



Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
Τομέας Σημάτων, Ελέγχου και Ρομποτικής
Εργαστήριο Όρασης Υπολογιστών, Επικοινωνίας Λόγου και Επεξεργασίας Σημάτων

Εκτίμηση της κατεύθυνσης του βλέμματος με χρήση Ενεργών Μοντέλων Εμφάνισης

Διπλωματική Εργασία
του
Πέτρου Ν. Κούτρα

Επιβλέπων: Πέτρος Μαραγκός
Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούλιος 2012



Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
Τομέας Σημάτων, Ελέγχου και Ρομποτικής
Εργαστήριο Όρασης Υπολογιστών, Επικοινωνίας Λόγου και
Επεξεργασίας Σημάτων

Εκτίμηση της κατεύθυνσης του βλέμματος με χρήση Ενεργών Μοντέλων Εμφάνισης

Διπλωματική Εργασία

του

Πέτρου Ν. Κούτρα

Επιβλέπων: Πέτρος Μαραγκός
Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 19η Ιουλίου 2012.

(Υπογραφή)

(Υπογραφή)

(Υπογραφή)

.....
Πέτρος Μαραγκός
Καθηγητής
Ε.Μ.Π.

.....
Κωνσταντίνος Τζαφέστας
Επίκουρος Καθηγητής
Ε.Μ.Π.

.....
Γεράσιμος Ποταμιάνος
Αναπληρωτής Καθηγητής
Παν/μίου Θεσσαλίας

Αθήνα, Ιούλιος 2012

(Υπογραφή)

.....
Πέτρος Ν. Κούτρας

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

Copyright © Πέτρος Ν. Κούτρας, 2012.

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

Ευχαριστίες

Θα ήθελα καταρχήν να ευχαριστήσω τον καθηγητή κ. Πέτρο Μαραγκό για την επίβλεψη αυτής της διπλωματικής εργασίας και για την ευκαιρία που μου έδωσε να την εκπονήσω στο εργαστήριο Όρασης Υπολογιστών, Επικοινωνίας Λόγου και Επεξεργασίας Σημάτων. Ευχαριστώ θερμά όλα τα μέλη του εργαστηρίου και ιδιαίτερα τον υποψήφιο διδάκτορα Ισίδωρο Ροδομαγουλάκη για την πολύτιμη βοήθεια τους. Επίσης, θα ήθελα να ευχαριστήσω το Γεώργιο Παπανδρέου για την παραχώρηση του πηγαίου κώδικα του MATLAB σχετικά με την υλοποίηση και προσαρμογή των AAMs που είχε υλοποιήσει στη διάρκεια του διδακτορικού του. Τέλος, θα ήθελα να ευχαριστήσω την οικογένεια μου και τους φίλους μου για την καθοδήγηση και την ηθική συμπαράσταση που μου προσέφεραν όλα αυτά τα χρόνια.

Περίληψη

Η συγκεκριμένη διπλωματική έχει σαν αντικείμενο την επίλυση προβλημάτων που προκύπτουν από την αναγνώριση και μοντελοποίηση του αναθρώπινου προσώπου, εστιάζοντας στο μεγαλύτερο μέρος της στην εκτίμηση της κατεύθυνσης του βλέματος (eye gaze). Επιπλέον, αναλύονται και γίνεται προσπάθεια να επιλυθούν και άλλα σημαντικά προβλήματα που σχετίζονται με τη μοντελοποίηση του προσώπου όπως η εκτίμηση της πόζας κεφαλιού και η αναγνώριση συναισθήματος με βάση χαρακτηριστικά του προσώπου του. Αρχικά, γίνεται μια συστηματική και συνοπτική επισκόπηση των μεθόδων της βιβλιογραφίας σχετικά με τα παραπάνω προβλήματα καθώς και περιγραφή των μεθόδων υλοποίησης και προσαρμογής των Ενεργών Μοντέλων Εμφάνισης (AAMs) σε νέες εικόνες. Στη συνέχεια, ακολουθεί η ανάλυση και υλοποίηση του συστήματος υπολογισμού του eye gaze ακολουθώντας δύο διαφορετικές προσεγγίσεις. Στην πρώτη από αυτές αναλύεται μια μέθοδος υπολογισμού του eye gaze που στηρίζεται στο γεωμετρικό μοντέλο του ματιού, η οποία παρέχει εκτίμηση υπό μορφή γωνιών. Για την ακριβέστερη εκτίμηση του eye gaze αναπτύσσεται, σαν ένα ενδιάμεσο στάδιο, μια τεχνική υπολογισμού της πόζας κεφαλιού. Σύμφωνα με τη δεύτερη προσέγγιση γίνεται κατηγοριοποίηση του gaze με χρήση μοντέλων μείγματος γκαουσιανών συναρτήσεων (GMMs) σε διακριτές κλάσεις, τόσο με βάση τις παραμέτρους του τοπικού AAM όσο και με χαρακτηριστικά Ιστογραμμάτων Προσανατολισμένης Κλίσης (HOGs), και προτείνεται τρόπος για την αντιστοίχιση των κλάσεων με γωνίες. Τέλος, γίνεται ανάπτυξη ενός συστήματος κατηγοριοποίησης του συναισθήματος σε διακριτές κλάσεις, βασιζόμενο στις παραμέτρους των AAMs και χρησιμοποιώντας σαν ταξινομητές GMMs αλλά και Κρυφά Μαρκοβιανά Μοντέλα (HMMs).

Λέξεις Κλειδιά

Εκτίμηση κατεύθυνσης βλέματος (eye gaze), ενεργά μοντέλα εμφάνισης, πόζα κεφαλιού, γεωμετρικό μοντέλο, ιστογράμματα προσανατολισμένης κλίσης, κατηγοριοποίηση συναισθήματος, μοντέλα μείγματος γκαουσιανών συναρτήσεων, κρυφά μαρκοβιανά μοντέλα.

Abstract

The aim of this thesis is to find solutions to problems that occur from the recognition and modeling of the human face, focusing primarily on the estimation of eye gaze. Moreover, a greater range of problems concerning from the face modeling, such as head pose estimation and emotion classification by facial features, are being analyzed and an effort is being made to resolve them. Initially, a systematic and compact review of the bibliography methods about these problems is made, as well as a description of the solution and fitting technique of Active Appearance Models (AAMs) into prototype pictures. Furthermore, the analysis and implementation of a computing system for eye gaze by two different points of view is being presented. In the first one, one eye gaze computing method based on a geometric model of eye is analyzed; this method provides angle form estimation. For a more accurate estimation, it is developed in an intermediate step one computing technique for the head pose. In the second one, gaze is being classified into discrete classes through Gaussian Mixture Models (GMM), using both the AAMs parameters, and the HOGs (Histograms of Oriented Gradients) features; and a matching way between classes and gaze angles. Finally, a system for discrete emotion classification, based on AAMs parameters using GMMs and Hidden Markov Models (HMMs) classifiers, is developed.

Keywords

Eye gaze estimation, active appearance models, head pose, geometric model, histograms of Oriented gradients, emotion classification, gaussian mixture models, hidden markov models.

Περιεχόμενα

Ευχαριστίες	7
Περίληψη	9
Abstract	11
Περιεχόμενα	13
Κατάλογος Σχημάτων	15
1 Εισαγωγή	21
1.1 Όραση Υπολογιστών	21
1.2 Το Eye Gaze στην όραση υπολογιστών	22
1.2.1 Το γενικότερο πρόβλημα της αναγνώρισης ενός αντικειμένου	22
1.2.2 Η αναγνώριση και μοντελοποίηση του ανθρώπινου προσώπου	22
1.2.3 Ο ρόλος του Eye Gaze	23
1.3 Περιεχόμενο και δομή της παρούσας εργασίας	24
1.3.1 Σκοπός της διπλωματικής	24
1.3.2 Συνεισφορά της διπλωματικής	25
1.3.3 Οργάνωση του περιεχομένου της διπλωματικής	25
2 Υπάρχουσα Έρευνα (Background)	29
2.1 Εισαγωγή	29
2.2 Τεχνικές εκτίμησης του eye gaze	30
2.2.1 Γενικά	30
2.2.2 Γεωμετρικά μοντέλα με χρήση υπέρυθρου φωτός	32
2.2.3 Γεωμετρικά μοντέλα με χρήση συμβατικών καμερών	34
2.2.4 Μέθοδοι βασισμένες σε ταίριασμα προτύπων και ανάλυση παλινδρόμησης	36
2.3 Τεχνικές αναγνώρισης/κατηγοριοποίησης συναισθήματος	39
2.3.1 Γενικά	39

2.3.2	Τεχνικές που βασίζονται στη χρήση AAMs	40
3	Ενεργά Μοντέλα Εμφάνισης	43
3.1	Εισαγωγή	43
3.2	Περιγραφή των AAMs	44
3.2.1	Μοντέλο Σχήματος	44
3.2.2	Μοντέλο Υφής	46
3.2.3	Δημιουργία πρωτότυπων εικόνων με χρήση AAMs	47
3.2.4	Εφαρμογή των AAMs σε νέες εικόνες	48
3.3	AAMs προσαρμοσμένα στο χρήστη	51
3.3.1	Κατασκευή των AAMs	51
3.3.2	Αποτελέσματα εφαρμογής των AAMs σε νέες εικόνες	55
3.4	Τοπικό AAM για τα μάτια	56
3.5	Γενικευμένο AAM για κάθε χρήστη	58
4	Εκτίμηση του Eye Gaze από χαρακτηριστικά σημεία του προσώπου και των ματιών	61
4.1	Εισαγωγή	61
4.2	Εκτίμηση του gaze από τη θέση της ίριδας	62
4.2.1	Γενική περιγραφή της μεθόδου	62
4.2.2	Μορφολογικό φιλτράρισμα	63
4.2.3	Μετασχηματισμός Hough και σχεδίαση ιστογραμμάτων	65
4.3	Εκτίμηση της πόζας κεφαλιού	69
4.3.1	Περιγραφή της μεθόδου	69
4.3.2	Υπολογισμός των γωνιών slant και tilt	72
4.3.3	Πειραματικά αποτελέσματα	75
4.4	Γεωμετρικό μοντέλο εκτίμησης του Eye Gaze	77
4.4.1	Περιγραφή του μοντέλου	77
4.4.2	Προσδιορισμός των ανατομικών σταθερών	80
4.4.3	Πειραματικά αποτελέσματα	82
4.5	Συμπεράσματα	87
5	Εκτίμηση του Eye Gaze με κατηγοριοποίησή του σε κλάσεις	89
5.1	Εισαγωγή	89
5.2	Κατηγοριοποίηση με βάση το τοπικό AAM	90
5.2.1	Περιγραφή των GMMs και του αλγορίθμου EM	90
5.2.2	Εφαρμογή της μεθόδου	92
5.3	Κατηγοριοποίηση με χρήση των HOGs	95
5.3.1	Περιγραφή των Histograms of Oriented Gradients	95
5.3.2	Εφαρμογή των HOGs σε εικόνες των ματιών	97
5.4	Υπολογισμός των γωνιών του gaze από την πιθανοφάνεια κάθε κλάσης	99

5.4.1	Περιγραφή του αλγορίθμου	99
5.4.2	Πειραματικά αποτελέσματα	101
5.5	Συμπεράσματα	103
6	Κατηγοριοποίηση Συναισθήματος	105
6.1	Εισαγωγή	105
6.2	Κρυφά Μαρκοβιανά Μοντέλα (HMMs)	106
6.2.1	Δομή ενός HMM	106
6.2.2	Υπολογισμός της πιθανοφάνειας του μοντέλου	108
6.2.3	Εκπαίδευση των παραμέτρων του μοντέλου	109
6.3	Πειραματικά Αποτελέσματα	111
6.3.1	Κατηγοριοποίηση με χρήση GMMs	111
6.3.2	Κατηγοριοποίηση με χρήση HMMs	114
6.3.3	Κατηγοριοποίηση χωρίς τη χρήση του ουδετέρου	117
6.4	Τροποποιήσεις της μεθόδου κατηγοριοποίησης	119
6.4.1	Συνδυασμός ανεξάρτητων μοντέλων για το σχήμα και την υφή	119
6.4.2	Προσαρμογή σε συγκεκριμένο χρήστη	122
6.5	Συμπεράσματα	125
7	Συμπεράσματα	127
7.1	Συμβολή της διπλωματικής εργασίας	127
7.2	Μελλοντικές ερευνητικές κατευθύνσεις	128
	Βιβλιογραφία	130

Κατάλογος Σχημάτων

2.1	Στάδια εκτίμησης του eye gaze. Έχει ληφθεί από το [13].	30
2.2	Πλήρες γεωμετρικό μοντέλο εκτίμησης του eye gaze. Έχει ληφθεί από το [12].	33
2.3	Αποτελέσματα εκτίμησης του gaze στον οδηγό ενός αυτοκινήτου. Έχει ληφθεί από το [15].	35
2.4	Προσέγγιση ιδιοχώρων για την εκτίμηση του eye gaze. Έχει ληφθεί από το [2].	38
2.5	Ταξινομητές βασισμένοι σε AAMs και SVMs. Έχει ληφθεί από το [32].	41
3.1	Τριγωνοποιημένο σχήμα βάσης s_0 και 3 πρώτα διανύσματα σχήματος σχεδιασμένα πάνω στον γράφο του s_0 . Έχει ληφθεί από το [19]. . . .	45
3.2	Εικόνα της υψής βάσης A_0 και 3 πρώτες ιδιοεικόνες A_i , ορισμένες εντός του γράφου βάσης s_0 . Έχει ληφθεί από το [19].	46
3.3	Διαδικασία δημιουργίας πρωτότυπης εικόνας μεταβάλλοντας τις παραμέτρους σχήματος και υψής του AAM. Έχει ληφθεί από το [19].	47
3.4	Παραδείγματα εικόνων νοηματικής γλώσσας από τις βάσεις δεδομένων της ENG του IEL και της BU400.	51
3.5	Μεταβολή των 3 πρώτων συνιστωσών σχήματος και υψής για τη βάση της ENG του IEL στο διάστημα $[-3\sqrt{\lambda_i}, 3\sqrt{\lambda_i}]$	53
3.6	Μεταβολή των 3 πρώτων συνιστωσών σχήματος για τη βάση BU400 στο διάστημα $[-3\sqrt{\lambda_i}, 3\sqrt{\lambda_i}]$	54
3.7	Παραδείγματα εφαρμογής του AAM για τη βάση δεδομένων της ENG του IEL.	55
3.8	Παραδείγματα εφαρμογής του AAM για τη βάση δεδομένων BU400. . .	56
3.9	Μεταβολή των 3 πρώτων συνιστωσών σχήματος και υψής του τοπικού μοντέλου των ματιών για τη βάση της ENG του IEL στο διάστημα $[-3\sqrt{\lambda_i}, 3\sqrt{\lambda_i}]$	57
3.10	Παραδείγματα εφαρμογής του τοπικού AAM των ματιών για τη βάση δεδομένων της ENG του IEL.	58
3.11	Παραδείγματα διαφορετικών προσώπων από τη βάση “The extended Cohn-Kanade (CK+)” του CMU.	59

3.12	Μεταβολή των 3 πρώτων συνιστωσών υψής του γενικού AAM για τη βάση “CK+” του CMU στο διάστημα $[-3\sqrt{\lambda_i}, 3\sqrt{\lambda_i}]$	59
4.1	Κατευθύνσεις του βλέματος στα αριστερά και στα δεξιά μαζί με τα αντίστοιχα ιστογράμματα φωτεινότητας. Έχει ληφθεί από το [35].	62
4.2	Παραδείγματα εφαρμογής του Top-Hat μετασχηματισμού και της εξαγωγής περιγράμματος για τη βάση εικόνων της ENG του IEL.	65
4.3	Πίνακες συσσώρευσης για τον κυκλικό μετασχηματισμό Hough.[Έχει ληφθεί από το NTUA.]	66
4.4	Παραδείγματα εφαρμογής του κυκλικού μετασχηματισμού Hough και κατασκευή ιστογραμμάτων φωτεινότητας για τη βάση εικόνων της ENG του IEL.	67
4.5	Σύγκριση των εκτιμήσεων για το κέντρο της ίριδας για τη βάση εικόνων της ENG του IEL (με μπλε χρώμα η εκτίμηση του AAM και με κόκκινο της παρούσας μεθόδου).	69
4.6	Εκτίμηση της πόζας κεφαλιού από τις παραμέτρους του αφινικού μετασχηματισμού ενός τραπεζίου ορισμένου στο επίπεδο του προσώπου.	70
4.7	Αντιμετώπιση της ασάφειας του tilt αξιοποιώντας τη θέση της μύτης. Έχει ληφθεί από το [9].	72
4.8	Παραδείγματα εφαρμογής της μεθόδου υπολογισμού της πόζας κεφαλιού για τη βάση δεδομένων της ENG του IEL.	75
4.9	Παραδείγματα εφαρμογής της μεθόδου υπολογισμού της πόζας κεφαλιού για τη βάση δεδομένων BU400.	76
4.10	Σφαιρικό γεωμετρικό μοντέλο για την εκτίμηση του eye gaze. Έχει ληφθεί από το [15].	77
4.11	Γεωμετρική μέθοδος υπολογισμού του eye gaze (κάτοψη του ματιού).	78
4.12	Τυπικές τιμές παραμέτρων (διαστάσεις σε mm) του ανθρώπινου ματιού για έναν μέσο ενήλικα άνθρωπο. Έχει ληφθεί από το διαδίκτυο.	81
4.13	Παραδείγματα εφαρμογής του γεωμετρικού μοντέλου εκτίμησης του eye gaze, συμπεριλαμβανομένης της πόζας κεφαλιού, για τη βάση δεδομένων της ENG του IEL.	83
4.14	Παραδείγματα εφαρμογής του γεωμετρικού μοντέλου εκτίμησης του eye gaze, χωρίς πληροφορία για την πόζα κεφαλιού, για τη βάση δεδομένων της ENG του IEL.	84
4.15	Παραδείγματα εφαρμογής του γεωμετρικού μοντέλου εκτίμησης του eye gaze, με εύρεση της θέσης της ίριδας χρησιμοποιώντας μορφολογικό φιλτράρισμα, για τη βάση δεδομένων της ENG του IEL.	85
5.1	Παραδείγματα από την κατηγοριοποίηση του eye gaze για τη βάση δεδομένων της ENG του IEL.	93

5.2	Ποσοστά επιτυχούς κατηγοριοποίησης του eye gaze με χρήση των παραμέτρων του AAM και μείγματος 2 γκαουσιανών για τη βάση δεδομένων της ENΓ του IEL.	94
5.3	Τύποι «κελιών» και μπλοκ που χρησιμοποιούνται για την κατασκευή των HOGs. Έχει ληφθεί από το [7].	96
5.4	Ποσοστά επιτυχούς κατηγοριοποίησης του eye gaze με χρήση των HOG χαρακτηριστικών και μείγματος 2 γκαουσιανών για τη βάση δεδομένων της ENΓ του IEL.	98
5.5	Παραδείγματα εφαρμογής της μεθόδου υπολογισμού των γωνιών διεύθυνσης του gaze από τις πιθανοφάνειες των GMM κάθε κλάσης με χρήση των AAM παραμέτρων, για τη βάση δεδομένων της ENΓ του IEL.	102
5.6	Παραδείγματα εφαρμογής της μεθόδου υπολογισμού των γωνιών διεύθυνσης του gaze από τις πιθανοφάνειες των GMM κάθε κλάσης με χρήση των HOG χαρακτηριστικών, για τη βάση δεδομένων της ENΓ του IEL.	103
6.1	Γραμμικό δεξιόστροφο (left-right) HMM	107
6.2	Αποτελέσματα κατηγοριοποίησης συναισθήματος με χρήση GMMs για τη βάση δεδομένων “CK+” του CMU σε σχέση με τον αριθμό των γκαουσιανών.	112
6.2	Αποτελέσματα κατηγοριοποίησης συναισθήματος με χρήση GMMs για τη βάση δεδομένων “CK+” του CMU σε σχέση με τον αριθμό των γκαουσιανών (Συνέχεια).	113
6.3	Αποτελέσματα κατηγοριοποίησης συναισθήματος με βάση τις παραμέτρους σχήματος και χρήση HMMs για τη βάση δεδομένων “CK+” του CMU σε σχέση με τον αριθμό των καταστάσεων και για $M = 1$	115
6.3	Αποτελέσματα κατηγοριοποίησης συναισθήματος με βάση τις παραμέτρους σχήματος και χρήση HMMs για τη βάση δεδομένων “CK+” του CMU σε σχέση με τον αριθμό των καταστάσεων και για $M = 1$ (Συνέχεια).	116
6.4	Αποτελέσματα κατηγοριοποίησης συναισθήματος με χρήση GMMs για τη βάση δεδομένων “CK+” του CMU σε σχέση με τον αριθμό των γκαουσιανών χωρίς να συμπληρωθεί το ουδέτερο συναίσθημα.	117
6.4	Αποτελέσματα κατηγοριοποίησης συναισθήματος με χρήση GMMs για τη βάση δεδομένων “CK+” του CMU σε σχέση με τον αριθμό των γκαουσιανών χωρίς να συμπληρωθεί το ουδέτερο συναίσθημα (Συνέχεια).	118

- 6.5 Αποτελέσματα κατηγοριοποίησης συναισθήματος με χρήση ανεξάρτητων GMMs για το σχήμα και την υφή, για τη βάση δεδομένων “CK+” του CMU, σε σχέση με τη μεταβολή της παραμέτρου θ και για αριθμό γκαουσιανών $M = 1$ 120
- 6.6 Αποτελέσματα κατηγοριοποίησης συναισθήματος με χρήση ανεξάρτητων GMMs για το σχήμα και την υφή, για τη βάση δεδομένων “CK+” του CMU, σε σχέση με τον αριθμό των γκαουσιανών και για $\theta = 0.05$ 120
- 6.7 Αποτελέσματα κατηγοριοποίησης συναισθήματος με χρήση ανεξάρτητων HMMs για το σχήμα και την υφή, για τη βάση δεδομένων “CK+” του CMU, σε σχέση με τον αριθμό των καταστάσεων, κρατώντας τα τελευταία 6 καρέ και για $\theta = 0.05$ και $M = 1$ 121
- 6.8 Αποτελέσματα κατηγοριοποίησης συναισθήματος με χρήση GMMs για τη βάση δεδομένων “CK+” του CMU σε σχέση με τον αριθμό των γκαουσιανών αφού πρώτα γίνει η αφαίρεση των παραμέτρων του ουδέτερου συναισθήματος. 123
- 6.8 Αποτελέσματα κατηγοριοποίησης συναισθήματος με χρήση GMMs για τη βάση δεδομένων “CK+” του CMU σε σχέση με τον αριθμό των γκαουσιανών αφού πρώτα γίνει η αφαίρεση των παραμέτρων του ουδέτερου συναισθήματος (Συνέχεια). 124
- 6.9 Αποτελέσματα κατηγοριοποίησης συναισθήματος με χρήση HMMs για τη βάση δεδομένων “CK+” του CMU, σε σχέση με τον αριθμό των καταστάσεων, κρατώντας τα τελευταία 6 καρέ GMMs, αφού πρώτα γίνει η αφαίρεση των παραμέτρων του ουδέτερου συναισθήματος και για $M = 2$ 124

Κεφάλαιο 1

Εισαγωγή

1.1 Όραση Υπολογιστών

Η όραση υπολογιστών είναι η επιστήμη και η τεχνολογία που κάνει τις μηχανές να “βλέπουν”, με την έννοια της εξαγωγής συμβολικής πληροφορίας από μια εικόνα ή ακολουθία εικόνων, η οποία είναι χρήσιμη για την επίλυση διάφορων προβλημάτων.[20] Στόχος είναι να μπορούμε να κατασκευάσουμε συστήματα που να μπορούν να αντιλαμβάνονται τον ορατό κόσμο με ανθρώπινη ευφυΐα, στόχος από τον οποίο σήμερα απέχουμε πολύ. Η όραση υπολογιστών έχει τις ρίζες της στην Τεχνητή Νοημοσύνη και από τότε έχει αναπτυχθεί με μεγάλους ρυθμούς συνδυάζοντας μαζί πολλά επιστημονικά πεδία, όπως η Επεξεργασία Σημάτων, η Αναγνώριση Προτύπων, τα Μαθηματικά, η Φυσική, η Νευροβιολογία, ο Αυτόματος Έλεγχος και Ρομποτική και άλλες ακόμα.

Η όραση υπολογιστών βρίσκει πολλές εφαρμογές σε διάφορους τομείς όπως η Επεξεργασία Εικόνων, η Ρομποτική, η Βιοιατρική Τεχνολογία, η Επικοινωνία Ανθρώπου-Υπολογιστή, η Τηλεπισκόπηση, τα Ευφυή Συστήματα, η Οργάνωση Πληροφορίας, ο Κινηματογράφος και οι άλλες Ψηφιακές Τέχνες. Τυπικά προβλήματα της όρασης υπολογιστών είναι:

- η Αναγνώριση Αντικειμένων,
- η Εκτίμηση Κίνησης,
- η τρισδιάστατη Ανακατασκευή Σκηνής από διάφορες εικόνες της
- η Αποκατάσταση Εικόνας.

1.2 Το Eye Gaze στην όραση υπολογιστών

1.2.1 Το γενικότερο πρόβλημα της αναγνώρισης ενός αντικειμένου

Το πρόβλημα της αναγνώρισης ενός αντικειμένου σε μία εικόνα ή βίντεο (ακολουθία εικόνων) αποτελεί ένα από τα βασικότερα προβλήματα της όρασης υπολογιστών που βρίσκει τεράστιες εφαρμογές στην επικοινωνία ανθρώπου υπολογιστή. Ο βασικός στόχος είναι να εξάγουμε συμβολικές πληροφορίες από τα δεδομένα, δηλαδή τις εικόνες, οι οποίες θα μας οδηγήσουν να αναγνωρίσουμε το ζητούμενο αντικείμενο και την κατάσταση στην οποία αυτό εμφανίζεται μέσα στην εικόνα (για παράδειγμα αν έχει υποστεί κάποιου είδους περιστροφή). Στη συνέχεια, σε αρκετές περιπτώσεις, το αρχικό πρόβλημα ανάγεται σε ένα απλό πρόβλημα της Αναγνώρισης Προτύπων. Στις περισσότερες περιπτώσεις είναι απαραίτητο να διαθέτουμε εξ' αρχής ένα αρκετά σημαντικό σύνολο εικόνων, το οποίο αποτελεί τα δεδομένα εκπαίδευσης. Τα βασικά στάδια ενός κλασικού προβλήματος αναγνώρισης είναι:

- Δημιουργία ενός μοντέλου που να αναπαριστά το ζητούμενο αντικείμενο με βάση το σύνολο των δεδομένων εκπαίδευσης.
- Επεξεργασία των νέων δεδομένων και προσαρμογή του μοντέλου σε αυτά.
- Εξαγωγή της συμβολικής πληροφορίας είτε απευθείας από τις παραμέτρους του μοντέλου, είτε από κάποια επεξεργασία τους.
- Αναγνώριση του αντικειμένου και της κατάστασης του με βάση τη συμβολική πληροφορία ή εφαρμογή κάποιας τεχνικής από την περιοχή της αναγνώρισης προτύπων με σκοπό την τελική αναγνώριση του αντικειμένου.

1.2.2 Η αναγνώριση και μοντελοποίηση του ανθρώπινου προσώπου

Ο άνθρωπος παράγοντας παίζει τον κυρίαρχο ρόλο σε οποιαδήποτε δραστηριότητα συμβαίνει στον κόσμο. Είναι επομένως πρωταρχικής σημασίας ζήτημα ο υπολογιστής να μπορεί να αναγνωρίζει την παρουσία του ανθρώπου σε μια σκηνή. Στις περισσότερες περιπτώσεις όμως δεν μας αρκεί μόνο αυτό αλλά πρέπει να ξέρουμε και σε τι κατάσταση βρίσκεται ο άνθρωπος καθώς και την αλληλεπίδρασή του με το υπόλοιπο περιβάλλον. Βασικό στοιχείο που καθορίζει την κατάσταση ενός ανθρώπου αποτελεί το πρόσωπο. Το πρόσωπο περιέχει τεράστια πληροφορία για έναν άνθρωπο αφού αποτελεί το κύριο στοιχείο για την αναγνώριση του. Περιέχει, επίσης, λόγω της τεράστιας εκφραστικότητας του και πολλά στοιχεία για την συναισθηματική

κατάσταση στην οποία βρίσκεται, την ομιλία του καθώς και το που εστιάζει την προσοχή του.

Βασικά προβλήματα που σχετίζονται με την αναγνώριση και μοντελοποίηση του ανθρώπινου προσώπου, πέρα του βασικού προβλήματος της αναγνώρισης του σε μία σκηνή, είναι η αναγνώριση της ταυτότητας ενός ατόμου, ο υπολογισμός της θέσης και της πόζας του κεφαλιού, η εκτίμηση για το πού κοιτάει (Eye Gaze) η αναγνώριση εκφράσεων του προσώπου και ο προσδιορισμός της συναισθηματικής κατάστασης του ατόμου. Ένα από τα κύρια χαρακτηριστικά του προσώπου αποτελούν τα μάτια, τα οποία μπορούν να μας δώσουν μεγάλη πληροφορία για όλες τις παραπάνω περιοχές.

1.2.3 Ο ρόλος του Eye Gaze

Τα μάτια είναι ένα κατεξοχήν όργανο εισροής πληροφοριών. Ο άνθρωπος με αυτά αντιλαμβάνεται τον εξωτερικό, κόσμο και αποκτά πληροφορίες τόσο γνωστικού όσο και ψυχαγωγικού τύπου μέσω κειμένου, εικόνων και βίντεο. Αποτελούν όμως και ένα μέσο εξαγωγής πληροφοριών, εκφράζοντας κυρίως την κατεύθυνση του βλέματος και το που εστιάζει ένα άτομο την προσοχή του. Επίσης, συμμετέχουν και στην εκδήλωση συναισθημάτων, όπως και όλο το υπόλοιπο πρόσωπο. Συγκεκριμένα, υπάρχουν συναισθήματα που εκδηλώνονται κυρίως με κάποια μεταβολή στην περιοχή των ματιών. Οι άνθρωποι, γενικά, μπορούν εύκολα να εκτιμήσουν την κατεύθυνση eye gaze, από την περιστροφή του κεφαλιού και τις γωνίες των ματιών.

Σαν κατεύθυνση του Eye Gaze μπορούμε να ορίσουμε τη γραμμή του βλέματος ή τα σημεία στα οποία εστιάζουμε. Τα σημεία εστίασης προκύπτουν αν επεκτείνουμε την γραμμή βλέματος μέχρι να συναντήσει το επίπεδο κάποιου αντικείμενου. Αναπτύσσοντας ένα σύστημα που να έχει τη δυνατότητα της αυτόματης εκτίμησης του eye gaze θα έχουμε κάνει ένα αρκετά σηματονικό βήμα στην προσπάθεια μας να κάνουμε έναν υπολογιστή να επικοινωνεί αποτελεσματικά με έναν άνθρωπο. Ο υπολογιστής, γνωρίζοντας την κατεύθυνση του Eye Gaze μπορεί προβλέπει μια επιθυμία ή συμπεριφορά του χρήστη και να δρα ανάλογα. Μπορούν έτσι να αποφευχθούν κάποιες δυσάρεστες καταστάσεις που προκαλούνται από έλλειψη προσοχής ή κούραση του ανθρώπου.

Εφαρμογές του Eye Gaze

Πολλές πεδία εφαρμογών μπορούν να ωφεληθούν από τη χρήση ενός συστήματος αυτόματης εξαγωγής eye gaze. Στα συστήματα ασφαλείας των αυτοκινήτων θα μπορούσε να ενσωματωθεί, ένα τέτοιο σύστημα προκειμένου να προειδοποιεί και να επαναφέρει τον οδηγό σε κατάσταση εγρήγορσης, εάν η προσοχή του αφαιρεθεί από κάτι άλλο ή αν εμφανίσει σημάδια κόπωσης.

Επίσης, ένα τέτοιο σύστημα θα μπορούσε να βρει εφαρμογή και στην εξυπηρέτηση ανθρώπων με ειδικές ανάγκες, για τους οποίους η ανάγκη για επικοινωνία με τον

υπολογιστή είναι αναγκαία αλλά όχι πάντοτε εύκολη. Θα μπορούσε, για παράδειγμα ο υπολογιστής να αναγνωρίζει τις ανάγκες και τις επιθυμίες ενός ανθρώπου με κινητικά προβλήματα από την μεταβολή του eye gaze και να τον εξυπηρετεί ανάλογα. Έτσι, ένα τέτοιο άτομο χρησιμοποιώντας τα μάτια του μπορεί να αντισταθμίσει κατά κάποιο τρόπο τα προβλήματα και τους περιορισμούς που προκύπτουν από την αναπηρία του.

Επιπλέον, το eye gaze μπορεί να χρησιμοποιηθεί και στην δημιουργία ενός συστήματος αυτόματης αναγνώρισης της νοηματικής γλώσσας, βοηθώντας έτσι και άτομα με προβλήματα ομιλίας και ακοής στην επικοινωνία τους με τον υπολογιστή και τις δυνατότητες που προσφέρει, όπως το διαδίκτυο, μια επικοινωνία που ακόμα και σήμερα είναι πάρα πολύ δύσκολη.

Τέλος, το eye gaze βρίσκει εφαρμογές και στον τομέα του marketing και της διαφήμισης. Μια ιστοσελίδα, για παράδειγμα, καταγράφοντας σε ποια προϊόντα και σημεία του site εστιάζει την προσοχή του ένας πελάτης κατά την πλοήγηση του σε αυτό, μπορεί να συγκεντρώσει χρήσιμες πληροφορίες που θα τον βοηθήσουν στον καλύτερο επασχεδιασμό ή βελτίωση του site, καθώς και στην πιο αποτελεσματική και στοχευμένη επιλογή και εμφάνιση των διαφημίσεων.

1.3 Περιεχόμενο και δομή της παρούσας εργασίας

1.3.1 Σκοπός της διπλωματικής

Σκοπός της διπλωματικής είναι η εφαρμογή διαφόρων τεχνικών της όρασης υπολογιστών και της αναγνώρισης προτύπων, με στόχο την επίλυση προβλημάτων που προκύπτουν από την αναγνώριση και μοντελοποίηση του ανθρώπινου προσώπου, καθώς και η σύγκριση τους με βάση τα αποτελέσματα που δίνουν. Το βασικό πρόβλημα με το οποίο ασχολείται, στο μεγαλύτερο μέρος της, η παρούσα διπλωματική είναι αυτό της εκτίμησης του eye gaze. Για το λόγο αυτό αναπτύσσονται και υλοποιούνται αρκετές διαφορετικές τεχνικές και στο τέλος γίνεται η σύγκρισή τους τόσο ποιοτικά όσο και ποσοτικά, εφόσον κάτι τέτοιο είναι εφικτό. Επιπλέον, εκτός από το πρόβλημα του eye gaze, αναλύονται και γίνεται προσπάθεια να επιλυθούν και άλλα σημαντικά προβλήματα που σχετίζονται με την οπτική μοντελοποίηση του ανθρώπινου προσώπου. Συγκεκριμένα, γίνεται αναφορά και προσπάθεια εκτίμησης της πόζας του κεφαλιού, σαν ένα επιμέρους πρόβλημα του υπολογισμού του eye gaze. Τέλος, αναλύεται το πρόβλημα της αναγνώρισης συναισθήματος και γίνεται προσπάθεια προσδιορισμού του συναισθήματος ενός ατόμου με βάση τα χαρακτηριστικά του προσώπου του.

1.3.2 Συνεισφορά της διπλωματικής

Η κύρια συνεισφορά της συγκεκριμένης διπλωματικής έγκειται στη συστηματική μελέτη των παραπάνω προβλημάτων και στην υλοποίηση και αξιολόγηση των διαφορετικών μεθόδων που οδηγούν στην επίλυση τους. Πιο συγκεκριμένα οι επιστημονικές της συνεισφορές συνοψίζονται στους εξής άξονες:

- Στην όσο το δυνατόν πιο ακριβή οπτική μοντελοποίηση του ανθρώπινου προσώπου και ιδιαίτερα της περιοχής των ματιών με χρήση των Ενεργών Μοντέλων Εμφάνισης (Active Appearance Models). Σκοπός είναι να προσαρμόζουμε το παραγόμενο μοντέλο σε νέες εικόνες και να εξάγουμε ένα σχετικά μικρό διάνυσμα παραμέτρων που να περιγράφει με ικανοποιητική ακρίβεια τις παραπάνω περιοχές.
- Στην ανάλυση και υλοποίηση ενός συστήματος υπολογισμού του eye gaze χρησιμοποιώντας ένα πλήθος διαφορετικών τεχνικών. Αναπτύσσονται τεχνικές του βασίζονται στη μοντελοποίηση του προσώπου με AAMs που αναφέρθηκε παραπάνω, αλλά και τεχνικές που χρησιμοποιούν τα AAMs ως ένα αρχικό μόνο στάδιο της μεθόδου. Επίσης, για την εκτίμηση του eye gaze χρησιμοποιούνται γεωμετρικές τεχνικές αλλά και μέθοδοι αναγνώρισης προτύπων.
- Στην υλοποίηση και ποσοτική αξιολόγηση ενός συστήματος αναγνώρισης συναισθήματος σε εικόνες, με χρήση AAMs και Κρυφών Μαρκοβιανών Μοντέλων (HMMs). Το σύστημα επιτυγχάνει να αναγνωρίζει ανάμεσα σε 7 διαφορετικά συναισθήματα, ενώ υπάρχει η δυνατότητα το σύστημα να προσαρμόζεται κάθε φορά στο νέο χρήστη, αυξάνοντας έτσι κατά πολύ την απόδοσή του.

1.3.3 Οργάνωση του περιεχομένου της διπλωματικής

Το περιεχόμενο της διπλωματικής είναι οργανωμένο σε 6 κεφάλαια ως εξής:

- Στο **κεφάλαιο 2** περιγράφονται συνοπτικά αρκετές μέθοδοι της βιβλιογραφίας σχετικά με τα θέματα της εκτίμησης του eye gaze και της αναγνώρισης συναισθήματος που εξετάζονται στην συγκεκριμένη διπλωματική. Σκοπός του κεφαλαίου δεν είναι να αναπτυχθούν διεξοδικά οι λεπτομέρειες της κάθε τεχνικής αλλά να επισυμανθούν τα κύρια στάδια των μεθόδων, καθώς και οι εφαρμογές τους. Ιδιαίτερη έμφαση δίνεται σε τεχνικές που βασίζονται στη χρήση των Active Appearance Models.
- Στο **κεφάλαιο 3** γίνεται μια σύντομη περιγραφή των Ενεργών Μοντέλων Εμφάνισης (AAMs), καθώς και της μεθόδου με την οποία γίνεται η προσαρμογή τους σε νέες εικόνες με ταχύτητα και ακρίβεια. Στη συνέχεια με χρήση

των AAMs μοντελοποιείται το ανθρώπινο πρόσωπο, σύμφωνα με τις βάσεις δεδομένων που θα χρησιμοποιηθούν, και παρουσιάζονται τα αποτελέσματα. Τέλος, επειδή το κύριο ενδιαφέρον της διπλωματικής εστιάζεται στην περιοχή των ματιών, υλοποιούμε και ένα τοπικό μοντέλο για την περιοχή αυτή, με βάση το αρχικό μοντέλο για ολόκληρο το πρόσωπο.

- Στο **κεφάλαιο 4** αναλύονται μέθοδοι εκτίμησης του eye gaze που δεν χρησιμοποιούν κάποια τεχνική κατηγοριοποίησής του σε κλάσεις. Αρχικά, περιγράφεται μια απλή τεχνική που βασίζεται σε μορφολογικό φιλτράρισμα της περιοχής των ματιών και κατασκευή ιστογραμμάτων για την περιοχή γύρω από την ίριδα, με περιορισμένα ωστόσο αποτελέσματα. Στη συνέχεια αναλύεται μια μέθοδος υπολογισμού του eye gaze που στηρίζεται στο γεωμετρικό μοντέλο του ματιού. Για την κατασκευή του μοντέλου αυτού χρησιμοποιούνται τα σημεία landmarks της προσαρμοζόμενης μάσκας σχήματος των AAMs. Επειδή για την ακριβέστερη εκτίμηση του eye gaze το γεωμετρικό μοντέλο χρειάζεται και την πόζα του κεφαλιού, αναπτύσσεται μια τεχνική υπολογισμού της πόζας χρησιμοποιώντας και πάλι σημεία της μάσκας σχήματος.
- Στο **κεφάλαιο 5** αναπτύσσονται μέθοδοι που βασίζονται στην χρήση τεχνικών της αναγνώρισης προτύπων με στόχο την κατηγοριοποίηση του eye gaze σε διακριτές κλάσεις. Συγκεκριμένα χρησιμοποιούνται μοντέλα μείγματος γκαουσιανών συναρτήσεων GMMs προκειμένου να κατατάξουμε το eye gaze σε 6 βασικές κατηγορίες, παίρνοντας έτσι μια αρκετά καλή γενική διακριτή εκτίμηση. Στη συνέχεια, αξιοποιώντας την λογαριθμική πιθανοφάνεια το eye gaze να ανήκει σε κάθε μια από τις 6 κλάσεις καταφέρνουμε να επιτύχουμε μια συνεχή εκτίμηση του gaze. Σαν είσοδος στους ταξινομητές χρησιμοποιούνται τόσο οι παράμετροι σχήματος και υψής του τοπικού active appearance μοντέλου, όσο και χαρακτηριστικά που προκύπτουν από την εφαρμογή των ιστογραμμάτων προσανατολισμένης κλίσης (Histograms of Oriented Gradients “HOGs”) στην περιοχή των ματιών.
- Στο **κεφάλαιο 6** αναλύεται το πρόβλημα της αναγνώρισης συναισθήματος. Για τον σκοπό αυτό χρησιμοποιούνται παράμετροι που προκύπτουν από την εφαρμογή των AAMs σε όλη την περιοχή του προσώπου, προκειμένου να επιτευχθεί κατηγοριοποίηση σε 7 βασικές κατηγορίες συναισθημάτων (θυμός, περιφρόνηση, αηδία, φόβος, ευτυχία, θλίψη, έκπληξη) καθώς και στο ουδέτερο συναίσθημα. Σαν ταξινομητές χρησιμοποιούνται Κρυφά Μαρκοβιανά Μοντέλα (HMMs) με διάφορους αριθμούς καταστάσεων και γκαουσιανών σε κάθε μια κατάσταση. Τέλος, παρόλο που το σύστημα να λειτουργεί ικανοποιητικά ανεξαρτήτως χρήστη, υπάρχει η δυνατότητα εύκολης προσαρμογής του σε έναν συγκεκριμένο χρήστη εκτοξεύοντας έτσι τα ποσοστά αναγνώρισης.

- Στο **κεφάλαιο 7** παρουσιάζονται τα συμπεράσματα που προκύπτουν από το σύνολο της διπλωματικής και συνοψίζονται οι επιστημονικές συνεισφορές της. Επίσης, αναφέρονται και κάποιες μελλοντικές κατευθύνσεις και προεκτάσεις της.

Κεφάλαιο 2

Υπάρχουσα Έρευνα (Background)

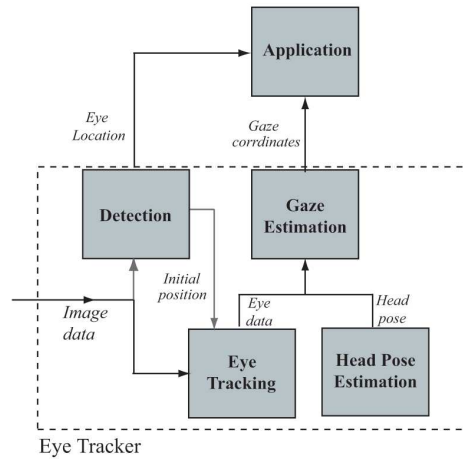
2.1 Εισαγωγή

Το πρόβλημα της μοντελοποίησης του ανθρώπινου ματιού και της εκτίμησης του eye gaze αποτελεί ένα από τα διαχρονικότερα και σημαντικότερα προβλήματα της όρασης υπολογιστών. Για το λόγο αυτό έχουν ήδη αναπτυχθεί και παρουσιαστεί πάρα πολλές διαφορετικές τεχνικές για την επίλυση τους, ανάλογα με τις επικρατούσες συνθήκες και τον σκοπό της εφαρμογής. Σε αρκετές από αυτές, παράλληλα με το πρόβλημα του eye gaze αναλύεται και το πρόβλημα εκτίμησης της πόζας κεφαλιού ως ένα ενδιάμεσο στάδιο της συνολικής μεθόδου.

Επιπλέον, ένα σημαντικό πεδίο στο οποίο βρίσκεται εφαρμογή η μοντελοποίηση των ματιών και γενικότερα του προσώπου είναι αυτό της αναγνώρισης συναισθήματος. Και για αυτό το σημαντικό θέμα έχει αναπτυχθεί ένα σημαντικό πλήθος τεχνικών, αρκετές από τις οποίες χρησιμοποιούν μεθόδους από τις παραπάνω περιοχές.

Στο κεφάλαιο αυτό παρουσιάζονται αρκετές μέθοδοι της διεθνούς βιβλιογραφίας σχετικά την εκτίμηση και παρακολούθηση του eye gaze αλλά και την αναγνώριση συναισθήματος. Σκοπός του κεφαλαίου δεν είναι να παρουσιάσει αναλυτικά τη μεθοδολογία που ακολουθεί η κάθε τεχνική, αλλά να αναδείξει τη γενική ιδέα και τις διαφορές της με άλλες παρόμοιες μεθόδους.

Έχει γίνει προσπάθεια οι διάφορες τεχνικές να παρουσιαστούν όσο γίνεται πιο κατηγοριοποιημένες ανάλογα με το είδος της μεθόδου που χρησιμοποιούν, με έμφαση σε μεθόδους που χρησιμοποιούν ενεργά μοντέλα εμφάνισης (AAM). Επίσης, αρκετό βάρος δίνεται και στις εφαρμογές που βρίσκουν οι τεχνικές αυτές καθώς και σε ενδεχόμενους περιορισμούς που μπορεί να έχουν. Τέλος, το πρόβλημα της εκτίμησης της πόζας προσεγγίζεται σαν ένα υποπρόβλημα του eye gaze και για αυτό δεν κρίνεται σκόπιμο να γίνει αναφορά σε μεθόδους αποκλειστικής εκτίμησης της



Σχήμα 2.1: Στάδια εκτίμησης του eye gaze. Έχει ληφθεί από το [13].

πόζας. Σε περίπτωση που κάποιος θέλει να αναζητήσει περισσότερες λεπτομέρειες για το συγκεκριμένο πρόβλημα, αρκετές μέθοδοι παρουσιάζονται από τους Murphy-Chutorian and Trivedi.[23]

2.2 Τεχνικές εκτίμησης του eye gaze

2.2.1 Γενικά

Ο όρος eye gaze μπορεί να γίνει αντιληπτός με δύο διαφορετικούς τρόπους: σαν την διεύθυνση του διανύσματος του gaze στο χώρο ή σαν το σημείο ενός σταθερού επιπέδου αναφοράς (π.χ. μια οθόνη) στο οποίο εστιάζουν τα μάτια μας. Και οι δύο αυτές προσεγγίσεις έχουν χρησιμοποιηθεί ευρέως, ανάλογα με τον σκοπό της εφαρμογής. Τα βασικά στάδια και επιμέρους προβλήματα που προκύπτουν κατά την εκτίμηση του eye gaze είναι η εύρεση και παρακολούθηση των ματιών και η εκτίμηση της πόζας κεφαλιού καθώς και ο συνδυασμός τους με χρήση κάποιου μοντέλου με στόχο τον τελικό υπολογισμό του gaze. Η όλη διαδικασία παρουσιάζεται συνοπτικά στο Σχήμα 2.1.

Γενικά, οι μέθοδοι που χρησιμοποιούνται για την εκτίμηση του eye gaze μπορούν να χωριστούν σε παρεμβατικές και μη παρεμβατικές. Οι παρεμβατικές τεχνικές είναι αυτές που απαιτούν τη χρήση επιπρόσθετου εξοπλισμού και άρα έτσι επεμβαίνουν στη συμπεριφορά του χρήστη. Τα πρόσθετα μέσα που χρησιμοποιούνται είναι συνήθως προσαρμοζόμενοι αισθητήρες στο πρόσωπο και τα μάτια, περιοριστικά συστήματα καταγραφής video που απαιτούν προκαθορισμένη θέση κεφαλιού και

κάμερες που τοποθετούνται στο κεφάλι. Οι τεχνικές αυτές δίνουν εξαιρετικής ποιότητας αποτελέσματα και έχουν χρησιμοποιηθεί αρκετά μέχρι σήμερα. Ωστόσο σιγά σιγά χάνουν έδαφος εξαιτίας του υψηλού κόστους του επιπρόσθετου εξοπλισμού, της έλλειψης ευελιξίας κινήσεων για τον χρήστη, καθώς και του ειδικού χώρου αλλά και συνθηκών που συνήθως απαιτούνται για την επιτυχημένη εφαρμογή τους.

Αντίθετα, οι μη παρεμβατικές μέθοδοι δεν χρησιμοποιούν κάποιο είδος επιπρόσθετου εξοπλισμού που να έρχεται σε άμεση επαφή με το χρήστη και να επηρεάζει τη συμπεριφορά του. Η συγκεκριμένη κατηγορία μεθόδων χρησιμοποιεί τεχνικές της Όρασης Υπολογιστών προκειμένου να εξάγει πληροφορία από ένα σύνολο εικόνων και παρουσιάζεται αρκετά ελπιδοφόρα για την δημιουργία ενός ευρείας χρήσης συστήματος αναγνώρισης του gaze με σχετικά χαμηλό κόστος. Γι' αυτό το λόγο στο συγκεκριμένο κεφάλαιο αναφέρονται μη παρεμβατικές τεχνικές μόνο. Οι μη παρεμβατικές τεχνικές διακρίνονται σε αυτές που χρησιμοποιούν υπέρυθρο φως και υπέρυθρες κάμερες και σε εκείνες που χρησιμοποιούν ορατό φως και συμβατικές κάμερες.

Για την εκτίμηση του gaze έχουν χρησιμοποιηθεί αρκετά είδη τεχνικών: 3Δ γεωμετρικά μοντέλα, 2Δ μοντέλα παλινδρόμησης βάσει χαρακτηριστικών, μοντέλα βασισμένα στην εμφάνιση.[13] Βασικά στοιχεία που πρέπει να ληφθούν υπόψιν κατά την επιλογή του κατάλληλου μοντέλου είναι: η ακρίβεια εκτίμησης, η ποιότητα των δεδομένων (εικόνες), η ανεξαρτησία από την πόζα κεφαλιού, το είδος και η ευκολία βαθμονόμησης, η ευελιξία στις κινήσεις του χρήστη, η ευρωστία σε μεταβολές του περιβάλλοντος.

Τα 3Δ γεωμετρικά μοντέλα συνδυάζονται συνήθως με τη χρήση υπέρυθρου φωτός και επιτυγχάνουν πολύ καλή ακρίβεια εκτίμησης. Ωστόσο, χρειάζονται εικόνες υψηλής ανάλυσης εστιασμένες στην περιοχή των ματιών, αφού ένα μικρό σφάλμα στην εκτίμηση κάποιας παραμέτρου μπορεί να οδηγήσει σε μεγάλα σφάλματα στον υπολογισμό του gaze. Μπορούν βέβαια να χρησιμοποιηθούν και με ορατό φως αλλά τότε η ποιότητα των αποτελεσμάτων περιορίζεται.

Τα μοντέλα παλινδρόμησης, μιας και αυτά βασίζονται σε εξαγωγή χαρακτηριστικών, είναι ευαίσθητα στα σφάλματα των χαρακτηριστικών που χρησιμοποιούν αλλά σε μικρότερο βαθμό. Επίσης, χρειάζονται κάποια εξωτερική εκτίμηση της πόζας κεφαλιού και η ακρίβεια τους είναι πιο περιορισμένη. Ωστόσο είναι αρκετά απλά στην υλοποίησή τους και χρησιμοποιούνται κυρίως στην εκτίμηση του σημείου ενδιαφέροντος.

Τέλος, τα μοντέλα βασισμένα στην εμφάνιση είναι πιο εύρωστα και μπορούν να λειτουργήσουν ικανοποιητικά και με πιο κακή ποιότητα εικόνας, μιας και δεν βασίζονται στην εξαγωγή χαρακτηριστικών αλλά χρησιμοποιούν πληροφορία από όλη την περιοχή των ματιών. Ωστόσο, με τον τρόπο αυτό δεν λαμβάνεται υπόψιν η πόζα κεφαλιού (καθολική εκτίμηση του gaze) αλλά μόνο η μεταβολή στα μάτια (τοπική εκτίμηση του gaze). Γι' αυτό το λόγω συνήθως απαιτείται η χρήση ενός εξωτερικού

συστήματος εκτίμησης της πόζας κεφαλιού. Επίσης, δεν απαιτούν την βαθμονόμηση των καμερών και των γεωμετρικών παραμέτρων αλλά η ακρίβεια εκτίμησης είναι πιο μικρή σε σχέση με τα προηγούμενα μοντέλα.

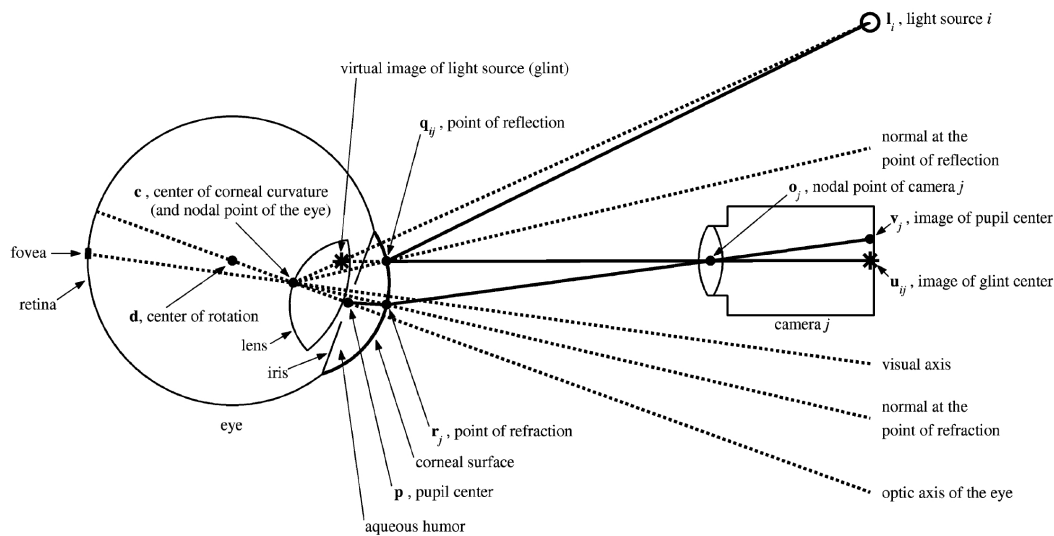
2.2.2 Γεωμετρικά μοντέλα με χρήση υπέρυθρου φωτός

Οι συγκεκριμένες τεχνικές εκτίμησης του gaze βασίζονται στην αλληλεπίδραση της δέσμης του υπέρυθρου φωτός με τα διάφορα στρώματα του ματιού και την καταγραφή των αποτελεσμάτων από κάμερες υπέρυθρων. Ο λόγος για τον οποίο χρησιμοποιούμε υπέρυθρες ακτίνες και όχι ακτίνες από το ορατό φάσμα είναι για να μην γίνονται αντιληπτές από το ανθρώπινο μάτι και να μην τραβούν την προσοχή, επηρεάζοντας το gaze.

Τα βασικά χαρακτηριστικά που πρέπει να προσδιοριστούν έτσι ώστε να προχωρήσει η κατασκευή του γεωμετρικού μοντέλου είναι το κέντρο της ίριδας ή της κόρης και το κέντρο της κερατοειδούς επιφάνειας. Η χρήση υπέρυθρου φωτός κάνει τον εντοπισμό των χαρακτηριστικών αυτών πιο εύκολο και πιο ακριβή. Όταν η δέσμη των ακτίνων είναι έχει την ίδια διεύθυνση με τον οπτικό άξονα της κάμερας τότε έχουμε σαν αποτέλεσμα μια φωτεινή ίριδα, ενώ σε αντίθετη περίπτωση η ίριδα εμφανίζεται σκοτεινή. Και στις δύο περιπτώσεις η ίριδα μπορεί να εντοπιστεί εύκολα με κάποια απλή τεχνική ανίχνευσης κύκλων ή ελλείψεων. Για τον προσδιορισμό του κέντρου του κερατοειδούς θα χρησιμοποιηθούν οι εικόνες Purkinje, που προκύπτουν από τις ανακλάσεις της φωτεινής δέσμης στα διάφορα στρώματα και συγκεκριμένα η πρώτη εικόνα Purkinje το λεγόμενο glint. Το glint εμφανίζεται σαν ένα έντονα φωτεινό σημείο στην περιοχή του ματιού και μπορεί και αυτό να εντοπιστεί εύκολα.

Έχοντας εντοπίσει τη θέση της ίριδας και του glint ο προσδιορισμός του κέντρου του κερατοειδούς και στη συνέχεια του οπτικού άξονα (optic axis) του ματιού γίνεται εφικτός με τη χρήση ενός γεωμετρικού μοντέλου αξιοποιώντας την γεωμετρία των δέσμεων φωτός, των κάμερων και του ματιού στο χώρο καθώς και τις ανατομικές παραμέτρους του ματιού. Συνήθως, η διεύθυνση του gaze ορίζεται σαν τον άξονα όρασης του ματιού (visual axis) ο οποίος διαφέρει λίγο από τον οπτικό άξονα. Το πλήρες γεωμετρικό μοντέλο απεικονίζεται στο Σχήμα 2.2.

Ο αριθμός των φωτεινών πηγών καθώς και των καμερών σχετίζεται άμεσα με τους περιορισμούς που θα έχει το γεωμετρικό μοντέλο καθώς και με τον αριθμό των σημείων που είναι απαραίτητα για την βαθμονόμηση του μοντέλου. Όταν έχουμε μια φωτεινή δέσμη και μια κάμερα τότε χρειάζεται ο προσδιορισμός αρκετών παραμέτρων για την βαθμονόμηση του μοντέλου ενώ πρέπει επίσης η πόζα του κεφαλιού να παραμένει σχεδόν σταθερή γύρω από μια προκαθορισμένη θέση. Αν χρησιμοποιηθούν πολλαπλές φωτεινές πηγές και μία κάμερα τότε μπορεί να προκύπτει σύστημα ανεξάρτητο της πόζας κεφαλιού. Τέλος, αν γίνει και χρήση πολλαπλών καμερών τότε οι παράμετροι που αφορούν τη γεωμετρία του ματιού δεν χρειάζεται να προσδιοριστούν από κάποιο στάδιο βαθμονόμησης. Ωστόσο, ακόμα και σε αυτή την



Σχήμα 2.2: Πλήρες γεωμετρικό μοντέλο εκτίμησης του eye gaze. Έχει ληφθεί από το [12].

περίπτωση είναι απαραίτητη η γνώση για τη γεωμετρία του χώρου και τη βαθμονόμηση των καμερών. Επίσης, σε κάθε περίπτωση χρειάζεται να προσδιοριστεί και η σχέση μεταξύ οπτικού άξονα και άξονα όρασης για την οποία ένα σημείο βαθμονόμησης είναι αρκετό.[12, 13]

Οι Guestrin and Eizenman[12] προχώρησαν στην υλοποίηση ενός συστήματος υπολογισμού του σημείου εστίασης σε μία οθόνη υπολογιστή χρησιμοποιώντας δύο πηγές υπέρυθρου φωτός και μία κάμερα υπέρυθρων με αρκετά καλής ακρίβειας αποτελέσματα. Πειραματικά αποτελέσματά τους έδειξαν ότι τυχόν αποκλίσεις και σφάλματα οφείλονται στη μη σφαιρικότητα του κερατοειδούς και στο θόρυβο κατά την εξαγωγή των χαρακτηριστικών.

Οι Ohno et al.[25] υλοποίησαν ένα ευρείας χρήσης σύστημα εκτίμησης του gaze με στόχο την επικοινωνία ανθρώπου-υπολογιστή σε καθημερινή βάση χωρίς τη χρήση κάποιου άλλου μέσου. Χρησιμοποίησαν ένα σύστημα με μια δέσμη υπέρυθρων και μια κάμερα καθώς και ένα αρκετά πιο απλό γεωμετρικό μοντέλο, υποθέτοντας σταθερή τη θέση του κεφαλιού μπροστά στην οθόνη. Έτσι, η βαθμονόμηση του συστήματος μπορεί να περιοριστεί σε δύο σημεία (θεωρώντας γνωστές και κάποιες ανατομικές παραμέτρους του ματιού), κάνοντας το έτσι πιο εύκολο στη χρήση από μη εξειδικευμένα άτομα.

Οι τεχνικές με χρήση υπέρυθρων δίνουν αρκετά ακριβή αποτελέσματα, ωστόσο η χρήση τους περιορίζεται σε εσωτερικούς και με σταθερή γεωμετρία χώρους ενώ έχουν και ένα σχετικά υψηλό χρηματικό κόστος.

2.2.3 Γεωμετρικά μοντέλα με χρήση συμβατικών καμερών

Οι μέθοδοι που χρησιμοποιούν συμβατικές κάμερες και φυσικό φως αρχίζουν να γίνονται αρκετά δημοφιλείς λόγω της ευκολίας υλοποίησής τους με απλό και σχετικά φτηνό εξοπλισμό. Η ακρίβεια των αποτελεσμάτων δεν είναι πολύ μεγάλη αλλά αυτό στις περισσότερες πρακτικές εφαρμογές δεν αποτελεί μεγάλο πρόβλημα. Τα γεωμετρικά μοντέλα που χρησιμοποιούνται είναι αρκετά πιο απλά από τα προηγούμενα και δεν απαιτούν συγκεκριμένη γεωμετρία χώρου. Τα πιο συνιθισμένα μοντέλα αρκούνται στην εκτίμηση του eye gaze από τη σχετική θέση της κόρης και του κέντρου του ματιού, το σχήμα του οποίου θεωρείται σφαιρικό. Συνήθως απαιτείται και εξωτερική πληροφορία για την πόζα κεφαλιού, έτσι ώστε να υπολογιστεί η σωστή θέση του κέντρου του ματιού και να στη συνέχεια το gaze στον 3D χώρο.

Γεωμετρικά μοντέλα με χρήση Active Appearance Models

Οι Ishikawa et al.[15] υλοποίησαν ένα σύστημα για την παρακολούθηση του gaze του οδηγού στο αυτοκίνητο με στόχο τη βελτίωση της ασφάλειας του, στηριζόμενοι στη χρήση των Active Appearance Models.[19]. Αρχικά, προχώρησαν στην εκτίμηση της πόζας κεφαλιού προσαρμόζοντας 3D AAM[34] στο πρόσωπο του χρήστη, ενώ και οι γωνίες των ματιών προσδιορίζονται από τα σημεία της μάσκας σχήματος των AAMs. Σε αυτό το σημείο είναι δυνατόν να υπολογιστεί η προβολή του κέντρου του ματιού πάνω στην εικόνα. Στη συνέχεια η θέση της κόρης μπορεί να βρεθεί με το ταίριασμα ενός μαύρου δίσκου και μιας έλλειψης στην εικόνα του ματιού. Τέλος, το 3D διάνυσμα του eye gaze υπολογίζεται χρησιμοποιώντας ένα γεωμετρικό μοντέλο που συνδυάζει όλες τις προηγούμενες πληροφορίες καθώς και κάποιες ανατομικές παραμέτρους για το μάτι, οι οποίες προκύπτουν κατά την αρχικοποίηση του συστήματος ζητώντας από το χρήστη να κοιτάξει σε προκαθορισμένα σημεία στο χώρο. Τα αποτελέσματά τους κρίνονται αρκετά ικανοποιητικά όπως φαίνεται και στα καρέ του Σχήματος 2.3. Επίσης, οι Kim and Kim[17] χρησιμοποιούν την ίδια μέθοδο εκτίμησης του eye gaze, εφαρμόζοντας τα αποτελέσματα για να ελέγξουν την κατεύθυνση των ματιών ενός μικρού ρομπότ.

Άλλες γεωμετρικές τεχνικές

Οι Matsumoto and Zelinsky[21] υλοποίησαν ένα σύστημα δύο καμερών (stereo cameras) για την εκτίμηση της πόζας κεφαλιού στον τρισδιάστατο χώρο. Για τον σκοπό αυτό χρησιμοποίησαν ένα 3D μοντέλο βασισμένο σε χαρακτηριστικά του προσώπου, και συγκεκριμένα τις άκρες των ματιών και του στόματος. Για τον υπολογισμό του eye gaze η θέση του κέντρου του ματιού υπολογίζεται από την πόζα κεφαλιού, ενώ το κέντρο της κόρης ανιχνεύεται από τις στέρεο εικόνες με χρήση του



Σχήμα 2.3: Αποτελέσματα εκτίμησης του gaze στον οδηγό ενός αυτοκινήτου. Έχει ληφθεί από το [15].

κυκλικού μετασχηματισμού Hough. Στη συνέχεια η διεύθυνση του gaze υπολογίζεται από το διάνυσμα που συνδέει το κέντρο του ματιού με το κέντρο της κόρης. Το σύστημα δίνει αρκετά καλά αποτελέσματα με ικανοποιητική ταχύτητα ακόμα και για μεγάλες μεταβολές στην πόζα κεφαλιού, ενώ εμφανίζεται και αρκετά εύρωστο σε μικρές επικαλύψεις του προσώπου.

Οι Wang et al.[37] προτείνουν μια μέθοδο υπολογισμού του gaze χρησιμοποιώντας μια μοναδική εικόνα που να εστιάζει αποκλειστικά στο ένα μάτι. Υποθέτοντας ότι το σχήμα της ίριδας στο χώρο είναι κύκλος, προχωρούν στην εκτίμηση του gaze υπολογίζοντας τη διεύθυνση του κύκλου αυτού από την ελλειπτική του προβολή στην εικόνα. Στην προβολική γεωμετρία μια έλλειψη μπορεί να προβληθεί πίσω στο χώρο σε δύο κύκλους με διαφορετικό προσανατολισμό. Στη συνέχεια χρησιμοποιώντας τον ανθρωπομετρικό περιορισμό το κέντρο της σφαίρας του ματιού να ισαπέχει από τις γωνίες του ματιού, επιλέγεται ο σωστός κύκλος και επομένως η σωστή διεύθυνση του gaze σαν το κάθετο διάνυσμα στο επίπεδο του κύκλου στο 3D χώρο. Η ακρίβεια της μεθόδου, όπως προκύπτει από τον προσδιορισμό σημείων εστίασης σε μια οθόνη κρίνεται αρκετά ικανοποιητική. Όσοσο, το σύστημα δεν συνυπολογίζει στην εκτίμησή του την πόζα κεφαλιού και γι' αυτό χρειάζεται ένα εξωτερικό σύστημα για το σκοπό αυτό.

Τέλος, οι Gee and Cipolla[9] προχωρούν στην εκτίμηση του eye gaze βασιζόμενοι αποκλειστικά στην πόζα κεφαλιού. Εστιάζουν κυρίως στην επίλυση του γεωμετρικού προβλήματος της εύρεσης του κάθετου διανύσματος στο επίπεδο του προσώπου, θεωρώντας γνωστά χαρακτηριστικά σημεία του προσώπου, χωρίς να ασχολούνται με το πρόβλημα της εύρεσης των σημείων αυτών.

2.2.4 Μέθοδοι βασισμένες σε ταίριασμα προτύπων και ανάλυση παλινδρόμησης

Οι μέθοδοι της συγκεκριμένης κατηγορίας συνήθως στηρίζονται στην προσαρμογή ενός μοντέλου σε απλές εικόνες του προσώπου και στην εξαγωγή ενός πλήθους συντελεστών που να περιγράφουν το πρόσωπο και τα μάτια. Στη συνέχεια γίνεται αντιστοίχιση του eye gaze με τους συντελεστές αυτούς χρησιμοποιώντας ανάλυση παλινδρόμησης ή κάποια άλλη μέθοδο ταίριασματος προτύπων. Η εκτίμηση του gaze που προκύπτει είναι σε αρκετές περιπτώσεις ανεξάρτητη της πόζας κεφαλιού, επειδή οι συντελεστές που περιγράφουν το πρόσωπο περιέχουν πληροφορία και για την θέση του κεφαλιού. Η ακρίβεια των αποτελεσμάτων της συγκεκριμένης κατηγορίας μεθόδων είναι λίγο πιο περιορισμένη, ωστόσο οι τεχνικές αυτές παρουσιάζουν μεγαλύτερη ευρωστία σε μεταβολές του περιβάλλοντος ενώ χρειάζονται και μικρότερης ποιότητας εικόνες.

Χρήση Active Appearance Models και Morphable Models

Οι Takatani et al.[33] εφαρμόζουν ένα μοντέλο ανάλυσης παλινδρόμησης στους συντελεστές που προκύπτουν από την προσαρμογή των AAMs σε εικόνες του προσώπου, για να εκτιμήσουν ταυτόχρονα την οριζόντια γωνία της πόζας κεφαλιού και του συνολικού eye gaze, συμπεριλαμβανομένης της πόζας. Ωστόσο, υπάρχουν αχρείαστοι συντελεστές που δυσχεραίνουν την σωστή εκμάθηση του πίνακα παραμέτρων της εξίσωσης παλινδρόμησης. Για να αποφύγουν τη χρήση των συντελεστών αυτών χρησιμοποιούν μια μέθοδο επιλογής μοντέλου, η οποία θα στηρίζεται στην ελαχιστοποίηση κάποιου κριτηρίου πιθανοφάνειας και θα κρατά μόνο το απαραίτητο σύνολο συντελεστών. Η προτεινόμενη αυτή μέθοδος συγκρίνεται με τεχνικές ξεχωριστού υπολογισμού της πόζας και του gaze καθώς και με άλλες συμβατικές γεωμετρικές μεθόδους και τα αποτελέσματά της είναι αρκετά καλύτερα. Επίσης, ελέγχεται η ακρίβεια της μεθόδου για διαφορετικά κριτήρια της μεθόδου επιλογής μοντέλου.

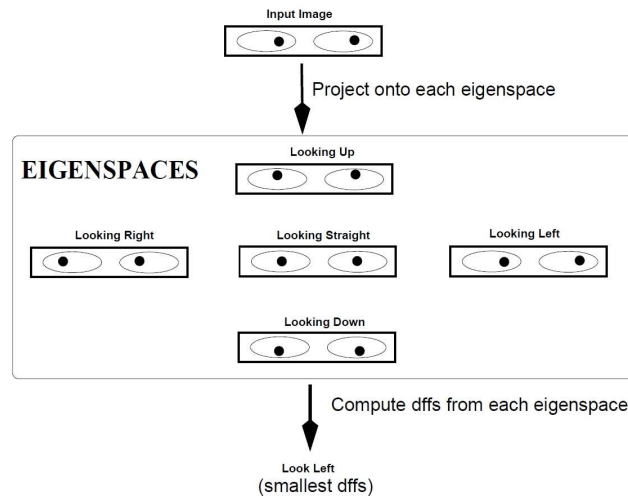
Οι Hansen et al.[14] προσπαθούν να εκτιμήσουν το eye gaze χρησιμοποιώντας μια κάμερα διαδικτύου χαμηλής ανάλυσης και προσαρμόζοντας AAMs για να εξάγουν πληροφορίες για τις γωνίες και την κόρη του ενός ματιού. Στη συνέχεια χρησιμοποιούν μια στοχαστική διαδικασία για να βρουν την αντιστοιχία μεταξύ των σημείων εστίασης σε μια οθόνη και των χαρακτηριστικών του ματιού στην εικόνα. Για να αυξήσουν την ανάλυση της εικόνας και την ακρίβεια υπολογισμού τοποθέτησαν την κάμερα κοντά στα μάτια, περιορίζοντας έτσι σε μεγάλο βαθμό τις κινήσεις του χρήστη. Η συγκεκριμένη μέθοδος χρησιμοποιήθηκε για τη δημιουργία ενός συστήματος πληκτρολόγησης με τη χρήση των ματιών, μιας εφαρμογής πολύ χρήσιμης για άτομα με ειδικές ανάγκες. Για το σκοπό αυτό υλοποιήθηκε και ένα σύστημα πρόβλεψης λέξεων βασισμένο σε Μαρκοβιανά μοντελά.

Τα Morphable Models[16] αποτελούν, όπως και τα Active Appearance Models [5, 6], παραμετρικά μοντέλα για την μοντελοποίηση αντικειμένων. Χρησιμοποιούν ένα σύνολο πρωτότυπων εικόνων για να περιγράψουν το σχήμα και την υφή μιας συγκεκριμένης κατηγορίας αντικειμένων, όπως το ανθρώπινο πρόσωπο και τα μάτια. Για την προσαρμογή τους σε νέες εικόνες χρησιμοποιείται συνήθως κάποια στοχαστική μέθοδος καθοδικής κλίσης (gradient descent). Οι Rikert and Jones[31] χρησιμοποίησαν 2D Morphable Models για να μοντελοποιήσουν της περιοχής των ματιών διαφορετικών προσώπων και για διάφορες μεταβολές της πόζας κεφαλιού, καταγράφοντας παράλληλα και τις συντεταγμένες του σημείου εστίασης του gaze στην οθόνη ενός υπολογιστή. Το σύστημα που προκύπτει είναι ανεξάρτητο του χρήστη και της πόζας κεφαλιού. Στη συνέχεια προχωρούν στην αντιστοίχιση των παραμέτρων του μοντέλου με τις συντεταγμένες της οθόνης με τη χρήση ενός νευρωνικού δικτύου. Σαν είσοδος στο νευρωνικό δίκτυο βρέθηκε ότι είναι καλύτερο να χρησιμοποιηθεί το άθροισμα των παραμέτρων σχήματος και υφής της κάθε πρωτότυπης εικόνας. Τα αποτελέσματα που προκύπτουν κρίνονται αρκετά ικανοποιητικά με εξαίρεση κάποιες περιπτώσεις όπου τα μάτια είναι πολύ κλειστά ή υπάρχουν γυαλιά.

Άλλα είδη τεχνικών

Οι Zhu and Yang[38] χρησιμοποιούν μια μέθοδο υπολογισμού του gaze σε επίπεδο υποπίξελ subpixel για να αντιμετωπίσουν την περιορισμένη ακρίβεια που παρουσιάζουν οι τεχνικές που στηρίζονται στον προσδιορισμό χαρακτηριστικών στοιχείων του ματιού στην αρχικής ανάλυσης εικόνα. Αρχικά, έδειξαν ότι ακόμα και αν τα χαρακτηριστικά αυτά ανιχνευθούν με τη μέγιστη δυνατή ακρίβεια (ενός pixel) η ακρίβεια στον υπολογισμό του eye gaze προκύπτει περιορισμένη. Αρχικά προσδιορίζονται οι γωνίες των ματιών και οι ακμές της ίριδας με απλά γραμμικά φίλτρα στην αρχική ανάλυση. Έπειτα, γύρω από τις περιοχές αυτές γίνεται ενδοσυμπλήρωση των απαραίτητων τιμών έτσι ώστε να προκύψει ένα ακριβέστερο πλέγμα τιμών. Η όλη διαδικασία επαναλαμβάνεται μέχρις ότου επιτευχθεί η ζητούμενη ακρίβεια. Στη συνέχεια το κέντρο της ίριδας υπολογίζεται ταιριάζοντας μια έλλειψη στα σημεία των ακμών που υπολογίστηκαν. Τέλος, για την τελική εκτίμηση του gaze χρησιμοποιείται γραμμική παλινδρόμηση για τον συσχετισμό της διεύθυνσης του gaze με το διάνυσμα που ενώνει το κέντρο της ίριδας με τη γωνία του ματιού. Τα αποτελέσματα προκύπτουν αρκετά πιο ακριβή τόσο από εκείνα των μεθόδων σε επίπεδο pixel όσο και από άλλων τεχνικών σε επίπεδο υποπίξελ subpixel.

Οι Bebis and Fujimura[2] προτείνουν μια τεχνική εκτίμησης του eye gaze βασισμένη σε ανάλυση ιδιοχώρων. Σκοπός είναι η χρησιμοποίηση της πληροφορίας για την πρόβλεψη της συμπεριφοράς του οδηγού του αυτοκινήτου και την προειδοποίησή του σε ενδεχόμενες μη ασφαλείς καταστάσεις. Η ακρίβεια εκτίμησης του gaze επομένως δεν είναι τόσο πρωταρχικής σημασίας όσο η ταχύτητα απόκρισης του συστήματος. Για το λόγο αυτό το eye gaze κατηγοριοποιείται σε 5 βασικές



Σχήμα 2.4: Προσέγγιση ιδιοχώρων για την εκτίμηση του eye gaze. Έχει ληφθεί από το [2].

διευθύνσεις (ευθεία, πάνω, κάτω, δεξιά, αριστερά), κάθε μια από τις οποίες μοντελοποιείται με έναν διαφορετικό ιδιοχώρο. Μια νέα διεύθυνση του gaze κατηγοριοποιείται σε μία από τις παραπάνω κλάσεις υπολογίζοντας την απόσταση από κάθε ιδιοχώρο και επιλέγοντας αυτή με την ελάχιστη απόσταση. Για την δημιουργία του κάθε ιδιοχώρου χρησιμοποιείται ένα σύνολο εικόνων εκπαίδευσης από την περιοχή των ματιών καθώς και Ανάλυση Πρωτεύουσων Συνιστωστών (PCA), με στόχο να παραμείνει ένας μικρότερος αριθμός ιδιοδιανυσμάτων που να περιγράφει όμως το μεγαλύτερο μέρος της μεταβλητότητας. Επίσης, τόσο οι εικόνες εκπαίδευσης όσο και οι νέες εικόνες είναι απαραίτητο να είναι ευθυγραμμισμένες. Γι' αυτό, σε ένα στάδιο προεπεξεργασίας, εφαρμόζεται στην εικόνα ένας αφινικός μετασχηματισμός βασισμένος στις θέσεις των ματιών και της μύτης. Η όλη διαδικασία φαίνεται σχηματικά στο Σχήμα 2.4. Τα αποτελέσματα που προκύπτουν είναι πολύ ικανοποιητικά ακόμα και όταν υπάρχουν μεταβολές στην πόζα κεφαλιού. Ωστόσο, όταν το άτομο δεν βρίσκεται πολύ κοντά στην κάμερα εμφανίζονται κάποιες περιπτώσεις λανθασμένης εκτίμησης. Η συγκεκριμένη μέθοδος παρόλο που μοιάζει αρκετά με τις τεχνικές που χρησιμοποιούν AAM και Morphable Models δεν κατατάσσεται στην ίδια κατηγορία γιατί δεν μοντελοποιεί ξεχωριστά το σχήμα και την υφή.

Τέλος, οι von Agris et al.[35] προτείνουν μια μέθοδο εκτίμησης της γραμμής του βλέματος που στηρίζεται στη διαφορά φωτεινότητας μεταξύ ίριδας, δέρματος και κερατοειδούς. Αρχικά, γίνεται ο εντοπισμός της ίριδας χρησιμοποιώντας τον κυκλικό μετασχημασμό Hough. Στη συνέχεια γίνεται δειγματοληψία στα σημεία μεταξύ της ίριδας και ενός κύκλου διπλάσιας διαμέτρου με σκοπό την κατασκευή ιστογραμμάτων

φωτεινότητας, τα οποία θα διαφέρουν ανάλογα με την κατεύθυνση του βλέματος. Έπειτα γίνεται αντιστοίχιση μεταξύ κατανομής φωτεινότητας και γραμμή βλέματος εκπαιδευοντας έναν ταξινομητή μέγιστης πιθανοφάνειας.

2.3 Τεχνικές αναγνώρισης/κατηγοριοποίησης συναισθήματος

2.3.1 Γενικά

Η έκφραση του προσώπου μπορεί να προσεγγιστεί τόσο σαν ένας συνδυασμός βασικών δράσεων του προσώπου όσο και σαν την έκφραση ενός βασικού συναισθήματος. Τα συστήματα που πραγματοποιούν κατηγοριοποίηση συναισθήματος ανεξάρτητα με τις κατηγορίες και χαρακτηριστικά αναγνώρισης που χρησιμοποιούν μπορούν να χωριστούν σε τρεις βασικές κατηγορίες: κατηγοριοποίηση με βάση συγκεκριμένα πρότυπα, κατηγοριοποίηση με χρήση νευρωνικών δικτύων και κατηγοριοποίηση με βάση συγκεκριμένους κανόνες.[27] Συνήθως, οι κλάσεις που χρησιμοποιούνται περιλαμβάνουν τα 6 βασικά συναισθήματα (θυμός, αγδία, φόβος, ευτυχία, θλίψη, έκπληξη) και το ουδέτερο συναίσθημα.

Στην πρώτη κατηγορία η κατηγοριοποίηση γίνεται σε συγκεκριμένες διακριτές κλάσεις που έχουν οριστεί κατά την φάση της εκπαίδευσης. Η νέα εικόνα συγκρίνεται με ένα σύνολο προτύπων, κάθε ένα από τα οποία αντιπροσωπεύει ένα συγκεκριμένο συναίσθημα, και επιλέγεται σαν αποτέλεσμα η κλάση με το καλύτερο αποτέλεσμα ταιριάσματος. Στην κατηγορία αυτή δεν είναι δυνατή η κατηγοριοποίηση σε ένα ενδιάμεσο συναίσθημα που δεν υπάρχει στις διακριτές κλάσεις, και έτσι η αναγνώριση γίνεται μεταξύ πεπερασμένων κατηγοριών. Υπάρχουν διάφορες τεχνικές που χρησιμοποιούν το παραπάνω γενικό πλαίσιο. Αρχικά μπορούν να χρησιμοποιηθούν σαν είσοδο του ταξινομητή οι παράμετροι των AAMs, στις οποίες όμως εκτός από πληροφορία για το συναίσθημα περιέχεται και πληροφορία για την ταυτότητα του ατόμου, την πόζα κεφαλιού και άλλα. Γι' αυτό το λόγο συνήθως εφαρμόζεται σε ένα προηγούμενο στάδιο γραμμική ανάλυση διακριτών συνιστώσων (Linear Discriminant Analysis LDA) με σκοπό τον διαχωρισμό της μεταβλητότητας των παραμέτρων. Μια άλλη μέθοδος βασίζεται στο γεγονός ότι άνθρωποι που μοιάζουν μεταξύ τους θα έχουν και παρόμοιες εκφράσεις. Γι' αυτό το λόγο αρχικά βρίσκειται το άτομο του συνόλου εκπαίδευσης που ταιριάζει καλύτερα στο νέο πρόσωπο και μετά γίνεται η αναγνώριση του συναισθήματος. Άλλες τεχνικές χρησιμοποιούν σαν χαρακτηριστικά αναγνώρισης παραμέτρους που προκύπτουν από κάποια επεξεργασία της εικόνας, όπως μιγαδικό Gabor μετασχηματισμό, ακολουθούμενου συνήθως από ανάλυση PCA για την μείωση της διάστασης του προβλήματος. Σαν ταξινομητές μπορούν να χρησιμοποιηθούν ταξινομητές που να ελαχιστοποιούν κάποια απόσταση (Support Vector Machines SVM), ταξινομητές μέγιστης πιθανοφάνειας (GMMs) ή νευρωνικά

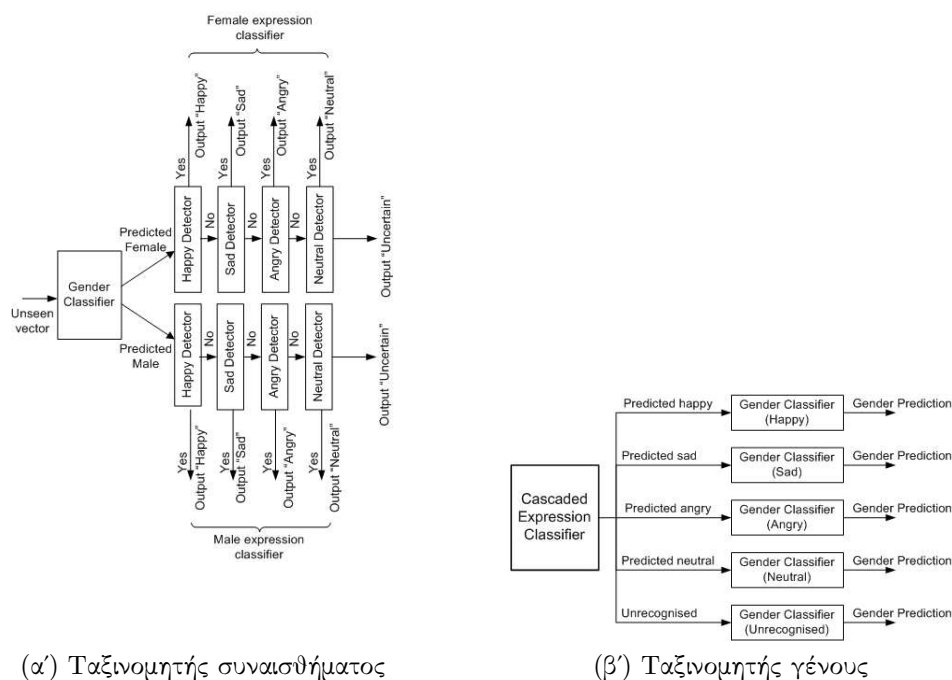
δίκτυα με διακριτή έξοδο. Τέλος, σε περιπτώσεις όπου η αναγνώριση πρέπει να γίνει σε ακολουθίες εικόνων χρησιμοποιούνται Κρυφά Μαρκοβιανά Μοντέλα για την μοντελοποίηση της κάθε κλάσης. Επίσης, αν η ακολουθία ξεκινά με το ουδέτερο συναίσθημα αυτό μπορεί να αφαιρεθεί με σκοπό την ελαχιστοποίηση της μεταβλητότητας των παραμέτρων αναγνώρισης λόγω του χρήσης.

Στις μεθόδους που χρησιμοποιούν νευρωνικά δίκτυα μπορεί να γίνει κατηγοριοποίηση και σε περισσότερες από μία κλάσεις ταυτόχρονα μιας και δεν γίνεται μοντελοποίηση της κάθε κατηγορίας με διαφορετικό ταξινομητή αλλά εκπαιδεύεται ένα νευρωνικό δίκτυο που δίνει σαν έξοδο την κλάση που ανήκει ένα συναίσθημα σαν μια συνεχή τιμή. Έτσι η κατηγοριοποίηση μπορεί να γίνει και μεταξύ πρωτότυπων συναισθημάτων, που να ανήκουν κάπου ενδιάμεσα σε σχέση με τα συναισθήματα που περιέχονται στο σύνολο εκπαίδευσης. Σαν είσοδο των νευρωνικών δικτύων μπορεί να χρησιμοποιηθούν παρόμοια χαρακτηριστικά με αυτά της προηγούμενης κατηγορίας μεθόδων. Επίσης, έχουν χρησιμοποιηθεί διάφορες τοπολογίες νευρωνικών δικτύων με σε γενικές γραμμές ικανοποιητικά αποτελέσματα.

Στις τεχνικές που βασίζονται στην εφαρμογή κανόνων η κατηγοριοποίηση γίνεται με βάση τις δράσεις του προσώπου που πραγματοποιούνται κατά την έκφραση ενός συναισθήματος. Αρχικά προδιορίζονται χαρακτηριστικά στοιχεία του προσώπου της εικόνας και υπολογίζεται η διαφορά μεταξύ αυτών και των ίδιων χαρακτηριστικών σε ένα πρόσωπο στο οποίο δεν υπάρχει κάποιο συναίσθημα. Στη συνέχεια χρησιμοποιώντας κανόνες του FACS (Facial Action Coding System) γίνεται κατηγοριοποίηση σε κλάσεις που αντιπροσωπεύουν δράσεις του προσώπου (Action Units AUs). Έπειτα η κατηγοριοποίηση του συναισθήματος στις 6 βασικές κατηγορίες πραγματοποιείται συγκρίνοντας την AUs περιγραφή της ζητούμενης έκφρασης του προσώπου με την AUs περιγραφή κάθε μίας από τις κλάσεις συναισθημάτων, χρησιμοποιώντας και πάλι συγκεκριμένους κανόνες.

2.3.2 Τεχνικές που βασίζονται στη χρήση AAMs

Οι Lucey et al.[18] εφαρμόζουν κατηγοριοποίηση του συναισθήματος από ακολουθίες εικόνων σε 7 βασικές κατηγορίες (θυμός, περιφρόνηση, αηδία, φόβος, ευτυχία, θλίψη, έκπληξη) χρησιμοποιώντας χαρακτηριστικά που προκύπτουν από την χρήση AAMs και επιλέγοντας γραμμικά Support Vector Machines (SVM) σαν ταξινομητές. Σαν παράμετροι αναγνώρισης χρησιμοποιούνται το κανονικοποιημένο σχήμα και η κανονικοποιημένη εμφάνιση. Το κανονικοποιημένο σχήμα προκύπτει από το μοντέλο σχήματος των AAMs χωρίς όμως να ληφθούν υπόψιν παράμετροι που σχετίζονται με τον γεωμετρικό μετασχηματισμό ομοιότητας. Επίσης, γίνεται κανονικοποίηση και ως προς το AU0 αφαιρώντας το σχήμα του πρώτου καρέ που αντιστοιχεί στο ουδέτερο συναίσθημα. Η κανονικοποιημένη εμφάνιση αναφέρεται στην εικόνα που προκύπτει αν κάθε μεταβολή του σχήματος κανονικοποιηθεί με βάση το μέσο σχήμα. Ορισμένα συναισθήματα όπως η αηδία, η έκπληξη και η ευτυχία μπορούν να αναγνωριστούν



(α) Ταξινομητής συναισθήματος

(β) Ταξινομητής γένους

Σχήμα 2.5: Ταξινομητές βασισμένοι σε AAMs και SVMs. Έχει ληφθεί από το [32].

με ακρίβεια και μόνο από τις παραμέτρους που σχετίζονται με το σχήμα, ενώ άλλα όπως ο θυμός, η θλίψη και ο φόβος αναγνωρίζονται καλύτερα από τα χαρακτηριστικά που αφορούν την υφή. Γι' αυτό κρίνεται καλύτερο να χρησιμοποιηθεί το συνολικό διάλυμα των χαρακτηριστικών των AAMs. Τέλος, το ουδέτερο συναίσθημα μπορεί να αναγνωριστεί εύκολα και με απόλυτη ακρίβεια χρησιμοποιώντας ένα απλό μέτρο που να εκφράζει την ενέργεια της μετατόπισης του σχήματος.

Οι Saatci and Town[32] προχωρούν στην κατηγοριοποίηση του συναισθήματος σε 4 βασικές κατηγορίες (ευτυχία, θυμός, θλίψη, ουδέτερο) καθώς του γένους του χρήστη χρησιμοποιώντας σαν είσοδο των ταξινομητών τις παραμέτρους που προκύπτουν από την εφαρμογή AAMs σε εικόνες προσώπων. Σαν ταξινομητές χρησιμοποιούνται Support Vector Machines, οι οποίοι τοποθετούνται σε αλυσιδωτή μορφή προκειμένου να βελτιστοποιηθεί το συνολικό ποσοστό ανγνώρισης. Επίσης, για το ίδιο σκοπό μπορεί να χρησιμοποιηθεί και ένας ταξινομητής γένους ακολουθούμενος από δύο ταξινομητές συναισθήματος, έναν για το αρσενικό και ένα για το θηλυκό. Από μια άλλη σκοπιά μπορούν να χρησιμοποιηθούν 4 ταξινομητές γένους που να ακολουθούν τον ταξινομητή συναισθήματος συνθέτοντας έτσι έναν ταξινομητή γένους βασισμένο στο συναίσθημα. Και οι δύο τοπολογίες ταξινομητών φαίνονται στο Σχήμα 2.5

Κεφάλαιο 3

Ενεργά Μοντέλα Εμφάνισης

3.1 Εισαγωγή

Τα Ενεργά Μοντέλα Εμφάνισης (Active Appearance Models) αρχικά προτάθηκαν από τους Cootes et al.[5, 6] και στη συνέχεια εξελίχθηκαν από τους Baker and Matthews[1, 19], οι οποίοι πρόσθεσαν γρήγορους και αποδοτικούς αλγόριθμους για την εφαρμογή των AAMs σε νέες εικόνες. Τα AAMs, όπως και τα Morphable Models[16], αποτελούν μη γραμμικά, γεννητικά, παραμετρικά μοντέλα για την περιγραφή μια συγκεκριμένης κατηγορίας αντικειμένων, όπως το ανθρώπινο πρόσωπο ή τα μάτια. Αποτελούν ένα συνδυασμό γραμμικών στατιστικών μοντέλων για το σχήμα και την υφή. Το συνολικό μοντέλο εμφάνισης που προκύπτει εκπαιδεύεται σε ένα σύνολο εικόνων έτσι ώστε το παραγόμενο στατιστικό μοντέλο να μπορεί να περιγράψει τη μεταβλητότητα του συνόλου εκπαίδευσης με ένα σχετικά μικρό πλήθος παραμέτρων που να ελέγχουν το σχήμα και την υφή. Στη συνέχεια τα AAMs μπορούν να χρησιμοποιηθούν για την δημιουργία νέων πρωτότυπων εικόνων μεταβάλλοντας τις παραμέτρους σχήματος και υφής ή να εφαρμοστούν σε άλλες εικόνες με στόχο την ανάλυση και την περιγραφή τους βάση των παραμέτρων των AAMs.

Τα AAMs αποτελούν το βασικό εργαλείο για την ανάπτυξη των διαφόρων μεθόδων υπολογισμού του eye gaze και κατηγοριοποίησης συναισθήματος, που παρουσιάζονται στην συγκεκριμένη διπλωματική. Κάποιες τεχνικές χρησιμοποιούν απευθείας τις παραμέτρους των AAMs, ενώ άλλες στηρίζονται στο αποτέλεσμα της προσαρμογής της μάσκας σχήματος προκειμένου να υπολογίσουν διάφορα χαρακτηριστικά του προσώπου. Επίσης, σε κάποιες μεθόδους τα AAMs μπορούν να χρησιμοποιηθούν σαν ένα αρχικό στάδιο προεπεξεργασίας πριν την εφαρμογή του κύριου αλγόριθμου. Από όλα αυτά κρίνεται απαραίτητο να γίνει μια συνοπτική αλλά περιεκτική αναφορά στα AAMs καθώς και στον αλγόριθμο προσαρμογής τους σε νέες εικόνες που χρησιμοποιείται στην παρούσα εργασία.

Αφού ολοκληρωθεί η θεωρητική περιγραφή των AAMs κατασκευάζονται συγκεκριμένα μοντέλα προκειμένου να χρησιμοποιηθούν ως βάση για την υλοποίηση των διαφόρων τεχνικών στα κεφάλαια που ακολουθούν. Πιο συγκεκριμένα αρχικά κατασκευάζονται μοντέλα του προσώπου προσαρμοσμένα στο χρήστη με στόχο την εκτίμηση της πόζας κεφαλιού και του eye gaze. Επιπρόσθετα, για τον υπολογισμό του gaze κατασκευάζεται και ένα τοπικό μοντέλο για την περιοχή των ματιών. Τέλος, δημιουργείται και ένα γενικευμένο μοντέλο του προσώπου, κατάλληλο για κάθε χρήστη, με σκοπό τη χρησιμοποίησή του στο πρόβλημα της κατηγοριοποίησης συναισθήματος.

3.2 Περιγραφή των AAMs

3.2.1 Μοντέλο Σχήματος

Στα ανεξάρτητα AAMs το σχήμα και η υφή μοντελοποιούνται ξεχωριστά. Το σχήμα των AAMs μπορεί να καθοριστεί από ένα γράφο και πιο συγκεκριμένα από τις θέσεις των κορυφών του γράφου. Άρα το σχήμα \mathbf{s} ενός AAMs ορίζεται σαν το διάνυσμα 2Δ συντεταγμένων των u κορυφών που σχηματίζουν το γράφο:

$$\mathbf{s} = (x_1, y_1, x_2, y_2, \dots, x_u, y_u)^T \quad (3.1)$$

Το μοντέλο σχήματος των AAMs χαρακτηρίζεται από γραμμική μεταβολή του σχήματος. Δηλαδή, το σχήμα \mathbf{s}_p μπορεί να εκφραστεί σαν ένα βασικό σχήμα \mathbf{s}_0 στο οποίο προστίθεται ένας γραμμικός συνδυασμός n διανυσμάτων σχήματος \mathbf{s}_i :

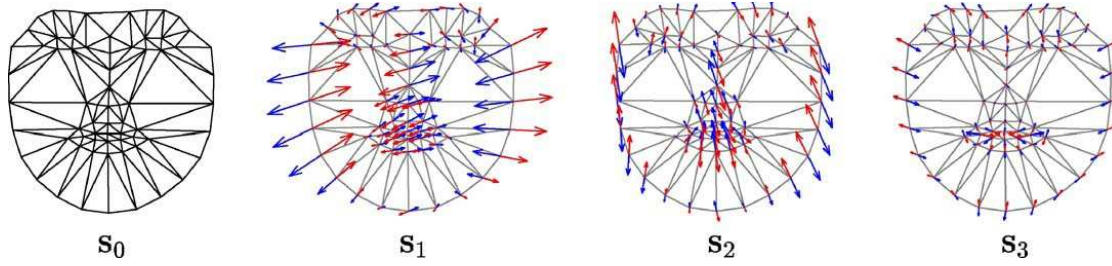
$$\mathbf{s}_p = \mathbf{s}_0 + \sum_{i=1}^n p_i \mathbf{s}_i \quad (3.2)$$

Οι συντελεστές p_i αποτελούν τις παραμέτρους σχήματος ενώ τα διανύσματα \mathbf{s}_i θεωρούνται ορθοκανονικά. Το μοντέλο σχήματος των AAMs κατασκευάζεται από ένα επισημασμένο σύνολο k εικόνων εκπαίδευσης, εφαρμόζοντας Principal Component Analysis (PCA) στους γράφους του σχήματος \mathbf{s}^j . Το σχήμα βάσης \mathbf{s}_0 είναι το μέσο σχήμα του συνόλου εκπαίδευσης:

$$\mathbf{s}_0 = \frac{1}{k} \sum_{j=1}^k \mathbf{s}^j \quad (3.3)$$

ενώ τα διανύσματα \mathbf{s}_i υπολογίζονται από τις ιδιοτιμές του πίνακα συμμεταβλητότητας:

$$S = \frac{1}{k-1} \sum_{j=1}^k (\mathbf{s}^j - \mathbf{s}_0)(\mathbf{s}^j - \mathbf{s}_0)^T \quad (3.4)$$



Σχήμα 3.1: Τριγωνοποιημένο σχήμα βάσης s_0 και 3 πρώτα διανύσματα σχήματος σχεδιασμένα πάνω στον γράφο του s_0 . Έχει ληφθεί από το [19].

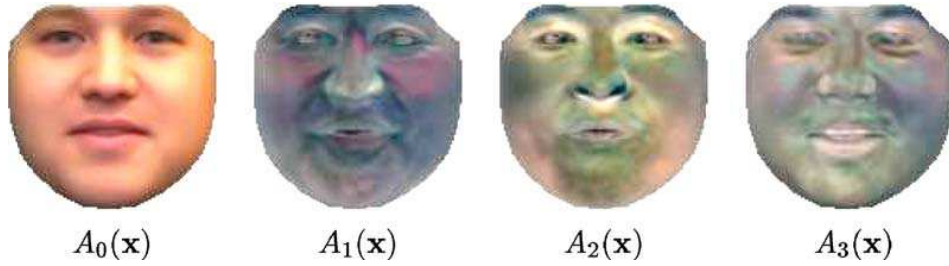
Συγκεκριμένα επιλέγονται τα n ιδιοδιανύσματα που αντιστοιχούν στις n μεγαλύτερες ιδιοτιμές του πίνακα S . Η κάθε ιδιοτιμή λ_i^{shape} αντιπροσωπεύει την μεταβλητότητα του συνόλου εκπαίδευσης στην κατεύθυνση του ιδιοδιανύσματος s_i . Μεταβάλλοντας τις παραμέτρους σχήματος p_i στο διάστημα $[-3\sqrt{\lambda_i^{shape}}, 3\sqrt{\lambda_i^{shape}}]$ προκύπτουν σχήματα ανάλογα με αυτά των εικόνων εκπαίδευσης. Ένα παράδειγμα μοντέλου σχήματος φαίνεται στο Σχήμα 3.1, στο οποίο έχουν σχεδιαστεί ο τριγωνοποιημένος γράφος του σχήματος βάσης καθώς και τα 3 πρώτα διανύσματα σχήματος πάνω στο αρχικό γράφο του s_0 .

Συνήθως, οι γράφοι του συνόλου εκπαίδευσης ευθυγραμμίζονται πρώτα χρησιμοποιώντας Προκρούστια ανάλυση πριν την εφαρμογή του PCA. Με το τρόπο αυτό αφαιρείται όλη η μεταβλητότητα που οφείλεται σε έναν καθολικό μετασχηματισμό σχήματος, όπως αλλαγή κλίμακας, μετατόπιση, καθολική περιστροφή γύρω από τον άξονα της κάμερας. Με τον τρόπο αυτό, κατά τη διαδικασία του PCA έχει παραμείνει μόνο η μεταβλητότητα που οφείλεται σε τοπική παραμόρφωση του σχήματος. Η παραμόρφωση αυτή μπορεί να περιγραφεί συναρτήσει των παραμέτρων p_i του μοντέλου από έναν μετασχηματισμό $\mathbf{W}(\mathbf{x}; \mathbf{p})$, ο οποίος αντιστοιχίζει κάθε κορυφή του σχήματος βάσης s_0 στην ανάλογη κορυφή του σχήματος s_p . Ο μετασχηματισμός που χρησιμοποιείται συνήθως είναι ένας τμηματικά αφινικός μετασχηματισμός (piecewise affine transform), βασισμένος στον τριγωνοποιημένο γράφο ή μία thin-plate spline αντιστοίχιση.

Για να συμπεριληφθεί στο μοντέλο του σχήματος η συνολική μεταβλητότητα του σχήματος χρησιμοποιείται ένας επιπλέον Μετασχηματισμός Ομοιότητας 4 βαθμών ελευθερίας:

$$\mathbf{S}_t(\mathbf{x}) = \begin{pmatrix} 1+t_1 & -t_2 \\ t_2 & 1+t_1 \end{pmatrix} \mathbf{x} + \begin{pmatrix} t_3 \\ t_4 \end{pmatrix} \quad (3.5)$$

Το εκτεταμένο διάνυσμα των παραμέτρων σχήματος $\tilde{\mathbf{p}} = [\mathbf{t}_{1:4}, \mathbf{p}_{1:n}]^T$ έχει μήκος $n+4$ και ορίζει τον συνδυασμένο μετασχηματισμό για την ολική παραμόρφωση $\mathbf{W}(\mathbf{x}; \tilde{\mathbf{p}}) = \mathbf{S}_t(\mathbf{W}(\mathbf{x}; \mathbf{p}))$, ο οποίος αντιστοιχίζει κάθε σημείο \mathbf{x} του προτύπου του μοντέλου στο αντίστοιχο σημείο της εικόνας βάσει των παραμέτρων $\tilde{\mathbf{p}}$ του μοντέλου.



Σχήμα 3.2: Εικόνα της υφής βάσης A_0 και 3 πρώτες ιδιοεικόνες A_i , ορισμένες εντός του γράφου βάσης \mathbf{s}_0 . Έχει ληφθεί από το [19].

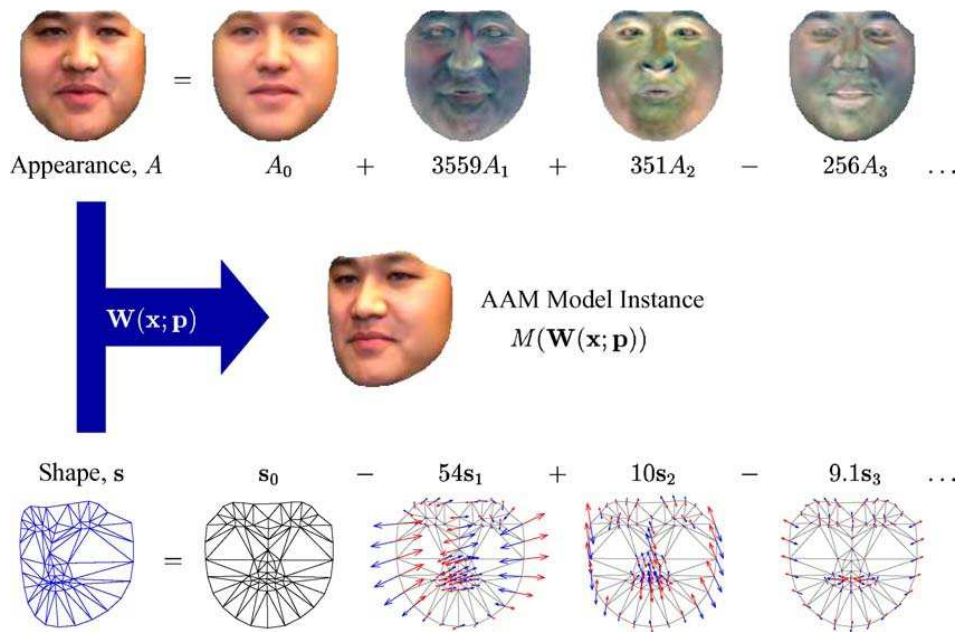
3.2.2 Μοντέλο Υφής

Η υφή αποτελεί το τμήμα εκείνο της εμφάνισης που σχετίζεται με τη φωτεινότητα ή το χρώμα ενός αντικειμένου, αφού πρώτα έχει γίνει κανονικοποίηση ως προς το σχήμα. Η υφή ενός ανεξάρτητου AAM ορίζεται μέσω του γράφου βάσης \mathbf{s}_0 . Αν θεωρηθεί ότι το σχήμα \mathbf{s}_0 εμπεριέχει και όλα τα pixels $\mathbf{x} = (x, y)^T$ που βρίσκονται εντός του γράφου του \mathbf{s}_0 , τότε η υφή ενός AAM αποτελεί μια εικόνα $A(\mathbf{x})$ ορισμένη στα pixels για τα οποία ισχύει $\mathbf{x} \in \mathbf{s}_0$. Το μοντέλο υφής των AAMs, σε αντιστοιχία με το μοντέλο σχήματος, χαρακτηρίζεται από γραμμική μεταβολή της υφής. Δηλαδή, η υφή $A_\lambda(\mathbf{x})$ μπορεί να εκφραστεί σαν μια βασική εικόνα υφής $A_0(\mathbf{x})$ στην οποία προστίθεται ένας γραμμικός συνδυασμός m εικόνων υφής $A_i(\mathbf{x})$:

$$A_\lambda(\mathbf{x}) = A_0(\mathbf{x}) + \sum_{i=1}^m \lambda_i A_i(\mathbf{x}) \quad \forall \mathbf{x} \in \mathbf{s}_0 \quad (3.6)$$

Οι συντελεστές λ_i αποτελούν τις παραμέτρους υφής ενώ οι εικόνες $A_i(\mathbf{x})$ θεωρούνται ορθοκανονικές. Το μοντέλο υφής των AAMs κατασκευάζεται εφαρμόζοντας PCA, με όμοιο τρόπο όπως και στο μοντέλο σχήματος, σε ένα σύνολο από k σχηματικά κανονικοποιημένων εικόνων εκπαίδευσης. Η κανονικοποίηση ως προς το σχήμα πραγματοποιείται προβάλλοντας τα pixels που βρίσκονται εντός του επισημασμένου γράφου μια εικόνας εκπαίδευσης στις αντίστοιχες θέσεις του γράφου βάσης \mathbf{s}_0 . Για τον σκοπό αυτό για κάθε εικόνα εκπαίδευσης I^j χρησιμοποιείται ο μετασχηματισμός $A^j(\mathbf{x}) = I^j(\mathbf{W}(\mathbf{x}; \tilde{\mathbf{p}}))$, όπως αυτός περιγράφηκε στην προηγούμενη ενότητα.

Έτσι, η υφή βάσης A_0 είναι η μέση κανονικοποιημένη εικόνα, ενώ οι εικόνες A_i αποτελούν τις m ιδιοεικόνες (eigenimages) που αντιστοιχούν στις m μεγαλύτερες ιδιοτιμές λ_i^{text} . Μεταβάλλοντας τις παραμέτρους υφής λ_i στο διάστημα $[-3\sqrt{\lambda_i^{text}}, 3\sqrt{\lambda_i^{text}}]$ προκύπτουν εικόνες υφής ανάλογες με αυτές των εικόνων εκπαίδευσης. Ένα παράδειγμα μοντέλου υφής φαίνεται στο Σχήμα 3.2, στο οποίο παρουσιάζεται η εικόνα της υφής βάσης A_0 καθώς και οι 3 πρώτες ιδιοεικόνες A_i , ορισμένες εντός του γράφου βάσης \mathbf{s}_0 .



Σχήμα 3.3: Διαδικασία δημιουργίας πρωτότυπης εικόνας μεταβάλλοντας τις παραμέτρους σχήματος και υφής του AAM. Έχει ληφθεί από το [19].

Συνήθως, προκειμένου να αφαιρεθεί η επίδραση των καθολικών συνθηκών φωτισμού, γίνεται φωτομετρική ευθυγράμμιση των σχηματικά κανονικοποιημένων εικόνων υφής πριν την εφαρμογή του PCA για το μοντέλο της υφής. Με τον τρόπο αυτό ορίζεται ένας καθολικός αφινικός μετασχηματισμός υφής που περιγράφει πολλαπλασιαστικές u_1 και προσθετικές u_2 αλλαγές στα επίπεδα φωτεινότητας:

$$T_{\mathbf{u}}(I) = (u_1 + 1)I + u_2 \quad (3.7)$$

Οι συνολικές παράμετροι που ελέγχουν την υφή μπορούν να συμπεριληφθούν σε ένα εκτεταμένο διάνυσμα παραμέτρων υφής: $\boldsymbol{\lambda} = [\mathbf{u}_{1:2}, \boldsymbol{\lambda}_{1:m}]^T$ έχει μήκος $m+2$.

3.2.3 Δημιουργία πρωτότυπων εικόνων με χρήση AAMs

Η μεταβολή στο σχήμα και την υφή ενός AAM περιγράφονται από τις εξισώσεις (3.2) και (3.6) αντίστοιχα. Ωστόσο, τα μοντέλα αυτά είναι ξεχωριστά μεταξύ τους και έτσι δεν είναι σαφές το πως μπορούν να συνδυαστούν έτσι ώστε να προκύψει ένα νέο στιγμιότυπο του αντικειμένου που περιγράφεται από το συγκεκριμένο AAM. Στο Σχήμα 3.3 παρουσιάζεται η διαδικασία δημιουργίας μιας πρωτότυπης εικόνας μεταβάλλοντας τις παραμέτρους σχήματος και υφής του AAM.

Μεταβάλλοντας τις παραμέτρους σχήματος $\mathbf{p} = (p_1, p_2, \dots, p_n)^T$ και μέσω της εξίσωσης (3.2) μπορούμε να δημιουργήσουμε ένα νέο σχήμα \mathbf{s}_p του AAM. Με ανάλογο τρόπο, μεταβάλλοντας τις παραμέτρους υφής $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_m)^T$, μπορούμε να δημιουργήσουμε μια νέα εικόνα υφής $A_{\boldsymbol{\lambda}}(\mathbf{x})$ ορισμένη στο εσωτερικό του γράφου του σχήματος βάσης \mathbf{s}_0 . Στη συνέχεια το νέο στιγμιότυπο του AAM, με παραμέτρους σχήματος και υφής \mathbf{p} και $\boldsymbol{\lambda}$ αντίστοιχα, δημιουργείται αντιστοιχίζοντας την υφή $A_{\boldsymbol{\lambda}}(\mathbf{x})$ από το γράφο του σχήματος βάσης \mathbf{s}_0 στο νέο σχήμα \mathbf{s}_p . Για τον σκοπό αυτό χρησιμοποιείται ο τμηματικά αφινικός μετασχηματισμός $\mathbf{W}(\mathbf{x}; \tilde{\mathbf{p}})$, που αναφέρθηκε σε προηγούμενη ενότητα. Η νέα εικόνα M που προκύπτει από το AAM περιγράφεται από την παρακάτω εξίσωση:

$$M(\mathbf{W}(\mathbf{x}; \mathbf{p})) = A_{\boldsymbol{\lambda}}(\mathbf{x}) \quad (3.8)$$

3.2.4 Εφαρμογή των AAMs σε νέες εικόνες

Ένα άλλο, επίσης σημαντικό θέμα είναι η εφαρμογή ενός AAM σε νέες εικόνες, δηλαδή η εύρεση των βέλτιστων παραμέτρων σχήματος \mathbf{p} και υφής $\boldsymbol{\lambda}$ που ελαχιστοποιούν τη διαφορά μεταξύ της νέας εικόνας $I(\mathbf{x})$ και της ανακατασκευασμένης εικόνας $M(\mathbf{W}(\mathbf{x}; \mathbf{p})) = A_{\boldsymbol{\lambda}}(\mathbf{x})$ βάσει του μοντέλου. Υπάρχουν δύο συστήματα αναφοράς στα οποία μπορεί να εκφραστεί η διαφορά αυτή, το πλαίσιο της εικόνας εισόδου I και το πλαίσιο αναφοράς του AAM. Για λόγους ευκολίας επιλέγεται το σύστημα συντεταγμένων του AAM, δηλαδή το σχήμα βάσης \mathbf{s}_0 . Στην περίπτωση αυτή η εικόνα λάθους θα δίνεται από τη σχέση:

$$\begin{aligned} E(\mathbf{x}) &= I(\mathbf{W}(\mathbf{x}; \mathbf{p})) - A_{\boldsymbol{\lambda}}(\mathbf{x}) \\ &= I(\mathbf{W}(\mathbf{x}; \mathbf{p})) - (A_0(\mathbf{x}) + \sum_{i=1}^m \lambda_i A_i(\mathbf{x})) \end{aligned} \quad (3.9)$$

Αν προσθέσουμε στην παραπάνω εξίσωση τον μετασχηματισμό ομοιότητας για το σχήμα και τον αφινικό μετασχηματισμό υφής τότε το συνδυασμένο διάνυσμα των προς βελτιστοποίηση παραμέτρων είναι $\mathbf{q} = [\tilde{\mathbf{p}}^T, \tilde{\boldsymbol{\lambda}}^T]^T$ και η εικόνα λάθους δίνεται από τη σχέση:

$$E(\mathbf{x}; \mathbf{q}) = T_{\mathbf{u}}(I(\mathbf{W}(\mathbf{x}; \tilde{\mathbf{p}}))) - (A_0(\mathbf{x}) + \sum_{i=1}^m \lambda_i A_i(\mathbf{x})) \quad (3.10)$$

Συνήθως, σαν ποσότητα προς ελαχιστοποίηση επιλέγεται η Ευκλείδεια νόρμα (άθροισμα τετραγώνων των διαφορών) $\|E(\mathbf{q})\|_2^2$ της εικόνας λάθους, ενώ άλλες πιο εύρωστες νόρμες μπορούν να χρησιμοποιηθούν στις περιπτώσεις που υπάρχει επικάλυψη του προσώπου από τα χέρια.[10] Το πρόβλημα βελτιστοποίησης που προκύπτει είναι γραμμικό ως προς τις παραμέτρους υφής και μη γραμμικό για τις

παραμέτρους σχήματος. Για την επίλυση του αναζητούμε ένα διάνυσμα ανανέωσης των παραμέτρων $d\mathbf{q} = [d\tilde{\mathbf{p}}^T, d\tilde{\boldsymbol{\lambda}}^T]^T$ που να ελαχιστοποιεί την νόρμα $\|E(\mathbf{q}, d\mathbf{q})\|_2^2$, θεωρώντας γνωστές τις παραμέτρους \mathbf{q} . Οι Baker and Matthews[1, 19] πρότειναν την ανάστροφη συνθετική (inverse compositional) τεχνική ανανέωσης των παραμέτρων σχήματος:

$$\mathbf{W}(\mathbf{x}; \tilde{\mathbf{p}}) \leftarrow \mathbf{W}(\mathbf{x}; \tilde{\mathbf{p}}) \circ \mathbf{W}^{-1}(\mathbf{x}; d\tilde{\mathbf{p}}) \equiv \mathbf{W}(\mathbf{W}^{-1}(\mathbf{x}; d\tilde{\mathbf{p}}); \tilde{\mathbf{p}}) \quad (3.11)$$

σε αντίθεση με την απλή προσθετική ανανέωση $\tilde{\mathbf{p}} \leftarrow \tilde{\mathbf{p}} + d\tilde{\mathbf{p}}$. Στην περίπτωση αυτή η εικόνα λάθους γράφεται στη μορφή:

$$E(\mathbf{x}; \mathbf{q}, d\mathbf{q}) = T_{\mathbf{u}}(I(\mathbf{W}(\mathbf{x}; \tilde{\mathbf{p}}))) - T_{d\mathbf{u}}(A_{\boldsymbol{\lambda}+d\boldsymbol{\lambda}}(\mathbf{W}(\mathbf{x}; d\tilde{\mathbf{p}}))) \quad (3.12)$$

Στη συνέχεια εφαρμόζοντας επέκταση Taylor πρώτης τάξης γύρω από το μηδέν έχουμε σύμφωνα με τη μέθοδο των Papandreou and Maragos[26]:

$$E(\mathbf{x}; \mathbf{q}, d\mathbf{q}) = E(\mathbf{x}; \mathbf{q}) - \left. \frac{\partial T(A_{\boldsymbol{\lambda}}(\mathbf{x}))}{\partial \mathbf{u}} d\mathbf{u} - \frac{\partial A_{\boldsymbol{\lambda}}(\mathbf{x})}{\partial \mathbf{x}} \frac{\partial \mathbf{W}(\mathbf{x})}{\partial \tilde{\mathbf{p}}} \right|_{\tilde{\mathbf{p}}=\mathbf{0}} d\tilde{\mathbf{p}} - [A_1(\mathbf{x}) \dots A_m(\mathbf{x})] d\boldsymbol{\lambda} \quad (3.13)$$

Οι εικόνες με τις οποίες πολλαπλασιάζονται οι ανανεώσεις των παραμέτρων αποκαλούνται εικόνες απότομης καθόδου (steepest descent) και εκφράζουν την αλλαγή που προκύπτει στην υφή ανανεώνοντας την κάθε κατηγορία παραμέτρων. Αν θέσουμε $d\boldsymbol{\lambda}' = d\boldsymbol{\lambda} - u_1 d\boldsymbol{\lambda}$ τότε στο νέο διάνυσμα παραμέτρων υφής $d\tilde{\boldsymbol{\lambda}}' = [d\mathbf{u}_{1:2}, d\boldsymbol{\lambda}'_{1:m}]^T$ αντιστοιχούν steepest descent εικόνες $A(\mathbf{x}) = [A_0(\mathbf{x}) 1 A_1 \dots A_m(\mathbf{x})]$. Αντίστοιχα αν θεωρήσουμε $M_i(\mathbf{x}) = \left. \frac{\partial A_i(\mathbf{x})}{\partial \mathbf{x}} \frac{\partial \mathbf{W}(\mathbf{x})}{\partial \tilde{\mathbf{p}}} \right|_{\tilde{\mathbf{p}}=\mathbf{0}}$ τότε για τις παραμέτρους σχήματος θα προκύψουν τα πρότυπα κίνησης (motion templates):

$$M_{\boldsymbol{\lambda}} = M_0 + \sum_{i=1}^m \lambda_i M_i \quad (3.14)$$

Έτσι το προς ελαχιστοποίηση κριτήριο παίρνει τη μορφή $\min \|E(\mathbf{q}) - H_{\boldsymbol{\lambda}} d\mathbf{q}'\|_2^2$, όπου $E(\mathbf{q})$ η τρέχουσα εικόνα λάθους, $H_{\boldsymbol{\lambda}} = [M_{\boldsymbol{\lambda}} \quad A]$ ο συνολικός πίνακας με τις steepest descent εικόνες και $d\mathbf{q}' = [d\tilde{\mathbf{p}}^T, d\tilde{\boldsymbol{\lambda}}'^T]^T$ το νέο διάνυσμα παραμέτρων. Η εκτίμηση ελαχίστων τετραγώνων για τις παραμέτρους δίνεται από τις σχέσεις:

$$d\mathbf{q}' = \Sigma_{\mathbf{q}} H_{\boldsymbol{\lambda}}^T E(\mathbf{q}), \quad \text{όπου} \quad (3.15)$$

$$\Sigma_{\mathbf{q}} = (H_{\boldsymbol{\lambda}}^T H_{\boldsymbol{\lambda}})^{-1} = \begin{bmatrix} M_{\boldsymbol{\lambda}}^T M_{\boldsymbol{\lambda}} & M_{\boldsymbol{\lambda}}^T A \\ A^T M_{\boldsymbol{\lambda}} & A^T A \end{bmatrix}^{-1} \quad (3.16)$$

Χρησιμοποιώντας τη σχέση $d\boldsymbol{\lambda} = d\boldsymbol{\lambda}' + u_1 d\boldsymbol{\lambda}$ για τις παραμέτρους υφής παίρνουμε το αρχικό διάνυσμα ανανέωσης παραμέτρων $d\mathbf{q} = [d\tilde{\mathbf{p}}^T, d\tilde{\boldsymbol{\lambda}}'^T]^T$.

Στη συνέχεια, οι m παράμετροι υφής λ ανανεώνονται προσθετικά $\lambda \leftarrow \lambda + d\lambda$, ενώ οι 2 παράμετροι \mathbf{u} του αφινικού μετασχηματισμού υφής ανανεώνονται σύμφωνα με την ανάστροφη συνθετική τεχνική $T_{\mathbf{u}} \leftarrow T_{\mathbf{u}} \circ T_{d\mathbf{u}}^{-1}$. Τέλος, οι $4 + n$ παράμετροι σχήματος $\bar{\mathbf{p}}$ ανανεώνονται, όπως ήδη αναφέρθηκε, με την inverse compositional τεχνική βάσει της σχέσης (3.11). Ο αλγόριθμος που προκύπτει από την παραπάνω διαδικασία είναι αρκετά αποδοτικός και ακριβής, αφού αρκετοί από τους παραπάνω υπολογισμούς μπορούν να γίνουν σε ένα στάδιο προεπεξεργασίας.

Επεκτάσεις του inverse compositional αλγορίθμου

Οι Papandreou and Maragos[26] πρόσθεσαν κάποιες επεκτάσεις στον inverse compositional αλγόριθμο εφαρμογής των AAM που αρχικά προτάθηκε από τους Baker and Matthews[1, 19]. Τα πρότυπα κίνησης (motion templates), όπως υπολογίστηκαν παραπάνω, περιέχουν όλη την πληροφορία της υφής οδηγώντας έτσι στον ταυτόχρονο (simultaneous) inverse compositional αλγόριθμο[11], ο οποίος είναι αρκετά ακριβής αλλά όχι τόσο γρήγορος. Αντίθετα, αν στα motion templates ληφθεί υπόψιν μόνο η μέση υφή ($M_{\lambda} = M_0$) τότε προκύπτει ο project-out inverse compositional αλγόριθμος[19], ο οποίος είναι εξαιρετικά γρήγορος αλλά έχει περιορισμένα αποτελέσματα σε περιπτώσεις με μεγάλη μεταβλητότητα στην υφή, όπως στα γενικά μοντέλα για το πρόσωπο.[11] Έτσι, οι Papandreou and Maragos[26] πρότειναν τον r τάξης inverse compositional αλγόριθμο, ο οποίος βρίσκεται κάπου ανάμεσα στους δύο παραπάνω αλγορίθμους. Σύμφωνα, με τον αλγόριθμο αυτό μπορούμε να συμπεριλάβουμε στα motion templates πληροφορία μόνο από τις r πρώτες ιδιοεικόνες υφής μεταβάλλοντας το $0 \leq r \leq m$:

$$M_{\lambda} \approx M_0 + \sum_{i=1}^k \lambda_i M_i, \quad 0 \leq r \leq m \quad (3.17)$$

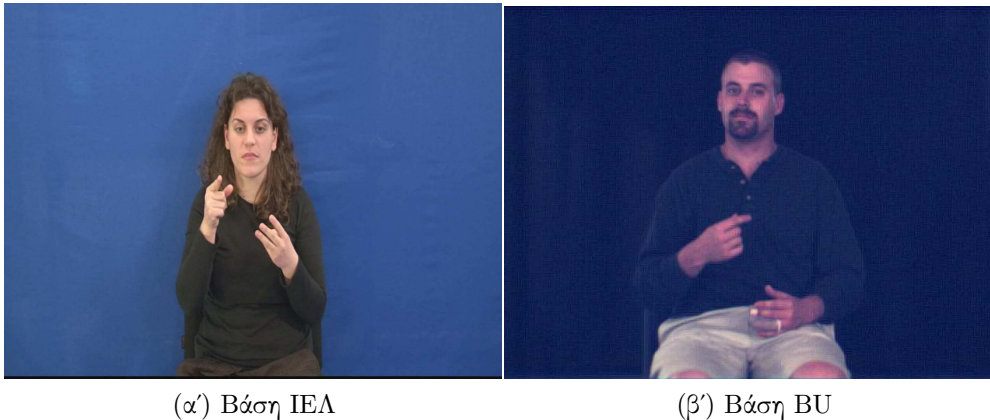
Με αυτο τον τρόπο μπορεί να προκύψει ένας γρήγορος όσο και ακριβής αλγόριθμος ανάλογα με τις ανάγκες που έχουμε κάθε φορά. Επίσης, παρατηρούμε ότι για $r = 0$ καταλήγουμε στον project-out αλγόριθμο[19], ενώ για $r = m$ έχουμε τον simultaneous αλγόριθμο.[11]

Επιπλέον, για να αυξήσουν την απόδοση του r τάξης αλγορίθμου, κυρίως σε ακολουθίες εικόνων από βίντεο, πρότειναν την περιοδική ανανέωση της βασικής εικόνας υφής A_0 με χρήση των ιδιοεικόνων υφής A_i που δεν συμπεριλήφθηκαν στον υπολογισμό των motion templates ($i > r$):

$$A'_0 = A_0 + \sum_{i=r+1}^m \lambda_i A_i \quad (3.18)$$

Περισσότερες και πιο αναλυτικές πληροφορίες σχετικά με τους αλγορίθμους εφαρμογής των AAM σε νέες εικόνες μπορεί κάποιος να βρει στα papers [1, 19, 11, 26].

3.3 AAMs προσαρμοσμένα στο χρήστη



Σχήμα 3.4: Παραδείγματα εικόνων νοηματικής γλώσσας από τις βάσεις δεδομένων της ENΓ του IEL και της BU400.

3.3.1 Κατασκευή των AAMs

Πριν προχωρήσουμε στην εφαρμογή των διαφόρων μεθόδων υπολογισμού του eye gaze είναι απαραίτητο να έχει προηγηθεί η οπτική μοντελοποίηση του ανθρώπινου προσώπου με χρήση των AAMs. Για το σκοπό αυτό αναζητούμε ένα σύνολο από εικόνες ή ακολουθίες εικόνων που να περιέχουν αρκετές μεταβολές τόσο στην κίνηση των ματιών όσο και στην πόζα κεφαλιού, οι οποίες να είναι και όσο το δυνατόν πιο φυσικές.

Τα βίντεο της νοηματικής γλώσσας (Σχήμα 3.4) αποτελούν μια τέτοια περίπτωση. Στη νοηματική γλώσσα η κίνηση του κεφαλιού και των ματιών παίζει πολύ σημαντικό ρόλο, αφού κάποια νοήματα ενώ μπορεί να μοιάζουν ίδια ως προς την κίνηση των χεριών διαφέρουν όμως λόγω της διαφορετικής μεταβολής στην πόζα κεφαλιού και στην κατεύθυνση του βλέματος. Επίσης, τα μάτια, αποτελούν ένα από τα κύρια μέσα για την αλλαγή στην προσωδία του νοηματικού λόγου καθώς και για να απευθύνουμε το λόγο σε κάποιο άλλο άτομο.

Υπάρχουν, όμως, και κάποιες τεχνικές δυσκολίες στη χρήση των βίντεο της νοηματικής γλώσσας. Αρχικά, στα περισσότερα βίντεο των βάσεων δεδομένων που υπάρχουν δεν υπάρχει επαρκής επισύμανση της πόζας κεφαλιού και του gaze για να ελέγξουμε την ορθότητα των αποτελεσμάτων που προκύπτουν από την εφαρμογή των διαφόρων μεθόδων. Επίσης, το σύστημα που θα σχεδιαστεί θα πρέπει να μην απαιτεί κάποια συγκεκριμένη βαθμονόμηση, μιας και τα βίντεο έχουν ήδη καταγραφεί σε κάποια χρονική στιγμή παρελθόν και δεν είναι δυνατόν να αντλήσουμε κάποια

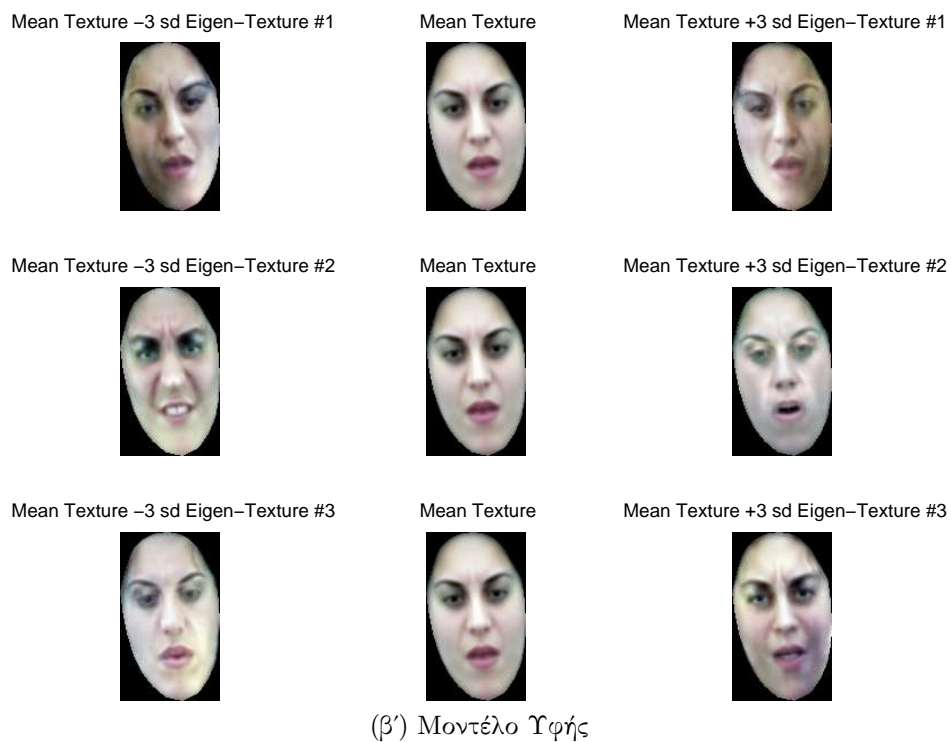
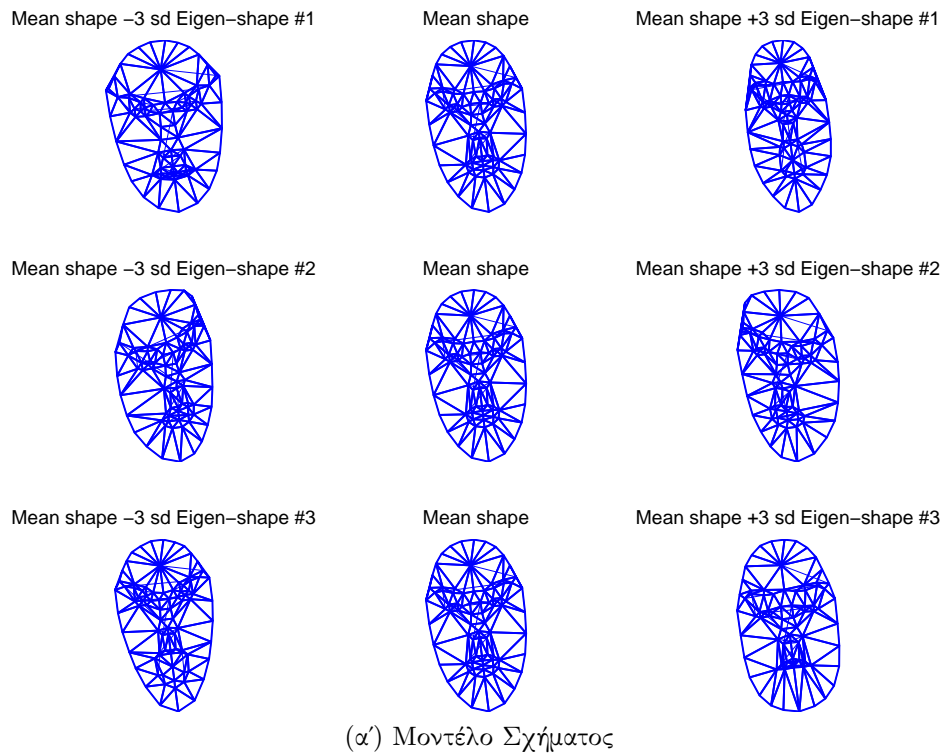
τέτοια πληροφορία. Τέλος, επειδή στη νοηματική γλώσσα η κίνηση των χεριών παίζει των κυρίαρχο ρόλο, ο νοηματιστής έχει βιντεοσκοπηθεί από τη μέση και πάνω. Αυτό έχει σαν αποτέλεσμα η ανάλυση της εικόνας στην περιοχή του προσώπου και πολύ περισσότερο κοντά στα μάτια να είναι περιορισμένη, ακόμα και στις περιπτώσεις που η αρχική ανάλυση του βίντεο είναι αρκετά καλή. Στην συγκεκριμένη εργασία θα χρησιμοποιήσουμε 2 βίντεο νοηματικής γλώσσας, ένα από τη βάση δεδομένων της Ελληνικής Νοηματικής Γλώσσας “ENΓ” του Ινστιτούτου Επεξεργασίας Λόγου “IEΛ-ILSP” (Σχήμα 3.4α’) και ένα από τη βάση του πανεπιστημίου της Βοστώνης “BU400” (Σχήμα 3.4β’).

Για κάθε ένα από τα βίντεο αυτά θα κατασκευάσουμε ένα διαφορετικό AAM, που να μοντελοποιεί όσο το δυνατόν περισσότερη από τη μεταβλητότητα του προσώπου και ιδιαίτερα των ματιών. Για να εκπαιδύσουμε τα AAMs θα χρειαστεί να επισημάνουμε χειροκίνητα με τα σημεία της μάσκας του προσώπου ένα σύνολο εικόνων. Χρησιμοποιούμε 86 σημεία για το γράφο του σχήματος προσώπου, μέσα στα οποία συμπεριλαμβάνονται και σημεία για το μέτωπο προκειμένου να αυξήσουμε την ακρίβεια του μοντέλου.

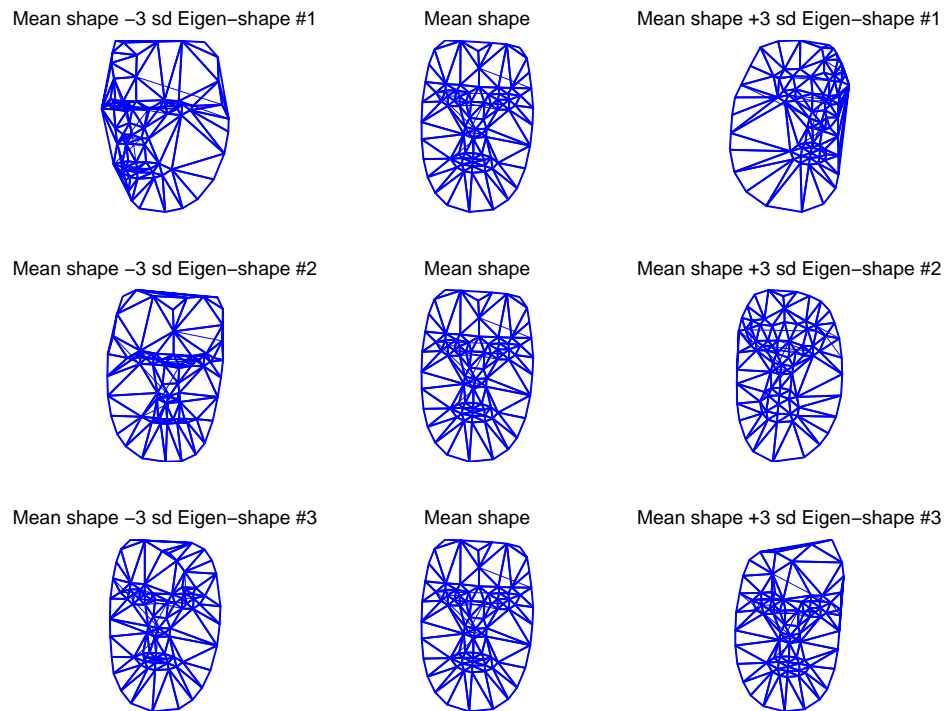
Αρχικά, ξεκινάμε την εκπαίδευση με ένα μικρό σύνολο εικόνων και συνεχίζουμε εφαρμόζοντας έναν Bootstrapping αλγόριθμο.[36] Αφού δημιουργήσουμε ένα AAM από το αρχικό σύνολο εκπαίδευσης, προχωράμε στην εφαρμογή του σε ένα μεγαλύτερο σύνολο εικόνων. Σε όσες από αυτές το αποτέλεσμα της εφαρμογής είναι καλό συμπεριλαμβάνονται στο σύνολο εκπαίδευσης προκειμένου να δημιουργηθεί ένα πιο αναλυτικό AAM. Η διαδικασία αυτή συνεχίζεται μέχρι να προκύψει ένα AAM που να ικανοποιεί τις προδιαγραφές μας.

Τόσο για την εκπαίδευση όσο και για την εφαρμογή των AAMs χρησιμοποιήθηκε ο ανοικτός κώδικας AAMtools του George Papandreou. Σαν ποσοστά της μεταβλητότητας σχήματος και υψής που επιλέγουμε να κρατήσουμε μετά από την εφαρμογή των αντίστοιχων PCA επιλέγουμε 95% και 85% αντίστοιχα. Οι 3 πρώτες συνιστώσες του μοντέλου σχήματος και υψής για τη βάση της ENΓ του IEΛ και για μεταβολές των αντίστοιχων συντελεστών στο διάστημα $[-3\sqrt{\lambda_i}, 3\sqrt{\lambda_i}]$ παρουσιάζονται στο Σχήμα 3.5.

Παρατηρώντας το μοντέλο του σχήματος διαπιστώνουμε τη σημαντική μεταβολή στη μάσκα σχήματος του προσώπου. Η μεταβολή αυτή σχετίζεται με διάφορα στοιχεία του προσώπου, όπως το άνοιγμα και κλείσιμο του στόματος και των ματιών (1^η και 3^η συνιστώσα), αλλά και με την πόζα του κεφαλιού όπως φαίνεται στην 2^η συνιστώσα. Όσο αφορά το μοντέλο υψής, αυτό δεν παρουσιάζει μεγάλες μεταβολές μιας και χρησιμοποιούμε μοντέλα προσαρμοσμένα στον κάθε χρήστη και άρα η μεταβλητότητα στην υφή του προσώπου είναι περιορισμένη. Ωστόσο, το μοντέλο υψής εκφράζει αλλαγές σε συγκεκριμένα σημεία του προσώπου, όπως τα μάτια και το στόμα, ενώ είναι απαραίτητο και για τη καλύτερη εφαρμογή του AAM σε νέες εικόνες.



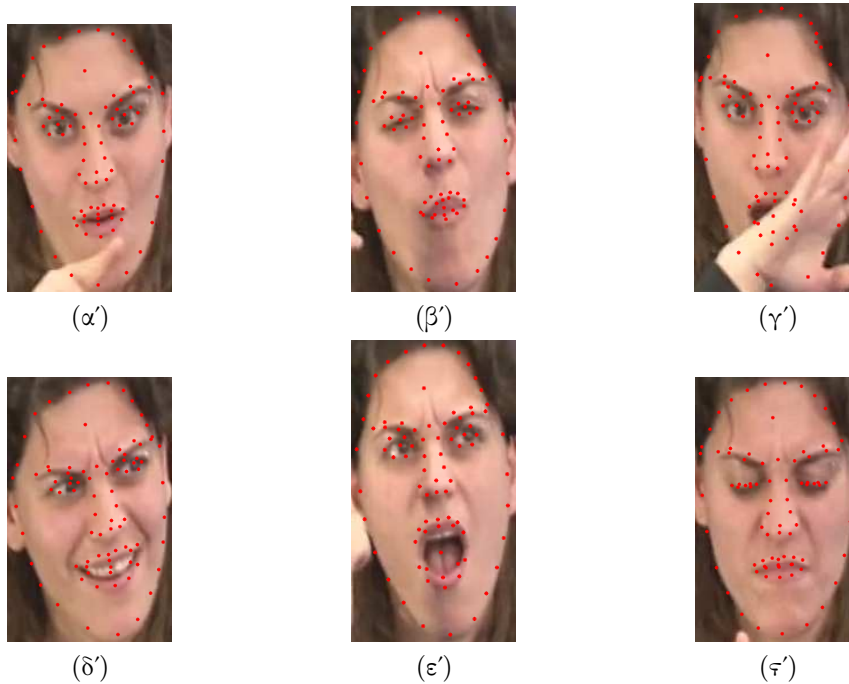
Σχήμα 3.5: Μεταβολή των 3 πρώτων συνιστωσών σχήματος και υφής για τη βάση της ENF του IEL στο διάστημα $[-3\sqrt{\lambda_i}, 3\sqrt{\lambda_i}]$.



Σχήμα 3.6: Μεταβολή των 3 πρώτων συνιστωσών σχήματος για τη βάση BU400 στο διάστημα $[-3\sqrt{\lambda_i}, 3\sqrt{\lambda_i}]$.

Για να φάνει καλύτερα η μεταβολή στο σχήμα λόγω της αλλαγής της πόζας κεφαλιού απεικονίζουμε στο Σχήμα 3.6 το αντίστοιχο μοντέλο σχήματος και για τη βάση BU400. Όπως φαίνεται, η συγκεκριμένη βάση εμφανίζει μεγαλύτερη μεταβλητότητα σε κινήσεις του κεφαλιού, η οποία περιγράφεται από τη μεταβολή των δύο πρώτων συνιστωσών σχήματος. Ωστόσο, η μεταβολή αυτή της πόζας κεφαλιού δεν παρουσιάζεται διαχωρισμένη στην οριζόντια και κατακόρυφη διεύθυνση, με αποτέλεσμα να μην είναι δυνατή η άμεση αντιστοίχιση των παραμέτρων του μοντέλου με τις γωνίες της κατεύθυνσης της πόζας.

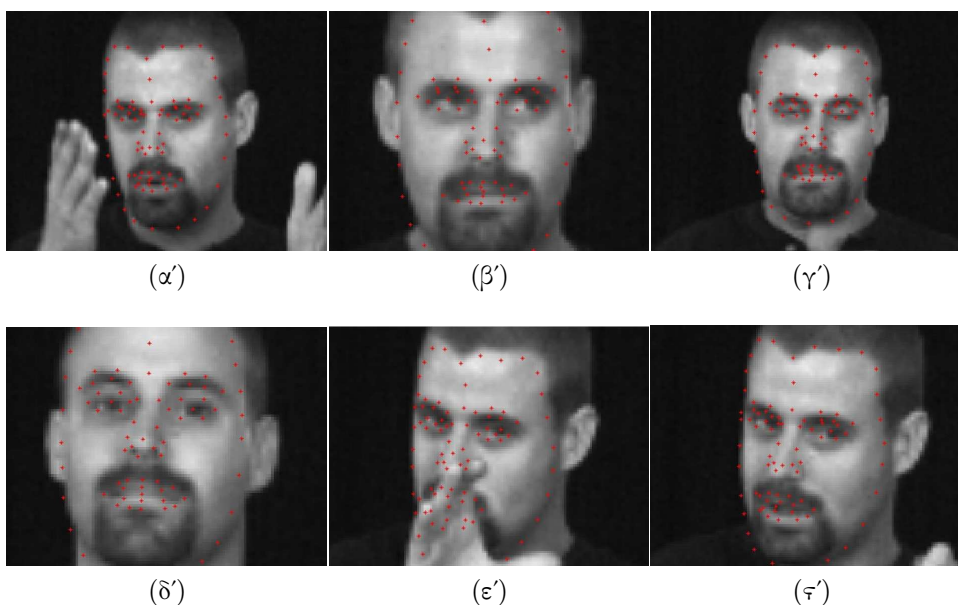
3.3.2 Αποτελέσματα εφαρμογής των AAMs σε νέες εικόνες



Σχήμα 3.7: Παραδείγματα εφαρμογής του AAM για τη βάση δεδομένων της ENΓ του ΙΕΛ.

Στη συνέχεια εφαρμόζουμε τα AAMs που κατασκευάσαμε σε νέες εικόνες από τις δύο βάσεις δεδομένων. Χρησιμοποιούμε τον αλγόριθμο προσαρμογής πλήρους τάξης $r = m$ (simultaneous inverse compositional), έτσι ώστε να προκύψει το καλύτερο δυνατόν αποτέλεσμα. Με τον τρόπο αυτό η αξιολόγηση των μεθόδων που ακολουθούν στα επόμενα κεφάλαια θα επηρεάζεται λιγότερο από τυχόν σφάλματα κατά την διαδικασία της εφαρμογής των AAMs. Στο Σχήμα 3.7 παρουσιάζονται κάποια παραδείγματα από την εφαρμογή του AAM για τη βάση εικόνων της ENΓ του ΙΕΛ. Στο Σχήμα 3.8 απεικονίζονται αντίστοιχα παραδείγματα για τη βάση δεδομένων BU440.

Παρατηρούμε πως το αποτέλεσμα της εφαρμογής για τη βάση της ENΓ του ΙΕΛ είναι αρκετά ικανοποιητικό για τις περιπτώσεις που δεν έχουμε επικαλύψεις του προσώπου από τα χέρια. Πιο συγκεκριμένα, για την περιοχή των ματιών για την οποία κυρίως ενδιαφερόμαστε, διαπιστώνουμε πως τα σημεία γύρω από τα βλέφαρα παρακολουθούν το άνοιγμα και κλείσιμο των ματιών ενώ και το σημείο πάνω στην ίριδα ακολουθεί με κάποια σχετική ακρίβεια την κίνηση της ίριδας. Επίσης, ακόμα και σε περιπτώσεις που έχουμε επικαλύψεις το αποτέλεσμα της εφαρμογής κοντά στα



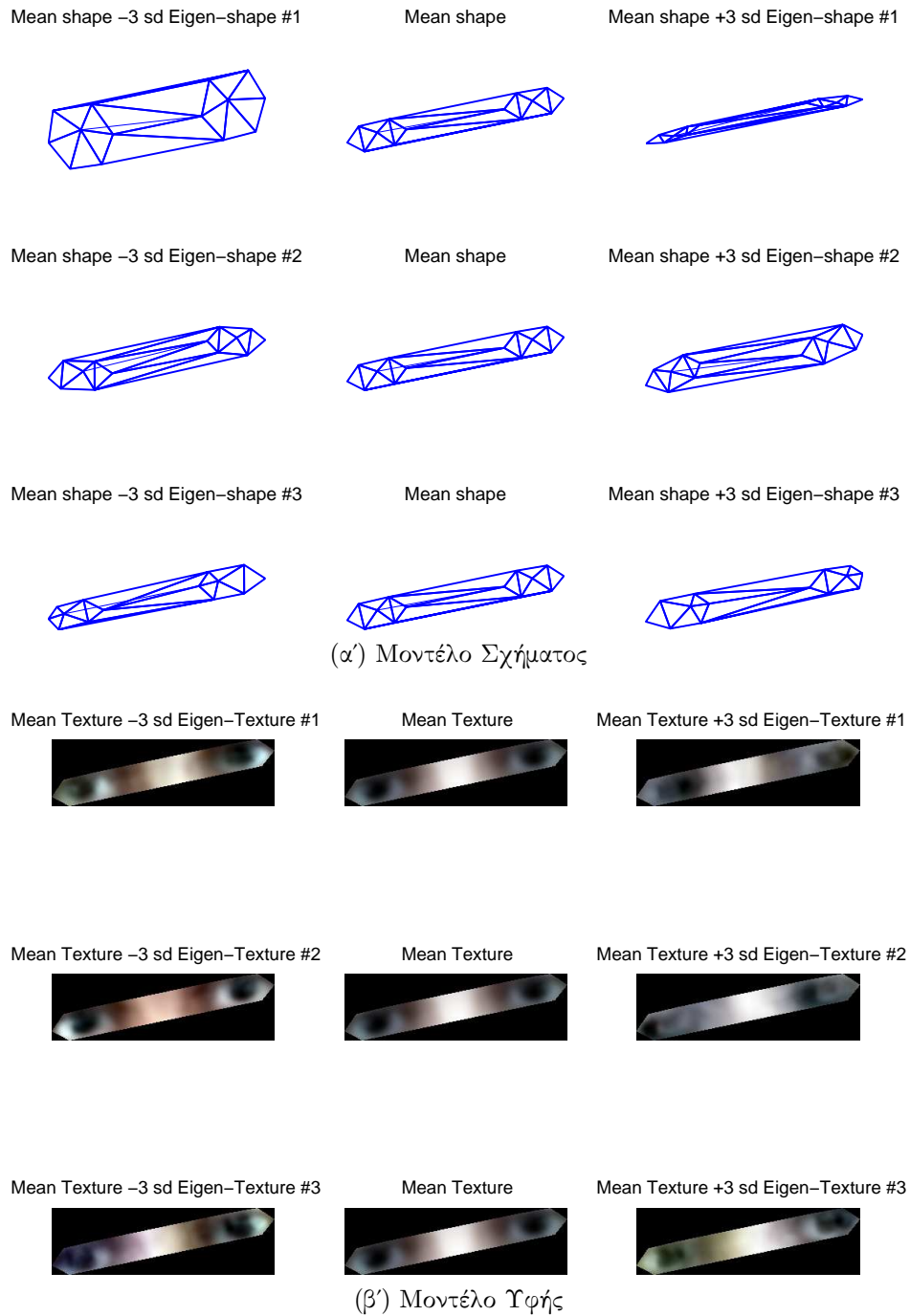
Σχήμα 3.8: Παραδείγματα εφαρμογής του AAM για τη βάση δεδομένων BU400.

μάτια παραμένει σε ικανοποιητικά επίπεδα, αφού οι επικαλύψεις λόγω των χεριών δεν φτάνουν ως το ύψος των ματιών. Το αποτέλεσμα αυτό μπορεί να βελτιωθεί λίγο παραπάνω εφαρμόζοντας και ένα τοπικό AAM για τα μάτια.

Για τη βάση BU400, ωστόσο, τα αποτελέσματα από την εφαρμογή των AAM δεν είναι ικανοποιητικά ακόμα και στις περιπτώσεις χωρίς επικαλύψεις. Όπως φαίνεται στις εικόνες 3.8β' και 3.8γ' μεταβολές στην περιοχή των ματιών δεν περιγράφονται με επιτυχία από τα σημεία της μάσκας προσώπου. Παρόλα αυτά, η προσαρμογή στο υπόλοιπο πρόσωπο είναι σχετικά ικανοποιητική, γι' αυτό η βάση αυτή μπορεί να χρησιμοποιηθεί στην εκτίμηση της πόζας κεφαλιού μιας και εμφανίζει μεγάλη μεταβλητότητα ως προς τις κινήσεις του κεφαλιού.

3.4 Τοπικό AAM για τα μάτια

Για να μοντελοποιήσουμε καλύτερα την περιοχή γύρω από τα μάτια, κατασκευάζουμε και εφαρμόζουμε ένα τοπικό AAM, αρχικοποιώντας τα σημεία του γράφου σχήματος από τα αντίστοιχα σημεία του μοντέλου για ολόκληρο το πρόσωπο. Στο Σχήμα 3.9 παρουσιάζονται οι μεταβολές των 3 πρώτων συνιστωσών σχήματος και υψής του τοπικού μοντέλου για τη βάση της ENΓ του IEL. Παρατηρούμε ότι η 1^η συνιστώσα του σχήματος μοντελοποιεί το άνοιγμα και κλείσιμο των ματιών, ενώ η 2^η και η 3^η περιγράφουν την κίνηση της ίριδας. Επιπλέον, και το μοντέλο υψής περιγράφει με επιτυχία μεταβολές στη θέση της ίριδας. Έτσι, λοιπόν, είναι δυνατόν να γίνει



Σχήμα 3.9: Μεταβολή των 3 πρώτων συνιστωσών σχήματος και υφής του τοπικού μοντέλου των ματιών για τη βάση της ENΓ του ΙΕΛ στο διάστημα $[-3\sqrt{\lambda_i}, 3\sqrt{\lambda_i}]$.

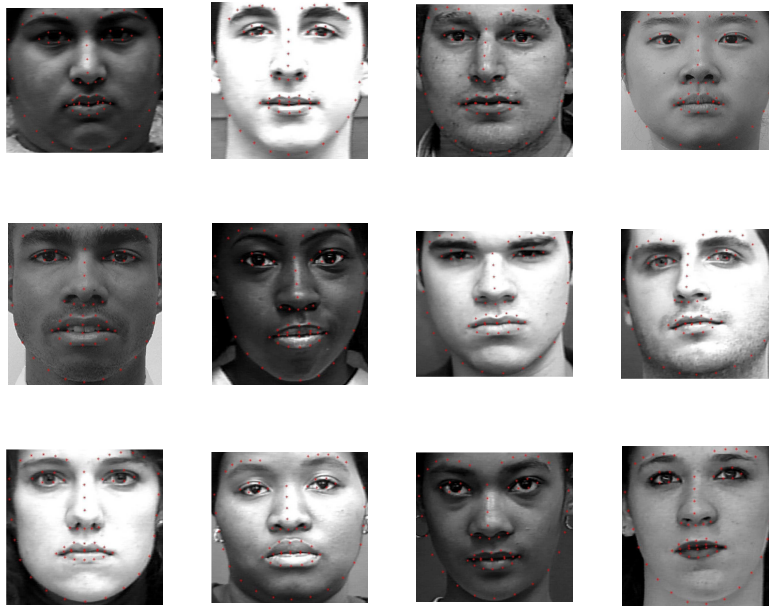


Σχήμα 3.10: Παραδείγματα εφαρμογής του τοπικού AAM των ματιών για τη βάση δεδομένων της ENG του IEL.

μια αντιστοίχιση μεταξύ των παραμέτρων του μοντέλου και της διεύθυνσης του eye gaze. Τέλος, στο Σχήμα 3.10 παρουσιάζονται κάποια παραδείγματα από την εφαρμογή του τοπικού μοντέλου των ματιών σε νέες εικόνες της βάσης της ENG του IEL. Παρατηρούμε μια μικρή βελτίωση στην προσαρμογή των σημείων των ματιών σε σχέση με το μοντέλο ολόκληρου του προσώπου που είχαμε πριν.

3.5 Γενικευμένο AAM για κάθε χρήστη

Για να μελετήσουμε το πρόβλημα της κατηγοριοποίησης συναισθήματος θα χρησιμοποιήσουμε την επισημειωμένη (μάσκα σχήματος με 68 σημεία) βάση εικόνων “The extended Cohn-Kanade (CK+)” του Carnegie Mellon University (CMU).[18] Στις εικόνες που περιέχονται εκφράζονται διάφορα συναισθήματα από ένα πλήθος διαφόρων ανθρώπων αρκετά διαφορετικών μεταξύ τους ως προς το φύλο, τη φυλετική καταγωγή και άλλα χαρακτηριστικά (Σχήμα 3.11). Για το λόγο αυτό θα εκπαιδεύσουμε ένα γενικό AAM κατάλληλο για κάθε άτομο, προσθέτοντας στο σύνολο εκπαίδευσης όσο το δυνατόν περισσότερες εικόνες διαφορετικών ατόμων. Έτσι, το μοντέλο θα αντιπροσωπεύει και τη μεταβλητότητα λόγω της ταυτότητας του ατόμου, η οποία θα περιγράφεται κυρίως από το μοντέλο υψής (Σχήμα 3.12).



Σχήμα 3.11: Παραδείγματα διαφορετικών προσώπων από τη βάση “The extended Cohn-Kanade (CK+)” του CMU.



Σχήμα 3.12: Μεταβολή των 3 πρώτων συνιστωσών υφής του γενικού ΑΑΜ για τη βάση “CK+” του CMU στο διάστημα $[-3\sqrt{\lambda_i}, 3\sqrt{\lambda_i}]$.

Κεφάλαιο 4

Εκτίμηση του Eye Gaze από χαρακτηριστικά σημεία του προσώπου και των ματιών

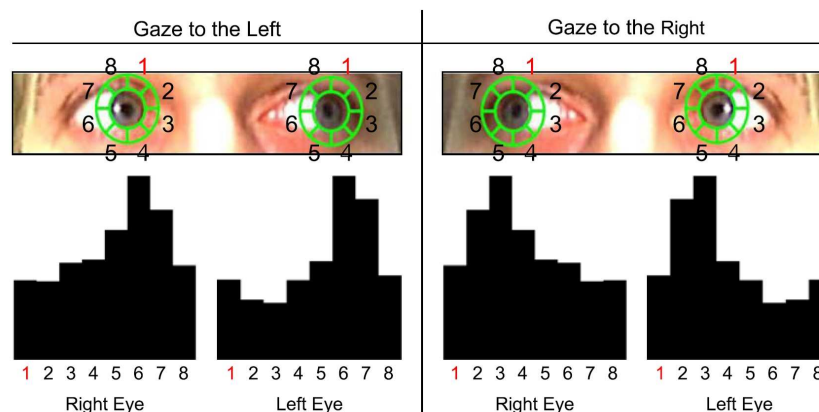
4.1 Εισαγωγή

Στο συγκεκριμένο κεφάλαιο θα παρουσιαστούν μέθοδοι για την εκτίμηση του eye gaze που δεν χρησιμοποιούν κάποια τεχνική κατηγοριοποίησής του σε διακριτές κλάσεις. Έτσι δεν είναι απαραίτητο να προχωρήσουμε στην εκπαίδευση κάποιου ταξινομητή από ένα σύνολο εκπαίδευσης με προκαθορισμένες κατευθύνσεις του gaze. Οι τεχνικές οι οποίες θα αναφερθούν χρησιμοποιούν ένα σύνολο από χαρακτηριστικά σημεία του προσώπου και των ματιών (ίριδα, γωνίες ματιών, άκρες του στόματος) προκειμένου να πραγματοποιήσουν εκτίμηση του eye gaze. Τα σημεία αυτά μπορούν να βρεθούν από την μάσκα σχήματος των AAMs, που περιγράφηκαν στο προηγούμενο κεφάλαιο.

Αρχικά παρουσιάζεται μια απλή τεχνική που προσπαθεί να κάνει μια εκτίμηση του gaze με βάση τη θέση της ίριδας. Αρχικά με χρήση μορφολογικού φιλτραρίσματος καθώς και του μετασχηματισμού Hough προσδιορίζεται το κέντρο της ίριδας, ενώ στη συνέχεια γίνεται δειγματοληψία στην περιοχή γύρω από αυτή και κατασκευάζονται ιστογράμματα φωτεινότητας.

Στη συνέχεια του κεφαλαίου θα αναφερθούμε στο πρόβλημα εκτίμησης της πόζας κεφαλιού. Για το σκοπό αυτό αναλύεται μια μέθοδος υπολογισμού των γωνιών στρέψης βασισμένη στις παραμέτρους ενός αφινικού μετασχηματισμού για την περιοχή του προσώπου. Έπειτα, περιγράφεται ένα γεωμετρικό μοντέλο για τα μάτια και γίνεται ο υπολογισμός του eye gaze χρησιμοποιώντας την πληροφορία για την πόζα κεφαλιού αλλά και χωρίς αυτήν. Τέλος, γίνεται μια αξιολόγηση των μεθόδων και καταγράφονται τα συμπεράσματα που προκύπτουν.

ΚΕΦΑΛΑΙΟ 4. ΕΚΤΙΜΗΣΗ ΤΟΥ EYE GAZE ΑΠΟ 62 ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΣΗΜΕΙΑ ΤΟΥ ΠΡΟΣΩΠΟΥ ΚΑΙ ΤΩΝ ΜΑΤΙΩΝ



Σχήμα 4.1: Κατευθύνσεις του βλέματος στα αριστερά και στα δεξιά μαζί με τα αντίστοιχα ιστογράμματα φωτεινότητας. Έχει ληφθεί από το [35].

4.2 Εκτίμηση του gaze από τη θέση της ίριδας

4.2.1 Γενική περιγραφή της μεθόδου

Όπως μπορεί εύκολα να παρατηρηθεί, όταν κάποιος αλλάζει την κατεύθυνση του βλέματος του η σφαίρα του ματιού περιστρέφεται και επομένως αλλάζει η θέση της ίριδας. Άρα, εντοπίζοντας τη σχετική θέση της ίριδας ως προς το υπόλοιπο ορατό μέρος της σφαίρας του ματιού μπορεί να γίνει μια αρχική εκτίμηση της κατεύθυνσης του gaze. Η διαδικασία αυτή μπορεί να πραγματοποιηθεί αξιοποιώντας την αντίθεση στη φωτεινότητα μεταξύ ίριδας, δέρματος και κερατοειδούς του ματιού.

Αρχικά πρέπει να βρεθεί η περιοχή εκείνη του ματιού που αντιπροσωπεύει την ίριδα. Έτσι, επιλέγεται το κόκκινο κανάλι από την εικόνα της περιοχής του ματιού επειδή συνήθως η ίριδα περιέχει σε μικρό βαθμό την κόκκινη απόχρωση. Στη συνέχεια με μορφολογικό φιλτράρισμα είναι δυνατόν να βρεθεί η δυαδική εικόνα για την περιοχή της ίριδας καθώς και το περίγραμμά της. Στη συνέχεια, θεωρώντας το σχήμα της ίριδας σαν έναν κύκλο, εφαρμόζουμε τον κυκλικό μετασχηματισμό Hough εντοπίζοντας έτσι το κέντρο της ίριδας μαζί με την ακτίνα της.

Για την αναγνώριση της κατεύθυνσης του gaze αναλύεται η κατανομή της φωτεινότητας στην περιοχή γύρω από την ίριδα και συγκεκριμένα μέσα σε έναν δακτύλιο που ορίζεται από τον κύκλο της ίριδας και ένα κύκλο διπλάσιας ακτίνας. Τέλος, ο δακτύλιος αυτός χωρίζεται σε επιμέρους τομείς, ανάλογα με την εκτίμηση του gaze που επιθυμούμε να έχουμε, και κατασκευάζονται τα αντίστοιχα ιστογράμματα φωτεινότητας. Τα ιστογράμματα αυτά προκύπτουν διαφορετικά για τις διάφορες κατευθύνσεις του βλέματος αντιπροσωπεύοντας την αντίθεση μεταξύ του λευκού χρώματος του κερατοειδούς και του χρώματος του δέρματος.

Στο Σχήμα 4.1 παρουσιάζονται δύο διαφορετικές κατευθύνσεις για το βλέμα, αριστερά και δεξιά (με βάση το πως κοιτάει η εικόνα), καθώς και τα αντίστοιχα ιστογράμματα. Παρατηρούμε, και για τα δύο μάτια, πως όταν η ίριδα πλησιάζει τις άκρες των ματιών τότε η τιμή του ιστογράμματος για τους αντίστοιχους τομείς μικραίνει λόγω του χρώματος του δέρματος, ενώ αυξάνεται στους τομείς της αντίθετης κατεύθυνσης όπου κυριαρχεί το λευκό χρώμα του κερατοειδούς.

4.2.2 Μορφολογικό φίλτράρισμα

Άρχικα, πριν την εφαρμογή οποιουδήποτε είδους φιλτραρίσματος, είναι απαραίτητο να βρεθεί η περιοχή των ματιών. Αυτό μπορεί να γίνει αρκετά εύκολα αξιοποιώντας τη μάσκα σχήματος του AAM που προσαρμόσαμε στο προηγούμενο κεφάλαιο. Αν θεωρήσουμε σαν (x_i^{eye}, y_i^{eye}) τα 7 σημεία της μάσκας που αντιστοιχούν στο κάθε μάτι, τότε η περιοχή του κάθε ματιού I_{eye} θα βρίσκεται μέσα σε ένα ορθογώνιο πλαίσιο που ορίζεται από την κάτω αριστερά και την πάνω δεξιά κορυφή:

$$Bottom_Left = \min_{i=1:7} (x_i^{eye}, y_i^{eye}) \quad (4.1)$$

$$Top_Right = \max_{i=1:7} (x_i^{eye}, y_i^{eye}) \quad (4.2)$$

Στη συνέχεια γίνεται μια εξομάλυνση της εικόνας εφαρμόζοντας median φίλτράρισμα ($I'_{eye} = med_W(I_{eye})$) προκειμένου να αποφευχθούν μεμονωμένες μεγάλες ή μικρές τιμές.

Η ίριδα αποτελεί μια ομογενή περιοχή της εικόνας με αρκετά χαμηλές τιμές φωτεινότητας. Άρα, αποτελεί “κοιλιάδα” για την εικόνα και μπορεί να προσδιοριστεί με την εφαρμογή ενός top-hat μετασχηματισμού:

$$valley(I'_{eye}) = (I'_{eye} \bullet B) - I'_{eye} \quad (4.3)$$

Αν στην εξίσωση (4.3) αντικαταστήσουμε το κλασικό closing με το reconstruction closing $\rho^+(m|I'_{eye})$, ορίζοντας σαν marker $m = I'_{eye} + h$, τότε μπορούμε να εντοπίσουμε “λεκάνες” της εικόνας με βάθος h :

$$basin(I'_{eye}) = \rho^+(I'_{eye} + h|I'_{eye}) - I'_{eye} \quad (4.4)$$

Το reconstruction closing ορίζεται με βάση το conditional erosion του marker m με ένα δίσκο B μοναδιαίας ακτίνας $\epsilon_B(m|f) = (m \ominus B) \vee f$ ως εξής:

$$\rho_B^+(g|f) = \lim_{n \rightarrow \infty} \epsilon_B^n(g|f) \quad (4.5)$$

Η σταθερά h προσδιορίζεται ως ένα ποσοστό του εύρους τιμών ($\max(I'_{eye}) - \min(I'_{eye})$) της εικόνας, κοντά στο 20% περίπου. Για να προκύψει

η δυαδική εικόνα I_{iris} που θα αντιπροσωπεύει την περιοχή της ίριδας γίνεται καταωφλιοποίηση σε ένα επίπεδο κοντά στο $(40\% - 50\%) \cdot h$. Επειδή η εικόνα I_{iris} μπορεί να περιέχει μικρές κυλίδες ή οπές, που μπορεί να δυσχεραίνουν την εύρεση της θέσης της ίριδας, εφαρμόζεται διαδοχικό closing και opening φιλτράρισμα:

$$I'_{iris} = (I_{iris} \bullet B) \circ B \quad (4.6)$$

Στη συνέχεια, προκειμένου να χρησιμοποιήσουμε σε επόμενο στάδιο τον κυκλικό μετασχηματισμό Hough, θα πρέπει να εξάγουμε το περίγραμμα της περιοχής της ίριδας. Αυτό μπορεί να γίνει χρησιμοποιώντας οποιοδήποτε γραμμικό τελεστή για τον υπολογισμό ακμών, αλλά στην συγκεκριμένη περίπτωση θα εφαρμόσουμε τον μη γραμμικό τελεστή ακμών:

$$edge(I'_{iris}) = (I'_{iris} \oplus B) - (I'_{iris} \ominus B) \quad (4.7)$$

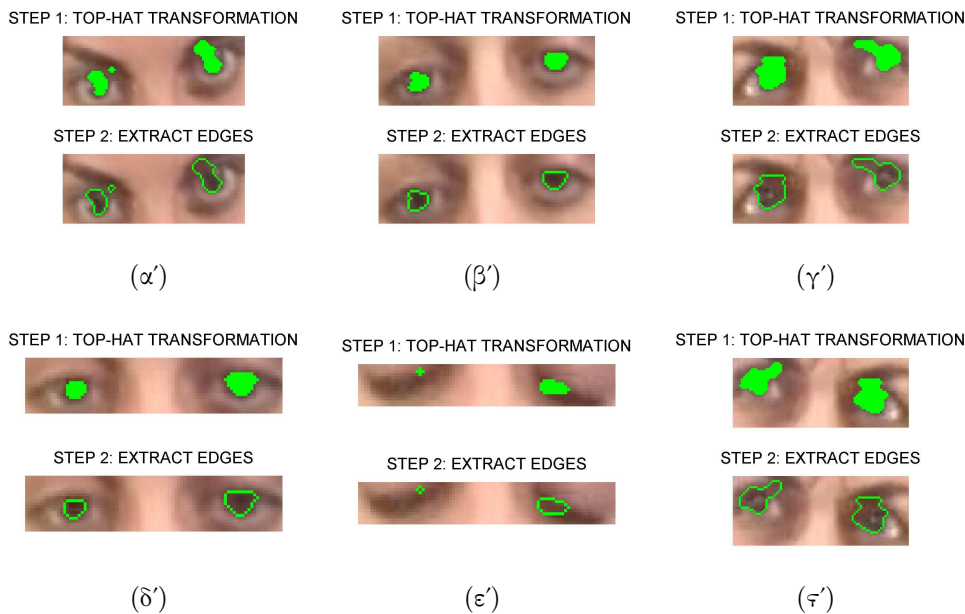
Επειδή είναι προτιμότερο οι ακμές πάνω στις οποίες θα εφαρμόσουμε το μετασχηματισμό Hough να περιέχουν όσο το δυνατόν λιγότερα pixels, εφαρμόζουμε στην εικόνα $E_{iris} = edge(I'_{iris})$ έναν μορφολογικό μετασχηματισμό λέπτυνσης:

$$E'_{iris} = Thin(E_{iris}) = \lim_{n \rightarrow \infty} [E_{iris} \odot (H, M)_{90n}]^n \quad (4.8)$$

$$E_{iris} \odot (H, M) = E_{iris} \setminus [E_{iris} \otimes (H, M)] \quad (4.9)$$

όπου το δομικό στοιχείο (H, M) αντιπροσωπεύει ένα ζευγάρι foreground-background προτύπων ενώ ο δείκτης $(H, M)_{90n}$ υποδηλώνει περιστροφή του στοιχείου κατά $90n^\circ$.

Στο Σχήμα 4.2 παρουσιάζονται παραδείγματα από τα δύο αυτά πρώτα στάδια επξεργασίας (του Top-Hat μετασχηματισμού και της εξαγωγής περιγράμματος) για τη βάση εικόνων της ENΓ του ΙΕΛ. Παρατηρούμε πως σε αρκετές περιπτώσεις (εικόνες 4.2β', 4.2δ') η περιοχή της ίριδας ανιχνεύεται με αρκετά ικανοποιητική ακρίβεια, ενώ σε άλλα παραδείγματα (4.2α', 4.2γ', 4.2ε') συμπεριλαμβάνονται στην δυαδική εικόνα και περιοχές εκτός ίριδας, όπως τα φρύδια ή σημεία της εικόνας με έντονη σκιά. Τέλος, η μέθοδος εμφανίζει προβλήματα και οδηγεί σε λανθασμένη εκτίμηση στην περίπτωση που τα μάτια είναι κλειστά και άρα δεν υπάρχει ορατό τμήμα της ίριδας προς ανίχνευση (εικόνα 4.2ε').

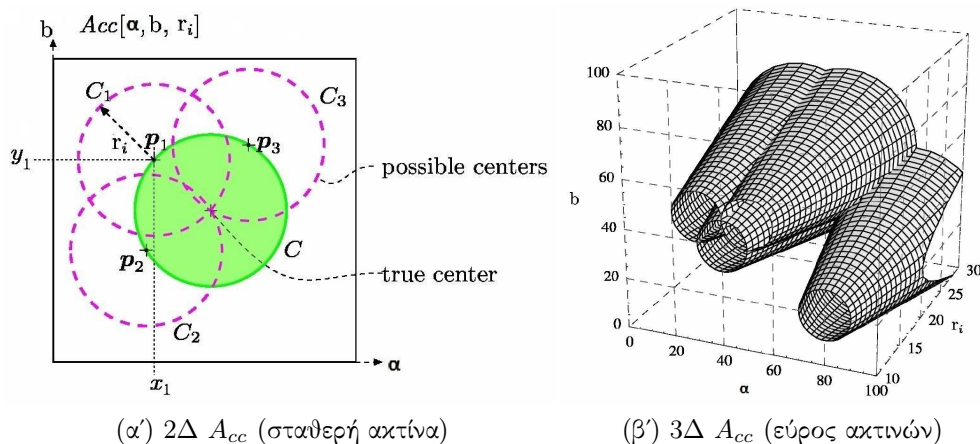


Σχήμα 4.2: Παραδείγματα εφαρμογής του Top-Hat μετασχηματισμού και της εξαγωγής περιγράμματος για τη βάση εικόνων της ENΓ του IEL.

4.2.3 Μετασχηματισμός Hough και σχεδίαση ιστογραμμάτων

Ο μετασχηματισμός Hough[28] αποτελεί μια ευρέως χρησιμοποιημένη τεχνική για τον εντοπισμό απλών σχημάτων, όπως ευθείες γραμμές και κύκλους, σε εικόνες. Ο συγκεκριμένος μετασχηματισμός εφαρμόζεται στη δυαδική εικόνα ακμών της αρχικής εικόνας και μετασχηματίζει κάθε σημείο από το επίπεδο (x, y) της εικόνας στο επίπεδο των παραμέτρων του ζητούμενου σχήματος.

Στην περίπτωση της ευθείας γραμμής, κάθε σημείο (x_i, y_i) που ανήκει στην ευθεία $y_i = ax_i + b$ μετασχηματίζεται στο επίπεδο (a, b) των παραμέτρων στην ευθεία $b = -x_i a + y_i$. Έτσι, για κάθε μη μηδενικό pixel της εικόνας ακμών υπολογίζεται η αντίστοιχη ευθεία και κάθε σημείο του επιπέδου (a, b) που ανήκει σε αυτήν αυξάνει την τιμή του κατά 1 μονάδα. Έτσι κατασκευάζεται ένας πίνακας συσσώρευσης $A_{cc}(a_i, b_i)$ (accumulator matrix) για το χώρο των παραμέτρων (Hough space), ο οποίος εκφράζει το κατά πόσο είναι δυνατόν να υπάρξει μια ευθεία με παραμέτρους (a_i, b_i) στην αρχική εικόνα. Στη συνέχεια, από τον πίνακα A_{cc} κρατούνται οι τιμές εκείνες που είναι μεγαλύτερες από ένα ποσοστό της μέγιστης τιμής του, συνήθως 40% – 50%. Έπειτα, ανιχνεύονται τα τοπικά μέγιστα του πίνακα και έτσι προκύπτουν οι πιο πιθανές ευθείες γραμμές της αρχικής εικόνας.



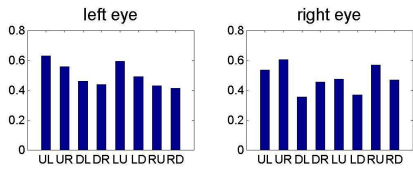
Σχήμα 4.3: Πίνακες συσσώρευσης για τον κυκλικό μετασχηματισμό Hough. [Έχει ληφθεί από το NTUA.]

Ακολουθώντας παρόμοια μεθοδολογία μπορεί να γίνει και ο εντοπισμός των κύκλων που μπορεί να περιλαμβάνονται σε μια εικόνα. Στην περίπτωση αυτή ένα σημείο (x_i, y_i) της εικόνας που ανήκει σε έναν κύκλο κέντρου (a, b) και ακτίνας r θα ικανοποιεί τη σχέση: $(x_i - a)^2 + (y_i - b)^2 = r^2$. Στον κυκλικό μετασχηματισμό Hough υπάρχουν 3 παράμετροι (a, b, r) που πρέπει να προσδιοριστούν, ενώ στις ευθείες γραμμές είχαμε μόνο 2. Έτσι ο πίνακας συσσώρευσης $A_{cc}(a_i, b_i, r_i)$ γίνεται 3Δ, σε αντίθεση με πριν που ήταν δύο διαστάσεων.

Για τον υπολογισμό ενός 2Δ πίνακα $A_{cc}(a_i, b_i, r_i)$ αρχικά θεωρούμε ότι η ακτίνα r_i παραμένει σταθερή. Στην περίπτωση αυτή, κάθε σημείο (x_i, y_i) της δυαδικής εικόνας αχμών αντιστοιχίζεται στον χώρο των παραμέτρων σε έναν κύκλο με κέντρο το σημείο (x_i, y_i) και ακτίνα r_i ($(a - x_i)^2 + (b - y_i)^2 = r^2$). Στη συνέχεια κατασκευάζεται ο 2Δ πίνακας A_{cc} με τον ίδιο τρόπο όπως και στην ευθεία γραμμή. Στο Σχήμα 4.3α' απεικονίζεται ο 2Δ πίνακας συσσώρευσης για κύκλους με σταθερή ακτίνα. Για τη δημιουργία του 3Δ πίνακα συσσώρευσης επαναλαμβάνεται η ίδια διαδικασία για ένα εύρος διαφορετικών ακτινών, με μέγιστη τιμή ένα ποσοστό της μικρότερης διάστασης της εικόνας (Σχήμα 4.3β'). Στη συνέχεια ακολουθείται για τον 3Δ A_{cc} η ίδια διαδικασία που περιγράφηκε και στις ευθείες γραμμές, προκειμένου να προσδιοριστούν οι παράμετροι των πιο πιθανών κύκλων της αρχικής εικόνας.

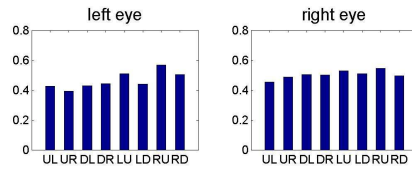
Ο κυκλικός μετασχηματισμός Hough μπορεί να χρησιμοποιηθεί για τον εντοπισμό της ίριδας του ματιού, μιας και το σχήμα της προσεγγίζει αρκετά αυτό του κύκλου.[4] Εφαρμόζοντας, λοιπόν, το μετασχηματισμό στις δυαδικές εικόνες του περιγράμματος της ίριδας, βρίσκουμε και σχεδιάζουμε τους κύκλους που είναι πιθανότερο να αντιπροσωπεύουν την ίριδα (Σχήμα 4.4 [κόκκινο χρώμα]).

STEP 3: FIND CIRCLES WITH HOUGH TRANSFORMATION



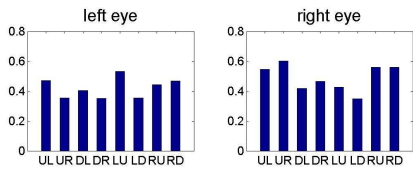
(α')

STEP 3: FIND CIRCLES WITH HOUGH TRANSFORMATION



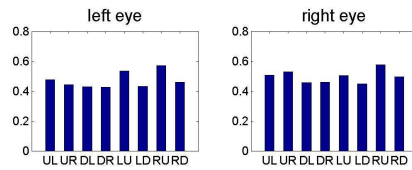
(β')

STEP 3: FIND CIRCLES WITH HOUGH TRANSFORMATION



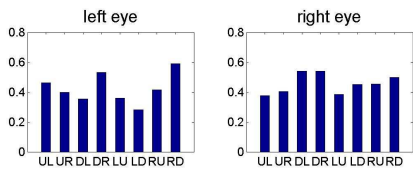
(γ')

STEP 3: FIND CIRCLES WITH HOUGH TRANSFORMATION



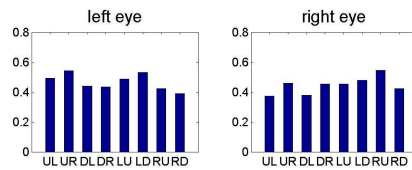
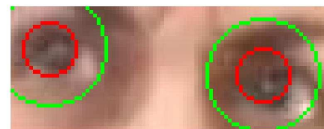
(δ')

STEP 3: FIND CIRCLES WITH HOUGH TRANSFORMATION



(ε')

STEP 3: FIND CIRCLES WITH HOUGH TRANSFORMATION



(ϵ')

Σχήμα 4.4: Παραδείγματα εφαρμογής του κυκλικού μετασχηματισμού Hough και κατασκευή ιστογραμμάτων φωτεινότητας για τη βάση εικόνων της ENT του IEL.

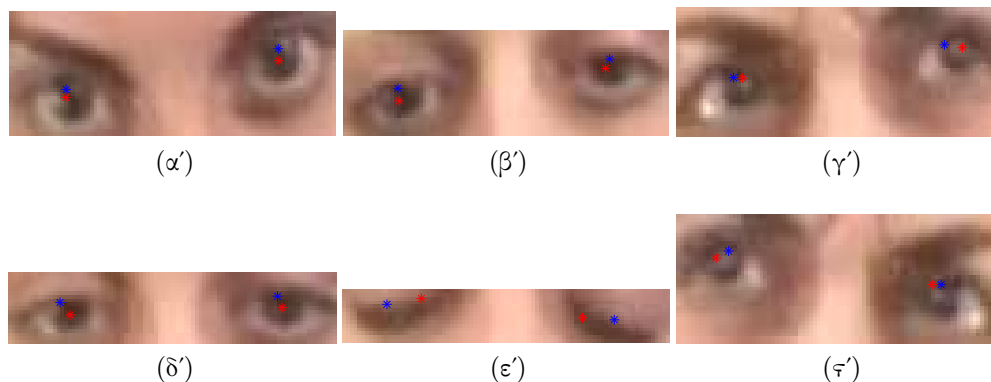
ΚΕΦΑΛΑΙΟ 4. ΕΚΤΙΜΗΣΗ ΤΟΥ EYE GAZE ΑΠΟ 68 ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΣΗΜΕΙΑ ΤΟΥ ΠΡΟΣΩΠΟΥ ΚΑΙ ΤΩΝ ΜΑΤΙΩΝ

Στη συνέχεια, σχεδιάζεται ένας ομόκεντρος κύκλος διπλάσιας ακτίνας [πράσινο χρώμα] και η περιοχή ανάμεσα στους δύο κύκλους χωρίζεται σε 8 τομείς: Up-Left(UL), Up-Right(UR), Down-Left(DL), Down-Right(DR), Left-Up(LU), Left-Down(LD), Right-Up(RU), Right-Down(RD). Έπειτα, κατασκευάζονται τα αντίστοιχα ιστογράμματα αφαιρώντας από τη μέγιστη τιμή της εικόνας τη μέση τιμή φωτεινότητας του κάθε τομέα. Παρατηρούμε, πως σε όλες τις περιπτώσεις (με εξαίρεση εκείνη που τα μάτια είναι κλειστά) ο κύκλος που περιέχει την ίριδα ανιχνεύεται με ικανοποιητική ακρίβεια. Ωστόσο, τα ιστογράμματα που προκύπτουν δεν είναι ικανά να δώσουν σαφή εκτίμηση της κατεύθυνσης του gaze. Παρατηρώντας τα ιστογράμματα, διαπιστώνουμε πως μόνο η στροφή της ίριδας προς τα πάνω καθώς και η ευθεία διεύθυνση μπορούν να γίνουν αντιληπτές από τα διαγράμματα. Αντίθετα, η στροφή προς τα αριστερά η τα δεξιά δεν είναι δυνατόν να περιγραφεί από τα ιστογράμματα, παρά μόνο σε κάποιες περιπτώσεις και όχι και για τα δύο μάτια (εικόνα 4.4γ’).

Η σχετική αυτή αποτυχία της μεθόδου οφείλεται, στο ότι παρόλο που ο κύκλος της ίριδας βρέθηκε αρκετά ικανοποιητικά, η ακρίβεια δεν ήταν αρκετή προκειμένου να κατασκευαστούν ιστογράμματα φωτεινότητας που να περιγράφουν τη γύρω αναλυτικά περιοχή. Ένας άλλος παράγοντας, είναι ότι οι τιμές φωτεινότητας επηρεάζονται έντονα από τις συνθήκες φωτισμού και τις σκιάσεις, με συνέπεια να οδηγούμαστε σε λανθασμένα αποτελέσματα σε σχέση με τα θεωρητικά αναμενόμενα. Από όλα αυτά γίνεται αντιληπτό ότι μια τόσο “τοπική” και απλή τεχνική, όπως αυτή που περιγράφηκε, δεν είναι πρακτικά δυνατόν να δώσει μια ικανοποιητική περιγραφή για το eye gaze, παρόλη την αρχική θεωρητική ανάλυση. Για το λόγο αυτό, είναι απαραίτητο να εφαρμόσουμε πιο σύνθετες τεχνικές που θα συνδυάζουν περισσότερες πληροφορίες από το πρόσωπο για την αποτελεσματικότερη εκτίμηση του gaze.

Σύγκριση της εκτίμησης του κέντρου της ίριδας με εκείνη του AAM

Παρόλο που τέλικος σκοπός της μεθόδου ήταν η κατασκευή ιστογραμμάτων, το μεγαλύτερο μέρος της αφιερώθηκε στην εύρεση της θέσης της ίριδας. Για το λόγο αυτό, κρίνεται σκόπιμο να συγκρίνουμε την εκτίμηση σχετικά με το κέντρο της ίριδας με εκείνη που προκύπτει από την μάσκα σχήματος του AAM. Στο Σχήμα 4.5 φαίνονται χαρακτηριστικά παραδείγματα που απεικονίζουν και τις δύο αυτές εκτιμήσεις. Παρατηρούμε πως στις περισσότερες εικόνες οι δύο εκτιμήσεις είναι σχετικά κοντά, με εκείνη μάλιστα της παρούσας μεθόδου να βρίσκεται πιο κοντά στην πραγματική θέση της ίριδας. Ωστόσο, στην περίπτωση που τα μάτια είναι κλειστά το AAM βρίσκει πιο σωστά την θεωρητικά αναμενόμενη θέση της ίριδας αφού συνδυάζει πληροφορία από ένα σύνολο εικόνων εκπαίδευσης. Έτσι, λοιπόν, συμπεραίνουμε πως η εκτίμηση της ίριδας που βρήκαμε με την συγκεκριμένη μέθοδο είναι αρκετά ακριβής έτσι ώστε να χρησιμοποιηθεί και σε επόμενες μεθόδους, είτε μόνη της είτε σε συνδυασμό με την εκτίμηση του AAM.



Σχήμα 4.5: Σύγκριση των εκτιμήσεων για το κέντρο της ίριδας για τη βάση εικόνων της ENF του IEL (με μπλε χρώμα η εκτίμηση του AAM και με κόκκινο της παρούσας μεθόδου).

4.3 Εκτίμηση της πόζας κεφαλιού

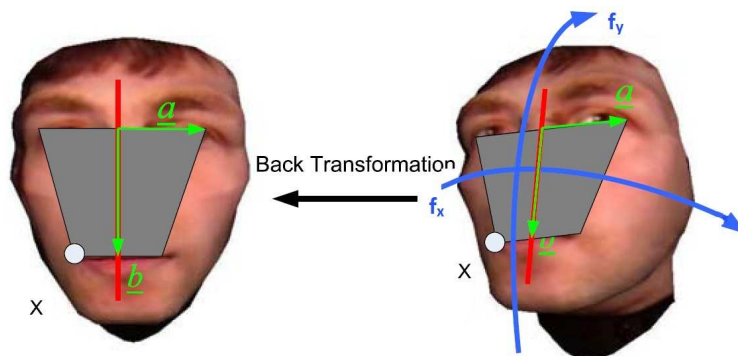
4.3.1 Περιγραφή της μεθόδου

Η εκτίμηση της πόζας του κεφαλιού αποτελεί ένα πρόβλημα που σχετίζεται άμεσα με τον υπολογισμό του eye gaze. Είναι φανερό πως ο άνθρωπος όταν θέλει να στρέψει κάπου το βλέμα του, συνήθως περιστρέφει πρώτα το κεφάλι του και στη συνέχεια επικεντρώνει την προσοχή του χρησιμοποιώντας τα μάτια του. Γι' αυτό το λόγο, η πόζα κεφαλιού μπορεί να θεωρηθεί σαν το καθολικό (global) gaze, ενώ οι κινήσεις των ματιών αποτελεί το τοπικό (local) gaze. Επομένως, για είναι ένα γεωμετρικό μοντέλο εκτίμησης του eye gaze αρκετά αποτελεσματικό και ακριβές, θα πρέπει να συνδυάζει πληροφορία για την πόζα του κεφαλιού προκειμένου να υπολογιστεί η πραγματική θέση του κέντρου της σφαίρας του ματιού.

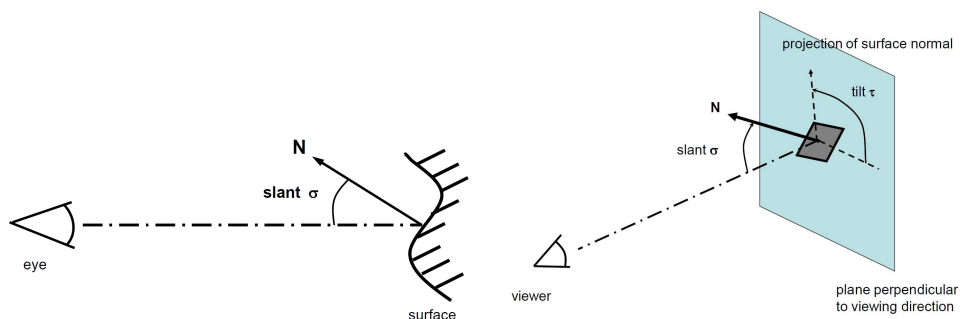
Σε προηγούμενο κεφάλαιο είδαμε ότι η μεταβολή της πόζας κεφαλιού μπορεί να περιγραφεί από τις δύο πρώτες συνιστώσες του μοντέλου σχήματος του AAM. Ωστόσο, με τον τρόπο αυτό δεν είναι εύκολος ο διαχωρισμός της οριζόντιας (roll) από την κατακόρυφη (pitch) στρέψη. Για το λόγο αυτό, θα χρησιμοποιήσουμε μια αναλυτική τεχνική για τον προσδιορισμό του κάθετου στο επίπεδο του προσώπου διανύσματος (face normal) και επομένως και των αντίστοιχων γωνιών.

Η μέθοδος αυτή βασίζεται στην παραμόρφωση που υφίσταται ένα τραπέζιο σχήμα, ορισμένο πάνω στο επίπεδο του προσώπου, όταν αλλάζει η πόζα κεφαλιού. Το τραπέζιο αυτό περιγράφεται από 4 σημεία της μάσκας σχήματος του AAM: τις εξωτερικές γωνίες των ματιών και του στόματος. Όταν το κεφάλι βρίσκεται σε ευθεία θέση τότε το τραπέζιο εμφανίζεται συμμετρικό. Όταν όμως αλλάζει η πόζα του, τότε η περιοχή ανάμεσα στο στόμα και τα μάτια, που ορίζεται από το τραπέζιο,

ΚΕΦΑΛΑΙΟ 4. ΕΚΤΙΜΗΣΗ ΤΟΥ EYE GAZE ΑΠΟ
70 ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΣΗΜΕΙΑ ΤΟΥ ΠΡΟΣΩΠΟΥ ΚΑΙ ΤΩΝ ΜΑΤΙΩΝ



(α') Τραπεζίο ορισμένο πάνω στο επίπεδο του προσώπου. Έχει ληφθεί από το [35].



(β') Ορισμός της γωνίας slant (σ)

(γ') Ορισμός της γωνίας tilt (τ)

Σχήμα 4.6: Εκτίμηση της πόζας κεφαλιού από τις παραμέτρους του αφινικού μετασχηματισμού ενός τραπεζίου ορισμένου στο επίπεδο του προσώπου.

παραμορφώνεται (Σχήμα 4.6α'). Για να υπολογίζουμε λοιπόν τις γωνίες στέψεις του κεφαλιού, αρκεί να προσδιορίσουμε τον αφινικό μετασχηματισμό U που προβάλλει το παραμορφωμένο τραπέζιο πίσω στην ευθεία διεύθυνση. Αν x είναι ένα σημείο του παραμορφωμένου επιπέδου, τότε το αντίστοιχο σημείο x' θα δίνεται από τη σχέση:

$$x' = Ux \quad (4.10)$$

Το τραπέζιο μπορεί να περιγραφεί από δύο διανύσματα που είναι κάθετα μεταξύ τους, όταν το πρόσωπο βρίσκεται στην ευθεία διεύθυνση. Το διάνυσμα a έχει αρχή το μέσο του ευθύγραμμου τμήματος που ενώνει τις δύο εξωτερικές γωνίες των ματιών και πέρας την εξωτερική γωνία του δεξιού ματιού. Το διάνυσμα b ξεκινά και αυτό από το ίδιο σημείο και καταλήγει στο μέσο του ευθύγραμμου τμήματος που ενώνει τις άκρες του στόματος. Προκειμένου τα δύο αυτά διανύσματα να προκύψουν κανονικοποιημένα, διαιρούμε με το μήκος του διανύσματος b . Επίσης, για την ευθεία διεύθυνση, υπολογίζεται ο λόγος μεταξύ του μήκους L_e του διανύσματος a και του

μήκους L_f του διανύσματος \mathbf{b} :

$$R_e = \frac{L_e}{L_f} \quad (4.11)$$

Έτσι τα διανύσματα \mathbf{a}' , \mathbf{b}' για την ευθεία πόζα θα δίνονται από τη σχέση:

$$[\mathbf{a}' \quad \mathbf{b}'] = \begin{bmatrix} R_e & 0 \\ 0 & -1 \end{bmatrix} \quad (4.12)$$

Ο αφινικός πίνακας μετασχηματισμού \mathbf{U} μπορεί λοιπόν να προσδιοριστεί με βάση τα κανονικοποιημένα διανύσματα \mathbf{a} , \mathbf{b} του παραμορφωμένου τραπεζίου από τη σχέση:

$$\mathbf{U}[\mathbf{a} \quad \mathbf{b}] = \begin{bmatrix} R_e & 0 \\ 0 & -1 \end{bmatrix} \Leftrightarrow \mathbf{U} = \begin{bmatrix} R_e & 0 \\ 0 & -1 \end{bmatrix} [\mathbf{a} \quad \mathbf{b}]^{-1} \quad (4.13)$$

Ο πίνακας \mathbf{U} μπορεί να διασπαστεί στην παρακάτω μορφή:

$$\mathbf{U} = \lambda_1 \mathbf{R}(\theta) \mathbf{P}(\lambda, \tau), \quad (4.14)$$

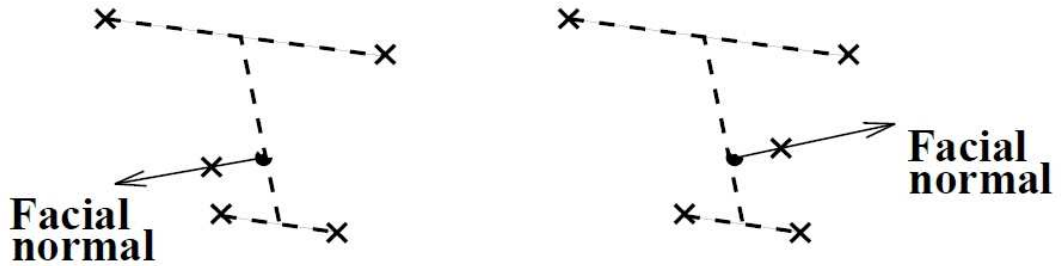
όπου λ_1 μια ισοτροπική κλιμάκωση, $\mathbf{R}(\theta)$ περιστροφή κατά γωνία θ γύρω από τον άξονα της κάμερας και $\mathbf{P}(\lambda, \tau)$ κλιμάκωση κατά λ στη διεύθυνση του $\boldsymbol{\tau} = (\cos \tau, \sin \tau)$.

Από τον όρο $\mathbf{P}(\lambda, \tau)$ μπορούν να προσδιοριστούν οι γωνίες slant (σ) και tilt (τ). Σαν slant (σ) ορίζεται η γωνία που σχηματίζει το κάθετο στο επίπεδο του προσώπου διάνυσμα \mathbf{N} (surface normal) με τον οπτικό άξονα της κάμερας (άξονας z) στον 3Δ χώρο (Σχήμα 4.6β'). Το tilt (τ) εκφράζει τη γωνία που σχηματίζει η προβολή στην εικόνα (επίπεδο xy) του διανύσματος \mathbf{N} με τον άξονα x (Σχήμα 4.6γ'). Η μέθοδος υλολογισμού τους θα περιγραφεί αναλυτικά σε επόμενη υποενότητα.

Γνωρίζοντας, λοιπόν, τις γωνίες slant και tilt το κάθετο διάνυσμα \mathbf{N} μπορεί να υπολογιστεί με βάση τη γεωμετρία των σχημάτων 4.5β' και 4.6γ' από τη σχέση:

$$\mathbf{N} = \begin{bmatrix} n_x \\ n_y \\ n_z \end{bmatrix} = \begin{bmatrix} \sin \sigma \cos \tau \\ \sin \sigma \sin \tau \\ \cos \sigma \end{bmatrix} \quad (4.15)$$

Λόγω της χρήσης τριγωνομετρικών συναρτήσεων για τον υπολογισμό των παραπάνω γωνιών, προκύπτει μια ασάφεια $\pm 180^\circ$ στην τιμή του tilt που επηρεάζει το πρόσημο της n_x συνιστώσας του διανύσματος \mathbf{N} . Η ασάφεια αυτή μπορεί να αντιμετωπιστεί χρησιμοποιώντας κάποιο επιπλέον σημείο του προσώπου, στη συγκεκριμένη περίπτωση αυτό της μύτης. Αν το σημείο της μύτης βρίσκεται στα αριστερά του διανύσματος \mathbf{b} , τότε το πρόσημο $\text{sign}(n_x)$ του n_x είναι αρνητικό και η γωνία τ γίνεται $180^\circ + \tau$, αλλιώς είναι θετικό και η γωνία τ παραμένει αμετάβλητη (Σχήμα 4.7).



Σχήμα 4.7: Αντιμετώπιση της ασάφειας του tilt αξιοποιώντας τη θέση της μύτης. Έχει ληφθεί από το [9].

Στο σημείο αυτό έχουμε υπολογίσει το κάθετο διάνυσμα \mathbf{N} χωρίς να λάβουμε υπόψιν την περιστροφή κατά μια γωνία θ γύρω από τον άξονα της κάμερας. Πολλαπλασιάζοντας, λοιπόν, με έναν πίνακα περιστροφής γύρω από τον άξονα z παίρνουμε το τελικό διάνυσμα \mathbf{N}' :

$$\mathbf{N}' = \begin{bmatrix} n'_x \\ n'_y \\ n'_z \end{bmatrix} = \begin{bmatrix} \cos(-\theta) & -\sin(-\theta) & 0 \\ \sin(-\theta) & \cos(-\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} n_x \\ n_y \\ n_z \end{bmatrix} \quad (4.16)$$

Τέλος, η πόζα κεφαλιού προσδιορίζεται από τις γωνίες οριζόντιας ϕ_x και κάθετης ϕ_y στρέψης που δίνονται από τις σχέσεις:

$$\phi_x = \arctan\left(\frac{n'_x}{n'_z}\right) \quad (4.17)$$

$$\phi_y = \arctan\left(\frac{n'_y}{n'_z}\right) \quad (4.18)$$

4.3.2 Υπολογισμός των γωνιών slant και tilt

Όπως ήδη αναφέρθηκε, οι γωνίες slant και tilt μπορούν να προσδιοριστούν από τις παραμέτρους του πίνακα αφινικού μετασχηματισμού \mathbf{U} , ο οποίος γράφεται στη μορφή της σχέσης (4.14). Ο όρος $\mathbf{P}(\lambda, \tau)$ είναι ένας συμμετρικός πίνακας που εκφράζει την κλιμάκωση κατά ένα παράγοντα λ στη διεύθυνση του ιδιοδιανύσματος $\tau = (\cos \tau, \sin \tau)$ του πίνακα $\mathbf{P}(\lambda, \tau)$ και μπορεί να γραφεί στη μορφή:

$$\mathbf{P}(\lambda, \tau) = \mathbf{R}(\tau) \begin{bmatrix} \lambda & 0 \\ 0 & 1 \end{bmatrix} \mathbf{R}(-\tau), \quad (4.19)$$

$$\text{όπου} \quad \mathbf{R}(\tau) = \begin{bmatrix} \cos \tau & -\sin \tau \\ \sin \tau & \cos \tau \end{bmatrix}. \quad (4.20)$$

Οι παράμετροι λ και τ μπορούν να προσδιοριστούν ορίζοντας το συμμετρικό και θετικά ορισμένο πίνακα $\mathbf{V} = \mathbf{U}^T \mathbf{U}$ [24], τα στοιχεία του οποίου θα είναι:

$$\mathbf{V} = \begin{bmatrix} \alpha & \beta \\ \beta & \gamma \end{bmatrix} \quad (4.21)$$

Ο πίνακας \mathbf{V} χρησιμοποιώντας τις σχέσεις (4.14), (4.19), (4.20) μπορεί να γραφεί και στη μορφή:

$$\mathbf{V} = \lambda_1^2 \mathbf{P}(\lambda^2, \tau). \quad (4.22)$$

Το trace και η ορίζουσα του πίνακα \mathbf{V} , με βάση τις παραπάνω εξισώσεις, θα δίνονται από τις σχέσεις:

$$\text{tr} \mathbf{V} = \alpha + \gamma = \lambda_1^2 (1 + \lambda^2) \quad (4.23)$$

$$\det \mathbf{V} = \alpha\gamma - \beta^2 = \lambda_1^4 \lambda^2 \quad (4.24)$$

Διαιρώντας τις παραπάνω σχέσεις κατά μέλη για να εξαληφθεί ο όρος λ_1^2 έχουμε:

$$\frac{(\text{tr} \mathbf{V})^2}{(\det \mathbf{V})} = \frac{(\alpha + \gamma)^2}{\alpha\gamma - \beta^2} = \frac{(1 + \lambda^2)^2}{\lambda^2} = 4\mu \quad (4.25)$$

Επιλύοντας την παραπάνω εξίσωση παίρνουμε $\lambda^2 = \{\lambda'^2, 1/\lambda'^2\}$, όπου:

$$\lambda'^2 = 2\mu - 1 + 2\sqrt{\mu(\mu - 1)} \quad (4.26)$$

Για τη γωνία περιστροφής tilt τ αντίστοιχα θα έχουμε:

$$\begin{bmatrix} \alpha & \beta \\ \beta & \gamma \end{bmatrix} = \lambda_1^2 \mathbf{R}(\tau) \begin{bmatrix} \lambda & 0 \\ 0 & 1 \end{bmatrix} \mathbf{R}(-\tau) \Leftrightarrow \quad (4.27)$$

$$\begin{bmatrix} \cos \tau & \sin \tau \\ -\sin \tau & \cos \tau \end{bmatrix} \begin{bmatrix} \alpha & \beta \\ \beta & \gamma \end{bmatrix} \begin{bmatrix} \cos \tau & -\sin \tau \\ \sin \tau & \cos \tau \end{bmatrix} = \lambda_1^2 \begin{bmatrix} \lambda & 0 \\ 0 & 1 \end{bmatrix} \quad (4.28)$$

Από τα μη διαγώνια στοιχεία παίρνουμε τη σχέση για τη γωνία τ :

$$-\sin^2 \tau \beta - \cos \tau \sin \tau (\alpha - \gamma) + \cos^2 \tau \beta = 0 \Rightarrow \tan 2\tau = \frac{2\beta}{\alpha - \gamma} \Leftrightarrow \quad (4.29)$$

$$2\tau^* = \text{Arctan} \left[\frac{2\beta}{\alpha - \gamma} \right] \quad (4.30)$$

Ωστόσο, λόγω της χρήσης του Arctan θα ισχύει:

$$2\tau = 2\tau^* \pm k\pi, \quad k = 0, 1, 2, \dots \quad \mu\epsilon \quad 2\tau^* \in [-\pi/2, \pi/2] \quad (4.31)$$

Έτσι προκύπτει μια ασάφεια $\pm\pi$, η οποία μπορεί να επιλυθεί αν γνωρίζουμε και το πρόσημο ενός εκ' των $\sin 2\tau$ και $\cos 2\tau$. Από τη σχέση (4.27) έχουμε:

$$\beta = \lambda_1^2(\lambda^2 - 1) \cos \tau \sin \tau = \frac{1}{2} \lambda_1^2(\lambda^2 - 1) \sin 2\tau. \quad (4.32)$$

Από εδώ προκύπτει πως $\text{sgn}(\sin 2\tau) = \text{sgn}(\beta/(\lambda^2 - 1))$ και επομένως η ασάφεια στην τιμή του 2τ αντιστοιχεί στις δύο λύσεις $\{\lambda'^2, 1/\lambda'^2\}$ για το λ^2 που βρήκαμε προηγουμένως. Επιλέγοντας $\lambda^2 = \lambda'^2$ μπορούμε να προσδιορίσουμε την ακριβή τιμή του 2τ στο διάστημα $[-\pi, \pi]$ από το πρόσημο του συντελεστή β και άρα θα έχουμε: $2\tau = 2\tau^* \pm 2k\pi$, $k = 0, 1, 2, \dots$. Επομένως, για τη γωνία (tilt) τ θα ισχύει:

$$\tau = \left(\frac{\tau^*}{2}\right) \pm k\pi, \quad k = 0, 1, 2, \dots \quad \text{με} \quad \tau \in [-\pi/2, \pi/2] \quad (4.33)$$

Εδώ προκύπτει πάλι μια ασάφεια $\pm\pi$, η οποία δεν μπορεί να επιλυθεί με βάση τις παραμέτρους του αφινικού μετασχηματισμού. Η ασάφεια αυτή της γωνίας τ μπορεί να αντιμετωπιστεί, όπως έχουμε ήδη δει, αξιοποιώντας τη σχετική θέση της μύτης στο πρόσωπο. Τέλος, επιλέγοντας $\lambda = \sqrt{\lambda'^2}$, η γωνία slant (σ) θα υπολογίζεται από τη σχέση:

$$\sigma = \text{arcsec}(\lambda) \quad (4.34)$$

Στο σημείο αυτό μπορούμε να προσδιορίσουμε και τις παραμέτρους (λ_1, θ) του αφινικού μετασχηματισμού. Γνωρίζοντας τις παραμέτρους (τ, σ) υπολογίζουμε τον πίνακα $\mathbf{P}(\lambda, \tau)$ μέσω της σχέσης (4.19). Τότε από την εξίσωση (4.14) θα έχουμε:

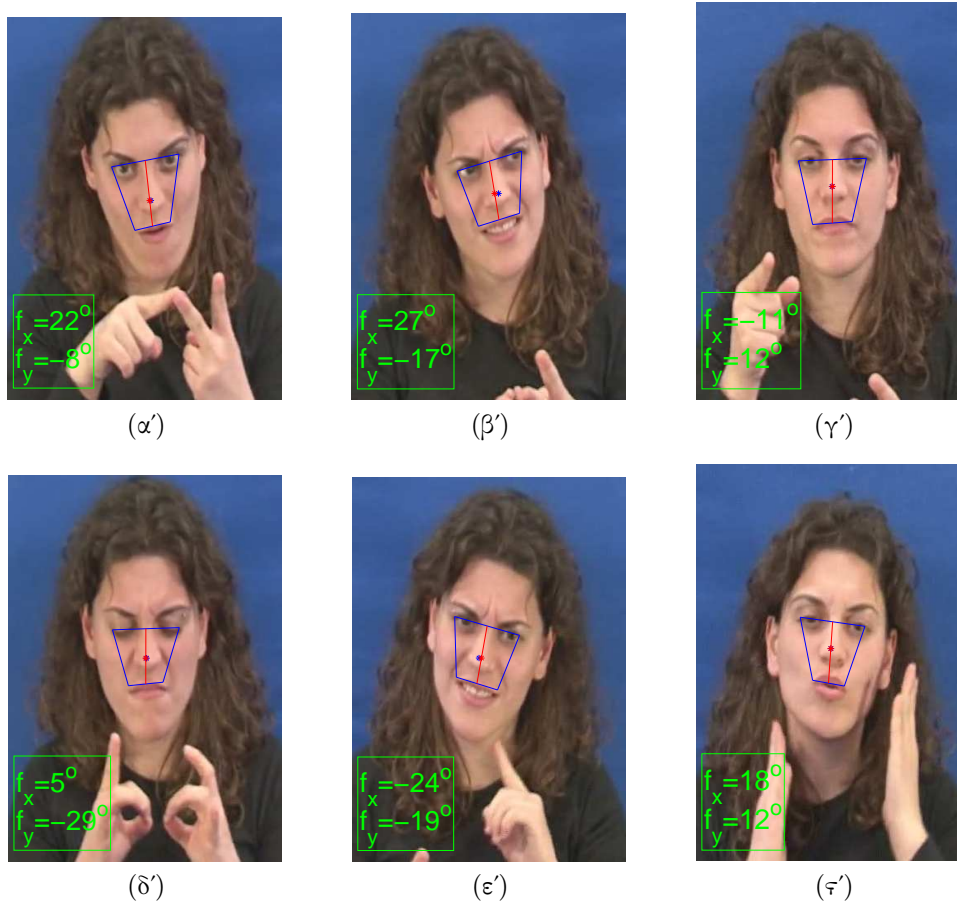
$$\begin{aligned} \lambda_1 \mathbf{R}(\theta) &= \mathbf{UP}^{-1}(\lambda, \tau) \Leftrightarrow \\ \begin{bmatrix} \lambda_1 \cos(\theta) & -\lambda_1 \sin(\theta) & 0 \\ \lambda_1 \sin(\theta) & \lambda_1 \cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix} &= \mathbf{UP}^{-1}(\lambda, \tau) \end{aligned} \quad (4.35)$$

Εξισώνοντας τα στοιχεία των δύο πινάκων παίρνουμε (όπου $\mathbf{UP}^{-1}[i, j]$ το (i, j) στοιχείο του γινομένου $\mathbf{UP}^{-1}(\lambda, \tau)$):

$$\frac{\lambda_1 \sin \theta}{\lambda_1 \cos \theta} = \frac{\mathbf{UP}^{-1}[2, 1]}{\mathbf{UP}^{-1}[1, 1]} \Leftrightarrow \theta = \arctan \left(\frac{\mathbf{UP}^{-1}[2, 1]}{\mathbf{UP}^{-1}[1, 1]} \right) \quad (4.36)$$

$$\text{και} \quad \lambda_1 = \frac{\mathbf{UP}^{-1}[1, 1]}{\cos \theta} \quad (4.37)$$

4.3.3 Πειραματικά αποτελέσματα



Σχήμα 4.8: Παραδείγματα εφαρμογής της μεθόδου υπολογισμού της πόζας κεφαλιού για τη βάση δεδομένων της ENΓ του IEL.

Στο σημείο αυτό, η μέθοδος εκτίμησης της πόζας κεφαλιού βάσει των γωνιών κάθετης (ϕ_x) και οριζοντίας (ϕ_y) στρέψης, που περιγράφηκε παραπάνω, εφαρμόζεται σε εικόνες από τις βάσεις δεδομένων της ENΓ του IEL και της BU400. Για τον υπολογισμό του λόγου R_e , θεωρήθηκε ότι το μέσο σχήμα του AAM αντιπροσωπεύει την ευθεία διεύθυνση της πόζας κεφαλιού.

Στο Σχήμα 4.8 παραθέτονται κάποια χαρακτηριστικά παραδείγματα από την εφαρμογή της μεθόδου στη βάση της ENΓ του IEL, προκειμένου να παρουσιαστούν τα αποτελέσματα της μεθόδου και να γίνει αξιολόγησή της με βάση την ανθρώπινη αντίληψη για την πόζα του κεφαλιού. Παρατηρούμε πως οι γωνίες στρέψης των παραπάνω εικόνων συμφωνούν σε γενικές γραμμές με την ανθρώπινη αίσθηση. Γενικές κατευθύνσεις της πόζας κεφαλιού, όπως το δεξιά/αριστερά και το πάνω/κάτω



Σχήμα 4.9: Παραδείγματα εφαρμογής της μεθόδου υπολογισμού της πόζας κεφαλιού για τη βάση δεδομένων BU400.

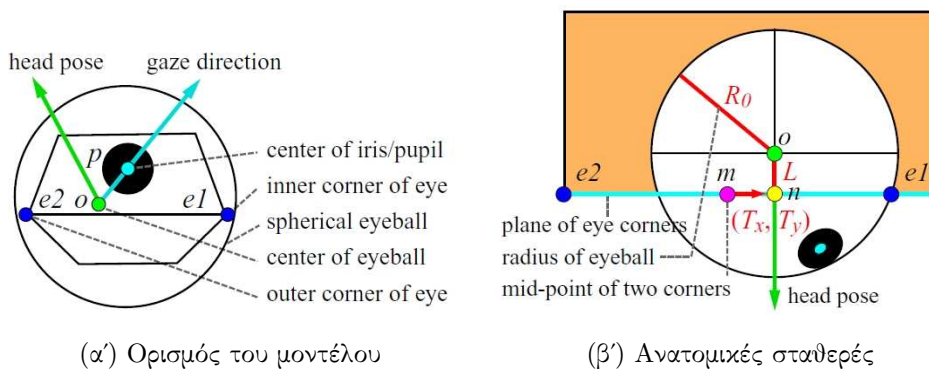
περιγράφονται αρκετά καλά. Επίσης, και η κατανομή των τιμών των γωνιών είναι αρκετά ρεαλιστική, παρόλο που μπορεί να υπάρχουν κάποια σχετικά μικρά σφάλματα της τάξης των 10° .

Στη συνέχεια, παρουσιάζονται και κάποια αποτελέσματα από τη βάση εικόνων BU400, η οποία εμφανίζει μεγαλύτερη μεταβλητότητα σχετικά με την πόζα κεφαλιού (Σχήμα 4.9). Και στην περίπτωση αυτή, παρατηρούμε πως το αποτέλεσμα που προκύπτει πειραματικά συμβαδίζει με την ανθρώπινη διαίσθηση, μιας και οι γενικές κατευθύνσεις της περιστροφής του κεφαλιού περιγράφονται αρκετά ικανοποιητικά από τις γωνίες ϕ_x, ϕ_y .

Ωστόσο, η συγκεκριμένη αναλυτική μέθοδος εκτίμησης της πόζας είναι αρκετά ευαίσθητη σε μεταβολές των σημείων που ορίζουν το τραπέζιο. Μικρές αποκλίσεις στις εκτιμήσεις των θέσεων των σημείων αυτών μπορεί να οδηγήσει σε αρκετά μεγάλα σφάλματα στις γωνίες στέψης, άρα και σε αποτυχία της μεθόδου. Επιπλέον, υπάρχουν και αλλαγές στην πόζα του κεφαλιού που δεν μπορούν να περιγραφούν από τις

παραμορφώσεις του τραπεζίου. Μια πιο σύνθετη και εύρωστη μέθοδος εκτίμησης θα μπορούσε να λύσει τα προβλήματα αυτά, ωστόσο στη συγκεκριμένη περίπτωση που η πόζα λειτουργεί συμπληρωματικά στην εκτίμηση του gaze η παρούσα μέθοδος κρίνεται αρκετή.

4.4 Γεωμετρικό μοντέλο εκτίμησης του Eye Gaze



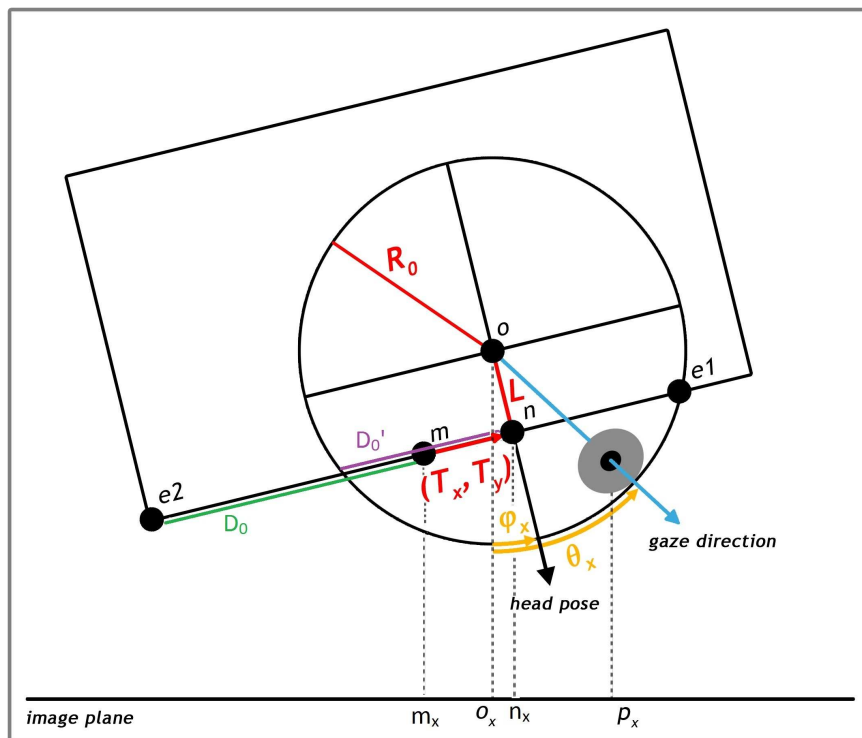
Σχήμα 4.10: Σφαιρικό γεωμετρικό μοντέλο για την εκτίμηση του eye gaze. Έχει ληφθεί από το [15].

4.4.1 Περιγραφή του μοντέλου

Στην ενότητα αυτή, θα γίνει περιγραφή ενός γεωμετρικού μοντέλου για την εκτίμηση της κατεύθυνσης του gaze, συνδυάζοντας και τη διαθέσιμη πληροφορία για την πόζα κεφαλιού.[15] Το συγκεκριμένο μοντέλο υποθέτει ότι το σχήμα του βολβού του ματιού είναι σφαιρικό, ενώ η εσωτερική (e1) και εξωτερική (e2) γωνία του κάθε ματιού έχουν ήδη εκτιμηθεί από τα σημεία της μάσκας σχήματος του AAM (Σχήμα 4.10α').

Ο αλγόριθμος υπολογισμού της διεύθυνσης του gaze μπορεί να διαιρεθεί σε δύο επιμέρους στάδια. Στην πρώτη φάση γίνεται η εκτίμηση του κέντρου της σφαίρας του ματιού πάνω στην εικόνα, αξιοποιώντας τις θέσεις των γωνιών του ματιού και την πόζα του κεφαλιού. Επίσης, είναι απαραίτητο να είναι γνωστές και κάποιες ανατομικές παράμετροι του ματιού (Σχήμα 4.10β'), οι οποίες μπορούν να έχουν προσδιοριστεί σε κάποιο προηγούμενο στάδιο. Στη συνέχεια γίνεται ο υπολογισμός των γωνιών που ορίζουν την κατεύθυνση του gaze, χρησιμοποιώντας τη θέση της ίριδας καθώς και το κέντρο της σφαίρας του ματιού, που υπολογίστηκε στο προηγούμενο στάδιο.

ΚΕΦΑΛΑΙΟ 4. ΕΚΤΙΜΗΣΗ ΤΟΥ EYE GAZE ΑΠΟ
78 ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΣΗΜΕΙΑ ΤΟΥ ΠΡΟΣΩΠΟΥ ΚΑΙ ΤΩΝ ΜΑΤΙΩΝ



Σχήμα 4.11: Γεωμετρική μέθοδος υπολογισμού του eye gaze (κάτοψη του ματιού).

Εκτίμηση του κέντρου της σφαίρας ματιού και της διεύθυνσης του gaze

Αρχικά, θεωρούμε ότι οι ανατομικές σταθερές που χρειάζονται είναι ήδη γνωστές και επιπλέον έχει γίνει και η αναγωγή τους στην κλίμακα της εικόνας. Σε επόμενη ενότητα θα περιγραφούν οι τρόποι με τους οποίους μπορεί να γίνει αυτή η διαδικασία. Οι ανατομικές σταθερές που θα χρησιμοποιηθούν, όπως φαίνεται και στο Σχήμα 4.10β' θα είναι:

- R_0 : ακτίνα της σφαίρας του ματιού
- (T_x, T_y) : σταθερή απόσταση πάνω στην εικόνα μεταξύ του μέσου των δύο γωνιών του ματιού και του κέντρου της σφαίρας ματιού
- L : το βάθος στο οποίο βρίσκεται το κέντρο του ματιού σε σχέση με το επίπεδο που περιέχει τις δύο γωνίες του ματιού

Το μέσο σημείο (m_x, m_y) του ευθυγράμμου τμήματος που ενώνει την εσωτερική $(e1_x, e1_y)$ και την εξωτερική $(e2_x, e2_y)$ γωνία του ματιού θα δίνεται από τη σχέση:

$$\begin{pmatrix} m_x \\ m_y \end{pmatrix} = \begin{pmatrix} \frac{e1_x + e2_x}{2} \\ \frac{e1_y + e2_y}{2} \end{pmatrix} \quad (4.38)$$

Για να υπολογίσουμε την προβολή του κέντρου της σφαίρας ματιού (o_x, o_y) πάνω στο επίπεδο της εικόνας θα πρέπει να προσθέσουμε στο μέσο σημείο (m_x, m_y) δύο όρους διόρθωσης, όπως φαίνεται και στην τοπολογία του Σχήματος 4.11. Ο πρώτος όρος περιέχει ένα σταθερό διάνυσμα (T_x, T_y) , το οποίο εκφράζει το γεγονός πως ακόμα και στην ευθεία διεύθυνση του κεφαλιού το μέσο των γωνιών δεν ταυτίζεται με την προβολή του κέντρου του ματιού. Αν λάβουμε υπόψιν και την πόζα του κεφαλιού (ϕ_x, ϕ_y) τότε η διορθωμένη εκτίμηση (n_x, n_y) του κέντρου του ματιού θα είναι:

$$\begin{pmatrix} n_x \\ n_y \end{pmatrix} = \begin{pmatrix} m_x \\ m_y \end{pmatrix} + \begin{pmatrix} T_x \cos \phi_x \\ T_y \cos \phi_y \end{pmatrix} \quad (4.39)$$

Ο δεύτερος όρος αντισταθμίζει το σφάλμα που προκύπτει από το γεγονός ότι το κέντρο του ματιού δεν βρίσκεται κατά ανάγκη στο επίπεδο που ορίζουν οι γωνίες του ματιού. Επομένως, τελική εκτίμηση (o_x, o_y) για τη θέση του κέντρου του ματιού πάνω στην εικόνα θα δίνεται:

$$\begin{pmatrix} o_x \\ o_y \end{pmatrix} = \begin{pmatrix} m_x \\ m_y \end{pmatrix} + \begin{pmatrix} T_x \cos \phi_x \\ T_y \cos \phi_y \end{pmatrix} + L \begin{pmatrix} -\sin \phi_x \\ -\sin \phi_y \end{pmatrix} \quad (4.40)$$

Στην παραπάνω εκτίμηση έχουμε συμπεριλάβει και πληροφορία για την πόζα κεφαλιού, η οποία θα πρέπει να έχει εξαχθεί από ένα άλλο εξωτερικό σύστημα σε προηγούμενο στάδιο. Εάν η εκτίμηση της πόζας δεν είναι διαθέσιμη ή η ακρίβεια της δεν είναι ικανοποιητική, τότε μπορούμε να μην τη χρησιμοποιήσουμε στο γεωμετρικό μοντέλο για το gaze. Στην περίπτωση αυτή ο υπολογισμός του gaze θα βασιστεί μόνο στη σχετική θέση ίριδας και κέντρου ματιού (local gaze) και η εξίσωση (4.40) γίνεται:

$$\begin{pmatrix} o_x \\ o_y \end{pmatrix} = \begin{pmatrix} m_x \\ m_y \end{pmatrix} + \begin{pmatrix} T_x \\ T_y \end{pmatrix} \quad (4.41)$$

Παρατηρούμε πως αν δεν γίνει χρήση της πόζας κεφαλιού μόνο ο πρώτος όρος διόρθωσης χρειάζεται για να προσδιοριστεί η σωστή θέση του κέντρου του ματιού. Στη συνέχεια, οι γωνίες διεύθυνσης του gaze (θ_x, θ_y) μπορούν να υπολογιστούν βάσει του Σχήματος 4.11 από τη σχέση:

$$\begin{pmatrix} \theta_x \\ \theta_y \end{pmatrix} = \begin{pmatrix} \arcsin \left(\frac{p_x - o_x}{\sqrt{R_0^2 - (p_y - o_y)^2}} \right) \\ \arcsin \left(\frac{p_y - o_y}{\sqrt{R_0^2 - (p_x - o_x)^2}} \right) \end{pmatrix} \quad (4.42)$$

Η θέση της ίριδας (p_x, p_y) μπορεί να προσδιοριστεί είτε από τη μάσκα σχήματος του AAM είτε με τη μέθοδο μορφολογικού φιλτραρίσματος που περιγράφηκε στην πρώτη ενότητα του κεφαλαίου.

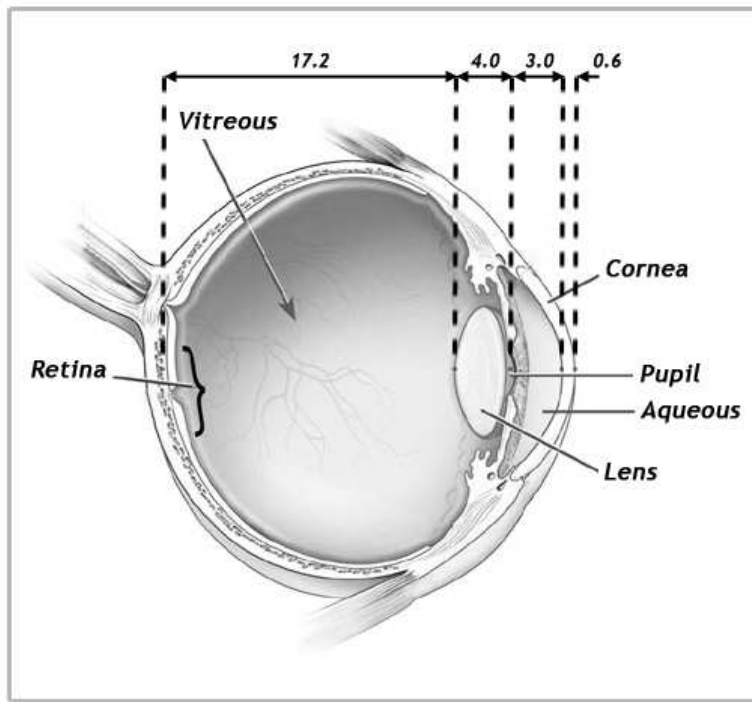
4.4.2 Προσδιορισμός των ανατομικών σταθερών

Στο σημείο αυτό θα περιγράψουμε τους τρόπους προσδιορισμού των ανατομικών σταθερών, οι οποίες θεωρήθηκαν γνωστές σε όλη την προηγούμενη ανάλυση. Ένας τρόπος είναι να υπολογιστούν οι σταθερές αυτές σε ένα στάδιο προεπεξεργασίας μέσα από μια διαδικασία εκπαίδευσης.[15] Για το σκοπό αυτό είναι απαραίτητο να διαθέτουμε ένα σύνολο εικόνων εκπαίδευσης με προκαθορισμένες κατευθύνσεις για την πόζα κεφαλιού και του gaze, ενώ οι σταθερές R_0, T_x, T_y, L υπολογίζονται με προσέγγιση ελαχίστων τετραγώνων. Οι τιμές που προκύπτουν κανονικοποιούνται έτσι ώστε να αντιστοιχούν σε μοναδιαία κλίμακα προσώπου. Για να πάρουμε τις τιμές ανηγμένες σε μια διαφορετική κλίμακα αρκεί να πολλαπλασιάσουμε κάθε μια σταθερά με τον συντελεστή κλιμάκωσης S . Ο συντελεστής S μπορεί να υπολογιστεί από τις παραμέτρους του AAM αλλά και από την απόσταση μεταξύ της εσωτερικής και της εξωτερικής γωνίας του ματιού:

$$S = \frac{\sqrt{(e1_x - e2_x)^2 + (e1_y - e2_y)^2}}{\cos \phi_x} \quad (4.43)$$

Παρόλο που η παραπάνω διαδικασία μπορεί να αποδώσει τις τιμές των ανατομικών σταθερών με αρκετή ακρίβεια, δεν είναι ξεκάθαρο αν είναι όντως απαραίτητη αλλά και αποδοτική μια τέτοια διαδικασία εκπαίδευσης σε ένα πρόβλημα εκτίμησης στεθερών ποσοτήτων. Ένας άλλος περιορισμός που υπάρχει στην όλη διαδικασία είναι αυτός της δημιουργίας του συνόλου εκπαίδευσης, μιας και δεν είναι πάντα βολικό να ζητηθεί από τον χρήστη η συμμετοχή του σε μια βαθμονόμηση του συστήματος με αρκετά σημεία. Επιπλέον, σε πραγματικά δεδομένα, όπως τα βίντεο της νοηματικής γλώσσας, υπάρχει και το πρόβλημα πως η καταγραφή τους έχει γίνει σε προγενέστερη χρονική στιγμή. Επομένως δεν υπάρχει η δυνατότητα να πάρουμε εικόνες με προκαθορισμένες κατευθύνσεις πόζας και gaze, αν δεν είχε γίνει στο παρελθόν πρόβλεψη για μια τέτοια βαθμονόμηση.

Μια άλλη μέθοδος που δεν αντιμετωπίζει τα παραπάνω προβλήματα είναι αναζητήσουμε μέσες τιμές για τις ανατομικές παραμέτρους R_0, L στην ιατρική βιβλιογραφία. Ωστόσο, στην περίπτωση αυτή υπάρχει το πρόβλημα ότι οι παράμετροι του ματιού διαφέρουν από άτομο σε άτομο και επηρεάζονται από παραγόντες όπως η ηλικία, η εθνικότητα, το γένος, η κατάσταση της υγείας κτλ. Στο Σχήμα 4.12 παρουσιάζονται κάποιες τυπικές τιμές των παραμέτρων του ανθρώπινου ματιού για έναν μέσο ενήλικα άνθρωπο. Από το σχήμα αυτό μπορούμε να θεωρήσουμε ότι η ακτίνα της σφαίρας του ματιού είναι $R_0 = 12.4mm$, ενώ το βάθος μεταξύ του



Σχήμα 4.12: Τυπικές τιμές παραμέτρων (διαστάσεις σε mm) του ανθρώπινου ματιού για έναν μέσο ενήλικα άνθρωπο. Έχει ληφθεί από το διαδίκτυο.

κέντρου του ματιού και του επιπέδου που ορίζεται από τις γωνίες του ματιού θα είναι $L = 4.5mm$. Από τις τιμές αυτές μπορεί να υπολογιστεί η απόσταση D'_0 μεταξύ σημείων τομής της σφαίρας ματιού με το επίπεδο των γωνιών και της προβολής του κέντρου της σφαίρας ματιού στο επίπεδο αυτό (Σχήμα 4.11): $D'_0 = \sqrt{R_0^2 - L^2}$. Στη συνέχεια βρίσκουμε το λόγο $LD = \left(\frac{L}{D'_0}\right)$ που σχετίζει την ανατομική στεθερά L με την απόσταση D'_0 .

Γνωρίζοντας, τις τιμές των παραμέτρων R_0, L οι ανατομικές σταθερές T_x, T_y μπορούν να προσδιοριστούν από ένα απλό αρχικό στάδιο βαθμονόμησης, κατά το οποίο ζητείται από το χρήστη να κοιτάζει ευθεία στο φακό της κάμερας κρατώντας τη θέση του κεφαλιού του παράλληλη στο επίπεδο της εικόνας. Σε περίπτωση που δεν μπορεί να πραγματοποιηθεί ούτε αυτή η απλή διαδικασία βαθμονόμησης, μπορεί να θεωρηθεί ότι το μέσο σχήμα του AAM αντιπροσωπεύει με κάποια σχετική προσέγγιση την ευθεία διεύθυνση τόσο του gaze όσο και της πόζας. Τότε οι γωνίες κατευθύνσης του gaze και της πόζας κεφαλιού θα είναι μηδενικές $((\theta_x, \theta_y) = (0, 0), (\phi_x, \phi_y) = (0, 0))$. Οι παράμετροι T_x, T_y , κανονικοποιημένες ως

προς την απόσταση $D_0 = S/2$, βάσει των σχέσεων (4.40) και (4.42), θα δίνονται:

$$T_x = \frac{p_x - m_x - L \sin \phi_x}{D_0 \cos \phi_x} = \frac{p_x - m_x}{D_0} \quad (4.44)$$

$$T_y = \frac{p_y - m_y - L \sin \phi_y}{D_0 \cos \phi_y} = \frac{p_y - m_y}{D_0} \quad (4.45)$$

Τέλος, για τον προσδιορισμό των ανατομικών σταθερών σε μια διαφορετική κλίμακα της εικόνας ακολουθείται η παρακάτω διαδικασία:

- Αρχικά υπολογίζεται η απόσταση $D_0 = S/2$ βάσει της σχέσης (4.43) και προσδιορίζονται οι σταθερές T_x, T_y στην κλίμακα της εικόνας ($T_x = D_0 T_x$ και $T_y = D_0 T_y$).

- Υπολογίζεται η απόσταση D'_0 πάνω στην νέα εικόνα βάσει της σχέσης:

$$D'_0 = \frac{\sqrt{(e1_x - (e2_x + 2T_x \cos \phi_x))^2 + ((e1_y + 2T_y \cos \phi_y) - (e2_y + 2T_y \cos \phi_y))^2}}{\cos \phi_x} \quad (4.46)$$

- Υπολογίζεται η τιμή της σταθεράς L στην κλίμακα της εικόνας πολλαπλασιάζοντας την παραπάνω απόσταση με τον λόγο LD : $L = LD \cdot D'_0$.
- Τέλος, υπολογίζεται η ακτίνα της σφαίρας του ματιού για τη νέα κλίμακα της εικόνας χρησιμοποιώντας τη σχέση:

$$R_0 = \sqrt{(D'_0)^2 + L^2} \quad (4.47)$$

4.4.3 Πειραματικά αποτελέσματα

Συνοψίζοντας, για την εφαρμογή του παραπάνω γεωμετρικού μοντέλου εκτίμησης του eye gaze είναι απαραίτητο να είναι διαθέσιμες πληροφορίες για τα παρακάτω στοιχεία:

- Θέσεις των εσωτερικών ($e1^{right}, e1^{left}$) και εξωτερικών γωνιών ($e2^{right}, e2^{left}$), οι οποίες προσδιορίζονται από τα αντίστοιχα σημεία του γράφου σχήματος του τοπικού AAM για τα μάτια.
- Θέση του κέντρου (p_x, p_y) της ίριδας, που μπορεί να εκτιμηθεί τόσο από το σχήμα του AAM όσο και με τη χρήση της τεχνικής μορφολογικού φιλτραρίσματος της ενότητας 4.2.2.
- Εκτίμηση της πόζας κεφαλιού (ϕ_x, ϕ_y) μέσω της τεχνικής της ενότητας 4.3.
- Προσδιορισμός των ανατομικών σταθερών R_0, T_x, T_y, L με βάση τη μέθοδο της ενότητας 4.4.2.



Σχήμα 4.13: Παραδείγματα εφαρμογής του γεωμετρικού μοντέλου εκτίμησης του eye gaze, συμπεριλαμβανομένης της πόζας κεφαλιού, για τη βάση δεδομένων της ENG του IEL.

Το συγκεκριμένο γεωμετρικό μοντέλο, λοιπόν, εφαρμόζεται σε εικόνες της νοηματικής γλώσσας από τη βάση της ENG του IEL με στόχο την εκτίμηση της κατεύθυνσης του βλέματος του νοηματιστή. Η πληροφορία αυτή είναι ιδιαίτερα σημαντική και μπορεί να χρησιμοποιηθεί σε συστήματα αυτόματης αναγνώρισης του νοηματικού λόγου. Για την εκτίμηση των ανατομικών σταθερών θεωρήθηκε ότι το μέσο σχήμα του AAM αντιστοιχεί στην ευθεία διεύθυνση του gaze και της πόζας κεφαλιού.

Στο Σχήμα 4.13 παρουσιάζονται κάποια ενδεικτικά αποτελέσματα από την εφαρμογή της μεθόδου, στην οποία χρησιμοποιείται εξωτερική πληροφορία για την πόζα κεφαλιού ενώ η θέση της ίριδας προσδιορίζεται από το αντίστοιχο σημείο στο γράφο σχήματος του AAM. Παρατηρούμε ότι σε γενικές γραμμές το αποτέλεσμα που προκύπτει από αυτή την αυτόματη μέθοδο συμφωνεί με αυτό που προκύπτει μετά από ανθρώπινη παρατήρηση. Τόσο οι γωνίες στρέψης θ_x, θ_y όσο και το διαισθητικό αποτέλεσμα, σχεδιάζοντας τα διανύσματα κατεύθυνσης πάνω στην εικόνα, δίνουν μια



Σχήμα 4.14: Παραδείγματα εφαρμογής του γεωμετρικού μοντέλου εκτίμησης του eye gaze, χωρίς πληροφορία για την πόζα κεφαλιού, για τη βάση δεδομένων της ENΓ του IEL.

αρκετά ικανοποιητική και ακριβής εκτίμηση για το eye gaze. Αξίζει να σημειωθεί πως η μέθοδος καταφέρνει να δώσει αληθοφανή αποτελέσματα και σε περιπτώσεις που τα μάτια είναι κλειστά (Σχήμα 4.13γ') ή σχεδόν κλειστά (Σχήμα 4.13ε'). Αυτό συμβαίνει γιατί παρόλο που η ίριδα στις περιπτώσεις αυτές δεν είναι εμφανής, η θεωρητικά αναμενόμενη θέση της μπορεί να εκτιμηθεί από το AAM, εάν έχουν συμπεριληφθεί στο σύνολο εκπαίδευσης ανάλογες εικόνες. Τέλος, η μέθοδος εφαρμόζεται σε κάθε ένα μάτι ξεχωριστά, δίνοντας έτσι διαφορετική εκτίμηση του gaze για κάθε μάτι. Ωστόσο, όπως φαίνεται και στα παραδείγματα, οι διαφορές μεταξύ των δύο αυτών εκτιμήσεων είναι πολύ μικρές (της τάξης των $0^\circ - 3^\circ$), γεγονός που συμβαδίζει και με την ανθρώπινη λογική.

Στο Σχήμα 4.14 παρουσιάζεται το αποτέλεσμα από την εφαρμογή της μεθόδου στα ίδια παραδείγματα από τη βάση της ENΓ του IEL, μόνο που τώρα στο γεωμετρικό



Σχήμα 4.15: Παραδείγματα εφαρμογής του γεωμετρικού μοντέλου εκτίμησης του eye gaze, με εύρεση της θέσης της ίριδας χρησιμοποιώντας μορφολογικό φιλτράρισμα, για τη βάση δεδομένων της ENG του IEL.

μοντέλο δεν έχει συμπεριληφθεί η πληροφορία για την πόζα κεφαλιού. Έτσι, η εκτίμηση που λαμβάνουμε αφορά μόνο το τοπικό (local) gaze, το οποίο περιγράφεται από την περιστροφή της σφαίρας του ματιού και τη σχετική θέση της ίριδας. Παρατηρούμε πως αν η προσοχή μας επικεντρωθεί μόνο στην περιοχή των ματιών, τότε τα αποτελέσματα της μεθόδου συμφωνούν αρκετά με την ανθρώπινη αντίληψη. Ωστόσο σε σχέση με πριν, σημειώνονται αρκετές διαφορές στην κατεύθυνση gaze. Σε κάποιες περιπτώσεις η πόζα κεφαλιού συμφωνεί με το local gaze ενισχύοντας την υπάρχουσα εκτίμηση, ενώ σε άλλες μπορεί να είναι αντίθετη δίνοντας έτσι αρκετά διαφορετική εκτίμηση για το γενικό (global) gaze. Από αυτό συμπεραίνουμε, πως η εκτίμηση του eye gaze μόνο από την περιοχή των ματιών δεν οδηγεί πάντα σε σωστά αποτελέσματα. Ωστόσο σε περιπτώσεις που η μεταβλητότητα της πόζας κεφαλιού είναι μικρή ή η εκτίμηση της δεν είναι ακριβής, τότε η μέθοδος εκτίμησης χωρίς χρήση της πόζας μπορεί να προτιμηθεί για να δώσει καλύτερα αποτελέσματα.

ΚΕΦΑΛΑΙΟ 4. ΕΚΤΙΜΗΣΗ ΤΟΥ EYE GAZE ΑΠΟ 86 ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΣΗΜΕΙΑ ΤΟΥ ΠΡΟΣΩΠΟΥ ΚΑΙ ΤΩΝ ΜΑΤΙΩΝ

Τέλος, στο Σχήμα 4.15 παρουσιάζονται οι ίδιες εικόνες από την εφαρμογή του γεωμετρικού μοντέλου στη βάση της ENΓ του IEL, χρησιμοποιώντας όμως για τον προσδιορισμό της θέσης της ίριδας τη μέθοδο μορφολογικού φιλτραρίσματος της ενότητας 4.2.2 και όχι το σημείο του γράφου του AAM. Εδώ παρουσιάζονται εμφανείς διαφορές και αποκλίσεις της εκτίμησης του gaze σε σχέση με τα θεωρητικά αναμενόμενα αποτελέσματα σύμφωνα με την ανθρώπινη αντίληψη. Σε αρκετές περιπτώσεις η μέθοδος αποτυγχάνει να εκτιμήσει έστω και τη γενική διεύθυνση του gaze, ενώ και στις υπόλοιπες υπάρχουν αρκετά σφάλματα κυρίως στην κάθετη διεύθυνση. Αυτό οφείλεται στην λανθασμένη εκτίμηση του κέντρου της ίριδας με τη μέθοδο μορφολογικού φιλτραρίσματος. Σε περιπτώσεις που η εκτίμηση για την είναι αρκετά ικανοποιητική τότε και τα αποτελέσματα για το gaze είναι αρκετά καλά (Σχήματα 4.15β' & 4.15δ').

Εκεί, όμως, που η μέθοδος αποτυγχάνει εντελώς είναι στις περιπτώσεις που η ίριδα καλύπτεται σε μεγάλο βαθμό από τα βλέφαρα, αφού το αποτέλεσμα που προκύπτει μπορεί να είναι τελείως τυχαίο (Σχήμα 4.15ε') ή να μην μπορεί να δοθεί καμία εκτίμηση (Σχήμα 4.15γ'). Αυτό οφείλεται στο γεγονός ότι εάν η ίριδα δεν φαίνεται στην εικόνα τότε η θέση της δεν μπορεί να εκτιμηθεί με το μορφολογικό φιλτράρισμα, σε αντίθεση με το AAM το οποίο χρησιμοποιεί προηγούμενη πληροφορία από ανάλογες εικόνες. Τέλος, παρατηρούνται και σημαντικές αποκλίσεις μεταξύ των δύο ματιών μιας και ο υπολογισμός για τη θέση της ίριδας γίνεται ξεχωριστά για τα δύο μάτια, ενώ στο AAM ο υπολογισμός των σημείων του γράφου του σχήματος γίνεται συνδυασμένα για όλο το πρόσωπο.

Από όλα τα παραπάνω, συμπεραίνουμε ότι είναι σαφώς προτιμότερο να χρησιμοποιήσουμε σαν θέση της ίριδας το σημείο που προκύπτει από την προσαρμογή του AAM παρά την εκτίμηση μέσω φιλτραρίσματος, μιας και όπως είδαμε η μέθοδος αυτή υπερτερεί συνολικά ενώ δίνει αποδεκτά αποτελέσματα ακόμα και σε “δύσκολες” περιπτώσεις.

4.5 Συμπεράσματα

Στο συγκεκριμένο κεφάλαιο αναφέρθηκαν μέθοδοι εκτίμησης του eye gaze και της πόζας κεφαλιού οι οποίες στηρίζονται στον προσδιορισμό χαρακτηριστικών σημείων του προσώπου πάνω στην εικόνα. Αρκετές από αυτές τις τεχνικές έδωσαν αρκετά ικανοποιητικά και ακριβή αποτελέσματα. Πιο συγκεκριμένα, μέθοδοι που βασίζονται στα AAMs, τα οποία μοντελοποιούν ολόκληρη την περιοχή του προσώπου, οδήγησαν σε αρκετά καλύτερα αποτελέσματα από τις “τοπικές” τεχνικές. Ωστόσο, οι συγκεκριμένες μέθοδοι επηρεάζονται σε πολύ μεγάλο βαθμό από τις εκτιμήσεις των σημείων του προσώπου που χρησιμοποιούν και έτσι δεν εμφανίζονται αρκετά εύρωστες σε εξωτερικούς παράγοντες, όπως μια μεγάλη στρέψη του κεφαλιού ή επικαλύψεις του προσώπου από τα χέρια.

Επειδή στις παραπάνω τεχνικές ο προσδιορισμός των απαιτούμενων σημείων γίνεται με χρήση AAM, οι παραπάνω περιορισμοί μπορούν να αντιμετωπιστούν σε ένα βαθμό σχεδιάζοντας καλύτερους αλγορίθμους προσαρμογής των AAMs χρησιμοποιώντας πιο εύρωστες νόρμες σφάλματος.[10] Επιπλέον, τα προβλήματα που αναφέρθηκαν μπορεί να αντιμετωπιστούν χρησιμοποιώντας, όπως θα δούμε στο επόμενο κεφάλαιο, μια διαφορετική προσέγγιση για την εκτίμηση του eye gaze, η οποία θα στηρίζεται στη συνολική εμφάνιση της περιοχής των ματιών και του προσώπου και όχι σε συγκριμένα σημεία και θα χρησιμοποιεί τεχνικές αναγνώρισης προτύπων αντί για κάποιο γεωμετρικό μοντέλο.

ΚΕΦΑΛΑΙΟ 4. ΕΚΤΙΜΗΣΗ ΤΟΥ EYE GAZE ΑΠΟ
88 ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΣΗΜΕΙΑ ΤΟΥ ΠΡΟΣΩΠΟΥ ΚΑΙ ΤΩΝ ΜΑΤΙΩΝ

Κεφάλαιο 5

Εκτίμηση του Eye Gaze με κατηγοριοποίησή του σε κλάσεις

5.1 Εισαγωγή

Στο προηγούμενο κεφάλαιο αναφερθήκαμε σε μεθόδους που βασίζονται στην εκτίμηση χαρακτηριστικών σημείων πάνω στο πρόσωπο, όπως η ίριδα και οι γωνίες των ματιών προκειμένου να υπολογίσουν τη διεύθυνση του eye gaze. Μια άλλη προσέγγιση του προβλήματος, η οποία θα παρουσιαστεί στο συγκεκριμένο κεφάλαιο, είναι να προσπαθήσουμε να εκτιμήσουμε το gaze συνδυάζοντας όλη την πληροφορία για την εμφάνιση γύρω από την περιοχή των ματιών, η οποία θα εμπεριέχει στοιχεία και για το σχήμα και για την υφή. Με τον τρόπο αυτό, το πρόβλημα προσδιορισμού του gaze ανάγεται σε ένα πρόβλημα αναγνώρισης προτύπων, για την επίλυση του οποίου μπορεί να χρησιμοποιηθεί οποιαδήποτε τεχνική κατηγοριοποίησης σε κλάσεις.

Σύμφωνα με τα παραπάνω, η εκτίμηση του eye gaze μπορεί να χωριστεί σε δύο επιμέρους στάδια: τον προσδιορισμό του διανύσματος με τις παραμέτρους που θα περιγράφουν την περιοχή των ματιών και τη χρησιμοποίηση του κατάλληλου ταξινομητή για την κατηγοριοποίηση της διεύθυνσης του gaze σε προκαθορισμένες κλάσεις. Για το λόγο αυτό, αρχικά θα πρέπει να δημιουργηθεί “χειροκίνητα” ένα σύνολο εικόνων με επισημειωμένες τις κατευθύνσεις του gaze, προκειμένου να γίνει η εκπαίδευση του ταξινομητή. Όσον αφορά την επιλογή του ταξινομητή, υπάρχουν αρκετά διαφορετικά είδη που μπορούν να χρησιμοποιηθούν, όπως Support Vector Machines (SVM), Νευρωνικά Δίκτυα και Μοντέλα Μείγματος Γκαουσιανών (GMMs), που είναι και αυτά που θα χρησιμοποιηθούν στη παρούσα διπλωματική.

Όσον αφορά το διάνυσμα των παραμέτρων που θα δοθεί σαν είσοδος στο ταξινομητή, στη βιβλιογραφία υπάρχουν αρκετοί διαφορετικοί περιγραφείς για την εμφάνιση μιας εικόνας. Από αυτούς ξεχωρίζουν τα AAMs, τα οποία έχουν αναλυθεί σε προηγούμενο κεφάλαιο, και τα Ιστογράμματα Προσανατολισμένης Κλίσης

(Histograms Oriented Gradients “HOGs”), τα οποία έχουν χρησιμοποιηθεί με επιτυχία στην αναγνώριση αντικειμένων σε εικόνες.[7, 8] Στο κεφάλαιο που ακολουθεί θα γίνει εφαρμογή δύο διαφορετικών μεθόδων για την εκτίμηση του eye gaze μέσω κατηγοριοποίησης σε διακριτές κλάσεις. Η μία στηρίζεται στη χρήση των παραμέτρων του τοπικού AAM για τα μάτια, ενώ η άλλη στην κατασκευή HOGs για την εικόνα της περιοχής των ματιών. Και στις δύο τεχνικές, περιλαμβάνεται η χρήση GMMs σαν μεθόδου κατηγοριοποίησης. Τέλος, το κεφάλαιο ολοκληρώνεται με την παρουσίαση ενός αλγορίθμου που αντιστοιχίζει τις πιθανοφάνειες κάθε κλάσης με τις γωνίες κάθετης και οριζόντιας διεύθυνσης του eye gaze.

5.2 Κατηγοριοποίηση με βάση το τοπικό AAM

5.2.1 Περιγραφή των GMMs και του αλγορίθμου EM

Όπως ήδη αναφέρθηκε, τα GMMs (Gaussian Mixture Models) αποτελούν το βασικό ταξινομητή που θα χρησιμοποιηθεί σε όλες τις μεθόδους κατηγοριοποίησης του gaze που θα αναπτυχθούν στη συνέχεια του κεφαλαίου. Γι’ αυτό κρίνεται σκόπιμο να γίνει πρώτα μια σύντομη και περιεκτική περιγραφή της δομής τους αλλά και τρόπου υπολογισμού των παραμέτρων τους, μέσω του αλγορίθμου Expectation Maximization (EM).[3]

Η βασική ιδέα των GMMs είναι ότι ένα σύνολο δεδομένων \mathbf{x} μπορεί να αναπαρασταθεί από ένα μείγμα K γκαουσιανών συναρτήσεων $\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ μέσης τιμής $\boldsymbol{\mu}_k$ και μεταβλητότητας $\boldsymbol{\Sigma}_k$:

$$p(\mathbf{x}) = \sum_{k=1}^K \pi_k \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \quad (5.1)$$

Ο όρος π_k αντιπροσωπεύει το βάρος της κάθε γκαουσιανής και ικανοποιεί τους περιορισμούς:

$$0 \leq \pi_k \leq 1, \quad \sum_{k=1}^K \pi_k = 1 \quad (5.2)$$

Στη συνέχεια, μπορεί να υπολογιστεί η “ευθύνη” (responsibility) της γκαουσιανής $\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ σαν το ποσοστό της συνολικής πληροφορίας των δεδομένων \mathbf{x} που μπορεί να εξηγηθεί χρησιμοποιώντας μόνο την k γκαουσιανή συνιστώσα του μοντέλου:

$$\gamma_k(\mathbf{x}) = \frac{\pi_k \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{j=1}^K \pi_j \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)} \quad (5.3)$$

Έστω τώρα ότι διαθέτουμε N διανύσματα δεδομένων εκπαίδευσης \mathbf{x}_n ($\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N]$), των οποίων τη μεταβλητότητα θέλουμε να περιγράψουμε με ένα GMM. Τότε, οι βέλτιστες παράμετροι $\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\pi}$ του μοντέλου θα προκύπτουν από τη μεγιστοποίηση της συνολικής λογαριθμικής πιθανοφάνειας (log likelihood):

$$\ln(p(\mathbf{X}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\pi})) = \sum_{n=1}^N \ln \left\{ \sum_{k=1}^K \pi_k \mathcal{N}(\mathbf{x}_n | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \right\} \quad (5.4)$$

Για την επίλυση του παραπάνω προβλήματος μπορεί να χρησιμοποιηθεί οποιαδήποτε αριθμητική μέθοδος βελτιστοποίησης ή ο αλγόριθμος Expectation-Maximization (EM)[22]:

Αλγόριθμος 1 Expectation-Maximization (EM)

1. Αρχικοποίηση των παραμέτρων $\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k$ υπολογίζοντας το μέσο όρο και τη τυπική απόκλιση των διανυσμάτων των δεδομένων εκπαίδευσης. Οι συντελεστές π_k αρχικοποιούνται ομοιόμορφα.
2. **Expectation-Βήμα.** Υπολογισμός της responsibility $\gamma_{nk}(\mathbf{x}_n)$ της κάθε γκαουσιανής συνιστώσας k για κάθε διάνυσμα \mathbf{x}_n με βάση τη Σχέση (5.3).
3. **Maximization-Βήμα.** Επαναπροσδιορισμός των παραμέτρων του μοντέλου χρησιμοποιώντας τις responsibilities που υπολογίστηκαν στο E-Βήμα:

$$\boldsymbol{\mu}_k^{new} = \frac{1}{N_k} \sum_{n=1}^N \gamma_{nk} \mathbf{x}_n \quad (5.5)$$

$$\boldsymbol{\Sigma}_k^{new} = \frac{1}{N_k} \sum_{n=1}^N \gamma_{nk} (\mathbf{x}_n - \boldsymbol{\mu}_k^{new})(\mathbf{x}_n - \boldsymbol{\mu}_k^{new})^T \quad (5.6)$$

$$\pi_n^{new} = \frac{N_k}{N} \quad (5.7)$$

$$\text{όπου,} \quad N_k = \sum_{n=1}^N \gamma_{nk} \quad (5.8)$$

4. Υπολογισμός της συνολικής λογαριθμικής πιθανοφάνειας με χρήση της σχέσης (5.4). Αν η διαφορά μεταξύ της τιμής που υπολογίστηκε και της τιμής που προέκυψε από την προηγούμενη επανάληψη είναι πολύ μικρή τότε ο αλγόριθμος τερματίζει, αλλιώς συνεχίζει στο Βήμα 2 και εκτελείται ακόμα μια επανάληψη.
-

5.2.2 Εφαρμογή της μεθόδου

Όπως αναφέρθηκε και στην εισαγωγή του κεφαλαίου, η ταξινόμηση του eye gaze μέσω GMMs αποτελεί μια επιβλεπόμενη διαδικασία, με την έννοια ότι θα πρέπει να οριστούν από το χρήστη οι κλάσεις στις οποίες θα γίνει η κατηγοριοποίηση. Για να γίνει αυτό, θα πρέπει να δημιουργηθεί “χειροκίνητα” ένα σύνολο εικόνων εκπαίδευσης για κάθε κλάση, η οποία θα αντιπροσωπεύει και μια διαφορετική διεύθυνση του gaze. Στην παρούσα υλοποίηση χρησιμοποιούνται 6 κλάσεις για την περιγραφή των βασικών κατευθύνσεων του gaze: “πάνω”, “κέντρο και προς τα πάνω”, “κάτω”, “κέντρο και προς τα κάτω”, “αριστερά”, “δεξιά”. Στο Σχήμα 5.1 παρουσιάζονται παραδείγματα εικόνων εκπαίδευσης για τις 6 παραπάνω κατηγορίες του eye gaze. Στη συνέχεια, για κάθε μια κλάση εκπαιδεύεται, με χρήση του αλγορίθμου EM, ένα διαφορετικό GMM, το οποίο θα περιγράφει την πιθανότητα κατάταξης του gaze στις παραπάνω κατηγορίες.

Σαν διάνυσμα χαρακτηριστικών, βάσει των οποίων θα γίνει η κατηγοριοποίηση, θα χρησιμοποιηθούν οι παράμετροι του τοπικού AAM για την περιοχή των ματιών, το οποίο έχει περιγραφεί στην υποενότητα 3.4. Σε αυτό θα περιλαμβάνονται οι παράμετροι $\mathbf{p}, \boldsymbol{\lambda}$ που ελέγχουν το σχήμα και την υφή του AAM, αλλά όχι οι παράμετροι εκείνες που αφορούν το μετασχηματισμό ομοιότητας και τον αφινικό μετασχηματισμό υφής.

Αφού γίνει η εκπαίδευση των GMMs, θα πρέπει να περιγραφεί και ο τρόπος με τον οποίο θα κατατάσσεται μια νέα εικόνα με παραμέτρους $\mathbf{x}_{n+1} = [\mathbf{p} \ \boldsymbol{\lambda}]^T$ στις παραπάνω κλάσεις. Αρχικά, για τα νέα δεδομένα υπολογίζεται η λογαριθμική πιθανοφάνεια για κάθε μία από τις 6 κατηγορίες του gaze:

$$\ln(p^{1:6}(\mathbf{x}_{n+1} | \boldsymbol{\mu}^{1:6}, \boldsymbol{\Sigma}^{1:6}, \boldsymbol{\pi}^{1:6})) = \ln \left\{ \sum_{k=1}^K \pi_k^{1:6} \mathcal{N}(\mathbf{x}_{n+1} | \boldsymbol{\mu}_k^{1:6}, \boldsymbol{\Sigma}_k^{1:6}) \right\} \quad (5.9)$$

Στη συνέχεια, η κατάταξη σε κλάσεις γίνεται εφαρμόζοντας την παρακάτω μέθοδο. Αρχικά, εντοπίζεται η πιο πιθανή κλάση, δηλαδή αυτή με τη μεγαλύτερη πιθανοφάνεια, μεταξύ των κατηγοριών: “πάνω”, “κέντρο και προς τα πάνω”, “κάτω”, “κέντρο και προς τα κάτω”, δίνοντας την κλάση για την κάθετη διεύθυνση C_{ver} . Επείτα, επαναλαμβάνεται η ίδια διαδικασία και για τις κλάσεις “κέντρο και προς τα πάνω”, “κάτω”, “κέντρο και προς τα κάτω”, “αριστερά”, “δεξιά” και προκύπτει έτσι η κλάση της οριζόντιας διεύθυνσης C_{hor} .

Αν στις κλάσεις $\{C_{ver}, C_{hor}\}$ δεν περιέχεται καμία από τις δύο κατηγορίες που εκφράζουν την οριζόντια διεύθυνση, τότε η νέα εικόνα κατατάσσεται σε δύο κλάσεις (π.χ. “πάνω” & “αριστερά”). Στην περίπτωση που μόνο μία εκ των $\{C_{ver}, C_{hor}\}$ περιέχει την ευθεία διεύθυνση τότε γίνεται κατηγοριοποίηση μόνο στην άλλη κλάση. Για παράδειγμα, αν $\{C_{ver}, C_{hor}\} = \{\text{“πάνω”}, \text{“κέντρο και προς τα κάτω”}\}$, τότε η κατηγοριοποίηση θα γίνει στην κλάση “πάνω”. Τέλος, αν και οι δύο μεταβλητές



(α') Κέντρο-Πάνω



(β') Κέντρο-Πάνω



(γ') Κέντρο-Κάτω



(δ') Πάνω & Δεξιά



(ε') Κάτω



(Ϛ') Κέντρο-Κάτω



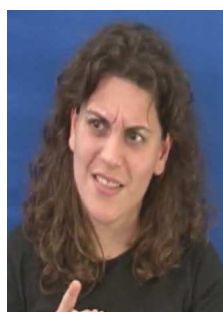
(ζ') Δεξιά & Πάνω



(η') Κέντρο-Κάτω



(θ') Πάνω & Αριστερά



(ι') Αριστερά & Πάνω

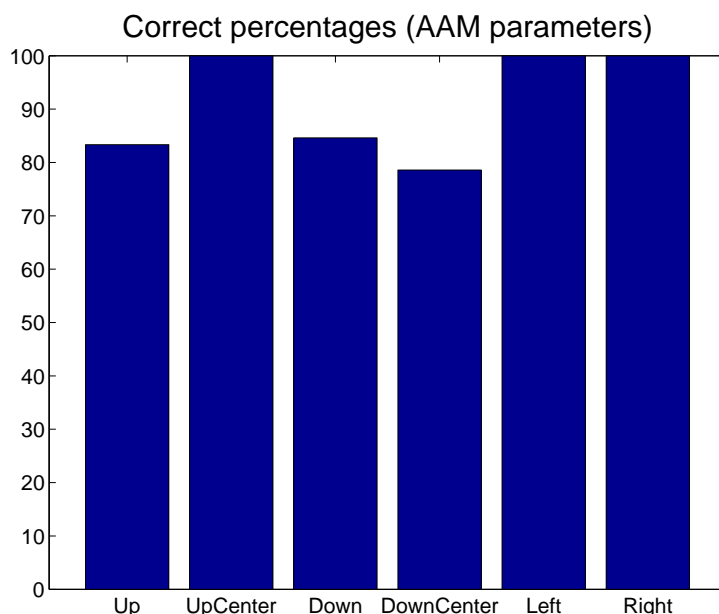


(ια') Κάτω



(ιβ') Κέντρο-Κάτω

Σχήμα 5.1: Παραδείγματα από την κατηγοριοποίηση του eye gaze για τη βάση δεδομένων της ENΓ του ΙΕΛ.



Σχήμα 5.2: Ποσοστά επιτυχούς κατηγοριοποίησης του eye gaze με χρήση των παραμέτρων του AAM και μείγματος 2 γκαουσιανών για τη βάση δεδομένων της ENG του IEL.

$\{C_{ver}, C_{hor}\}$ περιέχουν μια από τις 2 ευθείες διεύθυνσεις τότε επιλέγεται εκείνη με τη μεγαλύτερη πιθανοφάνεια.

Στο Σχήμα 5.2 παρουσιάζονται τα ποσοστά επιτυχής κατηγοριοποίησης του gaze σε κάθε μια από τις 6 κλάσεις. Σαν είσοδος του ταξινομητή επιλέχθηκαν οι παράμετροι του τοπικού AAM, ενώ χρησιμοποιήθηκαν και μοντέλα μείγματος δύο γκαουσιανών ($K = 2$). Παρατηρούμε πως τα ποσοστά αναγνώρισης αγγίζουν το απόλυτο για τις κατηγορίες “UpCenter”, “Left”, “Right”, ενώ και για τις υπόλοιπες κατηγορίες είναι αρκετά υψηλά. Αυτό συμβαίνει επειδή οι 3 πρώτες κατηγορίες περιέχουν πιο εμφανείς μεταβολές στην περιοχή των ματιών σε σχέση με τις υπόλοιπες. Συγκεκριμένα οι κλάσεις “Down”, “DownCenter” έχουν μικρές διαφορές μεταξύ τους, ενώ η κλάση “Up” μπορεί μερικές φορές να αναγνωριστεί σαν “UpCenter”. Στο σημείο αυτό, πρέπει να επιστημόνουμε πως όπου τα μάτια εμφανίζονται κλειστά, τότε θεωρούμε πως η κατεύθυνση του gaze είναι προς τα κάτω.

Όπως είδαμε, τα ποσοστά αναγνώρισης για αυτές τις πολύ βασικές κατηγορίες του gaze είναι αρκετά υψηλά, γεγονός που δείχνει την επιτυχία της μεθόδου. Αν, όμως, ο αριθμός των κλάσεων αυξηθεί τότε τα ποσοστά, όπως είναι αναμενόμενο, θα πέσουν αρκετά. Ωστόσο, ο διαχωρισμός του gaze σε περισσότερες κατηγορίες προϋποθέτει και τη δημιουργία των αντίστοιχων συνόλων εκπαίδευσης, κάτι που δεν είναι πάντα εύκολο μιας και ο πιο λεπτομερής διαχωρισμός του eye gaze είναι δύσκολος ακόμα

και για τον άνθρωπο.

Τέλος, θα πρέπει να αναφερθεί πως η επισημείωση της διεύθυνσης του gaze τόσο για τα δεδομένα εκπαίδευσης όσο και για τη διαδικασία της επαλήθευσης έγινε με βάση την προσωπική διαίσθηση. Αυτό έγινε επειδή δεν ήταν δυνατόν να βρεθεί μια επισημειωμένη βάση δεδομένων που να περιέχει υψηλής ανάλυσης εικόνες και οι μεταβολές στην κατεύθυνση του βλέματος να είναι όσο το δυνατόν πιο φυσικές.

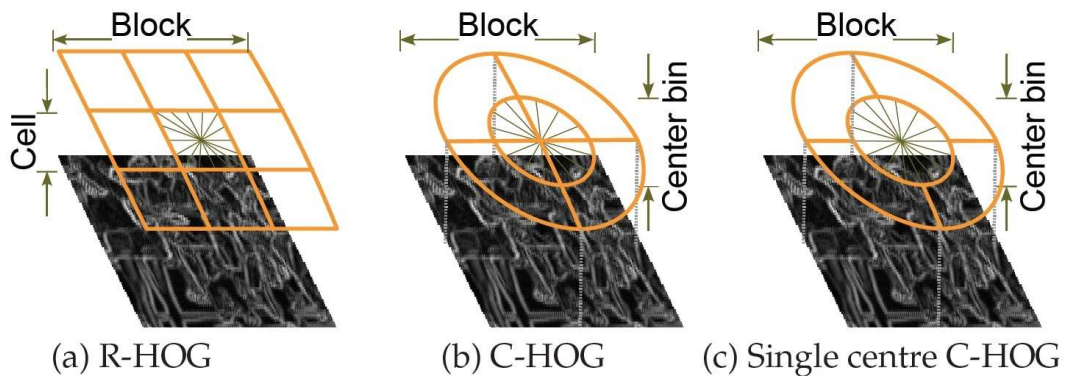
5.3 Κατηγοριοποίηση με χρήση των HOGs

5.3.1 Περιγραφή των Histograms of Oriented Gradients

Τα Ιστογράμματα Προσανατολισμένης Κλίσης (Histograms of Oriented Gradients “HOGs”) αποτελούν ένα σύνολο χαρακτηριστικών για την περιγραφή της δομής του σχήματος σε μια εικόνα και έχουν χρησιμοποιηθεί με μεγάλη επιτυχία στην αναγνώριση των αντικειμένων, όπως το ανθρώπινο σώμα, που περιέχονται σε μια εικόνα.[7, 8]. Τα HOGs παρέχουν μια πυκνή επικαλυπτόμενη περιγραφή των περιοχών μια εικόνας και υπολογίζονται σε ένα πυκνό πλέγμα ομοιόμορφα κατανεμημένων κελιών. Με τον τρόπο αυτό αναδεικνύουν την πληροφορία για το τοπικό σχήμα κωδικοποιώντας τις κατευθύνσεις της κλίσης της εικόνας σε ιστογράμματα. Τέλος, επιτυγχάνουν να είναι αρκετά αμετάβλητα ως προς τις αλλαγές της φωτεινότητας εφαρμόζοντας μια τοπική κανονικοποίηση των ιστογραμμάτων σε επικαλυπτόμενες περιοχές.

Υπάρχουν αρκετά είδη HOGs που χρησιμοποιούνται στην πράξη ανάλογα με τον τρόπο που γίνεται ο διαχωρισμός του πλέγματος κελιών και του βαθμού της επικάλυψης. Ωστόσο, τα βασικά στάδια του υπολογισμού της κλίσης στα σημεία της εικόνας και του ορισμού των διευθύνσεων βάσει των οποίων θα υπολογιστούν τα ιστογράμματα είναι τα ίδια για όλους τους τύπους HOGs.

Η αποτελεσματικότητα των HOGs σαν χαρακτηριστικά που περιγράφουν μια εικόνα εξαρτάται σε μεγάλο βαθμό από το τρόπο με τον οποίο υπολογίζεται η κλίση της εικόνας. Τόσο η οριζόντια όσο η κάθετη κλίση μιας εικόνας μπορούν να υπολογιστούν εφαρμόζοντας γκαουσιανή εξομάλυνση στην αρχική εικόνα και στη συνέχεια χρησιμοποιώντας γραμμικά φίλτρα διακριτών παραγώγων. Για την υλοποίησή τους μπορούν να χρησιμοποιηθούν διάφορες μάσκες, όπως μονοδιάστατες παράγωγοι κεντραρισμένες $([-1, 0, 1])$ ή μη $([-1, 1])$, μάσκες Sobel 3×3 , 2Δ διαγώνιες μάσκες παραγώγων $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ και $\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$. Επίσης, μπορούν να χρησιμοποιηθούν διάφορες σκάλες για την εξομάλυνση της εικόνας συμπεριλαμβανομένης και της $\sigma = 0$. Ωστόσο, σύμφωνα με τον Dalal et al.[7, 8] η απλή 1Δ μάσκα $[-1, 0, 1]$ και η μηδενική σκάλα εξομάλυνσης δίνουν καλύτερα



Σχήμα 5.3: Τύποι «κελιών» και μπλοκ που χρησιμοποιούνται για την κατασκευή των HOGs. Έχει ληφθεί από το [7].

αποτελέσματα. Στην περίπτωση που έχουμε έγχρωμες εικόνες, τότε η κλίση υπολογίζεται για κάθε κανάλι ξεχωριστά και στο τέλος κρατάμε το αποτέλεσμα εκείνο που εμφανίζει τη μεγαλύτερη νόρμα.

Το επόμενο στάδιο αφορά την κατασκευή των ιστογραμμάτων και αποτελεί το μη γραμμικό μέρος των HOGs. Για κάθε pixel της εικόνας υπολογίζεται η «ψήφος» με την οποία θα συνεισφέρει στο ιστόγραμμα κατευθυνόμενης κλίσης ανάλογα με την γωνία που εμφανίζει η κλίση στο συγκεκριμένο pixel. Στη συνέχεια συσσωρεύονται στα bins του ιστογράμματος οι «ψήφοι» από όλα τα pixels που ανήκουν σε μια τοπική περιοχή της εικόνας, η οποία ονομάζεται «κελί». Τα «κελιά» αυτά μπορούν να είναι ορθογώνια ή να αποτελούν τμήματα ενός κυκλικού δίσκου, όπως θα δούμε και στη συνέχεια. Τα bins του ιστογράμματος κατανέμονται ομοιόμορφα μεταξύ $0^\circ - 180^\circ$, στην περίπτωση που δεν λαμβάνεται υπόψιν το πρόσημο της κλίσης, και μεταξύ $0^\circ - 360^\circ$ αν αυτό συμπεριληφθεί. Η «ψήφος» αποτελεί συνάρτηση του πλάτους της κλίσης στο κάθε pixel και μπορεί να είναι το πλάτος αυτούσιο, το τετράγωνό του ή η τετραγωνική ρίζα του. Στη πράξη συνήθως χρησιμοποιείται το πλάτος της κλίσης ως έχει, ενώ για το ιστόγραμμα επιλέγονται 9 bins μεταξύ $0^\circ - 180^\circ$, δηλαδή χωρίς να ληφθεί υπόψιν το πρόσημο της κλίσης.

Η ένταση της κλίσης σε μια περιοχή της εικόνας επηράζεται έντονα από τις τοπικές μεταβολές της φωτεινότητας, γι' αυτό μια αποτελεσματική κανονικοποίηση είναι απαραίτητη προκειμένου να πετύχουμε καλή απόδοση για τον περιγραφέα βάσει των HOGs. Για το λόγο αυτό τα «κελιά» που αναφέρθηκαν προηγουμένως, οργανώνονται σε μεγαλύτερα τοπικά μπλοκ, κάθε ένα από τα οποία κανονικοποιείται ξεχωριστά. Έτσι το τελικό διάνυσμα των χαρακτηριστικών αποτελείται από τις κανονικοποιημένες αποκρίσεις των κελιών σε κάθε μπλοκ. Μάλιστα, προκειμένου να αυξηθεί η απόδοση, στις περισσότερες περιπτώσεις υπάρχουν επικαλύψεις μεταξύ των μπλοκ, έτσι ώστε η απόκριση του κάθε «κελιού» να περιλαμβάνεται περισσότερες φορές στο τελικό

διάνυσμα χαρακτηριστικών αλλά με διαφορετική κανονικοποίηση κάθε φορά.

Όσον αφορά τώρα τη γεωμετρία των μπλοκ, υπάρχουν δύο μεγάλες κατηγορίες: τα ορθογώνια ή τετραγωνικά μπλοκ (R-HOG), τα οποία χωρίζονται σε πλέγματα ορθογώνιων ή τετραγωνικών «κελιών», και τα κυκλικά (C-HOG), τα οποία χωρίζονται σε «κελιά» λογαριθμική-πολικής μορφής. Τα C-HOG μπορούν να χωριστούν σε δύο επιπλέον κατηγορίες ανάλογα με το αν το κεντρικό «κελί» είναι χωρισμένο ή όχι σε γωνιακούς τομείς. Στο Σχήμα 5.3 παρουσιάζονται οι τύποι μπλοκ που αναφέρθηκαν παραπάνω, καθώς και ο τρόπος χωρισμού τους σε κελιά.

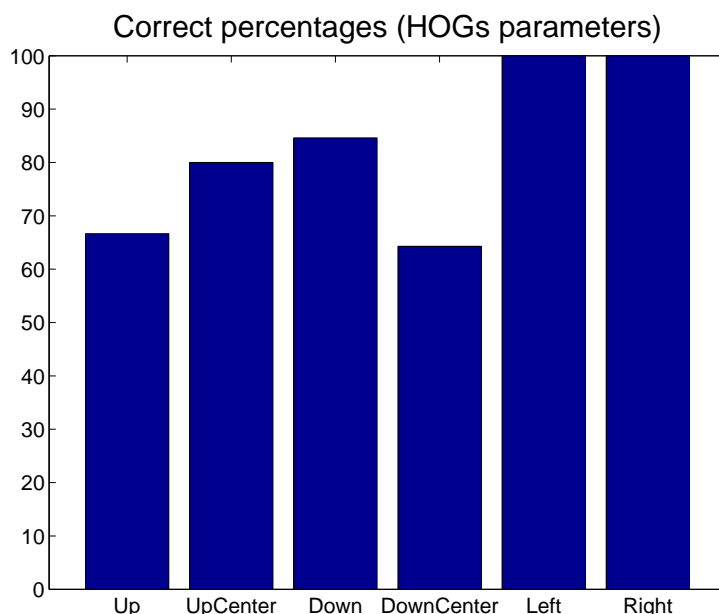
Τέλος, υπάρχουν αρκετοί διαφορετικοί τρόποι με τους οποίους μπορεί να γίνει η κανονικοποίηση σε κάθε ένα από τα μπλοκ των HOGs. Αν θεωρήσουμε σαν \mathbf{v} το μη κανονικοποιημένο διάνυσμα χαρακτηριστικών, $\|\mathbf{v}\|_k$ τη νόρμα 1^{ης} ή 2^{ης} τάξης και ϵ μια μικρή σταθερά τότε οι τρόποι κανονικοποίησης θα είναι:

- Η νόρμα 2^{ης} τάξης (L2): $\mathbf{v} \rightarrow \frac{\mathbf{v}}{\sqrt{\|\mathbf{v}\|_2^2 + \epsilon^2}}$
- Η νόρμα L2 ακολουθούμενη από περιορισμό των μέγιστων τιμών στο 0.2 και κανονικοποιώντας ξανά
- Η νόρμα 1^{ης} τάξης (L1): $\mathbf{v} \rightarrow \frac{\mathbf{v}}{(\|\mathbf{v}\|_1 + \epsilon)}$
- Η νόρμα L2 ακολουθούμενη από την τετραγωνική ρίζα: $\mathbf{v} \rightarrow \frac{\mathbf{v}}{\sqrt{\|\mathbf{v}\|_1 + \epsilon}}$

5.3.2 Εφαρμογή των HOGs σε εικόνες των ματιών

Για την κατηγοριοποίηση του gaze σε κλάσεις, μπορούμε να χρησιμοποιήσουμε σαν είσοδο στο ταξινομητή, εκτός από τις παραμέτρους του AAM και άλλα είδη χαρακτηριστικών. Τα HOGs χαρακτηριστικά έχουν εφαρμοστεί επιτυχώς στην αναγνώριση ανθρώπων σε εικόνες, και έτσι αναμένουμε να έχουν ανάλογη επιτυχία και στην αναγνώριση του eye gaze. Από τις διάφορες κατηγορίες HOGs, τα R-HOG είναι πιο δημοφιλή και έχουν δώσει ικανοποιητικά αποτελέσματα σε αρκετές διαφορετικές περιπτώσεις. Για το λόγο αυτό, είναι αυτά που θα χρησιμοποιήσουμε και στην παρούσα μέθοδο. Επίσης, σαν μέθοδος κανονικοποίησης επιλέγεται η απλή L1 νόρμα. Για την επαρκή περιγραφή ενός R-HOG είναι απαραίτητο να οριστούν 3 παράμετροι:

- το μέγεθος του μπλοκ σε αριθμό «κελιών», το οποίο επιλέγεται [2x2].
- το μέγεθος ενός κελιού σε pixels, το οποίο επιλέγεται ξεχωριστά για κάθε εικόνα με τέτοιο τρόπο ώστε να προκύψει ο ίδιος αριθμός «κελιών».
- ο αριθμός των bins, ο οποίος επιλέγεται να είναι 9, ενώ τα ιστογράμματα κατασκευάζονται για $0^\circ - 180^\circ$.



Σχήμα 5.4: Ποσοστά επιτυχούς κατηγοριοποίησης του eye gaze με χρήση των HOG χαρακτηριστικών και μείγματος 2 γκαουσιανών για τη βάση δεδομένων της ENG του IEL.

Για τον υπολογισμό των HOG χαρακτηριστικών είναι απαραίτητο να προσδιοριστεί η περιοχή των ματιών. Αυτό μπορεί να γίνει ορίζοντας ένα ορθογώνιο με βάση τα σημεία της μάσκας σχήματος του AAM, με ανάλογο τρόπο όπως στην ενότητα 4.2.2. Στην συνέχεια, για κάθε εικόνα προκύπτει ένα διάνυσμα από HOG χαρακτηριστικά με βάση το οποίο θα γίνει η κατηγοριοποίηση σε κλάσεις. Η μέθοδος που ακολουθείται τόσο για την εκπαίδευση όσο και για τη διαδικασία επαλήθευσης είναι η ίδια που εφαρμόστηκε στην προηγούμενη ενότητα για τις παραμέτρους του AAM.

Στο Σχήμα 5.4 παρουσιάζονται τα ποσοστά επιτυχούς κατηγοριοποίησης του gaze σε κάθε μια από τις 6 κλάσεις. Σαν είσοδος του ταξινομητή τώρα επιλέχθηκαν τα HOG χαρακτηριστικά, ενώ χρησιμοποιήθηκαν και πάλι μοντέλα μείγματος δύο γκαουσιανών ($K = 2$). Παρατηρούμε πως, παρόλο που σε γενικές γραμμές τα ποσοστά αναγνώρισης παραμένουν αρκετά υψηλά, αν εξαιρέσουμε τις κλάσεις “Left”, “Right” τα υπόλοιπα εμφανίζονται σχετικά πεσμένα σε σχέση με εκείνα της προηγούμενης ενότητας. Επίσης, πρέπει να αναφέρουμε πως το διάνυσμα των HOG χαρακτηριστικών είναι αρκετά μεγαλύτερο (81 τιμές) σε σχέση με εκείνο των AAM παραμέτρων (18 τιμές).

Από όλα αυτά, μπορούμε να συμπεράνουμε πως η χρήση των AAM σαν χαρακτηριστικά αναγνώρισης του gaze είναι προτιμότερη σε σχέση με τη χρήση HOGs, αφού εμφανίζει καλύτερα αποτελέσματα χρησιμοποιώντας αρκετά μικρότερο

αριθμό παραμέτρων. Το γεγονός αυτό είναι σε ένα βαθμό αναμενόμενο, μιας και τα HOGs αποτελούν ένα τοπικό εργαλείο για την περιγραφή του σχήματος μια εικόνας σε αντίθεση με τα AMMs που στηρίζονται στην μοντελοποίηση ενός συγκεκριμένου αντικειμένου αξιοποιώντας πληροφορίες και από άλλες παρόμοιες εικόνες.

5.4 Υπολογισμός των γωνιών του gaze από την πιθανοφάνεια κάθε κλάσης

5.4.1 Περιγραφή του αλγορίθμου

Στις προηγούμενες ενότητες του κεφαλαίου είδαμε διάφορους τρόπους με τους οποίους μπορεί να γίνει η κατηγοριοποίηση της διεύθυνσης του gaze σε κλάσεις, παίρνοντας με τον τρόπο αυτό μια αρκετά καλή διακριτή εκτίμηση. Ωστόσο, σε αρκετές περιπτώσεις είναι απαραίτητη μια συνεχής εκτίμηση για το gaze, μιας και η ακρίβεια που δίνουν οι 6 βασικές κατηγορίες δεν είναι αρκετή. Για το λόγο αυτό, προτείνεται μια μέθοδος η οποία αξιοποιεί την λογαριθμική πιθανοφάνεια το eye gaze να ανήκει σε μια από τις 6 βασικές κλάσεις προκειμένου να υπολογίσει τις γωνίες της οριζόντιας και κάθετης διεύθυνσης του gaze.

Η προτεινόμενη μέθοδος χρησιμοποιεί δύο διαφορετικές προσεγγίσεις για τον υπολογισμό των γωνιών. Η μια αφορά μεγάλες μεταβολές στις γωνίες του gaze και στηρίζεται στον προσδιορισμό της προβολής του διανύσματος του gaze στο επίπεδο της εικόνας από τους λόγους πιθανοφάνειας μεταξύ των κλάσεων. Η άλλη αφορά μικρές μεταβολές του gaze γύρω από την ευθεία διεύθυνση και υπολογίζεται με γραμμική συσχέτιση της πιθανοφάνειας και της γωνίας διεύθυνσης του gaze.

Για τον προσδιορισμό της γωνίας $angle_{ver}$ της κάθετης διεύθυνσης, αρχικά υπολογίζουμε τους παρακάτω λόγους πιθανοφάνειας ($loglik^i$ η πιθανοφάνεια κάθε κλάσης αφού αφαιρεθεί η ελάχιστη τιμή τους):

$$V_{up} = \frac{loglik^{Up} - loglik^{UpCenter}}{loglik^{Up}} \quad (5.10)$$

$$V_{down} = \frac{loglik^{Down} - loglik^{DownCenter}}{loglik^{Down}} \quad (5.11)$$

Εάν στους παραπάνω λόγους προκύψει αρνητική τιμή τότε αντί της τιμής αυτής θα πάρουν την τιμή μηδέν. Ο λόγος V_{up} θεωρούμε ότι εκφράζει την κάθετη προβολή του διανύσματος του gaze πάνω στην εικόνα στην περίπτωση που αυτό έχει κατεύθυνση προς τα πάνω και σαν ευθεία διεύθυνση θεωρηθεί αυτή του “κέντρο και πάνω”. Αντίστοιχα, ο όρος V_{down} θα εκφράζει την προβολή στην περίπτωση που η κατεύθυνση είναι προς τα κάτω και η ευθεία διεύθυνση αντιπροσωπεύεται από την κλάση “κέντρο και κάτω”. Επομένως, η συνολική κάθετη προβολή θα είναι $V_{updown} = V_{up} - V_{down}$,

ενώ η γωνία της κάθετης διεύθυνσης θα δίνεται:

$$angle_{ver} = \arctan (V_{updown}) \quad (5.12)$$

Στη συνέχεια, υπολογίζεται και ο λόγος μεταξύ των δύο κεντρικών διευθύνσεων:

$$D_{center} = \frac{\loglik^{UpCenter} - \loglik^{DownCenter}}{\max(\loglik^{UpCenter}, \loglik^{DownCenter})} \quad (5.13)$$

Επειδή κατά τον υπολογισμό της γωνίας $angle_{ver}$ θεωρήσαμε ότι υπάρχουν δύο διαφορετικές ευθείες διευθύνσεις, θα πρέπει να προσθεθεί στο αποτέλεσμα και μια γωνία διώρθωσης η οποία θα εκφράζει τη διαφορά μεταξύ των δύο κεντρικών κατευθύνσεων. Η γωνία αυτή προκύπτει πολλαπλασιάζοντας τον λόγο D_{center} με μια μικρή γωνία θ_{center} της τάξης των $7^\circ - 10^\circ$ και έτσι έχουμε τελικά για τη γωνία της κάθετης διεύθυνσης:

$$angle_{ver} = \arctan (V_{updown}) + D_{center} \cdot \theta_{center} \quad (5.14)$$

Αντίστοιχα, για την οριζόντια διεύθυνση υπολογίζονται οι λόγοι:

$$V_{left} = \frac{\loglik^{Left} - \loglik^{Center}}{\loglik^{Left}} \quad (5.15)$$

$$V_{right} = \frac{\loglik^{Right} - \loglik^{Center}}{\loglik^{Right}} \quad (5.16)$$

Σαν *Center* επιλέγεται η κεντρική διεύθυνση με την μεγαλύτερη πιθανοφάνεια, ενώ και πάλι αν προκύψουν αρνητικές τιμές οι λόγοι παίρνουν μηδενικές τιμές. Εδώ ο λόγος V_{left} θεωρούμε ότι εκφράζει την οριζόντια προβολή του διανύσματος του gaze αν η κατεύθυνση είναι προς τα αριστερά, ενώ ο λόγος V_{right} την προβολή στην περίπτωση που έχει κατεύθυνση προς τα δεξιά. Άρα η συνολική οριζόντια προβολή θα είναι $V_{leftright} = V_{left} - V_{right}$, ενώ η γωνία της κάθετης διεύθυνσης θα δίνεται:

$$angle_{hor} = \arctan (V_{leftright}) \quad (5.17)$$

Παρατηρώντας τις παραπάνω σχέσεις συμπεραίνουμε πως οι μέγιστες τιμές που μπορούν να πάρουν οι παραπάνω γωνίες είναι περίπου κοντά στις 45° , γεγονός που είναι σύμφωνο και με τις εφικτές γωνίες περιστροφής του ματιού.

Ωστόσο, η μέθοδος που αναπτύχθηκε παραπάνω εμφανίζει αδυναμία να εκτιμήσει γωνίες του gaze που βρίσκονται κοντά στην ευθεία διεύθυνση. Αυτό συμβαίνει, γιατί στην περίπτωση αυτή οι κεντρικές διευθύνσεις θα έχουν μεγαλύτερες πιθανοφάνειες και έτσι οι παραπάνω λόγοι θα προκύπτουν αρνητικοί δηλαδή μηδενικοί, όπως αναφέρθηκε και παραπάνω. Επομένως, θα πρέπει να γίνει μια τροποποίηση ώστε σε περίπτωση που και οι δύο λόγοι V_{up}, V_{down} ή V_{left}, V_{right} προκύπτουν αρνητικοί

να μπορεί να υπολογιστούν οι γωνίες του gaze. Έτσι, υπολογίζονται δύο γωνίες $small_{ver}$ και $small_{hor}$ για την περιγραφή μικρών μεταβολών του gaze γύρω από την ευθεία διεύθυνση, οι οποίες υπολογίζονται με γραμμική συσχέτιση της πιθανοφάνειας κάθε κλάσης:

$$small_{ver} = \left(\frac{loglik^{Up}}{loglik^{UpCenter}} - \frac{loglik^{Down}}{loglik^{DownCenter}} \right) \theta_{ver} \quad (5.18)$$

$$small_{hor} = \left(\frac{loglik^{Right} - loglik^{Left}}{loglik^{Center}} \right) \theta_{hor} \quad (5.19)$$

Οι γωνίες $\theta_{ver}, \theta_{hor}$ αποτελούν δύο μικρές γωνίες της τάξης των 5° , ενώ στη γωνία $small_{ver}$ θα πρέπει να προσθεθεί και η γωνία μεσαξύ των κεντρικών γωνιών. Επομένως, τελικά, οι γωνίες διεύθυνσης του gaze θα δίνονται από τις σχέσεις 5.12 και 5.17 στην περίπτωση μεγάλων μεταβολών, ενώ για μικρές μεταβολές γύρω από τη μηδενική θέση θα δίνονται από τις σχέσεις:

$$angle_{ver} = small_{ver} + D_{center} \cdot \theta_{center} \quad (5.20)$$

$$angle_{hor} = small_{hor} \quad (5.21)$$

5.4.2 Πειραματικά αποτελέσματα

Εφαρμόζουμε, λοιπόν, την παραπάνω μέθοδο σε εικόνες από τη βάση δεδομένων της ENΓ του ΙΕΛ. Αρχικά, υπολογίζονται, όπως και προηγουμένως, οι λογαριθμικές πιθανοφάνειες για τα GMMs κάθε κλάσης, χρησιμοποιώντας σαν διάνυσμα χαρακτηριστικών είτε τις παραμέτρους του AAM είτε τα HOG χαρακτηριστικά. Στη συνέχεια, όμως δεν γίνεται κατηγοριοποίηση του gaze στις 6 βασικές κατηγορίες, αλλά υπολογίζονται οι γωνίες διεύθυνσης σύμφωνα με τον παραπάνω αλγόριθμο. Στο Σχήμα 5.5 παρουσιάζονται κάποια παραδείγματα εφαρμογής της μεθόδου αν σαν είσοδο του ταξινομητή χρησιμοποιηθούν οι AAM παράμετροι, ενώ στο Σχήμα 5.6 απεικονίζονται τα ίδια παραδείγματα αλλά τώρα σαν παράμετροι αναγνώρισης έχουν χρησιμοποιηθεί τα HOG χαρακτηριστικά.

Παρατηρούμε πως σε γενικές γραμμές τα αποτελέσματα είναι αρκετά ικανοποιητικά και συμβαδίζουν σε μεγάλο βαθμό με την ανθρώπινη διαίσθηση. Σε σχέση με τα αντίστοιχα αποτελέσματα που προέκυψαν από τη χρήση του γεωμετρικού μοντέλου, παρατηρούνται αρκετές ομοιότητες αλλά και διαφορές. Συγκεκριμένα, οι γωνίες που προκύπτουν μέσω της πιθανοφάνειας εμφανίζονται λιγότερα ακριβείς και έχουν τη τάση να μεγενθύνουν τις ακραίες μεταβολές του gaze, όπως η στροφή προς τα πάνω/κάτω ή αριστερά/δεξιά. Από την άλλη πλευρά όμως, καταφέρνουν να δώσουν αποδεκτά αποτελέσματα και σε δύσκολες περιπτώσεις που έχουμε επικαλύψεις του προσώπου, στις οποίες το ευαίσθητο γεωμετρικό μοντέλο είναι δύσκολο να παράγει σωστή εκτίμηση.



Σχήμα 5.5: Παραδείγματα εφαρμογής της μεθόδου υπολογισμού των γωνιών διεύθυνσης του gaze από τις πιθανοφάνειες των GMM κάθε κλάσης με χρήση των AAM παραμέτρων, για τη βάση δεδομένων της ENΓ του IEL.

Σε σχέση τώρα με το διάνυσμα χαρακτηριστικών που χρησιμοποιείται κάθε φορά, φαίνεται ξεκάθαρα πως οι AAM παράμετροι υπερτερούν έναντι των HOG χαρακτηριστικών, μιας και τα τελευταία αποτυγχάνουν σε κάποιες περιπτώσεις να εκτίμησουν σωστά τη διεύθυνση του gaze (Εικόνες 5.6γ', 5.6ϕ'). Το γεγονός αυτό, είχε παρατηρηθεί και κατά τη διαδικασία κατηγοριοποίησης του gaze σε διακριτές κλάσεις, ωστόσο τώρα με τις γωνίες διεύθυνσης γίνεται πιο εμφανές.



Σχήμα 5.6: Παραδείγματα εφαρμογής της μεθόδου υπολογισμού των γωνιών διεύθυνσης του gaze από τις πιθανοφάνειες των GMM κάθε κλάσης με χρήση των HOG χαρακτηριστικών, για τη βάση δεδομένων της ENΓ του IEL.

5.5 Συμπεράσματα

Ανακεφαλαιώνοντας, στο συγκεκριμένο κεφάλαιο αναφέρθηκαν μέθοδοι εκτίμησης του eye gaze, οι οποίες στηρίζονται σε τεχνικές αναγνώρισης προτύπων για να κατηγοριοποιήσουν το gaze σε διακριτές κλάσεις. Χρησιμοποιήθηκαν δύο τύποι παραμέτρων αναγνώρισης, με τις AAM παραμέτρους να εμφανίζουν καλύτερα αποτελέσματα από τα HOG χαρακτηριστικά. Τόσο η διακριτή εκτίμηση του gaze μέσω των 6 βασικών κλάσεων, όσο η συνεχής εκτίμηση μέσω του υπολογισμού των γωνιών είναι αρκετά ικανοποιητικές στις περισσότερες περιπτώσεις. Επιπλέον, οι μέθοδοι του κεφαλαίου αυτού εμφανίζονται αρκετά εύρωστες σε αστάθμιστους παράγοντες, όπως οι επικαλύψεις του προσώπου, σε αντίθεση με τη γεωμετρική μέθοδο του προηγούμενου

κεφαλαίου. Βέβαια, για να κερδίσουν την ευρωστία αυτή, οι μέθοδοι αυτές χάνουν λίγο σε ακρίβεια, γεγονός που δεν δημιουργεί προβλήματα απαραίτητα μιας και η απαιτούμενη ακρίβεια εξαρτάται από την εκάστοτε εφαρμογή.

Επίσης, πρέπει να αναφερθεί πως οι παραπάνω μέθοδοι απαιτούν τη δημιουργία ενός συνόλου εκπαίδευσης με επισημειωμένη τη διεύθυνση του gaze, σε αντίθεση με το γεωμετρικό μοντέλο. Έτσι, δεν είναι εύκολη η προσαρμογή τους σε νέους χρήστες μιας και η διαδικασία επισημείωσης των δεδομένων είναι μια “χειροκίνητη” διαδικασία. Η αξιοποίηση δεδομένων εκπαίδευσης από άλλους χρήστες θα μπορούσε να αποτελέσει μια λύση, αλλά στην περίπτωση αυτή η αποτελεσματικότητα των μεθόδων θα μειωνόταν αισθητά.

Τέλος, όσον αφορά τον συνυπολογισμό της πόζας κεφαλιού στην εκτίμηση του gaze, αυτό μπορεί να γίνει με την κατάλληλη εκπαίδευση των ταξινομητών κάθε κλάσης. Στην περίπτωση της συγκεκριμένης διπλωματικής, έγινε προσπάθεια οι κατηγορίες του gaze να οριστούν από τη γενικότερη αίσθηση για την κατεύθυνση του βλέματος, συμπεριλαμβανομένης της πόζας κεφαλιού, και όχι κοιτώντας αποκλειστικά την περιοχή των ματιών. Επιπρόσθετα, η χρήση πληροφορίας από κάποια εξωτερική μέθοδο εκτίμησης της πόζας αποτελεί μια ακόμα αποδεκτή λύση.

Κεφάλαιο 6

Κατηγοριοποίηση Συναισθήματος

6.1 Εισαγωγή

Η αναγνώριση του ανθρώπινου συναισθήματος από υπολογιστή αποτελεί ένα ακόμα σημαντικό πρόβλημα της όρασης υπολογιστών, που βρίσκει πολλές εφαρμογές στην επικοινωνία ανθρώπου-μηχανής. Η αναγνώριση της συναισθηματικής κατάστασης ενός ατόμου αποτελεί ένα δύσκολο και σύνθετο πρόβλημα ακόμα και για τους επιστήμονες που ασχολούνται με την ψυχολογία, πόσο μάλλον για μια υπολογιστική μηχανή. Αυτό συμβαίνει γιατί πολλές φορές ο άνθρωπος μπορεί να έχει ένα μίγμα διαφορετικών συναισθημάτων ή μπορεί να θέλει να κρύψει τα πραγματικά του συναισθήματα αλλά και για πολλούς άλλους ακόμη λόγους. Γι' αυτό με τον όρο αναγνώριση συναισθήματος από υπολογιστή ουσιαστικά αναφερόμαστε στην κατηγοριοποίηση των ανθρώπινων συναισθημάτων σε έναν αριθμό διακριτών κλάσεων. Συνήθως, οι κλάσεις αυτές επιλέγονται έτσι ώστε να αντιπροσωπεύουν τις 7 βασικές κατηγορίες συναισθημάτων (θυμός, περιφρόνηση, αηδία, φόβος, ευτυχία, θλίψη, έκπληξη) καθώς και το ουδέτερο συναίσθημα.

Υπάρχουν αρκετοί τρόποι με τους οποίους μπορεί να γίνει η έκφραση ενός συναισθήματος όπως η αλλαγή στη χροιά της φωνής, η στάση του σώματος και των χεριών και οι μεταβολές στις εκφράσεις του προσώπου. Στο συγκεκριμένο κεφάλαιο θα επικεντρωθούμε στην έκφραση συναισθημάτων μέσω του προσώπου και συγκεκριμένα στα στοιχεία που προκύπτουν από την μοντελοποίησή του μέσω των AAMs. Κάθε πρόβλημα κατηγοριοποίησης μπορεί να χωριστεί σε δύο φάσεις: στην εξαγωγή των χαρακτηριστικών βάσει των οποίων θα γίνει η αναγνώριση και στην επιλογή του κατάλληλου ταξινομητή. Στην περίπτωση που θα εξετάσουμε, σαν διάνυσμα χαρακτηριστικών θα χρησιμοποιηθούν οι παράμετροι σχήματος και υψής που προκύπτουν από την προσαρμογή AAMs σε εικόνες

προσώπων με διαφορετικά συναισθήματα, ενώ σαν μέθοδος κατηγοριοποίησης χρησιμοποιούνται Κρυφά Μαρκοβιανά Μοντέλα (HMMs) με συνεχή συνάρτηση πυκνότητας πιθανότητας.

Στη συνέχεια του κεφαλαίου γίνεται μια σύντομη περιγραφή της δομής των (HMMs), του τρόπου υπολογισμού της πιθανοφάνειας του μοντέλου καθώς και της μεθόδου εκπαίδευσης. Έπειτα, παρουσιάζονται κάποια πειραματικά αποτελέσματα για τη βάση εικόνων “The extended Cohn-Kanade (CK+)” του Carnegie Mellon University (CMU)[18] σχετικά με την μέθοδο κατηγοριοποίησης συναισθήματος που εφαρμόζουμε. Επίσης, παρουσιάζονται κάποιες τροποποιήσεις της μεθόδου κατηγοριοποίησης που βελτιώνουν τη συνολική απόδοση και προσαρμόζουν εύκολα το σύστημα σε έναν συγκεκριμένο χρήστη αυξάνοντας έτσι κατά πολύ τα ποσοστά αναγνώρισης. Τέλος, καταγράφονται κάποια συμπεράσματα σχετικά με το σύνολο των μεθόδων που χρησιμοποιήθηκαν.

6.2 Κρυφά Μαρκοβιανά Μοντέλα (HMMs)

6.2.1 Δομή ενός HMM

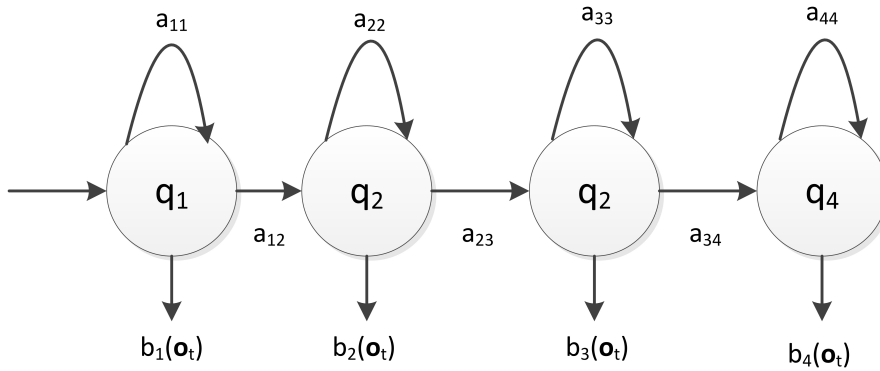
Ένα Κρυφό Μαρκοβιανό Μοντέλο αποτελεί μια μηχανή πεπερασμένων καταστάσεων, οι οποίες δεν είναι άμεσα παρατηρήσιμες παρά μόνο μέσω της ακολουθίας των παραγόμενων συμβόλων του μοντέλου.[29, 30] Για την περιγραφή ενός HMM απαιτείται ο ορισμός ενός συνόλου παραμέτρων:

- Σε κάθε χρονική στιγμή $t=1,2,\dots,T$ το μοντέλο βρίσκεται σε μία κατάσταση $q_t = \{q_1, q_2, \dots, q_N\}$, όπου N ο συνολικός αριθμός των καταστάσεων του μοντέλου. Σε κάθε κατάσταση i ορίζεται η πιθανότητα μετάβασης στην κατάσταση j : $\mathbf{A} = \{a_{ij}\} : a_{ij} = Pr(q_{t+1} = j | q_t = i)$, $1 \leq i, j \leq N$. Οι πιθανότητες a_{ij} θα πρέπει να ικανοποιούν τις σχέσεις:

$$a_{ij} \geq 0 \quad (6.1)$$

$$\sum_{j=1}^N a_{ij} = 1, \forall i \quad (6.2)$$

- Σε κάθε κατάσταση εξάγεται μια παρατήρηση \mathbf{o}_t από το σύνολο συμβόλων $V = \{v_1, v_2, \dots, v_{N_{symp}}\}$ των δυνατών παρατηρήσεων. Σε κάθε κατάσταση αντιστοιχεί μια διαφορετική συνάρτηση κατανομής για την πιθανότητα εξαγωγής του κάθε συμβόλου: $\mathbf{B} = \{b_j\} : b_j(\mathbf{o}_t) = Pr(\mathbf{o}_t | q_t = j)$. Στην περίπτωση HMM με συνεχή συνάρτηση πυκνότητας πιθανότητας αυτή θα



Σχήμα 6.1: Γραμμικό δεξιόστροφο (left-right) HMM με 4 καταστάσεις.

ορίζεται σαν το συνδυασμό M γκαουσιανών κατανομών:

$$b_j(\mathbf{o}_t) = \sum_{k=1}^M c_{jk} \mathcal{N}(\mathbf{o}_t, \boldsymbol{\mu}_{jk}, \mathbf{U}_{jk}), 1 \leq j \leq N \quad (6.3)$$

Οι παράγοντες $c_{jk}, \boldsymbol{\mu}_{jk}, \mathbf{U}_{jk}$ αντιπροσωπεύουν τους συντελεστές των γκαουσιανών, το διάνυσμα των μέσων όρων και τον πίνακα συμμεταβλητότητας αντίστοιχα για την k κατανομή και την j κατάσταση. Αν θεωρήσουμε ότι τα επιμέρους στοιχεία της παρατήρησης \mathbf{o}_t είναι ανεξάρτητα μεταξύ τους τότε ο πίνακας συμμεταβλητότητας θα προκύψει διαγώνιος. Επίσης, οι συντελεστές των γκαουσιανών θα πρέπει να ικανοποιούν τους παρακάτω περιορισμούς:

$$\sum_{k=1}^M c_{jk} = 1, \quad 1 \leq j \leq N \quad (6.4)$$

$$c_{jk} \leq 0, \quad 1 \leq j \leq N, \quad 1 \leq k \leq M \quad (6.5)$$

- Τέλος, ορίζεται και μια κατανομή πιθανότητας π_i η οποία εκφράζει την πιθανότητα το HMM να ξεκινά από την κατάσταση i :
 $\boldsymbol{\pi} = \{\pi_i\} : \pi_i = Pr(q_1 = i), 1 \leq i \leq N.$

Έτσι, λοιπόν, ένα HMM μπορεί να περιγραφεί πλήρως από το σύνολο παραμέτρων $\boldsymbol{\lambda} = (\mathbf{A}, \mathbf{B}, \boldsymbol{\pi})$.

Μια ειδική κατηγορία HMM αποτελούν τα γραμμικά δεξιόστροφα (left-right) HMMs. Στο Σχήμα 6.1 παρουσιάζεται ένα HMM αυτής της μορφής με 4 καταστάσεις. Τα HMMs αυτά έχουν χρησιμοποιηθεί με επιτυχία σε εφαρμογές αναγνώρισης φωνής, ωστόσο είναι κατάλληλα και για οποιοδήποτε πρόβλημα περιλαμβάνει χρονική εξέλιξη,

όπως αυτό της εκδήλωσης συναισθήματος μέσω εκφράσεων του προσώπου. Όπως παρατηρούμε, στα HMMs αυτά η ροή μετάβασης των καταστάσεων είναι δεξιόστροφη και δεν υπάρχει η δυνατότητα επιστροφής σε προηγούμενη κατάσταση. Επίσης, επιτρέπονται μεταβάσεις μόνο μεταξύ διαδοχικών καταστάσεων. Έτσι ο πίνακας μεταβάσεων \mathbf{A} του μοντέλου έχει τη μορφή:

$$a_{ij} = 0, \quad j < i \quad (6.6)$$

$$a_{ij} = 0, \quad j > i + 1 \quad (6.7)$$

ενώ οι αρχικές πιθανότητες των καταστάσεων θα είναι:

$$\pi_i = \begin{cases} 0, & i \neq 1 \\ 1, & i = 1 \end{cases} \quad (6.8)$$

6.2.2 Υπολογισμός της πιθανοφάνειας του μοντέλου

Έχοντας, λοιπόν, ένα HMM με παραμέτρους $\boldsymbol{\lambda} = (\mathbf{A}, \mathbf{B}, \boldsymbol{\pi})$ και μια ακολουθία παρατηρήσεων $\mathbf{O} = (\mathbf{o}_1, \mathbf{o}_2, \dots, \mathbf{o}_T)$, ένα από τα βασικά προβλήματα προς επίλυση είναι ο υπολογισμός της πιθανοφάνειας $Pr(\mathbf{O}|\boldsymbol{\lambda})$ για την ακολουθία παρατηρήσεων δεδομένου του μοντέλου. Επειδή ο απευθείας τρόπος υπολογισμού της παραπάνω πιθανότητας είναι αρκετά χρονοβόρος, χρησιμοποιείται είτε ο forward είτε ο backward αλγόριθμος.

Στον forward αλγόριθμο ορίζεται η μεταβλητή $\alpha_t(i)$, η οποία εκφράζει την πιθανότητα ενώ βρισκόμαστε στην κατάσταση q_i να παρουσιαστεί μια ακολουθία παρατηρήσεων μέχρι τη χρονική στιγμή t , δεδομένου του μοντέλου $\boldsymbol{\lambda}$:

$$\alpha_t(i) = Pr(\mathbf{o}_1, \mathbf{o}_2, \dots, \mathbf{o}_t, q_t = i | \boldsymbol{\lambda}) \quad (6.9)$$

Στη συνέχεια αναδρομικά κατασκευάζεται ο αλγόριθμος που παρουσιάζεται παρακάτω:

1. Αρχικοποίηση:

$$\alpha_1(i) = \pi_i b_i(\mathbf{o}_1), \quad 1 \leq i \leq N \quad (6.10)$$

2. Αναδρομή:

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^N \alpha_t(i) a_{ij} \right] b_j(\mathbf{o}_{t+1}), \quad \begin{cases} 1 \leq t \leq T-1 \\ 1 \leq j \leq N \end{cases} \quad (6.11)$$

3. Τερματισμός:

$$Pr(\mathbf{O}|\boldsymbol{\lambda}) = \sum_{i=1}^N \alpha_T(i) \quad (6.12)$$

Αντίστοιχα, στον backward αλγόριθμο ορίζεται η μεταβλητή $\beta_t(i)$, η οποία εκφράζει την πιθανότητα ενώ βρισκόμαστε στην κατάσταση q_i να παρουσιαστεί μια ακολουθία παρατηρήσεων από τη χρονική στιγμή $t + 1$ μέχρι το τέλος T , δεδομένου του μοντέλου λ :

$$\beta_t(i) = Pr(\mathbf{o}_{t+1}, \mathbf{o}_{t+2}, \dots, \mathbf{o}_T, q_t = i | \lambda) \quad (6.13)$$

Όπως και προηγουμένως, κατασκευάζεται αναδρομικά ο αλγόριθμος που παρουσιάζεται παρακάτω:

1. Αρχικοποίηση:

$$\beta_T(i) = 1, \quad 1 \leq i \leq N \quad (6.14)$$

2. Αναδρομή:

$$\beta_t(i) = \sum_{j=1}^N a_{ij} b_j(\mathbf{o}_{t+1}) \beta_{t+1}(j), \quad \begin{cases} t = T - 1, T - 2, \dots, 1 \\ 1 \leq i \leq N \end{cases} \quad (6.15)$$

3. Τερματισμός:

$$Pr(\mathbf{O} | \lambda) = \sum_{i=1}^N \beta_1(i) \quad (6.16)$$

6.2.3 Εκπαίδευση των παραμέτρων του μοντέλου

Ένα άλλο πιο δύσκολο και σύνθετο πρόβλημα είναι αυτό της εκπαίδευσης των HMM, δηλαδή της εύρεσης των βέλτιστων παραμέτρων που μεγιστοποιούν την πιθανοφάνεια του μοντέλου $Pr(\mathbf{O} | \lambda)$ για ένα σύνολο εκπαίδευσης. Παρόλο που δεν υπάρχει κάποιος αναλυτικός τρόπος υπολογισμού των παραμέτρων αυτών, ο προσδιορισμός τους επιτυγχάνεται μέσω του αλγορίθμου Baum-Welch, ο οποίος αποτελεί ειδική περίπτωση του EM (Expectation-Maximization) αλγορίθμου. Πρέπει να αναφερθεί ότι ο παραπάνω αλγόριθμος δεν βρίσκει το ολικό αλλά το τοπικό μέγιστο της πιθανότητας $Pr(\mathbf{O} | \lambda)$.

Αρχικά, ορίζεται η μεταβλητή $\xi_t(i, j)$ η οποία εκφράζει την πιθανότητα, δεδομένου του μοντέλου λ και της ακολουθίας παρατηρήσεων \mathbf{O} , να βρισκείται το HMM τη χρονική στιγμή t στην κατάσταση q_i και τη χρονική στιγμή $t + 1$ στην κατάσταση q_j :

$$\xi_t(i, j) = Pr(q_t = i, q_{t+1} = j | \mathbf{O}, \lambda) \quad (6.17)$$

Η παραπάνω πιθανότητα συναρτήσει των μεταβλητών α_t και β_t γράφεται:

$$\xi_t(i, j) = \frac{Pr(q_t = i, q_{t+1} = j, \mathbf{O} | \boldsymbol{\lambda})}{Pr(\mathbf{O} | \boldsymbol{\lambda})} = \quad (6.18)$$

$$= \frac{\alpha_t(i) a_{ij} b_j(\mathbf{o}_{t+1}) \beta_{t+1}(j)}{\sum_{i=1}^N \sum_{j=1}^N \alpha_t(i) a_{ij} b_j(\mathbf{o}_{t+1}) \beta_{t+1}(j)} \quad (6.19)$$

Επίσης, ορίζεται η μεταβλητή $\gamma_t(i)$ που εκφράζει την πιθανότητα να βρίσκεται το μοντέλο τη χρονική στιγμή t στην κατάσταση q_i , η οποία θα δίνεται από τη σχέση:

$$\gamma_t(i) = \sum_{j=1}^N \xi_t(i, j) \quad (6.20)$$

Ο αλγόριθμος Baum -Welch μεγιστοποιεί ως προς τις παραμέτρους $\bar{\boldsymbol{\lambda}}$ την παρακάτω βοήθητική συνάρτηση:

$$Q(\boldsymbol{\lambda}, \bar{\boldsymbol{\lambda}}) = \sum_{\mathbf{q}} Pr(\mathbf{q} | \mathbf{O}, \boldsymbol{\lambda}) \log [Pr(\mathbf{O}, \mathbf{q} | \bar{\boldsymbol{\lambda}})] \quad (6.21)$$

Η μεγιστοποίηση της συνάρτησης $Q(\boldsymbol{\lambda}, \bar{\boldsymbol{\lambda}})$ οδηγεί σε νέες βελτιωμένες τιμές παραμέτρων $\bar{\boldsymbol{\lambda}}$ που αυξάνουν την πιθανοφάνεια $Pr(\mathbf{O}, \mathbf{q} | \bar{\boldsymbol{\lambda}})$ του μοντέλου:

$$\max_{\bar{\boldsymbol{\lambda}}} [Q(\boldsymbol{\lambda}, \bar{\boldsymbol{\lambda}})] \Rightarrow Pr(\mathbf{O} | \bar{\boldsymbol{\lambda}}) \geq Pr(\mathbf{O} | \boldsymbol{\lambda}) \quad (6.22)$$

Οι ανανεωμένες τιμές των παραμέτρων του μοντέλου που προκύπτουν από τον παραπάνω αλγόριθμο δίνονται από τις σχέσεις:

$$\bar{a}_{ij} = \frac{\sum_{t=1}^{T-1} \xi_t(i, j)}{\sum_{t=1}^{T-1} \gamma_t(i)} \quad (6.23)$$

$$\bar{b}_j(k) = \frac{\sum_{t=1|o_t=v_k}^T \gamma_t(j)}{\sum_{t=1}^T \gamma_t(j)} \quad (6.24)$$

$$\bar{\pi}_i = \gamma_1(i) \quad (6.25)$$

Στην περίπτωση που η συνάρτηση κατανομής της πιθανότητας εξαγωγής μιας παρατήρησης αποτελείται από μείγμα γκαουσιανών συναρτήσεων ορίζεται η γενικευμένη μεταβλητή $\gamma_t(j, k)$ που εκφράζει την πιθανότητα να βρίσκεται το μοντέλο τη χρονική στιγμή t στην κατάσταση q_j και η k γκαουσιανή συνιστώσα να ευθύνεται για την παρατήρηση \mathbf{o}_t :

$$\gamma_t(j, k) = \left[\frac{\alpha_t(j)\beta_t(j)}{\sum_{j=1}^N \alpha_t(j)\beta_t(j)} \right] \cdot \left[\frac{c_{jk}\mathcal{N}(\mathbf{o}_t, \boldsymbol{\mu}_{jk}, \mathbf{U}_{jk})}{\sum_{m=1}^M c_{jm}\mathcal{N}(\mathbf{o}_t, \boldsymbol{\mu}_{jm}, \mathbf{U}_{jm})} \right] \quad (6.26)$$

Οι τιμές των παραμέτρων του μείγματος γκαουσιανών που προκύπτουν από τον αλγόριθμο βελτιστοποίησης θα δίνονται από τις σχέσεις:

$$\bar{c}_{jk} = \frac{\sum_{t=1}^T \gamma_t(j, k)}{\sum_{t=1}^T \sum_{k=1}^M \gamma_t(j, k)} \quad (6.27)$$

$$\bar{\boldsymbol{\mu}}_{jk} = \frac{\sum_{t=1}^T \gamma_t(j, k) \cdot \mathbf{o}_t}{\sum_{t=1}^T \gamma_t(j, k)} \quad (6.28)$$

$$\bar{\mathbf{U}}_{jk} = \frac{\sum_{t=1}^T \gamma_t(j, k) \cdot (\mathbf{o}_t - \bar{\boldsymbol{\mu}}_{jk})(\mathbf{o}_t - \bar{\boldsymbol{\mu}}_{jk})'}{\sum_{t=1}^T \gamma_t(j, k)} \quad (6.29)$$

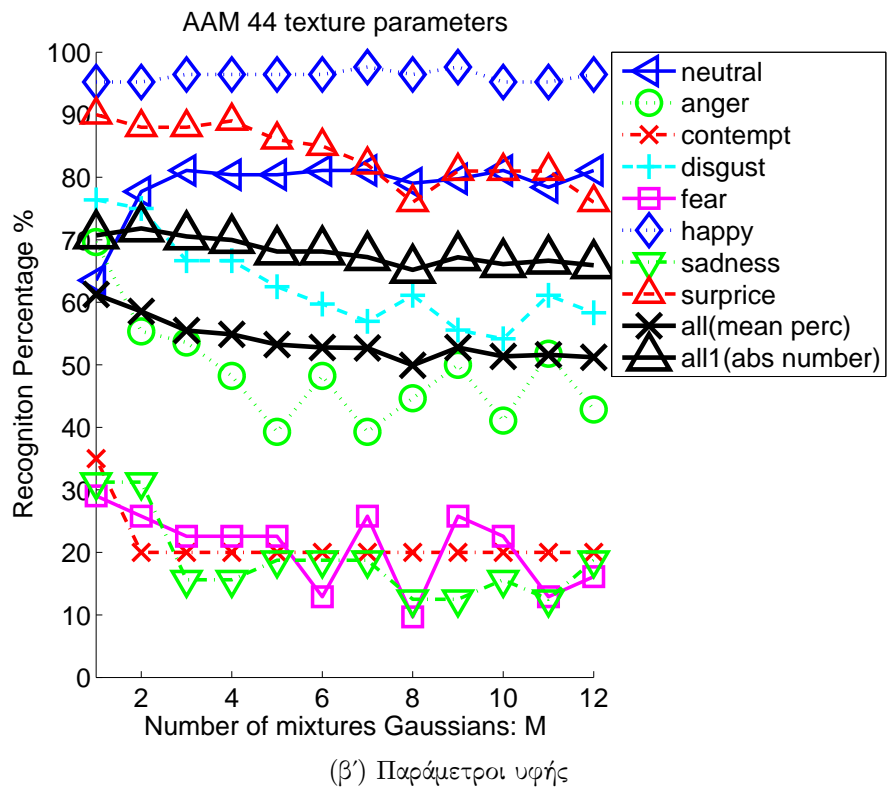
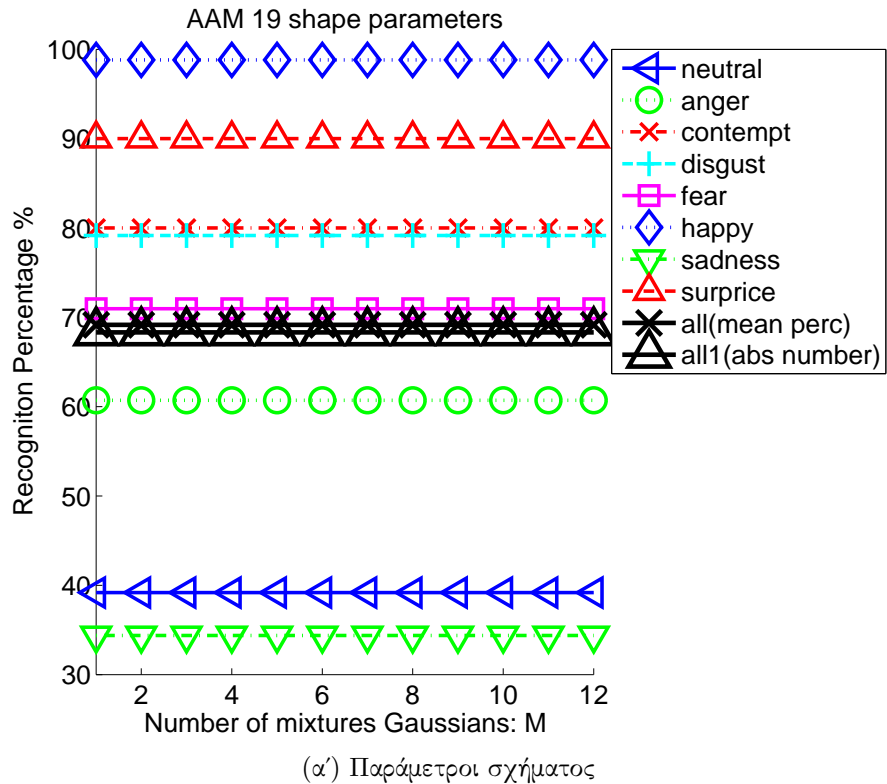
6.3 Πειραματικά Αποτελέσματα

6.3.1 Κατηγοριοποίηση με χρήση GMMs

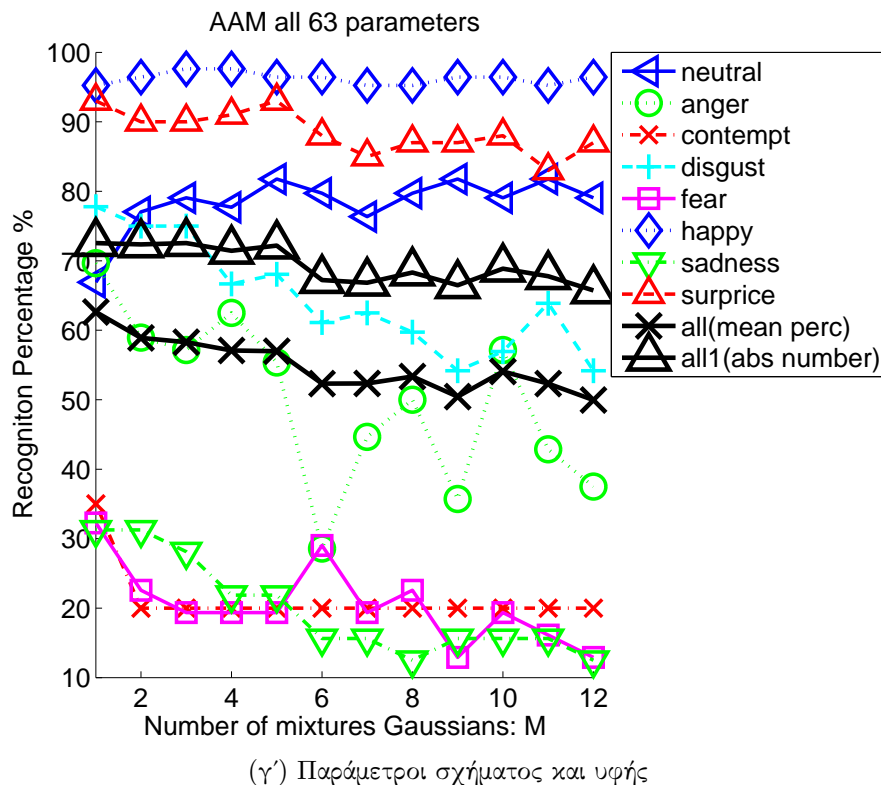
Αρχικά, θα προχωρήσουμε στην κατηγοριοποίηση του συναισθήματος με χρήση μείγματος γκαουσιανών συναρτήσεων (GMMs). Σαν βάση δεδομένων θα χρησιμοποιηθούν καρέ εικόνων από βίντεο της επισημασμένης βάσης δεδομένων “The extended Cohn-Kanade (CK+)” του Carnegie Mellon University (CMU)[18].

Ο διαχωρισμός θα γίνει σε 8 κλάσεις που αντιπροσωπεύουν τις 7 βασικές κατηγορίες συναισθημάτων (θυμός, περιφρόνηση, αηδία, φόβος, ευτυχία, θλίψη, έκπληξη) καθώς και το ουδέτερο συναίσθημα. Σαν διάνυσμα παραμέτρων θα χρησιμοποιηθούν οι παράμετροι σχήματος και υψής, ξεχωριστά αλλά και συνδυασμένα, που προκύπτουν από την προσαρμογή AAM στο τελευταίο καρέ του κάθε βίντεο, το οποίο και εκφράζει την κορύφωση του κάθε συναισθήματος.

Πιο συγκεκριμένα, για κάθε κατηγορία συναισθήματος εκπαιδεύεται ένα ξεχωριστό GMM, σύμφωνα με τις παραμέτρους του AAM, έτσι ώστε να μεγιστοποιείται η λογαριθμική πιθανοφάνεια του μοντέλου σε ολόκληρο το σύνολο εκπαίδευσης. Στη συνέχεια, για τις παραμέτρους κάθε νέας εικόνας υπολογίζεται η πιθανοφάνεια για κάθε ένα από τα 8 GMMs και η εικόνα κατατάσσεται στην κατηγορία συναισθήματος εκείνη της οποίας το αντίστοιχο GMM εμφανίζει τη μεγαλύτερη λογαριθμική πιθανοφάνεια.



Σχήμα 6.2: Αποτελέσματα κατηγοριοποίησης συναισθήματος με χρήση GMMs για τη βάση δεδομένων “CK+” του CMU σε σχέση με τον αριθμό των γκαουσιανών.



Σχήμα 6.2: Αποτελέσματα κατηγοριοποίησης συναισθήματος με χρήση GMMs για τη βάση δεδομένων “CK+” του CMU σε σχέση με τον αριθμό των γκαουσιανών (Συνέχεια).

Στο Σχήμα 6.2 παρουσιάζονται τα ποσοστά ορθής κατηγοριοποίησης με χρήση GMMs για κάθε διαφορετικό συναίσθημα σε σχέση με τον αριθμό των γκαουσιανών συναρτήσεων που χρησιμοποιούμε κάθε φορά. Σαν all(mean perc) ορίζεται ο μέσος όρος των επιμέρους ποσοστών κάθε κατηγορίας συναισθήματος, ενώ με τον όρο all1(abs number) αναφερόμαστε στο πόσοστο ορθής κατηγοριοποίησης για το σύνολο των εικόνων ανεξαρτήτως της κατηγορίας στην οποία ανήκουν. Επειδή, ο αριθμός των δεδομένων δεν είναι ο ίδιος για κάθε κατηγορία συναισθήματος οι δύο αυτοί μέσοι όροι μπορούν να προκύψουν διαφορετικοί.

Αρχικά, όπως φαίνεται και στο Σχήμα 6.2α' τα αποτελέσματα στην περίπτωση που χρησιμοποιούμε σαν είσοδο των GMMs τις παραμέτρους σχήματος μόνο προκύπτουν ανεξάρτητα του αριθμού των γκαουσιανών. Επίσης, παρατηρούμε πως τα συναισθήματα “ευτυχία” και “έκπληξη” αναγνωρίζονται με πολύ μεγάλη επιτυχία, ανεξάρτητα από το διάλυμα παραμέτρων που χρησιμοποιείται κάθε φορά. Αυτό οφείλεται στο γεγονός πως τα δύο αυτά συναισθήματα εμφανίζουν μεγάλες μεταβολές στο σχήμα του προσώπου και ιδιαίτερα στην περιοχή του στόματος.

Αντίθετα, σε λιγότερο εκφραστικά συναισθήματα όπως η “θλίψη”, ο “θυμός” και το ουδέτερο συναίσθημα η κατηγοριοποίηση με βάση τις παραμέτρους σχήματος γίνεται δυσκολότερη και με χειρότερα αποτελέσματα. Ωστόσο, στην περίπτωση του ουδέτερου συναισθήματος αν λάβουμε υπόψιν και τις παραμέτρους υψής τότε το ποσοστό αναγνώρισής του αυξάνεται σε αρκετά μεγάλο βαθμό (Σχήματα 6.2β’ & 6.2γ’).

Με τον τρόπο αυτό, όμως, περιλαμβάνεται στη ταξινόμηση και ένα μεγάλο ποσοστό πληροφορίας που έχει να κάνει με την ταυτότητα του προσώπου της εικόνας, όπως το φύλο, το χρώμα, η φυλετική καταγωγή. το γεγονός αυτό έχει σαν συνέπεια να μειώνονται τα ποσοστά σε άλλες κατηγορίες συναισθημάτων, όπως ο “φόβος” και η “περιφρόνηση”, αφού λόγω της αυξημένης πληροφορίας για την υψή αναγνωρίζονται σαν ουδέτερο συναίσθημα.

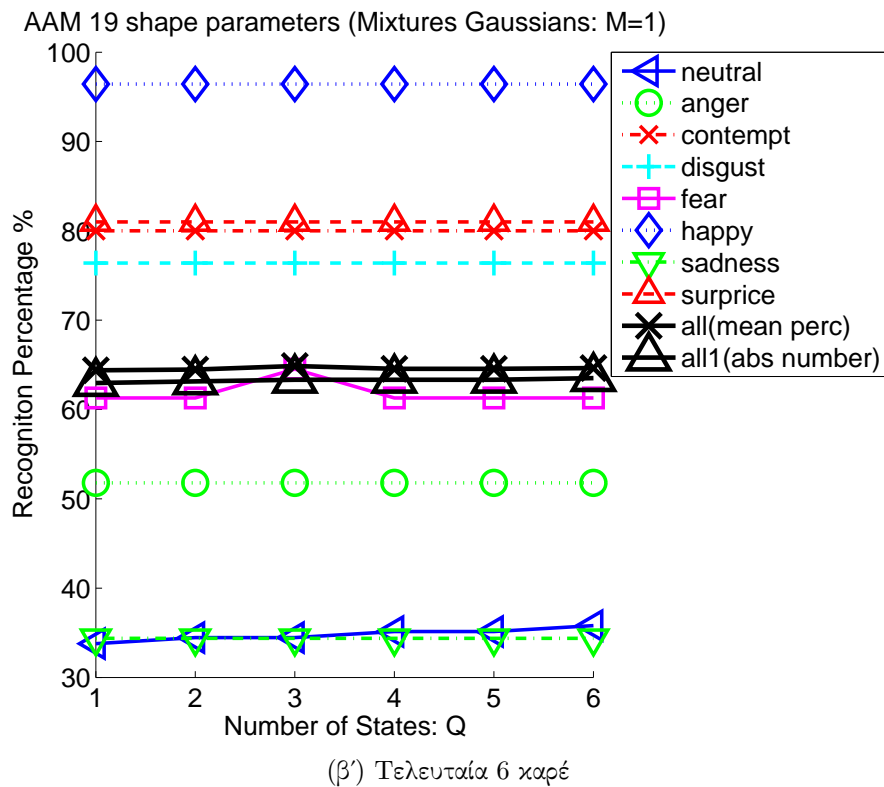
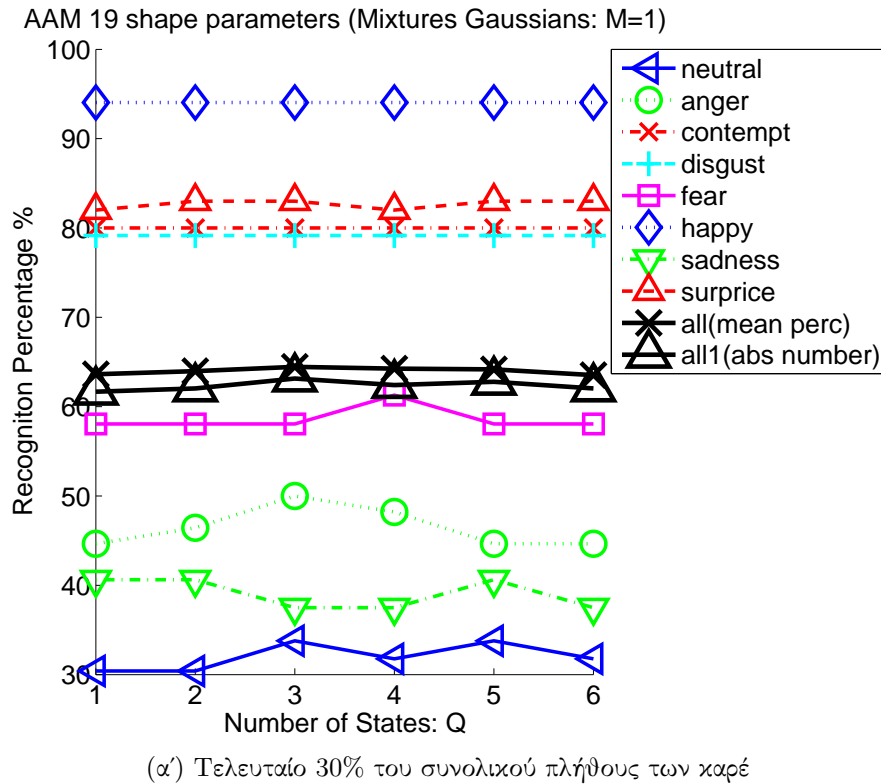
Άρα γίνεται φανερό πως χρησιμοποιώντας τις παραμέτρους σχήματος μόνο καταφέρνουμε να πάρουμε το καλύτερο δυνατόν αποτέλεσμα για τις περισσότερες κατηγορίες συναισθημάτων, γεγονός που επιβεβαιώνεται και από τους μέσους όρους all(mean perc) και all1(abs number).

6.3.2 Κατηγοριοποίηση με χρήση HMMs

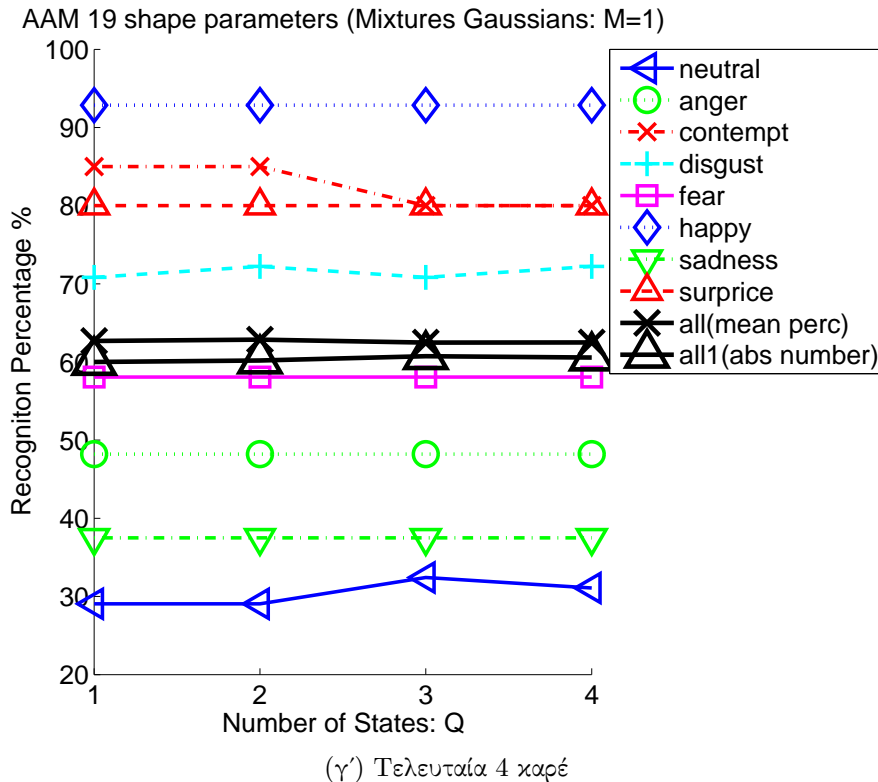
Στη συγκεκριμένη υποενότητα ο διαχωρισμός των συναισθημάτων σε κλάσεις θα γίνει με τη χρήση Κρυφών Μαρκοβιανών Μοντέλων (HMMs). Η διαδικασία που ακολουθουθείται είναι η ίδια με πριν μόνο που τώρα για κάθε μια κατηγορία συναισθήματος θα πρέπει να εκπαιδεύσουμε ένα ξεχωριστό HMM με συγκεκριμένο αριθμό καταστάσεων Q και πλήθος γκαουσιανών συναρτήσεων M για κάθε κατάσταση.

Σαν χαρακτηριστικά αναγνώρισης θα χρησιμοποιηθούν οι παράμετροι σχήματος του AMM, μιας και μας έδωσαν τα καλύτερα αποτελέσματα στην περίπτωση των GMMs. Σε αντίθεση όμως με την περίπτωση των GMMs όπου λαμβάναμε υπόψιν μόνο το τελευταίο καρέ, οι παράμετροι σχήματος θα υπολογιστούν για ένα σύνολο από τα τελευταία καρέ του βίντεο προκειμένου να μοντελοποιηθεί έτσι η χρονική εξέλιξη της κορύφωσης του συναισθήματος από τις καταστάσεις του HMM. Συγκεκριμένα, επιλέγονται 3 διαφορετικές περιπτώσεις σχετικά με τον πλήθος των καρέ: τελευταίο 30% του συνολικού τους πλήθους, τελευταία 6 και τελευταία 4.

Στο Σχήμα 6.3 παρουσιάζονται τα ποσοστά ορθής κατηγοριοποίησης για τις 3 παραπάνω περιπτώσεις σε σχέση με τον αριθμό των καταστάσεων των HMMs και για πλήθος γκαουσιανών συναρτήσεων για κάθε κατάσταση $M = 1$. Για την εκπαίδευση των HMMs χρησιμοποιήθηκε ο ανοικτός κώδικας HMM toolbox του Kevin Murphy.[22] Παρατηρούμε πως τα αποτελέσματα, όπως και στην περίπτωση των GMMs, σε ορισμένα συναισθήματα όπως “ευτυχία”, “έκπληξη” και περιφρόνηση αναγνωρίζονται με μεγάλη επιτυχία ενώ άλλα όπως “θλίψη”, ο “θυμός” και το ουδέτερο συναίσθημα δεν αναγνωρίζονται με ικανοποιητική ακρίβεια.



Σχήμα 6.3: Αποτελέσματα κατηγοριοποίησης συναισθήματος με βάση τις παραμέτρους σχήματος και χρήση HMMs για τη βάση δεδομένων “CK+” του CMU σε σχέση με τον αριθμό των καταστάσεων και για $M = 1$.

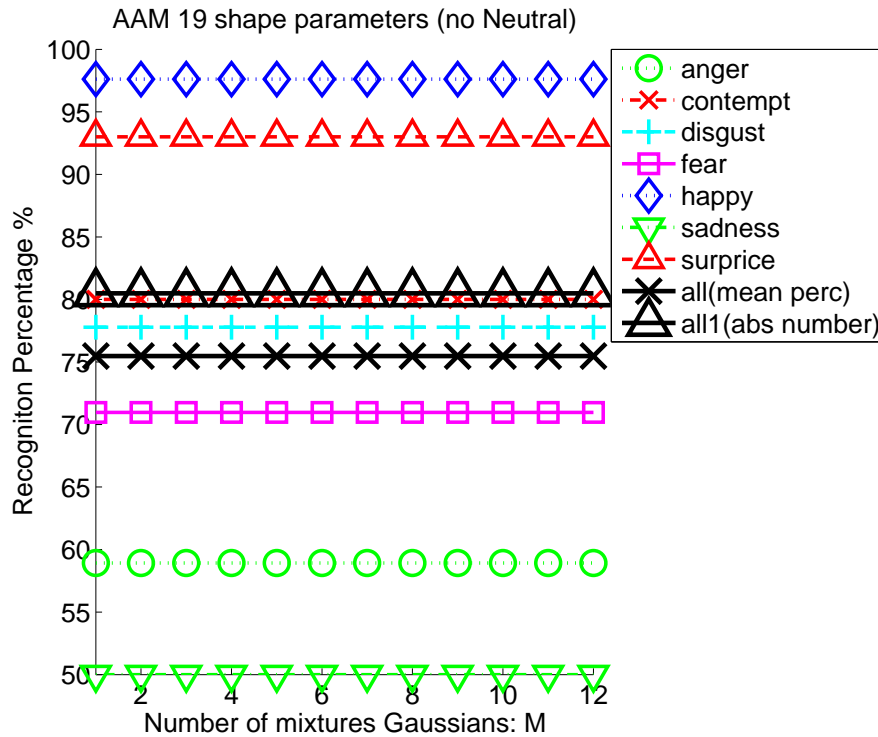


Σχήμα 6.3: Αποτελέσματα κατηγοριοποίησης συναισθήματος με βάση τις παραμέτρους σχήματος και χρήση HMMs για τη βάση δεδομένων “CK+” του CMU σε σχέση με τον αριθμό των καταστάσεων και για $M = 1$ (Συνέχεια).

Επίσης, το μέσο ποσοστό αναγνώρισης κυμαίνεται γύρω στο 60%, δηλαδή πιο χαμηλά από το 70% που είχαμε στην κατηγοριοποίηση με χρήση GMMs.

Τέλος, όσο αφορά το σύνολο των καρτέ που συμπεριλήφθηκε στην αναγνώριση δεν παρατηρούνται μεγάλες διαφορές στα αποτελέσματα, με τα 6 καρτέ, ωστόσο, να δίνουν ελαφρώς καλύτερα ποσοστά αναγνώρισης για περισσότερες κατηγορίες αποτελεσμάτων. Αυτό οφείλεται στο γεγονός πως καθώς στα τελευταία καρτέ πλησιάζουμε στην κορύφωση του συναισθήματος οι διαφορές στις εκφράσεις του προσώπου είναι αρκετά μεγάλες σε σχέση με τα αρχικά και ενδιάμεσα καρτέ.

Αξίζει να τονιστεί πως αν συμπεριλάβουμε στην αναγνώριση αρκετά μεγάλο αριθμό καρτέ (από 70% του συνολικού αριθμού και πάνω), τότε τα ποσοστά ορθής κατηγοριοποίησης μειώνονται ραγδαία με μέσο όρο γύρω κάτω από το 50%, μιας και με τον τρόπο αυτό πλησιάζουμε όλο και περισσότερο προς το ουδέτερο συναίσθημα.



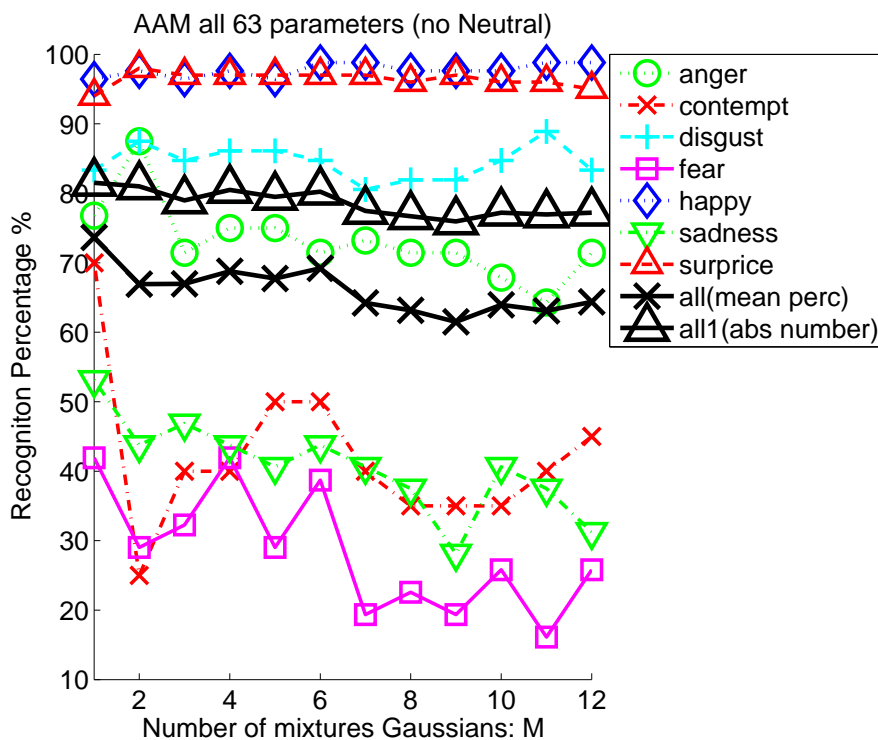
(α') Παράμετροι σχήματος

Σχήμα 6.4: Αποτελέσματα κατηγοριοποίησης συναισθήματος με χρήση GMMs για τη βάση δεδομένων “CK+” του CMU σε σχέση με τον αριθμό των γκαουσιανών χωρίς να συμπεληφθεί το ουδέτερο συναίσθημα.

6.3.3 Κατηγοριοποίηση χωρίς τη χρήση του ουδετέρου

Όπως είχαμε ήδη επισημάνει και στην κατηγοριοποίηση με GMMs, το ουδέτερο συναίσθημα δεν μπορεί να αναγνωριστεί με επαρκή ακρίβεια από τις παραμέτρους σχήματος του AAM αποκλειστικά. Στην περίπτωση που συμπεριληφθούν και οι παράμετροι υψής ναι μεν το ποσοστό ορθής κατηγοριοποίησής του αυξάνει σε μεγάλο βαθμό, ωστόσο μειώνονται τα ποσοστά από άλλες κατηγορίες συναισθημάτων οι οποίες χαρακτηρίζονται εσφαλμένα σαν ουδέτερο συναίσθημα.

Αυτό μας οδηγεί στην επανάληψη των παραπάνω πειραμάτων χωρίς να συμπεριλάβουμε αυτή τη φορά το ουδέτερο συναίσθημα σαν κλάση κατηγοριοποίησης. Στο Σχήμα 6.4 παρουσιάζονται τα ποσοστά ορθής κατηγοριοποίησης με χρήση GMMs χωρίς να συμπεριλάβουμε στις κλάσεις το ουδέτερο συναίσθημα. Παρατηρούμε πως όσο αναφορά την αναγνώριση με βάση τις παραμέτρους σχήματος τα ποσοστά στις περισσότερες κατηγορίες εμφανίζονται ελαφρώς αυξημένα (με το χειρότερο να βρίσκεται λίγο πάνω από το 50%), γεγονός που αντικατοπτρίζεται και



(β') Παράμετροι σχήματος και υφής

Σχήμα 6.4: Αποτελέσματα κατηγοριοποίησης συναισθήματος με χρήση GMMs για τη βάση δεδομένων “CK+” του CMU σε σχέση με τον αριθμό των γκαουσιανών χωρίς να συμπληρωθεί το ουδέτερο συναίσθημα (Συνέχεια).

στους αντίστοιχους μέσους όρους.

Στην περίπτωση τώρα που σαν διάνυσμα χαρακτηριστικών αναγνώρισης χρησιμοποιηθούν οι συνολικοί παράμετροι (σχήματος και υφής) του AAM, παρουσιάζεται μια σαφής βελτίωση στα αποτελέσματα της αναγνώρισης για τις περισσότερες κατηγορίες σε σχέση με εκείνα της υποένότητας 6.3.1. Παρόλα, αυτά συνεχίζει να υπάρχει μια μείωση στα ποσοστά που αφορούν τις κατηγορίες “φόβος” και “περιφρόνηση” από εκείνα που προκύπτουν μόνο από τις παραμέτρους σχήματος. Αντίθετα, σημειώνεται σημαντική αύξηση στα ποσοστά που αφορούν το “θυμό” και την “αηδία”.

6.4 Τροποποιήσεις της μεθόδου κατηγοριοποίησης

6.4.1 Συνδυασμός ανεξάρτητων μοντέλων για το σχήμα και την υφή

Όπως είδαμε στην ενότητα 6.3.3, ακόμα και αν παραλείψουμε την κλάση του ουδέτερου συναισθήματος, όταν συμπεριλαμβάνουμε στην αναγνώριση τις παραμέτρους υφής τα ποσοστά κάποιων κλάσεων μειώνονται αισθητά ενώ σε κάποια άλλα παρατηρείται αύξηση. Στόχος είναι να μπορέσουμε να εκμεταλλευτούμε τη βελτίωση που μας προσφέρει η χρήση των παραμέτρων υφής χωρίς να χειροτερέψουμε το αποτέλεσμα που προκύπτει αποκλειστικά από τις παραμέτρους σχήματος.

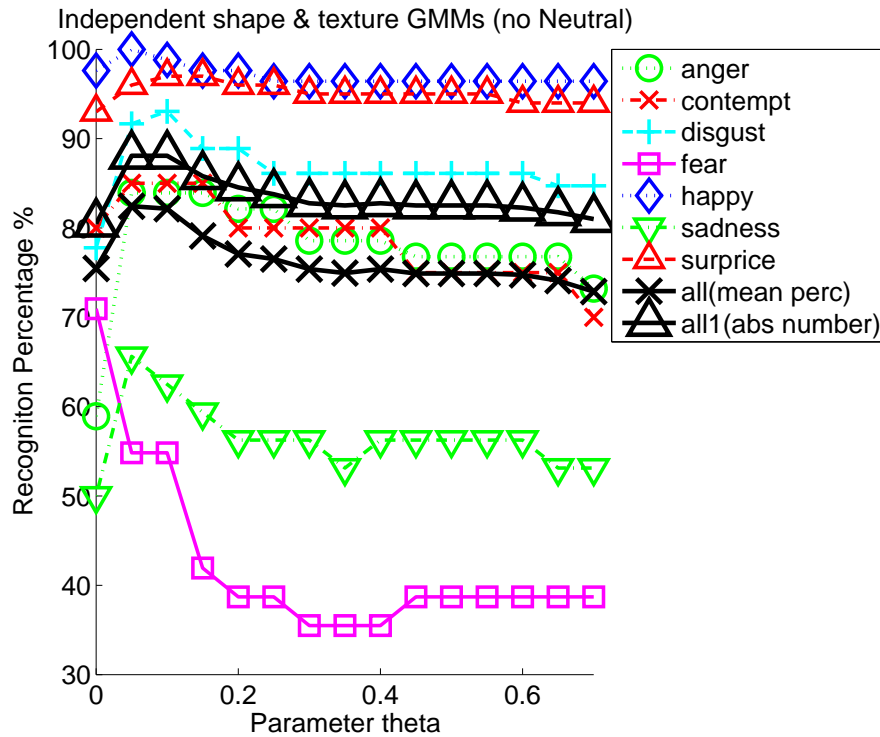
Για το σκοπό αυτό, για κάθε κατηγορία συναισθήματος εκπαιδεύουμε δύο ανεξάρτητα μοντέλα (GMM ή HMM) ένα για τις παραμέτρους σχήματος με πιθανοφάνεια $\ln(Pr_{shape}^{i=1:7}(\mathbf{O}|\boldsymbol{\lambda}))$ και ένα για τις παραμέτρους υφής αντίστοιχα με πιθανοφάνεια $\ln(Pr_{texture}^{i=1:7}(\mathbf{O}|\boldsymbol{\lambda}))$. Ο δείκτης $i = 1 : 7$ συμβολίζει την κλάση που αντιστοιχεί σε κάθε ένα από τα 7 βασικά συναισθήματα: 1[θυμός], 2[περιφρόνηση], 3[αηδία], 4[φόβος], 5[ευτυχία], 6[θλίψη], 7[έκπληξη].

Στη συνέχεια, για κάθε νέα εικόνα υπολογίζεται η πιθανοφάνεια του μοντέλου σχήματος για κάθε κλάση i και γίνεται κανονικοποίηση της ως προς τη μέγιστη τιμή της. Η ίδια διαδικασία επαναλαμβάνεται και για το μοντέλο υφής. Στο τέλος, η νέα εικόνα κατατάσσεται στην κατηγορία εκείνη η οποία εμφανίζει τη μεγαλύτερη συνολική πιθανοφάνεια:

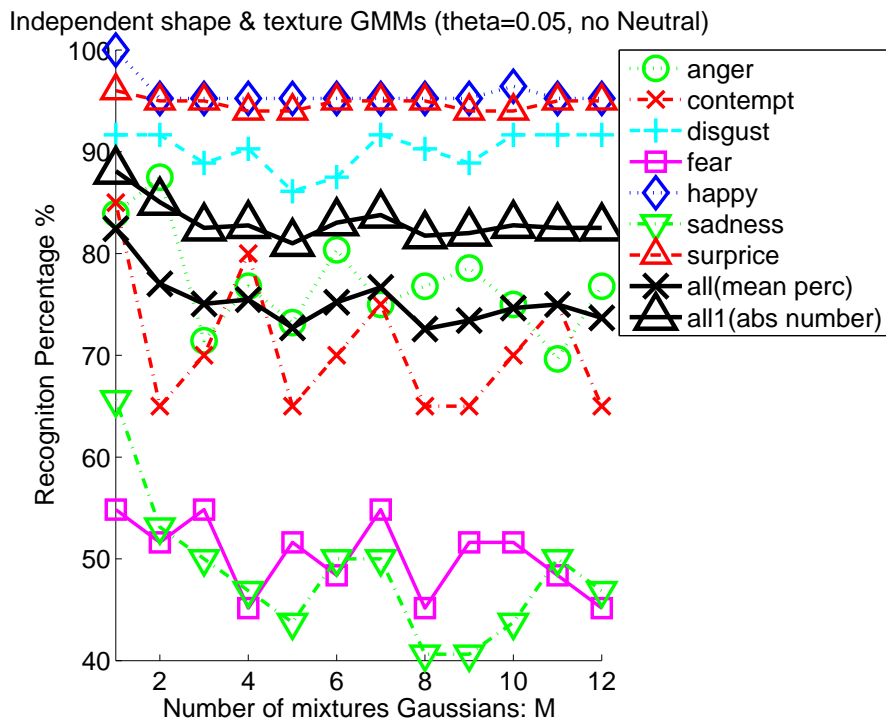
$$\ln(Pr_{all}^{i=1:7}(\mathbf{O}|\boldsymbol{\lambda})) = (1 - \theta) \frac{\ln(Pr_{shape}^{i=1:7}(\mathbf{O}|\boldsymbol{\lambda}))}{\max_i(\ln(Pr_{shape}^{i=1:7}(\mathbf{O}|\boldsymbol{\lambda})))} + \theta \frac{\ln(Pr_{texture}^{i=1:7}(\mathbf{O}|\boldsymbol{\lambda}))}{\max_i(\ln(Pr_{texture}^{i=1:7}(\mathbf{O}|\boldsymbol{\lambda})))} \quad (6.30)$$

Η τιμή $\theta \in [0, 1]$ ορίζει το ποσοστό με το οποίο θα συνεισφέρουν οι παράμετροι υφής στο τελικό αποτέλεσμα της αναγνώρισης. Έτσι με τα δύο ανεξάρτητα αυτά μοντέλα γίνεται εφικτό να δώσουμε διαφορετική βαρύτητα στο κάθε είδος παραμέτρων κάτι το οποίο δεν θα μπορούσε να γίνει με το ένα καθολικό μοντέλο. Στο Σχήμα 6.5 παρουσιάζονται τα ποσοστά αναγνώρισης με χρήση ανεξάρτητων GMMs για το σχήμα και την υφή για κάθε κατηγορία συναισθήματος σε σχέση με τη μεταβολή της παραμέτρου θ . Παρατηρούμε, πως τα καλύτερα αποτελέσματα προκύπτουν για αρκετά μικρές τιμές του θ , δηλαδή για ένα συνολικό μοντέλο αρκετά κοντά σε εκείνο που χρησιμοποιεί αποκλειστικά τις παραμέτρους σχήματος.

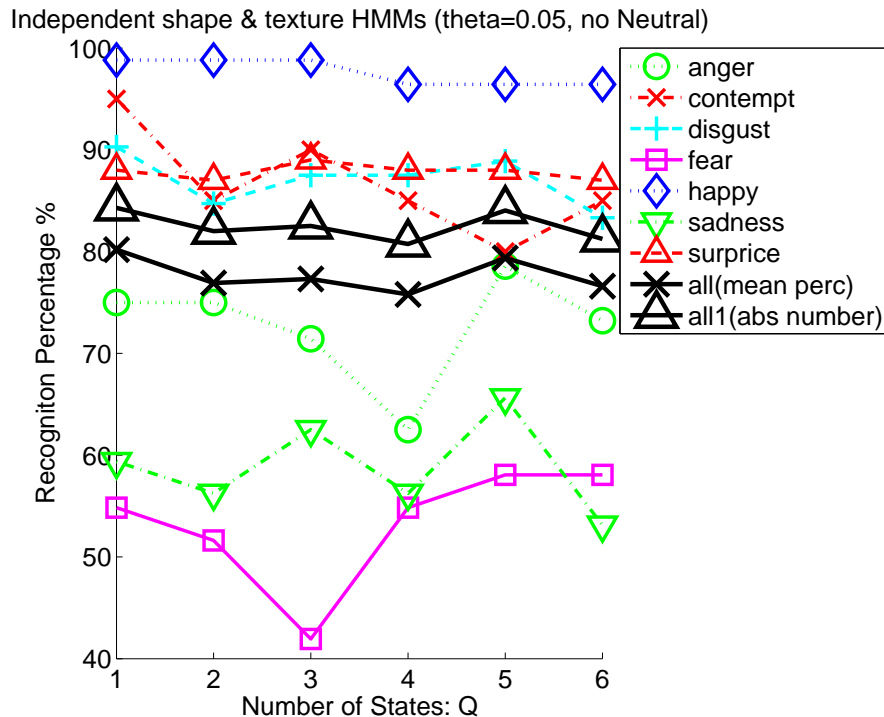
Στο Σχήμα 6.6 παρουσιάζονται τα ποσοστά ορθής κατηγοριοποίησης με χρήση ανεξάρτητων GMMs για το σχήμα και την υφή για κάθε διαφορετικό συναίσθημα σε σχέση με τον αριθμό των γκαουσιανών συναρτήσεων και για $\theta = 0.05$. Από τα δύο παραπάνω σχήματα παρατηρούμε πως σχεδόν όλες οι κατηγορίες συναισθημάτων



Σχήμα 6.5: Αποτελέσματα κατηγοριοποίησης συναισθήματος με χρήση ανεξάρτητων GMMs για το σχήμα και την υφή, για τη βάση δεδομένων “CK+” του CMU, σε σχέση με τη μεταβολή της παραμέτρου θ και για αριθμό γκαουσιανών $M = 1$.



Σχήμα 6.6: Αποτελέσματα κατηγοριοποίησης συναισθήματος με χρήση ανεξάρτητων GMMs για το σχήμα και την υφή, για τη βάση δεδομένων “CK+” του CMU, σε σχέση με τον αριθμό των γκαουσιανών και για $\theta = 0.05$.



Σχήμα 6.7: Αποτελέσματα κατηγοριοποίησης συναισθήματος με χρήση ανεξάρτητων HMMs για το σχήμα και την υφή, για τη βάση δεδομένων “CK+” του CMU, σε σχέση με τον αριθμό των καταστάσεων, κρατώντας τα τελευταία 6 καρέ και για $\theta = 0.05$ και $M = 1$.

αναγνωρίζονται με αρκετά μεγάλη ακρίβεια και με τους μέσους όρους να βρίσκονται μεταξύ 80% και 90%. Επίσης, βλέπουμε πως τα αρνητικά στοιχεία που προέκυπταν από τη εισαγωγή των παραμέτρων υφής έχουν μειωθεί σε ικανοποιητικό βαθμό χωρίς ωστόσο να χάσουμε τις θετικές επιδράσεις τους.

Στη συνέχεια, στο Σχήμα 6.7 παρουσιάζονται τα ποσοστά αναγνώρισης που προκύπτουν αν αντί για GMMs χρησιμοποιήσουμε HMMs κρατώντας τα τελευταία 6 καρέ και για $\theta = 0.05$ και $M = 1$. Όπως μπορούμε να δούμε, τα αποτελέσματα που προκύπτουν είναι πολύ καλύτερα από εκείνα της ενότητας 6.3.2, όπου χρησιμοποιήθηκαν οι παράμετροι σχήματος μόνο, και συγκρίσιμα με τα αποτελέσματα που είδαμε παραπάνω με τη χρήση GMMs.

6.4.2 Προσαρμογή σε συγκεκριμένο χρήστη

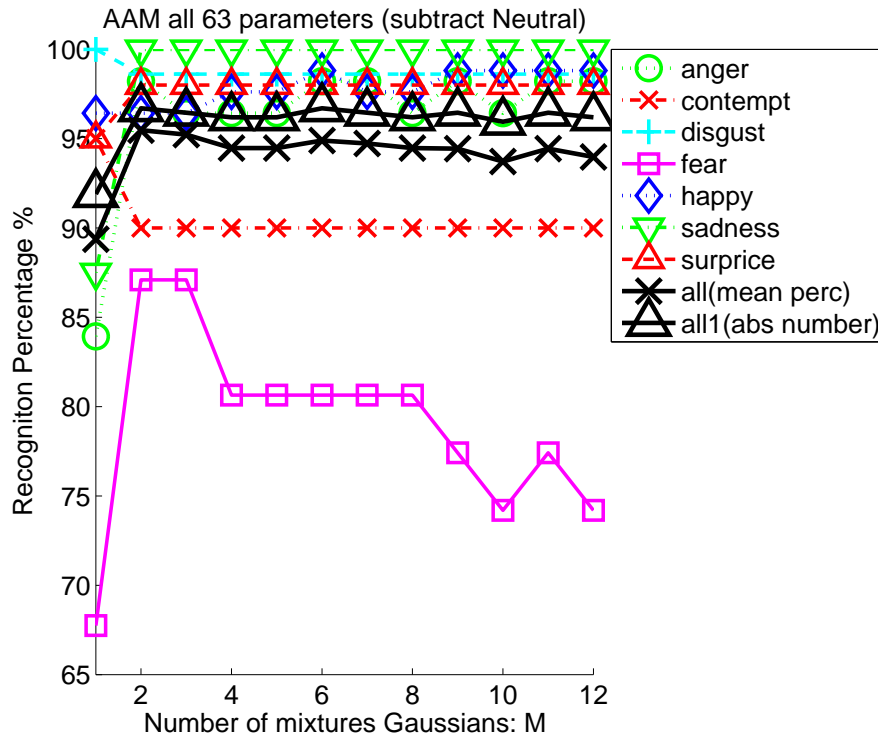
Όπως έχει ήδη αναφερθεί και σε προηγούμενες ενότητες, ένα σημαντικό μειονέκτημα της συνολικής μεθόδου με την οποία γίνεται η κατηγοριοποίηση των συναισθημάτων σε κλάσεις είναι ότι οι παράμετροι υψής κύριως αλλά και σχήματος όπως προκύπτουν από την προσαρμογή ενός AAM δεν περιέχουν αποκλειστικά πληροφορία σχετικά με την εκδήλωση του συναισθήματος. Έτσι είναι δυνατόν να περιλαμβάνεται πληροφορία σχετικά με την ταυτότητα του ατόμου που απεικονίζεται στο βίντεο και έχουν να κάνουν με το φύλο, το χρώμα, την ηλικία, τη φυλετική καταγωγή κτλ. Με την εκπαίδευση των μοντέλων με εικόνες από διαφορετικούς χρήστες μπορούμε να αντιμετωπίσουμε σε ένα βαθμό αυτό το πρόβλημα αλλά όχι όσο καλά θα θέλαμε, όπως μας έδειξαν και τα παραπάνω πειράματα.

Μια άλλη προσέγγιση είναι να προσπαθήσουμε να αφαιρέσουμε την επίδραση παραγόντων που δεν σχετίζονται με το συναίσθημα πριν προχωρήσουμε στην κατηγοριοποίηση. Μπορούμε να θεωρήσουμε ότι η έκφραση του προσώπου στο ουδέτερο συναίσθημα περιέχει όλη την απαιτούμενη πληροφορία σχετικά με την ταυτότητα του ατόμου και μηδενική πληροφορία σχετικά με το συναίσθημα. Έτσι, λοιπόν, αν αφαιρέσουμε από ένα AAM για μια εικόνα συναισθήματος το AAM του ίδιου ατόμου αλλά για το ουδέτερο συναίσθημα τότε προκύπτει ένα μοντέλο που περιγράφει τις μεταβολές των εκφράσεων του προσώπου λόγω της εκδήλωσης συναισθήματος.

Αν θεωρήσουμε ότι το σχήμα ενός AAM μπορεί να προκύψει από τον γραμμικό συνδυασμό $\mathbf{s}_p = \mathbf{s}_0 + \sum_{i=1}^n p_i \mathbf{s}_i$, τότε αφαιρώντας το σχήμα του προσώπου υπό ουδέτερο συναίσθημα θα έχουμε:

$$\begin{aligned}
 \mathbf{s}_p - \mathbf{s}_p^{\text{neutral}} &= \mathbf{s}_0 + \sum_{i=1}^n p_i \mathbf{s}_i - \left(\mathbf{s}_0 + \sum_{i=1}^n p_i^{\text{neutral}} \mathbf{s}_i \right) \\
 &= \sum_{i=1}^n (p_i^{\text{neutral}} + p_i^{\text{emotion}}) \mathbf{s}_i - \sum_{i=1}^n p_i^{\text{neutral}} \mathbf{s}_i \\
 &= \sum_{i=1}^n p_i^{\text{emotion}} \mathbf{s}_i
 \end{aligned} \tag{6.31}$$

Επομένως, η αφαίρεση των δύο παραπάνω μοντέλων σχήματος ισοδυναμεί με απλή αφαίρεση των αντίστοιχων παραμέτρων τους, ενώ με ανάλογο τρόπο αποδεικνύεται ότι το ίδιο ισχύει και για το μοντέλο υψής. Έτσι, λοιπόν, πριν προχωρήσουμε στην εκπαίδευση των μοντέλων για κάθε κλάση συναισθήματος αφαιρούμε από τις AAM παραμέτρους κάθε καρέ τις αντίστοιχες παραμέτρους του πρώτου καρέ στο οποίο εκφράζεται το ουδέτερο συναίσθημα. Σε μια πρακτική εφαρμογή, η παραπάνω διαδικασία θα μπορούσε να υλοποιηθεί ζητώντας από τον χρήστη να κοιτάξει στην



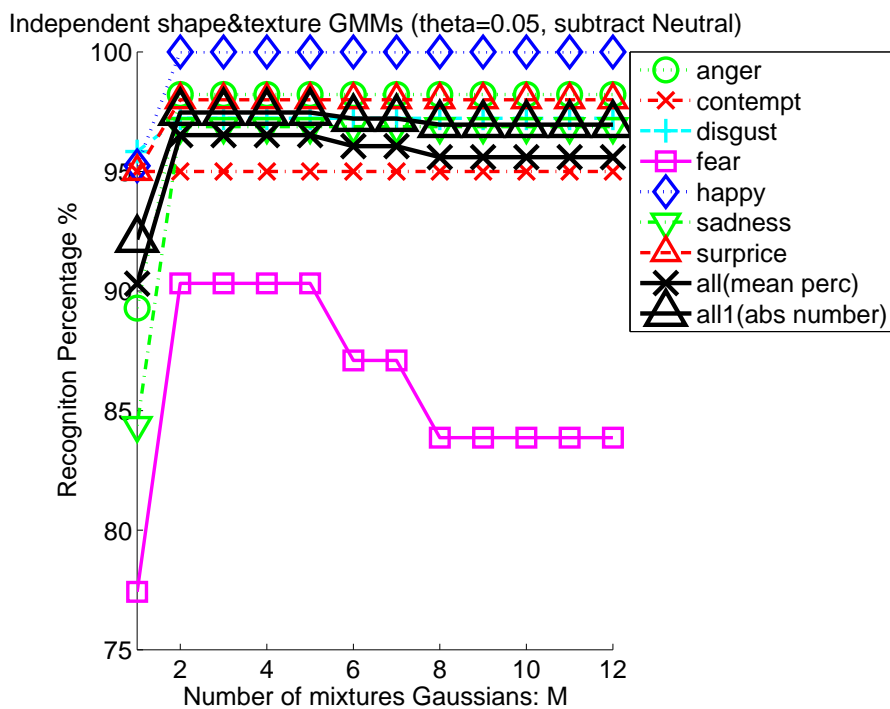
(α') Κοινό GMM για σχήμα και υφή

Σχήμα 6.8: Αποτελέσματα κατηγοριοποίησης συναισθήματος με χρήση GMMs για τη βάση δεδομένων “CK+” του CMU σε σχέση με τον αριθμό των γκαουσιανών αφού πρώτα γίνει η αφαίρεση των παραμέτρων του ουδέτερου συναισθήματος.

κάμερα χωρίς να εκφράσει κάποιο συναίσθημα. Με τον τρόπο αυτό θα γινόταν άμεση προσαρμογή του συστήματος σε οποιονδήποτε νέο χρήστη απαιτώντας μόνο μία εικόνα για τη βαθμονόμηση (one image calibration)

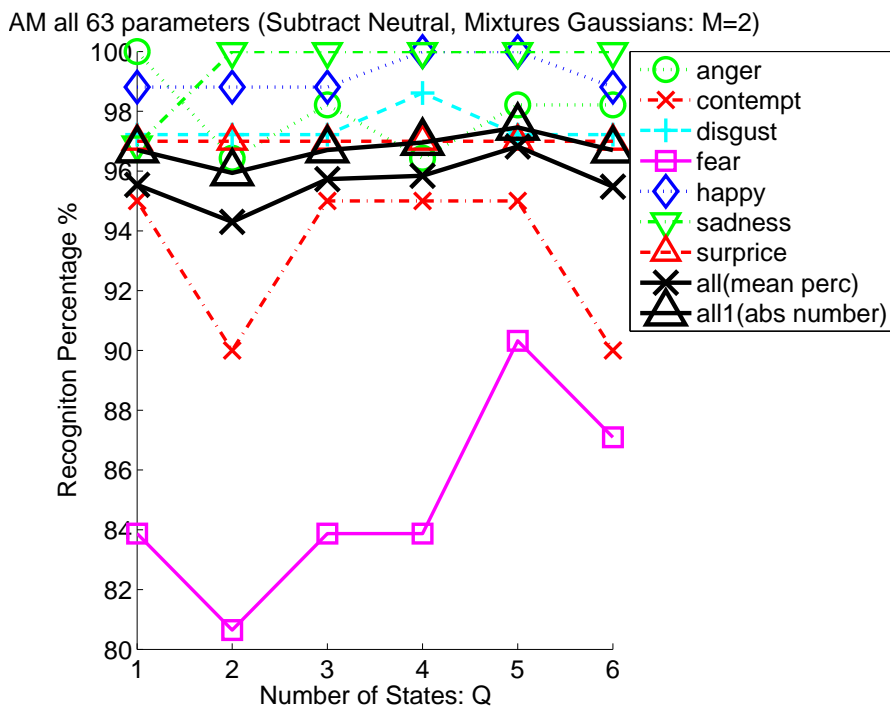
Στο Σχήμα 6.8α' παρουσιάζονται τα ποσοστά αναγνώρισης με χρήση GMMs για κάθε κατηγορία συναισθήματος σε σχέση με τον αριθμό των γκαουσιανών συναρτήσεων αφού πρώτα γίνει αφαίρεση των παραμέτρων του ουδέτερου συναισθήματος. Στο Σχήμα 6.8β' παρουσιάζονται τα ίδια ποσοστά αλλά αυτή τη φορά με χρήση ανεξάρτητων GMMs για το σχήμα και την υφή με $\theta = 0.05$ και $M = 1$. Παρατηρούμε, ότι και στις δύο περιπτώσεις τα ποσοστά αναγνώρισης όλων των κατηγοριών αυξάνονται σε μεγάλο βαθμό δίνοντας συνολικά ένα εξαιρετικό αποτέλεσμα (μέσοι όροι κοντά στο 90%), ακόμα και χωρίς τη χρήση ανεξάρτητων GMMs για το σχήμα και την υφή.

Τέλος, στο Σχήμα 6.9 παρουσιάζονται τα αποτελέσματα που προκύπτουν ακολουθώντας την παραπάνω διαδικασία χρησιμοποιώντας όμως HMM αντί για GMMs και λαμβάνοντας υπόψιν τα τελευταία 6 καρέ. Ο αριθμός των γκαουσιανών



(β') Ανεξάρτητα GMMs για σχήμα και υφή

Σχήμα 6.8: Αποτελέσματα κατηγοριοποίησης συναισθήματος με χρήση GMMs για τη βάση δεδομένων “CK+” του CMU σε σχέση με τον αριθμό των γκαουσιανών αφού πρώτα γίνει η αφαίρεση των παραμέτρων του ουδέτερου συναισθήματος (Συνέχεια).



Σχήμα 6.9: Αποτελέσματα κατηγοριοποίησης συναισθήματος με χρήση HMMs για τη βάση δεδομένων “CK+” του CMU, σε σχέση με τον αριθμό των καταστάσεων, κρατώντας τα τελευταία 6 καρέ GMMs, αφού πρώτα γίνει η αφαίρεση των παραμέτρων του ουδέτερου συναισθήματος και για $M = 2$.

συναρτήσεων για κάθε κλάση επιλέγεται $M = 2$, μιας και για την τιμή αυτή η GMM προσέγγιση δίνει το καλύτερο αποτέλεσμα. Παρατηρούμε πως τα ποσοστά που προκύπτουν αγγίζουν το τέλειο για όλες τις κατηγορίες, με τον μέσο όρο επιτυχούς αναγνώρισης να βρίσκεται κοντά στο 96%. Επιπλέον, το αποτέλεσμα που παίρνουμε με χρήση HMMs προκύπτει για πρώτη φορά καλύτερο από εκείνο που μας δίνουν τα GMMs.

6.5 Συμπεράσματα

Συνοψίζοντας, μπορούμε να πούμε ότι σε πρώτη φάση και σε δεδομένα που εκτός από την πληροφορία σχετικά με το συναίσθημα εμπεριέχεται και πληροφορία σχετικά με την ταυτότητα του προσώπου η χρήση των GMMs, και ιδιαίτερα των ανεξάρτητων GMMs για το σχήμα και την υφή, υπερτερούν έναντι των αντίστοιχων μεθόδων που χρησιμοποιούν HMMs. Το αποτέλεσμα που προκύπτει κρίνεται αρκετά ικανοποιητικό και μπορεί να χρησιμοποιηθεί σε εφαρμογές που απαιτούν ταχύτητα απόκρισης μια και η πολυπλοκότητα των GMMs είναι σημαντικά μικρότερη από εκείνη των HMMs.

Στην περίπτωση, όμως, που είναι εφικτό με κάποιο τρόπο (π.χ. με την απλή μεθοδολογία της ενότητας ;;) να γίνει διαχωρισμός της πληροφορίας που αφορά το συναίσθημα από αυτή που είναι συνολικά διαθέσιμη τότε τα HMMs “κερδίζουν” έναντι των GMMs δίνοντας ένα άψογο ως προς την ακρίβεια αποτέλεσμα. Αυτό είναι αναμενόμενο, μιας και η εκδήλωση ενός συναισθήματος μέσω εκφράσεων του προσώπου είναι ένα χρονικά εξελισσόμενο φαινόμενο και η μοντελοποίηση μέσω καταστάσεων προσεγγίζει περισσότερο την πραγματικότητα. Ωστόσο, αυτό δεν μπορεί να φανεί στις περιπτώσεις που η πληροφορία σχετικά με το συναίσθημα είναι αναμιγμένη με άλλων ειδών πληροφορία. Μπορεί να γίνει όμως εξαιρετικά εμφανές όταν έχουμε στη διάθεση μας χαρακτηριστικά που εκφράζουν αποκλειστικά την εκδήλωση ενός συναισθήματος.

Κεφάλαιο 7

Συμπεράσματα

7.1 Συμβολή της διπλωματικής εργασίας

Οι κύριοι άξονες στους οποίους συνεισφέρει η συγκεκριμένη διπλωματική εργασία συνοψίζονται στα εξής σημεία:

- Στην οπτική μοντελοποίηση του ανθρώπινου προσώπου και συγκεκριμένα της περιοχής των ματιών με χρήση των AMMs. Αρχικά, αφού έγινε περιγραφή των μεθόδων υλοποίησης και προσαρμογής τους σε νέες εικόνες, πραγματοποιήθηκε εκπαίδευση των AAMs με εικόνες της νοηματικής γλώσσας καθώς και με εικόνες που περιέχουν εκδήλωση συναισθημάτων. Στη συνέχεια, τα AMMs που προέκυψαν προσαρμόστηκαν σε νέες εικόνες, παρέχοντας έτσι μια συμπαγή περιγραφή τους με ένα σχετικά μικρό σύνολο παραμέτρων.
- Στην ανάλυση και υλοποίηση ενός συστήματος αυτόματου υπολογισμού του eye gaze που στηρίζεται κατά κύριο λόγο στην πληροφορία που παίρνουμε από την προσαρμογή του AMM στην περιοχή του προσώπου. Για το σκοπό αυτό ακολουθήθηκαν δύο διαφορετικές προσεγγίσεις. Για την πρώτη από αυτές έγινε αναλυτική περιγραφή και υλοποίηση ενός γεωμετρικού μοντέλου, στηριζόμενο σε σημεία της μάσκας σχήματος του AAM, το οποίο παρείχε εκτίμηση για το eye gaze υπό μορφή γωνιών. Αντίθετα, σύμφωνα με τη δεύτερη προσέγγιση έγινε κατηγοριοποίηση του gaze σε διακριτές κλάσεις σύμφωνα με τις παραμέτρους του AAM και προτάθηκε τρόπος για την αντιστοίχιση των κλάσεων με γωνίες. Και οι δύο μέθοδοι δίνουν αρκετά καλά αποτελέσματα που συμφωνούν και με την ανθρώπινη διαίσθηση.
- Στην ανάπτυξη ενός συστήματος κατηγοριοποίησης του συναισθήματος σε διακριτές κλάσεις. Η αναγνώριση βασίστηκε στις παραμέτρους AAMs, ενώ σαν ταξινομητές χρησιμοποιήθηκαν GMMs και HMMs. Όπως είδαμε, τα ποσοστά

επιτυχούς αναγνώρισης που προέκυψαν ήταν αρκετά υψηλά και σε ορισμένες περιπτώσεις ξεπερνούσαν το 90%.

Επιπρόσθετα, στα πλαίσια των παραπάνω βασικών καταυθύνσεων αναπτύχθηκαν και τα παρακάτω θέματα:

- Συστηματική και συνοπτική επισκόπηση των μεθόδων της βιβλιογραφίας σχετικά με τα προβλήματα της εκτίμησης του eye gaze και της κατηγοριοποίησης συναισθήματος.
- Στη περιγραφή και υλοποίηση μεθόδου υπολογισμού της πόζας κεφαλιού σαν ένα ενδιάμεσο στάδιο για την εκτίμηση του gaze. Η τεχνική που αναπτύχθηκε βασίστηκε στα σημεία της μάσκας σχήματος του AAM προκειμένου να υπολογίσει αναλυτικά τις γωνίες στρέψης του κεφαλιού. Τα αποτελέσματα που παρήγαγε η μέθοδος κρίθηκαν αρκετά ικανοποιητικά και χρησιμοποιήθηκαν στη συνέχεια στην εκτίμηση του eye gaze.
- Τέλος, έγινε μια περιγραφή και υλοποίηση των ιστογραμμάτων προσανατολισμένης κλίσης (HOGs), τα οποία εφαρμόστηκαν σε εικόνες της περιοχής των ματιών έτσι ώστε να προκύψει μια εναλλακτική περιγραφή τους πέρα από εκείνη των AAMs. Τα HOG χαρακτηριστικά χρησιμοποιήθηκαν για την κατηγοριοποίηση του gaze, με τα αποτελέσματα που προέκυψαν να υπολείπονται λίγο σε σχέση με εκείνα των AAMs.

7.2 Μελλοντικές ερευνητικές κατευθύνσεις

Παρόλο που στην παρούσα διπλωματική εργασία αναπτύχθηκαν αρκετές τεχνικές με αρκετά ικανοποιητικά αποτελέσματα, υπάρχουν ακόμα πολλά περιθώρια βελτίωσης και επέκτασης των μεθόδων εκτίμησης του gaze και της κατηγοριοποίησης συναισθήματος. Ορισμένες κατευθύνσεις στις οποίες θα μπορούσε να επικεντρωθεί μια μελλοντική έρευνα προτείνονται παρακάτω:

- Όσον αφορά την εκτίμηση του gaze θα μπορούσε να χρησιμοποιηθεί ένα σύνολο δεδομένων με επισημειωμένη την κατεύθυνση του gaze, έτσι ώστε να μπορεί να γίνει σύγκριση των αποτελεσμάτων που προέκυψαν από τις προηγούμενες μεθόδους όχι μόνο διαισθητικά αλλά και ποσοτικά. Επιπλέον, θα μπορούσε να γίνει προσαρμογή και επέκταση των μεθόδων, έτσι ώστε να δίνουν αποδεκτά αποτελέσματα ακόμα και με μεγάλες επικαλύψεις της περιοχής του προσώπου. Στο σημείο, αυτό θα μπορούσε να συμβάλει η λήψη εικόνας από δύο κάμερες καθώς και η δημιουργία ενός πληρέστερου γεωμετρικού μοντέλου από αυτό που χρησιμοποιήθηκε στην παρούσα εργασία.

- Έκφραση του eye gaze σαν συγκεκριμένα σημεία εστίασης της οπτικής προσοχής στο χώρο πέρα από τις γωνίες κατεύθυνσης καθώς και αντιστοίχιση της μεταβολής της κατεύθυνσης του gaze με το γλωσσικό περιεχόμενο ενός νοηματικού λόγου, σε συνδυασμό και με άλλα χαρακτηριστικά όπως οι κινήσεις των χεριών και του κεφαλιού.
- Σχετικά τώρα με το πρόβλημα της αναγνώρισης συναισθήματος, θα μπορούσε να γίνει εφαρμογή των μεθόδων που παρουσιάστηκαν σε εικόνες με πιο φυσικές εκφράσεις συναισθημάτων, όπως σε βίντεο της νοηματικής γλώσσας αλλά και στον προφορικό λόγο με ταυτόχρονη αξιοποίηση της οπτικής και ακουστικής πληροφορίας.

Βιβλιογραφία

- [1] Simon Baker and Iain Matthews. Lucas-Kanade 20 Years On: A Unifying Framework. *International Journal of Computer Vision*, Vol. 56, No. 3, pp. 221-255, March, 2004.
- [2] George Bebis, Kikuo Fujimura. An Eigenspace Approach to Eye-Gaze Estimation. *ISCA 13th International Conference on Parallel and Distributed Computing Systems*, 2000.
- [3] C.M. Bishop. Pattern Recognition and Machine Learning. *Springer*, 2006.
- [4] David B.L. Bong and Kok Houi Lim. Application of Fixed-Radius Hough Transform In Eye Detection. *International Journal of Intelligent Information Technology Application*, 2009.
- [5] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. *5th European Conference on Computer Vision 1998, volume 2, pages 484-498, 1998.*
- [6] T.F. Cootes and C.J. Taylor. Statistical models of appearance for computer vision. *Technical report, University of Manchester, Wolfson Image Analysis Unit, Imaging Science and Biomedical Engineering, United Kingdom, 1999.*
- [7] Navneet Dalal. Finding People in Images and Videos, PhD Thesis. *Institut National Polytechnique de Grenoble/INRIA Grenoble, Grenoble, July 2006.*
- [8] Navneet Dalal and Bill Triggs. Histograms of Oriented Gradients for Human Detection. *In Proceedings of IEEE Conference Computer Vision and Pattern Recognition , San Diego, USA, pages 886 - 893, June 2005.*
- [9] A.H. Gee and R. Cipolla. Determining the gaze of faces in images. *Technical Report CUED/F-INFENG/TR 174, Trumpington Street, Cambridge CB2 1PZ, England, 1994.*

- [10] Ralph Gross, Iain Matthews, and Simon Baker. Constructing and Fitting Active Appearance Models With Occlusion. *Proceedings of the IEEE Workshop on Face Processing in Video, June, 2004*.
- [11] Ralph Gross, Iain Matthews, and Simon Baker. Generic vs. person specific active appearance models. *Image and Vision Computing, Vol. 23, No. 11, pp. 1080-1093, November, 2005*.
- [12] Guestrin E.D., Eizenman M. General theory of remote gaze estimation using the pupil center and corneal reflections. *IEEE Transactions on Biomedical Engineering, 2006*.
- [13] Dan Witzner Hansen, Qiang Ji. In the eye of the beholder: a survey of models for eyes and gaze. *IEEE Transactions on Pattern Analysis and Machine Intelligence (2010)*.
- [14] Hansen D.W., Hansen J.P., Nielsen M., Johansen A.S., Stegmann M.B. Eye typing using Markov and active appearance models. *Proceedings of Sixth IEEE Workshop on Applications of Computer Vision, 2002. (WACV 2002)*.
- [15] Takahiro Ishikawa, Simon Baker, Iain Matthews, and Takeo Kanade. Passive Driver Gaze Tracking with Active Appearance Models. *Proceedings of the 11th World Congress on Intelligent Transportation Systems, October, 2004*.
- [16] Michael J. Jones, Tomaso Poggio. Multidimensional Morphable Models: A Framework for Representing and Matching Object Classes. *International Journal of Computer Vision 29(2): 107-131, 1998*.
- [17] Jungtae Kim, Daijin Kim. Gaze Tracking with Active Appearance Models. *In Proceeding of The 7th POSTECH-KYUTECH Joint Workshop On Neuroinformatics, Pohang, Korea, pp. 90-92, Aug. 2007*.
- [18] Patrick Lucey, Jeffrey F. Cohn, Takeo Kanade, Jason Saragih, Zara Ambadar, and Iain Matthews. The extended Cohn-Kanade (CK+): A complete dataset for action unit and emotion-specified expression. *In IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Workshops), pages 94-101, June 2010*.
- [19] I. Matthews, S. Baker. Active Appearance Models revisited. *International Journal of Computer Vision, 60(2):135-164, 2004*.
- [20] Π. Μαραγκός. *Ανάλυση Εικόνων και Όραση Υπολογιστών*. Ε.Μ.Π., Αθήνα, 2005.

-
- [21] Matsumoto Y., Zelinsky A. An algorithm for real-time stereo vision implementation of head pose and gaze direction measurement. *Proceedings of Fourth IEEE International Conference on Automatic Face and Gesture Recognition, 2000.*
- [22] Kevin Murphy. Hidden Markov Model (HMM) Toolbox for Matlab. *MIT 1998.*
- [23] Murphy-Chutorian E., Trivedi M.M. Head Pose Estimation in Computer Vision: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009.*
- [24] D.P. Mukherjee and A. Zisserman and J.M. Brady. Shape from Symmetry-Detecting and Exploiting Symmetry in Affine Images. *Philosophical Transactions of the Royal Society of London, volume 351, pages 77-106, 1995.*
- [25] T. Ohno, N. Mukawa, and A. Yoshikawa. Freegaze: a gaze tracking system for everyday gaze interaction. *ETRA '02 Proceedings of the 2002 symposium on Eye tracking research and applications, pages 125-132, 2002.*
- [26] G. Papandreou and P. Maragos. Adaptive and Constrained Algorithms for Inverse Compositional Active Appearance Model Fitting. *Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition (CVPR-08), Anchorage, AK, June 2008.*
- [27] Maja Pantic, J.M. Rothkrantz. Automatic Analysis of Facial Expressions: The State of the Art. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000*
- [28] J. Princen, J. Illingworth and J. Kittler. A formal definition of the Hough transform: Properties and relationships. *Journal of Mathematical Imaging and Vision, Volume 1, Number 2, 153-168, 1992.*
- [29] L. Rabiner. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proceedings of the IEEE, 77(2):257-286, 1989.*
- [30] L. Rabiner and B.H. Juang. Fundamental of Speech Recognition. *Prentice Hall, Englewood Cliffs, NJ, 1993.*
- [31] T.D. Rikert, M.J. Jones. Gaze Estimation Using Morphable Models. *Third IEEE International Conference on Automatic Face and Gesture Recognition, 1998.*
- [32] Yunus Saatci, Christopher Town. Cascaded Classification of Gender and Facial Expression using Active Appearance Models. *Seventh IEEE*

International Conference on Automatic Face and Gesture Recognition (FG'06), 2006

- [33] Manabu Takatani, Yasuo Arika, Tetsuya Takiguchi. Gaze Estimation Using Regression Analysis and AAMs Parameters Selected Based on Information Criterion. *International Workshop on Gaze Sensing and Interactions in conjunction with ACCV2010, pp. 1-10, 2010-11.*
- [34] J. Xiao, S. Baker, I. Matthews, and T. Kanade. Real-time combined 2D+3D active appearance models. *In IEEE Conference on Computer Vision and Pattern Recognition, 2004.*
- [35] Ulrich von Agris, Jorg Zieren, Ulrich Canzler, Britta Bauer, Karl-Friedrich Kraiss. Recent developments in visual sign language recognition. *Universal Access in the Information Society, pp. 323-362, 2008.*
- [36] T. Vetter, M. Jones, and T. Poggio. A Bootstrapping Algorithm for Learning Linear Models of Object Classes. *CBCL Paper 143/ AI Memo 1600, Massachusetts Institute of Technology, Cambridge, MA, February 1997.*
- [37] Wang J.G., Sung E., Ronda Venkateswarlu. Eye gaze estimation from a single image of one eye. *Proceedings of Ninth IEEE International Conference on Computer Vision, 2003.*
- [38] Jie Zhu, Jie Yang. Subpixel Eye Gaze Tracking. *Proceedings of Fifth IEEE International Conference on Automatic Face and Gesture Recognition, 2002*

