



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

Ανίχνευση Αντικειμένων σε Εικόνες και Ταξινόμηση Σκηνών με Χρήση
Περιγραφέων MPEG-7 και Τεχνικών Οπτικού Θησαυρού

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

Γεωργίου Κ. Τόλια

Επιβλέπων: Στέφανος Κόλλιας
Καθηγητής

Αθήνα, Νοέμβριος 2007



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ
ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

Ανίχνευση Αντικειμένων σε Εικόνες και Ταξινόμηση Σκηνών με Χρήση
Περιγραφών MPEG-7 και Τεχνικών Οπτικού Θησαυρού

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

Γεωργίου Κ. Τόλια

Επιβλέπων: Στέφανος Κόλλιας
Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 27^η Νοεμβρίου 2007.

.....
Στέφανος Κόλλιας
Καθηγητής Ε.Μ.Π.

.....
Σταφυλοπάτης Ανδρέας-Γεώργιος
Καθηγητής Ε.Μ.Π.

.....
Τσανάκας Παναγιώτης
Καθηγητής Ε.Μ.Π.

Αθήνα, Νοέμβριος 2007

.....
Γεώργιος Κ. Τόλιας

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

Copyright © Γεώργιος Κ. Τόλιας (2007) Εθνικό Μετσόβιο Πολυτεχνείο.

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα. Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

Περίληψη

Ο διαρκώς αυξανόμενος όγκος οπτικοακουστικού πολυμεσικού περιεχομένου τα τελευταία χρόνια έχει οδηγήσει στην ανάγκη αυτόματης εξαγωγής χαρακτηριστικών και αυτόματης περιγραφής του. Με χρήση διαφόρων περιγραφών, καθώς και εκείνων που ορίζονται από το πρότυπο MPEG-7, γίνεται η προσπάθεια περιγραφής της χαμηλού επιπέδου πληροφορίας του. Στα πλαίσια της παρούσας διπλωματικής παρουσιάζονται περιγραφείς οπτικής πληροφορίας, αναπτύσσεται μία εφαρμογή εξαγωγής αυτών βασισμένη στο eXperimentation Model του MPEG-7 με δυνατότητες εξαγωγής από πολλαπλές περιοχές που ορίζονται από χάρτη περιοχών και διερευνάται η συμβατή με το eXperimentation Model λειτουργία της. Έπειτα γίνεται μία προσέγγιση του προβλήματος που αποκαλείται σημασιολογικό κενό, της αντιστοίχισης δηλαδή χαρακτηριστικών χαμηλού επιπέδου σε έννοιες υψηλού επιπέδου. Χρησιμοποιείται μία τεχνική βασισμένη σε περιοχές εικόνας και δημιουργείται ένας οπτικός θησαυρός για την αναπαράσταση της γνώσης. Εκπαιδεύονται νευρωνικά δίκτυα για τον εντοπισμό των εννοιών υψηλού επιπέδου. Θεωρώντας κάθε εικόνα σαν ένα έγγραφο κειμένου και κάθε περιοχή σαν μία λέξη χρησιμοποιείται και η τεχνική Λανθάνουσας Σημασιολογικής Ανάλυσης. Αποτελέσματα παρουσιάζονται για όλες τις παραπάνω τεχνικές, για διαφορετικά μεγέθη θησαυρού με σκοπό την επίτευξη των καλύτερων δυνατών αποτελεσμάτων, καθώς και για πειράματα ταξινόμησης εικόνων.

Λέξεις Κλειδιά

MPEG-7, οπτικός θησαυρός, οπτικές περιγραφές, σημασιολογική ανάλυση εικόνας, εντοπισμός αντικειμένων, σημασιολογικό κενό, TRECVID, ταξινόμηση εικόνων, eXperimentation Model, λανθάνουσα σημασιολογική ανάλυση

Abstract

The enlarging audiovisual multimedia content during the last few years has emerged the need of automatic feature extraction and description of this content. With the use of various descriptors, including those defined by the MPEG-7 standard, its low level information is captured. In this diploma thesis MPEG-7 visual descriptors are examined and a descriptor extraction application is developed based on the MPEG-7 eXperimentation Model. This application is evaluated in order to verify its alignment to the XM. This application is then used within a high-level detection approach. A region-based technique is applied and a visual thesaurus is constructed to formalise knowledge. Neural-network detectors are trained in order to detect high-level concepts. Moreover, the utility of the well known Latent Semantic Analysis technique is investigated. The dataset of the TRECVID benchmark has been used for testing this techniques. Finally a car exterior/interior classification problem is also tackled. Extensive experimental results are presented for each of the aforementioned problems.

Keywords

MPEG-7, visual thesaurus, visual descriptors, semantic image analysis, concept detection, semantic gap, TRECVID, image classification, experimentation model, latent semantic analysis

Ευχαριστίες

Η παρούσα διπλωματική εργασία εκπονήθηκε κατά το ακαδημαϊκό έτος 2006- 2007 στο Εργαστήριο Ψηφιακής Επεξεργασίας Εικόνας, Βίντεο και Πολυμέσων (IVML) του Εθνικού Μετσόβιου Πολυτεχνείου. Θα ήθελα να ευχαριστήσω τον επιβλέποντα Καθηγητή κ. Στέφανο Κόλλια για την εμπιστοσύνη που μου έδειξε αναθέτοντάς μου την εργασία αυτή και για τη δυνατότητα που μου έδωσε να ασχοληθώ με το συγκεκριμένο ενδιαφέρον θέμα. Θα ήθελα επίσης να ευχαριστήσω τον Ερευνητή Δρ Ιωάννη Αβρίθη και τον Υποψήφιο Διδάκτορα Ευάγγελο Σπύρου, διότι συνέβαλαν στη συγγραφή αυτής της εργασίας με την καθοδήγηση, τις πολύτιμες συμβουλές και υποδείξεις τους.

Περιεχόμενα

I	Οπτικοί Περιγραφείς MPEG-7	17
1	Εισαγωγή στο MPEG-7	18
2	Οπτικοί Περιγραφείς	24
2.1	Περιγραφείς Χρώματος	24
2.1.1	Χρωματικοί Χώροι (Color Spaces)	24
2.1.2	Περιγραφέας Κύριων Χρωμάτων (Dominant Color Descriptor)	25
2.1.3	Κλιμακωτός Περιγραφέας Χρώματος (Scalable Color Descriptor)	27
2.1.4	Περιγραφέας Δομής χρώματος (Color Structure Descriptor)	29
2.1.5	Περιγραφέας Διάταξης Χρώματος (Color Layout Descriptor)	31
2.2	Περιγραφείς Υφής	33
2.2.1	Περιγραφέας Ομοιογενούς Υφής (Homogeneous Texture Descriptor)	33
2.2.2	Περιγραφέας Ιστογράμματος Ακμών (Edge Histogram Descriptor)	34
2.3	Περιγραφείς Σχήματος	36
2.3.1	Περιγραφέας Σχήματος με βάση το Σχήμα (Region-Based Shape Descriptor)	36
2.3.2	Περιγραφέας Σχήματος με βάση το Περίγραμμα (Contour-Based Shape Descriptor)	38
II	Εφαρμογή Εξαγωγής Περιγραφών	41
3	Εισαγωγή στο MPEG-7 eXperimentation Model	42
4	Εφαρμογή Εξαγωγής Περιγραφών: VDE	44
4.1	Υλοποίηση VDE	44
4.2	Εξαγωγή από Πολλαπλές Περιοχές Ενδιαφέροντος	45
4.2.1	Λεπτομέρειες στον κώδικα	49
4.3	Bugs και Επιπλέον Αλλαγές στον Κώδικα	55
4.3.1	Περιγραφέας Κύριων Χρωμάτων	55
4.3.2	Περιγραφέας Διάταξης Χρώματος	56
4.3.3	Περιγραφέας Ομοιογενούς Υφής	57
4.3.4	Περιγραφέας Δομής Χρώματος	59
4.3.5	Περιγραφέας Σχήματος με βάση το Περίγραμμα	59

4.3.6	Περιγραφείας Σχήματος με βάση το Σχήμα	61
5	Εγχειρίδιο Χρήσης της Εφαρμογής VDE	63
5.1	Μεταγλώττιση	63
5.1.1	Μεταγλώττιση των επιπλέον βιβλιοθηκών	63
5.1.2	Μεταγλώττιση της εφαρμογής VDE	64
5.2	Τρόπος Χρήσης VDE	65
5.2.1	Κλήση Της Εφαρμογής	65
5.2.2	Αναπαράσταση Περιοχών Ενδιαφέροντος	70
5.3	Παράμετροι Εισόδου και Δομή Περιγραφέων	73
6	Έλεγχος και Αξιολόγηση	79
6.1	Διαδικασία Ελέγχου και Αξιολόγησης	79
6.2	Αποτελέσματα	81
6.3	Σχολιασμός Αποτελεσμάτων	85
III	Εντοπισμός Εννοιών	87
7	Εισαγωγή	88
7.1	Περιγραφή του Προβλήματος	88
7.2	TRECVID	88
7.3	Περιορισμοί	90
7.4	Άλλες Τεχνικές	92
8	Εντοπισμός Εννοιών με Χρήση Οπτικού Θησαυρού	93
8.1	Εξαγωγή Χαρακτηριστικών Χαμηλού Επιπέδου	94
8.2	Κατασκευή Οπτικού Θησαυρού	95
8.3	Διανύσματα Αναπαράστασης	97
8.4	Λανθάνουσα Σημασιολογική Ανάλυση (Latent Semantic Analysis - LSA)	98
8.5	Εκπαίδευση Ανιχνευτών	101
8.6	Έξοδος Συστήματος	103
9	Πειράματα	104
9.1	Διαδικασία Μετρήσεων	104
9.2	Αποτελέσματα	106
9.2.1	Πειράματα με Σύνολο Βίντεο του TRECVID	106
9.2.2	Πειράματα με Σύνολο Εικόνων του COREL	122
9.3	Συμμετοχές στο TRECVID	128
10	Ταξινόμηση Σκηνών	129
10.1	Περιγραφή Προβλήματος	129
10.2	Πειράματα	130
11	Συμπεράσματα και Μελλοντικές Επεκτάσεις	133

A	Αρχεία Παραμέτρων και Αρχεία Εξόδου	138
A.1	Παραδείγματα Αρχείων Παραμέτρων	138
A.2	Παραδείγματα Αρχείων XML Εξόδου	142

Περιεχόμενα Πινάκων

2.1	Αποτελέσματα ANMRR για τον περιγραφέα κύριων χρωμάτων	27
2.2	Αποτελέσματα ANMRR για τον περιγραφέα κύριων χρωμάτων με χωρική συνοχή	27
2.3	Αποτελέσματα για τον περιγραφέα δομής χρώματος στον HMMD χρωματικό χώρο	31
2.4	Αποτελέσματα πειραμάτων για τον περιγραφέα σχήματος με βάση το σχήμα στην βάση δεδομένων με εμπορικά σχήματα	38
2.5	Αποτελέσματα πειραμάτων με τον περιγραφέα σχήματος με βάση το περίγραμμα	40
4.1	Μέθοδος Start των κλάσεων DescriptorServer στο XM, δεξιά αναφέρονται τα αρχεία στα οποία ορίζονται οι μέθοδοι που κάθε φορά καλούνται	51
4.2	Μέθοδος Start των κλάσεων DescriptorServer στο VDE, με κόκκινη γραμματοσειρά είναι ότι προστέθηκε	52
4.3	Μέθοδος StartExtracting των κλάσεων DescriptorExtraction στο VDE, με κόκκινη γραμματοσειρά είναι ότι προστέθηκε	55
4.4	Μέθοδος CreateSmallImage της κλάσης ColorLayoutExtraction στο VDE, με κόκκινη γραμματοσειρά είναι ότι προστέθηκε	58
5.1	Πίνακας συντομογραφιών για το αναγνωριστικό -p, αντιστοιχία ονομάτων αρχείων παραμέτρων με descriptorCode	66
5.2	Πίνακας αναγνωριστικών από την γραμμή εντολών	68
5.3	Πίνακας συντομογραφιών για το αναγνωριστικό -a, αντιστοιχία εφαρμογών εξαγωγής με descriptorCode	69
5.4	Παράμετροι και έξοδος για τον περιγραφέα κύριων χρωμάτων	74
5.5	Παράμετροι και έξοδος για τον κλιμακωτό περιγραφέα χρώματος	75
5.6	Παράμετροι και έξοδος για τον περιγραφέα διάταξης χρώματος	76
5.7	Παράμετροι και έξοδος για τον περιγραφέα δομής χρώματος	76
5.8	Έξοδος για τον περιγραφέα ιστογράμματος ακμών	77
5.9	Παράμετροι και έξοδος για τον περιγραφέα ομοιογενούς υφής	77
5.10	Έξοδος για τον περιγραφέα σχήματος με βάση το σχήμα	78
5.11	Έξοδος για τον περιγραφέα σχήματος με βάση το περίγραμμα	78
6.1	Αποτελέσματα ελέγχου με χάρτη περιοχών	85
9.1	Αριθμός χαρακτηριστικών καρέ που απεικονίζουν την κάθε έννοια	107
9.2	Στοιχεία πειράματος TRECVID	108

9.3	Μέση ακρίβεια για σύνολα εκπαίδευσης με διαφορετικό λόγο λ , με τονισμένη γραμματοσειρά είναι οι υψηλότερες τιμές	109
9.4	Αριθμός χαρακτηριστικών καρέ που απεικονίζουν ή όχι την κάθε έννοια για τα σύνολα εκπαίδευσης και ελέγχου	109
9.5	Ζεύγη precision-recall μεταβάλλοντας το κατώφλι, για την έννοια Βλάστηση	111
9.6	Κατώφλια για όλους τους ανιχνευτές, χωρίς LSA	111
9.7	Τελικά αποτελέσματα πειραμάτων χωρίς την χρήση LSA, με σύνολο ελέγχου όλες τις εικόνες	115
9.8	Τελικά αποτελέσματα πειραμάτων χωρίς την χρήση LSA, με σύνολο ελέγχου αποτελούμενο από 20% θετικά και 80% αρνητικά χαρακτηριστικά καρέ ($\lambda=4$)	115
9.9	Κατώφλια για όλους τους ανιχνευτές, με LSA	116
9.10	Τελικά αποτελέσματα πειραμάτων με την χρήση LSA	120
9.11	Τελικά αποτελέσματα πειραμάτων με την χρήση LSA, με σύνολο ελέγχου αποτελούμενο από 20% θετικά και 80% αρνητικά χαρακτηριστικά καρέ ($\lambda=4$)	120
9.12	Σύγκριση αποτελεσμάτων χωρίς και με LSA, με σύνολο ελέγχου αποτελούμενο από 20% θετικά και 80% αρνητικά χαρακτηριστικά καρέ ($\lambda=4$)	121
9.13	Στοιχεία πειράματος COREL	124
9.14	Αριθμός χαρακτηριστικών καρέ που απεικονίζουν την κάθε έννοια	124
9.15	Σύνολα εκπαίδευσης και ελέγχου	124
9.16	Κατώφλια για όλους τους ανιχνευτές	125
9.17	Τελικά αποτελέσματα πειραμάτων χωρίς την χρήση LSA	127
9.18	Τελικά αποτελέσματα πειραμάτων με την χρήση LSA	127
9.19	Αποτελέσματα συμμετοχών στο TRECVID 2007	128
10.1	Σύνολα εκπαίδευσης και ελέγχου για την ταξινόμηση σε εξωτερικό/εσωτερικό μέρος αυτοκινητού	130
10.2	Αποτελέσματα ταξινόμησης εικόνων αυτοκινητών	132
A.1	Περιεχόμενο αρχείου παραμέτρων για τον περιγραφέα κύριων χρωμάτων	138
A.2	Περιεχόμενο αρχείου παραμέτρων για τον κλιμακωτό περιγραφέα χρώματος	139
A.3	Περιεχόμενο αρχείου παραμέτρων για τον περιγραφέα διάταξης χρώματος	139
A.4	Περιεχόμενο αρχείου παραμέτρων για τον περιγραφέα δομής χρώματος	140
A.5	Περιεχόμενο αρχείου παραμέτρων για τον περιγραφέα ιστογράμματος ακμών	140
A.6	Περιεχόμενο αρχείου παραμέτρων για τον περιγραφέα ομοιογενούς υψής	140
A.7	Περιεχόμενο αρχείου παραμέτρων για τον περιγραφέα σχήματος με βάση το περίγραμμα	141
A.8	Περιεχόμενο αρχείου παραμέτρων για τον περιγραφέα σχήματος με βάση το σχήμα	141
A.9	Περιεχόμενο αρχείου εξόδου XML για τον περιγραφέα κύριων χρωμάτων για μία περιοχή	142
A.10	Περιεχόμενο αρχείου εξόδου XML για τον περιγραφέα διάταξης χρώματος για μία περιοχή	143
A.11	Περιεχόμενο αρχείου εξόδου XML για τον κλιμακωτό περιγραφέα χρώματος για χάρτη με 4 περιοχές	144
A.12	Περιεχόμενο αρχείου εξόδου XML για τον περιγραφέα δομής χρώματος για αρχείο κειμένου με συντεταγμένες για 4 ορθογώνιες περιοχές	145

A.13 Περιεχόμενο αρχείου εξόδου XML για τον περιγραφέα ιστογράμματος ακμών για μία περιοχή	146
A.14 Περιεχόμενο αρχείου εξόδου XML για τον περιγραφέα ομοιογενούς υφής για μία περιοχή	146
A.15 Περιεχόμενο αρχείου εξόδου XML για τον περιγραφέα σχήματος με βάση το σχήμα για χάρτη με 4 περιοχές	147
A.16 Περιεχόμενο αρχείου εξόδου XML για τον περιγραφέα σχήματος με βάση το περιγράμμα για μία περιοχή	148

Περιεχόμενα Σχημάτων

2.1	Αριστερά: Τα 4 πρώτα επίπεδα Δεξιά: Επίπεδα 5-8	28
2.2	Αποτελέσματα ανάκτησης με διαφορετικό αριθμό συντελεστών Haar (16-256) κβαντισμένων με διαφορετικό αριθμό bits. H-Rec είναι η ανάκτηση με το κατασκευασμένο ιστόγραμμα από Haar συντελεστές σε πλήρη ανάλυση	29
2.3	Εικόνες 2 χρωματικών επιπέδων με διαφορετική δομή χρώματος	30
2.4	Υπολογισμός περιγραφέα δομής χρώματος	31
2.5	Διαδικασία εξαγωγής του περιγραφέα διάταξης χρώματος	32
2.6	Αποτελέσματα περιγραφέα διάταξης χρώματος σε σύγκριση με το αντιπροσωπευτικό χρώμα βασισμένο σε πλέγμα	33
2.7	Διάταξη συχνότητας για την εξαγωγή του περιγραφέα ομοιογενούς υψής	34
2.8	Ορισμός υποεικόνας και μπλοκ	35
2.9	Ανάκτηση για φυσικές εικόνες με τον περιγραφέα ιστογράμματος ακμών	36
2.10	Ανάκτηση για “clip art” εικόνες με τον περιγραφέα ιστογράμματος ακμών	37
2.11	Παραδείγματα σχημάτων για τα οποία ο περιγραφέας σχήματος με βάση το σχήμα είναι εφαρμόσιμος	37
2.12	Παράδειγμα ανάκτησης ομοιότητας για τον περιγραφέα σχήματος με βάση το σχήμα στην βάση δεδομένων με εμπορικά σχήματα	39
2.13	Παραδείγματα σχημάτων για τα οποία ο περιγραφέας σχήματος με βάση το περίγραμμα είναι εφαρμόσιμος	39
2.14	Παράδειγμα ανάκτησης με βάση την ομοιότητα για τον περιγραφέα σχήματος με βάση το περίγραμμα	40
3.1	Εφαρμογές εξυπηρετητή και πελάτη στο XM	43
4.1	Εικόνα εισόδου και χάρτης περιοχών	46
4.2	Υποεικόνες και δυαδικές μάσκες στο μέγεθος του περικλείοντος ορθογώνιου για την κάθε περιοχή	47
4.3	Διαδικασίες εξαγωγής για XM	48
4.4	Διαδικασίες εξαγωγής για VDE	48
4.5	Συναρτήσεις XM - ορθογώνιο: κλάση, έλλειψη: μέθοδος, βέλος: κλάση καλεί κλάση, αρίθμηση με την σειρά που καλούνται οι κλάσεις	48
4.6	Συναρτήσεις VDE - Κόκκινη διακεκομμένη γραμμή: νέες κλάσεις και μέθοδοι, έντονη γραμματοσειρά ονόματα νέων μεθόδων και κλάσεων, έντονη και κόκκινη γραμματοσειρά: νέα ορίσματα	49

4.7	Διαδικασία εξαγωγής με χάρτη περιοχών	50
4.8	Διάγραμμα ροής για την Main	54
4.9	Μικροσκοπικές εικόνες κατασκευασμένες κατά την εξαγωγή του περιγραφέα διάταξης χρώματος, για την περιοχή θάλασσας του σχήματος 4.1 με δυαδική μάσκα (αριστερά), με χάρτη περιοχών (δεξιά)	57
4.10	Περιγραφή λειτουργίας μεθόδου smallRegionsHT	60
5.1	Αναπαράσταση περιοχών ενδιαφέροντος για το VDE	71
5.2	Συντεταγμένες ορθογώνιων περιοχών και αντίστοιχες σχεδιασμένες περιοχές ενδιαφέροντος	72
5.3	Τιμές εικονοστοιχείων δυαδικής μάσκας και η αντίστοιχη εικόνα	72
5.4	Τιμές εικονοστοιχείων χάρτη περιοχών και η αντίστοιχη εικόνα	73
6.1	Δυαδικές μάσκες κατασκευασμένες από τον χάρτη περιοχών	80
6.2	Τρόπος σάρωσης για τον περιγραφέα δομής χρώματος, κόκκινο πλαίσιο: περικλείον ορθογώνιο, πράσινο τετράγωνο: δομικό στοιχείο	82
6.3	Αρχική εικόνα-μάσκα και υποεικόνα με αντίστοιχη μάσκα	83
6.4	Περιγραφέας δομής χρώματος με δυαδική μάσκα και με χάρτη περιοχών για την ίδια περιοχή ενδιαφέροντος(πάνω) , κανονικοποιημένο (κάτω)	84
7.1	Παραδείγματα εικόνων που απεικονίζουν τις 9 έννοιες προς εντοπισμό	91
8.1	Αλγόριθμος εντοπισμού εννοιών υψηλού επιπέδου	94
8.2	Αποστάσεις ανάμεσα σε περιοχές εικόνας και τύπους περιοχής, αριστερά αποστάσεις ανάμεσα σε μία περιοχή της εικόνας και σε όλους τους τύπους περιοχής, δεξιά αποστάσεις ανάμεσα σε όλες τις περιοχές μία εικόνας και έναν τύπο περιοχής	98
8.3	Παραδείγματα διανυσμάτων αναπαράστασης	99
8.4	Δομή νευρωνικού δικτύου	102
9.1	Διάγραμμα μέσης ακρίβειας καθώς μεταβάλλεται το παράθυρο υπολογισμού της για την έννοια γραφείο	110
9.2	Διαγράμματα precision-recall καθώς η τιμή κατωφλίου μεταβάλλεται	112
9.3	Διαγράμματα μέσης ακρίβειας καθώς μεγαλώνει το σύνολο ελέγχου και μεγαλώνει το λ, μέχρι $\lambda=4$ θεωρείται λογικό για σύνολο ελέγχου ενός συστήματος	114
9.4	Διαγράμματα precision-recall και Average Precision καθώς αυξάνεται το μέγεθος παραθύρου, χωρίς και με LSA, για την έννοια βλάστηση	117
9.5	Διαγράμματα precision-recall και Average Precision καθώς αυξάνεται το μέγεθος παραθύρου, χωρίς και με LSA, για την έννοια εξωτερικός χώρος	118
9.6	Διαγράμματα Average Precision καθώς αυξάνεται το μέγεθος παραθύρου, χωρίς και με LSA, για τις έννοιες δρόμος και γραφείο	119
9.7	Εικόνες που απεικονίζουν την έννοια ουρανός(1-4) και την έννοια βλάστηση (5-8) οι οποίες εντοπίστηκαν επιτυχώς	122
9.8	Εικόνες που απεικονίζουν την έννοια ουρανός(1-4) και την έννοια βλάστηση (5-8) οι οποίες δεν εντοπίστηκαν	123

9.9	Εικόνες που δεν απεικονίζουν την έννοια ουρανός(1-4) ή την έννοια βλάστηση (5-8) οι οποίες εντοπίστηκαν λανθασμένα πως τις απεικονίζουν	123
9.10	Μερικές από τις εικόνες που χρησιμοποιήθηκαν από την συλλογή του Corel, εικόνες που απεικονίζουν ουρανό στην πρώτη σειρά, χιόνι στην δεύτερη, βλάστηση στην τρίτη και καμία από τις 3 έννοιες στην τέταρτη	125
9.11	Διαγράμματα Average Precision καθώς αυξάνεται το μέγεθος παραθύρου για τα τέσσερα μεγέθη θησαυρού, για τις έννοιες χιόνι και ουρανός	126
9.12	Διάγραμμα Average Precision καθώς αυξάνεται το μέγεθος παραθύρου για τα τέσσερα μεγέθη θησαυρού, για την έννοια βλάστηση	127
10.1	Μερικές από τις εικόνες που χρησιμοποιήθηκαν για ταξινόμηση, εικόνες που απεικονίζουν εξωτερικό μέρος στις δύο πρώτες σειρές και εσωτερικό στις δύο τελευταίες .	131
10.2	Αποτελέσματα ταξινόμησης εικόνων αυτοκινήτων	132

Μέρος Ι
Οπτικοί Περιγραφείς MPEG-7

Κεφάλαιο 1

Εισαγωγή στο MPEG-7

Ιδρυμένη το 1988 η ομάδα ειδικών κινούμενης εικόνας (Moving Picture Experts Group - MPEG) ανέπτυξε οπτικοακουστικά πρότυπα συμπίεσης. Τα πρότυπα αυτά άλλαξαν τον τρόπο με τον οποίο το οπτικοακουστικό υλικό παράγεται από πολλές εταιρίες, διανέμεται από πολλών ειδών κανάλια και χρησιμοποιείται από ποικιλία συσκευών. Το MPEG είναι πολύ στενά συνδεδεμένο με τον ψηφιακό ήχο και βίντεο.

Με την χρήση των MPEG-1 και MPEG-2 τα αρχεία ήχου γίνονται εύχρηστα και οι οπτικοί δίσκοι μπορούν να χρησιμοποιηθούν για να αποθηκεύσουν ταινίες. Το MPEG-4 διευρύνει την δυνατότητα παράδοσης ήχου και βίντεο σε νέα περιβάλλοντα όπως το διαδίκτυο και τα κινητά τηλέφωνα. Στις μέρες μας η παραγωγή πολυμεσικού υλικού είναι πιο εύκολη από ποτέ, χάρη σε ένα μεγάλο πλήθος ψηφιακών συσκευών (ψηφιακές φωτογραφικές μηχανές, ψηφιακές βιντεοκάμερες, σαρωτές κλπ.). Η διανομή του έγινε και αυτή ιδιαίτερα εύκολη εξαιτίας των προσωπικών υπολογιστών, του διαδικτύου και ιδιαίτερα των τριών αυτών προτύπων MPEG. Η ίδια ψηφιακή τεχνολογία που μεγαλώνει την ευκολία παραγωγής και δημοσιοποίησης του μπορεί επίσης να βοηθήσει στην ανάλυση, ταξινόμηση, αναζήτηση και ανάκτηση του. Με ή χωρίς αυτόματη υποστήριξη η πληροφορία για το περιεχόμενο είναι απαραίτητη για την εύρεση και διαχείριση του.

Οι μηχανές αναζήτησης που βασίζονται σε κείμενο απαιτούν από τον άνθρωπο να έχει χειροκίνητα περιγράψει το περιεχόμενο με λέξεις κλειδιά και ελεύθερους σχολιασμούς. Αυτό έχει δύο σημαντικά προβλήματα. Το πρώτο είναι πως αυτή η ανθρώπινη διαδικασία κοστίζει και θα κοστίζει ακόμα περισσότερο καθώς ο όγκος του περιεχομένου θα αυξάνεται. Το δεύτερο πως αυτή η περιγραφή υποκύπτει στην ανθρώπινη υποκειμενικότητα. Έτσι αναδεικνύεται η ανάγκη αυτόματης και αντικειμενικής περιγραφής, δεικτοδότησης και σχολιασμού της πολυμεσικής πληροφορίας με χρήση εργαλείων που θα εξάγουν αυτόματα οπτικοακουστικά χαρακτηριστικά από το περιεχόμενο για να αντικαταστήσουν ή συμπληρώσουν την χειροκίνητη και βασισμένη στο κείμενο περιγραφή. Αυτά τα αυτομάτως εξαχθέντα οπτικοακουστικά χαρακτηριστικά θα έχουν τρία προτερήματα σε σχέση με τον ανθρώπινο σχολιασμό: (1) παράγονται αυτόματα, (2) είναι πιο αντικειμενικά και ανεξάρτητα του θεματικού χώρου και (3) μπορούν να είναι εγγενή στο οπτικοακουστικό περιεχόμενο. Οι έμφυτες αυτές περιγραφές χρησιμοποιούν δεδομένα που δεν έχουν την μορφή κειμένου για να περιγράψουν το περιεχόμενο. Χρησιμοποιούν χρώμα, υφή, σχήμα, μελωδία κ.α. με έναν τρόπο που θα επιτρέψει στον χρήστη να κάνει αναζήτηση συγκρίνοντας τις περιγραφές. Μολονότι οι αυτόματες περιγραφές είναι πολύ χρήσιμες, πάντα υπάρχει η ανάγκη ενσωμάτωσης κειμένου για κάποιες έννοιες όπως εί-

ναι ο τίτλος. Τέτοιες περιγραφές βοηθούν τους χρήστες και τις εφαρμογές να αναγνωρίσουν, να ανακτήσουν και να φιλτράρουν οπτικοακουστική πληροφορία.

Το παραπάνω πρόβλημα και η τεχνολογική κατάσταση είχαν αναγνωριστεί από το MPEG μέχρι που αποφάσισε να ξεκινήσει ένα έργο προτυποποίησης, γενικά γνωστό σαν MPEG-7 [1] [2], της Διεπαφής Περιγραφής Πολυμεσικού Περιεχομένου (Multimedia Content Description Interface) το οποίο επίσημα ξεκίνησε το 1997. Ο απώτερος σκοπός του είναι να παρέχει διαχρηστικότητα ανάμεσα στα συστήματα και τις εφαρμογές που χρησιμοποιούνται στην δημιουργία, διαχείριση, διανομή και χρήση περιγραφών πολυμεσικού περιεχομένου.

Το MPEG-7 αποδείχθηκε ότι είναι ένα πολύ συμπαγές και αποτελεσματικό πρότυπο. Το ακουστικό και το οπτικό τμήμα παρέχουν προτυποποιημένους “ακουστικούς μόνο” και “οπτικούς μόνο” περιγραφείς, το τμήμα πολυμεσικού σχήματος περιγραφής [Multimedia Description Schemes - MDS] παρέχει προτυποποιημένα σχήματα περιγραφής που εμπλέκουν οπτικούς και ακουστικούς περιγραφείς. Η γλώσσα καθορισμού περιγραφής [Description Definition Language - DDL] [3] παρέχει μία προτυποποιημένη γλώσσα για να εκφράσει τα σχήματα περιγραφής και το τμήμα συστήματος [4] παρέχει την δυνατότητα χρήσης του προτύπου σε πρακτικά περιβάλλοντα.

Όπως τα MPEG-1 και MPEG-2 παρείχαν τα εργαλεία τα οποία συνέβαλαν στη τωρινή αφθονία οπτικοακουστικού υλικού, έτσι και το MPEG-7 παρέχει τα μέσα για να πλοηγηθεί κανείς σε αυτήν την αφθονία περιεχομένου. Ωστόσο όλες οι άλλες πρωτοβουλίες “μεταδεδομένων” έχουν αναπτυχθεί για να εξυπηρετήσουν τις συγκεκριμένες ανάγκες ενός επιχειρησιακού περιβάλλοντος. Το MPEG-7 δεν έχει καμία δέσμευση ούτε από τους τρόπους με τους οποίους οι εταιρίες μεταφέρουν και παραδίδουν το πολυμεσικό περιεχόμενο αλλά ούτε και από τους τρόπους με τους οποίους οι χρήστες το λαμβάνουν κι εδώ εναποτίθεται η πραγματική του αξία παρέχοντας γενική λύση αγνοώντας το περιβάλλον.

Συνοπτικά, το MPEG-7 ακολουθεί την επιτυχή ανάπτυξη των προηγούμενων προτύπων με μία βασική διαφορά: ενώ τα προηγούμενα πρότυπα εστίαζαν στην κωδικοποίηση και αναπαράσταση του οπτικοακουστικού περιεχομένου, αυτό εστιάζει στην περιγραφή του. Αναφέρεται στο περιεχόμενο με διάφορες μορφές όπως εικόνα, βίντεο, ήχος, ομιλία, γραφικά και τους συνδυασμούς τους. Σε ένα τόσο ευρέως διαδεδομένο, πολλαπλών συσκευών και πλούσιο στην πληροφορία περιβάλλον, υπάρχουν κρίσιμες ανάγκες για εργαλεία και συστήματα για δεικτοδότηση, αναζήτηση, φιλτράρισμα και διαχείριση οπτικοακουστικού περιεχομένου, αποθηκευμένου ή ζωντανού κι εδώ έρχεται να συνεισφέρει το MPEG-7. Τέτοιες υπηρεσίες είναι χρήσιμες όχι μόνο για τους τελικούς χρήστες αλλά και για τους παροχούς περιεχομένου ή υπηρεσιών. Το MPEG-7 επιτυγχάνει την μέγιστη διαχρηστικότητα και διευκολύνει την δημιουργία καινοτόμων εφαρμογών, ορίζοντας τη σύνταξη και τη σημασιολογία διαφόρων εργαλείων περιγραφής.

Παραδείγματα εφαρμογών στα οποία μπορεί να γίνει χρήση του προτύπου περιλαμβάνουν: επιλογή πολυμεσικής εκπομπής (προσωποποιημένο ραδιόφωνο, κανάλια τηλεόρασης), ψηφιακές βιβλιοθήκες (κατάλογος εικόνων, μουσικό λεξικό), ευρετήρια πολυμεσικών υπηρεσιών (yellow pages) και διαμόρφωση του πολυμεσικού υλικού (προσωποποιημένη υπηρεσία ηλεκτρονικών νέων). Η χρήση των περιγραφών του MPEG-7 έχει σαν αποτέλεσμα έναν ευέλικτο και κλιμακωτό σκελετό για την σχεδίαση υπηρεσιών στις οποίες μπορούν να έχουν πρόσβαση μία ποικιλία τερματικών συσκευών όπως κινητά τηλέφωνα, προσωπικοί υπολογιστές κ.α. Ανάμεσα σε αυτήν την ποικιλία πιθανών εφαρμογών οι περιγραφές του MPEG-7 επιτρέπουν στους χρήστες ή τις εφαρμογές να διεκπεραιώσουν τα ακόλουθα έργα:

- *Πολυμέσα*: Δημιουργία ενός προσωποποιημένου οδηγού ή περίληψης εκπαιδόμενου πολυμεσικού περιεχομένου ανάλογα με την προτίμηση του χρήστη και το ιστορικό χρήσης.
- *Αρχείο*: Δημιουργία περιγραφών ξεχωριστών στοιχείων ή όλης της συλλογής οπτικοακουστικού περιεχομένου και έπειτα ανταλλαγή περιεχομένου και αντίστοιχων περιγραφών ανάμεσα σε ιδιοκτήτες περιεχομένου και καταναλωτές.
- *Προσαρμογή*: Φιλτράρισμα και μετατροπή των πολυμεσικών ροών δεδομένων σε περιβάλλοντα περιορισμένων πόρων (κινητά δίκτυα και τερματικά) ταιριάζοντας τις προτιμήσεις του χρήστη, τους διαθέσιμους πόρους και τις περιγραφές περιεχομένου.
- *Μουσική-Ήχος*: Παίζοντας μερικές νότες στο πιάνο, επιστροφή ενός μουσικού κομματιού που περιέχει τη μελωδία που ταιριάζει με τις νότες.
- *Γραφικά*: Ζωγραφίζοντας ένα πρόχειρο σχέδιο, επιστροφή ενός συνόλου εικόνων που περιέχουν παρόμοια γραφικά, λογότυπα κ.α.
- *Κίνηση*: Με ένα δεδομένο σύνολο αντικειμένων βίντεο και περιγράφοντας την κίνηση και συσχέτιση μεταξύ των αντικειμένων, επιστροφή ενός συνόλου κινουμένων σχεδίων (animation) ή βίντεο που πληρούν τις περιγραφόμενες χωρικές και χρονικές συσχετίσεις.
- *Σενάριο*: Σε δεδομένο οπτικοακουστικό περιεχόμενο, περιγράφοντας τις ενέργειες, επιστροφή λίστας σεναρίων (τμήματα οπτικοακουστικού υλικού) όπου παρόμοια σενάρια λαμβάνουν χώρα.

Για την περιγραφή οπτικοακουστικού υλικού, το πρότυπο MPEG-7 καθορίζει ένα σύνολο από περιγραφείς οι οποίοι εξάγουν οπτική πληροφορία από εικόνες και βίντεο και δημιουργούν δομημένες και λεπτομερείς περιγραφές της. Οι περιγραφείς ορίζουν την σύνταξη και σημασιολογία χαρακτηριστικών οπτικοακουστικού περιεχομένου. Στο χαμηλό επίπεδο αφαίρεσης, οι περιγραφείς μπορούν να περιλάβουν *σχήμα*[5], *κίνηση*, *υφή*[6], *χρώμα*[6], και *κίνηση κάμερας*[7] για εικόνες/βίντεο, *ενέργεια*, *αρμονικότητα*, και *χροιά* για τον ήχο[8]. Στο υψηλό επίπεδο αφαίρεσης, μπορούν να περιλάβουν *γεγονότα*, *αφηρημένες έννοιες*, *είδη περιεχομένου* κ.α. Οι ακουστικοί και οπτικοί περιγραφείς αντιπροσωπεύουν συγκεκριμένα χαρακτηριστικά γνωρίσματα σχετικά με το ακουστικό και οπτικό περιεχόμενο αντίστοιχα. Οι γενικοί περιγραφείς εξετάζουν γενικότερα χαρακτηριστικά γνωρίσματα. Τα σχήματα περιγραφής επιτρέπουν την κατασκευή σύνθετων περιγραφών με τη διευκρίνιση της δομής και της σημασιολογίας των σχέσεων μεταξύ των περιγραφών ή των σχεδίων περιγραφής. Για παράδειγμα, το σχήμα περιγραφής για ένα τμήμα βίντεο μπορεί να καθορίσει τη σύνταξη και τη σημασιολογία των συστατικών στοιχείων όπως αποσύνθεση τμήματος, μεμονωμένες ιδιότητες τμήματος (π.χ., μήκος τμήματος, σχολιασμοί σε μορφή κειμένου), και σχέσεις μεταξύ των συστατικών τμημάτων.

Τα σχήματα περιγραφής, όπως στην περίπτωση των περιγραφών, μπορούν να ταξινομηθούν σε *ακουστικά*, *οπτικά*, ή *γενικά*. Τα γενικά σχήματα περιγραφής αντιπροσωπεύουν συνήθως τις γενικές μετα-πληροφορίες σχετικές με όλα τα είδη μέσων (ακουστικά, οπτικά, κείμενο, γραφικά, κ.λ.π. . .). Πέρα από την ύπαρξη περιγραφών ή σχημάτων περιγραφής που προέρχονται από το περιεχόμενο, το MPEG-7 επίσης περιλαμβάνει περιγραφείς και σχήματα περιγραφής σχετικά με τη δημιουργία, την παραγωγή, τη διαχείριση, και την πρόσβαση του οπτικοακουστικού περιεχομένου. Τέτοια μεταδεδομένα

μπορούν να συμπεριλάβουν τις πληροφορίες για το “σχήμα κωδικοποίησης” που χρησιμοποιείται για τη συμπίεση του περιεχομένου (π.χ., JPEG, MPEG-2), τα γενικά στοιχεία (πχ. μέγεθος, διάρκεια), όρους για πρόσβαση στο υλικό (π.χ., πληροφορίες δικαιωμάτων πνευματικής ιδιοκτησίας καθώς και οικονομικές πληροφορίες), ταξινόμηση (γονική εκτίμηση, και κατηγοριοποίηση περιεχομένου στις διάφορες προκαθορισμένες κατηγορίες), και τις συνδέσεις με άλλο σχετικό υλικό. Οι πληροφορίες αυτές μπορούν να βοηθήσουν το χρήστη να επιταχύνει την αναζήτηση.

Το MPEG-7 δίνει την δυνατότητα για περιγραφές οπτικοακουστικού περιεχομένου σε διαφορετικά αντιληπτικά και σημασιολογικά επίπεδα. Τα χαρακτηριστικά χαμηλού επιπέδου (όπως χρώμα) μπορούν να εξαχθούν αυτόματα, ενώ τα χαρακτηριστικά υψηλού επιπέδου που περιγράφουν τη σημασιολογική πληροφορία χρειάζονται ανθρώπινη αλληλεπίδραση.

Πρέπει να γίνει κατανοητό ότι για τους περιγραφείς και τα σχήματα περιγραφής, το MPEG-7 δεν διευκρινίζει πως να εξαχθούν αυτές οι περιγραφές αφήνοντας σχετική ευελιξία στις διάφορες εφαρμογές. Εντούτοις, απαραίτητη προϋπόθεση είναι η αναπαράσταση αυτών των περιγραφών να είναι σύμφωνη με το πρότυπο. Οι MPEG-7 δυαδικές ή μη δυαδικές περιγραφές μπορούν να προσπελαστούν, να κατανοηθούν και να χρησιμοποιηθούν από εφαρμογές που είναι σε θέση να τις αποκωδικοποιήσουν και να τις επεξεργαστούν. Ούτε το πως οι MPEG-7 περιγραφές θα χρησιμοποιηθούν για την περαιτέρω επεξεργασία οπτικοακουστικού υλικού δεν είναι προκαθορισμένο, κάτι που επίσης παρέχει την απαιτούμενη ευελιξία στις διάφορες εφαρμογές. Πρέπει επίσης να σημειωθεί ότι οι MPEG-7 περιγραφές μπορούν να συνυπάρξουν με το οπτικοακουστικό υλικό με το οποίο σχετίζονται στην ίδια ροή δεδομένων ή στο ίδιο σύστημα αποθήκευσης. Εναλλακτικά, οι περιγραφές θα μπορούσαν επίσης να βρεθούν οπουδήποτε αλλού, εφόσον είναι δυνατό να συνδεθεί αποτελεσματικά το οπτικοακουστικό υλικό με την αντίστοιχη περιγραφή του.

Ο αυξανόμενος όγκος των οπτικοακουστικών δεδομένων που είναι διαθέσιμα στις καθημερινές ζωές μας απαιτεί αποτελεσματικά συστήματα πολυμέσων. Τέτοια συστήματα πρέπει να διευκολύνουν την πρόσβαση, την αλληλεπίδραση και την παρουσίαση των σύνθετων και ετερογενών πληροφοριών. Τέτοιες ανάγκες συσχετίζονται με τα σημαντικά κοινωνικά και οικονομικά ζητήματα, και είναι επιτακτικές σε διάφορες περιπτώσεις επαγγελματικών και καταναλωτικών εφαρμογών. Τέτοιες εφαρμογές μπορεί να είναι οι ήδη υπάρχουσες οι οποίες βελτιώθηκαν με την χρήση του προτύπου MPEG-7 αλλά και νέες οι οποίες έγιναν πραγματικότητα χάρη στο συγκεκριμένο πρότυπο. Πιο συγκεκριμένα, τέτοιες εφαρμογές συναντώνται στις παρακάτω υπηρεσίες:

- *Εκπαίδευση* (Αναζήτηση ντοκιμαντέρ με βάση το περιεχόμενο τους σχετικά με το εκπαιδευτικό θέμα)
- *Δημοσιογραφία* (Αναζήτηση ομιλιών ενός συγκεκριμένου πολιτικού χρησιμοποιώντας το όνομά του, τη φωνή του ή το πρόσωπο του)
- *Τουριστικές Πληροφορίες*
- *Πολιτιστικές Πληροφορίες* (ιστορικά μουσεία, αίθουσες εκθέσεων τέχνης κ.α.)
- *Ψυχαγωγία* (αναζήτηση ενός παιχνιδιού, karaoke)
- *Υπηρεσίες Εξιχνίασης* (αναγνώριση ανθρωπίνων χαρακτηριστικών)
- *Γεωγραφικά Συστήματα Πληροφοριών*

- Απομακρυσμένα συστήματα αισθητήρων (χαρτογραφία, οικολογία, διαχείριση φυσικών πόρων κ.α.)
- Επίβλεψη (έλεγχος κυκλοφοριακής κίνησης κ.α.)
- Βιοιατρικές εφαρμογές (ταξινόμηση ιατρικού φωτογραφικού υλικού)
- Αγορές (αναζήτηση ρούχων συγκεκριμένης προτίμησης)
- Αρχιτεκτονική (κτιματομεσιτικά, σχέδιο εσωτερικού χώρου)
- Κοινωνικά (Υπηρεσίες γνωριμιών)
- Αρχεία ταινιών, βίντεο και ραδιοφωνικού υλικού (Αναζήτηση ραδιοφωνικού υλικού με την φωνή συγκεκριμένου ατόμου)

Τέλος κατά την διαδικασία προτυποποίησης του MPEG-7 διεξήχθησαν πειράματα για να συγκριθούν ανταγωνιστικές τεχνολογίες αλλά και να βελτιστοποιηθούν οι υιοθετημένες μέθοδοι. Για τους οπτικούς περιγραφείς μία εφαρμογή ανάκτησης αποδείχθηκε πως είναι η καλύτερη μέθοδος για να διεξαχθούν τα πειράματα. Ένα ικανοποιητικό αποτέλεσμα ανάκτησης βασισμένο σε ένα οπτικό χαρακτηριστικό είναι ενδεικτικό της καλής εκφραστικότητας του. Στα πειράματα σαν κύρια μέθοδος αξιολόγησης χρησιμοποιήθηκε το ερώτημα με παράδειγμα (query by example). Σε αυτή την μέθοδο ο αντίστοιχος κάθε φορά περιγραφέας εξάγεται από την εικόνα του ερωτήματος και έπειτα γίνεται ταίριασμα με τους περιγραφείς των εικόνων που βρίσκονται στην βάση δεδομένων. Για να είναι όσο πιο αντικειμενικές γίνεται οι συγκρίσεις αναπτύχθηκε ένα ποσοτικό μέτρο με βάση τις προδιαγραφές του συνόλου δεδομένων, του συνόλου ερωτημάτων και του αντίστοιχου συνόλου δεδομένης αλήθειας (ground-truth). Το ground-truth είναι ένα σύνολο από εικόνες παρόμοιες οπτικά για την δεδομένη εικόνα του ερωτήματος.

Για να οριστεί ένα αντικειμενικό ποσοτικό μέτρο για την αποτελεσματικότητα της ανάκτησης εικόνων δεδομένου ενός συνόλου ερωτημάτων και το αντίστοιχο ground-truth λήφθηκαν υπόψη τα παρακάτω:

- Υπάρχει η ανάγκη κανονικοποίησης λόγω του διαφορετικού μεγέθους ground-truth
- Πρέπει να επιβραβεύονται αλγόριθμοι που επιστρέφουν τα αντικείμενα του ground-truth σαν τα καλύτερα ταίριασμα.
- Πρέπει να τιμωρείται η απουσία κάθε χαμένου αντικειμένου από το ground-truth.
- Το μέτρο να λαμβάνει υπόψη την σειρά με την οποία τα αντικείμενα του ground-truth ανακτώνται.

Έτσι υιοθετήθηκε η εξής λύση: Θεωρώντας ένα ερώτημα q με ένα ground-truth μεγέθους $NG(q)$, η τάξη $Rank(k)$ της k -οστής εικόνας του ground-truth ορίζεται σαν η θέση στην οποία η εικόνα αυτή ανακτάται (μία τιμή τάξης ένα αντιστοιχεί στο καλύτερο ταίριασμα). Επιπλέον ορίζεται ένας αριθμός $K(q) \geq NG(q)$ ο οποίος προσδιορίζει τις "σχετικές τάξεις", έτσι ανάκτηση με τάξη μεγαλύτερη από $K(q)$ θεωρείται σαν αστοχία. Για εικόνες από το ground-truth που δεν ανακτήθηκαν με τάξη

μικρότερη ή ίση από $K(q)$ η κύρωση που επιβάλλεται πρέπει να είναι μεγαλύτερη ή ίση με $K(q)$. Μια κύρωση ίση με $K(q)$ θα έθετε σε πλεονεκτική θέση μεθόδους ανάκτησης με πολλές αστοχίες. Έτσι το $Rank(k)$ ορίζεται όπως στην σχέση (1.1).

$$Rank(k) = \begin{cases} Rank(k) & \text{if } Rank(k) \leq K(q) \\ 1.25 \cdot K(q) & \text{if } Rank(k) > K(q) \end{cases} \quad K(q) = \min\{4 \cdot NG(q), 2 \cdot \max[NG(q), \forall q]\} \quad (1.1)$$

η σχέση (1.2) δίνει την μέση τάξη (Average Rank - AVR)

$$AVR(q) = \frac{1}{NG(q)} \sum_{k=1}^{NG(q)} Rank(k) \quad (1.2)$$

για να μειωθεί η επιρροή των διαφορετικών μεγεθών ground-truth ορίστηκε και η τροποποιημένη τάξη ανάκτησης (Modified Retrieval Rank - MRR) που δίνεται από την σχέση (1.3)

$$MRR(q) = AVR(q) - 0.5 \cdot [1 + NG(q)] \quad (1.3)$$

Ο τελευταίος τύπος οδηγεί τελικά στην σχέση (1.4) η οποία δίνει την κανονικοποιημένη τροποποιημένη τάξη ανάκτησης (Normalized Modified Retrieval Rank - NMRR)

$$NMRR(q) = \frac{MRR(q)}{1.25 \cdot K(q) - 0.5 \cdot [1 + NG(q)]} \quad (1.4)$$

Από την σχέση (1.4) μπορεί να οριστεί η μέση κανονικοποιημένη τροποποιημένη τάξη ανάκτησης η οποία υπολογίζεται ανάμεσα σε όλα τα ερωτήματα q , όπου NQ το πλήθος τους. Η μέση κανονικοποιημένη τροποποιημένη τάξη ανάκτησης (Average Normalized Modified Retrieval Rank - ANMRR) δίνεται από την σχέση (1.5).

$$ANMRR = \frac{1}{NQ} \sum_{q=1}^{NQ} NQNMRR(q) \quad (1.5)$$

Τα παραπάνω μέτρα τελικά χρησιμοποιήθηκαν από το MPEG-7 για τα πειράματα ανάκτησης προκειμένου να διερευνηθεί η δυνατότητα των περιγραφέων να περιγράφουν αποτελεσματικά την οπτικοακουστική πληροφορία χαμηλού επιπέδου.

Κεφάλαιο 2

Οπτικοί Περιγραφείς

Σε αυτή την ενότητα αναλύονται οι περιγραφείς που αποτελούν μέρος της εφαρμογής που δημιουργήθηκε και θα περιγραφεί στο μέρος II. Παρουσιάζονται περιγραφείς χρώματος, υφής και σχήματος. Αρχικά αναφέρονται τα ποιοτικά χαρακτηριστικά για τον καθένα. Έπειτα τα μέτρα που χρησιμοποιούνται για τον υπολογισμό της απόστασης μεταξύ δύο περιγραφών και ο τρόπος εξαγωγής τους. Τέλος παρουσιάζονται αποτελέσματα πειραμάτων που αφορούν την απόδοση τους στην ανάκτηση εικόνων.

2.1 Περιγραφείς Χρώματος

2.1.1 Χρωματικοί Χώροι (Color Spaces)

Οι διάφοροι χρωματικοί χώροι που χρησιμοποιούνται στο MPEG-7 είναι Μονοχρωματικός, RGB, YCbCr, HSV και ο νέος HMMD. Ο RGB χώρος είναι ένα από τα πιο γνωστά χρωματικά μοντέλα και ορίζεται σαν μοναδιαίους κύβος στο καρτεσιανή σύστημα συντεταγμένων. Ένας χρωματικός χώρος που υιοθετήθηκε από τα προηγούμενα MPEG πρότυπα (MPEG-1,2,4) είναι ο YCbCr, ο οποίος ορίζεται σαν γραμμικός μετασχηματισμός από τον RGB. Υ είναι η συνιστώσα φωτεινότητας και Cb, Cr οι συνιστώσες που αντιστοιχούν στην διαφορά του μπλε και του κόκκινου χρώματος με την φωτεινότητα αντίστοιχα. Ο μονοχρωματικός περιλαμβάνει μόνο την Υ συνιστώσα του YCbCr. Ο HSV ορίζεται σαν μη γραμμικός αλλά αναστρέψιμος μετασχηματισμός από το RGB. Το H (Hue) είναι η χροιά του χρώματος και αναπαρίσταται από μία γωνία 0° έως 360° που διαχωρίζει την μία κατηγορία χρώματος από την άλλη (πχ. κίτρινο από κόκκινο). Το S (Saturation) είναι κορεσμός και καθορίζει το πόσο καθαρό είναι ένα χρώμα: πράσινο, κίτρινο, μπλε κλπ, με τιμές από 0 έως 1. Τέλος το V (Value) είναι η φωτεινότητα του χρώματος με τιμές από 0 έως 1. Στον χρωματικό χώρο HMMD (Hue-Max-Min-Diff) οι συνιστώσες του ορίζονται ως εξής από τον RGB χρωματικό χώρο:

$$Max = \max(R, G, B) \quad Min = \min(R, G, B) \quad Diff = Max - Min \quad (2.1)$$

Ο χρωματικός χώρος δεν είναι περιγραφέας αλλά είναι χρήσιμος για να γνωρίζουμε σε ποιον χώρο αναπαράστασης της εικόνας γίνονται οι υπολογισμοί για κάποιους περιγραφείς.

2.1.2 Περιγραφέας Κύριων Χρωμάτων (Dominant Color Descriptor)

Ένα σύνολο από αντιπροσωπευτικά χρώματα σε μια εικόνα ή σε μία περιοχή ενδιαφέροντος (region of interest) παρέχει μια συμπαγή αλλά αποτελεσματική περιγραφή. Στόχος είναι η χρήση σε εφαρμογές όπως ανάκτηση με βάση το χρώμα σε μεγάλες βάσεις δεδομένων. Όπως είναι προφανές τα χρώματα σε μια περιοχή μιας εικόνας είναι συγκεντρωμένα γύρω από έναν μικρό αριθμό αντιπροσωπευτικών χρωμάτων. Έτσι ο περιγραφέας αυτός αποτελείται από τα κύρια (αντιπροσωπευτικά) χρώματα, τα ποσοστά τους στην περιοχή, τη χωρική τους συνοχή (spatial coherency), και τη διακύμανσή τους (variance). Τα κύρια χρώματα μπορούν να αναπαρασταθούν στο τρισδιάστατο χώρο χρώματος αποφεύγοντας έτσι τα υψηλής διάστασης προβλήματα δεικτοδότησης που συνδέονται με το παραδοσιακό ιστόγραμμα χρώματος.

Για τον υπολογισμό του περιγραφέα κύριων χρωμάτων, τα χρώματα που είναι παρόντα σε μια δεδομένη εικόνα ή μια περιοχή ενδιαφέροντος πρώτα ομαδοποιούνται. Αυτό οδηγεί σε έναν μικρό αριθμό χρωμάτων. Στη συνέχεια υπολογίζονται τα ποσοστά αυτών των χρωμάτων. Προαιρετικά υπολογίζονται επίσης οι διακυμάνσεις των κύριων χρωμάτων. Επίσης υπολογίζεται μια τιμή χωρικής συνοχής που βοηθά στην διαφοροποίηση μεταξύ των μεγάλων ενιαίων περιοχών χρώματος εναντίον των χρωμάτων που είναι εξαπλωμένα σε όλη την εικόνα. Ο αριθμός των κυρίων χρωμάτων μπορεί να ποικίλει από εικόνα/περιοχή σε εικόνα/περιοχή και ένα μέγιστο οκτώ κύριων χρωμάτων μπορεί να χρησιμοποιηθεί για να περιγράψει την εικόνα/περιοχή. Η τιμή του ποσοστού κβαντίζεται σε 5 bits.

Ο περιγραφέας κύριων χρωμάτων ορίζεται από την σχέση (2.2):

$$DCD = \{(c_i, p_i, v_i), s\}, i = 1, 2, \dots, N \quad (2.2)$$

Όπου c_i είναι το i -οστό κύριο χρώμα, p_i το ποσοστό του, v_i η διακύμανσή του και s η χωρική συνοχή. Για τη σύγκριση μεταξύ δύο περιγραφέντων κύριων χρωμάτων, έστω DCD_1 και DCD_2 , το πρότυπο MPEG-7 ορίζει την συνάρτηση (2.3).

$$D^2(DCD_1, DCD_2) = \sum_{i=1}^{N_1} p_{1i}^2 + \sum_{j=1}^{N_2} p_{2j}^2 - \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} 2a_{1i,2j} p_{1i} p_{2j} \quad (2.3)$$

Στην σχέση (2.3), οι δείκτες 1 και 2 αντιστοιχούν στους περιγραφείς DCD_1 και DCD_2 , αντίστοιχα $a_{1i,2j}$ είναι οι συντελεστές ομοιότητας ανάμεσα σε δύο χρώματα c_1 και c_2 που ορίζονται από την σχέση (2.4)

$$a_{k,l} = \begin{cases} 1 - \frac{d_{k,l}}{d_{max}}, & d_{k,l} \leq T_d \\ 0, & d_{k,l} > T_d \end{cases} \quad (2.4)$$

Στην σχέση αυτή $d_{k,l} = \|c_k - c_l\|$ είναι η Ευκλείδεια απόσταση μεταξύ δύο χρωμάτων, T_d είναι το κατώφλι κάτω από το οποίο δύο χρώματα θεωρούνται παρόμοια και $d_{max} = a \cdot T_d$. Η σχέση αυτή μας δείχνει, επιπλέον, ότι δύο χρώματα θεωρούνται παρόμοια αν απέχουν τουλάχιστον T_d . Μια καλή επιλογή, στην πράξη, για την απόσταση T_d είναι μεταξύ 10 και 20 στον CIE-LUV χρωματικό χώρο [9]. Αντίστοιχα, μια καλή επιλογή για το a είναι μεταξύ 1-1.5. Τέλος οι διακυμάνσεις και η χωρική

συνοχή αγνοούνται για τον προσδιορισμό της απόστασης μεταξύ δύο περιγραφών κύριων χρωμάτων στην σχέση (2.3).

Μία παραλλαγή της απόστασης είναι με χρήση του πεδίου χωρικής συνοχής όπως στα πειράματα του MPEG-7 (2.5).

$$D_S = w_1 |s_1 - s_2| \cdot D + w_2 D \quad (2.5)$$

Όπου τα w_1 και w_2 είναι βάρη με προτεινόμενες τιμές 0.3 και 0.7 αντίστοιχα.

Ωστόσο, ο τύπος της απόστασης μπορεί να επεκταθεί, προκειμένου να συμπεριλάβει και την προαιρετική διακύμανση (2.6).

$$D_V = \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} p_{1i} p_{1j} f_{1i1j} + \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} p_{2i} p_{2j} f_{2i2j} - \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} 2p_{1i} p_{2j} f_{1i2j} \quad (2.6)$$

όπου

$$f_{x_i y_j} = \frac{1}{2\pi \sqrt{v_{x_i y_j}^{(l)} v_{x_i y_j}^{(u)} v_{x_i y_j}^{(v)}}} \exp \left[- \left(\frac{c_{x_i y_j}^{(l)}}{v_{x_i y_j}^{(l)}} + \frac{c_{x_i y_j}^{(u)}}{v_{x_i y_j}^{(u)}} + \frac{c_{x_i y_j}^{(v)}}{v_{x_i y_j}^{(v)}} \right) / 2 \right] \quad (2.7)$$

και

$$c_{x_i y_j}^{(l)} = (c_{x_i}^{(l)} - c_{y_j}^{(l)})^2, v_{x_i y_j}^{(l)} = (v_{x_i}^{(l)} + v_{y_j}^{(l)}) \quad (2.8)$$

Στις εξισώσεις (2.7) και (2.8), $c_{x_i}^{(l)}$ και $v_{x_i}^{(l)}$ είναι τιμές των κύριων χρωμάτων και των χρωματικών διακυμάνσεων, x και y δεικτοδοτούνε τους εξεταζόμενους (query) περιγραφείς και τους περιγραφείς στόχους, i, j δεικτοδοτούνε τις συνιστώσες του περιγραφέα και l, u, v οι συνιστώσες του χρωματικού χώρου.

Η διαδικασία εξαγωγής για τον περιγραφέα κύριων χρωμάτων χρησιμοποιεί τον γενικευμένο αλγόριθμο Lloyd (GLA) για να συσταδοποιήσει τις χρωματικές τιμές των εικονοστοιχείων (pixel). Συνίσταται η ομαδοποίηση να γίνεται σε έναν αντιληπτικά ομοιόμορφο χώρο όπως ο CIE-LUV. Αρχικά έχουμε μία συστάδα αποτελούμενη από όλα τα εικονοστοιχεία και ένα αντιπροσωπευτικό χρώμα υπολογισμένο σαν το κέντρο μάζας (centroid) της. Ο αλγόριθμος ύστερα ακολουθεί μία σειρά υπολογισμών κέντρων μάζας και ομαδοποιήσεων μέχρι να ικανοποιηθεί ένα κριτήριο (ελάχιστη διαστρέβλωση ή μέγιστος αριθμός επαναλήψεων). Οι ομάδες με την υψηλότερη διαστρέβλωση (distortion) διαιρούνται μέχρι να πέσει κάτω από ένα προκαθορισμένο κατώφλι η μέγιστη διαστρέβλωση ή να έχει δημιουργηθεί ο μέγιστος αριθμός ομάδων. Με τον όρο διαστρέβλωση εννοείται η διαφοροποίηση των χρωμάτων της ομάδας από το χρώμα του κέντρου. Τέλος υπολογίζονται τα ποσοστά των εικονοστοιχείων της εικόνας ή της περιοχής ενδιαφέροντος που ανήκουν σε κάθε κβαντισμένο χρώμα και κβαντίζονται ομοιόμορφα με 5 bits.

Μία σύγκριση του ANMRR (Average Normalized Modified Retrieval rate) για δύο διαφορετικούς μέσους όρους κύριων χρωμάτων φαίνεται στον πίνακα 2.1. Παρατηρείται ότι ικανοποιητικά αποτελέσματα πετυχαίνονται και για την βασική έκδοση του περιγραφέα και ότι σημαντική βελτίωση μπορεί να επιτευχθεί χρησιμοποιώντας τα προαιρετικά πεδία.

Ο πίνακας 2.2 δίνει αποτελέσματα χρησιμοποιώντας την παράμετρο χωρικής διακύμανσης και συγκρίνοντας με τον περιγραφέα κύριων χρωμάτων χωρίς διακύμανση. Αυτά τα πειράματα αναδεικνύουν πως η χρήση 5 bits για το πεδίο χωρικής συνοχής είναι μία λογική εξισορρόπηση ανάμεσα

Μέσος αριθμός χρωμάτων	ANMRR(D)	ANMRR(D_S)	ANMRR(D_V)
3	0.31	0.30	0.25
5	0.25	0.21	0.16

Πίνακας 2.1: Αποτελέσματα ANMRR για τον περιγραφέα κύριων χρωμάτων

στην πολυπλοκότητα και την αποτελεσματικότητα του περιγραφέα. Η αυξημένη πολυπλοκότητα του περιγραφέα ήταν ο κύριος παράγοντας που επιλέχθηκε μία μοναδική τιμή χωρικής συνοχής.

Αριθμός bits για την χωρική συνοχή	ANMRR	
	Πεδίο χωρικής συνοχής για όλα τα κύρια χρώματα	Πεδίο χωρικής συνοχής για κάθε κύριο χρώμα
5	0.221	
4	0.227	
3	0.246	
2	0.250	0.197
1	0.252	0.202
0	0.252 (χωρίς τιμή χωρικής συνοχής)	

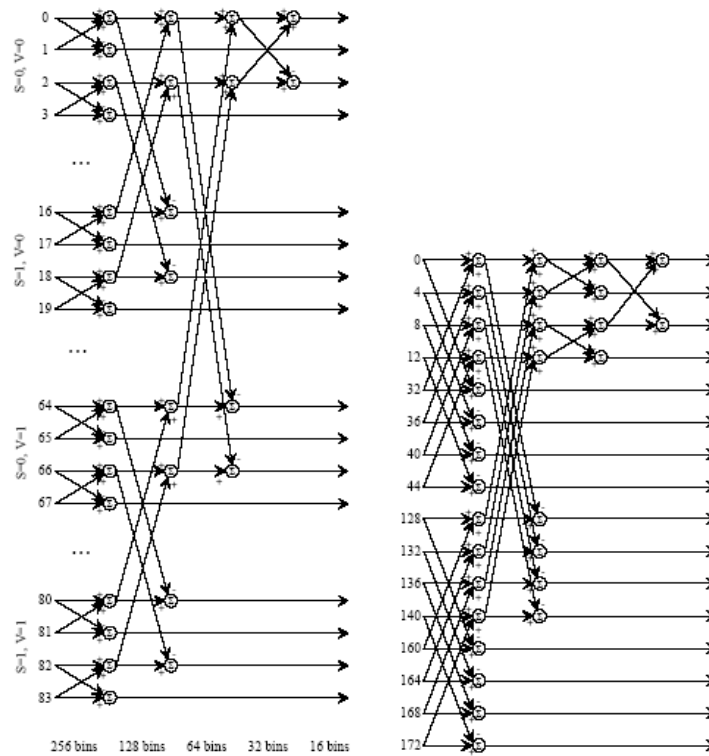
Πίνακας 2.2: Αποτελέσματα ANMRR για τον περιγραφέα κύριων χρωμάτων με χωρική συνοχή

2.1.3 Κλιμακωτός Περιγραφέας Χρώματος (Scalable Color Descriptor)

Ο κλιμακωτός περιγραφέας χρώματος επιτρέπει μια κλιμακωτή αναπαράσταση της περιγραφής, καθώς επίσης και την κλιμακωτή πολυπλοκότητα για την εξαγωγή χαρακτηριστικών γνωρισμάτων και τις διαδικασίες ταυρίσματος. Ο συγκεκριμένος περιγραφέας μπορεί να ερμηνευτεί σαν ένα σχήμα κωδικοποίησης βασισμένο στην μετατροπή Haar που εφαρμόζεται στις τιμές ενός ιστογράμματος χρώματος στον HSV χρωματικό χώρο. Οι τιμές του ιστογράμματος εξάγονται, κανονικοποιούνται, και μετατρέπονται μη γραμμικά σε μία ακέραια αναπαράσταση τεσσάρων bits, η οποία δίνει περισσότερη σημασία στις μικρές τιμές. Η μετατροπή Haar εφαρμόζεται πάνω στις ακέραιες τιμές τεσσάρων bits του ιστογράμματος. Η βασική μονάδα της μετατροπής αυτής αποτελείται από μια λειτουργία αθροίσματος (βαθυπερατό φίλτρο) και μια λειτουργία διαφοράς (υψιπερατό φίλτρο). Αθροίζοντας ζευγάρια γειτονικών κορυφών (bins) είναι ισοδύναμο με τον υπολογισμό ενός ιστογράμματος με τον μισό αριθμό κορυφών. Από τα αθροίσματα των κάθε δύο γειτονικών τιμών χροιάς (H) από ένα ιστογράμμα 256 κορυφών προκύπτει αναπαράσταση ενός ιστογράμματος 128 κορυφών. Αν επαναληφθεί αυτή η διαδικασία τότε προκύπτουν ιστογράμματα με 64, 32, 16 κορυφές. Αυτή η μέθοδος επιτυγχάνει

την πλήρη διαλειτουργικότητα μεταξύ των διαφορετικών αναλύσεων της αναπαράστασης χρώματος, που κυμαίνεται από 16 bits/ιστόγραμμα στο χαμηλό όριο ως περίπου 1000 bits/ιστόγραμμα στο υψηλό όριο. Φυσικά, η ακρίβεια της περιγραφής χαρακτηριστικών γνωρισμάτων εξαρτάται ιδιαίτερα από τον αριθμό των χρησιμοποιούμενων bits. Εντούτοις, τα πειράματα έχουν δείξει ότι ικανοποιητικά αποτελέσματα ανάκτησης είναι ακόμα επιτεύξιμα χρησιμοποιώντας μόνο 64 bits, ενώ τα άριστα αποτελέσματα μπορούν να επιτευχθούν χρησιμοποιώντας τη μέση ή πλήρη ανάλυση του περιγραφέα.

Οι συντελεστές διαφοράς (υψιπερατοί) της μετατροπής Haar εκφράζουν πληροφορίες που περιέχονται σε επίπεδα υψηλότερης ανάλυσης (με μεγαλύτερο αριθμό κορυφών). Ιστογράμματα φυσικών εικόνων συνήθως αναδεικνύουν επανάληψη ανάμεσα στις γειτονικά κορυφές. Αυτό μπορεί να ερμηνευτεί από την μη καθαρότητα (μικρή διακύμανση) των χρωμάτων που προκαλείται από μεταβλητό φωτισμό και εφέ σκιάς. Έτσι είναι αναμενόμενο πως οι συντελεστές διαφοράς που εκφράζουν την διαφορά μεταξύ γειτονικών κορυφών συνήθως έχουν μικρές τιμές.



Σχήμα 2.1: Αριστερά: Τα 4 πρώτα επίπεδα Δεξιά: Επίπεδα 5-8

Ο κλιμακωτός περιγραφέας χρώματος είναι ένα ιστόγραμμα που ορίζεται από τη σχέση (2.9).

$$SCD = [c_1, c_1, \dots, c_N] \quad (2.9)$$

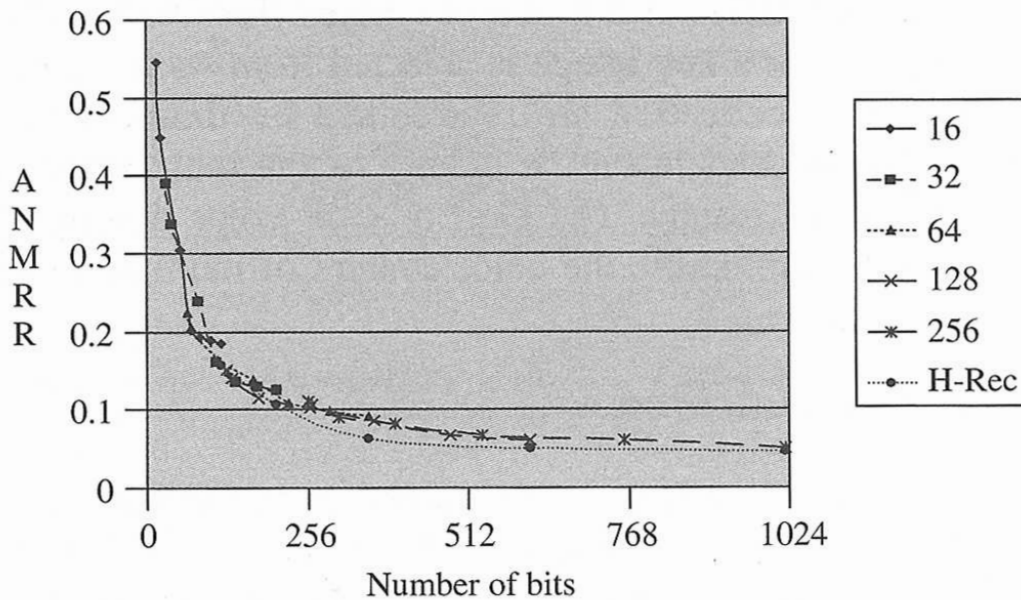
Όπου N είναι το μέγεθος του ιστογράμματος. Το πρότυπο MPEG-7 δεν ορίζει με αυστηρότητα κάποια συνάρτηση υπολογισμού απόστασης για το συγκεκριμένο περιγραφέα. Ωστόσο, στην πράξη

εφαρμόζεται με επιτυχία η πολύ γνωστή L1 απόσταση (2.10).

$$D(SCD_1, SCD_2) = \sum_{i=1}^N |c_{1i} - c_{2i}| \quad (2.10)$$

Η εξαγωγή αποτελείται από τον υπολογισμό ενός ιστογράμματος στον HSV χρωματικό χώρο ομοιόμορφα κβαντισμένο σε 256 κορυφές. Έπειτα οι τιμές του ιστογράμματος κβαντίζονται μη γραμμικά. Τέλος ο μετασχηματισμός Haar εφαρμόζεται σε αυτές τις τιμές όπως φαίνεται στο σχήμα 2.1.

Αποτελέσματα ανάκτησης που επιτεύχθηκαν από τον κλιμακωτό περιγραφέα χρώματος φαίνονται στο σχήμα 2.2.



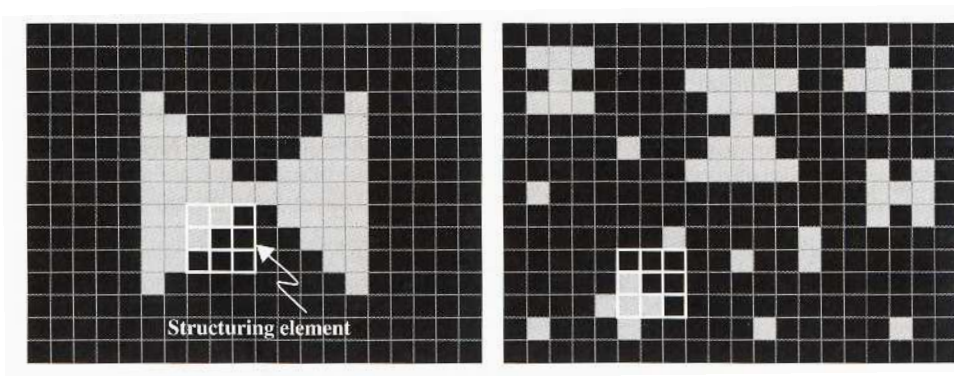
Σχήμα 2.2: Αποτελέσματα ανάκτησης με διαφορετικό αριθμό συντελεστών Haar (16-256) κβαντισμένων με διαφορετικό αριθμό bits. H-Rec είναι η ανάκτηση με το κατασκευασμένο ιστόγραμμα από Haar συντελεστές σε πλήρη ανάλυση

2.1.4 Περιγραφέας Δομής χρώματος (Color Structure Descriptor)

Ο περιγραφέας δομής χρώματος περιγράφει μία εικόνα ή μία περιοχή ενδιαφέροντος με βάση την κατανομή του χρώματος σε αυτήν με παρόμοιο τρόπο με ένα ιστόγραμμα χρώματος αλλά και την τοπική χωρική δομή του χρώματος. Έτσι μπορεί να ξεχωρίσει μεταξύ εικόνων που περιλαμβάνουν τα ίδια χρώματα, αλλά με διαφορετική δομή, κάτι στο οποίο ένα ιστόγραμμα χρώματος αποτυγχάνει. Η δομή ενός επιπέδου είναι ο βαθμός στον οποίο τα εικονοστοιχεία συγκεντρώνονται από κοινού

και σχηματίζουν συμπαγή αντικείμενα. Καταμετρείται ο αριθμός των φορών που ένα συγκεκριμένο χρώμα περιλαμβάνεται μέσα στο δομικό στοιχείο όσο εκείνο σαρώνει την εικόνα. Για να εξασφαλίσει διαλειτουργικότητα, το ιστόγραμμα δομών χρώματος υπολογίζεται στον HMMD χρωματικό χώρο.

Στο σχήμα 2.3 φαίνονται δύο εικόνες που αποτελούνται από δύο επίπεδα χρώματος, ένα γκρι κι ένα μαύρο. Το γκρι επίπεδο στην αριστερή εικόνα είναι υψηλά δομημένο ενώ στην δεξιά όχι. Κάθε εικόνα περιέχει 50 εικονοστοιχεία στο γκρι επίπεδο και 250 στο μαύρο. Οι δύο εικόνες είναι μη διακριτές με βάση τα ιστρογράμματα χρώματος 2 κορυφών, καθώς αυτά είναι πανομοιότυπα. Οι περιγραφείς δομής χρώματος τους όμως είναι πολύ διαφορετικοί κι έτσι είναι διακριτές π.χ. σε μία εφαρμογή ανάκτησης ή δεικτοδότησης.



Σχήμα 2.3: Εικόνες 2 χρωματικών επιπέδων με διαφορετική δομή χρώματος

Ο περιγραφέας δομής χρώματος είναι ένα ιστόγραμμα που ορίζεται από την σχέση (2.11).

$$CSD = [h_1, h_1, \dots, h_N] \quad (2.11)$$

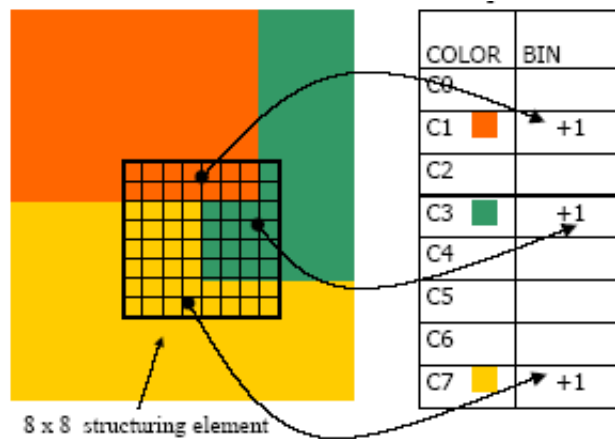
Όπου N είναι το μέγεθος του ιστογράμματος, δηλαδή ο αριθμός των χρωμάτων. Το πρότυπο MPEG-7 δεν ορίζει ούτε στην περίπτωση αυτή με αυστηρότητα κάποια συνάρτηση υπολογισμού απόστασης για τον συγκεκριμένο περιγραφέα. Ωστόσο, κι εδώ εφαρμόζεται με επιτυχία η πολύ γνωστή L1 απόσταση (2.12).

$$D(CSD_1, CSD_2) = \sum_{i=1}^N |h_{1i} - h_{2i}| \quad (2.12)$$

Ο περιγραφέας υπολογίζεται με την σάρωση όλων (ή ενός υποσυνόλου) των θέσεων στην εικόνα, ανακτώντας τα χρώματα c_m όλων των εικονοστοιχείων που περιλαμβάνονται στο δομικό στοιχείο το οποίο τοποθετείται σε κάθε θέση, και αυξάνοντας τις κορυφές $h(m)$ που αντιστοιχούν στο χρώμα c_m .

Για παράδειγμα, έστω ότι υπάρχουν 8 διαφορετικά χρώματα όπως στο σχήμα 2.4. Το δομικό στοιχείο είναι τετραγωνικό και έχει το μέγεθος 8×8 εικονοστοιχεία. Σε μια ορισμένη θέση της εικόνας το δομικό στοιχείο περιέχει μερικά εικονοστοιχεία με το χρώμα c_1 , μερικά με το χρώμα c_3 , και μερικά με το χρώμα c_6 τότε η κορυφή c_1 , η κορυφή c_3 και η κορυφή c_6 κάθε μία θα αυξάνονταν,

κατά ένα. Έτσι, σε αυτήν την θέση, ο περιγραφέας δομής χρώματος αυξάνεται κατά τρία συνολικά, κατά ένα για κάθε χρώμα παρόν στην περιοχή του δομικού στοιχείου.



Σχήμα 2.4: Υπολογισμός περιγραφέα δομής χρώματος

Τα παραπάνω που αφορούν την εξαγωγή του ιστογράμματος γίνονται σε εικόνα που αναπαρίσταται στον HMMD χρωματικό χώρο κβαντισμένο σε 256 χρώματα. Αν η εικόνα είναι σε άλλο χρωματικό χώρο τότε πρέπει να μετατραπεί σε αυτόν και στην συνέχεια να γίνει κβαντισμός πριν την εξαγωγή του ιστογράμματος. Ύστερα εάν λιγότερες από 256 κορυφές είναι επιθυμητές τότε οι κορυφές εννοποιούνται για να σχηματίσουν τον επιθυμητό αριθμό. Τέλος οι τιμές των κορυφών κβαντίζονται μη γραμμικά.

Στο πίνακα 2.3 φαίνεται η ευστοχία ανάκτησης για τα τέσσερα μήκη αναπαράστασης που έχουν οριστεί από το πρότυπο. Η μεγαλύτερη αναπαράσταση έχει και το καλύτερο αποτέλεσμα.

Μέγεθος περιγραφέα	ANMRR
256 bins	0.06799
128 bins	0.07613
64 bins	0.09374
32 bins	0.14438

Πίνακας 2.3: Αποτελέσματα για τον περιγραφέα δομής χρώματος στον HMMD χρωματικό χώρο

2.1.5 Περιγραφέας Διάταξης Χρώματος (Color Layout Descriptor)

Ο περιγραφέας διάταξης χρώματος είναι μία πολύ συμπαγής και σταθερή ως προς την ανάλυση αναπαράσταση χρώματος και μπορεί να χρησιμοποιηθεί για γρήγορη ανάκτηση εικόνων. Είναι σχεδιασμένος έτσι ώστε να αναπαριστά αποτελεσματικά τη χωρική κατανομή του χρώματος. Είναι

ιδιαίτερα χρήσιμος σε εφαρμογές ανάκτησης βασισμένη στη χωρική δομή, για παράδειγμα ανάκτηση βασισμένη σε σκίτσο (sketch-based retrieval) και αναγνώριση τμήματος βίντεο. Η ανάκτηση βασισμένη σε σκίτσο θεωρείται πολύ χρήσιμη, καθώς μπορεί να προσφέρει πολύ φιλικές διεπαφές προς τον χρήστη, ιδιαίτερα όταν η αναζήτηση είναι αρκετά γρήγορη. Εφαρμογές του περιγραφέα περιλαμβάνουν ταίριασμα από εικόνα σε εικόνα, από βίντεο σε βίντεο, από σκίτσο σε εικόνα και από σκίτσο σε βίντεο.

Ο περιγραφέας διάταξης χρώματος ορίζεται από την σχέση (2.13).

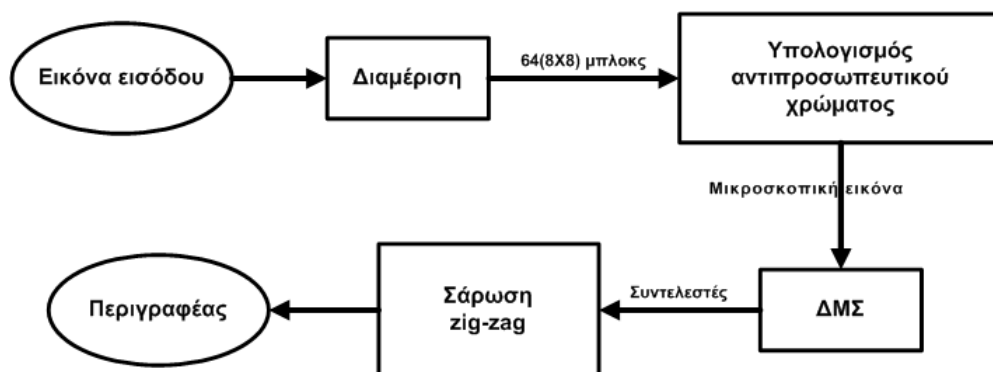
$$CLD = \left[\left\{ DY_{DC}, DY_{AC_i} \right\}, \left\{ DCr_{DC}, DCr_{AC_j} \right\}, \left\{ DCb_{DC}, DCb_{AC_k} \right\} \right] \quad (2.13)$$

Όπου τα i, j, k δηλώνουν τον αριθμό των AC συντελεστών και μπορούν να πάρουν τις τιμές 3, 6, 10, 15, 21, 28, 64. Για τη σύγκριση μεταξύ δύο περιγραφέων διάταξης χρώματος, το MPEG-7 προτείνει την συνάρτηση (2.14).

$$D(CLD_1, CLD_2) = \sqrt{\sum w_{yi}(DY_i^1 - DY_i^2)^2 + w_{rj}(DCr_j^1 - DCr_j^2)^2 + w_{bk}(DCb_k^1 - DCb_k^2)^2} \quad (2.14)$$

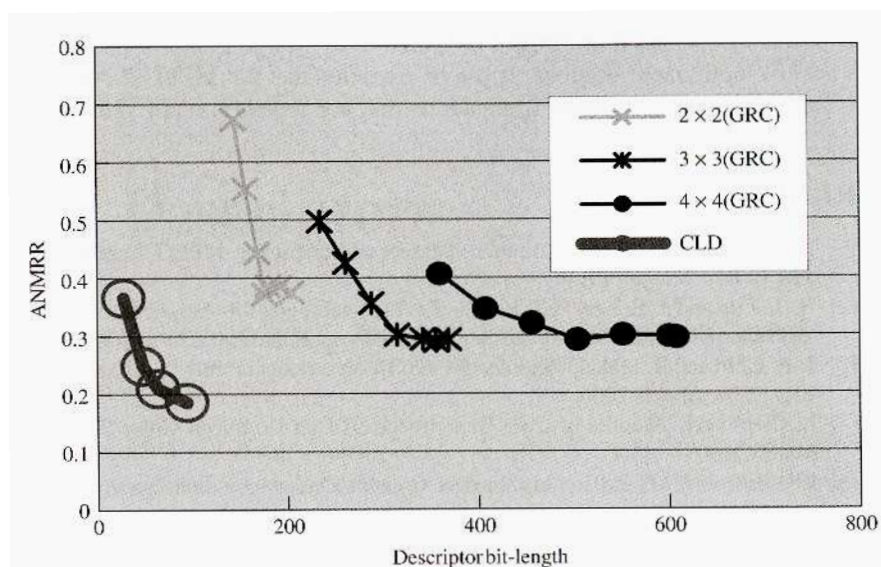
Όπου τα βάρη w_{yi}, w_{rj}, w_{bk} επιλέγονται από το χρήστη.

Στο πρώτο στάδιο της εξαγωγής η εικόνα εισόδου διαιρείται σε 64 μπλοκ για να είναι εγγυημένη η σταθερότητα ως προς την ανάλυση και την κλιμάκωση. Σε επόμενη φάση ένα αντιπροσωπευτικό χρώμα επιλέγεται για κάθε μπλοκ. Αυτό μπορεί να γίνει με οποιαδήποτε μέθοδο αλλά συνιστάται να χρησιμοποιείται η μέση τιμή των τιμών των εικονοστοιχείων σαν η αντιπροσωπευτική τιμή για κάθε μπλοκ, αφού είναι απλή αλλά και επαρκής μέθοδος. Το αποτέλεσμα είναι μία μικροσκοπική εικόνα μεγέθους 8×8 . Στο τρίτο στάδιο η κάθε μία από τις τρεις χρωματικές συνιστώσες μετατρέπεται με έναν 8×8 διακριτό μετασχηματισμό συνημιτόνου (ΔΜΣ- Discrete Cosine Transform - DCT) και έτσι προκύπτουν 3 σύνολα από 64 συντελεστές το κάθε ένα. Σαρώνονται με οδοντωτή τροχιά (zigzag scan) και οι πρώτοι συντελεστές χβαντίζονται μη ομοιόμορφα. Συνιστάται η χρήση 12 συντελεστών συνολικά, 6 για φωτεινότητα και 3 για κάθε χρωματικό κανάλι. Η παραπάνω διαδικασία φαίνεται στο σχήμα 2.5.



Σχήμα 2.5: Διαδικασία εξαγωγής του περιγραφέα διάταξης χρώματος

Στο σχήμα 2.6 φαίνεται η αποτελεσματικότητα του περιγραφέα σε εφαρμογή ανάκτησης εικόνας παρά το συμπαγές μέγεθος του. Η αποτελεσματικότητα του συγκρίνεται με εκείνη μίας παραδοσιακής τεχνικής βασισμένη σε πλέγμα όπου η εικόνα διαμερίζεται και αντιπροσωπευτικά χρώματα για κάθε διαμέρισμα χρησιμοποιούνται για να αναπαραστήσουν την διάταξη της. Διαφαίνεται πως ο περιγραφέας διάταξης χρώματος επιτυγχάνει καλύτερα αποτελέσματα.



Σχήμα 2.6: Αποτελέσματα περιγραφέα διάταξης χρώματος σε σύγκριση με το αντιπροσωπευτικό χρώμα βασισμένο σε πλέγμα

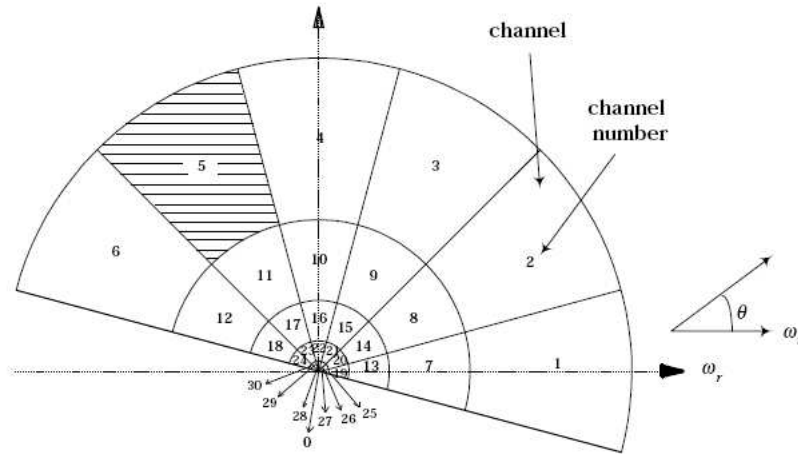
2.2 Περιγραφείς Υφής

2.2.1 Περιγραφέας Ομοιογενούς Υφής (Homogeneous Texture Descriptor)

Ο Περιγραφέας Ομοιογενούς Υφής παρέχει έναν ποσοτικό χαρακτηρισμό των περιοχών ομοιογενούς υφής για ανάκτηση ομοιότητας με βάση την υφή. Βασίζεται στον υπολογισμό των τοπικών χωρικών και συχνοτικών στατιστικών της υφής. Η εικόνα αρχικά φιλτράρεται με μια σειρά ευαίσθητων φίλτρων προσανατολισμού και κλίμακας. Υπολογίζεται η μέση και σταθερή απόκλιση των φιλτραρισμένων αποτελεσμάτων στον χώρο της συχνότητας. Προηγούμενη εκτενής εργασία για αυτόν τον περιγραφέα έχει δείξει ότι είναι αποτελεσματικός, και εύκολο να υπολογιστεί. Κατά τη διάρκεια των MPEG-7 πειραμάτων, συνειδητοποιήθηκε ότι η υπολογιστική πολυπλοκότητα αυτού του περιγραφέα μπορεί να μειωθεί σημαντικά με τον υπολογισμό των τιμών στην περιοχή συχνότητας παρά στη χωρική περιοχή.

Ο περιγραφέας ομοιογενούς υφής ορίζεται από την σχέση (2.15).

$$HTD = [f_{DC}, f_{SD}, e_1, e_2, \dots, e_N, d_1, d_2, \dots, d_N] \quad (2.15)$$



Σχήμα 2.7: Διάταξη συχνότητας για την εξαγωγή του περιγραφέα ομοιογενούς υφής

Όπου f_{DC} είναι η μέση και f_{SD} η τυπική απόκλιση των τιμών της εικόνας, N ο αριθμός των καναλιών στα οποία χωρίζει ο περιγραφέας τον χώρο συχνότητων και e_i και d_i η μέση ενέργεια και η απόκλιση ενέργειας του καθενός από αυτά. Η απόσταση μεταξύ δυο περιγραφέων ομοιογενούς υφής δίνεται από την σχέση (2.16).

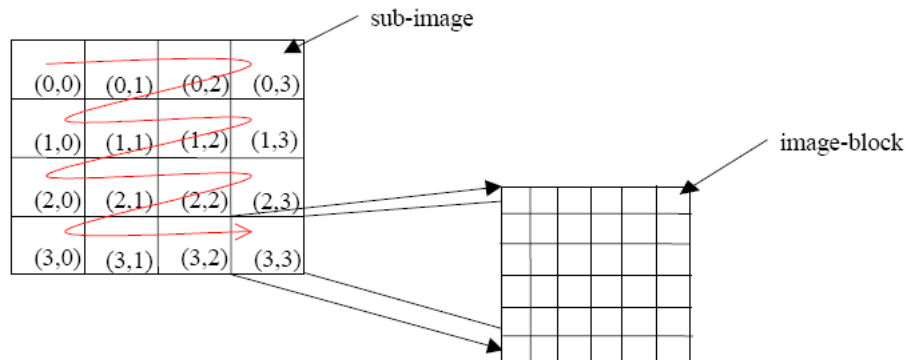
$$D(HTD_1, HTD_2) = \sum_{k=1}^N \left| \frac{HTD_2(k) - HTD_1(k)}{d_{norm}} \right| \quad (2.16)$$

Στη σχέση (2.16) ο παράγοντας κανονικοποίησης d_{norm} μπορεί να επιλεγθεί ελεύθερα από το χρήστη. Το πρότυπο MPEG-7 προτείνει την χρήση της τυπικής απόκλισης της βάσης που περιλαμβάνει τους περιγραφείς ομοιογενούς υφής.

Ο περιγραφέας ομοιογενούς υφής χαρακτηρίζει την υφή της εικόνας/περιοχής χρησιμοποιώντας την μέση ενέργεια και την απόκλιση ενέργειας από ένα σύνολο καναλιών συχνότητας. Το 2-Δ επίπεδο συχνότητας διαμερίζεται σε 30 κανάλια όπως φαίνεται στο σχήμα 2.7. Η μέση ενέργεια και η απόκλιση της υπολογίζονται σε κάθε ένα από αυτά τα 30 κανάλια συχνότητας (στον χώρο της συχνότητας). Πρέπει να σημειωθεί ότι η διαμέριση του επιπέδου συχνότητας είναι ομοιόμορφη κατά την γωνιακή κατεύθυνση (με βήμα μεγέθους 30°) αλλά μη ομοιόμορφη κατά την ακτινική κατεύθυνση. Πιο συγκεκριμένα η διαίρεση κατά την ακτινική κατεύθυνση είναι με κλίμακα οκτάβας. Το κίνητρο για μία τέτοια διαμέριση πηγάζει από προηγούμενη δουλειά [10]. Υπάρχουν επίσης στοιχεία που δηλώνουν πως η πρώιμη οπτική ανάλυση στον οπτικό φλοιό του εγκεφάλου μπορεί να μοντελοποιηθεί καλά χρησιμοποιώντας μία διάταξη συχνότητας παρόμοια με του σχήματος 2.7.

2.2.2 Περιγραφέας Ιστογράμματος Αχμών (Edge Histogram Descriptor)

Ο περιγραφέας ιστογράμματος αχμών καταγράφει τη χωρική κατανομή των αχμών. Η κατανομή των αχμών είναι μια καλή αναπαράσταση υφής, είναι χρήσιμη για το ταίριασμα από εικόνα σε εικόνα



Σχήμα 2.8: Ορισμός υποεικόνας και μπλοκ

όταν η υποκειμένη περιοχή δεν είναι ομοιογενής ως προς τις ιδιότητες της υφής. Εφόσον ο περιγραφέας βασίζεται στις πληροφορίες ακμών στην εικόνα, είναι καλός για ταίριασμα φυσικής εικόνας. Δεδομένου ότι οι ακμές διαδραματίζουν έναν σημαντικό ρόλο για την αντίληψη εικόνας, μπορεί να χρησιμεύσει στην ανάκτηση εικόνων με παρόμοιες σημασιολογικές έννοιες. Κατά συνέπεια, στοχεύει πρώτιστα στο ταίριασμα από εικόνα σε εικόνα (από παράδειγμα ή από σκίτσο), ειδικά για τις φυσικές εικόνες με ανομοιόμορφη κατανομή ακμών. Σε αυτό το πλαίσιο, η απόδοση ανάκτησης εικόνας μπορεί να βελτιωθεί σημαντικά, εάν ο περιγραφέας ιστογραμμάτων ακμών συνδυάζεται με άλλους περιγραφείς.

Ο περιγραφέας ιστογράμματος ακμών ορίζεται από την σχέση (2.17)

$$EHD = [e_1, e_2, \dots, e_{80}] \quad (2.17)$$

Ο συγκεκριμένος περιγραφέας έχει τη μορφή ιστογράμματος. Το πρότυπο MPEG-7 δεν προτείνει κάποια συγκεκριμένη συνάρτηση για τον υπολογισμό της απόστασης. Ωστόσο υπονοείται η χρήση της L1 συνάρτησης που χρησιμοποιείται ευρέως για ομοιότητα ιστογραμμάτων (2.18).

$$D(EHD_1, EHD_2) = \sum_{i=1}^{80} |e_{1i} - e_{2i}| \quad (2.18)$$

Για την εξαγωγή του απαιτείται ο εντοπισμός μη κατευθυντικών ακμών όπως επίσης και τεσσάρων ειδών κατευθυντικών ακμών (οριζόντιες, κάθετες, διαγώνιες με προσανατολισμό 45° , διαγώνιες με προσανατολισμό 135°). Η εικόνα διαιρείται σε 4×4 υποεικόνες κι έπειτα η κάθε μία από αυτές διαιρείται επιπλέον σε μη επικαλυπτόμενα τετράγωνα μπλοκ (σχήμα 2.8). Το μέγεθος του μπλοκ εξαρτάται από την ανάλυση της εικόνας. Ανεξαρτήτως του μεγέθους της εικόνας, όλη η εικόνα διαιρείται σε έναν προκαθορισμένο συνολικό αριθμό από μπλοκ. Πειράματα έδειξαν πως ένας αριθμός μπλοκ γύρω στο 1100 φαίνεται να συλλαμβάνει καλά τα χαρακτηριστικά της κατεύθυνσης των ακμών. Στην συνέχεια κάθε ένα από τα μπλοκ ταξινομείται σε μία από τις πέντε κατηγορίες ακμών ή σαν μπλοκ χωρίς ακμές, πράγμα το οποίο γίνεται θεωρώντας το κάθε μπλοκ σαν εικόνα 2×2 και εφαρμόζοντας ανιχνευτές ακμών. Αφού γίνει η ταξινόμηση τότε υπολογίζονται τα ιστογράμματα ακμών με 5 κορυφές, μία για κάθε είδος ακμής, για τις 16 υποεικόνες (συνολικά $16 \times 5 = 80$ κορυφές). Έπειτα

κάθε τιμή κανονικοποιείται ως προς το συνολικό αριθμό των μπλοκ στην υποεικόνα κι εφόσον υπάρχουν και μπλοκ χωρίς ακμές το άθροισμα των πέντε τιμών για την κάθε υποεικόνα είναι μικρότερο ή ίσο του 1. Έτσι μπορούμε να πούμε ότι ο περιγραφέας ιστογραμμάτων ακμών συνυπολογίζει και την πληροφορία για την κατανομή περιοχών χωρίς ακμές.



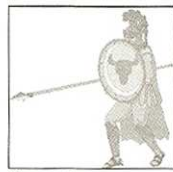
Σχήμα 2.9: Ανάκτηση για φυσικές εικόνες με τον περιγραφέα ιστογράμματος ακμών

Αποτελέσματα πειραμάτων έδειξαν πως ο περιγραφέας ιστογραμμάτων ακμών είναι αρκετά χρήσιμος στην ανάκτηση εικόνας, ιδιαίτερα σε εικόνες με μη ομοιόμορφη υφή και για "clip art" εικόνες. Τα σχήματα 2.9 και 2.10 δείχνουν παραδείγματα αποτελεσμάτων ανάκτησης για φυσικές εικόνες και για "clip art" εικόνες αντίστοιχα. Αυτός ο περιγραφέας μπορεί επίσης να χρησιμοποιηθεί για ανάκτηση εικόνας με βάση σχίτσο.

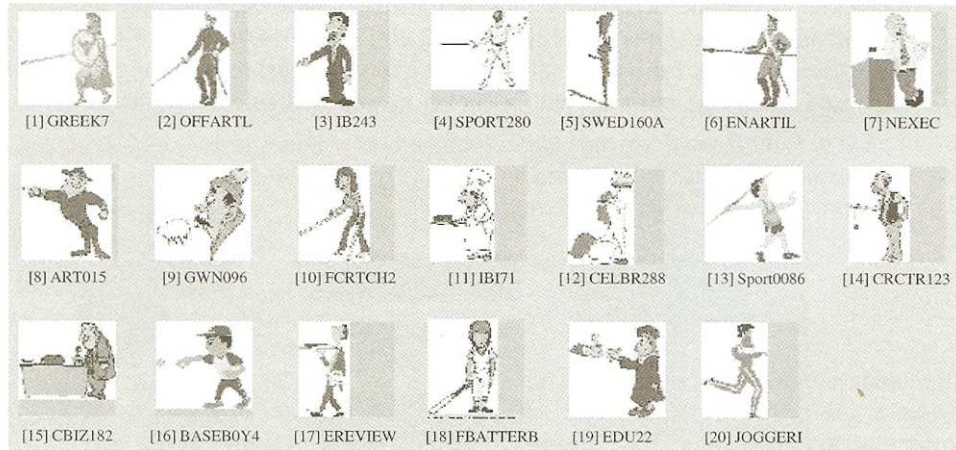
2.3 Περιγραφείς Σχήματος

2.3.1 Περιγραφέας Σχήματος με βάση το Σχήμα (Region-Based Shape Descriptor)

Ο περιγραφέας σχήματος με βάση το σχήμα εκφράζει τη κατανομή εικονοστοιχείων μέσα σε μια 2-Δ περιοχή αντικειμένου, μπορεί να περιγράψει σύνθετα αντικείμενα που αποτελούνται από πολλαπλάσιες ασύνδετες περιοχές καθώς επίσης και απλά αντικείμενα με ή χωρίς τρύπες (σχήμα 2.11). Συνυπολογίζει όλα τα εικονοστοιχεία που αποτελούν το σχήμα, δηλαδή και το περίγραμμα και τα εσωτερικά εικονοστοιχεία. Δεν επηρεάζεται από περιστροφή και κλιμάκωση. Μερικά σημαντικά



Query Image



Σχήμα 2.10: Ανάκτηση για “clip art” εικόνες με τον περιγραφέα ιστογράμματος ακμών

χαρακτηριστικά γνωρίσματα αυτού του περιγραφέα είναι τα ακόλουθα.

1. Δίνει έναν συμπαγή και αποδοτικό τρόπο περιγραφής των ιδιοτήτων πολλαπλών αποσυνδεδεμένων περιοχών ταυτόχρονα.
2. Μερικές φορές κατά τη διάρκεια της διαδικασίας της κατάτμησης, ένα αντικείμενο μπορεί να χωριστεί σε αποσυνδεδεμένες υποπεριοχές. Ένα τέτοιο αντικείμενο μπορεί ακόμα να ανακτηθεί, υπό τον όρο ότι οι πληροφορίες στις οποίες οι περιοχές αυτές χωρίστηκαν διατηρούνται και χρησιμοποιούνται κατά τη διάρκεια της εξαγωγής του περιγραφέα.
3. Ο περιγραφέας είναι εύρωστος στο θόρυβο κατάτμησης, π.χ., στον κρουστικό θόρυβο (salt and



Σχήμα 2.11: Παραδείγματα σχημάτων για τα οποία ο περιγραφέας σχήματος με βάση το σχήμα είναι εφαρμόσιμος

pepper noise).

Ο περιγραφέας σχήματος με βάση το σχήμα ορίζεται από την σχέση (2.19).

$$RSD = [M_i], i = 1, 2, \dots, 35 \quad (2.19)$$

Όπου M_i είναι οι κανονικοποιημένοι και κβαντισμένοι ART συντελεστές του περιγραφέα. Για τον υπολογισμό της απόστασης μεταξύ δύο περιγραφένων σχήματος με βάση το σχήμα, χρησιμοποιείται η σχέση (2.20) (εφαρμογή της L1 νόρμας).

$$D(RSD_1, RSD_2) = \sum_{i=1}^{35} \| M_{1i} - M_{2i} \| \quad (2.20)$$

Ο περιγραφέας δουλεύει αποσυνθέτοντας το σχήμα σε έναν αριθμό ορθογώνιων 2-Δ συναρτήσεων βάσης (μιγαδικές) ορισμένες από την Γωνιακή Ακτινική Μετατροπή (Angular Radial Transform - ART). Η τεχνική αυτή βασίζεται σε μια ορθογώνια μοναδιαία μετατροπή σε έναν μοναδιαίο δίσκο αποτελούμενο από τις πλήρεις ορθοκανονικές συναρτήσεις βάσης σε πολικές συντεταγμένες. Οι συντελεστές ART καθορίζονται από στην σχέση (2.21). F_{nm} είναι ένας συντελεστής ART τάξεως n και m , $f(\rho, \theta)$ είναι η εικόνα σε πολικές συντεταγμένες και $V_{nm}(\rho, \theta)$ είναι η συνάρτηση βάσης που είναι διαχωρίσιμη κατά μήκος των γωνιακών και ακτινικών κατευθύνσεων. Ο περιγραφέας αυτός αποτελείται από ένα σύνολο κανονικοποιημένων μετρών των μιγαδικών συντελεστών ART. Η σταθερότητα ως προς την περιστροφή αποκτάται με την χρήση του μέτρου των συντελεστών. Χρησιμοποιούνται 12 γωνιακές και 3 ακτινικές συναρτήσεις. Για κανονικοποίηση οι συντελεστές διαιρούνται με το μέτρο του συντελεστή ART τάξεως $n=0$ και $m=0$. Τέλος οι συντελεστές κβαντίζονται σε 4 bits ο καθένας.

$$F_{nm} = \langle V_{nm}(\rho, \theta), f(\rho, \theta) \rangle = \int_0^{2\pi} \int_0^1 V_{nm}^*(\rho, \theta) f(\rho, \theta) \rho d\rho d\theta \quad (2.21)$$

Στον πίνακα 2.4 φαίνονται αποτελέσματα πειραμάτων με τον περιγραφέα σχήματος βασισμένο στο σχήμα. Ικανοποιητικά αποτελέσματα προέκυψαν σε πειράματα για την επιρροή από την περιστροφή και την κλιμάκωση ενώ λίγο μειώθηκε η αποτελεσματικότητα σε πειράματα επιρροής από την προοπτική. Επίσης παράδειγμα αποτελεσμάτων ανάκτησης ομοιότητας φαίνονται στο σχήμα 2.12.

Κλιμάκωση	Περιστροφή	Προοπτική
86.20%	98.06%	62.25%

Πίνακας 2.4: Αποτελέσματα πειραμάτων για τον περιγραφέα σχήματος με βάση το σχήμα στην βάση δεδομένων με εμπορικά σχήματα

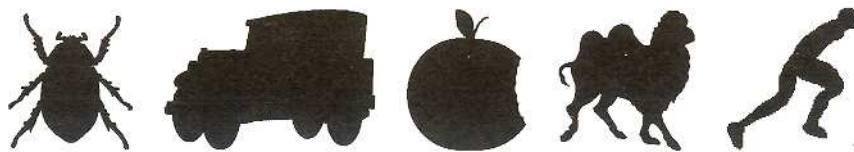
2.3.2 Περιγραφέας Σχήματος με βάση το Περίγραμμα (Contour-Based Shape Descriptor)

Ο περιγραφέας σχήματος με βάση το περίγραμμα εκφράζει τις ιδιότητες σχήματος του περιγράμματος του αντικειμένου. Τα αντικείμενα για τα οποία χαρακτηριστικά γνωρίσματα σχήματος



Σχήμα 2.12: Παράδειγμα ανάκτησης ομοιότητας για τον περιγραφέα σχήματος με βάση το σχήμα στην βάση δεδομένων με εμπορικά σχήματα

περιλαμβάνονται στο περίγραμμα περιγράφονται αποτελεσματικά από αυτόν τον περιγραφέα. Εάν ένα σύνθετο αντικείμενο αποτελείται από πολλαπλές χωριστές περιοχές, κάθε περιοχή των συστατικών περιγραμμάτων μπορεί να περιγραφεί χωριστά, χρησιμοποιώντας τον περιγραφέα σχήματος με βάση το περίγραμμα και ένα MPEG-7 σχήμα περιγραφής. Έχει διάφορες ενδιαφέρουσες ιδιότητες όπως: ευθυγράμμιση με την ανθρώπινη αντίληψη για την ομοιότητα μορφής, ευρωστία στις σημαντικές εύκαμπτες παραμορφώσεις και υποστήριξη για το ταίριασμα κάτω από τους μετασχηματισμούς προοπτικής.

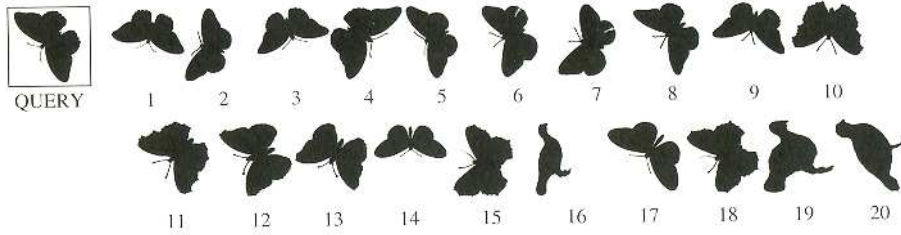


Σχήμα 2.13: Παραδείγματα σχημάτων για τα οποία ο περιγραφέας σχήματος με βάση το περίγραμμα είναι εφαρμόσιμος

Ο περιγραφέας σχήματος με βάση το περίγραμμα ορίζεται από την σχέση (2.22).

$$CBSD = [NoP, C, PrC, HPY, pX[], pY[]] \quad (2.22)$$

Ο περιγραφέας αυτός βασίζεται στην CSS (Curvature Scale Space) αναπαράσταση του περιγράμματος ενός σχήματος [11]. Η CSS αναπαράσταση “αποσυνθέτει” το περίγραμμα σε κοίλα και κυρτά μέρη ορίζοντας τα σημεία καμπής (μηδενικής καμπυλότητας). Γίνεται επαναληπτικά μία εξομάλυνση του περιγράμματος μέχρι να προκύψει σαν αποτέλεσμα ένα κυρτό περίγραμμα. Η CSS εικόνα δείχνει πως τα σημεία καμπής αλλάζουν όσο το περίγραμμα εξομαλύνεται και τείνει να γίνει κυρτό. Ο x -CSS-άξονας αντιστοιχεί στην θέση πάνω στο περίγραμμα (ωρολογιακά αρχίζοντας από ένα τυχαίο σημείο) και ο y -CSS-άξονας αντιστοιχεί στις τιμές μίας παραμέτρου εξομάλυνσης (όταν οι τιμές y αυξάνουν τότε και η εξομάλυνση αυξάνει). Ο περιγραφέας αποτελείται από την κυκλικότητα και την εκκεντρότητα του αρχικού αλλά και του εξομαλυμένου περιγράμματος, έναν δείκτη που δηλώνει τον αριθμό των κορυφών στην CSS εικόνα, το ύψος της υψηλότερης κορυφής και τα x και y κορυφών που μένουν.



Σχήμα 2.14: Παράδειγμα ανάκτησης με βάση την ομοιότητα για τον περιγραφέα σχήματος με βάση το περίγραμμα

Στην σχέση (2.22), το NoP δηλώνει τον αριθμό των κορυφών της CSS αναπαράστασης, το C περιλαμβάνει την κυκλικότητα (circularity) και την εκκεντρότητα (eccentricity) του περιγράμματος, το PrC περιλαμβάνει την κυκλικότητα και την εκκεντρότητα του εξομαλυμένου περιγράμματος, HPY είναι η απόλυτη τιμή του ύψους της υψηλότερης κορυφής, $pX[]$ είναι ένας πίνακας που περιλαμβάνει τις X -τιμές των θέσεων στο περίγραμμα ενός σχήματος και $pY[]$ τα ύψη των αντίστοιχων κορυφών. Για την απόσταση μεταξύ δύο περιγραφέντων σχήματος με βάση το περίγραμμα, το MPEG-7 χρησιμοποιεί την συνάρτηση (2.23).

$$D(CBSD_1, DCBSD_2) = 0.4 \frac{|C_1[0] - C_2[0]|}{\max\{C_1[0], C_2[0]\}} + 0.3 \frac{|C_1[1] - C_2[1]|}{\max\{C_1[1], C_2[1]\}} + M_{CSS} \quad (2.23)$$

Στη σχέση αυτή $C_1[0]$ και $C_2[0]$ είναι οι εκκεντρότητες των δύο περιγραφέντων και $C_1[1]$ και $C_2[1]$ είναι κυκλικότητες. Η ποσότητα M_{CSS} είναι μια L2 απόσταση ανάμεσα στις κορυφές, η οποία “τιμωρεί” τις κορυφές που δεν ταιριάζουν και δίνεται από την σχέση (2.24). Όπου στη σχέση αυτή, η άθροιση \sum_1 πραγματοποιείται για όλες τις κορυφές που ταιριάζουν, ενώ η \sum_2 άθροιση για όλες τις κορυφές που δεν ταιριάζουν. Κριτήριο για το αν δύο κορυφές ταιριάζουν ή όχι είναι η L2 απόσταση μεταξύ τους, η οποία δεν πρέπει να υπερβαίνει την τιμή 0.1.

$$M_{CSS} = \sum_1 ((pX[i] - pX[j])^2 + (pY[i] - pY[j])^2) + \sum_2 (pY[i])^2 \quad (2.24)$$

Στον πίνακα 2.5 φαίνονται αποτελέσματα πειραμάτων με τον περιγραφέα σχήματος με βάση το περίγραμμα όσον αφορά την σταθερότητα σε περιστροφή και κλιμάκωση αλλά και ανάκτηση ομοιότητας. Ένα παράδειγμα ανάκτησης με βάση την ομοιότητα δίνεται στο σχήμα 2.14.

Κλιμάκωση	Περιστροφή	Ομοιότητα
91.03%	100%	79.15%

Πίνακας 2.5: Αποτελέσματα πειραμάτων με τον περιγραφέα σχήματος με βάση το περίγραμμα

Μέρος II

Εφαρμογή Εξαγωγής Περιγραφών

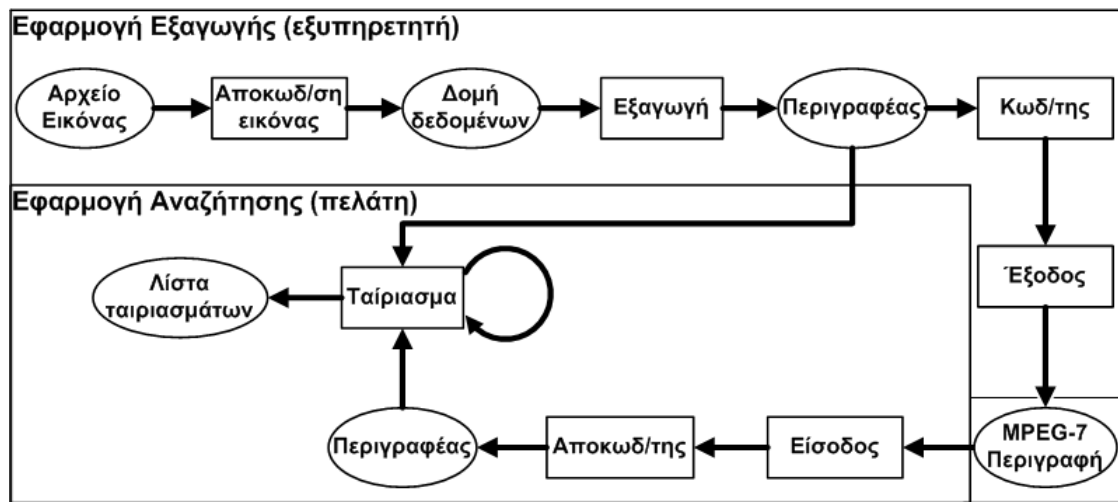
Κεφάλαιο 3

Εισαγωγή στο MPEG-7 eXperimentation Model

Το MPEG-7 eXperimentation Model (XM) [12] είναι ένα πλαίσιο εργασίας για όλο τον κώδικα αναφοράς για το πρότυπο MPEG-7. Ένας από τους κύριους σκοπούς του ήταν να παρέχει ένα προσχέδιο του ίδιου του προτύπου κατασκευάζοντας το καλύτερο δυνατό μοντέλο και να επιβεβαιώσει την λειτουργικότητα των βασικών συνιστωσών του. Υλοποιεί τις παρακάτω βασικές συνιστώσες του MPEG-7:

- Περιγραφείς (Descriptors)
- Σχήματα Περιγραφής (Description Schemes)
- Σχήματα κωδικοποίησης (Coding Schemes)
- Γλώσσα Καθορισμού Περιγραφής (Description Definition Language)
- Συνιστώσες συστήματος

Πέρα από την υλοποίηση των παραπάνω συνιστωσών, το MPEG-7 περιλαμβάνει και επιπλέον βιβλιοθήκες. Πιο συγκεκριμένα, για την ανάγνωση εικόνων/βίντεο χρησιμοποιείται η *ImageMagick* και για την εκτέλεση διαδικαστικού κώδικα πάνω στις βασικές δομές δεδομένων, όπως την δομή των εικόνων. Για την ανάγνωση ήχου χρησιμοποιείται η βιβλιοθήκη *Afsp* και για την έξοδο σε XML η *Xerces*. Οι δομές δεδομένων μαζί με τον διαδικαστικό κώδικα σχηματίζουν μία εφαρμογή. Για κάθε περιγραφέα και κάθε σχήμα περιγραφής υπάρχει τουλάχιστον μία εφαρμογή μέσα στο XM, με σκοπό να επιτρέψει την εξακρίβωση της καλής λειτουργικότητας των βασικών συνιστωσών. Οι εφαρμογές επιδεικνύουν επίσης τον τρόπο εξαγωγής των μεταδεδομένων από το περιεχόμενο, και την χρήση τους σε απλές εφαρμογές. Συνεπώς το XM υλοποιεί μόνο βασικούς και στοιχειώδεις τύπους εφαρμογών και όχι πραγματικές εφαρμογές. Επιπλέον η μοναδική επαφή του XM είναι μέσω γραμμής εντολών. Έτσι δεν είναι δυνατή οποιαδήποτε αλληλεπίδραση κατά την διάρκεια της εκτέλεσής της. Οι ενότητες του XM έχουν σχεδιαστεί με τέτοιο τρόπο ώστε όλες να χρησιμοποιούν συγκεκριμένες διεπαφές.



Σχήμα 3.1: Εφαρμογές εξυπηρετητή και πελάτη στο XM

Αυτό επιτρέπει την εύκολη περιήγηση μέσα σε όλες τις διαφορετικές ενότητες για τους διάφορους περιγραφείς και σχήματα περιγραφής μετά από κατανόηση της δομής μία φορά. Αφ' ετέρου, η χρήση των σταθερών διεπαφών επιτρέπει την επαναχρησιμοποίηση και τον συνδυασμό των μεμονωμένων εννοιών σε μεγαλύτερη εφαρμογή.

Στον σκελετό του XM οι εφαρμογές σχετίζονται με έναν συγκεκριμένο περιγραφέα ή σχήμα περιγραφής. Επειδή υπάρχουν πολλοί περιγραφείς και σχήματα περιγραφής προτυποποιημένα, υπάρχουν επίσης πολλές εφαρμογές ενσωματωμένες στον σκελετό. Οι εφαρμογές οι οποίες δημιουργούν τον περιγραφέα ή το σχήμα περιγραφής το οποίο δοκιμάζουν, ονομάζονται εφαρμογές εξαγωγής (extraction) ή εφαρμογές εξυπηρετητή (server). Ενώ εφαρμογές που χρησιμοποιούν τους περιγραφείς ή τα σχήματα περιγραφής κάτω από δοκιμασίες ελέγχου ονομάζονται εφαρμογές αναζήτησης (search) ή εφαρμογές πελάτη (client). Στο σχήμα 3.1 περιγράφονται και διαχωρίζονται αυτά τα δύο είδη εφαρμογών. Οι εφαρμογές εξυπηρετητή χρειάζονται εάν ο περιγραφέας ή το σχήμα περιγραφής είναι χαμηλού επιπέδου. Για υψηλού επιπέδου η αυτόματη εξαγωγή δεν είναι εφικτή.

Στο XM σαν είσοδο για την εξαγωγή μπορεί να δοθεί μία εικόνα σε ένα από τα format που υποστηρίζει η βιβλιοθήκη ImageMagick την οποία και χρησιμοποιεί για να διαβάσει εικόνες. Έτσι η εξαγωγή μπορεί να γίνει είτε σε όλη την εικόνα, είτε σε μία περιοχή ενδιαφέροντος εάν μαζί με την εικόνα δοθεί και μία δυαδική μάσκα (binary mask) η οποία και θα δηλώνει την επιθυμητή αυτή περιοχή.

Κεφάλαιο 4

Εφαρμογή Εξαγωγής Περιγραφών: VDE

4.1 Υλοποίηση VDE

Η εφαρμογή Visual Descriptor Extraction (VDE) δημιουργήθηκε με βάση το XM με σκοπό την παροχή νέων δυνατοτήτων όσον αφορά την εξαγωγή περιγραφών χαμηλού επιπέδου, τροποποιώντας τις εφαρμογές εξυπηρετητή που αναφέρθηκαν στο κεφάλαιο 2. Αυτές οι νέες δυνατότητες αφορούν κυρίως την εξαγωγή ενός περιγραφέα από περιοχή ή περιοχές ενδιαφέροντος αλλά και μία πιο ευέλικτη και φιλική προς τον χρήστη διεπαφή γραμμής εντολών. Το VDE περιέχει εφαρμογές εξαγωγής για τους παρακάτω περιγραφείς:

- Περιγραφέας Κύριων Χρωμάτων
- Κλιμακωτός Περιγραφέας Χρώματος
- Περιγραφέας Δομής Χρώματος
- Περιγραφέας Διάταξης Χρώματος
- Περιγραφέας Ομοιογενούς Υφής
- Περιγραφέας Ιστογράμματος Ακμών
- Περιγραφέας Σχήματος με βάση το Σχήμα
- Περιγραφέας Σχήματος με βάση το Περίγραμμα

Πιο συγκεκριμένα, στις εφαρμογές εξυπηρετητή που συμπεριλαμβάνονται στο VDE δόθηκε η δυνατότητα να γίνεται εξαγωγή των περιγραφών για όλες τις περιοχές ενδιαφέροντος που δηλώνονται σε έναν χάρτη περιοχών (region map) και τα μεταδεδομένα που παράγονται να δίνονται σε ένα αρχείο XML. Ο χάρτης περιοχών είναι μία εικόνα στις ίδιες διαστάσεις με την εικόνα εισόδου, τα εικονοστοιχεία της οποίας έχουν την τιμή της περιοχής στην οποία ανήκουν, από 0 μέχρι τον αριθμό των συνολικών περιοχών μειωμένο κατά ένα. Επίσης, στην εφαρμογή VDE ως περιοχές ενδιαφέροντος

μπορούν να οριστούν και ορθογώνιες (ενδεχομένως αλληλοεπικαλυπτόμενες) περιοχές, οι οποίες δηλώνονται από τις συντεταγμένες τους μέσα στην εικόνα εισόδου. Πέρα από τις παραπάνω βασικές αλλαγές έγιναν και κάποιες που αφορούν την διεπαφή της γραμμής εντολών. Στην εφαρμογή VDE μπορεί πλέον να δηλωθεί το όνομα ενός αρχείου εικόνας για είσοδο σε αντίθεση με τον περιορισμό του XM, στο οποίο απαιτείται η ύπαρξη αρχείου κειμένου με τα ονόματα αρχείων των εικόνων εισόδου, ακόμη και στην περίπτωση μοναδικής εικόνας εισόδου. Ένας άλλος περιορισμός του XM αφορά την προκαθορισμένη ονομασία του αρχείου δυαδικής μάσκας. Στην εφαρμογή VDE το αρχείο δυαδικής μάσκας μπορεί προαιρετικά να δηλωθεί σε μία νέα παράμετρο με τον ίδιο τρόπο που δηλώνονται τα ονόματα των αρχείων που ορίζουν τα άλλα δύο είδη περιοχών ενδιαφέροντος (χάρτης περιοχών και ορθογώνιες περιοχές). Τέλος έχει προστεθεί η δυνατότητα εξαγωγής περισσοτέρων του ενός περιγραφών με μία κλήση της εφαρμογής.

Η εφαρμογή VDE αποτελείται από 5 projects της Visual C++, τα οποία είναι τα εξής:

- XMLibAP: αρχεία των εφαρμογών.
- XMLibB: αρχεία εισόδου εικόνας κ.α.
- XMLibC: αρχεία που δηλώνουν δομές, σχήματα κωδικοποίησης κ.α.
- XMLibV: αρχεία με αλγόριθμους εξαγωγής και διαδικασίες δημιουργίας εξόδου σε XML.
- VDE: περιλαμβάνει όλα τα παραπάνω projects και την main κλάση.

4.2 Εξαγωγή από Πολλαπλές Περιοχές Ενδιαφέροντος

Σε αυτή την ενότητα θα παρουσιαστούν συνοπτικά οι διαδικασίες και οι δομές τις οποίες χρησιμοποιεί το XM για να κάνει την εξαγωγή των περιγραφών και ο τρόπος που αυτές χρησιμοποιήθηκαν για την σχεδίαση και την παροχή νέων λειτουργιών στην εφαρμογή VDE. Η λειτουργία για την οποία αξίζει να γίνει αυτή η αναφορά για τον εσωτερικό τρόπο λειτουργίας της εφαρμογής είναι εκείνη που αναφέρθηκε στο κεφάλαιο 3 σαν βασική αλλαγή και αφορά τη δυνατότητα να δηλώνονται πολλαπλές περιοχές ενδιαφέροντος.

Στο XM σαν είσοδο για να εξαχθούν οι περιγραφείς δίνονται εικόνες και προαιρετικά δυαδικές μάσκες που συνοδεύουν την κάθε εικόνα. Η ίδια διαδικασία επαναλαμβάνεται επαναληπτικά για κάθε εικόνα. Όσον αφορά την περίπτωση που υπάρχει και συνοδευτική δυαδική μάσκα για κάθε εικόνα τότε αρχικά η εικόνα διαβάζεται και φορτώνεται στην μνήμη και αναπαρίσταται από μία δομή την οποία χρησιμοποιεί το XM για τις εικόνες. Η δομή αυτή περιλαμβάνει τις τιμές των εικονοστοιχείων όλων των καναλιών της εικόνας, αλλά και περαιτέρω πληροφορίες για την εικόνα (πχ. διαστάσεις).

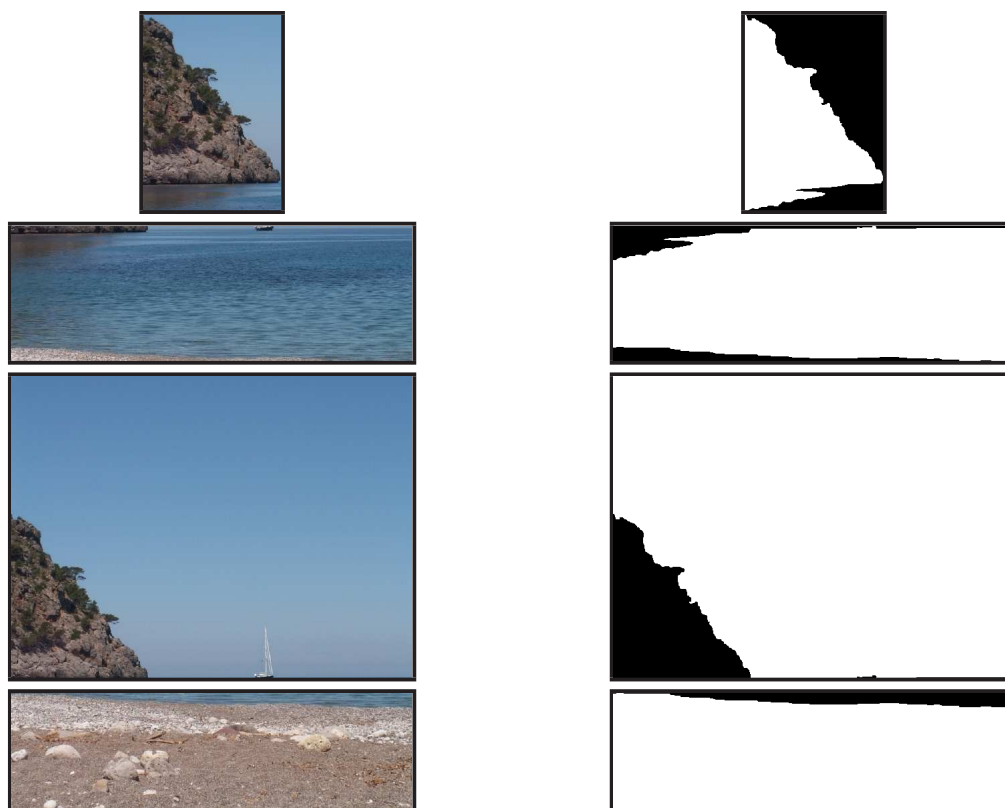
Ύστερα, διαβάζεται και φορτώνεται στην μνήμη και η δυαδική μάσκα που αντιστοιχεί στην εικόνα. Μία παρόμοια δομή με της εικόνας χρησιμοποιείται και για την δυαδική μάσκα που εδώ βέβαια έχει ένα μόνο κανάλι. Τα στοιχεία του έχουν τιμές 1 αν τα αντίστοιχα εικονοστοιχεία ανήκουν στην περιοχή ενδιαφέροντος και 0 αν όχι. Οι δύο δομές αυτές συνδέονται και συσχετίζονται κατά την συνέχεια της διαδικασίας, η οποία είναι η εξαγωγή του περιγραφέα και γίνεται για τα στοιχεία εκείνα τα οποία στην δομή της μάσκας έχουν τιμές 1. Πρέπει να αναφερθεί ότι η αναπαράσταση της εσωτερικής λειτουργίας της εφαρμογής γίνεται αφαιρετικά και όσο χρειάζεται για να εξηγηθεί απλά πώς κάποιες μονάδες της αρχικής εφαρμογής διατηρήθηκαν και αξιοποιήθηκαν από την εφαρμογή VDE.



Σχήμα 4.1: Εικόνα εισόδου και χάρτης περιοχών

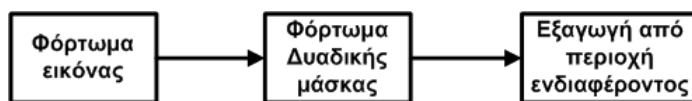
Στην περίπτωση της εφαρμογής VDE και πιο συγκεκριμένα όταν η εικόνα εισόδου συνοδεύεται από χάρτη περιοχών, αυτό που ήταν επιθυμητό ήταν να διατηρηθεί η μονάδα εξαγωγής (συναρτήσεις εξαγωγής) ως έχει. Άρα έπρεπε κάθε φορά να δίνεται σαν είσοδος σε αυτήν μία δομή εικόνας με μία συνδεδεμένη με αυτή δομή δυαδικής μάσκας, όπως περιγράφηκε νωρίτερα. Επίσης, ένα από τα ζητούμενα ήταν να μην χρησιμοποιείται κάθε φορά ολόκληρη η εικόνα κατά την εξαγωγή για κάθε περιοχή αλλά να δίνεται μόνο το απαραίτητο μέρος αυτής και αυτό για συντομία των υπολογισμών που διατρέχουν ολόκληρη την εικόνα. Έτσι με γνώμονα τα παραπάνω, στην περίπτωση χρήσης χάρτη περιοχών αρχικά διαβάζεται και φορτώνεται η εικόνα στην αντίστοιχη δομή. Έπειτα διαβάζεται η εικόνα που αντιστοιχεί στον χάρτη περιοχών και φορτώνεται σε μία δομή παρόμοια με εκείνη της δυαδικής μάσκας και συσχετίζεται με την δομή της αρχικής εικόνας. Η μέγιστη τιμή των στοιχείων του χάρτη είναι ίση με τον αριθμό των περιοχών ενδιαφέροντος μειωμένο κατά ένα (στον χάρτη οι περιοχές αριθμούνται ξεκινώντας από το μηδέν και αυξάνονται κατά ένα). Για κάθε περιοχή εκτελείται επαναληπτικά η ίδια διαδικασία. Υπολογίζεται πρώτα το περικλείον ορθογώνιο (bounding box) όλων των εικονοστοιχείων που ανήκουν σε μία περιοχή το οποίο αναπαρίσταται από τις συντεταγμένες του άνω αριστερά και κάτω δεξιά σημείου του ορθογωνίου μέσα στην αρχική εικόνα. Με βάση αυτές τις συντεταγμένες δημιουργείται στην μνήμη μία νέα εικόνα στο μέγεθος του περικλείοντος ορθογωνίου που περιέχει και όλη την αντίστοιχη περιοχή. Ακόμα, δημιουργείται και μία δομή δυαδικής μάσκας, η οποία συσχετίζεται με την νέα “μικρή” εικόνα, η οποία είναι στο ίδιο μέγεθος με το ορθογώνιο που υπολογίστηκε και οι τιμές της είναι 1 για τα εικονοστοιχεία της περιοχής και 0 για τα υπόλοιπα. Παραδείγματα εικόνας και χάρτη περιοχών καθώς και των νέων εικόνων και δυαδικών μασκών που δημιουργούνται στην μνήμη υπάρχουν στα σχήματα 4.1 και 4.2 αντίστοιχα. Τέλος κάθε φορά οι δύο νέες αυτές δομές που αντιστοιχούν στο τμήμα της αντίστοιχης περιοχής της εικόνας δίνονται για εξαγωγή, όπως δίνεται και μία αρχική εικόνα με την μάσκα της. Οι δύο διαδικασίες του XM και του VDE φαίνονται στα σχήματα 4.3 και 4.4.

Μια αναπαράσταση των συναρτήσεων που καλούνται και χρησιμοποιούνται από το XM και το

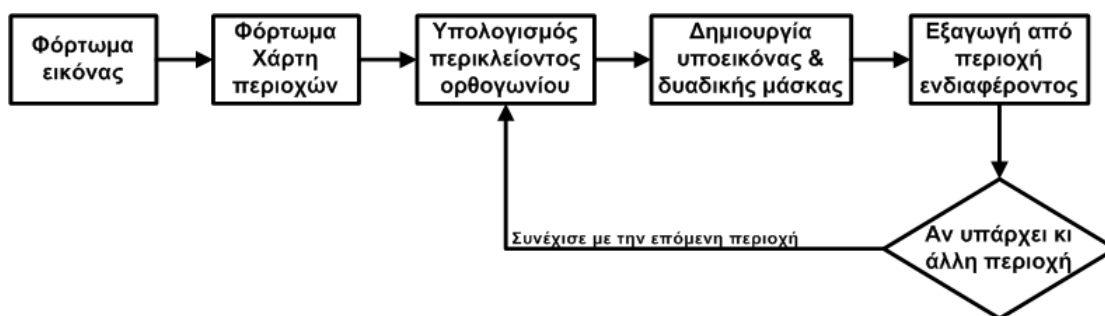


Σχήμα 4.2: Υποεικόνες και δυαδικές μάσκες στο μέγεθος του περικλείοντος ορθογωνίου για την κάθε περιοχή

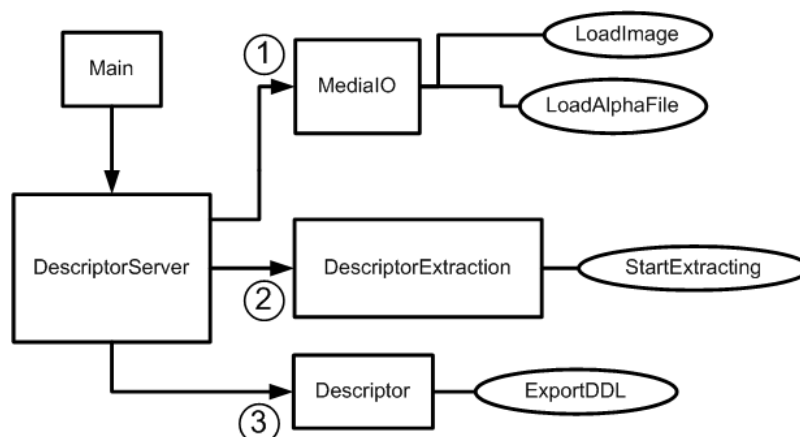
VDE φαίνεται στα σχήματα 4.5 και 4.6 αντίστοιχα. Στο σχήμα 4.5 για το XM απεικονίζεται η κλάση *DescriptorServer* που αφορά την εφαρμογή εξυπηρετητή για τον κάθε περιγραφέα (π.χ. *DominantColorServer*), η κλάση *MediaIO* για την είσοδο εικόνων με συναρτήσεις όπως *LoadImage* (για φόρτωμα εικόνας) και *LoadAlphaFile* (για φόρτωμα της δυαδικής μάσκας). Επίσης η κλάση *DescriptorExtraction* (π.χ. *DominantColorExtraction*) που αφορά την εξαγωγή του περιγραφέα η οποία ξεκινά με την συνάρτηση *StartExtracting* και η κλάση *Descriptor* (π.χ. *DominantColor*) για την εξαγωγή του περιγραφέα με την γλώσσα καθορισμού περιγραφής σε αρχείο XML με την συνάρτηση *ExportDDL*. Στο σχήμα 4.6 βλέπουμε τις αλλαγές που έχουν γίνει για την εξαγωγή περιγραφέα από χάρτη περιοχών. Προστέθηκε η νέα συνάρτηση *LoadRegionMapImage* για το φόρτωμα του χάρτη περιοχών στην μνήμη, η *GetNumberOfRegions* η οποία επιστρέφει τον αριθμό των περιοχών που περιέχονται σε ένα αρχείο χάρτη περιοχών και η *GetNumberOfBndboxes* η οποία επιστρέφει τον αριθμό των ορθογώνιων περιοχών που περιέχονται σε αρχείο κειμένου με τις συντεταγμένες τους. Επίσης δημιουργήθηκε μία νέα κλάση που αφορά την δημιουργία της νέας κάθε φορά υποεικόνας. Η συνάρτηση *GetBoundingBox* υπολογίζει τις συντεταγμένες του περικλείοντος ορθογωνίου για την εκάστοτε περιοχή και η *GetNextRegion* αφού καλέσει την *GetBoundingBox* δίνει στην μνήμη την υποεικόνα που αντιστοιχεί στη επόμενη περιοχή. Η *GetNextRegionFromBox* δίνει στην μνήμη την



Σχήμα 4.3: Διαδικασίες εξαγωγής για XM

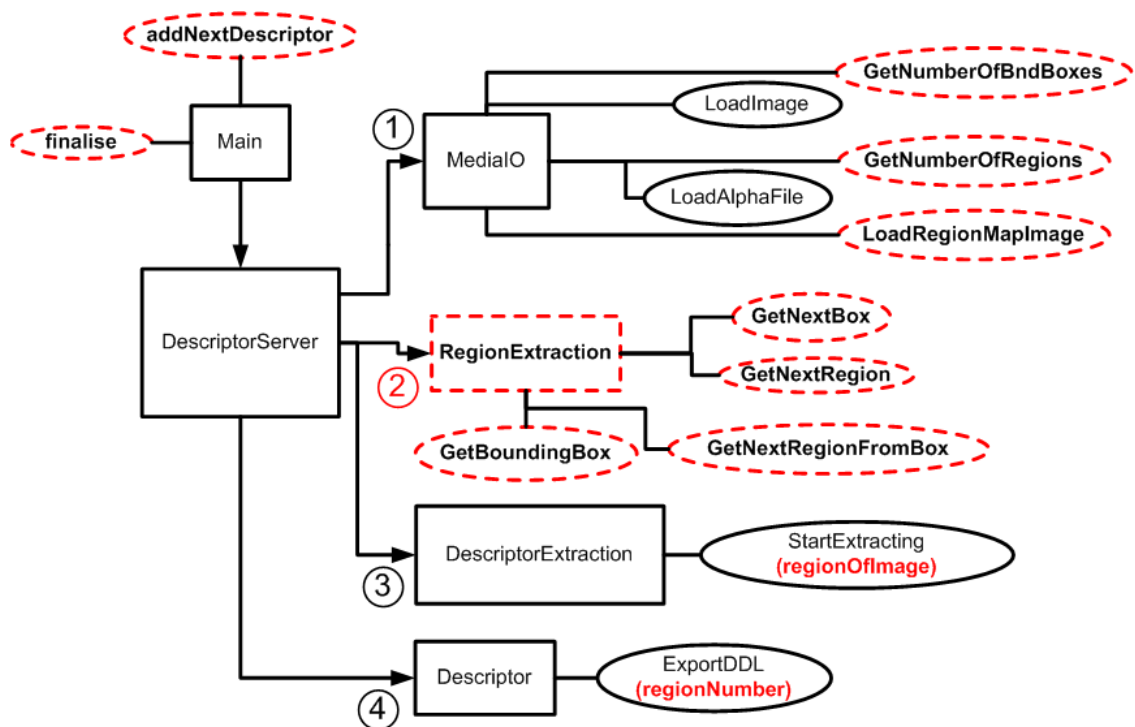


Σχήμα 4.4: Διαδικασίες εξαγωγής για VDE



Σχήμα 4.5: Συναρτήσεις XM - ορθογώνιο: κλάση, έλλειψη: μέθοδος, βέλος: κλάση καλεί κλάση, αρίθμηση με την σειρά που καλούνται οι κλάσεις

υποεικόνα που αντιστοιχεί στην ορθογώνια περιοχή που αντιστοιχεί σε μία τετράδα συντεταγμένων, η οποία τετράδα διαβάζεται από το αρχείο κειμένου με την συνάρτηση *GetNextBox*. Ακόμα τροποποιήθηκε η συνάρτηση *StartExtracting* να δέχεται σαν όρισμα την υποεικόνα από την οποία κάθε φορά θέλουμε να γίνει η εξαγωγή (δηλαδή την εικόνα που επιστρέφει η *GetNextRegion*) και η *ExportDDL* να δέχεται όρισμα τον αύξοντα αριθμό της περιοχής για να αποθηκεύεται στο XML αρχείο εξόδου. Μία μέθοδος *addNextDescriptor* καλείται από την *main* για να προσθέτει σε ένα string τις εξόδους από τους περιγραφείς που ζητούνται στην περίπτωση κλήσης για πολλούς περιγραφείς και η μέθοδος *finalise* για να δημιουργήσει το τελικό αρχείο XML με περιεχόμενο τα αποτελέσματα για όλους τους περιγραφείς. Η κλάση *DescriptorServer* έχει την μέθοδο *start* και είναι εκείνη η οποία καλεί τις μεθόδους των άλλων κλάσεων με την σειρά που φαίνεται και στα σχήματα. Το περιεχόμενο της όπως



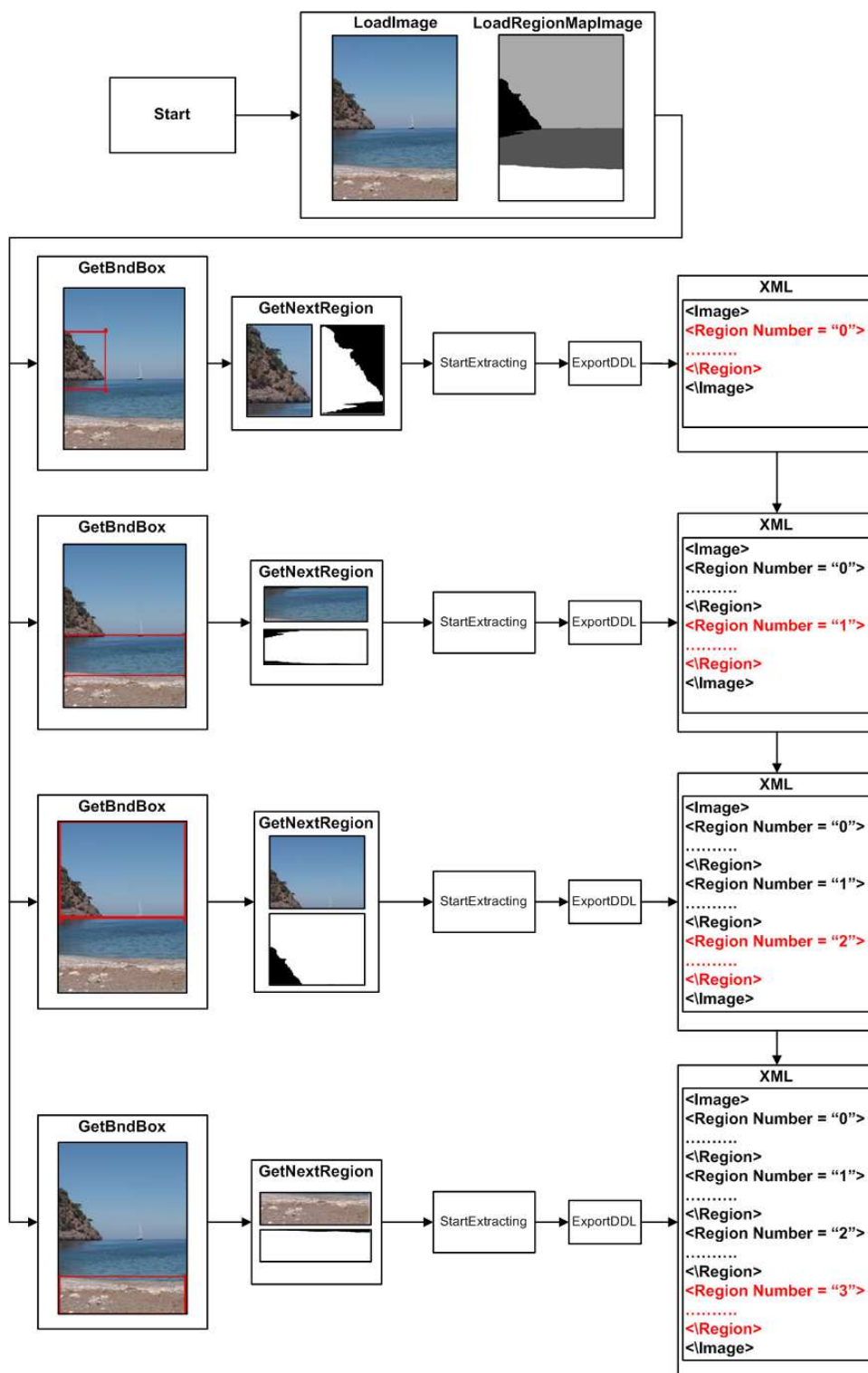
Σχήμα 4.6: Συναρτήσεις VDE - Κόκκινη διακεκομμένη γραμμή: νέες κλάσεις και μέθοδοι, έντονη γραμματοσειρά ονόματα νέων μεθόδων και κλάσεων, έντονη και κόκκινη γραμματοσειρά: νέα ορίσματα

ήταν πριν και όπως είναι τώρα στο VDE φαίνονται στους πίνακες 4.1 και 4.2.

4.2.1 Λεπτομέρειες στον κώδικα

Για κάθε περιγραφέα υπάρχει μία κλάση για την εφαρμογή εξαγωγής, που ονομάζεται *DescriptorSAppl.cpp* (ανήκουν στο project *XMLibAP*). Επίσης υπάρχει η κλάση η οποία είναι υπεύθυνη για τον υπολογισμό του περιγραφέα και ονομάζεται *DescriptorExtraction.cpp* (ανήκουν στο project *XMLibV*), και η κλάση *Descriptor.cpp* (ανήκουν επίσης στο project *XMLibV*) όπου *Descriptor* για τις παραπάνω κλάσεις είναι το όνομα του κάθε περιγραφέα (π.χ. *DominantColor*, *RegionShape*, κλπ.), δηλαδή όλες αυτές οι κλάσεις υπάρχουν για κάθε ένα από τους 8 περιγραφείς που είναι και το περιεχόμενο της εφαρμογής VDE. Έτσι, όταν γίνεται αναφορά για αλλαγές που έγιναν για παράδειγμα στο αρχείο *DescriptorExtraction.cpp*, αυτό σημαίνει ότι αλλαγές έγιναν και για τα 8 αρχεία (*DominantColorExtraction.cpp*, *RegionShapeExtraction.cpp*, *ScalableColorExtraction.cpp*, ... κλπ.).

Αρχικά θα αναφερθούν οι νέες μέθοδοι και κλάσεις που δημιουργήθηκαν προκειμένου να υλοποιηθούν οι νέες δυνατότητες. Στο αρχείο *Multimedia.cpp*, το οποίο ανήκει στα *BasicModules* του project *XMLibC*, δημιουργήθηκαν οι μέθοδοι



Σχήμα 4.7: Διαδικασία εξαγωγής με χάρτη περιοχών

<pre> DescriptorServer::Start(char *ListFile, char *Bitstream) { ... MediaIO.CreateDatabase(ListFile); for(int ientry=0;ientry<NDBaseEntries;ientry++){ if (!MediaIO.DBLoadImage()) break; ... theExtractionInterface->StartExtracting(); theDescriptorInterface->ExportDDL(theCollectionInterface); } ... } </pre>	<pre> [MediaIO.cpp] [MediaIO.cpp] [DescriptorExtraction.cpp] [Descriptor.cpp] </pre>
--	--

Πίνακας 4.1: Μέθοδος Start των κλάσεων DescriptorServer στο XM, δεξιά αναφέρονται τα αρχεία στα οποία ορίζονται οι μέθοδοι που κάθε φορά καλούνται

```

void MultiMedia::SetRegionType(int aRegionType)
int MultiMedia::SetRegionFileName(char *aFileName)
int MultiMedia::GetRegionType(void)
char *MultiMedia::GetRegionFileName(void)

```

Η πρώτη και η δεύτερη θέτουν τιμές στις νέες μεταβλητές *m_RegionFileName* και *m_RegionType* του αντικειμένου *Multimedia* ανάλογα με το τι έχει δηλωθεί από την γραμμή εντολών. Το *m_RegionType* θα πάρει τιμή 1 αν έχει δηλωθεί όνομα για αρχείο δυαδικής μίαςκας, 2 για χάρτη περιοχών, 3 για αρχείο κειμένου με λίστα συντεταγμένων και 0 αν δεν έχει δηλωθεί τίποτα από τα προηγούμενα. Αντίστοιχα το *m_RegionFileName* παίρνει τιμή το όνομα του αρχείου ή NULL για *m_RegionType*=0. Η τρίτη και τέταρτη μέθοδος επιστρέφουν τις τιμές των δύο αυτών μεταβλητών.

Μία νέα κλάση δημιουργήθηκε με το αρχείο *RegionExtraction.cpp* (*BasicModules* του project *XMLibB*). Περιέχει μεθόδους για την εύρεση του περιεχόμενου ορθογωνίου για την κάθε περιοχή ενδιαφέροντος όταν εκείνη δηλώνεται με χάρτη περιοχών, για το διάβασμα των συντεταγμένων για τις ορθογώνιες περιοχές από αρχείο, για την δημιουργία των υποεικόνων αλλά και κάποιες μεθόδους για διόρθωση του μεγέθους εικόνων λόγω κάποιων προβλημάτων που προέκυψαν. Τα προβλήματα αυτά και οι μέθοδοι που τα επιλύουν θα περιγραφούν στην υποενότητα 4.3. Για καλύτερη κατανόηση των μεθόδων αναφέρεται ότι *MomVop* είναι η δομή που χρησιμοποιείται από το XM για την αναπαράσταση της εικόνας και *BoundingBox* είναι μία δομή που δημιουργήθηκε για να αναπαραστήσει τις 4 συντεταγμένες μίας ορθογώνιας περιοχής, δηλαδή συντεταγμένες πάνω αριστερού και κάτω δεξιού άκρου. Το *RegionExtraction.cpp* περιέχει τις μεθόδους

```

BoundingBox *GetBndBox(MomVop *mainVop,int currentRegion);
MomVop *RegionExtraction::GetNextRegion(MomVop *mainVop,int currentRegion);
MomVop *RegionExtraction::GetNextRegionFromBox(MomVop *mainVop,
                                                BoundingBox *Box);
BoundingBox *RegionExtraction::GetNextBox(int *listOfBoxes,int i);

```

Η *GetNextRegion* δέχεται σαν όρισμα την αρχική εικόνα και το αύξοντα αριθμό της περιοχής, καλεί

<pre> DescriptorServer::Start(char *ListFile, char *Bitstream, char *RegionTypeFileName,int RegionType){ ... theMediaInterface->SetRegionType(RegionType); theMediaInterface->SetRegionFileName(RegionTypeFileName); ... MediaIO.CreateDatabase(ListFile); for(int ientry=0;ientry<NDBaseEntries;ientry++){ if (!MediaIO.DBLoadImage()) break; int *boxList=MediaIO.GetNumberOfBndBoxes(); int regions=MediaIO.GetNumberOfRegions(); if(RegionType==2 (RegionType==0 && regions>0)){ for(int iRegion=0;iRegion<regions;iRegion++){ nextRegionVop=RegionExtraction.GetNextRegion(theMediaInterface->GetImage(),iRegion); theExtractionInterface->StartExtracting(nextRegionVop); theDescriptorInterface->ExportDDL(theCollectionInterface,iRegion);} else if(RegionType==3 (RegionType==0) && boxList){ for(int iRegion=0;iRegion<boxList[0];iRegion++){ BoundingBox *nextBox=RegionExtraction.GetNextBox(boxList,iRegion); nextRegionVop=RegionExtraction.GetNextRegionFromBox(theMediaInterface->GetImage(),nextBox); theExtractionInterface->StartExtracting(nextRegionVop); theDescriptorInterface->ExportDDL(theCollectionInterface,iRegion);} else{ ... theExtractionInterface->StartExtracting(NULL); theDescriptorInterface->ExportDDL(theCollectionInterface,0); } ... } </pre>	<pre> [Multimedia.cpp] [Multimedia.cpp] [MediaIO.cpp] [MediaIO.cpp] [MediaIO.cpp] [RegionExtraction.cpp] [DescriptorExtraction.cpp] [Descriptor.cpp] [RegionExtraction.cpp] [RegionExtraction.cpp] [DescriptorExtraction.cpp] [Descriptor.cpp] [DescriptorExtraction.cpp] [Descriptor.cpp] </pre>
---	---

Πίνακας 4.2: Μέθοδος Start των κλάσεων DescriptorServer στο VDE, με κόκκινη γραμματοσειρά είναι ότι προστέθηκε

μέσα της την *GetBndBox* η οποία θα επιστρέψει τις συντεταγμένες για την περιοχή με τον αύξοντα αριθμό που δίνεται σαν όρισμα (*currentRegion*). Για το περικλείον ορθογώνιο που σχηματίζεται από τις συντεταγμένες αυτές η *GetNextRegion* θα επιστρέψει έναν δείκτη σε μία δομή εικόνας (*MomVop*) η οποία αντιστοιχεί στην υποεικόνα στο μέγεθος του περικλείοντος ορθογωνίου. Οι δύο αυτές χρησιμοποιούνται για την περίπτωση που υπάρχει χάρτης περιοχών. Αν υπάρχει αρχείο με συντεταγμένες για ορθογώνιες περιοχές τότε η *GetNextBox* διαβάζει από το αρχείο την *i*-οστή τετράδα συντεταγμένων και εκείνη θα δοθεί σαν όρισμα στην *GetNextRegionFromBox* για να επιστρέψει επίσης την υποεικόνα. Για την αναπαράσταση του περικλείοντος ορθογωνίου έχει οριστεί η δομή *BoundingBox*.

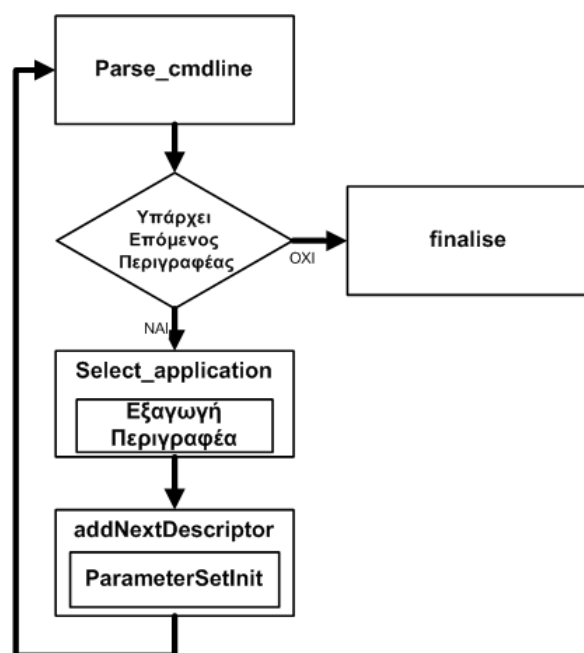
```
struct bndBox // structure for bounding box
{
    int left;
    int right;
    int top ;
    int bottom;
} ;
typedef struct bndBox BoundingBox;
```

Ακόμα πρέπει να αναφερθεί για την δομή *MomVop*, η οποία δηλώνεται στο αρχείο *momusys.h* (*AddressLib* του project *XMLibC*) ότι περιείχε ένα δείκτη (*MomVopExtend *pextend;*) στην κενή δομή *MomVopExtend* για πιθανή επέκταση της δομής. Έτσι χρησιμοποιήθηκε αυτή η δομή για τον χάρτη περιοχών.

```
struct momvop_extend
{
    MomImage *RegionMap;
    int TotalRegions;
};
typedef struct momvop_extend MomVopExtend;
```

MomImage είναι μία δομή που χρησιμοποιείται για το κάθε κανάλι της εικόνας (η *MomVop* περιέχει 3 τέτοιες δομές, μία για κάθε χρωματικό κανάλι). Στην προκειμένη *RegionMap* είναι ένας πίνακας με τις τιμές των pixel του χάρτη περιοχών. Ο πίνακας αυτός δημιουργείται από την μέθοδο *LoadRegionMapImage* που αναφέρθηκενωρίτερα. *TotalRegions* είναι ο αριθμός των περιοχών που δηλώνονται στο *RegionMap*. Επίσης στο αρχείο *vopio.c* (*AddressLib* του project *XMLibC*) υπάρχουν οι μέθοδοι *initvop* και *freevop*, για δέσμευση μνήμης για την δομή *MomVop* και για απελευθέρωση αντίστοιχα. Σε αυτές τις δύο προστέθηκε και ο κώδικας για τις αντίστοιχες λειτουργίες πάνω στην δομή *MomVopExtend* την οποία περιέχει η *MomVop*.

Οι μέθοδοι που καλούνται από την *main* κλάση (*main.cpp*) της εφαρμογής VDE περιέχονται στο αρχείο *version1.cpp* (*BasicModules* του project *XMLibAP*), μία εκ των οποίων είναι η *parse_cmdline*. Διαβάζει τις παραμέτρους από την γραμμική εντολών και δημιουργεί μία δομή *TXMApplicationParameterSet* με πεδία αντίστοιχα για κάθε παράμετρο. Η δομή αυτή δηλώνεται στο αρχείο *version1.h* και προστέθηκαν τα κατάλληλα πεδία για τις νέες παραμέτρους του VDE. Προστέθηκε και ο αντίστοιχος κώδικας στην *parse_cmdline* για το διάβασμα των παραμέτρων αυτών. Στην *main* προστέθηκε ένας



Σχήμα 4.8: Διάγραμμα ροής για την Main

βρόγχος επανάληψης, ο οποίος δίνει την δυνατότητα να γίνεται η εξαγωγή για πολλούς περιγραφείς, το αποτέλεσμα από κάθε εξαγωγή περιγραφέα αποθηκεύεται κανονικά σε ένα αρχείο XML, ενώ σε αυτή την περίπτωση το αρχείο διαβάζεται, έπειτα διαγράφεται και το περιεχόμενο του προστίθεται σε ένα string με την μέθοδο *addNextDescriptor*. Τέλος η μέθοδος *finalise* θα πάρει το συνολικό string σαν όρισμα και θα το αποθηκεύσει αρχείο XML με κατάλληλο όνομα. Οι μέθοδοι *addNextDescriptor* και *finalise* δημιουργήθηκαν στο *version1.cpp*. Επίσης η *ParameterSetInit* δημιουργήθηκε για να αρχικοποιεί την δομή *TXMApplicationParameterSet* μετά από κάθε εκτέλεση εξαγωγής. Η μέθοδος *select_application* είναι εκείνη η οποία καλεί την κάθε εφαρμογή εξαγωγής ανάλογα με τις παραμέτρους που έχουν δοθεί. Ο τρόπος με τον οποίο καλούνται όλες οι παραπάνω από την *main* φαίνεται στο σχήμα 4.8.

Τα αρχεία των 8 εφαρμογών είναι τα *DescriptorSAPpl.cpp*. Οι αλλαγές και για τα 8 αυτά αρχεία εντοπίζονται στην μέθοδο *Start* μέσα στην οποία αρχίζει και τελειώνει η εξαγωγή του περιγραφέα. Καλούνται οι δύο νέες μέθοδοι *GetNumberOfRegions()* και *GetNumberOfBndBoxes()*. Η πρώτη θα επιστρέψει τον αριθμό των περιοχών ενδιαφέροντος για τον χάρτη περιοχών ο οποίος θα έχει δηλωθεί από την γραμμή εντολών ή αν δεν έχει δηλωθεί περιοχή ενδιαφέροντος με κανέναν από τους 3 τρόπους από γραμμή εντολών τότε για τον χάρτη περιοχών με το προκαθορισμένο όνομα (ενότητα 5.2) αν αυτός υπάρχει. Η δεύτερη θα κάνει το ίδιο αλλά για τις περιοχές του αρχείου με συντεταγμένες. Η ροή εκτέλεσης θα συνεχιστεί με μία τριπλή συνθήκη για τις περιπτώσεις όπου έχουμε χάρτη περιοχών, αρχείο με συντεταγμένες και δυαδική μάσκα ή ολόκληρη εικόνα. Για τις δύο πρώτες καλούνται οι μέθοδοι που αναφέρθηκαν νωρίτερα για την δημιουργία των υποεικόνων για την κάθε περιοχή μέσα σε ένα βρόγχο επανάληψης με τόσες επαναλήψεις όσες και οι περιοχές. Για την εξαγωγή από κάθε περιοχή καλείται η *StartExtracting* με νέο όρισμα έναν δείκτη στην νέα υποπεριοχή ή NULL αν έχουμε

<pre> DescriptorExtraction::StartExtracting(MomVop *nVop) { ... if(nVop) ImageMedia=nVop; else ImageMedia=m_Media->GetImage(); ... Extract(); ... } </pre>	<p>[Multimedia.cpp]</p> <p>[DescriptorExtraction.cpp]</p>
---	---

Πίνακας 4.3: Μέθοδος StartExtracting των κλάσεων DescriptorExtraction στο VDE, με κόκκινη γραμματοσειρά είναι ότι προστέθηκε

δυναμική μάσκα ή ολόκληρη εικόνα. Για την αποθήκευση σε αρχείο XML καλείται η ExportDDL με νέο όρισμα τον αύξοντα αριθμό της περιοχής ώστε να αποθηκεύεται στο XML.

Η *StartExtracting* βρίσκεται στο αρχείο *DescriptorExtraction.cpp*. Η μόνη αλλαγή που έγινε στον κώδικα της φαίνεται στον πίνακα 4.3.

Το *nVop* είναι δείκτης σε δομή *MomVop*, δηλαδή σε εικόνα και χρησιμοποιείται για τις νέες υποπεριοχές που δημιουργήθηκαν. Η εξαγωγή γίνεται πάνω στην εικόνα που δείχνει ο δείκτης *ImageMedia*. Μέχρι τώρα εκείνος έδειχνε σε πάντα στην αρχική εικόνα. Τώρα θα δείχνει στην νέα υποεικόνα κάθε φορά εκτός από όταν στην *StartExtracting* δίνεται σαν όρισμα NULL, οπότε και έχουμε περίπτωση δυαδικής μάσκας από όλη την εικόνα ή εξαγωγή από όλη την αρχική εικόνα. Τότε θα δείχνει και πάλι στην αρχική εικόνα, στην οποία έναν δείκτη επιστρέφει η *m_Media->GetImage()*.

Η *ExportDDL* βρίσκεται στο αρχείο *Descriptor.cpp* και έχει σαν αποτέλεσμα την μετατροπή του περιγραφέα σε XML δομή και αποθήκευση του σε αρχείο. Το νέο όρισμα όπως ειπώθηκε σαν σκοπό έχει να αποθηκεύεται στο αρχείο για να είναι ευδιάκριτο κάθε περιγραφέας σε ποιά περιοχή αντιστοιχεί. Έτσι και ο κώδικας που προστέθηκε στην μέθοδο δημιουργεί ένα νέο “παιδί” στο XML δέντρο με όνομα *Region* και ιδιότητα τον αριθμό της περιοχής.

Συνοπτικά ο τρόπος με τον οποίο καλούνται οι μέθοδοι που αναφέρθηκαν ότι χρησιμοποιούνται για την εξαγωγή από χάρτη περιοχών, καθώς και το αποτέλεσμα τους πάνω σε εικόνα φαίνονται στο σχήμα 4.7.

4.3 Bugs και Επιπλέον Αλλαγές στον Κώδικα

4.3.1 Περιγραφέας Κύριων Χρωμάτων

Η εξαγωγή του περιγραφέα στην τελευταία έκδοση του XM (6.1) γίνεται με λανθασμένα αποτελέσματα. Έτσι έγιναν αλλαγές¹ που συγκεκριμένα αφορούν ένα λανθασμένο *casting* και μία διαίρεση με τον αριθμό όλων των εικονοστοιχείων της εικόνας και όχι μόνο της περιοχής ενδιαφέροντος. Οι

¹Το πρόβλημα έχει εντοπιστεί και διορθωθεί από τον *Nicolas Voisine*

αλλαγές που έγιναν στον κώδικα του XM6.1 εντοπίζονται στο αρχείο *DominantColorExtraction.cpp*, το οποίο ανήκει στο project *XMLibAP*. Στην μέθοδο *Extract* οι γραμμές

```
for( j=0;j<m_CurrSize; j++ )
m_Weights[j] /= imsize;
```

αντικαθίστανται από τις

```
unsigned char *pAlpha;
float ctpixel=0.;

for( i=0; i<imsize; i++ )
{
    pAlpha = &quantImageAlpha[i];
    if(!quantImageAlpha || *pAlpha)
        {
            ctpixel++;
        }
}
for( j=0; j<m_CurrSize; j++ )
    m_Weights[j] /= ctpixel;
```

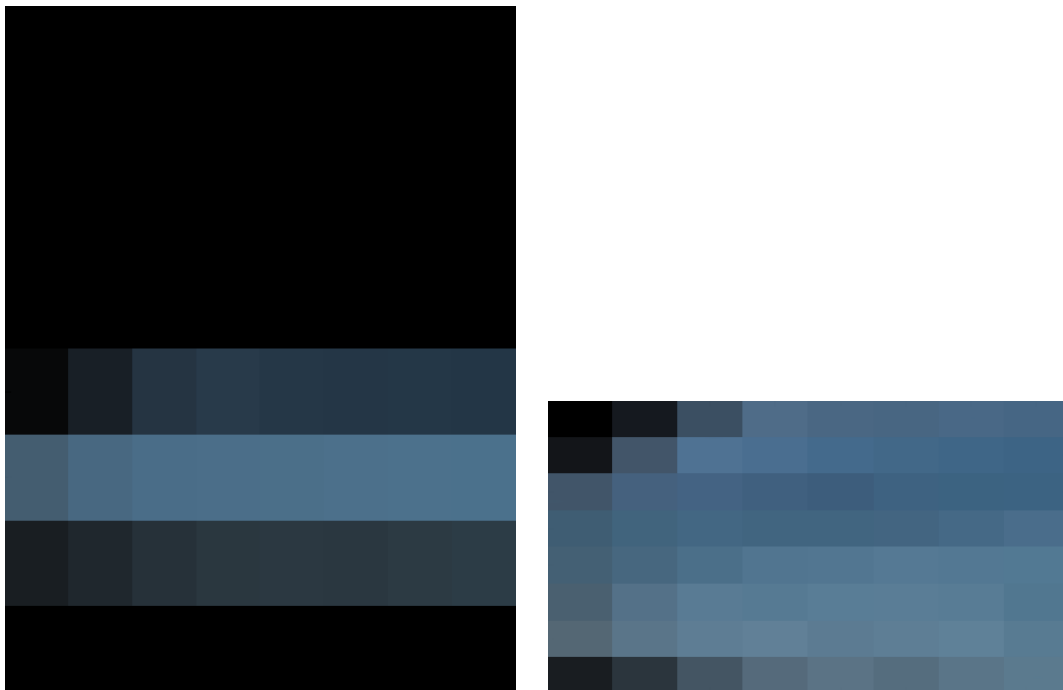
καθώς αφαιρείται και η εντολή `#include "imgutil.h"`. Στο αρχείο *imgutil.h*, το οποίο ανήκει στα *BasicModules* του project *XMLibC*, πρέπει να προστεθεί η εντολή `int lroundf(float x);` ανάμεσα στις γραμμές 63-65 ώστε να είναι

```
#ifdef _VISUALC_
int round(float x);
int lroundf(float x);
#endif
```

αλλιώς γίνεται λανθασμένο *casting*.

4.3.2 Περιγραφείας Διάταξης Χρώματος

Ο περιγραφείας διάταξης χρώματος δημιουργεί αρχικά μία μικροσκοπική εικόνα διαστάσεων 8×8 . Στο XM 6.1 η μετατροπή αυτή γινότανε πάνω σε ολόκληρη την εικόνα και όχι μόνο πάνω στην περιοχή η οποία μπορεί να δηλώνεται με δυαδική μάσκα, γεγονός το οποίο χαρακτηρίστηκε σαν *bug*. Αυτό είχε σαν αποτέλεσμα πανομοιότυπες περιοχές οι οποίες μπορεί να δίνονται σαν μέρος διαφορετικών εικόνων να δίνουν διαφορετικό περιγραφέα. Το ίδιο επίσης συνέβαινε σε περίπτωση που από την ίδια περιοχή γίνει εξαγωγή με δυαδική μάσκα ή με χάρτη περιοχών. Αυτό διότι με τον χάρτη περιοχών κατασκευάζεται η υποεικόνα στο μέγεθος του περικλείοντος παραθύρου. Για παράδειγμα για την εικόνα του σχήματος 4.1, όσον αφορά την περιοχή της θάλασσας, αν εκείνη δοθεί με δυαδική μάσκα ή με μία περιοχή μέσα στον χάρτη περιοχών όπως φαίνεται στο ίδιο σχήμα, τότε οι αντίστοιχες μικροσκοπικές εικόνες είναι εκείνες που φαίνονται στο σχήμα 4.9. Στην περίπτωση της δυαδικής



Σχήμα 4.9: Μικροσκοπικές εικόνες κατασκευασμένες κατά την εξαγωγή του περιγραφέα διάταξης χρώματος, για την περιοχή θάλασσας του σχήματος 4.1 με δυαδική μάσκα (αριστερά), με χάρτη περιοχών (δεξιά)

μάσκας περιλαμβάνονται περιοχές εκτός της περιοχής ενδιαφέροντος και υπολογίζονται σαν μαύρες κι έτσι επηρεάζουν το αποτέλεσμα. Στην περίπτωση της μικροσκοπικής εικόνας με χάρτη περιοχών φαίνεται ότι υπάρχει και σαφώς περισσότερη πληροφορία. Διακρίνεται έτσι η ανάγκη να υπολογίζεται η εικόνα 8×8 πάνω στην περιοχή η οποία δηλώνεται με δυαδική μάσκα, με χάρτη περιοχών και ορθογώνιες περιοχές για το VDE. Για τον σκοπό αυτό έγιναν αλλαγές στον κώδικα της μεθόδου *CreateSmallImage* του αρχείου *ColorLayoutExtraction.cpp*, οι οποίες παρουσιάζονται στον πίνακα 4.4. Έτσι τελικά υπολογίζεται το περικλείον ορθογώνιο της περιοχής της δυαδικής μάσκας το οποίο έχει σαν όρια τις συντεταγμένες $max_x, max_y, min_x, min_y$ και διαστάσεις $xsize, ysize$. Τελικά οι υπολογισμοί για την μικροσκοπική εικόνα γίνονται πάνε σε αυτή την περιοχή.

4.3.3 Περιγραφέας Ομοιογενούς Υφής

Η μονάδα εξαγωγής για αυτόν τον περιγραφέα στην έκδοση 6.1 του XM δεχόταν σαν είσοδο εικόνες με ελάχιστη διάσταση 128×128 pixels (εικονοστοιχεία). Το γεγονός αυτό όμως σε συνδυασμό με την τεχνική που ακολουθήθηκε για την εξαγωγή από διάφορες περιοχές, αυτή δηλαδή της δημιουργίας νέων υποεικόνων στην μνήμη προκάλεσε πρόβλημα. Αυτό γιατί για πολλές περιοχές σε έναν χάρτη περιοχών (ο οποίος συνήθως είναι προϊόν ενός εργαλείου κατάτμησης) το περικλείον ορθογώνιο είναι μικρότερο από 128×128 pixels. Έτσι σε αυτή την περίπτωση η υποεικόνα αυτή συμπληρώνεται με μηδενικά ώστε να φτάσει το μέγεθος αυτό και να μπορεί να δοθεί σαν είσοδος

```

ColorLayoutExtractionTool::CreateSmallImage(MomVop *src, short small_img[3][64]){
:
int  xsize, ysize;
int max_x = 0, max_y = 0, min_x = src->width-1, min_y = src->height-1;
int min_size, re_xsize, re_ysize;
if(src->a_chan){
for(j=0; j<src->height; j++){
for(i=0; i<src->width; i++){
if(src->a_chan->data->u[j*src->width+i]){
if(max_x<i) max_x = i;
if(max_y<j) max_y = j;
if(min_x>i) min_x = i;
if(min_y>j) min_y = j;
}}}
xsize = max_x - min_x + 1;
ysize = max_y - min_y + 1; }
else{
xsize = src->width;
ysize = src->height;
min_x = min_y = 0;
}
for( y=0; y <ysize; y++){
for( x=0; x < xsize; x++) {
y_axis = (int)(y/(ysize/8.0));
x_axis = (int)(x/(xsize/8.0));
k = y_axis * 8 + x_axis;
//these 3 lines replaced commands G = *pG++; B = *pB++; R = *pR++;
G = *(pG+(y+min_y)*src->width+(x+min_x));
B = *(pB+(y+min_y)*src->width+(x+min_x));
R = *(pR+(y+min_y)*src->width+(x+min_x));
if(src->a_chan){
//this line replaced command A = *pA++;
A = *(pA+(y+min_y)*src->width+(x+min_x));
if (A==0) continue; }
... }}
...
}

```

Πίνακας 4.4: Μέθοδος CreateSmallImage της κλάσης ColorLayoutExtraction στο VDE, με κόκκινη γραμματοσειρά είναι ότι προστέθηκε

στην μονάδα εξαγωγής. Αυτό δεν δημιουργεί πρόβλημα στην εξαγωγή του περιγραφέα, αφού δεν εξάγεται περιγραφέας από την πρόσθετη περιοχή με μία εξαίρεση σε πολύ μικρές περιοχές όπως θα δούμε στην ενότητα 6.2. Για τον σκοπό αυτό στο αρχείο *RegionExtraction.cpp*, το οποίο ανήκει στα *BasicModules* του project *XMLibB*, δημιουργήθηκε η μέθοδος

```
MomVop *RegionExtraction::smallRegionsHT(MomVop *nextRegionVop)
```

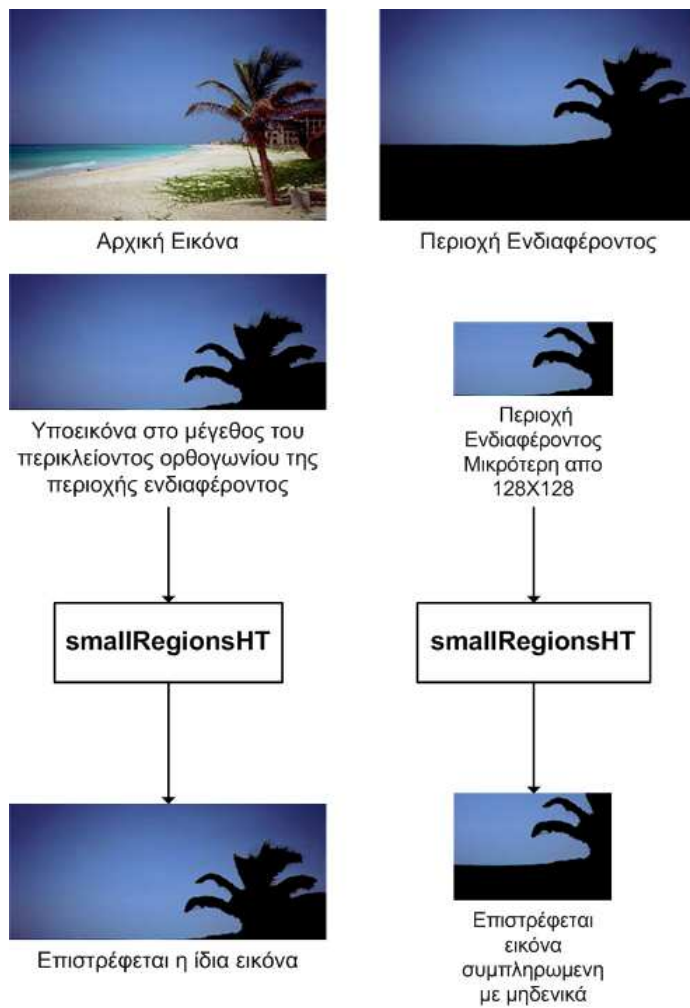
η οποία παίρνει σαν είσοδο έναν δείκτη σε δομή *MomVop*, δηλαδή δομή εικόνας, ελέγχοντας αν η εικόνα έχει διαστάσεις μικρότερες από 128×128 . Στην περίπτωση που ισχύει αυτό επιστρέφει έναν δείκτη στην ίδια εικόνα αλλά συμπληρωμένη με μηδενικά ώστε να φτάσει το κατάλληλο μέγεθος (128×128), αλλιώς επιστρέφει τον δείκτη που δόθηκε σαν είσοδο, την εικόνα ως είχε δηλαδή. Η μέθοδος αυτή καλείται στο αρχείο *HomoTextureSAppl.cpp* του project *XMLibC* για κάθε νέα υποπεριοχή που δημιουργείται από τον χάρτη περιοχών ή το αρχείο με τις ορθογώνιες περιοχές. Καλείται δηλαδή με είσοδο είτε την έξοδο της μεθόδου *GetNextRegion* που χρησιμοποιείται για χάρτη περιοχών, είτε της *GetNextRegionFromBox* που χρησιμοποιείται για αρχείο ορθογώνιων περιοχών. Η λειτουργία της μεθόδου παρουσιάζεται στο σχήμα 4.10.

4.3.4 Περιγραφέας Δομής Χρώματος

Το δομικό στοιχείο που σαρώνει ολόκληρη την εικόνα εισόδου κατά την εξαγωγή αυτού του περιγραφέα έχει διάσταση που εξαρτάται από αυτή της εικόνας. Το στοιχείο αυτό είναι τετράγωνο με ελάχιστο μέγεθος διάστασης ίσο με 8 εικονοστοιχεία. Στην έκδοση 6.1, στην περίπτωση που μία εικόνα εισόδου έχει τη μία διάσταση μικρότερη από 8 εικονοστοιχεία, η εφαρμογή τερματίζεται με σφάλμα και αυτό γιατί ένας δείκτης (pointer) δείχνει εκτός της εικόνας. Χρησιμοποιήθηκε κι εδώ η τεχνική συμπλήρωσης με μηδενικά. Αν και δεν φαίνεται χρήσιμη η εξαγωγή περιγραφής για εικόνες που έχουν την μία διάσταση τόσο μικρή, η διόρθωση αυτή είναι αναγκαία σε περιπτώσεις που η κατάτμηση της εικόνας έχει δώσει σαν αποτέλεσμα περιοχή με τέτοια διάσταση. Μία τέτοια περιοχή θα μπορούσε να ήταν μακρόστενη, για παράδειγμα 7×600 εικονοστοιχεία, και να αντιστοιχεί σε κάτι το πραγματικό όπως για παράδειγμα ένας δρόμος έτσι όπως θα φαινόταν σε μία αεροφωτογραφία. Για τον σκοπό αυτό παρομοίως με το πρόβλημα για τον περιγραφέα Ομοιογενούς Υφής δημιουργήθηκαν οι μέθοδοι *smallRegionsCST* και *smallRegionsCST2* στο αρχείο *RegionExtraction.cpp*. Η πρώτη καλείται παρομοίως με την *smallRegionsHT* μέσα στο αρχείο *ColorstructureSAppl.cpp*. Η δεύτερη δημιουργήθηκε για την περίπτωση εξαγωγής από δυαδική μάσκα, στην οποία περίπτωση και το XM6.1 τερματιζόταν με σφάλμα.

4.3.5 Περιγραφέας Σχήματος με βάση το Περίγραμμο

Για κάθε εφαρμογή εξαγωγής (εφαρμογή εξυπηρετητή) πριν την έναρξη του σταδίου εξαγωγής γίνονται κάποιες αρχικοποιήσεις, μερικές εκ των οποίων αφορούν και την έξοδο του αποτελέσματος σε αρχείο XML, όπως έχει δηλωθεί στις παραμέτρους του XM 6.1. Μέσα σε αυτές τις διαδικασίες παραμένει ένας δείκτης που κρατά το αρχείο αυτό ανοιχτό με αποτέλεσμα να μην μπορεί να διαγραφεί, πράξη αναγκαία στην εφαρμογή VDE. Αυτό διότι στην περίπτωση που ζητείται εξαγωγή για πολλούς περιγραφείς, δημιουργείται για τον καθένα η έξοδος σε XML αρχείο, διαβάζεται και αποθηκεύεται σε μία αλφαριθμητική σειρά χαρακτήρων (string) μαζί με τα αποτελέσματα των υπολοίπων. Έπειτα



Σχήμα 4.10: Περιγραφή λειτουργίας μεθόδου smallRegionsHT

το XML αρχείο για τον κάθε περιγραφέα διαγράφεται και το συνολικό αποτέλεσμα αποθηκεύεται σε ένα μοναδικό XML αρχείο. Στην περίπτωση του συγκεκριμένου περιγραφέα αυτό το αρχείο δεν μπορούσε να διαγραφεί. Το πρόβλημα αυτό διορθώθηκε για το VDE. Συγκεκριμένα στο αρχείο *ContourShapeSAppl.cpp* του project *XMLibApp*, οι αρχικοποιήσεις που γίνονται μέσα στην μέθοδο *ContourShapeServer::Start* αντικαταστάθηκαν με τις αρχικοποιήσεις που γίνονται για τον περιγραφέα κύριων χρωμάτων, αντικαθιστώντας βέβαια όπου *DominantColor* με *ContourShape*.

4.3.6 Περιγραφέας Σχήματος με βάση το Σχήμα

Ο τρόπος με τον οποίο πρέπει να λειτουργεί η εξαγωγή του περιγραφέα αυτού (όπως και του περιγραφέα σχήματος με βάση το περίγραμμα) είναι να υπολογίζει τον περιγραφέα για το σχήμα της περιοχής ενδιαφέροντος. Έτσι όταν δοθεί μία εικόνα και η αντίστοιχη δυαδική μάσκα, πρέπει να γίνεται εξαγωγή του περιγραφέα για την δυαδική μάσκα. Συγκεκριμένα για το σχήμα που αποτελείται από τα εικονοστοιχεία που έχουν τιμή 1. Στο XM6.1 αυτό δεν συνέβαινε ακριβώς έτσι αλλά ο υπολογισμός γινόταν για το σχήμα της υπόλοιπης περιοχής (εικονοστοιχεία με τιμή 0). Η εξαγωγή λανθασμένα γινόταν για την εικόνα που δινόταν σαν είσοδος και όχι για την αντίστοιχη δυαδική μάσκα η οποία ενδεχομένως συνόδευε την εικόνα. Επομένως για να γίνει εξαγωγή για το σχήμα της περιοχής ενδιαφέροντος έπρεπε άμεσα να δοθεί σαν είσοδος η δυαδική μάσκα. Επιπλέον λάθος ήταν ότι κατά την εξαγωγή θεωρούνταν δεδομένο ότι η μάσκα είχε τιμές 255 για την περιοχή ενδιαφέροντος πάντα και όχι 1. Τα δύο παραπάνω προβλήματα διορθώθηκαν στο VDE. Έτσι εξαγωγή γίνεται για το σχήμα της περιοχής ενδιαφέροντος, με όποιον τρόπο κι αν δηλώνεται αυτή, είτε τα εικονοστοιχεία της έχουν τιμές 1 είτε 255 (για να είναι και οπτικά αναγνωρίσιμη η περιοχή στην μάσκα). Οι αλλαγές στον κώδικα που έγιναν εντοπίζονται στο αρχείο *RegionShapeExtraction.cpp*, το οποίο ανήκει στο project *XMLib V*. Στις μεθόδους *FindCenterOfMass()*, *FindRadius()* και *ExtractCoefficients()* υπάρχουν οι εντολές

```
unsigned char *pImage = m_Media->y_chan->data->u;
unsigned char size = datasize(m_Media->y_chan);
```

οι οποίες θέτουν έναν δείκτη πάνω στην δομή της εικόνας για την οποία γινόταν η εξαγωγή υπολογίζοντας επίσης και το βήμα με το οποίο ο δείκτης θα σαρώνει την εικόνα, αντικαθίστανται από τις:

```
unsigned char *pImage;
unsigned char size;
if(m_Media->a_chan)
{
    pImage = m_Media->a_chan->data->u;
    size = datasize(m_Media->a_chan);
}
else
{
    pImage = m_Media->y_chan->data->u;
    size = datasize(m_Media->y_chan);
}
```

Έτσι όταν υπάρχει δυαδική μάσκα να συνοδεύει την εικόνα τότε η εξαγωγή θα γίνει για εκείνη και μόνο λόγω απουσίας δυαδικής μάσκας θα γίνει εξαγωγή από την εικόνα εισόδου. Σημειώνεται ότι δυαδική μάσκα στην προκειμένη περίπτωση μπορεί να είναι και εκείνη που προέκυψε από έναν χάρτη περιοχών. Επίσης στην μέθοδο *FindCenterOfMass()*, στην γραμμή 341 έχει προστεθεί η εντολή `if(*pImage==1 && m_Media->a_chan) *pImage=255;` η οποία εξασφαλίζει το γεγονός ότι θα γίνει σωστή εξαγωγή ακόμα και αν η δυαδική μάσκα είχε τιμές 1 στα εικονοστοιχεία της περιοχής. Οι τιμές 1 αντικαθίστανται με 255 όπως και θεωρείται ότι πρέπει να ισχύει στην υπόλοιπη διαδικασία εξαγωγής για συγκεκριμένο περιγραφέα. Τέλος για να διορθωθεί το εσφαλμένο γεγονός πως εξαγωγή γινόταν όχι για την περιοχή ενδιαφέροντος αλλά για όλη την υπόλοιπη περιοχή αντικαθίσταται στις 3 πάλι μεθόδους *FindCenterOfMass()*, *FindRadius()* και *ExtractCoefficients()* η εντολή `if(*pImage < 128)` με την `if(*pImage > 128)`.

Κεφάλαιο 5

Εγχειρίδιο Χρήσης της Εφαρμογής VDE

5.1 Μεταγλώττιση

Η υποενότητα αυτή περιλαμβάνει οδηγίες για το πως θα γίνει η μεταγλώττιση της εφαρμογής VDE και των απαραίτητων συνοδευτικών βιβλιοθηκών. Οι οδηγίες αυτές έχουν δοκιμασθεί σε λειτουργικό σύστημα Windows XP με εγκατεστημένο το Microsoft Visual C++ 6.0.

5.1.1 Μεταγλώττιση των επιπλέον βιβλιοθηκών

Για την μεταγλώττιση και εκτέλεση της εφαρμογής VDE, απαραίτητες είναι δύο επιπλέον βιβλιοθήκες. Οι βιβλιοθήκες αυτές είναι διαθέσιμες στην ιστοσελίδα του XM:

http://www.lis.ei.tum.de/research/bv/topics/mmdb/e_mpeg7.html

Ιστοσελίδες για ImageMagick και xerces αντίστοιχα

<http://www.imagemagick.org/script/index.php>

<http://xerces.apache.org/xerces-c/>

Είναι απαραίτητη η αντιγραφή τους στον ίδιο φάκελο με τον κώδικα της εφαρμογής VDE. Στο παρακάτω παράδειγμα θεωρούμε φάκελο εγκατάστασης τον C:\WORK.

- Απαιτείται το C-API (magick) του ImageMagick (για ανάκτηση και αποθήκευση οπτικών δεδομένων). Χρησιμοποιήθηκε η έκδοση 5.2 του ImageMagick. Πρέπει να γίνει μεταγλώττιση του workspace:

C:\WORK\ImageMagick\ImageMagick.dsw

Για το ImageMagick φάκελος για τις δυναμικές βιβλιοθήκες έχει οριστεί ο ImageMagick/bin και για τις στατικές ο ImageMagick/lib.

- Απαιτείται το Xerces (XML parser). Χρησιμοποιήθηκε το Apache Xerces Parser (version 1.6). Για την μεταγλώττιση:

- Αρχικά είναι απαραίτητη η δημιουργία του φακέλου `C:\WORK\xml\lib`
- Στην συνέχεια πρέπει να γίνει μεταγλώττιση του workspace:
`C:\WORK\xml\Projects\Win32\VC6\xerces-all\xerces-all.dsw`
- Οι στατικές και δυναμικές βιβλιοθήκες βρίσκονται στον φάκελο `xml/Build/Win32/VC6/Release` και όλα τα *.dll, *.lib αρχεία του φακέλου αυτού πρέπει να αντιγραφούν στον `C:\WORK\xml\lib`.
- Εάν η μεταγλώττιση έγινε σε debug mode τότε τα παραπάνω αρχεία θα βρεθούν στο `xml/Build/Win32/VC6/Debug` και θα πρέπει να αντιγραφούν στον `C:\WORK\xml\lib` αλλά και να γίνει μετονομασία της βιβλιοθήκης: `xerces-c_1D.lib` σε `xerces-c_1.lib` καθώς και `xerces-c_1D.exp` σε `xerces-c_1.exp`

5.1.2 Μεταγλώττιση της εφαρμογής VDE

Το .dsp αρχείο έχει οριστεί να έχει τους φακέλους των βιβλιοθηκών δίπλα στο VDE φάκελο. Έτσι, μία εντολή dir στον φάκελο εργασίας θα πρέπει να δίνει:

```
VDE
ImageMagick
Xml
```

Για να υπάρχει πρόσβαση στις στατικές και δυναμικές βιβλιοθήκες πρέπει να προστεθούν τα παρακάτω path των windows:

```
C:\WORK\xml\lib
C:\WORK\IVMLXM\lib
C:\WORK\ImageMagick\bin
```

αρκεί επίσης να αντιγραφούν τα .dll αρχεία από τους παραπάνω φακέλους μέσα στον φάκελο VDE. Η μεταγλώττιση της εφαρμογής VDE τώρα γίνεται με το αρχείο:

```
C:\WORK\VDE\XMWinExe\VDE.dsw
```

Δοκιμάστηκαν η debug και release έκδοση. Η τελευταία δίνει με τη μεταγλώττιση της ένα εκτελέσιμο αρχείο στο φάκελο `C:\VDE`. Το όνομα του μπορεί να επιλεγεί από τις ρυθμίσεις του project. Στα επόμενα θα το αναφέρουμε σαν `vde.exe`. Η debug έκδοση έχει σαν αποτέλεσμα ένα εκτελέσιμο αρχείο αρκετά μεγαλύτερο από την release περιλαμβάνοντας διαγνωστικά μηνύματα και εκτύπωση αποτελεσμάτων στην γραμμή εντολών από την εξαγωγή κάποιων περιγραφών.

5.2 Τρόπος Χρήσης VDE

5.2.1 Κλήση Της Εφαρμογής

Συνοπτικά θα μπορούσαμε να πούμε ότι υπάρχουν δύο τρόποι για την κλήση της εφαρμογής VDE. Ο πρώτος είναι μέσω ενός αρχείου παραμέτρων. Το αρχείο αυτό είναι ένα αρχείο κειμένου στο οποίο περιέχονται όλες οι παράμετροι που θα χρησιμοποιηθούν. Ο δεύτερος τρόπος είναι να δοθούν όλες οι παράμετροι από την γραμμική εντολών.

1. **Με χρήση αρχείου παραμέτρων (parameter file)** Ένας απλός τρόπος χρήσης της εφαρμογής VDE είναι με την χρήση ενός αρχείου παραμέτρων. Το όνομα του αρχείου δηλώνεται μετά το αναγνωριστικό `-p`. Μερικά δείγματα τέτοιων αρχείων δίνονται στον φάκελο `VDE/Parfiles/` όπου και συνίσταται να αποθηκεύονται τα αρχεία παραμέτρων. Ένα τέτοιο αρχείο περιέχει τις παραμέτρους που υποχρεωτικά χρειάζεται η εφαρμογή για εκτέλεση της καθώς και κάποιες ειδικές για κάθε περιγραφέα. Για συντομία αντί κάθε φορά να δηλώνεται όλο το μονοπάτι (path) και όνομα αρχείου ενός αρχείου παραμέτρων, όπως π.χ. `-p parfiles/DominantColor.par`, δίνεται η δυνατότητα να χρησιμοποιείται μόνο το όνομα του περιγραφέα που κάθε φορά θέλουμε πχ. `DominantColor` ή την συντομογραφία που αντιστοιχεί στον καθένα πχ. `DC`. Στην περίπτωση αυτή ως αρχείο παραμέτρων χρησιμοποιείται αυτό με το συγκεκριμένο όνομα και στον προαναφερθέντα φάκελο. Δηλαδή, στο παράδειγμα αυτό, όταν γράψουμε `-p DominantColor` ή `-p DC` θα διαβαστεί το `parfiles/DominantColor.par`.

Γενική Σύνταξη

```
vde.exe -p <parameter_files>
[-i <image> | -l <listfile>] [-b <output>]
[-vm <binary_mask> | -vp <region_map> | -vb <bounding_boxes_file>]
```

Στην παραπάνω γενική σύνταξη εντολής υποχρεωτικό είναι να εμφανίζεται το αναγνωριστικό `-p` με ένα ή περισσότερα αρχεία παραμέτρων ή τις αντίστοιχες συντομογραφίες τους (descriptor codes). Όλα τα υπόλοιπα αναγνωριστικά είναι προαιρετικά εκ των οποίων μόνο ένα κάθε φορά από τα `-i` και `-l` μπορεί να χρησιμοποιηθεί, καθώς και μόνο από τα `-vm`, `-vp` και `-vb`.

Στον πίνακα 5.1 δίνονται τα ονόματα και οι συντομογραφίες για κάθε περιγραφέα καθώς και τα αντίστοιχα ορισμένα εκ των προτέρων αρχεία παραμέτρων.

Δομή Αρχείου Παραμέτρων

Σαν παράδειγμα ενός αρχείου παραμέτρων δίνουμε παρακάτω το περιεχόμενο του `DominantColor.par` :

```
/* default application name*/
Application DominantColorServer

/* default database file name*/
```

Όνομα	Συντομογραφία	Αρχείο Παραμέτρων
ScalableColor	SC	Parfiles/ScalableColor.par
DominantColor	DC	Parfiles/DominantColor.par
ColorLayout	CL	Parfiles/ColorLayout.par
ColorStructure	CST	Parfiles/ColorStructure.par
HomogeneousTexture	HT	Parfiles/HomogeneousTexture.par
EdgeHistogram	EH	Parfiles/EdgeHistogram.par
ContourShape	CS	Parfiles/ContourShape.par
RegionShape	RS	Parfiles/RegionShape.par

Πίνακας 5.1: Πίνακας συντομογραφιών για το αναγνωριστικό -p, αντιστοιχία ονομάτων αρχείων παραμέτρων με descriptorCode

```
ListFile      Databases/mylist.lst

/* default bitstream file*/
/*Bitstream Bitstreams/DominantColor.xml*/

/* dominant color setting */
ColorSpacePresent      0
ColorQuantizationPresent  0
VariancePresent        1
SpatialCoherency       1
```

Οι παράμετροι Application και ListFile, είναι υποχρεωτικές και υπάρχουν σε όλους τους περιγραφείς. Ενώ οι NumberOfBitplanesDiscarded και NumberOfCoefficients είναι ειδικές παράμετροι για τον περιγραφέα κύριων χρωμάτων. Η παράμετρος Bitstream στο VDE υποστηρίζεται αλλά είναι προαιρετική, για αυτό και στο αρχείο παραμέτρων παραπάνω που δόθηκε σαν παράδειγμα είναι μέσα σε σχόλια.

Στην παράμετρο Application δηλώνεται το εργαλείο που θα χρησιμοποιηθεί δηλαδή ο περιγραφέας που θα εξαχθεί, όπως DominantColorServer εάν επιθυμούμε να εξαγάγουμε τον περιγραφέα κυρίου χρώματος. Στο Listfile δηλώνεται ένα αρχείο κειμένου που περιέχει λίστα με τα αρχεία εικόνων που θα δοθούν σαν είσοδο. Για παράδειγμα:

```
Image1.bmp
Image2.jpg
Image3.png
```

Στην παράμετρο Bitstream δηλώνεται το XML αρχείο στο οποίο αποθηκεύεται ο κάθε περιγραφέας. Αν δεν δηλωθεί, τότε η έξοδος παίρνει το όνομα της από το αρχείο Listfile ή από την μοναδική εικόνα που μπορεί να είναι η είσοδος. Παρουσιάζονται λεπτομέρειες σε επόμενη παράγραφο.

-p Γενική περιγραφή του **-p**:

- **Κλήση για έναν περιγραφέα**

Κλήση για έναν περιγραφέα είναι δυνατή με την χρήση μιας εντολής όπως η:

```
vde.exe -p Parfiles/DominantColor.par
```

Εφόσον το DominantColor.par είναι μέσα στο ./parfiles ισοδύναμες εντολές είναι οι :

```
vde.exe -p DominantColor
```

```
vde.exe -p DC
```

- **Κλήση για περισσότερους περιγραφείς**

Δυνατή είναι η εξαγωγή πολλών περιγραφέων με μία εκτέλεση και αποθήκευση τους σε ένα XML. Αυτό μπορεί να γίνει με μία εντολή όπως η:

```
vde.exe -p Parfiles/DominantColor.par Parfiles/RegionShape.par
```

Ισοδύναμες (εφόσον τα αντίστοιχα αρχεία παραμέτρων περιέχονται στον φάκελο Parfiles) είναι οι:

```
vde.exe -p DominantColor RegionShape
```

```
vde.exe -p DC RS
```

-l

Στην περίπτωση που είναι επιθυμητή η επιλογή Listfile από τον χρήστη, είναι δυνατή η παρακάμψη του Listfile που περιέχεται στα αρχεία παραμέτρων τόσο για έναν όσο και για περισσότερους περιγραφείς. Για την εκτέλεση του VDE στην περίπτωση αυτή προστίθεται στη γραμμή εντολών και το αναγνωριστικό **-l** ακολουθούμενο από το μονοπάτι και το όνομα αρχείου του νέου Listfile. Σε αυτή την περίπτωση το Listfile μπορεί να μην περιέχεται καθόλου στο αρχείο παραμέτρων. Ένα παράδειγμα είναι η παρακάτω εντολή:

```
vde.exe -p DC RS DC -l Databases/images.lst
```

-i

Στην περίπτωση που επιθυμείται η εξαγωγή περιγραφέων από μία εικόνα, μπορεί να δοθεί το μονοπάτι και το όνομα αρχείου της μετά από το αναγνωριστικό **-i**, για παράδειγμα με την εντολή:

```
vde.exe -p DC RS DC -i testset/inputImage.png
```

Εάν χρησιμοποιείται η παράμετρος **-i** τότε το Listfile είναι προαιρετικό να υπάρχει μέσα στο αρχείο παραμέτρων μέσα. Στην περίπτωση που δηλώνεται, παρακάμπτεται.

-b

Τέλος, είναι δυνατόν από την γραμμή εντολών να δηλωθεί Bitstream για έναν ή για πολλούς περιγραφείς. Για παράδειγμα με την παρακάτω εντολή θα αποθηκευτεί στο αρχείο Descriptors.xml μέσα στον φάκελο Bitstreams αποτέλεσμα της εξαγωγής του περιγραφέα κύριων χρωμάτων και του περιγραφέα σχήματος με βάση το σχήμα για την εικόνα inputImage.ppm.

```
vde.exe -p DC RS -i inputImage.ppm -b Bitstreams/Descriptors.xml
```

Τα -vm, -vp και -vb εξηγούνται παρακάτω στην περίπτωση χρήσης του -a.

2. Παράμετροι απο γραμμή εντολών (χωρίς χρήση αρχείου παραμέτρων)

Χωρίς την χρήση αρχείου παραμέτρων όλες οι παράμετροι ορίζεται απο τη γραμμή εντολών. Μπορούν να οριστούν μόνο οι γενικές παράμετροι για όλους τους περιγραφείς και όχι οι ειδικές. Για τις ειδικές παραμέτρους χρησιμοποιούνται οι προκαθορισμένες (default) τιμές.

Η παράμετρος Bitstream είναι και σε αυτή την περίπτωση προαιρετική. Ακόμα, κάθε φορά

Αναγνωριστικό	Σημασία
-a	Application
-b	Bitstream
-l	Listfile
-i	Imagefile
-vm	BinaryMask
-vp	RegionMap
-vb	BoundingBoxesFile

Πίνακας 5.2: Πίνακας αναγνωριστικών από την γραμμή εντολών

χρησιμοποιείται μία εκ των δύο παραμέτρων Listfile και Imagefile.

Γενική Σύνταξη

```
vde.exe -a <DescriptorServer>  
[-i <image> | -l <listfile>] [-b <output>]  
[-vm <binary_mask> | -vp <region_map> | -vb <bounding_boxes_file>]
```

Στην παραπάνω γενική σύνταξη εντολής υποχρεωτικό είναι να εμφανίζεται το αναγνωριστικό -a με το όνομα μιας εφαρμογής εξαγωγής περιγραφέα ή με την αντίστοιχη συντομογραφία (descriptor code). Όλα τα υπόλοιπα αναγνωριστικά είναι προαιρετικά εκ των οποίων μόνο ένα κάθε φορά από τα -i και -l μπορεί να χρησιμοποιηθεί, καθώς και μόνο από τα -vm, -vp και -vb.

-a Γενική περιγραφή του -a:

Παραδείγματα εντολών εκτέλεσης:

```
vde.exe -a ScalableColorServer -l Databases/images.lst
```

```
vde.exe -a ScalableColorServer -i testset/inputImage.png
```

Όπως με τις συντομογραφίες για τα αρχεία παραμέτρων έτσι και με τα ονόματα των εφαρμογών εξαγωγής στο αναγνωριστικό -a από την γραμμή εντολών υπάρχουν κάποιες αντιστοιχίες για συντομία, οι οποίες φαίνονται στον πίνακα 5.3.

Ισοδύναμες, δηλαδή, με την εντολή:

Όνομα	Συντομογραφία	Εφαρμογή
ScalableColor	SC	ScalableColorServer
DominantColor	DC	DominantColorServer
ColorLayout	CL	ColorLayoutServer
ColorStructure	CST	ColorStructureServer
HomogeneousTexture	HT	HomogeneousTextureServer
EdgeHistogram	EH	EdgeHistogramServer
ContourShape	CS	ContourShapeServer
RegionShape	RS	RegionShapeServer

Πίνακας 5.3: Πίνακας συντομογραφιών για το αναγνωριστικό -a, αντιστοιχία εφαρμογών εξαγωγής με descriptorCode

```
vde.exe -a ScalableColorServer -i inputImage.ppm
```

είναι οι:

```
vde.exe -a ScalableColor -i inputImage.ppm
```

```
vde.exe -a SC -i inputImage.ppm
```

-vm

Ο τρόπος με τον οποίο δηλώνονται οι περιοχές ενδιαφέροντος ορίζεται από τα αναγνωριστικά -vm, -vp και -vb. Χρησιμοποιούνται για να δηλωθούν τα ονόματα των αρχείων δυαδικής μάσκας, χάρτη περιοχών και αρχείου κειμένου με συντεταγμένες ορθογώνιων περιοχών αντίστοιχα. Αυτό φυσικά έχει νόημα όταν δηλώνεται μία συγκεκριμένη εικόνα για είσοδος (και όχι Listfile) και μπορεί έτσι να οριστεί και το συγκεκριμένο αρχείο για την περιοχή ή περιοχές ενδιαφέροντος. Ένα παράδειγμα εντολής για εξαγωγή από μία εικόνα και την αντίστοιχη δυαδική μάσκα είναι το εξής:

```
vde.exe -a ScalableColorServer -i inputImage.ppm -vm region.pgm
```

-vp

Ένα παράδειγμα εντολής για εξαγωγή από μία εικόνα και τον αντίστοιχο χάρτη περιοχών είναι το εξής:

```
vde.exe -a ScalableColorServer -i inputImage.ppm -vp regionMap.pgm
```

-vb

Ένα παράδειγμα εντολής για εξαγωγή από μία εικόνα και το αντίστοιχο αρχείο με λίστα συντεταγμένων για ορθογώνιες περιοχές είναι το εξής:

```
vde.exe -a ScalableColorServer -i inputImage.ppm -vb boundingBoxes.txt
```

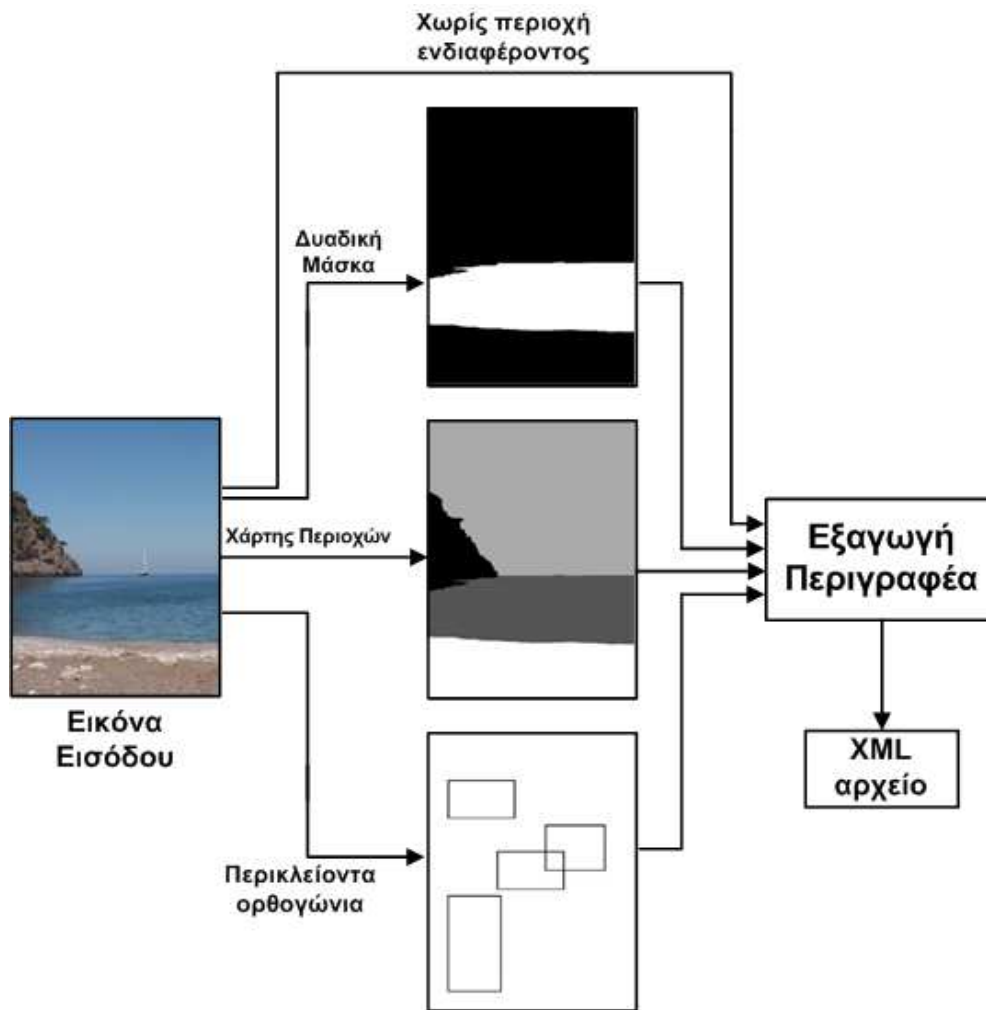
Εάν δηλωθεί από τη γραμμή εντολών Bitstream (με το αναγνωριστικό -b) τότε η έξοδος αποθηκεύεται με το όνομα που δηλώθηκε (και στο φάκελο που δηλώθηκε). Αυτό ισχύει με ή χωρίς χρήση

αρχείου παραμέτρων. Επίσης όταν γίνεται κλήση για έναν περιγραφέα και το αρχείο παραμέτρων του έχει μέσα την παράμετρο Bitstream τότε πάλι θα χρησιμοποιηθεί εκείνη για την ονομασία του αρχείου εξόδου. Στη συνέχεια αν τίποτα από τα παραπάνω δεν ισχύει και δηλωθεί μοναδική εικόνα εισόδου με το αναγνωριστικό `-i` τότε η έξοδος τοποθετείται στον φάκελο Bitstream και έχει το όνομα της εικόνας και επέκταση `.xml`. Για τα παραπάνω παραδείγματα αυτό θα ήταν `Bitstreams/inputImage.xml`. Η τελευταία περίπτωση είναι να μην δηλώνεται ούτε Bitstream ούτε ImageFile. Τότε η έξοδος τοποθετείται στον φάκελο Bitstream και έχει το όνομα του αρχείου κειμένου που δηλώθηκε στο Listfile και επέκταση `xml`. Για τα παραπάνω παραδείγματα αυτό θα ήταν `Bitstreams/images.xml`. Η ύπαρξη του φακέλου Bitstreams κάτω από τον ίδιο φάκελο που βρίσκεται και το εκτελέσιμο είναι υποχρεωτική μιας και οι εξοδοί έχουν προκαθοριστεί να αποθηκεύονται εκεί.

5.2.2 Αναπαράσταση Περιοχών Ενδιαφέροντος

- Σαν είσοδο, η εφαρμογή VDE μπορεί να δεχθεί μία εικόνα σε ένα από τα formats που μπορεί η βιβλιοθήκη ImageMagick να διαβάσει (συγκεκριμένα όλα τα formats που υποστηρίζει η ImageMagick βρίσκονται στην ιστοσελίδα <http://www.imagemagick.org/script/formats.php>). Αν δοθεί μόνη της η εικόνα τότε ο περιγραφέας θα εξαχθεί από ολόκληρη την περιοχή της.
- Άν είναι επιθυμητό η εξαγωγή να γίνει από μία μόνο συγκεκριμένη περιοχή της τότε πρέπει να τοποθετηθεί στον ίδιο φάκελο με την εικόνα μία δυαδική μάσκα (binary mask) σε ένα από τα formats `pgm`, `pbm` και `ppm` που θα ορίζει την περιοχή ενδιαφέροντος με όνομα αρχείου το ίδιο με της κύριας εικόνας ακολουθούμενο από `'_a'` (για εικόνα `image.jpg` η μάσκα θα είναι `image_a.pgm`). Είναι επίσης εφικτό με το αναγνωριστικό `-vm` να ορισθεί για δυαδική μάσκα οποιοδήποτε αρχείο, δίνοντας το μονοπάτι και το όνομα του.
- Στην περίπτωση που απαιτείται η εξαγωγή του περιγραφέα από διάφορες περιοχές της εικόνας ταυτόχρονα, οι περιοχές αυτές πρέπει να ορισθούν σε έναν χάρτη περιοχών (Region Map). Το σύνολο όλων αυτών των περιοχών πρέπει να καλύπτουν όλη την εικόνα. Το αρχείο αυτό τοποθετείται στον ίδιο φάκελο με την εικόνα και πρέπει να έχει ίδιο όνομα και επέκταση `.seg`. Είναι επίσης εφικτό με το αναγνωριστικό `-vr` να ορισθεί ως χάρτης περιοχών οποιοδήποτε αρχείο, δίνοντας το μονοπάτι και το όνομα του.
- Τέλος υπάρχει η τελευταία δυνατότητα σαν είσοδος να οριστούν ορθογώνιες περιοχές των οποίων οι συντεταγμένες δίνονται σε αρχείο με ίδιο όνομα με την εικόνα και επέκταση `.bbx` το οποίο βρίσκεται κάτω από τον ίδιο φάκελο με την εικόνα. Είναι επίσης εφικτό το αρχείο αυτό να ορισθεί με το αναγνωριστικό `-vb`. Οι περιοχές αυτές μπορούν να είναι αλληλοεπιβαλλόμενες.

Το σχήμα 5.1 αναπαριστά τους τέσσερις αυτούς τρόπους εξαγωγής περιγραφέα.

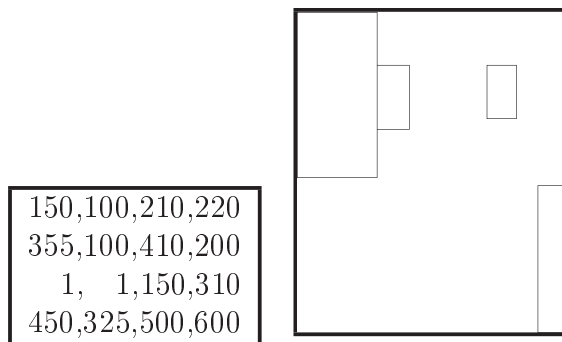


Σχήμα 5.1: Αναπαράσταση περιοχών ενδιαφέροντος για το VDE

Τα αρχεία εισόδου έχουν τις παρακάτω συγκεκριμένες δομές.

Διαδική Μάσκα(-vm): Εάν δεν δηλωθεί συγκεκριμένο όνομα αρχείου στο αναγνωριστικό `-vm` τότε θα πρέπει να έχει όνομα ίδιο με της εικόνας ακολουθούμενο με 'a' και μία από τις 3 επεκτάσεις (ppm, pgm, pbm). Τα εικονοστοιχεία που ανήκουν στην περιοχή από την οποία είναι επιθυμητή η εξαγωγή των περιγραφέων έχουν τιμή 1 ενώ τα άλλα 0.

Χάρτης Περιοχών(-vp): Εάν δεν δηλωθεί συγκεκριμένο όνομα αρχείου στο αναγνωριστικό `-vp` τότε θα πρέπει να έχει ίδιο όνομα με την εικόνα και επέκταση .seg. Σκοπός του είναι να παρέχει σε ένα αρχείο την πληροφορία για όλες τις μη επικαλυπτόμενες περιοχές από τις οποίες είναι επιθυμητή η εξαγωγή περιγραφέων. Η αρίθμηση τους αρχίζει από το μηδέν και συνεχίζει αυξάνοντας κατά 1. Έτσι κάθε εικονοστοιχείο θα έχει σαν τιμή τον αντίστοιχο ακέραιο ανά-

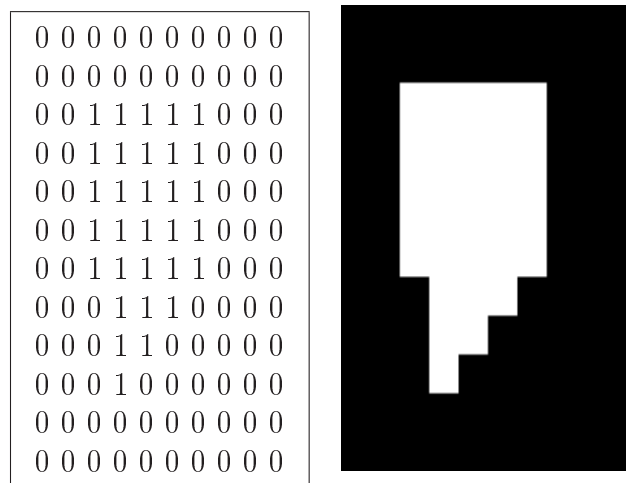


Σχήμα 5.2: Συντεταγμένες ορθογώνιων περιοχών και αντίστοιχες σχεδιασμένες περιοχές ενδιαφέροντος

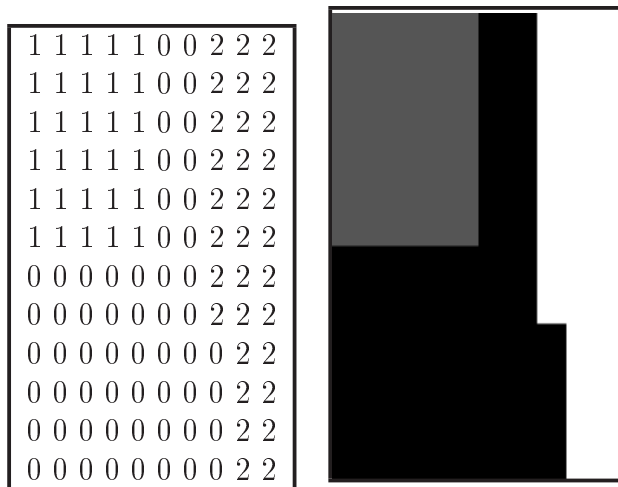
λογα με την περιοχή στην οποία ανήκει.

Αρχείο με λίστα ορθογώνιων περιοχών(-vb): Αυτό το αρχείο κειμένου περιλαμβάνει συντεταγμένες για το πάνω αριστερά και κάτω δεξιά άκρο τις περιοχής. Οι συντεταγμένες χωρίζονται με κόμμα (X1,Y1,X2,Y2) και οι συντεταγμένες κάθε περιοχής γράφονται σε νέα γραμμή. Εάν δεν δηλωθεί συγκεκριμένο όνομα αρχείου στο αναγνωριστικό -vb τότε θα πρέπει να έχει επίσης το ίδιο όνομα με την εικόνα και επέκταση .bvh και να βρίσκεται στον ίδιο φάκελο.

Για την ευκολότερη κατανόηση των τριών αυτών μορφών περιοχών δίνονται παραδείγματα στα σχήματα 5.2, 5.3 και 5.4.



Σχήμα 5.3: Τιμές εικονοστοιχείων δυαδικής μάσκας και η αντίστοιχη εικόνα



Σχήμα 5.4: Τιμές εικονοστοιχείων χάρτη περιοχών και η αντίστοιχη εικόνα

5.3 Παράμετροι Εισόδου και Δομή Περιγραφών

Στην παράγραφο αυτή παρουσιάζονται οι περιγραφείς με λίγα λόγια για την φύση τους, τις παραμέτρους που χρησιμοποιεί ο καθένας, καθώς και τα πεδία εξόδου και το νόημα τους. Παραδείγματα αρχείων παραμέτρων για κάθε περιγραφέα καθώς και αρχείων εξόδου XML βρίσκονται στο παράρτημα Α.

Περιγραφέας Κύριων Χρωμάτων

Ο περιγραφέας αυτός αποτελείται από τα αντιπροσωπευτικά χρώματα σε μία εικόνα/περιοχή (μέχρι 8 σε πλήθος), το ποσοστό τους στην εικόνα/περιοχή, την χωρική συνοχή των κύριων χρωμάτων και την χρωματική διακύμανση για κάθε χρώμα. Τα κυρίαρχα αυτά χρώματα υπολογίζονται πάνω σε κάθε εικόνα και δεν είναι εκ των προτέρων ορισμένα.

	Όνομα	Σχόλια/τιμές
Παράμετροι	ColorSpacePresent	Δεν χρησιμοποιείται
	ColorQuantizationPresent	Δεν χρησιμοποιείται
	VariancePresent	1: Υπολογίζεται για κάθε χρώμα η χρωματική διακύμανση και προστίθεται στην έξοδο (XML file). 0:Δεν υπολογίζεται και δεν υπάρχει αντίστοιχο πεδίο στο xml αρχείο.
	SpatialCoherency	1:Υπολογίζεται η χωρική συνοχή. 0: Δεν υπολογίζεται και παίρνει τιμή 0 στην έξοδο
Έξοδος	SpatialCoherency	Η τιμή της οποία δηλώνει την χωρική ομοιογένεια του Dominant Color στις εικόνες. Κβαντισμένη στο διάστημα 1-31. Το 1 αντιστοιχεί στην μικρότερη ομοιογένεια ενώ το 31 στην μεγαλύτερη. Το 0 χρησιμοποιείται για να δηλώσει ότι το στοιχείο αυτό δεν υπολογίζεται καθόλου
	Index	Οι τιμές του κύριου χρώματος
	Percentage	Το ποσοστό των pixel που έχουν το αντίστοιχο χρώμα
	ColorVariance	Διακύμανση για κάθε χρώμα

Πίνακας 5.4: Παράμετροι και έξοδος για τον περιγραφέα κύριων χρωμάτων

Κλιμακωτός περιγραφέας χρώματος

Ο περιγραφέας αυτός αποτελείται από το αποτέλεσμα της μετατροπής Haar πάνω σε ένα ιστόγραμμα χρώματος υπολογισμένο στον HSV χρωματικό χώρο.

	Όνομα	Σχόλια/τιμές
Παράμετροι	NumberOfCoefficients	Ορίζει τον αριθμό των συντελεστών που θα χρησιμοποιηθούν πιθανές τιμές 16, 32, 64, 128 και 256.
	numberOfBitplanesDiscarded	Ορίζει τον αριθμό των επιπέδων bits (bitplanes) που θα απορριφθούν για την αναπαράσταση για κάθε συντελεστή. Πιθανές τιμές 0, 1, 2, 3, 4, 6, 8. Με την τιμή 8 προκύπτει μία συμπαγή αναπαράσταση με μόνο το bit προσήμου.
Έξοδος	numberOfCoefficients	Ο αριθμός των συντελεστών που χρησιμοποιήθηκαν, είναι ακέραιος από 0 έως 4. Αντιστοιχεί σε $2^{(4+numberOfCoefficients)}$ στοιχεία.
	numberOfBitplanesDiscarded	Ο αριθμός των επιπέδων bits (bitplanes) που απορρίφθηκαν κατά την αναπαράσταση για κάθε συντελεστή. Πιθανές τιμές 0, 1, 2, 3, 4, 6, 8. Με την τιμή 8 προκύπτει μία συμπαγή αναπαράσταση με μόνο το bit προσήμου.
	Coefficients	Οι συντελεστές που χρησιμοποιούνται για την αναπαράσταση του κλιμακωτού περιγραφέα χρώματος, σε προσημασμένους ακέραιους. Αν το numberOfBitplanesDiscarded είναι 8 τότε οι συντελεστές είναι 0 ή 1.

Πίνακας 5.5: Παράμετροι και έξοδος για τον κλιμακωτό περιγραφέα χρώματος

Περιγραφέας διάταξης χρώματος

Ο Περιγραφέας Διάταξης Χρώματος συλλαμβάνει την χωρική διάταξη των κυρίαρχων χρωμάτων σε ένα πλέγμα που δημιουργείται στην περιοχή ενδιαφέροντος.

	Όνομα	Σχόλια/τιμές
Παράμετροι	NumberOfYCoeff	Ο αριθμός των Y συντελεστών (φωτεινότητας) που αποθηκεύονται στο XML file: 1 DC συντελεστής και NumberOfYCoeff-1 AC συντελεστές. Πιθανές τιμές: 3, 6, 10, 15, 21, 28, 64.
	NumberOfCCoeff	Ο αριθμός των Cr και Cb (χρωματικών) συντελεστών που αποθηκεύονται στο XML file : 1 DC συντελεστής και NumberOfCCoeff-1 AC συντελεστές για το καθένα (Cb και Cr). Πιθανές τιμές: 3, 6, 10, 15, 21, 28, 64.
Έξοδος	YDCCoeff	Ο πρώτος κβαντισμένος DCT συντελεστής της Y συνιστώσας σε μορφή ακεραίου.
	YACCoeff	Ο δεύτερος και οι επόμενοι κβαντισμένοι DCT συντελεστές της Y συνιστώσας σε μορφή ακεραίου.
	CbDCCoeff	Ο πρώτος κβαντισμένος DCT συντελεστής της Cb συνιστώσας σε μορφή ακεραίου.
	CbACCoeff	Ο δεύτερος και οι επόμενοι κβαντισμένοι DCT συντελεστές της Cb συνιστώσας σε μορφή ακεραίου.
	CrDCCoeff	Ο πρώτος κβαντισμένος DCT συντελεστής της Cr συνιστώσας σε μορφή ακεραίου.
	CrACCoeff	Ο δεύτερος και οι επόμενοι κβαντισμένοι DCT συντελεστές της Cr συνιστώσας σε μορφή ακεραίου.

Πίνακας 5.6: Παράμετροι και έξοδος για τον περιγραφέα διάταξης χρώματος

Περιγραφέας Δομής Χρώματος

Ο περιγραφέας αυτός καταγράφει με ένα ιστόγραμμα την κατανομή του χρώματος στην εικόνα αλλά και την τοπική χωρική δομή του. Είναι ικανός να ξεχωρίζει εικόνες/περιοχές με ίδια κατανομή χρώματος, στις οποίες όμως τα χρώματα έχουν διαφορετική δομή.

	Όνομα	Σχόλια/τιμές
Παράμετροι	ColorQuantSize	Κβαντισμός χρώματος (32, 64, 128 or 256)
Έξοδος	ColorQuant	Ορίζει τον αριθμό των τιμών που θα χρησιμοποιηθούν για τον περιγραφέα
	Values	Οι τιμές του περιγραφέα

Πίνακας 5.7: Παράμετροι και έξοδος για τον περιγραφέα δομής χρώματος

Περιγραφέας Ιστογράμματος ακμών

Ο περιγραφέας ιστογράμματος ακμών συλλαμβάνει τη χωρική κατανομή των ακμών και την αναπα-

ριστά με ένα ιστόγραμμα το οποίο δηλώνει πόσες και ποιας κατεύθυνσης ακμές εντοπίστηκαν.

	Όνομα	Σχόλια/τιμές
Έξοδος	BinCounts	Οι τιμές του περιγραφέα που αναπαριστά την κατανομή των ακμών στην εικόνα.

Πίνακας 5.8: Έξοδος για τον περιγραφέα ιστογράμματος ακμών

Περιγραφέας ομοιογενούς υφής

Ο περιγραφέας ομοιογενούς υφής αποτελείται από την μέση τιμή και την τυπική απόκλιση των τιμών μίας εικόνας, την ενέργεια και την απόκλιση ενέργειας των τιμών του μετασχηματισμού Fourier της εικόνας, παρέχοντας έτσι ένα χαρακτηρισμό των περιοχών ομοιογενούς υφής.

	Όνομα	Σχόλια/τιμές
Παράμετροι	layer	1: Υπολογίζεται η τυπική απόκλιση για κάθε κανάλι ενέργειας 0: Υπολογίζεται μόνο η ενέργεια του καναλιού
Έξοδος	Average	Μέση ένταση της εικόνας
	StandardDeviation	Τυπική απόκλιση της εικόνας
	Energy	Ενέργεια κάθε καναλιού
	Energy Deviation	Τυπική απόκλιση της ενέργειας κάθε καναλιού

Πίνακας 5.9: Παράμετροι και έξοδος για τον περιγραφέα ομοιογενούς υφής

Περιγραφέας σχήματος με βάση το σχήμα

Ο περιγραφέας σχήματος με βάση το σχήμα εκφράζει τη κατανομή εικονοστοιχείων μέσα σε μια 2-Δ περιοχή αντικειμένου, μπορεί να περιγράψει σύνθετα αντικείμενα που αποτελούνται από πολλαπλάσιες αποσυνδεδεμένες περιοχές καθώς επίσης και απλά αντικείμενα με ή χωρίς τρύπες.

	Όνομα	Σχόλια/τιμές
Έξοδος	MagnitudeOfART	Πίνακας κανονικοποιημένων και κβαντισμένων εντάσεων των 35 ART συντελεστών. Η κανονικοποίηση γίνεται διαιρώντας με την ένταση του μεγαλύτερου συντελεστή.

Πίνακας 5.10: Έξοδος για τον περιγραφέα σχήματος με βάση το σχήμα

Περιγραφέας Σχήματος με βάση το Περίγραμμα

Ο περιγραφέας σχήματος με βάση το περίγραμμα εκφράζει τις ιδιότητες σχήματος του περιγράμματος του αντικειμένου. Τα αντικείμενα για τα οποία χαρακτηριστικά γνωρίσματα σχήματος περιλαμβάνονται στο περίγραμμα περιγράφονται αποτελεσματικά από αυτόν τον περιγραφέα.

	Όνομα	Σχόλια/τιμές
Έξοδος	GlobalCurvature	2 γενικές παράμετροι για το περίγραμμα. Εκκεντρότητα και κυκλικότητα..
	PrototypeCurvature	Εκκεντρότητα και κυκλικότητα του πρωτότυπου περιγράμματος.
	HighestPeakY	Ύψος της υψηλότερης κορυφής.
	Peak	Σύνολο εναπομεινάντων κορυφών με πεδία peakX και peakY για την κάθε μία.

Πίνακας 5.11: Έξοδος για τον περιγραφέα σχήματος με βάση το περίγραμμα

Κεφάλαιο 6

Έλεγχος και Αξιολόγηση

6.1 Διαδικασία Ελέγχου και Αξιολόγησης

Μετά την ολοκλήρωση της δημιουργίας της εφαρμογής VDE ακολούθησε το στάδιο του ελέγχου των αποτελεσμάτων της εφαρμογής. Ήταν επιθυμητό να ερευνηθεί αν υπάρχουν διαφορές ανάμεσα στο VDE και στο XM6.1 όσον αφορά τους τρόπους χρήσης του τελευταίου δηλαδή με δυαδική μάσκα. Επίσης κρίθηκε αναγκαία η έρευνα αποκλίσεων για τις ίδιες περιοχές, αλλά με τους νέους τρόπους χρήσης, καθώς και η αιτία αυτών. Η διαδικασία που ακολουθήθηκε ήταν η εξής: συγκεντρώθηκε ένα μεγάλο σύνολο διαφόρων και ετερογενών εικόνων και έγινε εξαγωγή των κοινών περιγραφών και με τα δύο εργαλεία. Στην συνέχεια έγινε σύγκριση των δύο εξόδων για τις ίδιες περιοχές των εικόνων.

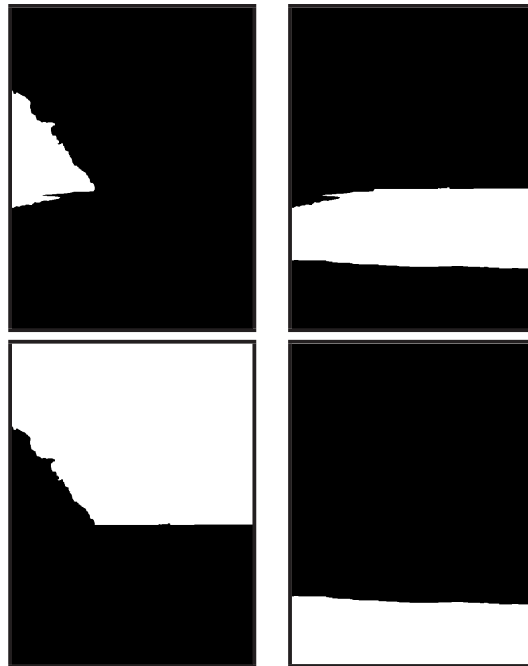
Πιο συγκεκριμένα οι συγκρίσεις που έγιναν ήταν οι παρακάτω:

1. Συγκρίσεις με χρήση δυαδικής μάσκας

Η δυνατότητα εξαγωγής περιγραφών από περιοχή ενδιαφέροντος της εικόνας υπάρχει στην έκδοση 6.1 του XM. Η συγκεκριμένη σύγκριση γίνεται για να ελεγχθεί αν μετά τις αλλαγές που έχουν γίνει το VDE παράγει πανομοιότυπα αποτελέσματα όπως και το XM6.1 σε μία λειτουργία η οποία διατηρήθηκε ως είχε. Χρησιμοποιήθηκε ένα σύνολο 50 εικόνων με μία δυαδική μάσκα για την κάθε μία.

2. Συγκρίσεις με χρήση χάρτη περιοχών

Σε αυτήν την φάση έγινε έλεγχος της νέας λειτουργίας του VDE, της εξαγωγής περιγραφών για όλες τις περιοχές που δηλώνονται μέσω ενός χάρτη περιοχών. Η σύγκριση γίνεται μεταξύ της εισόδου μίας περιοχής που ορίζεται σε έναν χάρτη περιοχών και της αντίστοιχης εξόδου για την ίδια περιοχή άλλα ορισμένη σε μία δυαδική μάσκα. Χρησιμοποιήθηκε το VDE και για την εξαγωγή με δυαδική μάσκα λόγω του αποτελέσματος της πρώτης σύγκρισης (ενότητα 6.2). Έτσι, στην δεύτερη φάση συγκρίνουμε αν το αποτέλεσμα του VDE για μία περιοχή ορισμένη σε έναν χάρτη περιοχών δίνει το ίδιο με αυτό της ίδιας περιοχής ορισμένης σε δυαδική μάσκα. Θα ήταν αναμενόμενο το αποτέλεσμα να είναι το ίδιο. Οι συγκρίσεις αυτές γίνονται για να ελεγχθεί αν οι διαφοροποιήσεις που έχουν γίνει στον εσωτερικό τρόπο λειτουργίας της



Σχήμα 6.1: Δυαδικές μάσκες κατασκευασμένες από τον χάρτη περιοχών

εφαρμογής, όπως εκείνες περιγράφηκαν στην ενότητα 4.2 επηρεάζουν την σωστή εξαγωγή των περιγραφών. Με λίγα λόγια το αποτέλεσμα για μία περιοχή που ορίζεται σε χάρτη περιοχών θα πρέπει να είναι το ίδιο με εκείνο αν από την εικόνα και τον αντίστοιχο χάρτη περιοχών (σχήμα 4.1) κατασκευαστεί μία νέα εικόνα στο μέγεθος του bounding box της περιοχής κι επίσης κατασκευαστεί η αντίστοιχη δυαδική μάσκα στο ίδιο μέγεθος πάλι (σχήμα 4.2) και έπειτα αυτές δοθούν σαν είσοδο στο εκτελέσιμο VDE. Αυτό έχει φανεί και στην πράξη μετά την εκτέλεση του παραπάνω για μερικές εικόνες. Χρησιμοποιήθηκαν 180 εικόνες μαζί με έναν χάρτη περιοχών για την κάθε μία ο οποίος ήταν προϊόν ενός εργαλείου κατάτμησης. Σαν είσοδο δόθηκαν εικόνα και χάρτης περιοχών σε πρώτη φάση και στην συνέχεια απο τον χάρτη περιοχών κατασκευάστηκαν όλες οι αντίστοιχες δυαδικές μάσκες για όλες τις περιοχές που εκείνος περιείχε (σχήμα 6.1. Στην συνέχεια δόθηκαν μία μία σαν είσοδο με την αρχική εικόνα (σχήμα 4.1). Έτσι η σύγκριση γινόταν ανάμεσα στο αποτέλεσμα της πρώτης εκτέλεσης και για την 3η π.χ. περιοχή και της 2ης εκτέλεσης και για την δυαδική μάσκα που αντιστοιχούσε στην 3η περιοχή.

3. Συγκρίσεις με λίστα ορθογώνιων περιοχών

Εδώ τα πράγματα είναι πολύ παρόμοια με την παραπάνω περίπτωση μόνο που οι περιοχές είναι πάντα ορθογώνιες και μπορεί να είναι και αλληλοεπικαλυπτόμενες. Δηλώνονται μέσω συνεταγμένων σε αρχείο κειμένου όπως έχει περιγραφεί στην ενότητα 5.2. Χρησιμοποιήθηκαν οι ίδιες 180 εικόνες με ένα αρχείο κειμένου να ορίζει περιοχές, κοινό για όλες. Όπως και πριν, για να γίνει η σύγκριση δημιουργήθηκαν δυαδικές μάσκες αντίστοιχες των περιοχών που δι-

νονταν σαν είσοδο η κάθε μία με την κάθε εικόνα.

6.2 Αποτελέσματα

Συγκρίσεις με χρήση δυαδικών μασκών Δίνοντας σαν είσοδο στις δύο εφαρμογές τις 50 εικόνες που αναφέρθηκαν (6.1) και συγκρίνοντας μετέπειτα τα αποτελέσματα οι δύο έξοδοι αποδείχθηκαν πανομοιότυπες.

Συγκρίσεις με χρήση χάρτη περιοχών Μετά από την διαδικασία που περιγράφηκε στην ενότητα 6.1 έγινε σύγκριση των αποτελεσμάτων. Προέκυψαν πανομοιότυπα αποτελέσματα για τους εξής περιγραφείς: περιγραφέας κύριων χρωμάτων, περιγραφέας διάταξης χρώματος, κλιμακωτός περιγραφέας χρώματος, περιγραφέας ομοιογενούς υψής και περιγραφέας σχήματος με βάση το σχήμα.

Ο περιγραφέας ιστογράμματος ακμών είναι ένας ακόμα περιγραφέας για τον οποίο η απόκλιση ήταν πάρα πολύ μικρή. Για να υπολογιστεί η μέγιστη απόσταση που μπορούν δύο διαφορετικές περιοχές να έχουν όσον αφορά τον κάθε περιγραφέα υπολογίζονται σαν ελάχιστες τιμές οι ελάχιστες που εμφανίζονται ανάμεσα στις 1298 περιοχές και για κάθε ξεχωριστή τιμή του περιγραφέα και αντιστοίχως οι μέγιστες. Απόκλιση εντοπίστηκε σε ένα ποσοστό 0.28% στο σύνολο των κορυφών. Ως προς την μέγιστη απόσταση η μέση απόκλιση στο σύνολο όλων των περιοχών είναι 0.04%.

Για τον περιγραφέα σχήματος με βάση το περίγραμμα απόκλιση εντοπίστηκε μόνο ανάμεσα στα δύο πρώτα στοιχεία του περιγραφέα (εκκεντρότητα και κυκλικότητα) και καμία διαφορά όσον αφορά τις υψηλότερες κορυφές ή τις υπόλοιπες κορυφές. Η απόκλιση αυτή υπήρχε σε 58 περιοχές από τις 1298 και η μέση τιμή της ως προς την μέγιστη απόσταση (υπολογισμένη μόνο για τις δύο πρώτες τιμές) είναι 0.2%.

Ο περιγραφέας δομής χρώματος με την ίδια λογική του υπολογισμού της απόκλισης με τους παραπάνω περιγραφείς παρουσίασε μέση απόκλιση 3%. Το γεγονός της απόκλισης εδώ οφείλεται στον τρόπο με τον οποίο εξάγεται ο περιγραφέας (2.1), όπου από τις διαστάσεις της εικόνας εξαρτάται τόσο το μέγεθος του δομικού στοιχείου όσο και το βήμα με το οποίο αυτό σαρώνει την εικόνα. Στην προκειμένη περίπτωση έχουμε από την μία ολόκληρη την εικόνα και από την άλλη την υποεικόνα στο μέγεθος του περικλείοντος ορθογωνίου της κάθε φορά περιοχής. Αυτό σημαίνει στοιχεία δόμησης διαφορετικού μεγέθους, διαφορετικού βήματος και διαφορετικών αρχικών θέσεων από τις οποίες αρχίζουν να σαρώνουν την περιοχή ενδιαφέροντος. Στην πρώτη περίπτωση η σάρωση αρχίζει από το πάνω αριστερά σημείο της εικόνας κινούμενο από αριστερά προς τα δεξιά και από πάνω προς τα κάτω. Κάποια στιγμή αρχίζει να σαρώνει την περιοχή ενδιαφέροντος αρχίζοντας από τα εξωτερικά εικονοστοιχεία της. Έπειτα η σάρωση προχωρά και στα υπόλοιπα εικονοστοιχεία. Στην δεύτερη περίπτωση αν για παράδειγμα οι ακμές της περιοχής είναι και ακμές του περικλείοντος ορθογωνίου τότε το δομικό στοιχείο δεν έχει την δυνατότητα να σαρώσει μόνο τα εξωτερικά εικονοστοιχεία της περιοχής. Στο σχήμα 6.3 έχουμε μία εικόνα και θεωρούμε την περιοχή της θάλασσας από την οποία θέλουμε να γίνει η εξαγωγή και φαίνεται στην πρώτη εικόνα. Στην δεύτερη βλέπουμε μία θέση του δομικού στοιχείου, με την οποία σαρώνονται κάποια εικονοστοιχεία της περιοχής. Στην θέση αυτή όμως δεν μπορεί να βρεθεί το στοιχείο όταν έχουμε την υποεικόνα στο μέγεθος του περικλείοντος ορθογωνίου (το οποίο φαίνεται με κόκκινο χρώμα). Στο σχήμα 6.4 φαίνεται η αναπαράσταση όλου



Σχήμα 6.2: Τρόπος σάρωσης για τον περιγραφέα δομής χρώματος, κόκκινο πλαίσιο: περι-
κλείον ορθογώνιο, πράσινο τετράγωνο: δομικό στοιχείο



Σχήμα 6.3: Αρχική εικόνα-μάσκα και υποεικόνα με αντίστοιχη μάσκα

του ιστογράμματος για μία συγκεκριμένη περιοχή με τους δύο τρόπους (δυναδική μάσκα και χάρτης περιοχών). Με μία κανονικοποίηση τα δύο αυτά ιστογράμματα σχεδόν ταυτίζονται.

Το μέγεθος του δομικού στοιχείου είναι E και το βήμα με το οποίο εκείνο σαρώνει την εικόνα είναι K , τα οποία δίνονται από τις σχέσεις (6.2) και (6.1).

$$K = 2^p \quad (6.1)$$

$$E = 8 * K \quad (6.2)$$

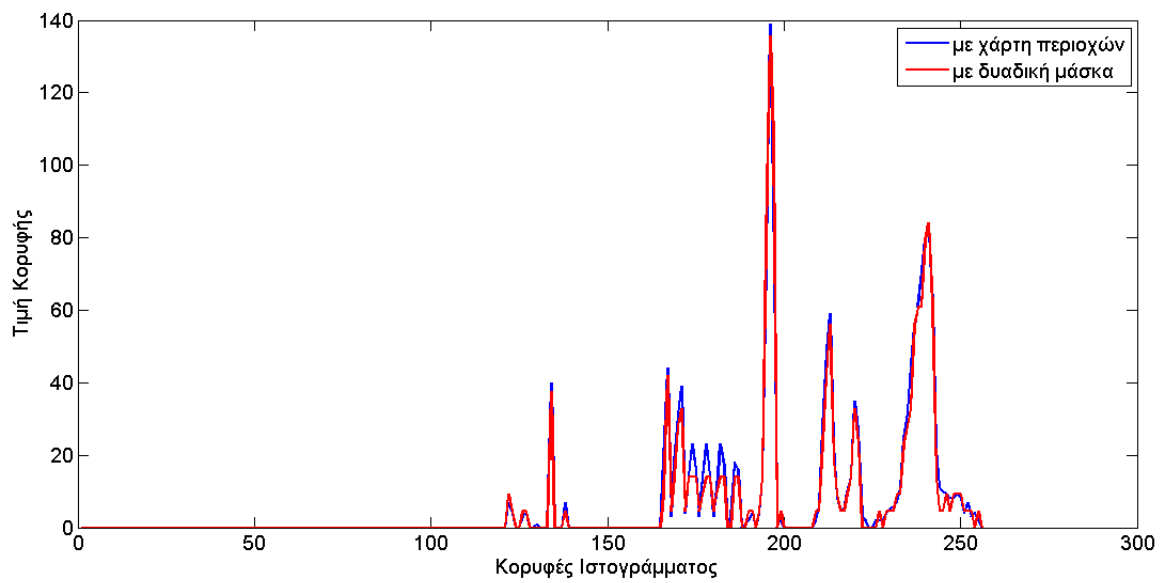
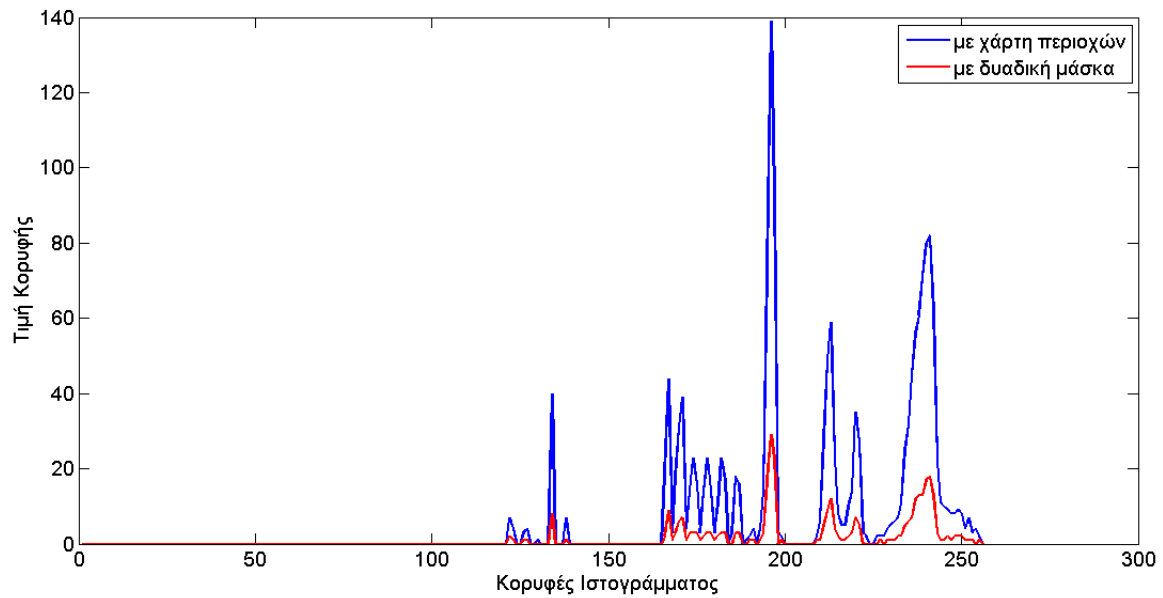
όπου $p = \text{round}(0.5 * \log_2(\text{πλάτος} * \text{ύψος})) - 8$ και αν $p < 0$, τότε $p = 0$.

Επίσης μετά την σάρωση όλης της εικόνας με το δομικό στοιχείο και την δημιουργία του ιστογράμματος, όλες οι τιμές του ιστογράμματος διαιρούνται με μία τιμή $S = ((\text{πλάτος} - E + K) / K) * ((\text{ύψος} - E + K) / K)$.

Το διάγραμμα του σχήματος 6.4 αντιστοιχεί στις εικόνες του σχήματος 6.3, όπου στην μία περίπτωση έχουμε την περιοχή πάνω σε όλη την εικόνα και στην άλλη την υποεικόνα και αντίστοιχη μάσκα που θα δημιουργηθούν στην μνήμη αν η ίδια περιοχή δοθεί με χάρτη περιοχών. Οι διαστάσεις της αρχικής εικόνας είναι 800×600 , το $p = 1$ και τελικά $K = 2$, $E = 16$ και $S = 115149$. Ενώ για την υποεικόνα η οποία έχει διαστάσεις 515×45 τελικά είναι $p = 0$, $K = 1$, $E = 8$ και $S = 19304$. Ο λόγος των δύο S είναι 5.9 και ο παράγοντας με τον οποίο πολλαπλασιάστηκε το ένα ιστόγραμμα στο σχήμα 6.4 ήταν 5, πολύ κοντά δηλαδή στον προαναφερθέντα λόγο. Η μικρή διαφορά μεταξύ 5-5.9 οφείλεται στο γεγονός του τρόπου σάρωσης με μικρές διαφορές όπως περιγράφηκε στην προηγούμενη παράγραφο, αλλά και στο διαφορετικό μέγεθος δομικού στοιχείου και βήματος.

Συνολικές Αποκλίσεις

Στον πίνακα 6.1 φαίνονται τα αποτελέσματα του ελέγχου όσον αφορά τον αριθμό των περιοχών για



Σχήμα 6.4: Περιγραφείας δομής χρώματος με δυαδική μάσκα και με χάρτη περιοχών για την ίδια περιοχή ενδιαφέροντος(πάνω) , κανονικοποιημένο (κάτω)

τις οποίες ο κάθε περιγραφέας παρουσίασε απόκλιση, το ποσοστό αυτών των περιοχών ως προς τον συνολικό αριθμό των περιοχών, την μέση ποσοστιαία απόκλιση ως προς τα μέγιστα μέτρα και το ποσοστό των πεδίων του κάθε περιγραφέα τα οποία είχαν απόκλιση (για παράδειγμα ο περιγραφέας ιστογράμματος ακμών έχει 80 πεδία).

Περιγραφέας	Λανθασμένες Περιοχές		Συνολικό Ποσοστό Λάθους	Ποσοστό Λανθασμένων Πεδίων
	DC	0	0 %	0 %
SC	0	0 %	0 %	0 %
CL	0	0 %	0 %	0 %
CST	1296	99.8 %	3 %	15.9 %
HT	0	0 %	0 %	0 %
EH	93	7 %	0.04 %	0.28 %
RS	0	0 %	0 %	0 %
CS	58	4 %	0.2 %	0.2 %

Πίνακας 6.1: Αποτελέσματα ελέγχου με χάρτη περιοχών

Συγκρίσεις με συντεταγμένες ορθογώνιων περιοχών Σε αυτό το μέρος του ελέγχου προέκυψαν πανομοιότυπα αποτελέσματα για τους εξής περιγραφείς: περιγραφέας κύριων χρωμάτων, κλιμακωτός περιγραφέας χρώματος, περιγραφέας διάταξης χρώματος, περιγραφέας ιστογράμματος ακμών, περιγραφέας ομοιογενούς υφής, περιγραφέας σχήματος με βάση το περίγραμμα και περιγραφέας σχήματος με βάση το σχήμα. Για τον ένα περιγραφέα ο οποίος είχε αποκλίσεις, εκείνες ήταν της ίδιας φύσεως και τάξεως με την διαδικασία συγκρίσεων με χρήση χάρτη περιοχών. Οι αποκλίσεις υπολογισμένες επί τις % με την ίδια μέθοδο όπως και παραπάνω είναι 2.3% για τον περιγραφέα δομής χρώματος.

6.3 Σχολιασμός Αποτελεσμάτων

Αρχικά πρέπει να αναφερθεί ότι για να γίνεται σωστή εξαγωγή των περιγραφέων εντοπίστηκαν και διορθώθηκαν bugs σε 3 περιγραφείς. Αυτοί οι περιγραφείς είναι: ο περιγραφέας κύριων χρωμάτων, ο περιγραφέας διάταξης χρώματος και περιγραφέα σχήματος με βάση το σχήμα. Στην έκδοση 6.1 του ΧΜ η εξαγωγή για τους παραπάνω περιγραφείς γινότανε με λανθασμένο τρόπο, ο οποίος εντοπίστηκε και τελικά διορθώθηκε. Οι διορθώσεις αυτές έχουν τελικά σαν αποτέλεσμα την πανομοιότυπη έξοδο των δύο τρόπων εξαγωγής περιγραφέα από μία συγκεκριμένη περιοχή, με δυαδική μάσκα και με χάρτη περιοχών.

Επιπλέον χρήσιμες αλλαγές ήταν εκείνες οι οποίες έγιναν για τον περιγραφέα ομοιογενούς υφής και τον περιγραφέα σχήματος με βάση το περίγραμμα. Είναι αλλαγές που αφορούν την εξαγωγή από μικρές περιοχές που ορίζονται σε χάρτη περιοχών ή και σε αρχείο με ορθογώνιες περιοχές. Περιοχές με την μία διάσταση μικρότερη από 128 εικονοστοιχεία για τον πρώτο περιγραφέα ή 8 εικονοστοιχεία για τον δεύτερο. Οι αλλαγές αυτές δίνουν την δυνατότητα η εξαγωγή από τέτοιων διαστάσεων περιοχές να είναι εφικτή.

Τελικά οι αποκλίσεις παρουσιάζονται για 3 περιγραφείς: τον περιγραφέα δομής χρώματος, τον περιγραφέα ιστογράμματος ακμών και τον περιγραφέα σχήματος με βάση το σχήμα. Οι αποκλίσεις για τους δύο τελευταίους όπως φάνηκε και στον πίνακα 6.1 είναι πολύ μικρές και ανεπαίσθητες προφανώς σαν διαφορά μιας οπτικής περιγραφής. Σημειώνεται ακόμα ότι οι αποκλίσεις για τους 2 αυτούς περιγραφείς δεν εμφανίστηκαν για ορθογώνιες περιοχές. Όσο για την απόκλιση του περιγραφέα δομής χρώματος σχολιάστηκε αρκετά και φαίνεται πως μία κανονικοποιημένη χρήση των περιγραφέων αυτών θα ήταν αποτελεσματική.

Μέρος ΙΙΙ
Εντοπισμός Εγνοιών

Κεφάλαιο 7

Εισαγωγή

7.1 Περιγραφή του Προβλήματος

Ο εντοπισμός εννοιών υψηλού επιπέδου (high-level concept detection)[13][14][15] σε βίντεο βοηθά στην οργάνωση, δεικτοδότηση, αναζήτηση και ανάκτηση του πολυμεσικού υλικού που συνεχώς αυξάνεται. Είναι ένας τομέας στον οποίο πολλή έρευνα γίνεται αλλά παραμένει ένα πρόβλημα που δεν μπορεί να επιλυθεί πλήρως, δεν μπορεί δηλαδή να βρεθεί μία μέθοδος που να αντιμετωπίζει κάθε περίπτωση. Όπως και κάθε τυπικό πρόβλημα αναγνώρισης έχει δύο όψεις. Η πρώτη είναι η εξαγωγή διαφόρων οπτικών χαρακτηριστικών από ένα βίντεο, όπως χρώμα, υφή, κίνηση και ήχος, μία διαδικασία η οποία συνήθως αποκαλείται εξαγωγή χαρακτηριστικών χαμηλού επιπέδου (low-level feature extraction) και έπειτα σχηματισμός μίας περιγραφής συνδυάζοντας τα. Η δεύτερη είναι η μέθοδος που χρησιμοποιείται για να αντιστοιχήσει αυτές τις περιγραφές χαμηλού επιπέδου σε κάποια έννοια υψηλού επιπέδου, ένα πρόβλημα το οποίο συνήθως αποκαλείται “Σημασιολογικό Κενό” (Semantic Gap)[16]. Πολλές προσεγγίσεις έχουν προταθεί για να γεφυρώσουν αυτό το κενό επιτρέποντας έτσι την εξαγωγή εννοιών υψηλού επιπέδου από πολυμεσικά έγγραφα, χρησιμοποιώντας και συνδυάζοντας ετερογενή οπτικοακουστικά χαρακτηριστικά. Έτσι από την μία έχει γίνει ανάπτυξη πολλών περιγραφών που σκοπεύουν στην εξαγωγή των χαρακτηριστικών χρώματος, υφής, κίνησης, σχήματος και ήχου των οπτικοακουστικών πολυμεσικών αρχείων. Από την άλλη έχει γίνει πολλή έρευνα όπου οι περιγραφές χαμηλού επιπέδου αντιστοιχίζονται σε έννοιες με τη χρήση πολλών τεχνικών εκμάθησης όπως νευρωνικά δίκτυα, ασαφή συστήματα, μηχανές διανυσμάτων στήριξης (Support Vector Machines - SVM) κ.α. Ο όρος “έννοια υψηλού επιπέδου” χρησιμοποιείται για να αποδώσει τον όρο “high-level concept”, εννοώντας εικόνες που απεικονίζουν κάτι συγκεκριμένο όπως *ουρανός*, *βλάστηση* κ.α. Στα επόμενα ο όρος “έννοια” θα χρησιμοποιείται για αναφορά σε τέτοιου είδους περιοχές και την απεικόνιση τους ή όχι σε εικόνες.

7.2 TRECVID

Ιδιαίτερη προσοχή αξίζει να δοθεί στην προσπάθεια που γίνεται για την αξιολόγηση των τεχνικών ανάκτησης πληροφορίας από το TREC. Ειδική ενότητα του TREC είναι το TRECVID που αφορά ανάκτηση πληροφορίας από βίντεο, συμμετοχή στο οποίο είναι και η δουλειά με τον αλγόριθμο που

παρουσιάζεται στο κεφάλαιο 8. Το TREC (Text REtrieval Conference) είναι ένα συνέδριο με θέμα την ανάκτηση κειμένου χρηματοδοτούμενο από το εθνικό ινστιτούτο προτύπων και τεχνολογίας (National Institute of Standards and Technology - NIST) των ΗΠΑ. Ξεκίνησε το 1992 με σκοπό να υποστηρίξει την κοινότητα ανάκτησης πληροφορίας παρέχοντας την απαραίτητη υποδομή για αξιολόγηση μεθοδολογιών ανάκτησης κειμένου ευρείας κλίμακας. Επιδιώξη του είναι να ενθαρρύνει την έρευνα πάνω στην ανάκτηση πληροφορίας βασισμένη σε μεγάλες συλλογές ελέγχου, να βελτιώσει την επικοινωνία ανάμεσα στην βιομηχανική, ακαδημαϊκή και κυβερνητική κοινότητα ανοίγοντας ανοιχτή συζήτηση για την ανταλλαγή ερευνητικών ιδεών. Επίσης, επιδιώκει να επιταχύνει την μετατροπή της τεχνολογίας από τα εργαστήρια έρευνας σε εμπορικά προϊόντα, αναδεικνύοντας ουσιαστικές βελτιώσεις σε μεθοδολογίες ανάκτησης για προβλήματα πραγματικού κόσμου, αλλά και να αυξήσει τη διαθεσιμότητα των κατάλληλων τεχνικών αξιολόγησης για χρήση από βιομηχανίες και ακαδημαϊκούς, συμπεριλαμβανομένης της ανάπτυξης νέων τεχνικών αξιολόγησης καλύτερα εφαρμόσιμων στα τωρινά συστήματα. Για κάθε συνέδριο, το NIST παρέχει ένα σύνολο ελέγχου από έγγραφα και ερωτήματα. Οι συμμετέχοντες χρησιμοποιούν πάνω σε αυτά τα δεδομένα τα δικά τους συστήματα ανάκτησης και επιστρέφουν μία λίστα ανακτημένων εγγράφων. Ύστερα τα ανακτημένα έγγραφα κρίνονται όσον αφορά την ορθότητα τους και γίνεται αξιολόγηση των αποτελεσμάτων. Ο κύκλος του συνεδρίου τελειώνει με μία ανοιχτή συζήτηση όπου οι συμμετέχοντες μοιράζονται τις εμπειρίες τους.

Το TREC το 2001 άρχισε να χρηματοδοτεί και μια νέα ενότητα με βίντεο, αφιερωμένη στην έρευνα στην αυτόματη κατάτμηση, δεικτοδότηση και ανάκτηση ψηφιακού βίντεο βασισμένη στο περιεχόμενο. Ξεκινώντας το 2003 αυτή η ενότητα έγινε ένα ανεξάρτητο έργο αξιολόγησης (benchmark) και ονομάστηκε “TRECVID” [17]. Κύριος στόχος του είναι να προωθήσει την πρόοδο σε ανάκτηση με βάση το περιεχόμενο από ψηφιακό βίντεο μέσω ανοιχτής και βασισμένης σε συγκεκριμένα μέτρα αξιολόγησης. Πρόκειται για μία αξιολόγηση εργαστηριακής τεχνοτροπίας που προσπαθεί να μοντελοποιήσει καταστάσεις πραγματικού κόσμου ή ενδεικτικές εργασίες που εμπλέκονται σε τέτοιου είδους καταστάσεις. Το 2006 συμπληρώθηκε για το TRECVID ο δεύτερος κύκλος αφιερωμένος σε ψηφιακό βίντεο ειδήσεων στα αγγλικά, αραβικά και κινέζικα. Συμπληρώθηκαν επίσης δύο χρόνια πιλοτικών μελετών στην εκμετάλλευση ακατέργαστου (raw) υλικού βίντεο. Το 2007 εισήλθε σε μία σχετική αλλά νέα και δύσκολη περιοχή. Μετά από τα τέσσερα χρόνια σε βίντεο ειδήσεων το TRECVID 2007 ελέγχει τις τρεις θεμελιώδεις εργασίες του (καθορισμός συνόρου στιγμιότυπου, εξαγωγή χαρακτηριστικών υψηλού επιπέδου, αναζήτηση) σε βίντεο που προέρχεται από αρχείο με τηλεημερίδες, επιστημονικά νέα, ντοκιμαντέρ, εκπαιδευτικά προγράμματα κ.α. με σκοπό την εφαρμογή προϋπαρχουσών τεχνολογιών και συστημάτων σε νέα είδη δεδομένων.

Το TRECVID παρέχει αρχικά στους συμμετέχοντες ένα σύνολο από βίντεο τα οποία είναι χωρισμένα σε πλάνα (shots), αυτό είναι το σύνολο ανάπτυξης (development set), όπως αποκαλείται από το NIST. Οι συμμετέχοντες στην συνέχεια πρέπει να δημιουργήσουν συνολικούς σχολιασμούς για τα πλάνα όσον αφορά τις έννοιες με τις οποίες ασχολείται το TRECVID. Αυτό σημαίνει ότι πρέπει να σημειωθεί το κατά πόσο ένα πλάνο απεικονίζει κάποια συγκεκριμένη έννοια ή όχι. Στην συνέχεια βάση αυτού του συνόλου και των σχολιασμών τους δημιουργούν συστήματα τα οποία πρέπει να είναι ικανά να αποφανθούν για το επόμενο σύνολο αν οι συγκεκριμένες έννοιες απεικονίζονται ή όχι. Το σύνολο αυτό είναι το σύνολο ελέγχου αποτελούμενο από βίντεο χωρισμένα και πάλι σε πλάνα. Οι συμμετέχοντες χρησιμοποιώντας τα συστήματα που ανέπτυξαν με τις διάφορες προσεγγίσεις ο καθένας, παραδίδουν αποτελέσματα στο NIST για το αν το κάθε πλάνο απεικονίζει τις

επιλεγμένες έννοιες και ποιες από αυτές απεικονίζει. Τέλος το TRECVID αξιολογεί αντικειμενικά τις μεθόδους αυτές. Περισσότερες πληροφορίες μπορούν να βρεθούν στην ιστοσελίδα του TRECVID <http://www-nlpir.nist.gov/projects/trecvid/>.

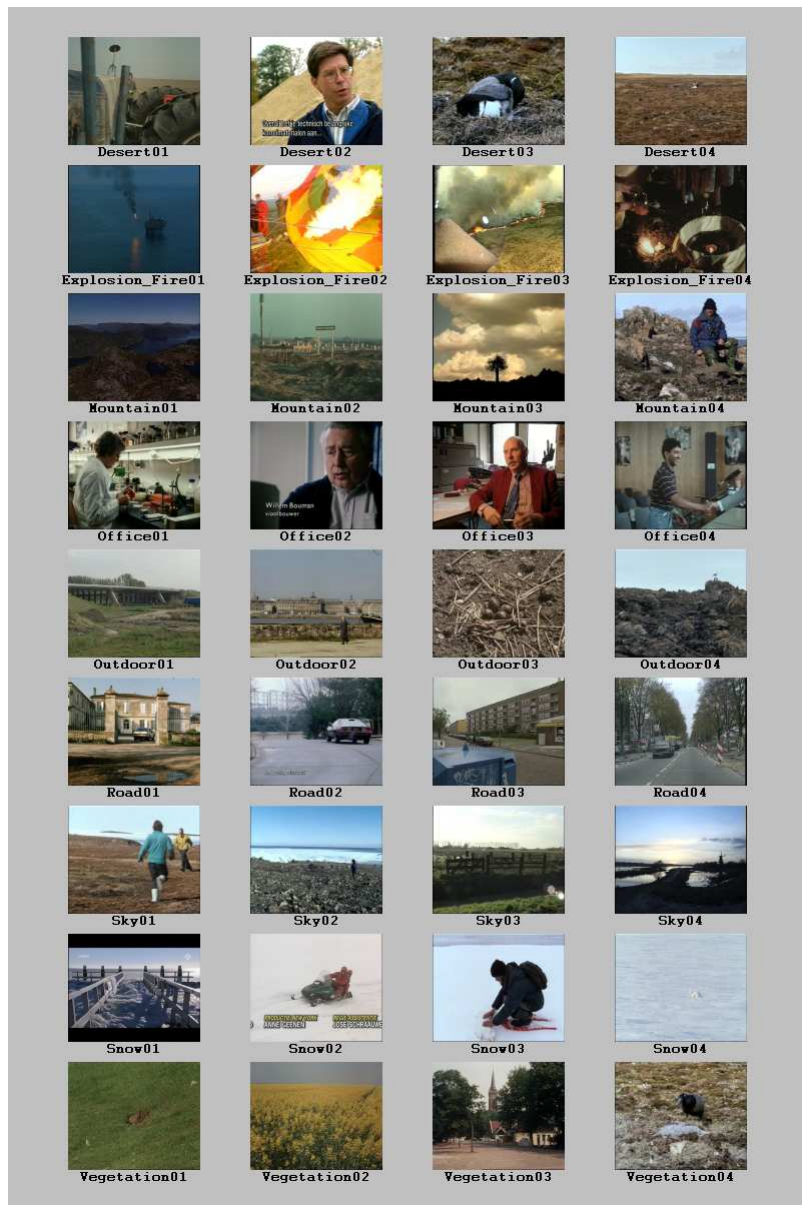
Οι έννοιες με τις οποίες ασχολείται το TRECVID είναι οι:

Sports	Government-Leader
Entertainment	Corporate-Leader
Weather	Police_Security
Court	Military
Office	Prisoner
Meeting	Animal
Studio	Computer_TV-screen
Outdoor	Flag-US
Building	Airplane
Desert	Car
Vegetation	Bus
Mountain	Truck
Road	Boat_Ship
Sky	Walking_Running
Snow	People-Marching
Urban	Explosion_Fire
Waterscape-Waterfront	Natural-Disaster
Crowd	Maps
Face	Charts
Person	

Σε αυτή την διπλωματική εργασία επιλέγονται 9 από τις έννοιες προς εντοπισμό οι οποίες είναι *vegetation* (βλάστηση), *snow* (χιόνι), *mountain* (βουνό), *road* (δρόμος), *office* (γραφείο), *outdoor* (εξωτερικός χώρος), *explosion_fire* (έκρηξη-φωτιά), *desert* (έρημος) και *sky* (ουρανός). Στο σχήμα 7.1 φαίνεται από ένα παράδειγμα εικόνας που απεικονίζει την κάθε έννοια. Από τις εικόνες αυτές αναδεικνύεται επίσης και η δυσκολία του συνόλου των βίντεο αλλά και οι έννοιες με τις οποίες ασχολείται το TRECVID, καθώς υπάρχουν εικόνες από ίδιες έννοιες με καθόλου όμοια οπτικά χαρακτηριστικά.

7.3 Περιορισμοί

Ένα κίνητρο της δουλειάς αυτής είναι και ο αυξανόμενος αριθμός των συνολικών σχολιασμών των εικόνων. Συνολικός σχολιασμός σημαίνει πως είναι απλώς γνωστό το αν η εικόνα απεικονίζει μία έννοια, χωρίς να είναι γνωστό ποια περιοχή αντιστοιχεί στην έννοια, σε αντίθεση με τους σχολιασμούς περιοχών όπου υπάρχει αντιστοιχία ανάμεσα σε συγκεκριμένες περιοχές και έννοιες. Αυτό συμβαίνει καθώς είναι πολύ πιο εύκολο για ένα χρήστη να σχολιάσει συνολικά την εικόνα δηλώνοντας πως περιέχει κάποιες συγκεκριμένες έννοιες, παρά να σημειώνει την κάθε περιοχή, καθώς κάθε εικόνα μπορεί να αποτελείται από πολλές περιοχές. Έτσι συλλογές εικόνων με σχολιασμούς σε επίπεδο περιοχών σπανίζουν.



Σχήμα 7.1: Παραδείγματα εικόνων που απεικονίζουν τις 9 έννοιες προς εντοπισμό

Επίσης για το συγκεκριμένο πρόβλημα χρειάζονται γενικές τεχνικές οι οποίες μπορούν να εφαρμοστούν για διαφορετικές έννοιες, χωρίς να αντιμετωπίζεται η κάθε έννοια με συγκεκριμένο τρόπο. Για παράδειγμα απαιτείται μία τεχνική η οποία μπορεί με τον ίδιο τρόπο να εφαρμοστεί για τις έννοιες *θάλασσα* και *ουρανός*.

Τέλος υπάρχει η ανάγκη για τεχνικές που θα μπορούν να εφαρμοστούν σε μεγάλο όγκο δεδομένων όπως και εκείνος του TRECVID. Αυτό σημαίνει ότι ο χρόνος περάτωσης της θα πρέπει να είναι

λογικός ακόμα και αν μιλάμε για πολλές ώρες βίντεο το οποίο θα χρησιμοποιηθεί για την εκπαίδευση των συστημάτων αλλά και για πολλές ώρες των βίντεο ελέγχου για τα οποία τα συστήματα θα αποφανθούν για την απεικόνιση των εννοιών.

7.4 Άλλες Τεχνικές

Ένα από τα πιο γνωστά συστήματα ανάλυσης και ανάκτησης πολυμεσικού υλικού που παρουσιάζεται στο [18], χρησιμοποιεί πολυτροπικές τεχνικές εκμάθησης για την μοντελοποίηση των σημασιολογικών εννοιών. Επίσης μία προσέγγιση βασισμένη σε περιοχές για την ανάκτηση περιεχομένου χρησιμοποιώντας την Λανθάνουσα Σημασιολογική Δεικτοδότηση (Latent Semantic Indexing) παρουσιάζεται στο [19], όπου φαίνεται να είναι κρίσιμη η επιλογή καθολικών ή τοπικών οπτικών χαρακτηριστικών. Έτσι εφαρμόζοντας ένα πλέγμα πάνω στην εικόνα χαρακτηριστικά χρώματος και υφής εξάγονται τοπικά [20]. Μία ακόμα τεχνική βασισμένη σε περιοχές παρουσιάζεται στο [21], η οποία χρησιμοποιεί γνώση στην μορφή οντολογίας. MPEG-7 οπτικά χαρακτηριστικά εξάγονται και συνδυάζονται ώστε να εντοπιστούν τελικά σημασιολογικές έννοιες.

Επιπλέον μία τεχνική με την χρήση υβριδικού θησαυρού παρουσιάζεται στο [22], όπου γίνεται αναγνώριση και ταυτοποίηση των αντικειμένων μέσα σε βίντεο ειδήσεων, με έμφαση στον εντοπισμό προσώπων και λογοτύπων καναλιών. Ένα λεξικό, χρησιμοποιείται σε μία προσέγγιση για ένα αλληλεπιδρών σύστημα ανάκτησης που παρουσιάζεται στο [23], όπου ο πυρήνας είναι ο αυτόματος εντοπισμός ενός λεξικού 101 εννοιών. Μία άλλη τεχνική για αναγνώριση υφής και αντικειμένων παρουσιάζεται στο [24], η οποία χρησιμοποιεί σταθερά ως προς την κλιμάκωση τοπικά χαρακτηριστικά σε συνδυασμό με έναν ταξινομητή. Τα SVM έχουν χρησιμοποιηθεί για ταξινόμηση εικόνων βασισμένη σε ιστογράμματα [25] και για εντοπισμό εννοιών όπως γκολ, κίτρινη κάρτα και αλλαγή σε υλικό με ποδοσφαιρικό περιεχόμενο [26]. Τέλος ένας αυτο-οργανούμενος χάρτης (Self-Organized Map - SOM) που χρησιμοποιεί MPEG-7 χαρακτηριστικά παρουσιάζεται στο [27].

Κεφάλαιο 8

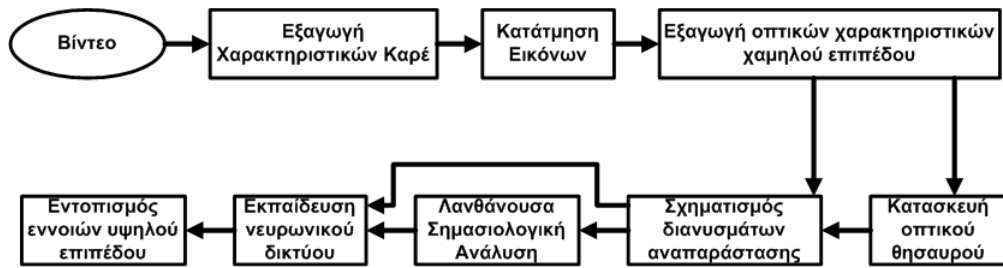
Εντοπισμός Εννοιών με Χρήση Οπτικού Θησαυρού

Το πρόβλημα του εντοπισμού εννοιών σε βίντεο σε αυτή την εργασία προσεγγίζεται με τον παρακάτω αλγόριθμο. Κάθε βίντεο χωρίζεται σε πλάνα (shots) και το κάθε πλάνο αναπαρίσταται από ένα χαρακτηριστικό καρέ (keyframe). Έτσι, το πρώτο βήμα είναι η εξαγωγή των χαρακτηριστικών καρέ από τα πλάνα.

Κάθε βίντεο συμβολίζεται με v_i και V είναι το σύνολο όλων των βίντεο. Επίσης s_i κάθε πλάνο, k_i το κάθε χαρακτηριστικό καρέ και S και K τα σύνολα όλων των πλάνων και χαρακτηριστικών καρέ αντίστοιχα. $S(v_i) \subset S$ είναι το σύνολο των πλάνων του βίντεο v_i . Το k_i είναι χαρακτηριστικό καρέ που αντιστοιχεί στο πλάνο s_i . Στην συνέχεια οι εικόνες που αντιστοιχούν σε αυτά τα χαρακτηριστικά καρέ δίνονται σαν είσοδο σε ένα εργαλείο κατάτμησης με βάση το χρώμα. Οι περιοχές που προκύπτουν είναι οι περιοχές για κάθε εικόνα. Κάθε περιοχή συμβολίζεται με r_i , R είναι το σύνολο όλων των περιοχών και $R(k_i) \subset R$ είναι το σύνολο των περιοχών του χαρακτηριστικού καρέ k_i . Συνοπτικά:

$$\begin{aligned}v_i &\in V, & i &= 1 \dots N_V \\s_i &\in S, & i &= 1 \dots N_S \\k_i &\in K, & i &= 1 \dots N_K = N_S \\r_i &\in R, & i &= 1 \dots N_R \\S(v_i) &\subset S, & i &= 1 \dots N_V \\R(k_i) &\subset R, & i &= 1 \dots N_K\end{aligned}$$

Οπτικά χαρακτηριστικά χαμηλού επιπέδου εξάγονται από όλες τις περιοχές κάθε εικόνας, οι οποίες έχουν προκύψει με κατάτμηση με βάση το χρώμα. Τα χαρακτηριστικά αυτά είναι MPEG-7 περιγραφείς χρώματος και υφής. Συνδυάζοντας τους περιγραφείς αυτούς σχηματίζεται ένα διάνυσμα χαρακτηριστικών (feature vector) για κάθε περιοχή. Χρησιμοποιώντας έπειτα ένα μεγάλο αριθμό από χαρακτηριστικά καρέ κι εφαρμόζοντας μία μέθοδο συσταδοποίησης (clustering), κατασκευάζεται ένας θησαυρός περιοχών (region thesaurus) ή αλλιώς οπτικός θησαυρός (visual thesaurus), ο οποίος περιέχει όλους τους τύπους περιοχής (regions types). Ο θησαυρός αυτός λειτουργεί σαν μια “βάση γνώσης” και διευκολύνει την συσχέτιση χαρακτηριστικών χαμηλού με υψηλού επιπέδου.



Σχήμα 8.1: Αλγόριθμος εντοπισμού εννοιών υψηλού επιπέδου

Για τους τύπους περιοχής επιλέγονται εκείνες οι περιοχές οι οποίες βρίσκονται πιο κοντά στα κέντρα (centroids) της κάθε συστάδας και κάθε τύπος αναπαρίσταται από ένα συγχωνευμένο διάνυσμα (διάνυσμα χαρακτηριστικών). Μετρώντας τις αποστάσεις των περιοχών της εικόνας από τους τύπους περιοχής του θησαυρού σχηματίζεται ένα διάνυσμα αναπαράστασης το οποίο αποτυπώνει την σημασιολογία της εικόνας. Έπειτα με την εφαρμογή της Λανθάνουσας Σημασιολογικής Ανάλυσης (Latent Semantic Analysis - LSA) μειώνεται η διάσταση των διανυσμάτων, καθώς το πρόβλημα οδηγείται σε έναν χώρο εννοιών (concept space). Τέλος ένα νευρωνικό δίκτυο εκπαιδεύεται να εντοπίζει την κάθε έννοια, βασισμένο στις τιμές των εικόνων των διανυσμάτων αναπαράστασης στο χώρο των εννοιών ή των ίδιων των διανυσμάτων αναπαράστασης εάν δεν έχει εφαρμοστεί LSA. Στις παρακάτω υποενότητες περιγράφονται αναλυτικά τα βήματα του αλγορίθμου που περιγράφηκε. Στο σχήμα 8.1 φαίνεται συνοπτικά όλος ο αλγόριθμος.

8.1 Εξαγωγή Χαρακτηριστικών Χαμηλού Επιπέδου

Για την αναπαράσταση των οπτικών χαρακτηριστικών χαμηλού επιπέδου ενός χαρακτηριστικού καρέ, χρησιμοποιούνται χαρακτηριστικά χρώματος και υφής. Για τα χαρακτηριστικά χρώματος χρησιμοποιούνται και οι τέσσερις περιγραφείς χρώματος που έχουν αναφερθεί στην ενότητα 2.1, δηλαδή κλιμακωτός περιγραφέας χρώματος, περιγραφέας κύριων χρωμάτων, περιγραφέας δομής χρώματος και περιγραφέας διάταξης χρώματος. Για τις ιδιότητες της υφής χρησιμοποιείται ο περιγραφέας ιστογράμματος ακμών. Το κάθε χαρακτηριστικό καρέ χωρίζεται σε περιοχές με ένα εργαλείο κατάτμησης με βάση το χρώμα και ύστερα για κάθε περιοχή εξάγονται 5 διανύσματα, ένα για κάθε περιγραφέα όπως αυτά έχουν περιγραφεί από τις εξισώσεις (2.2), (2.11), (2.13), (2.9) και (2.17). Συγκεκριμένα το διάνυσμα για τον περιγραφέα κύριων χρωμάτων περιέχει μόνο το χρώμα με το μεγαλύτερο ποσοστό εμφάνισης. Τα διανύσματα αυτά συγχωνεύονται σε ένα μοναδικό διάνυσμα, το οποίο χαρακτηρίζει την περιοχή και θα ονομάζεται διάνυσμα χαρακτηριστικών (feature vector). Το διάνυσμα χαρακτηριστικών που αντιστοιχεί σε μία περιοχή $r_i \in R$ δίνεται από την σχέση (8.1).

$$f_i = f(r_i) = [CLD(r_i), DCD(r_i), CSTD(r_i), SCD(r_i), EHD(r_i)], \quad r_i \in R \quad (8.1)$$

όπου $CLD(r_i)$ ο περιγραφέας διάταξης χρώματος για την περιοχή r_i , $DCD(r_i)$ ο περιγραφέας κύριων χρωμάτων για την περιοχή r_i κ.ο.κ. Το κάθε διάνυσμα χαρακτηριστικών συμβολίζεται με f_i και F είναι το σύνολο όλων των διανυσμάτων χαρακτηριστικών. Δηλαδή $f_i \in F, \quad i = 1 \dots N_F = N_R$.

Κάθε ένα από τα διανύσματα των περιγραφών που χρησιμοποιούνται έχει συγκεκριμένη διάσταση. Οι διαστάσεις αυτές συμβολίζονται με N_{CLD} , N_{DCD} , N_{CSTD} , N_{SCD} και N_{EHD} για τα διανύσματα του περιγραφέα διάταξης χρώματος, τον περιγραφέα κύριων χρωμάτων, τον περιγραφέα ιστογράμματος ακμών, τον περιγραφέα δομής χρώματος και τον κλιμακωτό περιγραφέα χρώματος αντίστοιχα. Επίσης η συνολική διάσταση του διανύσματος χαρακτηριστικών συμβολίζεται με N_{DF} . Για τις διαστάσεις αυτές ισχύει:

$$\begin{aligned} N_{CLD} &= 18 \\ N_{DCD} &= 4 \\ N_{CSTD} &= 256 \\ N_{SCD} &= 256 \\ N_{EHD} &= 80 \\ N_{DF} &= N_{CLD} + N_{DCD} + N_{CSTD} + N_{SCD} + N_{EHD} = 614 \end{aligned}$$

8.2 Κατασκευή Οπτικού Θησαυρού

Σαν αρχικές παρατηρήσεις από το σύνολο των χαρακτηριστικών καρέ μπορεί να επισημανθεί η ύπαρξη πολλών παρόμοιων χαρακτηριστικών καρέ (καθώς τα καρέ αυτά προέρχονται από τα ίδια βίντεο με μία μόνο μικρή διαφορά μεταξύ τους) αλλά και παρόμοιων περιοχών. Επίσης το γεγονός ότι εικόνες με την ίδια έννοια αποτελούνται από παρόμοιες περιοχές, δηλαδή έχουν παρόμοιες περιγραφές χαμηλού επιπέδου.

Είναι επιθυμητό να οργανωθούν όλες οι περιοχές όλων των εικόνων σε μία “βάση” με σκοπό να περιγράψουμε μία εικόνα σαν σύνολο από περιοχές. Έτσι εφαρμόζουμε έναν αλγόριθμο συσταδοποίησης (clustering) σε όλες τις περιοχές ή σε ένα υποσύνολο αυτών. Μία πρώτη παρατήρηση είναι ότι κάθε συστάδα (cluster) μπορεί να περιέχει περιοχές όχι απαραίτητα από την ίδια έννοια. Επίσης περιοχές από το ίδιο χαρακτηριστικό υψηλού επιπέδου (έννοια) μπορεί να έχουν καταταχθεί σε διαφορετικές συστάδες. Για παράδειγμα περιοχές που ανήκουν στη έννοια *βλάστηση* μπορεί να υπάρχουν σε περισσότερες από μία συστάδες διαφέροντας για παράδειγμα στο χρώμα των φύλλων των δέντρων. Κάθε στιγμιότυπο θα αναπαρίσταται από την περιοχή εκείνη που είναι πιο κοντά στο κέντρο (centroid) μίας συστάδας και θα αναφέρεται σαν “τύπος περιοχής”. Συγκεκριμένα θα αναπαρίσταται από το διάνυσμα χαρακτηριστικών της περιοχής αυτής, καθώς και ο αλγόριθμος συσταδοποίησης εφαρμόζεται πάνω στα διανύσματα χαρακτηριστικών του συνόλου των περιοχών. Επιπλέον σε μία συστάδα που περιλαμβάνει στιγμιότυπα της έννοιας *θάλασσα* είναι δυνατόν αυτά τα στιγμιότυπα να έχουν αναμειχθεί με εκείνες της έννοιας *ουρανός*, εφόσον μία τυπική περιοχή ουρανού είναι σχεδόν πανομοιότυπη με συγκεκριμένες περιοχές θάλασσας. Τελικά ο θησαυρός δίνεται από την σχέση (8.2).

$$T = \{w_i, \quad i = 1 \dots N_T\}, \quad w_i \subset R \quad (8.2)$$

$$\bigcup_i w = R, \quad i = 1 \dots N_T \quad (8.3)$$

$$\bigcap_{i,j} w = \emptyset, \quad i \neq j \quad (8.4)$$

όπου N_T ο αριθμός των τύπων περιοχής του θησαυρού αντιστοίχως και των clusters και w_i είναι το i -οστό cluster, το οποίο είναι ένα σύνολο από περιοχές εικόνων που ανήκουν στο R , όπως φαίνεται και στην σχέση (8.2). Τύπος περιοχής ή αλλιώς “οπτική λέξη” (visual word) του οπτικού θησαυρού ονομάζεται το κέντρο του κάθε cluster. Επίσης από τις σχέσεις (8.3) και (8.4) φαίνεται ότι η ένωση όλων των συστάδων δίνει όλο το σύνολο R , αυτό εφόσον όλες οι περιοχές έχουν χρησιμοποιηθεί για το clustering και πως διαφορετικές συστάδες δεν έχουν κοινές περιοχές.

Γενικά ένας θησαυρός συνδυάζει μία λίστα με τους όρους ενός δεδομένου πεδίου γνώσης κι ένα σύνολο από σχετικούς όρους για κάθε όρο στην λίστα τα οποία είναι τα συνώνυμα του τρέχοντα όρου. Στην προσέγγιση μας ο κατασκευασμένος *Θησαυρός Περιοχών* περιέχει όλους τους τύπους περιοχών οι οποίοι υπάρχουν μέσα στο σύνολο εκπαίδευσης. Επιλέγεται ο κάθε τύπος περιοχής να αναπαρίσταται από το διάνυσμα χαρακτηριστικών που είναι πιο κοντά στο διάνυσμα χαρακτηριστικών του κέντρου του κάθε cluster. Στην σχέση (8.5) φαίνεται ο τρόπος υπολογισμού του κάθε κέντρου σαν το “κέντρο μάζας” όλων των διανυσμάτων χαρακτηριστικών που ανήκουν στο ίδιο cluster. Όπου $|w_i|$ είναι το πλήθος των στοιχείων του w_i . Έτσι το διάνυσμα χαρακτηριστικών που επιλέγεται να αντιπροσωπεύει τον κάθε τύπο περιοχής δίνεται από την σχέση (8.6). Η απόσταση ανάμεσα σε δύο διανύσματα χαρακτηριστικών εξηγείται παρακάτω στη σχέση (8.7). Αυτοί οι τύποι περιοχών είναι οι κοντινότερες περιοχές στα κέντρα μάζας των συστάδων και όλα τα υπόλοιπα διανύσματα χαρακτηριστικών της συστάδας είναι τα “συνώνυμα” τους. Σκοπός του θησαυρού είναι να αποτελέσει ένα ενδιάμεσο βήμα στην αντιστοιχία χαρακτηριστικών χαμηλού και υψηλού επιπέδου (έννοιες) και να διευκολύνει την συσχέτιση τους.

$$z(w_i) = \frac{1}{|w_i|} \sum_{r \in w_i} f(r) \quad (8.5)$$

$$f(w_i) = f\left(\arg \min_{r \in w_i} \{d(f(r), z(w_i))\}\right) \quad (8.6)$$

Ο κάθε τύπος περιοχής λοιπόν αναπαρίσταται από ένα διάνυσμα χαρακτηριστικών το οποίο περιέχει όλη την εξαχθείσα πληροφορία χαμηλού επιπέδου για την περιοχή. Όπως είναι προφανές μία περιγραφή χαμηλού επιπέδου δεν περιέχει καμία σημασιολογική πληροφορία, αποτελεί απλώς μία τυποποιημένη αναπαράσταση των οπτικών χαρακτηριστικών. Σημασιολογική πληροφορία και μόνο περιέχει μία έννοια υψηλού επιπέδου. Ένας τύπος περιοχής περιγράφεται σαν κάτι ανάμεσα σε χαρακτηριστικά υψηλού και χαμηλού επιπέδου. Περιέχει τις απαραίτητες πληροφορίες για να περιγράψει τυποποιημένα τα χαρακτηριστικά υψής και χρώματος αλλά μπορεί επίσης να περιγράψει με μία περιγραφή χαμηλότερου επιπέδου από τις σημασιολογικές έννοιες υψηλού επιπέδου. Για παράδειγμα κάποιος μπορεί να χαρακτηρίσει ένα είδος περιοχής σαν *πράσινη περιοχή με τραχιά υφή*.

Για τον σχηματισμό του θησαυρού είναι απαραίτητος ένας αριθμός από διανύσματα χαρακτηριστικών από περιοχές εικόνων. Αν το σύνολο όλων των περιοχών δεν είναι πολύ μεγάλο τότε όλες οι περιοχές θα χρησιμοποιηθούν για τον σχηματισμό του θησαυρού. Συγκεκριμένα ο αλγόριθμος συσταδοποίησης θα εφαρμοστεί στα διανύσματα χαρακτηριστικών όλων των χαρακτηριστικών καρέ. Όταν το σύνολο των εικόνων είναι πολύ μεγάλο και δεν υπάρχει δυνατότητα να χρησιμοποιηθεί ολόκληρο για λόγους χρονοβόρων υπολογισμών αλλά και υπολογιστικής μνήμης τότε επιλέγεται ένα υποσύνολο αυτών. Ένας τρόπος με τον οποίο μπορεί να γίνει αυτό είναι να συλλέγονται περιοχές από εικόνες οι οποίες απεικονίζουν τις επιλεγμένες προς εντοπισμό έννοιες και στην συνέχεια με τυχαίο τρόπο κάποιες ακόμα περιοχές. Αυτό λοιπόν το υποσύνολο θα χρησιμοποιηθεί για να κατασκευασθεί

ο θησαυρός στην περίπτωση ενός μεγάλου αριθμού συνολικών περιοχών. Στην περίπτωση αυτή δεν ισχύει η σχέση (8.3) αλλά ισχύει ότι $\bigcup_i w \subset R$, $i = 1 \dots N_T$.

8.3 Διανύσματα Αναπαράστασης

Σε αυτή την ενότητα παρουσιάζεται ένας αλγόριθμος που χρησιμοποιήθηκε για να περιγραφεί η κάθε εικόνα, που είναι κατατημημένη σε περιοχές, με την βοήθεια του κατασκευασμένου οπτικού θησαυρού.

Πάνω στα διανύσματα της μορφής της εξίσωσης (8.1) εφαρμόζεται η Ευκλείδεια απόσταση για να μετρήσουμε την απόσταση μεταξύ δύο διαφορετικών περιοχών, όσον αφορά τα διανύσματα χαρακτηριστικών τους. Έστω δύο οποιεσδήποτε περιοχές $r_1, r_2 \in \mathcal{R}$, $R \subset \mathcal{R}$, έχουν διανύσματα χαρακτηριστικών $f_1, f_2 \in \mathcal{F}$, $F \subset \mathcal{F}$ και Ευκλείδεια απόσταση $d(f_1, f_2)$ η οποία δίνεται από την σχέση (8.7). F είναι το σύνολο των διανυσμάτων χαρακτηριστικών για το συγκεκριμένο σύνολο περιοχών, ενώ \mathcal{F} είναι όλος ο χώρος των διανυσμάτων χαρακτηριστικών. Το ίδιο ισχύει και για τα R και \mathcal{R} . Πρέπει να σημειωθεί ότι το MPEG-7 δεν ορίζει αυστηρά μέτρα απόστασης για του περιγραφείς, παρά μόνο προτείνει ενδεικτικά αυτούς τους τύπους που περιγράφηκαν στο κεφάλαιο 2 ώστε να αφήνει την ευελιξία να χρησιμοποιούνται και άλλοι τύποι αποστάσεων και να ελέγχεται η αποτελεσματικότητά τους.

$$d(f_1, f_2) = \sqrt{\sum_{i=1}^n (f_1^i - f_2^i)^2} \quad (8.7)$$

Έχοντας υπολογίσει την απόσταση της κάθε περιοχής της εικόνας από όλους τους τύπους περιοχής του κατασκευασμένου θησαυρού, το διάνυσμα αναπαράστασης του κάθε χαρακτηριστικού καρέ που σημασιολογικά περιγράφει το οπτικό περιεχόμενο της εικόνας σχηματίζεται κρατώντας την μικρότερη απόσταση από κάθε τύπο περιοχής του θησαυρού. Πιο συγκεκριμένα το διάνυσμα αναπαράστασης που περιγράφει το χαρακτηριστικό καρέ k_i είναι αυτό του τύπου (8.8).

$$m_i = [m_i(1), m_i(2), \dots, m_i(j), \dots, m_i(N_T)], \quad i = 1 \dots N_K \quad (8.8)$$

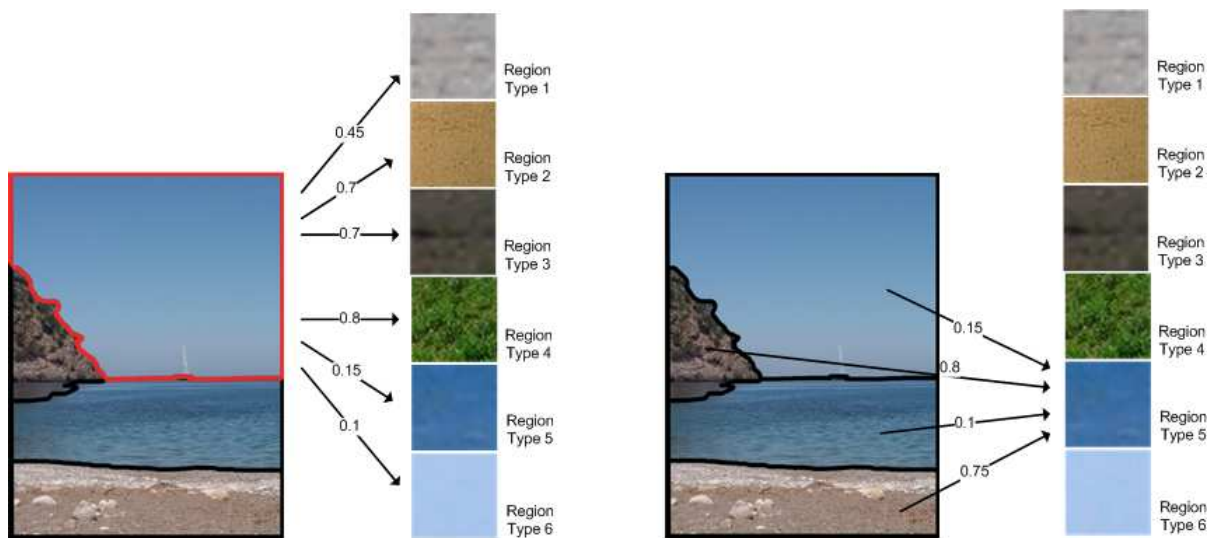
όπου

$$m_i(j) = \min_{r \in R(k_i)} \{d(f(w_j), f(r))\}, \quad i = 1 \dots N_K, \quad j = 1 \dots N_T \quad (8.9)$$

Κάθε διάνυσμα αναπαράστασης συμβολίζεται με $m_i \in M$, $i = 1 \dots N_K$, όπου M το σύνολο όλων των διανυσμάτων αναπαράστασης και m_i το διάνυσμα αναπαράστασης του χαρακτηριστικού καρέ k_i .

Στο σχήμα 8.2 έχουμε μία εικόνα χωρισμένη σε περιοχές και τους 6 τύπους περιοχών ενός οπτικού θησαυρού. Αριστερά παρουσιάζονται οι αποστάσεις για την περιοχή του ουρανού από κάθε τύπο περιοχής και δεξιά οι αποστάσεις κάθε περιοχής της εικόνας από τον τύπο περιοχής 5. Το διάνυσμα αναπαράστασης σχηματίζεται από τις μικρότερες αποστάσεις για κάθε τύπο περιοχής. Στην προκειμένη περίπτωση για τον τύπο περιοχής 5, η ελάχιστη απόσταση είναι 0.1. Το διάνυσμα αναπαράστασης για αυτή την εικόνα και για τον συγκεκριμένο θησαυρό θα είναι όπως φαίνεται στην σχέση (8.10).

$$m = [m(1), m(2), m(3), m(4), m(5), m(6)] \quad (8.10)$$



Σχήμα 8.2: Αποστάσεις ανάμεσα σε περιοχές εικόνας και τύπους περιοχής, αριστερά αποστάσεις ανάμεσα σε μία περιοχή της εικόνας και σε όλους τους τύπους περιοχής, δεξιά αποστάσεις ανάμεσα σε όλες τις περιοχές μία εικόνας και έναν τύπο περιοχής

όπου το $m(5)$ θα είναι 0.1. Υπολογίζοντας όλες τις αποστάσεις ανάμεσα σε όλες τις περιοχές της εικόνας (4) και σε όλους τους τύπους περιοχής (6), συνολικά δηλαδή $4 \times 6 = 24$ αποστάσεις σχηματίζεται και το διάνυσμα αναπαράστασης.

Στο σχήμα 8.3 φαίνονται τα διανύσματα αναπαράστασης για 5 εικόνες και για τον οπτικό θησαυρό 6 περιοχών, του οποίου οι περιοχές αναπαριστώνται στο σχήμα.

8.4 Λανθάνουσα Σημασιολογική Ανάλυση (Latent Semantic Analysis - LSA)

Σαν επόμενο στάδιο είναι η χρήση της λανθάνουσας σημασιολογικής ανάλυσης. Αυτή είναι μία τεχνική που πρωτοχρησιμοποιήθηκε στην επεξεργασία φυσικής γλώσσας, που εκμεταλλεύεται τις σχέσεις ανάμεσα σε ένα σύνολο εγγράφων και τους όρους τους οποίους περιέχουν, δημιουργώντας συνήθως ένα σύνολο εννοιών συσχετισμένες με τα έγγραφα και τους όρους. Για να γίνει η χρήση της τεχνικής αυτής η οποία εφαρμόζεται πάνω σε πίνακες που περιέχουν τον αριθμό των φορών τις οποίες κάθε λέξη εμφανίζεται σε κάθε έγγραφο, γίνεται ένας μετασχηματισμός των διανυσμάτων αναπαράστασης πριν από αυτό το στάδιο. Τα διανύσματα αναπαράστασης περιέχουν για κάθε τύπο περιοχής τις ελάχιστες αποστάσεις από όλες τις περιοχές της εικόνας. Έτσι ένας τύπος περιοχής που εμφανίζεται πολύ παρόμοιος σε μία εικόνα θα έχει χαμηλή τιμή απόστασης. Όλες οι αποστάσεις διαιρούνται με την μέγιστη απόσταση και αφαιρούνται από το 1. Έτσι κανονικοποιούνται στο $[0,1]$, με τιμές κοντά στο 1 να έχουν εκείνες οι περιοχές οι οποίες είναι πολύ όμοιες με τους τύπους περιοχής.



Εικόνα 1	[0.4	0.3	0.6	0.7	0.2	0.1]
Εικόνα 2	[0.5	0.7	0.1	0.8	0.3	0.4]
Εικόνα 3	[0.8	0.1	0.9	0.8	0.2	0.5]
Εικόνα 4	[0.1	0.7	0.3	0.6	0.2	0.4]
Εικόνα 5	[0.3	0.7	0.4	0.1	0.9	0.7]



Εικόνα 1



Εικόνα 2



Εικόνα 3



Εικόνα 4



Εικόνα 5

Σχήμα 8.3: Παραδείγματα διανυσμάτων αναπαράστασης

Πλέον τα διανύσματα m_i μετασχηματίζονται στα m'_i , τα οποία δίνονται στην σχέση (8.11).

$$m'_i = [m'_i(1), m'_i(2), \dots, m'_i(j), \dots, m'_i(N_T)], \quad i = 1 \dots N_K \quad (8.11)$$

$$m'_i(j) = 1 - \frac{m_i(j)}{d_{max}} \quad (8.12)$$

$$d_{max} = \max_{r \in R} \{d(f(w_j), f(r))\}, \quad j = 1 \dots N_T \quad (8.13)$$

Πιο συγκεκριμένα, έστω \mathcal{M} ένας πίνακας του οποίου το (i, j) στοιχείο περιγράφει την ύπαρξη του όρου i στο έγγραφο j . Σε αυτή την προσέγγιση, οι όροι αντιστοιχούν στα είδη περιοχής του θησαυρού και το έγγραφο σε μία δεδομένη εικόνα. Έτσι μία εικόνα θεωρείται ένα έγγραφο της οποίας οι λέξεις είναι οι τύποι περιοχής που αντιστοιχούν στις περιοχές της.

$$\mathcal{M} = \begin{pmatrix} m'_{11}(1) & \dots & m'_{N_K}(1) \\ \vdots & \ddots & \vdots \\ m'_{11}(N_T) & \dots & m'_{N_K}(N_T) \end{pmatrix} \quad (8.14)$$

όπου μία γραμμή $q_i^T = (m'_{11}(i), \dots, m'_{N_K}(i))$ είναι ένα διάνυσμα που περιγράφει την σχέση των όρων (τύποι περιοχής), συγκεκριμένα του τύπου περιοχής w_i , με το κάθε έγγραφο (εικόνα). Τα στοιχεία δηλαδή του διανύσματος γραμμής i είναι τα $m'_{ij}(i)$ στοιχεία των μετασχηματισμένων διανυσμάτων

αναπαράστασης m'_j , $j = 1 \dots N_K$. Κάθε στήλη του \mathcal{M} δηλαδή $m'_j = \begin{pmatrix} m'_{1j}(1) \\ \vdots \\ m'_{1j}(N_T) \end{pmatrix}$ είναι ένα

διάνυσμα το οποίο αντιστοιχεί σε ένα συγκεκριμένο έγγραφο(εικόνα) και περιγράφει την σχέση του με κάθε όρο (τύπος περιοχής). Αυτή η σχέση είναι η βεβαιότητα για την ύπαρξη του είδους περιοχής στην δεδομένη εικόνα. Είναι λοιπόν προφανές ότι το m'_j είναι το μετασχηματισμένο διάνυσμα αναπαράστασης της εικόνας αυτής. Επομένως ισχύει ότι $\mathcal{M} = [m_1^T, \dots, m_{N_K}^T]$.

Το εσωτερικό γινόμενο $q_i^T q_p$ ανάμεσα σε δύο διανύσματα όρων δίνει την συσχέτιση τους και ο πίνακας $\mathcal{M}\mathcal{M}^T$ περιέχει όλα αυτά τα εσωτερικά γινόμενα. Επιπλέον ο πίνακας $\mathcal{M}^T\mathcal{M}$ περιέχει όλα τα εσωτερικά γινόμενα ανάμεσα στα διανύσματα εγγράφων $m_i^T m_p$, περιγράφοντας την συσχέτιση τους πάνω στους όρους. Υποθέτοντας ότι υπάρχει μία αποσύνθεση του \mathcal{M} που δίνεται από την σχέση $\mathcal{M} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$, όπου \mathbf{U} και \mathbf{V} είναι ορθοκανονικοί πίνακες και ο $\mathbf{\Sigma}$ διαγώνιος. Αυτή η αποσύνθεση είναι γνωστή ως Αποσύνθεση Ιδιαζουσών Τιμών (Singular Value Decomposition -SVD) και απεικονίζεται αναλυτικά στην σχέση (8.15).

$$\mathcal{M} = \left((\mathbf{u}_1) \dots (\mathbf{u}_{N_T}) \right) \begin{pmatrix} \sigma_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_{N_T} \end{pmatrix} \left((\mathbf{v}_1) \dots (\mathbf{v}_{N_T}) \right)^T \quad (8.15)$$

όπου σ_i είναι οι ιδιοτιμές και $\mathbf{u}_i, \mathbf{v}_i$ είναι τα αριστερά και δεξιά ιδιοδιανύσματα αντίστοιχα.

Τώρα οι N_L μεγαλύτερες ιδιοτιμές μπορούν να κρατηθούν από τον $\mathbf{\Sigma}$ μαζί με τις αντίστοιχες στήλες του \mathbf{U} και σειρές του \mathbf{V} . Μία εκτίμηση του \mathcal{M} δίνεται από την σχέση (8.16)

$$\hat{\mathcal{M}} = \mathcal{M}_{N_L} = \mathbf{U}_{N_L} \mathbf{\Sigma}_{N_L} \mathbf{V}_{N_L}^T \quad (8.16)$$

Με αυτόν τον τρόπο επιτυγχάνεται μία μετάφραση των διανυσμάτων όρων και εγγράφων στον χώρο των “εννοιών”. Αν \hat{q}_i και \hat{m}_i είναι τα διανύσματα που δηλώνουν την ύπαρξη των όρων στις έννοιες και την σχέση ανάμεσα σε έγγραφα και έννοιες αντίστοιχα, τότε η μετατροπή ενός διανύσματος εγγράφων στον χώρο των εννοιών πετυχαίνεται με την χρήση του πίνακα Σ και \mathbf{U} με την μετατροπή:

$$\hat{m}_i = \Sigma_{N_L}^{-1} \mathbf{U}_{N_L}^T m'_i \quad (8.17)$$

Στην προσέγγιση της παρούσας εργασίας σαν ένα επόμενο βήμα τα διανύσματα αναπαράστασης του συνόλου εκπαίδευσης εξάγονται με την χρήση του θησαυρού περιοχών όπως εξηγήθηκε. Μετά εφαρμόζουμε η λανθάνουσα σημασιολογική ανάλυση κι έτσι ορίζονται οι πίνακες Σ και \mathbf{U} . Τελικά κάθε διάνυσμα αναπαράστασης μεταφέρθηκε στον χώρο των εννοιών.

Ο αλγόριθμος μπορεί να περιέχει το στάδιο της λανθάνουσας σημασιολογικής ανάλυσης, μπορεί και όχι. Δηλαδή τα διανύσματα αναπαράστασης μπορεί είτε να παραμείνουν ως έχουν είτε να μεταφερθούν στον χώρο των εννοιών με την τεχνική αυτή, πριν από το στάδιο της εκπαίδευσης των νευρωνικών δικτύων.

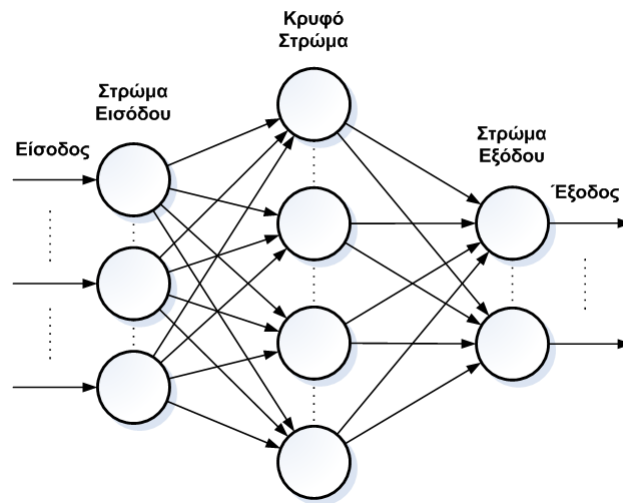
8.5 Εκπαίδευση Ανιχνευτών

Ένα Τεχνητό Νευρωνικό Δίκτυο[28], αποκαλούμενο συνήθως “νευρωνικό δίκτυο” είναι ένα μαθηματικό ή υπολογιστικό μοντέλο βασισμένο σε βιολογικά δίκτυα νευρώνων. Αποτελείται από ένα σύνολο αλληλοσυνδεδεμένων τεχνητών νευρώνων και επεξεργάζεται δεδομένα χρησιμοποιώντας μια συνδεομοστροφική προσέγγιση για υπολογισμούς. Στις περισσότερες περιπτώσεις ένα τεχνητό νευρωνικό δίκτυο είναι ένα προσαρμόσιμο σύστημα το οποίο αλλάζει την δομή του βασισμένο σε εξωτερικές ή εσωτερικές πληροφορίες οι οποίες ρέουν στο δίκτυο κατά την φάση της εκπαίδευσης. Ο όρος “νευρωνικό δίκτυο” μπορεί επίσης να σημαίνει συστήματα βασισμένα σε βιολογικές διεργασίες. Με πιο πρακτικούς όρους μπορεί να ειπωθεί ότι είναι μη γραμμικά στατιστικά μοντέλα δεδομένων και μπορούν να χρησιμοποιηθούν για να μοντελοποιήσουν πολύπλοκες σχέσεις μεταξύ εισόδων και εξόδων ή για να βρίσκουν τύπους σε σύνολα δεδομένων. “Μαθαίνουν” πως να μετατρέπουν τα δεδομένα εισόδου στην επιθυμητή απόκριση, έτσι χρησιμοποιούνται ευρέως για ταξινόμηση (classification). Παρά το γεγονός ότι δεν ενσωματώνουν καμία γνώση για τον θεματικό χώρο παρέχουν καλές επιδόσεις γενίκευσης. Η αρχική έμπνευση για την κατασκευή τους προήλθε από την μελέτη του κεντρικού νευρικού συστήματος και των νευρώνων του (συμπεριλαμβανομένου των αξόνων, δενδριτών και συναψέων) οι οποίοι αποτελούν ένα από τα πιο σημαντικά στοιχεία επεξεργασίας πληροφορίας του.

Η λειτουργία του νευρωνικού δικτύου μπορεί να συμβολισθεί συνολικά με την συνάρτηση:

$$f : Y \rightarrow Z \quad (8.18)$$

όπου Y ο χώρος εισόδου του νευρωνικού δικτύου και Z ο χώρος εξόδου του. Ο όρος δίκτυο μπορεί συσχετίζεται με το γεγονός ότι η συνάρτηση $f(y)$ ορίζεται σαν η σύνθεση πολλών άλλων συναρτήσεων $g_i(y)$, οι οποίες μπορούν επιπλέον να οριστούν σαν σύνθεση άλλων συναρτήσεων. Αυτό μπορεί να συμβολιστεί με μία δομή δικτύου όπου βέλη δηλώνουν τις εξαρτήσεις μεταξύ των μεταβλητών. Μία ευρέως διαδεδομένη σύνθεση είναι το μη γραμμικό άθροισμα με βάρη, όπου: $f(y) = K \left(\sum_i w_i g_i(y) \right)$, όπου K είναι μία προκαθορισμένη συνάρτηση.



Σχήμα 8.4: Δομή νευρωνικού δικτύου

Ένα είδος νευρωνικού δικτύου είναι τα προσοτροφοδοτούμενα νευρωνικά δίκτυα των οποίων οι συνδέσεις δεν σχηματίζουν απευθείας κύκλο. Ήταν το πρώτο και πιο απλού τύπου νευρωνικό δίκτυο που σχεδιάστηκε. Η πληροφορία σε αυτό μεταδίδεται μόνο προς την μία κατεύθυνση, από τους κόμβους εισόδου, μέσω των κόμβων κρυφού στρώματος και τέλος στους κόμβους εξόδου. Τα στρώματα φαίνονται στο σχήμα 8.4. Τα πολυστρωματικά Perceptrons (Multi Layer Perceptron - MLP) είναι ένα είδος προσοτροφοδοτούμενων νευρωνικών δικτύων τα οποία αποτελούνται από πολλαπλά στρώματα νευρώνων (στρώμα εισόδου, στρώμα εξόδου και τουλάχιστον ένα κρυφό στρώμα) τα οποία είναι συνδεδεμένα ώστε να γίνεται προσοτροφοδότηση της πληροφορίας. Όλοι οι νευρώνες σε κάθε στρώμα έχουν άμεσο σύνδεσμο με κάθε νευρώνα του επακόλουθου στρώματος. Κατά το γενικό θεώρημα προσέγγισης για τα νευρωνικά δίκτυα, οποιαδήποτε συνεχής συνάρτηση η οποία απεικονίζει ένα σύνολο πραγματικών αριθμών σε ένα σύνολο εξόδου πάλι πραγματικών αριθμών μπορεί να προσεγγιστεί από ένα πολυστρωματικό perceptron με μόνο ένα κρυφό στρώμα νευρώνων.

Χρησιμοποιείται μία ποικιλία τεχνικών εκπαίδευσης με πιο γνωστή εκείνη που ονομάζεται “back-propagation”, μία τεχνική επιβλεπόμενης μάθησης. Συνοπτικά η τεχνική αυτή αποτελείται από τα παρακάτω βήματα:

1. Παρουσίαση ενός συνόλου εκπαίδευσης.
2. Σύγκριση της εξόδου του δικτύου με την επιθυμητή και υπολογισμός του σφάλματος.
3. Υπολογισμός για κάθε νευρώνα της επιθυμητής εξόδου και του τοπικού σφάλματος.
4. Προσαρμογή των βαρών για κάθε νευρώνα ώστε να μειωθεί το τοπικό σφάλμα.
5. Επιβολή “ποινής” για τα τοπικά σφάλματα και ανάθεση “ευθύνης” σε νευρώνες με μεγάλα βάρη.
6. Επανάληψη όλων των βημάτων χρησιμοποιώντας την “ποινή” σαν σφάλμα τους κάθε νευρώνα.

8.6 Έξοδος Συστήματος

Για κάθε σημασιολογική έννοια εκπαιδεύτηκε ένα ξεχωριστό νευρωνικό δίκτυο τύπου MLP, λύνοντας έτσι ένα δυαδικό πρόβλημα, της ύπαρξης ή μη της συγκεκριμένης σημασιολογικής έννοιας. Η είσοδος του νευρωνικού είναι τα διανύσματα m_i ή \hat{m}_i , ανάλογα με το αν δεν έχει γίνει LSA ή αν έχει γίνει. Η έξοδος του κάθε νευρωνικού δικτύου κανονικοποιείται στο διάστημα $[0,1]$ και αυτή η τιμή αποτελεί το ποσοστό πεποίθησης (confidence) της ύπαρξης της έννοιας στην εικόνα. Τιμές κοντά στο 1 δηλώνουν υψηλή βεβαιότητα απεικόνισης ενώ τιμές κοντά στο 0 δηλώνουν υψηλή βεβαιότητα μη απεικόνισης. Τα ποσοστά αυτά θα χρησιμοποιηθούν στην διαδικασία αξιολόγησης της απόδοσης του συστήματος. Σαν επόμενο βήμα είναι η χρήση ενός κατωφλίου (threshold, με την οποία μετασχηματίζεται η έξοδος σε δυαδική απάντηση της ύπαρξης ή μη της κάθε έννοιας. Για κάθε νευρωνικό επιλέγεται η κατάλληλη τιμή κατωφλίου ώστε να δίνει το καλύτερο δυνατό αποτέλεσμα με τα μέτρα που χρησιμοποιούνται και περιγράφονται στην επόμενη ενότητα. Ο ανιχνευτής τελικά για κάθε έννοια δέχεται σαν είσοδο το σχηματισμένο για κάθε εικόνα (χαρακτηριστικό καρέ) διάνυσμα m_i ή \hat{m}_i και σαν έξοδος δίνεται η δυαδική απάντηση $\{0,1\}$ της απεικόνισης ή μη της έννοιας. Το αντίστοιχο ποσοστό πεποίθησης για κάθε χαρακτηριστικό καρέ k_i συμβολίζεται με $q_i \in [0,1]$.

Έστω c_i η κάθε έννοια που έχει επιλεχθεί να εντοπιστεί, C το σύνολο όλων των εννοιών και N_C το πλήθος τους. Η δυαδική απόφαση για ένα χαρακτηριστικό καρέ $k \in K$ για τον αν απεικονίζει ή όχι την έννοια c_i συμβολίζεται με $b_i(k)$. Το σύνολο των χαρακτηριστικών καρέ τα οποία τελικά εντοπίζονται πως απεικονίζουν την έννοια c_i συμβολίζονται με D_i . Δηλαδή:

$$b_i(k) \in \{0,1\}, \quad k \in K, \quad i = 1 \dots N_C \quad (8.19)$$

$$D_i = \{k \in K : b_i(k) = 1, \quad i = 1 \dots N_C\} \quad (8.20)$$

Με την χρήση δηλαδή ενός κατωφλίου h_j για την έννοια c_j πάνω στις τιμές q_i καταλήγουμε στο $b_j(k_i) = 0$ εάν $q_i \leq h_j$, ενώ $b_j(k_i) = 1$ εάν $q_i > h_j$.

Κεφάλαιο 9

Πειράματα

Διεξήχθησαν πειράματα για το πρόβλημα του αυτόματου εντοπισμού εννοιών σε εικόνες-βίντεο με τον αλγόριθμο που παρουσιάστηκε στο κεφάλαιο 8 και χρησιμοποιήθηκαν κάποια μέτρα για την αξιολόγηση της αποτελεσματικότητας του αλγορίθμου στην επίλυση του προβλήματος. Τα πειράματα αυτά διεξήχθησαν στο σύνολο δεδομένων του TRECVID αλλά και σε ένα σύνολο εικόνων από την συλλογή του Corel.

9.1 Διαδικασία Μετρήσεων

Μετά το στάδιο εκπαίδευσης των νευρωνικών δικτύων είναι πλέον έτοιμοι οι ανιχνευτές των εννοιών που έχουν επιλεγεί. Με την τεχνική αυτή δημιουργείται ένας ανιχνευτής για κάθε έννοια υψηλού επιπέδου. Για να γίνει ανίχνευση σε ένα σύνολο από εικόνες πρέπει για αυτές τις εικόνες να σχηματιστούν τα διανύσματα αναπαράστασης με την διαδικασία που έχει περιγραφεί στον αλγόριθμο της προηγούμενης ενότητας. Αν κατά την διαδικασία της εκπαίδευσης των νευρωνικών δικτύων είχε χρησιμοποιηθεί η τεχνική LSA τότε αυτό πρέπει να γίνει και με τα συγκεκριμένα διανύσματα χρησιμοποιώντας του ίδιους πίνακες αποσύνθεσης. Έτσι τελικά τα διανύσματα καταλήγουν στην ίδια διάσταση με εκείνα των διανυσμάτων εκπαίδευσης. Δίνονται στους ανιχνευτές σαν είσοδος και σαν έξοδος για το κάθε πλάνο s είναι η δυαδική απάντηση απεικόνισης ή όχι. Για την διαδικασία της αξιολόγησης βέβαια κάποια μέτρα εφαρμόζονται εκτός από αυτή την δυαδική τιμή και στα ποσοστά πεποίθησης απεικόνισης. Στο κάθε πλάνο s αντιστοιχεί το χαρακτηριστικό καρτέ k στο οποίο αντιστοιχούν και τα παραπάνω διανύσματα. Η δυαδική απάντηση για την έννοια c_i για ένα χαρακτηριστικό καρτέ k είναι το $b_i(k)$, το οποίο δίνεται στην εξίσωση (8.19). Έτσι για το σύνολο των εικόνων το σύστημα έχει αποφανθεί αν η κάθε έννοια απεικονίζεται στις εικόνες μία προς μία ή όχι. Από τις εικόνες αυτές είναι σχεδόν βέβαιο ότι κάποιες για τις οποίες η έξοδος δηλώνει απεικόνιση της έννοιας αυτές να μην απεικονίζουν τελικά την έννοια, όπως επίσης και το αντίθετο, δηλαδή εικόνες για τις οποίες το σύστημα απεφάνθη μη απεικόνιση, τελικά να είναι από εκείνες που απεικονίζουν την έννοια. Το ποσοστό πεποίθησης για κάθε χαρακτηριστικό καρτέ k_i συμβολίζεται με g_i .

Το σύνολο δεδομένης αλήθειας (ground-truth) περιλαμβάνει την πληροφορία για το ποιά από τα χαρακτηριστικά καρτέ απεικονίζουν τις επιλεγμένες προς εντοπισμό έννοιες και ποιά όχι. Έτσι για κάθε έννοια c_i το σύνολο των χαρακτηριστικών καρτέ χωρίζεται σε δύο υποσύνολα. Τα ένα

αποτελείται από τις εικόνες που απεικονίζουν την έννοια c_i και το άλλο από εκείνες που δεν την απεικονίζουν. Τα δύο αυτά υποσύνολα συμβολίζονται με G_i και \bar{G}_i αντίστοιχα:

$$G_i : \{k \in K : c_i \in C(k)\} \quad (9.1)$$

$$\bar{G}_i : \{k \in K : c_i \notin C(k)\} \quad (9.2)$$

όπου $C(k)$ είναι το σύνολο των εννοιών εκείνων οι οποίες απεικονίζονται στο χαρακτηριστικό καρέ k . Ακόμα από την έξοδο του ανιχνευτή για κάθε έννοια c_i σχηματίζεται το σύνολο $D_i = \{k \in K : b_i(k) = 1\}$ με εκείνα τα χαρακτηριστικά καρέ τα οποία δόθηκαν σαν είσοδος και για τα οποία εντοπίστηκε ότι απεικονίζουν την έννοια c_i .

Υπάρχει η ανάγκη λοιπόν να χρησιμοποιηθούν κάποια αντικειμενικά μέτρα για την αξιολόγηση της απόδοσης των ανιχνευτών. Τα μέτρα γενικά θα εφαρμοστούν πάνω στην δυαδική έξοδο του ανιχνευτή D_i αλλά και σε μία ταξινομημένη λίστα των ποσοστών πεποίθησης q_i . Τα χαρακτηριστικά καρέ ταξινομούνται με βάση τα q_i με φθίνουσα σειρά και σε κάθε ένα από αυτά αντιστοιχεί μία δυαδική τιμή x_i , η οποία είναι 1 εάν για το χαρακτηριστικό καρέ k στην i -οστή θέση της ταξινομημένης λίστας για την έννοια c_j ισχύει ότι $k \in G_j$, αλλιώς είναι 0. Αν το x_i είναι 1 τότε θα λέγεται ότι το χαρακτηριστικό καρέ στην i -οστή θέση της ταξινομημένης λίστας είναι “σχετικό” ή “θετικό” ως προς την απεικόνιση της έννοιας. “Αρνητικό” θα σημαίνει πως δεν απεικονίζει την έννοια.

Δύο γνωστά μέτρα εφαρμόσιμα σε προβλήματα σαν και το συγκεκριμένο είναι το μέτρο ακρίβειας (precision) και το μέτρο ανάκτησης (recall). Ορίζοντας σαν $|\cdot|$ το πλήθος των στοιχείων ενός συνόλου, τα μέτρα precision και recall δίνονται από τους τύπους (9.3) και (9.4) αντίστοιχα.

$$P_i = \frac{|D_i \cap G_i|}{|D_i|}, \quad i = 1 \dots N_C \quad (9.3)$$

$$R_i = \frac{|D_i \cap G_i|}{|G_i|}, \quad i = 1 \dots N_C \quad (9.4)$$

Με απλά λόγια το πρώτο μέτρο είναι το ποσοστό των εικόνων εκείνων που όντως απεικονίζουν την έννοια ως προς όλες εκείνες που το σύστημα αποφάνθηκε ότι την απεικονίζουν. Το δεύτερο είναι το ποσοστό των εικόνων που ανιχνεύθηκαν σωστά ότι απεικονίζουν την έννοια ως προς όλες τις εικόνες που την απεικονίζουν.

Έπειτα πάνω στην ταξινομημένη λίστα των χαρακτηριστικών καρέ όπως αναφέρθηκε πριν εφαρμόζεται το μέτρο της μέσης ακρίβειας (Average Precision). Για τις ανάγκες υπολογισμού της ορίζεται με την σχέση (9.5) ακρίβεια εάν πάρουμε τα m χαρακτηριστικά καρέ με τα μεγαλύτερα q , πάνω σε αυτά τα m καρέ.

$$p_m = \frac{1}{m} \sum_{k=1}^m x_k \quad (9.5)$$

Μπορεί να ερμηνευθεί ότι το p_m είναι η μέση τιμή των x_1, x_2, \dots, x_m , και συμβολίζεται με \bar{x}_m . Η μέση ακρίβεια ορίζεται σαν η μέση τιμή των ακριβειών μετά από κάθε σχετικό χαρακτηριστικό καρέ που συναντιέται στην λίστα. Μαθηματικά εκφράζεται από τον τύπο (9.6), ο οποίος αντιστοιχεί στην μέση ακρίβεια για την έννοια c_i με παράθυρο n .

$$AP_i^n = \frac{1}{|G_i|} \sum_{j=1}^n I(x_j) p_j = \frac{1}{|G_i|} \sum_{j=1}^n I(x_j) \bar{x}_j \quad (9.6)$$

όπου n είναι το πλήθος των χαρακτηριστικών καρέ από την λίστα με τα μεγαλύτερα q που επιλέγεται για να υπολογιστεί η μέση ακρίβεια, το παράθυρο δηλαδή πάνω στο οποίο υπολογίζεται. $I(x_i)$ είναι μία συνάρτηση για την οποία ισχύει $I(x_i) = x_i$ εφόσον $x_i \in \{0,1\}$ ¹. Άρα τελικά η μέση ακρίβεια δίνεται από τον τύπο (9.7).

$$AP_i^n = \frac{1}{|G_i|} \sum_{j=1}^n x_j p_j = \frac{1}{|G_i|} \sum_{j=1}^n \frac{x_j}{j} \sum_{k=1}^j x_k \quad (9.7)$$

Το μέτρο της μέσης ακρίβειας συσχετίζεται με το μέτρο της μέσης ακρίβειας (σχέση 1.2) το οποίο χρησιμοποιήθηκε στα πειράματα του MPEG-7 για την αξιολόγηση της δυνατότητας των οπτικών περιγραφών να περιγράψουν αποτελεσματικά τα οπτικά χαρακτηριστικά μιας εικόνας. Ομοίως και τα δύο αυτά μέτρα “ενδιαφέρονται” για την θέση όσον αφορά το confidence, στην οποία επιστρέφονται οι σχετικές εικόνες. Διαφορά τους είναι πως το Average Precision παίρνει μεγαλύτερες τιμές για καλύτερες αποδόσεις, ενώ το Average Rank μικρότερες.

Ένα ακόμα μέτρο είναι εκείνο που αποκαλείται r -precision, το οποίο είναι η ακρίβεια μετά από r ανακτημένες εικόνες. Τα πρώτα r δηλαδή χαρακτηριστικά καρέ στην ταξινομημένη λίστα. Το r -precision που αντιστοιχεί στην έννοια c_i δίνεται από την σχέση (9.8).

$$rP_i = \frac{1}{r_i} \sum_{j=1}^{r_i} x_j \quad (9.8)$$

Στα πειράματα που έγιναν από το MPEG-7 και συγκεκριμένα στον τύπο (1.1) ορίζεται το $K(q)$ το οποίο δίνει μια ανεκτικότητα στο ποιες εικόνες θεωρούνται σχετικές. Έτσι και στον τύπο (9.8) το r θα οριστεί έτσι ώστε η ακρίβεια να υπολογίζεται σε ένα μεγαλύτερο παράθυρο. Συγκεκριμένα:

$$r_i = \min\{4 \cdot |G_i|, 2 \cdot \max\{|G_j|\}\}, \quad j = 1 \dots N_C \quad (9.9)$$

Ένα ακόμα μέτρο είναι το κανονικοποιημένο μέτρο ανάκτησης (normalized recall) το οποίο ορίζεται από την σχέση (9.10).

$$nR_i = 1 - \frac{\sum_{j=1}^N jI(x_j) - \sum_{j=1}^{|G_i|} j}{|G_i|(N - |G_i|)} \quad (9.10)$$

όπου N ο αριθμός των πρώτων σχετικών εικόνων πάνω στις οποίες υπολογίζεται το μέτρο.

9.2 Αποτελέσματα

9.2.1 Πειράματα με Σύνολο Βίντεο του TRECVID

Τα βίντεο που δόθηκαν για το TRECVID κατά την αρχική φάση της εκπαίδευσης ήταν 110 στον αριθμό και το κάθε ένα χωρισμένο σε πλάνα. Για το κάθε πλάνο έγινε η εξαγωγή των χαρακτηριστικών καρέ κι έτσι προέκυψε ένα σύνολο από 18113 χαρακτηριστικά καρέ. Σε συνεργασία με

¹Το $I(x_i)$ ορίζεται με διαφορετικό τρόπο σε περιπτώσεις όπου το x_i είναι συνεχές.

υπόλοιπους συμμετέχοντες[29]² έγινε σχολιασμός των καρτέ αυτών κι έτσι προέκυψε το ground-truth. Ο σχολιασμός γίνεται συνολικά σε επίπεδο της εικόνας και όχι σε επίπεδο περιοχών όπως έχει ήδη ειπωθεί. Σχολιασμός γίνεται για τις έννοιες που έχουν επιλεγεί να εντοπιστούν, οι οποίες είναι χιόνι, ουρανός, έρημος, δρόμος, βλάστηση, βουνό, γραφείο, εξωτερικός χώρος και έκρηξη_φωτιά. Στον πίνακα 9.14 φαίνεται κατά το ground-truth πόσα από τα χαρακτηριστικά καρτέ απεικονίζουν την κάθε έννοια.

Έννοια	θετικά χαρ. καρτέ
c_i	πλήθος στοιχείων G_i
Έρημος (Desert)	52
Δρόμος (Road)	923
Ουρανός (Sky)	2146
Χιόνι (Snow)	112
Βλάστηση (Vegetation)	1939
Γραφείο (Office)	1419
Εξωτερικός χώρος (Outdoor)	5185
Έκρηξη_Φωτιά (Explosion_Fire)	29
Βουνό (Mountain)	97

Πίνακας 9.1: Αριθμός χαρακτηριστικών καρτέ που απεικονίζουν την κάθε έννοια

Με την κατάτμηση των χαρακτηριστικών καρτέ, η οποία έγινε με ένα εργαλείο κατάτμησης με βάση το χρώμα ρυθμισμένο ώστε να δίνει μεγάλες περιοχές, προκύπτει ένας συνολικός αριθμός από 345994 περιοχές. Έτσι έχουμε την περίπτωση που έχει ήδη αναφερθεί με ένα πολύ μεγάλο αριθμό περιοχών για την κατασκευή του θησαυρού. Τελικά οι περιοχές που επιλέγονται να χρησιμοποιηθούν με τον αλγόριθμο για clustering και με σκοπό την κατασκευή του θησαυρού είναι ένα υποσύνολο των συνόλων εκπαίδευσης που πρόκειται να δημιουργηθούν για την εκπαίδευση των ανιχνευτών και περιγράφονται στην συνέχεια. Αρχικά συμπεριλαμβάνονται περιοχές από εικόνες που ανήκουν στα G_i σύνολα για κάθε έννοια. Επιλέγονται οι περιοχές από όλα τα χαρακτηριστικά καρτέ για έννοιες με μικρό G_i , ενώ από μέρος των χαρακτηριστικών καρτέ για μεγάλα G_i . Στην συνέχεια συμπεριλαμβάνεται και ίσος αριθμός περιοχών με τον αριθμό των περιοχών από τα θετικά καρτέ, οι οποίες αντιστοιχούν σε εικόνες που επιλέχθηκαν τυχαία για να αναπαραστήσουν και τις οποιεσδήποτε άλλες έννοιες.

Το μέγεθος του θησαυρού N_T επιλέχθηκε να είναι 100 λόγω του μεγάλου συνόλου εικόνων και το N_L της τεχνικής LSA επιλέχθηκε 70 μειώνοντας έτσι αντίστοιχα την διάσταση των διανυσμάτων. Συνεπώς η διάσταση των διανυσμάτων αναπαράστασης είναι 100, καθώς 100 είναι και οι τύποι περιοχής, ενώ στα πειράματα που χρησιμοποιήθηκε η τεχνική LSA η διάσταση αυτή μειώθηκε σε 70 καθώς τα διανύσματα μεταφέρθηκαν στον χώρο των εννοιών. Για να γίνει μία αξιολόγηση της τεχνικής με το δεδομένο σύνολο ανάπτυξης για το οποίο σχηματίστηκε και σχολιασμός σε επίπεδο ολόκληρης της εικόνας, θα πρέπει να χωρισθεί σε δύο υποσύνολα για κάθε έννοια τα οποία θα αποτελέσουν το

²έγινε στα πλαίσια του project K-SPACE, όπως και η εξαγωγή χαρακτηριστικών καρτέ και η κατάτμηση τους

Συνολικός Αριθμός Βίντεο	N_V	110
Συνολικός Αριθμός Πλάνων	N_S	18113
Συνολικός Αριθμός Χαρ. Καρέ	N_K	18113
Συνολικός Αριθμός Περιοχών	N_R	345994
Συνολικός Αριθμός Διαν. Χαρακτ/ων	N_F	345994
Μέγεθος Οπτικού Θησαυρού	N_T	100
Μέγεθος Διαν. Μετά Από LSA	N_L	70

Πίνακας 9.2: Στοιχεία πειράματος TRECVID

σύνολο εκπαίδευσης και σύνολο ελέγχου για κάθε ανιχνευτή, έστω ότι αυτά τα δύο συμβολίζονται με TR_i και TE_i αντίστοιχα για την έννοια c_i . Ισχύει ότι:

$$TR_i \cup TE_i \subseteq K \quad (9.11)$$

$$TR_i \cap TE_i = \emptyset \quad (9.12)$$

Πειράματα για TRECVID χωρίς LSA

Συνολικά διεξήχθησαν τρεις σειρές πειραμάτων για την κάθε έννοια χωρίς την χρήση LSA.

Πειράματα με Σύνολο Εκπαίδευσης: Χρησιμοποιούνται σύνολα εκπαίδευσης με διαφορετικό αριθμό στοιχείων από το σύνολο \bar{G}_i για κάθε έννοια c_i και τελικά επιλέγεται το καταλληλότερο.

Πειράματα με Τιμή Κατωφλίσωσης: Για κάθε ανιχνευτή μεταβάλλεται η τιμή του κατωφλίου h_i και επιλέγεται εκείνο το οποίο δίνει βέλτιστα μέτρα precision και recall.

Πειράματα με Σύνολο Ελέγχου: Ο κάθε ανιχνευτής χρησιμοποιείται σε σύνολα εκπαίδευσης με συνεχώς αυξανόμενα στοιχεία για να φανεί πως η απόδοση του μεταβάλλεται.

Πειράματα με Σύνολο Εκπαίδευσης Για την κατασκευή των συνόλων εκπαίδευσης και ελέγχου το 70% του G_i καταχωρείται στο TR_i . Χρησιμοποιήθηκαν 5 διαφορετικά σύνολα εκπαίδευσης, όπου αν λ ο λόγος του πλήθους των στοιχείων του TR_i που προέρχονται από το \bar{G}_i προς το πλήθος των στοιχείων από το από το G_i , τότε στο κάθε ένα από αυτά αντιστοιχούν οι τιμές $\lambda = 1$, $\lambda = 2$, $\lambda = 3$, $\lambda = 4$ και $\lambda = 5$. Αυτό σημαίνει ότι τα σύνολα αυτά περιέχουν 50%, 66.6%, 75%, 80% και 83.3% αντίστοιχα ποσοστό στοιχείων από το σύνολο \bar{G}_i , δηλαδή αρνητικών χαρακτηριστικών καρέ στην απεικόνιση των εννοιών. Τα σύνολα ελέγχου TE_i που χρησιμοποιήθηκαν αποτελούνται από το 30% του G_i και από ισάριθμα στοιχεία από το \bar{G}_i , δηλαδή ισχύει $\lambda = 1$. Στην σχέση (9.13) δίνεται ο υπολογισμός του λ για το σύνολο εκπαίδευσης της έννοιας c_i . Ομοίως υπολογίζεται και για το σύνολο ελέγχου.

$$\lambda_i = \frac{|\{k \in K : k \in TR_i \cap \bar{G}_i\}|}{|\{k \in K : k \in TR_i \cap G_i\}|} \quad (9.13)$$

Το μέτρο που χρησιμοποιήθηκε είναι η μέση ακρίβεια (average precision) της σχέσης (9.6) με n ίσο με ολόκληρο το σύνολο ελέγχου για κάθε έννοια. Έτσι γίνεται επιλογή για κάθε ανιχνευτή του

αποδοτικότερου συνόλου εκπαίδευσης. Επιλέχθηκε να κατασκευαστούν αυτά τα 5 σύνολα ελέγχου καθώς για μερικές έννοιες όλο το development set του TRECVID δεν θα επέτρεπε περισσότερα χαρακτηριστικά καρέ να χρησιμοποιηθούν στο σύνολο εκπαίδευσης διότι πρέπει να χρησιμοποιηθούν και αρκετά χαρακτηριστικά καρέ στο σύνολο ελέγχου προκειμένου να δημιουργηθεί σύνολο της τάξεως εκείνων που δίνει και το TRECVID για έλεγχο. Τα αποτελέσματα δίνονται στον πίνακα 9.3, όπου για την έννοια εξωτερικός χώρος χρησιμοποιήθηκαν μόνο 3 σύνολα εκπαίδευσης, με $\lambda=1$, $\lambda=2$ και $\lambda=3$ γιατί τα θετικά χαρακτηριστικά καρέ ήταν πολλά.

	$\lambda=1$	$\lambda=2$	$\lambda=3$	$\lambda=4$	$\lambda=5$
Έρημος	0.6593	0.6994	0.3653	0.4775	0.6634
Δρόμος	0.5944	0.6091	0.5954	0.6062	0.6957
Ουρανός	0.6791	0.723	0.6883	0.7197	0.7369
Χιόνι	0.9144	0.9054	0.9293	0.9174	0.9504
Βλάστηση	0.7175	0.7731	0.7649	0.7522	0.7802
Γραφείο	0.6337	0.7073	0.7382	0.7077	0.7235
Εξωτερικός χώρος	0.6832	0.6842	0.6978	0	0
Έκρηξη_Φωτιά	0.3879	0.3679	0.3485	0.647	0.3827
Βουνό	0.6878	0.6119	0.5458	0.625	0.7662

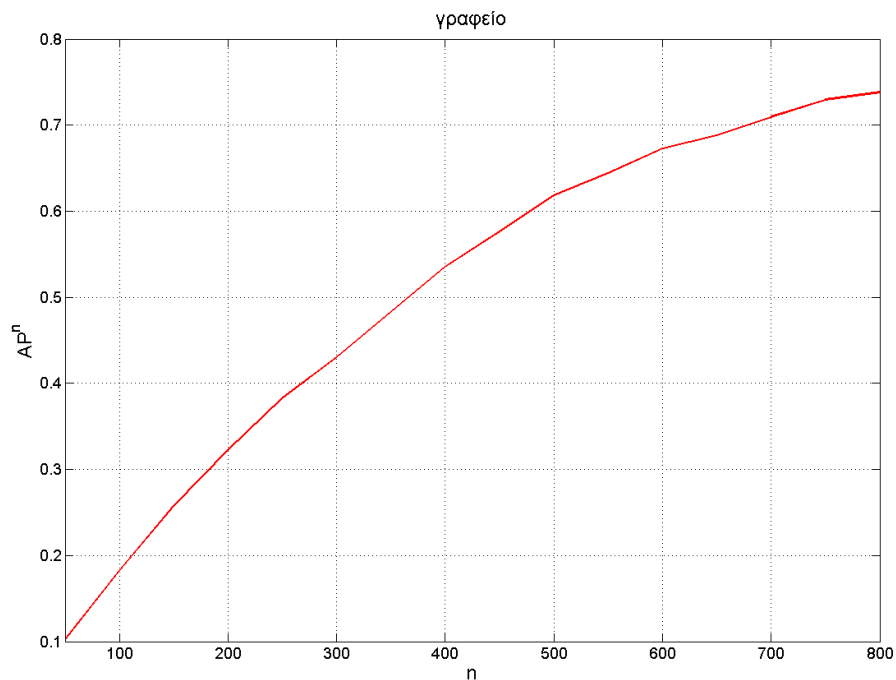
Πίνακας 9.3: Μέση ακρίβεια για σύνολα εκπαίδευσης με διαφορετικό λόγο λ , με τονισμένη γραμματοσειρά είναι οι υψηλότερες τιμές

Μέσα από αυτή την διαδικασία επιλέγονται τα καταλληλότερα σύνολα εκπαίδευσης για κάθε ανιχνευτή. Συνοπτικά τα χαρακτηριστικά καρέ που αποτελούν το κάθε σύνολο εκπαίδευσης και ελέγχου φαίνονται στον πίνακα 9.4.

c_i	TR_i		TE_i	
	G_i	\bar{G}_i	G_i	\bar{G}_i
Έρημος	36	72	16	16
Δρόμος	646	3230	277	277
Ουρανός	1502	7510	644	644
Χιόνι	78	390	34	34
Βλάστηση	1357	6785	582	582
Γραφείο	993	2979	426	426
Εξωτερικός χώρος	3629	10887	1556	1556
Έκρηξη_Φωτιά	20	80	9	9
Βουνό	68	340	29	29

Πίνακας 9.4: Αριθμός χαρακτηριστικών καρέ που απεικονίζουν ή όχι την κάθε έννοια για τα σύνολα εκπαίδευσης και ελέγχου

Θεωρητικά ένα νευρωνικό δίκτυο ταξινόμησης ανάμεσα σε δύο κατηγορίες πρέπει να εκπαιδεύεται με ισάριθμα πρότυπα από την κάθε μία κατηγορία, ώστε να “μαθαίνει” εξίσου και τις δύο. Στην



Σχήμα 9.1: Διάγραμμα μέσης ακρίβειας καθώς μεταβάλλεται το παράθυρο υπολογισμού της για την έννοια γραφείο

προκειμένη όμως περίπτωση υπάρχει ένα σύνολο δεδομένων με μεγάλη ποικιλομορφία. Έτσι για κάθε έννοια υπάρχει η κατηγορία εκείνη με τα πρότυπα (εικόνες) που απεικονίζουν την έννοια και η κατηγορία με εκείνες που δεν την απεικονίζουν. Η δεύτερη όμως κατηγορία περιέχει εικόνες με ποικίλα θέματα άρα και με ποικίλα οπτικά χαρακτηριστικά. Συνεπώς απαιτούνται περισσότερα πρότυπα από την δεύτερη κατηγορία για να “μάθει” το νευρωνικό δίκτυο και εκείνη. Αυτό αποδεικνύεται και στην πράξη με τα παραπάνω πειράματα όπου τα σύνολα εκπαίδευσης που αποδίδουν καλύτερα είναι εκείνα με περισσότερα αρνητικά χαρακτηριστικά καρέ από ότι θετικά.

Μία παρατήρηση που αξίζει να γίνει για το μέτρο της μέσης ακρίβειας που χρησιμοποιήθηκε είναι κοιτώντας το διάγραμμα του σχήματος 9.1, όπου παρουσιάζεται η μέση ακρίβεια ως προς το παράθυρο πάνω στο οποίο υπολογίζεται, δηλαδή το n . Καθώς αυξάνεται το n είναι προφανές ότι η μέση ακρίβεια θα αυξάνεται, διότι θα συναντιούνται όλο και περισσότερες σχετικές εικόνες. Μεγάλη κλίση καμπύλης υποδηλώνει πως εντοπίζονται πολλές σχετικές εικόνες σε συνεχόμενες θέσεις, ενώ πιο μικρή ότι μεταξύ των σχετικών παρεμβάλλονται και πολλές μη σχετικές.

Πειράματα με Τιμή Κατωφλίσωσης Χρησιμοποιούνται τα σύνολα εκπαίδευσης που επιλέχθηκαν από την πρώτη σειρά πειραμάτων και σαν σύνολα ελέγχου για κάθε έννοια χρησιμοποιούνται όλα

τα υπόλοιπα χαρακτηριστικά καρέ του K . Δηλαδή ισχύει:

$$TR_i \cup TE_i = K \quad (9.14)$$

$$TR_i \cap TE_i = \emptyset \quad (9.15)$$

Τα νευρωνικά δίκτυα εκπαιδεύονται με τα επιλεγμένα σύνολα εκπαίδευσης και σαν είσοδος μετά δίνεται το σύνολο ελέγχου. Πάνω στα ποσοστά πεποίθησης τα οποία για κάθε χαρακτηριστικό καρέ εξάγονται γίνεται μετατροπή με χρήση κατωφλίου σε δυαδικό $\{0,1\}$. Πάνω στην δυαδική αυτή έξοδο εφαρμόζονται τα μέτρα precision-recall. Το πείραμα γίνεται για τιμές κατωφλίου από 0 έως 0.9 με βήμα 0.1. Έτσι δημιουργούνται τα διαγράμματα σαν εκείνα του σχήματος 9.2 και τελικά η τιμή κατωφλίου που επιλέγεται είναι εκείνη η οποία δίνει τιμές πιο κοντά στο σημείο της καμπύλης όπου precision=recall. Είναι επιθυμητό και τα δύο μέτρα αυτά να είναι όσο τον δυνατό καλύτερα με την ίδια σημασία στο κάθε ένα και για αυτό το κατώφλι επιλέγεται με τον συγκεκριμένο τρόπο. Επίσης στον πίνακα 9.5 παρουσιάζονται οι τιμές για τα δύο αυτά μέτρα όσο το κατώφλι αυξάνεται για τον ανιχνευτή της έννοια *ουρανός*. Τελικά η τιμή κατωφλίου που επιλέγεται για τον ανιχνευτή της έννοιας *βλάστηση* είναι 0.4, το οποίο φαίνεται και από το σχήμα αλλά και από τον πίνακα των τιμών. Για την έννοια *ουρανός* επιλέγεται κατώφλι 0.3 ενώ τα κατώφλια για όλες τις έννοιες φαίνονται στον πίνακα 9.6.

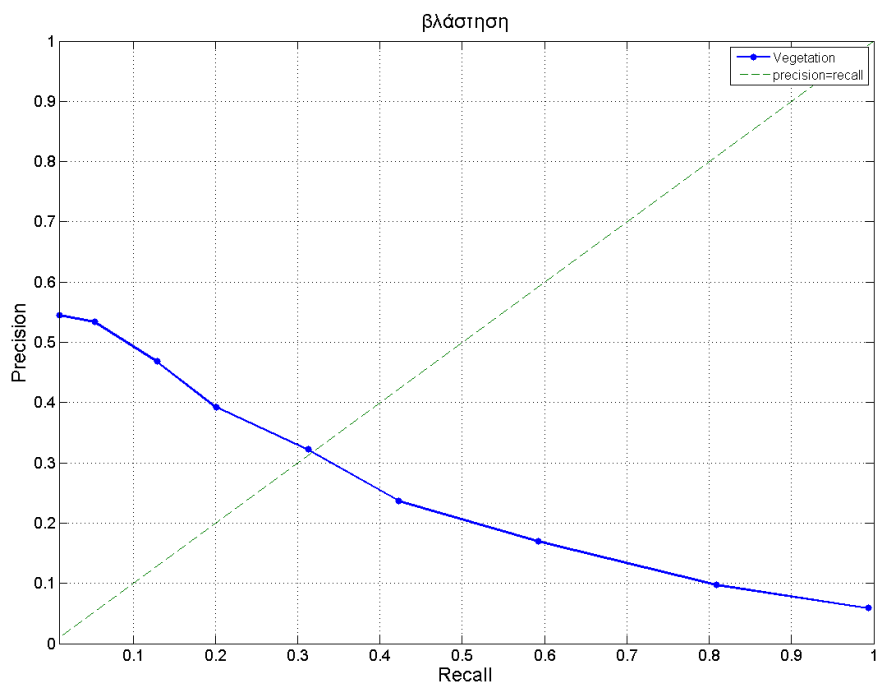
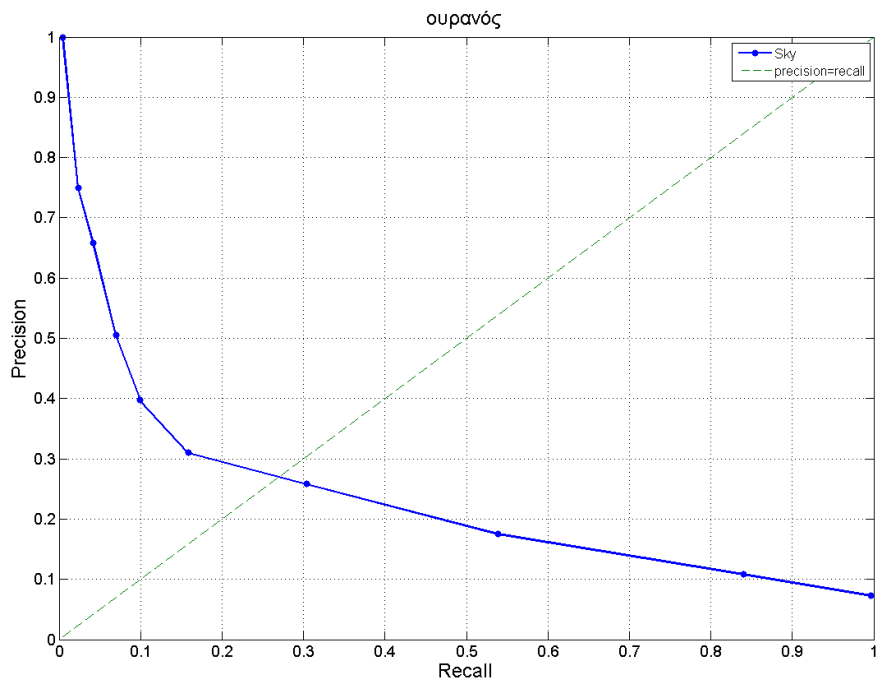
κατώφλι	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Precision	0.05	0.09	0.17	0.23	0.32	0.39	0.46	0.53	0.54	0.33
Recall	0.99	0.80	0.59	0.42	0.31	0.20	0.12	0.05	0.01	0.01

Πίνακας 9.5: Ζεύγη precision-recall μεταβάλλοντας το κατώφλι, για την έννοια Βλάστηση

Όσο πιο μεγάλη (κοντά στο 1) τιμή έχει ένα κατώφλι τόσο πιο μεγάλο είναι το μέτρο precision και πιο μικρό το μέτρο recall. Αυτό διότι επιλέγονται σαν θετικά χαρακτηριστικά καρέ ως προς την απεικόνιση της κάθε έννοιας εκείνα τα οποία έχουν υψηλά ποσοστά πεποίθησης κι έτσι υπάρχει μεγαλύτερη σιγουριά ότι εκείνα θα απεικονίζουν όντως την έννοια, για αυτό και το υψηλό precision. Με αυτό όμως τον τρόπο υπάρχει περισσότερη αυστηρότητα ως προς το τι θεωρείται θετικό και συνεπώς ένας μικρός μόνο αριθμός εικόνων επιλέγεται και αυτό οδηγεί σε χαμηλό recall εφόσον μέσα σε αυτόν τον μικρό αριθμό ανακτημένων εικόνων θα υπάρχει και μικρός αριθμός σχετικών εικόνων ως προς όλο το σύνολο των σχετικών εικόνων. Για μικρές (κοντά στο 0) τιμές κατωφλίου συμβαίνει το αντίθετο, δηλαδή χαμηλό precision και υψηλό recall. Συνεπώς η αποτελεσματικότητα του κάθε ανιχνευτή εξαρτάται άμεσα και σε μεγάλο βαθμό από την τιμή αυτή του κατωφλίου. Ο τρόπος επιλογής τιμής κατωφλίου που δίνει τιμές κοντά στο σημείο τομής της καμπύλης *precision-recall* με την καμπύλη *precision=recall* εξασφαλίζει την καλύτερη απόδοση συνυπολογίζοντας από κοινού και εξίσου τα δύο μέτρα.

Έρημος	Δρόμος	Ουρανός	Χιόνι	Βλάστηση	Γραφείο	Εξ.Χώρος	Έκ.-Φωτιά	Βουνό
0.8	0.5	0.3	0.6	0.4	0.5	0.3	0.2	0.8

Πίνακας 9.6: Κατώφλια για όλους τους ανιχνευτές, χωρίς LSA

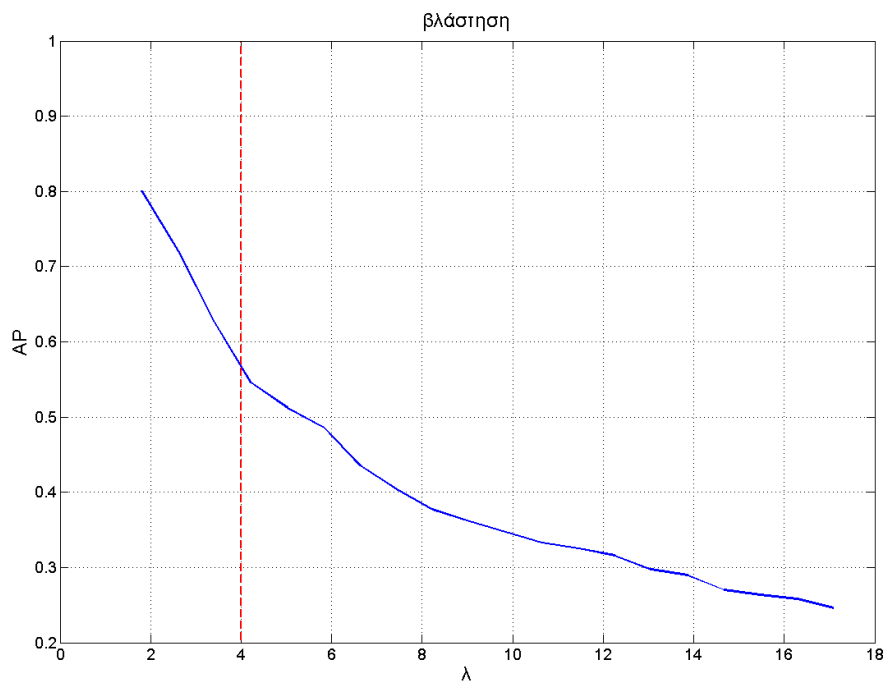
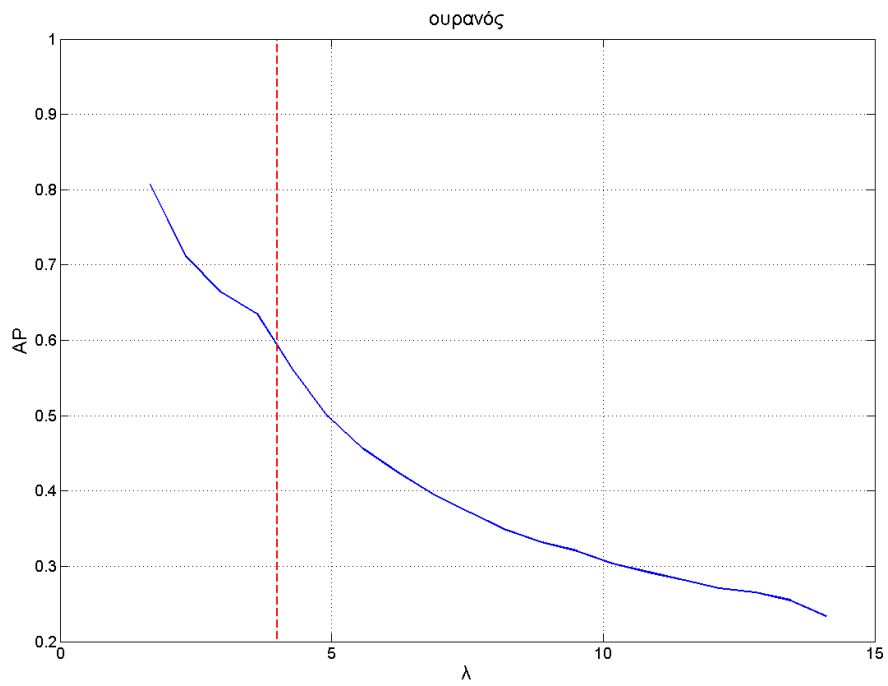


Σχήμα 9.2: Διαγράμματα precision-recall καθώς η τιμή κατωφλίου μεταβάλλεται

Πειράματα με Σύνολο Ελέγχου Τα σύνολα ελέγχου τώρα αποτελούνται όπως και νωρίτερα από το 30% του G_i και συμπληρώνονται κάθε φορά με ένα μέρος των αρνητικών χαρακτηριστικών καρέ (\bar{G}_i) που δεν ανήκουν στο TR_i . Αυτό γίνεται με ένα βήμα μέχρι να δημιουργηθεί το σύνολο ελέγχου που θα περιέχει και όλα τα k για τα οποία ισχύει ότι: $\{k \in \bar{G}_i, k \notin TR_i\}$. Κάθε φορά δηλαδή αυξάνεται ο λόγος λ για το σύνολο ελέγχου. Όλα τα ενδιαμέσα σύνολα ελέγχου δίνονται σαν είδος στους ανιχνευτές και στην έξοδο τους υπολογίζεται κάθε φορά η μέση ακρίβεια με παράθυρο όλο το σύνολο. Κατασκευάζονται έτσι διαγράμματα μέσης ακρίβειας ως προς τον λόγο λ του συνόλου ελέγχου. Στα διαγράμματα αυτά φαίνεται πως η μέση ακρίβεια μειώνεται γενικά καθώς μεγαλώνει το σύνολο ελέγχου, αλλά πιο συγκεκριμένα καθώς αυξάνεται ο λόγος των αρνητικών χαρακτηριστικών καρέ προς τα θετικά.

Με την χρήση ενός ανιχνευτή εννοιών πάνω σε ένα σύνολο χαρακτηριστικών καρέ αντιστοιχίζεται κάθε καρέ με μία δυαδική τιμή 1, 0 που αντιστοιχεί σε απεικόνιση ή μη της έννοιας. Από εκείνα τα καρέ τα οποία ανιχνεύθηκαν ότι εντοπίζουν την έννοια, δηλαδή έξοδος με τιμή 1, είναι σχεδόν βέβαιο ότι μερικά από αυτά δεν θα απεικονίζουν πραγματικά την συγκεκριμένη έννοια. Αυτό τελικά οδηγεί σε χαμηλότερη μέση ακρίβεια, το οποίο συμβαίνει καθώς οι λανθασμένοι αυτοί εντοπισμοί να γίνονται με ποσοστό πεποίθησης μεγαλύτερο από καρέ που πραγματικά απεικονίζουν την έννοια. Έτσι τα θετικά αυτά καρέ να εντοπίζονται σε χαμηλότερες θέσεις στην ταξινομημένη λίστα με βάση το ποσοστό πεποίθησης και να συνεισφέρουν λιγότερο στο άθροισμα όπως φαίνεται και στον τύπο υπολογισμού της μέσης ακρίβειας (9.7). Όταν τα αρνητικά χαρακτηριστικά καρέ είναι περισσότερα αυτό που συμβαίνει είναι να είναι και περισσότεροι αυτοί οι λανθασμένοι εντοπισμοί μεταξύ των πραγματικών εντοπισμών και η μέση ακρίβεια να πέφτει, γεγονός το οποίο φαίνεται στα διαγράμματα του σχήματος 9.3.

Τελικά Αποτελέσματα για TRECVID χωρίς LSA Τελικά το νευρωνικό δίκτυο για τον ανιχνευτή κάθε έννοιας εκπαιδεύτηκε με το σύνολο εκπαίδευσης το οποίο επιλέχθηκε στην πρώτη σειρά πειραμάτων. Χρησιμοποιήθηκε επίσης για τον κάθε ένα το κατώφλι που επιλέχθηκε στην δεύτερη σειρά πειραμάτων προκειμένου να υπολογιστούν τα μέτρα precision-recall πάνω στην δυαδική έξοδο. Πάνω σε σύνολο ελέγχου το οποίο αποτελείται από όλα τα χαρακτηριστικά καρέ τα οποία δεν ανήκουν στο σύνολο εκπαίδευσης υπολογίστηκαν και τα μέτρα μέσης ακρίβειας (average precision), r-ακρίβειας (r-precision) και το κανονικοποιημένο μέτρο ανάκτησης (normalized recall). Το μέτρο Average Precision υπολογίστηκε πάνω σε παράθυρο ίσο με το μισό του μεγέθους του συνόλου ελέγχου, διότι σε περίπτωση που υπολογιζόταν σε σταθερό παράθυρο αυτό θα έφερνε σε μειονεκτική θέση ανιχνευτές για έννοιες με λίγα θετικά καρέ. Το μέτρο r-precision υπολογίστηκε με ένα παράθυρο ανεκτικότητας όπως ορίστηκε σε προηγούμενη ενότητα(9.1). Όλα τα αποτελέσματα συνοψίζονται στον πίνακα 9.7. Η χρήση ενός τέτοιου συνόλου είναι λογικό να έχει χαμηλά αποτελέσματα καθώς στα περισσότερα σύνολα έχουμε ένα μεγάλο αριθμό αρνητικών χαρακτηριστικών καρέ με λίγα θετικά. Σε μερικές περιπτώσεις (όπως για τις έννοιες *έκρηξη_φωτιά*, *έρημος*, *χιόνι* και *βουνό*) αυτή η διαφορά είναι πολύ μεγάλη, εφόσον τα συνολικά θετικά χαρακτηριστικά καρέ ήταν πολύ λίγα. Για αυτό τον λόγο χρησιμοποιήθηκε κι ένα διαφορετικό σύνολο ελέγχου το οποίο αποτελείται από τον ίδιο αριθμό θετικών χαρακτηριστικών καρέ με το προηγούμενο σύνολο ελέγχου (τα οποία είναι και τα θετικά που φαίνονται στο πίνακα 9.14) και από αρνητικά χαρακτηριστικά καρέ τετραπλάσια σε αριθμό από τα θετικά. Χρησιμοποιήθηκε δηλαδή τελικά σύνολο ελέγχου με $\lambda = 4$. Έτσι ο πίνακας αποτελεσμάτων



Σχήμα 9.3: Διαγράμματα μέσης ακρίβειας καθώς μεγαλώνει το σύνολο ελέγχου και μεγαλώνει το λ , μέχρι $\lambda=4$ θεωρείται λογικό για σύνολο ελέγχου ενός συστήματος

(9.12) για το συγκεκριμένο σύνολο είναι και πιο κατάλληλος για να αξιολογηθεί η απόδοση του κάθε ανιχνευτή. Για να κατασκευαστεί ένα τέτοιο σύνολο ελέγχου για την έννοια εξωτερικός χώρος της οποίας τα θετικά καρέ ήταν πολλά και το τετραπλάσιο τους ξεπερνά το σύνολο όλων των καρέ, χρησιμοποιήθηκαν λιγότερα θετικά καρέ από τα συνολικά και τελικά τετραπλάσιος αριθμός αρνητικών κι έτσι δημιουργείται το σύνολο με $\lambda=4$. Ενώ για όλες τις έννοιες τα αποτελέσματα του πίνακα 9.7 είναι με σύνολα ελέγχου όπου το λ είναι πολύ μεγάλο, μόνο στην συγκεκριμένη έννοια εξωτερικός χώρος το λ είναι και μικρότερο του 4, λόγω της έλλειψης αρνητικών χαρακτηριστικών καρέ. Έτσι φαίνεται τελικά στον πίνακα 9.12 τα ποσοστά να πέφτουν αντί να ανεβαίνουν διότι τα θετικά εδώ είναι λιγότερα αλλά μόνο για αυτή την έννοια.

c_i	P	R	AP	rP	nR
Βλάστηση	0.322	0.313	0.232	0.154	0.771
Δρόμος	0.045	0.047	0.043	0.054	0.695
Έκρηξη-Φωτιά	0.000	0.000	0.001	0.000	0.667
Ουρανός	0.258	0.304	0.214	0.159	0.755
Χιόνι	0.013	0.412	0.008	0.000	0.734
Γραφείο	0.117	0.157	0.072	0.085	0.732
Έρημος	0.003	0.313	0.064	0.016	0.627
Εξωτερικός χώρος	0.683	0.510	0.515	0.475	0.731
Βουνό	0.003	0.379	0.037	0.009	0.628

Πίνακας 9.7: Τελικά αποτελέσματα πειραμάτων χωρίς την χρήση LSA, με σύνολο ελέγχου όλες τις εικόνες

c_i	P	R	AP	rP	nR
Βλάστηση	0.643	0.312	0.460	0.344	0.782
Δρόμος	0.295	0.046	0.280	0.228	0.709
Έκρηξη-Φωτιά	0.291	0.777	0.182	0.250	0.592
Ουρανός	0.571	0.304	0.436	0.367	0.778
Χιόνι	0.777	0.411	0.460	0.250	0.798
Γραφείο	0.446	0.157	0.318	0.284	0.734
Έρημος	0.333	0.312	0.287	0.250	0.603
Εξωτερικός χώρος	0.425	0.514	0.361	0.292	0.733
Βουνό	0.444	0.137	0.241	0.250	0.618

Πίνακας 9.8: Τελικά αποτελέσματα πειραμάτων χωρίς την χρήση LSA, με σύνολο ελέγχου αποτελούμενο από 20% θετικά και 80% αρνητικά χαρακτηριστικά καρέ ($\lambda=4$)

Πειράματα για TRECVID με LSA

Τα παραπάνω πειράματα διεξήχθησαν με σύνολα όπου κάθε χαρακτηριστικό καρέ αναπαρίσταται από το διάνυσμα αναπαράστασης το οποίο και έχει διάσταση όση και το μέγεθος του θησαυρού, 100 στην προκειμένη περίπτωση. Όσοι δηλαδή και οι τύποι περιοχής. Στα σύνολα εκπαίδευσης που επιλέχθηκαν εφαρμόζεται η τεχνική της Λανθάνουσας Σημασιολογικής Ανάλυσης και τα διανύσματα αναπαράστασης μεταφέρονται έτσι σε έναν άλλο χώρο, τον χώρο τον “εννοιών” και η διάσταση τους μειώνεται σε 70 από 100 στοιχεία. Η ίδια τεχνική εφαρμόστηκε και στο σύνολο ελέγχου το οποίο αποτελείται από όλα τα υπόλοιπα χαρακτηριστικά καρέ. Έτσι στην έξοδο του κάθε ανιχνευτή που αφορά το αντίστοιχο σύνολο εκπαίδευσης εφαρμόστηκαν τα διάφορα μέτρα για να γίνει η σύγκριση με εκείνα στα οποία δεν εφαρμόστηκε LSA.

Στο σχήμα 9.4 (πάνω) φαίνονται οι καμπύλες precision-recall για την έννοια βλάστηση καθώς το κατώφλι μεταβάλλεται. Οι δύο καμπύλες αντιστοιχούν σε αποτελέσματα χωρίς αλλά και με την χρήση LSA. Η καμπύλη που αντιστοιχεί στα αποτελέσματα χωρίς την χρήση της τεχνικής βρίσκεται συνεχώς πάνω από την άλλη γεγονός που αποδεικνύει ότι για όλες τις τιμές του κατωφλίου αποδίδει καλύτερα χωρίς LSA ο ανιχνευτής. Το συμπέρασμα αυτό επιβεβαιώνει και το διάγραμμα του σχήματος 9.4 (κάτω) για την μέση ακρίβεια καθώς το παράθυρο υπολογισμού μεγαλώνει. Η καμπύλη που αντιστοιχεί στην μη χρήση LSA είναι και πάλι από πάνω. Τα διαγράμματα αυτά συνεπώς δείχνουν ότι η τεχνική LSA δεν δίνει καλύτερα αποτελέσματα για τον ανιχνευτή της συγκεκριμένης όμως έννοιας. Ενώ για τον ανιχνευτή της έννοιας εξωτερικός χώρος αυτό που φαίνεται στο διάγραμμα της μέσης ακρίβειας (σχήμα 9.5) είναι πως οι σωστά εντοπισμένες εικόνες βρίσκονται στις πρώτες θέσεις της διατεταγμένης λίστας και για αυτό η μέση ακρίβεια είναι μεγαλύτερη για τα μικρά παράθυρα με την χρήση LSA. Παρόμοια και στην καμπύλη precision-recall φαίνεται κάτι παρόμοια, πως σωστά εντοπισμένες εικόνες δίνονται με μεγάλα ποσοστά πεποίθησης. Το συμπέρασμα αυτό πως η τεχνική LSA βελτιώνει για μερικούς από τους ανιχνευτές τις θέσεις στις οποίες οι σωστοί εντοπισμοί συμβαίνουν επιβεβαιώνεται από τα διαγράμματα μέσης ακρίβειας του σχήματος 9.6.

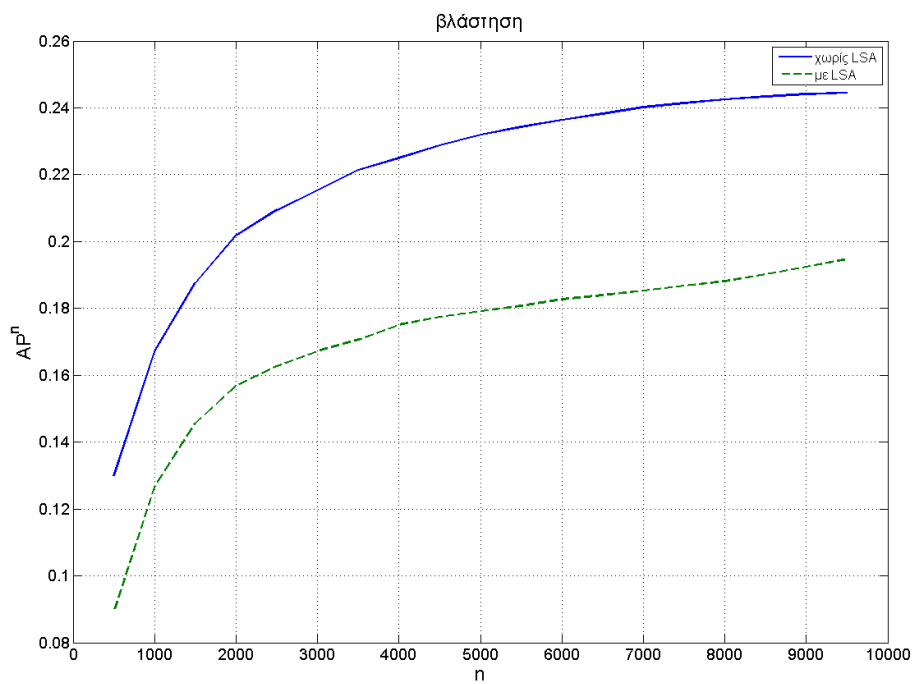
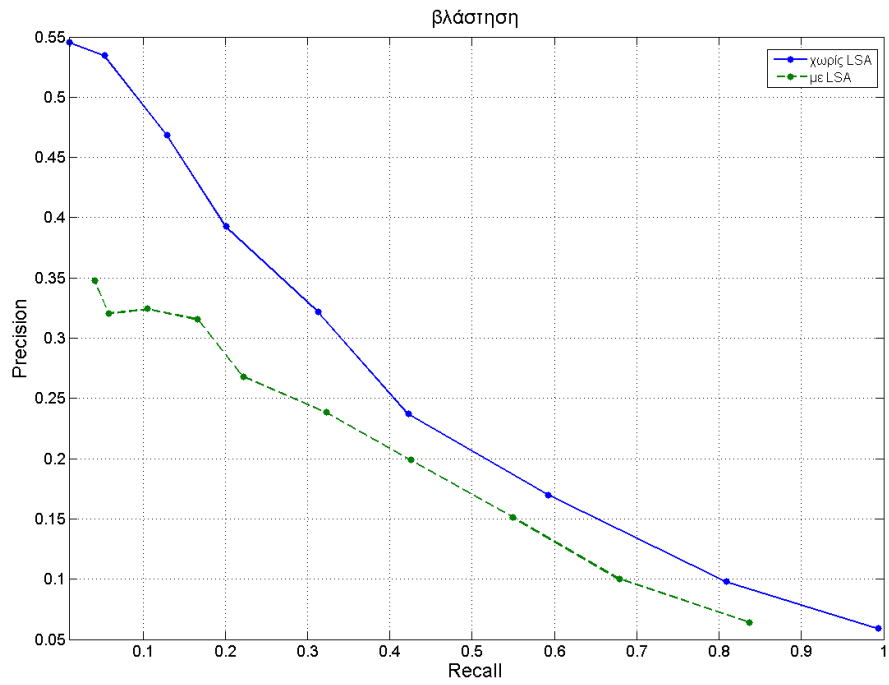
Η τεχνική LSA αποδεικνύεται χρήσιμη για μερικούς από τους ανιχνευτές όσον αφορά το γεγονός ότι ανάμεσα στις πρώτες εικόνες με το μεγαλύτερο ποσοστό πεποίθησης βρίσκονται περισσότεροι σωστοί εντοπισμοί από ότι χωρίς την χρήση LSA. Αυτό συμβαίνει καθώς η τεχνική αυτή λαμβάνει υπόψη τις λανθάνουσες σχέσεις μεταξύ των περιοχών και αυτές οι συσχετίσεις μπορεί να είναι πιο μεγάλες για μερικές έννοιες από κάποιες άλλες. Έτσι για τις έννοιες με υψηλές συσχετίσεις μεταξύ των περιοχών η τεχνική αποδίδει καλύτερα. Αυτό που συμβαίνει με την τεχνική αυτή είναι ενώ πριν υπήρχε η εμφάνιση του κάθε τύπου περιοχής να περιγράφει την εικόνα, τώρα αντί των τύπων περιοχής εντοπίζονται συσχετίσεις μεταξύ των περιοχών και χρησιμοποιούνται για την περιγραφή των εικόνων.

Οι τιμές των κατωφλίων επιλέγονται και πάλι για κάθε ανιχνευτή με τον ίδιο τρόπο και είναι εκείνες που φαίνονται στον πίνακα 9.9.

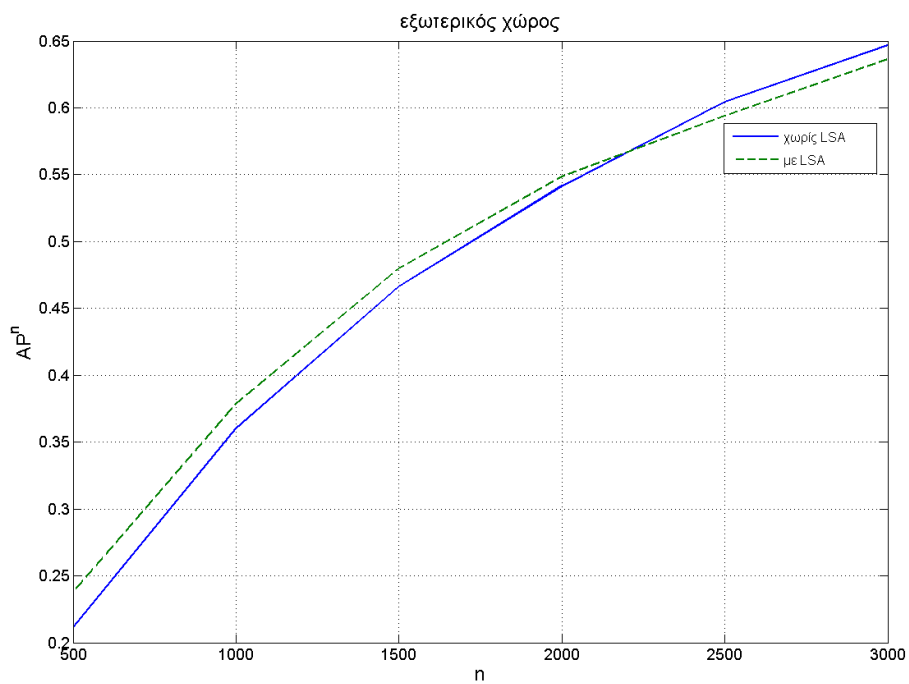
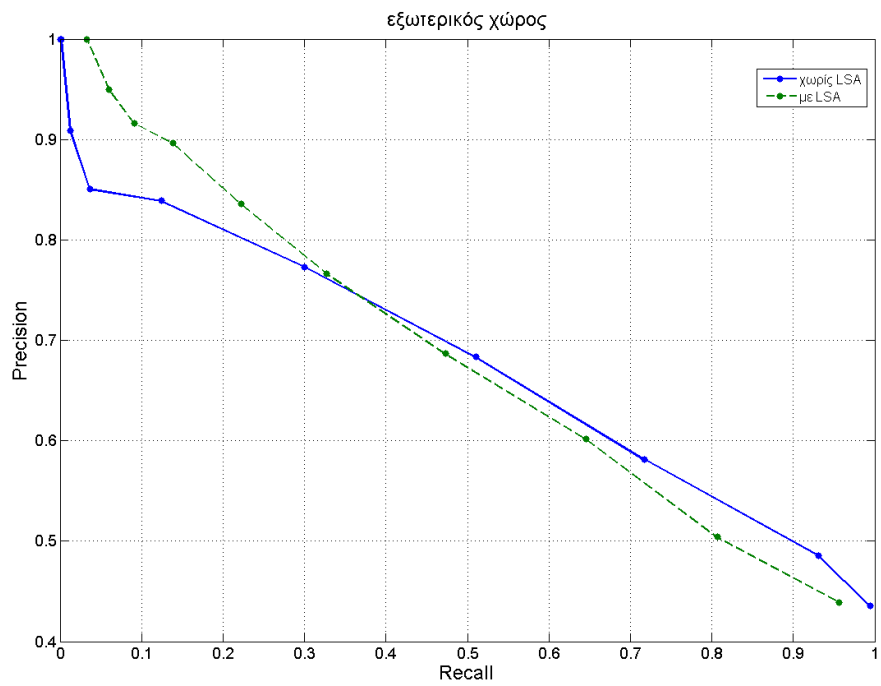
Έρημος	Δρόμος	Ουρανός	Χιόνι	Βλάστηση	Γραφείο	Εξ.Χώρος	Έκ.-Φωτιά	Βουνό
0.8	0.8	0.4	0.8	0.5	0.8	0.2	0.8	0.8

Πίνακας 9.9: Κατώφλια για όλους τους ανιχνευτές, με LSA

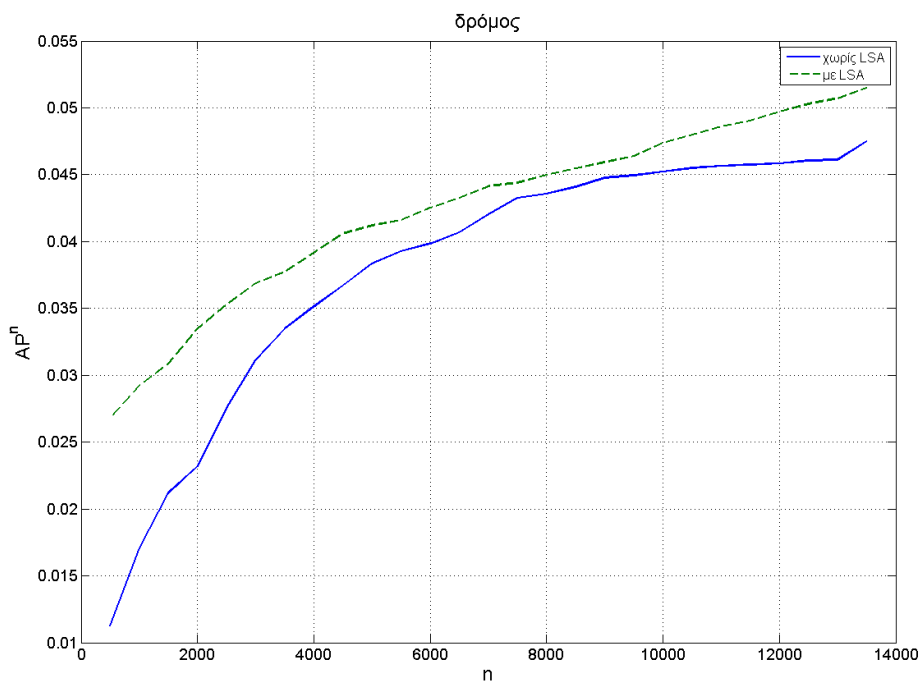
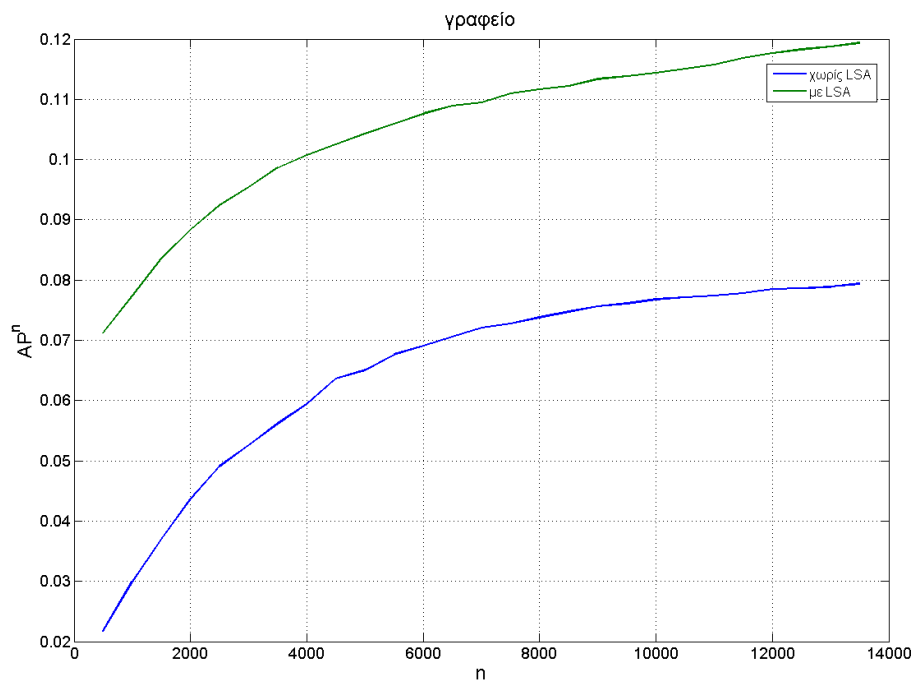
Τέλος στους πίνακες 9.10 και 9.11 παρουσιάζονται όλα τα μέτρα τα οποία χρησιμοποιήθηκαν και για τους ανιχνευτές χωρίς LSA, για σύνολο ελέγχου με όλα τα αρνητικά χαρακτηριστικά καρέ και



Σχήμα 9.4: Διαγράμματα precision-recall και Average Precision καθώς αυξάνεται το μέγεθος παραθύρου, χωρίς και με LSA, για την έννοια βλάστηση



Σχήμα 9.5: Διαγράμματα precision-recall και Average Precision καθώς αυξάνεται το μέγεθος παραθύρου, χωρίς και με LSA, για την έννοια εξωτερικός χώρος



Σχήμα 9.6: Διαγράμματα Average Precision καθώς αυξάνεται το μέγεθος παραθύρου, χωρίς και με LSA, για τις έννοιες δρόμος και γραφείο

για σύνολο ελέγχου με τετραπλάσια αρνητικά από τα θετικά αντίστοιχα. Μία τελευταία σύγκριση και σε αυτά οδηγεί στο τελικό συμπέρασμα πως η τεχνική LSA που χρησιμοποιήθηκε και αξιολογήθηκε οδηγεί σε καλύτερα αποτελέσματα εντοπισμού από εκείνα των ανιχνευτών χωρίς την χρήση της τεχνικής για μερικούς από τους ανιχνευτές. Παρατηρούνται καλύτερα αποτελέσματα σε έννοιες των οποίων οι εικόνες έχουν καλύτερα συσχετισμένες περιοχές. Για παράδειγμα η έννοια *χιόνι* αποτελείται από εικόνες, από τις οποίες οι περισσότερες καταλαμβάνονται από κατάλευκες περιοχές, έτσι παρατηρείται αισθητή βελτίωση με την τεχνική LSA. Σε αντίθετη περίπτωση έννοιες όπως *βουνό* όπου στο σύνολο των εικόνων παρατηρείται μεγαλύτερη ανομοιομορφία, καθώς δεν περιλαμβάνονται μόνο και χαρακτηριστικά τοπία βουνών, αλλά αντιθέτως και αρκετές φωτογραφίες οι οποίες απλώς έχουν ληφθεί σε ένα βουνό χωρίς όμως αυτό να γίνεται ιδιαίτερα αντιληπτό από κάποια συγκεκριμένα χαρακτηριστικά. Σαν αποτέλεσμα η τεχνική LSA αποδίδει χειρότερα στην συγκεκριμένη έννοια λόγω έλλειψης υψηλών συσχετίσεων ανάμεσα στις περιοχές του συνόλου των θετικών για την έννοια εικόνων.

c_i	P	R	AP	rP	nR
Βλάστηση	0.268	0.222	0.179	0.141	0.716
Δρόμος	0.036	0.051	0.044	0.035	0.622
Έκρηξη_Φωτιά	0.001	0.111	0.000	0.000	0.516
Ουρανός	0.288	0.207	0.184	0.136	0.693
Χιόνι	0.023	0.265	0.012	0.007	0.792
Γραφείο	0.095	0.148	0.110	0.148	0.674
Έρημος	0.001	0.438	0.063	0.016	0.549
Εξωτερικός χώρος	0.601	0.646	0.522	0.457	0.707
Βουνό	0.003	0.172	0.001	0.000	0.531

Πίνακας 9.10: Τελικά αποτελέσματα πειραμάτων με την χρήση LSA

c_i	P	R	AP	rP	nR
Βλάστηση	0.626	0.221	0.395	0.687	0.722
Δρόμος	0.400	0.050	0.210	0.250	0.638
Έκρηξη_Φωτιά	0.200	0.111	0.148	0.250	0.620
Ουρανός	0.559	0.271	0.372	0.320	0.706
Χιόνι	0.818	0.264	0.529	0.250	0.825
Γραφείο	0.406	0.147	0.285	0.250	0.673
Έρημος	0.215	0.687	0.246	0.218	0.602
Εξωτερικός χώρος	0.331	0.634	0.382	0.289	0.708
Βουνό	0.110	0.035	0.072	0.224	0.461

Πίνακας 9.11: Τελικά αποτελέσματα πειραμάτων με την χρήση LSA, με σύνολο ελέγχου αποτελούμενο από 20% θετικά και 80% αρνητικά χαρακτηριστικά καρέ ($\lambda=4$)

c_i	χωρίς LSA					με LSA				
	P	R	AP	rP	nR	P	R	AP	rP	nR
Βλάστηση	0.643	0.312	0.460	0.344	0.782	0.626	0.221	0.395	0.687	0.722
Δρόμος	0.295	0.046	0.280	0.228	0.709	0.400	0.050	0.210	0.250	0.638
Έκρ.Φωτιά	0.291	0.777	0.182	0.250	0.592	0.200	0.111	0.148	0.250	0.620
Ουρανός	0.571	0.304	0.436	0.367	0.778	0.559	0.271	0.372	0.320	0.706
Χιόνι	0.777	0.411	0.460	0.250	0.798	0.818	0.264	0.529	0.250	0.825
Γραφείο	0.446	0.157	0.318	0.284	0.734	0.406	0.147	0.285	0.250	0.673
Έρημος	0.333	0.312	0.287	0.250	0.603	0.215	0.687	0.246	0.218	0.602
Εξ.χώρος	0.425	0.514	0.361	0.292	0.733	0.331	0.634	0.382	0.289	0.708
Βουνό	0.444	0.137	0.241	0.250	0.618	0.110	0.035	0.072	0.224	0.461

Πίνακας 9.12: Σύγκριση αποτελεσμάτων χωρίς και με LSA, με σύνολο ελέγχου αποτελούμενο από 20% θετικά και 80% αρνητικά χαρακτηριστικά καρέ ($\lambda=4$)

Παραδείγματα Σωστών και Λανθασμένων Εντοπισμών

Σε αυτή την υποενότητα σχολιάζονται και δίνονται παραδείγματα εικόνων οι οποίες απεικονίζουν μία έννοια και είτε ανιχνεύθηκαν σωστά είτε όχι, επίσης εικόνων οι οποίες δεν απεικονίζουν μία έννοια αλλά ανιχνεύθηκαν λανθασμένα πως την απεικονίζουν. Οι έννοιες για τις οποίες θα παρουσιαστούν παραδείγματα είναι ο ουρανός και η βλάστηση.

Στο σχήμα 9.7 φαίνονται 4 εικόνες που απεικονίζουν την έννοια ουρανός (1-4) και 4 εικόνες την έννοια βλάστηση (5-8). Αυτές είναι παραδείγματα σωστής ανίχνευσης, δηλαδή το σύστημα για την έννοια ουρανός ανίχνευσε πως στις εικόνες (1-4) απεικονίζεται η συγκεκριμένη έννοια. Αντιστοίχως για την έννοια βλάστηση και τις εικόνες (5-8). Στις 4 πρώτες εικόνες παρατηρούνται χαρακτηριστικές περιοχές ουρανού από την σκοπιά του χρώματος και της υψής, οι οποίες και διευκολύνουν στον σωστό εντοπισμό. Το ίδιο συμβαίνει και στις 4 τελευταίες εικόνες, όπου υπάρχουν χαρακτηριστικές περιοχές πράσινης βλάστησης.

Στο σχήμα 9.8 φαίνονται 4 εικόνες που απεικονίζουν την έννοια ουρανός (1-4) και 4 εικόνες την έννοια βλάστηση (5-8). Αυτές είναι παραδείγματα λανθασμένης ανίχνευσης, δηλαδή το σύστημα για την έννοια ουρανός δεν ανίχνευσε πως στις εικόνες (1-4) απεικονίζεται η συγκεκριμένη έννοια. Αντιστοίχως για την έννοια βλάστηση και τις εικόνες (5-8). Για τον ουρανό και τις εικόνες 1 και 3 ισχύει πως τα τεχνητά χαρακτηριστικά τους τα οποία προσδίδουν και διαφορετικά χρώματα στον ουρανό συμβάλει στην λανθασμένη ανίχνευση. Το σύστημα το οποίο έχει εκπαιδευτεί με παραδείγματα ουρανού ο οποίος έχει γαλάζιο ή και γαλάζιο με γκρι (σύννεφα) χρώμα, αδυνατεί να ανιχνεύσει έναν ουρανό με μωβ (1) ή κίτρινο (3) χρώμα. Για τις εικόνες 1 και 2, ισχύει ότι η κατάτμηση των εικόνων ενώνει την περιοχή του ουρανού με άλλες περιοχές, όπως εκείνες των κλαδιών οπότε και επηρεάζονται τα οπτικά χαρακτηριστικά χαμηλού επιπέδου. Τέλος στην εικόνα 4 υπάρχει περιοχή ουρανού η οποία είναι εντελώς λευκή. Για τις εικόνες της βλάστησης και συγκεκριμένα για τις εικόνες 5, 6 και 8 φαίνονται μία πράσινη λωρίδα βλάστησης, ένα δέντρο και μία γλάστρα αντίστοιχα τα οποία εξαιτίας της μικρής και μη συμπαγής περιοχής που καταλαμβάνουν, στην κατάτμηση ενώνονται με άλλες περιοχές κι έτσι αλλοιώνονται τα οπτικά χαρακτηριστικά χαμηλού επιπέδου της περιοχής. Η εικόνα



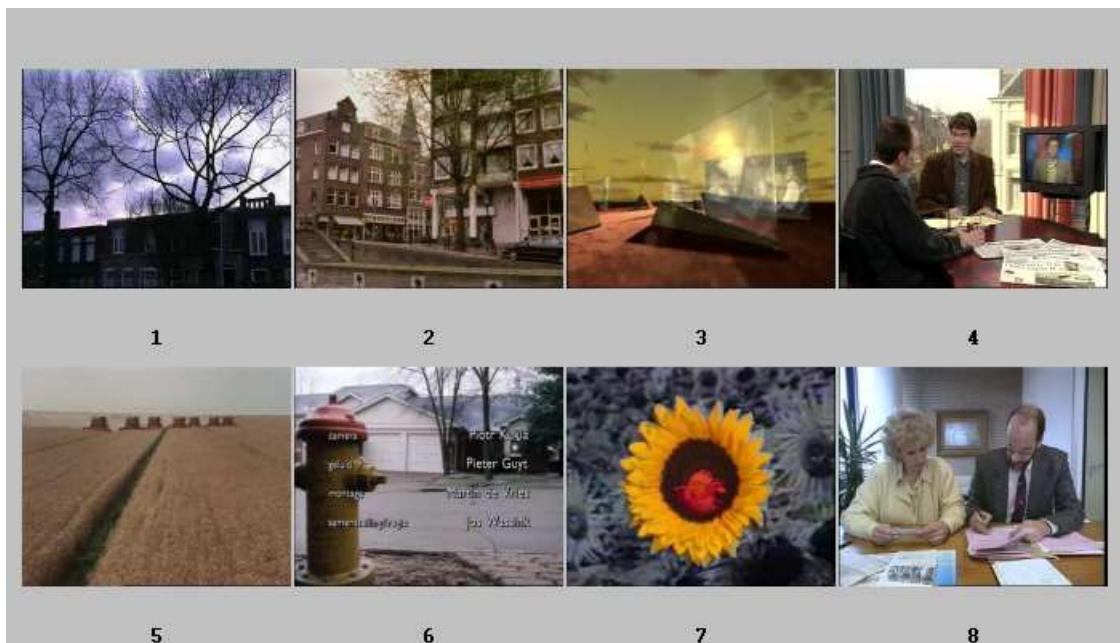
Σχήμα 9.7: Εικόνες που απεικονίζουν την έννοια ουρανό(1-4) και την έννοια βλάστηση (5-8) οι οποίες εντοπίστηκαν επιτυχώς

7 είναι ένα παράδειγμα εικόνας διαφορετικής από εκείνο με το οποίο το σύστημα έχει εκπαιδευτεί, δηλαδή πράσινες περιοχές βλάστησης και όχι κίτρινα λουλούδια τα οποία είναι μία εξαίρεση.

Στο σχήμα 9.9 φαίνονται 4 εικόνες που δεν απεικονίζουν την έννοια ουρανό (1-4) και 4 εικόνες που δεν απεικονίζουν την έννοια βλάστηση (5-8). Αυτές είναι παραδείγματα λανθασμένης ανίχνευσης, δηλαδή το σύστημα για την έννοια ουρανό ανίχνευσε πως στις εικόνες (1-4) απεικονίζεται η συγκεκριμένη έννοια. Αντιστοίχως για την έννοια βλάστηση και τις εικόνες (5-8). Οι εικόνες (1-4) φαίνεται πως περιέχουν γαλάζιες περιοχές οι οποίες πλησιάζουν στα χαρακτηριστικά των περιοχών ουρανού. Για την βλάστηση οι εικόνες 5 και 7 είναι παραδείγματα λάθους στο ground-truth. Αυτό διότι έχουν σημειωθεί ότι δεν απεικονίζουν την έννοια βλάστηση ενώ τελικά την απεικονίζουν. Το σύστημα ανίχνευσε επίσης πως την απεικονίζουν. Ενώ δηλαδή εδώ υπάρχει ένας σωστός εντοπισμός, στις μετρήσεις υπολογίζεται σαν λανθασμένος. Η εικόνα 8 περιέχει μία πράσινη περιοχή η οποία ενδεχομένως “μπερδεύει” το σύστημα. Τέλος η εικόνα 6 αν και δεν φαίνεται τόσο στην τυπωμένη της μορφή, έχει μία πράσινη απόχρωση καθώς και υφή η οποία θυμίζει της βλάστησης, έτσι κι ανιχνεύεται ως απεικόνιση της.

9.2.2 Πειράματα με Σύνολο Εικόνων του COREL

Ο ίδιος αλγόριθμος εφαρμόστηκε και σε ένα σύνολο από εικόνες της βάσης δεδομένων του Corel. Αυτό έγινε για να μπορέσει η μέθοδος να αξιολογηθεί καλύτερα με την χρήση ενός πιο “εύκολου” συνόλου εικόνων, αλλά και με ένα σύνολο από το οποίο θα μπορεί να υπάρχει και σύνολο ελέγχου με γνωστή απόκριση. Επίσης κάτι σημαντικό που προσφέρει αυτό το σύνολο εικόνων είναι ότι λόγω



Σχήμα 9.8: Εικόνες που απεικονίζουν την έννοια ουρανός(1-4) και την έννοια βλάστηση (5-8) οι οποίες δεν εντοπίστηκαν



Σχήμα 9.9: Εικόνες που δεν απεικονίζουν την έννοια ουρανός(1-4) ή την έννοια βλάστηση (5-8) οι οποίες εντοπίστηκαν λανθασμένα πως τις απεικονίζουν

του μικρού του μεγέθους είναι πιο εύκολο και γρήγορο να γίνουν διάφορα πειράματα αλλάζοντας για παράδειγμα το μέγεθος του θησαυρού και να επιλεγθεί έτσι το καταλληλότερο μέγεθος. Στο σχήμα 9.10 φαίνονται μερικές εικόνες από το σύνολο που χρησιμοποιήθηκε. Στην προκειμένη περίπτωση δεν υπάρχουν βίντεο από τα οποία γίνεται εξαγωγή χαρακτηριστικών καρτέ αλλά εικόνες απευθείας. Έτσι κάθε εικόνα εδώ είναι που συμβολίζεται με k_i .

Συνολικός Αριθμός Χαρ. Καρέ	N_K	750
Συνολικός Αριθμός Περιοχών	N_R	18150
Συνολικός Αριθμός Διαν. Χαρακτ/ων	N_F	18150
Μέγεθος Οπτικού Θησαυρού	N_T	50/40/30/20

Πίνακας 9.13: Στοιχεία πειράματος COREL

Οι έννοιες υψηλού επιπέδου που επιλέχθηκαν να εντοπιστούν είναι *ουρανός*, *βλάστηση* και *χιόνι*, έννοιες δηλαδή συμβατές και με εκείνες του TRECVID. Έτσι στο σύνολο των εικόνων υπάρχουν γενικά εικόνες που απεικονίζουν τουλάχιστον μία από αυτές τις έννοιες. Προστέθηκαν επίσης στο σύνολο κάποιες εικόνες οι οποίες δεν απεικονίζουν τίποτα από τα παραπάνω τρία. Συγκεκριμένα το σύνολο αποτελείται από 600 εικόνες οι οποίες απεικονίζουν τουλάχιστον μία από τις 3 παραπάνω έννοιες (497 ουρανός, 126 χιόνι και 342 βλάστηση) και 150 εικόνες των οποίων η απεικόνιση είναι διαφορετική με τις 3 έννοιες. Οι 750 εικόνες αυτές διαχωρίστηκαν σε σύνολο εκπαίδευσης και ελέγχου, καταχωρώντας 525 εικόνες στο πρώτο και 225 στο δεύτερο. Η σύνθεση των δύο αυτών συνόλων ανάλογα με το πόσες εικόνες περιέχουν που απεικονίζουν την έννοια και πόσες όχι φαίνεται στον πίνακα 9.15.

Έννοια	θετικά χαρ. καρτέ
c_i	πλήθος στοιχείων G_i
Ουρανός	497
Χιόνι	126
Βλάστηση	342

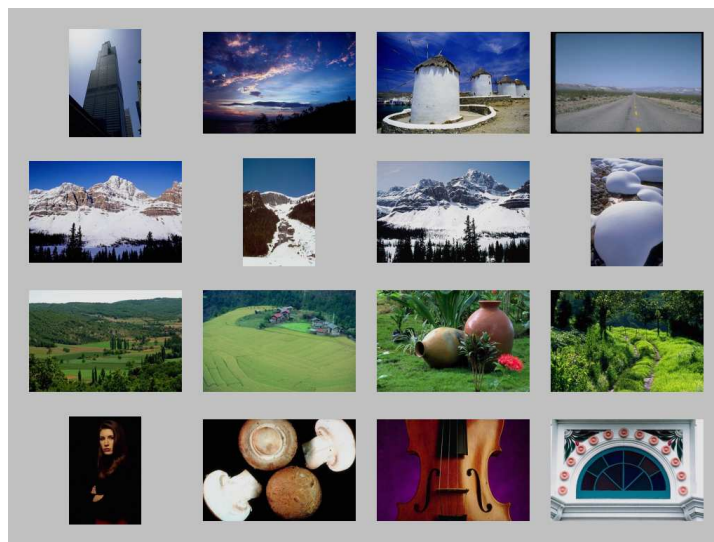
Πίνακας 9.14: Αριθμός χαρακτηριστικών καρτέ που απεικονίζουν την κάθε έννοια

c_i	TR_i		TE_i	
	G_i	\bar{G}_i	G_i	\bar{G}_i
Ουρανός	344	181	153	72
Χιόνι	88	147	38	187
Βλάστηση	236	289	106	119

Πίνακας 9.15: Σύνολα εκπαίδευσης και ελέγχου

Σε αυτά τα πειράματα το διάνυσμα χαρακτηριστικών (9.16) αποτελείται από ένα περισσότερο περιγραφέα ο οποίος δεν χρησιμοποιήθηκε στο TRECVID, τον περιγραφέα ομοιογενούς υψής.

$$f_i = f(r_i) = [CLD(r_i), DCD(r_i), CSTD(r_i), SCD(r_i), EHD(r_i), HTD(r_i)], \quad r_i \in R \quad (9.16)$$



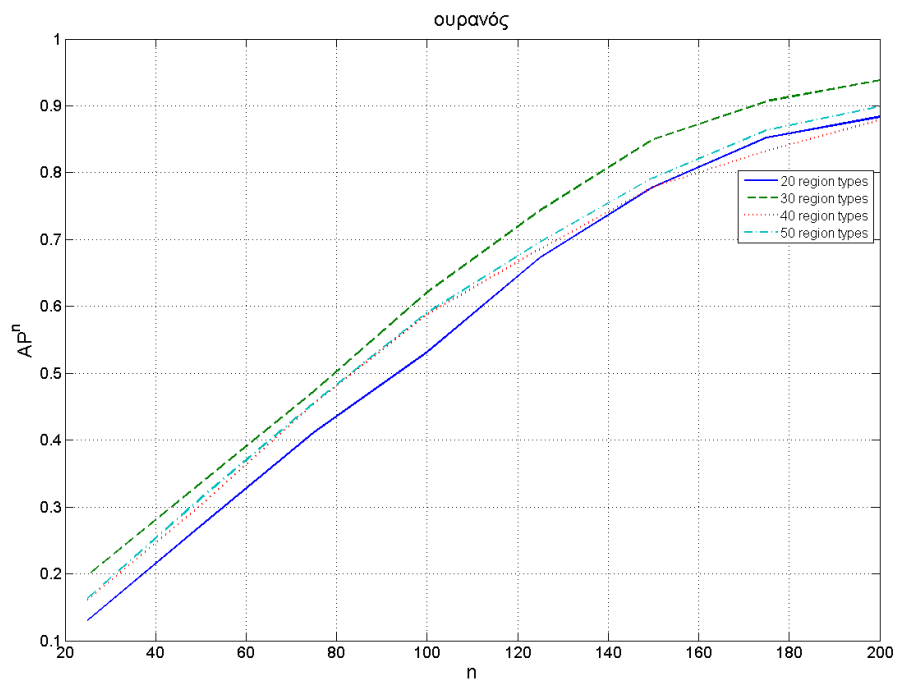
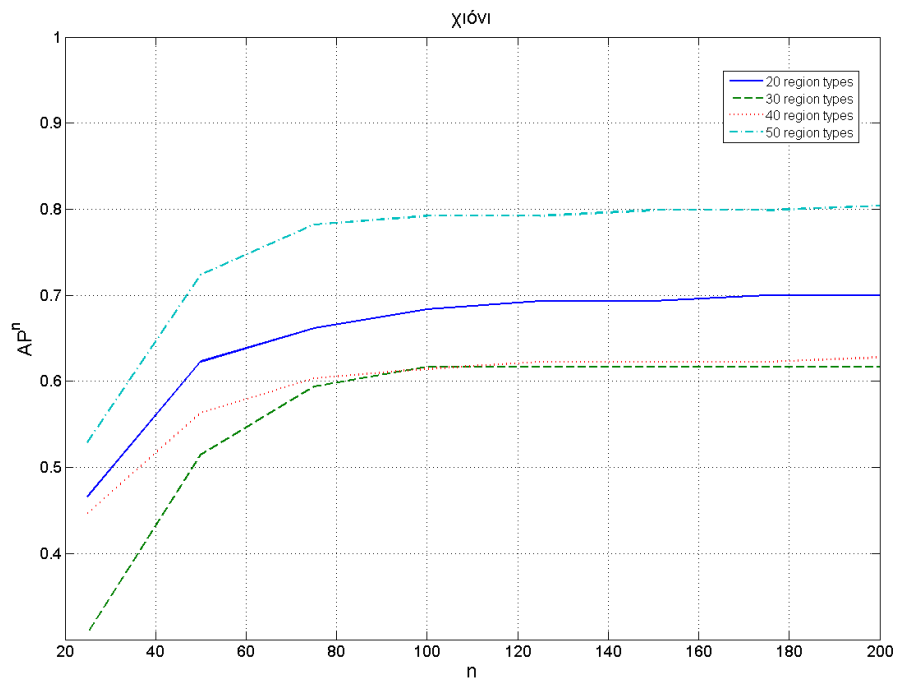
Σχήμα 9.10: Μερικές από τις εικόνες που χρησιμοποιήθηκαν από την συλλογή του Corel, εικόνες που απεικονίζουν ουρανό στην πρώτη σειρά, χιόνι στην δεύτερη, βλάστηση στην τρίτη και καμία από τις 3 έννοιες στην τέταρτη

Έγιναν πειράματα με θησαυρό 50, 40, 30 και 20 τύπων περιοχής. Αυτό γίνεται για να βρεθεί ο αριθμός εκείνος των τύπων περιοχών ο οποίος μπορεί καλύτερα να περιγράψει τις εικόνες του κάθε συνόλου και για τις κάθε φορά έννοιες. Για την κατασκευή του θησαυρού χρησιμοποιήθηκε όλο το σύνολο των περιοχών καθώς ήταν ικανοποιητικής διάστασης για να γίνει αυτό. Έτσι τα νευρωνικά δίκτυα για κάθε μία από τις 3 έννοιες εκπαιδεύτηκαν με 4 διαφορετικά σύνολα εκπαίδευσης. Το διαφορετικό ανάμεσα σε αυτά ήταν το μέγεθος του θησαυρού. Υπολογίστηκε τελικά το Average Precision καθώς το n αυξάνεται πάνω στα 4 αντίστοιχα σύνολα ελέγχου. Τα διαγράμματα στο σχήμα 9.12 δείχνουν της καμπύλες AP^n ως προς το n και από αυτές επιλέγεται και το καταλληλότερο μέγεθος θησαυρού. Τελικά τα μεγέθη θησαυρού τα οποία επιλέγονται είναι 50, 30 και 40 για τις έννοιες χιόνι, ουρανός και βλάστηση αντίστοιχα. Είναι προφανές ότι για κάθε έννοια διαφέρει ο αριθμός των χαρακτηριστικών περιοχών οι οποίες είναι κατάλληλες για να περιγράψουν τις εικόνες της έννοιας κατάλληλα.

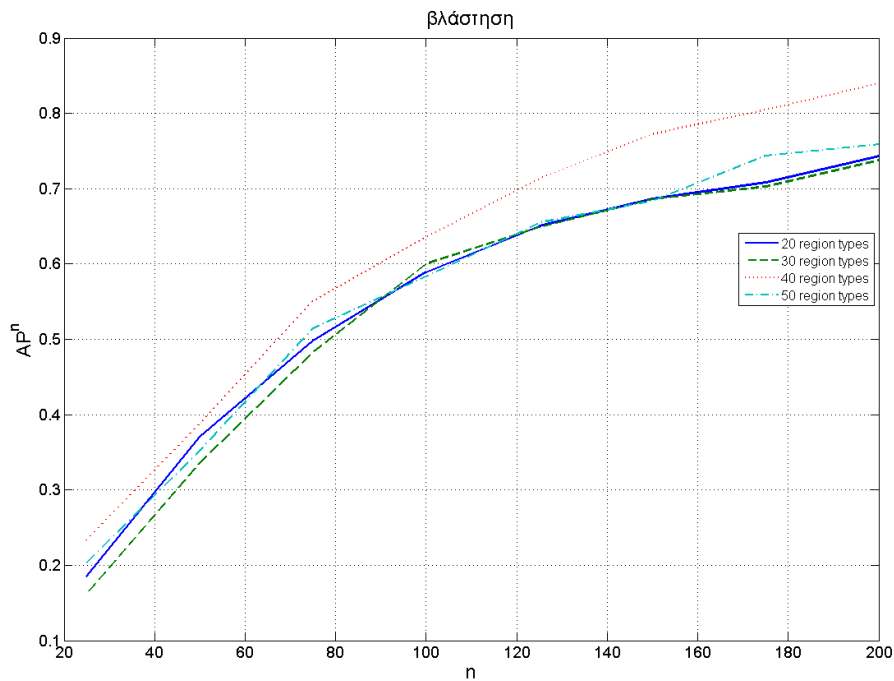
	Ουρανός	Χιόνι	Βλάστηση
χωρίς LSA	0.5	0.4	0.3
με LSA	0.5	0.4	0.3

Πίνακας 9.16: Κατώφλια για όλους τους ανιχνευτές

Τα κατώφλια επιλέχθηκαν για κάθε ανιχνευτή όπως φαίνονται στον πίνακα 9.16. Τα τελικά αποτελέσματα για τα μέτρα που χρησιμοποιήθηκαν για την αξιολόγηση των ανιχνευτών παρουσιάζονται στους πίνακες 9.17 και 9.18 χωρίς και με την χρήση LSA αντίστοιχα.



Σχήμα 9.11: Διαγράμματα Average Precision καθώς αυξάνεται το μέγεθος παραθύρου για τα τέσσερα μεγέθη θησαυρού, για τις έννοιες χιόνι και ουρανός



Σχήμα 9.12: Διάγραμμα Average Precision καθώς αυξάνεται το μέγεθος παραθύρου για τα τέσσερα μεγέθη θησαυρού, για την έννοια βλάστηση

c_i	P	R	AP	rP	nR
Βλάστηση	0.827	0.939	0.642	0.876	0.864
Ουρανός	0.681	0.789	0.791	0.46	0.923
Χιόνι	0.728	0.741	0.632	0.688	0.796

Πίνακας 9.17: Τελικά αποτελέσματα πειραμάτων χωρίς την χρήση LSA

c_i	P	R	AP	rP	nR
Βλάστηση	0.812	0.939	0.650	0.869	0.864
Ουρανός	0.538	0.736	0.668	0.432	0.883
Χιόνι	0.725	0.706	0.625	0.688	0.783

Πίνακας 9.18: Τελικά αποτελέσματα πειραμάτων με την χρήση LSA

	Γραφείο	Έρημος	Βουνό	Έκρηξη-Φωτιά
χωρίς LSA	124	119	128	64
με LSA	141	105	127	142

Πίνακας 9.19: Αποτελέσματα συμμετοχών στο TRECVID 2007

9.3 Συμμετοχές στο TRECVID

Όλα τα προηγούμενα πειράματα για το TRECVID έγιναν για να βγουν κάποια συμπεράσματα για τις τεχνικές που χρησιμοποιούνται με σκοπό τελικά να κατασκευαστούν τα συστήματα ανίχνευσης για κάθε έννοια και να χρησιμοποιηθούν για την συμμετοχή στο TRECVID 2007. Κατασκευάστηκαν έτσι σύνολα εκπαίδευσης για κάθε έννοια τα οποία περιείχαν όλα τα θετικά χαρακτηριστικά καρέ του development set για κάθε έννοια. Στόχος ήταν να χρησιμοποιηθεί όσο το δυνατό περισσότερη πληροφορία αλλά και με τον κατάλληλο τρόπο από το development set. Τα θετικά χαρακτηριστικά καρέ συνοδεύονται και από τον κατάλληλο αριθμό αρνητικών χαρακτηριστικών καρέ για κάθε έννοια. Τελικά εκπαιδεύεται ένα νευρωνικό για κάθε έννοια. Το σύνολο ελέγχου του TRECVID είναι και αυτό επίσης χωρισμένο σε πλάνα από τα οποία εξήχθησαν και πάλι χαρακτηριστικά καρέ. Στα εκπαιδευμένα συστήματα δόθηκαν σαν είσοδος τα χαρακτηριστικά αυτά καρέ κι έτσι στο NIST παραδίδονται τα ποσοστά πεποίθησης για κάθε πλάνο (δηλαδή αντίστοιχο χαρακτηριστικό καρέ). Το NIST για να αξιολογήσει τα αποτελέσματα επιλέγει έναν αριθμό εννοιών και ένα μέρος των πλάνων για τα οποία και θα δημιουργήσει τους σχολιασμούς. Υπολογίζει πάνω σε αυτό το μέρος των πλάνων και για κάθε έννοια την μέση ακρίβεια. Αποτελέσματα από τη δουλειά της διπλωματικής δόθηκαν για την συμμετοχή στο TRECVID 2007 χωρίς την χρήση LSA³[29] και με την χρήση LSA⁴[30]. Η κατάταξη των δύο αυτών συμμετοχών ανάμεσα στις 163 συμμετοχές φαίνεται στον πίνακα 9.19. Πρέπει να σημειωθεί ότι ανάμεσα στις συμμετοχές υπήρχαν πολλές οι οποίες προσέγγισαν το πρόβλημα πολυτροπικά, χρησιμοποιώντας δηλαδή όχι μόνο οπτικά χαρακτηριστικά αλλά και ακουστικά χαρακτηριστικά των βίντεο.

³ στο COST292

⁴ στο K-SPACE

Κεφάλαιο 10

Ταξινόμηση Σκηνών

10.1 Περιγραφή Προβλήματος

Στην ενότητα αυτή παρουσιάζεται το πρόβλημα της αυτόματης ταξινόμησης εικόνων σε κατηγορίες. Πιο συγκεκριμένα της ταξινόμησης ανάμεσα σε δύο κατηγορίες οπότε ανάγεται και σε δυαδική απόφαση για την μία ή την άλλη κατηγορία. Το πρόβλημα αυτό στην συγκεκριμένη προσέγγιση αντιμετωπίζεται με ολικές περιγραφές της εικόνας. Έστω ότι έχουμε N εικόνες οι οποίες όλες αντιστοιχούν σε μία από τις δύο κατηγορίες και κάθε εικόνα συμβολίζεται με k_i , ενώ το σύνολο όλων των εικόνων με K .

$$k_i \in K, \quad i = 1 \dots N \quad (10.1)$$

Περιγραφείς MPEG-7 χρώματος και υψής εξάγονται συνολικά από την κάθε εικόνα. Οι περιγραφείς που εξάγονται είναι οι: περιγραφέας κύριων χρωμάτων, κλιμακωτός περιγραφέας χρώματος, περιγραφέας διάταξης χρώματος, περιγραφέας δομής χρώματος, περιγραφέας ομοιογενούς υψής και περιγραφέας ιστογράμματος ακμών. Έτσι σε κάθε εικόνα αντιστοιχεί και ένα διάνυσμα για κάθε περιγραφέα. Συνολικά 6 δηλαδή διανύσματα για κάθε εικόνα k_i , τα: DCD_i , CLD_i , SCD_i , $CSTD_i$, EHD_i και HT_i . Τα διανύσματα για κάθε περιγραφέα είναι όπως εκείνα έχουν περιγραφεί από τις εξισώσεις (2.2), (2.11), (2.13), (2.9), (2.17) και (2.15). Για τον περιγραφέα κύριων χρωμάτων το διάνυσμα περιέχει μόνο το χρώμα με το μεγαλύτερο ποσοστό εμφάνισης. Επίσης δημιουργείται κι ένα συγχωνευμένο διάνυσμα το οποίο περιέχει όλους τους περιγραφείς για κάθε εικόνα, όπως φαίνεται στην σχέση (10.2).

$$f_i = [DCD_i, CSTD_i, CLD_i, SCD_i, EHD_i, HTD_i], \quad i = 1 \dots N \quad (10.2)$$

Μετά ανάλυση πρωτογενών συνιστωσών εφαρμόζεται σε όλα τα διανύσματα ανά περιγραφέα προκειμένου να αποσυσχετισθούν και προκύπτουν διανύσματα μικρότερης διάστασης. Με το διάνυσμα από κάθε περιγραφέα εκπαιδεύεται ένα νευρωνικό δίκτυο MLP για να ταξινομή τις εικόνες ανάμεσα στις δύο κατηγορίες. Κάτι ακόμα που μπορεί να γίνει για να υπάρξουν βελτιωμένα αποτελέσματα είναι να συγχωνευθούν μερικά από τα διανύσματα των περιγραφέων και να εκπαιδευτεί ένα νέο MLP με αυτά.

Οι κατηγορίες εδώ είναι δύο. Κατηγορία 1 το εξωτερικό μέρος αυτοκινήτου που συμβολίζεται με c_1 και κατηγορία 2 το εσωτερικό c_2 . Συμβολίζεται $b(k_i)$ την έξοδο του ταξινομητή για την εικόνα

k_i , η οποία είναι 1 για την κατηγορία 1 και 0 για την κατηγορία 2. Σχηματίζεται έτσι ένα σύνολο D_1 το οποίο περιέχει τις εικόνες οι οποίες ταξινομήθηκαν στην κατηγορία 1 και αντίστοιχα το D_2 για την κατηγορία 2.

$$b(k) \in \{0, 1\}, \quad k \in K \quad (10.3)$$

$$D_1 = \{k \in K : b(k) = 1\} \quad (10.4)$$

$$D_2 = \{k \in K : b(k) = 0\} \quad (10.5)$$

10.2 Πειράματα

Πειράματα έγιναν για την επίλυση του συγκεκριμένου προβλήματος με την τεχνική που περιγράφηκε πάνω σε ταξινόμηση εικόνων που απεικονίζουν είτε εσωτερικό μέρος αυτοκινήτου, είτε εξωτερικό. Χρησιμοποιήθηκε ένα σύνολο 1247 εικόνων, εκ των οποίων οι 1033 απεικονίζουν εξωτερικό μέρος αυτοκινήτου και οι 214 εσωτερικό. Το σύνολο αυτό διαχωρίστηκε με τυχαίο τρόπο σε σύνολο εκπαίδευσης και σύνολο ελέγχου, με 830 και 417 εικόνες το κάθε ένα αντίστοιχα. Μερικές από τις εικόνες που χρησιμοποιήθηκαν φαίνονται στο σχήμα 10.1. Τελικά το σύνολο εκπαίδευσης που προέκυψε περιέχει 690 εικόνες που απεικονίζουν εξωτερικό και 140 που απεικονίζουν εσωτερικό. Ενώ το σύνολο ελέγχου 343 εικόνες που απεικονίζουν εξωτερικό και 74 εσωτερικό. Όλα τα προηγούμενα συνοψίζονται στον πίνακα 10.1

	Σύνολο Εκπαίδευσης	Σύνολο Ελέγχου
Εξωτερικό	690	343
Εσωτερικό	140	74

Πίνακας 10.1: Σύνολα εκπαίδευσης και ελέγχου για την ταξινόμηση σε εξωτερικό/εσωτερικό μέρος αυτοκινήτου

$$G_i : \{k \in K : C(k) = c_i\} \quad (10.6)$$

Ορίζεται σαν G_1 το σύνολο των εικόνων οι οποίες πραγματικά ανήκουν στην κατηγορία 1, ενώ G_2 στην κατηγορία 2. G_i δηλαδή είναι το σύνολο δεδομένης αλήθειας (ground-truth) και δίνεται από την σχέση (10.6), όπου $C(k)$ είναι η κατηγορία στην οποία ανήκει η εικόνα k . Έτσι ορίζονται τα μέτρα precision και recall και για το συγκεκριμένο πρόβλημα, όπως φαίνονται στις σχέσεις 10.7 και 10.8.

$$P_i = \frac{|D_i \cap G_i|}{|D_i|} \quad (10.7)$$

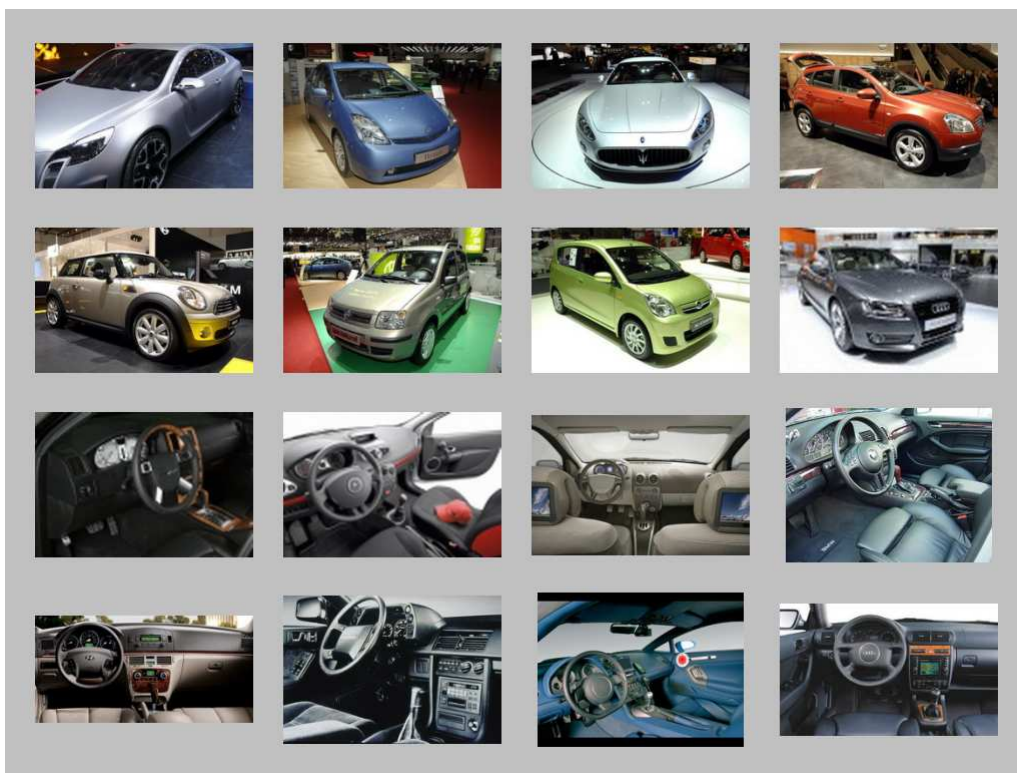
$$R_i = \frac{|D_i \cap G_i|}{|G_i|} \quad (10.8)$$

Ορίζοντας τα μέτρα precision και recall από κοινού για την ανάκτηση και των δύο κατηγοριών προκύπτουν οι σχέσεις (10.9) και (10.10).

Τελικά όμως ισχύει ότι $P_{1,2} = R_{1,2}$, διότι $D_1 \cup D_2 = G_1 \cup G_2 = K$

$$P_{1,2} = \frac{|(D_1 \cap G_1) \cup (D_2 \cap G_2)|}{|D_1 \cup D_2|} \quad (10.9)$$

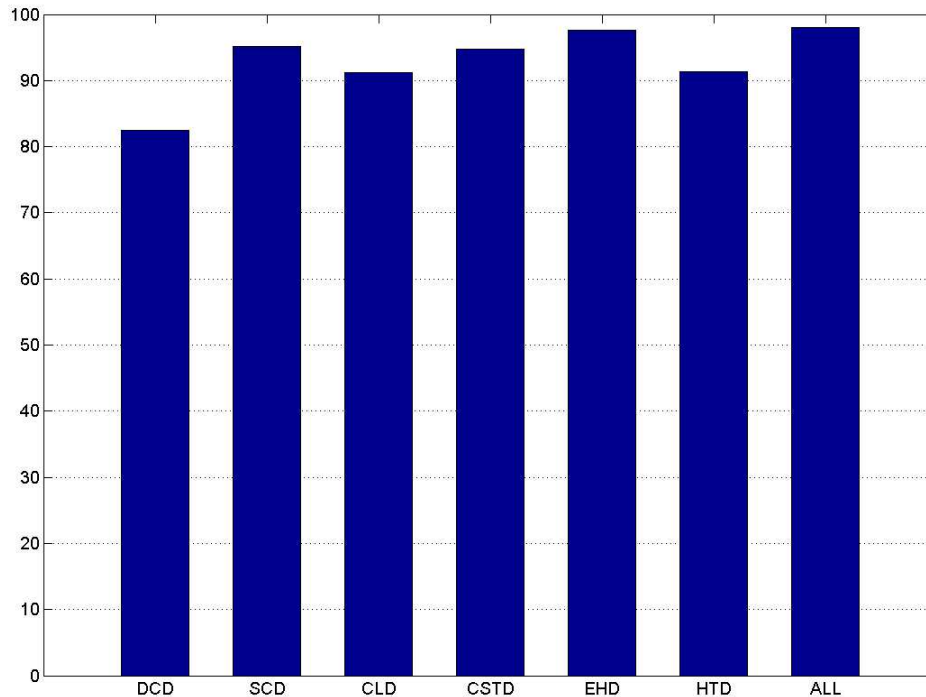
$$R_{1,2} = \frac{|(D_1 \cap G_1) \cup (D_2 \cap G_2)|}{|G_1 \cup G_2|} \quad (10.10)$$



Σχήμα 10.1: Μερικές από τις εικόνες που χρησιμοποιήθηκαν για ταξινόμηση, εικόνες που απεικονίζουν εξωτερικό μέρος στις δύο πρώτες σειρές και εσωτερικό στις δύο τελευταίες

Αρχικά οι 6 περιγραφείς εξήχθησαν από το σύνολο εικόνων και 6 διαφορετικά νευρωνικά δίκτυα MLP εκπαιδεύτηκαν με τα διανύσματα για κάθε περιγραφέα, καθώς και ένα ακόμα με το συγχωνευμένο διάνυσμα. Στην συνέχεια οι περιγραφείς εξήχθησαν και από τις εικόνες του συνόλου ελέγχου και δόθηκαν σαν είσοδο στα αντίστοιχα νευρωνικά δίκτυα. Στην έξοδο εφαρμόστηκε ένα κατώφλι και μετατράπηκε σε δυαδική απόκριση 0 ή 1 όπου το κάθε ένα αντιστοιχεί και στην κάθε μία από τις δύο κατηγορίες. Χρησιμοποιώντας την δυαδική αυτή έξοδο και το ground-truth υπολογίζονται τα μέτρα precision και recall όπως εκείνα περιγράφηκαν στην ενότητα 9.1. Τα αποτελέσματα παρουσιάζονται στον πίνακα 10.2 και το σχήμα 10.2.

Γενικά παρατηρούμε πολύ καλά αποτελέσματα για το εξωτερικό μέρος αυτοκινήτου και για το εσωτερικό με μία μικρή πτώση ενδεχομένως εξαιτίας των λιγότερων δειγμάτων στο σύνολο εκπαίδευσης. Μικρή εξαίρεση στα πολύ καλά αποτελέσματα αποτελεί ο ανιχνευτής βασισμένος στον περιγραφέα κύριων χρωμάτων, του οποίου το διάνυσμα αποτελείται από το χρώμα με το μεγαλύτερο ποσοστό εμφάνισης όπως έχει ήδη αναφερθεί. Φαίνεται λοιπόν λογικό να είναι λιγότερο δυνατό να γίνει μία σωστή ταξινόμηση ως προς αυτό το χαρακτηριστικό και μόνο. Ο λόγος που κρατήθηκε μόνο το ένα κυρίαρχο χρώμα είναι για να υπάρχει ίδια διάσταση των διανυσμάτων, καθώς ο περιγρα-



Σχήμα 10.2: Αποτελέσματα ταξινόμησης εικόνων αυτοκινήτων

Περιγραφέας	Εξωτερικό		Εσωτερικό		$P_{1,2}$
	P_1	R_1	P_2	R_2	
Περιγραφέας Κύριων Χρωμάτων	88%	92.4%	54.3%	41.8%	82.5%
Κλιμακωτός Περιγραφέας Χρώματος	97.3%	96.2%	83.3%	87.8%	95.2%
Περιγραφέας Διάταξης Χρώματος	95.5%	93.5%	72.8%	79.7%	91.2%
Περιγραφέας Δομής Χρώματος	96.2%	95.9%	81.3%	82.4%	94.7%
Περιγραφέας Ιστογράμματος Ακμών	98.5%	97.9%	90.7%	93.2%	97.6%
Περιγραφέας Ομοιογενούς Υφής	96.2%	96.7%	84.7%	82.4%	91.3%
Όλοι οι Περιγραφείς	98.2%	99.3%	95.7%	91.8%	98.0%

Πίνακας 10.2: Αποτελέσματα ταξινόμησης εικόνων αυτοκινήτων

φέας κύριων χρωμάτων για κάθε περιοχή ενδιαφέροντος εξάγει από 1 έως 8 κύρια χρώματα ανάλογα με την κάθε εικόνα. Τέλος παρατηρείται στην εκπαίδευση με το συγχωνευμένο διάνυσμα όλων των περιγραφέων μία βελτίωση της απόδοσης με ακόμα καλύτερα ποσοστά. Το διάνυσμα αυτό αποτελεί μία πιο πλήρη περιγραφή της εικόνας με συνολική πληροφορία για χρώμα αλλά και υφή.

Κεφάλαιο 11

Συμπεράσματα και Μελλοντικές Επεκτάσεις

Σε αυτήν την ενότητα σχολιάζονται τα αποτελέσματα των πειραμάτων που έγιναν για τον αυτόματο εντοπισμό εννοιών σε εικόνες και για ταξινόμηση εικόνων. Επίσης παρουσιάζονται πιθανές επεκτάσεις πάνω στην δουλειά που έγινε.

Για την περιγραφή της οπτικής πληροφορίας χαμηλού επιπέδου στην παρούσα δουλειά χρησιμοποιήθηκαν περιγραφείς MPEG-7. Με τα πειράματα που διεξήχθησαν ελέγχεται η δυνατότητα τους να περιγράψουν αποτελεσματικά την οπτική πληροφορία χαμηλού επιπέδου μίας εικόνας. Τα αποτελέσματα των πειραμάτων για τα δύο προβλήματα τα οποία διερευνήθηκαν, δηλαδή εντοπισμός εννοιών και ταξινόμηση εικόνων, έρχονται να ενισχύσουν ακόμα περισσότερο την θέση ότι οι MPEG-7 περιγραφείς είναι ικανοί για αυτή την περιγραφή. Ιδιαίτερα στα πειράματα της ταξινόμησης σκηνών αναδεικνύεται η δυνατότητα του κάθε περιγραφέα να περιγράφει με διαχωρίσιμο τρόπο την κάθε μία από τις δύο κατηγορίες εικόνων.

Τα πειραματικά αποτελέσματα για τον εντοπισμό εννοιών δείχνουν πως η αναπαράσταση των εικόνων με τα διανύσματα αναπαράστασης τα οποία περιέχουν τις ελάχιστες αποστάσεις ανάμεσα στις περιοχές των εικόνων και στις περιοχές ενός θησαυρού που περιέχει σημασιολογικά χαρακτηριστικά μπορεί να οδηγήσει σε επιτυχή εντοπισμό των εννοιών. Με την τεχνική αυτή λοιπόν ξεπερνιέται το πρόβλημα της έλλειψης σχολιασμών σε επίπεδο περιοχών για την κάθε εικόνα. Συνδυάζεται η πληροφορία χαμηλού επιπέδου από τις διάφορες περιοχές και οδηγεί σε έννοιες υψηλού επιπέδου. Επίσης μία αισθητή βελτίωση των αποτελεσμάτων του εντοπισμού για μερικές από τις έννοιες είναι εφικτή αν εφαρμοστεί η τεχνική LSA. Λαμβάνονται έτσι υπόψη οι λανθάνουσες σχέσεις μεταξύ των περιοχών της κάθε εικόνας και οι ισχυρότερες από αυτές χρησιμοποιούνται για την περιγραφή της. Αποδίδει καλύτερα ο εντοπισμός με αυτή την τεχνική σε έννοιες των οποίων οι εικόνες έχουν καλά συσχετισμένες περιοχές.

Η προσέγγιση με την τεχνική οπτικού θησαυρού έχει επίσης το πλεονέκτημα ότι είναι μία γενική λύση και όχι εξειδικευμένη στον εντοπισμό της κάθε έννοιας με συγκεκριμένο τρόπο. Εφαρμόζεται για διαφορετικές έννοιες με τον ίδιο ακριβώς τρόπο. Το κοινό το οποίο έχουν οι έννοιες για τις οποίες εφαρμόζεται είναι ότι οι περιοχές οι οποίες προσδίδουν σε κάθε εικόνα και την ιδιότητα της απεικόνισης της κάθε έννοιας δεν έχουν συγκεκριμένο σχήμα. Αυτό είναι και το στοιχείο που

επιτρέπει την γενική λύση, ενώ σε περιοχές που θα είχαν συγκεκριμένο σχήμα χρειάζεται μία πιο εξειδικευμένη τεχνική, για παράδειγμα για τον εντοπισμό προσώπου.

Ένα ακόμα συμπέρασμα από τα πειράματα στο σύνολο εικόνων του Corel είναι πως για κάθε έννοια είναι σημαντικός ο αριθμός των τύπων περιοχής του θησαυρού, οι οποίοι χρησιμοποιούνται για να περιγράψουν τις εικόνες της. Κάθε έννοια αποτελείται από εικόνες με περιοχές που έχουν κάποια παρόμοια οπτικά χαρακτηριστικά μεταξύ τους. Έτσι για κάθε έννοια είναι και διαφορετικός ο αριθμός των “οπτικών λέξεων” που αρκούν για να την περιγράψουν αποτελεσματικά αλλά και με συμπαγή τρόπο, χωρίς δηλαδή τύπους περιοχής οι οποίοι δεν παρουσιάζονται και σε μεγάλο ποσοστό ανάμεσα στις εικόνες άρα αποτελούν και μία μορφή θορύβου. Όταν το μέγεθος του θησαυρού αυξάνεται, τότε αυξάνεται και ο χρόνος υπολογισμού των διανυσμάτων αναπαράστασης για κάθε εικόνα. Μεγαλύτερο λοιπόν μέγεθος θησαυρού από το αναγκαίο δεν προσφέρει κάτι άλλο παρά μόνο υπολογιστικό κόστος.

Μελλοντική επέκταση της τεχνικής που χρησιμοποιήθηκε και διερευνήθηκε για τον εντοπισμό των εννοιών μπορεί να είναι μία μελέτη και των χωρικών σχέσεων των εικόνων μέσα στην κάθε εικόνα. Για παράδειγμα για τον εντοπισμό του ουρανού μπορεί να είναι πολύ χρήσιμο να είναι γνωστό ότι κάτω από κάποια περιοχή η οποία φαίνεται να είναι ουρανός, βρίσκεται μία πράσινη περιοχή. Αυτό προσδίδει μία επιπλέον βεβαιότητα ότι η περιοχή είναι ουρανός, κάτω από τον οποίο υπάρχει μία περιοχή βλάστησης και όχι για παράδειγμα ένας γαλάζιος τοίχος σε εσωτερικό χώρο.

Παρατηρώντας τις έννοιες οι οποίες επιλέχθηκαν να εντοπιστούν αλλά και έννοιες οι οποίες γενικά προκαλούν το ενδιαφέρον του αυτόματου εντοπισμού φαίνεται πως κάποιες έννοιες είναι υποσύνολα κάποιων άλλων. Για παράδειγμα ο ουρανός και η έρημος είναι υποσύνολα της έννοιας εξωτερικός χώρος, καθώς κάθε εικόνα από τις δύο πρώτες έννοιες είναι και εικόνα της τρίτης. Η ιδιότητα αυτή μπορεί να χρησιμοποιηθεί για την βελτίωση της χαμηλής απόδοσης κάποιων ανιχνευτών, με την βοήθεια ανιχνευτών καλύτερης απόδοσης για έννοιες που είναι υπερσύνολα των προηγούμενων. Έτσι στην έξοδο του ανιχνευτή για έρημο ο οποίος έχει χαμηλή απόδοση, μπορεί να εφαρμοστεί ο ανιχνευτής για εξωτερικό χώρο ή και το αντίστροφο. Βοηθά αυτό στον να αποκλειστούν επιπλέον λανθασμένες εικόνες.

Για τον εντοπισμό των εννοιών θα μπορούσαν επίσης να χρησιμοποιηθούν συνδυαστικά με τους ανιχνευτές της παρούσας εργασίας και ανιχνευτές οι οποίοι θα έχουν εκπαιδευτεί με συγκεκριμένες περιοχές για τις οποίες είναι γνωστό ότι απεικονίζουν κάποια έννοια. Για τον σκοπό αυτό χρειάζεται ένα σύνολο εικόνων με σχολιασμούς σε επίπεδο περιοχών σαν εκείνο του LabelMe (<http://labelme.csail.mit.edu/>). Έτσι κατασκευάζεται ένας ανιχνευτής ο οποίος μπορεί να αποκριθεί για κάθε περιοχή της υπό εξέταση εικόνας αν είναι περιοχή της συγκεκριμένης έννοιας και αυτή η πληροφορία να συνδυαστεί με την απόκριση των ανιχνευτών με βάση τις περιγραφές με τους τύπους περιοχής του θησαυρού.

Τέλος χρειάζεται να γίνει διερεύνηση για την ικανότητα του κάθε περιγραφέα από μόνος του να σχηματίζει το διάνυσμα χαρακτηριστικών. Όπως δηλαδή έγινε για την ταξινόμηση σκηνών όπου φαίνεται η περιγραφή των εικόνων από κάθε περιγραφέα χωριστά πόσο αποτελεσματικά γίνεται, έτσι να κατασκευάζεται και θησαυρός με διανύσματα χαρακτηριστικών από τον κάθε περιγραφέα χωριστά. Στην συνέχεια οι περιγραφείς να συνδυάζονται κατά διάφορους τρόπους και να σχηματίζουν συγχωνευθέντα διανύσματα χαρακτηριστικών, αναδεικνύοντας και πάλι την αποτελεσματικότερη αποτύπωση των χαρακτηριστικών χαμηλού επιπέδου.

Βιβλιογραφία

- [1] B.S. Manjunath, Philippe Salembier, and Thomas Sikora. *Introduction to MPEG-7*. JOHN WILEY & SONS, LTD.
- [2] S. Chang, T. Sikora, and A. Puri. Overview of the MPEG-7 standard.
- [3] J. Hunter. An Overview of the MPEG-7 Description Definition Language (DDL).
- [4] O. Avaro and P. Salembier. MPEG-7 Systems: Overview.
- [5] M. Bober. MPEG-7 Visual Shape Descriptors. *IEEE Transactions on Circuits and Systems for Video Technology*, 11(6), 2001.
- [6] B.S. Manjunath, J. Ohm, V.V. Vasudevan, and A. Yamada. Color and Texture Descriptors. *IEEE Transactions on Circuits and Systems for Video Technology*, 11(6):703–715, 2001.
- [7] S. Jeannin and A. Divakaran. MPEG-7 Visual Motion Descriptors.
- [8] S. Quackenbush. Overview of MPEG-7 audio.
- [9] R.C Gonzalez and R.E. Woods. *Digital Image Processing*. Prentice Hall, 2002.
- [10] Y.M. Ro and K.W. Yoo. Texture featurng and indexing using matching pursuit in radon space. 2:580–584, 1999.
- [11] F. Mokhtarian and M. Bober. *The Curvature Scale Space Representation: Theory, Applications, and MPEG-7 Standarization*. Kluwer Academic Publishers.
- [12] A. Yamada, M.Pickering, S. Jeannin, L.Cieplinski, J. Ohm, and M. Kim. N3914 - MPEG-7 Visual part of eXperimentation Model version 9.0. 2001.
- [13] E. Spyrou and Y.Avrithis. High-Level Concept Detection in Video Using a Region Thesaurus.
- [14] E. Spyrou and Y.Avrithis. A Region Thesaurus Approach for High-Level Concept Detection in the Natural Disaster Domain.
- [15] E. Spyrou, G. Koumoulos, Y. Avrithis, and S. Kollias. Using Local Region Semantics for Concept Detection in Video.

- [16] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-Based Image Retrieval at the End of the Early Years. pages 1349–1380.
- [17] A.F. Smeaton, P. Over, and W. Kraaij. Evaluation campaigns and TRECVID. In *MIR '06: Proceedings of the 8th ACM International Workshop on Multimedia Information Retrieval*, pages 321–330, New York, NY, USA, 2006. ACM Press. ISBN 1-59593-495-2.
- [18] IBM. MARVEL Multimedia Analysis and Retrieval System. IBM Research White paper.
- [19] F. Souvannavong, B. Merialdo, and B. Huet. Region-based video content indexing and retrieval. In *CBMI 2005, Fourth International Workshop on Content-Based Multimedia Indexing, June 21-23, 2005, Riga, Latvia, Jun 2005*.
- [20] S. Aksoy, A. Avci, E. Balcuk, O. Cavus, P. Duygulu, Z. Karaman, P. Kavak, C. Kaynak, E. Kucukayvaz, C. Ocalan, and P. Yildiz. Bilkent university at TRECVID 2005. 2005.
- [21] N. Voisine, S. Dasiopoulou, V. Mezaris, E. Spyrou, Th. Athanasiadis, I. Kompatsiaris, Y. Avrithis, and M. G. Strintzis. Knowledge-Assisted Video Analysis Using A Genetic Algorithm. In *6th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS 2005)*, April 13-15, 2005.
- [22] N. Boujemaa, F. Fleuret, V. Gouet, and H. Sahbi. Visual content extraction for automatic semantic annotation of video news. In *IS&T/SPIE Conference on Storage and Retrieval Methods and Applications for Multimedia, part of Electronic Imaging symposium*, January 2004.
- [23] C.G.M. Snoek, M. Worring, D.C. Koelma, and A.W.M. Smeulders. Learned Lexicon-driven Interactive Video Retrieval. 2006.
- [24] S. Lazebnik, C. Schmid, and J. Ponce. A discriminative framework for texture and object recognition using local image features. In *Towards category-level object recognition*. Springer, 2006. to appear.
- [25] O. Chapelle, P. Haffner, and V.N. Vapnik. Support vector machines for histogram-based image classification. *IEEE Transactions on Neural Networks*, 10(5):1055–1064, 1999.
- [26] C. G. M. Snoek and M. Worring. Time Interval based Modelling and Classification of Events in Soccer Video. In *Proceedings of the 9th Annual Conference of the advanced School for Computing and Imaging (ASCI)*, 2003.
- [27] J. Laaksonen, M. Koskela, and E. Oja. Picsom, Self-Organizing Image Retrieval with MPEG-7 Content Descriptors. 2002.
- [28] S. Haykin. *Neural Networks A Comprehensive Foundation*. Prentice Hall, 1999.
- [29] Q. Zhang, K. Chandramouli, U. Damnjanovic, T. Piatrik, E. Izquierdo M. Corvaglia, N. Adami, R. Leonardi, G. Yakin, S. Aksoy, U. Naci, A. Hanjalic, S. Vrochidis, A. Moutzidou, S. Nikolopoulos, V. Mezaris L. Makris, I. Kompatsiaris, E. Esen, A. Alatan, E. Spyrou, P. Kapsalas, G. Talias, P. Mylonas, Y. Avrithis, B. Reljin, G. Zajic, R. Jarina, M. Kuba,

- N. Aginamo, J. Goya, B. Mansencal, J. Benois-Pineau, A. M. G. Pinheiro, L. A. Alexandre, and P. Almeida. The COST292 experimental framework for TRECVID 2007. 2007.
- [30] P. Wilkins, T. Adamek, D. Byrne, G. J.F.Jones, H. Lee, G. Keenan, K. McGuinness, N. E. O Connor, A. F. Smeaton, A. Amin, Z. Obrenovic, R. Benmokhtar, E. Galmar, B. Huet, S. Essid, R. Landais, F. Vallet, G. Th. Papadopoulos, S. Vrochidis, V. Mezaris, I. Kompatsiaris, E. Spyrou, Y. Avrithis, R. Mörzinger, P. Schallauer, W. Bailer, T. Piatrik, K. Chandramouli, E. Izquierdo, Martin Haller, Lutz Goldmann, Amjad Samour, Andreas Cobet, Thomas Sikora, and P. Praks. K-Space at TRECVID 2007. 2007.

Παράρτημα Α

Αρχεία Παραμέτρων και Αρχεία Εξόδου

Σε αυτή την ενότητα δίνονται παραδείγματα αρχείων παραμέτρων για τις εφαρμογές εξυπηρετητή των 8 περιγραφών που είναι και το περιεχόμενο του VDE καθώς και περιεχόμενα των XML αρχείων εξόδου τους.

A.1 Παραδείγματα Αρχείων Παραμέτρων

```
Application DominantColorServer

/* default database file name*/
ListFile    Databases/mylist.lst

/* default bitstream file*/
/*Bitstream Bitstreams/DominantColor.xml*/

/* dominant color setting */
ColorSpacePresent      0
ColorQuantizationPresent  0
VariancePresent        0
SpatialCoherency       1
```

Πίνακας A.1: Περιεχόμενο αρχείου παραμέτρων για τον περιγραφέα κύριων χρωμάτων

```

/* default application name*/ Application ScalableColorServer

/* default database file name*/ ListFile    Databases/mylist.lst

/* default bitstream file*/ /*Bitstream
Bitstreams/ScalebleColor_64_3.mp7*/

/* NumberOfBitplanesDiscarded: 0, 1, 2, 3, 4, 6, 8 */
NumberOfBitplanesDiscarded 3

/* NumberOfCoefficients: 16, 32, 64, 128, 256      */
NumberOfCoefficients 64

```

Πίνακας A.2: Περιεχόμενο αρχείου παραμέτρων για τον κλιμακωτό περιγραφέα χρώματος

```

/* default application name*/ Application
ColorLayoutServer

/* default database file name*/ ListFile    Databases/images.lst

/* default bitstream file*/
/*Bitstream Bitstreams/ColorLayout.xml*/

/*Number of Coefficients 3, 6, 10, 15, 21, 28, 64*/
NumberOfYCoeff 6
NumberOfCCoeff 3

```

Πίνακας A.3: Περιεχόμενο αρχείου παραμέτρων για τον περιγραφέα διάταξης χρώματος

```

/* default application name*/ Application ColorStructureServer

/* default database file name*/ ListFile    Databases/myList.lst

/* default bitstream file*/
/*Bitstream Bitstreams/ColorStructure.xml*/

/* color quantization resolution*/ ColorQuantSize  256

```

Πίνακας A.4: Περιεχόμενο αρχείου παραμέτρων για τον περιγραφέα δομής χρώματος

```

/* default application name*/ Application EdgeHistogramServer

/* default database file name*/ ListFile
Databases/myList.lst

/* default bitstream file*/
/*Bitstream Bitstreams/EdgeHistogram.xml*/

```

Πίνακας A.5: Περιεχόμενο αρχείου παραμέτρων για τον περιγραφέα ιστογράμματος ακμών

```

/* default application name*/ Application HomogeneousTextureServer

/* default database file name*/ ListFile    Databases/myList.lst

/* default bitstream file*/ /*Bitstream Bitstreams/HT_Test.mp7*/

/* additional parameters*/
/* default layer */ /* (0:base-layer 32-components  1:full-layer
62components) */ layer  1

```

Πίνακας A.6: Περιεχόμενο αρχείου παραμέτρων για τον περιγραφέα ομοιογενούς υφής

```
/* default application name*/ Application ContourShapeServer  
  
/* default database file name*/ ListFile    Databases/myList.lst  
  
/* default bitstream file*/  
/*Bitstream Bitstreams/ContourShape.xml*/
```

Πίνακας A.7: Περιεχόμενο αρχείου παραμέτρων για τον περιγραφέα σχήματος με βάση το περίγραμμα

```
/* default application name*/ Application    RegionShapeServer  
  
/* default database file name*/ ListFile  
databases/mylist.lst  
  
/* default bitstream file*/  
/*Bitstream bitstreams/RegionShape.xml*/
```

Πίνακας A.8: Περιεχόμενο αρχείου παραμέτρων για τον περιγραφέα σχήματος με βάση το σχήμα

A.2 Παραδείγματα Αρχείων XML Εξόδου

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
- <Mpeg7 xmlns="http://www.mpeg7.org/2001/MPEG-7_Schema"
xmlns:xsi="http://www.w3.org/2000/10/XMLSchema-instance">
- <DescriptionUnit xsi:type="DescriptorCollectionType">
- <Descriptor xsi:type="DominantColorType">
- <Image Number="0">
  <SpatialCoherency>25</SpatialCoherency>
- <Value>
  <Percentage>19</Percentage>
  <Index>1 2 3</Index>
  <ColorVariance>0 0 0</ColorVariance>
</Value>
- <Value>
  <Percentage>12</Percentage>
  <Index>6 8 12</Index>
  <ColorVariance>0 0 0</ColorVariance>
</Value>
</Image>
</Descriptor>
</DescriptionUnit>
</Mpeg7>
```

Πίνακας A.9: Περιεχόμενο αρχείου εξόδου XML για τον περιγραφέα κύριων χρωμάτων για μία περιοχή

```

<?xml version="1.0" encoding="ISO-8859-1" ?>
- <Mpeg7 xmlns="http://www.mpeg7.org/2001/MPEG-7_Schema"
xmlns:xsi="http://www.w3.org/2000/10/XMLSchema-instance">
- <DescriptionUnit xsi:type="DescriptorCollectionType">
- <Descriptor xsi:type="ColorLayoutType">
- <Image Number="0">
  <YDCCoeff>6</YDCCoeff>
  <CbDCCoeff>38</CbDCCoeff>
  <CrDCCoeff>27</CrDCCoeff>
  <YACCCoeff5>16 16 16 16 16</YACCCoeff5>
  <CbACCCoeff2>16 16</CbACCCoeff2>
  <CrACCCoeff2>16 16</CrACCCoeff2>
  </Image>
</Descriptor>
</DescriptionUnit>
</Mpeg7>

```

Πίνακας Α.10: Περιεχόμενο αρχείου εξόδου XML για τον περιγραφέα διάταξης χρώματος για μία περιοχή

```

<?xml version="1.0" encoding="ISO-8859-1" ?>
- <Mpeg7 xmlns="http://www.mpeg7.org/2001/MPEG-7_Schema"
xmlns:xsi="http://www.w3.org/2000/10/XMLSchema-instance">
- <DescriptionUnit xsi:type="DescriptorCollectionType">
- <Descriptor xsi:type="ScalableColorType">
- <Image Number="0">
- <Region Number="0" NumberOfCoefficients="0"
NumberOfBitplanesDiscarded="3">
  <Coefficients>-8 3 8 0 -1 1 1 1 -3 2 3 0 0 0 -3 2</Coefficients>
  </Region>
- <Region Number="1" NumberOfCoefficients="0"
NumberOfBitplanesDiscarded="3">
  <Coefficients>-3 1 7 -2 0 1 0 1 -3 1 1 0 0 1 -3 2</Coefficients>
  </Region>
- <Region Number="2" NumberOfCoefficients="0"
NumberOfBitplanesDiscarded="3">
  <Coefficients>-1 -3 0 0 -2 1 0 2 -1 1 0 1 -4 1 -1 2</Coefficients>
  </Region>
- <Region Number="3" NumberOfCoefficients="0"
NumberOfBitplanesDiscarded="3">
  <Coefficients>0 -4 -1 0 -2 0 0 1 -2 1 0 1 -3 1 -1 2</Coefficients>
  </Region>
</Image>
</Descriptor>
</DescriptionUnit>
</Mpeg7>

```

Πίνακας Α.11: Περιεχόμενο αρχείου εξόδου XML για τον κλιμακωτό περιγραφέα χρώματος για χάρτη με 4 περιοχές


```

    <?xml version="1.0" encoding="ISO-8859-1" ?>
- <Mpeg7 xmlns="http://www.mpeg7.org/2001/MPEG-7_Schema"
xmlns:xsi="http://www.w3.org/2000/10/XMLSchema-instance">
- <DescriptionUnit xsi:type="DescriptorCollectionType">
- <Descriptor xsi:type="ColorStructureType">
- <Image Number="0">
- <Region Number="0" colorQuant="1">
  <Values>0 0 0 0 139 0 2 0 136 137 93 47 9 27 30 1 27
  51 102 184 8 13 2 1 12 19 28 13 5 3 2 62</Values>
  </Region>
- <Region Number="1" colorQuant="1">
  <Values>0 0 0 0 0 0 0 0 241 235 0 0 0 0 0 0 63 109 86
  189 229 203 24 8 13 158 131 41 15 0 0 53</Values>
  </Region>
- <Region Number="2" colorQuant="1">
  <Values>0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 230 0
  0 0 0 0 0 0 0 0 0 0 170</Values>
  </Region>
- <Region Number="3" colorQuant="1">
  <Values>0 0 0 0 118 0 41 0 207 194 77 43 5 9 15 3 140
  123 114 147 92 78 22 6 88 71 59 26 11 6 6 72</Values>
  </Region>
</Image>
</Descriptor>
</DescriptionUnit>
</Mpeg7>

```

Πίνακας A.12: Περιεχόμενο αρχείου εξόδου XML για τον περιγραφέα δομής χρώματος για αρχείο κειμένου με συντεταγμένες για 4 ορθογώνιες περιοχές

```

<?xml version="1.0" encoding="ISO-8859-1" ?>
- <Mpeg7 xmlns="http://www.mpeg7.org/2001/MPEG-7_Schema"
xmlns:xsi="http://www.w3.org/2000/10/XMLSchema-instance">
- <DescriptionUnit xsi:type="DescriptorCollectionType">
- <Descriptor xsi:type="EdgeHistogramType">
- <Image Number="0">
<BinCounts>0 0 0 0 1 2 2 6 1 4 1 1 0 3 3 0 0 0 0 0 3 0 2 0 2 5
0 5 1 3 4 0 1 2 3 2 0 1 4 3 3 0 2 1 2 3 1 1 3 1 1 3 2 1 2 2 3
2 2 2 2 0 0 1 1 1 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0</BinCounts>
</Image>
</Descriptor>
</DescriptionUnit>
</Mpeg7>

```

Πίνακας Α.13: Περιεχόμενο αρχείου εξόδου XML για τον περιγραφέα ιστογράμματος ακμών για μία περιοχή

```

<?xml version="1.0" encoding="ISO-8859-1" ?>
- <Mpeg7 xmlns="http://www.mpeg7.org/2001/MPEG-7_Schema"
xmlns:xsi="http://www.w3.org/2000/10/XMLSchema-instance">
- <DescriptionUnit xsi:type="DescriptorCollectionType">
- <Descriptor xsi:type="HomogeneousTextureType">
- <Image Number="0">
<Average>51</Average>
<StandardDeviation>58</StandardDeviation>
<Energy>130 146 166 193 161 133 133 149 173 194 155
130 113 136 142 168 129 119 97 107 134 139 125 99
80 64 104 109 105 70</Energy>
<EnergyDeviation>132 143 161 192 160 131 129 147 172
198 144 128 106 129 127 156 118 117 89 93 139 127 125
98 74 52 91 111 83 60</EnergyDeviation>
</Image>
</Descriptor>
</DescriptionUnit>
</Mpeg7>

```

Πίνακας Α.14: Περιεχόμενο αρχείου εξόδου XML για τον περιγραφέα ομοιογενούς υφής για μία περιοχή

```

<?xml version="1.0" encoding="ISO-8859-1" ?>
- <Mpeg7 xmlns="http://www.mpeg7.org/2001/MPEG-7_Schema"
xmlns:xsi="http://www.w3.org/2000/10/XMLSchema-instance">
- <DescriptionUnit xsi:type="DescriptorCollectionType">
- <Descriptor xsi:type="RegionShapeType">
- <Image Number="0">
- <Region Number="0">
  <MagnitudeOfART>14 15 8 10 6 15 10 15 13 0 12 4 5 4 11 7
  4 7 8 1 4 3 8 5 3 4 5 3 0 5 3 0 4 4 3</MagnitudeOfART>
</Region>
- <Region Number="1">
  <MagnitudeOfART>4 15 3 3 3 15 15 15 6 2 6 14 13 7 10 4 10
  6 4 3 3 5 4 11 10 3 5 2 7 10 9 3 2 1 3</MagnitudeOfART>
</Region>
- <Region Number="2">
  <MagnitudeOfART>7 15 6 7 2 15 15 12 11 2 10 9 5 5 12 7 9
  9 11 8 11 8 0 7 10 9 6 5 5 5 2 6 5 2 3</MagnitudeOfART>
</Region>
- <Region Number="3">
  <MagnitudeOfART>15 15 10 12 10 14 3 14 13 5 12 12 6 9 3 8
  3 11 2 11 9 3 7 2 0 2 10 5 6 5 4 2 6 5 1</MagnitudeOfART>
</Region>
</Image>
</Descriptor>
</DescriptionUnit>
</Mpeg7>

```

Πίνακας Α.15: Περιεχόμενο αρχείου εξόδου XML για τον περιγραφέα σχήματος με βάση το σχήμα για χάρτη με 4 περιοχές

```

<?xml version="1.0" encoding="ISO-8859-1" ?>
- <Mpeg7 xmlns="http://www.mpeg7.org/2001/MPEG-7_Schema"
xmlns:xsi="http://www.w3.org/2000/10/XMLSchema-instance">
- <DescriptionUnit xsi:type="DescriptorCollectionType">
- <Descriptor xsi:type="ContourShapeType">
- <Image Number="0">
  <GlobalCurvature>11 2</GlobalCurvature>
  <PrototypeCurvature>0 2</PrototypeCurvature>
  <HighestPeakY>25</HighestPeakY>
  <Peak peakX="49" peakY="3" />
  <Peak peakX="39" peakY="6" />
  <Peak peakX="30" peakY="5" />
  <Peak peakX="22" peakY="7" />
  <Peak peakX="9" peakY="7" />
  <Peak peakX="49" peakY="6" />
  <Peak peakX="46" peakY="5" />
  <Peak peakX="16" peakY="5" />
  <Peak peakX="61" peakY="7" />
</Image>
</Descriptor>
</DescriptionUnit>
</Mpeg7>

```

Πίνακας Α.16: Περιεχόμενο αρχείου εξόδου XML για τον περιγραφέα σχήματος με βάση το περίγραμμα για μία περιοχή