



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ

ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ

ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

Αποτελεσματικές Τεχνικές Εξαγωγής  
Χαρακτηριστικών και Ανάλυσης Εκφράσεων  
Προσώπου στην Επικοινωνία Ανθρώπου-Μηχανής

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

ΤΟΥ

**ΣΠΥΡΟΥ Β. ΙΩΑΝΝΟΥ**

Διπλωματούχου Ηλεκτρολόγου Μηχανικού  
& Μηχανικού Υπολογιστών Ε.Μ.Π. (2000)

Αθήνα, Μάιος 2006





ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ & ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ & ΥΠΟΛΟΓΙΣΤΩΝ

Αποτελεσματικές Τεχνικές Εξαγωγής  
Χαρακτηριστικών και Ανάλυσης Εκφράσεων  
Προσώπου στην Επικοινωνία Ανθρώπου-Μηχανής

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

του

ΣΠΥΡΟΥ Β. ΙΩΑΝΝΟΥ

Διπλωματούχου Ηλεκτρολόγου Μηχανικού  
& Μηχανικού Υπολογιστών Ε.Μ.Π. (2000)

Συμβουλευτική Επιτροπή: Στέφανος Κόλλιας  
Ανδρέας Σταφυλοπάτης  
Παναγιώτης Τσανάκας

Επταμελής εξεταστική επιτροπή

...	...	...
Σ. Κόλλιας	Α. Σταφυλοπάτης	Π. Τσανάκας
Καθηγητής Ε.Μ.Π.	Καθηγητής Ε.Μ.Π.	Καθηγητής Ε.Μ.Π.

...	...	...
Γ. Καραγιάννης	Ν. Μήτρου	Π. Μαραγκός
Καθηγητής Ε.Μ.Π.	Καθηγητής Ε.Μ.Π.	Καθηγητής Ε.Μ.Π.

...

Αν. Ντελόπουλος  
Επ. Καθηγητής  
Αριστοτέλειου Πανεπιστημίου

Αθήνα, Μάιος 2006

...

**ΣΠΥΡΟΣ Β. ΙΩΑΝΝΟΥ**

Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

© 2006

Με επιφύλαξη παντός δικαιώματος - All rights reserved

# Περιεχόμενα

<b>1</b>	<b>Εξαγωγή των χαρακτηριστικών του προσώπου</b>	<b>15</b>
1.1	Εισαγωγή .....	15
1.2	Επισκόπηση της προτεινόμενης μεθοδολογίας .....	20
1.3	Ανίχνευση του προσώπου και της περιστροφής του .....	22
1.4	Ανίχνευση των χαρακτηριστικών του προσώπου, και εξαγωγή των ορίων τους .....	26
1.4.1	Ανίχνευση ορίων οφθαλμών .....	29
1.4.1.1	Συνδυασμός μάσκας φωτεινότητας και χρώματος	29
1.4.1.2	Μάσκα βασιζόμενη σε ακμές .....	29
1.4.1.3	Μάσκα βασιζόμενη στην τυπική απόκλιση .....	31
1.4.1.4	Μάσκα βασιζόμενη στη φωτεινότητα .....	33
1.4.2	Ανίχνευση ορίων φρυδιών .....	33
1.4.3	Ρινικός εντοπισμός .....	36
1.4.4	Ανίχνευση στόματος .....	37
1.4.4.1	Κατασκευή μάσκας χειλιών και στόματος με νευρωνικό δίκτυο .....	37
1.4.4.2	Κατασκευή μάσκας με ανίχνευση ακμών .....	39
1.4.4.3	Μάσκα συγχώνευσης φωτεινότητας χειλιών και ακμών στόματος .....	40
1.5	Συγχώνευση των ενδιάμεσων масκών .....	43
1.5.1	Αξιολόγηση οφθαλμικών και στοματικών масκών .....	43
1.5.2	Συγχώνευση Μασκών .....	45
1.5.3	Εκτίμηση Εμπιστοσύνης για τις μάσκες Οφθαλμών, Στόματος και Φρυδιών .....	50
1.6	Πειραματικά Αποτελέσματα .....	55
1.6.1	Σχολιασμός των αποτελεσμάτων .....	57
1.7	Συμπεράσματα - Επεκτάσεις .....	59

<b>2</b>	<b>Εκτίμηση Συναισθημάτων με Βάση τις Εκφράσεις του Προσώπου</b>	<b>61</b>
2.1	Η μέθοδος πρόκλησης συναισθημάτων SAL.....	61
2.2	Αυτόματη Εκτίμηση Εκφράσεων .....	63
2.2.1	Από τις μάσκες στα χαρακτηριστικά σημεία .....	63
2.2.2	Από τα χαρακτηριστικά σημεία στις παραμέτρους κίνησης χαρακτηριστικών σημείων (FAP) .....	67
2.2.3	Εκφράσεις Προσώπου και Επικοινωνία Ανθρώπου - Μηχανής .....	69
2.2.3.1	Τροχός Συναισθημάτων της Whissel.....	70
2.3	Δυνατοτική Αποτίμηση Ασαφών Κανόνων.....	71
2.3.1	Εισαγωγή .....	71
2.3.2	Αποτίμηση των όρων.....	74
2.3.3	Πειραματικά Αποτελέσματα .....	77
2.4	Αποτίμηση Κανόνων με Ασαφές Νευρωνικό Δίκτυο.....	78
2.5	Συμπεράσματα .....	80
<b>3</b>	<b>Επεκτάσεις με Βάση την Ανάλυση Χειρονομιών</b>	<b>87</b>
3.1	Ανάλυση χειρονομιών .....	87
3.2	Ασαφής συγχώνευση πολλαπλών ειδών εισόδου για την κατάτμηση εικόνων και βίντεο .....	90
3.3	Η προτεινόμενη αρχιτεκτονική .....	93
3.3.1	Ο Αλγόριθμος RSST.....	94
3.3.2	Κατάτμηση με Βάση το Χρώμα .....	96
3.3.3	Κατάτμηση με Βάση την Κίνηση .....	96
3.3.4	Κατάτμηση με Βάση το Πεδίο Βάθους .....	97
3.4	Ασαφής Συγχώνευση Κατατμήσεων .....	98
3.5	Αποτελέσματα Κατάτμησης.....	101
3.5.1	Εξοπλισμός.....	101
3.5.2	Χρώμα .....	101
3.5.3	Κίνηση .....	104
3.5.4	Βάθος .....	104
3.5.5	Συγχώνευση.....	104
3.5.6	Αποτελέσματα .....	105
3.6	Επεκτάσεις .....	106
3.7	Συμπεράσματα .....	108
<b>4</b>	<b>Συμπεράσματα</b>	<b>115</b>
<b>A</b>	<b>Αποδόσεις όρων</b>	<b>119</b>

Βιβλιογραφία	121
Κατάλογος δημοσιεύσεων του συγγραφέα	131
Βιογραφικό Σημείωμα	135





# Σχήματα

1.1	Επισκόπηση του συστήματος .....	21
1.2	Περιοχές ανίχνευσης χαρακτηριστικών .....	23
1.3	Έξοδος του νευρωνικού δικτύου σε αριστερό οφθαλμό .....	25
1.4	Απεικόνιση Εισόδου Νευρωνικού Δικτύου .....	25
1.5	1 <sup>η</sup> οφθαλμική μάσκα .....	30
1.6	2 <sup>η</sup> οφθαλμική μάσκα .....	30
1.7	3 <sup>η</sup> οφθαλμική μάσκα .....	31
1.8	4 <sup>η</sup> οφθαλμική μάσκα .....	33
1.9	Μάσκα φρυδιών .....	34
1.10	Ανίχνευση μύτης .....	37
1.11	Μάσκα στόματος .....	38
1.12	1 <sup>η</sup> στοματική μάσκα .....	39
1.13	2 <sup>η</sup> στοματική μάσκα .....	40
1.14	περιοχή ανίχνευσης χαρ/κών στόματος .....	40
1.15	3 <sup>η</sup> στοματική μάσκα .....	42
1.16	Η διαδικασία συγχώνευσης масκών .....	43
1.17	Μια προβληματική μάσκα .....	43
1.18	Το μοντέλο της Δυναμικής Μηχανής Επιτροπής .....	49
1.19	Σφάλμα χειρωνακτικής κατάτμησης .....	56
1.20	Κατανομή Williams Index .....	57
2.1	Διαπροσωπίες .....	63
2.2	Εύρεση των 19 χαρ. σημείων σε συνεχόμενα καρέ και εκτίμηση της έκφρασης .....	65
2.3	Τα 19 εντοπισμένα χαρακτηριστικά σημεία σε ποικιλία εκφράσεων .....	66
2.4	Τα 19 εντοπισμένα χαρακτηριστικά σημεία σε ποικιλία εκφράσεων .....	67
2.5	Αποστάσεις υπολογισμού χαρακτηριστικών σημείων .....	69
2.6	Μονάδες ISO MPEG-4 .....	69
2.7	Ο χώρος ενεργοποίησης-συναισθήματος .....	71

2.8	Παραδείγματα καρέ τα οποία δεν αναλύθηκαν χωρίς τη χρήση προσαρμογής .....	78
2.9	Η διαπροσωπεία της υλοποίησης (Matlab) .....	80
2.10	Εύρεση των 19 χαρ. σημείων σε συνεχόμενα καρέ και εκτίμηση της έκφρασης .....	82
2.11	Εύρεση των 19 χαρ. σημείων σε συνεχόμενα καρέ και εκτίμηση της έκφρασης .....	83
2.12	Εύρεση των 19 χαρ. σημείων σε συνεχόμενα καρέ και εκτίμηση της έκφρασης .....	84
3.1	Γενική Αρχιτεκτονική του Συστήματος .....	94
3.2	Συγχωνευτικό Ασαφές Σύστημα Συμπερασμού .....	101
3.3	Ορισμός των ασαφών κατατμήσεων .....	102
3.4	Σχηματική αναπαράσταση των συσχετίσεων εισόδου-εξόδου .....	102
3.5	Ακολουθία 2, Καρέ 60, Μεγέθυνση.....	107
3.6	Ακολουθία 1, Καρέ 30 .....	109
3.7	Ακολουθία 1, Καρέ 31 .....	109
3.8	Ακολουθία 1, Καρέ 32 .....	109
3.9	Ακολουθία 1, Καρέ 33 .....	110
3.10	Ακολουθία 1, Καρέ 34 .....	110
3.11	Ακολουθία 1, Καρέ 35 .....	110
3.12	Ακολουθία 2, Καρέ 55 .....	111
3.13	Ακολουθία 2, Καρέ 56 .....	111
3.14	Ακολουθία 2, Καρέ 57 .....	111
3.15	Ακολουθία 2, Καρέ 58 .....	112
3.16	Ακολουθία 2, Καρέ 59 .....	112
3.17	Ακολουθία 2, Καρέ 60 .....	112
3.18	Χαρακτηριστικά σημεία προσώπου, χρονική εξέλιξη των FAP, εκτιμώμενη και χειρωνακτική κατηγοριοποίηση και προσωδιακά χαρακτηριστικά .....	113

# Πίνακες

1.1	Ορισμός Περιοχών Ανίχνευσης Χαρακτηριστικών .....	23
1.2	Σύνοψη των οφθαλμικών και στοματικών масκών .....	27
1.3	Προσαρμοζόμενα Κατώφλια .....	28
1.4	Κατώφλια .....	28
1.5	Συσχετιστικοί ανθρωπομετρικοί περιορισμοί .....	28
1.6	Ανθρωπομετρικές οφθαλμικές μετρήσεις αξιολόγησης масκών ....	46
1.7	Ανθρωπομετρικές μετρήσεις αξιολόγησης масκών στόματος .....	46
1.8	Παραδείγματα συγχώνευσης масκών στον αριστερό οφθαλμό, αντίστοιχες τιμές αξιολόγησης και FPs .....	51
1.9	Παραδείγματα συγχώνευσης масκών στον αριστερό οφθαλμό, αντίστοιχες τιμές αξιολόγησης και FPs .....	52
1.10	Παραδείγματα συγχώνευσης масκών στο στόμα, αντίστοιχες τιμές αξιολόγησης και FP .....	53
1.11	Ανθρωπομετρική Αξιολόγηση [32] για τη θέση και τις διαστάσεις του στόματος και των οφθαλμών. ....	55
1.12	Σύνοψη Αποτελεσμάτων .....	58
2.1	Τα χρησιμοποιούμενα χαρακτηριστικά σημεία (Feature Points) [5] .....	64
2.2	Παραδείγματα FAP και αντίστοιχων αποστάσεων .....	68
2.3	Παραδείγματα κανόνων και αντίστοιχα διαστήματα τιμών FAP σε μονάδες MPEG-4 .....	71
2.4	Παράδειγμα κανόνων με το αντίστοιχο εύρος τιμών των FAP σε μονάδες MPEG-4 .....	73
2.5	Αποτελέσματα αναγνώρισης εκφράσεων .....	77
2.6	Νέοι κανόνες που αποκτήθηκαν μέσω της προσαρμογής .....	79
2.7	Το πλήρες δείγμα κανόνων που χρησιμοποιήθηκε στο σύστημα αναγνώρισης εκφράσεων .....	81

3.1	Παραδείγματα αντιστοίχισης μεταξύ χειρονομιών και συναισθηματικών καταστάσεων .....	88
3.2	Το σύνολο των γλωσσικών κανόνων του ΑΣΚ .....	103

## ΠΕΡΙΛΗΨΗ

Η παρούσα διδακτορική διατριβή έχει κεντρικό αντικείμενο την ανάλυση ακολουθιών εικόνων και βίντεο με στόχο την αναγνώριση εκφράσεων του προσώπου. Στόχος της είναι η συνεισφορά στην ανάπτυξη ενός πλήρως αυτοματοποιημένου υπολογιστικού συστήματος το οποίο θα έχει τη δυνατότητα εξαγωγής των χαρακτηριστικών του προσώπου και χαρακτηρισμού των εκφράσεων. Σε αυτό το πλαίσιο καλείται να αντιμετωπίσει δύο βασικά προβλήματα: την αυτόματη εξαγωγή χαρακτηριστικών από το πρόσωπο και την αυτόματη αναγνώριση της έκφρασης, καλείται δηλαδή να συνδυάσει την επεξεργασία εικόνας χαμηλού επιπέδου με τα αποτελέσματα ψυχολογικών ερευνών για την αντίληψη της έκφρασης και του συναισθήματος.

Η διατριβή απαρτίζεται από τέσσερα κεφάλαια. Στο πρώτο κεφάλαιο εξετάζεται η εξαγωγή των χαρακτηριστικών του προσώπου. Γίνεται μια σύντομη ανασκόπηση στις τρέχουσες τεχνικές, δίνεται μια πιο αναλυτική περιγραφή των προβλημάτων της περιοχής και προτείνεται μία τεχνική συγχώνευσης, που επιτρέπει την αξιόπιστη εξαγωγή των αντικειμένων του προσώπου χωρίς χειρωνακτική αρχικοποίηση. Η τεχνική μπορεί να χρησιμοποιηθεί είτε μόνη της είτε ως το στάδιο αρχικοποίησης άλλων τεχνικών παρακολούθησης αντικειμένων και είναι ικανή να χειριστεί περιπτώσεις μη ακραίων συναισθημάτων έχοντας ικανοποιητικά αποτελέσματα σε μεγάλη ποικιλία ποιότητας/ανάλυσης εικόνων και προσώπων.

Στο δεύτερο κεφάλαιο εξετάζεται η εκτίμηση των παραμέτρων FAP (Facial Animation Parameters) του μοντέλου προσώπου MPEG-4 καθώς και η εκτίμηση της έκφρασης μέσω αυτών. Παρουσιάζεται ο τρόπος δημιουργίας των ακολουθιών ρεαλιστικών συναισθημάτων που αναλύθηκαν καθώς και ο τρόπος αναπαράστασης των συναισθημάτων. Παρουσιάζονται τέλος δύο τρόποι χρήσης των MPEG-4 FAP για την εκτίμηση της έκφρασης, ένας που βασίζεται στην δυνατοτική αποτίμηση ασαφών κανόνων και λαμβάνει υπόψη τους βαθμούς εμπιστοσύνης του σταδίου εξαγωγής χαρακτηριστικών και ένας που βασίζεται σε ασαφές νευρωνικό δίκτυο, ο οποίος παρέχει τη δυνατότητα προσαρμογής των

εκτιμήσεων σε συγκεκριμένους χρήστες.

Στο τρίτο κεφάλαιο, εξετάζεται η ανάλυση ακολουθιών εικόνων με στόχο τον εμπλουτισμό της εκτίμησης συναισθήματος και με πληροφορία σχετιζόμενη με την ανάλυση χειρονομιών. Στο κεφάλαιο εξετάζεται το πρώτο στάδιο της ανάλυσης, αυτό της κατάτμησης των εικόνων. Παρουσιάζονται τεχνικές κατάτμησης και εξαγωγής αντικειμένων με βάση το χρώμα, την κίνηση, καθώς και το πεδίο βάθους στην περίπτωση των στερεοσκοπικών ακολουθιών. Για το συνδυασμό των επιμέρους κατατμήσεων, προτείνεται μία τεχνική συγχώνευσης βασισμένη σε ασαφείς αποφάσεις, η οποία επιδέχεται σημασιολογικούς κανόνες για τη συγχώνευση των πολλαπλών κατατμήσεων και τη βελτίωση της ποιότητας της τελικής κατάτμησης. Στο τέταρτο κεφάλαιο καθώς και στο τέλος κάθε κεφαλαίου παρατίθενται αποτελέσματα των προτεινόμενων τεχνικών, καθώς και συμπεράσματα και προτάσεις για ζητήματα που θα μπορούσαν να αποτελέσουν τη βάση για μελλοντική έρευνα.

## ABSTRACT

The main research area of this Ph.D. thesis is facial expression recognition through facial feature analysis. A novel multicue feature extraction technique is proposed which is able to perform well under a large variation of facial image quality, color and resolution. MPEG-4 features are estimated and used for fuzzy rule-based facial expression and emotion recognition. Real data and results are presented, involving both extreme and intermediate expression/emotional states, obtained within the sensitive artificial listener HCI environment that was generated in the framework of related European projects. Two approaches are presented for facial expression recognition, a possibilistic fuzzy rule evaluation approach which is able to handle uncertainty provided by the feature extraction stage, and a neurofuzzy network, able to adapt its knowledge and estimations to specific users. Gesture analysis is also being considered as an additional cue, and towards this goal a multiple-cue segmentation technique is proposed, which can be used by higher-level gesture analysis systems. The proposed system investigates fuzzy data fusion techniques which are capable of integrating the results of multiple cue segmentations, and provide time consistent spatiotemporal image partitions corresponding to moving objects.





## Ευχαριστίες

Θα ήθελα να εκφράσω τις ειλικρινείς ευχαριστίες μου σε όλους όσους με βοήθησαν κατά τη διάρκεια της διδακτορικής μου διατριβής, τα τελευταία πέντε χρόνια. Καταρχάς θα ήθελα να ευχαριστήσω τον επιβλέποντα Καθηγητή, κ. Στέφανο Κόλλια, για την πολύτιμη βοήθεια και υποστήριξή του, τόσο σε επιστημονικό, όσο και σε προσωπικό επίπεδο. Τα σχόλια και η εποικοδομητική κριτική του στην ουσία και στο κείμενο της διατριβής και των δημοσιεύσεων βοήθησαν καθοριστικά στην βελτίωση της ποιότητάς τους. Η ελευθερία που μου διέθεσε και η αισιοδοξία που μου ενέπνευσε στην πορεία της διατριβής ήταν αποφασιστικές για την ανάπτυξη και την αξιοποίηση νέων ιδεών.

Θα ήθελα επίσης να ευχαριστήσω τους συναδέλφους μου στο Εργαστήριο Ψηφιακής Επεξεργασίας Εικόνων, Βίντεο και Πολυμέσων για τη βοήθειά τους στη διατήρηση ενός δημιουργικού περιβάλλοντος εργασίας και ιδιαίτερα εκείνους με τους οποίους συνεργάστηκα στο πλαίσιο της ερευνητικής μου εργασίας. Ειδικότερα είμαι υπόχρεος στους συναδέλφους Κώστα Καρπούζη, Γιάννη Αβρίθη, Γιώργο Στάμου, Μανόλη Γουάλλες, Βασίλη Τζουβάρα, Αμαρυλλίς Ραουζαίου, Νίκο Τσαπατσούλη, Θεόφιλο Μαΐλη και Νίκο Σίμου, για την πολύτιμη προσφορά και τις ιδέες τους στα θέματα που παρουσιάζονται στη διατριβή. Επίσης θα ήθελα να εκφράσω την εκτίμησή μου προς τους συναδέλφους Τάσο Ντελόπουλο, Νάσο Δροσόπουλο και Γιώργο Ακρίβα και για τις συνεργασίες που είχαμε κατά τη διάρκεια της διατριβής.

Οφείλω να εκφράσω τις θερμές ευχαριστίες μου στη φίλη μου Αναστασία Χαλκιοπούλου για τη συνεχή ενθάρρυνση που μου παρείχε στη διάρκεια της ερευνητικής μου δραστηριότητας, καθώς και για την υπομονή που επέδειξε κατά τη διάρκεια συγγραφής της διατριβής. Τέλος, είμαι ευγνώμων προς την οικογένειά μου για την ανεκτίμητη συμπαράστασή της σε όλη τη διάρκεια των σπουδών μου. Χωρίς την πολύπλευρη βοήθειά της η ολοκλήρωση της διδακτορικής διατριβής δε θα ήταν δυνατή.



# Κεφάλαιο 1

## Εξαγωγή των χαρακτηριστικών του προσώπου

### 1.1 Εισαγωγή

Η ανάλυση εκφράσεων και η αναγνώριση συναισθημάτων, μια θεματική περιοχή που παραδοσιακά ανήκει στην ψυχολογία, έχει κερδίσει τις τελευταίες δεκαετίες την προσοχή της κοινότητας των μηχανικών. Παρατηρείται τελευταία αυξημένο ενδιαφέρον προς τη βελτίωση των διαπροσωπικών ανθρώπου-μηχανής και τον εμπλουτισμό τους με συναισθηματικό περιεχόμενο. Οι ερευνητικές προσπάθειες μπορούν να συνοψιστούν στην ανάγκη ποσοτικής περιγραφής των εκφράσεων του προσώπου [1], αλλά και στην αυτοματοποίηση της διαδικασίας αναγνώρισης εκφράσεων [2] οι οποίες συνδέονται και με τη συναισθηματική κατάσταση [3].

Οι τεχνικές αναγνώρισης της έκφρασης μπορούν να χωριστούν σε δύο βασικές κατηγορίες: την ολιστική όπου τα πρόσωπα εξετάζονται στην ολότητά τους και την αναλυτική που βασίζεται σε εξαγωγή χαρακτηριστικών.

Οι ολιστικές προσεγγίσεις βασίζονται στη σύγκριση του προσώπου με κάποιο πρότυπο. Το πρότυπο αυτό μπορεί να είναι είτε μια εικόνα προσώπου, είτε διανύσματα χαρακτηριστικών που έχουν εξαχθεί αναλύοντας την εικόνα είτε κατά τμήματα είτε στην ολότητά της. Σύμφωνα με τις τεχνικές αυτές, λαμβάνεται ένας μεγάλος αριθμός από δείγματα εικόνων-προτύπων που παριστάνουν κάποια γνωστή εκ των προτέρων έκφραση στην κορύφωσή της και οι υπόλοιπες εικόνες συγκρίνονται με αυτές. Η πιο συνήθης τεχνική είναι η μείωση των διαστάσεων των προτύπων μέσω ανάλυσης βασικών συνιστωσών (principal component analysis) και η χρήση νευρωνικών δικτύων για τη σύγκριση. Μια από τις πιο γνωστές εργασίες που χρησιμοποιεί την παραπάνω τεχνική είναι η [87] η οποία αναφέρεται στα πρόσωπα-πρότυπα, με τον ευρέως πλέον χρησιμο-

ποιούμενο όρο *eigenfaces* (ιδιοπρόσωπα), λόγω των ιδιοδιανυσμάτων της ανάλυσης βασικών συνιστωσών που τα απαρτίζουν. Οι πρόσφατες ολιστικές τεχνικές ακολουθούν την ίδια γενική προσέγγιση και περιλαμβάνουν ταίριασμα προτύπων εικόνων φιλτραρισμένων με *Gabor wavelets* [65] ή ταίριασμα χαρτών οπτικής ροής [9]. Μια άλλη τεχνική περιλαμβάνει την αυτόματη τοποθέτηση ελαστικών γράφων στο πρόσωπο όπου οι κόμβοι των γράφων αντιστοιχούν σε συγκεκριμένες αποκρίσεις *Gabor wavelets* των περιοχών του προσώπου που αντιστοιχούν. Η αναγνώριση της έκφρασης γίνεται με σύγκριση των παραγόμενων γράφων με τους γράφους-πρότυπα όπως στην [25].

Οι ολιστικές τεχνικές αποτελούν μέχρι σήμερα τη συντριπτική πλειοψηφία των τεχνικών εκτίμησης έκφρασης, κυρίως λόγω του γεγονότος του ότι αντιμετωπίζουν το εν λόγω πρόβλημα σαν ένα τυπικό πρόβλημα ανίχνευσης προτύπων, συνεπώς το βάρος δίνεται κυρίως στους ταξινομητές (*classifiers*) οπότε δεν απαιτείται γνώση του τρόπου σχηματισμού των εκφράσεων, παρά μόνο η ύπαρξη χαρακτηρισμένων εικόνων προς εκμάθηση. Επιπλέον οι ολιστικές προσεγγίσεις προτιμώνται έναντι των αναλυτικών για δύο επιπλέον λόγους: την διαθεσιμότητα πολλών βάσεων με στατικές εικόνες και την παράκαμψη των επίπονων σταδίων προεπεξεργασίας που απαιτούνται πριν τη σύγκριση των εικόνων (κανονικοποίηση, ευθυγράμμιση). Τα πρότυπα που χρησιμοποιούνται ανήκουν στις έξι καθολικές ή αρχετυπικές εκφράσεις, χαρά, έκπληξη, φόβος, θυμός, αποστροφή, λύπη που περιγράφονται στην ενότητα 2.2.3, επομένως είναι πολύ δύσκολο τα συστήματα αυτά να γενικεύσουν στην περίπτωση των ενδιάμεσων συναισθημάτων. Το μεγαλύτερο πλεονέκτημα αυτών των μεθόδων είναι ότι μπορούν να εκπαιδευτούν και σε εικόνες χαμηλής ανάλυσης ή λεπτομέρειας, παραμένουν όμως ευαίσθητες στις διακυμάνσεις της έκφρασης, της γωνίας του προσώπου, του θορύβου και προαπαιτούν ακριβή κανονικοποίηση και ευθυγράμμιση των χαρακτηριστικών του προσώπου των υπό ανάλυση φωτογραφιών.

Στα αναλυτικά συστήματα από την άλλη, ανιχνεύεται η θέση και η μορφή των βασικών χαρακτηριστικών του προσώπου. Υπολογίζονται αποστάσεις μεταξύ προκαθορισμένων χαρακτηριστικών σημείων και συνήθως υπολογίζονται οι παράμετροι κάποιου μοντέλου προσώπου. Η εκτίμηση των εκφράσεων γίνεται εξετάζοντας τη μετατόπιση των χαρακτηριστικών του προσώπου συνήθως μέσω της αποτίμησης κανόνων έκφρασης. Η αυτόματη εκτίμηση των παραμέτρων ενός μοντέλου προσώπου αποτελεί ένα αρκετά δυσεπίλυτο πρόβλημα, και γι' αυτό το λόγο στη βιβλιογραφία έχουν προταθεί σχετικά λίγες λύσεις [4]. Η δυσκολία των μεθόδων αυτών έγκειται αφενός στο ότι προϋποθέτουν την ύπαρξη γνώσης για τον τρόπο που εκφράζονται τα συναισθήματα στο πρόσωπο και αφετέρου στο ότι απαιτούν την ακριβή ανίχνευση των χαρακτηριστικών του προσώπου ή

το ταίριασμα κάποιου μοντέλου σε αυτά. Το πλεονέκτημα αυτών των μεθόδων είναι ότι επιτρέπουν την άμεση επαλήθευση και εξαγωγή της γνώσης αναπαράστασης συναισθημάτων.

Οι κλασικές μέθοδοι κατάτμησης δεν μπορούν να χρησιμοποιηθούν για την κατάτμηση του προσώπου στα χαρακτηριστικά του, γιατί η κλίμακα των εικόνων στις πραγματικές ακολουθίες είναι τόσο μικρή που καθιστά την πληροφορία π.χ. χρώματος ή υφής στο εσωτερικό των χαρακτηριστικών μη εκμεταλλεύσιμη. Επιπλέον απαιτείται γνώση για τη συγκεκριμένη δομή των υπό ανίχνευση χαρακτηριστικών. Για την εξαγωγή των χαρακτηριστικών και το ταίριασμα μοντέλων, στη βιβλιογραφία έχουν προταθεί αρκετές απλές μέθοδοι που συνήθως βασίζουν την ανίχνευση της θέσης των χαρακτηριστικών του προσώπου στην πληροφορία του χρώματος ή στις προβολές φωτεινότητας [102] [63]. Για την ανίχνευση των «δύσκολων» χαρακτηριστικών όπως οι οφθαλμοί και το στόμα, έχουν προταθεί διάφορες τεχνικές που βασίζονται στην προσαρμογή απλών αλλά επαρκών μοντέλων που κατασκευάζονται ειδικά για κάθε χαρακτηριστικό. Τα συνήθη μοντέλα μπορεί να είναι τμήματα ελλείψεων, παραβολών ή πιο σύνθετα που προσαρμόζονται με κάποιο κριτήριο ελαχιστοποίησης ενέργειας ή με κάποια παραλλαγή του μετασχηματισμού hough [101]. Τα φίδια (snakes) γενικά δεν προτιμώνται λόγω της τοπικής επίδρασης των δυνάμεων που καθορίζουν το σχήμα τους, ενώ τα πλήρως ανελαστικά πρότυπα αδυνατούν να ταιριάζουν σε χαρακτηριστικά μεταβλητού σχήματος όπως το στόμα. Τα πιο ευέλικτα μοντέλα μοιάζουν να είναι όσα βασίζονται σε παραμορφούμενα πρότυπα. Ήδη εδώ και μια δεκαετία τα παραμορφούμενα πρότυπα προτείνονται για την προσαρμογή στοματικών και οφθαλμικών μοντέλων όπως στις [98][99][100]. Οι περισσότερες τεχνικές που έχουν προταθεί για το σκοπό αυτό, δε σχετίζονται αναγκαστικά με την αναγνώριση έκφρασης, αλλά επικεντρώνονται στο ταίριασμα μοντέλων χαρακτηριστικών του προσώπου, προτείνοντας συνήθως ένα απλό αλλά επαρκές μοντέλο περιγραφής των χαρακτηριστικών που εξετάζουν. Το επίκεντρο του ενδιαφέροντος τους είναι ο ορισμός ενός επαρκούς μοντέλου περιγραφής και όχι ο χώρος στον οποίο θα προσαρμοστεί το μοντέλο αυτό στις πραγματικές εικόνες, απλουστεύοντας κατ'αυτόν τον τρόπο το προς επίλυση πρόβλημα με την εισαγωγή περιορισμών που συρρικνώνουν τα χαρακτηριστικά του εξεταζόμενου δείγματος. Έτσι π.χ. οι [98][100][102] και [101] υποθέτουν εικόνες οφθαλμού υψηλής ευκρίνειας και ομοιόμορφου φωτισμού, ενώ στη [99] τα χαρακτηριστικά του δείγματος είναι τόσο περιορισμένα που το προτεινόμενο μοντέλο μοιάζει να προσαρμόζεται μόνο για την λήψη στατιστικών στοιχείων για τα γεωμετρικά στοιχεία των οφθαλμών. Επιπλέον υπάρχουν και μέθοδοι υπολογισμού παραμέτρων κάποιου μοντέλου προσώπου που βασίζονται στην ανίχνευση της κίνησης

σημείων στο πρόσωπο, χωρίς αυτά να αντιστοιχούν στα χαρακτηριστικά του προσώπου αλλά να αποτελούν σημεία κατάλληλα προς παρακολούθηση [44][88]. Ανεξάρτητα με το είδος του μοντέλου, στο χώρο τόσο της αρχικοποίησής τους αλλά και της προσαρμογής τους σε χαρακτηριστικά που είναι υπαρκτά σε μεγάλη ποικιλία εικόνων πραγματικών ακολουθιών, παρατηρείται ένα έλλειμμα προτάσεων.

Τα περισσότερα συστήματα ανάλυσης εκφράσεων χρησιμοποιούν το σύστημα FACS (Facial Action Coding System) από τους Ekman και Friesen [3] στο οποίο γίνεται μια προσπάθεια να περιγραφούν οι εκούσιες κινήσεις που παρατηρούνται στο πρόσωπο από ένα σύνολο χαρακτηριστικών παραμέτρων, οι οποίες δρουν είτε ατομικά, είτε σε συνδυασμό. Η ανάλυση δεν επεκτείνεται ιδιαίτερα στο θέμα των συνολικών εκφράσεων, ούτε στη σημασιολογία τους και την πιθανή σχέση με μετρήσιμες μεθόδους, αλλά επικεντρώνεται στη σύνδεση των στοιχειωδών αυτών δράσεων με δεδομένα φυσιολογίας του ανθρώπινου προσώπου. Το μοντέλο FACS περιγράφει τις εκφράσεις χρησιμοποιώντας 44 μονάδες δράσης - Action Units (AU) οι οποίες σχετίζονται με τη συστολή συγκεκριμένων προσωπικών μυών. Εκτός του FACS, οι μονάδες το συστήματος MPEG-4 [5] χρησιμοποιούνται συχνά για την μοντελοποίηση των εκφράσεων του προσώπου και των υποκείμενων συναισθημάτων. Το μοντέλο MPEG-4 παρέχει ένα εναλλακτικό τρόπο μοντελοποίησης των εκφράσεων ο οποίος έχει δεχθεί σημαντικές επιρροές από ψυχολογικές και νευροφυσιολογικές μελέτες. Το MPEG-4 εστιάζοντας κυρίως στην σύνθεση εκφράσεων και στο animation, ορίζει τις παραμέτρους κίνησης προσώπου - Facial Animation Parameters (FAPs) οι οποίες είναι στενά συνδεδεμένες με τα Action Units, τον πυρήνα του FACS. Μια σύγκριση μεταξύ των δύο συστημάτων βρίσκεται στο [6].

Τα περισσότερα συστήματα αναγνώρισης έκφρασης προσώπου προσπαθούν να αντιστοιχίσουν τις εκφράσεις κατευθείαν στις αρχέτυπες συναισθηματικές κατηγορίες, μη όντας ικανά να διαχειριστούν εκφράσεις που δημιουργούνται από ενδιάμεσες συναισθηματικές κατηγορίες ή εκφράσεις που δεν αντιστοιχούν σε συναισθήματα. Τελευταία έχουν προταθεί αρκετά συστήματα που είναι ικανά να διακρίνουν την ένταση των εκφράσεων [7][8][9][10], αλλά μόνο λίγα [4][11] έχουν τη δυνατότητα να πετύχουν ανάλυση βασισμένη σε κάποιο μοντέλο προσώπου (MPEG-4 ή FACS). Οι περισσότερες υπάρχουσες προσεγγίσεις εξαγωγής χαρακτηριστικών προσώπου είτε είναι σχεδιασμένες να αντιμετωπίσουν μειωμένη ποικιλία χαρακτηριστικών των εικόνων εισόδου, είτε απαιτούν χειρωνακτική αρχικοποίηση ή παρέμβαση. Πιο συγκεκριμένα, η [4] εξαρτάται από την οπτική ροή, οι [12][13][14][15][16] απαιτούν βίντεο υψηλής ανάλυσης και απουσία θορύβου, οι [17][18][19] εξαρτώνται από την πληροφορία χρώματος, οι [14][20] απαιτούν χει-

ρωνακτική αρχικοποίηση, η [11] απαιτεί την ύπαρξη ειδικών διακριτικών (markers) στα πρόσωπα, οι [21][13] απαιτούν τη χειρωνακτική επιλογή των χαρακτηριστικών σημείων (feature points) στο πρώτο καρέ, η [22] απαιτεί την ανάρτηση δύο καμερών στο κεφάλι και οι [23][24][25][26] απαιτούν ειδική εκπαίδευση ανά χρήστη ή ανά έκφραση είτε κατά την εξαγωγή χαρακτηριστικών είτε κατά την εκτίμηση έκφρασης. Από τις παραπάνω εργασίες, οι [24][22][7][26][12][20] παραθέτουν αποτελέσματα επιτυχίας μόνο για την ανάλυση έκφρασης και όχι για τη διαδικασία εξαγωγής χαρακτηριστικών. Επιπρόσθετα ελάχιστες προσεγγίσεις μπορούν να εφαρμοστούν σε σχεδόν πραγματικό χρόνο.

Οι συνήθεις απλές και γρήγορες μεθοδολογίες ανίχνευσης του προσώπου και των χαρακτηριστικών του σε ακολουθίες εικόνων, βασίζονται και στην ανίχνευση του χρώματος του δέρματος. Αυτό συνήθως επιτυγχάνεται με τον υπολογισμό της δεσμευμένης *a posteriori* πιθανότητας ενός pixel να ανήκει στην κλάση «δέρμα» στο χώρο  $C_r/C_b$ . Έχουν προταθεί αρκετά επιπλέον χρωματικά μοντέλα (color spaces) τα οποία εκμεταλλεύονται τα ιδιαίτερα χαρακτηριστικά των χαρακτηριστικών του προσώπου [27].

Τα συστήματα βίντεο από την άλλη, επεξεργάζονται την πληροφορία με τη μορφή τριών συνιστωσών εκ των οποίων η μία μεταφέρει τη φωτεινότητα (lightness-luma) και οι άλλες δύο τις χρωματικές συνιστώσες (chroma), αγνοώντας τη φωτεινότητα. Αυτά τα συστήματα εκμεταλλεύονται τη χαμηλή ικανότητα αντίληψης εναλλαγής χρωμάτων της ανθρώπινης όρασης: εάν η φωτεινότητα είναι πλήρης λεπτομερειών, οι λεπτομέρειες στις χρωματικές συνιστώσες μπορούν να μειωθούν με κάποια μέθοδο υποδειγματοληψίας. Αυτή η τεχνική χρησιμοποιείται σε όλα τα συστήματα μετάδοσης σήματος και αποθήκευσης βίντεο με εξαίρεση ελάχιστα format που χρησιμοποιούνται για αρχειακούς σκοπούς. Σχεδόν όλα τα υπάρχοντα πρότυπα μετάδοσης και καταγραφής βίντεο προϋποθέτουν μειωμένη χρωματική ανάλυση τόσο στην οριζόντια όσο και στην κάθετη διάσταση. Σήματα βίντεο που ακολουθούν το πρότυπο 4:2:0 όπως τα H-261, και MPEG-2 όπου οι συνιστώσες  $C_r$  και  $C_b$  έχουν υποδειγματοληφθεί κατά 2 οριζόντια και κάθετα, θεωρούνται ακόμα και σήμερα ως σήματα πολύ καλής ποιότητας. Η παρατηρούμενη ποιότητα είναι πράγματι πολύ καλή, αλλά εάν η ανάλυση του σήματος φωτεινότητας είναι αρκετά χαμηλή ή το πρόσωπο καταλαμβάνει μικρό μόνο ποσοστό της συνολικής επιφάνειας του καρέ, τότε συμβαίνει συχνά ολόκληρα χαρακτηριστικά του προσώπου να μοιράζονται την ίδια πληροφορία χρωματικής συνιστώσας, με αποτέλεσμα η τελευταία να καθίσταται ακατάλληλη για την ανάλυση των χαρακτηριστικών του προσώπου. Επιπρόσθετα η υπερέκθεση στην περιοχή του προσώπου, φαινόμενο αρκετά συχνό λόγω της υψηλής ανακλασιμότητας της περιοχής αυτής, αλλά και οι χρωματι-

κές αλλοιώσεις λόγω κωδικοποίησης και συμπίεσης, καθιστούν την αξιοποίηση των χρωματικών συνιστωσών ιδιαίτερα προβληματική σε πολλές πραγματικές ακολουθίες βίντεο. Για την αντιμετώπιση των παραπάνω προβλημάτων έχουν προταθεί διάφορες τεχνικές, όπως η [28], αλλά δεν έχουν παρατηρηθεί σημαντικές βελτιώσεις.

Στο πλαίσιο των ευρωπαϊκών προγραμμάτων τεχνολογίας και πληροφορίας ERMIS [29] και HUMANE [30], δημιουργήθηκε μια εκτενής οπτικοακουστική βάση δεδομένων η οποία αποτελείται από ανθρώπους που οδηγούνται σε εκφραστική αλληλεπίδραση με ειδικούς. Οι συμμετέχοντες στο πείραμα δεν προσποιούνται τις εκφράσεις τους, ως εκ τούτου το μεγαλύτερο μέρος του υλικού αποτελείται από ήπια συναισθήματα τα οποία είναι πολύ δύσκολο να εντοπιστούν ακόμα και από ειδικούς ψυχολόγους, ειδικά εάν αγνοήσουμε το ηχητικό σήμα.

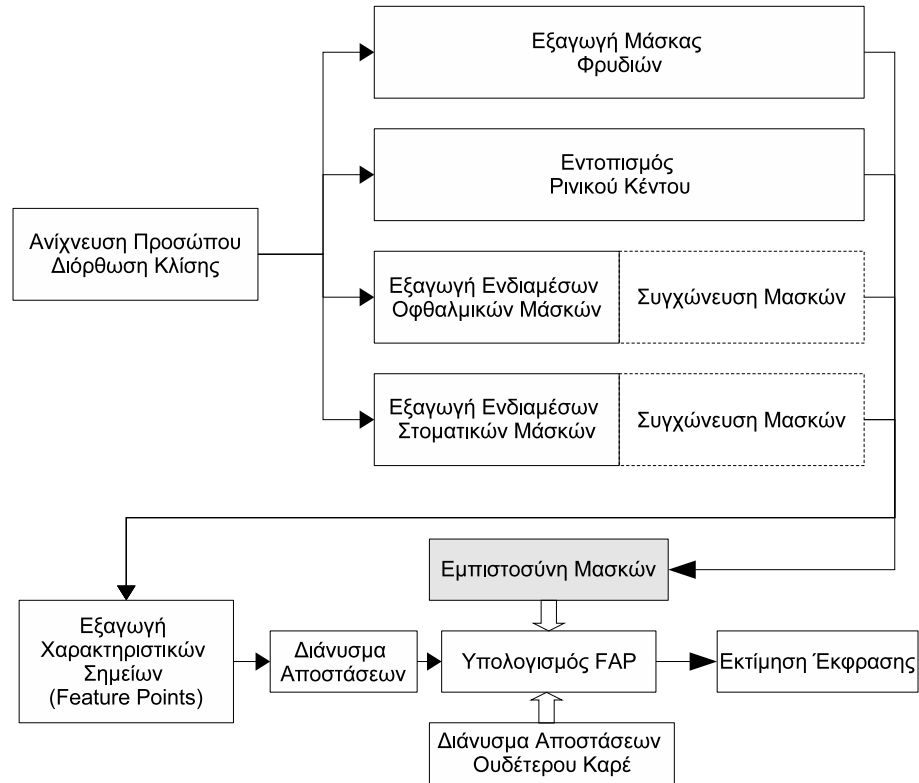
Οι βασικοί στόχοι του συστήματος ανάλυσης εκφράσεων που περιγράφεται στη συνέχεια, είναι η ικανότητά του να διαχειριστεί και να αναλύσει μη ακραίες εκφράσεις προσώπου καθώς και η ικανότητά του να διαχειρίζεται εκφράσεις από διαφορετικά πρόσωπα. Η περιγραφόμενη προσέγγιση δοκιμάστηκε σε πραγματική αλληλεπίδραση ανθρώπου-μηχανής χρησιμοποιώντας το σύστημα SAL (Sensitive Artificial Listener) [29][30], το οποίο περιγράφεται επίσης εν συντομία στην ενότητα 2.1. Προκειμένου να παρακαμφθεί το πρόβλημα της μεγάλης ποικιλίας στα χαρακτηριστικά και την ποιότητα των υπό ανάλυση εικόνων όσον αφορά στην ανάλυση χρώματος και φωτεινότητας, αναπτύχθηκε μια αναλυτική μέθοδος αναγνώρισης εκφράσεων η οποία επιτρέπει την αυτόματη αναγνώριση των χαρακτηριστικών σημείων σε εικόνες προσώπου, με στόχο τον υπολογισμό παραμέτρων του μοντέλου MPEG-4. Η μέθοδος που προτείνεται επιτρέπει το χαρακτηρισμό της εισόδου με τη βοήθεια προφίλ εκφράσεων. Η έκφραση εκτιμάται δια μέσου της ανάλυσης των MPEG FAPs [31], τα οποία υπολογίζονται δια της ανίχνευσης της κίνησης και της παραμόρφωσης ορισμένων χαρακτηριστικών του προσώπου, όπως είναι το στόμα, οι οφθαλμοί και τα φρύδια, υποθέτοντας όμως την διαθεσιμότητα και της «ουδέτερης έκφρασης» του υπό ανάλυση προσώπου. Η προτεινόμενη μέθοδος έχει την ικανότητα ανίχνευσης και των βασικών αλλά και των ενδιάμεσων εκφράσεων (π.χ. ανία, θυμός) [6] με τα αντίστοιχα επίπεδα έντασης και βαθμού εμπιστοσύνης των αποτελεσμάτων.

## **1.2 Επισκόπηση της προτεινόμενης μεθοδολογίας**

Μια επισκόπηση της προτεινόμενης μεθοδολογίας απεικονίζεται στο σχήμα 1.1 ενώ μια πιο αναλυτική επισκόπηση της διαδικασίας συγχώνευσης (που παριστάνεται με διακεκομμένες γραμμές) απεικονίζεται στο σχήμα 1.16. Αρχικά ανι-



## Κεφάλαιο 1. Εξαγωγή των χαρακτηριστικών του προσώπου



Σχήμα 1.1: Επισκόπηση του συστήματος

χνεύεται το πρόσωπο έτσι ώστε να γίνει μια αρχική εκτίμηση των θέσεων των χαρακτηριστικών του. Ανιχνεύεται και διορθώνεται επίσης η στροφή του προσώπου στο επίπεδο της κάμερας και το πρόσωπο κατατέμνεται στις εξής περιοχές: αριστερός οφθαλμός/φρύδι, δεξί οφθαλμός/φρύδι, μύτη και στόμα. Κάθε μία από τις τέσσερις περιοχές που στη συνέχεια θα αναφέρονται ως *περιοχές ανίχνευσης χαρακτηριστικών - ΠΑΧ* περιέχουν τα χαρακτηριστικά των οποίων τα όρια θέλουμε να εξάγουμε. Στο εσωτερικό των αντίστοιχων ΠΑΧ διενεργείται εξαγωγή χαρακτηριστικών υψηλότερης ακρίβειας για κάθε χαρακτηριστικό του προσώπου ξεχωριστά (οφθαλμοί, φρύδια, στόμα και μύτη). Χρησιμοποιώντας μια προσέγγιση πολλαπλών στοιχείων-προτύπων (multiple cue) δημιουργείται ένας μικρός αριθμός από ενδιάμεσες μάσκες χαρακτηριστικών · οι ενδιάμεσες μάσκες κάθε χαρακτηριστικού συγχωνεύονται στην τελική μάσκα κάθε χαρακτηριστικού. Η διαδικασία συγχώνευσης (fusion) λαμβάνει υπόψη της ανθρωπομετρικά κριτήρια [32] με στόχο να αξιολογήσει και να αναθέσει κατάλληλα βάρη στις επιμέρους ενδιάμεσες μάσκες. Η αξιολόγηση αυτή χρησιμοποιείται τόσο κατά τη συγχώνευση όσο και κατά την απόδοση μιας τιμής εμπιστοσύνης στην τελική μάσκα.

Ο υπολογισμός των FAP αξιώνει την ύπαρξη ενός καρέ όπου το υπό ανά-

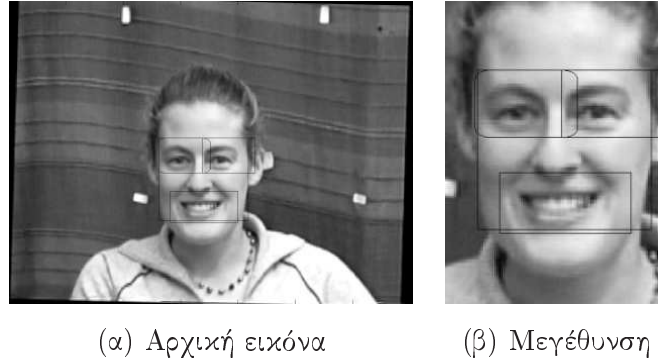
λυση πρόσωπο θα έχει ουδέτερη έκφραση. Αυτό το καρέ θα ονομάζεται εφεξής «ουδέτερο καρέ» και επιλέγεται είτε χειρωνακτικά από τις υπό ανάλυση ακολουθίες βίντεο είτε παρέχεται διαδραστικά στο σύστημα. Οι τελικές μάσκες κάθε χαρακτηριστικού χρησιμοποιούνται στην εξαγωγή 19 σημείων στο πρόσωπο, τα οποία αποτελούν υποσύνολο των MPEG-4 Feature Points, FPs [6]. Η θέση των 19 χαρακτηριστικών σημείων που εξάγονται από κάθε καρέ συγκρίνεται με την αντίστοιχη θέση τους στο ουδέτερο καρέ ώστε να εκτιμηθούν οι κινήσεις των χαρακτηριστικών του προσώπου και να εξαχθούν οι παράμετροι κίνησης προσώπου, FAPs. Οι τιμές εμπιστοσύνης για το κάθε FAP υπολογίζονται από τις αντίστοιχες τιμές των FP από τα οποία προκύπτει. Τα FAPs σε συνδυασμό με τις τιμές εμπιστοσύνης τους χρησιμοποιούνται στη συνέχεια για την εκτίμηση της έκφρασης του προσώπου.

### 1.3 Ανίχνευση του προσώπου και της περιστροφής του

Στην προτεινόμενη μέθοδο τα χαρακτηριστικά του προσώπου αρχικά ανιχνεύονται και εντοπίζονται. Το πρώτο βήμα λοιπόν του εντοπισμού του προσώπου και της εκτίμησης της γωνίας στροφής του περιγράφεται εν συνεχεία, και ακολουθείται από την ακριβή εξαγωγή των μασκών των χαρακτηριστικών του προσώπου που περιγράφεται στο κεφάλαιο 1.4. Στο παρόν στάδιο υποθέτουμε ότι μας διατίθεται η εικόνα του προσώπου σε ουδέτερη έκφραση.

Ο στόχος της διαδικασίας ανίχνευσης προσώπων είναι να προσδιορίσει εάν υπάρχουν πρόσωπα στην εικόνα και αν υπάρχουν να επιστραφεί η θέση και η έκταση που καταλαμβάνει το καθένα [33]. Η ανίχνευση μπορεί να γίνει με ποικίλους τρόπους [34],[35],[36]. Το παρόν σύστημα χρησιμοποιεί μη παραμετρική διαχωριστική ανάλυση με τη χρήση μηχανής διανυσμάτων υποστήριξης (support vector machine - SVM) η οποία ταξινομεί τις περιοχές της εικόνας σε περιοχές προσώπου και μη-προσώπου μειώνοντας τη διάσταση του προβλήματος εκπαίδευσης σε ένα κλάσμα της αρχικής, με αμελητέα απώλεια της απόδοσης [37],[29].

Το σύστημα εκπαιδεύτηκε σε 800 παραδείγματα προσώπων από την βάση δεδομένων 18 του NIST. Οι εικόνες ευθυγραμμίστηκαν όσον αφορά τις συντεταγμένες των ματιών και του στόματος και μετατράπηκαν στο απαιτούμενο μέγεθος. Στο σύνολο των αρχικών δεδομένων εφαρμόστηκαν μικρής κλίμακας διαταράξεις μετατόπισης και περιστροφής, επεκτείνοντάς το στα 16695 παραδείγματα.



Σχήμα 1.2: Περιοχές ανίχνευσης χαρακτηριστικών

Περιοχή	Τοποθεσία	Πλάτος	Ύψος
Οφθαλμοί και φρύδια	άνω τμήματα του προσώπου	$0.6W_f$	$0.5H_f$
Μύτη και στόμα	κάτω τμήμα του προσώπου	$W_f$	$0.5H_f$

Πίνακας 1.1: Ορισμός Περιοχών Ανίχνευσης Χαρακτηριστικών

Το βήμα της ανίχνευσης προσώπου παρέχει ένα παραλληλόγραμμο όριο μέσα στο οποίο περιέχεται το πρόσωπο όπως απεικονίζεται στην εικόνα 1.2. Το τελευταίο κατατμείται αρχικά στις περιοχές ανίχνευσης χαρακτηριστικών χρησιμοποιώντας στατικούς ανθρωπομετρικούς κανόνες (Εικόνα 1.2, Πίνακας 1.1) σε τρεις υπέρθετες περιοχές ενδιαφέροντος οι οποίες εμπεριέχουν τα χαρακτηριστικά του προσώπου αλλά και μέρος του φόντου. Οι περιοχές αυτές χρησιμοποιούνται στη συνέχεια για την αρχικοποίηση της διαδικασίας εξαγωγής των χαρακτηριστικών. Οι διαστάσεις του προσώπου και η ανάλυση της εικόνας δεν επηρεάζουν τη διαδικασία ορισμού των περιοχών αυτών εφόσον ορίζονται σε σχέση με το μέγεθος της έκτασης του προσώπου που παρέχεται από την διαδικασία ανίχνευσής του.

Η ορθότητα της εξαγωγής χαρακτηριστικών εξαρτάται και από την κλίση του προσώπου. Αυτό το σύστημα εμπλέκεται κυρίως με την στροφή στο επίπεδο της κάμερας (roll rotation) εφόσον είναι και η πιο συχνή που συναντάται σε πραγματικές ακολουθίες. Μικρή στροφή κατά τους δύο άλλους άξονες (yaw & pitch rotation) η οποία δεν οδηγεί σε απόκρυψη των χαρακτηριστικών του προσώπου δεν έχει σημαντικό αντίκτυπο στην εκτίμηση της έκφρασης. Η τεχνική ανίχνευσης προσώπου που περιγράφηκε προηγουμένως δύναται να αντιμετωπίσει επιτυχώς γωνίες μέχρι και  $30^\circ$ . Πρόκειται για ένα ικανοποιητικό εύρος κλίσεων, μέσα στο οποίο οι περιοχές ανίχνευσης χαρακτηριστικών είναι αρκετά μεγάλες ώστε οι οφθαλμοί να βρίσκονται εντός των περιοχών ανίχνευσης χαρακτηριστικών που ορίστηκαν στην αρχική κατάτμηση ενός περιστραμμένου προσώπου.

Για την εκτίμηση της γωνίας στροφής του προσώπου αρχικά εντοπίζεται ο αριστερός και ο δεξιός οφθαλμός στις αντίστοιχες περιοχές ανίχνευσης χαρακτηριστικών. Εν συνεχεία η κλίση υπολογίζεται από τον υπολογισμό της γωνίας που σχηματίζεται από το οριζόντιο επίπεδο και την ευθεία που ορίζουν τα κέντρα των δύο ματιών. Ο εντοπισμός των κέντρων των ματιών γίνεται μέσω μιας αποτελεσματικής τεχνικής που χρησιμοποιεί ένα πολυστρωματικό (multi-layer) με πρόσθια διάδοση (feed forward) νευρωνικό δίκτυο με σιγμοειδή συνάρτηση ενεργοποίησης. Αυτό το multi-layer perceptron (MLP) χρησιμοποιεί εκμάθηση Marquardt-Levenberg [38],[39] ενώ η βέλτιστη αρχιτεκτονική που βρέθηκε μέσω τεχνικής pruning να διαθέτει 20 κρυμμένα επίπεδα και 13 εισόδους. Για κάθε pixel στις περιοχές ανίχνευσης ματιών οι 13 είσοδοι αποτελούνται από τη συνιστώσα φωτεινότητας, τις δύο χρωματικές συνιστώσες  $C_r$  &  $C_b$  καθώς και τους 10 σημαντικότερους (με επιλογή zig-zag) συντελεστές του δισδιάστατου μετασχηματισμού συνημιτόνου (Discrete Cosine Transform - DCT) της περιβάλλουσας περιοχής, διαστάσεων  $M \times N$  pixels (εξίσωση 1.1, εικόνα 1.4).

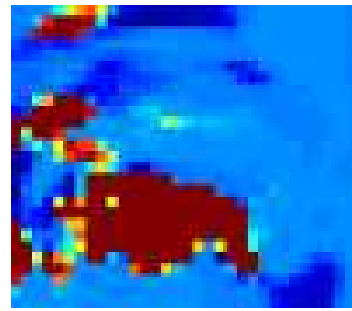
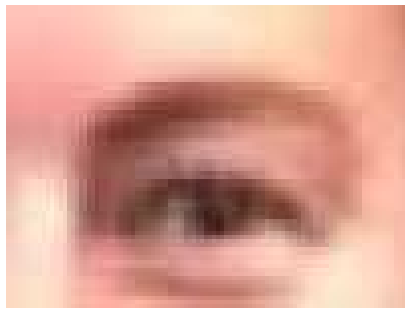
$$F(m, n) = \frac{2}{\sqrt{MN}} C(m) C(n) \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) \cos\left(\frac{(2x+1)m\pi}{2M}\right) \cos\left(\frac{(2y+1)n\pi}{2N}\right) \quad (1.1)$$

$$C(m), C(n) = \begin{cases} \frac{1}{\sqrt{2}} & m, n = 0 \\ 1 & \text{διαφορετικά} \end{cases}$$

Στη συγκεκριμένη περίπτωση επιλέχθηκε  $M, N = 8$ . Η χρήση εναλλακτικών χρωματικών χώρων όπως οι Lab, RGB και HSV για την εκπαίδευση του δικτύου δε μετέβαλε την διακριτική του ικανότητα. Το δίκτυο έχει δύο εξόδους, μία για την κάθε κλάση, «οφθαλμός» και «όχι-οφθαλμός» και εκπαιδεύτηκε σε περισσότερες από 100 χειρωνακτικά κατασκευασμένες μάσκες ματιών που ορίστηκαν σε εικόνες της βάσης ERMIS [29] ποικίλης ποιότητας, ανάλυσης και συνθηκών φωτισμού.

Η γενίκευση του δικτύου σε τυχαίες εικόνες εκτός του συνόλου εκμάθησης ήταν πολύ ικανοποιητική για τον εντοπισμό του οφθαλμού, όπως φαίνεται και στην εικόνα 1.3. Παρά ταύτα αδυνατεί να εκτιμήσει τα ακριβή όρια του οφθαλμού· ο ακριβέστερος εντοπισμός των ορίων αυτών για τον καθορισμό των  $FP$  περιγράφεται σε επόμενη ενότητα.

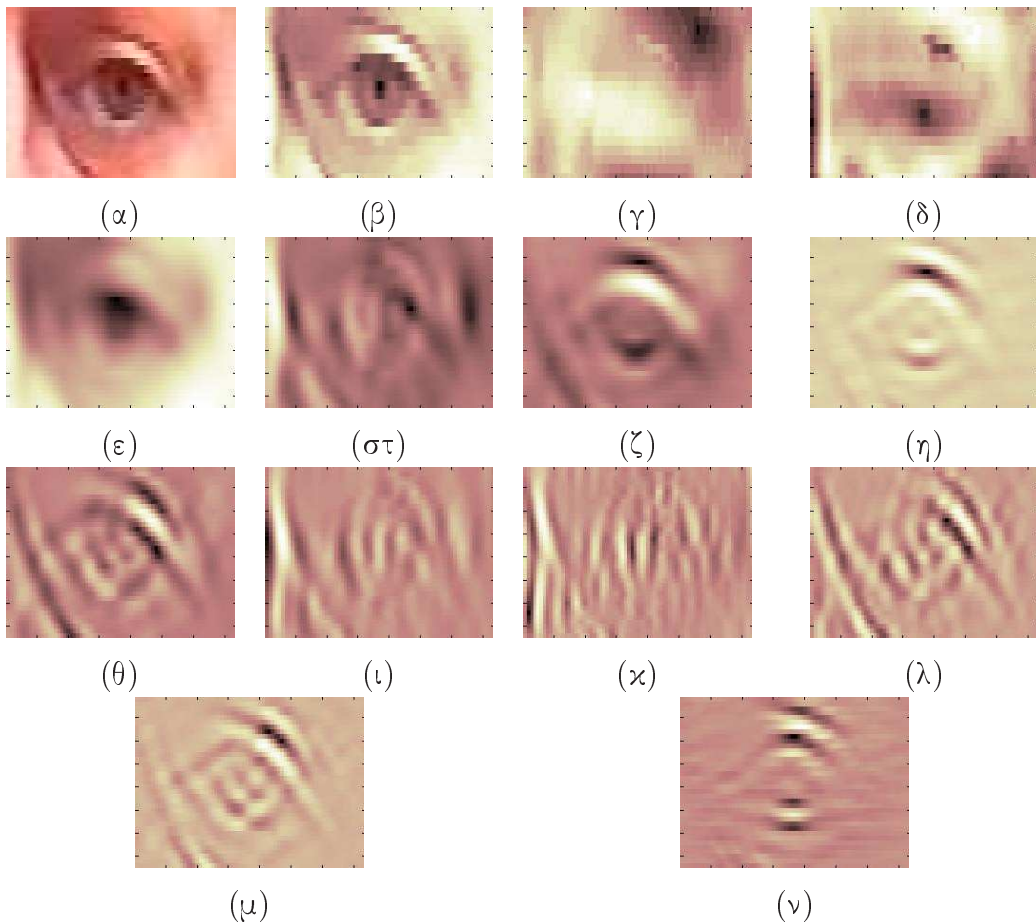
Προκειμένου να αυξηθεί η ταχύτητα και να μειωθούν οι απαιτήσεις μνήμης, οι οφθαλμοί δεν ανιχνεύονται σε κάθε καρέ με τη βοήθεια του νευρωνικού δικτύου. Αντ'αυτού, μετά την ανίχνευση των ματιών στο πρώτο καρέ, κατασκευάζονται δυο πρότυπα (templates) από το κανάλι της φωτεινότητας, ένα για



(α) είσοδος δικτύου

(β) έξοδος δικτύου, οι σκουρότερες περιοχές υποδηλώνουν υψηλότερη έξοδο

**Σχήμα 1.3:** Έξοδος του νευρωνικού δικτύου σε αριστερό οφθαλμό



(α) Τυπική έγχρωμη εικόνα διαστάσεων 80x44, (β)–(ν) Είσοδος νευρωνικού δικτύου, (β)(γ)(δ) Συνιστώσες Y/Cr/Cb, (ε)–(ν) οι 10 σημαντικότεροι συντελεστές DCT υπολογισμένοι σε 8x8 blocks που έχουν συγκολληθεί μεταξύ τους για λόγους απεικόνισης.

**Σχήμα 1.4:** Απεικόνιση Εισόδου Νευρωνικού Δικτύου

κάθε οφθαλμό, γύρω από το κέντρο του κάθε οφθαλμού και σε μικρή περιοχή. Το μέγεθος κάθε ίχνους είναι ίσο με την διακορική απόσταση (την απόσταση μεταξύ των οφθαλμών),  $D_{bp}$  - bupupil breadth. Στα ακόλουθα καρέ οι οφθαλμοί εντοπίζονται με απλό ταίριασμα προτύπων βρίσκοντας την περιοχή όπου ελαχιστοποιείται το άθροισμα των απολύτων διαφορών.

Όταν έχει υπολογιστεί η γωνία του κεφαλιού, αυτό περιστρέφεται σε όρθια θέση και ορίζονται εκ νέου περιοχές ανίχνευσης χαρακτηριστικών με τη χρήση των ίδιων κανόνων του πίνακα 1.1 ώστε να εξασφαλιστεί ότι τα χαρακτηριστικά του προσώπου περικλείονται στις αντίστοιχες περιοχές. Αυτές οι περιοχές ανίχνευσης, χρησιμοποιούνται για την εξαγωγή των χαρακτηριστικών όπως περιγράφεται παρακάτω.

## 1.4 Ανίχνευση των χαρακτηριστικών του προσώπου, και εξαγωγή των ορίων τους

Προκειμένου να υπολογιστούν τα FAPs απαιτείται η ακριβής εξαγωγή των ορίων των χαρακτηριστικών του προσώπου, δηλαδή ματιών φρυδιών και στόματος. Οι συνήθεις μέθοδοι εύρεσης των ορίων του οφθαλμού βασίζονται στην ανίχνευση των ιδιαίτερων χρωματικών χαρακτηριστικών της περιοχής [27], είτε στη χρήση προβολών φωτεινότητας [63], ή της αντίστροφης πιθανότητας χρώματος δέρματος, ή στην προσαρμογή οφθαλμικών μοντέλων [40],[16]. Η εξαγωγή των ορίων του στόματος στην περίπτωση του κλειστού στόματος είναι σχετικά απλή [39]. Για την περίπτωση του ανοικτού στόματος έχουν προταθεί αρκετές τεχνικές οι οποίες βασίζονται είτε στη φωτεινότητα [16][40] είτε στο χρώμα [17][41][42][27]. Η πληροφορία του χρώματος όμως είναι αρκετά ευαίσθητη στις συνθήκες φωτισμού και στα χαρακτηριστικά της κάμερας που χρησιμοποιείται. Μια συνήθης μεθοδολογία είναι το ταίριασμα κάποιου μοντέλου, συνήθως έλλειψης ή κύκλου, χρησιμοποιώντας παραλλαγές του μετασχηματισμού Hough ή ανίχνευση γωνιών [43]. Αυτή η τεχνική παρόλο που παρέχει ακριβή αποτελέσματα σε εικόνες υψηλής ανάλυσης, δεν δύναται να χρησιμοποιηθεί σε βίντεο χαμηλής ανάλυσης όπου απουσιάζει η υψίσυχη πληροφορία· η τελευταία αν και είναι απαραίτητη για την ανίχνευση των γωνιών και την παρακολούθηση (trackability) [44] των ορίων των χαρακτηριστικών, συνήθως χάνεται εξαιτίας της διακωδικοποίησης (transcoding) της αναλογικής πληροφορίας η της ψηφιακής συμπίεσης χαμηλής ποιότητας. Στην παρούσα προσέγγιση, η εξαγωγή της μάσκας των φρυδιών και της μύτης γίνεται σε ένα στάδιο, ενώ για τους οφθαλμούς και το στόμα, τα οποία ανιχνεύονται πιο δύσκολα, κατασκευάζονται πολλαπλές μάσκες, λαμβά-

## Κεφάλαιο 1. Εξαγωγή των χαρακτηριστικών του προσώπου

νοντας υπόψη τις ιδιαίτερες ιδιότητες της κάθε περιοχής. Οι πολλαπλές αυτές μάσκες εν συνεχεία συνδυάζονται για την κατασκευή των τελικών μασκών στόματος και ματιών όπως φαίνεται στην εικόνα 1.1. Ο πίνακας 1.2 συνοψίζει τους συμβολισμούς των εξαγομένων ενδιάμεσων μασκών για το στόμα και τους οφθαλμούς αντίστοιχα, προσφέροντας και μια σύντομη ποιοτική περιγραφή της κάθε μίας.

Περιγράφεται στην ενότητα	Ανιχνεύει	Εξαρτάται από	Απολήγει σε
1.4.1.1	Ίριδα, περιβάλλουσες σκούρες περιοχές, βλεφαρίδες.	$\mathbf{L}^e, \mathbf{M}_{nn}^e$	$\mathbf{M}_1^e$
1.4.1.2	Οριζόντιες ακμές βλεφάρων, άνωθεν και κατώθι του οφθαλμικού κέντρου	$\mathbf{L}^e$ , κέντρο του οφθαλμού	$\mathbf{M}_2^e$
1.4.1.3	Περιοχές με πυκνή υφή πλησίον της ίριδας	$\mathbf{L}^e$	$\mathbf{M}_3^e$
1.4.1.4	Περιοχή παρόμοιας φωτεινότητας με αυτή του οφθαλμού (μάσκα $\mathbf{M}_{nn}^e$ )	$\mathbf{L}^e, \mathbf{M}_{nn}^e$	$\mathbf{M}_4^e$
1.4.4.1	Χείλη και στόμα με παρόμοιες ιδιότητες με τα εκπαιδευθέντα από το «ουδέτερο καρέ»	Περιοχή ανίχνευσης στόματος (έγχρ.)	$\mathbf{M}_1^m$
1.4.4.2	Οριζόντιες ακμές χειλιών	$\mathbf{L}_m$	$\mathbf{M}_2^m$
1.4.4.3	Οριζόντια έκταση στόματος προκύπτουσα από την ανίχνευση των γωνιών των χειλιών, άνοιγμα στόματος προκύπτει από τις ακμές των χειλιών	$\mathbf{L}_m$	$\mathbf{M}_3^m$

**Πίνακας 1.2:** Σύνοψη των οφθαλμικών και στοματικών μασκών

Εφεξής θα χρησιμοποιούνται οι εξής συμβολισμοί: με  $\mathbf{M}_i^x$  θα συμβολίζεται η δυαδική μάσκα  $i$  του χαρακτηριστικού  $x$  του προσώπου όπου το  $x$  είναι  $e$  για τους οφθαλμούς,  $m$  για το στόμα,  $n$  για τη μύτη και  $b$  για τα φρύδια. Ός  $\mathbf{L}_x$  θα συμβολίζονται οι αντίστοιχες μάσκες φωτεινότητας. Επιπλέον, η αξιολόγηση του μεγέθους και της θέσης των ευρεθέντων χαρακτηριστικών εξαρτάται από μια σειρά ανθρωπομετρικών κριτηρίων τα οποία είναι ορισμένα ως ποσοστό μεγέθους επί κάποιας εύκολα μετρήσιμης διάστασης του προσώπου· αυτά

Κεφάλαιο 1. Εξαγωγή των χαρακτηριστικών του προσώπου

Κατώφλι	Τιμή	Εξαρτάται από
$t_b^e$	$\frac{1}{3} (2 \langle f_c(\mathbf{L}^e, \mathbf{M}_{nn}^e) \rangle + \min(\mathbf{L}^e))$	L
$t_E^b$	$\langle \mathbf{M}_{E_1}^b \rangle + \sqrt{\langle (\mathbf{M}_{E_1}^b)^2 \rangle - \langle \mathbf{M}_{E_1}^b \rangle^2}$	L
$t_{c_1}^m$	$\langle \mathbf{L}_m^{asfr} \rangle - \sqrt{\langle [\mathbf{L}_m^{asfr}]^2 \rangle - \langle \mathbf{L}_m^{asfr} \rangle^2}$	L
$t_1^m$	$\frac{1}{3} (2\overline{\mathbf{L}_m} + \min(\mathbf{L}_m))$	L
$t_1^n$	$\frac{1}{3} (\overline{\mathbf{L}_n} + 2 \min(\mathbf{L}_n))$	L
$t_2^m$	90%	NN output
$t_d^e$	90%	L

**Πίνακας 1.3:** Προσαρμοζόμενα Κατώφλια

Κατώφλι	Τιμή
$t_\sigma$	$10^{-3}$
$t_{vd}$	0.8
$L_x$ : Εικόνα φωτεινότητας του χαρ/κού $x$	

**Πίνακας 1.4:** Κατώφλια

Συμβολισμός	Τιμή	Αναφέρεται σε
$r_{asf}^m$	1%	$d_m^w$
$r_{b1}^m$	10%	$I_w$
$r_{c2}^m$	25%	$D_{bp}$
$r_{b2}^m$	50%	$I_w$
$r_c^e$	5%	$W_f$
$r_2^b$	5%	$D_{bp}$
$r_1^b$	30%	$D_{bp}$
$r_2^n$	10%	$D_{bp}$
$r_4^n$	15%	$D_{bp}$
$r_3^n$	20%	$D_{bp}$

**Πίνακας 1.5:** Συσχετιστικοί ανθρωπομετρικοί περιορισμοί



περιλαμβάνουν τα  $r_{asf}^m, r_c^e, r_1^b, r_2^b, r_{b1}^m, r_{b2}^m, r_2^n, r_3^n, r_4^n$  τα οποία περιγράφονται στον πίνακα 1.5, ενώ τα στατικά και προσαρμοζόμενα κατώφλια, που ορίζονται στο κείμενο, συνοψίζονται στους πίνακες 1.4 και 1.3.

### 1.4.1 Ανίχνευση ορίων οφθαλμών

#### 1.4.1.1 Συνδυασμός μάσκας φωτεινότητας και χρώματος

Οι οφθαλμικές μάσκες που προκύπτουν από το νευρωνικό δίκτυο όπως περιγράφηκε στην ενότητα 1.3, παριστάνονται ως  $\mathbf{M}_{nn}^e$ ) και μπορούν να βελτιωθούν λαμβάνοντας υπόψη το γεγονός ότι τα βλέφαρα είναι συνήθως σκουρότερα του δέρματος εξαιτίας των βλεφαρίδων και ότι συνήθως συνορεύουν με την ίριδα.

Αρχικά, η πληροφορία της φωτεινότητας στο εσωτερικό της περιοχής που προκύπτει από τη διαστολή της αρχικής μάσκας  $\mathbf{M}_{nn}^e$  χρησιμοποιείται για την εύρεση του κατωφλίου φωτεινότητας  $t_b^e$  (εξ. 1.2).

$$t_b^e = \frac{1}{3} (2 \langle f_c(\mathbf{L}^e, \mathbf{M}_{nn}^e) \rangle + \min(\mathbf{L}^e)) \quad (1.2)$$

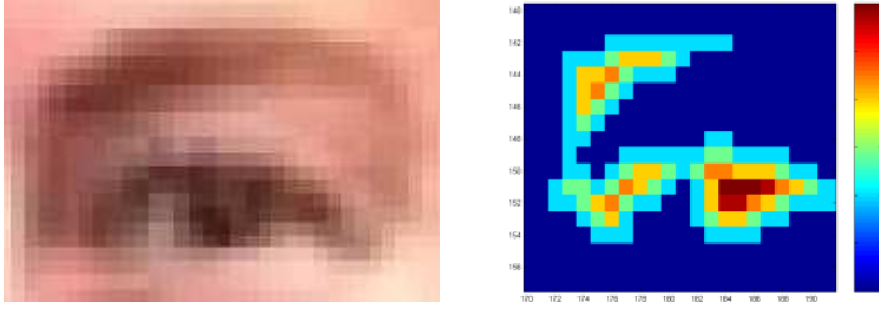
$$f_c(\mathbf{A}, \mathbf{B}) = \{c_{ij}\}, \quad c_{ij} = \begin{cases} a_{ij}, & b_{ij} \neq 0 \\ 0, & b_{ij} = 0 \end{cases} \quad (1.3)$$

όπου ως  $\mathbf{L}^e$  συμβολίζεται η συνιστώσα της φωτεινότητας της οφθαλμικής περιοχής αναζήτησης χαρακτηριστικών, το  $\langle \cdot \rangle$  δηλώνει τη μέση τιμή μιας περιοχής και το  $\min(\mathbf{X})$  την ελάχιστη τιμή σε μια περιοχή  $\mathbf{X}$ .

Εφαρμόζοντας το κατώφλι  $t_b^e$  στο  $\mathbf{L}^e$ , προκύπτει η μάσκα  $\mathbf{M}_{npp}^e$  η οποία περιέχει τα σκούρα αντικείμενα γύρω από το κέντρο του οφθαλμού και συγκεκριμένα τις βλεφαρίδες και την ίριδα. Από τα συνεκτικά αντικείμενα της  $\mathbf{M}_{npp}^e$  μπορεί να επιλεγεί εκείνο που περιέχει την ίριδα χρησιμοποιώντας ένα κριτήριο πάχους. Εφαρμόζοντας τον μετασχηματισμό απόστασης χρησιμοποιώντας ευκλείδεια μετρική, επιλέγεται το συνεκτικό αντικείμενο μέσα στο οποίο ο μετασχηματισμός αποκτά τη μέγιστη τιμή του  $DT_{max}$  και προκύπτει η μάσκα  $\mathbf{M}_1^e$  που απεικονίζεται στη εικόνα 1.5. Η τελευταία περιέχει την ίριδα και τις παρακείμενες βλεφαρίδες. Το σημείο όπου ο μετασχηματισμός απόστασης ισούται με  $DT_{max}$  αντιστοιχεί στο κέντρο της ίριδας.

#### 1.4.1.2 Μάσκα βασισμένη σε ακμές

Η δεύτερη μάσκα βασίζεται στην ανίχνευση των βλεφάρων. Τα τελευταία βρίσκονται εκατέρωθεν του οφθαλμικού κέντρου που έχει ήδη βρεθεί από το νευρωνικό δίκτυο. Τα βλέφαρα εντοπίζονται εύκολα χάρη στην οριζόντια κατεύθυνσή τους μέσω ανίχνευσης ακμών.



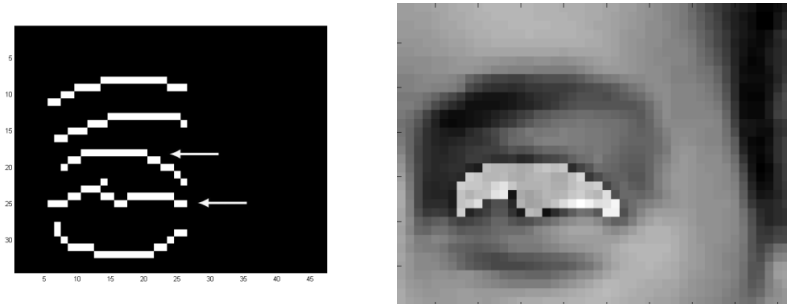
(α) αριστερός οφθαλμός (τμήμα). (β) η μάσκα  $M_1^e$  αριστερού οφθαλμού και τιμές μετ/μού απόστασης.

Σχήμα 1.5: 1<sup>η</sup> οφθαλμική μάσκα

Για το σκοπό αυτό έγινε χρήση του ανιχνευτή ακμών Canny [45] εξαιτίας των καλών ιδιοτήτων εντοπισμού που διαθέτει και της ικανότητάς του να ελαχιστοποιεί τις πολλαπλές αποκρίσεις σε μία μόνο ακμή. Ο τελεστής του Canny ακολουθεί τα τοπικά μέγιστα, συνήθως παράγοντας κλειστές καμπύλες· αυτές διασπώνται εύκολα χρησιμοποιώντας μορφολογικό opening με ακτινικό δομικό στοιχείο μήκους 3, οπότε προκύπτει η μάσκα  $M_{b_1}^e$ . Η αυτή μάσκα εμπλουτίζεται στη συνέχεια με το δεύτερο τροποποιημένο ανιχνευτή ακμών Canny. Ο τελευταίος εξετάζει την κλίση μόνο προς την κάθετη διεύθυνση ακολουθώντας τις μισές από τις δυνατές κατευθύνσεις κίνησης του τελεστή. Εφόσον η κατεύθυνση των ακμών είναι κάθετη στην κλίση, ο τελεστής αυτός ανιχνεύει κυρίως οριζόντιες ακμές οι οποίες αποθηκεύονται στην μάσκα  $M_{b_2}^e$ . Οι δύο δυαδικές μάσκες  $M_{b_1}^e$  και  $M_{b_2}^e$  εν συνεχεία συνδυάζονται:

$$M_{b_3}^e = M_{b_1}^e + M_{b_2}^e \quad (1.4)$$

παράγοντας τη μάσκα  $M_{b_3}^e$  η οποία απεικονίζεται στην εικόνα 1.6(α). Οι ακμές



(α) αποτέλεσμα τροποποιημένου ανιχνευτή ακμών canny

(β) η μάσκα  $M_2^e$

Σχήμα 1.6: 2<sup>η</sup> οφθαλμική μάσκα

εκατέρωθεν του οφθαλμικού κέντρου στη μάσκα  $M_{b_3}^e$  που σημειώνονται με βέλη

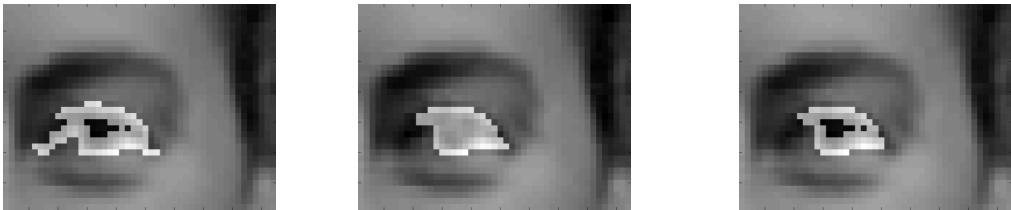
στην εικόνα 1.6(α) επιλέγονται ως βλέφαρα και ο χώρος ανάμεσά τους ως η μάσκα  $\mathbf{M}_2^e$  που απεικονίζεται στην εικόνα 1.6(β).

#### 1.4.1.3 Μάσκα βασιζόμενη στην τυπική απόκλιση

Δεδομένων των απαιτήσεων ταχύτητας αλλά και του γεγονότος ότι ο οφθαλμός καταλαμβάνει πολύ μικρή έκταση (σε pixels) μη επαναλαμβανόμενης υφής, αναπτύχθηκε μια απλή αλλά αποτελεσματική μέθοδος για την εκτίμηση μιας τρίτης οφθαλμικής μάσκας. Η μάσκα αυτή δημιουργείται με τεχνική επέκτασης περιοχών (region growing) και δίνει συνήθως πολύ καλά αποτελέσματα κατάτμησης που αντιστοιχούν επαρκώς στις υπάρχουσες ακμές. Η κατασκευή αυτής της μάσκας βασίζεται σε δύο στοιχεία:

- η υφή του προσώπου είναι πιο σύνθετη εντός της οφθαλμικής περιοχής και ιδιαίτερα στα όρια βλεφάρων, σκληρού χιτώνα και ίριδας από ότι στις προσκείμενες περιοχές.
- οι περιοχές αυτές χαρακτηρίζονται από αντικείμενα χαμηλότερης φωτεινότητας (βλέφαρα, ίριδα) σε σχέση με το περιβάλλον δέρμα.

Ένας άλλος τρόπος να περιγραφούν οι περιοχές αυτές τόσο στα όρια όσο και εντός του οφθαλμού είναι λοιπόν: περιοχές χαμηλής φωτεινότητας, γύρω από τις οποίες η φωτεινότητα μεταβάλλεται γρήγορα. Μια εκτιμήτρια της πολυπλοκότητας της περιοχής γύρω από ένα σημείο είναι η τυπική απόκλιση. Η μάσκα κατασκευάζεται συγκρίνοντας την τυπική απόκλιση σε μια μικρή περιοχή γύρω από ένα σημείο της περιοχής του οφθαλμού, με τη φωτεινότητα του σημείου αυτού. Με την κατάλληλη προσαρμογή των τιμών της φωτεινότητας που περιγράφεται παρακάτω, η μάσκα αυτή ανιχνεύει περιοχές χαμηλής μη ομοιόμορφης φωτεινότητας.



(α) η μάσκα  $\mathbf{M}_{s_{n,d}}^e$ ,  $n = 3$  (β) η μάσκα  $\mathbf{M}_{s_{n,d}}^e$ ,  $n = 6$  (γ) ο συνδυασμός τους,  $\mathbf{M}_3^e$

Σχήμα 1.7: 3<sup>η</sup> οφθαλμική μάσκα

Αρχικά υπολογίζεται η τυπική απόκλιση της συνιστώσας φωτεινότητας  $\mathbf{L}^e$  σε  $n \times n$  κυλιόμενα blocks οπότε προκύπτει η  $\mathbf{I}_{std_n}^e$  (σχέση 1.5). Η τελευταία κατωφλιώνεται διαδοχικά με όριο  $\frac{1}{d}\mathbf{L}^e$  οπότε προκύπτει η  $\mathbf{M}_{s_{n,d}}^e$ . Το  $d$  είναι

έναν διαδοχικά αυξανόμενος διαιρέτης. Όσο αυξάνεται, οι περιοχές στην μάσκα  $\mathbf{M}_{s_{n,d}}^e$  διευρύνονται τείνοντας να ενωθούν.

Η τιμή της τυπικής απόκλισης εξαρτάται και από το μέγεθος  $n$  του επιλεγόμενου παραθύρου. Το τελευταίο θα θέλαμε να είναι μικρότερο από το μέγεθος της ίριδας αλλά αρκετά μεγάλο, ώστε στις περιοχές των βλεφάρων να περιλαμβάνει και τις περιβάλλουσες περιοχές δέρματος. Οι δοκιμές που έγιναν σε πλήθος εικόνων διαφορετικής ανάλυσης αλλά και παρεχόμενης λεπτομέρειας έδειξαν ότι δεν είναι εφικτή η εκ των προτέρων εκτίμηση του κατάλληλου παραθύρου επομένως η διαδικασία εκτελείται με δύο διαφορετικά μεγέθη παραθύρου και τα ενδιαμέσα αποτελέσματα συνδυάζονται.

Η διαδικασία αυτή εκτελείται αρχικά για  $n = 3$ . Το κέντρο του οφθαλμού επιλέγεται στην πρώτη επανάληψη ως το κέντρο του μεγαλύτερου αντικειμένου. Στην επανάληψη  $i$  το εκτιμώμενο κέντρο συμβολίζεται ως  $\mathbf{c}_i$ . Στην περίπτωση που η εκτίμηση της οφθαλμικής επιφάνειας ξεπεράσει τα πραγματικά της όρια και ενωθεί με κάποιο γειτονικό αντικείμενο, τότε το κέντρο της θα μετατοπιστεί προς το νέο προσαρτημένο αντικείμενο. Επομένως μπορούμε να ορίσουμε ως κριτήριο τερματισμού της διαδικασίας διεύρυνσης την απόσταση του νέου κέντρου από το αρχικό, οπότε η διαδικασία συνεχίζεται όσο  $\|\mathbf{c}_1 - \mathbf{c}_i\| \leq r_c^e W_f$ , όπου το κριτήριο  $r_c^e W_f$  ορίζεται ως ποσοστό επί του ήδη υπολογισμένου εύρους του προσώπου και αναφέρεται στον πίνακα 1.5.

Έτσι προκύπτει η δυαδική μάσκα  $\mathbf{M}_{s_{3,f}}^e$  που απεικονίζεται στην εικόνα 1.7(α). Όπως φαίνεται στην εικόνα, η μάσκα περιλαμβάνει τα βλέφαρα και τις περιοχές της ίριδας στο όριο με το σκληρό χιτώνα, αλλά όχι την περιοχή του εσωτερικού της ίριδας όπου η φωτεινότητα δε μεταβάλλεται.

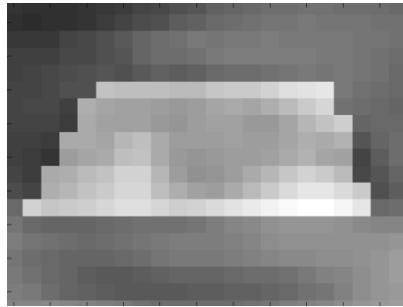
Η ίδια διαδικασία επαναλαμβάνεται για  $n = 6$  οπότε προκύπτει η μάσκα  $\mathbf{M}_{s_{6,f}}^e$  που απεικονίζεται στην εικόνα 1.7(β). Από το συνδυασμό των μασκών που περιγράφεται στην (1.5) προκύπτει η μάσκα  $\mathbf{M}_3^e$ , όπως στην περίπτωση της εικόνας 1.7(γ), εξασφαλίζοντας καλύτερο αποτέλεσμα σε περίπτωση ανωμαλιών όπως είναι ο θόρυβος που μπορεί να επιφέρει απότομες διακυμάνσεις στη φωτεινότητα.

$$\begin{aligned}
 \mathbf{L}^e &= \{l_{i,j}\} \\
 \mathbf{I}_{std_n}^e &= \{i_{n,i,j}\}, i_{n,i,j} = \sqrt{\langle l_{i,j}^2 \rangle - \langle l_{i,j} \rangle^2} \\
 m_{n,d,i,j} &= \begin{cases} 1, & \frac{l_{i,j}}{d} > i_{n,i,j} \\ 0, & \frac{l_{i,j}}{d} < i_{n,i,j} \end{cases}, n = 3, 6 \\
 \mathbf{M}_{s_{n,d}}^e &= \{m_{n,d,i,j}\}
 \end{aligned} \tag{1.5}$$

όπου το  $d \in (0, \max(\mathbf{L}^e)]$  και με  $\langle \cdot, \cdot \rangle_{i,j}$  συμβολίζεται ο μέσος σε περιοχή  $n \times n$  γύρω από το σημείο  $(i, j)$

#### 1.4.1.4 Μάσκα βασιζόμενη στη φωτεινότητα

Η τέταρτη και τελευταία ενδιάμεση οφθαλμική μάσκα δημιουργείται με στόχο την ανίχνευση του ορίου μεταξύ οφθαλμών και βλεφάρων. Η μάσκα αυτή βασίζεται στον εύκολο υπολογισμό της πιθανότητας η φωτεινότητα εντός της οφθαλμικής περιοχής  $\mathbf{L}^e$  να μοιάζει με αυτή στην περιοχή που ορίζεται από τη μάσκα  $\mathbf{M}_{nn}^e$  η οποία εξάγεται από το αρχικό νευρωνικό δίκτυο, θεωρώντας κανονική κατανομή. Από τον προκύπτοντα πίνακα πιθανοτήτων λαμβάνουμε τις πιθανότερες τιμές με χρήση του διαστήματος εμπιστοσύνης  $t_a^e$ . Το αποτέλεσμα είναι συνήθως μια μάσκα που ορίζει καθαρά τα όρια του οφθαλμού κατά την κάθετη διεύθυνση. Λόγω της απουσίας βλεφάρων εκατέρωθεν του οφθαλμού στην οριζόντια διεύθυνση αλλά και της παρουσίας σκιών από τη μύτη και τα φρύδια οι τιμές της φωτεινότητας εκτός του οφθαλμού μπορούν να ομοιάζουν με αυτές εντός, με αποτέλεσμα την σύνδεση της μάσκας με τμήματα εκτός του οφθαλμού στην οριζόντια διεύθυνση. Για να αποφευχθεί η χρήση στενότερου ορίου εμπιστοσύνης, προτιμάται η διατήρηση του υπάρχοντος, με ταυτόχρονη εκμετάλλευση του σχεδόν ελλειπτικού σχήματος του οφθαλμού, περιορίζοντας την έκταση της προκύπτουσας μάσκας στα λεπτότερα σημεία της εκατέρωθεν του κέντρου του οφθαλμού. Το κυρτό περίβλημα της προκύπτουσας μάσκας που απεικονίζεται στο παράδειγμα της εικόνας 1.8 συμβολίζεται ως  $\mathbf{M}_4^e$ .



η μάσκα  $\mathbf{M}_4^e$

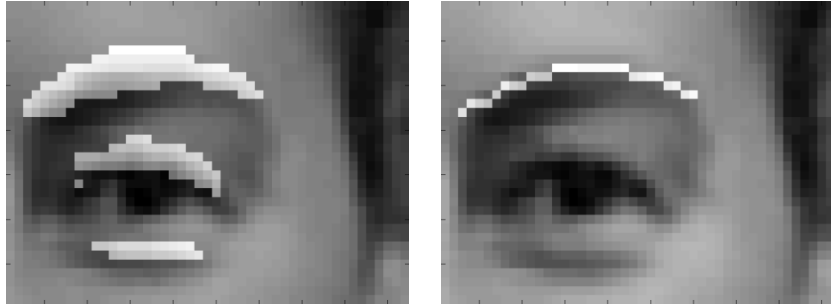
Σχήμα 1.8: 4<sup>η</sup> οφθαλμική μάσκα

#### 1.4.2 Ανίχνευση ορίων φρυδιών

Η ανίχνευση των φρυδιών δεν είναι ιδιαίτερα δύσκολη. Βασίζεται στο γεγονός ότι έχουν απλό κατευθυντικό σχήμα και βρίσκονται στο μέτωπο που χάρη

στην προεκβολή του έχει ομοιόμορφο φωτισμό. Καθένα από τα φρύδια ανιχνεύονται στις περιοχές ανίχνευσης χαρακτηριστικών οφθαλμών-φρυδιών όπως στο παράδειγμα της εικόνας 1.2.

Η συνιστώσα φωτεινότητας της περιοχής ανίχνευσης χαρακτηριστικών οφθαλμών / φρυδιών χρησιμοποιείται για την κατασκευή του χάρτη ακμών  $M_E^b$ , με χρήση της μορφολογικής παραγώγου, αφαιρώντας δηλαδή την διαβρωμένη από τη διεσταλμένη εικόνα (εξ. 1.6). Χρησιμοποιείται ακτινικό δομικό στοιχείο  $s$  μήκους  $r_2^b$  και στη συνέχεια κατωφλίωση όπως φαίνεται στο παράδειγμα της εικόνας 1.9(α) και την εξίσωση (1.6).



(α) υποψήφια φρύδια

(β) η τελική μάσκα  $M^b$

Σχήμα 1.9: Μάσκα φρυδιών

$$\begin{aligned}
 M_{E_1}^b &= \delta_s(L_e) - \varepsilon_s(L_e) \\
 t_E^b &= \left( \langle M_{E_1}^b \rangle + \sqrt{\langle (M_{E_1}^b)^2 \rangle - \langle M_{E_1}^b \rangle^2} \right) \\
 M_E^b &= M_{E_1}^b > t_E^b
 \end{aligned} \tag{1.6}$$

όπου τα  $\delta_s, \varepsilon_s$  δηλώνουν τους μορφολογικούς τελεστές της διαστολής (dilation) και της διάβρωσης (erosion) με δομικό στοιχείο  $s$  και ο τελεστής  $>$  δηλώνει κατωφλίωση για την κατασκευή της μάσκας  $M_E^b$ . Η συγκεκριμένη διαδικασία ανίχνευσης ακμών επελέγη ως κατάλληλη για τα φρύδια διότι δύναται να είναι κατευθυντική, διατηρεί το αρχικό μέγεθος του χαρακτηριστικού και μπορεί να συνδυαστεί με κατώφλι ώστε να αγνοηθούν μικρότερες διαταράξεις φωτεινότητας στο δέρμα όπως οι ρυτίδες.

Το σύνολο των συνεκτικών αντικειμένων στο χάρτη ακμών ελέγχεται έναντι μιας λίστας κριτηρίων τα οποία κατασκευάστηκαν από τη στατιστική ανάλυση των μηκών και της θέσης των φρυδιών σε 20 άτομα της βάσης δεδομένων ERMIS [29].

Αν θεωρήσουμε τα διανύσματα  $x$  και  $y$ , που περιέχουν τις συντεταγμένες των σημείων ενός αντικειμένου, τότε μπορούμε να υπολογίσουμε τους ημιάξονες της

Κεφάλαιο 1. Εξαγωγή των χαρακτηριστικών του προσώπου

έλλειψης που περικλείει το αντικείμενο μέσω ανάλυσης βασικών συνιστωσών (principal component analysis). Τα δύο ιδιοδιανύσματα (eigenvectors) του πίνακα συμμεταβλητότητας εκφράζουν τους κύριους άξονες (principal axes) του σχήματος, και συγκεκριμένα το ιδιοδιάνυσμα που αντιστοιχεί στη μέγιστη ιδιοτιμή (eigenvalue) εκφράζει τον πρωτεύοντα ή μείζονα άξονα (major axis), ενώ εκείνο που αντιστοιχεί στην ελάχιστη ιδιοτιμή εκφράζει τον δευτερεύοντα ή ελάσσονα άξονα (minor axis). Ο πίνακας συμμεταβλητότητας  $C$  των σημείων αυτών, δίνεται από τη σχέση:

$$C = \begin{pmatrix} C_{xx} & C_{xy} \\ C_{yx} & C_{yy} \end{pmatrix} \quad (1.7)$$

όπου τα στοιχεία του πίνακα υπολογίζονται μέσω των κεντρικών ροπών:

$$C_{xx} = \mu_{20} = \frac{m_{20}}{m_{00}} - \bar{x}^2 \quad (1.8)$$

$$C_{xy} = \mu_{11} = \frac{m_{20}}{m_{00}} - \bar{x}\bar{y} \quad (1.9)$$

$$C_{yy} = \mu_{02} = \frac{m_{02}}{m_{00}} - \bar{y}^2 \quad (1.10)$$

$$(1.11)$$

και το κέντρο βάρους υπολογίζεται ως:

$$\bar{x} = \frac{m_{10}}{m_{00}}, \bar{y} = \frac{m_{01}}{m_{00}} \quad (1.12)$$

τα μήκη των αξόνων υπολογίζονται με την εύρεση των ιδιοτιμών:

$$\begin{vmatrix} C_{xx} - \lambda & C_{xy} \\ C_{yx} & C_{yy} - \lambda \end{vmatrix} = 0 \quad (1.13)$$

$$\lambda^2 - \underbrace{(C_{xx} + C_{yy})}_{\text{trace}C} \lambda + \underbrace{(C_{xx}C_{yy} - C_{xy}^2)}_{\det C} = 0 \quad (1.14)$$

$$\lambda_{1,2} = \frac{\text{trace}(C)}{2} \pm \sqrt{\frac{\text{trace}(C)^2}{4} - \det(C)} \quad (1.15)$$

$$(1.16)$$

και η γωνία στροφής υπολογίζεται ως:

$$\Theta = \frac{1}{2} \arctan \left( \frac{C_{xy}}{C_{xx} - C_{yy}} \right) \quad (1.17)$$

Τα αντικείμενα των οποίων ο κύριος άξονας σχηματίζει γωνία μεγαλύτερη από  $30^\circ$  με το οριζόντιο επίπεδο αφαιρούνται από το σύνολο. Επίσης αφαιρούνται όσα έχουν κύριο άξονα μικρότερου μήκους από  $r_1^b$ , καθώς και όσα αποτελούνται

από στοιχεία με απόσταση από το κέντρο του οφθαλμού μεγαλύτερη από  $\frac{1}{2}r_1^b$ , όπου το κριτήριο  $r_1^b$  ορίζεται στον πίνακα 1.5 και εξαρτάται από τη διακορική απόσταση. Τα εναπομείναντα αντικείμενα θα περιέχουν το φρύδι και αντικείμενα στην περιοχή του οφθαλμού, (εικόνα 1.9α). Συνεπώς, επιλέγεται αυτό που περιέχει το στοιχείο με την ψηλότερη θέση κατά την κάθετη διεύθυνση, οπότε προκύπτει η μάσκα  $\mathbf{M}_{E_2}^b$ . Εφόσον η επιφάνεια των φρυδιών δε χρησιμοποιείται κατά τον υπολογισμό των FAPs, το αποτέλεσμα μπορεί να απλοποιηθεί χρησιμοποιώντας την εξίσωση (1.18) και παράγοντας τη μάσκα  $\mathbf{M}^b$  όπως φαίνεται στο παράδειγμα της εικόνας 1.9(β).

$$\begin{aligned} \mathbf{M}^b &= \{m_{i,j}\} \\ \mathbf{M}_E^b &= \{m_{i,j}^E\} \\ m_{i,j}^E &= \begin{cases} 1, & (m_{i,j} = 1) \wedge (m_{i,j'} \neq 1), j' < j \\ 0, & \text{otherwise} \end{cases} \end{aligned} \quad (1.18)$$

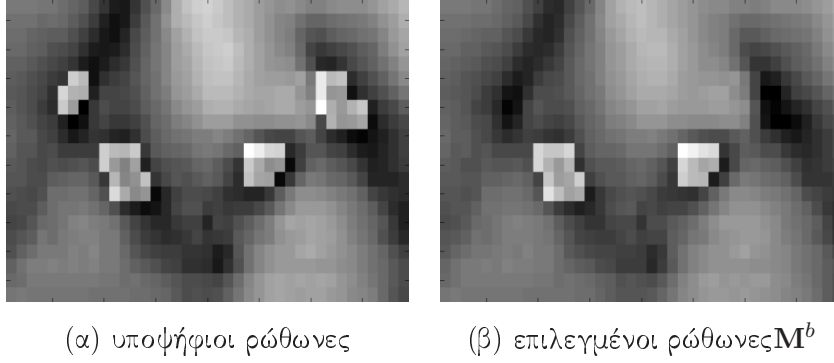
### 1.4.3 Ρινικός εντοπισμός

Η μύτη δε χρησιμοποιείται για την εκτίμηση των εκφράσεων από μόνη της, αλλά παρέχει ένα σταθερό και εύκολα εντοπίσιμο σημείο το οποίο μας διευκολύνει στη μέτρηση αποστάσεων για τον υπολογισμό των FAP (εικόνα 2.5). Με βάση το σκεπτικό αυτό επομένως δεν απαιτείται ο ακριβής υπολογισμός των ορίων της. Ο εντοπισμός της μύτης απαντάται συχνά σε εφαρμογές παρακολούθησης του προσώπου και συνήθως βασίζεται στον εντοπισμό των ρωθώνων. Οι τελευταίοι εντοπίζονται σχετικά εύκολα λόγω της χαμηλής φωτεινότητάς τους [46], με την προϋπόθεση όμως ότι είναι ορατοί από την κάμερα, επομένως η κάμερα πρέπει να βρίσκεται στο ύψος του προσώπου και χαμηλότερα. Για το ρινικό εντοπισμό χρησιμοποιείται η περιοχή άνωθεν της περιοχής ανίχνευσης χαρακτηριστικών του στόματος. Στην αντίστοιχη εικόνα φωτεινότητας εφαρμόζεται το κατώφλι  $t_1^n$ :

$$t_1^n = \frac{1}{3} (\langle \mathbf{L}_n \rangle + 2 \min(\mathbf{L}_n)), \mathbf{L}_n : \text{Εικόνα φωτεινότητας} \quad (1.19)$$

Σε περίπτωση ανομοιόμορφου φωτισμού μπορεί να σχηματιστούν μακριές σκιές εκατέρωθεν της μύτης. Για το λόγο αυτό με τη χρήση ανθρωπομετρικών στοιχείων [47] μειώνονται οι πιθανές επιλογές. Αρχικά αποκλείονται αντικείμενα κοντύτερα από  $r_2^n$  και μακρύτερα από  $r_3^n D_{bp}$ . Αυτός έχει αποδειχθεί ένας αποτελεσματικός τρόπος να μειωθούν τα υποψήφια αντικείμενα χωρίς να δημιουργούνται λανθασμένα αρνητικά αποτελέσματα. Κατ'αυτόν τον τρόπο κατασκευάζεται η μάσκα  $\mathbf{M}_1^n$  που απεικονίζεται στο παράδειγμα της εικόνας 1.10.





Σχήμα 1.10: Ανίχνευση μύτης

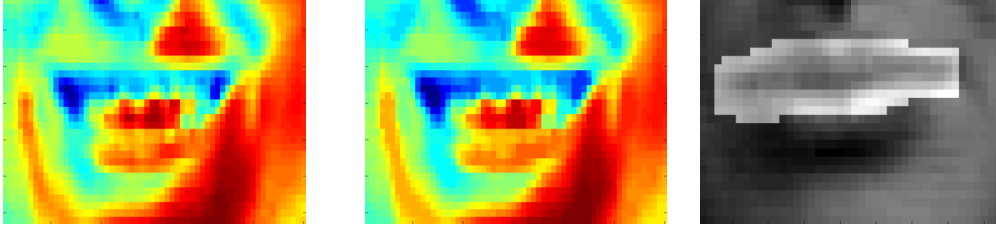
Οι οριζόντιες συντεταγμένες της μύτης μπορούν να προβλεφθούν από τις συντεταγμένες των οφθαλμικών κέντρων. Στη μάσκα  $\mathbf{M}_1^n$  οι συντεταγμένες του κέντρου κάθε συνεκτικού αντικειμένου συγκρίνονται στον οριζόντιο άξονα με την μέση διαρωθωνική απόσταση η οποία είναι κατά προσέγγιση  $r_4^n D_{bp}$  [47] και τα αντικείμενα με τις μεγαλύτερες αποστάσεις απορρίπτονται ως ακατάλληλα. Αυτά που επιλέγονται εισάγονται σε δύο λίστες. Η μία περιέχει τα υποψήφια αντικείμενα για τον αριστερό ρώθωνα και η άλλη για τον δεξιό. Η επιλογή γίνεται σύμφωνα με την εγγύτητά των αντικειμένων στον αριστερό η στο δεξιό οφθαλμό. Οι λίστες ταξινομούνται σύμφωνα με τη μέση φωτεινότητα των αντικειμένων και σε κάθε λίστα διατηρούνται τα δύο αντικείμενα με τη μικρότερη φωτεινότητα. Τέλος το μεγαλύτερο αντικείμενο από κάθε λίστα επιλέγεται ως ο αριστερός και ο δεξιός ρώθωνας αντίστοιχα, όπως φαίνεται και στο παράδειγμα της εικόνας 1.10(β). Το ρινικό κέντρο ορίζεται ως το ενδιάμεσο σημείο μεταξύ των ρωθώνων.

#### 1.4.4 Ανίχνευση στόματος

##### 1.4.4.1 Κατασκευή μάσκας χειλιών και στόματος με νευρωνικό δίκτυο

Αρχικά εκτελείται εξαγωγή του στοματικού ορίου στην περιοχή ανίχνευσης χαρακτηριστικών του στόματος (εικόνα 1.2). Για το σκοπό αυτό εκπαιδεύεται ένα πολυεπίπεδο perceptron προς αναγνώριση της στοματικής περιοχής χρησιμοποιώντας το «ουδέτερο» καρέ. Εφόσον στην ουδέτερη έκφραση το στόμα είναι κλειστό, διακρίνεται πάντα μια περιοχή χαμηλής φωτεινότητας μεταξύ των χειλιών, η οποία ανιχνεύεται ως εξής:

Η συνιστώσα της φωτεινότητας της αρχικής περιοχής ανίχνευσης χαρακτηριστικών του στόματος,  $\mathbf{L}^m$  (εικόνα 1.11α) αρχικά απλοποιείται προς μείωση του θορύβου και της περιττής πληροφορίας. Έτσι προκύπτει μια ομαλή εικόνα που αποτελείται κυρίως από μεγάλες και αριθμητικά ομοιόμορφες περιοχές ενδιαφέ-



(α) εικόνα φωτεινότητας  $\mathbf{L}^m$  (β) η μάσκα  $\mathbf{L}^m_{asfr}$  (γ) η τελική μάσκα  $\mathbf{M}^m_{t_1}$

Σχήμα 1.11: Μάσκα στόματος

ροντος. Για το σκοπό αυτό εκτελείται εναλλασσόμενο διαδοχικό φιλτράρισμα με ανακατασκευή (Alternating Sequential Filtering by Reconstruction, εξ. 1.20) στην  $\mathbf{L}_m$  οπότε προκύπτει η εικόνα  $\mathbf{L}^m_{asfr}$  όπως στο παράδειγμα 1.11(β). Η διαδικασία αυτή εξασφαλίζει τη διατήρηση των ορίων των αντικειμένων λόγω χρήσης συνεκτικών τελεστών [48].

$$\begin{aligned} f_{asfr}(I) &= \beta_n \alpha_n \dots \beta_2 \alpha_2 \beta_1 \alpha_1(I), n = 1, 2, \dots \\ \alpha_r(I) &= \rho^-(f \ominus rB|f), \beta_r(I) = \rho^+(f \oplus rB|f), r = 1, 2, \dots \end{aligned} \quad (1.20)$$

όπου με  $\rho^{+(-)}(g|f)$  συμβολίζεται το closing (opening) της  $f$  με ανακατασκευή και σηματοδευτή  $g$  και τα  $\oplus, \ominus$  δηλώνουν τους μορφολογικούς τελεστές της διαστολής και της διάβρωσης.

Η παραπάνω διαδικασία δημιουργεί επίπεδες περιοχές ανάλογης διάστασης με τη διάσταση του δομικού στοιχείου που χρησιμοποιείται. Για να αποφευχθεί η υπεραπλούστευση της εικόνας, το φίλτρο εφαρμόζεται με δομικό στοιχείο μέγιστης ακτίνας  $n \leq d_w^m \cdot r_{asf}^m$  όπου ως  $d_w^m$  συμβολίζεται το πλάτος της  $\mathbf{L}^m$ .

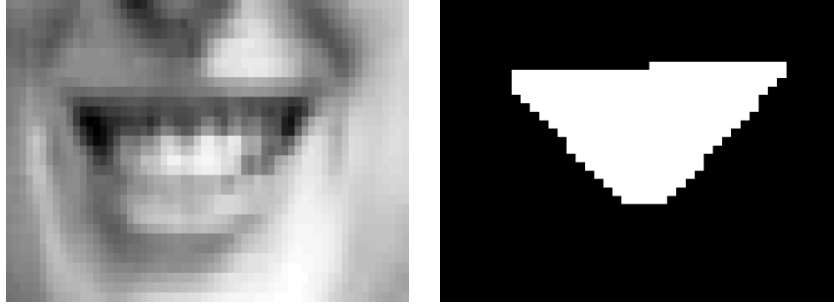
Όπως προαναφέρθηκε, στην ουδέτερη εικόνα διακρίνεται πάντα μια περιοχή χαμηλής φωτεινότητας μεταξύ των χειλιών η οποία μετά το φιλτράρισμα έχει εξομαλυνθεί και ανιχνεύεται εύκολα με κατωφλίωση. Η εικόνα μετατρέπεται σε δυαδική εφαρμόζοντας το κατώφλι  $t_1^m$ :

$$t_1^m = \frac{1}{3} (2\overline{\mathbf{L}^m} + \min(\mathbf{L}^m_{asfr})) \quad (1.21)$$

όπως φαίνεται στο παράδειγμα της εικόνας 1.11(γ). Για κάθε συνεκτικό αντικείμενο υπολογίζεται το μήκος του κυρίου άξονα μέσω ανάλυσης κυρίων συνιστωσών και το αντικείμενο με το μακρύτερο άξονα επιλέγεται ως το στόμα. Προκειμένου η μάσκα να συμπεριλάβει και τα χείλη, διαστέλλεται κάθετα και η προκύπτουσα, που συμβολίζεται με  $\mathbf{M}_t^m$  και φαίνεται στην εικόνα 1.11(γ), χρησιμοποιείται για την εκπαίδευση ενός νευρωνικού δικτύου στην αναγνώριση περιοχών στόματος και μη-στόματος. Η περιοχή που περιλαμβάνεται από τη

## Κεφάλαιο 1. Εξαγωγή των χαρακτηριστικών του προσώπου

μάσκα αντιστοιχεί στην κλάση στόμα και η περιοχή εκτός της μάσκας στην κλάση όχι-στόμα. Το πολυεπίπεδο perceptron που χρησιμοποιείται έχει 13 εισόδους και η αρχιτεκτονική του είναι όμοια με αυτό που χρησιμοποιήθηκε για την ανίχνευση των οφθαλμών.



(α) εικόνα φωτεινότητας

(β) μάσκα  $\mathbf{M}_1^m$

**Σχήμα 1.12:** 1<sup>η</sup> στοματική μάσκα

Το νευρωνικό δίκτυο που εκπαιδεύεται στο καρέ ουδέτερης έκφρασης, χρησιμοποιείται στη συνέχεια για την εκτίμηση της στοματικής περιοχής. Χρησιμοποιείται και πάλι η εικόνα φωτεινότητας, όπως στο παράδειγμα της εικόνας 1.12(α) και εφαρμόζεται κατώφλι  $t_2^m$  στην έξοδο του νευρωνικού δικτύου. Λαμβάνονται με αυτόν τον τρόπο εκτιμήσεις των περιοχών που θεωρούνται εντός της στοματικής περιοχής (αυτές με τη μεγαλύτερη βεβαιότητα). Το κυρτό περίβλημα αυτών των περιοχών συμβολίζεται με  $\mathbf{M}_1^m$  και αποτελεί μια εκτίμηση της περιοχής του στόματος όπως δεικνύεται και στο παράδειγμα της εικόνας 1.12(β).

### 1.4.4.2 Κατασκευή μάσκας με ανίχνευση ακμών

Για την κατασκευή αυτής της μάσκας το κανάλι φωτεινότητας απλοποιείται και πάλι όπως περιγράφηκε στην ενότητα 1.4.4.1 και υπολογίζεται η οριζόντια μορφολογική παράγωγος της  $\mathbf{L}_m$  με τον τρόπο που περιγράφηκε στην ενότητα 1.4.2. Η προκύπτουσα μάσκα συμβολίζεται ως  $\mathbf{M}_{b1}^m$  και το αντίστοιχο παράδειγμα βρίσκεται στην εικόνα 1.13(α).

Επιχειρείται στη συνέχεια ο η αφαίρεση των αντικειμένων της μάσκας που δεν αντιστοιχούν στο στόμα. Οι ρινικές συντεταγμένες έχουν ήδη βρεθεί. Τα συνεκτικά αντικείμενα της μάσκας  $\mathbf{M}_{b1}^m$  που βρίσκονται πολύ κοντά στη μύτη αποκλείονται. Εν συνεχεία αφαιρούνται όσα είναι πολύ μικρά (μικρότερα από  $r_{b1}^m I_w$  όπου το  $I_w$  είναι το πλάτος της μάσκας. Από τα εναπομείναντα αντικείμενα όσα απέχουν λιγότερο από  $r_{b2}^m I_w$  συνενώνονται με μορφολογικό closing με χρήση κυκλικού δομικού στοιχείου διαμέτρου  $r_{b2}^m I_w$ , οπότε προκύπτει η μάσκα  $\mathbf{M}_{b2}^m$



(α) αρχικός δυαδικός χάρτης ακμών (β) η προκύπτουσα μάσκα  $\mathbf{M}_{b_2}^m$

Σχήμα 1.13: 2<sup>η</sup> στοματική μάσκα



Σχήμα 1.14: περιοχή ανίχνευσης χαρ/κών στόματος

όπως στο παράδειγμα της εικόνας 1.13(β). Το αντικείμενο με το μεγαλύτερο μήκος στην οριζόντια διεύθυνση επιλέγεται ως η στοματική μάσκα  $\mathbf{M}_2^m$ .

#### 1.4.4.3 Μάσκα συγχώνευσης φωτεινότητας χειλιών και ακμών στόματος

Οι περισσότερες μέθοδοι ανίχνευσης ενός ανοικτού στόματος, στη βιβλιογραφία, βασίζονται στην ανίχνευση των χρωματικών χαρακτηριστικών της στοματικής κοιλότητας. Όμως τα δόντια της άνω γνάθου που εμφανίζονται κάτω από το άνω χείλος μεταβάλλουν την ομοιομορφία του χρωματικού κορεσμού όπως απεικονίζεται στην εικόνα 1.14.

Μια νέα μέθοδος περιγράφεται στη συνέχεια ώστε να αντιμετωπιστεί το προαναφερθέν πρόβλημα. Αρχικά βρίσκονται οι περιοχές χαμηλής φωτεινότητας στη συνιστώσα φωτεινότητας της περιοχής ανίχνευσης στόματος  $\mathbf{L}_m$ , με χρήση ενός χαμηλού κατωφλίου  $t_{cl}^m$  (εξ 1.22). Το αποτέλεσμα είναι μια εκτίμηση της περιοχής της στοματικής κοιλότητας ή της περιοχής μεταξύ των χειλιών στην περίπτωση κλειστού στόματος. Το προσαρμοστικό κατώφλι υπολογίζεται ως

εξής:

$$\mathbf{M}_{c1a}^m = \mathbf{L}_{asfr}^m < t_{c1}^m$$

$$t_{c1}^m = \left( \langle \mathbf{L}_{asfr}^m \rangle - \sqrt{\langle [\mathbf{L}_{asfr}^m]^2 \rangle - \langle \mathbf{L}_{asfr}^m \rangle^2} \right) \quad (1.22)$$

όπου ο τελεστής  $<$  δηλώνει τη διαδικασία της κατωφλίωσης.

Από την προκύπτουσα δυαδική μάσκα αφαιρούνται τα συνεκτικά αντικείμενα που συνορεύουν με τα εξωτερικά άκρα της εικόνας, απορρίπτοντας έτσι αντικείμενα του φόντου. Προκύπτει έτσι η μάσκα  $\mathbf{M}_{c1}^m$  που στο παράδειγμα απεικονίζεται στην εικόνα 1.15(α). Εξετάζονται τώρα δύο διακριτές περιπτώσεις: ή δεν υπάρχουν εμφανή δόντια και η στοματική περιοχή αποτελείται από μία συνεκτική σκοτεινή περιοχή (περίπτωση 1) ή τα δόντια είναι εμφανή και υπάρχουν δύο περιοχές χαμηλής φωτεινότητας εκατέρωθεν των δοντιών (περίπτωση 2) αντίστοιχες με αυτές της εικόνας 1.14 που αντιστοιχούν στις δύο γωνίες του χειλιών. Πρέπει να σημειωθεί ότι αυτή η περίπτωση απαντάται ακόμα και σε εκτεταμένα χαμόγελα. Εν συνεχεία το συνεκτικό αντικείμενο με τη μεγαλύτερη επιφάνεια επιλέγεται από τη  $\mathbf{M}_{c1}^m$  και υπολογίζεται το κέντρο βάρους του. Εάν το τελευταίο βρίσκεται κοντά στο ρινικό κέντρο στον οριζόντιο άξονα επιλέγεται η περίπτωση 1, αλλιώς βρισκόμαστε στην περίπτωση 2 όπου υπάρχουν δύο περιοχές χαμηλής φωτεινότητας εκατέρωθεν των δοντιών. Η οριζόντια εγγύτητα προς το ρινικό κέντρο εκτιμάται από το κατώφλι  $r_{c2}^m D_{bp}$  που υπολογίζεται ως ποσοστό της διαχωρικής απόστασης και του πλάτους του στόματος. Οι δύο περιπτώσεις είναι αρκετά διακριτές. Στην δεύτερη περίπτωση επιλέγεται το δεύτερο σε μέγεθος αντικείμενο, και κατασκευάζεται μια δυαδική μάσκα η οποία περιέχει ένα αντικείμενο ανάλογο με την πρώτη περίπτωση και δύο αντικείμενα στην δεύτερη. Το κυρτό περίβλημα των αντικειμένων αποτελεί τη μάσκα  $\mathbf{M}_{c2}^m$  που στο παράδειγμα απεικονίζεται στην εικόνα 1.15(β).

Οι γωνίες των χειλιών, που ανιχνεύονται στη δεύτερη περίπτωση, αποτελούν μια ισχυρή εκτίμηση της οριζόντιας έκτασης του στόματος αλλά δεν επαρκούν για τον προσδιορισμό του ανοίγματός του. Για το σκοπό αυτό η μάσκα  $\mathbf{M}_{c2}^m$  επεκτείνεται ώστε να περιλάβει και το κάτω χείλος. Κατασκευάζεται επιπλέον ένας χάρτης οριζόντιων ακμών στην περιοχή ανίχνευσης χαρακτηριστικών στόματος και συμβολίζεται με  $\mathbf{M}_{c3}^m$  (εικόνα 1.15γ). Η μάσκα αυτή περιέχει και αντικείμενα σε εγγύτητα με το κάτω μέρος του στόματος. Οι δύο μάσκες  $\mathbf{M}_{c2}^m$  και  $\mathbf{M}_{c3}^m$  πρέπει να συγχωνευτούν. Ένας αποτελεσματικός τρόπος είναι να κρατηθούν από κάθε μάσκα τα αντικείμενα που έχουν κοινά στοιχεία και στις δύο μάσκες καθώς και αυτά που βρίσκονται σε εγγύτητα μεταξύ τους (μεταξύ των

μασκών). Εφόσον η  $\mathbf{M}_{c_2}^m$  δύναται στην περίπτωση 2 να μην περιέχει αντικείμενα που ανήκουν στα χαμηλότερα μέρη του στόματος, επεκτείνεται προς τα κάτω με χρήση ενός μη συμμετρικού ακτινικού δομικού στοιχείου οπότε προκύπτει η μάσκα  $\mathbf{M}_{c_{2d}}^m$ . Οι μάσκες συνδυάζονται με μορφολογική ανακατασκευή [49] ως εξής: χρησιμοποιείται ως είσοδος μάσκα με τα στοιχεία που βρίσκονται σε οποιαδήποτε από τις δύο μάσκες  $\mathbf{M}_{c_3}^m$  και  $\mathbf{M}_{c_{2d}}^m$  και τα κοινά στοιχεία τους, στοιχεία δηλαδή που βρίσκονται και στις δύο μάσκες, ως σημαδευτές (εξ. 1.23). Η τελική μάσκα  $\mathbf{M}_3^m$  για το παράδειγμα απεικονίζεται στην εικόνα 1.15(δ).

$$\begin{aligned} \mathbf{M}_{c_{23}}^m &= f_a(\mathbf{M}_{c_3}^m, \mathbf{M}_{c_{2d}}^m) \\ f_a(\mathbf{A}, \mathbf{B}) &= \{c_{ij}\}, c_{ij} = \{a_{ij}b_{ij}\} \\ f_o(\mathbf{A}, \mathbf{B}) &= \{c_{ij}\}, c_{ij} = \left\{ \begin{array}{l} a_{ij}, b_{ij} = 0 \\ b_{ij}, a_{ij} = 0 \end{array} \right\} \\ \mathbf{M}_3^m &= \rho(\mathbf{M}_{c_{23}}^m, f_o(\mathbf{M}_{c_{2d}}^m, \mathbf{M}_{c_3}^m)) \end{aligned} \quad (1.23)$$

όπου με  $\rho(\mathbf{B}, \mathbf{A})$  συμβολίζεται η ανακατασκευή του  $\mathbf{A}$  με σημαδευτή  $\mathbf{B}$ .



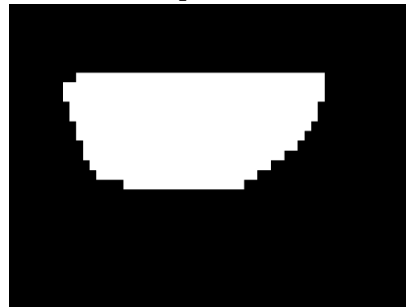
(α) μάσκα  $\mathbf{M}_{c_1}^m$  χωρίς αντικείμενα φόντου



(β) μάσκα  $\mathbf{M}_{c_2}^m$  με εμφανή δόντια

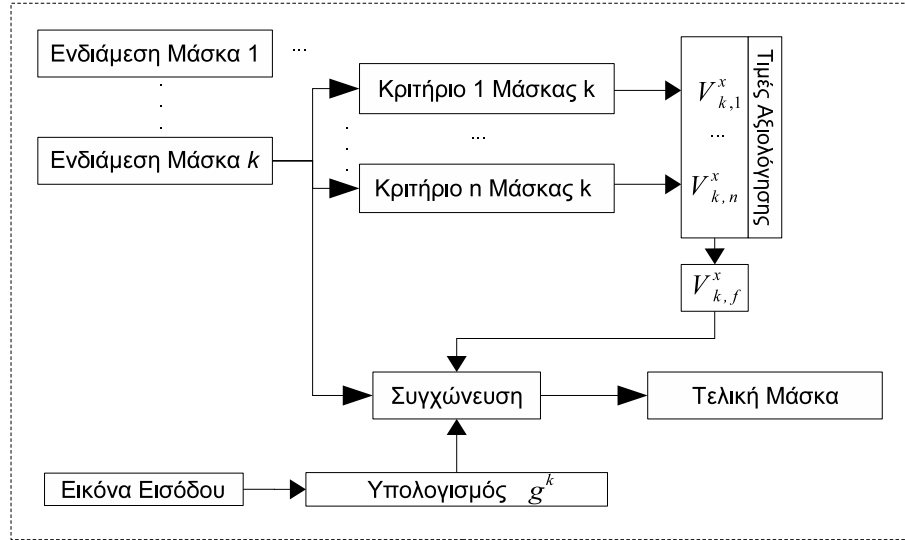


(γ) μάσκα οριζόντιων ακμών  $\mathbf{M}_{c_3}^m$



(δ) τελική μάσκα  $\mathbf{M}_3^m$

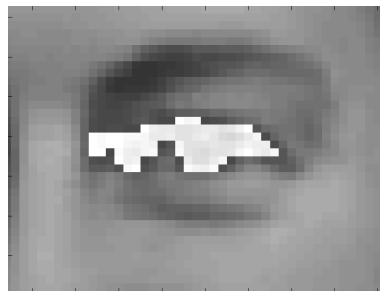
Σχήμα 1.15: 3<sup>η</sup> στοματική μάσκα



Σχήμα 1.16: Η διαδικασία συγχώνευσης масκών

## 1.5 Συγχώνευση των ενδιάμεσων масκών

Οι ενδιάμεσες μάσκες των χαρακτηριστικών του προσώπου πρέπει να συγχωνευθούν για την παραγωγή μιας τελικής μάσκας για κάθε χαρακτηριστικό. Τα πιο συχνά προβλήματα των ανιχνευμένων масκών, που απαντώνται κυρίως σε εικόνες χαμηλής ποιότητας, περιλαμβάνουν σύνδεση με τα όρια άλλων περιοχών ή μετατόπιση της μάσκας λόγω θορύβου, όπως απεικονίζεται στην εικόνα 1.17. Υπάρχουν περιπτώσεις όπου οι μάσκες παρέχουν εντελώς λανθασμένες εκτιμήσεις. Τα προβλήματα στην είσοδο όπως αλλαγές στη φωτεινότητα και ο θόρυβος κωδικοποίησης/συμπίεσης δεν μπορούν να προβλεφθούν επομένως οι μάσκες πρέπει να επαναξιολογούνται σε κάθε καρέ.



Σχήμα 1.17: Μια προβληματική μάσκα

### 1.5.1 Αξιολόγηση οφθαλμικών και στοματικών масκών

Οι προτεινόμενοι αλγόριθμοι που παρουσιάζονται στην ενότητα 1.4 παράγουν μία μάσκα  $\mathbf{M}^b$  για κάθε φρύδι, τις ρινικές συντεταγμένες, τέσσερις ενδιάμεσες

μάσκες για καθέναν από τους οφθαλμούς  $\mathbf{M}_{1..4}^e$  και τρεις ενδιάμεσες στοματικές μάσκες  $\mathbf{M}_{1..3}^m$ .

Οι τέσσερις οφθαλμικές μάσκες για κάθε μάτι και οι τρεις στοματικές μάσκες πρέπει συγχωνευτούν σε μια τελική μάσκα για κάθε χαρακτηριστικό. Εφόσον η αξιολόγηση των ενδιάμεσων μασκών μπορεί να γίνει μόνο στις πλήρεις μάσκες, δε γίνεται να δοθούν διαφορετικές τιμές εμπιστοσύνης σε διαφορετικά τμήματα της κάθε ενδιάμεσης μάσκας, επομένως κάθε pixel των μασκών αυτών μοιράζεται την ίδια τιμή εμπιστοσύνης. Για την αξιολόγηση των ενδιάμεσων μασκών, προτείνεται ο έλεγχός τους με ένα σύνολο ανθρωπομετρικών κριτηρίων. Ορισμένα από τα κριτήρια αυτά σχετίζονται είτε με μεταβλητά χαρακτηριστικά του προσώπου είτε με την αισθητική, επομένως δεν είναι δυνατή η εφαρμογή αυστηρής ανθρωπομετρικής κρίσης των μασκών.

Για κάθε μάσκα  $k$  του χαρακτηριστικού  $x$ , εφαρμόζεται ένα σύνολο κριτηρίων αξιολόγησης  $V_{k,i}^x$ , όπου  $i$  ο δείκτης του κάθε κριτηρίου. Οι τιμές των κριτηρίων συνδυάζονται σε μια τελική τιμή αξιολόγησης  $V_{k,f}^x$  για την κάθε ενδιάμεση μάσκα. Το κάθε κριτήριο παρέχει ένα μέτρο της ομοιότητας του σχήματος και της θέσης του χαρακτηριστικού του προσώπου σε σχέση με τις αναμενόμενες τιμές του. Οι αναμενόμενες τιμές ορίζονται από το [32] αλλά και από εικόνες που έχουν εξαχθεί από 20 άτομα της βάσης δεδομένων [29]. Σε κάθε μάσκα προσαρτεία μία τιμή αξιολόγησης στο διάστημα  $[0, 1]$  με τις μεγαλύτερες τιμές να δηλώνουν εγγύτητα στις αναμενόμενες τιμές.

Όλες οι τιμές αξιολόγησης βασίζονται σε αποστάσεις οι οποίες ορίζονται στον πίνακα 1.5.1 Με τη χρήση των ορισμών αυτών, οι οφθαλμικές μάσκες αξιολογούνται βάσει τεσσάρων κριτηρίων που ορίζονται στον πίνακα 1.5.1 και οι οποίες σχετίζονται με τις οφθαλμικές διαστάσεις, τις αποστάσεις μεταξύ των οφθαλμών και τις αποστάσεις μεταξύ τμημάτων των οφθαλμών και των φρυδιών. Τέλος η αξιολόγηση των στοματικών μασκών βασίζεται σε τέσσερα κριτήρια που ορίζονται στον πίνακα 1.7.

Στη συνέχεια, η τιμή αξιολόγησης του κριτηρίου  $i$  για τη μάσκα  $k$  του χαρακτηριστικού  $x$  συμβολίζεται ως  $V_{k,i}^x \in [0, 1]$ , όπου οι τιμές του  $V_{k,i}^x$  κανονικοποιούνται στο  $[0, 1]$ , δηλαδή εάν  $V_{k,i}^x > 1$  τότε  $V_{k,i}^x = 1$  και εάν  $V_{k,i}^x < 0$  τότε  $V_{k,i}^x = 0$ .

Επιθυμούμε οι μάσκες με πολύ χαμηλή τιμή των κριτηρίων αξιολόγησης να εξαιρεθούν από τη διαδικασία συγχώνευσης και κατά συνέπεια να μη συνεισφέρουν στον υπολογισμό της τιμής αξιολόγησης της τελικής μάσκας για κάθε χαρακτηριστικό. Ως εκ τούτου, αγνοούμε τις μάσκες για τις οποίες ισχύει:  $V_{k,f}^x < (t_{vd} \cdot \langle V_{k,i}^x \rangle_i)$ .

Η τιμή αξιολόγησης της τελικής μάσκας  $k$  υπολογίζεται επομένως ως ο



Αποστάσεις Αξιολόγησης Μασκών	
$d_1$	Απόσταση της ανώτερης οριζόντιας οφθαλμικής συντεταγμένης από το κάτω μέσο του φρυδιού
$d_2$	Οφθαλμικό Εύρος
$d_3$	Οφθαλμικό Ύψος
$d_4$	Απόσταση των κάθετων συντεταγμένων του οφθαλμικού κέντρου και του μέσου του φρυδιού
$d_5$	Οφθαλμικό Εύρος
$d_6$	$D_{bp}$ , Διακορική Απόσταση
$d_7$	Απόσταση των κάθετων συντεταγμένων του οφθαλμικού κέντρου και του μέσου του στόματος
$d_8$	Εύρος Στόματος
$d_9$	Ύψος Στόματος
$d_{10}$	Απόσταση sellion - στομίου
$d_{11}$	Απόσταση Sellion - υπορρινίου

μέσος των υπολοίπων τιμών:

$$V_{k,f}^x = \langle V_{k,i'}^x \rangle_{i'}, i' : V_{k,i'}^x \geq t_{vd} \langle V_{k,i}^x \rangle_i, i \in N_n \quad (1.24)$$

### 1.5.2 Συγχώνευση Μασκών

Καθεμία από τις ενδιάμεσες μάσκες αντιπροσωπεύει την καλύτερη προσπάθεια της αντίστοιχης μεθόδου με την οποία κατασκευάστηκε. Οι πολλαπλές οφθαλμικές και στοματικές ενδιάμεσες μάσκες πρέπει να συγχωνευθούν για την παρασκευή μιας τελικής μάσκας για κάθε χαρακτηριστικό. Η συγχώνευση βασίζεται στην υπόθεση ότι έχοντας πολλαπλές μάσκες για κάθε χαρακτηριστικό μειώνεται η πιθανότητα να είναι όλες άκυρες εφόσον καθεμία παράγει διαφορετικά πρότυπα λαθών. Έχει αποδειχθεί στη θεωρία μηχανών επιτροπής (committee machine) [50] [51] ότι για την επιθυμητή έξοδο  $t$  το συνδυαστικό λάθος  $y_{comb} - t$  από διαφορετικές μηχανές  $f_i$  είναι χαμηλότερο από το μέσο λάθος:

Κεφάλαιο 1. Εξαγωγή των χαρακτηριστικών του προσώπου

Μεταβλητή	Κριτήριο	Περιγραφή
$V_{k,1}^e$	$1 -  (d_1)/(d_6/4) - 1 $	Κάθετη απόσταση οφθαλμών-φρυδιών
$V_{k,2}^e$	$1 -  1 - (d_2/d_6)/0.49 $	Αναλογία οφθαλμικού εύρους/διακορικής απόστασης
$V_{k,3}^e$	$0.3 - (d_3 - d_2)/d_2$	Αναλογία εύρους/πλάτους οφθαλμών
$V_{k,4}^e$	$1 -  d_4 /d_5$	Οριζόντια ευθυγράμμιση οφθαλμών-φρυδιών
Σημ: (Οφθαλμικό Εύρος)/(Διακορική Απόσταση)=0.49 [32]		

**Πίνακας 1.6:** Ανθρωπομετρικές οφθαλμικές μετρήσεις αξιολόγησης μασκών

Μεταβλητή	Κριτήριο	Περιγραφή
$V_{k,1}^m$	$1 -  d_7 /d_6$	Οριζόντια απόσταση στοματικού και οφθαλμικού κέντρου.
$V_{k,2}^m$	$1 - \left  \frac{d_8}{d_6} \frac{1}{0.82} - 1 \right $	Αναλογία στοματικού εύρους και διακορικής απόστασης
$V_{k,3}^m$	$1 \text{ εάν } d_9 < (1.3d_6) \text{ αλλιώς } d_9/(1.3d_6)$	Αναλογία στοματικού ύψους και διακορικής απόστασης
$V_{k,4}^m$	$1 - \left  1 - \frac{(d_{10} - d_{11})}{d_6} \frac{1}{0.344} \right $	Ρινική απόσταση από το άνω χείλος

Σημ. 1: (Διαχειλικό Εύρος)/(Διακορική Απόσταση)=0.82

Σημ. 2: (Μήκος στομίου - Υπορίνιου)/(Διακορική Απόσταση)=0.344 [33] [32]

**Πίνακας 1.7:** Ανθρωπομετρικές μετρήσεις αξιολόγησης μασκών στόματος

$$\begin{aligned} y_{comb} &= \frac{1}{M} \sum y_i \\ (y_{comb} - t)^2 &= \frac{1}{M} \sum_i (y_i - t)^2 - \frac{1}{M} \sum_i (y_i - y_{comb})^2 \end{aligned} \quad (1.25)$$

Εφόσον οι ενδιάμεσες μάσκες διαθέτουν μια τιμή αξιολόγησης η οποία αναπαριστά την «αληθοφάνεια» του να αποτελούν πραγματικές μάσκες για το χαρακτηριστικό που αντιπροσωπεύουν, φαίνεται φυσικό να συνδυαστούν δίνοντας μεγαλύτερη εμπιστοσύνη σε αυτές με τη μεγαλύτερη τιμή αξιολόγησης από τη μία και από την άλλη αγνοώντας αυτές οι οποίες σίγουρα δεν θα συνεισφέρουν θετικά στο αποτέλεσμα. Επιπλέον είναι επιθυμητό να ευνοηθούν οι αλγόριθμοι εξαγωγής масκών εξετάζοντας τις ιδιότητες τις εκάστοτε εισόδου, δηλαδή να δοθεί μεγαλύτερη εμπιστοσύνη στις μάσκες που βασίζονται στο χρώμα όταν ανιχνευθεί ότι η είσοδος παρέχει χρωματικές συνιστώσες καλής ποιότητας και στις μάσκες που προέρχονται από το νευρωνικό δίκτυο όταν το πρόσωπο έχει αρκετή ανάλυση ώστε το τελευταίο να παρέχει ακριβή ανίχνευση των ορίων.

Όσον αφορά την ποιότητα των εικόνων εισόδου, δύο παράμετροι λαμβάνονται υπόψη: η ανάλυση της εικόνας και η ποιότητα του χρώματος. Όσον αφορά την ποιότητα, μη συνθετικά δεδομένα εκμάθησης είναι πολύ δύσκολο να αποκτηθούν, βρέθηκε όμως ότι καλή εκτιμήτρια της ποιότητας είναι η διασπορά των χρωματικών συνιστωσών στην περιοχή του δέρματος του προσώπου: πολύ μεγάλες τιμές διακύμανσης στις χρωματικές συνιστώσες υποδηλώνουν την παρουσία χρωματικού θορύβου ή και υπερφωτισμού. Επομένως οι  $\sigma_{cr}$ ,  $\sigma_{cb}$  είναι μικρότερες από  $t_\sigma$  όταν οι χρωματικές συνιστώσες έχουν ικανοποιητική ποιότητα και πολύ μεγαλύτερες στην αντίθετη περίπτωση.

Όσον αφορά την ανάλυση, βρέθηκε ότι το προτεινόμενο ν.δ. αποδίδει πολύ ικανοποιητικά σε ακολουθίες όπου  $D_{bp} > t_r$  pixels, όπου με  $D_{bp}$  συμβολίζεται η διαχωρική απόσταση.

Ακολούθως οι τελικές μάσκες για τον αριστερό και δεξιό οφθαλμό και το στόμα θα συμβολίζονται ως:  $\mathbf{M}_f^{eL}, \mathbf{M}_f^{eR}, \mathbf{M}_f^m$ .

Για την ενδιάμεση μάσκα  $k$  του χαρακτηριστικού  $x$ , η μεταβλητή  $V_{k,f}^x$  ορίζει ποιες μάσκες ευνοούνται σύμφωνα με τις τελικές τιμές αξιολόγησής τους και η μεταβλητή  $g^k$  ποιοί αλγόριθμοι ευνοούνται σύμφωνα με τα χαρακτηριστικά της εισόδου. Επιπλέον κάθε στοιχείο (pixel) στην τελική μάσκα  $\mathbf{M}_f^x$  συμβολίζεται ως  $m_f^x$  και κάθε στοιχείο στην ενδιάμεση μάσκα  $k$ ,  $\mathbf{M}_k^x$  ως  $m_k^x$ ,  $k \in N_n$  όπου οι συντεταγμένες δεν σημειώνονται για λόγους απλότητας.

Επιπλέον επιθυμούμε τη πραγματοποίηση της συγχώνευσης των масκών σε επίπεδο pixel. Συνεπώς τα pixel της τελικής μάσκας δεν θα προέρχονται υποχρεωτικά από τις ίδιες ενδιάμεσες μάσκες. Κάθε pixel στην τελική μάσκα θα

φέρει την τιμή αξιολόγησης  $v_f^x$  που θα αντανακλά τις τιμές αξιολόγησης και την καταλληλότητα των αλγορίθμων ανίχνευσης των μασκών από τις οποίες προήλθε. Οι τιμές  $v_f^x$  για όλα τα pixels στην τελική μάσκα θα σχηματίζουν τις τιμές αξιολόγησης των τελικών μασκών  $V_f^x$ .

Ας ορίσουμε τη συνάρτηση μεταξύ του  $m_f^x \in \{0, 1\}$ ,  $v_f^x \in [0, 1]$  και του  $m_k^x \in \{0, 1\}$  ως

$$v_f^x = f(m_k^x, V_{k,f}^x, g^k) \quad (1.26)$$

$$m_f^x = F(v_f^x) \quad (1.27)$$

οι προηγούμενες απαιτήσεις μπορούν να εκφραστούν ως εξής:

1. Εάν όλες οι μάσκες  $k$  συμφωνούν ότι το pixel  $m_k^x$  δεν ανήκει στο χαρακτηριστικό  $x$ , τότε αυτό πρέπει να αντανακλάται στο αποτέλεσμα της συγχώνευσης ανεξάρτητα από τις τιμές αξιολόγησης  $V_{k,f}^x$ :

$$\forall k \in N_n, m_k^x = 0 \Rightarrow m_f^x = 0 \quad (1.28)$$

2. Η μεταβλητή  $g^k$  πρέπει να είναι κανονικοποιημένη σύμφωνα με τον αριθμό των μασκών:

$$\sum_{k=1}^n g^k = n \quad (1.29)$$

3. Εάν όλες οι μάσκες  $k$  συμφωνούν ότι το pixel  $m_f^x$  ανήκει στο χαρακτηριστικό  $x$  με μέγιστη βεβαιότητα, αυτό πρέπει να αντανακλάται στο αποτέλεσμα της συγχώνευσης:

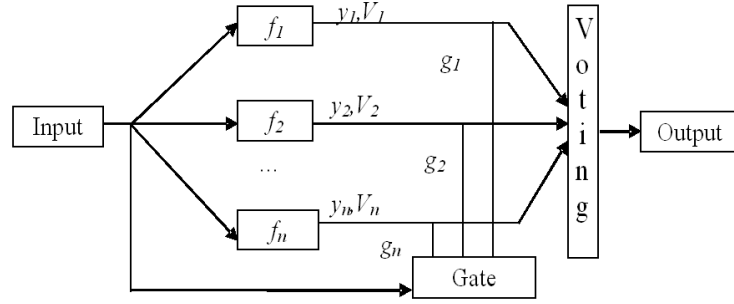
$$\text{Εάν } \forall k \in N_n, m_k^x = 1 \wedge V_{k,f}^x = 1 \Rightarrow m_f^x = 1, v_f^x = 1 \quad (1.30)$$

4. Εάν όλες οι μάσκες  $k$  αποτύχουν, τότε κανένα στοιχείο στην τελική μάσκα δεν πρέπει να ανήκει στο χαρακτηριστικό  $x$ :

$$\forall k \in N_n, V_{k,f}^x = 0 \Rightarrow m_f^x = 0 \quad (1.31)$$

5. Εάν αποτύχει μία μάσκα, τότε το αποτέλεσμα πρέπει να εξαρτάται μόνο από τις υπόλοιπες:

$$\exists k_0 \in N_n : V_{k_0,f}^x = 0 \Rightarrow m_f^x = \bigvee_{k \in N_n - \{k_0\}} f(m_k^x, V_{k,f}^x, g^k) \quad (1.32)$$



Σχήμα 1.18: Το μοντέλο της Δυναμικής Μηχανής Επιτροπής

6. Η συγχώνευση με μάσκες υψηλότερης αξιολόγησης πρέπει να παράγει υψηλότερη εμπιστοσύνη στις μάσκες εξόδου, στα pixels που προέρχονται από τις συγκεκριμένες μάσκες:

$$\begin{aligned} &\text{Εάν } V_{k_0,f}^{x_1} > V_{k_0,f}^{x_2} \text{ και } \forall k \in N_n - \{k_0\} \text{ και ισχύει } V_{k,f}^{x_1} = V_{k,f}^{x_2} \\ &\text{και το ίδιο ισχύει για όλα τα } m_k^{x_j} \neq 0, g^{k,j}, j=1,2 \text{ τότε } v_f^{x_2} > v_f^{x_1} \end{aligned} \quad (1.33)$$

7. Εάν μια μάσκα προέρχεται από καταλληλότερο αλγόριθμό για τη συγκεκριμένη είσοδο, τότε τα pixels που προέρχονται από αυτή στην τελική μάσκα πρέπει να συσχετιστούν με μεγαλύτερες τιμές εμπιστοσύνης:

$$\begin{aligned} &\text{Εάν } g^{k,1} > g^{k,2} \text{ και } \forall k \in N_n - \{k_0\} \text{ και ισχύει } V_{k,f}^{x_1} = V_{k,f}^{x_2} \\ &\text{και το ίδιο ισχύει και για όλα τα } m_k^{x_j} \neq 0, V_{k,f}^{x_j}, j=1,2 \text{ τότε } v_f^{x_2} > v_f^{x_1} \end{aligned} \quad (1.34)$$

Για να ικανοποιηθούν οι παραπάνω απαιτήσεις, προτείνεται μια διαδικασία συγχώνευσης η οποία βασίζεται στη ιδέα των Δυναμικών Μηχανών Επιτροπής, που αναπαρίστανται στο σχήμα 1.18. Σε μια στατική μηχανή επιτροπής, το βάρος της κάθε ψήφου είναι ανάλογο ή καθορίζεται από το αντίστοιχο ποσοστό σφάλματός της σε ένα δείγμα επικύρωσης. Στις δυναμικές μηχανές επιτροπής, η ίδια η είσοδος εμπλέκεται στην απονομή βαρών μέσω μιας διάταξης πύλης η οποία χρησιμοποιείται για την μεταβολή των βαρών δυναμικά λαμβάνοντας υπόψη την είσοδο.

Ως είσοδος της μηχανής θεωρούνται οι ενδιαμέσες μάσκες  $\mathbf{M}_k^x$ , οι τιμές  $V_{k,f}^x$  θεωρούνται ως οι τιμές εμπιστοσύνης κάθε εισόδου και η μεταβλητή  $g^k$  αναλαμβάνει το ρόλο της πύλης (gate). Οι τελικές μάσκες  $\mathbf{M}_f^{xL}, \mathbf{M}_f^{eR}$  και  $\mathbf{M}_f^m$  θεωρούνται ως η έξοδος της μηχανής.

Κάθε pixel  $m_f^x$  στην τελική μάσκα  $\mathbf{M}_f^x$  υπολογίζεται από τις  $n$  μάσκες ως εξής:

$$f(m_k^x; V_{k,f}^x, g^k) = \frac{1}{n} \sum_{k=1}^n m_k^x V_{k,f}^x g^k \quad (1.35)$$

$$F(v_f^x) = \begin{cases} 0, & v_f^x < (\langle V_f^x \rangle | v_f^x > 0) \\ 1, & \text{διαφορετικά} \end{cases} \quad (1.36)$$

Ο ρόλος της μεταβλητής πύλης  $g^k$  είναι η εύνοια των αλγορίθμων εξαγωγής μασκών που λαμβάνουν υπόψη τους το χρώμα,  $(\mathbf{M}_1^e, \mathbf{M}_1^m)$  στις εικόνες υψηλής ποιότητας χρώματος και ανάλυσης. Η μεταβλητή  $g^i$  ορίζεται ως εξής:

$$g^k = \begin{cases} n - \left(\frac{n-1}{n}\right), k=1, D_{bp} > t_r, \sigma_{cr} < t_\sigma, \sigma_{cb} < t_\sigma \\ \frac{1}{n}, k \neq 1, D_{bp} > t_r, \sigma_{cr} < t_\sigma, \sigma_{cb} < t_\sigma \\ 1, \text{διαφορετικά} \end{cases} \quad (1.37)$$

όπου  $D_{bp}$  η διακορική απόσταση,  $\sigma_{cr}, \sigma_{cb}$  η τυπική απόκλιση των συνιστωσών χρώματος  $C_r, C_b$  αντίστοιχα εντός της περιοχής του προσώπου. Δεν είναι δύσκολο να διαπιστώσουμε ότι οι εξισώσεις (1.35)–(1.37) ικανοποιούν τις (1.34)–(1.34)

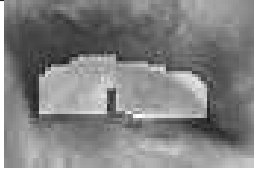


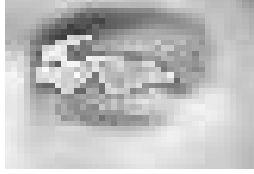

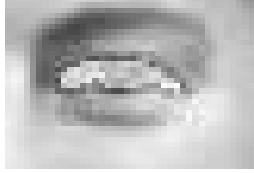


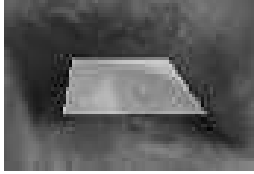

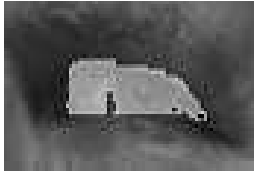
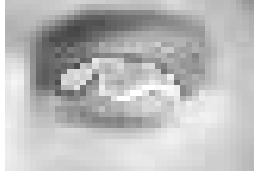
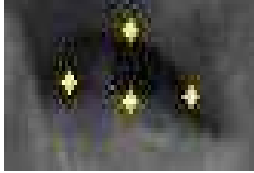
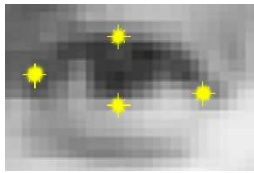
Οι πίνακες 1.8 1.9 και 1.10 απεικονίζουν παραδείγματα συγχώνευσης για τον αριστερό οφθαλμό και το στόμα αντίστοιχα, στα οποία κάποιες από τις μάσκες είναι προβληματικές. Οι τιμές αξιολόγησης αναφέρονται στην αντίστοιχη τιμή αξιολόγησης των μασκών, ενώ η διακορική απόσταση αναφέρεται ως ενδεικτική της ανάλυσης και κατ'επέκταση της ποιότητας των ακολουθιών βίντεο. Στους πίνακες παριστάνονται και τα αντίστοιχα FP στις τελικές μάσκες για την οπτική επαλήθευση των αποτελεσμάτων.

### 1.5.3 Εκτίμηση Εμπιστοσύνης για τις μάσκες Οφθαλμών, Στόματος και Φρυδιών

Οι τιμές εμπιστοσύνης των μασκών υποδηλώνουν το μέτρο κατά το οποίο γνωρίζουμε ότι ένα χαρακτηριστικό έχει εξαχθεί σωστά και επομένως το βαθμό κατά τον οποίο η ανάλυση έκφρασης πρέπει να βασιστεί στο συγκεκριμένο χαρακτηριστικό. Συνεπώς οι τιμές εμπιστοσύνης μεταβιβάζονται από τις μάσκες στα αντίστοιχα FP και FAP και το στάδιο εκτίμησης της έκφρασης.

Για την εκτίμηση της τιμών εμπιστοσύνης κρίνεται η ομοιότητα κάθε χαρακτηριστικού με τις μέσες τιμές διαστάσεων χαρακτηριστικών που λαμβάνονται από το [32]. Εφόσον αντίστοιχες τιμές για τα φρύδια δεν ήταν διαθέσιμες στην βιβλιογραφία, οι μετρήσεις εμπλουτίζονται με πληροφορία που αφορά τη χρο-




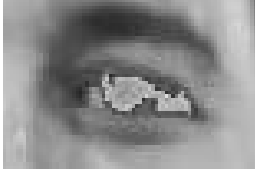

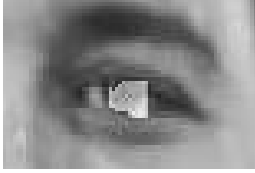


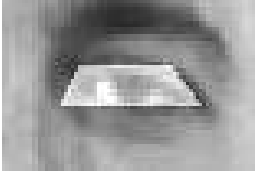
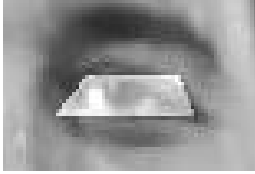
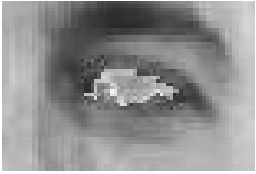
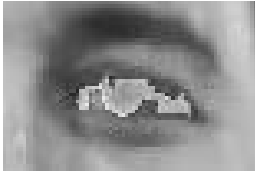
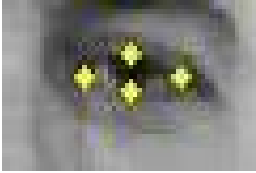
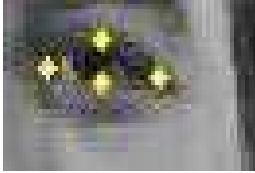
Κεφάλαιο 1. Εξαγωγή των χαρακτηριστικών του προσώπου

Μάσκα	rd-12259 *	$V_f^e$	al-27 *	$V_f^e$
$M_{nn}^e$				
$M_1^e$		0.823		0.839
$M_2^e$		0.763		0.810
$M_3^e$		0.716		0.787
$M_4^e$		0.826		0.872
$M_f^e$				
FPs				
	$D_{bp}$ : 96 px		$D_{bp}$ : 40 px	

Η διακορική απόσταση  $D_{bp}$  αναφέρεται ως ενδεικτική της ανάλυσης

\*Αριθμός ακολουθίας στις Β.Δ. [29] [30]

**Πίνακας 1.8:** Παραδείγματα συγχώνευσης масκών στον αριστερό οφθαλμό, αντίστοιχες τιμές αξιολόγησης και FPs

Μάσκα	kk-1002 *	$V_f^e$	kk-1998*	$V_f^e$
$M_{nn}^e$				
$M_1^e$		0.825		0.813
$M_2^e$		0.782		0.581
$M_3^e$		0.866		0.733
$M_4^e$		0.883		0.917
$M_f^e$				
FPS				

$D_{bp}$ : 58 px









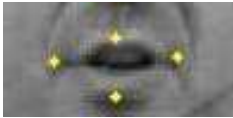
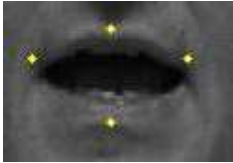
$D_{bp}$ : 58 px

Η διακορική απόσταση  $D_{bp}$  αναφέρεται ως ενδεικτική της ανάλυσης

\*Αριθμός ακολουθίας στις Β.Δ. [29] [30]

**Πίνακας 1.9:** Παραδείγματα συγχώνευσης масκών στον αριστερό οφθαλμό, αντίστοιχες τιμές αξιολόγησης και FPS



Μάσκα	kk-1014	$V_f^m$	Rd-1113	$V_f^m$
$M_1^m$		0.820		0.538
$M_2^m$		0.868		0.752
$M_3^m$		0.828		0.821
$M_f^m$				
FPs				

**Πίνακας 1.10:** Παραδείγματα συγχώνευσης масκών στο στόμα, αντίστοιχες τιμές αξιολόγησης και  $FP$

νική διαστασιολογική σταθερότητα των χαρακτηριστικών του προσώπου και τη συμμετρία του προσώπου.

Οι τιμές εμπιστοσύνης που συνδέονται με κάθε τελική μάσκα συμβολίζονται ως  $C^e, C^b, C^m \in [0, 1]$  και βρίσκονται στο διάστημα  $[0..1]$ . Για τη μύτη δεν εξάγονται τιμές εμπιστοσύνης και έτσι θεωρείται πάντα ότι  $C^n = 1$ . Αυτές οι τιμές παράγονται από ένα σύνολο κριτηρίων τα οποία συμπληρώνουν τις τελικές τιμές αξιολόγησης  $V_f$  που χρησιμοποιούνται για τη συγχώνευση. Τα κριτήρια αυτά σχετίζονται με:

1. τη χρονική σταθερότητα των διαστάσεων, παράγοντας τις  $C_{med}^b, C_{med}^e$
2. τη συμμετρία του προσώπου, παράγοντας την  $C_s^e$
3. συμφωνία με τις ανθρωπολογικές μέσες μετρήσεις, παράγοντας τις  $C_1^e, C_2^e, C_1^m$ .

Οι τιμές αυτές υπολογίζονται ως εξής:

1. Εξαιρουμένου του στόματος, το εύρος των χαρακτηριστικών του προσώπου παραμένει σχεδόν σταθερό ακόμα και σε έντονες εκφράσεις. Το μετρούμενο εύρος για τα φρύδια  $w_i^b$  και κάθε οφθαλμό  $w_i^{eL}, w_i^{eR}$  εξετάζεται σε κάθε καρέ  $i$  και υπολογίζεται η ενδιάμεση τιμή  $\tilde{w}^x$  των τελευταίων 10 καρέ για το χαρακτηριστικό  $x$ . Σε κάθε καρέ, η ομοιότητα μεταξύ του  $w_i^x$  και του  $\tilde{w}^x$  στα τελευταία 10 καρέ χρησιμοποιείται ως εκτίμηση της  $C_{med}^b$  για τα φρύδια και της  $C_{med}^e$  για τα μάτια:

$$C_{med,i}^x = 1 - |w_i^x - med(w_j^x, j = i - 10..i)| (w_i^x)^{-1} \quad (1.38)$$

2. Αξιοποιώντας τη συμμετρία του προσώπου μπορεί να εκτιμηθεί η ομοιότητα μεταξύ των άνω τμημάτων αριστερού και δεξιού βλεφάρου, η οποία συμβολίζεται με  $C_s^e \in [0, 1]$ . Ας ορίσουμε ως  $\mathbf{X}^L, \mathbf{X}^R$  τα διανύσματα που περιέχουν τις οριζόντιες συντεταγμένες των άνω τμημάτων των δύο βλεφάρων. Η τιμή  $C_s$  που δηλώνει την ομοιότητά τους μπορεί να υπολογιστεί ως ο δισδιάστατος συντελεστής συσχέτισης μεταξύ των δύο διανυσμάτων:

$$C_s^e = \frac{\sum_n ((\mathbf{X}_n^L - \langle \mathbf{X}^L \rangle) (\mathbf{X}_n^R - \langle \mathbf{X}^R \rangle))}{\sqrt{\left( \sum_n (\mathbf{X}_n^L - \langle \mathbf{X}^L \rangle)^2 \right) \left( \sum_n (\mathbf{X}_n^R - \langle \mathbf{X}^R \rangle)^2 \right)}} \quad (1.39)$$

3. Από τη σύγκριση των ανθρωπομετρικών διαστάσεων του [32] και των μασκών προκύπτουν οι τιμές  $C^e, C^m, C^b$ . Ο τρόπος υπολογισμού των  $C_1^e, C_2^e, C_1^m$  περιγράφεται στον πίνακα 1.11.

Περιγραφή	Μεταβλητή
Διαενδοκανθιακό Εύρος	$D_7^a$
Διαεξωκανθιακό Εύρος	$D_5^a$
Διαχειλικό Εύρος	$D_{10}^a$
$\frac{1}{2}(D_5^a - D_7^a)$	$D_{ew}^a$
Θέσεις/αποστάσεις οφθαλμών	$C_1^e = 1 -  D_5^{an} - D_5^n /D_5^{an}$
Οφθαλμικό Έυρος	$C_2^e = 1 -  D_{ew}^{an} - D_{ew}^n /D_{ew}^{an}$
Στοματικό Έυρος	$C_1^m = 1 -  D_{10}^{an} - D_{10}^n /D_{10}^{an}$
$D_i^{an} = D_x^a/D_7^a$ ο δείκτης $a$ δηλώνει ότι η αντίστοιχη απόσταση $i$ προκύπτει από το [32] και ο δείκτης $n$ ότι η τιμή έχει κανονικοποιηθεί διαιρώντας με την $D_7^a$	

**Πίνακας 1.11:** Ανθρωπομετρική Αξιολόγηση [32] για τη θέση και τις διαστάσεις του στόματος και των οφθαλμών.

Οι τιμές εμπιστοσύνης των χαρακτηριστικών εκτιμώνται από το μέσο των προηγούμενων κριτηρίων και των τελικών τιμών αξιολόγησης των μασκών ως εξής:

$$C^e = \langle V_f^e, C_1^e, C_2^e, C_s^e, C_{med}^e \rangle \quad (1.40)$$

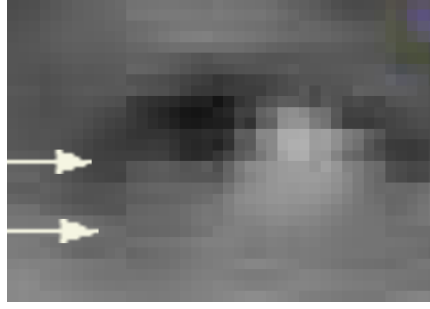
$$C^m = \langle V_f^m, C_1^m \rangle \quad (1.41)$$

$$C^b = C_{med}^b \quad (1.42)$$

$$(1.43)$$

## 1.6 Πειραματικά Αποτελέσματα

Η ανίχνευση και εξαγωγή χαρακτηριστικών του προσώπου μπορεί να θεωρηθεί ως υποκατηγορία του γενικότερου προβλήματος της κατάτμησης εικόνων (image segmentation), δηλαδή κατάτμηση εικόνων στα χαρακτηριστικά του προσώπου. Σύμφωνα με το Zhang [58] οι αλγόριθμοι κατάτμησης μπορούν να αξιολογηθούν με δύο τρόπους: αναλυτικά ή εμπειρικά. Οι αναλυτικές μέθοδοι εξετάζουν τους ίδιους τους χρησιμοποιούμενους αλγόριθμους εξετάζοντας τις αρχές, τις απαιτήσεις, τις σκοπιμότητες, την πολυπλοκότητα κ.ο.κ. των ίδιων των αλγορίθμων. Παρότι αυτές οι μέθοδοι παρέχουν μια αξιολόγηση των αλγορίθμων η οποία είναι ανεξάρτητη της υλοποίησης αυτής καθαυτής ή της επιλογής και διάταξης των δεδομένων εισόδου, πολύ λίγες ιδιότητες των αλγορίθμων δύναται να προσδιοριστούν ή είναι πρακτικό να προσδιοριστούν μέσω της αναλυτικής μελέτης. Από την άλλη οι εμπειρικές μέθοδοι μπορούν να χωριστούν σε δύο υποκατηγορίες: τις εμπειρικές μεθόδους «επιτυχίας» (goodness) οι οποίες χρησιμοποιούν κάποιο



Ποιό είναι το κάτω βλέφαρο;

Πραγματικό μέγεθος περίπου  $30 \times 15\text{px}$

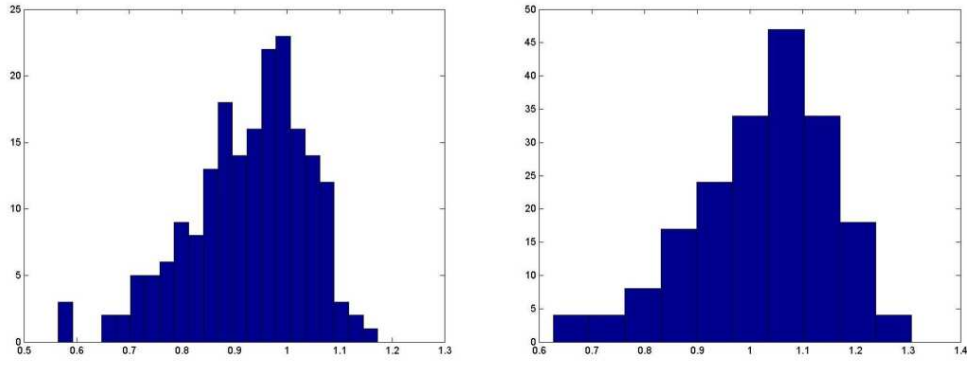
**Σχήμα 1.19:** Σφάλμα χειρωνακτικής κατάτμησης

συγκεκριμένο μέτρο αποτελεσματικότητας για την αξιολόγηση της επίδοσης του αλγόριθμου και τις εμπειρικές μεθόδους «ασυμφωνίας» (discrepancy) οι οποίες μετρούν τις ασυμφωνίες μεταξύ του αποτελέσματος του αυτόματου αλγόριθμου και κάποιας ιδεατά κατατμημένης εικόνας. Ο Zhang εξετάζει μια σειρά από απλούς τρόπους μέτρησης της ασυμφωνίας· εάν θεωρήσουμε την κατάτμηση εικόνας ως μια διαδικασία ταξινόμησης pixels, τότε μόνο ένα κριτήριο είναι εφαρμόσιμο εδώ: ο αριθμός των pixels που έχουν υποστεί λάθος ταξινόμηση σε κάθε χαρακτηριστικό του προσώπου.

Παρόλο που η χειρωνακτική εξαγωγή των χαρακτηριστικών του προσώπου δεν απαιτεί αναγκαστικά ιδιαίτερη εξειδίκευση, είναι φανερό ότι ειδικά σε εικόνες χαμηλής ανάλυσης, η χειρωνακτική κατηγοριοποίηση εισάγει σφάλμα (βλ. εικόνα 1.19). Είναι λοιπόν επιθυμητό να δημιουργηθεί ένας (μικρός) αριθμός από χειρωνακτικές παρατηρήσεις (κατηγοριοποιήσεις/κατατμήσεις) προκειμένου να αξιολογηθεί η διακύμανση μεταξύ τους. Ένας τρόπος αντιστάθμισης που λαμβάνει υπόψη του τη διακύμανση αυτή, είναι ο δείκτης του Williams (Williams' Index -  $WI$ ) [59] ο οποίος συγκρίνει τη συμφωνία μιας κατηγοριοποίησης με την από κοινού συμφωνία των υπολοίπων. Μια επεκταμένη έκδοση του  $WI$  η οποία αντιμετωπίζει την περίπτωση πολλών μεταβλητών μπορεί να βρεθεί στο [60]. Ο τροποποιημένος δείκτης του Williams  $I'$  διαιρεί το μέσο αριθμό συμφωνιών (αντίστροφες ασυμφωνίες,  $D_{j,j'}$ ) μεταξύ του υπολογιστή (παρατηρητής 0) και των  $n - 1$  πραγματικών παρατηρητών ( $j$ ) με τον μέσο αριθμό συμφωνιών μεταξύ των πραγματικών παρατηρητών:

$$WI = \frac{\frac{1}{n} \sum_{j=1}^n \frac{1}{D_{0,j}}}{\frac{2}{n(n-1)} \sum_j \sum_{j':j'>j} \frac{1}{D_{j,j'}}} \quad (1.44)$$

και στην περίπτωσή μας ορίζουμε τη μέση ασυμφωνία μεταξύ των δύο παρατη-



(α) μέσος WI στα μάτια και στο στόμα

(β) μέσος WI στα φρύδια

Σχήμα 1.20: Κατανομή Williams Index

ρητών  $j, j'$  ως:

$$D_{j,j'} = \frac{1}{D_{bp}} \|M_j^x \vee M_{j'}^x\| \quad (1.45)$$

Από ένα σύνολο δεδομένων περίπου 50000 καρέ, επιλέχθηκαν στην τύχη περίπου 250 καρέ στα οποία σημαδεύτηκαν χειρωνακτικά τα 19 feature points, από δύο παρατηρητές έκαστο. Ο δείκτης του Williams υπολογίστηκε χρησιμοποιώντας την εξίσωση (1.44) για κάθε χαρακτηριστικό και κάθε καρέ ξεχωριστά. Τιμή του δείκτη ίση με 0 δηλώνει ότι η αυτόματα εξαγόμενη μάσκα (του υπολογιστή-παρατηρητή) είναι απείρως μακριά από αυτή των άλλων. Τιμή μεγαλύτερη της μονάδας δηλώνει ότι η αυτόματα εξαγόμενη μάσκα διαφωνεί λιγότερο με τις χειρωνακτικά εξαγόμενες από ότι αυτές διαφωνούν μεταξύ τους. Η κατανομή του μέσου  $WI$  υπολογισμένη στους δύο οφθαλμούς και στο στόμα για κάθε καρέ βρίσκεται στην εικόνα 1.20(α) ενώ στην εικόνα 1.20(β) βρίσκεται ο μέσος  $WI$  υπολογισμένος για τα δύο φρύδια. Τα αποτελέσματα συνοψίζονται στον πίνακα 1.6. Για τα μάτια και το στόμα, ο  $WI$  υπολογίζεται και για τις τελικές αλλά και για κάθε μία από τις ενδιάμεσες μάσκες. Με  $WI_x$  συμβολίζεται ο  $WI$  για τη μάσκα  $x$  ενώ  $WI_f$  είναι ο  $WI$  της τελικής μάσκας για κάθε χαρακτηριστικό. Ο μέσος  $WI$  υπολογισμένος σε όλα τα καρέ της μάσκας  $x$  συμβολίζεται ως  $\langle WI_x \rangle$ .

### 1.6.1 Σχολιασμός των αποτελεσμάτων

Στην στήλη 7 του πίνακα 1.6 εμφανίζονται τα ποσοστά που δείχνουν ότι η συγχώνευση των масκών οδήγησε σε βελτίωση του  $WI$  ενώ οι στήλες 8 και 9 εμφανίζουν το μέσο  $WI$  σε καρέ όπου το αποτέλεσμα της συγχώνευσης ήταν καλύτερο και χειρότερο από τις αρχικές μάσκες αντίστοιχα. Κάποιος μπορεί να παρασυρθεί στο συμπέρασμα ότι κάποιες μέθοδοι εξαγωγής χαρακτηριστι-

Αλγόρ. <sup>1</sup>	Μάσκα	$\langle WI_x \rangle$	$\langle WI_f \rangle$	$\frac{\langle WI \rangle_f}{\langle WI \rangle_x}$	$\sigma^2$	% των καρέ όπου $WI_f > WI_x$	$\langle WI \rangle$ στα καρέ όπου $WI_f < WI_x$	$\langle WI \rangle$ στα καρέ όπου $WI_x < WI_f$	$\langle WI \rangle$ στα χει- ρό- τερα 5% καρέ <sup>4</sup>
1	2	3	4	5	6	7	8	9	10
Α.Οφθ.									
NN <sup>2</sup>	0	0.677		1.287	0.103	74.2	0.697	0.885	0.351
1.4.1.1	1	0.701		1.216	0.056	78.8	0.731	0.868	0.414
1.4.1.2	2	0.821	0.838	1.029	0.027	82.4	0.770	0.887	0.459
1.4.1.3	3	0.741		1.131	0.057	76.2	0.811	0.847	0.265
1.4.1.4	4	0.870		0.979	0.026	44.3	0.812	0.867	0.427
	f	0.838		1.000					0.475
Δ.Οφθ.									
NN <sup>2</sup>	0	0.800		1.093	0.020	75.2	0.672	0.946	0.411
1.4.1.1	1	0.718		1.243	0.084	81.4	0.674	0.929	0.352
1.4.1.2	2	0.774	0.875	1.140	0.021	58.2	0.836	0.883	0.396
1.4.1.3	3	0.650		1.346	0.028	84.5	0.632	0.920	0.305
1.4.1.4	4	0.893		0.982	0.02	48.4	0.778	0.996	0.418
	f	0.875		1.000					0.429
Στόμα									
1.4.4.1	1	0.763		1.051	0.046	59.2	0.752	0.772	0.288
1.4.4.2	2	0.823	0.780	0.963	0.038	44.8	0.721	0.852	0.345
1.4.4.3	3	0.570		1.446	0.204	96.9	0.510	0.793	0.220
	f	0.780		1.000					0.359
Φρύδια <sup>3</sup>									
A		1.034							
Δ		1.013							

Ως  $WI_x$  συμβολίζεται ο  $WI$  για τη μάσκα  $x$  και ως  $WI_f$  ο  $WI$  για την τελική μάσκα κάθε χαρακτηριστικού. Με  $\langle \cdot \rangle$  συμβολίζεται ο μέσος όρος υπολογισμένος σε όλα τα χαρακτηριστικά σε όλα τα καρέ, με  $\langle \cdot \rangle_f$  ο μέσος όρος υπολογισμένος σε όλες τις τελικές μάσκες σε όλα τα καρέ, ενώ με  $\langle \cdot \rangle_x$  συμβολίζεται ο μέσος όρος υπολογισμένος για τη μάσκα  $x$  σε όλα τα καρέ.

<sup>1</sup>Βλέπε αντίστοιχη ενότητα

<sup>2</sup>Οφθαλμική μάσκα  $\mathbf{M}_{nn}^e$ , προκύπτουσα απευθείας από την έξοδο του ν.δ.

<sup>3</sup>Χρησιμοποιώντας τη μάσκα φρυδιών  $\mathbf{M}_{E_2}^b$ , πριν τη λέπτυνση

<sup>4</sup> $\langle WI \rangle$  στο 5% του συνόλου των καρέ με το μικρότερο  $WI$ .

Πίνακας 1.12: Σύνοψη Αποτελεσμάτων

κών έχουν καλύτερα αποτελέσματα από το αποτέλεσμα της συγχώνευσης· αυτό ίσως να συμπεραίνεται εξετάζοντας μόνο τις μέσες τιμές αλλά δεν ισχύει εάν εξεταστούν τα μεμονωμένα καρέ: οι διαφορετικές μέθοδοι αποδίδουν καλύτερα για διαφορετικές εικόνες εισόδου και οι μέσοι όροι των αποτελεσμάτων που μοιάζει να ευνοούν κάποιες μεθόδους έναντι των άλλων εξαρτώνται από την επιλογή και την κατανομή του είδους των εικόνων εισόδου. Αυτό γίνεται φανερό εξετάζοντας την απόκλιση των αποτελεσμάτων μεταξύ του αριστερού και του δεξιού ματιού για την ίδια μάσκα αλλά και από τις τιμές της στήλης 8: η μέση τιμή του δείκτη Williams στα καρέ όπου η μάσκα 4 απέδωσε καλύτερα από το αποτέλεσμα της συγχώνευσης είναι και πάλι λίγο χαμηλότερο από το συνολικό μέσο  $WI$ , επομένως μπορεί να φαίνεται ότι η μάσκα αυτή αποδίδει καλύτερα στα συγκεκριμένα δεδομένα, αλλά η βελτίωση δεν είναι σημαντική. Αυτό σημαίνει ότι ακόμα και αν εξετάσουμε την ίδια ακολουθία, οι μέσες τιμές μπορεί να εμφανίζονται λίγο βελτιωμένες για ένα είδος μάσκας αλλά εάν τα αποτελέσματα στηρίζονταν μόνο σε αυτή τη μάσκα δεν θα παρεχόταν εναλλακτική λύση εάν αυτή η μάσκα δεν απέδιδε καλά. Αυτό αποτυπώνεται στην στήλη 10 όπου είναι έκδηλο ότι όταν χρησιμοποιούνται οι τελικές (συγχωνευμένες) μάσκες, οι χειρότερες περιπτώσεις είναι κατά μέσο όρο καλύτερες από τη χειρότερη επίδοση της μάσκας με τον καλύτερο μέσο δείκτη Williams.

Εκτός αυτού ο στόχος δεν ήταν να βρεθεί ο καλύτερος τρόπος εξαγωγής χαρακτηριστικών, αλλά να βρεθούν ορισμένοι αποδοτικοί τρόποι και να συνδυαστούν με έξυπνο τρόπο σε σχέση και με τον τύπο της εισόδου. Ένα ασφαλές συμπέρασμα που μπορεί να εξαχθεί για τις διάφορες μάσκες είναι ότι εξετάζοντας τις διαφορές τιμών μεταξύ των στηλών 8 και 9, φαίνεται ότι για παράδειγμα η μάσκα 0 αποδίδει καλύτερα σε «πολύ δύσκολες» περιπτώσεις καρέ, όπου ο συνολικός μέσος  $WI$  είναι μόνο 0.69.

Περαιτέρω αποτελέσματα βρίσκονται στις εικόνες 2.3, 2.4, 2.10, 2.11, 2.12 και 2.2 του επόμενου κεφαλαίου. Στις εικόνες 2.10, 2.11 και 2.12 απεικονίζονται συνολικά τέσσερα συνεχή τμήματα της ακολουθίας «rodgy» και στην εικόνα 2.2 ένα συνεχές τμήμα της ακολουθίας «allyssa» όπου σε κάθε καρέ τα 19 χαρακτηριστικά σημεία (feature points) έχουν ευρεθεί εκ νέου, δηλαδή η κάθε εκτίμηση δεν βασίζεται στην προηγούμενη όπως θα γινόταν στην περίπτωση της παρακολούθησης (tracking).

## 1.7 Συμπεράσματα - Επεκτάσεις

Σε αντίθεση με την πλειονότητα των μεθόδων της βιβλιογραφίας που αναφέρονται στηνγια την εξαγωγή των χαρακτηριστικών του προσώπου, που εισάγουν

## *Κεφάλαιο 1. Εξαγωγή των χαρακτηριστικών του προσώπου*

πολλούς περιορισμούς στις πιθανές εφαρμογές, όπως η χρήση εικόνων πολύ υψηλής ανάλυσης, η ανάρτηση μίας ή περισσότερων καμερών στο κεφάλι, η χρήση υπέρυθρων καμερών για την ανίχνευση των οφθαλμών, η προϋπαρξη τρισδιάστατων μοντέλων του κεφαλιού ή η χειρωνακτική αρχικοποίηση των αλγορίθμων εξαγωγής των χαρακτηριστικών, η τεχνική που εξετάσθηκε στο κεφάλαιο αυτό μπορεί να χρησιμοποιηθεί ακόμα και σε πραγματικές ακολουθίες που περιέχουν εικόνες προσώπου χαμηλής ανάλυσης με πλήθος ποιοτικών χαρακτηριστικών, όπως η κακή ποιότητα χρώματος ή η ύπαρξη υπερφωτισμένων περιοχών (ψαλιδισμός).

Η τεχνική προτάθηκε στο κεφάλαιο αυτό μπορεί να χρησιμοποιηθεί εκτός της αναγνώρισης εκφράσεων και σε εφαρμογές όπως είναι η ταυτοποίηση προσώπων ή η τηλεδιάσκεψη με πολύ χαμηλούς ρυθμούς μετάδοσης όπου θα μεταδίδονται μόνο οι παράμετροι κάποιου μοντέλου προσώπου. Επιπλέον η τεχνική αυτή μπορεί να χρησιμοποιηθεί ως πρώτο στάδιο για την αρχικοποίηση αλγορίθμων παρακολούθησης σημείων, όπως είναι η ευρέως χρησιμοποιούμενη [44], ή ακόμη και την αρχική ακριβή προσαρμογή αναλυτικών δισδιάστατων ή τρισδιάστατων μοντέλων προσώπου. Επιπλέον μπορεί να χρησιμοποιηθεί για την αρχικοποίηση παραμορφούμενων προτύπων [97], των οποίων η αρχικοποίηση αποτελεί βασικό περιορισμό. Η χρήση της παρακολούθησης των εξαγόμενων σημείων μπορεί να επιταχύνει σημαντικά τη διαδικασία, οπότε η εκ νέου εκτίμηση της θέσης των σημείων μπορεί να επιτελείται όταν εκτιμάται ότι η παρακολούθηση απέτυχε, αλλά και να προσφέρει εξομάλυνση των μικρών διακυμάνσεων της θέσης των 19 σημείων που προκαλούνται από την μεταβολή των χαρακτηριστικών της εικόνας και τον θόρυβο. Επιπλέον η αρχικοποίηση 3-Δ μοντέλου προσώπου θα προσδώσει στο σύστημα μεγαλύτερη ευρωστία σε μεγάλες κλίσεις του προσώπου ιδιαίτερα όταν υπάρχει και απόκρυψη σημείων, αν και η περίπτωση αυτή δεν έχει ιδιαίτερο νόημα για την εκτίμηση εκφράσεων.

□



## Κεφάλαιο 2

# Εκτίμηση Συναισθημάτων με Βάση τις Εκφράσεις του Προσώπου

### 2.1 Η μέθοδος πρόκλησης συναισθημάτων SAL

Η μέθοδος SAL (Sensitive Artificial Listener) που αναπτύχθηκε στο πλαίσιο των προγραμμάτων ERMIS και HUMAINE [29][30], αποτελεί την επέκταση του συστήματος ELIZA [57] που δημιουργήθηκε ως αποτέλεσμα έρευνας στο πεδίο της τεχνητής νοημοσύνης από τον Weizenbaum τη δεκαετία του 60. Το σύστημα ELIZA προσομοιώνει τη μέθοδο θεραπείας του Carl Rogers κατά την οποία οι ασθενείς αναφέρουν τα προβλήματά τους σε έναν ακροατή, ο οποίος απαντά κατάλληλα με στόχο να προκαλέσει την συνέχιση της συζήτησης χωρίς όμως να εκφέρει κρίση ή σχόλια. Στις πρώτες υλοποιήσεις της ιδέας του Weizenbaum οι χρήστες αλληλεπιδρούσαν με το σύστημα χρησιμοποιώντας μηνύματα κειμένου μέσω του πληκτρολογίου και ο υπολογιστής απαντούσε με μηνύματα κειμένου στην οθόνη. Το πρόγραμμα χρησιμοποιούσε ελάχιστους λεκτικούς και γραμματικούς κανόνες ώστε να αναγνωρίζει συγκεκριμένες λέξεις-κλειδιά και επαναχρησιμοποιούσε τμήματα φράσεων που εισήγαγε ο χρήστης στις ερωτήσεις του. Αργότερα το σύστημα εμπλουτίστηκε με σύνθεση φωνής και συνθετικά μοντέλα προσώπων. Στην πλευρά της εισόδου, το πληκτρολόγιο παραμένει η βασική συσκευή εισαγωγής πληροφορίας εφόσον η αυτόματη αναγνώριση φυσικού λόγου δεν έχει ωριμάσει ακόμα σε επίπεδο που να είναι ικανή να αναγνωρίζει συναισθηματικά φορτισμένη ομιλία.

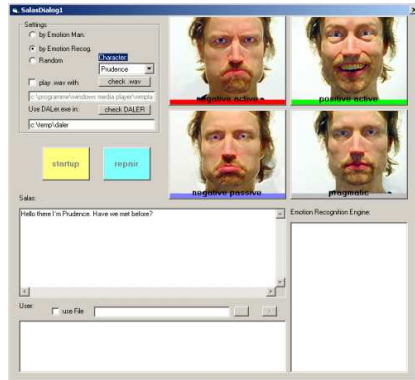
Ο τρόπος καταγραφής αποτελεί εγγενώς προβληματική περιοχή αυτής της μεθόδου. Λαμβάνοντας υπόψη ότι πρέπει να καταγράφεται και η φωνή αλλά και

η εικόνα, είναι φανερό η ανάγκη συμβιβασμού μεταξύ των απαιτήσεων ποιότητας εγγραφής που ορίζει η επεξεργασία σήματος και της μη-παρεμβολής που απαιτεί η ψυχολογία. Εάν κάποιος είναι πολύ προσεκτικός με την ποιότητα εγγραφής, οι χρήστες θα αισθανθούν περιορισμένοι και είναι μάλλον απίθανο να εμφανίσουν την επιθυμούμενη καθημερινή ήπια συναισθηματική έκφραση. Από την άλλη, οι αλγόριθμοι ανάλυσης εικόνων και ομιλίας, δεν αναμένεται να λειτουργήσουν ικανοποιητικά σε παρουσία αυθόρμητων κινήσεων του κεφαλιού και των χεριών, μειωμένο φωτισμό και ήπιας μουσικής υπόκρουσης. Επιπλέον θέματα ανακύπτουν από τις αλληλοσυγκρουόμενες απαιτήσεις των διαφορετικών μέσων: ενώ τα μικρόφωνα που αναρτώνται στο κεφάλι μπορεί να προτιμώνται στην ανάλυση ομιλίας, έχουν καταστρεπτικές συνέπειες στην αυτόματη εξαγωγή χαρακτηριστικών του προσώπου.

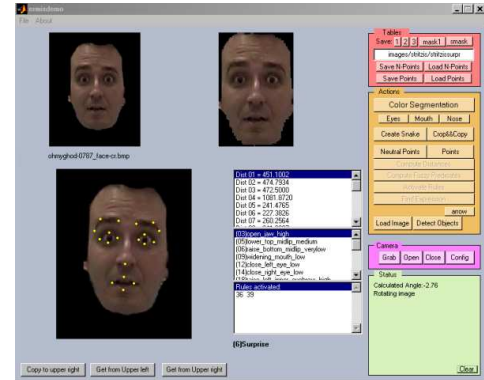
Η υλοποίηση του SAL αποτελείται κυρίως από μια εφαρμογή λογισμικού η οποία σχεδιάστηκε ώστε να επιτρέπει στους χρήστες να μεταπηδούν στις διάφορες συναισθηματικές καταστάσεις. Περιέχει τέσσερις «προσωπικότητες» (εικ 2.1α) οι οποίες ακούν το χρήστη και απαντούν σε αυτό που λέει ανάλογα με τα συναισθηματικά χαρακτηριστικά που διαθέτει η κάθεμία. Αυτό που προσδιορίζει τις προσωπικότητες αυτές είναι ότι κάθε μία προσπαθεί να οδηγήσει το χρήστη προς μία κατάσταση-στόχο:

- Η Poppy είναι χαρούμενη και προσπαθεί να οδηγήσει το χρήστη προς μία θετική συναισθηματική κατάσταση.
- Ο Spike είναι επιθετικός και προσπαθεί να κάνει το χρήστη να υιοθετήσει μία οργισμένη, διενεκτική διάθεση.
- Ο Obadiah καταθλιπτικός και προσπαθεί να οδηγήσει το χρήστη προς μία παθητική - πεσιμιστική στάση.
- Η Prudence είναι πραγματίστρια, και προσπαθεί να περάσει στο χρήστη προς μία ρεαλιστική νοοτροπία.

Ο πυρήνας του συστήματος αποτελείται από μια σειρά απαντήσεων σχεδιασμένες να χρησιμοποιηθούν όταν ο χρήστης βρίσκεται σε συγκεκριμένη συναισθηματική κατάσταση και έχει επιλέξει συγκεκριμένη «προσωπικότητα» προς συνομιλία. Η αρχική καταγραφή αφορούσε 20 χρήστες με τους οποίους παρήχθησαν περίπου 200 λεπτά δεδομένων. Η δεύτερη σειρά καταγραφών αφορούσε τέσσερις χρήστες οι οποίοι συμμετείχαν σε δύο συνεδρίες έκαστος οπότε παρήχθησαν 160 λεπτά δεδομένων, δηλαδή συνολικά 360 λεπτά. Και οι δύο σειρές δεδομένων περιείχαν αναλογία φύλων 50/50. Τα δεδομένα που παρήχθησαν χρησιμοποιήθηκαν για



(α) Διαπροσωπεία SAL



(β) Διαπροσωπεία ανάλυσης έκφρασης (MATLAB)

Σχήμα 2.1: Διαπροσωπείες

την εξαγωγή χαρακτηριστικών του προσώπου και την εκτίμηση συναισθήματος με τις μεθόδους που παρουσιάζονται στην παρούσα διατριβή. Στην εικόνα 2.1(β) απεικονίζεται η διαπροσωπεία της αυτόματης εκτίμησης εκφράσεων (MatLab).

## 2.2 Αυτόματη Εκτίμηση Εκφράσεων

Μια επισκόπηση της διαδικασίας αναγνώρισης έκφρασης παρουσιάζεται στο σχήμα 1.1: αρχικά επιλέγονται 19 χαρακτηριστικά σημεία (Feature Points - FP) από τις αντίστοιχες μάσκες χαρακτηριστικών. Τα FP αυτά πρέπει να συγκριθούν με τα αντίστοιχα που προκύπτουν από το καρέ ουδέτερης έκφρασης, προκειμένου να υπολογιστεί η μετατόπισή τους και να υπολογιστούν τα αντίστοιχα Facial Animation Parameters (FAPs). Τα FAPs εν συνεχεία χρησιμοποιούνται για την αξιολόγηση των προφίλ εκφράσεων τα οποία παρέχουν την εκτιμώμενη έκφραση.

### 2.2.1 Από τις μάσκες στα χαρακτηριστικά σημεία

Οι συντεταγμένες των ακραίων αριστερών, δεξιών, άνω και κάτω ορίων των τελικών μασκών  $M_f^{eL}, M_f^{eR}, M_f^m$ , των αριστερών δεξιών και άνω ορίων των  $M_f^{bL}, M_f^{bR}$ , καθώς και οι συντεταγμένες του ρινικού κέντρου επιλέγονται για τον ορισμό των 19 FPs που απεικονίζονται στον πίνακα 2.1 και τις εικόνες 2.3, 2.4, 2.10, 2.11.

Στο χαρακτηριστικό σημεία  $x$  εκχωρείται η τιμή εμπιστοσύνης  $C_x^{FP}$  κληρονομώντας την αντίστοιχη τιμή ( $C^e, C^m, C^b, C^n$ ) της τελικής μάσκας από την οποία προήλθε.

FP #	MPEG-4 FP	όνομα FP
01	4.5	Outer point of Left eyebrow
02	4.3	Middle point of Left eyebrow
03	4.1	Inner point of Left eyebrow
04	4.6	Outer point of Right eyebrow
05	4.4	Middle point of Right eyebrow
06	4.2	Inner point of Right eyebrow
07	3.7	Outer point of Left eye
08	3.11	Inner point of Left eye
09	3.13	Upper point of Left eyelid
10	3.9	Lower point of Left eyelid
11	3.12	Outer point of Right eye
12	3.8	Inner point of Right eye
13	3.14	Upper point of Right eyelid
14	3.10	Lower point of Right eyelid
15	9.15	Nose point
16	8.3	Left corner of mouth
17	8.4	Right corner of mouth
18	8.1	Upper point of mouth
19	8.2	Lower point of mouth

**Πίνακας 2.1:** Τα χρησιμοποιούμενα χαρακτηριστικά σημεία (Feature Points) [5]

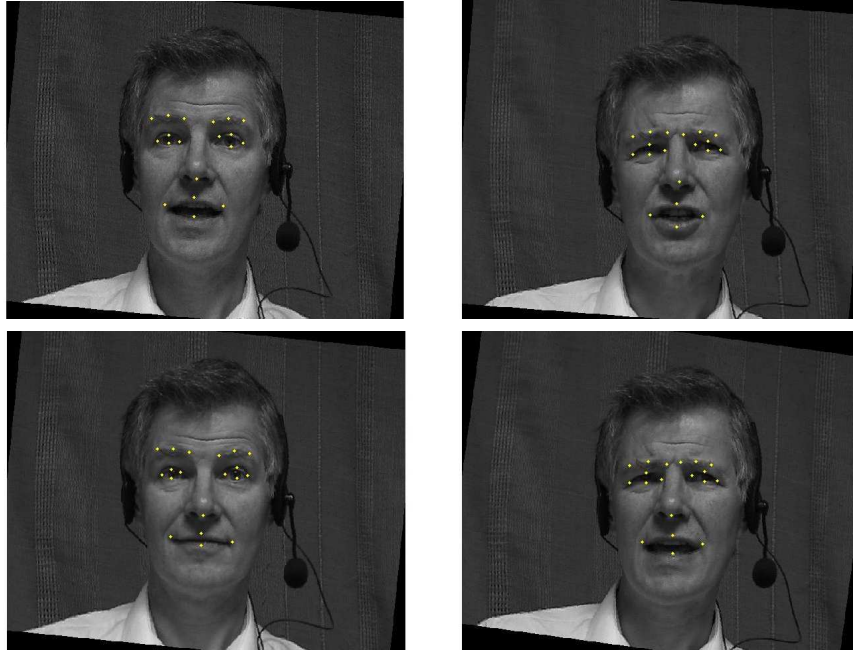


Σχήμα 2.2: Εύρεση των 19 χαρακτηριστικών σημείων σε συνεχόμενα καρέ (χωρίς tracking). Διαχωρική απόσταση: 40px



Έχει επιτελεστεί αυτόματη διόρθωση της κλίσης

Σχήμα 2.3: Τα 19 εντοπισμένα χαρακτηριστικά σημεία σε ποικιλία εκφράσεων



Έχει επιτελεστεί αυτόματη διόρθωση της κλίσης

Σχήμα 2.4: Τα 19 εντοπισμένα χαρακτηριστικά σημεία σε ποικιλία εκφράσεων

### 2.2.2 Από τα χαρακτηριστικά σημεία στις παραμέτρους κίνησης χαρακτηριστικών σημείων (FAP)

Τα σύνολα των Facial Definition Parameters (FDP) και των Facial Animation Parameters (FAP) σχεδιάστηκαν στο πλαίσιο του MPEG-4 με στόχο τη δυνατότητα ορισμού σχημάτων προσώπου και υφής για τη σύνθεση των εκφράσεων και της ομιλίας. Λειτουργώντας αντίστροφα, είναι δυνατόν να εκτιμηθούν οι τιμές των FAP μέσω της ανάλυσης της κίνησης ορισμένων χαρακτηριστικών σημείων (FP), τα οποία αποτελούν υποσύνολο των FDP.

Ο ποσοτικός προσδιορισμός των FAPs μπορεί να επιτευχθεί υπολογίζοντας το 25-διάστατο διάνυσμα αποστάσεων  $D_v$ , το οποίο περιέχει τις κάθετες και οριζόντιες αποστάσεις μεταξύ των 19 FP και οι οποίες απεικονίζονται στο σχήμα 2.5. Οι μονάδες των αποστάσεων δεν είναι pixels, αλλά οι κανονικοποιημένες μονάδες MPEG-4 που είναι ανεξάρτητες της κλίμακας: ENS, MNS, MW, IRISD and ES [5] οι οποίες αντιστοιχούν στις αποστάσεις του σχήματος 2.6. Οι βάσεις των μονάδων εκτιμώνται απευθείας από τις αποστάσεις των αντίστοιχων FP στο καρέ ουδέτερης έκφρασης. Π.χ. η μονάδα ES υπολογίζεται ως  $|FP_9, FP_{13}|$ , ενώ με  $FP_{em}$  συμβολίζεται το ενδιάμεσο σημείο μεταξύ των οφθαλμών.

Αρχικά κατασκευάζεται ένα διάνυσμα αποστάσεων για το καρέ ουδέτερης έκφρασης ( $D_v^n$ ) και ένα διάνυσμα  $D_v$  για κάθε ένα από τα υπόλοιπα καρέ. Τα FAP υπολογίζονται σε κάθε καρέ με τη σύγκριση των δύο διανυσμάτων από-



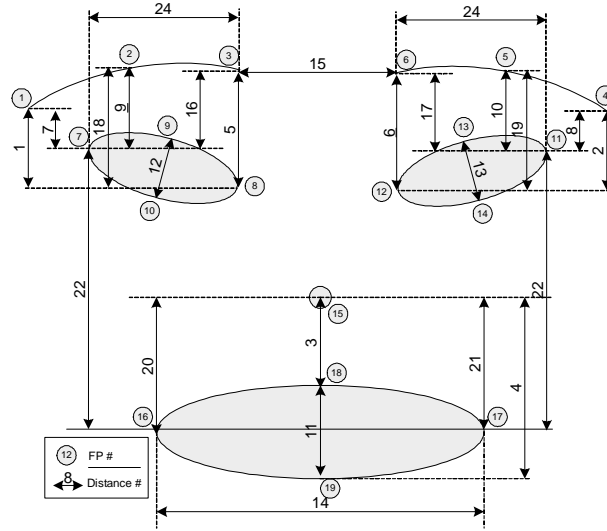
Συμβολισμός FAP	Όνομα FAP	Απόσταση (α.α.)
F <sub>3</sub>	open_jaw	11
F <sub>4</sub>	lower_top_midlip	3
F <sub>5</sub>	raise_bottom_midlip	4
F <sub>6</sub> +F <sub>7</sub>	widening_mouth	14
F <sub>19</sub> +F <sub>21</sub>	close_left_eye	12
F <sub>20</sub> +F <sub>22</sub>	close_right_eye	13
F <sub>31</sub>	raise_left_inner_eyebrow	5,16
F <sub>32</sub>	raise_right_inner_eyebrow	6,17
F <sub>33</sub>	raise_left_medium_eyebrow	18,9
F <sub>34</sub>	raise_right_medium_eyebrow	19,10
F <sub>35</sub>	raise_left_outer_eyebrow	7,1
F <sub>36</sub>	raise_right_outer_eyebrow	8,2
F <sub>37</sub>	squeeze_left_eyebrow	24
F <sub>38</sub>	squeeze_right_eyebrow	25
F <sub>37</sub> +F <sub>38</sub>	squeeze_eyebrows	15
F <sub>59</sub>	raise_left_outer_cornerlip	22
F <sub>60</sub>	raise_right_outer_cornerlip	23

**Πίνακας 2.2:** Παραδείγματα FAP και αντίστοιχων αποστάσεων

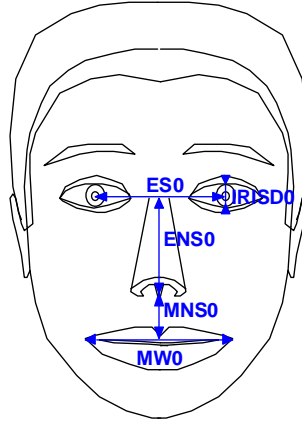
στασης,  $D_v^n$  και  $D_v$ . Κάθε FAP εξαρτάται από ένα ή περισσότερα στοιχεία του διανύσματος απόστασης, κατά συνέπεια κάποια FAP είναι υπερ-ορισμένα. Ο λόγος υπολογισμού των FAP  $F_{31} \cdots F_{36}$  με χρήση περισσότερων αποστάσεων από τις απαραίτητες είναι η αύξηση της ευρωστίας του υπολογισμού η οποία επιτυγχάνεται λαμβάνοντας υπόψη τις τιμές εμπιστοσύνης κάθε απόστασης. Τα στοιχεία του διανύσματος  $D_v$  υπολογίζονται με τη μέτρηση των αποστάσεων του σχήματος 2.5. Η αβεβαιότητα που εμπεριέχεται στις συντεταγμένες των FP αντανακλάται στα αντίστοιχα FAP, ως εκ τούτου οι χρησιμοποιούμενες πλεονάζουσες αποστάσεις για τον υπολογισμό των FAP σταθμίζονται σύμφωνα με την εμπιστοσύνη  $C^x$  της αντίστοιχης μάσκας από την οποία προέρχονται.

Η τιμή εμπιστοσύνης του FAP  $i$  συμβολίζεται με  $C_i^{FAP}$  και υπολογίζεται ως  $C_i^{FAP} = \langle C_Y^{FP} \rangle$ ,  $\mathbf{Y}$ : σύνολο των FPs που χρησιμοποιήθηκαν στην εκτίμηση του FAP  $i$ . Οι αντιστοιχίσεις μεταξύ των FAP και των αντίστοιχων στοιχείων του διανύσματος αποστάσεων φαίνονται στον πίνακα 2.2.





Σχήμα 2.5: Αποστάσεις υπολογισμού χαρακτηριστικών σημείων



$$\begin{aligned}
 FP_{em} &= \frac{1}{2}(FP_7 + FP_{11}) \\
 ES &= |FP_9, FP_{13}|/1024 \\
 ENS &= |FP_{em}, FP_{15}|/1024 \\
 MW &= |FP_{16}, p_{17}|/1024 \\
 MNS &= |FP_{15}, FP_{18}|/1024 \\
 IRISD &= 0.2ES
 \end{aligned}$$

Σχήμα 2.6: Μονάδες ISO MPEG-4

### 2.2.3 Εκφράσεις Προσώπου και Επικοινωνία Ανθρώπου - Μηχανής

Σε προγενέστερη έρευνα για την αναγνώριση εκφράσεων, δημιουργήθηκε ένα σύστημα βασισμένο σε κανόνες οι οποίοι χαρακτηρίζουν τη συναισθηματική κατάσταση του χρήστη σε σχέση με τις έξι καθολικές ή αρχετυπικές εκφράσεις, χαρά, έκπληξη, φόβο, θυμό, αποστροφή, λύπη. Για κάθε μία από αυτές εκφράσεις δημιουργήθηκαν 41 κανόνες βασισμένοι στα MPEG-4 FAPs [6], οι οποίοι προέκυψαν από την ανάλυση των FAPs των εκφράσεων των προσώπων της βάσης δεδομένων του Ekman [6]. Το σύνολο δεδομένων του Ekman περιέχει αρκετές εικόνες-παραλλαγές για κάθε μια από τις έξι αρχετυπικές εκφράσεις, οι οποίες είναι όμως μάλλον υπερβολικά έντονες. Ως αποτέλεσμα, οι κανόνες που εξήχθησαν από αυτά τα δεδομένα δεν αποδίδουν καλά εάν χρησιμοποιηθούν σε

πραγματικά διαδραστικά περιβάλλοντα ανθρώπου-μηχανής.

### 2.2.3.1 Τροχός Συναισθημάτων της Whissel

Ένα από τα πιο σημαντικά στοιχεία στην ανάλυση των εκφράσεων είναι η εύρεση χαρακτηρισμού για τις διάφορες συναισθηματικές καταστάσεις. Δεδομένου όμως ότι ο ίδιος ο άνθρωπος χαρακτηρίζει τις εκφράσεις κατά τρόπο αρκετά σύνθετο, ένας απλός χαρακτηρισμός δεν είναι αρκετός, επιβάλλεται η εύρεση μίας συνεχούς αναπαράστασης των συναισθημάτων που θα προσεγγίζει όσο το δυνατόν περισσότερο την ανθρώπινη αντίληψη επί του θέματος. Οι ψυχολόγοι έχουν εξετάσει ένα ευρύ φάσμα συναισθημάτων, αλλά πολύ λίγες από αυτές τις μελέτες έχουν αποτελέσματα αξιοποιήσιμα στο χώρο της γραφικής και της υπολογιστικής όρασης. Η Whissel [52] προσέγγισε το θέμα με το επίπεδο activation-emotion ("τροχός συναισθημάτων"), το οποίο αποτελεί μία αναπαράσταση, απλή αλλά συγχρόνως ικανή να «περιγράψει» ένα ευρύ φάσμα συναισθημάτων, η οποία βασίζεται σε δύο μεγέθη:

- Valence: Το κοινό χαρακτηριστικό των διαφόρων συναισθηματικών καταστάσεων αποτελεί το γεγονός ότι είναι αποτέλεσμα της επίδρασης που έχουν στο άτομο οι συνάνθρωποί του ή οι διάφορες καταστάσεις που βιώνει. Η επίδραση αυτή είναι είτε θετική είτε αρνητική.
- Activation level (επίπεδο δραστηριότητας): Ερευνητές έχουν επισημάνει το γεγονός ότι το άτομο, βιώνοντας τις διάφορες συναισθηματικές καταστάσεις, παρουσιάζει τάση συγκεκριμένων αντιδράσεων. Βασιζόμενοι στην παρατήρηση αυτή έγινε προσπάθεια να χαρακτηριστεί η έκφραση με βάση την "ενεργοποίηση" του προσώπου.

Οι άξονες στον «τροχό συναισθημάτων» αντικατοπτρίζουν τα δύο αυτά μεγέθη. Ο κατακόρυφος άξονας είναι ο άξονας της «ενεργοποίησης» (activation), ενώ ο οριζόντιος άξονας είναι ο άξονας της «αποτίμησης» του συναισθήματος (evaluation).

Ως επέκταση της ιδέας του τροχού μπορούμε να θεωρήσουμε τα βασικά συναισθήματα ως σημεία στην περιφέρεια του κύκλου [64](εικ. 2.7). Επομένως γίνεται φανερό η ανάγκη για δημιουργία κανόνων που να περιγράφουν τα τρία πρώτα τεταρτημόρια του τροχού συναισθημάτων - το τέταρτο στερείται συναισθημάτων. Για να επιτευχθεί αυτό, οι κινήσεις των μυών του προσώπου μεταφράστηκαν σε FAP ενώ τα σύνολα FAP της κάθε έκφρασης σε κάθε τεταρτημόριο επαληθεύτηκαν πειραματικά. Στη συνέχεια τα διαστήματα τιμών του κάθε FAP υπολογίστηκαν αναλύοντας πραγματικούς διαλόγους και ακολουθίες



Σχήμα 2.7: Ο χώρος ενεργοποίησης-συναισθήματος

βίντεο. Στον πίνακα 2.3 περιέχονται τρία παραδείγματα κανόνων που δημιουργήθηκαν με την παραπάνω μέθοδο.

Κανόνας	Τεταρτημόριο
$F_6 \in [160, 240]$ , $F_7 \in [160, 240]$ , $F_{12} \in [260, 340]$ , $F_{13} \in [260, 340]$ , $F_{19} \in [-449, -325]$ , $F_{20} \in [-426, -302]$ , $F_{21} \in [325, 449]$ , $F_{22} \in [302, 426]$ , $F_{33} \in [70, 130]$ , $F_{34} \in [70, 130]$ , $F_{41} \in [130, 170]$ , $F_{42} \in [130, 170]$ , $F_{53} \in [160, 240]$ , $F_{54} \in [160, 240]$	(++)
$F_{16} \in [45, 155]$ , $F_{18} \in [45, 155]$ , $F_{19} \in [-330, -200]$ , $F_{20} \in [-330, -200]$ , $F_{31} \in [-200, -80]$ , $F_{32} \in [-194, -74]$ , $F_{33} \in [-190, -70]$ , $F_{34} \in [-190, -70]$ , $F_{37} \in [65, 135]$ , $F_{38} \in [65, 135]$	(-+)
$F_3 \in [400, 560]$ , $F_5 \in [-240, -160]$ , $F_{19} \in [-630, -570]$ , $F_{20} \in [-630, -570]$ , $F_{21} \in [-630, -570]$ , $F_{22} \in [-630, -570]$ , $F_{31} \in [460, 540]$ , $F_{32} \in [460, 540]$ , $F_{33} \in [360, 440]$ , $F_{34} \in [360, 440]$ , $F_{35} \in [260, 340]$ , $F_{36} \in [260, 340]$ , $F_{37} \in [60, 140]$ , $F_{38} \in [60, 140]$	(-+)

Πίνακας 2.3: Παραδείγματα κανόνων και αντίστοιχα διαστήματα τιμών FAP σε μονάδες MPEG-4

## 2.3 Δυνατοτική Αποτίμηση Ασαφών Κανόνων

### 2.3.1 Εισαγωγή

Οι ασαφείς κανόνες και τα συστήματα που βασίζονται σε αυτούς έχουν χρησιμοποιηθεί εκτενώς στο παρελθόν για την ανάπτυξη έμπειρων συστημάτων, καθώς

προσφέρουν μια αναπαράσταση γνώσης που βοηθά στην αποτύπωση της εμπειρίας και της γνώσης που χρησιμοποιείται για την καθοδήγηση ενός ευφυούς συστήματος. Το αποτέλεσμα είναι συστήματα ικανά να επεξεργαστούν περίπλοκα δεδομένα σε πολύ λίγο χρόνο και να αντιδρούν κατάλληλα χρησιμοποιώντας την ενσωματωμένη εμπειρία τους. Οι λαμβανόμενες μετρήσεις αποτυπώνονται σε λεκτικές μεταβλητές υψηλού επιπέδου, οι οποίες με τη σειρά τους χρησιμοποιούνται για να επιτευχθεί μια πολύ γρήγορη προσέγγιση της ιδανικής απόκρισης. Οι βασικές απαιτήσεις των συστημάτων αυτών είναι:

- η ιδανική απόκριση να είναι συνεχής συνάρτηση των παραμέτρων εισόδου
- η ορθότητα της γνώσης που περιέχεται στους κανόνες
- η διαθεσιμότητα των εισόδων

Αν και αυτές μοιάζουν ίσως με απαιτήσεις που ικανοποιούνται εύκολα, εντούτοις υπάρχουν περιπτώσεις στις οποίες οι δύο από αυτές τις υποθέσεις δεν ισχύουν απαραίτητα και όπου όμως η χρήση συστημάτων ασαφών σχέσεων είναι επιθυμητή. Στην περίπτωση μας θέλουμε να χρησιμοποιήσουμε ένα ασαφές σύστημα κανόνων για την εκτίμηση του συναισθήματος ενός καρέ χρησιμοποιώντας ως είσοδο τα αναγνωρισμένα FAPs. Η διαθεσιμότητα των FAP όμως εξαρτάται από τη λειτουργία του συστήματος εξαγωγής χαρακτηριστικών, συνεπώς είναι πιθανό για ορισμένα καρέ είτε να μην είναι διαθέσιμα ως μεταβλητές εισόδου, ή να είναι διαθέσιμα με πολύ χαμηλό βαθμό εμπιστοσύνης.

Για την αντιμετώπιση του ενδεχομένου ελλιπούς ή αβέβαιης εισόδου, σε αυτή την ενότητα προτείνεται ένα δυνατοτικό μοντέλο αποτίμησης ασαφών κανόνων.

Προκειμένου να χρησιμοποιηθούν αυτοί οι κανόνες σε ένα σύστημα που καλείται να αντιμετωπίσει το συνεχές χώρο συναισθήματος-δραστηριότητας, αντικαταστάθηκαν τα διαστήματα τιμών των FAPs με τα ασαφή διαστήματα *high*, *medium*, *low* μετά από την κανονικοποίησή τους. Ένα παράδειγμα των αρχικών κανόνων δίνεται στον πίνακα 2.4, ενώ το πλήρες σύνολο των κανόνων μπορεί να βρεθεί στον πίνακα 2.7 και στο [53].

Για την εκμετάλλευση της γνώσης που περιέχεται την βάση ασαφών κανόνων και την πληροφορία που εξάγεται από κάθε καρέ με τη μορφή εκτιμήσεων των τιμών των FAP, ανακύπτουν τα εξής ζητήματα:

- Οι βαθμοί ενεργοποίησης των FAP πρέπει να ληφθούν υπόψη κατά την εκτίμηση του συνολικού αποτελέσματος.
- Πρέπει να αντιμετωπιστεί η περίπτωση όπου κάποια FAPs δεν έχουν εκτιμηθεί ή έχουν εκτιμηθεί με χαμηλούς βαθμούς εμπιστοσύνης

Κανόνας	Τεταρτημόριο
$F_6 \in [160,240], F_7 \in [160,240], F_{12} \in [260,340],$ $F_{13} \in [260,340], F_{19} \in [-449,-325], F_{20} \in [-426,-302],$ $F_{21} \in [325,449], F_{22} \in [302,426], F_{33} \in [70,130],$ $F_{34} \in [70,130], F_{41} \in [130,170], F_{42} \in [130,170],$ $F_{53} \in [160,240], F_{54} \in [160,240]$	(++)
$F_{16} \in [45,155], F_{18} \in [45,155], F_{19} \in [-330,-200], F_{20} \in [-$ $330,-200], F_{31} \in [-200,-80], F_{32} \in [-194,-74], F_{33} \in [-190,-$ $70], F_{34} \in [-190,-70], F_{37} \in [65,135], F_{38} \in [65,135]$	(-+)
$F_3 \in [400,560], F_5 \in [-240,-160], F_{19} \in [-630,-570],$ $F_{20} \in [-630,-570], F_{21} \in [-630,-570], F_{22} \in [-630,-570],$ $F_{31} \in [460,540], F_{32} \in [460,540], F_{33} \in [360,440],$ $F_{34} \in [360,440], F_{35} \in [260,340], F_{36} \in [260,340],$ $F_{37} \in [60,140], F_{38} \in [60,140]$	(-+)

**Πίνακας 2.4:** Παράδειγμα κανόνων με το αντίστοιχο εύρος τιμών των FAP σε μονάδες MPEG-4

Τα συστήματα ασαφών κανόνων είναι έμπειρα συστήματα που περιέχουν τη γνώση με μορφή κανόνων όπως είναι ο ακόλουθος [54]:

$$IF \ x_1, x_2, \dots, x_n \ THEN \ y \quad (2.1)$$

$$y = t(x_1, x_2, \dots, x_n) \quad (2.2)$$

όπου  $t$  είναι μια ασαφής  $t$ -νόρμα, όπως είναι το ελάχιστο:

$$t(x_1, x_2, \dots, x_n) = \min(x_1, x_2, \dots, x_n) \quad (2.3)$$

το αλγεβρικό γινόμενο:

$$t(x_1, x_2, \dots, x_n) = x_1 \cdot x_2 \cdot \dots \cdot x_n \quad (2.4)$$

το φραγμένο άθροισμα

$$t(x_1, x_2, \dots, x_n) = x_1 + x_2 + \dots + x_n + 1 - n \quad (2.5)$$

κ.ο.κ. Άλλη μια γνωστή μέθοδος αποτίμησης κανόνων περιγράφεται στο [55] η οποία χρησιμοποιεί σταθμισμένα αθροίσματα αντί  $t$ -νόρμας προκειμένου να συνδυάσει πληροφορία προερχόμενη από διαφορετικούς κανόνες:

$$y = w_1x_1 + w_2x_2 + \dots + w_nx_n \quad (2.6)$$

Και οι δύο προσεγγίσεις έχουν εξεταστεί και εδραιωθεί στον τομέα του ασαφούς αυτομάτου ελέγχου. Παραταύτα δεν κρίνονται επαρκείς στην περίπτωση της εκτίμησης εκφράσεων προσώπου: το βασικό τους μειονέκτημα είναι ότι υποθέτουν πως όλοι οι πρόγονοι είναι γνωστοί, δηλαδή ότι όλα τα χαρακτηριστικά έχουν εξαχθεί και μετρηθεί επιτυχώς. Στην προκειμένη περίπτωση όμως, τα FAP μπορούν να έχουν εκτιμηθεί με πολύ χαμηλό βαθμό εμπιστοσύνης ή να μην έχουν εκτιμηθεί καθόλου λόγω χαμηλής ποιότητας εικόνας, απόκρυψης, θορύβου κ.ο.κ. Συνεπώς απαιτείται ένας πιο ευέλικτος τρόπος αποτίμησης των κανόνων ο οποίος να είναι ικανός να ενσωματώσει και τέτοιου είδους αβεβαιότητα. Επιπλέον η δεύτερη από τις συμβατικές προσεγγίσεις που αναφέρονται έχει το επιπλέον μειονέκτημα ότι δύναται να παρέχει υψηλή ενεργοποίηση της εξόδου της ακόμα και στην περίπτωση που απουσιάζει κάποιο σημαντικό πρότερο τμήμα του κανόνα. Στη συγκεκριμένη περίπτωση όμως η δεδομένη μη ενεργοποίηση ενός συγκεκριμένου χαρακτηριστικού του προσώπου μπορεί να είναι αρκετή για να αποκλείσει μια σειρά από πιθανές εκφράσεις με μεγάλη βεβαιότητα.

### 2.3.2 Αποτίμηση των όρων

Στην  $t$ -νόρμα που περιγράφεται στην εξίσωση 2.1, οι όροι εισόδου με τις μικρότερες τιμές επηρεάζουν κατά μέγιστο βαθμό την τελική τιμή  $y$  ενώ όροι εισόδου με τιμές κοντά στη μονάδα έχουν αμελητέα επίδραση στην τελική τιμή του  $y$ . Έχοντας αυτό υπόψη, μπορεί να απαιτηθεί μόνο όροι που έχουν μεγάλη τιμή βεβαιότητας να επιτρέπεται να έχουν μικρές τιμές στην παραπάνω διαδικασία. Πιο συγκεκριμένα απαιτείται ο βαθμός  $k(x)$  κατά τον οποίο λαμβάνεται υπόψη ο όρος εισόδου  $x$  να είναι χαμηλός, δηλαδή το συμπλήρωμά του  $c(k(x))$  είναι υψηλό, μόνο όταν η εμπιστοσύνη  $x^c$  κατά την οποία γνωρίζουμε την τιμή του  $x$  είναι υψηλή και η τιμή του  $x$  είναι χαμηλή. Αυτό μπορεί να εκφραστεί ως εξής:

$$c(k(x)) = t(x^c, c(x)) \quad (2.7)$$

όπου το  $c$  δηλώνει το ασαφές συμπλήρωμα. Εφαρμόζοντας το νόμο του de Morgan βρίσκουμε ότι ο βαθμός κατά τον οποίο θεωρείται ο όρος  $x$  είναι:

$$k(x) = s(c(x^c), x) \quad (2.8)$$

όπου  $s$  είναι μια ασαφής  $s$ -νόρμα. Διαπιστώνεται εύκολα ότι η εξίσωση 2.8 ικανοποιεί τις επιθυμητές οριακές συνθήκες:

- όταν  $x^c \rightarrow 1$  τότε  $c(x^c) \rightarrow 0$  και  $k(x) \rightarrow x$ , δηλαδή ο όρος εισόδου θεωρείται κανονικά.

- όταν  $x^c \rightarrow 0$  τότε  $c(x^c) \rightarrow 1$  και  $k(x) \rightarrow 1$ , δηλαδή δεν επιτρέπεται στον όρο εισόδου να επηρεάσει την αποτίμηση του κανόνα.

Ο τύπος που συνοψίζει όσα προαναφέρθηκαν βασίζεται σε αυτόν της τυπικής προσέγγισης, με τη διαφορά ότι οι όροι εισόδου συμμετέχουν με κατά το βαθμό θεώρησής τους:

$$y = t(k(x_1), k(x_2), \dots, k(x_n)) \quad (2.9)$$

Εύκολα συμπεραίνεται ότι στην περίπτωση που όλοι οι όροι εισόδου είναι γνωστοί με εμπιστοσύνη 1, ο κανόνας θα αποτιμηθεί όπως και στην κλασική περίπτωση. Από την άλλη, όταν όλοι οι όροι είναι γνωστοί με βαθμό εμπιστοσύνης 0, όταν δηλαδή δεν διατίθεται κανενός είδους πληροφορία, ο κανόνας θα αποτιμηθεί με βαθμό ένα. Με αυτή τη μέθοδο, ο βαθμός ενεργοποίησης ενός κανόνα μπορεί να ερμηνευτεί με δυνατοτικό τρόπο, δηλαδή ως ο βαθμός κατά τον οποίο η αντίστοιχη έξοδος είναι δυνατή σύμφωνα με τη διαθέσιμη πληροφορία. Στη βιβλιογραφία ο βαθμός αυτός αναφέρεται ως αληθοφάνεια ή εφικτότητα (plausibility) [54]. Όσο αφορά στην εμπιστοσύνη στην έξοδο του συστήματος, στην κλασική μέθοδο, όπου υποθέτουμε απόλυτη εμπιστοσύνη στην είσοδο, θεωρούμε απλά πως οι βαθμοί ενεργοποίησης είναι σωστοί. Στην επέκταση που παρουσιάζουμε εδώ, όπου δεχόμαστε το ενδεχόμενο η είσοδος να είναι ατελής και αβέβαιη, δεν είναι λογικό να υποθέτουμε πως αυτή η αβεβαιότητα δεν διαδίδεται και στην έξοδο. Αντίθετα, το εκτιμώμενο επίπεδο ενεργοποίησης  $y$  είναι πληροφοριακά πλήρες μόνο όταν συνοδεύεται και από τον αντίστοιχο βαθμό βεβαιότητας  $y^c$ . Η εμπιστοσύνη στην έξοδο υπολογίζεται με βάση την εμπιστοσύνη στους διάφορους όρους της εισόδου ως εξής:

$$y^c = \frac{x_1^c + x_2^c + \dots + x_n^c}{n} \quad (2.10)$$

Ο ορισμός του  $y^c$  με αυτό τον τρόπο έχει το επιθυμητό αποτέλεσμα ότι το  $y^c = 0$  είναι ταυτόσημο με την πλήρη έλλειψη πληροφορίας, εφόσον μπορεί να συμβεί μόνο όταν όλοι οι όροι εισόδου είναι γνωστοί με εμπιστοσύνη μηδέν. Αυτή η ιδιότητα είναι ουσιώδης για το δυνατοτικό συλλογισμό.

Όταν χρησιμοποιούμε τη μέθοδο αποτίμησης της σχέσης (2.2) συχνά απομακρύνουμε την ασάφεια απλά επιλέγοντας την έξοδο του κανόνα με την υψηλότερη ενεργοποίηση και αγνοώντας τους υπόλοιπους κανόνες. Αυτό σημαίνει πως σιωπηρά δίνουμε στους βαθμούς ενεργοποίησης μια πιθανοτική χροιά. Αυτό φαίνεται και από το ότι:

- Ο κανόνας με την υψηλότερη ενεργοποίηση θεωρείται πιο πιθανός από τους άλλους.

- Οι περιπτώσεις στις οποίες δύο διαφορετικοί κανόνες ενεργοποιούνται σε υψηλά και παρόμοια επίπεδα δεν θεωρούνται ξεκάθαρες και οι δύο κανόνες θεωρούνται σχεδόν εξίσου πιθανοί.

Από την άλλη πλευρά αξίζει να σημειωθεί πως η έξοδος ενός συστήματος κανόνων, αν και έχει μια πιθανοτική χροιά, δεν μπορεί να χρησιμοποιηθεί σαν πιθανότητα, καθώς το άθροισμα των ενεργοποιήσεων μη συμβατών κανόνων δύναται να ξεπερνά τη μονάδα. Όμοια, δεν μπορούμε να ισχυριστούμε πως η προτεινόμενη μεθοδολογία αποτίμησης δίνει στην έξοδό της δυνατοτικά μέτρα. Έχει όμως ξεκάθαρα μια δυνατοτική χροιά. Η ενεργοποίηση  $y$  του κανόνα έχει μια δυνατοτική ερμηνεία που αντιστοιχεί στην εφικτότητα του κανόνα.

Προκειμένου να υπάρξει μια πλήρης δυνατοτική αναπαράσταση ενός ενδεχομένου όπως είναι η ισχύς ενός κανόνα, μαζί με την αληθοφάνεια των προφίλ εκφράσεων πρέπει να αποτιμηθεί και το αντίστοιχο μέτρο πίστης (belief), δηλαδή το μέτρο κατά το οποίο οι διαθέσιμες μαρτυρίες υποδεικνύουν ότι το υπό εξέταση προφίλ έκφρασης είναι παρόν στο υπό εξέταση σύνολο εισόδου. Ο βαθμός πίστης θα πρέπει να είναι υψηλός όταν υπάρχει επαρκής πληροφορία κατά την αποτίμηση του κανόνα η οποία να υποδεικνύει ότι ο κανόνας πρέπει να ενεργοποιηθεί. Η ποσότητα της πληροφορίας που ήταν διαθέσιμη κατά την αποτίμηση του κανόνα παρέχεται από την τιμή εμπιστοσύνης, ενώ ο βαθμός κατά τον οποίο αυτή η πληροφορία υποδεικνύει ότι ο συγκεκριμένος κανόνας πρέπει να ενεργοποιηθεί παρέχεται από το επίπεδο ενεργοποίησης. Η πλήρης λοιπόν δυνατοτική αναπαράσταση παρέχεται ως:

$$Bel = t(y, y^c) \quad (2.11)$$

$$Pl = y \quad (2.12)$$

Οι οριακές συνθήκες είναι:

- $Bel = Pl = 1$ , το οποίο εκδηλώνεται όταν  $y = y^c = 1$  και υποδηλώνει την απόλυτη πεποίθηση ότι το συγκεκριμένο προφίλ είναι αυτό που ταιριάζει πλήρως στο παρατηρούμενο πρόσωπο.
- $Bel = Pl = 0$ , το οποίο εκδηλώνεται όταν  $y = 0$  και υποδηλώνει την απόλυτη πεποίθηση ότι το υπό εξέταση προφίλ δεν ταιριάζει στο παρατηρούμενο πρόσωπο.
- $Bel = 0, Pl = 1$ , το οποίο εκδηλώνεται όταν  $y = 1, y^c = 0$  και υποδηλώνει πλήρη άγνοια.



### 2.3.3 Πειραματικά Αποτελέσματα

Εφόσον το σύνολο δεδομένων που παρήχθησαν στα πλαίσια των ερευνητικών προγραμμάτων [29] και [30] δημιουργήθηκε οδηγώντας τους συμμετέχοντες σε συναισθηματικό διάλογο, οι εκφράσεις του προσώπου που εμπεριέχονται σε αυτά δεν είναι προσπονητές αλλά κυρίως φυσικές, Αξιολογήθηκαν περίπου 30000 καρτέ. Τα αποτελέσματα της ανάλυσης έκφρασης με τη δυνατοτική μέθοδο συγκρίθηκαν με αυτά της χειρωνακτικής κατηγοριοποίησης [61] (η οποία έγινε λαμβάνοντας υπόψη και τον ήχο) και τα αποτελέσματα παρουσιάζονται στον πίνακα 2.5.

Ασαφείς Κανόνες	Δυνατοτική προσέγγιση	Διαφωνία Χαρακτηρισμών
65.1%	78.4%	20.01%

**Πίνακας 2.5:** Αποτελέσματα αναγνώρισης εκφράσεων

Όπως φαίνεται στην τελευταία στήλη του πίνακα 2.5, ακόμα και οι ειδικοί ψυχολόγοι που χαρακτήρισαν τις ακολουθίες στο χώρο ενεργοποίησης-αποτίμησης, διαφωνούν κατά μέσο όρο μία φορά κάθε πέντε καρτέ, κάτι το οποίο είναι ενδεικτικό του διφορούμενου της διαδικασίας. Από τη συγκριτική μελέτη της απόδοσης των συμβατικών προτεινόμενων συστημάτων όπως φαίνεται στις πρώτες δύο στήλες, είναι επίσης φανερό ότι η προτεινόμενη προσέγγιση αποδίδει καλύτερα. Τέλος είναι σκόπιμο να τονιστεί ότι επιτυγχάνεται ποσοστό επιτυχίας στην ταξινόμηση της τάξης του 78%, με βάση μόνο τους ασαφείς κανόνες που έχουν οριστεί από ειδικούς χωρίς τη χρήση βαρών για τους όρους εισόδου των κανόνων. Η εισαγωγή της έννοιας της σημαντικότητας στους όρους των κανόνων αναμένεται να βελτιώσει περαιτέρω τα ποσοστά επιτυχίας.

Στις εικόνες 2.10, 2.11, 2.12 και 2.2 απεικονίζονται τέσσερα τμήματα της ακολουθίας «rodgy» και στην εικόνα 2.2 ένα συνεχές τμήμα από την ακολουθία «allyssa». Τα τμήματα είναι συνεχή και σε κάθε καρτέ τα 19 χαρακτηριστικά σημεία (feature points) έχουν ευρεθεί εκ νέου, δηλαδή η κάθε εκτίμηση δεν βασίζεται στην προηγούμενη όπως θα γινόταν στην περίπτωση της παρακολούθησης (tracking). Κάτω από κάθε καρτέ απεικονίζεται η αντίστοιχη εκτίμηση της έκφρασης, όπου οι πρώτες τρεις ράβδοι αντιστοιχούν στην έξοδο του συστήματος για τα τρία πρώτα τεταρτημόρια και η τέταρτη στην έξοδο για την ουδέτερη κατάσταση.

## 2.4 Αποτίμηση Κανόνων με Ασαφές Νευρωνικό Δίκτυο

Προκειμένου να εισαχθεί στο σύστημα αποτίμησης κανόνων η δυνατότητα προσαρμοστικότητας σε διαφορετικούς χρήστες, χρησιμοποιήθηκε η ιδέα του νευρωνικού ασαφούς δικτύου το οποίο περιγράφεται στο [53].

Στη συγκεκριμένη περίπτωση αναλύθηκαν τα δεδομένα ενός μόνο χρήστη ώστε να γίνει εμφανής η προσαρμοστική δυνατότητα της προτεινόμενης μεθόδου σε κάθε χρήστη. Εφόσον η βάση δεδομένων είναι οπτικοακουστική, τα δεδομένα χωρίστηκαν σε *tunes* [65]. Τα *tunes* ορίζουν χρονικές περιόδους μεταξύ δύο παύσεων στην ομιλία, διάρκειας μεγαλύτερης των 180msec. Μέσα σε κάθε *tune* επιλέχθηκαν τα καρέ τα οποία ήταν περισσότερο εκφραστικά στο οπτικό μέρος, δηλαδή αυτά με τιμές FAP που απέκλιναν περισσότερο από το καρέ ουδέτερης έκφρασης και ταυτόχρονα διέθεταν υψηλούς βαθμούς εμπιστοσύνης. Δημιουργήθηκε με αυτό τον τρόπο ένα σύνολο από 100 καρέ τα οποία επαναχαρακτηρίστηκαν οπτικά (απουσία ήχου) με τη μέθοδο των τεταρτημορίων συναισθημάτων. Από αυτά επιλέχθηκαν 54 καρέ, καταναμημένα ομοιόμορφα στα τρία τεταρτημόρια (το τεταρτημόριο (+, -) δεν περιείχε δεδομένα και αγνοήθηκε. Το δείγμα αυτό χρησιμοποιήθηκε κυρίως για λόγους εκπαίδευσης/προσαρμογής του δικτύου. Επιπλέον δημιουργήθηκε ένα δεύτερο σύνολο με περίπου 930 καρέ τα οποία επίσης χαρακτηρίστηκαν οπτικά, και χρησιμοποιήθηκαν ως το δείγμα ελέγχου.



Σχήμα 2.8: Παραδείγματα καρέ τα οποία δεν αναλύθηκαν χωρίς τη χρήση προσαρμογής

Αρχικά εφαρμόστηκε το γενικευμένο νευροασαφές μοντέλο στα παραπάνω δείγματα. Η επίδοσή του ήταν περίπου 58%, η οποία είναι μεν αρκετά μεγαλύτερη από το ποσοστό τυχαιότητας αλλά δεν λήφθηκαν υπόψη οι τρόποι έκφρασης του συγκεκριμένου χρήστη. Στην εικόνα 2.8 εμφανίζονται τρία καρέ στα οποία η αναγνώριση με τους αρχικούς κανόνες απέτυχε. Για παράδειγμα η εικόνα 2.8(α) ανήκει σύμφωνα με το χαρακτηρισμό της στο τεταρτημόριο (+, +) αλλά ο κανόνας που ενεργοποιήθηκε άνηκε στο δεύτερο τεταρτημόριο, (-, +), κάτι

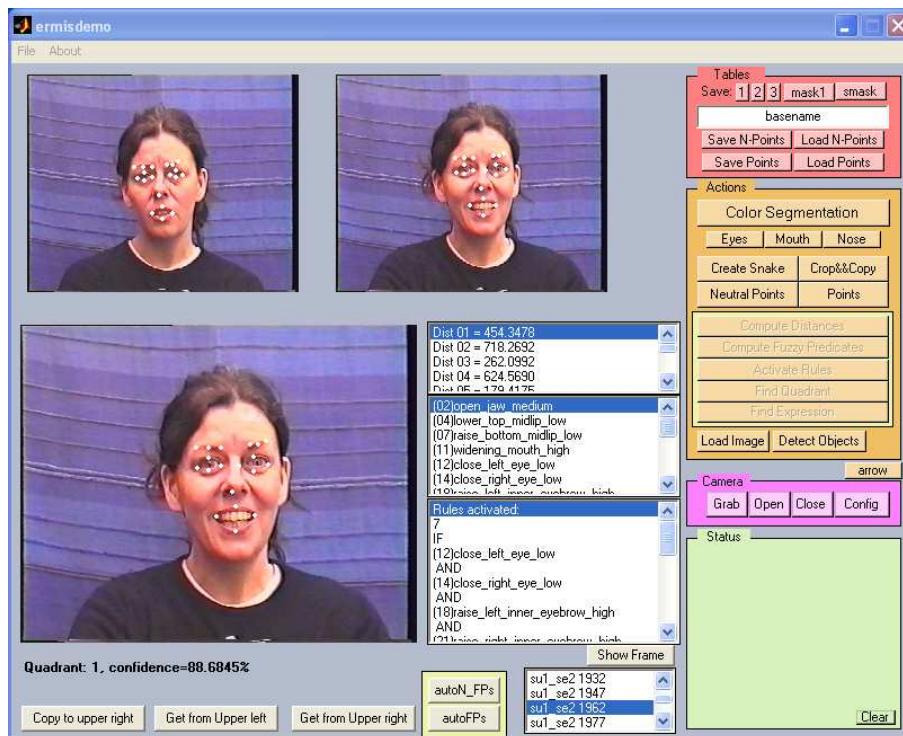
που μπορεί να αποδοθεί στο μοναδικό τρόπο με τον οποίο αυτός ο χρήστης εκφράζει τα συναισθήματά του, ο οποίος απαιτεί την προσαρμογή των αρχικών κανόνων.

Ακολούθως περιγράφεται η διαδικασία που ακολουθήθηκε προκειμένου να εξαχθούν επιπλέον κανόνες για την προσαρμογή της γνώσης του νευροασαφούς συστήματος. Αρχικά εκτελέστηκε ιεραρχική συσσωρευτική ομαδοποίηση (hierarchical agglomerative clustering) των 54 δειγμάτων FAP, με το καθένα να περιέχει 17 τιμές FAP. Μια συναθροιστική απόσταση Mahalanobis χρησιμοποιήθηκε για την ανίχνευση των υποκείμενων προτύπων και τη δημιουργία ομάδων όμοιων δειγμάτων. Με αυτό τον τρόπο λήφθηκαν 10 ομάδες και τα αντίστοιχα κέντρα τους χρησιμοποιήθηκαν μαζί με την επιθυμητή απόκριση για τον ορισμό των κανόνων με τους οποίους το νευροασαφές σύστημα θα προσαρμοζόταν εν συνεχεία τη γνώση του στο συγκεκριμένο χρήστη. Κατά την ομαδοποίηση αγνοήθηκαν οι τιμές του FAP που αντιστοιχούν στο άνοιγμα του στόματος ( $FAP_1$ ) ώστε να αγνοηθεί η επίδραση της ομιλίας στη διαδικασία της ομαδοποίησης. Πρέπει να σημειωθεί ότι κατά την προσαρμογή του δικτύου η δομή των κανόνων δεν μεταβάλλεται: προσαρμόζονται μόνο τα ενεργοποιούμενα κατηγορήματα των κανόνων εισόδου χωρίς να δημιουργούνται νέα, προστατεύοντας έτσι την αρχική γνώση που έχει παραχθεί από ειδικούς.

Κανόνας	FAPs	Τεταρτημόριο
5	$F_3\_H + F_4\_M + F_5\_L + [F_{53}+F_{54}]\_M + [F_{19}+F_{21}]\_L$ $+ [F_{20}+F_{22}]\_L + F_{31}\_H + F_{32}\_H + F_{33}\_H + F_{34}\_H$ $+ F_{37}\_M + F_{38}\_M + F_{59}\_M + F_{60}\_M$	(-, +)
29	$F_3\_L + F_4\_L + F_5\_H + [F_{53}+F_{54}]\_M + [F_{19}+F_{21}]\_M$ $+ [F_{20}+F_{22}]\_M + F_{31}\_M + F_{32}\_M + F_{33}\_L + F_{34}\_L$ $+ [F_{37}+F_{38}]\_M$	(-, -)
39	$F_3\_L + F_4\_L + F_5\_H + [F_{53}+F_{54}]\_L + [F_{19}+F_{21}]\_H$ $+ [F_{20}+F_{22}]\_H + F_{31}\_L + F_{32}\_L + F_{33}\_L + F_{34}\_L +$ $[F_{37}+F_{38}]\_H + F_{59}\_H + F_{60}\_H$	(+, +)

**Πίνακας 2.6:** Νέοι κανόνες που αποκτήθηκαν μέσω της προσαρμογής

Η διαδικασία της προσαρμογής δημιούργησε ένα σύνολο νέων κανόνων, τρεις από τους οποίους εμφανίζονται στον πίνακα 2.6. Οι αλλαγές που επιτεύχθηκαν με την προσαρμογή στο συγκεκριμένο χρήστη γίνονται αντιληπτές με τη σύγκριση των πινάκων 2.6 και 2.7. Η παραπάνω διαδικασία εφαρμόστηκε στο πλήρες δείγμα δοκιμής. Η απόδοση της αναγνώρισης αυξήθηκε στο 78% και κρίνεται πολύ ικανοποιητική λαμβάνοντας υπόψη ότι αναλύονται μη ακραία συναισθήματα. Το αρχικό σύνολο κανόνων εμφανίζεται στον πίνακα 2.7.



Σχήμα 2.9: Η διαπροσωπεία της υλοποίησης (Matlab)

Στην εικόνα 2.9 απεικονίζεται η διαπροσωπεία του υλοποιημένου συστήματος εκτίμησης έκφρασης. Διακρίνονται το «ουδέτερο» και το «εκφραστικό» καρέ, τα χαρακτηριστικά σημεία που εκτιμήθηκαν αυτόματα, το διάγραμμα αποστάσεων, η είσοδος του νευροασαφούς και οι ενεργοποιημένοι κανόνες. Κάτω αριστερά εμφανίζεται το αποτέλεσμα (τεταρτημόριο 1, εμπιστοσύνη 0.88).

## 2.5 Συμπεράσματα

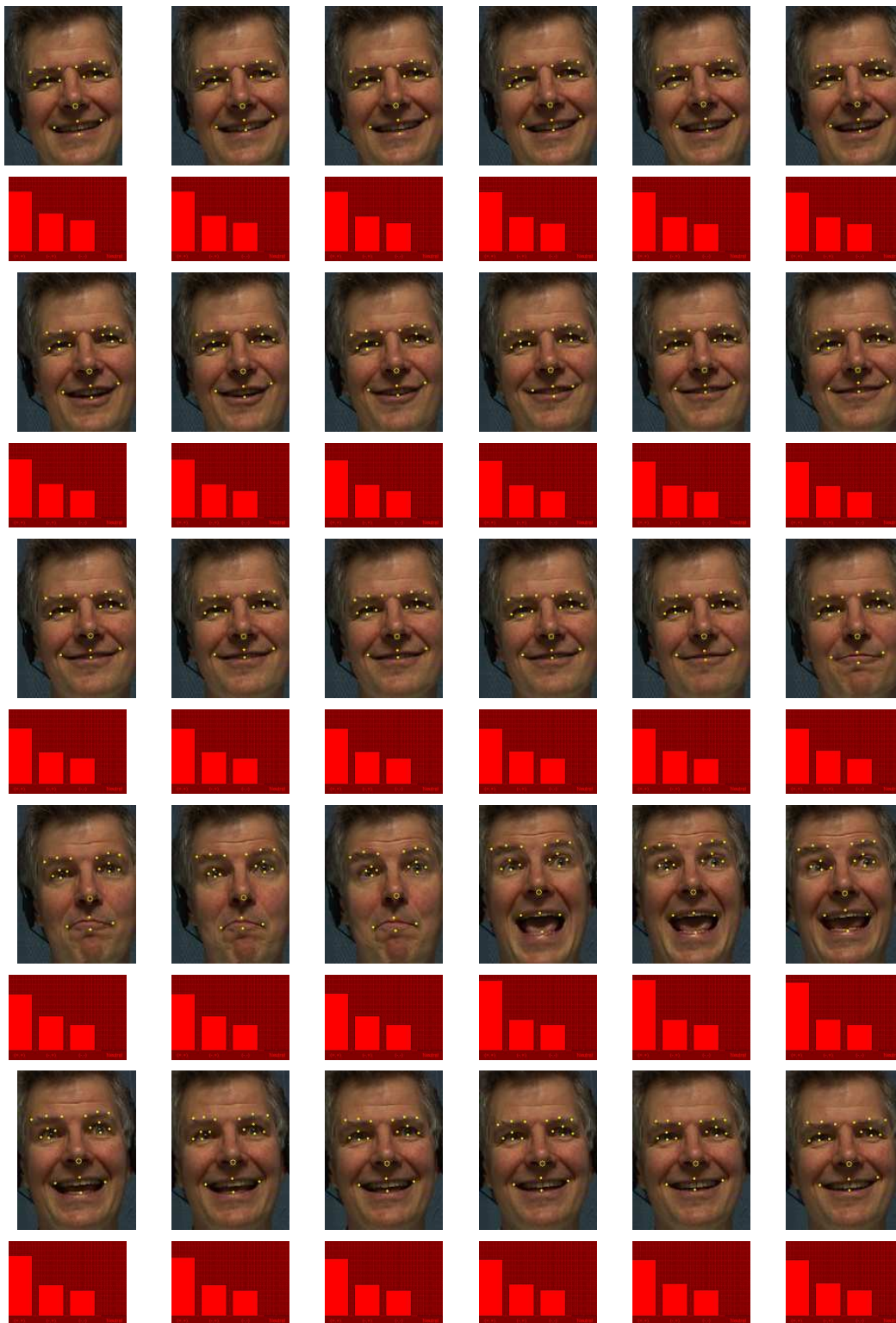
Δύο εγγενή προβλήματα των συστημάτων αναγνώρισης εκφράσεων που βασίζονται στις αναλυτικές μεθόδους, που χρησιμοποιούν δηλαδή κανόνες σχετιζόμενους με τις κινήσεις των χαρακτηριστικών του προσώπου, είναι η εξάρτηση από τα σφάλματα του συστήματος ανίχνευσης χαρακτηριστικών σημείων του προσώπου και οι περιορισμοί που επιβάλλει η αναγκαστική ομαδοποίηση των εκφράσεων κατά τη διατύπωση των κανόνων εκφράσεων, με αποτέλεσμα τη μειωμένη ικανότητα προσαρμογής των συστημάτων αναγνώρισης στους ιδιαίτερους τρόπους έκφρασης των συναισθημάτων του καθενός.

Για την αντιμετώπιση των λαθών του συστήματος εξαγωγής των χαρακτηριστικών σημείων, μπορεί να ληφθεί υπόψη η έξοδος εμπιστοσύνης του συστήματος αυτού και με αυτό το στόχο, προτάθηκε σε αυτό το κεφάλαιο η αποτί-

## Κεφάλαιο 2. Εκτίμηση Συναισθημάτων με Βάση τις Εκφράσεις του Προσώπου

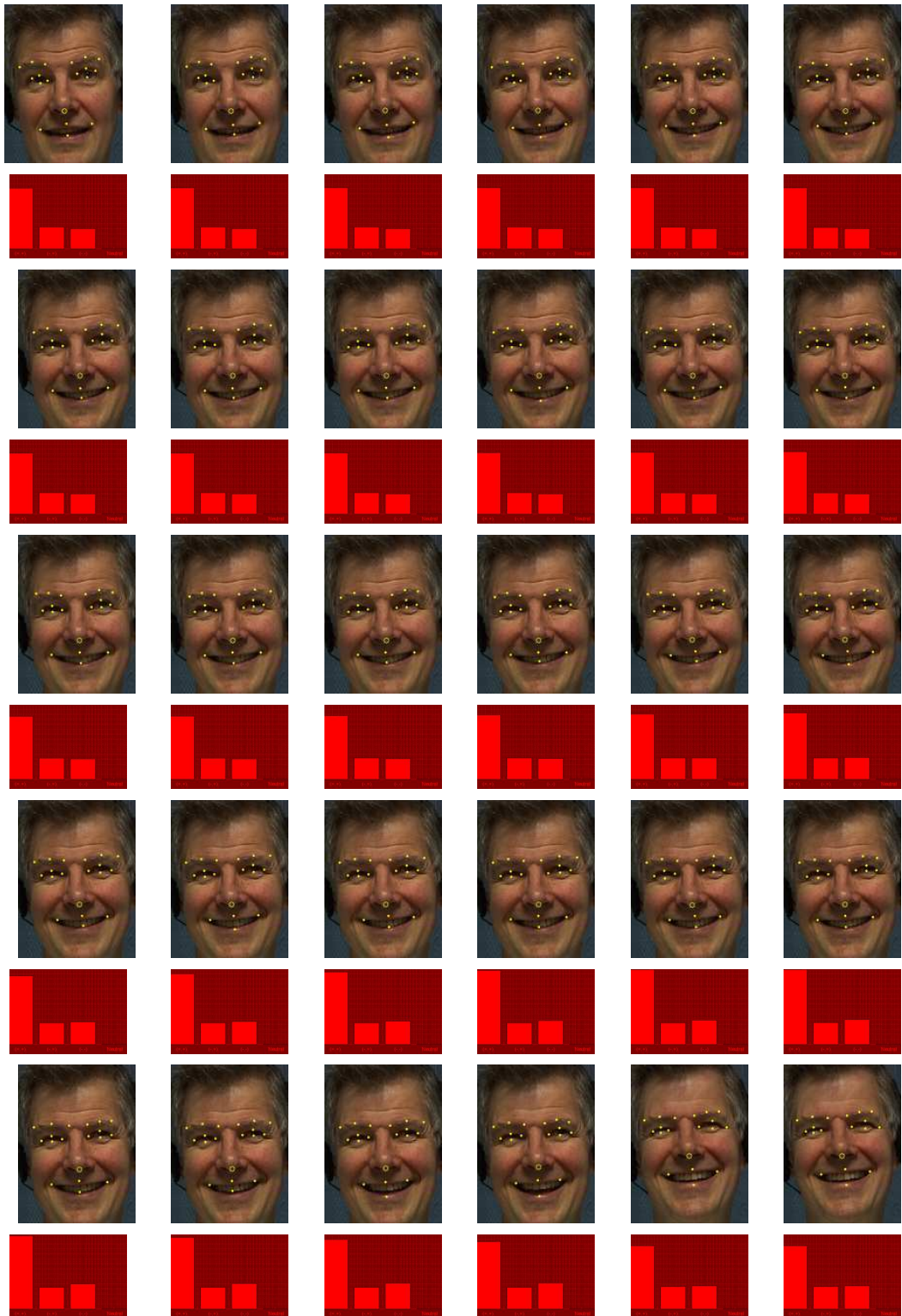
Rule	FAP	Quadrant
1	$F_{3H} + F_{4L} + F_{5VL} + [F_{53} + F_{54}]_H + [F_{19} + F_{21}]_H + [F_{20} + F_{22}]_H$	(+, +)
2	$F_{3M} + F_{4L} + F_{5L} + [F_{53} + F_{54}]_H + [F_{19} + F_{21}]_H + [F_{20} + F_{22}]_H$	(+, +)
3	$F_{3M} + F_{4L} + F_{5H} + [F_{53} + F_{54}]_H + [F_{19} + F_{21}]_H + [F_{20} + F_{22}]_H$	(+, +)
4	$F_{3H} + F_{4L} + F_{5L} + [F_{53} + F_{54}]_H + [F_{19} + F_{21}]_H + [F_{20} + F_{22}]_H$	(+, +)
5	$F_{3L} + F_{4M} + F_{5H} + [F_{53} + F_{54}]_M + [F_{19} + F_{21}]_H + [F_{20} + F_{22}]_H + F_{31M} + F_{32M} + F_{33M} + F_{34M} + F_{37M} + F_{38M} + F_{59H} + F_{60H}$	(+, +)
6	$F_{3H} + F_{4L} + F_{5VL} + [F_{53} + F_{54}]_M + [F_{19} + F_{21}]_H + [F_{20} + F_{22}]_H$	(+, +)
7	$F_{3L} + F_{4L} + F_{5H} + [F_{53} + F_{54}]_H + [F_{19} + F_{21}]_H + [F_{20} + F_{22}]_H + [F_{37} + F_{38}]_M + F_{59H} + F_{60H}$	(+, +)
8	$F_{3H} + F_{5VL} + [F_{53} + F_{54}]_L + [F_{19} + F_{21}]_L + [F_{20} + F_{22}]_L + F_{31H} + F_{32H} + F_{33H} + F_{34H} + F_{35H} + F_{36H} + F_{37L} + F_{38L} + [F_{37} + F_{38}]_L$	(+, +)
9	$F_{3H} + F_{5VL} + [F_{53} + F_{54}]_M + [F_{19} + F_{21}]_L + [F_{20} + F_{22}]_L + F_{31H} + F_{32H} + F_{33H} + F_{34H} + F_{35H} + F_{36H} + F_{37L} + F_{38L}$	(+, +)
10	$F_{3M} + F_{5L} + [F_{53} + F_{54}]_L + [F_{19} + F_{21}]_L + [F_{20} + F_{22}]_L + F_{31H} + F_{32H} + F_{33H} + F_{34H} + F_{35H} + F_{36H}$	(+, +)
11	$F_{3H} + F_{5VL} + [F_{53} + F_{54}]_L + [F_{19} + F_{21}]_L + [F_{20} + F_{22}]_L + F_{31H} + F_{32H} + F_{33H} + F_{34H} + F_{35M} + F_{36M}$	(+, +)
12	$F_{3H} + F_{5VL} + [F_{53} + F_{54}]_L + [F_{19} + F_{21}]_L + [F_{20} + F_{22}]_L + F_{31M} + F_{32M} + F_{33H} + F_{34H} + F_{35M} + F_{36M}$	(+, +)
13	$F_{3L} + F_{4M} + F_{5H} + F_{31L} + F_{32L} + F_{33L} + F_{34L} + F_{37H} + F_{38H} + [F_{37} + F_{38}]_H + F_{59M} + F_{60M}$	(-, +)
14	$F_{3L} + F_{4M} + F_{5L} + F_{31L} + F_{32L} + F_{33L} + F_{34L} + F_{37M} + F_{38M} + [F_{37} + F_{38}]_H$	(-, +)
15	$F_{3L} + F_{4M} + F_{5H} + F_{31M} + F_{32M} + F_{33L} + F_{34L} + F_{37H} + F_{38H} + [F_{37} + F_{38}]_H$	(-, +)
16	$F_{3L} + F_{4M} + F_{5L} + F_{31L} + F_{32L} + F_{33L} + F_{34L} + F_{37H} + F_{38H} + [F_{37} + F_{38}]_H + F_{59M} + F_{60M}$	(-, +)
17	$F_{3H} + F_{4L} + F_{5VL} + [F_{53} + F_{54}]_L + F_{31M} + F_{32M} + F_{33L} + F_{34L} + F_{35L} + F_{36L} + F_{37H} + F_{38H} + [F_{37} + F_{38}]_H + F_{59L} + F_{60L}$	(-, +)
18	$F_{3H} + F_{4M} + F_{5VL} + [F_{53} + F_{54}]_L + F_{31L} + F_{32L} + F_{33L} + F_{34L} + F_{35L} + F_{36L} + F_{37H} + F_{38H} + [F_{37} + F_{38}]_H$	(-, +)
19	$F_{3H} + F_{4M} + F_{5VL} + [F_{53} + F_{54}]_L + F_{31L} + F_{32L} + F_{33L} + F_{34L} + [F_{37} + F_{38}]_H + F_{59L} + F_{60L}$	(-, +)
20	$F_{3H} + F_{4L} + F_{5VL} + [F_{53} + F_{54}]_L + F_{31M} + F_{32M} + F_{33M} + F_{34M} + F_{37H} + F_{38H} + [F_{37} + F_{38}]_H + F_{59M} + F_{60M}$	(-, +)
21	$F_{3M} + F_{4L} + F_{5L} + [F_{53} + F_{54}]_L + F_{31L} + F_{32L} + F_{33L} + F_{34L} + F_{37H} + F_{38H} + [F_{37} + F_{38}]_H + F_{59M} + F_{60M}$	(-, +)
22	$F_{3L} + F_{4M} + F_{5H} + F_{31M} + F_{32M} + F_{33L} + F_{34L} + F_{37H} + F_{38H} + [F_{37} + F_{38}]_M$	(-, +)
23	$F_{3L} + [F_{53} + F_{54}]_M + F_{5H} + [F_{19} + F_{21}]_L + [F_{20} + F_{22}]_L + F_{31M} + F_{32M} + F_{33M} + F_{34M}$	(-, +)
24	$F_{3M} + [F_{19} + F_{21}]_L + [F_{20} + F_{22}]_L + F_{31H} + F_{32H} + F_{37M} + F_{38M}$	(-, +)
25	$F_{3M} + F_{4M} + F_{5H} + [F_{19} + F_{21}]_L + [F_{20} + F_{22}]_L + F_{33M} + F_{34M} + F_{35H} + F_{36H}$	(-, +)
26	$F_{3M} + F_{5L} + [F_{19} + F_{21}]_H + [F_{20} + F_{22}]_H + F_{31H} + F_{32H} + F_{33M} + F_{34M} + F_{35M} + F_{36M} + [F_{37} + F_{38}]_H$	(-, -)
27	$F_{3M} + F_{4L} + F_{5L} + [F_{53} + F_{54}]_L + [F_{19} + F_{21}]_H + [F_{20} + F_{22}]_H + F_{31L} + F_{32L} + F_{33L} + F_{34L} + [F_{37} + F_{38}]_M$	(-, -)
28	$F_{3M} + F_{4L} + F_{5H} + [F_{53} + F_{54}]_L + [F_{19} + F_{21}]_H + [F_{20} + F_{22}]_H + F_{31L} + F_{32L} + F_{33L} + F_{34L} + F_{35L} + F_{36L} + F_{37M} + F_{38M} + [F_{37} + F_{38}]_L$	(-, -)
29	$F_{3L} + F_{4L} + F_{5H} + [F_{53} + F_{54}]_L + [F_{19} + F_{21}]_H + [F_{20} + F_{22}]_H + F_{31M} + F_{32M} + F_{33M} + F_{34M} + [F_{37} + F_{38}]_M$	(-, -)
30	$F_{3L} + F_{4L} + F_{5H} + [F_{53} + F_{54}]_L + [F_{19} + F_{21}]_H + [F_{20} + F_{22}]_H + F_{31L} + F_{32L} + F_{33L} + F_{34L} + [F_{37} + F_{38}]_H$	(-, -)
31	$F_{3L} + F_{4L} + F_{5H} + [F_{53} + F_{54}]_M + [F_{19} + F_{21}]_H + [F_{20} + F_{22}]_H + F_{31L} + F_{32L} + F_{33L} + F_{34L} + [F_{37} + F_{38}]_M$	(-, -)
32	$F_{3L} + F_{4M} + F_{5H} + [F_{53} + F_{54}]_L + [F_{19} + F_{21}]_H + [F_{20} + F_{22}]_H + F_{31L} + F_{32L} + F_{33L} + F_{34L} + [F_{37} + F_{38}]_H + F_{59H}$	(-, -)
33	$F_{3M} + F_{4L} + F_{5H} + [F_{53} + F_{54}]_M + [F_{19} + F_{21}]_H + [F_{20} + F_{22}]_H + F_{31M} + F_{32M} + F_{33M} + F_{34M} + F_{35M} + F_{36M} + [F_{37} + F_{38}]_H + F_{60H}$	(-, -)
34	$F_{3L} + F_{4L} + [F_{53} + F_{54}]_L + F_{31M} + F_{32M} + F_{33M} + F_{34M} + F_{35M} + F_{36M} + F_{37M} + F_{38M} + [F_{37} + F_{38}]_M$	(-, -)
35	$F_{3H} + F_{4L} + F_{5VL} + [F_{53} + F_{54}]_L + [F_{19} + F_{21}]_H + [F_{20} + F_{22}]_H + F_{31L} + F_{32L} + F_{33L} + F_{34L} + F_{35L} + F_{36L} + F_{37H} + F_{38H} + [F_{37} + F_{38}]_H + F_{59H}$	(-, -)
36	$F_{3L} + F_{4M} + F_{5H} + [F_{53} + F_{54}]_L + [F_{19} + F_{21}]_H + [F_{20} + F_{22}]_H + F_{31L} + F_{32L} + F_{33L} + F_{34L} + [F_{37} + F_{38}]_H + F_{60H}$	(-, -)
37	$F_{3M} + F_{4L} + F_{5H} + [F_{53} + F_{54}]_M + [F_{19} + F_{21}]_H + [F_{20} + F_{22}]_H + F_{31M} + F_{32M} + F_{33M} + F_{34M} + F_{35M} + F_{36M} + [F_{37} + F_{38}]_H + F_{59H}$	(-, -)
38	$F_{3L} + F_{4L} + F_{5H} + [F_{53} + F_{54}]_M + [F_{19} + F_{21}]_H + [F_{20} + F_{22}]_H + F_{31L} + F_{32L} + F_{33L} + F_{34L} + [F_{37} + F_{38}]_H + F_{59H} + F_{60H}$	(-, -)
39	$F_{3L} + F_{4L} + F_{5H} + [F_{53} + F_{54}]_L + [F_{19} + F_{21}]_H + [F_{20} + F_{22}]_H + F_{31M} + F_{32M} + F_{33M} + F_{34M} + [F_{37} + F_{38}]_H + F_{59H} + F_{60H}$	(-, -)
40	$F_{3H} + F_{4L} + F_{5VL} + [F_{53} + F_{54}]_L + [F_{19} + F_{21}]_H + [F_{20} + F_{22}]_H + F_{31L} + F_{32L} + F_{33L} + F_{34L} + F_{35L} + F_{36L} + F_{37H} + F_{38H} + [F_{37} + F_{38}]_H + F_{60H}$	(-, -)
41	$F_{3L} + F_{4M} + F_{31M} + F_{32M} + F_{33M} + F_{34M} + F_{35M} + F_{36M} + F_{37M} + F_{38M} + [F_{37} + F_{38}]_M + F_{59M} + F_{60M}$	Neutral

**Πίνακας 2.7:** Το πλήρες δείγμα κανόνων που χρησιμοποιήθηκε στο σύστημα αναγνώρισης εκφράσεων

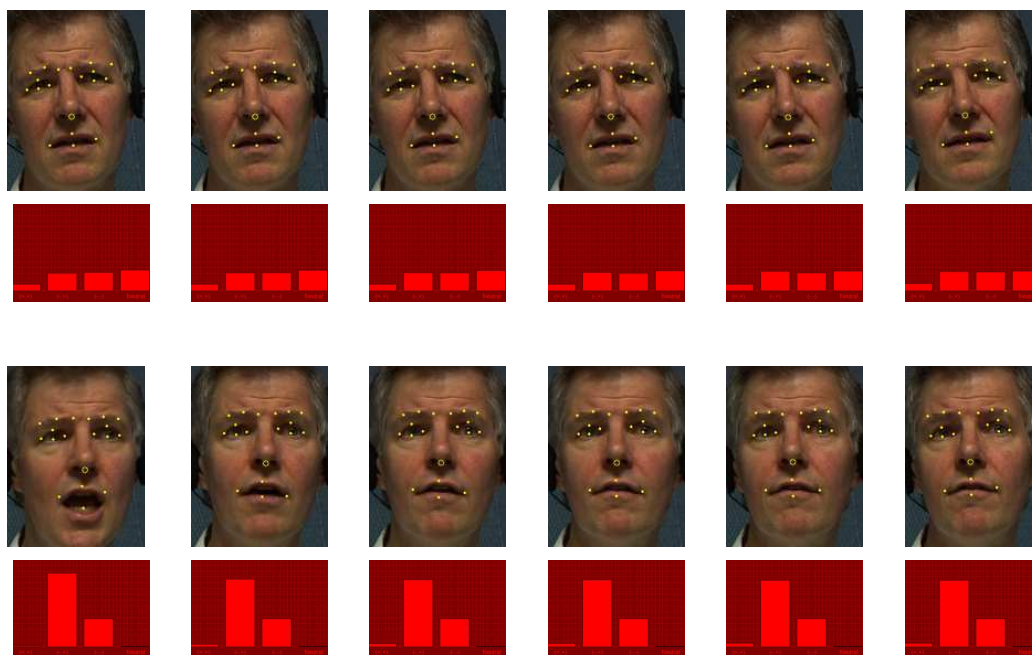


Σχήμα 2.10: Εύρεση των 19 χαρακτηριστικών σημείων σε συνεχόμενα καρέ (χωρίς tracking) και εκτίμηση της έκφρασης. Οι ράβδοι παριστάνουν τη συμμετοχή στα 3 τεταρτημόρια και στην ουδέτερη έκφραση. Διαφορική απόσταση: 96px





Σχήμα 2.11: Εύρεση των 19 χαρακτηριστικών σημείων σε συνεχόμενα καρέ (χωρίς tracking) και εκτίμηση της έκφρασης. Οι ράβδοι παριστάνουν τη συμμετοχή στα 3 τεταρτημόρια και στην ουδέτερη έκφραση. Διαχωρική απόσταση: 96px



Σχήμα 2.12: Εύρεση των 19 χαρακτηριστικών σημείων σε συνεχόμενα καρέ (χωρίς tracking) και εκτίμηση της έκφρασης. Οι ράβδοι παριστάνουν τη συμμετοχή στα 3 τεταρτημόρια και στην ουδέτερη έκφραση. Η κάθε γραμμή αντιστοιχεί σε διαφορετικό τμήμα της ακολουθίας. Διαφορική απόσταση: 96px



## *Κεφάλαιο 2. Εκτίμηση Συναισθημάτων με Βάση τις Εκφράσεις του Προσώπου*

μηση των κανόνων με ένα σύστημα δυνατοτικής αναπαράστασης της γνώσης, το οποίο έχει τη δυνατότητα να θεωρεί τα πρότερα τμήματα των κανόνων εκφράσεων, σύμφωνα με το βαθμό εμπιστοσύνης στην ορθότητα της εξαγωγής των χαρακτηριστικών προσώπου από τα οποία εξαρτώνται. Τα αποτελέσματα είναι ενθαρρυντικά και δείχνουν σαφή βελτίωση της εκτίμησης με τη χρήση αυτής της τεχνικής.

Όσον αφορά στην αντιμετώπιση των περιορισμών που εισάγει η ομαδοποίηση των εκφράσεων σε κανόνες, προτείνεται η βελτίωσή τους λαμβάνοντας υπόψη τις ιδιαιτερότητες στην εκφραστικότητα των εκάστοτε χρηστών του συστήματος αναγνώρισης έκφρασης. Για το λόγο αυτό προτάθηκε η προσαρμογή των κανόνων εκφράσεων σε συγκεκριμένο χρήστη με τη χρήση ενός νευροασαφούς συστήματος αποτίμησης. Τα αποτελέσματα που παρουσιάζονται σε αυτό το κεφάλαιο είναι ενθαρρυντικά και δείχνουν ότι με την προσαρμογή των ενεργοποιούμενων κατηγορημάτων των κανόνων εισόδου, υπήρξε σαφής βελτίωση στο ποσοστό αναγνώρισης των εκφράσεων του συγκεκριμένου χρήστη.

□



## Κεφάλαιο 3

# Επεκτάσεις με Βάση την Ανάλυση Χειρονομιών

### 3.1 Ανάλυση χειρονομιών

Η έρευνα εστιάζεται σε πέντε βασικές μεθόδους αναγνώρισης της συναισθηματικής κατάστασης ενός χρήστη: την ανάλυση εκφράσεων προσώπου, την αναγνώριση ομιλίας, την εκτίμηση των προσωδιακών χαρακτηριστικών της φωνής όπως είναι η χροιά και η ένταση, την ανάλυση χαρακτηριστικών της φυσιολογίας (π.χ. καρδιακοί παλμοί, συχνότητα αναπνοών, αγωγιμότητα δέρματος) και την ανάλυση χειρονομιών.

Στο κεφάλαιο αυτό παρουσιάζεται μια τεχνική κατάτμησης εικόνων και ακουλουθιών βίντεο, η οποία μπορεί να χρησιμοποιηθεί από συστήματα υψηλότερου επιπέδου για την ανίχνευση και κατηγοριοποίηση των χειρονομιών. Η ανάλυση των χειρονομιών, αξιοποιούμενη σε συνδυασμό με την ανάλυση εκφράσεων του προσώπου, μπορεί να χρησιμοποιηθεί για την δημιουργία ενός εύρωστου συστήματος εκτίμησης της εκφραστικότητας.

Η ανάλυση και ερμηνεία των χειρονομιών αποτελεί τα τελευταία χρόνια ένα σημαντικό κομμάτι της διαδραστικότητας ανθρώπου-μηχανής. Πολλές φορές μια απλή κίνηση μεταδίδει άμεσα ένα μήνυμα που μπορεί να δηλώνει μια συναισθηματική κατάσταση όπως π.χ. η τοποθέτηση των χεριών επάνω στα αυτιά δηλώνει ότι ο χρήστης δεν θέλει να ακούσει τίποτε άλλο. Τέτοια μηνύματα εκφράζονται πολλές φορές πιο φυσικά με χειρονομίες παρά με το λόγο. Οι σημαντικές χειρονομίες μπορούν να αποτελούνται είτε από κινήσεις των χεριών είτε από τη στατική θέση τους. Συνήθως οι στατικές θέσεις των χεριών εκφράζουν ορισμένες έννοιες μέσω συγκεκριμένων διατάξεων του χεριού και των δακτύλων, ενώ οι χρονικά εξελισσόμενες χειρονομίες αναπαριστούν συνήθως

Συναίσθημα	Χειρονομία
Χαρά	Χειροκρότημα υψηλής συχνότητας
Λύπη	Στάση με τα χέρια επάνω από το κεφάλι
Θυμός	Σήκωμα χεριού με υψηλή ταχύτητα, γρήγορες κινήσεις χεριών μικρής έκτασης
Φόβος	Στάση με τα χέρια επάνω από το κεφάλι, γρήγορες κινήσεις χεριών μικρής έκτασης, πρόταση του χεριού
Αποστροφή	Σήκωμα χεριού με χαμηλή ταχύτητα, χειροκρότημα χαμηλής συχνότητας
Έκπληξη	Στάση με τα χέρια επάνω από το κεφάλι

**Πίνακας 3.1:** Παραδείγματα αντιστοίχισης μεταξύ χειρονομιών και συναισθηματικών καταστάσεων

κάποιες ενέργειες. Ορισμένες φορές οι στατικές θέσεις των χεριών αποτελούν μεταβατικές καταστάσεις στις χρονικά εξελισσόμενες χειρονομίες. Εφόσον οι χρονικά εξελισσόμενες χειρονομίες εμπεριέχουν τις έννοιες της τροχιάς, της ταχύτητας και της επιτάχυνσης, η μοντελοποίησή τους είναι ιδιαίτερα περίπλοκη, δεδομένης της ποικιλίας τους και η εύρεση ενός κατάλληλου μοντέλου περιγραφής χειρονομιών αποτελεί ακόμα ανοικτό ζήτημα.

Οι χειρονομίες στις περισσότερες περιπτώσεις δεν είναι αρκετά σαφείς ώστε να υποδείξουν από μόνες τους την ύπαρξη κάποιου συναισθήματος. Μπορούν όμως να χρησιμοποιηθούν ενισχυτικά σε συνδυασμό με άλλες μεθόδους αναγνώρισης συναισθήματος όπως η ανάλυση φωνής και η ανάλυση εκφράσεων του προσώπου. Συγκεκριμένα, λαμβάνοντας υπόψη το συνολικό νοηματικό πλαίσιο μιας συνομιλίας, ορισμένες χειρονομίες μπορούν να συσχετιστούν εύκολα με κάποιες εκφράσεις, π.χ. το γρήγορο χειροκρότημα μπορεί εκφράζει χαρά και ικανοποίηση. Τα ποσοτικά χαρακτηριστικά που μπορούν να εξαχθούν από τις κινήσεις των χεριών, μπορούν επιπλέον να χρησιμοποιηθούν ως ποσοτικοί περιγραφείς μιας συναισθηματικής κατάστασης. Για παράδειγμα, με την αύξηση της ταχύτητας και του πλάτους του χειροκροτήματος, είναι εφικτός ο διαχωρισμός μεταξύ της απλής ικανοποίησης και της χαράς. Στον πίνακα 3.1 απεικονίζεται μια αντιστοίχιση μεταξύ ορισμένων ανιχνεύσιμων χειρονομιών και των έξι αρχετυπικών εκφράσεων.

Δεδομένου του σημασιολογικού επικοινωνιακού πλαισίου, οι κατηγορίες των χειρονομιών που αντιστοιχούν στην ίδια συναισθηματική κατάσταση μπορούν να θεωρηθούν ως ανεξάρτητες μαρτυρίες για την ύπαρξη του αντίστοιχου συναισθήματος. Στον πίνακα 3.1 φαίνεται ότι οι χειρονομίες μπορούν να αντιστοιχούν σε

περισσότερες από μία συναισθηματικές κατηγορίες. Για παράδειγμα η ανίχνευση της χειρονομίας «Στάση με τα χέρια επάνω από το κεφάλι» μπορεί να χρησιμοποιηθεί για την ενίσχυση της εμπιστοσύνης του συστήματος αναγνώρισης εάν το ήδη αναγνωρισμένο συναισθηματικό από την ανάλυση εκφράσεων εμπίπτει σε μία από τις κατηγορίες «λύπη, φόβος, έκπληξη» ή να την αποδυναμώσει εάν εμπίπτει στις κατηγορίες «χαρά, θυμός, αποστροφή». Επιπλέον η ανίχνευση χειροκροτήματος υψηλής συχνότητας μπορεί να ενισχύσει άμεσα την εμπιστοσύνη του συστήματος αναγνώρισης έκφρασης εάν οι ήδη αναγνωρισμένες εκφράσεις εμπίπτουν στην κατηγορία «χαρά» ενώ η ανίχνευση χειροκροτήματος χαμηλής συχνότητας που συνήθως υποδηλώνει ειρωνεία, μπορεί να ενισχύσει την ανίχνευση εκφράσεων του προσώπου που σχετίζονται με την αποστροφή.

Γενικότερα, οι ανθρώπινες χειρονομίες αποτελούνται από τις καθολικές κινήσεις των χεριών και από τις τοπικές κινήσεις των δακτύλων. Για τον εντοπισμό των χεριών, δύο είδη χαρακτηριστικών χρησιμοποιούνται ευρέως, τα χαρακτηριστικά χρώματος δέρματος και τα χαρακτηριστικά κίνησης [91] [92]. Το χρώμα του δέρματος παρέχει έναν αποτελεσματικό και αποδοτικό τρόπο εντοπισμού της θέσης των χεριών. Για την ανάλυση της αρθρωτής κίνησης των χεριών με πλήρη βαθμό ελευθερίας, πρέπει να ανιχνευτεί τόσο η κίνηση του χεριών όσο και η τοπικές κινήσεις των δακτύλων. Για το σκοπό αυτό έχουν προταθεί διάφορες τεχνικές όπως τεχνικές που βασίζονται σε δισδιάστατα παραμορφούμενα πρότυπα [93], ροπές εικόνων διαφορών, ανίχνευση των άκρων των δακτύλων, ιστογράμματα προσανατολισμού ακμών και ταίριασμα προτύπων σε χώρο ιδιοδιανυσμάτων. Τελευταία έχουν προταθεί και τεχνικές προσαρμογής τρισδιάστατων μοντέλων όπως η [90] που αντιμετωπίζει το πρόβλημα της απόκρυψης (occlusion), μαθαίνοντας και μαντεύοντας τις παραμέτρους του τρισδιάστατου μοντέλου βάσει παραλλαγών του περιγράμματος των χεριών.

Σε αντίθεση με τις εικόνες προσώπου που συνήθως στερούνται λεπτομέρειας και ανάλυσης, οι εικόνες όπου εμφανίζεται το μεγαλύτερο μέρος του ανθρώπινου σώματος έχουν την κατάλληλη κλίμακα ώστε να επιδέχονται πολλών ειδών καταταμήσεις, όπως είναι η κατάτμηση βάθους, χρώματος, κίνησης κ.λπ. Στις επόμενες ενότητες περιγράφεται ένα σύστημα το οποίο επιδέχεται σημασιολογικούς κανόνες για τη συγχώνευση πολλαπλών καταταμήσεων. Το σύστημα στοχεύει στο συνδυασμό των βέλτιστων χαρακτηριστικών που παρέχει το κάθε είδος κατάτμησης σε μια βελτιωμένη τελική κατάτμηση, η οποία μπορεί να χρησιμοποιηθεί ως το πρώτο στάδιο σε εφαρμογές εντοπισμού των μελών του σώματος και ανάλυσης χειρονομιών.

## 3.2 Ασαφής συγχώνευση πολλαπλών ειδών εισόδου για την κατάτμηση εικόνων και βίντεο

Η συγχώνευση πολλαπλών κατατμήσεων έχει αποδειχθεί απαραίτητο εργαλείο για την αυτόματη αντικειμενοστρεφή κατάτμηση και κωδικοποίηση βίντεο, κυρίως στο πλαίσιο του πρότυπου MPEG-4 [85]. Ορισμένα από τα κορυφαία συστήματα ανάλυσης και κωδικοποίησης, πετυχαίνουν χωροχρονική κατάτμηση περιεχομένου σε ακολουθίες βίντεο και εικόνες ενοποιώντας πολλαπλά στοιχεία-πρότυπα (multiple cue fusion) όπως είναι η κατάτμηση με βάση το χρώμα, την κίνηση, την υφή και το βάθος αλλά και λαμβάνοντας υπόψη αντισταθμισμένες κατατμήσεις προηγούμενων καρέ βίντεο, κυρίως για την παρακολούθηση αντικειμένων [66, 74]. Η διαδικασία της συγχώνευσης επιτυγχάνεται με την παραγωγή ιεραρχικών δομών κατατμήσεων των εικόνων, κατασκευή δέντρων αποφάσεων και ευριστικών μεθόδων επεξεργασίας της κατάτμησης βασισμένους σε κανόνες. Σε πολλές περιπτώσεις οι τρέχουσες προσεγγίσεις εστιάζουν κυρίως στην κωδικοποίηση βίντεο, συνήθως μέσω κριτηρίων ρυθμού-απώλειας (rate-distortion) [82]. Άλλες προσεγγίσεις επιτελούν ταυτόχρονη συγχώνευση και κατάτμηση με αποτέλεσμα να μην επεκτείνονται εύκολα και σε επιπλέον είδη εισόδου [81], παρόλο που μπορεί να παρέχουν εύρωστο διαχωρισμό μεταξύ αντικειμένων με διαφορετικά πρότυπα κινήσεων.

Τα επιπλέον είδη εισόδου εκτός του χρώματος, δηλαδή του βάθους, της υφής του σχήματος και της κίνησης είναι συνήθως απαραίτητα για τον ακριβή εντοπισμό αντικειμένων. Κάθε επιπλέον είδος εισόδου μπορεί να προσφέρει αυξημένη ακρίβεια κατάτμησης σε συγκεκριμένες περιπτώσεις όπου τα άλλα αποτυγχάνουν. Επιπλέον, κάθε κατάτμηση στη πολλές περιπτώσεις παριστάνεται από μια ομάδα διαμερισμών με διαφορετικά επίπεδα λεπτομέρειας, σχηματίζοντας έτσι ένα δέντρο διαμερισμών το οποίο προέρχεται από επεξεργασία της εισόδου σε πολλαπλές αναλύσεις. Η γενικευμένη διαχείριση και επεξεργασία πιο περίπλοκων δομών αναπαράστασης γίνεται ακόμα πιο δύσκολη, ιδιαίτερα παρουσία αβεβαιότητας.

Επιπλέον η διαδικασία συγχώνευσης θα πρέπει να αλληλεπιδρά με τον εντοπισμό αντικειμένων, την κατηγοριοποίηση και την αναγνώρισή τους. Τα γνωσιακά συστήματα όρασης απαιτούν εκ των προτέρων γνώση για τα αντικείμενα προκειμένου να αποδίδουν αποτελεσματικά [69]. Η εφαρμογή αυστηρών συσχετίσεων μεταξύ των τμημάτων μιας εικόνας κάνει το σύστημα πολύ ευαίσθητο στο θόρυβο· είναι σκόπιμο επομένως να εισαχθεί ένα μέτρο ασάφειας σε κάθε επίπεδο της αναπαράστασης του κατακερματισμού. Η εισαγωγή της ασάφειας, ο συνδυασμός πολλών στοιχείων, η προσαρμογή του συστήματος, η χρήση ευφυίας

και η χρήση της εκ των προτέρων γνώσης μπορεί να επιτρέψει στους αλγορίθμους συγχώνευσης κατακερματισμένων στοιχείων να αποδίδουν αποτελεσματικά σε γενικευμένα συστήματα υπολογιστικής όρασης ακόμα και σε πραγματικές συνθήκες [83].

Τα παραπάνω ζητήματα πυροδότησαν την εξέλιξη των ασαφών συστημάτων συγχώνευσης βασισμένα σε κανόνες [80] τα οποία μπορούν να διαχειριστούν περίπλοκες αναπαραστάσεις δένδρων διαμερισμού και να διευθετήσουν την ολοκλήρωση πολλών τύπων εισόδου όπως η φωτεινότητα, το χρώμα, η υφή, η κίνηση και το βάθος. Τα συστήματα που βασίζονται σε ασαφείς κανόνες συγχώνευσης με ενσωματωμένη σύνθετη γνώση αναπαράστασης μπορούν να επιτρέψουν την εύρωστη κατάτμηση εικόνων και βίντεο σε αντίθεση με τις συμβατικές τεχνικές συγχώνευσης οι οποίες βασίζονται σε κριτήρια rate-distortion ή στοιχειώδεις ευριστικούς κανόνες. Μπορούν να επιτρέψουν επίσης τον ευκολότερο ορισμό κανόνων παρακολούθησης πολλαπλών αντικειμένων, αντιμετωπίζοντας τις περιπτώσεις παραλλαγών στην εμφάνιση, απόκρυψης/εμφάνισης, και αρθρωτής κίνησης των αντικειμένων.

Το κεφάλαιο αυτό εστιάζεται στην συγχώνευση των αποτελεσμάτων από κατατμήσεις διαφορετικών εισόδων και στην εξαγωγή ασαφών αποφάσεων για τη διαχείριση των ασυνεπειών και τη βελτίωση της ακρίβειας κατάτμησης. Η προτεινόμενη μέθοδος λαμβάνει υπόψη της την εκ των προτέρων γνώση που σχετίζεται με τα αναμενόμενα χαρακτηριστικά των διαφορετικών τμαχίων (segments). Για παράδειγμα η πληροφορία της κίνησης σχετίζεται με τα κινούμενα μέρη των αντικειμένων, η πληροφορία του βάθους ορίζει τα επίπεδα/όψεις των αντικειμένων αλλά μάλλον ασαφώς και η πληροφορία του χρώματος παρέχει ακριβή αλλά όρια αλλά υπερκατατμημένα αντικείμενα. Ο στόχος της παρουσιαζόμενης τεχνικής είναι η συγχώνευση των επιμέρους ξεχωριστών κατατμήσεων, η ομαδοποίηση των περιοχών ομάδων σε αντικείμενα και η αποκόμιση μιας αναπαράστασης υποψήφιων αντικειμένων προβάλλοντας και συνδυάζοντας τα παραγόμενα τμήματα των διαφόρων εισόδων.

Πιο συγκεκριμένα, οι γενικές ιδιότητες των διαφορετικών ειδών εισόδου δίνονται παρακάτω, φανερώνοντας ότι οι κατατμήσεις που δημιουργούνται από τις διάφορες εισόδους δεν μπορούν να αξιοποιηθούν μεμονωμένα για την σημασιολογική εξαγωγή αντικειμένων:

- η κατάτμηση χρώματος/έντασης, απορρέει μέσω κριτηρίων χωρικής ομοιογένειας, επιτρέπει πολύ ακριβή καθορισμό των περιγραμμάτων των περιοχών αλλά αποτυγχάνει στον εντοπισμό αντικειμένων που απαρτίζονται από ποικιλία χρωματικών χαρακτηριστικών.

- η κατάτμηση υφής, επιτρέπει την εξαγωγή περιοχών με συγκεκριμένα πρότυπα χρώματος/φωτεινότητας αλλά επίσης δημιουργεί μεγάλο αριθμό τεμαχίων ανά αντικείμενο
- η κατάτμηση κίνησης, συνήθως βασίζεται σε παραμετρικά μοντέλα κίνησης, παρέχει πιο περιορισμένο αριθμό περιοχών αλλά παρέχει χονδρικά όρια περιοχών εξαιτίας σφαλμάτων ταιριάσματος και αποκρύψεων, ενώ μπορεί να κατατμήσει αντικείμενα με αρθρωτή κίνηση
- η κατάτμηση βάθους, είτε μέσω στερεοσκοπικής ανάλυσης είτε μέσω σχετικής κίνησης μεταξύ γειτονικών περιοχών, επιτυγχάνει αξιόπιστη προσέγγιση των αληθινών αντικειμένων, παρά ταύτα αποτυγχάνει στον εντοπισμό απομακρυσμένων ακίνητων αντικειμένων ενώ οι περιοχές βάθους που παράγει διαθέτουν επίσης ανακριβή όρια
- η κινησιολογικά αντισταθμισμένη διαμέριση προηγούμενου καρέ, παρέχει μια καθαρή ένδειξη για τις αναμενόμενες θέσεις των αντικειμένων και είναι απαραίτητη για την χρονική παρακολούθηση αντικειμένων και τον εντοπισμό νέων αντικειμένων αλλά χρήζει διόρθωση με τη χρήση των διαμερίσεων που προέρχονται από τις άλλες εισόδους

Γενικά, οι διαμερίσεις εικόνων που προέρχονται από τις κατατμήσεις χρώματος, υφής, κίνησης και βάθους συγκροτούν μια ιεραρχική δενδρική δομή με την έννοια ότι η μια περιοχή βάθους αποτελείται από μία ή περισσότερες περιοχές κίνησης, οι περιοχές κίνησης περιέχουν αρκετές περιοχές χρώματος κ.ο.κ. Η συγχώνευση μπορεί λοιπόν να επιτευχθεί με προβολή και ιεραρχική ομαδοποίηση των διαμερίσεων χαμηλότερου επιπέδου, βασιζόμενη στις διαμερίσεις υψηλότερου επιπέδου. Εν τούτοις αυτή δεν είναι η μοναδική περίπτωση εφόσον ένα αντικείμενο μπορεί να αποσυντίθεται σε ανεξάρτητες μη αλληλεπικαλυπτόμενες διαμερίσεις με βάση το χρώμα ή την κίνηση. Για το λόγο αυτό η προτεινόμενη μέθοδος συγχώνευσης πραγματοποιεί έξυπνη ομαδοποίηση βασιζόμενη στην εκ των προτέρων γνώση.

Συγκεκριμένα με βάση τις υπάρχουσες προσπάθειες συγχώνευσης κατατμήσεων χρώματος και κίνησης, κατασκευάζεται ένα αρχικό σύνολο ευρετικών κανόνων, όπως το εξής

- Η προβολή μιας περιοχής χρώματος σε περιοχές κίνησης καθορίζει το βαθμό συμμετοχής της σε κάθε περιοχή κίνησης
- Η ομαδοποίηση περιοχών χρώματος που ανήκουν στην ίδια περιοχή κίνησης παρέχει μια ενιαία περιοχή με ακριβή όρια

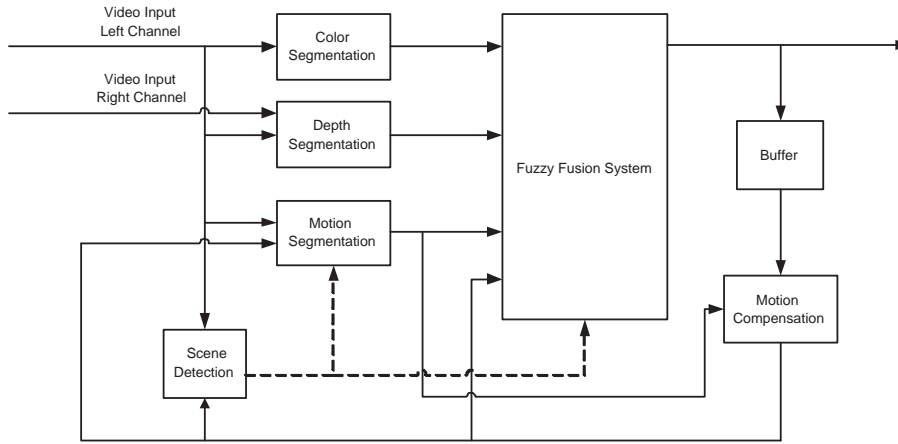


- Η προβολή τέτοιων περιοχών σε περιοχές κίνησης όπου έχει γίνει αντιστάθμιση της κίνησης (motion compensation) επιτρέπει τη σύνδεση περιοχών με περιοχές προηγούμενων καρέ (tracking)
- Οι ασυμφωνίες μεταξύ των περιοχών κίνησης και των περιοχών αντισταθμισμένης κίνησης σε προηγούμενα καρέ δεικνύει την εισαγωγή νέων αντικειμένων ή το διαχωρισμό υπάρχοντων λόγω αρθρωτής κίνησης

Αυτό το σύνολο εμπλουτίζεται κατάλληλα για να συμπεριλάβει την πληροφορία του βάθους. Το συγκεκριμένο σύστημα που παρουσιάζεται εστιάζεται στο χρώμα την κίνηση το βάθος και την κινησιολογικά αντισταθμισμένη διαμέριση του προηγούμενου καρέ. Εντούτοις η διατύπωση της προτεινόμενης τεχνικής συγχώνευσης επιτρέπει την ενσωμάτωση οποιουδήποτε αριθμού διαφορετικών εισόδων με ενιαίο και συνεπή τρόπο. Για την εκτέλεση της συγχώνευσης επιστρατεύονται τεχνικές τεχνητής νοημοσύνης και συγκεκριμένα ασαφή συστήματα κανόνων [79]. Επιπλέον είναι δυνατή η προσαρμογή επιτρέποντας προσθήκη αφαίρεση και ενημέρωση των κανόνων. Παρουσιάζονται τέλος πειράματα σε φυσικές ακολουθίες βίντεο ώστε να δειχθούν τα αποτελέσματα της προτεινόμενης τεχνικής.

### 3.3 Η προτεινόμενη αρχιτεκτονική

Προκειμένου να εκμεταλλευτούμε την κατάτμηση βάθους στη διαδικασία της συγχώνευσης, εκτελείται εκτίμηση βάθους σε στερεοσκοπικές ακολουθίες βίντεο αποτελούμενες από δύο κανάλια. Κάθε στερεοσκοπική ακολουθία αρχικά χωρίζεται σε δύο ακολουθίες, μία για το αριστερό και μία για το δεξί κανάλι. Το αριστερό κανάλι χρησιμοποιείται για την κατάτμηση του βάθους, της κίνησης και του χρώματος ενώ το δεξί κανάλι χρησιμοποιείται μόνο για την κατάτμηση βάθους και τον υπολογισμό του χάρτη απόκλισης (disparity map). Για κάθε καρέ, όταν όλα τα υποσυστήματα παραγάγουν τις τελικές κατατμήσεις τους, αυτές προωθούνται στο ασαφές σύστημα συγχώνευσης το οποίο παράγει τον τελικό χάρτη κατακερματισμού. Ο τελικός αυτός χάρτης επανατροφοδοτείται στο σύστημα ώστε να χρησιμοποιηθεί για την ανίχνευση των αντικειμένων αλλά και για την ανίχνευση και αντιστάθμιση της κίνησης στο επόμενο καρέ [76][77]. Όλες οι μονάδες κατάτμησης χρησιμοποιούν τον αλγόριθμο M-RSST για την κατασκευή των διαμερίσεων. Το διάγραμμα ροής του προτεινόμενου συστήματος φαίνεται στο σχήμα 3.1.



Σχήμα 3.1: Γενική Αρχιτεκτονική του Συστήματος

### 3.3.1 Ο Αλγόριθμος RSST

Ως βάση για την κατάτμηση εικόνων στο πλαίσιο της διατριβής χρησιμοποιήθηκε ο αλγόριθμος Recursive Shortest Spanning Tree (RSST) [94]. Παρά τη σχετικά μεγάλη υπολογιστική του πολυπλοκότητα, θεωρείται ένα ισχυρό εργαλείο κατάτμησης εικόνων συγκρινόμενος με άλλες τεχνικές [95], όπως color clustering, pyramidal region growing καθώς και το μορφολογικό αλγόριθμο watershed [96] και στη συγκεκριμένη περίπτωση προτιμάται λόγω της δυνατότητας εύκολου ορισμού του κριτηρίου συγχώνευσης περιοχών.

Κατά την αρχικοποίηση του αλγορίθμου η εικόνα εισόδου  $I$ , διαστάσεων  $M \times N$  χωρίζεται σε  $M \times N$  περιοχές (regions) ή τμήματα (segments) μεγέθους 1 pixel και για κάθε περιοχή παράγονται σύνδεσμοι (links) με τις τέσσερις γειτονικές της περιοχές. Γειτονικές μιας περιοχής  $X$  θεωρούνται οι περιοχές εκείνες που απέχουν 1 pixel από τη  $X$  και συνδέονται με τη  $X$  οριζόντια ή κατακόρυφα, αλλά όχι διαγώνια (4-connected). Σε κάθε σύνδεσμο ανατίθεται ένα βάρος (weight) ίσο με τη απόσταση (distance) των δύο περιοχών τις οποίες συνδέει. Η απόσταση αυτή μπορεί να ορισθεί με οποιοδήποτε τρόπο και καθορίζει ουσιαστικά την προτεραιότητα με την οποία οι περιοχές ενώνονται μεταξύ τους. Στη συνέχεια όλοι οι σύνδεσμοι ταξινομούνται κατά αύξουσα σειρά απόστασης, έτσι ώστε ο πρώτος σύνδεσμος με τη μικρότερη απόσταση να αντιστοιχεί στις πιο «κοντινές» περιοχές.

Κατά την επαναληπτική φάση του RSST πραγματοποιείται αναδρομική συγχώνευση (merging) επιλέγοντας κάθε φορά τις δύο γειτονικές περιοχές με τη μικρότερη απόσταση και εφαρμόζοντας τις ακόλουθες ενέργειες σε κάθε επανάληψη:

1. συγχωνεύονται οι δύο πλησιέστερες γειτονικές περιοχές, και επαναυπολο-

γίζονται οι ιδιότητες των νέων περιοχών

2. επαναυπολογίζονται και ταξινομούνται τα βάρη των συνδέσμων της νέας περιοχής με όλες τις γειτονικές της
3. αφαιρούνται οι τυχόν διπλοί σύνδεσμοι που προκύπτουν από τη συγχώνευση

Η επαναληπτική διαδικασία τερματίζεται όταν είτε ο συνολικός αριθμός περιοχών είτε η ελάχιστη απόσταση γίνει ίση με ένα προκαθορισμένο όριο ή κατώφλι (threshold). Το κατώφλι απόστασης είναι συνήθως προτιμότερο διότι το πλήθος τμημάτων που προκύπτει από την κατάτμηση εξαρτάται από τα αντικείμενα που υπάρχουν στην εικόνα. Η υπολογιστική πολυπλοκότητα του RSST καθορίζεται σε μεγάλο βαθμό από την επιλογή του αλγορίθμου ταξινόμησης των βαρών, ο οποίος αποτελεί το βασικό παράγοντα καθυστέρησης. Για το λόγο αυτό έχει προταθεί πολυδιακριτική υλοποίηση (multiresolution implementation) του αλγορίθμου, που ονομάζεται Multiresolution RSST (MRSST) [71] η οποία εφαρμόζει αναδρομικά τον αλγόριθμο RSST σε εικόνες αυξανόμενης ανάλυσης. Αρχικά πραγματοποιείται ανάλυση της εικόνας εισόδου  $I$  σε πολλαπλές αναλύσεις (multiresolution decomposition) με χαμηλότερο επίπεδο ανάλυσης  $L_0$  και κατασκευάζεται μία ιεραρχία εικόνων  $I(0) = I, I(1), \dots, I(L_0)$ . Δημιουργείται συνεπώς μία κόλινος πυραμίδα εικόνων (truncated image pyramid) κάθε επίπεδο της οποίας έχει το ένα τέταρτο των pixels του αμέσως χαμηλότερου επιπέδου.

Η αρχικοποίηση του αλγορίθμου RSST εφαρμόζεται στην εικόνα χαμηλότερης ανάλυσης  $I(L_0)$

και στη συνέχεια ξεκινά μία επαναληπτική διαδικασία που περιλαμβάνει τα παρακάτω βήματα:

1. οι περιοχές της τρέχουσας εικόνας συγχωνεύονται αναδρομικά χρησιμοποιώντας την επαναληπτική φάση του αλγορίθμου RSST
2. τα συνοριακά pixels των περιοχών που προκύπτουν διαίρουνται σε τέσσερις νέες περιοχές το καθένα. οι χρωματικές συνιστώσες των νέων περιοχών προέρχονται από την εικόνα του αμέσως επόμενου υψηλότερου επιπέδου ανάλυσης
3. υπολογίζονται και ταξινομούνται τα βάρη των νέων συνδέσμων των νέων περιοχών

Η παραπάνω διαδικασία συγχώνευσης / διαίρεσης (splitting) επαναλαμβάνεται μέχρι και την εικόνα υψηλότερης ανάλυσης  $I(0)$ . Η συγχώνευση των περιοχών

της τελευταίας που προκύπτει από το βήμα 1 δίνει και το τελικό αποτέλεσμα κατάτμησης του αλγορίθμου M-RSST.

Όπως φαίνεται από την παραπάνω περιγραφή, η πραγματική συγχώνευση τμημάτων πραγματοποιείται στο χαμηλότερο επίπεδο ανάλυσης  $I(L_0)$  [71] όπου και ο συνολικός αριθμός τμημάτων είναι ο μικρότερος. Στα υπόλοιπα επίπεδα ανάλυσης ουσιαστικά προσδιορίζονται οι συνοριακές περιοχές κάθε τμήματος με μεγαλύτερη ακρίβεια, χωρίς να επηρεάζεται το εσωτερικό του, έτσι είναι αναμενόμενο ο M-RSST να είναι σημαντικά ταχύτερος από τον RSST.

### 3.3.2 Κατάτμηση με Βάση το Χρώμα

Η στερεοσκοπική ακολουθία αποτελείται από δύο ακολουθίες, μία για το αριστερό και μία για το δεξί κανάλι, οι οποίες απέχουν μεταξύ τους  $1/50sec$ . Όπως προαναφέρθηκε οι κατατμήσεις γίνονται στο αριστερό κανάλι, και το δεξί χρησιμοποιείται μόνο για την κατάτμηση βάθους. Συνεπώς, για την κατάτμηση χρώματος, ο αλγόριθμος M-RSST εφαρμόζεται απ'ευθείας στο αριστερό κανάλι. Χρησιμοποιώντας το χρωματικό χώρο RGB για την αναπαράσταση των χρωματικών συνιστωσών, η απόσταση δύο γειτονικών περιοχών (segments)  $S_1$  και  $S_2$  ορίζεται ως:

$$d^C(S_1, S_2) = \left[ (R_{S_1} - R_{S_2})^2 + (G_{S_1} - G_{S_2})^2 + (B_{S_1} - B_{S_2})^2 \right]^{1/2} \frac{A_{S_1} A_{S_2}}{A_{S_1} + A_{S_2}} \quad (3.1)$$

όπου  $R_{S_i}$ ,  $G_{S_i}$  και  $B_{S_i}$  είναι οι μέσες χρωματικές συνιστώσες R,G,B στο εσωτερικό της περιοχής  $S_i$  και  $A_{S_i}$  είναι ο αριθμός των pixels εντός της περιοχής.

### 3.3.3 Κατάτμηση με Βάση την Κίνηση

Για το πρόβλημα της εκτίμησης κίνησης (motion estimation) και κατάτμησης ακολουθιών εικόνων με βάση την κίνηση έχουν προταθεί αρκετές διαφορετικές τεχνικές, όπως τεχνικές που βασίζονται στη διαφορά φωτεινότητας μεταξύ διαδοχικών καρέ, τεχνικές εκτίμησης οπτικής ροής (optical flow), καθώς και τεχνικές ταυτόχρονης εκτίμησης και κατάτμησης κίνησης [68, 86]. Για το πρόβλημα της εκτίμησης κίνησης (motion estimation) και κατάτμησης ακολουθιών εικόνων με βάση την κίνηση έχουν προταθεί αρκετές διαφορετικές τεχνικές, όπως τεχνικές που βασίζονται στη διαφορά φωτεινότητας μεταξύ διαδοχικών καρέ, τεχνικές εκτίμησης οπτικής ροής (optical flow), καθώς και τεχνικές ταυτόχρονης εκτίμησης και κατάτμησης κίνησης

Στη συγκεκριμένη προσέγγιση χρησιμοποιείται ένας αλγόριθμος block matching για την κατασκευή του πεδίου κίνησης. Σε περιοχές χαμηλής πληροφορίας

υφής, το παραγόμενο πεδίο κίνησης μπορεί να είναι ιδιαίτερα θορυβώδες, επομένως είναι απαραίτητο ένα βήμα ύστερης επεξεργασίας του πεδίου κατά το οποίο εφαρμόζεται το φίλτρο median στα διανύσματα κίνησης λόγω της ταχύτητάς του και της ικανότητάς του να διατηρεί τα όρια των αντικειμένων.

Η κατάτμηση με βάση την κίνηση πραγματοποιείται διαιρώντας κάθε καρέ σε περιοχές ομοιόμορφης κίνησης. Χρησιμοποιείται και πάλι ο αλγόριθμος M-RSST, ενώ για τον υπολογισμό αποστάσεων μεταξύ περιοχών χρησιμοποιούνται διαφορές διανυσμάτων κίνησης αντί για χρωματικές διαφορές. Για τους σκοπούς της κατάτμησης με βάση την κίνηση η απόσταση  $d^M(S_1, S_2)$  μεταξύ δύο γειτονικών περιοχών  $X$  και  $Y$  ορίζεται, ως το γινόμενο της Ευκλείδειας απόστασης των μέσων διανυσμάτων κίνησης των δύο περιοχών και ενός όρου που είναι συνάρτηση του μεγέθους των δύο περιοχών σε pixels και ευνοεί τη συγχώνευση των μικρών περιοχών.

$$d^M(S_1, S_2) = \left[ (X_{S_1} - X_{S_2})^2 + (Y_{S_1} - Y_{S_2})^2 \right]^{1/2} \frac{A_{S_1} A_{S_2}}{A_{S_1} + A_{S_2}} \quad (3.2)$$

όπου  $X_{S_i}, Y_{S_i}$  οι μέσες συντεταγμένες  $x, y$  των διανυσμάτων κίνησης της περιοχής  $S_i$ .

### 3.3.4 Κατάτμηση με Βάση το Πεδίο Βάθους

Οι τρισδιάστατες ακολουθίες επιτρέπουν τη αποδοτικότερη διαχείριση των αντικειμένων με την εκμετάλλευση της πληροφορίας βάθους που προκύπτει από τη στερεοσκοπική ανάλυση. Επιπλέον, το πρόβλημα της κατάτμησης σε αντικείμενα αντιμετωπίζεται με μεγαλύτερη ακρίβεια, αφού τα αντικείμενα αποτελούνται συνήθως από περιοχές εικόνας που ανήκουν στο ίδιο επίπεδο βάθους, επομένως οι τρισδιάστατες ακολουθίες επιτρέπουν τη αποδοτικότερη διαχείριση των αντικειμένων με την εκμετάλλευση της πληροφορίας βάθους που προκύπτει από τη στερεοσκοπική ανάλυση. [67, 72].

Όταν μία τρισδιάστατη σκηνή προβάλλεται στο επίπεδο μιας εικόνας μέσω μιας κάμερας, η πληροφορία του βάθους χάνεται, αφού η απεικόνιση αυτή είναι ένας μετασχηματισμός από τρισδιάστατο χώρο σε δισδιάστατο. Συγκεκριμένα, μέσω της προοπτικής προβολής, κάθε 2-Δ σημείο της εικόνας αντιστοιχεί στο σύνολο όλων των συγγραμικών σημείων του 3-Δ χώρου που ανήκουν στη ευθεία η οποία διέρχεται από το σημείο αυτό και από το εστιακό σημείο της κάμερας [72]. Αν δεν είναι διαθέσιμη κάποια επιπλέον πληροφορία σχετικά με το πραγματικό 3-Δ σημείο που αντιστοιχεί στο σημείο της εικόνας, δεν είναι δυνατή η υλοποίηση ενός αντίστροφου μετασχηματισμού για την πλήρη ανάκτηση του 3-Δ σημείου.

Οι στερεοσκοπικές ακολουθίες αποτελούνται από δύο κανάλια, οι οποίες λαμβάνονται από δύο κάμερες οι οποίες παρέχουν δύο απόψεις της ίδιας σκηνής. Αν η αντιστοιχία μεταξύ δύο 2-Δ σημείων  $(x_1, y_1)$  και  $(x_2, y_2)$  είναι γνωστή, αν δηλαδή είναι γνωστό ότι προέρχονται και τα δύο από την προοπτική προβολή του ίδιου 3-Δ σημείου στα επίπεδα των δύο εικόνων, τότε η εκτίμηση του βάθους είναι δυνατή. Στο πλαίσιο αυτής της διατριβής χρησιμοποιήθηκε η τεχνική που προτάθηκε στο [73] για την κατασκευή ενός χάρτη απόκλισης από δύο εικόνες. Η κατάτμηση βάθους γίνεται με εφαρμογή του αλγορίθμου M-RSST στο χάρτη απόκλισης [84] χρησιμοποιώντας την ακόλουθη απόσταση :

$$d^D(S_1, S_2) = \left| D_{S_1} - D_{S_2} \right| \frac{A_{S_1} A_{S_2}}{A_{S_1} + A_{S_2}} \quad (3.3)$$

όπου  $D_{S_i}$  η μέση τιμή απόκλισης στο εσωτερικό της περιοχής  $S_i$ .

### 3.4 Ασαφής Συγχώνευση Κατατμήσεων

Έχουν εξαχθεί ως τώρα οι περιοχές χρώματος, κίνησης και βάθους χρησιμοποιώντας τον αλγόριθμο M-RSST. Το επόμενο βήμα είναι η ένωση των περιοχών χρώματος και κίνησης εφόσον αυτές προσδιορίζουν τα όρια των αντικειμένων με καλή ακρίβεια. Επιπλέον οι περιοχές βάθους παρέχουν έναν λιγότερο λεπτομερή διαχωρισμό των σημασιολογικών αντικειμένων επειδή συνήθως τα αντικείμενα απαρτίζονται από περιοχές που βρίσκονται σε παρόμοιο βάθος. Χρησιμοποιώντας λοιπόν τις περιοχές βάθους ως περιοχές συγκράτησης, προβάλλονται σε αυτές οι αντίστοιχες περιοχές χρώματος και κίνησης. Βρίσκεται εν συνεχεία η αντίστοιχη περιοχή βάθους στην οποία ανήκε κάθε περιοχή χρώματος και συνεχώνονται οι περιοχές χρώματος ιδίου βάθους σε μια νέα μεγαλύτερη περιοχή.

Η μονάδα ασαφούς συγχώνευσης κατατμήσεων (ΑΣΚ) δέχεται ως είσοδο το χρώμα, την κίνηση, το βάθος και την κινησιολογικά αντισταθμισμένη εικόνα της προηγούμενης κατάτμησης και παρέχει στην έξοδο το τελικό προϊόν της συγχώνευσης των κατατμήσεων.

Παρέχεται στη συνέχεια η τυπική μαθηματική περιγραφή του αλγορίθμου.

$$\begin{aligned} S^C &= \{s_1^C, s_2^C, \dots, s_N^C\} \\ S^M &= \{s_1^M, s_2^M, \dots, s_K^M\} \\ S^D &= \{s_1^D, s_2^D, \dots, s_L^D\} \\ S^P &= \{s_1^P, s_2^P, \dots, s_Q^P\} \end{aligned}$$

όπου οι είσοδοι είναι οι κατατμήσεις (segmentations) χρώματος  $S^C$ , κίνησης

$S^M$ , βάθους  $S^D$ , και αυτή του προηγούμενου καρέ  $S^P$ , ενώ με

$$S^F = \{s_1^F, s_2^F, \dots, s_U^F\}$$

συμβολίζεται η έξοδος (τελική κατάτμηση συγχώνευσης). Για λόγους απλούστευσης δεν συμβολίζεται η παράμετρος του χρόνου, δηλαδή υπονοείται χρόνος  $t$  εκτός και αν αναφέρεται διαφορετικά. Οι παραπάνω κατατμήσεις αποτελούν ουσιαστικά *σαφείς* (*crisp*) διαμερίσεις των εικόνων.

Η μονάδα ΑΣΚ προσπαθεί να εκμεταλλευτεί την πληροφορία των διαμερίσεων  $S^C, S^M, S^D$  και  $S^P$  προκειμένου να βελτιώσει την ακρίβεια και το σημασιολογικό πλαίσιο της κατάτμησης. Εφόσον η κατάτμηση χρώματος παρέχει τα πιο ακριβή όρια είναι λογικό να χρησιμοποιηθεί η  $S^C$  ως βάση για την κατασκευή της  $S^F$ . Επιπλέον υποτίθεται ότι η  $S^C$  παρέχει μια υπερκατατμημένη διαμέριση των εικόνων. Επομένως η ΑΣΚ ουσιαστικά παράγει μια πιο «ενοποιημένη» διαμέριση  $S^F$  σε σχέση με την  $S^C$ , επομένως ισχύει:

$$\forall i \in \mathbb{N}_N \quad \text{και} \quad j \in \mathbb{N}_U \quad \text{έχουμε} \quad s_i^C \subseteq s_j^F$$

όπου

$$N_a = 1, 2, \dots, a.$$

Για τον υπολογισμό της κατάτμησης  $S^F$  από τις  $S^C, S^M, S^D$  και  $S^P$  χρησιμοποιείται ο αλγόριθμος M-RSST. Συνοπτικά, λαμβάνονται όλα τα δυνατά ζεύγη περιοχών που παρέχονται από την  $S^C$  και υπολογίζονται οι αποστάσεις τους. Στη συνέχεια τα δύο κοντινότερα ενώνονται (εάν η απόστασή τους είναι μικρότερη από ένα κατώφλι) και η διαδικασία επαναλαμβάνεται μέχρι να μην ενωθούν νέες περιοχές. Προφανώς η ευφυΐα της μεθόδου εξαρτάται από τον τρόπο μέτρησης των παραπάνω αποστάσεων. Προκειμένου να περιγραφεί η διαδικασία αναλυτικά, ορίζονται αρχικά ορισμένοι τελεστές.

Για κάθε ζεύγος περιοχών  $s_1, s_2$  ο τελεστής προβολής  $R(s_1, s_2)$  ορίζεται ως:

$$R(s_1, s_2) = \frac{|s_1 \cap s_2|}{|s_1|} \quad (3.4)$$

όπου με  $|\cdot|$  συμβολίζεται ο πληθάριθμος του συνόλου, στην περίπτωση αυτή ο αριθμός των pixels της περιοχής.

Δεδομένης της περιοχής  $s$  και ενός συνόλου περιοχών  $S$ , ο τελεστής θ-γειτνίασης  $I_\theta(s, S)$  ορίζεται ως:

$$I_\theta(s, S) = \{i \in N_{|S|} : |R(s, s_i) - \max_{j \in N_{|S|}} R(s, s_j)| \leq \theta\}$$

όπου  $\theta$  είναι ένα κατώφλι και  $R(\cdot, \cdot)$  ο τελεστής προβολής που ορίζεται στην εξίσωση (3.4).

Η ομοειδής απόσταση δύο περιοχών, δηλαδή η απόσταση δύο περιοχών του ίδιου τύπου εισόδου, ορίζεται σε οποιαδήποτε είσοδο με τη βοήθεια του τελεστή  $\theta$ -γειτνίασης. Για παράδειγμα η απόσταση κίνησης  $D^M(s_i^C, s_j^C)$  δύο περιοχών χρώματος  $s_i^C, s_j^C, (i, j \in \mathbb{N}_N)$  ορίζεται ως:

$$D^M(s_i^C, s_j^C) = a \cdot b \cdot \delta_1 - a \cdot b \cdot \delta_2$$

όπου

$$a = \max_{k \in \mathbb{N}_K} R(s_i^C, s_k^M)$$

$$b = \max_{k \in \mathbb{N}_K} R(s_j^C, s_k^M)$$

$$\delta_1 = \delta(|I_\theta(s_i^C, S^M) \cap I_\theta(s_j^C, S^M)|)$$

$$\delta_2 = \delta(|[I_\theta(s_i^C, S^M) \cup I_\theta(s_j^C, S^M)] - [I_\theta(s_i^C, S^M) \cap I_\theta(s_j^C, S^M)]|)$$

$$\delta = \begin{cases} 1, & \text{if } x \neq 0 \\ 0, & \text{if } x = 0 \end{cases}$$

Η απόσταση  $D(s_i^C, s_j^C)$  μεταξύ δύο περιοχών χρώματος  $s_i^C$  και  $s_j^C$ , που χρησιμοποιείται στη διαδικασία κατάτμησης M-RSST της μονάδας ΑΣΚ, προκύπτει από τις αποστάσεις

$$\text{κίνησης } (D^M(s_i^C, s_j^C))$$

$$\text{βάθους } (D^D(s_i^C, s_j^C))$$

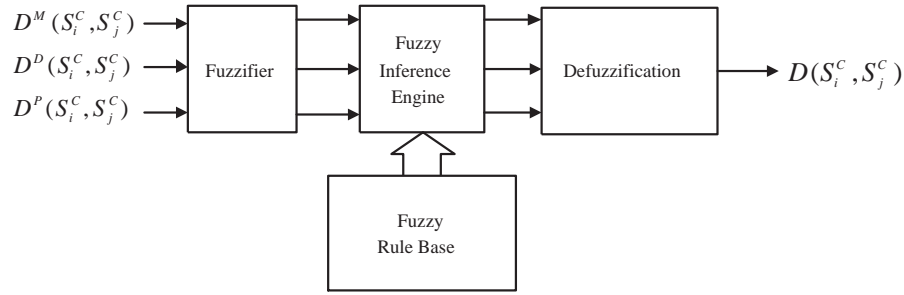
$$\text{προηγούμενου καρέ } (D^P(s_i^C, s_j^C))$$

οι οποίες υπολογίζονται με την παραπάνω διαδικασία. Αυτό αποτελεί στην ουσία το «έξυπνο» τμήμα της μονάδας ΑΣΚ το οποίο παρέχει στο σύστημα την ικανότητα να συγχωνεύει διαφορετικά είδη εισόδου χρησιμοποιώντας εκ των προτέρων γνώση με τη μορφή ασαφών γλωσσικών κανόνων [78]. Στο σχήμα 3.2 παριστάνεται η δομή του συστήματος ασαφούς συμπερασμού το οποίο υλοποιεί την παραπάνω ιδέα. Το σύστημα δέχεται τρεις εισόδους: τις αποστάσεις που βασίζονται στην κίνηση και στο βάθος και τις αποστάσεις του προηγούμενου καρέ. Οι ασαφείς τιμές (μεταξύ -1 και 1) παριστάνουν το βαθμό κατά τον οποίο μια συγκεκριμένη είσοδος «συνιστά» στο ΑΣΚ να παραγάγει υψηλή τιμή «ενότητας» μεταξύ δύο περιοχών χρώματος (έξοδος).

Οι γλωσσικοί κανόνες που χρησιμοποιούνται αντιστοιχίζουν τις μεταβλητές εισόδου σε αυτές της εξόδου, για παράδειγμα τις διαμερίσεις

$$D^M(s_i^C, s_j^C), D^D(s_i^C, s_j^C) \quad \text{και} \quad D^P(s_i^C, s_j^C)$$





Σχήμα 3.2: Συγχωνευτικό Ασαφές Σύστημα Συμπερασμού

στη διαμέριση  $D(s_i^C, s_j^C)$  (σχήμα 3.3). Στον πίνακα 3.2 συνοψίζεται το σύνολο των κανόνων. Τέλος στο σχήμα 3.4 παριστάνεται η σχηματική επιφανειακή αναπαράσταση της αντιστοιχίας εισόδου εξόδου.

## 3.5 Αποτελέσματα Κατάτμησης

### 3.5.1 Εξοπλισμός

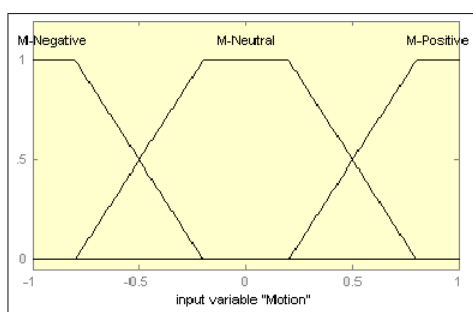
Ο αλγόριθμος δοκιμάστηκε σε πραγματικές τρισδιάστατες ακολουθίες εικόνων. Οι ακολουθίες λήφθηκαν με το φακό «SX2000 Stereo-Optix» της *NuView* συνδεδεμένο σε μια απλή MiniDV κάμερα. Ο φακός αυτός προβάλλει δύο εικόνες στον κανονικό φακό της κάμερας μέσω δύο καθρεπτών. Οι δύο εικόνες δεν καταγράφονται ταυτόχρονα αλλά διαδοχικά λόγω της πεπλεγμένης μορφής του σήματος βίντεο. Οι ζυγές γραμμές του παραγόμενου βίντεο αντιστοιχούν στην αριστερή όψη και οι μονές στη δεξιά. Η εστιακή απόσταση ρυθμίζεται μέσω μηχανικών μέσων. Δυστυχώς η μέθοδος αυτή μειώνει την κάθετη ανάλυση στο μισό ( $720 \times 288$ ), άλλα το αποτέλεσμα κρίνεται αρκετά ικανοποιητικό για τη χρήση που προορίζεται.

Αρχικά αναλύονται οι παραγόμενες στερεοσκοπικές ακολουθίες και για κάθε ζεύγος καρέ (αριστερό-δεξιό) εκτιμάται ένας χάρτης απόκλισης. Με εφαρμογή του αλγορίθμου M-RSST στο χάρτη αυτό παράγεται η κατάτμηση βάθους. Οι κατατμήσεις χρώματος και κίνησης παράγονται από την εικόνα αριστερής όψης.

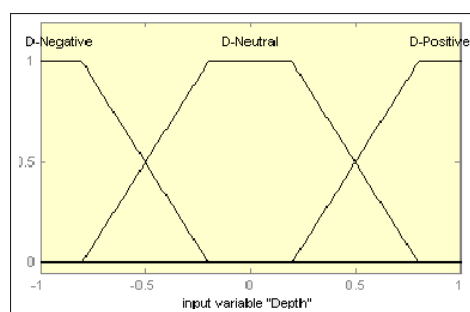
### 3.5.2 Χρώμα

Εφόσον θέλουμε τα όρια της κατάτμησης χρώματος να διαθέτουν όρια ακριβείας, επιλέγουμε την κατάτμηση σε ανάλυση επιπέδου 1 (διαστάσεων  $2 \times 2$

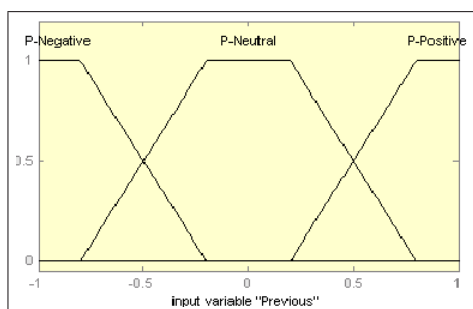
### Κεφάλαιο 3. Επεκτάσεις με Βάση την Ανάλυση Χειρονομιών



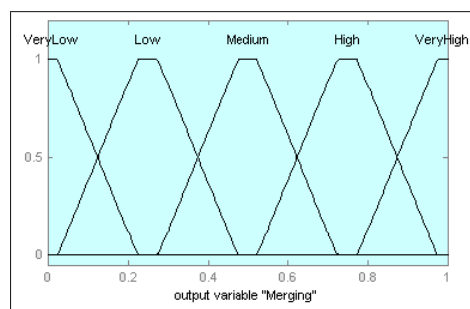
Κίνηση



Βάθος

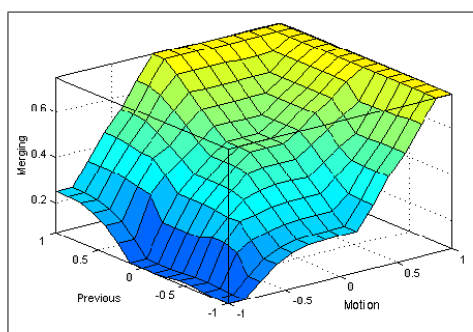


Προηγ. Καρέ

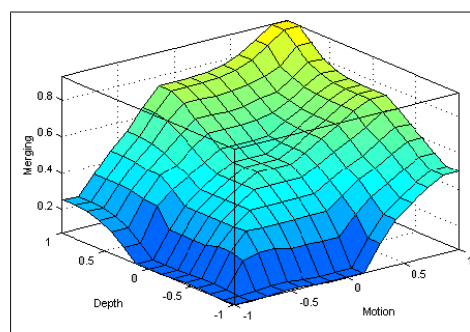


Συγχώνευση

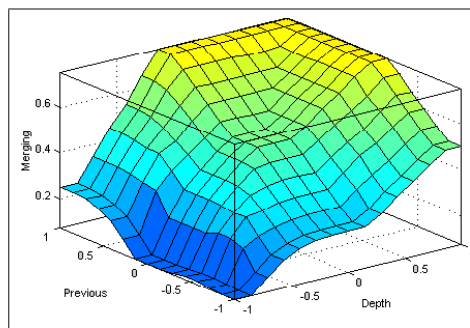
Σχήμα 3.3: Ορισμός των ασαφών καταταμήσεων



Κίνηση - Προηγ. Καρέ



Κίνηση - Βάθος



Βάθος - Προηγ. Καρέ

Σχήμα 3.4: Σχηματική αναπαράσταση των συσχετίσεων εισόδου-εξόδου

If Motion is	and Depth is	and Previous is	then Merging is
M-Negative	D-Negative	P-none	VeryLow
M-Negative	D-Neutral	P-Negative	VeryLow
M-Negative	D-Neutral	P-Neutral	VeryLow
M-Negative	D-Neutral	P-Positive	Low
M-Negative	D-Positive	P-Negative	VeryLow
M-Negative	D-Positive	P-Neutral	Low
M-Negative	D-Positive	P-Positive	Medium
M-Neutral	D-Negative	P-Negative	VeryLow
M-Neutral	D-Negative	P-Neutral	VeryLow
M-Neutral	D-Negative	P-Positive	Low
M-Neutral	D-Neutral	P-Negative	Low
M-Neutral	D-Neutral	P-Neutral	M-Positive
M-Neutral	D-Neutral	P-Positive	High
M-Neutral	D-Positive	P-Negative	Medium
M-Neutral	D-Positive	P-Neutral	High
M-Neutral	D-Positive	P-Positive	High
M-Positive	D-Negative	P-Negative	Low
M-Positive	D-Negative	P-Neutral	Medium
M-Positive	D-Negative	P-Positive	Medium
M-Positive	D-Neutral	P-none	High
M-Positive	D-Positive	P-none	VeryHigh

**Πίνακας 3.2:** Το σύνολο των γλωσσικών κανόνων του ΑΣΚ

pixels) στον M-RSST. Επιπλέον επιθυμούμε την υπερκατάτμηση της εικόνας, συνεπώς επιλέγουμε ένα πολύ χαμηλό κατώφλι για τη διαδικασία ένωσης περιοχών. Στη συγκεκριμένη περίπτωση επιλέχθηκε κατώφλι απόστασης 0.05%, στο χρωματικό χώρο RGB αλλά οποιοδήποτε χαμηλό κατώφλι θα παρήγαγε παρόμοια αποτελέσματα. Ένα αποτέλεσμα κατάτμησης παριστάνεται σε μεγέθυνση στην εικόνα 3.5 όπου σημειώνονται και τα όρια των περιοχών.

### 3.5.3 Κίνηση

Για τον υπολογισμό των διανυσμάτων κίνησης χρησιμοποιείται ένας αλγόριθμος block-matching. Η περιοχή αναζήτησης στη συγκεκριμένη περίπτωση ορίστηκε στα 10 pixels και στις δύο κατευθύνσεις, με block μεγέθους  $3 \times 3$  pixels. Παρόλο το μικρό μέγεθος block το πεδίο κίνησης είναι αρκετά ακριβές. Κατά τη διαδικασία ταιριάσματος των blocks, εάν η ελάχιστη απόσταση ενός block από το υπό δοκιμή block είναι λιγότερο από 20% μεγαλύτερη από το block μηδενικής κίνησης τότε θεωρείται ακίνητο. Στο τέλος εφαρμόζεται το φίλτρο ενδιάμεσης τιμής στα διανύσματα κίνησης, πριν τη διαδικασία κατάτμησης. Η τελευταία πραγματοποιείται με μέγιστο επίπεδο ανάλυσης  $L_0 = 3$  (ανάλυση block  $8 \times 8$  pixels) για την επιτάχυνση της διαδικασίας.

### 3.5.4 Βάθος

Ο χάρτης απόκλισης (disparity map) εκτιμάται από τις εικόνες αριστερής και δεξιάς όψης με τρόπο παρόμοιο με τον υπολογισμό των διανυσμάτων κίνησης αλλά υποθέτοντας μόνο οριζόντια μετατόπιση. Ο υπολογισμός του βάθους από το χάρτη απόκλισης δεν προσφέρει επιπλέον πληροφορία στην κατάτμηση και ο χάρτης απόκλισης είναι αρκετός. Η κατάτμηση γίνεται χρησιμοποιώντας ανάλυση  $4 \times 4$  pixels.

### 3.5.5 Συγχώνευση

Η διαδικασία συγχώνευσης χρησιμοποιεί τις κατατμήσεις χρώματος, κίνησης, βάθους και την κατάτμηση του προηγούμενου καρέ ως είσοδο. Για το πρώτο καρέ όπου δεν υπάρχει προηγούμενη κατάτμηση χρησιμοποιείται ως είσοδος ένας λευκός χάρτης. Η συγχώνευση επενεργεί στις περιοχές χρώματος λαμβάνοντας υπόψη της πλέον τις ασαφείς αποστάσεις των περιοχών αυτών.

### 3.5.6 Αποτελέσματα

Ένα παράδειγμα της διαδικασίας συγχώνευσης απεικονίζεται στην εικόνα 3.5, όπου φαίνεται επίσης το αποτέλεσμα των επιμέρους κατατμήσεων χρώματος κίνησης και βάθους, καθώς και η κατάτμηση του προηγούμενου καρέ. Περισσότερα παραδείγματα βρίσκονται στις εικόνες 3.12–3.11 για δύο ακολουθίες βίντεο.

Στην πρώτη ακολουθία κινείται μόνο το χέρι. Στις εικόνες 3.6–3.11 φαίνεται ένα παράδειγμα 6 καρέ. Όπως φαίνεται κατασκευάζονται μέχρι δύο περιοχές κίνησης στην παλάμη και στο χέρι οι οποίες όταν συνενώνονται με την πληροφορία βάθους και χρώματος παράγονται οι τελικές εικόνες. Το τμήμα της εικόνας κάτω από το δεξί χέρι (κάτω αριστερά) θεωρείται ότι βρίσκεται σε διαφορετική περιοχή από το υπόλοιπο φόντο επειδή οι αντίστοιχες περιοχές χρώματος δεν είναι γειτονικές: χωρίζονται από τις περιοχές χρώματος του χεριού. Στην εικόνα 3.6 ανιχνεύθηκαν δύο περιοχές κίνησης λόγω της μεγαλύτερης γραμμικής ταχύτητας της παλάμης. Εξαιτίας αυτού, η παλάμη ανατίθεται σε διαφορετική περιοχή στο αποτέλεσμα της συγχώνευσης (εικ. 3.6) και αυτός ο διαχωρισμός μεταβιβάζεται και στις επόμενες κατατμήσεις. Εάν αυτός ο διαχωρισμός είναι επιθυμητός, ο αλγόριθμος M-RSST μπορεί να ρυθμιστεί ώστε να είναι πιο ευαίσθητος και να παράγει περισσότερες περιοχές. Εάν η συμπεριφορά αυτή δεν είναι επιθυμητή θα μπορούσε να χρησιμοποιηθεί ένα πλήρες παραμετρικό μοντέλο κίνησης για την ανίχνευση της κυκλικής κίνησης και τη διατήρηση του μπράτσου και της παλάμης σε μία περιοχή.

Εξαιτίας της ανακρίβειας της κατάτμησης βάθους, μπορεί κάποιος να παρατηρήσει μια επιπλέον περιοχή στην εικόνα 3.9 κοντά στην παλάμη (με μαύρο χρώμα). Αυτή η περιοχή δεν συγχωνεύτηκε με την υπόλοιπη παλάμη εξαιτίας της προηγούμενης κατάτμησης και της κατάτμησης κίνησης και επομένως δε διαδόθηκε στα υπόλοιπα καρέ.

Στη δεύτερη ακολουθία η κάμερα είναι ακίνητη και το άτομο κινείται από αριστερά προς τα δεξιά με το αριστερό του χέρι προτεταμένο. Στις εικόνες 3.12–3.17 φαίνεται ένα παράδειγμα 6 συνεχόμενων καρέ. Σε αυτό το δείγμα η κατάτμηση χρώματος παράγει περίπου 140–150 περιοχές, η κατάτμηση κίνησης δύο, μία για το κινούμενο άτομο και μία για το φόντο, ενώ η κατάτμηση βάθους παρέχει τρεις με πέντε περιοχές: μία για το φόντο, μία για το σώμα, μία για το χέρι και επιπλέον ορισμένες λανθασμένες ειδικά σε περιοχές με μικρή πληροφορία σε οριζόντιες ακμές.

Τα περιγράμματα στο αποτέλεσμα της συγχώνευσης του παραδείγματος είναι μεγαλύτερης ακρίβειας από τα περιγράμματα των ξεχωριστών πηγών. Αυτό

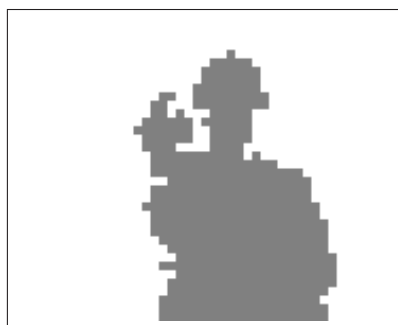
δεν οφείλεται στην υψηλότερη ανάλυση της κατάτμησης χρώματος αλλά στη συγχώνευση των κατατμήσεων χρώματος και βάθους· μια κατάτμηση βάθους σε υψηλότερη ανάλυση δεν θα είχε την ακρίβεια ορίων της κατάτμησης χρώματος. Ο άνθρωπος στο αποτέλεσμα της συγχώνευσης διαχωρίστηκε επιτυχώς από το φόντο. Το χέρι του, έχοντας διαφορετικό βάθος από το σώμα χωρίστηκε σε διαφορετική περιοχή. Στο καρέ 59 της εικόνας 3.16 το χέρι διαχωρίστηκε σε δύο περιοχές εξαιτίας της κάθετης επέκτασης του τμήματος του χεριού της περιοχής βάθους.

### **3.6 Επεκτάσεις**

Εκτός από την ανάλυση χειρονομιών, δυο πεδία στα οποία μπορεί να στραφεί η περαιτέρω έρευνα για την βελτίωση της εκτίμησης συναισθήματος είναι η αξιοποίηση πληροφορίας σχετιζόμενης με τη χρονική εξέλιξη των FAP και η επέκταση των συστημάτων μάθησης με προσωδιακά χαρακτηριστικά της φωνής. Στην εικόνα 3.18 παρουσιάζονται δύο καρέ που ανήκουν στο ίδιο tune και απέχουν κατά 11 καρέ. Στο μέσο αριστερά εμφανίζεται η εξέλιξη των τιμών των FAP που σχετίζονται με το στόμα, και στο πράσινο πλαίσιο προσωδιακά χαρακτηριστικά του tune στο οποίο ανήκει το απεικονιζόμενο καρέ, τα οποία σχετίζονται με την τονικότητα της φωνής. Στον κύκλο σημειώνεται με κίτρινη κουκκίδα η έξοδος του νευροασαφούς συστήματος αυτόματης εκτίμησης συναισθήματος, ενώ οι άλλες τέσσερις κουκκίδες αντιστοιχούν στις εκτιμήσεις των τεσσάρων χαρακτηριστών. Η παρουσιαζόμενη πληροφορία δεν είναι προς το παρόν εκμεταλλεύσιμη προτού δοθεί λύση σε μια σειρά από προβλήματα. Όσον αφορά την ανάλυση της ομιλίας, η κατάτμηση σε tunes με βάση τις παύσεις δεν επαρκεί για την εξαγωγή τονικών χαρακτηριστικών, αλλά θα ήταν προτιμότερη η κατάτμηση σε αυτοτελείς νοηματικές φράσεις. Όσον αφορά την ανάλυση της χρονικής εξέλιξης των FAP, πρέπει να οριστεί το χρονικό παράθυρο της μελέτης τους. Κάποια FAP όπως π.χ. αυτά που αντιστοιχούν στο ανασήκωμα των φρυδιών, μπορούν να μεταβάλλονται πολύ γρήγορα, γιατί μερικές φορές χρησιμοποιούνται για τη μετάδοση κάποιου μηνύματος, ενώ τα ίδια FAP μπορεί αν μελετηθούν σε μεγαλύτερο παράθυρο να παρουσιάσουν μια αργή μεταβολή που θα υποδεικνύει την ύπαρξη κάποιου συναισθήματος. Η επιλογή των χρονικών παραθύρων μελέτης των FAP, τα οποία μπορεί να είναι πολλαπλά και αλληλοεπικαλυπτόμενα, καθώς και ο συγχρονισμός τους με τα χαρακτηριστικά της φωνής αποτελούν ανοικτά ερευνητικά ζητήματα.



Κατάτμηση Χρώματος



Κατάτμηση Κίνησης



Κατάτμηση Βάθους



Προηγούμενη Κατάτμηση



Αποτέλεσμα Συγχώνευσης

Σχήμα 3.5: Ακολουθία 2, Καρέ 60, Μεγέθυνση

### 3.7 Συμπεράσματα

Σε αυτό το κεφάλαιο εξετάστηκε η ενσωμάτωση πολλαπλών ειδών κατατιμήσεων με τη χρήση ασαφών αποφάσεων για τη βελτίωση της ποιότητας της κατάτιμησης. Ένας τρόπος περαιτέρω βελτίωσης της ακρίβειας του αλγορίθμου είναι η αύξηση της μνήμης του, η χρήση δηλαδή περισσότερων του ενός προηγούμενων καρτέ, μειώνοντας με αυτόν τον τρόπο τη διάδοση σφαλμάτων. Η ασάφεια θα μπορούσε επίσης να εισαχθεί στις ίδιες τις διαμερίσεις έτσι ώστε η μετάβαση μεταξύ γειτονικών περιοχών να είναι σταδιακή. Επιπλέον ασάφεια θα μπορούσε να εισαχθεί και στην κατασκευή των ίδιων των δέντρων διαμερίσεων και στα βάρη των συνδέσμων τους έτσι ώστε να δοθεί η δυνατότητα για γλωσσική υπόσταση στους περιγραφείς χαμηλού επιπέδου (low level descriptors) παρέχοντας την αριθμητική ερμηνεία των συμπερασματικών κανόνων.

Όσον αφορά τη χρήση των ευρεθεισών περιοχών σε ένα σύστημα ανάλυσης χειρονομιών, μια πιθανή επέκταση είναι η δυναμική προσαρμογή των κανόνων του ΑΣΚ με τη σύνδεσή τους με περιγραφές αντικειμένων ανώτερου επιπέδου που μπορούν να περιλαμβάνουν χαρακτηριστικά όπως το σχήμα. Τα χαρακτηριστικά αυτά μπορούν να ληφθούν υπόψη στη συνάρτηση κόστους είτε κατά τη διαδικασία συγχώνευσης, είτε μετά το πέρας της για τη συνέχιση της συγχώνευσης με στόχο το ταίριασμα μοντέλων [103] στις προκύπτουσες διαμερίσεις. Έτσι εκτός από το χρώμα δέρματος, την κίνηση και το βάθος μπορεί να δοθεί η δυνατότητα εκμετάλλευσης της εξαγόμενης πληροφορίας για την εκτίμηση των παραμέτρων αρθρωτών ανθρώπινων μοντέλων για την ανάλυση χειρονομιών.



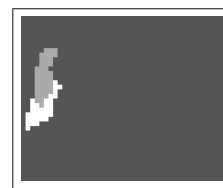
Κεφάλαιο 3. Επεκτάσεις με Βάση την Ανάλυση Χειρονομιών



Καρέ 30



Κατάτμηση Χρώματος



Κατάτμηση Κίνησης



Κατάτμηση Βάθους



Προηγούμενη Κατάτμηση

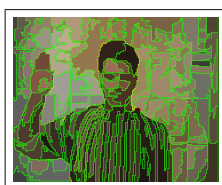


Αποτέλεσμα Συγχώνευσης

Σχήμα 3.6: Ακολουθία 1, Καρέ 30



Καρέ 31



Κατάτμηση Χρώματος



Κατάτμηση Κίνησης



Κατάτμηση Βάθους



Προηγούμενη Κατάτμηση



Αποτέλεσμα Συγχώνευσης

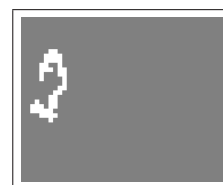
Σχήμα 3.7: Ακολουθία 1, Καρέ 31



Καρέ 32



Κατάτμηση Χρώματος



Κατάτμηση Κίνησης



Κατάτμηση Βάθους



Προηγούμενη Κατάτμηση



Αποτέλεσμα Συγχώνευσης

Σχήμα 3.8: Ακολουθία 1, Καρέ 32

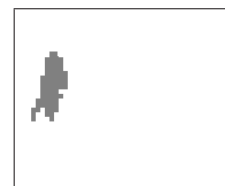
### Κεφάλαιο 3. Επεκτάσεις με Βάση την Ανάλυση Χειρονομιών



Καρέ 33



Κατάτμηση Χρώματος



Κατάτμηση Κίνησης



Κατάτμηση Βάθους



Προηγούμενη Κατάτμηση

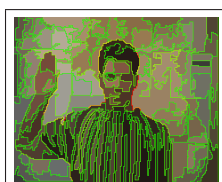


Αποτέλεσμα Συγχώνευσης

Σχήμα 3.9: Ακολουθία 1, Καρέ 33



Καρέ 34



Κατάτμηση Χρώματος



Κατάτμηση Κίνησης



Κατάτμηση Βάθους



Προηγούμενη Κατάτμηση



Αποτέλεσμα Συγχώνευσης

Σχήμα 3.10: Ακολουθία 1, Καρέ 34



Καρέ 35



Κατάτμηση Χρώματος



Κατάτμηση Κίνησης



Κατάτμηση Βάθους



Προηγούμενη Κατάτμηση



Αποτέλεσμα Συγχώνευσης

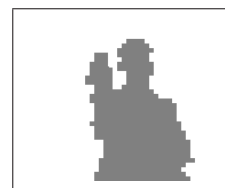
Σχήμα 3.11: Ακολουθία 1, Καρέ 35



Καρέ 55



Κατάτμηση Χρώματος



Κατάτμηση Κίνησης



Κατάτμηση Βάθους



Προηγούμενη Κατάτμηση

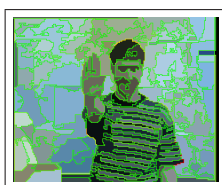


Αποτέλεσμα Συγχώνευσης

Σχήμα 3.12: Ακολουθία 2, Καρέ 55



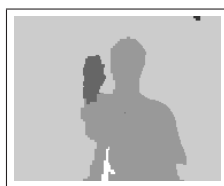
Καρέ 56



Κατάτμηση Χρώματος



Κατάτμηση Κίνησης



Κατάτμηση Βάθους



Προηγούμενη Κατάτμηση

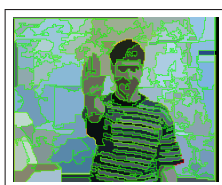


Αποτέλεσμα Συγχώνευσης

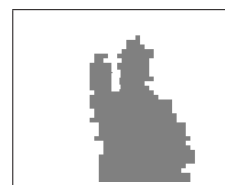
Σχήμα 3.13: Ακολουθία 2, Καρέ 56



Καρέ 57



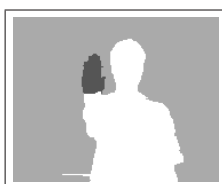
Κατάτμηση Χρώματος



Κατάτμηση Κίνησης



Κατάτμηση Βάθους



Προηγούμενη Κατάτμηση

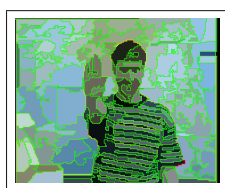


Αποτέλεσμα Συγχώνευσης

Σχήμα 3.14: Ακολουθία 2, Καρέ 57



Καρέ 58



Κατάτμηση Χρώματος



Κατάτμηση Κίνησης



Κατάτμηση Βάθους



Προηγούμενη Κατάτμηση



Αποτέλεσμα Συγχώνευσης

Σχήμα 3.15: Ακολουθία 2, Καρέ 58



Καρέ 59



Κατάτμηση Χρώματος



Κατάτμηση Κίνησης



Κατάτμηση Βάθους



Προηγούμενη Κατάτμηση

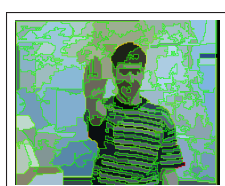


Αποτέλεσμα Συγχώνευσης

Σχήμα 3.16: Ακολουθία 2, Καρέ 59



Καρέ 60



Κατάτμηση Χρώματος



Κατάτμηση Κίνησης



Κατάτμηση Βάθους



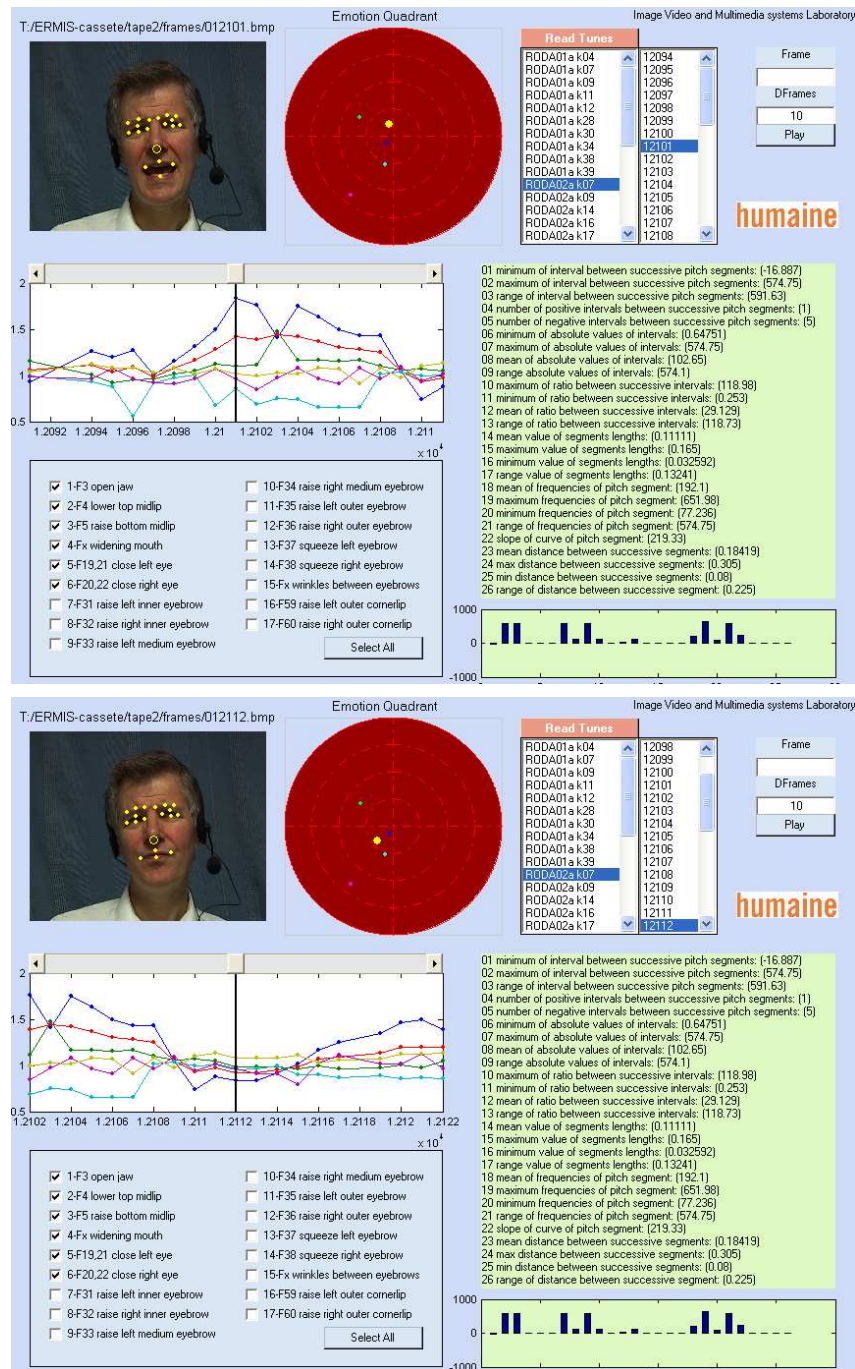
Προηγούμενη Κατάτμηση



Αποτέλεσμα Συγχώνευσης

Σχήμα 3.17: Ακολουθία 2, Καρέ 60

### Κεφάλαιο 3. Επεκτάσεις με Βάση την Ανάλυση Χειρονομιών



Σχήμα 3.18: Χαρακτηριστικά σημεία προσώπου, χρονική εξέλιξη των FAP, εκτιμώμενη και χειρωνακτική κατηγοριοποίηση και προσωδιακά χαρακτηριστικά

### *Κεφάλαιο 3. Επεκτάσεις με Βάση την Ανάλυση Χειρονομιών*

## Κεφάλαιο 4

### Συμπεράσματα

Από τους τρόπους με τους οποίους εκφράζονται τα ανθρώπινα συναισθήματα μέσω του σώματος, μπορούμε να παρατηρήσουμε άμεσα αυτά που σχετίζονται με τη φυσιολογία, με τη φωνή, με τις κινήσεις και με τις εκφράσεις προσώπου. Μεταξύ των παρατηρούμενων συναισθηματικών εκφάνσεων και του μηχανισμού που παράγει τα συναισθήματα αυτά, υπάρχει ακόμα ένα μεγάλο γνωστικό χάσμα, το οποίο θα γεφυρωθεί μελλοντικά με την πρόοδο των γνωστικών θεωριών και της γενετικής. Μέχρι να επιτευχθεί η επιθυμητή σύνδεση της παρατήρησης (top-down approaches) και του μηχανισμού παραγωγής συναισθημάτων (bottom-up approaches), οι εκτιμήσεις μας για την παρατηρούμενη συναισθηματική κατάσταση θα απέχουν από το να είναι ακριβείς, με εξαίρεση τις εκφάνσεις ορισμένων ακραίων συναισθημάτων που εκφράζονται με καθολικό τρόπο.

Στην παρούσα διατριβή προτάθηκαν λύσεις για την εκτίμηση συναισθήματος μέσω της οπτικής παρατήρησης. Όσον αφορά στον τομέα της εκτίμησης των εκφράσεων προσώπου, η συντριπτική πλειοψηφία των προγενέστερων τεχνικών βασίζεται στην σύγκριση των εικόνων με εικόνες-πρότυπα εκφράσεων, αντιμετωπίζοντας το πρόβλημα ως ένα τυπικό πρόβλημα ταιριάσματος προτύπων. Στην παρούσα διατριβή αντιθέτως έγινε προσπάθεια εκτίμησης των εκφράσεων μέσω της χρήσης κανόνων περιγραφής της κίνησης των χαρακτηριστικών του προσώπου. Αυτή η αναλυτική προσέγγιση, επιτρέπει την άμεση επαλήθευση της γνώσης μας, αλλά και εξαγωγή νέας γνώσης για την αναπαράσταση των εκφράσεων. Οι τιμές των πρότερων τμημάτων των κανόνων, αυτών δηλαδή που αναφέρονται στη θέση των χαρακτηριστικών του προσώπου, εκτιμώνται με τεχνικές που προτείνονται στο πρώτο κεφάλαιο της διατριβής. Οι τεχνικές που προτείνονται αφορούν τον εντοπισμό και την εξαγωγή μασκών για τα φρύδια, τη μύτη, τους οφθαλμούς και το στόμα, με ταυτόχρονη εκτίμηση της εγκυρότητας των αποτελεσμάτων με τη μορφή βαθμών εμπιστοσύνης. Οι τεχνικές που



προτείνονται μπορούν να βρουν χρήση εκτός από την εκτίμηση συναισθήματος και σε άλλες εφαρμογές όπως είναι η αναγνώριση προσώπων, ενώ μπορούν να συνδυαστούν και με τεχνικές παρακολούθησης αντικειμένων και ταιριάσματος μοντέλων, λειτουργώντας ως το στάδιο αρχικοποίησης και επανάκαμψης αυτών. Οι τεχνικές είναι υλοποιημένες στο περιβάλλον Matlab, και απαιτούν στην παρούσα υλοποίησή τους, περίπου 30 δευτερόλεπτα για την εξαγωγή όλων των χαρακτηριστικών σε ένα καρέ.

Επιπλέον προτείνονται μέθοδοι αντιμετώπισης των δύο εγγενών αδυναμιών του συστήματος αναγνώρισης εκφράσεων, όπως είναι τα σφάλματα του συστήματος ανίχνευσης χαρακτηριστικών σημείων του προσώπου και οι περιορισμοί που επιβάλλει η αναγκαστική ομαδοποίηση των εκφράσεων κατά τη διατύπωση των κανόνων εκφράσεων.

Για την αντιμετώπιση των λαθών του συστήματος εξαγωγής των χαρακτηριστικών σημείων μπορεί να ληφθεί υπόψη η έξοδος εμπιστοσύνης του συστήματος αυτού και με αυτό το στόχο, προτάθηκε στο δεύτερο κεφάλαιο η αποτίμηση των κανόνων με ένα σύστημα δυνατοτικής αναπαράστασης της γνώσης, το οποίο έχει τη δυνατότητα να θεωρεί τα πρότερα τμήματα των κανόνων εκφράσεων, σύμφωνα με το βαθμό εμπιστοσύνης στην ορθότητα της εξαγωγής των χαρακτηριστικών προσώπου από τα οποία εξαρτώνται. Τα αποτελέσματα είναι ενθαρρυντικά και δείχνουν σαφή βελτίωση της εκτίμησης με τη χρήση αυτής της τεχνικής.

Όσον αφορά στην αντιμετώπιση των περιορισμών που εισάγει η ομαδοποίηση των εκφράσεων σε κανόνες, προτείνεται η βελτίωσή τους λαμβάνοντας υπόψη τις ιδιαιτερότητες στην εκφραστικότητα των εκάστοτε χρηστών του συστήματος αναγνώρισης έκφρασης. Για το λόγο αυτό προτάθηκε η προσαρμογή των κανόνων εκφράσεων σε συγκεκριμένο χρήστη με τη χρήση ενός νευροασαφούς συστήματος αποτίμησης. Τα αποτελέσματα που παρουσιάζονται στο δεύτερο κεφάλαιο είναι ενθαρρυντικά και δείχνουν ότι με την προσαρμογή των ενεργοποιούμενων κατηγορημάτων των κανόνων εισόδου, υπήρξε σαφής βελτίωση στο ποσοστό αναγνώρισης των εκφράσεων του συγκεκριμένου χρήστη.

Στο πλαίσιο της εκτίμησης της εκφραστικότητας ενός ατόμου μέσω της οπτικής παρατήρησης, προτάθηκε ένα σύστημα κατάταξης εικόνων με στόχο την ανάλυση χειρονομιών. Το σύστημα, που περιγράφεται στο τρίτο κεφάλαιο, επιδέχεται ασαφείς σημασιολογικούς κανόνες για τη συγχώνευση πολλαπλών κατατιμήσεων που βασίζονται στο χρώμα, στην κίνηση και στο βάθος. Οι περιοχές που εξάγονται από το σύστημα μπορούν να χρησιμοποιηθούν για την ανάλυση χειρονομιών από συστήματα υψηλότερου επιπέδου. Με την κατάλληλη εισαγωγή νέων κανόνων περιγραφής στο προτεινόμενο σύστημα, οι περιοχές μπορούν να



#### Κεφάλαιο 4. Συμπεράσματα

συνδεθούν με περιγραφές αντικειμένων ανώτερου επιπέδου για την εκτίμηση των παραμέτρων αρθρωτών ανθρώπινων μοντέλων με στόχο την ανάλυση χειρονομιών.

Το δείγμα εικόνων που χρησιμοποιήθηκε στην παρούσα διατριβή, αποτελείτο από μεγάλες ακολουθίες μη επιτηδευμένων εκφράσεων, σε αντίθεση με τη συνήθη πρακτική εκπαίδευσης και δοκιμής σε μικρό δείγμα εικόνων με μικρό εκφραστικό εύρος, γεγονός που δικαιολογεί τα συχνά αναφερόμενα στη βιβλιογραφία ποσοστά επιτυχίας αναγνώρισης εκφράσεων που αγγίζουν το 100%. Ο μεγάλος αριθμός των δυνατών μη ακραίων εκφράσεων, η διαφοροποίησή τους ανάμεσα στα άτομα αλλά και η μη απαραίτητη σύνδεσή τους με κάποιο συναισθηματικό, δημιουργούν εγγενείς περιορισμούς σε οποιαδήποτε προσπάθεια αυτόματης κατηγοριοποίησης αυτών των παρατηρούμενων εκφράσεων. Η πρόοδος της τεχνολογίας με τη βελτίωση της ανάλυσης των εικόνων θα μειώσει αφενός το σφάλμα ανίχνευσης των χαρακτηριστικών σημείων και πιθανώς να επιτρέψει την μοντελοποίηση και εκμετάλλευση και άλλων χαρακτηριστικών όπως ο σχηματισμός ρυτίδων από τις εκφράσεις, αλλά πρέπει να ληφθεί υπόψη ότι ακόμα και ο χειρονακτικός συναισθηματικός χαρακτηρισμός των ακολουθιών παρουσιάζει πρακτικά μεγάλη ασυνέπεια, ιδιαίτερα όταν εκτελείται λαμβάνοντας υπόψη μόνο την οπτική πληροφορία. Η ταυτόχρονη εκμετάλλευση επιπλέον οπτικής πληροφορίας όπως οι χειρονομίες, αλλά και η ταυτόχρονη χρήση χαρακτηριστικών της φωνής, αποτελούν σήμερα αντικείμενα έρευνας που θα περιορίσουν τις πιθανές εξόδους των συστημάτων αναγνώρισης εκφράσεων και θα αυξήσουν το ποσοστό επιτυχούς αναγνώρισης, βελτιώνοντας παράλληλα και το εύρος των εφαρμογών τους.

□



# Παράρτημα Α

## Αποδόσεις όρων

ανάλυση βασικών συνιστωσών	principal component analysis
απόκρυψη	occlusion
διακωδικοποίηση	transcoding
εναλλασσόμενο διαδοχικό φιλτράρισμα με ανακατασκευή	alternating sequential filtering by re- construction
ιδιοπρόσωπα	eigenfaces
ιδιοδιάνυσμα	eigenvector
μονάδες δράσης	Action Units (AU)
παρακολούθηση	tracking
παράμετροι κίνησης προσώπου	facial animation parameters (FAPs)
ταξινομητής	classifier
χαρακτηριστικά σημεία προσώπου	feature points (FPs)
χάρτης απόκλισης	disparity map
χώρος ενεργοποίησης-αποτίμησης	activation-evaluation space

□



# Βιβλιογραφία

- [1] A. Mehrabian, "Communication without words", Psychology Today, vol. 2, no. 4, pp. 53-56, 1968.
- [2] B. Fasel, et al, "Automatic Facial Expression Analysis: A Survey", Pattern Recognition, 36, 2003, pp. 259-275.
- [3] P. Ekman and W. V. Friesen, The facial Action Coding System: A Technique for the Measurement of Facial Movement. San Francisco: Consulting Psychologists Press, 1978
- [4] Ying-li Tian, Takeo Kanade, J. F. Cohn, Recognizing Action Units for Facial Expression Analysis, IEEE Transactions on PAMI, Vol.23, No.2, February 2001
- [5] A. Murat Tekalp, Jorn Ostermann, Face and 2-D mesh animation in MPEG-4, Signal Processing: Image Communication 15 (2000) 387-421
- [6] A. Raouzaoui and N. Tsapatsoulis and K. Karpouzis and S. Kollias, Parameterized facial expression synthesis based on MPEG-4, Eurasip Journal on Applied Signal Processing, Vol. 2002, No 10, 2002, pp. 1021-1038.
- [7] I.A. Essa and A.P. Pentland, "Coding, Analysis, Interpretation, and Recognition of Facial Expressions," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pp. 757-763, July 1997.
- [8] A. Lanitis, C. Taylor, T. Cootes, and T. Ahmed. Automatic interpretation of human faces and hand gestures using flexible models. In FG, Zurich, 1995, pp 98-103.
- [9] Y. Yacoob and L.S. Davis, "Recognizing Human Facial Expression from Long Image Sequences Using Optical Flow" IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 18, no. 6, June 1996, pp. 636-642.

- [10] C. L. Lisetti and D. E. Rumelhart, Facial expression recognition using a neural network, in: Proc. of the 11<sup>th</sup> International Flairs Conference. AAAI Press, 1998.
- [11] S. Kaiser and T. Wehrle, Automated coding of facial behaviour in human-computer interactions with facs, Journal of Nonverbal Behavior, 16(2), 1992, pp. 67–83.
- [12] G. Edwards, T. Cootes, C. Taylor, Face recognition using active appearance models, Proceedings of the 5<sup>th</sup> European Conference on Computer Vision (ECCV), Vol 2, University of Freiburg, Germany, 1998, pp. 581-695
- [13] J.F. Cohn, A.J. Zlochower, J.J Lien and T. Kanade, Feature-point Tracking by Optical Flow Discriminates Subtle Differences in Facial Expression, Proc. Int'l Conf. Automatic Face and Gesture Recognition, pp. 396-401, 1998
- [14] M.J. Black and Y. Yacoob, "Recognizing Facial Expressions in Image Sequences Using Local Parameterized Models of Image Motion," Int'l J. Computer Vision, vol. 25, no. 1, Oct. 1997, pp. 23-48.
- [15] Kin-Man Lam and Hong Yan, An Analytic-to-Holistic Approach for Face Recognition Based on a Single Frontal View, IEEE PAMI, Vol. 20, No.7, July 1998
- [16] Hua Gu, Guangda Su, Cheng Du, Feature Points Extraction from Face Images, IEEE Transactions on Neural Networks, vol.14, pp 195-200.
- [17] Leung et al, Lip image segmentation using fuzzy clustering incorporating an elliptic shape function, IEEE Trans. on image processing, vol.13, No.1, January 2004
- [18] N. Sarris, N. Grammalidis, M. G. Strintzis, FAP Extraction Using Three-Dimensional Motion Estimation, IEEE Transactions on CSVT, Vol.12, No.10, October 2002
- [19] Y. Tian, T. Kanade, and J. Cohn, Robust Lip Tracking by Combining Shape, Color and Motion, in: Proc. of the 4th Asian Conference on Computer Vision (ACCV'00), January, 2000.

- [20] N. Sebe, M.S. Lew, I. Cohen, Y. Sun, T. Gevers, T.S. Huang, Authentic Facial Expression Analysis, International Conference on Automatic Face and Gesture Recognition (FG'04), Seoul, Korea, May 2004, pp. 517-522.
- [21] D. DeCarlo, D. Metaxas, The integration of optical flow and deformable models with applications to human face shape and motion estimation, CVPR 96, 1996, pp. 231-238
- [22] M. Pantic, L.J.M Rothkrantz, Expert system for automatic analysis of facial expressions, Image and Vision Computing vol 18, 2000, pp. 881-905.
- [23] T. Cootes, G. Edwards, C. Taylor, Active appearance models, IEEE PAMI 23 (6), 2001, pp. 681-685
- [24] C. Huang, Y. Huang, Facial expression recognition using model-based feature extraction and action parameters classification, J. Visual Commun. Image Representation 8 (3) (1997), pp. 278-290.
- [25] M. Lyons, J. Budynek, S. Akamatsu, Automatic classification of single facial images, IEEE PAMI 21(12) 1999
- [26] H. Hong, H. Neven, and C. von der Malsburg, Online Facial Expression Recognition Based on Personalised Galleries, Proc. Int'l Conf. Automatic Face and Gesture Recognition, 1998, pp. 354-359.
- [27] Rein-Lien Hsu, Mohamed Abdel-Mottaleb, Anil K. Jain, Face Detection in Color Images, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.24, No.5, May 2002
- [28] S. McKenna, Y. Raja, and S. Gong, Tracking Color Objects using Adaptive Mixture Models, Image and Vision Computing, 1998, pp. 225-231
- [29] ERMIS, Emotionally Rich Man-machine Intelligent System IST-2000-29319 (<http://www.image.ntua.gr/ermis>)
- [30] HUMAINE IST, Human-Machine Interaction Network on Emotion, 2004-2007, <http://www.emotion-research.net>
- [31] ISTFACE - MPEG-4 Facial Animation System – Version 3.3.1 Gabriel Abrantes, (Developed in the context of the European Project ACTS MoMuSys © 97-98 Instituto Superior Tecnico).

- [32] J. W. Young, Head and Face Anthropometry of Adult U.S. Civilians, FAA Civil Aeromedical Institute, 1963-1993 (final report 1993)
- [33] M. H. Yang, D. Kriegman, N. Ahuja, "Detecting Faces in Images:A Survey", PAMI, Vol.24(1), 2002, pp. 34-58.
- [34] C. Papageorgiou, M. Oren and T. Poggio, A general framework for object detection. In international Conference on Computer Vision, 1998.
- [35] P. Viola, M. Jones, Rapid Object Detection using a Boosted Cascade of Simple Features. Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on , Volume: 1 , 8-14 Dec. 2001 Pages:I-511 - I-518 vol.1
- [36] I. Fasel, B. Fortenberry, and J. R. Movellan, A generative framework for real-time object detection and classification, Computer Vision and Image Understanding, Volume 98 , Issue 1 (April 2005), pp. 182 – 210.
- [37] R. Fransens, Jan De Prins, SVM-based Nonparametric Discriminant Analysis, An Application to Face Detection, Ninth IEEE International Conference on Computer Vision Volume 2, October 13 - 16, 2003
- [38] S. Kollias and D. Anastassiou, An adaptive least squares algorithm for the efficient training of artificial neural networks, IEEE Transactions on Circuits and Systems, Volume: 36 , Issue: 8 , Aug. 1989 pp.1092-1101
- [39] M. T. Hagan, and M. Menhaj, Training feedforward networks with the Marquardt algorithm, IEEE Transactions on Neural Networks, vol. 5, no. 6, 1994, pp. 989-993.
- [40] Lijun Yin, Generating Realistic Facial Expressions with Wrinkles for Model-Based Coding, Computer Vision and Image Understanding 84, (2001), pp. 201-240.
- [41] M. J. Lyons, M. Haehnel & N. Tetsutani, The Mouthesizer: A Facial Gesture Musical Interface, Conference Abstracts, Siggraph 2001, Los Angeles, p. 230.
- [42] S.Arca, P.Campadelli, R. Lanzarotti, An automatic feature-based face recognition system, in: Proc. of the 5th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS 2004)



- [43] Kin-Man Lam and Hong Yan, Locating And Extracting the Eye in Human Face Images, *Pattern Recognition* Vol.29, No.5, 1996, pp. 771-779.
- [44] C. Tomasi and T. Kanade, Detection and Tracking of Point Features, Carnegie Mellon University Technical Report CMU-CS-91-132, April 1991.
- [45] J. Canny, A Computational Approach to Edge Detection, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol 8, No. 6, Nov 1986.
- [46] D. O. Gorodnichy. On Importance of Nose for Face Tracking, in: *Proc. Intern. Conf. on Automatic Face and Gesture Recognition (FG'2002)*, Washington DC, May 20-21, 2002.
- [47] Laser scan and facial anthropometry, *British Journal of Plastic Surgery*, (48), 1995, pp. 551-558.
- [48] L. Vincent, Morphological Grayscale Reconstruction in Image Analysis: Applications and Efficient Algorithms, *IEEE Trans. Image Processing*, vol. 2, no. 2, 1993, pp. 176-201.
- [49] L. Vincent, Morphological Grayscale Reconstruction in Image Analysis: Applications and Efficient Algorithms, *IEEE Transactions on Image Processing*, Vol. 2, No. 2, April, 1993, pp. 176-201.
- [50] A. Krogh, J. Vedelsby, Neural network ensembles, cross validation, and active learning, In Tesauro, G., Touretzky, D., Leen, T., *Advances in Neural Information Processing Systems*, 7, Cambridge, MA. 1995 MIT Press, pp. 231-238.
- [51] V. Tresp, Committee machines, In Yu Hen Hu and Jenq-Nen Hwang, editors, *Handbook for Neural Network Signal Processing*. CRC Press, 2001
- [52] C. M. Whissel, The dictionary of affect in language, R. Plutchnik and H. Kellerman (Eds) *Emotion: Theory, research and experience: vol 4, The measurement of emotions*. Academic Press, New York, 1989.
- [53] S. Ioannou, A. Raouzaïou, V. Tzouvaras, T. Mailis, K. Karpouzis, S. Kollias, Emotion recognition through facial expression analysis based on a neurofuzzy network, *Neural Networks*, Elsevier, Volume 18, Issue 4, May 2005, pp. 423-435.

- [54] G. Klir, B. Yuan, Fuzzy Sets and Fuzzy Logic, Theory and Applications, New Jersey, Prentice Hall, 1995.
- [55] M.A. Lee, H. Takagi, Integrating design stages of fuzzy systems using genetic algorithms, proceedings of IEEE International conference on fuzzy systems, 1993.
- [56] Wallace M. and Kollias S. "Possibilistic Evaluation of Extended Fuzzy Rules in the Presence of Uncertainty" FUZZ-IEEE 2005, May 22-25, Reno, Nevada, USA
- [57] J. Weizenbaum, ELIZA – A Computer Program For the Study of Natural Language Communication Between Man and Machine, Communications of the ACM, Volume 9, Number 1, 1966, pp. 36-35.
- [58] Y. J. Zhang, A Survey on Evaluation Methods for Image Segmentation, Pattern Recognition, Vol 29, No. 8, 1996, pp.1334-1346.
- [59] G. W. Williams, Comparing the joint agreement of several raters with another rater, Biometrics, vol 32, pp. 619-627, 1976
- [60] V. Chalana and Y. Kim, A Methodology for Evaluation of Boundary Detection Algorithms on Medical Images, IEEE Transactions on Medical Imaging, Vol.16, No.5 October 1997
- [61] R. Cowie, E.D. Cowie, S. Savvidou, E. McMahon, M. Sawey, M. Schröder, 'FEELTRACE': An instrument for recording perceived emotion in real time, in proc: ISCA Workshop on Speech and Emotion-2000, Northern Ireland, pp. 19-24.
- [62] D. Cristinacce, T. F. Cootes, A comparison of shape constrained facial feature, in: proc. FGR 2004, Seoul, pp. 375-380.
- [63] G. Votsis, A. Drosopoulos and S. Kollias, A Modular Approach to Facial Feature Segmentation on Real Sequences, Signal Processing: Image Communication, Vol. 18 (1), (2003), pp. 67-89
- [64] R. Plutchik, Emotion: A psychoevolutionary synthesis, Harper and Row, New York, 1980.
- [65] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz and J. Taylor, Emotion Recognition in Human-Computer Interaction, IEEE Signal Processing Magazine, no. 1, January 2001, pp.32-80

- [66] T. Meier and K. N. Ngan Automatic Segmentation of Moving Objects for Video Object Plane Generation IEEE Trans. Circuits and Systems for Video Technology, Vol. 8, No. 5 pp. 525-538 1998
- [67] M. Pardas and P.Salembier 3-D morphological segmentation and motion estimation for image segmentation Signal Processing, vol. 38 pp. 31-43 Sept. 1994
- [68] M. M. Chang, A. M. Tekalp, and M. I. Sezan Motion field segmentation using an adaptive MAP criterion Proc. of IEEE Int. Conf. Acoust., Speech, Signal Processing ICASSP'93 Minneapolis, MN, Apr. 1993, vol. V pp.33-36
- [69] D. Geiger and A. Yuille A common framework for image segmentation Int. Journal Comput. Vision, vol 6 pp.227-243 1991
- [70] O. J. Morris, M. J. Lee and A. G. Constantinides Graph Theory for Image Analysis: an Approach based on the Shortest Spanning Tree IEE Proceedings, Vol. 133 pp.146-152 April 1986
- [71] Y. Avrithis, A. Doulamis, N. Doulamis and S. Kollias A Stochastic Framework for Optimal Key Frame Extraction from MPEG Video Databases Computer Vision and Image Understanding July 1999
- [72] R. C. Gonzalez and R. E. Woods, Digital Image Processing, Addison-Wesley, 1992
- [73] D. Tzovaras, N. Grammalidis and M. G. Strintzis Disparity Field and Depth Map Coding for Multiview 3D Image Generation Image Communication, No. 11 pp. 205-230 1998
- [74] C. Gu and M.-C. Lee Semiautomatic Segmentation and Tracking of Semantic Video Objects IEEE Trans. on Circuits and Systems for Video Technology, Vol. 8, No. 5 pp. 572-584 Sept. 1998
- [75] A.A. Alatan, L. Onural, M. Wollborn, R. Mech, E. Tuncel, and T. Sikora Image Sequence Analysis for Emereging Interactive Multimedia Services - The European COST 211 Framework IEEE Trans. on Circuits and Systems for Video Technology, Vol. 8, No. 7 pp. 802-813 Nov. 1998
- [76] Chong-Wah Ngo, Ting-Chuen Pong, and R. T. Chin Video Partitioning by Temporal Slice Coherency IEEE Trans. on Circuits and Systems for Video Technology, Vol. 11, No. 8 pp. 941-953 Aug. 2001

- [77] A. Smolic, T. Sikora, and J.R. Ohm Long-Term Global Motion Estimation and Its Application for Sprite Coding, Content Description, and Segmentation IEEE Trans. on Circuits and Systems for Video Technology, Vol. 9, No. 8 pp. 1227-1242 Dec. 1999
- [78] G.B. Stamou and S.G. Tzafestas Fuzzy Relation Equations and Fuzzy Inference Systems: An Inside Approach IEEE Transactions on Systems, Man and Cybernetics - PART B: Cybernetics, Vol. 29, No. 6 Dec. 1999
- [79] G.J. Klir and B. Yuan Fuzzy Sets and Fuzzy Logic: Theory and Applications Englewood Cliffs, NY: Prentice Hall, 1995
- [80] L.A. Zadeh Fuzzy logic=computing with words IEEE Transactions on Fuzzy Systems pp. 103-111 1996
- [81] E. Saber and A. Tekalp and G. Bozdagi Fusion of Color and Edge Information for Improved Segmentation and Edge Linking Image and Vision Computing vol 15 pp. 769-780 1997
- [82] P.Salembier, F.Marques, M. Pardas, J.R.Morros, I. Corset, S.Jeannin, L.Bouchard, F.Meyer, and B.Marcotegui Segmentation-Based Video Coding System Allowing the Manipulation of Objects IEEE Trans. on Circuits and Systems for Video Technology, Vol. 7, No. 1 pp. 60-74 Feb. 1997
- [83] I. Koprinska, S. Carrato Temporal video segmentation: A survey Signal Pricessing: Image Communication 16 pp. 477-500 2001
- [84] N. Doulamis, A. Doulamis, Y. Avrithis, K. Ntalianis, and S. Kollias Efficient Summarization of Stereoscopic Video Sequences IEEE Trans. Circuits and Systems for Video Technology, Vol. 10, No. 4 pp. 501- 517 June 2000
- [85] T. Sikora The MPEG-4 Video Standard Verification Model IEEE Trans. Circuits and Systems for Video Technology, Vol. 7, No. 1 pp. 19-31 Feb. 1997
- [86] A.M. Tekalp Digital Video Processing Prentice Hall 1995
- [87] M. Turk and A. Pentland, Eigenfaces for recognition, Journal of Cognitive Neuroscience, 3(1):71-86, Mar. 1991.

- [88] T. Otsuka, J. Ohya, Extracting facial motion parameters by tracking feature points, Proceedings of First International Conference on Advanced Multimedia Content Processing, Osaka, Japan, 1998, pp. 442-453.
- [89] Jain, A.K. Fundamentals of Digital Image Processing. Englewood Cliffs, NJ: Prentice-Hall, 1989.
- [90] A. Imai, N. Shimada, Y. Shirai, 3-D hand posture recognition by training contour variation, Sixth IEEE International Conference on Automatic Face and Gesture Recognition, 17-19 May 2004 pp. 895-900, 2004
- [91] R. Kjeldsen, J. Kender, Finding skin in color images, Proc. 2nd International Conference on Automatic Face and Gesture Recognition, 1996, pp. 312-317
- [92] W. Freeman, C.D. Weissman, Television Control by Hand Gestures, Proc. International Workshop on Automatic Face and Gesture Recognition, Zurich Switzerland, 1995, pp. 179-183.
- [93] T. Darrell, I. Essa, A. Pentland, Task-Specific Gesture Analysis in Real-Time Using Interpolated Views, IEEE Trans. Pattern Analysis and Machine Intelligence, 18(12), 1996, pp. 1236-1242
- [94] O.J Morris, M. J Lee, A. G. Constantinides, "Graph Theory for Image Analysis: An Approach Based on the Shortest Spanning Tree", IEE Proceedings, Vol. 133, pp. 146-152, April 1986.
- [95] P.J Mulroy, "Video Content Extraction: Review of Current Automatic Segmentation Algorithms", in Proc. of Workshop on Image Analysis and Multimedia Interactive Systems (WIAMIS), Louvain-la-Neuve, Belgium, June 1997.
- [96] P. Salembier, M. Pardas, "Hierarchical morphological segmentation for image sequence coding", IEEE Trans. Image Processing, Vol. 3, No. 5, pp. 639-651, Sept. 1994.
- [97] Z. Xue, S.Z. Li, and E.K. Teoh, Ai-eigensnake: An affine-invariant deformable contour model for object matching, Image and Vision Computing, vol. 20, no. 2, pp. 784, 2002
- [98] A.L. Yuille, D.S. Cohen, P.W. Hallinan, Feature Extraction from Faces Using Deformable Templates, International Journal Computer Vision, vol. 8, no. 2, pp. 99-111, Aug. 1992.

- [99] N. Funabiki, M. Isogai, T. Higashino, M. Oda, An Eye-Contour Extraction Algorithm from Face Image using Deformable Template Matching, *Memoirs of the Faculty of Engineering, Okayama University*, Vol.40, pp.78-82, January, 2006
- [100] P. Kuo, J. Hannah, An improved eye feature extraction algorithm based on deformable templates, *IEEE International Conference on Image Processing (ICIP)*, pp. II-1206-9, 2005
- [101] T. Takegami, T. Gotoh, S. Kagei, R. Minamikawa-Tachino, A Hough Based Eye Direction Detection Algorithm without On-site Calibration, *Proc. VIIth Digital Image Computing: Techniques and Applications*, 10-12 Dec. 2003, Sydney A new method for eye extraction from facial image
- [102] Xiaoyun Deng, Chip-Hong Chang, E. Brandle, Second IEEE International Workshop on Electronic Design, Test and Applications, 28-30 Jan. 2004 pp. 29 - 34
- [103] S. Sclaroff, L. Liu ,Deformable Shape Detection and Description via Model-Based Region Grouping, *Pattern Analysis and Machine Intelligence*, Vol. 23, No. 5, pp. 475-489, May 2001

# Κατάλογος δημοσιεύσεων του συγγραφέα

## ΔΙΕΘΝΗ ΠΕΡΙΟΔΙΚΑ

- [J1] S. Ioannou, K. Karpouzis and S. Kollias, Robust Feature Detection for Facial Expression Recognition, International Journal of Image and Video Processing, Hindawi Publishing, Submitted
- [J2] M. Wallace, S. Ioannou, A. Raouzaïou, K. Karpouzis, S. Kollias, “Dealing with Feature Uncertainty in Facial Expression Recognition Using Possibilistic Fuzzy Rule Evaluation”, International Journal of Intelligent Systems Technologies and Applications, accepted for publication
- [J3] S. Ioannou, A. Raouzaïou, V. Tzouvaras, T. Mailis, K. Karpouzis, S. Kollias, “Emotion recognition through facial expression analysis based on a neurofuzzy network”, Special Issue on Emotion: Understanding & Recognition, Neural Networks, Elsevier, Volume 18, Issue 4, Pages 423-435, May 2005
- [J4] K. Karpouzis, G. Moschovitis, K. Ntalianis, S. Ioannou, and S. Kollias, “Web Access To Large Audiovisual Assets Based On User Preferences”, Multimedia Tools and Applications, vol. 22, no. 3, pp. 215-234, Kluwer Academic Publishers, March 2004

## ΚΕΦΑΛΑΙΑ ΣΕ ΒΙΒΛΙΑ

- [B5] K. Karpouzis, A. Drosopoulos, S. Ioannou, A. Raouzaïou, N. Tsapatsoulis, S. Kollias, “Facial and Body Feature Extraction for Emotionally-Rich HCI”, Mehdi Khosrow-Pour (ed.), Encyclopedia of Information Science and Technology, Idea Group Publishers, pp. 1180 - 1185, 2005

- [B6] K. Karpouzis, A. Raouzaïou, A. Drosopoulos, S. Ioannou, T. Balomenos, N. Tsapatsoulis and S. Kollias, “Facial expression and gesture analysis for emotionally-rich man-machine interaction”, N. Sarris, M. Strintzis, (eds.), 3D Modeling and Animation: Synthesis and Analysis Techniques, pp. 175-200, Idea Group Publ, 2004.
- [B7] S. Ioannou, Y. Avrithis, G. Stamou and S. Kollias, “Fuzzy Data Fusion For Multiple Cue Image And Video Segmentation”, Fuzzy Technologies and Applications, E. Kerre (Editor), Springer-Verlag, 2002.

## **ΔΙΕΘΝΗ ΣΥΝΕΔΡΙΑ**

- [C8] S. Ioannou, L. Kessous, K. Karpouzis, “Multimodal emotion recognition on naturalistic data”, International Conference on Artificial Neural Networks (ICANN), Athens, Greece, 10-14 September 2006
- [C9] S. Ioannou, M. Wallace and S. Kollias, “Intelligent Facial Analysis and Expression Recognition”, International Joint Conference on Neural Networks (IJCNN), Vancouver, Canada, July 2006.
- [C10] S. Ioannou, M. Wallace, K. Karpouzis and S. Kollias, “A Robust Scheme for Facial Analysis and Expression Recognition”, Measuring Behavior 2005, 5th International Conference on Methods and Techniques in Behavioral Research, Wageningen, The Netherlands, 30 August - 2 September 2005
- [C11] S. Ioannou, M. Wallace, K. Karpouzis, A. Raouzaïou and S. Kollias, “Combination of Multiple Extraction Algorithms in the Detection of Facial Features”, Proceedings of the IEEE International Conference on Image Processing (ICIP), Genova, Italy, September 2005
- [C12] S. Ioannou, M. Wallace, K. Karpouzis, A. Raouzaïou and S. Kollias, “Confidence-Based Fusion of Multiple Feature Cues for Facial Expression Recognition”, FUZZ-IEEE 2005, May 22-25, Reno, Nevada, USA
- [C13] R. Cowie, E. Douglas-Cowie, J. Taylor, S. Ioannou, M. Wallace, S. Kollias, “An Intelligent System for Facial Emotion Recognition”, IEEE International Conference on Multimedia & Expo, July 6-8, 2005, Amsterdam, The Netherlands



- [C14] A. Raouzaïou, S. Ioannou, G. Akrivas, K. Karpouzis and S. Kollias, “Adaptation of Expression Analysis Based on Evaluation Principles”, Fourth European Symposium on Intelligent Technologies and their implementation on Smart Adaptive Systems (Eunite 2004), Aachen, Germany, June 2004.
- [C15] S. Ioannou, A. Raouzaïou, K. Karpouzis, M. Pertselakis, N. Tsapatsoulis, S. Kollias, “Adaptive Rule-Based Facial Expression Recognition”, G. Vouros, T. Panayiotopoulos (Eds.), Lecture Notes in Artificial Intelligence, Vol. 3025, Springer-Verlag, pp. 466 - 475, 2004.
- [C16] S. Ioannou, A. Raouzaïou, K. Karpouzis and S. Kollias, “Adaptation of Facial Feature Extraction and Rule Generation in Emotion-Analysis Systems”, International Joint Conference on Neural Networks (IJCNN 2004), Budapest, Hungary, July 2004.
- [C17] T. Balomenos, A. Raouzaïou, S. Ioannou, A. Drosopoulos, K. Karpouzis, and S. Kollias, “Emotion Analysis In Man-Machine Interaction Systems”, Workshop on Multimodal Interaction and Related Machine Learning Algorithms (MLMI04), Geneva
- [C18] T. Balomenos, A. Raouzaïou, S. Ioannou, A. Drosopoulos, K. Karpouzis, S. Kollias, “Emotion Analysis in Man-Machine Interaction Systems”, Samy Bengio, Herve’ Bourlard (Eds.), Machine Learning for Multimodal Interaction, Lecture Notes in Computer Science, Vol. 3361, 2004, pp. 318 - 328, Springer-Verlag .
- [C19] A. Raouzaïou, S. Ioannou, K. Karpouzis, N. Tsapatsoulis, S. Kollias, R. Cowie, “An Intelligent Scheme for Facial Expression Recognition”, O. Kaynak (ed.), Artificial Neural Networks and Neural Information Processing, Lecture notes in Computer Science 2714, Springer, pp. 1109 ÷ 1116, 2003.
- [C20] A. Raouzaïou, S. Ioannou, K. Karpouzis, S. Kollias, R. Cowie , “Facial Expression Analysis”, Eunite 2003, Third European Symposium on Intelligent Technologies, Oulu, Finland
- [C21] A. Drosopoulos, T. Mpalomenos, S. Ioannou, K. Karpouzis and S. Kollias, “Emotionally-rich Man-machine Interaction Based on Gesture Analysis”, Human-Computer Interaction International 2003, 22 - 27 June, Crete, Greece, vol. 4, pp. 1372 - 1376

- [C22] A. Raouzaïou, S. Ioannou, K. Karpouzis, N. Tsapatsoulis, S. Kollias, R. Cowie , “An Intelligent Scheme for Facial Expression Recognition”, Proceedings of the International Conference on Artificial Neural Networks (ICANN), Istanbul, Turkey, June 2003
- [C23] S. Ioannou, G. Akrivas, “Archiving multimedia content descriptions: an early adaptation of MPEG-7”, Proceedings of Int. Workshop on Very Low Bitrate Video Coding (VLBV01)
- [C24] I. Varlamis, M. Vazirgiannis, P. Poulos, G. Akrivas, S. Ioannou, “X-Database. A middleware for collaborative video annotation, storage and retrieval”, proceedings of the 8th Panhellenic Conference. Cyprus 2001.
- [C25] G. Akrivas, S. Ioannou, E. Karakoulakis, K. Karpouzis, Y. Avrithis, A. Delopoulos, S. Kollias, I. Varlamis and M. Vaziriannis, “An Intelligent System for Retrieval and Mining of Audiovisual Material Based on the MPEG-7 Description Schemes”, Eunite 2001, European Symposium on Intelligent Technologies, Tenerife, Spain
- [C26] K. Ntalianis, S. Ioannou, K. Karpouzis, G. Moschovitis and S. Kollias, “Visual Information Retrieval from Annotated Large Audiovisual Assets Based on User Profiling and Collaborative Recommendations”, Proceedings of the IEEE International Conference on Multimedia and Exposition (ICME01), Tokyo, Japan
- [C27] S. Ioannou, G. Moschovitis, K. Ntalianis, K. Karpouzis and S. Kollias, “Effective Access to Large Audiovisual Assets Based on User Preferences”, Proceedings of the ACM Multimedia 2000, Los Angeles, CA, USA, November 2000

# Βιογραφικό Σημείωμα

Ο Σπύρος Ιωάννου του Βασιλείου γεννήθηκε στις 6 Σεπτεμβρίου 1975 στην Αθήνα. Αποφοίτησε από την Ελληνογαλλική Σχολή Αγ. Παρασκευής τον Ιούλιο του 1993. Την ίδια χρονιά έγινε δεκτός στη Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του ΕΜΠ. Ακολούθησε τις πλήρεις ροές Υπολογιστών, Λογισμικού και Δικτύων. Το Δεκέμβριο του 2000 έγινε δεκτός από τη Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του ΕΜΠ για εκπόνηση διδακτορικής διατριβής στον τομέα της Ψηφιακής Επεξεργασίας Εικόνας.

## Διδακτική Εμπειρία

- Σεμινάρια στο UNIX, εισαγωγικά και για διαχειριστές συστημάτων, Υπολογιστικό Κέντρο ΕΜΠ ακαδημαϊκά έτη 1995-96
- Με την ιδιότητα του Υποψήφιου διδάκτορα, διοργάνωση και επίβλεψη εργαστηριακών ασκήσεων για το μάθημα Προγραμματιστικές Τεχνικές (C, PASCAL), Τμήμα Ηλεκτρολόγων Μηχ/κών & Μηχανικών Υπολογιστών, Ε.Μ.Π. (2000 - )
- Με την ιδιότητα του Υποψήφιου διδάκτορα, υποστήριξη του εργαστηριακού μαθήματος Τεχνολογία Εικόνας & Πολυμέσων, Τμήμα ΗΜΜΗΥ Ε.Μ.Π. (2000 - )
- Σεμινάρια στο Κ.Ε.Κ Εθνικό Κέντρο Πολιτιστικών και Επαγγελματικών Εφαρμογών για λογαριασμό της OTENET Α.Ε. με τίτλο Τεχνικοί IP δικτύων - Διαχειριστές Πληροφοριακών Συστημάτων, 2001

## Ξένες Γλώσσες

Αγγλικά: Cambridge First Certificate in English

Γαλλικά: Diplôme d'Études en Langue Française (DEL F)

Diplôme Approfondi de Langue Française (DAL F)

### **Επαγγελματικά Σωματεία**

- Μέλος του Τεχνικού Επιμελητηρίου Ελλάδος
- Μέλος του Institute of Electrical and Electronic Engineers (IEEE)

### **Συμμετοχή σε Ερευνητικά - Αναπτυξιακά Προγράμματα**

- Έργο ΕΠΕΤ II - ΠΑΝΟΡΑΜΑ (Σύστημα Έξυπνης Αναζήτησης Και Εξόρυξης Από Βάσεις Με Οπτικοακουστικά Δεδομένα). Συμμετοχή στη σχεδίαση και στην υλοποίηση των υποσυστημάτων εισαγωγής, αναζήτησης και ανάκτησης δεδομένων. (1999-2001)
- Έργο ΓΓΤΠ (Σύστημα Ψηφιακής Αρχαιοθήκης του Οπτικοακουστικού Υλικού του Υπουργείου Τύπου & ΜΜΕ). Συμμετοχή στη σχεδίαση και στην υλοποίηση των υποσυστημάτων εισαγωγής, αναζήτησης και ανάκτησης δεδομένων. (1999-2000)
- IST-FAETHON, (Unified Intelligent Access to Heterogeneous Audiovisual Content)
- IST ERMIS (Emotionally Rich Man-machine Intelligent System) (2002-2004) Συμμετοχή στους τομείς που σχετίζονται με την ανάλυση συναισθήματος από εικόνες.
- IST HUMAINE (Human-Machine Interaction Network on Emotion) Συμμετοχή στους τομείς που σχετίζονται με την ανάλυση συναισθήματος από εικόνες.

□