Πανεπιστήμιο Κρήτης
Σχολή Θετικών Επιστημών
Τμήμα Επιστήμης Υπολογιστών

Ανάλυση Περιεχομένου Video

Ηλίας Κ. Γκρίνιας

Μεταπτυχιακή Εργασία

Ηράκλειο, Σεπτέμβριος 1999

© Copyright 1999 by Elias K. Grinias
All Rights Reserved

Πανεπιστήμιο Κρήτης
Σχολή Θετικών Επιστημών
Τμήμα Επιστήμης Υπολογιστών

Ανάλυση Περιεχομένου Video

Εργασία που υποβλήθηκε από τον
Ηλία Κ. Γκρίνια
ως μερική εκπλήρωση των απαιτήσεων για την απόκτηση
ΜΕΤΑΠΤΥΧΙΑΚΟΥ ΔΙΠΛΩΜΑΤΟΣ ΕΙΔΙΚΕΥΣΗΣ

Συγγραφέας:

Ηλίας Κ. Γκρίνιας
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

Σεπτέμβριος 1999

Αφιερώνεται στη φυσική
και πνευματική μου οικογένεια

Ανάλυση Περιεχομένου Video

Ηλίας Κ. Γκρίνιας

Μεταπτυχιακή Εργασία

Τμήμα Επιστήμης Υπολογιστών
Πανεπιστήμιο Κρήτης

Περίληψη

Στις μέρες μας, είναι δεδομένη η πρόοδος στην ανάπτυξη δικτύων ευρείας μετάδοσης η οποία συνοδεύεται από την ανάπτυξη σταθμών εργασίας μεγάλης υπολογιστικής ισχύος. Η πρόοδος αυτή, δίνει πλέον τη δυνατότητα αποδοτικής επεξεργασίας της πληροφορίας των πολυμέσων, με αποτέλεσμα να αυξάνεται διαρκώς η σημασία της ανάπτυξης συστημάτων αποδοτικής διαχείρισης της οπτικής πληροφορίας. Η διαχείριση συνίσταται σε μεθόδους επεξεργασίας των δεδομένων video και εικόνων, που οδηγούν στη μείωση του απαιτούμενου χώρου αποθήκευσης και στην αποδοτική δεικτοδότηση και ανάκτηση του περιεχομένου τους. Η διαφορετική φύση της οπτικής πληροφορίας σε σχέση με άλλες πηγές ψηφιακών δεδομένων, όπως είναι το κείμενο, έχει οδηγήσει στη ανάπτυξη μεθόδων δεικτοδότησης και ανάκτησης που βασίζονται στο περιεχόμενο της οπτικής πληροφορίας.

Στην εργασία που παρουσιάζεται στη συνέχεια, περιγράφεται ένα σύστημα εξαγωγής χαρακτηριστικών βάσει του περιεχομένου του video και των εικόνων που το αποτελούν. Ο σχεδιασμός του συστήματος ακολουθεί τις τρέχουσες προσεγγίσεις αποθήκευσης/ανάκτησης και επεξεργασίας της οπτικής πληροφορίας. Στο πρώτο μέρος της εργασίας παρουσιάζεται μια μέθοδος που συνίσταται στην ταχεία τμηματοποίηση του video στις σκηνές από τις οποίες αποτελείται. Η μέθοδος εκτελείται ταυτόχρονα με τη διαδικασία αυτόματης εξαγωγής των αντιπροσωπευτικών εικόνων κάθε σκηνής με σκοπό τη συμπλήρωση των εισόδων ενός ιεραρχικού μοντέλου δεδομένων το οποίο γενικά ακολουθεί το πρότυπο μοντέλο δεικτοδότησης video MPEG7. Η μέθοδος εφαρμόστηκε σε video τα οποία έχουν συμπιεστεί σύμφωνα με το πρότυπο MPEG.

Στο δεύτερο μέρος της εργασίας παρουσιάζονται μια μέθοδος εξαγωγής αντικειμένων από εικόνες βάσει του χρώματος τους και μια μέθοδος εξαγωγής και

παρακολούθησης αντικειμένων στο χρόνο βάσει της κίνησής τους. Οι μέθοδοι βασίζονται στον αλγόριθμο τμηματοποίησης εικόνων Seeded Region Growing (SRG). Αρχικά, παρουσιάζεται η επέκταση του καθοδηγούμενου από το χρήστη αλγορίθμου SRG για την τμηματοποίηση εικόνων χρώματος. Στη συνέχεια, προτείνεται μια μέθοδος αυτόματης εξαγωγής περιοχών χρώματος η οποία επίσης στηρίζεται στον αλγόριθμο SRG. Ο αλγόριθμος εφαρμόστηκε σε εικόνες προσώπων. Τα αποτελέσματα περιλαμβάνουν τη σύγκριση των χώρων χρώματος GLHS και Lab. Τα εξαγόμενα αντικείμενα μπορούν να αποτελέσουν στη συνέχεια την είσοδο του συστήματος τμηματοποίησης και παρακολούθησης αντικειμένων βάσει της κίνησής τους.

Επόπτης: Γιώργος Τζιρίτας
Αναπληρωτής Καθηγητής Επιστήμης
Υπολογιστών
Πανεπιστήμιο Κρήτης

Video Content Analysis

Elias K. Grinias

Master of Science Thesis

Computer Science Department
University of Crete

Abstract

Nowadays, the progress in the development of broadband networks which is followed by the concurrent progress in building high-powered workstations, has increased the data manipulation capabilities. This advent, provides the ability to efficiently parse the multimedia information which in turn, increases the importance of efficient optical information management systems. The management consists of video and image manipulation methods, which facilitate the decreasing of the data storage capacity as well as the efficient indexing and retrieval of their content. The special nature of the optical information compared to other digital data sources, such as the text, has lead to the development of indexing/retrieval methods that are based on the optical information content.

In this thesis, a feature extraction system is described which is based on the content of the video and its images. The design of the system follows the current approaches on the indexing/retrieval and processing of the optical information. In the first part of the thesis, a fast video segmentation procedure to scenes, is presented. The method cooperates with a key-frame extraction procedure in order to fill the entries of a hierarchical data model, which in general, follows the video indexing data model of the standard MPEG7. The method has been applied on MPEG compressed video.

In the second part of the thesis, two object extraction methods are presented. The first one is based on the object color, while the second uses the object motion information, to segment the image. Both of them, are based on the segmentation algorithm Seeded Region Growing (SRG). First, the extension of the user-guided SRG algorithm is described, in order to be used in the color image segmentation. Then, an automatic region extraction method which is based on the color of images is proposed. The algorithm is performed on images

with faces. The results include the comparison between the systems GLHS and Lab. The extracted regions may then be used as input to the motion segmentation and tracking system.

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

Ευχαριστίες

Ξεκινώντας αυτή την ενότητα, που είναι η μόνη στην οποία μπορεί κανείς να ξεφύγει για λίγο από τον καθαρά ορθολογιστικό τρόπο γραφής, νιώθω την ανάγκη να ευχαριστήσω τους γονείς μου Κώστα και Αναστασία και τον κατά ένα χρόνο μικρότερο αδελφό μου Γιάννη, οι οποίοι με δίδαξαν τί ακριβώς σημαίνει να είναι κανείς ‘‘προνομιούχος’’, δουλεύοντας όλοι τους σκληρά, για να εξασφαλίσουν σε μένα τη δυνατότητα να ξεκινήσω και να συνεχίσω τις σπουδές μου. Θα ήθελα να αναφερθώ ιδιαίτερα στον πατέρα μου Κώστα, ο οποίος έχοντας περάσει τα παιδικά του χρόνια ορφανός στη διάρκεια της Κατοχής, με έχει βοηθήσει στη συνεχή προσπάθειά μου να διακρίνω το αληθινό από το ψεύτικο και το σημαντικό από το εφήμερο, μέσα από μακρές συζητήσεις όπου η σχέση γονιού-παιδιού αντικαθίσταται ανεπαίσθητα από τη σχέση φίλου προς φίλο.

Μιλώντας για αληθινές αξίες, θα ήθελα να ευχαριστήσω από βάθους καρδιάς, τον άνθρωπο που πίστεψε ότι είμαι ικανός να ξεκινήσω και να τελειώσω τις μεταπτυχιακές μου σπουδές. Αναφέρομαι στον κ. Γιώργο Τζιρίτα, ο οποίος έχοντας καταφέρει να συνδυάσει με επιτυχία τη σκληρή, ουσιαστική δουλειά με μια ευτυχισμένη οικογενειακή ζωή, αποτελεί για μένα κάτι πιο σημαντικό από τον επόπτη της μεταπτυχιακής μου εργασίας: ένα πρότυπο ζωής που αμφιβάλλω αν σταθώ ποτέ ικανός να φτάσω. Αφού τον ευχαριστήσω και πάλι για την ουσιαστική συμβολή του σε όλη τη διάρκεια της εργασίας, εύχομαι ό,τι καλύτερο για τον ίδιο, τη σύζυγό του και τα υπέροχα παιδιά τους.

Στο σημείο αυτό θα ήθελα να ευχαριστήσω και τους άλλους δύο καθηγητές της επιτροπής εξέτασης της μεταπτυχιακής μου εργασίας κ.κ. Απόστολο Τραγανίτη και Πάνο Τραχανιά, καταρχήν για την υπομονή που είχαν να παρακολουθήσουν την παρουσίασή της αλλά και για τις εύστοχες παρατηρήσεις και διορθώσεις τους σε αρκετά από τα θέματα που παρουσιάστηκαν.

Στη συνέχεια θα ήθελα να ευχαριστήσω, όλους όσους έμμεσα ή άμεσα συνετέλεσαν στο δύσκολο έργο της συγγραφής αυτής της εργασίας, ξεκινώντας από το συγκάτοικό μου Σταύρο Χαρίση, για την υπομονή και διακριτικότητα με την οποία με αντιμετώπισε όταν σε κάποιες στιγμές τα πράγματα δυσκόλευαν για μένα, αλλά και για την προτροπή του να ξεκινήσω τις μεταπτυχιακές μου σπουδές και την υποστήριξή του κατά τη διάρκειά τους. Το ίδιο σημαντική ήταν και η συμβολή της Αθηνάς Σαντορινιαίου η οποία συνετέλεσε επιπλέον στο να παρουσιαστεί αυτή η εργασία, προσφέροντας --μαζί

με την αμέριστη συμπαράστασή της-- το φορητό υπολογιστή της.

Σημαντικές υπήρξαν και οι στιγμές που πέρασα με τους συναδέλφους μεταπτυχιακούς φοιτητές Σπύρο Λιάπη, Μαρία Τράκα, Ελευθερία Τζόβα και Γιώργο Χαλκιαδάκη τις μέρες της συγγραφής τόσο της δικής μου όσο και των δικών τους εργασιών. Τους ευχαριστώ όλους, ιδιαίτερα για την ψυχολογική στήριξη που μου πρόσφεραν και εύχομαι να εκπληρωθούν οι στόχοι που έχουν θέσει για την πορεία τους από εδώ και στο εξής. Εύχομαι στο Σπύρο να βρεί επιτέλους αυτό που ψάχνει εναγωνίως τον τελευταίο καιρό, έστω και αν η Μαρία δε συμφωνεί με τις επιλογές του και στο Γιώργο να δει το όνειρό του για μια νέα κοινωνία να εκπληρώνεται, αν και δε νομίζω πως θα ήθελα να ζήσω σ' αυτή.

Θα ήταν νομίζω παράλειψη να μην ευχαριστήσω τους νέους αξιόλογους ανθρώπους που ο κ. Τζιρίτας έχει επιλέξει για τη στελέχωση της ομάδας εργασίας του, στην οποία ανήκω και γω. Θα ήθελα να σταθώ σε ότι αφορά την ολοκλήρωση της εργασίας στον Christophe Garcia, για την ουσιαστική συζήτηση που είχαμε κατά καιρούς πάνω στα ζητήματα της τμηματοποίησης του video και σε όλους τους άλλους για την έμπρακτη υποστήριξη τους, ιδιαίτερα από τη Θέμιδα Ζαμάνη, το Νίκο Κομοντάκη, τους νέους μεταπτυχιακούς Γιώργο Σημαντήρη και Μπάμπη Αποστολίδη, τον Ευτύχη Σηφάκη, τον Κώστα Παναγιωτάκη, το Δημήτρη Δημάκη αλλά και τον ''ξενητεμένο'' Γιώργο Τζανετάκη, τους οποίους έχω τη χαρά να γνωρίζω καλύτερα από τους υπόλοιπους.

Εύχομαι επίσης στον Ξενοφώντα Ζαμπούλη, με τον οποίο είχαμε μια σειρά ''έντονων'' συζητήσεων για τις υπάρχουσες δυνατότητες επεξεργασίας του χρώματος, να συνεχίσει με τον ίδιο ζήλο το διδακτορικό του και στον κάτοχο μεταπτυχιακού διπλώματος Αντώνη Χατζησταματίου, ο οποίος βοήθησε στην επεξεργασία ορισμένων από τις εικόνες που παρουσιάζονται στην εργασία, καλή σταδιοδρομία από δώ και στο εξής.

Τελειώνοντας θα ήθελα να ευχαριστήσω το Χρήστο (Chris) Αργύρη, τη Ράνια Καραξάνη και το Βασίλη Λύκο, για τις ατέρμονες αλλά πάντα ενδιαφέρουσες συζητήσεις μας περί ανέμων και υδάτων, οι οποίες, σε συνάρτηση βέβαια με όσα γνώρισα και πέρασα με όλους όσους αναφέρθηκαν, είχαν ως αποτέλεσμα να εδραιωθεί πλέον σταθερά η πεποίθηση μέσα μου πως...

''δεν είμαι μόνος...
όλα είναι δρόμος...''

Περιεχόμενα

Περίληψη **i**

Abstract **iii**

Ευχαριστίες **v**

Περιεχόμενα **vii**

Κατάλογος Σχημάτων **xi**

1 Γενική Εισαγωγή **1**

1.1 Κίνητρα της Εργασίας 1

1.2 Συστήματα Αποθήκευσης και Ανάκτησης Περιεχομένου Video 2

1.2.1 Προηγούμενες Προσεγγίσεις Δεικτοδότησης 2

1.2.2 Ιδιαιτερότητες της Οπτικής Πληροφορίας 3

1.2.3 Τρέχουσες Προσεγγίσεις Δεικτοδότησης 3

1.2.4 Εξαγωγή Χαρακτηριστικών Βάσεων Δεδομένων Οπτικής Πληροφορίας 5

1.3 Οργάνωση της Εργασίας 7

2 Αναπαράσταση Πληροφορίας Χρώματος **11**

2.1 Μέτρηση Χρώματος (Colorimetry) 11

2.2 Χώροι ή Συστήματα Χρώματος 13

2.2.1 Ιδιότητες Χώρων Χρώματος 14

2.2.2 Ο Χώρος Χρώματος RGB 14

2.2.3 Πρότυποι Γραμμικοί Μετασχηματισμοί Χρώματος 15

2.2.4 Χώροι Χρώματος της CIE 15

2.3 Αναπαράσταση Πληροφορίας Έντασης και Χρώματος Εικόνων 16

2.3.1 Κβαντισμός (Quantization) Χρώματος 16

2.3.2 Ιστογράμματα 17

2.3.3 Ιστογράμματα με Συνδυασμό Χρώματος και Άλλων Χαρακτηριστικών 17

2.3.4 Ομοιότητα Εικόνων βάσει της Παράστασης Κατανομής Χαρακτηριστικών Τους . . . 18

3 Τμηματοποίηση και Δεικτοδότηση του Περιεχομένου Video 21
3.1 Εισαγωγή . . . 21
3.2 Βασικές Αρχές Σχεδιασμού της Υλοποίησης . . . 22
3.3 Ιεραρχικό Μοντέλο Δεικτοδότησης Δεδομένων Video . . . 23
3.4 Κωδικοποίηση MPEG-1 . . . 24

4 Ανίχνευση Αλλαγής Σκηνών του Video 31
4.1 Εισαγωγή-Ανασκόπηση . . . 31
4.1.1 Τύποι Μεταβάσεων Σκηνών . . . 31
4.1.2 Προηγούμενες Προσεγγίσεις . . . 32
4.2 Ο Αλγόριθμος Ανίχνευσης Ασυνεχειών . . . 34
4.2.1 Σύνδεση με το Σύστημα Δεικτοδότησης του Video . . . 34
4.2.2 Μέτρηση (Αν)ομοιότητας Εικόνων του Video . . . 36
4.2.3 Σειρά Προσπέλασης των Εικόνων του Video . . . 37
4.2.4 Πειραματικά Αποτελέσματα . . . 39
4.2.5 Συμπεράσματα-Επεκτάσεις . . . 41

5 Εξαγωγή Αντιπροσωπευτικών Εικόνων 47
5.1 Εισαγωγή . . . 47
5.2 Ο Αλγόριθμος Εξαγωγής Αντιπροσωπευτικών Εικόνων . . . 48
5.2.1 Σύνδεση με τον Αλγόριθμο Ανίχνευσης Ασυνεχειών . . . 48
5.2.2 Μέθοδος Σειριακής Ομαδοποίησης . . . 49
5.2.3 Σειριακή Ομαδοποίηση Εικόνων Βάσει Ιστογραμμάτων . . . 51
5.2.4 Συμπλήρωση του Μοντέλου Δεικτοδότησης . . . 53
5.2.5 Πειραματικά Αποτελέσματα . . . 54
5.2.6 Συμπεράσματα-Επεκτάσεις . . . 56

6 Εξαγωγή Αντικειμένων του Video 59
6.1 Εισαγωγή . . . 59
6.2 Αλγόριθμοι Τμηματοποίησης . . . 60
6.3 Ο Αλγόριθμος Seeded Region Growing (SRG) . . . 61

7 Εξαγωγή Ομοιόμορφων ως προς το Χρώμα Περιοχών 67
7.1 Εισαγωγή-Ανασκόπηση . . . 67
7.2 Τμηματοποίηση Χρώματος βάσει του SRG . . . 68
7.2.1 Περιγραφή . . . 68
7.2.2 Κριτήριο Ομοιότητας Χρώματος Γραμμικών Χώρων . . . 69
7.2.3 Πειραματικά Αποτελέσματα . . . 71

7.3 Εφαρμογή: Αυτόματη Εξαγωγή Συνεκτικών Περιοχών Χρώματος 76
7.3.1 Περιγραφή 76
7.3.2 Αυτόματη Εξαγωγή Αρχικών Περιοχών 76
7.3.3 Συγχώνευση Περιοχών 78
7.3.4 Πειραματικά Αποτελέσματα-Σύγκριση GLHS και Lab στην Εξαγωγή Αρχικών Περιοχών 80
7.3.5 Συμπεράσματα-Επεκτάσεις 92

8 Παρακολούθηση Αντικειμένων Video **101**
8.1 Εισαγωγή-Ανασκόπηση 101
8.2 Τμηματοποίηση Κίνησης με Χρήση του SRG 103
8.3 Παρακολούθηση Αντικειμένων στο Χρόνο 105
8.4 Σταθεροποίηση των Συνόρων των Αντικειμένων 106
8.4.1 Εξομάλυνση Συνόρων των Αντικειμένων (Boundary Smoothing) . . 106
8.4.2 Υπολογισμός ‘‘Μέσου’’ Σχήματος (Shape Averaging) 107
8.5 Πειραματικά Αποτελέσματα 107
8.6 Συμπεράσματα-Επεκτάσεις 111

A Μετασχηματισμοί του Χώρου Χρώματος RGB **113**
A.1 Μετασχηματισμός RGB σε YIQ 113
A.2 Μετασχηματισμός RGB σε YCbCr 113
A.3 Μετασχηματισμός RGB σε XYZ 113
A.4 Μετασχηματισμός RGB σε Lab 114
A.5 Μετασχηματισμός RGB σε Luv 114

Βιβλιογραφία **115**

Κατάλογος Σχημάτων

1.1 Αποθήκευση και ανάκτηση εικόνων και video. 8
1.2 Εξαγωγή και αποθήκευση βάσει του περιεχομένου της οπτικής πληροφορίας 9

2.1 Ορατό φάσμα φωτός των χρωμάτων. (Πηγή: http://www.linocolor.www) . 12
2.2 Αντιληπτική αναπαράσταση του χώρου χρωμάτων 13
2.3 Χρωματικό διάγραμμα του χώρου XYZ. (Πηγή: http://www.linocolor.com) 19
2.4 Χρωματικό διάγραμμα του χώρου Lab. (Πηγή: http://www.linocolor.com) . 19
2.5 Ο χώρος χρώματος Lab (Πηγή: http://www.linocolor.www) 19

3.1 Το Μοντέλο Δεικτοδότησης του Video 24
3.2 Κωδικοποίηση I-εικόνων (Πηγή: http://fas.sfu.ca/cs/undergrad/CourseMaterials) 26
3.3 Κωδικοποίηση P-εικόνων (Πηγή: http://fas.sfu.ca/cs/undergrad/CourseMaterials) 27
3.4 Κωδικοποίηση B-εικόνων (Πηγή: http://fas.sfu.ca/cs/undergrad/CourseMaterials) 28
3.5 Σειρά εμφάνισης και αποκωδικοποίησης εικόνων video σε μορφή MPEG1 29

4.1 Λεπτομέρειες Αλγορίθμου Ανίχνευσης Ασυνεχειών 35
4.2 Σειρά Προσπέλασης Εικόνων του Video 39
4.3 Απόσταση ιστογραμμάτων έντασης των διαδοχικών I-εικόνων του VIDEO1 40
4.4 Απόσταση ιστογραμμάτων έντασης και χρώματος των διαδοχικών I-εικόνων του VIDEO1 41
4.5 Πρώτη και τελευταία εικόνα των δεκατριών σκηνών του VIDEO1 42
4.6 Απόσταση ιστογραμμάτων έντασης των διαδοχικών I-εικόνων του VIDEO2 43
4.7 Απόσταση ιστογραμμάτων έντασης και χρώματος των διαδοχικών I-εικόνων του VIDEO2 43
4.8 Πρώτη και τελευταία εικόνα των εννέα σκηνών του VIDEO2 44

5.1 Λεπτομέρειες Αλγορίθμου Εξαγωγής Αντιπροσωπευτικών Εικόνων 49
5.2 Μέθοδος Σειριακής Ομαδοποίησης 50

5.3 Εξαγόμενες αντιπροσωπευτικές εικόνες για την τελευταία σκηνή του VIDEO2 με χρήση ιστογραμμάτων έντασης. 55
5.4 Η μοναδική αντιπροσωπευτική εικόνα για την τελευταία σκηνή του VIDEO2 με χρήση ιστογραμμάτων έντασης-χρώματος. 55

6.1 Αναπαράσταση ενός βήματος του SRG 63
6.2 Ορισμός περιοχών εκκίνησης του αλγορίθμου SRG (κάτω εικόνα) για την πάνω εικόνα 66

7.1 Ιδιότητα αναλογίας χρωμάτων (α) στο RGB και (β) στο YCbCr 70
7.2 Απόσταση διανύσματος χρώματος από ευθεία απόχρωσης, στο χώρο χρώματος YCbCr 72
7.3 Εικόνα με αντικείμενα που παρουσιάζουν διακυμάνσεις στη φωτεινότητά τους και/ή έντονα χρώματα (ΕΙΚΟΝΑ1). 72
7.4 Τμηματοποίηση μέσω του SRG για την ΕΙΚΟΝΑ1, στους χώρους χρωμάτων YCbCr (α) YIQ (b) Lab (γ) και Luv (δ) 73
7.5 Τμηματοποίηση βάσει της αναλογίας χρωμάτων στο RGB (α), στο YCbCr (β) και στο YIQ (γ) 75
7.6 Αυτόματη τμηματοποίηση εικόνων χρώματος 77
7.7 Εξαγωγή κυρίαρχων χρωμάτων της πάνω εικόνας (ΕΙΚΟΝΑ2): στο χώρο GLHS (κάτω αριστερά) και στο χώρο Lab (κάτω δεξιά) 82
7.8 Αρχικές περιοχές βάσει των κυρίαρχων χρωμάτων του GLHS (πάνω εικόνα) και του Lab (κάτω εικόνα). 84
7.9 Επέκταση περιοχών από τον SRG μέσω της αρχικοποίησης στο GLHS (αριστερή εικόνα) και του Lab (δεξιά εικόνα). 85
7.10 Συγχώνευση περιοχών που έχουν προκύψει από την αρχικοποίηση στο GLHS (αριστερή εικόνα) και στο Lab (δεξιά εικόνα) 85
7.11 Εξαγωγή κυρίαρχων χρωμάτων για την ΕΙΚΟΝΑ1 στο χώρο GLHS (αριστερή εικόνα) και στο χώρο Lab (δεξιά εικόνα). 87
7.12 Εξαγωγή αρχικών περιοχών για την ΕΙΚΟΝΑ1 στο χώρο GLHS (αριστερά) και στο χώρο Lab (δεξιά). 87
7.13 Τμηματοποίηση στους χώρους GLHS (πάνω αριστερά) και Lab (πάνω δεξιά) και συγχώνευση (κάτω εικόνα) στο χώρο Lab για την ΕΙΚΟΝΑ1 88
7.14 Τμηματοποίηση (μεσαία εικόνα) και εφαρμογή του κριτηρίου συνόρου (κάτω εικόνα) για την πάνω εικόνα (ΕΙΚΟΝΑ3). 91
7.15 ΕΙΚΟΝΑ4 94
7.16 Κυρίαρχα χρώματα (α), τμηματοποίηση (β) και συγχώνευση περιοχών (γ) για το GLHS (1) και για το Lab (2) για την ΕΙΚΟΝΑ4 95
7.17 ΕΙΚΟΝΑ5 96
7.18 Κυρίαρχα χρώματα (α), τμηματοποίηση (β) και συγχώνευση περιοχών (γ) για το GLHS (1) και για το Lab (2) για την ΕΙΚΟΝΑ5 97

7.19 ΕΙΚΟΝΑ6 98
7.20 Κυρίαρχα χρώματα (α), τμηματοποίηση (β) και συγχώνευση περιοχών (γ) για το GLHS (1) και για το Lab (2) για την ΕΙΚΟΝΑ6 99

8.1 Εικόνα 51 της ακολουθίας *Foreman*. 108
8.2 Αποτέλεσμα παρακολούθησης για την Εικόνα 51 της ακολουθίας *Foreman*. 108
8.3 Ιεραρχική παράσταση του αποτελέσματος της παρακολούθησης για την Εικόνα 51 της ακολουθίας *Foreman*. 109
8.4 Εικόνα 66 της ακολουθίας εικόνων *CoastGuard*. 109
8.5 Αποτέλεσμα τμηματοποίησης για την Εικόνα 66 της ακολουθίας *CoastGuard*. 110
8.6 Βελτιωμένο αποτέλεσμα της ακολουθίας εικόνων *CoastGuard*. 111
8.7 Ιεραρχική παράσταση του μικρού ταχύπλοου, του μεγάλου ταχύπλοου και του background που αποτελούν την εικόνα 66 της ακολουθίας *CoastGuard*. 111

.

Κεφάλαιο 1

Γενική Εισαγωγή

1.1 Κίνητρα της Εργασίας

Η ανάπτυξη των δικτύων μετάδοσης ευρείας κλίμακας (broadband networks) παράλληλα με την αύξηση της υπολογιστικής ισχύος των σταθμών εργασίας και την καθιέρωση προτύπων συμπίεσης καθιστούν απαραίτητη την ύπαρξη συστημάτων διαχείρισης της --κατά κύριο λόγο οπτικής-- πληροφορίας των πολυμέσων.

Επιπλέον, το γεγονός ότι η πληροφορία του video ειδικότερα, απαιτεί μεγάλο χώρο αποθήκευσης και σημαντική υπολογιστική ισχύ για την επεξεργασία της, δημιουργεί την ανάγκη αποδοτικής δεικτοδότησης (indexing) κατά την αποθήκευσή της (storage), ώστε να επιτυγχάνεται η γρήγορη και ταυτόχρονα περιεκτική ανάκτησή της (retrieval), με σκοπό τη χρήση της σε τομείς που εμφανίστηκαν σχετικά πρόσφατα και αναπτύσσονται με ταχείς ρυθμούς όπως είναι:

- η μάθηση από απόσταση (distance learning),
- η τηλεϊατρική,
- η αλληλοδραστική (interactive) τηλεόραση,
- οι ψηφιακές βιβλιοθήκες,
- το ‘‘video επί παραγγελία’’ (video on demand-VOD) και
- τα συστήματα γεωγραφικής πληροφορίας.

Από την καθημερινή χρήση των βάσεων δεδομένων οπτικής πληροφορίας στους παραπάνω τομείς, έχει γίνει σαφές ότι οι παρεχόμενες υπηρεσίες πρέπει να είναι σε θέση να υποστηρίζουν εφαρμογές όπως:

- η ηλεκτρονική διανομή φωτογραφικού υλικού,
- η υποστήριξη στρατηγικών αναζήτησης εικόνων και video μέσω του Παγκόσμιου Διαδικτύου (Internet),

- η ηλεκτρονική δημιουργία και ενημέρωση ‘‘εγγράφων’’ πολυμέσων,
- η μεταβολή του περιεχομένου του video (video editing and re-purposing),
- η υποστήριξη ηλεκτρονικών περιοδικών κ.τ.λ.

Τίθεται επομένως το ζήτημα της επιλογής των χαρακτηριστικών των εικόνων, έτσι ώστε ο χρήστης να βρίσκει στη βάση τις εικόνες ή τα επιμέρους τμήματα εικόνων που τον ενδιαφέρουν γρήγορα και αποτελεσματικά. Επιπρόσθετα, το ανακτώμενο υλικό, θα μπορούσε να είναι τέτοιο που να διευκολύνει την περαιτέρω χρήση του, σε εφαρμογές που προσδιορίζονται από τα ενδιαφέροντα του χρήστη.

Στην εργασία που παρουσιάζεται εδώ, περιγράφονται τα βασικά μέρη ενός συστήματος ανάλυσης του περιεχομένου του video, που ως στόχο έχει την αποδοτική (σε χρόνο επεξεργασίας και απαιτήσεις αποθήκευσης) εξαγωγή χαρακτηριστικών , ικανών να χρησιμεύσουν

- στην κατασκευή ενός μοντέλου δεικτοδότησης δεδομένων που να επιτρέπει τη γρήγορη και περιεκτική ανάκτησή του και
- στην περαιτέρω επεξεργασία εικόνων ή σκηνών του video βάσει αυτών των χαρακτηριστικών.

1.2 Συστήματα Αποθήκευσης και Ανάκτησης Περιεχομένου Video

Σε ότι ακολουθεί αναλύονται διεξοδικά οι ανάγκες που οδήγησαν στην προσέγγιση δεικτοδότησης video μέσω των εξαγόμενων χαρακτηριστικών του περιεχομένου του. Μέσα από τη συζήτηση αυτή τίθεται και το περιβάλλον-πλαίσιο εφαρμογής των επιμέρους τμημάτων του συστήματος που υλοποιήθηκε.

1.2.1 Προηγούμενες Προσεγγίσεις Δεικτοδότησης

Στο παρελθόν, οι τεχνικές που χρησιμοποιήθηκαν για τη συλλογή και δεικτοδότηση της πληροφορίας (κυρίως κειμένου) είχαν ως βάση τους την αποθήκευσή της ως διακριτών αλφαριθμητικών μονάδων (strings), ώστε να ικανοποιεί τις απαιτήσεις ανάκτησής της που επιβάλλουν οι λεγόμενες *δομημένες γλώσσες ερωτήσεων* (Structured Query Languages). Επιπρόσθετα, αναπτύχθηκαν τεχνικές αναζήτησης βάσει περιεχομένου (content based searching) που στηρίζονται ακριβώς σ’ αυτή την αλφαριθμητική αναπαράσταση της πληροφορίας. Επομένως, μια άμεση προσέγγιση της δεικτοδότησης video και εικόνων, θα μπορούσε να είναι η χρήση των τεχνικών κειμένου για την αναπαράσταση του περιεχομένου τους μέσω λέξεων-κλειδιών (key-words) και ιδιοτήτων (attributes). Η ‘‘ευεργετική’’ συνέπεια αυτής της προσέγγισης, είναι ότι η οπτική πληροφορία προσπελαύνεται χρησιμοποιώντας γνωστές γλώσσες ερωτήσεων όπως είναι για παράδειγμα η SQL. Από την άλλη όμως, είναι γνωστό πως αυτού του τύπου η αναπαράσταση ενός τεραστίου

όγκου δεδομένων όπως αυτός των πολυμέσων, απαιτεί σημαντικό επιπρόσθετο χώρο αποθήκευσης και αρκετό χρόνο επεξεργασίας με απαραίτητη μάλιστα την παρουσία του διαχειριστή της βάσης δεδομένων.

Μια άλλη προσέγγιση δεικτοδότησης της οπτικής πληροφορίας σχετίζεται με τις τεχνικές ανάλυσης και κατανόησης εικόνων (image analysis and understanding) Η δεικτοδότηση του video/εικόνας σ' αυτή την περίπτωση γίνεται με τη διάκρισή τους αρχικά σε κλάσεις-κατηγορίες εφαρμόζοντας γνωστές τεχνικές αναγνώρισης προτύπων. Στη συνέχεια τα εξαγόμενα χαρακτηριστικά, αντιστοιχίζονται σε ένα περιορισμένο σύνολο στοιχειωδών προτύπων της βάσης δεδομένων για καθένα από τα οποία προυπάρχει μια συμβολική περιγραφή (descriptors). Η συνολική περιγραφή της όλης σκηνής σ' αυτή την περίπτωση στη βάση δεδομένων, είναι μια ακολουθία τέτοιων περιγραφέων, βάσει των οποίων γίνονται η αναζήτηση και ανάκτησή της. Όπως όμως αναφέρεται στο [1], ''η όλη διαδικασία είναι πολύπλοκη και απαιτεί σημαντικό υπολογιστικό χρόνο''.

1.2.2 Ιδιαιτερότητες της Οπτικής Πληροφορίας

Όπως αναφέρεται και στα [1] [2] όπου επιχειρείται μια συνολική προσέγγιση του θέματος διαχείρισης της προς αποθήκευση οπτικής πληροφορίας, επιπρόσθετες δυσκολίες που μπορεί να παρουσιασθούν από τις παραπάνω προσεγγίσεις έχουν να κάνουν με την ίδια τη φύση της οπτικής πληροφορίας --έναντι για παράδειγμα της πληροφορίας κειμένου-- και με τον τρόπο που ο άνθρωπος αντιλαμβάνεται την πληροφορία αυτή. Προβλήματα που διαπιστώνονται από την καθημερινή χρήση των βάσεων δεδομένων είναι μεταξύ των άλλων:

- Το γεγονός ότι η οπτική πληροφορία γίνεται αντιληπτή με διαφορετικό τρόπο από τους χρήστες της και επομένως η αναπαράσταση όλων αυτών των διαφορετικών εκδοχών ερμηνείας μέσω key-words καθίσταται σχεδόν αδύνατη.
- Το γεγονός ότι η αναπαράσταση κειμένου δεν είναι ικανή να εκφράσει πάντα αποδοτικά την εξέλιξη του περιεχομένου του video στο χρόνο καθώς και τη σημασιολογική σχέση μεταξύ οντοτήτων των εικόνων.
- Το συμπέρασμα που έχει εξαχθεί μετά από έρευνα [2], ότι ο τυχών τεράστιος χώρος δεδομένων περιορίζει αισθητά την ικανότητα του χρήστη να κατανοήσει το περιεχόμενο των δεδομένων και τον τρόπο αποδοτικής αναζήτησης που πρέπει να χρησιμοποιήσει. Επομένως, η λύση είναι να δοθεί η δυνατότητα στο χρήστη να γνωρίσει το σχήμα της βάσης δεδομένων με διάφορους εύκολα κατανοητούς τρόπους.

1.2.3 Τρέχουσες Προσεγγίσεις Δεικτοδότησης

Για όλους αυτούς τους λόγους, αρκετές τεχνικές έχουν προταθεί πρόσφατα στη βιβλιογραφία οι οποίες όσον αφορά τη δεικτοδότηση εικόνων, διακρίνονται ανάλογα με την

προσέγγιση που ακολουθείται σε δύο βασικές κατηγορίες:

- Προσέγγιση δεικτοδότησης βάσει του περιεχομένου (content indexing approach) που βασίζεται σε χαρακτηριστικά όπως το *χρώμα*, το *σχήμα*, ή η *υφή*, τα οποία ουσιαστικά χρησιμοποιούνται και ως δείκτες του μοντέλου.

- Δομική (structural) προσέγγιση, όπου οι εικόνες αναπαρίστανται ως ιεραρχία περιοχών, αντικειμένων, ή μερών αντικειμένων. Η προσέγγιση αυτή βασίζεται στη σχέση των διαφόρων αντικειμένων μιας σκηνής στο χώρο και στην εναλλαγή αυτής της *τοπολογικής* σχέσης τους στο χρόνο. Ακριβώς γι' αυτό το λόγο ενδείκνυται και για την παρακολούθηση (tracking) και καταγραφή της θέσης τους στο χρόνο, μια τεχνική που όπως συζητείται πιο κάτω οδηγεί σε σημαντική συμπίεση της πληροφορίας δεικτοδότησης.

Αρκετά συστήματα δεικτοδότησης χρησιμοποιούν από κοινού και τις δύο προσεγγίσεις, με στόχο την ανάκτηση εικόνων μέσω ερωτήσεων που βασίζονται σε χαρακτηριστικά τόσο της εικόνας ως συνόλου όσο και των επιμέρους τμημάτων-οντοτήτων της.

Από την άλλη πλευρά, μια αποδοτική δεικτοδότηση του video, απαιτεί ανάλυση τόσο της χωρικής όσο και της χρονικής πληροφορίας που περιέχει. Γι' αυτό το λόγο, το video χωρίζεται αρχικά σε επιμέρους στοιχειώδη τμήματα που ονομάζονται *σκηνές* (shots). Μία σκηνή αποτελεί μια ακολουθία εικόνων η οποία παρουσιάζει συνέχεια στο περιεχόμενό της και αναπαριστά μία ανεξάρτητη ενέργεια στο χώρο ή το χρόνο [3]. Σκοπός αυτής της τμηματοποίησης είναι η εξαγωγή αυτοτελών ως προς τη συνέχειά τους τμημάτων του video, ώστε να είναι δυνατός ο αποδοτικός χειρισμός του περιεχομένου τους τόσο στο χώρο όσο και στο χρόνο. Η χωρική πληροφορία της σκηνής συχνά αναπαρίσταται μέσω *αντιπροσωπευτικών εικόνων* ή αλλιώς εικόνων-κλειδιών (key-frames), από τις οποίες εξάγονται χαρακτηριστικά που βασίζονται και στις δύο προσεγγίσεις που αναφέρθηκαν πιο πάνω. Χαρακτηριστικά όπως είναι η κίνηση της κάμερας και των αντικειμένων μιας σκηνής, σχετίζονται άμεσα με τη διάσταση του χρόνου η οποία ενέχεται επιπλέον στο video σε σχέση με τις εικόνες και των οποίων η εξαγωγή (στις περιπτώσεις βέβαια που αυτό είναι εφικτό) μπορεί να οδηγήσει σε ακόμα αποδοτικότερη συμπίεση της αποθήκευσης του video και της πληροφορίας δεικτοδότησής του.

Το σχήμα δεικτοδότησης που μόλις περιγράφηκε είναι ιδανικό και για τους σκοπούς της *περιήγησης* (browsing) στο περιεχόμενο των σκηνών, μια δραστηριότητα που ταιριάζει απόλυτα με τη χρήση του video ως πηγής υλικού δεικτοδότησης. Ο όρος περιήγηση, αναφέρεται στην ''ανεπίσημη'' εξέταση του περιεχομένου χωρίς πολλές φορές κάποιο ιδιαίτερο σκοπό. Κατ' αυτή την έννοια η περιήγηση μοιάζει σε πολλά σημεία με την ανάκτηση των δεδομένων, αν και όπως αναφέρεται στο [4], λόγω της φύσης της αποτελεί πολλές φορές τον καλύτερο τρόπο εξαγωγής αποτελεσμάτων μιας γενικής ερώτησης και βοηθάει στην κατανόηση και διόρθωση ερωτήσεων προς τη βάση οι οποίες δεν είχαν το αναμενόμενο αποτέλεσμα.

1.2.4 Εξαγωγή Χαρακτηριστικών Βάσεων Δεδομένων Οπτικής Πληροφορίας

Είναι γνωστό ότι η ανάπτυξη των βάσεων δεδομένων αποθήκευσης γενικά της πληροφορίας συνίσταται σε δύο μέρη: την *αποθήκευση* και την *ανάκτηση.* Αποτελεί επίσης αδιαμφισβήτητο γεγονός ότι τα δύο αυτά συστατικά μέρη βρίσκονται σε άμεση συσχέτιση μεταξύ τους. Ειδικότερα, όπως φαίνεται στο σχ. 1.1, μια βάση δεδομένων οπτικής πληροφορίας αποτελείται από

- το γραφικό περιβάλλον για τη διευκόλυνση του χρήστη (user interface)
- τις διαδικασίες εξαγωγής των χαρακτηριστικών του video
- την οργάνωση των δεικτών στην πληροφορία του video που επιτρέπει τη χρήση αποδοτικών στρατηγικών αναζήτησης και
- τη διαχείριση της βάσης δεδομένων ώστε να υποστηρίζει:
  - ▷ το διαχωρισμό προγραμμάτων και δεδομένων
  - ▷ την αφαιρετική ως προς τη σύλληψη της, παράσταση της δομής της βάσης ώστε να γίνεται εύκολα αντιληπτή από το χρήστη
  - ▷ τις πολλαπλές όψεις της παρουσίασης των δεδομένων και
  - ▷ τη διασφάλιση συνέπειας της βάσης.

Με βάση το παραπάνω σχήμα, κατά τη διαδικασία ανάκτησης το σύστημα αναλύει ερωτήσεις (queries) των χρηστών, εξάγει το κατάλληλο διάνυσμα χαρακτηριστικών και στη συνέχεια εκτελείται η αναζήτηση υπολογίζοντας την ομοιότητα του διανύσματος ερώτησης με τα αντίστοιχα των αντιπροσωπευτικών εικόνων και του video τα οποία έχουν αποθηκευθεί στη βάση δεδομένων. Αρκετά συστήματα βάσεων δεδομένων έχουν προταθεί στη βιβλιογραφία τα οποία, μιλώντας γενικά, ακολουθούν το παραπάνω σχήμα, με κάποιες διαφορές κυρίως ως προς τη στρατηγική αναζήτησης και την επιλογή των κριτηρίων ομοιότητας των εικόνων --αναφέρονται ενδεικτικά το σύστημα δεικτοδότησης και ανάκτησης βάσει περιεχομένου QBIC [1] [5] που αποτελεί μέρος των προϊόντων της IBM, το Photobook [6], το σύστημα VisualSEEk [7] [8] και τέλος το σύστημα DiVAN [2] [9] στο οποίο έχουν ενσωματωθεί οι μέθοδοι αλλαγής σκηνών και επιλογής αντιπροσωπευτικών εικόνων που παρουσιάζονται στην εργασία. Σημαντική είναι επίσης η εργασία που επιτελείται γενικότερα στον τομέα της ανάκτησης οπτικής πληροφορίας βάσει περιεχομένου, από το Ινστιτούτο Πληροφορικής του Ιδρύματος Τεχνολογίας και Ερευνας (ITE). Ενδεικτικά αναφέρεται η πρόσφατη εργασία πάνω στο θέμα από τους J. Sporring, Ξ. Ζαμπούλη, Π. Τραχανιά και Σ. Ορφανουδάκη στην οποία προτείνεται μια μέθοδος σύγκρισης σχημάτων η οποία μπορεί να χρησιμοποιηθεί για ανάκτηση εικόνων [10].

[1] Query By Image Content
[2] Distributed Audio-Visual Archives Network

Ειδικότερα το δεύτερο συστατικό (εξαγωγή χαρακτηριστικών) το οποίο αποτελεί το αντικείμενο αυτής εδώ της εργασίας, αποτελείται από τις εξής διαδικασίες (βλ. σχ. 1.2):

1. *Ανίχνευση της αλλαγής σκηνών* (scene change (or shot) detection) που αναφέρεται στη βιβλιογραφία και με τον όρο *ανίχνευση ασυνεχειών* (cut detection) κατά την οποία το video τμηματοποιείται σε επιμέρους αυτόνομες ως προς τη δράση τους σκηνές.

2. *Εξαγωγή αντιπροσωπευτικών εικόνων* (key-frame extraction) των σκηνών που προκύπτουν. Οι εικόνες αυτές θεωρείται ότι αντιπροσωπεύουν (βάσει προεπιλεγμένων χαρακτηριστικών) το υπό δεικτοδότηση περιεχόμενο των σκηνών στο χώρο.

3. *Εξαγωγή των χωρικών χαρακτηριστικών* (spatial features) των αντιπροσωπευτικών εικόνων τα οποία ανάλογα με τη φύση τους μπορούν να χρησιμοποιηθούν ακολουθώντας τόσο την προσέγγιση δεικτοδότησης βάσει του περιεχομένου όσο και τη δομική προσέγγιση.

4. *Εξαγωγή των χρονικών χαρακτηριστικών* (temporal features) των σκηνών, όπως είναι ο τύπος της κίνησης της κάμερας ή/και των αντικειμένων της σκηνής.

5. *Προεπεξεργασία των εικόνων* η οποία συνίσταται σε διεργασίες όπως η αποσυμπίεση κωδικοποιημένων εικόνων, ή η εφαρμογή φίλτρων για τη βελτίωση των εικόνων.

Τα εξαγόμενα χαρακτηριστικά διακρίνονται σε *βασικά* (primitive) και *λογικά* (logical). Τα βασικά χαρακτηριστικά όπως είναι για παράδειγμα το χρώμα, τα σύνορα ή το κεντροειδές ενός σχήματος είναι από τη φύση τους *ποσοτικά* και εξάγονται αυτόματα ή ημι-αυτόματα. Αντίθετα το λογικά χαρακτηριστικά διακρίνονται από τα βασικά από το γεγονός ότι παρέχουν μια αφαιρετική αναπαράσταση της οπτικής πληροφορίας των εικόνων σε διάφορα επίπεδα λεπτομέρειας και γι' αυτό είναι *ποιοτικά*. Συνήθως αυτά τα χαρακτηριστικά εξάγονται με το χέρι (manually).

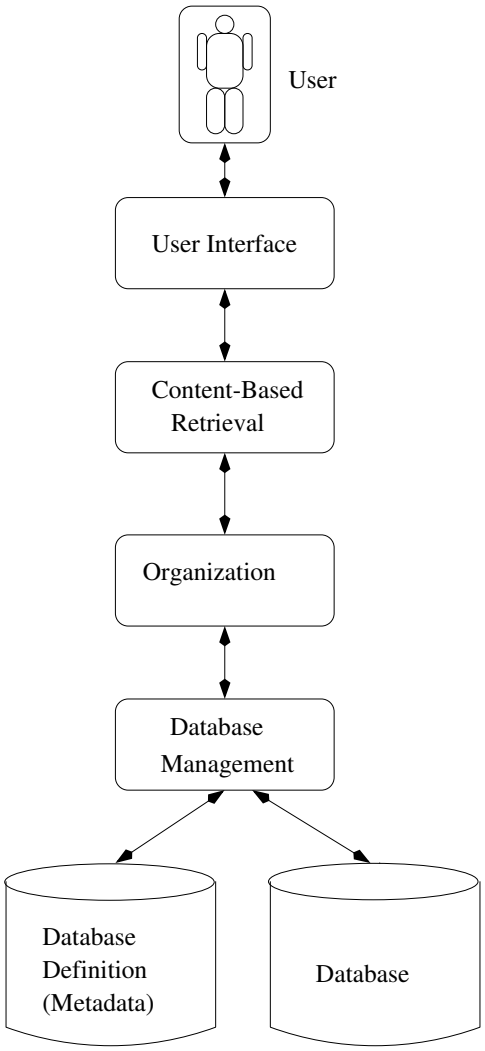
Το αντικείμενο της εργασίας που ακολουθεί, είναι η εξαγωγή χαρακτηριστικών που μπορούν να χρησιμοποιηθούν στη συνέχεια, για την αφαιρετική παράσταση των εικόνων ή σκηνών, μέσω των αντικειμένων τους. Έτσι, γίνεται δυνατή η ανάκτηση εικόνων βάσει της σχέσης των αντικειμένων αυτών μεταξύ τους στο χώρο ή τη μεταβολή αυτής της σχέσης τους στο χρόνο. Επιπλέον, η ανάκτηση μπορεί να βασίζεται σε χαρακτηριστικά ή ιδιότητες των αντικειμένων αντί όλης της εικόνας. Από την άλλη πλευρά, η ανίχνευση αλλαγής σκηνών και η εξαγωγή αντιπροσωπευτικών εικόνων βασίζονται στη μέτρηση του ιστογράμματος του χρώματος και υπό αυτή την έννοια, εξάγονται και βασικά χαρακτηριστικά των εικόνων τα οποία είναι δυνατό να αποθηκευθούν περαιτέρω για τις ανάγκες ανάκτησης.

1.3 Οργάνωση της Εργασίας

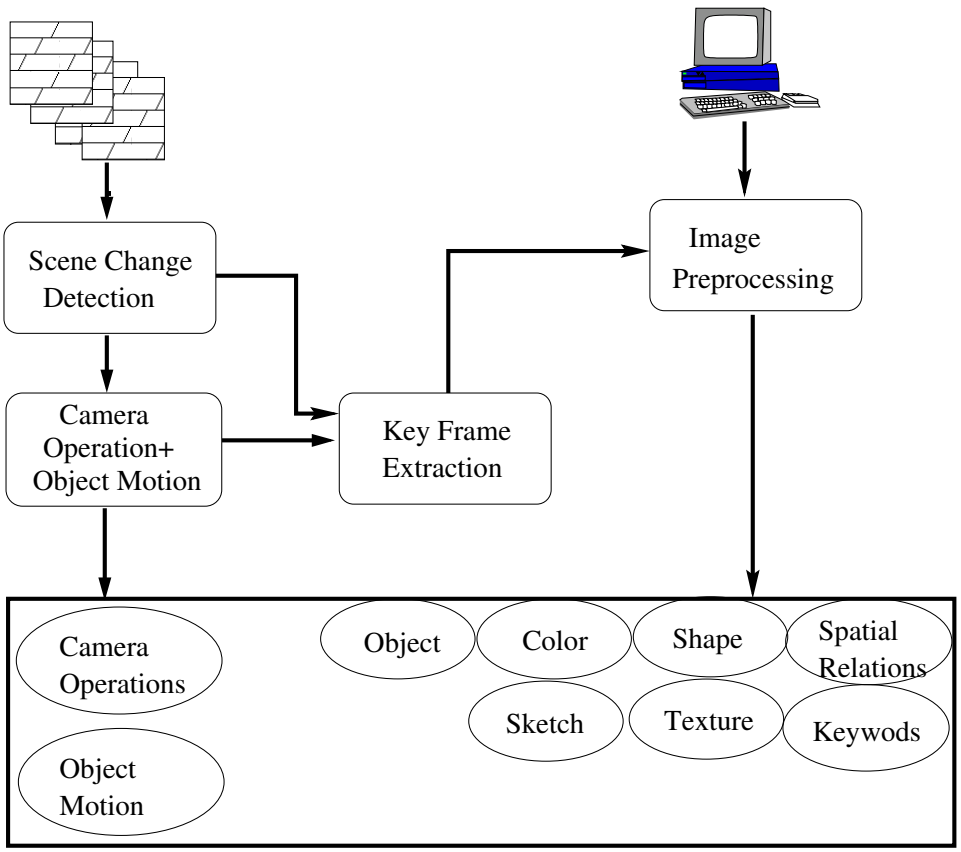
Η εργασία έχει οργανωθεί ως εξής: στο επόμενο κεφάλαιο, περιγράφονται οι βασικές έννοιες γύρω από το χρώμα, ως πληροφορίας μετάδοσης και επεξεργασίας εικόνων.

Στο κεφάλαιο 3 περιγράφεται ένα ιεραρχικό μοντέλο δεικτοδότησης δεδομένων του video, του οποίου η δομή συμπληρώνεται από τις διαδικασίες ανίχνευσης αλλαγής σκηνών και εξαγωγής των αντιπροσωπευτικών εικόνων του video οι οποίες περιγράφονται στα κεφάλαια 4 και 5 αντίστοιχα. Οι δύο αυτές διαδικασίες σχετίζονται άμεσα με το μοντέλο δεδομένων που καλούνται να συμπληρώσουν βάσει του περιεχομένου του video.

Στο δεύτερο μέρος αυτής της εργασίας περιγράφονται μέθοδοι εξαγωγής αντικειμένων που βασίζονται σε μια τεχνική τμηματοποίησης εικόνων η οποία αναλύεται στο κεφάλαιο 6. Στο κεφάλαιο 7 παρουσιάζεται μια μέθοδος εξαγωγής αντικειμένων των εικόνων βάσει του χρώματος τους, ενώ στο κεφάλαιο 8, περιγράφεται ένας αλγόριθμος παρακολούθησης (tracking) αντικειμένων στο χρόνο με βάση την ταχύτητά τους. Οι δύο τεχνικές προσφέρονται για την ανάλυση των αντιπροσωπευτικών εικόνων του video ή/και για τη μείωση του χωρικού και χρονικού πλεονασμού (spatial and temporal redundancy) που χαρακτηρίζει την πληροφορία των σκηνών του video.



Σχήμα 1.1: Αποθήκευση και ανάκτηση εικόνων και video.



Σχήμα 1.2: Εξαγωγή και αποθήκευση βάσει του περιεχομένου της οπτικής πληροφορίας

Κεφάλαιο 2

Αναπαράσταση Πληροφορίας Χρώματος

Το χρώμα αποτελεί μια βασική πηγή πληροφορίας των εικόνων και έχει χρησιμοποιηθεί τόσο για τη δεικτοδότηση των εικόνων βάσει του περιεχομένου τους όσο και για την ανάλυση των οντοτήτων που περιέχουν. Το κεφάλαιο αυτό παρουσιάζει τις βασικές αρχές ανάλυσης του χρώματος ως σήματος *μετάδοσης* και *επεξεργασίας* των εικόνων.

2.1 Μέτρηση Χρώματος (Colorimetry)

Από την εποχή του Νεύτωνα ακόμα, είναι γνωστό ότι ένα ευρύ φάσμα χρωμάτων είναι δυνατό να παραχθεί από ένα σύνολο τριών στοιχειωδών (χρωμάτων) (primaries): ένας ζωγράφος για παράδειγμα θα μπορούσε να δημιουργήσει τις μπογιές του αναμειγνύοντας το φως των τριών *αφαιρετικών* (subtractive) στοιχειωδών χρωμάτων: κόκκινο, κίτρινο και μπλε. Οι οθόνες της τηλεόρασης παράγουν τα χρώματα αναμειγνύοντας το φως των τριών *αθροιστικών* (additive) στοιχειωδών χρωμάτων: κόκκινο, πράσινο και μπλε. Αν και θα μπορούσε να χρησιμοποιηθεί ο συνδυασμός δύο συστημάτων στοιχειωδών --μιλώντας γενικά ένα κόκκινο-πορτοκαλί και ένα πράσινο-μπλε-- η εικόνα που προκύπτει δε μοιάζει τόσο φυσική και είναι γνωστό ότι αυτός ο συνδυασμός δεν υπήρξε ποτέ επιτυχημένος στην τηλεόραση.

Βάσει της θεωρίας των τριών στοιχειωδών, έχει επιβεβαιωθεί πειραματικά ότι η ηλεκτρομαγνητική ακτινοβολία $C(\lambda)$ στο εύρος του ορατού φωτός ($\lambda \in \{380nm, 780nm\}$) γίνεται αντιληπτή ως χρώμα ή χρωματικό φως. Το χρώμα γίνεται αντιληπτό από το ανθρώπινο μάτι, διαμέσου τριών ανεξάρτητων δεκτών-αισθητήρων φωτός με συναρτήσεις απορρόφησης στο φασματικό πεδίο $S_r(\lambda)$, $S_g(\lambda)$, $S_b(\lambda)$. Οι αποκρίσεις των δεκτών αυτών γίνονται μέγιστες στις περιοχές κίτρινου-πράσινου (λ=700nm), πράσινου (λ=546.1nm) και μπλε (λ=435.8nm) αντίστοιχα, του φάσματος του ορατού φωτός με σημαντική επικάλυψη μεταξύ των δεκτών S_r και S_g (σχ. 2.1). Αντιστοιχίζοντας σε κάθε έναν από τους δέκτες μια συνάρτηση απόκρισης $P_k(\lambda)$ ($k \in \{r, g, b\}$), η αναπαράσταση

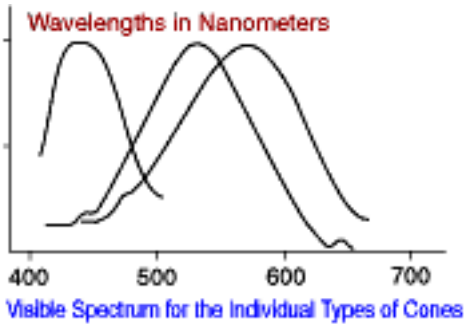
του ορατού φωτός οποιουδήποτε χρώματος $C(\lambda)$ προκύπτει από τη γραμμική υπέρθεση των P_k [11], ως εξής: το χρώμα $C(\lambda)$ ορίζεται να είναι αυτό που παράγει τις αποκρίσεις (R, G, B) που προκύπτουν από τη σχέση

$$C(\lambda) = R\bar{P}_r(\lambda) + G\bar{P}_g(\lambda) + B\bar{P}_b(\lambda) \tag{2.1}$$

όπου $\bar{P}_k(\lambda)$, είναι οι αντίστοιχες κανονικοποιημένες συναρτήσεις απόκρισης των $P_k(\lambda)$ στο λευκό φως αναφοράς (reference white light) $W(\lambda)$ σύμφωνα με τη σχέση

$$W(\lambda) = \bar{P}_r(\lambda) + \bar{P}_g(\lambda) + \bar{P}_b(\lambda) \tag{2.2}$$

Έτσι κάθε χρώμα αναπαρίσταται από τα τρία στοιχειώδη χρώματα (R, G, B).



Σχήμα 2.1: Ορατό φάσμα φωτός των χρωμάτων. (Πηγή: http://www.linocolor.www)

Ο χώρος χρωμάτων που προκύπτει συνδυάζοντας τα τρία χρώματα περιγράφεται από ένα τρίγωνο σε ένα ειδικό χάρτη χρωμάτων με την ιδιότητα ότι κάθε βασικό χρώμα αποτελεί μια κορυφή του τριγώνου. Ο κλασικός χάρτης χρωμάτων που χρησιμοποιήθηκε από τις πρώτες μέρες της τηλεόρασης προτάθηκε το 1931 από τη Commission Internationale de L' Eclairage (CIE) [12]. Ο χάρτης αυτός ονομάζεται *χρωματικό διάγραμμα* (chromaticity diagram) και το τρίγωνο των χρωμάτων που είναι δυνατό να αναπαραχθούν από τα τρία στοιχειώδη χρώματα αποτελεί τη λεγόμενη *κλίμακα χρωμάτων* (color gamut).

Το χρωματικό διάγραμμα ήταν το αποτέλεσμα μιας σειράς νέων ιδεών που εισήχθηκαν από τη CIE στην ανάλυση των χρωμάτων, μία από τις οποίες ήταν η απομόνωση της φωτεινότητας (luminance) ή της φαινόμενης φωτεινότητας (brightness) από τη χρωματικότητα (chrominance) του κάθε χρώματος. Το χρωματικό διάγραμμα αντιστοιχεί στα χρωματικά συστατικά των χρωμάτων και για δοσμένη τιμή φωτεινότητας Y είναι ένα χρωματικό επίπεδο (chrominance plane) --υπόχωρος του χώρου χρωμάτων που προκύπτει από τα χρωματικά επίπεδα του συνόλου των δυνατών τιμών φωτεινότητας.

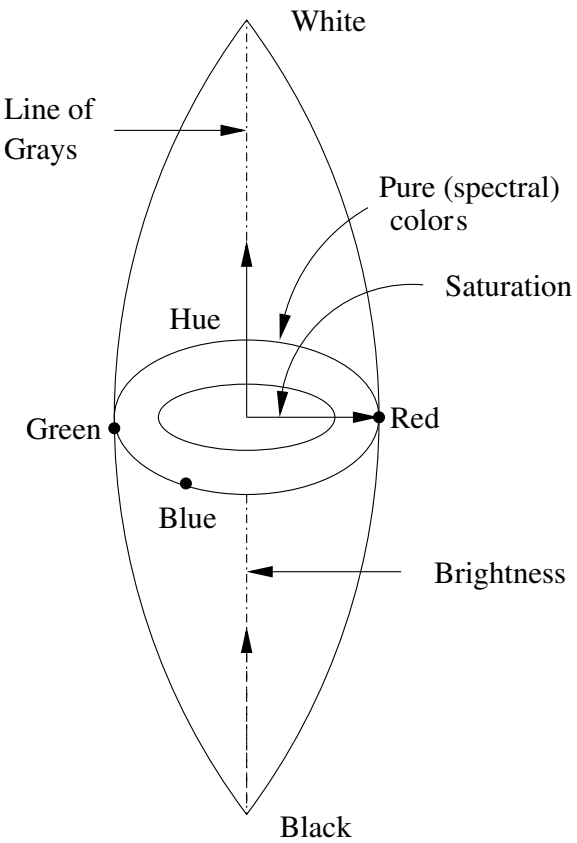
Η ανάλυση αυτή των χρωμάτων ακολουθεί τη φυσική διάκριση των τριών συστατικών των χρωμάτων όπως αυτές γίνονται αντιληπτές (perceived) από το ανθρώπινο μάτι. Αναφερόμενοι στο σχήμα 2.2, τα συστατικά αυτά είναι τα εξής:

Φαινόμενη (Apparent) Φωτεινότητα (brightness). Είναι η ιδιότητα της οπτικής αντίληψης που αντιστοιχεί στο κατά πόσο ένα χρώμα είναι περισσότερο ή λιγότερο φωτεινό.

Απόχρωση (Hue). Είναι η ιδιότητα της οπτικής αντίληψης του κατά πόσο ένα χρώμα είναι όμοιο με ένα ή δύο από τα χρώματα κόκκινο, πράσινο και μπλε.

Κορεσμός (Saturation). Είναι η ιδιότητα της οπτικής αίσθησης που αντιστοιχεί στο κατά πόσο ένα χρώμα εμφανίζεται περισσότερο ή λιγότερο όμοιο με την απόχρωσή του. Ο κορεσμός μετράει το βαθμό στον οποίο ένα χρωματικό φως διαφέρει από ένα μή χρωματικό φως ανεξάρτητα από τη φαινόμενη φωτεινότητά τους.

Στο σχήμα 2.2, φαίνεται η γραφική αναπαράσταση αυτών των συστατικών. Η μετάβαση από το μαύρο στο άσπρο καταγράφεται στον κατακόρυφο άξονα της φαινόμενης φωτεινότητας, η απόχρωση αντιστοιχεί στη μεταβολή της γωνίας πάνω σε κύκλους δοσμένης ακτίνας, ενώ ο κορεσμός μεταβάλλεται κατά την ακτίνα των ομόκεντρων κύκλων.



Σχήμα 2.2: Αντιληπτική αναπαράσταση του χώρου χρωμάτων

2.2 Χώροι ή Συστήματα Χρώματος

Σ' ότι ακολουθεί παρουσιάζονται γνωστοί χώροι χρωμάτων (color spaces or systems) που χρησιμοποιήθηκαν στην εργασία και οι μετασχηματισμοί από το χώρο χρωμάτων RGB σε κάθε έναν από αυτούς. Η ποικιλία αυτή των χώρων χρώματος οφείλεται στο ότι δεν

υπάρχει κάποιο παγκόσμια αποδεκτό σύστημα χρωμάτων, αλλά και στο γεγονός ότι η αντίληψη των χρωμάτων είναι υποκειμενική. Προηγείται όλων αυτών, μια σύντομη παρουσίαση των ιδιοτήτων των χώρων χρώματος που κυρίως μας ενδιαφέρουν για εφαρμογές στην επεξεργασία εικόνων.

2.2.1 Ιδιότητες Χώρων Χρώματος

Μέχρι σήμερα, έχουν προταθεί αρκετά συστήματα συντεταγμένων χρώματος (color coordinate systems). Κάθε ένα από αυτά προορίζεται να ικανοποιήσει ανάγκες που προκύπτουν από τη χρήση του χρώματος τόσο στην επεξεργασία όσο και στην μετάδοση και εμφάνιση (display) εικόνων. Αν και η σύγκριση αυτών των συστημάτων είναι πολλές φορές δύσκολη εξ αιτίας των διαφορετικών λόγων ύπαρξής τους, υπάρχουν ορισμένες ιδιότητες που όσον αφορά τη χρήση του χρώματος στην επεξεργασία εικόνων τουλάχιστο, θα ήταν καλό οι χώροι αυτοί να πληρούν.

Σύμφωνα με το [13], οι ιδιότητες αυτές είναι οι εξής:

Ιδιότητα 1: *Ομοιομορφία (uniformity).* Η μέτρηση του πόσο κοντά βρίσκονται δύο χρώματα (proximity metric), αντιστοιχεί στην αντιλαμβανόμενη από τον άνθρωπο (αν)ομοιότητά τους.

Ιδιότητα 2: *Πληρότητα (completeness).* Ο χώρος χρωμάτων περιλαμβάνει όλα τα χρώματα που γίνονται αντιληπτά ως διακριτά.

Ιδιότητα 3: *Συμπύκνωση (compactness).* Κάθε χρώμα στο χώρο χρωμάτων διακρίνεται (ή γίνεται αντιληπτό ως διακριτό) από τα υπόλοιπα χρώματα.

Ιδιότητα 4: *Φυσικότητα (naturalness).* Ο χώρος χρωμάτων παρέχει τη φυσική διαίρεση των χρωμάτων στις τρεις αντιληπτές ιδιότητές τους: φαινόμενη φωτεινότητα, απόχρωση και κορεσμό.

2.2.2 Ο Χώρος Χρώματος RGB

Ο χώρος χρωμάτων που χρησιμοποιείται από τις οθόνες υπολογιστών-τηλεόρασης για την εμφάνιση γραφικών είναι ο γνωστός τρισδιάστατος χώρος RGB στον οποίο κάθε χρώμα περιγράφεται από μία τριπλέτα χρωμάτων (R, G, B) σύμφωνα με τη σχέση (2.1). Ο χώρος αυτός είναι πλήρης, αφού περιλαμβάνει όλα τα ορατά χρώματα. Από την άλλη πλευρά, εύκολα διαπιστώνεται ότι δεν παρουσιάζει ομοιομορφία, αφού η απόσταση των χρωμάτων σ' αυτό το σύστημα συντεταγμένων δεν αντιστοιχεί στη διάκρισή τους όπως αυτή γίνεται αντιληπτή από τον άνθρωπο. Παρόλα αυτά, λόγω της πληρότητας του, μπορεί να χρησιμοποιηθεί ως η βάση μετάβασης σε χώρους που επίσης είναι πλήρεις, χωρίς αναγκαστικά οι μετασχηματισμοί αυτοί να είναι γραμμικοί. Στο Παράρτημα Α, δίνονται οι μετασχηματισμοί που χρησιμοποιήθηκαν για τη μετάβαση από το χώρο RGB,

στους γραμμικούς ή ομοιόμορφους χώρους, ανάλογα με τις ανάγκες της επεξεργασίας του χρώματος.

2.2.3 Πρότυποι Γραμμικοί Μετασχηματισμοί Χρώματος

Χρησιμοποιώντας το χρωματικό διάγραμμα της CIE ως οδηγό, η National Television System Committee (NTSC), όρισε τη μετάδοση σημάτων χρώματος στη μορφή φωτεινότητα-χρωματικότητα, αντί της μορφής των τριών συστατικών του χρώματος όπως αναπαρίσταται από τις οθόνες της τηλεόρασης. Ο νέος χώρος χρώματος που εισήχθηκε από την NTSC είναι το YIQ και συνδέεται γραμμικά με τον χώρο RGB.

Στην Ευρώπη, αναπτύχθηκαν αργότερα δύο πρότυπα τηλεόρασης, το Phase-Alternation-Line (PAL) και το Sequentiel Couleura Memoire (SECAM) που και τα δύο υποστήριζαν τον χώρο χρωμάτων YUV. Το ψηφιακό ισοδύναμο του YUV είναι το YCbCr το οποίο χρησιμοποιείται για τη μετάδοση χρώματος από τους υπολογιστές.

Και στα τρία συστήματα, η συνιστώσα Y αντιστοιχεί στη φωτεινότητα ενώ οι άλλες συνιστώσες στη χρωματικότητα των χρωμάτων. Πρέπει να τονιστεί, ότι η μόνη διαφορά των συστημάτων YIQ και YUV είναι μια περιστροφή 33 μοιρών στο UV-επίπεδο. Επιπλέον, στο σύστημα YCbCr οι ψηφιακές συνιστώσες Cb, Cr, αντιστοιχούν στις αναλογικές συνιστώσες U, V, του YUV, αν και υπολογίζονται μέσω διαφορετικών συντελεστών κλίμακας.

Αν και οι γραμμικοί μετασχηματισμοί είναι απλοί, πλήρεις και όπως θα δούμε σε επόμενα κεφάλαια χαρακτηρίζονται από ενδιαφέρουσες στατιστικές ιδιότητες, δεν παράγουν φυσικούς ομοιόμορφους χώρους χρώματος.

2.2.4 Χώροι Χρώματος της CIE

Έχοντας ως στόχο τη δημιουργία ομοιόμορφων χώρων χρώματος, η CIE όρισε το 1976 δύο χώρους χρώματος που ονομάστηκαν CIE 1976 Lab και CIE 1976 Luv. Αρχικά ο χώρος RGB μετασχηματίζεται γραμμικά στο χώρο XYZ. Ο χώρος αυτός ορίστηκε από τη CIE για να ξεπεραστεί το πρόβλημα των τριπλετών χρωμάτων που δεν είναι δυνατό να αναπαραχθούν από το πρώτο χρωματικό διάγραμμα της CIE (σχ. 2.3). Αν και το χρωματικό διάγραμμα του XYZ είναι μή ρεαλιστικό, διευκολύνει τους υπολογισμούς μέτρησης χρώματος.

Ο γραμμικός μετασχηματισμός του RGB στο XYZ, ακολουθείται από έναν ή περισσότερους μή γραμμικούς μετασχηματισμούς ώστε να προκύψει το σύστημα CIE που μας ενδιαφέρει.

Η σημαντική ιδιότητα αυτών των χώρων είναι ότι --όπως υποστηρίζεται από τη CIE-- η αντιληπτή απόσταση μεταξύ δύο χρωμάτων, αντιστοιχεί στην ευκλείδια απόσταση των διανυσμάτων τους. Αυτό επιτυγχάνεται μετατρέποντας το χρωματικό διάγραμμα του χώρου XYZ, σε ένα ομοιόμορφο χώρο για κάθε τιμή φαινόμενης φωτεινότητας. Για το Lab, το αντίστοιχο χρωματικό διάγραμμα φαίνεται στο σχ. 2.4, ενώ ο τρισδιάστατος

χώρος χρώματος που προκύπτει φαίνεται στο σχ. 2.5. Ο ισχυρισμός αυτός όμως, όπως αναφέρεται στο [14], μένει να αποδειχθεί στην πράξη.

Οι χώροι χρώματος της CIE είναι πλήρεις και δίνουν την ίδια έμφαση στις τρεις παραμέτρους αντίληψης του χρώματος: τη σχετική φωτεινότητα (συνιστώσα L), την απόχρωση και τον κορεσμό (όπως θα μπορούσαν να εξαχθούν από τα επίπεδα χρώματος (a,b) και (u,v) αντίστοιχα για κάθε τιμή του L). Ωστόσο, δεν έχει επιβεβαιωθεί ακόμη εάν και κατά πόσο οι χώροι αυτοί περιλαμβάνουν μόνο τα αντιληπτά ως διαφορετικά από το ανθρώπινο μάτι χρώματα, δηλαδή κατά πόσο πληρούν τις απαιτήσεις συμπύκνωσης των πραγματικά διακριτών χρωμάτων.

2.3 Αναπαράσταση Πληροφορίας Έντασης και Χρώματος Εικόνων

Περιγράφεται στη συνέχεια ο τρόπος παράστασης των χαρακτηριστικών της έντασης και του χρώματος τα οποία χρησιμοποιούνται για τη δεικτοδότηση στατικών εικόνων και τον έλεγχο της ομοιότητας τους βάσει των χρωμάτων που περιέχουν. Τα χαρακτηριστικά αυτά αποτελούν ένα ελάχιστο υποσύνολο των χαρακτηριστικών που χρησιμοποιούνται από τα συστήματα ανάκτησης βάσει περιεχομένου [15]. Αφού αναφερθούν κάποιες μέθοδοι μείωσης του πλεονασμού που --σε αντίθεση με την ένταση-- χαρακτηρίζει την πληροφορία του χρώματος, παρουσιάζονται στη συνέχεια δύο προσεγγίσεις παράστασης

- της κατανομής των επιλεγμένων χρωμάτων ή τιμών έντασης σε όλη την εικόνα (ιστογράμματα) και
- της ολικής κατανομής της τοπικής (local) χωρικής συσχέτισης αυτών, γύρω από κάθε σημείο (συσχετογράμματα-correlograms) [16].

2.3.1 Κβαντισμός (Quantization) Χρώματος

Είναι γεγονός ότι το χρώμα αποτελεί μια πλούσια πληροφορία: σε αντίθεση με τις τιμές της φωτεινότητας που αναπαρίστανται ψηφιακά από ένα περιορισμένο αριθμό τιμών που είναι συνήθως 256, η διακριτοποίηση του κάθε άξονα του RGB σε 256 τιμές δίνει πάνω από 1.67 εκατομμύρια διακριτά χρώματα.

Για να είμαστε σε θέση να διαχειριστούμε αποδοτικά την πληροφορία του χρώματος, απαιτείται ο κβαντισμός των χρωμάτων της εικόνας ώστε οι μετρήσεις να περιλαμβάνουν ένα μικρότερο αριθμό χρωμάτων.

Οι μέθοδοι κβαντισμού των χρωμάτων ποικίλουν. Για παράδειγμα μια απλή μορφή κβαντισμού αποτελεί η διαίρεση καθενός από τους άξονες του χώρου χρωμάτων σε υποδιαστήματα που περιλαμβάνουν τον ίδιο σταθερό αριθμό τιμών. Επιπρόσθετα, στη βιβλιογραφία έχουν προταθεί αρκετές τεχνικές που αποσκοπούν σε μια πιο αιτιοκρατική επιλογή των χρωμάτων κβαντισμού, βάσει των χρωμάτων που εμφανίζονται στην εικόνα.

Ο Σ. Λιάπης στο [17] προτείνει μια μέθοδο επιλογής των λεγομένων *κυρίαρχων χρωμάτων* (dominant colors) της εικόνας, που στηρίζεται στον αλγόριθμο k-means. Στην ίδια εργασία περιλαμβάνεται γενικότερα μια ανασκόπηση των μεθόδων που έχουν κατά καιρούς προταθεί πάνω στο ίδιο αντικείμενο.

Τα χρώματα που προκύπτουν από τον κβαντισμό χρησιμοποιούνται στη συνέχεια για την οργάνωση της μέτρησης του χρώματος των εικόνων.

2.3.2 Ιστογράμματα

Τα ιστογράμματα χρησιμοποιούνται για την αναπαράσταση της κατανομής της υπό θεώρηση διακριτής πληροφορίας όπως για παράδειγμα είναι η ένταση ή το χρώμα των εικόνων. Η μέτρηση συχνά περιλαμβάνει την πληροφορία όλης της εικόνας ή μιας περιοχής της ή αντικειμένου της.

Θεωρώντας M διακριτές τιμές του μετρούμενου χαρακτηριστικού και ότι η μέτρηση αφορά το σύνολο των σημείων της εικόνας $I(x, y)$, το ιστόγραμμα H υπολογίζεται από τη σχέση:

$$H(m) = \sum_{x=0}^{N_L-1} \sum_{y=0}^{N_C-1} \begin{cases} 1 & \text{εάν I}(x, y) = m \\ 0 & \text{αλλιώς} \end{cases} \tag{2.3}$$

όπου $N_L \times N_C$ είναι οι διαστάσεις της εικόνας.

Στην περίπτωση των ιστογραμμάτων χρώματος, τα M διακριτά χρώματα (που αναφέρονται και ως bins) είναι δυνατό να εξαχθούν μέσω των μεθόδων κβαντισμού που περιγράφηκαν πιο πάνω και αυτό είναι που γίνεται στην πράξη, ώστε να είμαστε σε θέση να μετρήσουμε αξιόπιστα την κατανομή των χρωμάτων της εικόνας. Μ' αυτό τον τρόπο λύνονται τα προβλήματα αποθήκευσης και επεξεργασίας των ιστογραμμάτων χρώματος και γίνεται αποδοτική η σύγκριση των ιστογραμμάτων χρώματος των εικόνων.

Τα ιστογράμματα χρησιμοποιούνται ευρέως από όλα τα συστήματα ανάκτησης βάσει περιεχομένου, γιατί ο υπολογισμός τους είναι απλός και γρήγορος. Από την άλλη πλευρά, είναι δυνατό δύο εικόνες με εντελώς διαφορετικό περιεχόμενο να έχουν όμοια ιστογράμματα. Η πιθανότητα εμφάνισης αυτού του δυσάρεστου φαινομένου αυξάνεται καθώς αυξάνεται ο αριθμός των δεικτοδοτούμενων εικόνων της βάσης. Στο [18] αναλύονται σε θεωρητικό επίπεδο τόσο οι δυνατότητες όσο και οι προϋποθέσεις σωστής χρήσης των ιστογραμμάτων για τη δεικτοδότηση εικόνων.

Η μέθοδος των συσχετογραμμάτων που περιγράφεται στη συνέχεια, ξεπερνάει σε αρκετές περιπτώσεις τα προβλήματα που παρουσιάζουν τα ιστογράμματα.

2.3.3 Ιστογράμματα με Συνδυασμό Χρώματος και Άλλων Χαρακτηριστικών

Μια σημαντική προσέγγιση στη χρήση χαρακτηριστικών των εικόνων για τη διαδικασία της ανάκτησης, αποτελούν τα λεγόμενα συσχετογράμματα (correlograms) τα οποία μετρούν τη συσχέτιση των χρωμάτων σε μια γειτονιά κάθε σημείου. Τα συσχετογράμματα

προτάθηκαν από το Zabih κ. άλ. [16] ως μια μέθοδος που μειώνει την πιθανότητα εμφάνισης του προβλήματος των ιστογραμμάτων όπως περιγράφηκε πιο πάνω.

Σε αντίθεση με τα ιστογράμματα, τα συσχετογράμματα λαμβάνουν υπόψη την κατανομή της πληροφορίας στη διάσταση του χώρου, αφού η μέτρηση της συχνότητας εμφάνισης δεν αναφέρεται στο σύνολο της εικόνας αλλά στην τοπική συσχέτιση μιας γειτονιάς σημείων βάσει του μετρούμενου χαρακτηριστικού. Το χαρακτηριστικό αυτό μπορεί να είναι το χρώμα η ένταση ή οποιοδήποτε άλλο βασικό χαρακτηριστικό των εικόνων σε επίπεδο pixels μπορεί να χρησιμοποιηθεί για τη σύγκριση της ομοιότητας εικόνων.

2.3.4 Ομοιότητα Εικόνων βάσει της Παράστασης Κατανομής Χαρακτηριστικών Τους

Καθώς, όπως ήδη ειπώθηκε, οι τεχνικές τόσο των ιστογραμμάτων όσο και των συσχετογραμμάτων βασίζονται σε μετρήσεις της (τοπικής ή μή) συχνότητας εμφάνισης χαρακτηριστικών, οποιοδήποτε από τα γνωστά μέτρα σύγκρισης διακριτών κατανομών ενδείκνυται για τη σύγκριση τους. Στη συνέχεια παρουσιάζονται αυτά τα στατιστικά μέτρα όπως εφαρμόζονται στη σύγκριση των ιστογραμμάτων H_1 και H_2:

$$\delta_1(H_1, H_2) = \sum_{m=1}^{M} |H_1(m) - H_2(m)| \tag{2.4}$$

$$\delta_2(H_1, H_2) = \sum_{m=1}^{M} \frac{(H_1(m) - H_2(m))^2}{H_1(m) + H_2(m)} \tag{2.5}$$

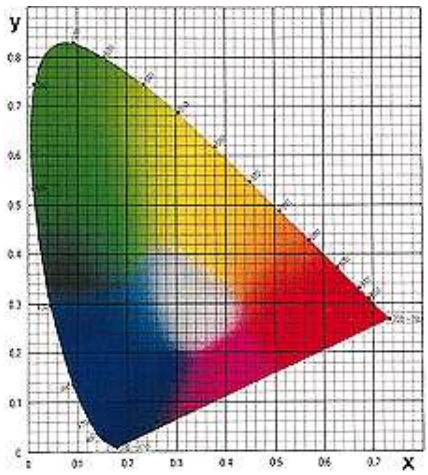
$$\delta_3(H_1, H_2) = max_k\{|\sum_{m=1}^{k} H_1(m) - H_2(m)|\} \tag{2.6}$$

$$\delta_4(H_1, H_2) = 1 - \frac{\sum_{m=1}^{M} \min\{H_1(m), H_2(m)\}}{\sum_{m=1}^{M} H_2(m)} \tag{2.7}$$

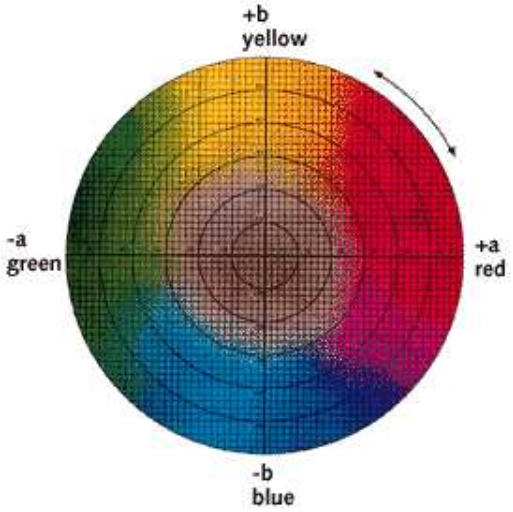
όπου M είναι ο αριθμός των εισόδων-διακριτών τιμών του ιστογράμματος.

Από τα κριτήρια αυτά, η εξίσωση (2.5) είναι το στατιστικό κριτήριο χ^2 ενώ η εξίσωση (2.6) είναι το στατιστικό κριτήριο Kolmogorov-Smirnov. Τέλος η εξίσωση (2.7) βασίζεται στην κανονικοποιημένη τομή των δύο ιστογραμμάτων.

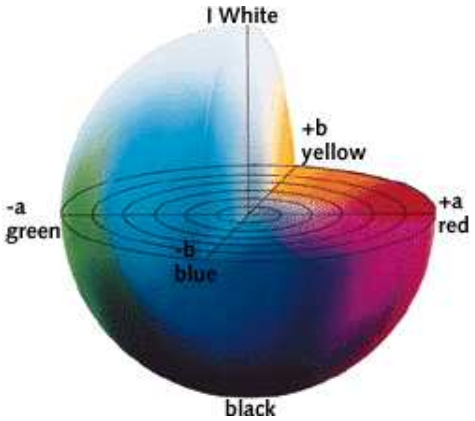
Τα κριτήρια αυτά χρησιμοποιήθηκαν στην εργασία για τη μέτρηση της ομοιότητας εικόνων μέσω των ιστογραμμάτων τους τόσο κατά την ανίχνευση αλλαγής των σκηνών του video όσο και για την εξαγωγή των αντιπροσωπευτικών εικόνων, όπως περιγράφεται στα επόμενα τρία κεφάλαια. Σ' αυτά, εκτός των άλλων, συζητιούνται οι λόγοι που οδήγησαν στην επιλογή των ιστογραμμάτων για τη μέτρηση της ομοιότητας εικόνων.



Σχήμα 2.3: Χρωματικό διάγραμμα του χώρου XYZ. (Πηγή: http://www.linocolor.com)



Σχήμα 2.4: Χρωματικό διάγραμμα του χώρου Lab. (Πηγή: http://www.linocolor.com)



Σχήμα 2.5: Ο χώρος χρώματος Lab (Πηγή: http://www.linocolor.www)

Κεφάλαιο 3

Τμηματοποίηση και Δεικτοδότηση του Περιεχομένου Video

3.1 Εισαγωγή

Στο κεφάλαιο αυτό παρουσιάζονται οι βασικές αρχές και οι προδιαγραφές του σχεδιασμού του συστήματος δεικτοδότησης video που υλοποιήθηκε. Στην προσέγγιση που περιγράφεται αναλυτικά στα κεφάλαια 4, 5, οι διεργασίες της εξαγωγής αντιπροσωπευτικών εικόνων και της ανίχνευσης αλλαγής των σκηνών εκτελούνται ‘‘παράλληλα’’: οι αντιπροσωπευτικές εικόνες εξάγονται, ενώ η διεργασία ανίχνευσης αλλαγής σκηνών ‘‘επιχειρεί’’ να προσδιορίσει την πρώτη εικόνα της επόμενης σκηνής. Ταυτόχρονα, οι εξαγώμενες από τον αλγόριθμο αντιπροσωπευτικές εικόνες καθώς και τα όρια των σκηνών χρησιμοποιούνται για την συμπλήρωση της δομής του ιεραρχικού μοντέλου δεδομένων δεικτοδότησης του video (hierarchical video indexing data model) το οποίο φαίνεται στο σχ. 3.1 και περιγράφεται αναλυτικά στο υποκεφάλαιο 3.3. Η κατασκευή αυτού του σχήματος, η οποία προφανώς αποτελεί το στόχο της ολικής υλοποίησης, δίνεται ως έξοδος από τον συνολικό αλγόριθμο. Από αυτή την άποψη, το σύστημα είναι ανάλογο του συστήματος δεικτοδότησης video που περιγράφεται στο [4], αφού και σ’ αυτό οι διαδικασίες ανίχνευσης ασυνεχειών και εξαγωγής αντιπροσωπευτικών εικόνων συνδέονται άμεσα μεταξύ τους.

Ο αλγόριθμος εφαρμόστηκε σε video που έχουν κωδικοποιηθεί (συμπιεστεί) βάσει του προτύπου κωδικοποίησης MPEG1 του οποίου μια συνοπτική περιγραφή δίνεται στο υποκεφάλαιο 3.4.

Το δεύτερο μέρος της εργασίας που καλύπτει το θέμα εξαγωγής χωρικών και χρονικών χαρακτηριστικών των σκηνών περιλαμβάνει τα κεφάλαια 6, 7 και 8. Ειδικότερα στο κεφάλαιο 6, παρουσιάζεται το θέμα γενικότερα της τμηματοποίησης εικόνων για την εξαγωγή αντικειμένων της οποίας μια εφαρμογή θα μπορούσε να είναι η χρήση των αντικειμένων για τη δεικτοδότηση εικόνων/video. Στη συνέχεια περιγράφονται δύο μέθοδοι που η πρώτη αφορά στην εξαγωγή και η δεύτερη στην παρακολούθηση των

αντικειμένων στο χρόνο (κεφ. 7,8 αντίστοιχα).

Στο υποκεφ. 3.2 που ακολουθεί δίνονται οι απαιτήσεις των εφαρμογών για τις οποίες σχεδιάστηκαν οι διεργασίες συμπλήρωσης της δομής δεδομένων και οι λόγοι που οδήγησαν στις επιλογές που έγιναν για μια υλοποίηση που να καλύπτει πληρέστερα αυτές τις απαιτήσεις.

3.2 Βασικές Αρχές Σχεδιασμού της Υλοποίησης

Το σχήμα παραλληλίας στην ανίχνευση αλλαγής σκηνών και την εξαγωγή των αντιπροσωπευτικών εικόνων με την ταυτόχρονη συμπλήρωση του μοντέλου δεδομένων του video που συνοπτικά περιγράφηκε στην εισαγωγή του κεφαλαίου, είναι ιδανικό για εφαρμογές *πραγματικού χρόνου* όπου η ανάλυση του περιεχομένου πρέπει να γίνεται με τρόπο αποδοτικό ως προς την ακρίβεια των αποτελεσμάτων αλλά κυρίως σύντομο.

Ειδικότερα το σύστημα δεικτοδότησης που περιγράφεται εδώ, προορίζεται να χρησιμοποιηθεί σε εφαρμογές που απαιτούν την ανάλυση του video σε σκηνές και εικόνες-κλειδιά *κατά τη διάρκεια μετάδοσής του* και επομένως κάνουν αναγκαία τη χρήση κατά το δυνατό αυτόματων και σύντομων σε διάρκεια εκτέλεσης μεθόδων. Βάσει αυτών των απαιτήσεων και αγνοώντας προς στιγμήν το σημαντικό ρόλο του μοντέλου δεδομένων, κατά το σχεδιασμό των αλγορίθμων ανίχνευσης ασυνεχειών και εξαγωγής αντιπροσωπευτικών εικόνων έχουν ληφθεί υπόψη οι παρακάτω παράγοντες:

- Αποφυγή της ‘‘περιττής’’ αποσυμπίεσης των εικόνων.
- Χρήση χαρακτηριστικών των εικόνων τα οποία από τη μια εξάγονται γρήγορα και από την άλλη δίνουν αξιόπιστα αποτελέσματα κατά την εφαρμογή τους σε video ‘‘με έντονη’’ δραστηριότητα των αντικειμένων των σκηνών.
- Άμεση συμπλήρωση των πεδίων του μοντέλου δεδομένων που αντιστοιχούν στην πληροφορία που εξάγεται από την εφαρμογή των δύο αλγορίθμων. Με το χαρακτηρισμό ‘‘άμεση’’, εννοούμε την αποθήκευση της πληροφορίας στη βάση αμέσως μόλις εξαχθεί.
- Δυνατότητα επέκτασης των αλγορίθμων με τη χρήση (σε εφαρμογές που αυτό είναι εφικτό) επιπλέον χαρακτηριστικών αλλά και διαφορετικών μέσων πληροφορίας, όπως για παράδειγμα είναι ο ήχος, με στόχο τόσο την επαλήθευση της τμηματοποίησης όσο και την πιο πλήρη δεικτοδότηση του υλικού των πολυμέσων.
- Δυνατότητα επέκτασης της χρήσης του συστήματος σε μή συμπιεσμένο video.

Από τα παραπάνω γίνεται σαφές ότι οι αλγόριθμοι ανίχνευσης ασυνεχειών και εξαγωγής αντιπροσωπευτικών εικόνων συνδέονται άμεσα με το μοντέλο δεδομένων που καλούνται να συμπληρώσουν βάσει του περιεχομένου του video. Επομένως, και η δομή του μοντέλου θα πρέπει να είναι τέτοια που να επιτρέπει την αποδοτική εκτέλεση των

δύο διεργασιών. Το μοντέλο δεδομένων που περιγράφεται στη συνέχεια προβλέπεται στο μέλλον να αποτελέσει το πρότυπο σχήμα αποθήκευσης του περιεχομένου του video και ικανοποιεί σε μεγάλο βαθμό αυτές τις απαιτήσεις.

3.3 Ιεραρχικό Μοντέλο Δεικτοδότησης Δεδομένων Video

Το μοντέλο δεδομένων που χρησιμοποιήθηκε για τις ανάγκες της εργασίας είναι ένα ιεραρχικό μοντέλο δεικτοδότησης δεδομένων, του οποίου ένα στιγμιότυπο φαίνεται στο σχήμα 3.1. Το μοντέλο ακολουθεί γενικότερα το πρότυπο μοντέλο δεικτοδότησης των δεδομένων video του MPEG7 και έτσι μετατρέπεται εύκολα ώστε να είναι συμβατό με εφαρμογές δεικτοδότησης video υψηλότερου επιπέδου, οι οποίες ακολουθούν το ίδιο πρότυπο.

Οπως φαίνεται στο σχήμα, η ιεραρχική παράσταση μιας ακολουθίας εικόνων διακρίνεται σε τρία επίπεδα:

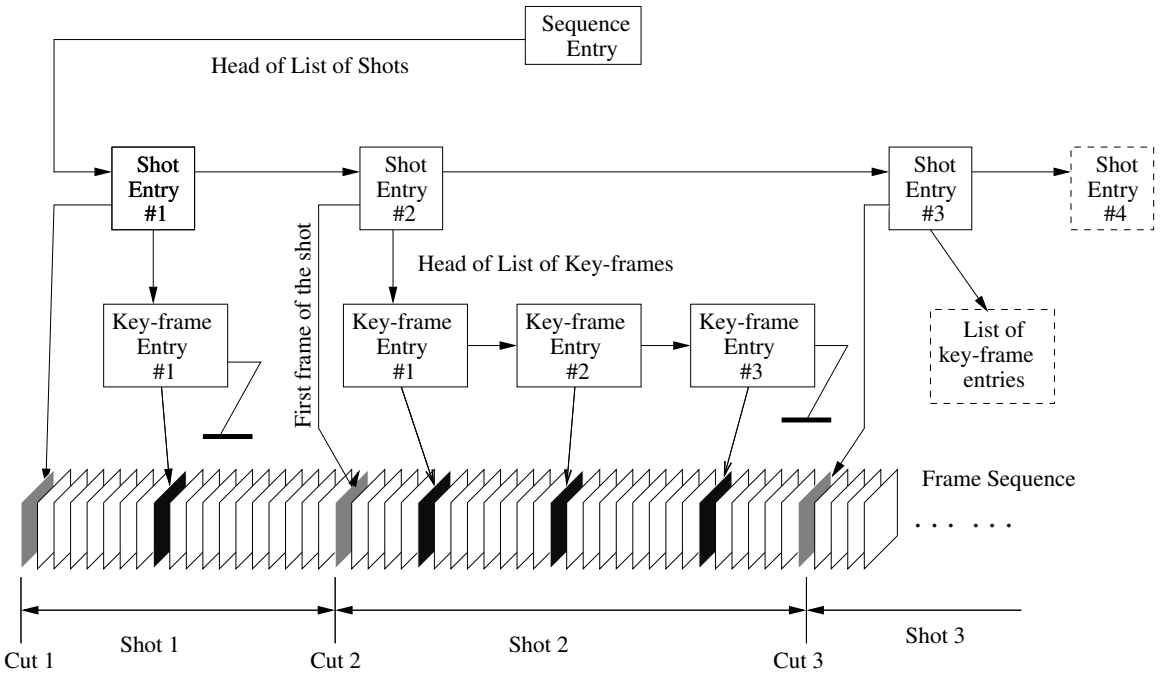
α. Επίπεδο Ακολουθίας: αποτελείται από μια δομή στην οποία διατηρείται πληροφορία για τα επί μέρους τμήματα του video όπως εξάγονται από τον αλγόριθμο ανίχνευσης αλλαγής σκηνών. Αυτή η είσοσος μπορεί να είναι μέλος μιας λίστας ακολουθιών οι οποίες δεικτοδοτούνται από ένα δείκτη δεδομένων υψηλότερου επιπέδου (για παράδειγμα, στο πρότυπο μοντέλο δεικτοδότησης του MPEG7, ένα σύνολο ακολουθιών μπορεί να είναι τα στοιχειώδη συστατικά ενός ‘‘θέματος’’ (topic) ή ‘‘εγγράφου’’ (document)). Ενα από τα πεδία της εισόδου ακολουθίας είναι ένα δείκτης στη λίστα των εισόδων των σκηνών που περιγράφονται αμέσως μετά.

β. Επίπεδο Σκηνής: το επίπεδο αυτό περιλαμβάνει μια λίστα των εισόδων οι οποίες κρατούν την πληροφορία των σκηνών της ακολουθίας όπως είναι ο αριθμός εικόνας (frame number) της πρώτης εικόνας της σκηνής και η λίστα των αντιπροσωπευτικών της εικόνων. Σε αντίθεση με το πρότυπο μοντέλο του MPEG7, οι σκηνές δεν υποδιαιρούνται σε επιμέρους τμήματα. Με άλλα λόγια, η λίστα των σκηνών βρίσκεται ιεραρχικά ακριβώς πάνω από το χαμηλότερο επίπεδο της ιεραρχίας, το οποίο είναι το ίδιο το video. Αυτό έχει ως συνέπεια, οι είσοδοι αυτού του επιπέδου να διατηρούν επιπλέον τη χρονική πληροφορία των σκηνών όπως είναι ο τύπος της κίνησης της κάμερας.

γ. Επίπεδο Εικόνας: είναι το χαμηλότερο επίπεδο της ιεραρχίας το οποίο αποτελείται από τις εικόνες της ακολουθίας.

Πρέπει να σημειωθεί ότι η λίστα των αντιπροσωπευτικών εικόνων δεν αντιστοιχεί σε κάποιο επίπεδο του μοντέλου. Είναι απλά μια λίστα δεικτοδότησης του περιεχομένου της σκηνής η οποία περιγράφει συνοπτικά τη χρονική του μεταβολή κατά τη διάρκειά της. Κάθε αντιπροσωπευτική εικόνα διατηρεί πληροφορία δεικτοδότησης της εικόνας

στην οποία αναφέρεται. Αυτή η πληροφορία συνίσταται κατά κύριο λόγο σε χαρακτηριστικά ομοιότητας των εικόνων που χρησιμοποιούνται για την ανάκτησή τους.



Σχήμα 3.1: Το Μοντέλο Δεικτοδότησης του Video

Οι διαδικασίες ανίχνευσης αλλαγής σκηνών και εξαγωγής εικόνων-κλειδιών παρέχουν, η μεν πρώτη την πληροφορία για την πρώτη εικόνα (όρια) της σκηνής η δε δεύτερη, τις αντιπροσωπευτικές εικόνες της σκηνής, αμέσως μόλις αυτή η πληροφορία εξαχθεί, όπως θα συζητηθεί στα κεφάλαια 4 και 5.

Σε ότι αφορά την εξαγωγή των αντικειμένων της σκηνής και των χωρικών και χρονικών χαρακτηριστικών τους, όπως περιγράφεται στα κεφάλαια 7 και 8, αυτά είναι δυνατό να αποθηκευθούν στη δομή που αντιστοιχεί στις αντιπροσωπευτικές εικόνες, προσδίδοντάς τους μια αφαιρετική ιεραρχική διάσταση. Η διάταξη αυτή ταιριάζει στην ανάκτηση του περιεχομένου του video ακολουθώντας τη δομική προσέγγιση όπως ήδη αναφέρθηκε στο κεφάλαιο της Γενικής Εισαγωγής. Εξίσου σημαντική θα μπορούσε να είναι αυτή η διάταξη στην επεξεργασία εκ των υστέρων (post-production) των σκηνών του video.

3.4 Κωδικοποίηση MPEG-1

Το MPEG (Moving Picture Experts Group) είναι ένα πρότυπο συμπίεσης video, με το οποίο μπορούν να επιτευχθούν υψηλοί ρυθμοί συμπίεσης και γι' αυτό έχει γίνει δημοφιλές σε πολλές εφαρμογές. Τα τελευταία χρόνια, πράγματι υπάρχει η τάση τα δεδομένα του

video να αποθηκεύονται σε MPEG format. Γενικά μιλώντας, η συμπίεση των δεδομένων που επιτυγχάνεται από το MPEG, βασίζεται στη χρήση της υπάρχουσας τεχνολογίας εκτίμησης της κίνησης (motion compensation technology) η οποία έχει εισαχθεί επιπλέον σ' αυτό το πρότυπο και το διακρίνει από τις παραδοσιακές μεθόδους επέκτασης της συμπίεσης στατικών εικόνων (still images) σε video. Άλλωστε αρκετές από τις βασικές προσεγγίσεις συμπίεσης στατικών εικόνων εμπεριέχονται και στο MPEG1.

Σύμφωνα με το Διεθνές Πρότυπο ISO/IEC 11178 [19], το video το οποίο έχει κωδικοποιηθεί με βάση τον αλγόριθμο κωδικοποίησης MPEG1 είναι δυνατό να περιλαμβάνει έναν ή περισσότερους από τους παρακάτω τύπους εικόνων:

- I-εικόνες (intra-coded frames) οι οποίες κωδικοποιούνται ανεξάρτητα, χωρίς καμιά αναφορά σε άλλες εικόνες του video. Η κωδικοποίησή τους βασίζεται στη μείωση του *χωρικού πλεονασμού* που επιτυγχάνεται με την εφαρμογή κατά σειρά

  ▷ του διακριτού μετασχηματισμού συνημιτόνου (DCT)

  ▷ κβαντισμού

  ▷ κωδικοποίησης μήκους κατά την εκτέλεση (run length encoding)

  ▷ και κωδικοποίησης εντροπίας.

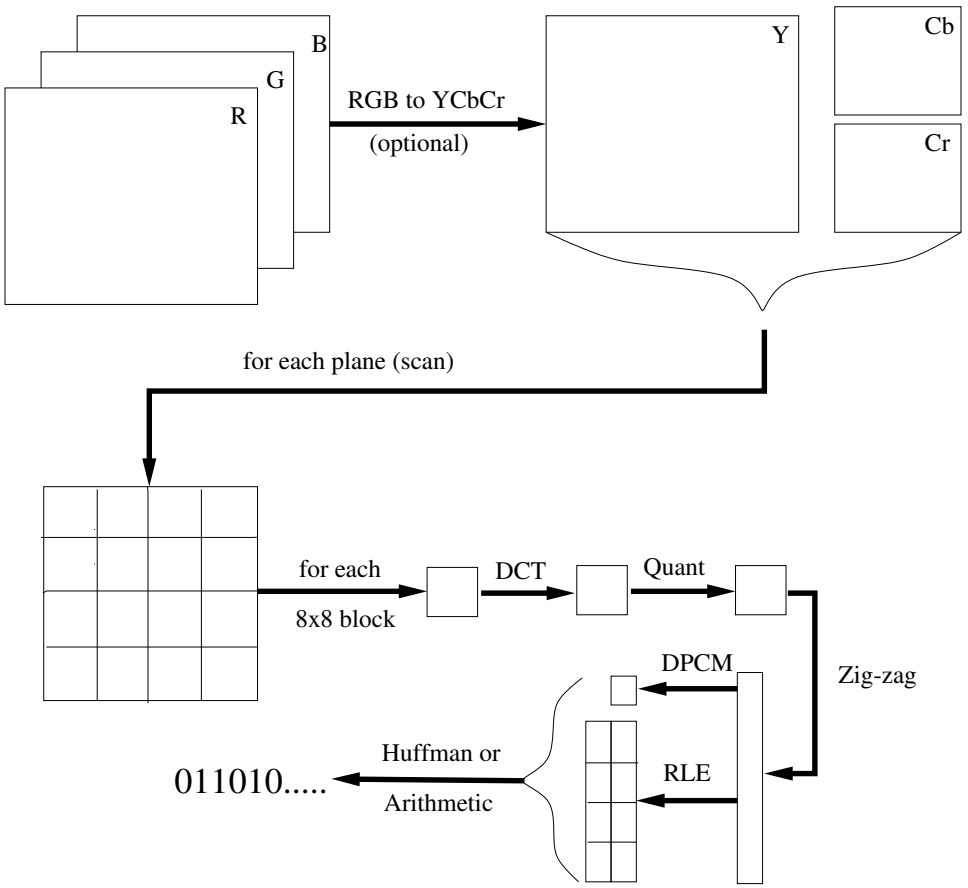
  σε επίπεδο block διαστάσεων 8×8. Η διαδικασία συμπίεσης φαίνεται στο σχ. 3.2. Στο σχήμα, η είσοδος σ' αυτή τη διαδικασία είναι έγχρωμες εικόνες σε μορφή YCbCr. Αν και κάτι τέτοιο δεν απαιτείται γενικά, όπως αναφέρεται στο τέλος της ενότητας ο (απο)κωδικοποιητής που χρησιμοποιήθηκε δέχεται τις εικόνες σ' αυτή τη μορφή.

- P-εικόνες (predicative coded frames) οι οποίες κωδικοποιούνται χρησιμοποιώντας εκτίμηση κίνησης από την προηγούμενη I- ή P-εικόνα, ώστε να επιτευχθεί μείωση του *χρονικού πλεονασμού* που αποτελεί και το στόχο του B τύπου εικόνων.

- B-εικόνες (bidirectional predictive coded frames) για την κωδικοποίηση των οποίων εφαρμόζεται εκτίμηση κίνησης με αναφορά τόσο στην προηγούμενη όσο και στην επόμενη I- ή P-εικόνα.

- D-εικόνες (DC-coded frames), τα οποία κωδικοποιούνται με χρήση μόνο του DC συντελεστή των block τους, που αντιστοιχεί στην πληροφορία χαμηλών συχνοτήτων που εξάγεται από τον DCT και αντιστοιχεί στη μέση τιμή του block. Χρησιμοποιούνται σπάνια αφού δεν επιτρέπεται να συνυπάρχουν με τις I-, B-, P-εικόνες.

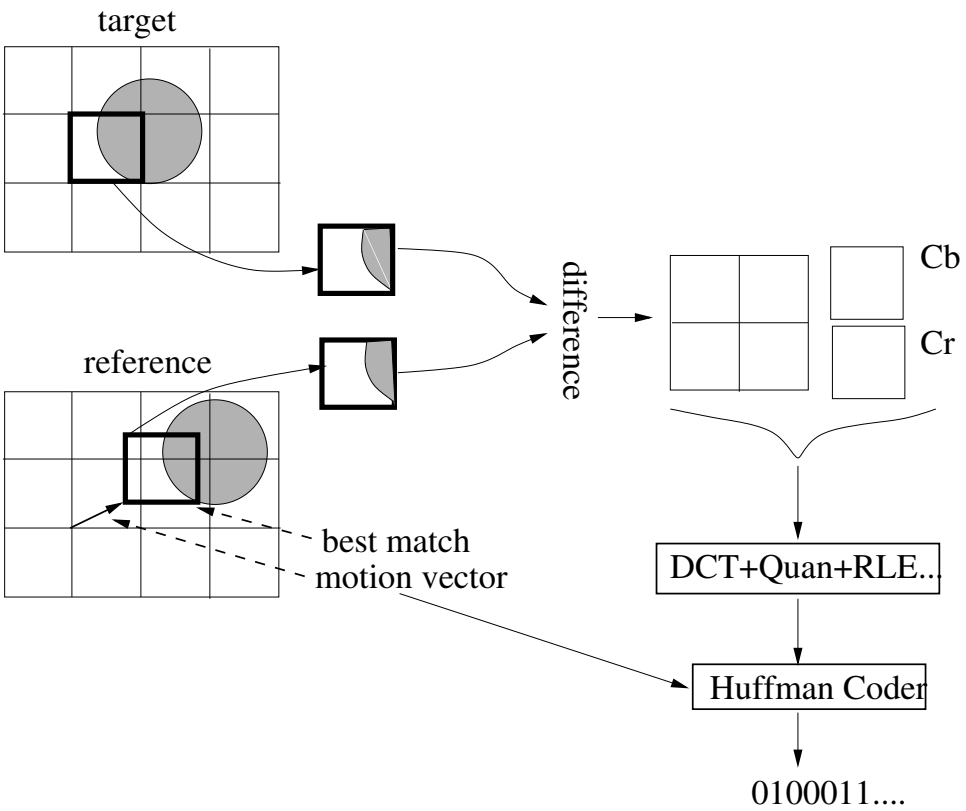
Στην πραγματικότητα, οι αποφάσεις για τον τρόπο κωδικοποίησης λαμβάνονται σε επίπεδο block: η εικόνα υποδιαιρείται σε επιμέρους block 8×8 τα οποία ομαδοποιούνται ανά τέσσερα στα λεγόμενα macro-block διαστάσεων 16×16. Τα block αποτελούν τις μονάδες εφαρμογής των αλγορίθμων συμπίεσης που αναφέρθηκαν κατά την περιγραφή



Σχήμα 3.2: Κωδικοποίηση I-εικόνων (Πηγή: http://fas.sfu.ca/cs/undergrad/CourseMaterials)

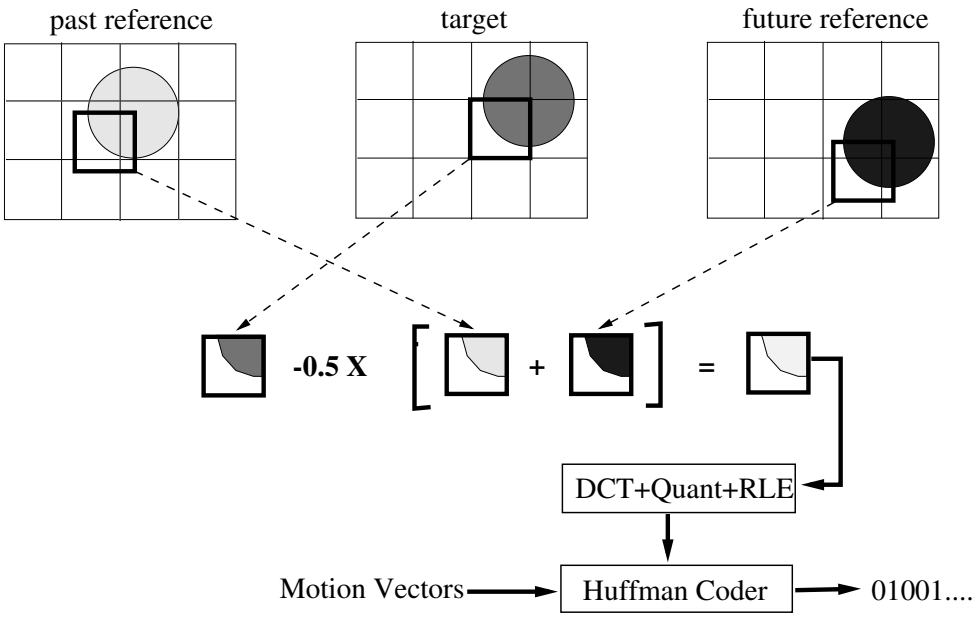
των I-εικόνων. Όσον αφορά την εκτίμηση της ταχύτητας στην περίπτωση των P-, B-εικόνων αυτή εφαρμόζεται στα macro-block τους, με χρήση block matching τεχνικών.

Στην περίπτωση των P-εικόνων η απόφαση που λαμβάνεται σε επίπεδο macro-block είναι το εάν αυτό θα κωδικοποιηθεί όπως τα macro-block των I-εικόνων (intra-coded macro-block) (σχ. 3.2) ή θα χρησιμοποιηθεί η εκτίμηση της ταχύτητας που έχει εξαχθεί. Στην πρώτη περίπτωση, η είσοδος της κωδικοποίησης είναι η πληροφορία της έντασης και του χρώματος του macro-block, όπως πριν. Αντίθετα στη δεύτερη, η πληροφορία που κωδικοποιείται είναι το διάνυσμα ταχύτητας καθώς και το αντίστοιχο block του απολύτου λάθους στην εκτίμηση της τιμής των σημείων, όπως υπολογίζονται από τη μέθοδο εκτίμησης ταχύτητας του macro-block (σχ. 3.3). Η απόφαση βασίζεται σε στατιστικά κριτήρια μέτρησης του κόστους κωδικοποίησης των δύο μεθόδων σε bits.



Σχήμα 3.3: Κωδικοποίηση P-εικόνων (Πηγή: http://fas.sfu.ca/cs/undergrad/CourseMaterials)

Κατ' αναλογία, στην περίπτωση των B-εικόνων, θα πρέπει να καθοριστεί εάν θα χρησιμοποιηθεί η εκτίμηση της ταχύτητας με αναφορά στην προηγούμενη και την επόμενη I-, P-εικόνα, ή μόνο η εκτίμηση που αναφέρεται στη μία από τις δύο όπως φαίνεται στο σχήμα 3.4 ή τέλος αν το macro-block θα κωδικοποιηθεί ως έχει (intra-coded macro-block). Η απόφαση όπως και πριν, βασίζεται σε μετρήσεις του κόστους κωδικοποίησης της κάθε μεθόδου στο συγκεκριμένο macro-block. Είναι επομένως δυνατό, block των P-, B-εικόνων να έχουν κωδικοποιηθεί με τη μέθοδο κωδικοποίησης των λεγομένων intra-coded macro-block που είναι τα μόνα που απαντώνται στις I-εικόνες.



Σχήμα 3.4: Κωδικοποίηση Β-εικόνων (Πηγή: http://fas.sfu.ca/cs/undergrad/CourseMaterials)

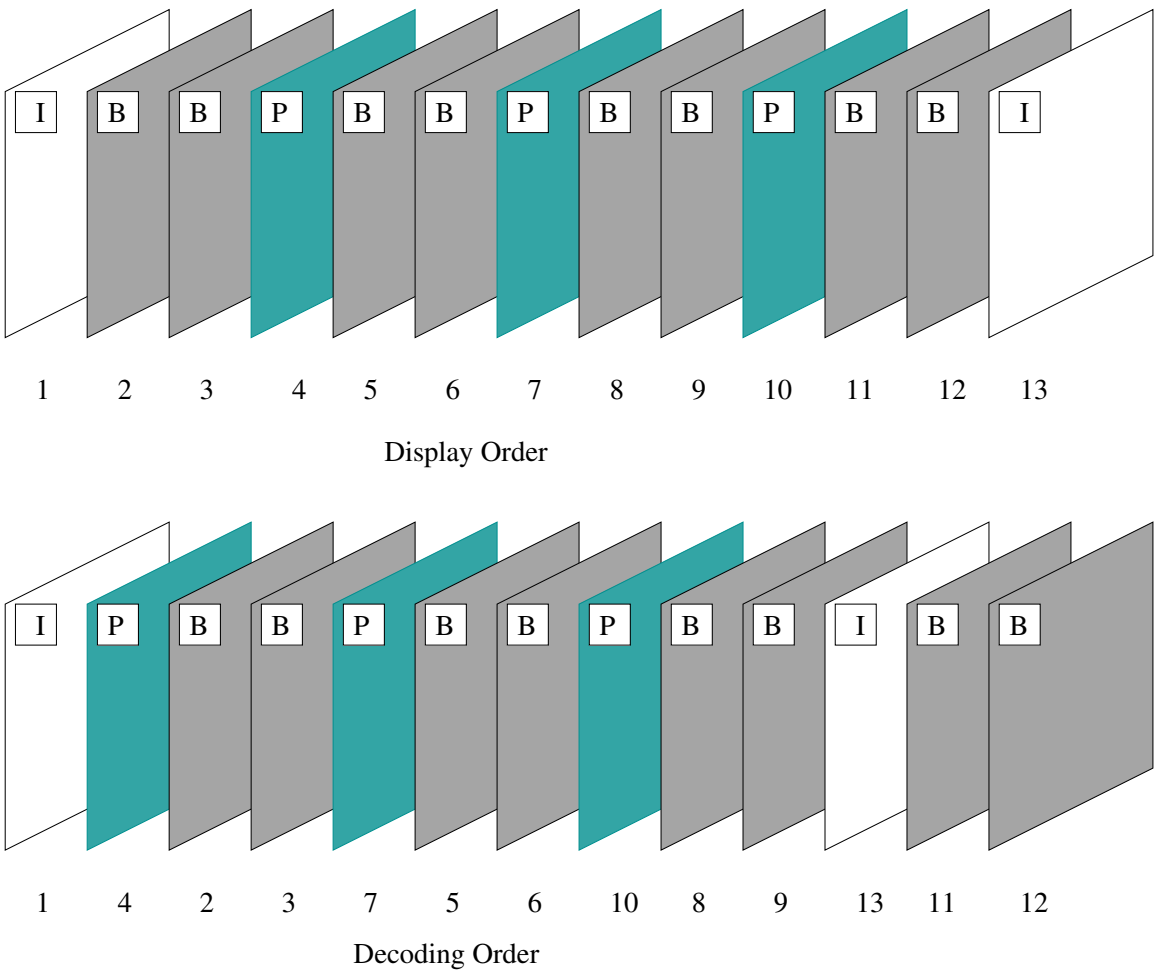
Προφανώς, το MPEG1 video θα πρέπει να περιλαμβάνει τουλάχιστον Ι-εικόνες. Η αναλογία του αριθμού των Ι-, Ρ-, Β-εικόνων καθώς και η ακολουθία εμφάνισής τους προσδιορίζονται εξ αρχής. Η τρέχουσα τάση η οποία φαίνεται ότι ικανοποιεί και τις απαιτήσεις μετάδοσης είναι η αναλογία 1 : 3 : 8 των αριθμών των Ι-, Ρ-, Β-εικόνων αντίστοιχα. Η σειρά εμφάνισης (display order) και η σειρά αποκωδικοποίησης (decoding order), των εικόνων που έχουν κωδικοποιηθεί μ' αυτή την αναλογία φαίνεται στο σχ. 3.5.

Οπως θα δούμε στο κεφάλαιο που ακολουθεί, αρκετές φορές οι αποφάσεις που έχουν ληφθεί κατά την κωδικοποίηση του video, χρησιμοποιούνται για την ανίχνευση ασυνεχειών.

Τέλος, πρέπει να τονισθεί ότι οι αποκωδικοποιημένες εικόνες δίνονται από τον αλγόριθμο αποκωδικοποίησης στο σύστημα χρώματος YCbCr με υποδειγματοληψία των συνιστωσών του χρώματος κατά 2 σε κάθε διάσταση της εικόνας (το ίδιο σχήμα χρησιμοποιείται και από τον αλγόριθμο συμπίεσης στατικών εικόνων JPEG). Οι λόγοι που οδήγησαν σ' αυτή την επιλογή είναι οι εξής:

- θεωρητικά τουλάχιστον, οι τρείς συνιστώσες του είναι στατιστικά ασυσχέτιστες
- και επιπλέον η σημαντική πληροφορία συγκεντρώνεται στον άξονα Υ της φωτεινότητας. Σ' αυτή την ιδιότητα του YCbCr στηρίζεται άλλωστε η απόφαση υποδειγματοληψίας των συνιστωσών του χρώματος.

Στην πράξη το όφελος είναι ότι επιτυγχάνεται σημαντική οικονομία σε μνήμη, υπολογιστικό χρόνο και χώρο αποθήκευσης λόγω της υποδειγματοληψίας των συνιστωσών του χρώματος, χωρίς σημαντική απώλεια πληροφορίας των εικόνων.



Σχήμα 3.5: Σειρά εμφάνισης και αποκωδικοποίησης εικόνων video σε μορφή MPEG1

Κεφάλαιο 4

Ανίχνευση Αλλαγής Σκηνών του Video

4.1 Εισαγωγή-Ανασκόπηση

Η ανίχνευση αλλαγής σκηνών, είναι συνήθως το πρώτο βήμα στην ανάλυση του video. Στόχος της είναι η τμηματοποίηση του video σε επιμέρους σκηνές (shots) οι οποίες χαρακτηρίζονται από ενιαία αυτόνομη δράση στο χώρο και στο χρόνο.

4.1.1 Τύποι Μεταβάσεων Σκηνών

Ένας τόσο γενικός ορισμός της έννοιας ‘‘σκηνή’’, αφήνει πολλά περιθώρια ποιοτικής ερμηνείας των αποτελεσμάτων αλγορίθμων που προσπαθούν να καθορίσουν τα χρονικά όρια τους. Το πρόβλημα γίνεται ακόμα πιο σύνθετο, εάν σκεφτεί κανείς την ποικιλία πιθανών τρόπων εναλλαγής των σκηνών που εμφανίζεται στην τεχνολογία του video. Οι μεταβάσεις αυτές προκύπτουν κατά τη διαδικασία δημιουργίας video από τη συνένωση σκηνών και διακρίνονται γενικά σε δύο βασικές κατηγορίες:

- *Απότομες (abrupt) μεταβάσεις* που προκύπτουν από τη συνένωση ανεξαρτήτων ως προς το περιεχόμενό τους θεμάτων video, κατά τη διαδικασία κατασκευής του video. Χαρακτηριστικό παράδειγμα αποτελεί το video ειδήσεων που μεταδίδεται μέσω δορυφόρων και το οποίο περιλαμβάνει σημαντικό αριθμό διαδοχικών και διαφορετικών θεμάτων βάσει των γεγονότων που έχει καταγράψει η κάμερα σ’ όλο τον κόσμο. Οι μεταβάσεις αυτές δεν περιγράφονται από κάποιο μαθηματικό μοντέλο και γι’ αυτό κατά την ανίχνευσή τους αναζητάται κατά τη διάρκεια προβολής του video μια ποσοτικά μεγάλη μεταβολή σε επίπεδο σύγκρισης διαδοχικών εικόνων. Η ασυνέχεια (cut) αυτή, ορίζεται ως το σημείο μετάβασης από τη μια σκηνή στην επόμενή της.
- *Βαθμιαίες (gradual) μεταβάσεις*, οι οποίες προκύπτουν μεν από τη συνένωση ανεξαρτήτων σε περιεχόμενο σκηνών όπως και πριν, αλλά επιπλέον στα όρια τους

έχουν εφαρμοστεί διάφορες γνωστές διαδικασίες μεταβολής του περιεχομένου του video (video editing), με σκοπό ακριβώς την εξάλειψη των απότομων μεταβάσεων από τη μία σκηνή στην επόμενή της. Γνωστές τεχνικές που οδηγούν σε βαθμιαίες μεταβάσεις είναι οι εξής:

- ▷ *Βαθμιαία Εξαφάνιση (fade out) ή Εμφάνιση (fade in)* της σκηνής. Οι όροι αυτοί δηλώνουν, ο μεν πρώτος, την οπτική διαδικασία που οδηγεί στο προοδευτικό ‘‘σβήσιμο’’ ή ‘‘σκοτείνιασμα’’ της σκηνής, έως ότου οι τελευταίες εικόνες γίνουν εντελώς σκοτεινές, ο δε δεύτερος την προοδευτική μετάβαση της σκηνής από το σκοτάδι στο φως.
- ▷ *Διάλυση (dissolve)* της σκηνής, η οποία προκύπτει από το συνδυασμό της προοδευτικής εξαφάνισης μιας σκηνής, με την ταυτόχρονη προοδευτική εμφάνιση της επόμενής της.

Εκτός των οπτικών εφέ, βαθμιαίες μεταβάσεις προκύπτουν στην πράξη και από τα λεγόμενα χωρικά εφέ (spatial effects) όπως για παράδειγμα είναι το λεγόμενο wipe κατά το οποίο μία γραμμή (κάθετη, οριζόντια ή διαγώνια) κινείται κατά μήκος της εικόνας, με τη νέα σκηνή να εμφανίζεται σταδιακά πίσω της.

4.1.2 Προηγούμενες Προσεγγίσεις

Αρκετές τεχνικές έχουν προταθεί τα τελευταία χρόνια στη βιβλιογραφία, για την ανίχνευση τόσο των απότομων όσο και των βαθμιαίων μεταβάσεων μεταξύ διαδοχικών σκηνών. Μια πρόσφατη ανασκόπηση αυτών των μεθόδων περιλαμβάνεται στο [20], όπως επίσης και στα [1] [2], τα οποία όπως έχει ήδη αναφερθεί, επιχειρούν μια γενικότερη ανασκόπηση της τεχνολογίας δεικτοδότησης του video.

Ειδικότερα, οι τεχνικές που αναφέρονται στην ανίχνευση απότομων αλλαγών βασίζονται στη σύγκριση δύο διαδοχικών (ή σχετικά κοντινών) εικόνων του video θεωρώντας ένα ή περισσότερα χαρακτηριστικά τους. Προφανώς ο τύπος του χαρακτηριστικού καθορίζει και το κριτήριο βάσει του οποίου αποφασίζεται αν έχει ανιχνευθεί ή όχι, κάποια ασυνέχεια ή μετάβαση. Αυτό που αναμένεται στην περίπτωση των απότομων αλλαγών, είναι το κριτήριο ανομοιότητας να δώσει μια ‘‘πολύ μεγάλη’’ τιμή γεγονός που αποτελεί μια ποιοτική ένδειξη της ύπαρξης ασυνέχειας στο χρονικό διάστημα μεταξύ των δύο εξεταζόμενων εικόνων. Η διαφορά των διαφόρων τεχνικών, έγκειται κυρίως στο πεδίο στο οποίο εφαρμόζονται. Βάσει αυτής της διαφοροποίησης, οι τεχνικές αυτές διακρίνονται σε:

- Τεχνικές που στηρίζονται σε χαρακτηριστικά μή κωδικοποιημένων ή πλήρως αποκωδικοποιημένων εικόνων. Η κατηγορία αυτή περιλαμβάνει μεθόδους όπως
  - ▷ σύγκριση ιστογραμμάτων έντασης/χρώματος και
  - ▷ τεχνικές σύγκρισης σε επίπεδο σημείων, block ή συνδυασμών τους.

- Τεχνικές που εφαρμόζονται σε χαρακτηριστικά από το πεδίο κωδικοποίησης-συμπίεσης του video όπως
  - ▷ οι συντελεστές DC ή ένας αριθμός συντελεστών του DCT των block των εικόνων και
  - ▷ για MPEG-κωδικοποιημένα video, οι αποφάσεις που έχουν ληφθεί κατά την κωδικοποίηση των macro-block, των P-, B-εικόνων.
- Υβριδικές τεχνικές όπως για παράδειγμα η εφαρμογή σύγκρισης ιστογραμμάτων στο πεδίο των συντελεστών DC των block των εικόνων.

Για τη σύγκριση των ιστογραμμάτων χρησιμοποιούνται συνήθως στατιστικά κριτήρια όπως οι εξισώσεις (2.4)-(2.7). Χαρακτηριστικό της ευρείας χρήσης των ιστογραμμάτων, είναι ότι έχουν χρησιμοποιηθεί επίσης και για την ανίχνευση βαθμιαίων αλλαγών σκηνών από μια τεχνική που ονομάζεται twin comparison (βλ. αναφορά [70] στο [1]). Η ευρεία χρήση τους δικαιολογείται από το ότι παραμένουν αναλλοίωτα (invariant)

- σε σχετικά μεγάλες κινήσεις των αντικειμένων και της κάμερας και
- σε μικρές διαφορές στο background από εικόνα σε εικόνα,

ενώ υπολογίζονται απλά και γρήγορα. Παρόλα αυτά, σε περιπτώσεις όπου η κίνηση είναι σημαντική, αποτυγχάνουν δίνοντας λάθος ανιχνεύσεις. Στο [21], συγκρίνονται ως προς την αποτελεσματικότητά τους στην ανίχνευση ασυνεχειών, τα ιστογράμματα διαφόρων συστημάτων χρώματος, χρησιμοποιώντας για τη σύγκρισή τους γνωστά κριτήρια.

Αρκετά ενδιαφέρουσες είναι επίσης οι τεχνικές που βασίζονται στις αποφάσεις που έχουν ληφθεί κατά την κωδικοποίηση των P-, B-εικόνων, λόγω κυρίως του χρόνου που εξοικονομείται από την αποφυγή της πλήρους αποκωδικοποίησης των εικόνων. Μια τεχνική που ανήκει σ' αυτή την κατηγορία περιγράφεται στο [22]. Το πρόβλημα αυτών των μεθόδων είναι ότι οι αποφάσεις του τρόπου κωδικοποίησης των macro-block, παίρνονται από τον κωδικοποιητή με στόχο τη βελτιστοποίηση της κωδικοποίησης και γι' αυτό δεν είναι ασφαλές να χρησιμοποιούνται για τη μέτρηση της ομοιότητας των P-, B-εικόνων με τις εικόνες αναφοράς τους.

Τέλος, αξίζει να αναφερθεί, η ενδιαφέρουσα προσέγγιση που ακολουθείται στο [23] για τον προσδιορισμό μεταβάσεων και των δύο τύπων. Το ενδιαφέρον είναι ότι η ανίχνευση βασίζεται στο ποσοστό μεταβολής των ακμών, από εικόνα σε εικόνα. Αυτό που παρατηρήθηκε είναι ότι, κατά τη διάρκεια της μετάβασης, οι ακμές έντασης της νέας σκηνής εισέρχονται στην εικόνα σε εντελώς διαφορετική θέση, σε σχέση με τη θέση των ακμών της προηγούμενης σκηνής. Έτσι, οι απότομες αλλαγές σκηνής, τα fades, και τα dissolves, ανιχνεύονται μελετώντας τις τιμές μεγίστου του ποσοστού μεταβολής των ακμών, σε ένα σταθερό παράθυρο εικόνων. Τα wipes προσδιορίζονται από την κατανομή των εισερχομένων και εξερχομένων σημείων ακμής (edge pixels). Για τη μείωση των λάθος ανιχνεύσεων, υπολογίζεται αρχικά η ολική κίνηση (global motion)

των εικόνων. Από τα αποτελέσματα εξάγεται το συμπέρασμα ότι πράγματι ο αλγόριθμος συμπεριφέρεται σταθερά σε όλους τους τύπους μεταβάσεων. Το μειονέκτημα της μεθόδου είναι ο μεγάλος μέσος χρόνος επεξεργασίας των εικόνων κάτι που καθιστά δυσχερή την εφαρμογή της σε video μεγάλης διάρκειας και δικαιολογεί την ερευνητική τάση στο πεδίο της τμηματοποίησης video, για όσο το δυνατό πιο απλές μεθόδους σύγκρισης των εικόνων.

Τα ίδια συμπεράσματα ισχύουν και για τη μέθοδο που περιγράφεται στο [24], όπου η ανίχνευση και κατηγοριοποίηση τύπου των μεταβάσεων, προκύπτουν από τον ταυτόχρονο προσδιορισμό του τύπου της κίνησης της κάμερας. Μ' αυτό τον τρόπο μειώνεται σημαντικά ο αριθμός των λάθος ανιχνεύσεων, το υπολογιστικό κόστος όμως παραμένει υψηλό.

4.2 Ο Αλγόριθμος Ανίχνευσης Ασυνεχειών

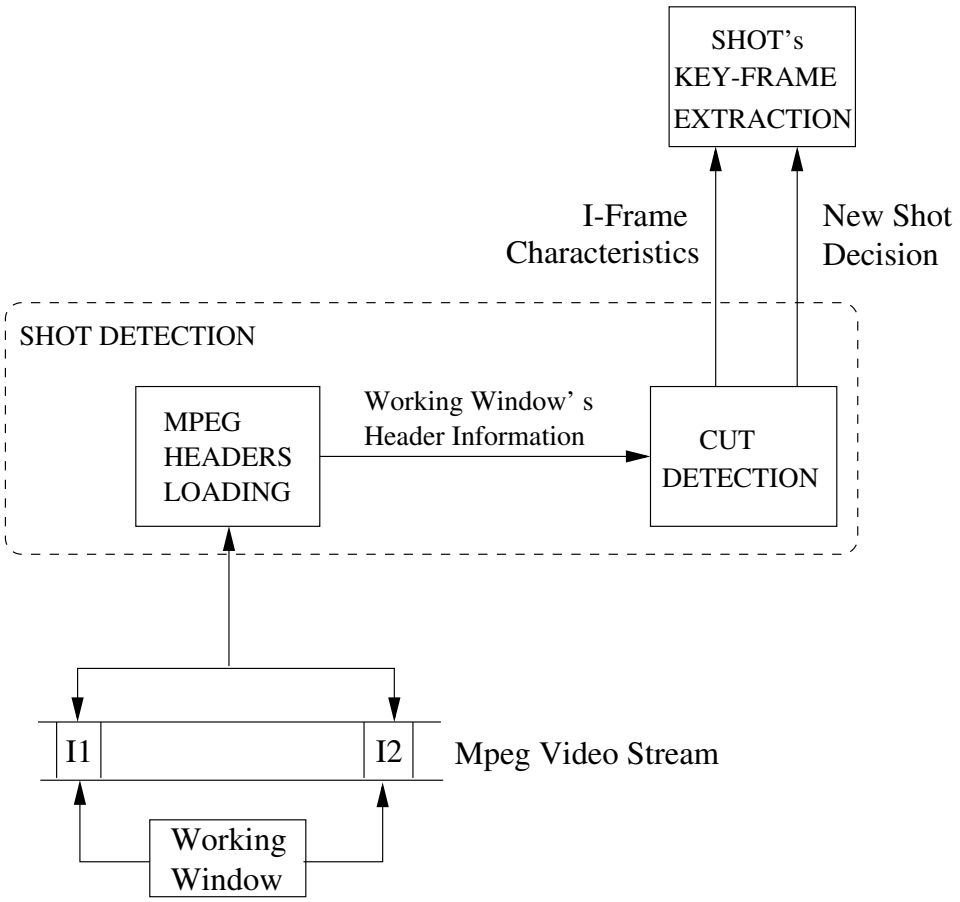
Ο αλγόριθμος ανίχνευσης ασυνεχειών που παρουσιάζεται στη συνέχεια εφαρμόζεται στο επίπεδο πλήρως αποκωδικοποιημένων εικόνων. Η είσοδος του είναι το video το οποίο πρόκειται να τμηματοποιηθεί, σε μορφή MPEG1. Για τη σύγκριση των εικόνων χρησιμοποιήθηκαν τα ιστογράμματα έντασης/χρώματος, ενώ για να αποφευχθεί η συνεχής αποκωδικοποίηση των εικόνων υλοποιήθηκε μια top-down μέθοδος προσπέλασης τους. Οι επιλογές σχεδιασμού του αλγορίθμου, έγιναν με βάση της ανάγκες εφαρμογών πραγματικού χρόνου όπως περιγράφεται στο προηγούμενο κεφάλαιο.

4.2.1 Σύνδεση με το Σύστημα Δεικτοδότησης του Video

Όπως αναφέρθηκε ήδη η συνολική προσέγγιση χωρίζεται σε δύο κύρια μέρη, την ανίχνευση ασυνεχειών-αλλαγής σκηνών του video και την εξαγωγή αντιπροσωπευτικών εικόνων της κάθε σκηνής. Το σχήμα 4.1 δείχνει ακριβώς αυτή τη δομή του συνολικού αλγορίθμου δίνοντας έμφαση στις λεπτομέρειες της ανίχνευσης αλλαγής σκηνών.

Όπως φαίνεται στο σχήμα 4.1, αρχικά ''διαβάζονται'' από το κωδικοποιημένο αρχείο εισόδου MPEG1 οι επικεφαλίδες (headers) των εικόνων που περιλαμβάνονται στο υπό εξέταση διάστημα. Κάθε επικεφαλίδα περιλαμβάνει εκτός των άλλων, πληροφορία για τη χρονική σειρά τους μέσα στο video καθώς και για τη μέθοδο που χρησιμοποιήθηκε για την κωδικοποίησης τους. Αυτές οι δύο παράμετροι καθορίζουν τη σειρά προσπέλασης των ζευγών εικόνων πάνω στις οποίες εφαρμόζεται το κριτήριο ανίχνευσης ασυνεχειών που αποτελεί το δεύτερο βήμα του αλγορίθμου, όπως φαίνεται στο σχ. 4.1.

Κλείνοντας την περιγραφή της διεργασίας ανίχνευσης ασυνεχειών, θα πρέπει να τονιστεί η σχέση ''παραλληλίας'' της με τη διεργασία εξαγωγής αντιπροσωπευτικών εικόνων όπως φαίνεται και στο σχ. 4.1: με το τέλος της ανίχνευσης ασυνέχειας τα χαρακτηριστικά των δύο I-frames που έχουν ήδη εξαχθεί για τις ανάγκες αυτής της διαδικασίας καθώς και η απόφαση για το αν ανιχνεύθηκε ή όχι νέα σκηνή, ''περνιούνται''



Σχήμα 4.1: Λεπτομέρειες Αλγορίθμου Ανίχνευσης Ασυνεχειών

για να χρησιμοποιηθούν επιπλέον από τη διεργασία εξαγωγής αντιπροσωπευτικών εικόνων όπως περιγράφεται στο επόμενο κεφάλαιο.

4.2.2 Μέτρηση (Αν)ομοιότητας Εικόνων του Video

Όπως αναφέρθηκε και στην εισαγωγή, τα ιστογράμματα τόσο της έντασης όσο και του χρώματος των εικόνων έχουν χρησιμοποιηθεί με σημαντική επιτυχία στην ανίχνευση ασυνεχειών.

Οι τεχνικές που βασίζονται στα ιστογράμματα υπολογίζουν ως διαφορά δύο εικόνων f_1 και f_2, την απόσταση των αντίστοιχων ιστογραμμάτων έντασης ή χρώματος H_1 και H_2. Ασυνέχεια στο δοσμένο διακριτό χρονικό διάστημα Δn ανιχνεύεται εφόσον η απόσταση $\delta_{CD}(f_1, f_2)$ των εικόνων f_n και $f_{n+\Delta n}$ του video, ξεπερνάει κάποιο κατώφλι t_{CD}.

Αρχικά θεωρήθηκαν μόνο τα ιστογράμματα έντασης. Σ' αυτή την περίπτωση για τη μέτρηση της απόστασης μεταξύ των εικόνων $\delta_{CD}(f_1, f_2)$ χρησιμοποιείται μία από τις εξισώσεις (2.4)-(2.7). Συμβολίζοντας ως δ_i την εξίσωση που έχει επιλεγεί, η απόσταση μεταξύ των δύο εικόνων γίνεται:

$$\delta_{CD}^{Y}(f_1, f_2) = \delta_i(\mathrm{H}_1^{\mathrm{Y}}, \mathrm{H}_2^{\mathrm{Y}}) \tag{4.1}$$

όπου $\mathrm{H_1}^{\mathrm{Y}}$,$\mathrm{H_2}^{\mathrm{Y}}$ είναι τα ιστογράμματα έντασης των εικόνων f_1 και f_2 αντίστοιχα. Τα συμπεράσματα από την εφαρμογή αυτής της απόστασης είναι τα συμπεράσματα που αναφέρθηκαν στην εισαγωγή για τη χρήση των ιστογραμμάτων στην ανίχνευση ασυνεχειών: αποτελούν μια σταθερή, γρήγορη μέθοδο ανίχνευσης ασυνεχειών και οδηγούν σε λάθος ανιχνεύσεις μόνο στην περίπτωση πραγματικά έντονης κίνησης των αντικειμένων ή της κάμερας.

Στη συνέχεια έγινε προσπάθεια συνδυασμού της πληροφορίας των ιστογραμμάτων έντασης και χρώματος στην ανίχνευση των ασυνεχειών. Προς αυτή την κατεύθυνση θεωρήθηκαν οι χώροι χρώματος YCbCr και YIQ. Ο πρώτος είναι ο χώρος χρώματος στον οποίο δίνονται οι εικόνες από τον αλγόριθμο αποκωδικοποίησης MPEG και επομένως δεν απαιτείται επιπλέον μετασχηματισμός. Ο δεύτερος προκύπτει από τον YCbCr με μετασχηματισμό μόνο του υποδειγματοληπτημένου χρωματικού χώρου CbCr στον IQ. Επιπλέον ο μετασχηματισμός αυτός είναι γραμμικός και έτσι το χρονικό κόστος της διαδικασίας μετασχηματισμού είναι ελάχιστο. Στη συνέχεια της διαδικασίας υπολογίζεται για κάθε εικόνα το δισδιάστατο ιστόγραμμα H^{Chr} θεωρώντας 40 bins για κάθε χρωματική συνιστώσα (συνολικά 1600 bins) όπως ακριβώς περιγράφεται στο [21]. Ως μια πρώτη προσέγγιση, η απόσταση δύο εικόνων ορίστηκε ως εξής:

$$\delta_{CD}^{Y,Chr}(f_1, f_2; \alpha) = (1-\alpha)\,\delta_i(\mathrm{H}_1^{\mathrm{Y}}, \mathrm{H}_2^{\mathrm{Y}}) + \alpha(4\delta_i(\mathrm{H}_1^{Chr}, \mathrm{H}_2^{Chr})) \tag{4.2}$$

όπου H_1^{Chr}, H_2^{Chr} είναι τα ιστογράμματα χρώματος των εικόνων f_1, f_2 αντίστοιχα και α $(0 \leq \alpha \leq 1)$ είναι ένας συντελεστής συμμετοχής του χρώματος στην απόσταση των εικόνων. Από την εφαρμογή του κριτηρίου συμπεραίνεται ότι:

- τα ιστογράμματα χρώματος αποτελούν σε σχέση με τα ιστογράμματα έντασης μια πιο συνοπτική περιγραφή των εικόνων λόγω

  ▷ της υποδειγματοληψίας που έχουν υποστεί οι χρωματικές συνιστώσες και

  ▷ του κβαντισμού των χρωμάτων κατά τον υπολογισμό του ιστογράμματος.

  Η διαπίστωση αυτή ενισχύεται από το γεγονός ότι η σημαντική πληροφορία και στα δύο συστήματα χρώματος (YCbCr και YIQ) συσσωρεύεται στη συνιστώσα Y της έντασης. Έτσι, η δυναμική των αποστάσεων των ιστογραμμάτων χρώματος ακολουθεί γενικά μια πιο ομαλή μεταβολή σε σχέση με αυτή των ιστογραμμάτων έντασης.

- Λόγω των όσων περιγράφηκαν μόλις, για τιμές του α κοντά στο 1, αποφεύγονται μεν οι λάθος ανιχνεύσεις που οφείλονται σε μεγάλες κινήσεις, χάνονται όμως αρκετές σωστές ανιχνεύσεις.

- για τιμές του α κοντά στο 0.5 τα αποτελέσματα είναι ανάλογα εκείνων που λαμβάνονται από τη σύγκριση μόνο των ιστογραμμάτων έντασης. Παρόλα αυτά είναι δυνατό να χαθούν ανιχνεύσεις (miss-detection) σε περιπτώσεις ασυνεχειών όπου καταγράφεται μεταβολή στη γεωμετρία της εικόνας από σκηνή σε σκηνή στο ιστόγραμμα έντασης, χωρίς ταυτόχρονη μεταβολή του χρώματος (και άρα του ιστογράμματος του). Αυτό το συμπέρασμα ενισχύεται από το γεγονός ότι η γεωμετρία του χώρου στα συστήματα αυτά αποτυπώνεται στην Y συνιστώσα.

  Γι' αυτούς τους λόγους, προτιμήθηκε η απόσταση που δίνεται από τη σχέση:

$$\delta_{CD}(f_1, f_2) = s(\delta_i(\mathrm{H}_1^{\mathrm{Y}}, \mathrm{H}_2^{\mathrm{Y}}), 4\delta_i(\mathrm{H}_1^{Chr}, \mathrm{H}_2^{Chr})) \, \delta_{CD}^{Y,Chr}(f_1, f_2; 0.5) \tag{4.3}$$

όπου

$$s(x, y) = -\frac{x}{x+y}\log_2\left(\frac{x}{x+y}\right) - \frac{y}{x+y}\log_2\left(\frac{y}{x+y}\right) \tag{4.4}$$

είναι μια συνάρτηση μέτρησης της ομοιότητας των αποστάσεων των ιστογραμμάτων έντασης και χρώματος. Ποιοτικά, η εξ. 4.3 εκφράζει αυτό που αναμένεται στην περίπτωσης πραγματικής ασυνέχειας: τόσο η απόσταση των ιστογραμμάτων έντασης όσο και αυτή των ιστογραμμάτων χρώματος να είναι μεγάλες και σχετικά όμοιες μεταξύ τους. Μ' αυτό τον τρόπο αποφεύγονται τόσο οι λάθος ανιχνεύσεις, όσο και οι μή ανιχνεύσεις, που προκαλούνται αν δοθεί η δυνατότητα στο χρήστη του αλγορίθμου να ρυθμίζει την παράμετρο α συμμετοχής του χρώματος στο κριτήριο ανομοιότητας.

4.2.3 Σειρά Προσπέλασης των Εικόνων του Video

Ο αλγόριθμος ανίχνευσης ασυνεχειών εφαρμόζεται στο χρονικό διάστημα μεταξύ δύο διαδοχικών I-frames (το οποίο αναφέρεται ως ''παράθυρο εργασίας'' στο σχ. 4.2). Η υπόθεση που γίνεται είναι ότι αυτό το διάστημα είναι αρκετά μικρό ώστε να περιλαμβάνει

ένα μόνο σημείο ασυνέχειας (το αν και κατά πόσο αυτή η υπόθεση ευσταθεί, συζητείται στην παράγραφο 4.2.5). Η σειρά με την οποία εξετάζονται οι εικόνες μεταξύ των δύο I-frames, φαίνεται στο σχ. 4.2 το οποίο αναπαριστά σχηματικά τις δυνατές επιλογές ζευγών εικόνων κατά τη διάρκεια εκτέλεσης του αλγορίθμου που σε μορφή ψευδοκώδικα έχει ως εξής:

B1. (I-επίπεδο) Εφαρμογή της ανίχνευσης ασυνέχειας θεωρώντας το επόμενο ζεύγος των I-frames (στο σχ. 4.2 οι εικόνες f_1 και f_7) με χρήση μιας από τις εξισώσεις απόστασης ιστογραμμάτων που περιγράφηκαν παραπάνω. Εάν η απόσταση μεταξύ των δύο εικόνων είναι μικρότερη από το κατώφλι εκτελείται το παρόν βήμα στο επόμενο διάστημα μεταξύ δύο I-frames (στο σχ. 4.2 στο διάστημα μεταξύ της εικόνας f_7 και της επόμενης I-εικόνας), αλλιώς εκτελείται το βήμα B2.

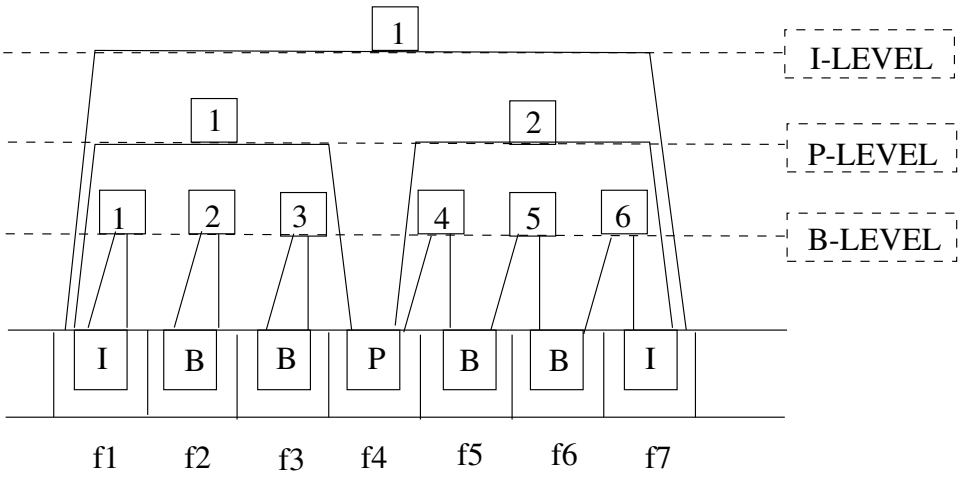
B2. (P-επίπεδο) Επιβεβαίωση της ανίχνευσης ασυνέχειας στο χρονικό διάστημα μεταξύ των δύο επομένων διαδοχικών I-, P-frames (οι εικόνες f_1,f_4), που περιλαμβάνονται στο υπό εξέταση διάστημα των I-frames:

   B2.1 Εφαρμογή της ανίχνευσης ασυνέχειας μεταξύ των δύο I-, P-frames όπως και στο βήμα B1. Εάν η ένδειξη ασυνέχειας του προηγουμένου επιπέδου δεν επιβεβαιώνεται (με άλλα λόγια εάν η απόσταση μεταξύ των υπό εξέταση εικόνων δεν ξεπερνάει το κατώφλι), εκτελείται το βήμα B2 στο επόμενο χρονικά ζεύγος διαδοχικών I-, P-frames (εικόνες f_4,f_7), αλλιώς εκτελείται το βήμα B2.2.

   B2.2 (B-επίπεδο) Προσπάθεια προσδιορισμού του ακριβούς χρονικού σημείου της ασυνέχειας στο τρέχον υποδιάστημα, μετρώντας την απόσταση μεταξύ των διαδοχικών ζευγών εικόνων. Εάν η ασυνέχεια δεν επιβεβαιωθεί εκτελείται το βήμα B2 στο υποδιάστημα που ορίζεται από το επόμενο ζεύγος I-, P-frames, αλλιώς εκτελείται το βήμα B1 στο επόμενο διάστημα του I-επιπέδου.

Μ' αυτό τον τρόπο, όπως φαίνεται και στο σχ. 4.2, το διάστημα μεταξύ των I-frames υποδιαιρείται σε μικρότερα υποδιαστήματα βάσει της MPEG κωδικοποίησης. Ζεύγη εικόνων που βρίσκονται σε χαμηλότερα επίπεδα προσπέλασης, εξετάζονται μόνο εφόσον στο αμέσως ανώτερό τους επίπεδο έχει ανιχνευθεί ασυνέχεια στη ροή του video. Για παράδειγμα, ανίχνευση ασυνέχειας στα χρονικά διαστήματα 1, 2, 3 στο B-επίπεδο εφαρμόζεται μόνο εάν η εξέταση της ασυνέχειας στα υποδιαστήματα 1 των I-, P-επιπέδων έχει δώσει θετικό αποτέλεσμα (''ένδειξη'' ασυνέχειας). Ο αλγόριθμος τερματίζει (πάντα για το διάστημα μεταξύ του τρέχοντος ζεύγους I-frames) στο πρώτο σημείο ασυνέχειας που θα βρεθεί στο B-επίπεδο (''επιβεβαίωση'' ασυνέχειας). Το χρονικό αυτό σημείο-εικόνα του video οριοθετεί το τέλος μιας σκηνής και την αρχή της επόμενής της.

Το βασικό πλεονέκτημα αυτού του top-down σχήματος προσπέλασης των εικόνων είναι ότι μειώνει αισθητά την ''περιττή'' αποκωδικοποίηση και επομένως το χρόνο επεξεργασίας του video, γεγονός που το κάνει πολύ αποδοτικό στις εφαρμογές πραγματικού



Σχήμα 4.2: Σειρά Προσπέλασης Εικόνων του Video

χρόνου όπου είναι γνωστό πως ο χρόνος αποτελεί βασικό κριτήριο αποδοτικότητας των αλγορίθμων. Στο καθαρά προγραμματιστικό μέρος του ο αλγόριθμος έχει σχεδιαστεί έτσι ώστε κάθε εικόνα του video να αποκωδικοποιείται το πολύ μία φορά κατά τη διάρκεια της επεξεργασίας του.

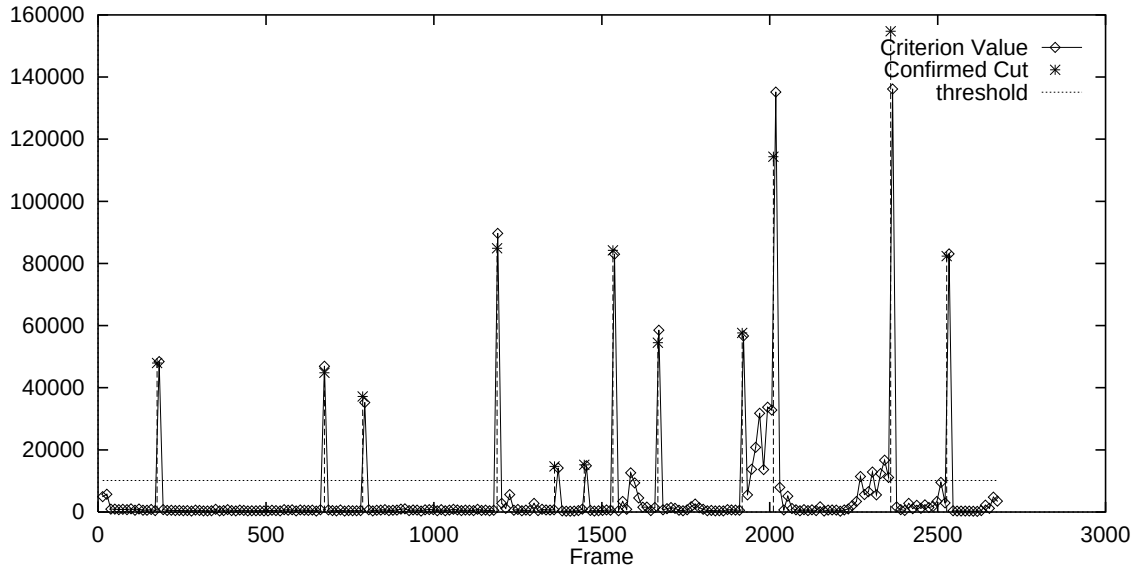
4.2.4 Πειραματικά Αποτελέσματα

Ο αλγόριθμος έχει ελεγχθεί σε video MPEG1 που ευγενικά παραχωρήθηκαν από την Ελληνική Ραδιοφωνία και Τηλεόραση (ΕΡΤ), στα πλαίσια του προγράμματος επεξεργασίας και δεικτοδότησης του περιεχομένου video DiVAN [9]. Σ' ότι ακολουθεί, παρουσιάζονται τα αποτελέσματα για δύο από αυτά τα video, τα οποία για ευκολία έχουν ονομασθεί VIDEO1 και VIDEO2 αντίστοιχα.

Το θέμα του πρώτου video, είναι οι εξαγγελίες του προέδρου των ΗΠΑ, σχετικά με την απαγόρευση εφαρμογής της μεθόδου της κλωνοποίησης στον άνθρωπο. Η απόσταση ιστογραμμάτων που χρησιμοποιήθηκε σε όλα τα παραδείγματα είναι η εξίσωση (2.5) αφού το κριτήριο αυτό παρουσιάζει αντοχή σε μεγάλες κινήσεις αντικειμένων καθώς και στις μεταβολές του background. Κοινό σε όλα τα παραδείγματα είναι επίσης και το κατώφλι ασυνέχειας, το οποίο τέθηκε στο 0.1 του αριθμού των σημείων της εικόνας.

Στο σχ. 4.3, φαίνεται η απόσταση δ_{CD}^{Y} των ιστογραμμάτων της έντασης των διαδοχικών I-εικόνων, που περιέχονται στις 2683 εικόνες του video. Με αστερίσκο σημειώνονται οι τιμές του κριτηρίου για τις I-,P- ή B-εικόνες στις οποίες ανιχνεύθηκε ασυνέχεια. Η πρώτη και η τελευταία εικόνα καθεμιάς από τις 13 σκηνές του video που ανιχνεύθηκαν όλες σωστά, φαίνονται στο σχήμα 4.5, (από αριστερά προς τα δεξιά και από πάνω προς τα κάτω). Από τις αποστάσεις των ιστογραμμάτων του σχήματος 4.3, διαπιστώνεται ότι οι σκηνές δεν περιλαμβάνουν μεγάλες μεταβολές στο περιεχόμενό τους μεταξύ διαδοχικών I-εικόνων, εκτός από αυτό των σκηνών 10 και 11 (εικόνες 1918 έως 2359), στις οποίες υπάρχουν αντικείμενα που κινούνται με μεγάλη ταχύτητα

(δεύτερο ζεύγος εικόνων της πέμπτης γραμμής και πρώτο ζεύγος της έκτης αντίστοιχα, του σχήματος 4.5). Αν και σε αρκετές περιπτώσεις στο διάστημα αυτό η απόσταση μεταξύ δύο διαδοχικών I-εικόνων ξεπερνάει το κατώφλι, η ανίχνευση δεν επιβεβαιώνεται στη συνέχεια στο διάστημα μεταξύ των δύο διαδοχικών I-εικόνων.

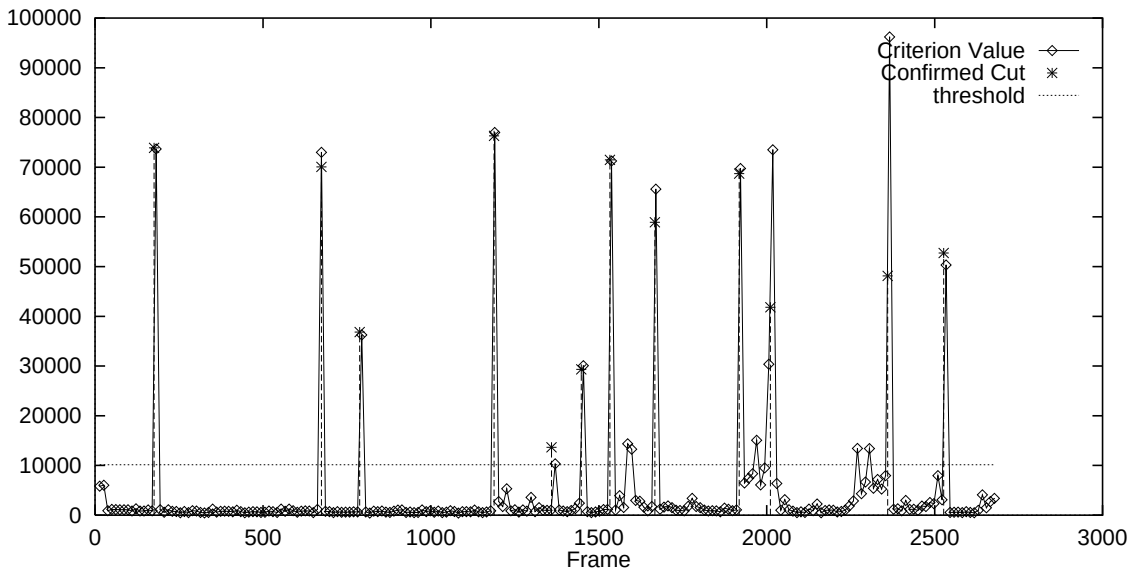


Σχήμα 4.3: Απόσταση ιστογραμμάτων έντασης των διαδοχικών I-εικόνων του VIDEO1

Στο σχ. 4.4 φαίνεται η τιμή της απόστασης δ_{CD}, όπου έχει συνδυαστεί η απόσταση του ιστογράμματος έντασης με αυτή του ιστογράμματος χρώματος. Καμία από τις ανιχνεύσεις που επιβεβαιώθηκαν στην περίπτωση των ιστογραμμάτων της έντασης δεν έχει χαθεί. Εάν χρησιμοποιηθεί η εξίσωση $\delta_{CD}^{Y,Chr}$, με $\alpha \geq 0.4$, χάνεται η ανίχνευση ασυνέχειας μεταξύ των σκηνών 6 και 7. Το ακριβές χρονικό σημείο της ασυνέχειας είναι η εικόνα 1359. Ο λόγος είναι ότι οι εικόνες 1358 και 1359 είναι όμοιες χρωματικά, ενώ και η απόσταση των ιστογραμμάτων έντασης μόλις που ξεπερνάει το κατώφλι t_{CD} (βλ. σχ. 4.3).

Το VIDEO2 περιλαμβάνει σκηνές με μεγάλη κίνηση των αντικειμένων και μεταβολές στο background. Αυτό διαπιστώνεται εύκολα, τόσο από τις μεγάλες αποστάσεις των ιστογραμμάτων έντασης (σχ. 4.6) και έντασης-χρώματος (σχ. 4.7) που υπολογίζονται όπως και για το VIDEO1, όσο και από το γεγονός ότι σε πολλές από τις σκηνές η πρώτη και τελευταία εικόνα διαφέρουν σημαντικά μεταξύ τους (σχ. 4.8) Πάντως και στις δύο περιπτώσεις, δε χάνονται οι 8 αλλαγές σκηνής που υπάρχουν στο video. Συγκρίνοντας τη γραφική παράσταση των αποστάσεων των ιστογραμμάτων της έντασης με αυτή των ιστογραμμάτων έντασης-χρώματος, προκύπτει ότι ο συνυπολογισμός του χρώματος, ομαλοποιεί τις αποστάσεις των εικόνων. Αυτή η διαπίστωση μπορεί να χρησιμοποιηθεί όπως εξηγείται στο επόμενο κεφάλαιο για την εξαγωγή των αντιπροσωπευτικών εικόνων.

Ο χρόνος εκτέλεσης του συστήματος (μαζί μ' αυτόν της εξαγωγής αντιπροσωπευ-



Σχήμα 4.4: Απόσταση ιστογραμμάτων έντασης και χρώματος των διαδοχικών I-εικόνων του VIDEO1

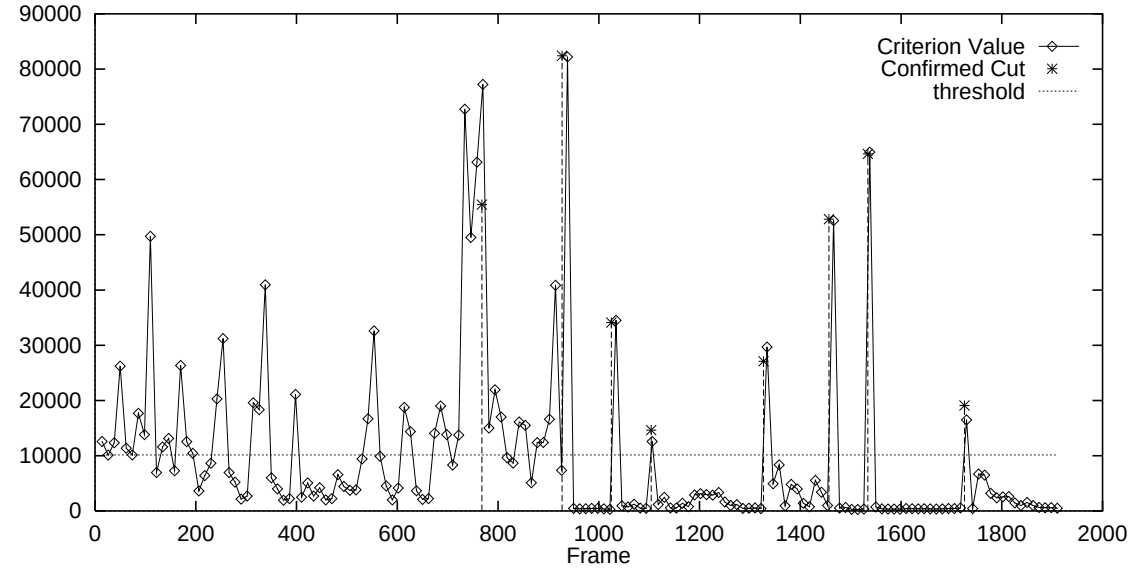
τικών εικόνων), εξαρτάται από το περιεχόμενο του video. Εάν το video περιλαμβάνει έντονη δραστηριότητα αντικειμένων και μεταβολή στο background των σκηνών, η διαδικασία ανίχνευσης ασυνεχειών, εφαρμόζεται συνήθως μέχρι το επίπεδο των B-εικόνων, έστω και εάν τελικά η ασυνέχεια δεν επιβεβαιωθεί. Πάντως, σε κάθε περίπτωση, ο μέσος χρόνος απόκρισης είναι περίπου 1 sec/εικόνα.

4.2.5 Συμπεράσματα-Επεκτάσεις

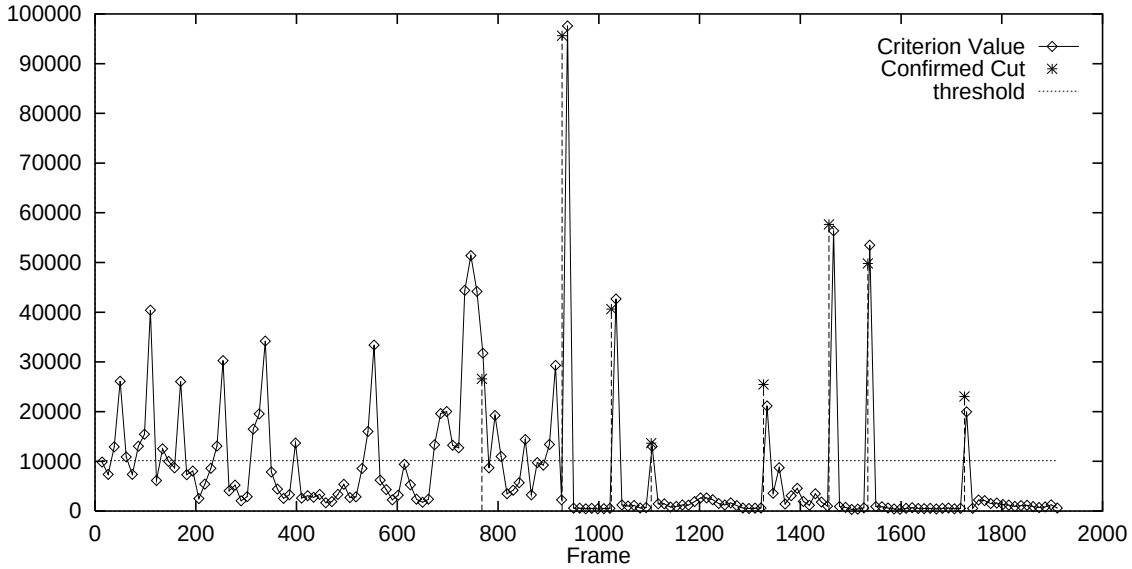
Ο αλγόριθμος ανίχνευσης αλλαγής σκηνών, σχεδιάστηκε έτσι ώστε να ακολουθεί τις βασικές αρχές που αναπτύχθηκαν στο προηγούμενο κεφάλαιο. Έτσι, αναφέροντας τις κύριες επιλογές που έγιναν, η χρήση ιστογραμμάτων έγινε λόγω της ανθεκτικότητας που παρουσιάζουν σε μεγάλες μεταβολές του περιεχομένου που προκαλείται από μεγάλες κινήσεις των αντικειμένων. Ένας σημαντικός επιπρόσθετος παράγοντας είναι ο σύντομος χρόνος υπολογισμού τους αλλά και το γεγονός ότι από τη φύση τους οι απότομες μεταβάσεις κάθε άλλο παρά αντιστοιχούν σε κάποιο μαθηματικό μοντέλο.

Όσον αφορά την υπόθεση που γίνεται για την ύπαρξη μιας μόνο ασυνέχειας στο χρονικό διάστημα μεταξύ δύο I-εικόνων, αυτή δικαιολογείται από το γεγονός ότι αυτό το διάστημα είναι, στην πλειονότητα των περιπτώσεων, μικρό. Για παράδειγμα στα video MPEG στα οποία οποία εφαρμόστηκε το σύστημα, η αναλογία του αριθμού των I-, P-, B-εικόνων είναι όπως αναφέρθηκε στο προηγούμενο κεφάλαιο 1 : 3 : 8. Λαμβάνοντας υπόψη ότι ο ρυθμός μετάδοσης του video για την Ευρώπη είναι 25 εικόνες/sec, το χρονικό διάστημα των 12 εικόνων αντιστοιχεί σε κάτι λιγότερο από 0.5 sec. και αυτή είναι η τάση που τείνει να επικρατήσει και στις εφαρμογές μετάδοσης video. Επομένως σ' ένα τόσο μικρό διάστημα είναι απίθανο να περιέχονται περισσότερες της μιας ασυνέχειες. Στην

Σχήμα 4.5: Πρώτη και τελευταία εικόνα των δεκατριών σκηνών του VIDEO1



Σχήμα 4.6: Απόσταση ιστογραμμάτων έντασης των διαδοχικών I-εικόνων του VIDEO2



Σχήμα 4.7: Απόσταση ιστογραμμάτων έντασης και χρώματος των διαδοχικών I-εικόνων του VIDEO2

Σχήμα 4.8: Πρώτη και τελευταία εικόνα των εννέα σκηνών του VIDEO2

περίπτωση που συμβαίνει κάτι τέτοιο, έχουμε μια ‘‘σκηνή’’ η οποία δεν ξεπερνάει το μισό sec. και την οποία ο αλγόριθμος αποτυγχάνει να ανιχνεύσει. Το θέμα που τίθεται είναι το πόσο σημαντικό μπορεί να είναι το περιεχόμενο μιας σκηνής, την οποία ο χρήστης δεν προλαβαίνει καν να διακρίνει.

Επιπλέον, από τη χρήση του αλγορίθμου, γίνεται σαφές πως το ‘‘ενοχλητικό’’, όσον αφορά τη δεικτοδότηση, πρόβλημα --ακόμα και με το top-down σχήμα που χρησιμοποιήθηκε-- είναι η ανίχνευση αλλαγής εκεί όπου δεν υπάρχει μετάβαση (false positives), συνήθως σε σημεία μεγάλης κίνησης και όχι η μή ανίχνευση αλλαγής σκηνών (miss-detection). Το σχόλιο μάλιστα των ειδικών του INA [1], που προέκυψε από τη χρήση του αλγορίθμου στην πράξη, είναι ότι θα μπορούσαν να αγνοούνται ως μή σημαντικές, ανιχνεύσεις που οριοθετούν ‘‘σκηνές’’ των οποίων η διάρκεια δεν ξεπερνάει κάποιο αριθμό δευτερολέπτων.

Τα όσα λέχθηκαν για την επιλογή του ‘‘παραθύρου εργασίας’’, δικαιολογούν και την εφαρμογή του top-down σχήματος προσπέλασης των εικόνων του οποίου τα πλεονεκτήματα κυρίως από την άποψη του υπολογιστικού χρόνου, παρατέθηκαν αμέσως μετά την περιγραφή του. Το πλεονέκτημα αυτό ισχύει και στην περίπτωση μή συμπιεσμένων video αρκεί να εφαρμοστεί ένα ανάλογο σχήμα προσπέλασης των εικόνων βάσει (εύλογων) χρονικών αποστάσεων μεταξύ τους. Ο χρόνος που εξοικονομείται σ’ αυτή την περίπτωση προέρχεται αποκλειστικά από την αποφυγή εφαρμογής του κριτηρίου ανίχνευσης που, γι’ αυτόν ακριβώς το λόγο, θα μπορούσε να περιλαμβάνει συγκρίσεις μεταξύ χαρακτηριστικών των εικόνων των οποίων η εξαγωγή απαιτεί περισσότερο χρόνο υπολογισμού σε σύγκριση με αυτής των ιστογραμμάτων που χρησιμοποιήθηκαν εδώ.

Τέλος, πρέπει να επισημανθεί ότι αν και ο αλγόριθμος δεν έχει σχεδιασθεί για τον προσδιορισμό βαθμιαίων μεταβάσεων, η ύπαρξή τους καταγράφεται στη βάση, αφού όπου εμφανίζονται τέτοιου τύπου μεταβάσεις, εξάγονται αντιπροσωπευτικές εικόνες που περιγράφουν πλήρως τη συντελούμενη μεταβολή.

[1] Institut National de l’Audiovisuel: το μεγαλύτερο ίδρυμα αρχειοθέτησης υλικού video της Γαλλίας και βασικός συνεργάτης στο πρόγραμμα DiVAN

Κεφάλαιο 5

Εξαγωγή Αντιπροσωπευτικών Εικόνων

5.1 Εισαγωγή

Η εξαγωγή των αντιπροσωπευτικών εικόνων είναι ότι ακολουθεί την ανίχνευση των ορίων των σκηνών στις εφαρμογές επεξεργασίας του video, που έχουν ως σκοπό την κατασκευή μιας δομής δεικτοδότησης που να επιτρέπει την αποδοτική ανάκτηση του περιεχομένου του ενώ παράλληλα να ικανοποιεί τις απαιτήσεις για όσο το δυνατόν μικρότερο χώρο αποθήκευσης αυτού του περιεχομένου. Με τον όρο *αντιπροσωπευτική εικόνα* (μτφρ. του αγγλικού όρου key-frame), αναφερόμαστε στις εικόνες εκείνες που αναπαριστούν καλύτερα το περιεχόμενο της κάθε σκηνής σύμφωνα πάντα με κάποια προεπιλεγμένα χαρακτηριστικά των εικόνων. Οι εικόνες αυτές συντελούν στην *αφαιρετική* περιγραφή του video, από τη στιγμή που ένα τμήμα του αντιστοιχίζεται κατ' αυτόν τον τρόπο σε ένα μικρό αριθμό εικόνων. Έτσι, η δεικτοδότηση μπορεί να περιλαμβάνει μόνο αυτές τις εικόνες και επομένως οι ερωτήσεις στη βάση να απευθύνονται σ' αυτές. Επιπλέον, εάν ο χώρος αποθήκευσης είναι περιορισμένος, αρκεί να διατηρούνται on-line μόνο αυτές οι εικόνες.

Μια άλλη προσέγγιση πάνω στο ίδιο θέμα περιγραφής του περιεχομένου του video είναι η χρήση των λεγομένων εικόνων-μωσαϊκού (mosaic images) οι οποίες αποθηκεύουν το περιεχόμενο της σκηνής σε μία μόνο εικόνα. Η τεχνική αυτή έχει εφαρμοσθεί μέχρι σήμερα με σημαντική επιτυχία, κυρίως σε σκηνές στις οποίες η κάμερα κινείται ομαλά και όπου δεν εμπλέκονται πολύπλοκες κινήσεις αντικειμένων σ'αυτές. Πάντως, η έρευνα σ' αυτή την κατεύθυνση συνεχίζεται και τα αποτελέσματα σε πιο ''δύσκολες'' --από την άποψη κίνησης κάμερας και αντικειμένων-- σκηνές είναι ομολογουμένως ενθαρρυντικά. Σημαντική είναι η εργασία που έχει γίνει πάνω στο αντικείμενο από τη Μ. Τράκα [25]. Εκεί, αναλύονται διεξοδικά και οι δυσκολίες που ανακύπτουν.

Επιστρέφοντας στη συζήτηση για τις εικόνες-κλειδιά, πρέπει να επισημανθεί ότι το ιδανικό θα ήταν η εξαγωγή τους να γίνεται αυτόματα, και να βασίζεται στο περιεχόμενό

τους, έτσι ώστε να διατηρείται στη βάση δεδομένων το σημαντικό περιεχόμενο και να μειώνεται ο πλεονασμός: θεωρητικά τουλάχιστο, για την εξαγωγή των αντιπροσωπευτικών εικόνων θα πρέπει να χρησιμοποιούνται σημασιολογικά χαρακτηριστικά (semantic primitives), όπως αντικείμενα, ενέργειες και γεγονότα της υπό δεικτοδότηση σκηνής. Παρόλα αυτά η πρόοδος που έχει συντελεστεί στην προσπάθεια αυτόματης εξαγωγής εικόνων-κλειδιών είναι μικρή και συνήθως αυτό που γίνεται είναι να επιλέγεται από πριν ένας μέγιστος επιτρεπτός αριθμός εικόνων-κλειδιών για κάθε σκηνή (για παράδειγμα η πρώτη, μεσαία και τελευταία εικόνα της σκηνής ή πιθανή επιλογή τους από το χρήστη). Επιπλέον, επειδή μια τόσο γενική σημασιολογική ανάλυση της οπτικής πληροφορίας δεν είναι ακόμη εφικτή, χρησιμοποιούνται στην πράξη χαμηλού επιπέδου χαρακτηριστικά των εικόνων που εξάγονται εύκολα ή είναι ήδη διαθέσιμα όπως για παράδειγμα η ένταση και το χρώμα των σημείων των εικόνων.

Βάσει αυτών των περιορισμών, ο αλγόριθμος που ακολουθεί εξάγει τις αντιπροσωπευτικές εικόνες βασιζόμενος στα ιστογράμματα χρώματος και έντασης των εικόνων, που έχουν ήδη υπολογιστεί για τις ανάγκες του αλγορίθμου ανίχνευσης ασυνεχειών, ο οποίος βρίσκεται συγχρόνως σε εξέλιξη.

5.2 Ο Αλγόριθμος Εξαγωγής Αντιπροσωπευτικών Εικόνων

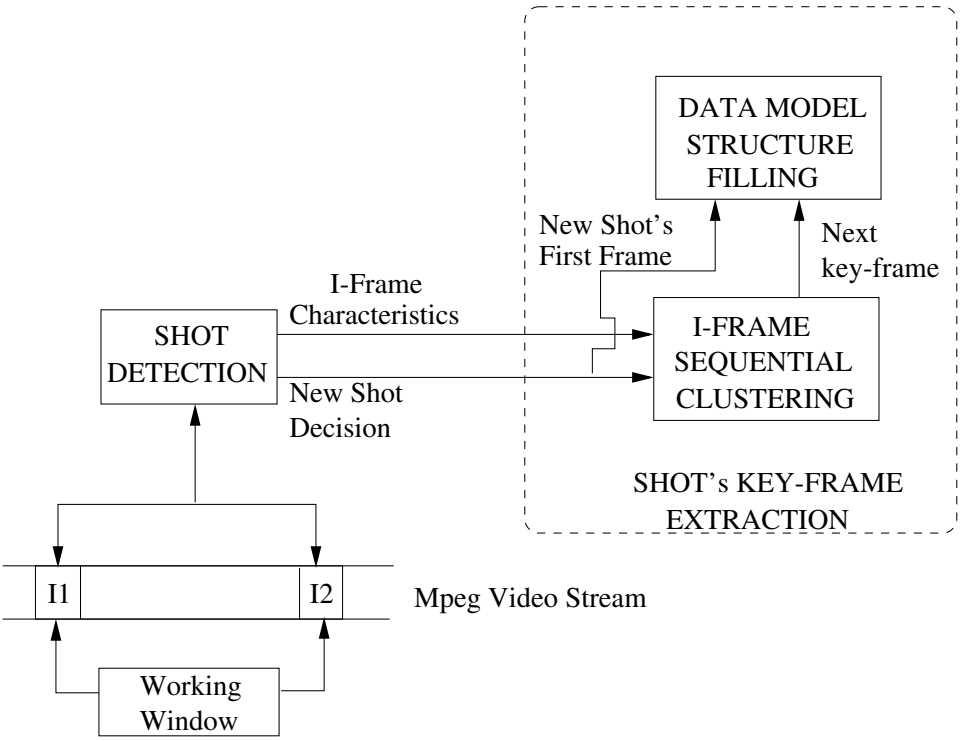
5.2.1 Σύνδεση με τον Αλγόριθμο Ανίχνευσης Ασυνεχειών

Όπως ήδη έχει ειπωθεί, η εξαγωγή των αντιπροσωπευτικών εικόνων κάθε σκηνής συμπληρώνει την πληροφορία δεικτοδότησής της στο μοντέλο δεδομένων του video. Το σχήμα 5.1, δίνει τη συνολική εικόνα της περιγραφόμενης προσέγγισης εστιάζοντας στις λεπτομέρειες που αφορούν το μέρος της εξαγωγής αντιπροσωπευτικών εικόνων και στις παραμέτρους αλληλεπίδρασης των επιμέρους τμημάτων του αλγορίθμου. Θεωρώντας την κατάσταση του ολικού σχήματος η οποία έχει προκύψει μετά την ανίχνευση ασυνέχειας στο χρονικό ‘‘παράθυρο’’ μεταξύ των εικόνων I_1 και I_2, φαίνεται σχηματικά ότι η έξοδος της που αποτελείται από

- την απόφαση ανίχνευσης ή μη ασυνέχειας
- τα χαρακτηριστικά της εικόνας I_2 τα οποία έχουν ήδη εξαχθεί --στην περίπτωσή μας το ιστόγραμμα έντασης/χρώματος
- και τέλος, στην περίπτωση διαπίστωσης ασυνέχειας το ακριβές χρονικό σημείο --αριθμό εικόνας-- αυτής.

δίνεται στην διεργασία εξαγωγής αντιπροσωπευτικών εικόνων που στόχο της έχει την εξαγωγή της επόμενης απ’ αυτές, κατά τη διάρκεια εκτέλεσης της ανίχνευσης του τέλους της τρέχουσας σκηνής. Για το σκοπό αυτό, ακολουθήθηκε η προσέγγιση σειριακής ομαδοποίησης (sequential clustering) των I-frames της υπό επεξεργασία σκηνής η οποία

περιγράφεται στην επόμενη παράγραφο. Το χαρακτηριστικό βάσει του οποίου γίνεται η ομαδοποίηση είναι η πληροφορία του ιστογράμματος των εικόνων, αποφεύγοντας έτσι τον υπολογισμό επιπλέον χαρακτηριστικών. Οι λόγοι που οδήγησαν σ' αυτές τις επιλογές, σχετίζονται άμεσα με τις ανάγκες της εφαρμογής του σε πραγματικό χρόνο για την οποία έχει σχεδιαστεί ο αλγόριθμος και επεξηγούνται στην παρ. 3.2.



Σχήμα 5.1: Λεπτομέρειες Αλγορίθμου Εξαγωγής Αντιπροσωπευτικών Εικόνων

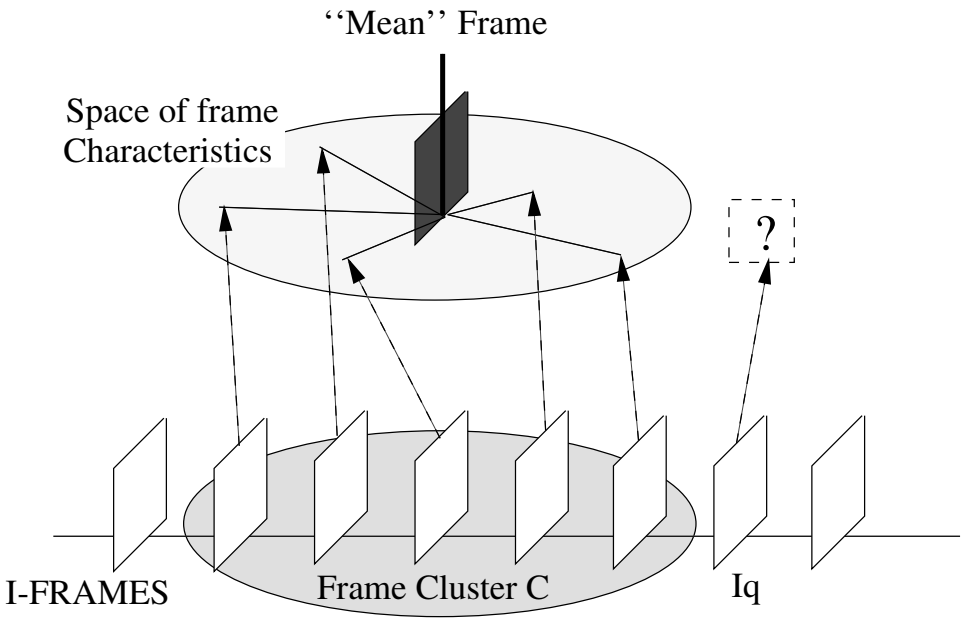
Η εξαγωγή των αντιπροσωπευτικών εικόνων γίνεται αυτόματα και βασίζεται στα ιστογράμματα των I-εικόνων όπως περιγράφεται πιο κάτω.

5.2.2 Μέθοδος Σειριακής Ομαδοποίησης

Η μέθοδος σειριακής ομαδοποίησης (sequential clustering) περιλαμβάνει μία μόνο ομάδα C των υποψηφίων αντιπροσωπευτικών εικόνων που επεκτείνεται σειριακά με την εισαγωγή σ' αυτή της I-εικόνας που ακολουθεί το τελευταίο της στοιχείο-εικόνα στη διάταξη εμφάνισης του video. Έτσι, η ομάδα C είναι το σύνολο K διαδοχικών I-εικόνων που ορίζεται ως

$$C=\{C_{mf} : C_{f_1}, C_{f_2}, \ldots, C_{f_K}\} \tag{5.1}$$

όπου ως C_{mf} συμβολίζεται μία ''ιδεατή'' εικόνα που αντιστοιχεί στο ''μέσο'' διάνυσμα των αντίστοιχων χαρακτηριστικών των εικόνων C_{f_k} $(1 \leq k \leq K)$ και σημειώνεται ως ''Μέση Εικόνα'' στο σχ. 5.2.



Σχήμα 5.2: Μέθοδος Σειριακής Ομαδοποίησης

Επιπλέον, θεωρούμε την απόσταση $\delta_{int}(C_{f_k}, C_{mf})$ μιας από τις C_{f_k} εικόνες μέλη του συνόλου από την εικόνα C_{mf}. Προφανώς, η απόσταση αυτή αποτελεί συνάρτηση των μετρούμενων χαρακτηριστικών. Βάσει αυτής της απόστασης, η αντιπροσωπευτική εικόνα C_{KF} επιλέγεται να είναι η εικόνα της ομάδας που ικανοποιεί τη σχέση

$$C_{KF}=\arg\min_{C_{f_k}}\{\delta_{int}(C_{f_k}, C_{mf})\} \tag{5.2}$$

Το σχήμα 5.2 περιγράφει σχηματικά ένα στιγμιότυπο της διεργασίας της ομαδοποίησης. Τα I-frames τα οποία στην τρέχουσα φάση της ομαδοποίησης περιλαμβάνονται στο σύνολο C, εμφανίζονται στο σχήμα μέσα σε κύκλο. Επιπλέον φαίνεται η αντιστοίχιση καθεμιάς απ' αυτές τις εικόνες στο αντίστοιχο σημείο της του χώρου χαρακτηριστικών σύμφωνα με το διάνυσμα χαρακτηριστικών τους. Στην περίπτωση που μόνο τα ιστογράμματα χρησιμοποιούνται για την ομαδοποίηση ο δηλούμενος ως ''Χώρος Χαρακτηριστικών'' του σχ. 5.2 είναι ο χώρος των Μ-διάστατων ιστογραμμάτων. Επιπρόσθετα, η απόσταση της κάθε εικόνας από την υπολογισμένη ''Μέση Εικόνα'' είναι η γραμμή που ξεκινά από το αντίστοιχο σημείο της του χώρου χαρακτηριστικών και τελειώνει στο σημείο που βρίσκεται η εικόνα των ''μέσων'' χαρακτηριστικών της ομάδας, στο ''Χώρο Χαρακτηριστικών''.

Ας υποθέσουμε τώρα ότι η διαδικασία ανίχνευσης ασυνεχειών στο διάστημα μεταξύ της εικόνας C_{f_K} και της επόμενής της σε διάταξη εμφάνισης μή ομαδοποιημένης I-εικόνας (η οποία σημειώνεται ως I_q στο σχ. 5.2) έχει ολοκληρωθεί. Θεωρώντας ότι δεν έχει ανιχνευθεί ασυνέχεια σ' αυτό το διάστημα, η εικόνα I_q ή συμπεριλαμβάνεται στο τρέχον σύνολο C ή θα πρέπει να δημιουργηθεί ένα νέο σύνολο, του οποίου το πρώτο στοιχείο θα είναι η εικόνα I_q. Η απόφαση βασίζεται στην απόσταση $\delta_{ext}(I_q, C_{mf})$ της I_q

από τη ''Μέση Εικόνα'' του συνόλου C που επίσης εξαρτάται από τα χαρακτηριστικά των εικόνων:

Περίπτωση 1 Εάν η τιμή της απόστασης ξεπερνάει ένα προεπιλεγμένο κατώφλι t_{NewC}, ένα νέο σύνολο δημιουργείται. Η δημιουργία του περιλαμβάνει δύο βήματα:

1. Εξαγωγή της αντιπροσωπευτικής εικόνας C_{FK} (εξ. 5.2) βάσει του τρέχοντος περιεχομένου του συνόλου C.
2. Αρχικοποίηση του συνόλου C σε ένα μονοσύνολο που περιλαμβάνει την εικόνα I_q.

Περίπτωση 2 Αλλιώς,

1. Η I_q συμπεριλαμβάνεται στο σύνολο.
2. Η ''Μέση Εικόνα'' επανατοποθετείται στο χώρο χαρακτηριστικών επανα-υπολογίζοντας το διάνυσμα χαρακτηριστικών της.

Από την άλλη μεριά, εάν η I_q είναι η πρώτη εικόνα μιας νέας σκηνής, δεν υπάρχει λόγος να υπολογιστεί η απόστασή της από την ομάδα C, από τη στιγμή που αυτή η ομάδα αποτελεί ένα υποσύνολο των εικόνων της προηγούμενης σκηνής. Σ' αυτή την περίπτωση, απλά εφαρμόζεται το βήμα 1, της Περίπτωσης 1, από το οποίο προκύπτει η τελευταία αντιπροσωπευτική εικόνα της προηγούμενης σκηνής και το βήμα 2 της Περίπτωσης 1 το οποίο αρχικοποιεί τη διαδικασία ομαδοποίησης για την εξαγωγή της πρώτης εικόνας-κλειδί της νέας σκηνής.

5.2.3 Σειριακή Ομαδοποίηση Εικόνων Βάσει Ιστογραμμάτων

Εφόσον το χαρακτηριστικό ομαδοποίησης είναι το ιστόγραμμα της έντασης $H^{\mathrm{Y}}_{f_k}$ των εικόνων C_{f_k}, το ιστόγραμμα H^{Y}_{mf} της ''Μέσης Εικόνας'' υπολογίζεται ως εξής:

$$H^{\mathrm{Y}}_{mf}(m) = \frac{1}{K}\sum_{k=1}^{K} H^{\mathrm{Y}}_{f_k}(m) \tag{5.3}$$

όπου ($1 \leq m \leq M_Y$) και M_Y είναι ο αριθμός τιμών της έντασης. Στην περίπτωση αυτή, οι αποστάσεις δ_{int} και δ_{ext}, υπολογίζονται χρησιμοποιώντας μία από τις εξισώσεις (2.4)-(2.7), έστω δ_i και το κατώφλι t_{NewC} τίθεται σε ένα ποσοστό του t_{CD} της ανίχνευσης ασυνεχειών. Αυτό που παρατηρείται σ' αυτή την περίπτωση, είναι ότι εξάγεται μεγάλος αριθμός εικόνων-κλειδιών, όταν η κίνηση των αντικειμένων της εικόνας είναι πραγματικά σημαντική ή υπάρχουν μεγάλες αλλαγές στη φωτεινότητα από εικόνα σε εικόνα της σκηνής. Εφόσον όπως ειπώθηκε ήδη, ο στόχος αυτής της διαδικασίας είναι η όσο το δυνατό καλύτερη αντιπροσώπευση της σκηνής, ώστε οι εικόνες αυτές να αποτελέσουν μια συνοπτική περιγραφή του περιεχομένου του video, η εκτίμηση για το αν το αποτέλεσμα είναι ικανοποιητικό, θα πρέπει να βασιστεί στο κατά πόσο κανείς θεωρεί

ότι αλλάζει το περιεχόμενο της σκηνής, λόγω της κίνησης ή της αλλαγής φωτεινότητας. Ειδικότερα στις περιπτώσεις όπου η κίνηση των αντικειμένων είναι μεγάλη, η εξαγωγή μικρού αριθμού εικόνων για εφαρμογές όπως για παράδειγμα το browsing, δεν είναι ικανή πάντα να περιγράψει ικανοποιητικά τη μεταβολή του περιεχομένου της σκηνής στη διάρκεια προβολής της. Από την άλλη πλευρά, η εξαγωγή μεγάλου αριθμού εικόνων μπορεί να οδηγήσει σε προβλήματα που έχουν να κάνουν με την αποθήκευση τόσο των ίδιων όσο και της πληροφορίας δεικτοδότησής τους. Καταλήγουμε επομένως στο συμπέρασμα ότι η εκτίμηση του αποτελέσματος βασίζεται σε υποκειμενικά κριτήρια και σχετίζεται με την εφαρμογή για την οποία προορίζεται ο αλγόριθμος, αλλά και τις δυνατότητες απόδοσης της βάσης.

Επιπρόσθετα, εισάγοντας στα χαρακτηριστικά της κλάσης τα ιστογράμματα χρώματος, παρατηρήθηκε ότι είναι δυνατό κάτω από ορισμένες προϋποθέσεις να μειωθεί ο αριθμός των εξαγωμένων αντιπροσωπευτικών εικόνων. Στην περίπτωση αυτή το μέσο ιστόγραμμα χρώματος $H_{mf}^{Chr}(m)$, υπολογίζεται ανάλογα με εκείνο της έντασης (εξ. 5.3) όπου $(1 \leq m \leq M_{Chr})$ και M_{Chr} είναι ο αριθμός των bins του ιστογράμματος.

Συμβολίζοντας ως δ_i την απόσταση ιστογραμμάτων που έχει επιλεγεί τόσο για την ανίχνευση αλλαγής σκηνών, όσο και για την εξαγωγή των εικόνων-κλειδιών, ο υπολογισμός της απόστασης της εικόνας της κλάσης C_{f_k} από τη ''Μέση Εικόνα'' C_{mf}, δίνεται από τη σχέση:

$$\delta_{int}(C_{f_k}, C_{mf}) = \frac{1}{s_1(\delta_i(H_{f_k}^{Y}, H_{mf}^{Y}), 4\delta_i(H_{f_k}^{Chr}, H_{mf}^{Chr}))} \delta_{CD}^{Y,Chr}(C_{f_k}, C_{mf}; 0.5) \tag{5.4}$$

όπου $\delta_{CD}^{Y,Chr}$ είναι η απόσταση που δίνεται από τη σχέση (4.2) και

$$s_1(x, y) = 1 - \frac{|x - y|}{x + y + 1} \tag{5.5}$$

είναι, όπως και στην περίπτωση ανίχνευσης ασυνεχειών, μια συνάρτηση μέτρησης της ομοιότητας δύο θετικών τιμών, η οποία προτείνεται στο [16] για τη μέτρηση ομοιότητας των εισόδων των συσχετογραμμάτων. Στην περίπτωση που εξετάζεται, οι τιμές αυτές είναι οι αποστάσεις των ιστογραμμάτων χρώματος και έντασης των δύο εικόνων. Ποιοτικά, ο ορισμός της απόστασης δ_{int} κατ' αυτόν το τρόπο, οδηγεί στην επιλογή (μέσω της εξ. 5.2) ως αντιπροσωπευτικής, της εικόνας της κλάσης που έχει την ιδιότητα ότι η απόστασή της από τη ''Μέση Εικόνα'', τόσο ως προς το χρώμα όσο και ως προς την ένταση είναι μικρές και σχετικά όμοιες. Έτσι, αποφεύγεται η εξαγωγή εικόνων που περιγράφουν την κλάση μόνο ως προς το ένα χαρακτηριστικό.

Με την ίδια λογική, η απόσταση της εικόνας I_q από την ''Μέση Εικόνα'' ορίζεται από τη σχέση

$$\delta_{ext}(I_q, C_{mf}; \alpha) = s_1(\delta_i(H_q^{Y}, H_{mf}^{Y}), 4\delta_i(H_q^{Chr}, H_{mf}^{Chr}))\, \delta_{CD}^{Y,Chr}(I_q, C_{mf}; \alpha) \tag{5.6}$$

όπου το α όπως και στη διαδικασία ανίχνευσης ασυνεχειών εκφράζει το ποσοστό συμμετοχής του χρώματος στην απόσταση των εικόνων. Αυτό που υπονοείται από την

παραπάνω εξίσωση, είναι ότι η εικόνα I_q θεωρείται ότι δεν ανήκει στην κλάση C, εφόσον οι αποστάσεις των ιστογραμμάτων έντασης και χρώματος αυτής από τη ''Μέση Εικόνα'', είναι σχετικά μεγάλες και όμοιες μεταξύ τους.

Σε αντίθεση με την ανίχνευση ασυνεχειών, εδώ χρησιμοποιείται η πιο ''απαιτητική'' ως προς την ομοιότητα των αποστάσεων των ορισμάτων της (εν προκειμένω των ιστογραμμάτων χρώματος και έντασης) συνάρτηση σύγκρισης s_1. Η διαφοροποίηση αυτή δικαιολογείται από το γεγονός ότι η τάξη των αποστάσεων που αναμένονται στην κάθε διαδικασία διαφέρει. Για παράδειγμα, στην ανίχνευση ασυνεχειών, μια απόσταση των ιστογραμμάτων έντασης των εικόνων πολύ μεγαλύτερη από το κατώφλι t_{CD} που ακολουθείται από μια απόσταση των ιστογραμμάτων του χρώματος κοντά στο κατώφλι, συνήθως δηλώνει ότι υπάρχει ασυνέχεια. Αυτή την ποιοτική διαπίστωση εκφράζει το κριτήριο s του προηγούμενου κεφαλαίου. Αντίθετα, εδώ μας ενδιαφέρουν αποστάσεις ιστογραμμάτων που ''δυναμικά'' ξεπερνούν το κατώφλι t_{NewC} τόσο στην ένταση όσο και στο χρώμα και το γεγονός ότι οι αποστάσεις αυτές φράσσονται κατά κάποιο τρόπο από το κατώφλι ασυνεχειών, μας επιτρέπει να μετρήσουμε πιο αυστηρά την ομοιότητά τους.

Τέλος, αυτό που διαπιστώνεται είναι ότι αυξάνοντας το συντελεστή α της συμμετοχής των ιστογραμμάτων χρώματος στην ολική απόσταση, μειώνεται ο αριθμός των αντιπροσωπευτικών εικόνων που ατυχώς εξάγονται λόγω μεγάλης κίνησης των αντικειμένων η δυναμική της οποίας, φαίνεται ότι καταγράφεται κυρίως στη συνιστώσα της έντασης των εικόνων. Μάλιστα, υπάρχουν περιπτώσεις που για τιμές του α μεγαλύτερες του 0.8, ο αριθμός των εξαγόμενων εικόνων που προκύπτουν όταν χρησιμοποιείται μόνο η ένταση των εικόνων, υποδιπλασιάζεται. Η διαπίστωση αυτή ισχύει τόσο για το σύστημα χρωμάτων YCbCr, όσο και για το σύστημα YIQ. Αν και όπως ειπώθηκε το αποτέλεσμα της διαδικασίας είναι υποκειμενικό, μπορούμε να πούμε ότι το αποτέλεσμα που παίρνουμε από το YIQ είναι καλύτερο από το YCbCr στο εξής σημείο: σε σκηνές με μεγάλες μεταβολές του περιεχομένου λόγω κίνησης, το YIQ πετυχαίνει να περιγράψει το περιεχόμενο αυτό με λιγότερες εικόνες από ότι το YCbCr. Για το λόγο αυτό προτιμήθηκε το σύστημα YIQ.

Στην παράγραφο που ακολουθεί, περιγράφεται ο τρόπος με τον οποίο ενσωματώνεται η μέθοδος σειριακής ομαδοποίησης στο συνολικό αλγόριθμο καθώς και η αλληλεπίδρασή της με τη διαδικασία ανίχνευσης ασυνεχειών με στόχο την εξαγωγή της απαραίτητης πληροφορίας δεικτοδότησης του video.

5.2.4 Συμπλήρωση του Μοντέλου Δεικτοδότησης

Σε κάθε περίπτωση, εάν μια νέα ομάδα πρέπει να δημιουργηθεί, είτε διότι η απόσταση της I_2 εικόνας του σχ. 5.2 από την τρέχουσα ομάδα ξεπερνάει το κατώφλι t_{NewC} είτε επειδή ανιχνεύθηκε η αρχή νέας σκηνής, υπολογίζεται η αντιπροσωπευτική εικόνα της τρέχουσας ομάδας και δίνεται για δεικτοδότηση στο κομμάτι της εφαρμογής που υλοποιεί τη συμπλήρωση της δομής του μοντέλου δεδομένων του video. Η δεικτοδότηση

περιλαμβάνει την ενημέρωση της ‘‘Λίστας Εικόνων-Κλειδιών’’ της τρέχουσας ‘‘Εισόδου Σκηνής’’ του μοντέλου (βλ. και σχ. 3.1). Επιπρόσθετα, εάν έχει ανιχνευθεί η αρχή νέας σκηνής, τερματίζεται η συμπλήρωση της ‘‘Λίστας Εικόνων-Κλειδιών’’ της τρέχουσας σκηνής και δημιουργείται η επόμενη ‘‘Είσοδος Σκηνής’’ η οποία εισάγεται στο τέλος της ‘‘Λίστας Σκηνών’’ της τρέχουσας ακολουθίας εικόνων. Η τρίτη παράμετρος της εξόδου της διαδικασίας ανίχνευσης ασυνεχειών η οποία αναφέρεται ως ‘‘Πρώτη Εικόνα Νέας Σκηνής’’ στο σχ. 3.1, είναι ακριβώς ο δείκτης στην πρώτη εικόνα της νέας σκηνής του σχ. 5.1.

Με άλλα λόγια, οι διαδικασίες ανίχνευσης αλλαγής σκηνών και εξαγωγής εικόνων-κλειδιών, ‘‘συνεργάζονται’’ έχοντας ως τελικό στόχο τη συμπλήρωση ενός στιγμιοτύπου του μοντέλου δεδομένων που αντανακλά με τρόπο δομημένο, περιεκτικό αλλά και αποδοτικό το περιεχόμενο του video.

5.2.5 Πειραματικά Αποτελέσματα

Σ’ ότι ακολουθεί, παρουσιάζεται η επίδραση των ιστογραμμάτων χρώματος, στη μείωση του αριθμού των εξαγωμένων αντιπροσωπευτικών εικόνων. Το αποτέλεσμα αφορά την τελευταία σκηνή του VIDEO2 του προηγούμενου κεφαλαίου. Η πρώτη και η τελευταία εικόνα αυτής της σκηνής φαίνονται στο σχ. 4.8. Το περιεχόμενο της σκηνής αλλάζει σημαντικά, αφού εισέρχεται σ’ αυτή ένα νέο αντικείμενο. Αυτό έχει ως αποτέλεσμα στην περίπτωση που χρησιμοποιείται μόνο το ιστόγραμμα της έντασης των εικόνων, να εξάγονται τέσσερις αντιπροσωπευτικές εικόνες, οι οποίες φαίνονται στο σχ. 5.3 και περιγράφουν τη σταδιακή είσοδο του αντικειμένου στη σκηνή. Το κριτήριο σύγκρισης ιστογραμμάτων είναι το ίδιο όπως και για την ανίχνευση ασυνεχειών, και το κατώφλι t_{NewC} τέθηκε 0.05. Εάν χρησιμοποιηθεί το κριτήριο δ_{ext} με α=0.8 και η σχέση δ_{int} μέτρησης ομοιότητας των εικόνων της ομάδας, εξάγεται μόνο μία αντιπροσωπευτική εικόνα, αυτή του σχ. 5.4. Η εικόνα αυτή αποτελεί όντως μία συνοπτική περιγραφή της όλης μεταβολής. Επομένως είναι φανερό ότι οι μεταβολές λόγω ταχύτητας καταγράφονται κυρίως στη συνιστώσα της έντασης. Το αν αυτή η εικόνα είναι περιεκτική ως προς το περιεχόμενο που αντιπροσωπεύει, βασίζεται σε καθαρά υποκειμενικά κριτήρια. Θεωρώντας ότι στη βάση αποθηκεύονται η πρώτη, η τελευταία και οι εξαγόμενες αντιπροσωπευτικές εικόνες της κάθε σκηνής, μπορούμε να πούμε πως η εικόνα αυτή δεν αφήνει κενά στην περιγραφή του περιεχομένου. Άλλωστε, το γεγονός ότι επιτρέπεται στο χρήστη να ρυθμίζει το συντελεστή συμμετοχής του χρώματος στο κριτήριο δ_{ext}, του δίνει την ελευθερία να αποφασίσει αν θα αρκεστεί σε μια συνοπτική περιγραφή του περιεχομένου ή σε μια καλύτερη (και πιο δαπανηρή σε χώρο αποθήκευσης) περιγραφή της όλης μεταβολής.

Όπως προκύπτει από τα αποτελέσματα, ο αλγόριθμος επιτυγχάνει υψηλό βαθμό συμπίεσης του περιεχομένου του video. Χρησιμοποιώντας μόνο την πληροφορία της έντασης, εξάγεται συνήθως 2-3% των εικόνων του video. Η χρήση επιπλέον του χρώματος οδηγεί στη μείωση αυτού του ποσοστού περίπου στο 1% των εικόνων του video.

Σχήμα 5.3: Εξαγόμενες αντιπροσωπευτικές εικόνες για την τελευταία σκηνή του VIDEO2 με χρήση ιστογραμμάτων έντασης.

Σχήμα 5.4: Η μοναδική αντιπροσωπευτική εικόνα για την τελευταία σκηνή του VIDEO2 με χρήση ιστογραμμάτων έντασης-χρώματος.

5.2.6 Συμπεράσματα-Επεκτάσεις

Μια πρώτη παρατήρηση που μπορεί να γίνει για την προσέγγιση που περιγράφηκε, είναι ότι σκηνές-μέρη του video με το ίδιο περιεχόμενο οι οποίες δεν είναι διαδοχικές στη γραμμή του χρόνου, αναπαρίστανται από το σχήμα που δόθηκε πιο πάνω από περίπου όμοιες εικόνες-κλειδιά. Από τη στιγμή που οι εικόνες αυτές κατά κάποιο τρόπο, περιγράφουν τη χρονική εναλλαγή του περιεχομένου του video, θεωρείται πως δεν είναι ‘‘περιττές’’ (από την άποψη της αύξησης των απαιτήσεων αποθήκευσης) και επομένως καταγράφονται στο μοντέλο δεδομένων. Άλλωστε, θα ήταν χρονοβόρο και χωρίς σημασία για έναν αλγόριθμο που αποσκοπεί στην συμπλήρωση της δομής του video με ένα πέρασμα κατά τη διάρκεια μετάδοσης, να γίνεται έλεγχος της ομοιότητας του περιεχομένου των σκηνών καθώς αυτές εξάγονται. Κάτι τέτοιο μπορεί βέβαια να αποτελέσει το αντικείμενο ενός άλλου αλγορίθμου με τη χρήση μάλιστα επιπλέον χαρακτηριστικών ομοιότητας των αντιπροσωπευτικών εικόνων των σκηνών που έχουν ήδη εξαχθεί.

Όπως είδαμε, ο αλγόριθμος χρησιμοποιεί για την εξαγωγή των αντιπροσωπευτικών εικόνων τα ιστογράμματα έντασης και χρώματος των I-εικόνων τα οποία έχουν ήδη υπολογιστεί για τις ανάγκες της ανίχνευσης ασυνεχειών. Αυτό έχει ως αποτέλεσμα να μην απαιτείται επιπλέον χρόνος υπολογισμού άλλων χαρακτηριστικών αυξάνοντας τις χρονικές επιδόσεις του συστήματος. Το γεγονός ότι χρησιμοποιούνται μόνο οι I-εικόνες δεν αποτελεί μειονέκτημα αν υπολογίσει κανείς δύο παράγοντες:

- την καλύτερη ποιότητα κωδικοποίησης των I-εικόνων σε σχέση μ’ εκείνη των P-, B-εικόνων που εξασφαλίζει την εξαγωγή χαρακτηριστικών χωρίς μεγάλη επίδραση θορύβου και

- το μικρό διάστημα χρονικής υποδειγματοληψίας του video που όπως περιγράφηκε στο προηγούμενο κεφάλαιο είναι περίπου μία εικόνα ανά 0.5 sec.

Το σχήμα θα μπορούσε να επεκταθεί, ώστε να χρησιμοποιούνται και άλλα χαρακτηριστικά των εικόνων, όπως αυτά που περιγράφονται στο [4]. Η χρήση όμως επιπλέον χαρακτηριστικών, είναι δυνατό να οδηγήσει σε μεγάλο χρόνο απόκρισης του συνολικού συστήματος τμηματοποίησης του video και συνήθως αποφεύγεται σε συστήματα πραγματικού χρόνου.

Μια επέκταση του συνολικού αλγορίθμου ανίχνευσης ασυνεχειών και εξαγωγής αντιπροσωπευτικών εικόνων, είναι να εξάγονται επιπλέον χαρακτηριστικά, όπως είναι τα κυρίαρχα χρώματα και τα πρόσωπα που πιθανώς περιέχονται, αμέσως μετά την εξαγωγή της κάθε αντιπροσωπευτικής εικόνας. Η επέκταση αυτή μπορεί να γίνει άμεσα, ενσωματώνοντας στο σύστημα, αλγορίθμους εξαγωγής κυρίαρχων χρωμάτων όπως αυτός που προτείνεται στο [17], εξαγωγής προσώπων [26] και άλλων χαρακτηριστικών περιγραφής του περιεχομένου. Αν και κάτι τέτοιο αυξάνει σημαντικά το υπολογιστικό

κόστος, θα μπορούσε να δοθεί η δυνατότητα στο χρήστη να επιλέξει αυτή τη διαδικασία, για τις ανάγκες σχετικών εφαρμογών.

Κεφάλαιο 6

Εξαγωγή Αντικειμένων του Video

6.1 Εισαγωγή

Το κεφάλαιο αυτό αποτελεί την εισαγωγή στις μεθόδους εξαγωγής αντικειμένων από εικόνες ή από μια σειρά εικόνων του video βάσει κάποιων χαρακτηριστικών που τα διαφοροποιούν από το background τους ή άλλα αντικείμενα.

Επιχειρώντας μια σύνδεση με το σχήμα δεικτοδότησης του video, μπορούμε να πούμε ότι η εξαγωγή των αντικειμένων θα μπορούσε να είναι η συνέχεια των αλγορίθμων ανίχνευσης αλλαγής σκηνών και εξαγωγής αντιπροσωπευτικών εικόνων έτσι ώστε να συμπληρώνει την προς δεικτοδότηση πληροφορία των σκηνών του video. Η πληροφορία αυτή αφορά τόσο στη θέση και τη σχέση των αντικειμένων μεταξύ τους στο χώρο όσο και στη μεταβολή αυτών των παραμέτρων κατά τη διάρκεια εξέλιξης της σκηνής. Θεωρώντας το σχ. 3.1 του μοντέλου της δεικτοδότησης, η πληροφορία διάκρισης των αντικειμένων κάθε σκηνής θα μπορούσε να κρατείται στην αντίστοιχη είσοδο της λίστας των σκηνών του video.

Επιπλέον, αξίζει να σημειωθεί ότι οι τεχνικές που θα παρουσιαστούν δίνουν μια αφαιρετική διάσταση περιγραφής των εικόνων, όπου τα εξαγόμενα αντικείμενα αποτελούν τις οντότητες στο χώρο. Έχοντας αυτή τη δομική πληροφορία, γίνεται εφικτή η σύγκριση των εικόνων, στη βάση της δομικής (structural) προσέγγισης ανάλυσης των χαρακτηριστικών τους, σε αντίθεση με την προσέγγιση που αναφέρεται σε βασικά χαρακτηριστικά που εξάγονται για το σύνολο της εικόνας, όπως περιγράφηκε στο κεφάλαιο 2. Ευρεία είναι επίσης στις μέρες μας, η χρήση αυτής της δομικής παράστασης για τη δημιουργία οπτικών εφέ, μεταβάλλοντας το περιεχόμενο των εικόνων ή video, μια διαδικασία που είναι γνωστή ως εκ των υστέρων παραγωγή (post-production).

Στο υποκεφάλαιο 6.3, αναλύεται ο αλγόριθμος Seeded Region Growing, ο οποίος χρησιμοποιείται στα κεφάλαια 7 και 8 που ακολουθούν, για την εξαγωγή ομοιόμορφων περιοχών βάσει του χρώματος τους και την παρακολούθηση αντικειμένων στο χρόνο (tracking).

6.2 Αλγόριθμοι Τμηματοποίησης

Η τμηματοποίηση των εικόνων αποτελεί ένα βασικό πρόβλημα της Μηχανικής Όρασης με σημαντικές πρακτικές εφαρμογές. Γνωστές προσεγγίσεις πάνω στο θέμα είναι οι εξής:

1. Τεχνικές εφαρμογής φίλτρων όπως ο ανιχνευτής ακμών του Canny [27],
2. Οι μέθοδοι των snakes [28], των balloons [29] [30] και των level sets [31] [32]
3. Οι τεχνικές επέκτασης και συγχώνευσης περιοχών (region growing and merging) [33] [34] και
4. Οι προσεγγίσεις ολικής βελτιστοποίησης (global optimization) οι οποίες βασίζονται σε συναρτήσεις ενέργειας ή σε στατιστικά κριτήρια που απορρέουν από το νόμο του Bayes [35] [36] και την Περιγραφή Ελαχίστου Μήκους (Minimum Description Length) [37].

Όπως αναφέρεται στο [38], ‘‘μια κοινή ιδιότητα αυτών των αλγορίθμων είναι ότι όλοι κάνουν υποθέσεις για την υπό τμηματοποίηση εικόνα, ελέγχουν χαρακτηριστικά και παίρνουν αποφάσεις με τη χρήση άμεσα ή έμμεσα κατωφλιών. Η βασική διαφορά τους έγκειται στο πεδίο εφαρμογής αυτών των υποθέσεων, ελέγχων και αποφάσεων’’. Μιλώντας γενικά, οι προσεγγίσεις των snakes, των balloons και της ολικής βελτιστοποίησης αποτελούν προεκτάσεις των αριθμητικών μεθόδων ελαχιστοποίησης συναρτήσεων --στην προκειμένη περίπτωση συναρτήσεων ενέργειας-- από μετρήσεις πάνω σε περιφέρειες (contours) ή περιοχές σημείων της εικόνας.

Μια διαφορετική προσέγγιση ακολουθείται από τη μέθοδο των level sets, κατά την οποία επιχειρείται η κατασκευή της λύσης πάνω στο πλέγμα της εικόνας του προβλήματος επίλυσης εξισώσεων καμπυλών, αντί της επίλυσής τους μέσω αριθμητικών μεθόδων. Από την εφαρμογή του αλγορίθμου σε ζητήματα που αφορούν τόσο την τμηματοποίηση εικόνων όσο και ευρύτερα σε διάφορα προβλήματα ομαδοποίησης δεδομένων διαπιστώνεται ότι η τεχνική των level sets, κάτω από ορισμένες προϋποθέσεις είναι γρήγορη και αποτελεσματική. Η μέθοδος εφαρμόστηκε με επιτυχία στην ανίχνευση και παρακολούθηση κινουμένων αντικειμένων από τους Ε. Σηφάκη και Γ. Τζιρίτα [39], στα πλαίσια των δραστηριοτήτων του ΙΤΕ, όπως αναφέρεται και στο κεφάλαιο εξαγωγής και παρακολούθησης αντικειμένων.

Στη βιβλιογραφία αναφέρονται συχνά και αλγόριθμοι που προκύπτουν από το συνδυασμό μιας ή περισσότερων από τις παραπάνω προσεγγίσεις, όπως είναι ο αυτόματος αλγόριθμος Region Competition [38], στον οποίο γίνεται προσπάθεια ενοποίησης των αλγορίθμων των snakes, επέκτασης περιοχών και της ελαχιστοποίησης ολικών Bayes/MDL στατιστικών κριτηρίων. Η βασική ιδέα είναι ο ‘‘συναγωνισμός’’ των περιοχών κατά μήκος των κοινών συνόρων τους. Ο συναγωνισμός αυτός οδηγεί στην επέκταση περιοχών σε βάρος κάποιων άλλων τα οποία είναι δυνατόν να εξαλειφθούν εφόσον δεν πληρούν κάποιες προϋποθέσεις με βάση στατιστικά κριτήρια ομαδοποίησης των σημείων τους.

Η εξάλειψη περιοχών αποτελεί μια επιπλέον ιδιότητα αυτού του αλγορίθμου σε σχέση με τους άλλους και είναι ακριβώς αυτή η ιδιότητα που επιτρέπει την αυτοματοποίηση του: αυτό που χρειάζεται είναι απλά η επιλογή περισσότερων αρχικών περιοχών από τον αριθμό των αντικειμένων, με την προοπτική ο αλγόριθμος να εξαλείψει όσες απ' αυτές δεν αντιστοιχούν σε πραγματικές, ομοιόμορφες περιοχές της εικόνας.

Όπως θα φανεί και στην επόμενη παράγραφο όπου περιγράφεται ο αλγόριθμος Seeded Region Growing, σημαντικό ρόλο στην τελική τμηματοποίηση, παίζει η αρχικοποίηση αυτών των αλγορίθμων πάνω στο πλέγμα της εικόνας. Όπως ακριβώς συμβαίνει γενικότερα στα προβλήματα ελαχιστοποίησης συναρτήσεων κόστους ή επίλυσης εξισώσεων, μια ατυχής αρχικοποίηση των παραμέτρων εισόδου μπορεί να οδηγήσει στην τερματισμό της διαδικασίας τμηματοποίησης σε κάποιο τοπικό ελάχιστο.

6.3 Ο Αλγόριθμος Seeded Region Growing (SRG)

Ο αλγόριθμος Seeded Region Growing (SRG) [33], ανήκει στην κατηγορία αλγορίθμων επέκτασης περιοχών. Οι αλγόριθμοι αυτοί είναι μή επαναληπτικοί και αυτό τους κάνει αποδοτικούς από άποψη χρόνου. Η βασική ιδέα αυτών των αλγορίθμων, είναι η επέκταση αρχικών περιοχών, με την εισαγωγή σ' αυτές γειτονικών τους σημείων βάσει κάποιων χαρακτηριστικών των περιοχών.

Η μέθοδος SRG, είναι μια μέθοδος τμηματοποίησης μιας εικόνας με βάση ένα σύνολο σημείων που ονομάζονται ''σπόροι'' (seeds). Η αρχικοποίησή του γίνεται με έναν αριθμό σπόρων, οι οποίοι ομαδοποιούνται σε n σύνολα, έστω $R_1, R_2, \ldots, R_n$, ξένων μεταξύ τους. Η επιλογή αυτών των συνόλων είναι σημαντική για τη συνέχεια της διαδικασίας, αφού κατά κάποιο τρόπο ορίζει ποιό είναι το χαρακτηριστικό της περιοχής που θα προκύψει για κάθε σύνολο και τί θεωρείται ''άσχετο'' ή θόρυβος προς αυτή. Αφού δοθούν αυτά τα σύνολα, ο αλγόριθμος βρίσκει έναν κατάλληλο χωρισμό της εικόνας σε περιοχές, με την ιδιότητα ότι κάθε συνδεδεμένο στοιχείο μιας περιοχής ανήκει σε ένα μόνο από τα R_i. Έτσι μεγιστοποιείται η ομοιομορφία των περιοχών.

Η διαδικασία τμηματοποίησης συνεχίζεται επαναληπτικά, ξεκινώντας από τα αρχικά σύνολα $R_1, R_2, \ldots, R_n$. Στην αρχή για κάθε ένα από τα σύνολα R_i υπολογίζονται τα χαρακτηριστικά (features) βάσει των οποίων γίνεται η τμηματοποίηση. Ο υπολογισμός βασίζεται στα αρχικά τους σημεία.

Κάθε βήμα του αλγορίθμου περιλαμβάνει την είσοδο ενός σημείου σε ένα από τα παραπάνω σύνολα. Ας θεωρήσουμε τώρα την κατάσταση στην οποία βρίσκονται τα σύνολα μετά από m βήματα. Έστω ότι T είναι το σύνολο των σημείων που δεν έχουν εισαχθεί σε κάποιο από τα R_i, αλλά συνορεύουν με ένα απ'αυτά

$$\mathrm{T}=\left\{ x \notin \bigcup_{i=1}^{n} R_i | N(x) \cap \bigcup_{i=1}^{n} R_i \neq \emptyset \right\}$$

όπου $N(x)$ είναι το σύνολο των αμέσων γειτόνων του σημείου x (για παράδειγμα η

γειτονιά των 8 σημείων που άμεσα συνδέονται με το x). Εάν για κάποιο $x \in T$, ισχύει ότι το N(x) περιλαμβάνει σημεία που ανήκουν σε ένα μόνο από τα R_i, τότε ορίζουμε ως $i(x) \in \{1, 2, \ldots, n\}$ το δείκτη του μοναδικού αυτού συνόλου και ορίζουμε το $\delta(x; R_{i(x)})$ να είναι μια μέτρηση του πόσο όμοιο (ή διαφορετικό) ως προς τα χαρακτηριστικά του συνόλου $R_{i(x)}$ είναι το σημείο x.

Εάν η γειτονιά $N(x)$ περιέχει σημεία που ανήκουν σε περισσότερα από ένα σύνολα, μπορούμε να θεωρήσουμε το σημείο x ως συνοριακό και να το εντάξουμε στο σύνολο B των συνοριακών σημείων που έχουν βρεθεί μέχρι το βήμα m.

Το βήμα m + 1 ολοκληρώνεται, εισάγοντας το σημείο $z \in T$, του οποίου η απόσταση από το μοναδικό σύνολο με το οποίο συνορεύει είναι η ελάχιστη --σε σχέση μ' αυτή των άλλων σημείων του συνόλου T, από τα αντίστοιχα γειτονικά τους σύνολα-- στο σύνολο $R_{i(z)}$. Στη συνέχεια υπολογίζονται ξανά τα χαρακτηριστικά του συνόλου $R_{i(z)}$ και περνάμε στο επόμενο βήμα. Ο αλγόριθμος τελειώνει όταν όλα τα σημεία έχουν εισαχθεί σε κάποιο σύνολο από τα R_i, ή στο σύνολο B. Το αποτέλεσμα είναι η τμηματοποίηση της εικόνας σε n ξένα μεταξύ τους σύνολα.

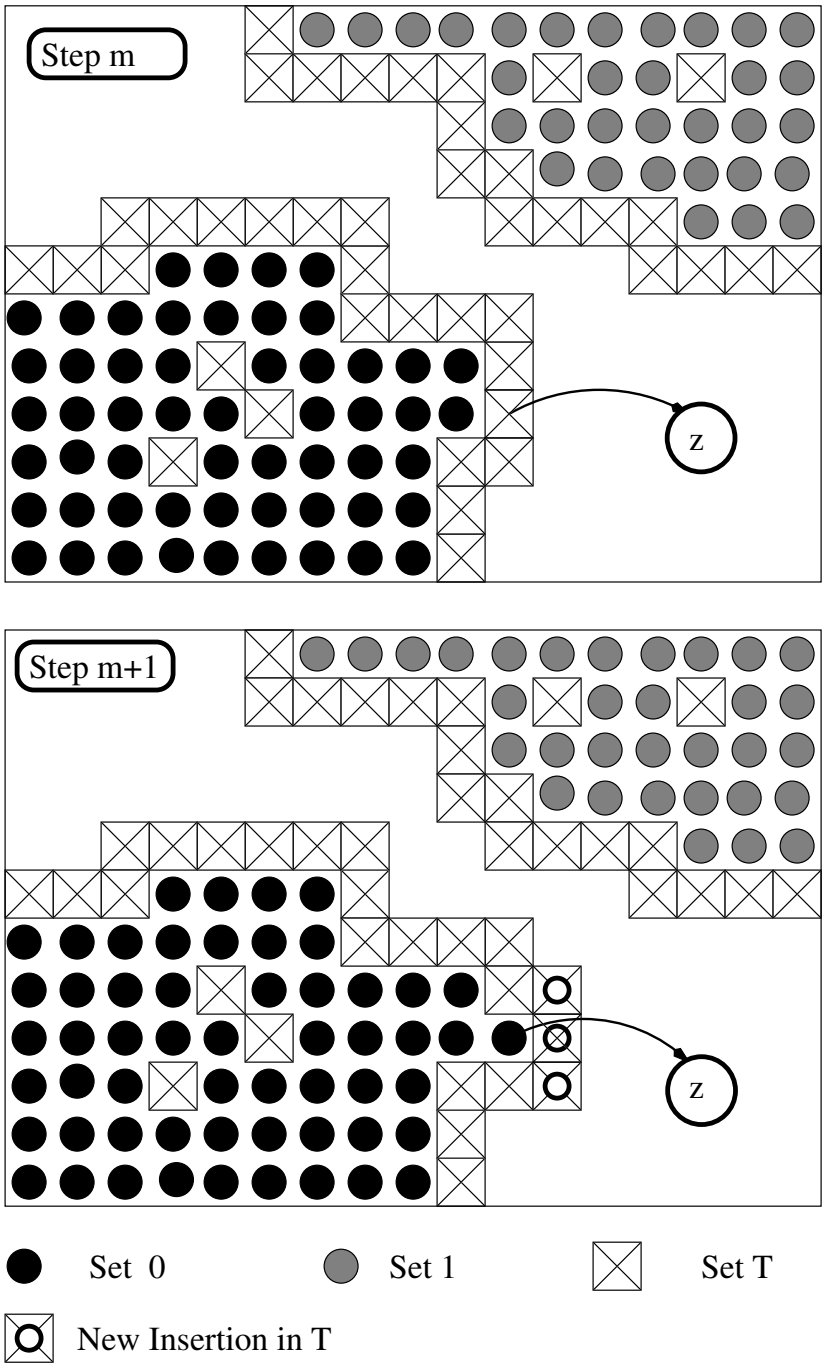
Ό,τι περιγράφηκε φαίνεται παραστατικά στο σχ. 6.1, για δύο επεκτεινόμενες περιοχές, τις 0 και 1. Η κατάσταση που έχει προκύψει μέχρι το βήμα m, περιγράφεται στην πάνω εικόνα του σχήματος. Οι μαύροι και γκρι κύκλοι αντιστοιχούν στα σημεία που ήδη ανήκουν στα σύνολα 0 και 1 αντίστοιχα. Με τετράγωνα συμβολίζονται τα σημεία που έχουν εισαχθεί στο σύνολο T, ενώ υποτίθεται ότι το σημείο του συνόλου T με την ελάχιστη απόσταση από το σύνολο με το οποίο γειτονεύει είναι αυτό που ορίζεται ως z. Τα βήματα που γίνονται για να περάσουμε στο βήμα m + 1, φαίνονται στην κάτω εικόνα. Το σημείο z εισήχθηκε στο σύνολο 0, ενώ οι γείτονές του που δεν ανήκουν στο σύνολο T, εισάγονται σ' αυτό. Τα σημεία αυτά συμβολίζονται ως τετράγωνα με κύκλο στο κέντρο τους. Έτσι, τελειώνει το βήμα m + 1.

Για την υλοποίηση του SRG, είναι απαραίτητη η χρήση μιας δομής δεδομένων που να διατηρεί τα αντικείμενα τα οποία περιέχει, ταξινομημένα ως προς κάποια ιδιότητά τους (attribute). Σ' ό,τι αφορά τον αλγόριθμο SRG τα αντικείμενα αυτά είναι οι συντεταγμένες των σημείων και η ιδιότητα είναι το κριτήριο ομοιότητας. Η δομή αυτή ονομάζεται παραδοσιακά ''σειριακά ταξινομημένη λίστα'' (sequentially sorted list -- SSL). Για να είναι ο αλγόριθμος αποδοτικός θα πρέπει η λίστα αυτή να επιτρέπει τη διατήρηση της διάταξης των στοιχείων της μετά από εισαγωγή και την έξοδο του ελαχίστου από αυτή, στον ελάχιστο δυνατό χρόνο. Για το λόγο αυτό ενδείκνυται η χρήση δομών όπως είναι τα AVL δένδρα ή οι Ουρές Προτεραιότητας, πολύ περισσότερο που ο χρόνος εκτέλεσης του αλγορίθμου αναλώνεται στο μεγαλύτερο ποσοστό του, στις εισαγωγές σημείων και στην έξοδο του ελαχίστου από τη λίστα.

Έχοντας αυτή τη δομή,ο αλγόριθμος SRG σε ψευδοκώδικα είναι ο εξής :

B1 Εισαγωγή των σπόρων στα σύνολα στα οποία έχουν αρχικά ομαδοποιηθεί.

B2 Εισαγωγή των σημείων που είναι γείτονες των σπόρων στην SSL.



Σχήμα 6.1: Αναπαράσταση ενός βήματος του SRG

Β3 Όσο υπάρχουν σημεία στην SSL κάνε τα εξής :

Β3.1 Εξαγωγή του πρώτου σημείου της SSL (σημείο y).

Β3.2 Εξέταση των γειτόνων του y :

---Εάν όλοι οι γείτονες του y οι οποίοι έχουν εισαχθεί σε κάποιο σύνολο (διάφορο του B), ανήκουν στο ίδιο σύνολο R_i κάνε τα εξής :

Β3.2.1 Εισαγωγή του y στο R_i.

Β3.2.2 Υπολογισμός των χαρακτηριστικών τμηματοποίησης του συνόλου R_i.

Β3.2.3 Είσοδος των γειτόνων του y x, οι οποίοι

α) δεν ανήκουν σε κάποιο σύνολο

β) δεν έχουν εισαχθεί στην SSL

στην SSL βάσει του κριτηρίου ομοιότητας $\delta(x; R_i)$.

---αλλιώς

Β3.2.4 Είσοδος του y στο σύνολο B.

Το αποτέλεσμα της τμηματοποίησης είναι ο πίνακας $i(p)$ αντιστοίχισης κάθε σημείου p, στο σύνολο στο οποίο ανήκει.

Μια πρώτη παρατήρηση σε ότι αφορά τον αλγόριθμο στην πρώτη του μορφή για εικόνες έντασης (βλ. [33]), είναι ότι δεν επαναυπολογίζεται η απόσταση σημείων της λίστας από το γειτονικό τους σύνολο, κάθε φορά που μεταβάλλονται τα χαρακτηριστικά του με την είσοδο σ' αυτό νέων σημείων. Η υπόθεση που γινόταν είναι ότι η μεταβολή αυτή είναι μικρή διότι

- η επιλογή των αρχικών συνόλων έχει γίνει έτσι ώστε να περιγράφουν ικανοποιητικά το περιεχόμενο της περιοχής που προορίζονται να καταλάβουν και,
- η δομή του αλγορίθμου είναι τέτοια που δεν επιτρέπει την είσοδο σημείων σε κάποιο σύνολο εάν η απόστασή τους απ' αυτό είναι μεγάλη, εφόσον υπάρχουν γειτονικά σημεία του συνόλου με μικρότερες αποστάσεις απ' αυτό. Με άλλα λόγια, τα αρχικά σύνολα επεκτείνονται τηρώντας την αρχή της συνεκτικότητας των σημείων τους, σε περιοχές όσο το δυνατόν πιο ομοιόμορφες ως προς αυτά (μεγιστοποίηση της ομοιομορφίας). Σημεία των οποίων τα χαρακτηριστικά δεν εκφράζονται από κάποιο σύνολο προκύπτουν συνήθως διότι δεν υπάρχει αρχικό σύνολο-αντιπρόσωπος της περιοχής στην οποία ανήκουν και αναγκαστικά εισάγονται σε κάποιο από τα ήδη υπάρχοντα, στο τέλος της διαδικασίας εισαγωγής. Αυτή η ''ετεροχρονισμένη'' εισαγωγή επιτυγχάνεται με τη χρήση της λίστας.

Συμπεραίνοντας, η μή ανανέωση της λίστας των σημείων μειώνει σημαντικά το χρόνο εκτέλεσης χωρίς να προκαλεί προβλήματα εάν τα αρχικά σύνολα αντιστοιχούν στις ομοιόμορφες περιοχές που εμφανίζονται στην εικόνα. Διαφορετικά υπάρχει κίνδυνος η

μέθοδος να σταματήσει σε κάποιο τοπικό ελάχιστο, κάτι που έχει γενικότερα διαπιστωθεί για τους αλγορίθμους επέκτασης περιοχών.

Εάν, παρόλα αυτά, η τμηματοποίηση δεν είναι ικανοποιητική, όπως αναφέρεται στο [33], μπορεί να εφαρμοστεί ο αλγόριθμος SRG, σε περιοχές που πρέπει να τμηματοποιηθούν περαιτέρω ή η συγχώνευση (merging) ομοιόμορφων περιοχών σε μία.

Τέλος, είναι προφανές ότι καμία υπόθεση δε γίνεται για το σχήμα των αρχικών συνόλων. Έτσι, στις εφαρμογές που χρησιμοποιήθηκε ο αλγόριθμος τα αρχικά σύνολα είχαν μία από τις τρεις μορφές:

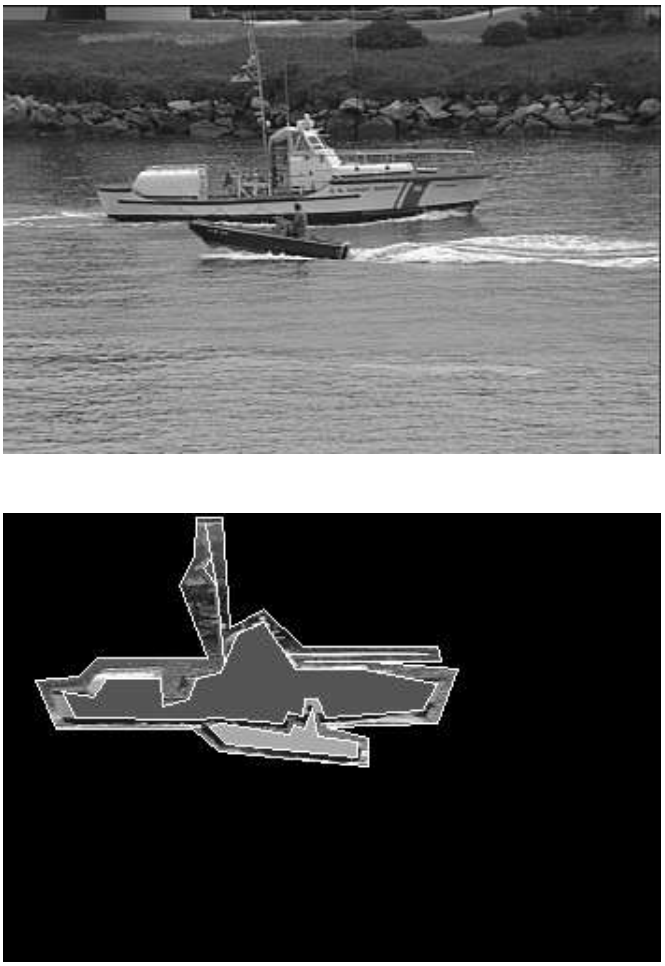
- Τετράγωνα μεταβλητής διάστασης στο εσωτερικό των περιοχών.
- Κλειστές πολυγωνικές περιοχές, οι οποίες ακολουθούν κατά προσέγγιση τα όρια των αντικειμένων όπως φαίνεται στο σχ. 6.2. Στην κάτω εικόνα του σχήματος φαίνεται η αρχικοποίηση κατ' αυτό τον τρόπο για τα δύο αντικείμενα της πάνω εικόνας και το background. Αν και κάποιος θα μπορούσε να παρατηρήσει ότι η εφαρμογή του αλγορίθμου σ' αυτή την περίπτωση γίνεται τετριμμένη, θα άλλαζε ίσως άποψη αν ήταν αναγκασμένος να κάνει ο ίδιος την οριοθέτηση των συνόρων των αντικειμένων με το χέρι, στη λεπτομέρεια που δίνεται από τον αλγόριθμο.
- Ημιτελείς περιοχές που προκύπτουν από την εφαρμογή άλλων αλγορίθμων και δίνονται στον αλγόριθμο SRG για την ολοκλήρωση της τμηματοποίησης.

Στις δύο πρώτες περιπτώσεις, η είσοδος δίνεται από το χρήστη.

Από την εφαρμογή του αλγορίθμου σε διάφορα προβλήματα τμηματοποίησης έχει διαπιστωθεί ότι τα μειονεκτήματά του είναι γενικά τα προβλήματα που παρουσιάζουν οι αλγόριθμοι επέκτασης περιοχών. Πιο συγκεκριμένα, είναι γνωστό ότι οι αλγόριθμοι αυτοί αποτυγχάνουν να εξάγουν περιοχές που είναι επιμήκεις (όπως γραμμές μικρού πάχους) ή γενικότερα μικρού εμβαδού. Στην περίπτωση που ο συγκεκριμένος αλγόριθμος χρησιμοποιηθεί όπως περιγράφεται πιο πάνω, το πρόβλημα έγκειται στο

- αν οι διαστάσεις της περιοχής είναι τέτοιες που να επιτρέπουν την ''εμφύτευση'' σπόρων στο εσωτερικό της με μία από τις μεθόδους που περιγράφηκαν πιο πάνω και
- στο κατά πόσο αυτοί οι σπόροι περιγράφουν ικανοποιητικά τα χαρακτηριστικά της περιοχής.

Καταλήγουμε επομένως στο συμπέρασμα ότι το πραγματικό πρόβλημα μετατοπίζεται στην περίπτωση του SRG, στην αδυναμία επιλογής αντιπροσωπευτικών σπόρων (αρχικοποίησης) σε περιοχές που θα μπορούσαν λόγω της μορφής τους να χαρακτηριστούν ακμές ή θόρυβος, το οποίο όμως αποτελεί μέρος του προβλήματος μέτρησης χαρακτηριστικών σε μια ανομοιόμορφη γειτονιά των σημείων στις ακμές της εικόνας.



Σχήμα 6.2: Ορισμός περιοχών εκκίνησης του αλγορίθμου SRG (κάτω εικόνα) για την πάνω εικόνα

Κεφάλαιο 7

Εξαγωγή Ομοιόμορφων ως προς το Χρώμα Περιοχών

7.1 Εισαγωγή-Ανασκόπηση

Στο κεφάλαιο αυτό, περιγράφεται η εφαρμογή του αλγορίθμου SRG στην τμηματοποίηση εικόνων, με βάση το χρώμα των αντικειμένων. Η χρήση της μεθόδου προκύπτει από όσα ειπώθηκαν στο προηγούμενο κεφάλαιο γενικά για τις εφαρμογές των αλγορίθμων τμηματοποίησης.

Γενικά είναι γνωστό, ότι η αντίληψη του χρώματος των εικόνων, εξαρτάται από παραμέτρους, όπως είναι ο φωτισμός του περιβάλλοντος χώρου ή η απόσταση θέασης. Επίσης, είναι δεδομένη η επίδραση της αντανάκλασης του φωτός στην εικόνα των αντικειμένων. Η επίδραση αυτή έχει να κάνει με τη θέση και την απόσταση της πηγής του φωτός, αλλά και με τη γεωμετρία και το υλικό του αντικειμένου, το οποίο δέχεται το φως [27]. Ειδικότερα στις εφαρμογές τμηματοποίησης χρώματος, το πρόβλημα που δημιουργείται από την αντανάκλαση του φωτός σε σχετικά λείες επιφάνειες (όπως μεταλλικά αντικείμενα ή το δέρμα), είναι ότι στα περισσότερα συστήματα αντίληψης χρώματος, καταγράφονται μεγάλες διαφορές φωτεινότητας των σημείων της περιοχής του αντικειμένου, χωρίς να παρατηρείται το ίδιο στις συνιστώσες χρώματος. Ανάλογη είναι και η επίδραση σκιών πάνω στα αντικείμενα. Από την άλλη πλευρά, σε αρκετές περιπτώσεις εμφανίζονται περιοχές, που χαρακτηρίζονται από υψηλή φωτεινότητα και έλλειψη χρώματος (highlights).

Όλα αυτά κάνουν την τμηματοποίηση χρώματος, ένα αρκετά δύσκολο πρόβλημα, ιδιαίτερα σε ότι αφορά τον ορισμό της ομοιομορφίας των περιοχών στα γνωστά συστήματα μέτρησης χρώματος. Στις προσεγγίσεις μοντελοποίησης του χρώματος αντικειμένων που αναφέρονται στη βιβλιογραφία, συνήθως γίνονται υποθέσεις για το υλικό του αντικειμένου και τις συνθήκες φωτεινότητας [40], παράμετροι που είναι δύσκολο να ελεγχθούν στην τμηματοποίηση εικόνων γενικού περιεχομένου που σκοπό έχει την τμηματοποίηση σε ομοιόμορφες, ξένες μεταξύ τους, περιοχές. Μια ενδιαφέρουσα πρόταση

στο ζήτημα τόσο του υπολογισμού των παραμέτρων χρώματος των αντικειμένων όσο και της τμηματοποίησης, αποτελεί η εφαρμογή του αλγορίθμου Region Competition στην τμηματοποίηση του χρώματος [38], αν και πρέπει να σημειωθεί ότι τα αποτελέσματα αφορούν απλές εικόνες πάνω σε ευδιάκριτα ως προς το χρώμα τους αντικείμενα. Μια διαφορετική προσέγγιση στο θέμα εξαγωγής χρωματικών περιοχών ακολουθείται στο [41], η οποία βασίζεται στη διάκριση των περιοχών χρώματος σε ομοιόμορφες και περιοχές που περιέχουν υφή ή/και ακμές. Η χονδρική (coarse) αυτή διάκριση των περιοχών επιτρέπει τη χρήση τους για την ανάκτηση εικόνων. Αν και η εργασία δε σχετίζεται άμεσα με την τμηματοποίηση χρώματος, βοηθάει στην κατανόηση της διαφοράς των μεθόδων που αποσκοπούν στην αφαιρετική παράσταση των εικόνων μέσω οντοτήτων, σε σχέση με αυτές που στοχεύουν στη χρήση του περιεχομένου για συγκρίσεις εικόνων κατά την ανάκτηση.

Σε ό,τι ακολουθεί, παρουσιάζεται ο βασικός αλγόριθμος SRG, ο οποίος μπορεί να χρησιμοποιηθεί ως έχει, αρκεί να υπάρχει ένα γραφικό περιβάλλον που να διευκολύνει την είσοδο των αρχικών συνόλων από το χρήστη. Στη συνέχεια, περιγράφεται μια μέθοδος αυτόματης εξαγωγής ομοιόμορφων περιοχών χρώματος. Η μέθοδος αυτή χρησιμοποιήθηκε για την εξαγωγή χρώματος προσώπων από εικόνες. Η πληροφορία αυτή μπορεί να χρησιμοποιηθεί στη συνέχεια για τις ανάγκες δεικτοδότησης των εικόνων. Η αυτοματοποίηση της μεθόδου, αναφέρεται στην ύπαρξη μιας διαδικασίας εξαγωγής των αρχικών περιοχών η οποία αναλύεται πιο κάτω.

7.2 Τμηματοποίηση Χρώματος βάσει του SRG

7.2.1 Περιγραφή

Η περιγραφή που ακολουθεί, αναφέρεται στην εφαρμογή του SRG ως ενός *καθοδηγούμενου από το χρήστη* (user-guided), προγράμματος τμηματοποίησης Για την τμηματοποίηση χρώματος, η καθοδήγηση αυτή αναφέρεται

- στην επιλογή του χώρου χρώματος στον οποίο θα εφαρμοστεί η τμηματοποίηση,
- στην είσοδο των αρχικών συνόλων από σπόρους $R_1, R_2, \ldots, R_n$. Η είσοδος είναι δυνατό να γίνει με τους τρόπους που αναφέρθηκαν στο προηγούμενο κεφάλαιο. Είτε δηλαδή ως τετράγωνα, είτε ως πολυγωνικές κλειστές περιοχές στο εσωτερικό των αντικειμένων.

Στην προσέγγιση που παρουσιάζεται, χρησιμοποιήθηκαν για την τμηματοποίηση των εικόνων γνωστά συστήματα χρώματος τα οποία έχουν χρησιμοποιηθεί το καθένα για διαφορετικούς σκοπούς ανάλογα με τις ιδιότητές του, στη διαχείριση και ανάλυση δεδομένων. Έτσι, δίνεται η δυνατότητα στο χρήστη να επιλέξει ένα από τα εξής συστήματα χρώματος:

- RGB

- YCbCr
- YIQ
- Lab
- Luv

Ως χαρακτηριστικό κάθε συνόλου R_i της τμηματοποίησης, θεωρείται η μέση τιμή (μ_1, μ_2, μ_3) της κάθε συνιστώσας του χρώματος των σημείων του, στο σύστημα που έχει επιλεγεί. Στη συνέχεια εφαρμόζεται ο SRG, θεωρώντας ότι η απόσταση $\delta(p; R_i)$ του σημείου p=(x, y) από το σύνολο R_i, δίνεται από την ευκλείδια απόσταση

$$\delta(x, R_i) = \sqrt{\sum_{k=1}^{3} (I_k(p) - \mu_k)^2} \tag{7.1}$$

όπου $(I_1(p), I_2(p), I_3(p))$ είναι το διάνυσμα χρώματος του σημείου p.

Εφόσον ο χρήστης έχει επιλέξει κάποιον από τους γραμμικούς χώρους χρώματος (RGB, YCbCr ή YIQ), του παρέχεται η δυνατότητα τμηματοποίησης της εικόνας, μειώνοντας την επίδραση της φωτεινότητας. Η επίδραση αυτή είναι δεδομένη, αφού οι χώροι YCbCr και YIQ, έχουν σχεδιαστεί έτσι ώστε η σημαντική πληροφορία να καταγράφεται (στατιστικά) στη συνιστώσα της φωτεινότητας και όχι στη χρωματικότητα. Για το χώρο RGB, η μέθοδος αυτή αποτελεί ένα τρόπο μετάβασης του, από χώρο εμφάνισης στην οθόνη, των χρωμάτων, σε χώρο μέτρησης χρώματος. Η δυνατότητα αυτή βασίζεται σε μια ιδιότητα αυτών των χώρων η οποία χρησιμοποιείται ευρέως για την επεξεργασία της φωτεινότητας και του κορεσμού των εικόνων χρώματος [42]. Στην ενότητα 7.2.2 παρουσιάζεται η μέτρηση ομοιότητας σημείου από περιοχή, σύμφωνα μ' αυτή την ιδιότητα.

Εφόσον το αποτέλεσμα δεν είναι ικανοποιητικό, δίνεται η δυνατότητα στο χρήστη να το διορθώσει αλληλοδραστικά

- συγχωνεύοντας περιοχές που θεωρεί όμοιες (merging procedure)
- τμηματοποιώντας επιπλέον δύο ή περισσότερες περιοχές (sub-spliting procedure).

Η διαδικασία είναι δυνατό να επαναληφθεί όσες φορές επιθυμεί ο χρήστης. Παρέχοντας αυτή τη δυνατότητα, ξεπερνιούνται προβλήματα που σχετίζονται με τυχόν λάθος αποτελέσματα λόγω της διαφοράς φωτεινότητας περιοχών των αντικειμένων.

7.2.2 Κριτήριο Ομοιότητας Χρώματος Γραμμικών Χώρων

Στην ενότητα αυτή, περιγράφεται μια μέθοδος μέτρησης της απόστασης δύο χρωμάτων, η οποία μπορεί να εφαρμοστεί στην περίπτωση που ο χώρος που έχει επιλεγεί είναι το RGB ή ένας χώρος από τους YCbCr ή YIQ, οι οποίοι προκύπτουν γραμμικά από το RGB.

Η μέτρηση βασίζεται σε μια γνωστή ιδιότητα που προκύπτει από τη θέση των χρωμάτων στον κύβο του χώρου RGB, η οποία διατηρείται και στους χώρους που προκύπτουν από αυτόν μέσω γραμμικών μετασχηματισμών.

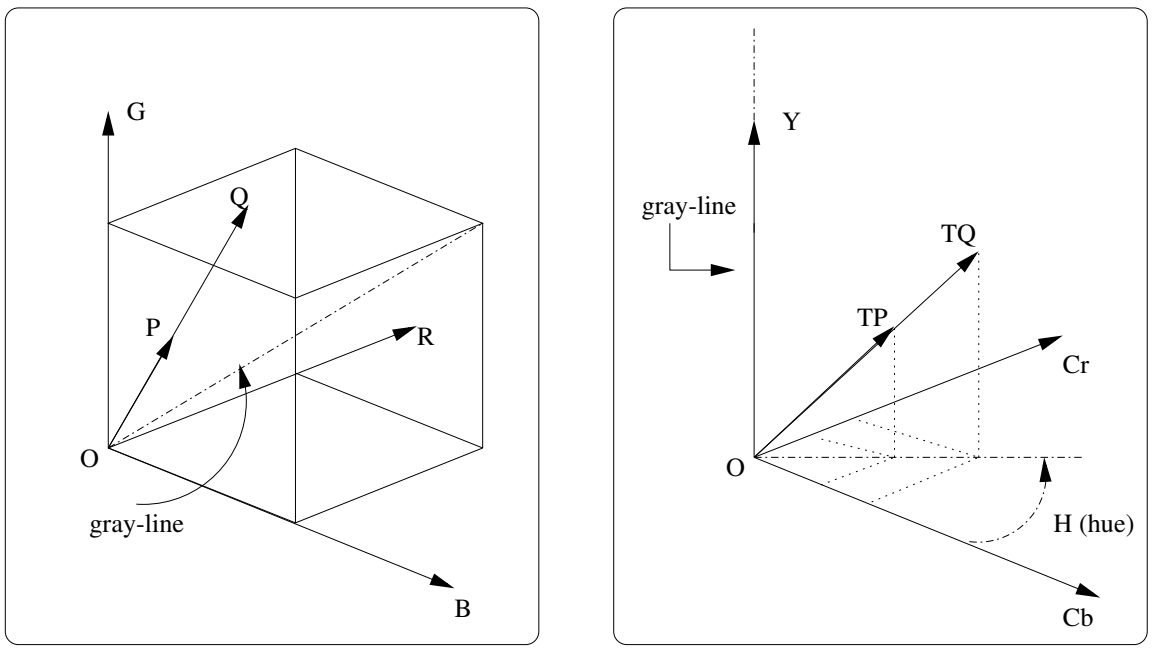
Η ιδιότητα αφορά χρώματα του χώρου RGB, τα οποία παράγονται από το ίδιο μοναδιαίο διάνυσμα. Στο σχ. 7.1(α), τα σημεία $P(R_1, G_1, B_1)$ και $Q(R_2, G_2, B_2)$ είναι δύο διανύσματα χρώματος μ' αυτή την ιδιότητα. Εφόσον τα χρώματα αυτά συνδέονται γραμμικά, υπάρχει συντελεστής γ έτσι ώστε να ισχύει:

$$R_2 = \gamma R_1 \tag{7.2}$$

$$G_2 = \gamma G_1 \tag{7.3}$$

$$B_2 = \gamma B_1 \tag{7.4}$$

Ορίζοντας την απόχρωση και τον κορεσμό με βάση τη γραμμή του άσπρου-μαύρου (gray-line) της εικόνας, τα χρώματα αυτά αντιστοιχούν στην ίδια απόχρωση αλλά σε διαφορετικό κορεσμό και φωτεινότητα. Στο σχ. 7.1(β), παρατηρούμε ότι η αναλογία αυτή των χρωμάτων ισχύει και στο χώρο YCbCr για τα χρώματα $TP(Y_1, Cb_1, Cr_1)$ και T$Q(Y_2, Cb_2, Cr_2)$, που προκύπτουν για τα χρώματα P και Q αντίστοιχα και μάλιστα με τον ίδιο συντελεστή αναλογίας γ. Στην πραγματικότητα η ιδιότητα αυτή ισχύει για οποιοδήποτε χώρο προκύπτει γραμμικά από το χώρο RGB καθώς επίσης και για τις χρωματικές συνιστώσες (a, b) στο χώρο Lab μόνο που ο συντελεστής αναλογίας είναι πλέον $\gamma^{\frac{1}{3}}$.



(α) (β)

Σχήμα 7.1: Ιδιότητα αναλογίας χρωμάτων (α) στο RGB και (β) στο YCbCr

Όπως φαίνεται στο σχ. 7.2, βάσει αυτής της ιδιότητας, είναι δυνατό να θεωρηθεί ως απόσταση $\delta(p;R)$ του χρώματος ενός σημείου p=(x,y) από τη μέση τιμή του συνόλου R $\vec{\mu}$=(μ_1,μ_2,μ_3), η απόσταση του διανύσματος χρώματος $\vec{I}$=(I_1,I_2,I_3) του σημείου από την ευθεία ϵ που ορίζει το διάνυσμα χρώματος $\vec{\mu}$. Στο σχ. 7.2, η απόσταση αυτή συμβολίζεται ως $\delta(I,\mu)$. Αρχικά, υπολογίζεται η προβολή $\alpha\vec{\mu}$ του διανύσματος $\vec{I}$ στο διάνυσμα $\vec{\mu}$. Ο συντελεστής α προκύπτει από τη σχέση:

$$\alpha=\frac{\vec{I}\bullet\vec{\mu}}{\|\vec{\mu}\|^2} \tag{7.5}$$

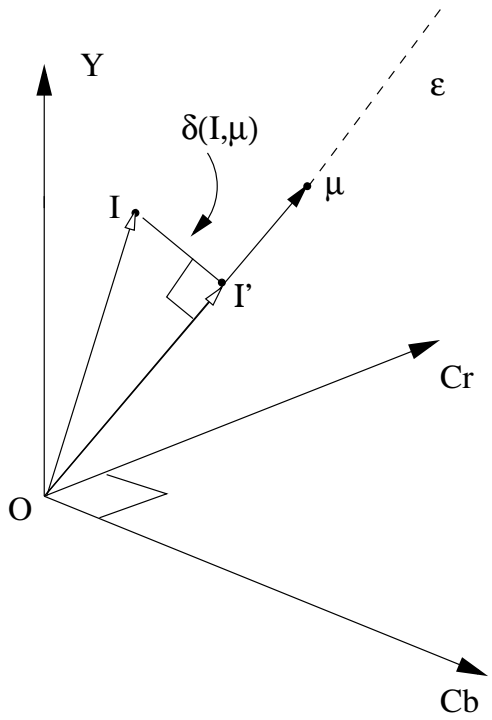
όπου $\bullet$ είναι ο τελεστής του εσωτερικού γινομένου δύο διανυσμάτων και $\|\ \|$ το μέτρο του διανύσματος. Το διάνυσμα $\alpha\vec{\mu}$, συμβολίζεται στο σχ. 7.2 ως I'. Στη συνέχεια υπολογίζεται η απόσταση $\delta(p;R)$ από τη σχέση:

$$\delta(p;R)=\|\vec{I}-\alpha\vec{\mu}\| \tag{7.6}$$

Αυτή είναι η απόσταση του διανύσματος $\vec{I}$ από την προβολή του πάνω στο διάνυσμα $\vec{\mu}$ της μέσης τιμής του συνόλου. Μ' αυτό τον τρόπο, το χαρακτηριστικό των συνόλων είναι το μοναδιαίο διάνυσμα του μέσου διανύσματος των χρωμάτων τους, ενώ όλα τα χρώματα που βρίσκονται στην ευθεία που ορίζει θεωρούνται όμοια. Η ευθεία αυτή όπως ειπώθηκε, αντιστοιχεί σε σταθερή απόχρωση των χρωμάτων, ανεξαρτήτως φωτεινότητας και κορεσμού. Επιπλέον, παρέχεται η δυνατότητα να γίνουν μετρήσεις απόστασης των χρωμάτων και σε χώρους που δεν έχουν σχεδιαστεί γι' αυτό το σκοπό όπως είναι οι γραμμικοί χώροι. Το μειονέκτημα αυτής της μεθόδου είναι πως το χρώμα των περιοχών της εικόνας θα πρέπει να είναι τέτοιο που να επιτρέπει τον υπολογισμό του διανύσματος απόχρωσης. Είναι γνωστό πως κάτι τέτοιο δεν μπορεί να γίνει για άχρωμες περιοχές των οποίων το μέσο διάνυσμα βρίσκεται κοντά στη γραμμή άσπρου-μαύρου (άχρωμες περιοχές). Από τα αποτελέσματα φαίνεται πως οι περιοχές αυτές δε μπορούν να ''συναγωνιστούν'' την επέκταση των περιοχών με έντονα χρώματα, αφού εξαρχής το υπολογιζόμενο διάνυσμά τους δεν ανταποκρίνεται στο χρώμα τους. Αυτό που θα μπορούσε να γίνει είναι να εξαιρούνται από τη διαδικασία της επέκτασης σημεία των οποίων το χρώμα δεν ικανοποιεί κάποιες απαιτήσεις ως προς την απόχρωση και να απορρίπτονται αρχικές περιοχές που περιλαμβάνουν τέτοια σημεία σε μεγάλο ποσοστό.

7.2.3 Πειραματικά Αποτελέσματα

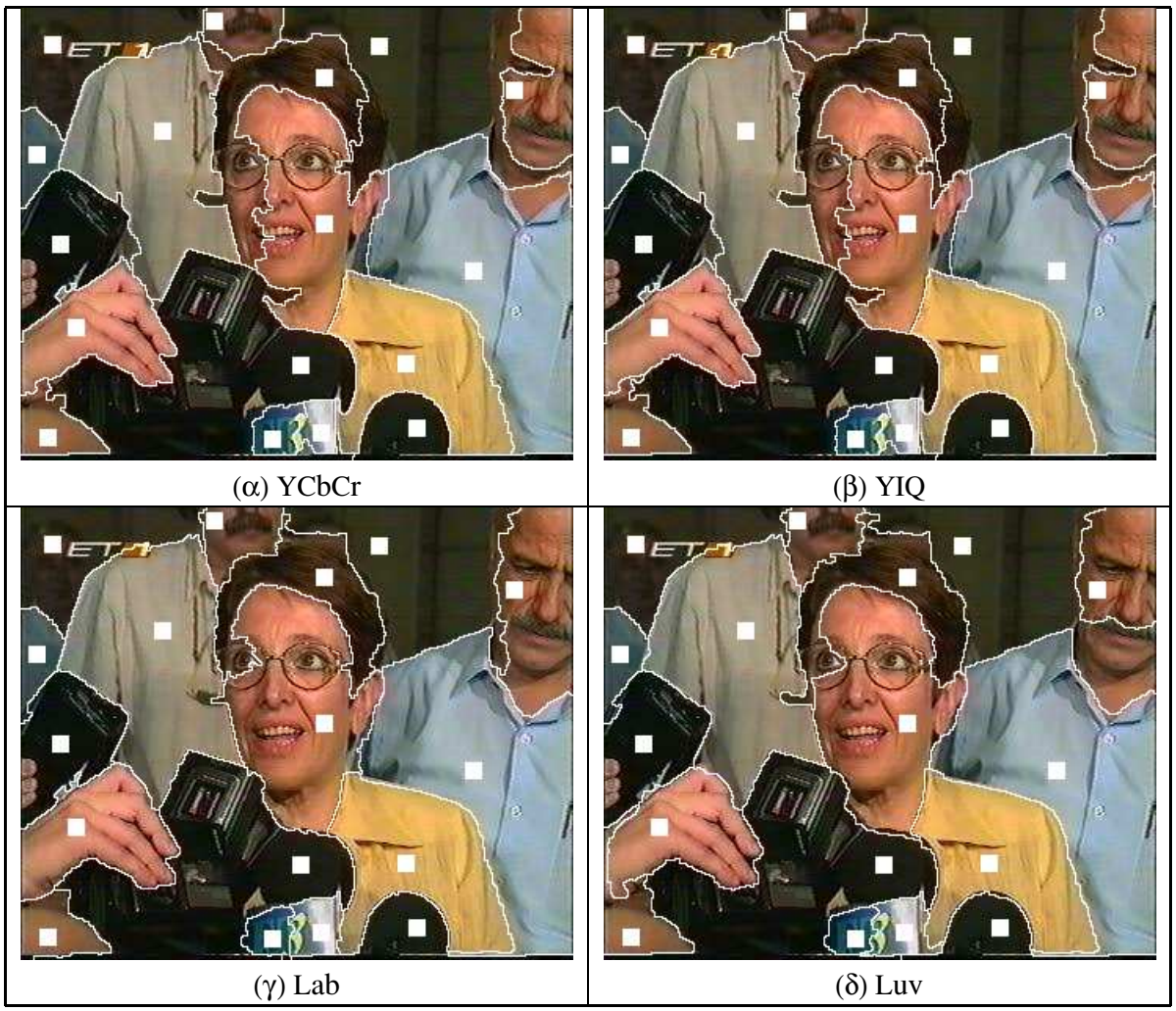
Στην ενότητα αυτή παρουσιάζεται το αποτέλεσμα της μεθόδου τμηματοποίησης σε μια εικόνα η οποία περιλαμβάνει περιοχές με διακυμάνσεις στη φωτεινότητά τους αλλά και έντονα χρώματα. Η εικόνα αυτή, όπως και οι εικόνες στις οποίες εφαρμόστηκε ο αυτόματος αλγόριθμος τμηματοποίησης, προέρχονται από τα video που όπως έχει αναφερθεί προσφέρθηκαν από την ΕΡΤ στα πλαίσια του προγράμματος DiVAN. Η εικόνα αυτή φαίνεται στο σχ. 7.3 και για ευκολία έχει ονομαστεί ΕΙΚΟΝΑ1. Για την εξαγωγή των αποτελεσμάτων, χρησιμοποιήθηκαν μηχανήματα ULTRA Sparc.



Σχήμα 7.2: Απόσταση διανύσματος χρώματος από ευθεία απόχρωσης, στο χώρο χρώματος YCbCr

Σχήμα 7.3: Εικόνα με αντικείμενα που παρουσιάζουν διακυμάνσεις στη φωτεινότητά τους και/ή έντονα χρώματα (ΕΙΚΟΝΑ1).

Στο σχ. 7.4 φαίνεται το αποτέλεσμα της τμηματοποίησης της εικόνας στους χώρους χρώματος YCbCr και YIQ, στις εικόνες (α) και (β) αντίστοιχα. Τα λευκά τετράγωνα στις εικόνες αυτές είναι οι ‘‘σπόροι’’-αρχικές περιοχές που έχουν δοθεί από το χρήστη, ενώ η λευκή περιφέρεια αντιστοιχεί στα σύνορα των συνόλων που έχουν προκύψει από την τμηματοποίηση. Εντελώς ανάλογα, στις εικόνες (γ) και (δ) του σχ. 7.4, φαίνονται τα αποτελέσματα της τμηματοποίησης για τους χώρους χρώματος Lab και Luv αντίστοιχα. Όπως προκύπτει από αυτά τα αποτελέσματα ο χώρος Lab παρουσιάζει τη μεγαλύτερη αντοχή στις επιδράσεις της φωτεινότητας και ακολουθείται ως προς αυτή την ιδιότητα, από το χώρο Luv.



(α) YCbCr (β) YIQ

(γ) Lab (δ) Luv

Σχήμα 7.4: Τμηματοποίηση μέσω του SRG για την ΕΙΚΟΝΑ1, στους χώρους χρωμάτων YCbCr (α) YIQ (b) Lab (γ) και Luv (δ)

Αντίθετα οι χώροι YCbCr και YIQ, σε αρκετές περιπτώσεις βασίζονται στη φωτεινότητα των αντικειμένων και όχι στο χρώμα τους, με αποτέλεσμα τόσο το χαρακτηριστικό των περιοχών όσο και οι αποστάσεις σημείων από σύνολα να επηρεάζονται από τη συνιστώσα της φωτεινότητας σε μεγάλο βαθμό. Έτσι, ο αλγόριθμος βρίσκει τα όρια

των περιοχών ως προς τη φωτεινότητα και όχι ως προς το χρώμα. Η εισαγωγή επιπλέον πληροφορίας του χρώματος όπως η διασπορά, δε δίνει σωστά αποτελέσματα τμηματοποίησης από τον SRG, αφού έχει παρατηρηθεί ότι είναι δυνατό να ''ευνοηθούν'' αρχικές περιοχές μεγάλης διασποράς έναντι άλλων και να επεκταθούν σε περιοχές που είναι εντελώς ανόμοιες ως προς το χρώμα.

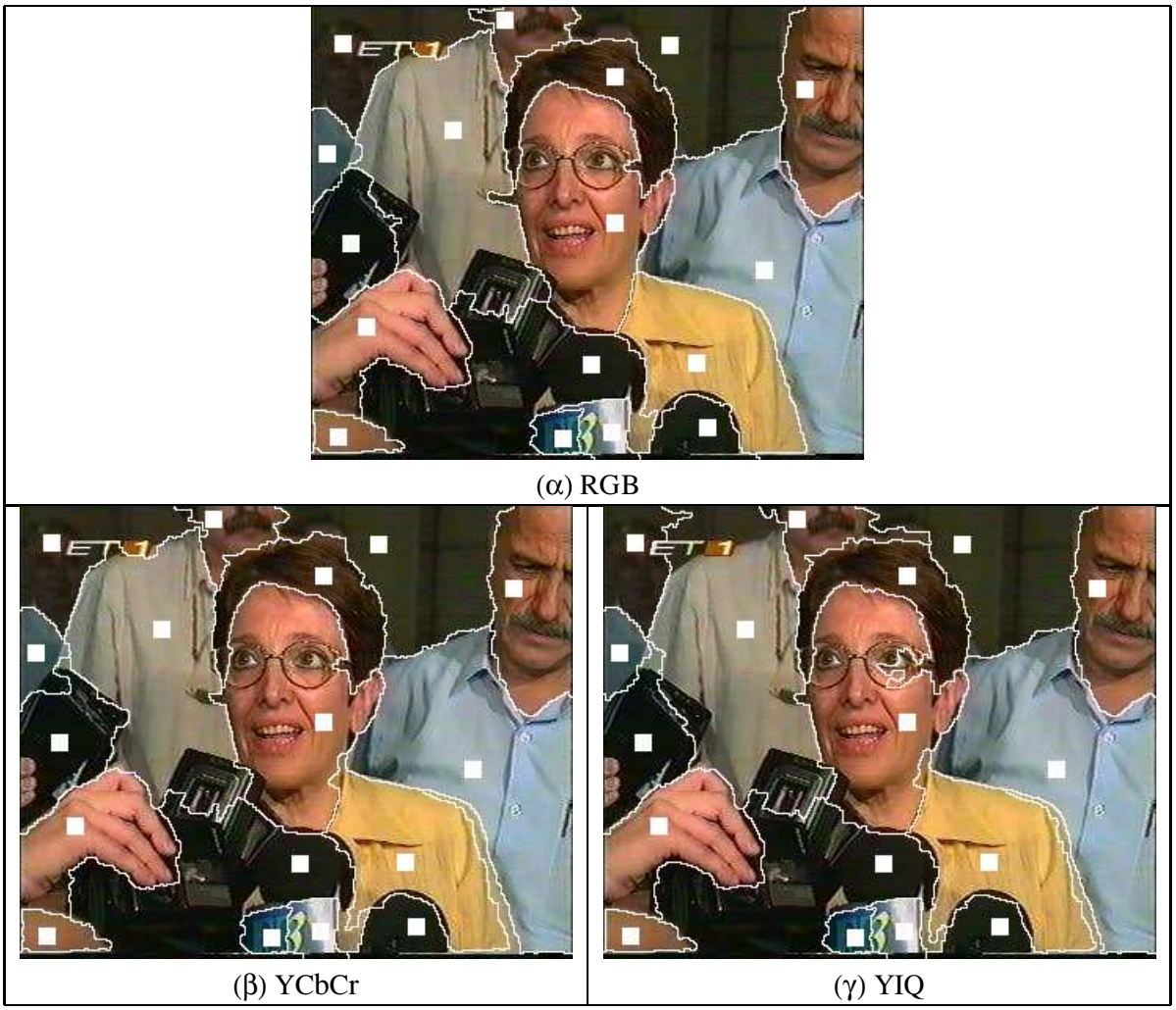
Εφαρμόζοντας στη συνέχεια την τμηματοποίηση βάσει του κριτηρίου απόστασης που δίνεται από τη σχέση (7.6) στην ίδια εικόνα και με τις ίδιες αρχικές περιοχές, το αποτέλεσμα είναι αυτό που φαίνεται στο σχ. 7.5. Στις εικόνες (α), (β) και (γ) φαίνεται το αποτέλεσμα που προκύπτει για τους χώρος RGB, YCbCr και YIQ αντίστοιχα. Τα συμπεράσματα είναι κοινά και για τους τρεις χώρους:

- Τα σύνορα μεταξύ περιοχών που διακρίνονται από το χρώμα τους, προσδιορίζονται με ακρίβεια. Η διατύπωση αυτή εκφράζει με διαφορετικό τρόπο το συμπέρασμα ότι η επίδραση της φωτεινότητας έχει εξαλειφθεί. Αυτό δίνει τη δυνατότητα επέκτασης χρωματικών περιοχών, χωρίς τον κίνδυνο τερματισμού της σε κάποιο τοπικό ελάχιστο. Το τοπικό ελάχιστο στην περίπτωση της τμηματοποίησης χρώματος από τον SRG, αντιστοιχεί σε ακμές φωτεινότητας στο εσωτερικό των περιοχών, όπως αυτές που περιέχονται στα πρόσωπα της ΕΙΚΟΝΑ1. Όπως φαίνεται από τα αποτελέσματα οι αρχικές χρωματικές περιοχές εύκολα ξεπερνούν πλέον αυτές τις ακμές και επεκτείνονται σταματώντας στα όρια διαφορετικών ως προς το χρώμα περιοχών.

- Αχρωμες αρχικές περιοχές (όπως οι περιοχές μαύρου χρώματος της εικόνας) μειονεκτούν ως προς την επέκτασή τους, αφού η απόχρωση των σημείων τους διασπείρεται γύρω από τον άξονα άσπρου-μαύρου και δεν είναι δυνατό να οριστεί η κλίση της ευθείας ϵ του σχ. 7.2, γι' αυτές τις περιοχές. Έτσι, δίνεται η δυνατότητα σε χρωματικές περιοχές να επεκταθούν εισάγοντας στο σύνολό τους σημεία των άχρωμων περιοχών. Κατ' αυτό τον τρόπο, η τμηματοποίηση αποτυγχάνει να καθορίσει τα όρια δύο άχρωμων περιοχών, ή μιας άχρωμης και μιας χρωματικής περιοχής. Εκτός από τη λύση που περιγράφεται στην προηγούμενη ενότητα για τις περιοχές αυτές, το αποτέλεσμα μπορεί να βελτιωθεί χρησιμοποιώντας την αλληλοδραστική διαδικασία επιπλέον τμηματοποίησης ή/και συγχώνευσης περιοχών.

Από τα παραπάνω γίνεται σαφές, ότι το κριτήριο δε δίνει σωστά αποτελέσματα, για εικόνες όπου η φωτεινότητα επικρατεί του χρώματος. Σ' αυτές τις περιπτώσεις, είναι προτιμότερο να επιλεγεί ένας από τους ομοιόμορφους χώρους χρώματος.

Τέλος, πρέπει να σημειωθεί ότι αν και κατά την επέκταση εξετάζονται σχεδόν όλα τα σημεία της εικόνας, ο χρόνος της τμηματοποίησης είναι περίπου 3sec.



(α) RGB

(β) YCbCr

(γ) YIQ

Σχήμα 7.5: Τμηματοποίηση βάσει της αναλογίας χρωμάτων στο RGB (α), στο YCbCr (β) και στο YIQ (γ)

7.3 Εφαρμογή: Αυτόματη Εξαγωγή Συνεκτικών Περιοχών Χρώματος

7.3.1 Περιγραφή

Περιγράφεται στη συνέχεια ένα σύστημα αυτόματης τμηματοποίησης των εικόνων χρώματος, σε ομοιόμορφες περιοχές χρώματος. Όπως φαίνεται στο σχήμα 7.6, η σημαντική διαφορά σε σχέση με την προσέγγιση καθοδήγησης της τμηματοποίησης από το χρήστη έγκειται στην είσοδο των αρχικών συνόλων στον αλγόριθμο SRG που εδώ εξάγονται αυτόματα, μέσω μιας διαδικασίας που στηρίζεται στα κυρίαρχα χρώματα της εικόνας. Αφού καθοριστούν τα αρχικά σύνολα, δίνονται ως είσοδος στον SRG για την ολοκλήρωση της τμηματοποίησης της εικόνας. Η έξοδος του SRG είναι η τμηματοποιημένη εικόνα σε ξένες μεταξύ τους περιοχές. Στη συνέχεια εφαρμόζεται μια αυτόματη τεχνική συγχώνευσης των περιοχών που είναι όμοιες μεταξύ τους. Σ' αυτό το τελευταίο στάδιο, εμφανίζονται προβλήματα ως προς τη σύγκριση των συνόλων, λόγω διαφοράς στη φωτεινότητα περιοχών των αντικειμένων, όπως θα περιγραφεί στο αντίστοιχο υποκεφάλαιο.

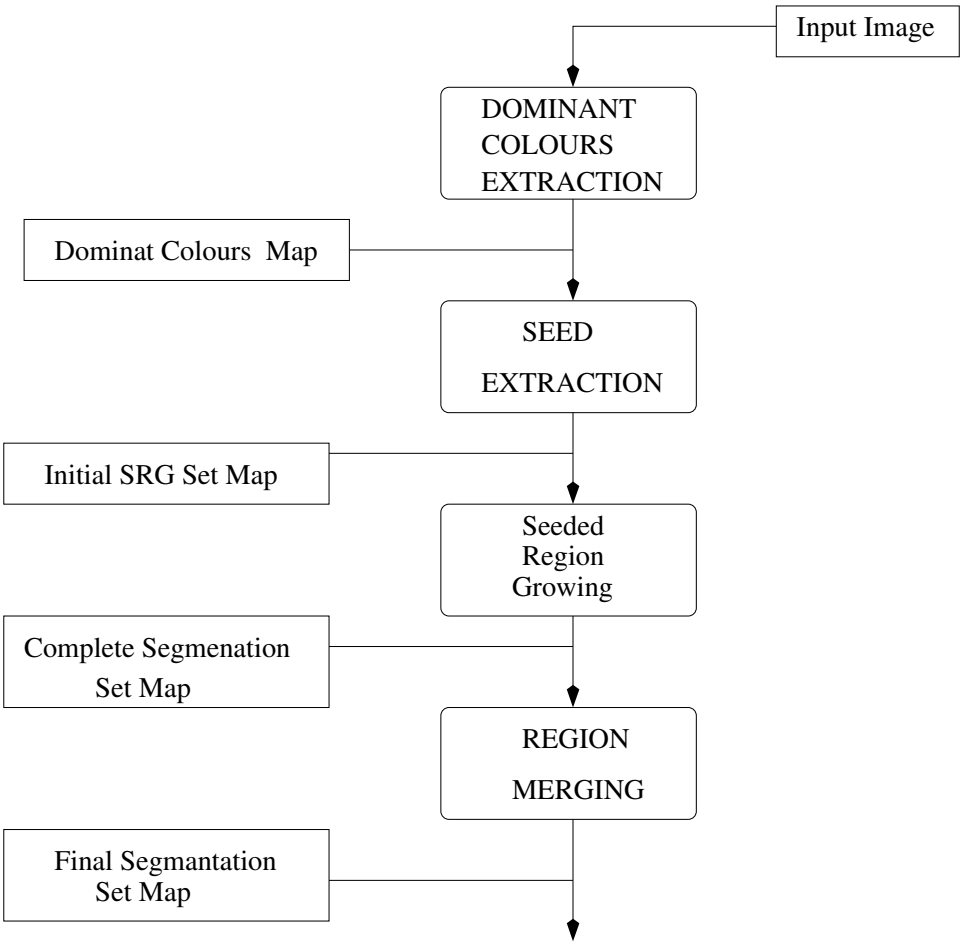
Για την τμηματοποίηση επιλέχθηκε ο χώρος χρωμάτων Lab λόγω της ομοιομορφίας που παρουσιάζει, γεγονός που επιτρέπει την αποδοτική χρήση της ευκλείδιας απόστασης των χρωμάτων. Ακολουθεί η περιγραφή των τμημάτων του συστήματος εξαγωγής ομοιόμορφων περιοχών της εικόνας.

7.3.2 Αυτόματη Εξαγωγή Αρχικών Περιοχών

Η εξαγωγή των αρχικών συνόλων από σπόρους που απαιτούνται για την εφαρμογή του SRG, βασίζεται στα κυρίαρχα χρώματα (dominant colors) της εικόνας βάσει της μεθόδου που περιγράφεται από το Σ. Λιάπη στο [17]. Στην αρχική υλοποίηση, τα χρώματα αυτά εξαγόταν στο χώρο χρώματος GLHS (Generalized Light Hue Saturation) [43]. Η χαρακτηριστική ιδιότητα αυτού του συστήματος, είναι ότι παρέχει μεγαλύτερη διάκριση των χρωμάτων χαμηλής φωτεινότητας στο χρωματικό κώνο του LHS, από ότι το ίδιο το LHS. Στη συνέχεια, η μέθοδος μετατράπηκε ώστε να εξάγονται τα κυρίαρχα χρώματα της εικόνας και στο χώρο Lab. Τα συμπεράσματα από τη χρήση των δύο συστημάτων δίνονται στο αντίστοιχο υποκεφάλαιο.

Αφού εξαχθούν τα κυρίαρχα χρώματα της εικόνας στο χώρο που έχει επιλεγεί, η διαδικασία εξαγωγής αρχικών σπόρων περιλαμβάνει τα εξής βήματα:

- Αρχικά, δημιουργείται ένας δυαδικός πίνακας b στις διαστάσεις της εικόνας ο οποίος ενημερώνεται ως εξής: για κάθε σημείο της εικόνας p, υπολογίζεται η απόσταση του από το κυρίαρχο χρώμα (DC_1, DC_2, DC_3) στο οποίο αντιστοιχεί



Σχήμα 7.6: Αυτόματη τμηματοποίηση εικόνων χρώματος

από τη σχέση:

$$\delta(p, DC) = \frac{1}{|W(p)|} \sum_{q \in W(p)} \sqrt{\sum_{l=1}^{3} (I_l(q) - DC_l)} \tag{7.7}$$

όπου $|W(p)|$ είναι το πλήθος των σημείων ενός παραθύρου $W(p)$ με κέντρο το σημείο p και $(I_1(p), I_2(p), I_3(p))$ είναι το διάνυσμα χρώματος του σημείου.

Εάν η απόσταση αυτή είναι μικρότερη από ένα κατώφλι t_{assign}, τότε τίθεται $b(p)$ =1, διαφορετικά $b(p)$ =0. Το αποτέλεσμα αυτής της διαδικασίας, είναι ότι προκύπτουν ομοιόμορφες σε χρώμα περιοχές οι οποίες είναι ξένες μεταξύ τους.

- Στη συνέχεια διατρέχεται ο πίνακας b, έτσι ώστε να καταγραφούν οι M_{init} ξένες μεταξύ τους περιοχές που προκύπτουν από τα σύνολα των σημείων p με $b(p)$ =1, τα οποία είναι συνδεδεμένα μεταξύ τους.

- Στο τελευταίο βήμα αυτής της διαδικασίας, απορρίπτονται περιοχές με πλήθος σημείων μικρότερο από ένα κατώφλι t_{size} και τα M_{fin}=n σύνολα-περιοχές που απομένουν αποτελούν τα αρχικά σύνολα $R_1, R_2, \ldots, R_{M_{fin}}$ που δίνονται ως αρχική τμηματοποίηση στον SRG.

Σκοπός της αρχικοποίησης αυτής, είναι η εξαγωγή ομοιόμορφων περιοχών συνδεδεμένων μεταξύ τους σημείων που

- δε βρίσκονται κοντά σε ακμές,

- είναι αρκετά μεγάλες ώστε να είναι αξιόπιστη η εκτίμηση των παραμέτρων του χρώματός τους και

- είναι ομοιόμορφες

Το επόμενο βήμα είναι η ολοκλήρωση της τμηματοποίησης από τον SRG, όπως περιγράφηκε στο προηγούμενο υποκεφάλαιο, με είσοδο τα σύνολα που προκύπτουν και θεωρώντας το χώρο χρώματος Lab. Το αποτέλεσμα της τμηματοποίησης είναι ο πίνακας απεικόνισης $i(p)$ κάθε σημείου p, στο σύνολο που τελικά ανήκει.

Πριν παρουσιαστεί το επόμενο βήμα της συγχώνευσης περιοχών, πρέπει να τονιστεί ο σημαντικός ρόλος της αρχικοποίησης των περιοχών στην τμηματοποίηση, αλλά κυρίως στη διαδικασία συγχώνευσης των περιοχών. Η αρχικοποίηση αυτή βασίζεται στα κυρίαρχα χρώματα της εικόνας, τα οποία διαφέρουν ανάλογα με το ποιό σύστημα χρησιμοποιείται για την εξαγωγή τους. Η διαφορά αυτή έγκειται τόσο στην κατανομή των κυρίαρχων χρωμάτων στο χώρο της εικόνας, όσο και στη μεταξύ τους απόσταση, στο δοσμένο χώρο χρώματος.

7.3.3 Συγχώνευση Περιοχών

Το τελευταίο στάδιο της συνολικής προσέγγισης είναι η συγχώνευση περιοχών που προκύπτουν από την τμηματοποίηση, οι οποίες είναι όμοιες ως προς το χρώμα. Γενικά

μιλώντας, εάν ο σκοπός μας είναι η εξαγωγή αντικειμένων από τις περιοχές που έχουν προκύψει από την τμηματοποίηση, η διαδικασία αυτή είναι αρκετά σύνθετη, ιδιαίτερα σε περιπτώσεις που επιχειρείται η συγχώνευση περιοχών του ίδιου αντικειμένου, τα οποία παρουσιάζουν μεγάλες διακυμάνσεις ως προς τη φωτεινότητά τους, λόγω γεωμετρίας, αντανάκλασης ή σκιών. Χαρακτηριστικό παράδειγμα, αποτελούν οι περιοχές του προσώπου σε εικόνες που περιλαμβάνουν πρόσωπα. Η συγχώνευση περιοχών στις περιπτώσεις αυτές, γίνεται με βάση κάποια *a-priori* πληροφορία του χρώματος ή και άλλων χαρακτηριστικών που σχετίζονται με τις συγκεκριμένες εφαρμογές [26] [44], όπως για παράδειγμα είναι το σχήμα (το οποίο για τα πρόσωπα είναι δεδομένο).

Η μέθοδος συγχώνευσης που χρησιμοποιήθηκε εδώ, βασίζεται στο χρώμα των περιοχών στο σύστημα Lab και στην πληροφορία των συνόρων τους. Αρχικά, υπολογίζεται ο χάρτης γειτνίασης των περιοχών R_i $(1 \leq i \leq M_{fin})$, της τμηματοποίησης. Στη συνέχεια, για κάθε ζεύγος γειτονικών συνόλων R_i, R_j, ορίζεται ως απόσταση τους $\delta(R_i, R_j)$, η ευκλείδια απόσταση των μέσων τιμών των χρωμάτων τους. Το κάθε βήμα της μεθόδου συγχώνευσης περιλαμβάνει τη συγχώνευση των γειτονικών περιοχών R_i, R_j με την ελάχιστη απόσταση δ_{min}:

- εάν $\delta_{min} < t_{merge}$, όπου t_{merge} είναι ένα δοσμένο κατώφλι,

  ▷ τα δύο σύνολα ενώνονται, εφόσον ισχύει επιπλέον και η σχέση

  $$\frac{B_{ij}}{B_i + B_j} > t_{bound} \tag{7.8}$$

  όπου B_i,B_j είναι το μήκος του συνόρου των συνόλων R_i, R_j αντίστοιχα, B_{ij} είναι το μήκος του κοινού συνόρου τους και t_{bound} είναι ένα δοσμένο κατώφλι. Η σχέση (7.8) δεν επιτρέπει την ένωση περιοχών των οποίων το κοινό σύνορο είναι μικρό σε σχέση με το συνολικό μήκος τους. Επιπλέον εάν τεθεί t_{bound}=0, η απόφαση για τη συγχώνευση βασίζεται αποκλειστικά στο χρώμα των περιοχών.

  Στη συνέχεια για το νέο σύνολο που προκύπτει από την ένωση των R_i, R_j, και θεωρώντας ότι αντικαθιστά ένα από αυτά στον πίνακα απεικόνισης $i(p)$ των σημείων σε σύνολα, υπολογίζονται η μέση τιμή του χρώματος του και βάσει αυτής, η απόστασή του από όλους τους γείτονες του.

- αλλιώς ο αλγόριθμος τερματίζεται.

Το πρόβλημα που εμφανίζεται, θεωρώντας την ευκλείδια απόσταση και στις τρεις συνιστώσες του χρώματος, έχει αναφερθεί ήδη: προκύπτει, όπως θα δούμε στα αποτελέσματα, από τη μεγάλη διακύμανση στη φωτεινότητα περιοχών ακόμα και στο Lab που όπως είναι γνωστό είναι ένα ‘‘συμπαγές’’ σύστημα χρωμάτων. Παρόλα αυτά το σύστημα δίνει καλά αποτελέσματα σε εικόνες όπου οι περιοχές-αντικείμενα είναι ομοιόμορφες χρωματικά.

Για τις ανάγκες τις υλοποίησης, δοκιμάστηκαν επιπλέον κριτήρια ομοιότητας δύο περιοχών. Το πρώτο από αυτά, είναι η σύγκριση των γωνιών μεταξύ των διανυσμάτων μέσης τιμής σε κάθε συνιστώσα χρώματος δύο συνόλων. Το δεύτερο είναι το κριτήριο συγχώνευσης περιοχών που προτείνεται στο [38]. Πρόκειται για ένα στατιστικό κριτήριο Περιγραφής Ελαχίστου Μήκους (MDL). Σύμφωνα με αυτό, συγχωνεύονται οι δύο περιοχές που η ένωσή τους προκαλεί τη μεγαλύτερη μείωση μιας συνάρτησης κόστους. Το κόστος είναι ανάλογο του μήκους κωδικοποίησης των συνόρων μεταξύ των περιοχών και αντιστρόφως ανάλογο της εκ των υστέρων (a-posteriori) πιθανότητας εμφάνισης του κάθε συνόλου. Το χαρακτηριστικό ως προς το οποίο γίνεται η τμηματοποίηση είναι το μοναδιαίο διάνυσμα απόχρωσης στο χώρο RGB κάθε συνόλου, όπως περιγράφηκε στην ενότητα 7.2.2. Αφού υπολογιστεί αυτό το διάνυσμα, ως εκ των υστέρων πιθανότητα κάθε συνόλου, θεωρείται η δισδιάστατη κατανομή των διαφορών του χρώματος των σημείων από την ευθεία που ορίζει το μοναδιαίο διάνυσμα. Η κατανομή αυτή, όπως αναφέρεται στο άρθρο, στις περισσότερες των περιπτώσεων περιγράφεται από την κατανομή Gauss.

Τα κριτήρια αυτά (όπως και το απλό κριτήριο που προτείνεται στην ενότητα 7.2.2), αποτυγχάνουν να διακρίνουν περιοχές με διαφορές στη φωτεινότητα, ενώ στο δεύτερο από αυτά, το μοντέλο δεν περιγράφει σωστά μεγάλες μεταβολές στη φωτεινότητα αντικειμένων. Εκτός αυτού, η χρήση του αυξάνει το υπολογιστικό κόστος, χωρίς ουσιαστική βελτίωση των αποτελεσμάτων.

Στην επόμενη ενότητα, παρουσιάζονται τα αποτελέσματα από την εξαγωγή των κυρίαρχων χρωμάτων τόσο στο χώρο GLHS, όσο και στο χώρο Lab και συζητούνται κάποιες επεκτάσεις που μπορεί να έχει ο αλγόριθμος στη δεύτερη περίπτωση αλλά όχι στην πρώτη.

7.3.4 Πειραματικά Αποτελέσματα-Σύγκριση GLHS και Lab στην Εξαγωγή Αρχικών Περιοχών

Στη συζήτηση που ακολουθεί, δίνονται ορισμένα αποτελέσματα που έχουν προκύψει από την εφαρμογή του συστήματος σε εικόνες που περιέχουν πρόσωπα και έχουν ληφθεί υπό διαφορετικές συνθήκες φωτισμού και με διαφορετικό περιβάλλον γύρω από αυτά, σε αντίθεση με τις εικόνες πειραμάτων, στις οποίες ο φωτισμός είναι καλός και συνήθως ο υπολογισμός παραμέτρων του χρώματος είναι ακριβής και ακολουθεί κάποιο μοντέλο. Ο σκοπός ήταν να εξετασθεί η δυνατότητα αυτόματης ανίχνευσης των περιοχών χρώματος των προσώπων, μέσω ενός γενικότερου αλγορίθμου τμηματοποίησης των εικόνων. Μία από τις εικόνες αυτές είναι η ΕΙΚΟΝΑ1, η οποία χρησιμοποιήθηκε στα αποτελέσματα της καθοδηγούμενης από το χρήστη τμηματοποίησης. Καθώς οι εικόνες αυτές είναι εικόνες video, έχουν υποστεί διπλή συμπίεση, την πρώτη κατά την κωδικοποίηση των video στα οποία ανήκουν και τη δεύτερη στην αποσυμπίεση, αφού αποθηκεύονται σε μορφή JPEG. Από την εφαρμογή του συστήματος, εξάγονται κάποια συμπεράσματα για τις δυνατότητες διάκρισης αντικειμένων σε φυσικές εικόνες από το χρώμα τους ή με τη

βοήθεια κάποιων άλλων χαρακτηριστικών των περιοχών.

Όπως ειπώθηκε ήδη, η αρχικοποίηση μέσω των κυρίαρχων χρωμάτων παίζει σημαντικό ρόλο για τα επόμενα βήματα του αλγορίθμου. Σκοπός της είναι να δώσει στο επόμενο στάδιο του συστήματος που είναι η τμηματοποίηση μέσω του SRG, αρχικές υπό επέκταση περιοχές, που αποτελούν ομοιόμορφα και κατά το δυνατό μεγάλα ως προς το μέγεθος σύνολα σημείων. Τα σύνολα αυτά, με τον τρόπο που εξάγονται, μπορεί να πει κανείς ότι είναι η απεικόνιση του κάθε κυρίαρχου χρώματος στο χώρο της εικόνας, σε συνεκτικές, ομοιόμορφες περιοχές σημείων.

Ο αλγόριθμος εξαγωγής κυρίαρχων χρωμάτων, ομαδοποιεί τα σημεία σε ένα δοσμένο αριθμό κλάσεων βάσει του αλγορίθμου ομαδοποίησης k-means, θεωρώντας ως χαρακτηριστικό των κλάσεων τη μέση τιμή του χρώματός τους στις τρεις συνιστώσες του χώρου. Για την αρχικοποίηση, αρχικά θεωρήθηκε ο χώρος χρωμάτων GLHS, ο οποίος προκύπτει από το χώρο LHS και έχει σχεδιαστεί για την καλύτερη διάκριση των χρωμάτων χαμηλής φωτεινότητας που είναι τοποθετημένα κοντά στη βάση του χώρου LHS. Στην περίπτωση που ο σκοπός της εφαρμογής είναι η εξαγωγή των κυρίαρχων χρωμάτων για εφαρμογές όπως η ανάκτηση εικόνων, τα αποτελέσματα στο χώρο αυτό είναι ικανοποιητικά, όπως φαίνεται στην κάτω αριστερά εικόνα του σχ. 7.7 που δείχνει την εξαγωγή των κυρίαρχων χρωμάτων για την πάνω εικόνα του σχήματος (ΕΙΚΟΝΑ2) . Στην εικόνα αυτή, κάθε σημείο αναπαρίσταται με τη μέση τιμή του κυρίαρχου χρώματος στο οποίο αντιστοιχεί, όπως προκύπτει από την εφαρμογή του αλγορίθμου για 16 κλάσεις χρώματος. Ωστόσο, η εξαγωγή των κυρίαρχων χρωμάτων σ' αυτή την περίπτωση, δείχνει ότι το σύστημα βασίζεται περισσότερο στη συνιστώσα της φωτεινότητας και όχι σ' αυτές του χρώματος. Έτσι το απεικονιζόμενο πρόσωπο, διαιρείται σε τρεις χρωματικές περιοχές που αντιστοιχούν σε τρία διαφορετικά κυρίαρχα χρώματα, ταυτόχρονα όμως κάθε κυρίαρχο χρώμα αντιστοιχεί στις διακυμάνσεις της φωτεινότητας του προσώπου (συνιστώσα L) ενώ το μετρούμενο χρώμα τους (συνιστώσες H,S) είναι όμοιο.

Σε μια τέτοια περίπτωση, η εξαγωγή αρχικών προς επέκταση περιοχών, με τον τρόπο που περιγράφηκε πιο πάνω, οδηγεί συνήθως σε μεγάλο αριθμό μικρών περιοχών για το συγκεκριμένο αντικείμενο, που είναι ακριβώς αυτό που θα έπρεπε να αποφευχθεί για την εφαρμογή της τμηματοποίησης και εξαγωγής αντικειμένων. Οι λόγοι είναι οι εξής:

- είναι δύσκολο αυτές οι υπό-περιοχές να συγχωνευθούν στη συνέχεια σε μια ενιαία χρωματική περιοχή, όπως θα αναλυθεί πιο κάτω και
- οι μεγάλες περιοχές επιτρέπουν το σταθερό υπολογισμό των αρχικών παραμέτρων επέκτασης των περιοχών.

Γι' αυτό το λόγο, χρησιμοποιήθηκε στη συνέχεια ο χώρος Lab για την εξαγωγή των κυρίαρχων χρωμάτων. Όπως αναφέρθηκε στο κεφ. 2, ο χώρος χρώματος Lab παρουσιάζει ομοιομορφία ως προς την κατανομή των χρωμάτων, γεγονός που επιτρέπει τη χρήση της ευκλείδιας απόστασης για τη μέτρηση της διαφοράς τους. Από τη χρήση

Σχήμα 7.7: Εξαγωγή κυρίαρχων χρωμάτων της πάνω εικόνας (ΕΙΚΟΝΑ2): στο χώρο GLHS (κάτω αριστερά) και στο χώρο Lab (κάτω δεξιά)

του για την εξαγωγή των κυρίαρχων χρωμάτων για 16 χρωματικές κλάσεις όπως και πριν, προκύπτει ότι στο χώρο αυτό η επίδραση της φωτεινότητας ελαττώνεται αισθητά. Αυτή η διαπίστωση φαίνεται οπτικά στην κάτω δεξιά εικόνα του σχ. 7.7. Το πρόσωπο σ' αυτή την εικόνα αντιστοιχεί σε μια περιοχή σημείων, όπως συμβαίνει και για αρκετά από τα άλλα αντικείμενα της εικόνας. Το συμπέρασμα από αυτό το παράδειγμα, όπως και για το επόμενο που ακολουθεί, είναι ότι μεγάλες τιμές φωτεινότητας των περιοχών, μετρούνται και ομαδοποιούνται στο Lab ως (σχετικά) μεγάλος κορεσμός (saturation) με όμοια απόχρωση (hue). Μια παρατήρηση που πρέπει να γίνει σε σχέση με την εξαγωγή των αρχικών περιοχών, είναι ότι σκοτεινές περιοχές όπως αυτές της πάνω εικόνας του σχ. 7.7, διακρίνονται μεταξύ τους, λόγω της διαφοράς στις χρωματικές συνιστώσες τους και όχι στη φωτεινότητα, στο σύστημα Lab. Είναι γεγονός ότι για τιμές φωτεινότητας κοντά στη βάση του συστήματος, στο χώρο αυτό, τα χρώματα εξακολουθούν να διαφέρουν ως προς την απόχρωσή τους κυρίως αλλά και ως προς τον κορεσμό (saturation) με αποτέλεσμα να μην ικανοποιείται η ιδιότητα της συμπύκνωσης (βλ. κεφ 2): τα χρώματα διακρίνονται από το σύστημα ενώ αντίθετα δεν είναι διακριτά από το ανθρώπινο μάτι. Παρόλα αυτά, οι πραγματικά μεγάλες αποστάσεις μεταξύ περιοχών είναι όπως θα φανεί πιο κάτω οι αποστάσεις που οφείλονται στη φωτεινότητα και όχι στο χρώμα των αντικειμένων.

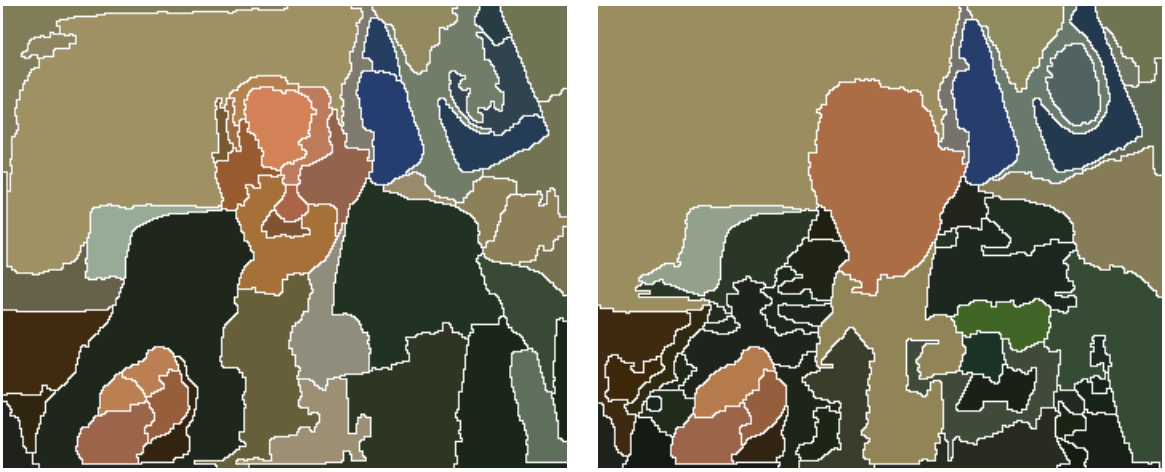
Τα βήματα που ακολουθούν την εξαγωγή των κυρίαρχων χρωμάτων, εφαρμόζονται στο σύστημα Lab. Οι αρχικές περιοχές που προκύπτουν για κάθε ένα από τα συστήματα αυτά, για την πάνω εικόνα του σχ. 7.7, φαίνονται στο σχ. 7.8. Τα σημεία λευκού χρώματος είναι αυτά τα οποία δεν αντιστοιχούν σε κάποια περιοχή μετά το τέλος της διαδικασίας αρχικοποίησης. Το παράθυρο W που χρησιμοποιήθηκε είναι διαστάσεων 11×11 και το κατώφλι t_{assign} τέθηκε στο 10.

Στο σχ. 7.8, φαίνεται ότι η αρχικοποίηση μέσω του GLHS (πάνω εικόνα) περιλαμβάνει αρκετές ανεξάρτητες περιοχές του προσώπου, ενώ αντίθετα, στο χώρο Lab (κάτω εικόνα) υπάρχει μία μόνο περιοχή.

Η τμηματοποίηση που προκύπτει μέσω του SRG, φαίνεται στο σχήμα 7.9, για την αρχικοποίηση τόσο στο χώρο GLHS (αριστερή εικόνα), όσο και στο χώρο Lab (δεξιά εικόνα). Η άσπρη περιφέρεια αντιστοιχεί στο σύνορο μεταξύ των περιοχών, ενώ κάθε περιοχή καλύπτεται από τη μέση τιμή του χρώματός της.

Το επόμενο βήμα, είναι η συγχώνευση περιοχών που έχουν προκύψει από τα δύο αποτελέσματα. Εδώ φαίνεται η διαφορά μεταξύ των δύο χώρων. Το σχήμα 7.10 δείχνει το αποτέλεσμα της συγχώνευσης για την αρχικοποίηση μέσω του GLHS (αριστερή εικόνα) και του Lab (δεξιά εικόνα). Και στις δύο περιπτώσεις, το κατώφλι t_{merge} τέθηκε ίσο με 10, ενώ δε χρησιμοποιήθηκε η πληροφορία του συνόρου (t_{bound}=0). Όπως φαίνεται στην αριστερή εικόνα του σχ. 7.10, η μέθοδος συγχώνευσης δεν ενώνει τις περιοχές του προσώπου και γενικότερα περιοχές αντικειμένων με μεγάλες μεταβολές φωτεινότητας. Οι περιοχές αυτές προέκυψαν, όπως ειπώθηκε ήδη, από τη διαδικασία εξαγωγής κυρίαρχων χρωμάτων μέσω του GLHS και γενικά, εφόσον προκύψουν, είναι δύσκολο να ενωθούν σε μια ενιαία περιοχή εκτός εάν το κατώφλι t_{merge} τεθεί σε μια

Σχήμα 7.8: Αρχικές περιοχές βάσει των κυρίαρχων χρωμάτων του GLHS (πάνω εικόνα) και του Lab (κάτω εικόνα).



Σχήμα 7.9: Επέκταση περιοχών από τον SRG μέσω της αρχικοποίησης στο GLHS (αριστερή εικόνα) και του Lab (δεξιά εικόνα).

Σχήμα 7.10: Συγχώνευση περιοχών που έχουν προκύψει από την αρχικοποίηση στο GLHS (αριστερή εικόνα) και στο Lab (δεξιά εικόνα)

μεγάλη τιμή, με κίνδυνο συγχώνευσης ανόμοιων περιοχών. Αντίθετα στη δεξιά εικόνα του ίδιου σχήματος, βλέπουμε ότι η αρχικοποίηση μέσω του Lab δεν παρουσιάζει ιδιαίτερα προβλήματα στην ένωση των σκοτεινών περιοχών που έχουν προκύψει, ενώ δεν τίθεται καν θέμα συγχώνευσης περιοχών με μεγάλες διακυμάνσεις στη φωτεινότητα, αφού εξ αρχής ομαδοποιούνται σε μια ενιαία περιοχή. Μ' αυτό τον τρόπο, αποφεύγονται λάθος συγχωνεύσεις που οφείλονται σε μεγάλες τιμές του κατωφλιού t_{merge}.

Παρόλα αυτά και στο Lab είναι δυνατό να εμφανιστούν κυρίαρχα χρώματα που αντιστοιχούν σε σκιάσεις ή μεταβολές στη φωτεινότητα των αντικειμένων. Το δεύτερο παράδειγμα, αναφέρεται στην εικόνα του σχ. 7.3 (ΕΙΚΟΝΑ1). Στις δύο εικόνες του σχ. 7.11 φαίνονται τα κυρίαρχα χρώματα που έχουν προκύψει, στα συστήματα GLHS και Lab αντίστοιχα. Μετά την τμηματοποίηση από τον SRG το αποτέλεσμα είναι αυτό της πάνω εικόνας του σχ. 7.13. Για την ένωση των περιοχών του προσώπου απαιτείται να τεθεί

t_{merge}=15 (κάτω εικόνα του σχ. 7.13). Αυτή η διακύμανση στην τιμή του κατωφλιού είναι ένας από τους κύριους λόγους για τους οποίους η μέθοδος δεν μπορεί να είναι πλήρως αυτόματη, ανεξάρτητα από την αρχικοποίηση. Το σημαντικό όμως στο Lab είναι, όπως φαίνεται στη δεξιά εικόνα του σχ. 7.12, ότι οι ανεξάρτητες περιοχές του προσώπου που προκύπτουν από την αρχικοποίηση, ''ανήκουν'' όλες στο ίδιο κυρίαρχο χρώμα και επιπλέον έχουν εξαλειφθεί περιοχές που ανήκουν σε σκιές. Επομένως, μια επέκταση του αλγορίθμου θα μπορούσε να είναι η εξής: εάν μετά το τέλος της τμηματοποίησης υπάρχουν περιοχές οι οποίες

- ξεκίνησαν να επεκτείνονται ως εκπρόσωποι στο χώρο της εικόνας του ίδιου κυρίαρχου χρώματος και
- στο τέλος της τμηματοποίησης είναι ''γείτονες''

τότε αυτές να ενώνονται, υπό ορισμένες συνθήκες. Μ' αυτό τον τρόπο, κάθε αρχική περιοχή αφήνεται να επεκταθεί με βάση τα τοπικά χαρακτηριστικά της περιοχής από τον SRG, ενώ στο τέλος συγχωνεύεται όπως περιγράφηκε, θεωρώντας ότι οι συνδεδεμένες πλέον περιοχές που προκύπτουν αποτελούν υποσύνολα μιας ευρύτερης περιοχής ενιαίου χρώματος. Η παραπάνω μέθοδος τμηματοποιεί σωστά, την εικόνα του σχ. 7.3 για την αρχικοποίηση στο Lab, θέτοντας t_{merge}=7.5 για γειτονικές περιοχές που δεν ανήκαν αρχικά στο ίδιο κυρίαρχο χρώμα. Η τμηματοποίηση στο GLHS δε δίνει τέτοιες δυνατότητες αφού τα αρχικά σύνολα αντιστοιχούν σε διαφορετικά κυρίαρχα χρώματα στην επιφάνεια του προσώπου. Επιπλέον, σε μια μελλοντική επέκταση του αλγορίθμου, θα μπορούσαν να τεθούν περιορισμοί στην ένωση αυτών των περιοχών, όπως το κριτήριο του συνόρου που περιγράφηκε ήδη ή η εξής απλή συνθήκη: μια περιοχή A, συγχωνεύεται με γειτονικές της που έχουν προκύψει από το ίδιο κυρίαρχο χρώμα, εκτός εάν η ελάχιστη απόσταση $\delta_{min}(A, B)$ προκύπτει για μια γειτονική της περιοχή B, η οποία έχει προκύψει

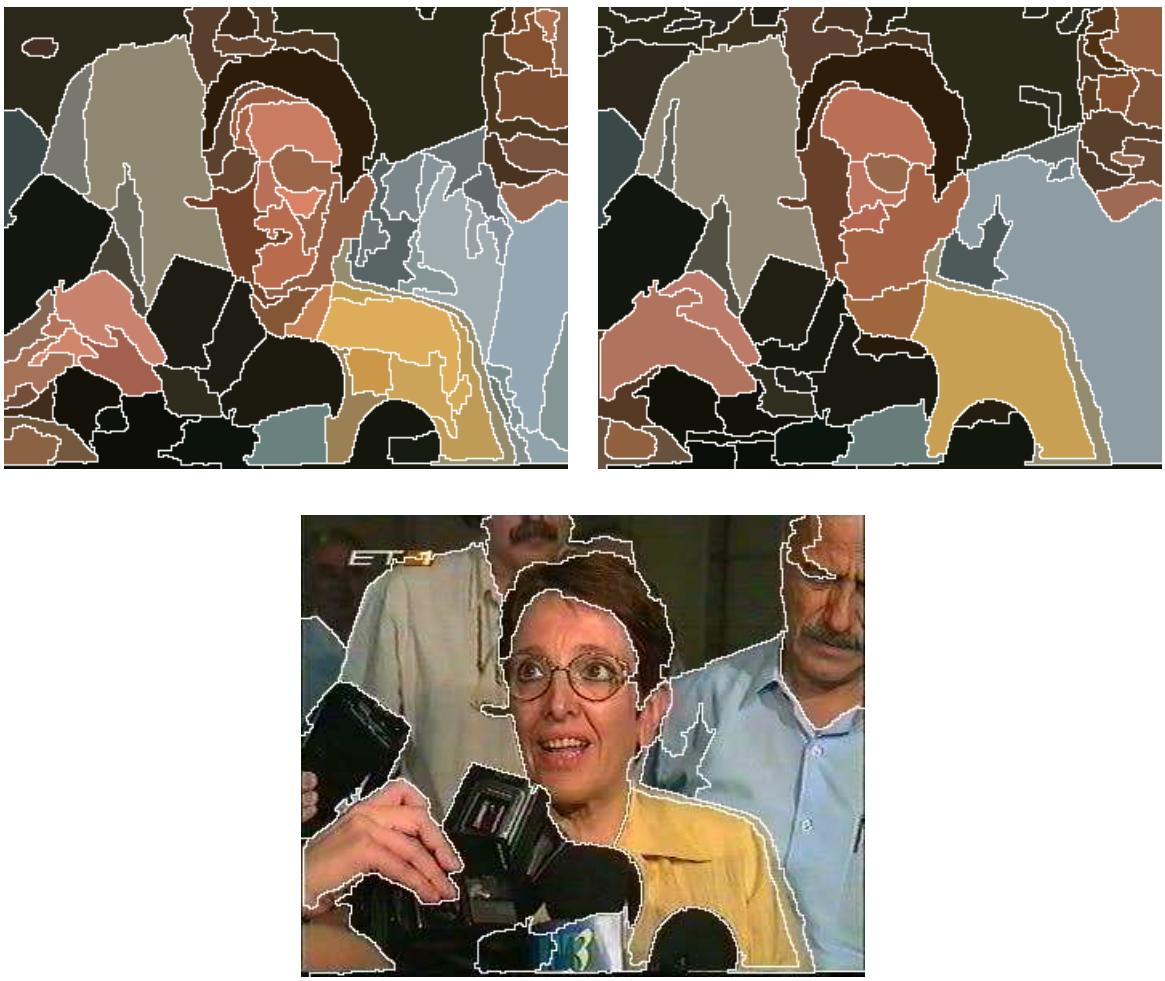
- είτε από διαφορετικό κυρίαρχο χρώμα,
- ή συγχώνευση περιοχών που προήλθαν από διαφορετικά κυρίαρχα χρώματα.

Σ' αυτή την περίπτωση, το σύνολο A, συγχωνεύεται με την περιοχή B, εάν βεβαίως ισχύει ότι $\delta_{min}(A, B) < t_{merge}$.

Η μεγάλη τιμή στην οποία πρέπει να τεθεί το κατώφλι συγχώνευσης γι' αυτή την εικόνα, οφείλεται στο γεγονός ότι τα τμήματα του προσώπου που προκύπτουν μετά την τμηματοποίηση, παρουσιάζουν μεγάλες διαφορές στο χρώμα και τη φωτεινότητά τους. Παρόλα αυτά, όπως φαίνεται στο σχ. 7.13 στο Lab μόνο το αριστερό μέρος του προσώπου της γυναίκας διαφέρει σημαντικά από τα άλλα τμήματα τα οποία συγχωνεύονται με t_{merge}=7.5, σε αντίθεση με το χώρο GLHS στον οποίο το πρόσωπο έχει κατακερματιστεί σε υποτμήματα φωτεινότητας και διαφορετικού κορεσμού. Αντίθετα υπάρχουν εικόνες στις οποίες οι διαφορές των αντικειμένων τόσο στο χρώμα όσο και στη φωτεινότητα είναι μικρές και αν τεθεί $t_{merge} > 5$, οι περιοχές που όντως αντιστοιχούν

Σχήμα 7.11: Εξαγωγή κυρίαρχων χρωμάτων για την ΕΙΚΟΝΑ1 στο χώρο GLHS (αριστερή εικόνα) και στο χώρο Lab (δεξιά εικόνα).

Σχήμα 7.12: Εξαγωγή αρχικών περιοχών για την ΕΙΚΟΝΑ1 στο χώρο GLHS (αριστερά) και στο χώρο Lab (δεξιά).



Σχήμα 7.13: Τμηματοποίηση στους χώρους GLHS (πάνω αριστερά) και Lab (πάνω δεξιά) και συγχώνευση (κάτω εικόνα) στο χώρο Lab για την ΕΙΚΟΝΑ1

σε αντικείμενα συγχωνεύονται. Η αντικατάσταση της απόστασης από άλλα κριτήρια δε λύνει το πρόβλημα στη γενική περίπτωση, αφού συνήθως πρέπει να γίνουν υποθέσεις για την κατανομή της πληροφορίας χρώματος και φωτεινότητας, οι οποίες δεν ισχύουν γενικά. Το σημαντικό πρόβλημα είναι ότι ενώ οι περιοχές αυτές είναι όμοιες ως προς την απόχρωση, όταν εξετάζονται ανεξάρτητα η μία από την άλλη (και καθώς συγχωνεύονται) διαφέρουν σημαντικά ως προς τη φωτεινότητα αλλά και ως προς τον κορεσμό, ανεξαρτήτως συστήματος χρώματος. Όπως θα δούμε στα συμπεράσματα, η εξέταση της ομοιότητας των περιοχών μόνο ως προς την απόχρωση ως μιας πιθανής λύσης του προβλήματος, πρέπει να γίνεται με προσοχή, ώστε να μην επιτραπεί σε εντελώς ανόμοιες περιοχές να ενωθούν.

Το ζήτημα που τίθεται στη συνέχεια, είναι σε ποιές περιπτώσεις αποτυγχάνει η εξαγωγή των κυρίαρχων χρωμάτων από το Lab. Για το Lab, ο k-means αλγόριθμος αργεί να συγκλίνει, λόγω της δυσκολίας εύρεσης των ορίων μεταξύ των κυρίαρχων χρωμάτων μικρής φωτεινότητας όπως οι σκοτεινές περιοχές στην ΕΙΚΟΝΑ2. Για τα πειραματικά αποτελέσματα χρησιμοποιήθηκαν 20 επαναλήψεις, οι οποίες για ορισμένες εικόνες δεν είναι αρκετές. Σε πολλές περιπτώσεις απαιτούνται έως 80 επαναλήψεις πριν τη σύγκλιση (βλ. [17]). Δυστυχώς, η εφαρμογή του αλγορίθμου εξαγωγής των κυρίαρχων χρωμάτων για ένα τόσο μεγάλο αριθμό επαναλήψεων επιβαρύνει σημαντικά το χρόνο απόκρισης του συστήματος. Πάντως η επίδραση της φωτεινότητας μειώνεται ακόμα περισσότερο όταν ο αλγόριθμος έχει συγκλίνει. Αντίθετα, στο χώρο GLHS η επίδραση της φωτεινότητας συνεχίζει να υφίσταται μέχρι και τη σύγκλιση, ενώ ο αριθμός των επαναλήψεων ως τη σύγκλιση, όπως και για το Lab, εξαρτάται από την εικόνα. Πάντως, στις περισσότερες εικόνες, είναι της τάξης του αριθμού επαναλήψεων για το Lab. Το γενικότερο συμπέρασμα είναι ότι η αρχικοποίηση μέσω του GLHS, οδηγεί σε τμηματοποίηση που βασίζεται κυρίως στη φωτεινότητα, ακόμα και των χρωματικών περιοχών. Έτσι, σε περιπτώσεις όπως τα πρόσωπα, η τμηματοποίηση που δίνει ο SRG περιλαμβάνει υπο-περιοχές φωτεινότητας που είναι δύσκολο να ενωθούν. Το Lab είναι πιο ανθεκτικό στις επιδράσεις της φωτεινότητας και συνήθως δίνει μεγαλύτερες (και λιγότερες) αρχικές περιοχές τμηματοποίησης, αφού ομαλοποιεί τις διαφορές λόγω φωτεινότητας. Αυτό σημαίνει ότι απαιτούνται λιγότερες συγχωνεύσεις σ' αυτή την περίπτωση, από ότι στο GLHS. Αυτή η ιδιότητα του Lab, μπορεί να χρησιμοποιηθεί στην περίπτωση ειδικότερα ανίχνευσης προσώπων από τις εικόνες, από μεθόδους σαν αυτή που περιγράφεται στο [26].

Στο σχήμα 7.14, φαίνεται η χρήση του κατωφλιού του συνόρου στο να αποφεύγεται η ένωση περιοχών που το μήκος του κοινού τους συνόρου είναι μικρό, σε σχέση με το μήκος της καθεμιάς, ανεξάρτητα από το εάν οι περιοχές παρουσιάζουν ομοιομορφία ως προς το χρώμα τους ή όχι. Η τμηματοποίηση για την πάνω εικόνα του σχήματος (ΕΙΚΟΝΑ3), φαίνεται στη μεσαία εικόνα. Θέτοντας t_{merge}=10 και t_{bound}=0.05 το αποτέλεσμα της συγχώνευσης είναι αυτό της κάτω εικόνας του σχήματος 7.14. Σ' αυτή την εικόνα το σύνολο του προσώπου δε συγχωνεύθηκε με την περιοχή του καθίσματος πίσω του όπως

θα συνέβαινε εάν δεν είχε τεθεί το κριτήριο του συνόρου.

Ακολουθούν στη συνέχεια, τρία αποτελέσματα τμηματοποίησης εικόνων βάσει του αλγορίθμου που περιγράφηκε. Το περιβάλλον σ' αυτές τις εικόνες είναι εντελώς διαφορετικό, όπως και οι συνθήκες λήψης των προσώπων σ' αυτές. Οι εικόνες αναφέρονται στη συνέχεια ως ΕΙΚΟΝΑ4, ΕΙΚΟΝΑ5 και ΕΙΚΟΝΑ6 και φαίνονται στα σχήματα 7.15, 7.17 και 7.19 αντίστοιχα. Στα σχήματα7.16, 7.18 και 7.20 φαίνονται τα βασικά στάδια της τμηματοποίησης που αντιστοιχούν σε κάθε μία από τις εικόνες. Επιπλέον, στα σχήματα αυτά παρουσιάζονται τα αποτελέσματα βάσει της αρχικοποίησης στους χώρους GLHS (αριστερή στήλη εικόνων) και Lab (δεξιά στήλη εικόνων). Τα στάδια που παρουσιάζονται για κάθε χώρο είναι η εξαγωγή των κυρίαρχων χρωμάτων ((α) εικόνα κάθε στήλης), η τμηματοποίηση μετά την εφαρμογή του SRG ((β) εικόνα κάθε στήλης) και το αποτέλεσμα μετά τη συγχώνευση περιοχών ((γ) εικόνα σε κάθε στήλη). Ο αλγόριθμος εξαγωγής των χρωμάτων εφαρμόζεται ως τη σύγκλιση όπως περιγράφεται στο [17]. Μ' αυτό τον τρόπο φαίνονται οι αδυναμίες του συστήματος GLHS το οποίο, όπως φαίνεται στις αντίστοιχες εικόνες, προκαλεί τον κατακερματισμό της εικόνας σε μικρά τμήματα, που διαφέρουν ως προς τη φωτεινότητά τους. Αντίθετα στο χώρο Lab, οι περιοχές που προκύπτουν είναι κατεξοχήν χρωματικές περιοχές. Για την εξαγωγή των αρχικών περιοχών, το παράθυρο W είχε διαστάσεις 11×11, ενώ το κατώφλι t_{assign} τέθηκε στο 10 για όλες τις εικόνες.

Η ΕΙΚΟΝΑ4, περιλαμβάνει άχρωμες περιοχές, οι οποίες είναι δύσκολο να διαχωριστούν λόγω μικρής πληροφορίας των συνιστωσών a και b. Στις εικόνες 7.16 .1(α) και 7.16 .2(β), φαίνεται το αποτέλεσμα της τμηματοποίησης, βάσει της αρχικοποίησης στο GLHS και στο Lab. Τέλος στις εικόνες (γ) του ίδιου σχήματος, φαίνεται η συγχώνευση των περιοχών που προκύπτουν από κάθε τμηματοποίηση. Για τη συγχώνευση τέθηκε t_{merge}=7 (t_{bound}=0), αφού για μεγαλύτερη τιμή ενώνονται ξένες περιοχές. Η ευκολία στην ένωση αυτών των περιοχών οφείλεται στην έλλειψη πληροφορίας χρώματος.

Τα ίδια συμπεράσματα προκύπτουν και για την τμηματοποίηση της ΕΙΚΟΝΑ5: η αρχικοποίηση μέσω του GLHS, οδηγεί σε περιοχές που διαφέρουν ως προς τη φωτεινότητα. Μάλιστα γι' αυτή την εικόνα η τμηματοποίηση που προκύπτει για το πρόσωπο, αντιστοιχεί ακριβώς στις περιοχές της φωτεινότητας με αποτέλεσμα αυτές να μη συγχωνεύονται εφόσον τεθεί t_{merge}=7.5. Αυτή η τιμή χρησιμοποιήθηκε για τη συγχώνευση που φαίνεται στις εικόνες 7.18 .1(γ) και 7.18 .2(γ) για την αρχικοποίηση στους δύο χώρους (με t_{bound}=0).

Τα στάδια της τμηματοποίησης για την ΕΙΚΟΝΑ6 φαίνονται στο σχ. 7.20. Για τη συγχώνευση περιοχών τέθηκε t_{merge}=7 και t_{bound}=0.05, αφού οι περιοχές αυτές αντιστοιχούν σε φωτεινά χρώματα που βρίσκονται σε μικρή απόσταση στο χώρο Lab, ο οποίος χρησιμοποιείται για τα στάδια μετά την αρχικοποίηση. Η εικόνα περιλαμβάνει επίσης περιοχές χωρίς χρώμα. Πάντως το αποτέλεσμα και γι' αυτή την εικόνα είναι καλύτερο εφόσον η αρχικοποίηση γίνει στο χώρο Lab.

Ο χρόνος επεξεργασίας του συνολικού συστήματος αναλώνεται στο μεγαλύτερο

Σχήμα 7.14: Τμηματοποίηση (μεσαία εικόνα) και εφαρμογή του κριτηρίου συνόρου (κάτω εικόνα) για την πάνω εικόνα (ΕΙΚΟΝΑ3).

μέρος του από τις διαδικασίες αρχικοποίησης. Η εξαγωγή των κυρίαρχων χρωμάτων απαιτεί περίπου 2 sec/επανάληψη, ενώ η εξαγωγή των αρχικών περιοχών μπορεί να διαρκέσει έως και 3 min, για παράθυρο W διαστάσεων 11×11. Η εφαρμογή του SRG, απαιτεί γύρω στο 1 sec, ενώ ο χρόνος απόκρισης της συγχώνευσης εξαρτάται από τον αριθμό των περιοχών που ενώνονται (και άρα από το σύστημα χρώματος στο οποίο γίνεται η εξαγωγή των κυρίαρχων χρωμάτων), καθώς και από το αν χρησιμοποιείται ή όχι η πληροφορία του συνόρου. Συνήθως δεν ξεπερνάει τα 20sec.

7.3.5 Συμπεράσματα-Επεκτάσεις

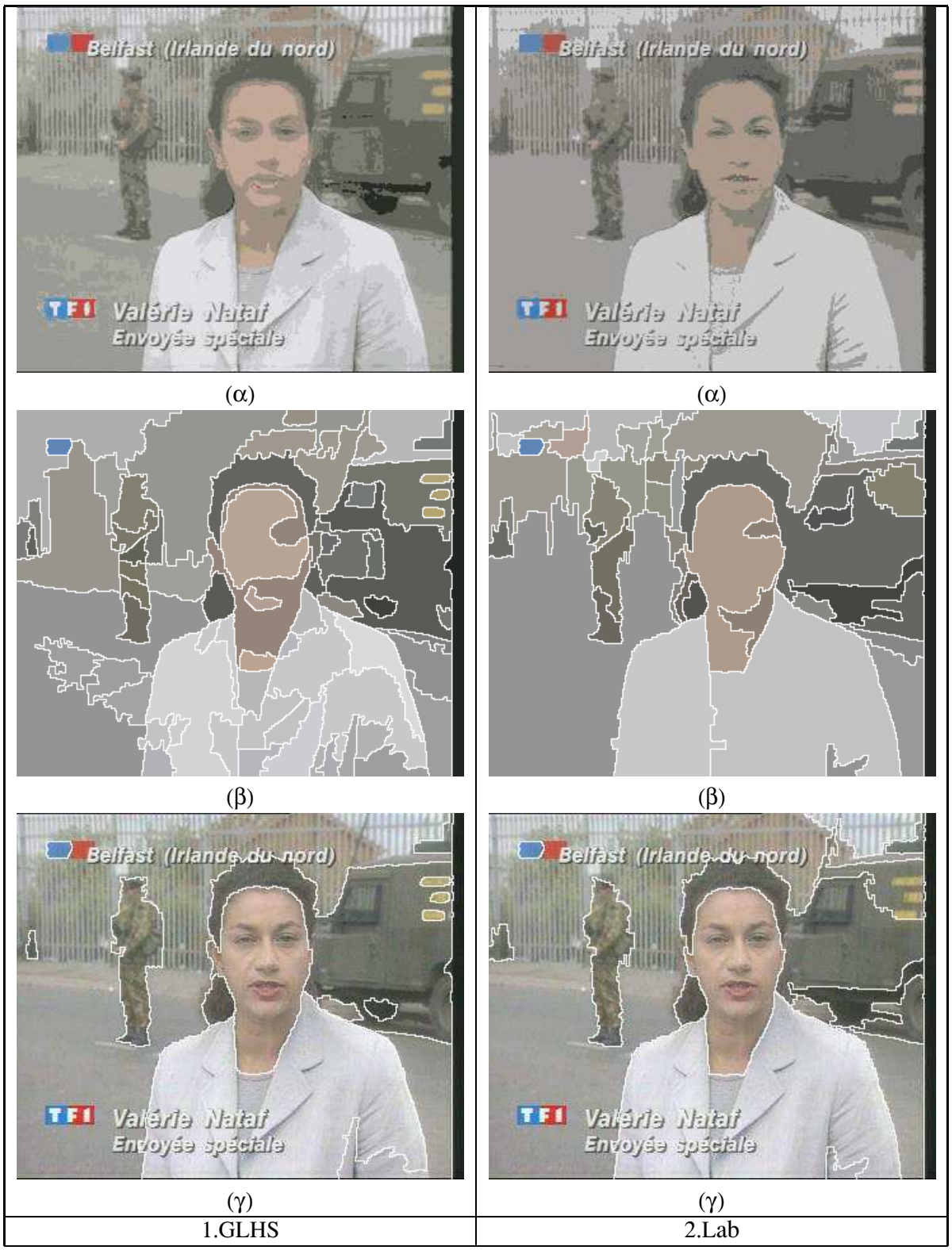
Ο αλγόριθμος που παρουσιάστηκε πιο πάνω, θα μπορούσε να επεκταθεί με τη χρήση επιπλέον χώρων χρώματος εφαρμογής του, μεθόδων αρχικοποίησης και κριτηρίων συγχώνευσης των περιοχών που προκύπτουν από την τμηματοποίηση μέσω του SRG.

Για την αρχικοποίηση, εκτός της πληροφορίας των κυρίαρχων χρωμάτων, μπορούν να χρησιμοποιηθούν γνωστές μέθοδοι εξαγωγής υποψηφίων περιοχών ''σπόρων'', όπως είναι το κριτήριο των ροπών και γνωστά στατιστικά κριτήρια σύγκρισης κατανομών, όπως το χ^2 και το Kolmogorov-Smirnov, τα οποία χρησιμοποιήθηκαν για τη σύγκριση των ιστογραμμάτων στα κεφάλαια τα σχετικά με τη δεικτοδότηση του video. Σκοπός των μεθόδων αυτών είναι να εξεταστεί το κατά πόσο κάθε σημείο της εικόνας, είναι κατάλληλο να ενταχθεί σε μια αρχική περιοχή του αλγορίθμου επέκτασης. Κατά την εφαρμογή των μεθόδων αυτών, αρχικά μετράται η πληροφορία του χαρακτηριστικού που προκύπτει από τη δειγματοληψία σε ένα παράθυρο με κέντρο το κάθε σημείο. Στη συνέχεια εξετάζεται αν η δειγματοληπτούμενη πληροφορία ακολουθεί μια δοσμένη κατανομή, η οποία στην απλούστερη περίπτωση είναι η κατανομή Gauss (βλ. [45]). Έτσι, αντί να εξετάζεται μόνο η ομοιομορφία μεταξύ ενός σημείου και της περιοχής του, ελέγχεται επιπλέον και το αν η πληροφορία γύρω από ένα σημείο ακολουθεί μια δοσμένη κατανομή.

Για τη συγχώνευση των περιοχών είναι δυνατό να χρησιμοποιηθούν επίσης επιπλέον κριτήρια. Η πρώτη κατηγορία κριτηρίων αφορά την πληροφορία που εξάγεται από τα σύνορα των περιοχών όπως είναι το μήκος τους το οποίο χρησιμοποιήθηκε εδώ. Γνωστές ευριστικές μέθοδοι που βασίζονται στην εύρεση των ''αδυνάτων'' σημείων του συνόρου, περιγράφονται στο [11]. Μια δεύτερη προσέγγιση, προκύπτει από τη μελέτη των περιοχών με μεγάλες μεταβολές φωτεινότητας, όπως είναι τα πρόσωπα. Αυτό που παρατηρείται σ' αυτές είναι ότι παρουσιάζουν ομοιομορφία στη χρωματικότητά τους και κυρίως στην απόχρωσή τους. Θα μπορούσε επομένως η συγχώνευση να βασίζεται στη σύγκριση ή των δισδιάστατων κατανομών στο χρωματικό επίπεδο, ή της κατανομής της απόχρωσης μόνο. Και στις δύο περιπτώσεις η σύγκριση μπορεί να γίνει με τα στατιστικά κριτήρια σύγκρισης κατανομών, που αναφέρθηκαν πιο πάνω. Μάλιστα, τα κριτήρια αυτά, εκτός από τη μέτρηση της ομοιότητας των ιστογραμμάτων, παρέχουν πληροφορία για την πιθανότητα με την οποία οι δύο περιοχές ακολουθούν την ίδια

κατανομή (στο [46] παρέχεται η υλοποίηση αυτών των μεθόδων για ιστογράμματα μονοδιάστατων και δισδιάστατων κατανομών). Η εξάλειψη πάντως της φωτεινότητας από το κριτήριο συγχώνευσης ενέχει τον κίνδυνο συγχώνευσης περιοχών που γίνονται αντιληπτές ως άχρωμες. Ο κίνδυνος αυτός είναι μεγαλύτερος στην περίπτωση του Lab λόγω της πυκνότητας της εμφάνισης των χρωμάτων των φυσικών εικόνων που περιγράφηκε στην προηγούμενη παράγραφο, και τη σημαντική μείωση της επίδρασης της φωτεινότητας. Έτσι, χρώματα τα οποία είναι τοποθετημένα σε εύρος φωτεινότητας 40 έως 55 βρίσκονται σε μικρή (ευκλείδια) απόσταση στο χώρο Lab, αν και γίνονται αντιληπτά ως διαφορετικά χρώματα. Σ' αυτή την περίπτωση απαιτείται καλή γνώση του χώρου χρώματος, ώστε να μη συγχωνεύονται περιοχές που με βάση το κριτήριο είναι όμοιες χρωματικά, αλλά στην πραγματικότητα, διαφέρουν ως προς τη φωτεινότητά τους.

Σχήμα 7.15: ΕΙΚΟΝΑ4

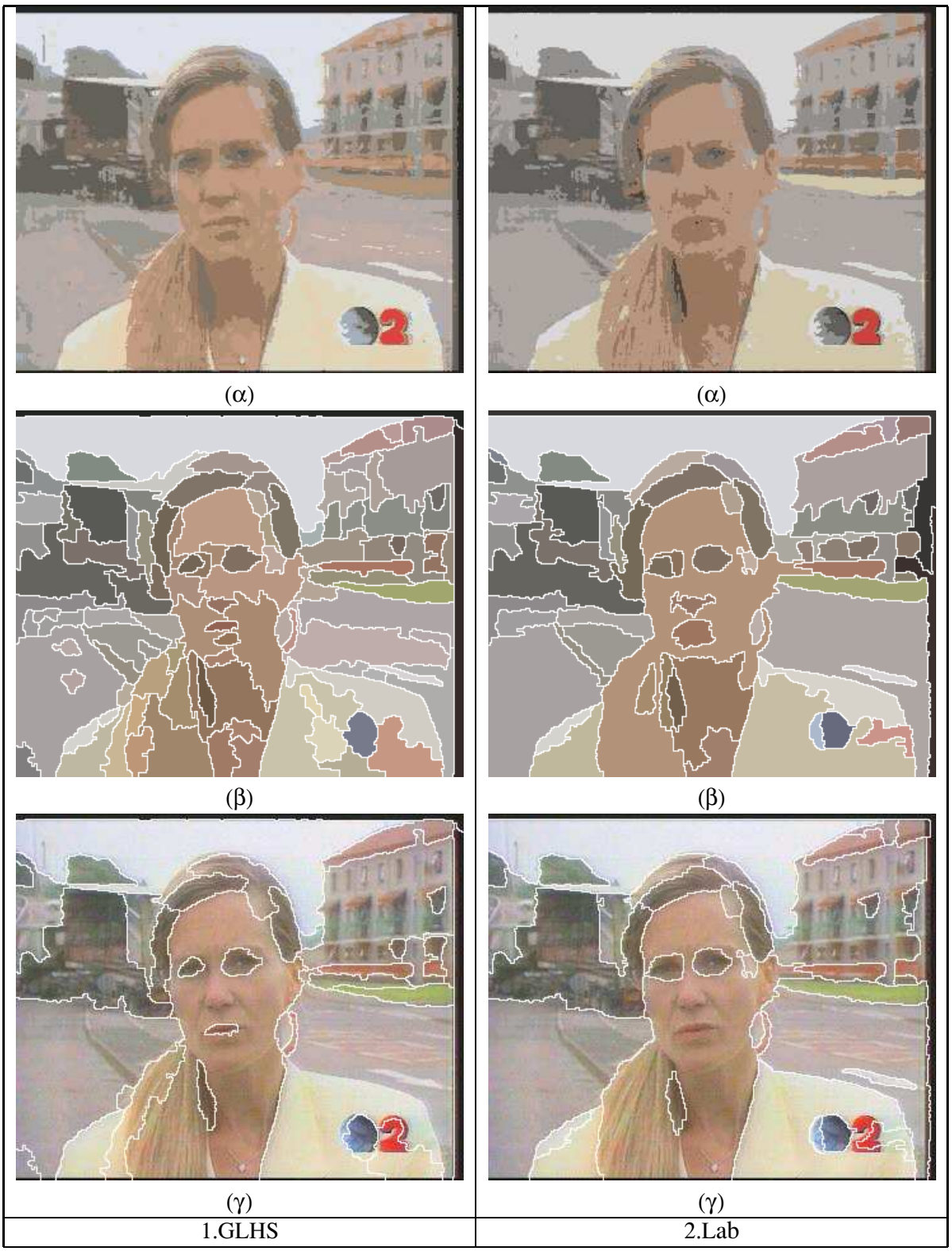


Σχήμα 7.16: Κυρίαρχα χρώματα (α), τμηματοποίηση (β) και συγχώνευση περιοχών (γ) για το GLHS (1) και για το Lab (2) για την ΕΙΚΟΝΑ4

Σχήμα 7.17: ΕΙΚΟΝΑ5

Σχήμα 7.18: Κυρίαρχα χρώματα (α), τμηματοποίηση (β) και συγχώνευση περιοχών (γ) για το GLHS (1) και για το Lab (2) για την ΕΙΚΟΝΑ5

Σχήμα 7.19: ΕΙΚΟΝΑ6



Σχήμα 7.20: Κυρίαρχα χρώματα (α), τμηματοποίηση (β) και συγχώνευση περιοχών (γ) για το GLHS (1) και για το Lab (2) για την ΕΙΚΟΝΑ6

Κεφάλαιο 8

Παρακολούθηση Αντικειμένων Video

8.1 Εισαγωγή-Ανασκόπηση

Η τμηματοποίηση κίνησης δισδιάστατων εικόνων (2-D motion segmentation) είναι ένα σημαντικό βήμα στην ανάλυση ακολουθιών εικόνων. Τα αποτελέσματά της χρησιμοποιούνται για τον προσδιορισμό των τρισδιάστατων χαρακτηριστικών των αντικειμένων της σκηνής καθώς και για την κωδικοποίηση που βασίζεται σε χαρακτηριστικά των αντικειμένων (feature-oriented coding) [47] [27].

Στη δεύτερη περίπτωση ειδικότερα, πρέπει να επισημανθεί ο σημαντικός ρόλος της τμηματοποίησης κίνησης στη μείωση του χρονικού πλεονασμού (temporal redundancy) που όντως παρουσιάζεται σε αρκετές περιπτώσεις στις σκηνές του video. Το γεγονός αυτό κάνει ιδανική την εφαρμογή της τμηματοποίησης κίνησης σε συστήματα δεικτοδότησης του video αφού επιτυγχάνεται η συνεχής περιγραφή του περιεχομένου των σκηνών σε όλη τη διάρκειά τους, μειώνοντας σημαντικά τις απαιτήσεις σε χώρο αποθήκευσης. Τα τελευταία χρόνια σημαντικές εφαρμογές της τμηματοποίησης κίνησης συναντώνται επίσης στη διαχείριση video βάσει αντικειμένων, η οποία αποσκοπεί στη δυνατότητα χρήσης των αντικειμένων μετά την εξαγωγή τους (post-production) όπως είδαμε, για τη δημιουργία ειδικών εφέ. Με την είσοδο της πληροφορίας κίνησης των αντικειμένων στη δημιουργία νέων video (video editing), τα εφέ αυτά είναι ιδιαίτερα εντυπωσιακά και έχουν αποκτήσει τεράστια σημασία και από οικονομικής απόψεως.

Στο παρελθόν, έχουν προταθεί αρκετές τεχνικές στη βιβλιογραφία για την τμηματοποίηση του δισδιάστατου πεδίου κίνησης, των οποίων οι στόχοι είναι τόσο η σωστή εκτίμηση της κίνησης των αντικειμένων, όσο και η ακρίβεια στην τμηματοποίηση που προκύπτει. Μια πρόσφατη ανασκόπηση αυτών των τεχνικών αποτελεί το [48]. Η μέτρηση των παραμέτρων κίνησης στο δισδιάστατο πεδίο της εικόνας, βασίζεται στις χωροχρονικές μεταβολές (spatio-temporal variations) της έντασης των σημείων της, οι οποίες παρέχουν τη μόνη δυνατή πηγή μέτρησης. Βάσει αυτών των μεταβολών υπολογίζεται ένα διανυσματικό πεδίο ταχυτήτων σε κάθε σημείο της εικόνας, το οποίο ονομάζεται *οπτική ροή* (optical flow) και αποτελεί μέτρηση της φαινόμενης κίνησης (apparent motion), σε

αντιδιαστολή με την πραγματική κίνηση (real motion), η οποία δεν είναι πάντα δυνατό να προσδιοριστεί από αυτές τις μετρήσεις. Οι μεταβολές αυτές εκτός από την κίνηση των αντικειμένων, εξαρτώνται και από παράγοντες όπως τα χαρακτηριστικά της τρισδιάστατης σκηνής που επηρεάζουν το σχήμα και τις φωτομετρικές ιδιότητες της επιφάνειάς τους.

Είναι γνωστό [47], ότι για να υπολογιστούν οι παράμετροι της ταχύτητας, πρέπει να έχουν καθοριστεί τα όρια του αντικειμένου. Από την άλλη πλευρά, η γνώση των παραμέτρων ταχύτητας των αντικειμένων, αποτελεί σημαντικό χαρακτηριστικό τμηματοποίησης της εικόνας σε αντικείμενα και άρα της εύρεσης των ορίων τους. Αυτή η ισχυρή εξάρτηση παραμέτρων κίνησης και τμηματοποίησης χαρακτηρίζεται απόλυτα από το γνωστό πρόβλημα ''της κότας και του αυγού''. Οι μέθοδοι τμηματοποίησης κίνησης επομένως, αποσκοπούν στο ''σπάσιμο'' αυτής της αλληλεξάρτησης. Στην κατεύθυνση αυτή, γίνεται χρήση της *χωρικής συνάφειας* (spatial coherence) που χαρακτηρίζει το πεδίο ταχυτήτων η οποία δίνει την επιπλέον πληροφορία που απαιτείται για την τμηματοποίηση. Σύμφωνα μ' αυτή την αρχή, η υπολογιζόμενη οπτική ροή για κάθε αντικείμενο, είναι μια συνεχής διανυσματική συνάρτηση ως προς τις συντεταγμένες τις εικόνας. Ο υπολογισμός του δισδιάστατου διανύσματος οπτικής ροής (u, v) σε κάθε σημείο δίνεται από τη γνωστή *εξίσωση οπτικής ροής*:

$$I_x(x, y; t)\ u + I_y(x, y; t)\ v + I_t(x, y; t) = 0 \tag{8.1}$$

όπου $I_x(x, y; t)$, $I_y(x, y; t)$ είναι οι μερικές χωρικές παράγωγοι της έντασης του σημείου (x, y) και $I_t(x, y; t)$ η χρονική παράγωγος του σημείου. Αυτό που πρέπει να γίνει επομένως είναι να καθοριστούν τα σημεία ασυνεχειών του πεδίου ταχυτήτων. Όμως συχνά τα σημεία αυτά, δίνουν την υφή, η οποία αποτελεί τη χρήσιμη πληροφορία υπολογισμού της κίνησης των αντικειμένων και όχι τα όρια τους στο χώρο.

Από την άλλη πλευρά, σε αρκετές εφαρμογές, γίνεται ανίχνευση των κινουμένων περιοχών, χωρίς την εκτίμηση των παραμέτρων της κίνησης. Σ' αυτή την κατηγορία ανήκει η μέθοδος που προτείνεται στο [39], στα πλαίσια των δραστηριοτήτων του ΙΤΕ. Η μέθοδος αυτή, περιλαμβάνει αρχικά την ανίχνευση αλλαγής (change detection) της έντασης των σημείων μεταξύ δύο διαδοχικών εικόνων. Στη συνέχεια, προσδιορίζονται τα ακριβή όρια των αντικειμένων, χρησιμοποιώντας τον αλγόριθμο των level sets [31] [32].

Στη μέθοδο που περιγράφεται πιο κάτω (βλ. και [49]), θεωρείται ότι κάθε αντικείμενο της εικόνας (συμπεριλαμβανομένου και του background) κινείται, εκτελώντας μια μεταφορική κίνηση στο επίπεδο, η οποία μπορεί να περιγραφεί από ένα μοντέλο των 2 συνιστωσών του διανύσματος της ταχύτητάς του. Η εκτίμηση των παραμέτρων βασίζεται στην υπόθεση που γίνεται ότι η φωτεινότητα (luminance) μιας περιοχής παραμένει σταθερή ενώ η θέση του μπορεί να αλλάζει στη διάρκεια του χρόνου. Αυτή είναι η γνωστή *αρχή διατήρησης των δεδομένων* (data conservation constraint). Έχοντας υπόψη αυτή την υπόθεση, οι παράμετροι της κίνησης εκτιμώνται μέσω μιας τεχνικής *region*

matching που αποτελεί επέκταση της τεχνικής *block matching* σε περιοχές οποιουδήποτε σχήματος.

Ο αλγόριθμος SRG, έχει τροποποιηθεί κατάλληλα για τη χρήση του στην τμηματοποίηση της κίνησης και χρησιμοποιείται για δύο σκοπούς: για την τμηματοποίηση του πεδίου ταχυτήτων δύο διαδοχικών στο χρόνο εικόνων και για την παρακολούθηση των αντικειμένων που έχουν ήδη εξαχθεί σε όλη τη διάρκεια της σκηνής. Στην πρώτη περίπτωση οι ''σπόροι'' των περιοχών δίνονται από το χρήστη, ενώ στη δεύτερη εισάγεται μια *ιεραρχική διάταξη* των αντικειμένων της σκηνής η οποία χρησιμοποιείται όπως θα εξηγηθεί παρακάτω για την αυτόματη εξαγωγή των αρχικών περιοχών. Πρέπει να τονιστεί ότι αυτή η ιεραρχική παράσταση των αντικειμένων δίνεται από το χρήστη και δεν αποτελεί πληροφορία που εξάγεται από τον αλγόριθμο τμηματοποίησης.

Η τμηματοποίηση δεν είναι πάντοτε ικανοποιητική. Ένα πρόβλημα που προκύπτει όταν εφαρμόζουμε το σύστημα σε ακολουθίες είναι ότι τα σύνορα εμφανίζονται ασταθή στο χρόνο. Το αποτέλεσμα γίνεται χειρότερο όταν το σύστημα εφαρμόζεται σε εικόνες που περιλαμβάνουν είτε στατικά αντικείμενα σε κινούμενο background, είτε επικαλύψεις. Για να βελτιωθεί το αποτέλεσμα, υλοποιήθηκαν δύο διαδικασίες σταθεροποίησης των συνόρων των αντικειμένων που προκύπτουν από την τμηματοποίηση. Η πρώτη ''εξομαλύνει'' τα σύνορα. Στη δεύτερη γίνεται η υπόθεση ότι το σχήμα των αντικειμένων δεν αλλάζει σημαντικά στο χρόνο και έτσι υπολογίζεται ένα ''μέσο'' σχήμα για κάθε αντικείμενο.

Το κεφάλαιο έχει οργανωθεί ως εξής: στο υποκεφ. 8.2, παρουσιάζεται ο αλγόριθμος SRG όπως έχει τροποποιηθεί για την τμηματοποίηση κίνησης. Στο υποκ. 8.3 περιγράφεται η χρήση του SRG στην παρακολούθηση της αρχικής τμηματοποίησης στο χρόνο. Στη συνέχεια ακολουθεί η περιγραφή των διαδικασιών επεξεργασίας των συνόρων που σκοπό έχουν τη βελτίωση της τμηματοποίησης στο χρόνο (υποκ. 8.4). Τέλος στην παράγραφο 8.6, παρατίθενται οι πιθανές προεκτάσεις του αλγορίθμου, ενώ στην παράγραφο 8.5 δίνονται τα αποτελέσματα της τμηματοποίησης στις ακολουθίες *Foreman* και *CoastGuard*.

8.2 Τμηματοποίηση Κίνησης με Χρήση του SRG

Στην παρούσα παράγραφο περιγράφεται η αρχική τμηματοποίηση όπου ο χρήστης δίνει ένα πολύγωνο που προσεγγίζει το σχήμα κάθε αντικειμένου, όπως είδαμε στο υποκεφάλαιο 6.3. Έτσι, οι υποθέσεις που γίνονται είναι ότι

- ο αριθμός των αντικειμένων είναι γνωστός και
- δίνεται επιπλέον ένα υποσύνολο των σημείων κάθε περιοχής.

Τα ημιτελή αυτά σύνολα αποτελούν την είσοδο του SRG και το υποσύνολο σημείων της κάθε περιοχής, χρησιμοποιείται στη συνέχεια για τον υπολογισμό του δισδιάστατου διανύσματος της ταχύτητάς της, χρησιμοποιώντας μια τεχνική *region matching* (RM) η

οποία δεν είναι τίποτε άλλο από την εφαρμογή της γνωστής τεχνικής *block matching* σε περιοχές οποιουδήποτε σχήματος αντί blocks. Η τεχνική αυτή επιλέχθηκε --έναντι ανάλογων τεχνικών που βασίζονται στην εκτίμηση της οπτικής ροής-- λόγω της χαρακτηριστικής ''ευρωστίας'' που παρουσιάζει στην εκτίμηση των παραμέτρων κίνησης. Εφόσον σ' αυτή την πρώτη τμηματοποίηση δεν υπάρχει προηγούμενη εκτίμηση για την κίνηση των αντικειμένων, συχνά η τεχνική εκτίμησης της ταχύτητας εφαρμόζεται σε αρκετά μεγάλη περιοχή αναζήτησης, γεγονός που μπορεί να οδηγήσει σε μεγάλο χρονικό κόστος. Το καθησυχαστικό γεγονός είναι ότι η πρώτη τμηματοποίηση εφαρμόζεται μόνο μία ή δύο φορές σ' όλη την ακολουθία εικόνων και επιπλέον υπάρχουν τεχνικές για τη μείωση του κόστους υπολογισμού εκτίμησης της ταχύτητας της μεθόδου [47]. Το διάνυσμα κίνησης που υπολογίζεται κατ' αυτό τον τρόπο αποτελεί το βασικό χαρακτηριστικό τμηματοποίησης των αντικειμένων, με την εφαρμογή του SRG.

Για την εφαρμογή του SRG στην τμηματοποίηση του πεδίου κίνησης, το κριτήριο (αν)ομοιότητας ενός σημείου $p(x, y)$ που δεν έχει ακόμη εισαχθεί σε κάποιο σύνολο, από το γειτονικό του σύνολο A με διάνυσμα ταχύτητας $(\hat{u}, \hat{v})$ ορίζεται να είναι:

$$\delta_{IM}(p) = |I_k(x, y) - I_{k-1}(x - \hat{u}, y - \hat{v})| \tag{8.2}$$

όπου I_k,I_{k-1}, είναι οι αντίστοιχες εικόνες έντασης τις χρονικές στιγμές k και $k-1$. Πάντως, στις περιπτώσεις που κάποιο αντικείμενο κινείται με τέτοιο τρόπο που η κίνησή του δεν περιγράφεται από το απλό μοντέλο μεταφορικής κίνησης δύο παραμέτρων, το κριτήριο που χρησιμοποιείται είναι το εξής:

$$\delta_I(p) = |I_k(x, y) - I_{k-1}(x - \hat{u}, y - \hat{v})| + \lambda \left| I_k(x, y) - \mu_{W(p)\cap R} \right| \tag{8.3}$$

όπου $\mu_{W(p)\cap R}$ είναι η μέση ένταση των σημείων της περιοχής R σ' ένα μικρό παράθυρο με κέντρο το p. Αντικαθιστώντας τις εικόνες έντασης με εκείνες του χρώματος στις εξισώσεις (8.2),(8.3) καταλήγουμε σε δύο όμοια με τα παραπάνω κριτήρια τα οποία συμβολίζονται ως δ_{CM}, δ_C. Βάσει αυτών των κριτηρίων, εφαρμόζεται στη συνέχεια ο αλγόριθμος SRG όπως περιγράφηκε στο προηγούμενο κεφάλαιο.

Αυτό που πρέπει να σημειωθεί είναι ότι σε αντίθεση με την εφαρμογή του SRG στην τμηματοποίηση εικόνων που στηρίζεται σε άλλα χαρακτηριστικά, όπως είναι η ένταση ή το χρώμα, το διάνυσμα ταχύτητας της κάθε περιοχής εκτιμάται μόνο μια φορά, πριν εφαρμοστεί ο αλγόριθμος επέκτασης των περιοχών. Η υπόθεση που γίνεται επομένως είναι ότι η αρχική περιοχή του κάθε αντικειμένου είναι αρκετά μεγάλη και αντιπροσωπευτική ώστε να είμαστε σε θέση να υπολογίσουμε με ακρίβεια την ταχύτητα της κίνησής της. Το γεγονός αυτό μειώνει σημαντικά το χρονικό κόστος των υπολογισμών.

Στο τέλος αυτής της διαδικασίας επέκτασης, σε κάθε σημείο p αντιστοιχεί μια ετικέτα $i(p) \in \{1, 2, \ldots, n, b\}$, όπου b είναι ένας ειδικός δείκτης του συνόλου B των συνοριακών σημείων. Η έξοδος του αλγορίθμου περιλαμβάνει επίσης το διάνυσμα κίνησης του κάθε συνόλου-αντικειμένου. Αυτή η έξοδος δίνεται στη συνέχεια στον αλγόριθμο παρακολούθησης των αντικειμένων στο χρόνο, όπως περιγράφεται στο επόμενο υποκεφάλαιο.

8.3 Παρακολούθηση Αντικειμένων στο Χρόνο

Σε ό,τι ακολουθεί περιγράφεται η μέθοδος παρακολούθησης του αποτελέσματος της αρχικής τμηματοποίησης --όπως αυτή έχει αποτυπωθεί στον πίνακα i_0 αντιστοίχισης των σημείων σε σύνολα (set map)-- κατά τη διάρκεια ενός αριθμού εικόνων της ακολουθίας. Υποθέτουμε ότι το αποτέλεσμα αυτό έχει αναπαραχθεί μέχρι την εικόνα $k-1$ και ότι ο πίνακας αντιστοίχισης που το περιγράφει είναι ο i_{k-1}. Σκοπός μας είναι η εξαγωγή της απεικόνισης i_k που αντιστοιχεί στην εικόνα k (k-τμηματοποίηση) για τα n σύνολα $R_1, R_2, \ldots, R_n$.

Και σ' αυτή την περίπτωση η τμηματοποίηση στηρίζεται στον αλγόριθμο SRG. Οι αρχικές περιοχές για την k-τμηματοποίηση εξάγονται από το ''χάρτη'' i_{k-1}. Επιπλέον, για τις ανάγκες της παρακολούθησης, θεωρούμε ότι τα σύνολα διατάσσονται σύμφωνα με το βάθος τους από την κάμερα (depth ordering) σε αντίθεση με την επίπεδη αναπαράσταση των συνόλων που υποδηλώνεται από τον SRG. Αναλυτικά, η διάταξή τους, η οποία δίνεται ως είσοδος από το χρήστη, είναι η εξής:

$$\forall i, j \in \{1, 2, \ldots, n\}, \quad \text{το } R_i \text{ κινείται πίσω από το } R_j \text{ εάν και μόνο εάν} \quad i < j \quad (8.4)$$

Υπό αυτή τη θεώρηση, κάθε σύνολο-περιοχή αντιστοιχεί σε ένα επίπεδο βάθους, κάθε ένα από τα οποία κινείται στο δισδιάστατο επίπεδο εκτελώντας προσεγγιστικά μια σταθερή μεταφορική κίνηση.

Έχοντας διαθέσιμη αυτή τη διάταξη, για κάθε σύνολο R_i $(i > 1)$ που περιλαμβάνεται στον πίνακα απεικόνισης i_{k-1}, (εκτός από την περιοχή του background R_0) και από το πιο απομακρυσμένο από την κάμερα στο πιο κοντινό σ' αυτή, εφαρμόζονται οι ακόλουθες διεργασίες, ώστε να εξαχθούν τα αρχικά σύνολα του πίνακα απεικόνισης i_k:

1. Εκτίμηση της μετατόπισης $(\hat{u}_k, \hat{v}_k)_i$ του R_i από την εικόνα $k-1$ στην εικόνα k εφαρμόζοντας την τεχνική RM (πιθανώς με ακρίβεια μισού pixel) σε μια μικρή περιοχή αναζήτησης γύρω από το διάνυσμα ταχύτητας του R_i, $(\hat{u}_{k-1}, \hat{v}_{k-1})_i$. Η υπόθεση που γίνεται έμμεσα είναι ότι κάθε σύνολο κινείται με σταθερή ταχύτητα ή μικρή επιτάχυνση από μια εικόνα στην επόμενή της.

2. Επέκταση (dilation) του συνόλου των συνοριακών σημείων B από την οποία προκύπτουν τα μή αντιστοιχισμένα σε κάποιο σύνολο σημεία όπως απαιτείται από τον SRG. Ο βαθμός της επέκτασης προσδιορίζεται από το χρήστη. Αν και αυτή η προσέγγιση είναι απλή και ταχύτατη, εξαλείφει σημαντικά ''λεπτά'' επιμήκη τμήματα σημείων που είναι δυνατό να εμφανίζονται στη μορφολογία των συνόλων. Για εικόνες που περιλαμβάνουν τέτοια αντικείμενα έχει υλοποιηθεί μια πιο σύνθετη διαδικασία επέκτασης: σημεία του αντικειμένου R_i, των οποίων η εξάλειψη οδηγεί τοπικά σε αποσύνδεση των μερών του, δε διαγράφονται, ενώ αντίθετα σημεία που είναι υποψήφια για διαγραφή και δεν ανήκουν στο σύνολο R_i, εξαλείφονται όπως και στην περίπτωση της απλής επέκτασης.

3. Μετατόπιση του ‘‘συρρικνωμένου’’ αντικειμένου R_i, από την εικόνα $k-1$ στην εικόνα k, σύμφωνα με το εκτιμώμενο διάνυσμα μετατόπισης $(\hat{u}_k, \hat{v}_k)_i$.

Το τελευταίο βήμα, πριν την εκτέλεση του SRG, είναι η εκτίμηση του διανύσματος ταχύτητας του background (δεν πρέπει να ξεχνάμε ότι σε αντίθεση με άλλες τεχνικές τμηματοποίησης ο SRG εφαρμόζεται σε όλες τις περιοχές της εικόνας και για κάθε μία από αυτές είναι ανάγκη να υπολογιστεί το χαρακτηριστικό ως προς το οποίο γίνεται η τμηματοποίηση).

Τελικά εφαρμόζεται ο SRG δίνοντας ως έξοδο τον (συμπληρωμένο) πίνακα απεικόνισης της τμηματοποίησης i_k.

8.4 Σταθεροποίηση των Συνόρων των Αντικειμένων

Αυτό που παρατηρείται κατά την εφαρμογή του παραπάνω σχήματος παρακολούθησης, είναι ότι τα σύνορα των αντικειμένων εμφανίζονται μετατοπισμένα από εικόνα σε εικόνα, για παράδειγμα στις περιπτώσεις αντικειμένων που δε κινούνται.

Για τη σταθεροποίηση του συνόρου των αντικειμένων στο χρόνο, έχουν υλοποιηθεί δύο διαδικασίες επεξεργασίας του σχήματος τους. Η πρώτη εξομαλύνει το σύνορο των αντικειμένων ενώ η δεύτερη υπολογίζει ένα μέσο σχήμα για κάθε αντικείμενο όπως περιγράφεται αμέσως πιο κάτω.

8.4.1 Εξομάλυνση Συνόρων των Αντικειμένων (Boundary Smoothing)

Η διαδικασία που περιγράφεται εδώ εφαρμόζεται μόνο στο σύνολο B των σημείων του συνόρου. Αρχικά, τα σημεία των συνόρων καταγράφονται, ακολουθώντας για παράδειγμα την κατεύθυνση των δεικτών του ρολογιού. Κατ’ αυτό τον τρόπο, προκύπτει μία διατεταγμένη λίστα αυτών των σημείων:

$$B_{ord} = \{ (x_1, y_1), (x_2, y_2), \ldots, (x_L, y_L) \} \tag{8.5}$$

όπου L είναι ο αριθμός των σημείων του συνόρου. Στη συνέχεια δοθέντος ενός μονοδιάστατου συμμετρικού φίλτρου χαμηλών συχνοτήτων

$$h = [h_n] \ (-N \leq n \leq N)$$

υπολογίζεται η νέα θέση $(\hat{x}_i, \hat{y}_i)$ του κάθε σημείου (x_i, y_i) του συνόλου B_ord ως εξής:

$$\begin{bmatrix} \hat{x}_i \\ \hat{y}_i \end{bmatrix} = \frac{1}{2N+1} \begin{bmatrix} \sum_{k=-N}^{N} h_k x_{i-k} \\ \sum_{k=-N}^{N} h_k y_{i-k} \end{bmatrix}$$

Σημεία του συνόρου που μετά το τέλος της διαδικασίας έχουν χάσει την απεικόνισή τους σε κάποιο σύνολο, εισάγονται στο σύνολο με το οποίο γειτονεύουν.

Η παραπάνω διαδικασία είναι δυνατό να επαναληφθεί αρκετές φορές έως ότου επιτευχθεί η επιθυμητή εξομάλυνση του συνόρου. Από την άλλη πλευρά η κατάχρηση της

εφαρμογής της μπορεί να ‘‘αλλοιώσει’’ το σχήμα κάποιων αντικειμένων εξομαλύνοντας πέραν του επιθυμητού τις γωνίες που τυχαίνει να περιέχουν. Η μέθοδος που ακολουθεί ξεπερνάει αυτό το πρόβλημα.

8.4.2 Υπολογισμός ‘‘Μέσου’’ Σχήματος (Shape Averaging)

Αυτή η μέθοδος βελτιώνει την απεικόνιση i_k της k-τμηματοποίησης χρησιμοποιώντας την πληροφορία των m προηγούμενων πινάκων απεικόνισης που συμβολίζονται στο εξής ως

$$i_{k-m}, i_{k-m+1}, \ldots, i_{k-1}$$

.

Για μια ακόμα φορά, θεωρείται ότι τα αντικείμενα ακολουθούν την ιεραρχική αντιστοίχισή τους σε επίπεδα βάθους (Εξ. 8.4). Στη συνέχεια, για κάθε σύνολο R_j $(j > 1)$, και από το πιο κοντινό στην κάμερα ως το πιο απομακρυσμένο από αυτή, εφαρμόζονται τα εξής βήματα:

1. Πρώτα κάθε πίνακας απεικόνισης i_{k-l} $(1 \leq l \leq m)$ μετατοπίζεται στην αντίστοιχη θέση του στον πίνακα i_k σύμφωνα με το συνολικό διάνυσμα μετατόπισης του αντικειμένου R_j από την εικόνα $k - l$ στην εικόνα k. Κατ’ αυτό τον τρόπο, προκύπτουν m επιπλέον απεικονίσεις για την τρέχουσα τμηματοποίηση, βάσει της ταχύτητας του αντικειμένου R_j, που αποτελούν περιγραφές του σχήματός του κατά τη διάρκεια των m + 1 τμηματοποιήσεων.

2. Έχοντας αυτούς τους μετατοπισμένους πίνακες, εφαρμόζεται στη συνέχεια ο κανόνας της πλειοψηφίας (majority law), ώστε να προκύψει η τελική απεικόνιση του συνόλου R_j στον πίνακα τμηματοποίησης i_k.

Η διαδικασία τελειώνει με την εξαγωγή του νέου συνόλου συνοριακών σημείων τα οποία είναι απαραίτητα για την εξαγωγή των αρχικών συνόλων της $(k$ + 1) τμηματοποίησης που ακολουθεί.

8.5 Πειραματικά Αποτελέσματα

Ο αλγόριθμος υλοποιήθηκε σε C και η απόδοσή του ελέγχθηκε σε μηχανήματα εργασίας ULTRA Sparc. Σ’ ότι ακολουθεί παρουσιάζονται τα αποτελέσματα που προέκυψαν από την τμηματοποίηση δύο ακολουθιών εικόνων.

Ο αλγόριθμος εφαρμόστηκε αρχικά στις 250 εικόνες της ακολουθίας *Foreman*. Οι διαστάσεις των εικόνων είναι 352×288. Η εικόνα 51 αυτής της ακολουθίας φαίνεται στο σχ. 8.1. Σ’ αυτή την ακολουθία, η κάμερα προσπαθεί να ‘‘εστιάσει’’ στο πρόσωπο του *Foreman*, ενώ αυτός κινείται, εκτελώντας μια σύνθετη κίνηση η οποία όμως τον περισσότερο χρόνο είναι μεταφορική όταν μετρείται από εικόνα σε εικόνα. Έτσι, το background του *Foreman* κινείται επίσης.

Σχήμα 8.1: Εικόνα 51 της ακολουθίας *Foreman*.

Ο αλγόριθμος παρακολούθησης εφαρμόστηκε στα μέρη της ακολουθίας όπου η ταχύτητα των αντικειμένων περιγράφεται από το μοντέλο μεταφορικής κίνησης που θεωρήσαμε έως τώρα. Για την εξαγωγή των αρχικών συνόλων που απαιτούνται ως είσοδος από τον SRG, χρησιμοποιήθηκε η απλή επέκταση του συνόρου, αφού δεν υπήρχε κίνδυνος αποκοπής κάποιου σημαντικού μέρους της μορφολογίας των υπό τμηματοποίηση αντικειμένων. Στην τμηματοποίηση, το κριτήριο ομοιότητας που χρησιμοποιήθηκε είναι το δ_C διότι η ένταση των εικόνων δε βοηθά στη διάκριση των αντικειμένων. Η τιμή της παραμέτρου λ τέθηκε στο 0.4. Το αποτέλεσμα της παρακολούθησης για την εικόνα 51 φαίνεται στο σχ. 8.2. Σ' αυτό το σχήμα, η άσπρη περιφέρεια αντιστοιχεί στο σύνορο μεταξύ του επιπέδου του *Foreman* και του επιπέδου του background (βλ. εικόνα στο σχ. 8.3).

Σχήμα 8.2: Αποτέλεσμα παρακολούθησης για την Εικόνα 51 της ακολουθίας *Foreman*.

Σχήμα 8.3: Ιεραρχική παράσταση του αποτελέσματος της παρακολούθησης για την Εικόνα 51 της ακολουθίας *Foreman*.

Από την άλλη πλευρά, όταν η κίνηση του *Foreman* δεν είναι μεταφορική, χρησιμοποιείται η πρώτη τμηματοποίηση μεταξύ δύο εικόνων, με το ίδιο κριτήριο ομοιότητας όπως πριν, αλλά με μεγάλο λ, το οποίο δείχνει ότι η τμηματοποίηση βασίζεται κυρίως στο χρώμα.

Μετά από κάθε τμηματοποίηση, εφαρμόζεται η διαδικασία εξομάλυνσης των συνόρων δύο φορές, για τη διατήρηση της σταθερότητας τους στο χρόνο. Ο μέσος υπολογιστικός χρόνος ανά εικόνα είναι περίπου 25 sec συμπεριλαμβανομένης και της διαδικασίας εξομάλυνσης του συνόρου.

Σχήμα 8.4: Εικόνα 66 της ακολουθίας εικόνων *CoastGuard*.

Το σύστημα εφαρμόστηκε επίσης στις 300 εικόνες της ακολουθίας *CoastGuard* των οποίων οι διαστάσεις είναι 352×240. Στις πρώτες 120 εικόνες της ακολουθίας, δύο ταχύπλοα κινούνται το ένα αντίθετα στο άλλο, όπως φαίνεται στο σχ. 8.4. Το

ολικό σχήμα είναι δυνατό να περιγραφεί από τρία επίπεδα, ένα για το background, ένα για το μεγάλο ταχύπλοο και ένα για το μικρό, μ' αυτή τη διάταξη σύμφωνα με την απόστασή (βάθος) τους από την κάμερα. Στις υπόλοιπες εικόνες το μικρό ταχύπλοο σταδιακά βγαίνει από την εικόνα, το μεγάλο ταχύπλοο εμφανίζεται ακίνητο, ενώ το background μεταβάλλεται λόγω της ''εστίασης'' της κάμερας στο ταχύπλοο (camera pan). Το σημαντικό σε ότι αφορά αυτή την ακολουθία είναι ότι κίνηση των αντικειμένων και του background είναι μεταφορική, γεγονός που επιτρέπει την παρακολούθηση της πρώτης τμηματοποίησης για μεγάλο αριθμό διαδοχικών εικόνων και άρα μειώνει τον απαιτούμενο αριθμό επανεκκίνησης της όλης διαδικασίας παρακολούθησης.

Τα αρχικά σύνολα για τον SRG εξάγονται από την ''υπό συνθήκη'' επέκταση του συνόρου, γεγονός που επιτρέπει τη διατήρηση της συνδεσιμότητας μεταξύ του καταρτιού και του κυρίου μέρους του πλοίου. Το κριτήριο ομοιότητας που χρησιμοποιήθηκε είναι το δ_{IM}. Αυτό το κριτήριο εφαρμόζεται στα σημεία που είναι γείτονες των συνόλων του μεγάλου ταχύπλοου, και του backgroynd. Για τα σημεία-γείτονες του μικρού ταχύπλοου, χρησιμοποιείται το κριτήριο δ_I με λ=0.4, αφού η ένταση αυτού του συνόλου είναι ομοιόμορφη. Αυτή η επιλογή των κριτηρίων βελτιώνει το αποτέλεσμα της τμηματοποίησης στις εικόνες που το μικρό ταχύπλοο κινείται μπροστά από το μεγάλο.

Σχήμα 8.5: Αποτέλεσμα τμηματοποίησης για την Εικόνα 66 της ακολουθίας *CoastGuard.*

Μετά από κάθε τμηματοποίηση, η διαδικασία εξομάλυνσης των συνόρων εφαρμόζεται μόνο μία φορά, γιατί η επανάληψη της εφαρμογής της αλλοιώνει το σχήμα των αντικειμένων. Επιπλέον στα μέρη της ακολουθίας που το μεγάλο ταχύπλοο είναι σχεδόν ακίνητο, αυτή η διαδικασία δε δίνει την επιθυμητή σταθερότητα του συνόρου. Για αυτούς τους λόγους, η επεξεργασία του αποτελέσματος της τμηματοποίησης περιλαμβάνει και την εκτίμηση του ''μέσου'' σχήματος του κάθε αντικειμένου. Αυτή η διαδικασία όχι μόνο σταθεροποιεί το σύνορο, αλλά βελτιώνει σημαντικά το αποτέλεσμα τμηματοποίησης στις εικόνες που εμφανίζονται επικαλύψεις (occlusions) μεταξύ αντικειμένων, δίνοντας μια πολύ καλή ιεραρχική παράσταση βάθους των αντικειμένων. Η επίδραση αυτής της μεθόδου φαίνεται στις εικόνες των σχημάτων 8.5 και 8.6. Το σχήμα 8.5 δείχνει

το αποτέλεσμα της τμηματοποίησης για την εικόνα 66 και το σχ. 8.6 την επίδραση της εφαρμογής του φίλτρου σχήματος σ'αυτό το αποτέλεσμα, χρησιμοποιώντας την πληροφορία που έχει εξαχθεί από τα m=12 προηγούμενα αποτελέσματα της τμηματοποίησης. Στο σχ. 8.7, βλέπουμε τα τρία επίπεδα που προκύπτουν από το βελτιωμένο αποτέλεσμα του σχ. 8.6. Τέλος, ο συνολικός (τμηματοποίησης και βελτίωσης) μέσος χρόνος ανά εικόνα είναι περίπου 40sec για m=12.

Σχήμα 8.6: Βελτιωμένο αποτέλεσμα της ακολουθίας εικόνων *CoastGuard.*

Σχήμα 8.7: Ιεραρχική παράσταση του μικρού ταχύπλοου, του μεγάλου ταχύπλοου και του background που αποτελούν την εικόνα 66 της ακολουθίας *CoastGuard.*

8.6 Συμπεράσματα-Επεκτάσεις

Στο κεφάλαιο αυτό παρουσιάστηκε ένα σύστημα τμηματοποίησης που βασίζεται στην κίνηση των αντικειμένων μιας ακολουθίας εικόνων. Στη συνέχεια είδαμε τον τρόπο με τον οποίο γίνεται η παρακολούθηση του αρχικού αποτελέσματος σε μια σειρά διαδοχικών εικόνων.

Για τις ανάγκες της παρακολούθησης, εισήχθηκε μια ιεραρχική παράσταση των αντικειμένων, ανάλογα με το βάθος τους, έτσι ώστε να είμαστε σε θέση να χειριστούμε, καταστάσεις που προκύπτουν από την εφαρμογή του συστήματος σε ακολουθίες με πολλαπλές κινήσεις, κινούμενο background και αλληλοεπικαλύψεις αντικειμένων.

Τέλος, παρουσιάστηκαν τρόποι βελτίωσης του αποτελέσματος της παρακολούθησης που αποσκοπούν σε ένα καλύτερο οπτικό αποτέλεσμα, και ταυτόχρονα βελτιώνουν τις αρχικές συνθήκες εφαρμογής της τμηματοποίησης μέσω του SRG.

Ο χρήστης καθοδηγεί το όλο σύστημα, αφού παρέχει τα αρχικά ημιτελή σύνολα-είσοδο του SRG, τη διάταξη ως προς το βάθος τους και τις παραμέτρους των κριτηρίων τμηματοποίησης. Επίσης επιλέγει μία ή και τις δύο από τις μεθόδους επεξεργασίας του συνόρου που περιγράφηκαν. Εναλλακτικά, η παρακολούθηση θα μπορούσε να αφορά αποτελέσματα που έχουν προκύψει από άλλες εφαρμογές τμηματοποίησης, όπως είναι αυτή που περιγράφηκε στο προηγούμενο κεφάλαιο, που βασίζεται στο χρώμα περιοχών ή αντικειμένων. Μ' αυτό τον τρόπο, τα αρχικά σύνολα εισάγονται αυτόματα, απομένει όμως στο χρήστη να δώσει επιπλέον ως είσοδο, τη διάταξη βάθους των αντικειμένων.

Το σύστημα παρακολούθησης αντικειμένων βάσει της ταχύτητάς τους που περιγράφηκε σ' αυτό το κεφάλαιο, θα μπορούσε να επεκταθεί και για αντικείμενα που η κίνησή τους δεν ακολουθεί το απλό μοντέλο μεταφορικής κίνησης για το οποίο υλοποιήθηκε.

Τέλος, πρέπει να σημειωθεί η γενικότερη χρησιμότητα των μεθόδων παρακολούθησης για το βαθμό συμπίεσης που προσφέρουν, μειώνοντας σημαντικά το χρονικό πλεονασμό που συνήθως χαρακτηρίζει τις σκηνές του video.

Παράρτημα Α

Μετασχηματισμοί του Χώρου Χρώματος RGB

Σε ό,τι ακολουθεί, παρουσιάζονται οι μετασχηματισμοί από το χώρο χρώματος RGB, στους χώρους που χρησιμοποιήθηκαν για την επεξεργασία των εικόνων .

Α.1 Μετασχηματισμός RGB σε YIQ

Ο μετασχηματισμός από τα στοιχειώδη χρώματα του RGB όπως αυτά ορίστηκαν από την NTSC [1], είναι ο εξής:

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.274 & -0.322 \\ 0.211 & -0.523 & 0.312 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{A.1}$$

Α.2 Μετασχηματισμός RGB σε YCbCr

Ο μετασχηματισμός από το χώρο RGB της NTSC στο YCbCr είναι ο εξής:

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.1687 & -0.3313 & 0.5 \\ 0.5 & -0.4187 & 0.0813 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{A.2}$$

Α.3 Μετασχηματισμός RGB σε XYZ

Ο πίνακας μετασχηματισμού από το RGB της NTSC στο XYZ, εξαρτάται από την τιμή του λευκού που έχει επιλεγεί ως λευκό χρώμα αναφοράς (reference white color). Θεωρώντας ότι ως λευκό χρώμα αναφοράς έχει επιλεγεί ‘‘ο συννεφιασμένος ουρανός το μεσημέρι’’

[1] Η NTSC (αντίθετα από τη CIE) όρισε τα στοιχειώδη χρώματα του RGB, έτσι ώστε και τα τρία να παίρνουν τιμές από 0.0 έως 1.0)

(μτφρ. του ‘‘overcast sky at noon’’) όπως ορίστηκε από τη CIE, ο μετασχηματισμός δίνεται από τη σχέση:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.607 & 0.174 & 0.201 \\ 0.299 & 0.587 & 0.114 \\ 0.0 & 0.066 & 1.117 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{A.3}$$

A.4 Μετασχηματισμός RGB σε Lab

Ο χώρος χρώματος XYZ, χρησιμοποιείται ως το ενδιάμεσο στάδιο για το μετασχηματισμό από το RGB στα συστήματα χρώματος Lab και Luv. Για το χώρο Lab, ο μετασχηματισμός δίνεται από τις σχέσεις:

$$L=116\left(\frac{Y}{Y_0}\right)^{\frac{1}{3}} - 16 \quad \text{εάν } \frac{Y}{Y_0} \geq 0.008856 \tag{A.4}$$

$$L=903.3\frac{Y}{Y_0} \quad \text{εάν } \frac{Y}{Y_0} < 0.008856 \tag{A.5}$$

$$a=500\left[\left(\frac{X}{X_0}\right)^{\frac{1}{3}} - \left(\frac{Y}{Y_0}\right)^{\frac{1}{3}}\right] \tag{A.6}$$

$$b=200\left[\left(\frac{Y}{Y_0}\right)^{\frac{1}{3}} - \left(\frac{Z}{Z_0}\right)^{\frac{1}{3}}\right] \tag{A.7}$$

A.5 Μετασχηματισμός RGB σε Luv

Οι σχέσεις που δίνουν τα χρώματα στο χώρο Luv, είναι οι εξής:

$$L=116\left(\frac{Y}{Y_0}\right)^{\frac{1}{3}} - 16 \quad \text{εάν } \frac{Y}{Y_0} \geq 0.008856 \tag{A.8}$$

$$L=903.3\frac{Y}{Y_0} \quad \text{εάν } \frac{Y}{Y_0} < 0.008856 \tag{A.9}$$

$$u=13L(u' - u_0) \tag{A.10}$$

$$v=13L(v' - v_0) \tag{A.11}$$

όπου

$$u'=\frac{4X}{X + 15Y + 3Z} \tag{A.12}$$

$$v'=\frac{9Y}{X + 15Y + 3Z} \tag{A.13}$$

και

$$u_0=\frac{4X_0}{X_0 + 15Y_0 + 3Z_0} \tag{A.14}$$

$$v_0=\frac{9Y_0}{X_0 + 15Y_0 + 3Z_0} \tag{A.15}$$

Στις παραπάνω σχέσεις οι όροι X_0, Y_0 και Z_0, αντιστοιχούν στις τιμές που προκύπτουν για το λευκό χρώμα αναφοράς που επιλέχθηκε, όπως αναφέρθηκε πιο πάνω.

Βιβλιογραφία

[1] F. Idris and S. Panchanathan. Review of image and video indexing techniques. *Journal of Visual Comminication and Image Representation*, 8:146--166, June 1997.

[2] G. Ahanger and T. Little. A survey of technologies for parsing and indexing digital video. *Journal of Visual Comminication and Image Representation*, 7:28--43, March 1996.

[3] T.A. Smith G. Davenport and N. Pincever. Cinematic primitives for multimedia. *IEEE Computer Graphics Applications*, pages 67--74, July 1991.

[4] S.W. Smoliar H.J. Zhang, C.Y. Low and J.H. Wu. Video parsing, retrieval and browsing: An intergrated and content-based solution. *Proc. 4th ACM Conference on Multimedia*, pages 15--24, 1995.

[5] H. Sawhney W. Niblack et al. M. Flickner. Query by image and video content: The qbic system. *IEEE Computer*, 28:23--32, September 1995.

[6] R. Picard A. Pentland and S. Sclaroff. Photobook: Content-based manipulation of image databases. *International Journal of Computer Vision*, 18:233--254, June 1996.

[7] J.R. Smith and S.-F. Chang. Tools and techniques for color image retrieval. *to appear in Proc. IS&T/SPIE, Storage and Retrieval for Image and Video Database IV*, February 1996.

[8] J.R. Smith and S.-F. Chang. Visualseek: A fully automated content-based image query system. *ACM Multimedia Conference*, pages 87--98, November 1996.

[9] E. Veneau P. Bouthemy, C. Garcia R. Ronfard G. Tziritas and D. Zugaj. 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.

[10] P.E. Trahanias J. Sporring, X. Zabulis and S.C. Orphanoudakis. Shape similarity by piecewise linear alignment. *to appear in Proc. of ACCV 2000.*

[11] A.K. Jain. *Fundamentals of Digital Image Processing*. Prentice-hall, 1989.

[12] *International Commission on Illumination.* http://www.hike.te.chiba-u.ac.jp/ikeda/CIE.

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

[14] W.K. Pratt. *Digital Image Processing.* John Wiley and Sons, 1991.

[15] V.N. Gudivada and V.V. Raghavan. Content-based image retrieval systems. *IEEE Computer*, 28:18--22, September 1995.

[16] 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.

[17] Σ. Λιάπης. *Εξαγωγή Χαρακτηριστικών Υφής και Χρώματος για Τμηματοποίηση και Ανάκληση Εικόνων.* Μεταπτυχιακή Εργασία, Τμήμα Επιστήμης Υπολογιστών, Πανεπιστήμιο Κρήτης, 1999.

[18] M. Stricker and M. Swain. The capacity and the sensitivity of color histogram indexing. *Department of Computer Science, University of Chicago, Technical Report*, TR 94-05, 1994.

[19] *International Organization for Standardization ISO/IEC 11172 (MPEG).*

[20] A.K. Elmagarmid H. Jiang, A.(S.) Helal and A. Joshi. Scene change detection techniques for video database systems. *Multimedia Systems*, 6:186--195, 1998.

[21] U. Gargi R. Kasturi, S.H. Strayer and S. Antani. An evaluation of color histogram based methods in video indexing. *Department of Computer Science and Engineering, The Pennsylvania State University, Technical Report*, CSE-96-053, September 1996.

[22] R. Muranoi J. Zhao, R. Hayasaka and Y. Matsushita. A mpeg video structure analysis scheme and its application to hierarchical video browser. *Matsusita Laboratory, Dept. of Instrumentation Engg., Keio University, Technical Report*, TSM-S003-A.

[23] J. Miller R. Zabih and K. Mai. A feature-based algorithm for detecting and classifying scene breaks. *Proc. 4th ACM Conference on Multimedia*, pages 189--200, 1995.

[24] M. Gelgon P. Bouthemy and F. Ganansia. A unified approach to shot change detection and camera motion characterization. *INRIA, Technical Report*, No. 3304.

[25] Μ. Τράκα. *Πανοραμική Σύνθεση Όψεων από μια Ακολουθία Εικόνων.* Μεταπτυχιακή Εργασία, Τμήμα Επιστήμης Υπολογιστών, Πανεπιστήμιο Κρήτης, 1999.

[26] 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.

[27] E. Trucco and A. Verri. *Introductory Techniques for 3-D Computer Vision*. Prentice Hall, 1998.

[28] A. Witkin M. Kass and D. Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, 4:7--38, 1990.

[29] E. Segawa G. Xu and S. Tsuji. Robust active contours with insensitive parameters. *Pattern Recognition*, 27:879--884, 1994.

[30] R. Ronfard. Region based strategies for active contour models. *International Journal of Computer Vision*, 13, 1994.

[31] R. Malladi and J.A. Sethian. An o(nlogn) algorithm for shape modeling. *Applied Mathematics*, 93:9389--9392, September 1996.

[32] J.A. Sethian. A fast marching level set method for monotonically advancing fronts. *Applied Mathematics*, 93:1591--1595, February 1996.

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

[34] A. Gupta A. Leonardis and R. Bajcsy. Segmentation of range images as the search for geometric parametric models. *International Journal of Computer Vision*, 14:253--277, 1995.

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

[36] S. Geman D. Geman and P. Dong. Boundary detection by constrained optimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23:609--628, 1990.

[37] Y.G. Leclerc. Constructing simple stable descriptors for image partitioning. *International Journal of Computer Vision*, 3:73--102, 1989.

[38] S.C. Zhu and A. Yuille. Region competition: Unifying snakes, region growing, and bayes/mdl for multiband image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18:884--900, September 1996.

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

[40] S.A. Shafer G.J. Klinker and T. Kanade. A physical approach to color image understanding. *International Journal of Computer Vision*, 4:7--38, 1990.

[41] K.N. Plataniotis D. Androutsos and A. N. Venetsanopoulos. Image region extraction for content-based image retrieval. *in Proc. of European Signal Processing Conference IX*, September 1998.

[42] C.C. Yang and J.J. Rodriguez. Efficient luminance and saturation processing techniques for color images. *Journal of Visual Communication and Image Representation*, 8:263--277, September 1997.

[43] H. Levkowitz and G.T. Herman. Glhs: A generalized lightness, hue, and saturation color model. *Graphical Models and Image Processing*, 55:271--285, July 1993.

[44] K.N. Plataniotis N. Herodotou and A.N. Venetsanopoulos. A color segmentation and classification scheme for facial image and video retrieval. *in Proc. of European Signal Processing Conference IX*, September 1998.

[45] S.C. Zhu and A. Yuille. A unified theory for image segmentation: Region competition and its analysis. *Harvard Robotics Laboratory Technical Report*, TR 95-7.

[46] *Numerical Recipes in C: The Art of Scientific Computing*. http://www.ulib.org/webRoot/Books/NumericalRecipes/bookc.html.

[47] G. Tziritas and C. Labit. *Motion Analysis and Image Sequence Coding*. Elsevier, 1994.

[48] A. Mitiche and P. Bouthemy. Computation and analysis of image motion: A synopsis of current problems and methods. *International Journal of Computer Vision*, 19:29--55, July 1996.

[49] I. Grinias and G. Tziritas. Motion segmentation and tracking using a seeded region growing method. *in Proc. of European Signal Processing Conference IX*, September 1998.