Πανεπιστήμιο Κρήτης
Σχολή Θετικών Επιστημών
Τμήμα Επιστήμης Υπολογιστών

Εξαγωγή χαρακτηριστικών υφής και χρώματος για τμηματοποίηση και ανάκληση εικόνων

Σπύρος Α. Λιάπης

Μεταπτυχιακή Εργασία

Ηράκλειο, Σεπτέμριος 1999

© Copyright 1999 by Spyros A. Liapis
All Rights Reserved

Πανεπιστήμιο Κρήτης
Σχολή Θετικών Επιστημών
Τμήμα Επιστήμης Υπολογιστών

Εξαγωγή χαρακτηριστικών υφής και χρώματος για τμηματοποίηση και ανάκληση εικόνων

Εργασία που υποβλήθηκε από τον
Σπύρο Α. Λιάπη
ως μερική εκπλήρωση των απαιτήσεων για την απόκτηση
ΜΕΤΑΠΤΥΧΙΑΚΟΥ ΔΙΠΛΩΜΑΤΟΣ ΕΙΔΙΚΕΥΣΗΣ

Συγγραφέας:

Σπύρος Α. Λιάπης
Τμήμα Επιστήμης Υπολογιστών
Πανεπιστήμιο Κρήτης

Εισηγητική Επιτροπή:

Γιώργος Τζιρίτας, Αναπληρωτής Καθηγητής,Επόπτης

Απόστολος Τραγανίτης, Αναπληρωτής Καθηγητής, Μέλος

Πάνος Τραχανιάς, Αναπληρωτής Καθηγητής, Μέλος

Δεκτή:

Πάνος Κωνσταντόπουλος
Πρόεδρος Επιτροπής Μεταπτυχιακών Σπουδών

Σεπτέμριος 1999

Αφιερώνεται στην οικογένεια μου
και στους παππούδες μου που έφυγαν

Εξαγωγή χαρακτηριστικών υφής και χρώματος για τμηματοποίηση και ανάκληση εικόνων

Σπύρος Α. Λιάπης

Μεταπτυχιακή Εργασία

Τμήμα Επιστήμης Υπολογιστών
Πανεπιστήμιο Κρήτης

Περίληψη

Καθώς ο ρυθμός δημιουργίας εικόνων αυξάνεται συνεχώς με την πάροδο του χρόνου είναι απαραίτητη η δημιουργία τρόπων διαχείρισης (ευρετηριαμός, ανάκληση) της οπτικής πληροφορίας με βάση το περιεχόμενο της, *Content Based Image Retrieval* CBIR. Το περιεχόμενο των εικόνων δεν μπορεί να περιγραφεί με λέξεις καθώς η περιγραφή είναι υποκειμενική για κάθε άνθρωπο. Άλλωστε αυτό επαληθεύεται από το ρητό: "*μια εικόνα αξίζει όσο χίλιες λέξεις* ". Για την περιγραφή του περιεχομένου πρέπει να γίνει χρήση αριθμητικών χαρακτηριστικών τα οποία είναι συγκρίσιμα. Στην παρούσα εργασία μελετάται η χρήση της υφής και του χρώματος για την περιγραφή της εικόνας.

Παρουσιάζονται τρεις προσεγγίσεις για τον χαρακτηρισμό της υφής: Descrite Wavelet Frames ανάλυση, MR-SAR και τα φίλτρα Gabor. Για την σύγκριση των μεθόδων γίνονται εκτεταμένα πειράματα σε μεγάλο αριθμό από εικόνες που περιέχουν υφή. Τα πειράματα βασίζονται σε κατηγοριοποιήσεις των εικόνων με βάση τα χαρακτηριστικά τους. Επίσης παρουσιάζεται ένας αλγόριθμος τμηματοποίησης εικόνων χρησιμοποιώντας χαρακτηριστικά υφής με βάση την DWF ανάλυση. Ο αλγόριθμος είναι μερικά αυτόνομος καθώς χρειάζεται τον αριθμό των διαφορετικών περιεχομένων της εικόνας. Γίνεται παράθεση πειραματικών αποτελεσμάτων σε συνθετικές εικόνες που περιέχουν υφή αλλά και σε εικόνες με φυσικές σκηνές.

Επίσης όσον αφορά τα χαρακτηριστικά του χρώματος παρουσιάζεται ένας αλγόριθμος εξαγωγής των κυρίων χρωμάτων εικόνων. Η εξαγωγή αυτών βασίζεται στα συστήματα χρώματων HSV, και Lab, τα οποία και συγκρίνονται. Παρουσιάζονται οι δυνατότητες που παρέχουν τα κύρια χρώματα για την περιγραφή εικόνων, καθώς και μέτρα σύγκρισης των χαρακτηριστικών των χρωμάτων. Στην συνέχεια γίνονται πειράματα ανάκτησης συνδυάζοντας χαρακτηριστικά υφής και χρώματος στις ίδιες εικόνες

που χρησιμοποιήθηκαν στην κατηγοριοποίηση υφής. Τέλος παρουσιάζονται τρόποι εξαγωγής τμημάτων από την εικόνα με βάση το χρώμα και την υφή για την περιγραφή του περιεχομένου.

Επόπτης: Γιώργος Τζιρίτας
Αναπληρωτής Καθηγητής
Επιστήμης Υπολογιστών
Πανεπιστήμιο Κρήτης

Texture and colour feature extraction for image segmentation and retrieval

Spyros A. Liapis

Master of Science Thesis

Computer Science Department
University of Crete

Abstract

Now days the generation rate of image is increasing rapidly. So it is necessary to develop ways of manipulating (indexing, retrieval) the visual information by its content. This is called *Content Based Image Retrieval* (CBIR). The content of an image can not be described with words, because the description made by a man is subjective. Never the less this is confirmed by the old saying: "*an image is worthy as one thousand words*". In order to describe the image content, must be used arithmetic features that will be comparative. In this work texture and color are used for image description.

Three ways of texture characterization are presented: Discrete Wavelet Frames, MR-SAR and Gabor filters. For comparing these texture features, experiments are performed on a large number of textured images. These experiments are based on image classifications, according to the above texture features. Also an image segmentation algorithm based on DWF texture features is proposed. This algorithm is partially supervised because the number of different texture contents in the image is needed. Experimental results of the segmentation algorithm are performed on synthetic textured images and on images of physical scenes.

As for color features an algorithm of image dominant color extraction is proposed. The extraction is based on HSV and Lab color systems, which are compared. The capabilities of dominant color extraction in image description, and distance metrics for comparing them are presented. Also experiments of image retrieval based on texture and color features are performed, on the same images that where used in texture classification. Finally numerous ways of segment extraction from the image according to texture and color are proposed for content description.

Supervisor: George Tziritas
Associate Professor of Computer Science
University of Crete

Ευχαριστίες

Πρώτα από όλους θα ήθελα να ευχαριστήσω τους γονείς μου Αλέκο και Κατερίνα όπως και τις γιαγιάδες Αθηνίτσα και Σταματούλα, που χωρίς αυτούς δεν θα βρισκόμουν στο σημείο αυτό. Με την ανατροφή και τις αρχές που μου έδωσαν κατάφερα εκτός από κάτοχος μεταπτυχιακού διπλώματος, να γίνω άτομο χρήσιμο για τους συνανθρώπους μου. Ελπίζω ως ένα σημείο να εκπλήρωσα τις προσδοκίες τους...

Επίσης ευχαριστώ τον επόπτη της εργασίας αυτής Γιώργο Τζιρίτα, που χωρίς την συμβολή του δεν θα είχε ολοκληρωθεί η εργασία. Τα τελευταία 3 χρόνια είχαμε μια άψογη συνεργασία από την οποία κέρδισα πάρα πολλά πράγματα. Εκτός από γνώσεις στο αντίκειμενο της εργασίας, απέκτησα βασικές αρχές που θα μου χρησιμέψουν στην υπόλοιπη μου ζωή. Ο Γιώργος Τζιρίτας είναι ένα υπόδειγμα επιστήμονα, καθηγητή αλλά και οικογενειάρχη. Του εύχομαι να παραμείνει όπως είναι.

Στο σημείο αυτό θα ήθελα να ευχαριστήσω τα μέλη της επιτροπής κ. Απόστολο Τραγανίτη και κ. Πάνο Τραχανιά για τις επισημάνσεις τους και για τον χρόνο τους.

Δεν θα μπορούσα να παραλείψω, πρώτα από όλους τους φίλους και συμφοιτητές, να ευχαριστήσω τον Ηλία Γκρίνια, ο οποίος βοήθησε κατά τα μέγιστα στην πραγματοποίηση της εργασίας. Οι ατελείωτες συζητήσεις μας για θέματα της εργασίας και μη ήταν πάντα εποικοδομητικές - μάλλον όχι πάντα, έχει το ελάττωμα να ανήκει στους ”αθλητικά” μισητούς. Θα τον παρακαλούσα να προσέξει από εδώ και πέρα τον εαυτό του για να μην έχουμε και άλλα κενά διαστήματα.

Ακόμη ευχαριστώ όλους τους φίλους με τους οποίους πέρασα αξέχαστες μέρες και νύχτες στο Ηράκλειο τα τελευταία 6 χρόνια. Θα μου λείψουν πολύ κατά την αποχώρηση μου από το Ηράκλειο όπου πέρασα τα πιο όμορφα χρόνια της ζωής μου - τα φοιτητικά. Ιδιαίτερα θα μου λείψουν οι νύχτες που πέρασα στο Ηράκλειο. Στον Δημήτρη, στον Γιώργο, στον ξενιτεμένο Θάνο, στον Κώστα, στον Νίκο και στους υπόλοιπους ... εύχομαι να βρουν αυτό που ψάχνουν στην ζωή. Μερικοί νομίζω ότι το έχουν βρει ήδη.

Επίσης θέλω να ευχαριστήσω τις φίλες μου Μαρία και Ελευθερία, οι οποίες μου συμπαραστάθηκαν ψυχολογικά στους δύσκολους για μένα καιρούς στο Ηράκλειο. Νομίζω ότι το κατάφεραν με το παραπάνω. Από την πλευρά μου θα τους ευχηθώ να ”χαλαρώσουν” λίγο. Το άγχος κάνει κακό ... Μαρία.

Τέλος ευχαριστώ την Βαγγελίτσα και τον Παναγιώτη που με βοήθησαν πάρα πολύ κατά την παραμονή μου στο Ηράκλειο.

” Η ζωή είναι ένα παιχνίδι ... παίξτε όσο μπορείτε. ”

Σπύρος Α. Λιάπης

Περιεχόμενα

Περίληψη **i**

Abstract **iii**

Ευχαριστίες **v**

Περιεχόμενα **vii**

Κατάλογος Σχημάτων **xi**

Κατάλογος Πινάκων **xv**

1 Εισαγωγή **1**

1.1 Κίνητρο υλοποίησης της εργασίας 1

1.2 Το πρόβλημα 2

1.3 Ερευνητικές προσεγγίσεις 3

1.4 Η οργάνωση της εργασίας 5

2 Χαρακτηριστικά υφής **7**

2.1 Γενικά 7

2.2 Ανάλυση Discrete Wavelet Frames (DWF) 8

2.2.1 Καθορισμός φίλτρων και ιδιότητες αυτών 8

2.2.2 Wavelet Frames 9

2.2.3 2-Δ επέκταση των DWF για εξαγωγή χαρακτηριστικών υφής 12

2.3 Φίλτρα Gabor, Gabor Wavelets 14

2.4 MR-SAR μοντέλο 16

2.4.1 SAR μοντέλο 18

2.4.2 Επέκταση σε πολλαπλή ανάλυση 18

3 Ταξινόμηση υφής **21**
3.1 DWF χαρακτηρισμός της υφής και κριτήριο ταξινόμησης 21
3.2 Gabor φίλτρα 23
3.3 MR-SAR 24
3.4 Πρότυπο πείραμα 24
3.5 Επιπλέον πειραματικά αποτελέσματα 31

4 Τμηματοποίηση υφής **35**
4.1 Αυτόματος υπολογισμός των διανυσμάτων διασπορών για κάθε κλάση . 37
4.1.1 Αρχικές προσεγγίσεις 37
4.1.2 Απόρριψη μη ομοιογενών block 38
4.1.3 Ομαδοποίηση (Clustering) 38
4.2 Κατάταξη σημείων στις κλάσεις. Ένωση συνεκτικών περιοχών 40
4.3 Πειραματικά αποτελέσματα 42

5 Χρωματικά χαρακτηριστικά **49**
5.1 Γενικά 49
5.2 Χρωματικά συστήματα 50
5.3 Εξαγωγή κυρίων χρωμάτων 54
5.3.1 Ιεραρχική ομαδοποίηση 54
5.3.2 Μοντέλο μίξης κατανομών 55
5.3.3 K-means ομαδοποίηση 56
5.3.3.1 Βασικός αλγόριθμος 56
5.3.3.2 Αρχικοποίηση 57
5.3.3.3 Επεξεργασία εκ των υστέρων 58
5.3.3.4 Πειραματικά αποτελέσματα 58
5.4 Περιγραφή εικόνας με βάση το χρώμα 60
5.4.1 Χρωματικά χαρακτηριστικά 64
5.4.2 Μέτρα σύγκρισης χαρακτηριστικών 66

6 Κατηγοριοποίηση με τη χρήση χρώματος και υφής **69**
6.1 Χρήση κοινής απόστασης χρώματος και υφής 69
6.2 Συνδυασμός εκ των υστέρων 72

7 Εξαγωγή τμημάτων με τη χρήση χρώματος και υφής **75**
7.1 Εξαγωγή τμημάτων ξεχωριστά σε χρώμα και υφή 75

7.1.1 Ανίχνευση υφής 76
7.2 Τμηματοποίηση με χρώμα και υφή 77

8 Επίλογος - Συμπεράσματα **81**
8.1 Μελλοντική εργασία 82

A Αλγόριθμος χαλάρωσης ετικετών **85**

Βιβλιογραφία **88**

Κατάλογος Σχημάτων

1.1 Τυπική αρχιτεκτονική μιας βάσης δεδομένων με εικόνες. 5

2.1 Αποκρίσεις στον χώρο των συχνοτήτων που αποκόπτονται σε 5 επίπεδα της ανάλυσης. 10

2.2 Discrete Wavelet Transform - 1D decomposition, όπου s_i, d_i προσέγγιση και υπόλοιπο στο επίπεδο i αντίστοιχα 11

2.3 Discrete Wavelet Frame 1D decomposition, όπου s_i, d_i προσέγγιση και υπόλοιπο στο επίπεδο i αντίστοιχα 13

2.4 Discrete Wavelet Frame 2-D ανάλυση: h κατωδιαβατό, g ανωδιαβατό . . . 14

2.5 Τρόπος διχωρισμού των συχνοτήτων στην Discrete Wavelet Frame 14

2.6 Gabor Wavelets frequency decomposition: στο σχήμα φαίνονται οι περιοχές συχνοτήτων των λεπτομερειών σε κάθε επίπεδο, και της προσέγγισης δεύτερο επιπέδου (μεσαίος κύκλος). Συμμετρικοί κύκλοι (δηλαδή στην ίδια κατεύθυνση) στο αντίστοιχο επίπεδο ανάλυσης ανήκουν στην ίδια συνιστωσα της ανάλυσης (coefficient) διότι έχουμε φίλτρα άρτιας συμμετρίας 17

2.7 Πολλαπλής ανάλυσης SAR μοντέλο με χρήση Gauss πυραμίδας L επιπέδων. 19

3.1 Η βάση με τις 55 εικόνες που περιέχουν υφή από το VisTex του MIT. 26

3.2 Καμπύλη επιτυχίας, συναρτήσει του αριθμού εικόνων που ανακλήθηκαν από την βάση, για την DWF ανάλυση. Γίνεται σύγκριση όταν χρησιμοποιούμε την Bhattacharya με Laplace και με Gauss κατανομή άλλα και την Maximum Likelyhood. 27

3.3 Καμπύλη επιτυχίας, συναρτήσει του αριθμού εικόνων που ανακλήθηκαν από την βάση, για όλες τις μεθόδους (Gabor, DWF (με Laplace), και MR-SAR). 28

4.1 Όλα τα στάδια του αυτόματου αλγόριθμου τμηματοποίησης. 36

4.2 Ομαδοποίηση: κάθε κύκλος παριστάνει μια τρέχουσα κλάση.Κάθε φορά ενώνονται οι πιο κοντινές. Αρχικά έχουμε 6 κλάσεις που προκύπτουν από χαρακτηριστικά 6 block-κλάσεων. Τελικά υπολογίζονται τα αντιπροσωπευτικά χαρακτηριστικά των δύο ζητούμενων κλάσεων 40
4.3 Η αρχική συνθετική εικόνα με 5 υφές και οι αρχικές κατατάξεις με βάση την απόσταση 42
4.4 Οι κατατάξεις μετά το φίλτρο ενδιάμεσης τιμής και η τελική εικόνα μετά τον ICM, αντίστοιχα για την συνθετική εικόνα με τις πέντε διαφορετικές υφές. 43
4.5 Τα μη ομοιογενή block και τα block που χρησιμοποιήθηκαν για την εξαγωγή των χαρακτηριστικών κάθε κλάσης, αντίστοιχα για την συνθετική εικόνα με τις πέντε διαφορετικές υφές. 43
4.6 Η αρχική εικόνα και η τελική τμηματοποιημένη, αντίστοιχα για την συνθετική εικόνα με δύο διαφορετικές υφές (D9 D3). 44
4.7 Η αρχική εικόνα και η τελική τμηματοποιημένη, αντίστοιχα για την συνθετική εικόνα με 4 διαφορετικές υφές (D19 D9 D3 D5). 44
4.8 Η αρχική εικόνα και η τελική τμηματοποιημένη, αντίστοιχα για την συνθετική εικόνα με τρεις διαφορετικές υφές. 45
4.9 Η αρχική εικόνα και η τελική τμηματοποιημένη, αντίστοιχα για την συνθετική εικόνα με δύο διαφορετικές υφές. 45
4.10 Η αρχική εικόνα και οι τελικές τμηματοποιημένες, με 2 και 3 κλάσεις αντίστοιχα για την εικόνα με τα βότσαλα. 47
4.11 Τα block που χρησιμοποιούνται για την εξαγωγή των παραμέτρων κάθε κλάσης για κάθε περίπτωση (2 και 3 κλάσεις). 47

5.1 Αναπαράσταση του κώνου των LHS χρωματικών συστημάτων. 52
5.2 Η αρχική εικόνα και οι απεικονίσεις της με την εξαγωγή κυρίων χρωμάτων με το HSV και Lab 61
5.3 Η αρχική εικόνα και οι απεικονίσεις της με την εξαγωγή κυρίων χρωμάτων με το HSV και Lab 62
5.4 Η αρχική εικόνα και οι απεικονίσεις της με την εξαγωγή κυρίων χρωμάτων με το HSV και Lab 63
5.5 Παράθεση των εξαχθέντων κυρίων χρωμάτων με το HSV πάνω και Lab κάτω, για την που περιέχει τον δρόμο με τα διερχόμενα αυτοκίνητα . . . 64

5.6 Παράθεση των εξαχθέντων κυρίων χρωμάτων με το HSV πάνω και Lab κάτω, για την εικόνα που περιέχει την κυρία που αγοράζει υφάσματα . . 64
5.7 Παράθεση των εξαχθέντων κυρίων χρωμάτων με το HSV πάνω και Lab κάτω, για την εικόνα που περιέχει τον Bill Clinton 64

6.1 Καμπύλη επιτυχίας, συναρτήσει του αριθμού εικόνων που ανακλήθηκαν από την βάση, κάνοντας χρήση κοινής απόστασης για χρώμα και για υφή. 71
6.2 Καμπύλη επιτυχίας, συναρτήσει του αριθμού εικόνων που ανακλήθηκαν από την βάση, κάνοντας χρήση του χρώματος και της υφής ξεχωριστά. . 72
6.3 Καμπύλη επιτυχίας, συναρτήσει του αριθμού εικόνων που ανακλήθηκαν από την βάση, κάνοντας χρήση διαφορετικής απόστασης για χρώμα και για υφή και συνδυάζοντας τες εκ των υστέρων με ποσοστόσεις. 74

7.1 Διάκριση τρόπων περιγραφής εικόνας με τμηματοποίηση με βάση την υφή και το χρώμα 76
7.2 Η αρχική εικόνα και οι τελική τμηματοποιημένη, με 3 κλάσεις για την εικόνα με τα βότσαλα κάνοντας και χρήση χρώματος 78

A.1 Γειτονιά δεύτερης τάξης και οι πιθανές επιλογές παρεών 85

.

Κατάλογος Πινάκων

2.1 Μάσκα γειτονιάς δεύτερης τάξεως για το pixel στην θέση (0,0) 18
2.2 Σύνολα γειτονιών $\mathcal{D}_l$ πολλαπλών επιπέδων για το MR-SAR μοντέλο. 20

3.1 Αναλυτικά αποτελέσματα για τα ποσοστό επιτυχίας στις επιμέρους εικόνας της βάσης για κάθε μέθοδο. 32
3.2 Αποτέσματα της DWF ανάλυσης με Bhattacharya απόσταση για το πείραμα με τις Brodatz εικόνες και 50 τυχαία εικονίδια ανά εικόνα. 33
3.3 Αποτέσματα της DWF ανάλυσης με ML απόσταση για το πείραμα με τις Brodatz εικόνες και 50 τυχαία εικονίδια ανά εικόνα. 34

.

Κεφάλαιο 1

Εισαγωγή

1.1 Κίνητρο υλοποίησης της εργασίας

Ο ρυθμός δημιουργίας των εικόνων από διάφορες πηγές αυξάνεται συνεχώς με την πάροδο του χρόνου. Πηγές παραγωγής εικόνων είναι δορυφόροι, στρατιωτικές αναγνωριστικές πτήσεις, η αστυνομία για την καταχώρηση κακοποιών με βάση το πρόσωπο είτε τα δακτυλικά αποτυπώματα, κάμερες παρακολούθησης χώρων, ιατρικά όργανα, σύστημα οικιακής διασκέδασης, μέσα επικοινωνιών όπως τηλεοπτικοί σταθμοί. Αυτομάτως συμπεραίνουμε ότι άνθρωποι διάφορων επαγγελμάτων από επιστήμονες έως ένα απλό χρήστη ενός προσωπικού υπολογιστή, έρχονται σε επαφή με την οπτική πληροφορία, που παράγεται από τις παραπάνω πηγές, για διαφορετικούς σκοπούς. Οπότε αναγκαία θεωρείται η δημιουργία τρόπων διαχείρισης (ευρετηριασμός, ανάκληση) της οπτικής πληροφορίας με βάση το περιεχόμενο της, Content Based Image Retrieval (CBIR) [1]. Ένα σύστημα CBIR οφείλει να αρχειοθετήσει αποτελεσματικά και με ευφιές τρόπο την τεράστια σε όγκο οπτική πληροφορία που βρίσκεται στα αποθηκευτικά μέσα, ώστε να παρέχονται στον χρήστη διαφανείς λειτουργίες δεικτοδότησης και ψαξίματος των εικόνων. Με την εξέλιξη της τεχνολογίας το κόστος της μνήμης και των αποθηκευτικών μέσων μειώνονται, καθώς το μέγεθος αυτών αυξάνει με γρήγορους ρυθμούς. Γεγονός που κάνει εφικτή την πραγματοποίηση εφαρμογών που εμπλέκονται με διαχείριση εικόνων, αφού οι εφαρμογές αυτές απαιτούν μεγάλη υπολογιστική ισχύ, μνήμη και αποθηκευτικό χώρο. Απαιτήσεις που ήταν απαγορευτικές για τις εφαρμογές αυτές τα προηγούμενα χρόνια. Τα τελευταία χρόνια αρχίζει και γίνεται εφικτή η υλοποίηση εφαρμογών όπως:

- διαχείριση πληροφορίας αρχείων φιλμ τηλεοπτικών σταθμών (video/multimedia database).
- διαχείριση βάσεων λογοτύπων (trademarks) για την κατοχύρωση πνευματικών, εμπορικών δικαιωμάτων (copyright).
- γεωγραφικά συστήματα πληροφορίας, ψηφιακές βιβλιοθήκες (digital libraries) [2].
- επιστημονικές βάσεις δεδομένων, σε τομείς όπως η βιολογία ιατρική.
- σχεδίαση υφασμάτων και μόδας.
- πρόβλεψη καιρού.
- επιβολή νόμου και διερεύνηση εγκλημάτων (εγκληματολογικές βάσεις).
- σχεδίαση και διακόσμηση εσωτερικών χώρων.
- διαχείριση βάσεων εκθεμάτων των γκαλλερύ και των μουσείων.

1.2 Το πρόβλημα

Τα τελευταία χρόνια γίνεται έρευνα για την ανεύρεση τρόπων για την περιγραφή και την διαχείριση της σημασιολογικής πληροφορίας, που περιέχεται σε μια εικόνα, και την εισαγωγή της σε μια βάση δεδομένων. Οι κοινές βάσεις δεδομένων που διαχειρίζονται κείμενο και αριθμούς έχουν μελετηθεί εξαντλητικά τις τελευταίες δεκαετίες. Οι αριθμοί και το κείμενο έχουν ένα μεγάλο πλεονέκτημα, το ότι μπορούν να καταμετρηθούν να ταξινομηθούν και να αποθηκευτούν με τον τρόπο αυτό σε αποθηκευτικά μέσα. Σε αντίθεση, για τις εικόνες και το video δεν υπάρχει κάποια γλώσσα περιγραφής τους, οπότε σε μια σχεσιακή βάση μπορεί να αποθηκευτεί μια εικόνα αλλά δεν μπορεί να γίνει επερώτηση σε αυτήν με βάση το περιεχόμενο της. Αλγόριθμοι για την δεικτειοδότηση εικόνων είναι είναι ακόμα σε μη ικανοποιητικό στάδιο για να ικανοποιήσουν το CBIR. Το πρόβλημα είναι ότι το περιεχόμενο των εικόνων δεν μπορεί να περιγραφεί με κείμενο. Επερωτώντας σε πληροφορία κειμένου, το τι σημαίνει αυτό που διατυπώνεται, συγκριτικά με τις εικόνες, είναι αυστηρά συγκεκριμένο. Σε αντίθεση επερωτώντας πληροφορία εικόνας, το περιεχόμενο της μπορεί να ληφθεί διαφορετικά με πολλούς τρόπους από διαφορετικούς χρήστες. Το κείμενο έχει λίγους βαθμούς ελευθερίας όσον αφορά το νόημα του. Αν χαρακτηρισθεί/επερωτηθεί μια εικόνα με βάση το κείμενο, θα διαπιστωθεί ευθύς αμέσως ότι υπάρχουν πάρα πολλοί χαρακτηρισμοί επερωτήσεις που

μπορούν να γίνουν με βάση το περιεχόμενο της. Το γεγονός αυτό συμβαδίζει με το γνωστό ρητό: *"Μια εικόνα αξίζει όσο χίλιες λέξεις"*. Αλλά και πάλι αν σκεφτεί κανείς ότι μια εικόνα μπορεί να έχει αναρίθμητα αντικείμενα με διάφορες ιδιότητες και τοπολογικές σχέσεις μεταξύ τους, που και αυτές είναι υποκειμενικές, τότε συμπεραίνουμε ότι ούτε οι χίλιες λέξεις είναι αρκετές.

1.3 Ερευνητικές προσεγγίσεις

Για την πραγματοποίηση του CBIR υπάρχουν δύο βασικές τάσεις όσον αφορά την περιγραφή των εικόνων : με "γνωρίσματα" και με εξαγωγή "χαρακτηριστικών"/αναγνώριση αντικειμένων (*attribute based representation* vs *feature extraction/object recognition subsystem*).

Στην πρώτη προσέγγιση το περιεχόμενο της εικόνας μοντελοποιείται ως ένα σύνολο από γνωρίσματα τα οποία εξάγονται την επίβλεψη χρήστη. Οπότε μετέπειτα είναι δυνατή η διαχείριση τους από μια συμβατική βάση δεδομένων (π.χ. σχεσιακή) [3]. Τα συστήματα αυτά εμπεριέχουν ένα μεγάλο βαθμό γενίκευσης στην περιγραφή της εικόνας. Αυτό έχει το μειονέκτημα ότι όσο μεγαλύτερη είναι η αφαιρετικότητα στην περιγραφή τόσο λιγότερο ευέλικτο θα είναι το σύστημα στην δημιουργία επερωτήσεων. Η μέθοδος αυτή υιοθετήθηκε από ερευνητές στον τομέα των βάσεων δεδομένων.

Η δεύτερη προσέγγιση εξαρτάται από ένα υποσύστημα εξαγωγής χαρακτηριστικών και αναγνώρισης αντικειμένων. Τα χαρακτηριστικά που εξάγονται περιγράφουν και αντιπροσωπεύουν την εικόνα. Για το ψάξιμο ανάμεσα στο χαρακτηριστικά αυτά και για το καλύτερο ταίριασμα με τις προδιαγραφές του χρήστη, επιστρατεύονται μέθοδοι αναγνώρισης προτύπων. Αυτή η προσέγγιση αυτοματοποιεί τον χαρακτηρισμό των εικόνων κατά την είσοδο της εικόνας στην βάση δεδομένων. Όμως τεχνικές εξαγωγής χαρακτηριστικών πλήρως αυτόματες είτε είναι πολύ πολύπλοκες και έχουν μεγάλο υπολογιστικό κόστος, είτε τείνουν να λειτουργούν για ένα πολύ συγκεκριμένο πεδίο δεδομένων εφαρμογών και όχι στην γενική περίπτωση. Η μέθοδος αυτή αναπτύχθηκε κυρίως από επιστήμονες στο τομέα της υπολογιστικής όρασης και της επεξεργασίας εικόνων.

Όπως τέθηκε το πρόβλημα στην προηγούμενη ενότητα οι επερωτήσεις που γίνονται σε φυσική γλώσσα πρέπει να μεταφραστούν σε γλώσσα κατανοητή για το σύστημα ανάκλησης. Γλώσσα που να χρησιμοποιεί χαμηλού επιπέδου χαρακτηριστικά (**low-level/primitive features**) της εικόνας και να συνθέτει λογικά χαρακτηριστικά ση-

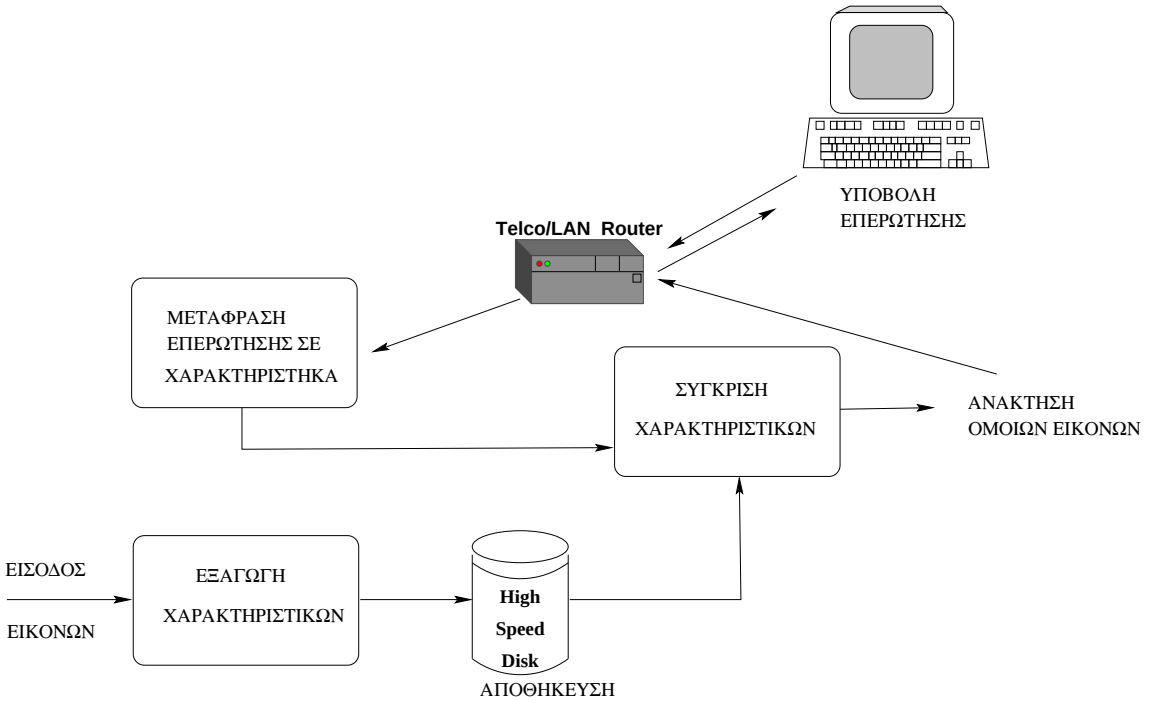
μασιολογικού χαρακτήρα με βάση το περιεχόμενο από τα πρωτογενή χαρακτηριστικά. Αμέσως διαφαίνεται η αναγκαιότητα της εξαγωγής των low-level features από την εικόνα και η μελέτη περιγραφής αυτών αλλά και συγκρίσεως αυτών. Χαρακτηριστικά που χρησιμοποιούνται για την περιγραφή των εικόνων και από τις δυο προσεγγίσεις είναι: χρώμα, υφή, σχήμα, κείμενο, κίνηση (αν πρόκειται για video), χωρικοί περιορισμοί (spatial contraints).

Στην παρούσα εργασία μελετάται η εξαγωγή χαρακτηριστικών που βασίζονται στην υφή (**texture**) και στο χρώμα (**colour**), ακολουθώντας την δεύτερη προσέγγιση (feature extraction). Το χρώμα και η υφή είναι δύο χαρακτηριστικά από τα πιο σημαντικά στην περιγραφή του περιεχομένου μιας εικόνας. Καθώς βασιζόμενοι σε αυτά μπορούμε να εντοπίσουμε αντικείμενα η να απομονώσουμε και να χαρακτηρίσουμε σημασιολογικά ξεχωριστές περιοχές στην εικόνα. Δυστυχώς αλγόριθμοι μηχανικής όρασης και επεξεργασίας εικόνων προς αυτήν την κατεύθυνση είναι πολύπλοκοι δεν λειτουργούν πλήρως αυτόματα, ή όταν λειτουργούν τείνουν να δουλεύουν για να συγκεκριμένο πεδίο δεδομένων, όπως προαναφέρθηκε. Στην εργασία αυτή διερευνώνται αλγόριθμοι τμηματοποίησης που να χρησιμοποιούν χρώμα και υφή ξεχωριστά ή και συνδυασμό αυτών όσο το δυνατόν αυτόματα. Επίσης μελετώνται οι κατάλληλοι περιγραφείς (descriptors) για χρώμα και για υφή και συγκρίνονται μεταξύ τους με πειράματα ταξινομήσεων επιστρατεύοντας και γνώσεις από αναγνώριση προτύπων (pattern recognition).

Στο σχήμα 1.1 φαίνεται μια τυπική αρχιτεκτονική μιας βάσεως δεδομένων από εικόνες. Όπου ένας χρήστης από οποιοδήποτε τερματικό συνδεδεμένο σε δίκτυο μπορεί να υποβάλει επερώτηση μέσω παραδείγματος είτε μέσω λεκτικών γνωρισμάτων. Η επέρωτηση αυτή μεταφράζεται από την βάση σε χαρακτηριστικά που χρησιμοποιούνται για ευρετηριασμό, π.χ. χαρακτηριστικά υφών και χρώματος. Επίσης οι εικόνες που εισάγονται στην βάση επεξεργάζονται από αλγορίθμους εξαγωγής των προηγούμενων χαρακτηριστικών. Τα χαρακτηριστικά αποθηκεύονται, με την μορφή μεταδεδομένων (metadata) σε ένα αποθηκευτικό μέσο (σκληρός δίσκος). Τα χαρακτηριστικά της επερώτησης συγκρίνονται με τα χαρακτηριστικά στην βάση. Με αυτόν τον τρόπο ανακτώνται οι ομοιότερες εικόνες και μεταδίδονται στον χρήστη. Θέμα που θίγεται στο σύστημα αυτό, είναι η σημασιολογική γλώσσα μετάφρασης της επερώτησης (όταν αυτή δεν γίνεται με εικόνα υπόδειγμα), όπου εμπλέκονται όσοι ασχολούνται με σημασιολογικό ευρετηριασμό και τεχνητή νοημοσύνη. Επίσης θίγεται η μετάδοση και η αποθήκευση όπου ασχολούνται επιστήμονες από τον χώρο των δικτύων και θεωρίας πληροφορίας και κωδικοποίησης. Στα κομμάτια της εξαγωγής χαρακτηριστικών και σύγκρισης γίνεται

χρήση αλγορίθμων ψηφιακής επεξεργασίας εικόνων και ανγώρισης προτύπων. Σε αυτά τα δύο κομμάτια επικεντρώνεται η παρούσα εργασία.

Γνωστά συστήματα βάσεων δεδομένων με εικόνες είναι QBIC [4], Virage [5], VisualSeek, PhotoBook [6], Chabot [3].



Σχήμα 1.1: Τυπική αρχιτεκτονική μιας βάσης δεδομένων με εικόνες.

1.4 Η οργάνωση της εργασίας

Στο κεφάλαιο 2 εξετάζονται χαρακτηριστικά υφών (texture features). Προτείνεται η χρήση Discrete Wavelet Frames (DWF) ανάλυσης που βασίζεται σε πολυκαναλικό φιλτράρισμα των εικόνων (multichannel filtering) που διαχωρίζει τις υφές με βάση την απόκριση τους σε διάφορες περιοχές συχνοτήτων. Επίσης παρουσιάζονται τα Gabor φίλτρα, τα οποία είναι της ίδιας οικογένειας με την DWF, και Multiresolution Simultaneous Autoregressive model (MR-SAR) μέθοδος που βασίζεται σε μοντέλα αυτο-παλινδρόμισης (autoregressive models). Στο κεφάλαιο 3 παρουσιάζεται η επίδοση των παραπάνω χαρακτηριστικών υφής σε πειράματα ταξινόμησης σε διάφορες βάσεις με υφές, αφού πρώτα παρουσιαστούν τα κατάλληλα κριτήρια ομοιότητας ή απόστασης ανά περίπτωση. Στο κεφάλαιο 4 παρουσιάζεται ο αλγόριθμός τμηματοποίησης που αναπτύχθηκε με βάση τα DWF features για τμηματοποίηση με βάση την υφή. Στο κεφάλαιο 5 παρουσιάζονται

τρόποι περιγραφής της εικόνας με βάση το χρώμα (εξαγωγή κυρίων χρωμάτων / dominant colour extraction), με διάφορα συστήματα χρωμάτων. Στο κεφάλαιο 6 παρουσιάζεται η ταξινόμηση εικόνων με βάση το χρώμα και την υφή. Εξετάζονται τρόποι συνδυασμού αυτών καθώς μελετάται το πόσο βελτιώνονται τα ποσοστά επιτυχίας με την επιπρόσθετη χρήση του χρώματος. Στο κεφάλαιο 7 προτείνονται τρόποι εξαγωγής τμημάτων με βάση την υφή και το χρώμα για την περιγραφή του περιεχομένου της εικόνας. Επίσης παρατίθεται η τμηματοποίηση εικόνων με την προσθήκη του χρώματος στον αλγόριθμο τμηματοποίησης, όπου και γίνεται η σύγκριση στην περίπτωση που δεν έχουμε χρώμα.

Κεφάλαιο 2

Χαρακτηριστικά υφής

2.1 Γενικά

Υφή μπορεί να χαρακτηρισθεί οποιοδήποτε περιοδικό ομοιογενές πρότυπο στην εικόνα. Για παράδειγμα υφή μπορεί να χαρακτηρισθεί ένα ριγέ ύφασμα, μια περιοχή με γρασίδι, μια ξύλινη επιφάνεια κ.τ.λ. Το πρότυπο αυτό δεν μπορεί περιγραφεί με βάση την κατανομή της φωτεινότητας αλλά με το *πώς μεταβάλλεται η φωτεινότητα περιοδικά στα pixel της εικόνας.*

Στο παρελθόν έχουν προταθεί πληθώρα από στατιστικές προσεγγίσεις για την ανάλυση και επεξεργασία εικόνων που περιέχουν υφή [7],[8],[9], οι οποίοι βασίζονται σε χαρακτηριστικά που απορρέουν από συσχετίσεις σημείων και από μοντέλα τυχαίων πεδίων (random fields). Αργότερα προτάθηκαν μέθοδοι χρησιμοποιώντας χωρικά φίλτρα τοπικού χαρακτήρα [10],[11]. Ωστόσο σε αυτές τις προσεγγίσεις του προβλήματος υπάρχουν μειονεκτήματα, όπως μεγάλο υπολογιστικό κόστος, τα οποία λύνονται με την χρήση του μετασχηματισμού Wavelet [12], [13]. Έχουν συγκροτηθεί δύο κύριες οικογένειες/κατηγορίες μεθόδων που έχουν επικρατήσει τα τελευταία χρόνια: **multichannel filtering** ([14]) (Wavelets [12], Wavelet Packets [12], Wavelet Frames [15], Gabor φίλτρα [16] [17]) και **random fields** (Markov Random Fields, SAR και MR-SAR models [18]).

Σε αυτήν την εργασία το πρόβλημα του χαρακτηρισμού της υφής προσεγγίζεται με την Discrete Wavelet Frame ανάλυση, η οποία ανήκει στην πρώτη κατηγορία μεθόδων. Ο σκοπός της ανάλυσης είναι να καθοριστούν χαρακτηριστικά για κάθε υφή έτσι ώστε κάθε μια να μπορεί να οριστεί μονοσήμαντα. Από την φύση της υφής τέτοια διάκριση μπορεί να γίνει στο πεδίο των συχνοτήτων, όπου κάθε υφή έχει διαφορετική συμπεριφορά, δηλαδή διαφορετικές συχνότητες συνεισφέρουν, όσον αφορά το μέγεθος και την κατεύ-

θυνση, για την σύνθεση της κάθε υφής. Οπότε για τον σκοπό αυτό χρησιμοποιούνται φίλτρα που αποκόπτουν διαφορετικές ζώνες συχνοτήτων. Προσεγγίσεις σε multi-channel filtering διαφοροποιούνται κυρίως στα φίλτρα και στον τρόπο που τα χρησιμοποιούν. Οι μέθοδοι στην κατηγορία αυτή βασίζονται στην θεώρηση ότι η υφή είναι περιοδικό σήμα και ανεξάρτητο στην μετατόπιση (translation invariant).

Στις παρακάτω ενότητες παρουσιάζονται τα DWF, τα Gabor φίλτρα (δεύτερη κατηγορία), και τα πολλαπλής ανάλυσης ταυτόχρονης αυτοπαλινδρόμισης μοντέλα (Multi Resolution Simultaneous Auto Regressive models, MR-SAR), προσέγγιση που ανήκει στην δεύτερη κατηγορία.

2.2 Ανάλυση Discrete Wavelet Frames (DWF)

2.2.1 Καθορισμός φίλτρων και ιδιότητες αυτών

Η ανάλυση, όπως περιγράφηκε, πρέπει να διασπάει τις συχνότητες του σήματος εισόδου σε χαμηλές και υψηλές συχνότητες. Οπότε για τον σκοπό αυτό χρησιμοποιούνται φίλτρα h, g, όπου το h είναι το κατωδιαβατό (lowpass) φίλτρο και το g ανωδιαβατό (highpass) φίλτρο. Τα φίλτρα αυτά έχουν την ιδιότητα να είναι συμπληρωματικά έτσι ώστε να καλύπτουν όλο το φάσμα των συχνοτήτων. Οπότε έχοντας το h φίλτρο το g ορίζεται στις συχνότητες και στον χρόνο από τις εξής σχέσεις αντίστοιχα:

$$\left.\begin{array}{rcl} G(z) & = & zH(-z^{-1}) \\ g(n) & = & (-1)^{(1-n)}\, h(1-n) \end{array}\right\} \tag{2.1}$$

όπου τα G, H είναι οι μετασχηματισμοί $\mathcal{Z}$ των h, g στον χώρο των συχνοτήτων αντίστοιχα. Επίσης για τον Wavelet μετασχηματισμό χρειαζόμαστε την τέλεια ανακατασκευή του αρχικού σήματος από τα φίλτρα αυτά. Για να είναι εφικτό αυτό τότε το h να ικανοποιεί την συνθήκη των quadrature mirror filter που έχει ως εξής:

$$H(z)\, H(z^{-1}) + H(-z)\, H(-z^{-1}) = 1 \tag{2.2}$$

Στην εφαρμογή χρησιμοποιήθηκε σαν φίλτρο h που περιγράφεται κάτωθι με την συνάρτηση μεταφοράς του :

$$H(z) = \frac{z^2 + 4z + 6 + 4z^{-1} + z^{-2}}{16} \tag{2.3}$$

Τα πρότυπα αυτά φίλτρα μπορούν να επεκταθούν υπερδειγματοληπτώντας (upsampling) με παράγοντα 2 και να παράγουν καινούργια φίλτρα με επαναληπτικό τρόπο. Τότε

τα παραγόμενα φίλτρα θα έχουν ως εξής στον χώρο των συχνοτήτων και στον χρόνο αντίστοιχα:

$$\left.\begin{array}{rcl} H_{i+1}(z) & = & H(z^{2^i})\, H_i(z) \\ G_{i+1}(z) & = & G(z^{2^i})\, H_i(z) \end{array}\right\} \tag{2.4}$$

Όπου το i είναι ο δείκτης του τρέχοντος παραγόμενου φίλτρου, και με αρχική τιμή $H_0(z)$ =1. Ισοδύναμα στο χρόνο η σχέση γράφεται

$$\left.\begin{array}{rcl} h_{i+1}(k) & = & [h]_{\uparrow 2^i} * h_i(k) \\ g_{i+1}(k) & = & [g]_{\uparrow 2^i} * h_i(k) \end{array}\right\} \tag{2.5}$$

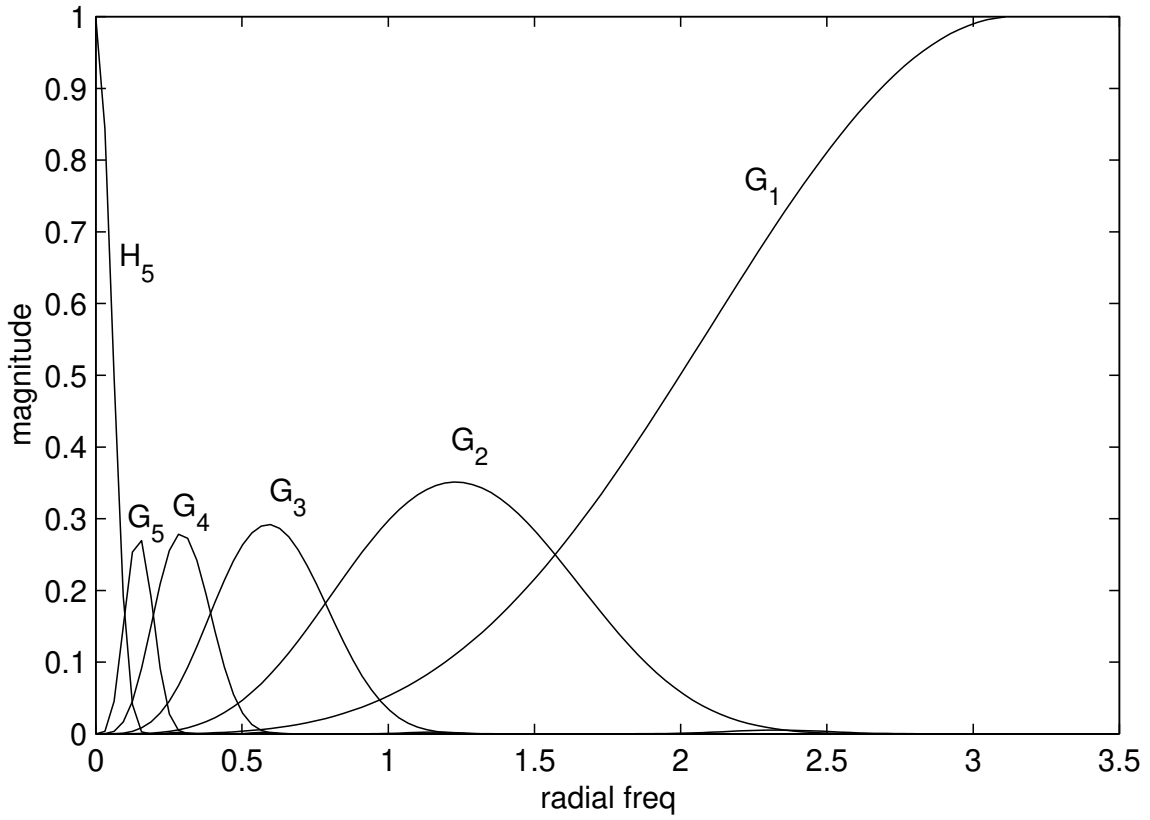
Όπου ο συμβολισμός $[h]_{\uparrow m}$ σημαίνει υπερδειγματοληψία με παράγοντα m, δηλαδή παρεμβολή $m-1$ μηδενικών ενδιάμεσα από κάθε συντελεστή του φίλτρου. Τέτοια φίλτρα μπορούν να αποσυνθέσουν ένα σήμα στο πεδίο των συχνοτήτων σε πολλά επίπεδα συχνοτήτων. Πιο συγκεκριμένα όσον αφορά την υλοποίηση κάθε επίπεδο απέχει από το προηγούμενο του και το επόμενο του κατά μια οκτάβα (απόσταση συχνοτήτων f_2, f_1 σε οκτάβες είναι $log_2(\frac{f_2}{f_1})$), επειδή χρησιμοποιείται το φίλτρο (2.3). Αυτό συμβαίνει γιατί το τρέχον κατωδιαβατό φίλτρο με το συμπληρωματικό του χωρίζουν τον χώρο των συχνοτήτων στην μέση σε κάθε επίπεδο της παραπάνω σχέσης. Σε κάθε επίπεδο ανάλυσης των συχνοτήτων διασπώνται όλο και πιο χαμηλές συχνότητες, κρατώντας με το $[g]_{\uparrow 2^i}$ φίλτρο το περιεχόμενο του αρχικού σήματος στο επίπεδο αυτό i των συχνοτήτων. Ενώ με το $[h]_{\uparrow 2^i}$ αποκόπτονται οι χαμηλότερες συχνότητες (h, g συμπληρωματικά) για να αναλυθούν στο επόμενο επίπεδο από το i (i + 1). Επίσης το ότι η παραπάνω σειρά φίλτρων καλύπτει όλο το πεδίο των συχνοτήτων επιβεβαιώνεται από την σχέση:

$$|H_i(e^{j2\pi f})|^2 + \sum_{k=1}^{i} |G_k(e^{j2\pi f})|^2 = 1 \tag{2.6}$$

Στο σχήμα 2.1 φαίνονται οι ζώνες των συχνοτήτων που αποκόπτονται σε κάθε επίπεδο του μετασχηματισμού. Τα επίπεδα της ανάλυσης που αναπαρίστανται στο σχήμα είναι 5, οπότε σύμφωνα με την 2.4, φαίνονται οι αποκρίσεις των $H_5(z)$, $G_5(z)$, $G_4(z)$, $G_3(z)$, $G_2(z)$, $G_1(z)$.

2.2.2 Wavelet Frames

Χρησιμοποιώντας μια σειρά διακριτών φίλτρων, όπως αυτά που περιγράφηκαν στην προηγούμενη ενότητα, μπορεί να επιτευχθεί η κατασκευή μίας ορθογώνιας Wavelet αποσύνθεσης (orthogonal Wavelet decomposition). Έτσι για αυτό τον σκοπό μπορούμε



Σχήμα 2.1: Αποκρίσεις στον χώρο των συχνοτήτων που αποκόπτονται σε 5 επίπεδα της ανάλυσης.

να θεωρήσουμε τις διακριτές κανονικοποιημένες συναρτήσεις βάσεις (basis functions) :

$$\left.\begin{array}{l} \varphi_{i,l}(k) = 2^{i/2} h_i(k - 2^i l) \\ \psi_{i,l}(k) = 2^{i/2} g_i(k - 2^i l) \end{array}\right\} \tag{2.7}$$

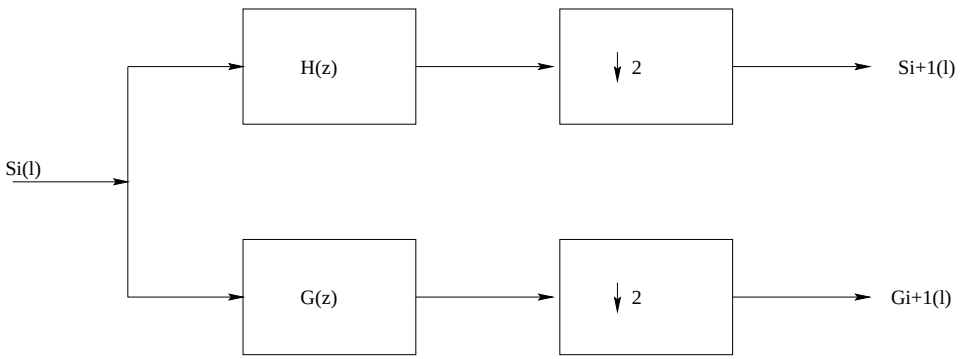
όπου i, l είναι οι δείκτες κλιμάκωσης και μετατόπισης αντίστοιχα. Οι συναρτήσεις αυτές παράγουν τους χώρους V_i, W_i, όπου V, W ονομάζονται χώροι προσέγγισης και υπολοίπου αντίστοιχα. Βάση των ιδιοτήτων των φίλτρων, όπως περιγράφηκαν, έχουμε $V_{i-1} = V_i \oplus W_i$ και $V_i \perp W_i$. Οπότε έχουμε μια ορθοκανονική βάση των χώρων αυτών παραγόμενη από τις φ, ψ. Για να υπολογισθεί η ορθογώνια προβολή ενός διακριτού σήματος στους χώρους αυτούς, δεν έχουμε παρά να υπολογίσουμε το εσωτερικό γινόμενο του σήματος με τις αντίστοιχες συναρτήσεις βάσεις. Έτσι μπορούμε να κατασκευάσουμε μια διακριτή Wavelet ανάλυση με βάθος ανάλυσης I, για την οικογένεια συναρτήσεων $< \varphi_I, \psi_1, \psi_2, \ldots, \psi_I >$, η οποία ανάλυση περιγράφεται από την παρακάτω εξίσωση

$$x(k) = \sum_{l \in Z} s_I(l)\, \varphi_{I,l} + \sum_{i=1}^{I} \sum_{l \in Z} d_i(l)\, \psi_{i,l} \tag{2.8}$$

όπου s_I είναι η προσσέγγιση στο βάθος ανάλυσης I και d_i οι λεπτομέρειες (υπόλοιπα) της ανάλυσης **Wavelet** του αρχικού σήματος. Τα $s_I, d_i i$=1..I (**wavelet coefficients**) υπολογίζονται από τον εξής αλγόριθμο αποσύνθεσης που προκύπτει από τα παραπάνω :

$$\left.\begin{array}{rcl} s_I(l) & = & 2^{I/2} [h_I^T * x]_{\downarrow 2I}(l) \\ d_i(l) & = & 2^{i/2} [g_i^T * x]_{\downarrow 2i}(l) \end{array}\right\} \tag{2.9}$$

όπου το σύμβολο $x(k)^T = x(-k)$, και $[.]_{\downarrow m}$ δειγματολειψία ανά m δείγματα. Ο αλγόριθμος αυτός φαίνεται και στο σχήμα (2.2).



Σχήμα 2.2: Discrete Wavelet Transform - 1D decomposition, όπου s_i, d_i προσέγγιση και υπόλοιπο στο επίπεδο i αντίστοιχα

Όμως παρατηρείται ότι μια ακέραια μετατόπιση του υπό ανάλυση σήματος θα έχει σαν αποτέλεσμα μια μη τετριμμένη τροποποίηση του μετασχηματισμού Wavelet, ο οποίος περιγράφηκε παραπάνω. Οπότε αν χρησιμοποιόταν η ανάλυση αυτή για να εξαχθούν χαρακτηριστικά υφής δεν θα ήταν αυτή επαρκής γιατί έχει θεωρηθεί η υφή σαν σήμα ανεξάρτητο κατά την μετατόπιση, δηλαδή δεν εκμεταλλεύεται η ανάλυση την περιοδικότητα της υφής. Έτσι για να ξεπερασθεί το πρόβλημα αυτό χρειαζόμαστε μια ανάλυση ανεξάρτητη από την μετατόπιση, και αυτόν τον λόγο εισάγονται τα Wavelet Frames (κυματομορφές πλαισίου). Οι οποίες βασίζονται στην ιδέα του υπολογισμού του μετασχηματισμού Wavelet για όλες τις δυνατές μετατοπίσεις του σήματος εισόδου. Αυτό θα απαιτούσε την θεώρηση μιας οικογένειας από φίλτρα που θα έχουν ως εξής :

$$< g_1(k-l), ..., g_I(k-l), h_I(k-l) > leZ \tag{2.10}$$

όπου l ακέραια μετατόπιση. Θεώρηση που μας οδηγεί στην ανάλυση **Discrete Wavelet Frames** με Wavelet Frame coefficients που περιγράφονται από την σχέση

$$x(k) = \sum_{leZ} s_I(l)\, h_I(k-l) + \sum_{i=1}^{I} \sum_{leZ} d_i(l)\, g_i(k-l) \tag{2.11}$$

η οποία συνεπάγει την σχέση :

$$x(k) = s_I(k) * h_I + \sum_{i=1}^{I} d_i(k) * g_i \tag{2.12}$$

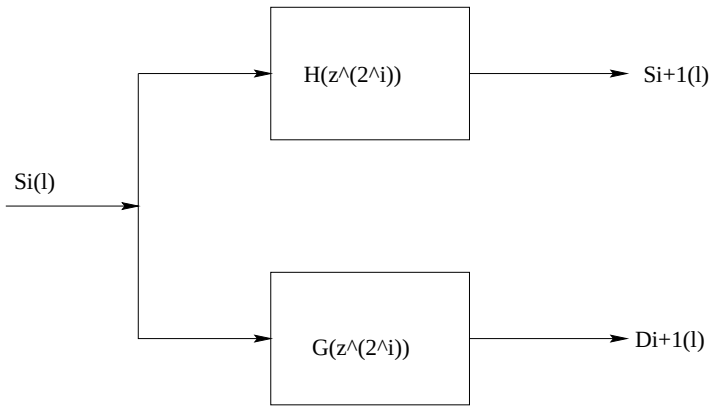
η σχέση αυτή περιγράφει τον αλγόριθμο ανακατασκευής του αρχικού σήματος από τα wavelet frame coefficients. Οι συνιστώσες (coefficients) της ανάλυσης μπορούν υπολογιστούν από έναν γρήγορο επαναληπτικό αλγόριθμο :

$$\left.\begin{aligned} s_{i+1}(k) &= [h]_{\uparrow 2^i} * s_i(k) \\ d_{i+1}(k) &= [g]_{\uparrow 2^i} * s_i(k) \end{aligned}\right\} \tag{2.13}$$

Αυτές οι δύο εξισώσεις είναι οι εξισώσεις αποσύνθεσης του αρχικού σήματος. Εδώ φαίνεται η διαφορά των Wavelet frames από την κλασική ανάλυση Wavelet όπου το σήμα κατωδειγματοληπτόταν (**down-sample**) ενώ τώρα υπερδειγματοληπτείται (**up-sample**) το φίλτρο, και στις δύο περιπτώσεις για την εξαγωγή λεπτομερειών που αντιπροσωπεύουν χαμηλότερες συχνότητες από το αρχικό σήμα. Ο αλγόριθμος αυτός φαίνεται και σχηματικά με διάγραμμα ροής στο σχήμα (2.3).

2.2.3 2-Δ επέκταση των DWF για εξαγωγή χαρακτηριστικών υφής

Η ανάλυση που περιγράφηκε για να χρησιμοποιηθεί σε εικόνες πρέπει να επεκταθεί στις δύο διαστάσεις. Αυτό μπορεί να επιτευχθεί αν πάρουμε σαν βάση τις συναρτήσεις



Σχήμα 2.3: Discrete Wavelet Frame 1D decomposition, όπου s_i, d_i προσέγγιση και υπόλοιπο στο επίπεδο i αντίστοιχα

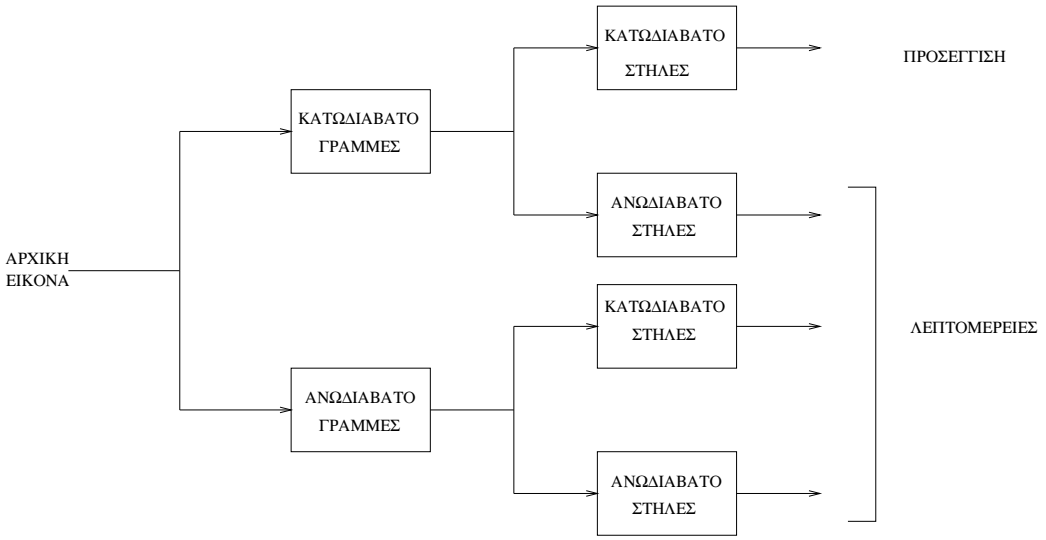
που προκύπτουν από όλους τους συνδυασμούς των γινομένων των συναρτήσεων βάσεων φ, ψ σε κάθε κατεύθυνση. Δηλαδή έχουμε :

$$\begin{aligned} \Phi(x,y) &= \phi(x)\,\phi(y) \quad & \Psi_1(x,y) &= \phi(x)\,\psi(y) \\ \Psi_2(x,y) &= \psi(x)\,\phi(y) \quad & \Psi_3(x,y) &= \psi(x)\,\psi(y) \end{aligned} \tag{2.14}$$

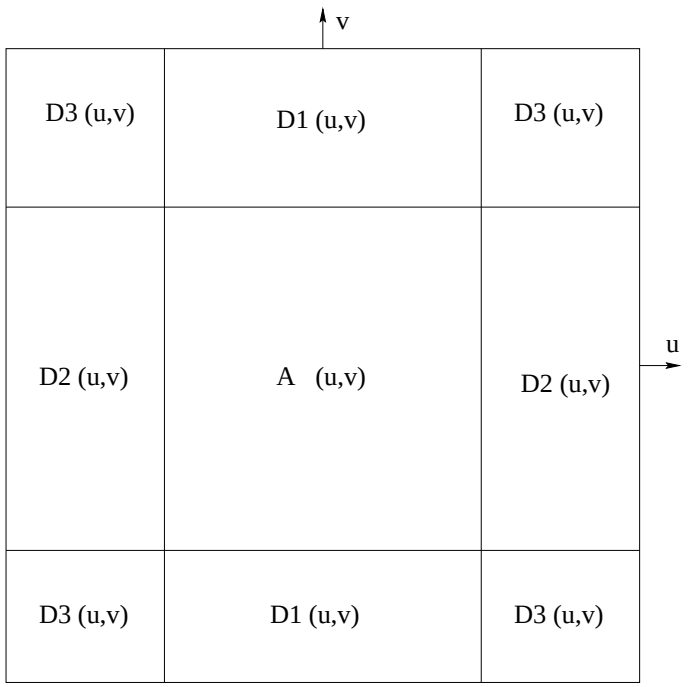
όπου τα $\Phi, \Psi_1, \Psi_2, \Psi_3$ είναι οι 2-Δ wavelet συναρτήσεις βάσης όπου οι φ, ψ ορίζονται στην (2.7). Επειδή είναι διαχωρίσιμες η ανάλυση έχει λιγότερο υπολογιστικό κόστος, αφού οι γραμμές και οι στήλες τις εικόνας επεξεργάζονται ξεχωριστά με μονοδιάστατα φίλτρα. Ο αλγόριθμος της 2-Δ αποσύνθεσης για εικόνες έχει ως εξής :

$$\left.\begin{aligned} d_{1,i+1}(k,l) &= [h]_{\uparrow 2^i}(k) * [g]_{\uparrow 2^i}(l) * s_i(k,l) \\ d_{2,i+1}(k,l) &= [g]_{\uparrow 2^i}(k) * [h]_{\uparrow 2^i}(l) * s_i(k,l) \\ d_{3,i+1}(k,l) &= [g]_{\uparrow 2^i}(k) * [g]_{\uparrow 2^i}(l) * s_i(k,l) \\ s_{i+1}(k,l) &= [h]_{\uparrow 2^i}(k) * [h]_{\uparrow 2^i}(l) * s_i(k,l) \end{aligned}\right\} \tag{2.15}$$

όπου το (k,l) είναι ένα σημείο της εικόνας, $s_0(k,l)$ η αρχική εικόνα, $[\,]_{\uparrow m}$ είναι υπερδειγματοληψία με παράγοντα m, $d_{1,i+1}, d_{2,i+1}, d_{3,i+1}$ είναι οι λεπτομέρειες βάθους ανάλυσης i + 1 (επίπεδο i + 1) και s_{i+1} είναι η προσέγγιση της ανάλυσης στο i + 1 επίπεδο. Ο αλγόριθμος παρουσιάζεται σχηματικά στο σχήμα (2.4), ενώ στο σχήμα (2.5) φαίνεται πως η ανάλυση αυτή διαχωρίζει τις συχνότητες, όπου A, D_1, D_2, D_3 είναι οι περιοχές των συχνοτήτων που αποκόπτονται σε κάθε επίπεδο ανάλυσης. Με A συμβολίζονται οι συχνότητες της προσέγγισης (wavelet coefficient s_{i+1}). Ενώ με D_1, D_2, D_3 συμβολίζονται οι κατακόρυφες, οριζόντιες και οι διαγώνιες συχνότητες (wavelet coefficients d_1, d_2, d_3) αντίστοιχα.



Σχήμα 2.4: Discrete Wavelet Frame 2-D ανάλυση: h κατωδιαβατό, g ανωδιαβατό



Σχήμα 2.5: Τρόπος διχωρισμού των συχνοτήτων στην Discrete Wavelet Frame

2.3 Φίλτρα Gabor, Gabor Wavelets

Όπως περιγράφηκε η ανάλυση Discrete Wavelet Frames (DWF) σε κάθε επίπεδο συχνοτήτων που αναλύει προσθέτει 3 καινούργια wavelet coefficients ένα που απομονώνει τις οριζόντιες συχνότητες, ένα που απομονώνει τις κατακόρυφες συχνότητες και

τέλος ένα που απομονώνει τις διαγώνιες συχνότητες. Υπάρχει όμως ένα μειονέκτημα, *δεν διαχωρίζονται οι διαγώνιες συχνότητες*, διότι θεωρούνται ενιαία οι συχνότητες στις κατευθύνσεις 45^o, 135^o. Αυτό έχει σαν συνέπεια, π.χ. σε μια εικόνα προς τμηματοποίηση, η οποία περιέχει δύο υφές που διαφέρουν μόνο στην κατεύθυνση των διαγωνίων συχνοτήτων, να μην μπορεί η ανάλυση DWF να διαχωρίζει τις δύο περιοχές. Αυτή η διάκριση μπορεί να γίνει χρησιμοποιώντας φίλτρα Gabor [16]. Τα Gabor φίλτρα έχουν την ιδιότητα να κόβουν μια συγκεκριμένη περιοχή συχνοτήτων και σε συγκεκριμένη κατεύθυνση, έτσι ώστε να μπορούν να καλύψουν όλο τον χώρο των συχνοτήτων. Τα φίλτρα αυτά διαφέρουν στην **συχνότητα αποκοπής**, στην **κατεύθυνση** και στο **εύρος** των συχνοτήτων που αποκόπτουν. Στην εργασία αυτή υλοποιήθηκαν Gabor φίλτρα άρτιας συμμετρίας. Η κρουστική απόκριση ενός τέτοιου φίλτρου είναι :

$$h(x,y) = exp\{-\frac{1}{2}[\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}]\} * cos(2\pi u_0 X) \tag{2.16}$$

με

$$X = xcos(\theta) + ysin(\theta) \tag{2.17}$$

Όπου u_0 είναι το κέντρο της περιοχής των συχνοτήτων που θέλουμε να απομονώσουμε, $\sigma_x\sigma_y$ είναι οι σταθερές χώρου του Gaussian φακέλου. Όπου θ είναι η διεύθυνση των συχνοτήτων (orientation) ως προς τον οριζόντιο άξονα x. Μετασχηματίζοντας στις συχνότητες με Fourier μετασχηματισμό για θ=0 (orientation στον x-άξονα) έχουμε :

$$H(u,v) = A(exp\{-\frac{1}{2}[\frac{(u-u_0)^2}{\sigma_u^2} + \frac{v^2}{\sigma_v^2}]\} + exp\{-\frac{1}{2}[\frac{(u+u_0)^2}{\sigma_u^2} + \frac{v^2}{\sigma_v^2}]\}) \tag{2.18}$$

Όπου $\sigma_u = \frac{1}{2\pi\sigma_x}$, $s_v = \frac{1}{2\pi\sigma_y}$ και $A = 2\pi\sigma_x\sigma_y$.

Για την κατάταξη ενός σημείου υφής χρειάζεται πληροφορία στον χώρο των συχνοτήτων αλλά επίσης και στον χώρο (spatial). Αυτό είναι γιατί φίλτρα με μικρό εύρος στις συχνότητες διαχωρίζουν καλύτερα τις υφές. Ενώ στον χώρο φίλτρα με μικρό εύρος δίνουν καλύτερο localization. Αυτές οι δύο ιδιότητες είναι αντιστρόφως συνδεδεμένες λόγο αντιστρόφως ανάλογης σχέσης χώρου-συχνοτήτων. Τα Gabor φίλτρα μπορούν και συνδυάζουν και τις δύο αυτές ιδιότητες.

Η επιλογή των παραμέτρων έγινε έτσι ώστε τα φίλτρα να χωρίζουν τις συχνότητες με μικρότερη επικάλυψη και να καλύπτουν όλο το φάσμα των συχνοτήτων. Έτσι με τα κριτήρια αυτά το εύρος των συχνοτήτων B_r και το εύρος της κατεύθυνσης (orientation) B_θ για έχουμε επικάλυψη των στις συχνότητες στο μισό του μεγίστου της συνάρτησης μεταφοράς (half-peak support), δίνονται από τις παρακάτω σχέσεις:

$$B_r = log_2(\frac{u_0 + S\sigma_u}{u_0 - S\sigma_u}) \tag{2.19}$$

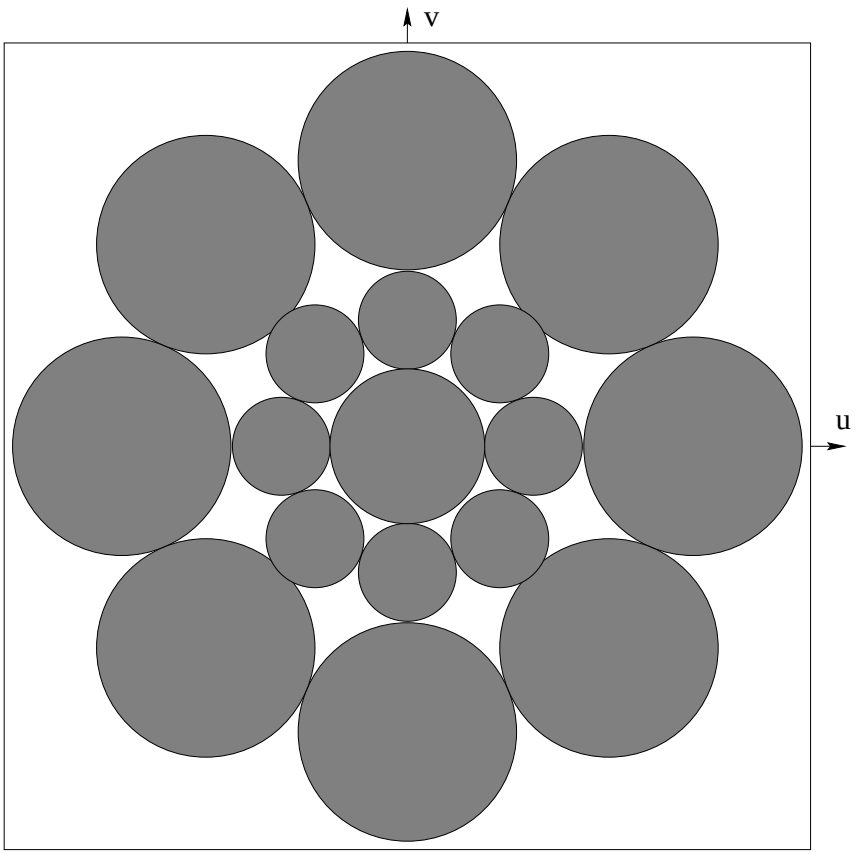
$$B_\theta = 2tan^{-1}(\frac{S\sigma_v}{u_0}) \quad (2.20)$$

Όπου το B_r μετριέται σε οκτάβες και το B_θ σε μοίρες. Στην υλοποίηση η παράμετρος S επιλέχθηκε ίσον με $(2ln2)^{\frac{1}{2}}$.

Στην εφαρμογή χρησιμοποιείται σαν εύρος γωνίας συχνοτήτων B_θ 45^o, οπότε έχουμε 4 διαφορετικές κατευθύνσεις συχνοτήτων που διαφέρουν κατά 45^o με $\theta = 0^o, 45^o, 90^0, 135^o$. Σαν αρχική συχνότητα u_0 χρησιμοποιείται η $\frac{N}{4} * \sqrt{2}$ (κύκλους$*$(εύρος$_\varepsilon$ικόνας)$^{-1}$), όπου N ο αριθμός των pixels στις στήλες, θεωρώντας τις συχνότητες στο διάστημα $[-N/2 \;\; N/2]$. Για τους παραπάνω τύπους η τιμή αυτή του u_0 κανονικοποιείται στην μονάδα (διαίρεση με N) φέρνοντας τις συχνότητες στο διάστημα $[-1/2 \;\; 1/2]$. Αυτό γίνεται για να ισχύουν οι παραπάνω σχέσεις και για να εφαρμόζεται η ανάλυση και σε εικόνες με διαφορετικό αριθμό στηλών γραμμών. Χρησιμοποιώντας τους παραπάνω τύπους παίρνουμε τα φίλτρα που απομονώνουν τις πιο υψηλές συχνότητες σε 4 διαφορετικές διευθύνσεις.Για να πάρουμε χαμηλότερες συχνότητες το u_0 διαιρείται δια του 2 έτσι ώστε οι δύο συχνότητες να απέχουν κατά μια οκτάβα σε κάθε επίπεδο. Επίσης μια οκτάβα είναι το εύρος συχνοτήτων (frequency bandwidth) B_r έτσι ώστε να καλύπτουν όλο τον χώρο των συχνοτήτων και να μην επικαλύπτονται τα επιμέρως φίλτρα σε κάθε επίπεδο ανάλυσης. Παίρνοντας όλο και μικροτέρα u_0, διαιρώντας πάλι με το 2, συχνότητες αναλύονται σε πολλά επίπεδα και σε 4 διαφορετικά orientation σε κάθε επίπεδο. Οπότε για μια ανάλυση σε 3 επίπεδα θα έχουμε 12 φιλτραρισμένες εικόνες (gabor wavelets coefficients). Το πως χωρίζονται οι συχνότητες για δύο επίπεδο ανάλυσης φαίνεται στο σχήμα (2.6) στο οποίο επίσης φαίνεται πως οι λεπτομέρειες διαφέρουν σε κάθε επίπεδο μια οκτάβα ($u_1 = u_0/2$). Επίσης για τον υπολογισμό τον συντελεστών των φίλτρων αφαιρώντας την μέση τιμή των συντελεστών κάνουμε έτσι τα φίλτρα να έχουν μέση τιμή μηδέν οπότε και τα coefficients θα έχουν την ίδια μέση τιμή, έτσι ώστε να έχουμε ταύτιση με τα DWF. Ακόμα οι συντελεστές κανονικοποιούνται διαιρώντας σε κάθε επίπεδο με το $A = 2\pi\sigma_x\sigma_y$. Οπότε για κάθε σημείο της εικόνας έχουμε ένα χαρακτηριστικό διάνυσμα το οποίο δίνει την συνεισφορά της κάθε συχνότητας, που αναλύθηκε, στο σημείο αυτό της εικόνας, όπως ακριβώς και στην ανάλυση DWF.

2.4 MR-SAR μοντέλο

Η μέθοδος αυτή είναι μια υποπερίπτωση τυχαίων πεδίων Markov (MRF). Είναι μια στατιστική προσέγγιση που χαρακτηρίζει την υφή με βάση αριθμητικά στατιστικά



Σχήμα 2.6: Gabor Wavelets frequency decomposition: στο σχήμα φαίνονται οι περιοχές συχνοτήτων των λεπτομερειών σε κάθε επίπεδο, και της προσέγγισης δεύτερο επιπέδου (μεσαίος κύκλος). Συμμετρικοί κύκλοι (δηλαδή στην ίδια κατεύθυνση) στο αντίστοιχο επίπεδο ανάλυσης ανήκουν στην ίδια συνιστωσα της ανάλυσης (coefficient) διότι έχουμε φίλτρα άρτιας συμμετρίας

γνωρίσματα. Για παράδειγμα η SAR μέθοδος ταιριάζει ένα στατιστικό μοντέλο αυτό-παλλινδόμισης στην εικόνα βρίσκοντας τις **παραμέτρους του μοντέλου**, με τις οποίες χαρακτηρίζεται η υφή.

2.4.1 SAR μοντέλο

Έχουν προταθεί διάφορες περιπτώσεις του SAR μοντέλου όπως SAR ανεξάρτητο της περιστροφής (RISAR) και κυκλικό SAR (CAR) [18]. Στην παρούσα εργασία εξετάζεται το πιο βασικό μοντέλο SAR, από το οποίο προκύπτουν τα υπόλοιπα όπως και το MRSAR.

Έστω ότι $g(s)$ είναι η φωτεινότητα στο σημείο s=$(s1, s2)$ μιας εικόνας διαστάσεως $M \times M$, με $s1, s2$=$1, \ldots, M$. Το SAR μοντέλο μπορεί να εκφρασθεί ως:

$$g(s) = \mu + \sum_{r \in \mathcal{D}} \theta(r)\, g(s + r) + \epsilon(s) \tag{2.21}$$

όπου $\mathcal{D}$ είναι ένα σύνολο από γειτονικά pixel από το s. Η πιο κοινή γειτονιά (ή **μάσκα**) που χρησιμοποιείται ευραίως είναι η δευτέρας τάξεως που φαίνεται στο πίνακα 2.1.

(-1,1)	(0,1)	(1,1)
(-1,0)	(0,0)	(1,0)
(-1,-1)	(0,-1)	(1,-1)

Πίνακας 2.1: Μάσκα γειτονιάς δεύτερης τάξεως για το pixel στην θέση (0,0)

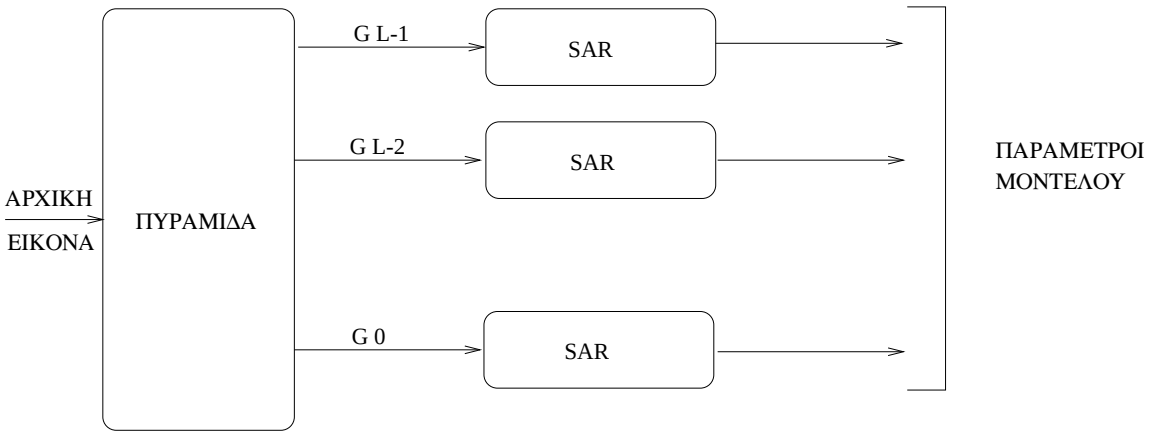
Στην εξίσωση (2.21), $\epsilon(s)$ είναι μια ανεξάρτητη Gauss τυχαία μεταβλητή με μέση τιμή μηδέν και διασπορά σ^2. $\theta(r)$, όπου το $r \in \mathcal{D}$, είναι οι παράμετροι του μοντέλου που χαρακτηρίζουν την εξάρτηση ενός σημείου με τα γειτονικά του, και μ είναι εξαρτώμενο από την μέση τιμή της φωτεινότητας της εικόνας. Όταν έχουμε $\theta(r) = \theta(-r)$ τότε το μοντέλο είναι συμμετρικό. Όλες οι παράμετροι του μοντέλου (μ, σ, $\theta(r)$) μπορούν να υπολογιστούν σε ένα δεδομένο παράθυρο στην εικόνα κάνοντας χρήση της μεθόδου του σφάλματος ελαχίστων τετραγώνων (Least Squares Error, LSE). Κατά την υλοποίηση η μέση τιμή αφαιρείται από το μοντέλο και δεν λαμβάνεται υπ' όψην στις παραμέτρους, δηλαδή σαν να αφαιρούμε την μέση φωτεινότητα από όλες τις θέσεις στην εικόνα.

2.4.2 Επέκταση σε πολλαπλή ανάλυση

Υπάρχουν δύο βασικές δυσκολίες που έχει το μοντέλο SAR στην εφαρμογή του. Το πρώτο είναι η επιλογή της γειτονιάς, στην οποία τα pixel θεωρούνται ως εξαρτόμενα.

Το δεύτερο είναι η επιλογή του κατάλληλου παραθύρου για τον υπολογισμό των παραμέτρων, στο οποίο η υφή μπορεί να θεωρηθεί ομοιογενής. Στις πιο πολλές προσεγγίσεις χρησιμοποιείται γειτονιά σταθερού μεγέθους (fixed-size neighborhood), η οποία κάνει την προσέγγιση μη προσαρμόσιμη σε όλες τις περιπτώσεις (nonadaptive). Για παράδειγμα σε εικόνες με υφή "λεπτών" χαρακτηριστικών (fine texture), μια μικρή γειτονία και ένα μικρό παράθυρο είναι αρκετά για τον χαρακτηρισμό της υφής, αλλά σε άλλες μπορεί να είναι απαραίτητη η χρήση και μικρών και μεγαλών γειτονιών και παραθύρων, έτσι ώστε να εξαχθεί πληρφορία σε διαφορετική κλίμακα και ανάλυση.

Ωστόσο δεν μπορούμε να χρησιμοποιήσουμε μεγάλες γειτονιές για όλες τις υφές. Έχει παρατηρηθεί από πειράματα ταξινόμησης ότι, η χρήση όλο και μεγαλύτερης γειτονιάς από ένα σημείο και πέρα ρίχνει το ποσοστό επιτυχίας. Αυτό συμβαίνει γιατί επιπρόσθετα σημεία στην μάσκα δεν προσθέτουν πληροφορία η οποία μπορεί να διαφοροποιήσει την υφή, αλλά συσχετίζουν σημεία τα οποία οποία δεν είναι εξαρτόμενα. Αυτό έχει σαν αποτέλεσμα να μειώνεται η δύναμη των παραμέτρων της γειτονιάς οι οποίες έχουν την διαφοροποιητική ικανότητα.



Σχήμα 2.7: Πολλαπλής ανάλυσης SAR μοντέλο με χρήση Gauss πυραμίδας L επιπέδων.

Η απόρροια των παραπάνω δείχνει ότι είναι ανώφελη η χρήση μεγάλων γειτονιών. Αυτό βέβαια δεν σημαίνει ότι σημεία της εικόνας που βρίσκονται μακριά είναι ανεξάρτητα μεταξύ τους σε όλες τις περιπτώσεις. Η λύση για την αντιμετώπιση του προβλήματος βασίζεται στην εξαγωγή παραμέτρων σε πολλαπλά επίπεδα αναλύσεων (**multi-resolution**).

Μια προσέγγιση της λύσης είναι η αναπαράσταση της εικόνας με μια Gauss πυραμίδα. Σε κάθε επίπεδο της πυραμίδας G_l η τρέχουσα εικόνα προκύπτει από την

προηγούμενη με την εφαρμογή ενός κατωδιαβατού φίλτρου και μιας δειγματολειψίας. Έπειτα σε κάθε επίπεδο ανάλυσης, μπορεί να εφαρμοστεί ένα μοντέλο SAR με σταθερή μάσκα και παράθυρο για την εξαγωγή παραμέτρων (σχήμα 2.7). Οπότε σε υψηλότερο επίπεδο ανάλυσης της πυραμίδας (fine resolution) εξάγονται παράμετροι που αντιστοιχούν σε λεπτά χαρακτηριστικά, ενώ σε χαμηλότερα (coarse resolution) παράμετροι που αντιστοιχούν σε βασικά δομικά στοιχεία της υφής, χονδρά χαρακτηριστικά (texture primitives). Αυτή η προσέγγιση έχει το μειονέκτημα της χρήσης των φίλτρων το οποίο προσθέτει υπολογιστικό κόστος και επίσης δυσκολία στην ανίχνευση τον συνόρων στην περίπτωση της τμηματοποίησης. Επίσης υπάρχει δυσκολία στην αντιστοίχιση κάθε σημείου στην αρχική εικόνα λόγω της αλλαγής της κλίμακας, σε κάθε επίπεδο το μέγεθος των εικόνων αλλάζει.

d2			d2			d2
	d1		d1		d1	
		d0	d0	d0		
d2	d1	d0	x	d0	d1	d2
		d0	d0	d0		
	d1		d1		d1	
d2			d2			d2

Πίνακας 2.2: Σύνολα γειτονιών $\mathcal{D}_l$ πολλαπλών επιπέδων για το MR-SAR μοντέλο.

Μια άλλη προσέγγιση είναι να κρατηθεί η ανάλυση της εικόνας σταθερή και να αλλάζει η μάσκα. Όσο μεγαλύτερο επίπεδο τόσο κράταμε πιο μακρινά σημεία στην γειτονιά. Στον πίνακα 2.2 φαίνονται πια σημεία χρησιμοποιούνται για την εξαγωγή παραμέτρων σε κάθε επίπεδο, $\mathcal{D}_l$ για το επίπεδο l, με $\mathcal{D}_l = d_l$. Όσον αφορά το μέγεθος του παραθύρου, θα μπορούσε κάλλιστα να χρησιμοποιηθεί μεταβλητό μέγεθος. Αφού όσο μικρότερο το επίπεδο τόσο πιο λεπτά χαρακτηριστικά άρα τόσο πιο λίγα σημεία για να υπολογίσουμε τις αντίστοιχες παραμέτρους. Αλλά αυτό δυσχεραίνει τις επιδόσεις της τμηματοποίησης, στον εντοπισμό των συνόρων, γιατί σε μεγάλυτερα επίπεδα το παράθυρο μεγαλώνει πολύ. Έτσι χρησιμοποιήθηκε μέγεθος παραθύρου σταθερό.

Κεφάλαιο 3

Ταξινόμηση υφής

3.1 DWF χαρακτηρισμός της υφής και κριτήριο ταξινόμησης

Η DWF ανάλυση μπορεί να εφαρμοστεί σε εικόνες που περιέχουν υφή, και να προκύψει το ακόλουθο αντιπροσωπευτικό διάνυσμα

$$y(k,l) =< y_1(k,l) , \ldots , y_{N-1}(k,l) , y_N(k,l) > \tag{3.1}$$

όπου κάθε στοιχείο $y(k,l)$ έχει υπολογιστεί σύμφωνα με την ανάλυση που περιγράφεται από την (2.15) και η διάσταση του διανύσματος είναι N=3I + 1 όπου I είναι το πλήθος των επιπέδων της ανάλυσης. Τα πρώτα $N-1$ στοιχεία του διανύσματος είναι και οι λεπτομέρειες της ανάλυσης ενώ το τελευταίο στοιχείο είναι η προσέγγιση της ανάλυσης στο I επίπεδο. Σε κάθε επίπεδο ανάλυσης προστείθονται τρεις καινούργιες συνιστώσες που διαφέρουν στην κατεύθυνση της συχνότητας που αντιπροσωπεύουν. Έτσι ανάλογα με την τιμή σε μια συγκεκριμένη συνιστώσα του διανύσματος των χαρακτηριστικών, μπορούμε να μετρήσουμε την συνεισφορά συχνοτήτων συγκεκριμένου μεγέθους και κατεύθυνσης (μέγεθος και κατεύθυνση εξαρτώνται από την συνιστώσα του διανύσματος), σε ένα σημείο (k,l) της εικόνας. Επομένως διαφορετικές υφές μπορούν να διακριθούν από το μέγεθος και την κατεύθυνση των συχνοτήτων, που συνεισφέρουν για την κατασκευή τους, και μόνο. Για αυτόν τον λόγο στην εργασία αυτήν η διάκριση διαφορετικών υφών βασίζεται μόνο στις $N-1$ λεπτομέρειες και όχι στην συνιστώσα της προσέγγισης. Επίσης ένας άλλος λόγος που η προσέγγιση παραλείπεται βασίζεται στην περίπτωση που δύο εικόνες με την ίδια υφή έχουν διαφορετική διασπορά στην προσέγγιση όταν οι αρχικές εικόνες έχουν διαφορά στην αντίθεση (contrast).

Οπότε κάθε υφή χαρακτηρίζεται από τις διασπορές των $N-1$ συνιστωσών $y_i(k,l)$ με τις υψηλότερες συχνότητες , με διάνυσμα διασπορών σ_i^2(i=$1,\ldots,N-1$). Αυτός ο

χαρακτηρισμός είναι εφικτός γιατί υπολογίζοντας τον πίνακα συνδιασπορών (covariance matrix) διαπιστώνεται ότι συσχετίσεις των συνιστωσών είναι πολύ μικρές, κοντά στο μηδέν, οπότε μπορούν να θεωρηθούν μηδέν. Αυτό επαληθεύει και το αληθές της ανάλυσης αφού τα φίλτρα σχεδιάστηκαν για να απομονώνουν διαφορετικές περιοχές συχνοτήτων με όσον το δυνατόν μικρότερες επικαλύψεις. Το γεγονός αυτό βοηθάει πολύ στην απλοποίηση του κριτηρίου που προτείνεται παρακάτω έτσι μειώνεται κατά πολύ το υπολογιστικό κόστος του αλγορίθμου. Επίσης πρέπει να σημειωθεί ότι οι λεπτομέρειες στο διάνυσμα των χαρακτηριστικών έχουν μηδενική μέση τιμή (γιατί $G(1) = H(-1) = 0$).

Υποθέτοντας τώρα ότι έχουμε Gauss συνάρτηση πυκνότητας πιθανότητας στις εικόνες της ανάλυσης το κριτήριο μέγιστης πιθανοφάνειας (maximum likelihood, ML) δίνει την απόσταση μιας υφής y από κατηγορία (κλάση) εικόνων υφής j.

$$d_j(y) = \sum_{(k,l)} \sum_{i=1}^{N-1} \left(\frac{y_i^2(k,l)}{\sigma_{i,j}^2} + \log \sigma_{i,j}^2 \right) = \sum_{i=1}^{N-1} \left(\frac{S_i^2}{\sigma_{i,j}^2} + \log \sigma_{i,j}^2 \right) \tag{3.2}$$

όπου το πρώτο άθροισμα αναφέρεται σε όλα τα σημεία της εξεταζόμενης προς κατηγοριοποίηση υφής, $\sigma_{i,j}^2$ είναι η διασπορά της i συνιστώσας της κλάσης j, και S^2 το διάνυσμα διασπορών που χαρακτηρίζει τη υπό εξέταση υφή y. Το κριτήριο αυτό προκύπτει από την εξίσωση

$$d_k(x) = (x - m_k)^T C_k^{-1} (x - m_k) + \log(det(C_k)) \tag{3.3}$$

η οποία υπολογίζει την απόσταση ενός σημείου με διάνυσμα ανάλυσης x από κλάση k όπου C_k ο πίνακας συνδιασποράς (covariance matrix) της κλάσης, m_k διάνυσμα μέσων τιμών σε κάθε επίπεδο (ανάλυσης) της κλάσης, και όπου A^T είναι ανάστροφος του A. Η απλοποίηση του κριτηρίου είναι εμφανής εάν ληφθεί υπόψην, ότι το διάνυσμα των μέσων τιμών έχει μηδενικές τιμές αφού κάθε wavelet coefficient έχει μέση τιμή μηδέν, και ότι από τον πίνακα συνδιασπορών κρατούνται μόνο τα στοιχεία της διαγωνίου (δηλαδή οι διασπορές σε κάθε επίπεδο ανάλυσης) αφού οι συσχετίσεις (correlation) μεταξύ των συνιστωσών είναι αμελητέες (δηλαδή τα στοιχεία εκτός διαγωνίου έχουν πολύ μικρές τιμές).

Στο σημείο αυτό φαίνεται το πλεονέκτημα της ανάλυσης αυτής έναντι του μετασχηματισμού Wavelet. Τώρα πλέον κάθε σημείο χαρακτηρίζεται από ένα διάνυσμα της ανάλυσης ενώ με τον μετασχηματισμό Wavelet δεν γίνεται αυτό λόγο της υποδειγματοληψίας.

Επίσης εξετάστηκε και ένα πιο ενδιαφέρον κριτήριο η απόσταση Bhattacharya. Αυτό είναι ελκυστικό διότι συνδέεται στενά με ελάχιστο λάθος κατά Bayes ταξινόμησης

στην περίπτωση δύο κλάσεων, αφού οι συγκρίσεις και στα παρακάτω πείραματα γίνονται ανά δύο. Πιο συγκεκριμένα είναι ειδική περίπτωση του Chernoff φράγματος της πιθανότητας λάθους κατά Bayes [19]. Κατόπιν στατιστικής ανάλυσης δείχθηκε ότι οι συνιστώσες της ανάλυσης που αντιπροσωπεύουν τις υψηλές συχνότητες (όχι προσεγγίσεις) ακολουθούν κατανομή που μπορεί να περιγραφεί από την γενικευμένη Gauss [12].

$$p(y) = \frac{c}{2\sigma\sqrt{\frac{1}{c}}e^{-(\frac{|y|}{c})^c}} \tag{3.4}$$

όπου η παράμετρος σ σχετίζεται με την διασπορά και η c με την οξύτητα συνάρτησης πυκνότητας πιθανότητας. Για c=2 έχουμε Gauss ενώ για c=1 Laplace. Αφού πρακτικά όπως προαναφέρθηκε δεν έχουμε συσχέτιση αναμεταξύ συνιστωσών της ανάλυσης, η απόσταση εκφράζεται ως:

$$d_j^B(x) = \frac{1}{c}\sum_{i=1}^{N-1} ln\frac{S_i^c + \sigma_{i,j}^c}{2\sqrt{S_i^c\sigma_{i,j}^c}} \tag{3.5}$$

Στην παρούσα εργασία χρησιμοποιείται η Bhattacharya απόσταση, και διερευνούνται οι επιδόσεις της για τιμές του c = (2,1), δηλαδή θεωρώντας Gauss ή Laplace κατανομή αντίστοιχα. Το κριτήριο ML χρησιμοποιείται περισσότερο για την τμηματοποίηση καθώς κατηγοριοποιεί σημείο σε κλάση.

Συνοψίζοντας μπορούμε πλέον να κατηγοριοποιήσουμε μια υφή προς εξέταση σε κάποια από τις εικόνες με υφή, που βρίσκονται σε μια βάση δεδομένων με εικόνες, χρησιμοποιώντας την DWF ανάλυση. Αυτό γίνεται υπολογίζοντας τα χαρακτηριστικά διανύσματα διασπορών για κάθε εικόνα στην βάση. Κάνοντας χρήση του παραπάνω κριτηρίου 3.5 μπορούμε με βάση την τιμή του, να διατάξουμε τις εικόνες στην βάση με σειρά ομοιότητας. Όσο πιο μικρή η τιμή τόσο πιο όμοιες είναι οι εικόνες που συγκρίνονται.

3.2 Gabor φίλτρα

Η διαφορά των φίλτρων Gabor από τη ανάλυση DWF είναι ότι διαχωρίζει τις δύο διαγώνιες συχνότητες, σε αντίθεση με την DWF που τις αποκόπτει σε μια συνιστώσα. Για να διαπιστωθεί, στα πειράματα που ακολουθούν σε πιο κάτω ενότητες, πόσο συνεισφέρει ο διαχωρισμός των διαγώνιων συχνοτήτων στο ποσοστό επιτυχίας, κρατούνται οι αποκρίσεις των DWF φίλτρων. Απλώς οι συνιστώσες που αφορούν τις διαγώνιες αντικαθιστούνται με τις αντίστοιχες δύο αποκρίσεις των φίλτρων Gabor που αφορούν τις δυο διαγώνιες του τρέχοντος επιπέδου ανάλυσης. Αυτό γίνεται γιατί τα φίλτρα Gabor

δεν είναι διαχωρίσιμα και αργεί πολύ κατά τον συγκερασμό. Οπότε χρησιμοποιείται η ίδια μεθοδολογία με πριν για την κατηγοριοποίηση.

3.3 MR-SAR

Για τον χαρακτηρισμό υφής με MRSAR μεταβλητής μάσκας χρησιμοποιήθηκαν διάφορες τιμές του αριθμού επιπέδων ανάλυσης. Οι γειτονιές, για το SAR μοντέλο σε κάθε επίπεδο, που χρησιμοποιήθηκαν φαίνονται στο σχήμα 2.2. Το μέγεθος του παραθύρου είναι όλη η εικόνα καθώς οι εικόνες αφορούν μια υφή. Τα χαρακτηριστικά του μοντέλου είναι οι παράμετροι της μάσκας καθώς και η τυπική απόκλιση του σφάλματος στον υπολογισμό των παραμέτρων με βάση την μέθοδο του σφάλματος των ελαχίστων τετραγώνων. Οπότε σε κάθε επίπεδο ο αριθμός των χαρακτηριστικών είναι 5, επειδή γίνεται χρήση συμμετρικού μοντέλου.

Για το κριτήριο ομοιότητας χρησιμοποιείται η απόσταση Mahalanobis:

$$d(x, y) = (x - y)^T C^{-1} (x - y) \tag{3.6}$$

που x, y είναι τα χαρακτηριστικά δύο υφών, και C είναι ο κοινός πίνακας συνδιασποράς όλων των χαρακτηριστικών. Όσο πιο μικρή τιμή του κριτηρίου τόσο πιο πολύ μοιάζουν οι δύο υφές.

3.4 Πρότυπο πείραμα

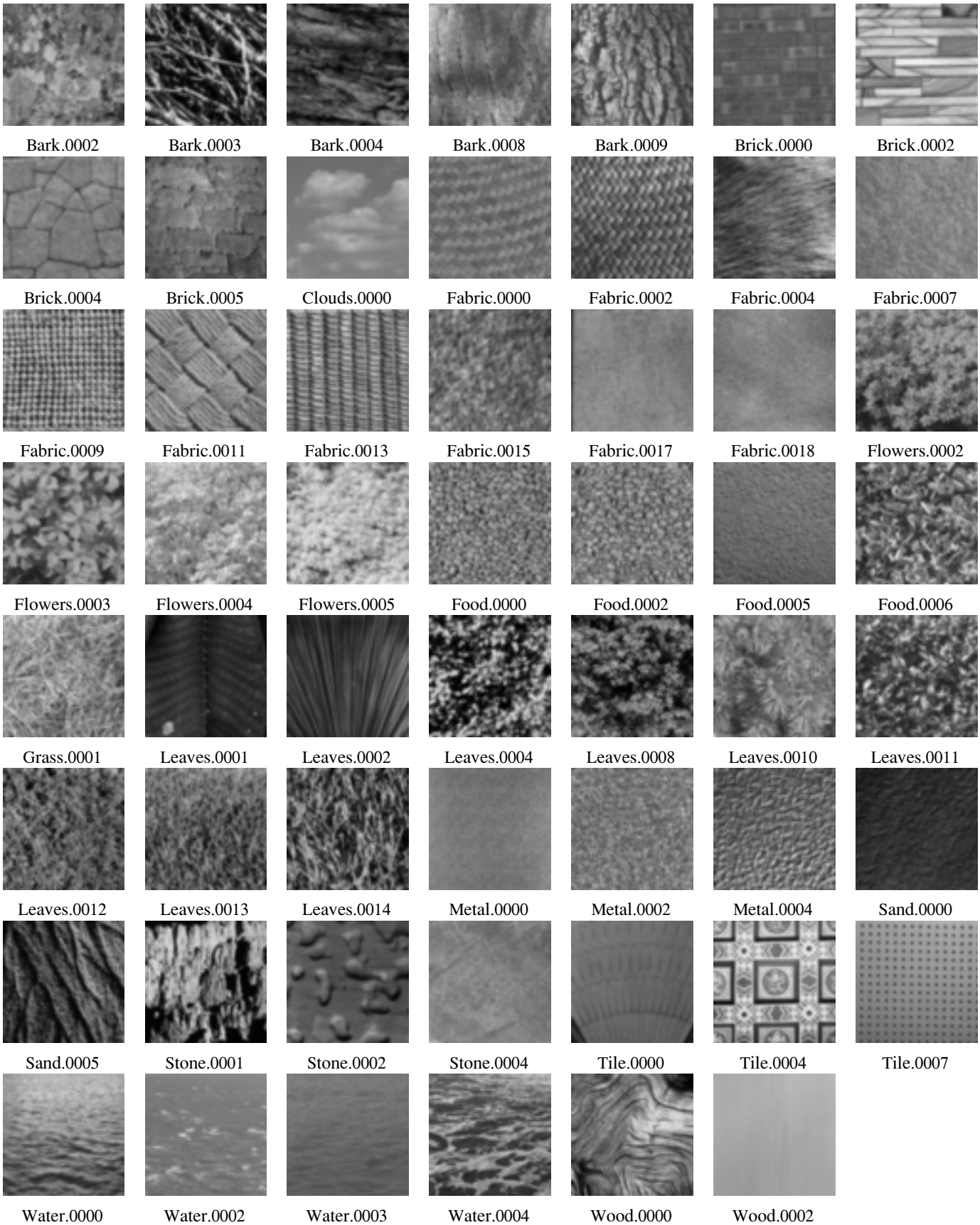
Για την σύγκριση των επιδόσεων των διαφόρων χαρακτηριστικών υφής γίνεται χρήση ενός πρότυπου πειράματος ταξινόμησης [20]. Όλες οι εικόνες της βάσης κόβονται σε ισάριθμα τμήματα ίσου μεγέθους, κάτι που προϋποθέτει όλες οι εικόνες να έχουν ίσες διαστάσεις. Οπότε και δημιουργείται ένας μεγάλος αριθμός από εικονίδια. Κάθε εικονίδιο χρησιμοποιείται για την ανάκληση των κοντινότερων εικονιδίων από την βάση, εκτός από τον εαυτό του, χρησιμοποιώντας τα αντίστοιχα χαρακτηριστικά και αντίστοιχο κριτήριο ανάλογα με την μέθοδο (DWF, MRSAR, Gabor). Σε κάθε αριθμό ανακληθέντων εικονιδίων, που πρέπει να είναι μεγαλύτερος ή ίσος από τον σταθερό αριθμό εικονιδίων ανά εικόνα, υπολογίζεται το ποσοστό επιτυχίας. Δηλαδή το ποσοστό των σωστών εικονιδίων που ανακλήθηκαν στο το σύνολο των σωστών. Με σωστό εικονίδιο να είναι το εικονίδιο που ανήκει στην ίδια εικόνα με αυτό που χρησιμοποιείται για ανάκληση. Κάνοντας την διαδικασία αυτή για όλα τα εικονίδια μπορεί να εξαχθεί γραφικά το ολικό ποσοστό επιτυχίας σε κάθε αριθμό εικονιδίων που ανακλήθηκαν.

Η βάση που χρησιμοποιήθηκε είναι ένα υποσύνολο της VisTex από το MIT [21]. Η βάση αυτή αποτελείται από φυσικές υφές, που φωτογραφήθηκαν σε φυσικές συνθήκες. Οι υφές είναι από ξύλα, φυτά, λουλούδια, τροφή, άμμο, κλαδιά, θάλασσα, ουρανό, υφάσματα και άλλες θεματικές περιοχές.

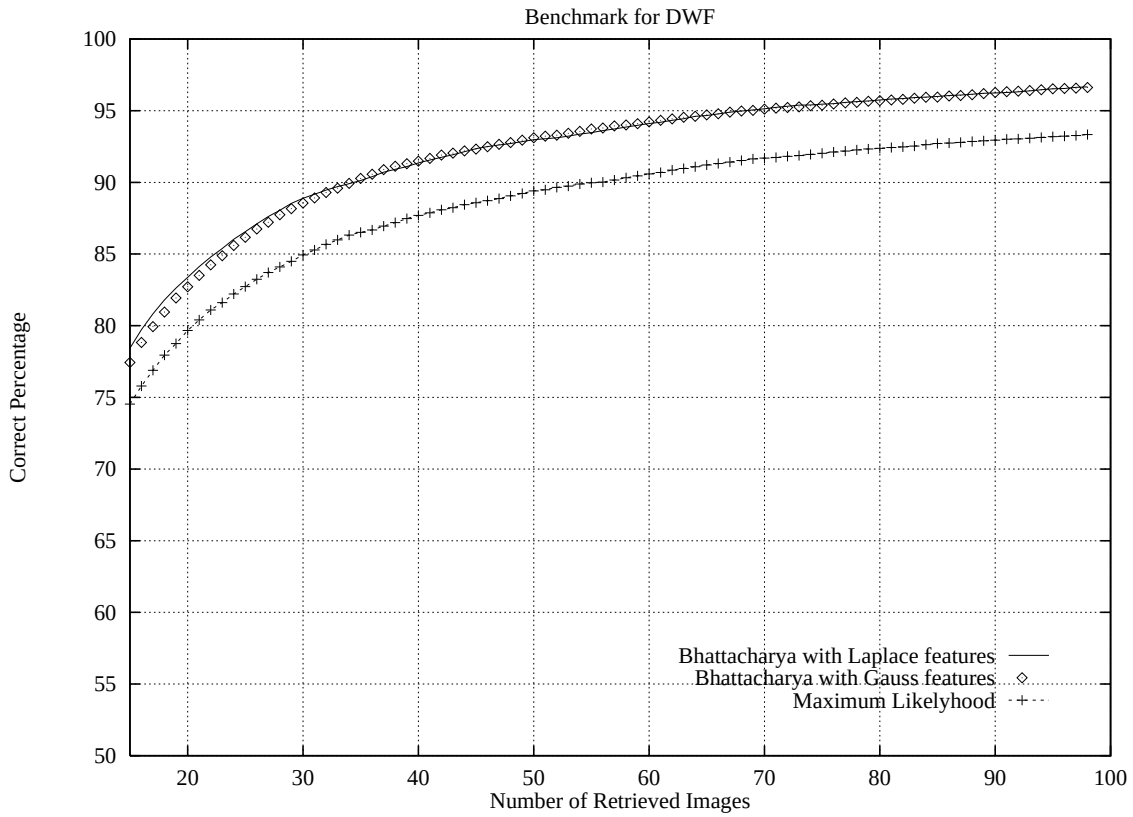
Από αυτές επιλέχθηκαν οι περισσότερο ομοιογενείς καθώς το πείραμα εξαρτάται από την ανομοιογένεια των εικόνων. Το πείραμα εξετάζει αν από ένα κομμάτι μια εικόνας μπορούμε να βρούμε τα υπόλοιπα τα οποία θα είναι παρόμοια, δηλαδή κατά πόσο κάποια χαρακτηριστικά υφής ανταπεξέρχονται στην ανομοιογένεια μιας εικόνας. Αν μια εικόνα είναι πολύ ανομοιογενής δεν έχει νόημα να συμπεριληφθεί στο πείραμα. Από την επιλογή αυτή έμειναν στην βάση 55 εικόνες οι οποίες φαίνονται στο σχήμα 3.1. Αυτές κόβονται σε κομάτια των 128×128 δημιουργώντας μια βάση από 880 εικονίδια, καθώς κάθε εικόνα έχει μέγεθος 512×512.

Στον σχήμα 3.2 φαίνονται τα ποσοστά επιτυχίας στο πείραμα με την DWF ανάλυση, κάνοντας χρήση της Bhattacharya για κατανομή Gauss και Laplace, και της απόστασης μέγιστης πιθανοφάνειας. Η ανάλυση έγινε σε 5 επίπεδα, δημιουργώντας διάνυσμα χαρακτηριστικών διάστασης 15. Από ότι φαίνεται στο σχήμα η Laplace λειτουργεί λίγο καλύτερα από την Gauss για την Bhattacharya. Συγκεκριμένα έχουν μια ποσοστιαία μονάδα διάφορα όταν έχουν ανακληθεί 15 εικονίδια από την βάση (76.8% έναντι 75.9%). Ενώ το ποσοστό επιτυχίας γίνεται 73.0% στην περίπτωση της απόστασης βασισμένης στην μέγιστη πιθανοφάνεια. Αυτό δείχνει, όπως προαναφέρθηκε, την καλύτερη σχεδίαση της Bhattacharya στην περίπτωση που έχουμε την ελαχιστοποίηση του σφάλματος όταν έχουμε δύο κλάσεις. Όπως στην περίπτωση εδώ που οι αποστάσεις υπολογίζονται ανά ζεύγη κλάσεων. Επίσης η Bhattacharya έχει την ιδιότητα της συμμετρικότητας, που δεν έχει η ML απόσταση. Αυτό σημαίνει ότι, αν υπολογίζουμε την απόσταση της υφής x από την υφή y, είναι η ίδια όταν υπολογίσουμε την απόσταση της y από την x. Όσο αυξάνεται ο αριθμός των εικόνων που ανακλήθηκαν τόσο αυξάνεται και το ποσοστό επιτυχίας, πράγμα που ισχύει εκ κατασκευής του πειράματος.

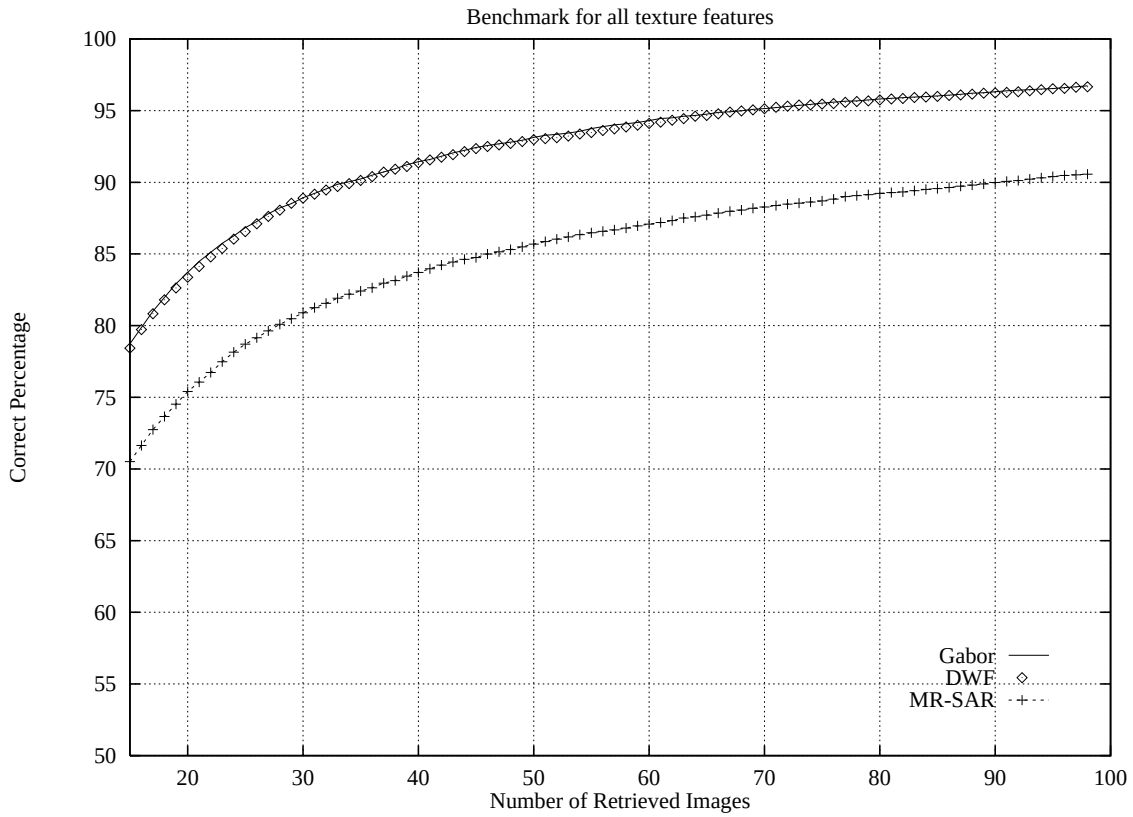
Επίσης το πείραμα εφαρμόστηκε για την MR-SAR προσέγγιση για διάφορα επίπεδα ανάλυσης. Δοκιμάστηκαν τρεις διαφορετικές τιμές για τον αριθμό των επιπέδων, 3, 4 και 5. Ο αριθμός των χαρακτηριστικών, που εξάγονται σε κάθε επίπεδο είναι 5, έχοντας 15, 20 και 25 διάσταση στο διάνυσμα χαρακτηριστικών για τους αριθμούς επιπέδων 3, 4, και 5 αντίστοιχα. Διαπιστώθηκε ότι για 3 και 4 επίπεδα ανάλυσης είχαμε τα μεγαλύτερα ποσοστά επιτυχίας, 69.22% και 69.28% αντίστοιχα. Από εκεί και μετά το ποσοστό φθίνει, έχοντας στα 5 επίπεδα ποσοστό επιτυχίας 68.9%. Από το τρίτο επίπεδο στο τέταρτο η



Σχήμα 3.1: Η βάση με τις 55 εικόνες που περιέχουν υφή από το VisTex του MIT.



Σχήμα 3.2: Καμπύλη επιτυχίας, συναρτήσει του αριθμού εικόνων που ανακλήθηκαν από την βάση, για την DWF ανάλυση. Γίνεται σύγκριση όταν χρησιμοποιούμε την Bhattacharya με Laplace και με Gauss κατανομή άλλα και την Maximum Likelyhood.



Σχήμα 3.3: Καμπύλη επιτυχίας, συναρτήσει του αριθμού εικόνων που ανακλήθηκαν από την βάση, για όλες τις μεθόδους (Gabor, DWF (με Laplace), και MR-SAR).

αύξηση στο ποσοστό επιτυχίας είναι οριακή έως μηδαμινή, ενώ επιπλέον έχουμε αύξηση του υπολογιστικού κόστους, καθώς προστείθονται 5 τιμές χαρακτηριστικών. Επίσης αυξάνεται το υπολογιστικό κόστος στον υπολογισμό του αντιστρόφου κοινού πίνακα συνδιασποράς, ο οποίος χρησιμοποιείται στην απόσταση Mahalanobis (3.6). Το ότι δεν αυξάνεται το ποσοστό, αλλά μειώνεται κιόλας, οφείλεται στο ότι συσχετίζονται σημεία τα οποία δεν προσφέρουν πλέον διαχωριστική δυνατότητα, παρά μόνο μειώνουν την δύναμη των ήδη υπαρχόντων χαρακτηριστικών από τα προηγούμενα επίπεδα.

Στον σχήμα 3.3 φαίνονται τα ποσοστά επιτυχίας στο πείραμα με όλες τις μεθόδους DWF, φίλτρα Gabor, MR-SAR. Το βάθος της ανάλυσης είναι 5 για την DWF και τα φίλτρα Gabor, έχοντας έτσι διάνυσμα χαρακτηριστικών διαστάσεως 15 και 20 αντίστοιχα. Το επίπεδα ανάλυσης για την DWF είναι 3, έχοντας έτσι και αυτή διάσταση 15 στο διάνυσμα χαρακτηριστικών. Όπως φαίνεται η συνεισφόρα του διαχωρισμού των διαγώνιων συχνοτήτων είναι πολύ μικρή της τάξεως της μισής ποσοστιαίας μονάδας. Ενώ και οι δύο μέθοδοι πολυκαναλικού φιλτραρίσματος υπερέχουν της MR-SAR που ανήκει στην οικογένεια των τυχαίων πεδίων, με DWF 76.8%, Gabor 77.2% και MR-SAR 69.2% ποσοστό επιτυχίας.

Στον πίνακα 3.1 παρατείθονται τα αναλυτικά αποτελέσματα για τα ποσοστά επιτυχίας σε κάθε εικόνα από τις 55 της βάσης. Από τα ποσοστά του πίνακα μπορούμε να διαπιστώσουμε το ότι κατά μέσο όρο η DWF π.χ είναι πολύ καλύτερη από την MR-SAR δεν σημαίνει ότι υπερτερεί σε ποσοστό επιτυχίας σε όλες τις επιμέρους εικόνες. Αντιθέτως σε πολλές περιπτώσεις η MR-SAR είναι συντριπτικά καλύτερη από τις άλλες δύο μεθόδους. Η MR-SAR μέθοδος έχει καλύτερα αποτελέσματα στις περιπτώσεις όπου έχουμε μη δομημένες υφές, που δεν έχουν κυρίαρχες συνιστώσες στο πεδίο των συχνότητων, δηλαδή αυτές που είναι ”χαοτικές” κατά μια έννοια. Αυτό συμβαίνει λόγω κατασκευής της μεθόδου, η οποία στα χαρακτηριστικά της ”αποθηκεύει” τον τρόπο που αλληλεπιδρούν (συσχετίζονται) γειτονικά σημεία. Προσέγγιση εντελώς διαφορετική από την διαχώριση των υφών ανάλογα με την συμπεριφορά τους στον χώρο των συχνοτήτων. Επίσης η MR-SAR όπως και στις δομημένες υφές, και στις πολύ ”αδρές” (coarse) υφές αδυνατεί να εξάγει την απαιτούμενη πληροφορία, αφού ”χρησιμοποιεί” συσχετίσεις τοπικά σε γειτονικά σημεία. Για τον σκοπό αυτό η διασπορά του λάθους στον υπολογισμό των παραμέτρων της μάσκας, με βάση την μέθοδο των ελαχίστων τετραγώνων, δεν είναι αρκετή για να εξαχθεί η ανάλογη πληροφορία. Αν και πολλαπλών επιπέδων ανάλυσης η μάσκα εξάγει πληροφορία από συν πλην τρία pixel, και από εκεί και πέρα οι συσχετίσεις δεν προσφέρουν την απαιτούμενη διαχωριστική ικανότητα. Ενώ το μέγεθος του φίλτρου

στην DWF φτάνει 65 pixel στο πέμπτο επίπεδο.

Έτσι στις υφές Brick.0002, Leaves.0001, Leaves.0002, Leaves.0010, Wood.0002, Fabric.0000, Fabric.0002, Fabric.0004, Fabric.0009, Fabric.0011, Flowers.0003, όπου είναι δομημένες και έχουν χαρακτηριστικές συνιστώσες στο πεδίο των συχνοτήτων, τα DWF και φίλτρα Gabor υπερτερούν. Επίσης δεν έχει καλές επιδόσεις η MR-SAR, στις υφές, Tile.0000, Tile.0004, Grass.0001, Leaves.0014, όπου έχουν πολύ αδρά χαρακτηριστικά. Ωστόσο η MRSAR υπερτερεί στις υφές Bark.0009, Fabric.0007, Fabric.0015, Fabric.0018, Food.0005, Leaves.0011, Leaves.0014, Leaves.0008, Stone.0004, Metal.0000, Meatal.0004, Clouds.0000, οι οποίες δεν είναι δομημένες και είναι σχετικά με λεπτά χαρακτηριστικά (fine). Επίσης συμπτωματικά η MR-SAR προσέγγιση έχει καλύτερη επίδοση στην Stone.0001 όπου η υφή είναι με αδρά χαρακτηριστικά αλλά με μεγάλη ανομοιογένεια σε αυτά, οπότε επηρεάζει την επίδοση των Gabor και DWF. Η υφή αυτή όμως διατηρεί σχετικά τις τοπικές συσχετίσεις σχετικά ίδιες στις περισσότερες θέσεις της εικόνας, δίνοντας καλύτερο ποσοστό στην MR-SAR.

Το συμπέρασμα που βγαίνει από τις παραπάνω διαπιστώσεις είναι ότι οι υφές διακρίνονται καλύτερα κυρίως στον ”χώρο” των συχνοτήτων και όχι με τις τοπικές συσχετίσεις. Όμως αυτό δεν είναι αρκετό, γιατί η επίδοση των μεθόδων με πολυκαναλικά φίλτρα αδυνατούν να ικανοποιήσουν όλες τις περιπτώσεις. Οπότε ιδανική θα ήταν μια προσέγγιση που θα συνδύαζε και τις δύο προσεγγίσεις, ανάλογα με την περίσταση, αν έχουμε δομημένη ή μη υφή. Προς αυτήν την κατεύθυνση κινήθηκαν οι Picard και Liu, για να βελτιώσουν την επίδοση του MR-SAR μοντέλου, χρησιμοποιώντας την Wold αποσύνθεση σε συνδυασμό με την MR-SAR [20]. Η αποσύνθεση αυτή δεν είναι τίποτα άλλο από την εύρεση κυρίαρχων αρμονικών στο πεδίο των συχνοτήτων των υφών. Για τον συνδυασμό πρότειναν ένα κριτήριο για να δείχνει το κατά πόσο μια υφή είναι δομημένη. Το κριτήριο αυτό το δοκίμασαν στην Brodatz βάση δεδομένων και δήλωσαν καλή συμπεριφορά αυτού. Εφαρμόζοντας όμως στην παρούσα εργασία το κριτήριο αυτό στις 55 εικόνες από το VisTeX δεν είχε καλή συμπεριφορά. Τα στατιστικά των τιμών του στις εικόνες δεν έδιναν την επιθυμητή μικτή (δομημένων και μη) κατανομή με δύο μέγιστα (bimodal). Επίσης το Wold μοντέλο σε καμία περίπτωση δεν μπορεί να συγκριθεί με τα χαρακτηριστικά των πολυκαναλικών μεθόδων, καθώς χρησιμοποιεί ένα πολύ μικρό μέρος του φάσματος για να εξάγει κυρίαρχα μόνο χαρακτηριστικά.

Όσον αφορά την βελτίωση που προσφέρουν τα Gabor φίλτρα, αυτή είναι εμφανής στις περιπτώσεις όπου έχουμε μια από τις δύο κύριες διαγωνίες κατευθύνσεις κυρίαρχη στην εικόνα, όπως στις Bark.0003, Fabric.0004, αλλά και τις δύο αλλά να είναι κάπως

εμφανείς (δομημένες), να ξεχωρίζουν μεταξύ τους και να μην είναι ”χαοτικές”, όπως στις υφές Tile.0007, Metal.0004. Στις υφές Grass.0001 και Stone.0004, παρατηρείται μείωση με τον διαχωρισμό των διαγωνίων συχνοτήτων, γιατί αυτές αναμιγνίονται με ακανόνιστο τρόπο χωρίς καμία δομή, όποτε ο διαχωρισμός δεν προσφέρει καμία επιπλέον πληροφορία για το περιεχόμενο. Στο σημείο αυτό πρέπει να επισημανθεί το μεγάλο μειονέκτημα που έχουν τα φίλτρα Gabor, το οποίο είναι το μεγάλο υπολογιστικό κόστος καθώς δεν είναι διαχωρίσιμα. Όσο μεγαλώνει το επίπεδο τόσο στενότερες ζώνες συχνοτήτων αλλά και χαμηλότερες συχνότητες αποκόπτονται δημιουργώντας έτσι μεγάλο εύρος φίλτρου πχ. 50 pixel. Σε αντίθεση με την DWF όπου τα φίλτρα είναι διαχωρίσιμα, οι επεξεργασία γίνεται σε κάθε κατεύθυνση ξεχωριστά. Και επιπλέον αν και τα φίλτρο φθάνει να έχει μέγεθος στο πέμπτο επίπεδο 65, ο χρόνος που αναλίσκουν οι συγκερασμοί ανά επίπεδο παραμένει ο ίδιος καθώς οι μηδενικές τιμές είναι οι ίδιες σε όλα τα φίλτρα. Απλώς αγνοούνται οι πολλαπλασιασμοί με τους μηδενικούς συντελεστές των φίλτρων, αφού είναι σε γνωστή θέση. Το γεγονός αυτό κάνει την DWF ανάλυση πιο ελκυστική.

3.5 Επιπλέον πειραματικά αποτελέσματα

Στην ενότητα αυτήν παρουσιάζονται πειραματικά αποτελέσματα που έγιναν για την DWF ανάλυση και την Brodatz βάση υφών [22]. Τα πειράματα αφορούν ένα υποσύνολο της Brodatz βάσης, καθώς υπήρχαν διαθέσιμα μόνο 20 δείγματα από το άλμπουμ, σε ένα σύνολο από 110 περίπου εικόνες.

Το πείραμα που έγινε είναι παρόμοιο με το πείραμα της προηγούμενης ενότητας. Πάλι γίνεται η χρήση εικονιδίων μεγέθους 128$\times$ 128, που εξάγονται από την αρχική εικόνα, για την εξαγωγή ποσοστών επιτυχίας. Η διαφορά είναι ότι εξάγονται 50 εικονίδια από κάθε εικόνα με τυχαίο τρόπο. Ακόμα η κάθε αρχική εικόνα κλάση έχει ένα ενιαίο διάνυσμα χαρακτηριστικών το οποίο εξάγεται από ολόκληρη την εικόνα. Κάθε εικονίδιο από τα 1000 (50$\times$ 20) στο σύνολο κατηγοριοποιείται ως προς τις 20 αρχικές εικόνες της βάσης. Χρησιμοποιήθηκαν οι αποστάσεις της μέγιστης πιθανοφάνειας και η Bhattacharya υποθέτοντας κατανομή χαρακτηριστικών Gauss. Τα αποτελέσματα αναλυτικά φαίνονται στους πίνακες 3.2, 3.3 για την Bhattacharya και μέγιστη πιθανοφάνεια αντίστοιχα. Ένας ”ενδιάμεσος” αριθμός στους πίνακες σημαίνει πόσα εικονίδια από την υφή στην τρέχουσα γραμμή κατατάχθηκαν στην υφή που βρίσκεται στην τρέχουσα στήλη. Για αυτό και τα στοιχεία στην κύρια διαγώνιο του πίνακα είναι μεγαλύτερα γιατί αντιπροσωπεύουν τις σωστές κατατάξεις. Οι δύο τελευταίες στήλες είναι το σύνολο των εικονιδίων σε κάθε

Όνομα υφής	DWF	Gabor	MR-SAR
Bark.0002	78.8	82.5	79.6
Bark.0003	54.2	69.6	53.8
Bark.0004	77.9	77.5	82.1
Bark.0008	56.7	59.2	59.6
Bark.0009	37.9	41.7	56.7
Brick.0000	87.9	86.2	88.8
Brick.0002	81.2	80.4	37.5
Brick.0004	76.7	76.2	52.5
Brick.0005	90.0	92.1	72.5
Clouds.0000	93.3	93.3	100.0
Fabric.0000	79.2	75.0	31.2
Fabric.0002	86.7	86.2	50.0
Fabric.0004	72.1	76.2	63.3
Fabric.0007	97.5	98.8	100.0
Fabric.0009	98.8	99.6	92.9
Fabric.0011	78.3	86.2	52.5
Fabric.0013	100.0	100.0	97.1
Fabric.0015	80.4	87.1	90.0
Fabric.0017	98.8	100.0	100.0
Fabric.0018	90.0	84.6	100.0
Flowers.0002	51.2	52.9	44.2
Flowers.0003	93.3	93.3	43.3
Flowers.0004	66.7	78.8	62.5
Flowers.0005	57.1	60.8	69.2
Food.0000	59.2	57.1	61.2
Food.0002	53.8	50.0	55.8
Food.0005	81.2	80.4	96.7
Food.0006	90.0	92.1	60.8
Grass.0001	91.7	81.2	53.8
Leaves.0001	78.3	75.8	40.0
Leaves.0002	89.2	82.1	34.6
Leaves.0004	86.2	92.1	96.2
Leaves.0008	65.4	62.5	69.2
Leaves.0010	57.5	54.2	29.6
Leaves.0011	52.5	51.2	92.1
Leaves.0012	75.4	72.1	70.8
Leaves.0013	61.2	59.2	60.0
Leaves.0014	76.2	71.7	36.7
Metal.0000	79.6	75.4	90.8
Metal.0002	98.3	99.6	100.0
Metal.0004	79.2	94.6	100.0
Sand.0000	96.2	96.7	89.2
Sand.0005	61.2	57.9	48.8
Stone.0001	57.9	59.6	67.1
Stone.0002	83.3	82.9	71.2
Stone.0004	88.3	84.2	90.4
Tile.0000	77.1	75.4	50.8
Tile.0004	98.8	98.3	55.8
Tile.0007	88.3	94.6	100.0
Water.0000	57.9	56.7	99.6
Water.0002	64.6	65.0	69.2
Water.0003	98.8	97.9	90.8
Water.0004	60.4	60.8	85.4
Wood.0000	34.6	29.6	20.4
Wood.0002	100.0	100.0	41.2
Σύνολο	76.8	77.2	69.2

Πίνακας 3.1: Αναλυτικά αποτελέσματα για τα ποσοστό επιτυχίας στις επιμέρους εικόνας της βάσης για κάθε μέθοδο.

εικόνα και το ποσοστό επιτυχίας σε αυτή. Το συνολικό ποσοστό επιτυχίας είναι 98.8% για την Bhattacharya και 98.5% για την απόσταση με μέγιστη πιθανοφάνεια. Επίσης και εδώ παρατηρείται η υπεροχή της Bhattacharya έναντι της ML. Ακόμη το ότι τα ποσοστά είναι πολύ υψηλά οφείλεται στον μικρό αριθμό κλάσεων 20, έναντι 55 στην VisTex. Και επειδή οι υφές στην Brodatz είναι πιο καλά δομημένες και πιο ομοιόμορφες κάνοντας πολύ πιο εύκολο το πείραμα. Το πρότυπο πείραμα της προηγούμενης ενότητας εφαρμόστηκε και σε αυτήν την βάση για την DWF ανάλυση και με απόσταση Bhattacharya και Gauss κατανομή χαρακτηριστικών δίνοντας ποσοστό επιτυχίας 92.0% στον ελάχιστο δυνατό αριθμό εικόνων που ανακλήθηκαν, επιβεβαιώνοντας έτσι την ευκολία του πειράματος. Περισσότερα πειράματα για την βάση με τις Brodatz έχουν δημοσιευθεί σε παλαιότερη εργασία [23].

	d1	d3	d4	d5	d6	d9	d10	d11	d12	d14	d15	d17	d18	d19	d37	d41	d51	d52	d72	d95	σύνολο	ποσοστό
d1	50	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	50	100.0
d3	0	50	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	50	100.0
d4	0	0	50	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	50	100.0
d5	0	0	0	50	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	50	100.0
d6	0	0	0	0	50	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	50	100.0
d9	0	0	0	0	0	50	0	0	0	0	0	0	0	0	0	0	0	0	0	0	50	100.0
d10	0	0	0	0	0	0	43	0	7	0	0	0	0	0	0	0	0	0	0	0	50	86.0
d11	0	0	0	0	0	0	0	50	0	0	0	0	0	0	0	0	0	0	0	0	50	100.0
d12	0	0	0	0	0	0	1	0	49	0	0	0	0	0	0	0	0	0	0	0	50	98.0
d14	0	0	0	0	0	0	0	0	0	50	0	0	0	0	0	0	0	0	0	0	50	100.0
d15	0	0	0	0	0	0	0	0	2	0	48	0	0	0	0	0	0	0	0	0	50	96.0
d17	0	0	0	0	0	0	0	0	0	0	0	50	0	0	0	0	0	0	0	0	50	100.0
d18	0	0	0	0	0	0	0	0	0	0	0	0	50	0	0	0	0	0	0	0	50	100.0
d19	0	0	0	0	0	0	0	0	0	0	0	0	0	50	0	0	0	0	0	0	50	100.0
d37	0	0	0	0	0	0	0	0	0	0	0	0	0	0	50	0	0	0	0	0	50	100.0
d41	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	49	0	0	0	0	50	98.0
d51	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	50	0	0	0	50	100.0
d52	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	50	0	0	50	100.0
d72	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	49	0	50	98.0
d95	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	50	50	100.0

Πίνακας 3.2: Αποτέσματα της DWF ανάλυσης με Bhattacharya απόσταση για το πείραμα με τις Brodatz εικόνες και 50 τυχαία εικονίδια ανά εικόνα.

	d1	d3	d4	d5	d6	d9	d10	d11	d12	d14	d15	d17	d18	d19	d37	d41	d51	d52	d72	d95	σύνολο	ποσοστό
d1	50	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	50	100.0
d3	0	50	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	50	100.0
d4	0	0	50	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	50	100.0
d5	0	0	0	50	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	50	100.0
d6	0	0	0	0	50	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	50	100.0
d9	0	0	0	0	0	50	0	0	0	0	0	0	0	0	0	0	0	0	0	0	50	100.0
d10	0	0	0	0	0	0	44	0	6	0	0	0	0	0	0	0	0	0	0	0	50	88.0
d11	0	0	0	0	0	0	0	50	0	0	0	0	0	0	0	0	0	0	0	0	50	100.0
d12	0	0	0	0	0	0	1	0	49	0	0	0	0	0	0	0	0	0	0	0	50	98.0
d14	0	1	0	0	0	0	0	0	0	49	0	0	0	0	0	0	0	0	0	0	50	98.0
d15	0	0	0	0	0	0	0	0	4	0	46	0	0	0	0	0	0	0	0	0	50	92.0
d17	0	0	0	0	0	0	0	0	0	0	0	50	0	0	0	0	0	0	0	0	50	100.0
d18	0	0	0	0	0	0	0	0	0	0	0	0	50	0	0	0	0	0	0	0	50	100.0
d19	0	0	0	0	0	0	0	0	0	0	0	0	0	50	0	0	0	0	0	0	50	100.0
d37	0	0	0	0	0	0	0	0	0	0	0	0	0	0	50	0	0	0	0	0	50	100.0
d41	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	49	0	0	0	0	50	98.0
d51	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	50	0	0	0	50	100.0
d52	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	50	0	0	50	100.0
d72	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	48	1	50	96.0
d95	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	50	50	100.0

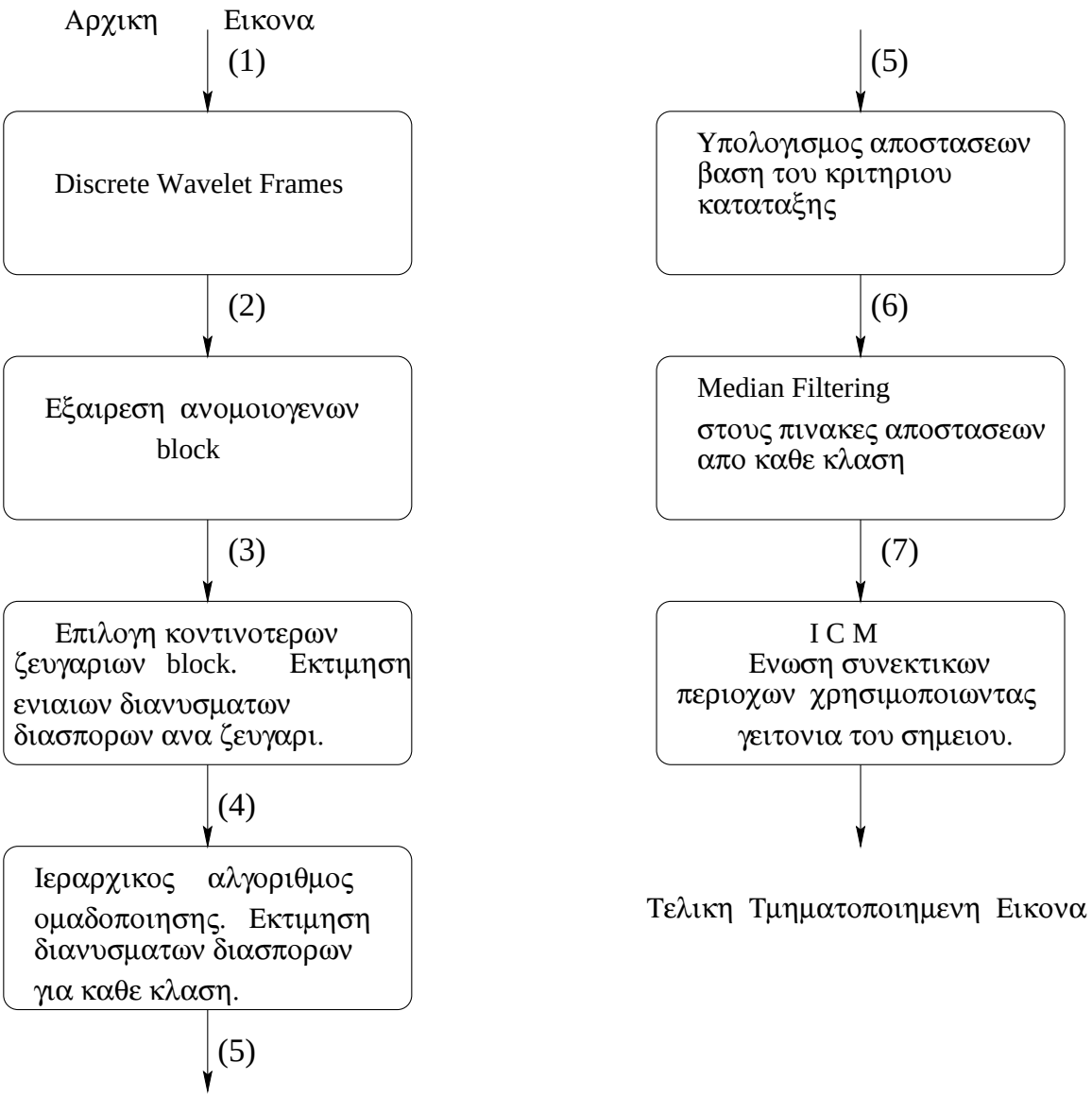
Πίνακας 3.3: Αποτέσματα της DWF ανάλυσης με ML απόσταση για το πείραμα με τις Brodatz εικόνες και 50 τυχαία εικονίδια ανά εικόνα.

Κεφάλαιο 4

Τμηματοποίηση υφής

Για την τμηματοποίηση εικόνων, που περιέχουν έναν αριθμό από διαφορετικές υφές χρησιμοποιείται η ανάλυση DWF μια που αυτή έδειξε να έχει αξιόλογα αποτελέσματα στην κατηγοριοποίηση. Ζητούμενο είναι να εντοπισθούν στην εικόνα τα τμήμα όπου βρίσκεται η κάθε υφή. Σε αυτό το κεφάλαιο παρουσιάζεται ένας αλγόριθμος τμηματοποίησης βασισμένος στην DWF ανάλυση. Στο παρελθόν έχουν προταθεί αλγόριθμοι τμηματοποίησης που βασίζονται σε νευρωνικά δίκτυα [24], και στατιστικές προσεγγίσεις μεγιστοποιώντας ένα κριτήριο πιθανότητας εκ των υστέρων (a posteriori) -προσέγγιση Bayes [25], σε ομαδοποίησεις (clustering) [16] [26], και τεχνικές μεταβλητών συνόρων (region growing, snakes) [27]. Στην παρούσα εργασία για την τμηματοποίηση χρησιμοποιούνται στατιστικές τεχνικές (όπως η χρήση του κριτηρίου της μέγιστης πιθανοφάνειας και τυχαίων πεδίων Markov για την κατατάξη σημείων και την χαλάρωση των ετικετών αντίστοιχα) σε συνδυασμό με τεχνικές ομαδοποίησης (όπως ιεραρχική ομαδοποίηση για τον υπολογισμό των αρχικών παραμέτρων) [28].

Το κεφάλαιο χωρίζεται σε τρεις ενότητες. Στην πρώτη περιγράφεται πως υπολογίζονται τα διανύσματα των διασπορών, που χαρακτηρίζουν κάθε κλάση-τμήμα στην εικόνα, αυτόματα χωρίς την βοήθεια του χρήστη. Στην δεύτερη υποενότητα περιγράφεται, βάση των διανυσμάτων αυτών, πως κάθε σημείο της εικόνας κατατάσσεται σε μία από τις κλάσεις και πως καταλήγει ο αλγόριθμος στην τελική εικόνα αποτέλεσμα. Η όλη διαδικασία που ακολουθείται παρουσιάζεται στο σχήμα (4.1). Στην τρίτη ενότητα παρουσιάζονται πειραματικά αποτελέσματα σε συνθετικές εικόνες υφών αλλά και σε φυσικές σκηνές.



Σχήμα 4.1: Όλα τα στάδια του αυτόματου αλγόριθμου τμηματοποίησης.

4.1 Αυτόματος υπολογισμός των διανυσμάτων διασπορών για κάθε κλάση

4.1.1 Αρχικές προσεγγίσεις

Στον αλγόριθμο που υλοποιήθηκε για την τμηματοποίηση θεωρείται γνωστό το πόσες κλάσεις-υφές περιέχονται στην υπό τμηματοποίηση εικόνα, δηλαδή δίνει ο χρήστης τον αριθμό αυτόν στην γραμμή εντολών όταν αρχίζει το πρόγραμμα. Οπότε το πρόβλημα της τμηματοποίησης έχει ως εξής: γνωρίζοντας πόσες διαφορετικές υφές υπάρχουν σε μια εικόνα να βρεθούν οι περιοχές όπου βρίσκεται η κάθε μια. Για να κατατάξουμε σημεία σε κλάσεις-υφές χρειάζεται να υπολογίσουμε αποστάσεις από τις κλάσεις, οι οποίες χρειάζονται για να υπολογιστούν τα διανύσματα διασποράς κάθε κλάσης (αρχικές παράμετροι). Άρα πρέπει να βρεθούν οι αντιπροσωπευτικές περιοχές κάθε κλάσης για τον υπολογισμό των διανυσμάτων διασποράς.

Σε πρώτη φάση η εικόνα χωρίζονται σε block μεγέθους 32×32 pixels, και σε κάθε block υπολογίζεται το διάνυσμα διασπορών. Δοκιμάστηκαν διάφοροι αλγόριθμοι για την εύρεση των block που αντιπροσωπεύουν κάθε κλάση έτσι ώστε από αυτά να υπολογιστούν τα αντιπροσωπευτικά διανύσματα διασπορών. Ένας από αυτούς βασίζεται στην ιδέα ότι τα αντιπροσωπευτικά block π.χ. για δύο κλάσεις θα είναι αυτά που τα διανύσματα διασπορών τους θα απέχουν περισσότερο. Αυτός αλγόριθμος χωλαίνει γιατί ένα από τα δύο block που θα επιλεγούν θα είναι block που θα βρίσκεται πάνω στο σύνορο των δύο υφών, δηλαδή θα περιέχει και τις δύο υφές. Αυτό συμβαίνει γιατί εμφανίζεται μεγάλη διασπορά στο διάνυσμα διασπορών του block αυτού στην κατεύθυνση κάθετη του συνόρου επειδή εκεί έχουμε διαφορετικό περιεχόμενο, πράγμα που το κάνει το block αυτό να είναι μακριά από όλα τα άλλα. Το ίδιο συμβαίνει και αν ακολουθηθεί η αντίστροφη πορεία δηλαδή εάν βρίσκουμε ενιαίο διάνυσμα διασπορών για block απέχουν λιγότερο ομαδοποιώντας αυτά τα block, και ολοκληρώνοντας την διαδικασία όταν έχουμε αριθμό ομάδων ίδιο με τον αριθμό των κλάσεων (αρχίζουμε από ομάδες όσα και τα block της εικόνας). Αυτός ο ιεραρχικός αλγόριθμος ομαδοποίησης [29] έχει ακριβώς το ίδιο πρόβλημα με τον προηγούμενο αλγόριθμο με τα block που περιέχουν πολλές υφές δηλαδή που βρίσκονται σε σύνορο.

4.1.2 Απόρριψη μη ομοιογενών block

Άρα με κατά κάποιο τρόπο πρέπει να εξαιρούνται τα μη ομοιογενή block. Άρα τίθεται θέμα ομοιογένειας των block και πως μπορεί να διαπιστωθεί εάν ένα block είναι ομοιογενές ή όχι. Για αυτό τον λόγο βρέθηκε ένα κριτήριο ομοιογένειας που περιγράφεται από την εξίσωση :

$$H_b = \frac{1}{\#pixels} \sum_{p \in block} \sum_{i=1}^{N-1} \frac{(y_{p,i}^2 - \sigma_{b,i}^2)^2}{\sigma_{b,i}^2} \tag{4.1}$$

όπου b το block που εξετάζεται, p ένα pixel που ανήκει στο block αυτό, και N ο αριθμός των συνιστωσών του διανύσματος διασποράς (η προσέγγιση παραλείπεται). Το κριτήριο αυτό βασίζεται στην ιδέα ότι εάν ένα block περιέχει δύο υφές η μέση διαφορά του $y_i^2(k,l)$ από την διασπορά σ_i^2 θα είναι μεγαλύτερη από ότι σε ένα block με μια υφή (αυτή η διαφορά έχει νόημα γιατί $\sigma^2 = E\{x^2\}$ αφού $E\{x\}=0$ για κάθε συνιστώσα της ανάλυσης). Η διαίρεση με το εκάστοτε σ_i^2 γίνεται για λόγους κανονικοποίησης, έτσι ώστε να έχει νόημα η σύγκριση των τιμών του κριτηρίου για διαφορετικά block εικόνας (κάθε block έχει διαφορετικό $\sigma_{b,i}^2$). Οπότε αν ένα block έχει πιο μεγάλη τιμή κριτηρίου σε σχέση με τα γειτονικά του το αυτό είναι ένα πιθανό ανομοιογενές block το οποίο περιέχει σύνορο. Δεν είναι όμως πάντα ανομοιογενές ένα τέτοιο block, γιατί κάλλιστα θα μπορούσε να είναι απλώς ένα τοπικό μέγιστο. Από πληθώρα πειραμάτων διαπιστώθηκε ότι ένα block είναι ανομοιογενές όταν τα δύο γειτονικά του (είτε στην κατακόρυφη είτε στην οριζόντια διεύθυνση) έχουν άθροισμα τιμών ομοιογένειας, βάση της (4.1), μικρότερο από την τιμή ομοιογένειας του block που ερευνούμε. Με αυτό το κριτήριο, αν προφανώς εφαρμοστεί για όλα τα block της εικόνας, αφαιρούμε σχεδόν όλα τα ανομοιογενή block. Αυτά τα ελάχιστα block με σύνορο, που μπορεί να μείνουν, δεν επηρεάζουν το αποτέλεσμα της ομαδοποίησης που θα εφαρμοστεί παρακάτω για τον υπολογισμό των διανυσμάτων διασπορών των κλάσεων, γιατί αυτά που έμειναν δεν έχουν τόσο μεγάλες διασπόρες, όσο αυτά που απορρίφθηκαν, λόγω του κριτηρίου ομοιογένειας.

4.1.3 Ομαδοποίηση (Clustering)

Αφού έχουμε διακρίνει τα μη ομοιογενή block, δουλεύουμε τώρα με τα υπόλοιπα. Επειδή πιθανόν να έχουμε block με σύνορο σε αυτά που έχουν απομείνει, όπως προαναφέρθηκε, βρίσκουμε τα K (= (αριθμός block) / 4) ζευγάρια πού είναι πιο κοντά μεταξύ τους. Με αυτόν τον τρόπο θα παρθούν μόνο αντιπροσωπευτικά ζευγάρια block για όλες τις κλάσης της εικόνας. Για να εκτιμηθεί πιο block είναι πιο κοντά σε ένα άλλο, ανάλογα με το διάνυσμα διασπορών του καθενός block, χρησιμοποιήθηκε η εξής

απόσταση μεταξύ δύο block:

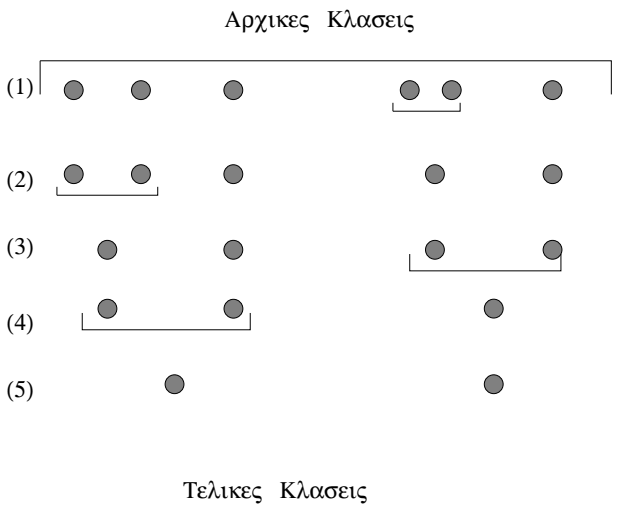
$$d_{b_1,b_2} = \sum_{i=1}^{N-1} \frac{|\sigma_{1,i}^2 - \sigma_{2,i}^2|}{\sigma_{1,i}^2 + \sigma_{2,i}^2} \tag{4.2}$$

όπου N ο αριθμός των συνιστωσών του διανύσματος των διασπορών,και όπου σ_1^2, σ_2^2 τα διανύσματα διασπορών των b_1, b_2 αντίστοιχα. Η διαίρεση στο άθροισμα αποσκοπεί σε κανονικοποίηση για να έχει νόημα να συγκριθούν οι εν μέρη αποστάσεις. Αφού εντοπισθούν τα ζευγάρια αυτά υπολογίζονται ενιαία διανύσματα διασπορών, δηλαδή υπολογίζοντας για καθένα από αυτά το διάνυσμα διασπορών για όλα τα σημεία των δύο block. Δεδομένων των K διανυσμάτων διασπορών εφαρμόζουμε τον ιεραρχικό αλγόριθμο ομαδοποίησης που προαναφέρθηκε στην προηγούμενη υποενότητα. Έχοντας αρχικά κλάσεις όσες και τα ζευγάρια των block (K το πλήθος), ενώνουμε σε κάθε επανάληψη του αλγορίθμου τα πλησιέστερα διανύσματα διασπορών, υπολογίζοντας από όλα τα σημεία, που αντιπροσωπεύει το κάθε διάνυσμα, το ενιαίο διάνυσμα διασπορών. Επομένως στην επόμενη επανάληψη κάθε φορά θα υπάρχει και μια κλάση λιγότερη. Όταν ο αριθμός των κλάσεων φτάσει να είναι ίσος με τον αριθμό των κλάσεων των διαφορετικών υφών που υπάρχουν στην εικόνα, τότε αλγόριθμος τερματίζει και έχουμε τα αντιπροσωπευτικά διανύσματα διασπορών για κάθε κλάση. Για την απόσταση μεταξύ των διανυσμάτων διασπορών δυο κλάσεων, για την ομαδοποίηση δεν χρησιμοποιείται η παραπάνω εξίσωση (4.2) αλλά η εξής:

$$d_{c_1,c_2} = \sqrt{\frac{nb_1 \times nb_2}{nb_1 + nb_2}} \sum_{i=1}^{N-1} \frac{|\sigma_{1,i}^2 - \sigma_{2,i}^2|}{\sigma_{1,i}^2 + \sigma_{2,i}^2} \tag{4.3}$$

όπου οι συμβολισμοί είναι οι ίδιοι με την (4.2), εκτός από τα nb_1, nb_2 που είναι ο αριθμός των block που αντιπροσωπεύουν οι κλάσεις c_1, c_2 αντίστοιχα. Ο παράγοντας που προστέθηκε στην απόσταση αποσκοπεί στο να μειώσει την απόσταση κλάσεων που έχουν μεγάλη διαφορά στον αριθμό των block. Δηλαδή το κριτήριο αυτό έχει την τάση να ενώνει μεγάλες ομάδες block με μικρές ομάδες block, και όχι να ενώνει κλάσεις που αντιπροσωπεύουν μεγάλο αριθμό block. Αυτό έχει σαν αποτέλεσμα να επιλέγονται οι πιο αντιπροσωπευτικές περιοχές για κάθε κλάση και να δίνονται καλύτερα αντιπροσωπευτικά διανύσματα. Δηλαδή είναι πιο πιθανό δυο πολυπληθής ομάδες block που έχουν ομαδοποίηθει μέχρι στιγμής να αντιπροσωπεύουν διαφορετικές κλάσεις παρά να ανήκουν τελικά στην ίδια. Η ιεαρχική ομαδοποίηση που περιγράφηκε παρουσιάζεται στο σχήμα (4.2).

Στο παραπάνω κριτήριο ομοιότητας block χρησιμοποιείται μόνο η διασπορά. Δηλαδή βρίσκει όμοια τα block που έχουν όμοιες διασπορές στις λεπτομέρειες ή και στην προσέγγιση. Αν όμως θέλουμε να κατατάξουμε διαφορετικά δύο block που περιέχουν



Σχήμα 4.2: Ομαδοποίηση: κάθε κύκλος παριστάνει μια τρέχουσα κλάση.Κάθε φορά ενώνονται οι πιο κοντινές. Αρχικά έχουμε 6 κλάσεις που προκύπτουν από χαρακτηριστικά 6 block-κλάσεων. Τελικά υπολογίζονται τα αντιπροσωπευτικά χαρακτηριστικά των δύο ζητούμενων κλάσεων

ίδιες υφές αλλά διαφορετικής φωτεινότητας, τότε σύμφωνα με τα παραπάνω θα τις θεωρεί όμοιες. Για αυτό τον λόγο κάλλιστα μπορεί να χρησιμοποιηθεί η μέση τιμή της προσέγγισης στο κριτήριο. Αυτό γίνεται με την χρήση της απόστασης Bhattacharya (5.11), καθώς αυτή συνδυάζει την μέση τιμή και και τις διασπόρες και σε ξεχωριστούς όρους. Επίσης όπως αποδείχθηκε και στο κεφάλαιο 3 έχει πολύ καλή επίδοση στην κατηγοριοποίηση. Στα πειραματικά αποτελέσματα παρουσιάζονται ανάλογα αποτελέσματα.

4.2 Κατάταξη σημείων στις κλάσεις. Ένωση συνεκτικών περιοχών

Εφαρμόζοντας τα παραπάνω εκτιμούμε τα αντιπροσωπευτικά διανύσματα διασπορών για κάθε διαφορετική υφή που βρίσκεται στην εικόνα. Έχοντας την πληροφορία αυτή μπορούμε να κατατάξουμε κάθε σημείο σε μια από τις κλάσεις-υφές της εικόνας. Αυτό γίνεται χρησιμοποιώντας το κριτήριο (3.2) αλλά μόνο για την απόσταση ενός σημείου από χαρακτηριστικό υφής (διάνυσμα διασπορών) και όχι για σύνολο σημείων. Για τον λόγο αυτό δεν χρησιμοποιήθηκε Bhattacharya, που είχε καλύτερη επίδοση, επειδή δεν μπορεί να υπολογίσει απόσταση χαρακτηριστικών ενός σημείου από τα χαρακτηριστικά μιας κλάσης. Οπότε έχουμε:

$$d_j(y) = \sum_{i=1}^{N-1} \left(\frac{y_i^2(k,l)}{\sigma_{i,j}^2} + \log \sigma_{i,j}^2 \right) + \frac{(y_N(k,l) - \mu_{Nj})^2}{\sigma_{N,j}^2} + \log \sigma_{N,j}^2 \qquad (4.4)$$

όπου $y_{k,l}$ το διάνυσμα των τιμών της ανάλυσης για το σημείο αυτό με συντεταγμένες k, l , N ο αριθμός των συνιστωσών της ανάλυσης, και σ_j^2 το διάνυσμα διασπορών της κλάσης. Οι δύο τελευταίοι όροι χρησιμοποιούνται στην περίπτωση που γίνεται χρήση των χαρακτηριστικών της προσέγγισης, με μ_{Nj} να είναι η μέση τιμή της προσέγγισης (για i=N). Άρα σύμφωνα με την απόσταση αυτή ένα σημείο θα ανήκει στην κλάση αυτή όπου η (4.4) δίνει την μικρότερη τιμή. Άρα υπολογίζουμε τις τιμές της απόστασης για κάθε σημείο από κάθε κλάση.

Επειδή το κριτήριο που εφαρμόζεται είναι στατιστικό και σημειακό θα έχουμε αρκετές λάθος κατατάξεις σημείων διασκορπισμένες στην εικόνα. Οπότε για την βελτίωση των κατατάξεων χρησιμοποιείται ένα φίλτρο ενδιάμεσης τιμής, το οποίο εφαρμόζεται σε κάθε πίνακα αποστάσεων προς μια συγκεκριμένη κλάση. Το φίλτρο εφαρμόζεται σε ένα παράθυρο 15×15 pixels. Αυτό το φιλτράρισμα έχει σαν αποτέλεσμα σε γειτονικά σημεία οι αποστάσεις προς μια κλάση να έχουν την ίδια τιμή έτσι εξαφανίζονται τα στατιστικά σφάλματα. Επίσης ένας λόγος που επιλέχθηκε το φίλτρο ενδιάμεσης τιμής είναι να μην επηρεασθούν τα σύνορα των κλάσεων. Όσον αφορά την υλοποίηση του ενδιάμεσου φίλτρου χρησιμοποιήθηκε ένας γρήγορος αλγόριθμος πολυπλοκότητας O(N). Το σημείο αυτό του αλγορίθμου είναι που καθυστερεί τον αλγόριθμο περισσότερο και για αυτόν τον λόγο έγινε η βελτιστοποίηση.

Ακόμη όμως απομένουν λάθος κατατάξεις στα σημεία της εικόνας. Αυτό συμβαίνει λόγο των ανομοιομορφιών που παρουσιάζονται σε πολλές υφές σε ορισμένες μικρές περιοχές τους, το οποίο και δημιουργεί μικρές οπές στην εικόνα κατάταξης των σημείων. Αυτό διορθώνεται με την εφαρμογή του επαναληπτικού αλγορίθμου ICM για την συνένωση συνεκτικών περιοχών [30]. Αυτός ο αλγόριθμος για την κατάταξη κάποιου σημείου χρησιμοποιεί τις απόστασεις από τις κλάσεις, μετά από το φιλτράρισμα, αλλά συνάμα λαμβάνοντας υπόψην το που έχουν καταταχθεί τα γειτονικά σημεία.Ο ICM αλγόριθμος χρησιμοποιεί μια παράμετρο, που δίνεται από τον χρήστη στην γραμμή εντολών, η οποία περιγράφει κατά πόσο θα ληφθεί υπόψην για την κατάταξη ενός σημείου η γειτονία του. Αυτό περιγράφεται από την εξής εξίσωση :

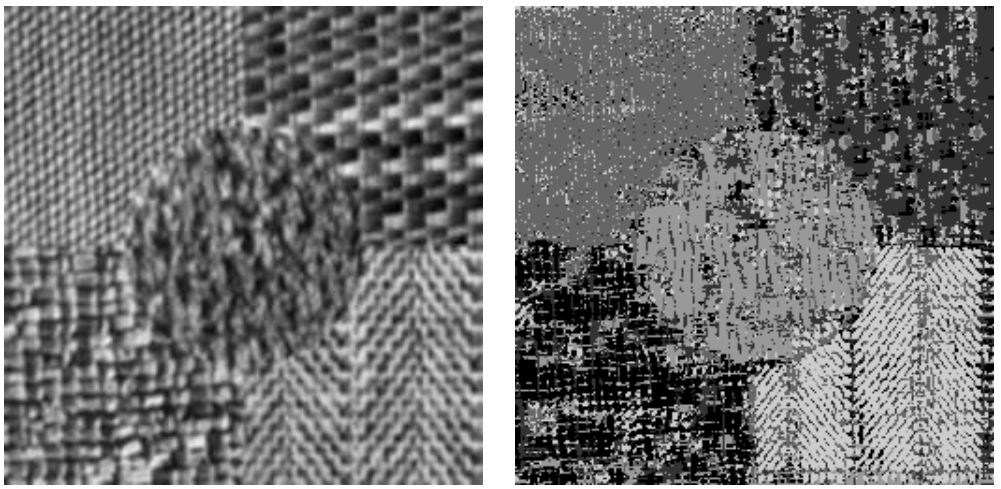
$$icm_dist_j(y) = d_j(y) + \lambda * (8 - 2 * M_j) \tag{4.5}$$

όπου λ η παράμετρος του ICM , 8 είναι ο αριθμός των γειτονικών σημείων, M_j είναι ο αριθμός των γειτονικών σημείων που έχουν καταταχθεί στην κλάση j, $d_j(y)$ είναι η απόσταση από την (4.4) μετά την εφαρμογή του φίλτρου, $icm_dist_j(y)$ είναι η απόσταση που χρησιμοποιεί ο ICM για την κατατάξη ενός σημείου που περιγράφεται από το διάνυσμα y της ανάλυσης. Οπότε επαναλήπτικά ο ICM ελαχιστοποιεί την παραπάνω

απόσταση για όλα τα σημεία της εικόνας. Αυτό έχει σαν αποτέλεσμα οι οπές και τυχαία μικρολάθη να καλυφθούν και να δημιουργηθεί η τελική εικόνα κατατάξεων, όπου έχουν τμηματοποιηθεί οι διαφορετικές περιοχές της αρχικής εικόνας με βάση την υφή. Περισσότερα για το πως μοντελοποιείται το πρόβλημα ανάθεσης ετικετών παρουσιάζονται στο παράρτημα Α.

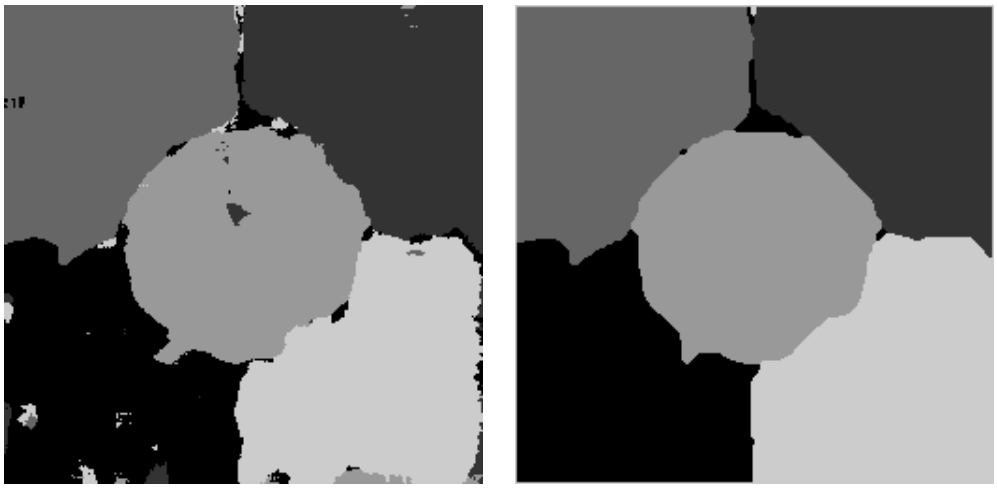
4.3 Πειραματικά αποτελέσματα

Ο αλγόριθμος της τμηματοποίησης εφαρμόστηκε σε εικόνες συνθετικές αλλά και σε εικόνες από φυσικές σκηνές. Στο σχήμα 4.3 φαίνεται η αρχική συνθετική εικόνα που περιέχει 5 διαφορετικές υφές από το Brodatz album (D77, D55, D84, D17 και D24), καθώς και δεξιά της οι κατατάξεις με βάση την απόσταση 4.4. Στην αριστερή εικόνα του 4.4 φαίνονται οι κατατάξεις μετά το φιλτράρισμα με το ενδιάμεσο φίλτρο, και στην δεξιά μετά την εφαρμογή του ICM. Επίσης στο 4.5 παρουσιάζονται οι εικόνες που δείχνουν τα block που χαρακτηριστικών ανομοιογενή και τα block που τελικά επιλέχθηκαν σύμφωνα με τον αλγόριθμο αρχικοποίησης για να υπολογιστούν οι αρχικές παράμετροι της κάθε κλάσης.

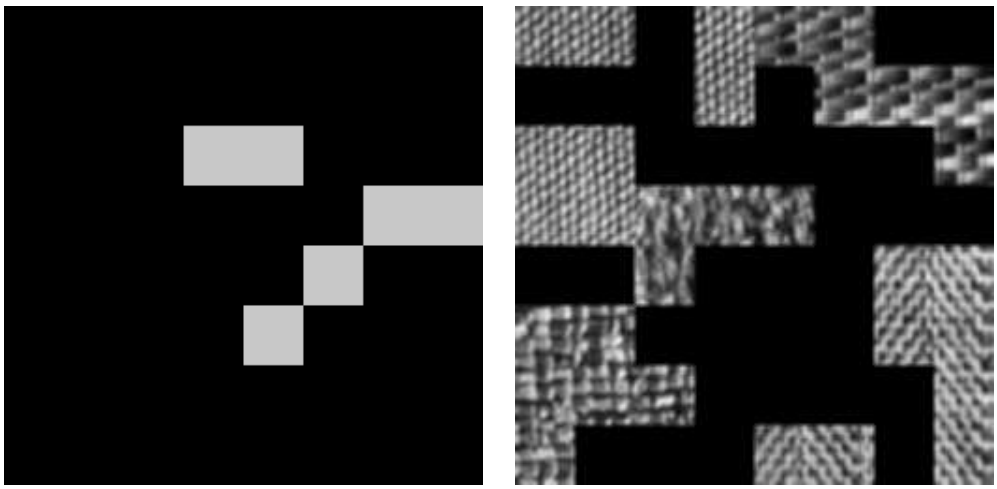


Σχήμα 4.3: Η αρχική συνθετική εικόνα με 5 υφές και οι αρχικές κατατάξεις με βάση την απόσταση

Τα block που εξαιρούνται από το κριτήριο ανομοιογένειας όντως είναι αυτά στα σύνορα που έχουν την μεγαλύτερη αντίθεση μεταξύ των διαφορετικών υφών όπως στα σύνορα της D55 με την D17 και την D23, στα σύνορα των D23, D77 και D23, D17. Επίσης τα block που χρησιμοποιούνται για τον υπολογισμό των παραμέτρων κάθε κλάσης είναι αντιπροσωπευτικά για κάθε μια από αυτήν όπως φαίνεται στην δεξιά εικόνα του

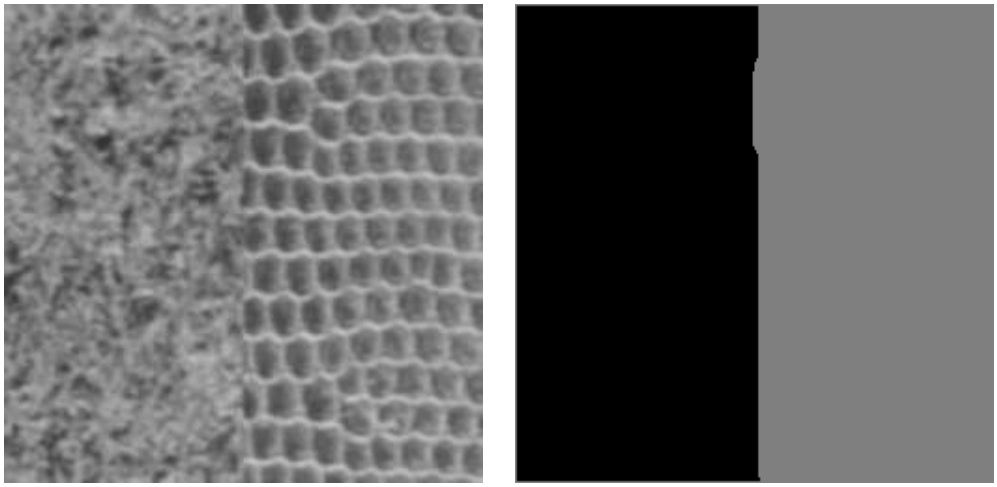


Σχήμα 4.4: Οι κατατάξεις μετά το φίλτρο ενδιάμεσης τιμής και η τελική εικόνα μετά τον ICM, αντίστοιχα για την συνθετική εικόνα με τις πέντε διαφορετικές υφές.

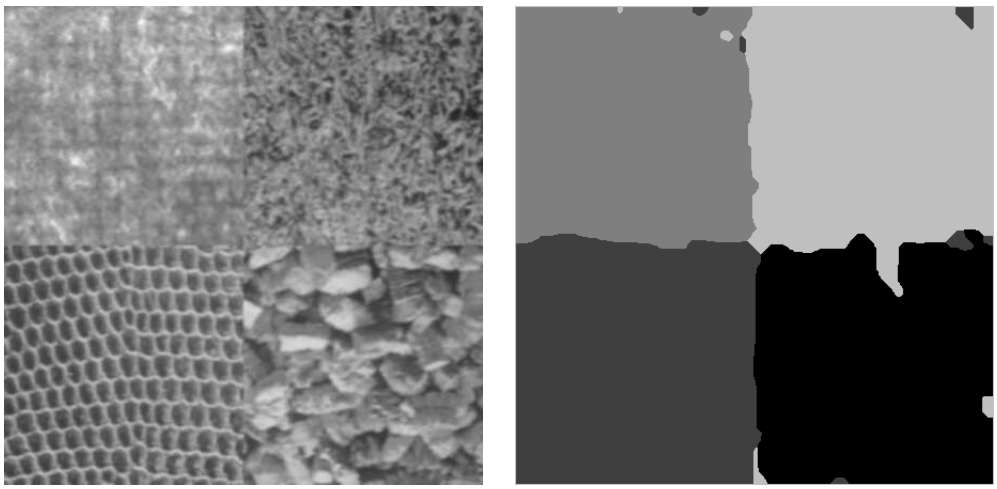


Σχήμα 4.5: Τα μη ομοιογενή block και τα block που χρησιμοποιήθηκαν για την εξαγωγή των χαρακτηριστικών κάθε κλάσης, αντίστοιχα για την συνθετική εικόνα με τις πέντε διαφορετικές υφές.

σχήματος 4.5. Η ανάλυση έγινε σε 3 επίπεδα δημιουργώντας 9 τιμές για το διάνυσμα χαρακτηριστικών. Και αυτό γιατί δεν χρησιμοποιήθηκε η προσέγγιση, όπως και στις υπόλοιπες εικόνες με συνθετικές υφές. Στα σχήματα 4.6, 4.7 φαίνονται οι αρχικές εικόνες και τελικές τμηματοποιημένες για συνθετικές εικόνες με 2 και 4 διαφορετικές υφές αντίστοιχα. Οι αναλύσεις έγιναν σε 4 επίπεδα, και στις δύο εικόνες.

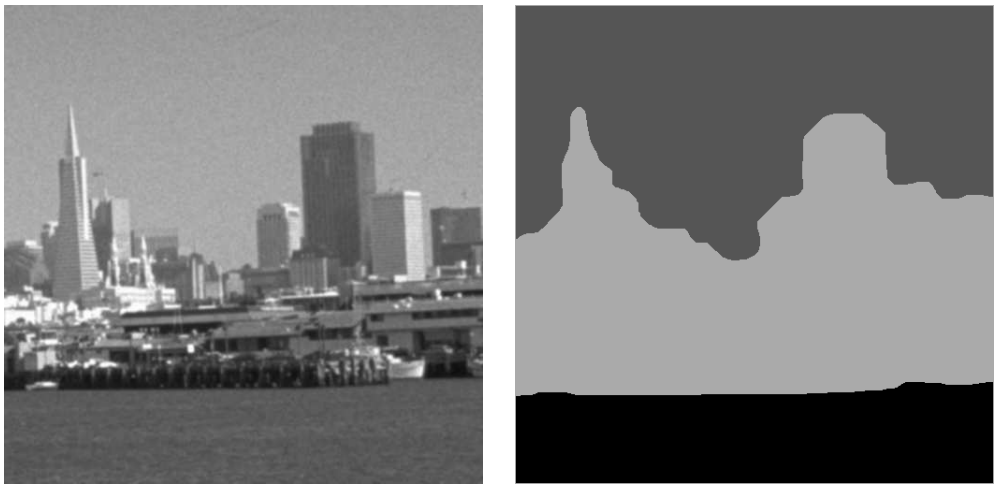


Σχήμα 4.6: Η αρχική εικόνα και η τελική τμηματοποιημένη, αντίστοιχα για την συνθετική εικόνα με δύο διαφορετικές υφές (D9 D3).

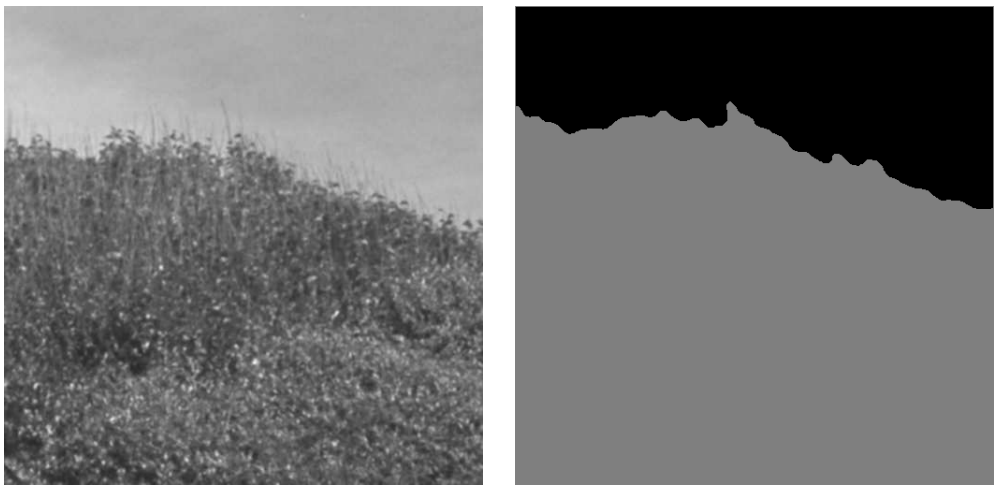


Σχήμα 4.7: Η αρχική εικόνα και η τελική τμηματοποιημένη, αντίστοιχα για την συνθετική εικόνα με 4 διαφορετικές υφές (D19 D9 D3 D5).

Ο αλγόριθμος εφαρμόστηκε και σε εικόνες με φυσικές σκηνές. Στα σχήματα 4.8, 4.9 φαίνονται τα αποτελέσματα για δύο φυσικές σκηνές. Η πρώτη περιέχει την πόλη του San Francisco μαζί με θάλασσα και ουρανό. Ενώ η δεύτερη περιέχει μία πλαγιά με χόρτα, και ουρανό. Τα επίπεδα ανάλυσης είναι 3 και για τις δύο εικόνες.



Σχήμα 4.8: Η αρχική εικόνα και η τελική τμηματοποιημένη, αντίστοιχα για την συνθετική εικόνα με τρεις διαφορετικές υφές.



Σχήμα 4.9: Η αρχική εικόνα και η τελική τμηματοποιημένη, αντίστοιχα για την συνθετική εικόνα με δύο διαφορετικές υφές.

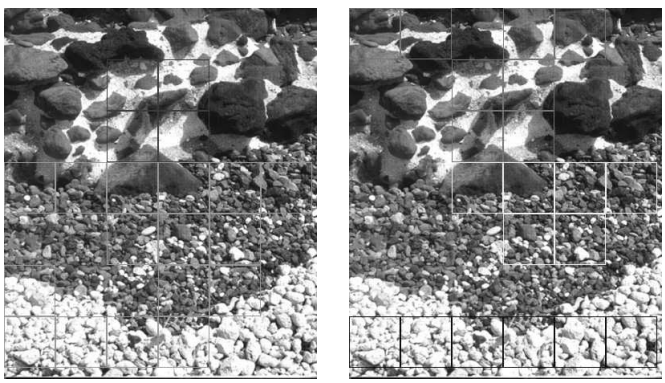
Στις εικόνες με τις φυσικές σκηνές καλύτερα αποτελέσματα εξάχθηκαν κάνοντας χρήση της συνιστώσας της προσέγγισης. Στην ουσία χρησιμοποιούμε την μέση τιμή και την διασπορά της προσέγγισης στο κριτήριο 4.4. Η μέση τιμή της προσέγγισσης είναι ίδια με την μέση τιμή της αρχικής εικόνας ($H(z)\,|_{z=1}=1$). Με την χρήση της μέσης τιμής βελτιώνονται τα αποτελέσματα στην τμηματοποίηση στις εικόνες με φυσικές σκηνές γιατί παίζει μεγάλο ρόλο η φωτεινότητα (πχ. διαχωρισμός ουρανού θάλλασας). Σε αντίθεση με τις συνθετικές που χρησιμοποιήθηκαν που είναι ομοιογενείς όσον αφορά την περιοδικότητα τους και την δομή τους, αλλά μερικές διαφέρουν εσωτερικά από περιοχή σε περιοχή στην φωτεινότητα πράγμα που κάνει την προσθήκη της προσέγγισης στον χαρακτηρισμό μη βελτιωτική για τις εικόνες αυτές. Επίσης η προσέγγιση δεν βοηθάει πάντα καθώς πρέπει βάσει του κριτηρίου να ακολουθεί κατανομή Gauss, που δεν συμβαίνει πάντα.

Επίσης ο αλγόριθμος τμηματοποίησης εφαρμόστηκε και σε άλλη μια εικόνα που περιέχει φυσική σκηνή. Η εικόνα αυτή είναι μία ειδική περίπτωση καθώς περιέχει βότσαλα από κάποια παραλία. Υπάρχουν τρεις ζώνες από βότσαλα η πρώτη ζώνη είναι πέτρες μεγάλου μέγεθους, στην δεύτερη ζώνη ψιλό βότσαλο σκούρου χρώματος ενώ η τρίτη περιέχει και αυτή ψιλό βοτσαλάκι, αλλά ανοιχτής απόχρωσης σχεδόν άσπρο. Στην περίπτωση που θέλουμε να διαχωρίσουμε τις τρεις ζώνες το κριτήριο 4.3 για την ομαδοποίηση block κατά την αρχικοποίηση δεν είναι αρκετό, γιατί δεν κάνει χρήση των μεσέων τιμών κάθε block στην συνιστώσα της προσέγγισης. Το κριτήριο αυτό είναι αναγκαίο όταν θα θέλαμε να τμηματοποιήσουμε την εικόνα σε δύο ζώνες, ενώνοντας σαν μία περιοχή τα ψιλά βότσαλα, επειδή συγκρίνει διασπορές και όχι μέσες τιμές. Στην αντίθετη περίπτωση όμως είναι αναγκαία η χρήση της απόστασης Bhattacharya. Στο σχήμα 4.10 φαίνονται η αρχική εικόνα καθώς και οι τελικές τμηματοποιημένες με δύο και τρία τμήματα (κάνοντας χρήση της 4.3 και της Bhattacharya αντίστοιχα).

Στο σχήμα 4.11 φαίνονται τα block που χρησιμοποιούνται για την αρχικοποίηση σε κάθε περίπτωση. Παρατηρούμε πως με την Bhattacharya το άσπρο και το μαύρο λεπτό βότσαλο ενώνονται ενώ στην αντίθετη περίπτωση αποτελούν ξεχωριστές κλάσεις (η φωτεινότητα στα όρια των block στο σχήμα είναι διαφορετική). Επίσης το μέγεθος του block στο παράδειγμα αυτό είναι 64×64. Χρησιμοποιείται πιο μεγάλο block λόγω της μεγάλης ανομοιογένειας που παρατηρείται στην πάνω ζώνη.

Τελειώνοντας το κεφάλαιο αυτό, διαπιστώθηκε ότι ο αλγόριθμος που αναπτύχθηκε έχει καλή επίδοση στην τμηματοποίηση εικόνων με βάση την υφή. Ωστόσο έχει το μειονέκτημα ότι δεν είναι εντελώς αυτόνομος καθώς χρειάζεται των αριθμό των κλάσεων στην

Σχήμα 4.10: Η αρχική εικόνα και οι τελικές τμηματοποιημένες, με 2 και 3 κλάσεις αντίστοιχα για την εικόνα με τα βότσαλα.



Σχήμα 4.11: Τα block που χρησιμοποιούνται για την εξαγωγή των παραμέτρων κάθε κλάσης για κάθε περίπτωση (2 και 3 κλάσεις).

εικόνα. Η εύρεση του αριθμού των διαφορετικών υφών στο παρελθόν έχει μελετηθεί από τον A. Jain [16]. Επίσης χρειάζεται και την παράμετρο λ για τον αλγόριθμο χαλάρωσης ετικετών ο οποίος δίνεται από τον χρήστη αλλά εκτιμάται εύκολα με τον πειραματισμό. Μια ενδεικτική τιμή είναι το 1.0. Ένα μειονέκτημα του είναι ότι η χρήση block δεν το κάνει προσαρμόσιμο σε όλες τις εικόνες, γιατί θα υπάρχουν εικόνες που θα χρειάζονται διαφορετικό μέγεθος block ανάλογα με το ”coarseness” της υφής που περιέχεται στην εικόνα και με το χώρο που καταλαμβάνει σε αυτήν.

Στο τελευταίο παράδειγμα φαίνεται ότι παρά τις καλές αρχικοποιήσεις τα αποτελέσματα παρουσιάζουν αρκετές λάθος κατατάξεις (οπές), οι οποίες δεν εξουδετερώνονται από τον ICM ούτε από την χρήση ενδιάμεσου φίλτρου στις τιμές της απόστασης. Αυτό οφείλεται στην αδυναμία του στατιστικού κριτηρίου να κατατάξει σημεία όταν έχουμε ανομοιογενείς υφές. Πράγματι η μεγάλη ανομοιογένεια που υπάρχει στην πάνω κλάση δεν μας αφήνει να μιλάμε για υφή στην περιοχή αυτή, παρόλα αυτά το κριτήριο δείχνει να τα καταφέρνει εν μέρη. Για την βελτίωση των αποτελεσμάτων είναι απαραίτητη η χρήση πιο ανθεκτικών (robust) μεθόδων όπως level set αλγορίθμους ή αλγορίθμους επέκτασης περιοχών (region growing). Περιπτώσεις αλγορίθμων αυτών είναι ο αλγόριθμος γρήγορου βηματισμού [31] και ο SRG [32], όπου ξεκινούν από τις πιο σίγουρες κατατάξεις με βάση την εκτίμηση της απόστασης και επεκτείνονται.

Κεφάλαιο 5

Χρωματικά χαρακτηριστικά

5.1 Γενικά

Το χρώμα παίζει σημαντικό ρόλο στην ζωή του ανθρώπου. Κατά' αρχήν είναι ένα από τα σημαντικότερα οπτικά ερεθίσματα μαζί με την κίνηση, και το βάθος. Διαφορετικά χρώματα έχουν διαφορετική επίδραση στον άνθρωπο, όπως το κόκκινο δηλώνει απαγόρευση, ή το κίτρινο προσοχή, κίνδυνο. Επίσης το κόκκινο είναι πιο συναρπαστικό χρώμα, το πράσινο ξεκουράζει, ενώ το μπλε το πιο ευχάριστο. Από τα βάθη των αιώνων ο άνθρωπος χρησιμοποιούσε το χρώμα στην καθημερινή του ζωή για να κατασκευάσει ενδύματα να δηλώσει συναισθήματα και άλλα. Φυσικά υλικά από τα οποία μπορούσαν να χρωματίζουν οτιδήποτε άλλο υλικό σπάνιζαν στα παλαιότερα χρόνια (χρωστικές ουσίες, pigments). Τέτοια υλικά άξιζαν το βάρος τους σε χρυσάφι εκείνες τις εποχές. Ολόκληρη η οικονομία πόλεων στηριζόταν σε τέτοια υλικά, όπως η πορφύρα στα χρόνια του Βυζαντίου. Μόλις τον 19ο αιώνα επιτεύχθηκε η παραγωγή απόχρωσης με συνθετικά μέσα, οπότε και άλλαξε η όψη του κόσμου. Από τότε και έπειτα άνθισε βιοτεχνία έτοιμων ενδυμάτων σε πολλές νέες αποχρώσεις, η έγχρωμη φωτογραφία, η έγχρωμη τηλεόραση, ψυχολόγοι μελετούν την επίδραση διάφορων χρωμάτων στον άνθρωπο, και η αγορά έχει γεμίζει από προϊόντα που έχουν ότι αποχρώσεις επιθυμεί ο πελάτης. Σήμερα σειρά έχουν πάρει οι υπολογιστές αναπαριστώντας τα εκατομμύρια χρώματα της φύσης.

Αιώνες τώρα ο άνθρωπος προσπαθούσε να εξηγήσει την ύπαρξη του χρώματος και να το μελετήσει. Ως τον 17ο αιώνα ακόμα διατηρούνταν η θεωρία του Αριστοτέλη που έλεγε ότι όλα τα χρώματα παράγονταν με την μίξη του μαύρου και του άσπρου. Έπειτα ο Leonardo da Vinci προσπάθησε να μετρήσει ποια είναι τα βασικά χρώματα αλλά δεν

μπορούσε να αποφασίσει αν αυτά είναι έξι ή οκτώ. Αυτά έως ότου ο Isaac Newton, το 1672 ανακαλύψει ότι το λευκό φως συντειθόταν από όλα τα ορατά χρώματα του φάσματος [33]. Πράγματι το χρώμα δεν υπάρχει στον φυσικό κόσμο, παρά μόνο στο μάτι και στο μυαλό του ανθρώπου, στην φύση υπάρχει η ηλεκτρομαγνητική ακτινοβολία. Έτσι χρώμα είναι *το αντιληπτό αποτέλεσμα του φωτός, στην ορατή περιοχή του φάσματος, που έχει μήκος κύματος από 400 nm έως 700 nm, πέφτοντας αμφιβλιστροειδή του ματιού.* Ο ανθρώπινος αμφιβλιστροειδής έχει τρεις τύπους από χρωματικούς φωτολήπτες που ονομάζονται κωνικά κύτταρα (cones), τα οποία ανταποκρίνονται στην ακτινοβολία με διαφορετικές φασματικές καμπύλες απόκρισης. Για αυτό τον λόγο τρείς τιμές είναι αρκετές για να περιγράψουν ένα χρώμα, παρέχοντας τις αντίστοιχες φασματικές συναρτήσεις που χρησιμοποιούνται (τριχρωματική όραση). Με το αντικείμενο αυτό ασχολείται η επιστήμη της χρωματομετρίας [33] [34] [35] [36].

5.2 Χρωματικά συστήματα

Ένα χρωματικό σύστημα (**colour system** ή **colour space**) *είναι μια μέθοδος με την οποία μπορούμε να καθορίσουμε, να δημιουργήσουμε και να οπτικοποιήσουμε χρώμα.* Σαν άνθρωποι μπορούμε να να ορίσουμε ένα χρώμα από τις ιδιότητες του όπως φωτεινότητα (brightness), απόχρωση (hue), και κορεσμός (saturation). Ένας υπολογιστής ορίζει ένα χρώμα με βάση τις διεγέρσεις των κόκκινων πράσινων και μπλε φωσφόρων στον καθοδικό σωλήνα της οθόνης. Ένας εκτυπωτής ορίζει ένα χρώμα με βάση την αντανάκλαση και την απορρόφηση του κυανού, της ματζέντας, του κίτρινου και του μαύρου μελανιού στο χαρτί. Ένα χρώμα αναπαρίσταται με τρεις συνιστώσες, οι οποίες καθορίζουν την θέση του στο συγκεκριμένο χρωματικό σύστημα. Οπότε οι συνιστώσες δεν δείχνουν πως είναι το χρώμα οπτικά αλλά που βρίσκεται στο χρωματικό σύστημα. Υπάρχουν πάρα πολλά χρωματικά συστήματα. Ένα λόγος είναι ότι χρειάζονται διαφορετικά συστήματα για διαφορετικές εφαρμογές, γιατί μερικά συστήματα λειτουργούν καλύτερα σε κάποιες εφαρμογές από κάποια άλλα είτε ένα χρωματικό σύστημα φτιάχτηκε για τη συγκεκριμένη εφαρμογή. Σύστημα όπως το RGB (Red Green Blue) και αλλά συστήματα έχουν φτιαχτεί για γραφικά υπολογιστών. Άλλα είναι φτιαγμένα για την εύκολη δημιουργία χρωμάτων αλλά και χρήση από τον άνθρωπο, όπως το LHS (Lightness Hue, Saturation). Ακόμα συστήματα έχουν κατασκευαστεί για να είναι ομοιόμορφα αντιληπτά κατά τον άνθρωπο (perceptual uniformity), όπως το Lab [37].

Τα χρωματικά σύστημα μπορούν να διαχωριστούν σε εξαρτημένα από την συ-

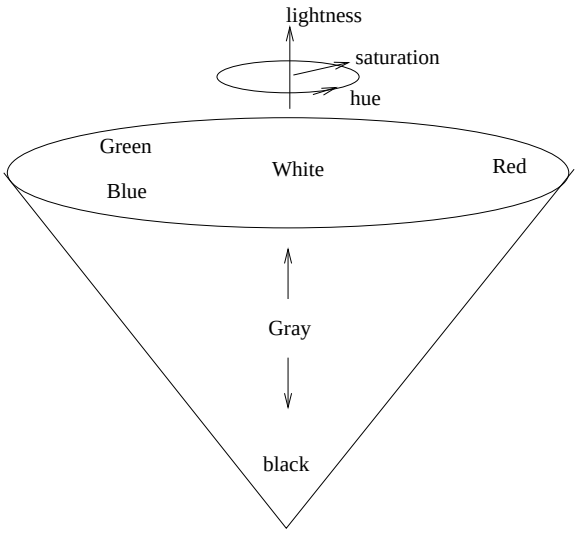
σκευή και μη, σε προσθετικά (additive) και αφαιρετικά (subtractive), και σε ομοιόμορφα αντιληπτά από τον άνθρωπο και μη. Εξαρτημένα από την συσκευή (device dependent) είναι τα συστήματα στα οποία τα χρώματα που παράγονται εξαρτώνται από την ρύθμιση της συσκευής με την οποία γίνεται η οπτικοποίηση, όσον αφορά την αντίληψη από τον άνθρωπο. Σε αντίθεση μη εξαρτόμενα από την συσκευή (device independent) είναι τα συστήματα όπου το χρώμα φαίνεται το ίδιο όπου και αν αυτό παρασταθεί. Το RGB παράδειγμα είναι device dependent, ενώ το Lab δεν είναι. Προσθετικά είναι τα συστήματα στα οποία διαφορετικές ακτίνες φωτός αναμιγνύονται και δίνουν ένα χρώμα. Π.χ ένα αντικείμενο αντανακλά ακτίνες όπου αυτές συνδυάζονται και δίνουν την αίσθηση του χρώματος του αντικειμένου στον άνθρωπο. Τα βασικά χρώματα (primary colours) είναι το κόκκινο το πράσινο και το μπλε. Αφαιρετικά συστήματα είναι αυτά στα οποία το πως θα φαίνεται ένα σημείο χρωματικά εξαρτάται από τις αποχρώσεις που απορροφά αυτό από το φάσμα. Τα βασικά χρώματα σε αφαιρετικό σύστημα είναι το κυανό η ματζέντα και το κίτρινο (CMY cyan magenta, yellow). Το RGB με το CMY είναι συμπληρωματικά μεταξύ τους. Με αφαιρετικά συστήματα δουλεύουν κυρίως οι φωτογράφοι και οι καλλιτέχνες γιατί χρησιμοποιούν χρωστικές που απορροφούν αποχρώσεις για να δημιουργήσουν χρώματα στην παλέτα τους, π.χ η μπλε βαφή είναι μπλε γιατί όταν το φως ”χτυπήσει” στην επιφάνεια της το pigment απορροφά (ή αφαιρεί) όλο το χρωματικό φάσμα εκτός του μπλε το οποίο αντανακλάται στο μάτι. Ομοιόμορφο αντιληπτό από τον άνθρωπο είναι ένα σύστημα, εάν μια μικρή αλλαγή στην τιμή μιας χρωματικής συνιστώσας, είναι προσεγγιστικά η ίδια στην αντίληψη του χρώματος από τον άνθρωπο.

Σημαντικά χρωματικά συστήματα τα οποία χρησιμοποιούνται και στην παρούσα εργασία είναι:

- RGB το οποίο είναι προσθετικό σύστημα βασισμένο εκ κατασκευής στην τριχρωματική θεωρία. Είναι εύκολο στην υλοποίηση μη γραμμικό (not perceptually uniform), εξαρτημένο από την συσκευή αφού έχει σχεδιαστεί για τις απαιτήσεις των συσκευών (πολύ κοντά στο hardware). Δύσκολο στην χρήση από τον άνθρωπο. Χρησιμοποιείται στις τηλεοπτικές κάμερες, γραφικά υπολογιστών, στον καθοδικό σωλήνα οθονών, και σε άλλες συσκευές.

- LHS (Lightness Hue Saturation) το οποίο ανήκει στην οικογένεια των συστημάτων που διαχωρίζουν την φωτεινότητα από την απόχρωση και τον κορεσμό [38]. Διαφέρουν μόνο στο πώς υπολογίζουν την φωτεινότητα. Μερικά από αυτά είναι HCI (intensity), HSV (value), HVC TSD (hue saturation darkness) και άλλα. Τα συστήματα αυτά είναι γραμμικοί μετασχηματισμοί του RGB οπότε και αυτά είναι εξαρτημένα

από την συσκευή. Είναι μη γραμμικά και παρουσιάζουν χρωματικές ασυνέχειες. Ένα βασικό πλεονέκτημα που έχει η οικογένεια των συστημάτων αυτών είναι ο διαχωρισμός της φωτεινότητας, ο οποίος επιτρέπει την χωριστή επεξεργασία φωτεινότητας και των χρωματικών συνιστωσών χρήσιμο σε πολλές εφαρμογές. Επίσης είναι εκ κατασκευής διαισθητικό για τον άνθρωπο γιατί με αυτά ακριβώς τα τρία γνωρίσματα ορίζει ο άνθρωπος τα χρώματα. Για να εφαρμόσουμε απευθείας



Σχήμα 5.1: Αναπαράσταση του κώνου των LHS χρωματικών συστημάτων.

ευκλείδια απόσταση μεταξύ σημείων δεν αρκούν οι τιμές του συστήματος όπως είναι, καθώς στο LHS έχουμε στην ουσία πολικές συντεταγμένες. Για αυτόν τον λόγο εφαρμόζεται ένας μη γραμμικός μετασχηματισμός, που δεν είναι τίποτα άλλο από την μετατροπή σε καρτεσιανές συντεταγμένες, με αρχή των αξόνων στην βάση του κώνου σχήμα 5.1, [26].

$$< l, h, s > \longrightarrow < l, l\ s\ cos(2\pi h)\ , l\ s\ sin(2\pi h) > \tag{5.1}$$

όπου τα l φωτεινότητα, h απόχρωση, s κορεσμός είναι κανονικοποιημένα στο διάστημα [0,1]. Τα χαμηλά σημεία στον κώνο είναι κοντύτερα μεταξύ τους. Αυτό συμβαδίζει με την ανθρώπινη αντίληψη καθώς όσο πιο λιγότερο φως τόσο πιο δύσκολα μπορεί ο άνθρωπος να διαχωρίσει χρώματα, τείνοντας έτσι προς την ομοιομορφία στην αντίληψη. Στην παρούσα εργασία επιλέχθηκε το HSV σύστημα από την οικογένεια LHS συστημάτων.

- Lab: Το 1931 CIE, (Commission Internationale de l'Eclairage) [39], παίζει σημαντικό ρόλο στον ορισμό των χρωματικών συστημάτων. Στην προσπάθεια να κατασκευά-

σει ένα χρωματικό σύστημα το οποίο να είναι perceptual uniform, δημιουργεί τον XYZ χρωματικό σύστημα, το οποίο καλείται και ”κανονικό”. Το σύστημα αυτό αναπαρίσταται συχνά με ένα διδιάστατο χρωματικό διάγραμμα που παριστάνει τις θέσεις όλων των αποχρώσεων, και έχει το σχήμα ενός ιστιοπλοϊκού πανιού. Ωστόσο το χρωματικό πρότυπο βασίζεται σε φανταστικά βασικά χρώματα που δεν υπάρχουν πραγματικά. Είναι καθαρά θεωρητικά και αυτό κάνει το σύστημα ανεξάρτητο συσκευής. Η κατασκευή του βασίστηκε στις καμπύλες απόκρισης των χρωματικών φωτοληπτών του ανθρώπινου ματιού. Επειδή οι αποκρίσεις αλλάζουν από άνθρωπο σε άνθρωπο η CIE καθόρισε ένας πρότυπο μέσο παρατηρητή κατόπιν πολλών πειραμάτων, κάνοντας την καταγραφή των αποκρίσεων των χρωμάτων αντικειμενική. Για την αναπαράσταση των χρωμάτων χρησιμοποιούνται 3 συνιστώσες X,Y,Z, στις οποίες τα βασικά χρώματα σχηματίζουν ένα χρωματικό τρισδιάστατο τρίγωνο. Τα βασικά χρώματα ενώνονται δηλαδή με ευθείες γραμμές και το χρωματικό διάγραμμα που προαναφέρθηκε σχηματίζεται με την προβολή των χρωμάτων στο επίπεδο του τριγώνου που σχηματίζεται με το κόκκινο και το πράσινο.

Ωστόσο η πρώτη αυτή προσπάθεια με το XYZ έχει δύο βασικά μειονεκτήματα: το ένα ότι είναι δύσκολο να προστεθεί η φωτεινότητα, και το άλλο ότι υπάρχει μια διάκριση ανάμεσα στις λαμβανόμενες διαφορές χρωμάτων και στη πραγματική διαφορά στο σύστημα. Για αυτό το λόγο από το XYZ φτιάχθηκε το Lab. Χρησιμοποιείται ένας μη γραμμικός μετασχηματισμός μετασχηματίζοντας το χρωματικό διάγραμμα στον τρισδιάστατο χώρο. Αυτός ο μετασχηματισμός έχει σαν αποτέλεσμα το Lab, να είναι σχεδόν ομοιόμορφο στην κατανομή των χρώματων για τον άνθρωπο, και να μπορεί να εφαρμοστεί η ευκλείδια απόσταση. Επίσης έχει την ιδιότητα οποιοσδήποτε συνδυασμός (a,b) σε σταθερή τιμή του L, να έχει πάντα την ίδια φωτεινότητα στην αντίληψη. Το σχήμα του συστήματος στον χώρο είναι σαν ένα σακί το οποίο στενεύει όσο χαμηλώνει η φωτεινότητα, μιμούμενο τις ιδιότητες του HSV, όχι όμως τόσο όσο το HSV, όπως θα φανεί από τα πειραματικά αποτελέσματα. Επιπλέον όμως το Lab διατηρεί την ιδιότητα του XYZ να είναι ανεξάρτητο συσκευής.

5.3 Εξαγωγή κυρίων χρωμάτων

Η προσέγγιση που ακολουθείται στην εργασία αυτή, για την περιγραφή της εικόνας με βάση το χρώμα, βασίζεται στην εξαγωγή των κυρίων χρωμάτων από την εικόνα (dominant colour extraction). Με τον όρο αυτό εννοείται η εξαγωγή ”κυρίαρχων” χρωμάτων της εικόνας, δηλαδή των χρωμάτων που επικρατούν στην εικόνα. Τα χρώματα αυτά, επιθυμητό είναι να συνοδεύονται από των αριθμό των pixel που κατέχουν στην εικόνα καθώς και τις περιοχές που καταλαμβάνουν. Άρα το πρόβλημα ανάγεται σε ”μερική” τμηματοποίηση, γιατί δεν είναι απαραίτητο να γνωρίζουμε τα ακριβή όρια των χρωματικών περιοχών. Αυτό βέβαια εξαρτάται από την εφαρμογή, αλλά για την ανάκληση η μη ακριβής τμηματοποίηση δεν θα επηρέαζε τον χακτηρισμό της εικόνας. Στις ακόλουθες υποενότητες παρουσιάζονται τρεις προσεγγίσεις που δοκιμάστηκαν. Η πρώτες δύο είναι η ιεραρχική ομαδοποίηση, και τμηματοποίηση με μοντέλο μίξης κατανομών (mixture decomposition) για να καταλήξουμε στην προτεινόμενη μέθοδο με ομαδοποίηση βασισμένη σε K-means.

5.3.1 Ιεραρχική ομαδοποίηση

Η μέθοδος βασίζεται στον κλασικό αλγόριθμο ιεραρχικής ομαδοποίησης από την αναγνώρηση προτύπων ([29]). Τα *αρχικά χρώματα* ενώνονται κατά ζεύγη και αντικαθίστανται με ένα χρώμα που είναι η μέση τιμή τους. Κάθε φορά γίνεται μια τέτοια ένωση. Τα χρώματα που ενώνονται κάθε φορά είναι τα πιο κοντινά σύμφωνα με κάποιο κριτήριο διαφοράς χρώματων σε ένα σύστημα χρώματων. Η διαδικασία σταματά όταν έχουμε φτάσει στον επιθυμητό αριθμό χρώματων, καθώς τα χρώματα σε κάθε επανάληψη μειώνονται κατά ένα. Αυτό που δεν έχει καθοριστεί είναι ποια θα είναι τα αρχικά χρώματα. Για αυτόν τον λόγο η εικόνα χωρίζεται σε ισομεγέθη τμήματα (block), όπου το κάθε τμήμα θα έχει ένα χρώμα αντιπρόσωπο, που θα υπολογίζεται από την μέση τιμή όλων των χρωμάτων των pixel σε αυτό. Η χρήση του block κάνει την εξαγωγή των χαρακτηριστικών χρωμάτων πολύ αδρά (coarse) και αυτά δεν ανταποκρίνονται στα αντιπροσωπευτικά χρώματα της εικόνας, όσο το μέγεθος του block μεγαλώνει. Ενώ ο χρόνος που απαιτείται αυξάνει πάρα πολύ όσο μικραίνει το block, με σκοπό να εξαχθούν πιο ”καθαρά” χρώματα. Αυτό γίνεται γιατί για να γίνει μία ένωση πρέπει να υπολογιστούν οι αναμεταξύ αποστάσεις όλων των χρωμάτων. Τουλάχιστον την πρώτη φορά μόνο, γιατί σε κάθε επίπεδο μπορούν να κρατούνται οι παλιές τιμές των αποστάσεων και να υπολογίζονται οι αποστάσεις από το καινούργιο χρώμα (αυτό που

προήλθε από την προηγούμενη ένωση). Αλλά και πάλι είναι πολύ αργό γιατί χρειάζονται πάρα πολλές επαναλήψεις, όταν μικραίνει το μέγεθος του block.

5.3.2 Μοντέλο μίξης κατανομών

Η μέθοδος αυτή έχει χρησιμοποιηθεί σε πολλές περιπτώσεις όπου απαιτείται τμηματοποίηση [40] [41]. Επίσης έχει χρησιμοποιηθεί για τμηματοποίηση με συνδυασμό χρώματος και υφής [26]. Για τους σκοπούς της εργασίας χρησιμοποιήθηκε η μίξη κατανομών για τμηματοποίηση με μόνο το χρώμα. Η προσέγγιση αυτή θεωρεί τα τμήματα της εικόνας σαν ξεχωριστές κατανομές που αναμιγνύονται μεταξύ τους. Στην προκυμένη περίπτωση θεωρούμε ότι κλάσεις-χρώματα ακολουθούν κατανομή Gauss. Η λύση του προβλήματος ανάγεται στην εύρεση των παραμέτρων των κατανομών με μέση τιμή μ_i, πίνακα συνδιασποράς Σ_i, apriori πιθανότητα $P(\omega_i)$ της κάθε κλάσης ω_i, με i=1$\ldots C$, όπου C ο αριθμός των κλάσεων. Σύμφωνα με το κριτήριο της μέγιστης πιθανοφάνειας (Maximum Likelyhood) οι συνθήκες για την εκτίμηση των βέλτιστων λύσεων με βάση επαναληπτικό αλγόριθμο είναι [29]:

$$P(\omega_i) = \frac{1}{N}\sum_{k=1}^{N} P(\omega_i)\,|x_k) \tag{5.2}$$

$$\mu_i = \frac{\sum_{k=1}^{N} P(\omega_i)\,|x_k)\,x_k}{\sum_{k=1}^{N} P(\omega_i)\,|x_k)} \tag{5.3}$$

$$\Sigma_i = \frac{\sum_{k=1}^{N} P(\omega_i)\,|x_k)\,(x_k-\mu_i)\,(x_k-\mu_i)^t}{\sum_{k=1}^{N} P(\omega_i)\,|x_k)} \tag{5.4}$$

με

$$\left.\begin{aligned} P(\omega_i|x_k) &= \frac{p(x_k|\omega_i)P(\omega_i)}{\sum_{j=1}^{N} p(x_k|\omega_j)P(\omega_j)} \\ &= \frac{|\Sigma_i|^{-\frac{1}{2}} exp\,[-\frac{1}{2}(x_k-\mu_i)^t\Sigma_i^{-1}(x_k-\mu_i)]\,P(\omega_i)}{\sum_{j=1}^{C}|\Sigma_j|^{-\frac{1}{2}} exp[-\frac{1}{2}(x_k-\mu_j)^t\Sigma_j^{-1}(x_k-\mu_j)]\,P(\omega_j)} \end{aligned}\right\} \tag{5.5}$$

όπου N είναι ο αριθμός των δεδομένων διανυσμάτων χρώματος x (διάστασης τρία), στην προκειμένη περίπτωση ο αριθμός των pixel στην εικόνα. Όπου $p(x_k|\omega_i)$ συνάρτηση πυκνότητας πιθανότητας της κλάσης ω_i, και $P(\omega_i|x_k)$ η εκ των υστέρων πιθανότητα της κλάσης ω_i, δηλαδή ποια η πιθανότητα να έχουμε κλάση ω_i όταν παρατηρούμε χρώμα x_k. Η επαναληπτική μέθοδος δουλεύει ως εξής:

1. θεωρώντας τον αριθμό των κλάσεων δεδομένο, αρχικοποιούμε με τιμές από το ιστόγραμμα τις αρχικές μέσες τιμές και τις πιθανότητες των κλάσεων ως ισοπίθανες.

2. υπολογισμός με την σχέση (5.4) του πίνακα συνδιασποράς.

3. εκτίμηση της εκ των υστέρων πιθανότητας με βάση την (5.5)

4. κατάταξη των pixel με βάση την μέγιστη εκ των υστέρων πιθανότητα.

5. υπολογισμός μέσων τιμών με την (5.3), πιθανοτήτων κλάσεων με την (5.2) και πίνακα συνδιασποράς με την (5.4). Μετά ακολουθείτε το βήμα (3).

Η επαναληπτική διαδικασία σταματά όταν ο αλγόριθμος συγκλίνει στην λύση όταν δηλαδή σταματίσουν να αλλάζουν οι μέσες τιμές. Το μειονέκτημα που έχει είναι ότι αργεί να συγκλίνει, χρειάζεται σχετικά πολλές επαναλήψεις. Επαναλήψεις που κοστίζουν αφού στους υπολογισμούς εμπεριέχονται αντιστροφές πινάκων για κάθε κλάση, καθώς και πολλοί υπολογισμοί για κάθε pixel ξεχωριστά. Όσο πιο πολλές κλάσεις τόσο πιο πολύ επιβραδύνεται ο χρόνος μιας επανάληψης ενός αλγόριθμου. Επίσης κλάσεις δεν έχουν απαραίτητα κατανομή Gauss όπως θεωρήθηκε. Το χρωματικό σύστημα που χρησιμοποιήθηκε είναι το HSV μετασχηματιμένο όπως στην 5.1.

5.3.3 K-means ομαδοποίηση

5.3.3.1 Βασικός αλγόριθμος

Επειδή δεν έχουμε πάντα κατανομές κλάσεων που να προσεγγίζουν την Gauss κατανομή, και επειδή η προσέγγιση με την μίξη κατανομών επίσης είναι αργή για μια εφαρμογή ανάκλησης και ευρετηριασμού εικόνων, προτιμήθηκε η εξαγωγή των κυρίαρχων χρωμάτων να γίνει με τον αλγόριθμο ομαδοποίησης K-means [29]. Ο K-means έχει το πλεονέκτημα ότι δεν χρειάζεται παραδοχές για τις κατανομές των κλάσεων. Επίσης είναι απλά εφαρμόσιμος και υλοποιήσιμος. Τα βήματα του K-means είναι απλά και προκύπτουν από την ελαχιστοποίηση του τετραγωνικού σφάλματος. Αρχικά θεωρούμε γνωστό τον αριθμό και τις τιμές των αντιπροσώπων χρώματων, αρχικοποιημένες εκ των προτέρων. Τα βήματα είναι τα εξής:

1. κατάταξη των pixel με βάση το χρώμα τους συγκρίνοντας τα με όλους τους αντιπροσώπους. Γίνεται χρήση μέτρου διαφοράς χρώματων σε επιλεγμένο σύστημα χρωμάτων. Το κάθε pixel κατατάσσεται στην κλάση όπου έχει την μικρότερη τιμή το μέτρο.

2. Υπολογισμός εκ νέου, μετά την αρχικοποίηση, των αντιπροσώπων ως την μέση τιμή όλων των χρωμάτων των pixel που ανήκουν στις αντίστοιχες κλάσεις.

3. αν οι μέσες τιμές δεν μεταβληθούν (χρήση κάποιου κάποιου κατωφλιού) τότε τερματίζεται ο αλγόριθμος με κύρια χρώματα τους αντιπροσώπους στην τελευταία επανάληψη, και σαν τμηματοποιημένη εικόνα οι κατατάξεις όπως έχουν, αλλιώς επιστροφή στο βήμα (1).

Αν σε κάποια επανάληψη μια κλάση μείνει χωρίς σημεία τότε αυτή αυτομάτως εξαιρείται από την διαδικασία και τα κυρίαρχα χρώματα μειώνονται κατά ένα. Η απόσταση μεταξύ χρωμάτων που χρησιμοποιείται είναι η ευκλείδια. Τα συστήματα χρωμάτων που ενσωματώθηκαν στον αλγόριθμο είναι το μετασχηματισμένο HSV [26] και το Lab, συστήματα στα οποία έχει νόημα να χρησιμοποιήσουμε ευκλείδια απόσταση καθώς είναι perceptually uniform.

5.3.3.2 Αρχικοποίηση

Μέχρι στιγμής οι τιμές των αντιπροσώπων των κλάσεων στο ξεκίνημα του επαναληπτικού αλγορίθμου θεωρήθηκαν γνωστές. Για να γίνει η αυτόματη η εξαγωγή των κυρίαρχων χρωμάτων αναπτύχθηκε μια μέθοδος αρχικοποίησης από το ιστόγραμμα των χρωμάτων της εικόνας. Όσον αφορά τον αριθμό των κυρίαρχων χρωμάτων θεωρείται σταθερός ίσος με 16. Ο αριθμός αυτός εκτιμήθηκε μέσω της εφαρμογής του αλγορίθμου, αφού πειραματικά διαπιστώθηκε ότι ο αριθμός αυτός είναι αρκετός για να περιγράψει τα βασικά χρώματα στην εικόνα. Στην περίπτωση που στην πραγματικότητα θα χρειαζόντουσαν πιο λίγα χρώματα, γίνεται μια επεξεργασία εκ των υστέρων (post-processing) που περιγράφεται στην επόμενη υποενότητα.

Όσον αφορά τον υπολογισμό των τιμών τους υπολογίζεται το ιστόγραμμα χρωμάτων (**colour histogram**) στο RGB σύστημα χρώματων. Το ιστόγραμμα υπολογιζόμενο υπόκειται κβαντισμό σε κάθε συνιστώσα, έτσι ώστε να έχουμε οκτώ επίπεδα τιμών σε κάθε συνιστώσα, προκύπτωντας 512 διαφορετικά χρώματα. Ο κβαντισμός είναι ομοιόμορφος, δηλαδή το αρχικό διάστημα τιμών (0-255) κάθε συνιστώσας διαιρείται σε 8 ισομήκη διαστήματα, με κάθε ένα από αυτά να αντιστοιχείται με το μέσο του διαστήματος στην αρχική κλίμακα. Η διαδικασία αυτή είναι υποχρεωτική γιατί χωρίς κβαντισμό έχουμε 256^3 διαφορετικά χρώματα κάνοντας το ιστόγραμμα πολύ αραιό (sparse), γιατί τα περισσότερα χρώματα του συστήματος δεν θα υπήρχαν στην εικόνα. Αυτά που υπήρχαν θα είχαν πολύ μικρές συχνότητες εμφάνισης κάνοντας το ιστόγραμμα δύσχρηστο για την εξαγωγή αρχικών χρωμάτων. Επίσης τα 512 χρώματα είναι αρκετά για να καλύψουν, εν μέρη, την αντίληψη του ανθρώπινου μάτιου.

Έχοντας υπολογίσει το κβαντισμένο ιστόγραμμα χρωμάτων, δηλαδή τις συχνότητες εμφάνισης σε κάθε κβαντισμένη χρωματική θέση (histogram bin), κάθε θέση υπόκειται σε ένα φιλτράρισμα. Επιτρέπονται bins στα οποία ποσοστιαία συχνότητα εμφάνισης ξεπερνάει ένα κατώφλι, στην προκειμένη περίπτωση εμπειρικά 0.05%. Επίσης επιτρέπεται ένας μέγιστος αριθμός από 100 χρώματα να διέλθουν από την διαδικασία αυτή. Αυτό γίνεται για την σταθεροποίηση του χρόνου εκτέλεσης, αφού στην συνέχεια απαιτείται ο υπολογισμός όλων των ανά δύο αποστάσεων των χρωμάτων. Επίσης 100 διαφορετικά χρώματα με τις μεγαλύτερες εμφανίσεις στην εικόνα είναι επαρκή πρακτικά.

Τα χρώματα ομαδοποιούνται με βάση τον ιεραρχικό αλγόριθμο ομαδοποίησης [29] που έχει αναφερθεί και σε προηγούμενες ενότητες. Σε κάθε επανάληψη του αλγορίθμου υπολογίζονται οι ανά δύο αποστάσεις χρωμάτων και ενώνονται τα κοντινότερα χρώματα, δημιουργώντας ένα νέο που είναι το σταθμισμένο άθροισμα με βάση τη ποσοστιαία συχνότητα εμφάνισης κάθε χρώματος. Μετά την ένωση οι ποσοστιαίες συχνότητες εμφάνισης αθροίζονται για να υπολογιστεί η συχνότητα του νέου χρώματος. Οι επαναλήψεις σταματούν εφ' όσον φτάσουμε στα 16 χρώματά με τις ομαδοποιήσεις. Οι συνάρτηση απόστασης χρωμάτων είναι η ευκλείδια και δεν υπολογίζεται στο RGB σύστημα χρωμάτων αλλά στο Lab ή στο μετασχηματισμένο HSV.

5.3.3.3 Επεξεργασία εκ των υστέρων

Με την περάτωση των επαναλήψεων παρατηρήθηκε ότι πολλές φορές κλάσεις έχουν χρώματα τα οποία μοιάζουν πάρα πολύ μεταξύ τους. Οπότε εκ των υστέρων ομαδοποιούνται τέτοια κυρίαρχα χρώματα ως εξής: υπολογίζονται όλες οι αποστάσεις μεταξύ τους και ιεραρχικά ενώνονται κάθε φορά τα δύο κοντινότερα χρώματα, εάν η απόσταση τους είναι κάτω από ένα κατώφλι το οποίο εκτιμήθηκε εμπειρικά. Από τα δύο χρώματα αυτά με βάση τα pixel που αντιπροσωπεύουν στην εικόνα υπολογίζεται ο μέσος αντιπρόσωπος και ξαναγίνεται η ίδια διαδικασία έως ότου δεν υπάρχουν χρώματα που ενώνονται με βάση το κατώφλι. Επίσης παράλληλα ενώνονται και οι ετικέτες (labels) κατατάξεων στην τμηματοποιημένη εικόνα, όταν ενώνουμε χρώματα.

5.3.3.4 Πειραματικά αποτελέσματα

Η παραπάνω μέθοδος για την εξαγωγή κυρίων χρωμάτων εφαρμόστηκε σε διάφορες εικόνες για να αξιολογηθεί η επίδοση του αλγορίθμου. Επίσης χρησιμοποιήθηκαν δύο χρωματικά συστήματα, το καρτεσιανό HSV (εξίσωση 5.1), και το Lab, για να συγκριθούν οι επιδόσεις τους. Τα κατώφλια που χρησιμοποιήθηκαν για την εκ των υστέρων ένωση

χρώματων είναι 0.007 και 58 αντίστοιχα, τα οποία εξάχθηκαν πειραματικά. Η επιλογή των τιμών είναι δύσκολη, καθώς έγινε με τέτοιο τρόπο ώστε να μην ενώνονται περιοχές με διαφορετικά χρώματα που χρήζουν σημασίας. Ένα τέτοιο όριο δεν είναι σαφές καθώς σε όποιο σημείο και να μπει αυτό πάντα θα υπάρχουν χρώματα τα οποία θα έπρεπε να ενωθούν και δεν ενώθηκαν, και αντίστροφα που ενώθηκαν ενώ δεν έπρεπε. Το αν θα γίνει η ένωση είναι υποκειμενικό και εξαρτάται από το τι επιθυμείται κάθε φορά για τον χαρακτηρισμό της εικόνας, το οποίο ξεφεύγει από την ομοιογένεια στην αντίληψη από τον άνθρωπο που διαθέτουν εν μέρη τα συστήματα χρωμάτων.

Στο σχήματα 5.2, 5.3, 5.4 φαίνονται οι αρχικές εικόνες, και μετά απεικονιζόμενες με τα κύρια χρώματα τους, κάνοντας χρήση HSV και Lab αντίστοιχα. Στο σχήματα 5.5 5.6 5.7 φαίνονται τα κύρια χρώματα που εξάχθηκαν και με τα δύο συστήματα χρωμάτων για τα προηγούμενα τρία πειράματα αντίστοιχα και για τα δύο συστήματα χρωμάτων.

Όπως φαίνεται από τα αποτελέσματα τα δύο συστήματα έχουν αντίθετη συμπεριφορά. Το HSV δίνει μεγάλη έμφαση στην φωτεινότητα, όπως φαίνεται και στην σχέση 5.1, όπου υπάρχει και στους τρεις όρους. Αυτό έχει σαν αποτέλεσμα να χωρίζονται αποχρώσεις με βάση την φωτεινότητα, όπως στο σχήμα 5.2 στον ουρανό και στο σχήμα 5.3 στο πουκάμισο, όπου το πουκάμισο χωρίζεται ανάλογα με του που είναι πιο σκούρο ή πιο ανοικτό. Σε αντίθεση με το Lab όπου ομαδοποιεί όλο το πουκάμισο, γιατί παίζει μεγάλο ρόλο η απόχρωση από ότι η φωτεινότητα. Επίσης για αυτό το λόγο ενώνει αποχρώσεις που βρίσκονται στο πρόσωπα και κάνει όλο το πρόσωπο να έχει ενιαία απόχρωση όπως στο σχήμα 5.4. Το τι από τα δύο είναι καλύτερο εξαρτάται με το τι θέλουμε να κάνουμε μετά, π.χ. αν έπρεπε να χρησιμοποιήθει για την εξαγωγή προσώπων τότε είναι προτιμότερο να γίνει η χρήση του Lab. Σε άλλες περιπτώσεις όμως όπου δύο αντικείμενα έχουν κοντινή απόχρωση αλλά διαφέρουν στην φωτεινότητα, δεν διαχωρίζονται με το Lab. Σε πολλές περιπτώσεις όμως αυτό που κάνει το Lab είναι πιο κοντά πολλές φορές στο περιεχόμενο της εικόνας, καθώς απομονώνει αντικείμενα πρόσωπα με ενιαία απόχρωση. Το Lab έχει όμως ένα μεγάλο μειονέκτημα ότι οι τιμές στις χρωματικές συνιστώσες να είναι πολύ συμπυκνωμένες, δηλαδή να έχουν μικρή διασπορά. Μπορεί να έχει μεγάλο εύρος δυνατών τιμών στις συνιστώσες (a,b) αλλά οι τιμές που συναντώνται στις εικόνες είναι μαζεμένες σε μικρή περιοχή. Για παράδειγμα μπορεί η συνιστώσα (a) να έχει πεδίο τιμών [-137,96], αλλά παίρνει τιμές στο [-30,30]. Οι μεγάλες κατά απόλυτον τιμές για τα (a,b) αφορούν πολύ κορεσμένα χρώματα (με μεγάλο saturation), όπως ένα πολύ έντονο κόκκινο που δύσκολα συναντάται σε μία φυσική σκηνή. Αυτή η μικρή διασπορά χρωματικών τιμών είναι υπεύθυνη βέβαια για την ομαδοποίηση των όμοιων

αποχρώσεων ανεξαρτήτου μικρής αλλαγής στην φωτεινότητα, αλλά πολλές φορές ενώνει διαφορετικές αποχρώσεις, οι οποίες θα έπρεπε να είναι δύο ξεχωριστά κύρια χρώματα. Αυτό συναντάται στο σχήμα 5.2 για το Lab, όπου ενώνεται μικρού κορεσμού κίτρινο με μικρού κορεσμού πράσινο.

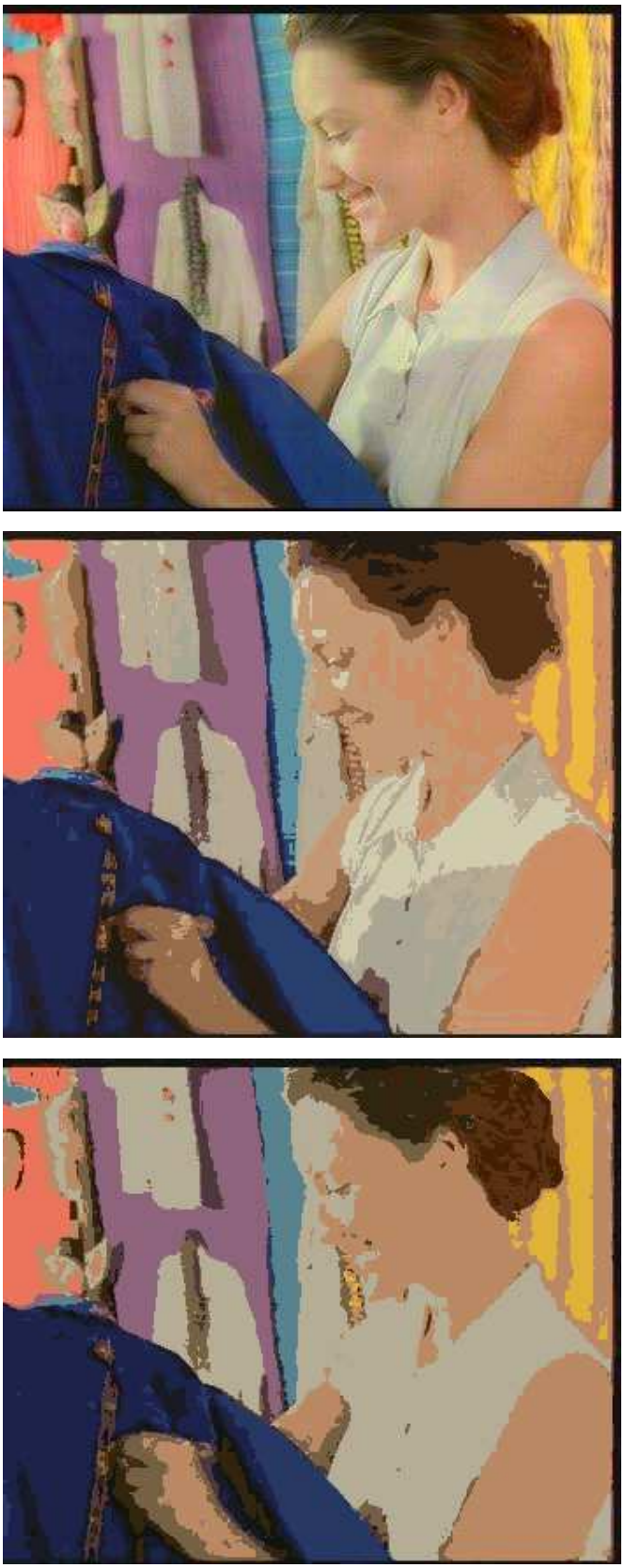
Το μεγαλύτερο μειονέκτημα είναι ότι δεν ενώνει των χαμηλών φωτεινοτήτων χρώματα. Όπως φαίνεται στα σχήματα 5.5, 5.6, 5.7 υπάρχουν πολλά κυρίαρχα χρώματα τα οποία στο μάτι φαίνονται ίδια επειδή είναι χαμηλής φωτεινότητας. Πράγματι όσο πιο λίγο φως τόσο πιο δύσκολα το ανθρώπινο μάτι μπορεί να διακρίνει διαφορετικές αποχρώσεις. Πχ. σε πολύ λίγο φως το μαύρο φαίνεται ίδιο με ένα σκούρο μπλε ή ένα σκούρο κόκκινο. Το Lab δεν λαμβάνει αυτό υπόψην του, όποτε εξάγονται αρκετά κυρίαρχα χρώματα τα οποία έχουν διαφορετική απόχρωση αλλά στο μάτι φαίνονται ίδια. Σε αντίθεση το HSV με τον τρόπο που μετασχηματίζεται αλλά και την δομή του, έχοντας σχήμα κώνου, αυτό που κάνει είναι να φέρνει πιο κοντά τα χρώματα όσο πιο πολύ μειώνεται η φωτεινότητα, δηλαδή όσο κατεβαίνουμε τον κώνο. Στα σχήματα 5.5, 5.6, 5.7 τα κυρίαρχα χρώματα του HSV δεν περιέχουν πολλές διαφορετικές σκοτεινές αποχρώσεις όπως στο Lab.

Επιπλέον το Lab αργεί περισσότερο να συγκλίνει από το HSV. Συνήθως χρειάζονται διπλάσιες επαναλήψεις, 35 με 75 κατά μέσο όρο αντίστοιχα. Αυτό έχει να κάνει βέβαια και την κλίμακα τιμών των δύο συστημάτων καθώς το HSV έχει πολύ μικρή κλίμακα τιμών, στο [0,1] η φωτεινότητα και [-1,1] οι άλλες δύο. Εν τω μεταξύ το κριτήριο σύγκλισης είναι το άθροισμα των τετραγωνικών διαφορών των παλαιών με των νέων αντιπροσώπων.

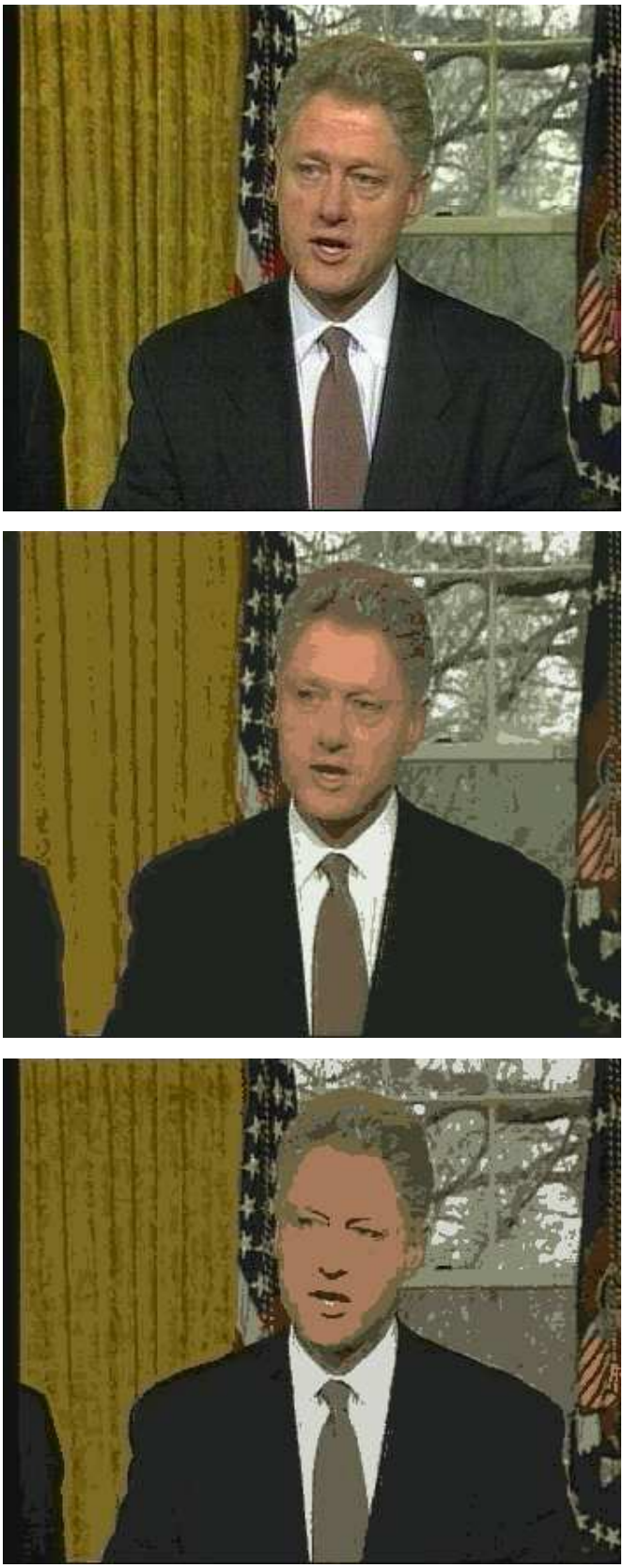
5.4 Περιγραφή εικόνας με βάση το χρώμα

Έχοντας εξάγει τα κύρια χρώματα σε μια εικόνα και τις περιοχές στις οποίες βρίσκονται στην εικόνα, μπορούμε να εξάγουμε χαρακτηριστικά, υψηλότερου σημασιολογικού επιπέδου, που να βασίζονται σε αυτά και να περιγράψουμε το περιεχόμενο της εικόνας. Επίσης να κατασκευάσουμε τρόπους σύγκρισης των χαρακτηριστικών αυτών. Η αμέσως επόμενη υποενότητα εξετάζει τις δυνατότητες που παρέχει η εξαγωγή των κυρίαρχων χρωμάτων όσον αφορά την περιγραφή του χρωματικού περιεχομένου της εικόνας. Ενώ η μεθεπόμενη παραθέτει μέτρα (κριτήρια) για την μέτρηση διαφοράς ή ομοιότητας μεταξύ των χαρακτηριστικών.

Σχήμα 5.2: Η αρχική εικόνα και οι απεικονίσεις της με την εξαγωγή κυρίων χρωμάτων με το HSV και Lab



Σχήμα 5.3: Η αρχική εικόνα και οι απεικονίσεις της με την εξαγωγή κυρίων χρωμάτων με το HSV και Lab



Σχήμα 5.4: Η αρχική εικόνα και οι απεικονίσεις της με την εξαγωγή κυρίων χρωμάτων με το HSV και Lab

Σχήμα 5.5: Παράθεση των εξαχθέντων κυρίων χρωμάτων με το HSV πάνω και Lab κάτω, για την που περιέχει τον δρόμο με τα διερχόμενα αυτοκίνητα

Σχήμα 5.6: Παράθεση των εξαχθέντων κυρίων χρωμάτων με το HSV πάνω και Lab κάτω, για την εικόνα που περιέχει την κυρία που αγοράζει υφάσματα

5.4.1 Χρωματικά χαρακτηριστικά

Ο πιο κοινός τρόπος, για να χαρακτηρίσουμε τις χρωματικές περιοχές που έχουν εξαχθεί, είναι ο υπολογισμός των αντιπροσώπων τους, ως την μέση τιμή όλων των pixel που ανήκουν στην περιοχή. Επίσης μπορεί να γίνει χρήση της διασποράς του χρώματος έτσι ώστε να προστεθεί η πληροφορία για το πόσο είναι διασπαρμένη χρωματικά μια περιοχή από το κυρίαρχο χρώμα. Ακόμα είναι δυνατή και η χρήση μεγαλύτερης τάξης στατιστικών όπως ροπές τρίτης τάξης [42].

Μια σημαντική πληροφορία όσον αφορά το περιεχόμενο είναι το πόσο αραιή (sparseness) είναι μια κλάση, δηλαδή πόσο διασπαρμένη βρίσκεται στο επίπεδο της εικόνας. Μια τέτοια πληροφορία θα μπορούσε να περιγραφεί από την μέση τιμή των συντεταγμένων των pixel και από την τυπική απόκλιση τους από αυτήν. Την πληροφορία αυτήν δεν μπορεί να περιγραφεί από την κλασικότερη και παλαιότερη μέθοδο περιγραφής εικόνας με βάση το χρώμα, τα έγχρωμα ιστογράμματα [43]. Ένα ιστόγραμμα θα κατέταζε όμοιες εικόνες με διαφορετικό περιεχόμενο όπως την εικόνα της γαλλικής και της αμερικάνικης σημαίας, καθώς έχουν περίπου ίδιες συνεισφορές στο μπλε στο άσπρο

Σχήμα 5.7: Παράθεση των εξαχθέντων κυρίων χρωμάτων με το HSV πάνω και Lab κάτω, για την εικόνα που περιέχει τον Bill Clinton

και στο κόκκινο. Αφού μόνο αυτήν την πληροφορία μπορεί να παρέχει το ιστόγραμμα, δηλαδή την συνεισφορά κάθε χρώματος στην εικόνα, και όχι το πως κατανέμεται κάθε χρώμα στην εικόνα.

Ο Pass προσφάτως για να προσθέσει τέτοια πληροφορία στο ιστόγραμμα, εισήγαγε τα χρωματικά συναφή διανύσματα (Color Coherence Vector, CCV). Η μέθοδος είναι μια επέκταση του ιστογράμματος με την διαφορά ότι εισάγει χωρική πληροφορία σε αυτό. Κάθε bin j του ιστογράμματος αντικαθίσταται από ένα ζευγάρι τιμών (α, β). Όπου α είναι ο αριθμός των pixel που ανήκουν σε συναφείς (συμπαγείς) μεγάλες περιοχές της εικόνας (coherent pixels) του χρώματος που αντιπροσωπεύει το bin j, ενώ β είναι ο αριθμός των μη συναφών (incoherent pixels). Το άθροισμα των δύο αυτών τιμών δίνει την τιμή του bin στο αρχικό ιστόγραμμα. Ο Zabih για τον ίδιο λόγο κατασκεύασε τα έγχρωμα διαγράμματα συσχέτισης (colour correlograms, συσχετογράμματα) [44]. Το colour correlogram εκφράζει την χωρική συσχέτιση ζευγαριών χρώματων με την απόσταση. Μια θέση του συσχετογράμματος αντιπροσωπεύει την πιθανότητα από ένα χρώμα c_i σε απόσταση κ να έχουμε χρώμα c_j. Ακόμη έχει προταθεί η χρήση διδιάστατων ψευδο-κρυμένων μοντέλων Markov (2D pseudo-hidden Markov model, PHMM) για την εξαγωγή παραμέτρων που θα εκφράζουν τις χρωματικές συσχετίσεις στο επίπεδο της εικόνας [45].

Επίσης έχουν προταθεί μέθοδοι εξαγωγής χρωματικών περιοχών οι οποίες να έχουν σημασιολογικό περιεχόμενο [46]. Τέτοιες περιοχές οφείλουν να είναι συμπαγείς και να έχουν ένα υπολογίσιμο αριθμό από pixel. Με αυτό τον τρόπο αυτομάτως απορρίπτονται μικρές περιοχές στην εικόνα με τα κύρια χρώματα, που δεν παρέχουν ουσιαστική πληροφορία για το περιεχόμενο της εικόνας. Αποθηκεύονται σαν χαρακτηριστικά οι θέσεις στην εικόνα των περιοχών αυτών, τα χρώματα που περιέχουν καθώς και τα μέγεθη τους. Η προσέγγιση αυτή δίνει την δυνατότητα επερωτήσεων που να αφορούν την τοπολογική σχέση χρωματικών περιοχών στην εικόνα, με την χρήση χωρικών βάσεων δεδομένων. Για παράδειγμα "βρες μου μια εικόνα που να περιέχει πάνω μπλε και κάτω πράσινο". Περιγράφοντας έτσι μια εικόνα η οποία να περιέχει πιθανόν ένα πράσινο τοπίο με γαλάζιο ουρανό πάνω από αυτό. Ή επίσης να θέλουμε μια εικόνα να έχει αλλεπάλληλες εναλλαγές περιοχών μπλε και άσπρου ψάχνοντας έτσι για την ελληνική σημαία.

Μια εφαρμογή που χρησιμοποιεί άμεσα την εξαγωγή κυρίων χρώματων είναι σύγκριση λογοτύπων (trademarks) εταιρειών, έτσι ώστε να εντοπίζονται απομιμήσεις. Αυτό γίνεται χρησιμοποιώντας ένα τρόπο σύγκρισης των εξαχθέντων κυρίων χρωμάτων. Η σύγκριση που προτείνεται, κατ' αρχήν υπολογίζει όλα τις αναμεταξύ αποστάσεις των χρωμάτων. Η απόσταση αυτή μπορεί να βασίζεται σε διάφορα χαρακτηριστικά των

εξαχθέντων χρωματικών κλάσεων, όπως μέσες τιμές και διασπορές χρωμάτων, πλήθος σημείων, ή το πόσο συμπαγείς είναι όσον αφορά την κατανομή των σημείων στην εικόνα. Όταν υπολογιστούν όλες οι αναμεταξύ αποστάσεις, με κάποιο μέτρο σύγκρισης, τις αποθηκεύουμε σε ένα πίνακα A(i,j), όπου η θέση (i,j) περιέχει την απόσταση του κυρίαρχου χρώματος i της πρώτης εικόνας από το κυρίαρχο χρώμα j της δεύτερης εικόνας. Κάθε φορά βρίσκουμε το ελάχιστο στοιχείο του πίνακα A και το κρατάμε. Μετά σβήνουμε την i γραμμή και την j στήλη υποθέτοντας ότι οι δείκτες των γραμμών αντιστοιχούν στα κύρια χρώματα της πρώτης εικόνας, ενώ οι στήλες της δεύτερης. Σβήνοντας την γραμμή i και την στήλη j σημαίνει ότι ταιρίαξε καλύτερα από όλα τα άλλα χρώματα το i και το j. Όποτε αυτά βγουν εκτός της υπόλοιπης διαδικασίας συγκρίσεως. Μετά ακολουθείται η ίδια διαδικασία ψάχνοντας την επόμενη ελάχιστη τιμή στα στοιχεία που έμειναν στον πίνακα. Η διαδικασία αυτή συνεχίζεται έως ότου τελειώσουν οι δείκτες στις γραμμές ή στις στήλες, γιατί δύο εικόνες δεν έχουν απαραίτητα τον ίδιο αριθμό κυρίαρχων χρώματων. Η συνολική απόσταση των δύο εικόνων είναι το άθροισμα όλων των ελάχιστων τιμών που επιλέχθηκαν.

Ο παρών αλγόριθμος εξαγωγής κυρίαρχων χρωμάτων αναπτύχθηκε για το πρόγραμμα DIVAN [47] όπου εκτός για τον χρωματικό χαρακτηρισμό σκηνών (scene/shot) και πλαισίων κλειδιών (key-frame) ενός video [48], χρησιμοποιείται για την ανίχνευση προσώπων [49], ανίχνευση κειμένου στην εικόνα.

5.4.2 Μέτρα σύγκρισης χαρακτηριστικών

Το πιο κοινό μέτρο ομοιότητας για την σύγκριση ιστογραμμάτων είναι η τομή δύο ιστογραμμάτων.

$$I(h_1, h_2) = \frac{\sum_j min(h_1(j), h_2(j))}{\sum_j h_1(j)} \tag{5.6}$$

όπου h_1, h_2, είναι τα δύο υπό σύγκριση ιστογράμματα, i είναι ο δείκτης που διατρέχει όλα τα bin. Άλλο μέτρο που έχει χρησιμοποιηθεί για ιστογράμματα είναι το χ^2 τεστ:

$$\chi^2 = \sum \frac{(h_1(j) - h_2(j))^2}{(h_1(j) + h_2(j))} \tag{5.7}$$

Το μέτρο αυτό είναι μέτρο απόστασης, σε αντίθεση με την τομή, δηλαδή μικρές τιμές σημαίνουν μεγάλες ομοιότητες. Επίσης χρησιμοποιούνται τετραγωνικού τύπου αποστάσεις (Quadratic) όπως στο σύστημα QBIC [50].

$$d_{hist} = (h_1 - h_2)^t A (h_1 - h_2) \tag{5.8}$$

όπου A=$[a_{ij}]$ είναι ένας πίνακας συσχετίσεων μεταξύ δύο χρωμάτων i, j στο ιστόγραμμα. Επίσης για την διαφορά ιστογραμμάτων χρησιμοποιείται η λογαριθμισμένη εξίσωση ταιριάσματος κατανομών στην οποία βασίζεται η απόσταση Bhattacharya [19]:

$$d_{hist}=log(\sum_i \sqrt{h_1(i)\,h_2(i)}) \tag{5.9}$$

Στην περίπτωση αυτή χρειάζεται τα ιστογράμματα να είναι κανονικοποιημένα $\sum_i h(i)$ =1, δηλαδή να είναι συνάρτηση πυκνότητας πιθανότητας.

Όσον αφορά μέτρα για διαφορές χρωματικών χαρακτηριστικών (στον $\mathcal{R}^n$) είναι η ευκλείδια (L_2 metric):

$$d_{L_2}(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \tag{5.10}$$

με $x, y \in \mathcal{R}^n$. Επίσης χρησιμοποιούνται και στατιστικά κριτήρια όπως η απόσταση Bhattacharya για την ομοιότητα χρωματικών περιοχών, κάνοντας χρήση της παραδοχής ότι η κατανομή του χρώματος στην περιοχή ακολουθεί κατανομή Gauss.

$$d_b=\sum_{i=1}^{n}(\frac{1}{2}\frac{\sigma_{1,i}^2 - \sigma_{2,i}^2}{\sqrt{\sigma_{1,i}^2\sigma_{2,i}^2}} + \frac{1}{2}\frac{(\mu_{1,i} - \mu_{2,i})^2}{\mu_{1,i} - \mu_{2,i}}) \tag{5.11}$$

όπου $\mu_1, \mu_2 \in \mathcal{R}^n$ είναι οι μέσες τιμές των χρωματικών περιοχών (dominant color), και $\sigma_1, \sigma_2 \in \mathcal{R}^n$ είναι οι διασπορές των χρωμάτων στις περιοχές αυτές. Σε όλες τις περιπτώσεις που έχουμε συγκρίσεις χρωματικών τιμών τότε το n έχει τιμή 3. Επίσης χρησιμοποιείται η απόσταση συνημιτόνου ιδιαίτερα στο RGB σύστημα χρωμάτων, όπου δεν έχει νόημα η εφαρμογή της ευκλείδιας [51].

$$\theta(x,y) = arccos(\frac{x, y^t}{|x||y|}) \tag{5.12}$$

Στην ουσία αυτή η απόσταση υπολογίζει την γωνία των δύο χρωματικών διανυσμάτων, όσο πιο μικρή η γωνία τόσο πιο κοντά είναι τα διανύσματα αυτά. Αυτή η απόσταση έχει νόημα για το RGB όπου η κατεύθυνση του χρωματικού διανύσματος δηλώνει εν μέρη την απόχρωση.

Κεφάλαιο 6

Κατηγοριοποίηση με τη χρήση χρώματος και υφής

Σε αυτό το κεφάλαιο γίνεται διερεύνηση του κατά πόσο βοηθά το χρώμα στην κατηγοριοποίηση εικόνων σε συνδυασμό με την υφή. Προτείνονται τρόποι για τον συνδυασμό των χαρακτηριστικών του χρώματος και της υφής. Στο παρελθόν έχουν προταθεί τρόποι για τον συνδυασμό του χρώματος και της υφής από κοινού με ένα διάνυσμα χαρακτηριστικών ([26] , [52]) και κοινή εξίσωση απόστασης. Τα χαρακτηριστικά μπορεί να μην είναι κοινά πχ. διάνυσμα διασπορών για υφή και ιστογράμματα για χρώμα ([53] [52]). Επίσης μπορεί να χρησιμοποιηθεί ξεχωριστή απόσταση διαφορετικών χαρακτηριστικών και εκ των υστέρων συνδυασμός υφής και χρώματος με βάση τις επιμέρους κατατάξεις των αποστάσεων. Όπως πρότεινε η Picard [20] για τον συνδυασμό Wold χαρακτηριστικών και MRSAR χαρακτηριστικών. Στο κεφάλαιο αυτό γίνονται πειράματα με κοινή απόσταση και μη, συνδυάζοντας την υφή και το χρώμα στην βάση VisTex από το MIT που παρουσιάστηκε στο κεφάλαιο 3. Το πείραμα ταξινόμησης είναι το ίδιο με το πρότυπο πείραμα που παρουσιάστηκε στην ενότητα 3.4. Με αυτόν τον τρόπο μπορεί να διαπιστωθεί κατά πόσο και ποια χαρακτηριστικά του χρώματος βοηθούν στην βελτίωση του ποσοστού επιτυχίας.

6.1 Χρήση κοινής απόστασης χρώματος και υφής

Μια πρώτη προσέγγιση, στον συνδυασμό υφής και χρώματος, είναι να προσθέσουμε τις χρωματικές συνιστώσες (a,b) του συστήματος χρωμάτων Lab. Θεωρώντας ότι τα χαρακτηριστικά αυτά του χρώματος ακολουθούν κατανομή είτε Gauss είτε Laplace,

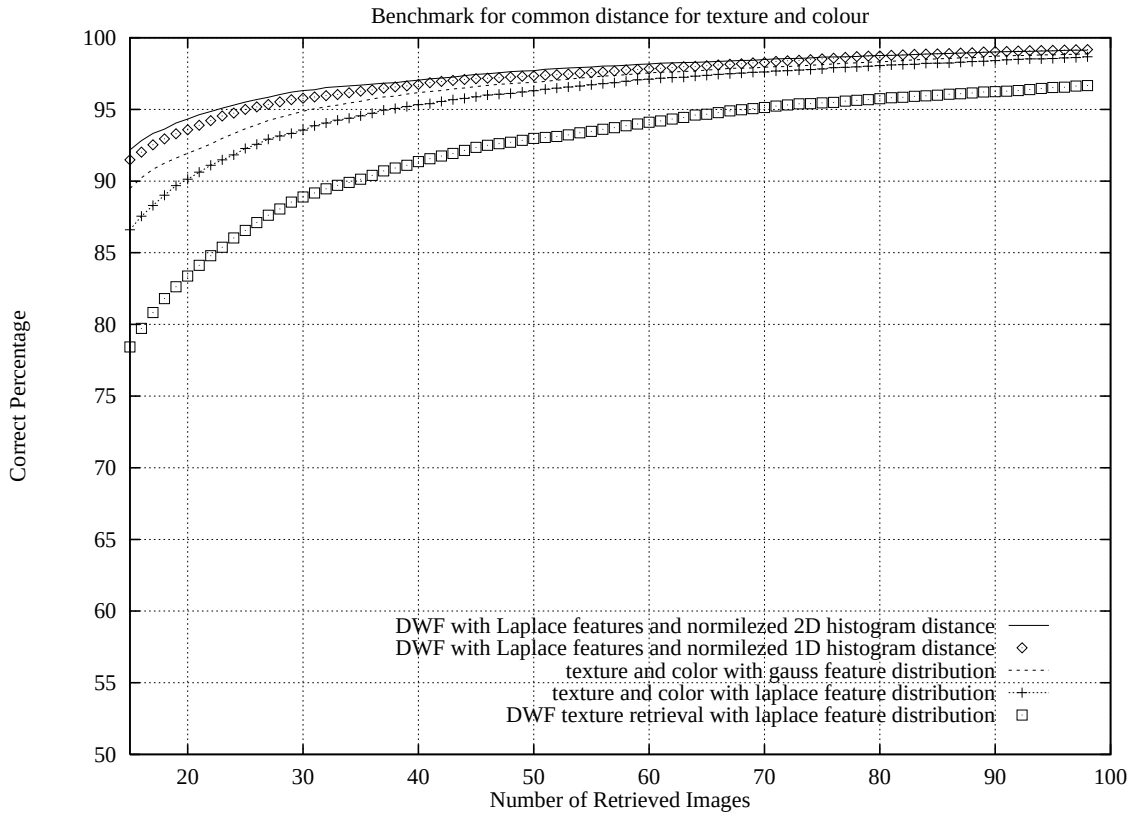
μπορούμε να χρησιμοποιήσουμε κοινό διάνυσμα χαρακτηριστικών. Στο διάνυσμα χαρακτηριστικών της υφής προστείθονται οι διασπορές των (a,b) ξεχωριστά. Έτσι μπορεί κάλλιστα να γίνει η χρήση της απόστασης Bhattacharya όπως και πριν. Μόνο που για τις δύο τελευταίες συνιστώσες προστίθεται ακόμα ένας όρος στην απόσταση που αφορά την μέση τιμή αυτών (εξίσωση 5.11). Μια που τα χαρακτηριστικά δεν έχουν απαραίτητα μέση τιμή μηδέν. Η προσέγγιση αυτή όμως έχει τον περιορισμό ότι οι εικόνες πρέπει να έχουν κατανομή τύπου Gauss, σε κάθε συνιστώσα από τις δύο. Δηλαδή πρέπει οι αποχρώσεις να συγκεντρώνονται γύρω από συγκεκριμένη απόχρωση σε κάθε συνιστώσα. Αυτό έχει βάση στο συγκεκριμένο πείραμα γιατί οι εικόνες από το VisTex είναι ομοιογενείς κατά βάσην και έχουν οι περισσότερες μια κύρια απόχρωση.

Αν όμως η βάση είναι χωρίς περιορισμούς (unconstrained), δηλαδή γενικές εικόνες και όχι εικόνες που περιέχουν υφή, τότε πρέπει να χρησιμοποιηθεί άλλος τρόπος χαρακτηρισμού του χρώματος. Ένας τέτοιος τρόπος είναι η χρήση ιστογραμμάτων. Ακόμη και αν στην συγκεκριμένη βάση που οι κατανομές προσεγγίζουν, έστω οι περισσότερες, την Gauss, πάλι είναι πιο ελκυστική η χρήση του ιστογράμματος καθώς ταιριάζει ακριβώς τα δεδομένα στην εικόνα και όχι τα στατιστικά τους. Για να χρησιμοποιηθούν τα ιστογράμματα των χρωματικών συνιστωσών μαζί με το διάνυσμα χαρακτηριστικών της υφής, πρέπει να γίνει χρήση μιας κανονικοποιημένης απόστασης ιστογραμμάτων. Όποτε χρησιμοποιώντας την απόσταση 5.9, μπορούμε να αθροίσουμε τους όρους της Bhattacharya για την υφή με τους δύο όρους για την απόσταση των ιστογραμμάτων των συνιστωσών a και b, υιοθετώντας έτσι ενιαία απόσταση χαρακτηριστικών. Προϋπόθεση για να είναι κανονικοποιημένη η απόσταση των ιστογραμμάτων είναι το άθροισμα όλων των τιμών του κάθε ιστογράμματος να είναι μονάδα.

Βέβαια χρησιμοποιώντας ξεχωριστά τις συνιστώσες του χρώματος θεωρούμε ότι δεν συσχετίζονται. Με αυτό τον τρόπο χάνουμε πολύ πληροφορία για το πώς κατανέμονται οι αποχρώσεις στον διδιάστατο χώρο, δηλαδή πως συνδυάζονται μεταξύ τους οι τιμές των (a,b), και ποιες αποχρώσεις πραγματικά υπάρχουν. Για αυτό τον λόγο είναι θεμιτή η χρήση ενός κανονικοποιημένου διδιάστατου ιστογράμματος των (a,b) ζευγών. Οι δυνατές θέσεις (bins) του ιστογράμματος αυτού είναι περίπου στις 53.000. Για αυτόν τον λόγο κβαντίζουμε τις δυνατές τιμές για να ελαττώσουμε το μέγεθος του ιστογράμματος. Δοκιμάστηκαν δύο ομοιόμορφοι κβάντισμοί με 256 θέσεις (16 επίπεδα σε κάθε συνιστώσα) και 1024 (32 επίπεδα σε κάθε συνιστώσα). Προτιμήθηκε ο δεύτερος γιατί μέσω πειραματικών αποτελεσμάτων είχε καλύτερη επίδοση. Αυτό γιατί το Lab συγκεντρώνει τις περισσότερες αποχρώσεις σε ένα μικρό υποσύνολο των δυνατών απο-

χρώσεων. Για αυτόν τον λόγο ένας κβαντισμός σε περισσότερα επίπεδα είναι αναγκαίος.

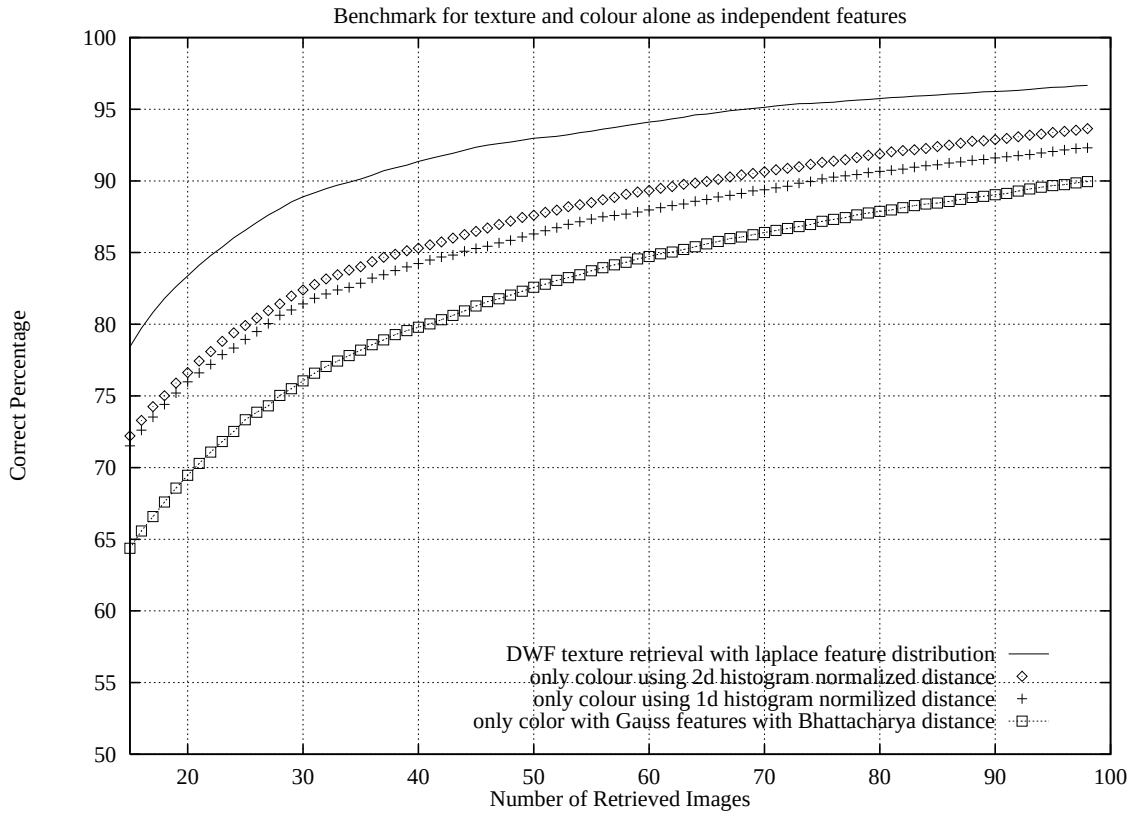
Στο σχήμα 6.1 φαίνονται οι επιδόσεις των παραπάνω μεθόδων για το πρότυπο πείραμα ανάκλησης εικόνων (ενότητα 3.4) για την βάση με τις 55 εικόνες από το VisTex 3.1. Πρώτο σε ποσοστό επιτυχίας έρχεται, όπως αναμενόταν, ο συνδυασμός με χρήση διδιάστατου ιστογράμματος, μετά τα μονοδιάστατα ιστογράμματα των a,b συνιστωσών, μετά η ενιαία Bhattacharya με Gauss κατανομή χαρακτηριστικών και μετά η ίδια με Laplace κατανομή, με ποσοστά επιτυχίας 91.3%, 90.6%, 88.5%, 85.3% αντίστοιχα. Επίσης παρατίθεται στο γράφημα αυτό και η καμπύλη με την ανάλυση DWF 5 επιπέδων, και Laplace κατανομή χαρακτηριστικών με ποσοστό 76.8%, η οποία έχει το καλύτερο αποτέλεσμα για χαρακτηριστικά υφής και μόνο, όπως παρουσιάσθηκε στην ενότητα 3.4. Η παράθεση του αποτελέσματος αυτού γίνεται για να φανεί πόσο ενισχύει το ποσοστό επιτυχίας η προσθήκη του χρώματος.



Σχήμα 6.1: Καμπύλη επιτυχίας, συναρτήσει του αριθμού εικόνων που ανακλήθηκαν από την βάση, κάνοντας χρήση κοινής απόστασης για χρώμα και για υφή.

Επίσης έγινε το πείραμα με τα παραπάνω χαρακτηριστικά χωρίς τον συνδυασμό της υφής. Αυτό έχει σαν σκοπό να διαπιστωθεί πιο χαρακτηριστικό έχει την καλύτερη επίδοση από μόνο του. Και επίσης να συμπερανθεί αν η υφή είναι πιο κυρίαρχο

χαρακτηριστικό, για τα συγκεκριμένα δεδομένα, από το χρώμα. Τα αποτελέσματα του πειράματος αυτού φαίνεται στο σχήμα 6.2. Με σειρά ποσοστού επιτυχίας είναι: υφή με DWF πέντε επιπέδων και Laplace κατανομή χαρακτηριστικών, διδιάστατο ιστόγραμμα των (a,b) συνιστωσών, μονοδιάστατα ιστογράμματα των a,b ξεχωριστά, και παραδοχή Gauss κατανομής των συνιστωσών a,b, με Bhattacharya απόσταση με ποσοστά επιτυχίας 76.8%, 71.1%, 70.2%, και 62.8% αντίστοιχα. Για την συγκεκριμένη περιορισμένη βάση δεδομένων η υφή παίζει μεγαλύτερο ρόλο από το χρώμα όπως αναμενόταν. Αλλά στην γενικότερη περίπτωση έχουμε το χρώμα να είναι πιο ισχυρό χαρακτηριστικό από την υφή [52].



Σχήμα 6.2: Καμπύλη επιτυχίας, συναρτήσει του αριθμού εικόνων που ανακλήθηκαν από την βάση, κάνοντας χρήση του χρώματος και της υφής ξεχωριστά.

6.2 Συνδυασμός εκ των υστέρων

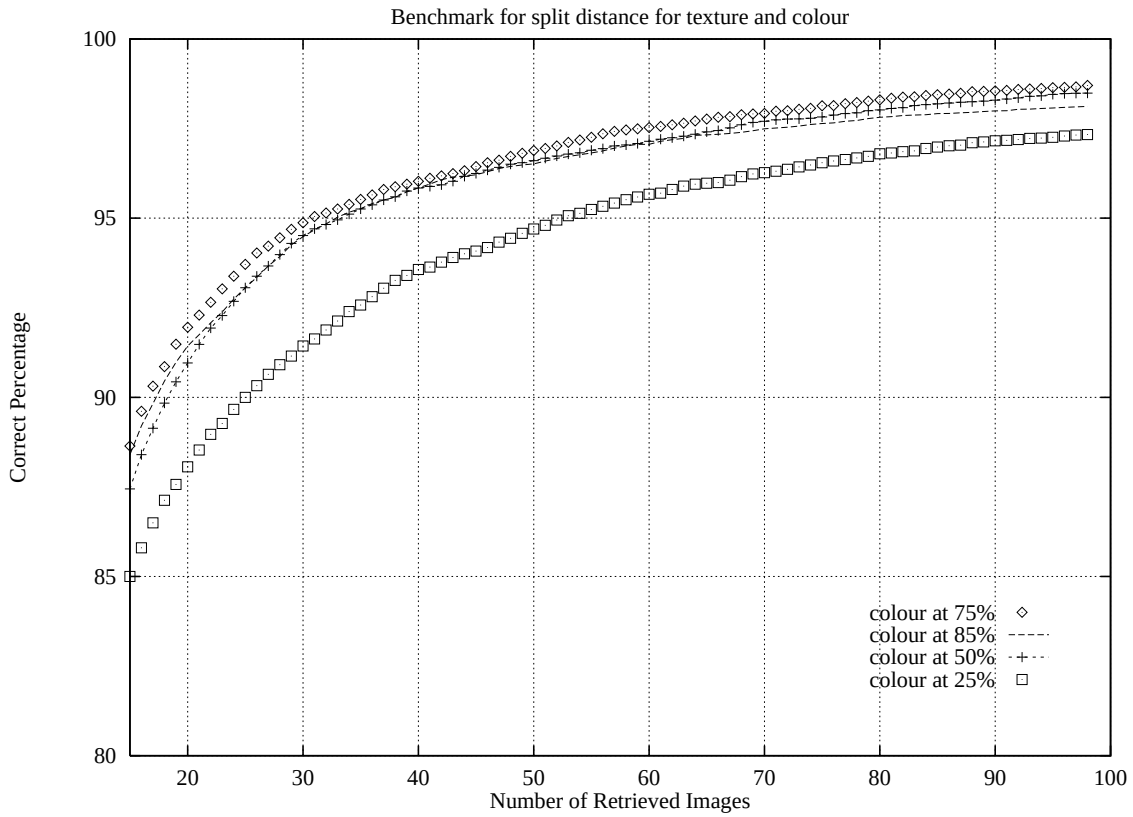
Επίσης για τον συνδυασμό του χρώματος και της υφής μπορεί να χρησιμοποιηθεί διαφορετική απόσταση διαφορετικών χαρακτηριστικών, και να γίνει ο συνδυασμός με βάση τις κατατάξεις των επιμέρους αποστάσεων με χρήση δεδομένης ποσόστωσης. Πιο

αναλυτικά χρησιμοποιούμε για να κατατάξουμε την εικόνα υπολογίζοντας αρχικά τις τιμές και των δύο αποστάσεων με βάση το χρώμα και με βάση την υφή. Ταξινομούμε τις τιμές των αποστάσεων ξεχωριστά με βάση την μικρότερη. Μετά υπολογίζουμε μια νέα τιμή-απόσταση από την σειρά κατάταξης που βρίσκεται μια εικόνα σε κάθε απόσταση:

$$\mathcal{D}(i,j) = per_t \times \mathcal{O}(dist_t(i,j)) + per_c \times \mathcal{O}(dist_c(i,j)) \quad (6.1)$$

όπου $\mathcal{O}(x)$ είναι η σειρά του στοιχείου x αν ταξινομηθούν όλες οι τιμές ως προς την μικρότερη, όπου $dist_c(i,j)$ είναι η απόσταση της i εικόνας από την j με βάση το χρώμα και $dist_t(i,j)$ η απόσταση i εικόνας από την j με βάση την υφή. Οι τιμές per_t, per_c είναι οι τιμές ποσόστωσης των δύο όρων στην σχέση, και ισχύει per_t + per_c=1. Αν θέλουμε να έχουμε ανάκληση επί ίσοις όροις με 50% για κάθε χαρακτηριστικό τότε η τιμή και των δύο παραμέτρων είναι 0.5. Με βάση αυτό το κριτήριο (6.1) γίνεται η τελική κατατάξη με βάση το χρώμα και την υφή. Στο πειράματα που ακολουθούν με βάση αυτήν προσέγγιση χρησιμοποιήθηκε DWF ανάλυση πέντε επιπέδων με Laplace κατανομή χαρακτηριστικών για την υφή και μονοδιάστατα ιστογράμματα των a,b συνιστωσών για το χρώμα. Τα ποσοστά για την υφή που χρησιμοποιήθηκαν είναι 0%, 25%, 50% 75% 85% των οποίων οι καμπύλες επιτυχίας φαίνονται στο σχήμα 6.3. Το μέγιστο ποσοστό επιτυχίας είναι 87.3% με 75% να συνεισφέρει η υφή. Πράγμα αναμενόμενο καθώς από τα προηγούμενα πειράματα φάνηκε ότι για το συγκεκριμένο σύνολο δεδομένων η υφή παίζει μεγαλύτερο ρόλο στην αναγνώριση. Βέβαια δεν κατάφερε να φτάσει το ποσοστό επιτυχίας με κοινή απόσταση, όταν χρησιμοποιήθηκαν τα ίδια χαρακτηριστικά (90.6%). Γεγονός που κάνει την ενιαία απόσταση πιο αποτελεσματική.

Ωστόσο η προσέγγιση αυτή έχει το πλεονέκτημα ότι το ποσοστό μπορεί να καθορίζεται από τον χρήστη με τι κριτήριο θέλει να ανακτήσει την εικόνα που αναζητείται. Μπορεί να θέλει αυτές που μοιάζουν περισσότερο στο χρώμα και λιγότερο στην υφή. Αυτό γίνεται πολύ εύκολα καθορίζοντας την ποσόστωση. Επίσης έχει το πλεονέκτημα ότι μπορεί να προσαρμοστεί οποιαδήποτε μέθοδος σύγκρισης με βάση το χρώμα ή την υφή με οποιαδήποτε χαρακτηριστικά. Πχ. θα μπορούσε να συγκρίνονται οι εικόνες με βάση τα κυρίαρχα χρώματά τους χρησιμοποιώντας ένα κριτήριο όπως αυτό στην προηγούμενο καεφάλαιο για ταίριασμα ομάδων κυρίων χρωμάτων, ή το συσχετόγραμμα και ούτο καθ' εξής.



Σχήμα 6.3: Καμπύλη επιτυχίας, συναρτήσει του αριθμού εικόνων που ανακλήθηκαν από την βάση, κάνοντας χρήση διαφορετικής απόστασης για χρώμα και για υφή και συνδυάζοντας τες εκ των υστέρων με ποσοστόσεις.

Κεφάλαιο 7

Εξαγωγή τμημάτων με τη χρήση χρώματος και υφής

Στο κεφάλαιο αυτό εξετάζονται τρόποι εξαγωγής τμημάτων της εικόνας, με βάση την υφή και το χρώμα, για τον χαρακτηρισμό του περιεχομένου της εικόνας. Ένας τρόπος για να γίνει αυτό είναι είναι η τμηματοποίηση χρησιμοποιώντας από κοινού χαρακτηριστικά χρώματος και υφής. Αλλά είναι δύσκολο ένας αλγόριθμος να μπορέσει να ικανοποιήσει όλες τις περιπτώσεις των εικόνων. Δηλαδή όταν δεν έχουμε περιορισμένο σύνολο από εικόνες (unconstrained). Πολλές εικόνες δεν έχει νόημα να τμηματοποιηθούν με υφή γιατί δεν περιέχουν πληροφορία που μπορεί να θεωρηθεί υφή, που είναι και συνηθισμένο. Αλλά σε άλλες περιπτώσεις η υφή παίζει πρωτεύοντα ρόλο για τον διαχωρισμό των διαφορετικών περιεχομένων της εικόνας. Ή επίσης να χρειάζεται να γίνει τμηματοποίηση χρησιμοποιώντας και τα δύο χαρακτηριστικά. Θεωρητικά η τρίτη περίπτωση είναι η πιο γενική καθώς ένα αντικείμενο χαρακτηρίζεται από τον συνδυασμό του χρώματος και της υφής του. Στο σχήμα 7.1 φαίνονται οι τρεις διακρίσεις ως προς τον χαρακτηρισμό της εικόνας. Η διάκριση αυτή μπορεί να γίνει με καθοδήγηση (supevised), γιατί είναι πολύ δύσκολο να γίνει αυτόματα η τμηματοποίηση και η επιλογή από έναν αλγόριθμο.

7.1 Εξαγωγή τμημάτων ξεχωριστά σε χρώμα και υφή

Αυτή η προσέγγιση ακολουθείται και στο QBIC ([53]) καθώς τα χαρακτηριστικά εξάγονται αυτόματα αλλά και ημιαυτόματα χωριστά σε υφή και χρώμα. Επίσης οι Ανδρούτσος, Πλατανιώτης και Βενετσανόπουλος [51], προσπάθησαν να διαχωρίσουν

το περιεχόμενο της εικόνας με βάση αν περιέχει ένα συγκεκριμένο χρώμα ή περιέχει μεταβολές στο χρώμα και στην φωτεινότητα οπότε μπορεί να θεωρηθεί υφή. Έπειτα περιεχόμενο χαρακτηρίζεται από τον επί μέρους χαρακτηρισμό των εξαχθέντων περιοχών με βάση το χρώμα και την υφή αντίστοιχα. Το κριτήριο που χρησιμοποιήθηκε ονομάζεται *δραστηριότητα χρώματος* (colour activity) και βασίζεται στο άθροισμα των γωνιών διανυσμάτων στο RGB σύστημα οι οποίες εκφράζουν απόχρωση. Το μειονέκτημα της προσέγγισης αυτής είναι ότι δεν μπορεί να χαρακτηρισθεί μια περιοχή με βάση το χρώμα και την υφή ταυτόχρονα και επίσης κάπου που έχουμε πολλές μεταβολές δεν σημαίνει αυτόματα ότι η περιοχή αυτή περιέχει υφή.

Σχήμα 7.1: Διάκριση τρόπων περιγραφής εικόνας με τμηματοποίηση με βάση την υφή και το χρώμα

7.1.1 Ανίχνευση υφής

Επίσης ένας τρόπος να περιγράψουμε την εικόνα ξεχωριστά για υφή και χρώμα είναι να χρησιμοποιηθεί ένας αλγοριθμος ανίχνευσης υφής. Το χρώμα μπορεί να περιγραφεί με τις μεθόδους που προτάθηκαν στο κεφάλαιο 5. Οι αλγόριθμοι τμηματοποίησης υφής όπως διαπιστώθηκε και στο κεφάλαιο 4 δεν μπορούν να λειτουργήσουν σωστά για όλες τις εικόνες γενικού περιεχομένου, μια που θα υπάρχει περιεχόμενο που δεν θα μπορεί να θεωρηθεί υφή. Για αυτόν τον λόγο μπορεί να χρησιμοποιηθεί ένας αλγόριθμος ανίχνευσης υφής, ο οποίος θα ανιχνεύει τις περιοχές της εικόνας που μπορούν να χαρακτηρισθούν με βάση την υφή, δηλαδή περιοχές που έχουν κάποιο επαναλαμβανόμενο πρότυπο. Έχοντας εξάγει τις περιοχές αυτές θα μπορούμε να τις τμηματοποιήσουμε με έναν αλγόριθμο τμηματοποίησης και να χαρακτηρίσουμε κάθε τμήμα.

Πρώτοι έθεσαν το ερώτημα *εάν υπάρχει υφή στην εικόνα* οι K. Karu A. Jain [54]. Στην πρώτη αυτή προσέγγιση για την ανίχνευση υφής στην εικόνα αναπτύχθηκε ένας αλγοριθμος που βασίζεται στην εξαγωγή τοπικών μεγίστων και ελαχίστων σε γραμμές και στήλες. Ένα σημείο στην εικόνα χαρακτηρίζεται σαν τοπικό ακρότατο (local

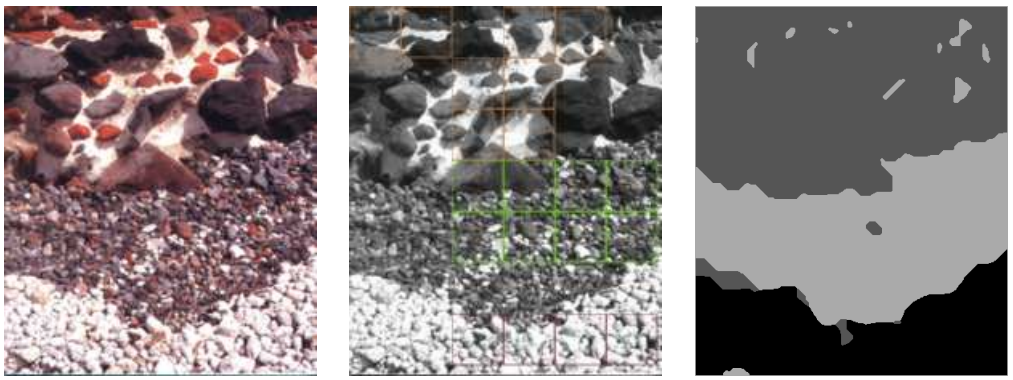
extremum) εάν, είναι ταυτόχρονα τοπικό μέγιστο ή ελάχιστο στις γραμμές και στις στήλες. Εάν ένα κομμάτι εικόνας περιέχει υφή, κρίνεται από την πυκνότητα των τοπικών ακροτάτων. Αν έχουμε πολύ μεγάλη πυκνότητα τότε είναι θόρυβος, εάν πολύ μικρή τότε μια ομοιογενής περιοχή. Στην ενδιάμεση περίπτωση έχουμε περιεχόμενο υφής. Κατά την εφαρμογή του διαπιστώθηκε ότι το κριτήριο αυτό δεν μπορεί να εντοπίσει σωστά τις περιοχές υφής. Αυτό γιατί τα ακρότατα παρέχουν το κατά πόσο ομοιογενής είναι η περιοχή. Ωστόσο με αυτόν τον τρόπο περιοχές που περιέχουν ακμές αλλά και άλλου είδους μη ομοιογενής περιοχές μπορούν να θεωρηθούν υφές. Επίσης το κριτήριο αυτό δεν παρέχει την πληροφορία το κατά πόσο δομημένη είναι η περιοχή που εξετάζεται πράγμα που χαρακτηρίζει μια υφή.

Στην εργασία αυτήν έγινε μια απόπειρα ανίχνευσης υφής με την χρήση της ανάλυσης DWF. Κάθε σημείο στην εικόνα χαρακτηρίζεται από την διασπορά των τιμών της ανάλυσης σε μια μικρή γειτονιά (π.χ. 5×5 και 9×9). Για την διάκριση του εάν ένα διάνυσμα χαρακτηριστικών είναι υφή η όχι, χρησιμοποιήθηκε ένα νευρωνικό δίκτυο ευθείας τροφοδοσίας (feed forward neural network). Για την εκπαίδευση του νευρωνικού δικτύου χρησιμοποιήθηκε ο αλγόριθμος οπίσθιας μετάδοσης (back-propagation) [55]. Μεγάλος αριθμός από σημεία υφής και αντιπροσωπευτικά μη-υφής δίνονται για εκπαίδευση. Η προσέγγιση αυτή είχε καλά αποτελέσματα για μερικές περιπτώσεις δηλαδή ανίχνευε την υφή τουλάχιστον στις περιπτώσεις που ανάλογο περιεχόμενο είχε δοθεί και στην διαδικασία εκπαίδευσης. Όποτε η διαδικασία αυτή απαιτεί πολύ καλό καθορισμένο σύνολο δεδομένων εκμάθησης του νευρωνικού δικτύου. Ωστόσο δεν αποφεύγονται οι λάθος ανιχνεύσεις, που είναι περιοχές που περιέχουν ακμές. Αυτό γίνεται γιατί τα χαρακτηριστικά έχουν επικάλυψη, πράγμα που κάνει την ανίχνευση προβληματική. Σε αυτό συμβάλλει οι επιλογή της μικρής περιοχής εξαγωγής των χακτηριστικών και η χρήση φίλτρων που δεν είναι τοπικού χαρακτήρα γιατί χρησιμοποιούν πληροφορία από σημεία πολύ μακριά από την περιοχή ενδιαφέροντος. Σε μια πρόσφατη εργασία των Malik και Belongie γίνεται εξαγωγή τμημάτων από φυσικές σκηνές κάνοντας χρήση χαρακτηριστικών υφής και ακμών, με ικανοποιητικά αποτελέσματα [56].

7.2 Τμηματοποίηση με χρώμα και υφή

Για την τμηματοποίηση εικόνων με βάση το χρώμα και την υφή χρησιμοποιήθηκε ο αλγόριθμος που παρουσιάστηκε για την υφή στο κεφάλαιο 4, ο οποίος επεκτάθηκε και για το χρώμα. Η επέκταση είναι μια πρώτη προσέγγιση του προβλήματος και γίνεται με

την προσθήκη στο διάνυσμα χαρακτηριστικών του αλγορίθμου δύο επιπλέον συνιστωσών. Η συνιστώσες αυτές πρέπει να χαρακτηρίζουν την απόχρωση καθώς η φωτεινότητα συμπεριλαμβάνεται ήδη με την χρήση της προσέγγισης. Για τον λόγο αυτόν χρησιμοποιήθηκαν οι δύο χρωματικές συνιστώσες του Lab συστήματος. Για να γίνει αυτή η προσθήκη έγιναν δύο πολύ σημαντικές παραδοχές οι οποίες θα σχολιαστούν στην συνέχεια. Η πρώτη είναι ότι θεωρούμε πως οι χρωματικές συνιστώσες, όπως άλλωστε και οι συνιστώσες της DWF ανάλυσης, ακολουθούν κατανομή Gauss. Ενώ οι δεύτερη είναι ότι όλες οι συνιστώσες είναι μεταξύ τους ασυσχέτιστες. Οι προσθήκη γίνεται σε όλα τα κριτήρια όπως και στην περίπτωση της μέσης τιμής της προσέγγισης. Για την ομοιότητα των block χρησιμοποιείται η απόσταση Bhattacharya. Ο αλγόριθμος εφαρμόστηκε στις ίδιες εικόνες με αυτές που χρησιμοποιήθηκαν στα πειραματικά αποτελέσματα του κεφαλαίου 4 με φυσικές σκηνές. Στο σχήμα 7.2 φαίνεται η εικόνα με τα βότσαλα στην έγχρωμη μορφή της και η τμηματοποιημένη εικόνα. Το αποτέλεσμα είναι καλύτερο από αυτό χωρίς χρώμα (4.10). Αυτό οφείλεται στο ότι οι συνιστώσες a,b ακολουθούν



Σχήμα 7.2: Η αρχική εικόνα και οι τελική τμηματοποιημένη, με 3 κλάσεις για την εικόνα με τα βότσαλα κάνοντας και χρήση χρώματος

κατανομή κοντά σε Gauss. Οπότε οι μέσες τιμές και οι διασπορές των χρωματικών συνιστωσών συνεισφέρουν στην διόρθωσή των λάθος κατατάξεων που δημιουργούνται με την χρήση μόνο των χαρακτηριστικών υφής. Όμως δεν συναντάται συχνά να έχουμε κατανομή Gauss στις χρωματικές συνιστώσες. Το ίδιο πρόβλημα απαντάται και στην χρήση την προσέγγισης της DWF ανάλυσης. Π.χ. μία συνεκτική σημασιολογικά περιοχή μπορεί να περιέχει διάφορες αποχρώσεις μακριά μεταξύ τους κάνοντας την κατανομή του να μπορεί να είναι οτιδήποτε. Οπότε πολλές φορές τα δεδομένα δεν θα ταιριάζουν στην μοντελοποίηση του προβλήματος όπως έχει. Το γεγονός αυτό επαληθεύεται από την εφαρμογή του αλγορίθμου στην εικόνα που περιέχει την πόλη του San Francisco, σχήμα 4.8. Η περιοχή της πόλης είναι πολύ ανομοιογενής έχοντας έτσι μεγάλη διασπορά

στις χρωματικές συνιστώσες, οι οποίες δεν ακολουθούν κατανομή Gauss. Αυτό έχει σαν αποτέλεσμα η κλάση που αντιπροσωπεύει την πόλη να παίρνει τα περισσότερα σημεία της κλάσης που αντιπροσωπεύει τον ουρανό καθώς η πρώτη κλάση έχει πολύ μεγάλη διασπορά σε σχέση με την άλλη. Επομένως το κριτήριο δεν ταιριάζει στις εικόνες που θέλουμε να τμηματοποιήσουμε.

Σαν μελλοντική προέκταση της προσπάθειας αυτής πρέπει να βρεθούν χαρακτηριστικά που να περιγράφουν την κατανομή της χρωματικής πληροφορίας. Αυτό μπορεί να γίνει χρησιμοποιώντας εξαγωγή κυρίων χρωμάτων ή την χρησιμοποίηση τοπικών ιστογραμμάτων [53]. Κάθε σημείο θα περιγράφεται από την τοπική κατανομή των χρωμάτων σε μια προκαθορισμένη γειτονιά. Αυτό βέβαια έχει το μειονέκτημα ότι εισάγει ένα υπολογιστικό κόστος το οποίο πρέπει να αντιμετωπισθεί. Επίσης για την αυτόματη εφαρμογή του αλγορίθμου πρέπει να αναπτυχθεί ένας αλγόριθμος για την εύρεση των διαφορετικών τμημάτων στην εικόνα.

Στο Berkley οι Malik, Carson, Belongie, και Greenspan για την ανάπτυξη ενός συστήματος ευρετηριασμού και ανάκλησης εικόνων πρότειναν την αναπαράσταση Blobword εικόνων [26]. Η αναπαράσταση αυτή βασίζεται σε ένα αυτόματο αλγόριθμο που βασίζεται σε χαρακτηριστικά υφής και χρώματος. Ο αλγόριθμος κάνει χρήση της EM (expectation maximization) ομαδοποίηση για αποσύνθεσης κατανομών. Και εδώ γίνεται η παραδοχή ότι τα χαρακτηριστικά ακολουθούν κατανομή Gauss. Βέβαια το σύνολο δεδομένων που χρησιμοποιούνται είναι περιορισμένο, δηλαδή στο σύνολο αυτό δεν περιέχονται εικόνες γενικού περιεχομένου. Για την αποφυγή συνιστωσών που δεν ακολουθούν κατανομή Gauss, όπως η φωτεινότητα και οι χρωματικές συνιστώσες, η εικόνα λυαίνεται με κάποιο κατωδιαβάτο φίλτρο. Π.χ. στην περίπτωση μιας εικόνας που περιέχει ζέβρα οι ρίγες χάνονται με την λύανση οπότε η κατανομή των φωτεινοτήτων πάνω στην ζέβρα ”πηγαίνει” προς την Gauss. Επίσης χρησιμοιούνται μόνο τρία χαρακτηριστικά για την υφή, ανισοτροπία , πολικότητα και αντίθεση. Ένας λόγος που χρησιμοποιούνται τόσα λίγα χαρακτηριστικά είναι λόγω του μεγάλου κόστους στην αντιστροφή του πίνακα συνδιασποράς στον EM.

Κεφάλαιο 8

Επίλογος - Συμπεράσματα

Στην εργασία αυτή παρουσιάστηκαν τρόποι εξαγωγής χαρακτηριστικών υφής και χρώματος και η εφαρμογή αυτών σε ανάκληση και τμηματοποίησης εικόνων.

Όσον αφορά την υφή παρουσιάστηκαν και συγκρίθηκαν εκπρόσωποι των δύο κυριότερων μεθόδων για την ανάλυση και χαρακτηρισμό της υφής. Εξετάστηκαν η DWF τα Gabor φίλτρα από την κατηγορία του πολυκαναλικού φιλτραρίσματος, και το MR-SAR μοντέλο που ανήκει στην κατηγορία των τυχαίων πεδίων. Με την περάτωση εκτεταμένων πειραμάτων διαπιστώθηκε ότι η πρώτη κατηγορία μεθόδων υπερτερεί. Η DWF ανάλυση και τα φίλτρα Gabor είχαν καλύτερη επίδοση από την MR-SAR προσέγγιση. Τα πειράματα έδειξαν ότι οι μέθοδοι με πολυκαναλικό φιλτράρισμα έχουν καλύτερη επίδοση στις εικόνες που περιέχουν δομημένες υφές. Ενώ με την χρήση MR-SAR έχουμε καλύτερο ποσοστό επιτυχίας για τις μη δομημένες υφές, δηλαδή αυτές που έχουν ακανόνιστη δομή.

Αναπτύχθηκε ένας αλγόριθμος τμηματοποίησης κάνοντας χρήση της DWF ανάλυσης. Ο αλγόριθμος είναι ημιαυτόματος καθώς χρειάζεται τον αριθμό των διαφορετικών περιεχομένων στην εικόνα. Ο αλγόριθμος εφαρμόστηκε με ικανοποιητικά αποτελέσματα σε εικόνες που περιέχουν υφή. Εκτός από συνθετικές εικόνες που περιέχουν υφή, έγινε εφαρμογή και σε φυσικές σκηνές που έχουν έστω και ανομοιογενές περιεχόμενο υφής. Τα αποτελέσματα ήταν ικανοποιητικά αρκεί η εικόνα που τμηματοποιείται να έχει περιοχές με κάποιο επαναλαμβανόμενο (περιοδικό) πρότυπο που μπορεί να θεωρηθεί υφή.

Όσον αφορά τα χρωματικά χαρακτηριστικά υλοποιήθηκε ένα αλγόριθμος εξαγωγής κυρίων χρωμάτων εικόνων. Ο αλγόριθμος αυτός κάνει χρήση της K-means ομαδοποίησης. Γίνεται χρήση των HSV (μετασχηματισμένο) και Lab όπου και συγκρίνονται. Από τα πειράματα διαπιστώθηκε ότι το μετασχηματισμένο HSV λαμβάνει υπόψην του περισσότερο την φωτεινότητα. Αυτό έχει σαν αποτέσμα να διασπάσει περιοχές που

έχουν ομοιόμορφες αποχρώσεις, σε αντίθεση με το Lab που παίζει μεγάλο η απόχρωση από την φωτεινότητα. Αλλά το μετασχηματισμένο HSV έχει το πλεονέκτημα ότι ενώνει διαφορετικές αποχρώσεις σε χαμηλές φωτεινότητες, το οποίο είναι συνεπές με την αντίληψη του ανθρώπου καθώς όσο πιο χαμηλός ο φωτισμός τόσο δύσκολα αντιλαμβάνεται διαφορές στην απόχρωση. Σε αντίθεση το Lab δεν ομαδοποιεί διαφορετικές αποχρώσεις σε χαμηλές φωτεινότητες. Επίσης παρουσιάζονται τρόποι χαρακτηρισμού του χρωματικού περιεχόμενο με βάση την εξαγωγή κυρίων χρώματων, καθώς και εξισώσεις αποστάσεων για την σύγκριση χρωμάτων και χαρακτηριστικών αυτών.

Επίσης διενεργήθηκαν πειράματα για κατηγοριοποίηση εικόνων με τον συνδυασμό χαρακτηρισμό υφής και χρώματος. Οι εικόνες που χρησιμοποιήθηκαν περιέχουν υφές από έγχρωμες φυσικές σκηνές. Από τα αποτελέσματα, για αυτό το σύνολο δεδομένων, διαπιστώθηκε ότι η υφή παίζει σημαντικότερο ρόλο στην αναγνώριση. Ωστόσο το ποσοστό επιτυχίας αυξήθηκε σημαντικά με την προσθήκη του χρώματος, χρησιμοποιώντας και τον συνδυασμό και των δυο χαρακτηριστικών.

Επίσης παρουσιάστηκαν προσεγγίσεις για τμηματοποίηση έγχρωμων εικόνων με χρήση χρώματος και υφής. Παρουσιάζεται μια πρώτη προσέγγιση επεκτείνοντας των αλγόριθμο που προτάθηκε για τμηματοποίηση υφής. Εξετάζονται οι περιορισμοί που έχει όσον αφορά τις εικόνες που μπορεί να τμηματοποίησει, γιατί η κατανομή του χρώματος σε κάθε τμήμα μοντελοποιείται με την κατανομή Gauss.

8.1 Μελλοντική εργασία

Όσον αφορά τον χαρακτηρισμό υφής θα μπορούσε να εξετασθεί ένας τρόπο συνδυασμού του DWF και MR-SAR. Η DWF ανάλυση έχει καλύτερα αποτελέσματα για δομημένες υφές ενώ η MR-SAR καλύτερα για μη δομημένες, οπότε αν και το ποσοστό αναγνώρισης είναι ικανοποιητικό, μπορεί να βελτιωθεί με τον συνδυασμό των δύο μεθόδων ανά περίπτωση. Ο συνδυασμός μπορεί να γίνει με τον τρόπο που προτείνει η Picard για συνδυασμό Wold και MR-SAR [20]. Πρέπει να βρεθεί όμως ένα κριτήριο που να διαχωρίζει τις δομημένες και τις μη δομημένες υφές. Επίσης αξίζει να εξεταστεί η ομοιότητα υφών ανεξαρτήτως κλίμακας. Δηλαδή να εξαχθούν χαρακτηριστικά που να είναι ανεξάρτητα με την κλίμακα. Αυτό γιατί μπορεί σε ορισμένες περιπτώσεις να θέλαμε μια ζέβρα να την ομαδοποιούσμε με μια άλλη η οποία να βρίσκεται λίγο πιο μακριά στην εικόνα οπότε αλλάζει αυτομάτως κλίμακα το περιεχόμενό της σε συχνότητες.

Όσον αφορά την τμηματοποίηση με υφή και χρώμα πρέπει να ευρεθεί ένας

αλγόριθμος εξαγωγής του αριθμού των διαφορετικών περιεχομένων στην εικόνα. Επίσης μπορεί να γίνει η προσθήκη διαφορετικής μοντελοποίησης της προσέγγισης και των χρωματικών συνιστωσών, καθώς δεν ακολουθούν πάντα κατανομή Gauss. Μια πιθανή προσέγγιση είναι η χρήση τοπικών ιστογραμμάτων. Επίσης θα άξιζε να χρησιμοποιειθεί για την εξαγωγή κυρίων χρωμάτων το Lab σε συνδιασμό με το HSV. Στην περίπτωση που έχουμε χαμηλές φωτεινότητες τα χρώματα να ενωποιούνται εκ των υστέρων με το HSV. Τέλος πρέπει να επεκταθεί η κατηγοριοποίηση για μη περιορισμένο σύνολο εικόνων, κάνοντας χρήση των κυρίων χρωμάτων και των χαρακτηριστικών υφής.

Παράρτημα Α

Αλγόριθμος χαλάρωσης ετικετών

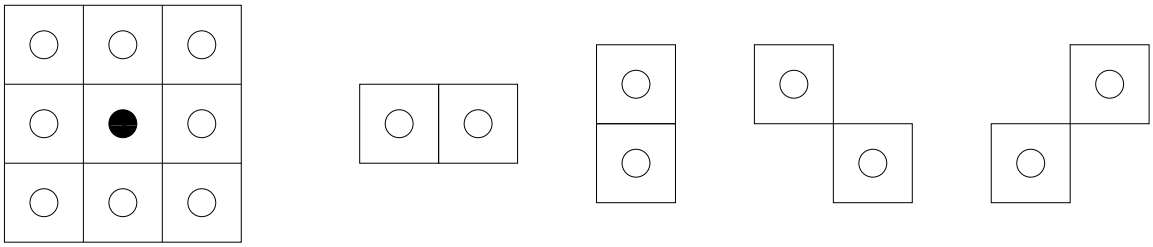
Για την διόρθωση των λάθος κατατάξεων που προκαλούνται από το στατιστικό κριτήριο, γίνεται χρήση ενός αλγόριθμου χαλάρωσης ετικετών. Ο αλγόριθμος αυτός είναι μια Bays προσέγγιση η οποία βασίζεται στο μοντέλο τυχαίων πεδίων Markov (MRF) των ετικετών.

Πολλά προβλήματα σε ανάλυση εικόνων μπορούν να μοντελοποιηθούν με την ανάθεση ετικετών χρησιμοποιώντας συμφραζόμενη πληροφορία. Σε αυτό το στατιστικό μοντέλο έχουμε:

- ένα σύνολο από θέσεις της εικόνας S = (k,l);
- ένα σύνολο από πιθανές ετικέτες για κάθε θέση L = $l_1, l_2, \ldots, l_q$;
- ένα σύνολο από πιθανές παρατηρήσεις Y = y(k,l), για κάθε θέση.
- μια σχέση γειτνίασης, G, στις θέσεις.

Το έγκειται στην ανάθεση ετικετών για κάθε θέση, με τέτοιο τρόπο ώστε η λύση να είναι συνεπής με τις παρατηρήσεις αλλά και με του περιορισμούς που εισάγουν οι σχέσεις γειτνίασης.

Έστω c μια παρέα (κλίκα) της γειτονιάς G, σχήμα Α.1. Μια καθολική διακριτή



Σχήμα Α.1: Γειτονιά δεύτερης τάξης και οι πιθανές επιλογές παρεών

ανάθεση ετικετών ω αναθέτει μια ετικέτα $\omega(k,l) \in$ L. Ένα MRF [57] ορίζεται από τα δυναμικά $V_c(\omega)$, για κάθε δυνατή παρά c και για κάθε δυνατή ανάθεση ετικετών. Ακολουθώντας το θεώρημα *Hammersley-Clifford* και την αντιστοιχία μεταξύ των των MRF και των κατανομών Gibbs [58], η πιθανότητα μιας καθολικής ανάθεσης ετικετών δίνεται από την παρακάτω σχέση [57]:

$$P(\omega) \stackrel{\triangle}{=} \frac{1}{Z} e^{-\frac{1}{T}U(\omega)} \tag{A.1}$$

όπου T είναι μια σταθερά κανονικοποίησης, και

$$U(\omega) \stackrel{\triangle}{=} \sum_{c \in C} V_c(\omega) \tag{A.2}$$

$$Z = \sum_{\omega} e^{-\frac{1}{T}U(\omega)} \tag{A.3}$$

Στην παραπάνω σχέση, που ονομάζεται και συνάρτηση διαχωρισμού, το C δηλώνει το σύνολο όλων των συνδεδεμένων παρεών λαμβάνοντας υπόψην το ορισμό της γειτονιάς G, και Z είναι μια σταθερά κανονικοποίησης. Μιλώντας με στατιστικούς όρους, U είναι η συνάρτηση κόστους (ή ενέργειας) του συστήματος, ενώ $V_c()$ ονομάζεται συνάρτηση δυναμικού, και αντιστοιχεί στην συνεισφορά των τοπικών αλληλεπιδράσεις στην καθολική ενέργεια του συστήματος. Ένα σημαντικό θέμα στην στην διεργασία αυτή είναι να εκμεταλλευτούμε την γνώση εκ των πρωτερών με την τις διαθέσιμες παρατηρήσεις, έτσι ώστε να δημιουργηθεί μια νέα μορφή της συνάρτησης κόστους (ή ενέργειας). Αυτή η μορφή είναι ένας συνδυασμός μεταξύ των χωρικών ιδιοτήτων του πεδίου ετικετών και της καταλληλότητας των παρατηρήσεων, σε σχέση με τις ετικέτες. Ακολουθώντας την υπόθεση η από κοινού πυκνότητα πιθανότητας του πεδίου των ετικετών και του πεδίου των των χαρακτηριστικών της υφής είναι ανάλογη με το ακόλουθο:

$$P(\omega) \prod_{(k,l) \in S} e^{-\delta(y(k,l);\omega(k,l))} \tag{A.4}$$

όπου η καταλληλότητα ενός διανύσματος χαρακτηριστικών δίνεται από την εξίσωση 4.4. Η μεγιστοποίηση της εκ των υστέρων πιθανότητας είναι ισοδύναμη με με την μεγιστοποίηση της από κοινού πυκνότητάς πιθανότητας, και οδηγεί στην ελαχιστοποίηση της συνάρτησης κόστους που δίνεται από την σχέση

$$U(\omega) \stackrel{\triangle}{=} \frac{1}{T} \sum_{c \in C} V_c(\omega) + \sum_{(k,l) \in S} \delta(\omega(k,l), y(k,l)) \tag{A.5}$$

$$\stackrel{\triangle}{=} U_1(\omega) + U_2(\omega, y) \tag{A.6}$$

όπου ο όρος $\delta()$ εκφράζει την απαίτηση της καταλληλότητας. Για την ανάθεση των ετικετών στην περίπτωση της τμηματοποίησης υφής ο δεύτερος όρος αντικαθίσταται από άθροισμα των αποστάσεων από μια κλάση (δηλαδή για μια ετικέτα) σε όλα τα σημεία της εικόνας.

Μια κοινή τεχνική για την λύση τέτοιων προβλημάτων είναι η χρήση του κριτηρίου της μεγιστοποίησης της εκ των υστέρων πιθανότητας, δηλαδή της μεγιστοποίησης της εκ των υστέρων κατανομής των ετικετών, δεδομένου των παρατηρήσεων, το οποίο είναι ισοδύναμο με την ελαχιστοποίηση της συνάρτησης κόστους. Η ελαχιστοποίηση γίνεται με έναν καθοριστικό (deterministic) αλγόριθμο, επαναλαμβανομένων τοπικών επικρατουσών τιμών ICM (em Iterated Conditional Models)[59].

Βιβλιογραφία

[1] V. N. Gudivada V. V. Raghavan. Content based image retrieval systems. *IEEE Computer*, 28, Sep 1995.

[2] Special issue on digital libraries. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 18(8), Aug 1996.

[3] V. Ogle M. Stonebraker. Chabot: Retrieval from a relational database of images. *IEEE Computer*, 28(9):40--48, Sep 1995.

[4] W. Niblack J Ashley M. Flickner, H. Sawhney. Query byimage and video content: the qbic system. *IEEE Computer*, 28(9):23--32, Sep 1995.

[5] A. Gupta R. Jain. Visual information retrieval. *Comm. Assoc. Com. Mach.*, 40(5):70--79, May 1997.

[6] A. Pentland, R. W. Picard, and S. Sclaroff. Photobook: Content-based manipulation of image databases. *M.I.T. Media Laboratory Perceptual Computing Technical Report No. 255*, November 1993.

[7] P.C. Chen and T. Pavlidis. Segmentation by texture using correlation. *IEEE Transactions on PAMI*, 5:64--69, January 1983.

[8] R.L. Kashyap, R. Chellappa, and A. Khotanzad. Texture classification using features derived form random field models. *Pattern Recognition Letters*, 1:43--50, 1982.

[9] M. Unser. Local linear transforms for texture measurements. *Signal Processing*, 11:61--79, July 1986.

[10] M. Porat and Y. Y. Zeevi. Localized texture processing in vision: Analysis and synthesis in gaborian space. *IEEE Trans. Biomed. Eng.*, 36:115--129, 1989.

[11] A.C. Bovic, M. Clark, and W.S. Geisler. Multichannel texture analysis using localized spatial filters. *IEEE Transactions on PAMI*, 12:55--73, January 1990.

[12] S. G. Malat. A theory of multiresolution signal decomposition: The wavelet representation. *IEEE Trans. Patt. Anal. Machine Intell.*, 11:674--693, January 1989.

[13] O. Rioul and M. Vetterli. Wavelets and signal processing. *IEEE Signal Processing Mag.*, 8:11--38, October 1991.

[14] J. Randen J.H. Husey. Filtering for texture classification: a comparetive study. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 21(4):837--842, Apr 1999.

[15] M. Unser. Texture classification and segmentation using wavelet frames. *IEEE Trans. on Image Processing*, 4:1549--1560, November 1995.

[16] A. K. Jain F. Farrokhnia. Unsupervised segmentation using gabor filters. *Pattern Recognition*, 24(12):1167--1186, Aug 1991.

[17] B.S. Manjunath W.Y. Ma. Texture features for browsing and retrival of image data. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 18(8):837--842, Aug 1996.

[18] J. Mao A. Jain. Texture classification and segmentation using multiresolution simultaneous autoregressive models. *Pattern Recognition*, 25(2):173--188, 1992.

[19] T. Young K.S. Fu. *Handbook of pattern recognition and image processing*. Acedemic Press, 1986.

[20] F. Liu R. W. Picard. Periodicity directionality and randomness wold features for image modeling retrieval. *Pattern Recognition*, 18(7):722--733, Jul 1996.

[21] VisTex. http://www-white.media.mit.edu/vismod/imagery/visiontexture/vistex.html. In *MIT Media Lab.*

[22] P Brodatz. *A Photografic Album for Artists and Designers1*. Dover, New York, 1966.

[23] S. Liapis, N. Alvertos, and G. Tziritas. Maximum likelihood texture classification and bayesian texture segmentation using discrete wavelet frames. *International Conference in Digital Signal Processing DSP97*, 2:1107--1110, July 1997.

[24] A. Jain Kalle Karu. Learning texture discrimination masks. *IEEE Pattern Analysis and Machine Intelligence*, 18(2):195--205, Feb 1996.

[25] M. L. Corner E. J Delp. Segmentation of textured image using a multiresolution gaussian autoregressive model. *IEEE Trans. on Image Processing*, 8(3):408--420, Mar 1999.

[26] S. Belongie C. Carson H. Greenspan J. Malik. Color- and texture-based image segmentation using em and its application to content based image retrieval. In *Int. Conf. Comp. Vis.*, 1998.

[27] N. Paragios R. Deriche. Geodesic active regions for supervised texture segmentation. In *ICCV*, volume II, pages 926--932, Corfu, Greece, Sep 1999.

[28] N. Alvertos G. Tziritas S. Liapis. Unsupervised texture segmentation using discrete wavelet frames. In *EUSIPCO*, Rhodes, Greece, Sep 1998.

[29] R. Duda and P. Hart. *Pattern Classification and Scene Analysis.* New York: J. Wiley & Sons, 1973.

[30] Νικόλαος Κ. Παραγυιός. *Δυναμική ανίχνευση καιεντοπισμός κινούμενων αντικειμένων σε ακολουθίες εικόνων.* Μεταπτυχιακή εργασία, Πανεπιστήμιο Κρήτης τμήμα Επιστήμης Υπολογιστών, Ηράκλειο, 1996.

[31] E. Sifakis and G. Tziritas. Fast marching to moving object location. *International Conference on Scale Space Theories in Computer Vision*, 1999.

[32] R. Adams and L. Bischof. Seeded region growing. *Pattern Analysis and Machine Intelligence*, 16, June 1994.

[33] G. A. Agoston. *Color theory and its application in art and design.* Springer-Verlag, Berlin Heidelberg, 1979.

[34] W. Ostwald. *Color Science.* Windsor and Newton, London, 1931.

[35] W. Ostwald. *The Color Primer.* Van Nostrand-Reinhold, New York, 1969.

[36] Munsell Color Company. *The Munsell Book of Color.* Munsell Color Co., Baltimore MD, 1976.

[37] A. Jain. *Fundamentals of Digital Image Processing.* Prentice Hall, 1989.

[38] H. Levkowitz G.T. Herman. Glhs: A generalized lightness hue and saturation color model. In *GMIP*, volume 55(4), pages 271--, Jul 1993.

[39] Commition Internationale de l'Eclairage (CIE). Cie recommmendations on uniform color spaces - color difference equations, psychometric color terms. *CIE Publication 15, E-1.3.1(1971)/TC-1.3, Supplement*, 2:9--12, 1978.

[40] A. Marzouki W. Pieczynski Y. Delignon. Esimation of generilized mixtures and its application in image segmention. *IEEE Trans. on Image Processing*, 6(10):1364--1373, Oct 1997.

[41] W. Pieczynski A. Peng. Adaptive mixture esimation and usupervised local bayesian image segmention. *IEEE Trans. on Image Processing*, 6(10):1364--1373, Oct 1997.

[42] C.Y. Low S.W. Smoliar J.H. Wu H.J Zhang. Video parsing, retrieval and browsing: An intergrated and content-based solution. In *ACM Multimedia*, 1995.

[43] M. Swain and D. Ballard. Color indexing. *Int. J. Comp Vis.*, 7(1):11--32, 1991.

[44] M. Mitra W.-J. Zhu J. Huang, S.R. Kumar and R. Zabih. Image indexing using color correlograms. *IEEE Conference on Computer Vision and Pattern Recognition*, pages 762--768, 1997.

[45] L.L. Wang S.N. Wang H.C. Lin. Color image retrieval based on hidden markov models. *IEEE Trans. on Pattern Analysis Machine Intell.*, 6(2):332--339, Feb 1997.

[46] J. R. Smith S. F. Chang. Single color extraction and image query. *IEEE Int Conf. on Image Processing*, pages 528--531, 1995.

[47] Ronfard R. Tziritas G. Veneau E. Bouthemy P., Garcia C. and Zugaj D. Scene segmentation and image feature extraction in the divan video indexing and retrieval architecture. *in Proc. of the Third International Conference on Visual Information Systems*, pages 244--252, June 1999.

[48] Ηλίας Γκρίνιας. *Ανάλυση περιεχομένου Video.* Μεταπτυχιακή εργασία, Πανεπιστήμιο Κρήτης τμήμα Επιστήμης Υπολογιστών, Ηράκλειο, 1999.

[49] C. Garcia and G. Tziritas. Face detection using quantized skin color regions merging and wavelet packet analysis. *IEEE Transactions on Multimedia*, 1:264--277, September 1999.

[50] W. Equitz M. Flickner W. Niblack J. Hafner, H. Sawhney. Efficient color histogram indexing for quadratic form distance functions. *IEEE Trans Pattern Analysis Machine Intell.*, 17(7):729--735, Jul 1995.

[51] K.N. Plataniotis A.N Venetsanopoulos D. Androutsos. Image region extraction for content-based image retrieval. In *EUSIPCO*, Rhodes, Greece, Sep 1998.

[52] J.R. Smith. Integrated spatial and feature image systems: Retrieval, analysis and compression. *Graduate School of Arts and Sciences, Columbia University, Phd Thesis*, 1997.

[53] M.D. Flickner J.L. Hafner D. Lee W. Niblack D. Petkovic J. Ashley, R. Barber. Automatic and semiautomatic methods for image annotaion and retrieval in QBIC. In *Proceedings of IST/SPIE Conference on Storage and Retrieval for Image and Video Databases III*, volume 2420, pages 24--35, 1995.

[54] Anil K. Jain Ruud M. Bolle Kalle Karu. Is there any texture in the image? *Pattern Recognition*, 29(9):1437--1446, 1996.

[55] Simon Haykin. *Neural Networks.* Prentice Hall, 1999.

[56] Serge Belongie Jitendra Malik. Finding boundaries in natural images: A new method using point descriptors and area completion. *ECCV*, I:751--766, Jun 1998.

[57] S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Trans. on Pattern Analysis Machine Inteligence*, 6:721--741, 1984.

[58] J. Besag. Spatial interaction and the statistical analysis of latice systems (with duscussion). *Journal of Royal Statistics Society*, B:36:192--326, 1974.

[59] J. Besag. On the statistical analysis of dirty images. *Journal of Royal Statistics Society*, 48:259--302, 1986.