	0,078
700	0,748	0,078	0,604	0,115	0,708	0,088
800	0,749	0,078	0,633	0,100	0,759	0,105
900	0,784	0,073	0,594	0,131	0,739	0,104
1000	0,704	0,083	0,656	0,105	0,782	0,097



Διαγρ. 8.2 Μέση Ακρίβεια που επιτεύχθηκε με διαφορετικές μεθοδολογίες σε 10 ανακλημένες εικόνες

Πιν. 8.3 Μέση Ακρίβεια που επιτεύχθηκε με διαφορετικές μεθοδολογίες σε 50 ανακλημένες εικόνες

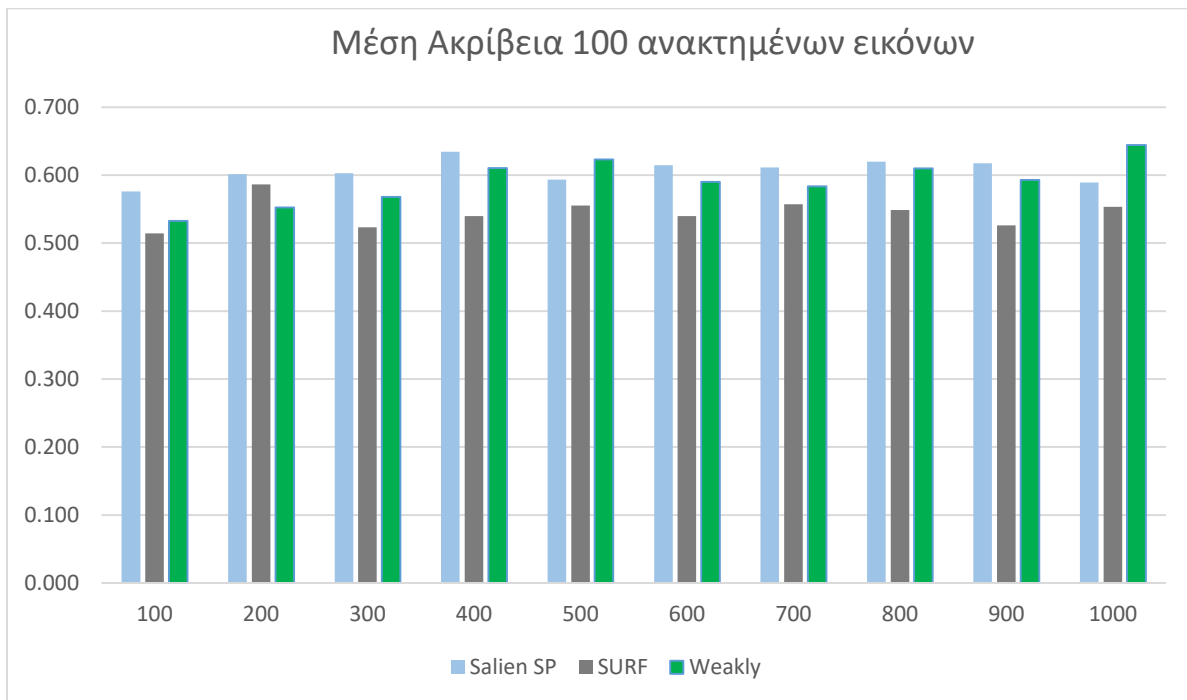
Μέγεθος Λεξιλογίου	Salient SP		SURF		Shuai et al. (Weakly)	
	Mean 50	stdev	Mean 50	stdev	Mean 50	stdev
100	0,615	0,088	0,539	0,073	0,551	0,090
200	0,662	0,081	0,606	0,067	0,593	0,083
300	0,665	0,076	0,547	0,083	0,606	0,078
400	0,701	0,089	0,562	0,080	0,669	0,067
500	0,653	0,094	0,596	0,081	0,695	0,085
600	0,671	0,091	0,572	0,088	0,639	0,079
700	0,672	0,084	0,578	0,075	0,639	0,081
800	0,683	0,076	0,584	0,075	0,677	0,097
900	0,689	0,095	0,556	0,088	0,649	0,100
1000	0,638	0,078	0,583	0,090	0,708	0,089



Διαγρ. 8.3 Μέση Ακρίβεια που επιτεύχθηκε με διαφορετικές μεθοδολογίες σε 50 ανακλημένες εικόνες

Πιν. 8.4 Μέση Ακρίβεια που επιτεύχθηκε με διαφορετικές μεθοδολογίες σε 100 ανακλημένες εικόνες

Μέγεθος Λεξιλογίου	Salient SP		SURF		Shuai et al. (Weakly)	
	Mean 100	stdev	Mean 100	stdev	Mean 100	stdev
100	0,576	0,074	0,514	0,057	0,533	0,066
200	0,601	0,078	0,587	0,052	0,552	0,072
300	0,603	0,083	0,523	0,063	0,568	0,067
400	0,634	0,093	0,540	0,061	0,611	0,076
500	0,594	0,089	0,555	0,071	0,623	0,094
600	0,615	0,086	0,540	0,070	0,590	0,075
700	0,611	0,085	0,557	0,058	0,584	0,080
800	0,620	0,083	0,549	0,064	0,610	0,097
900	0,617	0,098	0,526	0,068	0,593	0,091
1000	0,589	0,073	0,553	0,070	0,644	0,090



Διαγρ. 8.4 Μέση Ακρίβεια που επιτεύχθηκε με διαφορετικές μεθοδολογίες σε 50 ανακλημένες εικόνες.

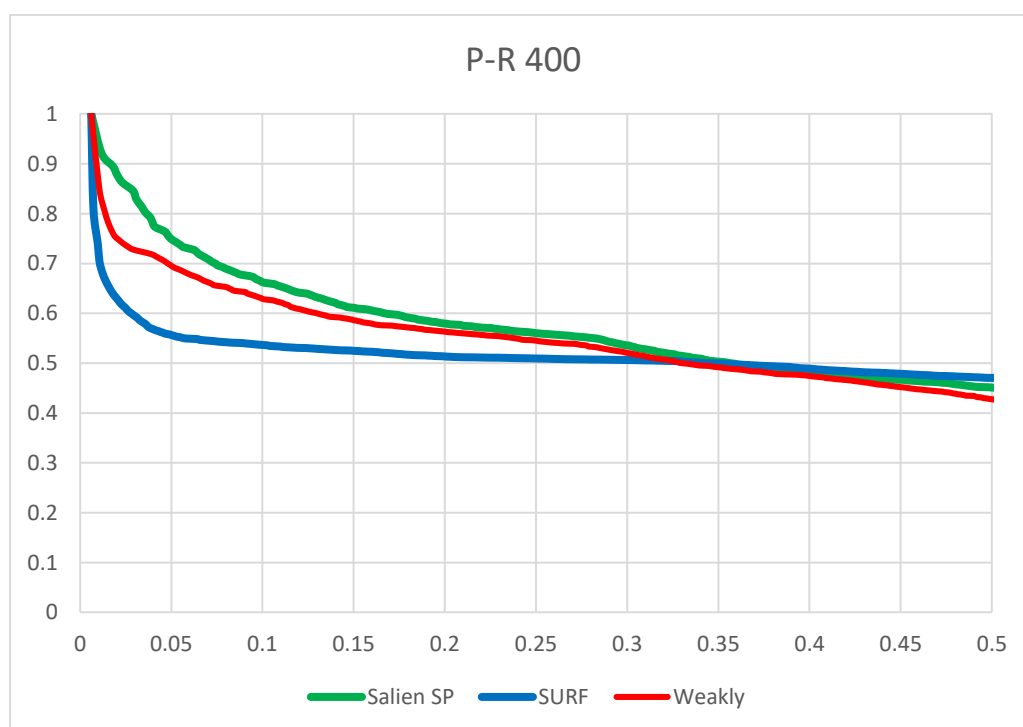
Από τα διαγράμματα φαίνεται ότι η προτεινομένη τεχνική έχει καλύτερη απόδοση σε λεξιλόγια με μικρό αριθμό λέξεων και αποδίδει αρκετά καλά σε σχέση με τις άλλες δυο τεχνικές που εξετάζονται, σε μεγαλύτερου μεγέθους λεξιλόγια. Οι τοπικοί περιγραφείς SURF φαίνεται από τα διαγράμματα ότι δεν είναι κατάλληλοι για την ανάκληση ενδοσκοπικών εικόνων με βάση το περιεχόμενο διότι έχουν χαμηλή μέση ακρίβεια για όλα τα εξεταζόμενα μεγέθη των λεξιλογίων. Τα χαρακτηριστικά Weakly εμφανίζουν καλύτερη απόδοση από τους τοπικούς περιγραφείς SURF, αλλά χαμηλότερη απόδοση από την

προτεινομένη τεχνική, η τεχνική Weakly φαίνεται να αποδίδει τα βέλτιστα όταν το μέγεθος του λεξιλογίου είναι το μέγιστο που εξετάζεται.

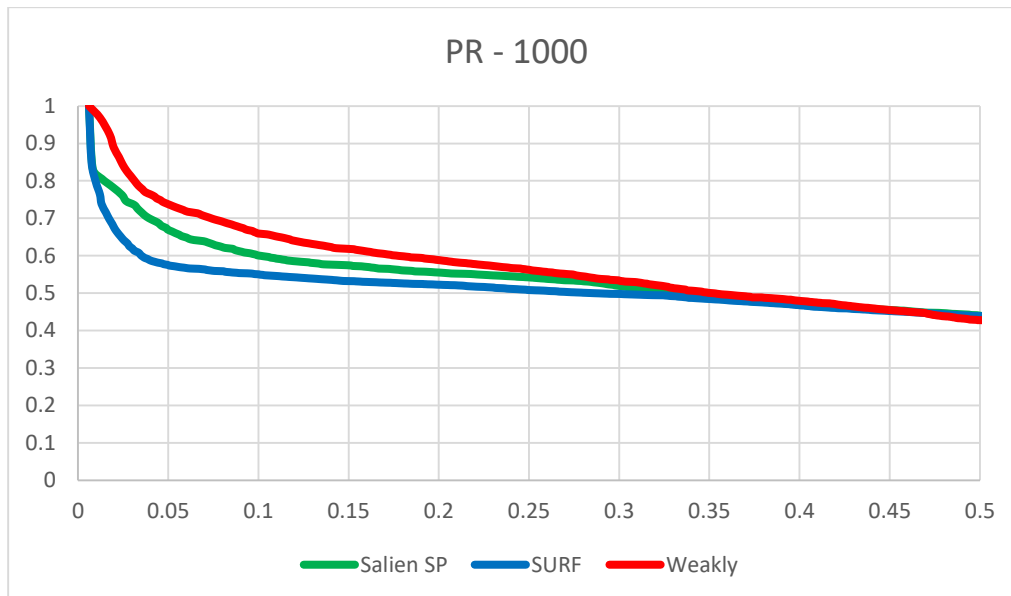
8.3.2 Καμπύλες ακρίβειας – ανάκλησης

Οι καμπύλες ακρίβειας – ανάκλησης (P-R curves) χρησιμοποιούνται κατά κόρον για την εξέταση της απόδοσης ενός συστήματος ανάκλησης [34,35,49]. Οι καμπύλες PR παρέχουν πολλές πληροφορίες για την απόδοση του συστήματος και είναι εύκολα κατανοητές. Όμως, η κλίση της καμπύλης εξαρτάται από το ερώτημα που θα δοθεί στο σύστημα, δύσκολα ερωτήματα θα έχουν ως αποτέλεσμα το σύστημα να έχει χαμηλότερη ακρίβεια και ανάκληση και χαμηλή καμπύλη στο διάγραμμα, αντίστροφα εύκολα ερωτήματα θα έχουν υψηλότερη καμπύλη στο διάγραμμα. Είναι σύνηθες στην αξιολόγηση των συστημάτων να χρησιμοποιείται μέρος της καμπύλης PR [34,35,39, 41,50].

Για την αξιολόγηση των χαρακτηριστικών που προτάθηκαν καθώς και του συστήματος ανάκλησης που δημιουργήθηκε στα πλαίσια της πτυχιακής εργασίας, έγιναν τυχαία ερωτήματα για κάθε μέγεθος του λεξιλογίου που δημιουργήθηκε και υπολογίστηκε η μέση ακρίβεια και ανάκληση του συστήματος όπως φαίνεται στα διαγράμματα παρακάτω. Στη συνέχεια παραθέτονται τα διαγράμματα για το μέγεθος του λεξιλογίου που υπήρχε ιδιαίτερο ενδιαφέρον, δηλαδή εκεί που τα ποσοστά ακρίβειας, μεγιστοποιούνται για κάθε τεχνική δηλαδή για μέγεθος του λεξιλογίου 400 και 1000.



Διαγρ. 8.5 Ακρίβεια- Ανάκληση, μέγεθος λεξιλογίου 400

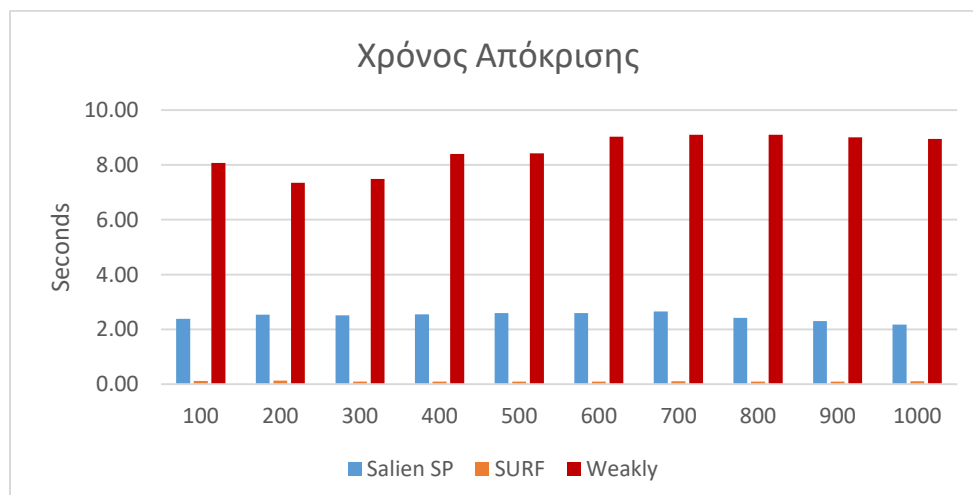


Διαγρ. 8.5 Ακρίβεια- Ανάκληση, μέγεθος λεξιλογίου 1000

8.3.3 Αξιολόγησης της ταχύτητας ανάκλησης εικόνων

Η ταχύτητα ανάκλησης είναι ένα σημαντικό μέρος της αξιολόγησης ενός συστήματος ανάκλησης, καθώς πέρα από την ακρίβεια του συστήματος είναι απαραίτητη και η άμεση απόκριση του. Η αξιολόγηση της ταχύτητας της ανάκλησης συνήθως απουσιάζει από τα συστήματα που έχουν δημιουργηθεί με σκοπό την εφαρμογή τους σε ιατρικές εικόνες [34] σε αντίθεση με τα συστήματα γενικού σκοπού αλλά είναι ένα μέτρο το οποίο πρέπει να λαμβάνεται υπόψιν για τη συνολική απόδοση του.

Για την αξιολόγηση του προτεινόμενου συστήματος έγιναν 100 τυχαία ερωτήματα και ζητήσαμε από το σύστημα να μας ανακτήσει 50 όμοιες εικόνες, ο χρόνος ξεκινάει όταν η εικόνα εισέρχεται στο σύστημα και σταματάει όταν έχει τελειώσει η διαδικασία της ανάκλησης. Στο διάγραμμα (Διαγρ. 8.6) που ακολουθεί παραθέτονται οι χρόνοι απόκρισης (sec) για κάθε μέγεθος του λεξιλογίου που δημιουργήθηκε συγκρίνοντας την προτεινόμενη τεχνική εξαγωγής χαρακτηριστικών με την τεχνική Weakly.

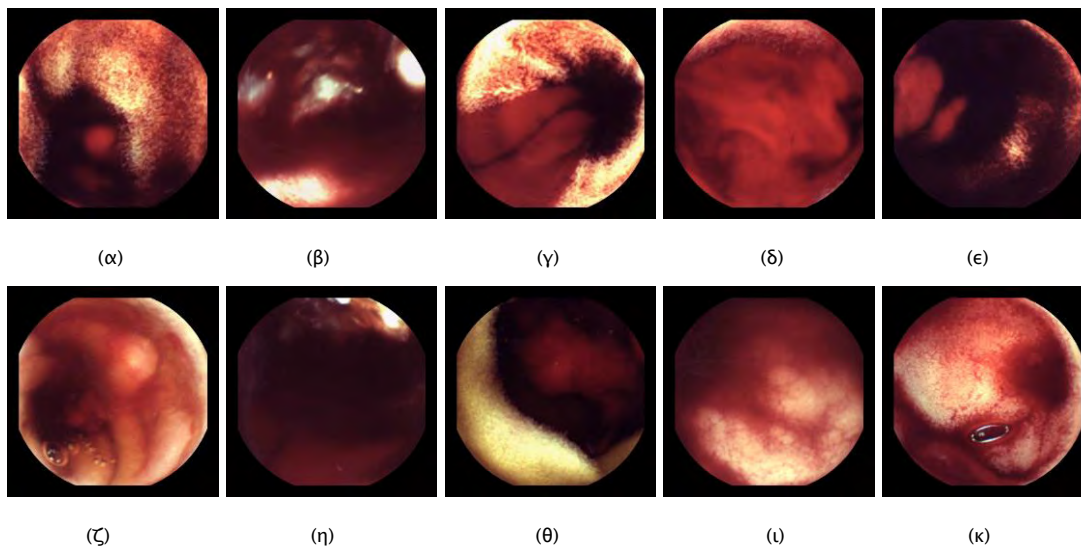
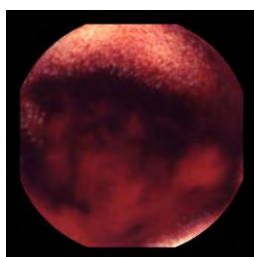


Διαγρ. 8.6 Μέσος χρόνος απόκρισης του συστήματος

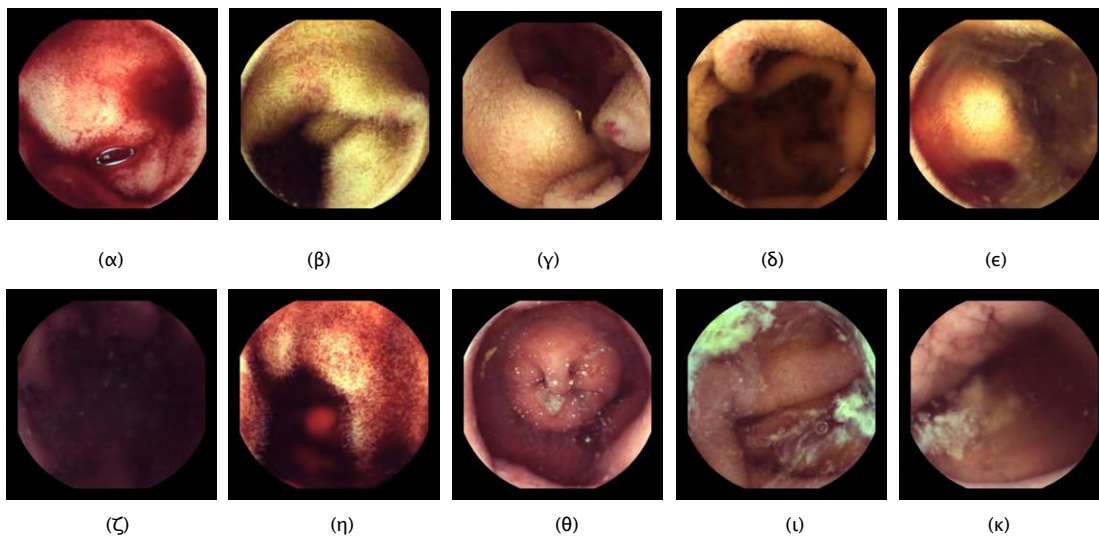
8.3.4 Παραδείγματα ανάκλησης εικόνων

Στη συνέχεια της εξέτασης της απόδοσης του συστήματος, παραθέτουμε ενδεικτικά παραδείγματα των ανακλημένων εικόνων που επιστέφει το προτεινόμενο σύστημα για κάθε εξεταζόμενη μέθοδο εξαγωγής χαρακτηριστικών. Παρομοίως, εισήγαμε στο σύστημα τυχαίες εικόνες και ζητήσαμε να μας επιστέψει τις δέκα κοντινότερες εικόνες με την εικόνα επερώτηση.

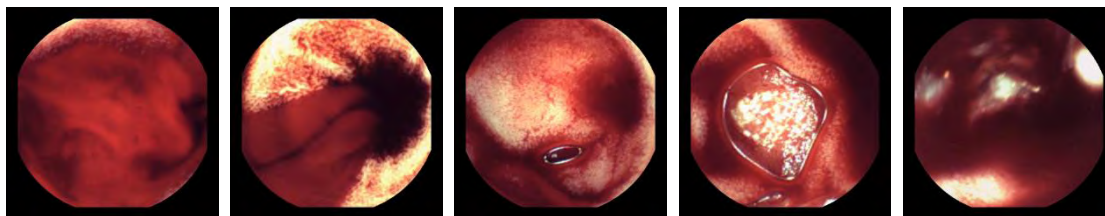
Παράδειγμα 8.1 : Εικόνα επερώτηση: Αγγειακή αλλοίωση



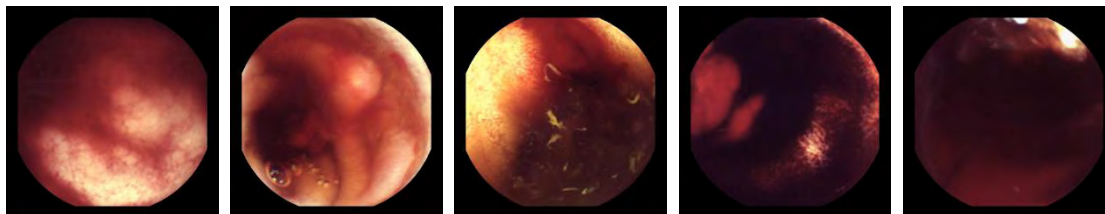
Ανακλημένες εικόνες με την προτεινόμενη μέθοδο (Salient SP), (α-ε, η-κ) αγγειακές αλλοιώσεις, (ζ) πολύποδας,



Ανακλημένες εικόνες με τη μέθοδο SURF (α-δ), είναι αγγειακές αλλοιώσεις, οι υπόλοιπες είναι εικόνες από φυσιολογικό ιστό



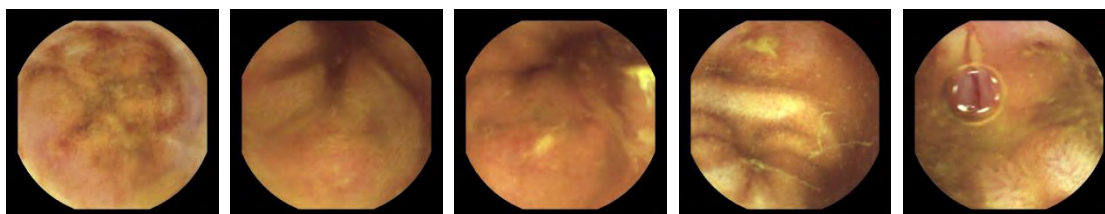
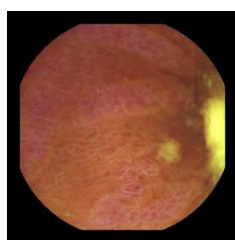
(α) (β) (γ) (δ) (ε)



(ζ) (η) (θ) (ι) (κ)

Ανακλημένες εικόνες με τη μέθοδο Shuai et al (Weakly), (η) πολύποδας, οι υπόλοιπες εικόνες είναι αγγειακές αλλοιώσεις

Παράδειγμα 8.2 Φλεγμονώδης εντερική νόσος (inflammatory)

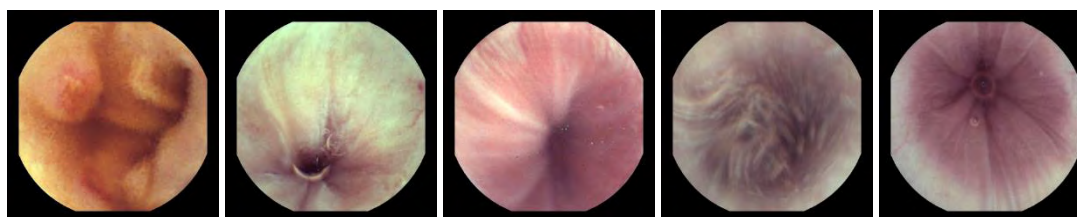


(α) (β) (γ) (δ) (ε)

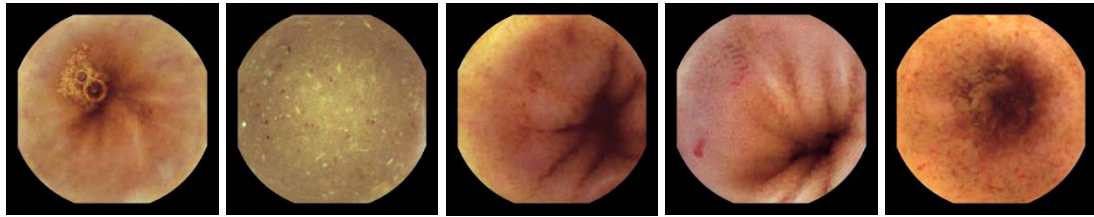


(ζ) (η) (θ) (ι) (κ)

Ανακλημένες εικόνες με την προτεινόμενη μέθοδο (Salient SP), (β,γ,ε,θ,κ) φλεγμονώδης νόσος, (α,γ,ι) φυσιολογικός ιστός, (ζ) αγγειεκτασία, (η) πολύποδας

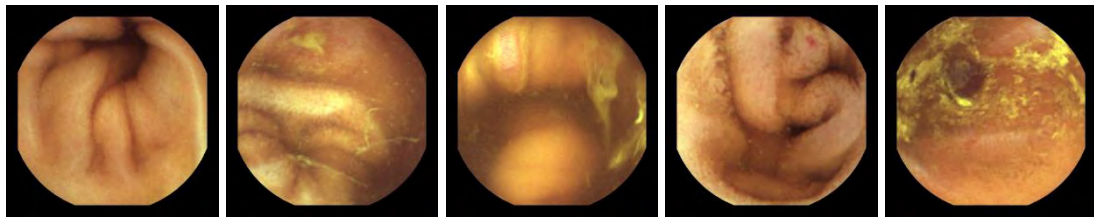


(α) (β) (γ) (δ) (ε)

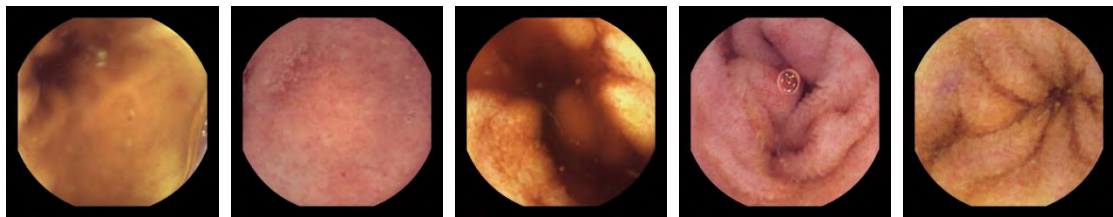


(ζ) (η) (θ) (ι) (κ)

Ανακλημένες εικόνες με τη μέθοδο SURF, (α) φλεγμονώδης νόσος, (ι) αγγειεκτασία, υπόλοιπες εικόνες είναι φυσιολογικός ιστός



(α) (β) (γ) (δ) (ε)



(ζ) (η) (θ) (ι) (κ)

Ανακλημένες εικόνες με τη μέθοδο Shuai et al (Weakly), (ζ) φλεγμονώδης νόσος, υπόλοιπες εικόνες είναι φυσιολογικός ιστός

Τα παραδείγματα μας δείχνουν οπτικά τα αποτελέσματα της προτεινόμενης μεθόδου εξαγωγής χαρακτηριστικών και των συγκρινόμενων μεθόδων. Παρατηρούμε ότι η προτεινόμενη μέθοδος έχει μεγαλύτερη ακρίβεια στις ανακλημένες εικόνες και επιβεβαιώνεται και οπτικά τα αποτελέσματα από τις καμπύλες ακρίβειας ανάκλησης και των προηγούμενων διαγραμμάτων.

Συμπεράσματα – Μελλοντική εργασία

9.1 Συμπεράσματα

Σε αυτή τη πτυχιακή εργασία παρουσιάστηκε ένα νέο σύστημα ανάκλησης εικόνων με βάση το περιεχόμενο για ενδοσκοπικές εικόνες. Σκοπός του συστήματος είναι η αξιόπιστη και αποδοτική ανάκληση εικόνων. Για την επίτευξη του στόχου, εφαρμόζεται μια νέα τεχνική για την εξαγωγή τοπικών ιδιοτήτων από περιοχές της εικόνας. Η μέθοδος εξαγωγής αξιοποιεί αλγορίθμους τμηματοποίησης εικόνας, εύρεσης σημείων ενδιαφέροντος και καινοτόμες τεχνικές για την αναπαράσταση των διανυσμάτων χαρακτηριστικών και δημιουργίας ευρετηρίου. Το προτεινόμενο σύστημα εκμεταλλεύεται την αποδοτική μέθοδο τμηματοποίησης SLIC για το διαχωρισμό της εικόνας σε μικρότερα τμήματα, τον αξιόπιστο αλγόριθμο εύρεση σημείων ενδιαφέροντος SURF και συνδυάζει αυτές τις μεθόδους με το μοντέλο Bag of Visual Words για την αποδοτική ανάκληση ενδοσκοπικών εικόνων. Για τη ταχύτατη ανάκληση, αποθήκευση και αναζήτηση των εικόνων δημιουργήθηκε ένα ευρετήριο χρησιμοποιώντας τη μηχανή αναζήτησης Apache Lucene.

Η προτεινόμενη μέθοδος συγκρίθηκε με ήδη υπάρχουσες μεθοδολογίες εξαγωγής χαρακτηριστικών που υπάρχουν στη βιβλιογραφία και έχουν εφαρμοστεί σε ενδοσκοπικές εικόνες. Η σύγκριση των μεθόδων επικύρωσε την αποτελεσματικότητα της προτεινομένης μεθόδου έναντι των άλλων μεθόδων που έχουν προταθεί. Στη σύγκριση των μεθόδων εξαγωγής χρησιμοποιήθηκαν όμοιες τεχνικές για τη δημιουργία οπτικών λέξεων και ευρετηρίου. Η προτεινόμενη μέθοδος πλεονεκτεί έναντι των συγκρινόμενων μεθόδων στα ακόλουθα :

- Καλύτερη εκμετάλλευση των σημείων ενδιαφέροντος
- Υψηλότερη ακρίβεια ανάκλησης εικόνων σε σχέση και με τις δυο μεθόδους που συγκρίθηκαν και ιδιαίτερα με τη μέθοδο SURF
- Μικρότερο μέγεθος διανύσματος χαρακτηριστικών σε σχέση και με τις δυο μεθόδους που συγκρίθηκαν
- Μεγαλύτερη Ταχύτητα ανάκλησης εικόνων σε σχέση με την μέθοδο Weakly

9.2 Μελλοντική Εργασία

Το σύστημα ανάκλησης που δημιουργήθηκε προσφέρεται για την ανάπτυξη και σύγκριση νέων μεθοδολογιών εξαγωγής χαρακτηριστικών για την ανάκληση ενδοσκοπικών εικόνων. Η δημιουργία καινοτόμων χαρακτηριστικών μπορεί να οδηγήσει στη βελτίωση της

απόδοσης του συστήματος. Ωστόσο επειδή εφαρμόζονται αλγόριθμοι τμηματοποίησης και εύρεση τοπικών περιγραφών, μπορεί να μελετηθεί η ακρίβεια της ανάκλησης του συστήματος εξετάζοντας νέους αλγορίθμους. Συνεπώς η μελλοντική έρευνα μπορεί να επικεντρωθεί στη χρήση νέων τεχνικών για τμηματοποίηση και εξαγωγή τοπικών περιγραφών.

Άλλος ένας τομέας διερεύνησης είναι η δημιουργία οπτικών λέξεων. Το μοντέλο Bag of Visual Words στο οποίο βασίζεται το σύστημα για τη δημιουργία οπτικών λέξεων εφαρμόζει τον αλγόριθμο συσταδοποίησης k-means. Η εφαρμογή νέων αλγορίθμων συσταδοποίησης ή αλγορίθμων ομαδοποίησης είναι ένας ακόμα τομέας που μπορεί να επικεντρωθεί μελλοντική η έρευνα του συστήματος.

Επιπροσθέτως, η έρευνα μπορεί να επικεντρωθεί στη καλύτερη αξιοποίηση των πληροφοριών των υπερεικονοστοιχείων. Τα υπερεικονοστοιχεία περιέχουν σημαντικές πληροφορίες για την εικόνα και αντιπροσωπεύουν τα εικονοστοιχεία μίας ορισμένης περιοχής της εικόνας, συνεπώς θα μπορούσαν να χρησιμοποιηθούν τεχνικές που λαμβάνουν υπόψη τα γειτονικά υπερεικονοστοιχεία ή να λαμβάνουν το σύνολο υπερεικονοστοιχείων σαν μια νέα εικόνα.

Αναφορές

- [1] W. Niblack, R. Barber, W. Equitz, M. Flickner, E. Glasman, D. Pektovic, P.Yanker, C. Faloutsos, G. Taubin, The QBIC project: Querying images by content using color, texture and shape, in: Proceedings of the SPIE Storage and Retrieval for Image and Video Databases, San Jose, CA, 1994.
- [2] A.Pentland, R.W.Picard, S.Sclaro, Photobook: Tools for content based manipulation of image databases, International Journal of Computer Vision 18 (3) (1996) 233–254.]
- [3] Y. Kalantidis, G. Tolias, Y. Avrithis, M. Phinikettos, E. Spyrou, P. Mylonas, S. Kollias. VIRaL: Visual Image Retrieval and Localization. In Multimedia Tools and Applications, Springer, Volume 51, Number 2, pp. 555–592, February 2011.
- [4] William Horsthemke, Daniela Raicu, and Jacob Furst, Task-Oriented Medical Image Retrieval
- [5] Lokesh Setia, Alexandra Teynor, Alaa Halawani and Hans Burkhardt , Image Classification using Cluster Cooccurrence Matrices of Local Relational Features, October 26–27, 2006
- [6] Mohammad Reza Zare, Henning Muller: A MEDICAL X-RAY IMAGE CLASSIFICATION AND RETRIEVAL SYSTEM in PACIS 2016 Proceeding, 6–27–2016
- [7] Ashnil Kumar, jinman Kim, Weidong Cai, Michael Fulham, Dagan Feng: Content-Based Medical Image Retrieval: A Survey of Applications to Multidimensional and Multimodality Data in J Digit Imaging
- [8] Dimitris K. Iakovidis, Member, IEEE, Nikos Pelekis, Evangelos E. Kotsifakos, Ioannis Kopanakis, Haralampos Karanikas, and Yannis Theodoridis : A Pattern Similarity Scheme for Medical Image Retrieval in IEEE TRANSACTIONS ON INFORMATION TECHNOLOGY IN BIOMEDICINE, VOL. 13, NO. 4, JULY 2009
- [9] Qian X, Tagare HD, Fulbright RK, Long R, Antani S: Optimal embedding for shape indexing in medical image databases. Med Image Anal 14(3):243–54, 2010
- [10] Trupti S. Atre, K.V.Metre, MIRS: Text Based and Content Based Image Retrieval in International Journal of Engineering Science and Innovative Technology (IJESIT) Volume 3, Issue 4, July 2014
- [11] Sameer Antani L. Rodney Long George R. Thoma A Biomedical Information System for Combined Content-Based Retrieval of Spine X-ray Images and Associated Text Information.
- [12] Napel SA, Beaulieu CF, Rodriguez C, Cui J, Xu J, Gupta A, et al: Automated retrieval of CT images of liver lesions on the basis of image similarity: Method and preliminary results. Radiology 256(1):243–52, 2010
- [13] Unay D, Ekin A, Jasinschi R: Local structure-based region-of interest retrieval in brain MR images. IEEE Trans Inf Technol Biomed 14(4):897–903, 2010
- [14] Chi-Ren Shyu, Carla E. Brodley, Avinash C. Kak, Akio Kosaka, ASSERT: A Physician-in-the-Loop Content-Based Retrieval System for HRCT Image Databases in Computer Vision and Image Understanding Vol. 75, Nos. 1/2, July/August, pp. 111–132, 1999

- [15] Essam A. El-Kwae, Haifeng Xu, and Mansur R. Kabuka Content-Based Retrieval in Picture Archiving and Communication Systems
- [16] R. Chbeir, Y. Amghar, A. Flory, MIMS: A prototype for medical image retrieval, in: Recherche d'Informations Assistee par Ordinateur (RIA0'2000) Computer Assisted Information Retrieval, Vol. 1, Paris, France, 2000.
- [17] Keyzers D, Dahmen J, Ney H, Wein BB, Lehmann TM: Statistical framework for model-based image retrieval in medical applications. J Electron Imaging 12(1):59-68, 2003
- [18] Menglin Jiang, Shaoting Zhang, Hongsheng Li, and Dimitris N. Metaxas, Computer-Aided Diagnosis of Mammographic Masses Using Scalable Image Retrieval in IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, VOL. 62, NO. 2, FEBRUARY 2015
- [19] Quelled G, Lamard M, Cazuguel G, Roux C, Cochener B: Case retrieval in medical databases by fusing heterogeneous information. IEEE Trans Med Imaging 30(1):108-18, 2011
- [20] Quelled G, Lamard M, Bekri L, Cazuguel G, Roux C, Cochener B: Medical case retrieval from a committee of decision trees. IEEE Trans Inf Technol Biomed 14(5):1227-35, 2010
- [21] Christian Beecks, Klaus Schoeffmann, Mathias Lux, Merih Seran Uysal, Thomas Seidl, Endoscopic Video Retrieval: A Signature-based Approach for Linking Endoscopic Images with Video Segments in International Symposium on Multimedia IEEE 2015
- [22] Manish Chowdhury, Malay Kumar Kundu Endoscopic Image Retrieval System Using Multi-scale Image Features in Proceedings of the 2nd International Conference on Perception and Machine Intelligence 2015
- [23] Ashish Oberoi and Manpreet Singh Content Based Image Retrieval System for Medical Databases (CBIR-MD) - Lucratively tested on Endoscopy, Dental and Skull Images in IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 3, No 1, May 2012
- [24] Oana Astrid VATAMANU, Mirela FRANDES, Diana LUNGEANU, Gheorghe-Ioan MIHALAS Content Based Image Retrieval Using Local Binary Pattern Operator and Data Mining Techniques
- [25] Shunren Xia , Dingfei Ge , Weirong Mo , Zanchao Zhang A Content-Based Retrieval System for Endoscopic Images in Proceedings of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference Shanghai, China, September 1-4, 2005
- [26] S. Antani, L. R. Long, G. R. Thoma, A biomedical information system for combined content based retrieval of spine xray images and associated text information, in: Proceedings of the 3rd Indian Conference on Computer Vision, Graphics and Image Processing (ICVGIP 2002), Ahamdabad, India, 2002.
- [27] T. Ikeda, M. Hagiwara, Content based image retrieval system using neural networks, International Journal of Neural Systems 10 (5) (2000) 417-424.
- [28] M. J. Egenhofer, Spatial query by sketch, in: Proceedings of the IEEE Symposium on Visual Languages (VL 1996), Boulder, CO, USA, 1996, pp. 6|67.
- [29] J. Ze Wang, G. Wiederhold, O. Firschein, S. Xin Wei, Wavelet based image indexing techniques with partial sketch retrieval capability, in: Proceedings of the Fourth Forum on Research and Technology Advances in Digital Libraries, Washington D.C., 1997, pp. 13-24.
- [30] C. Le Bozec, E. Zapletal, M.-C. Jaulent, D. Heudes, P. Degoulet, Towards content based image retrieval in HIS-integrated PACS, in: Proceedings of the Annual Symposium of the American Society for Medical Informatics (AMIA), Los Angeles, CA, USA, 2000, pp. 477-481.

- [31] J. Sivic, A. Zisserman, Video Google: a text retrieval approach to object matching in videos, in: Proceedings of the Ninth IEEE International Conference on Computer Vision, 2003, pp. 1470–1477.
- [32] H. Jegou, F. Perronnin, M. Douze, J. Sanchez, P. Perez, C. Schmid, Aggregating local image descriptors into compact codes, *IEEE Trans. Pattern Anal. Mach. Intell.* 34 (9) (2012) 1704–1716.
- [33] F. Perronnin, C.R. Dance, Fisher kernels on visual vocabularies for image categorization, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2007, pp. 1–8.
- [34] **Ahmad Alzu'bi, Abbas Amira, Naeem Ramzan Semantic content-based image retrieval: A comprehensive study in** *J. Vis. Commun. Image R.* 32 (2015) 20–54
- [35] Henning Muller, Nicolas Michoux, David Bandon and Antoine Geissbuhler A Review of Content Based Image Retrieval Systems in Medical Applications Clinical Benefits and Future Directions.
- [36] M. R. Ogiela, R. Tadeusiewicz, Semantic oriented syntactic algorithms for content recognition and understanding of images in medical databases, in: Proceedings of the second International Conference on Multimedia and Exposition (ICME'2001), IEEE Computer Society, IEEE Computer Society, Tokyo, Japan, 2001, pp. 621-624.
- [37] M.-C. Jaulent, C. Le Bozec, Y. Cao, E. Zapletal, P. Degoulet, A property concept frame representation for flexible image content retrieval in histopathology databases, in: Proceedings of the Annual Symposium of the American Society for Medical Informatics (AMIA), Los Angeles, CA, USA, 2000.
- [38] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine Susstrunk, SLIC Superpixels, EPFL Technical Report 149300, June 2010.
- [39] **R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk, "Slic superpixels compared to state-of-the-art superpixel methods," Pattern Analysis and Machine Intelligence,** *IEEE Transactions on*, vol. 34, no. 11, pp. 2274–2282, 2012.
- [40] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*. Elsevier/Academic Press, 2008.
- [41] Zhihua Xia, Member, IEEE, Xinhui Wang, Liangao Zhang, Zhan Qin, Xingming Sun, and Kui Ren, A Privacy-preserving and Copy-deterrence Content-based Image Retrieval Scheme in Cloud Computing in: *IEEE TRANSCATION ON INFORMATION FORENSIC AND SECURITY*
- [42] **D. K. Iakovidis and A. Koulaouzidis, "Automatic lesion detection in capsule endoscopy based on color saliency: closer to an essential adjunct for reviewing software,"** *Gastrointestinal endoscopy*, vol. 80, no. 5, pp. 877– 883, 2014.
- [43] **D. K. Iakovidis and A. Koulaouzidis, "Automatic lesion detection in wireless capsule endoscopy—a simple solution for a complex problem,"** in *Image Processing (ICIP), 2014 IEEE International Conference on*. IEEE, 2014, pp. 2236–2240.
- [44] Iakovidis D.K., Smailis C.V (2011). Efficient Semantically-Aware Annotation of Images. *IEEE International Conference of Imaging Systems and Techniques (IST)* (pp. 146-149).Penang, Malaysia.
- [45] **Shuai Wang, Yang Cong, Huijie Fan, Lianqing Liu, Xiaoqiu Li, Yunsheng Yang, Yandong Tang, Member, Huaici Zhao, and Haibin Yu, "Computer-Aided Endoscopic Diagnosis Without Human Specific Labeling" in :TBME 2015**
- [46] **V. D. S. Ke et al., "Evaluating color descriptors for object and scene recognition,"** *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 9, pp.1582–1596, 2010.

- [47] Dimitris K. Iakovidis, Senior Member, IEEE, Dimitris Chatzis, Panos Chrysanthopoulos, and Anastasios Koulaouzidis, "Blood Detection in Wireless Capsule Endoscope Images based on Salient Superpixels" in EMBC 2015
- [48] D. Chatzis, P. Chrysanthopoulos, D. K. Iakovidis¹, A. Koulaouzidis "COLOR BASED SEGMENTATION OF CAPSULE ENDOSCOPY IMAGES FOR AUTOMATED LESION SIZE MEASUREMENTS" in: UEG 2015
- [49] Henning Muller, Wolfgang Muller, David McG. Squire, Stephane Marchad-Maillet, Thierry Pun "Performance evaluation in content-based image retrieval: overview and proposals", in : Pattern Recognition Letters (2001) 593-601
- [50] Jing-Ming Guo, IEEE, and Heri Prasetyo "Content-Based Image Retrieval Using Features Extracted From Halftoning-Based Block Truncation Coding" in : IEEE Transactions on Image Processing 2015
- [51] Lux Mathias, Savvas A. Chatzichristofis. Lire: Lucene Image Retrieval – An Extensible Java CBIR Library. In proceedings of the 16th ACM International Conference on Multimedia, pp. 1085-1088, Vancouver, Canada, 2008
- [52] Bay, H., A. Ess, T. Tuytelaars, and L. Van Gool. "SURF: Speeded Up Robust Features." Computer Vision and Image Understanding (CVIU). Vol. 110, No. 3, pp. 346–359, 2008.
- [53] T. Ojala, M. Pietikäinen, and D. Harwood, A Comparative Study of Texture Measures with Classification Based on Feature Distributions, Pattern Recognition, 29 (1), (1996) 51-59
- [54] Lowe, David G. (1999). "Object recognition from local scale-invariant features" (PDF). Proceedings of the International Conference on Computer Vision. 2. pp. 1150–1157
- [55] H. Tamura, S. Mori, T. Yamawaki, Textural features corresponding to visual perception, IEEE Trans. Syst., Man Cybernet. 8 (6) (1978) 460–473.
- [56] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural Features for Image Classification", IEEE Trans. on Systems, Man, and Cybernetics, Vol. SMC-3, No.6, November 1973, pp.610-621.
- [57] Z. Wu, R. Leahy, An optimal graph theoretic approach to data clustering: theory and its application to image segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 15 (11) (1993) 1101–1113.
- [58] Y. Boykov, O. Veksler, R. Zabih, Fast approximate energy minimization via graph cuts, IEEE Trans. Pattern Anal. Mach. Intell. 23 (11) (2001) 1222–1239.
- [59] T.F. Cootes, G.J. Edwards, C.J. Taylor, Active appearance models, IEEE Trans. Pattern Anal. Mach. Intell. 23 (6) (2001) 681–685.
- [60] R. Beichel, H. Bischof, F. Leberl, M. Sonka, Robust active appearance models and their application to medical image analysis, IEEE Trans. Med. Imag. 24 (9) (2005) 1151–1169.
- [61] T. Pavlidis, Structural Pattern Recognition, Springer, New York, 1980.
- [62] H. Samet, The quadtree and related hierarchical data structures, ACM Comput. Surv. (CSUR) 16 (2) (1984) 187–260.
- [63] T.F. Cootes, C.J. Taylor, D.H. Cooper, J. Graham, Active shape models-their training and application, Comput. Vis. Image Understand. 61 (1) (1995) 38–59.
- [64] H. Li, O. Chutatape, Automated feature extraction in color retinal images by a model based approach, IEEE Trans. Biomed. Eng. 51 (2) (2004) 246–254.
- [65] M. Sezgin, Survey over image thresholding techniques and quantitative performance evaluation, J. Electron. Imag. 13 (1) (2004) 146–168.
- [66] G. Salton, A. Wong, and C. S. Yang (1975), "A Vector Space Model for Automatic Indexing," Communications of the ACM, vol. 18, nr. 11, pages 613–620

- [67] Y. Rubner, C. Tomasi, L.J. Guibas, The earth mover's distance as a metric for image retrieval, *Int. J. Comp. Vis.* 40 (2) (2000) 99–121.
- [68] C. Gollan. Nichtlineare verformungsmodelle für die bilderkennung. diploma thesis, 2003.
- [69] Kien A. Hua, Khanh Vu, kai Jung-Hwan Oh. Sammatch: a flexible and efficient sampling-based image retrieval technique for large image databases. *Sto MULTIMEDIA '99: Proceedings of the seventh ACM international conference on Multimedia (Part 1)*, serDec 225–234, New York, NY, USA, 1999. ACM Press.
- [70] Bentley, J. L. (1975). "Multidimensional binary search trees used for associative searching". *Communications of the ACM.* 18 (9): 509
- [71] Guttman, A. (1984). "R-Trees: A Dynamic Index Structure for Spatial Searching". *Proceedings of the 1984 ACM SIGMOD international conference on Management of data – SIGMOD '84 (PDF)*. p. 47
- [72] Codd, E.F. (1970). "A Relational Model of Data for Large Shared Data Banks". *Communications of the ACM.* 13 (6): 377–387
- [73] Y. Liu, D. Zhang, G. Lu, WY. Ma, A survey of content-based image retrieval with high-level semantics, *Pattern Recog.* 40 (1) (2007) 262–282.
- [74] X.Y. Wang, B.B. Zhang, H.Y. Yang, Content-based image retrieval by integrating color and texture features, *Multim. Tools Appl.* 68 (3) (2014) 545–569.
- [75] Nobuyuki Otsu (1979). "A threshold selection method from gray-level histograms". *IEEE Trans. Sys., Man., Cyber.* 9 (1): 62–66

