



ΠΑΝΕΠΙΣΤΗΜΙΟ ΜΑΚΕΔΟΝΙΑΣ
ΟΙΚΟΝΟΜΙΚΩΝ ΚΑΙ ΚΟΙΝΩΝΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΕΦΑΡΜΟΣΜΕΝΗΣ ΠΛΗΡΟΦΟΡΙΚΗΣ

Ανάκτηση Εικόνας με Βάση το Περιεχόμενο, Το πρότυπο MPEG-7
Μελέτη Περιπτώσεων: Alipr.com

Διπλωματική εργασία
ΤΟΥ
Ιωάννη Χ. Μασμανίδη
mai2907

Επιβλέπων: Αθανάσιος Μανισάρης
Καθηγητής

Θεσσαλονίκη 2010

Περίληψη

Στην παρούσα διπλωματική εργασία γίνεται μια ανασκόπηση της βιβλιογραφίας σχετικά με την ανάκτηση της εικόνας από το ξεκίνημα των συστημάτων ανάκτησης μέχρι σήμερα. Πλέον η πληροφορία που διανέμεται μέσω του διαδικτύου είναι τεράστια και ένα μεγάλο μέρος της εργασίας αναλύει τα μέσα τα οποία υπάρχουν για την ταξινόμηση της πολυμεσικής πληροφορίας και συγκεκριμένα της εικόνας.

Επίσης γίνεται μια παρουσίαση του γενικού προτύπου για τα πολυμέσα MPEG-7 και μελετούνται κάποιες περιπτώσεις ιστοσελίδων μηχανών αναζήτησης εικόνων.

Λέξεις Κλειδιά

ανάκτηση εικόνας με βάση το περιεχόμενο, MPEG-7, οπτικός θησαυρός, περιγραφείς, σημασιολογικό κενό, ανάκτηση εικόνων με βάση χαρακτηριστικά περιοχών, ομαδοποίηση, alipr, yahoo image, bing image, google image

ΕΥΧΑΡΙΣΤΙΕΣ

Θα ήθελα να ευχαριστήσω όλους αυτούς που με βοήθησαν κατά την διάρκεια της συμμετοχής μου ως φοιτητής στο μεταπτυχιακό τμήμα σπουδών της εφαρμοσμένης πληροφορικής και ιδιαίτερος τον καθηγητή κ. Μανισάρη Αθανάσιο που με βοήθησε ούτως ώστε να μπορέσω να ολοκληρώσω τις σπουδές μου στο τμήμα αυτό εμπιστευόμενός μου μία διπλωματική εργασία.

Θα ήθελα επίσης να ευχαριστώ τους γονείς μου που με βοήθησαν όλον αυτόν τον καιρό και οικονομικά και πνευματικά καθώς επίσης και την κοπέλα μου για την ανοχή που έδειξε και την ηρεμία που μου προσέδωσε για να μπορέσω να ολοκληρώσω τις σπουδές μου.

ΠΕΡΙΕΧΟΜΕΝΑ

ΕΙΣΑΓΩΓΗ.....	1
ΚΕΦΑΛΑΙΟ 1Ο - ΠΛΗΡΟΦΟΡΙΑ.....	1
ΔΟΜΗΜΕΝΕΣ ΒΑΣΕΙΣ ΔΕΔΟΜΕΝΩΝ.....	1
ΑΔΟΜΗΤΕΣ ΒΑΣΕΙΣ ΔΕΔΟΜΕΝΩΝ.....	1
ΑΝΑΚΤΗΣΗ ΠΛΗΡΟΦΟΡΙΑΣ.....	2
ΑΝΑΚΤΗΣΗ ΠΟΛΥΜΕΣΙΚΗΣ ΠΛΗΡΟΦΟΡΙΑΣ.....	2
ΥΠΟΛΟΓΙΣΤΗΣ ΚΑΙ ΑΝΘΡΩΠΟΣ.....	3
ΑΝΑΚΤΗΣΗ ΕΙΚΟΝΑΣ.....	3
ΚΕΦΑΛΑΙΟ 2Ο - ΑΝΑΚΤΗΣΗ ΕΙΚΟΝΑΣ.....	4
ΙΣΤΟΡΙΚΗ ΑΝΑΔΡΟΜΗ.....	4
ΕΙΚΟΣΤΟΣ ΠΡΩΤΟΣ ΑΙΩΝΑΣ.....	5
ΣΧΕΤΙΚΗ ΑΝΑΤΡΟΦΟΛΟΤΗΣΗ.....	6
ΚΕΦΑΛΑΙΟ 3Ο - ΤΕΧΝΙΚΕΣ ΣΥΣΤΗΜΑΤΩΝ ΑΝΑΚΤΗΣΗΣ ΕΙΚΟΝΑΣ.....	8
ΓΕΝΙΚΗ ΔΟΜΗ.....	8
ΤΕΧΝΙΚΕΣ ΣΥΝΤΑΙΡΙΑΣΜΑΤΟΣ.....	10
ΤΕΧΝΙΚΕΣ ΔΕΙΚΤΟΛΟΤΗΣΗΣ.....	11
ΤΕΧΝΙΚΕΣ ΧΡΗΣΗΣ ΛΕΞΙΚΟΥ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ.....	12
ΤΥΠΟΙ ΕΡΩΤΗΜΑΤΩΝ.....	14
ΟΠΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ.....	16
ΚΕΦΑΛΑΙΟ 4Ο - ΑΝΑΛΥΣΗ ΕΙΚΟΝΑΣ.....	18
ΚΑΤΗΓΟΡΙΟΠΟΙΗΣΗ ΩΣ ΠΡΟΣ ΤΗΝ ΠΛΗΡΟΦΟΡΙΑ.....	18
ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ.....	19
ΧΡΩΜΑ.....	20
ΙΣΤΟΡΙΚΟ.....	20
ΜΟΝΤΕΛΑ.....	21
ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΕΙΚΟΝΩΝ.....	22
ΚΑΤΑΤΜΗΣΗ ΕΙΚΟΝΑΣ.....	22

ΚΑΤΑΤΜΗΣΗ JSEG.....	23
ΚΑΤΑΤΜΗΣΗ BLOBWORLD	24
ΆΛΛΟΙ ΑΛΓΟΡΙΘΜΟΙ.....	24
ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΕΙΚΟΝΩΝ ΧΑΜΗΛΟΥ ΕΠΙΠΕΔΟΥ.....	24
ΧΑΡΑΚΤΗΡΙΣΤΙΚΟ ΧΡΩΜΑΤΟΣ.....	24
ΧΡΩΜΑΤΙΚΑ ΜΟΝΤΕΛΑ Ή ΣΥΣΤΗΜΑΤΑ.....	25
ΧΡΩΜΑΤΙΚΟ ΣΥΣΤΗΜΑ.....	26
RGB ΧΡΩΜΑΤΙΚΟ ΣΥΣΤΗΜΑ.....	26
CMYK.....	28
ΣΧΗΜΑ.....	29
ΥΦΗ (TEXTURE).....	30
ΧΩΡΙΚΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ.....	31
ΚΕΦΑΛΑΙΟ 50 - ΤΟ ΠΡΟΤΥΠΟ MPEG7.....	32
MPEG-7: ΕΙΣΑΓΩΓΗ.....	32
MPEG-1.....	33
MPEG-2.....	33
MPEG-4.....	33
ΧΩΡΟΣ ΕΝΔΙΑΦΕΡΟΝΤΩΝ ΤΟΥ MPEG-7.....	34
ΠΡΟΤΥΠΟΠΟΙΗΣΗ ΠΟΛΥΜΕΣΩΝ.....	35
ΕΦΑΡΜΟΓΕΣ	36
MPEG-7 ΚΑΙ ΒΑΣΕΙΣ ΔΕΔΟΜΕΝΩΝ.....	38
ΤΙ ΠΕΡΙΛΑΜΒΑΝΟΥΝ ΟΙ ΠΡΟΔΙΑΓΡΑΦΕΣ.....	39
ΕΦΑΡΜΟΓΗ ΉΧΟΥ.....	40
ΩΘΗΣΗ ΑΠΩΘΗΣΗ.....	41
ΑΠΑΙΤΗΣΕΙΣ ΕΦΑΡΜΟΓΩΝ.....	42
Η ΠΕΡΙΓΡΑΦΗ ΤΟΥ MPEG-7.....	43
ΣΤΟΧΟΙ ΤΟΥ MPEG-7:.....	43
ΠΕΡΙΓΡΑΦΗ ΠΟΛΥΜΕΣΙΚΟΥ ΠΕΡΙΕΧΟΜΕΝΟΥ (MULTIMEDIA CONTENT DESCRIPTION).....	43
ΠΕΡΙΓΡΑΦΕΙΣ.....	44
ΡΥΘΜΙΣΤΙΚΟΙ ΚΑΝΟΝΕΣ.....	46
ΕΥΕΛΙΞΙΑ ΔΙΑΧΕΙΡΙΣΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ (FLEXIBILITY IN DATA MANAGEMENT).....	47
Η ΚΑΘΟΛΙΚΟΤΗΤΑ ΤΩΝ ΠΗΓΩΝ ΤΩΝ ΔΕΔΟΜΕΝΩΝ (GLOBALIZATION OF DATA RESOURCES).....	48
ΟΡΟΛΟΓΙΑ ΤΟΥ MPEG-7.....	49
ΔΕΔΟΜΕΝΑ (DATA).....	49
ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ (FEATURES).....	49

ΠΕΡΙΓΡΑΦΕΙΣ ΚΑΙ ΤΙΜΕΣ ΠΕΡΙΓΡΑΦΕΩΝ (DESCRIPTORS AND DESCRIPTORS VALUES).....	50
ΣΧΗΜΑΤΑ ΠΕΡΙΓΡΑΦΕΩΝ (DESCRIPTION SCHEMES).....	51
ΠΕΡΙΓΡΑΦΗ.....	53
ΚΩΔΙΚΟΠΟΙΗΜΕΝΗ ΠΕΡΙΓΡΑΦΗ (CODED DESCRIPTION).....	53
ΓΛΩΣΣΑ ΠΡΟΣΔΙΟΡΙΣΜΟΥ ΤΗΣ ΠΕΡΙΓΡΑΦΗΣ (DESCRIPTION DEFINITION LANGUAGE, DDL).....	53
ΣΥΣΤΗΜΑΤΑ	54
Η ΔΟΜΗ ΤΟΥ ΜΡΕG-7.....	54
ΕΡΓΑΛΕΙΑ ΠΕΡΙΓΡΑΦΗΣ ΕΙΚΟΝΩΝ.....	57
ΠΕΡΙΓΡΑΦΕΑΣ ΧΡΩΜΑΤΟΣ.....	57
ΠΕΡΙΓΡΑΦΕΑΣ ΣΧΗΜΑΤΟΣ.....	58
ΚΕΦΑΛΑΙΟ 60 - ΜΕΛΕΤΗ ΠΕΡΙΠΤΩΣΕΩΝ.....	59
ΑΝΑΚΤΗΣΗ ΜΕ ΒΑΣΗ ΤΟ ΚΕΙΜΕΝΟ.....	60
GOOGLE, YAHOO, BING.....	60
ΑΝΑΚΤΗΣΗ ΕΙΚΟΝΑΣ ΜΕ ΒΑΣΗ ΤΟ ΠΕΡΙΕΧΟΜΕΝΟ.....	63
ALIPR.....	63
ΧΡΩΜΑ.....	63
ΥΦΗ.....	64
ΣΤΗΝ ΠΡΑΞΗ.....	65
ΕΠΙΛΟΓΟΣ.....	69
ΒΙΒΛΙΟΓΡΑΦΙΑ.....	70

Εισαγωγή

Η τεχνολογική ανάπτυξη που παρατηρείται στην σύγχρονη εποχή έχει οδηγήσει στην δημιουργία, αποθήκευση και διακίνηση τεράστιου όγκου δεδομένων. Η επανάσταση του διαδικτύου έχει οξυνθεί το πρόβλημα της πρόσβασης και διαθεσιμότητας όλης αυτής της πληροφορίας. Οι έρευνες που διεξάγονται πλέον για την διαχείριση της πληροφορίας είναι τεράστιες παρ' όλ' αυτά τα αποτελέσματα των ερευνών αυτών δύσκολα καταφέρνουν να περιγράψουν με ευκολία την παρεχόμενη και συνεχή πληροφορία.

Οξύτερο είναι το πρόβλημα σήμερα της λεγόμενης πολυμεσικής πληροφορίας. Με τον όρο πολυμεσική πληροφορία αναφερόμαστε σε οποιοδήποτε δημιούργημα περιλαμβάνει εικόνα, ήχο και κινούμενη εικόνα ή συνδυασμό όλων αυτών.

Η εικόνα σήμερα, περισσότερο από κάθε άλλη εποχή αποτελεί ένα βασικό μέσο επικοινωνίας. Μπορεί να αποδώσει με τον καλύτερο δυνατό τρόπο νοήματα και ιδέες χωρίς τα προβλήματα των γλωσσικών περιορισμών.

Πολλοί τομείς της ανθρώπινης δραστηριότητας βασίζονται στην εικόνα και στις δυνατότητες που αυτή παρέχει. Η ιατρική, η πληροφορική, η δημοσιογραφία, είναι μερικοί μόνο από τους πολλούς τομείς στους οποίους η εικόνα παίζει πολύ σημαντικό ρόλο και βοηθάει στην αντιμετώπιση ή στην επίλυση προβλημάτων. Πέρα, όμως, από σημαντικός ο ρόλος της εικόνας είναι και ιδιαίτερα ψυχαγωγικός. Αυτό κάνει την εικόνα τον σημαντικότερο ίσως παίκτη της πολυμεσικής πληροφορίας.

Το μεγαλύτερο πρόβλημα που χρήζει αντιμετώπισης στην εποχή μας είναι σίγουρα η ανάκτηση της πολυμεσικής πληροφορίας από συστήματα βάσεων δεδομένων. Οι παλιές κατηγοριοποιήσεις με την χρήση απλών συστημάτων βάσεων δεδομένων ή απλά με την οπτική ή την αναγνώριση με βάση το κείμενο που χρησιμοποιούνταν έχουν πλέον απαρχαιωθεί και δεν προσφέρουν κανενός είδους διαχείρισης κι εύρεσης της κατάλληλης πληροφορίας λόγω του τεραστίου όγκου τον οποίο βρέθηκαν να διαχειρίζονται. Από παράδειγμα ενός τέτοιου απαρχαιωμένου συστήματος θα μπορούσε να είναι το σύστημα μια βάσης δεδομένων με φωτογραφίες

διαφόρων ασθενειών με μερικές εκατοντάδες μόνο από αυτές που στην σημερινή εποχή και με την βοήθεια των σύγχρονων φωτογραφικών μέσων θα πρέπει να διαχειρίζεται χιλιάδες ή ακόμα και εκατομμύρια τέτοιου είδους φωτογραφίες.

Στην παρούσα εργασία γίνεται μια σε βάθος ανάλυση των χαρακτηριστικών που απαρτίζουν μια εικόνα καθώς επίσης και πως μπορεί αυτή η πληροφορία να γίνει διαχειρίσιμη. Παράλληλα γίνεται μια παρουσίαση των συστημάτων ανάκτησης δεδομένων που υπάρχουν σήμερα καθώς και του προτύπου MPEG7

ΚΕΦΑΛΑΙΟ 1^ο - ΠΛΗΡΟΦΟΡΙΑ

Με τον όρο πληροφορία, χαρακτηρίζουμε οποιοδήποτε είδος αλφαριθμητικά όπως για παράδειγμα ένα όνομα ή έναν αριθμό τηλεφώνου μέχρι την πολυμεσική πληροφορία δηλαδή τις εικόνες, τους ήχους και τις κινούμενες εικόνες.

Υπάρχουν δύο κύριες κατηγορίες τέτοιων βάσεων δεδομένων που χαρακτηρίζονται από τον τρόπο με τον οποίο δομούνται. Οι δύο αυτές κατηγορίες είναι οι Δομημένες Βάσεις Δεδομένων και οι Αδόμητες Βάσεις Δεδομένων.

Δομημένες Βάσεις Δεδομένων

Στις δομημένες βάσεις δεδομένων η ανάκτηση οποιασδήποτε πληροφορίας γίνεται με γλώσσες προγραμματισμού 5ης γενιάς όπως είναι η SQL. Με την γλώσσα SQL μπορούμε να δημιουργήσουμε ερωτήματα έτσι ώστε να εξάγουμε τις πληροφορίες που χρειαζόμαστε. Τα ερωτήματα είναι μικρά κομμάτια κώδικα, με τα οποία ο χρήστης περιγράφει αυτό που θέλει να εξάγει από μία βάση δεδομένων με συνοπτικό κώδικα. Ο κώδικας αυτός συνήθως είναι πολύ μικρός παρ' όλ' αυτά περιγράφει με πολύ συγκεκριμένο και ακριβή τρόπο την πληροφορία που θέλουμε να εξάγουμε.

Αδόμητες Βάσεις Δεδομένων

Οι αδόμητες ή μη-δομημένες βάσεις δεδομένων όπως αναφέρονται συχνά στην βιβλιογραφία είναι αυτές οι οποίες δεν ανήκουν στην πρώτη κατηγορία. Με απλά λόγια είναι οι βάσεις δεδομένων όπου δεν υπάρχει συνοχή μεταξύ των δεδομένων τα οποία αποθηκεύονται σε αυτές οπότε είναι και τρομακτικά δύσκολο το να τεθεί με κάποιον συγκεκριμένο τρόπο ένα ερώτημα για να εξαχθεί η κατάλληλη πληροφορία. Το μεγαλύτερο και παράλληλα διασημότερο αδόμητο σύνολο δεδομένων είναι το Διαδίκτυο όπου επικρατεί γενικότερα ένα είδους αναρχίας και γι' αυτό τον λόγο η ανάκτηση της πληροφορίας από ένα τέτοιο σύστημα είναι πολύ δύσκολη.

Ανάκτηση Πληροφορίας

Με τον όρο Ανάκτηση πληροφορίας ονομάζουμε την εξαγωγή μιας πληροφορίας, από μια βάση δεδομένων. Η εξόρυξη της πληροφορίας από μία βάση δεδομένων μπορεί να γίνει με διάφορους τρόπους αναλόγως με το είδος της βάσης δεδομένων που πραγματευόμαστε

Στην παρούσα έρευνα θα γίνει μία ανάλυση της εξαγωγής πληροφορίας από μία πολυμεσική βάση δεδομένων και πιο συγκεκριμένα η έρευνά μας θα επικεντρωθεί στην εύρεση της κατάλληλης οπτικής πληροφορίας με την χρήση εικόνων.

Ανάκτηση πολυμεσικής πληροφορίας

Η ανάκτηση πολυμεσικής πληροφορίας (Multimedia information retrieval) αποτελεί επιστημονική έρευνα απ' όλες τις απόψεις. Το ερώτημα που τίθεται πολύ συχνά είναι το εξής: γιατί είναι καλή η γνώση εάν αυτή δεν είναι δυνατόν να εντοπιστεί με εύκολο τρόπο; Έτσι, λοιπόν, η πρόκληση είναι μεγάλη όπως ακριβώς πρέπει « να γίνει καθημερινή ευκολία η αποτύπωση, αποθήκευση, εύρεση και χρήση πολυμέσων στο υπολογιστικό περιβάλλον» [Rowe and Jain 2005] ¹.

Στην πραγματικότητα στη σημερινή εποχή το βασικότερο πρόβλημα είναι πως μπορούμε να ενεργοποιήσουμε και να βελτιώσουμε την Ανάκτηση Πολυμεσικής Πληροφορίας. Μέθοδοι και υλοποιήσεις του υπάρχουν πολλές και είναι απαραίτητες όταν το κείμενο απουσιάζει παντελώς. Ακόμη όμως κι όταν το κείμενο είναι παρών οι μέθοδοι αυτοί θα πρέπει να βελτιστοποιηθούν έτσι ώστε να είναι ακόμη τελειότερη και ακριβής η εύρεση ενός πολυμέσου από μια μεγάλη συλλογή.

¹ Rowe, L.A. And Jain, R. 2005. ACM SIGMM retreat report on future directions in multimedia research. *ACM Transactions on Multimedia Computing, Communications, and Application 1(1)*, 3-13.

Υπολογιστής και άνθρωπος

Η διαφορά μεταξύ ανθρώπου και υπολογιστή βγάζει νικητή τον άνθρωπο στον τομέα της αναγνώρισης των πολυμέσων. Για να μπορέσει ένας υπολογιστής να καταλάβει ακόμη και την παραμικρή απεικόνιση ενός δέντρου σε ένα βίντεο θα πρέπει να χρησιμοποιήσει τεράστιους αλγόριθμους και πολύ καλές υλοποιήσεις αυτών, εν αντιθέσει με τον άνθρωπο όπου η αναγνώριση ενός δέντρου είναι πάρα πολύ εύκολη υπόθεση.

Η ανάλυση ενός οπτικού ή ηχητικού ερεθίσματος από τον ανθρώπινο εγκέφαλο είναι μία διεργασία η στην οποία υπόκειται από την ημέρα της υπάρξής του κάτι το οποίο δεν μπορεί να ενεργοποιήσει με απλό τρόπο ένας υπολογιστής.

Ανάκτηση εικόνας

Η διάδοση της ψηφιακής φωτογραφικής μηχανής έφερε επανάσταση στον τομέα της φωτογραφίας. Με αυτόν τον τρόπο πλέον ο καθένας μπορεί να απεικονίζει και να μοιράζεται μέσω των υπάρχοντων υποδομών τις φωτογραφίες του με εύκολο και απλό τρόπο. Έτσι έχουμε φτάσει στο σημείο να δημιουργούνται στο διαδίκτυο τεράστιες συλλογές οι οποίες αυξάνονται συνεχώς και σε όγκο αλλά και σε πλήθος. Σε αυτό το σημείο έρχεται η έρευνα αυτή για την ανάκτηση πληροφοριών να μας δώσει κάποια στοιχεία για το πώς μπορούν όλες αυτές οι πληροφορίες να αποθηκευτούν αλλά και να εξερευνηθούν μέσω στο χαστικό διαδίκτυο.

ΚΕΦΑΛΑΙΟ 2^ο - ΑΝΑΚΤΗΣΗ ΕΙΚΟΝΑΣ

Ιστορική αναδρομή

Η έρευνα για ψηφιακή γνώση ξεκίνησε πριν από πολλές δεκαετίες όταν υπήρξε η ιδέα για ψηφιοποίηση σε όλους τους τομείς της επιστήμης για αρχή και αργότερα και σε όλους τους τομείς της κοινωνίας. Πριν από αυτήν την εποχή η ανάκτηση εικόνας υπήρχε μόνο σε θεωρητικές συζητήσεις μεταξύ των επιστημόνων. Από την θεωρητική πλευρά, περιοχές όπως η τεχνητή νοημοσύνη, θεωρίες βελτιστοποίησης, οπτικοποίηση των υπολογιστών και αναγνώριση υποδειγμάτων βοήθησαν σημαντικά στην δημιουργία του μαθηματικού υποβάθρου στον τομέα της ανάκτησης εικόνας. Η ψυχολογία και σχετικές με αυτήν περιοχές όπως η αισθητική και η εργονομία παρέχουν τις βασικές αρχές για την αλληλεπίδραση μεταξύ των χρηστών. Επιπρόσθετα, εφαρμογές για αναζήτηση εικόνων σε μια βάση με εικόνες σε χαρτί προϋπήρχαν σε απλοϊκές μορφές για την αναγνώριση προσώπων και χαρακτήρων γραφής.

Στα πρώιμα χρόνια της Ανάκτησης Εικόνας (ΑνΕικ) έχουν γραφεί διάφορα βιβλία για την επιτυχία τέτοιων συστημάτων με την χρήση ηλεκτρονικού υπολογιστή. Τα βιβλία των Ballard και Brown (1982) [1], Levine (1985) [2], και οι Haralick και Shapiro (1993) [3], εστιάζουν στους αλγόριθμους που χρησιμοποιούνται για την ΑνΕικ με βάση τις ομοιότητες μεταξύ εικόνων. Υλοποιήσεις των παραπάνω αλγορίθμων μπορούν να θεωρηθούν συστήματα όπως το QBIC, Flickner et al. (1995) [4] και το Virage, Bach, et al. (1996) [5] τα οποία υλοποιήθηκαν στα μέσα τις δεκαετίες του 1990. Μέσα σε λίγα χρόνια η βασική ιδέα της ομοιότητας μεταξύ εικόνων μεταφέρθηκε σε διάφορους δικτυακούς τόπους μηχανών αναζήτησης εικόνων. Έτσι έχουμε υλοποιήσεις όπως το Webseek, Smith και Chang (1997) [6] και το Webseer, Frankel, et al. (1996) [7]. Μεγάλη προσπάθεια έγινε επίσης και στον τομέα των επιχειρήσεων όπου η αναζήτηση με βάση τις ομοιότητες εικόνων ήταν η βάση για την δημιουργία των βάσεων δεδομένων της Informix, τις επεκτάσεις της ευρέως γνωστής βάσης δεδομένων IBM DB2 καθώς επίσης και της βάσης

δεδομένων Oracle (Oracle Cartridges όπως έμεινε γνωστό) Bliujute, et al. (1999) [8], Egas, et al. (1999) [9] έτσι ώστε να μπορεί να γίνει ευκολότερη η χρήση των εικόνων στον τομέα των επιχειρήσεων.

Εικοστός πρώτος αιώνας

Ξεκινώντας από τις αρχές του εικοστού πρώτου αιώνα οι ερευνητές παρατήρησαν ότι οι χρήση αλγορίθμων εύρεσης ομοιοτήτων είχε περισσότερο χαρακτήρα διαίσθησης παρά ήταν φιλική προς τον χρήστη όπως περίμεναν.

Θα μπορούσε κανείς να παρατηρήσει ότι αυτά τα συστήματα τα οποία είχαν δημιουργηθεί από επιστήμονες μέσω ερευνητικών προγραμμάτων μπορούσαν να χρησιμοποιηθούν αποτελεσματικά μόνο από τους ίδιους τους δημιουργούς τους. Η καινούργια φιλοσοφία ήταν να δημιουργηθούν συστήματα φιλικά προς τον χρήστη που θα μπορούσαν να παρέχουν την πληροφορία από βάσεις δεδομένων και συλλογές σε όλων τον κόσμο. Για την υλοποίηση όμως τέτοιων συστημάτων τα συστήματα επόμενης γενιάς θα έπρεπε να καταλάβουν την σημασιολογία των ερωτημάτων όχι με την έννοια της σε χαμηλό επίπεδο λειτουργίας του υπολογιστή. Το πρόβλημα αυτό ονομάστηκε «Γεφυροποίηση των Σημασιολογικών κενών». Από την πλευρά της αναγνώρισης υποδειγμάτων αυτό σήμαινε την μετάφραση των χαμηλού επιπέδου οπτικών περιεχομένων μιας εικόνας σε υψηλού επιπέδου αρχές ή όρους που θα γίνονταν εύκολα διαχειρίσιμοι από τον χρήστη. Παραδείγματα θεωρητικών υποβάθρων που ασχολήθηκαν με τον συγκεκριμένο τομέα υπάρχουν μόνο για την αναγνώριση προσώπων από τον Rowley και άλλοι (1996) [10] και από τους Lew και Huijismans (1996) [11]. Ίσως το νεότερο σύστημα AvEικ με βάση το περιεχόμενο που αναπτύχθηκε και προσπαθούσε να καλύψει το σημασιολογικό κενό στον τομέα της αναζήτησης με ερωτήματα, της ευρετηριοποίησης και ταξινόμησης ήταν η μηχανή αναζήτησης ImageScape, Lew (2000) [12].

Στην μηχανή αναζήτησης εικόνων ImageScape, ο χρήστης μπορούσε να δημιουργήσει απευθείας ερωτήματα με πληθώρα οπτικών αντικειμένων όπως ουρανός, δέντρα, νερό κτλ, χρησιμοποιώντας εικονίδια τα οποία βρισκονταν

στον δικτυακό τόπο ως βάση της αναζήτησής του που ξεπερνούσαν σε αριθμό τα 10 εκατομμύρια. Το σύστημα χρησιμοποιούσε αλγόριθμους πληροφοριών επέστρεφε τις εικόνες εκείνες που χρησιμοποιούσαν σε μεγαλύτερο βαθμό τα χαρακτηριστικά που ζητούσε ο χρήστης.

Το καθοριστικότερο βήμα για την γεφύρωση του σημασιολογικού κενού στα συστήματα ΑνΕικ έγινε με την χρήση της σχετικής ανατροφοδότησης. Ο χρήστης με αυτόν τον τρόπο κατεύθυνε το σύστημα με ερωτήματα προς τον στόχο της αναζήτησής του.

Σχετική ανατροφοδότηση

Με την δημιουργία των συστημάτων ανάκτησης παρουσιάστηκαν δύο βασικά προβλήματα. Το πρώτο είχε να κάνει με το σημασιολογικό κενό, δηλαδή με την έλλειψη αντιστοιχίας μεταξύ της εξαγόμενης οπτική πληροφορίας και της ερμηνείας αυτής στον χρήστη. Το δεύτερο πρόβλημα ήταν η υποκειμενικότητα της ανθρώπινης αντίληψης όσον αφορά το περιεχόμενο της εικόνας, για παράδειγμα είχε να κάνει με το ενδιαφέρον του χρήστη για ένα χρώμα ή για μια υφή.

Έχοντας αυτά υπόψη, μεταφερθήκαμε από τα απλά ερωτήματα των πρώτων συστημάτων ανάκτησης στην επόμενη γενιά συστημάτων με την οποία ο χρήστης είχε μια συνεχή ανατροφοδότηση των ερωτημάτων έτσι ώστε να επιτύχει το δυνατόν καλύτερα τον σκοπό του και να ανακτήσει την καταλληλότερη εικόνα.. Το σύστημα αυτό ουσιαστικά θέτει τον άνθρωπο στο «κέντρο των εξελίξεων» και ονομάζεται human in the loop.

Η διαδραστική αυτή διαδικασία του χρήστη με το σύστημα με την οποία ο χρήστης κρίνει τα αποτελέσματα της αναζήτησης και ανατροφοδοτεί το σύστημα με περισσότερες λεπτομέρειες ονομάζεται σχετική ανατροφοδότηση (Relevance Feedback)

Ιστορικά η μέθοδος της σχετικής ανατροφοδότησης ορίστηκε στην απλούστερη μορφή της από τον Rocchio, το 1971, ο οποίος πρότεινε την μετακίνηση του σημείου ερωτήματος πιο κοντά σε σχετικά αποτελέσματα και πιο μακριά από τα άσχετα. Από το 1971 μέχρι σήμερα χρησιμοποιώντας αυτό

ως βάση έχουν δημιουργηθεί πολλές μέθοδοι που προσαρμόζονται στην ανάκτηση εικόνων κι έχουν, φυσικά, ξεφύγει από αυτήν την θεωρία που απασχολούσε μόνο ένα σημείο.

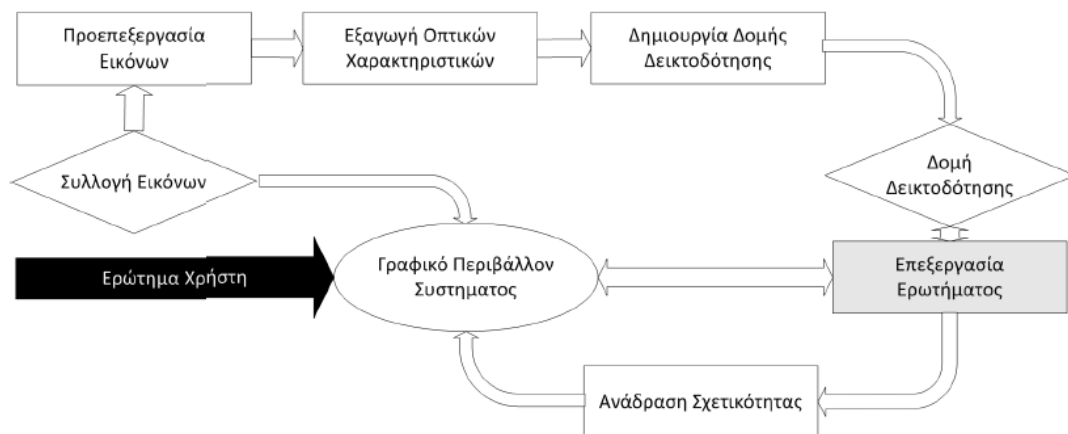
Επίσης ο Sclaroff στο σύστημα ImageRover χρησιμοποιεί αποστάσεις Minkowski για να υλοποιήσει την σχετική ανατροφοδότηση, δίνοντας την δυνατότητα στον χρήστη να επιλέγει συνεχώς της εικόνες εκείνες που είναι σχετικές με το αποτέλεσμα που ζητάει Sclaroff et al. (2001) [13]. Ο Chen χρησιμοποιεί μια διαφορετική προσέγγιση χρησιμοποιώντας μια κλάση που ονομάζεται SVM και σύμφωνα με την οποία γίνεται ουσιαστικά εκμάθηση της ανατροφοδότησης Chen et al. (2001) [14], ενώ οι Tieu και Viola χρησιμοποιούν τον αλγόριθμο εκμάθησης AdaBoost με τον οποίο καταφέρνουν με περιορισμένο αριθμό σετ εκμάθησης να έχουν σχετικά καλά αποτελέσματα Tieu & Viola (2000) [15].

Οι Rui και Huang χρησιμοποιούν ένα πολυμεσικό μοντέλο των αντικειμένων το οποίο περιέχει διαφορετικά χαρακτηριστικά με πολλαπλές αναπαραστάσεις τους, εκ των οποίων η κάθε μία από αυτές φέρει κάποιον συγκεκριμένο ποσοστιαίο συντελεστή ως προς την βαρύτητα που έχει στην συγκεκριμένη αναζήτηση. Χρησιμοποιώντας διαδοχικά ερωτήματα και σύμφωνα με τις επιλογές του χρήστη, οι συγκεκριμένοι συντελεστές μεταβάλλονται, με ταυτόχρονη μεταβολή και των αποτελεσμάτων αναζήτησης Rui et al. (2001) [16].

ΚΕΦΑΛΑΙΟ 3^ο - ΤΕΧΝΙΚΕΣ ΣΥΣΤΗΜΑΤΩΝ ΑΝΑΚΤΗΣΗΣ ΕΙΚΟΝΑΣ

Γενική Δομή

Για να προχωρήσουμε περαιτέρω στην ανάλυση των συστημάτων ΑνΕικ θα πρέπει να δούμε μια γενική δομή ενός τέτοιου συστήματος. Παρακάτω παρουσιάζεται μία απεικόνιση της δομής αυτής παίρνοντας ως βάση το περιεχόμενο της εικόνας καθώς και τα κύρια χαρακτηριστικά που αυτή απεικονίζει.



Μία βάση δεδομένων εικόνων ξεκινάει με την συλλογή των εικόνων που την απαρτίζουν. Ο χρήστης μπορεί να δημιουργεί μια δική του βάση δεδομένων ή να προσθέτει καινούργιες εικόνες σε μία υπάρχουσα βάση στο διαδίκτυο για παράδειγμα. Έτσι η συλλογή των εικόνων συνεχώς αυξάνεται και είναι αναγκαία η χρήση των μεθόδων ΑνΕικ για να μπορέσουμε να εξάγουμε τις σωστές πληροφορίες από την βάση.

Σε μία βάση δεδομένων εικόνων τα περιεχόμενά της μπορούν να υποστούν αρχικά κάποια επεξεργασία έτσι ώστε να μπορέσουμε στη συνέχεια να εξάγουμε με μεγαλύτερη ακρίβεια τα περιεχόμενά των εικόνων που αναζητούμε. Θα πρέπει να διευκρινίσουμε εδώ ότι σε αυτό το στάδιο θα μπορούσε να γίνει και κάποιου είδους φιλτράρισμα ή εξομάλυνση των χαρακτηριστικών των εικόνων με σκοπό την εξάλειψη του θορύβου και γενικότερα την απλοποίηση των εικόνων. Επιπρόσθετα στο συγκεκριμένο

στάδιο μπορεί να γίνει και κατάτμηση των εικόνων σε περιοχές έτσι ώστε να γίνει η εξαγωγή των χαρακτηριστικών της εικόνας σε κομμάτια και να υλοποιηθεί ευκολότερα η διαδικασία αυτή.

Στο επόμενο στάδιο γίνεται προσπάθεια να εξαχθούν τα περιεχόμενα μιας εικόνας είτε ως ολότητα είτε από συγκεκριμένες περιοχές εάν στο παραπάνω στάδιο έχει πραγματοποιηθεί η κατάτμησης της. είναι η εξαγωγή των οπτικών χαρακτηριστικών είτε από ολόκληρη την εικόνα είτε από περιοχές της είτε από κάποια σημεία ενδιαφέροντος.

Στο τρίτο στάδιο γίνεται ουσιαστικά δεικτοδότηση των εικόνων αυτών. Με την έννοια αυτή εννοούμε την παραγωγή ενός είδους υπογραφής ή αποτυπώματος όπως ονομάζεται από πολλούς για να γίνει ξεχωριστή η κάθε εικόνα του συστήματός μας. Η υπογραφή αυτή ουσιαστικά περιγράφει το περιεχόμενο της εικόνας με βάση τα χαρακτηριστικά που έχουν αναγνωριστεί έως τώρα. Οι περιγραφές αυτές όλων των εικόνων πρέπει να αποθηκευτούν αποδοτικά, σε μια βάση δεδομένων για παράδειγμα, ώστε να μπορεί έπειτα το σύστημα να χρησιμοποιήσει όλη αυτή την πληροφορία.

Επιπλέον, μπορεί να γίνει κάποια ομαδοποίηση των περιγραφών από την οποία να προκύψει ένα λεξικό διάφορων χαρακτηριστικών με βάση τα οποία θα χαρακτηριστεί κάθε εικόνα. Παρακάτω θα δούμε τεχνικές δεικτοδότησης και τεχνικές χρήσης του οπτικού θησαυρού.

Ο πυρήνας του συστήματος ανάκτησης ενεργοποιείται με την είσοδο ενός ερωτήματος από τον χρήστη. Εκεί γίνεται το συνταίριασμα των χαρακτηριστικών του ερωτήματος με τα χαρακτηριστικά των εικόνων της συλλογής και ως έξοδο έχουμε κάποιες εικόνες ταξινομημένες με αύξουσα σειρά ως προς κάποιο κριτήριο απόστασης από την εικόνα - ερώτημα. Σε πολλά σύγχρονα συστήματα δίνεται η δυνατότητα στον χρήστη να κρίνει τα αποτελέσματα του ερωτήματος θετικά ή αρνητικά και με αυτόν τον τρόπο έχει το σύστημα ένα είδος ανατροφοδότησης στην οποία έχουμε αναφερθεί παραπάνω.

Τεχνικές συνταιριάσματος

Όσο σημαντικό είναι να έχουμε λειτουργικούς και διακριτούς περιγραφείς των χαρακτηριστικών μιας εικόνας, τόσο σημαντικό είναι να έχουμε και έναν αποτελεσματικό τρόπο για να μετράμε την ομοιότητα μεταξύ δύο εικόνων.

Εκτενής έρευνα και ανάλυση πάνω στην ομοιότητα των χαρακτηριστικών κάνει ο Jolion ο οποίος παρουσιάζει τις πιο διαδεδομένες μεθόδους για τον υπολογισμό των ομοιοτήτων μιας εικόνας Jolion (2001) [17]. Η ομοιότητα μεταξύ δύο διανυσματικών χαρακτηριστικών όπως αναφέρει (Feature Vectors), τα οποία μπορεί να περιέχουν χαρακτηριστικά από μια εικόνα ως ολότητα, μια περιοχή της εικόνας ή μια περιοχή γύρω από κάποιο σημείο ενδιαφέροντος, μπορεί να εκφραστεί ως με τον παρακάτω γενικευμένο τύπο:

$$S_{q;db} = d(F_q; F_{db})$$

όπου F_q και F_{db} είναι δύο διανυσματικά χαρακτηριστικά. Το πρώτο από αυτά αντιπροσωπεύει την εικόνα που έχει θέσει ως ερώτημα ο χρήστης και το δεύτερο αντιπροσωπεύει μία από τις εικόνες της βάσης δεδομένων. Όπου $d(i,j)$ ορίζεται κάποια συνάρτηση απόστασης.

Στην περίπτωση όπου ο χώρος των χαρακτηριστικών είναι διανυσματικός, αποτελεσματικό κριτήριο ομοιότητας είναι η Ευκλείδεια απόσταση. Έρευνες έδειξαν ότι δεν υπάρχει ένα γενικά αποτελεσματικότερο κριτήριο - μέτρο απόστασης, αλλά η επιλογή πρέπει να γίνει σε σχέση με την συλλογή εικόνων που διαθέτουμε και την φύση των περιγραφικών διανυσμάτων που θέλουμε να συγκρίνουμε Sebe et al. (2000) [18].

Για να μπορέσουμε να εξάγουμε συμπέρασμα για την απόσταση μεταξύ των ιστογραμμάτων δύο εικόνων έχουν προταθεί διάφορες μέθοδοι που συγκρίνουν τις σχετικές κορυφές των ιστογραμμάτων όπως η απόσταση Minkowski. Προτάθηκε επίσης η χρήση αθροιστικών ιστογραμμάτων καθώς υποστηρίζεται ότι τα αποτελέσματά τους δεν έχουν μεγάλη ευαισθησία στον θόρυβο Stricker & Orengo (1995) [19].

Ο πιο γνωστός και πολυχρησιμοποιημένος τρόπος χειρισμού των ιστογραμμάτων είναι η μετατροπή τους σε διάνυσμα. Έτσι, ένα ιστόγραμμα με

k κορυφές, αναπαρίστανται με ένα k-διάστατο διάνυσμα όπως φαίνεται παρακάτω

$[f_1; f_2, \dots, f_k]$, όπου f η τιμή της κάθε κορυφής

Βασικό μειονέκτημα αυτής της μεθόδου είναι ότι με αυτόν τον τρόπο αγνοείται η πληροφορία της θέσης από όπου εξήχθησαν οι κορυφές του ιστογράμματος [Rubner et al., 1998]. Αυτό οδήγησε στο να προταθεί το 1998 από τους Rubner και Tomasi η χρήση της Απόστασης των Περιγητών στην Γη (Earth Movers Distance - EMD), για την χρήση της οποίας αναπαρίστανται τα ιστογράμματα σαν διανύσματα που περιέχουν ζεύγη τιμών και χωρικής πληροφορίας ταυτόχρονα,

$[(z_1; f_1); (z_2; f_2), \dots, (z_k; f_k)]$, όπου f είναι η τιμή της κάθε κορυφής και z η θέση από το κέντρο της κορυφής Rubner et al., (2000) [20] .

Μία ακόμη πρόταση για την χρήση της απόστασης είναι η απόσταση Mahalanobis, η οποία χρησιμοποιήθηκε κυρίως σε συστήματα ανάκτησης κατά περιοχές όπως το Blobworld για να συγκρίνει πιο σύνθετες δομές (Blobs) που περιέχουν χαρακτηριστικά χρώματος, υφής και χωρικής κατανομής μαζί Carson et al., (1997), [21], Carson et al. (2002) [22].

Τεχνικές δεικτοδότησης

Στα πρώτα συστήματα ανάκτησης οι εξαγόμενες από τις εικόνες πληροφορίες σώζονταν απλά σε αρχεία ή αποτελούσαν εγγραφές σε κάποια βάση δεδομένων. Η απόδοση και των δύο μεθόδων, από την σκοπιά της υπολογιστικής αποτελεσματικότητας, έπεφτε με τον καιρό καθώς το μέγεθος των συλλογών αυξανόταν με γεωμετρική πρόοδο. Επίσης, τα διανυσματικά χαρακτηριστικά που εξάγονται από τις εικόνες της συλλογής έχουν συνήθως μεγάλη διάσταση. Αυτό είχε ως αποτέλεσμα, κυρίως σε μεγάλες συλλογές, να είναι πλέον απαραίτητη η χρήση κάποιας έξυπνης μορφής δεικτοδότησης ή και κάποιας μορφής ομαδοποίησης της πληροφορίας.

Αρκετές φορές προτού εφαρμοστεί κάποια μέθοδος δεικτοδότησης, δοκιμάζονταν κάποια τεχνική για να ελαττωθεί η διάσταση των διανυσμάτων,

όπως για παράδειγμα κάποια μορφή της Ανάλυσης Πρωτευόντων Συστατικών (Principal Component Analysis - PCA), Ng & Sedighian, (1996) [23].

Ακόμη, τα ευρετήρια που χρησιμοποιούνται πλέον για τον υπολογισμό της ομοιότητας περιέχουν δομές δέντρων για να επιτύχουν λογαριθμική πολυπλοκότητα. Δομές όπως τα k-d δέντρα, τα k-d δέντρα προτεραιότητας, τα R-δέντρα χρησιμοποιήθηκαν και χρησιμοποιούνται σε διάφορες φάσεις της διαδικασίας ανάκτησης Egas et al., (1999) [24], Scott & Shyu, (2003) [25].

Άλλες μέθοδοι που προτάθηκαν περιλαμβάνουν τον κβαντισμό των διανυσμάτων (vector quantization) Ye & Xu (2003) [26] ή την χρήση δύο ειδών υπογραφών για κάθε εικόνα, με το ένα είδος να περιγράφει τα αντικείμενα που περιέχει η εικόνα και το άλλο είδος να περιγράφει την χωρική κατανομή στην εικόνα El-Kwae & Kabuka (2000) [27]. Τέλος πολλές είναι οι τεχνικές που χρησιμοποιούν κάποιας μορφής hashing, δηλαδή προβολή - απεικόνιση των διανυσμάτων των εικόνων σε άλλους χώρους μικρότερης διάστασης Greene et al. (1994) [28].

Τεχνικές χρήσης λεξικού χαρακτηριστικών

Η εξαγωγή χαρακτηριστικών από περιοχές εικόνων ή σημεία ενδιαφέροντος και όχι από ολόκληρη την εικόνα αποδίδει έναν τεράστιο όγκο της εξαγόμενης πληροφορίας. Για παράδειγμα από 1000 μόνο εικόνες μπορεί να πάρουμε 60000 σημεία ενδιαφέροντος συνολικά, με το κάθε ένα να αναπαρίσταται από ένα διάνυσμα χαρακτηριστικών. Έτσι όταν το μέγεθος της συλλογής αυξάνει στην τάξη του 10^5 , οι διαδικασίες που εκτελούνται κατ' επανάληψη στην ανάκτηση όπως η αναζήτηση και εξαγωγή αποστάσεων, ακόμα και με έναν αποδοτικό τρόπο δεικτοδότησης είναι ιδιαίτερα χρονοβόρες.

Ως λύση αυτού του προβλήματος έχει προταθεί η ομαδοποίηση (clustering) των διανυσμάτων σε ομάδες (clusters). Κάθε ομάδα αντιπροσωπεύεται συνήθως από το κέντρο της. Μέθοδοι ομαδοποίησης υπάρχουν πολλές με πλέον διαδεδομένη την k-means. Χαρακτηριστικό της μεθόδου αυτής είναι ότι ο αριθμός των ομάδων πρέπει να καθοριστεί εκ των προτέρων. Άλλες μέθοδοι που έχουν χρησιμοποιηθεί είναι οι kernel mapping, hierarchical clustering και metric learning. Οι μέθοδοι ομαδοποίησης συνεχίζουν να

εξελίσσονται μέχρι σήμερα και να βρίσκουν εφαρμογή σε συστήματα ανάκτησης εικόνων και αντικειμένων Philbin et al. (2007) [29].

Στην περίπτωση ανάκτησης αντικειμένων, απαιτείται και κατηγοριοποίηση περιοχών των εικόνων ανάλογα με τα αντικείμενα που περιέχουν. Αυτό προϋποθέτει να έχει προηγηθεί πριν τη χρήση του συστήματος μια φάση εκπαίδευσης. Πολλές μέθοδοι ομαδοποίησης και εκπαίδευσης αναλύονται στο βιβλίο *Elements of statistical learning* Hastie et al. (2001) [30].

Στην περίπτωση που εξάγονται χαρακτηριστικά όπως το χρώμα και η υφή από περιοχές της εικόνας (που συνήθως προέρχονται από κατάτμηση) και εκτελεστεί κάποιος αλγόριθμος ομαδοποίησης στα διανυσματικά χαρακτηριστικά, θα πάρουμε κάποιους τύπους περιοχών, κάθε έναν από τους οποίους θα χαρακτηρίζει το κέντρο της ομάδας. Μπορεί, έτσι, να αναπαρασταθεί η κάθε εικόνα με ένα διάνυσμα που θα περιέχει τις ελάχιστες αποστάσεις των περιοχών της από τον κάθε τύπο περιοχής Spyrou et al. (2008) [31].

Στην περίπτωση που τα διανύσματα χαρακτηριστικών προέρχονται από περιοχές γύρω από σημεία ενδιαφέροντος της εικόνας, με την εφαρμογή κάποιου είδους ομαδοποίησης μπορεί να παραχθεί ένα λεξικό χαρακτηριστικών μικρών περιοχών (*visual vocabulary*). Κάθε στοιχείο του λεξικού, δηλαδή κάθε μια από τις παραγόμενες ομάδες, θα εκφράζει, σε περίπτωση που χρησιμοποιηθούν περιγραφείς τύπου SIFT, μια περιοχή ως προς τις κατευθύνσεις των παραγώγων των φωτεινότητων μέσα σε αυτή.

Στο σύστημα ανάκτησης αντικειμένων μέσα από καρέ ταινιών που χρησιμοποιείται στην ιστοσελίδα «Video Google», οι Sivic και Zisserman χρησιμοποιούν τον αλγόριθμο *k-means* για ομαδοποίηση και δημιουργία λεξικού χαρακτηριστικών. Στη συνέχεια για την αναπαράσταση των καρέ, επιστρατεύουν τεχνικές συχνότητας εμφάνισης όρων που χρησιμοποιούνται για την ανάκτηση κειμένου από μηχανές αναζήτησης παίρνοντας σαν λέξεις τα στοιχεία του λεξικού χαρακτηριστικών. Τέλος, χρησιμοποιούν μια τεχνική χωρικής ανεκτικότητας σύμφωνα με την οποία ένα συνταίριασμα σημείων είναι σωστό αν στην γύρω περιοχή υπάρχουν κι άλλα στοιχεία που ταιριάζουν Sivic & Zisserman (2003) [32].

Μια ιδιαίτερα πρόσφατη τάση είναι να μην αντιστοιχείται κάθε σημείο μιας εικόνας σε μία μονάχα χαρακτηριστική λέξη. Πρόσφατα προτάθηκαν μεταξύ άλλων τρεις τεχνικές στις οποίες η αντιστοίχιση σημείων και λέξεων χαρακτηριστικών δεν γίνεται ένα προς ένα ντετερμινιστικά, αλλά είτε μέσω πυρήνων γκαουσιανών Gemert et al. (2008) [33], είτε με επιπρόσθετη πληροφορία για την σχετική θέση σημείου- λέξης Jegou et al., (2008) [34], είτε με την αντιστοίχιση παραπάνω από μια λέξεων χαρακτηριστικών σε κάθε σημείο Philbin et al., (2008) [35].

Τύποι ερωτημάτων

Στον χρήστη ενός συστήματος ανάκτησης εικόνων, μπορούν να παρέχονται διάφοροι τρόποι με τους οποίους αυτός θα εκφράσει το ερώτημά του ώστε να πάρει τα επιθυμητά αποτελέσματα. Είναι πολύ χρήσιμο να μπορεί ο χρήστης να διαλέξει μεταξύ διαφόρων μορφών θέσης του ερωτήματος, για να μπορέσει να βρει τον καλύτερο ανάλογα με την περίπτωση. Μερικές μορφές ερωτήματος είναι οι εξής:

- **Ερώτημα με κείμενο:** Ο χρήστης δημιουργεί το ερώτημά του πληκτρολογώντας κάποια λέξη ή λέξεις που σχετίζονται με το περιεχόμενο της εικόνας. Η μέθοδος αυτή ονομάζεται «Ανάκτηση Εικόνας με Βάση το Κείμενο» και είναι η πλέον διαδεδομένη σήμερα. Χρησιμοποιείται από τις πιο γνωστές μηχανές αναζήτησης όπως την Yahoo, την Google και την Bing(Με πλέον εξελιγμένα χαρακτηριστικά από τις υπόλοιπες). Όπως είναι προφανές, στην περίπτωση αυτών των μηχανών αναζήτησης, η αναζήτηση δεν γίνεται με με βάση το περιεχόμενο, αλλά με μετα-δεδομένα τριγύρω από την εικόνα, όπως για παράδειγμα από το κείμενο που την συνοδεύει. Για να γίνει αναζήτηση με βάση το περιεχόμενο από ερώτημα κείμενου, θα πρέπει να προϋπάρχει γνώση που συνδέει τα χαμηλού επιπέδου χαρακτηριστικά της εικόνας με τις υψηλού επιπέδου έννοιες, κάτι το οποίο δεν χρησιμοποιείται από τις μηχανές αυτές.
- **Ερώτημα με εικόνα:** Ο χρήστης είτε ανεβάζει μια δικιά του εικόνα είτε διαλέγει κάποια από την συλλογή και ψάχνει για παρόμοιες. Το

σύστημα πρέπει να αναγνωρίσει το περιεχόμενο της εικόνας του χρήστη και να το συγκρίνει με το αντίστοιχο περιεχόμενο των εικόνων της συλλογής. Αυτή είναι και η πλέον διαδεδομένη μορφή ερωτήματος στην ανάκτηση εικόνων με βάση το περιεχόμενό τους.

- **Ερώτημα με βάση μια περιοχή της εικόνας:** Ο χρήστης είτε ανεβάζει μια δικιά του εικόνα είτε διαλέγει κάποια από την συλλογή και μέσα σε αυτήν επιλέγει την περιοχή που τον ενδιαφέρει. Προϋπόθεση είναι να υπάρχει κάποιο είδος κατάτμησης των εικόνων ώστε να μπορούν να απομονωθούν οι διάφοροι τύποι περιοχών αποδοτικά. Αυτού του είδους τα ερωτήματα φιλοξενούν τα συστήματα ανάκτησης αντικειμένων, καθώς η περιοχή ενδιαφέροντος μπορεί να είναι ένα ζώο ή αντικείμενο, το οποίο πρέπει να απομονωθεί από την υπόλοιπη μη ενδιαφέρουσα εικόνα.
- **Ερώτημα με ζωγραφική:** Στο περιβάλλον αλληλεπίδρασης με τον χρήστη του συστήματος μπορεί να υπάρχει μια περιοχή σχεδίασης στην οποία ο χρήστης σχεδιάζει απλά και χοντροκομμένα το επιθυμητό ερώτημα. Το σύστημα έπειτα, εισάγει στην βάση του το σκίτσο ως εικόνα, με μεθόδους ανάλυσης σχημάτων εξάγει τα περιεχόμενά του και εκτελεί την αναζήτηση.
- **Ερώτημα με σύνθεση περιοχών:** Ο χρήστης μπορεί να φτιάξει ένα προσχέδιο της δομής των εικόνων που αναζητά θέτοντας ποιους τύπους περιοχών θέλει να υπάρχουν και μερικές φορές και σε ποιο μέρος της εικόνας. Για παράδειγμα θα μπορεί να ζητήσει από το σύστημα εικόνες με το φεγγάρι στο πάνω μέρος της εικόνας με ένα δέντρο στ' αριστερά και κάποιο ζώο στο κάτω δεξιά μέρος της εικόνας.
- **Σύνθετο ερώτημα:** Μπορεί να είναι συνδυασμό κάποιων από τις παραπάνω μεθόδους, κυρίως μείξη εικόνας και κειμένου. Για παράδειγμα μια εικόνα που ανεβάζει ο χρήστης και απεικονίζει ένα αυτοκίνητο μπροστά από τον πύργο του Άιφελ, θα μπορεί να συνοδευτεί από τις λέξεις car, citroen και c2 για να μπορέσει ίσως να καταλάβει το σύστημα ότι ο χρήστης δεν ενδιαφέρεται για εικόνες που περιέχουν το γνωστό σιδερένιο μνημείο του Παρισιού, αλλά θέλει

εικόνες που περιέχουν το αυτοκίνητο που δείχνει η εικόνα. Η παραπάνω μορφή ερωτήματος έχει άμεση σχέση με την χρήση της ανατροφοδότησης.

Ως συμπέρασμα της ενότητας αυτής μπορούμε να πούμε ότι υπάρχουν αμέτρητες μορφές ερωτημάτων που η κάθε μία προσαρμόζεται από τον δημιουργό στα συστήματα που τις χρησιμοποιεί. Σε πιο καινούριες δημοσιεύσεις εισάγονται νέες μορφές ερωτημάτων. Για παράδειγμα προτείνεται από τον Assfalg η χρήση ερωτημάτων που περιέχουν τρισδιάστατα μοντέλα, καθώς υποστηρίζει ότι τα δυδιάστατα ερωτήματα δεν περιγράφουν σωστά την χωρική κατανομή των αντικειμένων στην εικόνα Assfalg et al. (2002) [36]. Σε μια άλλη καινοτόμα εφαρμογή, ο Kaster προτείνει την χρήση χειρονομιών και ομιλίας σαν ερωτήματα και σαν ανατροφοδότηση για την ανάκτηση εικόνων Kaster et al. (2003) [37].

Οπτικά αποτελέσματα

Το οπτικό αποτέλεσμα της αναζήτησης είναι ένας από τους πιο σημαντικούς παράγοντες που μπορούν να οδηγήσουν στην αποδοχή και δημοτικότητα ενός συστήματος ανάκτησης. Ο Datta (2008) κατηγοριοποιεί τα συστήματα αυτά με βάση τα οπτικά τους αποτελέσματα ως εξής:

- Ταξινόμηση αποτελεσμάτων με βάση την σχετικότητα ως προς το ερώτημα.
- Χρονολογική ταξινόμηση των αποτελεσμάτων.
- Ομαδοποίηση των οπτικών αποτελεσμάτων.
- Ιεραρχική δομή των οπτικοποιημένων αποτελεσμάτων.
- Σύνθεση όλων των παραπάνω

Η ταξινόμηση των αποτελεσμάτων με βάση την σχετικότητα ως προς το ερώτημα είναι ο πιο συνηθισμένος τρόπος εμφάνισης των αποτελεσμάτων, σύμφωνα με τον οποίο τα αποτελέσματα ταξινομούνται με βάση ένα μέτρο απόστασης από την εικόνα - ερώτημα.

Η χρονολογική ταξινόμηση των αποτελεσμάτων είναι αποδοτική σε εφαρμογές οργάνωσης προσωπικών συλλογών εικόνων.

Η ομαδοποίηση των οπτικών αποτελεσμάτων εφαρμόστηκε από τον Chen στο σύστημα CLUE, Chen et al. (2005) [38].

Η Ιεραρχική δομή των οπτικοποιημένων αποτελεσμάτων εμφανίζει τα αποτελέσματα σε κάποια ιεραρχική δομή όπως για παράδειγμα με μορφή δέντρου.

Μία ακόμη πρόταση για την απεικόνιση των αποτελεσμάτων είναι η τρισδιάστατη απεικόνιση τους η οποία προτείνεται από το σύστημα 3D Mars όπου οι εικόνες εμφανίζονται σε τρισδιάστατο πλέγμα σαν σε εφαρμογή εικονικής πραγματικότητας Nakazato & Huang [39]

Ένα από τα σημαντικότερα δεδομένα για την δημιουργία του γραφικού περιβάλλοντος αλληλεπίδρασης με τον χρήστη ενός συστήματος ανάπτυξης είναι να μελετηθεί και να κατανοηθεί το πώς ο χρήστης συνήθως διαχειρίζεται τις εικόνες που έχει και πως διατυπώνει και περιγράφει αυτό που θέλει να πάρει ως αποτέλεσμα. Έρευνα πάνω στην ανθρώπινη συμπεριφορά ως προς την διάχυση του οπτικού υλικού έχουν διεξάγει οι Rodden και Wood [Rodden & Wood, 2003] και ο Cunningham αναλύει το πώς ο άνθρωπος περιγράφει την ανάγκη του για οπτική πληροφορία Cunningham et al. (2004) [40], Cunningham & Masoodian, (2006) [41].

Τα τελευταία χρόνια περιορισμούς στον τρόπο δημιουργίας των ερωτημάτων θέτουν οι σύγχρονες φορητές υπολογιστικές συσκευές, όπως κινητά τηλέφωνα ή PDA. Τα τελευταία χρόνια οι χρήστες που χρησιμοποιούν φορητές συσκευές γίνονται όλο και πιο συχνά χρήστες απομακρυσμένων συστημάτων ανάκτησης πολυμέσων. Σε τέτοια συστήματα όμως, λειτουργίες όπως η κύλιση και γενικότερα η περιήγηση στο περιβάλλον δυσχεραίνεται από το μικρό μέγεθος της οθόνης. Έρευνα για την αποτελεσματικότητα συστημάτων ανάκτησης σε τέτοιου είδους συσκευές υπάρχει με προσεγγίσεις βασισμένες στην προσοχή του χρήστη Chen et al. (2003) [42], και στην προσωποποίηση των σελίδων Bertini et al., (2005) [43].

ΚΕΦΑΛΑΙΟ 4^ο - ΑΝΑΛΥΣΗ ΕΙΚΟΝΑΣ

Έχοντας ερευνήσει την βιβλιογραφία για τον τρόπο με τον οποίο μπορεί να πραγματοποιηθεί η ανάκτηση εικόνων στα σημερινά συστήματα είναι η ώρα να αναλύσουμε τον τρόπο με τον οποίο μπορεί να γίνει η διαχείριση μιας εικόνας και τα χαρακτηριστικά που αυτή περιλαμβάνει.

Κατηγοριοποίηση ως προς την πληροφορία

Μία εικόνα όπως και όλες οι κωδικοποιημένες πληροφορίες έχει κάποια χαρακτηριστικά γνωρίσματα τα οποία την αποτελούν. Όταν μιλάμε για εικόνες στο μυαλό μας έρχονται χρώματα, τοπία, διάφορες απεικονίσεις κι αν το δούμε κι από την πλευρά του υπολογιστή ημερομηνία που τραβήχτηκε μια φωτογραφία, ανάλυση κτλ. Όλα αυτά τα χαρακτηριστικά μπορούμε να τα εντάξουμε σε δύο κατηγορίες.

Στα χαρακτηριστικά που δεν έχουν άμεση σχέση με την απεικόνιση και το περιεχόμενο μιας εικόνας αλλά έχουν σχέση τα τυπικά υπολογιστικά χαρακτηριστικά της όπως το μέγεθός της, η ημερομηνία αποθήκευσής της, ο δημιουργός της κτλ.

Στα χαρακτηριστικά τα οποία έχουν να κάνουν με αυτό καθ' αυτό το περιεχόμενο της εικόνας και τα οποία χωρίζονται σε δύο υποκατηγορίες

Στα χαρακτηριστικά χαμηλού επιπέδου ή ποιοτικά χαρακτηριστικά με το οποίο εννοούμε τα χρώματα που απαρτίζουν την εικόνα, τα σχήματα που υπάρχουν σε αυτήν, την υφή της κá.

Και στα χαρακτηριστικά υψηλού επιπέδου ή το περιεχόμενο της εικόνας όπως για παράδειγμα το τι αυτή απεικονίζει αντικειμενικά.

Μία ακόμη κατηγορία θα μπορούσε να είναι τα ψυχολογικά χαρακτηριστικά τα οποία προκαλεί η εικόνα. Με τον όρο αυτό εννοούμε τα συναισθήματα τα οποία μας προκαλούνται με την παρατήρησή της όπως χαρά, λύπη κτλ. Τα χαρακτηριστικά αυτά όμως είναι υποκειμενικά και δύσκολα θα μπορούσαν να στοιχειοθετηθούν και να μας δώσουν μία αντικειμενική πληροφορία για την

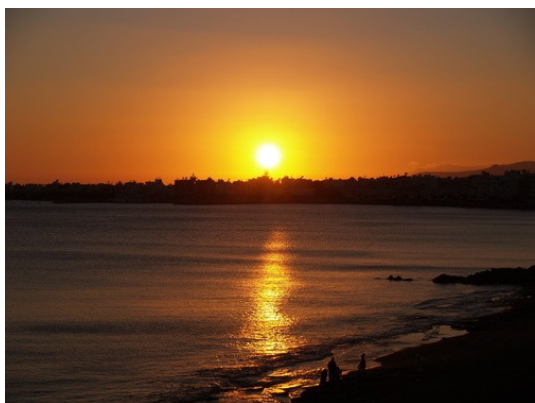
εικόνα. Παρ' όλ' αυτά υπάρχουν συστήματα ανάκτησης τα οποία προσπαθούν να καταχωρήσουν τα συναισθήματα αυτά και να εξάγουν ένα κοινό συμπέρασμα για κάθε εικόνα που βρίσκεται στο σύστημά του.

Χαρακτηριστικά

Τα χαρακτηριστικά μπορούν να εξαχθούν είτε από ολόκληρη την εικόνα (global features) είτε από ομάδες εικονοστοιχείων (pixel) (περιοχές της εικόνας (region features)). Τα πλέον χρησιμοποιούμενα χαρακτηριστικά είναι αυτά που κωδικοποιούν το χρώμα, την υφή, το σχήμα ή την περιοχή γύρω από κάποια σημεία ενδιαφέροντος της εικόνας. Οι κατηγορίες αυτές θα περιγραφούν περιληπτικά παρακάτω.

Στην περίπτωση που τα χαρακτηριστικά εξάγονται από ολόκληρη την εικόνα, επιδιώκεται να κωδικοποιηθούν συνολικά τα χαρακτηριστικά της. Με αυτόν τον τρόπο η εικόνα αναπαρίσταται ολόκληρη με ένα πολυδιάστατο διάνυσμα, που συνήθως έχει διάσταση της τάξης του 10^2 . Τα χαρακτηριστικά που εξάγονται από ολόκληρη την εικόνα χρησιμοποιούνται όλο και λιγότερο τα τελευταία χρόνια καθώς δεν είναι αποτελεσματικά σε εικόνες με πυκνό οπτικό περιεχόμενο.

Με τα γενικά χαρακτηριστικά μπορεί να κωδικοποιηθεί αποτελεσματικά μια εικόνα ηλιοβασιλέματος, αλλά όχι μια εικόνα με πολύχρωμα λουλούδια. Η αριστερή εικόνα αναπαρίσταται αποδοτικά με χρωματικού περιγραφείς εξαγόμενους από ολόκληρη την εικόνα (για παράδειγμα με κάποιον χρωματικό περιγραφέα) καθώς ολόκληρο το σημασιολογικό της περιεχόμενο μπορεί να περιγραφεί αποδοτικά μονάχα με την έννοια ηλιοβασίλεμα. Για την



εικόνα στα δεξιά θα χρειαστούν πιο εξελιγμένες τεχνικές εξαγωγής για να κωδικοποιηθεί αξιοποιήσιμα το περιεχόμενο της.

Χρώμα

Ιστορικό

Το χρώμα είναι ένα από το πιο διαδεδομένο χαρακτηριστικό μιας εικόνας. Η έρευνα για τα χρώματα ξεκίνησε ουσιαστικά από τα πρώτα βήματα των συστημάτων ανάκτησης εικόνων Wang et al. (1997) [44].

Μεγάλη έμφαση δόθηκε τότε στην εξερεύνηση διαφόρων χρωματικών χώρων (color spaces) που φαίνεται να προσεγγίζουν καλύτερα τον τρόπο με τον οποίο το ανθρώπινο μάτι αντιλαμβάνεται το χρώμα σε σχέση με τον κλασικό RGB χώρο ο οποίος ήταν ευρέως διαδεδομένος μέχρι εκείνη την στιγμή. Μελετήθηκε επίσης το πώς μεταβάλλεται το χρώμα σε συνάρτηση με τις αλλαγές στον φωτισμό, την θέση της κάμερας και την κλίση των επιφανειών και έχουν γίνει προσπάθειες να αναπτυχθούν μοντέλα που να λαμβάνουν υπόψη τις μεταβολές αυτές Sebe & Lew (2001) [45].

Το μοντέλο της αντίθετης αναπαράστασης χρωμάτων, για παράδειγμα Οι Swain & Ballard (1991) [46], κατάφερε να λάβει υπόψη τη μεταβολή της φωτεινότητας (brightness) σε έναν άξονα, με αποτέλεσμα να υπάρχει η δυνατότητα να απαλειφθεί αυτή η συνιστώσα.

Το ιστόγραμμα μιας εικόνας ή των χρωματικών της αποδόσεων είναι μια πάγια τεχνική που χρησιμοποιήθηκε σε πολλά συστήματα ανάκτησης της προηγούμενης δεκαετίας, όπως για παράδειγμα το σύστημα QBIC, Flickner et al. (1995) [47] της IBM. Στατιστικά το ιστόγραμμα χρωμάτων υποδηλώνει την από κοινού πιθανότητα των εντάσεων των τριών χρωματικών καναλιών. Μια εναλλακτική του ιστογράμματος αυτού είναι η χρήση των χρωματικών στιγμιοτύπων (Color moments), Stricker & Orengo (1995) [48].

Με το πέρασμα των χρόνων η έρευνα εστιάστηκε στην εξαγωγή μιας όσο το δυνατόν καλύτερης περίληψης των χρωμάτων της εικόνας, λαμβάνοντας υπόψη και την χωρική τους κατανομή. Το πρότυπο MPEG-7 περιλαμβάνει

σύγχρονους χρωματικούς περιγραφείς οι οποίοι θα αναλυθούν σε επόμενο κεφάλαιο.

Μοντέλα

Ο πιο απλός τρόπος για να αναπαρασταθεί το χρώμα ενός pixel μιας εικόνας στην ψηφιακή επεξεργασία είναι με έναν φυσικό αριθμό, από 0 ως N θέτοντας όπου N τον αριθμός χρωμάτων που χρησιμοποιείται στην παλέτα της εκάστοτε εικόνας. Υπάρχουν όμως πολλά άλλα μοντέλα που μπορούν να μας δώσουν μια ακριβέστερη περιγραφή των χρωματικών ερεθισμάτων (color stimulate). Αυτά τα μοντέλα είναι τα εξής/

- Μοντέλο μέτρησης χρώματος τα οποία μετράνε τον βαθμό απορρόφησης της αντανάκλαση της προσπίπτουσας φωτεινής ακτινοβολίας. Στην κατηγορία αυτή συναντάμε διαγράμματα που παρουσιάζουν φάσμα με βαθμούς αντανάκλασης, όπως το «χρωματικό διάγραμμα»
- Μοντέλα νευροφυσιολογίας τα οποία βασίζονται σε έρευνα της λειτουργίας του ανθρώπινου ματιού. Στα μοντέλα αυτά συναντάμε την έννοια των «κόνων». Ο κώνοι είναι τα κύτταρα του ματιού που ερεθίζονται διαφορετικού φασματικού επιπέδου ακτινοβολία το καθένα. Θα μπορούσαμε να τα χωρίσουμε αυτά σε τρεις κατηγορίες σύμφωνα με το χρώμα το οποίο αντιδρούν το καθένα από αυτά. Στα μοντέλα αυτά ανήκει το RGB (Red-Green-Blue) μοντέλο. Αν τα χωρίζαμε σε τέσσερις κατηγορίες θα μπορούσαμε για παράδειγμα να μιλήσουμε για το CMYK (Cyan-Magenta-Yellow-Black) μοντέλο.
- Ψυχολογικά μοντέλα τα οποία βασίζονται στις αντιδράσεις που προκαλούν τα χρώματα στην ψυχολογία του ανθρώπου. Τέτοιο μοντέλο είναι το hue-saturation-brightness, όπου η χρωματική απόδοση, η ζωντάνια του χρώματος και η φωτεινότητά του προκύπτουν με βάση αυτό το χαρακτηριστικό.

Χαρακτηριστικά Εικόνων

Από όλα αυτά τα χαρακτηριστικά της εικόνας, μπορούμε να πάρουμε ωφέλιμη πληροφορία, που θα μας χρησιμεύσει στην ανάκτηση εικόνας. Για τους διάφορους τύπους δεδομένων έχουν αναπτυχθεί διαφορετικά συστήματα και τεχνικές ανάκτησης.

Η αναγνώριση των χαρακτηριστικών χαμηλού επιπέδου μιας εικόνας αποτελούν την βάση για τα συστήματα ανάκτησης εικόνας με βάση το περιεχόμενο. Τα χαρακτηριστικά αυτά μπορούν να εξαχθούν από ολόκληρη την εικόνα ή από συγκεκριμένες περιοχές της. Όπως έχει διαπιστωθεί ο χρήστης συνήθως ενδιαφέρεται για μια συγκεκριμένη περιοχή μιας εικόνας παρά για ολόκληρη την εικόνα, γι' αυτό άλλωστε και τα περισσότερα συστήματα ανάκτησης εικόνων χρησιμοποιούν την μέθοδο της ανάκτησης συγκεκριμένων περιοχών. Η ανάκτηση με βάση τα συνολικά χαρακτηριστικά θεωρείται ευκολότερη. Η ανάκτηση της εικόνας με την χρήση περιοχών έχει αποδειχθεί ότι είναι πιο κοντά στον τρόπο αντίληψης του ανθρώπου, F. Jing et al. (2003) [49]

Για να εκτελέσουμε την Ανάκτηση εικόνας με βάση συγκεκριμένες περιοχές- όπως ονομάζεται- θα πρέπει πρωτ' απ' όλα να χρησιμοποιήσουμε μία τεχνική για την κατάτμηση της εικόνας όπως ονομάζεται η κοπή της εικόνας σε περιοχές. Έπειτα, χαμηλού επιπέδου χαρακτηριστικά όπως το χρώμα, η υφή, το σχήμα ή τα χωρικά δεδομένα μπορούν να εξαχθούν από τις μεμονωμένες περιοχές. Η ομοιότητα μεταξύ δύο εικόνων ορίζεται ως βάση τα χαρακτηριστικά των περιοχών.

Κατάτμηση εικόνας

Η αυτόματη κατάτμηση εικόνας είναι μια δύσκολη διαδικασία. Πληθώρα τεχνικών έχει προταθεί στο παρελθόν όπως η κατάτμηση με καμπύλες (curve evolution), H. Feng et al. (2001) [50] , η διάχυση της ενέργειας (energy diffusion) Ma & Majunath (1997) [51], και ο γραφικός τεμαχισμός (graph partitioning). Πολλές τεχνικές κατάτμησης δουλεύουν καλά για εικόνες που περιέχουν ομοιογενείς χρωματικά περιοχές. Αυτές όμως αναφέρονται μόνο σε

συστήματα ανάκτησης που δουλεύουν μόνο με χρώματα, Stanchev et al (2003) [52], Hua et al (1999) [53]

Παρ' όλ' αυτά, οι φυσικές φωτογραφίες είναι πολύ πλούσιες και σε χρώματα και σε υφές. και οι περισσότερες φωτογραφίες με φυσικό περιεχόμενο μπορούν να θεωρηθούν ως μωσαϊκό από περιοχές με διαφορετικά χρώματα και υφές. Η υφή είναι μια σημαντική παράμετρος για τον ορισμό του περιεχομένου μιας εικόνας και άλλωστε όπως έχει ερευνηθεί είναι πάρα πολύ δύσκολη η αναγνώρισή της. Πολλοί αλγόριθμοι κατάτμησης απαιτούν την εκτίμηση των παραμέτρων της υφής για να μπορέσουν να λειτουργήσουν σωστά, Deng & Manjunath (2001) [54]

Κατάτμηση JSEG

Η κατάτμηση «JSEG», Deng & Manjunath (2001) [54], λύνει αυτά τα προβλήματα. Αντί να κάνει μία εκτίμηση ενός συγκεκριμένου μοντέλου για την υφή, εξετάζει την ομοιογένεια των χρωμάτων. Η μέθοδος αυτή αποτελείται από δύο βήματα. Στο πρώτο βήμα τα χρώματα των εικόνων κβαντοποιούνται σε μερικές κλάσεις. Έτσι αντικαθίσταται το κάθε χρώμα με μία ετικέτα από την κλάση την οποία αντιπροσωπεύεται. Μπορούμε με αυτόν τον τρόπο να δημιουργήσουμε έναν χάρτη κλάσεων της εικόνας. Έπειτα η χωρική κατάτμηση εκτελείται σύμφωνα με τον χάρτη κλάσεων. Ο αλγόριθμος έπειτα παράγει ομοιογενείς περιοχές χρωμάτων και υφών και χρησιμοποιείται σε πολλά συστήματα [49, 50]. Οι παρακάτω εικόνες παρέχουν ένα παράδειγμα της εικόνας και του αποτελέσματος.



14 regions



11 regions

JSEG μέθοδος

Κατάτμηση Blobworld

Η κατάτμηση Blobworld Carson et al (2002) [55], είναι ένας ακόμη ευρέως γνωστός αλγόριθμος κατάτμησης . Feng et al (2003) [56], Shi et al (2004) [57]. . Αντλείται από την ομαδοποίηση των κουκίδων και συνδέεται με ένα επίπεδο χρωμάτος-υφής-θέσης. Στο πρώτο βήμα η κατανομή του χρώματος, της υφής, και της θέσης των χαρακτηριστικών μοντελοποιείται με την χρήση Γκαουσιανών μοντέλων. Έπειτα παρεμβάλλεται ο αλγόριθμος προσδοκώμενης μεγιστοποίησης (expectation maximization, EM) για να εκτίμηση τις παραμέτρους του μοντέλου. Το αποτέλεσμα της περιοχής κουκίδα-ομάδα παρέχει την κατάτμηση της εικόνας. Οι περιοχές που εξάγονται ως αποτέλεσμα αντιπροσωπεύουν το αντικείμενο.

Άλλοι αλγόριθμοι

Κάποια συστήματα ανάκτησης χρησιμοποιούν δικούς τους αλγόριθμους για την κατάτμηση των εικόνων χρησιμοποιώντας το χρώμα και την υφή ή και τα δύο μαζί. Αυτοί οι αλγόριθμοι συνήθως βασίζονται στην ομαδοποίηση κατά k-means της κουκίδας ή του block.

Χαρακτηριστικά εικόνων χαμηλού επιπέδου

Η εξαγωγή των χαρακτηριστικών γνωρισμάτων χαμηλού επιπέδου της εικόνας αποτελεί τη βάση των συστημάτων ανάκτησης. Παρακάτω θα εστιάσουμε στα γνωρίσματα που χρησιμοποιούνται από τα συστήματα ανάκτησης.

Χαρακτηριστικό χρώματος

Το χρώμα ή αλλιώς το χαρακτηριστικό χρώματος χρησιμοποιείται ευρέως για την ανάκτηση εικόνων. Τα χρώματα ορίζονται με βάση ένα συγκεκριμένο χρωματικό σύστημα. Πολλές χρωματικές παλέτες ή χρωματικά συστήματα είναι διαθέσιμα για χρήση και συνήθως χρησιμοποιούνται σε διαφορετικές εφαρμογές. Οι χρωματικές παλέτες που συνήθως είναι πιο κοντά στην ανθρώπινη αντίληψη και χρησιμοποιούνται ευρέως είναι τα RGB, LAB, LUV, HSV (HSL) [53], [54], [55]. Κοινά χαρακτηριστικά χρωμάτων ή περιγραφείς

στα συστήματα ανάκτησης περιλαμβάνουν χρωματικά ιστογράμματα, χρωματικές στιγμές, χρωματικές διανυσματικές συνοχές κτλ. Το πρότυπο MPEG-7 όπως θα περιγράψουμε παρακάτω συμπεριλαμβάνει τα επικρατέστερα χρώματα, τις χρωματικές δομές και τις χρωματικές διατάξεις.

Τα περισσότερα από τα χαρακτηριστικά χρώματος αν και περιγράφουν αποδοτικά το χρώμα, δεν συσχετίζονται άμεσα με την υψηλού επιπέδου σημασιολογία. Για την κατάλληλη χαρτογράφηση των περιοχών χρώματος, μερικά συστήματα χρησιμοποιούν το μέσο χρώμα όλων των κουκίδων σε μια περιοχή ως χαρακτηριστικό γνώρισμα χρώματός της. Η κατάτμηση συνήθως συνεπάγεται χωρισμό της εικόνας σε περιοχές που είναι χρωματικά ομοιογενείς, παρόλα αυτά λόγω της ανακρίβειας της κατάτμησης το κυρίαρχο χρώμα μιας περιοχής κάποιες φορές απέχει πολύ το μέσο χρώμα. Άρα για το μέσο χρώμα υπάρχουν δύο εκδοχές, το μέσο χρώμα και το κυρίαρχο να είναι παρόμοια ή να είναι πολύ διαφορετικά.

Η επιλογή των χαρακτηριστικών χρώματος που θα επιλέξουμε για μια επιτυχημένη ανάκτηση εικόνας εξαρτάται από τα αποτελέσματα της κατάτμησης. Για παράδειγμα αν οι περιοχές που προκύπτουν από την κατάτμηση δεν είναι ομοιογενείς ως προς το χρώμα, τότε δεν θα μας χρησίμευε το μέσο χρώμα. Εδώ πρέπει να σημειωθεί ότι στα περισσότερα συστήματα ανάκτησης οι εικόνες δεν είναι προ επεξεργασμένες. Δεδομένου ότι οι εικόνες συχνά αλλοιώνονται από τον θόρυβο, καλό θα ήταν να εφαρμόσουμε ένα φίλτρο που να απομακρύνει τον θόρυβο χρώματος ώστε να βελτιώσουμε την ακρίβεια της ανάκτησης. Άρα η επεξεργασία είναι σημαντική, ιδικά όταν τα αποτελέσματα χρησιμοποιούνται για την απόδοση προσώπων.

Χρωματικά μοντέλα ή συστήματα

Παρακάτω θα δούμε κάποια από τα χαρακτηριστικά που διαισθάνεται ο άνθρωπος βλέποντας τα χρώματα σε μία εικόνα. Αυτά είναι τα εξής.

- Φωτεινότητα (Luma και Luminance), την οποία αντιλαμβάνεται ο άνθρωπος ως αίσθηση για το αν μια περιοχή εκθέτει περισσότερη ή λιγότερη από την απόχρωσή του χρώματος.

- Κορεσμός (saturation), γίνεται αντιληπτή ως η πληρότητα χρώματος μιας περιοχής σε σχέση με τη λαμπρότητα ενός αναφερόμενου λευκού, δηλαδή το ότι σε κάποιες περιοχές τις εικόνες το χρώμα είναι πιο 'ζωντανό' από άλλες.
- Χρωματική Απόχρωση (hue), είναι η ανθρώπινη αίσθηση με την οποία μια περιοχή εκθέτει διαφορετικούς χρωματικούς τόνους.
- Λάμψη (brightness), με την οποία εννοούμε την πληρότητα χρώματος μιας περιοχής σε σχέση με τη λάμψη του.
- Χρωματικότητα (Lightness), η οποία χωρίζεται σε, Χρώμα (chroma), και είναι η αίσθηση της λαμπρότητας μιας περιοχής σε σχέση με μια άσπρη αναφορά στη σκηνή και σε Πληρότητα χρώματος (colourfulness), η οποία είναι η ανθρώπινη αίσθηση σύμφωνα με την οποία μια περιοχή εμφανίζεται να είναι όμοια με ένα, ή με ποσοστά από δυο, από τα αντιλαμβανόμενα χρώματα κόκκινο, κίτρινο, πράσινο και μπλε.

Τα παραπάνω συνθέτουν τις αισθήσεις του ανθρώπινου ματιού που οδήγησαν στην δημιουργία των χρωματικών χώρων στους οποίους θα κάνουμε μία σύντομη αναφορά παρακάτω

Χρωματικό σύστημα

Οι χρωματικοί χώροι ουσιαστικά χρησιμοποιούνται για να περιγράψουν τον τρόπο κωδικοποίησης της χρωματικής πληροφορίας σε μια εικόνα. Το κάθε χρωματικό μοντέλο ορίζει ένα σύστημα συντεταγμένων και έναν υποχώρο μέσα σε αυτό. Το κάθε χρώμα είναι μοναδικό και αναπαρίσταται σε ένα μοναδικό σημείο.

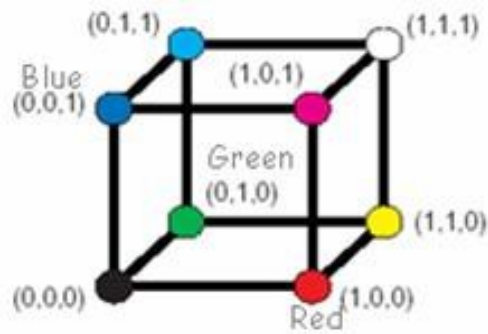
RGB χρωματικό σύστημα

Το πρότυπο χρώματος RGB είναι ένα προσθετικό πρότυπο στο οποίο τα χρώματα κόκκινο, πράσινο και μπλε (χρώματα που χρησιμοποιούνται συχνά σε προσθετικά χρωματικά πρότυπα) συνδυάζονται με διάφορους τρόπους για να αναπαραχθούν άλλα χρώματα.[1] Το όνομα του προτύπου και η σύντηξη

RGB προέρχονται από τα τρία βασικά χρώματα, το κόκκινο (Red), πράσινο (Green), και το μπλε (Blue). Αυτά τα τρία χρώματα δεν πρέπει να συγχέονται με τα τρία ανακλαστικά χρώματα κόκκινο, μπλε, και κίτρινο, τα οποία αναφέρονται στον χώρο των τεχνών ως βασικά χρώματα.

Ένα χρώμα στο πρότυπο χρώματος RGB μπορεί να περιγραφεί με το προσδιορισμό του πόσο κάθε ένα από το κόκκινο, πράσινο και μπλε χρώματα συμπεριλαμβάνεται. Κάθε ένα μπορεί να ποικίλει μεταξύ του ελάχιστου (καθόλου χρώμα) και του μέγιστου (πλήρης ένταση). Εάν όλα τα χρώματα είναι στο ελάχιστο το αποτέλεσμα είναι μαύρο. Εάν όλα τα χρώματα είναι στο μέγιστο, το αποτέλεσμα είναι το άσπρο. Τα χρώματα μπορούν να περιγραφούν ποσοτικά με διάφορους τρόπους:

- Οι επιστήμονες του χρώματος συχνά τοποθετούν τα χρώματα στην κλίμακα 0 (ελάχιστο) έως 1 (μέγιστο). Πολλοί μαθηματικοί τύποι που σχετίζονται με το χρώμα χρησιμοποιούν αυτές τις τιμές. Π.χ. το μέγιστο κόκκινο είναι 1,0,0 για Κόκκινο, Πράσινο, Μπλε.
- Οι τιμές χρώματος μπορούν να γραφτούν επίσης ως ποσοστά, από 0% (ελάχιστο) ως 100% (μέγιστο). Το μέγιστο κόκκινο είναι 100%, 0%, 0%.
- Οι τιμές χρώματος μπορούν να γραφτούν ως αριθμοί στην κλίμακα 0 έως 255, απλά με τον πολλαπλασιασμό της κλίμακας 0.0 έως 1.0 με 255. Αυτό το μοντέλο απαντάται συνήθως στην πληροφορική, όπου οι προγραμματιστές προτιμούν να αποθηκεύουν κάθε αξία χρώματος σε ένα byte (8 bit). Αυτή η σύμβαση έχει γίνει τόσο διαδεδομένη ώστε πολλοί συγγραφείς την θεωρούν αυτονόητη και δεν παρέχουν το σωστό υπόβαθρο αναφοράς. Το μέγιστο κόκκινο είναι το 255,0,0.
- Η ίδια σειρά, 0 έως 255, γράφεται μερικές φορές σε δεκαεξαδικό, και ίσως με ένα πρόθεμα (π.χ. #). Επειδή οι δεκαεξαδικοί αριθμοί σε αυτήν την κλίμακα μπορούν να γραφτούν με ένα σταθερό σχήμα δύο ψηφίων, το μέγιστο κόκκινο #FF, #00, #00 μπορεί να γραφτεί και σαν #ff0000. Αυτή η σύμβαση χρησιμοποιείται στα χρώματα στο διαδίκτυο και θεωρείται επίσης από μερικούς συγγραφείς αυτονόητη.



CMYK

Το χρωματικό μοντέλο CMYK είναι ένα αφαιρετικό χρωματικό μοντέλο. Χρησιμοποιεί τέσσερα βασικά χρώματα, το κυανό, το ματζέντα, το κίτρινο και το μαύρο. Χρησιμοποιείται επίσης για την εξαγωγή χρωμάτων και ανοιχτόχρωμα υπόβαθρα κυρίως τα λευκά. Το μοντέλο αυτό είναι αφαιρετικό γιατί το χρώμα αφαιρεί την λάμψη από το λευκό υπόβαθρο κι έτσι δημιουργείται η αίσθηση ενός διαφορετικού χρώματος.

Η διαφορά των δύο μοντέλων RGB και CMYK είναι ότι ενώ στο RGB με την τριάδα (0,0,0) αναπαριστούμε το μαύρο δηλαδή την πλήρη απορρόφηση του φωτός και η (1,1,1) το λευκό δηλαδή την πλήρη αντανάκλαση του φωτός, στο CMYK μοντέλο ισχύει ακριβώς το αντίστροφο. Δηλαδή εδώ αντί να προσθέτουμε φως στο Μαύρο, μετράμε πόσο χρώμα αφαιρούμε από το Λευκό. Άρα στο CMYK με το 1,1,1 εννοούμε το μαύρο (απορροφάται όλο το χρώμα).

Προσεγγιστικά, για μετατροπές από το ένα μοντέλο στο άλλο έχουμε τις εξής απλές εξισώσεις:

$$R = 1 - C \quad (1)$$

$$G = 1 - M \quad (2)$$

$$B = 1 - Y \quad (3)$$

Πρακτικά το CMYK μοντέλο χρησιμοποιείται περισσότερο στους εκτυπωτές, οι οποίοι κατασκευάζουν το κάθε χρώμα σαν μείξη των Κυανό, Πορφυρό, Κίτρινο.

Σχήμα

Για να πετύχουμε σε ένα σύστημα ανάκτηση εικόνας με βάση το σχήμα αυτό συνεπάγεται ότι έχουμε πρόσβαση σε σχήματα που έχουν εξαχθεί από πραγματικές εικόνες. Οπότε ουσιαστικά πρέπει πρώτα να αναλύσουμε τις εικόνες αφού τα πρωτογενή δεδομένα που είναι διαθέσιμα είναι τα ίδια τα πολυμέσα που θέλουμε να ανακτήσουμε. Τα σχήματα συνήθως είναι τα περιγράμματα των αντικειμένων που εμφανίζονται στις εικόνες. Εδώ πρέπει να προσέξουμε, γιατί θέλουμε να υπάρχει ανεξαρτησία της μεθόδου ανάκτησης από τα ίδια τα δεδομένα ανάκτησης.

Το σχήμα (shape) αποτελεί ένα άλλο βασικό γνώρισμα για την ανάκτηση εικόνας με βάση το περιεχόμενο. Τα φυσικά αντικείμενα αναγνωρίζονται πέρα από το χρώμα τους κι από το σχήμα τους. Ένας αριθμός χαρακτηριστικών γνωρισμάτων των σχημάτων (ανεξάρτητα από το μέγεθος ή τον προσανατολισμό) των αντικειμένων της εικόνας εξάγονται για κάθε αποθηκευμένη εικόνα από το σύστημα. Όταν διεξάγεται μια αναζήτηση το σύστημα ανακαλεί αυτές τις εικόνες που τα χαρακτηριστικά τους ταιριάζουν περισσότερο με αυτά του ερωτήματος υπολογίζοντας τα χαρακτηριστικά της εικόνας που δίνεται στο ερώτημα. Γενικά οι σχηματικές αναπαραστάσεις διαιρούνται σε δύο κατηγορίες, σ' αυτές που βασίζονται στα όρια του σχήματος (boundary-based) και σ' αυτές που βασίζονται στην περιοχή (region-based). Οι πρώτες χρησιμοποιούν μόνο τα εξωτερικά όρια του σχήματος, το περίγραμμά του ενώ οι δεύτερες χρησιμοποιούν τη συνολική περιοχή του σχήματος. Οι ερωτήσεις κατά την αναζήτηση γίνονται είτε με τη δήλωση μιας εικόνας (ή κάποιου σχήματος αυτής) ως υπόδειγμα είτε με τη σχεδίαση ενός σχήματος από το χρήστη.

Παρακάτω παρουσιάζονται μερικά από τα πιο γνωστά συστήματα Ανάκτησης Εικόνας και Βίντεο που εκμεταλλεύονται Τεχνικές Σύγκρισης Σχημάτων:

1. Το σύστημα Query by Image Content (QBIC) της IBM παρέχει δυνατότητες ανάκτησης εικόνας μέσω ενός συνόλου κριτηρίων επιλογής που βασίζονται στο χρώμα, την υφή, τα σχετική θέση και το σχήμα αντικειμένων. Τα πολυμέσα ανακτώνται με βάση σχήματα που είτε έχουν εξαχθεί από άλλη εικόνα είτε σχεδιάζονται από το χρήστη του συστήματος.

2. Το σύστημα VIR, της Virage, χρησιμοποιεί πληροφορίες σχήματος για την ανάκτηση εικόνων.
3. Το σύστημα Excalibur Visual Retrieval Ware SDK (παρέχει δυνατότητες ανάκτησης εικόνας με βάση διάφορα χαρακτηριστικά, ανάμεσα στα οποία και πληροφορίες σχήματος).

Υφή (texture)

Η υφή είναι από τα βασικά χαρακτηριστικά της εικόνας, η οποία όμως δεν μπορεί να οριστεί με τρόπο κατάλληλο όπως η χρωματική απόδοση (χρώμα), Mezaris et al (2003) [58], Stanchev (2003) [59]. Αρκετά από τα συστήματα ανάκτησης δεν λαμβάνουν υπόψη τους την υφή. Παρ' όλ' αυτά η υφή παρέχει σημαντικές πληροφορίες για την κατηγοριοποίηση της εικόνας καθώς περιγράφει το περιεχόμενο πολλών πραγματικών εικόνων. Τέτοια παραδείγματα είναι η φλούδα των φρούτων, τα σύννεφα, τα δέντρα, τα τούβλα και τα υφάσματα. Οπότε η υφή είναι ένα πολύ σημαντικό χαρακτηριστικό για τον ορισμό της υψηλού επιπέδου σημασιολογίας μιας εικόνας για την καλύτερη ανάκτησή της. Ο διαχωρισμός της υφής γίνεται σε δύο κατηγορίες, στην φυσική υφή και στην οπτική υφή

Με τον όρο φυσική υφή εννοούμε την απτές μεταβλητές ενός αντικειμένου. Δηλαδή την αίσθηση που μας προκαλείτε με το άγγιγμα της επιφάνειας ενός αντικειμένου. Παραδείγματα υπάρχουν πολλά, όπως η τραχύτητα της επιφάνειας ενός ξύλου και η αναγνώριση των στρώσεων του κορμού, η αντίληψη των κόκκων της άμμου κτλ.

Με τον όρο οπτική υφή εννοούμε την παραισθητική αντίληψη της φυσικής υφής. Από μικρό παιδί ο άνθρωπος μαθαίνει να αναγνωρίζει αντικείμενα μόλις τα δει ανακτώντας ουσιαστικά την πληροφορία της υφής από τον εγκέφαλό του και αναγνωρίζοντάς την μόνο αφού την έχει αντιληφθεί έστω και μία φορά στην ζωή του. Κάθε υλικό και κάθε επιφάνεια έχει την δική της υφή.

Χωρικά Χαρακτηριστικά

Εκτός από το χρώμα την υφή και το σχήμα, τα χωρικά χαρακτηριστικά είναι επίσης ένας ακόμη χρήσιμος τρόπος για την ταξινόμηση των περιοχών. Για παράδειγμα, ο ουρανός και η θάλασσα σε πολλές περιπτώσεις έχουν παρόμοια χρώματα αλλά και υφή, αλλά η χωρική κατανομή τους είναι σίγουρο ότι διαφέρει αφού ο ουρανός βρίσκεται πάντα στο πάνω μέρος της εικόνας ενώ αντίθετα η θάλασσα βρίσκεται στο κάτω μέρος.

Τα χωρικά χαρακτηριστικά συνήθως είναι εύκολο να χαρακτηριστούν από τις λέξεις άνω, κάτω, πάνω από σύμφωνα μάλιστα και με την περιοχή της φωτογραφίας στην οποία βρίσκονται.

ΚΕΦΑΛΑΙΟ 5^ο - ΤΟ ΠΡΟΤΥΠΟ MPEG7

MPEG-7: Εισαγωγή

Το MPEG-7 ή αλλιώς «Περιβάλλον Περιγραφής Πολυμεσικής Πληροφορίας» (Multimedia Content Description Interface), αποτελεί ένα επίσημο πρότυπο από τον Φεβρουάριο του 2002 και είναι το πρότυπο που προβλέπεται από το MPEG (Moving Pictures Expert Group) για την διαχείριση και την περιγραφή της πολυμεσικής πληροφορίας.

Η πληροφορία αυτή περιλαμβάνει πολλές μορφές, όπως κείμενο, σταθερές εικόνες, γραφικά, μοντέλα 3D, ήχο, ομιλία, video και συνδυασμούς όλων των προηγούμενων. Οι γενικοί στόχοι που καλείται να εξυπηρετήσει το πρότυπο είναι δύο:

1. Να γίνει ευκολότερος ο στόχος της ανάκτησης της πολυμεσική πληροφορίας και
2. ο δεύτερος στόχος είναι η επίτευξη της διαλειτουργικότητας μεταξύ των εφαρμογών και του λογισμικού που χρησιμοποιούν οι εφαρμογές, και όλα αυτά να πραγματοποιηθούν με την χρήση των μεταδεδομένων.

Πριν ξεκινήσουμε την περιγραφή της δομής και των τρόπων για την επίτευξη των παραπάνω στόχων, θα ήταν καλό να γίνει μία σύντομη περιγραφή των προηγούμενων προτύπων που αναπτύχθηκαν από το MPEG, δηλαδή το MPEG-1, MPEG-2 και MPEG-4 καθώς και το πλαίσιο για την δημιουργία του MPEG-7.

Το MPEG είναι μία ομάδα του Διεθνούς Οργανισμού Προτυποποίησης (ISO, International Organization for Standardization), που σκοπός της είναι η ανάπτυξη προτύπων για συμπίεση, επεξεργασία, κωδικοποίηση και διαχείριση εικόνων, ήχου και των συνδυασμών τους, που καθιερώνονται διεθνώς. Πολύ γνωστά πρότυπα που έχουν δημιουργεί είναι τα MPEG-1, MPEG-2, MPEG-3(το γνωστό σε όλους mp3), MPEG-4(αντίστοιχα το γνωστό mp4), MPEG-7 και MPEG-21.

MPEG-1

Το MPEG-1 αφορά την συμπίεση ήχου και εικόνας. Μέρος του προτύπου είναι και το MPEG-1 Audio που αφορά αποκλειστικά τον ήχο και επίσης είναι το πρώτο διεθνές πρότυπο για την ψηφιακή συμπίεση ήχου υψηλής πιστότητας. Στο πρότυπο αυτό ανήκει και το παγκοσμίως γνωστό MP3 (MPEG- Layer 3), που είναι το αποτέλεσμα της αναζήτησης για μεγαλύτερη συμπίεση και ταυτόχρονα καλή ποιότητα ήχου.

MPEG-2

Το MPEG-2 , είναι το πρότυπο για την ψηφιακή τηλεόραση και αποτελεί την συνέχεια του MPEG-1 ως προς τις επιπλέον δυνατότητες που παρέχει. Σαν μία συμβατική συνέχεια λοιπόν του MPEG-1 υποστηρίζει μορφές video (video formats) και έναν αριθμό άλλων προηγμένων χαρακτηριστικών, όπως τέτοιων που να μπορούν να υποστηρίξουν τον τύπο HDTV (High Definition TV) καθώς επίσης αποτέλεσε και την βάση για το DVD Video (Digital Versatile Disc). Το MPEG-2 παρέχει επίσης ένα σύστημα με ορισμούς (definitions) για το πώς το video, ο ήχος και άλλα δεδομένα συνδυάζονται σε ένα ή πολλαπλά stream που να είναι κατάλληλα για την αποθήκευση και μετάδοσή τους.

MPEG-4

Το MPEG-4 αφορά την παραγωγή, κατανομή (distribution) και πρόσβαση του πολυμεσικού περιεχομένου. Παρόλο που το MPEG-1 και MPEG-2 κάλυπταν ένα αρκετά μεγάλο εύρος εφαρμογών, όπως η ψηφιακή τηλεόραση και η διαδραστική κινούμενη εικόνα (interactive video), σύντομα έγινε αντιληπτό ότι οι εφαρμογές που έχουν σχέση με τα πολυμέσα έχουν περισσότερες απαιτήσεις. Αναπτύχθηκε λοιπόν το MPEG-4 με σκοπό να παρέχει ένα προτυποποιημένο σύνολο τεχνολογιών, που θα καταστήσει ικανή την ενοποίηση (integration) της παραγωγής, κατανομής, και τα παραδείγματα για την πρόσβαση στο περιεχόμενο (content access paradigms), και τα διαδραστικά πολυμέσα για την ανάπτυξη της χρήσης των εφαρμογών πολυμέσων στο Internet.

Θα πρέπει επίσης να σημειώσουμε ότι το MPEG-4 επιτρέπει την επισύναψη μετά-δεδομένων (metadatas) σχετικά με το περιεχόμενο επάνω στο

αντικείμενο. Επομένως ο χρήστης μπορεί να χρησιμοποιήσει αυτή τη ροή δεδομένων (data stream) με την πληροφορία του περιεχομένου του αντικειμένου (Object Content Information, OCI) για να στείλει τα μετα-δεδομένα ταυτόχρονα με το περιεχόμενο του MPEG-4. Ωστόσο στο συγκεκριμένο πρότυπο δεν υπάρχει κάποια τυποποιημένη δομή και καθορισμένη μορφή που να αφορά τα μετα-δεδομένα.

Χώρος ενδιαφερόντων του MPEG-7

Ενώ τα προηγούμενα πρότυπα του MPEG είχαν ως στόχο την καλύτερη παρουσίαση και αναπαράσταση της πληροφορίας, ο στόχος του MPEG-7 είναι να παρέχει ένα προτυποποιημένο περιβάλλον για την περιγραφή της πληροφορίας αυτής. Στόχος δηλαδή είναι η πλήρης περιγραφή της πολυμεσικής πληροφορίας, τα μετα-δεδομένα («τα δεδομένα για τα δεδομένα»), και όχι το περιεχόμενό της. Τα μετα-δεδομένα είναι απαραίτητα γιατί διευκολύνουν την ανταλλαγή, την αναζήτηση, την ανάκτηση και το φιλτράρισμα των πληροφοριών που υπάρχουν στα πολυμεσικά τεκμήρια..

Το κίνητρο για την ανάπτυξη του MPEG-7 δόθηκε από τις ξεκάθαρες πλέον τάσεις στην τεχνολογία, στην αγορά και στους χρήστες. Με την εξάπλωση των εργαλείων για την δημιουργία οπτικού ή/και ακουστικού περιεχομένου και την ψηφιοποίησή του και με την ευρεία χρήση των πολυμέσων σε ολοένα και περισσότερους τομείς (πέρα από την προσωπική χρήση) δημιουργήθηκε ένα περιβάλλον, όπου παράγεται όλο και περισσότερη πληροφορία οπτικοακουστικού περιεχομένου. Η ίδια η φύση αυτής της πληροφορίας υπαγορεύει νέους τρόπους ως προς την επεξεργασία, την οργάνωση και την αξιοποίηση της.

Όπως είναι, λοιπόν, φυσιολογικό αναπτύσσονται νέα συστήματα και εργαλεία που να καλύπτουν τις ανάγκες για συλλογή, διαχείριση, ευρετηρίαση, και αναζήτηση πολυμεσικού περιεχομένου. Εκτός από τις ανάγκες των χρηστών, τέτοιες λύσεις πρέπει να δοθούν και για τους διάφορους παραγωγούς των υπηρεσιών και για τους δημιουργούς του περιεχομένου. Γι' αυτό λοιπόν και το MPEG-7 προκειμένου να επιτύχει τη μέγιστη δια-λειτουργικότητα και να επιτρέψει την δημιουργία καινοτόμων εργαλείων και εφαρμογών, παρέχει ένα

δια-λειτουργικό περιβάλλον, στο οποίο ορίζεται η δομή και η σημασιολογία ποικίλων εργαλείων περιγραφής.

Κάθε ένα από αυτά τα εργαλεία περιγραφής μπορεί να έχει σχεδιαστεί για κάποια εξειδικευμένη ή γενική μορφή πληροφορίας (π.χ. για ήχο, για εικόνα ή για πολυμέσα), πλευράς (aspects) (π.χ. δομική, σημασιολογική) και εφαρμογής (π.χ. για μηχανή αναζήτησης ή για πλοήγηση). Ως απάντηση στις ανάγκες αυτές έχουν πραγματοποιηθεί από διάφορες ομάδες και οργανισμούς αρκετές δραστηριότητες για τον καθορισμό δια-λειτουργικών πλαισίων (frameworks) και για τον καθορισμό της αναπαράστασης της περιγραφής των μετα-δεδομένων (representations for metadata descriptions.) Κάποιες αξιοσημείωτες εργασίες, που έχουν σχέση με το MPEG-7, περιλαμβάνουν το Dublin Core, το Society of Motion Picture and Television Engineers (SMPTE) και το European Broadcasting Union (EBU), που ωστόσο απαιτούν ακόμα αρκετό συγχρονισμό.

Μέσα σε αυτό το πλαίσιο λοιπόν αναπτύχθηκε το MPEG-7. Το MPEG-7 δε στοχεύει σε κάποια συγκεκριμένη εφαρμογή, αλλά θα μπορούσε να διατυπωθεί ότι τα στοιχεία που προτυποποιεί υποστηρίζουν μία ευρεία σειρά εφαρμογών. Αυτή είναι επίσης και η διαφορά του σε σχέση με κάποια άλλα πρότυπα που αφορούν τα μετα-δεδομένα, ότι δηλαδή στοχεύει στο να είναι ένα γενικό (generic) πρότυπο και όχι στο να χρησιμοποιείται σε μία και μόνο εφαρμογή ή σε ένα μόνο πεδίο δραστηριοτήτων.

Προτυποποίηση πολυμέσων

Οι πηγές των πολυμέσων (multimedia sources) στην σημερινή κοινωνία κατέχουν ένα σημαντικό ρόλο και χρησιμοποιούνται σε πολλές από τις δραστηριότητές μας. Γι' αυτόν τον λόγο υπάρχει η ανάγκη για περαιτέρω επεξεργασία τους. Αυτή η ανάγκη καθιστά απαραίτητη όχι μόνο την ανάπτυξη προτυποποιημένων τρόπων για την αναπαράσταση της πολυμεσικής πληροφορίας, αλλά και τρόπων για την περιγραφή της. Για παράδειγμα θα μπορούσε να χρησιμοποιηθεί το εξής: ένας αισθητήρας εικόνας μπορεί να παράγει οπτικά δεδομένα (visual data) και να είναι συνδεδεμένη σε αυτά πληροφορία σχετική με τον χρόνο και τα φυσικά μεγέθη. Μία συσκευή

επεξεργασίας μπορεί να επεξεργαστεί την πληροφορία αυτή είτε τοπικά είτε απομακρυσμένα και να επαληθεύσει, εάν τηρούνται κάποιες συνθήκες του προγράμματος.

Ακόμα μία συσκευή για εγγραφή video (a video recording device) μπορεί να δέχεται περιγραφές της πολυμεσικής πληροφορίας που είναι συνδεδεμένες με ένα πρόγραμμα και που θα μπορούσε να εγγραφεί σε αυτήν, όπως για παράδειγμα να καταγράψει το δελτίο ειδήσεων αποκλείοντας όμως τις αθλητικές ειδήσεις.

Ακόμη θα μπορούσαμε να κάνουμε περιγραφή των προϊόντων σε μία εταιρεία, ώστε μία μηχανή να μπορεί να εξάγει αποτελέσματα ακόμη και σε όχι δομημένα (unstructured) queries των καταναλωτών. Όλες αυτές οι δυνατότητες που παρέχει το πρότυπο στηρίζονται στη χρήση των περιγραφών. Τα διάφορα μέρη του MPEG-7 (που θα αναλυθούν στη συνέχεια), είτε συλλογικά είτε μεμονωμένα, παρέχουν ένα πλήρες σύνολο με εργαλεία περιγραφής πολυμέσων, με σκοπό να δημιουργηθούν οι περιγραφές. Στο επόμενο κεφάλαιο γίνεται μια προσπάθεια να καταγραφούν οι εφαρμογές του MPEG-7 .

Εφαρμογές

Όπως προαναφέρθηκε στόχος του MPEG-7 είναι να προάγει την διαλειτουργικότητα ανάμεσα στα συστήματα και στις εφαρμογές που χρησιμοποιούνται για την δημιουργία, διαχείριση, κατανομή και χρήση των περιγραφών του οπτικοακουστικού περιεχομένου. Αυτές οι περιγραφές βοηθήσουν τον χρήστη και τις εφαρμογές στον προσδιορισμό της πληροφορίας, στην ανάκτηση και στο να φιλτράρουν (filter) αυτήν την πληροφορία. Το MPEG-7 μπορεί να χρησιμοποιηθεί τόσο σε επιστημονικές όσο και σε απλές εφαρμογές που αφορούν έναν καταναλωτή, όπως

- Στην εκπαίδευση
- Στην δημοσιογραφία (στην αναζήτηση κάποιας ομιλίας ενός πολιτικού, χρησιμοποιώντας το όνομά του ή την φωνή του.)

- Σε πολιτισμικές υπηρεσίες (όπως σε ένα μουσείο, σε μία αίθουσα τέχνης)
- Στην ψυχαγωγία (για διάφορα παιχνίδια)
- Σε συστήματα Γεωγραφικής Πληροφόρησης.
- Σε βιοϊατρικές εφαρμογές
- Σε ερευνητικές υπηρεσίες (για την αναγνώριση ανθρώπινων χαρακτηριστικών)
- Στην αρχιτεκτονική και στην διακόσμηση εσωτερικών χώρων
- Για πληροφορίες σε σχέση με τον τουρισμό
- Σε κοινωνικές εφαρμογές
- Σε αρχεία video, μιας ταινίας ή ενός ραδιοφωνικού σταθμού.

Παραδείγματα εφαρμογών αποτελούν οι ψηφιακές βιβλιοθήκες (ένας κατάλογος χρηστών με βάση την εικόνα τους), υπηρεσίες του χρυσού οδηγού με την χρήση πολυμέσων, επεξεργασία με την χρήση πολυμέσων (multimedia editing) (για παράδειγμα υπηρεσίες για εξατομίκευση των ειδήσεων σε ηλεκτρονική μορφή).

Δεδομένης όλης αυτής της ποικιλίας των εφαρμογών, οι περιγραφές του MPEG-7 δίνουν την δυνατότητα στους χρήστες ή στις εφαρμογές να εκτελέσουν τις παρακάτω εργασίες που αναφέρονται παρακάτω.

Στην κατηγορία των πολυμέσων βοηθούν στην δημιουργία ενός εξειδικευμένου προγράμματος ενός οδηγού ή μιας περίληψης από κάποιο τηλεοπτικό οπτικό ή/ και ακουστικό περιεχόμενο σύμφωνα με τις προτιμήσεις του εκάστοτε χρήστη. Ακόμη βοηθούν στην δημιουργία μιας βάσης δεδομένων εικόνων όπου μπορεί να γίνει η ανάκτηση της πολυμεσικής πληροφορίας.

Στα αρχεία δημιουργούν περιγραφές από κάποια μεμονωμένα στοιχεία μιας συλλογής ή ακόμα και από ολόκληρη την συλλογή από οπτικοακουστικό περιεχόμενο και γίνεται ανταλλαγή των περιγραφών αυτών με τους ιδιοκτήτες του περιεχομένου, τους οργανισμούς ή τους καταναλωτές.

Στην προσαρμοστικότητα, φιλτράρουν και μετατρέπουν τις ροές των πολυμέσων (multimedia streams) σε περιβάλλοντα περιορισμένων πόρων (resource limited environments) ταιριάζοντας τις προτιμήσεις των χρηστών, τους διαθέσιμους πόρους και τις περιγραφές του περιεχομένου.

Στον ήχο, δίνουν την δυνατότητα να παίξουμε μερικές νότες και να μας επιστραφεί μία λίστα από μουσικά κομμάτια που να περιλαμβάνουν αυτόν τον ρυθμό, συνδυάζοντας κάπως τις νότες. Παράδειγμα τέτοιας εφαρμογής είναι το TrackIt, το οποίο χρησιμοποιείται από κινητά τηλέφωνα για να αναγνωρίσει ένα συγκεκριμένο μουσικό κομμάτι.

Στα γραφικά μπορούμε να σχεδιάσουμε μερικές γραμμές στην οθόνη και να μας επιστρέψει ένα σύνολο από εικόνες, που να περιέχουν παρόμοια σχέδια ή εικόνες, δηλαδή ανάκτηση μέσω κάποιου σκίτσου

Στην κίνηση, έχοντας ένα σύνολο από αντικείμενα video (video objects), μπορούμε να περιγράψουμε κινήσεις και σχέσεις ανάμεσα στα αντικείμενα και να μας επιστραφεί μία λίστα από video clips που να εκπληρώνουν τις χωρικές και χρονικές σχέσεις.

Στο σενάριο (scenario) μπορούμε σε ένα οπτικοακουστικό περιεχόμενο που μας δίνεται να περιγράψουμε τις ενέργειες και να λάβουμε μία λίστα από σενάρια, στα οποία θα περιλαμβάνονται παρόμοιες πράξεις.

Οι εφαρμογές του MPEG-7 που προβλέπονται από το ISO, καθώς και οι προδιαγραφές για την εκπλήρωσή τους είναι ποικίλες και καλύπτουν πολλούς τομείς. Για μία καλύτερη κατανόηση του τι δυνατότητες παρέχει το MPEG-7 και ποιες είναι οι λειτουργίες του, θα παρατεθούν κάποιες εφαρμογές που έχουν σχέση με την εικόνα. Αυτό δεν δηλώνει κάποια προτεραιότητα ως προς τις εφαρμογές, απλά εκπληρώνει τους στόχους αυτής της διπλωματικής εργασίας.

MPEG-7 και βάσεις δεδομένων

Σε ότι αφορά λοιπόν το οπτικοακουστικό περιεχόμενο, το MPEG-7 ξεκίνησε σαν ένα πρότυπο που θα μπορούσε να καταστήσει το περιεχόμενο αυτό εύκολο στην αναζήτηση, όπως συμβαίνει και στην απλή κειμενική (textual)

πληροφορία. Πράγματι μπορεί να γίνει εύκολα αντιληπτό, ότι η δομή και η αυστηρότητα που απαιτείται για την περιγραφή ακόμα και της ελάχιστης πολυμεσικής πληροφορίας, είναι αρκετά πιο δύσκολη σε σχέση με ένα απλό, γραπτό κείμενο.

Οι μέθοδοι ανάκτησης επομένως πρέπει να περιλαμβάνουν βάσεις δεδομένων, οπτικοακουστικά αρχεία και τις υπηρεσίες του Internet, που βασίζονται στον Παγκόσμιο Ιστό (παράδειγμα, αναζήτηση μιας εικόνας από μία μηχανή αναζήτησης). Τυπικό παράδειγμα αυτής της κατηγορίας αποτελούν τα αρχεία από ταινίες και διάφορες τηλεοπτικές εκπομπές, που αποθηκεύουν τεράστιες ποσότητες πολυμεσικού περιεχομένου σε διάφορα μέσα (CD-ROM, DVD κ.τ.λ.) μαζί με ακριβή περιγραφική πληροφορία (μετα-δεδομένα). Αυτά τα μετα-δεδομένα βρίσκονται αποθηκευμένα σε βάσεις δεδομένων, όπου υπάρχει ένα αρκετά μεγάλο ενδιαφέρον για την αποθήκευση και ανταλλαγή των περιγραφών που θα μπορούσαν να διασφαλίσουν:

1. Την διαλειτουργικότητα ανάμεσα σε χειριστές αρχείων video
2. Την ουσιαστική σημασία των μετα-δεδομένων και
3. Μία ευρύτερη εξάπλωση των δεδομένων τόσο στο επιστημονικό όσο και στο ευρύ κοινό.

Τι περιλαμβάνουν οι προδιαγραφές

Οι προδιαγραφές του προτύπου του MPEG-7 για τέτοιου είδους εφαρμογές περιλαμβάνουν:

- Περιγραφές πλήρους κειμένου και δομημένα πεδία (περιγραφές της βάσης δεδομένων).
- Έναν μηχανισμό με τον οποίο διαφορετικές περιγραφές του MPEG-7 θα μπορούν να υποστηρίξουν την ικανότητα για δια-λειτουργικότητα ανάμεσα σε διαφορετικές σημασιολογίες
- Έναν αυτόνομο μηχανισμό

- Μία δομή για να χειρίζεται πολλαπλές εκδόσεις του ίδιου τεκμηρίου κατά τα διάφορα στάδια της παραγωγής του και περιγραφές που να μπορούν να εφαρμοστούν σε πολλαπλά αντίγραφα του ίδιου υλικού.

Εφαρμογή Ήχου

Μία παρόμοια κατάσταση συναντάμε και στις βάσεις δεδομένων που αποθηκεύουν ήχο (audio databases). Η μουσική βιομηχανία προσπαθεί να προσεγγίσει όσο το δυνατό περισσότερους πελάτες, από τους οποίους ο καθένας μπορεί να έχει ποικίλες μουσικές προτιμήσεις. Οι ερευνητές-καταναλωτές μπορεί να αναζητούν ένα τραγούδι βασισμένοι σε μία σχετική απόδοση του τραγουδιού αυτού από ένα κατάστημα ή από το σπίτι τους. Εναλλακτικά θα μπορούσαν να ζητήσουν ένα μουσικό κομμάτι με χαρακτηριστικά που ήδη γνωρίζουν (όπως για παράδειγμα το όνομα του τραγουδιστή) και από εκείνο το σημείο μπορούν να ακούσουν το κατάλληλο δείγμα (και ίσως να δουν σε αυτό συνδεδεμένη πληροφορία, όπως τους στίχους ή το video clip) και να το αγοράσουν.

Οι εξειδικευμένες απαιτήσεις για τέτοιου είδους εφαρμογές, που έχουν σχέση με τον ήχο είναι οι εξής

- Ένας μηχανισμός που να μπορεί να υποστηρίξει μελωδία και άλλα μουσικά χαρακτηριστικά, που θα επιτρέπουν το περιθώριο λάθους, για κάποιον που θα κάνει αναζήτηση ενός μουσικού τραγουδιού με βάση ένα απλό μουρμουρητό (query-by-humming).
- Έναν μηχανισμό που θα υποστηρίξει τους περιγραφείς (descriptors) που βασίζονται στην πληροφορία που είναι συνδεδεμένη με τα δεδομένα
- Θα πρέπει να υποστηρίζονται σχήματα περιγραφής (description schemes) που να περιέχουν περιγραφείς οπτικών, ηχητικών και άλλων χαρακτηριστικών, καθώς και να παρέχει συνδέσμους ανάμεσα στα διάφορα μέσα.
- Εκτός από την μουσική βιομηχανία, ενδιαφέρουσες εφαρμογές που έχουν σχέση με τον ήχο είναι βάσεις δεδομένων που περιέχουν ομιλίες ιστορικών προσώπων.

Ωθηση Απώθηση

Μία άλλη εφαρμογή που παρέχει το MPEG-7 είναι ότι υποστηρίζει τρόπους απόκτησης πληροφοριών ώθησης – απώθησης (pull and push information). Υπάρχουν δύο μεγάλες ομάδες πληροφορίας. Στην πρώτη ομάδα ανήκουν οι πληροφορίες για τις οποίες συλλέγουμε όλο και περισσότερα στοιχεία γι' αυτές ("pull" information) και στην δεύτερη ομάδα ανήκουν οι πληροφορίες από τις οποίες εξάγουμε στοιχεία, ανάλογα με το τι ακριβώς ζητάμε ("push" information). Στην πρώτη κατηγορία ανήκει η γενική έρευνα (search), ενώ στην δεύτερη κατηγορία ανήκει το φιλτράρισμα (filtering) της πληροφορίας, που είναι το αντίστροφο της έρευνας.

Και οι δύο αυτοί τομείς επωφελούνται αρκετά από την χρήση του MPEG-7, ενώ βρίσκουν εφαρμογή σε τηλεοπτικές μεταδόσεις (broadcasting) και στις μεταδόσεις μέσω ιστού (Webcasting). Αυτά τα πεδία έχουν ξεκάθαρες απαιτήσεις από το πρότυπο και σχετίζονται περισσότερο με περιγραφές ροής (streaming description) , παρά με στατικές περιγραφές, που αποθηκεύονται σε βάσεις δεδομένων.

Παραδείγματα τέτοιων εφαρμογών θα μπορούσαν να είναι, από την πλευρά του χρήστη, μία εξατομικευμένη επιλογή οπτικοακουστικών προγραμμάτων. Από την πλευρά του δημιουργού, παράδειγμα θα μπορούσε να είναι η εξειδικευμένη επιλογή πρωτογενούς υλικού από κάποια αρχεία. Οι προδιαγραφές του MPEG-7 για τέτοιου είδους εφαρμογές περιλαμβάνουν:

- Υποστήριξη για χρήση περιγραφών και περιγραφικών σχημάτων που επιτρέπουν την χρήση πολλαπλών γλωσσών.
- Μηχανισμούς με τους οποίους κάποιο μέσο (a media object) θα μπορεί να αναπαρίσταται από ένα σύνολο εννοιών που μπορεί να στηρίζονται στην περιοχή (locality) ή στην γλώσσα.
- Υποστήριξη για ικανοποιητικούς διαδραστικούς χρόνους ανταπόκρισης (efficient interactive response times).

Ως προς αυτές τις εφαρμογές, ένα ακόμη ενδιαφέρον στοιχείο, είναι ότι η επιλογή και το φιλτράρισμα της πληροφορίας διευκολύνουν την δυνατότητα πρόσβασης στην διαθέσιμη πληροφορία σε όλους τους χρήστες και ειδικά σε

όσους έχουν κάποια αναπηρία, όπως νοητική, κινητική, οπτική. Με το να παρέχουμε την αναπαράσταση της πληροφορίας ενεργά (active) μπορούμε να ξεπεράσουμε τέτοια προβλήματα. Το στοιχείο κλειδί σε αυτές τις περιπτώσεις είναι να επιτρέψουμε να υπάρξουν πολλαπλές μορφές επικοινωνίας, ώστε να παρουσιαστεί η πληροφορία με τον κατάλληλο κάθε φορά τρόπο ανάλογα με την ικανότητα του κάθε χρήστη.

Με την εύκολη πρόσβαση στις ακολουθίες ενός video (video sequences) που παρέχεται μέσα από τις βάσεις δεδομένων, το MPEG-7 μπορεί να προωθήσει την χρήση ήχου και εικόνας σε τομείς που έχουν σχέση με την εκπαίδευση. Τέτοιοι τομείς είναι η ιστορία, η τέχνη κ.ά. Όσον αφορά για παράδειγμα την τέχνη, οι προτυποποιημένοι περιγραφείς μπορούν να αποφέρουν μία συλλογή από σχετικά τεκμήρια μέσα στην αίθουσα, για συγκεκριμένες κατηγορίες θεατρικών έργων ή συγκεκριμένων μουσικών. Μπορούν επομένως οι χρήστες με αυτόν τον τρόπο να συγκρίνουν διάφορες παραγωγές θεατρικών έργων ή μουσικών κομματιών.

Απαιτήσεις εφαρμογών

Οι απαιτήσεις που προβλέπονται από το MPEG-7 και που απορρέουν από τις παραπάνω εφαρμογές είναι οι εξής:

- Μηχανισμούς με τους οποίους οι περιγραφείς μπορούν να συνδέονται με εξωτερική πληροφορία (external information), όπως οι υπηρεσίες του World Wide Web και η HTML (Hypertext Markup Language).
- Υποστήριξη για δια-λειτουργικότητα μεταξύ των περιγραφέων.
- Η ικανότητα να επιτρέπονται λειτουργίες σε πραγματικό χρόνο σε συνδυασμό με τις βάσεις δεδομένων.
- Δείκτες (pointers) που να αναφέρονται απευθείας στα δεδομένα, ώστε να δίνεται η δυνατότητα στους χρήστες να επεξεργάζονται τα πολυμέσα.

Η περιγραφή του MPEG-7

Στόχοι του MPEG-7:

Οι στόχοι που καλύπτει και προβλέπεται να καλύψει το MPEG-7 είναι τρεις:

- Την περιγραφή του πολυμεσικού περιεχομένου.
- Την ευέλικτη διαχείριση των δεδομένων.
- Την δια-λειτουργικότητα των συστημάτων

Περιγραφή πολυμεσικού περιεχομένου (Multimedia Content Description)

Ο πιο σημαντικός στόχος του MPEG-7 (ο ουσιαστικός λόγος για την ανάπτυξή του) είναι να παρέχει ένα σύνολο με προτυποποιημένες μεθόδους και εργαλεία για τις διάφορες κατηγορίες του πολυμεσικού περιεχομένου. Εκτός από την διάκριση του υλικού σε κατηγορίες οπτικού και ακουστικού ή συνδυασμό αυτών, γίνεται και διάκριση στις κλάσεις περιγραφών (description classes).

Όταν αναφερόμαστε σε κλάσεις περιγραφών, στην πραγματικότητα εννοούμε τις διάφορες πιθανές πλευρές που θα μπορούσε να καλύψει μία περιγραφή οπτικοακουστικού περιεχομένου. Ένα βασικό στοιχείο είναι ότι υπάρχουν πολλοί τρόποι για να περιγράψουμε μία οντότητα, ανάλογα με το πως αυτή θα χρησιμοποιηθεί. Για αυτόν τον λόγο το MPEG-7 πρέπει να συμβιβάσει αυτές τις περιγραφές και να τις καταστήσει συμπληρωματικές την μία ως προς την άλλη, παρά να χρησιμοποιηθεί αποκλειστικά μόνο μία.

Υπάρχουν τέσσερις θεμελιώδεις κλάσεις για την περιγραφή του περιεχομένου, που σχετίζονται περισσότερο με τα δεδομένα, δηλαδή με το υλικό που πρόκειται να περιγραφεί, παρά μεταξύ τους. Οι τέσσερις αυτές κατηγορίες είναι:

- Η εγγραφή (transcriptive),
- Η φυσική
- Η νοητική και

- Η περιγραφή που βασίζεται στο μέσο (medium-based description).

Στην κορυφή αυτών των σχημάτων βρίσκεται η περιγραφή της αρχιτεκτονικής τους, που σχεδιάζει τις σχέσεις ανάμεσα σε μεγάλες ενότητες δεδομένων καθώς και τις σχέσεις μεταξύ των περιγραφών και μέσα στις ίδιες τις περιγραφές. Η περιγραφή της σήμανσης (the annotative description), όπως ονομάζεται, βρίσκεται στην κορυφή όλων των κατηγοριών και αγγίζει την κάθε μία ξεχωριστά.

Το πιο πιθανό πάντως είναι ότι στην πραγματικότητα κάθε περιγραφή που θα χρησιμοποιείται στο MPEG-7 θα περιλαμβάνει μόνο μία ή δύο από αυτές τις κλάσεις. Στη συνέχεια θα αναλυθούν οι διαφορετικοί πιθανοί τύποι περιγραφών που υπάρχουν με περισσότερη λεπτομέρεια.

Περιγραφείς

Ο «περιγραφέας που βασίζεται στο μέσο» (medium based) είναι εκείνος που περιγράφει το μέσο με το οποίο εκφράζονται τα δεδομένα. Τέτοιος περιγραφέας θα μπορούσε να είναι ο ρυθμός δειγματοληψίας ενός ψηφιακού αρχείου (sampling rate), εάν είναι αναλογική ή ψηφιακή η πηγή, σε ποιο σημείο κόβονται οι σκηνές, σε ποιο σημείο εστιάζει κάθε φορά η κάμερα, και θα μπορούσαν να παρατεθούν και πολλά άλλα στοιχεία. Ένας περιγραφέας που βασίζεται στο μέσο μπορεί να πραγματεύεται τέτοια είδη στοιχείων χαμηλού επιπέδου, επιφανειακά χαρακτηριστικά (surface features) που περιγράφουν την εγγραφή ή την επαναπροβολή (playback) του ίδιου του μέσου. Υπάρχουν αρκετές τεχνικές για να αποκτηθεί αυτού του είδους η περιγραφή, μέσω της ανάλυσης του ήχου ή της εικόνας. Η κωδικοποίηση αυτών των περιγραφών θα μπορούσε να γίνει πολύ πιο εύκολα κατά την επεξεργασία της δημιουργίας του υλικού.

Ο «φυσικός περιγραφέας» (physical) είναι ο περιγραφέας που ενδεχομένως θα μπορούσε να συνδυαστεί και με την υποκειμενική πλευρά και αυτή η προσέγγιση μπορεί να καλύψει όλες εκείνες τις πλευρές που έχουν σχέση με τα υπολογιστικά χαρακτηριστικά, που δεν ανταποκρίνονται στην ανθρώπινη αντίληψη. Πρακτικά μιλώντας θα μπορούσαμε να εξάγουμε αυτά τα χαρακτηριστικά εύκολα από τα ακατέργαστα πολυμεσικά δεδομένα. Αυτά τα

δεδομένα έχουν αδιαμφισβήτητες τιμές και καλά εδραιωμένους αλγόριθμους για την εξαγωγή των δεδομένων.

Ο «περιγραφέας που βασίζεται στην υποκειμενική αντίληψη» (perceptual class): Η υποκειμενική αντίληψη τεμαχίζει τα μέσα σε αντικείμενα, δηλαδή αναφέρεται σε περιγραφές τέτοιων χαρακτηριστικών όπως το χρώμα, την υφή και το σχήμα. Τα αντικείμενα περιγράφουν την εικόνα ή τον ίδιο τον ήχο και δεν πραγματεύονται γενικά την δομή της καταγραφόμενης (ή δημιουργούμενης) σκηνής μέσα στα δεδομένα.

Ο «περιγραφέας της εγγραφής ή η κλάση της εγγραφής (transcription class): τυπικά αναπαριστά μία επαναδόμηση (reconstruction) της δομής του κόσμου, όπως αυτή συλλαμβάνεται από τα δεδομένα.. Στην μουσική αυτή η κλάση εξυπηρετεί σαν την πραγματική αντιγραφή της μουσικής, δηλαδή οι νότες που παίζονται. Φυσικά το να καταφέρνουμε τον εμπλουτισμό της δομής από ένα απλό video ή ήχο (χωρίς κάποια άλλη πληροφορία) προάγει την σημερινή τεχνολογία. Υπάρχουν περισσότεροι από ένας τρόποι για να αποκτήσουμε μία τέτοια περιγραφή.

Ο «περιγραφέας της αρχιτεκτονικής» βρίσκεται πάνω από τις τρεις προαναφερθείσες. Αυτός ο περιγραφέας περιέχει τα δομικά στοιχεία των άλλων κλάσεων και επομένως τα δεδομένα που αυτές περιγράφουν. Αυτή η κλάση περιγράφει την περιοχή των τεκμηρίων, στην οποία ένας χρήστης ενδιαφέρεται να μάθει μόνο τις σχέσεις μεταξύ των τμημάτων και όχι το τι αυτά περιέχουν. Αυτή είναι επίσης και η κλάση που καθορίζει την συντακτική δομή, που πρέπει να στηρίζεται απαραίτητα σε χαμηλότερα σημασιολογικά επίπεδα. Αυτό το είδος της περιγραφής μπορεί να αποκτηθεί με πολλούς τρόπους. Οι περιγραφές μπορούν να παραχθούν με το χέρι, μπορούν να είναι υποπροϊόντα μίας αυτοματοποιημένης ανάλυσης ή να υποδηλώνονται σε ένα χειρόγραφο αντίγραφο και απλά να εξυπηρετούν στο να κάνουν αυτές τις σχέσεις πιο ξεκάθαρες.

Ο «περιγραφέας της σήμανσης (annotative) βρίσκεται τοποθετημένος στη κορυφή όλων των άλλων κλάσεων, όπως επίσης στην κορυφή βρίσκονται και τα ίδια τα δεδομένα. Είναι ο τομέας της σήμανσης και της ανάλυσης άλλων ήδη υπαρχόντων μετα-δεδομένων που πραγματοποιούνται από τον ίδιο τον

άνθρωπο. Στην πραγματική της χρήση, η περιγραφή αυτή, στη μουσική, υπήρξε ο χώρος για τους μουσικολόγους, στον οποίο μπορούσαν να πραγματοποιήσουν τα σχόλιά τους για διάφορα χαρακτηριστικά ενός μουσικού κομματιού, όπως γενικά ζητήματα μουσικολογικής ανάλυσης. Συνηθισμένα ζητήματα που αφορούν την μουσικολογία περιλαμβάνουν μουσικές φόρμες (που αναφέρονται στην αρχιτεκτονική του μουσικού κομματιού), τις ίδιες τις νότες (που συνδέονται με το αντίγραφο του μουσικού κομματιού), τα σχόλια σχετικά με το συναισθηματικό περιεχόμενο της μουσικής (που αναφέρονται στα ίδια τα δεδομένα) και τις σχέσεις που υπάρχουν με άλλα μουσικά κομμάτια (που συνδέονται με άλλες περιγραφές).

Ρυθμιστικοί κανόνες

Το MPEG-7 προσπαθεί να παρέχει εκείνους τους ρυθμιστικούς κανόνες για να υποστηρίξει τις απαιτήσεις που έχουν οι διαφορετικές κλάσεις περιγραφών. Είναι αξιοπρόσεκτο ότι παρά την πολυπλοκότητα των εφαρμογών, που μπορεί να φαίνεται μεγάλη και αποθαρρυντική, η ποικιλία των επιπέδων περιγραφής παρέχει ένα πλήρες σύνολο, που επιτρέπει ευέλικτους και εκφραστικούς τρόπους, για την επαρκή αναπαράσταση του περιεχομένου μέσω κάποιας διαμορφωμένης δομής.

Για να επαναχρησιμοποιήσουμε τους περιγραφείς του MPEG-7 με ικανοποιητικό τρόπο, ο χρήστης θα χρειαστεί να τους προσαρμόσει στις συγκεκριμένες ανάγκες του. Αυτό οδηγεί στην τροποποίηση και επεξεργασία των ήδη υπάρχοντων δομών. Η επεξεργασία της δομής του τεκμηρίου θα ευνοήσει διαδικασίες όπως την διάβαση (traversal) και την επεξεργασία των ιεραρχικών δέντρων, συνδεδεμένες μεταξύ τους λίστες και άλλες εργασίες.

Αυτό που προβλέπεται όμως είναι ότι η βαθύτερη δομή των δεδομένων και η σύνθεσή τους θα παραμείνουν ανεξάρτητα από τους εφαρμοσμένους μηχανισμούς εξαγωγής δεδομένων (applied extraction mechanisms). Με άλλα λόγια οι δομές του MPEG-7 παρέχουν ένα πλαίσιο περιγραφών το οποίο είναι ανεξάρτητο από την εφαρμογή που προκύπτει κάθε φορά και στο οποίο οι μηχανισμοί εξαγωγής θα μπορούν να του αποδίδουν στιγμιότυπα (instantiate).

Οποιαδήποτε χαρακτηριστικά περιγράφουν ένα οπτικοακουστικό τεκμήριο εξάγονται είτε αυτοματοποιημένα με την χρήση ενός αλγορίθμου που θα χρησιμοποιεί ο υπολογιστής είτε με την σήμανση που πραγματοποιείται από κάποιον άνθρωπο που είναι ειδικός σε τέτοιες εφαρμογές.

Για να πραγματοποιηθεί με αυτοματοποιημένο τρόπο μια τέτοια εργασία απαιτείται μία επίσημη προδιαγραφή της οντότητας ή των χαρακτηριστικών που θα εξαχθούν. Η προδιαγραφή αυτή μπορεί να είναι ατομική (atomic) ή μπορεί να προέρχεται από την εξαγωγή ενός αριθμού χαρακτηριστικών. Παραδείγματα τέτοιων χαρακτηριστικών από τον χώρο της εικόνας θα μπορούσε να είναι η σύνθεση μίας εικόνας.

Συμπερασματικά, αυτό που γίνεται κατανοητό, είναι ότι οι περιγραφείς του MPEG-7 κατέχουν ένα κεντρικό ρόλο στην λειτουργία του προτύπου και για αυτόν τον λόγο υπάρχει και η πρόβλεψη όσο το δυνατό περισσότερων περιγραφέων με σκοπό να καλυφθούν περισσότερες εφαρμογές με τον καλύτερο και πιο ακριβή τρόπο.

Στη συνέχεια ακολουθεί η ανάλυση του δεύτερου στόχου του MPEG-7, που είναι η ευελιξία στην διαχείριση των δεδομένων.

Ευελιξία διαχείριση των δεδομένων (Flexibility in data management)

Το MPEG-7 παρέχει ένα πλαίσιο όπου επιτρέπεται να γίνονται αναφορές στα τμήματα ενός τεκμηρίου, σε ολόκληρο το τεκμήριο, ή σε μία σειρά από τεκμήρια. Με την ευελιξία στην διαχείριση των δεδομένων στηρίζεται και η πολλαπλή μορφή που μπορεί να έχει ένα τεκμήριο. Αυτό σημαίνει ότι θα μπορούσε να είναι πιθανή η περιγραφή πολυμεσικού περιεχομένου με τέτοιο τρόπο που να επιτρέπει να θέτουμε ερωτήματα βασισμένα σε ηχητικά δεδομένα και να ανακτούμε οπτικά δεδομένα, καθώς και το αντίστροφο.

Επιπρόσθετα, η ευελιξία αυτή βοηθά στο να διασφαλιστεί η μακροβιότητα του προτύπου, που σημαίνει ότι τα σχήματα των περιγραφέων που χρησιμοποιούνται για μία συγκεκριμένη εργασία, μπορούν να τροποποιηθούν για μία διαφορετική, αλλά σχετική εφαρμογή. Το MPEG-7 πραγματεύεται εφαρμογές που μπορούν να αποθηκευτούν (online ή offline) ή να είναι με την

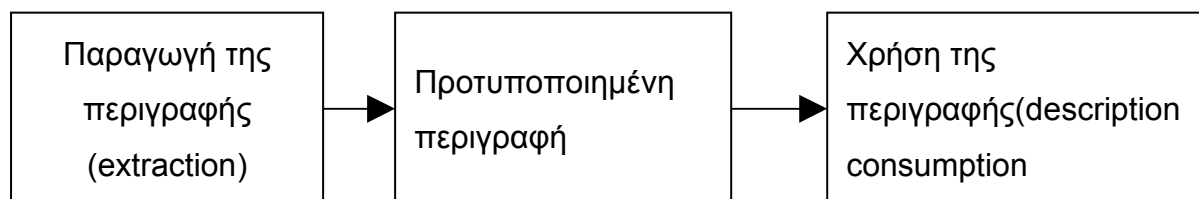
μορφή της ροής (streamed) (όπως για παράδειγμα συμβαίνει στην αναμετάδοση) και μπορούν να λειτουργήσουν ταυτόχρονα σε περιβάλλοντα πραγματικού χρόνου και σε περιβάλλοντα μη πραγματικού χρόνου .

Η καθολικότητα των πηγών των δεδομένων (Globalization of Data Resources).

Οι περιγραφείς του MPEG-7 μπορούν να εντοπισθούν μαζί με το συνδεδεμένο με αυτές οπτικοακουστικό υλικό στην ίδια ροή δεδομένων (data stream) ή στο ίδιο αποθηκευτικό σύστημα όμως οι περιγραφείς θα μπορούσαν να υπάρξουν και αλλού. Όταν το περιεχόμενο και οι περιγραφείς του δεν βρίσκονται μαζί, θα μπορούσαν να αποδειχθούν χρήσιμοι μηχανισμοί που συνδέουν το οπτικό ακουστικό υλικό μαζί με τους περιγραφείς τους. Αυτοί οι σύνδεσμοι μπορούν να χρησιμεύουν και στις δύο κατευθύνσεις. Ο συνδυασμός της ευελιξίας και της καθολικότητας επιτρέπει τόσο στους ανθρώπους όσο και στις μηχανές να ανταλλάσσουν, να ανακτούν και να επαναχρησιμοποιούν το σχετικό υλικό.

Το βασικό είναι ότι το MPEG-7 υπάρχει για να εξυπηρετεί αυτόν τον στόχο. Η προτυποποίηση κυρίως αναζητεί τρόπους για να φτάσει πέρα από κάθε απλή, αυθαίρετη λύση και να παρέχει όχι μόνο ένα πλαίσιο, αλλά ένα στέρεο μέσο με το οποίο οι λύσεις που προτείνουν οι βιομηχανίες θα μπορούσαν να συνεργαστούν. Εάν το MPEG-7 γίνει ακόμη πιο εξειδικευμένο ή αρκετά γενικό ως προς τις εφαρμογές του, τότε αυτός ο στόχος, που είναι και κύριος, θα αποτύχει.

Η παρακάτω εικόνα απεικονίζει την διαδικασία της επεξεργασίας του MPEG-7 με στόχο να εξηγήσει την λειτουργία του προτύπου με έναν απλό τρόπο. Στη συνέχεια ακολουθεί η επόμενη ενότητα, που έχει σχέση με την παρουσίαση και ανάλυση των ρυθμιστικών στοιχείων του MPEG-7.



Εικόνα 1: Διαδικασία επεξεργασίας του MPEG-7.

Ορολογία του MPEG-7

Στην παραπάνω ενότητα αναφέρθηκαν οι εφαρμογές στις οποίες μπορεί να χρησιμοποιηθεί το MPEG-7. Προκειμένου να χειριστεί μία τέτοια σειρά από εφαρμογές, το MPEG-7 παρέχει ορισμένα ρυθμιστικά στοιχεία, που περιλαμβάνουν τους Περιγραφείς (Descriptors), τα Σχήματα Περιγραφής (Description Schemes), την Γλώσσα Προσδιορισμού της Περιγραφής (Description Definition Language) και προδιαγραφές που πραγματεύονται ζητήματα σχετικά με τα συστήματα (systems) του MPEG-7. Στη συνέχεια θα διασαφηνιστεί τι σημαίνουν αυτοί οι όροι μέσα στο MPEG-7 και για να γίνει καλύτερη η περιγραφή και η κατανόησή τους θα χρησιμοποιηθεί η ορολογία, όπως αυτή ορίζεται μέσα από το έγγραφο που ρυθμίζει τις προδιαγραφές του MPEG-7 (ISO/MPEG N2859, MPEG-7 Requirements Document).

Δεδομένα (Data)

Ο πλέον στοιχειώδης και απαραίτητος ορισμός για να ξεκινήσουμε το κεφάλαιο αυτό είναι τα δεδομένα. (data). Ως δεδομένα στο MPEG-7 ορίζεται η οπτικοακουστική πληροφορία η οποία θα χρησιμοποιηθεί στο MPEG-7 άσχετα με την αποθήκευση, την κωδικοποίηση, την εμφάνιση, την μετάδοση, το μέσο ή την τεχνολογία.

Αυτός ο ευρύς ορισμός περιλαμβάνει τα γραφικά, την στατική ή κινούμενη εικόνα, το video, τα film, τον ήχο, την μουσική, μία ομιλία και γενικά οτιδήποτε μπορεί να αποτελέσει ένα οπτικοακουστικό τεκμήριο. Παραδείγματα δεδομένων για το MPEG-7 αποτελούν ένα DVD, ένα CD που μπορεί να περιλαμβάνει ομιλίες ιστορικών προσώπων ή απλά ήχο, μία εικόνα αποτυπωμένη σε χαρτί ή ακόμα ένα πρόγραμμα εγκατάστασης πολυμέσων στο Web.

Χαρακτηριστικά (Features)

Ως χαρακτηριστικό ορίζεται ένα διακριτό χαρακτηριστικό των δεδομένων που δηλώνει κάτι σε κάποιον.

Το χαρακτηριστικό δεν είναι ένας καθιερωμένος όρος της επεξεργασίας σημάτων, αλλά είναι πιο πολύ κάτι που βρίσκεται στο μυαλό μας. Για αυτό,

λοιπόν, σε ένα σύνολο δεδομένων δεν μπορούμε να συγκρίνουμε τα χαρακτηριστικά μεταξύ τους, χωρίς να τους αποδώσουμε πρώτα μία αναπαράσταση τους (δηλαδή τους περιγραφείς, descriptors), που να εμπεριέχει σημασία, και χωρίς την απόδοση στιγμιοτύπων σε αυτά (instantiations), δηλαδή τις τιμές που έχει ένας περιγραφέας (descriptor value).

Παραδείγματα χαρακτηριστικών αποτελούν το είδος σε ένα μουσικό κομμάτι, ο τίτλος μιας ταινίας και οι ηθοποιοί, χρώμα, η υφή και το σχήμα σε μία εικόνα.

Περιγραφείς και Τιμές Περιγραφέων (Descriptors and Descriptors Values)

Ως περιγραφέας ορίζεται η τιμή ενός χαρακτηριστικού. Ένας περιγραφέας (descriptor) καθορίζει την σύνταξη και την σημασιολογία της αναπαράστασης του χαρακτηριστικού.

Για να λειτουργήσει ένας περιγραφέας στο MPEG-7, πρέπει να προσδιορίσει με ακρίβεια την σημασιολογία του χαρακτηριστικού, τον συνδεδεμένο με αυτόν τύπο δεδομένων, τις επιτρεπτές τιμές και μία ερμηνεία των τιμών των περιγραφέων (descriptor values). Ένα παράδειγμα θα μπορούσε να είναι το εξής. Το χρώμα για παράδειγμα είναι ένα αλφαριθμητικό. Ο τύπος δεδομένων μπορεί να είναι σύνθετος, το οποίο σημαίνει ότι μπορεί να διαμορφωθεί με το να συνδέσουμε αλυσιδωτά πολλαπλές τιμές ενός τύπου δεδομένων, π.χ. RGB color: {integer, integer, integer}.

Το MPEG-7 πραγματεύεται διάφορα επίπεδα αφαίρεσης. Στο χαμηλότερο επίπεδο αφαίρεσης, οι περιγραφείς μπορεί να περιλαμβάνουν χαρακτηριστικά όπως η μορφή, το χρώμα, το μέγεθος για τις εικόνες και το video, η αρμονία και το ηχόχρωμα για την μουσική.

Στο υψηλό επίπεδο μπορεί να βρίσκονται τα γεγονότα που συμβαίνουν, οι σχέσεις μεταξύ των ατόμων, κάποιες αφηρημένες έννοιες. Οι περιγραφείς ήχου και εικόνας αναπαριστούν εξειδικευμένα χαρακτηριστικά που σχετίζονται με ακουστικό και οπτικό περιεχόμενο αντίστοιχα. Υπάρχουν επίσης και οι

γενικοί περιγραφείς (generic descriptions) που περιγράφουν τα γενικά χαρακτηριστικά.

Είναι πιθανό να έχουμε ποικίλους περιγραφείς που να αναπαριστούν ένα μοναδικό χαρακτηριστικό και αυτό σημαίνει ότι πραγματεύεται διαφορετικές σχετικές απαιτήσεις. Παραδείγματα πολλαπλών περιγραφέων (multiple descriptors) για ένα χαρακτηριστικό αποτελούν οι αριθμητικές λίστες και τα ιστογράμματα με χρώματα.

Σε ένα σύνολο δεδομένων (ή και σε ένα υποσύνολό τους) μία τιμή ενός περιγραφέα (description value) είναι η απόδοση της τιμής (instantiation) που δίδεται σε αυτόν. Οι τιμές συνδυάζονται μέσω ενός Σχήματος Περιγραφέων (Description Scheme) ώστε να δημιουργήσουν μία περιγραφή.

Σχήματα Περιγραφέων (Description Schemes)

Ένα σχήμα περιγραφέων προδιαγράφει την δομή και την σημασιολογία των σχέσεων ανάμεσα στα συστατικά στοιχεία του, τα οποία μπορεί να είναι ταυτόχρονα περιγραφείς και σχήματα περιγραφής.

Η διάκριση ανάμεσα στα σχήματα περιγραφέων και στους περιγραφείς είναι η εξής. Ένας περιγραφέας πραγματεύεται την αναπαράσταση και την παρουσίαση ενός χαρακτηριστικού. Το σχήμα περιγραφέων από την άλλη ασχολείται με την δομή μιας περιγραφής. Γίνεται αντιληπτό λοιπόν ότι οι περιγραφείς και τα σχήματα περιγραφής αναφέρονται σε διαφορετικές πλευρές και για αυτό αποτελούν δύο πολύ διαφορετικές έννοιες.

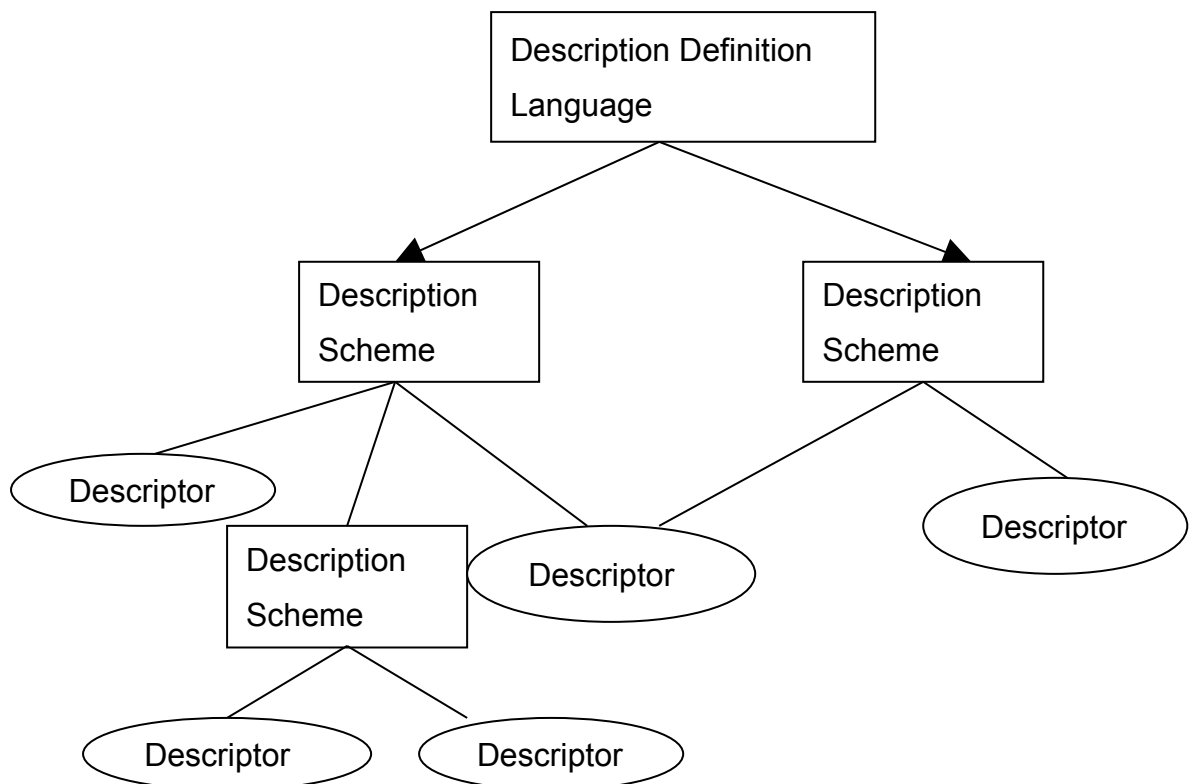
Τα σχήματα περιγραφέων επιτρέπουν την δημιουργία πολύπλοκων περιγραφών με το να προσδιορίζουν την δομή και την σημασιολογία των σχέσεων ανάμεσα στους περιγραφείς και στα σχήματά τους. Για παράδειγμα, ένα σχήμα περιγραφής για ένα τμήμα μιας εικόνας μπορεί να προσδιορίζει την σύνταξη και την σημασιολογία των συστατικών στοιχείων του, όπως την βαθύτερη αποσύνθεση των τμημάτων, τα χαρακτηριστικά κάθε τμήματος ξεχωριστά και τις σχέσεις ανάμεσα στα συστατικά στοιχεία.

Όπως συμβαίνει και με τους περιγραφείς, έτσι και τα σχήματα περιγραφέων μπορούν να κατηγοριοποιηθούν σε σχήματα για τον ήχο, την εικόνα, ή να είναι γενικά. Τα γενικά Σχήματα συνήθως αναπαριστούν μία γενική μετα-

πληροφορία που σχετίζεται με όλα τα είδη των μέσων (ήχο, εικόνα, κείμενο, γραφικά, κ.τ.λ.)

Σε μία περιγραφή εκτός από τους Περιγραφείς και τα Σχήματα Περιγραφής που προέρχονται εσωτερικά από το κείμενο, το MPEG-7 περιλαμβάνει και Περιγραφείς και Σχήματα Περιγραφής που έχουν σχέση με την δημιουργία, την παραγωγή και την διαχείριση του οπτικοακουστικού υλικού . Αυτά τα μετα-δεδομένα μπορεί να περιέχουν πληροφορίες για τις προϋποθέσεις της πρόσβασης στο υλικό (για παράδειγμα, πληροφορία για τα δικαιώματα της πνευματικής ιδιοκτησίας του υλικού), για την ταξινόμηση (να περιλαμβάνει ταξινόμηση του υλικού σε έναν αριθμό προκαθορισμένων κατηγοριών) και συνδέσμους με άλλα σχετικά στοιχεία (όπου και η πληροφορία αυτή πιθανό να κάνει πιο σύντομη την έρευνα του χρήστη).

Η εικόνα που ακολουθεί αναπαριστά τις πιθανές σχέσεις που υπάρχουν ανάμεσα στους Περιγραφείς και στα Σχήματα Περιγραφής.



Εικόνα 2: Πιθανές σχέσεις ανάμεσα σε Περιγραφείς και Σχήματα Περιγραφής.

Περιγραφή

Ως περιγραφή ορίζεται μία ικανότητα που αποτελείται από ένα Σχήμα περιγραφής και από το σύνολο των στιγμιοτύπων των περιγραφών που περιγράφουν τα δεδομένα.

Μία περιγραφή περιλαμβάνει ή αναφέρεται σε ένα Σχήμα Περιγραφής που του έχει αποδοθεί (instantiated) πλήρως ή μερικώς μία τιμή. Με άλλα λόγια μία περιγραφή είναι η απόδοση τιμής (instance) ενός Σχήματος Περιγραφής.

Κωδικοποιημένη Περιγραφή (Coded Description)

Ως κωδικοποιημένη περιγραφή ορίζεται: «Η περιγραφή η οποία έχει κωδικοποιηθεί, για να εκπληρώσει τις σχετικές προδιαγραφές, όπως η αποτελεσματική συμπίεση, η ελαστικότητα στα λάθη, η τυχαία προσπέλαση και ούτω καθεξής.

Γλώσσα Προσδιορισμού της Περιγραφής (Description Definition Language, DDL)

Ορισμός: «Η DDL επιτρέπει την δημιουργία νέων Σχημάτων Περιγραφής, και πιθανόν, και Περιγραφών. Επιτρέπει επίσης την επέκταση και την τροποποίηση των ήδη υπαρχόντων Σχημάτων Περιγραφής.»

Η DDL στο MPEG-7 επιτρέπει τον ευέλικτο προσδιορισμό των περιγραφών και των Σχημάτων Περιγραφής βασισμένη στο XML Schema. Οι περιγραφείς και τα σχήματα Περιγραφής που υπάρχουν είναι ανεξάρτητα της εφαρμογής στην οποία χρησιμοποιούνται. Όταν είναι απαραίτητο να περιγράψουμε περιεχόμενο από συγκεκριμένα πεδία, όπως από τον τομέα των ειδήσεων ή τα films, υπάρχει συχνά η ανάγκη για επέκταση και για εξειδίκευση των ειδικών εργαλείων του MPEG-7 και η χρήση της DDL για να προσδιορίσουμε τα εξειδικευμένα ή επιπρόσθετα εργαλεία.

Με την υιοθέτηση του XML Schema σαν την βάση για την Γλώσσα Προσδιορισμού της Περιγραφής του MPEG-7, το αποτέλεσμα είναι η διευκόλυνση της διαλειτουργικότητας με την χρήση μιας κοινής, γενικής και ισχυρής μορφής αναπαράστασης της πληροφορίας που περιγράφεται. Αντίστροφα, διευκολύνεται και η χρήση των εργαλείων του MPEG-7 για την

επέκταση των ήδη υπαρχόντων εφαρμογών της XML με τις λειτουργίες που διαθέτουν οι περιγραφές του πολυμεσικού περιεχομένου.

Συστήματα

Τα Συστήματα παρέχουν εργαλεία για να υποστηρίξουν την πολυπλεξία, τον συγχρονισμό των περιγραφών με το περιεχόμενο, τους μηχανισμούς μεταφοράς των περιγραφών και την κωδικοποίηση της αναπαράστασης της πληροφορίας με κειμενική ή δυαδική μορφή για ικανοποιητική αποθήκευση μετάδοση, διαχείριση και προστασία της πνευματικής ιδιοκτησίας των περιγραφών.

Η δομή του MPEG-7

Το πρότυπο του MPEG-7 αποτελείται από επτά (7) διαφορετικά μέρη. Κάθε ένα από αυτά τα μέρη διαπραγματεύεται και διαφορετικά ζητήματα και αυτό μας επιτρέπει να χρησιμοποιούμε κάθε φορά την ομάδα που εμείς θέλουμε ανάλογα με την εφαρμογή και (την ομάδα) που θα είναι σύμφωνη με την προσέγγιση των εργαλείων του MPEG-7. Επίσημα το πρότυπο του MPEG-7 αναφέρεται σαν ISO 15938 και οργανώνεται ως εξής:

ISO/IEC 15 938-1: MPEG-7 Systems, Συστήματα στο MPEG-7. Το πρώτο μέρος του προτύπου, προσδιορίζει τις λειτουργίες του προτύπου στο επίπεδο του συστήματος. Οι λειτουργίες αυτές είναι η προετοιμασία των περιγραφών για το MPEG-7 για την ικανοποιητική αποθήκευση και διάδοσή τους, ο συγχρονισμός του περιεχομένου και των περιγραφών και η ανάπτυξη των ανάλογων προσαρμοσμένων αποκωδικοποιητών.

ISO/IEC 15 938-2: MPEG-7 Description Definition Language, Γλώσσα Προσδιορισμού της Περιγραφής. Η γλώσσα αυτή είναι μία προτυποποιημένη γλώσσα για την δημιουργία νέων Περιγραφών ή Σχημάτων Περιγραφής, όπως επίσης και για την επέκταση των ήδη υπαρχόντων Σχημάτων Περιγραφής και Περιγραφών.

ISO/IEC 15 938-3: MPEG-7 Visual, Περιγραφή οπτικού περιεχομένου στο MPEG-7. Αυτό το μέρος προδιαγράφει ένα σύνολο από προτυποποιημένους Περιγραφείς και Σχήματα Περιγραφών. Οι Περιγραφείς οπτικού

περιεχομένου πραγματεύονται συνήθως εξειδικευμένα χαρακτηριστικά όπως το χρώμα, η υφή, το μέγεθος και η κίνηση της εικόνας. Τα Σχήματα Περιγραφής για το οπτικό περιεχόμενο απαιτούν συχνά την χρήση άλλων περιγραφών χαμηλού επιπέδου ή βοηθητικών στοιχείων, όπως τη δομή (την χωρική διάταξη, την εμφάνιση), την οπτική γωνία (πολλαπλές πλευρές), τον εντοπισμό (εντοπισμό της εικόνας), και τον χρόνο (χρονική διαδοχή).

Έχει προτυποποιηθεί ένας αριθμός περιγραφών για κάθε χαρακτηριστικό, όπως οι περιγραφείς για τα χρώματα, για την πυκνότητα, για την μορφή, την κίνηση και άλλοι περιγραφείς. Τέτοιοι περιγραφείς μπορεί να είναι οι περιγραφείς για την κλιμάκωση του χρώματος, για την μορφή της εικόνας, για την τροχιά της κίνησης και η αναγνώριση του προσώπου σε κάθε κατηγορία αντίστοιχα.

Επίσης έχει προσδιοριστεί και ένα σχήμα Περιγραφής για τον εντοπισμό της πληροφορίας, που ονομάζεται SpatioTemporal locator και συντίθεται από άλλα Σχήματα Περιγραφής, όπως το FigureTrajectory Description Scheme και το ParameterTrajectory Description Scheme.

ISO/IEC 15 938-4: MPEG-7 Audio, Περιγραφή Ακουστικού Περιεχομένου. Σε αυτό το μέρος προσδιορίζεται ένα προτυποποιημένο σύνολο Περιγραφών και Σχημάτων Περιγραφής. Οι Περιγραφείς του ήχου στο MPEG-7 πραγματεύονται τέσσερις κλάσεις ηχητικών σημάτων, που είναι: η καθαρή ομιλία, η καθαρή μουσική, καθαρές επιδράσεις του ήχου (pure sound effects) και αυθαίρετα ίχνη ήχου (arbitrary soundtracks). Οι περιγραφείς και τα Σχήματα Περιγραφής πραγματεύονται χαρακτηριστικά του ήχου, όπως τη σιωπή, το περιεχόμενο του προφορικού λόγου, το ηχόχρωμα, την μελωδία. Όπως οι περιγραφείς για το οπτικό υλικό, έτσι και οι περιγραφείς για το ακουστικό περιεχόμενο, απαιτούν συχνά την χρήση άλλων χαμηλού επιπέδου περιγραφών και ένα πλαίσιο Περιγραφής του Ήχου.

Παραδείγματα προτυποποιημένων Περιγραφών για τα ποικίλα χαρακτηριστικά του ήχου, μπορούν να είναι τα εξής: Περιγραφείς για την σιωπή, περιγραφείς για περιεχόμενο προφορικού λόγου, περιγραφείς για το ηχόχρωμα, για την επίδραση του ήχου και τον τύπο της μελωδίας, όπως οι περιγραφείς για τον τύπο σιωπής, για τον τύπο του ομιλητή στο προφορικό

περιεχόμενο, για τον τύπο του ηχοχρώματος του μουσικού οργάνου και για τον τύπο του ρυθμού σε κάθε κατηγορία αντίστοιχα.

ISO/IEC 15 938-5: MPEG-7 Multimedia Description Schemes, Σχήματα Περιγραφής Πολυμέσων. Τα Σχήματα Περιγραφής Πολυμέσων στο MPEG-7, προδιαγράφουν ένα πλαίσιο εργασίας υψηλού επιπέδου, που επιτρέπει την γενική περιγραφή όλων των ειδών των πολυμέσων, περιλαμβάνοντας δεδομένα ήχου, εικόνας και κειμένου. Περισσότερο εκτενής αναφορά θα γίνει στη συνέχεια, στην ενότητα για τα Σχήματα Περιγραφής Πολυμέσων.

ISO/IEC 15 938-6: MPEG-7 Reference Software, Σχετικό Λογισμικό. Αυτό το μέρος έχει σκοπό να παρέχει τις εφαρμογές των σχετικών μερών του MPEG-7 και είναι γνωστό σαν experimentation software (XM). Σε αυτό το μέρος περιγράφονται το απαραίτητο λογισμικό, το λειτουργικό περιβάλλον του MPEG-7 και οι περιορισμοί της πνευματικής ιδιοκτησίας (copyright). Παρόλο που περιλαμβάνεται σε αυτό κάποιο λογισμικό για την εξαγωγή των περιγραφών, η εστίαση γίνεται πιο πολύ στην δημιουργία των bitstream των περιγραφών και των Σχημάτων Περιγραφής με μία ρυθμιστική σύνταξη, παρά στην απόδοση των εργαλείων. Μέχρι στιγμής τα συστατικά που περιλαμβάνει ανήκουν σε τέσσερις κατηγορίες: Στον αναλυτή (parser) της DDL, στον αναλυτή της εγκυρότητας της DDL (validation parser), στους περιγραφείς για οπτικό περιεχόμενο, στους περιγραφείς για ακουστικό περιεχόμενο και στα Σχήματα Περιγραφής Πολυμέσων.

ISO/IEC 15 938-7: MPEG-7 Conformance, στοχεύει να παρέχει τις κατευθυντήριες γραμμές και διαδικασίες για να δοκιμάσει την συμβατότητα των εφαρμογών του MPEG-7. Για αυτό το μέρος του προτύπου δεν υπάρχουν ακόμα πολλά στοιχεία, γιατί οι έρευνες βρίσκονται σε πολύ αρχικό στάδιο.

Επειδή στην παρούσα εργασία το ενδιαφέρον είναι για την γενική περιγραφή της πληροφορίας, για το πώς αυτή γίνεται και όχι για κάποια συγκεκριμένη εφαρμογή ή για την συμβατότητα αυτών των περιγραφών, θα εστιάσουμε σε συγκεκριμένα μέρη του προτύπου. Στην ενότητα που ακολουθεί, θα γίνει η ανάλυση του πρώτου, του δεύτερου και του πέμπτου μέρους του προτύπου, δηλαδή η ανάλυση των Συστημάτων του MPEG-7, της Γλώσσας

Προσδιορισμού της Περιγραφής και των Σχημάτων Περιγραφής των Πολυμέσων.

Εργαλεία περιγραφής εικόνων

Μετά την γενική δομή και περιγραφή του προτύπου MPEG-7 και αφού έχουμε κατανοήσει την διαδικασία με την οποία αυτό δουλεύει, είναι αναγκαίο να περιγραφούν τα εργαλεία που αυτό παρέχει για την διαχείριση των εικόνων

Ένας επεξεργαστής διαχείρισης της εικόνας σύμφωνα με το περιεχόμενό της (image content-driven (CDP) pre-processor) δημιουργείται για να ενεργοποιήσει τους κατάλληλους μηχανισμούς του προτύπου MPEG-7 για την αναγνώριση χαρακτηριστικών και περιεχομένου σε μια εικόνα. Καθορίζει αυτόματα το αν υπάρχουν συγκεκριμένα χαρακτηριστικά γνωρίσματα όπως το χρώμα, η υφή ή το σχήμα κι έπειτα λειτουργεί έτσι ώστε να παράγει τους αντίστοιχους περιγραφείς. Τα πιο χαρακτηριστικά γνωρίσματα ενός τέτοιου συστήματος είναι ότι δεν υπάρχουν περιττά χαρακτηριστικά μεταφέροντας την πληροφορία των χαρακτηριστικών στους περιγραφείς. Πειράματα έχουν δείξει ότι συστήματα που χρησιμοποιούν τέτοιους επεξεργαστές καταφέρνουν να κατηγοριοποιούν την εικόνα ως προς κάποιος χαρακτηριστικό με μεγάλη ακρίβεια.

Οι οπτικοί περιγραφείς του προτύπου MPEG-7 για μια εικόνα όπως έχει προαναφερθεί είναι το χρώμα, η υφή και το σχήμα. Οι περιγραφείς αυτοί που λειτουργούν και ως εργαλεία του προτύπου περιγράφονται παρακάτω.

Περιγραφέας Χρώματος

Το εργαλείο του περιγραφέα χρώματος στο MPEG-7 περιέχει τον χρωματικό χώρο ή μοντέλο, την κβαντοποίηση του χρώματος, τα επικρατέστερα χρώματα, την βαθμίδα κλιμακωτών χρωμάτων, το χρωματικό αποτέλεσμα, την χρωματική δομή κ.ά. Τα εργαλεία αυτά μπορούν να χρησιμοποιηθούν για την έξυπνη ανάκτηση εικόνας μόνο υπό την προϋπόθεση ότι το υποκειμενικό περιεχόμενο που πρέπει να περιγραφεί είναι καλά ενημερωμένο και σωστά χρησιμοποιημένο εξαρχής.

Οι χρωματικοί περιγραφείς συνήθως εσωκλείουν και χρωματικά στατιστικά όπως για παράδειγμα είναι το ιστόγραμμα. Ο μαθηματικός τύπος που εκφράζει το ιστόγραμμα είναι ο εξής:

$$CD = [h(0); h(1); \dots ; h(nC)]$$

Όπου το nC είναι ο αριθμός των ορθογωνίων που περιγράφουν το εύρος των χρωμάτων με δυναμικό τρόπο. Με άλλα λόγια, το CD είναι απλά ένα διάνυσμα χαρακτηριστικών που αναπαριστά τις χρωματικές ιδιότητες ολόκληρης της εικόνας.

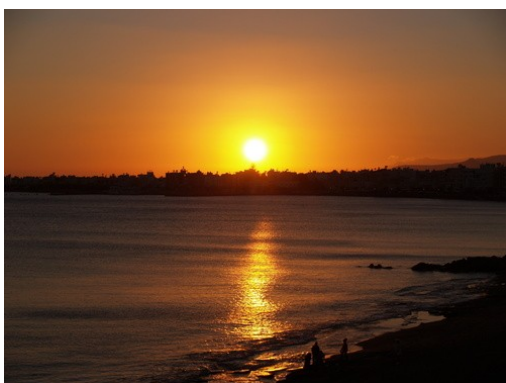
Περιγραφέας Σχήματος

Το εργαλείο του περιγραφέα σχήματος σε δισδιάστατες εικόνες συνδυάζει περιγραφείς με βάση μια περιοχή και περιγραφείς με βάση την περίμετρο μιας περιοχής. Για δισδιάστατες δυαδικές εικόνες, οι περιγραφείς σχήματος μιας περιοχής είναι προβολές με περιεχόμενο σχήματα σε έναν δίσκο με πολικές συντεταγμένες.

ΚΕΦΑΛΑΙΟ 6^ο - ΜΕΛΕΤΗ ΠΕΡΙΠΤΩΣΕΩΝ

Στο τρέχων κεφάλαιο θα αναπτύξουμε την έρευνα που έχει γίνει για τα διαθέσιμα συστήματα ανάκτησης εικόνων. Θα γίνει μια σύντομη παρουσίαση για τα συστήματα που χρησιμοποιούν ως βάση τους το κείμενο για την ανάκτηση μιας εικόνας και θα δοθεί μεγάλη βάση στα συστήματα αυτά που χρησιμοποιούν μια άλλη εικόνα για να ανακτήσουν μία συνώνυμή τους.

Αυτή τη στιγμή, λόγω και της μεγάλης εξάπλωσης του διαδικτύου οι διαδικτυακές εφαρμογές που ασχολούνται με την ανάκτηση εικόνας είναι αρκετές στην πραγματικότητα όμως οι εφαρμογές αυτές οι οποίες θα ονομάζονται από εδώ και πέρα μηχανές αναζήτησης εικόνων είναι ακόμη σε νηπιακό στάδιο ως προς τα αποτελέσματα που αυτές παρέχουν στον χρήστη. Όπως θα δούμε άλλωστε με την μελέτη των περιπτώσεων που διαπραγματεύεται η εργασία αυτή. Στην έρευνά μας έχουμε χρησιμοποιήσει δύο διαφορετικές εικόνες ως προς και τους τρεις χαρακτηριστικούς περιγραφείς τους. Έχουμε χρησιμοποιήσει τις εικόνες αυτές που έχουμε αναφέρει και σε προηγούμενο κεφάλαιο όπου διαπιστώσαμε ότι η ανάκτηση της πρώτης είναι σχετικά εύκολος ο τρόπος ανάκτησης και όπου της δεύτερης είναι αρκετά δύσκολο αφού υπάρχει πληθώρα διαφορετικών χρωμάτων, υφών και σχημάτων όπως φαίνεται και στις δύο εικόνες παρακάτω.



Ηλιοβασίλεμα



Μπουκέτο με λουλούδια

Ανάκτηση με βάση το κείμενο

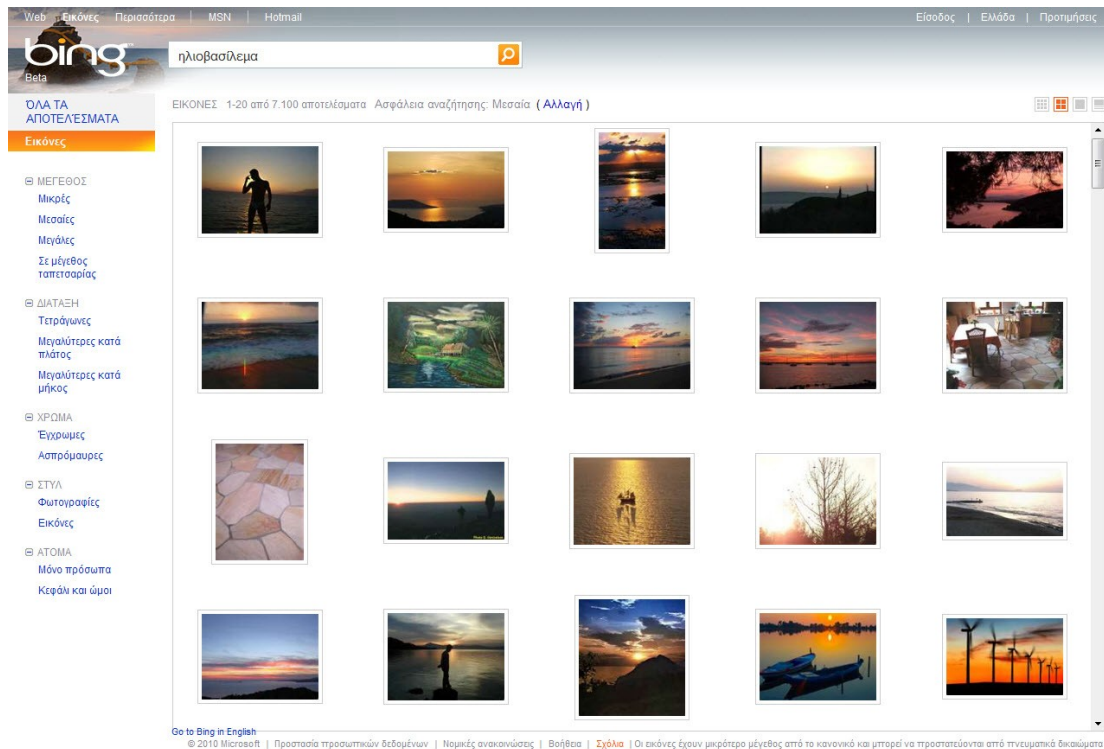
Google, yahoo, bing

Οι μηχανές αναζήτησης google, yahoo και bing είναι οι τρεις πλέον γνωστές μηχανές αναζήτησης ιστοσελίδων. Πέρα όμως από αυτό στην αναζήτηση εικόνων ο μόνος τρόπος με τον οποίο μπορεί κάποιος να αναζητήσει εικόνες είναι μέσω κειμένου.

Όπως βλέπουμε στην παρακάτω εικόνα θέσαμε το ερώτημα ηλιοβασίλεμα στην μηχανή αναζήτησης yahoo και στην μηχανή αναζήτησης bing.

The screenshot shows the Yahoo! search interface. At the top, there are navigation links for 'Web', 'Images', 'Video', 'Local', 'Shopping', 'News', and 'More'. The search bar contains the text 'ηλιοβασίλεμα' and a 'Search' button. Below the search bar, there are filters for 'SafeSearch: OFF', 'Show only: Wallpaper', and 'Less Filters'. The 'Size' filter is set to 'Wallpaper', and the 'Color' filter is set to 'Color'. The search results are displayed in a grid of 18 images, each with a thumbnail, filename, and source URL. The search bar at the bottom of the page also contains the text 'ηλιοβασίλεμα' and a 'Search' button.

Image	Filename	Source
	214462623_956c8..._o.jpg	nenyaki.blogspot.com
	myconoskiv0.jpg	esoteric.gr
	skopelos sunset.jpg	greekproperties.net
	DSC00065.jpg	skifun.gr
	fourtouniasmen..._sa.jpg	remallaclub.gr
	DSC00066.jpg	skifun.gr
	DSC00064.jpg	forum.snowreport.gr
	thailand2009_3.jpg	jokkel.com
	100_0569.jpg	forum.snowreport.gr
	161.jpg	alexiou-forum.gr
	100_0579.jpg	forum.snowreport.gr
	mar2002_lgbsuntree.jpg	colinfahy.com
	galissas.jpg	vasweb.vassilikos.gr
	thailand2009_2.jpg	jokkel.com
	An Ubuntu Sunsi..._ax.png	nothing2hide.net
	picture029pm3.jpg	snowclub.gr
	singapore_24.jpg	tolanic.com
	raff1.JPG	blogs.sch.gr



Τα πρώτα αποτελέσματα φαίνονται παραπάνω. Μας εμφανίζουν, λοιπόν, και οι τρεις μηχανές αναζήτησης εικόνες οι οποίες έχουν σημειωθεί με την λέξη την οποία δώσαμε εμείς προς αναζήτηση.

Επιπλέον και οι τρεις αυτές μηχανές εμφανίζουν κάποια επιπλέον φίλτρα όπως τα ονομάζουν με τα οποία μπορούμε να έχουμε μία πιο ολοκληρωμένη λύση για την εικόνα την οποία αναζητούμε. Έτσι βλέπουμε να έχουν αναζήτηση με βάση το μέγεθος της εικόνας, με βάση τις χρωματικές αποδόσεις μόνο ως προς την επιλογή του άσπρου-μαύρου και πολύχρωμου μοντέλου κ.ά. επιλογές.

Στην αναζήτηση του bing παρατηρούμαι ότι υπάρχει στα δεξιά επιλογή για τον αν θέλουμε να υπάρχουν άνθρωποι στην φωτογραφία ή όχι. Φυσικά αυτό δεν είναι αναγνώριση μέσω κάποιου χαρακτηριστικού απλά υπάρχουν ετικέτες που διευκρινίζουν το αν υπάρχει άνθρωπος ή όχι στην εικόνα.

Αντίστοιχα η αναζήτηση του yahoo έχει κι αυτή κάποια επιπλέον φίλτρα αναζήτησης δίνοντας κι αυτή βάση στο χρώμα και στο μέγεθος της εικόνας.

Στην αναζήτηση του google υπάρχει η δυνατότητα και κάποιον επιπλέον χαρακτηριστικών

Google ηλιοβασιλεμα Αναζήτηση

Ασφαλής αναζήτηση Περιορισμένη ▾

Παγκόσμιος ιστός Εικόνες Πορτοκαλί Απόκρυψη επιλογών

Αποτελέσματα 1 - 20 από περίπου 90.600 για ηλιοβασιλεμα. (0,24 δευτερόλεπτα)

Όλα τα αποτελέσματα

Εικόνες

Βίντεο
Ειδήσεις
Ιταλόγιο
Επιγραφές
Βιβλία
Συζητήσεις

Όλα τα μεγέθη

Μεγαίου μεγέθους
Μεγάλες
Εικονίδια
Μεγαλύτερα από ...
Ακριβώς ...

Οποιοδήποτε τύπος

Πρόσωπα
Φωτογραφία
Clip art
Σχεδιασμός γραμμής

Οποιοδήποτε χρώμα

Πλήρη χρώματα

Μαύρο και άσπρο

Συγκεκριμένο χρώμα

■ ■ ■ ■ ■ ■ ■ ■ ■ ■

Επαναφορά επιλογών

Ηλιοβασιλεμα

gr.hotellium.com/Sunset Πλήρης προσφορά ξενοδοχεία. Καλύτερη τιμή Guaranteed.

Σύνδεσμος διαφημιζόμενου



V - Dieses Wallpaper als 1024 x 768 - 84k - jpg wallpaper-area.to
[Εύρεση παρόμοιων εικόνων](#)



Aug 20 2009 8:20 AM 500 x 332 - 90k - jpg myspace.com
[Εύρεση παρόμοιων εικόνων](#)



V - Dieses Wallpaper als 1024 x 768 - 69k - jpg wallpaper-area.to
[Εύρεση παρόμοιων εικόνων](#)



Sonnenuntergang 650 x 488 - 192k - jpg gratis-foto.eu
[Εύρεση παρόμοιων εικόνων](#)



ΗΛΙΟΒΑΣΙΛΕΜΑ 457 x 343 - 82k - jpg blish.in.com
[Εύρεση παρόμοιων εικόνων](#)



Ονειρεμένο 1024 x 768 - 188k - jpg chlomos.gr
[Εύρεση παρόμοιων εικόνων](#)



Sonnenuntergang 650 x 487 - 200k - jpg gratis-foto.eu
[Εύρεση παρόμοιων εικόνων](#)



Χλωμός Κέρκυρα - 800 x 504 - 150k - jpg chlomos.gr
[Εύρεση παρόμοιων εικόνων](#)



Ηλιοβασιλεμα στην 400 x 287 - 17k - jpg aktolyiakas.blogspot.com
[Εύρεση παρόμοιων εικόνων](#)



καλούν στο 480 x 640 - 251k - jpg estiatona.gr
[Εύρεση παρόμοιων εικόνων](#)



Sonnenuntergang 330 x 440 - 47k - jpg la-palma-fotolog.de
[Εύρεση παρόμοιων εικόνων](#)



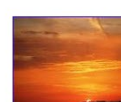
So bereiten Sie ein 384 x 517 - 30k - jpg kammerbibel.de
[Εύρεση παρόμοιων εικόνων](#)



φωτο Αθηνά Πικ 400 x 370 - 19k - jpg volkastoneiro.blogspot.com
[Εύρεση παρόμοιων εικόνων](#)



Χρυσό τό 400 x 300 - 21k - jpg www.taxiarhes.blogspot.com
[Εύρεση παρόμοιων εικόνων](#)



Ηλιοβασιλεμα από 1024 x 768 - 56k - jpg visitcorinthia.gr
[Εύρεση παρόμοιων εικόνων](#)

Το επιπλέον χαρακτηριστικό είναι ότι μπορείς να επιλέξεις το κύριο χρώμα της εικόνας. Αυτή η ενέργεια δείχνει να χρησιμοποιεί κάποιον σχετικό περιγραφέα του jpeg7 όμως δυστυχώς είναι δύσκολο να το καταλάβουμε από την στιγμή που το google δεν είναι open source και δεν μπορούμε να δούμε τους αλγόριθμους που χρησιμοποιεί.

Ανάκτηση Εικόνας Με Βάση Το Περιεχόμενο

Στην ενότητα αυτή θα εξετάσουμε μία επιτυχημένη και πρωτοποριακή χρήση του προτύπου MPEG-7 με την χρήση της μηχανής αναζήτησης Alipr.

Alipr

Το alipr προέρχεται από τα αρχικά Automatic Linguistic Indexing of Pictures - Real Time και στην πραγματικότητα κάνει αυτό ακριβώς που λέει και η ονομασία του. Προσθέτει σε πραγματικό χρόνο σημειώσεις ή υπογραφές όπως έχουμε αναφέρει σε προηγούμενα κεφάλαια και κατόπιν με βάση την συγκεκριμένη διαδικασία τοποθετεί τις εικόνες σε ομάδες ανάλογα με το περιεχόμενό τους.

Το μεγαλύτερο πρόβλημα που έχει προκύψει μέχρι σήμερα από την διαδικασία αυτή είναι ότι ενώ αρκετά συστήματα δουλεύουν κατ' αυτόν τον τρόπο στην πραγματικότητα η αναζήτηση πολλές φορές επιστρέφει λάθη. Αυτό συμβαίνει γιατί η σημειώσεις προστίθενται από τον δημιουργό της εικόνας και με αυτόν τον τρόπο έχουν πολλά υποκειμενικά στοιχεία μέσα τους.

Το alipr όπως αναφέραμε χρησιμοποιεί ένα είδος υπογραφής για να αναγνωρίσει την φωτογραφία. Αυτή η υπογραφή δεν έχει καμία σχέση με τις σημειώσεις τις οποίες πληκτρολογεί ή επιλέγει ο χρήστης από λιστα για την φωτογραφία την οποία ανεβάζει. Οι υπογραφή που χρησιμοποιεί το alipr προκύπτει από δύο χαρακτηριστικά της εικόνας. Το χρώμα και την υφή. Το σχήμα δεν συμμετέχει σε αυτήν την διαδικασία γιατί είναι ασαφή μέχρι σήμερα τα αποτελέσματα που μπορεί να δώσει ένα σύστημα με υπογραφή που βασίζεται στο σχήμα.

Χρώμα

Για να εξάγουμε το χρώμα από την υπογραφή τα στοιχεία του μοντέλου RGB μετατρέπονται για κάθε κουκίδα σε LUV χρωματικά στοιχεία. Τα τρισδιάστατα διανυσματικά χρώματα ομαδοποιούνται με τον αλγόριθμο k-means. Ο

αριθμός των ομάδων στο k-means καθορίζεται δυναμικά υπολογίζοντας την μέση τιμή μεταξύ των διανυσμάτων των ομάδων. Τακτοποιώντας τις ετικέτες της κάθε ομάδας από κουκίδες μέσα στην εικόνα σύμφωνα με την θέση τους σε αυτήν καταφέρνουμε να κατατμήσουμε την εικόνα. Αναφερόμαστε στην συλλογή των κουκίδων που βρίσκεται στην ίδια ομάδα ως περιοχή. Για κάθε περιοχή υπολογίζονται η μέση τιμή του διανύσματος του χρώματος και το ποσοστό των κουκίδων που το περιέχουν σε σχέση με ολόκληρη την εικόνα. Έτσι η χρωματική υπογραφή εμφανίζεται με τον μαθηματικό τύπο:

$$\{(v(1), p(1)), (v(2), p(2)), \dots, (v(m), p(m))\},$$

Όπου $v(j)$ η μέση τιμή του χρωματικού διανύσματος, και $p(j)$ η συσχετιζόμενη πιθανότητα, και m ο αριθμός των περιοχών.

Υφή

Χρησιμοποιώντας συντελεστές κυμάτων σε πολύ υψηλή περιοχή συχνότητας καταφέρνει το σύστημα να εξάγει το χαρακτηριστικό της υφής. Κάνοντας χρήση της μετατροπής κυμάτων Daubechies-4 [9] πάνω στο στοιχείο L (το οποίο είναι η ένταση της εικόνας) καταφέρνουμε να αποσυνθέσουμε την εικόνα σε τέσσερις περιοχές συχνότητων που τις ονομάζει το Daubechies-4: LL, LH, HL και HH. Οι περιοχές αυτές ανταποκρίνονται στα ίδιες χωρικές θέσεις μέσα στην εικόνα και ομαδοποιούνται σε τρισδιάστατα διανυσματικά χαρακτηριστικά υφών.

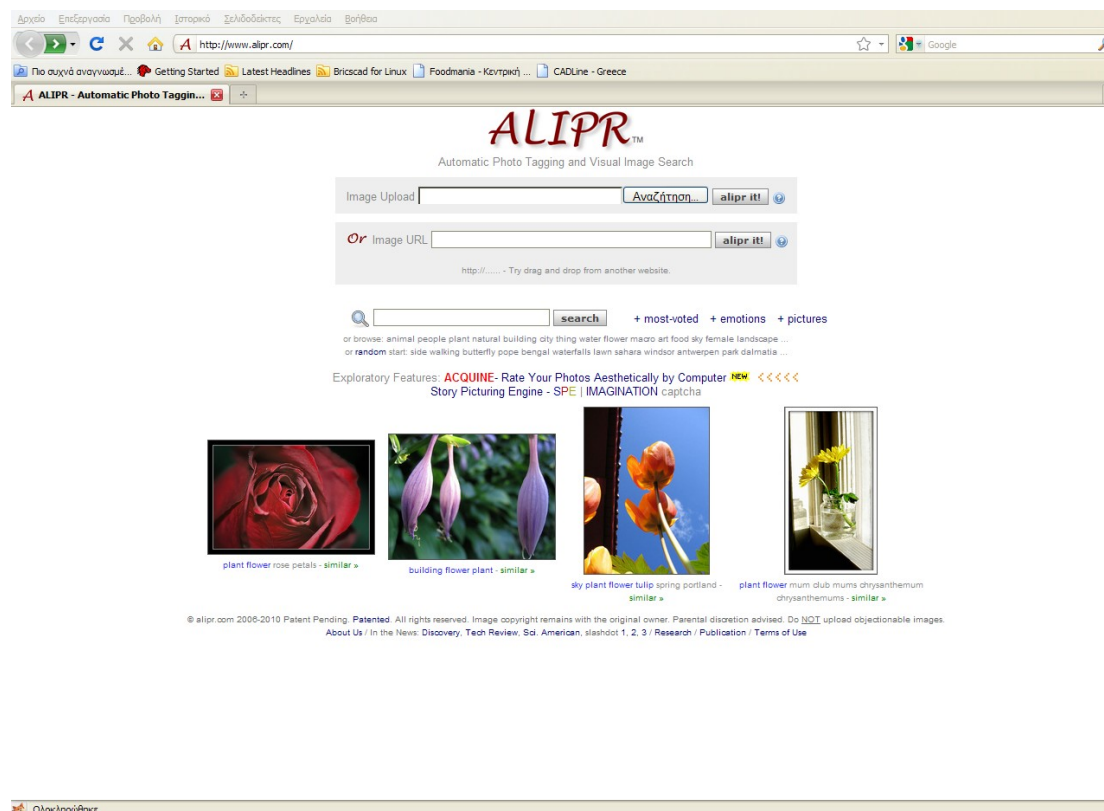
Εάν μία εικόνα περιέχει $n_r \times n_c$ κουκίδες, ο συνολικός αριθμός των υφών που

την απαρτίζουν είναι $\frac{n_r}{2} \times \frac{n_c}{2}$. Κατά την τυποποίηση της υφής, χρησιμοποιούνται οι απόλυτες τιμές των συντελεστών των κυμάτων. Στα διανυσματικά χαρακτηριστικά της υφής εφαρμόζεται η ομαδοποίηση K-means για την εξαγωγή των κυρίων σχεδίων αυτών των διανυσμάτων.

Με αυτόν τον τρόπο εξάγεται και υφή μιας εικόνας και σε συνδυασμό με το χαρακτηριστικό του χρώματος αλλά και τις σημειώσεις του χρήστη αποτελούν την υπογραφή της εικόνας.

Στην πράξη

Η ιστοσελίδα alipr.com επιτρέπει στους χρήστες να ανεβάσουν μία εικόνα ή να δείξουν έναν δικτυακό τόπο με εικόνες και να προσθέσουν σημειώσεις σε αυτές σε πραγματικό χρόνο. Το περιβάλλον του alipr.com φαίνεται στην παρακάτω εικόνα.



Αφού εισαχθεί η εικόνα στο σύστημα ο χρήστης έχει την δυνατότητα να επιλέξει σημειώσεις για την εικόνα που ανεβάζει επιλέγοντας απλά τα κουτάκια που βρίσκονται δίπλα από κάθε λέξη περιγραφής της εικόνας. Η λέξεις αυτές προκύπτουν από το σύστημα χρησιμοποιώντας τις τεχνικές που αναφέραμε παραπάνω αλλά και κάποιες επιπλέον τεχνικές με τις οποίες αναγνωρίζονται αντικείμενα στις εικόνες. Επιπλέον ο χρήστης μπορεί να πληκτρολογήσει και την δική του ξεχωριστή σημείωση στο κουτάκι που βρίσκεται κάτω από τις υπόλοιπες λέξεις όπως φαίνεται και στην εικόνα.


Top 15 Computer-Predicted Tags

ALIPR is like a child trying to learn about the world. Please help us to teach ALIPR. Check those correctly annotated words.

- indoor food art man-made flower
 plant dining people royal_guard barbecue
 old poster animal wild_life cloth

 Thought of other terms missed by ALIPR? Please add here, separated by commas '':
 and make the picture searchable

Optional information:

 Picture title

 URL to see related pictures

 Copyright (hypertext ok)

© alipr.com 2006-2007 Patent Pending. All rights reserved. Do NOT upload objectionable images. Pictures may be subject to copyright.

Όπως παρατηρούμαι είναι χαρακτηριστικό το κουμπί “teach alipr” το οποίο μας δείχνει ότι αυτήν την στιγμή το σύστημα μαθαίνει από εμάς ταυτόχρονα με τις δικές του επιλογές. Το αποτέλεσμα της αναζήτησης αυτή αφού επιλέξουμε τα κατάλληλα χαρακτηριστικά της εικόνας είναι αυτό που βλέπουμε παρακάτω. Έχουν εμφανιστεί εικόνες από λουλούδια που εν μέρει δείχνουν ότι η αναζήτησή μας είναι πετυχημένη. Όμως, στην πραγματικότητα δεν μπορούμε να πούμε ότι είναι τεράστια η επιτυχία αυτή αφού οι εικόνες που πήραμε είναι οι περισσότερες εικόνες από μεμονωμένα λουλούδια.



Thanks! Please upload another picture.

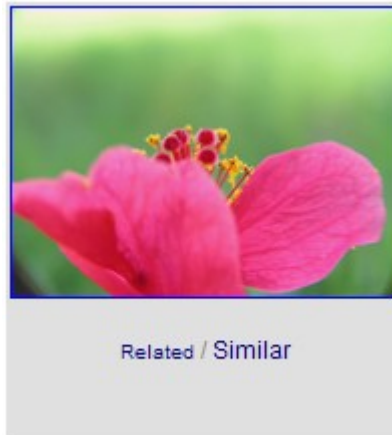
 Image Upload

 Or Image URL

 The following pictures may be related. Click Related to explore further or [search for visually similar pictures](#). **NEW**


© alipr.com 2006-2007 Patent Pending. All rights reserved. Do NOT upload objectionable images. Pictures may be subject to copyright.

Παρ' όλ' αυτά μπορούμε να δείξουμε στο σύστημα εάν κάποιες από τις εικόνες που μας έβγαλε αντιστοιχεί σε αυτές που θέλαμε να αναζητήσουμε. Έτσι κάτω από κάθε εικόνα υπάρχουν δύο λέξεις. Η λέξη Συσχετιζόμενες (Related) και η λέξη Όμοιες (Similar) όπως φαίνεται και στην παρακάτω εικόνα,



Οι λέξεις αυτές ουσιαστικά μας δίνουν την δυνατότητα να ορίσουμε αν η συγκεκριμένη εικόνα είναι συσχετιζόμενη με την εικόνα την οποία διαλέξαμε και να ανατροφοδοτήσουμε το σύστημα έτσι ώστε να μας εμφανίσει νέες εικόνες που να ανήκουν σε μια ομαδοποιημένη κατηγορία που να συμπεριλαμβάνει στοιχεία από την εικόνα την οποία δώσαμε στο σύστημα αλλά και από την εικόνα την οποία επιλέξαμε ως συσχετιζόμενη.

Επίσης εάν επιλέξουμε την εντολή Όμοια, τότε θα δούμε ότι το σύστημα θα μας εμφανίσει εικόνες οι οποίες είναι όμοιες με την εικόνα την οποία έχουμε επιλέξει και δεν έχουν σχέση με την εικόνα την οποία έχουμε ανεβάσει εμείς στο σύστημα.

Στην πράξη επιλέγουμε να μας εμφανίσει το σύστημα τις όμοιες εικόνες με την παραπάνω και το αποτέλεσμα είναι το εξής:

The screenshot shows the ALIPR website interface. At the top, there is a search bar with the text 'ALIPR' and 'Keyword(s):'. Below the search bar, there are several icons and links: 'search', 'upload', 'tags', 'title', 'Search', and 'Search'. There are also links for '+ most-voted', '+ emotions', and '+ pictures'. The main content area displays a grid of 15 visually similar images, each with a small icon and the text 'Related / Similar' below it. The images are arranged in three rows of five. The footer contains the text 'Visual Similarity Search By S-I-M-P-L-I-C-I-T-Y' and '© alipr.com 2006-2007. Automatic Linguistic Indexing of Pictures - Real Time. Patent Pending. Patented. All rights reserved. About Us. Pictures may be subject to copyright.'

Το αποτέλεσμα όπως παρατηρούμε κι εμείς είναι άκρως ικανοποιητικό αφού κατάφερε η αναζήτηση να μας επιστρέψει σωστά αποτελέσματα και ως προς το χρώμα και ως προς την υφή κι επειδή στην συγκεκριμένη περίπτωση η βάση του συστήματος δεν είναι εμπλουτισμένη με πολλές εικόνες παρομοίων χρωμάτων και υφής, κατάφερε το σύστημα ανάκτησης να καλύψει και το χαρακτηριστικό του σχήματος χωρίς όμως αυτό να είναι ο σκοπός του.

ΕΠΙΛΟΓΟΣ

Ο τομέας της ανάκτησης πολυμεσικής πληροφορίας έχει πολύ δρόμο ακόμα για να μπορέσει να δώσει αποτελέσματα τα οποία θα είναι αδιαμφισβήτητα σωστά. Η αναζήτηση μιας πληροφορίας γενικότερα δεν είναι εύκολο πόσο μάλλον όταν πρόκειται για πολυμεσική πληροφορία.

Βήματα προόδου έχουν γίνει μετά την προτυποποίηση κάποιων κανόνων με το MPEG7, όμως, παρατηρούμαι ότι στο πεδίο της εικόνας λίγες είναι οι μηχανές αναζήτησης οι οποίες παράγουν αποτελέσματα σχετικά με την αναζήτησή μας. Έτσι, λοιπόν, υπάρχει ακόμη τεράστιο περιθώριο έρευνας στον τομέα αυτό. και σίγουρα οι λύσεις που θα μας προσφέρονται στο μέλλον θα είναι πολύ καλύτερες από τις σημερινές.

ΒΙΒΛΙΟΓΡΑΦΙΑ

- [1]** Ballard, D.H. And Brown, C.M. 1982. *Computer Vision*. Prentice Hall, New Jersey, USA. Levine, M. 1985. *Vision in Man and Machine*, Mcgraw Hill, Columbus.
- [2]** Haralick, R.M. And Shapiro, L.G. 1993. *Computer and Robot Vision*. Addison-Wesley, New York, USA.
- [3]** Flickner, M. Sawhney, H. Niblack, W. Ashley, J. Qian Huang Dom, B. Gorkani, M. Hafner, J. Lee, D. Petkovic, D. Steele, D. Yanker, P. 1995. Query by image and video content: the QBIC system, *IEEE Computer*, September, 23-32.
- [4]** Bach, J.R., Fuller, C., Gupta, A., Hampapur, A., Horowitz, B., Humphrey, R., Jain, R., And Shu, C.F. 1996. Virage image search engine: An open framework for image management. In *Proceedings of the SPIE Storage and Retrieval for Still Image and Video Databases*, California, USA, 76-87.
- [5]** Smith, J. R. And Chang, S.F. 1997. Visually Searching the Web for Content. *IEEE Multimedia* 4(3), 12- 20.
- [6]** Frankel, C., Swain, M.J., And Athitsos, V. 1996. WebSeer: An Image Search Engine for the World Wide Web. University of Chicago Technical Report 96-14, University of Chicago, USA.
- [7]** Bliujute, R., Saltenis, S., Slivinskas, G., And Jensen, C.S. 1999. Developing a DataBlade for a New Index. In *Proceedings of IEEE International Conference on Data Engineering*, IEEE, Sydney, March 1999, 314-323.
- [8]** Egas, R., Huijsmans, N., Lew, M.S., And Sebe, N. 1999. Adapting k-d Trees to Visual Retrieval. In *Proceedings of the International Conference on Visual Information Systems*, Amsterdam, June 1999, A. SMEULDERS AND R. JAIN, Eds., 533-540.
- [9]** Rowley, H., Baluja, S., And Kanade, K. 1996. Human Face Detection in Visual Scenes. *Advances in Neural Information Processing Systems 8* (Proceedings of NIPS), Denver, USA, November, 875-881.
- [10]** Lew, M.S. And Huijsmans, N. 1996. Information Theory and Face Detection. In *Proceedings of the International Conference on Pattern Recognition*, Vienna, Austria, 601-605.
- [11]** Lew, M.S. 2000. Next Generation Web Searches for Visual Content. *IEEE Computer*, November, 46-53.

- [12]** Sclaroff, S., La Cascia, M., Sethi, S., And Taycher, L. 2001. Mix and Match Features in the ImageRover Search Engine. In Principles of Visual Information Retrieval, M.S. LEW, Ed. Springer-Verlag, London, 259-277.
- [13]** Chen, Y., Zhou, X.S., And Huang, T.S. 2001. One-class SVM for Learning in Image Retrieval, In Proceedings of IEEE International Conference on Image Processing, Thessaloniki, Greece, October, 815-818.
- [14]** Tieu, K. And Viola, P. 2004. Boosting Image Retrieval, International Journal of Computer Vision 56(1), 17-36.
- [15]** Rui, Y. And Huang, T.S. 2001. Relevance Feedback Techniques in Image Retrieval. In Principles of Visual Information Retrieval, M.S. LEW, Ed. Springer-Verlag, London, 219-258.
- [16]** Jolion, J.M. 2001. Feature Similarity. In Principles of Visual Information Retrieval, M.S. LEW, Ed. Springer-Verlag, London, 122-162.
- [17]** Sebe, N., And Lew, M.S. 2001. Color Based Retrieval, Pattern Recognition Letters 22(2), 223-230.
- [18]** Stricker, M. & Orengo, M. [Similarity of color images](#) (1995). : 381-392.
- [19]** Rubner, Y., Tomasi, C., & Guibas, L. J. (2000). The earth mover's distance as a metric for image retrieval. Int. J. Comput. Vision, 40(2), 99 - 121.
- [20]** Carson, C., Belongie, S., Greenspan, H., & Malik, J. (1997). Region-based image querying. In CAIVL '97: Proceedings of the 1997 Workshop on Content-Based Access of Image and Video Libraries (CBAIVL '97) (pp.~42). Washington, DC, USA: IEEE Computer Society.
- [21]** Carson, C., Belongie, S., Greenspan, H., & Malik, J. (2002). Blobworld: Image segmentation using expectation-maximization and its application to image querying. IEEE Trans. Pattern Anal. Mach. Intell., 24(8), 1026-1038.
- [22]** Ng, R. T. & Sedighian, A. (1996). Evaluating multidimensional indexing structures for images transformed by principal component analysis. volume 2670 (pp. 50-61): SPIE.
- [23]** Egas, R., Huijismans, D. P., Lew, M. S., & Sebe, N. (1999). Adapting k-d trees to visual retrieval. In VISUAL '99: Proceedings of the Third International Conference on Visual Information and Information Systems (pp. 533-540). London, UK: Springer-Verlag.

- [24]** Scott, G. J. & Shyu, C.-R. (2003). Ebs k-d tree: An entropy balanced statistical k-d tree for image databases with ground-truth labels. In CIVR (pp. 467-476).
- [25]** Ye, H. & Xu, G. (2003). Fast search in large-scale image database using vector quantization. In CIVR (pp. 477-487).
- [26]** El-Kwae, E. A. & Kabuka, M. R. (2000). Efficient content-based indexing of large image databases. *ACM Trans. Inf. Syst.*, 18(2), 171-210.
- [27]** Greene, D., Parnas, M., & Yao, F. (1994). Multi-index hashing for information retrieval. In *Foundations of Computer Science, 1994 Proceedings., 35th Annual Symposium on* (pp. 722-731).
- [28]** Philbin, J., Chum, O., Isard, M., Sivic, J., & Zisserman, A. (2007). Object retrieval with large vocabularies and fast spatial matching. In *Proc. CVPR*.
- [29]** Hastie, T., Tibshirani, R., & Friedman, J. H. (2001). *The Elements of Statistical Learning*. Springer.
- [30]** Spyrou, E., Tolia, G., Mylonas, P., & Avrithis, Y. (2008). A semantic multimedia analysis approach utilizing a region thesaurus and LSA. *wiamis*, 0, 8-11.
- [31]** Sivic, J. & Zisserman, A. (2003). Video google: A text retrieval approach to object matching in videos. In *ICCV '03: Proceedings of the Ninth IEEE International Conference on Computer Vision* (pp. 1470). Washington, DC, USA: IEEE Computer Society.
- [32]** Gemert, J. C. V., Geusebroek, J.-M., Veenman, C. J., & Smeulders, A. W. (2008). Kernel codebooks for scene categorization. In D. Forsyth, P. Torr, & A. Zisserman (Eds.), *Computer Vision { ECCV 2008, volume 5304 of Incs* (pp. 696-709).: Springer.
- [33]** Jegou, H., Douze, M., & Schmid, C. (2008). Hamming embedding and weak geometric consistency for large scale image search. In D. Forsyth, P. Torr, & A. Zisserman (Eds.), *Computer Vision { ECCV 2008, volume 5302 of Incs* (pp. 304-317).: Springer.
- [34]** Philbin, J., Chum, O., Isard, M., Sivic, J., & Zisserman, A. (2008). Lost in quantization: Improving particular object retrieval in large scale image databases. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.
- [35]** Assfalg, J., Bimbo, A. D., & Pala, P. (2002). Three-dimensional interfaces for querying by example in content-based image retrieval. *IEEE Transactions on Visualization and Computer Graphics*, 8(4), 305-318.

- [36]** Kaster, T., Pfeifer, M., & Bauckhage, C. (2003). Combining speech and haptics for intuitive and efficient navigation through image databases. In *ICMI '03: Proceedings of the 5th international conference on Multimodal interfaces* (pp. 180-187). New York, NY, USA: ACM.
- [37]** Chen, Y., Wang, J., & Krovetz, R. (2005). Clue: cluster-based retrieval of images by unsupervised learning. *Image Processing, IEEE Transactions on*, 14(8), 1187-1201.
- [38]** Nakazato, M. & Huang, T. (2001). 3d mars: immersive virtual reality for content-based image retrieval. *Multimedia and Expo, 2001. ICME 2001. IEEE International Conference on*, (pp. 44-47).
- [39]** Cunningham, S. J., Bainbridge, D., & Masoodian, M. (2004). How people describe their image information needs: a grounded theory analysis of visual arts queries. In *JCDL '04: Proceedings of the 4th ACM/IEEE-CS joint conference on Digital libraries* (pp. 47-48). New York, NY, USA: ACM.
- [40]** Cunningham, S. J. & Masoodian, M. (2006). Looking for a picture: an analysis of everyday image information searching. In *JCDL '06: Proceedings of the 6th ACM/IEEE-CS joint conference on Digital libraries* (pp. 198-199). New York, NY, USA: ACM.
- [41]** Chen, L.-Q., Xie, X., Fan, X., Ma, W.-Y., Zhang, H., & Zhou, H.-Q. (2003). A visual attention model for adapting images on small displays. *Multimedia Syst.*, 9(4), 353-364.
- [42]** Bertini, E., Cali', A., Catarci, T., Gabrielli, S., & Kimani, S. (2005). Interaction-based adaptation for small screen devices. In L. Ardissono, P. Brna, & A. Mitrovic (Eds.), *User Modeling*, volume 3538 of *Lecture Notes in Computer Science* (pp. 277-281).: Springer.
- [43]** Wang, J., Jann Yang, W., & Acharya, R. (1997). Color clustering techniques for color-content-based image retrieval from image databases. *icmcs*, 00, 442.
- [44]** Sebe, N. & Lew, M. S. (2001). Color-based retrieval. *Pattern Recogn. Lett.*, 22(2), 223-230.
- [45]** Swain, M. J. & Ballard, D. H. (1991). Color indexing. *Int. J. Comput. Vision*, 7(1), 11-32.
- [46]** Flickner, M., Sawhney, H., Niblack, W., Ashley, J., Huang, Q., Dom, B., Gorkani, M., Hafner, J., Lee, D., Petkovic, D., Steele, D., & Yanker, P. (1995). Query by image and video content: The qbic system. *Computer*, 28(9), 23-32.
- [47]** Y. Song, W. Wang, A. Zhang, Automatic annotation and retrieval of images, *J. World Wide Web* 6 (2) (2003) 209–231.

- [48]** F. Jing, M. Li, L. Zhang, H.-J. Zhang, B. Zhang, Learning in regionbased image retrieval, Proceedings of the International Conference on Image and Video Retrieval (CIVR2003), 2003, pp. 206–215.
- [49]** H. Feng, D.A. Castanon, W.C. Karl, A curve evolution approach for image segmentation using adaptive flows, Proceedings of the International Conference on Computer Vision (ICCV'01), 2001, pp. 494–499.
- [50]** W.Y. Ma, B.S. Manjunath, Edge flow: a framework of boundary detection and image segmentation, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1997, pp. 744–749.
- [51]** P.L. Stanchev, D. Green Jr., B. Dimitrov, High level color similarity retrieval, Int. J. Inf. Theories Appl. 10 (3) (2003) 363–369.
- [52]** K.A. Hua, K. Vu, J.-H. Oh, SamMatch: a flexible and efficient sampling-based image retrieval technique for large image databases, Proceedings of the Seventh ACM International Multimedia Conference (ACM Multimedia'99), November 1999, pp. 225–234.
- [53]** Y. Deng, B.S. Manjunath, Unsupervised segmentation of color-texture regions in images and video, IEEE Trans. Pattern Anal. Mach. Learn. (PAMI) 23 (8) (2001) 800–810.
- [54]** C. Carson, S. Belongie, H. Greenspan, J. Malik, Blobworld: image segmentation using expectation-maximization and its application to image querying, IEEE Trans. Pattern Anal. Mach. Intell. 8 (8) (2002) 1026–1038.
- [55]** H. Feng, T.-S. Chua, A bootstrapping approach to annotating large image collection, Workshop on Multimedia Information Retrieval in ACM Multimedia, November 2003, pp. 55–62.
- [56]** R. Shi, H. Feng, T.-S. Chua, C.-H. Lee, An adaptive image content representation and segmentation approach to automatic image annotation, International Conference on Image and Video Retrieval (CIVR), 2004, pp. 545–554.
- [57]** V. Mezaris, I. Kompatsiaris, M.G. Strintzis, An ontology approach to object-based image retrieval, Proceedings of the ICIP, vol. II, 2003, pp. 511–514.
- [58]** P. Stanchev, Using image mining for image retrieval, IASTED Conference “Computer Science and Technology,” Cancun, Mexico, May 2003, pp. 214–218.