

ΑΝΑΛΥΣΗ ΕΓΧΡΩΜΩΝ ΕΙΚΟΝΩΝ ΓΙΑ ΕΦΑΡΜΟΓΕΣ ΑΝΑΚΤΗΣΗΣ ΒΑΣΕΙ ΠΕΡΙΕΧΟΜΕΝΟΥ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

Βασίλειος Σεβετλίδης

ΤΕΧΝΟΛΟΓΙΚΟ ΕΚΠΑΙΔΕΥΤΙΚΟ ΙΔΡΥΜΑ
ΣΧΟΛΗ ΤΕΧΝΟΛΟΓΙΚΩΝ ΕΦΑΡΜΟΓΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ & ΕΠΙΚΟΙΝΩΝΙΩΝ

ΑΝΑΛΥΣΗ ΕΓΧΡΩΜΩΝ ΕΙΚΟΝΩΝ ΓΙΑ
ΕΦΑΡΜΟΓΕΣ ΑΝΑΚΤΗΣΗΣ ΒΑΣΕΙ ΠΕ-
ΡΙΕΧΟΜΕΝΟΥ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

Βασίλειος Σεβετλίδης, Α.Μ. 2478

ΣΕΡΡΕΣ, 2013

Υπεύθυνη Δήλωση: Βεβαιώνω ότι είμαι συγγραφέας αυτής της πτυχιακής εργασίας και ότι κάθε βοήθεια την οποία είχα για την προετοιμασία της, είναι πλήρως αναγνωρισμένη και αναφέρεται στην πτυχιακή εργασία. Επίσης έχω αναφέρει τις όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών ή λέξεων, είτε αυτές αναφέρονται ακριβώς είτε παραφρασμένες. Επίσης βεβαιώνω ότι αυτή η πτυχιακή εργασία προετοιμάστηκε από εμένα προσωπικά ειδικά για τις απαιτήσεις του προγράμματος σπουδών του τμήματος Πληροφορικής & Επικοινωνιών του Τ.Ε.Ι. Σερρών.

ΑΝΑΛΥΣΗ ΕΓΧΡΩΜΩΝ ΕΙΚΟΝΩΝ ΓΙΑ ΕΦΑΡΜΟΓΕΣ ΑΝΑΚΤΗΣΗΣ ΒΑΣΕΙ ΠΕΡΙΕΧΟΜΕΝΟΥ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

Βασίλειος Σεβετλίδης, Α.Μ. 2478

Επιβλέπων: Δρ. Γεώργιος Παυλίδης
Επιστημονικός Συνεργάτης ΤΕΙ Σερρών

ΤΕΧΝΟΛΟΓΙΚΟ ΕΚΠΑΙΔΕΥΤΙΚΟ ΙΔΡΥΜΑ ΣΕΡΡΩΝ
ΣΧΟΛΗ ΤΕΧΝΟΛΟΓΙΚΩΝ ΕΦΑΡΜΟΓΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ & ΕΠΙΚΟΙΝΩΝΙΩΝ

*στους γονείς μου Ηλία και Μαλαματένια, για την αφειδώλετη υποστήριξη
και την διακαή αγάπη τους.*

Περίληψη

Η παρούσα εργασία είχε ως αντικείμενο μελέτης την ανάκτηση έγχρωμων στατικών δισδιάστατων εικόνων βάσει περιεχομένου. Στόχος της εργασίας ήταν η ανάπτυξη μιας μεθοδολογίας και ενός περιγραφέα έγχρωμων εικόνων που να βασίζεται στην αναγνώριση καθολικών χαρακτηριστικών των εικόνων και να μην επηρεάζονται από την περιστροφή και την κλιμάκωση. Ο περιγραφέας θα μπορούσε να χρησιμοποιηθεί για εφαρμογές ανάκτησης βάσει περιεχομένου αποδίδοντας τουλάχιστον όσο οι περιγραφείς που έχουν προταθεί στο πλαίσιο του προτύπου MPEG-7. Μετά από εκτενή βιβλιογραφική αναζήτηση, βασική έρευνα στα θέματα της ανάκτησης βάσει περιεχομένου και πειραματική ανάπτυξη με τη χρήση διεθνώς αναγνωρισμένων εργαλείων λογισμικού, αναπτύχθηκε μια πρωτότυπη μεθοδολογία εξαγωγής καθολικών χαρακτηριστικών και η δημιουργία μιας νέας υπογραφής για έγχρωμες. Ο περιγραφέας έγχρωμων στατικών εικόνων Color Gravity Descriptor (CGD), ο οποίος αναπτύχθηκε, βασίζεται στην υπόθεση πως, εάν μειωθεί η χρωματική παλέτα σε κάποια βασικά χρώματα και έπειτα εξαχθούν κατάλληλα χαρακτηριστικά σχετιζόμενα με την τοποθέτηση των εν λόγω χρωμάτων στο χωρικό επίπεδο της εικόνας, μπορεί να επιτευχθεί ταύτιση όμοιων εικόνων. Στη μεθοδολογία εξαγωγής του CGD περιλαμβάνεται αλγόριθμος μείωσης χρωμάτων, επιλογής, μετατροπής και εξαγωγής χαρακτηριστικών και τελικής διαμόρφωσης της υπογραφής κάθε εικόνας. Η μέθοδος δοκιμάστηκε πειραματικά με χρήση τυπικών βάσεων δεδομένων για εφαρμογές ανάκτησης βάσει περιεχομένου και αποδείχθηκε ότι αποδίδει κατά μέσο όρο καλύτερα αποτελέσματα ανάκτησης από τις μεθόδους που περιλαμβάνονται στο πρότυπο MPEG-7.

Abstract

This work focuses on the content-based retrieval of two-dimensional static color images. The aim was to develop a methodology and a descriptor for color images based on the recognition of global characteristics of the images that will not be affected by rotation and scaling. The descriptor could be used for content-based retrieval applications yielding at least as good performance as the one provided by descriptors that have been proposed in the context of the MPEG-7 standard. After an extensive literature research, basic research on content-based retrieval and experimental development using internationally recognized software tools, a novel methodology has been developed that extracts global image features and creates a novel signature for color images. The Color Gravity Descriptor (CGD), which was developed as a result of this research, was based on the assumption that if image colors are reduced to a fixed palette and then appropriate characteristics associated with the location of these colors on that image plane are extracted, matching of similar images can be accomplished. The methodology of forming the CGD includes a color reduction algorithm, selection, extraction and transformation of features and the final formation of the signature of each image. The method was extensively tested using standard image databases and proved to provide overall average better retrieval results than methods included in the MPEG-7 standards.

Πρόλογος

Η ανάλυση έγχρωμων εικόνων για την εξαγωγή χαρακτηριστικών που να αντικατοπτρίζουν το περιεχόμενό τους σαν μια μορφή ψηφιακής υπογραφής αποτελεί ένα δυναμικό πεδίο έρευνας. Στόχος της παρούσας εργασίας είναι η ανάπτυξη μεθοδολογίας ανάλυσης δισδιάστατων έγχρωμων στατικών εικόνων για την εξαγωγή χαρακτηριστικών και η υλοποίηση συστήματος ανάκτησης βάσει περιεχομένου που να εκμεταλλεύεται τα εν λόγω χαρακτηριστικά.

Η ενασχόληση με την παρούσα ερευνητική εργασία είχε ως αποτέλεσμα ένα πολύτιμο συνονθύλευμα εμπειριών. Οι εμπειρίες αυτές επηρεάζουν τόσο τις επαγγελματικές όσο και τις προσωπικές μου αποφάσεις, καθώς δοκίμασα τις αντοχές μου στην επίμονη και συστηματική μελέτη.

Σε αυτό το σημείο θέλω να ευχαριστήσω γονείς και φίλους, οι οποίοι μου παρείχαν άφθονη υλική και ηθική υποστήριξη όσο καιρό μου χρειάστηκε για να ολοκληρώσω τη παρούσα εργασία. Επίσης, ευχαριστώ τον επιβλέποντα καθηγητή Γεώργιο Παυλίδη, που δίχως την συνεχή καθοδήγηση του δεν θα είχα κερδίσει τίποτα από όλα αυτά και προπάντων έναν μέντορα.

*Βασίλειος Σεβετλίδης
Σέρρες, 2013*

ΕΙΚΟΝΕΣ

Εικόνα 1. Κάθετη τομή οφθαλμού.....	22
Εικόνα 2. Αριστερά: διάγραμμα πυκνότητας κωνίων και ραβδίων, δεξιά: διάγραμμα κατανομής των φωτοϋποδοχών πάνω στον αμφιβληστροειδή	22
Εικόνα 3 Αριστερά, η απόκριση των κωνίων και ραβδίων στο ορατό φάσμα φωτός. Δεξιά, ο συνδυασμός των τριών βασικών χρωμάτων	23
Εικόνα 4 Καμπύλες απόκρισης εξισώσεων	24
Εικόνα 5. Ο χρωματικός χώρος CIExyY	25
Εικόνα 6. Ο κύβος RGB, και οι τιμές των παραμέτρων στις γωνίες	26
Εικόνα 7. Χρωματική διαβάθμιση του RGB χώρου	26
Εικόνα 8. Αριστερά, οι συνιστώσες του YCbCr, δεξιά, τα κανάλια Y, Cb, και Cr	27
Εικόνα 9. Ο ανάποδος κώνος του μοντέλου HSI.....	29
Εικόνα 10. Ο διπλός κώνους του HMMD	30
Εικόνα 11. Ο χρωματικός χώρος CIELAB	30
Εικόνα 12. Κατωφλίωση με τη μέθοδο Otsu	33
Εικόνα 13 Αποτελέσματα μείωσης χρωμάτων σε έγχρωμη εικόνα	34
Εικόνα 14. Τυπικό σύστημα CBIR	36
Εικόνα 15. Γραφικές παραστάσεις precision/recall	42
Εικόνα 16 Αριστερά η αρχική έγχρωμη εικόνα και δεξιά απεικονίσεις με χρήση 3-8 κυρίαρχων χρωμάτων	45
Εικόνα 17. Βασική μονάδα του μετασχηματισμού Haar	46
Εικόνα 18. Σχηματικό διάγραμμα της δημιουργίας του SCD.....	46
Εικόνα 19. Αντιπροσωπευτικά χρώματα εικόνας	47
Εικόνα 20. Παραδείγματα της διαδικασίας GoF/GoP	49
Εικόνα 21. Διάγραμμα ροής της μεθόδου που αναπτύχθηκε	52
Εικόνα 22 Οι χώροι των μοντέλων RGB και HSV	53
Εικόνα 23 Τομή κώνου HSV και άποψη κορυφής	53
Εικόνα 24 Χρωματική τμηματοποίηση κυκλικού δίσκου του χρωματικού χώρου HSV	55
Εικόνα 25 Κατάτμηση εικόνας, με χρήση масκών.....	57
Εικόνα 26 Αρχική και εικόνες μειωμένης χρωματικότητας για $N_H = \{6, 12, 24\}$ και $N_G = \{3\}$	58
Εικόνα 27 Αποτελέσματα μείωσης χρωμάτων σε πραγματικές εικόνες..	61
Εικόνα 28. Παράδειγμα δύο εικόνων διαφορετικής ανάλυσης που παρουσιάζουν το ίδιο μοτίβο	62
Εικόνα 29 Γράφος υπογραφής	64
Εικόνα 30 Δείγμα αποτελεσμάτων ανάκτησης, από την ομάδα Αφρικανοί	67
Εικόνα 31 Δείγμα αποτελεσμάτων ανάκτησης, από την ομάδα Παραλίες	68
Εικόνα 32 Δείγμα αποτελεσμάτων ανάκτησης, από την ομάδα Λεωφορεία	69

Εικόνα 33 Δείγμα αποτελεσμάτων ανάκτησης, από την ομάδα Δεινόσαυροι	70
Εικόνα 34 Δείγμα αποτελεσμάτων ανάκτησης, από την ομάδα Ελέφαντες	71
Εικόνα 35 Δείγμα αποτελεσμάτων ανάκτησης, από την ομάδα Μνημεία	72
Εικόνα 36 Δείγμα αποτελεσμάτων ανάκτησης, από την ομάδα Άλογα...	73
Εικόνα 37 Δείγμα αποτελεσμάτων ανάκτησης, από την ομάδα Φαγητά.	74
Εικόνα 38 Δείγμα αποτελεσμάτων ανάκτησης, από την ομάδα Βουνά...	75
Εικόνα 39 Δείγμα αποτελεσμάτων ανάκτησης, από την ομάδα Τριαντάφυλλα.....	76
Εικόνα 40. Γράφημα precision/recall για το ερώτημα στην κατηγορία «Αφρικανοί».....	77
Εικόνα 41. Γράφημα precision/recall για το ερώτημα στην κατηγορία «Παραλίες».....	77
Εικόνα 42. Γράφημα precision/recall για το ερώτημα στην κατηγορία «Λεωφορεία».....	78
Εικόνα 43. Γράφημα precision/recall για το ερώτημα στην κατηγορία «Δεινόσαυροι»	79
Εικόνα 44. Γράφημα precision/recall για το ερώτημα στην κατηγορία «Ελέφαντες».....	79
Εικόνα 45. Γράφημα precision/recall για το ερώτημα στην κατηγορία «Μνημεία»	80
Εικόνα 46. Γράφημα precision/recall για το ερώτημα στην κατηγορία «Άλογα»	80
Εικόνα 47. Γράφημα precision/recall για το ερώτημα στην κατηγορία «Φαγητό»	81
Εικόνα 48. Γράφημα precision/recall για το ερώτημα στην κατηγορία «Βουνά»	82
Εικόνα 49. Γράφημα precision/recall για το ερώτημα στην κατηγορία «Ρόδα».....	82
Εικόνα 50. MAP μεταξύ των περιγραφών του MPEG-7 και της μεθόδου CGD	85
Εικόνα 51 Σύγκριση αποτελεσμάτων ανάκτησης, από την ομάδα λεωφορεία	88
Εικόνα 52 Σύγκριση αποτελεσμάτων ανάκτησης, από την ομάδα ελέφαντες.....	91
Εικόνα 53 Σύγκριση αποτελεσμάτων ανάκτησης, από την ομάδα φαγητά	93
Εικόνα 54 Σύγκριση αποτελεσμάτων ανάκτησης, από την ομάδα βουνά	96
Εικόνα 55 Σύγκριση αποτελεσμάτων ανάκτησης, από την ομάδα τριαντάφυλλα	98

ΠΙΝΑΚΕΣ

Πίνακας I. Υπολογισμός των παραμέτρων του HMMD.....	29
Πίνακας II. Απαιτούμενος αριθμός bits, για την αποθήκευση εικόνας....	31
Πίνακας III. Καταστάσεις χαρακτηρισμού ανάκτησης ενός αποτελέσματος.....	41
Πίνακας IV. Χρωματικοί χώροι που χρησιμοποιεί το πρότυπο MPEG-7	44
Πίνακας V. MAP ανά κατηγορία στη BΔ Wang	83
Πίνακας VI P@20 ανά κατηγορία στη BΔ Wang.....	84
Πίνακας VII. NMRR ανά κατηγορία, στη BΔ Wang.....	85

ΠΕΡΙΕΧΟΜΕΝΑ

ΕΙΣΑΓΩΓΗ	17
1.1 Εισαγωγή.....	17
1.2 Στόχος της εργασίας	17
1.3 Ο ρόλος της εικόνας για τον άνθρωπο	17
1.4 Η διάδοση της εικόνας στη σημερινή εποχή.....	18
1.5 Το πρότυπο MPEG-7.....	20
Η ΨΗΦΙΑΚΗ ΕΙΚΟΝΑ	21
2.1 Η εικόνα και το περιεχόμενο της	21
2.1.1 Το ανθρώπινο μάτι	21
2.1.2 Η τριχρωματική θεωρία.....	23
2.1.3 Ο χρωματικός χώρος CIEXYZ.....	23
2.1.4 Ο χρωματικός χώρος CIExyY	24
2.1.5 Χρωματικά μοντέλα	25
2.1.5.1 Το χρωματικό μοντέλο RGB.....	26
2.1.5.2 Το χρωματικό μοντέλο CMYK	27
2.1.5.3 Το χρωματικό μοντέλο YCrCb	27
2.1.5.4 Το χρωματικό μοντέλο HSI	28
2.1.5.5 Το χρωματικό μοντέλο HMMD	29
2.1.5.6 Το χρωματικό μοντέλο CIELab	30
2.1.6 Τι είναι η ψηφιακή εικόνα.....	31
2.1.6.1 Τεχνικές μείωσης αποχρώσεων.....	31
2.2 Ανάκτηση εικόνας βάσει περιεχομένου	34
2.2.1 Γενικά	34
2.2.2 Ιστορική Αναδρομή.....	35
2.2.3 Χαρακτηριστικές προσεγγίσεις στο CBIR	35
2.2.4 Συστατικά στοιχεία, ενός συστήματος CBIR	36
2.2.5 Τύποι χαρακτηριστικών.....	36
2.2.6 Χαρακτηριστικά χαμηλού επιπέδου	37
2.2.6.1 Χαρακτηριστικά χρώματος	37
2.2.6.2 Χαρακτηριστικά Υφής	38
2.2.7 Μέτρα ομοιότητας.....	39
2.2.8 Αξιολόγηση συστήματος.....	40
2.2.8.1 Μέτρο αξιολόγησης Precision/Recall	41
2.2.8.2 Μέτρο αξιολόγησης ANMRR.....	42
2.2.9 Περιγραφείς οπτικού περιεχομένου στο MPEG-7.....	43
2.2.9.1 Color Space Descriptor	43
2.2.9.2 Color Quantization Descriptor	44
2.2.9.3 Dominant Color Descriptor	44
2.2.9.4 Scalable Color Descriptor	45
2.2.9.5 Color Layout Descriptor.....	46

2.2.9.6 Color Structure Descriptor	47
2.2.9.7 Group of Frames/Group of Pictures Color Descriptor.....	49
ΣΥΣΤΗΜΑ ΑΝΑΚΤΗΣΗΣ ΒΑΣΕΙ ΧΡΩΜΑΤΟΣ.....	51
3.1 Άποψη συστήματος	51
3.2 Προεπεξεργασία	52
3.3 Εξαγωγή χαρακτηριστικών	61
3.5 Μέτρο ομοιότητας	65
3.6 Αξιολόγηση συστήματος.....	65
3.6.1 Πειραματική διαδικασία	65
3.6.2 Παρουσίαση αποτελεσμάτων	67
3.6.3 Σύγκριση αποτελεσμάτων.....	83
3.7 Συμπεράσματα.....	99
ΒΙΒΛΙΟΓΡΑΦΙΑ.....	101

Εισαγωγή

1.1 Εισαγωγή

Η ανάλυση έγχρωμων εικόνων για την εξαγωγή χαρακτηριστικών που να αντικατοπτρίζουν το περιεχόμενό τους σαν μια μορφή ψηφιακής υπογραφής αποτελεί ένα δυναμικό πεδίο έρευνας και ανάπτυξης με άμεση συσχέτιση με την ανάπτυξη του Web 3.0 και των εφαρμογών σημασιολογικού ιστού. Στόχος της έρευνας είναι η εύρεση αλγορίθμων και τεχνικών που να καθιστούν δυνατή την εξαγωγή χαρακτηριστικών τα οποία στη συνέχεια να χρησιμοποιούνται για την ταύτιση εικόνων ή για την ανάκτηση εικόνων βάσει περιεχομένου (χρωματικού, δομικού ή, γενικότερα, σημασιολογικού). Ένα από τα ζητούμενα στην εν λόγω έρευνα είναι η επίτευξη εξαγωγής χαρακτηριστικών ανεξάρτητων του μεγέθους και της περιστροφής της εικόνας (rotation and scale invariant).

1.2 Στόχος της εργασίας

Στόχος της παρούσας εργασίας ήταν η ανάπτυξη μεθοδολογίας ανάλυσης δισδιάστατων έγχρωμων στατικών εικόνων για την εξαγωγή καθολικών χαρακτηριστικών που να μην επηρεάζονται από την περιστροφή ή την κλιμάκωση και η υλοποίηση συστήματος ανάκτησης βάσει περιεχομένου που να εκμεταλλεύεται τα εν λόγω χαρακτηριστικά. Η έρευνα επικεντρώθηκε κυρίως στο χρωματικό περιεχόμενο χωρίς να απορρίπτει το περιεχόμενο υφής αν αυτό κριθεί αναγκαίο.

1.3 Ο ρόλος της εικόνας για τον άνθρωπο

Εικόνα ορίζεται ως η αναπαράσταση ενός πραγματικού ή φανταστικού αντικειμένου με πλαστικά μέσα, η οπτική αντίληψη μιας σκηνής ή η εντύπωση ενός δρώμενου. Ο σύγχρονος άνθρωπος μπορεί έτσι να έρχεται σε επαφή με τον πολιτισμό των προγόνων του, οι οποίοι αποτύπωσαν τις δραστηριότητές τους, σε κάποια εικόνα. Σπηλαιογραφίες από την Ύστερη Παλαιολιθική εποχή, περίπου 65.000-35.000 π.Χ., συμπληρώνουν τις ειδικασίες μας, για τον τρόπο ζωής τους. Η τέχνη της τοιχογραφίας ως χωρικός διάκοσμος αναπτύσσεται τη Μεσομινωική περίοδο (2000-1930 π.Χ.) και ήταν σύμβολο πολυτέλειας και ευμάρειας. Οι τοιχογραφικές θεματο-

λογίες ποίκιλαν ανάλογα με το χώρο που διακοσμούσαν. Σήμερα αποτελούν ισχυρές ενδείξεις, για τις θρησκευτικές τελετουργίες, τις πομπές και το κυνήγι, καθώς και η ίδια η τεχνική της τοιχογραφίας αποτελεί απόδειξη της εξέλιξης του ανθρώπινου πνεύματος. Εξαίρεση δεν αποτελεί η αγγειογραφία, η οποία εμφανίζεται περίπου το 6-4π.Χ. Στη μορφή αυτή, οι εικόνες αποτυπώνουν τον άνθρωπο και τα αισθήματα του. Κατά το Μεσαίωνα, μετά την ασύλληπτη καταστροφή έργων τέχνης, εντοπίζονται ξανά εικόνες με μορφή συμβόλων. Την εποχή του Βυζαντίου, η θεματική απεικόνιση περιορίζεται στις αγιογραφίες. Όμως ακόμη και αυτή η περιορισμένη θεματολογία προσφέρει αρκετές πληροφορίες, για την ανθρώπινη πνευματική και κοινωνική πορεία. Αμέσως μετά, έρχεται η Αναγέννηση, γύρω στο 15^ο-16^ο αιώνα μ.Χ. Στην περίοδο αυτή, η τέχνη της ζωγραφικής απογειώνεται. Ο άνθρωπος αρχίζει και αποτυπώνει τα καθημερινά δράματα και τις ανθρώπινες ανησυχίες στους καμβάδες του. Το μέσο απεικόνισης μιας σκηνής, ως τότε ήταν η ζωγραφική, πράγμα που σήμαινε ότι η πραγματικότητα που αποτυπωνόταν στον καμβά, ήταν συνυφασμένη με την υποκειμενικότητα του καλλιτέχνη. Οι πίνακες ήταν ακριβοί και χρονοβόροι σε αντίθεση με τη λήψη φωτογραφίας. Περίπου στα τέλη του 19^{ου} αιώνα, τοποθετείται η εφεύρεση της πρώτης μηχανής λήψεως φωτογραφιών, αν και οι αρχές της φωτογραφίας ήταν γνωστές από την αρχαιότητα (π.χ. camera obscura). Το αποτέλεσμα ήταν η αποτύπωση της εικόνας να πάρει πλέον άλλες διαστάσεις. Η απεικόνιση κάποιου θέματος δεν ήταν πλέον μόνο προνόμιο των καλλιτεχνών της ζωγραφικής, αλλά και των κατόχων φωτογραφικών μηχανών. Το εύρος των θεματικών εννοιών του περιεχομένου των εικόνων οριοθετήθηκε από τη φαντασία και την οπτική γωνία των φωτογράφων. Στη σημερινή εποχή, η εικόνα δεν αφορά μόνο την εικαστική αντίληψη του ανθρώπου, αλλά έχει περισσότερες προεκτάσεις. Οι δημοσιογράφοι καταγράφουν στο φακό τους κάποιο δράμα. Οι γραφίστες διαφημίζουν προϊόντα και υπηρεσίες μέσω της εικόνας. Οι ιατροί φτάνουν σε πορίσματα για την υγεία του ασθενούς, βάσει ειδικών ιατρικών εικόνων. Οι αστυνομικοί στη προσπάθεια της πρόληψης και παρεμπόδισης του εγκλήματος χρησιμοποιούν εικόνες, για ταυτοποίηση δακτυλικών αποτυπωμάτων ή προσώπων. Η σύντομη αυτή ιστορική αναδρομή δείχνει μόνο κατά ένα ελάχιστο πόσο σημαντικό ρόλο παίζει η εικόνα στη ζωή του ανθρώπου ως μέσο επικοινωνίας και πόσο σημαντική είναι η εκμετάλλευσή της πληροφορίας του περιεχομένου της.

1.4 Η διάδοση της εικόνας στη σημερινή εποχή

Στα τέλη του 20^{ου} αιώνα, με την καθιέρωση του διαδικτύου και της ψηφιοποίησης των δεδομένων, το πλήθος των εικόνων στον παγκόσμιο ιστό World Wide Web (www), υπολογίστηκε στα 10-30 εκατομμύρια εικόνες [1]. Δύο δεκαετίες αργότερα, με την ανάπτυξη των κοινωνικών δικτύων η ανάρτηση ψηφιακών φωτογραφιών στο διαδίκτυο αγγίζει τα 55 εκατομ-

μύρια ημερησίως **Error! Reference source not found.**, ενώ ένας διεθνούς χρήσης ιστότοπος σχετικός με τη φωτογραφία μετράει 65 δισεκατομμύρια εικόνες [1]. Από τα νούμερα αυτά γίνεται περισσότερο από ποτέ αισθητή η ανάγκη για συστήματα αναζήτησης, ταξινόμησης, περιγραφής και αποδοτικής αποθήκευσης των ψηφιακών εικόνων. Η ανάγκη της διαχείρισης των ψηφιακών μέσων, απασχόλησε τόσο την επιστημονική όσο και τη βιομηχανική κοινότητα σχεδόν από το 1992 [3]. Τη χρονιά εκείνη, όπως αναφέρουν οι Smeulders et al., “το Εθνικό Ίδρυμα Ερευνών των Η.Π.Α. (USNSF) οργάνωσε ένα συνέδριο με σκοπό την αναγνώριση των τομέων που θα έπρεπε να προσανατολιστούν οι ερευνητές, για τα οπτικού-περιεχομένου συστήματα διαχείρισης, τα οποία θα ήταν χρήσιμα ως επιστημονικές, βιομηχανικές, ιατρικές περιβαλλοντικές, εκπαιδευτικές, ψυχαγωγικές κ.α. εφαρμογές” [5]. Καρποί αυτού του ενδιαφέροντος, ενδεικτικά αποτελούν τα παρακάτω συστήματα:

- **QBIC:** Το QBIC (Query By Image Content) είναι το πρώτο εμπορικό σύστημα ανάκτησης εικόνων βάσει περιεχομένου. Σχεδιάστηκε και υλοποιήθηκε από την IBM το 1995. Η δομή και οι τεχνικές του συστήματος επηρέασαν τα επόμενα συστήματα ανάκτησης εικόνων. Η ανάκτηση γίνεται μέσω εικόνων-παραδειγμάτων, εικόνες σχεδιασμένες από το χρήστη, μοτίβα, τύπο κάμερας και επιλεγμένων χρωμάτων [6].
- **VisualSEEk:** Το VisualSEEk είναι ένα σύστημα ανάκτησης εικόνων βάσει περιεχομένου, το οποίο υλοποιήθηκε από το πανεπιστήμιο Columbia. Χρησιμοποιεί χαρακτηριστικά υφής που βασίζονται στο μετασχηματισμό κυματιδίων, καθώς και χρωματικά και χωρικά χαρακτηριστικά, για την περιγραφή του περιεχομένου των εικόνων. Ήταν το πρώτο σύστημα, το οποίο συνδύασε τα παραπάνω χαρακτηριστικά, για την ανάκτηση των αποτελεσμάτων [8].
- **Photobook:** Το Photobook υλοποιήθηκε από το Massachusetts Institute of Technology. Το σύστημα αυτό δίνει τη δυνατότητα στο χρήστη να ψάχνει και να οργανώνει γρήγορα τις εικόνες του μέσω μίας Βάσης Δεδομένων Τεχνητής Νοημοσύνης. Οι εικόνες περιγράφονται τόσο από κείμενο όσο και από το περιεχόμενο της εικόνας [9].
- **Google Image Search:** Η εταιρεία Google, μέχρι πρότινος, χρησιμοποιούσε κείμενο περιγραφής και μεταδεδομένα, για την ανάκτηση εικόνων. Από το 2012 δίνει τη δυνατότητα στο χρήστη να αναζητήσει εικόνες και μέσω παραδείγματος, χρησιμοποιώντας το περιγραφείς περιεχομένου **Error! Reference source not found.**
- **Caliph and Emir:** Το Caliph and Emir είναι ένα σύστημα ανάκτησης εικόνων βασισμένο στους περιγραφείς περιεχομένου του προτύπου MPEG-7. Το εργαλείο Caliph (common and light

weight photo annotation), υποστηρίζει χαρακτηριστικά χαμηλού επιπέδου και μεταδεδομένα. Τα μεταδεδομένα αποθηκεύονται σύμφωνα με το σχήμα MPEG-7 XML. Το εργαλείο Emir (experimental metadata based image retrieval), αναζητά σε συλλογές του MPEG-7, οι οποίες έχουν δημιουργηθεί από το Caliph και υποστηρίζει αναζήτηση βάσει περιεχομένου χαρακτηριστικών χαμηλού επιπέδου και αναζήτηση βάσει κειμένου και λέξεων-κλειδιών[11].

1.5 Το πρότυπο MPEG-7

Παράλληλα με τα παραπάνω συστήματα, πραγματοποιήθηκε και μια προσπάθεια για προτυποποίηση. Αποτέλεσμα της προσπάθειας αυτής είναι το διεθνές πρότυπο MPEG-7 (ISO/IEC 15938 – Multimedia Content Description Interface, από το 2002). Ειδικότερο ενδιαφέρον για την παρούσα εργασία παρουσιάζει το Μέρος 3 του προτύπου "ISO/IEC 15938-3:2002 – Multimedia Content Description Interface – Part 3: Visual". Το όνομα MPEG είναι ακρώνυμο του Moving Picture Experts Group και υπάγεται στο πλαίσιο του Διεθνούς Οργανισμού Προτυποποίησης (ISO). Η επιτροπή αυτή λειτουργεί με σκοπό την ανάπτυξη προτύπων για τη κωδικοποίηση οπτικών, ακουστικών, καθώς και το συνδυασμό αυτών μέσων. Το πρότυπο MPEG-7, σε αντίθεση με άλλα πρότυπα της οικογένειας MPEG δε σχετίζεται με την κωδικοποίηση πολυμεσικών δεδομένων αλλά με την περιγραφή του περιεχομένου τους. Το πλούσιο σύνολο εργαλείων που προσφέρει, επιτρέπει την υποστήριξη διάφορων πιθανών εφαρμογών διαχείρισης δεδομένων και υλικού [12]. Τα συστατικά του προτύπου είναι:

1. **Οι περιγραφείς**, οι οποίοι ορίζουν τη σύνταξη και τη σημασιολογία των χαρακτηριστικών και των σχημάτων περιγραφής
2. **Η γλώσσα προσδιορισμού περιγραφής**, η οποία ορίζει τη σύνταξη των εργαλείων της περιγραφής και επιτρέπει την επέκτασή τους
3. **Τα εργαλεία συστήματος**, τα οποία είναι σημαντικά για την υποστήριξη κωδικοποίησης της αναπαράστασης του πολυμεσικού περιεχομένου

Το MPEG-7 χρησιμοποιείται ευρέως σε εφαρμογές περιγραφής πολυμεσικού περιεχομένου καθώς και ανάκτησης βάσει περιεχομένου όπως:

- Αρχεία οπτικοακουστικών μέσων
- Τηλεοπτικά δίκτυα
- Οικιακή ψυχαγωγία
- Οπτικοακουστικές βιβλιοθήκες
- Ηλεκτρονικό εμπόριο

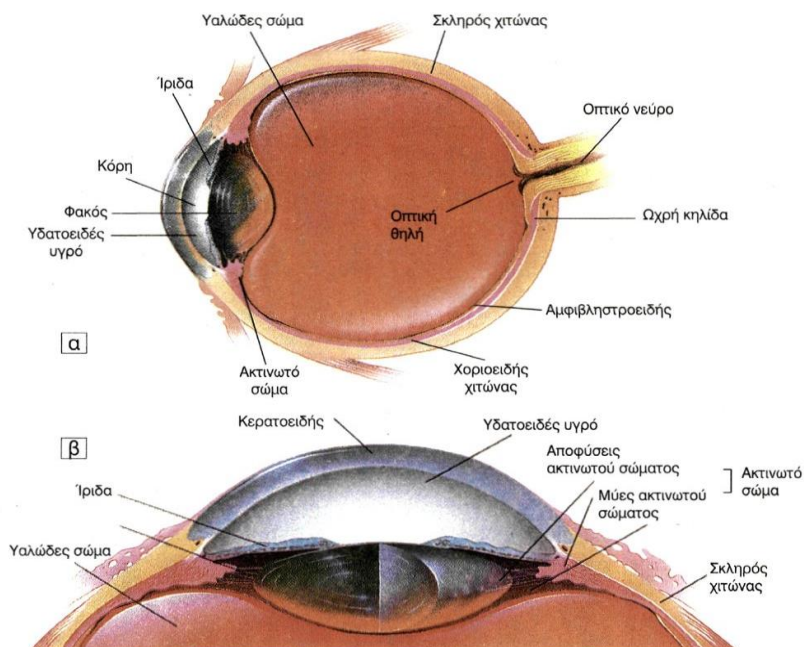
Η ψηφιακή εικόνα

2.1 Η εικόνα και το περιεχόμενο της

Η ανθρώπινη αντίληψη και ανάλυση παίζει σημαντικό ρόλο στην επιλογή της προσέγγισης, όσον αφορά στην επεξεργασία της εικόνας. Ενδιαφέρον παρουσιάζουν τόσο «τεχνικά» χαρακτηριστικά της όρασης, όπως η διακριτική ικανότητα και η προσαρμογή σε αλλαγές φωτισμού, περιστροφής του σκηνικού και απομάκρυνσης των αντικειμένων, όσο και αντιληπτικά χαρακτηριστικά που σχετίζονται με την αναγνώριση του περιεχομένου μιας εικόνας από τον ανθρώπινο εγκέφαλο. Επομένως, η κατανόηση της λειτουργίας των οφθαλμών και της οπτικής αντίληψης του ανθρώπου είναι καθοριστική για την ανάλυση κάθε σχετικού προβλήματος.

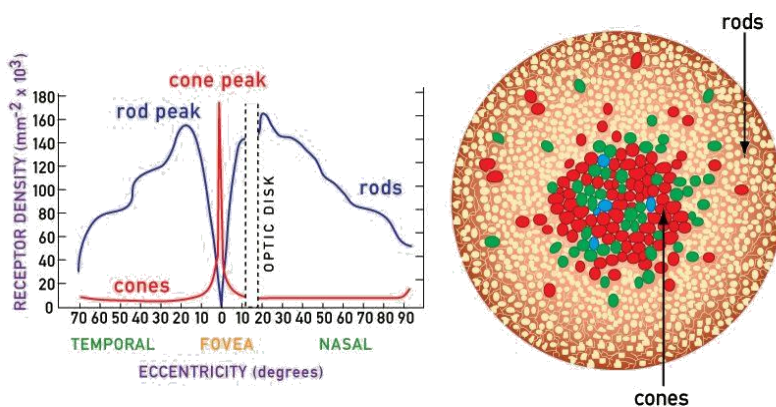
2.1.1 Το ανθρώπινο μάτι

Το μάτι έχει σχεδόν σφαιρικό σχήμα και περιβάλλεται από τρεις μεμβράνες. Οι δύο εξωτερικές είναι ο κερατοειδής και ο σκληρός χιτώνας. Η εσωτερική μεμβράνη είναι ο αμφιβληστροειδής χιτώνας και καλύπτει το οπίσθιο τμήμα του ματιού. Στην Εικόνα 1 παρουσιάζεται η κάθετη τομή ενός αριστερού ματιού. Όταν το μάτι εστιάσει κατάλληλα, το φως που ανακλάται από το αντικείμενο εμπρός από το φακό απεικονίζεται στον αμφιβληστροειδή. Πάνω σε αυτή τη μεμβράνη, υπάρχουν ειδικά κύτταρα, οι φωτοϋποδοχείς [13]. Οι φωτοϋποδοχείς αυτοί είναι δύο ειδών, τα κωνία και τα ραβδία. Αυτά τα κύτταρα είναι υπεύθυνα για την αντίληψη τόσο της φωτεινότητας όσο και της χρωματικότητας του φωτός. Τα κωνία μετρούνται σε ποσότητες των 6-7 εκατομμυρίων και βρίσκονται στο κεντρικό τμήμα του αμφιβληστροειδούς. Είναι τριών ειδών και είναι εξαιρετικά ευαίσθητα στο χρώμα. Σε αυτά αποδίδεται η ικανότητα του ανθρώπου να διακρίνει το χρώμα και τις λεπτομέρειες στο κέντρο της όρασης. Τα ραβδία έχουν μεγαλύτερο πλήθος, ο αριθμός τους κυμαίνεται μεταξύ 75-150 εκατομμύρια και είναι κατανομημένα στην επιφάνεια του αμφιβληστροειδούς με μεγαλύτερη παρουσία στην περιφέρεια. Τα ραβδία είναι υπεύθυνα για την περιφερειακή όραση και είναι ευαίσθητα στη διακύμανση της φωτεινότητας και όχι στο χρώμα. Αυτό τα καθιστά ικανά να διακρίνουν εικόνες ακόμη και με χαμηλό φωτισμό.



Εικόνα 1. Κάθετη τομή οφθαλμού

Στην Εικόνα 2 φαίνεται η πυκνότητα των κωνίων και ραβδίων γύρω από το οπτικό νεύρο. Στο δεξιό μέρος της εικόνας, φαίνεται η κατανομή των ραβδίων με κίτρινο χρώμα, ενώ με κόκκινο, μπλε και πράσινο χρώμα απεικονίζονται τα κωνία.



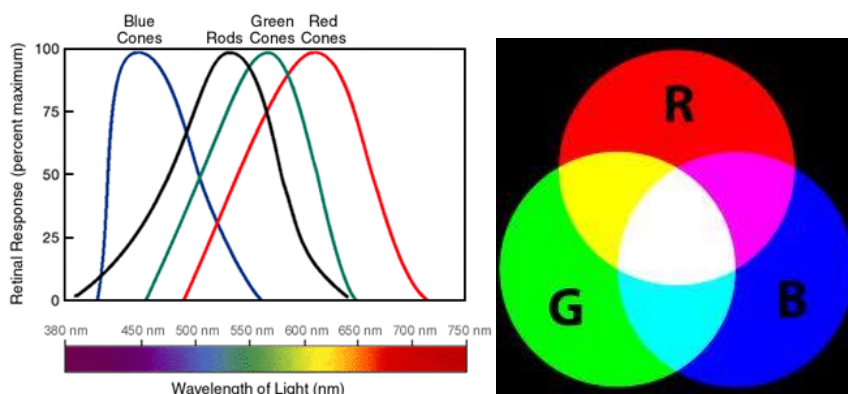
Εικόνα 2. Αριστερά: διάγραμμα πυκνότητας κωνίων και ραβδίων, δεξιά: διάγραμμα κατανομής των φωτοϋποδοχών πάνω στον αμφιβληστροειδή

2.1.2 Η τριχρωματική θεωρία

Ο T. Young το 1802, διατύπωσε τη θεωρία του, πως στο μάτι υπάρχουν τριών ειδών φωτοϋποδοχείς, οι οποίοι είναι υπεύθυνοι για την αντίληψη του χρώματος (κωνία)

[14]. Η κάθε μία ομάδα είναι ευαίσθητη σε διαφορετικό πεδίο του μήκους κύματος του ορατού φάσματος του φωτός. Αργότερα, ο H. Helmholtz, επέκτεινε τη θεωρία αυτή και διατύπωσε πως οι τρεις ομάδες είναι συγκεκριμένα ευαίσθητες στο μπλε, πράσινο και κόκκινο φως. Ωστόσο, το κάθε χρώμα μπορεί να περιγραφεί μαθηματικά σε διαφορετικές ποσότητες των τριών βασικών χρωμάτων. Στην Εικόνα 3 Εικόνα 3 Αριστερά, η απόκριση των κωνίων και ραβδίων στο ορατό φάσμα φωτός. Δεξιά, ο συνδυασμός των τριών βασικών χρωμάτων

φαίνεται η απόκριση των κωνίων στο ορατό φάσμα του φωτός και τα χρώματα που προκύπτουν από το συνδυασμό των τριών βασικών χρωμάτων (στο ονομαζόμενο προσθετικό σύστημα RGB).



Εικόνα 3 Αριστερά, η απόκριση των κωνίων και ραβδίων στο ορατό φάσμα φωτός. Δεξιά, ο συνδυασμός των τριών βασικών χρωμάτων

Η θεωρία των Young -Helmoltz, η οποία καλείται και τριχρωματική θεωρία, έδωσε το έναυσμα στη CIE (Commission internationale de l'éclairage) το 1931 να προτυποποιήσει τις τυπικές τιμές του μήκους κύματος των τριών βασικών χρωμάτων. Έτσι, μετά από μία πειραματική διαδικασία, κατέληξαν πως το τυπικό μπλε έχει μήκος κύματος 420–440nm, το πράσινο 530–540nm, και το κόκκινο 560–580nm.

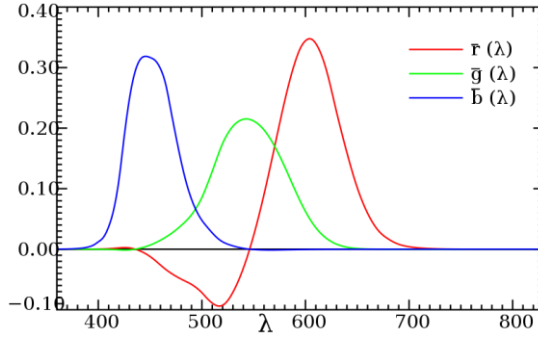
2.1.3 Ο χρωματικός χώρος CIEXYZ

Από τη παραπάνω θεωρία, πως κάθε χρώμα αποτελείται από τη μίξη των τριών βασικών χρωμάτων, προκύπτουν οι εξισώσεις (2.1.1)–(2.1.3), για λ το μήκος κύματος και $I(\lambda)$ την κατανομή του φάσματος, ενώ στην Εικόνα 4 παρουσιάζονται οι αντίστοιχες καμπύλες απόκρισης.

$$X = \int_{380nm}^{780nm} I(\lambda)x(\lambda)d\lambda \quad (2.1.1)$$

$$Y = \int_{380nm}^{780nm} I(\lambda)y(\lambda)d\lambda \quad (2.1.2)$$

$$Z = \int_{380nm}^{780nm} I(\lambda)z(\lambda)d\lambda \quad (2.1.3)$$



Εικόνα 4 Καμπύλες απόκρισης εξισώσεων

2.1.4 Ο χρωματικός χώρος CIExyY

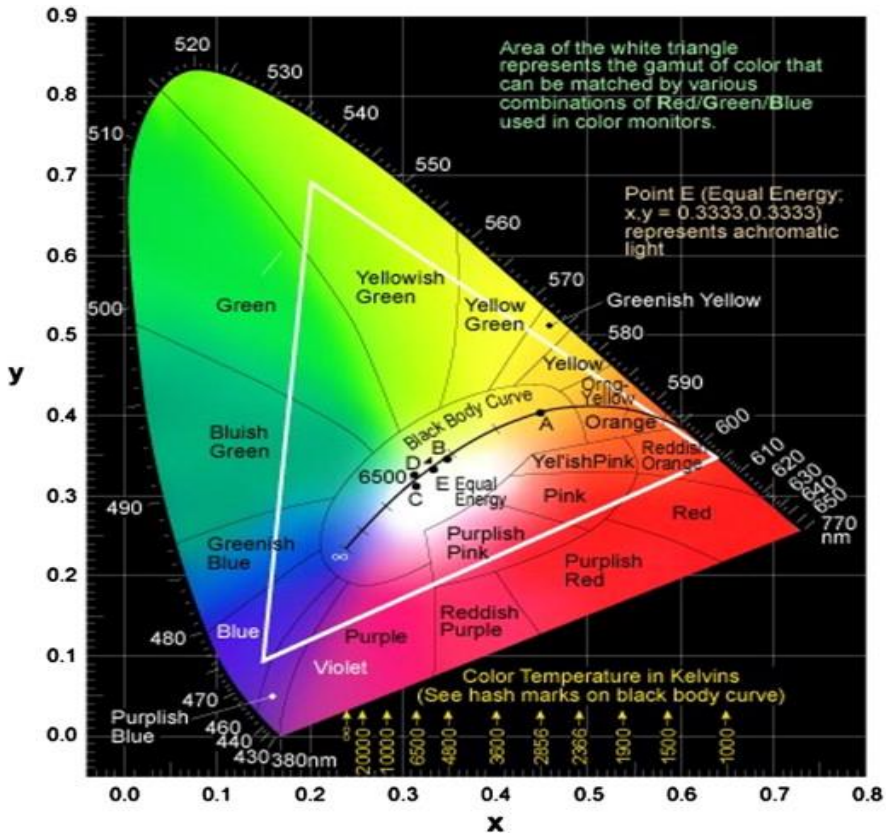
Ο χώρος CIEXYZ είναι ένας χώρος, στον οποίο απεικονίζονται τα χρώματα ως προσθετικές ποσότητες. Ένας άλλος τρόπος απεικόνισης του χρωματικού χώρου είναι ο CIExyY, ο οποίος περιλαμβάνει και τις έννοιες όπως η φωτεινότητα, η απόχρωση και ο κορεσμός. Η φωτεινότητα είναι το μέτρο «αχρωματικότητας» της έντασης. Η απόχρωση αναπαριστά το κυρίαρχο μήκος κύματος από το μείγμα των κυμάτων φωτός. Και τέλος, ο κορεσμός αναφέρεται στη ποσότητα του λευκού ή αλλιώς στη σχετική καθαρότητα του χρώματος [15]. Ο συνδυασμός του κορεσμού και της απόχρωσης ονομάζεται χρωματικότητα και επομένως ένα χρώμα μπορεί να χαρακτηρίζεται από τη φωτεινότητα και τη χρωματικότητα του. Οι εξισώσεις που προκύπτουν, συμβολίζουν το κάθε χρώμα:

$$x = \frac{X}{X + Y + Z} \quad (2.1.4)$$

$$y = \frac{Y}{X + Y + Z} \quad (2.1.5)$$

$$z = \frac{Z}{X + Y + Z} = 1 - x - y \quad (2.1.6)$$

Στην Εικόνα 5, φαίνεται το διάγραμμα χρωματικότητας του χώρου CIExyY. Στην περιοχή που περικλείεται από το λευκό τρίγωνο τοποθετούνται οι αποχρώσεις που μπορούν να αναπαράγουν οι τυπικές έγχρωμες οθόνες με την κλασική προσθετική μέθοδο RGB. Στο διάγραμμα διαχωρίζονται επίσης οι περιοχές των βασικών χρωμάτων που αναγνωρίζονται ακτινικά από την κεντρική λευκή περιοχή προς τις παρυφές με αντίστοιχη σήμανση των μηκών κύματος. Στην κεντρική περιοχή όπου οι τιμές των x, y είναι και οι δύο ίσες με 0.333 υπάρχει το σημείο αχρωματικού φωτός (λευκό στο σημείο E). Τέλος στο διάγραμμα τοποθετούνται και οι ενδείξεις θερμοκρασίας χρώματος όπως προκύπτουν κατ' αντιστοιχία από την ακτινοβολία μέλανος σώματος στα αντίστοιχα μήκη κύματος.



Εικόνα 5. Ο χρωματικός χώρος CIExyY

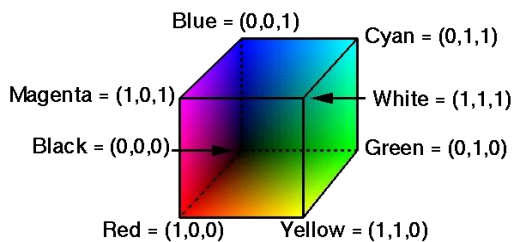
2.1.5 Χρωματικά μοντέλα

Τα χρωματικά μοντέλα έχουν σκοπό τη διευκόλυνση του προσδιορισμού ενός χρώματος με γενικά αποδεκτό τρόπο. Ο προσδιορισμός αυτός γίνεται εφικτός μέσω ενός συστήματος συντεταγμένων. Τα περισσότερα χρωματικά μοντέλα είναι προσανατολισμένα στο υλικό, όπως οι εκτυπωτές και οι οθόνες. Τα πιο συχνά χρησιμοποιούμενα χρωματικά μοντέλα είναι τα παρακάτω, τα οποία και αναλύονται στη συνέχεια:

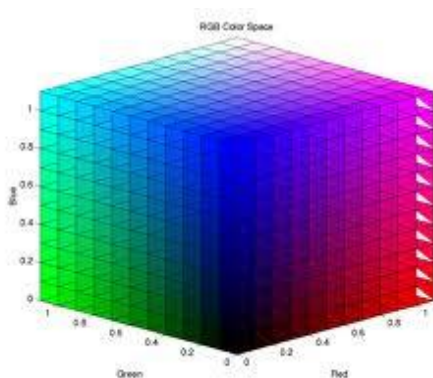
- RGB: ευρεία χρήση στις έγχρωμες οθόνες και κάμερες
- CMYK: έγχρωμη εκτύπωση
- HSI, HLS: περιγραφή των χρωμάτων σύμφωνα με την ανθρώπινη αντίληψη
- YCbCr: τηλεοπτικές μεταδόσεις, πρότυπα MPEG-1, MPEG-2, MPEG-4 και JPEG με ταύτιση με την οπτική αντίληψη
- HMMD: ταύτιση με οπτική αντίληψη για χρήση από το MPEG-7
- CIELab: μέγιστη ταύτιση με την ανθρώπινη αντίληψη

2.1.5.1 Το χρωματικό μοντέλο RGB

Στο χρωματικό μοντέλο RGB, το χρώμα περιγράφεται στις τρεις κύριες συνιστώσες του, Red (κόκκινο), Green (πράσινο) και Blue (μπλε). Το μοντέλο αυτό στηρίζεται στο σύστημα καρτεσιανών συντεταγμένων. Ο χρωματικός χώρος που αντιστοιχεί σε αυτό το μοντέλο σε σχέση με το CIExyY φαίνεται στην Εικόνα 5, ως ο χώρος μέσα στο λευκό τρίγωνο. Ο χώρος του μοντέλου αυτού είναι ένας κύβος, όπως φαίνεται στην Εικόνα 6 και τα τρία βασικά χρώματα τοποθετούνται σε τρεις γωνίες [17]. Οι υπόλοιπες γωνίες αντιστοιχούν στα δευτερεύοντα χρώματα κυανό, μωβ και κίτρινο. Το μαύρο και το άσπρο, αντιστοιχούν στην αρχή του συστήματος των συντεταγμένων και στο πέρας της διαγωνίου από την αρχή των αξόνων έως το σημείο συνάντησης των ακμών CMY αντίστοιχα. Οι τιμές που μπορούν να πάρουν οι τρεις συντελεστές είναι στο διάστημα $[0, 1]$. Ο χώρος μέσα στον κύβο, ορίζει όλους τους συνδυασμούς των χρωμάτων που προκύπτουν από τα τρία βασικά χρώματα, όπως φαίνεται στην Εικόνα 7. Το συνολικό πλήθος των χρωμάτων σε μία RGB εικόνα με βάθος χρώματος 24 bits ανά εικονοστοιχείο (24 bpp) είναι $(2^8)^3 = 16777216$ ή 16,7 εκατομμύρια χρώματα.



Εικόνα 6. Ο κύβος RGB, και οι τιμές των παραμέτρων στις γωνίες



Εικόνα 7. Χρωματική διαβάθμιση του RGB χώρου

2.1.5.2 Το χρωματικό μοντέλο CMYK

Στην προηγούμενη ενότητα ειπώθηκε πως τα χρώματα κυανό, μωβ και κίτρινο, αποτελούν τα δευτερεύοντα ή συμπληρωματικά χρώματα του φωτός [17]. Η μετατροπή μεταξύ του RGB και του CMY γίνεται μέσω της απλής εξίσωσης (2.1.7).

$$\begin{bmatrix} C \\ M \\ Y \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (2.1.7)$$

2.1.5.3 Το χρωματικό μοντέλο YCrCb

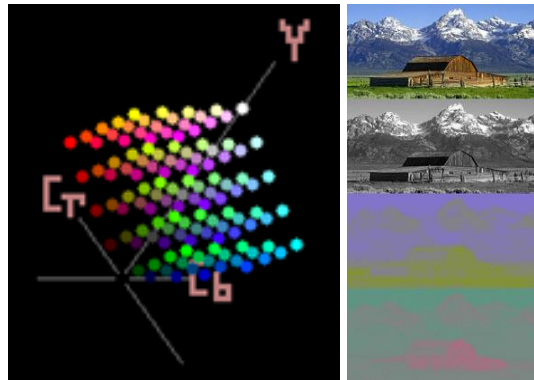
Το μοντέλο YCbCr χρησιμοποιείται στη τηλεοπτική μετάδοση σήματος, σε ψηφιακές κάμερες, στα πρότυπα JPEG, MPEG-1, MPEG-2, MPEG-4, και σε ένα μεγάλο εύρος εφαρμογών, οι οποίες έχουν ανάγκη για αποδοτική μετάδοση ή αποθήκευση οπτικού σήματος [17]. Οι παράμετροι που χρησιμοποιεί είναι η φωτεινότητα (Y-luma), και η χρωματικότητα, με Cr να συμβολίζει τη χρωματική απόκλιση μεταξύ κόκκινων και πράσινων αποχρώσεων και Cb να συμβολίζει τη χρωματική απόκλιση μπλε και κίτρινων αποχρώσεων αντίστοιχα. Το πεδίο τιμών φωτεινότητας (Y) είναι στο διάστημα $[0,1]$ από το σκοτεινό προς το φωτεινό, ενώ το πεδίο τιμών της χρωματικότητας (Cb, Cr) είναι στο διάστημα $[-0.5,0.5]$. Οι εξισώσεις (2.1.8)-(2.1.10) πραγματοποιούν τη μετατροπή από το μοντέλο RGB στο YCbCr. Οι παράμετροι R', G', B' είναι κανονικοποιημένες τιμές, με εύρος $[0,1]$. Οι σταθερές K_r και K_b έχουν τιμές 0.299 και 0.144 αντίστοιχα.

$$Y' = K_R R' + (1 - K_R - K_B) G' + K_B B' \quad (2.1.8)$$

$$P_B = \frac{1}{2} \frac{B' - Y'}{1 - K_B} \quad (2.1.9)$$

$$P_R = \frac{1}{2} \frac{R' - Y'}{1 - K_R} \quad (2.1.10)$$

Στην Εικόνα 8 φαίνεται ο χώρος συντεταγμένων του μοντέλου YCbCr και ένα παράδειγμα των καναλιών του σε πραγματική έγχρωμη εικόνα.



Εικόνα 8. Αριστερά, οι συνιστώσες του YCbCr, δεξιά, τα κανάλια Y, Cb, και Cr

2.1.5.4 Το χρωματικό μοντέλο HSI

Τα χρωματικά μοντέλα που παρουσιάστηκαν στις παραπάνω παραγράφους είναι προσανατολισμένα στις ανάγκες του υλικού. Πέρα από αυτά έχουν σχεδιαστεί και χρωματικά μοντέλα, τα οποία είναι πιο κοντά στην ανθρώπινη αντίληψη. Δύο από αυτά είναι τα HIS και HSV, όπου Hue η χρωματικότητα, Saturation ο κορεσμός του χρώματος και Intensity ή Value η ένταση του φωτός με διαφορετικό αλλά αντίστοιχο τρόπο. Ένα τέτοιο χρωματικό μοντέλο διαχωρίζει την πληροφορία της έντασης από τη χρωματική πληροφορία, πράγμα που το καθιστά ιδανικό για ανάπτυξη αλγορίθμων με βάση τις περιγραφές των χρωμάτων [17]. Παρακάτω συνοψίζονται οι εξισώσεις μετασχηματισμού από το μοντέλο RGB στο μοντέλο HSI.

$$R' = \frac{R}{255} \quad (2.1.11)$$

$$G' = \frac{G}{255} \quad (2.1.12)$$

$$B' = \frac{B}{255} \quad (2.1.13)$$

$$C_{max} = \max(R', G', B') \quad (2.1.14)$$

$$C_{min} = \min(R', G', B') \quad (2.1.15)$$

$$\Delta = C_{max} - C_{min} \quad (2.1.16)$$

$$H = \begin{cases} 60^\circ \times \left(\frac{G' - B'}{\Delta} \text{mod} 6 \right), C_{max} = R' \\ 60^\circ \times \left(\frac{B' - R'}{\Delta} + 2 \right), C_{max} = G' \\ 60^\circ \times \left(\frac{R' - G'}{\Delta} + 4 \right), C_{max} = B' \end{cases} \quad (2.1.17)$$

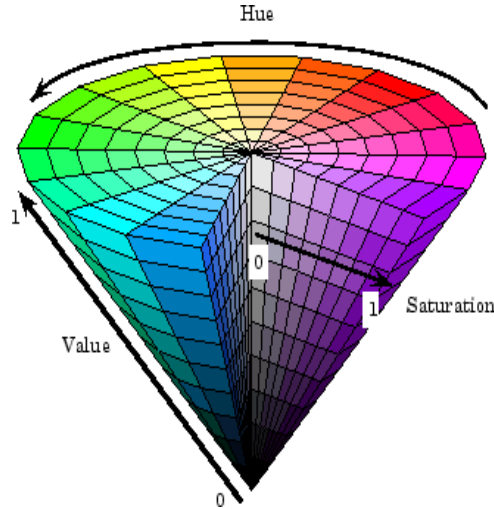
$$S = \begin{cases} 0, & \Delta = 0 \\ \frac{\Delta}{C_{max}}, & \Delta \neq 0 \end{cases} \quad (2.1.18)$$

$$I = C_{max} \quad (2.1.19)$$

Από τις εξισώσεις (2.1.11)-(2.1.19) προκύπτουν ως πεδία τιμών

- Για τη Χρωματικότητα H $[0^\circ, 360^\circ]$ μοίρες
- Για το Κορεσμό S $[0, 1]$, από την απουσία έως τη πλήρη καθαρότητα του χρώματος
- Για την Ένταση I $[0, 1]$, από την έλλειψη έντασης (μαύρο) έως τη μέγιστη ένταση (λευκό)

Από τα παραπάνω προκύπτει πως ο χώρος του HSI αναπαρίσταται ως ένας ανάποδος κώνος, όπως φαίνεται στην Εικόνα 9.



Εικόνα 9. Ο ανάποδος κώνος του μοντέλου HSI

2.1.5.5 Το χρωματικό μοντέλο HMMD

Το HMMD (Hue-Max-Min-Diff) είναι και αυτός ένας χώρος που στοχεύει να δώσει πιστή αναπαράσταση χρώματος σύμφωνα με την ανθρώπινη αντίληψη του χρώματος [17]. Ο όρος της χρωματικότητας είναι όμοιος με εκείνον που ορίστηκε στο μοντέλο HSI και είναι ως εξής:

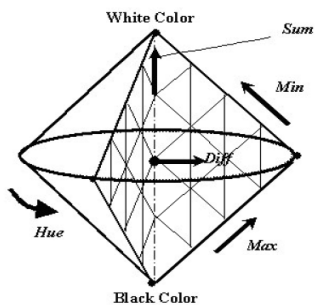
$$H = \begin{cases} 60^\circ \times \left(\frac{G' - B'}{\Delta} \bmod 6 \right), C_{max} = R' \\ 60^\circ \times \left(\frac{B' - R'}{\Delta} + 2 \right), C_{max} = G' \\ 60^\circ \times \left(\frac{R' - G'}{\Delta} + 4 \right), C_{max} = B' \end{cases} \quad (2.1.20)$$

Οι παράμετροι Max, Min, και Diff υπολογίζονται όπως παρακάτω:

Πίνακας Ι. Υπολογισμός των παραμέτρων του HMMD

Εξίσωση	Πεδίο Τιμών	Περιγραφή	
$Max = \max(R', G', B')$	[0,1]	Πόσο κοντά είναι στο μαύρο	(2.1.21)
$Min = \min(R', G', B')$	[0,1]	Πόσο κοντά είναι στο λευκό	(2.1.22)
$Diff = Max - Min$	[0,1]	Πόσο καθαρό είναι ένα χρώμα	(2.1.23)
$Sum = \frac{Max + Min}{2}$	[0,1]	Πόσο φωτεινό είναι ένα χρώμα	(2.1.24)

Η αναπαράσταση του χρωματικού μοντέλου στο χώρο παραπέμπει σε ένα διπλό κώνο, όπως φαίνεται στην Εικόνα 10.



Εικόνα 10. Ο διπλός κώνους του HMMD

2.1.5.6 Το χρωματικό μοντέλο CIELab

Τα μοντέλα παρουσιάστηκαν στις προηγούμενες παραγράφους, απομονώνουν το χαρακτηριστικό της χρωματικότητας από την ένταση ή τη φωτεινότητα. Εάν όμως το επιθυμητό χαρακτηριστικό είναι η ευαισθησία στις χρωματικές εναλλαγές, οι χρωματικά ομοιόμορφοι χώροι, προσφέρουν αυτή τη λύση [17]. Τέτοιοι χώροι είναι οι CIEL*u*v και CIEL*a*b. Τα δύο αυτά χρωματικά συστήματα έχουν μικρή διαφορά μεταξύ τους και στηρίζονται στην αντιληπτική φωτεινότητα L^* , καθώς και σε δύο συνιστώσες, οι οποίες καθορίζουν την ανάμιξη κόκκινου-πράσινου και κίτρινου-μπλε. Το πιο καθιερωμένο μεταξύ των δύο, είναι το CIEL*a*b. Ο μετασχηματισμός από το RGB στο CIELAB γίνεται σε δύο βήματα:

1. με γραμμικό μετασχηματισμό από το RGB στο CIEXYZ
2. με μη γραμμικό μετασχηματισμό από το CIEXYZ στο CIELAB.

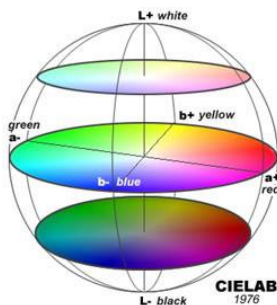
$$L^* = 116f(Y/Y_n) - 16 \quad (2.1.25)$$

$$a^* = 500[f(X/X_n) - f(Y/Y_n)] \quad (2.1.26)$$

$$b^* = 200[f(Y/Y_n) - f(Z/Z_n)] \quad (2.1.27)$$

$$f(t) = \begin{cases} t^{1/3} & \text{if } t > \left(\frac{6}{29}\right)^3 \\ \frac{1}{3} \left(\frac{29}{6}\right)^2 t + \frac{4}{29} & \text{otherwise} \end{cases} \quad (2.2.28)$$

Στην Εικόνα 11 φαίνεται αναπαράσταση του συστήματος CIELAB, που αντιστοιχεί σε σφαίρα.



Εικόνα 11. Ο χρωματικός χώρος CIELAB

2.1.6 Τι είναι η ψηφιακή εικόνα

Η ψηφιακή εικόνα είναι ένα δισδιάστατο σήμα της μορφής $f \rightarrow f(x, y)$. Υπάρχουν, βασικά, τρία είδη ψηφιακών εικόνων, οι δυαδικές εικόνες, οι εικόνες κλίμακας του γκρι και οι έγχρωμες εικόνες. Στις έγχρωμες εικόνες μπορούμε να ξεχωρίσουμε τις τυπικές εικόνες τριών καναλιών αλλά και τις πολυφασματικές εικόνες που μπορούν να αποτελούνται από πολλά χρωματικά κανάλια. Οι τελευταίες δεν αποτελούν αντικείμενο της παρούσας εργασίας. Ακόμη στις έγχρωμες εικόνες μπορούμε να εντοπίσουμε και τις εικόνες παλέτας, οι οποίες είναι επίσης ειδική περίπτωση εκτός του σκοπού της παρούσας εργασίας αν και θα μπορούσε μια επέκταση της μεθόδου που αναπτύχθηκε να καλύψει και αυτές τις εικόνες.

Οι δυαδικές εικόνες είναι εικόνες που περιέχουν μόνο δύο τόνους, λευκό και μαύρο. Οι εικόνες της κλίμακας του γκρι, καλύπτουν το φάσμα μεταξύ του μαύρου και του λευκού και σε τυπικές εφαρμογές με εικόνες των 8 bpp οι τιμές των δύο άκρων είναι 0 και 255 αντίστοιχα. Με συνολικά 256 τιμές, η πληροφορία ανά εικονοστοιχείο, είναι 1 byte. Οι έγχρωμες εικόνες ακολουθούν τη τριχρωματική θεωρία, επομένως η χρωματική αναπαράσταση αποτελείται από τρία κανάλια χρώματος με τιμές 0 έως 255 (για 8 bpp ανά κανάλι), από το σκοτεινότερο προς το φωτεινότερο. Αν κάθε χρώμα κωδικοποιηθεί με έναν αριθμό, τότε η ψηφιακή εικόνα περιγράφεται από ένα πίνακα διαστάσεων $N \times M$ και περιεχόμενο τις τιμές των χρωμάτων της εικόνας [17]. Τέλος, το σύνολο των χρωμάτων που μπορεί να χρησιμοποιηθεί σε μια εικόνα, ονομάζεται χρωματική παλέτα.

2.1.6.1 Τεχνικές μείωσης αποχρώσεων

Μία εικόνα I διαστάσεων $M \times N$ και πλήθους αποχρώσεων $G=2^m$ χρειάζεται $b = M * N * m \text{ bits}$, για να αποθηκευτεί, όπου m το βάθος χρώματος. Στο Πίνακα II, γίνεται φανερό η ανάγκη για μείωση τόσο των διαστάσεων όσο και των αποχρώσεων των ψηφιακών εικόνων για αποδοτική αποθήκευση και δικτυακή μετάδοση. Οι προσπάθειες για μείωση της πληροφορίας της εικόνας προφανώς δεν πρέπει να συνδέονται με υποβάθμιση της ποιότητάς της.

Πίνακας II. Απαιτούμενος αριθμός bits, για την αποθήκευση εικόνας

Τύπος εικόνας	N	M	m	bits	bytes
Δυαδική	256	256	1	65.536	8.192
Αποχρώσεων του γκρι	256	256	8	524.288	65.536
Έγχρωμη	256	256	24	1.572.684	196.608
Δυαδική	512	512	1	262.144	32.768
Αποχρώσεων του γκρι	512	512	8	2.097.152	262.144
Έγχρωμη	512	512	24	6.291.456	786.432

Στην εργασία [18] παρουσιάζεται η μέθοδος Otsu, η οποία είναι μια τεχνική μείωσης του πλήθους των αποχρώσεων του γκρι, με τη χρήση κατωφλίου. Το κριτήριο, το οποίο χρησιμοποιείται είναι η μεγιστοποίηση του διαχωρισμού μεταξύ των φωτεινών και των σκοτεινών περιοχών. Έστω, μια εικόνα με $L=256$ επίπεδα φωτεινότητας. Το πλήθος των εικονοστοιχείων με επίπεδο φωτεινότητας i συμβολίζεται ως n_i , ενώ το πλήθος των εικονοστοιχείων συμβολίζεται ως $N = n_1 + n_2 + \dots + n_L$. Το ιστόγραμμα της εικόνας κανονικοποιείται ως εξής: $p_i = \frac{n_i}{N}, p_i \geq 0, \sum_{i=1}^L p_i = 1$. Στη συνέχεια, χωρίζονται τα εικονοστοιχεία σε δύο κλάσεις το φόντο και τα αντικείμενα, C_0, C_1 με $C_0 = \{g_1, g_2, \dots, g_k\}$ και $C_1 = \{g_{k+1}, g_{k+2}, \dots, g_L\}$. Η πιθανότητα εμφάνισης κάποιου εικονοστοιχείου σε μια από τις δύο κλάσεις είναι

$$\omega_0 = P(C_0) = \sum_{i=1}^k p_i = \omega(k) \quad (2.1.29)$$

$$\omega_1 = P(C_1) = \sum_{i=k+1}^L p_i = 1 - \omega_0 \quad (2.1.30)$$

Υπολογίζονται και τα κέντρα των κλάσεων:

$$\mu_0 = \sum_{i=1}^k i p_i = \frac{\mu(k)}{\omega(k)} \quad (2.1.31)$$

$$\mu_1 = \sum_{i=k+1}^L i p_i = \frac{\mu_L - \mu(k)}{1 - \omega(k)} \quad (2.1.32)$$

$$\text{όπου } \mu(k) = \sum_{i=1}^k i p_i \text{ και } \mu_L = \mu(L) = \sum_{i=1}^L i p_i \quad (2.1.33)$$

Οι μεταβλητότητες προκύπτουν από τις δευτέρου βαθμού κεντρικές ροπές:

$$\sigma_0^2 = \sum_{i=1}^k \frac{(i - \mu_0)^2 p_i}{\omega_0} \quad (2.1.34)$$

$$\sigma_1^2 = \sum_{i=1}^L \frac{(i - \mu_1)^2 p_i}{\omega_1} \quad (2.1.35)$$

Τέλος, το βέλτιστο κατώφλι υπολογίζεται με μία από τις τρεις συναρτήσεις:

$$\lambda(k) = \frac{\sigma_B^2}{\sigma_w^2} \quad \eta(k) = \frac{\sigma_B^2}{\sigma_L^2} \quad \delta(k) = \frac{\sigma_L^2}{\sigma_w^2} \quad (2.1.36)$$

$$(2.1.37)$$

$$(2.1.38)$$

$$\sigma_w^2 = \omega_0 \sigma_0^2 + \omega_1 \sigma_1^2 \quad (2.1.39)$$

$$\sigma_B^2 = \omega_0 (\mu_0 - \mu_L)^2 + \omega_1 (\mu_1 - \mu_L)^2 = \omega_0 \omega_1 (\mu_1 - \mu_0)^2 \quad (2.1.40)$$

$$\sigma_L^2 = \sum_{i=1}^L (i - \mu_L)^2 p_i \quad (2.1.41)$$

Στη Εικόνα 12 φαίνεται παράδειγμα κατωφλίωσης με τη μέθοδο Otsu.



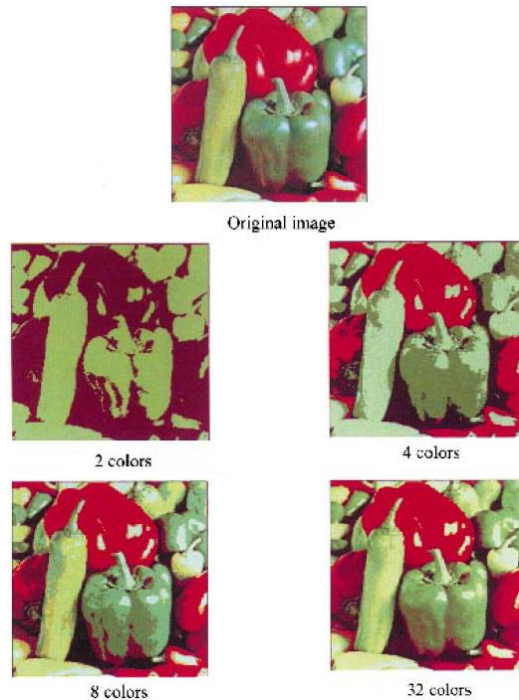
Εικόνα 12. Κατωφλίωση με τη μέθοδο Otsu

Για τις έγχρωμες εικόνες η διαδικασία της μείωσης των χρωμάτων είναι πιο περίπλοκη και συνήθως λύνεται με τη χρήση ταξινομητών. Οι ταξινομητές χρησιμοποιούν τόσο την πληροφορία του χρώματος όσο και τοπικά χαρακτηριστικά. Στην [19] αναπτύσσεται μία μέθοδος, η οποία βασίζεται στη χρήση δύο τεχνητών νευρωνικών αυτοδιοργανούμενων δικτύων, τα οποία συνδέονται διαδοχικά. Το πρώτο είναι ένα δίκτυο Ανάλυσης Κυρίων Συνιστοσών (PCA) [20], ενώ το δεύτερο είναι ένας Αυτοδιοργανώμενος Πίνακας Χαρακτηριστικών Kohonen (SOFM) [21]. Το νευρωνικό δίκτυο τροφοδοτείται με τοπικά χαρακτηριστικά και χρωματική πληροφορία. Τα τοπικά χαρακτηριστικά είναι η υφή, η αντίθεση, και γενικά οποιοδήποτε χωρικό χαρακτηριστικό θα μπορούσε να εξαχθεί και να ποσοτικοποιηθεί.

Το νευρωνικό δίκτυο PCA επιτρέπει την εύρεση αντιπροσωπευτικών προτύπων πολυδιάστατων διανυσμάτων, ίσου ή μικρότερου μεγέθους με εκείνων της εισόδου, στρέφοντας το σύστημα συντεταγμένων ώστε οι πρώτοι άξονες να αντιστοιχούν στα δεδομένα μεγαλύτερης διασποράς.

Το νευρωνικό δίκτυο SOFM έχει ως κύρια λειτουργία την ταξινόμηση ενός συνόλου διανυσμάτων εισόδου σε ένα μονοδιάστατο, και σε κάποιες περιπτώσεις το πολύ σε δισδιάστατο, διακριτό πίνακα. Για την εκπαίδευση του νευρωνικού ταξινομητή, χρησιμοποιείται ένα υποσύνολο της εικόνας, το οποίο προκύπτει έπειτα από τεχνικές δειγματοληψίας. Μετά την εκπαίδευση του ταξινομητή, σαρώνεται πάλι η εικόνα, αλλά αναπλάθεται μόνο με τα χρώματα που υπέδειξε ο ταξινομητής. Στη μέθοδο αυτή είναι δυνατό να παραληφθούν τοπικά χαρακτηριστικά και να χρησιμοποιηθεί μόνο η χρωματική πληροφορία, η οποία ικανοποιεί τη χρωματική κατα-

νομή της αρχικής εικόνας. Στην Εικόνα 13 παρουσιάζεται παράδειγμα της μεθόδου.



Εικόνα 13 Αποτελέσματα μείωσης χρωμάτων σε έγχρωμη εικόνα

Γενικά από την αντίστοιχη βιβλιογραφία [22]-[24],[27] προκύπτει ότι οι αλγόριθμοι, οι οποίοι στοχεύουν στη μείωση του πλήθους των χρωμάτων με βέλτιστο και δυναμικό τρόπο, χρησιμοποιούν ταξινομητές. Η παραγόμενη παλέτα των χρωμάτων δεν είναι σταθερή για όλες τις εικόνες, αλλά προσαρμόζεται ανάλογα, ώστε η τελική εικόνα να είναι όσο το δυνατό χρωματικά πλησιέστερη ως προς την αρχική.

2.2 Ανάκτηση εικόνας βάσει περιεχομένου

2.2.1 Γενικά

Ο κλάδος της Ανάκτησης Εικόνας Βάσει Περιεχομένου (Content Based Image Retrieval ή CBIR) είναι συγγενής των κλάδων της Ψηφιακής Επεξεργασίας Εικόνας, της Θεωρίας της Πληροφορίας, και της Μηχανικής Οράσεως. Για αυτόν το λόγο, μοιράζεται πολλά στοιχεία με αυτές. Ο όρος, CBIR φαίνεται να αποδίδεται στον Kato, ο οποίος το χρησιμοποιεί, για να περιγράψει τα πειράματά του, ως προς την αυτόματη ανάκτηση εικόνων από μία Βάση Δεδομένων (ΒΔ) [25].

2.2.2 Ιστορική Αναδρομή

Οι αρχικές προσεγγίσεις στο CBIR έγιναν στις αρχές της δεκαετίας του 1980. Δε στηρίζονταν στα οπτικά χαρακτηριστικά της εικόνας, αλλά στη χειροκίνητη περιγραφή των εικόνων με κείμενο από τους χρήστες. Η ανάκτηση των εικόνων γίνονταν με τεχνικές που απαντώνται στον κλάδο της ανάκτησης εγγράφων βάσει κειμένου.

Στη δεκαετία του 1990, λόγω του Internet και της ψηφιακής λήψης φωτογραφιών, ο όγκος των εικόνων αυξήθηκε ραγδαία. Ανάλογη αύξηση παρατηρήθηκε και στις επιστημονικές εργασίες γύρω από το CBIR [29],[31]. Οι προσεγγίσεις αυτές εκμεταλλεύονται τα οπτικά χαρακτηριστικά του περιεχομένου των εικόνων, όπως το χρώμα, την υφή, και τα σχήματα [34]-[36]. Τα χαρακτηριστικά που εξάγονται, έχουν τη μορφή πολυδιάστατων διανυσμάτων, τα οποία και αποθηκεύονται σε αναλόγως σχεδιασμένες ΒΔ. Η ομοιότητα και η ταξινόμηση των εικόνων βασίζεται στην απόσταση, η οποία προκύπτει μεταξύ των διανυσμάτων περιγραφής των οπτικών χαρακτηριστικών [37].

Μέσα σε αυτό το διάστημα αναπτύχθηκε και το πρότυπο MPEG-7, το οποίο μεταξύ άλλων παρέχει αλγορίθμους εξαγωγής χαρακτηριστικών, μέτρα σύγκρισης και γλώσσα περιγραφής των διανυσμάτων, για το οπτικό περιεχόμενο των εικόνων.

Οι πλέον πρόσφατες μελέτες επικεντρώθηκαν στην ταξινόμηση των αποτελεσμάτων, μέσω της αναφοράς στη κρίση του χρήστη (relevance feedback) [42],[43]. Τα αποτελέσματα που παράγονται από τα εν λόγω συστήματα είναι περισσότερο ομοιόμορφα και νοηματικά συνδεδεμένα μεταξύ τους.

Τέλος, η τρέχουσα ερευνητική κατεύθυνση στοχεύει τεχνικές γύρω από τους λεγόμενους “σάκους οπτικών λέξεων” (bags of visual words) και το “οπτικό λεξιλόγιο” (visual vocabulary). Οι τεχνικές αυτές είναι εμπνευσμένες από τη κοινότητα της ανάκτησης κειμενικού περιεχομένου (text based retrieval) και επεξεργάζονται τοπικά χαρακτηριστικά των εικόνων [48].

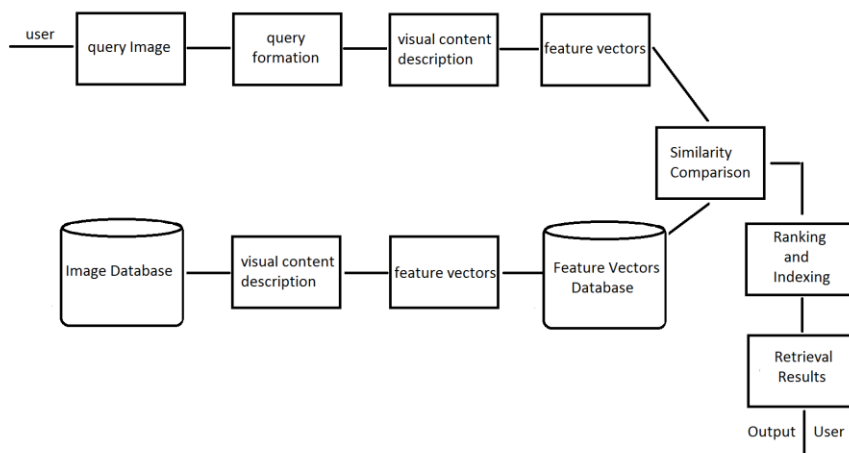
2.2.3 Χαρακτηριστικές προσεγγίσεις στο CBIR

Το CBIR χωρίζεται σε δύο κύριες προσεγγίσεις, τη συνεχή και διακριτή. Η συνεχής προσέγγιση, είναι η πιο συχνά χρησιμοποιούμενη. Η περιγραφή μιας εικόνας γίνεται από ένα ή περισσότερα διανύσματα χαρακτηριστικών και η σύγκριση τους γίνεται με διαφορετικά μέτρα ομοιότητας. Η μικρότερη απόσταση βαθμολογείται υψηλότερα, ως μεγαλύτερη ομοιότητα. Η συνεχής προσέγγιση, εκμεταλλεύεται τον ταξινομητή κοντινότερων γειτόνων (Nearest Neighbor Classifier). Η διακριτή προσέγγιση, περιγράφει την εικόνα με ένα πολυδιάστατο διάνυσμα χαρακτηριστικών. Το διάνυσμα αυτό, ονομάζεται και χώρος χαρακτηριστικών (feature space). Η προσέγγιση αυτή, μοιάζει με την ανάκτηση κειμένου, καθώς κάθε χαρα-

κτηριστικό παραλληλίζεται με μία λέξη, φτιάχνοντας τελικά ένα σάκο χαρακτηριστικών (bag of features) και αντιμετωπίζεται αναλόγως [47].

2.2.4 Συστατικά στοιχεία, ενός συστήματος CBIR

Ένα τυπικό σύστημα CBIR, έχει τη μορφή, που φαίνεται στην Εικόνα 14.



Εικόνα 14. Τυπικό σύστημα CBIR

Η λειτουργία του συστήματος γίνεται με μία καθορισμένη σειρά ενεργειών. Ο χρήστης εισάγει μία εικόνα, ως ερώτημα. Η εικόνα αυτή στη συνέχεια περνάει από τη φάση της προεπεξεργασίας, στην οποία τελούνται μετασχηματισμοί χρωματικών χώρων και εφαρμογή φίλτρων. Στη συνέχεια, εξάγονται τα επιθυμητά χαρακτηριστικά και σχηματίζεται ο χώρος χαρακτηριστικών, από τα διανύσματα περιγραφής. Αφού εξαχθούν τα χαρακτηριστικά από την εικόνα του ερωτήματος, ανακτώνται παρόμοια χαρακτηριστικά από μία Βάση Δεδομένων Χαρακτηριστικών, η οποία σχηματίστηκε με ανάλογο τρόπο από μία Βάση Δεδομένων Εικόνων. Το διάνυσμα περιγραφής του ερωτήματος συγκρίνεται με εκείνα που ανασύρθηκαν από τη Βάση Δεδομένων μέσω μέτρων ομοιότητας. Οι αποστάσεις που προκύπτουν ταξινομούνται και συνδέονται με τις εικόνες της Βάσης Δεδομένων. Τέλος, στο χρήστη παρουσιάζονται τα αποτελέσματα, τα οποία προέκυψαν από τις προηγούμενες ενέργειες.

2.2.5 Τύποι χαρακτηριστικών

Γενικά, χαρακτηριστικό μπορεί να θεωρηθεί οποιοδήποτε μετρήσιμο μέγεθος, το οποίο εξάγεται από μία εικόνα. Η επιλογή των χαρακτηριστικών, είναι ίσως η πιο σημαντική διεργασία ενός συστήματος CBIR. Τα χαρακτηριστικά πρέπει να είναι τέτοια, ώστε να περιγράφουν κατά το δυ-

νατόν καλύτερα το περιεχόμενο της εικόνας. Τα χαρακτηριστικά, τα οποία εξάγονται από μία εικόνα, χωρίζονται σε τρία επίπεδα:

- *Χαμηλό Επίπεδο*: αποτελείται από τα πρωτογενή χαρακτηριστικά μιας εικόνας. Τα χαρακτηριστικά αυτά είναι το χρώμα, η υφή, και τα σχήματα, τα οποία προκύπτουν κατευθείαν από τις τιμές των φωτεινότητων της εικόνας
- *Μεσαίο Επίπεδο*: αποτελείται από περιοχές ή δυαδικά αντικείμενα (blobs), τα οποία προκύπτουν μετά από κατάτμηση της εικόνας
- *Υψηλό επίπεδο*: περιέχει σημειολογική πληροφορία, για το περιεχόμενο μίας εικόνας, όπως η αναγνώριση αντικειμένων, ο σκοπός της ύπαρξης τους και η συσχέτιση με άλλα αντικείμενα, εντός της εικόνας

Σε οποιοδήποτε επίπεδο και αν ανήκουν τα χαρακτηριστικά που επιλέγονται για εξαγωγή, θεωρείται πολύ σημαντικό να ικανοποιούν κάποιες ιδιότητες:

- Μικρό υπολογιστικό κόστος
- Αξιοπιστία
- Διακριτικότητα
- Προσαρμογή με τη διαδικασία ταξινόμησης
- Ανεξάρτητο μετατόπισης, περιστροφής, και κλίμακας

2.2.6 Χαρακτηριστικά χαμηλού επιπέδου

Στα χαρακτηριστικά χαμηλού επιπέδου ανήκουν το χρώμα και υφή. Τα δύο αυτά χαρακτηριστικά προκύπτουν απευθείας από τις τιμές φωτεινότητας μιας εικόνας. Η πληροφορία του χρώματος λαμβάνεται από τις τιμές των τριών καναλιών R, G, και B. Ενώ για τη πληροφορία της υφής είτε το κάθε κανάλι αντιμετωπίζεται ως ξεχωριστή εικόνα αποχρώσεων του γκρι είτε γίνεται ένας μετασχηματισμός από την έγχρωμη εικόνα σε γκρι κλίμακα.

2.2.6.1 Χαρακτηριστικά χρώματος

Υπάρχουν διάφορες προσεγγίσεις, για τον τρόπο χειρισμού της πληροφορίας που προκύπτει από το χρώμα σε μια εικόνα και είναι το πιο διαδεδομένο χαρακτηριστικό, το οποίο χρησιμοποιείται στον τομέα του CBIR. Η πληροφορία του χρώματος, μπορεί να χωριστεί σε δύο κύριες κατηγορίες, την απόχρωση και τη φωτεινότητα. Η αναπαράσταση της απόχρωσης και της φωτεινότητας με ιστογράμματα, δίνει τη δυνατότητα της άντλησης στατιστικών πληροφοριών, για την κατανομή του χρώματος στην εικόνα. Ένα διάνυσμα χαρακτηριστικών μπορεί να δημιουργηθεί, από τις πληροφορίες οι οποίες προκύπτουν άμεσα από τα ιστογράμματα, όπως οι μονοδιάστατες ροπές, η εντροπία και κάποια κοινά στατιστικά χαρακτηριστικά, όπως φαίνονται στις εξισώσεις (2.2.1)-(2.2.12).

Η κατανομή πιθανότητας	$P(b) = \frac{N(b)}{M}$	(2.2.1)
Ροπές	$m_i = \sum_{b=0}^{L-1} b^i P(b)$	(2.2.2)
Απόλυτες Ροπές	$\hat{m}_i = \sum_{b=0}^{L-1} b ^i P(b)$	(2.2.3)
Κεντρικές ροπές	$\mu_i = \sum_{b=0}^{L-1} (b - m_1)^i P(b)$	(2.2.4)
Απόλυτες Κεντρικές Ροπές	$\hat{\mu}_i = \sum_{b=0}^{L-1} b - m_1 ^i P(b)$	(2.2.5)
Εντροπία	$H = - \sum_{b=0}^{L-1} P(b) \log_2[P(b)]$	(2.2.6)
Μέση Τιμή	$m_1 = \sum_{b=0}^{L-1} b P(b)$	(2.2.7)
Τυπική απόκλιση	$s = \left[\sum_{b=0}^{L-1} (b - m_1)^2 P(b) \right]^{1/2}$	(2.2.8)
Μεταβλητότητα	$\mu_2 = \sum_{b=0}^{L-1} (b - m_1)^2 P(b)$	(2.2.9)
Λοξότης	$\mu_3 = \sum_{b=0}^{L-1} (b - m_1)^3 P(b)$	(2.2.10)
Κύρτωση	$\mu_4 - 3 = \sum_{b=0}^{L-1} (b - m_1)^4 P(b)$	(2.2.11)
Ενέργεια	$E = \sum_{b=0}^{L-1} [P(b)]^2$	(2.2.12)

2.2.6.2 Χαρακτηριστικά Υφής

Η υφή είναι ένα χαρακτηριστικό, το οποίο γίνεται άμεσα αντιληπτό από τον άνθρωπο, αλλά είναι δύσκολο να οριστεί. Ο Hawkins [49], προσπαθώντας να διατυπώσει έναν ορισμό, κατέληξε πως η υφή παρουσιάζει τρεις ιδιότητες:

1. εμφάνιση τοπικής δομής σε μία περιοχή συγκρίσιμη ως προς το μέγεθος της εικόνας
2. η δομή αποτελείται από μη-τυχαία τοποθέτηση μερών της
3. τα μέρη είναι ομοιόμορφες οντότητες με περίπου ίδιες διαστάσεις

Οι πίνακες συνεμφάνισης (co-occurrence matrices), αποτελούν ένα στατιστικό εργαλείο, με το οποίο γίνεται εφικτή η περιγραφή της υφής σε μία εικόνα γκρι αποχρώσεων. Ο πίνακας συνεμφάνισης περιγράφει τη πιθανότητα διασύνδεσης μεταξύ μίας γκρι φωτεινότητας σε μία άλλη, δεδομένου ενός δομικού στοιχείου d και μίας κατεύθυνσης θ :

$$p_{d,\theta}(i, j) = \frac{C_{d,\theta}(i, j)}{\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} C_{d,\theta}(i, j)} \quad (2.2.13)$$

όπου,

$$C_{d,\theta} = (m, n), (u, v) \in N \times N : f(m, n) = j, f(u, v) = i, |(m, n) - (u, v)| = d, \angle((m, n), (u, v)) = \theta$$

Από ένα πίνακα συνεμφάνισης διαστάσεων $N \times M$, εξάγονται επιπλέον χαρακτηριστικά όπως η Εντροπία, η Ενέργεια, και η Αντίθεση, όπως ορίζονται από τις σχέσεις (2.2.14)-(2.2.23).

$$\text{Ενέργεια} \quad \sum_i \sum_j p_{\theta,d}^2(i, j) \quad (2.2.14)$$

$$\text{Αντίθεση} \quad \sum_i \sum_j \frac{p_{\theta,d}(i, j)}{1 + (i - j)^2} \quad (2.2.15)$$

$$\text{Ομογένεια} \quad \sum_i \sum_j (i - j)^2 p_{\theta,d}(i, j) \quad (2.2.16)$$

$$\text{Αντίστροφη δια-} \quad \sum_i \sum_j \frac{p_{\theta,d}(i, j)}{|i - j|^2}, i \neq j \quad (2.2.17)$$

$$\text{φορά ροπής} \quad - \sum_i \sum_j p_{\theta,d}(i, j) \log p_{\theta,d}(i, j) \quad (2.2.18)$$

$$\text{Εντροπία} \quad - \sum_i \sum_j p_{\theta,d}(i, j) \log p_{\theta,d}(i, j) \quad (2.2.19)$$

$$\text{Συσχέτιση} \quad \frac{\sigma_x \sigma_y}{\sum_i \sum_j (i - \mu_x)(j - \mu_y) p_{\theta,d}(i, j)} \quad (2.2.20)$$

$$\sigma_y^2 = \sum_i (i - \mu_y)^2 \sum_j p_{\theta,d}(i, j) \quad (2.2.21)$$

$$\sigma_x^2 = \sum_i (i - \mu_x)^2 \sum_j p_{\theta,d}(i, j) \quad (2.2.22)$$

$$\mu_x = \sum_i i \sum_j p_{\theta,d}(i, j) \quad (2.2.23)$$

$$\mu_y = \sum_j j \sum_i p_{\theta,d}(i, j) \quad (2.2.23)$$

$\mu = \mu_x = \mu_y$ for symmetric matches

2.2.7 Μέτρα ομοιότητας

Το τελευταίο τμήμα ενός συστήματος CBIR ευθύνεται για την ταύτιση «όμοιων» εικόνων και την αναφορά των αποτελεσμάτων ταύτισης. Συχνά, η ομοιότητα των εικόνων προκύπτει από τις αποστάσεις μεταξύ των διανυσμάτων περιγραφής. Οι αποστάσεις είναι αποτελέσματα γνωστών μέτρων ομοιότητας/ανομοιότητας ή μέτρων ειδικά κατασκευασμένων, ώστε να προσαρμόζονται στα χαρακτηριστικά που έχουν επιλεγεί [51].

Minkowski distance: αποτελεί το πιο δημοφιλές μέτρο απόστασης μεταξύ των διανυσμάτων περιγραφής. Το μέτρο ορίζεται ως εξής:

$$d_p(i, j) = \left(\sum_{k=1}^m |x_{ik} - x_{jk}|^p \right)^{\frac{1}{p}} \quad (2.2.24)$$

Τα x_i , και x_j είναι χαρακτηριστικά των περιγραφών δύο εικόνων I , και J αντίστοιχα. Όσο πιο μεγάλο είναι το p , τόσο μεγαλύτερη έμφαση δίνεται στην ανομοιομορφία μεταξύ των ζευγαριών.

City Block ή Manhattan distance: ονομάζεται η ειδική περίπτωση του μέτρου Minkowski, όταν η απόσταση των διανυσμάτων είναι το άθροισμα των απόλυτων διαφορών.

$$d_p(i, j) = \left(\sum_{k=1}^m |x_{ik} - x_{jk}| \right) \quad (2.2.25)$$

Ευκλείδεια απόσταση: ονομάζεται η ειδική περίπτωση του μέτρου Minkowski, όταν η απόσταση των διανυσμάτων είναι η τετραγωνική ρίζα του αθροίσματος των διαφορών των τετραγώνων τους, και ορίζεται ως εξής:

$$d_p(i, j) = \left(\sum_{k=1}^m |x_{ik} - x_{jk}|^2 \right)^{\frac{1}{2}} \quad (2.2.26)$$

Chebyshev ή Chessboard distance: Η απόσταση μεταξύ των διανυσμάτων, προκύπτει από τη μεγαλύτερη τιμή μεταξύ της διαφοράς των ζευγών.

$$d_p(i, j) = \max(|x_{i1} - x_{j1}|, |x_{i2} - x_{j2}|, \dots, |x_{im} - x_{jm}|) \quad (2.2.27)$$

Canberra distance: αποτελεί σταθμισμένη εκδοχή του μέτρου Manhattan. Το άθροισμα της εξίσωσης είναι μηδενικό μόνο εάν τα x_{jk} και x_{ik} είναι ίσα.

$$p_{d,\theta}(i, j) = \frac{C_{d,\theta}(i, j)}{\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} C_{d,\theta}(i, j)} \quad (2.2.28)$$

Τα παραπάνω μέτρα καθορίζουν την απόσταση μεταξύ των διανυσμάτων, συνεπώς και την ανομοιότητα. Η ομοιότητα μεταξύ των διανυσμάτων, προκύπτει από τη σχέση $S = 1 - D$, για κανονικοποιημένη τιμή απόστασης D .

2.2.8 Αξιολόγηση συστήματος

Αφού το σύστημα σχεδιαστεί και υλοποιηθεί, είναι σημαντικό να μετρηθεί και η απόδοση του. Η αξιολόγηση του συστήματος γίνεται με βάσει τα αποτελέσματα που επιστρέφονται, ως προς τη θεματική συνάφεια. Για να επιτευχθεί αυτό, χρησιμοποιούνται συγκεκριμένες Βάσεις Δεδομένων Εικόνων, οι οποίες έχουν κατασκευαστεί ειδικά για αυτό το λόγο. Αυτές οι Βάσεις Δεδομένων λέγονται και Β.Δ. αναφοράς (benchmark). Το σύστημα εκτελεί τις διεργασίες του και αφού επιστραφούν τα αποτελέσματα σε κατάταξη, οι αλγόριθμοι και τα μέτρα αξιολόγησης μετρούν την αποτελεσματικότητα του συστήματος. Σε αυτή την ενότητα, αναφέρονται τα βασικότερα μέτρα και οι πιο γνωστές μέθοδοι αξιολόγησης.

2.2.8.1 Μέτρο αξιολόγησης Precision/Recall

Το μέτρο αυτό, μετράει το ποσοστό των επιτυχημένων ανακτήσεων του συστήματος [50],[52]. Ορίζεται ως εξής: εφόσον, υπάρχει ο χάρτης αληθείας για τη συνάφεια των αποτελεσμάτων (Ground Truth), κάθε αποτέλεσμα της ανάκτησης μπορεί να ανήκει σε μία από τέσσερις πιθανές περιπτώσεις αληθείας (όπως φαίνεται και στον Πίνακα III).

- **True Positive:** σωστή ανάκτηση δείγματος
- **True Negative:** σωστή απόρριψη δείγματος
- **False Positive:** εσφαλμένη ανάκτηση δείγματος
- **False Negative:** εσφαλμένη απόρριψη δείγματος

Πίνακας III. Καταστάσεις χαρακτηρισμού ανάκτησης ενός αποτελέσματος

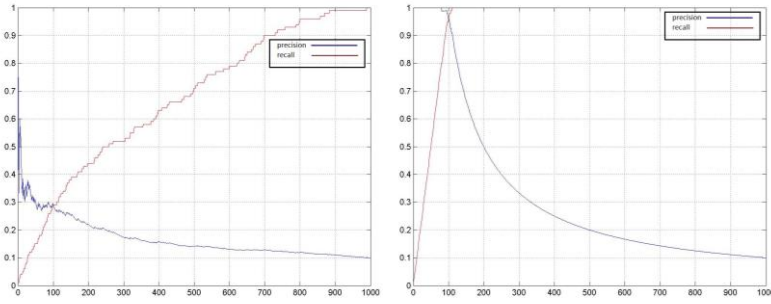
πραγματική κλάση (παρατηρούμενη)	
προβλεπόμενη κλάση (αναμενόμενη)	tp (true positive) Αληθές αποτέλεσμα
	fp (false positive) Μη αναμενόμενο αποτέλεσμα
	fn (false negative) Έλλειψη αποτελέσματος
	tn (true negative) Ορθή έλλειψη αποτελέσματος

Οι καταστάσεις αυτές είναι χρήσιμες, για να καθοριστεί το ποσοστό επιτυχίας του συστήματος. Η επιτυχία του συστήματος μετράται με τα μέτρα της ακριβείας (precision), και της ανάκλησης (recall), τα οποία ορίζονται όπως παρακάτω:

$$\text{Recall} = \frac{tp}{tp + fn} \quad (2.2.29)$$

$$\text{Precision} = \frac{tp}{tp + fp} \quad (2.2.30)$$

Οι τιμές των μέτρων αυτών, μπορούν να απεικονιστούν με γραφικές παραστάσεις, δημιουργώντας τις καμπύλες ακριβείας/ανάκλησης. Το σημείο τομής των καμπυλών απεικονίζει τη γενική απόδοση του συστήματος. Στη Εικόνα 15 φαίνονται δύο γραφικές παραστάσεις καμπυλών precision-recall, με μπλε και κόκκινο χρώμα αντίστοιχα. Από το σημείο τομής, της κάθε παράστασης μπορεί κανείς να συμπεράνει εάν το σύστημα είναι αποτελεσματικό ως προς την ανάκτηση σωστών αποτελεσμάτων. Η αριστερά παράσταση έχει σημείο τομής κοντά στο 0.3, το οποίο σημαίνει ότι η ανάκτηση των εικόνων είναι σχετικά «προβληματική», καθώς είναι κάτω του μετρίου (εάν μέτριο θεωρηθεί η τιμή 0.5). Σε αντίθεση, η δεξιά παράσταση έχει σημείο τομής κοντά στο 1, προφανώς το οποίο σημαίνει πως η ανάκτηση των αποτελεσμάτων είναι κοντά στην ιδανική.



Εικόνα 15. Γραφικές παραστάσεις precision/recall

Η μέθοδος της precision/recall δεν αξιολογεί τη κατάταξη των αποτελεσμάτων του συστήματος, αλλά την ικανότητα του να ανακτά σωστά αποτελέσματα. Στο σημείο αυτό θα πρέπει να αναφερθεί, πως το αποτέλεσμα της μεθόδου είναι άμεσα εξαρτώμενο από το είδος των συναφών αποτελεσμάτων (ground truth). Εάν ο εν λόγω είδος δεν είναι κατάλληλα επιλεγμένο, τότε ενδέχεται η αξιολόγηση του συστήματος να μην απεικονίζει τη πραγματικότητα. Ένα ακόμη πρόβλημα που αντιμετωπίζει η συγκεκριμένη μέθοδος είναι το διαφορετικό πλήθος των συναφών αποτελεσμάτων. Για κάποια ερωτήματα τα συναφή αποτελέσματα μπορεί να είναι δεκαπλάσια από κάποια άλλα. Επίσης, σε γενικές γραμμές, ένας χρήστης δε ζητάει ποτέ περισσότερα από 20 αποτελέσματα [62]. Επομένως, με τον υπολογισμό της μέσης τιμής ακρίβειας (mean average precision-MAP) για τα πρώτα M (top- N) αποτελέσματα, αντιμετωπίζεται και αυτό το ζήτημα. Η τιμή MAP υπολογίζεται ως εξής [55]:

$$MAP = \frac{1}{|Q|} \left(\sum_{Q_i} \frac{1}{|R|} \left(\sum_{j=1}^n rel(D_j) \sum_{k=1}^j \frac{rel(D_j)}{r_j} \right) \right) \quad (2.2.31)$$

Q: πλήθος ερωτημάτων

R: πλήθος συναφών απαντήσεων

D: πλήθος απαντήσεων στη κατάταξη r και n ανακτηθέντα αποτελέσματα

rel: συνάρτηση διαπίστωσης ομοιότητας αποτελεσμάτων, εκχωρεί 1 αν είναι όμοια

2.2.8.2 Μέτρο αξιολόγησης ANMRR

Το ANMRR προέρχεται από τον όρο Average Normalized Modified Ranked Retrieval. Σε αντίθεση με τη μέθοδο precision/recall, το μέτρο αυτό αξιολογεί το σύστημα ως προς τη κατάταξη των αποτελεσμάτων, που ανακτήθηκαν. Η τιμή του ANMRR καθορίζεται όπως παρακάτω [56] για $NG(q)$ το πλήθος εικόνων αληθείας (ground truth), για το ερώτημα q και $R(k)$ τη σειρά κατάταξης της εικόνας k στα αποτελέσματα ανάκτησης.

$$K(q) = \min(4 * NG(q), 2 * GTM), \quad \text{όπου } GTM = \max\{NG(q)\} \forall q \quad (2.2.31)$$

$$Rank(k) = \begin{cases} R(k) & \text{if } R(k) \leq K(q) \\ (K + 1) & \text{if } R(k) \geq K(q) \end{cases} \quad (1)$$

$$AVR(q) = \sum_{k=1}^{NG(q)} \frac{Rank(k)}{NG(q)} \quad (2)$$

$$MRR(q) = AVR(q) - 0.5 - \frac{NG(q)}{2} \quad (3)$$

$$NMRR(q) = \frac{MRR(q)}{K+0.5*NG(q)} \quad (4)$$

$$ANMRR(q) = \frac{1}{Q} \sum_{q=1}^Q NMRR(q) \quad (5)$$

2.2.9 Περιγραφείς οπτικού περιεχομένου στο MPEG-7

Όπως αναφέρθηκε προηγουμένως, το πρότυπο MPEG-7 προσφέρει περιγραφείς οπτικού περιεχομένου, οι οποίοι είναι ευρέως χρησιμοποιούμενοι σε εφαρμογές. Στις εργασίες [45],[54],[39],[40], γίνεται λεπτομερής αναφορά στο πρότυπο MPEG-7 και στους οπτικούς περιγραφείς. Οι περιγραφείς αυτοί είναι επτά στο σύνολο, αλλά χρησιμοποιούνται μόνο οι πέντε από αυτούς, καθώς ο *Color Space Descriptor* και ο *Color Quantization Descriptor* θεωρούνται βασικοί και είναι ενσωματωμένοι στους υπόλοιπους. Επιγραμματικά αναφέρονται τα ονόματα των περιγραφέων, καθώς στη συνέχεια θα αναλυθούν εκτενέστερα οι ορισμοί, και οι λειτουργίες τους (εξαγωγή χαρακτηριστικών και σύγκριση ομοιότητας):

- CSD - Color Space Descriptor
- CQD - Color Quantization Descriptor
- DCD - Dominant Color Descriptor
- SCD - Scalable Color Descriptor
- CLD - Color Layout Descriptor
- CSD - Color Structure Descriptor
- GoF/GoP - Group of Frames/Group of Pictures Color Descriptor

2.2.9.1 Color Space Descriptor

Ο περιγραφέας αυτός καθορίζει την επιλογή του χρωματικού χώρου, που θα χρησιμοποιηθεί από άλλον περιγραφέα. Οι χρωματικοί χώροι που ορίστηκαν στο πρότυπο MPEG-7 είναι ο μονοχρωματικός, ο RGB, ο HSV, ο HMMD και ο YCbCr (Πίνακας IV).

Πίνακας IV. Χρωματικοί χώροι που χρησιμοποιεί το πρότυπο MPEG-7

Περιγραφέας	Χρωματικός χώρος
SCD	HSV
CSD	HMMD
CLD	YCbCr
DCD	Οποιοσδήποτε έχει εγκριθεί από το MPEG-7

2.2.9.2 Color Quantization Descriptor

Ο CQD εξυπηρετεί την ανάγκη για διακριτή απεικόνιση των χρωμάτων της εικόνας, ορίζοντας τα επίπεδα της μείωσης των χρωμάτων για κάθε χρωματικό κανάλι.

2.2.9.3 Dominant Color Descriptor

Ο βασικός στόχος του DCD είναι η εφαρμογή του σε ανάκτηση εικόνων όμοιου περιεχομένου και στην περιήγηση Βάσεων Δεδομένων με εικόνες. Ο ορισμός του είναι ως εξής:

$$F = \{ \{c_i, p_i, v_i\}, s \}, (i = 1, 2, \dots, N)$$

όπου N είναι το πλήθος των κυρίαρχων χρωμάτων. Το πλήθος των κυρίαρχων χρωμάτων N μπορεί να διαφέρει από εικόνα σε εικόνα, ενώ ο περιορισμός κατά μέγιστο σε οκτώ χρώματα έχει χαρακτηριστεί αποδοτικός. Κάθε κυρίαρχο χρώμα με τιμή c_i , είναι ένα διάνυσμα, το οποίο αντιστοιχεί σε κάποιο χρωματικό χώρο. Το ποσοστό p_i , είναι το πλήθος των εικονοστοιχείων, που αντιπροσωπεύουν το χρώμα c , καθώς και $\sum p_i = 1$. Η διακύμανση της τιμής v_i περιγράφει τη διακύμανση των χρωματικών τιμών των εικονοστοιχείων γύρω από το κυρίαρχο χρώμα. Η χωρική συνοχή s είναι ένας απλός αριθμός, ο οποίος αντιπροσωπεύει τη χωρική ομογενοποίηση των κυρίαρχων χρωμάτων στην εικόνα. Η εξαγωγή του διανύσματος χαρακτηριστικών γίνεται σύμφωνα με τις εξισώσεις:

$$D_i = \sum_n h(n) \|x(n) - c_i\|^2, x(n) \in C_i \quad (2.2.37)$$

$$c_i = \frac{\sum h(n)x(n)}{\sum h(n)}, x(n) \in C \quad (2.2.38)$$

Η σύγκριση μεταξύ των δύο διανυσμάτων γίνεται ως εξής:

$$F_1 = \{ \{c_{1i}, p_{1i}, v_{1i}\}, s_1 \}, (i = 1, 2, \dots, N_1) \quad (2.2.39)$$

$$F_2 = \{ \{c_{2i}, p_{2i}, v_{2i}\}, s_2 \}, (i = 1, 2, \dots, N_2) \quad (2.2.40)$$

$$D^2(F_1, F_2) = \sum_{i=1}^{N_1} p_{1i}^2 + \sum_{j=1}^{N_1} p_{2j}^2 - \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} 2a_{1i,2j} p_{1i} p_{2j} \quad (2.2.41)$$

$$a_{k,l} = \begin{cases} 1 - \frac{d_{k,l}}{d_{max}} & , \quad d_{k,l} \leq T_d \\ 0 & , \quad d_{k,l} > T_d \end{cases} \quad (2.2.42)$$

$$d_{k,l} = \|c_k - c_l\| \quad (2.2.43)$$

Στην Εικόνα 16 παρουσιάζεται παράδειγμα έγχρωμης εικόνας και 6 διαφορετικών εκδοχών της μετά από μείωση αντίστοιχα σε 3, 4, 5, 6, 7, 8 χρώματα.



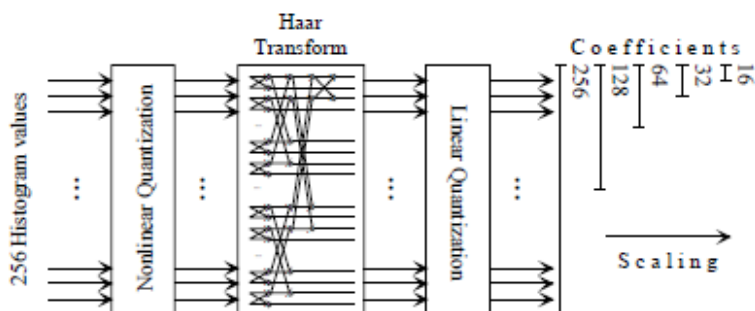
Εικόνα 16 Αριστερά η αρχική έγχρωμη εικόνα και δεξιά απεικονίσεις με χρήση 3-8 κυρίων χρωμάτων

2.2.9.4 Scalable Color Descriptor

Ο SCD μπορεί να ερμηνευτεί ως μια μέθοδος κωδικοποίησης εικόνων βασισμένη σε μετασχηματισμό Haar, ο οποίος εφαρμόζεται στις τιμές του ιστογράμματος των εικόνων στο χρωματικό χώρο HSV. Οι τιμές του ιστογράμματος κανονικοποιούνται και κβαντίζονται σε 4-bit αναπαράσταση, δίνοντας μεγαλύτερη έμφαση στις μικρότερες τιμές. Ο μετασχηματισμός Haar εφαρμόζεται στις ακέραιες τιμές του ιστογράμματος επαναληπτικά έως ότου το ιστόγραμμα αποτελείται πλέον μόνο από 16 bins (4 για το κανάλι H και 2 για τα κανάλια S,V). Στην Εικόνα 17 παρουσιάζεται η βασική μονάδα εφαρμογής του μετασχηματισμού Haar που λειτουργεί ως ένα στοιχειώδες χαμηλοπερατό και ένα υψηλοπερατό φίλτρο, ενώ στην Εικόνα 18 το πλήρες σχηματικό διάγραμμα της δημιουργίας του SCD.



Εικόνα 17. Βασική μονάδα του μετασχηματισμού Haar



Εικόνα 18. Σχηματικό διάγραμμα της δημιουργίας του SCD

2.2.9.5 Color Layout Descriptor

Ο CLD είναι περιεκτικός και ανεξάρτητος ανάλυσης, για υψηλής ταχύτητας ανάκτηση εικόνων. Είναι ιδιαίτερα χρήσιμος στις εφαρμογές που στηρίζονται στη δομή της χωρικής πληροφορίας. Επίσης, σε αντίθεση με τους υπόλοιπους περιγραφείς, επιτρέπει την αναζήτηση εικόνων βάσει ερωτήματος τύπου σκίτσου. Η εξαγωγή των χαρακτηριστικών γίνεται εφαρμόζοντας το διακριτό μετασχηματισμό συνημίτονου (DCT) σε ένα δισδιάστατο πίνακα (8X8 εικονοστοιχείων) από τα αντιπροσωπευτικά χρώματα της εικόνας στο χρωματικό χώρο YCbCr. Η διαδικασία αποτελείται από τέσσερα βήματα: κατάντηση της εικόνας, εντοπισμός αντιπροσωπευτικών χρωμάτων, μετασχηματισμός τύπου DCT, και τέλος μια μη γραμμική μείωση των συντελεστών που προκύπτουν με σάρωση ζιγκ-ζαγκ. Στο πρώτο στάδιο η αρχική εικόνα διαιρείται σε 64 τμήματα, ώστε να πετύχει ανεξαρτησία κλίμακας, και ανεξαρτησία ανάλυσης. Στο επόμενο στάδιο, το κυρίαρχο χρώμα του κάθε μπλοκ επιλέγεται ως αντιπροσωπευτικό (Εικόνα 19).



Εικόνα 19. Αντιπροσωπευτικά χρώματα εικόνας

Στο τρίτο στάδιο, κάθε κανάλι του χρωματικού χώρου μετασχηματίζεται από ένα 8x8 DCT, έτσι παράγονται τρία σετ από 64 συντελεστές DCT. Τέλος, στο τέταρτο στάδιο σαρώνονται με ζιγκ-ζιγκ και οι πρώτοι συντελεστές επιλέγονται, ώστε να μειωθούν με μη-γραμμικό τρόπο, χρησιμοποιώντας 64 και 32 επίπεδα για τους DC και AC συντελεστές. Το πρότυπο επιτρέπει την κλιμακωτή απεικόνιση των χαρακτηριστικών, αν και έχει συσταθεί η χρήση 12 συντελεστών, 6 για τη φωτεινότητα (Y) και 3 για κάθε χρωματικό κανάλι (Cb, Cr).

2.2.9.6 Color Structure Descriptor

Ένας περιγραφέας, ο οποίος εκμεταλλεύεται τη χωρική πληροφορία του χρώματος είναι ο CSD. Ο CSD περιγράφει την εικόνα τόσο χρησιμοποιώντας τη κατανομή του χρώματος, όσο και τοπικές δομές χωρικής πληροφορίας του χρώματος. Η ευαισθησία του περιγραφέα στη πληροφορία των τοπικών δομών, τον κάνει ικανό να αντιλαμβάνεται συγκεκριμένα χαρακτηριστικά, τα οποία δεν μπορεί να αξιοποιήσουν περιγραφείς βασισμένοι σε ιστογράμματα. Αυτό επιτυγχάνεται με τη χρήση ενός δομικού στοιχείου, το οποίο ανιχνεύει τη γειτονική κατανομή του χρώματος. Ο CSD μοιάζει σε μορφή με ιστόγραμμα χρώματος, αλλά σημασιολογικά διαφέρει. Συγκεκριμένα είναι ένας μονοδιάστατος πίνακας από 8-bit κβαντισμένες τιμές:

$$CSD = \bar{h}_s(m), m \in \{1, 2, \dots, M\} \quad (2.2.44)$$

όπου το M έχει επιλεχθεί από το σετ $\{256, 128, 64, 32\}$, και s είναι η κλίμακα του δομικού στοιχείου, το οποίο χρησιμοποιείται.

Η εξαγωγή του διανύσματος χαρακτηριστικών γίνεται σε τρία βήματα:

- Ένα ιστόγραμμα χρώματος-δομής με 256 τιμές εξάγεται από την εικόνα, η οποία έχει υποστεί μείωση χρωμάτων στον HMMD χρω-

ματικό χώρο. Σύμφωνα με πειράματα, ένα δομικό στοιχείο 4x4 εικονοστοιχείων είναι το ιδανικό για τη διαδικασία της εξαγωγής του ιστογράμματος. Για να αποτραπεί η σύγχυση με τις διαφορετικές διαστάσεις των εικόνων, εφαρμόζεται υποδειγματοληψία με συντελεστή $K = 2^p$, όπου το p ορίζεται ως:

$$p = \max\{0, \lfloor \log_2 \sqrt{W - H} - 7.5 \rfloor\} \quad (2.2.45)$$

με W και H το μήκος και το πλάτος της εικόνας αντίστοιχα, και όπου $\lfloor \cdot \rfloor$ είναι ο τελεστής *floor*.

- Το δεύτερο βήμα ορίζει πως, εάν $N < 256$, τότε είναι επιθυμητό οι τιμές να ενοποιηθούν, ώστε να προκύψει ένα ιστόγραμμα N -ιμών. Όταν ο CSD έχει μήκος $N \in \{128, 64, 32\}$, τότε το ιστόγραμμα των 256 τιμών πρέπει να μειωθεί σε μήκος με ενοποίηση τιμών. Η διαδικασία, για τη γενική περίπτωση της μείωσης από M τιμές σε $N < M$, περιγράφεται ως ακολούθως:

Στη περίπτωση που $M=256$ και $N \in \{128, 64, 32\}$, έστω ότι $P = \{p_1, p_2, \dots, p_m\}$ και $Q = \{q_1, q_2, \dots, q_N\}$, δύο κβαντίσεις του χρωματικού χώρου S , όπου p_m και q_n είναι τα κελιά των κβαντισμών και $M > N$. Ο κβαντισμός υπάγεται στις εξής συνθήκες.

$$\bigcup_{m=1}^M p_m = S = \bigcup_{n=1}^N q_n \quad (2.2.46)$$

$$\text{για κάθε } n, q_n = \bigcup_{m \in J_n} p_m, \text{ όπου } J_n = \{n_1, n_2, \dots, n_{k_n}\} \quad (2.2.47)$$

και $J_i \cap J_j = \emptyset$ for $i \neq j$

Η πρώτη συνθήκη εξασφαλίζει ότι και το P και το Q καλύπτουν τον χώρο S . Η δεύτερη συνθήκη έμμεσα ορίζει τον αύξοντα αριθμό του J_n . Η τρίτη συνθήκη είναι συνέπεια του γεγονότος τα κελιά που προκύπτουν από τη μείωση των τιμών είναι από μόνα τους ανεξάρτητα. Τέλος, η ενοποίηση των κελιών ορίζεται ως:

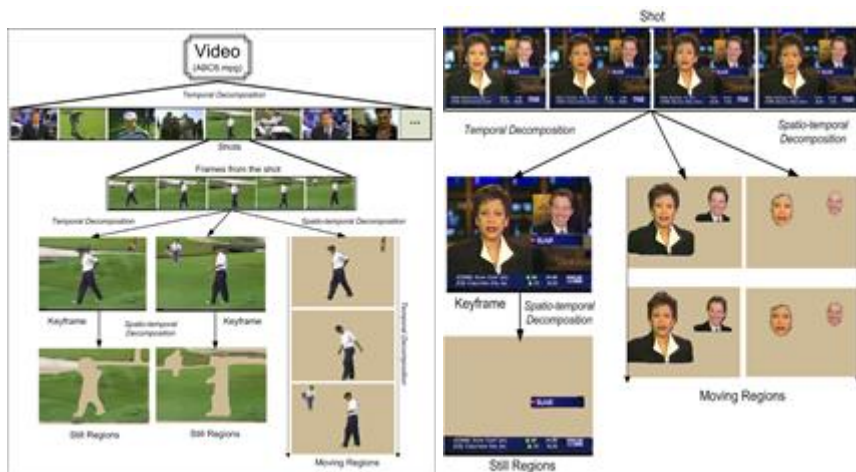
$$h_S^N(n) = \sum_{m \in J_n} h_S^M(m), \quad n \in \{1, 2, \dots, N\} \quad (2.2.47)$$

- Στο τρίτο και τελευταίο βήμα, κάθε μία από τις N τιμές πρέπει να κανονικοποιηθεί στο διάστημα των τιμών $[0, 1]$, και έπειτα εφαρμόζεται μη-γραμμική μείωση των κανονικοποιημένων τιμών σε 8 τιμές, σύμφωνα με τις στατιστικές των περιπτώσεων από τις εικόνες. Η μη-γραμμική μείωση των τιμών, προέκυψε από ευρεστικά πειράματα και αυξάνει το ποσοστό επιτυχίας της ανάκτησης εικόνων του CSD.

Για τη σύγκριση και το ταίριασμα των όμοιων διανυσμάτων, εφαρμόζεται ότι και σε άλλους περιγραφείς που έχουν τη μορφή ιστογράμματος. Ο CSD χρησιμοποιεί την l_1 -norm.

2.2.9.7 Group of Frames/Group of Pictures Color Descriptor

Οι περιγραφείς GoF/GoP παράγονται από τη συγκέντρωση ιστογραμμάτων από πολλές εικόνες ή από καρέ βίντεο. Τα ιστογράμματα αυτά συναθροίζονται στον περιγραφέα SCD. Κάθε ιστόγραμμα υπολογίζεται ξεχωριστά και είναι βασισμένο στη μείωση χρωμάτων στο χρωματικό χώρο HSV. Τρεις διαφορετικοί τρόποι υπάρχουν για τη συνάθροιση χρωματικών ιστογραμμάτων: της μέσης τιμής, του μέσου όρου, και της τομής. Στο παραγόμενο ιστόγραμμα εφαρμόζεται ένας μετασχηματισμός τύπου Haar, ώστε να μετατραπεί στον περιγραφέα SCD. Το ιστόγραμμα μέσου όρου (average) υπολογίζεται συνυπολογίζοντας τα ιστογράμματα της ομάδας (group) και κανονικοποιώντας κάθε τιμή του ιστογράμματος με το συνολικό πλήθος των τιμών N . το μειονέκτημα του ιστογράμματος μέσου όρου είναι η επιρροή που ασκείται από τις ακραίες τιμές. Το ιστόγραμμα ενδιάμεσης τιμής (median), εξαλείφει προβλήματα τυχαίων μέγιστων τιμών, επικαλύψεις κειμένου κ.α.. Το μειονέκτημα του ιστογράμματος ενδιάμεσης τιμής, είναι η υπολογιστική του πολυπλοκότητα. Τέλος, το ιστόγραμμα τομής (intersection) παράγεται από τον υπολογισμό της ελάχιστης τιμής για κάθε μία από τις N τιμές του γκρουπ των ιστογραμμάτων. Το ιστόγραμμα που προκύπτει αναδεικνύει περισσότερο τα λιγότερο όμοια χρωματικά μοτίβα παρά τη χρωματική κατανομή στις εικόνες. Η σύγκριση των διανυσμάτων περιγραφής του GoF/GoP, όπως είναι αναμενόμενο είναι όμοιο με τη διαδικασία της σύγκρισης του SCD περιγραφέα. Στην Εικόνα 20 δίνεται παράδειγμα της διαδικασίας GoF/GoP.



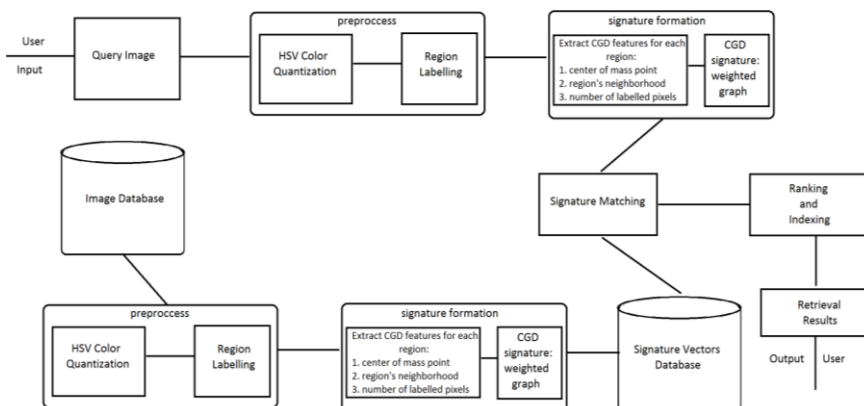
Εικόνα 20. Παραδείγματα της διαδικασίας GoF/GoP

Σύστημα ανάκτησης βάσει χρώματος

3.1 Άποψη συστήματος

Η ανάλυση ψηφιακών εικόνων με σκοπό την εξαγωγή χαρακτηριστικών, για τη περιγραφή του οπτικού τους περιεχομένου απασχόλησε πολλούς επιστήμονες τα προηγούμενα έτη. Καθένας τους, ακολούθησε διαφορετική προσέγγιση, για να επιτευχθεί ο παραπάνω στόχος. Στη συγκεκριμένη εργασία, η πηγή έμπνευσης ήταν τα λόγια ενός ψυχολόγου και συνιδρυτή της σχολής Gestalt, του Max Wertheimer [58]: *«στέκομαι στο παράθυρο και βλέπω ένα σπίτι, δέντρα, ουρανό. Θεωρητικά μπορώ να πω υπάρχουν 327 φωτεινότητες και αποχρώσεις. Έχω "327"; Όχι. Έχω ουρανό, σπίτι, και δέντρα»*. Η διαπίστωση του Wertheimer, οδηγεί στο συμπέρασμα πως εάν ομαδοποιηθεί η πληθώρα των αποχρώσεων στο βασικό τους χρώμα, το νόημα του θέματος που απεικονίζεται διατηρείται.

Ο περιγραφέας έγχρωμων στατικών εικόνων (Color Gravity Descriptor – CGD), ο οποίος αναπτύχθηκε σε αυτή την εργασία, βασίζεται στην υπόθεση πως, εάν μειωθεί η χρωματική παλέτα σε κάποια βασικά χρώματα και έπειτα από το μετασχηματισμό εξαχθούν κατάλληλα χαρακτηριστικά σχετιζόμενα με την τοποθέτηση των εν λόγω χρωμάτων στο χωρικό επίπεδο της εικόνας, μπορεί να επιτευχθεί ταύτιση όμοιων εικόνων. Στο πλαίσιο παρουσίασης του εν λόγω περιγραφέα, στις παραγράφους που ακολουθούν περιγράφεται πλήρως η μεθοδολογία που αναπτύχθηκε. Στη μεθοδολογία περιλαμβάνεται ο αλγόριθμος μείωσης χρωμάτων, επιλογής, μετατροπής και εξαγωγής χαρακτηριστικών και διαμόρφωσης της υπογραφής της εικόνας. Αναλύεται η επιλογή του μέτρου ομοιότητας μεταξύ υπογραφών εικόνων και παρουσιάζονται τα αποτελέσματα που προκύπτουν από τον πειραματισμό σε μία πρότυπη Βάση Δεδομένων εικόνων. Τέλος γίνεται σύγκριση με τα τρέχοντα πρότυπα και παρουσιάζονται τα συμπεράσματα που προκύπτουν. Στην Εικόνα 21 παρουσιάζεται ένα απλοποιημένο διάγραμμα ροής των διεργασιών που εμπλέκονται στη μέθοδο που αναπτύχθηκε.



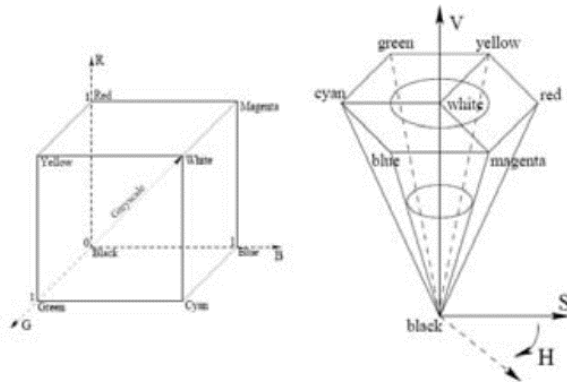
Εικόνα 21. Διάγραμμα ροής της μεθόδου που αναπτύχθηκε

3.2 Προεπεξεργασία

Συνήθως, μία εικόνα αποδίδεται στο μοντέλο RGB. Το μοντέλο αυτό, έχει τρία κανάλια, τα οποία περιλαμβάνουν την ένταση του φωτός στα τρία βασικά κανάλια χρώματος. Κάθε κανάλι μπορεί να πάρει τιμές από 0 έως 1. Η έννοια της φωτεινότητας, της χρωματικής καθαρότητας, καθώς και η αναπαράσταση κάποιου χρώματος, είναι εξαρτημένες από το συνδυασμό των τιμών των τριών καναλιών. Επομένως, το μοντέλο αυτό, δεν είναι ιδανικό, για τη διαδικασία της ομαδοποίησης χρωμάτων και αναπαράστασης με άλλα, αντιπροσωπευτικά χρώματα. Ένα ικανοποιητικό μοντέλο για αυτή τη διεργασία, είναι κάποιο στο οποίο το χρώμα περιγράφεται ξεχωριστά και αποκλειστικά από μία μεταβλητή, απλοποιώντας τη διαδικασία της ομαδοποίησης. Τέτοια μοντέλα είναι η οικογένεια των HSV, HSL και HSI. Σε αυτά, το χρώμα περιγράφεται από τη μεταβλητή H, η οποία εκφράζεται σε μοίρες γωνίας, όπως φαίνεται στην Εικόνα 22 για την περίπτωση του HSV. Επιλέγοντας το μοντέλο HSV, μεταβλητές H, S και V εκφράζουν το χρώμα, τη χρωματική καθαρότητα και τη φωτεινότητα αντίστοιχα. Αυτά τα οποία γίνονται άμεσα κατανοητά από το σχήμα του ανάποδου κώνου που αναπαριστά τη δομή του χρωματικού χώρου (Εικόνα 23) είναι τα εξής:

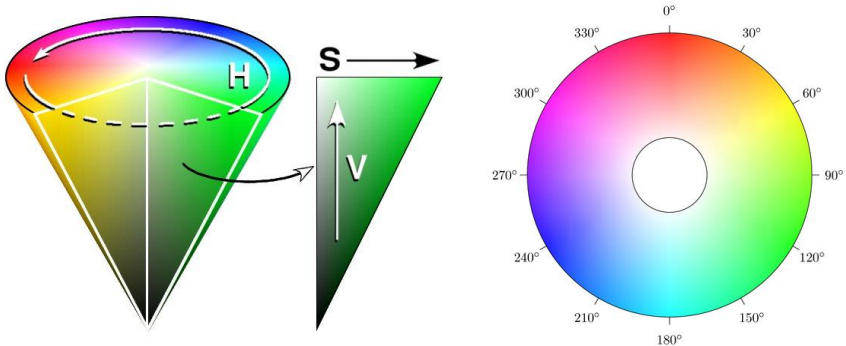
- Μια οριζόντια τομή του κώνου, αποτελεί έναν κυκλικό δίσκο, ο οποίος έχει κέντρο στο μηδέν $S=0$ και ακτίνα 1, ανεξάρτητα από τη τιμή του V, εφόσον οι τιμές είναι κανονικοποιημένες. Στη περιφέρεια του κύκλου, ορίζονται τα αμιγή χρώματα
- Το χρώμα είναι ανεξάρτητο της έντασης της φωτεινότητας
- Ανά 60 μοίρες υπάρχει και ένα από τα 3 βασικά χρώματα και 3 από τα συμπληρωματικά τους
- Όσο πιο χαμηλά είναι στον κώνο ένα χρώμα ($V \rightarrow 0$) τόσο πιο σκοτεινό εμφανίζεται

- Όσο πιο ψηλά είναι στον κώνο ένα χρώμα ($V \rightarrow 1$) τόσο πιο φωτεινό εμφανίζεται
- Όσο πιο κοντά στα όρια του κώνου είναι ένα χρώμα ($S \rightarrow 1$) τόσο πιο αμιγές χρώμα εμφανίζεται
- Όσο πιο κοντά στον κατακόρυφο άξονα V είναι ένα χρώμα ($S \rightarrow 0$) τόσο πιο συγκεχυμένο ή γκριζό εμφανίζεται



Εικόνα 22 Οι χώροι των μοντέλων RGB και HSV

Όλα τα παραπάνω φαίνονται και σε μία οριζόντια διατομή του κώνου στην Εικόνα 23.



Εικόνα 23 Τομή κώνου HSV και άποψη κορυφής

Το ανθρώπινο μάτι, εξαιτίας της μη-γραμμικής του απόκρισης στη φωτεινότητα (βλπ. gamma correction¹) και στην αντίθεση (βλπ. contrast sensitivity function²), μετά από κάποιο σημείο αντιλαμβάνεται κάθε σκοτεινό

¹ Περιγραφή της μη-γραμμικής απόκρισης της ανθρώπινης όρασης στη φωτεινότητα και των απαιτούμενων προσαρμογών των συστημάτων προβολής, στη διεύθυνση <http://www.cambridgeincolour.com/tutorials/gamma-correction.htm>

² Περιγραφή και επίδειξη της contrast sensitivity function στη διεύθυνση <http://sunburst.usd.edu/~schieber/coglab/CSFIntro.html>

χρώμα ως μαύρο ενώ κάθε πολύ φωτεινό γκριζο ως άσπρο, όπως και κάθε μη-αμιγές (συγκεχυμένο) χρώμα ως γκριζο. Αυτό γίνεται φανερό στην Εικόνα 23, στην οποία φαίνεται πως όταν το V τείνει στο μηδέν, η ανθρώπινη αντίληψη κατατάσσει τα χρώματα στο μαύρο, ενώ όσο το S τείνει στο 0 τόσο πιο γκριζο φαίνεται το χρώμα. Αντίθετα, όσο το V τείνει στο 1, τόσο πιο φωτεινό γίνεται το χρώμα και μπορεί να γίνει αντιληπτό ευκολότερα. Επίσης, όσο το S τείνει στο 1, το χρώμα γίνεται καθαρότερο. Στην ειδική περίπτωση που το S τείνει στο 0 και το V τείνει στο 1, απεικονίζεται το λευκό.

Λαμβάνοντας υπόψη τις παρατηρήσεις αυτές προκύπτει ο παρακάτω αλγόριθμος μείωσης χρωμάτων:

1. Μετατροπή της εικόνας στο χρωματικό χώρο HSV
2. Τμηματοποίηση του HSV χώρου σε N_H τμήματα χρώματος και N_G τμήματα γκριζων τόνων (συμπεριλαμβανομένων λευκού και μαύρου)
3. Ταξινόμηση κάθε εικονοστοιχείου σε κάποια από τις N_H χρωματικές κλάσεις μέσω βαθμωτού κβαντισμού χρωμάτων
4. Ταξινόμηση των εικονοστοιχείων στις N_G κλάσεις γκριζων τόνων μέσω εφαρμογής ευρεστικών κατωφλίων
5. Δημιουργία ευρετηρίου (index) για κάθε κλάση (labeling)

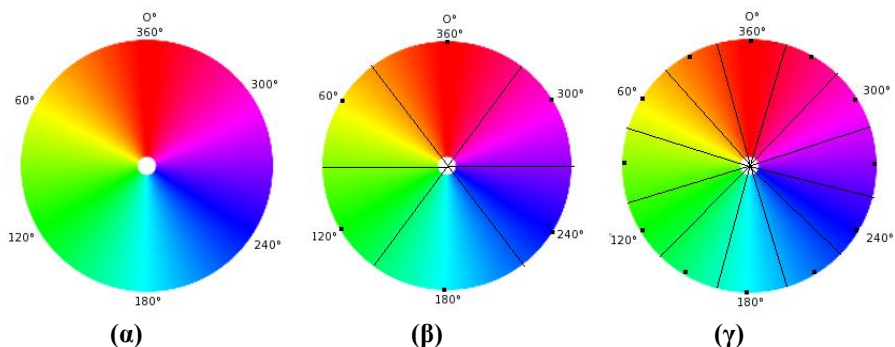
Έστω N_H το πλήθος των περιοχών χρώματος και N_G το πλήθος των περιοχών τόνων του γκρι στο HSV στις οποίες θέλουμε να αναπαραστήσουμε την αρχική εικόνα. Ισχύει προφανώς $N = N_H + N_G$, με N το συνολικό πλήθος των χρωμάτων στο HSV. Ορίζονται τα διανύσματα:

$$h = \{ x_{K_1} : 0^\circ \leq x_{K_1} \leq 360^\circ, \quad x_{K_1+1} = x_{K_1} + \frac{360}{N_H}, x_0 = 0 \} \quad (3.1.1)$$

$$g = \{ y_{K_2} : 0 \leq y_{K_2} \leq 1, y_{K_2+1} = y_{K_2} + \frac{1}{N_G - 1}, y_0 = 0 \} \quad (3.1.2)$$

$$\mu \varepsilon K_1 = \{1, 2, \dots, N_H - 1\} \text{ και } K_2 = \{1, 2, \dots, N_G - 1\}$$

Οι τιμές των N_H και N_G καθορίζουν το αποτέλεσμα της μείωσης των χρωμάτων. Το εύρος των χώρων (bins) που δημιουργούνται με την τμηματοποίηση του χώρου φαίνεται στην Εικόνα 24, στην οποία απεικονίζεται ένας κυκλικός δίσκος από μία οριζόντια τομή του χρωματικού χώρου (α), ενώ στο (β) και (γ) η τμηματοποίηση που προκύπτει για $N_H = 6$ και για $N_H = 12$ αντίστοιχα.



Εικόνα 24 Χρωματική τμηματοποίηση κυκλικού δίσκου του χρωματικού χώρου HSV

Αρχικά, ορίζεται το πλήθος των χρωμάτων και των αποχρώσεων του γκρι, έστω $N_H = 6$ και $N_G = 3$, αντίστοιχα. Έτσι, προκύπτουν τα διανύσματα: $h = \{0, 60, 120, 180, 240, 300\}$ και $g = \{0, 0.5, 1\}$. Για απλότητα στον αλγόριθμο, παράλληλα με αυτά τα διανύσματα παράγονται και δύο ακόμη διανύσματα με απλούς δείκτες, με ένα-προς-ένα αντιστοιχία με τα h και g : $h' = \{1, 2, 3, 4, 5, 6\}$ και $g' = \{1, 2, 3\}$. Οι τιμές αυτές θα χρησιμοποιηθούν στη συνέχεια για το labeling των εικονοστοιχείων της εικόνας. Οι γωνίες που περιλαμβάνει το διάνυσμα h απεικονίζονται στην Εικόνα 24(β). Στη συνέχεια με βαθμωτό κβαντισμό του χρώματος (Hue) κάθε εικονοστοιχείου, κάθε εικονοστοιχείο μετατρέπεται στον αντίστοιχο δείκτη του h' . Έτσι, για κάθε δείκτη μπορούν να εξαχθούν δυαδικές εικόνες-μάσκες που αναπαριστούν τα εικονοστοιχεία που αναπαρίστανται από το συγκεκριμένο δείκτη. Ο ομοιόμορφος βαθμωτός κβαντισμός των χρωμάτων γίνεται με τον απλούστερο κβαντιστή, τη λειτουργία round, όπως περιγράφει η (3.1.3).

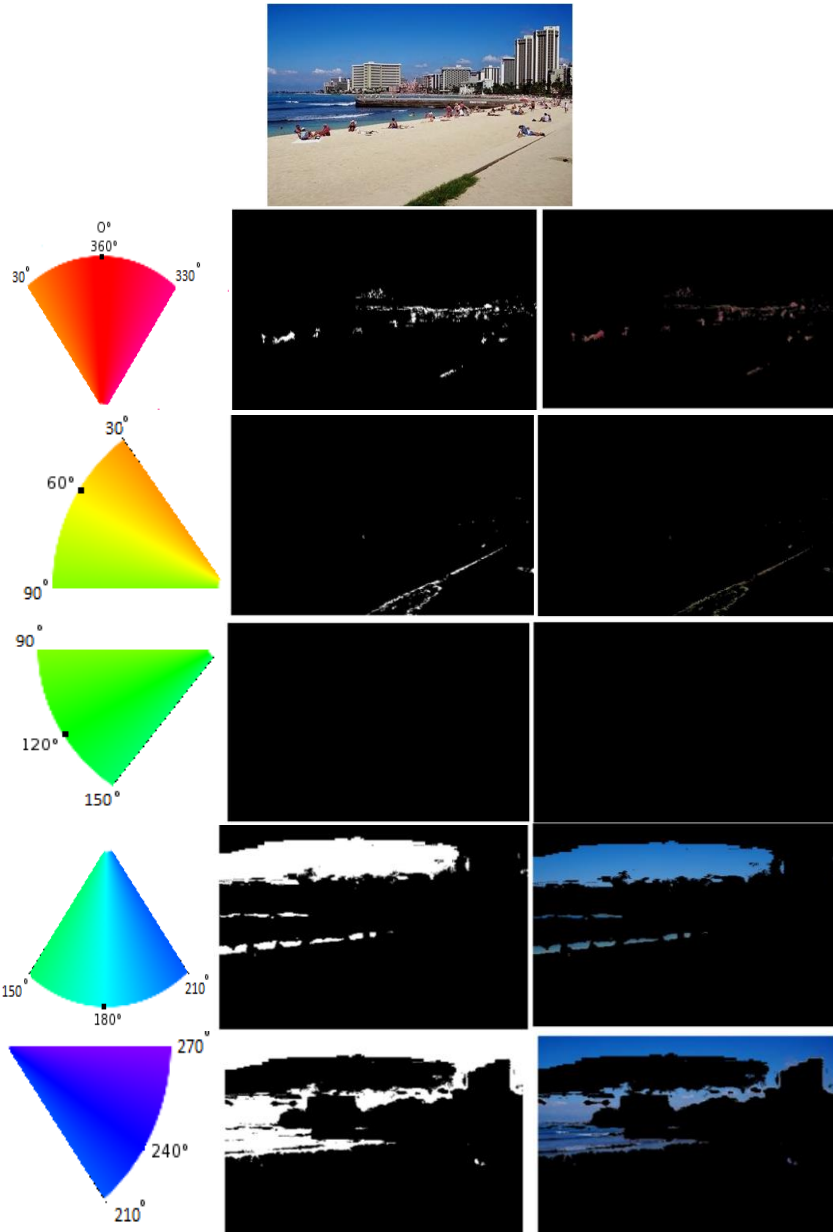
$$Q: I_{x,y} \rightarrow J_{x,y}: J_{x,y} = \lfloor I_{x,y} \rfloor_{x,y \text{ σ υ ν τ ε τ α γ μ έ ν ε ς ε ι κ ό ν α ς}} \quad (3.1.3)$$

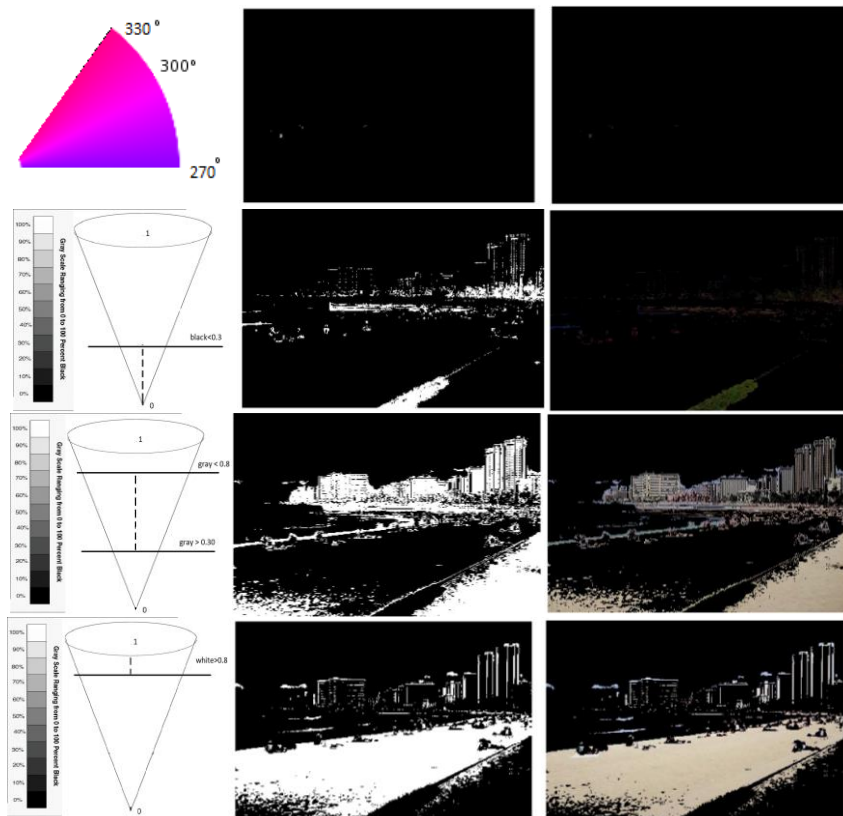
όπου I η εικόνα εισόδου και J η κβαντισμένη εικόνα.

Με αυτό τον απλό κβαντιστή ουσιαστικά πραγματοποιείται η τμηματοποίηση του χώρου χρώματος που περιγράφεται στην Εικόνα 24 με ταυτόχρονη ταξινόμηση των τιμών χρώματος των εικονοστοιχείων στην τιμή του μέσου του τόξου στον κυκλικό τομέα στον οποίο εντοπίζεται..

Ένα παράδειγμα του αποτελέσματος της διαδικασίας φαίνεται στην Εικόνα 25 η οποία έχει υποστεί μείωση χρωμάτων σε 6 βασικά χρώματα και τρεις περιοχές γκρίζων τόνων που αντιστοιχούν μαύρο σε τιμές φωτεινότητας μικρότερες του 0.3, γκρίζο σε τιμές μεταξύ 0.3 και 0.8 και λευκό σε τιμές μεγαλύτερες του 0.8.. Στα αποτελέσματα παρουσιάζονται η περιοχή χρώματος, η μάσκα χρώματος και η εικόνα που αντιστοιχεί, έτσι για παράδειγμα στην ομάδα με τιμή 1, φαίνεται το αντιπροσωπευτικό χρώμα, το

οποίο είναι στην περιοχή $[0^\circ, 30^\circ) \cup [330^\circ, 360^\circ]$. Τέλος συμπεριλαμβάνονται οι τρεις γκρίζες περιοχές που εντοπίζονται.





Εικόνα 25 Κατάτμηση εικόνας, με χρήση μασκών

Εάν αυξηθεί το πλήθος των χρωμάτων N στο διάνυσμα h , τότε η εικόνα με τη μειωμένη παλέτα χρωμάτων, χρωματικά φαίνεται περισσότερο όμοια με την αρχική. Η τιμή του πλήθους στη συγκεκριμένη εργασία είναι $N_H = 8$ και προέκυψε μετά από σειρά πειραμάτων. Στην Εικόνα 26, παρουσιάζεται μία εικόνα με διαφορετικές εκδοχές πλήθους τιμών χρώματος.

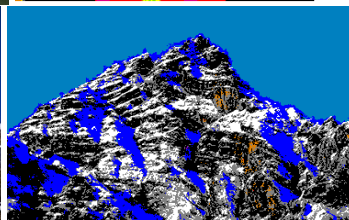
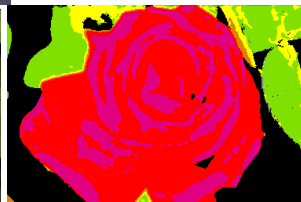


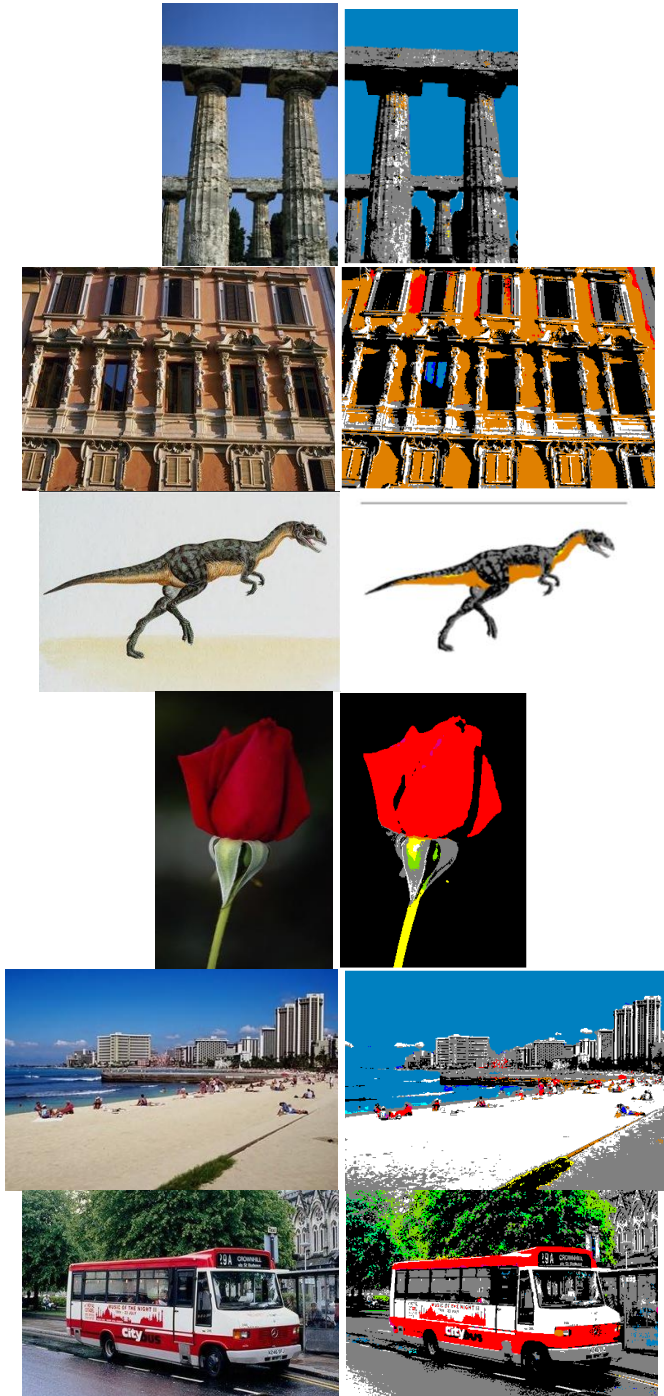
Εικόνα 26 Αρχική και εικόνες μειωμένης χρωματικότητας για $N_H = \{6, 12, 24\}$ και $N_G = \{3\}$

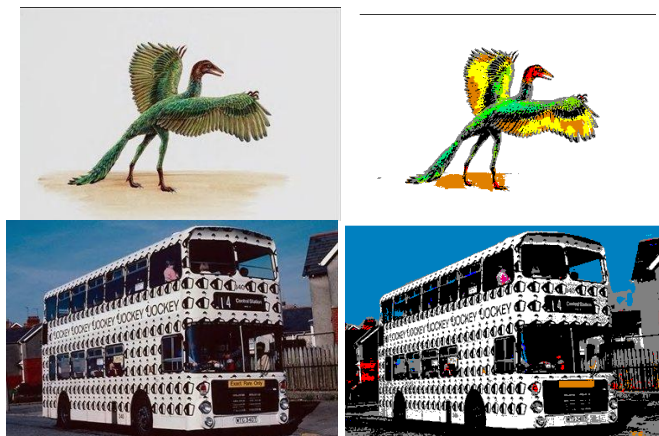
Ο κβαντισμός χρώματος των εικόνων που πραγματοποιείται, έχει ως αποτέλεσμα χαρακτηριστικά όπως η υφή και η φωτεινότητα να μην παίζουν ρόλο κατά την εξαγωγή των χαρακτηριστικών. Τονίζεται στο σημείο αυτό ότι τα χρώματα, τα οποία χρησιμοποιήθηκαν για την αναπαράσταση, δεν είναι τα πλέον αντιπροσωπευτικά, απλά αποτελούν μια βάση για τον κβαντισμό όλων των εικόνων στην ίδια χρωματική παλέτα.

Στην Εικόνα 27 παρουσιάζονται δείγματα αποτελεσμάτων της μεθόδου μείωσης χρωμάτων που εφαρμόζεται (με τη μορφή ζευγαριών αρχικής εικόνας και εικόνας μετά τη μείωση χρωμάτων) και η οποία δημιουργήθηκε αποκλειστικά για τους σκοπούς της παρούσας εργασίας. Το πλήθος των χρωμάτων είναι $N_H = 12$, και το πλήθος των τόνων του γκρι είναι $N_G = 3$.









Εικόνα 27 Αποτελέσματα μείωσης χρωμάτων σε πραγματικές εικόνες

3.3 Εξαγωγή χαρακτηριστικών

Το κυριότερο τμήμα σε ένα σύστημα ανάκτησης εικόνων βάσει περιεχομένου, είναι αυτό που εξάγει τα χαρακτηριστικά τα οποία θα επιλεχθούν. Τα χαρακτηριστικά αυτά σχηματίζουν το διάνυσμα περιγραφής του περιεχομένου της εικόνας. Το διάνυσμα περιγραφής, αφού σχηματιστεί, χαρακτηρίζεται ως υπογραφή της εικόνας, και με το κατάλληλο μέτρο ομοιότητας χρησιμοποιείται για σύγκριση μεταξύ εικόνων. Υπάρχουν κάποιες ιδιότητες που χαρακτηρίζουν ένα καλό χαρακτηριστικό όπως, η αξιοπιστία, η διακριτικότητα, η προσαρμογή με τη διαδικασία ταξινόμησης, και η ανεξαρτησία από τη περιστροφή ή την κλιμάκωση.

Μεγάλο μέρος της πληροφορίας που έχει μία έγχρωμη εικόνα, είναι το ίδιο της το χρώμα. Έχουν παρουσιαστεί προσεγγίσεις, οι οποίες αξιοποιούν την πληροφορία αυτή, όπως για παράδειγμα η σύγκριση ιστογραμμάτων χρώματος, και ο αλγόριθμος CLD. Οι αλγόριθμοι αυτοί, ενώ επιτυγχάνουν ικανοποιητικά αποτελέσματα, παρουσιάζουν κάποια μειονεκτήματα. Η σύγκριση ιστογραμμάτων δεν αξιοποιεί χωρική πληροφορία. Ο CLD, ενώ χρησιμοποιεί χωρική πληροφορία, δεν είναι ανεξάρτητος περιστροφής. Επομένως, τα χαρακτηριστικά τα οποία επιλέχθηκαν στην παρούσα εργασία, έχουν στόχο να εκμεταλλευτούν τη χρωματική πληροφορία, να εκφράσουν χωρικές συσχετίσεις και να επιλύσουν προβλήματα ευρωστίας στην περιστροφή και κλιμάκωση.

Κεντρικό πρόβλημα στην παρούσα εργασία ήταν να εξεταστεί η υπόθεση πως, εάν μια κλάση αντιπροσωπευτικού χρώματος χαρακτηρίζεται από την επιφάνεια που καλύπτει στην εικόνα, καθώς και από το σημείο του κέντρου μάζας και τη διασπορά της, τότε με την κατάλληλη συσχέτιση αυτών, μπορεί να επιτευχθεί σύγκριση και ανάκτηση όμοιων εικόνων.

Το πλήθος των εικονοστοιχείων ενός χρώματος δίνεται από τον τύπο (3.1.4) και αποτελούν μια ποσοτική χρωματική πληροφορία, η οποία υπο-

νοεί την ποσοστιαία συμμετοχή του εν λόγω χρώματος στην εικόνα. Τα κέντρα μάζας, τα οποία δίνονται από τον τύπο (3.1.5), είναι στην ουσία η μέση θέση του εν λόγω χρώματος στην εικόνα. Η θέση ενός χρώματος στην εικόνα, αποτελεί χωρικό δεδομένο. Πέρα από τα δύο αυτά μεγέθη απαιτείται και η πληροφορία της χωρικής κατανομής του χρώματος μέσα στην εικόνα. Για αυτό το λόγο ο τύπος (3.1.6) υπολογίζει την επιφάνεια που καλύπτει ένα χρώμα δημιουργώντας μια ορθογώνια περιοχή που περιλαμβάνει ακριβώς όλα τα εικονοστοιχεία του εν λόγω χρώματος.

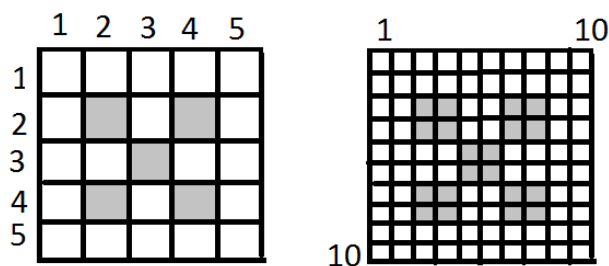
$$\text{Επιφάνεια χρώματος} \quad c = \sum_{\forall I(x,y)=1} I(x,y) \quad (3.1.4)$$

$$\text{Κέντρο μάζας} \quad p(\bar{x}, \bar{y}): \bar{x} = \frac{\sum_{\forall I(x,y)=1} x}{c}, \bar{y} = \frac{\sum_{\forall I(x,y)=1} y}{c} \quad (3.1.5)$$

$$\text{Παράθυρο επικάλυψης} \quad b = (\max(x) - \min(x) + 1)(\max(y) - \min(y) + 1) \quad (3.1.6)$$

όπου I η εικόνα εισόδου, x, y οι συντεταγμένες και N, M οι αντίστοιχες διαστάσεις της εικόνας.

Η εξαγωγή αυτών των χαρακτηριστικών είναι το πρώτο βήμα για το σχηματισμό της υπογραφής, του περιγραφέα. Τα δεδομένα που εξάγονται από την εικόνα, για να χαρακτηριστούν «ικανοποιητικά», πρέπει να υπακούουν σε κάποια κριτήρια, δύο από αυτά είναι η ανεξαρτησία ως προς την κλιμάκωση και την περιστροφή. Το πλήθος των εικονοστοιχείων που περιλαμβάνει ένα χρώμα, είναι εξαρτώμενο από τις διαστάσεις μιας εικόνας. Στο παράδειγμα που ακολουθεί στην Εικόνα 28, φαίνεται το ίδιο μοτίβο σε δύο εικόνες με διαφορετική ανάλυση (κλιμάκωση).



Εικόνα 28. Παράδειγμα δύο εικόνων διαφορετικής ανάλυσης που παρουσιάζουν το ίδιο μοτίβο

Για τις εικόνες στην Εικόνα 28 προκύπτουν τα εξής μεγέθη:

Διαστάσεις	Επιφάνεια (c)	Κέντρο μάζας (p)	Εμβαδόν επικάλυψης (b)
5x5	5	(3, 3)	9

10x10	20	(5.5, 5.5)	36
-------	----	------------	----

Για να αντιμετωπιστεί το πρόβλημα της διαφορετικής κλίμακας κανονικοποιούνται οι τιμές c , b , και το κέντρο βάρους p . Οι εξισώσεις των κανονικοποιήσεων φαίνονται παρακάτω:

$$\text{Επιφάνεια χρώματος} \quad C_{norm} = \frac{c}{N \cdot M} \quad (3.1.7)$$

$$\text{Κέντρο μάζας} \quad p(\bar{x}_{norm}, \bar{y}_{norm}): \bar{x}_{norm} = \left\lfloor 10 \cdot \frac{\bar{x}}{N} \right\rfloor / 10, \quad (3.1.8)$$

$$\bar{y}_{norm} = \left\lfloor 10 \cdot \frac{\bar{y}}{M} \right\rfloor / 10$$

$$\text{Παράθυρο επικάλυψης} \quad b_{norm} = \left(\max\left(\frac{x}{N}\right) - \min\left(\frac{x}{N}\right) + 1 \right) \cdot \left(\max\left(\frac{y}{M}\right) - \min\left(\frac{y}{M}\right) + 1 \right) \quad (3.1.9)$$

Έτσι, πλέον οι τιμές για τη παραπάνω εικόνα είναι ίσες, όπως φαίνεται στον παρακάτω πίνακα.

Διαστάσεις	Επιφάνεια (c)	Κέντρο μάζας (p)	Εμβαδόν επικάλυψης (b)
5x5	5/25=0.2	(0.6, 0.6)	0.36
10x10	20/100=0.2	(0.6, 0.6)	0.36

Ένα άλλο πρόβλημα ανακύπτει καθώς το κέντρο μάζας είναι ευαίσθητο ως προς την περιστροφή της εικόνας. Αντίθετα, η απόσταση μεταξύ δύο σημείων, είναι ανεξάρτητη περιστροφής. Επομένως για τη διατήρηση της ιδιότητας ανεξαρτησίας από περιστροφή, η υπογραφή της εικόνας, δεν είναι δυνατόν να περιέχει τα σημεία των κέντρων μάζας, αλλά θα μπορούσε να περιλαμβάνει τις αποστάσεις μεταξύ των εν λόγω σημείων. Οι αποστάσεις υπολογίζονται με μέτρο την Ευκλείδεια απόσταση. Από τις αποστάσεις προκύπτει ένας τετραγωνικός πίνακας D τάξης N , όπου $N = N_H + N_G$.

Τα άλλα δύο χαρακτηριστικά, το πλήθος των εικονοστοιχείων κάθε χρώματος και η ποσοστιαία επιφάνεια κάλυψης της εικόνας, πρέπει να συνδυαστούν τόσο μεταξύ τους όσο και με τις αποστάσεις, ώστε να γίνει συμπαγής η υπογραφή. Ο συνδυασμός έχει το σκεπτικό της δημιουργίας ενός συντελεστή βάρους w , ο οποίος θα εφαρμόζεται ανά χρώμα στις αποστάσεις των κέντρων βάρους. Το w εκφράζει τη συνεκτικότητα της ομάδας ενός χρώματος και υπολογίζεται αρχικά ως $w = \sqrt{C_{norm} \cdot b_{norm}}$. Ο συντελεστής συνεκτικότητας w εννοιολογικά είναι ανεξάρτητος περιστροφής, αλλά κανονικοποιείται στη συνέχεια ως προς το πλήθος των εικονοστοιχείων της επικαλυπτόμενης επιφάνειας που δεν ανήκουν την ομάδα, ώστε να γίνει ανεξάρτητος κλίμακας. Τέλος, για να αποδοθεί ομοιόμορφη βαρύτητα κατά το συνδυασμό του βάρους w με τις αποστάσεις

D , κανονικοποιείται και ως προς το συνολικό άθροισμα όλων των βαρών. Με αυτό τον τρόπο προκύπτει ο παρακάτω τύπος υπολογισμού:

$$w_i = \frac{\sqrt{c_{norm_i} * b_{norm_i}}}{\sqrt{b_{norm_i} - c_{norm_i}}}, \text{ όπου } i = \{1, 2, \dots, N\}, N = N_H + N_G \quad (3.1.10)$$

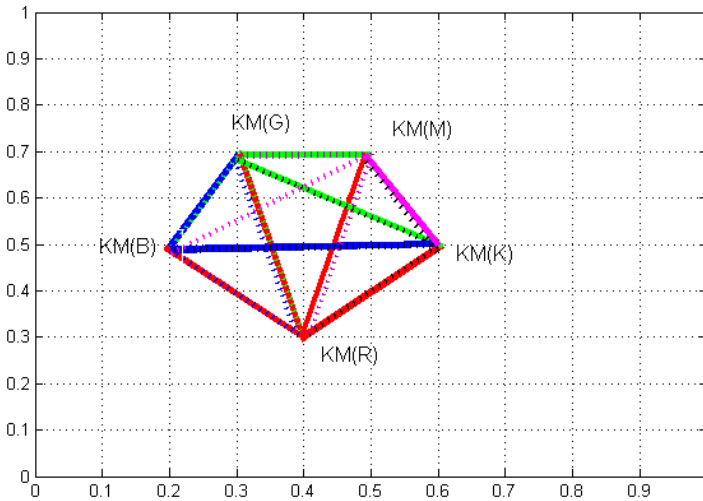
$$W_i = \frac{w_i}{\sum w}, \text{ όπου } i = \{1, 2, \dots, N\}, N = N_H + N_G \quad (3.1.11)$$

Τέλος, ο συνδυασμός του συντελεστή συνεκτικότητας και των αποστάσεων των κέντρων μάζας αποτελεί και το διάνυσμα χαρακτηριστικών ή αλλιώς την υπογραφή της εικόνας. Η υπογραφή s είναι ένα διάνυσμα με μήκος N^2 και σχηματίζεται ως εξής:

$$s = \bigcup_{i=1}^N \left(\bigcup_{j=1}^N (W_i \cdot D_j) \right) \quad (3.1.11)$$

όπου $i = \{1, 2, \dots, N\}, j = \{1, 2, \dots, N\}, N = N_H + N_G$

Ο σχηματισμός των αποστάσεων μεταξύ των κέντρων μάζας δημιουργεί ένα γράφο. Ο γράφος αυτός είναι ανεξάρτητος περιστροφής, αφού οι ακμές του έχουν τιμή την ποσοστιαία απόσταση μεταξύ των κόμβων (κέντρα μάζας), η οποία παραμένει αναλλοίωτη σε περίπτωση αλλαγής κλίμακας. Τέλος, ο γράφος είναι ανεξάρτητος περιστροφής, αφού όλα τα χαρακτηριστικά του, ακμές, γωνίες μεταξύ ακμών και οι κόμβοι παραμένουν επίσης σε αναλογία. Τα παραπάνω φαίνονται στο διάγραμμα που ακολουθεί στην Εικόνα 29.



Εικόνα 29 Γράφος υπογραφής

Το σύστημα που περιγράφηκε βασίζεται κατά κάποιο τρόπο στις μάζες των χρωμάτων και στη θέση που καταλαμβάνουν αυτές στο επίπεδο των εικόνων. Για το λόγο αυτό η μέθοδος και ο αντίστοιχος περιγραφέας έλαβε την ονομασία **Color Gravity Descriptor (CGD)** και όλα τα πειραματικά αποτελέσματα που παρουσιάζονται για την αξιολόγηση της αποδοτικότητας χρησιμοποιούν αυτή τη διακριτική ονομασία.

3.5 Μέτρο ομοιότητας

Ως μέτρο για την ομοιότητα των υπογραφών, χρησιμοποιείται το άθροισμα των απόλυτων διαφορών ή αλλιώς City Block (2.2.25). Μετά από πειραματισμούς με διάφορα μέτρα επιλέχθηκε το συγκεκριμένο λόγω της απλότητας και ταχύτητας που προσφέρει σε μια υλοποίηση λογισμικού και φυσικά εξαιτίας του ικανοποιητικού αποτελέσματος που παρέχει.

3.6 Αξιολόγηση συστήματος

Με βάση τη θεωρητική ανάλυση και προσέγγιση στην ανάπτυξη του CGD πραγματοποιήθηκε και αντίστοιχη υλοποίηση με χρήση του λογισμικού MATLAB. Το σύστημα υλοποιήθηκε εξ ολοκλήρου στο εν λόγω περιβάλλον και η αποδοτικότητά του δοκιμάστηκε με τυπικές βάσεις δεδομένων (ground truth data) για εφαρμογές CBIR. Πολύ σημαντική είναι η αξιολόγηση του συστήματος γιατί ουσιαστικά με αυτό τον τρόπο φαίνεται η αποδοτικότητα και η ποιότητα του. Ως μέτρα αξιολόγησης, όπως και σχεδόν σε κάθε αντίστοιχη περίπτωση, επιλέχθηκαν τα μέτρα αποδοτικότητας ανάκτησης precision/recall.

3.6.1 Πειραματική διαδικασία

Για να γίνει ο έλεγχος της αποδοτικότητας του συστήματος απαιτήθηκαν τρία συστατικά στοιχεία:

1. Βάση Δεδομένων με διάφορες έγχρωμες εικόνες (benchmark και ground truth data)
2. Επιλογή παλέτας
3. Μέτρα απόδοσης και ποιότητας
4. Αποτελέσματα από γνωστούς αλγορίθμους και σύγκριση

Με το πέρασμα των χρόνων και με το αυξανόμενο ενδιαφέρον της επιστημονικής κοινότητας γύρω από το πεδίο του CBIR, σχηματίστηκαν διάφορες βάσεις δεδομένων, ώστε να χρησιμοποιούνται κατά τον έλεγχο των συστημάτων. Αυτές οι βάσεις δεδομένων, χαρακτηρίζονται ως benchmarks. Για τον έλεγχο του συστήματος χρησιμοποιήθηκε η ΒΔ Wang [60], η οποία αποτελείται από 10 θεματικά σετ 100 εικόνων. Η ΒΔ τροποποιήθηκε ελαφρώς, αντικαθιστώντας 10 τυχαίες εικόνες (μία από κάθε σετ) με αντίγραφα τους μετά από περιστροφή, ώστε να εξετασθεί εάν έχει επιτευχθεί ανεξαρτησία από την περιστροφή και να γίνει εν τέλει

σύγκριση αυτού του χαρακτηριστικού σε σχέση με τους περιγραφείς του MPEG-7. Ως προς τη δομή της ΒΔ, αν και η θεματική ενότητα ενός σετ είναι η ίδια, το περιεχόμενο των εικόνων μπορεί να ερμηνευτεί με διαφορετικούς τρόπους. Για παράδειγμα, εάν ερωτηθεί η ΒΔ για εικόνες με ουρανό, τα σωστά αποτελέσματα θα ήταν η παραλία, το μνημείο, το βουνό και οι ελέφαντες, ακόμη και αν αυτές οι εικόνες δεν αποτελούν το ίδιο εννοιολογικό σύνολο. Τον προβληματισμό αυτό προσέγγισε ο Keister [59]. Αυτή η ΒΔ είναι σχεδιασμένη, ώστε να διαπιστώνεται η απόδοση της ανάκτησης ενός συστήματος, ως προς το εννοιολογικό περιεχόμενο. Συνεπώς, όλα τα αποτελέσματα των μέτρων απόδοσης προϋποθέτουν ως *ground truth* το εννοιολογικό σύνολο ως ομάδα (π.χ. λεωφορεία, δεινόσαυροι, κτλ.). Στο σύνολο των πειραμάτων η παλέτα των χρωμάτων που χρησιμοποιείται για την περιγραφή των εικόνων, θεωρήθηκε πως θα πρέπει να περιοριστεί στα βασικά χρώματα, ώστε να αποδειχθεί ο ισχυρισμός του Wertheimer. Ως εκ τούτου το πλήθος των χρωμάτων είναι έξι ($N_H=6$): κόκκινο, κίτρινο, πράσινο, κυανό, μπλε, και μωβ. Η κλίμακα του γκρι εκφράστηκε με τρεις τόνους ($N_G=3$) με το μαύρο να αντιστοιχεί σε τιμές φωτεινότητας μικρότερες του 0.3, το γκρίζο σε τιμές μεταξύ 0.3 και 0.8 και το λευκό σε τιμές μεγαλύτερες του 0.8. Έτσι, το συνολικό πλήθος των χρωμάτων N ισούται με εννέα (9). Τα αποτελέσματα που επιδεικνύονται στη συνέχεια είναι αντιπροσωπευτικά της μεθόδου. Για κάθε ερώτημα (query) φαίνεται η εικόνα του ερωτήματος και οι εικόνες που ανακτήθηκαν με τη σειρά της κατάταξής τους, με σκοπό να γίνεται εμφανές στον αναγνώστη το ποιοτικό αποτέλεσμα της ανάκτησης. Επίσης, απεικονίζονται οι καμπύλες Precision-Recall για το σύνολο των εικόνων της ΒΔ, ώστε να διακρίνεται το σημείο τομής των καμπυλών. Η επίδοση της μεθόδου αναλύεται για κάθε ομάδα της ΒΔ, και συγκρίνεται με τις αντίστοιχες επιδόσεις των περιγραφέων του MPEG-7. Η σύγκριση παρουσιάζεται σε πίνακες με τις τιμές των μέτρων αποδοτικότητας Precision-Recall, και ANMRR.

3.6.2 Παρουσίαση αποτελεσμάτων

	1: 0.000000		2: 1.453831		3: 1.480276		4: 1.700950		5: 1.751134
	6: 1.764896		7: 1.807892		8: 1.816209		9: 1.828705		10: 1.831891
	11: 1.843041		12: 1.862868		13: 1.877010		14: 1.905394		15: 1.913287
	16: 1.932008		17: 1.937835		18: 1.949340		19: 1.975639		20: 1.978526

Εικόνα 30 Δείγμα αποτελεσμάτων ανάκτησης, από την ομάδα Αφρικανοί



Εικόνα 31 Δείγμα αποτελεσμάτων ανάκτησης, από την ομάδα Παραλίες



Εικόνα 32 Δείγμα αποτελεσμάτων ανάκτησης, από την ομάδα Λεωφορεία

	3: 0.588596		8: 0.718118		13: 0.959843		18: 1.043428	
	4: 0.603658		9: 0.761712		14: 0.987100		19: 1.065351	
	5: 0.641353		10: 0.797869		15: 0.994865		20: 1.084532	
1: 0.000000	2: 0.386031		7: 0.659748		12: 0.943446		17: 1.041382	
6: 0.654304					11: 0.839464		16: 1.035634	

Εικόνα 33 Δείγμα αποτελεσμάτων ανάκτησης, από την ομάδα Δεινόσαυροι

1: 0.000000		2: 0.427408			3: 1.302313		4: 1.989893		5: 2.208508	
6: 2.248274		7: 2.627440			8: 2.665427		9: 2.667787		10: 2.866450	
11: 2.970454		12: 3.005250			13: 3.071071		14: 3.152130		15: 3.241689	
16: 3.252509		17: 3.281137			18: 3.303490		19: 3.312544		20: 3.325294	

Εικόνα 34 Δείγμα αποτελεσμάτων ανάκτησης, από την ομάδα Ελέφαντες

	1: 0.000000				
6: 1.485234	7: 1.488647	8: 1.547556	9: 1.639817	10: 1.649337	
					
11: 1.665856	12: 1.667975	13: 1.720786	14: 1.822894	15: 1.831800	
					
16: 1.851785	17: 1.854013	18: 1.876056	19: 1.891548	20: 1.893295	
					

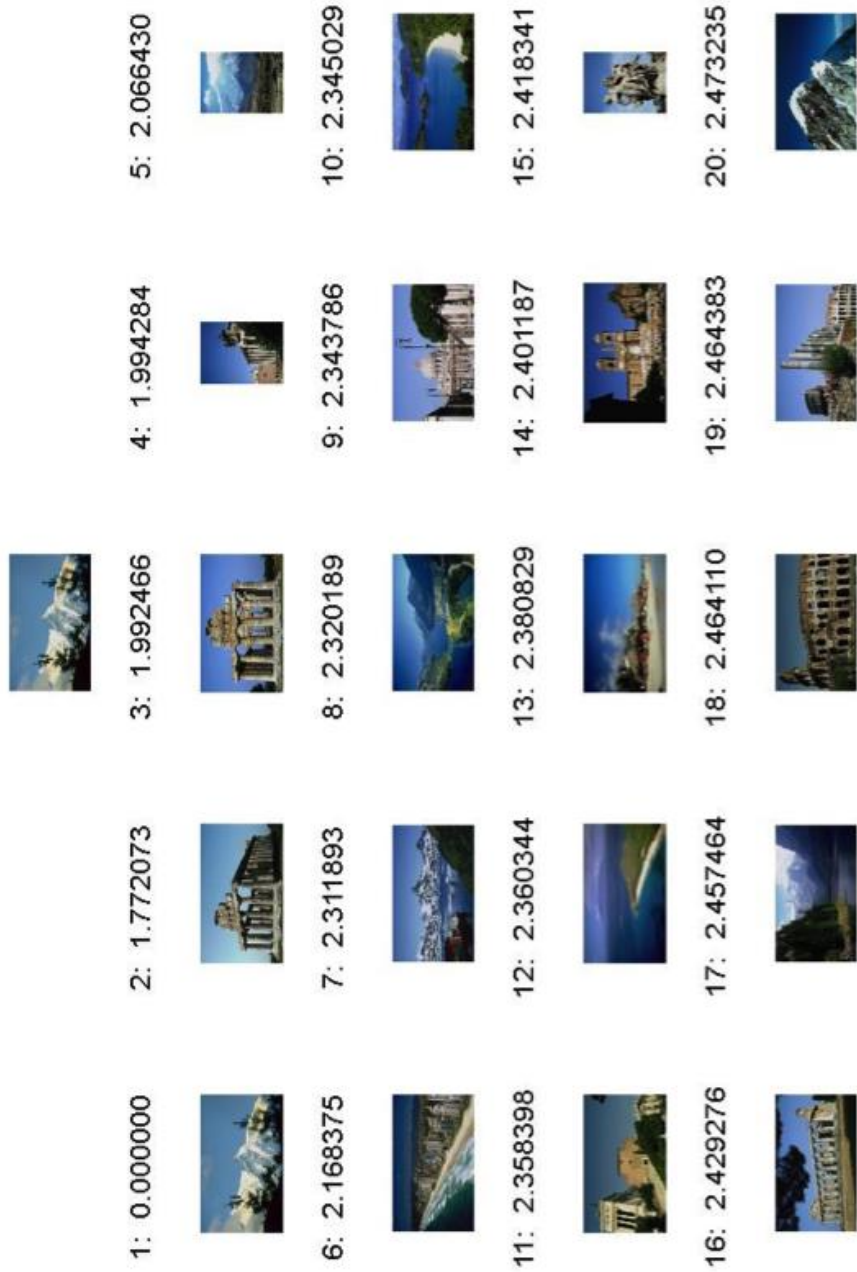
Εικόνα 35 Δείγμα αποτελεσμάτων ανάκτησης, από την ομάδα Μνημεία



Εικόνα 36 Δείγμα αποτελεσμάτων ανάκτησης, από την ομάδα Άλογα

1: 0.000000		2: 1.510879		3: 1.556471		4: 1.606499		5: 1.684455	
6: 1.704508		7: 1.716675		8: 1.720868		9: 1.759485		10: 1.778032	
11: 1.789569		12: 1.858555		13: 1.865400		14: 1.872630		15: 1.873657	
16: 1.920859		17: 1.981787		18: 1.986049		19: 1.989378		20: 1.997751	

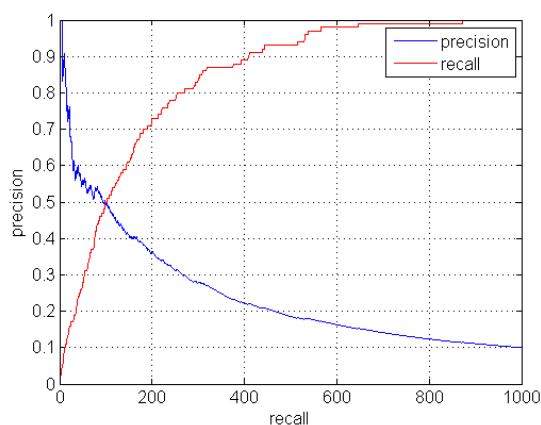
Εικόνα 37 Δείγμα αποτελεσμάτων ανάκτησης, από την ομάδα Φαγητά



Εικόνα 38 Δείγμα αποτελεσμάτων ανάκτησης, από την ομάδα Βουνά

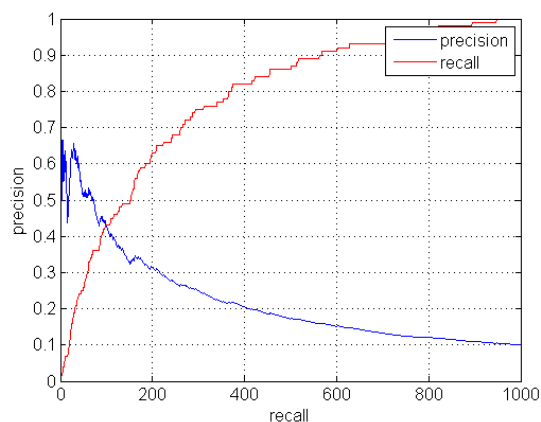


Εικόνα 39 Λείγμα αποτελεσμάτων ανάκτησης, από την ομάδα Τριαντάφυλλα



Εικόνα 40. Γράφημα precision/recall για το ερώτημα στην κατηγορία «Αφρικανοί»

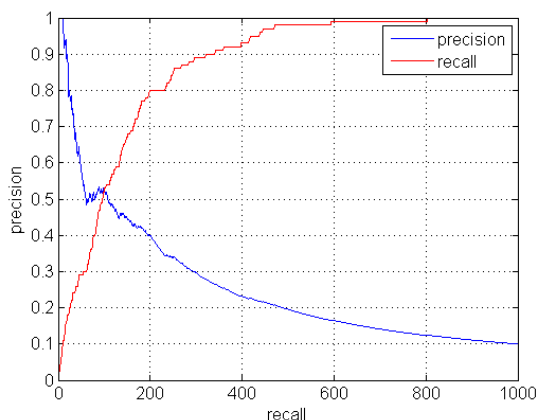
Το γράφημα στην Εικόνα 40 είναι το αποτέλεσμα του ερωτήματος με εικόνα από την κατηγορία «Αφρικανοί». Η πληθώρα των χρωμάτων και ο συνδυασμός τους μέσα στην εικόνα, καθιστά το ερώτημα δύσκολο σε σχέση με τις υπόλοιπες εικόνες του σετ. Το σημείο τομής των καμπυλών είναι στο 0.5, πράγμα το οποίο σημαίνει πως βρέθηκαν περίπου οι μισές σωστές εικόνες σε πλήθος ανακτήσεων ίσο με το σύνολο των σωστών εικόνων.



Εικόνα 41. Γράφημα precision/recall για το ερώτημα στην κατηγορία «Παραλίες»

Το γράφημα στην Εικόνα 41 είναι το αποτέλεσμα του ερωτήματος με εικόνα από την κατηγορία «παραλίες». Στα αποτελέσματα παρουσιάζονται εσφαλμένες εμφανίσεις εικόνων μνημείων ή βουνά. Αν και εννοιολογικά το σφάλμα εντοπίστηκε σωστά από τις καμπύλες precision/recall, σε μια διαφορετική ερμηνεία του περιεχομένου των εικόνων οι καμπύλες θα ή-

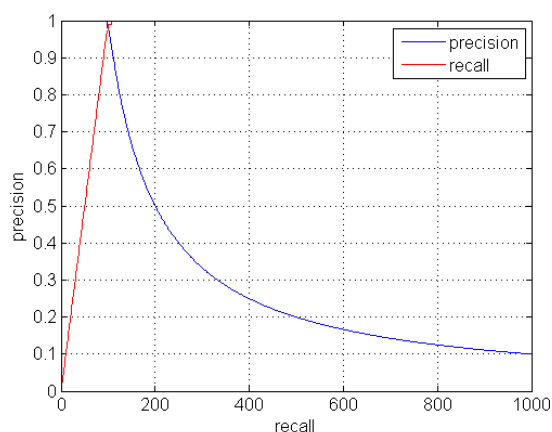
ταν διαφορετικές. Για παράδειγμα, εάν είχε ζητηθεί εικόνα με περιεχόμενο ουρανό ή «πολύ μπλε και λίγο πράσινο», καθώς στο ερώτημα υπάρχουν κατεξοχήν αυτοί οι συνδυασμοί, τότε δε θα είχε σημειωθεί κανένα σφάλμα τουλάχιστον στα 20 πρώτα αποτελέσματα. Το χρωματικό περιεχόμενο των 20 πρώτων αποτελεσμάτων είναι ομοιόμορφο.



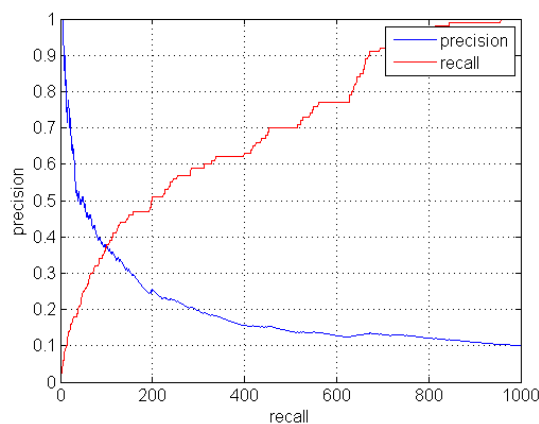
Εικόνα 42. Γράφημα precision/recall για το ερώτημα στην κατηγορία «Λεωφορεία»

Η γράφημα στην Εικόνα 42 είναι το αποτέλεσμα του ερωτήματος με εικόνα από την κατηγορία «λεωφορεία». Το σετ αυτό, περιέχει λεωφορεία με διάφορους χρωματικούς συνδυασμούς. Επομένως, το σημείο τομής των καμπυλών, το οποίο είναι λίγο παραπάνω από 0.5, μας δείχνει μια σχεδόν άριστη ανάκτηση, καθώς στα πρώτα 20 αποτελέσματα παρουσιάζονται μόνο 2 σφάλματα. Αν το ερώτημα είχε προσδιορισθεί ως «κόκκινο λεωφορείο», τότε τα αποτελέσματα της ανάκτησης θα ήταν επιτυχημένα.

Στην ομάδα «δεινόσαυροι» τα αποτελέσματα του συστήματος της ανάκτησης εικόνων βάσει περιεχομένου, σχεδόν πάντα μοιάζουν όπως αυτό στην Εικόνα 43, δηλαδή από κάθε άποψη άριστα. Τόσο στα πρώτα 20 αποτελέσματα, όσο και στα γραφήματα των καμπυλών precision/recall φαίνεται η επιτυχία. Το σημείο τομής είναι σχεδόν στο 1, ενώ στο γράφημα της σχέσης μεταξύ των δεικτών απόδοσης precision/recall, το εμβαδόν κάτω από τη καμπύλη είναι σχεδόν 1 επίσης.

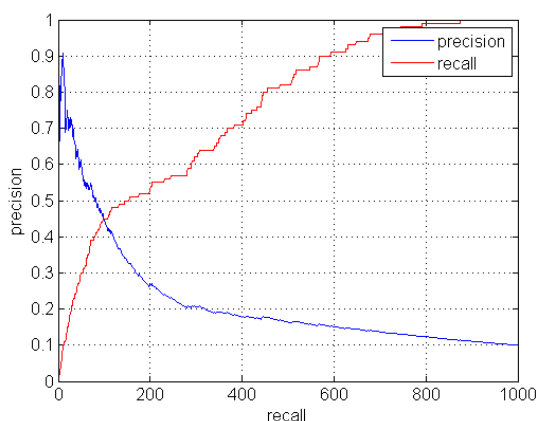


Εικόνα 43. Γράφημα precision/recall για το ερώτημα στην κατηγορία «Δεινόσαυροι»



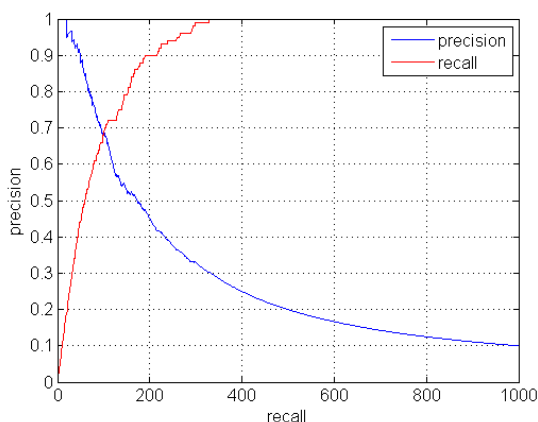
Εικόνα 44. Γράφημα precision/recall για το ερώτημα στην κατηγορία «Ελέφαντες»

Το γράφημα στην Εικόνα 44 είναι το αποτέλεσμα του ερωτήματος με εικόνα από την κατηγορία «ελέφαντες». Η κατηγορία αυτή είναι δύσκολη, λόγω τόσο του μεταβλητού πλήθους των ελεφάντων στις εικόνες όσο και του παρασκηνίου στο οποίο απεικονίζονται. Επομένως, η ποιότητα της ανάκτησης δεν μπορεί να γίνει πλήρως εμφανής από τις καμπύλες precision/recall. Παρόλα αυτά, τα πρώτα αποτελέσματα υποδεικνύουν πως με τους δεδομένους χρωματικούς συνδυασμούς, το θέμα «ελέφαντες» με το συγκεκριμένο φόντο βρίσκεται στις πρώτες θέσεις, άρα η ανάκτηση κρίνεται σε κάποιο βαθμό επιτυχής.



Εικόνα 45. Γράφημα precision/recall για το ερώτημα στην κατηγορία «Μνημεία»

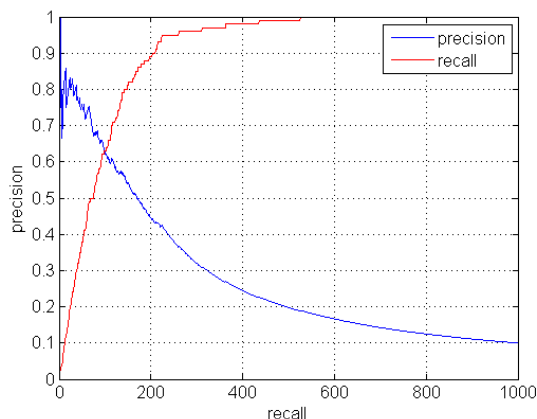
Το γράφημα στην Εικόνα 45 είναι το αποτέλεσμα του ερωτήματος με εικόνα από την κατηγορία των «μνημείων». Το σημείο τομής των καμπυλών είναι σχεδόν στο 0.45. Εάν εξετασθεί το τρίτο αποτέλεσμα στην πρώτη εικοσάδα, η ομοιότητα μεταξύ του ερωτήματος και του αποτελέσματος είναι εμφανής. Ακόμη και το ίδιο το σύνολο των «μνημείων» παρουσιάζει εννοιολογική δυσκολία, καθώς σε αυτό ανήκουν τόσο κτήρια, όσο και αγάλματα ή αγίδες. Με διαφορετικό φωτισμό και γενικότερα φόντο, καθιστούν πολύ δύσκολη την εύρεση ομοιότητας, πέρα της εννοιολογικής.



Εικόνα 46. Γράφημα precision/recall για το ερώτημα στην κατηγορία «Άλογα»

Το γράφημα στην Εικόνα 46 είναι το αποτέλεσμα του ερωτήματος με εικόνα από την κατηγορία «άλογα». Τα αποτελέσματα της ανάκτησης, συνήθως μοιάζουν με το παραπάνω, δηλαδή ο δείκτης επιτυχίας είναι πολύ

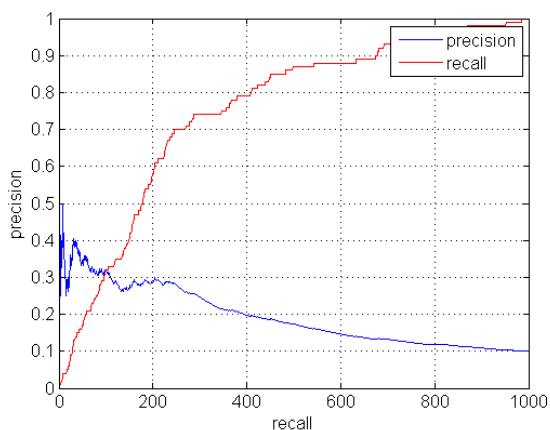
υψηλός. Ο λόγος της επιτυχίας είναι ότι σε αυτό το σετ εικόνων, τόσο το εννοιολογικό θέμα όσο και η απεικόνιση του είναι ομοιόμορφη. Το εμβαδόν της καμπύλης της σχέσης μεταξύ των δύο δεικτών (κάτω δεξιά εικόνα) καταλαμβάνει το μεγαλύτερο τμήμα του γραφήματος, πράγμα που επαληθεύει και τον σχολιασμό.



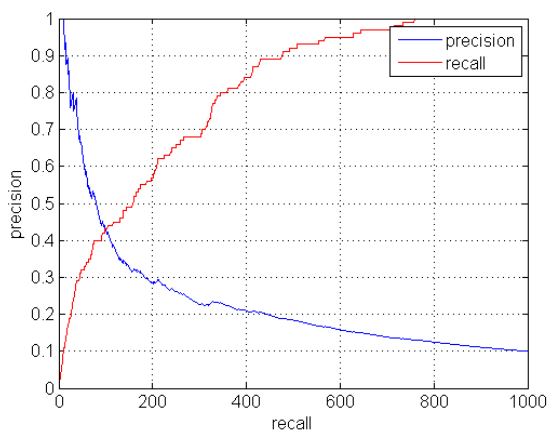
Εικόνα 47. Γράφημα precision/recall για το ερώτημα στην κατηγορία «Φαγητό»

Το γράφημα στην Εικόνα 47 είναι το αποτέλεσμα του ερωτήματος με εικόνα από την κατηγορία «φαγητό». Η ποικιλομορφία των εικόνων αυτής της ομάδας είναι μεγάλη. Παρόλα αυτά, τα αποτελέσματα είναι χρωματικά ομοιόμορφα, αλλά και οι καμπύλες δίνουν καλά στοιχεία. Τόσο το εμβαδόν κάτω από τη καμπύλη είναι ενδεικτικό των καλών αποτελεσμάτων (εικόνα κάτω δεξιά), όσο και η τιμή του σημείου τομής είναι αρκετά υψηλό σε σχέση με τη δυσκολία του σετ. Σε αυτά τα αποτελέσματα φαίνεται πως το σύστημα που σχεδιάστηκε σε αυτή την εργασία μπορεί να ανταπεξέλθει σε δύσκολα ερωτήματα.

Το γράφημα στην Εικόνα 48 είναι το αποτέλεσμα του ερωτήματος με εικόνα από την κατηγορία των «βουνών», ενώ είναι εννοιολογικά απλή καθώς το περιεχόμενο της αποτελείται από ουρανό και βουνά άλλοτε χιονισμένα και άλλοτε όχι, οι τιμές των δεικτών precision-recall είναι πάντα χαμηλές. Ο λόγος βέβαια είναι πως σε αρκετές εικόνες άλλων σετ υπάρχει το ίδιο φόντο (ουρανός) ή βουνά, τα οποία μοιάζουν αρκετά με κτίρια, όπως φαίνεται στα αποτελέσματα ανάκτησης ή όπως σχολιάστηκε προηγουμένως σε παρόμοιο παράδειγμα.



Εικόνα 48. Γράφημα precision/recall για το ερώτημα στην κατηγορία «Βουνά»



Εικόνα 49. Γράφημα precision/recall για το ερώτημα στην κατηγορία «Ρόδα»

Το γράφημα στην Εικόνα 49 είναι το αποτέλεσμα του ερωτήματος με εικόνα από την κατηγορία των «ρόδων», το οποίο όπως φαίνεται εντοπίζει κάποιες δυσκολίες, λόγω των διαφορετικών χρωμάτων των λουλουδιών. Άλλα τριαντάφυλλα είναι κόκκινα, και άλλα είναι κίτρινα, αυτό οδηγεί σε απόρριψη κάποιων αποτελεσμάτων που ενώ είναι εννοιολογικά σωστά, χρωματικά δεν είναι όμοια. Αυτό φαίνεται στα αποτελέσματα τόσο των εικόνων, όσο και στα γραφήματα των δεικτών precision-recall. Στα γραφήματα φαίνεται μία μέτρια επίδοση, ενώ στη πραγματικότητα η ανάκτηση θα έπρεπε να αφορά μόνο τα πορτοκαλί τριαντάφυλλα, τα οποία και έχουν τις πρώτες θέσεις στην κατάταξη των αποτελεσμάτων.

3.6.3 Σύγκριση αποτελεσμάτων

Σε επόμενο στάδιο της αξιολόγησης του περιγραφέα CGD που αναπτύχθηκε πραγματοποιείται σύγκριση των αποτελεσμάτων ανάκτησης που επιτυγχάνει με αυτά που επιτυγχάνουν οι γνωστοί αλγόριθμοι περιγραφής οπτικού περιεχομένου του προτύπου MPEG-7. Τα αποτελέσματα των οπτικών περιγραφέων του προτύπου ήταν διαθέσιμα από την εφαρμογή img(Rummager) [61], η οποία αποτελεί υλοποίηση στο πλαίσιο διδακτορικής έρευνας. Η εφαρμογή δοκιμάστηκε στην ίδια ΒΔ, στην οποία έλαβαν μέρος τα πειράματα της προτεινόμενης μεθόδου. Στους πίνακες που ακολουθούν παρουσιάζονται τα συγκριτικά αποτελέσματα από τους περιγραφείς CLD, SCD, CHD, και CGD (ο περιγραφέας που αναπτύχθηκε). Τα μέτρα σύγκρισης είναι η μέση τιμή ακρίβειας στο σημείο τομής των καμπυλών precision/recall στον Πίνακα V, η μέση τιμή ακρίβειας κατά την εικοστή ανάκτηση (MAP@20) στον Πίνακα VI, και το ANMRR στον Πίνακα VII. Τέλος, οι εικόνες που ανακτώνται από τους περιγραφείς επιδεικνύονται σε αντιπαράθεση, ώστε να γίνει αντιληπτό στον αναγνώστη, ποια είναι τα συγκριτικά οπτικά αποτελέσματα των περιγραφέων.

Πίνακας V. MAP ανά κατηγορία στη ΒΔ Wang

Σετ\περιγραφέας	CLD	CHD	SCD	CGD
Αφρικανοί	0.3291	0.3859	0.3867	0.3642
Παραλίες	0.3060	0.2255	0.1863	0.2960
Λεωφορεία	0.2805	0.2508	0.4854	0.4593
Δεινόσαυροι	0.9915	0.9394	0.6083	0.9494
Ελέφαντες	0.2966	0.2088	0.2363	0.2461
Φαγητό	0.3572	0.2933	0.1859	0.4632
Αλογα	0.7503	0.3116	0.3093	0.8031
Μνημεία	0.2276	0.1774	0.2883	0.2914
Βουνά	0.3461	0.1546	0.1687	0.2377
Ρόδα	0.4618	0.4001	0.3503	0.3506
Μέσος Όρος	0.4347	0.3347	0.3206	0.4461

Στον Πίνακα V απεικονίζονται οι τιμές της μέσης ακρίβειας, στην περίπτωση κατά την οποία είναι επιθυμητή η ανάκτηση μίας ολόκληρη ομάδας εικόνων (100 εικόνες του σετ). Είναι εμφανές ότι η μέθοδος περιγραφής που αναπτύχθηκε είναι είτε πρώτη σε απόδοση είτε δεύτερη με μικρή σχετική διαφορά ενώ ο μέσος όρος της προτεινόμενης μεθόδου, είναι υψηλότερος από όλους τους περιγραφείς του MPEG-7.

Πίνακας VI P@20 ανά κατηγορία στη BA Wang

Σετ\περιγραφείας	CLD	CHD	SCD	CGD
Αφρικανοί	0.434	0.637	0.545	0.496
Παραλίες	0.417	0.363	0.281	0.44
Λεωφορεία	0.404	0.337	0.721	0.596
Δεινόσαυροι	0.998	0.973	0.713	0.991
Ελέφαντες	0.462	0.408	0.341	0.439
Φαγητό	0.523	0.417	0.305	0.619
Άλογα	0.899	0.599	0.515	0.926
Μνημεία	0.355	0.284	0.436	0.495
Βουνά	0.450	0.214	0.246	0.352
Ρόδα	0.670	0.696	0.531	0.64
Μέσος Όρος	0.561	0.493	0.463	0.599

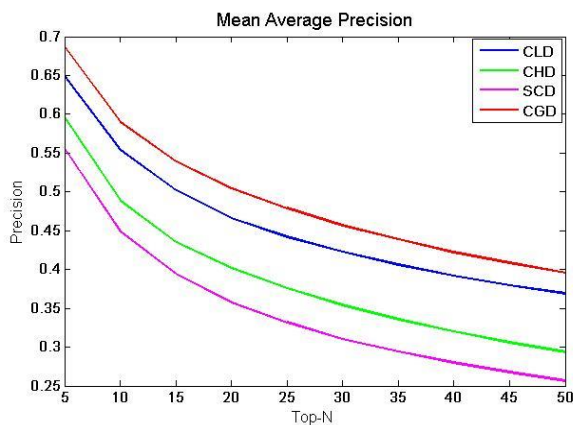
Ένας χρήστης συνήθως δεν επεκτείνεται σε αποτελέσματα που υπερβαίνουν τα 20 πρώτα ανακτηθέντα. Αυτό οδηγεί στην εξέταση της περίπτωσης του Πίνακα VI, δηλαδή τις τιμές της ακρίβειας στα πρώτα 20 ανακτηθέντα αποτελέσματα. Τα αποτελέσματα φαίνονται κι εδώ ενθαρρυντικά, καθώς σχεδόν στα μισά από τα σετ, η ακρίβεια του CGD, είναι καλύτερη από εκείνη των περιγραφών του MPEG-7, ενώ ο μέσος όρος της ακρίβειας για τις πρώτες 20 ανακτήσεις, είναι σημαντικά πιο μεγάλος από τους υπόλοιπους περιγραφείς. Είναι πλέον φανερό πως ο CGD είναι ανταγωνιστικός ως προς τους υπόλοιπους περιγραφείς, αφού πετυχαίνει να έχει καλύτερα αποτελέσματα κατά μέσο όρο τόσο στα πρώτα 20 αποτελέσματα όσο και σε όλα τα πιθανά αποτελέσματα.

Στον Πίνακα VII γίνεται μία σύγκριση της επίδοσης των περιγραφών με διαφορετικό μέτρο. Στο μέτρο ANMRR η μικρότερη τιμή είναι η καλύτερη. Στα ανά κατηγορίες αποτελέσματα, ο CGD έχει ανταγωνιστικά αποτελέσματα, πράγμα το οποίο σημαίνει πως η απόδοση του είναι ικανοποιητική, γεγονός που καταδεικνύεται και στο συνολικό αποτέλεσμα με τιμή 0.5499, η οποία είναι η καλύτερη μεταξύ των μέσων όρων όλων των μεθόδων.

Πίνακας VII. NMRR ανά κατηγορία, στη BA Wang

Σετ\περιγραφείας	CLD	CHD	SCD	CGD
Αφρικανοί	0.6563	0.5952	0.6012	0.6173
Παραλίες	0.6761	0.7497	0.7993	0.6870
Λεωφορεία	0.7127	0.7355	0.5068	0.5416
Δεινόσαυροι	0.0272	0.0946	0.4171	0.0857
Ελέφαντες	0.6938	0.7496	0.7497	0.7376
Φαγητό	0.6420	0.6797	0.7913	0.5399
Άλογα	0.2708	0.6598	0.6673	0.2306
Μνημεία	0.7525	0.8109	0.6892	0.6644
Βουνά	0.6430	0.8365	0.8201	0.7495
Ρόδα	0.5471	0.5968	0.6388	0.6453
ANMRR	0.5621	0.6508	0.6681	0.5499

Το γεγονός της καλής απόδοσης της μεθόδου CGD αποτυπώνεται και στη γραφική παράσταση της Μέσης Τιμής Ακρίβειας (Εικόνα 50). Η κόκκινη καμπύλη, η οποία αναφέρεται στην προτεινόμενη τεχνική περιγραφής των εικόνων, είναι υψηλότερη από τις άλλες, άρα και τα ποσοστά της ακρίβειας κατά την ανάκτηση είναι καλύτερα, από τους περιγραφείς του MPEG-7.



Εικόνα 50. MAP μεταξύ των περιγραφών του MPEG-7 και της μεθόδου CGD

Αν και τα παραπάνω μέτρα εκφράζουν την απόδοση των μεθόδων περιγραφής με τιμές, δεν αρκετό για να αξιολογηθεί η ποιότητα τους. Πολλές φορές κάποιες εικόνες ενώ είναι στο ίδιο σετ, το περιεχόμενό τους δεν είναι καθόλου ίδιο ως προς κάποια χαρακτηριστικά, τόσο χαμηλού επιπέ-

δου όπως τα χρώμα όσο και υψηλότερου επιπέδου, όπως το πλήθος των αντικειμένων στο προσκήνιο ή αλλαγή του παρασκήνιου. Έχει αναφερθεί προηγουμένως, πως το περιεχόμενο μιας εικόνας μπορεί να ερμηνευθεί διαφορετικά από άνθρωπο με άνθρωπο, καθώς και η ίδια εικόνα μπορεί να έχει διαφορετική σημασία για τον ίδιο σε κάποια άλλη χρονική στιγμή. Για αυτόν τον λόγο, παρουσιάζονται στη συνέχεια οι πρώτες 20 ανακτήσεις των περιγραφών που συγκρίθηκαν, για το ίδιο ερώτημα. Σκοπός αυτής της αντιπαραβολής, είναι η αξιολόγηση των αποτελεσμάτων από τον αναγνώστη, καθώς η ερμηνευτική ικανότητα και η αντίληψη του ανθρώπου, είναι ο απώτερος στόχος κάθε αλγορίθμου.

Ένα δείγμα από τα αποτελέσματα της ανάκτησης είναι αυτό της ομάδας των λεωφορείων που ακολουθεί. Σύμφωνα με τις καμπύλες Top-N, φαίνεται πως όλες οι μέθοδοι περιγραφής είναι κοντά, εκτός της μεθόδου Color Histogram. Σε ένα σχετικά ομοιόμορφο σετ, όπως αυτό, σημασία έχει η ανάκτηση να έχει πετύχει τόσο το εννοιολογικό περιεχόμενο, όσο και τη χρωματική συνάφεια μεταξύ των αποτελεσμάτων. Η μέθοδος που προτείνεται, έχει πετύχει τόσο την εννοιολογική περιγραφή, όπως υποδεικνύουν οι καμπύλες, όσο και τη χρωματική συνάφεια, όπως φαίνονται από τα αποτελέσματα.

Results of CGD retrieval

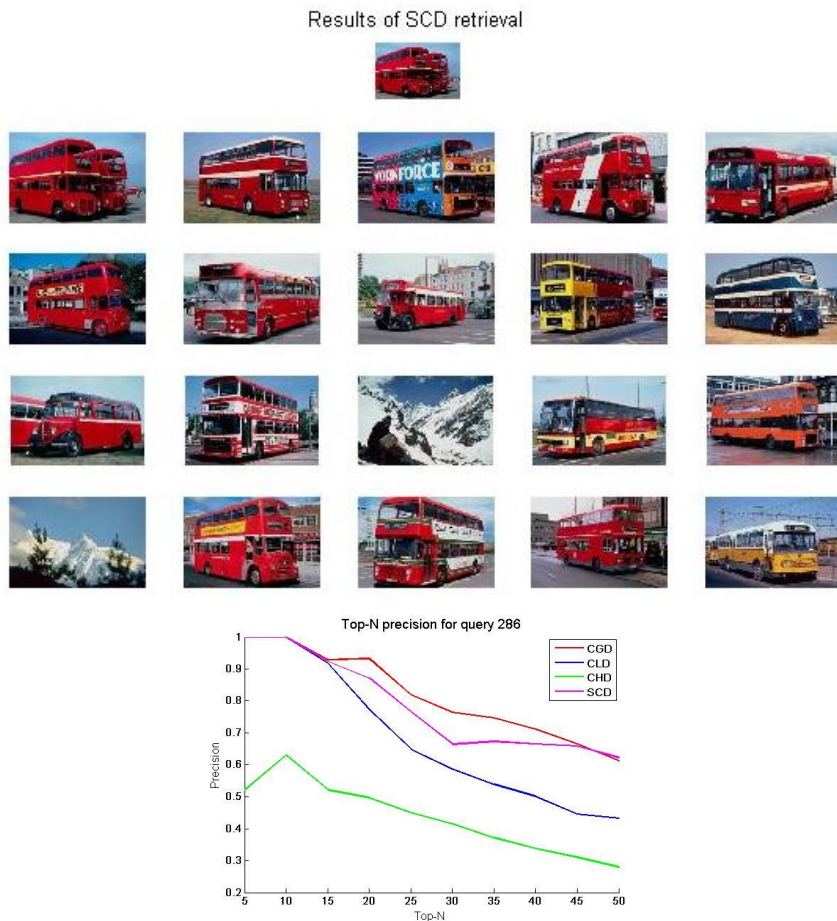


Results of CLD retrieval



Results of CHD retrieval

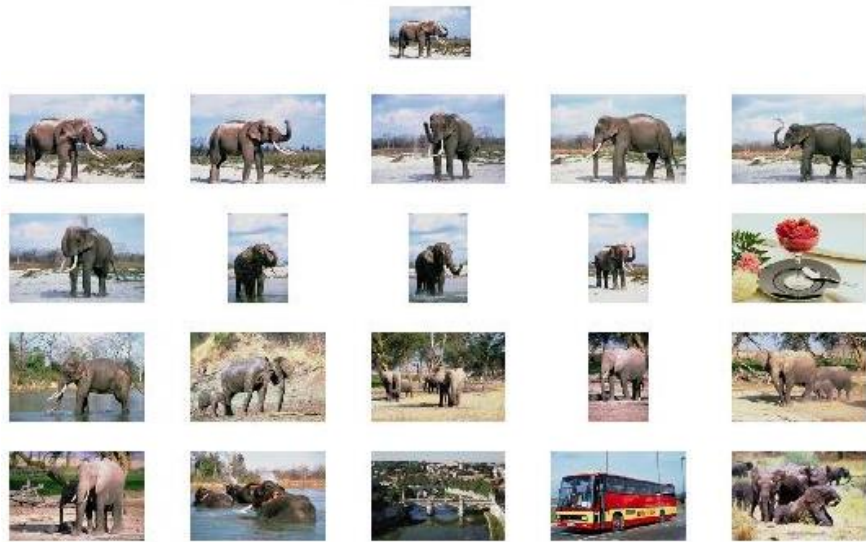




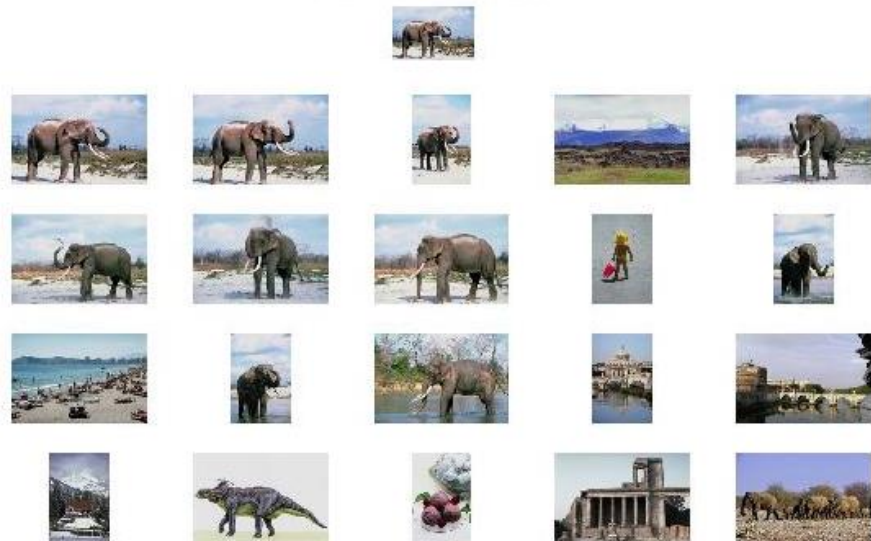
Εικόνα 51 Σύγκριση αποτελεσμάτων ανάκτησης, από την ομάδα λεωφορεία

Στην επόμενη ομάδα εικόνων, δεν υπάρχει ομοιογένεια στα χρώματα παρά στην έννοια που απεικονίζουν. Το παρασκήνιο είναι συχνά διαφορετικό, και αυτό έχει ως αποτέλεσμα αλγόριθμοι, που χειρίζονται μόνο χαρακτηριστικά χαμηλού επιπέδου, όπως το χρώμα, το σχήμα, και την υφή, να μην είναι ικανοί να πετύχουν τόσο μεγάλη ακρίβεια στις ανακτήσεις. Παρόλα αυτά, στο σετ υπάρχουν σχετικά όμοιες εικόνες, όπως του δείγματος, και είναι αναμενόμενο οι περιγραφείς να τις βρουν με επιτυχία. Η προτεινόμενη μέθοδος όχι μόνο τις βρήκε επιτυχώς, αλλά τις κατέταξε και με σωστή σειρά.

Results of CGD retrieval



Results of CLD retrieval

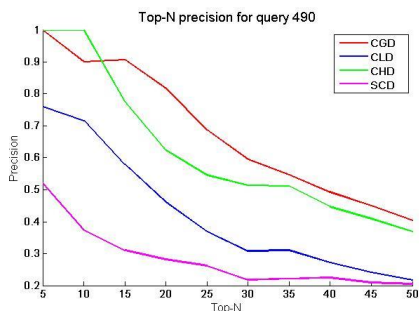


Results of CHD retrieval



Results of SCD retrieval





Εικόνα 52 Σύγκριση αποτελεσμάτων ανάκτησης, από την ομάδα ελέφαντες

Η ομάδα των πιάτων με φαγητά είναι από τις πιο δύσκολες ομάδες της ΒΔ. Η δυσκολία έγκειται τόσο στην ανομοιομορφία του παρασκηνίου όσο και στη ποικιλομορφία του προσκηνίου. Παρόλα αυτά, η προτεινόμενη μέθοδος παραμένει σταθερή στη σωστή κατάταξη των εικόνων, που ανέκτησε από τη ΒΔ. Από αυτό το παράδειγμα φαίνεται ο σημαντικός ρόλος της χρωματικής ομοιότητας σε συνδυασμό με τα χωρικά δεδομένα. Οι καμπύλες, παρουσιάζουν την επιτυχή κατάταξη των εικόνων.

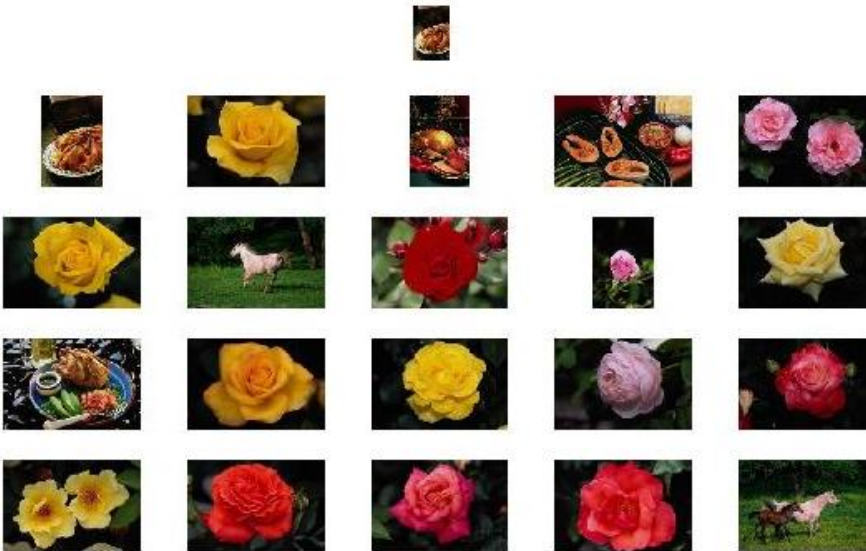
Results of CGD retrieval

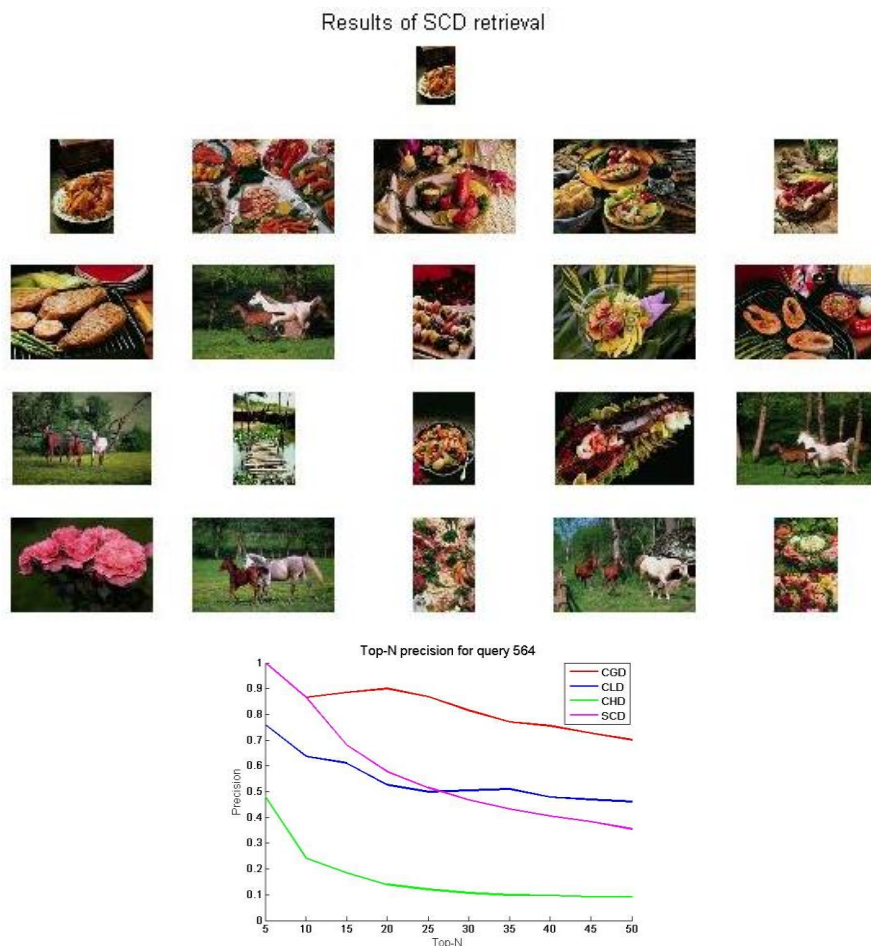


Results of CLD retrieval



Results of CHD retrieval





Εικόνα 53 Σύγκριση αποτελεσμάτων ανάκτησης, από την ομάδα φαγητά

Το δείγμα που εξετάζεται στη συνέχεια είναι από τη συλλογή των βουνών. Η συλλογή αυτή εννοιολογικά είναι συνεκτική, και σε μεγάλο μέρος ομοιόμορφη. Παρόλα αυτά, χαμηλά ποσοστά ανάκτησης παρουσιάζονται λόγω της σύγχυσης των εικόνων που απεικονίζουν βουνά με εκείνων που απεικονίζουν θαλάσσια και παραλιακά τοπία λόγω του κοινού τους σημείου, του ουρανού. Τόσο από τις καμπύλες όσο και από τα οπτικά αποτελέσματα, φαίνεται η επιτυχία της μεθόδου σε αντίθεση με τους περιγραφείς του MPEG-7, οι οποίοι φαίνεται να καταλήγουν μέχρι και σε χρωματικά σφάλματα.

Results of CGD retrieval



Results of CLD retrieval

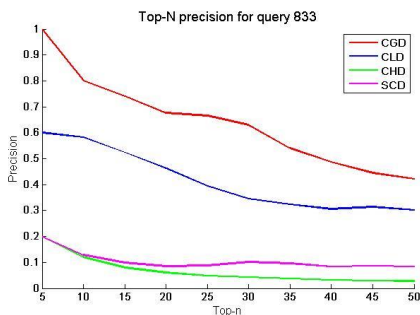


Results of CHD retrieval



Results of SCD retrieval





Εικόνα 54 Σύγκριση αποτελεσμάτων ανάκτησης, από την ομάδα βουνά

Η τελευταία ομάδα που εξετάζεται είναι των ρόδων. Η ομάδα αυτή παρουσιάζει ομοιογένεια ως προς το περιεχόμενο, καθώς το αντικείμενο αυτό καθ' εαυτό είναι παρόμοιο και απεικονίζεται σε παραπλήσιο παρασκήνιο. Η χρωματική ποικιλομορφία είναι το μόνο χαρακτηριστικό, το οποίο μπορεί να καταστήσει μη επιτυχημένη, την ανάκτηση όμοιων εικόνων. Από τη γραφική παράσταση, αλλά και από τα οπτικά αποτελέσματα, φαίνεται η αποδοτικότητα και η επιτυχία της μεθόδου CGD.

Results of CGD retrieval

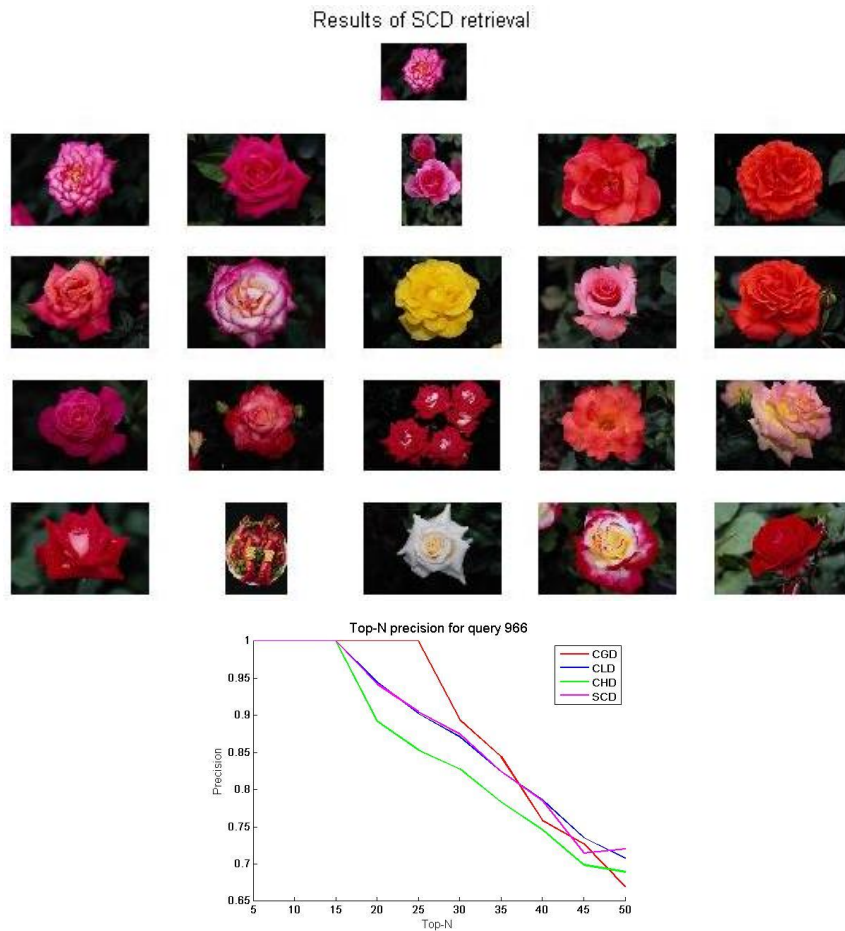


Results of CLD retrieval



Results of CHD retrieval





Εικόνα 55 Σύγκριση αποτελεσμάτων ανάκτησης, από την ομάδα τριαντάφυλλα

3.7 Συμπεράσματα

Στο πλαίσιο της παρούσας εργασίας μελετήθηκαν μέθοδοι εξαγωγής χαρακτηριστικών και ανάπτυξης συμπαγών περιγραφών έγχρωμων στατικών δισδιάστατων εικόνων. Αποτέλεσμα της έρευνας που πραγματοποιήθηκε είναι ένας νέος περιγραφέας έγχρωμων εικόνων με την ονομασία Color Gravity Descriptor (CGD), ο οποίος μπορεί να χαρακτηριστεί αρκετά επιτυχημένος δεδομένης της απλότητας εξαγωγής του και των αποτελεσμάτων που επιτυγχάνει. Η μέθοδος περιλαμβάνει προεπεξεργασία των εικόνων για μείωση των χρωμάτων, η οποία μπορεί να χαρακτηριστεί απλοϊκή και κάπως αυστηρή ως προς το γεγονός ότι χρησιμοποιεί «σκλήρα» κατώφλια, τα οποία δεν προσαρμόζονται με ευφυή τρόπο στη περίπτωση της κάθε εικόνας. Η τελική υπογραφή της εικόνας καταφέρνει να συνδυάσει επιτυχημένα τόσο τα δεδομένα του χρώματος όσο και χωρική πληροφορία. Αξιοσημείωτο στη μέθοδο είναι ότι η περιγραφή της εικόνας είναι ανεξάρτητη περιστροφής και κλιμάκωσης. Οι σταθεροί μήκους υπογραφές που προκύπτουν καθιστούν τη ανάκτηση ταχύτατη. Τα αποτελέσματα είναι ιδιαίτερα ικανοποιητικά καθώς είναι συγκρίσιμα και κατά μέσο όρο καλύτερα (σε όρους MAP) από αυτά που επιτυγχάνουν οι περιγραφείς που περιλαμβάνονται στο καθιερωμένο διεθνές πρότυπο MPEG-7. Ένα από τα πλεονεκτήματα της μεθόδου αυτής είναι η χρωματική ομοιομορφία των αποτελεσμάτων, δηλαδή μια σταθερή συνάφεια μεταξύ των χρωμάτων της εικόνας ερωτήματος με τις εικόνες της ανάκτησης. Η ομοιομορφία των αποτελεσμάτων σε συνδυασμό με τα αποτελέσματα των μέτρων απόδοσης precision-recall, και MAP, αποδεικνύουν την αρχική υπόθεση της παρούσας εργασίας. Η μέθοδος, προφανώς, έχει πολλά περιθώρια περαιτέρω βελτίωσης. Αρχικά, η μέθοδος της μείωσης των χρωμάτων, θα μπορούσε να είναι περισσότερο ελαστική και προσαρμοζόμενη ως προς τη θέσπιση των κατωφλίων, έτσι ώστε η εικόνα που προκύπτει έπειτα από την επεξεργασία να είναι όσο το δυνατό πιο κοντά στην αρχική. Παράλληλα εξετάζεται η πιθανότητα βελτίωσης των αποτελεσμάτων με την εφαρμογή προεπεξεργασίας με εφαρμογή φίλτρων εξομάλυνσης (Gaussian) για περαιτέρω προσαρμογή στο ανθρώπινο σύστημα όρασης και οπτικής αντίληψης. Συμπληρωματικά, ο συνυπολογισμός της υψής, σημείων αναφοράς (salient points) ή και σχημάτων θα μπορούσε να οδηγήσει σε πιο ασφαλή συμπεράσματα, και καλύτερα αποτελέσματα. Επίσης, η ενσωμάτωση τεχνικών όπως relevance feedback, θα μπορούσε να καταφέρει ακόμη καλύτερα αποτελέσματα στο τμήμα της εννοιολογικής συνάφειας. Το πεδίο της ανάκτησης εικόνων βάσει περιεχομένου, είναι ένας δυναμικός κλάδος, καθώς τα πολυμεσικά αρχεία πληθαίνουν και οι ανάγκες για πιο απαιτητικά αποτελέσματα αναζήτησης και ανάκτησης βάσει περιεχομένου είναι ανάλογες. Η μέθοδος που αναπτύχθηκε στο πλαίσιο της παρούσας εργασίας, με τις κατάλληλες βελτιώσεις μπορεί να προσφέρει περαιτέρω στην εξέλιξη του πεδίου.

Βιβλιογραφία

- [1] Wu “Efficient Statistical Computations For Optimal Color Quantization”, in Graphics Gems, vol. II, edited, by James Arvo, Academic Press, Inc., Cambridge, MA, pp. 126-133., 1991
- [2] Instagram press, <http://instagram.com/press/>, τελευταία επίσκεψη: 1 Οκτ. 2013.
- [3] Tumblr press, <http://www.tumblr.com/press>, τελευταία επίσκεψη: 20 Αυγ. 2013.
- [4] A.W.M. Smeulders, M. Worring, S. Santini, A. Gupta, R. Jain, "Content-Based Image eTrieval at the End of the Early Years", vol.22 no. 12, DEC 2000
- [5] Proc. US NSF Workshop Visual Information Management Systems, R. Jain, ed., 1992
- [6] S. Arya, D.M. Mount, N.S. Netanyahu, R. Silverman, and A.Y. Wu, "An Optimal Algorithm for Approximate Nearest Neighborhood Searching," Proc. Symp. Discrete Algorithms, pp. 573-582, 1994
- [7] J. Huang, Q. Dom, B. Gorkani, M. Hafner, J. Lee, D. Petkovic, D. Steele, D. Flickner, "Query by image and video content: the qbic system", Computer, 28(9):23ύ32, 1995
- [8] J.R. Smith, S.F. Chang, "Visual Seek: a fully automated content based image query system", Proceedings of the fourth ACM international conference on Multimedia" pp. 87-98, 1997
- [9] A. Petland, RW Picard, S Sclaroff, "Photobook: content base manipulation of image databases", International Journal of Computer Vision, 18(3):233-254, 1996
- [10] Google press, <http://www.google.com/press/>, τελευταία επίσκεψη: 10 Αυγ. 2013.
- [11] M. Lux, "Caliph & Emir: Mpeg-7 photo annotation and retrieval", Proceedings of the seventeen ACM international conference on Multimedia, pp. 925-926, 2009
- [12] ISO/IEC/JTC1/SC29/WG11, "Description of core experiments for mpeg-7 color/texture descriptors", 1999

- [13] D.A. Atchison, G. Smith, "Optics of the Human Eye", Butterworth-Heinemann, Boston, Mass. 2000
- [14] Hecht, S., The development of Thomas Young's theory of color vision. 1930. J. Optical Society of America (JOSA), Vol. 20, Issue 5, pp. 231-231, May 1930.
- [15] M. Stone, "A Field Guide to Digital Color", A.K. Peters, Natick, Ma, 2003
- [16] W.K. Pratt, "Digital Image Processing", A Wiley interscience publication, 1991
- [17] Rafael.C.Gonzalez, Richard E.Woods Digital image processing, 3rd edition, Pearson education, 2008
- [18] N. Otsu, "A Threshold selection method from gray-level histograms" IEEE Tran. on System man and Cybernetics, SMC-9(1), pp. 62-69, 1979
- [19] N. Papamarkos, C. Strouthopoulos, I. Andreadis, "Multithresholding of color and gray-level images through a neural network technique", Image and Vision Computing, Vol 18, pp. 213-222, 2000
- [20] S. Haykin, "Neural Networks: A Comprehensive Foundation", Macmillan College Publishing Company, Ney York, 1994
- [21] Kohonen, T.: The self-organizing map. Proceedings of IEEE vol. 78, no.9. (1990) 1464-1480
- [22] L. Fausett, Fundamentals of Neural Networks. Prentice Hall. pp. 169-187, 1994
- [23] K. Zagoris, N. Papamarkos, I. Koustoudis "Color Reduction using the Combination of the Kohonen Self-Organized Feature Map and the Gustafson-Kessel Fuzzy Algorithm", Transactions on Machine Learning and Data Mining, Vol.1, No 1, pp. 31 – 46, 2008
- [24] Papamarkos, N., Atsalakis, A., Strouthopoulos, C.: Adaptive color reduction. IEEE Transactions on Systems, Man and Cybernetics Part B: Cybernetics 32 (2002)
- [25] T. Kato, "Database architecture for content-based image retrieval" in Image Storage and Retrieval Systems (Jambardino, A A and Niblack, W R , eds), Proc SPIE 1662, 112-123, 1992
- [26] Gervautz, M., Purgathofer, W.: A simple method for color quantization: Octree quantization. Graphics Gems, A. S. Glassner, Ed. Academic New York (1990) 287–293
- [27] A.T. Ghanbarian, E. Kabir, and N.M. Charkari, Color reduction based on ant colony, Pattern Recognition Letters 28 (2007), 1383–1390
- [28] Wu (1991): "Efficient Statistical Computations For Optimal Color Quantization", in Graphics Gems, vol. II, edited, by James Arvo, Academic Press, Inc., Cambridge, MA, pp. 126-133.

- [29] Sclaroff S et al (1997) "ImageRover: a content-based image browser for the World-Wide Web" in Proceedings of IEEE Workshop on Content-Based Access of Image and Video Libraries, San Juan, Puerto Rico, June 1997, 2-9
- [30] H. Tamura and N. Yokoya, Image database systems: A survey, Pattern Recognition 17(1), 1984
- [31] S.-K. Chang and A. Hsu, Image information systems: Where do we go from here?IEEE Trans. on Knowledge and Data Engineering 4(5), 1992.
- [32] J.R. Smith, S.F. Chang, "Image Retrieval Evaluation", IEEE Workshop on Content Based Access of Image and Video Libraries", pp. 112-113, 1998
- [33] T. Chang and C.-C. J. Kuo, Texture analysis and classification with tree-structured wavelet transform, IEEE Trans. Image Proc. 2(4), 429-441, 1993.
- [34] D. Zhong and S.F. Chang, "Video Object Model and Segmentation for Content-Based Video Indexing," IEEE International Symposium on Circuits and Systems, Hong Kong, June 1997
- [35] W. Y. Ma and B. S. Manjunath, "Texture Features and Learning Similarity", IEEE Proceedings CVPR '96, pp. 425-430, 1996.
- [36] Y. Rui, A. C. She, and T. S. Huang, Modified fourier descriptors for shape representation—a practical approach, in Proc. of First International Workshop on Image Databases and Multi Media Search, 1996.
- [37] D. White and R. Jain, Algorithms and strategies for similarity retrieval, in TR VCL-96-101, University of California, San Diego, 1996.
- [38] D. White and R. Jain, Similarity indexing: Algorithms and performance, inProc. SPIE Storage and Retrieval for Image and Video Databases, 1996.
- [39] D. Zier, J. -R. Ohm : "Common Datasets and Queries in MPEG-7 Color Core Experiments ", ISO/IEC JTC1/SC29/WG11 (MPEG) document no. M5060, Melbourne, October 1999.
- [40] J.R. Ohm, L. Cieplinski, H.J. Kim, S. Krishnamachari, B. S. Manjunath, Dean S. Messing, Akio Yamada, "The mpeg-7 Color Descriptors",
- [41] C. S. Lee, W.-Y. Ma and H. J. Zhang, "Information Embedding Based on User's Relevance Feedback for Image Retrieval", Invited paper, SPIE Int. Conf. Multimedia Storage and Archiving Systems IV, Boston, pp.19-22, Sep. 1999.
- [42] Donna Harman, "Relevance Feedback Revisited", Proceedings of the Fifteenth Annual International ACM SIGIR conference on Research and development in information retrieval, pp. 1-10, 1992.

- [43] Gerald Salton and Chris Buckley, "Improving Retrieval Performance by Relevance Feedback", *Journal of the American Society for Information Science*, pp. 288-297, Vol. 41, No. 4, 1990.
- [44] D. Brahmi and D. Ziou. Improving CBIR systems by integrating semantic features. In *Proc. RIAO*, pages 291–305, Vaucluse, France, April 2004.
- [45] D. Messing, P. van Beek, J.H. Errico, "The MPEG-7 Colour Structure Descriptor: Image description using colour and local spatial information", *IEEE International Conference on Image Processing (ICIP 2001)*, Thessaloniki, Greece, October 2001
- [46] H. Ma et al. Bridging the semantic gap between image contents and tags. *IEEE TMM*, 2010
- [47] L. Wu et al. Semantics-preserving bag-of-words models for efficient image annotation. In *ACM workshop on LSMMRM*, 2009
- [48] Savvas A. Chatzichristofis, Chryssanthi Iakovidou, Yiannis S. Boutalis, Oge Marques: Co.Vi.Wo.: Color Visual Words Based on Non-Predefined Size Codebooks. *IEEE T. Cybernetics* 43(1): 192-205 (2013)
- [49] J. K. Hawkins, "Textural Properties for Pattern Recognition", in *Picture Proceeding and Psychopictorics*, B.C. Lipkin and A. Rosenfeld, Eds. Academic Press, New York, pp. 347-370, 1970
- [50] Raghavan, V., Bollmann, P., & Jung, G. S.. A, "critical investigation of recall and precision as measures of retrieval system performance". *ACM Trans. Inf. Syst.*, 7, 205{229. (1989)
- [51] Manning, C., & Schutze, H. "Foundations of statistical natural language processing. MIT Press. (1999)
- [52] Joachims, T. (2005). A support vector method for multivariate performance measures. *Proceedings of the 22nd International Conference on Machine Learning*. ACM Press.
- [53] Sylvie Philipp-Foliguet, Julien Gony, and Philippe-Henri Gosselin. "Frebir: An image retrieval system based on fuzzy region matching". *Comput. Vis. Image Underst.*, 113(6):693–707, 2009.
- [54] Jose M. Martinez, *Mpeg-7 overview* <http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm>
- [55] Jesse Davis and Mark Goadrich. 2006. The relationship between Precision-Recall and ROC curves. In *Proceedings of the 23rd international conference on Machine learning (ICML '06)*. ACM, New York, NY, USA, 233-240.
- [56] Ndjiki-Nya, J. Restat, T. Meiers, J.R. Ohm, A. Seyferth, and R. Sniehotta, "Subjective evaluation of the MPEG-7 retrieval accuracy measure (ANMRR)," in *52nd Int'l. MPEG Conf.*, Geneva, May 2000, ISO/IEC JTC1/SC29/WG11, doc. no. M6029

- [57] M. Stricker and M. Orengo, "Similarity of color images," SPIE Proc. Storage and Retrieval of Image and Video Databases, vol. 2420, pp. 381–392, 1995
- [58] M. Wertheimer, "Laws of Organization in Perceptual Forms", *Psychologische Forschung*, 4, pp. 301-350, (1923)
- [59] Keister, L H (1994) "User types and queries: impact on image access systems." In *Challenges in indexing electronic text and images* (Fidel, R et al., eds). ASIS, 1994, 7-22
- [60] J. Li, J. Z. Wang, "Automatic linguistic indexing of pictures by a statistical modeling approach," *IEEE Trans. Pattern Analysis and Machine Intelligence*, 25(9):1075-1088, 2003
- [61] Savvas A. Chatzichristofis, Yiannis S. Boutalis, and Mathias Lux. 2009. *Img(Rummager): An Interactive Content Based Image Retrieval System*. In *Proceedings of the 2009 Second International Workshop on Similarity Search and Applications (SISAP '09)*, Tomas Skopal and Pavel Zezula (Eds.). IEEE Computer Society, Washington, DC, USA, 151-153.
- [62] P. Sirotkin, "An evaluation of popular relevance metrics ", *Information und Wissen: global, sozial und frei Proc. des 12. Internationalen Symposiums für Informationswissenschaft VWH, Boizenburg*, Pages 24-35, 2011

[Αυτή η σελίδα είναι εσκεμμένα λευκή]

ΑΝΑΛΥΣΗ ΕΓΧΡΩΜΩΝ ΕΙΚΟΝΩΝ ΓΙΑ ΕΦΑΡΜΟΓΕΣ ΑΝΑΚΤΗΣΗΣ ΒΑΣΕΙ

ΠΕΡΙΕΧΟΜΕΝΟΥ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

Βασίλειος Σεβελίδης

Η παρούσα εργασία είχε ως αντικείμενο μελέτης την ανάκτηση έγχρωμων στατικών δισδιάστατων εικόνων βάσει περιεχομένου. Στόχος της εργασίας ήταν η ανάπτυξη μιας μεθοδολογίας και ενός περιγραφέα έγχρωμων εικόνων που να βασίζεται στην αναγνώριση καθολικών χαρακτηριστικών των εικόνων και να μην επηρεάζονται από την περιστροφή και την κλιμάκωση. Ο περιγραφέας θα μπορούσε να χρησιμοποιηθεί για εφαρμογές ανάκτησης βάσει περιεχομένου αποδίδοντας τουλάχιστον όσο οι περιγραφείς που έχουν προταθεί στο πλαίσιο του προτύπου MPEG-7. Μετά από εκτενή βιβλιογραφική αναζήτηση, βασική έρευνα στα θέματα της ανάκτησης βάσει περιεχομένου και πειραματική ανάπτυξη με τη χρήση διεθνώς αναγνωρισμένων εργαλείων λογισμικού, αναπτύχθηκε μια πρωτότυπη μεθοδολογία εξαγωγής καθολικών χαρακτηριστικών και η δημιουργία μιας νέας υπογραφής για έγχρωμες. Ο περιγραφέας έγχρωμων στατικών εικόνων Color Gravity Descriptor (CGD), ο οποίος αναπτύχθηκε, βασίζεται στην υπόθεση πως, εάν μειωθεί η χρωματική παλέτα σε κάποια βασικά χρώματα και έπειτα εξαχθούν κατάλληλα χαρακτηριστικά σχετιζόμενα με την τοποθέτηση των εν λόγω χρωμάτων στο χωρικό επίπεδο της εικόνας, μπορεί να επιτευχθεί ταύτιση όμοιων εικόνων. Στη μεθοδολογία εξαγωγής του CGD περιλαμβάνεται αλγόριθμος μείωσης χρωμάτων, επιλογής, μετατροπής και εξαγωγής χαρακτηριστικών και τελικής διαμόρφωσης της υπογραφής κάθε εικόνας. Η μέθοδος δοκιμάστηκε πειραματικά με χρήση τυπικών βάσεων δεδομένων για εφαρμογές ανάκτησης βάσει περιεχομένου και αποδείχθηκε ότι αποδίδει κατά μέσο όρο καλύτερα αποτελέσματα ανάκτησης από τις μεθόδους που περιλαμβάνονται στο πρότυπο MPEG-7.

ANALYSIS OF COLOUR IMAGES FOR CBIR APPLICATIONS

THESIS

Vasileios Sevetlidis

This work focuses on the content-based retrieval of two-dimensional static color images. The aim was to develop a methodology and a descriptor for color images based on the recognition of global characteristics of the images that will not be affected by rotation and scaling. The descriptor could be used for content-based retrieval applications yielding at least as good performance as the one provided by descriptors that have been proposed in the context of the MPEG-7 standard. After an extensive literature research, basic research on content-based retrieval and experimental development using internationally recognized software tools, a novel methodology has been developed that extracts global image features and creates a novel signature for color images. The Color Gravity Descriptor (CGD), which was developed as a result of this research, was based on the assumption that if image colors are reduced to a fixed palette and then appropriate characteristics associated with the location of these colors on that image plain are extracted, matching of similar images can be accomplished. The methodology of forming the CGD includes a color reduction algorithm, selection, extraction and transformation of features and the final formation of the signature of each image. The method was extensively tested using standard image databases and proved to provide overall average better retrieval results than methods included in the MPEG-7 standards.

