

# **Ανίχνευση Προσώπων και Χαρακτηριστικών Προσώπου**

Γ. Σημαντήρης, Ι. Μαυρικάκης, Η. Γκρίνιας,  
C. Garcia και Γ. Τζιρίτας

Τεχνική αναφορά  
CSD-TR-2001-02

3 Μάϊου 2001

## Περίληψη

Στην αναφορά αυτή ευρίσκονται συγκεντρωμένες όλες οι εργασίες που έχουν γίνει από τους συγγραφείς σχετικά με την ανίχνευση προσώπων και χαρακτηριστικών προσώπου. Με δοσμένο ότι η αναζήτηση προσώπου γίνεται σε οποιαδήποτε εικόνα, το πρώτο στάδιο συνίσταται στην ανίχνευση χρωμάτων προσώπου. Γι' αυτό αναπτύχθηκαν τρεις μέθοδοι: προσδιορισμός γεωμετρικού τόπου από δεδομένα, δίκτυο διανυσμάτων υποστήριξης και πολυστρωματικό νευρωνικό δίκτυο. Μετά την ανίχνευση χρωμάτων σχηματίζονται υποψήφιες περιοχές για πρόσωπα που ελέγχονται ως προς το σχήμα με βάση ανθρωπομορφικά κριτήρια. Το τελικό στάδιο βασίζεται στην πολυκαναλική ανάλυση κυματιδιακών πλαισίων. Μέσω αυτής της ανάλυσης εξάγονται τοπικά χαρακτηριστικά υφής που αξιοποιούνται για την ταξινόμηση σε πρόσωπο ή μη. Η ανίχνευση χαρακτηριστικών προσώπου περιλαμβάνει κατ' αρχήν επιμέρους ανιχνευτές ματιών και στόματος. Χρησιμοποιεί επιπλέον μία προσαρμοζόμενη στο πρόσωπο φόρμα.

# *Face and Face Feature Detection*

G. Simantiris, I. Mavrikakis, I. Grinias, C. Garcia, and G. Tziritas

Technical Report  
CSD-TR-2001-02

3rd May 2001

We report our work on face detection and face feature detection. Given that we consider images under nonconstrained scene conditions, the first stage of our method consists on skin colour detection. Three techniques have been developed and tested: construction of skin colour domain from image data, training of a support vector machine, and training of a multilayer perceptron. Candidate regions are then tested according to their shape. Finally, a wavelet packet analysis is performed, and local features are extracted, which are then used for the classification in human face or not. The face feature detection is based on eyes and mouth detectors. In addition, a deformable template is used for the verification of human face presence.

# Περιεχόμενα

<b>1</b>	<b>Εισαγωγή</b>	<b>1</b>
<b>2</b>	<b>Ανίχνευση περιοχών χρώματος ανθρώπινου δέρματος</b>	<b>3</b>
2.1	Κατασκευή γεωμετρικού τόπου χρωμάτων . . . . .	3
2.1.1	Κβαντισμός χρωμάτων . . . . .	3
2.1.2	Ταξινόμηση του χρώματος του ανθρώπινου δέρματος . . . . .	4
2.2	Δίκτυο διανυσμάτων υποστήριξης . . . . .	5
2.2.1	Επιλογή και αρχιτεκτονική του νευρωνικού δικτύου . . . . .	6
2.2.2	Επιφάνειες διαχωρισμού για την ταξινόμηση . . . . .	9
2.3	Πολυστρωματικό νευρωνικό δίκτυο . . . . .	13
2.3.1	Περιγραφή του νευρωνικού δικτύου . . . . .	13
2.3.2	Λειτουργία του δικτύου και τελικά κριτήρια. . . . .	13
<b>3</b>	<b>Υποψήφιες περιοχές προσώπου και έλεγχος του σχήματός τους</b>	<b>16</b>
3.1	Αλγόριθμος συγχώνευσης . . . . .	16
3.2	Ανθρωπομορφικά κριτήρια . . . . .	17
<b>4</b>	<b>Έλεγχος υφής με ανάλυση σε κυματιδιακά πλαίσια</b>	<b>19</b>
4.1	Πολυκαναλική ανάλυση σε κυματιδιακά πλαίσια . . . . .	20
4.2	Ταξινόμηση των χαρακτηριστικών διανυσμάτων . . . . .	24
<b>5</b>	<b>Αποτελέσματα ανίχνευσης προσώπων</b>	<b>27</b>
<b>6</b>	<b>Ανίχνευση χαρακτηριστικών προσώπου για απόρριψη λάθος ανιχνεύσεων</b>	<b>29</b>
6.1	Αρχικοποίηση της μεθόδου . . . . .	30
6.2	Ελαστική αντιστοίχιση με φόρμες . . . . .	33
6.3	Ταξινόμηση των διανυσμάτων περιγραφής . . . . .	38
<b>7</b>	<b>Ανίχνευση χαρακτηριστικών προσώπων με ελαστική αντιστοίχιση με φόρμες</b>	<b>47</b>
7.1	Παρουσίαση μεθόδου . . . . .	47
7.2	Ανίχνευση κοιλάδων -- μοντελοποίηση ματιού και στόματος . . . . .	48
7.2.1	Μοντελοποίηση ματιού . . . . .	51
7.2.2	Μοντελοποίηση στόματος . . . . .	55
7.3	Ελαστική αντιστοίχιση φόρμας, δημιουργία διανυσμάτων περιγραφής και εφαρμογή κριτηρίων ελέγχου . . . . .	56
7.3.1	Περιγραφή της γεωμετρικής φόρμας . . . . .	56
7.3.2	Στατιστικά μέτρα της κυματιδιακής ανάλυσης . . . . .	58
7.3.3	Ελαστική αντιστοίχιση και εφαρμογή ελέγχων . . . . .	59
7.3.4	Ταξινόμηση . . . . .	61
7.3.5	Τελική επιλογή . . . . .	62
<b>8</b>	<b>Επίλογος</b>	<b>72</b>

# Κεφάλαιο 1

## Εισαγωγή

Η αυτόματη ανίχνευση ανθρώπινων προσώπων γίνεται ολοένα και πιο σημαντική σε πολλές εφαρμογές. Μετά την επιτυχημένη ανίχνευση του ανθρώπινου προσώπου θα μπορούσε να ακολουθήσει η αναγνώρισή του μέσα από μία βάση προσώπων, για να προσδιοριστεί η ταυτότητά του, να ανιχνευτούν τα επιμέρους χαρακτηριστικά του προσώπου (όπως είναι τα μάτια, η μύτη και το στόμα) ή να βρεθεί η θέση και η κατεύθυνσή του στον τρισδιάστατο χώρο. Θα μπορούσε επίσης να ξεκινήσει μία παρακολούθηση του προσώπου ή και των επιμέρους χαρακτηριστικών του στο χρόνο, έτσι ώστε να προσδιοριστεί η συνολική του κίνηση ή οι επιμέρους κινήσεις των χαρακτηριστικών του για εφαρμογές στην επίβλεψη σκηνών, ή στην αναγνώριση της ομιλίας και των ανθρώπινων εκφράσεων.

Στα προβλήματα της ανίχνευσης και αναγνώρισης ανθρώπινων προσώπων έχει αποδοθεί μεγάλη σημασία τα τελευταία χρόνια. Παρόλο που η ανίχνευση σχετίζεται στενά με την αναγνώριση σαν ένα συχνά απαραίτητα προκαταρκτικό βήμα, μεγαλύτερη προσοχή έχει δοθεί κυρίως στην αναγνώριση. Σημαντική πρόοδος σημειώθηκε κυρίως κάτω από σταθερές συνθήκες, όπως μικρές αποκλίσεις στη φωτεινότητα, στις εκφράσεις και στη στάση του προσώπου. Αναλυτικές έρευνες παρουσιάζονται στα άρθρα [5], [38]. Όλες αυτές οι μέθοδοι μπορούν πάνω κάτω να ταξινομηθούν σε δύο κατηγορίες : γεωμετρική αντιστοίχιση χαρακτηριστικών και αντιστοίχιση με φόρμα (template matching). Στην πρώτη κατηγορία εξάγονται κάποια γεωμετρικά κριτήρια για σαφή χαρακτηριστικά προσώπου όπως είναι τα μάτια, η μύτη, το στόμα και το σαγόκι [3], [9]. Στη δεύτερη κατηγορία η εικόνα του ανθρώπινου προσώπου, που αναπαρίσταται σαν ένας δισδιάστατος πίνακας από τιμές φωτεινότητας, συγκρίνεται με μία ή περισσότερες φόρμες που εκπροσωπούν ένα πρόσωπο. Οι πιο πρώιμες μέθοδοι για αντιστοίχιση με φόρμες (template matching) βασίζονται στη συσχέτιση και άρα είναι υπολογιστικά ακριβές και απαιτούν ταυτόχρονα μεγάλη αποθηκευτική χωρητικότητα. Πριν από μερικά χρόνια χρησιμοποιήθηκε επιτυχώς και η ανάλυση σε κύριες συνιστώσες (principal components analysis, PCA), ή αλλιώς γνωστή ως μετασχηματισμός Karhunen-Loewe (KLT), και επιτεύχθηκε μείωση διαστάσεων (dimensionality reduction) [1], [20], [26], [33], [35]. Επίσης δοκιμάστηκαν και χρησιμοποιήθηκαν και άλλες μέθοδοι, όπως ταξινόμηση με νευρωνικά δίκτυα [8], [27], αλγεβρικές ροπές (algebraic moments) [16], ισοπυκνωτικές γραμμές (isodensity lines) [25], και προσαρμοζόμενες φόρμες (deformable model of templates) [21], [39].

Στις περισσότερες προσεγγίσεις αναγνώρισης προσώπου η ύπαρξη και η θέση του ανθρώπινου προσώπου στην επεξεργαζόμενη εικόνα είναι γνωστή εξ αρχής. Αυτό σημαίνει ότι υπάρχει ελάχιστη ανάγκη να ανιχνευτεί και να εξακριβωθεί το πρόσωπο. Όμως σε βάσεις εικόνων και βίντεο δεν τίθενται γενικά περιορισμοί ως προς τον αριθμό, τη θέση, το μέγεθος και τον προσανατολισμό του ανθρώπινου προσώπου και ταυτόχρονα οι συνθήκες φωτισμού και το φόντο της εικόνας δεν είναι προκαθορισμένες και γνωστές. Περαιτέρω η χρωματική πληροφορία παρόλο που είναι αρκετά χρήσιμη για την ανίχνευση δε βρίσκει μεγάλη χρησιμότητα στην αναγνώριση. Τα τελευταία χρόνια παρατηρείται αυξανόμενη δραστηριότητα στην ανάπτυξη αλγορίθμων για να εντοπίζονται πρόσωπα σε εικόνες όπου είναι γνωστή η ύπαρξη προσώπου ή να ανιχνεύονται πρόσωπα σε μη ελεγχόμενες εικόνες [28], [38]. Οι γνωστές μέθοδοι μπορούν να ταξινομηθούν γενικά σε τρεις ευρύτερες κατηγορίες : τοπική ανίχνευση χαρακτηριστικών προσώπου, αντιστοίχιση με φόρμες (template matching) και αναλλοίωτες εικόνων. Στην πρώτη περίπτωση εφαρμόζονται χαμηλού επιπέδου αλγόριθμοι υπολογιστικής όρασης

για να εντοπιστούν χαρακτηριστικά προσώπου (μάτια, μύτη, στόμα, κτλ.). Κατόπιν χρησιμοποιούνται στατιστικά μοντέλα ανθρώπινων προσώπων για να γίνει μία ταξινόμηση αυτών [4], [10], [17], [31], [40], [41]. Στη δεύτερη περίπτωση διάφορες φόρμες συσχέτισης χρησιμοποιούνται για να ανιχνευτούν τοπικά υποχαρακτηριστικά που μπορούν να θεωρηθούν είτε σταθερά στην εμφάνιση (view based eigenspaces, [26]) είτε προσαρμόσιμα (deformable templates, [39]). Στην τρίτη περίπτωση τα συστήματα υποθέτουν ότι υπάρχει μία χωρική σχέση που είναι κοινή και πιθανώς μοναδική σε όλα τα πρότυπα προσώπων, και διατηρείται ακόμα και κάτω από διαφορετικές συνθήκες απεικόνισης. Αντί να ανιχνεύονται τα πρόσωπα ακολουθώντας κάποια σειρά από κανόνες σχεδιασμένους από ανθρώπους, εναλλακτικές λύσεις βασίζονται σε νευρωνικά δίκτυα [21], [29], [32], που έχουν το πλεονέκτημα να μαθαίνουν τους απαιτούμενους κανόνες από μία συλλογή αντιπροσωπευτικών παραδειγμάτων ανθρώπινων προσώπων, αλλά έχουν και το μεγάλο μειονέκτημα να είναι υπολογιστικά ακριβά και δύσκολα στην εκπαίδευση, λόγω της δυσκολίας να χαρακτηριστούν αντιπροσωπευτικά παραδείγματα της κατηγορίας “μη πρόσωπο”.

Ακολουθούν τα στάδια που αναπτύχθηκαν για να αποτελέσουν ένα σύστημα ανίχνευσης του ανθρώπινου προσώπου σε ψηφιακές έγχρωμες εικόνες, χωρίς περιορισμούς στη σκηνή, όπως είναι η παρουσία πολύπλοκου φόντου και φωτισμού, και όπου οι περισσότερες μέθοδοι που βασίζονται στα τοπικά χαρακτηριστικά προσώπου δεν είναι σταθερές. Σε μία πρώτη εργασία [13] αναπτύχθηκε μία μέθοδος ανίχνευσης προσώπων ειδικά για να επεξεργάζεται μεγάλους όγκους δεδομένων από βίντεο και εικόνες και να συμβαδίζει με απαιτήσεις υψηλής απόδοσης. Το σύστημα που παρουσιάζεται εδώ έχει εμπλουτιστεί σημαντικά (βλ. και [14] για περισσότερες λεπτομέρειες). Ο αλγόριθμος ανιχνεύει πρόσωπα χρησιμοποιώντας κβαντισμένες περιοχές ως προς το χρώμα του ανθρώπινου προσώπου, συγχώνευση (merging) και ανάλυση με πακέτα πλαισίων κυματιδίων. Ανθρώπινα πρόσωπα ανιχνεύονται αυτόματα σε έγχρωμες ψηφιακές εικόνες, χωρίς σκηνικούς περιορισμούς (όπως είναι η παρουσία πολύπλοκου φόντου και ανεξέλεγκτης φωτεινότητας). Αρχικά εφαρμόζεται μία ομαδοποίηση χρώματος (color clustering) για να εξαχθούν τα κυρίαρχα χρώματα και να κβαντιστεί μετέπειτα η εικόνα σύμφωνα με αυτά τα νέα και μειωμένα χρώματα. Ύστερα ακολουθεί μία τμηματοποίηση βασισμένη στο χρώμα. Το επόμενο στάδιο της συγχώνευσης εφαρμόζεται επαναληπτικά πάνω στην ομάδα των χρωμάτων, που ανήκουν σε αυτήν του χρώματος του ανθρώπινου προσώπου, στην κβαντισμένη ως προς το χρώμα εικόνα. Έτσι εξάγεται μια σειρά από υποψήφιες περιοχές προσώπου. Στις περιοχές αυτές εφαρμόζονται κριτήρια σχετικά με το μέγεθος και το σχήμα του προσώπου και αναλύεται η υφή του προσώπου σε κάθε υποψήφια περιοχή χρησιμοποιώντας μία ανάλυση πακέτων πλαισίων κυματιδίων με τελικό στόχο την ανίχνευση του ανθρώπινου προσώπου. Κάθε υποψήφια περιοχή χαρακτηρίζεται από μία σειρά εικόνων που περιέχουν τους συντελεστές κυματιδίων. Μία σειρά από απλά στατιστικά δεδομένα εξάγεται από αυτούς τους συντελεστές για να σχηματιστούν διανύσματα περιγραφής του υποψηφίου προσώπου και στη συνέχεια ένα κατάλληλο στατιστικό μέτρο, που εξάγεται με βάση την απόσταση Bhattacharyya, χρησιμοποιείται για την ταξινόμηση, συγκρίνοντας έτσι το διάνυσμα περιγραφής με κάποια πρότυπα διανύσματα περιγραφής προσώπου, τα οποία δημιουργήθηκαν ύστερα από μία διαδικασία εκπαίδευσης.

Η μέθοδος επεκτείνεται στη συνέχεια για να αντιμετωπίσει ζητήματα ανίχνευσης χαρακτηριστικών προσώπου. Εισάγονται γι' αυτό το λόγο τεχνικές ανίχνευσης ματιών, στόματος και μύτης, με ταυτόχρονο έλεγχο μιας προσαρμοζόμενης στο όλο πρόσωπο φόρμας. Η πολυκαναλική ανάλυση με χρήση κυματιδιακών πλαισίων αποτελεί και εδώ την κύρια μεθοδολογική προσέγγιση.

## Κεφάλαιο 2

# Ανίχνευση περιοχών χρώματος ανθρώπινου δέρματος

Το πρώτο στάδιο της ανίχνευσης αποτελεί η εύρεση πιθανών περιοχών προσώπου στην εικόνα. Αυτό γίνεται λαμβάνοντας υπόψη την πληροφορία του χρώματος και συγκεκριμένα του χρώματος του ανθρώπινου δέρματος. Και αυτό γιατί μία τέτοια πληροφορία μπορεί να μειώσει σημαντικά το χώρο αναζήτησης, περιορίζοντας την εφαρμογή των υπόλοιπων σταδίων του αλγορίθμου σε τμήματα της εικόνας, αντί για ολόκληρη, που θα ήταν υπολογιστικά ακριβό. Θα δούμε τρεις μεθόδους που αναπτύχθηκαν και δοκιμάσθηκαν για την ανίχνευση του χρώματος του ανθρώπινου δέρματος σε έγχρωμες εικόνες. Οι δύο τελευταίες σχετίζονται, καθώς έχουν να κάνουν με νευρωνικά δίκτυα που επιλέχθηκαν και εκπαιδεύτηκαν προσεκτικά για το συγκεκριμένο σκοπό, εξοικονομώντας το δυνατό σε υπολογιστικό κόστος και κερδίζοντας σε απόδοση.

### 2.1 Κατασκευή γεωμετρικού τύπου χρωμάτων

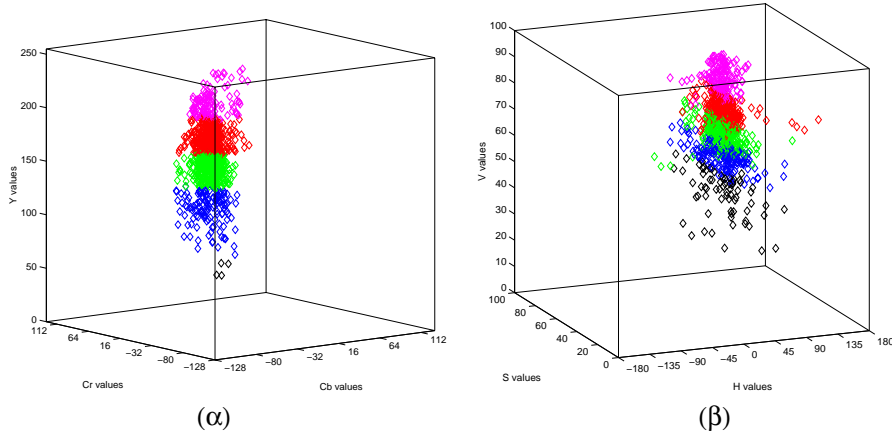
Πριν την ταξινόμηση των χρωμάτων σε αυτά που αντιστοιχούν στο ανθρώπινο δέρμα και σε χρώματα που δεν αντιστοιχούν σ' αυτό, εφαρμόζεται αρχικά ένας κβαντισμός χρωμάτων, ώστε να βελτιωθεί η ταξινόμηση ομοιογενοποιώντας τα διάφορα χρωματικά τμήματα της αρχικής εικόνας.

#### 2.1.1 Κβαντισμός χρωμάτων

Εφαρμόζεται ο διανυσματικός κβαντισμός για να κβαντιστούν τα χρώματα της αρχικής εικόνας σε έναν περιορισμένο αριθμό, τα λεγόμενα κυρίαρχα χρώματα. Η βασική ιδέα συνοψίζεται στην απεικόνιση ενός χρωματικού διανύσματος  $x$  πάνω σε ένα άλλο χρωματικό διάνυσμα  $y_i$ . Το  $y_i$  αποτελεί μέρος ενός λεξικού κύριων χρωμάτων αντιπροσωπεύοντας μία ομάδα χρωμάτων  $C_i$  στο χρωματικό χώρο. Κάθε κυρίαρχο χρώμα είναι ουσιαστικά η μέση τιμή όλων των χρωμάτων που ανήκουν στην  $C_i$ . Για να δημιουργηθεί το λεξικό των κύριων χρωμάτων υπολογίζεται το τρισδιάστατο ιστογράμμα από την αρχική έγχρωμη εικόνα στον χρωματικό χώρο  $HSV$ . Τα τμήματα του ιστογράμματος (bins) υφίστανται ιεραρχική συγχώνευση σύμφωνα με την ευκλείδεια απόσταση στο χώρο ( $SV \cos(H)$ ,  $SV \sin(H)$ ,  $V$ ) μετατρέποντας το χρωματικό χώρο  $HSV$  σε έναν κώνο (βλ. και [2]) για να δημιουργηθεί ένα αρχικό σύνολο από ομάδες χρωμάτων  $C_i$ . Το τελικό λεξικό των κύριων χρωμάτων προκύπτει χρησιμοποιώντας τον επαναληπτικό αλγόριθμο  $K$ -means, γνωστό και ως αλγόριθμο των  $Linde-Buzo-Gray$  ( $LBG$ ) [23]. Σε κάθε επανάληψη κάθε χρωματικό διάνυσμα (pixel) ανατίθεται στην πιο κοντινή ομάδα χρωμάτων σύμφωνα με βάση την απόστασή του από το χρωματικό διάνυσμα μέσης τιμής της ομάδας, και ακολουθεί η επανενημέρωση του χρωματικού διανύσματος μέσης τιμής της ομάδας αυτής. Ο αλγόριθμος αυτός τερματίζει μετά από σύγκλιση. Ακολουθεί η συγχώνευση των χρωματικών ομάδων σύμφωνα με μία προκαθορισμένη μέγιστη απόσταση. Τελικά σε κάθε σημείο της αρχικής εικόνας δίδεται η μέση τιμή της ομάδας χρωμάτων στην οποία αυτό θα ανήκει μετά το πέρας του αλγορίθμου.

## 2.1.2 Ταξινόμηση του χρώματος του ανθρώπινου δέρματος

Μετά τον κβαντισμό της αρχικής εικόνας το δεύτερο στάδιο αυτής της φάσης έχει στόχο την ταξινόμηση των κυρίαρχων χρωμάτων σε χρώματα που ανήκουν στο χρώμα του ανθρώπινου δέρματος και σε αυτά που δεν ανήκουν. Για το σκοπό αυτό χρησιμοποιήθηκε ένα σύνολο εκμάθησης κομματιών-προτύπων του χρώματος του δέρματος τα οποία συλλέχθηκαν από διάφορες έγχρωμες ψηφιακές εικόνες και σκηνές από βίντεο, αριθμούν τα 950 και καλύπτουν ένα ευρύ φάσμα του ζητούμενου χρώματος (διάφορες φυλές ανθρώπων, διάφορες συνθήκες φωτεινότητας). Στο σημείο αυτό πρέπει να επισημανθεί ότι χρησιμοποιήθηκαν δύο χρωματικά συστήματα ( $YCbCr$ ,  $HSV$ ) ενώ παράλληλα οι σχετικές έρευνες επεκτείνονται και σε άλλα ( $Lab$ ,  $Luv$ ). Στο Σχήμα 2.1 παραδείγματα του χρώματος του ανθρώπινου δέρματος αναπαρίστανται στα συστήματα  $YCbCr$  και  $HSV$ . Μπορεί εύκολα να παρατηρηθεί ότι αποτελούν ένα αρκετά συμπαγές σύνολο και στα δύο συστήματα. Σκοπός του σταδίου αυτού είναι να προσεγγιστεί το τρισδιάστατο περίβλημα αυτών των συνόλων.



Σχήμα 2.1: Κατανομές παραδειγμάτων του χρώματος του ανθρώπινου δέρματος στο χρωματικό σύστημα (α)  $YCbCr$  και (β)  $HSV$ .

Ξεκινώντας από το σύστημα  $YCbCr$  παρατηρούμε ότι η συνιστώσα  $Y$  (που αποτελεί την τιμή της έντασης) δεν επηρεάζει την κατανομή στο επίπεδο που ορίζουν οι  $Cb$ ,  $Cr$  συνιστώσες. Μάλιστα τα παραδείγματα σχηματίζουν ένα μικρό και αρκετά συμπαγές σύνολο στο επίπεδο  $CbCr$ . Είναι δυνατόν να γίνει ταξινόμηση του χρώματος του ανθρώπινου δέρματος άμεσα σ' αυτό το επίπεδο (βλ. και [36]) χωρίς να ληφθεί υπόψη η τιμή της έντασης  $Y$ . Στην υλοποίησή μας χρησιμοποιείται και η συνιστώσα  $Y$ , έτσι ώστε να μπορεί να αποδίδει και κάτω από έντονες συνθήκες φωτισμού, αφού παρατηρήθηκε ότι η κατανομή αποδεικνύεται τελικά διαφορετική για τα άκρα της έντασης  $Y$ , τα οποία αντίστοιχα είναι οι τιμές γύρω από το 50 (για το σκοτεινό) και γύρω από το 240 (για το πολύ φωτεινό). Δεδομένου του αρκετού θορύβου που περιέχεται στα παραδείγματα λόγω της ύπαρξης σκοτεινών περιοχών του προσώπου (γένεια, φρύδια, μαλλιά), αποφασίστηκε να προσεγγιστούν τα όρια του συνόλου των χρωμάτων που ανήκουν στο ανθρώπινο δέρμα χρησιμοποιώντας επίπεδα, τα οποία και μπορούν πολύ εύκολα να προσαρμοστούν σε αντίθεση με αυτόματες μεθόδους ομαδοποίησης. Τα επίπεδα αυτά βρέθηκαν με διαδοχικές ρυθμίσεις σύμφωνα με τα αποτελέσματα της κατάταξης. Οι εξισώσεις που ορίζουν τα επίπεδα που βρέθηκαν είναι οι :

$$Cr \geq -2(Cb + 24); Cr \geq -(Cb + 17); Cr \geq -4(Cb + 32); Cr \geq 2.5(Cb + \theta_1);$$

$$Cr \geq \theta_3; Cr \geq 0.5(\theta_4 - Cb); Cr \leq \frac{220 - Cb}{6}; Cr \leq \frac{4}{3}(\theta_2 - Cb);$$

$$\text{Av } (Y > 128) \theta_1 = -2 + \frac{256 - Y}{16}; \theta_2 = 20 - \frac{256 - Y}{16}; \theta_3 = 6; \theta_4 = -8;$$

$$\text{Av } (Y \leq 128) \theta_1 = 6; \theta_2 = 12; \theta_3 = 2 + \frac{Y}{32}; \theta_4 = -16 + \frac{Y}{16};$$



Σε αντίθεση με το χρωματικό χώρο  $YCbCr$ , το σύνολο του χρώματος του ανθρώπινου δέρματος είναι λιγότερο συμπαγές στο χρωματικό χώρο  $HSV$ . Σε λίγες έρευνες, όπως στην [34], χρησιμοποιούνται μόνο οι προβολές στο επίπεδο  $HS$  και η κατάταξη γίνεται χρησιμοποιώντας κατάλληλα κατώφλια. Αυτά όμως αποδεικνύονται ανίκανα να αντιμετωπίσουν αποκλίσεις στις συνθήκες φωτεινότητας της εικόνας. Χρησιμοποιήθηκε η ίδια προσέγγιση που εφαρμόστηκε στο χρωματικό χώρο  $YCbCr$  και οι εξισώσεις που ορίζουν τα επίπεδα που περικλείουν το σύνολο του χρώματος του ανθρώπινου δέρματος είναι οι :

$$S \geq 10; V \geq 40; S \leq -H - 0.1V + 110; H \leq -0.4V + 75;$$

$$\text{αν } (H \geq 0) S \leq 0.08(100 - V)H + 0.5V; \text{ αλλιώς } S \leq 0.5H + 35;$$

Τα αποτελέσματα ταξινόμησης είναι αρκετά ισοδύναμα και στους δύο χρωματικούς χώρους, όπως έδειξαν εξαντλητικά πειράματα. Παρατηρήθηκε όμως ότι παρόλο που το σύνολο του χρώματος του ανθρώπινου δέρματος είναι πιο συμπαγές στο επίπεδο  $CbCr$  απ' ό,τι στο  $HS$ , τα προαναφερθέντα επίπεδα προσαρμόζονται πολύ πιο εύκολα στο μοντέλο  $HSV$  λόγω της άμεσης πρόσβασης στο  $H$  (*Hue*), που κυρίως κωδικοποιεί τη διάκριση ως προς το χρώμα.

Στο Σχήμα 2.2 παρατίθενται παραδείγματα εικόνων όπου εξάχθηκαν τα κυρίαρχα χρώματα και στη συνέχεια ακολούθησε η ταξινόμηση του χρώματος του ανθρώπινου δέρματος. Φαίνεται η αποτελεσματικότητα του σταδίου αυτού, αλλά επίσης και η ανεπάρκειά του, αφού μπορεί να υπάρχουν τμήματα του φόντου με χρώμα παρόμοιο του χρώματος του ανθρώπινου δέρματος. Έτσι απαιτούνται κι άλλα στάδια επεξεργασίας που βασίζονται σε ανθρωπομορφικά κριτήρια και ανάλυση της υφής του προσώπου.



Σχήμα 2.2: Αρχικές εικόνες (πρώτη γραμμή), κυρίαρχα χρώματα (δεύτερη γραμμή), τμηματοποίηση ως προς το χρώμα του ανθρώπινου δέρματος ( τρίτη γραμμή).

## 2.2 Δίκτυο διανυσμάτων υποστήριξης

Σε αντίθεση με την προηγούμενη η μέθοδος αυτή εκπαιδεύει ένα νευρωνικό δίκτυο (βλ. και [15]) για την απευθείας κατάταξη του χρώματος του ανθρώπινου δέρματος. Έγινε προσπάθεια να τμηματοποιηθούν οι περιοχές των χρωμάτων στα διάφορα χρωματικά συστήματα σε περιοχές που ανήκουν σε πρόσωπο και μη. Το νευρωνικό δίκτυο που επιλέχθηκε γι' αυτό το σκοπό στηρίζεται στις λεγόμενες μηχανές διανυσμάτων υποστήριξης (*support vector machines*) για λόγους που αναφέρονται παρακάτω.

## 2.2.1 Επιλογή και αρχιτεκτονική του νευρωνικού δικτύου

Σε μη ελεγχόμενες εικόνες, χωρίς δηλαδή σκηνηικούς περιορισμούς, με γενικά πολύπλοκο φόντο και άγνωστη φωτεινότητα, η περιοχή των χρωμάτων που ανήκει στο χρώμα του ανθρώπινου δέρματος μπορεί γενικά να είναι αρκετά δυσδιάκριτη σε σχέση με την περιοχή των χρωμάτων του φόντου. Σκοπός, λοιπόν, του νευρωνικού δικτύου που θα επιλεγεί γι' αυτήν την εργασία πρέπει να είναι ο βέλτιστος διαχωρισμός των δύο κλάσεων (χρώμα προσώπου και μη χρώμα προσώπου). Το δίκτυο των διανυσμάτων υποστήριξης παρέχει τη βέλτιστη αυτή λύση, χωρίς να δημιουργείται κάποιο πρόβλημα, αφού στην αρχιτεκτονική του δικτύου δεν λαμβάνεται υπόψη η τοπική γειτνίαση των δεδομένων. Τα παραδοσιακά δίκτυα, όπως αυτό του *multilayer perceptron*, στηρίζουν την εκπαίδευσή τους στην ελαχιστοποίηση του σφάλματος καθώς προσπαθούν να βελτιώσουν την επίδοσή τους στο σύνολο εκπαίδευσης. Αντιθέτως, το δίκτυο διανυσμάτων υποστήριξης προσπαθεί να βρει τη βέλτιστη λύση γεωμετρικά, μειώνοντας την πιθανότητα λαθεμένης κατάταξης (ελαχιστοποίηση του δομικού ρίσκου). Το δίκτυο καταφέρνει να συμπυκνώσει όλη την πληροφορία του συνόλου εκπαίδευσης σε λίγα δεδομένα, τα λεγόμενα διανύσματα υποστήριξης (*support vectors*). Έτσι σχηματίζεται μια οικογένεια επιφανειών σχετικά μικρής πολυπλοκότητας. Το δίκτυο προσπαθεί επαναληπτικά να βρει την ελάχιστη απόσταση μεταξύ των δύο κλάσεων, ελαχιστοποιώντας παράλληλα την πιθανότητα λάθους κατάταξης. Αυτό επιτυγχάνεται με την εύρεση των κατάλληλων διανυσμάτων υποστήριξης.

Τα διανύσματα υποστήριξης, όπως αναφέρθηκε, αποτελούν μία συμπυκνωμένη μορφή ολόκληρου του συνόλου των δεδομένων εκπαίδευσης ( $x_i \in \mathbb{R}^n$ ). Χρησιμοποιώντας μόνο την πληροφορία που περιέχουν τα διανύσματα αυτά, το νευρωνικό δίκτυο εκπαιδεύεται στην εύρεση επιφανειών για το διαχωρισμό των δύο κλάσεων ( $y_i \in \{-1, 1\}$ , με  $-1$  να αποτελεί την κλάση μη χρώμα προσώπου και  $1$  την κλάση χρώμα προσώπου). Στην περίπτωση που τα δεδομένα είναι γραμμικά διαχωρίσιμα, ή τουλάχιστον υποθέτουμε ότι είναι, ισχύει :

$$w^T x_i + b \leq -1, \text{ αν } y_i = -1, \quad w^T x_i + b \geq 1, \text{ αν } y_i = 1 \Rightarrow y_i(w^T x_i + b) \geq 1 \quad (2.1)$$

όπου  $w$  το βάρος των νευρώνων του δικτύου και  $b$  η προκατάληψη (bias), ενώ η διαχωριστική επιφάνεια των δύο κλάσεων δίδεται από τη σχέση  $w^T x + b = 0$ . Η απόσταση επομένως ενός σημείου  $x_i$  από το επίπεδο διαχωρισμού δίνεται από τη σχέση  $d_i = \frac{w^T x_i + b}{\|w\|}$  ή ισοδύναμα  $y_i d_i \geq \frac{1}{\|w\|}$ , απ' όπου βλέπουμε ότι η ελάχιστη απόσταση μεταξύ των σημείων  $x_i$  και του επιπέδου διαχωρισμού είναι  $\frac{1}{\|w\|}$ . Έχοντας υπόψη τα παραπάνω μπορούμε να πούμε ότι η βέλτιστη επιφάνεια διαχωρισμού είναι το υπερεπίπεδο που μεγιστοποιεί την απόσταση των κοντινότερων σημείων προς αυτό. Επειδή η απόσταση των κοντινότερων σημείων είναι  $\frac{1}{\|w\|}$ , μπορούμε να πούμε ότι το πρόβλημα ανάγεται στην ελαχιστοποίηση του όρου  $\frac{1}{2} w^T w$ , υπό τον περιορισμό  $y_i(w^T x_i + b) \geq 1$ , για  $i = 1, 2, \dots, N$ . Το παραπάνω πρόβλημα μπορεί να λυθεί με τη μέθοδο των πολλαπλασιαστών Lagrange  $a = (a_1, a_2, \dots, a_N)$ . Ισοδύναμα το πρόβλημα γράφεται ως

$$\Lambda(w, b, a) = \frac{1}{2} w^T w - \sum_{i=1}^N a_i \{y_i(w^T x_i + b) - 1\}$$

ζητώντας να βρούμε το ελάχιστο της συνάρτησης. Παραγωγίζοντας ως προς  $b$  και  $w$  προκύπτει η σχέση

$$\Lambda(a) = \sum_{i=1}^N a_i - \frac{1}{2} \sum_{i=1}^N a_i a_j y_i y_j x_i^T x_j$$

την οποία προσπαθούμε να μεγιστοποιήσουμε ως προς  $a$ . Έτσι καταλήγουμε στη μεγιστοποίηση του όρου

$$-\frac{1}{2} a^T D a + \sum_{i=1}^N a_i \quad \text{υπό τον περιορισμό} \quad \sum_{i=1}^N a_i y_i = 0, a \geq 0,$$

που αποτελεί και το δυαδικό ισοδύναμο του προηγούμενου. Ο  $D$  είναι ένας  $N \times N$  πίνακας με στοιχεία  $D_{ij} = y_i y_j x_i^T x_j$ . Παραγωγίζοντας την  $\Lambda$  ως προς  $w$  προκύπτει ότι

$$w = \sum_{i=1}^N a_i y_i x_i, \quad i = 1, \dots, N,$$

ενώ το  $b$  προκύπτει από τη σχέση  $a_i \{y_i(w^T x_i + b) - 1\} = 0, i = 1, \dots, N$ . Διανύσματα υποστήριξης είναι τα  $x_i$  για τα οποία ισχύει η παραπάνω σχέση, και  $a_i \geq 0$ . Το πρόβλημα της κατάταξης των σημείων περιορίζεται απλά στην εύρεση του προσήμου της παράστασης  $w^T x + b$ .

Από την άλλη μεριά, αν δεν ισχύει η διαχωριστικότητα των δύο κλάσεων, το προηγούμενο παράδειγμα γενικεύεται εισάγοντας  $N$  μη αρνητικές μεταβλητές  $\xi = (\xi_1, \xi_2, \dots, \xi_N)$  έτσι ώστε

$$y_i(w^T x_i + b) \geq 1 - \xi_i, i = 1, 2, \dots, N \quad (2.2)$$

Αν ένα σημείο  $x_i$  ικανοποιεί την ανισότητα 2.1, τότε το  $\xi_i$  είναι 0 και η 2.2 καταλήγει στην 2.1. Διαφορετικά στην 2.1 προσθέτουμε τον όρο  $-\xi_i$  οπότε καταλήγουμε στην 2.2. Η επίλυση του προβλήματος ανάγεται στην ελαχιστοποίηση του όρου

$$\frac{1}{2} w^T w + C \sum_{i=1}^N \xi_i, \text{ υπό τον περιορισμό } y_i(w^T x_i + b) \geq 1 - \xi_i,$$

για  $i = 1, 2, \dots, N$  και  $\xi_i \geq 0$ . Ο δεύτερος όρος της πρώτης συνθήκης μας δίνει ένα μέτρο σύγκρισης των συνολικών λαθών κατάταξης και οδηγεί στην εύρεση μίας διαχωριστικής επιφάνειας που είναι λιγότερο ευαίσθητη στην ύπαρξη θορύβου στο σύνολο εκπαίδευσης. Η παράμετρος  $C$  μπορεί να θεωρηθεί ως όρος εξομάλυνσης. Σε αναλογία με το προηγούμενο πρόβλημα των διαχωρίσιμων κλάσεων το δυαδικό ισοδύναμο πρόβλημα των μη διαχωρίσιμων κλάσεων είναι η μεγιστοποίηση του όρου

$$-\frac{1}{2} a^T D a + \sum_{i=1}^N a_i, \text{ υπό τον περιορισμό } \sum_{i=1}^N a_i y_i = 0, 0 \leq a_i \leq C, i = 1, 2, \dots, N.$$

Ανάλογα έχουμε

$$w = \sum_{i=1}^N a_i y_i x_i, i = 1, \dots, N,$$

ενώ το  $b$  μπορεί να υπολογιστεί από τις υποθέσεις

$$a_i(y_i(w^T x_i + b) - 1 + \xi_i) = 0 \quad (2.3)$$

και

$$(C - a_i)\xi_i = 0 \quad (2.4)$$

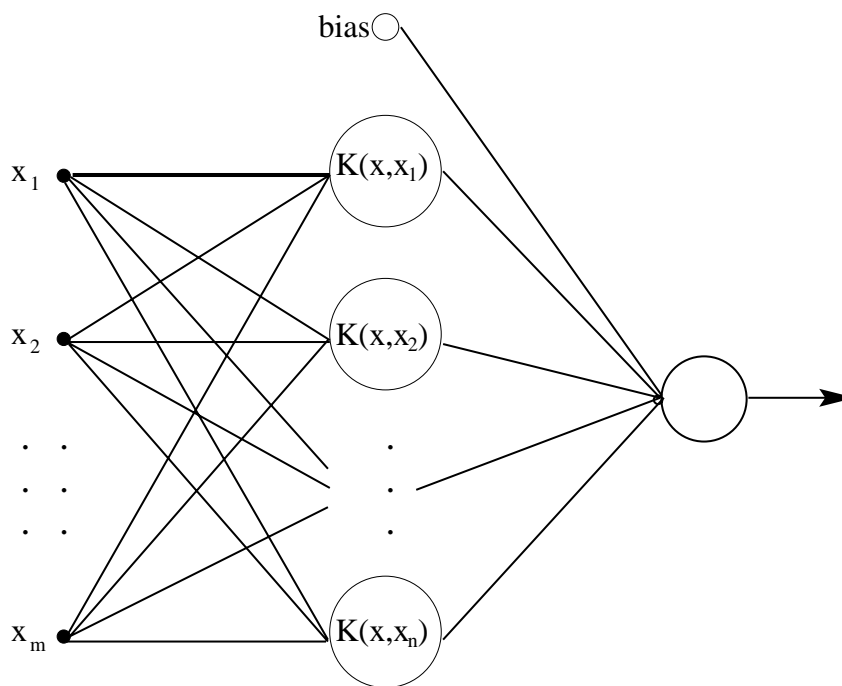
Από τους υπολογισμούς των  $a_i$  και  $\xi_i$  μπορούμε να εξάγουμε πληροφορίες για τα διανύσματα υποστήριξης. Αν  $a_i < C$  από την Εξίσωση (2.4) προκύπτει ότι  $\xi_i = 0$ , και από την 2.3 προκύπτει ότι το διάνυσμα υποστήριξης βρίσκεται σε απόσταση  $\frac{1}{\|w\|}$  από το επίπεδο διαχωρισμού. Αυτά τα διανύσματα ονομάζονται margin vectors. Τέλος, τα διανύσματα για τα οποία ισχύει  $a_i = C$ , βρίσκονται λάθος κατεταγμένα εάν  $\xi_i > 1$ , ενώ απέχουν απόσταση μικρότερη από  $\frac{1}{\|w\|}$  από το επίπεδο διαχωρισμού αν  $0 < \xi_i \leq 1$ .

Στις περισσότερες περιπτώσεις ο γραμμικός διαχωρισμός των δεδομένων δεν έχει μεγάλη πρακτική εφαρμογή. Αντιθέτως, οι μη γραμμικές επιφάνειες διαχωρισμού είναι αποδοτικότερες. Για την επίλυση του προβλήματος του μη γραμμικού διαχωρισμού απαιτείται ο μετασχηματισμός των δεδομένων ( $x_i \in \mathbb{R}^n$ ) σε ένα χώρο Hilbert μέσω μιας συνάρτησης  $\phi$ . Έχοντας μετασχηματίσει τα δεδομένα, το πρόβλημα αντιμετωπίζεται όπως στην προηγούμενη περίπτωση, αντικαθιστώντας απλά τα στοιχεία του πίνακα  $D$  με αυτά της εξίσωσης  $D_{ij} = y_i y_j \phi^T(x_i) \phi(x_j)$ , απ' όπου και βρίσκονται τα  $(w, b)$ . Αντίστοιχα η εξίσωση

$$w^T \phi(x) + b = \sum_{i=1}^N a_i y_i \phi^T(x_i) \phi(x) + b = 0 \quad (2.5)$$

δίνει την επιφάνεια διαχωρισμού. Το εσωτερικό γινόμενο  $\phi^T(x_i) \phi(x_j) = K(x_i, x_j)$  ονομάζεται πυρήνας. Παραδείγματα πυρήνων είναι αυτό του πολυωνυμικού  $K(x, y) = (1 + x^T y)^d$  και του γκαουσιανού  $K(x, y) = e^{-\frac{\|x-y\|^2}{2\sigma^2}}$ . Οι δύο παραπάνω πυρήνες επιλέχτηκαν επειδή πληρούν τα κριτήρια του θεωρήματος Mercer.

Για να σχεδιαστεί ένα δίκτυο του τύπου *radial-basis function* ή *multilayer perceptron* απαιτούνται συνήθως πειράματα και ευρηματικές ιδέες για τον αριθμό των κρυφών επιπέδων και των κόμβων που πρέπει να έχει το δίκτυο ώστε να διαχωρίζει βέλτιστα τα δεδομένα. Αντιθέτως στο δίκτυο διανυσμάτων υποστήριξης, με την εύρεση των πολλαπλασιαστών Lagrange και τα αντίστοιχα διανύσματα υποστήριξης έχουμε ταυτόχρονα και την αρχιτεκτονική του δικτύου όπως φαίνεται και στο Σχήμα 2.3.



Είσοδος του $m$ -διάστατου διανύσματος	Το κρυφό στρώμα των $n$ πυρήνων που αντιστοιχούν στα $n$ διανύσματα υποστήριξης	Η έξοδος ισούται με το γραμμικό συνδυασμό των πολλαπλασιαστών Lagrange με τους πυρήνες
--	---	--

Σχήμα 2.3: Αρχιτεκτονική νευρωνικού δικτύου με διανύσματα υποστήριξης.

Ένα άλλο χαρακτηριστικό του δικτύου που το κάνει να ξεχωρίζει από τα προηγούμενα, είναι ότι η πολυπλοκότητά του μπορεί εύκολα να ελεγχθεί θέτοντας συγκεκριμένους περιορισμούς ως παραμέτρους στο πρόβλημα βελτιστοποίησης. Αντιθέτως στα άλλα δίκτυα, η πολυπλοκότητα ελέγχεται κρατώντας τον αριθμό των κρυφών στρωμάτων και κόμβων μικρό. Οι παράμετροι που καθορίζουν την πολυπλοκότητα είναι ο όρος εξομάλυνσης  $C$  όπως παρουσιάζεται στο πρόβλημα της εύρεσης των διανυσμάτων υποστήριξης και ειδικά για την περίπτωση των γκαουσιανών πυρήνων η διασπορά  $\sigma^2$ . Γενικά όσο πιο μικρή είναι η παράμετρος  $C$  τόσο πιο απλό είναι το δίκτυο (μικρή διάσταση Vapnik-Chervonenkis  $VCdim$ ), ενώ για μεγάλες τιμές τείνει να υπερεξειδικευτεί (μεγάλη διάσταση  $VCdim$ ). Για μικρές τιμές του  $C$  το δίκτυο βρίσκει μεγάλο περιθώριο (margin) μεταξύ των κλάσεων με πολλές λάθος κατατάξεις, ενώ για μεγάλες τιμές το περιθώριο γίνεται όλο και πιο μικρό ελαχιστοποιώντας τις λάθος κατατάξεις. Τα παραπάνω φαίνονται στον πίνακα 2.1 για το σύνολο εκπαίδευσης του χρωματικού συστήματος  $Luv$  που αποτελείται από 2520 δεδομένα.

Όσον αφορά τη διασπορά  $\sigma^2$  παρατηρούμε ότι όσο πιο μεγάλη είναι τόσο πιο λίγα σφάλματα έχουμε σε συνδυασμό με ένα λογικό αριθμό από διανύσματα υποστήριξης. Το  $\sigma^2$  εκφράζει τη διασπορά της κατανομής, επομένως για μικρά  $\sigma^2$  το δίκτυο τείνει να υπερεξειδικευτεί στα δεδομένα (κάθε δεδομένο να προσεγγίζεται και από μία συνάρτηση) με αποτέλεσμα στη γενίκευση να υπάρχει σημαντικό πρόβλημα για δεδομένα που απέχουν πολύ από τα δεδομένα εκπαίδευσης του δικτύου. Τα παραπάνω φαίνονται στον πίνακα 2.2 πάλι με το ίδιο σύνολο δεδομένων όπως και παραπάνω.

Πρέπει να επισημανθεί σ' αυτό το σημείο, ότι ο όρος  $VCdim$  δίνει ένα μέτρο δυνατότητας του

$C$	$\frac{1}{2\sigma^2}$	Σφάλματα	Διανύσματα υποστήριξης	$VCdim$	Επαναλήψεις
10	0.001	511	1186	817	1
100	0.001	491	1148	12400	1
500	0.001	421	1097	84871	1
1000	0.001	416	1068	148766	1

Πίνακας 2.1: Επίδραση του όρου εξομάλυνσης  $C$  στον αριθμό των σφαλμάτων, των διανυσμάτων υποστήριξης και την πολυπλοκότητα του δικτύου

$C$	$\frac{1}{2\sigma^2}$	Σφάλματα	Διανύσματα υποστήριξης	$VCdim$	Επαναλήψεις
10	0.1	93	1487	12293	1
100	0.01	299	909	14474	1
500	0.001	491	1148	12400	1
1000	0.0001	544	1256	2919	1

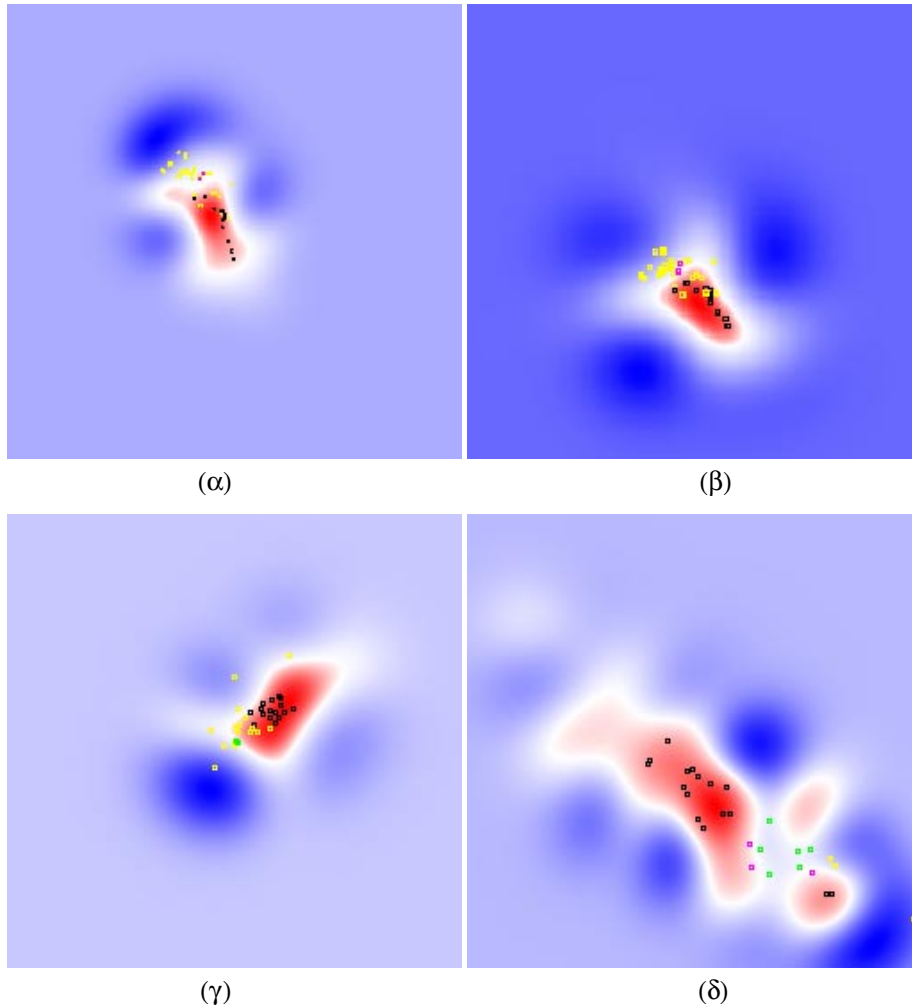
Πίνακας 2.2: Επίδραση της διασποράς στον αριθμό των σφαλμάτων, των διανυσμάτων υποστήριξης και την πολυπλοκότητα του δικτύου

δικτύου να κατατάσει σωστά τα δεδομένα. Η παράμετρος αυτή είναι ίση με το μέγιστο αριθμό δεδομένων από το σύνολο εκπαίδευσης που μπορούν να μαθευτούν από το δίκτυο χωρίς κανένα λάθος και για όλες τις πιθανές δυαδικές τιμές των συναρτήσεων ταξινόμησης (classification functions).

## 2.2.2 Επιφάνειες διαχωρισμού για την ταξινόμηση

Στην προηγούμενη ενότητα αναφέρθηκαν οι λόγοι για τους οποίους επιλέχθηκε το συγκεκριμένο νευρωνικό δίκτυο. Ο κυριότερος είναι ότι καταφέρνει να συμπυκνώσει όλη την πληροφορία του συνόλου δεδομένων εκπαίδευσης σε λίγα μόνο δεδομένα, σχηματίζοντας έτσι μία οικογένεια επιφάνειων σχετικά μικρής πολυπλοκότητας. Στην απλοποίηση των επιφανειών πολύ σημαντικό ρόλο μπορεί να έχει και η επιλογή του χρωματικού συστήματος που θα χρησιμοποιηθεί. Στη συγκεκριμένη έρευνα δοκιμάστηκαν τα χρωματικά συστήματα  $RGB$ ,  $HSV$ ,  $YCbCr$ ,  $Luv$  και  $Lab$ . Από τις δοκιμές που έγιναν συμπερασματικά θα μπορούσαμε να πούμε ότι πιο απλές ήταν οι επιφάνειες που προέκυψαν από τα συστήματα  $Luv$  και  $Lab$ . Ακολουθούν αυτές του  $YCbCr$ , του  $HSV$  και τέλος αυτή του  $RGB$ . Το αποτέλεσμα για τα τρία πρώτα χρωματικά συστήματα ήταν κάπως αναμενόμενο αφού σε αυτά, τα χρώματα είναι αποσυσχετισμένα μεταξύ τους κάτι το οποίο δεν ισχύει στα συστήματα  $HSV$  και  $RGB$ . Το  $Lab$  και το  $Luv$  είναι καλύτερα από το  $YCbCr$  για το λόγο ότι οι περιοχές προσώπων καταλαμβάνουν μικρότερη έκταση στην πρώτη συνιστώσα  $L$  με συνεκτικότερες επιφάνειες απ' ότι σε αυτή του  $YCbCr$  στην συνιστώσα  $Y$ . Η εικόνα του Σχήματος 2.4(α) αντιστοιχεί στη διαχωριστική επιφάνεια του  $Luv$  συστήματος για  $L = 70$ . Οι γραμμές της εικόνας απεικονίζουν τον άξονα  $u$ , ενώ οι στήλες τον άξονα  $v$ . Τα μαύρα σημεία ανήκουν σε πρόσωπα ενώ τα κίτρινα σε περιοχές του φόντου. Τα μωβ είναι τα θετικά διανύσματα υποστήριξης, τα πράσινα είναι τα αρνητικά. Τα κίτρινα σημεία, τα οποία βρίσκονται στην περιοχή των χρωμάτων του προσώπου, είναι σημεία που επαναληπτικά αφαιρούνται από το σύνολο εκπαίδευσης, χωρίς να λαμβάνουν μέρος στην διαδικασία της εκμάθησης του δικτύου. Το άνω αριστερά σημείο της εικόνας αντιστοιχεί στο  $(\min u, \min v)$ , ενώ το κάτω δεξιά στο  $(\max u, \max v)$ . Επίσης η επιφάνεια αυτή προέκυψε ύστερα από εκπαίδευση 15 επαναλήψεων με παραμέτρους  $C = 100$  και  $\frac{1}{2\sigma^2} = 0.001$ . Το ίδιο ισχύει και για τα άλλα χρωματικά συστήματα των οποίων οι διαχωριστικές επιφάνειες απεικονίζονται στο Σχήμα 2.4 αντίστοιχα στις εικόνες (β), (γ) και (δ). Γενικά για μικρές τιμές της πρώτης συνιστώσας (αντίστοιχα  $L$ ,  $L$ ,  $Y$ ,  $R$ ) οι επιφάνειες δεν είναι συνεκτικές. Αυτό μεταφράζεται στην ύπαρξη κενών σε περιοχές προσώπων. Αυτά όμως τα επίπεδα είναι λίγα στα συστήματα  $Lab$  και  $Luv$ . Αντιθέτως στα συστήματα  $YCbCr$  και  $RGB$ , όπου οι περιοχές προσώπων εκτείνονται σε μεγάλο εύρος στην  $Y$  και  $R$  συνιστώσα, δημιουργείται πρόβλημα

στον προσδιορισμό ενός απλού γεωμετρικού τόπου.



Σχήμα 2.4: Οι διαχωριστικές επιφάνειες για τα χρωματικά συστήματα (α)  $Luv$  ( $L = 70$ ,  $u$  γραμμές και  $v$  στήλες), (β)  $Lab$  ( $L = 70$ ,  $a$  γραμμές και  $b$  στήλες), (γ)  $YCbCr$  ( $Y = 170$ ,  $Cb$  γραμμές και  $Cr$  στήλες), (δ)  $RGB$  ( $R = 170$ ,  $G$  γραμμές και  $B$  στήλες).

Για την εκπαίδευση του νευρωνικού δικτύου χρησιμοποιήθηκε μία βάση δεδομένων από 2000 δείγματα, δηλαδή χρωματικά διανύσματα. Από αυτά τα 1000 ήταν θετικά, που σημαίνει είχαν το χρώμα του ανθρώπινου δέρματος, δηλαδή σημεία προσώπου, και 1000 αρνητικά, σημεία του φόντου. Κατά την εκπαίδευση του δικτύου σε κάθε επανάληψη αφαιρούνταν δεδομένα τα οποία χαρακτηρίστηκαν ως δεδομένα προσώπου ενώ στην πραγματικότητα δεν ήταν. Αυτή η προσπάθεια έγινε αφενός για να ελαττωθεί ο αριθμός των διανυσμάτων υποστήριξης και αφετέρου για να απλοποιηθεί η επιφάνεια διαχωρισμού. Αναλυτικά στατιστικά στοιχεία παρουσιάζονται στον πίνακα 2.3.

Με το μικρότερο αριθμό διανυσμάτων υποστήριξης οι πράξεις λιγοστεύουν προκειμένου να ληφθεί η απόφαση για το αν ένα σημείο ανήκει σε περιοχή προσώπου ή όχι. Ενώ απλοποιώντας την επιφάνεια μπορούμε να οδηγηθούμε σε ένα γεωμετρικό τόπο έτσι ώστε σε χρόνο  $O(1)$  να μπορεί να αποφασιστεί αν ένα σημείο βρίσκεται μέσα σε αυτόν ή όχι. Αφαιρώντας αρνητικά παραδείγματα μέσα από θετικές περιοχές τείνουμε να κάνουμε τις δύο κλάσεις διαχωρίσιμες εξαλείφοντας επαναληπτικά δύσκολα δεδομένα που αυξάνουν την πολυπλοκότητα του δικτύου. Με αυτόν τον τρόπο το σύστημα υπολογισμού γενικεύεται πιο πολύ χωρίς να χάνει πρόσωπα σε εικόνες, με μειονέκτημα όμως να χαρακτηρίζει ως πρόσωπα και ορισμένες περιοχές του φόντου.

$C$	$\frac{1}{2\sigma^2}$	Σφάλματα	Διανύσματα υποστήριξης	$VCdim$	Επαναλήψεις
100	0.001	491	1148	12400	1
100	0.001	331	840	19900	2
100	0.001	280	790	27487	3
100	0.001	220	732	27708	5
100	0.001	103	451	22886	10
100	0.001	39	250	14724	15

Πίνακας 2.3: Επίδραση του αριθμού επαναλήψεων στον αριθμό των σφαλμάτων, των διανυσμάτων υποστήριξης και την πολυπλοκότητα του δικτύου

Στις φυσικές εικόνες υπάρχουν πολλά χρώματα που απέχουν πολύ από τις επιφάνειες διαχωρισμού που καθορίζει το δίκτυο. Επομένως πολλά σημεία της εικόνας μπορούν να χαρακτηριστούν ως μη πρόσωπα με λίγους μονάχα ελέγχους, χωρίς να χρειάζονται οι πολύπλοκοι υπολογισμοί του νευρωνικού δικτύου για την ίδια απόφαση. Για να αποφευχθεί λοιπόν η μεγάλη υπολογιστική πολυπλοκότητα του νευρωνικού δικτύου βρέθηκε ένα ορθογώνιο παραλληλεπίπεδο που περικλείει την τρισδιάστατη επιφάνεια την οποία καθορίζει το νευρωνικό. Έτσι, με έναν απλό προέλεγχο όσα σημεία βρίσκονται έξω από αυτό χαρακτηρίζονται κατευθείαν ως αρνητικά, ενώ για τα υπόλοιπα που βρίσκονται εντός του παραλληλεπίπεδου χρησιμοποιείται το δίκτυο για την κατάταξή τους. Με την ίδια λογική μπορεί να βρεθεί ένα γεωμετρικό σχήμα το οποίο να βρίσκεται ολόκληρο μέσα στην τρισδιάστατη επιφάνεια που ορίζει το δίκτυο και να χαρακτηρίζονται απευθείας με απλούς ελέγχους τα πολύ κοντινά σε πρόσωπα σημεία, χωρίς να είναι αναγκαίο και πάλι το δίκτυο. Τα σημεία τα οποία καλείται να κατατάξει το νευρωνικό δίκτυο θα βρίσκονται πλέον σ' ένα φλοιό που σχηματίζουν το εξωτερικό γεωμετρικό σχήμα και το εσωτερικό.

Σε γενικές γραμμές τα αποτελέσματα της εφαρμογής του νευρωνικού δικτύου είναι περίπου ίδια για όλα τα χρωματικά συστήματα με ελαφρώς καλύτερα χρησιμοποιώντας  $Luv$  και  $Lab$ . Στο Σχήμα 2.5 παρατίθενται αποτελέσματα κατάταξης του χρώματος του ανθρώπινου δέρματος με χρήση του νευρωνικού δικτύου στα χρωματικά συστήματα  $Luv$ ,  $Lab$ ,  $YCbCr$  και  $RGB$ , με παράμετρος  $C = 100$ ,  $\frac{1}{2\sigma^2} = 0.001$  και με 15 επαναλήψεις. Στην πρώτη στήλη παρατίθενται οι αρχικές εικόνες και στις επόμενες ακολουθούν τα αποτελέσματα κατάταξης στα αντίστοιχα χρωματικά συστήματα. Στα αποτελέσματα αυτά οι άσπρες περιοχές είναι θετικά καταταγμένες από το νευρωνικό, οι μαύρες περιοχές είναι αρνητικά καταταγμένες, ενώ οι σκουροπράσινες περιοχές είναι περιοχές μη προσώπων που αποφασίζονται από τους ελέγχους σύμφωνα με το προαναφερθέν ορθογώνιο παραλληλεπίπεδο.

Τα αποτελέσματα της κατάταξης όπως προκύπτουν από την εφαρμογή του νευρωνικού δικτύου, είναι κατανοητό, όπως επισημάνθηκε και στην προηγούμενη μέθοδο ότι δεν επαρκούν για την ανίχνευση προσώπων. Πρέπει να εισαχθούν και άλλα κριτήρια που τελικά θα εγγυώνται βέλτιστες αποφάσεις και τα οποία περιγράφονται στα στάδια που ακολουθούν παρακάτω.





Σχήμα 2.5: Αποτελέσματα κατάταξης του χρώματος του ανθρώπινου δέρματος. Στην πρώτη στήλη φαίνονται οι αρχικές εικόνες και ακολουθούν στις επόμενες τα αποτελέσματα για τα χρωματικά συστήματα *Luv*, *Lab*, *YCbCr*, *RGB* αντίστοιχα.



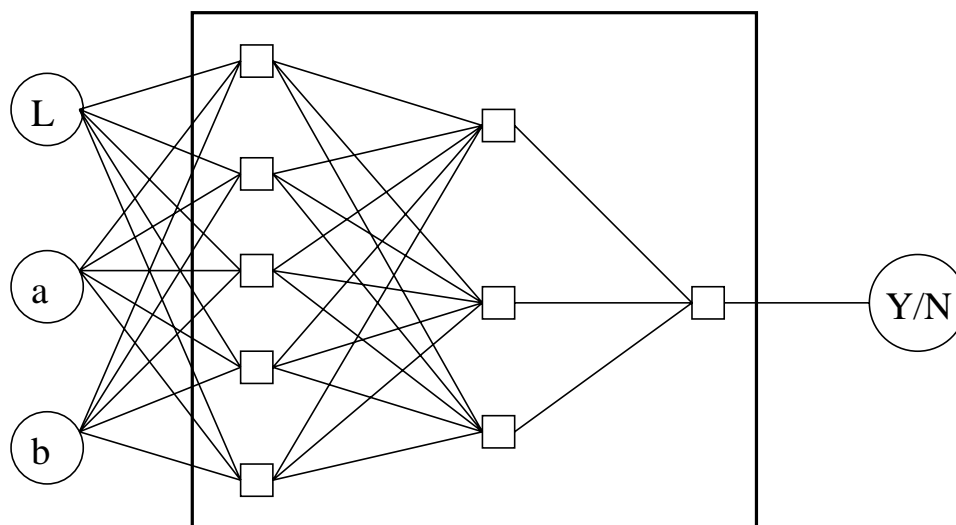
## 2.3 Πολυστρωματικό νευρωνικό δίκτυο

Η μέθοδος αυτή χρησιμοποιεί επίσης ένα νευρωνικό δίκτυο για να κατατμήσει το χρώμα του ανθρώπινου δέρματος από την εικόνα. Αυτήν τη φορά το νευρωνικό χρησιμοποιεί την κλασική μέθοδο back-propagation για τη λήψη της απόφασης σε αντίθεση με το νευρωνικό που περιγράφεται στην υποενότητα 2.2, και το οποίο είναι σαφώς πιο πολύπλοκο. Για την τμηματοποίηση επιλέχθηκε ο χώρος χρωμάτων *Lab* λόγω της ομοιομορφίας που παρουσιάζει, γεγονός που επιτρέπει την αποδοτική χρήση της ευκλείδειας απόστασης των χρωμάτων. Ακολουθεί η περιγραφή του νευρωνικού δικτύου, ο τρόπος λειτουργίας του και παρουσιάζονται αποτελέσματα τα οποία και μπορούν να χρησιμοποιηθούν ως μέτρο σύγκρισης με τις υπόλοιπες δύο μεθόδους που παρουσιάστηκαν.

### 2.3.1 Περιγραφή του νευρωνικού δικτύου

Όπως ήδη αναφέρθηκε, στη μέθοδος αυτή για τον προσδιορισμό των χρωμάτων προσώπου, χρησιμοποιήθηκε ένα πολυστρωματικό νευρωνικό δίκτυο εκπαιδευόμενο με τον αλγόριθμο back-propagation [15]. Όπως φαίνεται και στο Σχήμα 2.6, το νευρωνικό δίκτυο αποτελείται από τρία επίπεδα με αριθμό νευρώνων 5, 3, 1, σε κάθε επίπεδο. Η αρχιτεκτονική αυτή κρίθηκε ως η καταλληλότερη μετά από σειρά πειραμάτων με τα δεδομένα εκπαίδευσης του δικτύου. Η είσοδος του νευρωνικού είναι η τριπλέτα των χρωμάτων στο σύστημα *Lab*. Η έξοδος του νευρωνικού είναι η απόφαση για το αν το χρώμα της εισόδου είναι (Y) ή όχι (N) χρώμα προσώπου. Επίσης χρησιμοποιήθηκε και ένα δεύτερο κριτήριο απόφασης, πέρα από την ίδια την επιλογή του δικτύου, που διευρύνει κατά κάποιο λόγο αυτήν την επιλογή, και που θα περιγραφεί παρακάτω.

Το σύνολο εκπαίδευσης του δικτύου αποτελούν 2000 δεδομένα, από τα οποία τα 1000 είναι χρώματα προσώπου και τα υπόλοιπα 1000 είναι χρώματα που δεν ανήκουν σε πρόσωπα, αλλά αποτελούν διάφορες περιοχές του φόντου. Τα δεδομένα εξήχθησαν από μία βάση 100 εικόνων με πρόσωπα και οι εικόνες της οποίας χρησιμοποιήθηκαν γενικότερα για τον έλεγχο της απόδοσης των αλγορίθμων ανίχνευσης προσώπων.



Σχήμα 2.6: Back-propagation νευρωνικό δίκτυο προσδιορισμού χρωμάτων προσώπου.

### 2.3.2 Λειτουργία του δικτύου και τελικά κριτήρια.

Ο αλγόριθμος του back-propagation είναι γνωστός από τη θεωρία. Χρησιμοποιεί δύο περάσματα, όπου κατά το πρώτο το σήμα εισόδου περνάει από τους νευρώνες δίδοντας ένα διάλυμα απόκρισης σύμφωνα με τη συνάρτηση μεταφοράς *tansig*, η οποία δίνει αποκρίσεις που ανήκουν στο διάστημα  $[-1, 1]$ . Στο δεύτερο πέρασμα, σήματα λάθους, που υπολογίστηκαν στην έξοδο, σύμφωνα με μία

συνάρτηση κόστους  $E$ , μεταφέρονται πίσω και επιτρέπουν να διορθωθούν τα βάρη σε κάθε νευρώνα. Η έκδοση που επιλέχτηκε για το σκοπό μας χρησιμοποιεί και αδράνεια μαζί με μεταβλητό ρυθμό μάθησης. Έτσι οι αλλαγές στα βάρη γίνονται σύμφωνα με :

$$\Delta W = \alpha \cdot \Delta W_{\text{πρ}} + \rho \cdot \alpha \cdot \frac{\Delta E}{\Delta W} \quad (2.6)$$

όπου  $\Delta W$  είναι οι αλλαγές που πρέπει να πραγματοποιηθούν,  $\alpha$  η σταθερά της αδράνειας, που επιλέχτηκε 0.9,  $\Delta E$  είναι η συνάρτηση κόστους, που υπολογίζει το λάθος και που πρέπει να ελαχιστοποιηθεί (μετράει δηλαδή την απόδοση του δικτύου), και  $\Delta W_{\text{πρ}}$  είναι οι αλλαγές στα βάρη που έγιναν κατά το προηγούμενο πέρασμα. Ο ρυθμός μάθησης  $\rho$  έχει τεθεί 0.01, ενώ για κάθε εποχή αν η συνάρτηση κόστους μειώνεται ο ρυθμός αυξάνεται κατά τον παράγοντα 1.05. Στην αντίθετη περίπτωση, όταν δηλαδή η συνάρτηση κόστους αυξάνεται και μάλιστα περισσότερο από έναν παράγοντα 1.04, ο ρυθμός μάθησης ελαττώνεται κατά 0.7, και ταυτόχρονα η αλλαγή, που είχε ως αποτέλεσμα αυτήν την αύξηση της συνάρτησης κόστους, δεν πραγματοποιείται.

Αφού το νευρωνικό δίκτυο εκπαιδεύτηκε ώστε να συγκλίνει, στη συνέχεια χρησιμοποιείται για την κατάταξη κάθε σημείου της εικόνας προς επεξεργασία σε χρώμα ανθρώπινου δέρματος ή όχι. Όπως προαναφέρθηκε, η είσοδος του δικτύου αποτελείται από τις τρεις συνιστώσες του χρώματος στο χρωματικό σύστημα *Lab*. Χρησιμοποιήθηκαν δύο κριτήρια τελικής επιλογής. Κατά το πρώτο μία είσοδος θεωρείται θετική, δηλαδή ότι το σημείο εισόδου ανήκει σε χρώμα δέρματος, αν η έξοδος του δικτύου είναι μεγαλύτερη ή ίση του μηδενός, και ενδεικτικά αποτελέσματα παρουσιάζονται στο Σχήμα 2.7. Με το δεύτερο κριτήριο επιλογής ουσιαστικά διευρύνεται η απόκριση του δικτύου και μία είσοδος θεωρείται θετική αν η έξοδος είναι μεγαλύτερη ή ίση του μηδενός (όπως και προηγούμενα) ή αν η έξοδος είναι μεγαλύτερη του  $-0.7$  και ταυτόχρονα η συνιστώσα  $L$  είναι μεγαλύτερη του 65. Αποτελέσματα αυτού του κριτηρίου επιλογής φαίνονται στο Σχήμα 2.8. Τα δύο κριτήρια επιλογής μπορούν να συγκριθούν μεταξύ τους παρατηρώντας τα αντίστοιχα αποτελέσματα.



Σχήμα 2.7: Αποτελέσματα κατάταξης του χρώματος του ανθρώπινου δέρματος με κατφλι εξόδου ίσου με το μηδέν. Χρησιμοποιήθηκε το χρωματικό σύστημα *Lab* και παρουσιάζονται με άσπρο τα σημεία που τμηματοποιήθηκαν ως σημεία δέρματος και με μαύρο τα αρνητικά σημεία, δηλαδή αυτά που δεν ανήκουν σε ανθρώπινο δέρμα κατά το νευρωνικό.



Σχήμα 2.8: Αποτελέσματα κατάταξης του χρώματος του ανθρώπινου δέρματος χρησιμοποιώντας ως κριτήριο επιλογής: η έξοδος του δικτύου να είναι μεγαλύτερη ή ίση του μηδενός, ή η έξοδος να είναι μεγαλύτερη του  $-0.7$  αν η συνιστώσα  $L$  είναι μεγαλύτερη του 65. Χρησιμοποιήθηκε το χρωματικό σύστημα  $Lab$  και παρουσιάζονται με άσπρο τα σημεία που κατατάχθηκαν ως σημεία δέρματος και με μαύρο τα αρνητικά σημεία, δηλαδή αυτά που δεν ανήκουν σε ανθρώπινο δέρμα κατά το νευρωνικό.

## Κεφάλαιο 3

# Υποψήφιες περιοχές προσώπου και έλεγχος του σχήματός τους

Αφού, λοιπόν, έχει προηγηθεί στην εικόνα η ταξινόμηση των χρωματικών περιοχών σε χρώμα ανθρώπινου δέρματος ή όχι, χρησιμοποιώντας μία από τις τρεις παραπάνω μεθόδους, οι περιοχές με το χρώμα του ανθρώπινου δέρματος πρέπει να συνενωθούν, έτσι ώστε αυτές να σχηματίσουν ενδεχόμενες περιοχές προσώπου. Αυτές είναι που θα χρησιμοποιηθούν τελικά για την κατάταξή τους σε πρόσωπο ή όχι με βάση την πολυκαναλική ανάλυση σε κυματιδιακά πλαίσια και την ανάλυση της υφής του προσώπου.

Σε προηγούμενη εργασία [13] παρουσιάζεται μία μέθοδος ανίχνευσης προσώπων η οποία εφαρμόστηκε στα I frames που εξάχθηκαν από MPEG-1 βίντεο. Σε επίπεδο μακρο-μπλοκ (macro-block level,  $16 \times 16$  pixels) χρησιμοποιείται φιλτράρισμα του χρώματος του ανθρώπινου δέρματος και έτσι η κάθε εμφάνιση αυτού του χρώματος συμπληρώνει μία δυαδική μάσκα (σε επίπεδο μακρο-μπλοκ πάντοτε) η οποία τμηματοποιείται σε μη επικαλυπτόμενες παραλληλόγραμμες περιοχές που περιέχουν γειτονικές περιοχές του χρώματος του ανθρώπινου δέρματος. Ξεκινώντας από τη μεγαλύτερη τέτοια υποψήφια για πρόσωπο περιοχή μειώνεται επαναληπτικά το μέγεθός της ώστε να ερευνηθούν όλοι οι δυνατοί λόγοι δύο διαστάσεων και θέσεων σε κάθε περιοχή.

Στο προτεινόμενο εδώ σχήμα προηγείται ένα στάδιο συγχώνευσης στο σύνολο των ομοιογενών περιοχών του χρώματος του ανθρώπινου δέρματος στην κβαντισμένη ως προς το χρώμα εικόνα. Έτσι τελικά έχουμε ένα σύνολο από υποψήφιες περιοχές προσώπων χωρίς να χρειάζεται να ερευνηθούν όλοι οι δυνατοί λόγοι δύο διαστάσεων και θέσεων σε κάθε περιοχή. Αυτή η ενέργεια μας δίνει μεγαλύτερη ακρίβεια στον εντοπισμό του προσώπου και βοηθάει στην τμηματοποίηση των προσώπων από το υπόλοιπο φόντο, ειδικά σε περιπτώσεις όπου το φόντο περιέχει περιοχές χρωματικά όμοιες με το ανθρώπινο πρόσωπο.

### 3.1 Αλγόριθμος συγχώνευσης

Η κατετμημένη σε χρώμα ανθρώπινου δέρματος εικόνα αποτελείται πλέον από μικρές ομοιογενείς περιοχές ( $R_i$ ) λόγω του κβαντισμού που προηγήθηκε. Στόχος του τμήματος αυτού είναι, όπως προαναφέρθηκε, η δημιουργία υποψήφιας περιοχών προσώπου με επαναληπτική συγχώνευση των γειτονικών ομοιογενών περιοχών του χρώματος του ανθρώπινου δέρματος. Το αποτέλεσμα της συγχώνευσης αυτής είναι η επαναληπτική δημιουργία τέτοιων περιοχών, ενώ ταυτόχρονα διαχωρίζουμε μεταξύ των διαφόρων αποχρώσεων του χρώματος του ανθρώπινου δέρματος στις διάφορες γειτονικές περιοχές. Αυτό μας επιτρέπει την τμηματοποίηση του προσώπου από ένα χρωματικά όμοιο φόντο του.

Ο αλγόριθμος της συγχώνευσης αρχικοποιείται υπολογίζοντας ένα γράφο γειτονικών περιοχών (region adjacency graph), όπου κάθε κόμβος αναπαριστά μία ομοιογενή περιοχή του χρώματος του ανθρώπινου δέρματος της κβαντισμένης εικόνας. Το κριτήριο συνδετικότητας των κόμβων υπολογίζεται μεταξύ των αντίστοιχων περιγεγραμμένων παραλληλογράμμων κάθε ομοιογενούς περιοχής. Η συνδετικότητα αποτιμάται χρησιμοποιώντας μία απόσταση μεταξύ των κορυφών των περιγεγραμμένων

παραλληλογράμμων. Η μέθοδος αυτή είναι πολύ πιο γρήγορη από την εκτίμηση της συνδετικότητας σε επίπεδο pixels.

Έστω  $C_0$  είναι ένα σύνολο από ομοιογενείς περιοχές του χρώματος του ανθρώπινου δέρματος  $R_i$ ,  $i \in [1 \dots N]$ . Αρχικά το κριτήριο συνδετικότητας εφαρμόζεται για τη δημιουργία του γράφου. Εφόσον επιλέξαμε να δουλέψουμε με τα περιγεγραμμένα παραλληλόγραμμα υλοποιήθηκε ένας γρήγορος αλγόριθμος, ο οποίος θεωρεί την κανονική απόσταση μεταξύ των κορυφών των αποσυνδεδεμένων τεταγώνων ή αποφασίζει αν τα παραλληλόγραμμα αυτά επικαλύπτονται. Γι'αυτό το λόγο ορίζονται γειτονικές περιοχές. Το κριτήριο συγχώνευσης μεταξύ των περιγεγραμμένων παραλληλογράμμων γειτονικών περιοχών  $R_i, R_j$  βασίζεται σε μία μέγιστη επιτρεπόμενη απόσταση  $D_r(R_i, R_j)$  η οποία σηματοδοτεί τη χρωματική ανομοιότητα εντός των παραλληλογράμμων. Παρατηρήθηκε ότι οι διάφορες χρωματικές περιοχές σε ένα δεδομένο πρόσωπο είναι σταθερές ανάλογα με το  $H$  και χρωματικές διαφορές εμφανίζονται κυρίως στις συνιστώσες  $V$  και  $S$ , λόγω αποκλίσεων στην πρόσπτωση του φωτός πάνω στην επιφάνεια του προσώπου. Έτσι επιλέχθηκε το κριτήριο της χρωματικής ανομοιότητας να είναι το εξής :

$$D_r(R_i, R_j) = \alpha|H_i - H_j| + \beta|S_i - S_j| + \gamma|V_i - V_j| \quad (3.1)$$

όπου  $(H_i, S_i, V_i)$  είναι το μέσο διάνυσμα ου χρώματος του ανθρώπινου δέρματος εντός του περιγεγραμμένου παραλληλογράμμου  $R_i$ . Τα  $\alpha, \beta, \gamma$ , είναι βάρη, με το  $\alpha$  να έχει επιλεγεί αρκετά μεγαλύτερο από τα  $\beta, \gamma$ , ώστε να αποδίδεται μεγαλύτερη σημασία στις διαφορές στο  $H$  (που, όπως έχει επισημανθεί, κωδικοποιεί κυρίως το χρώμα του ανθρώπινου δέρματος).

Το σύνολο των υποψηφίων περιοχών προσώπου δημιουργείται επαναληπτικά ξεκινώντας από το σύνολο  $C_0$  και το γράφο των γειτονικών περιοχών του. Λαμβάνεται το σύνολο  $C_1$  συγχωνεύοντας συμβατές γειτονικές περιοχές του  $C_0$ . Ύστερα κάθε νέο σύνολο συγχωνευμένων περιοχών  $C_k$  ( $k \in [1 \dots K]$ ) λαμβάνεται συγχωνεύοντας επαναληπτικά το σύνολο των συγχωνευμένων περιοχών  $C_{k-1}$  με το αρχικό σύνολο  $C_0$  σύμφωνα με τον ακόλουθο τρόπο (κάθε γράφος γειτονικών περιοχών υπολογίζεται πριν από κάθε επανάληψη) :

$$\forall k \in [1 \dots K], \forall R_i \in C_0, \forall R_j \in C_{k-1}, \quad C_k = \bigcup R_i \cup R_j \quad (3.2)$$

αν επιτρέπεται  $D_r(R_i, R_j)$  και  $\{R_i, R_j\}$  είναι γειτονικές περιοχές.

Στη συγκεκριμένη υλοποίηση εκτελείται η επαναληπτική συγχώνευση με ένα μέγιστο αριθμό επαναλήψεων  $K = 3$ , ώστε να αποφεύγεται χρονική πολυπλοκότητα. Ακόμα παρατηρήθηκε, ότι το στάδιο της συγχώνευσης δίνει τις αναμενόμενες υποψήφιας περιοχές προσώπου εντός αυτών των τριών επαναλήψεων. Τελικά, η υποψήφια περιοχή προσώπου  $CF$  περιέχει το αρχικό σύνολο περιοχών  $C_0$  και το επαναληπτικά δημιουργημένο σύνολο συγχωνευμένων περιοχών  $C_k$  ( $k \in [1 \dots K]$ ), δηλ.  $CF = \bigcup_{k=0}^K C_k$ .

Στο Σχήμα 3.1 παρουσιάζεται η διαδικασία της συγχώνευσης σε πραγματικά παραδείγματα. Για κάθε παράδειγμα οι τέσσερις πρώτες εικόνες (στην αντίστοιχη γραμμή) περιέχουν τις ομοιογενείς περιοχές που αντιστοιχούν στα τέσσερα κυρίαρχα χρώματα ανθρώπινου δέρματος. Οι περιοχές που αντιστοιχούν στο λαμβανόμενο υπόψη χρώμα ανθρώπινου δέρματος απεικονίζονται με άσπρο. Για λόγους αναγνωσιμότητας τα περιγεγραμμένα παραλληλόγραμμα δεν ζωγραφίστηκαν γύρω από τις περιοχές αυτές. Στην πέμπτη εικόνα (στην αντίστοιχη γραμμή πάντοτε) φαίνονται τα περιγεγραμμένα αυτά παραλληλόγραμμα των υποψηφίων περιοχών προσώπου που δημιουργήθηκαν από τη διαδικασία της συγχώνευσης και περιέχονται στο σύνολο  $CF$ . Τα πρόσωπα που ανιχνεύτηκαν τελικά φαίνονται στην τελευταία εικόνα. Όπως μπορεί να παρατηρηθεί στην πέμπτη εικόνα, κάποια από τα περιγεγραμμένα παραλληλόγραμμα στο σύνολο  $CF$  είναι πολύ μικρά ή έχουν ένα λόγο διαστάσεων που δεν συμβαδίζει με αυτόν του ανθρώπινου προσώπου. Έτσι οδηγούμαστε στην εφαρμογή περιορισμών σχετικών με το σχήμα, την αποδεκτή διάσταση, την ομοιογένεια του χρώματος και την υφή του προσώπου, έτσι ώστε να ταξινομήσουμε κατάλληλα κάθε υποψήφια περιοχή του συνόλου  $CF$  σε πρόσωπο ή μη πρόσωπο.

## 3.2 Ανθρωπομορφικά κριτήρια

Όπως είδαμε τα ανθρωπομορφικά κριτήρια παίζουν μεγάλο ρόλο στην ταξινόμηση μίας υποψήφιας περιοχής. Δεν μπορεί οποιοδήποτε σχήμα να είναι πρόσωπο. Το σχήμα του ανθρώπινου προσώπου

μπορεί να προσεγγιστεί και να θεωρηθεί ως χαρακτηριστικό που το ξεχωρίζει και το διακρίνει. Έτσι, τα κριτήρια σχετικά με το σχήμα και τη διάσταση που θα εφαρμοστούν παρακάτω βοηθούν στην απομάκρυνση των περισσότερων μη πιθανών υποψήφιων περιοχών προσώπου.

Για την επίτευξη γρήγορου υπολογισμού αναλύουμε το σχήμα του προσώπου θεωρώντας μόνο το λόγο διαστάσεων των περιγεγραμμένων παραλληλογράμμων κάθε περιοχής. Ο λόγος αυτός περιορίζεται στο διάστημα  $[0.9, 2.1]$ , που θεωρείται αρκετά ευρύ και ικανό για να συμβαδίζει με διάφορους προσανατολισμούς και πόζες.

Επιπλέον εφαρμόζονται περιορισμοί που έχουν να κάνουν με τις επιτρεπόμενες για ένα πρόσωπο διαστάσεις. Η ελάχιστη διάσταση ενός επιτρεπόμενου προσώπου στην εικόνα φράσσεται από  $80 \times 48$  pixels. Φυσικά η μεγαλύτερη διάσταση που μπορεί να έχει περιορίζεται μονάχα από τις διαστάσεις τις εικόνας. Εφαρμόζοντας, λοιπόν, τους περιορισμούς αυτούς ελαττώνεται σημαντικά ο αριθμός των υποψηφίων περιοχών προσώπου  $CF$ .

Ο περιορισμός που αφορά την ομοιογένεια του χρώματος τίθεται σε εφαρμογή υπολογίζοντας την πυκνότητα των pixels που έχουν το χρώμα του ανθρώπινου δέρματος σε κάθε υποψήφια περιοχή  $R_i$ . Ορίζονται δύο υποπεριοχές, η εξωτερική και η εσωτερική αντίστοιχα, μέσα σε κάθε  $R_i$ . Η εξωτερική υποπεριοχή αντιστοιχεί στη συνοριακή περιοχή της  $R_i$  (το πλάτος της οποίας είναι ένα ποσοστό του πλάτους του περιγεγραμμένου παραλληλογράμμου, τυπικά 15%) και η εσωτερική περιοχή αντιστοιχεί στο εσωτερικό υπόλοιπο τμήμα της  $R_i$ . Μία περιοχή  $R_i$  γίνεται αποδεκτή σαν τελική υποψήφια περιοχή προσώπου εάν η πυκνότητα των pixels που έχουν το χρώμα του ανθρώπινου δέρματος σε ολόκληρη την περιοχή  $R_i$  ( $p$ ) είναι πάνω από ένα κατώφλι που τίθεται στο 0.4 και ταυτοχρόνως η πυκνότητα στην εσωτερική περιοχή ( $p_{in}$ ) είναι πάνω από το κατώφλι  $0.7p + 0.3$ . Η ομοιογένεια στην εσωτερική περιοχή πρέπει να είναι υψηλότερη απ' ό,τι σε ολόκληρη την  $R_i$  δεδομένου ότι στην εσωτερική περιοχή της ανήκει το κεντρικό τμήμα του προσώπου, χωρίς το φόντο, ενώ σε ολόκληρη την  $R_i$  μπορεί να περιλαμβάνονται μαλλιά και τμήματα του φόντου (αφού πρόκειται πάντα για παραλληλόγραμμα).



Σχήμα 3.1: Ομοιογενείς περιοχές του χρώματος του ανθρώπινου δέρματος, υποψήφιες περιοχές προσώπων, τελικά ανιχνευθέντα πρόσωπα.

## Κεφάλαιο 4

# Ελεγχος υφής με ανάλυση σε κυματιδιακά πλαίσια

Ο βασικός σκοπός του τελικού αυτού σταδίου είναι η ταξινόμηση των περιοχών  $R_i$ , που πέρασαν τους παραπάνω περιορισμούς στις δύο κατηγορίες, κλάσεις : πρόσωπα και μη πρόσωπα. Η ταξινόμηση αυτή στοχεύει στην απαλοιφή λαθών (περιοχές που στην πραγματικότητα δεν είναι πρόσωπα αλλά πέρασαν τα παραπάνω κριτήρια και άρα αποτελούν ακόμα υποψήφιας περιοχές) που προκλήθηκαν από αντικείμενα χρωματικά όμοια με το ανθρώπινο δέρμα και με περίπου όμοιο λόγο διαστάσεων με αυτόν του ανθρώπινου προσώπου, όπως μπορεί να είναι άλλα εκτεθειμένα τμήματα του ανθρώπινου σώματος ή τμήματα του φόντου. Για το σκοπό αυτό εφαρμόζεται μία μέθοδος ανάλυσης σε πακέτα κυματιδιακών πλαισίων, όμοια με εκείνη που προτείνεται για αναγνώριση προσώπων στο [12]. Χρησιμοποιώντας πλαίσια κυματιδίων γίνεται μία ανάλυση του επιπέδου φωτεινότητας της εικόνας και εξάγονται χαρακτηριστικά διανύσματα από ένα σύνολο συντελεστών, που προκύπτουν από τα πακέτα των κυματιδιακών πλαισίων, σε κάθε περιοχή  $R_i$ . Το επίπεδο φωτεινότητας είναι το επίπεδο  $Y$  της αρχικής έγχρωμης εικόνας, η οποία μετασχηματίστηκε από το χρωματικό σύστημα  $HSV$  στο  $YCbCr$ . Αυτή η εικόνα έντασης μεταφέρει πληροφορία σχετικά με την υφή των προσώπων η οποία και θα διατηρηθεί στους συντελεστές των κυματιδίων.

Η ανάλυση σε πλαίσια κυματιδίων θεωρήθηκε κατάλληλη για το σκοπό της εργασίας, διότι μας προμηθεύει με μία πολυκλιμακωτή ανάλυση της εικόνας, ταυτόχρονα στο χώρο και στις συχνότητες. Ισχυρά επιχειρήματα για τη χρήση πολυκλιμακωτής ανάλυσης μπορούν να βρεθούν σε ψυχοφυσικές έρευνες, οι οποίες αποδεικνύουν ότι το ανθρώπινο σύστημα όρασης επεξεργάζεται τις εικόνες με έναν πολυκλιμακωτό τρόπο.

Την τελευταία δεκαετία τα κυματίδια έγιναν αρκετά δημοφιλή και ένα νέο ενδιαφέρον γεννήθηκε και αυξάνει ολοένα γύρω από αυτό το θέμα. Ο κυριότερος λόγος είναι ότι δημιουργήθηκε πρόσφατα ένα πλήρες πλαίσιο εργασίας [9], [24] κυρίως σε ό,τι αφορά τη δημιουργία βάσεων κυματιδίων.

Ένα ενδιαφέρον χαρακτηριστικό των κυματιδίων είναι η ευκαμψία που τα διακρίνει : υπάρχουν διάφορες βάσεις από τις οποίες κάποιος πορεί να διαλέξει εκείνη που ταιριάζει περισσότερο σε κάποια εφαρμογή. Στο [6] χρησιμοποιείται ένα κριτήριο βασισμένο στην εντροπία για την επιλογή των βέλτιστων βάσεων, τη γενική πάντοτε περίπτωση. Όμως το θέμα αυτό είναι ακόμα ανοιχτό. Ειδικά για την περίπτωση προσώπων στην εικόνα, μπορούν να επικρατούν πειραματικές θεωρήσεις στην επιλογή της βάσης.

Σχετικά με την υπολογιστική πολυπλοκότητα που έχει η κυματιδιακή ανάλυση πρέπει να επισημάνουμε ότι είναι γραμμική με τον αριθμό ( $N$ ) των συντελεστών που υπολογίζονται ( $O(N)$ ). Άλλοι μετασχηματισμοί, στην πιο γρήγορή τους υλοποίηση τείνουν σε  $N \times \log N$  πολυπλοκότητα.

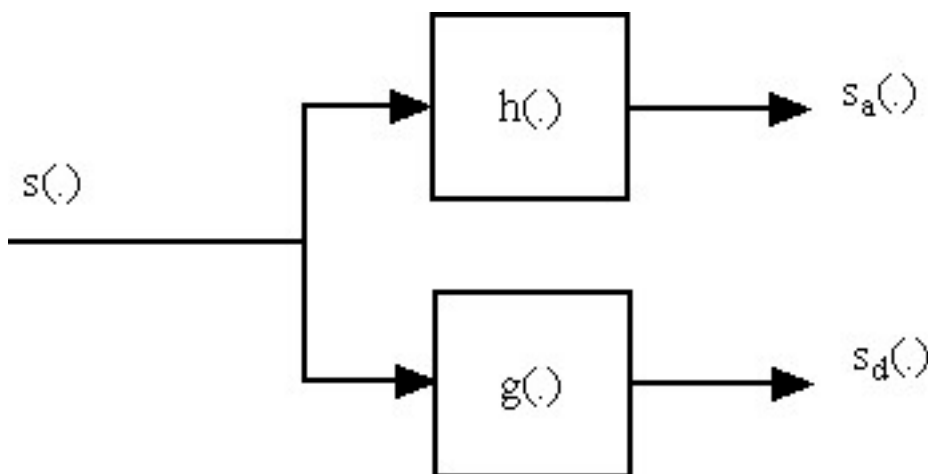
## 4.1 Πολυκαναλική ανάλυση σε κυματιδιακά πλαίσια

Η διακριτή σειρά κυματιδίων για ένα συνεχές σήμα  $s(t)$  ορίζεται από την παρακάτω εξίσωση :

$$c_{n,k} = \frac{1}{2^{\frac{n}{2}}} \int s(t) \psi^*(2^{-n}t - k) dt \quad (4.1)$$

Η συνάρτηση  $\psi(t)$  αποτελεί την κυματιδιακή μήτρα, η οποία ικανοποιεί κάποια κριτήρια παραδεκτότητας (admissibility criteria) και εξασφαλίζει μία πλήρη, χωρίς πλεονασμούς και ορθογωνική αναπαράσταση του σήματος. Ο διακριτός μετασχηματισμός πλαισίων κυματιδίων (DWF) προκύπτει από την παραπάνω σειρά και ισοδυναμεί με διαδοχικές αναλύσεις του σήματος χρησιμοποιώντας ένα ζευγάρι φίλτρων  $h(\cdot)$  και  $g(\cdot)$ , όπως φαίνεται και στο Σχήμα 4.1. Το βαθυπερατό φίλτρο  $h(\cdot)$  προμηθεύει τις προσεγγίσεις του σήματος, ενώ το υψιπερατό  $g(\cdot)$  δίνει τις λεπτομέρειες.

Η επέκταση σε δύο διαστάσεις (για εικόνες) γίνεται εφαρμόζοντας τα δύο αυτά φίλτρα ξεχωριστά στις δύο κατευθύνσεις. Ο συγκεκριασμός με το βαθυπερατό φίλτρο παράγει μία εικόνα προσέγγισης και ο συγκεκριασμός με το υψιπερατό σε συγκεκριμένες κατευθύνσεις οδηγεί στις επονομαζόμενες εικόνες των λεπτομερειών.



Σχήμα 4.1: Δυαδική ανάλυση ενός σήματος  $s$ .

Στην κλασική ανάλυση πλαισίων κυματιδίων η εικόνα χωρίζεται σε προσεγγίσεις και λεπτομέρειες. Οι προσεγγίσεις μετά αναλύονται περαιτέρω σε ένα δεύτερο επίπεδο προσεγγίσεων και λεπτομερειών, αφού πρώτα γίνει μία δυαδική υπερδειγματοληψία των φίλτρων, δηλαδή παρεμβολή ενός μηδενικού ενδιάμεσα από κάθε συντελεστή του κάθε φίλτρου, κ.ο.κ. Για ένα  $n$ -οστό επίπεδο ανάλυσης η εικόνα διαχωρίζεται σύμφωνα με τον ακόλουθο τρόπο :

$$\begin{aligned} A_n &= [H_x]_{\uparrow 2^n} * [H_y]_{\uparrow 2^n} * A_{n-1} \\ D_{n1} &= [H_x]_{\uparrow 2^n} * [G_y]_{\uparrow 2^n} * A_{n-1} \\ D_{n2} &= [G_x]_{\uparrow 2^n} * [H_y]_{\uparrow 2^n} * A_{n-1} \\ D_{n3} &= [G_x]_{\uparrow 2^n} * [G_y]_{\uparrow 2^n} * A_{n-1} \end{aligned} \quad (4.2)$$

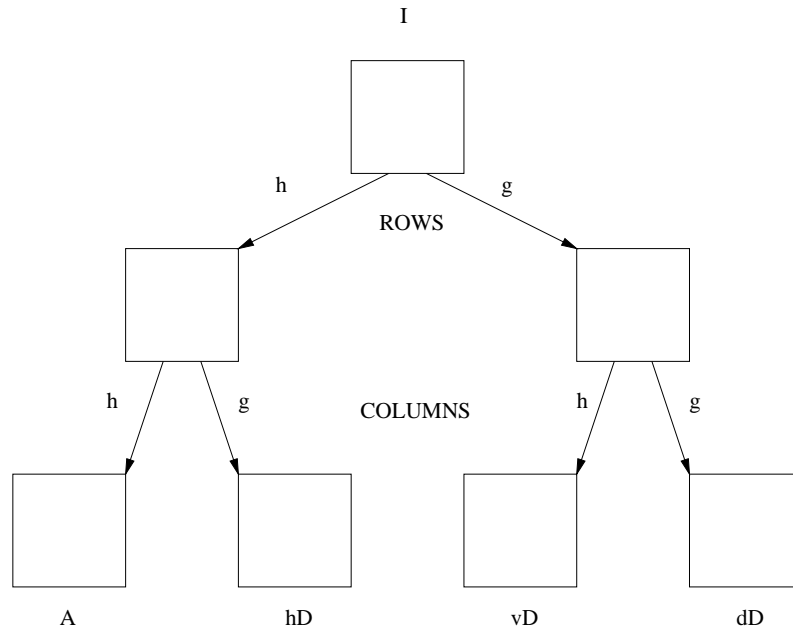
όπου  $*$  είναι ο συντελεστής συγκεκριασμού,  $\uparrow 2^n$  σημαίνει  $n$ -φορές δυαδική υπερδειγματοληψία και  $A_0 = I(x, y)$  είναι η αρχική εικόνα, που αποτελεί και το επίπεδο 0 της ανάλυσης. Η εικόνα των προσεγγίσεων στο επίπεδο  $n$ ,  $A_n$  λαμβάνεται από βαθυπερατό φιλτράρισμα. Αντίθετα από υψιπερατό φιλτράρισμα σε συγκεκριμένη κατεύθυνση λαμβάνονται οι εικόνες των λεπτομερειών  $D_{ni}$  ( $i = 1, 2, 3$ , για την οριζόντια, την κάθετη και τη διαγώνια κατεύθυνση αντίστοιχα) και περιέχουν κατευθυντικές λεπτομέρειες στην κλίμακα  $n$ . Η αρχική εικόνα αντιπροσωπεύεται έτσι από ένα σύνολο υπο-εικόνων σε διάφορες κλίμακες (δηλαδή τα διάφορα επίπεδα) :  $\{A_n, D_{ni}\}$ .

Η ανάλυση σε πλαίσια κυματιδίων χρησιμοποιήθηκε σ' αυτήν την εργασία με τη μορφή πακέτων. Η ανάλυση σε πακέτα κυματιδιακών πλαισίων είναι μία γενίκευση της κλασικής ανάλυσης πλαισίων



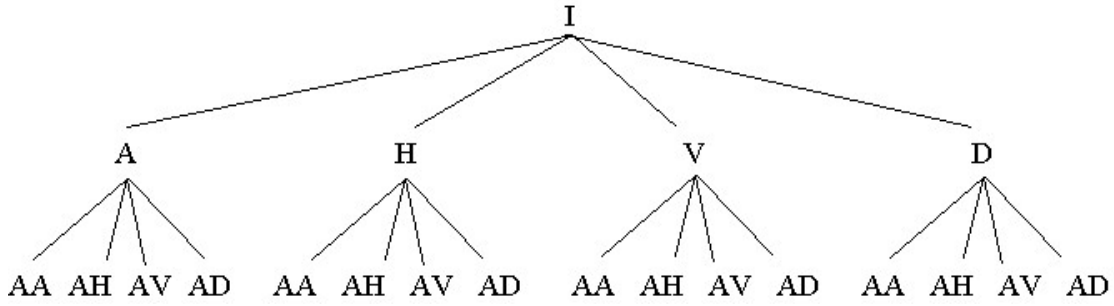
κυματιδίων, η οποία προσφέρει μία πλουσιότερη ανάλυση του σήματος. Αυτό σημαίνει ότι αναλύονται περαιτέρω και οι εικόνες των λεπτομερειών μαζί με την εικόνα των προσεγγίσεων του κάθε επιπέδου. Το Σχήμα 4.2 δείχνει πως προκύπτουν οι εικόνες των προσεγγίσεων και των λεπτομερειών για το πρώτο επίπεδο ανάλυσης με πλαίσια κυματιδίων (από μία αρχική εικόνα εξάγονται τέσσερις υπο-εικόνες που την περιγράφουν). Η έννοια των πακέτων θα γίνει καλύτερα κατανοητή από το Σχήμα 4.3, ενώ ένα παράδειγμα της ανάλυσης σε πλαίσια κυματιδίων μέχρι το επίπεδο 3 φαίνεται στο Σχήμα 4.4 (παρουσιάζονται μόνο οι εικόνες που προκύπτουν κάθε φορά από τις προσεγγίσεις του αντίστοιχου επιπέδου, δηλαδή η δεύτερη γραμμή αποτελείται από τις : A, H, V, D, η δεύτερη γραμμή προέκυψε από την ανάλυση της A και αποτελείται από τις : AA, AH, AV, AD, και η τρίτη γραμμή προέκυψε από την ανάλυση της AA και αποτελείται από τις : AAA, AAH, AAV, AAD, σύμφωνα με την αναπαράσταση του Σχήματος 4.3).

Για το επόμενο επίπεδο, κάθε μία από τις τέσσερις αυτές υπο-εικόνες θεωρείται σαν "αρχική" και εφαρμόζεται και πάλι η όλη διαδικασία με τα φίλτρα  $h(\cdot)$ ,  $g(\cdot)$ , που όμως πρέπει να έχουν δυαδικά υπερδειγματοληφθεί, πως περιγράφηκε παραπάνω. Έτσι θα έχουμε πάλι τέσσερις εικόνες για κάθε μία που αναλύθηκε και το δεύτερο επίπεδο θα περιλαμβάνει συνολικά 16 εικόνες που θα περιγράφουν την αρχική. Η ανάλυση μπορεί να συνεχιστεί μέχρι οποιοδήποτε επιθυμητό επίπεδο  $n$ , με ένα τελικό σύνολο εικόνων  $4^n$  να περιγράφουν την αρχική. Η προσέγγιση αυτή θεωρήθηκε κατάλληλη διότι κρατάει όσο το δυνατόν περισσότερη πληροφορία. Αυτό οφείλεται στο ότι οι τελικές υπο-εικόνες έχουν τις διαστάσεις της αρχικής (δεν έχουμε λοιπόν απώλεια πληροφορίας λόγω υποδειγματοληψίας όπως θα συνέβαινε στην κλασική ανάλυση κυματιδίων) και άρα ακόμα και μικρού μεγέθους πρόσωπα δεν αλλοιώνονται. Ταυτόχρονα τα χαρακτηριστικά προσώπου επαυξάνονται και διατηρούν και τα κανονικά τους μεγέθη, κάτι που σίγουρα ενδιαφέρει για κάποια περαιτέρω επεξεργασία.



Σχήμα 4.2: Το πρώτο επίπεδο της ανάλυσης με κυματιδικά πλαίσια χρησιμοποιώντας τα αρχικά φίλτρα  $h, g$ .

Το Σχήμα 4.5 δείχνει τα δύο φίλτρα  $h$  και  $g$ , βαθυπερατό και υψιπερατό αντίστοιχα, που χρησιμοποιήθηκαν για τη διακριτή ανάλυση πλαισίων κυματιδίων. Όπως προαναφέρθηκε ήδη, η επιλογή των πιο κατάλληλων φίλτρων για μία εφαρμογή σαν κι αυτή βασίζεται κυρίως σε πειραματικά αποτελέσματα. Οι περιορισμοί που λήφθηκαν υπόψη για τη δημιουργία των φίλτρων σχετίζονται στενά με την έρευνα που περιγράφεται στο [30]. Σύμφωνα με αυτήν και την [37], τα  $h, g$  δημιουργήθηκαν ως συζυγή τετραγωνικά φίλτρα. Οι συντελεστές καθώς και οι διαστάσεις των δύο φίλτρων επιλέχθηκαν εμπειρικά.



Σχήμα 4.3: Το δέντρο της πολυκαναλικής ανάλυσης σε κυματιδιακά πλαίσια μέχρι το επίπεδο 2. A : οι προσεγγίσεις, H : οι οριζόντιες λεπτομέρειες, V : οι κάθετες λεπτομέρειες, D : οι διαγώνιες λεπτομέρειες. Όλες οι υποεικόνες κάθε επιπέδου υπόκεινται ξανά στην ανάλυση.

Ο μετασχηματισμός Z του ζεύγους των φίλτρων είναι :

$$\begin{aligned} H(z) &= 0.853 + 0.377(z + z^{-1}) - 0.111(z^2 + z^{-2}) - 0.024(z^3 + z^{-3}) + 0.038(z^4 + z^{-4}) \\ G(z) &= -z^{-1}H(-z^{-1}) \end{aligned} \quad (4.3)$$

Ο μετασχηματισμός Fourier των φίλτρων αυτών μπορεί να βρεθεί στο Σχήμα 4.6. Το βαθυπερατό φίλτρο είναι συμμετρικό και το ζεύγος είναι ορθογώνιο :

$$\sum_n h(n-2k)g(n-2l) = 0, \quad \forall(k,l) \quad (4.4)$$

Εφόσον ενδιαφερόμαστε να έχουμε αποσυσχετισμένες αποκρίσεις των φίλτρων θεωρούμε δύο ποσοτικά κριτήρια της ιδιότητας αυτής. Το συντελεστή ιδανικής ανάλυσης (antialiasing), που ορίζεται ως :

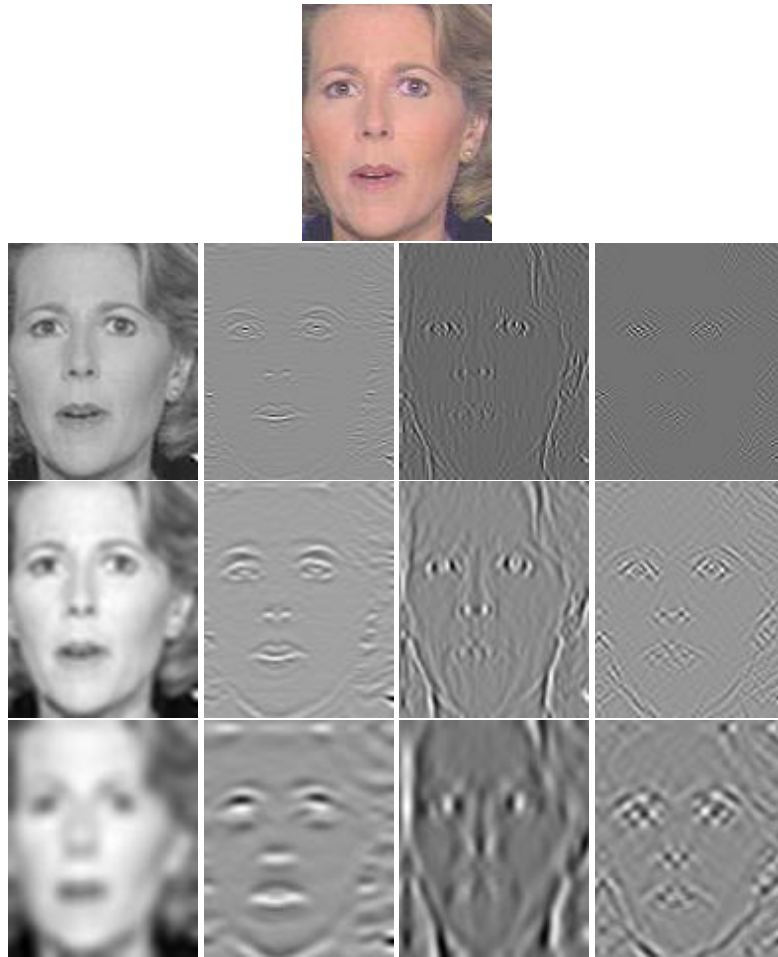
$$n = \frac{\int_0^{\frac{1}{2}} |\mathcal{H}(f)|^2 df}{\int_0^{\frac{1}{2}} |\mathcal{H}(f)|^2 df} = 0.87 \quad (4.5)$$

όπου  $H(f)$  είναι ο μετασχηματισμός Fourier του φίλτρου  $h(n)$  με ιδανική τιμή 1. Επίσης ορίζουμε και ένα συντελεστή συσχέτισης μεταξύ των σημάτων των προσεγγίσεων και των λεπτομερειών ως εξής :

$$\rho = \frac{\sum_{k,l} h(k)g(l)\gamma_s(k-l)}{\sqrt{\sum_{k,l} h(k)h(l)\gamma_s(k-l)} \sqrt{\sum_{k,l} g(k)g(l)\gamma_s(k-l)}} \quad (4.6)$$

όπου  $\gamma_s(\cdot)$  είναι η συνάρτηση αυτοσυνδιασποράς (autocovariance function) του σήματος εισόδου. Για ένα ασυσχέτιστο (λευκός θόρυβος) ή πλήρως συσχετισμένο σήμα εισόδου, είναι φανερό ότι το  $\rho$  έχει τιμή μηδέν. Υπολογίσαμε το  $\rho$  υπό την προϋπόθεση μιας πρώτου επιπέδου διεργασίας Markov για το  $s(t)$  και η μέγιστη τιμή του βρέθηκε να είναι 0.06. Είναι επαρκές μικρή θεωρώντας ότι η υπόθεση της ενδοζώνιας αποσυσχέτισης ισχύει πρακτικά.

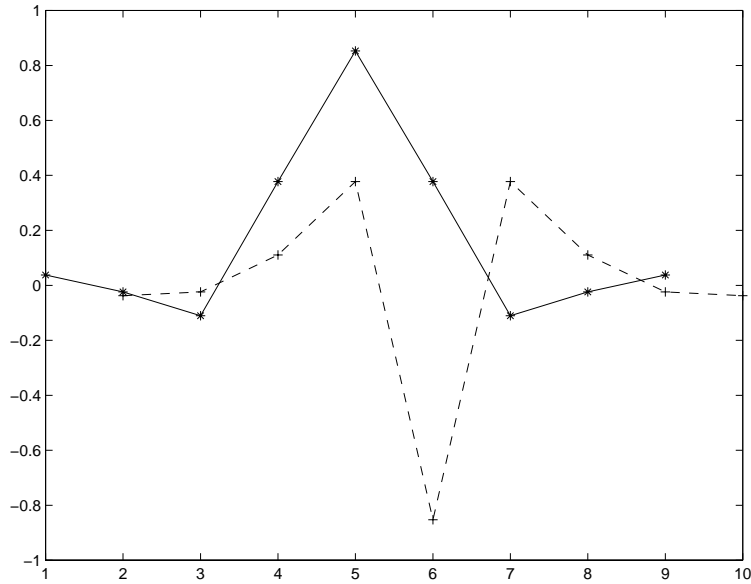
Η επιλογή του επιπέδου της παραπάνω ανάλυσης θα μπορούσε να γίνει με βάση κάποιο κριτήριο εντροπίας, κρατώντας παράλληλα την πιο σημαντική πληροφορία. Στην παρούσα περίπτωση το επίπεδο της ανάλυσης εξαρτάται από το μέγεθος των  $R_i$  περιοχών που επεξεργάζονται. Συγκεκριμένα εφαρμόστηκε η ανάλυση με πλαίσια κυματιδίων σε μορφή πακέτων σε 50 περιοχές προσώπων, οι οποίες κόπηκαν με το χέρι, και περιλαμβάνουν ένα μεγάλο αριθμό διαφορετικών περιπτώσεων, όπων είναι οι διάστασεις, η χρωματική πληροφορία, η φωτεινότητα, η θέση και ο προσανατολισμός του προσώπου. Τα παραδείγματα αυτά ταξινομήθηκαν σε δύο κατηγορίες, ανάλογα με το ύψος τους και χρησιμοποιήθηκαν για τη δημιουργία πρότυπων διανυσμάτων. Αν ένα πρόσωπο έχει ύψος μικρότερο



Σχήμα 4.4: Αρχική εικόνα προσώπου και η ανάλυσή της σε πλαίσια κυματιδίων μέχρι το επίπεδο 3 (των προσεγγίσεων μόνο). Από αριστερά προς τα δεξιά παρουσιάζονται οι εικόνες των προσεγγίσεων, των οριζοντίων, καθέτων και διαγωνίων λεπτομερειών όπως προκύπτουν για κάθε επίπεδο.

των 128 pixels θεωρείται μεσαίο, ενώ θεωρείται μεγάλο αν το ύψος του ξεπερνάει τα 128 pixels. Αφού μελετήθηκαν όλα τα ενδιάμεσα αποτελέσματα (για κάθε επίπεδο της ανάλυσης) διαπιστώθηκε ότι η ανάλυση μπορεί να σταματήσει στο επίπεδο 3, αφού τα επόμενα επίπεδα δεν παρέχουν πιο σημαντική πληροφορία. Σύμφωνα με τις διαστάσεις ενός υποψηφίου προσώπου αυτό μπορεί να περιγραφεί με ένα σύνολο  $k$  πινάκων από κυματιδιακούς συντελεστές. Για τα μεγάλα πρόσωπα χρησιμοποιείται το τελευταίο επίπεδο της ανάλυσης, δηλαδή το τρίτο, που σημαίνει  $k = 4^3 = 64$  πίνακες (μία εικόνα με προσεγγίσεις και 63 με λεπτομέρειες). Για τα μεσαία πρόσωπα χρησιμοποιείται το δεύτερο επίπεδο, δηλαδή  $k = 4^2 = 16$  πίνακες (μία εικόνα με προσεγγίσεις και 15 με λεπτομέρειες). Όπως είναι φανερό πρόκειται για ένα αρκετά μεγάλο όγκο δεδομένων, από τον οποίο πρέπει να εξαχθεί η διαχωρίζουσα πληροφορία, πετυχαίνοντας ταυτόχρονα και μείωση διαστάσεων.

Σε προηγούμενη εργασία μας, που αφορούσε την αναγνώριση προσώπων [12], αναφέρονται καλά αποτελέσματα τα οποία επετεύχθησαν εξάγοντας στατιστικές πληροφορίες από τους συντελεστές των κυματιδίων από διάφορες περιοχές του πίνακα (εικόνα) των προσεγγίσεων και ολόκληρων των πινάκων των λεπτομερειών. Στη συγκεκριμένη εργασία κάθε περιοχή  $R_i$  θεωρείται με βάση το περιγραμμένο παραλληλόγραμμο της, το οποίο και χωρίζεται σε τέσσερα ισομεγέθη μέρη : πάνω αριστερά ( $top_l$ ), πάνω δεξιά ( $top_r$ ), κάτω αριστερά ( $bottom_l$ ), κάτω δεξιά ( $bottom_r$ ). Εξάγοντας στατιστικές αποκλίσεις για τους συντελεστές της ανάλυσης είμαστε σε θέση να γνωρίζουμε πληροφορία σχετική με την υφή του προσώπου, η οποία σχετίζεται με τα διάφορα μέρη του, όπως τα μάτια, τη μύτη και το στόμα. Για την περιγραφή της υφής του προσώπου επιλέχθηκαν τυπικές αποκλίσεις. Έτσι εξάγονται οι



Σχήμα 4.5: Τα φίλτρα που χρησιμοποιήθηκαν για την πολυκαναλική ανάλυση πλαισίων κυματιδίων. Με τη συνεχή γραμμή απεικονίζεται το βαθυπερατό φίλτρο  $h$ , ενώ με την διακεκομμένη το υπερπερατό  $g$ .

τυπικές αποκλίσεις  $\sigma_{top_l}, \sigma_{top_r}, \sigma_{bottom_l}, \sigma_{bottom_r}$  των κυματιδιακών συντελεστών, που αντιστοιχούν στις τέσσερις προαναφερόμενες περιοχές της εικόνας των προσεγγίσεων (από το αντίστοιχο με το μέγεθος του προσώπου επίπεδο της ανάλυσης). Επιπλέον εξάγονται και από τις  $m$  εικόνες των λεπτομερειών ( $m \in \{15, 63\}$ ) οι αντίστοιχες τυπικές αποκλίσεις  $\sigma_i$  ( $i = 4, \dots, k; k = m + 3$ ) από όλη την  $R_i$  περιοχή.

Για κάθε υποψήφιο πρόσωπο, λοιπόν, εξάγονται χαρακτηριστικά διανύσματα τα οποία περιέχουν ένα μέγιστο  $4 + m$  συνιστωσών (τέσσερις τυπικές αποκλίσεις για την εικόνα των προσεγγίσεων και  $m$  τυπικές αποκλίσεις για τις εικόνες των λεπτομερειών). Τα διανύσματα αυτά περιγράφονται ως εξής :  $V = \cup_{i=0}^{m+3} \{\sigma_i\}$ , όπου οι δείκτες  $i = 0, 1, 2, 3$  αντιστοιχούν στις τυπικές αποκλίσεις  $\sigma_{top_l}, \sigma_{top_r}, \sigma_{bottom_l}, \sigma_{bottom_r}$ .

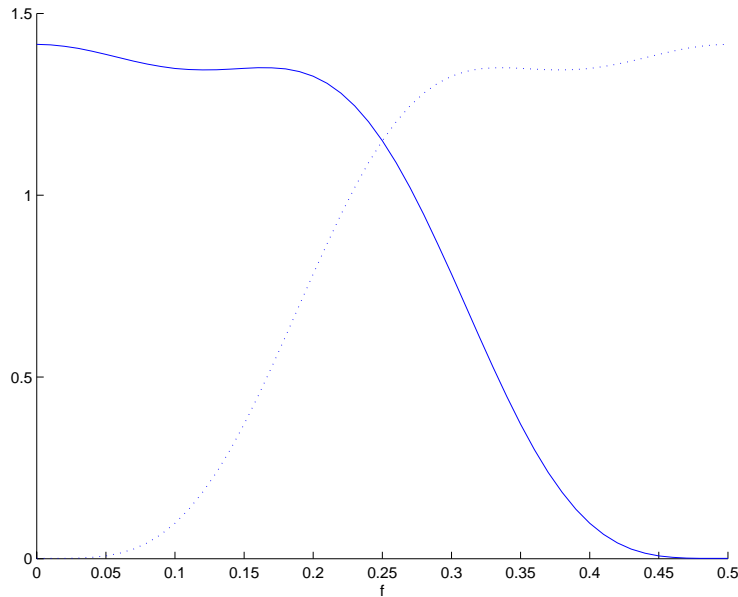
## 4.2 Ταξινόμηση των χαρακτηριστικών διανυσμάτων

Το διάνυσμα χαρακτηριστικών, που δημιουργήθηκε όπως είδαμε στην προηγούμενη ενότητα, χρησιμοποιείται για την ταξινόμηση της υποψήφιας περιοχής προσώπου  $R_i$  στις δύο κατηγορίες : πρόσωπο, μη πρόσωπο. Από τη βάση δεδομένων που αποτελείται από πρόσωπα που εξάχθηκαν με το χέρι (βλ. παραπάνω) και τα οποία ταξινομήθηκαν στις δύο κατηγορίες (μεσαία, μεγάλα), εξάγονται χαρακτηριστικά διανύσματα και δημιουργείται ένα πρότυπο χαρακτηριστικό διάνυσμα για κάθε μία από τις κατηγορίες αυτές.

Το πρόβλημα της ταξινόμησης πλέον ανάγεται σε ένα πρόβλημα αναγνώρισης προτύπων. Ο απώτερος στόχος είναι να ταξινομηθούν άγνωστα πρότυπα με τη χαμηλότερη δυνατή πιθανότητα λάθους ταξινόμησης. Στην περίπτωση μας, όπου πρόκειται για δύο πιθανές κλάσεις που μπορεί να ταξινομηθεί ένα διάνυσμα χαρακτηριστικών η πιθανότητα  $e$  των λάθους ταξινόμησης είναι :

$$e = \frac{1}{2} \left[ 1 - \int |p(X|\omega_1)P(\omega_1) - p(X|\omega_2)P(\omega_2)|dX \right] \quad (4.7)$$

Περισσότερες πληροφορίες για την πιθανότητα λάθους μπορούν να βρεθούν στο [11], απ' όπου και γίνεται κατανοητό ότι η πιθανότητα αυτή δεν μπορεί να αποτιμηθεί εύκολα και προτείνεται μία σειρά από εναλλακτικά κριτήρια. Ένα από αυτά βασίζεται σε πιθανοκρατικές μετρήσεις αποστάσεων. Η



Σχήμα 4.6: Ο μετασχηματισμός Fourier των φίλτρων που χρησιμοποιήθηκαν. Με τη συνεχή γραμμή απεικονίζεται το βαθυπερατό φίλτρο  $h$ , ενώ με την διακεκομμένη το υψιπερατό  $g$ .

πιθανότητα λάθους  $\epsilon$  θα μεγιστοποιηθεί, όταν η έκφραση υπό ολοκλήρωση θα είναι μηδέν, δηλαδή όταν οι συναρτήσεις πυκνότητας πιθανότητας επικαλύπτονται πλήρως, και, θα είναι μηδέν, όταν δεν επικαλύπτονται καθόλου. Η έκφραση υπό ολοκλήρωση μπορεί να θεωρηθεί ως η πιθανοκρατική απόσταση μεταξύ των δύο συναρτήσεων πυκνότητας πιθανότητας.

Ένα σημαντικό ζήτημα, λοιπόν, αποτελεί η επιλογή της πιθανοκρατικής απόστασης μεταξύ των δύο συναρτήσεων πυκνότητας πιθανότητας. Αυτή η επιλογή μπορεί να γίνει σύμφωνα με το μοντέλο των συναρτήσεων πυκνότητας πιθανότητας. Όσον αφορά το πρόβλημα της ταξινόμησης που αντιμετωπίζουμε, στατιστικές αναλύσεις των πειραματικών αποτελεσμάτων έδειξαν ότι η κατανομή πιθανότητας των εικόνων (πινάκων) των συντελεστών που προκύπτουν από την ανάλυση των κυματιδιακών πλαισίων είναι μία γενικευμένη γκαουσιανή (Gaussian) :

$$p(x) = \frac{c}{2\sigma\Gamma\left(\frac{1}{c}\right)} e^{-\frac{|x|}{\sigma}} \quad (4.8)$$

όπου η παράμετρος  $\sigma$  είναι η τυπική απόκλιση και  $c$  απεικονίζει την αιχμηρότητα της συνάρτησης πυκνότητας πιθανότητας. Για  $c = 2$  παίρνουμε τη γκαουσιανή συνάρτηση και για  $c = 1$  τη λαπλασιανή.

Μία από τις πιο δημοφιλείς πιθανοκρατικές αποστάσεις στον τομέα της αναγνώρισης προτύπων είναι η απόσταση *Bhattacharyya* [11], [18], [19]. Η απόσταση *Bhattacharyya* στην περίπτωση της γενικευμένης γκαουσιανής δίδεται ως εξής :

$$B_c = \frac{1}{c} \ln \left( \frac{\sigma_1^c + \sigma_2^c}{2\sqrt{\sigma_1^c \sigma_2^c}} \right) \quad (4.9)$$

με  $c$  όπως έχει περιγραφεί παραπάνω.

Στην περίπτωση μας συγκρίνουμε δύο χαρακτηριστικά διανύσματα  $V_k$  (ένα διάνυσμα που αντιπροσωπεύει ένα υποψήφιο πρόσωπο) και  $V_l$  (ένα πρότυπο διάνυσμα προσώπου). Η σύγκριση αυτή μπορεί να γίνει μετρώντας την απόσταση κάθε συνιστώσας του ενός διανύσματος από την αντίστοιχη του άλλου, οπότε η ολική απόσταση είναι το άθροισμα των επιμέρους αποστάσεων των συνιστωσών των δύο διανυσμάτων.

Στην περίπτωση της πολυκαναλικής ανάλυσης σε κυματιδιακά πλαίσια, που χρησιμοποιούμε, κάνουμε την πιθανή παραδοχή ότι οι εικόνες των προσεγγίσεων κατανέμονται σύμφωνα με το

γκαουσιανό νόμο, ενώ οι εικόνες που περιέχουν τους συντελεστές των λεπτομερειών κατανέμονται σύμφωνα με το λαπλασιανό. Δεν λαμβάνονται υπόψη οι διαφορές των μέσων τιμών των εικόνων των προσεγγίσεων, ώστε να έχουμε μία μέτρηση απόστασης ανεξάρτητη από τις συνθήκες φωτισμού της αρχικής εικόνας. Λόγω του σχεδιασμού των φίλτρων οι μέσες τιμές για τις εικόνες των λεπτομερειών είναι μηδέν.

Σύμφωνα με την απόσταση *Bhattacharyya* προκύπτει η απόσταση *D* μεταξύ δύο χαρακτηριστικών διανυσμάτων  $V_k$  και  $V_l$  ως εξής :

$$D(V_k, V_l) = \frac{1}{2} \sum_{i=0}^3 \ln \left( \frac{\sigma_{ik}^2 + \sigma_{il}^2}{2\sigma_{ik}\sigma_{il}} \right) + \sum_{i=4}^{m+3} \left( \frac{\sigma_{ik} + \sigma_{il}}{2\sqrt{\sigma_{ik}\sigma_{il}}} \right) \quad (4.10)$$

Η ταξινόμηση λοιπόν γίνεται αποτιμώντας την απόσταση *D* του χαρακτηριστικού διανύσματος  $V_k$  της κάθε  $R_i$  περιοχής από το πρότυπο χαρακτηριστικό διάνυσμα της αντίστοιχης κατηγορίας μεγέθους. Κάθε υποψήφια περιοχή προσώπου  $R_i$  ταξινομείται με βάση την απόσταση *D* του χαρακτηριστικού της διανύσματος από το πρότυπο διάνυσμα (της ίδιας πάντα κατηγορίας προσώπων). Η περιοχή  $R_i$  θεωρείται πρόσωπο αν η *D* είναι κάτω από ένα κατώφλι  $T_{HD}$ , ενώ διαφορετικά θεωρείται ως μη πρόσωπο. Οι τιμή του κατωφλιού  $T_{HD}$  είναι 6.0 για την κατηγορία των μεσαίων προσώπων και 8.0 για τα μεγάλα πρόσωπα. Τα κατώφλια αυτά προέκυψαν και κρίθηκαν καταλληλότερα κατόπιν ενός μεγάλου αριθμού πειραμάτων για κάθε μία από τις δύο κατηγορίες.

Προβλήματα που μπορεί να προκύψουν λόγω επικάλυψης των περιοχών  $R_i$  λύνονται ως εξής. Το σύνολο των περιοχών προσώπου  $R_i$  που επικαλύπτονται περισσότερο από 25% και στις δύο διαστάσεις ταξινομούνται, από την μικρότερη στη μεγαλύτερη, σύμφωνα με τις κανονικοποιημένες αποστάσεις  $\frac{D}{hw}$  (όπου *h* το ύψος και *w* το πλάτος της περιοχής) σχετικά με το μέγεθος της περιοχής  $R_i$ . Μετά από αυτήν τη διαδικασία επιλέγεται η περιοχή που κατατάσσεται στην πρώτη θέση και οι υπόλοιπες απορρίπτονται.

## Κεφάλαιο 5

# Αποτελέσματα ανίχνευσης προσώπων

Η μέθοδος της ανίχνευσης προσώπων που αναπτύχθηκε αποτιμήθηκε χρησιμοποιώντας μία βάση εικόνων οι οποίες εξήχθησαν από διάφορα MPEG-1 βίντεο. Συγκεκριμένα περιέχει 100 εικόνες, οι περισσότερες από τις οποίες εξήχθησαν από ειδήσεις, διαφημίσεις, ταινίες και εξωτερικές λήψεις. Αυτές οι 100 εικόνες περιέχουν 104 πρόσωπα (με μέγεθος πάνω από το όριο που αναφέρθηκε), ενώ 10 εικόνες δεν περιέχουν καθόλου πρόσωπα. Οι εικόνες καλύπτουν τις περισσότερες περιπτώσεις τις οποίες καλείται να αντιμετωπίσει ο αλγόριθμος. Στον Πίνακα 5.1 αναφέρεται μία πιο αναλυτική περιγραφή του περιεχομένου των εικόνων της βάσης σύμφωνα με τα χαρακτηριστικά του προσώπου.

Στάση Προσώπου	Αριθμός
Μετωπική	48
Ημιμετωπική $\pm 15^\circ$	25
Πλάγια $\pm 45^\circ$	18
Στραμμένα $\pm 25^\circ$	13
<b>Σύνολο</b>	<b>104</b>

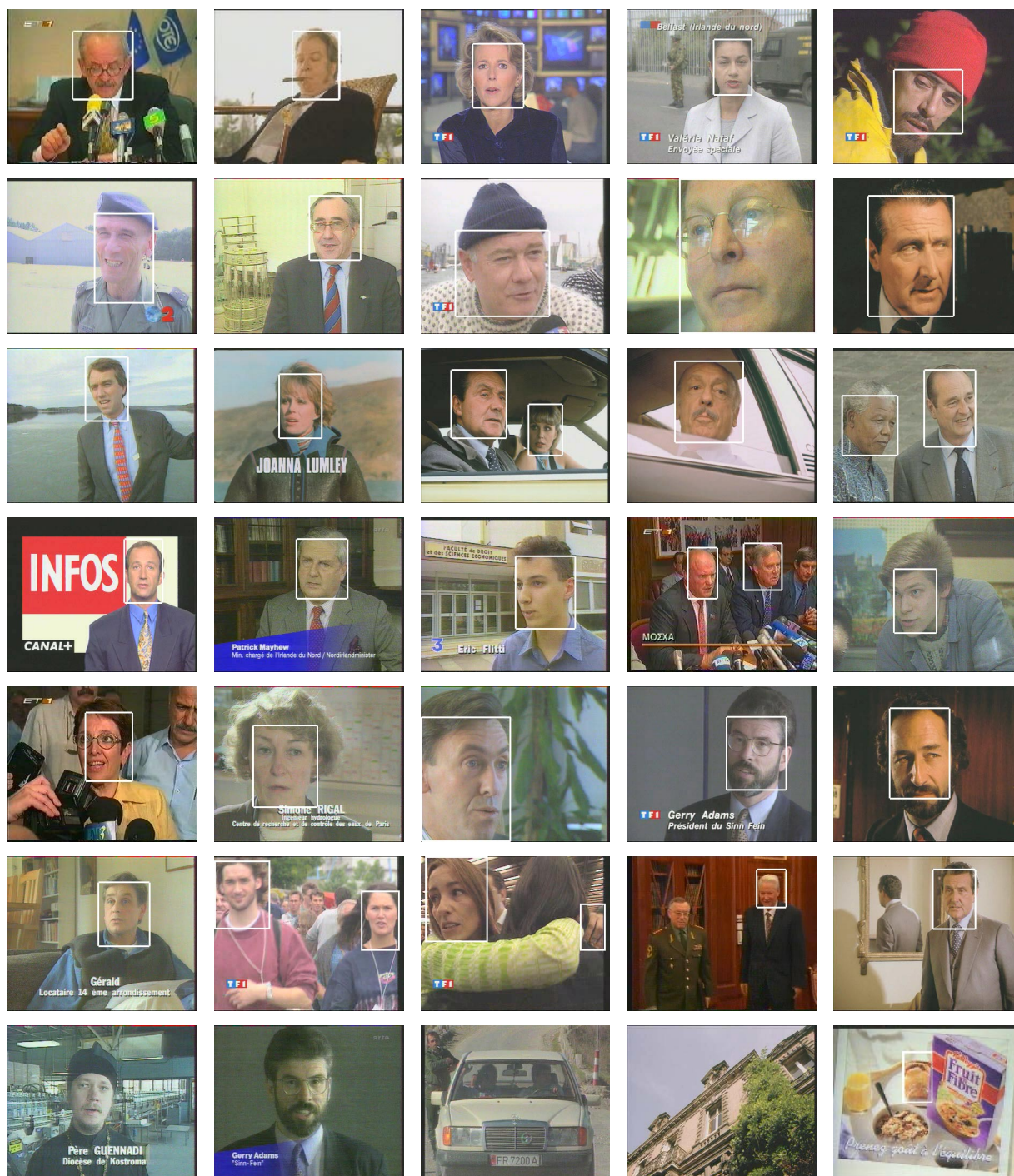
Πίνακας 5.1: Χαρακτηριστικά των προσώπων του συνόλου δοκιμών.

Στο Σχήμα 5.1 παρουσιάζουμε μερικά αποτελέσματα της ανίχνευσης. Τα παραδείγματα περιέχουν έγχρωμες εικόνες με πολλαπλά πρόσωπα διαφορετικών μεγεθών, χρωμάτων και θέσεων, καθώς και εικόνες που δεν περιέχουν πρόσωπα. Στον Πίνακα 5.2 αναφέρονται τα στατιστικά αποτελέσματα του αλγορίθμου, όπως προέκυψαν μετά την ολοκλήρωση των πειραμάτων χρησιμοποιώντας ολόκληρη τη βάση δοκιμών.

Ποιότητα Ανίχνευσης	Αποτελέσματα
Καλά προσδιορισμένα πρόσωπα	90 (91.83%)
Μέτρια προσδιορισμένα πρόσωπα	8 (8.17%)
<b>Ανιχνευμένα πρόσωπα</b>	<b>98 (94.23%)</b>
Μη ανιχνευμένα πρόσωπα	6 (5.76%)
Λάθος ανιχνεύσεις	20

Πίνακας 5.2: Αποτελέσματα ανίχνευσης προσώπων στο σύνολο δοκιμών.





Σχήμα 5.1: Αποτελέσματα ανίχνευσης προσώπων της προτεινόμενης μεθόδου από ένα υποσύνολο του συνόλου δοκιμών. Οι εικόνες εμπεριέχουν πρόσωπα διαφόρων βαθμών δυσκολίας που καλείται να αντιμετωπίσει ο αλγόριθμος.



## Κεφάλαιο 6

# Ανίχνευση χαρακτηριστικών προσώπου για απόρριψη λάθος ανιχνεύσεων

Εισαγωγικά αναφερθήκαμε στη σημασία ενός αποδοτικού συστήματος ανίχνευσης προσώπων. Επισημάνθηκε η έρευνα που έχει γίνει παγκοσμίως στην προσπάθεια ανάπτυξης συστημάτων αναγνώρισης προσώπων, τα οποία όμως υποθέτουν την ύπαρξη προσώπου στην υπό επεξεργασία εικόνα και μάλιστα σε θέση γνωστή και με πόζα μετωπική ή ημιμετωπική, ώστε να μπορέσουν να επιτύχουν. Στην προσπάθεια να απαλλαγούμε από αυτά τα πρότυπα αναπτύχθηκε μία μέθοδος κατάταξης του ανιχνευθέντος από το προηγούμενο σύστημα προσώπου στην κατηγορία μετωπική ή ημιμετωπική στάση, ώστε να τροφοδοτείται με αυτόν τον τρόπο απευθείας ένα σύστημα αναγνώρισης προσώπων. Επίσης με τη μέθοδο αυτή γίνεται μία προσπάθεια επαλήθευσης της αναγνώρισης έτσι ώστε να μπορούν μετέπειτα να απαλειφθούν λάθος ανιχνεύσεις, κάποιες από τις οποίες μπορούν να παρατηρηθούν στο Σχήμα 5.1.

Τα χαρακτηριστικά προσώπου παίζουν ένα σημαντικό ρόλο στην περιγραφή του ανθρώπινου προσώπου. Χαρακτηριστικά όπως τα μάτια, τα φρύδια, η μύτη, τα μάγουλα, το σαγόι μπορούν να χαρακτηρίσουν ισχυρά ένα ανθρώπινο πρόσωπο. Το μέγεθος και το σχήμα τους περιγράφουν το πρόσωπο και του αποδίδουν μία μοναδικότητα. Έτσι δεν πρέπει να μας φανεί παράξενο ότι τα χαρακτηριστικά προσώπου χρησιμοποιήθηκαν και χρησιμοποιούνται σε πολλές προσεγγίσεις αναγνώρισης και ανίχνευσης προσώπων, σε κάποιες με μεγάλη επιτυχία, σε άλλες με λιγότερη από την αρχικά αναμενόμενη. Θεωρώντας για παράδειγμα τη μέθοδο αναγνώρισης προσώπων που περιγράφεται παρακάτω, τα χαρακτηριστικά προσώπου χρησιμοποιούνται για τον προσδιορισμό περιοχών στο πρόσωπο, από τις οποίες παράγονται διανύσματα περιγραφής εξάγοντας απλά στατιστικά δεδομένα (βλ. και [12] σελ. 7-8). Τα πρόσωπα που χρησιμοποιήθηκαν για τα πειράματα ήταν εξαρχής γνωστά στην εικόνα προς επεξεργασία και έτσι ήταν και σίγουρο ότι βρίσκονται σε μετωπική ή ημιμετωπική στάση και μόνο. Το νέο πλεονέκτημα που προσφέρει η προτεινόμενη μέθοδος είναι, ότι όχι μόνο ξεδιαλέγει αυτόματα τα κατάλληλα πρόσωπα (δηλαδή αυτά που βρίσκονται σε μετωπική ή ημιμετωπική στάση) και τροφοδοτεί με αυτά τον αλγόριθμο αναγνώρισης, αλλά ταυτόχρονα είναι και γνωστές οι θέσεις των χαρακτηριστικών και έτσι μπορούν αμέσως να προσδιοριστούν οι επιθυμητές περιοχές. Λόγω της ποικιλίας και πολυπλοκότητας που τα διακρίνουν, τα χαρακτηριστικά προσώπου δεν είναι εύκολο να βρεθούν και να διαχωριστούν αυτόματα, αλλά μία αρκετά υποσχόμενη μέθοδος είναι αυτή που αναπτύχθηκε. Ο όρος “χαρακτηριστικά προσώπου” θα χρησιμοποιείται στο εξής για να δηλώνει τα δύο μάτια, τη μύτη και το στόμα.

Σε παλαιότερη εργασία [12] έχει αναπτυχθεί ένα σύστημα για την αναγνώριση προσώπων όταν αυτά βρίσκονται σε μετωπική ή ημιμετωπική στάση από μία βάση με εικόνες προσώπων. Η μέθοδος αποδίδει κάτω από σχεδόν σταθερή φωτεινότητα και βασίζεται στην πολυκαναλική ανάλυση κυματιδίων των εικόνων των προσώπων. Κάθε πρόσωπο εντοπίζεται πρώτα ακριβώς στην εικόνα και στη συνέχεια περιγράφεται από ένα υποσύνολο εικόνων, οι οποίες προέρχονται από την πολυκαναλική ανάλυση και απεικονίζουν συντελεστές κυματιδίων. Με αυτούς τους συντελεστές, οι οποίοι χαρακτη-

ρίζουν την υφή του προσώπου, δημιουργούνται συμπαγή διανύσματα περιγραφής χρησιμοποιώντας απλά στατιστικά δεδομένα. Μετά χρησιμοποιείται ένα ιδιαίτερα αποδοτικό στατιστικό μέτρο, το οποίο εξάγεται από την απόσταση Bhatthacharyya, για να ταξινομήσει τα διανύσματα περιγραφής σε κλάσεις ατόμων, δηλαδή “αναγνωρίζει” το πρόσωπο από τη βάση δεδομένων.

## 6.1 Αρχικοποίηση της μεθόδου

Η προτεινόμενη μέθοδος για την ανίχνευση των χαρακτηριστικών προσώπου χρησιμοποιεί την πολυκαναλική ανάλυση σε κυματιδιακά πλαίσια, όπως αυτή περιγράφεται στην ενότητα 4.1. Η κάθε εικόνα προς επεξεργασία αναλύεται μέχρι το 3ο επίπεδο για τους λόγους που αναφέρονται. Οι κυματιδιακοί συντελεστές που περιέχονται στα αποτελέσματα της κυματιδιακής ανάλυσης δίνουν ένα ισχυρό μέσο περιγραφής της υφής του προσώπου. Πριν ξεκινήσει η περιγραφή της μεθόδου κρίνεται σκόπιμο να αναφερθούν κάποιες υποθέσεις που έγιναν. Ο τομέας της ανίχνευσης και αναγνώρισης προσώπων, καθώς και τα ζητήματα που σχετίζονται με αυτόν, όπως η ανίχνευση χαρακτηριστικών προσώπων, είναι αρκετά πολύπλοκος και η θεώρηση κατάλληλων υποθέσεων εξ’ αρχής αποδεικνύεται πολύ χρήσιμη για τα επόμενα βήματα. Η βάση δεδομένων με τα πρόσωπα που χρησιμοποιήθηκαν για την εκπαίδευση και δοκιμή του αλγορίθμου είναι τα αποτελέσματα της μεθόδου ανίχνευσης προσώπων, όπως αυτή περιγράφηκε στις προηγούμενες ενότητες. Στατιστικά στοιχεία για τα πρόσωπα αυτά αναφέρονται στους Πίνακες 5.1 και 5.2. Ο αλγόριθμος εκπαίδευτηκε κυρίως με τα καλά προσδιορισμένα πρόσωπα. Μετά από διάφορα πειράματα με το εκπαιδευτικό αυτό σύνολο κρίθηκε σκόπιμο να προσδιοριστούν τρεις κατάλληλες κλίμακες προσώπων : μικρά, μεσαία και μεγάλα. Ένα πρόσωπο θεωρείται ότι ανήκει στην κατηγορία μικρό όταν το ύψος (height) και το πλάτος (width) του περιορίζονται από τις ανισότητες (6.1). Αντίστοιχα ένα πρόσωπο θεωρείται μεσαίο, μεγάλο όταν το ύψος και πλάτος του περιορίζονται από τις ανισότητες (6.2), (6.3). Οι διαστάσεις μετρώνται σε pixels σε εικόνες CIF διαστάσεων  $352 \times 288$ .

$$60 \leq height < 84 \quad , \quad 45 \leq width < 70 \quad (6.1)$$

$$84 \leq height < 137 \quad , \quad 70 \leq width < 109 \quad (6.2)$$

$$137 \leq height < 161 \quad , \quad 109 \leq width < 129 \quad (6.3)$$

Ακολουθώντας αυτούς τους κανόνες, κάθε πρόσωπο θα πρέπει να ταξινομηθεί σε μία από τις παραπάνω κατηγορίες. Στο Σχήμα 6.1 φαίνονται αντιπροσωπευτικά πρόσωπα για κάθε μία από τις κατηγορίες αυτές. Αν κάποιο πρόσωπο δεν πληρεί τους παραπάνω όρους, π.χ. είναι πολύ μικρό ή πολύ μεγάλο τότε αυτό απορρίπτεται. Θεωρούμε ότι κάτω από ένα ελάχιστο μέγεθος δεν υπάρχει αξιοποιήσιμη πληροφορία, για την περίπτωση που έχουμε να κάνουμε με πολύ μικρά πρόσωπα. Όσον αφορά τώρα τα πολύ μεγάλα, δεν μπορούμε παρά να θέσουμε κι εκεί έναν περιορισμό, εφόσον στη γενική περίπτωση η εικόνα προς επεξεργασία δεν περιορίζεται ως προς το μέγεθος και άρα το ίδιο ισχύει και για το πιθανό πρόσωπο που μπορεί να εμπεριέχεται σ’ αυτήν. Η μέθοδος, λοιπόν, θα πρέπει να εξειδικευτεί για έναν άπειρο αριθμό κατηγοριών μεγέθους, πράγμα αδύνατον. Στη συγκεκριμένη περίπτωση μας ενδιαφέρει η επεξεργασία εικόνων που προέρχονται από ακολουθίες βίντεο (MPEG-1). Οι εικόνες που εξάγονται από τέτοιες ακολουθίες έχουν ένα περιορισμένο μέγεθος και άρα και τα πρόσωπα που μπορεί να περιέχονται σ’ αυτές. Με αυτά τα δεδομένα λοιπόν, οι τρεις κλίμακες που περιγράφηκαν παραπάνω θεωρούνται ικανές να συμπεριλάβουν όλα τα πιθανά πρόσωπα (εκτός της ακραίας περίπτωσης όπου το πρόσωπο καταλαμβάνει το μεγαλύτερο μέρος της εικόνας).

Έχοντας καταλήξει στις κατηγορίες μεγέθους των προσώπων στη συνέχεια ορίζονται γεωμετρικές φόρμες για κάθε μία από αυτές. Και πάλι πειράματα μας έδωσαν τα κατάλληλα γεωμετρικά κριτήρια για να δημιουργήσουμε τις κατάλληλες αυτές φόρμες, οι οποίες θα χρησιμοποιηθούν για να συνδιάσουν τα χαρακτηριστικά προσώπου. Μία τέτοια φόρμα φαίνεται στο Σχήμα 6.2 και τα χαρακτηριστικά της οποίας αναφέρονται παρακάτω. Μελετώντας το σύνολο εκπαίδευσης αποκομίσθηκαν μέσες αποστάσεις που ενώνουν τα διάφορα χαρακτηριστικά. Αυτές οι αποστάσεις διαφέρουν φυσικά από κλίμακα σε κλίμακα, αποδείχθηκαν όμως αρκετά αξιόπιστες. Η πρώτη απόσταση ( $D_{eyes}$ ) σχετίζεται με τα δύο μάτια. Είναι η γεωμετρική απόσταση (σε pixels) που ενώνει τα κέντρα των δύο οφθαλμών. Οι δύο επόμενες αποστάσεις ενώνουν την πρώτη απόσταση με τα αντίστοιχα κέντρα της μύτης ( $D_{nose}$ ) και του στόματος ( $D_{mouth}$ ). Αυτές οι αποστάσεις είναι, όπως



Σχήμα 6.1: Αντιπροσωπευτικά πρόσωπα για τις κατηγορίες μικρό, μεσαίο, μεγάλο (από αριστερά προς δεξιά).

είπαμε, μέσες αποστάσεις, που αποκομίσθηκαν από τη βάση εκπαίδευσης για κάθε μία από τις τρεις κατηγορίες μεγεθών προσώπων. Οι διαστάσεις των ορθογωνίων παραλληλογράμμων που περιέχουν τα χαρακτηριστικά του προσώπου, τα οποία και φαίνονται επίσης στο Σχήμα 6.2, επιλέχθηκαν και πάλι ύστερα από πειράματα με το σύνολο εκπαίδευσης. Έτσι, τα ορθογώνια παραλληλόγραμμα που περιέχουν τα δύο μάτια και τη μύτη έχουν τις ίδιες διαστάσεις, αλλά είναι μικρότερα από αυτό που περιέχει το στόμα, αφού έτσι το απαιτούν τα σχήματα των χαρακτηριστικών. Όπως αναφέρθηκε, οι αποστάσεις που ενώνουν τα χαρακτηριστικά διαφέρουν από κατηγορία σε κατηγορία, το ίδιο και οι διαστάσεις των ορθογωνίων παραλληλογράμμων. Σύμφωνα με τις τρεις κατηγορίες θα υπάρχουν και τρεις αντίστοιχες γεωμετρικές φόρμες. Τα χαρακτηριστικά, λοιπόν, της κάθε γεωμετρικής φόρμας συνοψίζονται ως εξής :

$$\left. \begin{array}{l} D_{eyes} = 26 \quad , \quad D_{nose} = 19 \quad , \quad D_{mouth} = 33 \\ R1H = 15 \quad , \quad R1W = 19 \\ R2H = 13 \quad , \quad R2W = 25 \end{array} \right\} \quad (6.4)$$

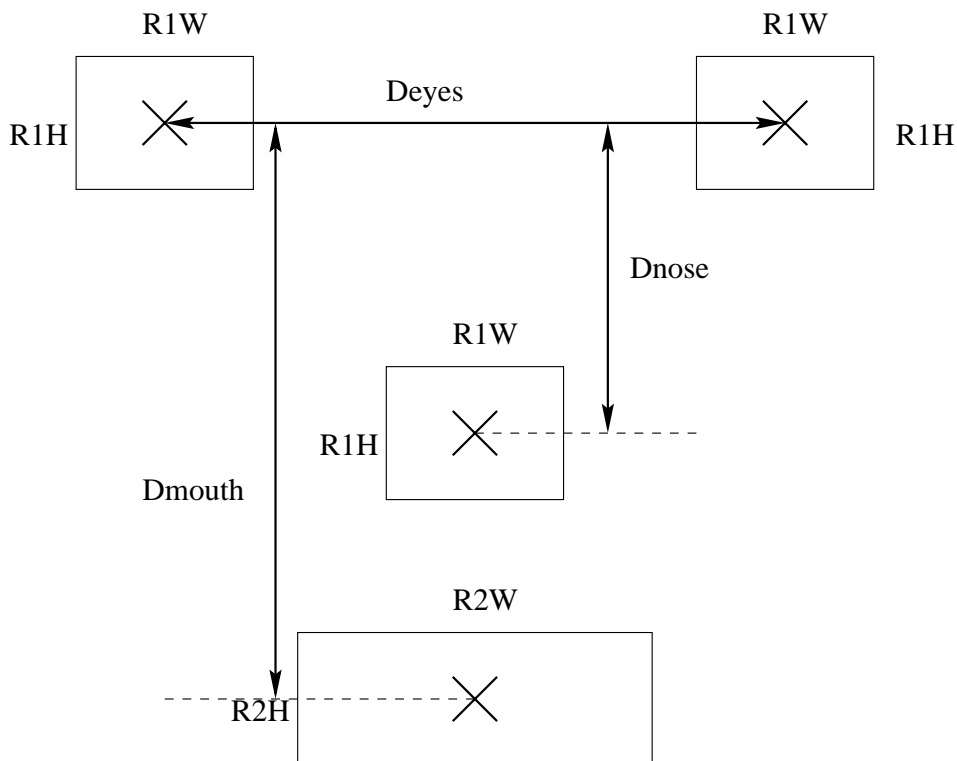
$$\left. \begin{array}{l} D_{eyes} = 39 \quad , \quad D_{nose} = 28 \quad , \quad D_{mouth} = 47 \\ R1H = 21 \quad , \quad R1W = 25 \\ R2H = 19 \quad , \quad R2W = 37 \end{array} \right\} \quad (6.5)$$

$$\left. \begin{array}{l} D_{eyes} = 57 \quad , \quad D_{nose} = 38 \quad , \quad D_{mouth} = 67 \\ R1H = 27 \quad , \quad R1W = 34 \\ R2H = 25 \quad , \quad R2W = 44 \end{array} \right\} \quad (6.6)$$

για τη μικρή 6.4, τη μεσαία 6.5 και τη μεγάλη 6.6 φόρμα αντίστοιχα, όπου  $R1$  είναι το ορθογώνιο παραλληλόγραμμα που περιέχει τα μάτια και τη μύτη και  $R2$ , αυτό που περιέχει το στόμα. Τα  $H$  και  $W$  αναφέρονται αντίστοιχα στο ύψος και στο πλάτος του αντίστοιχου ορθογωνίου παραλληλόγραμμου. Σ' αυτό το σημείο θα πρέπει να επισημάνουμε ότι το γεωμετρικό σχήμα που θα περιέχει--περικλείει τα χαρακτηριστικά προσώπου μπορεί να είναι οποιοδήποτε. Το ορθογώνιο παραλληλόγραμμα όμως κρίθηκε ως το πιο κατάλληλο, εφόσον μας παρέχει ευκολίες στους γεωμετρικούς υπολογισμούς και ταυτόχρονα δεν χάνει καμία πληροφορία του χαρακτηριστικού, αφού φροντίσαμε να είναι αρκετά μεγάλο για κάθε κατηγορία.

Χρησιμοποιώντας την πολυκαναλική ανάλυση σε κυματιδιακά πλαίσια αποκομίζουμε έναν αριθμό από εικόνες προσεγγίσεων και λεπτομερειών, που περιέχουν κυματιδιακούς συντελεστές, οι οποίοι περιγράφουν πλήρως την αρχική εικόνα, δηλαδή το πρόσωπο στη δική μας περίπτωση. Ως γνωστόν η ανάλυση μπορεί να συνεχιστεί μέχρι οποιοδήποτε επίπεδο. Όμως δεν έχει νόημα, εφόσον από κάποιο επίπεδο και κάτω δεν υπάρχει πλέον διαφορετική πληροφορία. Μετά από έρευνες καταλήξαμε στο συμπέρασμα ότι μία ανάλυση μέχρι το επίπεδο τρία είναι αρκετή. Αποδείχτηκε ότι αυτό το επίπεδο είναι καταλληλότερο για τις τρεις κατηγορίες μεγεθών των προσώπων, εφόσον περιέχει ίση πληροφορία γι' αυτές και δεν χάνει καθόλου πληροφορία για καμία κατηγορία.

Η υπολογιστική ταχύτητα της μεθόδου καθώς και η προσεγγιστική και μόνο ακρίβεια των ανιχνευ-



Σχήμα 6.2: Αποστάσεις που ενώνουν τα διάφορα χαρακτηριστικά προσώπου και τα ορθογώνια παραλληλόγραμμα που τα περικλείουν σχηματίζοντας τη γεωμετρική φόρμα στη γενικευμένη της μορφή.

θέντων χαρακτηριστικών που ζητείται συντελούν στη χρήση μόνο δύο πινάκων της κυματιδιακής ανάλυσης. Από το τελικό επίπεδο της ανάλυσης (επίπεδο 3), επιλέγονται οι πίνακες των συντελεστών των οριζοντίων και καθέτων λεπτομερειών της προσέγγισης του προηγούμενου επιπέδου (τα φύλλα AAH και AAV του δέντρου της πολυκαναλικής ανάλυσης πλαίσιας κυματιδίων) για να σχηματίσουν με την απλή έννοια του μέτρου το λεγόμενο μετρικό πίνακα  $M$  ως εξής :

$$M_{ij} = \sqrt{x^2 + y^2}, \quad \text{όπου} \quad x = \frac{\sum_{m=i-1}^{i+1} \sum_{n=j-1}^{j+1} hD_{mn}}{9} \quad \text{και} \quad y = \frac{\sum_{m=i-1}^{i+1} \sum_{n=j-1}^{j+1} vD_{mn}}{9} \quad (6.7)$$

όπου τα  $hD$  και  $vD$  συμβολίζουν τις οριζόντιες και κάθετες λεπτομέρειες του τρίτου επιπέδου της ανάλυσης. Παραδείγματα τέτοιων πινάκων μπορούν να βρεθούν στο Σχήμα 6.3, οι συντελεστές των οποίων βέβαια υποβλήθηκαν σε ανακλιμάκωση από 0 έως 255 για να γίνουν ορατοί την εικόνα.

Εφόσον δε γνωρίζουμε εξ αρχής την ακριβή θέση του προσώπου στην εικόνα, η εφαρμογή της φόρμας στο κάθε σημείο της αποτελεί μία υπολογιστικά χρονοβόρα διαδικασία, αλλά και δυσκολεύει περισσότερο τη διαδικασία ταξινόμησης. Για την αποφυγή αυτών των προβλημάτων επινοήθηκε ένας τρόπος να περιοριστούν τα σημεία εφαρμογής της γεωμετρικής φόρμας. Αξιοποιούμε το γεγονός ότι η πολυκαναλική ανάλυση σε κυματιδιακά πλαίσια τονίζει σημαντικά τα χαρακτηριστικά προσώπου. Στο μετρικό πίνακα, λοιπόν, σημεία που ανήκουν σε χαρακτηριστικά προσώπου θα έχουν ιδιαίτερα υψηλή τιμή ("μέτρο", με την έννοια του μετρικού πίνακα). Ένα χαλαρό κριτήριο χρησιμοποιείται για να ξεχωρίσει αυτά τα σημεία. Η όλη διαδικασία μεταφράζεται στην εφαρμογή ενός κατωφλιού  $C$  στο μετρικό πίνακα  $M_{ij}$ , έτσι ώστε όσα σημεία έχουν τιμή μεγαλύτερη απ' το κατώφλι να θεωρούνται θετικά, και τα υπόλοιπα αρνητικά. Η νέα αυτή πληροφορία συνοψίζεται σε μία δυαδική μάσκα, που μας δίνει τα σημεία στα οποία θα εφαρμοστεί το ταίριασμα της φόρμας.

Για να εφαρμοστεί η φόρμα στα σημεία αυτά ορίστηκε το σημείο αναφοράς να είναι το αριστερό

μάτι (όπως βλέπει κάποιος το πρόσωπο στην εικόνα, ή αλλιώς το δεξί μάτι του απεικονιζόμενου προσώπου). Αυτό σημαίνει ότι σε κάθε θετικό σημείο που εφαρμόζεται η γεωμετρική φόρμα, αντιστοιχείται πρώτα το αριστερό μάτι και στη συνέχεια με επαναληπτικό αναζητείται το ταίριασμα και των υπόλοιπων χαρακτηριστικών. Λόγω αυτού του γεγονότος δεν θα βρούμε ποτέ θετικά σημεία στα δεξιά και κάτω όρια της εικόνας, αφού αν προσπαθήσουμε να αντιστοιχίσουμε εκεί τη φόρμα, ακόμα κι αν κάποιο σημείο είναι θετικό, δεν πρόκειται ποτέ να βρεθούν τα υπόλοιπα χαρακτηριστικά, αφού ένα μέρος της φόρμας θα βρίσκεται εκτός εικόνας. Το κατώφλι  $C$  βρέθηκε μετά από πειράματα και η τιμή του μπορεί να ποικίλλει ανάλογα με την ποιότητα των εικόνων (αν είναι θορυβώδεις θα χρειαστεί σίγουρα ένα πολύ πιο χαμηλό κατώφλι απ' ό,τι όταν είναι άριστης ποιότητας) και από την υπολογιστική πολυπλοκότητα που θέλουμε να κερδίσουμε, χάνοντας όμως έτσι σημεία αναφοράς που μπορεί να αποβούν μοιραία (σε ένα βίντεο όπου ξέρουμε ότι τα ίδια πρόσωπα εμφανίζονται συχνά, π.χ. μία συνέντευξη, και τίθεται το θέμα της αναγνώρισης αυτών μπορούμε άνετα να θέσουμε ένα πολύ υψηλό κατώφλι αφού σε κάποια από τις εμφανίσεις ενός προσώπου σίγουρα θα ικανοποιείται). Σίγουρα δεν υπάρχει τρόπος να κρατηθούν στη μάσκα μόνο σημεία που ανήκουν σε χαρακτηριστικά προσώπου, αφού με την ανάλυση πλαισίων κυματιδίων ενισχύονται και άλλα χαρακτηριστικά, όπως ακμές, περιγράμματα, θόρυβος. Βέβαια η απόρριψη μπορεί να γίνει στη συνέχεια. Παραδείγματα των προαναφερθέντων δυαδικών μασκών μπορούν να βρεθούν στο Σχήμα 6.3.

## 6.2 Ελαστική αντιστοίχιση με φόρμες

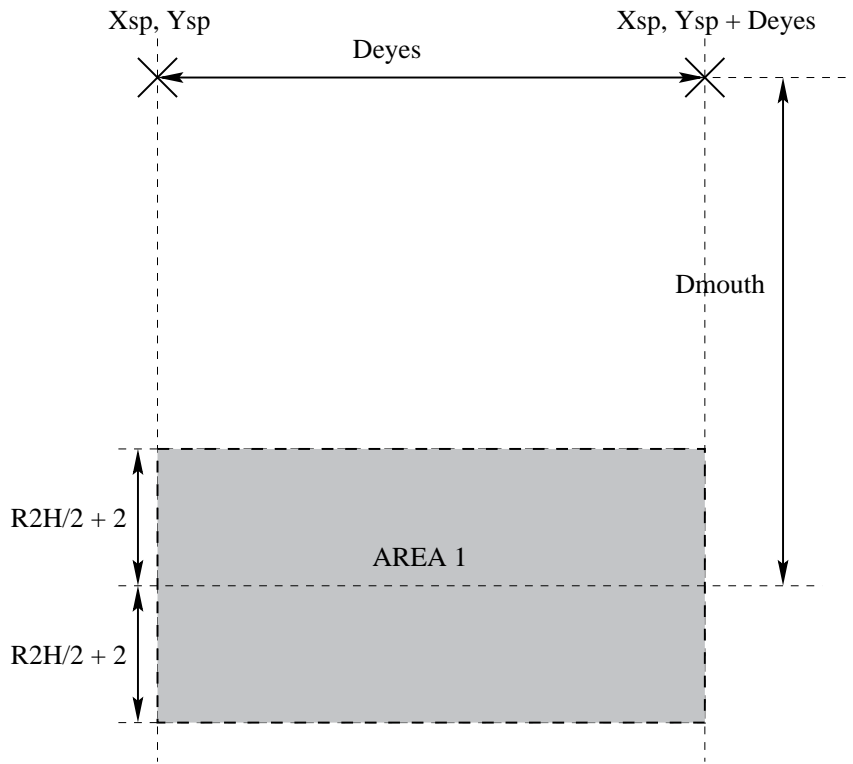
Στην προηγούμενη ενότητα είδαμε πώς από μία αρχική εικόνα εξάγεται μία δυαδική μάσκα στα θετικά σημεία της οποίας θα εφαρμοστεί η γεωμετρική φόρμα. Είπαμε επίσης ότι σημείο αναφοράς της γεωμετρικής φόρμας θεωρούμε το αριστερό μάτι. Έτσι σύμφωνα με το σημείο αναφοράς η φόρμα θα προσπαθήσει να αντιστοιχίσει και τα υπόλοιπα χαρακτηριστικά. Η μέθοδος ξεκινάει εφαρμόζοντας τη μικρότερη φόρμα με σημείο αναφοράς το αριστερό μάτι στο θετικό σημείο. Η αντιστοίχιση των σημείων γίνεται ελαστικά αυξομειώνοντας τις αποστάσεις σύμφωνα με κάποια σφάλματα, ώστε να καλύψουν σχεδόν όλες τις μετωπικές και ημιμετωπικές πόζες του ανθρώπινου προσώπου. Θυμίζουμε ότι το ζητούμενο αποτέλεσμα είναι η ταξινόμηση του προσώπου σε μετωπικό-ημιμετωπικό και μη, για να τροφοδοτηθεί σύστημα αναγνώρισης. Η ελαστική αυτή αντιστοίχιση αποφέρει έναν αριθμό διανυσμάτων που αποτελούνται από απλά στατιστικά κριτήρια για τις διάφορες περιοχές, όπως τα μάτια, τη μύτη, το στόμα, τα μάγουλα. Στη συνέχεια εφαρμόζονται η μεσαία και η μεγάλη φόρμα με μοναδικό περιορισμό αυτές να χωράνε στην εικόνα. Όλα τα διανύσματα περιγραφής που αποκομίζονται με αυτόν τον τρόπο κρατούνται ξεχωριστά για την κάθε φόρμα. Η διαδικασία αυτή επαναλαμβάνεται για όλα τα θετικά σημεία της δυαδικής μασκας. Παρακάτω περιγράφεται η διαδικασία της εξαγωγής διανυσμάτων.



Σχήμα 6.3: Εικόνες προσώπων, οι μετρικοί τους πίνακες (δεύτερη στήλη) και οι τελικές δυαδικές μάσκες που προκύπτουν ( τρίτη στήλη).

Στην περιγραφή της γεωμετρικής φόρμας αναφερθήκαμε στις αποστάσεις  $D_{eyes}$ ,  $D_{nose}$ ,  $D_{mouth}$  και στα ορθογώνια παραλληλόγραμμα που περικλείουν τα χαρακτηριστικά προσώπου (Σχήμα 6.2). Είναι φανερό ότι το σχήμα του στόματος είναι κατάλληλο για να περικλείεται από ορθογώνιο παραλληλόγραμμα. Πειράματα όμως που έγιναν με διάφορα μεγέθη διαφόρων διαστάσεων απέδειξαν ότι το ορθογώνιο παραλληλόγραμμα είναι επίσης πιο κατάλληλο και για να περικλείσει τα μάτια και τη μύτη.

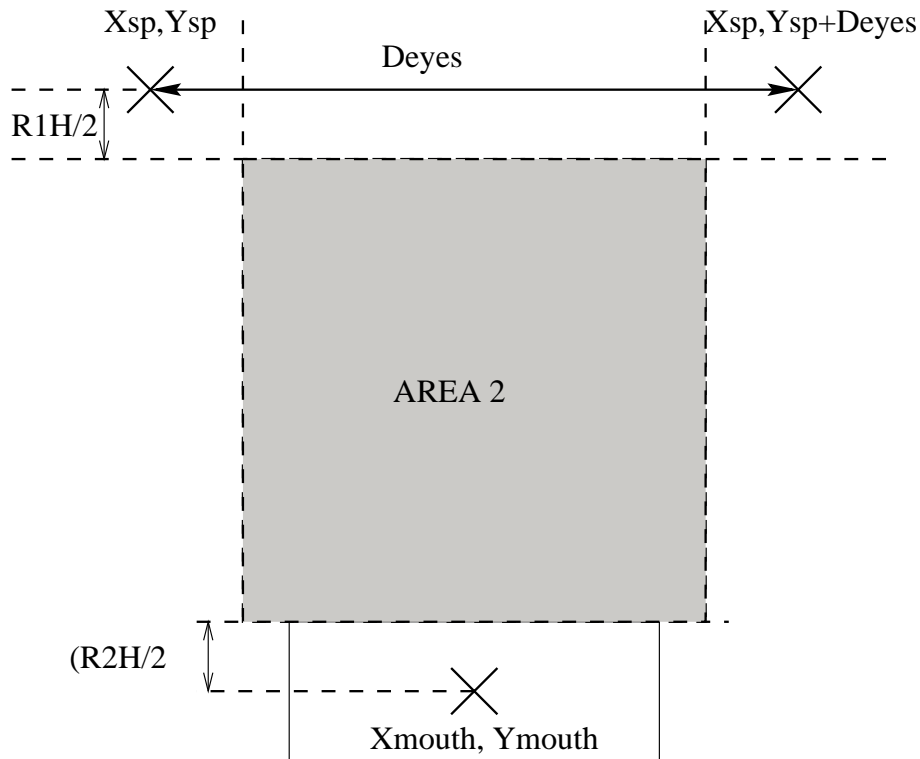
Θεωρώντας ένα σημείο αναφοράς (θετικό σημείο της μάσκας που αντιστοιχείται στο αριστερό μάτι) και το οποίο έχει συντεταγμένες  $(x_{sp}, y_{sp})$  η μέθοδος χρησιμοποιεί τις αντίστοιχες για κάθε μέγεθος φόρμας αποστάσεις που ορίζονται στις Εξισώσεις (6.4), (6.5), (6.6), για να χωρίσει την πιθανή περιοχή προσώπου που υπάρχει στην εικόνα σε τέσσερις περιοχές. Οι περιοχές αυτές αντιστοιχούν σε πιθανά υπάρχοντα χαρακτηριστικά προσώπου κι από αυτές θα εξάγει απλά στατιστικά δεδομένα για τη δημιουργία διανυσμάτων περιγραφής που θα αντιστοιχούν στο συγκεκριμένο σημείο αναφοράς. Η πρώτη περιοχή αφορά το στόμα και εξαρτάται μόνο από το σημείο αναφοράς και τις προαναφερόμενες αποστάσεις, Σχήμα 6.4. Η δεύτερη περιοχή αφορά τη μύτη και εξαρτάται από το σημείο αναφοράς, τις αποστάσεις αλλά και από το στόμα που βρέθηκε προηγουμένως, Σχήμα 6.5.



Σχήμα 6.4: Η περιοχή αναζήτησης του υποψηφίου στόματος (γκρίζο χρώμα).

Και στις δύο παραπάνω περιπτώσεις αναζήτησης χαρακτηριστικού προσώπου υπολογίζονται τα μέσα τοπικά μέγιστα για το στόμα και τη μύτη και αποθηκεύονται στα διανύσματα περιγραφής μαζί με τις γεωμετρικές θέσεις των σημείων που αντιστοιχούν στα μέγιστα αυτά. Αυτό γίνεται περνώντας το αντίστοιχο ορθογώνιο παραλληλόγραμμα που περιγράφει το κάθε αντίστοιχο χαρακτηριστικό πάνω από τις περιοχές αυτές. Αφού αναζητούνται πρόσωπα σε μετωπικές στάσεις, μόνο τα δικά τους χαρακτηριστικά θα πρέπει να ταιριάζουν με τις θέσεις της φόρμας, πάντοτε με κάποιο σφάλμα βέβαιο.

Πρώτα αναζητείται το στόμα. Υπολογίζεται στη γεωμετρική περιοχή ένα ( $AREA_1$ ) το μέσο τοπικό μέγιστο ( $LAM_m$ ) του στόματος από τους συντελεστές που προκύπτουν με μία ενίσχυση του μετρικού πίνακα  $M$ . Ο μετρικός πίνακας ενισχύεται στην περιοχή αυτή με ένα ποσοστό των οριζοντίων συντελεστών αφού το στόμα αναλύεται καλύτερα την εικόνα των οριζοντίων λεπτομερειών. Ο υπολογισμός αυτός γίνεται περνώντας πάνω από την περιοχή το κατάλληλο ορθογώνιο, η οποία όμως



Σχήμα 6.5: Η περιοχή αναζήτησης της υποψηφίας μύτης (γκρίζο χρώμα).

θα ενισχυθεί όπως αναφέρθηκε παραπάνω ως εξής :

$$Area_{1ij} = M_{ij} + 0.35|hD_{ij}| \quad (6.8)$$

όπου  $Area_{1ij}$  είναι η ενισχυμένη με τις οριζόντιες λεπτομέρειες τιμή του pixel  $i,j$  της περιοχής ένα. Το 35% της αντίστοιχης τιμής των οριζόντιων λεπτομερειών του pixel είναι αρκετή ενίσχυση. Το ίδιο ισχύει και για τη μύτη, αν αντικαταστήσουμε τις οριζόντιες λεπτομέρειες με τις κάθετες :

$$Area_{2ij} = M_{ij} + 0.35|vD_{ij}| \quad (6.9)$$

Θα βρίσκεται το μέσο τοπικό μέγιστο περνώντας ένα παράθυρο πάνω από την ενισχυμένη περιοχή δύο και θα βρίσκεται ταυτόχρονα και το σημείο της μύτης.

Σε αντίθεση με τις δύο περιοχές που περιγράφηκαν παραπάνω η περιοχή τρία ( $AREA_3$ ) δεν ενισχύεται αλλά μένει αυτή του μετρικού πίνακα  $M$ . Η περιοχή αυτή αντιστοιχεί στο δεξί μάτι (όπως βλέπουμε το πρόσωπο στην εικόνα ή αλλιώς στο αριστερό μάτι του απεικονιζόμενου προσώπου), Σχήμα 6.6. Στην περιοχή αυτή διασχίζουμε το ορθογώνιο που περιέχει το δεξί μάτι (βλ. και Σχήμα 6.2) και σε κάθε σημείο υπολογίζεται το μέσο μέγιστο (average maximum  $AM$ ), η συσχέτιση με το ορθογώνιο του αριστερού ματιού (το σημείο αναφοράς της φόρμας) και αποθηκεύονται στο διάνυσμα περιγραφής μαζί με την απόσταση και τη γωνία των σημείων των πιθανών ματιών και τις θέσεις τους. Επίσης αποθηκεύεται το μέσο μέγιστο του περιεχομένου του ορθογωνίου του σημείου αναφοράς. Για ένα, λοιπόν, σημείο αναφοράς έχουμε πολλά διανύσματα περιγραφής στα οποία όμως αλλάζουν τα στατιστικά του δεξιού ματιού, ενώ τα υπόλοιπα μένουν ίδια. Αυτό γιατί η εύρεση του δεύτερου ματιού είναι ιδιαίτερα δύσκολη και ευαίσθητη και δεν κρίθηκε σκόπιμο να προσεγγιστεί με ένα και μόνο στατιστικό στοιχείο όπως το στόμα και η μύτη, που είναι πιο εύκολα προσεγγίσιμα.

Η συσχέτιση του σημείου αναφοράς και του υποψηφίου δεξιού ματιού υπολογίζεται σε ορθογώνια λίγο μεγαλύτερα από αυτά που ορίστηκαν για να περικλείουν τα δύο μάτια (Σχήμα 6.2) και στα οποία υπολογίζονται τα μέσα μέγιστα ( $AM$ ), και αυτό γιατί θέλουμε να περιλάβουμε όσο το δυνατόν περισσότερη πληροφορία από τον περίγυρο των ματιών, αφού στην περίπτωση που αναζητάμε



(μετωπική ή ημιμετωπική στάση προσώπου) τα μάτια είναι αρκετά ξεκάθαρα :

$$CRH = \frac{RH + 6}{2} \quad , \quad CRW = \frac{RW + 6}{2} \quad (6.10)$$

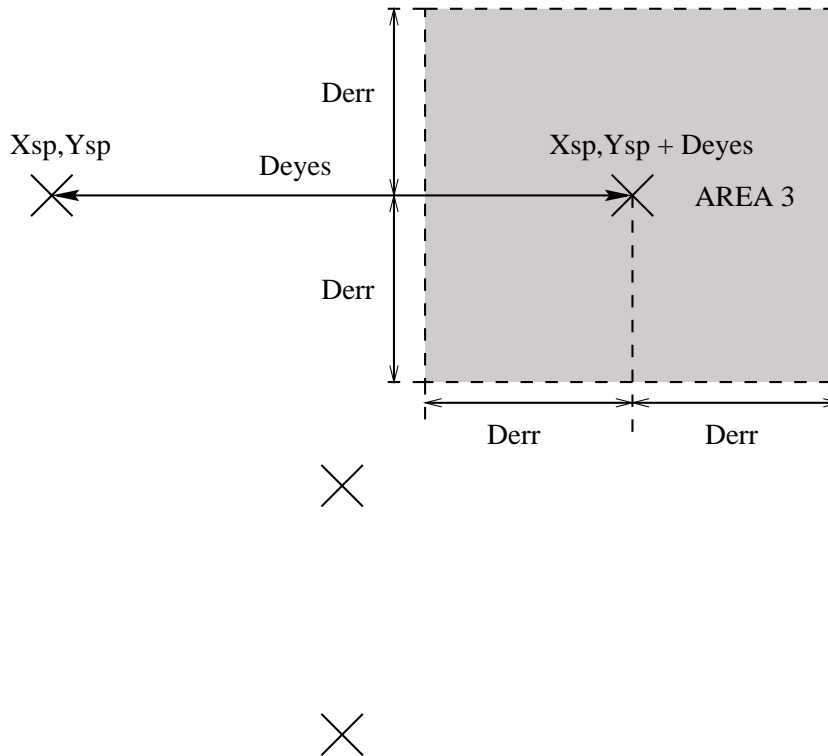
όπου  $CRH$ ,  $CRW$  είναι το ύψος και το πλάτος του ορθογωνίου συσχέτισης και ορίζονται με βάση το ορθογώνιο που περιέχει καθένα από τα δύο μάτια (με ύψος, πλάτος αντίστοιχα  $RH$ ,  $RW$ ). Τυπικά μεγέθη των ορθογωνίων που περιέχουν τα μάτια για τα διάφορα μεγέθη αναφέρονται στις Εξισώσεις (6.4), (6.5), (6.6). Η συσχέτιση δύο ίσων περιοχών  $T$ ,  $M$ , που περιέχουν η κάθε μία  $N$  pixels ορίζεται ως :

$$cor = \frac{covTM}{\sqrt{covTT \cdot covMM}} \quad (6.11)$$

όπου  $covTM$  είναι η συνδιασπορά :

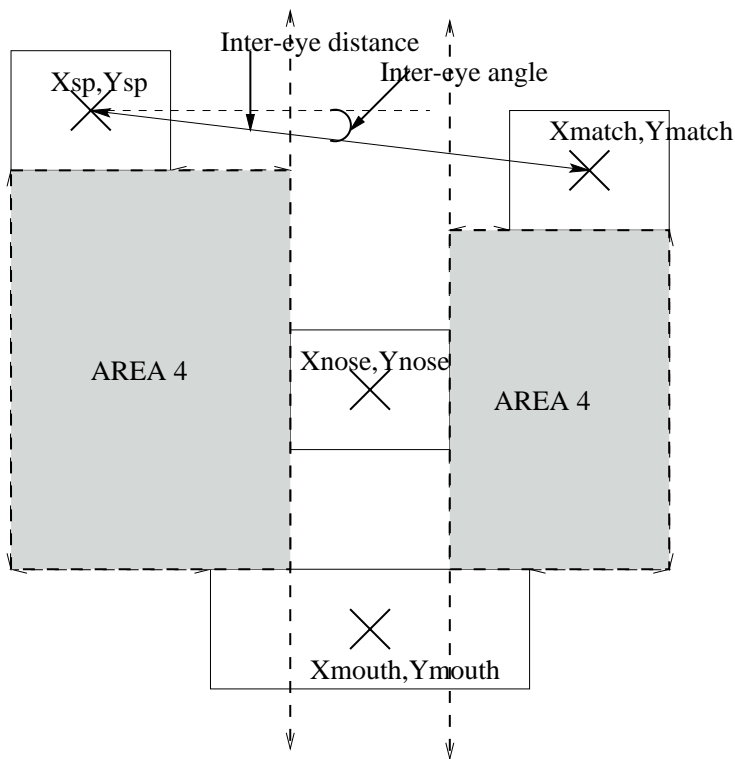
$$covTM = \frac{\sum_{i=1}^N [(T(t) - mean(T))(M(t) - mean(M))]}{N - 1} \quad (6.12)$$

η οποία είναι κανονικοποιημένη από  $-1$  (οι δύο περιοχές έχουν αντίθετη συσχέτιση) ως  $1$  (οι περιοχές είναι πλήρως συσχετισμένες), και το  $0$  δηλώνει πλήρως ασυσχέτιση.



Σχήμα 6.6: Η περιοχή των πιθανών δεξιών ματιών (γκρίζο χρώμα).

Τέλος, για κάθε ένα από τα υποψήφια δεξιά μάτια ορίζεται μία τέταρτη περιοχή ( $AREA_4$ ) που αφορά τα μάγουλα, Σχήμα 6.7. Η περιοχή αυτή εξαρτάται από όλα τα χαρακτηριστικά που βρέθηκαν μέχρι στιγμής. Τα μάγουλα ενός προσώπου σε μετωπική ή ημιμετωπική στάση είναι αρκετά ομοιογενείς. Έτσι υπολογίζεται στην περιοχή αυτή η διασπορά, η οποία θα πρέπει να είναι μικρή αν υπάρχει ομοιογένεια στην περιοχή αυτή. Η διασπορά αυτή επίσης περιλαμβάνεται στο διάλυσμα περιγραφής.



Σχήμα 6.7: Η υποψήφια περιοχή για τα πιθανά μάγουλα (γκρίζο χρώμα).

Τα διάφορα στατιστικά στοιχεία, που περιγράφηκαν παραπάνω, συνθέτουν διάνυσμα περιγραφής, όπως ήδη αναφέρθηκε :

$$V_i = [x_{sp} \ y_{sp} \ x_m \ y_m \ x_n \ y_n \ x_{mo} \ y_{mo} \ D \ A \ LAM_{sp} \ LAM_m \ Cor \ LAM_n \ LAM_{mo} \ Var] \quad (6.13)$$

όπου κάθε ζεύγος  $x, y$  αντιστοιχεί σε μία θέση στην εικόνα,  $D$  είναι η απόσταση των ματιών,  $A$  είναι η γωνία των ματιών,  $LAM$  το μέσο τοπικό μέγεθος,  $Cor$  η συσχέτιση μεταξύ των δύο ματιών,  $Var$  η διασπορά στα μάγουλα και τα  $sp, m, n, mo$  αναφέρονται αντίστοιχα στα χαρακτηριστικά : σημείο αναφοράς (αριστερό μάτι -- starting point), δεξί μάτι (σημείο αντιστοίχισης -- match point), μύτη (nose) και στόμα (mouth). Η γωνία των δύο ματιών ( $A$ ) πρέπει να βρίσκεται στο διάστημα  $[-30^\circ, 30^\circ]$  για να έχουμε την επιθυμητή μετωπική ή ημιμετωπική στάση. Τα διανύσματα  $V_i$  δημιουργούνται από την ελαστική αντιστοίχιση της φόρμας, όπως περιγράφηκε παραπάνω. Τονίζουμε και πάλι ότι για κάθε ένα θετικό σημείο της μάσκας γίνεται αντιστοίχιση από τη μικρότερη φόρμα στη μεγαλύτερη (όσο αυτή είναι βέβαια εφικτή). Για κάθε μία φόρμα υπολογίζονται τα διανύσματα περιγραφής. Τα διανύσματα περιγραφής για κάθε φόρμα είναι πολλά λόγω της δυσκολίας που αντιμετωπίζουμε με την εύρεση του δεξιού ματιού. Έχουμε, λοιπόν, πολλά διανύσματα με ίδια τα στοιχεία  $x_{sp}, y_{sp}, x_n, y_n, x_{mo}, y_{mo}, LAM_{sp}, LAM_n, LAM_{mo}$ , αλλά με διαφορετικά τα  $x_m, y_m, D, A, LAM_m, Cor, Var$  αφού λαμβάνουμε υπόψη ανά δύο όλα τα σημεία της περιοχής τρία για να έχουμε όσο το δυνατόν καλύτερα αποτελέσματα.

### 6.3 Ταξινόμηση των διανυσμάτων περιγραφής

Ο σκοπός της μεθόδου αυτής είναι να βρεί τα πρόσωπα σε μετωπική ή ημιμετωπική στάση και να απορρίψει τα υπόλοιπα καθώς και λάθος ανιχνεύσεις. Τα διανύσματα περιγραφής, που περιγράφηκαν παραπάνω, πρέπει λοιπόν να ταξινομηθούν, ώστε να προκύψει το ζητούμενο. Το στάδιο αυτό αναλαμβάνει να ταξινομήσει τα διανύσματα χρησιμοποιώντας μία σειρά κριτηρίων που αποκομίστηκαν από τη βάση εκπαίδευσης. Εφαρμόζει τα κριτήρια στα διανύσματα και σταδιακά απορρίπτει

όσα δεν τα εκπληρούν, ενώ τα εναπομείναντα θα μας δώσουν τρία σύνολα διανυσμάτων, ένα για κάθε κατηγορία μεγέθους, από τα οποία θα κρατηθεί το αριθμητικά υπερέχον σύνολο, σύμφωνα με το οποίο θα προσδιοριστούν τα χαρακτηριστικά προσώπου. Όπως είπαμε η όλη μέθοδος χρησιμοποιείται για να τροφοδοτήσει ένα σύστημα αναγνώρισης το οποίο λειτουργεί αυτόνομα και υποθέτει μόνο την ύπαρξη προσώπου στην εικόνα και τη θέση του. Έτσι η ανίχνευση των χαρακτηριστικών δεν παίζει μεγάλη σημασία και άρα δεν προσδοκούμε να τα προσδιορίσουμε με πολύ ακρίβεια. Θέλουμε να ξέρουμε μόνο αν το πρόσωπο είναι μετωπικό-ημιμετωπικό ή όχι (αρκετά στραμμένο πρόσωπο ή ανύπαρκτο στην εικόνα πρόσωπο). Ακολουθούν τα στάδια εφαρμογής των κριτηρίων για την ταξινόμηση των διανυσμάτων.

Τα στοιχεία που συνθέτουν ένα διάνυσμα περιγραφής περνούν μία σειρά από ελέγχους, οι οποίοι κυρίως είναι έλεγχοι κατωφλίων. Από τη βάση εκπαίδευσης αποκομίσθηκαν κατώφλια, τα οποία πρέπει να ικανοποιούν τα στοιχεία, για κάθε κατηγορία μεγέθους. Πρέπει να σημειώσουμε ότι τα τέσσερα τοπικά μέσα μέγιστα που περιλαμβάνει ένα διάνυσμα ( $LAM_{sp}$ ,  $LAM_m$ ,  $LAM_n$ ,  $LAM_{mo}$ ) υπόκεινται σε κανονικοποίηση και χρησιμοποιούνται για τους υπολογισμούς που ακολουθούν οι νέες τους τιμές ( $LAM'_{sp}$ ,  $LAM'_m$ ,  $LAM'_n$ ,  $LAM'_{mo}$ ):

$$meanLAM = \frac{LAM_{sp} + LAM_m + LAM_n + LAM_{mo}}{4} \quad (6.14)$$

$$LAM'_{sp} = \frac{LAM_{sp}}{meanLAM}, LAM'_m = \frac{LAM_m}{meanLAM}, LAM'_n = \frac{LAM_n}{meanLAM}, LAM'_{mo} = \frac{LAM_{mo}}{meanLAM} \quad (6.15)$$

οπότε και το διάνυσμα περιγραφής γίνεται :

$$V_i = [x_{sp} \ y_{sp} \ x_m \ y_m \ x_n \ y_n \ x_{mo} \ y_{mo} \ D \ A \ LAM'_{sp} \ LAM'_m \ Cor \ LAM'_n \ LAM'_{mo} \ Var] \quad (6.16)$$

Το πρώτο κριτήριο αφορά την ομοιογένεια στα μάγουλα. Είπαμε ότι τα μάγουλα είναι αρκετά ομοιογενείς περιοχές εφόσον το πρόσωπο έχει μία μετωπική ή ημιμετωπική στάση και άρα η διασπορά που μετράται στην περιοχή τέσσερα πρέπει να είναι αρκετά μικρή. Έτσι έχουμε :

$$V_i(16) < Variance_{thres} \quad (6.17)$$

Εφαρμόζοντας το κριτήριο αυτό παρατηρούμε ότι ο αριθμός των διανυσμάτων μειώνεται αρκετά, και άρα αποτελεί ένα πολύ ισχυρό κριτήριο.

Τα δύο μάτια ενός προσώπου είναι αρκετά όμοια χαρακτηριστικά και άρα τα τοπικά μέσα μέγιστα που υπολογίστικαν πρέπει να είναι αριθμητικά κοντά. Ελέγχουμε μία κανονικοποιημένη διαφορά των δύο τοπικών μέσων μεγίστων που αφορούν τα μάτια :

$$\frac{|V_i(11) - V_i(12)|}{\max\{V_i(11), V_i(12)\}} < Difference_{thres} \quad (6.18)$$

Επίσης, όσον αφορά τα μάτια, ελέγχεται και η συσχέτισή τους :

$$V_i(13) < Correlation_{thres} \quad (6.19)$$

Πρέπει να σημειώσουμε ότι παρόλο που αρχικά πιστεύαμε πολύ στο κριτήριο αυτό, στη συνέχεια αποδείχθηκε κάπως αναξιόπιστο, αφού δύο όμοιες περιοχές πολύ εύκολα παρουσιάζουν μεγάλη συσχέτιση μεταξύ τους. Τέτοιες περιοχές μπορεί να προέρχονται από το φόντο, αλλά κι από περιοχές του ίδιου του προσώπου, όπως είναι τα μάγουλα, τα μαλλιά, που καθιστούν το κριτήριο όχι απόλυτα αξιόπιστο, αλλά παρ' όλα αυτά βοηθάει στην όλη διαδικασία της ταξινόμησης.

Τέλος, συγκρίνουμε χρησιμοποιώντας την ευκλείδεια απόσταση τα τέσσερα τοπικά μέσα μέγιστα που αποκομίσθηκαν (στην κανονικοποιημένη τους μορφή) με τέσσερα πρότυπα μέγιστα που προέκυψαν από τη βάση εκπαίδευσης. Η κανονικοποιημένη μορφή αποδείχθηκε αρκετά χρήσιμη, αφού όλα τα πρόσωπα της βάσης εκπαίδευσης παρουσίασαν όμοιες τιμές. Το πρότυπο διάνυσμα που βρέθηκε για τα τοπικά μέσα μέγιστα είναι :

$$PV = [0.864759 \ 0.990789 \ 0.956072 \ 1.128381] \quad (6.20)$$

και η ευκλείδεια απόσταση των τεσσάρων τοπικών μέσων μεγίστων, στην κανονικοποιημένη τους πάντα μορφή, από τις τιμές του πρότυπου διανύσματος είναι :

$$ED_i = \sqrt{(PV(1) - V_i(11))^2 + (PV(2) - V_i(12))^2 + (PV(3) - V_i(14))^2 + (PV(4) - V_i(15))^2} \quad (6.21)$$

Το κριτήριο, λοιπόν, που θα πρέπει να ικανοποιείται είναι :

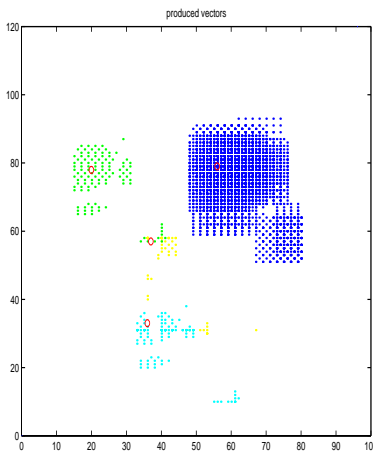
$$ED_i < Distance_{thres} \quad (6.22)$$

Τα Σχήματα 6.8, 6.9, 6.10 δείχνουν πόσο συνεισφέρει καθένα από τα παραπάνω κριτήρια στην απόρριψη διανυσμάτων που δεν αντιστοιχούν στα χαρακτηριστικά προσώπου. Στις εικόνες των σχημάτων φαίνονται με πράσινο τα σημεία που βρέθηκαν να αντιστοιχούν στο αριστερό μάτι, με σκούρο μπλε αυτά του δεξιού ματιού, με κίτρινο τα σημεία της μύτης και με ανοιχτό μπλε αυτά του στόματος. Επίσης σε κάθε εικόνα οι κόκκινοι κύκλοι δείχνουν τη θέση των πραγματικών χαρακτηριστικών προσώπου, ενώ στην τελευταία, με τα μαύρα τετράγωνα φαίνονται οι προσεγγίσεις που βρίσκει η μέθοδος υπολογίζοντας απλά την ενδιάμεση τιμή από τα σημεία που έμειναν για το κάθε χαρακτηριστικό.

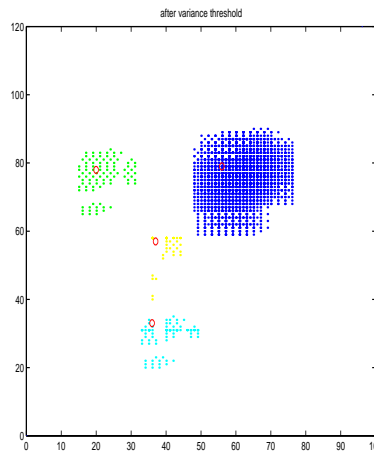
Μετά τη μείωση των διανυσμάτων μπορούμε να παρατηρήσουμε δύο περιπτώσεις. Μπορεί να μη μείνουν καθόλου διανύσματα, και άρα δεν έχουμε σημεία ή μπορεί να μείνουν. Στην πρώτη περίπτωση η έλλειψη διανυσμάτων περιγραφής σημαίνει ότι δεν υπάρχει πρόσωπο στην εικόνα, ή ότι το πρόσωπο που περιέχεται δεν είναι σε στάση μετωπική-ημιμετωπική και άρα δεν είναι κατάλληλο για περαιτέρω επεξεργασία, όπως π.χ. αναγνώριση (Σχήμα 6.11). Στη δεύτερη περίπτωση, μένουν διανύσματα περιγραφής μετά την όλη εφαρμογή των κριτηρίων και σκοπός μας είναι να διαπιστώσουμε αν πράγματι περιγράφουν ένα πρόσωπο σε μετωπική-ημιμετωπική στάση ή όχι.



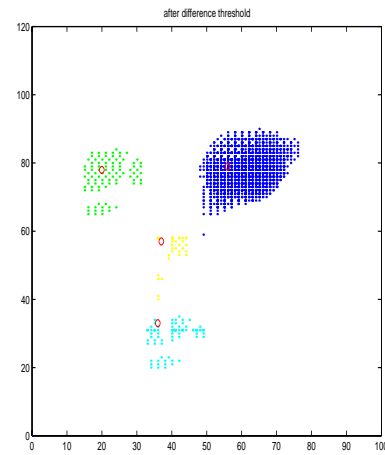
(α)



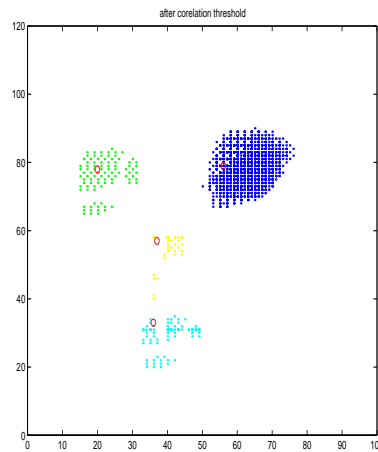
(β)



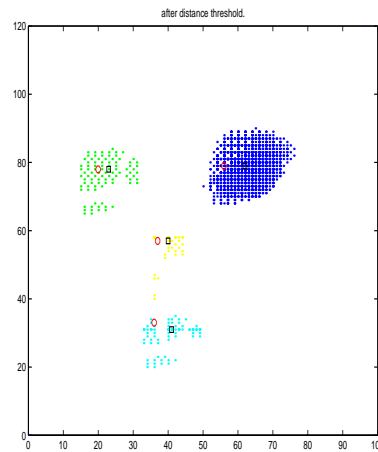
(γ)



(δ)



(ε)

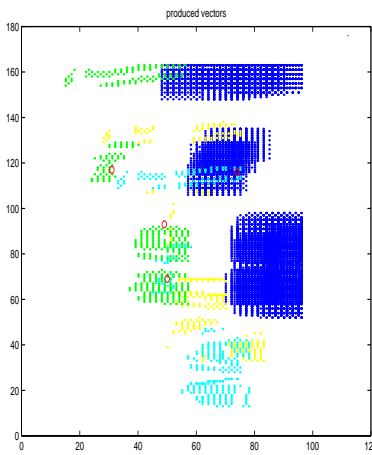


(στ)

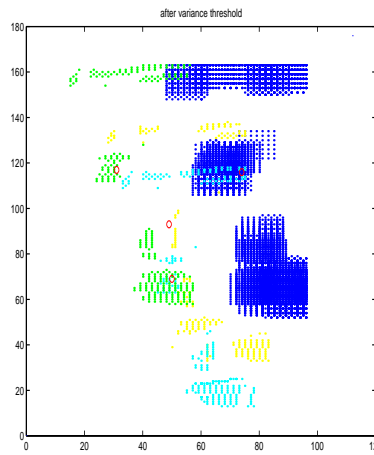
Σχήμα 6.8: Τα διανύσματα περιγραφής (β) που παράγονται από την αρχική εικόνα (α) και η διαδικασία της μείωσής τους. Με τη σειρά τα εφαρμοζόμενα κριτήρια είναι : η ομοιογένεια στα μάγουλα (γ), η διαφορά των τοπικών μέσων μεγίστων των ματιών (δ), η συσχέτιση των ματιών (ε) και η ευκλείδεια απόσταση (στ). Με τους κόκκινους κύκλους οι θέσεις των πραγματικών χαρακτηριστικών (μάτια, μύτη, στόμα) και με τα μαύρα τετράγωνα οι προσεγγίσεις της μεθόδου.



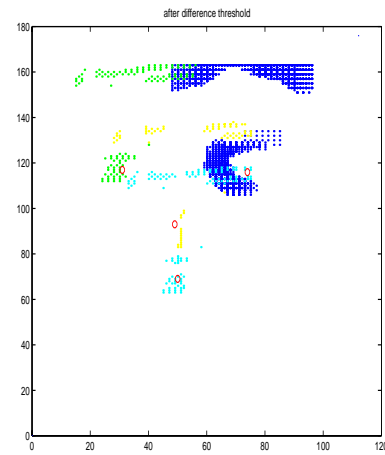
(α)



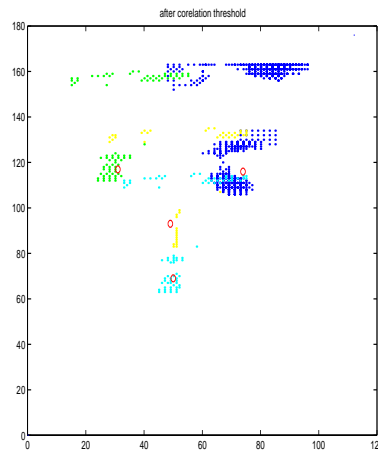
(β)



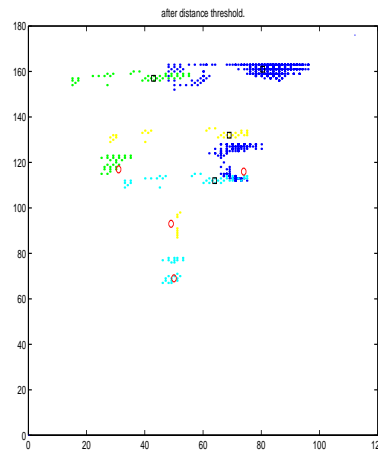
(γ)



(δ)



(ε)

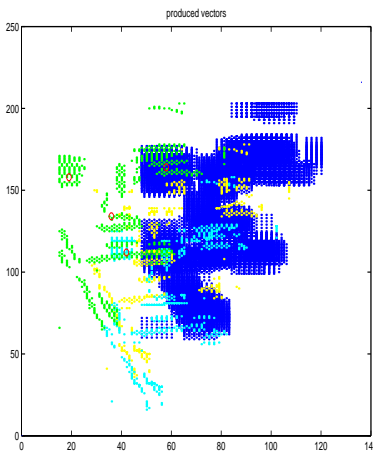


(στ)

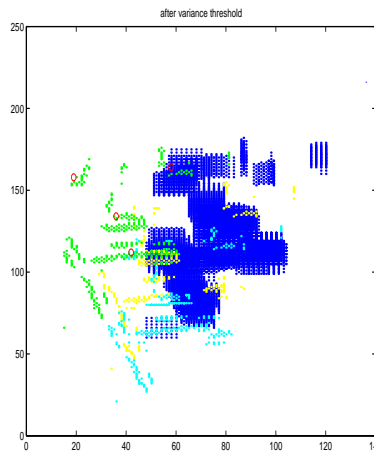
Σχήμα 6.9: Τα διανύσματα περιγραφής (β) που παράγονται από την αρχική εικόνα (α) και η διαδικασία της μείωσής τους. Με τη σειρά τα εφαρμοζόμενα κριτήρια είναι : η ομοιογένεια στα μάγουλα (γ), η διαφορά των τοπικών μέσων μεγίστων των ματιών (δ), η συσχέτιση των ματιών (ε) και η ευκλείδεια απόσταση (στ). Με τους κόκκινους κύκλους οι θέσεις των πραγματικών χαρακτηριστικών (μάτια, μύτη, στόμα) και με τα μαύρα τετράγωνα οι προσεγγίσεις της μεθόδου.



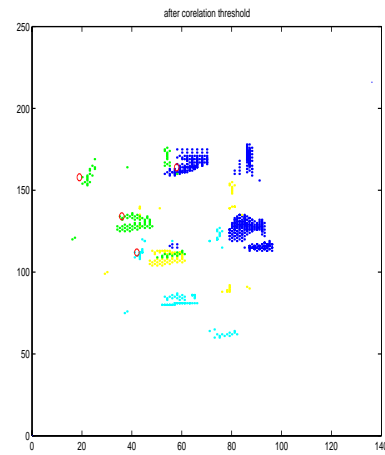
(α)



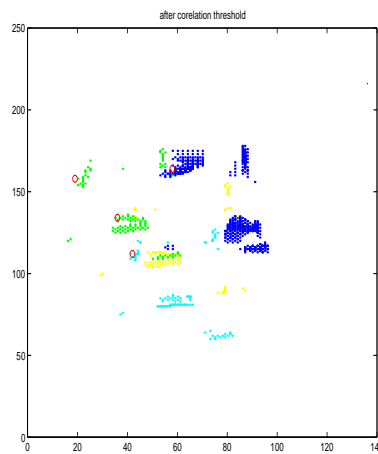
(β)



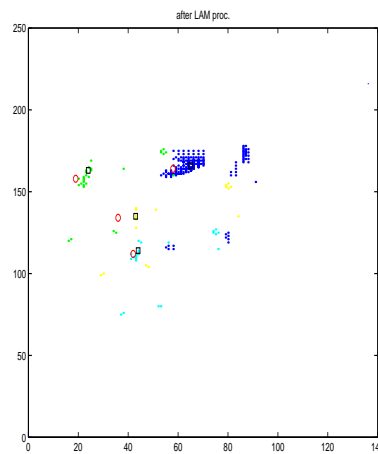
(γ)



(δ)



(ε)



(στ)

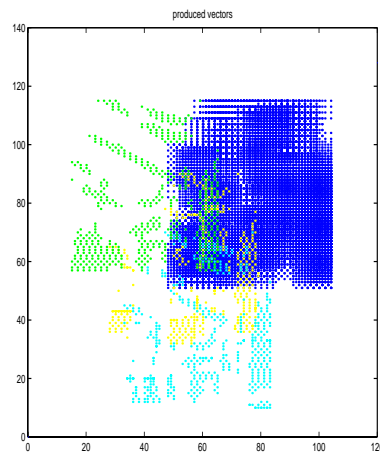
Σχήμα 6.10: Τα διανύσματα περιγραφής (β) που παράγονται από την αρχική εικόνα (α) και η διαδικασία της μείωσής τους. Με τη σειρά τα εφαρμοζόμενα κριτήρια είναι : η ομοιογένεια στα μάγουλα (γ), η διαφορά των τοπικών μέσων μεγίστων των ματιών (δ), η συσχέτιση των ματιών (ε) και η ευκλείδεια απόσταση (στ). Με τους κόκκινους κύκλους οι θέσεις των πραγματικών χαρακτηριστικών (μάτια, μύτη, στόμα) και με τα μαύρα τετράγωνα οι προσεγγίσεις της μεθόδου.



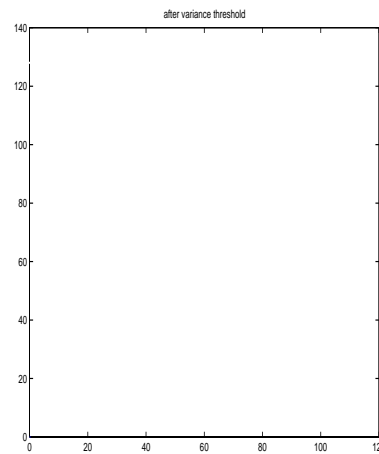
(α)



(β)



(γ)



(δ)

Σχήμα 6.11: Η αρχική εικόνα (α) δεν περιέχει πρόσωπο (πραγματικές διαστάσεις). Η αρχική εικόνα μεγενθυμένη (β) και τα διανύσματα περιγραφής που παράγονται (γ). Μετά κι από το πρώτο κριτήριο (δ) δεν μένουν πλέον διανύσματα. Η εικόνα απορρίπτεται.

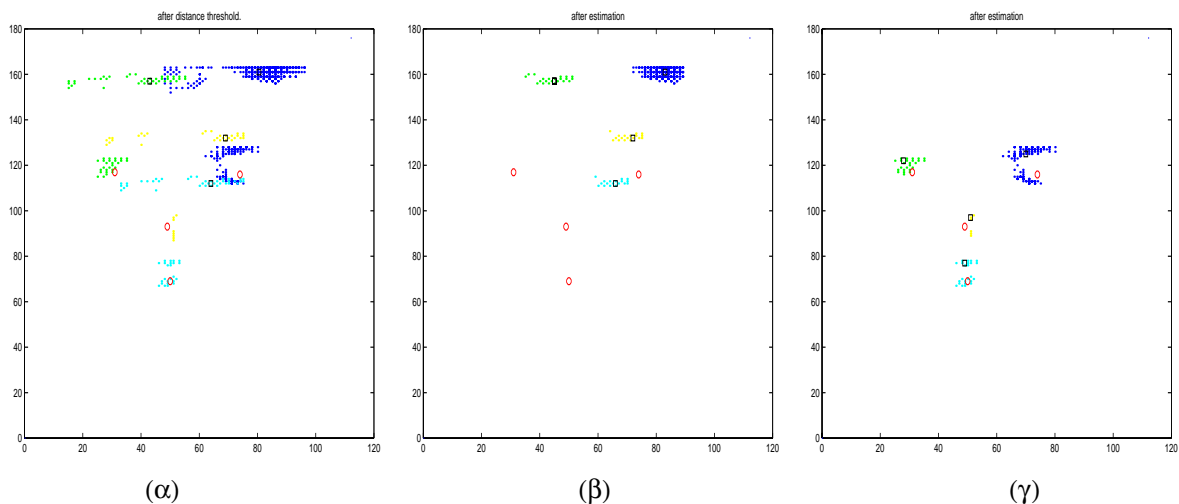
Στην περίπτωση που μένουν διανύσματα περιγραφής μετά το τέλος της εφαρμογής των κριτηρίων εφαρμόστηκαν δύο τρόποι για την ταξινόμησή τους. Και στις δύο περιπτώσεις τα σημεία που δίνουν τα διανύσματα περιγραφής αποτελούν όπως είδαμε σύνολα στον διδιάστατο χώρο της εικόνας. Με τον πρώτο τρόπο δίνουμε στο κάθε σημείο από ένα βάρος. Το βάρος είναι ο αριθμός των διανυσμάτων περιγραφής που έμειναν και που περιέχουν αυτό το σημείο (το σημείο μπορεί να περιέχεται πολλές φορές στα διανύσματα περιγραφής που έμειναν). Συνήθως παρατηρείται μεγάλη συγκέντρωση των σημείων στις περιοχές που βρίσκονται τα χαρακτηριστικά προσώπου. Για το λόγο αυτό χρησιμοποιούμε τις ενδιάμεσες τιμές του κάθε συνόλου για να προσεγγίσουμε τα χαρακτηριστικά. Αν οι ενδιάμεσες τιμές που υπολογίζονται δεν ικανοποιούν τη γεωμετρική φόρμα που αντιστοιχεί στην κατηγορία (θυμίζουμε ότι κρατάμε τα διανύσματα περιγραφής της κατηγορίας που έχει τα περισσότερα εναπομείναντα), αλλά και αν ο αριθμός των σημείων για κάθε σύνολο είναι υπερβολικά υψηλός, εφαρμόζεται επαναληπτικά μία μέθοδο σθεναρής εκτίμησης. Στο κάθε σύνολο  $c_i$  ( $i = 1, 2, 3, 4$  για το αριστερό μάτι, το δεξί μάτι, τη μύτη και το στόμα αντίστοιχα), κρατάμε τα σημεία που περιέχονται εντός του κύκλου με ακτίνα :

$$R_c = 1.2 \cdot \text{median}\{R_{c,j}\} \quad (6.23)$$

δηλαδή κρατάμε τα σημεία γύρω από τις ενδιάμεσες τιμές. Υπολογίζουμε τις νέες ενδιάμεσες τιμές και ελέγχουμε ξανά τη γεωμετρία και τον αριθμό των σημείων για κάθε σύνολο. Σε περίπτωση που απορριφθεί και πάλι συνεχίζουμε μέχρι να μη μείνει ένας ικανός αριθμός σημείων ο οποίος δεν θα μπορεί να μας δώσει ένα παράγοντα σιγουριάς της απόφασής μας. Αν δεν οδηγηθούμε σε σωστή γεωμετρία των σημείων γυρίζουμε στην αρχή και απορρίπτουμε όλα τα σημεία που κρατήθηκαν

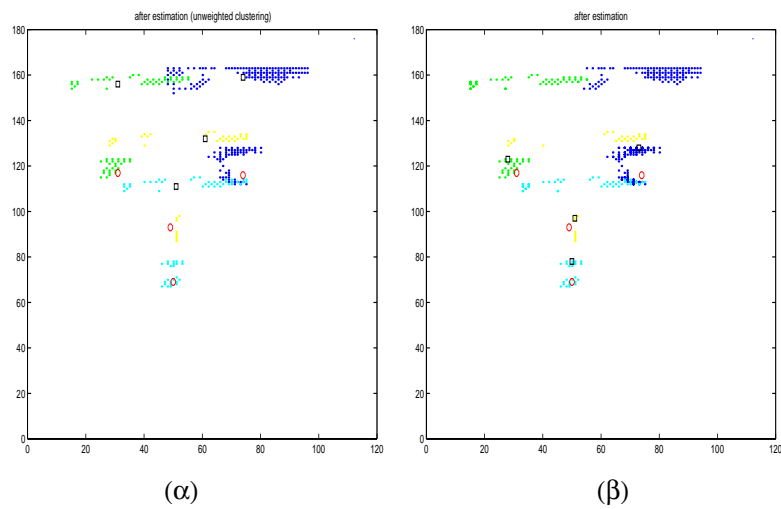


αρχικά εντός του κύκλου, εφαρμόζοντας την ίδια διαδικασία στα υπόλοιπα (Σχήμα 6.12). Αν οδηγηθούμε σε σωστή γεωμετρία κρατάμε την εικόνα για περαιτέρω επεξεργασία, ενώ στην αντίθετη περίπτωση απορρίπτεται. Ο βασικός σκοπός είναι να απορριφθούν πολλά σημεία, αλλά να μείνει ένας ικανός αριθμός για κάθε σύνολο που να προσδίδει σιγουριά στην απόφαση. Η όλη διαδικασία που περιγράφηκε παραπάνω μπορεί να εφαρμοστεί και χωρίς να προσδώσουμε βάρη στα σημεία με παρόμοια αποτελέσματα (Σχήμα 6.13).



Σχήμα 6.12: Η εφαρμογή των γεωμετρικών κριτηρίων και της σθεναρής εκτίμησης για την επίτευξη του στόχου. Η πρώτη, (α), εικόνα είναι η 22(στ). Παρατηρούμε ότι η ενδιάμεσες τιμές (μαύρα τετράγωνα) βρίσκονται μακριά από τα πραγματικά χαρακτηριστικά (κόκκινοι κύκλοι). Δεν ικανοποιείται η γεωμετρία και εφαρμόζεται το πρώτο στάδιο της σθεναρής εκτίμησης κρατώντας τα σημεία κοντά στις ενδιάμεσες τιμές. Επαναυπολογίζονται οι ενδιάμεσες τιμές. Και πάλι δεν ικανοποιείται η γεωμετρία (β), αλλά ταυτόχρονα τα σημεία που μένουν είναι λίγα (μη ικανός αριθμός) για να προσδώσουν σιγουριά στην απόφαση. Ξαναγυρίζουμε στην αρχή και απορρίπτονται όσα σημεία κρατήθηκαν αρχικά, κρατώντας τα υπόλοιπα (γ). Με τον υπολογισμό των ενδιάμεσων τιμών πετυχαίνουμε σωστή γεωμετρία και το αποτέλεσμα δεν απορρίπτεται.

Η μέθοδος αυτή βρίσκεται ακόμα στα αρχικά της στάδια. Σε περιπτώσεις που η εικόνα εισόδου δεν περιέχει πρόσωπο δείχνει αρκετά σταθερή συμπεριφορά με το να απορρίπτει όλα τα διανύσματα περιγραφής και μη αφήνοντας διανύσματα με την εφαρμογή των κριτηρίων από τα πρώτα κιόλας στάδια (βλ. Σχήμα 6.11). Σκοπός της είναι η απόρριψη λάθος ανιχνεύσεων και η τροφοδότηση συστήματος αναγνώρισης μόνο με πρόσωπα. Η αντιστοίχιση με φόρμες που εφαρμόζεται αποτέλεσε βάση ανάπτυξης μιας αξιόπιστης μεθόδου ανίχνευσης χαρακτηριστικών προσώπου, η οποία περιγράφεται στην επόμενη ενότητα.



Σχήμα 6.13: Εφαρμογή γεωμετρικών κριτηρίων και σθεναρής εκτίμησης χωρίς βάρη στα σημεία. Στην πρώτη, (α), εικόνα παρόλο που ικανοποιείται η γεωμετρία ο αριθμός των σημείων είναι μεγάλος για κάθε σύνολο. Εφαρμόζεται η μέθοδος σθεναρής εκτίμησης και λαμβάνουμε το αποτέλεσμα της δεύτερης εικόνας, όπου ικανοποιείται η γεωμετρία και ο αριθμός των σημείων είναι ικανός να προσδώσει σιγουριά.

## Κεφάλαιο 7

# Ανίχνευση χαρακτηριστικών προσώπων με ελαστική αντιστοίχιση με φόρμες

Τα χαρακτηριστικά προσώπου παίζουν, όπως είδαμε, ένα σημαντικό ρόλο τόσο στην αναγνώριση όσο και στην ανίχνευση του προσώπου. Παρακάτω παρουσιάζεται μία μέθοδος ανίχνευσης χαρακτηριστικών προσώπων σε έγχρωμες ψηφιακές εικόνες, χωρίς σκηνηικούς περιορισμούς. Με βάση την εύρεση των χαρακτηριστικών αυτών επιχειρείται και μία ανίχνευση του προσώπου.

### 7.1 Παρουσίαση μεθόδου

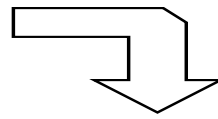
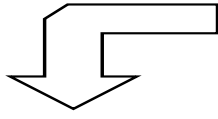
Η μέθοδος βασίζεται στην πολυκαναλική ανάλυση σε κυματιδικά πλαίσια και στην ελαστική αντιστοίχιση με φόρμες. Η κυματιδική ανάλυση είναι η ίδια με αυτήν που περιγράφεται στην ενότητα 4.1. Από τις εικόνες της ανάλυσης, που περιέχουν τους συντελεστές που περιγράφουν πλήρως την αρχική εικόνα, επιλέγονται ορισμένες για να εξαχθούν κριτήρια που θα χρησιμεύσουν στην περιγραφή των χαρακτηριστικών, όπως θα δούμε και παρακάτω. Για να ξεκινήσει η μέθοδος ανιχνεύει τα πιθανά μάτια που υπάρχουν στην εικόνα, όπως περιγράφεται στην επόμενη ενότητα. Τα σημεία αυτά θα αποτελέσουν σημεία αναφοράς για την αντιστοίχιση της φόρμας. Επιπλέον χρησιμοποιείται η κατάταξη του χρώματος του ανθρώπινου δέρματος, που πιθανώς να περιέχεται στην εικόνα, σχηματίζοντας μία δυαδική μάσκα. Η κατάταξη αυτή μπορεί να γίνει με μία από τις μεθόδους που περιγράφονται στην ενότητα 2. Η κατάταξη χρησιμεύει στην ελάττωση των προσπαθειών αντιστοίχισης της φόρμας, χρησιμοποιώντας σαν σημεία αναφοράς μόνο όσα περιβάλλονται από χρώμα του ανθρώπινου δέρματος, αφού τα χαρακτηριστικά βρίσκονται στο πρόσωπο και άρα γύρω τους θα υπάρχει το αντίστοιχο χρώμα. Περιορίζοντας έτσι τα σημεία αναφοράς κερδίζεται πολύ υπολογιστικό κόστος που κάνει τον αλγόριθμο εξαιρετικά γρήγορο. Στο κάθε σημείο αναφοράς αντιστοιχείται ελαστικά μία γεωμετρική φόρμα και εξάγονται διάφορα στατιστικά στοιχεία από σημεία που προκύπτουν από την αντιστοίχιση, που υποβάλλονται σε μία σειρά ελέγχων. Τα σημεία που μένουν τελικά ταξινομούνται με μία διαδικασία ομαδοποίησης ώστε να προκύψει το τελικό αποτέλεσμα. Μία συνοπτική περιγραφή της μεθόδου μπορεί να παρατηρηθεί στα Σχήματα 7.1, 7.2.

Η πολυκαναλική ανάλυση σε κυματιδικά πλαίσια περιγράφεται λεπτομερώς στην ενότητα 4.1. Τα φίλτρα είναι τα ίδια καθώς και το επίπεδο της ανάλυσης για τους λόγους που αναφέρονται. Στο σημείο αυτό, απλά θα θέλαμε να επισημάνουμε ότι χρησιμοποιούνται κάποιες από τις υποεικόνες της ανάλυσης για την εξαγωγή κριτηρίων. Οι εικόνες αυτές φαίνονται στο Σχήμα 7.3. Όσον αφορά την κατάταξη του χρώματος του ανθρώπινου δέρματος περιγράφηκε αναλυτικά στην ενότητα 2. Στη συγκεκριμένη μέθοδο χρησιμοποιήθηκαν τα αποτελέσματα κατάταξης του χρωματικού συστήματος HSV.

## 7.2 Ανίχνευση κοιλάδων -- μοντελοποίηση ματιού και στόματος

Το πρώτο στάδιο της μεθόδου αποτελείται από μία ανίχνευση των ματιών και του στόματος. Για το σκοπό αυτό χρησιμοποιείται η εικόνα ΑΗΗ της ανάλυσης, που αποδεικνύεται ότι περιγράφει ιδιαίτερα καλά τα χαρακτηριστικά αυτά, όπως παρατηρούμε και στο Σχήμα 7.4. Η ανίχνευση του ματιού και του στόματος γίνεται συσχετίζοντας τα πρότυπα σήματα των χαρακτηριστικών αυτών, τα οποία αποκομίσθηκαν μοντελοποιώντας τα χαρακτηριστικά αυτά από το σύνολό τους στη βάση εκπαίδευσης, σε ένα παράθυρο  $40 \times 40$  pixels με όλη την εικόνα των συντελεστών ΑΗΗ. Το παράθυρο επιλέχθηκε αρκετά μεγάλο ώστε να συμπεριλαμβάνει και τα δύο χαρακτηριστικά αυτά σε όλα τους τα μεγέθη (για το διαχωρισμό των επιμέρους χαρακτηριστικών σε τρία βασικά μεγέθη βλ. ενότητα 6.1). Αυτό έγινε για τη μείωση της υπολογιστικής πολυπλοκότητας, αφού έτσι χρειάζεται μία μόνο επανάληψη της συσχέτισης σε όλη την εικόνα αντί για τρεις που θα απαιτούσαν οι τρεις κατηγορίες μεγέθους των χαρακτηριστικών. Τα πειράματα απέδειξαν ότι τα αποτελέσματα ήταν αρκετά καλά, ώστε να δικαιολογούν την παραδοχή αυτή. Η μοντελοποίηση καθώς και όλα τα πειράματα έγιναν πάνω στη βάση εκπαίδευσης που χρησιμοποιήθηκε και για την ανάπτυξη του συστήματος ανίχνευσης προσώπων (ενότητες 2, 3, 4, 5, 6). Παρακάτω ακολουθούν λεπτομέρειες τόσο για τη μοντελοποίηση του ματιού, όσο και του στόματος.

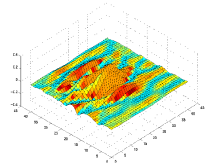
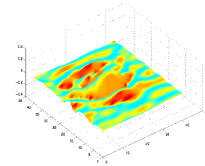
Αρχική εικόνα



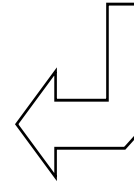
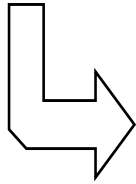
Κατάτμηση χρώματος



Κυματιδιακή ανάλυση (A, H, V)



Ελαστική αντιστοίχιση φόρμας



Ταξινόμηση χρησιμοποιώντας :

- Ανίχνευση ματιών και στόματος
- Κριτήρια διασποράν των συντελεστών της κυματιδιακής ανάλυσης για τα μάτια, τη μύτη και το στόμα
- Τοπικά κριτήρια για τα μάγουλα και τον περίγυρο του στόματος
- Απόσταση Bhattacharyya
- Συσχέτιση των ματιών και ομοιογένεια του όλου προσώπου



Αποτέλεσμα αντιστοίχισης : αρκετές καλές αποκρίσεις



Κόκκινο : αριστερό μάτι  
Μωβ : δεξί μάτι  
Πράσινο : μύτη  
Μπλε : στόμα

Σχήμα 7.1: Σχηματική αναπαράσταση της προτεινόμενης μεθόδου (1ο μέρος).

### Σύνολο ανιχνεύσεων



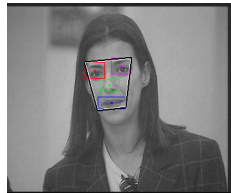
Κόκκινο : αριστερό μάτι  
Μωβ : δεξί μάτι  
Πράσινο : μύτη  
Μπλε : στόμα



### Ομαδοποίηση αποκρίσεων



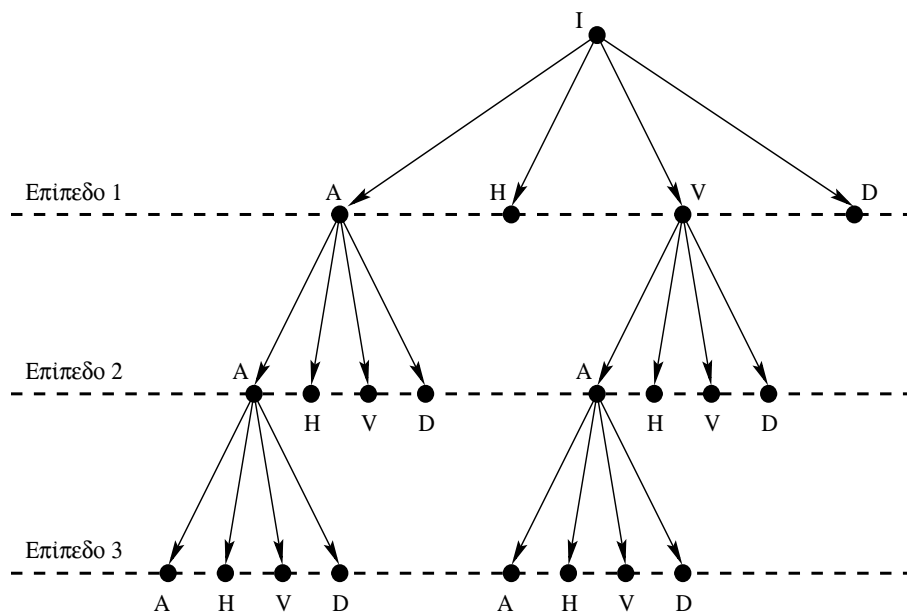
### Επιλογή της τελικής αντιστοίχισης



Κριτήρια: 

- Αριθμός ψήφων
- Ομοιογένεια

Σχήμα 7.2: Σχηματική αναπαράσταση της προτεινόμενης μεθόδου (2ο μέρος).



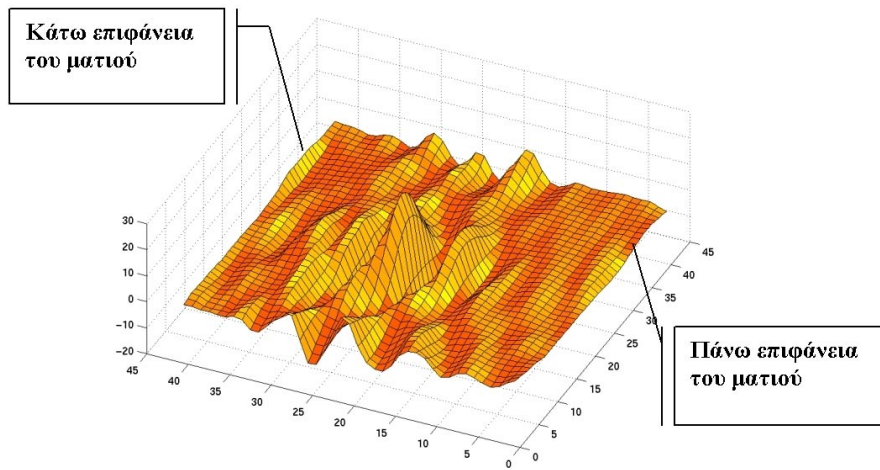
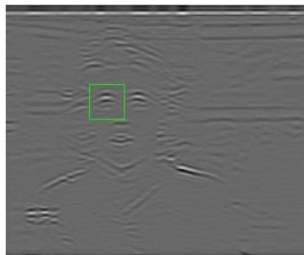
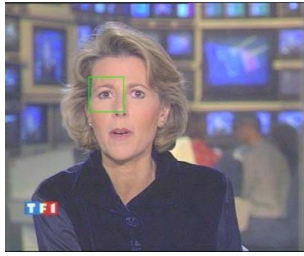
Σχήμα 7.3: Το δέντρο της κυματιδιακής ανάλυσης και οι συντελεστές των κυματιδιακών πλαίσιων σε γενική μορφή.



Σχήμα 7.4: Αρχική εικόνα (αριστερά), λεπτομέρειες AAA (κέντρο) και οριζόντια χαρακτηριστικά AHH (δεξιά) στο τρίτο επίπεδο ανάλυσης.

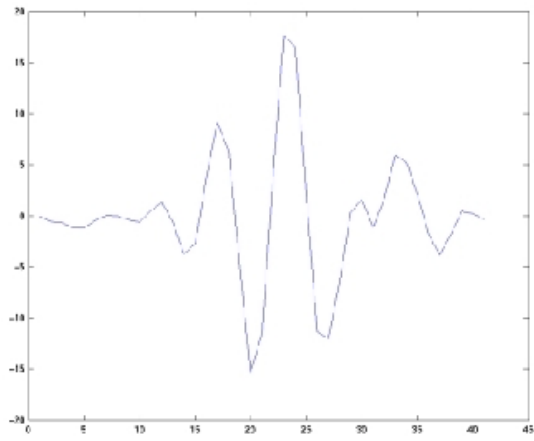
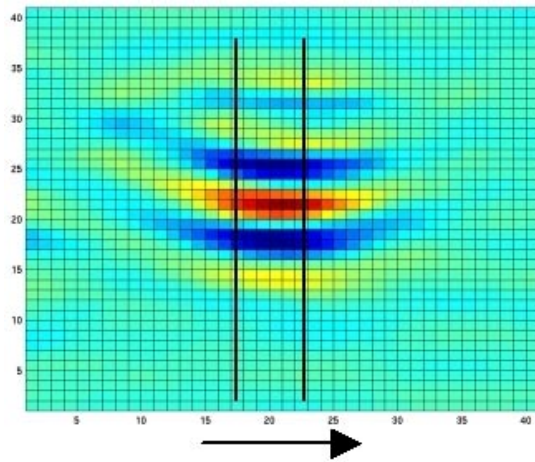
### 7.2.1 Μοντελοποίηση ματιού

Για τη μοντελοποίηση του ματιού χρησιμοποιήθηκαν όλα τα μάτια που συναντώνται στη βάση εκπαίδευσης και είναι καλά ορισμένα. Με σειρά πειραμάτων αποδείχθηκε ότι η εικόνα AHH των κυματιδιακών συντελεστών της ανάλυσης περιέχει την καταλληλότερη πληροφορία για το μάτι. Το παράθυρο  $40 \times 40$  επιλέχθηκε για να περιλαμβάνει όλα τα μεγέθη. Το παράθυρο αυτό δείνει μία τρισδιάστατη επιφάνεια του ματιού (Σχήμα 7.5), από την οποία αθροίζοντας σταθμισμένα στην οριζόντια κατεύθυνση σε μία ευρύ ζώνη του κέντρου του παραθύρου αποκομίζουμε ένα σήμα που περιγράφει χαρακτηριστικά το μάτι (Σχήμα 7.6). Η άθροιση γίνεται σταθμικά δίνοντας βάρος 1 στα κεντρικά σημεία της ζώνης και ελαττώνοντας το βάρος όσο απομακρυνόμαστε από το κέντρο ώστε να συμπεριληφθεί όσο το δυνατόν λιγότερη πληροφορία από τα πιθανά όρια-περιγράμματα του ματιού. Από όλα τα σήματα που ελήφθησαν σχηματίστηκε το μέσο σήμα που περιγράφει το μέσο μάτι, τα χαρακτηριστικά του οποίου χρησιμοποιήθηκαν σαν κριτήριο επιλογής (Σχήμα 7.7) κατά τη συσχέτιση του παραθύρου του μέσου σήματος διασχίζοντας την εικόνα AHH. Αποτελέσματα της ανίχνευσης ματιού με τη μέθοδο αυτή μπορούν να παρατηρηθούν στο Σχήμα 7.8.

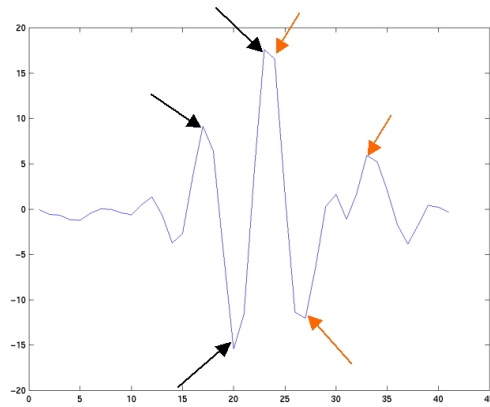


Σχήμα 7.5: Η 3Δ επιφάνεια του ματιού, όπως προκύπτει από τους συντελεστές ΑΗΗ της κυματιδιακής ανάλυσης.

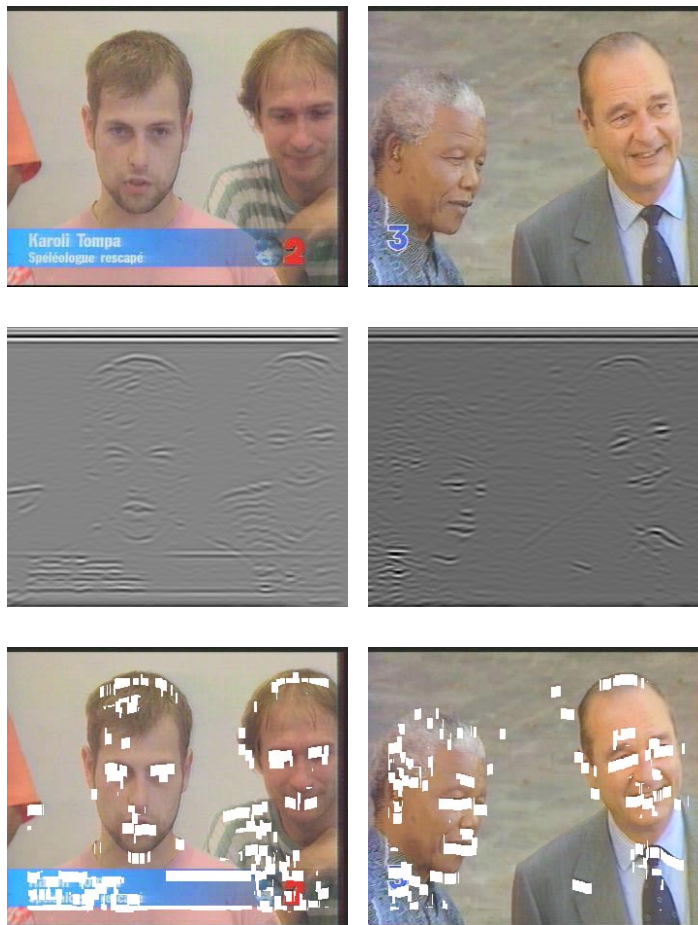




Σχήμα 7.6: Άθροιση κατά την οριζόντια κατεύθυνση και το μονοδιάστατο σήμα του ματιού που προκύπτει.



Σχήμα 7.7: Τα χαρακτηριστικά του σήματος (βελιάκια) που χρησιμοποιούνται ώστε να σχηματιστεί το μέσο μάτι και στα οποία τίθενται κατώφλια για την ανίχνευση.

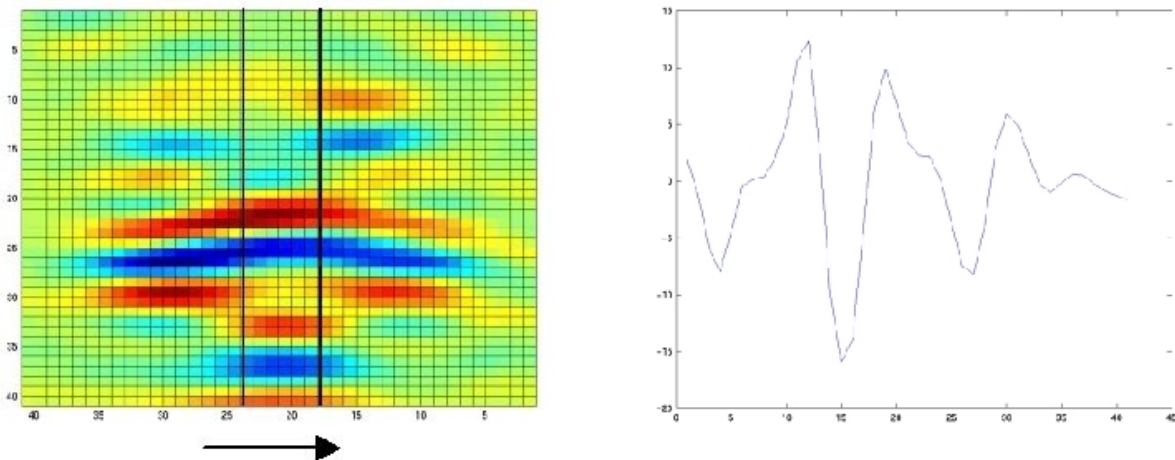


Σχήμα 7.8: Αρχική εικόνα (πάνω), εικόνα των συντελεστών ΑΗΗ της κυματιδιακής ανάλυσης, όπου και γίνεται η συσχέτιση με το μέσο σήμα ματιού (κέντρο) και αποτελέσματα της ανίχνευσης (κάτω). Με άσπρο χρώμα τα πιθανά μάτια που ανιχνεύτηκαν.

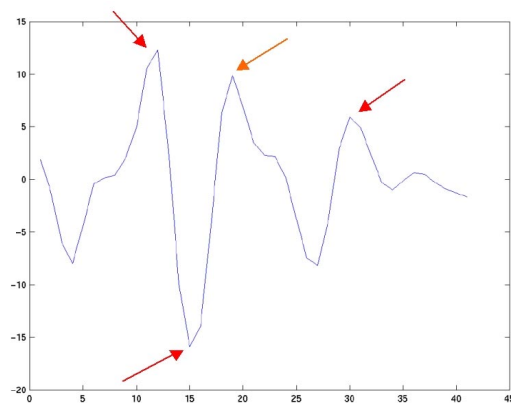
## 7.2.2 Μοντελοποίηση στόματος

Για την ανίχνευση του στόματος ακολουθήθηκε παρόμοια διαδικασία. Από το σύνολο εκπαίδευσης εξάχθηκαν όλα τα καλά ορισμένα στόματα, πάλι χρησιμοποιώντας ένα παράθυρο  $40 \times 40$  pixels, που όπως είπαμε είναι αρκετά μεγάλο για να περιέχει όλα τα μεγέθη. Από αυτά με σταθμική άθροιση στην οριζόντια κατεύθυνση παίρνουμε τα δισδιάστατα σήματα που τα περιγράφουν και από τα χαρακτηριστικά των οποίων κατασκευάστηκε το μέσο σήμα στόματος που χρησιμοποιείται στη συσχέτιση, Σχήματα 7.9, 7.10.

Στο σημείο αυτό σημειώνουμε ότι η μέθοδος επεξεργάζεται κάθε θετικό σημείο ξεχωριστά. Βρίσκοντας κάποιο θετικό σημείο στην εικόνα AAH, ένα υποψήφιο μάτι δηλαδή, αυτό ελέγχεται με βάση τη δυαδική μάσκα που προέκυψε από την κατάταξη του χρώματος του ανθρώπινου δέρματος. Μόνο αν το σημείο αυτό περιβάλλεται από ένα ικανό αριθμό σημείων που έχουν το χρώμα του ανθρώπινου δέρματος γίνεται δεκτό και συνεχίζεται η επεξεργασία, εφόσον τα μάτια βρίσκονται στο πρόσωπο. Διαφορετικά απορρίπτεται. Στη συνέχεια ακολουθεί η περιγραφή των σταδίων επεξεργασίας για ένα θετικό σημείο όπως προέκυψε από την ανίχνευση ματιών και που βέβαια πληρεί την παραπάνω προϋπόθεση, δηλαδή περιβάλλεται από χρώμα του ανθρώπινου δέρματος.



Σχήμα 7.9: Άθροιση και εξαγωγή του μονοδιάστατου σήματος του στόματος.



Σχήμα 7.10: Τυπικό μονοδιάστατο σήμα και τα χαρακτηριστικά που το περιγράφουν (βελάκια) και από τα οποία σχηματίζεται το μέσο σήμα στόματος.

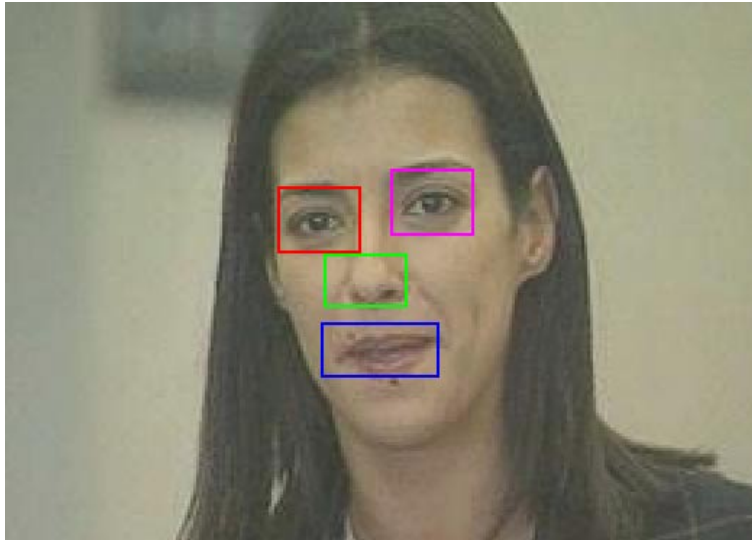
## 7.3 Ελαστική αντιστοίχιση φόρμας, δημιουργία διανυσμάτων περιγραφής και εφαρμογή κριτηρίων ελέγχου

Στην ενότητα αυτή περιγράφουμε τη μέθοδο ελαστικής αντιστοίχισης με μία γεωμετρική φόρμα ξεκινώντας με σημείο αναφοράς το αριστερό μάτι. Κάθε σημείο που βρέθηκε θετικό, με την έννοια που περιγράφεται στην προηγούμενη ενότητα, θεωρείται πλέον σημείο αναφοράς, δηλαδή αριστερό μάτι, και η γεωμετρική φόρμα αντιστοιχείται με βάση αυτό. Η ελαστική αντιστοίχιση της φόρμας δίνει ένα διάνυσμα περιγραφής του υποψηφίου προσώπου για τις θέσεις των υπόλοιπων πιθανών χαρακτηριστικών του προσώπου, δηλαδή του δεξιού ματιού, της μύτης και του στόματος, που υπολογίζονται σύμφωνα με τη γεωμετρία της φόρμας. Θα ακολουθήσουν κριτήρια ελέγχου, τόσο στο ίδιο το διάνυσμα, όσο και σε άλλα στατιστικά μέτρα, και εφόσον όλοι οι έλεγχοι αποδειχθούν θετικοί, η διάταξη αυτή κρατείται για περαιτέρω επεξεργασία στο τελευταίο στάδιο του συτήματος. Διαφορετικά απορρίπτεται. Σημειώνουμε την ύπαρξη μίας μόνο φόρμας η οποία όμως εφαρμόζεται, όπως θα δούμε και παρακάτω, ώστε να καλύψει όλες τις κατηγορίες μεγέθους των προσώπων.

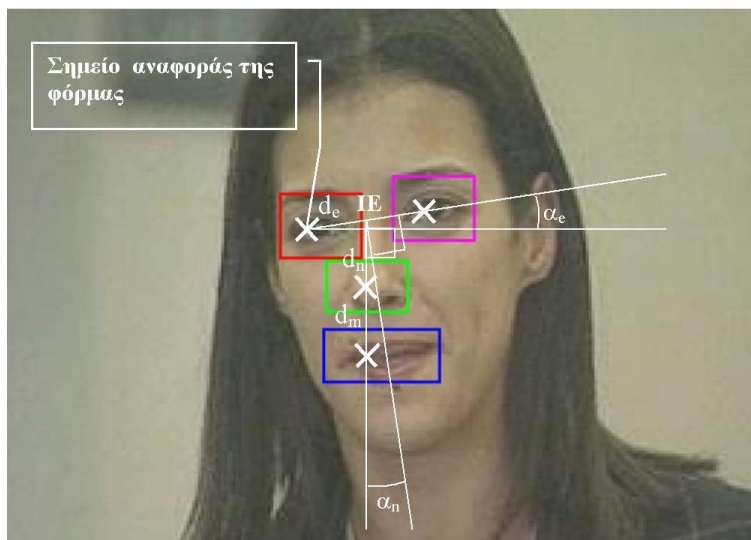
### 7.3.1 Περιγραφή της γεωμετρικής φόρμας

Η δημιουργία μιας γεωμετρικής φόρμας, με την αντιστοίχιση της οποίας θα γίνει απόπειρα εύρεσης των χαρακτηριστικών προσώπων, αποτελεί μία εξαιρετικά δύσκολη πρόκληση. Πρέπει να ληφθούν πολλοί παράγοντες υπόψη ώστε το αποτέλεσμα της ανίχνευσης να είναι όσο το δυνατόν καλύτερο. Το πρόσωπο είναι ένας ιδιαίτερα δύσκολος τομέας, αφού μπορεί να εμφανιστεί με εξαιρετικά πολύπλοκους τρόπους στην εικόνα από διάφορους προσανατολισμούς, θέσεις και μεγέθη, μέχρι και να περιλαμβάνει στοιχεία που να αλλοιώνουν τα χαρακτηριστικά του, όπως είναι οι μορφασμοί, τα γένεια, τα γυαλιά. Αντιλαμβανόμαστε τις δυσκολίες τις οποίες καλούμαστε να αντιμετωπίσουμε. Δεν είναι δυνατόν μία μέθοδος να μπορεί να λύσει όλες αυτές τις δυσκολίες, αλλά καλείται να αντιμετωπίσει πολλές από αυτές. Η δημιουργία μίας γεωμετρικής φόρμας πρέπει, λοιπόν, να γίνει με ιδιαίτερη προσοχή, ώστε να έχει επιτυχία η χρήση της.

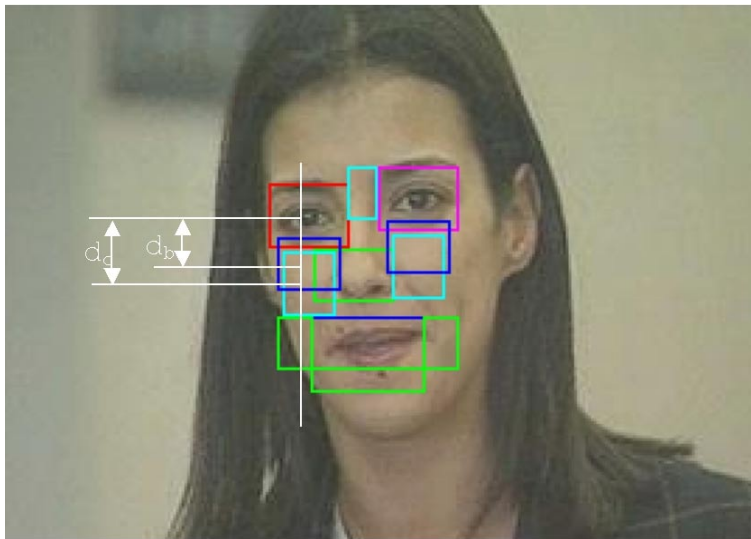
Η φόρμα που θα χρησιμοποιήσουμε δημιουργήθηκε ύστερα από μελέτη ενός αρκετά ευρέως συνόλου δεδομένων εκπαίδευσης, που περιλαμβάνουν πρόσωπα με ποικίλους βαθμούς δυσκολίας. Περιλαμβάνει τέσσερα βασικά συστατικά, τα οποία είναι τα ορθογώνια που περιέχουν τα δύο μάτια, τη μύτη και το στόμα (Σχήμα 7.11), χαρακτηριστικά τα οποία και προσπαθούμε να ανιχνεύσουμε. Η φόρμα αυτή αυξομειώνεται σύμφωνα με έναν παράγοντα κλιμάκωσης διατηρώντας τις αναλογίες της, ώστε να μπορεί να ταιριάζει σε όλα τα πρόσωπα ανεξαρτήτως του μεγέθους τους. Οι αναλογίες της φόρμας εξαρτώνται από τρεις αποστάσεις. Η πρώτη είναι η απόσταση των δύο ματιών,  $d_e$ , η δεύτερη, η απόσταση του μέσου της  $d_e$  από τη μύτη,  $d_n$ , και η τρίτη, η απόσταση του μέσου της  $d_e$  από το στόμα,  $d_m$ . Η γεωμετρική φόρμα επίσης είναι περιστρεφόμενη, ώστε να αντιμετωπίζει επιτυχώς και σταμμένα πρόσωπα, μέχρι μίας ορισμένης γωνίας φυσικά. Ορίζονται δύο γωνίες περιστροφής, η  $\alpha_e$ , που αφορά τη γωνία που μπορεί να έχουν τα μάτια μεταξύ τους και η  $\alpha_n$ , που αφορά τη στρέψη της μύτης και του στόματος (Σχήμα 7.12). Πρέπει να σημειώσουμε ότι η μύτη εξαρτάται πάντοτε από το στόμα. Επίσης ορίζεται και ένας αριθμός από βοηθητικά συστατικά της φόρμας, που είναι και πάλι ορθογώνια και αφορούν το σημείο ανάμεσα στα μάτια, την περιοχή κάτω από τα μάτια, τα μάγουλα, το δεξιό και αριστερό όριο του στόματος και το κάτω μέρος του στόματος (Σχήμα 7.13).



Σχήμα 7.11: Τα τέσσερα βασικά συστατικά της γεωμετρικής φόρμας.



Σχήμα 7.12: Αποστάσεις και γωνίες που ορίζουν τη γεωμετρική φόρμα.



Σχήμα 7.13: Ολοκληρωμένη φόρμα.

Όπως ήδη αναφέραμε η φόρμα πρέπει να συμπεριφέρεται έτσι ώστε να ταιριάζει σε πρόσωπα διαφόρων μεγεθών, καθώς επίσης και σε στραμμένα πρόσωπα. Η αυξομείωση και περιστροφή της φόρμας γίνεται με βάση τις αποστάσεις και γωνίες που ορίστηκαν παραπάνω. Από τα δεδομένα εκπαίδευσης, μετά από σειρά πειραμάτων, προέκυψαν τα εξής στατιστικά για τα παραπάνω κριτήρια :

- η απόσταση  $d_e$  ποικίλει από 20 μέχρι 60 pixels
- η γωνία  $\alpha_e$  βρίσκεται μεταξύ  $-20^\circ$  και  $+20^\circ$
- η γωνία  $\alpha_n$  βρίσκεται εντός του διαστήματος  $[-2\alpha_e, 2\alpha_e]$
- η απόσταση  $d_n$  βρίσκεται εντός του διαστήματος  $[0.6d_e, 0.8d_e]$
- η απόσταση  $d_m$  είναι πάντοτε σταθερή και ισούται με  $2d_n$

Τα ορθογώνια που περιβάλλουν τα χαρακτηριστικά προσώπου καθώς και τα βοηθητικά ορθογώνια που ορίστηκαν παραπάνω (βλ. και Σχήμα 7.13) μεταβάλλονται ανάλογα με την απόσταση των ματιών  $d_e$  σύμφωνα με :

$$WinSize_i = 0.03 \cdot win_i \cdot d_e \quad (7.1)$$

όπου  $win$  είναι σταθερές τιμές που έχουν αρχικά τα ορθογώνια αυτά και αναφέρονται στον πίνακα 7.1. Αυτό είναι κατανοητό αφού όσο μεγαλώνει το πρόσωπο μεγαλώνουν και τα χαρακτηριστικά του. Στην αρχή της περιγραφής της μεθόδου αναφέρθηκε ότι χρησιμοποιείται ένα παράθυρο  $40 \times 40$  pixels ώστε να βρεθούν με συσχέτιση τα μάτια και το στόμα, παράθυρο που βρέθηκε ικανό να περιλαμβάνει όλων των μεγεθών χαρακτηριστικά. Τονίζουμε ότι αυτό είναι κάτι το εντελώς προσεγγιστικό και είναι απαραίτητο για την εύρεση σημείων αναφοράς, ώστε να μπορεί να ξεκινήσει η ελαστική αντιστοίχιση της φόρμας, και επίσης συνεισφέρει στη μείωση του υπολογιστικού κόστους του αλγορίθμου. Σε αυτό το στάδιο ενδιαφερόμαστε ιδιαίτερα για τη λεπτομέρεια, οπότε δεν θα θέλαμε να έχουμε ένα μεγάλο παράθυρο για ένα μικρό χαρακτηριστικό ή το αντίστροφο, πράγμα το οποίο θα ενέπλεκε επιπλέον άσχετη πληροφορία στους υπολογισμούς των κριτηρίων και θα δυσκόλευε την ταξινόμηση. Για την αποφυγή τέτοιων φαινομένων τα ορθογώνια προσαρμόζονται στο μέγεθος των χαρακτηριστικών.

### 7.3.2 Στατιστικά μέτρα της κυματιδιακής ανάλυσης

Από τους συντελεστές που δίνει η πολυκαναλική ανάλυση σε κυματιδιακά πλαίσια εξάγονται στατιστικά δεδομένα με τα οποία σχηματίζονται διανύσματα περιγραφής του προσώπου. Συγκεκριμένα, τα δεδομένα αυτά αφορούν διασπορές που υπολογίζονται εντός των ορθογωνίων που περιγράφηκαν

Χαρακτηριστικά	Πλάτος	Ύψος
Μάτι	25	20
Μύτη	25	16
Στόμα	35	16
Μάγουλα	16	20
Περιοχή υπό των ματιών	20	16
Όρια του στόματος (δεξιά, αριστερά)	12	16
Περιοχή υπό του στόματος	35	6

Πίνακας 7.1: Βασικές διαστάσεις των ορθογωνίων που περιλαμβάνουν τα αντίστοιχα χαρακτηριστικά.

παραπάνω. Βέβαια δεν χρησιμοποιούνται όλοι οι συντελεστές. Από πειράματα που διεξάχθηκαν στη βάση εκπαίδευσης καταλήξαμε σε μία σειρά συντελεστών που περιγράφουν καλύτερα το κάθε χαρακτηριστικό. Οι παρατηρήσεις μας έδειξαν ότι το σχήμα των χαρακτηριστικών παίζει μεγάλο ρόλο, όπως και περιμέναμε. Τα μάτια και το στόμα, λόγω του οριζοντίου σχήματός τους, περιγράφονται καλύτερα στα οριζόντια χαρακτηριστικά της ανάλυσης, σε αντίθεση με τη μύτη που περιγράφεται καλύτερα στα κάθετα. Τελικά καταλήξαμε να χρησιμοποιήσουμε 16 κατάλληλους συντελεστές για το κάθε μάτι, 6 για τη μύτη και 8 για το στόμα, από τους οποίους εξάγονται διασπορές. Στο Σχήμα 7.14 φαίνονται οι συντελεστές που χρησιμοποιήθηκαν για το μάτι.

Οι διασπορές που εξάγονται από τους συντελεστές που κρίθηκαν κατάλληλοι σχηματίζουν ένα διάνυσμα περιγραφής του προσώπου  $X = [\sigma_1^2, \dots, \sigma_{45}^2]$ , όπου  $\sigma_i^2$  η διασπορά που αντιστοιχεί σε έναν από τους συντελεστές της κυματιδιακής ανάλυσης που περιγράφηκαν παραπάνω και αν  $i = [1, \dots, 16]$  αφορούν το αριστερό μάτι (σημείο αναφοράς), αν  $i = [17, \dots, 32]$  αφορούν το δεξί μάτι, αν  $i = [33, \dots, 38]$  αφορούν τη μύτη και τέλος αν  $i = [39, \dots, 45]$  αφορούν το στόμα. Οι διασπορές αυτές θα πρέπει να είναι φραγμένες σύμφωνα με κάποια όρια, τα οποία υπολογίστικαν λεπτομερώς από όλο το σύνολο της εκπαίδευσης και επαληθεύτηκαν πειραματικά. Συγκεκριμένα υπολογίστηκαν διανύσματα που φράσουν άνω και κάτω ένα οποιοδήποτε διάνυσμα περιγραφής που ανήκει σε πρόσωπο. Όπως θα δούμε και παρακάτω η ελαστική αντιστοίχιση της γεωμετρικής φόρμας γίνεται σύμφωνα με τρία μεγέθη προσώπων. Για κάθε μέγεθος από το σύνολο εκπαίδευσης υπολογίστηκαν τα κατάλληλα διανύσματα--φράγματα. Επειδή ο υπολογισμός των διασπορών είναι υπολογιστικά ακριβός, ο έλεγχος αν το διάνυσμα ανήκει εντός των ορίων του φράγματος γίνεται σταδιακά για τα στοιχεία του ανάλογα με τα υπόλοιπα κριτήρια που εφαρμόζονται.

### 7.3.3 Ελαστική αντιστοίχιση και εφαρμογή ελέγχων

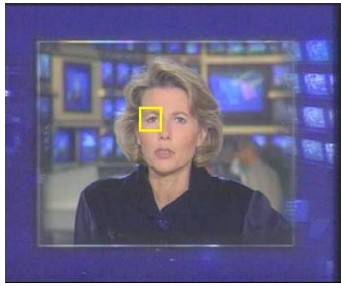
Είπαμε παραπάνω ότι η φόρμα πρέπει να ταιριάζει σε πρόσωπα διαφόρων μεγεθών. Τα πρόσωπα χωρίστηκαν σε τρεις κατηγορίες: μικρά, μεσαία και μεγάλα. Αντίστοιχα και η φόρμα προσαρμόζεται στις κατηγορίες αυτές και λαμβάνει τα μεγέθη: μικρή (0), μεσαία (1) και μεγάλη (2) σύμφωνα με:

- μικρή (0): η απόσταση των ματιών είναι 25 με 35 pixels
- μεσαία (1): η απόσταση των ματιών είναι 35 με 50 pixels
- μεγάλη (2): η απόσταση των ματιών είναι 50 με 80 pixels

Πριν να εφαρμοστεί η φόρμα στο σημείο αναφοράς, αυτό περνάει από ελέγχους, ώστε να περιοριστεί όσο το δυνατόν γίνεται η υπολογιστικά ακριβή αντιστοίχιση. Τα κριτήρια των ελέγχων αποκομίστηκαν από πειραματικά αποτελέσματα που λήφθηκαν από το σύνολο εκπαίδευσης με βάση την ίδια τη φύση που διακατέχει ένα πρόσωπο. Όλα τα φράγματα καθώς και τα κριτήρια εκπαιδεύτηκαν από το εκπαιδευτικό σύνολο ξεχωριστά για κάθε κατηγορία.

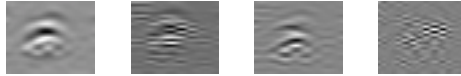
Πρώτα ελέγχονται τοπικά κριτήρια που αφορούν τα τμήματα υπό του αριστερού ματιού (σημείο αναφοράς) και την περιοχή στο αριστερό μάγουλο. Συγκεκριμένα οι διασπορές σ' αυτά τα τμήματα





### Κυματιδιακή Ανάλυση

Επίπεδο 2 :



Επίπεδο 3 :



Σχήμα 7.14: Κυματιδιακοί συντελεστές που αφορούν την περιγραφή του ματιού.

πρέπει να είναι μικρές, αφού πρόκειται για ομοιογενείς περιοχές. Για τα τμήματα υπό των ματιών χρησιμοποιούνται οι συντελεστές AH, ενώ για τα μάγουλα υπολογίζονται οι διασπορές στους συντελεστές AH και AHH.

Στη συνέχεια γίνεται μία σύγκριση της ενέργειας που εμπεριέχεται στην περιοχή του αριστερού ματιού με αυτήν της περιοχής στο αριστερό μάγουλο στην εικόνα των φωτεινότητων. Στη γενική περίπτωση τα μάγουλα είναι πιο φωτεινά από τα μάτια και άρα :

$$W_{(\text{αριστερό μάγουλο})} > W_{(\text{αριστερό μάτι})} \quad (7.2)$$

όπου  $W$  είναι η ενέργεια που υπολογίζεται σε μία περιοχή  $A$  σύμφωνα με :

$$W(A) = \frac{1}{N} \sum_{x,y} A_{x,y}^2 \quad (7.3)$$

Επίσης ελέγχεται και η ενέργεια όσον αφορά το αριστερό μάτι στους συντελεστές AH και AV. Όπως ήδη αναφέρθηκε, το μάτι περιγράφεται καλύτερα από τις οριζόντιες λεπτομέρειες και άρα η ενέργεια στα οριζόντια χαρακτηριστικά θα πρέπει να υπερέχει, δηλαδή :

$$W_{AH} > W_{AV} \quad (7.4)$$

Αφού ικανοποιηθούν τα παραπάνω κριτήρια, προχωράμε στην εξαγωγή των διασπορών του αριστερού ματιού, και ελέγχουμε αν είναι κατάλληλα φραγμένες.

Τέλος, αν ικανοποιηθούν και τα φράγματα των διασπορών ελέγχεται η περιοχή ακριβώς πάνω από τη μύτη, δηλαδή το τμήμα του προσώπου μεταξύ των ματιών, σε σχέση με το αριστερό μάτι. Εφόσον είναι πιο ομοιογενείς περιοχή θα έχει υψηλότερη ενέργεια από αυτήν του ματιού :

$$W_{(\text{περιοχή πάνω από μύτη})} > W_{(\text{αριστερό μάτι})} \quad (7.5)$$

Μόνο αν κάθε σημείο αναφοράς ικανοποιεί τα παραπάνω κριτήρια αρχίζει η ελαστική αντιστοίχιση της φόρμας σε κάθε της μέγεθος με σφάλματα όσον αφορά τις αποστάσεις και τις γωνίες. Εντός



αυτών των σφαλμάτων εφαρμόζονται για κάθε πιθανό δεξί μάτι όλοι οι παραπάνω έλεγχοι από τους οποίους πέρασε και το σημείο αναφοράς (το αριστερό μάτι δηλαδή). Ελέγχεται δηλαδή και πάλι αν γύρω του υπάρχει αρκετό χρώμα του ανθρώπινου δέρματος, αν συσχετίζεται κατάλληλα με το μέσο σήμα του ματιού, αν οι διασπορές του τμήματος από κάτω από το δεξί μάτι καθώς και του τμήματος στο δεξί μάγουλο είναι μικρές. Επίσης συγκρίνεται η ενέργεια του δεξιού ματιού με αυτήν στο δεξί μάγουλο, η ενέργεια που εμπεριέχεται στις οριζόντιες λεπτομέρειες με αυτήν στις κάθετες και εξάγονται οι διασπορές του δεξιού ματιού που πρέπει να φράσσονται κατάλληλα (βλ. και παραπάνω για το αριστερό μάτι).

Στη συνέχεια και εφόσον ικανοποιούνται όλα τα παραπάνω κριτήρια και κατά τη διάρκεια της αντιστοίχισης ελέγχεται η συσχέτιση που παρουσιάζουν τα μάτια. Η συσχέτιση δύο ίσων περιοχών  $T, M$  δίδεται από τον τύπο :

$$cor(T, M) = \frac{cov(T, M)}{\sqrt{cov(T, T) \cdot cov(M, M)}} \quad (7.6)$$

και είναι κανονικοποιημένη δίνοντας  $-1$  για αντίθετη συσχέτιση,  $0$  για πλήρως ασυσχέτιση και  $1$  για πλήρως συσχέτιση των δύο περιοχών. Σημειώνουμε ότι η συσχέτιση γίνεται και ως προς τη γωνία και ως προς τη συμμετρία.

Εφόσον ικανοποιούνται πλέον όλα τα κριτήρια που τέθηκαν για τα μάτια, αναζητούμε τη μύτη σύμφωνα με την κατάλληλη απόσταση και γωνία. Βέβαια πριν, προηγείται ένας έλεγχος για το αν υπάρχουν αρκετά σημεία του χρώματος του ανθρώπινου δέρματος εντός των ορίων της φόρμας, διαφορετικά δεν θα έχει νόημα να αναζητούμε χαρακτηριστικά προσώπου, αν δεν υπάρχει πιθανότητα να υπάρχει πρόσωπο. Αφού, λοιπόν, ικανοποιείται και αυτό, εξάγονται οι διασπορές που αφορούν τη μύτη και ελέγχονται αν βρίσκονται εντός των φραγμάτων. Αν βρεθεί “κατάλληλη μύτη” προχωράμε στην αναζήτηση στόματος.

Όπως αναφέρθηκε σε προηγούμενη ενότητα το στόμα μοντελοποιήθηκε όπως και τα μάτια. Έτσι, αντί να λάβουμε το στόμα απευθείας από το συγκεκριμένο σημείο, σύμφωνα με την κατάλληλη απόσταση και γωνία, συσχετίζουμε το μέσο σήμα του στόματος με το σημείο αυτό εφαρμόζοντας έναν επιπλέον έλεγχο. Το στόμα, όπως και τα μάτια, περιγράφεται καλύτερα από τις οριζόντιες λεπτομέρειες. Για το λόγο αυτό συγκρίνεται η ενέργεια στις οριζόντιες λεπτομέρειες με αυτήν στις κάθετες, που πρέπει να είναι μεγαλύτερη, με την ίδια ακριβώς διαδικασία που έγινε και για τα μάτια. Όσον αφορά το στόμα ελέγχεται και η περιφέρειά του πριν εξαχθούν και οι διασπορές που το αφορούν για να ελεγχθούν ως προς τα φράγματα.

### 7.3.4 Ταξινόμηση

Μετά από την όλη διαδικασία των ελέγχων, αφού επιτύχει η συγκεκριμένη διάταξη σημείων, ταξινομείται με βάση το όλο διάνυσμα περιγραφής της. Χρησιμοποιείται και η απόσταση Bhattacharyya, για λόγους που αναφέρθηκαν στην ενότητα 4.2. Υπολογίζεται η απόσταση του διανύσματος περιγραφής από το πρότυπο διάνυσμα περιγραφής της αντίστοιχης κατηγορίας. Τα πρότυπα διανύσματα περιγραφής δημιουργήθηκαν από το όλο σύνολο της εκπαίδευσης από τη μέση τιμή των επιμέρους διανυσμάτων περιγραφής που αντιστοιχούν στην τέλεια αντιστοίχιση της φόρμας στα σωστά χαρακτηριστικά. Αν η απόσταση είναι μικρή η διάταξη θεωρείται διάταξη που ανήκει σε χαρακτηριστικά προσώπου, διαφορετικά απορρίπτεται.

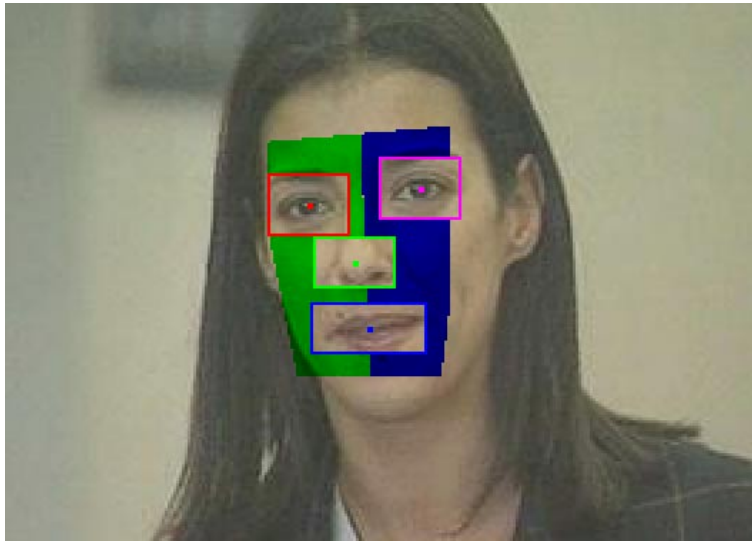
Επιπλέον της απόστασης Bhattacharyya χρησιμοποιείται και ένα κριτήριο ομοιογένειας που πρέπει να πληρεί το υποψήφιο πρόσωπο που ορίζεται από τα χαρακτηριστικά. Η φόρμα χωρίζεται σε δύο περιοχές, όπως φαίνεται στο Σχήμα 7.15, που αναπαριστούν το αριστερό και δεξί μέρος του προσώπου, πλην των χαρακτηριστικών. Η ομοιογένεια υπολογίζεται σύμφωνα με :

$$H = var\{\cup mean_{3 \times 3}\{A(x, y)\} | (x, y) \in D\} \quad (7.7)$$

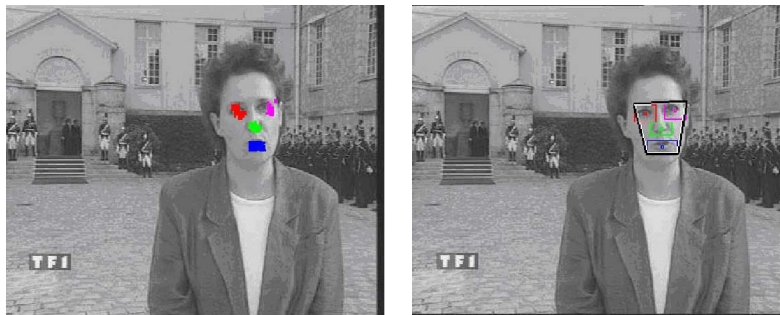
για κάθε μία από τις δύο περιοχές ξεχωριστά στην εικόνα των φωτεινοτήτων και απαιτείται τουλάχιστον μία εκ των δύο να είναι μικρή.

### 7.3.5 Τελική επιλογή

Είναι φανερό, ότι παρά την πολλαπλή εφαρμογή ελέγχων περισεύουν αρκετές υποψήφιες διατάξεις. Όμως, παρατηρούμε ότι συγκεντρώνονται κυρίως γύρω από τα πραγματικά χαρακτηριστικά. Από αυτές τις διατάξεις το ζητούμενο είναι να μείνει μία η οποία θα ταιριάζει όσο το δυνατόν καλύτερα στα πραγματικά χαρακτηριστικά. Αρχικά όλες οι διατάξεις συγκεντρώνονται σε μία λίστα, από το μικρότερο μέγεθος στο μεγαλύτερο, όπου οι πιο κοντινές δίνουν μόνο μία διάταξη, με βάση τη μέση τιμή των γεωμετρικών χαρακτηριστικών των φορμών (Σχήμα 7.16). Αν μείνουν διατάξεις με εμφανώς διαφορετικό μέγεθος, αυτές θα είναι αρκετά λιγότερες σε αριθμό. Μπορούμε να παρατηρήσουμε περιπτώσεις όπου μία φόρμα εμπεριέχεται μέσα σε άλλη (Σχήμα 7.17) ή όπου δύο ή περισσότερες φόρμες τέμνονται (Σχήμα 7.18).



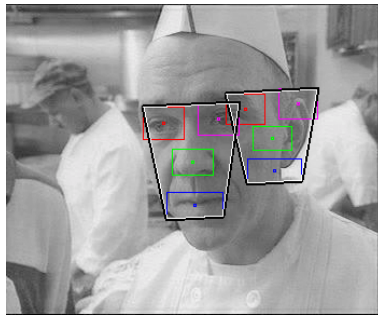
Σχήμα 7.15: Διαχωρισμός του προσώπου σε αριστερό και δεξί μέρος για τον υπολογισμό της ομοιογένειας.



Σχήμα 7.16: Διατάξεις της ίδιας τάξης μεγέθους δίνουν μία διάταξη με βάση τη μέση τιμή των γεωμετρικών χαρακτηριστικών των φορμών.



Σχήμα 7.17: Παράδειγμα διατάξεων η μία μέσα στην άλλη.



Σχήμα 7.18: Παράδειγμα διατάξεων που τέμνονται.

Στην περίπτωση αυτή μετράμε τον αριθμό των διατάξεων που απαρτίζουν τις φόρμες αυτές, και αν υπάρχει εμφανής διαφορά διαλέγεται εκείνη που απαρτίζεται από το μεγαλύτερο. Αν η διαφορά είναι μικρή, η τελική μας διάταξη θα είναι αυτή που θα παρουσιάσει την καλύτερη ομοιογένεια σε ολόκληρο το πρόσωπο, πλην των χαρακτηριστικών. Παράδειγμα της διαδικασίας αυτής απεικονίζεται στο Σχήμα 7.19.

Στα Σχήματα 7.20, 7.21 και 7.22 που ακολουθούν αναπαρίσταται και γραφικά η όλη μέθοδος της ανίχνευσης των χαρακτηριστικών προσώπου και κατά συνέπεια η ανίχνευση του προσώπου. Με κόκκινο αναπαρίστανται τα σημεία που απορρίπτονται από το ένα στάδιο στο άλλο.

Στον Πίνακα 7.2 δίδονται συγκεντρωτικά αποτελέσματα της προτεινόμενης μεθόδου τόσο στο σύνολο εκπαίδευσης όσο και σε ένα σύνολο ελέγχου. Το σύνολο των εικόνων εξάχθηκαν από τηλεοπτικά προγράμματα της ελληνικής και της γαλλικής τηλεόρασης. Στον Πίνακα 7.3 συγκρίνονται τα αποτελέσματα της μεθόδου μας με εκείνα ενός γνωστού αλγορίθμου [29] που διατίθεται για πειράματα στο Διαδίκτυο. Στα Σχήματα 7.23, 7.24, 7.25, 7.26 δίδονται χαρακτηριστικά αποτελέσματα ανίχνευσης προσώπων της προτεινόμενης μεθόδου.



Σχήμα 7.19: Παράδειγμα ομαδοποίησης διατάξεων που περιέχουν η μία την άλλη και που τέμνονται (από δεξιά προς αριστερά και από πάνω προς τα κάτω).



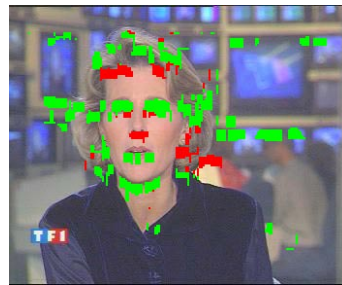
(α)



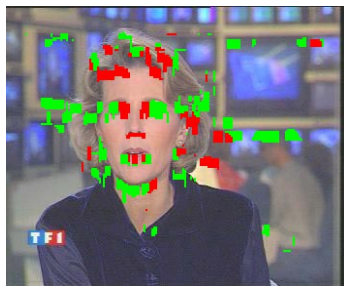
(β)



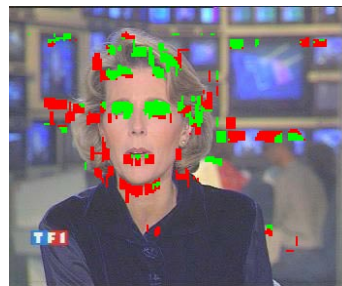
(γ)



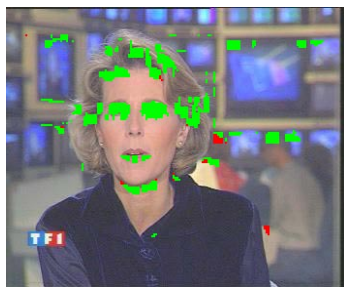
(δ)



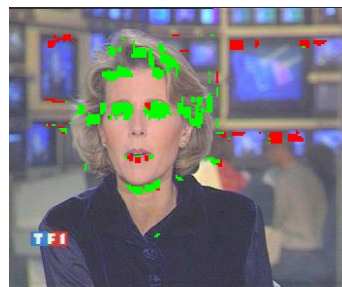
(ε)



(στ)

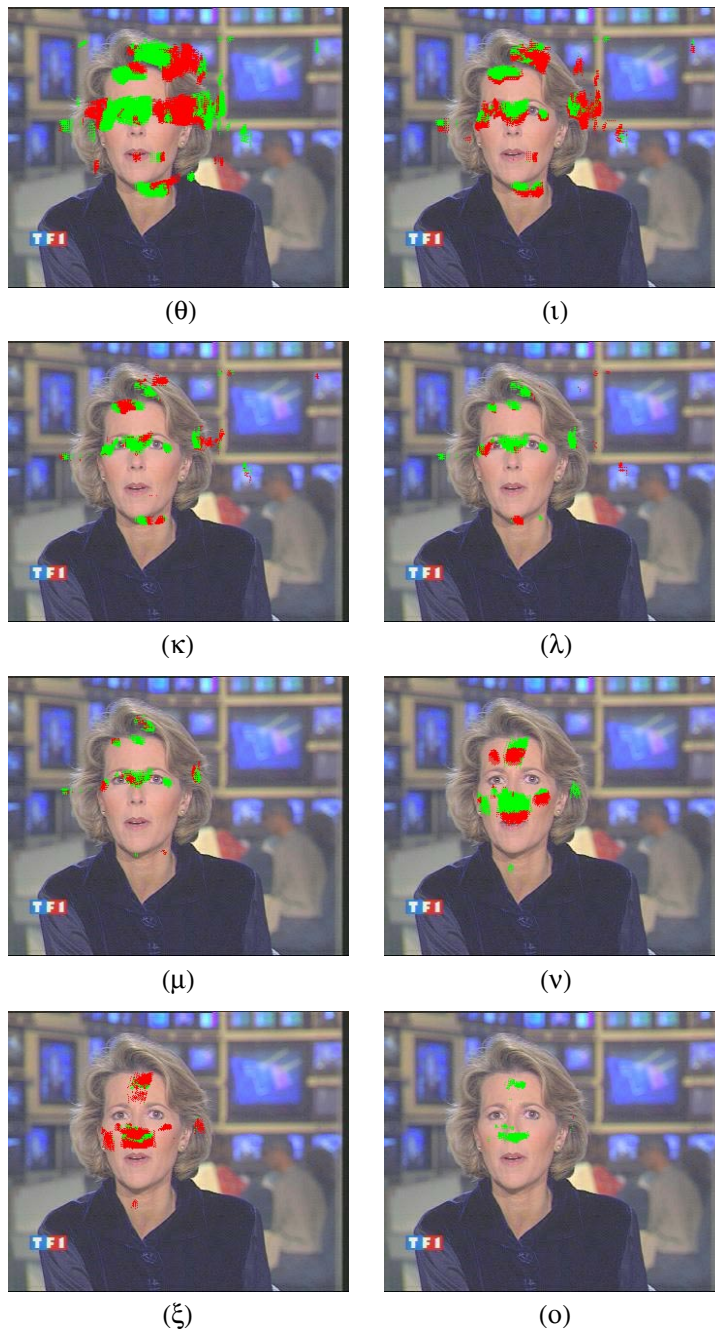


(ζ)



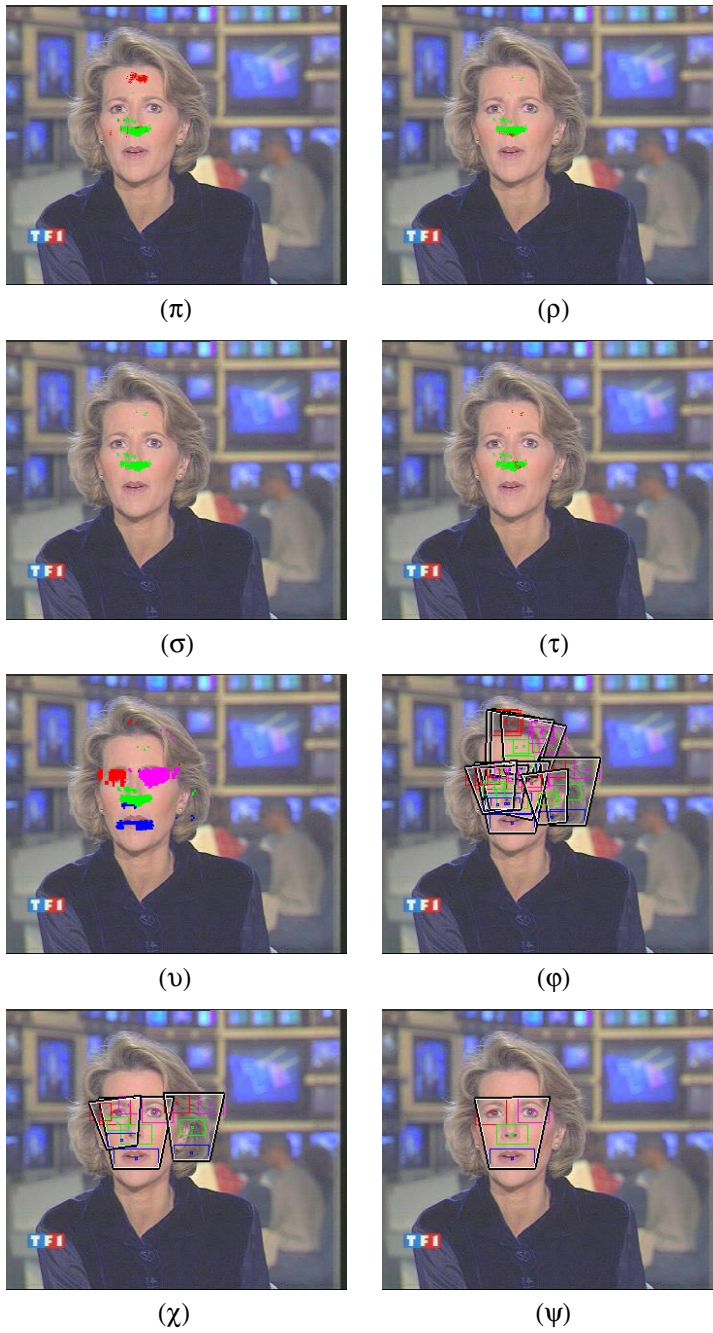
(η)

Σχήμα 7.20: Αρχικά στάδια της μεθόδου και κριτήρια σημείου αναφοράς : (α) η αρχική εικόνα, (β) κατάταξη χρώματος, (γ) ανίχνευση ματιών, (δ) διασπορά της περιοχής υπό του αριστερού ματιού, (ε) διασπορά στο αριστερό μάγουλο, (στ) ενέργεια στο αριστερό μάτι και μάγουλο, (ζ) ενέργεια οριζοντίων και καθέτων λεπτομερειών του αριστερού ματιού, (η) διασπορές αριστερού ματιού.



Σχήμα 7.21: Ενδιάμεσα στάδια αλγορίθμου για το δεξί μάτι και το στόμα : (θ) σύγκριση ενεργειών ματιού και περιοχής άνω της μύτης, (ι) εκκίνηση ελέγχων δεξιού ματιού, (κ) ενέργεια οριζοντίων και καθέτων λεπτομερειών δεξιού ματιού, (λ) διασπορές δεξιού ματιού, (μ) συσχέτιση αριστερού και δεξιού ματιού, (ν) έλεγχος χρώματος και διασπορές της μύτης, (ξ) ανίχνευση στόματος, (ο) ενέργεια οριζοντίων και καθέτων λεπτομερειών στόματος.





Σχήμα 7.22: Τελικά στάδια του αλγορίθμου : (π) περιφέρεια στόματος, (ρ) διασπορές στόματος, (σ) απόσταση Bhattacharyya, (τ) ομοιογένεια, (υ) τελικό αποτέλεσμα ελαστικής αντιστοίχισης, (φ) μέση τιμή των διατάξεων μετά την τελική αντιστοίχιση, (χ) ομαδοποίηση των διατάξεων, (ψ) επιλογή τελικής αντιστοίχισης.

	Δεδομένα εκμάθησης	Δεδομένα ελέγχου
Αριθμός εικόνων, προσώπων	46,51	88,82
Μετωπική θέση (μεγέθη 0,1,2)	10,23,8	24
Ημιμετωπική θέση (μεγέθη 0,1,2)	4,5,1	16
Σκυμμένο πρόσωπο	0	4
Περιστραμμένο πρόσωπο	0	11
Εικόνες χωρίς πρόσωπα	0	10

Πίνακας 7.2: Αποτελέσματα ανίχνευσης στα σύνολα εκπαίδευσης και ελέγχου.

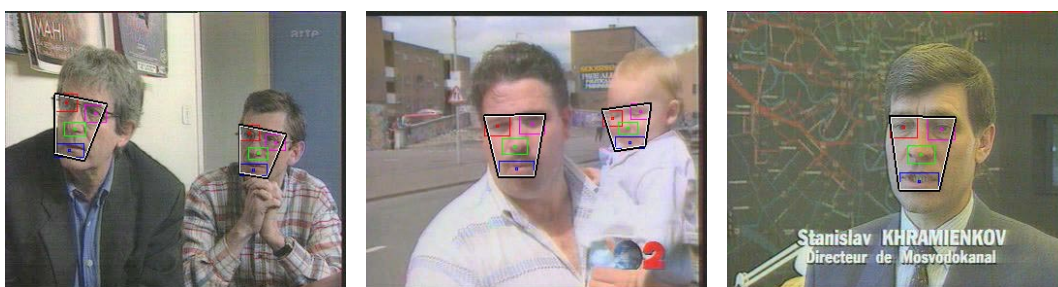
	Προτεινόμενη μέθοδος	Face detector
Μετωπική θέση	36/39	37/39
Ημιμετωπική θέση	22/24	20/24
Σκυμμένο πρόσωπο	2/4	0/4
Περιστραμμένο πρόσωπο	8/11	3/11
Λαθεμένες ανιχνεύσεις	6	9
Συνολικά πρόσωπα	68/78 = 87%	60/78 = 76%

Πίνακας 7.3: Συγκριτικά αποτελέσματα ανάμεσα στην προτεινόμενη μέθοδο και αυτή του CMU.

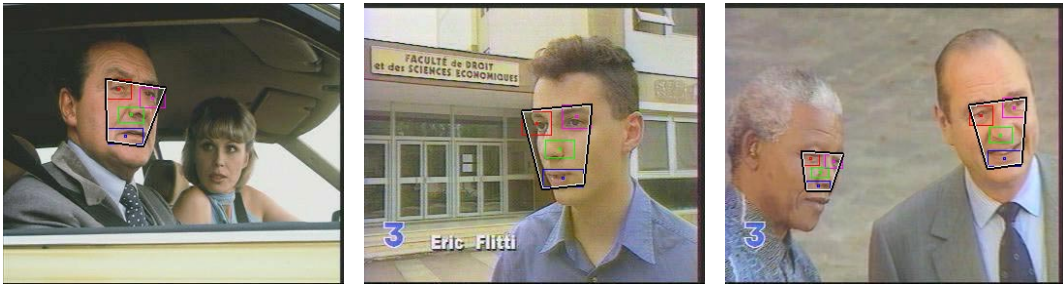




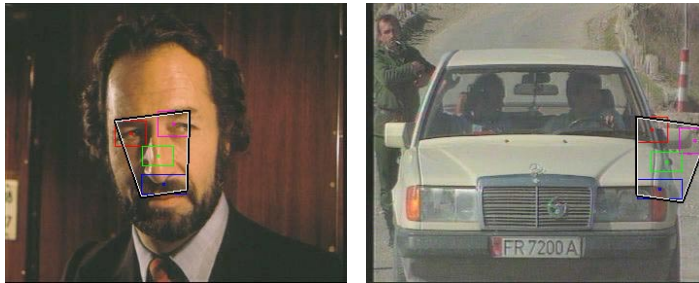
Σχήμα 7.23: Παραδείγματα ανίχνευσης προσώπων διαφόρων μεγεθών.



Σχήμα 7.24: Παραδείγματα ανίχνευσης προσώπων σε περίπου μετωπική στάση.



Σχήμα 7.25: Παραδείγματα ανίχνευσης στραμμένων προσώπων.



Σχήμα 7.26: Παραδείγματα σφαλμάτων της μεθόδου.

## Κεφάλαιο 8

# Επίλογος

Στο σύνολο της εργασίας παρουσιάστηκε μία νέα μέθοδος για την ανίχνευση ανθρώπινων προσώπων σε έγχρωμες εικόνες χωρίς σκηνικούς περιορισμούς, όπως είναι η παρουσία πολύπλοκου φόντου και άγνωστης φωτεινότητας. Τα στάδια περιλαμβάνουν :

- Κβαντισμό χρωμάτων και εξαγωγή των κυρίαρχων χρωμάτων της αρχικής εικόνας.
- Φιλτράρισμα του χρώματος του ανθρώπινου δέρματος χρησιμοποιώντας τα χρωματικά μοντέλα HSV ή YCbCr για την εξαγωγή περιοχών που έχουν το χρώμα του ανθρώπινου δέρματος. Μία δεύτερη μέθοδος κατάταξης του χρώματος του ανθρώπινου δέρματος χρησιμοποιεί είτε ένα νευρωνικό δίκτυο με διανύσματα υποστήριξης είτε ένα πολυστρωματικό νευρωνικό δίκτυο.
- Επαναληπτική συγχώνευση στο σύνολο των ομοιογενών, ως προς το χρώμα του ανθρώπινου δέρματος, περιοχών για την εξαγωγή υποψήφιας περιοχών προσώπου.
- Εφαρμογή περιορισμών σύμφωνα με ανθρωπομορφικά κριτήρια.
- Εφαρμογή πολυκαναλικής ανάλυσης σε κυματιδικά πλαίσια σε κάθε υποψήφια περιοχή προσώπου για την εξαγωγή απλών στατιστικών δεδομένων που περιγράφουν την υφή του προσώπου.
- Ταξινόμηση των υποψήφιας περιοχών εξάγοντας χαρακτηριστικά διανύσματα και χρησιμοποιώντας την απόσταση Bhattacharyya συγκρίνοντάς τα με κάποια πρότυπα χαρακτηριστικά διανύσματα που δημιουργήθηκαν από ένα σύνολο προσώπων ύστερα από ένα εκπαιδευτικό στάδιο.

Για ένα σύνολο 100 εικόνων που περιέχαν 104 πρόσωπα (σε διάφορες θέσεις, προσανατολισμούς, κλίσεις και με διάφορες χρωματικές πληροφορίες και συνθήκες φωτεινότητας) επετεύχθησαν 94.23% σωστής ανίχνευσης προσώπων, 20 λάθος ανιχνεύσεις και 5.76% μη ανιχνευμένων προσώπων. Η προτεινόμενη μέθοδος είναι εξαιρετικά αποδοτική κυρίως για πρόσωπα μεγάλου μεγέθους και η κατάταξη του χρώματος του ανθρώπινου δέρματος βελτιώνει ιδιαίτερα την ταχύτητα της ανίχνευσης προμηθεύοντας υποψήφιας περιοχές, και ταυτόχρονα αποκόπτει το κυρίως πρόσωπο από το φόντο του, ακόμα κι αν αυτό περιέχει χρωματική πληροφορία όμοια με εκείνη του προσώπου. Η πολυκαναλική ανάλυση σε πλαίσια κυματιδίων δίνει ένα ισχυρό σύστημα ανίχνευσης προσώπου, ακόμα κι αν δεν τίθεται απολύτως κανένας περιορισμός σχετικά με το πρόσωπο (εκτός ενός ελάχιστου απαιτούμενου μεγέθους). Επιπλέον πολύ γρήγορες υλοποιήσεις της ανάλυσης αυτής μπορούν να βρεθούν σε μορφή hardware συμβαδίζοντας με τις ανάγκες ανίχνευσης σε πραγματικό χρόνο.

Επίσης παρουσιάστηκε μία μέθοδο σύνδεσης του συστήματος ανίχνευσης με ένα αυτόνομο σύστημα αναγνώρισης που προϋποθέτει τη θέση του προσώπου στην εικόνα και τη μετωπική ή ημιμετωπική του πόζα. Η επεξεργασία αυτή εφαρμόζεται μετά την ανίχνευση προσώπων. Χρησιμοποιήθηκε η πολυκαναλική ανάλυση σε κυματιδικά πλαίσια για την εξαγωγή δεδομένων που με απλά στατιστικά κριτήρια οδηγούν σε συνδυασμό με μία ελαστική αντιστοίχιση γεωμετρικής φόρμας στην κατά

προσέγγιση ανίχνευση των χαρακτηριστικών του προσώπου και στη συνέχεια στη διαφοροποίηση των μετωπικών-ημιμετωπικών προσώπων από τα στραμένα πρόσωπα και τις λάθος ανιχνεύσεις.

Τέλος, αναπτύχθηκε μία αξιόπιστη μέθοδος ανίχνευσης χαρακτηριστικών προσώπων σε ψηφιακές έγχρωμες εικόνες, χωρίς σκηνικούς περιορισμούς. Χρησιμοποιείται η πολυκαναλική ανάλυση σε κυματιδιακά πλαίσια και η κατάταξη του χρώματος του ανθρώπινου δέρματος. Μοντελοποιήθηκαν τα μάτια και το στόμα. Δημιουργήθηκε κατάλληλη γεωμετρική φόρμα για το σκοπό της ανίχνευσης των χαρακτηριστικών. Ερευνήθηκαν κατάλληλα στατιστικά κριτήρια και μέθοδοι που θα οδηγήσουν μία ελαστική αναζήτηση φόρμας στο επιθυμητό αποτέλεσμα. Η μέθοδος χρησιμοποιήθηκε και για την ανίχνευση προσώπων με πολύ καλή επιτυχία.

# Βιβλιογραφία

- [1] P. Belhumeur, J. Hespanha, D. Kriegman. Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19(7), pp. 711--720, 1997.
- [2] S. Belongie, C. Carson, H. Greenspan, J. Malik. Color- and texture-based image segmentation using EM and its application to content-based image retrieval. *Proc. 6th IEEE Int. Conf. Computer Vision*, Jan. 1998.
- [3] R. Brunelli, T. Poggio. Face Recognition: Features versus Templates. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 15(10), pp. 1042--1052, 1993.
- [4] M.C. Burl, T.K. Leung, P. Perona. Face localization via shape statistics. Presented at *Int. Workshop on Automatic Face and Gesture Recognition*, June 1995.
- [5] R. Chellappa, C.L. Wilson, S. Sirohey. Human and Machine Recognition of faces: A survey. in: *Proceedings of IEEE*, 83(5), 705--740, 1995.
- [6] R.R. Coifman, M.V. Wickerhauser. Entropy-based algorithms for best basis selection. *IEEE Trans. on Information Theory*, 38, pp. 713--718, 1992.
- [7] J. Cox, Y.J. Ghosen, P. Yianilos. Feature-based face recognition using mixture distance. *Proc. Computer Vision and Pattern Recognition*, pp. 209--216, 1996.
- [8] Y. Dai, Y. Nakano. Recognition of facial images with low resolution using a Hopfield memory model. *Pattern Recognition*, 31(2), pp. 159--167, 1998.
- [9] I. Daubechies. The Wavelet Transform, Time-Frequency Localization and Signal Analysis. *IEEE Trans. on Information Theory*, 36(5), pp. 961--1005, 1990.
- [10] A. Eleftheradis and A. Jacquin. Model-assisted coding of video conferencing sequences at low bit rates. In *Proceedings of IEEE Int. Symp. Circuits and Systems*, pp. 3.177--3.180, 1994.
- [11] Y. Fu. Handbook of Pattern Recognition and Image processing. *New York: Academic*, 1986.
- [12] C. Garcia, G. Zikos, G. Tziritas. Wavelet Packet Analysis for Face Recognition. To appear in *Image and Vision Computing*, 1999.
- [13] C. Garcia, G. Zikos, G. Tziritas. Face Detection in Color Images using Wavelet Packet Analysis. *Proc. of the IEEE Intern. Conf. on Multimedia Computing and Systems*, vol. I, pp. 703--708, Florence, June 1999.
- [14] C. Garcia, G. Tziritas. Face Detection Using Quantized Skin Color Regions Merging and Wavelet Packet Analysis. *Proc. of the IEEE Trans. on Multimedia*, vol. I, no. 3, pp. 264--277, September 1999.
- [15] S. Haykin. *Neural Networks A Comprehensive Foundation*. *New Jersey : Prentice Hall*, 1999.
- [16] Z.Q. Hong. Algebraic feature extraction of image for recognition. *Pattern Recognition*, 24(3), pp. 211--219, 1991.

- [17] S.-H. Jeng, H.Y.M. Yao, C.C. Han, M.Y. Chern and Y.T. Liu. Facial Feature Detection Using Geometrical Face Model: An Efficient Approach. *Pattern Recognition*, 31(3), pp. 273--282, 1998.
- [18] T. Kailath. The divergence and Bhattacharyya distance measures in signal selection. *IEEE Trans. Commun.*, vol. COM--15, pp. 52--60, 1967.
- [19] D. Kazakos. The Bhattacharyya distance and detection between Markov chains. *IEEE Trans. Inform. Theory*, vol. IT--24, pp. 747--754, June 1978.
- [20] M. Kirby, L. Sirovich. Application of the Karhunen-Loeve Procedure and the Characterization of Human Faces. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 12(1), pp. 103--108, 1990.
- [21] A. Lanitis, C.J. Taylor, T. F. Cootes. Automatic Interpretation and Coding of Face Images Using Flexible Models. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19(7), pp. 743--756, 1997.
- [22] S.-H. Lin, S.-Y. Kung, L.-J. Lin. Face Recognition/Detection by Probabilistic Decision-Based Neural Network. *IEEE Trans. on Neural Networks*, 8(1), pp. 114--131, 1997.
- [23] Y. Linde, A. Buzo, R.M. Gray. An algorithm for vector quantizer design. *IEEE Trans. Commun.*, vol. COM--28, pp. 84--95, 1980
- [24] S. Mallat. A theory of multi-resolution signal decomposition: the wavelet representation. *IEEE Trans. on Pattern Recognition and Machine Intelligence*, 11, pp. 674--693, 1989.
- [25] O. Nakamura, S. Mathur, T. Minami. Identification of human faces based on isodensity maps. *Pattern Recognition*, 24(3), pp. 263--272, 1991.
- [26] A. Pentland, R.W. Picard, S. Sclaroff. Photobook: Content-Based Manipulation of Image Databases. in: *Proc. of the SPIE Storage and Retrieval and Video Databases II*, 1994.
- [27] J.L. Perry, J.M. Carney. Human Face Recognition Using a Multilayer Perceptron. in: *IJCNN*, pp. 413--416, 1990.
- [28] P.J. Phillips, R. McCabe, R. Chellappa. Biometric image processing and recognition. in: *Proc. of the IX European Signal Processing Conference*, I, 1998.
- [29] H. A. Rowley, S. Baluja, and T. Kanade. Neural network-based face detection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 20, pp. 23--28, 1998.
- [30] M. Smith, T. Barnwell. Exact reconstruction techniques for tree structured subband coders. *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP--34, pp. 434--441, Mar. 1986.
- [31] K. Sobottka and I. Pitas. A novel method for automatic face segmentation, facial feature extraction and tracking. *Signal Processing: Image Communication*, 12, pp. 263--281, 1998.
- [32] K.K. Sung and T. Poggio. Example-based learning for view-based human face detection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 20, pp. 39--51, 1998.
- [33] D.L. Swets, J. Weng. Using Discriminant Eigenfeatures for Image Retrieval. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 18(8), pp. 831--836, 1996.
- [34] S. Tsekeridou, I. Pitas. Facial features extraction in frontal views using biometric analogies. *Proc. IX European Signal Processing Conference*, vol. 1, 1998.
- [35] M. Turk, A. Pentland. Eigenfaces for Recognition. *Journal of Cognitive Science*, 3(1), pp. 71--86, 1991.
- [36] H. Wang, S.-F. Chang. A highly efficient system for automatic face region detection in MPEG video. *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, no. 4, pp. 615--628, 1997.

- [37] P.H. Westerink, J. Biemond, D.E. Boeke. Subband coding of color images”, Subband Image Coding. *J.W. Woods, Ed. Boston, MA : Kluwer*, pp. 193--228, 1991.
- [38] C.L. Wilson, C.S. Barnes, R. Chellappa, S.A. Sirohey. Face Recognition Technology for Law Enforcement Applications. *NISTIR 5465, U.S. Department of Commerce*, 1994.
- [39] L. Wiskott, J.M. Fellous, N. Kruger, C. Von der Malsburg. Face Recognition by Elastic Bunch Graph Matching. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19(7), pp. 775--779, 1997.
- [40] G. Yang and T.S. Huang. Human face detection in a complex background. *Pattern Recognition*, 27(1), pp. 55--63, 1994.
- [41] K.C. Yow, C. Cipolla. Feature-based human face detection. *Image and Vision Computing*, 15, pp. 713--735, 1997.