



Εθνικό Μετσόβιο Πολυτεχνείο

**ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ**

**ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ
ΕΡΓΑΣΤΗΡΙΟ ΕΙΚΟΝΩΝ, ΒΙΝΤΕΟ ΚΑΙ ΠΟΛΥΜΕΣΙΚΩΝ ΣΥΣΤΗΜΑΤΩΝ**

**Αναγνώριση και Μοντελοποίηση της Προσοχής και
του Ενδιαφέροντος του Χρήστη σε Περιβάλλοντα μη
Παρεμβατικής Αλληλεπίδρασης**

Διδακτορική Διατριβή

του

ΣΤΥΛΙΑΝΟΥ Π. ΑΣΤΕΡΙΑΔΗ

Διπλωματούχου Ηλεκτρολόγου Μηχανικού &
Μηχανικού Υπολογιστών Α.Π.Θ. (2004)

Αθήνα, Μάρτιος 2011



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ & ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ
ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ
ΕΡΓΑΣΤΗΡΙΟ ΕΙΚΟΝΩΝ, ΒΙΝΤΕΟ
ΚΑΙ ΠΟΛΥΜΕΣΙΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

Αναγνώριση και Μοντελοποίηση της Προσοχής και του Ενδιαφέροντος του Χρήστη σε Περιβάλλοντα μη Παρεμβατικής Αλληλεπίδρασης

Διδακτορική Διατριβή

του

ΣΤΥΛΙΑΝΟΥ Π. ΑΣΤΕΡΙΑΔΗ

Διπλωματούχου Ηλεκτρολόγου Μηχανικού &
Μηχανικού Υπολογιστών Α.Π.Θ. (2004)

Συμβουλευτική Επιτροπή: Στέφανος Κόλλιας
Ανδρέας-Γεώργιος Σταφυλοπάτης
Κώστας Καρπούζης

Εγκρίθηκε από την επταμελή εξεταστική επιτροπή την 3η Μαρτίου 2011.

...
Σ. Κόλλιας
Καθηγητής Ε.Μ.Π.

...
Α.-Γ. Σταφυλοπάτης
Καθηγητής Ε.Μ.Π.

...
Π. Μαραγκός
Καθηγητής Ε.Μ.Π.

...
Γ. Ματστρος
Επ. Καθηγητής Ε.Μ.Π.

...
Κ. Καρπούζης
Ερευνητής Β,
Ε.Π.Ι.Σ.Ε.Υ., Ε.Μ.Π

...
Γ. Στάμου
Λέκτορας Ε.Μ.Π

...
Α. Ντελόπουλος
Επ. Καθηγητής Α.Π.Θ.

Αθήνα, Μάρτιος 2011

...

ΑΣΤΕΡΙΑΔΗΣ Π. ΣΤΥΛΙΑΝΟΣ

Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Α.Π.Θ.

© 2011 - 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.

Περιεχόμενα

Περιεχόμενα	i
Κατάλογος Σχημάτων	v
Κατάλογος Πινάκων	xi
Πρόλογος	xvii
Περίληψη	xvii
Abstract	xvii
Κατάλογος Απόδοσης Ακρωνυμίων	xix
Κατάλογος Απόδοσης Όρων	xxi
1 Εισαγωγή	3
1.1 Ερευνητικό πλαίσιο και συνεισφορά της διατριβής	4
1.2 Δομή της διατριβής	6
2 Εντοπισμός χαρακτηριστικών προσώπου με Διανυσματικά Πεδία Απόστασης	11
2.1 Ερευνητικό πλαίσιο στην εύρεση χαρακτηριστικών προσώπου	11
2.1.1 Γενικά	11
2.1.2 Χαρακτηριστικές Εργασίες	12
2.2 Εύρεση χαρακτηριστικών προσώπου με χρήση Διανυσματικών Πεδίων Απόστασης	16
2.2.1 Εύρεση περιοχής προσώπου με χαρακτηριστικά Haar και ταίριασμα ελλειψοειδούς	17
2.2.2 Εύρεση περιοχών ματιών	17
2.2.3 Εύρεση περιοχής στόματος	19
2.3 Εύρεση χαρακτηριστικών σημείων με πληροφορία φωτεινότητας και ακμών	20
2.3.1 Εύρεση κέντρων ματιών	21
2.3.2 Εύρεση γωνιών ματιών και βλεφάρων	24
2.3.3 Εύρεση φρυδιών	26
2.3.4 Εύρεση ρωθώνων	26
2.3.5 Εύρεση γωνιών στόματος	27
2.4 Πειραματικά αποτελέσματα	28
2.5 Συμπεράσματα	31

3	Εκτίμηση περιστροφής κεφαλιού με χρήση οπτικής ροής	33
3.1	Ερευνητικό πλαίσιο	33
3.1.1	Ολιστικές μέθοδοι (μέθοδοι εμφάνισης)	35
3.1.2	Τοπικές τεχνικές	37
3.1.3	Ανάκτηση κίνησης προσώπου	39
3.1.4	Παραμετροποιημένα μοντέλα.....	40
3.2	Υπάρχουσες Βάσεις ακολουθιών Περιστροφής Κεφαλιού και Κατευθυντικότητας Ματιών στη Βιβλιογραφία	41
3.2.1	Εκτενής βάση προσώπων	41
3.2.2	XM2VTS βάση προσώπων.....	42
3.2.3	Βάση προσώπων «Yale», έκδοση B	43
3.2.4	Βάση προσώπων Πανεπιστημίου Βοστόνης (BU)	43
3.2.5	CVL βάση προσώπων	43
3.2.6	INDIAN βάση προσώπων.....	45
3.2.7	Βάση Pointing'04	45
3.3	Βάση ακολουθιών Περιστροφής Κεφαλιού και Κατευθυντικότητας Ματιών (HPEG).....	47
3.3.1	Βάσεις περιστροφής κεφαλιού και κατευθυντικότητας βλέμματος	47
3.3.2	Περιγραφή της βάσης	47
3.3.3	Εξαγωγή δεδομένων αναφοράς	49
3.3.4	Πειραματική επαλήθευση χρησιμότητας βάσης	51
3.3.4.1	Εκτίμηση περιστροφής κεφαλιού	51
3.3.4.2	Εκτίμηση κατευθυντικότητας βλέμματος	51
3.4	Παρακολούθηση χαρακτηριστικών	51
3.5	Εκτίμηση περιστροφής κεφαλιού	54
3.6	Πειράματικά αποτελέσματα	57
3.6.1	Πειράματα στην XM2VTS	57
3.6.1.1	Επιτυχία στην αναγνώριση περιστροφής κεφαλιού	59
3.6.2	Πειράματα στην βάση του Πανεπιστημίου της Βοστόνης	61
3.7	Συμπεράσματα	61
4	Σύντηξη τοπικής και ολιστικής πληροφορίας για εκτίμηση περιστροφής κεφαλιού	63
4.1	Ερευνητικό πλαίσιο	64
4.2	Χρήση τοπικής πληροφορίας με Διανυσματικά Πεδία Απόστασης	65
4.2.1	Παρακολούθηση προσώπου με προσαρμοστική κατωφλίωση	67
4.2.2	Παρακολούθηση χαρακτηριστικών προσώπου με χρήση Διανυσματικών Πεδίων Απόστασης	68
4.2.2.1	Παρακολούθηση ματιών	68
4.2.2.2	Παρακολούθηση στόματος	70
4.2.2.3	Βελτιστοποίηση της παρακολούθησης	71
4.2.3	Εκτίμηση οριζόντιας, κατακόρυφης και παράλληλης στο επίπεδο της κάμερας περιστροφής	71
4.2.3.1	Σύγκριση παρακολούθησης χαρακτηριστικών προσώπου με Διανυσματικά Πεδία Απόστασης και Οπτική Ροή	73
4.2.4	Πειραματικά αποτελέσματα	75

4.2.4.1	Αποτελέσματα στη βάση του Πανεπιστημίου της Βοστόνης	75
4.2.4.2	Αποτελέσματα στη βάση HPEG	75
4.2.4.3	Αποτελέσματα με εξαγωγή και χρήση χαρακτηριστικών καρτέ	75
4.3	Χρήση ολιστικής πληροφορίας με Συνελικτικά Νευρωνικά Δίκτυα	77
4.3.1	Προτεινόμενη αρχιτεκτονική.....	79
4.3.2	Διαδικασία εκπαίδευσης	80
4.3.3	Εκτίμηση οριζόντιας και κατακόρυφης περιστροφής κεφαλιού με χρήση Συνελικτικών Νευρωνικών Δικτύων	81
4.4	Σύντηξη ολιστικής με τοπική πληροφορία με χρήση σταθμισμένου μέσου	83
4.4.1	Αποτελέσματα στη βάση του Πανεπιστημίου της Βοστόνης	83
4.5	Σύντηξη ολιστικής με τοπική πληροφορία με χρήση μπεϋζιανών δικτύων	87
4.5.1	Μπεϋζιανά μοντέλα συλλογιστικής από ένα επίπεδο πληροφορίας	87
4.5.2	Μπεϋζιανά μοντέλα συλλογιστικής από περισσότερα επίπεδα πληροφορίας	88
4.5.3	Θόρυβος μέτρησης από τοπική πληροφορία	88
4.5.4	Θόρυβος μέτρησης από ολιστική πληροφορία	89
4.5.5	Παράμετροι και εκπαίδευση δικτύου.....	89
4.5.6	Πειραματικά αποτελέσματα: Υπολογισμός οριζόντιας περιστροφής κεφαλιού	90
4.6	Συμπεράσματα	90
5	Εκτίμηση ενδιαφέροντος σε περιβάλλον επικοινωνίας Ανθρώπου-Υπολογιστή με χρήση οπτικής ροής	93
5.1	Ερευνητικό πλαίσιο	94
5.1.1	Επικοινωνία ανθρώπου με εικονικό πράκτορα-ρομπότ	96
5.1.2	Μοντελοποίηση ανθρώπινης συμπεριφοράς σε συνθήκες οδήγησης	97
5.1.3	Βιβλιογραφική επισκόπηση μεθόδων κατευθυντικότητας βλέμματος ματιών	99
5.2	Προσοχή και ενασχόληση με χρήση κίνησης κεφαλιού	102
5.2.1	Αντιστοίχιση με συντεταγμένες οθόνης	103
5.2.2	Κατευθυντικότητα και επίπεδα προσήλωσης	103
5.2.3	Αντικείμενα Εικονικής Προσήλωσης	104
5.2.4	Επίπεδα Ενδιαφέροντος	105
5.2.5	Πειραματικά αποτελέσματα	106
5.2.5.1	Πείραμα 1	106
5.2.5.2	Πείραμα 2	110
5.2.6	Πορίσματα	112
5.3	Από το βλέμμα στην αναγνώριση κατάστασης χρήστη.....	113
5.3.1	Περιγραφή της εφαρμογής	115
5.3.2	Καταγραφή και επισημείωση της πλατφόρμας δοκιμών	118
5.3.3	Υπολογισμός χαρακτηριστικών ανάλυσης κίνησης κεφαλιού και προσώπου, για εκτίμηση επιπέδων προσοχής και νευρικότητας	119
5.3.3.1	Εκτίμηση βλέμματος ματιών	120
5.3.3.2	Υπολογισμός περισσότερων χαρακτηριστικών	120

5.3.4	Στατιστικός υπολογισμός καταλληλότερων χαρακτηριστικών ...	121
5.3.5	Πειραματικά αποτελέσματα	121
5.3.5.1	Μοντέλο Takagi-Sugeno-Kang	122
5.3.5.2	Ασαφές σύστημα συμπερασμού για εκτίμηση επιπέδων προσήλωσης	123
5.3.5.3	Ασαφές σύστημα συμπερασμού για εκτίμηση επιπέδων νευρικότητας	128
5.4	Συμπεράσματα	131
6	Χρήση γνώσης προσοχής χρήστη για εύρεση περιστροφής κεφαλιού. Επέκταση με χρήση κυλινδρικών μοντέλων για ανάκτηση βλέμματος ματιών	133
6.1	Ερευνητικό πλαίσιο	133
6.2	Χρήση γνώσης προσοχής χρήστη για εύρεση περιστροφής κεφαλιού ...	134
6.2.1	Υπολογισμός επιπέδων προσοχής χρήστη	135
6.2.2	Επισημείωση επιπέδων προσοχής	135
6.2.3	Εκπαίδευση νευρο-ασαφούς συστήματος για εκτίμηση επιπέδων προσοχής	135
6.2.4	Χρήση νευρο-ασαφούς συστήματος για δυναμική σύντηξη ολιστικής και τοπικής πληροφορίας	137
6.2.5	Πειραματικά αποτελέσματα: Υπολογισμός οριζόντιας περιστροφής κεφαλιού	137
6.3	Κατάσταση χρήστη με χρήση τοπικής και ολιστικής πληροφορίας για εκτίμηση περιστροφής κεφαλιού, και κυλινδρικών μοντέλων για υπολογισμό κατευθυντικότητας βλέμματος	138
6.3.1	Εύρεση κατευθυντικότητας ματιών με χρήση κυλινδρικών μοντέλων	138
6.3.2	Επισημείωση επιπέδων προσοχής της Βάσης του Πανεπιστημίου της Βοστώνης, με χρήση κατευθυντικότητας βλέμματος ματιών	141
6.3.3	Στατιστική σημαντικότητα κατευθυντικότητας ματιών και περιστροφής κεφαλιού	141
6.3.4	Πειραματικά αποτελέσματα	142
6.4	Συμπεράσματα	144
7	Συμπεράσματα και μελλοντικές επεκτάσεις	145
7.1	Συμπεράσματα	145
7.1.1	Αναγνώριση χαρακτηριστικών	145
7.1.2	Μοντελοποίηση επιπέδων προσήλωσης χρήστη	147
7.2	Μελλοντικές επεκτάσεις	148
	Βιβλιογραφία	151
	Κατάλογος δημοσιεύσεων συγγραφέα	165
	A Παράμετροι Ασαφών Συστημάτων	167

Κατάλογος Σχημάτων

2.1	Μέθοδος ανίχνευσης προσώπου: Από ολόκληρη την εικόνα, βρίσκεται η έξοδος της μεθόδου Viola-Jones, και, στην εκτιμώμενη περιοχή, ο Πίνακας Αποστάσεων περιοχής προσώπου στον οποίο εκτιμάται το αποτέλεσμα ταυρίσματος έλλειψης.....	17
2.2	Παράδειγμα εξαγωγής διανυσματικού πεδίου για περιοχή ματιού. (α) εικονίδιο ματιού, (β) χάρτης αποστάσεων, (γ) κάθετες συντεταγμένες διανυσματικού πεδίου, (δ) οριζόντιες συντεταγμένες διανυσματικού πεδίου	18
2.3	(α) Μέσες κάθετες συντεταγμένες (β) Μέσες οριζόντιες συντεταγμένες	18
2.4	(α) Περιοχή προσώπου (β) Χάρτης ακμών Canny (γ) οριζόντιες συντεταγμένες του διανυσματικού πεδίου (δ) κάθετες συντεταγμένες του διανυσματικού πεδίου	19
2.5	(α) περιοχή στόματος (β) κάθετες συντεταγμένες του διανυσματικού πεδίου (γ) οριζόντιες συντεταγμένες του διανυσματικού πεδίου.....	21
2.6	Μέσο διανυσματικό πεδίο (α) μέσες κάθετες συντεταγμένες περιοχών στόματος (β) μέσες οριζόντιες συντεταγμένες περιοχών στόματος.....	21
2.7	Διαδικασία αφαίρεσης αντανακλάσεων: (α) περιοχή ματιού, (β) δυαδική περιοχή ματιού, (γ) δυαδική περιοχή ματιού με αφαιρεμένες τις περιοχές αντανακλάσεων, (δ) ασπρόμαυρη περιοχή ματιού με αφαιρεμένες τις περιοχές αντανακλάσεων.....	22
2.8	Παραδείγματα περιοχών ματιών και περιοχές όπου γίνεται η αναζήτηση του κέντρου του ματιού	23
2.9	Παραδείγματα περιοχών αναζήτησης κέντρου ματιού και οι εικόνες των παραγώγων κατά μήκος γραμμών και στηλών.	23
2.10	(α) Αρχική εκτίμηση του κέντρου του ματιού (β),(γ) περιοχές που χρησιμοποιούνται για βελτίωση του αποτελέσματος	24
2.11	(α) Εκτίμηση θέσης κέντρου ματιού μετά από την πρώτη βελτίωση, (β) τελική θέση κέντρου ματιού	24
2.12	Τιμές Παραγώγων Συναρτήσεων Προβολής για υποπεριοχές στις οποίες ανιχνεύονται τα κατακόρυφα όρια (άκρες) του ματιού	25
2.13	Διαδικασία βελτιστοποίησης εύρεσης γωνίας ματιού με υπόθεση συγγραμικότητας γωνιών ματιών	26
2.14	Περιοχές ανίχνευσης χαρακτηριστικών σημείων στα φρύδια και στη μύτη.....	27
2.15	Περιοχές στόματος και οι αντίστοιχες τιμές του καναλιού hue	28
2.16	(α) Περιοχή στόματος, (β) Δυαδική εικόνα καναλιού hue με μικρές ενιαίες περιοχές, (γ) Αντίστοιχη δυαδική εικόνα χωρίς τις μικρές περιοχές (δ) περιοχή στόματος με γωνίες σημειωμένες	28

2.17	Άκρες και κέντρα περιοχών ματιών	29
2.18	Ακρίβεια εύρεσης για 6 χαρακτηριστικά σημεία στην XM2VTS για διάφορα κατώφλια T επιτυχίας	30
2.19	Ακρίβεια εύρεσης για σημείο ανάμεσα στους ρώθωνες στην XM2VTS για διάφορα κατώφλια T επιτυχίας	31
3.1	Οι τρεις βαθμοί ελευθερίας στην εκτίμηση περιστροφής κεφαλιού: οριζόντια περιστροφή (<i>yaw</i>), κατακόρυφη (<i>pitch</i>), και περιστροφή στο επίπεδο της κάμερας (<i>roll</i>)	34
3.2	Παραδείγματα από την Διευρυμένη Βάση Προσώπων.....	42
3.3	Παραδείγματα από την XM2VTS	43
3.4	Παραδείγματα από τη βάση Yale	44
3.5	Παραδείγματα εικόνων από τη βάση του Πανεπιστημίου της Βοστώνης .	44
3.6	Παράδειγμα από τη βάση CVL	45
3.7	Παράδειγμα από τη βάση προσώπων INDIAN.....	46
3.8	Παράδειγμα από τη βάση προσώπων INDIAN.....	46
3.9	Παραδείγματα από την πρώτη συνεδρία	48
3.10	Παραδείγματα από τη δεύτερη συνεδρία	48
3.11	(α) Αρχική δυαδική μάσκα, όπως υπολογίστηκε από τις τιμές των καναλιών a^* και b^* ; (β) Δυαδική μάσκα, μετά το λογικό <i>AND</i> μεταξύ της εικόνας (α) και του δυαδικού χάρτη φωτεινότητας L , (γ) τελική δυαδική μάσκα των φωτοδιδώδων μετά από χρήση της γεωμετρίας τους .	49
3.12	Παράδειγμα ακολουθίας της δεύτερης συνεδρίας, και αποτελέσματα στην εκτίμηση κατευθυντικότητας βλέμματος.....	52
3.13	Παράδειγμα διόρθωσης παρακολούθησης χαρακτηριστικών σε πραγματικό χρόνο. Τα περισσότερα σημεία στο πρόσωπο «χάνονται», όμως, αφού ο χρήστης επιστρέψει σε εμπρόσθια θέση, ανακτώνται με την επαναρχικοποίηση. Η λευκή γραμμή αντιστοιχεί στο διάνυσμα περιστροφής	55
3.14	Παράδειγμα διόρθωσης παρακολούθησης χαρακτηριστικών σε πραγματικό χρόνο. Εδώ αποκρύφτηκε το σημείο που αντιστοιχεί στην εξωτερική γωνία του δεξιού φρυδιού. Η θέση του ανακτάται, όμως, εφόσον υπάρχει επαναρχικοποίηση. Η λευκή γραμμή αντιστοιχεί στο διάνυσμα περιστροφής	56
3.15	(α) Οριζόντια περιστροφή κεφαλιού, (β) Κατακόρυφη περιστροφή κεφαλιού: πραγματικές, εκτιμώμενες τιμές	58
3.16	Παραδείγματα τιμών οριζόντιας και κάθετης συνιστώσας διανύσματος περιστροφής κεφαλιού και χαρακτηριστικά καρέ της ακολουθίας	60
4.1	Σύνοψη της μεθόδου εκτίμησης περιστροφής κεφαλιού με χρήση ΔΠΑ: Αρχικά εντοπίζεται το πρόσωπο και τα χαρακτηριστικά σημεία σε αυτό. Ακολουθώς, παρακολουθείται η περιοχή του προσώπου και, μέσα σε αυτή, τα Διανυσματικά Πεδία Απόστασης χαρακτηριστικών περιοχών. Τα όρια της περιοχής του προσώπου, σε σχέση με τα χαρακτηριστικά σημεία χρησιμοποιούνται για να υπολογιστεί το διάνυσμα περιστροφής του προσώπου. Εάν το μέτρο του διανύσματος είναι μικρότερο από κάποιο μέγεθος, τότε ο αλγόριθμος επανεκκινείται.	66
4.2	Περιοχή που χρησιμοποιείται για την εξαγωγή πληροφορίας χρωματικότητας	67

4.3	Επισκόπηση της διαδικασίας επιλογής του κατωφλίου T , για κατηγοριοποίηση των εικονοστοιχείων, με βάση την εξίσωση 4.1: για τη συγκεκριμένη ακολουθία, αποφασίστηκε κατωφλίωση στο $T = 0.15$, αφού ο συνολικός αριθμός των εικονοστοιχείων, οι τιμές κορεσμού των οποίων, είναι κοντά σε αυτήν της αρχικά επιλεγμένης περιοχής δέρματος, προσεγγίζουν τον πραγματικό αριθμό των εικονοστοιχείων του προσώπου.	69
4.4	(α) τιμές κορεσμού αρχικής εικόνας (β) κατωφλιωμένες τιμές κορεσμού (γ) Υποψήφια να ανήκουν σε περιοχή προσώπου εικονοστοιχεία C_{fp} , όπως εξήχθησαν μετά από μορφολογικό άνοιγμα (δ) Υποψήφια να ανήκουν σε δερματική περιοχή εικονοστοιχεία C_{sp} ; (ε) Υποψήφια να ανήκουν σε περιοχή προσώπου εικονοστοιχεία μετά το λογικό <i>AND</i> μεταξύ C_{fp} και C_{sp} ; (στ) τελική μάσκα προσώπου, μετά τους μορφολογικούς μετασχηματισμούς.	69
4.5	(α) Περιοχή ματιού στο καρέ k , (β) Περιοχή αναζήτησης ματιού στο καρέ $k + 1$; (γ) ΔΠΑ σφάλμα μεταξύ (α) και (β)	70
4.6	(α) Περιοχή εντοπισμού στόματος (β) Κανάλι χρωματικότητας περιοχής εντοπισμού στόματος (γ) Πίνακας οριζόντιων ακμών περιοχής εντοπισμού στόματος (δ) Χρωματικότητα πολλαπλασιασμένη με πίνακα οριζόντιων ακμών	71
4.7	Εκτίμηση Διανύσματος Περιστροφής Κεφαλιού (εξισώσεις 4.7 και 4.8)	73
4.8	Παραδείγματα περιοχών προσώπου με τα αντίστοιχα χαρακτηριστικά σημεία που χρησιμοποιούνται για εκτίμηση περιστροφής κεφαλιού	74
4.9	Παράδειγμα οβελιαίας περιστροφής στη βάση BU.	74
4.10	Εκτίμηση (α) οριζόντιας και (β) κατακόρυφης περιστροφής κεφαλιού με χρήση διαφορετικών κατωφλίων T για το χαρακτηρισμό επιτυχούς υπολογισμού, για Διανυσματικά Πεδία και Οπτική Ροή.	76
4.11	Αρχιτεκτονική Συνελικτικού Νευρωνικού Δικτύου για εκτίμηση περιστροφής κεφαλιού	80
4.13	α) Περιοχή προσώπου με βάση το χρωματικό πρότυπο β) Μέρος περιοχής προσώπου, όπως χρησιμοποιείται από τα ΣΝΔ.....	81
4.14	Περιστροφή προσώπου στο επίπεδο της κάμερας, για εξαγωγή γωνιών γύρω από τον οριζόντιο και κατακόρυφο άξονα	82
4.15	Μέθοδος εκτίμησης περιστροφής προσώπου με χρήση Διανυσματικών Πεδίων Απόστασης και Συνελικτικών Νευρωνικών Δικτύων: Αρχικά εξάγεται η περιοχή του προσώπου και, βάσει δείγματος από το πρόσωπο, παρακολουθείται η περιοχή του προσώπου, η οποία χρησιμοποιείται σαν είσοδος στα Συνελικτικά Δίκτυα. Όταν η περιστροφή είναι σχεδόν μηδενική, εκκινείται εντοπισμός/παρακολούθηση προσώπου με Διανυσματικά Πεδία Απόστασης. Ο αλγόριθμος μπορεί να επανεκκινείται όταν η περιστροφή υπολογίζεται μηδενική. Επιπλέον, μεγάλες διαφορές στα αποτελέσματα ολιστικής από τοπική μέθοδο ομαλοποιούνται επανεκκινώντας το σύστημα σε κατάλληλη χρονική στιγμή.	84
4.16	Εκτιμώμενες γωνίες περιστροφής στη βάση BU και αντίστοιχες τιμές δεδομένων αναφοράς, με χρήση ολιστικής και τοπικής πληροφορίας. ...	86
4.17	Γενική μορφή μπεϋζιανού δικτύου για ένα επίπεδο πληροφορίας	88
4.18	Μπεϋζιανό Δίκτυο για δύο επίπεδα πληροφορίας	88

4.19	Αρχιτεκτονική μπεϋζιανού δικτύου που χρησιμοποιείται για σύντηξη ολιστικής (ΣΝΔ) με τοπική πληροφορία (ΔΠΑ)	89
5.1	Απεικόνιση του σεναρίου εικονικού πωλητή και αντικειμένων προς πώληση και των περιεχομένων σκηνής με συμβολισμό Αντικειμένων Εικονικής Προσήλωσης. Κάθε Αντικείμενο Εικονικής Προσήλωσης κρατάει ιστορικό των χρονικών στιγμών και το βαθμό στον οποίο κάποιος χρήστης ενδιαφέρθηκε για αυτό.	104
5.2	3Δ καμπύλη που απεικονίζει τη μέση κίνηση ανά χρήστη.....	107
5.3	Μέση τιμή οριζόντιας συνιστώσας διανύσματος περιστροφής κεφαλιού .	107
5.4	Αντιστοίχιση διανύσματος περιστροφής κεφαλιού σε συντεταγμένες οθόνης	108
5.5	Απεικόνιση του σεναρίου, όπου ο πράκτορας ακολουθάει διαφορετικά πρότυπα συμπεριφοράς όσον αφορά το βλέμμα του σε σχέση με τα αντικείμενα για τα οποία μιλάει. Στο κέντρο της οθόνης, ο χρήστης έχει τη δυνατότητα να έχει οπτική επαφή με την εικόνα του, καθώς το αντικείμενο προς το οποίο κοιτάζει είναι μαρκαρισμένο με ένα κόκκινο κύκλο. Δεξιά φαίνεται το ιστορικό της κατευθυντικότητας του βλέμματος ενός χρήστη κατά τη διάρκεια ενός σεναρίου.	111
5.6	Θετικές απαντήσεις στα ερωτηματολόγια στην ερώτηση (αριστερά) αν ενδιαφέρεται ο πράκτορας για τον πράκτορα, αν (κέντρο) ο πράκτορας ενδιαφέρεται για τον χρήστη και (δεξιά) αν ο χρήστης ενδιαφέρεται για τον πράκτορα. Οι απαντήσεις αφορούν και στις τέσσερις συνθήκες του πειράματος	112
5.7	Παράδειγμα περίπτωσης συνθήκης 4, στο οποίο φαίνεται ότι ο χρήστης ακολούθησε τη μη-προφορική πληροφορία του πράκτορα για να στρέψει το βλέμμα του. Στη συγκεκριμένη περίπτωση, ο χρήστης στρέφει το βλέμμα του στο παιγνιομόχλιο (κάτω-κέντρο) ενώ ο πράκτορας αναφέρεται στο ποντίκι (κάτω-αριστερά). Αργότερα, ο χρήστης έστρεψε το βλέμμα του στο σωστό αντικείμενο.	114
5.8	Επέκταση του συστήματος εκμάθησης LTSA IEEE	117
5.9	Μέτρο διανύσματος περιστροφής κεφαλιού κατά τη διάρκεια ακολουθίας.	120
5.10	Μέτρο διανύσματος κατευθυντικότητας βλέμματος ματιών κατά τη διάρκεια ακολουθίας.	121
5.11	(α) Χαρακτηριστικά που χρησιμοποιούνται στο σενάριο Προσεκτικός-Απρόσεκτος (β) Χαρακτηριστικά που χρησιμοποιούνται στο σενάριο Νευρικός-Ήρεμος	122
5.12	Ασαφές TSK και αντίστοιχη έξοδος για είσοδο Περιστροφή Κεφαλιού=0.105, Βλέμμα=0.115 και Απόσταση=1.05	124
5.13	Ασαφές TSK και αντίστοιχη έξοδος για είσοδο Περιστροφή Κεφαλιού=0.052, Βλέμμα=0.33 και Απόσταση=1.05.....	125
5.14	Ασαφές TSK και αντίστοιχη έξοδος για είσοδο Περιστροφή Κεφαλιού=0.428, Βλέμμα=0.455 και Απόσταση=0.96	125
5.15	Αντιστοίχιση Περιστροφής Κεφαλιού και Βλέμματος με έξοδο Μη-Προσήλωση	126
5.16	Αντιστοίχιση Περιστροφής Κεφαλιού και Αλλαγές αποστάσεων χρήστη-κάμερας με έξοδο Μη-Προσήλωση.....	127

5.17	Αντιστοίχιση Βλέμματος και Αλλαγές αποστάσεων χρήστη-κάμερας με έξοδο <i>Μη-Προσήλωση</i>	127
5.18	Ασαφές ΤSK και αντίστοιχη έξοδος για είσοδο Περιστροφή Κεφαλιού=0.051,οριζόντια ταχύτητα=0.077 και κάθετη ταχύτητα=0.058	128
5.19	Ασαφές ΤSK και αντίστοιχη έξοδος για είσοδο Περιστροφή Κεφαλιού=0.257, οριζόντια ταχύτητα=0.324 και κάθετη ταχύτητα=0.276	129
5.20	Ασαφές ΤSK και αντίστοιχη έξοδος για είσοδο Περιστροφή Κεφαλιού=0.020,οριζόντια ταχύτητα=0.418 και κάθετη ταχύτητα=0.251	129
5.21	Αντιστοίχιση Περιστροφής Κεφαλιού και κάθετης ταχύτητας κίνησης κεφαλιού με έξοδο <i>Νευρικότητα</i>	130
5.22	Αντιστοίχιση Περιστροφής Κεφαλιού και οριζόντιας ταχύτητας κίνησης κεφαλιού με έξοδο <i>Νευρικότητα</i>	130
6.1	Χαρακτηριστικές εικόνες της βάσης του Πανεπιστημίου της Βοστόνης, στις οποίες οι επισημειωτές θεώρησαν ότι ο χρήστης στην (α) έχει βαθμό προσοχής 7% και στη (β) 10%.	136
6.2	Χαρακτηριστικό στιγμιότυπο διαδικασίας επισημείωσης της βάσης του Πανεπιστημίου Βοστόνης, ως προς τα επίπεδα προσοχής (η επίδραση της κατευθυντικότητας βλέμματος έχει αποκλεισθεί).	136
6.3	Μπεϋζιανό Δίκτυο με χρήση πληροφορίας προσοχής χρήστη	137
6.4	Χαρακτηριστικό στιγμιότυπο διαδικασίας επισημείωσης της βάσης του Πανεπιστημίου Βοστόνης, ως προς τα επίπεδα προσοχής (η επίδραση της κατευθυντικότητας βλέμματος δεν έχει αποκλεισθεί).	139
6.5	Σύνοψη προτεινόμενης αρχιτεκτονικής με χρήση ολιστικής και τοπικής πληροφορίας για εκτίμηση περιστροφής κεφαλιού και κατευθυντικότητας βλέμματος με χρήση κυλινδρικών μοντέλων.	140
6.6	Υπολογισμός διανύσματος κατευθυντικότητας βλέμματος ματιών. Οι θέσεις των ματιών στην περιστραμμένη εικόνα συγκρίνονται με αυτές στη μετωπική θέση, αφού εξαλειφθεί η επίδραση της περιστροφής του κεφαλιού.	141
6.7	Παραδείγματα επισημειωμένων εικόνων, με χρήση, τόσο περιστροφής κεφαλιού, όσο και κατευθυντικότητας ματιών (α,γ), και αποκλειστικά περιστροφής κεφαλιού (β,δ). Τα άτομα στις εικόνες (α) και (γ) χαρακτηρίστηκαν με μέσο βαθμό επιπέδων προσοχής 53% και 60%, αντίστοιχα, ενώ οι αντίστοιχες φιλτραρισμένες εκδόσεις (β) και (δ) με 7% και 10%.	142
6.8	Έξοδος συστήματος συμπερασμού, εξαρτώμενη από τις τιμές της οριζόντιας περιστροφής και της κατευθυντικότητας βλέμματος.	143

Κατάλογος Πινάκων

2.1	Ακρίβεια της μεθόδου και χρόνοι εντοπισμού χαρακτηριστικών ανά καρτέ	30
2.2	Ακρίβεια εντοπισμού χαρακτηριστικών σημείων στα μάτια, για κριτήρια επιτυχίας $T=0.1$ και $T=0.15$ (οι εικόνες είναι σε διάσταση 360×288)	31
3.1	Βάσεις εικόνων/βίντεο που χρησιμοποιούνται στη βιβλιογραφία για εκτίμηση περιστροφής κεφαλιού	51
3.2	Ακρίβεια-Ανάκτηση- f -measure για τις τέσσερις ακραίες καταστάσεις περιστροφής κεφαλιού	60
3.3	Χρόνοι παρακολούθησης χαρακτηριστικών ανά καρτέ	61
3.4	Μέσο Απόλυτο Σφάλμα στην εκτίμηση οριζόντιας (yaw), κατακόρυφης ($pitch$) και γύρω από τον κάθετο στο επίπεδο της κάμερας περιστροφή ($roll$) στην βάση του Πανεπιστημίου της Βοστώνης, για μεμονωμένους χρήστες	61
4.1	Εύρεση περιστροφής κεφαλιού με χρήση παρακολούθησης χαρακτηριστικών με Οπτική Ροή και Διανυσματικά Πεδία Απόστασης	75
4.2	Σύγκριση αποτελεσμάτων χρήσης ΔΠΑ στην BU	75
4.3	Αποτελέσματα χρήσης ΔΠΑ στην HPEG	75
4.4	αποτελέσματα εκτίμησης οριζόντιας και κατακόρυφης περιστροφής του κεφαλιού στη βάση HPEG με χρήση παρακολούθησης με ΔΠΑ και χαρακτηριστικών καρτέ	77
4.5	Πίνακας συνδεσμολογίας επιπέδου $S2$ (γραμμές) με στρώμα συνέλιξης $C3$ (στήλες). Κάθε X αντιστοιχεί σε υπάρχουσα σύνδεση μεταξύ εξόδου του $S2$ και εισόδου του $C3$	79
4.6	Μέσα Τετραγωνικά Σφάλματα στη BU για Διανυσματικά Πεδία Απόστασης, Συνελικτικά Νευρωνικά Δίκτυα, σύντηξη των δύο, και άλλες μεθόδους της βιβλιογραφίας	85
4.7	Μέσα απόλυτα Σφάλματα στη BU για Διανυσματικά Πεδία Απόστασης, Συνελικτικά Νευρωνικά Δίκτυα, σύντηξη των δύο, και άλλες μεθόδους της βιβλιογραφίας	85
4.8	Σύγκριση διαφόρων τεχνικών σύντηξης ολιστικής και τοπικής πληροφορίας για υπολογισμό οριζόντιας περιστροφής κεφαλιού	90
5.1	Πίνακας σύγκρισης σφαλμάτων για όλους τους χρήστες	109
5.2	Πίνακας Σύγκρισης για κάθε χρήστη ξεχωριστά	110
5.3	Κατηγοριοποίηση Πιθανών Καταστάσεων Χρήστη	116

5.4	Ακρίβεια νευρο-ασαφούς στην εκτίμηση επιπέδων προσήλωσης στη βάση του Agent-Dysl με χρήση δύο διαφορετικών συνόλων χαρακτηριστικών	127
5.5	Ακρίβεια νευρο-ασαφούς στην εκτίμηση επιπέδων νευρικότητας στη βάση του Agent-Dysl με χρήση δύο διαφορετικών συνόλων χαρακτηριστικών	131
A.1	Ασαφές σύστημα ενότητας 5.3.5.2: Είσοδοι, Κανόνας 1	167
A.2	Ασαφές σύστημα ενότητας 5.3.5.2: Έξοδος, Κανόνας 1	167
A.3	Ασαφές σύστημα ενότητας 5.3.5.2: Είσοδοι, Κανόνας 2	167
A.4	Ασαφές σύστημα ενότητας 5.3.5.2: Έξοδος, Κανόνας 2	167
A.5	Ασαφές σύστημα ενότητας 5.3.5.2: Είσοδοι, Κανόνας 3	168
A.6	Ασαφές σύστημα ενότητας 5.3.5.2: Έξοδος, Κανόνας 3	168
A.7	Ασαφές σύστημα ενότητας 5.3.5.2: Είσοδοι, Κανόνας 4	168
A.8	Ασαφές σύστημα ενότητας 5.3.5.2: Έξοδος, Κανόνας 4	168
A.9	Ασαφές σύστημα ενότητας 5.3.5.2: Είσοδοι, Κανόνας 5	168
A.10	Ασαφές σύστημα ενότητας 5.3.5.2: Έξοδος, Κανόνας 5	168
A.11	Ασαφές σύστημα ενότητας 5.3.5.2: Είσοδοι, Κανόνας 6	169
A.12	Ασαφές σύστημα ενότητας 5.3.5.2: Έξοδος, Κανόνας 6	169
A.13	Ασαφές σύστημα ενότητας 5.3.5.3: Είσοδοι, Κανόνας 1	170
A.14	Ασαφές σύστημα ενότητας 5.3.5.3: Έξοδος, Κανόνας 1	170
A.15	Ασαφές σύστημα ενότητας 5.3.5.3: Είσοδοι, Κανόνας 2	170
A.16	Ασαφές σύστημα ενότητας 5.3.5.3: Έξοδος, Κανόνας 2	170
A.17	Ασαφές σύστημα ενότητας 5.3.5.3: Είσοδοι, Κανόνας 3	170
A.18	Ασαφές σύστημα ενότητας 5.3.5.3: Έξοδος, Κανόνας 3	171
A.19	Ασαφές σύστημα ενότητας 6.3.4: Είσοδοι, Κανόνας 1	172
A.20	Ασαφές σύστημα ενότητας 6.3.4: Έξοδος, Κανόνας 1	172
A.21	Ασαφές σύστημα ενότητας 6.3.4: Είσοδοι, Κανόνας 2	172
A.22	Ασαφές σύστημα ενότητας 6.3.4: Έξοδος, Κανόνας 2	172
A.23	Ασαφές σύστημα ενότητας 6.3.4: Είσοδοι, Κανόνας 3	172
A.24	Ασαφές σύστημα ενότητας 6.3.4: Έξοδος, Κανόνας 3	173
A.25	Ασαφές σύστημα ενότητας 6.3.4: Είσοδοι, Κανόνας 4	173
A.26	Ασαφές σύστημα ενότητας 6.3.4: Έξοδος, Κανόνας 4	173

ΠΡΟΛΟΓΟΣ

Ο χώρος της όρασης υπολογιστών μου είχε κεντρίσει το ενδιαφέρον από τα χρόνια μου ως προπτυχιακός φοιτητής στο Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχ/κών Υπολογιστών, της πολυτεχνικής σχολής του ΑΠΘ. Η ώριμη ενασχόλησή μου με την επεξεργασία εικόνας και βίντεο, ωστόσο, ξεκίνησε το 2004, όταν ξεκίνησα το διετές μεταπτυχιακό πρόγραμμα του τμήματος της Πληροφορικής ΑΠΘ, με κατεύθυνση στα ψηφιακά μέσα. Στα πλαίσια του μεταπτυχιακού προγράμματος είχα την τύχη να ενασχοληθώ με ερευνητικά θέματα, στα πλαίσια διεθνών προγραμμάτων, και να διαπιστώσω την ερευνητική μου κλίση στο χώρο. Το Δεκέμβριο του 2006, είχα τη χαρά να γίνω δεκτός για εκπόνηση διδακτορικής διατριβής στο Εργαστήριο Ψηφιακής Επεξεργασίας Εικόνας, Βίντεο και Πολυμέσων, και να εργασθώ επάνω σε θέματα επικοινωνίας ανθρώπου υπολογιστή, επικεντρωμένα σε όραση υπολογιστών, με πληθώρα άμεσα πρακτικών εφαρμογών, γεγονός που κατέστησε την έρευνα ακόμα πιο προκλητική.

Σε όλη αυτήν την πορεία, είχα την ιδιαίτερη χαρά να έρχομαι, σε πολύ τακτικά χρονικά διαστήματα, σε επαφή με καταξιωμένους επιστήμονες άλλων ιδρυμάτων και, κυρίως του εξωτερικού, με τους οποίους έχουν ξεκινήσει συνεργασίες οι οποίες έπαιζαν καθοριστικό ρόλο στη διαμόρφωση και ωρίμανση της έρευνάς μου. Επιπλέον, οι πολλές ευκαιρίες που μου δόθηκαν να παρουσιάζω, ανά τακτά χρονικά διαστήματα την εξέλιξη της ερευνητικής μου δραστηριότητας σε διεθνή συνέδρια, αλλά και να ενημερώνομαι με τον καλύτερο δυνατό τρόπο για τις εξελίξεις στο αντικείμενό μου, αποτέλεσαν καίριο και ανεκτίμητο παράγοντα στην εξέλιξη της δουλειάς μου.

Κατά τη διάρκεια της εκπόνησης της διατριβής μου, είχα πάντοτε πλάι μου τον επιβλέποντα καθηγητή ΕΜΠ, κ. Στέφανο Κόλλια. Θα ήθελα να τον ευχαριστήσω θερμά για την πολύτιμη καθοδήγηση, αλλά και υποστήριξή του οποτεδήποτε αυτό χρειαζόταν. Ιδιαίτερες ευχαριστίες, οφείλω, και στον καθ. ΕΜΠ κ. Ανδρέα Σταφυλοπάτη. Μεγάλο ρόλο στην ολοκλήρωση της πορείας μου αυτής, έπαιξε ο διδάκτωρ ΕΜΠ και Ερευνητής β' ΕΠΙΣΕΥ, κ. Κώστας Καρπούζης. Η συνεργασία μας ήταν καθημερινή, και με παρότρυνε πάντοτε για νέες ερευνητικές περιοχές, στις οποίες ενδεχομένως, αρκετές φορές να δίσταζα να εισέλθω. Για την άριστη επιστημονικά συνεργασία μας, αλλά και τη φιλική του διάθεση, του οφείλω θερμό ευχαριστώ.

Επίσης, θα ήθελα να ευχαριστήσω θερμά τους Λώρη Μαλατέστα, Αμαρυλλίδα Ραουζαίου, Παρασκευή Τζούβελη, Γιώργο Καρυδάκη, διδάκτορες ΕΜΠ, τον Σταύρο Αποστόλου, διπλωματούχο Ηλεκτρολόγο Μηχανικό ΕΜΠ, αλλά και τον Christopher Peters, λέκτορα πανεπιστημίου Coventry, για την άριστη και φιλική συνεργασία που είχαμε όλα αυτά τα χρόνια, η οποία σε αρκετές περιπτώσεις οδήγησε σε ερευνητικές εργασίες που, θέλω να πιστεύω πως μπορούν να ωθήσουν την περιοχή λίγο ακόμα μπροστά. Ένα ευχαριστώ, ακόμη, στον Δημήτρη Σουφλερό, με τον οποίο είχαμε μία πολύ καλή συνεργασία σε επίπεδο διπλωματικής εργασίας, η οποία οδήγησε σε

πρωτότυπη και δημοσιευμένη έρευνα.

Τέλος, ένα θερμό ευχαριστώ στους γονείς μου Κλεοπάτρα και Παναγιώτη, στην αδερφή μου Κωνσταντία, και στη φίλη μου, Σ. Φουτσιτζή που, αν και για πρώτη φορά σε απόσταση από εμένα, τους ένιωθα όλους πάντα δίπλα μου.

Στυλιανός Αστεριάδης
Αθήνα, Μάρτιος 2011

ΠΕΡΙΛΗΨΗ

Το αντικείμενο έρευνας της διατριβής αυτής σχετίζεται με τα βήματα που αφορούν στην αναγνώριση και μοντελοποίηση συμπεριφοράς σε περιβάλλοντα αλληλεπίδρασης με μηχανή, όταν δεν υπάρχει συγκεκριμένη γνώση και εξειδικευμένος εξοπλισμός, πέρα από μία απλή διαδικτυακή κάμερα. Η έρευνα επικεντρώθηκε στην πηγή πληροφορίας που σχετίζεται με την ανάλυση και κίνηση του προσώπου. Συνήθως τέτοια συστήματα στηρίζονται σε παραδοχές σχετικά με την τοπολογική θέση του χρήστη, τις παραμέτρους της κάμερας, ή σε χρησιμοποίηση εξειδικευμένου υλικού. Πέρα από την προσπάθεια που έγινε στο πλαίσιο της διατριβής αυτής προς την κατεύθυνση αποφυγής τέτοιων παραδοχών, βασικός άξονας ήταν ένα ολοκληρωμένο σύστημα, τα επιμέρους στοιχεία του οποίου δεν θα βασίζονται σε μια στατική αλληλουχία, αλλά σε εκ βάθους μελέτη της καταλληλότητας και προσαρμοστικότητάς τους κατά τη διάρκεια της αλληλεπίδρασης. Προτείνονται επιμέρους αρχιτεκτονικές, στοχευμένες σε ένα περιβάλλον μη παρεμβατικό, που ενθαρρύνει τον αυθορμητισμό κινήσεων και φυσικότητα στον φωτισμό και το φόντο. Μεγάλη πρόκληση προς αυτήν την κατεύθυνση ήταν η παραπέρα μοντελοποίηση των μετρούμενων μεγεθών, για την εκπαίδευση συστημάτων τέτοιων που να προσεγγίζουν την ανθρώπινη αντίληψη ως προς την ενασχόληση χρήστη σε περιβάλλοντα αλληλεπίδρασης, με βάση την ανάλυση των χαρακτηριστικών και των κινήσεων της κεφαλής του. Τα επιμέρους θεωρητικά συμπεράσματα επιβεβαιώνονται από πειραματικά αποτελέσματα σε "απαιτητικές" βάσεις, η υποκειμενικότητα στην επισήμειωση των οποίων, σε αρκετές περιπτώσεις, προκάλεσε σύγχυση ακόμα και στον ανθρώπινο παράγοντα. Ωστόσο, τα αποτελέσματα ενθαρρύνουν την ανάπτυξη μη παρεμβατικών μεθοδολογιών μοντελοποίησης οπτικής επικοινωνίας βασισμένης σε ανάλυση και κίνηση προσώπου, σε πληθώρα εφαρμογών στο γενικότερο πλαίσιο της συναισθηματικής υπολογιστικής.

ABSTRACT

The research field, covered by this thesis is strongly related to recognition and modelling of human behavior in Human-Computer-Interaction environments, using face analysis as input modality. Research focuses on those cases where no specific knowledge regarding the set up, or specialized equipment exists, apart from simple hardware, like a common web-camera. Normally, such systems are based on admissions regarding user position, camera parameters or specific hardware. Beyond the effort to avoid such admissions, one of the basic principles of this thesis was research on a series of components, not statically positioned in the architecture, but dynamically emphasized throughout each process. Each component architecture has been worked on independently, aimed at non intrusive environments, encouraging spontaneity in movements, as well as unpretending lighting conditions and background. One of the large challenges towards these directions, on a second level, was modelling of extracted facial data, in order to train systems that would try to imitate human perception in terms of engagement in human-computer interaction scenarios. Experimental results on "demanding" datasets (annotated and developed for the purposes of this research) highlighted the prospect of employing non intrusive mechanisms for inferring engagement based on non verbal communication, using face analysis.

Κατάλογος Απόδοσης Ακρωνυμίων

BU dataset	:	Boston University dataset
DVF	:	Distance Vector Field
ΔΠΑ	:	Διανυσματικό Πεδίο Αποστάσεων
CNN	:	Convolutional Neural Networks
ΣΝΔ	:	Συνελικτικά(ο) Νευρωνικά(ό) Δίκτυα(ο)
AAM	:	Active Appearance Models
HPEG dataset	:	Head Pose and Eye Gaze dataset
PRFR	:	Pairwise Reinforcement of Feature Responses
GPF	:	General Projection Functions
VPF	:	Variance Projection Functions
IPF	:	Integral Projection Functions
HOG	:	Histogram of Oriented Gradients
SVD	:	Singular Value Decomposition
CHM	:	Cylinder Head Model
VAO	:	Virtual Attention Object
ToM	:	Theory of Mind
ANFIS	:	Adaptive-Network-based Fuzzy Inference Systems
SFM	:	Structure From Motion
ECA	:	Embodied Conversational Agent
PERCLOS	:	PER- cent eyelid CLOSure

Κατάλογος Απόδοσης Όρων

Action Units	:	μονάδες δράσης
Active Appearance Models	:	Μοντέλα Ενεργού Εμφάνισης
alertness	:	εγρήγορση
attention	:	προσοχή
Bayesian Networks	:	Μπεϋζιανά Δίκτυα
confusion matrix	:	πίνακας σύγχυσης
connected component labelling	:	επισηματοθέτηση συνεκτικών περιοχών
Convolutional Neural Networks	:	Συνελικτικά Νευρωνικά Δίκτυα
context	:	πλαίσιο, πλαίσιο γνώσης
co-occurrence graph	:	γράφος συμπτώσεων
Distance Vector Fields	:	Διανυσματικά Πεδία Απόστασης
engagement	:	ενασχόληση
false alarm	:	ψευδής συναγερμός
ground truth	:	δεδομένα αναφοράς
heatmap	:	θερμικός χάρτης
hue	:	χρωματικότητα
late fusion	:	εκ των υστέρων σύντηξη
leave one out	:	πρωτόκολλο παράλειψης στιγμιοτύπου
manifold	:	πολύπτυχο
non-rigid models	:	παραμετροποιημένα μοντέλα
rigid motion model	:	συμπαγή μοντέλα κίνησης
saturation	:	κορεσμός
segmentation	:	τιμηματοποίηση
shared attention senario	:	σενάριο προσοχής από κοινού
social understanding	:	κοινωνική κατανόηση
tansig	:	υπερβολική εφαπτομένη
template matching	:	ταίριασμα προτύπων
Theory of Mind	:	Θεωρία του Νού

Κεφάλαιο 1

Εισαγωγή

Με τη ραγδαία εξέλιξη των υπολογιστικών συστημάτων, αλλά και τη διείσδυσή τους σε όλους τους τομείς της ζωής του σύγχρονου ανθρώπου, η επιστήμη που μελετά την Επικοινωνία Ανθρώπου-Μηχανής έχει γνωρίσει ιδιαίτερα μεγάλη άνθηση. Θέματα όπως επεξεργασία και κατανόηση ανθρώπινου λόγου, ζητήματα πρακτικών και αποτελεσματικών διεπαφών, τόσο σε επίπεδο υλικού, όσο και λογισμικού, έρευνα γύρω από την αναγνώριση οπτικής πληροφορίας που παρέχεται από τον άνθρωπο και το περιβάλλον του, κατέχουν πολύ υψηλή θέση στις ερευνητικές δραστηριότητες πολλών ομάδων παγκοσμίως. Στο ίδιο πλαίσιο, πολλές έρευνες γίνονται γύρω από την αποτελεσματικότερη μοντελοποίηση της συμπεριφοράς χαρακτήρων του ίδιου του εικονικού περιβάλλοντος. Μέσα από έρευνες, έχει αποδειχτεί ότι, όσο περισσότερο πλησιάζει η συμπεριφορά ενός ψηφιακού πράκτορα σε αυτήν του ανθρώπου, τόσο πιο αυθόρμητη και αβίαστη συμπεριφορά θα έχει ο ίδιος ο άνθρωπος. Αυτός ο αυθορμητισμός είναι άμεση συνάρτηση και της ίδιας της διεπαφής μέσω της οποίας ο χρήστης καλείται να επικοινωνήσει με κάποιον εικονικό πράκτορα ή ένα πολυμεσικό υλικό, γενικότερα.

Μία από τις σημαντικότερες πηγές πληροφορίας στην επικοινωνία μεταξύ των ανθρώπων είναι, πέρα από την έκφραση του προσώπου, η κατευθυντικότητα του βλέμματος του συνομιλητή και οι μικρές - ίσως και ανεπαίσθητες - κινήσεις του κεφαλιού. Η ιδέα αυτή αποτελεί αντικείμενο μελέτης και στο πεδίο επικοινωνίας ανθρώπου-εικονικού πράκτορα. Η κατευθυντικότητα βλέμματος, βέβαια, αποτελεί πολύτιμη πηγή πληροφόρησης και σε άλλου είδους συστήματα όπου η εκτίμηση των επιπέδων προσοχής του χρήστη είναι καίριας σημασίας (πχ. οδήγηση).

Για τους παραπάνω λόγους, στην παρούσα διατριβή μελετώνται ζητήματα αναγνώρισης κατευθυντικότητας βλέμματος, με κύρια έμφαση στην περιστροφή του κεφαλιού. Δίνεται ιδιαίτερη έμφαση στη λειτουργικότητα τέτοιων συστημάτων κάτω από ρεαλιστικές συνθήκες, όπως μη ελεγχόμενος φωτισμός, αυθόρμητες κινήσεις χρηστών, πολύπλοκο, ρεαλιστικό φόντο, και απόκρυψη χαρακτηριστικών. Απαιτήση της συγκεκριμένης εργασίας είναι να μπορεί να ενσωματωθεί σε απλά υπολογιστικά συστήματα, που κάνουν χρήση απλής δικτυακής, μη βαθμονομημένης κάμερας, αποφεύγοντας συγκεκριμένες, ενδεχομένως περισσότερες της μίας μηχανής λήψης, πολύπλοκες διατάξεις. Μελετάται η δυνατότητα χρήσης τέτοιων αρχιτεκτονικών σε εφαρμογές διεπαφών ανθρώπου-υπολογιστή, μοντελοποιώντας τη συμπεριφορά του χρήστη, ορίζοντας μία σειρά καταστάσεων, οι οποίες, εκτός από το βλέμμα, λαμβάνουν υπόψη και επιπλέον πληροφορία κίνησης χαρακτηριστικών προσώπου. Στο πλαίσιο της διατριβής προτείνονται αλγόριθμοι και επαληθεύεται η λειτουργικότητά τους με τις απαιτήσεις που τίθενται εδώ, ενώ προτείνονται πρωτότυπες αρχιτεκτονικές εντοπισμού και παρα-

κολούθησης χαρακτηριστικών προσώπου, αναγνώρισης και μοντελοποίησης προσοχής χρήστη, οι οποίες κινούνται στα πλαίσια του απαιτούμενου ρεαλισμού και έχουν στόχο να είναι ανταγωνιστικές με δημοσιευμένες τεχνικές, αντίστοιχης ή χαμηλότερης ακρίβειας, οι οποίες λειτουργούν με συγκεκριμένους περιορισμούς.

1.1 Ερευνητικό πλαίσιο και συνεισφορά της διατριβής

Η περιοχή του προσώπου είναι πηγή πλούσιας, μη προφορικής πληροφορίας στην καθημερινή ζωή του ανθρώπου. Τα παιδιά, από βρεφική κιόλας ηλικία, μαθαίνουν να αναγνωρίζουν [105] την κατευθυντικότητα του βλέμματος των ατόμων που τα φροντίζουν, η οποία δρα σαν ένα φίλτράρισμα των σημαντικών γεγονότων ή σκηνών στον περιβάλλοντα χώρο. Αυτή η ιδιότητα ακολουθεί τον άνθρωπο για την υπόλοιπη ζωή του: Η κατευθυντικότητα του βλέμματος των άλλων, είναι πάντα ένα συνειδητό ή μη, ερέθισμα, στο να κατευθύνει ο άνθρωπος το δικό του βλέμμα προς κατεύθυνση τέτοια, ώστε να κοιτάζει κάτι ενδιαφέρον, να αποφύγει κάποιον κίνδυνο, κλπ. Πολύ σημαντική είναι η πληροφορία του βλέμματος και σε συζητήσεις. Πολλοί είναι αυτοί που χρησιμοποιούν το βλέμμα τους, είτε για να δείξουν ότι προτίθενται να πάρουν το λόγο, είτε για να αναθέσουν το λόγο σε κάποιον συνομιλητή τους. Επίσης, πέρα από το βλέμμα, οι κινήσεις του κεφαλιού μεταφέρουν πολύτιμα μηνύματα σε μια συζήτηση: Καταφατικό νεύμα από μέρους ακροατή ενθαρρύνει τον ομιλητή να συνεχίσει, χωρίς να χρειαστεί ο τελευταίος να διακόψει το λόγο του για να ρωτήσει αν τα λεγόμενά του γίνονται κατανοητά. Σημαντική, επίσης, είναι η κινησιολογία του κεφαλιού και για τη μεταφορά κι άλλων νοημάτων: Συμφωνία, σύγχυση, αντίθεση, προσοχή, περιέργεια, είναι κάποια μόνο βασικά νοήματα που κάποιος μπορεί να μεταφέρει απλά και μόνο με το βλέμμα του.

Μηνύματα, ωστόσο, μεταφέρονται, ακόμα και με υπερβολικές κινήσεις του κεφαλιού, αφού, αυτές μπορούν να είναι συνώνυμες με συμβατικούς τρόπους υπόδειξης κάποιου να κοιτάζει προς ένα σημείο χρησιμοποιώντας το δάχτυλο. Πολλές φορές, κάποια απότομη κίνηση του κεφαλιού μπορεί να είναι σημάδι έκπληξης ή ξαφνικής ανησυχίας. Τέτοια σημάδια είναι πολύ δύσκολο να παραβλεφθούν, ακόμα και αν ο συνοδευτικός λόγος είναι αντιφατικός με τις κινήσεις του κεφαλιού ή του βλέμματος [91].

Καθόλου αμελητέος, ωστόσο, δεν είναι και ο ρόλος της έννοιας της *αμοιβαίας προσοχής*: Πρόκειται για εκείνη τη συνθήκη όπου δύο άνθρωποι κοιτάζονται μεταξύ τους και, έτσι, εγκαθιδρύεται αμοιβαίο ενδιαφέρον για το θέμα συζήτησης. Η έννοια της αμοιβαίας προσοχής, ωστόσο, δεν αναφέρεται κατ' ανάγκη στο διάλογο: ενδέχεται να αφορά και άλλες καταστάσεις, όπως αυτή του πεζού ο οποίος, πριν διασχίσει τον δρόμο, κοιτάζει να βεβαιωθεί εάν ο διερχόμενος οδηγός αντιλήφθηκε την παρουσία του, κοιτάζοντάς τον.

Όσον αφορά στις τεχνολογικές εφαρμογές, η αυτόματη αναγνώριση κατευθυντικότητας βλέμματος του χρήστη, και η παραπέρα μοντελοποίησή της, ερευνάται, ήδη, σε πλειάδα καταστάσεων: Εκτίμηση της προσοχής οδηγού [35], περιβάλλοντα μάθησης [10], επικοινωνία ανθρώπου-εικονικών πρακτόρων [141, 140, 20] και ηλεκτρονικά παιχνίδια [68] είναι κάποιες από τις εφαρμογές όπου η χρήση της γνώσης της κατευθυντικότητας βλέμματος χρηστών παίζει μεγάλο ρόλο. Τα τελευταία χρόνια, επίσης, υπάρχει έντονη ερευνητική δραστηριότητα επάνω σε θέματα ενδιαφέροντος χρήστη σε πολυμεσικά υλικά (κυρίως ιστοσελίδες), με χρήση των λεγόμενων *θερμικών χαρτών* (heatmaps) [155, 150]: μέσω της παρακολούθησης του βλέμματος του χρήστη,

ελέγχονται τα σημεία εκείνα στα οποία δείχνει το μεγαλύτερο ενδιαφέρον, με στόχο τη βελτιστοποίηση της σχεδίασης της εκάστοτε εφαρμογής (πχ διαφημίσεις, ενημερώσεις, κλπ).

Για την υποστήριξη των παραπάνω εφαρμογών, στο εμπόριο ήδη βρίσκονται πολλά προϊόντα παρακολούθησης βλέμματος, προερχόμενο, είτε από τα μάτια ¹, είτε από την κίνηση του κεφαλιού [187] (δεν φαίνεται, ωστόσο, σε επίπεδο υλικού, να υπάρχει διάταξη εξαγωγής βλέμματος από μάτια και περιστροφή κεφαλής ταυτοχρόνως). Τέτοιες διατάξεις χρησιμοποιούν, συνήθως, μαγνητικούς αισθητήρες, κάμερες ευαίσθητες στο υπέρυθρο φως και εικόνες αρκετά μεγάλης ανάλυσης, επικεντρωμένες στα μάτια. Τέτοιες συσκευές, ενώ δίνουν αρκετά καλή ακρίβεια στα αποτελέσματα, είτε χρειάζεται με κάποιο τρόπο ο χρήστης να τις «φορέσει», είτε να ελέγχει τις κινήσεις του κεφαλιού του, και για αυτούς τους λόγους, ενδέχεται να μειώνουν τη φυσικότητα στην κίνησή του. Σε λιγότερο εμπορικά προϊόντα, και περισσότερο μεθοδολογίες υπό έρευνα, συναντώνται συστήματα που χρησιμοποιούν συγκεκριμένα διάταξη μηχανών λήψης [99], με συγκεκριμένα χαρακτηριστικά ή ανάγκη για εκ των προτέρων βαθμονόμηση [22, 188].

Λαμβάνοντας υπόψιν τα παραπάνω, η παρούσα διατριβή ασχολείται με τη μελέτη μεθοδολογιών αναγνώρισης και μοντελοποίησης προσοχής χρήστη, με κύρια παράμετρο το βλέμμα, αλλά και με εκμετάλλευση δευτερευόντων χαρακτηριστικών (ποιοτική εκτίμηση κίνησης, μεταβολές των αποστάσεων μεταξύ χαρακτηριστικών προσώπου, ταχύτητα κίνησης κεφαλιού, κλπ) σε μη παρεμβατικά, μη βαθμονομημένα περιβάλλοντα, με μη επιτηδευμένο φωτισμό και φόντο, απαλλαγμένα από την ανάγκη εκπαίδευσης ξεχωριστά για κάθε χρήστη, και με τη χρήση απλής, δικτυακής κάμερας.

Αρχικά ερευνήθηκε η υλοποίηση συστήματος ανάκτησης περιστροφής του προσώπου, ως δείκτης κατευθυντικότητας προσοχής. Ελέγχθηκε η δυνατότητα υιοθέτησης αλγοριθμικών μεθόδων παρακολούθησης χαρακτηριστικών σημείων και η εφαρμοσιμότητά τους σε πραγματικά περιβάλλοντα. Κατόπιν, ιδιαίτερη βαρύτητα δίνεται στη μελέτη συστήματος που συνδυάζει τοπική με ολιστική πληροφορία. Στην περίπτωση της τοπικής πληροφορίας, χρησιμοποιείται μετασχηματισμός Διανυσματικών Πεδίων Απόστασης (ΔΠΑ) ως τρόπος αξιοποίησης τοπολογικής πληροφορίας προερχόμενης από κάθε εικονοστοιχείο της περιοχής του προσώπου, ενώ γίνεται εκμετάλλευση των αλλαγών στις θέσεις χαρακτηριστικών σημείων στο πρόσωπο σε σχέση με τα όριά του, για υπολογισμό της περιστροφής. Με τον τρόπο αυτό, έως ένα σημείο, το σύστημα είναι σε μεγαλύτερο βαθμό εξαρτώμενο από τον παράγοντα ευρωστία, και λιγότερο από ανάγκη για ακριβή παρακολούθηση. Ως ολιστική πληροφορία έγινε χρήση Συνελκτικών Νευρωνικών Δικτύων (ΣΝΔ) [92]. Με την επιλογή των ΣΝΔ, το μεγαλύτερο κέρδος είναι ότι επιτυγχάνεται υψηλός βαθμός ανεκτικότητας σε μη απόλυτη ευθυγράμμιση ανάμεσα σε δεδομένα εκπαίδευσης και δεδομένα εισόδου, αλλά και η δυνατότητα κατάλληλης αρχιτεκτονικής μιας σειράς υπο-ταξινομητών οι οποίοι λαμβάνουν υπόψιν πληροφορία κίνησης προσώπου της τρέχουσας χρονικής στιγμής.

Στο πλαίσιο της αναγνώρισης ερευνήθηκε το πρόβλημα της ανάκτησης βλέμματος ματιών, με το πρόσωπο υπό γωνία, και χρήση απλής διαδικτυακής κάμερας. Στη μέχρι σήμερα βιβλιογραφία, το πρόβλημα της ανάκτησης κατευθυντικότητας βλέμματος ματιών έχει προσεγγιστεί κυρίως σε επίπεδο υλικού, στο οποίο αναφέρονται διατά-

¹http://www.tobii.com/scientific_research/products_services/eye_tracking_hardware/tobii_t60_t120_eye_trackers.aspx
<http://www.seeingmachines.com/product/facelab/>
<http://www.smivision.com/en/eye-gaze-tracking-systems/products/3d-vog.html>

Ξεις που κάνουν χρήση υπέρυθρου φωτός, μικροκάμερες τοποθετημένες πολύ κοντά στα μάτια και πολυκάμερα συστήματα, τοποθετημένα σε συγκεκριμένα σημεία ως προς τα μάτια του χρήστη. Τέτοια συστήματα, συνήθως απαιτούν ο χρήστης να είναι σταθερός ως προς αυτά και, τυχόν μετακίνησή του, ενδέχεται να απαιτεί εκ νέου βαθμονόμησή τους. Σε επίπεδο λογισμικού, με χρήση μονοκάμερων συστημάτων και φυσικό φωτισμό, στη βιβλιογραφία έχουν προταθεί λίγες μέθοδοι συνδυαστικού χαρακτήρα περιστροφής προσώπου και ανάκτησης βλέμματος ματιών [174],[76], ενώ, στην πλειονότητά τους, τα προτεινόμενα συστήματα αναφέρονται σε ανάκτηση κατευθυντικότητα βλέμματος ματιών με ευαισθησία σε κινήσεις του κεφαλιού [83],[160],[167].

Το αντικείμενο μελέτης της μοντελοποίησης στην διατριβή είναι η αναπαράσταση της κατάστασης του χρήστη ως προς εργασία που έχει μπροστά του. Δανειζόμενοι αρχές και τεχνικές που χρησιμοποιούν οι άνθρωποι για να κατανοήσουν, να προβλέψουν και να χειριστούν προθέσεις και συμπεριφορές άλλων ανθρώπων, μέσα από τη «*Θεωρία του Νου*» (Theory of Mind - ToM) [14], αναπτύχθηκαν συστήματα ασαφούς λογικής, τα οποία αντιστοίχησαν τα μετρούμενα μεγέθη με καταστάσεις σχετικές με *ενασχόληση ή/και προσοχή* του χρήστη απέναντι σε κάποιο πολυμεσικό υλικό. Ερευνήθηκε, επιπλέον, η δυνατότητα μελέτης αλληλεπίδρασης χρήστη με εικονικό πράκτορα υπό θεωρήσεις διαφορετικών σεναρίων και μελετήθηκε μία ιεραρχική δόμηση των επιπέδων ενδιαφέροντος του χρήστη, ξεκινώντας από απλή εκτίμηση μετρούμενων μεγεθών και καταλήγοντας στην έννοια της ενασχόλησης. Μέσα από τα παραπάνω, καθώς και στατιστική επεξεργασία των πλέον χρήσιμων μεγεθών για χαρακτηρισμό καταστάσεων, μέσα από την παρούσα δουλειά, προτείνεται στη βιβλιογραφία συγκεκριμένη μοντελοποίηση της έννοιας προσοχής από μη προφορικά χαρακτηριστικά, η οποία ελπίζουμε ότι θα αποτελέσει αφετηρία για μελλοντικές εργασίες.

Στο στάδιο της σύντηξης τοπικού και ολιστικού επιπέδου πληροφορίας για εκτίμηση περιστροφής κεφαλιού, η μία μέθοδος ενημερώνεται και εκμεταλλεύεται πληροφορία προερχόμενη από την άλλη, ενώ, για βελτιωμένα αποτελέσματα, ερευνήθηκε η δυνατότητα χρήσης πιθανοτικού δικτύου, το οποίο δυναμικά, και ανάλογα με δευτερεύουσες πληροφορίες προερχόμενες από το στάδιο της ανάλυσης συνδέει τις δύο πηγές καταλήγοντας σε συνολική κατανομή πιθανότητας στον κόμβο εξόδου. Μέσα από πειράματα, διαφάνηκε πως η έννοια της *προσοχής* σα γνώση πλαισίου, δύναται να αποτελέσει κατάλληλη πληροφορία πλαισίου ώστε να δράσει υποστηρικτικά στη δυναμική σύντηξη της τοπικής και ολιστικής πληροφορίας.

1.2 Δομή της διατριβής

Η δομή της διατριβής είναι η ακόλουθη:

Στο κεφάλαιο 2 περιγράφεται η διαδικασία εντοπισμού χαρακτηριστικών προσώπου στο αρχικό καρέ μιας ακολουθίας εικόνων, όπου θεωρείται ότι ο χρήστης κοιτάζει την κάμερα μετωπικά. Μετά από μία βιβλιογραφική επισκόπηση του χώρου, περιγράφεται ο τρόπος με τον οποίο εντοπίζονται 19 χαρακτηριστικά σημεία στο πρόσωπο. Για την εύρεσή τους, επιλέχθηκε μία ιεραρχική μεθοδολογία στην οποία πρώτα εντοπίζεται η περιοχή του προσώπου, με απώτερο στόχο το σύστημα να είναι ανεξάρτητο κλιμάκωσης. Για την εύρεση των χαρακτηριστικών χρησιμοποιήθηκε μετασχηματισμός βασισμένος περισσότερο σε τοπολογική πληροφορία, παρά σε χρωματικότητα-φωτεινότητα, έτσι ώστε να μην υπάρχει εξάρτηση από συγκεκριμένες συνθήκες φωτισμού. Στο κεφάλαιο παρουσιάζονται αναλυτικά πειραματικά αποτελέσματα, τα οποία, στην πορεία, αποδείχτηκαν αρκετά ικανοποιητικά για τις εφαρμογές που περιγράφο-

νται παρακάτω.

Το κεφάλαιο 3 πραγματεύεται θέματα αναγνώρισης περιστροφής κεφαλιού με χρήση μεθόδων όρασης υπολογιστών. Αρχικά γίνεται μία επισκόπηση των σχετικών εργασιών και των προσεγγίσεων που ακολουθούνται, περιγράφονται τα δεδομένα που είναι διαθέσιμα στην ερευνητική κοινότητα που ασχολείται με το αντικείμενο και περιγράφεται η βάση ακολουθιών εικόνων που αναπτύχθηκε στο πλαίσιο της παρούσας διατριβής. Η βάση που αναπτύχθηκε συνδυάζει κατευθυντικότητα βλέμματος, τόσο από περιστροφή κεφαλιού, όσο και από αυτήν των ματιών, σε ακολουθίες εικόνων. Μέχρι τη συγγραφή της διατριβής, δεν υπήρχε στο διαδίκτυο διαθέσιμη τέτοια βάση σε μορφή ακολουθιών. Τέλος, περιγράφονται τα βήματα εξαγωγής περιστροφής κεφαλιού με χρήση 2Δ αλγορίθμων και κανόνων επανεκκίνησης για ενσωμάτωση ακραίων περιπτώσεων κινήσεων και απόκρυψης χαρακτηριστικών, με σκοπό το σύστημα να είναι εύρωστο σε αυθόρμητες και απρόβλεπτες κινήσεις των χρηστών. Λαμβάνεται μέριμνα ώστε να υπάρχει διάκριση μεταξύ περιστροφικής και μεταφορικής κίνησης, αλλά και το σύστημα να είναι ανεξάρτητο κλίμακας, ενώ για βελτίωση των αποτελεσμάτων, χρησιμοποιούνται φίλτρα Kalman. Η επιβεβαίωση της καταλληλότητας της μεθόδου γίνεται σε δύο ευρέως διαδεδομένες βάσεις, την XM2VTS [104] και την βάση του Πανεπιστημίου της Βοστώνης (Boston University - BU) [22]. Στο στάδιο αυτό, δεν ενδιέφερε τόσο η ακρίβεια σε μοίρες, όσο η ποιοτική ευρωστία, η οποία σχετίζεται άμεσα με τις συνθήκες των πειραμάτων (φωτισμός, φόντο, ταχύτητα κίνησης) και τα αποτελέσματα που περιγράφονται στα αντίστοιχα κεφάλαια των πρακτικών εφαρμογών.

Στο κεφάλαιο 4 περιγράφεται ένα πρωτότυπο σύστημα, το οποίο ερευνήθηκε για την εξαγωγή πιο λεπτομερών αποτελεσμάτων στην εκτίμηση των τριών βαθμών ελευθερίας της περιστροφής του κεφαλιού, με χρήση, τόσο τοπικής, όσο και ολιστικής πληροφορίας. Η μέθοδος βασίζεται σε παρακολούθηση χαρακτηριστικών προσώπου με χρήση Διανυσματικών Πεδίων Απόστασης [7] και τις αποστάσεις τους από τα όρια του προσώπου, και Συνελικτικά Νευρωνικά Δίκτυα [93] για αξιοποίηση ολιστικής πληροφορίας. Για την παρακολούθηση του προσώπου προτείνεται ένας αλγόριθμος προσαρμοστικής κατωφλίωσης, με κατώφλι που εξάγεται με αυτόματο τρόπο και είναι προσαρμοσμένο σε κάθε χρήστη. Επιπλέον, για την ενσωμάτωση γνώσης της αναμενόμενης φυσιολογίας του προσώπου, η μέθοδος παρακολούθησης χαρακτηριστικών συνδέθηκε με έναν γκαουσιανό όρο που μοντελοποιεί το λόγο απόστασης ανάμεσα στα μάτια και ανάμεσα στα μάτια και το στόμα. Για την επαληθεύση της αποτελεσματικότητας της τοπικής τεχνικής χρησιμοποιήθηκαν δύο βάσεις, ενώ έγινε σύγκριση με άλλες γνωστές μεθόδους της βιβλιογραφίας. Τέλος, επαληθεύτηκε ότι ο αλγόριθμος παρακολούθησης χαρακτηριστικών που χρησιμοποιήθηκε είναι αξιόπιστος, συγκρίνοντάς τον με τον ευρέως διαδεδομένο αλγόριθμο οπτικής ροής [19]. Επίσης, μελετώνται τα Συνελικτικά Νευρωνικά Δίκτυα (Convolutional Neural Network - CNN)[93] ως αρχιτεκτονική αξιοποίησης ολιστικής πληροφορίας. Με είσοδο την περιοχή του προσώπου, καθώς και την γωνία υπό την οποία βρίσκεται περιστραμμένο το κεφάλι στο προηγούμενο καρέ, χρησιμοποιείται συγκεκριμένο υποσύνολο τέτοιων υπο-ταξινομητών δύο εξόδων. Οι υποταξινομητές είναι εκπαιδευμένοι με τέτοιο τρόπο, ώστε να διαχωρίζουν μεταξύ τους γειτονικές κλάσεις στο χώρο των γωνιών περιστροφής κεφαλής. Με τον τρόπο αυτό, επιτυγχάνεται, τόσο μικρότερη υπολογιστική πολυπλοκότητα (χρησιμοποιούνται μόνο γειτονικών κλάσεων ταξινομητές) αλλά και μεγαλύτερη ακρίβεια, αφού πρακτικά απίθανες έξοδοι αποκλείονται από τη θεώρηση των αποτελεσμάτων. Στο τέλος του κεφαλαίου προτείνεται μία σειρά μεθόδων σύ-

ντηξης των δύο επιπέδων πληροφορίας (τοπικό και ολιστικό). Μέσα από πειράματα, γίνεται φανερό ότι μεγάλη ακρίβεια λαμβάνεται από ένα πιθανοτικό μοντέλο, το οποίο μοντελοποιεί την αξιοπιστία του κάθε επιπέδου.

Στο κεφάλαιο 5 μελετώνται ζητήματα μοντελοποίησης μη παρεμβατικής αλληλεπίδρασης. Αρχικά γίνεται μία περιγραφή των σχετικών εργασιών και του τρόπου με τον οποίο μοντελοποιείται το βλέμμα για να εκτιμηθούν τα επίπεδα προσοχής και ενασχόλησης. Κατόπιν, υλοποιήθηκε σενάριο αμοιβαίας προσοχής μεταξύ χρήστη και εικονικού πράκτορα με χρήση της πλατφόρμας διασύνδεσης *Psychone*. Προτείνεται πολυεπίπεδη αναγνώριση ενδιαφέροντος χρήστη, ξεκινώντας με απλή αντιστοίχιση σε συντεταγμένες οθόνης και καταλήγοντας σε επίπεδα ενδιαφέροντος, λαμβάνοντας υπόψη το περιεχόμενο της αλληλεπίδρασης. Στο πλαίσιο της συγκεκριμένης εφαρμογής, πραγματοποιήθηκαν δύο πειράματα: Στο πρώτο αναλύεται η καταλληλότητα της μεθοδολογίας του κεφαλαίου 3, ενώ στο δεύτερο περιγράφεται πείραμα στο οποίο έλαβαν μέρος άτομα από το εργαστήριο σε ένα σενάριο εικονικού πωλητή. Μελετήθηκαν διαφορετικές εκδόσεις του σεναρίου, κάθε μία από τις οποίες μοντελοποιούσε διαφορετικές συμπεριφορές του πράκτορα, και μελετήθηκαν, τόσο το βλέμμα του κάθε χρήστη, όσο και η αντιληπτική του ικανότητα στα διαφορετικά σενάρια. Ερευνάται, επίσης, η δυνατότητα να χρησιμοποιηθούν μετρούμενα χαρακτηριστικά για εκτίμηση επιπέδων προσοχής του χρήστη, στην εκπαιδευτική διαδικασία. Χρησιμοποιήθηκε μία βάση παιδιών με μαθησιακές δυσκολίες, επισημειωμένη με δεδομένα αναφοράς σχετικά με την αντιμετώπισή τους απέναντι σε μαθησιακό υλικό. Η έρευνα αυτή έγινε στο πλαίσιο του έργου *AgentDysl*² και αποσκοπούσε στη δημιουργία μαθησιακής πλατφόρμας στην οποία η παρουσίαση του υλικού προσαρμόζεται στο προφίλ και την τρέχουσα κατάσταση του χρήστη. Στην εφαρμογή αυτή χρησιμοποιήθηκαν χαρακτηριστικά και πέραν της περιστροφής του κεφαλιού, όπως η κατευθυντικότητα του βλέμματος των ματιών, οι μετατοπίσεις του κεφαλιού κάθετα στο επίπεδο της κάμερας, κινήσεις των φρυδιών και του στόματος και ταχύτητα κίνησης του κεφαλιού. Καθώς, στο κεφάλαιο αυτό, η προσοχή θεωρείται ότι προέρχεται, τόσο από το κεφάλι, όσο και από τα μάτια, έγινε η σχετική βιβλιογραφική έρευνα. Μελετήθηκε η καταλληλότητα του κάθε χαρακτηριστικού και εκπαιδεύτηκε ένα νευροασαφές σύστημα για την κατηγοριοποίηση των χρηστών σε καταστάσεις.

Στο κεφάλαιο 6 χρησιμοποιείται το υβριδικό σύστημα που προτάθηκε στο κεφάλαιο 4, με μία προτεινόμενη τεχνική ανάκτησης βλέμματος ματιών, βασισμένη σε εξουδετέρωση της περιστροφής του προσώπου με χρήση κυλινδρικών μοντέλων και εικονική περιστροφή τους. Η περιστροφή του προσώπου σε συνδυασμό με την κατευθυντικότητα των ματιών, δίνονται σαν είσοδοι σε νευροασαφές δίκτυο αντίστοιχο με αυτό του κεφαλαίου 5, με στόχο τον υπολογισμό επιπέδων προσοχής του χρήστη. Ο έλεγχος έγινε στη βάση του Πανεπιστημίου της Βοστώνης [22], η οποία, πέρα από την επισημείωση σε περιστροφή κεφαλιού, με την οποία ήδη διατίθεται, επισημειώθηκε από εθελοντές, σε μελέτη που στήθηκε για τους σκοπούς αυτής της έρευνας, ως προς τα επίπεδα προσήλωσης στα οποία αντιλαμβάνονται ότι βρίσκονται οι συμμετέχοντες. Παρόμοια επισημείωση έγινε και εξαιρώντας την επίδραση της κατευθυντικότητας βλέμματος: Η επισημείωση αυτή χρησίμευσε ως είσοδος στην εκπαίδευση νευροασαφούς συστήματος, με εισόδους την περιστροφή κεφαλιού και τις κανονικοποιημένες αυξομειώσεις στην απόσταση των ματιών. Η εκτίμηση επιπέδων προσοχής για άγνωστες γωνίες περιστροφής και αποστάσεις ματιών, αξιοποιήθηκε ως γνώση η οποία αποτέλεσε κόμβο εισόδου σε μπεύζιανό δίκτυο σύντηξης ολιστικής με τοπική

²www.agent-dysl.eu/

πληροφορία. Με τον τρόπο αυτό, επιχειρείται να κλείσει ο βρόχος ανάμεσα στην εκτίμηση περιστροφής κεφαλιού και επιπέδων προσήλωσης. Μοντελοποιήθηκαν, με χρήση του μπεϋζιανού δικτύου σύντηξης πηγών πληροφορίας, οι συνθήκες εκείνες (επίπεδα προσήλωσης) κάτω από τις οποίες υπερτερεί η τοπική έναντι της ολιστικής πληροφορίας, και το αντίστροφο. Μέσα από πειράματα έγινε εμφανές πως, με χρήση αυτής της γνώσης, είναι δυνατό να προσεγγιστεί η σωστή κατανομή εξόδου στον αντίστοιχο κόμβο ενός μπεϋζιανού δικτύου, με στόχο τον υπολογισμό περιστροφής κεφαλιού. Κάτι τέτοιο έδωσε αρκετά καλύτερη ακρίβεια από ό,τι αντικαθιστώντας τη γνώση επιπέδων προσοχής με την υπολογισμένη γωνία περιστροφής του προηγούμενου καρέ, δίνοντας ώθηση σε έρευνα που θα συσχετίσει γνώση αναμενόμενων επιπέδων προσοχής του χρήστη, με μεθόδους αναγνώρισης κίνησης κεφαλιού.

Τέλος, στο κεφάλαιο 7, συνοψίζεται η συνεισφορά της διατριβής και αναφέρονται πιθανές μελλοντικές ερευνητικές προεκτάσεις, στις οποίες η παρούσα έρευνα μπορεί να παίζει σημαντικό ρόλο.

Κεφάλαιο 2

Εντοπισμός χαρακτηριστικών προσώπου με Διανυσματικά Πεδία Απόστασης

2.1 Ερευνητικό πλαίσιο στην εύρεση χαρακτηριστικών προσώπου

2.1.1 Γενικά

Στην πρόσφατη βιβλιογραφία, πολύ μεγάλο εύρος τεχνικών έχει κάνει την εμφάνισή του στο χώρο της ανίχνευσης προσώπων και χαρακτηριστικών προσώπου. Χάρη στη σχέση αυτών των περιοχών με πολλές πλευρές της καθημερινής, σύγχρονης ζωής, αλλά, επίσης, εξαιτίας του ότι το ανθρώπινο πρόσωπο έχει ευμετάβλητη και πολυποίκιλη υφή, πολλές διαφορετικές προσεγγίσεις έχουν προταθεί. Οι μέθοδοι αυτές μπορούν να χαρακτηριστούν ανάλογα με τα κριτήρια που χρησιμοποιούνται σε κάθε περίπτωση: Γεωμετρία του προσώπου, φωτεινότητα, χρωματική πληροφορία, καθώς και σχήμα χαρακτηριστικών προσώπου, είναι τα κυριότερα είδη πληροφορίας που χρησιμοποιούνται για την ανίχνευση χαρακτηριστικών προσώπου. Στη συντριπτική τους πλειοψηφία, οι δημοσιευμένες μέθοδοι βασίζονται σε μια αρχική εκτίμηση της θέσης του προσώπου. Αυτού του είδους οι τεχνικές δίνουν πολύ καλύτερα αποτελέσματα, γενικά, από εκείνες που προσπαθούν να ανιχνεύσουν χαρακτηριστικά (μάτια, στόμα, μύτη, κλπ.) σε μια εικόνα, χωρίς πληροφορία θέσης προσώπου. Στο πρώτο είδος προσεγγίσεων, προβλήματα κλιμάκωσης γενικά δεν υπάρχουν, αφού αυτά λύνονται, ουσιαστικά, στο στάδιο της ανίχνευσης προσώπου. Ωστόσο, στο δεύτερο είδος προσεγγίσεων, γίνεται μια αρχική εκτίμηση, εκτός της πιθανής θέσης του προσώπου, και σχετικά με το πιθανό μέγεθος του στόματος, των ματιών, της μύτης, κλπ. Συνήθως, προσαρμοσμένες στις βάσεις εικόνων που ελέγχουν, οι μέθοδοι αυτές δίνουν καλά αποτελέσματα, γι'αυτό και προτείνονται μονάχα για συγκεκριμένους σκοπούς, όπως, π.χ. έλεγχος εγρήγορσης οδηγών αυτοκινήτων [154], έλεγχος ταυτότητας, κ.α.

Οι μέθοδοι που χρησιμοποιούν ταίριασμα προτύπων (template matching) είναι, συνήθως, χρονοβόρες, ωστόσο, δίνουν πολύ πιο επιτυχημένα αποτελέσματα από τις μεθόδους που δε χρησιμοποιούν. Συνήθως, οι τεχνικές ταίριασματος προτύπων προσπαθούν να συγκρίνουν μοντέλα με κάποια περιοχή του προσώπου, χρησιμοποιώντας:

- Χαρακτηριστικά ενός μοντέλου ([34], [75],[30],[44],[142]), ή

- Συγκεκριμένες ιδιότητες ([183],[189],[39]).

Μέθοδοι που δε χρησιμοποιούν τεχνικές ταιριάσματος προτύπων συνήθως βασίζονται στο γεγονός ότι τα χαρακτηριστικά ενός προσώπου αντιστοιχούν σε σκούρες περιοχές σε σχέση με το υπόλοιπο πρόσωπο [16], [149],[129],[89], ή στο γεγονός ότι τα χαρακτηριστικά του προσώπου έχουν τις περισσότερες φορές, διαφορετική χρωματική πληροφορία από αυτήν του δέρματος [154]. Η δομή του προσώπου είναι, συνήθως, ένας πολύ καλός τρόπος για επαλήθευση αποτελεσμάτων, ή για απόρριψη ψευδών συναγερωμών (false alarms). Ο μετασχηματισμός Hough είναι, επίσης, ένας κοινός τρόπος για εύρεση ματιών, καθώς τα μάτια είναι το πλέον αξιόπιστο χαρακτηριστικό στην εύρεσή του από τα υπόλοιπα. Επιπλέον, τα μάτια, μπορούν να αποτελέσουν μία πολύ καλή αφετηρία στην ανίχνευση των υπολοίπων χαρακτηριστικών. Ωστόσο, ο μετασχηματισμός Hough δουλεύει καλά μόνο σε εικόνες αρκετά καλής ανάλυσης, ενώ αποτυγχάνει στις περιπτώσεις εκείνες όπου το πρόσωπο είναι στραμμένο προς τα δεξιά ή τα αριστερά, ή όταν τα μάτια είναι μερικώς κλειστά [39].

Τα κριτήρια που χρησιμοποιούνται για την αξιολόγηση του κάθε αλγορίθμου είναι πολλά. Σε πολλές μεθόδους, εξαρτώνται ξεκάθαρα από το μέγεθος των εικόνων που χρησιμοποιούνται, το είδος του χαρακτηριστικού που ανιχνεύεται (π.χ. περιοχή γύρω από το μάτι, κέντρο του ματιού, κλπ.). Αυτό βέβαια, κάνει δύσκολη τη σύγκριση των διαφόρων αλγορίθμων μεταξύ τους. Ένα αντικειμενικό κριτήριο προτάθηκε από τον Jesorsky [75]. Το κριτήριο αυτό είναι ανεξάρτητο κλίμακας, καθώς γίνεται κανονικοποίηση με την μεταξύ των ματιών απόσταση. Ουσιαστικά, αυτό που εκτιμάται, είναι το μέγιστο σφάλμα στην εύρεση κέντρων ματιών, σε σχέση με την πραγματική απόσταση μεταξύ των ματιών. Μια παραλλαγή αυτού του κριτηρίου χρησιμοποιήθηκε στο [30], όπου συνεκτιμάται η επιτυχία στην ανεύρεση περισσότερων από τα κέντρα των ματιών χαρακτηριστικών. Επιπλέον, ένας ακόμη παράγοντας που δυσκολεύει στη σύγκριση των αλγορίθμων είναι το γεγονός ότι ελέγχονται σε διαφορετικές βάσεις εικόνων, διαφορετικού βαθμού δυσκολίας. Για παράδειγμα, η BioID [17] αποτελείται από «φυσικές», θα έλεγε κανείς, εικόνες, με την έννοια ότι ο φωτισμός είναι πολύπλοκος, οι άνθρωποι που φωτογραφήθηκαν είχαν διάφορες εκφράσεις στο πρόσωπό τους, και δεν διατηρούσαν απόλυτα συγκεκριμένη απόσταση από την κάμερα. Ωστόσο, στις περισσότερες περιπτώσεις, οι βάσεις προέρχονται από εικόνες κεφάλι-ώμοι (head-shoulders databases), με ομοιόμορφο και απλό φόντο.

2.1.2 Χαρακτηριστικές Εργασίες

Στη δουλειά που περιγράφεται στο [16] οι συγγραφείς εντοπίζουν τις θέσεις έξι χαρακτηριστικών σημείων (μάτια, φρύδια, μύτη, στόμα) αφού έχει προηγηθεί εύρεση προσώπων στην εικόνα, με χρήση χρωματικότητας δέρματος στο χώρο ΥΙQ, και ακόλουθη χρήση γενετικού αλγορίθμου για πιο εξειδικευμένη εύρεση. Ακολούθως, χρησιμοποιείται ισοστάθμιση ιστογράμματος και μία τεχνική απομάκρυνσης θορύβου. Οι συγγραφείς κάνουν χρήση της ιδέας ότι τα χαρακτηριστικά προσώπου έχουν χαμηλότερες τιμές φωτεινότητας από το υπόλοιπο πρόσωπο και βρίσκουν ένα κατάλληλο κατώφλι για την κατάτμηση της εικόνας. Με αυτόν τον τρόπο, δημιουργούν μία δυαδική εικόνα, στην οποία, μετά από τη χρήση των κατάλληλων μορφολογικών μετασχηματισμών, κάθε τμήμα αντιστοιχίζεται και σε ένα χαρακτηριστικό, ανάλογα με τη θέση του στο πρόσωπο.

Στο σύστημα της δουλειάς [149] ανιχνεύονται τα μάτια, το στόμα και το σαγόνι. Για την εύρεση του κεφαλιού και του προσώπου οι συγγραφείς χρησιμοποιούν απλή

κατωφλίωση στους πίνακες ακμών, ενώ κάνουν χρήση προβολών στον x και στον y άξονα του πίνακα ακμών για την εύρεση των παραπάνω χαρακτηριστικών. Για το σαγόνι χρησιμοποιούν ελλείψεις, αν και, όπως αναφέρεται στο αντίστοιχο άρθρο, η εύρεσή του πολλές φορές αποτυγχάνει. Ακόμη, σε περίπτωση που η παραπάνω μεθοδολογία δώσει χαρακτηριστικά σημεία τα οποία βρίσκονται πολύ κοντά μεταξύ τους, κάνοντας αδύνατο τον διαχωρισμό τους, οι συγγραφείς κάνουν χρήση φίλτρων Gabor για την εύρεση των ματιών και, ξεκινώντας από αυτές τις θέσεις και μια σειρά ευριστικών κανόνων, εντοπίζουν και τα υπόλοιπα χαρακτηριστικά.

Στο [129], χρησιμοποιείται μία σειρά bottom hat μορφολογικών μετασχηματισμών για την εύρεση σκοτεινών περιοχών (οι οποίες, συνήθως, αντιστοιχούν σε μάτια και χείλη) σε υποψήφιας ελλειπτικές περιοχές προσώπου, και γεωμετρικοί περιορισμοί χρησιμοποιούνται για την εύρεση υποψήφιας χαρακτηριστικών προσώπου. Μετά από μια ακολουθία ευριστικών κανόνων [57, 25], οι υποψήφιας περιοχές αξιολογούνται για το αν μπορούν να αντιστοιχούν σε μάτια ή στόμα. Οι ευριστικοί αυτοί κανόνες, χρησιμοποιούν μέγιστο και ελάχιστο πιθανό μέγεθος του προσώπου, σχετικές μέγιστες και ελάχιστες διαστάσεις ματιών και στόματος, καθώς και λόγους ύψους προς πλάτος για αυτά τα χαρακτηριστικά. Με αυτόν τον τρόπο ορίζονται τριάδες ματιών-στόματος. Από αυτές, κρατιέται εκείνη η οποία βρίσκεται όσο πιο κοντά στο κέντρο της θεωρούμενης έλλειψης γίνεται. Επιπλέον, ελέγχεται αν ο άξονας που συνδέει τα κέντρα του κάθε ζεύγους ματιών διαφέρει σημαντικά σε κλίση, από τον μικρό άξονα της θεωρούμενης έλλειψης.

Στην έρευνα που δίνεται στο [52], τα χαρακτηριστικά του προσώπου βρίσκονται μετά την ανεύρεση ελλειπτικών περιοχών στις οποίες είναι πολύ σημαντική η ύπαρξη χρώματος που μοιάζει πολύ με αυτό του δέρματος. Για κάθε εικονοστοιχείο υπολογίζεται μία μεταβλητή, ανεξάρτητη της φωτεινότητας, χρησιμοποιώντας το κανάλι του κόκκινου και πράσινου χρώματος, διαφεμένων με τη φωτεινότητα. Οι τιμές αυτές, με χρήση εκπαιδευμένων μοντέλων, αντιστοιχίζουν σε κάθε εικονοστοιχείο την πιθανότητα με την οποία μπορεί να ανήκει σε πρόσωπο. Στα εντοπισμένα πρόσωπα, με βάση τιμές φωτεινότητας, βρίσκονται χαρακτηριστικά προσώπου. Κατόπιν, υπολογίζεται η συνέλιξη με την πρώτη και δεύτερη γκαουσιανή παράγωγο σε αυτές τις περιοχές και, έτσι, στο κάθε εικονοστοιχείο, αποδίδεται ένα 5-διάστατο διάνυσμα. Ένα σύνολο από εικόνες εκπαίδευσης χρησιμοποιείται για να υπολογιστούν τέτοιες γκαουσιανές. Τα αντίστοιχα διανύσματα ομαδοποιούνται με τον αλγόριθμο των K μέσων. Οι συστάδες που προκύπτουν χρησιμοποιούνται για να χαρακτηριστούν περιοχές των προσώπων σε μάτια, στόμα, μύτη ή σαγόνι. Περεταίρω γεωμετρική ανάλυση χρησιμοποιείται για να αποφασιστεί εάν μία περιοχή από εικονοστοιχεία που ανήκουν σε συγκεκριμένη συστάδα, αντιστοιχεί ή όχι, στο αντίστοιχο χαρακτηριστικό.

Στο [34], χρησιμοποιείται μία προσέγγιση πολλών επιπέδων για τον εντοπισμό 17 χαρακτηριστικών σημείων επάνω στο πρόσωπο. Τα σημεία αυτά είναι σημεία ενδιαφέροντος γύρω από τα μάτια, το στόμα, τη μύτη, τα φρύδια και το σαγόνι. Αρχικά, το πρόσωπο εντοπίζεται με χρήση του ταξινομητή [171]. Ο ίδιος ταξινομητής εκπαιδεύεται και με χαρακτηριστικά προσώπου (διαφορετικός ταξινομητής για κάθε χαρακτηριστικό). Τα αποτελέσματα βελτιώνονται με τη χρήση ενός περιορισμού σχήματος που προτείνεται. Ο περιορισμός αυτός ονομάζεται «Ενίσχυση Αποκρίσεων Χαρακτηριστικών Ανά Δύο» (Pairwise Reinforcement of Feature Responses, PRFR). Σύμφωνα με αυτή τη μέθοδο, μία ανά δύο κατανομή ορίζεται σαν την κατανομή της πραγματικής θέσης του χαρακτηριστικού i , δεδομένης της καλύτερης θέσης του χαρακτηριστικού j , όπως αυτή προέκυψε από τον αντίστοιχο ταξινομητή. Έτσι, μαθαίνονται $17^2=289$

ιστογράμματα $H_{ij}(\mathbf{x}_i - \mathbf{x}_j)$ για όλα τα πιθανά ζεύγη ανίχνευσης χαρακτηριστικού j και πραγματικών θέσεων χαρακτηριστικών i , για να εκτιμηθεί η ανά δύο κατανομή $P_{ij}(\mathbf{x}_i|\mathbf{x}_j)$. Ο υπολογισμός των ιστογραμμάτων γίνεται σε ένα σύνολο εικόνων επαλήθευσης. Μία βελτίωση της μεθόδου μπορεί να γίνει χρησιμοποιώντας μια παραλλαγή της μεθόδου AAM (Active Appearance Model)[30], με τη χρήση τεσσάρων τιμών για κάθε εικονοστοιχείο: Αυτές είναι οι κανονικοποιημένες παράγωγοι στην x και y κατεύθυνση και το κατά πόσο το εικονοστοιχείο αντιστοιχεί σε γωνία («cornerness») και σε ακμή («edgeness»).

Στο άρθρο [75], χρησιμοποιείται μία τεχνική τριών βημάτων για την εύρεση κέντρων ματιών. Αρχικά υπολογίζεται η απόσταση Hausdorff ανάμεσα στον πίνακα ακμών περιοχών της ολικής εικόνας και εκπαιδευμένο μοντέλο πίνακα ακμών για την εύρεση του προσώπου. Σε δεύτερο επίπεδο ακολουθεί η ίδια τεχνική για την εύρεση της περιοχής του πάνω μέρους του προσώπου. Τέλος, με χρήση Νευρωνικών Δικτύων, οι συγγραφείς βρίσκουν τα ακριβή κέντρα των ματιών.

Στη δουλειά που περιγράφουν οι Zhou *et al* [192], χρησιμοποιούνται Γενικευμένες Συναρτήσεις Προβολής (Generalized Projection Functions, GPF) για να βρεθούν κέντρα ματιών σε περιοχές που βρέθηκαν με βάση τη μέθοδο που προτείνεται στο [182]. Οι GPFs είναι γραμμικοί συνδυασμοί συναρτήσεων που λαμβάνουν υπόψη το μέσο όρο και την μεταβλητότητα φωτεινότητας κατά μήκος γραμμών και στηλών της εικόνας. Η χρήση τους στο συγκεκριμένο άρθρο, έγινε για να βρεθούν τα άκρα των ματιών, ενώ τα κέντρα των ματιών θεωρήθηκαν ως τα σημεία στο κέντρο των περιοχών που ορίζονται με αυτόν τον τρόπο.

Στη δουλειά που προτείνεται στο [173], το πρόσωπο αρχικά βρίσκεται με χρήση χρωματικότητας δέρματος και αντίστοιχη κατάτμηση. Ακολουθούν μορφολογικοί μετασχηματισμοί συστολής και διαστολής, και το περίγραμμα του προσώπου βρίσκεται με χρήση αλγορίθμου ταιριάσματος έλλειψης. Στην εργασία αυτή, χρησιμοποιήθηκε η φωτεινότητα, για να βρεθούν μάτια και ρώθωνες, ως οι περιοχές με τη χαμηλότερη φωτεινότητα. Ο βέλτιστος συνδυασμός των υποψήφιας περιοχών υπολογίστηκε βάσει γεωμετρικών περιορισμών. Στην εργασία αυτή, οι συγγραφείς εντοπίζουν και την περιοχή του στόματος με χρήση συναρτήσεων προβολής και μετασχηματισμούς ακμών.

Στο [44], χρησιμοποιείται μία τεχνική δύο σταδίων: Μετά την αρχική εκτίμηση της θέσης του προσώπου, δίκτυα Gabor Wavelet (GWN's) χρησιμοποιούνται για να συγκριθεί ένα πρόσωπο με ένα άλλο στη βάση εκπαίδευσης. Ένας αφφινικός μετασχηματισμός επίσης υπολογίζεται για να ληφθούν υπόψη και διαφορετικές γωνίες κλίσης του προσώπου. Με βάση τον τελικό σχηματισμό, αποφασίζεται μία αρχική εκτίμηση των θέσεων χαρακτηριστικών προσώπου. Ένα δεύτερο επίπεδο από GWN's, εκπαιδευμένο για κάθε ένα από 8 χαρακτηριστικά (γωνίες στόματος, ρώθωνες, γωνίες ματιών) χρησιμοποιείται για την εύρεση χαρακτηριστικών περιοχών στην περιοχή ενός προσώπου.

Στο [142], χρησιμοποιούνται νευρωνικά δίκτυα για την ανίχνευση μικρών περιοχών στα μάτια. Χρησιμοποιώντας πληροφορία γεωμετρικής δομής του ματιού, τα αποτελέσματα του νευρωνικού δικτύου μπορούν να υποστούν επεξεργασία για να βελτιωθεί η ανίχνευση. Επιπλέον, για να αντιμετωπιστούν ζητήματα πολύπλοκου φωτισμού, κλιμάκωσης και πολυπλοκότητας σχημάτων, αντί για τιμές των εικονοστοιχείων, χρησιμοποιήθηκαν τοπικές παράγωγοι των τιμών της φωτεινότητας.

Στην εργασία [70], σε εικόνες ακμών χρησιμοποιήθηκε ένας αλγόριθμος εύρεσης ελλειπτικών σχημάτων, ούτως ώστε να βρεθούν πρόσωπα σε εικόνες. Χρησιμο-

ποιήθηκε ποικιλία μεγεθών στην έλλειψη, όμως όχι και κλίσεων. Ωστόσο, μετά την βέλτιστη τοποθέτηση ελλείψεων, θεωρούνται και περιπτώσεις με γωνία. Αυτό γίνεται ψάχνοντας για συμμετρίες, όχι μόνο γύρω από τον κύριο άξονα της έλλειψης, αλλά και υπό γωνία εκδόσεων του. Ο έλεγχος συμμετριών γίνεται μέσα σε μια κεντρική περιοχή στην έλλειψη που σχετίστηκε καλύτερα με κάποια περιοχή μέσα στην εικόνα.

Στο [39], χρησιμοποιείται ο μετασχηματισμός Hough για την εύρεση της ίριδας του ματιού. Αφού βρεθεί το ένα από τα δύο μάτια σαν το καλύτερο αποτέλεσμα του μετασχηματισμού, ερευνάται η θέση του δεύτερου ματιού αριστερά και δεξιά από το πρώτο, και μόνο μέσα στα όρια μιας προκαθορισμένης ζώνης. Κατόπιν, ψάχνονται συμμετρικές περιοχές, για να δηλωθεί εάν ένα ζεύγος κυκλικών περιοχών ανήκει ή όχι σε μάτια. Τα πρόσωπα που χρησιμοποιήθηκαν στο άρθρο για τα πειράματα, ήταν δεδομένων διαστάσεων και, επομένως, οι παράμετροι που χρησιμοποιήθηκαν ήταν κατάλληλα προσαρμοσμένες.

Στο άρθρο [154], αναλύεται ένα σύστημα για έλεγχο της εγρήγορης οδηγών αυτοκινήτων. Για να ελεγχθεί η συγκέντρωσή τους, η εύρεση χαρακτηριστικών προσώπου πρέπει να είναι το πρώτο βήμα. Σε ολόκληρη την εικόνα, πρώτα ανιχνεύονται τα χείλη, με κριτήριο την χρωματική τους πληροφορία. Ακολούθως, χρωματική πληροφορία δέρματος χρησιμοποιήθηκε για να βρεθεί το υπόλοιπο πρόσωπο και, κενά στο επάνω μέρος του σήμαιναν μάτια. Η μέθοδος έδωσε καλά αποτελέσματα σε ακολουθίες βίντεο που δημιουργήθηκαν από τους συγγραφείς. Ωστόσο, βασίστηκε στο γεγονός ότι η κάμερα που παρακολουθεί τους οδηγούς στα οχήματά τους, είναι πάντοτε σε σταθερή θέση. Συνεπώς, είναι πολύ εύκολο να γίνουν υποθέσεις σχετικά με τα μεγέθη των χαρακτηριστικών, αφού τα προβλήματα κλιμάκωσης είναι σχεδόν ανύπαρκτα.

Στο [118], οι συγγραφείς χρησιμοποιούν μοντέλα γραμμικής παλινδρόμησης για την εύρεση της περιοχής ματιών σε ήδη εντοπισμένα πρόσωπα, ενώ, για πιο ακριβή εντοπισμό της ίριδας, χρησιμοποιούν Μηχανές Διανυσμάτων Υποστήριξης. Οι Μηχανές Διανυσμάτων Υποστήριξης χρησιμοποιούν βάρη στα εικονοστοιχεία, ώστε να δίνεται μεγαλύτερη βαρύτητα στα εικονοστοιχεία εκείνα τα οποία είναι πιο σημαντικά στην εύρεση της ίριδας. Τα διανύσματα υποστήριξης και τα βάρη, εκπαιδεύτηκαν ταυτόχρονα, ενώ τα αποτελέσματα είναι παρόμοια με αυτά των κλασικών Μηχανών Διανυσμάτων Υποστήριξης, με χαμηλότερο, όμως, υπολογιστικό κόστος.

Στο [37], οι συγγραφείς χρησιμοποιούν μία σειρά μεθόδων για να βρουν χαρακτηριστικά σημεία στο πρόσωπο, σε ήδη εντοπισμένες περιοχές του: Για την εύρεση των κέντρων των ματιών, χρησιμοποιούν τη μέθοδο ταξινόμησης που προτείνεται στο [193], και η οποία αντιμετωπίζει προβλήματα διαχωρισμού δύο κλάσεων, αντιμετωπίζοντας κάθε κλάση σαν σύνολο υποκλάσεων που βρίσκονται με τον αλγόριθμο K μέσων. Παρόμοια διαδικασία ακολουθούν και για την εύρεση γωνιών ματιών, στόματος και περιοχής της μύτης. Το περίγραμμα της μύτης βρίσκεται από την προβολή της παραγώγου της εικόνας στους δύο άξονες, ενώ οι ρώθωνες αντιστοιχούν στα σημεία χαμηλότερης φωτεινότητας. Για την εύρεση του στόματος χρησιμοποιείται παρόμοια λογική με αυτήν του περιγράμματος της μύτης, με την επιπλέον πληροφορία των καναλιών χρωματικότητας και κορεσμού. Πληροφορία στο σύστημα HSV χρησιμοποιήθηκε και για την εύρεση χρωματικότητας δέρματος και, με αφετηρία το αντίστοιχο αποτέλεσμα, οι συγγραφείς έβρισκαν τα φρύδια και το σαγόκι.

Στο [65], οι συγγραφείς εντοπίζουν χαρακτηριστικά προσώπου σε ήδη εντοπισμένες περιοχές του, με χρήση πληροφορίας προερχόμενης από διάφορα χαρακτηριστικά. Αρχικά, τα μάτια βρίσκονται με χρήση Νευρωνικών Δικτύων, εκπαιδευμένα, τόσο

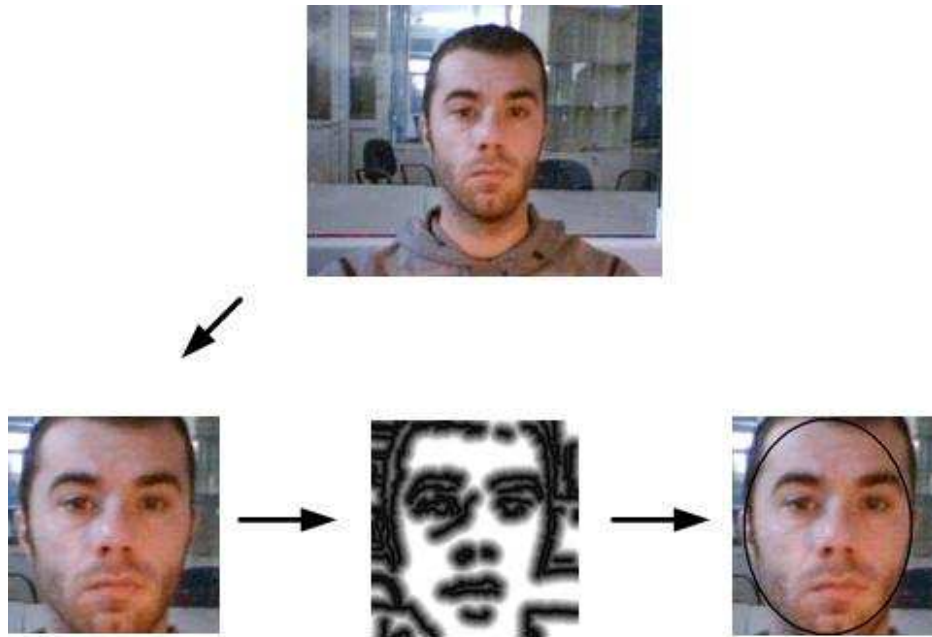
σε χρωματική, όσο και σε συχνотική πληροφορία. Ακολουθώς, πιο λεπτομερής εύρεση των ματιών γίνεται με χρήση πληροφορίας φωτεινότητας, ακμών, μορφολογικών μετασχηματισμών και μεγάλης τυπικής απόκλισης για περιοχές υψηλής πολυπλοκότητας. Παρόμοια διαδικασία ακολουθείται και για την εύρεση του στόματος, όπου γεωμετρικοί περιορισμοί χρησιμοποιούνται, ώστε να αποκλειστούν ψευδείς συναγερμοί. Για την εύρεση των φρυδιών χρησιμοποιείται πληροφορία ακμών, ενώ για την εύρεση των ρωθώνων, γίνεται χρήση πληροφορίας φωτεινότητας. Η συγχώνευση των μασκών για εκείνα τα χαρακτηριστικά στα οποία χρησιμοποιήθηκε πληροφορία από πολλά κανάλια, έγινε με χρήση τιμών εμπιστοσύνης, ώστε να δοθεί περισσότερη βαρύτητα σε εκείνες τις μάσκες που ικανοποιούν συγκεκριμένους περιορισμούς σε ό,τι αφορά τις συντεταγμένες τους.

Οι συγγραφείς στο [85] προτείνουν μία μέθοδο εύρεσης χαρακτηριστικών προσώπων με χρήση διανυσμάτων που ξεκινούν από σημεία δειγματοληψίας μέσα σε περιοχές προσώπου. Με χρήση Ιστογραμμάτων Προσανατολισμένων Ακμών, οι συγγραφείς εκπαιδεύουν τοπικούς περιγραφητές σε αυτά τα σημεία, και τους αντιστοιχίζουν διανύσματα προς συγκεκριμένα χαρακτηριστικά προσώπου. Με αυτόν τον τρόπο, δημιουργούν επεκτεταμένα πρότυπα, τα οποία, ακολουθώς, αντιστοιχίζονται με τα αντίστοιχα σημεία εικόνων εισόδου και, με βάση τα εκπαιδευμένα επεκτεταμένα πρότυπα και τα διανύσματα που τους έχουν αποδοθεί στη φάση της εκπαίδευσης, υπολογίζονται οι θέσεις χαρακτηριστικών σημείων στο πρόσωπο.

Στο κεφάλαιο αυτό προτείνεται η χρήση μιας τεχνικής ταιριάσματος προτύπων, η οποία χρησιμοποιεί εύρεση προσώπου με αυστηρή εύρεση των ορίων, σαν προεπεξεργαστικό στάδιο. Η πληροφορία που χρησιμοποιείται έχει, σε μεγάλο βαθμό, τοπολογικό χαρακτήρα, ώστε να εξασφαλίζεται ο μεγαλύτερος δυνατός βαθμός ευρωστίας σε διάφορες συνθήκες φωτισμού. Τα παραπάνω είναι καίριας σημασίας για τις αρχιτεκτονικές που μας ενδιαφέρουν στα επόμενα, καθώς, τόσο ο παράγοντας κλιμάκωσης, όσο και οι συνθήκες φωτισμού επιβάλλεται να μην αποτελούν περιοριστικούς παράγοντες στα προτεινόμενα συστήματα.

2.2 Εύρεση χαρακτηριστικών προσώπου με χρήση Διανυσματικών Πεδίων Απόστασης

Παρακάτω περιγράφεται η διαδικασία που ακολουθήθηκε για τον εντοπισμό χαρακτηριστικών σημείων στο πρόσωπο, τα οποία, στη συνέχεια του αλγορίθμου χρησιμοποιούνται, με βάση τη θέση τους, ώστε να παρακολουθούνται και να γίνεται εκτίμηση της περιστροφής του κεφαλιού, καθώς και υπολογισμός και άλλων κινήσεων στο πρόσωπο. Η διαδικασία αυτή ακολουθείται σε κάθε εκκίνηση ακολουθίας καρέ, αλλά και όποτε, όπως θα φανεί παρακάτω, το σύστημα απαιτεί επανεκκίνηση, για να μην υπάρχει συσσώρευση σφάλματος. Η μέθοδος εύρεσης χαρακτηριστικών προσώπου χρησιμοποιεί εντοπισμό προσώπου, αρχικά, έτσι ώστε να υπάρχει δυνατότητα να αντιμετωπίζονται ποικίλοι παράγοντες κλιμάκωσης. Κατόπιν, ακολουθείται μία ιεραρχική διαδικασία, στην οποία βρίσκονται περιοχές χαρακτηριστικών και, σε ένα ανώτερο επίπεδο βρίσκονται χαρακτηριστικά σημεία.



Σχήμα 2.1: Μέθοδος ανίχνευσης προσώπου: Από ολόκληρη την εικόνα, βρίσκεται η έξοδος της μεθόδου Viola-Jones, και, στην εκτιμώμενη περιοχή, ο Πίνακας Αποστάσεων περιοχής προσώπου στον οποίο εκτιμάται το αποτέλεσμα ταιριάσματος έλλειψης

2.2.1 Εύρεση περιοχής προσώπου με χαρακτηριστικά Haar και τείριασμα ελλειψοειδούς

Το πρόσωπο αρχικά εντοπίζεται χρησιμοποιώντας την μέθοδο Boosted Cascade, που περιγράφεται στο [171]. Η μέθοδος αυτή χρησιμοποιεί τον αλγόριθμο Adaboost boosting για να επιλέξει και να συνδυάσει ένα σύνολο χαρακτηριστικών τύπου Haar σε περιοχές εικόνων, ενώ ένας συνδυασμός από ταξινομητές συνδυασμένοι σειριακά (cascade) απορρίπτει περιοχές που δεν αντιστοιχούν σε πρόσωπα και διοχετεύει υπολογιστική ενέργεια σε περιοχές που είναι πιο πιθανόν να αντιστοιχούν σε πρόσωπα. Χρησιμοποιήθηκε, επιπλέον, ένα βήμα ταιριάσματος έλλειψης [8] στο πρόσωπο, με χρήση του πίνακα αποστάσεων, ώστε να βρεθούν τα αυστηρά όρια της περιοχής του προσώπου (σχήμα 2.1).

2.2.2 Εύρεση περιοχών ματιών

Η προτεινόμενη μέθοδος για εντοπισμό περιοχών ματιών συνοψίζεται ως εξής: Αρχικά, ο ανιχνευτής ακμών Canny εφαρμόζεται στην έξοδο του ανιχνευτή προσώπου. Κατόπιν, για κάθε εικονοστοιχείο, υπολογίζεται το διάνυσμα εκείνο που δείχνει προς την κοντινότερη ακμή. Έτσι, σε κάθε εικονοστοιχείο, αποδίδονται δύο τιμές, οι συντεταγμένες αυτού του διανύσματος x και y . Με αυτόν τον τρόπο, αντί για τις τιμές φωτεινότητας των εικονοστοιχείων, σε αυτήν την μέθοδο προτείνεται και χρησιμοποιείται η έννοια του Διανυσματικού Πεδίου Απόστασης, όπως προέρχεται από

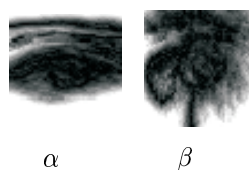
εικόνα ακμών. Το μέγεθός του είναι το ίδιο με αυτό της αρχικής εικόνας, όμως τα εικονοστοιχεία χαρακτηρίζονται από τα διανύσματα, όπως περιγράφηκαν παραπάνω. Πιο συγκεκριμένα, αυτά τα διανύσματα κωδικοποιούν για κάθε εικονοστοιχείο πληροφορία αναφορικά με την γεωμετρική του θέση σε σχέση με γειτονικές ακμές. Το διανυσματικό πεδίο μπορεί, επίσης, να παρασταθεί από δύο πίνακες (εικόνες) που αναπαριστούν τις οριζόντιες και κάθετες συντεταγμένες. Αυτοί οι πίνακες εξάγονται από τον χάρτη αποστάσεων της εικόνας, ο οποίος αποθηκεύει, για κάθε εικονοστοιχείο την απόστασή του, στον οριζόντιο και κατακόρυφο άξονα, από το κοντινότερο εικονοστοιχείο ακμής. Για τον εντοπισμό των ματιών, χρησιμοποιήθηκε ένα σύνολο πρότυπων εικονιδίων που αντιστοιχούν σε περιοχές ματιών. Ακολουθώντας, βρέθηκαν οι μέσοι πίνακες των οριζόντιων και κάθετων συντεταγμένων των πεδίων αυτών των περιοχών. Οι πίνακες αυτοί χρησιμοποιούνται, ακολούθως, για την εύρεση παρόμοιων πεδίων που αντιστοιχούν στο πρόσωπο υπό μελέτη. Η εικόνα 2.2 απεικονίζει ένα παράδειγμα περιοχής ματιού, τον αντίστοιχο χάρτη αποστάσεων, όπως επίσης και τους πίνακες των οριζόντιων και κάθετων συντεταγμένων του διανυσματικού πεδίου.



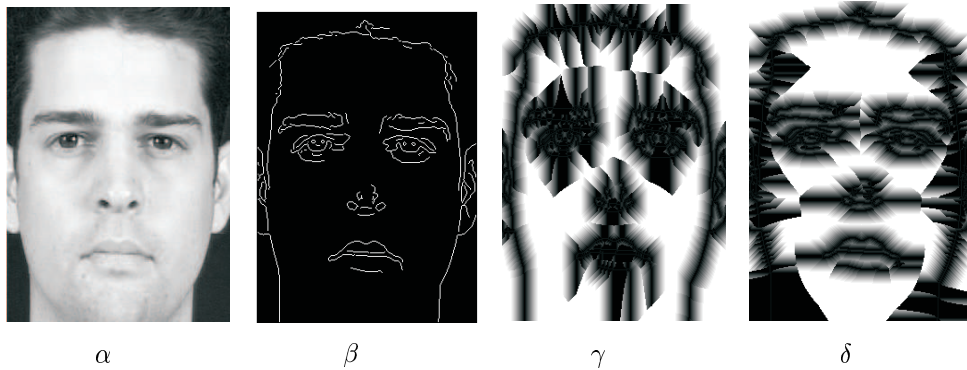
Σχήμα 2.2: Παράδειγμα εξαγωγής διανυσματικού πεδίου για περιοχή ματιού. (α) εικονίδιο ματιού, (β) χάρτης αποστάσεων, (γ) κάθετες συντεταγμένες διανυσματικού πεδίου, (δ) οριζόντιες συντεταγμένες διανυσματικού πεδίου

Στα πειράματά μας, $k=58$ περιοχές δεξιών και αριστερών ματιών χρησιμοποιήθηκαν για τον υπολογισμό των μέσων πινάκων συντεταγμένων x και y , όπως φαίνεται στο σχήμα 2.3. Όλες αυτές οι εικόνες επιλέχθηκαν έτσι ώστε να περιέχουν το φρύδι στο επάνω μέρος και τις άκρες του ματιού στο πλευρικά όριά τους. Κατόπιν, όλες οι εικόνες κλιμακώθηκαν σε ίδιες διαστάσεις, $N=26$ και $M=26$.

Όπως ήδη αναφέρθηκε, ένας ανιχνευτής προσώπου εφαρμόζεται πριν το στάδιο εντοπισμού ματιών. Η περιοχή του προσώπου κλιμακώνεται σε διαστάσεις 150×120 γραμμές επί στήλες, αντίστοιχα. Το μέγεθος αυτό βρέθηκε να είναι ικανοποιητικό στα πειράματά μας, καθώς η βασική γεωμετρική πληροφορία που χρειάζεται για τον εντοπισμό περιοχών ματιών, σε αυτές τις διαστάσεις, διατηρείται. Επόμενο βήμα είναι η εφαρμογή του ανιχνευτή ακμών Canny με υψηλό κατώφλι 50 και χαμηλό 20. Αυτές οι τιμές επιτρέπουν τον εντοπισμό μονάχα των πιο δυνατών ακμών, όπως είναι αυτές που αντιστοιχούν στα μάτια και στα φρύδια. Η εικόνα 2.4α απεικονίζει ένα πρόσωπο, όπως αυτό εντοπίστηκε από τον αλγόριθμο εύρεσης προσώπων. Οι εικόνες 2.4β, 2.4γ, 2.4δ, δείχνουν τον πίνακα ακμών Canny καθώς και τις οριζόντιες και κάθετες συντεταγμένες του διανυσματικού πεδίου, αντίστοιχως, της περιοχής που αντιστοιχεί στο πρόσωπο.



Σχήμα 2.3: (α) Μέσες κάθετες συντεταγμένες (β) Μέσες οριζόντιες συντεταγμένες



Σχήμα 2.4: (α) Περιοχή προσώπου (β) Χάρτης ακμών Canny (γ) οριζόντιες συντεταγμένες του διανυσματικού πεδίου (δ) κάθετες συντεταγμένες του διανυσματικού πεδίου

Για να βρεθούν οι περιοχές των ματιών, εξετάζονται όλες οι περιοχές μεγέθους $N \times M$ και τα αντίστοιχα διανυσματικά πεδία συγκρίνονται με το μέσο διανυσματικό πεδίο, όπως αυτό υπολογίστηκε από τα πρότυπα εικονίδια περιοχών ματιών. Η σύγκριση βασίζεται στην παρακάτω μετρική απόστασης, η οποία είναι απλοποίηση αυτής που χρησιμοποιήθηκε στο [103]:

$$E_{L_2} = \sum_{i \in R_k} \|v_i - m_i\|^2 \quad (2.1)$$

όπου $\|\cdot\|$ είναι η L_2 νόρμα. Ειδικά για μια περιοχή $N \times M$ R_k , η προηγούμενη εξίσωση δηλώνει το άθροισμα των ευκλείδειων αποστάσεων μεταξύ των διανυσμάτων u_i της εκάστοτε υποψήφιας περιοχής και των αντίστοιχων m_i του υπολογισμένου μέσου διανυσματικού πεδίου, του αντίστοιχου ματιού (δεξί ή αριστερό) που ψάχνεται. Η υποψήφια περιοχή στο πρόσωπο που ελαχιστοποιεί την ποσότητα E_{L_2} ορίζεται σαν την περιοχή του δεξιού ή αριστερού ματιού, ανάλογα με το ποιο είναι αυτό που αναζητείται. Για να γίνει γρηγορότερος ο αλγόριθμος, μπορεί να χρησιμοποιηθεί η γνώση των γενικών περιοχών των ματιών στο πρόσωπο. Έτσι, για το δεξί μάτι, η ανίχνευση γίνεται μονάχα στο πάνω δεξιά μέρος του προσώπου και για το αριστερό στο πάνω αριστερά.

2.2.3 Εύρεση περιοχής στόματος

Για την ανίχνευση της περιοχής του στόματος, ακολουθείται μία μέθοδος παρόμοια με αυτήν για την ανίχνευση των περιοχών των ματιών. Τα αποτελέσματα της εύρεσης των κέντρων των ματιών (περιγράφονται στο υποκεφάλαιο 2.3.1) μπορούν να χρησιμοποιηθούν για να οριστεί μία περιοχή μέσα στην οποία θα αναζητηθούν οι υποψήφιες περιοχές στόματος. Πειραματικά, αποδείχτηκε ότι, αν το πρόσωπο έρθει σε μέγεθος 150×120 , η περιοχή του στόματος μπορεί να αναζητηθεί σε μια ζώνη με τα ακόλουθα χαρακτηριστικά: Το επάνω όριο της είναι σε μια απόσταση d_{up} χαμηλότερα από το σημείο το ενδιάμεσο στα κέντρα των ματιών. Εδώ, d_{up} είναι 1,2 φορές η απόσταση μεταξύ των ματιών, όπως βρέθηκε στο προηγούμενο βήμα. Το κάτω όριο της ζώνης μέσα στην οποία θα αναζητηθεί η περιοχή του στόματος είναι σε απόσταση d_{low} από το προηγούμενο σημείο, όπου d_{low} είναι 1,8 φορές η απόσταση μεταξύ των κέντρων των ματιών, όπως βρέθηκαν προηγουμένως. Επιπλέον, η οριζόντια θέση του κέντρου των υποψήφιων περιοχών, δεν επιτρέπεται να είναι περισσότερο από 4 εικονοστοιχεία αριστερά ή δεξιά από την οριζόντια θέση του σημείου του ενδιάμεσου των δύο ματιών.

Για τον υπολογισμό του μέσου διανυσματικού πεδίου (δηλ. των μέσων x και y συντεταγμένων), χρησιμοποιήθηκαν 16 πρότυπα εικονίδια στόματος. Αυτά τα εικονίδια κλιμακώθηκαν σε διαστάσεις $N_m \times M_m$, με $N_m=13$ και $M_m=36$. Οι παράμετροι για την ανίχνευση ακμών με τη μέθοδο Canny, για τη δημιουργία των διανυσματικών πεδίων από τα εικονίδια αυτά, επιλέχθηκαν έτσι ώστε να υπολογίζονται ακόμα και ασθενείς ακμές. Σε αυτή τη φάση, οι περιοχές στόματος μπορούν να αναζητούνται, ακριβώς όπως γίνεται και στην περίπτωση των ματιών. Δηλαδή, να συγκριθούν τα μέσα διανυσματικά πεδία των περιοχών εκπαίδευσης, με τα αντίστοιχα των υποψήφιων περιοχών στη ζώνη που περιγράφτηκε παραπάνω. Ωστόσο, η εύρεση του στόματος είναι δυσκολότερη για πολλούς λόγους: α) Πολλές φορές το χρώμα του στόματος είναι παρόμοιο με αυτό του δέρματος, β) αντίστοιχα και η φωτεινότητα δεν διαφέρει πολύ, γ) γενειάδες σε πολλές περιπτώσεις αποκρύπτουν μέρος των χειλιών, παραμορφώνοντας, με αυτόν τον τρόπο, το σχήμα τους. Έτσι, η σύγκριση με το μέσο διανυσματικό πεδίο, που περιγράφει το μέσο σχήμα χειλιών, σε μερικές τέτοιες περιπτώσεις αποτυγχάνει. Για όλους αυτούς τους λόγους, ένας επιπρόσθετος παράγοντας εισέρχεται στην εξίσωση 2.1. Ο παράγοντας αυτός είναι το άθροισμα όλων των τιμών των εικονοστοιχείων του πίνακα των οριζόντιων ακμών της εκάστοτε υποψήφιας περιοχής. Οι οριζόντιες ακμές υπολογίστηκαν με τη χρήση μόνο των οριζόντιων μασκών Sobel, με χαμηλό κατώφλι, ώστε να επιτρέπεται ακόμα και σε ασθενείς ακμές να διαφαίνονται. Ο επιπρόσθετος αυτός παράγοντας είναι χρήσιμος εξαιτίας της οριζοντιοειδούς φύσης του σχήματος των χειλιών. Έτσι, μεγάλη συγκέντρωση οριζόντιων ακμών σε κάποια περιοχή σημαίνει πιθανή ύπαρξη χειλιών. Ο παράγοντας αυτός πολλαπλασιάζεται με ένα βάρος η ώστε η μέση τιμή του στην ζώνη υποψήφιων περιοχών να είναι ίδιας τάξης μεγέθους με την μέση τιμή της απόστασης E_{L_2} των υποψήφιων περιοχών από το μέσο διανυσματικό πεδίο:

$$E_{L_2}^{mouth} = \sum_{i \in R_k} \|v_i - m_i\|^2 + \eta * \sum_{i \in R_k} I_i^{horizontalEdges}, \quad (2.2)$$

όπου $I_i^{horizontalEdges}$ είναι η τιμή του εικονοστοιχείου i στον πίνακα ακμών, της υποψήφιας περιοχής R_k . Πιο συγκεκριμένα, η τιμή $I_i^{horizontalEdges}$ είναι μηδέν εάν το στοιχείο i ανήκει σε ακμή και, διαφορετικά, είναι ίση με τη μονάδα. Επιπλέον, η παράμετρος η υπολογίστηκε από δεδομένα εκπαίδευσης σύμφωνα με την εξίσωση 2.3,

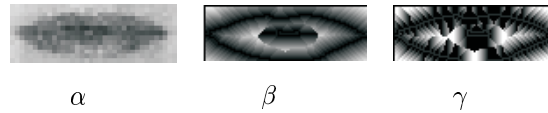
$$\eta = \frac{1}{L} \cdot \sum_{k=1}^L \left(\sum_{i \in R_k} \|v_i - m_i\|^2 \right) / \left(\sum_{i \in R_k} I_i^{horizontalEdges} \right) \quad (2.3)$$

όπου L ο αριθμός των υποψήφιων περιοχών στόματος στην ζώνη όπου αναζητείται το στόμα.

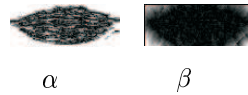
Ένα παράδειγμα περιοχής στόματος, που χρησιμοποιήθηκε για τον υπολογισμό του μέσου διανυσματικού πεδίου και των αντίστοιχων πινάκων οριζόντιων και κάθετων συντεταγμένων, φαίνεται στο σχήμα 2.5. Οι μέσοι πίνακες οριζόντιων και κάθετων συντεταγμένων φαίνονται στο σχήμα 2.6.

2.3 Εύρεση χαρακτηριστικών σημείων με πληροφορία φωτεινότητας και ακμών

Μετά τον εντοπισμό περιοχών ματιών και στόματος, τα κέντρα των ματιών και οι άκρες των χειλιών εντοπίζονται μέσα σε αυτές.



Σχήμα 2.5: (α) περιοχή στόματος (β) κάθετες συντεταγμένες του διανυσματικού πεδίου (γ) οριζόντιες συντεταγμένες του διανυσματικού πεδίου



Σχήμα 2.6: Μέσο διανυσματικό πεδίο (α) μέσες κάθετες συντεταγμένες περιοχών στόματος (β) μέσες οριζόντιες συντεταγμένες περιοχών στόματος

2.3.1 Εύρεση κέντρων ματιών

Η κάθε περιοχή ματιών που βρέθηκε σύμφωνα με τα παραπάνω, επανέρχεται στο κανονικό της μέγεθος N_{eye} , M_{eye} , το ανάλογο, δηλαδή, της αρχικής εικόνας από την οποία προήλθε. Πριν τον εντοπισμό των κέντρων των ματιών, εφαρμόζεται ένα προκαταρκτικό στάδιο. Καθώς οι αντανακλάσεις του φωτός επάνω στο μάτι επηρεάζουν αρνητικά τα αποτελέσματα και εμφανίζονται συχνά σε φωτογραφίες προσώπων, είναι απαραίτητο ένα προεπεξεργαστικό βήμα αφαίρεσής τους. Τέτοιου είδους αντανακλάσεις είναι συνήθως μικρές πολύ φωτεινές περιοχές επάνω στο μάτι. Έτσι, ο τρόπος αφαίρεσής τους είναι ο ακόλουθος: Η περιοχή του ματιού (εικόνα 2.7α) πρώτα μετατρέπεται σε δυαδική (εικόνα 2.7β) με κατωφλίωση. Η μέθοδος της κατωφλίωσης έγκειται στην επιλογή ενός κατωφλίου τέτοιου, ώστε να μεγιστοποιείται η μεταξύ των κλάσεων μεταβλητότητα [124]. Οι κλάσεις αυτές είναι τα τελικά λευκά και μαύρα μέρη της δυαδικής εικόνας.

ΕΠΙΛΟΓΗ ΒΕΛΤΙΣΤΟΥ ΚΑΤΩΦΛΙΟΥ ΓΙΑ ΔΗΜΙΟΥΡΓΙΑ ΔΥΑΔΙΚΗΣ ΕΙΚΟΝΑΣ

Σε μία ασπρόμαυρη εικόνα, εάν N είναι ο αριθμός των εικονοστοιχείων, τα επίπεδα του γκρι είναι από 1 έως L , και f_i ο αριθμός των εικονοστοιχείων φωτεινότητας i , τότε η πιθανότητα ένα εικονοστοιχείο να έχει φωτεινότητα i είναι $p_i = f_i/N$. Εάν C_1 είναι η κλάση των εικονοστοιχείων με $i < t$ και C_2 η κλάση με $i > t$ τότε, οι κατανομές του επιπέδου του γκρι για τις δύο κλάσεις είναι:

$$C_1: p_1/\omega_1(t), \dots, p_t/\omega_1(t)$$

και

$$C_2: p_{t+1}/\omega_2(t), \dots, p_L/\omega_2(t),$$

όπου

$$\omega_1 = \sum_{i=1}^t p_i \tag{2.4}$$

και

$$\omega_2 = \sum_{i=t+1}^L p_i \tag{2.5}$$

Οι μέσες φωτεινότητες των κλάσεων C_1 και C_2 είναι:

$$\mu_1 = \sum_{i=1}^t (ip_i/\omega_1(t)) \quad (2.6)$$

και

$$\mu_2 = \sum_{i=t+1}^L ip_i/\omega_2(t) \quad (2.7)$$

Έστω ότι η μέση φωτεινότητα σε όλη την εικόνα είναι μ_τ . Η μεταξύ των κλάσεων μεταβλητότητα της κατωφλιωμένης εικόνας ορίζεται ως:

$$\sigma_\beta^2 = \omega_1(\mu_1 - \mu_\tau)^2 + \omega_2(\mu_2 - \mu_\tau)^2 \quad (2.8)$$

Για διαχωρισμό μιας ασπρόμαυρης εικόνας σε δύο κλάσεις, ο Otsu απέδειξε ότι το βέλτιστο κατώφλι είναι αυτό που μεγιστοποιεί την παραπάνω τιμή:

$$t^* = \arg_{1 \leq t < L} \text{Max}(\sigma_\beta^2(t)) \quad (2.9)$$

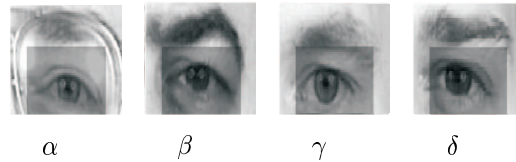
Ακολουθώντας, όλες οι μικρές λευκές περιοχές της δυαδικής εικόνας, που καταλαμβάνουν λιγότερο από το 1% της περιοχής του ματιού θεωρούνται περιοχές αντανάκλασης φωτός στο μάτι (βλ. εικόνα 2.7γ δυαδικής περιοχής με αφαιρεμένες τις μικρές αυτές περιοχές) και οι φωτεινότητές τους στην αντίστοιχη ασπρόμαυρη εικόνα αντικαθίστανται από τον μέσο όρο της γύρω περιοχής. Το αποτέλεσμα είναι η περιοχή του ματιού με τις περισσότερες αντανακλάσεις αφαιρεμένες (εικ. 2.7δ):



Σχήμα 2.7: Διαδικασία αφαίρεσης αντανακλάσεων: (α) περιοχή ματιού, (β) δυαδική περιοχή ματιού, (γ) δυαδική περιοχή ματιού με αφαιρεμένες τις περιοχές αντανακλάσεων, (δ) ασπρόμαυρη περιοχή ματιού με αφαιρεμένες τις περιοχές αντανακλάσεων

Παρατηρώντας τις εικόνες των ματιών που χρησιμοποιήθηκαν για τον υπολογισμό των μέσων διανυσματικών πεδίων, φαίνεται ότι τα μάτια βρίσκονται κάτω από το κέντρο της αντίστοιχης περιοχής. Ανάλογη τοπολογία, προφανώς, θα έχουν και οι περιοχές των ματιών, όπως αυτές εντοπίζονται σύμφωνα με τα όσα περιγράφηκαν στο αντίστοιχο υποκεφάλαιο. Το επάνω μέρος αυτών των περιοχών αντιστοιχεί στα φρύδια, ενώ τα πλευρικά μέρη τους αντιστοιχούν σε δέρμα ή, πιθανώς, μέρη του σκελετού πιθανών γυαλιών. Συνεπώς, το κέντρο του ματιού μπορεί να αναζητηθεί μέσα σε μια περιοχή που καλύπτει το κατώτερο 60% της περιοχής του ματιού και αποκλείει το δεξί και αριστερό μέρος αυτής της περιοχής (15% του δεξιού και του αριστερού μέρους της περιοχής αποκλείονται). Η πληροφορία σε αυτήν την περιοχή προέρχεται από το ίδιο το μάτι και όχι από το φρύδι ή το σκελετό των γυαλιών, όπως φαίνεται στην εικόνα (2.8).

Καθώς γύρω από το κέντρο του ματιού υπάρχει μεγάλη μεταβλητότητα όσον αφορά στις φωτεινότητες κατά μήκος του οριζόντιου και του κάθετου άξονα, υπολογίζονται

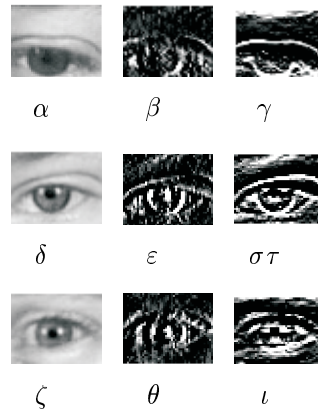


Σχήμα 2.8: Παραδείγματα περιοχών ματιών και περιοχές όπου γίνεται η αναζήτηση του κέντρου του ματιού

οι εικόνες $D_x(x, y)$ και $D_y(x, y)$ των απόλυτων διακριτών παραγώγων φωτεινότητας (Εικόνα 2.9):

$$D_x(x, y) = |I(x, y) - I(x - 1, y)| \quad (2.10)$$

$$D_y(x, y) = |I(x, y) - I(x, y - 1)| \quad (2.11)$$



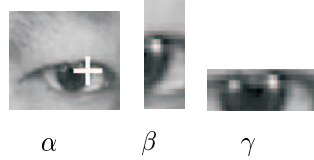
Σχήμα 2.9: Παραδείγματα περιοχών αναζήτησης κέντρου ματιού και οι εικόνες των παραγώγων κατά μήκος γραμμών και στηλών.

Ακολουθώς, τα περιεχόμενα της εικόνας οριζόντιων παραγώγων προβάλλονται στον κάθετο άξονα και τα περιεχόμενα της εικόνας καθέτων παραγώγων προβάλλονται στον οριζόντιο άξονα:

$$D_x^p(y) = \sum_x D_x(x, y) \quad (2.12)$$

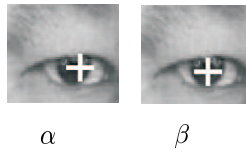
$$D_y^p(x) = \sum_y D_y(x, y) \quad (2.13)$$

Οι τέσσερις κάθετες και οι τέσσερις οριζόντιες γραμμές που αντιστοιχούν στις τέσσερις μεγαλύτερες κάθετες και οριζόντιες προβολές (δηλ. οι γραμμές που διαπερνούν τις πλέον έντονες ακμές) επιλέγονται. Η διατομή των μέσων τους ορίζει μια αρχική εκτίμηση του κέντρου του ματιού (Εικόνα 2.10). Κάνοντας χρήση του γεγονότος ότι το κέντρο του ματιού αντιστοιχεί στο κέντρο της πιο σκούρας υποπεριοχής, το προηγούμενο αποτέλεσμα μπορεί να βελτιωθεί: Βρίσκεται η σκοτεινότερη στήλη μιας περιοχής ύψους $0.4N_{eye}$ και πλάτους $0.15M_{eye}$ εικονοστοιχεία. Το ίδιο γίνεται και σε μια περιοχή $0.15N_{eye} \times 0.4M_{eye}$ (Εικόνα 2.10γ) για την εύρεση της σκορύτερης γραμμής. Οι νέες αυτές συντεταγμένες θα δώσουν μια βελτιωμένη θέση του κέντρου του ματιού (Εικόνα 2.11α). Ως σκοτεινότερη στήλη (γραμμή) σε μια εικόνα, εδώ θεωρείται εκείνη με το μικρότερο άθροισμα των στοιχείων της.



Σχήμα 2.10: (α) Αρχική εκτίμηση του κέντρου του ματιού (β), (γ) περιοχές που χρησιμοποιούνται για βελτίωση του αποτελέσματος

Για ακόμα πιο βελτιωμένα αποτελέσματα, σε μια περιοχή διαστάσεων $0.4N_{eye} \times M_{eye}$ γύρω από το σημείο που βρέθηκε από την προηγούμενη αναζήτηση, αναζητείται το σκουρότερο παράθυρο διαστάσεων $0.25N_{eye} \times 0.25M_{eye}$, ενώ το τελικό κέντρο του ματιού θεωρείται πως βρίσκεται στο κέντρο αυτού του παραθύρου. Η τελική επιλογή του κέντρου φαίνεται στην εικόνα 4.4στ .



Σχήμα 2.11: (α) Εκτίμηση θέσης κέντρου ματιού μετά από την πρώτη βελτίωση, (β) τελική θέση κέντρου ματιού

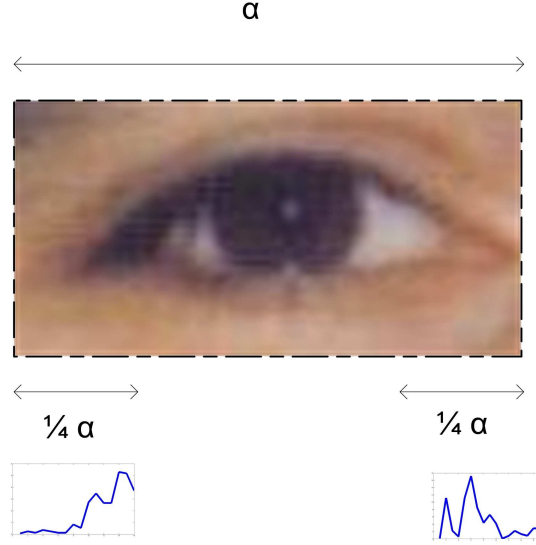
2.3.2 Εύρεση γωνιών ματιών και βλεφάρων

Για την εύρεση των γωνιών των ματιών, καθώς και των βλεφάρων (δεξιά, αριστερή γωνία και επάνω όρια ματιών), χρησιμοποιείται μία τεχνική παρόμοια με αυτήν που περιγράφεται στην εργασία των Zhou *et al* [192]. Μετά την εύρεση των κέντρων των ματιών, χρησιμοποιούνται μικρές περιοχές γύρω τους, για να βρεθούν τα υπόλοιπα χαρακτηριστικά σημεία, με χρήση Γενικευμένων Συναρτήσεων Προβολής (GPFs). Οι Γενικευμένες Συναρτήσεις Προβολής είναι γραμμικός συνδυασμός των Ολοκληρωτικών Συναρτήσεων Προβολών (Integral Projection Functions, IPF) και των Συναρτήσεων Προβολής Διασποράς (Variance Projection Functions, VPF). Γενικά, οι συναρτήσεις προβολής είναι συναρτήσεις, οι τιμές των οποίων αλλάζουν απότομα στα όρια περιοχών σε μια εικόνα. Για παράδειγμα, στην παρακάτω ανίσωση, εάν μια συγκεκριμένη τιμή είναι μεγαλύτερη από ένα κατώφλι, μπορεί να ειπωθεί ότι βρέθηκε ένα κάθετο όριο.

$$\max\left\{\frac{\partial PF_v(x)}{\partial x}\right\} > T \quad (2.14)$$

Στην παραπάνω ανίσωση, x είναι οι συντεταγμένες του οριζόντιου άξονα, και PF_v μία κάθετη συνάρτηση προβολής, ενώ T είναι το κατώφλι που χρησιμοποιείται. Οι Ολοκληρωτικές Συναρτήσεις Προβολών ορίζονται από τις εξισώσεις (2.15) και (2.16) (εδώ, οι μέσες ολοκληρωτικές συναρτήσεις προβολής), κατά μήκος διαστημάτων $[y_1, y_2]$ και $[x_1, x_2]$:

$$IPF'_v(x) = \frac{1}{y_2 - y_1} \sum_{y_1}^{y_2} I(x, y) \quad (2.15)$$



Σχήμα 2.12: Τιμές Παραγώγων Συναρτήσεων Προβολής για υποπεριοχές στις οποίες ανιχνεύονται τα κατακόρυφα όρια (άκρες) του ματιού

$$IPF'_h(y) = \frac{1}{x_2 - x_1} \sum_{x_1}^{x_2} I(x, y) \quad (2.16)$$

Οι μέσες οριζόντιες και κάθετες συναρτήσεις προβολής IPF'_h και IPF'_v αντίστοιχα, θεωρούν τη μέση φωτεινότητα σε μια εικόνα ανά στήλη και γραμμή. Εδώ, $I(x, y)$ είναι η φωτεινότητα. Παρομοίως, οι Συναρτήσεις Προβολών Διασποράς (οριζόντιες VPF_h και κάθετες VPF_v) χρησιμοποιούν την διασπορά κατά μήκος των αξόνων σε μια εικόνα:

$$VPF_v(x) = \frac{1}{y_2 - y_1} \sum_{y_1}^{y_2} [I(x, y_i) - IPF'_v(x)] \quad (2.17)$$

$$VPF_h(y) = \frac{1}{x_2 - x_1} \sum_{x_1}^{x_2} [I(x_i, y) - IPF'_h(y)] \quad (2.18)$$

Σε αυτό το άρθρο [192], προτάθηκε η GPF. Πρόκειται για γραμμικό συνδυασμό των δύο παραπάνω ειδών συναρτήσεων προβολών και ορίζεται ως εξής:

$$GPF_v(x) = (1 - \alpha) \cdot IPF'_v(x) + \alpha \cdot VPF_v(x) \quad (2.19)$$

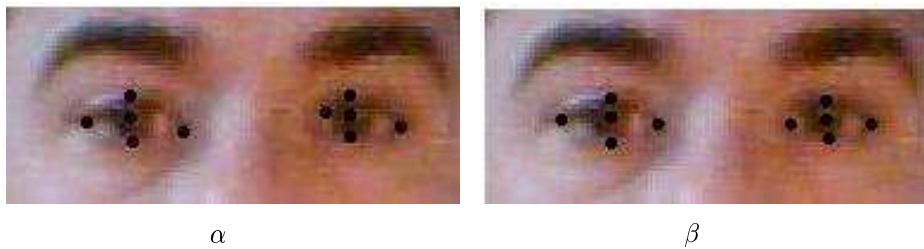
$$GPF_h(y) = (1 - \alpha) \cdot IPF'_h(y) + \alpha \cdot VPF_h(y) \quad (2.20)$$

όπου $0 \leq \alpha \leq 1$.

Στην περίπτωση της εύρεσης των πάνω και κάτω ορίων των ματιών, στην παρούσα διατριβή, χρησιμοποιήθηκαν τα κέντρα τους σαν αφετηρίες και εξήχθησαν περιοχές διαστάσεων $N_{eye}/2$ και M_{eye} γύρω από αυτά. Για την εύρεση των κάθετων ορίων (δεξιά και αριστερή γωνία του κάθε ματιού), χρησιμοποιήθηκαν υποπεριοχές τέτοιες ώστε να ελαχιστοποιηθεί ο κίνδυνος να θεωρηθούν σαν όρια του ματιού αυτά της ίριδας (σχήμα 2.12).

Στο σχήμα 2.12 φαίνεται η διαδικασία εύρεσης των κάθετων ορίων και οι υποπεριοχές που χρησιμοποιήθηκαν με τις αντίστοιχες τιμές των Γενικευμένων Συναρτήσεων Προβολής. Αντίστοιχη είναι η διαδικασία για την εύρεση των οριζόντιων ορίων με υποπεριοχές ορισμένες στο επάνω και κάτω μέρος των κέντρων των ματιών. Τα όρια εξάγονται για εκείνες τις θέσεις όπου παρατηρούνται τα μέγιστα των παραγώγων των προβολών. Όσον αφορά στην κάθετη συνιστώσα των κατακόρυφων ορίων, αυτή προέκυψε από τη θέση της χαμηλότερης τιμής φωτεινότητας στην αντίστοιχη στήλη. Ανάλογη διαδικασία ακολουθήθηκε και για την οριζόντια συνιστώσα των οριζόντιων ορίων (βλέφαρα).

Για πιο ακριβή εύρεση της αριστερής και δεξιάς γωνίας των ματιών, εδώ, ακολουθήθηκε και ένα ακόμη βήμα: Χρησιμοποιώντας τις συντεταγμένες των τεσσάρων γωνιών και των κέντρων των ματιών, βρέθηκε η πλέον κατάλληλη ευθεία $x_i = l \times y_i + c$ που περνάει πιο κοντά από όλα τα σημεία. Οι άγνωστες παράμετροι l και c βρέθηκαν με την επίλυση συστήματος διάσπασης ιδιόμορφων τιμών (Singular Value Decomposition-SVD) για τα 6 σημεία. Βρέθηκε το σημείο με τη μεγαλύτερη απόσταση από αυτήν την ευθεία και οι παράμετροι l' και c' επαναπροσδιορίζονται για τα υπόλοιπα σημεία, ενώ αυτό που αποκλείστηκε τοποθετείται επάνω στην ευθεία $x_i = l' \times y_i + c'$, σε απόσταση από την άλλη γωνία του ίδιου ματιού ίση με αυτήν που έχει μετρηθεί για το άλλο μάτι. Παράδειγμα της διαδικασίας φαίνεται στο σχήμα 2.3.2.



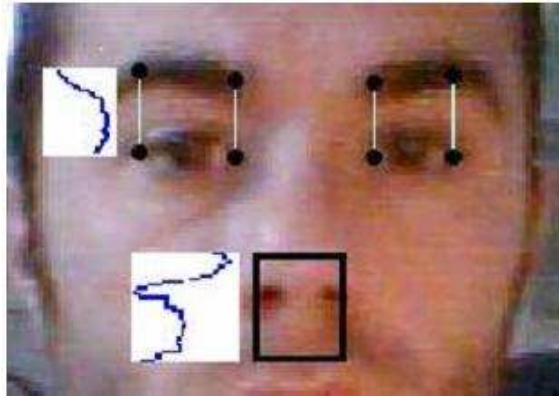
Σχήμα 2.13: Διαδικασία βελτιστοποίησης εύρεσης γωνίας ματιού με υπόθεση συγγραμμικότητας γωνιών ματιών

2.3.3 Εύρεση φρυδιών

Μετά την εύρεση των γωνιών των ματιών, δεν είναι δύσκολο να βρεθούν χαρακτηριστικά σημεία επάνω στα φρύδια, εφόσον είναι ευδιάκριτα. Καθώς δεν είναι απαραίτητη η εύρεση συγκεκριμένων σημείων επάνω στα φρύδια, παρά μόνο βοηθητικών, ώστε να υποστηρίζεται σωστή παρακολούθηση όλων των χαρακτηριστικών του προσώπου, με βάση γεωμετρικούς περιορισμούς (βλ. ενότητα 3.4), αρκεί να βρεθούν κάποια σημεία επάνω σε αυτά. Τα φρύδια είναι χαρακτηριστικά που παρουσιάζουν αρκετά μεγάλη αντίθεση σε σχέση με την περιοχή γύρω τους, καθώς και αρκετή λεπτομέρεια ώστε να είναι αρκετά χρήσιμα για σκοπούς παρακολούθησης προσώπου. Στην παρούσα διατριβή, ξεκινώντας από κάθε μία από τις τέσσερις γωνίες ματιών, τα σημεία στα φρύδια βρίσκονταν ως αυτά με τη χαμηλότερη φωτεινότητα σε ένα ευθύγραμμο τμήμα (το οποίο τέθηκε ίσο με το πλάτος του κάθε ματιού).

2.3.4 Εύρεση ρωθώνων

Παρόμοια διαδικασία με την εύρεση των φρυδιών ακολουθήθηκε και για την εύρεση των ρωθώνων. Συμβολίζοντας με (y_m, x_m) τις συντεταγμένες του σημείου στο ενδιαμέσο



Σχήμα 2.14: Περιοχές ανίχνευσης χαρακτηριστικών σημείων στα φρύδια και στη μύτη

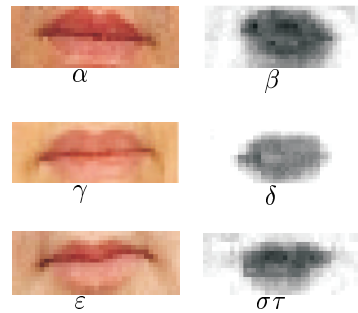
των δύο ματιών, και με d_e την απόσταση ανάμεσα στα κέντρα των ματιών, οι ρωθώνες ανιχνεύονται στο εσωτερικό της περιοχής που ορίζεται από τα όρια της εξίσωσης 2.21 (y και x είναι οι στήλες και οι γραμμές αντίστοιχα).

$$\begin{aligned}
 y_{left} &= y_m - 0.15 \times d_e \\
 x_{up} &= x_m + 0.50 \times d_e \\
 y_{right} &= y_m + 0.15 \times d_e \\
 x_{down} &= y_m + d_e
 \end{aligned}
 \tag{2.21}$$

Η γραμμή χαμηλότερης φωτεινότητας αυτής της περιοχής δίνει μια καλή εκτίμηση της κάθετης θέσης των ρωθώνων, και για παρακολούθηση χρησιμοποιείται το μέσο σημείο του αντίστοιχου ευθύγραμμου τμήματος. Στο σχήμα 2.14 φαίνονται χαρακτηριστικές περιοχές που χρησιμοποιήθηκαν για την εύρεση των φρυδιών και της μύτης.

2.3.5 Εύρεση γωνιών στόματος

Για την εύρεση των γωνιών του στόματος, χρησιμοποιείται χρωματική πληροφορία. Γένια ή μουστάκι, ή η δυσκολία να διαχωριστεί δέρμα από στόμα σε ασπρόμαυρες εικόνες κάνουν την εύρεση των γωνιών του στόματος δύσκολη διαδικασία. Ωστόσο, το στοιχείο hue (χρωματικότητα) των περιοχών του στόματος μπορεί να χρησιμοποιηθεί πολύ αποτελεσματικά για να βρεθούν οι άκρες των χειλιών. Το κανάλι χρωματικότητας στην περιοχή των χειλιών έχει πολύ χαμηλές τιμές. Η εικόνα (2.15) δείχνει αποτελέσματα εύρεσης στοματικής περιοχής και απομονωμένο το αντίστοιχο κανάλι χρωματικότητας. Είναι προφανές ότι, ακόμα και σε περιπτώσεις δύσκολου διαχωρισμού των χειλιών, ή ακόμα και στην περίπτωση όπου ο εντοπισμός της περιοχής δεν ήταν καλός (δηλ. όταν το στόμα δεν είναι κεντρικά τοποθετημένο στην περιοχή του) το κανάλι χρωματικότητας μπορεί να χρησιμοποιηθεί πολύ αποτελεσματικά στο να διαχωριστεί η περιοχή του στόματος. Για να βρεθούν οι γωνίες του στόματος, το κανάλι χρωματικότητας, εάν αντιμετωπιστεί σαν μία ασπρόμαυρη εικόνα, μπορεί να μετατραπεί σε δυαδική εικόνα με κατωφλίωση, όπως περιγράφηκε στο υποκεφάλαιο 2.3.1. Με αυτόν τον τρόπο, είναι πιθανό να επιτευχθεί καλός διαχωρισμός μεταξύ της περιοχής των χειλιών και της γύρω περιοχής. Το χρώμα των χειλιών (κόκκινο, αλλά όχι πορφυρό) έχει σχεδόν πάντα τιμή χρωματικότητας μικρότερη από αυτήν του δέρματος. Σε μερικές περιπτώσεις (γυναίκες με μακιγιάζ), το κανάλι χρωματικότητας έχει μεγάλες τιμές (μεγαλύτερες από 300°). Σε αυτές τις περιπτώσεις, η άσπρη



Σχήμα 2.15: Περιοχές στόματος και οι αντίστοιχες τιμές του καναλιού hue

περιοχή στην αντίστοιχη δυαδική εικόνα είναι αυτή που αντιστοιχεί στα χείλη και η μαύρη στο δέρμα. Σε κάθε περίπτωση, η εξωτερική περιοχή της δυαδικής εικόνας θεωρείται το δέρμα και η εσωτερική τα χείλη.

Επιπλέον, σε ορισμένες περιπτώσεις, μικρές περιοχές έχουν απόχρωση παρόμοια με αυτήν των χειλιών, ενώ δεν ανήκουν σε αυτά. Το αποτέλεσμα είναι, στην κατωφλίωση, να αποδοθούν στην ίδια κλάση με αυτήν των χειλιών. Αυτές οι περιοχές πρέπει να απορριφθούν πριν γίνει η αναζήτηση των ακρών του στόματος, καθώς μπορεί να επηρεάσουν αρνητικά το αποτέλεσμα. Για την ακρίβεια, όλες οι ενιαίες περιοχές της κλάσης των χειλιών, με μέγεθος μικρότερο του 15% της περιοχής του στόματος, αφαιρούνται. Παράδειγμα τέτοιας περίπτωσης φαίνεται στην εικόνα 2.16. Η εικόνα 2.16α δείχνει την περιοχή του στόματος όπως βρέθηκε από τον αντίστοιχο αλγόριθμο, η εικ. 2.16β δείχνει την αντίστοιχη δυαδική εικόνα του καναλιού χρωματικότητας και η εικόνα 2.16γ δείχνει την δυαδική εικόνα με τα στοιχεία που, λανθασμένα αποδόθηκαν στην κλάση του στόματος, να έχουν αφαιρεθεί.



Σχήμα 2.16: (α) Περιοχή στόματος, (β) Δυαδική εικόνα καναλιού hue με μικρές ενιαίες περιοχές, (γ) Αντίστοιχη δυαδική εικόνα χωρίς τις μικρές περιοχές (δ) περιοχή στόματος με γωνίες σημειωμένες

Η εύρεση των άκρων του στόματος γίνεται σαρώνοντας την δυαδική εικόνα του στόματος και βρίσκοντας τα πλέον ακραία σημεία της κλάσης του. Στην εικόνα 2.16δ φαίνεται το αποτέλεσμα της αναζήτησης των γωνιών του στόματος.

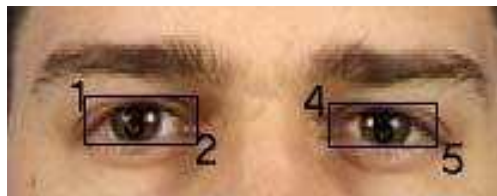
2.4 Πειραματικά αποτελέσματα

Για να αξιολογηθεί η ικανότητα της προτεινόμενης μεθοδολογίας εντοπισμού χαρακτηριστικών προσώπου, έγιναν πειράματα στις 1180 στατικές εικόνες της XM2VTS [104], όπου οι χρήστες κοιτάζουν ευθέως προς την κάμερα. Το σύνολο δεδομένων αποτελείται από 4 εικόνες για κάθε ένα από τα 295 άτομα, ενώ κάθε εικόνα απεικονίζει ένα άτομο. Οι 4 συνεδρίες για την κάθε εικόνα από το κάθε άτομο, απείχαν ένα μήνα χρονικά μεταξύ τους. Αξίζει, ακόμη, να τονιστεί ότι στις περισσότερες από το ένα τρίτο των εικόνων, τα άτομα φορούσαν γυαλιά, ενώ όλες οι εικόνες ελήφθησαν υπό ελεγχόμενες συνθήκες φωτισμού, σε ομοιόμορφο, μπλε φόντο. Πιο ειδικά, το

ενδιαφέρον στο συγκεκριμένο πείραμα, έχει να κάνει με την επιτυχία εύρεσης των κέντρων των ματιών, αλλά και των περιοχών των ματιών. Για το σκοπό αυτό, το κριτήριο που χρησιμοποιήθηκε ήταν ανάλογο με τη μετρική σφάλματος που προτείνεται στο [34]. Πιο συγκεκριμένα, σύμφωνα με αυτή τη μετρική, μετράται ο μέσος όρος των αποστάσεων όλων των χαρακτηριστικών από την πραγματική τους θέση, και αυτός κανονικοποιείται με την πραγματική απόσταση των ματιών. Το παραπάνω συνοψίζεται στην εξίσωση 2.22:

$$m_e = \frac{1}{ns} \sum_{i=1}^{i=n} d_i < T \quad (2.22)$$

όπου d_i είναι η απόσταση μεταξύ της πραγματικής θέσης του i -οστού χαρακτηριστικού σημείου και αυτού που εντοπίστηκε, s είναι η πραγματική απόσταση μεταξύ των κέντρων των ματιών, και n ο αριθμός των χαρακτηριστικών σημείων. Με άλλα λόγια, το παραπάνω κριτήριο υπονοεί ότι η ανίχνευση χαρακτηριστικών είναι επιτυχής εάν η τιμή m_e είναι χαμηλότερη από το κατώφλι T . Πιο συγκεκριμένα, εάν ο μέσος όρος των αποστάσεων πραγματικών και υπολογισμένων θέσεων χαρακτηριστικών, είναι μικρότερος από ένα συγκεκριμένο κλάσμα της απόστασης μεταξύ των κέντρων των ματιών. Στην περίπτωσή μας, τα χαρακτηριστικά είναι η επάνω δεξιά και κάτω αριστερά γωνία της περιοχής των ματιών, καθώς και τα κέντρα των ματιών, όπως φαίνεται στην εικόνα 2.17. Να σημειωθεί εδώ, ότι οι θέσεις των γωνιών ελήφθησαν από την y (οριζόντια) θέση των δεξιών/αριστερών γωνιών των ματιών και την x (κάθετη) των επάνω/κάτω γωνιών των ματιών, όπως παρέχονται από τα δεδομένα αναφοράς της βάσης.



Σχήμα 2.17: Άκρες και κέντρα περιοχών ματιών

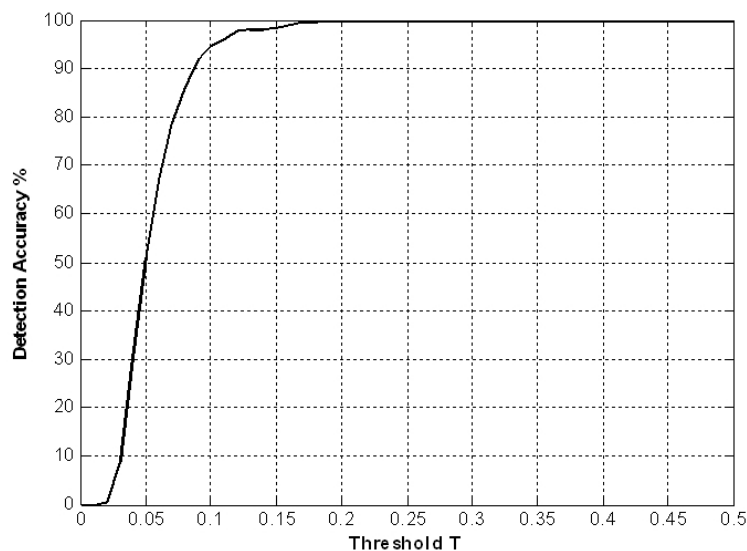
Καθώς το σύστημα πρέπει να πληρεί την προϋπόθεση της λειτουργίας σε πραγματικό χρόνο, και η μέθοδος παρακολούθησης απαιτείται να επανεκκινείται συχνά (βλ. ενότητα 3.4), είναι απαραίτητο, το βήμα του εντοπισμού χαρακτηριστικών να είναι όσο λιγότερο υπολογιστικά πολύπλοκο γίνεται. Για το σκοπό αυτό, ελέγχθηκε η ακρίβεια της μεθόδου για διάφορους παράγοντες κλιμάκωσης της εικόνας και κατώφλι (βλ. εξίσωση 2.22) $T=0.1$. Στον πίνακα 2.1 φαίνεται η επιτυχία της μεθόδου για διάφορους παράγοντες κλιμάκωσης, καθώς και ο απαιτούμενος χρόνος ανά καρέ. Οι εικόνες που χρησιμοποιήθηκαν έχουν διαστάσεις από 288×230 μέχρι 720×576 εικονοστοιχεία. Τα πειράματα έγιναν σε υπολογιστή Pentium 4, 2.80GHz.

Από τον πίνακα 2.1 μπορεί να φανεί ότι η ακρίβεια με την οποία εντοπίζονται τα χαρακτηριστικά του προσώπου είναι σταθερή με τη μεταβολή των διαστάσεων των εικόνων για τους παράγοντες κλιμάκωσης που εξετάστηκαν, ενώ οι απαιτήσεις σε χρόνο ποικίλουν σημαντικά, καθιστώντας, τόσο τον εντοπισμό, όσο και την παρακολούθηση των χαρακτηριστικών αρκετά χρονοβόρες διαδικασίες για μεγάλους παράγοντες κλιμάκωσης. Στο σχήμα 2.18 φαίνονται τα αποτελέσματα του εντοπισμού χαρακτηριστικών για διάφορα κατώφλια T , για τα 6 χαρακτηριστικά σημεία του σχήματος

Πίνακας 2.1: Ακρίβεια της μεθόδου και χρόνοι εντοπισμού χαρακτηριστικών ανά καρτέ

εικονοστοιχεία	ακρίβεια στον εντοπισμό	Μέσος απαιτούμενος χρόνος για εντοπισμό
255×230	96.16%	202ms
360×288	95.66%	239ms
432×346	95.40%	330ms
504×403	95.65%	395ms
576×461	94.67%	480ms
648×518	95.37%	587ms
720×576	94.66%	646ms

2.17.



Σχήμα 2.18: Ακρίβεια εύρεσης για 6 χαρακτηριστικά σημεία στην XM2VTS για διάφορα κατώφλια T επιτυχίας

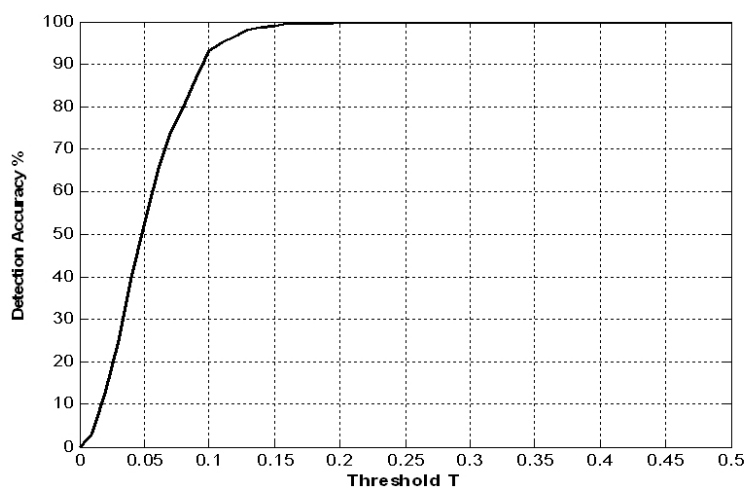
Στον πίνακα 2.2 φαίνεται η ακρίβεια έντοπισμού για το κάθε χαρακτηριστικό ξεχωριστά, για $T=0.1$ και $T=0.15$.

Επιπλέον, ελέγχθηκε η ακρίβεια στον εντοπισμό της μύτης και, πιο συγκεκριμένα, του σημείου ανάμεσα στους ρώθωνες. Στο σχήμα 2.19 φαίνονται τα ποσοστά επιτυχίας εντοπισμού της μύτης για διάφορα κατώφλια T . Για $T=0.1$, η ακρίβεια εντοπισμού είναι 93.77%.

Αντίστοιχα, για την εύρεση στόματος, οι περιοχές βρέθηκαν επιτυχώς στο 96.10% των περιπτώσεων. Αντίστοιχα, για τις γωνίες, συνολικά η επιτυχία ήταν 97.20% και 80.60% για $T=0.25$ και $T=0.1$ αντίστοιχα. Το γεγονός ότι η επιτυχία στην εύρεση των σημείων στο στόμα είναι χαμηλότερη από αυτή των ματιών οφείλεται ότι σαν παράγοντας κανονικοποίησης στην εξίσωση 2.22 χρησιμοποιήθηκε η απόσταση

Πίνακας 2.2: Ακρίβεια εντοπισμού χαρακτηριστικών σημείων στα μάτια, για κριτήρια επιτυχίας $T=0.1$ και $T=0.15$ (οι εικόνες είναι σε διάσταση 360×288)

	$T=0.1$	$T=0.15$
Επάνω γωνία -δεξί μάτι	91.97%	98.58%
κέντρο -δεξί μάτι	92.24%	98.76%
κάτω γωνία-δεξί μάτι	88.80%	98.04%
Επάνω γωνία -αριστερό μάτι	93.60%	98.84%
κέντρο -αριστερό μάτι	94.93%	98.70%
κάτω γωνία -αριστερό μάτι	87.36%	98.31%



Σχήμα 2.19: Ακρίβεια εύρεσης για σημείο ανάμεσα στους ρώθωνες στην XM2VTS για διάφορα κατώφλια T επιτυχίας

ανάμεσα στις γωνίες του στόματος, η οποία είναι μικρότερη από αυτήν ανάμεσα στα μάτια.

2.5 Συμπεράσματα

Στο παρόν κεφάλαιο μελετήθηκαν τα προβλήματα και οι απαιτήσεις των αλγορίθμων αρχικοποίησης της περιοχής του προσώπου και χαρακτηριστικών σημείων σε αυτό. Καθώς οι απαιτήσεις του συστήματος είναι τέτοιες που θα πρέπει να εξασφαλίζεται ευρωστία σε διάφορες συνθήκες φωτισμού και ανάλυσης εικόνας, η εκκίνηση του αλγορίθμου παρακολούθησης κίνησης κεφαλιού θα πρέπει να φροντίζει να τις ικανοποιεί σε όσο το δυνατόν μεγαλύτερο βαθμό. Έτσι, θεωρώντας ότι το πρόβλημα του εντοπισμού προσώπου έχει αντιμετωπιστεί πολύ αποτελεσματικά με τη μέθοδο Viola-Jones, μπορεί να χρησιμοποιηθεί για τον προσδιορισμό του. Ωστόσο, καθώς υπάρχει αρκετά μεγάλη μεταβλητότητα στα όρια της περιοχής του προσώπου, όπως αυτή βρίσκεται με τη διαδικασία Viola-Jones, συμπληρωματικά μπορεί να χρησιμοποιηθεί ένα βήμα ταιριάσματος έλλειψης, ώστε να βρεθούν αυστηρά όρια για την περιοχή του προσώπου. Με αυτόν τον τρόπο, η παραπέρα εύρεση των χαρακτηριστικών γίνεται σε συγκεκρι-

μένες περιοχές, αφού είναι γνωστά τα ακριβή όρια του προσώπου. Για την εύρεση χαρακτηριστικών σημείων, τα Διανυσματικά Πεδία Απόστασης παρέχουν ένα είδος πληροφορίας, αρκετά εύρωστο σε διάφορες συνθήκες φωτισμού, αφού κωδικοποιούν γεωμετρική πληροφορία, η οποία, μάλιστα, κωδικοποιείται καθ' όλη την έκταση των περιοχών αναζήτησης, και όχι σε συγκεκριμένα σημεία (πχ ακμές χαρακτηριστικών). Η εύρεση χαρακτηριστικών σημείων, έχοντας σαν αφετηρίες συγκεκριμένες περιοχές καθίσταται πιο επιτυχής σε σχέση με εάν γινόταν αναζήτησή τους σε μεγαλύτερες περιοχές. Τα πειραματικά αποτελέσματα των μεθόδων που χρησιμοποιήθηκαν για εύρεση χαρακτηριστικών σημείων προσώπου έδειξαν ότι μπορούν να χρησιμοποιηθούν με μεγάλο βαθμό ασφαλείας στα επόμενα βήματα της διατριβής.

Κεφάλαιο 3

Εκτίμηση περιστροφής κεφαλιού με χρήση οπτικής ροής

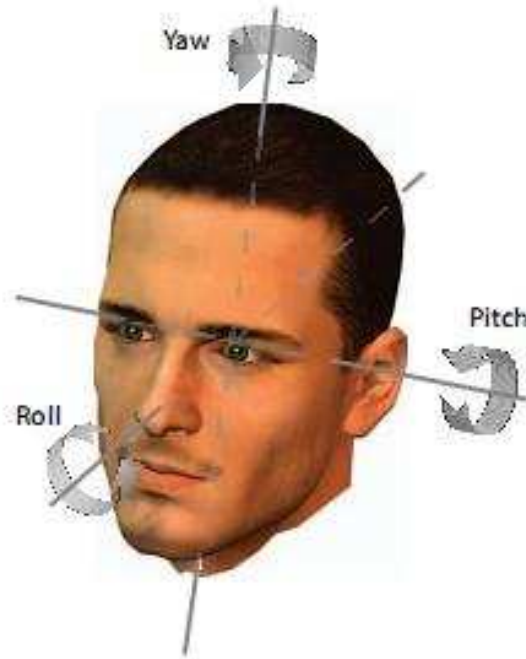
Στο κεφάλαιο αυτό μελετάται το πρόβλημα του υπολογισμού της περιστροφής του προσώπου, με χρήση 2Δ τεχνικών επεξεργασίας βίντεο. Στο πλαίσιο αυτό, προτείνεται και μελετάται ένα σύστημα που ξεκινά από μετωπική στάση και, μετά τον εντοπισμό χαρακτηριστικών σημείων, ξεκινάει η παρακολούθηση των θέσεών τους σε όλα τα υπόλοιπα καρέ της ακολουθίας. Ενσωματώνονται κανόνες τέτοιοι, που να το επανεκκινούν, όποτε αυτό κρίνεται αναγκαίο, ενώ λαμβάνεται μέριμνα ώστε οι μετρήσεις να είναι κανονικοποιημένες, ώστε να είναι ανεξάρτητες από αλλαγές κλιμακώσεων, όταν αυτές συμβαίνουν κατά τη διάρκεια ροής της ακολουθίας.

Ο όρος *εκτίμηση περιστροφής κεφαλιού*, συνήθως αναφέρεται στον υπολογισμό των τριών βαθμών ελευθερίας γύρω από τη βάση του λαιμού. Πιο συγκεκριμένα, έχει να κάνει με την οριζόντια περιστροφή (*yaw*), την κατακόρυφη (*pitch*), και την περιστροφή στο επίπεδο της κάμερας (*roll*) (βλ. εικόνα 3.1). Με τον υπολογισμό αυτών των παραμέτρων (ή, ανάλογα με την εφαρμογή, και υποσυνόλων αυτών), μπορούν να εξαχθούν σημαντικά συμπεράσματα αναφορικά με την κατευθυντικότητα του βλέμματος, ειδικά σε συνθήκες τέτοιες που τα μάτια δεν είναι αρκετά ορατά, ή η ανάλυση της εικόνας/ακολουθίας δεν είναι αρκετά υψηλή ώστε να εξαχθεί πληροφορία βλέμματος μόνο από τα μάτια.

Στην παρούσα διατριβή, ως περιστροφή κεφαλιού, θεωρούνται εκείνες οι παράμετροι περιστροφής, ως προς την κάμερα [114]. Πιο συγκεκριμένα, η περιστροφή του κεφαλιού θεωρείται μηδενική όταν το πρόσωπο φαίνεται να είναι συμμετρικό ως προς τον κατακόρυφο άξονα, και το τρίγωνο που σχηματίζεται ανάμεσα στα μάτια και το στόμα αντιστοιχεί σε επίπεδο παράλληλο με το επίπεδο της εικόνας [49]. Πρακτικά, εδώ, χρησιμοποιείται ο αλγόριθμος των Viola-Jones [171] για εντοπισμό μετωπικών προσώπων. Η γωνία περιστροφής παράλληλα με το επίπεδο του προσώπου ορίζεται αμελώντας το προοπτικό φαινόμενο (*weak perspective model*), ως η γωνία που σχηματίζεται ανάμεσα στην ευθεία των ματιών και τον οριζόντιο άξονα [59], [97].

3.1 Ερευνητικό πλαίσιο

Οι μεθοδολογίες που έχουν απασχολήσει την πρόσφατη βιβλιογραφία χρησιμοποιούν διαφορετικές προσεγγίσεις, οι οποίες ποικίλουν, είτε σε σχέση με τις ανάγκες σε υλικό, είτε τους αλγοριθμικούς περιορισμούς. Μία μεγάλη κατηγοριοποίηση των μεθοδολογιών εκτίμησης περιστροφής κεφαλιού θα είχε να κάνει με την αναγκαιότητα



Σχήμα 3.1: Οι τρεις βαθμοί ελευθερίας στην εκτίμηση περιστροφής κεφαλιού: οριζόντια περιστροφή (yaw), κατακόρυφη (pitch), και περιστροφή στο επίπεδο της κάμερας (roll)

σε εξειδικευμένο υλικό, όπως ειδικά κράνη ή μαγνητικούς αισθητήρες (παρεμβατικές μέθοδοι) και τις μεθόδους εκείνες όπου ο χρήστης δεν απαιτείται να έρθει σε επαφή με συγκεκριμένες συσκευές (μη παρεμβατικές μέθοδοι). Η παρούσα βιβλιογραφική επισκόπηση επικεντρώνεται κυρίως στη δεύτερη κατηγορία μεθόδων. Ωστόσο, κάποιες από αυτές τις μεθόδους, ακόμα και αν είναι μη παρεμβατικές, απαιτούν εξειδικευμένο υλικό ή διάταξη καμερών, πηγών υπέρυθρου φωτισμού, κλπ. Ακόμη, πολλά συστήματα χρειάζονται γνώση των εσωτερικών παραμέτρων των καμερών που χρησιμοποιούν, ή, ακόμη, διατάξεις πολλών καμερών, με σκοπό να εκμεταλλευθούν στερεοσκοπικές ιδιότητες για την εξαγωγή 3D παραμέτρων στάσης κεφαλιού. Στις περιπτώσεις εκείνες όπου δεν υπάρχει γνώση των παραμέτρων της/των κάμερας/καμερών, οι προσεγγίσεις αυτές αποτυγχάνουν να δώσουν ακριβή αποτελέσματα για κινήσεις που ξεφεύγουν από τις υποθέσεις του συστήματος (πχ κίνηση κατά μήκος του z-άξονα).

Αναφορικά με τη φύση των αλγορίθμων, κυρίως για διατάξεις της μιας μηχανής λήψης, ένας περαιτέρω διαχωρισμός θα είχε να κάνει με τις βασικές αρχές τους, το είδος των χαρακτηριστικών που χρησιμοποιούν, και τους περιορισμούς που εισάγονται από αυτά. Αν και δεν υπάρχει σαφής διάκριση σε αυτό το επίπεδο, στην παρούσα εργασία προτείνεται ένας διαχωρισμός βασισμένος σε συγκεκριμένα χαρακτηριστικά:

- **Ολιστικές μέθοδοι (μέθοδοι εμφάνισης):** Σε αυτήν την κατηγορία συναντώνται τεχνικές που, στη γενική μορφή τους, συγκρίνουν την περιοχή του προσώπου με εκπαιδευμένα μοντέλα, καθένα από τα οποία αντιστοιχεί σε μια περιστροφή. Η τελική περιστροφή του κεφαλιού, τέλος, βρίσκεται, είτε με παλινδρόμηση, είτε με ταξινόμηση. Ένα βασικό μειονέκτημα τέτοιων μεθόδων, είναι ότι απαιτούν η περιοχή του προσώπου να είναι αρκετά καλά ευθυγραμμισμένη με τα μοντέλα εκπαίδευσης, κάτι που απαιτεί πολύ καλή κατάτμηση της περιοχής του προσώπου. Συνεπώς, αν και τέτοιες τεχνικές είναι απαλλαγμένες από το πρόβλημα της ακριβούς εύρεσης χαρακτηριστικών προσώπου (μάτια,

στόμα, κλπ), καταλήγουν, ουσιαστικά στην εύρεση σημείων της περιφέρειας του προσώπου.

- **Τοπικές τεχνικές:** Σε αυτήν την κατηγορία γίνεται χρήση της θέσης συγκεκριμένων χαρακτηριστικών σημείων, και χρησιμοποιούνται οι σχετικές τους θέσεις (πληροφορία γεωμετρίας προσώπου) για τον υπολογισμό του προσανατολισμού του προσώπου. Αν και η χρήση τοπολογικής πληροφορίας είναι άμεση στην αντίληψή της, απαιτεί ακριβή εντοπισμό (και, ενδεχομένως, παρακολούθηση), χαρακτηριστικών σημείων σε μεγάλης διάρκειας ακολουθίες καρέ. Ακόμη και μικρά σφάλματα στην εκτίμηση θέσεων μπορεί να επιφέρουν μεγάλη απόκλιση από τις πραγματικές τιμές γωνιών περιστροφής. Ένα ακόμη ζήτημα που ανακύπτει με τις τοπικές τεχνικές, είναι αυτό της απόκρυψης κάποιων σημείων, οπότε και, είτε οι θέσεις τους είναι άγνωστες, ή αυθαίρετες.
- **Ανάκτηση κίνησης προσώπου:** Στην κατηγορία αυτή ανήκουν εκείνες οι μέθοδοι, στις οποίες παρακολουθείται η περιοχή του προσώπου συνολικά, και η περιστροφή του προσώπου ανακτάται από τη σύγκριση διαδοχικών καρέ. Τα αποτελέσματα από αυτές τις μεθόδους είναι, συνήθως, πολύ ακριβή, όμως απαιτείται η γνώση εσωτερικών παραμέτρων της κάμερας ή/και, της απόστασης του χρήστη από αυτήν. Επιπλέον, καθώς σε τέτοια συστήματα, υπάρχει πάντοτε ο κίνδυνος του εσωτερικού ελαχίστου σε λάθος περιοχές, θα πρέπει να υποστηρίζονται από συμπληρωματικούς αλγορίθμους, συνήθως κατάτμησης προσώπου.
- **Παραμετροποιήσιμα μοντέλα:** Η χρήση εκπαιδευμένων, παραμετροποιήσιμων μοντέλων, για αντιστοίχιση με χαρακτηριστικά σημεία του προσώπου και του περιγράμματός του, προσφέρει μεγάλη ευελιξία για αναγνώριση περιστροφής κεφαλιού σε συνθήκες έντονης μεταβλητότητας στα χαρακτηριστικά. Τέτοια μοντέλα ερευνώνται συστηματικά τα τελευταία χρόνια. Πέρα από την αναγκαιότητα για συστηματική και προσεκτική εκπαίδευσή τους, ίσως, ο κυριότερος παράγοντας που θα πρέπει να ληφθεί υπόψη στη χρήση τους, είναι η θέση αρχικοποίησής τους, καθώς υπάρχει μεγάλη πιθανότητα πτώσης σε τοπικό ελάχιστο εάν γίνει εκκίνηση από σημείο μακριά από το πρόσωπο, ή από κλίμακα διαφορετικής τάξης μεγέθους από αυτήν της περιοχής του προσώπου.

Παρακάτω γίνεται μία παράθεση και περιγραφή των κυριότερων και πιο πρόσφατων μελετών ανάκτησης περιστροφής προσώπου, για κάθε μία από τις ανωτέρω κατηγορίες.

3.1.1 Ολιστικές μέθοδοι (μέθοδοι εμφάνισης)

Οι Gourier *et al* [54] κάνουν εκτίμηση της πόζας του προσώπου στο σύνολο εικόνων Pointing 04 [53]. Για τον εντοπισμό του προσώπου, χρησιμοποιείται πληροφορία χρωματικότητας, ενώ κανονικοποιημένα εικονίδια φωτεινότητας χρησιμοποιούνται για την εκπαίδευση του συστήματος με πρότυπες πόζες, εκπαιδύοντας γραμμικές μνήμες αυτοσυσχέτισης, υπολογισμένες χρησιμοποιώντας κανόνα Widrow-Hoff. Οι συγγραφείς σύγκριναν δύο σχήματα ταξινόμησης: 1) Οριζόντιες γωνίες - με μεταβλητή κατακόρυφη, και κατακόρυφες - με μεταβλητή οριζόντια (ανεξάρτητη τεχνική) 2) Οριζόντιες και κατακόρυφες σε κοινό σχήμα εκπαίδευσης (κάθε κλάση αντιστοιχεί σε συγκεκριμένη δυάδα οριζόντιας και κατακόρυφης γωνίας) Οι συγγραφείς αναφέρουν ένα μέσο απόλυτο σφάλμα 10.3° και 15.9° για οριζόντια και κάθετη γωνία, αντίστοιχα, χρησιμοποιώντας την τεχνική (1) και 10.1° και 16.8° για την τεχνική (2). Για περεταίρω

έλεγχο της αποδοτικότητας της μεθόδου, οι συγγραφείς σύγκριναν τα αποτελέσματα, με αυτά που προέρχονται από την ανθρώπινη ικανότητα εκτίμησης της πόζας του κεφαλιού. Για το σκοπό αυτό, ζήτησαν από δύο γκρουπ ατόμων να εκτιμήσουν γωνίες περιστροφής κεφαλιού στο ίδιο σύνολο εικόνων. Το πρώτο γκρουπ είχε δει προηγουμένως (γκρουπ C) εικόνες με προσημειωμένες πραγματικές γωνίες περιστροφής, ενώ το δεύτερο (γκρουπ NC) δεν είχε τέτοια γνώση. Οι εκτιμήσεις των δύο γκρουπ όσον αφορά στην οριζόντια περιστροφή δεν διέφεραν σημαντικά, όμως το γκρουπ C είχε καλύτερη ακρίβεια στην εκτίμηση της κατακόρυφης γωνίας. Εάν θέλει να συγκρίνει κανείς την ανθρώπινη ικανότητα με αυτήν της προτεινόμενης αρχιτεκτονικής, το σύστημα έχει καλύτερη απόδοση- οι άνθρωποι συνολικά πέτυχαν εκτίμηση 11.85° κατά μέσο όρο. Όμως το σύστημα δεν ήταν τόσο ικανοποιητικό για κατακόρυφες γωνίες, αφού οι άνθρωποι είχαν ακρίβεια 10.9°

Στην εργασία που περιγράφεται από τους Murphy *et al* [113] προτείνεται ένα σύστημα για τον υπολογισμό της πόζας του κεφαλιού για οδηγούς αυτοκινήτων. Χρησιμοποιούν μία εξειδικευμένη κάμερα, ευαίσθητη σε φωτισμό με συχνότητα κυματιδίων κοντά στην υπέρυθη, καθώς και τον αντίστοιχο φωτισμό. Αρχικά το πρόσωπο εντοπίζεται με χρήση τριών ταξινομητών Adaboost Cascades, εκπαιδευμένοι να εντοπίζουν πρόσωπα μη περιστραμμένα, αλλά και αριστερό και δεξί προφίλ. Τα δυνατά εύρη τιμών στα οποία μπορεί να γίνεται εντοπισμός, είναι $[-30^\circ, 20^\circ]$ για κάθετη περιστροφή, και $[-80^\circ, 80^\circ]$ για οριζόντια. Παλινδρομητές υποστήριξης διανυσμάτων (Support Vector Regressors) χρησιμοποιούνται ακολούθως, εκπαιδευμένοι σε οριζόντια και κάθετη περιστροφή. Τα χαρακτηριστικά που χρησιμοποιούνται για τους παλινδρομητές είναι ιστογράμματα προσανατολισμού παραγώγου (gradient orientation histograms) [96]. Τα ιστογράμματα είναι τοπικά και τριών διαστάσεων: Θέση (x,y) και προσανατολισμός παραγώγου και αποτελούν περίπτωση των περιγραφητών SIFT. Το σύστημα έχει ελεγχθεί μέσα σε πραγματικές συνθήκες αυτοκινήτου (ένα VW Passat), αλλά και σε συνθήκες εργαστηρίου, όπου το σφάλμα σε κάθετη και οριζόντια κίνηση, αντίστοιχα, ήταν 5.58° και 6.40° αντίστοιχα.

Στη δουλειά που περιγράφεται στο [88], οι συγγραφείς χρησιμοποιούν Μηχανές Διανυσμάτων Υποστήριξης για τον εντοπισμό του προσώπου. Με βάση την παρατήρηση ότι οι πόζες του προσώπου αντιστοιχούν σε ένα συνεχές πολύπτυχο μέσα στη σφαίρα ενός ιδιοχώρου, οι πόζες διακρίνονται μεταξύ τους από το αντίστοιχο ιδιοδιάνυσμα μέσα στη σφαίρα. Έτσι, ο αλγόριθμος, με τη χρήση ενός ταξινομητή του κοντινότερου γείτονα, αποδίδει σε κάθε εντοπιζόμενο πρόσωπο και την αντίστοιχη πόζα. Η μέθοδος επαληθεύτηκε σε τρία πρόσωπα, και αναφέρονται αποτελέσματα για οριζόντια περιστροφή, και περιστροφή παράλληλη στο επίπεδο της κάμερας.

Στο [156], οι συγγραφείς χρησιμοποιούν νευρωνικά δίκτυα για την εκτίμηση του προσανατολισμού του προσώπου. Ωστόσο, ένα από τα μειονεκτήματα της συγκεκριμένης δουλειάς, είναι ότι η βάση εκπαίδευσης και επαλήθευσης είναι η ίδια. Κάτι τέτοιο, προφανώς, βελτίωσε αρκετά τα αποτελέσματα, αφού οι συνθήκες κατάτμησης των περιοχών των προσώπων ήταν παρόμοιες για όλους τους χρήστες και, συνεπώς, η είσοδος στο νευρωνικό δίκτυο ήταν κατάλληλα ευθυγραμμισμένη.

Στο [172], στήθηκε ένα σύστημα τεσσάρων καμερών μέσα σε χώρο συγκεκριμένων διαστάσεων. Η έξοδος της κάθε κάμερας χρησιμοποιείται για να τροφοδοτήσει ένα νευρωνικό δίκτυο, ώστε να υπολογιστεί η οριζόντια γωνία περιστροφής όταν κάποιος στέκεται στο κέντρο του χώρου. Με τη χρήση πίνακα σύγκυσης του κάθε δικτύου, για κάθε εκτιμώμενη πόζα, εκτιμάται η εκ των υστέρων πιθανότητα της κάθε πόζας. Η πόζα εκείνη, η οποία θα δώσει και το μεγαλύτερο άθροισμα όλων των εκ των υστέρων

πιθανοτήτων, είναι και η τελική έξοδος του συστήματος.

3.1.2 Τοπικές τεχνικές

Στο [50], ο λόγος της απόστασης μεταξύ των εξωτερικών γωνιών των ματιών και της απόστασης του στόματος από το σημείο ανάμεσα στα μάτια χρησιμοποιείται για να μοντελοποιήσει το πρόσωπο, έτσι ώστε να βρεθεί ο προσανατολισμός του. Για το σκοπό αυτό, χρησιμοποιείται αφφινικός μετασχημμός και, για κάθε πόζα, το σύστημα δίνει δύο αντιδιαμετρικές λύσεις, εκ των οποίων, η σωστή επιλέγεται χρησιμοποιώντας τις θέσεις των φρυδιών: Ανάλογα με την κάθε πιθανή περιστροφή, γίνεται σχετική υπόθεση της αναμενόμενης θέσης των φρυδιών. Εάν ικανοποιούνται συγκεκριμένες ιδιότητες σε ό,τι αφορά στις τιμές των εικονοστοιχείων, τότε επιλέγεται η μία ή η άλλη κατεύθυνση περιστροφής.

Στο [51], χρησιμοποιείται η σχετική θέση των χαρακτηριστικών προσώπου σε σχέση με το υπόλοιπο πρόσωπο. Για το σκοπό αυτό, το πρόσωπο μοντελοποιείται ως ελλειψοειδές συγκεκριμένης χρωματικότητας, και οι θέσεις των ματιών χρησιμοποιούνται για να βρεθεί η οριζόντια γωνία του προσώπου.

Στο [97] οι συγγραφείς μελετάνε τις αλλαγές στη συμμετρία του προσώπου ως προς τον κάθετο άξονα, έτσι ώστε να βγάλουν συμπεράσματα σχετικά με την οριζόντια κίνηση του κεφαλιού. Για το σκοπό αυτό, μέσα στην περιοχή ενδιαφέροντος του προσώπου, ακολουθούν μία λογική κατά γραμμές. Πιο συγκεκριμένα, κάθε γραμμή της εικόνας του προσώπου, τη θεωρούν ένα μονοδιάστατο σήμα, το οποίο και συνελίσσεται με φίλτρα Gabor 5 διαφορετικών συχνοτήτων, με σκοπό να μειωθεί ο θόρυβος και να εξαχθούν τοπικά χαρακτηριστικά. Ακολουθώς, εφαρμόζεται μετασχηματισμός Fourier. Η παρούσα δουλειά αναλύει τα φανταστικά και πραγματικά μέρη του μετασχηματισμού, με σκοπό να διαφανούν συμμετρίες μέσα από αυτά. Χρησιμοποιήθηκε Γραμμική Διακριτή Ανάλυση (LDA) για εκπαίδευση, ενώ η κατηγοριοποίηση έγινε με τη μέθοδο του κοντινότερου γείτονα. Η μέθοδος εφαρμόστηκε στα δεδομένα CAS-PEAL [47], τα οποία αποτελούνται από διακριτές πόζες σε διαστήματα των 15° , καθώς και σε δεδομένα συνεχούς χρόνου (βίντεο) τα οποία δημιουργήθηκαν για τις ανάγκες των συγκεκριμένων πειραμάτων. Σε ό,τι αφορά στο πρώτο σύνολο δεδομένων, η ακρίβεια ήταν 91.42%, ενώ για το δεύτερο, το μέσο τετραγωνικό σφάλμα υπολογίστηκε στις 4.4° . Οι συγγραφείς, ακόμη, αναφέρουν ποιοτικά αποτελέσματα και σε διαφορετικές βάσεις.

Στο [99] χρησιμοποιείται ένα σύστημα δύο καμερών, το οποίο δέχεται σαν είσοδο λήψεις του προσώπου από δύο διαφορετικές γωνίες. Θεωρώντας κάμερες οπής, καθώς και ορθογραφική προβολή, ο σκοπός είναι να υπολογιστούν οι πίνακες περιστροφής των δύο καμερών. Οι πίνακες περιστροφής βρίσκονται με τη βοήθεια τριών σημείων στο πρόσωπο, για τα οποία γίνεται αντιστοίχιση και στις δύο κάμερες. Βασικό μειονέκτημα της μεθόδου είναι ότι πρέπει να βρίσκονται σημεία τέτοια, που να είναι ορατά και από τις δύο κάμερες.

Στο [117], χρησιμοποιείται πρόβλεψη με γραμμική παρεμβολή για να βρεθούν αρχικά περιοχές μέσα στις οποίες θα βρίσκονται τα μάτια, σε ήδη εντοπισμένες περιοχές προσώπου, με χρήση της μεθόδου Viola [171]. Κατόπιν, χρησιμοποιούνται μηχανές διανυσμάτων υποστήριξης για να βρεθεί η θέση του κέντρου της ίριδας. Στη δουλειά αυτή, οι συγγραφείς χρησιμοποιούν τη μάθηση των παραμέτρων των SVMs και των βαρών των χαρακτηριστικών ταυτοχρόνως, σε ένα κοινό πλαίσιο βελτιστοποίησης, με σκοπό να πετύχουν παρόμοια αποτελέσματα με μηχανές υποστήριξης που δεν χρη-

σιμοποιούν βάρη στα χαρακτηριστικά, αλλά με υψηλότερο υπολογιστικό κόστος. Οι συγγραφείς αναφέρουν αποτελέσματα στη βάση CMU [161], όπου απλά προσπαθούν να διακρίνουν μεταξύ προσώπων που κοιτάζουν ευθεία και προσώπων στραμμένων προς τα επάνω. Αναφέρουν, δε, ένα συνολικό ποσοστό ακριβείας 95.5% στην ταξινόμηση αυτή.

Στο [120], οι συγγραφείς εκμεταλλεύονται τη λύση του προβλήματος της «Προοπτικής n σημείων» για την εύρεση της περιστροφής του κεφαλιού. Πιο συγκεκριμένα, βασισμένοι σε παρακολούθηση σημείων, αναπαράγουν το τρισδιάστατο μοντέλο του προσώπου σε κάθε καρέ, και αυτό συγκρίνεται με το τρισδιάστατο μοντέλο, όπως αυτό εξήχθη στην αρχή της ακολουθίας, πριν ξεκινήσει η παρακολούθηση σημείων. Με βάση αυτή τη σύγκριση, και θεωρώντας γνωστές τις παραμέτρους της κάμερας, οι συγγραφείς υπολογίζουν πραγματικές αποστάσεις μεταξύ σημείων για την τρέχουσα και την αρχική πόζα. Η μέθοδος ελέγχθηκε σε δεδομένα από ζώα (κουκουβάγια και χαμελέων), με χρήση σημείων που τοποθετήθηκαν χειροκίνητα στα καρέ, αλλά και σε ανθρώπους, με σημεία των οποίων οι συντεταγμένες βρίσκονταν με χρήση περιγραφητών SIFT. Οι συγγραφείς αναφέρουν καλύτερα αποτελέσματα για την πρώτη περίπτωση. Για την εξαγωγή των αποτελεσμάτων, χρησιμοποιήθηκαν δεδομένα αναφοράς, τα οποία υπολογίστηκαν με τη βοήθεια μαγνητικών αισθητήρων. Μετατοπίσεις παράλληλες και κάθετες ως προς το επίπεδο της κάμερας, λήφθηκαν, επίσης, υπόψη.

Στο [121], γίνεται εύρεση 10 σημείων στο πρόσωπο, για τον υπολογισμό της περιστροφής του κεφαλιού, αλλά και της κατευθυντικότητας του βλέμματος. Ο απώτερος στόχος της δουλειάς αυτής, είναι να ανευρίσκονται οι χρονικές περίοδοι εκείνες, κατά τις οποίες, ένας ομιλητής ψεύδεται. Για την παρακολούθηση της περιοχής του προσώπου, οι συγγραφείς χρησιμοποιούν την τεχνική που περιγράφεται στο [80], όπου τα μάτια αρχικά εντοπίζονται με βάση την περιοδικότητα του ανοιγοκλεισίματός τους. Οι θέσεις των ματιών είναι η αφετηρία με βάση την οποία βρίσκονται περεταίρω σημεία: Οι άκρες των ματιών βρίσκονται από τον πίνακα ακμών της περιοχής του ματιού, το κέντρο της ίριδας από τις προβολές της φωτεινότητας της περιοχής του ματιού, ενώ οι ρώθωνες και οι άκρες του στόματος εντοπίζονται ως οι περιοχές με τη χαμηλότερη φωτεινότητα σε συγκεκριμένες περιοχές. Ακολούθως, μοντέλα «εικονικών ελατηρίων» (virtual springs) χρησιμοποιούνται για την εξαγωγή της πόζας του κεφαλιού: Μοντέλα προσώπου περιστρέφονται με τέτοιο τρόπο, ώστε η ελαστική ενέργεια των ελατηρίων προς την κατεύθυνση των θέσεων των χαρακτηριστικών προσώπου να μειώνεται.

Στο [119] γίνεται χρήση της συμμετρίας του προσώπου για την εκτίμηση της περιστροφής του κεφαλιού γύρω από τον κάθετο άξονα. Για την εκτίμηση της γωνίας του προσώπου, εκτιμάται η γωνία του τριγώνου που σχηματίζεται από την προβολή του τριγώνου μάτια-στόμα στο επίπεδο της εικόνας με αυτό του αντίστοιχου ισοσκελούς, με μία κοινή πλευρά στην προβολή. Για το σκοπό αυτό ανιχνεύονται μάτια και στόμα, με βάση την απομόνωση ελλείψεων στο χρώμα του δέρματος, και ανίχνευση χαρακτηριστικών προσώπου με βάση τη φωτεινότητά τους και τη φυσιολογία του ανθρώπινου προσώπου. Η μέθοδος δοκιμάστηκε στη βάση M2VTS για την ανίχνευση των χαρακτηριστικών προσώπου, ενώ οι συγγραφείς υποστηρίζουν πως, επιτυχής εύρεση των χαρακτηριστικών συνεπάγεται και επιτυχής εύρεση της οριζόντιας γωνίας περιστροφής. Παρόλα αυτά, δεν δίνονται συγκεκριμένα αποτελέσματα σε ό,τι αφορά στην επιτυχία του αλγορίθμου να εντοπίσει χαρακτηριστικά σημεία προσώπου σε περιπτώσεις αρκετά μεγάλων περιστροφών σε σχέση με την ευθεία στάση.

Στο [176], οι συγγραφείς προτείνουν μία μεθοδολογία για την παρακολούθηση της

περιστροφής του κεφαλιού. Το κεφάλι ανιχνεύεται, ως το μοναδικό κινούμενο αντικείμενο στην ακολουθία καρέ (χρησιμοποιείται απλή αφαίρεση φωτεινότητων), ενώ γίνεται κατάτμηση του προσώπου στον χρωματικό υποχώρο «κανονικοποιημένο πράσινο» όπου, σύμφωνα με τους συγγραφείς, μπορεί εύκολα να εξαχθεί το πρόσωπο. Επιπλέον, με τη χρήση μιας σειράς φίλτρων (Sobel, Gabor, std), εξάγεται και η θέση του στόματος. Για την εκτίμηση της θέσης του κεφαλιού, χρησιμοποιείται ο αλγόριθμος Camshift [4], ενώ η θέση του στόματος παρακολουθείται με χρήση της μεθόδου Lucas-Kanade. Η πόζα του κεφαλιού, τέλος, υπολογίζεται από την απόκλιση της θέσης του στόματος, σε σχέση με το κέντρο της έλλειψης, καθώς και την αρχική του θέση, όταν ο χρήστης ήταν στραμμένος προς την κάμερα. Η γωνία περιστροφής παράλληλα με το επίπεδο της κάμερας, υπολογίζεται από τον μεγάλο άξονα της έλλειψης. Μεταξύ των μειονεκτημάτων της μεθόδου, είναι το ότι ο αλγόριθμος Camshift είναι αρκετά ευαίσθητος σε αλλαγές των συνθηκών φωτισμού. Στο άρθρο δεν αναφέρονται πειραματικά αποτελέσματα, παρά μόνο ότι το σύστημα μπορεί να λειτουργήσει σε πραγματικό χρόνο.

Στο [188], χρησιμοποιούνται προβολές κύκλων γύρω από τις τρισδιάστατες θέσεις χαρακτηριστικών προσώπου. Το αποτέλεσμα είναι ελλείψεις στο επίπεδο της εικόνας, ενώ η πόζα του κεφαλιού εξάγεται από τον μετασχηματισμό του αντίστοιχου κύκλου, ώστε να προκύψει η έλλειψη. Μειονέκτημα της μεθόδου, είναι ότι απαιτείται πρότερη γνώση των εσωτερικών παραμέτρων της κάμερας, προκειμένου να βρεθεί ο αφρινικός μετασχηματισμός που δίνει την έλλειψη.

Στο [189], θεωρείται ορθογραφική προβολή για την εκτίμηση πόζας κεφαλιού, βασισμένη σε χαρακτηριστικά προσώπου, ενώ χρησιμοποιείται ένα γραμμικό σύστημα εξισώσεων βασισμένο στις συντεταγμένες χαρακτηριστικών προσώπου στην εμπρόσθια και την περιστραμμένη έκφραση του προσώπου. Στην παραπάνω διαδικασία, δεν λαμβάνεται υπόψη πληροφορία βάθους, και ο προσανατολισμός του επιπέδου που ορίζεται από το πρόσωπο υπολογίζεται με βάση τις σχετικές αποστάσεις των χαρακτηριστικών προσώπου.

3.1.3 Ανάκτηση κίνησης προσώπου

Ο Cascia *et al.* [22] μοντελοποιούν το κεφάλι με κυλινδρικά μοντέλα για να εξάγουν παραμέτρους κίνησης. Πιο συγκεκριμένα, σε αυτή τη δουλειά, το πρόσωπο αρχικά εντοπίζεται χρησιμοποιώντας έναν ανιχνευτή προσώπου, και ακολούθως αντιστοιχίζεται σε μια κυλινδρική επιφάνεια, θεωρώντας τις παραμέτρους της κάμερας γνωστές. Ακόλουθα καρέ «παραμορφώνονται» ώστε να αντιστοιχίζονται με το βέλτιστο τρόπο στην υφή αναφοράς, όπως υπολογίζεται στην αρχή του αλγορίθμου. Με αυτόν τον τρόπο υπολογίζονται οι τρισδιάστατες παράμετροι κίνησης του κεφαλιού (ή, με άλλα λόγια, εξάγονται οι παράμετροι κίνησης του κυλίνδρου). Εξαιτίας της προοπτικής απεικόνισης, δεν έχουν όλα τα εικονοστοιχεία την ίδια τιμή εμπιστοσύνης κατά τη διάρκεια της αντιστοίχισης. Αυτό λαμβάνεται υπόψη εφαρμόζοντας ένα «μοντέλο εμπιστοσύνης» στα τρίγωνα του κυλίνδρου, χρησιμοποιώντας το σχήμα τους, όπως αυτό εμφανίζεται στην τρέχουσα υφή σε σχέση με την εικόνα εισόδου. Η μέθοδος λαμβάνει υπόψη και αλλαγές στο φωτισμό, μοντελοποιώντας το σφάλμα μεταξύ της τρέχουσας εικόνας σε σχέση με την υφή αναφοράς [56].

Στο [72], οι συγγραφείς προτείνουν χρήση κυλινδρικών μοντέλων με εύρεση θέσεων χαρακτηριστικών σημείων. Μετά τον αρχικό εντοπισμό της θέσης του προσώπου, η περιστροφή του υπολογίζεται με τη χρήση της μεθόδου ευθυγράμμισης προ-

σώπου Μπεϋζιανού Εφαπτομενικού Μοντέλου Σχήματος (Bayesian Tangent Shape Model - BTSM) [191]. Κατόπιν, χρησιμοποιείται ο αλγόριθμος SIFT [96] για την εξαγωγή και αντιστοίχιση σημείων σε διαδοχικά καρέ, καθώς και κανονικοποιημένη συσχέτιση. Τα χαρακτηριστικά αυτά χρησιμοποιούνται για να ενημερώσουν μια βάση η οποία δημιουργείται σε πραγματικό χρόνο, ενώ συγκρίνουν τις περιοχές γύρω από αυτά με εκείνες οι οποίες έχουν ήδη αποθηκευθεί σε προηγούμενα καρέ. Τέλος, χρησιμοποιούνται φίλτρα Kalman για να υπολογιστεί η διαφορά σε περιστροφή/μετατόπιση μεταξύ διαδοχικών καρέ, αλλά και μεταξύ του τρέχοντος καρέ και του πλέον όμοιού του, από την βάση η οποία ενημερώνεται κατά τη διάρκεια της ακολουθίας. Για την επαλήθευση της ευρωστίας της μεθόδου χρησιμοποιήθηκε η βάση του Πανεπιστημίου της Βοστώνης, ενώ για κάθετη, οριζόντια, και παράλληλη στο επίπεδο της κάμερας περιστροφή, τα αποτελέσματα ήταν 3.7° , 4.6° και 2.1° αντιστοίχως.

3.1.4 Παραμετροποιήσιμα μοντέλα

Οι συγγραφείς στο [31], χρησιμοποιούν Μοντέλα Ενεργού Εμφάνισης [30] για την εκτίμηση της στρέψης του κεφαλιού γύρω από τον κάθετο άξονα (οι συγγραφείς αναφέρουν ότι η εκτίμηση της γωνίας γύρω από τον οριζόντιο άξονα είναι παρόμοιο πρόβλημα). Τα Μοντέλα Ενεργού Εμφάνισης είναι πρακτικά ένα δίκτυο κόμβων και συνδέσμων, που μπορούν να μοντελοποιήσουν το σχήμα και την υφή ενός συνόλου εκπαίδευσης και, με αυτόν τον τρόπο, χρησιμοποιούνται για παρακολούθηση προσώπου, βρίσκοντας μία σειρά παραμέτρων οι οποίες «παραμορφώνουν» το δίκτυο ώστε να κάνει την καλύτερη δυνατή αντιστοίχιση με μια καινούρια εικόνα. Οι συγγραφείς χρησιμοποίησαν μοντέλο παλινδρόμησης μεταξύ γνωστών γωνιών και παραμέτρων των μοντέλων, όπως εξήχθησαν από δεδομένα εκπαίδευσης για την εύρεση άγνωστων γωνιών.

Στο [32], οι συγγραφείς χρησιμοποιούν ένα επεκτεταμένο φίλτρο Kalman, καθώς και παραμέτρους κίνησης μυϊκών ομάδων του προσώπου, με τη χρήση 3-D AMB AAMs [33], με σκοπό την ταυτόχρονη εύρεση της περιστροφής του κεφαλιού, αλλά και της έκφρασης του προσώπου. Τα 3-D AMB AAMs είναι παρόμοια με τα τρισδιάστατα Μοντέλα Ενεργού Εμφάνισης (Active Appearance Models, AAMs) με τη διαφορά ότι υιοθετούν περιορισμούς που έχουν να κάνουν με αναμενόμενες κινήσεις του κεφαλιού. Πιο συγκεκριμένα, χρησιμοποιούν Μονάδες Δράσης (Action Units), οι οποίες παίρνουν τιμές στο διάστημα $[-1,1]$, και ενθαρρύνουν τιμές και συνδυασμούς τιμών τέτοιους, που να συνάδουν με το αναμενόμενο. Με τη χρήση της εξόδου από τα μοντέλα 3-D AMB AAMs, οι συγγραφείς χρησιμοποιούν ένα επεκτεταμένο φίλτρο Kalman για να υπολογίσουν την περιστροφή του κεφαλιού (όπως επίσης και μονάδες δράσης για την εύρεση τη έκφρασης του προσώπου). Το διάλυμα κατάστασης που προκύπτει χρησιμοποιείται για την ενημέρωση ενός υποσυνόλου των παραμέτρων του Μοντέλου Ενεργού Εμφάνισης και, με αυτόν τον τρόπο, βελτιστοποιείται η αναγνώριση της θέσης των χαρακτηριστικών. Στο άρθρο αναφέρονται αποτελέσματα σε συνθετικά δεδομένα, όπως επίσης και στη βάση BU. Ωστόσο, πρέπει να τονιστεί ότι, από τις 45 ακολουθίες που αποτελούν τη βάση BU, αποτελέσματα αναφέρονται μόνο για τις 3.

Στο [55] χρησιμοποιούνται χειροκίνητα επιλεγμένα χαρακτηριστικά για να ξεκινήσει η παρακολούθησή τους, με τη χρήση εξειδικευμένων ανά χρήστη Μοντέλων Ενεργού Εμφάνισης (AAMs). Για την αναδόμηση της τρισδιάστατης δομής του προσώπου χρησιμοποιούνται μέθοδοι εύρεσης δομής από κίνηση (SFM) [162][136], ενώ

γνώση του άξονα συμμετρίας του προσώπου βοηθά στην αντιστοίχιση χαρακτηριστικών σημείων σε συγκεκριμένες συντεταγμένες. Ακολούθως, χρησιμοποιείται ο αλγόριθμος RANSAC, καθώς και γνωστές αντιστοιχίσεις 3D σημείων σε 2D συντεταγμένες για την εξαγωγή του πίνακα περιστροφής, ο οποίος πληροφορεί σχετικά με την περιστροφή του κεφαλιού. Προβολές από τις τρεις στις δύο διαστάσεις, για διάφορα σημεία, ελέγχουν ποια είναι τα πλέον κατάλληλα. Η μέθοδος χρειάζεται μόνο μία, μη βαθμονομημένη κάμερα, όμως δεν παρουσιάζονται συγκριτικά αποτελέσματα με άλλες μεθόδους, εκτός από μια σύγκριση με της μέθοδο εξαγωγής δομής από κίνηση.

Οι συγγραφείς στο [67], εκμεταλλεύονται τις ιδιότητες των μοντέλων ενεργού εμφάνισης (AAMs) για να εξάγουν πληροφορία περιστροφής κεφαλιού και κατευθυντικότητας ματιού, έτσι ώστε να εκτιμήσουν το βαθμό συγκέντρωσης οδηγού αυτοκινήτου. Για το σκοπό αυτό, το πρόσωπο παρακολουθείται με τη χρήση των τρισδιάστατων στοιχείων των AAMs και η κατευθυντικότητα του κεφαλιού βρίσκεται χρησιμοποιώντας το μοντέλο που συνδέει το διδιάστατο στο τρισδιάστατο μοντέλο. Καθώς τα μοντέλα ενεργού εμφάνισης μπορεί να εμπεριέχουν σημεία γύρω από το μάτι, οι συγγραφείς παρακολουθούν σε κάθε καρέ τις γωνίες των ματιών, και επομένως, οι ίριδες των ματιών βρίσκονται με απλές τεχνικές ταύτισης με πρότυπο και ανίχνευση ακμών.

3.2 Υπάρχουσες Βάσεις ακολουθιών Περιστροφής Κεφαλιού και Κατευθυντικότητας Ματιών στη Βιβλιογραφία

Στη βιβλιογραφία αναφέρεται πληθώρα βάσεων αναφορικά με την εκτίμηση περιστροφής κεφαλιού. Ωστόσο, είναι χαρακτηριστικό ότι δεν υπάρχουν συγκεκριμένα κριτήρια στησίματος ή επιλογής βάσης για επαλήθευση αποτελεσμάτων που προκύπτουν από τους διάφορους αλγορίθμους. Έτσι, υπάρχουν βάσεις που στήθηκαν σε ελεγχόμενες συνθήκες (πχ. φωτισμός, κίνηση, ανάλυση κάμερας), βάσεις που αποτελούνται αποκλειστικά από στατικές εικόνες, μην παρέχοντας, έτσι, τη δυνατότητα σε αλγορίθμους που στηρίζονται σε τεχνικές βίντεο να επαληθευθούν σε αυτές, αλλά και βάσεις που δεν ικανοποιούν παραδοχές που μπορεί να επιβάλλονται από τους ίδιους τους αλγορίθμους (πχ εκκίνηση ακολουθίας από εμπρόσθια περιστροφή του κεφαλιού). Ακόμη, πέραν αρκετών εξαιρέσεων, σε πολλές εργασίες που αναφέρονται στην βιβλιογραφία, αναφέρονται βάσεις οι οποίες δημιουργήθηκαν στα πλαίσια των συγκεκριμένων άρθρων και δεν έχουν δημοσιευτεί. Κάποιες από τις πλέον διαδεδομένες βάσεις στη βιβλιογραφία, για εκτίμηση πόζας κεφαλιού αναφέρονται παρακάτω.

3.2.1 Εκτενής βάση προσώπων

Η βάση αυτή [175] αποτελείται από στατικές, έγχρωμες εικόνες 20 ατόμων, σε κάθε ένα από τα οποία αντιστοιχούν 111 εικόνες (2220 συνολικά), και δημιουργήθηκε σε ελεγχόμενες συνθήκες φωτισμού (με τρεις προβολείς) και σε ομοιόμορφο φόντο (λευκό πανί). Για την εξαγωγή των γωνιών αναφοράς της περιστροφής του προσώπου, χρησιμοποιήθηκε ένας δείκτης λήξερ, για τον οποίο ζητήθηκε στους συμμετέχοντες να τον κατευθύνουν σε συγκεκριμένα σημεία στον τοίχο. Κατόπιν, εφαρμόστηκε σε έναν τρίποδα, δεύτερη συσκευή σταθμήμετρου λήξερ, το οποίο οι δημιουργοί της βάσης το έστρεφαν προς συγκεκριμένα σημεία, ελεγχόμενων γωνιών σε σχέση με



Σχήμα 3.2: Παραδείγματα από την Διευρυμένη Βάση Προσώπων

τα επισημειωμένα σημεία στον τοίχο. Έτσι, οι συμμετέχοντες έστρεφαν αρχικά το κεφάλι τους στα επισημειωμένα στον τοίχο σημεία και, κατόπιν, με σταθερό κεφάλι, έστρεφαν το βλέμμα τους στο σημείο που έδειχνε ο δεύτερος δείκτης λέιζερ. Ακολουθώντας αυτή την τακτική έχουν καταγραφεί για κάθε διαφορετική γωνία του κεφαλιού εννιά διαφορετικές συνθήκες βλέμματος, από -40° έως 40° με βήμα 10° σε σχέση με την επικρατούσα θέση του κεφαλιού. Οι πόζες του κεφαλιού έχουν καταγραφεί για οριζόντια περιστροφή του κεφαλιού από 0° έως 90° με βήμα 10° . Το ίδιο έχει γίνει και για τις κάθετες πόζες του κεφαλιού για κλίση από -20° έως 20° και για τις θέσεις με οριζόντια γωνία 0° , 30° και 60° . Κάθε πόζα έχει ληφθεί σε συνδυασμό με διαφορετικές κάθετες αποκλίσεις του βλέμματος (-20° , 0° , $+20^\circ$) σε σχέση με την θέση του κεφαλιού. Τέλος, για να αποφευχθεί η καταγραφή του ανοιγοκλεισίματος των ματιών, έχουν γίνει τρεις συνεχόμενες λήψεις για κάθε συνδυασμό, ενώ στο τέλος διάλεξαν την καλύτερη εικόνα από τις τρεις. Η ανάλυση τους είναι 1600×1200 pixels και αποθηκεύτηκαν σε jpeg κωδικοποίηση (καλή ποιότητα). Η βάση παρέχεται δωρεάν στο κοινό για καταφόρτωση από την ιστοσελίδα της στο διαδίκτυο. Στο σχήμα 3.2 φαίνεται παράδειγμα από τη βάση.

3.2.2 XM2VTS βάση προσώπων

Η βάση XM2VTS [104] αποτελεί χαρακτηριστικό παράδειγμα βάσης που δημιουργήθηκε σε ελεγχόμενες συνθήκες, ενώ, μεταξύ άλλων, παρέχονται ακολουθίες ατόμων που κάνουν προκαθορισμένη κίνηση με το κεφάλι. Εκτός από τις ελεγχόμενες συνθήκες σε φωτισμό και ανάλυση καρέ, οι κινήσεις των συμμετεχόντων είναι αργές και προκαθορισμένες, ενώ περιορίζονται σε μια οριζόντια περιστροφή του κεφαλιού, ξεκινώντας από την εμπρόσθια θέση. Συμμετείχαν 295 εθελοντές και παρήχθησαν βίντεο διαστάσεων 720×480 εικονοστοιχείων. Ωστόσο, δεν παρέχονται δεδομένα αναφοράς σχετικά με περιστροφή κεφαλιού, ενώ η βάση δεν διατίθεται ελεύθερα.



Σχήμα 3.3: Παραδείγματα από την XM2VTS

3.2.3 Βάση προσώπων «Yale», έκδοση B

Αποτελείται από 5760 εικόνες στην κλίμακα του γκρι, 10 ατόμων, κάθε ένα από τα οποία έχει 9 διαφορετικές πόζες σε 64 διαφορετικές συνθήκες φωτισμού [186]. Οι εικόνες παρέχονται με ανάλυση 640×480 , ενώ η βάση συνοδεύεται από δεδομένα αναφοράς όσον αφορά τις γωνίες περιστροφής και παρέχεται ελεύθερα. Παραδείγματα της βάσης φαίνονται στην εικόνα 3.4.

3.2.4 Βάση προσώπων Πανεπιστημίου Βοστώνης (BU)

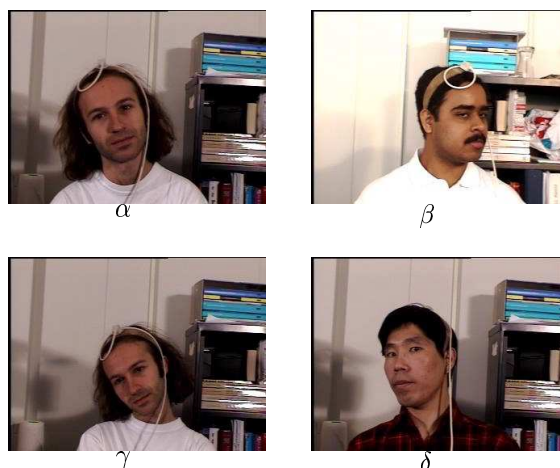
Αποτελείται από 45 ακολουθίες βίντεο [22], στις οποίες έλαβαν μέρος 9 άτομα, από 5 ακολουθίες το κάθε ένα. Οι συμμετέχοντες έκαναν ελεύθερη κίνηση του κεφαλιού τους προς όποια κατεύθυνση επιθυμούσαν, και με όποια ταχύτητα. Δεδομένα αναφοράς ελήφθησαν με τη χρήση μαγνητικού αισθητήρα, ενώ οι λήψεις έγιναν σε εσωτερικό χώρο, και με μη-ομοιόμορφο φόντο, στο οποίο, ωστόσο, δεν υπάρχει δραστηριότητα κίνησης. Τα καρέ είναι ανάλυσης 320×240 σε ρυθμό 30 καρέ ανά δευτερόλεπτο. Παραδείγματα εικόνων της βάσης φαίνονται στο σχήμα 3.5.

3.2.5 CVL βάση προσώπων

Στη βάση αυτή [127] έλαβαν μέρος 114 άτομα, και ελήφθησαν 7 εικόνες για το καθένα, οι τρεις εκ των οποίων ήταν μετωπική στάση προς την κάμερα, ενώ οι άλλες τέσσερις αντιστοιχούσαν σε οριζόντιες περιστροφές $-90^\circ, -45^\circ, 45^\circ, 90^\circ$. Η ανάλυση των φωτογραφιών είναι 640×480 , παρέχονται δωρεάν, ενώ ελήφθησαν σε ελεγχόμενες συνθήκες φωτισμού και σε ομοιόμορφο φόντο. Παράδειγμα της βάσης φαίνεται στο σχήμα 3.6.



Σχήμα 3.4: Παραδείγματα από τη βάση Yale



Σχήμα 3.5: Παραδείγματα εικόνων από τη βάση του Πανεπιστημίου της Βοστώνης



Σχήμα 3.6: Παράδειγμα από τη βάση CVL

3.2.6 INDIAN βάση προσώπων

Στη βάση [71] έλαβαν μέρος 40 άτομα, και για το καθένα ελήφθησαν 11 φωτογραφίες με διαφορετικές περιστροφές κεφαλιού. Οι φωτογραφίες ελήφθησαν σε ομοιογενές φόντο, ενώ παρέχεται πληροφορία μοναχά για τον γενικό προσανατολισμό κεφαλιού, και όχι για συγκεκριμένες γωνίες. Αναλυτικά, οι κατευθύνσεις του κεφαλιού που περιλαμβάνονται στην βάση: Μετωπική, αριστερή, δεξιά, πάνω, κάτω, πάνω-αριστερή, πάνω-δεξιά και οι εκφράσεις χαμόγελο, γέλιο, λύπη/αηδία. Οι εικόνες είναι αποθηκευμένες σε αρχεία jpeg, ενώ η ανάλυσή τους είναι 640×480 pixels σε κλίμακα του γκρι (256 επίπεδα). Η βάση διατίθεται δωρεάν για ερευνητικούς σκοπούς. Παραδείγματα της βάσης φαίνονται στην εικόνα 3.7.

3.2.7 Βάση Pointing'04

Η βάση [53] αποτελείται από 15 σετ εικόνων. Το κάθε σετ περιέχει 2 σειρές των 93 εικόνων του ίδιου ατόμου σε διαφορετικές στάσεις του κεφαλιού. Υπάρχουν συνολικά 15 άτομα στην βάση, που φορούν γυαλιά ή όχι και έχουν διαφορετικούς τόνους δέρματος. Η κατεύθυνση του κεφαλιού καθορίζεται από 2 γωνίες (οριζόντια - κάθετη), που ποικίλει από -90° έως $+90^\circ$. Παρέχονται δεδομένα αναφοράς οριζόντιας και κάθετης περιστροφής και, για να καθοριστούν οι διαφορετικές γωνίες στις στάσεις του κεφαλιού έγινε βαθμονόμηση στον τοίχο του δωματίου και τοποθετήθηκαν σημάδια στις γωνίες που αναφέρθηκαν παραπάνω ώστε να γνωρίζει κάθε άτομο που πρέπει να κοιτάξει σε κάθε νέα λήψη. Η βάση διανέμεται δωρεάν για οποιαδήποτε χρήση. Παράδειγμα της βάσης φαίνεται στην εικόνα 3.8.



Σχήμα 3.7: Παράδειγμα από τη βάση προσώπων INDIAN



Σχήμα 3.8: Παράδειγμα από τη βάση προσώπων INDIAN

3.3 Βάση ακολουθιών Περιστροφής Κεφαλιού και Κατευθυντικότητας Ματιών (HPEG)

Στα πλαίσια της παρούσας έρευνας, με απαίτηση περιβάλλοντα μη ελεγχόμενων συνθηκών και εύκολα προσβάσιμο υλικό, δημιουργήθηκε μία βάση, η βάση ακολουθιών Περιστροφής Κεφαλιού και Κατευθυντικότητας Ματιών (Head Pose - Eye Gaze dataset HPEG). Πρόσθετος στόχος στα δεδομένα αυτά, ήταν να ενσωματωθούν ακολουθίες κίνησης κεφαλιού, στις οποίες θα λαμβάνεται υπόψιν ταυτόχρονη κίνηση του κεφαλιού με αυτή των ματιών. Πέρα από στατικές εικόνες που συνδυάζουν περιστροφή κεφαλιού με κατευθυντικότητα βλέμματος, την περίοδο που δημιουργήθηκε η HPEG, δεν υπήρχε διαθέσιμο κάποιο αντίστοιχο σύνολο δεδομένων (δηλ. ακολουθίες που να εμπεριέχουν ταυτόχρονα κίνηση κεφαλιού και βλέμματος). Η βάση που δημιουργήθηκε στοχεύει στο να χρησιμοποιηθεί από αλγορίθμους υπολογισμού κατευθυντικότητας βλέμματος σε περιβάλλοντα μίας μηχανής λήψης, και που δεν εκμεταλλεύονται γνώση σχετικά με τις εσωτερικές παραμέτρους της.

3.3.1 Βάσεις περιστροφής κεφαλιού και κατευθυντικότητας βλέμματος

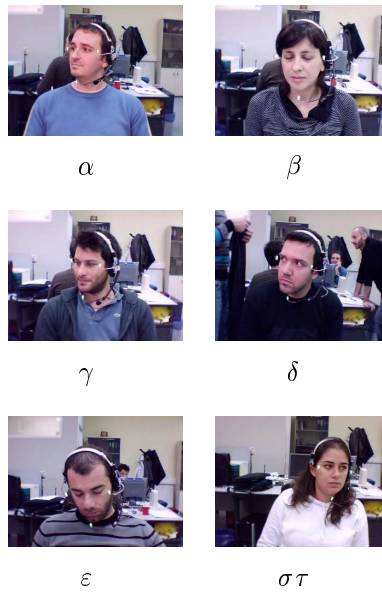
Τα δεδομένα που περιγράφονται στο [175] προσφέρουν πληροφορία παρόμοια με αυτήν της προτεινόμενης βάσης. Στη βάση αυτή, η εξαγωγή των δεδομένων αναφοράς της περιστροφής κεφαλιού έγινε ζητώντας από τους συμμετέχοντες να εφαρμόσουν μία συσκευή παραγωγής ακτίνας λέιζερ στο κεφάλι τους. Κατόπιν, τους ζητήθηκε να περιστρέψουν το κεφάλι τους προς συγκεκριμένα σημεία, ώστε η ακτίνα να συμπίσει σε αυτά. Με σταθερό κεφάλι υπό διάφορες γωνίες, περιστρέφοντα τα μάτια τους σε συγκεκριμένες θέσεις. Έτσι, για ένα σύνολο 20 συμμετεχόντων, δημιουργήθηκε μία βάση 2220 στατικών εικόνων, με διαφορετικούς συνδυασμούς περιστροφής κεφαλιού και βλέμματος ματιών. Βασική διαφορά της βάσης αυτής από την HPEG είναι ότι η βάση στο [175] έχει ελεγχόμενο φόντο, καθώς και το ότι προσφέρεται μονάχα υπό μορφή στατικών εικόνων, κι έτσι καθίσταται ακατάλληλη για αλγορίθμους που χρησιμοποιούν τεχνικές βίντεο για εξαγωγή πληροφορίας.

3.3.2 Περιγραφή της βάσης

Η βάση HPEG, η οποία δημιουργήθηκε σε συνθήκες εσωτερικού χώρου του εργαστηρίου, αποτελείται από 20 έγχρωμες ακολουθίες. Στη βάση συμμετείχαν 10 άτομα - 8 άνδρες, 2 γυναίκες -, ενώ για κάθε έναν έγιναν δύο λήψεις. Για τη μαγνητοσκόπηση χρησιμοποιήθηκε μία απλή κάμερα διαδικτύου ¹, η οποία μπορεί να κάνει λήψη ακολουθιών μεγέθους έως και 1280×960 εικονοστοιχείων, με ρυθμό 30 fps (καρέ ανά δευτερόλεπτο). Το λογισμικό που χρησιμοποιήθηκε για την καταγραφή των ακολουθιών ήταν το Hercules Webcam Station Evolution SE, ενώ χρησιμοποιήθηκαν τα προτερότερα χαρακτηριστικά λειτουργίας στην ισορροπία λευκού, στον κορεσμό χρώματος, τη φωτεινότητα και την αντίθεση. Όλες οι ακολουθίες ελήφθησαν με ανάλυση 640×480 εικονοστοιχεία, με ρυθμό 30fps, σε μορφή μη συμπιεσμένου avi αρχείου.

Για την πρώτη λήψη, ο συμμετέχων πρέπει να κοιτάζει την κάμερα μετωπικά και, ακολούθως, μπορεί να κινηθεί προς οποιαδήποτε κατεύθυνση επιθυμεί. Η κίνηση του

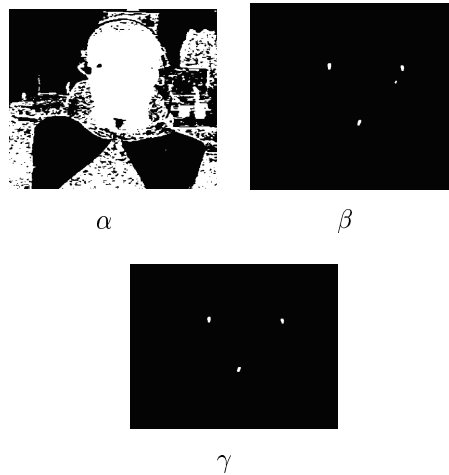
¹Hercules Dualpix Exchange webcam



Σχήμα 3.9: Παραδείγματα από την πρώτη συνεδρία



Σχήμα 3.10: Παραδείγματα από τη δεύτερη συνεδρία



Σχήμα 3.11: (α) Αρχική δυαδική μάσκα, όπως υπολογίστηκε από τις τιμές των καναλιών a^* και b^* ; (β) Δυαδική μάσκα, μετά το λογικό AND μεταξύ της εικόνας (α) και του δυαδικού χάρτη φωτεινότητας L , (γ) τελική δυαδική μάσκα των φωτοδιόδων μετά από χρήση της γεωμετρίας τους

κεφαλιού του μπορεί να είναι, είτε μεταφορική, είτε περιστροφική, ή και τα δύο ταυτόχρονα. Η κάμερα τοποθετήθηκε σε απόσταση περίπου ενός μέτρου από τον συμμετέχοντα, και σε ύψος περίπου το ίδιο με το επίπεδο των ματιών του, ενώ φαίνεται και μέρος του κορμού του σώματος. Επιπλέον, μεγάλο μέρος των καρέ καλύπτεται από το φόντο, το οποίο είναι πραγματικός χώρος εργασίας, με κίνηση σε μερικές από τις ακολουθίες. Οι συνθήκες φωτισμού είναι ανάλογες με τις αναμενόμενες σε έναν εσωτερικό χώρο, και παραμένουν σταθερές για όλες τις λήψεις. Συνολικά, κάθε ακολουθία της πρώτης λήψης διαρκεί 200 καρέ. Παραδείγματα από την πρώτη λήψη φαίνονται στην εικόνα 3.9.

Για τη δεύτερη λήψη, η οποία τραβήχτηκε στις ίδιες συνθήκες με την πρώτη, οι συμμετέχοντες έπρεπε να ακολουθήσουν μία κίνηση που να συνδυάζει περιστροφική κίνηση του κεφαλιού με περιστροφική κίνηση των ματιών, όπως φαίνεται στο σχήμα 3.10. Η κάμερα έχει τοποθετηθεί σε απόσταση περίπου 30 εκατοστών από τον συμμετέχοντα. Το πρωτόκολλο που ακολουθήθηκε ήταν αυτό που φαίνεται στο σχήμα 3.10: Στο καρέ 3.10α, το άτομο κοιτάζει ευθεία μπροστά, τόσο με τα μάτια, όσο και με τη στρέψη του κεφαλιού του. Στα καρέ 3.10β και 3.10γ, το κεφάλι εξακολουθεί να είναι στραμμένο μετωπικά σε σχέση με την κάμερα, όμως τα μάτια ακολουθούν μία κίνηση από αριστερά προς τα δεξιά (ή το αντίστροφο). Στο καρέ 3.10δ, το άτομο έχει στρέψει το κεφάλι στα δεξιά του, με τα μάτια να κοιτάζουν ευθεία σε σχέση με το κεφάλι, ενώ στα καρέ 3.10ε και 3.10στ, το κεφάλι παραμένει σταθερό, με τα μάτια, όμως, να κάνουν μία κίνηση, από τα αριστερά (βλέμμα στραμμένο στην κάμερα) προς τα δεξιά. Ο αριθμός των καρέ ανά ακολουθία, για τη δεύτερη λήψη, δεν είναι σταθερός, και ποικίλλει, από 349 έως 430 καρέ ανά ακολουθία.

3.3.3 Εξαγωγή δεδομένων αναφοράς

Για την εξαγωγή των δεδομένων αναφοράς της περιστροφής του κεφαλιού, χρησιμοποιήθηκαν 3 φωτοδιόδοι, οι οποίες τοποθετήθηκαν στο κεφάλι του εκάστοτε συμμετέχοντα. Ο σκοπός ήταν να γίνει πετυχημένη και ακριβής παρακολούθησή τους ώστε, με αξιόπιστο τρόπο, να είναι διαθέσιμες οι θέσεις τους καθόλη τη διάρκεια της κάθε

ακολουθίας. Με βάση τη θέση των φωτοδιόδων, όπως θα αναλυθεί παρακάτω, έγινε η εξαγωγή των δεδομένων αναφοράς των γωνιών του κεφαλιού. Χρησιμοποιήθηκαν φωτοδιόδοι μεγάλης έντασης γωνίας δέσμης 15° , ενώ έγινε χρήση και μεταβλητής αντίστασης για να ρυθμιστεί η φωτεινότητα των φωτοδιόδων σε τέτοια επίπεδα, ούτως ώστε, α) να είναι ορατές υπό φυσιολογικές συνθήκες φωτισμού και β) να μη μεταβάλλεται το χρώμα τους (να μη φαίνονται λευκές λόγω έντονης φωτεινότητας). Επιπλέον, η περιορισμένη δέσμη φωτός τους έχει το μειονέκτημα ότι, με την περιστροφή του κεφαλιού, τις καθιστά μη ορατές. Για το σκοπό αυτό, τοποθετήθηκαν προς τα επάνω, και περιτυλίχθηκαν με ημι-διαφανές υλικό, ώστε να διαχέεται το φως ομοιόμορφα, και να μην παραμορφώνεται το σχήμα τους με την περιστροφή του κεφαλιού. Επίσης, επιλέχθηκαν φωτοδιόδοι πράσινου φωτός, για να μην μοιάζουν με πιθανά τεχνουργήματα (συνήθως κόκκινης ή μπλε απόχρωσης) μιας κάμερας χαμηλής ποιότητας. Επιπλέον, το πράσινο επιλέχθηκε ως χρώμα αρκετά διαφορετικό από αυτό του δέρματος, αλλά και του περιβάλλοντος των συγκεκριμένων συνθηκών.

Τα δεδομένα αναφοράς για την περιστροφή του προσώπου (οριζόντιες και κατακόρυφες γωνίες περιστροφής), υπολογίστηκαν με βάση τις θέσεις των φωτοδιόδων σε κάθε καρέ των ακολουθιών. Για το σκοπό αυτό, χρησιμοποιήθηκε ένας ημι-αυτόματος τρόπος, με την έννοια ότι στο πρώτο καρέ της κάθε ακολουθίας καθορίστηκαν οι θέσεις των φωτοδιόδων χειροκίνητα, ενώ στα ακόλουθα καρέ, η θέση τους εξήχθη με τεχνικές παρακολούθησης. Πιο συγκεκριμένα, ο χρωματικός χώρος των καρέ μετασχηματίστηκε στον χρωματικό χώρο $L^*a^*b^*$ και, ακολούθως, υπολογίστηκαν οι μέσες τιμές των περιοχών των φωτοδιόδων σε κάθε ένα από τα τρία κανάλια. Παρομοίως, υπολογίστηκαν οι μέσες τιμές στα τρία κανάλια και για την υπόλοιπη έκταση των καρέ (*non-LED* περιοχές). Σε ακόλουθα καρέ, έγινε σύγκριση της τιμής του κάθε εικονοστοιχείου στα κανάλια L^* , a^* και b^* με αυτήν των LED και *non-LED* περιοχών. Τα κανάλια χρωματικότητας (a^* και b^*) αντιμετωπίστηκαν σε κοινό σύστημα συντεταγμένων, και τα εικονοστοιχεία που βρίσκονταν πιο κοντά στις μέσες τιμές των φωτοδιόδων από αυτές των *non-LED* περιοχών θεωρήθηκαν υποψήφια εικονοστοιχεία φωτοδιόδων. Παρόμοια διαδικασία ακολουθήθηκε και για το κανάλι της φωτεινότητας μεμονωμένα. Το αποτέλεσμα ήταν δύο χάρτες, οι οποίοι, αθροίζόμενοι με το λογικό *AND*, έδωσαν περιορισμένο αριθμό υποψήφιων περιοχών φωτοδιόδων.

Επιπλέον, χρησιμοποιήθηκε πληροφορία σχετικά με το μέγεθος και τη θέση της κάθε φωτοδιόδου στο προηγούμενο καρέ, ενώ περισσότερες *non-LED* περιοχές αποκλείστηκαν εάν δεν συμφωνούσαν με αυτήν την πληροφορία. Η παραπάνω διαδικασία συνοψίζεται στο σχήμα 3.11. Αξίζει να σημειωθεί ότι οι φωτοδιόδοι τοποθετήθηκαν ελαφρώς πιο μπροστά από το επίπεδο του προσώπου, ώστε να είναι συνεχώς ορατές για τις περιστροφές που πραγματοποίησαν οι συμμετέχοντες.

Με την γνώση της θέσης των φωτοδιόδων σε κάθε καρέ, ο υπολογισμός των γωνιών περιστροφής του προσώπου μπορεί να γίνει με ευκολία. Για το σκοπό αυτό, χρησιμοποιήθηκαν οι εξισώσεις που περιγράφονται στο [189] για την εξαγωγή της οριζόντιας και κατακόρυφης περιστροφής του κεφαλιού σε κάθε καρέ. Για τον υπολογισμό των δεδομένων αναφοράς, ωστόσο, εδώ χρησιμοποιήθηκαν οι θέσεις των φωτοδιόδων και όχι οι θέσεις του στόματος και των φρυδιών, όπως έγινε στο [189].

Στον πίνακα 3.1 φαίνονται συγκεντρωτικά τα σημαντικότερα χαρακτηριστικά των βάσεων που αναφέρθηκαν στο κεφάλαιο αυτό, έτσι ώστε να γίνει μια συνοπτική επισκόπηση των πλεονεκτημάτων/μειονεκτημάτων της κάθε μίας.

Πίνακας 3.1: Βάσεις εικόνων/βίντεο που χρησιμοποιούνται στη βιβλιογραφία για εκτίμηση περιστροφής κεφαλιού

	Βίντεο	Δεδομένα αναφοράς	Δωρεάν διάθεση	Έγχρωμη	Πολύπλοκο φόντο	Βλέμμα ματιών
Διευρυμένη		✓	✓	✓		✓
XM2VTS	✓			✓		
Yale		✓	✓		✓	
CVL		✓	✓	✓		
INDIAN			✓			
Pointing04		✓	✓	✓		
Πανεπ. Βοστώνης	✓	✓	✓	✓	✓	
HPEG	✓	✓	✓	✓	✓	✓

3.3.4 Πειραματική επαλήθευση χρησιμότητας βάσης

3.3.4.1 Εκτίμηση περιστροφής κεφαλιού

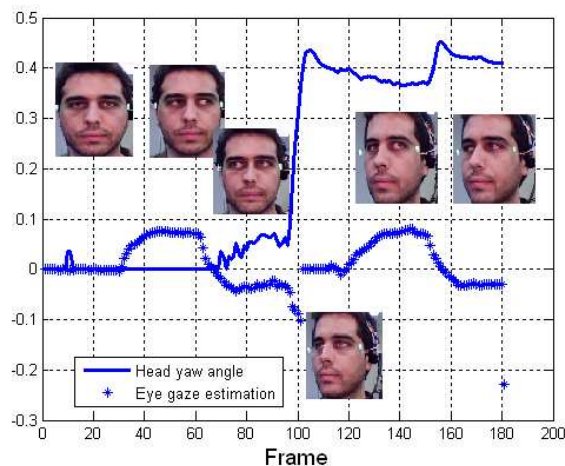
Με σκοπό να επιβεβαιωθεί η χρηστική αξία της βάσης HPEG, εξετάστηκαν τα αποτελέσματα εκτίμησης γωνιών περιστροφής του αλγορίθμου παρακολούθησης χαρακτηριστικών που θα περιγραφεί στο κεφάλαιο 4. Ο ίδιος αλγόριθμος εξετάστηκε και σε άλλη, ευρέως διαδεδομένη βάση, την βάση του Πανεπιστημίου της Βοστώνης [22]. Τα αποτελέσματα στις δύο βάσεις, όπως θα φανεί και στο αντίστοιχο κεφάλαιο είναι συγκρίσιμα, τόσο μεταξύ τους, όσο και με άλλες μεθόδους της βιβλιογραφίας. Κάτι τέτοιο δείχνει δύο πράγματα: Πρώτον, ότι η μέθοδος που χρησιμοποιήθηκε για την εκτίμηση των γωνιών είναι αρκετά αξιόπιστη, αφού δίνει σχετικά μικρό σφάλμα στην βάση του Πανεπιστημίου της Βοστώνης. Καθώς δίνει αντίστοιχης τάξης μεγέθους σφάλμα και στην HPEG, μπορεί να βγει το συμπέρασμα ότι ο τρόπος με τον οποίο υπολογίστηκαν τα δεδομένα αναφοράς για τη βάση, παρέχει αρκετά μεγάλη αξιοπιστία, καθιστώντας τη βάση κατάλληλη για επαλήθευση της λειτουργικότητας αντίστοιχων συστημάτων.

3.3.4.2 Εκτίμηση κατευθυντικότητας βλέμματος

Η προτεινόμενη βάση παρέχει, όπως έχει αναλυθεί, σύνολο 10 ακολουθιών ατόμων που, ταυτόχρονα με περιστροφή προσώπου, κινούν και τα μάτια τους. Με παρακολούθηση του κέντρου των ματιών και της σχετικής τους θέσης αναφορικά με τα κατακόρυφα όρια του προσώπου (βλ. κεφάλαιο 4), έγινε κάποια αρχική εκτίμηση της κατευθυντικότητας των ματιών. Στο σχήμα 3.12 φαίνεται χαρακτηριστικό παράδειγμα της μεταβλητότητας της οριζόντιας κατευθυντικότητας του βλέμματος των ματιών, και οι αντίστοιχες τιμές της οριζόντιας περιστροφής κεφαλιού.

3.4 Παρακολούθηση χαρακτηριστικών

Για την παρακολούθηση των χαρακτηριστικών του προσώπου, μπορεί να χρησιμοποιηθεί ένας πυραμιδικός αλγόριθμος Lucas-Kanade [19] δύο επιπέδων, με προσαρμοστική



Σχήμα 3.12: Παράδειγμα ακολουθίας της δεύτερης συνεδρίας, και αποτελέσματα στην εκτίμηση κατευθυντικότητας βλέμματος ματιών.

περιοχή αναζήτησης, διαστάσεων όσο το ένα πέμπτο της κάθε γραμμής και στήλης της περιοχής του προσώπου του ανώτερου επιπέδου της εικόνας. Με αυτό τον τρόπο, η παρακολούθηση των χαρακτηριστικών έχει τη δυνατότητα να προσαρμόζεται στην κλίμακα με την οποία απεικονίζεται το πρόσωπο στην ακολουθία. Μέσα από πειραματικές διαδικασίες, φάνηκε γρήγορα, ότι ο αλγόριθμος από μόνος του είναι αδύνατον να παρακολουθήσει το πρόσωπο για μια αρκετά μεγάλη ακολουθία καρέ. Κάτι τέτοιο δεν οφείλεται μονάχα στη μικρή ανάλυση των ακολουθιών, που είναι και το ζητούμενο στην παρούσα δουλειά (web-cameras), ή στις μη ελεγχόμενες συνθήκες φωτισμού και κινήσεων. Αντιθέτως, ακόμα και σε ακολουθίες αρκετά μεγάλης ανάλυσης και πολύ καλής ποιότητας χρώματος, ο αλγόριθμος αποτυγχάνει να παρακολουθεί τα χαρακτηριστικά για αρκετά μεγάλο αριθμό καρέ, καθιστώντας έτσι το πρόβλημα της εύρεσης περιστροφής κεφαλιού δύσκολο. Για το σκοπό αυτό, μπορεί να υιοθετηθεί μία σειρά κανόνων. Ενώ οι ιδέες πίσω από την εφαρμογή τους βασίζονται σε απλές αρχές γεωμετρίας, φυσιολογίας και αναμενόμενης κίνησης του κεφαλιού, αποδεικνύεται ότι οι κανόνες αυτοί είναι άκρως απαραίτητοι για την ευρωστία του συστήματος και την ανάκτηση αξιόπιστων αποτελεσμάτων.

1. Θεωρώντας ως p_x και p_y τις συνιστώσες του διανύσματος περιστροφής (βλ. παρακάτω), όταν ο χρήστης έχει το κεφάλι του στραμμένο ευθεία, ισχύει ότι $|p_x| < 0.1$ και $|p_y| < 0.1$. Όταν αρχίζει να περιστρέφει το κεφάλι του γύρω από τον κάθετο άξονα, το $|p_y|$ αυξάνει σε μέγεθος και το $|p_x|$ παραμένει σχετικά σταθερό. Όταν ο χρήστης αρχίσει να επιστρέφει προς την ευθεία κατάσταση, το μάτι εκείνο το οποίο έχει αποκρυφτεί λόγω της περιστροφής, δεν μπορεί να ανακτηθεί. Επιπλέον, και το σημείο ανάμεσα στους ρώθωνες πολλές φορές δεν μπορεί να ανακτηθεί. Ωστόσο, το μάτι εκείνο το οποίο ήταν ορατό καθ'όλη τη διάρκεια της κίνησης, εξακολουθεί να παρακολουθείται επιτυχώς από τον αλγόριθμο. Κάτω από αυτές τις συνθήκες, το μέτρο του διανύσματος περιστροφής εξακολουθεί να μειώνεται, χάρη στο μάτι το οποίο είναι ορατό, ενώ τα ση-

μεία εκείνα τα οποία αποκρύφτηκαν, συνήθως έχουν αντιστοιχηθεί με κάποια σταθερά σημεία στο φόντο. Συμβολίζοντας με n το καρέ εκείνο, στο οποίο ο χρήστης έχει επιστρέψει στην ευθεία θέση ως προς την κάμερα, τότε, κάτω από κανονικές συνθήκες, το μέτρο του διανύσματος περιστροφής $\|\bar{p}\|$, ήταν κατά πολύ μεγαλύτερο στο καρέ $n-n_t$, όταν ο χρήστης είχε στραμμένο το κεφάλι του. Επιπλέον, όταν ο χρήστης βρίσκεται σε ευθεία θέση, η μεταβλητότητα του $\|\bar{p}\|$ είναι πολύ μικρή σε σχέση με αυτή που είχε όσο ήταν εν κινήσει. Εάν οι παραπάνω δύο συνθήκες ισχύουν (εξισώσεις 3.1, 3.2), τότε συμπεραίνεται ότι ο χρήστης βρίσκεται σε ευθεία στάση σε σχέση με την κάμερα, και το σύστημα μπορεί να επανεκκινηθεί.

$$\|\bar{p}(n)\| < thr_1 \times \|\bar{p}(n - n_{t1})\| \quad (3.1)$$

$$var(\|\bar{p}_{n-n_{t2}:n}\|) < thr_2 \quad (3.2)$$

Στα πειράματά μας θεωρήσαμε $n_{t1}=10$, $n_{t2}=7$, $thr_1=0.7$ και $thr_2=0.05$. Εάν ικανοποιούνται οι συνθήκες αυτές, όμως στην πραγματικότητα ο χρήστης δεν έχει επιστρέψει σε ευθεία στάση, χρησιμοποιώντας μονάχα αλγόριθμο Viola-Jones τέτοιον που να εντοπίζει πρόσωπο μόνο σε μετωπική στάση, η εύρεση προσώπου αποτυγχάνει, αποφεύγοντας με αυτόν τον τρόπο την επανεκκίνηση του συστήματος από λάθος σημείο αναφοράς.

2. Για μετακινήσεις του κεφαλιού παράλληλες στο επίπεδο της κάμερας, ο λόγος της απόστασης μεταξύ των οφθαλμών και της απόστασης του σημείου ανάμεσα στα μάτια από το στόμα παραμένει σταθερός. Το μέγεθος αυτό μπορεί να παρακολουθείται συνεχώς κατά τη διάρκεια της ακολουθίας και, εάν βρίσκεται σε συγκεκριμένα όρια σε σχέση με την τιμή του στο καρέ αρχικοποίησης, τότε θεωρείται μηδενική περιστροφή.
3. Επίσης, μπορούν να εφαρμοστούν κανόνες που λαμβάνουν υπόψη την φυσιολογία του προσώπου. Μετά από γρήγορες περιστροφές του προσώπου, η παρακολούθηση των σημείων συχνά αποτυγχάνει, και οι αποστάσεις μεταξύ των οφθαλμών, ή των οφθαλμών και της μύτης, παίρνουν μη φυσιολογικές τιμές, οπότε και το σύστημα μπορεί να επανεκκινηθεί. Όπως και στην περίπτωση του πρώτου κανόνα, εάν τα παραπάνω ισχύουν χωρίς το πρόσωπο να έχει επιστρέψει σε ευθεία στάση, τότε δεν μπορεί να γίνει ανίχνευση προσώπου, μέχρι ο χρήστης να ξανατοποθετηθεί αναλόγως. Οι υποκανόνες που μπορούν να χρησιμοποιηθούν είναι οι ακόλουθοι:

- $|x_n - x_m| < 1.5 \times |y_{el} - y_{er}|$
- $y_{el} > y_{er}$
- $y_n < y_{el}$
- $y_n > y_{er}$

Όπου (y_{el}, x_{el}) , (y_{er}, x_{er}) και (y_n, x_n) , (y_m, x_m) είναι οι οριζόντιες και κάθετες συντεταγμένες του αριστερού και δεξιού ματιού και του σημείου ανάμεσα στους ρόθωνες και ανάμεσα στα μάτια, αντίστοιχα. Εάν κάποιος από τους υποκανόνες

δεν ικανοποιείται σε κάποια φάση της ακολουθίας, τότε το σύστημα επανεκκινείται. Μέσα από πειράματα, φάνηκε ότι οι συγκεκριμένοι υποκανόνες ήταν οι πλέον χρησιμοποιούμενοι και σε θέση να επαναφέρουν το σύστημα σε αξιόπιστα αποτελέσματα κάτω από πολλές συνθήκες: Όταν ο χρήστης φύγει από το οπτικό πεδίο της κάμερας, όταν κάνει απότομες μετακινήσεις, όταν για κάποια δευτερόλεπτα αποκρυφτεί το πρόσωπο (πχ αν περάσει το χέρι από μπροστά), αλλά και σε περιπτώσεις που η κάμερα μετακινηθεί στιγμιαία.

4. Επιπλέον περιορισμοί μπορούν να επιβληθούν λαμβάνοντας υπόψη το γεγονός ότι αναμένεται όλα τα σημεία να έχουν μετατοπίσεις της ίδιας τάξης μεγέθους. Έτσι, σε κάθε καρέ υπολογίζεται η μέση μετατόπιση όλων των σημείων και βρίσκονται εκείνα τα σημεία των οποίων οι μετατοπίσεις είναι μικρότερες ή μεγαλύτερες συγκεκριμένων κλασμάτων (t_a, t_b αντίστοιχα) της μέσης. Ξαναυπολογίζεται η μέση μετατόπιση όλων των σημείων χωρίς να λαμβάνεται υπόψη των έκτοπων σημείων, τα οποία επαναπροσδιορίζονται με βάση τη νέα μετατόπιση. Η διαδικασία συνεχίζεται μέχρι να μην υπάρχει κανένα σημείο που να χρειάζεται να επαναπροσδιοριστεί. Θέτοντας $t_a=0.3$ και $t_b=1.5$, βρέθηκε ότι, για κάθε καρέ, ο αλγόριθμος αυτός συγκλίνει ύστερα από 7 με 10 επαναλήψεις κατά μέσο όρο.

3.5 Εκτίμηση περιστροφής κεφαλιού

Με τη συγκεκριμένη μέθοδο εκτίμησης περιστροφής κεφαλιού, αν και χρησιμοποιούνται οι θέσεις χαρακτηριστικών σημείων στο πρόσωπο, η ευρωστία του συστήματος δεν εξαρτάται απόλυτα από τον ακριβή προσδιορισμό της θέσης. Με άλλα λόγια, η μέθοδος εξαρτάται σε μεγαλύτερο βαθμό από την εύρωστη παρακολούθηση των σημείων, παρά τον ακριβή προσδιορισμό της θέσης τους. Αυτό οφείλεται στο ότι ο επιτυχής υπολογισμός του διανύσματος περιστροφής κεφαλιού είναι περισσότερο θέμα των σχετικών μετατοπίσεων αυτών των σημείων, παρά των σχετικών μεταξύ τους θέσεων. Αυτή η ιδιότητα συμβάλει στην ευρωστία της μεθόδου, ακόμα και κάτω από συνθήκες χαμηλών διαστάσεων ακολουθιών, με μη ιδανικό φωτισμό. Σε τέτοιες ακολουθίες, ο εντοπισμός των σημείων ενδεχομένως να μην είναι όσο ακριβής θα ήταν σε βίντεο μεγαλύτερων διαστάσεων και πιο ελεγχόμενων συνθηκών φωτισμού. Ωστόσο, εφόσον η παρακολούθηση των σημείων είναι ικανοποιητική, και οι κανόνες που περιγράφονται στο 3.4 επαναφέρουν το σύστημα, όποτε κρίνεται αναγκαίο, η εκτίμηση της περιστροφής του κεφαλιού είναι ικανοποιητική. Ένα από τα μειονεκτήματα της μεθοδολογίας είναι ότι απαιτεί ο χρήστης να κοιτάζει την κάμερα μετωπικά (η οριζόντια και η κάθετη γωνία περιστροφής να είναι 0) στην αρχή. Αυτό απαιτείται επειδή, όπως θα φανεί παρακάτω, σαν καρέ αναφοράς λαμβάνεται το αρχικό, και όλοι οι υπολογισμοί γίνονται με βάση τις θέσεις των σημείων σε αυτό. Για να αποφευχθεί ο εντοπισμός προσώπου περιστραμμένου με κάποια γωνία, χρησιμοποιείται αλγόριθμος Viola-Jones εκπαιδευμένος μονάχα σε δεδομένα προσώπων στραμμένα ευθεία προς την κάμερα. Έτσι, ακόμα και αν μια ακολουθία ξεκινάει με κάποιο πρόσωπο μη στραμμένο ευθεία προς την κάμερα, αυτό δεν εντοπίζεται, μέχρι να περιστραφεί ευθεία προς αυτήν.

Για την εκτίμηση του διανύσματος πόζας κεφαλιού, χρησιμοποιείται η 2Δ μετατόπιση ενός σημείου αναφοράς επάνω στο πρόσωπο, όπως αυτή προβάλλεται στο επίπεδο της εικόνας. Εδώ, χρησιμοποιήθηκε το μέσο των δύο οφθαλμών. Σε κάθε καρέ υπολογίζεται η μετατόπισή του σημείου αυτού σε σχέση με τη θέση που έχει στο



Σχήμα 3.13: Παράδειγμα διόρθωσης παρακολούθησης χαρακτηριστικών σε πραγματικό χρόνο. Τα περισσότερα σημεία στο πρόσωπο «χάνονται», όμως, αφού ο χρήστης επιστρέψει σε εμπρόσθια θέση, ανακτώνται με την επαναρχικοποίηση. Η λευκή γραμμή αντιστοιχεί στο διάνυσμα περιστροφής

καρέ αναφοράς. Η οριζόντια συνιστώσα της μετατόπισης αντιστοιχεί στην περιστροφή του κεφαλιού γύρω από τον κατακόρυφο άξονα, ενώ η κάθετη συνιστώσα αντιστοιχεί στην περιστροφή γύρω από τον οριζόντιο άξονα. Κανονικοποίηση της μετατόπισης αυτής με την απόσταση μεταξύ των οφθαλμών, όπως αυτή υπολογίστηκε στο καρέ αναφοράς, μπορεί να εξασφαλίσει σωστή εκτίμηση της πόζας για διάφορους παράγοντες κλιμάκωσης. Έτσι, κάθε φορά που επανεκκινείται το σύστημα, κρατιέται η καινούρια απόσταση μεταξύ των οφθαλμών και όλοι οι υπολογισμοί γίνονται με βάση αυτήν, ενσωματώνοντας, έτσι, στο πρόβλημα, και συνθήκες στις οποίες ο χρήστης κάνει μεταφορική κίνηση κάθετη στο επίπεδο της κάμερας. Όπως θα φανεί και στα επόμενα κεφάλαια, το μέτρο και η κατευθυντικότητα του διανύσματος περιστροφής του κεφαλιού $\bar{p}=(p_x, p_y)$, μπορεί να χρησιμοποιηθεί με επιτυχία σε προβλήματα όπου απαιτείται η παρακολούθηση της προσοχής του χρήστη σε κάποιο πολυμεσικό περιεχόμενο, σε συνθήκες οδήγησης, στην επικοινωνία με κάποιον ψηφιακό πράκτορα, σε περιβάλλοντα εκπαίδευσης, κλπ.

Στις εικόνες 3.13 και 3.14 φαίνονται χαρακτηριστικά παραδείγματα εύρεσης πόζας κεφαλιού και τα σημεία που παρακολουθούνται. Επίσης, φαίνονται συνθήκες επανεκκίνησης του συστήματος. Στο σχήμα 3.13, οι θέσεις των χαρακτηριστικών του χρήστη, ενώ δεν έχουν αποκρυφτεί, έχουν τοπολογία που δε συμφωνεί με την αναμενόμενη της ανθρώπινης φυσιολογίας. Με την ενεργοποίηση του κατάλληλου κανόνα, το σύστημα επαναρχικοποιείται όταν ο χρήστης επιστρέψει στην εμπρόσθια θέση. Στην εικόνα 3.14 φαίνεται ότι, ύστερα από μία περιστροφή του κεφαλιού του χρήστη προς τα δεξιά του, αποκρύφτηκε ο ένας οφθαλμός, η θέση του οποίου, όμως, ανακτήθηκε μετά την επανεκκίνηση του συστήματος.

Για την εύρεση της γωνίας που σχηματίζεται με περιστροφή του κεφαλιού γύρω από τον κάθετο στο επίπεδο της κάμερας άξονα, υπολογίζεται η γωνία που σχηματίζεται από την ευθεία που ενώνει τους οφθαλμούς με τον οριζόντιο άξονα.

Οι συνιστώσες του διανύσματος περιστροφής φιλτράρονται με ένα φίλτρο Kalman. Σαν καταστάσεις στο φίλτρο Κάλμαν στο καρέ n θεωρούνται οι οριζόντιες και κάθετες συνιστώσες $\mathbf{p}_n = (p_{x,n}, p_{y,n})$ του διανύσματος περιστροφής κεφαλιού, καθώς και οι αντίστοιχες συνιστώσες της ταχύτητας, $\mathbf{u}_n = (u_{x,n}, u_{y,n})$. Συνεπώς, το διάνυ-



Σχήμα 3.14: Παράδειγμα διόρθωσης παρακολούθησης χαρακτηριστικών σε πραγματικό χρόνο. Εδώ αποκρύφτηκε το σημείο που αντιστοιχεί στην εξωτερική γωνία του δεξιού φρυδιού. Η θέση του ανακτάται, όμως, εφόσον υπάρχει επαναρχικοποίηση. Η λευκή γραμμή αντιστοιχεί στο διάνυσμα περιστροφής

σμα κατάστασης τη χρονική στιγμή n θα έχει τη μορφή $\mathbf{x}_n = (p_{x,n}, p_{y,n}, u_{x,n}, u_{y,n})$. Σύμφωνα με τη θεωρία Kalman [102] το διάνυσμα κατάστασης \mathbf{x}_{n+1} που αντιστοιχεί στο καρέ $n + 1$ σχετίζεται γραμμικά με το τρέχον διάνυσμα κατάστασης \mathbf{x}_n , με ένα μοντέλο συστήματος της μορφής που ορίζεται από την εξίσωση 3.3.

$$\mathbf{x}_{n+1} = \Phi \mathbf{x}_n + \mathbf{w}_n \quad (3.3)$$

όπου Φ ο πίνακας μετάβασης καταστάσεων και \mathbf{w}_n ο θόρυβος του συστήματος, κανονικής κατανομής $\mathbf{w}_n \sim N(0, \mathbf{Q})$.

Μπορεί να θεωρηθεί αρκετά μικρή μεταβολή του διανύσματος περιστροφής μεταξύ διαδοχικών καρέ, ώστε να υποθεθεί ομοιόμορφη μεταβολή κατά τη διάρκεια της ακολουθίας, οπότε και μπορεί να υιοθετηθεί γραμμικό μοντέλο, και ο πίνακας μεταβάσεων καταστάσεων να παραμετροποιηθεί ως εξής:

$$\Phi = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Θεωρώντας ότι οι μεταβλητές παρατηρήσεων είναι οι δύο πρώτες (συνιστώσες του διανύσματος περιστροφής) του διανύσματος κατάστασης, και ότι το μοντέλο παρατήρησης περιγράφεται από την εξίσωση 3.4.

$$\mathbf{z}_n = \mathbf{H} \mathbf{x}_n + \mathbf{v}_n \quad (3.4)$$

τότε

$$\mathbf{H} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

με \mathbf{H} τον πίνακα που σχετίζει την τρέχουσα κατάσταση με το διάνυσμα παρατήρησης, και \mathbf{v}_n το θόρυβο μέτρησης, ο οποίος θεωρείται πως αντιστοιχεί σε κανονική κατανομή $\mathbf{v}_n \sim N(0, \mathbf{R})$.

Θεωρώντας κανονικές κατανομές στις μεταβλητές καταστάσεων και στις πιθανότητες μεταβάσεων, ο αλγόριθμος Κάλμαν προσφέρει τη δυνατότητα υιοθέτησης αναδρομικών τύπων, με τους οποίους μπορεί να υπολογιστεί το διάνυσμα καταστάσεων από τις παρατηρήσεις. Σαν αρχικό διάνυσμα καταστάσεων θεωρήσαμε μηδενικές συνιστώσες του διανύσματος περιστροφής κεφαλιού, και σαν ταχύτητα, την διαφορά των συνιστωσών μεταξύ των δύο πρώτων καρέ (ο αλγόριθμος ξεκίνησε στο δεύτερο καρέ).

$$\begin{aligned} p_{x,0} &= 0 \\ p_{y,0} &= 0 \\ u_{x,0} &= p_{x,1} - p_{x,0} \\ u_{y,0} &= p_{y,1} - p_{y,0} \end{aligned}$$

Καθώς ο πίνακας συμμεταβλητότητας σφάλματος του διανύσματος καταστάσεων δεν είχε ιδιαίτερη επίδραση στα αποτελέσματα, τέθηκε αρχικά ίσος με τον μοναδιαίο 4×4 . Επίσης, θεωρώντας ότι ο θόρυβος του συστήματος είναι περίπου 0.2° και, διαιρώντας τον με τον παράγοντα κλιμάκωσης που αντιστοιχίζει τις μετρήσεις με τα δεδομένα αναφοράς, προκύπτει θόρυβος του συστήματος περίπου 10^{-2} . Αντίστοιχος θόρυβος θεωρήθηκε και για τις καταστάσεις της ταχύτητας. Έτσι, ο πίνακας συμμεταβλητότητας του θορύβου του συστήματος που θεωρήθηκε, ήταν ο ακόλουθος:

$$Q = \begin{pmatrix} 10^{-4} & 0 & 0 & 0 \\ 0 & 10^{-4} & 0 & 0 \\ 0 & 0 & 10^{-4} & 0 \\ 0 & 0 & 0 & 10^{-4} \end{pmatrix}$$

Ο πίνακας λαθών μετρήσεων υπολογίστηκε με βάση τις μετρήσεις του αλγορίθμου και τις πραγματικές τιμές γωνιών περιστροφής (διαιρεμένες με τον παράγοντα κλιμάκωσης που αναφέρθηκε προηγουμένως). Έτσι, στα πειράματα, θεωρήθηκε το μέγιστο των μεταβλητοτήτων σφάλματος μετρήσεων (αυτό έδωσε τα καλύτερα αποτελέσματα) για τις δύο παρατηρούμενες καταστάσεις, και ο πίνακας που θεωρήθηκε ήταν ο εξής:

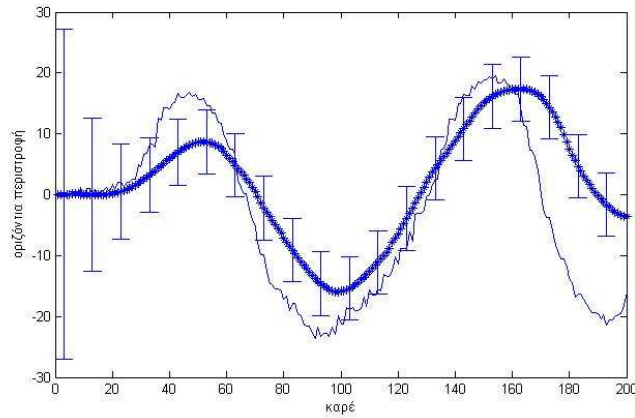
$$R = \begin{pmatrix} 2.38 & 0 \\ 0 & 0.46 \end{pmatrix}$$

Παραδείγματα γωνιών περιστροφής, οι πραγματικές τους τιμές και οι μετρούμενες, φαίνονται στο σχήμα 3.15. Στις γραφικές παραστάσεις είναι σημειωμένες με κάθετες γραμμές και οι τιμές της μεταβλητότητας εκτίμησης σε κάποια καρέ. γίνετα προφανές ότι, όσο η ακολουθία προχωράει, η αντίστοιχη μεταβλητότητα σταθεροποιείται.

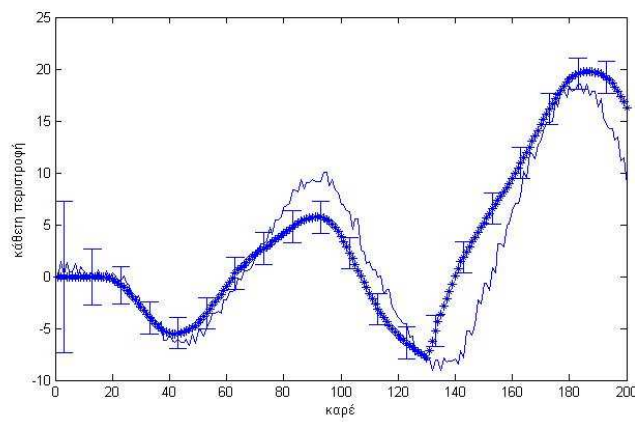
3.6 Πειράματικά αποτελέσματα

3.6.1 Πειράματα στην XM2VTS

Τα παραπάνω επαληθεύτηκαν στη βάση XM2VTS [104], η οποία αποτελεί ένα ευρέως διαδεδομένο σύνολο δεδομένων οπτικοακουστικού χαρακτήρα, για τον έλεγχο μίας αρκετά μεγάλης γκάμας αλγορίθμων αναγνώρισης φωνής, εντοπισμού χαρακτηριστικών προσώπου, περιστροφής κεφαλιού, κλπ. Μεταξύ άλλων, η XM2VTS εμπεριέχει



α



β

Σχήμα 3.15: (α) Οριζόντια περιστροφή κεφαλιού, (β) Κατακόρυφη περιστροφή κεφαλιού: πραγματικές, εκτιμούμενες τιμές

ακολουθίες εικόνων από 295 άτομα. Κάθε άτομο, ξεκινώντας από την θέση που κοιτάζει την κάμερα χωρίς να έχει στραμμένο το κεφάλι κατά κανέναν τρόπο (μετωπική στάση), γυρίζει προς τα αριστερά του, ύστερα προς τα δεξιά του, κατόπιν, αφού έρθει στην αρχική θέση, στρέφει το κεφάλι προς τα πάνω και, ακολούθως, προς τα κάτω. Κάθε ακολουθία τερματίζει με τον χρήστη να επιστρέφει στην αρχική του θέση, δηλαδή να κοιτάζει την κάμερα ευθέως. Για υπολογισμό των αποτελεσμάτων, όλες οι ακολουθίες υπέστησαν μία κλιμάκωση κατά 0,5, δηλαδή, από 720×480 εικονοστοιχεία, χρησιμοποιήθηκε διάσταση 360×240 . Αξιίζει να τονιστεί ότι στη βάση δεν παρέχονται δεδομένα αναφοράς σχετικά με τις γωνίες περιστροφής των ατόμων κατά τη διάρκεια των ακολουθιών.

3.6.1.1 Επιτυχία στην αναγνώριση περιστροφής κεφαλιού

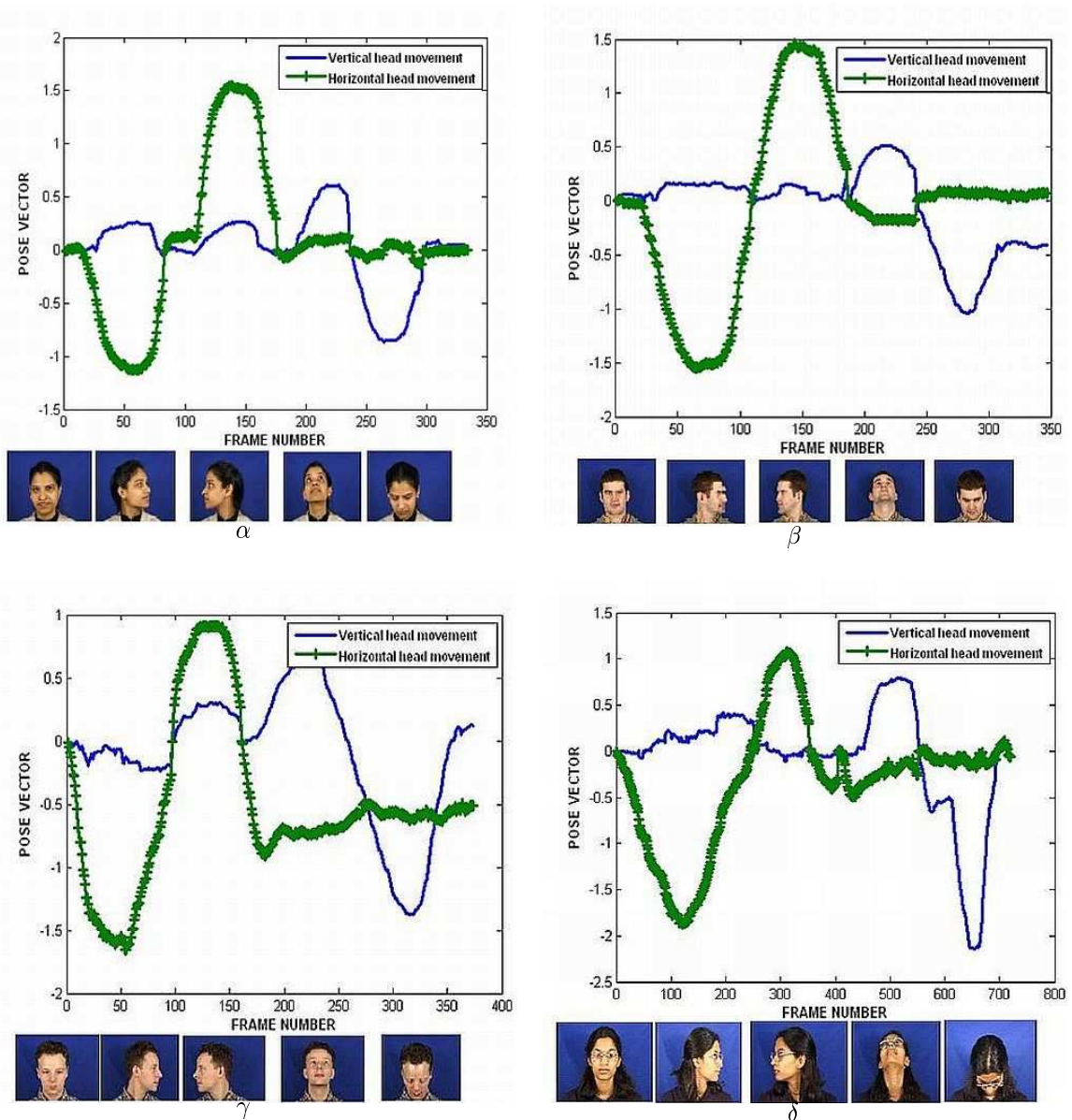
Η XM2VTS, όπως αναφέρθηκε και προηγουμένως, δεν παρέχει δεδομένα αναφοράς με συγκεκριμένες γωνίες, για τη σύγκριση των εκτιμώμενων γωνιών περιστροφής. Ωστόσο, έγινε μία προσπάθεια για μία ποιοτική, έστω, εκτίμηση του αλγορίθμου στη συγκεκριμένη βάση. Έτσι, υπολογίστηκαν τα διανύσματα περιστροφής κεφαλιού (σχήματα 3.16). Για τις οριζόντιες κινήσεις, δεξιές (αριστερές) περιστροφές αντιστοιχίστηκαν σε αρνητικά (θετικά) πρόσημα της συντεταγμένης p_y του διανύσματος περιστροφής, ενώ για κάθετες κινήσεις, περιστροφές του κεφαλιού προς τα πάνω (κάτω), αντιστοιχίστηκαν σε θετικές (αρνητικές) τιμές της συντεταγμένης p_x του διανύσματος περιστροφής του κεφαλιού. Ο συνολικός αριθμός ακολουθιών που χρησιμοποιήθηκαν για να ελεγχθεί η συγκεκριμένη προσέγγιση στην εκτίμηση περιστροφής κεφαλιού, ήταν 1180 (4 ακολουθίες για κάθε ένα από τα 295 άτομα που συμμετείχαν στη βάση), ενώ σημειώθηκαν χειροκίνητα τα καρέ εκείνα κατά τα οποία ή κάθε μία από τις τέσσερις περιστροφές φτάνει στο ακρότατό της. Για τη συγκεκριμένη βάση, η εκτίμηση θεωρείται επιτυχής αν, με τα κατάλληλα κατώφλια στις συντεταγμένες των διανυσμάτων περιστροφής του κεφαλιού, είναι δυνατός ο προσδιορισμός των καρέ εκείνων στα οποία η εκάστοτε περιστροφή του κεφαλιού βρίσκεται στο ακρότατο σημείο της. Έτσι, για τις οριζόντιες περιστροφές, θεωρήθηκε ότι, αν η συνιστώσα p_y πάρει απόλυτη τιμή μεγαλύτερη από το 90% της απόστασης των κέντρων των ματιών (όπως υπολογίστηκε στο καρέ του εντοπισμού), η οριζόντια περιστροφή είναι πλήρης. Αντίστοιχα, για την κάθετη συνιστώσα της περιστροφής του κεφαλιού, αυτό το ποσοστό τέθηκε στο 45%. Τα παραπάνω κατώφλια έδωσαν τα αποτελέσματα που περιγράφονται στον πίνακα 3.2, όπου αναφέρονται τα αποτελέσματα σε ανάκτηση και ακρίβεια, καθώς και το f -measure για κάθε περιστροφή. Επιτυχής προσδιορισμός του χρονικού σημείου περιστροφής θεωρήθηκε όταν, οι συνθήκες που περιγράφονται παραπάνω για την κάθε κίνηση, ικανοποιούνται σε χρονικά διαστήματα $\pm 1/4$ περιστροφής (≈ 25 καρέ) από το αντίστοιχο ακρότατο σημείο των δεδομένων αναφοράς.

Στο σχήμα 3.16, φαίνονται παραδείγματα τιμών που παίρνουν κάθετες και οριζόντιες συνιστώσες του διανύσματος περιστροφής, καθώς και τα αντίστοιχα καρέ της αντίστοιχης ακολουθίας.

Στον πίνακα 3.3 δίνονται και οι απαιτούμενοι χρόνοι για παρακολούθηση των χαρακτηριστικών για διάφορους παράγοντες κλιμάκωσης, όπως μετρήθηκαν σε υπολογιστή 3.2Ghz.

Πίνακας 3.2: Ακρίβεια-Ανάκτηση- f -measure για τις τέσσερις ακραίες καταστάσεις περιστροφής κεφαλιού

	αριστερή περιστροφή	δεξιά περιστροφή	προς τα επάνω περιστροφή	προς τα κάτω περιστροφή
ακρίβεια	0.958	0.978	0.860	0.860
ανάκτηση	0.899	0.922	0.948	0.948
f -measure	0.924	0.946	0.901	0.901



Σχήμα 3.16: Παραδείγματα τιμών οριζόντιας και κάθετης συνιστώσας διανύσματος περιστροφής κεφαλιού και χαρακτηριστικά καρέ της ακολουθίας

Πίνακας 3.3: Χρόνοι παρακολούθησης χαρακτηριστικών ανά καρτέ

εικονοστοιχεία	Μέσος απαιτούμενος χρόνος για παρακολούθηση ανά καρτέ
255×192	6.5ms
360×240	10ms
432×288	15ms
504×336	19.5ms
576×384	27.5ms
648×432	36ms
720×576	84ms

Πίνακας 3.4: Μέσο Απόλυτο Σφάλμα στην εκτίμηση οριζόντιας (yaw), κατακόρυφης (pitch) και γύρω από τον κάθετο στο επίπεδο της κάμερας περιστροφή (roll) στην βάση του Πανεπιστημίου της Βοστόνης, για μεμονωμένους χρήστες

χρήστης	οριζ. περιστροφή	κατακ. περιστροφή	περιστροφή στο επίπεδο της εικόνας
jam	7.43°	3.80°	4.73°
jim	9.15°	3.94°	5.37°
llm	10.5°	7.22°	4.73°
ssm	6.63°	4.71°	7.80°
vam	8.01°	5.69°	7.62°
Σύνολο	8.35°	5.07°	6.05°

3.6.2 Πειράματα στην βάση του Πανεπιστημίου της Βοστόνης

Για να ελεγχθεί η ικανότητα του συστήματος να υπολογίσει με αξιοπιστία τις τιμές των γωνιών περιστροφής του κεφαλιού, έγιναν πειράματα στη βάση του Πανεπιστημίου της Βοστόνης [22]. Στη βάση παρέχονται δεδομένα αναφοράς για περιστροφή του κεφαλιού γύρω από τον οριζόντιο και τον κάθετο νοητό άξονα του κεφαλιού, αλλά και τη γωνία που σχηματίζεται όταν γίνεται περιστροφή του κεφαλιού γύρω από τον άξονα τον κάθετο στο επίπεδο της εικόνας. Τα 5 άτομα που συμμετείχαν στη δημιουργία της βάσης πραγματοποιούν φυσικές και γρήγορες κινήσεις, τόσο περιστροφικές, όσο μεταφορικές. Στον πίνακα 3.4 φαίνονται τα αποτελέσματα για κάθε χρήστη ξεχωριστά. Για την αντιστοίχιση του διανύσματος περιστροφής με τις γωνίες που δίνονται από τη βάση σαν δεδομένα αναφοράς έγινε πολλαπλασιασμός με έναν παράγοντα, κοινό για όλες τις ακολουθίες.

3.7 Συμπεράσματα

Στο κεφάλαιο αυτό δόθηκε μία περιγραφή των σχετικών εργασιών γύρω από θέματα εκτίμησης περιστροφής προσώπου. Δόθηκε έμφαση κυρίως σε συστήματα μιας μη-

χανής λήψης, χωρίς ιδιαίτερες απαιτήσεις σε εξοπλισμό ή γνώσεις παραμέτρων κάμερας ή περιβάλλοντος. Ακολούθως, παρουσιάστηκαν τα κυριότερα σύνολα εικόνων/ακολουθιών που χρησιμοποιούνται διεθνώς από ερευνητικές ομάδες σε θέματα αναγνώρισης κατεύθυνσης βλέμματος χρήστη, και μία βάση που αναπτύχθηκε στο πλαίσιο της παρούσας διατριβής. Η βάση αυτή αναπτύχθηκε σε μη επιτηδευμένο περιβάλλον, ενώ, για πρώτη φορά στη βιβλιογραφία παρουσιάστηκε σύνολο δεδομένων ακολουθιών καρτέ που ενσωματώνουν πληροφορία, τόσο από περιστροφή προσώπου, όσο κι από κατευθυντικότητα βλέμματος. Καθώς απώτερος στόχος είναι να εξεταστεί το κατά πόσο μη παρεμβατικές μέθοδοι εξαγωγής βλέμματος χρήστη μπορούν να ενσωματωθούν σε εφαρμογές επικοινωνίας ανθρώπου-μηχανής, στο σημείο αυτό προτάθηκε μία μεθοδολογία ανάκτησης των τριών βαθμών ελευθερίας της περιστροφής του κεφαλιού. Δεν δίνεται ιδιαίτερη έμφαση στην ακρίβεια σε μοίρες, αλλά στην ευρωστία του συστήματος σε μεγάλες ακολουθίες εικόνων, όπως επίσης στην ικανότητα του συστήματος να αντιστοιχίζεται σε περιοχές ενδιαφέροντος στην οθόνη υπολογιστή και εκτός αυτής, κάτι το οποίο θα φανεί σε επόμενο κεφάλαιο.

Κεφάλαιο 4

Σύντηξη τοπικής και ολιστικής πληροφορίας για εκτίμηση περιστροφής κεφαλιού

Στο κεφάλαιο αυτό μελετάται ο υπολογισμός των τριών βαθμών περιστροφής κεφαλιού με χρήση σύντηξης τοπικής και ολιστικής πληροφορίας. Τα χαρακτηριστικά του προσώπου (τοπική πληροφορία) παρακολουθούνται, αφού πρώτα γίνει ο εντοπισμός τους ακολουθώντας τη μεθοδολογία που αναπτύσσεται στο κεφάλαιο 2, ενώ μελετάται η χρήση χαρακτηριστικών καρτέ για την ανάκτηση σωστών παραμέτρων περιστροφής σε περιπτώσεις απόκλισης. Για την παρακολούθηση των χαρακτηριστικών του προσώπου, χρησιμοποιούνται Διανυσματικά Πεδία Απόστασης, με σκοπό να αξιοποιηθεί περισσότερο τοπολογική πληροφορία, προερχόμενη από κάθε εικονοστοιχείο, παρά πληροφορία φωτεινότητας. Κατόπιν, γίνεται χρήση ολιστικής πληροφορίας, η οποία αντιστοιχεί στην περιοχή του προσώπου, όπως εξάγεται από τον αλγόριθμο παρακολούθησής του. Για την εξαγωγή της γωνίας του προσώπου με χρήση ολιστικής πληροφορίας χρησιμοποιούνται Συνελικτικά Νευρωνικά Δίκτυα, τα οποία, χάρη στην αρχιτεκτονική τους, επιτρέπουν τη γρήγορη εκμάθηση, χωρίς την ανάγκη πολλών δεδομένων εκπαίδευσης. Επιπλέον, τα Συνελικτικά Νευρωνικά Δίκτυα (ΣΝΔ) (Convolutional Neural Networks - CNNs) παρουσιάζουν ευρωστία σε μη απόλυτη ευθυγράμμιση ανάμεσα στα δεδομένα εκπαίδευσης και τα δεδομένα εισόδου. Τα δύο επίπεδα πληροφορίας συντίθενται με σταθμισμένο μέσο και, μέσα από μετρήσεις, φαίνεται ότι το συνολικό αποτέλεσμα βελτιώνεται σε σχέση με αυτό της κάθε ροής πληροφορίας ξεχωριστά. Τέλος, μελετάται η σύντηξη των δύο ροών πληροφορίας, με χρήση Μπεϋζιανών δικτύων, στα οποία έγινε χρήση των μετρήσεων, καθώς και εκτιμήσεων της αξιοπιστίας τους. Μέσα από πειράματα για υπολογισμό της οριζόντιας περιστροφής, φαίνεται η βελτίωση που παρέχεται μέσα από τη χρήση μπεϋζιανής λογικής, δεδομένου ότι μοντελοποιούνται οι συνθήκες εκείνες κάτω από τις οποίες υπερέρχει η μία έναντι της άλλης ροής, ανάλογα με το περιεχόμενο. Παρουσιάζονται συγκριτικά δεδομένα, ούτως ώστε να διαφανεί ότι, πέρα από την ευρωστία της προτεινόμενης μεθόδου σε διάφορες συνθήκες, επιτυγχάνεται και μεγάλος βαθμός ακρίβειας, σε σχέση με πρόσφατες μεθόδους στη βιβλιογραφία, που ασχολούνται με το θέμα της ανάκτησης περιστροφής κεφαλιού.

4.1 Ερευνητικό πλαίσιο

Πέρα από τις μεθόδους εκτίμησης περιστροφής κεφαλιού που παρουσιάστηκαν στο κεφάλαιο 3, στη βιβλιογραφία συναντώνται, ακόμη, πληθώρα συνδυαστικών μεθόδων, σε μια προσπάθεια κάλυψης των μειονεκτημάτων της μιας μεθόδου, με τα αντίστοιχα πλεονεκτήματα που προσφέρονται από την άλλη.

Στο άρθρο των Sung *et al* [158], οι συγγραφείς συνδυάζουν τα AAMs (Μοντέλα Ενεργού Εμφάνισης) με κυλινδρικά μοντέλα κεφαλιού (Cylinder Head Models - CHMs) [185], σε μια προσπάθεια να συνδυαστεί ο τοπικός χαρακτήρας των AAMs, με τις ιδιότητες συνολικής κίνησης που προκύπτουν από τα κυλινδρικά μοντέλα. Αρχικά, μετά τον εντοπισμό του προσώπου, εφαρμόζεται αναζήτηση ενός 2D+3D AAM μοντέλου [184] και οι παράμετροι που προκύπτουν από την εφαρμογή του μοντέλου αποτελούν είσοδο σε μια αρχική εκτίμηση της εφαρμογής του κυλινδρικού μοντέλου, το οποίο, εν συνεχεία, ενημερώνει τις παραμέτρους συνολικής κίνησης. Με βάση αυτό, τα AAMs εφαρμόζονται στο ακόλουθο καρέ. Η προσέγγιση AAM+CHM ελέγχθηκε σε δύο βάσεις εικόνων και αποδείχτηκε ότι αποδίδει καλύτερα σε σχέση με την 2D+3D AAM προσέγγιση. Τα ποσοστά επιτυχίας παρακολούθησης βελτιώνονται σε σχέση με τα AAMs. Ωστόσο, το «πεδίο κάλυψης πόζας» (το εύρος των γωνιών που μπορούν να εκτιμηθούν), δεν ξεπερνά τις 45° για οριζόντια γωνία.

Οι Morency *et al.* [110] χρησιμοποιούν μία τεχνική που ονόμασαν Γενικευμένα Προσαρμοστικά, εξαρτώμενα από την λήψη Μοντέλα Εμφάνισης (Generalized adaptive View-Based Appearance Model - GAVAM), τα οποία είναι μία επέκταση των AVAM [108]. Στη συγκεκριμένη δουλειά, μετά την κατάκτηση της περιοχής του προσώπου, η πόζα του εξάγεται με χρήση τριών βασικών ιδεών: Ένας στατικός εκτιμητής πόζας (με χρήση Haar χαρακτηριστικών), ένας διαφορικός αλγόριθμος παρακολούθησης (με χρήση του περιορισμού οπτικής ροής) και ένα σύνολο καρέ-κλειδιών, τα οποία ανανεώνονται κατά τη διάρκεια της ακολουθίας, με σκοπό να τροφοδοτούν συνεχώς μία βάση τέτοια, η οποία να καλύπτει χαρακτηριστικά καρέ διακριτών κινήσεων του κεφαλιού. Η μέθοδος εξετάστηκε στην βάση του Πανεπιστημίου της Βοστώνης [22]. Ο αλγόριθμος διαφορικής παρακολούθησης κίνησης, χρησιμοποιεί τον τρισδιάστατο αλγόριθμο Lucas-Kanade, και γι' αυτό, τα αποτελέσματα της προτεινόμενης μεθοδολογίας αναμένεται να εξαρτώνται από τις παραμέτρους της κάμερας. Οι συγγραφείς θέτουν κάποια τιμή για την εστιακή απόσταση και προκύπτουν ικανοποιητικά αποτελέσματα. Ωστόσο, δεν παρουσιάζεται κάποια μελέτη, η οποία να πιστοποιεί πως τα αποτελέσματα θα ήταν εξίσου καλά και για διαφορετικές τιμές της εστιακής απόστασης.

Στο [148], το πρόσωπο αρχικά εντοπίζεται με τη χρήση Μηχανών Διανυσμάτων Υποστήριξης (Support Vector Machines) σε υποχώρους της εικόνας (οι οποίοι εκπαιδεύτηκαν με Ανάλυση Ανεξάρτητων Συνιστωσών), σε περιοχές διαφόρων κλιμακώσεων και επικαλυπτόμενες. Η εκτίμηση της περιστροφής του κεφαλιού, γίνεται με τρισδιάστατο κύλινδρο, με βάση τον οποίο μοντελοποιείται το κεφάλι. Παρακολουθούνται οι συντεταγμένες συγκεκριμένων χαρακτηριστικών του κεφαλιού, και η κίνηση της περιστροφής γίνεται με αντιστοίχιση των θέσεων μεταξύ διαδοχικών καρέ, με τη χρήση συμπαγών μοντέλων κίνησης.

Στο [177], οι συγγραφείς χρησιμοποιούν ταξινομητές Viola-Jones, εκπαιδευμένους να ανιχνεύουν πρόσωπα σε διαφορετικές γωνίες περιστροφής. Επιπλέον, χρησιμοποιούν μία μέθοδο ανίχνευσης χαρακτηριστικών προσώπου (μάτια, μύτη, στόμα) και, με βάση τις θέσεις τους, απομονώνουν την περιοχή του προσώπου και την χρησιμο-

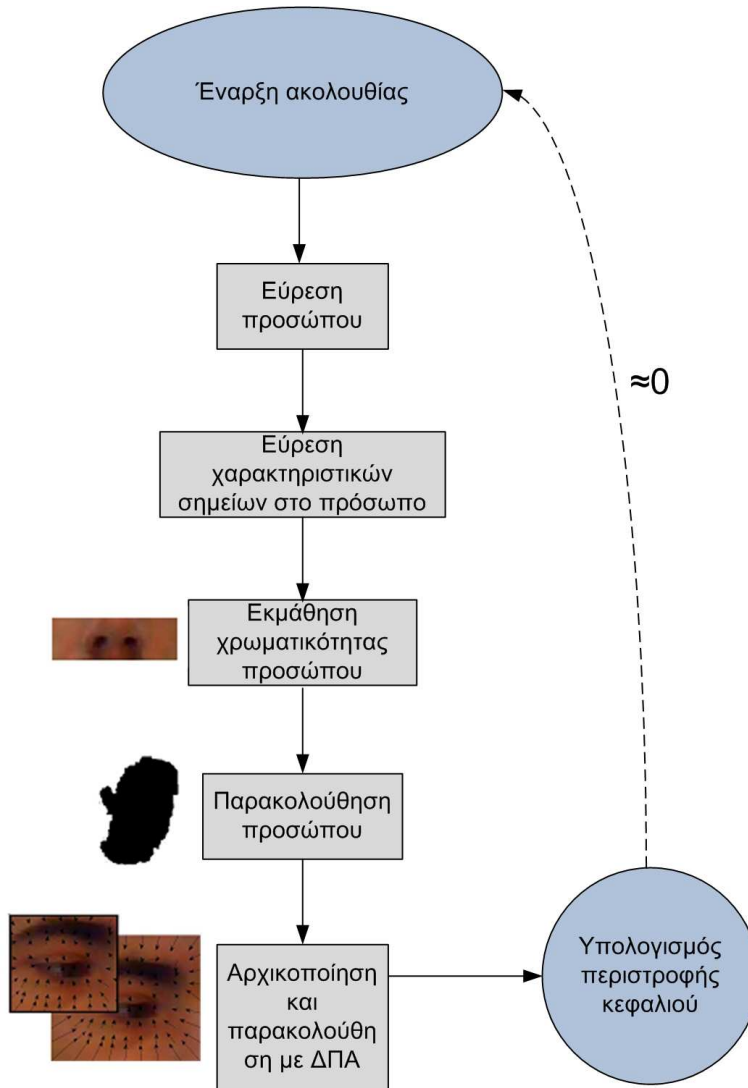
ποιούν ως είσοδο στους ταξινομητές. Για εκπαίδευση των ταξινομητών, οι συγγραφείς χρησιμοποίησαν πρωτόκολλα ενός-εναντίον-όλων και ενός-εναντίον-ενός και, μαζί με τις θέσεις των χαρακτηριστικών, δημιούργησαν μοντέλα παλινδρόμησης, από τα οποία εξάγουν τις τελικές γωνίες περιστροφής του προσώπου. Για εκπαίδευση χρησιμοποίησαν δεδομένα που αντιστοιχούν σε περιστροφές -45° έως 45° , τόσο για κάθετη, όσο και για οριζόντια περιστροφή, ενώ τα δεδομένα που χρησιμοποιήθηκαν ήταν τυχαίες εικόνες από το διαδίκτυο, έτσι ώστε να ενσωματωθούν κατά το δυνατό περισσότερες συνθήκες φωτισμού. Το μέσο τετραγωνικό σφάλμα που προέκυψε ήταν 5.82° , 5.65° και 2.96° για οριζόντια, κάθετη και παράλληλη με το επίπεδο της εικόνας περιστροφή, αντίστοιχα. Ο έλεγχος της απόδοσης της μεθόδου έγινε στην βάση που δημιουργήθηκε από τις παραπάνω εικόνες ¹.

Στο [182], προτείνεται μία τεχνική δύο σταδίων. Το σύστημα λειτουργεί σε στατικές εικόνες και χρησιμοποιεί την τεχνική Viola-Jones [171] για εύρεση προσώπου, σαν βήμα αρχικοποίησης. Για το σκοπό αυτό, έγινε εκπαίδευση εννιά μοντέλων σε εννιά διαφορετικές πόζες, ενώ το μοντέλο για την εύρεση του προσώπου επιλέγεται αρχικά χειροκίνητα. Το πρώτο στάδιο της μεθόδου συνίσταται στο μετασχηματισμό του προσώπου στο χώρο των Gabor κυματιδίων, σε διαφορετικές αναλύσεις. Ακολούθως, εξετάστηκαν προβολές των μετασχηματισμών σε υποχώρους PDA, FDA, KPCA και KDA. Λόγω της μη γραμμικής φύσης του προβλήματος, η μέθοδος KDA κρίθηκε ως η πλέον κατάλληλη. Σε όλα τα επίπεδα ανάλυσης της εικόνας, χρησιμοποιήθηκε πρωτόκολλο πλειοψηφίας και έτσι εκτιμήθηκε ένα αρχικό εύρος 15° μέσα στο οποίο αναμενόταν να κινείται η υπό εκτίμηση πόζα. Με χρήση γκαουσιανού πυρήνα, αυτό το βήμα έδωσε επιτυχία 94%. Πιο αναλυτική αναζήτηση της πόζας, μέσα στα όρια της αρχικής εκτίμησης έγινε με χρήση ελαστικών γράφων [180]. Οι γράφοι μπορούν να προσαρμοστούν μόνο τοπικά, ενώ για ένα κβαντισμένο χώρο πόζας (-90° έως 90° για οριζόντια και -45° έως 60° για κάθετη) σε 86 κλάσεις, το σύστημα είχε επιτυχία 75,4% σε μια βάση 28 ατόμων.

4.2 Χρήση τοπικής πληροφορίας με Διανυσματικά Πεδία Απόστασης

Αρχικά, εντοπίζεται το πρόσωπο, και μία περιοχή δέρματος μέσα σε αυτό χρησιμοποιείται ως δείγμα για την παρακολούθησή του σε μετέπειτα καρέ. Η χρήση αλγορίθμου εντοπισμού προσώπου στην έναρξη του συστήματος, εξασφαλίζει ευρωστία σε αλλαγές κλιμάκωσης, καθώς το πρόσωπο μπορεί να εντοπιστεί σε διάφορες κλίμακες, ενώ η παρακολούθησή του γίνεται βάσει χρωματικότητας, προσαρμοζόμενη στον κάθε χρήστη ξεχωριστά. Περιοχές χαρακτηριστικών του προσώπου, παρακολουθούνται συγκρίνοντας τα ΔΠΑ (Διανυσματικά Πεδία Απόστασης) τους, σε διαδοχικά καρέ. Τα Διανυσματικά Πεδία Απόστασης των χαρακτηριστικών αναζητούνται μόνο μέσα στα όρια του προσώπου (επιτρέπεται μόνο το 50% της περιοχής τους να βρίσκεται εκτός ορίων, θεωρώντας ότι αυτό συμβαίνει σε εικόνες όπου το άτομο έχει μεγάλη οριζόντια γωνία περιστροφής). Βάσει των θέσεων των χαρακτηριστικών, υπολογίζεται η περιστροφή του προσώπου. Η σύνοψη του αλγορίθμου φαίνεται στο Σχήμα 4.1.

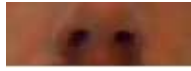
¹mplab.ucsd.edu



Σχήμα 4.1: Σύνοψη της μεθόδου εκτίμησης περιστροφής κεφαλιού με χρήση ΔΠΑ: Αρχικά εντοπίζεται το πρόσωπο και τα χαρακτηριστικά σημεία σε αυτό. Ακολούθως, παρακολουθείται η περιοχή του προσώπου και, μέσα σε αυτή, τα Διανυσματικά Πεδία Απόστασης χαρακτηριστικών περιοχών. Τα όρια της περιοχής του προσώπου, σε σχέση με τα χαρακτηριστικά σημεία χρησιμοποιούνται για να υπολογιστεί το διάνυσμα περιστροφής του προσώπου. Εάν το μέτρο του διανύσματος είναι μικρότερο από κάποιο μέγεθος, τότε ο αλγόριθμος επανεκκινείται.

4.2.1 Παρακολούθηση προσώπου με προσαρμοστική κατωφλίωση

Οι ιδέες που συζητήθηκαν στο κεφάλαιο που αναλύει τον εντοπισμό χαρακτηριστικών σημείων στο πρόσωπο, θα χρησιμοποιηθούν για την παρακολούθηση των ανιχνευμένων χαρακτηριστικών προσώπου. Για την εφαρμογή περιορισμών στις θέσεις των χαρακτηριστικών προσώπου, σε κάθε καρέ, χρησιμοποιούνται τα όρια του προσώπου, και τα παρακολουθούμενα χαρακτηριστικά περιορίζονται μέσα σε αυτή την περιοχή. Για το σκοπό αυτό, κάθε φορά που ανιχνεύεται ένα πρόσωπο, εξάγεται μία περιοχή C_{skin} στο κέντρο (Εικόνα 4.2). Η περιοχή αυτή, κατόπιν, χρησιμεύει για τη μοντελοποίηση της χρωματικότητας του δέρματος του εκάστοτε χρήστη, με σκοπό την μετέπειτα επιτυχή παρακολούθηση του προσώπου του. Τα πειράματα έδειξαν ότι το κανάλι του κορεσμού (saturation) μπορεί να αποδειχθεί ικανό στο να ξεχωρίζει το πρόσωπο από το φόντο². Οι τιμές του κορεσμού εικονοστοιχείων υποψήφιων να ανήκουν στο πρόσωπο, αναμένεται να είναι κοντά στη μέση τιμή του s_M της περιοχής C_{skin} . Με αυτόν τον τρόπο, δημιουργείται μία δυαδική εικόνα C_{fp} (Εξίσωση 4.1)



Σχήμα 4.2: Περιοχή που χρησιμοποιείται για την εξαγωγή πληροφορίας χρωματικότητας

$$C_{fp} = \{x \in \Omega : \|s_M - s_x\| < T\} \quad (4.1)$$

όπου Ω είναι το σύνολο όλων των εικονοστοιχείων που ανήκουν στο καρέ, x είναι τα υποψήφια να ανήκουν στην περιοχή του προσώπου εικονοστοιχεία, s_x η τιμή κορεσμού τους και T ένα κατώφλι, το οποίο, όσο μεγαλύτερο είναι, τόσο περισσότερα εικονοστοιχεία θα μπουν στην κατηγορία «πρόσωπο», με αποτέλεσμα μεγάλες περιοχές του φόντου να θεωρούνται πρόσωπο, ενώ, όσο μικρότερο είναι, τόσο λιγότερα εικονοστοιχεία θα θεωρηθούν ότι ανήκουν στο πρόσωπο (σχεδόν όλα μέσα στην περιοχή του προσώπου, όμως πολλά εικονοστοιχεία που ανήκουν στο πρόσωπο, κατηγοριοποιούνται διαφορετικά). Η εύρεση ενός προσαρμοστικού κατωφλίου θα συζητηθεί παρακάτω. Μετά την εξαγωγή της δυαδικής εικόνας C_{fp} , ακολουθεί δυαδικό άνοιγμα, ώστε να απομακρυνθούν μικρές περιοχές, που λανθασμένα αποδίδονται στην περιοχή του προσώπου.

Το κατώφλι T επιλέγεται αυτόματα, για κάθε χρήστη, στο αρχικό βήμα, του εντοπισμού του προσώπου (υποκεφάλαιο 2.2.1), σύμφωνα με τις εξισώσεις 4.2 και 4.3:

$$T = \arg \min_{0.05 < T < 0.35} \left(\sum_{x \in \Omega} \delta(k_x) - Face_{size} \right) \quad (4.2)$$

με

$$k_x = \begin{cases} 0, & \|s_M - s_x\| \leq T \\ 1, & \|s_M - s_x\| > T \end{cases} \quad (4.3)$$

με δ την συνάρτηση Kronecker delta και $Face_{size}$ το μέγεθος της περιοχής του κεφαλιού, όπως ορίζεται από την έλλειψη που περιέχει το πρόσωπο στο βήμα του εντοπισμού. Η παραπάνω διαδικασία έχει σαν αποτέλεσμα την επιλογή του κατωφλίου αυτόματα, για κάθε χρήστη, συνθήκες φωτισμού, και μέγεθος κεφαλιού σε

²μπορούν να χρησιμοποιηθούν και διαφορετικά κανάλια ή συνδυασμοί αυτών

σχέση με το πεδίο λήψης της κάμερας. Συνεπώς, το σύστημα διευκολύνεται στο να προσαρμόζεται σε οποιοσδήποτε συνθήκες σε σχέση με το φωτισμό και τη θέση του χρήστη. Σύμφωνα με τις εξισώσεις 4.2 και 4.3, το T επιλέγεται με βάση την εξής υπόθεση: Στο πρώτο καρέ, αναμένεται ότι ο αριθμός των εικονοστοιχείων με τιμές κορεσμού κοντά στο μέσο όρο αυτών της περιοχής C_{skin} , είναι κοντά στον αριθμό των εικονοστοιχείων της ανιχνευμένης περιοχής του προσώπου. Η παραπάνω υπόθεση φαίνεται εποπτικά στο σχήμα 4.3, όπου το βέλτιστο κατώφλι T που θα χρησιμοποιηθεί στην εξίσωση 4.1, βασίζεται στο μέγεθος της περιοχής του προσώπου στην αρχικοποίηση. Με σκοπό την παραπέρα μείωση των υποψήφίων να ανήκουν στο πρόσωπο εικονοστοιχείων, μπορούν να χρησιμοποιηθούν και οι κανόνες που περιγράφονται στο [84], έτσι ώστε να δημιουργηθούν συστάδες στο χρωματικό χώρο RGB . Σύμφωνα με τους συγγραφείς στο [84], ένας δυαδικός χάρτης C_{sp} , υποψήφίων να ανήκουν σε δερματική περιοχή εικονοστοιχείων, χτίζεται με βάση τους παρακάτω κανόνες:

$$\begin{aligned}
 C_{sp} = \{x \in \Omega : \\
 R(x) > 95 \text{ AND } G(x) > 40 \text{ AND } B(x) > 20 \\
 \text{AND } \max(R(x), G(x), B(x)) - \min(R(x), G(x), B(x)) > 15 \text{ AND} \\
 \|R(x) - G(x)\| > 15 \text{ AND } R(x) > G(x) \text{ AND } R(x) > B(x)\} \quad (4.4)
 \end{aligned}$$

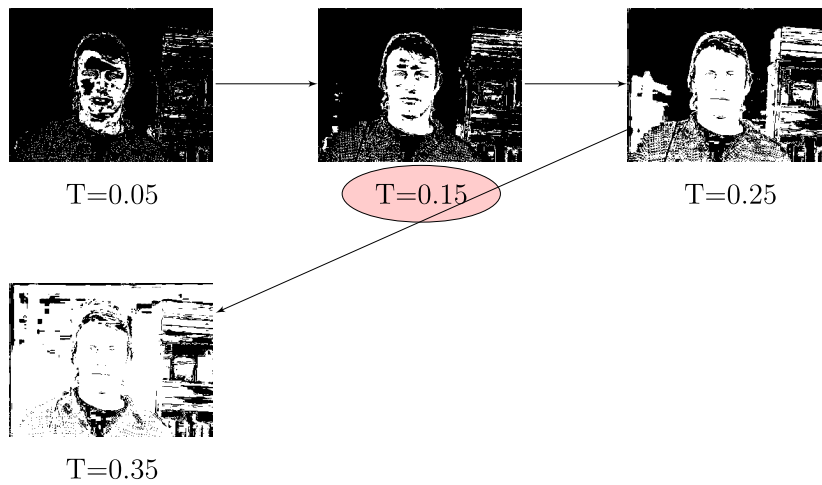
με R, G και B τα τρία βασικά κανάλια χρωματικότητας. Οι δυαδικοί χάρτες C_{fp} και C_{sp} , ακολούθως, συνδυάζονται με χρήση του λογικού AND , ενώ ακολουθεί δυαδικό κλείσιμο (με χρήση δομικού στοιχείου 10×10 , δηλαδή το 0.13% του μεγέθους του καρέ των εικόνων στις οποίες έγιναν πειράματα). Το παραπάνω βήμα, γίνεται με σκοπό να απομακρυνθούν μικρές περιοχές, όπως τα μάτια. Τέλος, στην προτεινόμενη μέθοδο, χρησιμοποιείται επισηματοθέτηση συνεκτικών περιοχών (connected component labelling) [58] και επιλέγεται η μεγαλύτερη περιοχή ως αυτή που θα αντιπροσωπεύει τελικά το κεφάλι. Συνοπτικά, τα βήματα που χρησιμοποιούνται για την παρακολούθηση προσώπου, φαίνονται στις εικόνες 4.3 και 4.4.

4.2.2 Παρακολούθηση χαρακτηριστικών προσώπου με χρήση Διανυσματικών Πεδίων Απόστασης

Με αφετηρία τον εντοπισμό των περιοχών του προσώπου, των ματιών και του στόματος, ξεκινά η παρακολούθηση και των χαρακτηριστικών προσώπου, με χρήση των Διανυσματικών Πεδίων Απόστασης (ΔΠΑ) (Distance Vector Field - DVF).

4.2.2.1 Παρακολούθηση ματιών

Για κάθε μάτι, υπολογίζεται το αντίστοιχο ΔΠΑ $f_{k,i}$ της περιοχής του R_k , στο καρέ k και τη θέση του i . Για το επόμενο καρέ $k+1$, υπολογίζεται εκείνη η νέα θέση $i+p$ του αντίστοιχου ΔΠΑ που ελαχιστοποιεί την απόσταση L_2 . Η περιοχή R_s στην οποία υπολογίζονται υποψήφια υποπεριοχές στο καρέ $k+1$ είναι επεκτεταμένη σε σχέση με την περιοχή R_k του χαρακτηριστικού στο καρέ k (εδώ, έγινε επέκταση κατά 33.3% του ύψους και του πλάτους της R_k , στον οριζόντιο και στον κάθετο άξονα, αντίστοιχα). Το διάνυσμα κίνησης \mathbf{p} , που θα δώσει τη νέα θέση της περιοχής του ματιού, δίνεται από την παρακάτω σχέση:



Σχήμα 4.3: Επισκόπηση της διαδικασίας επιλογής του κατωφλίου T , για κατηγοριοποίηση των εικονοστοιχείων, με βάση την εξίσωση 4.1: για τη συγκεκριμένη ακολουθία, αποφασίστηκε κατωφλίωση στο $T = 0.15$, αφού ο συνολικός αριθμός των εικονοστοιχείων, οι τιμές κορεσμού των οποίων, είναι κοντά σε αυτήν της αρχικά επιλεγμένης περιοχής δέρματος, προσεγγίζουν τον πραγματικό αριθμό των εικονοστοιχείων του προσώπου.

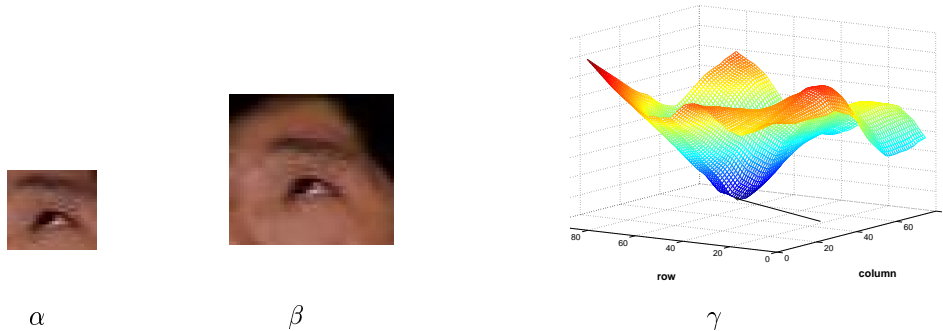


Σχήμα 4.4: (α) τιμές κορεσμού αρχικής εικόνας (β) κατωφλιωμένες τιμές κορεσμού (γ) Υποψήφια να ανήκουν σε περιοχή προσώπου εικονοστοιχεία C_{fp} , όπως εξήχθησαν μετά από μορφολογικό άνοιγμα (δ) Υποψήφια να ανήκουν σε δερματική περιοχή εικονοστοιχεία C_{sp} ; (ε) Υποψήφια να ανήκουν σε περιοχή προσώπου εικονοστοιχεία μετά το λογικό AND μεταξύ C_{fp} και C_{sp} ; (στ) τελική μάσκα προσώπου, μετά τους μορφολογικούς μετασχηματισμούς.

$$\mathbf{p} = \arg \min_{\mathbf{x}} \sum_{x \in R_s} \|f_{k,i} - f_{k+1,i+x}\|_2 \quad (4.5)$$

όπου R_s η περιοχή αναζήτησης.

Στην εικόνα 4.5 απεικονίζεται μία εντοπισμένη περιοχή ματιού στο καρέ k , η αντίστοιχη περιοχή αναζήτησης της νέας περιοχής του ματιού στο καρέ $k + 1$, και οι τιμές τη απόστασης L_2 του ΔΠΑ στο καρέ k σε σχέση με υποψήφιες περιοχές στα υπολογισμένα ΔΠΑ που αντιστοιχούν στο καρέ $k + 1$.



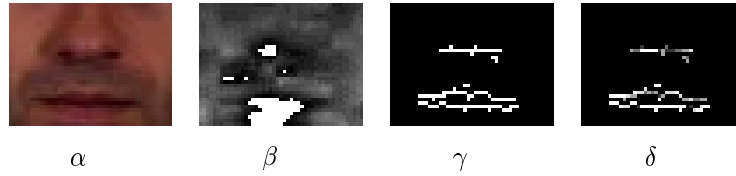
Σχήμα 4.5: (α) Περιοχή ματιού στο καρέ k , (β) Περιοχή αναζήτησης ματιού στο καρέ $k + 1$; (γ) ΔΠΑ σφάλμα μεταξύ (α) και (β)

Σε κάθε εντοπισμό περιοχής ματιού, υπολογίζεται η νέα θέση του κέντρου του ματιού μέσα σε αυτήν. Κατόπιν, η περιοχή του εύρους του ΔΠΑ «κεντράρεται» γύρω από το κέντρο του ματιού. Για την εύρεση του κέντρου του ματιού, ακολουθείται η ίδια διαδικασία με αυτήν που παρουσιάζεται στο κεφάλαιο 2. Τα πειράματα έδειξαν ότι, υιοθετώντας αυτό το βήμα, αποφεύγεται λανθασμένη παρακολούθηση των ματιών καθώς, ακόμη κι αν τα ΔΠΑ παρουσιάζουν μια τάση να ακολουθούν λανθασμένη διαδρομή, ο «εξαναγκασμός» να κεντράρονται γύρω από το κέντρο των ματιών, εξασφαλίζει την τοποθέτησή τους στο σωστό σημείο.

4.2.2.2 Παρακολούθηση στόματος

Η παρακολούθηση του στόματος, με την αποκλειστική χρησιμοποίηση του μετασχηματισμού ΔΠΑ δεν είναι εξίσου ακριβής όπως αυτή των ματιών. Αυτό οφείλεται στο ότι, αρκετές φορές, η χρωματικότητα των χειλιών δεν είναι αρκετά καλά διαφοροποιημένη σε σχέση με αυτή του δέρματος. Κάτι τέτοιο επιδεινώνεται σε περίπτωση που υπάρχει ομιλία, οπότε το περίγραμμα των χειλιών ενδέχεται να αλλάζει δραστικά μεταξύ διαδοχικών καρέ (και, συνεπώς να επηρεάζεται και το σχήμα που προκύπτει από τον μετασχηματισμό ΔΠΑ). Για τους παραπάνω λόγους, στην παρακολούθηση του στόματος, χρησιμοποιούνται τα εξής χαρακτηριστικά: Η χρωματικότητα (hue) και η συγκέντρωση οριζόντιων ακμών. Ο συνδυασμός των δύο χαρακτηριστικών γίνεται με πολλαπλασιασμό της δυαδικής εικόνας οριζόντιων ακμών με τις τιμές της χρωματικότητας στην περιοχή εντοπισμού (εικόνα 4.6)). Η περιοχή με τις μεγαλύτερες τιμές του προκύπτοντος συνδυασμού είναι και η καινούρια θέση της περιοχής του στόματος. Η μάσκα που χρησιμοποιείται προέρχεται από αυτήν που χρησιμοποιήθηκε στο βήμα του εντοπισμού, και με προεκβολή (η παρεμβολή) έρχεται σε διαστάσεις τέτοιες, ώστε να διατηρεί τις αναλογίες της σε σχέση με το πραγματικό μέγεθος του προσώπου (όπως αποτυπώνεται στην εικόνα), ενώ η περιοχή αναζήτησης προκύπτει όπως αυτές της

αναζήτησης θέσης ματιών. Μέσα από δοκιμές φάνηκε ότι το παραπάνω βήμα ήταν σημαντικό στη σωστή παρακολούθηση της θέσης του στόματος.



Σχήμα 4.6: (α) Περιοχή εντοπισμού στόματος (β) Κανάλι χρωματικότητας περιοχής εντοπισμού στόματος (γ) Πίνακας οριζόντιων ακμών περιοχής εντοπισμού στόματος (δ) Χρωματικότητα πολλαπλασιασμένη με πίνακα οριζόντιων ακμών

4.2.2.3 Βελτιστοποίηση της παρακολούθησης

Κατά τη διάρκεια της παρακολούθησης των χαρακτηριστικών, μπορεί να χρησιμοποιηθεί ένας ακόμη παράγοντας, ο οποίος μοντελοποιεί την τοπολογία των ματιών σε σχέση με το στόμα, για διάφορες πόζες. Στο σημείο αυτό, γίνεται η υπόθεση ότι ο λόγος μεταξύ της απόστασης ανάμεσα στα μάτια και της απόστασης του σημείου ανάμεσα στα μάτια με το στόμα ακολουθεί κανονική κατανομή $f(\mu, \sigma^2)$ με μ και σ ο μέσος όρος και η τυπική απόκλιση, αντίστοιχα. Επειδή, όμως, σε κάθε πρόσωπο οι αναλογίες διαφέρουν, σαν μέσος όρος εδώ θεωρήθηκε ο αντίστοιχος λόγος του κάθε ατόμου, όταν κοιτάζει την κάμερα μετωπικά. Η τυπική απόκλιση, ωστόσο, υπολογίστηκε από μία βάση εικόνων με γνωστές παραμέτρους περιστροφής και συντεταγμένες χαρακτηριστικών [51]. Ο παραπάνω εξτρά παράγοντας μετατρέπει την εξίσωση 4.5 ως ακολούθως:

$$\begin{aligned}
 \mathbf{p} &= \\
 &= \arg \min_{\mathbf{x}} \left(\sum_{i \in R_s} \|f_{k,i} - f_{k+1,i+x}\|_2 f(d_{k+1,x}; \mu, \sigma^2)^{-1} \right) \\
 &= \arg \min_{\mathbf{x}} \left(\sum_{i \in R_s} \|f_{k,i} - f_{k+1,i+x}\|_2 e^{\frac{(d_{k+1,x} - \mu)^2}{2\sigma^2}} \right) \quad (4.6)
 \end{aligned}$$

Στην παραπάνω εξίσωση, $d_{k+1,x}$ είναι ο λόγος της απόστασης ανάμεσα στα μάτια και αυτής του μέσου των ματιών σε σχέση με το στόμα, στο καρέ $k+1$, με μετατόπιση x του υπό παρακολούθηση χαρακτηριστικού σε σχέση με το καρέ k . Η παραπάνω εξίσωση χρησιμοποιείται όταν γίνεται παρακολούθηση του κάθε ματιού ξεχωριστά, ενώ σαν θέσεις των άλλων χαρακτηριστικών χρησιμοποιεί αυτές του προηγούμενου καρέ k για την εκτίμηση του όρου $d_{k+1,x}$.

□

4.2.3 Εκτίμηση οριζόντιας, κατακόρυφης και παράλληλης στο επίπεδο της κάμερας περιστροφής

Σύμφωνα με τα πειράματα του Wilson [179], οι θέσεις των χαρακτηριστικών σημείων του προσώπου σε σχέση με τα όρια του προσώπου παίζουν πολύ σημαντικό ρόλο στην ανθρώπινη αντίληψη της περιστροφής του κεφαλιού [51]. Γνωρίζοντας, για κάθε καρέ

k , τα όρια της περιοχής του προσώπου, το σημείο ανάμεσα στα μάτια $E_k = (E_{x,k}, E_{y,k})$ και το κέντρο του στόματος $M_k = (M_{x,k}, M_{y,k})$, η οριζόντια (κάθετη) περιστροφή, υπολογίζεται ως εξής: Μοντελοποιείται σαν τη μέση διαφορά στη σχετική απόσταση των σημείων E και M από τα κατακόρυφα C_{1x}, C_{2x} (οριζόντια C_{1y}, C_{2y}) όρια της περιοχής του προσώπου, σε σχέση με κάποιο καρέ στο οποίο ο χρήστης κοιτάζε την κάμερα μετωπικά. Για την σύνθεση των τιμών που προκύπτουν από το στόμα και τα μάτια, χρησιμοποιείται γραμμική παλινδρόμηση.

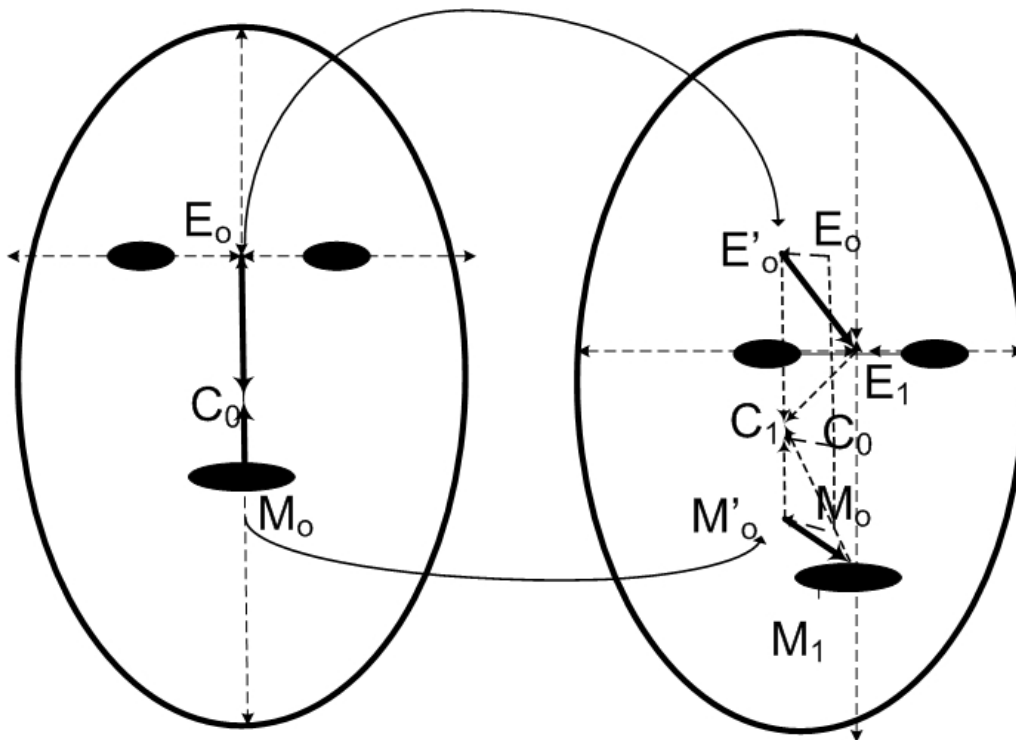
Το παραπάνω φαίνεται στο σχήμα 4.7 και τις εξισώσεις 4.7 και 4.8, ενώ με $y_{DVF,k}$ και $p_{DVF,k}$ συμβολίζονται οι τιμές της οριζόντιας και κατακόρυφης περιστροφής στο καρέ k .

$$\begin{aligned}
 y_{DVF,k} &= b_{1y} \cdot \left[\frac{(E_{x,k} - C_{1x,k}) + (E_{x,k} - C_{2x,k})}{2 \cdot d_{eyes,0}} \right. \\
 &\quad \left. - \frac{(E_{x,0} - C_{1x,0}) + (E_{x,0} - C_{2x,0})}{2 \cdot d_{eyes,0}} \right] \\
 &\quad + b_{2y} \cdot \left[\frac{(M_{x,k} - C_{1x,k}) + (M_{x,k} - C_{2x,k})}{2 \cdot d_{eyes,0}} \right. \\
 &\quad \left. - \frac{(M_{x,0} - C_{1x,0}) + (M_{x,0} - C_{2x,0})}{2 \cdot d_{eyes,0}} \right] \quad (4.7)
 \end{aligned}$$

$$\begin{aligned}
 p_{DVF,k} &= b_{1p} \cdot \left[\frac{(E_{y,k} - C_{1y,k}) + (E_{y,k} - C_{2y,k})}{2 \cdot d_{eyes,0}} \right. \\
 &\quad \left. - \frac{(E_{y,0} - C_{1y,0}) + (E_{y,0} - C_{2y,0})}{2 \cdot d_{eyes,0}} \right] \\
 &\quad + b_{2p} \cdot \left[\frac{(M_{y,k} - C_{1y,k}) + (M_{y,k} - C_{2y,k})}{2 \cdot d_{eyes,0}} \right. \\
 &\quad \left. - \frac{(M_{y,0} - C_{1y,0}) + (M_{y,0} - C_{2y,0})}{2 \cdot d_{eyes,0}} \right] \quad (4.8)
 \end{aligned}$$

όπου b_{1y}, b_{2y} και b_{1p}, b_{2p} οι συντελεστές της παλινδρόμησης που χρησιμοποιήθηκαν για τη σύντηξη της πληροφορίας που προέρχεται από το σημείο ανάμεσα στα μάτια και το κέντρο της περιοχής του στόματος για οριζόντια και κατακόρυφη περιστροφή αντίστοιχα. Για να ληφθούν υπόψη διαφορές ανάμεσα στη φυσιολογία των διαφόρων χρηστών, όλα τα μεγέθη, πριν τη συγχώνευσή τους με γραμμική παλινδρόμηση, μπορούν να κανονικοποιηθούν με την απόσταση ανάμεσα στα μάτια (όπως υπολογίστηκε στο καρέ 0, όπου γίνεται ο εντοπισμός προσώπου) $d_{eyes,0}$. Η κανονικοποίηση αυτή μπορεί να γίνει κάθε φορά που επαναρχικοποιείται το σύστημα, κι έτσι να επιτυγχάνονται αξιόπιστα αποτελέσματα, ακόμα και κατά τη μεταφορική κίνηση του χρήστη κατά μήκος του z άξονα. Όπως φαίνεται από τις παραπάνω εξισώσεις, η σχέση μεταξύ των υπολογισμένων συντεταγμένων των χαρακτηριστικών και της εκτίμησης περιστροφής κεφαλιού είναι γραμμική. Για το φιλτράρισμα του θορύβου, τα σήματα y_{DVF} και p_{DVF} που προκύπτουν, μπορούν να φιλτράρονται με φίλτρο FIR τάξης N (χρησιμοποιήθηκε $N=12$, με συντελεστές 0,3 και 0,2 για το τρέχον και το προηγούμενο καρέ, και 0,05 για όλα τα υπόλοιπα). Η παραπάνω μεθοδολογία εξηγείται διαισθητικά από το σχήμα 4.8.

Ο υπολογισμός της περιστροφής παράλληλα στο επίπεδο της κάμερας είναι πιο άμεσος: προκύπτει απευθείας από τη γωνία που ορίζεται από το ευθύγραμμο τμήμα



Σχήμα 4.7: Εκτίμηση Διανύσματος Περιστροφής Κεφαλιού (εξισώσεις 4.7 και 4.8)

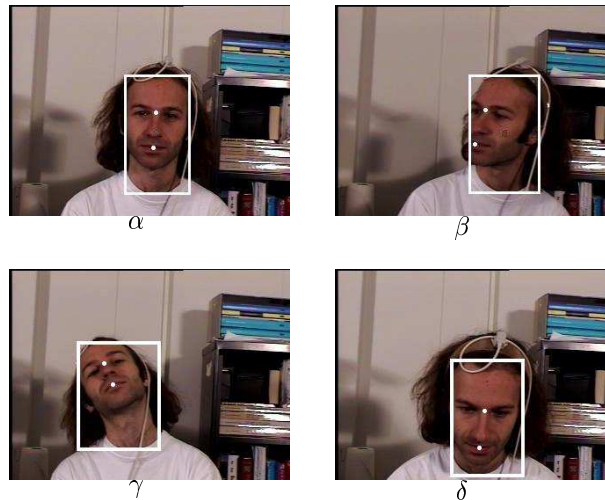
που συνδέει τα κέντρα των ματιών και τον οριζόντιο άξονα (σχήμα 4.9). Οι τιμές και πάλι φιλτράρονται, όπως έγινε για την οριζόντια και κατακόρυφη γωνία περιστροφής.

4.2.3.1 Σύγκριση παρακολούθησης χαρακτηριστικών προσώπου με Διανυσματικά Πεδία Απόστασης και Οπτική Ροή

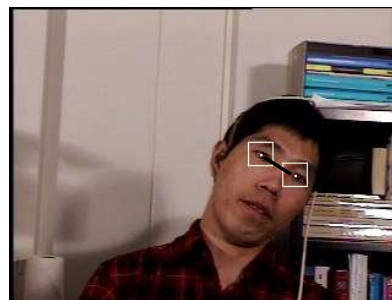
Η χρήση των Διανυσματικών Πεδίων Απόστασης για την παρακολούθηση των χαρακτηριστικών προσώπου συγκρίθηκε, ως προς την καταλληλότητά της, με οπτική ροή. Για το σκοπό αυτό, χρησιμοποιήθηκαν τα δεδομένα HPEG [9], καθώς η βάση αυτή δεν υλοποιήθηκε κάτω από ελεγχόμενες συνθήκες φωτισμού και φόντου και, επιπλέον, χρησιμοποιήθηκε μία απλή δικτυακή κάμερα για τη λήψη των ακολουθιών.

Μετά την εύρεση των χαρακτηριστικών προσώπου στο αρχικό καρέ της ακολουθίας (βλ. κεφάλαιο 2), χρησιμοποιήθηκε ο αλγόριθμος Lucas-Kanade [19] για την παρακολούθηση των θέσεων των ματιών, ενώ για τη θέση του στόματος χρησιμοποιήθηκε η μεθοδολογία που περιγράφεται στο 4.2.2.2. Οι περιοχές παρακολούθησης ήταν ίδιες με αυτές που αντιστοιχούν στον εντοπισμό και, με τον τρόπο που περιγράφτηκε στο 4.2.1, κι εδώ χρησιμοποιήθηκαν τα όρια του προσώπου για να περιοριστούν οι υποψήφιες περιοχές. Επιπλέον, ο γκαουσιανός όρος που περιγράφεται στην εξίσωση 4.6 εισήχθη και εδώ (εξίσωση 4.9), έτσι ώστε να δοθεί έμφαση στις αναμενόμενες, σύμφωνα με την ανθρώπινη γεωμετρία χαρακτηριστικών, θέσεις ματιών, ενώ, σε κάθε καρέ, η θέση των περιοχών των ματιών ενημερωνόταν ώστε να κεντράρεται γύρω από το κεντρικό σημείο της ίριδας.

$$\begin{aligned}
 \mathbf{p} &= \\
 &= \arg \min_{(u,v)} \left(\sum_{i \in R_k} \|(I_x \cdot u + I_y \cdot v + I_k)\|_2 e^{-\frac{(d_{k+1,v}-\mu)^2}{2\sigma^2}} \right)
 \end{aligned}
 \tag{4.9}$$



Σχήμα 4.8: Παραδείγματα περιοχών προσώπου με τα αντίστοιχα χαρακτηριστικά σημεία που χρησιμοποιούνται για εκτίμηση περιστροφής κεφαλιού



Σχήμα 4.9: Παράδειγμα οβελιαίας περιστροφής στη βάση BU.

με $\mathbf{v} = (u, v)$ η μετατόπιση του χαρακτηριστικού σε συνεχόμενα καρέ, I_x, I_y, I_k οι παράγωγοι της φωτεινότητας ως προς τη θέση (x, y) και το χρόνο k . Όπως και προηγουμένως, έτσι κι εδώ, η μεταβλητή $d_{k+1, \mathbf{v}}$ είναι ο λόγος μεταξύ της απόστασης των ματιών και της κατακόρυφης απόστασης μεταξύ ματιών και στόματος, στο καρέ $k+1$, και για μετατόπιση (u, v) του υπό παρακολούθηση χαρακτηριστικού. Η βέλτιστη μετατόπιση βρίσκεται με λύση της 4.9 χρησιμοποιώντας τον αλγόριθμο Levenberg-Marquardt [147].

Στον πίνακα 4.1 φαίνεται το μέσο τετραγωνικό σφάλμα (RMS error) στην εκτίμηση περιστροφής κεφαλιού με χρήση Οπτικής Ροής και Διανυσματικών Πεδίων Απόστασης. Φαίνεται ότι, για τις συνθήκες που ορίστηκαν, η παρακολούθηση χαρακτηριστικών με χρήση Διανυσματικών Πεδίων, γενικά δίνει καλύτερα αποτελέσματα από ό,τι με οπτική ροή. Κάτι τέτοιο μπορεί να οφείλεται στο ότι τα διανυσματικά πεδία κωδικοποιούν πληροφορία γεωμετρίας, ενώ στην περίπτωση της οπτικής ροής αναζητούνται ομοιότητες στη χρωματικότητα, κάτι το οποίο ενδέχεται να έχει περισσότερες από μία λύσεις. Αντίθετα, η αναζήτηση παρόμοιων δομών (περιοχή ματιού και φρυδιού), φαίνεται να έχει πιο περιορισμένο αριθμό λύσεων, με την πιθανότητα σφάλματος να αυξάνει όσο πιο ενθόρυβο γίνεται το περιβάλλον (για το λόγο αυτό αποκλείστηκε το φόντο).

Στο σχήμα 4.10 δίνεται μία γραφική παράσταση επιτυχούς εκτίμησης περιστροφής κεφαλιού με χρήση διαφορετικών κατωφλίων T για το χαρακτηρισμό μιας εκτίμησης ως επιτυχούς ή όχι, σε σχέση με τα δεδομένα αναφοράς, τόσο με χρήση Διανυσματι-

Πίνακας 4.1: Εύρεση περιστροφής κεφαλιού με χρήση παρακολούθησης χαρακτηριστικών με Οπτική Ροή και Διανυσματικά Πεδία Απόστασης

	Σφάλμα: Οπτική Ροή	Σφάλμα: ΔΠΑ
Οριζ. Περιστροφή	8.39°	6.65°
Κατακ. Περιστροφή	5.51°	5.59°
Μέσος Όρος	6.95°	6.12°

Πίνακας 4.2: Σύγκριση αποτελεσμάτων χρήσης ΔΠΑ στην BU

	Σφάλμα: ΔΠΑ
οριζ. περιστροφή	5.72° (rms) / 4.39°(mae)
κατακ. περιστροφή	4.89° (rms) / 3.87°(mae)
περιστροφή στο επίπεδο της εικόνας	3.56° (rms) / 2.61°(mae)

κών Πεδίων Απόστασης, όσο και με Οπτική Ροή.

4.2.4 Πειραματικά αποτελέσματα

4.2.4.1 Αποτελέσματα στη βάση του Πανεπιστημίου της Βοστώνης

Κατά την επαλήθευση των αποτελεσμάτων, τα βάρη των εξισώσεων 4.7 και 4.8, εκπαιδεύτηκαν σε όλα τα βίντεο εκτός από εκείνα που αντιστοιχούν στο υπό εξέταση άτομο. Έτσι, από τα 5 άτομα και 9 ακολουθίες για το κάθε ένα, το εκάστοτε μοντέλο παλινδρόμησης εκπαιδεύτηκε με χρήση 36 ακολουθιών. Στον πίνακα 4.2 φαίνονται τα *RMS* (και απόλυτα *MAE*) σφάλματα στην εκτίμηση περιστροφής, όπως εξήχθησαν με τη χρήση των εξισώσεων 4.7 και 4.8.

4.2.4.2 Αποτελέσματα στη βάση HPEG

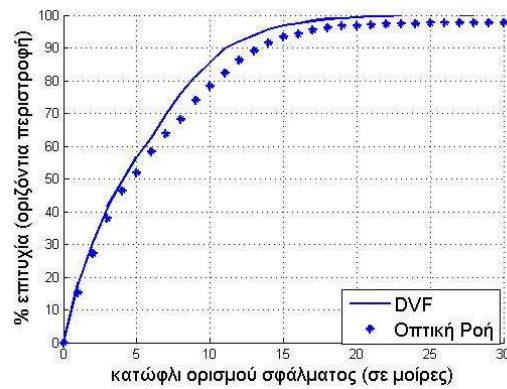
Αντίστοιχα είναι τα αποτελέσματα και στην HPEG. Τα αποτελέσματα (*RMS* σφάλματα) για οριζόντια και κατακόρυφη περιστροφή φαίνονται στον πίνακα 4.3.

4.2.4.3 Αποτελέσματα με εξαγωγή και χρήση χαρακτηριστικών καρέ

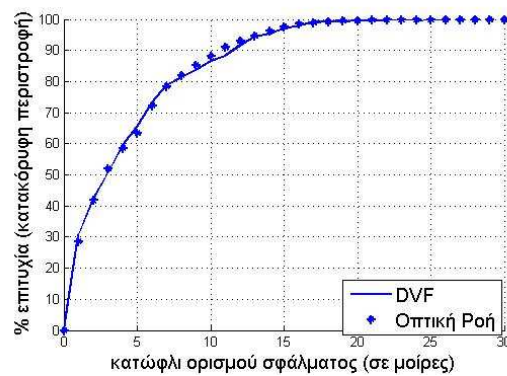
Για την περαιτέρω βελτίωση των αποτελεσμάτων, μπορεί να γίνει χρήση ολόκληρης της περιοχής που περικλείει το δέρμα του προσώπου. Πιο αναλυτικά, το εύρος των οριζόντιων και κάθετων γωνιών χωρίζεται σε 10 κλάσεις και, την πρώτη φορά που εμφανίζεται κάποιο καρέ, στο οποίο αντιστοιχίζεται οριζόντια και κάθετη γωνία περιστροφής πολύ κοντά στο κέντρο μιας κλάσης, το αντίστοιχο οριοθετικό πλαίσιο του

Πίνακας 4.3: Αποτελέσματα χρήσης ΔΠΑ στην HPEG

	Σφάλμα: ΔΠΑ
οριζ. περιστροφή	6.65°(rms)
κατακ. περιστροφή	5.59°(rms)



α



β

Σχήμα 4.10: Εκτίμηση (α) οριζόντιας και (β) κατακόρυφης περιστροφής κεφαλιού με χρήση διαφορετικών κατωφλίων T για το χαρακτηρισμό επιτυχούς υπολογισμού, για Διανυσματικά Πεδία και Οπτική Ροή.

Πίνακας 4.4: αποτελέσματα εκτίμησης οριζόντιας και κατακόρυφης περιστροφής του κεφαλιού στη βάση HPEG με χρήση παρακολούθησης με ΔΠΑ και χαρακτηριστικών καρτέ

	Σφάλμα: ΔΠΑ+Χαρακτ. Καρέ
οριζ. περιστροφή	6.59°(rms)
κατακ. περιστροφή	5.33°(rms)

προσώπου θεωρείται ως το αντιπροσωπευτικό καρτέ της κλάσης και ενημερώνεται η βάση των χαρακτηριστικών καρτέ. Τα οριοθετικά πλαίσια των προσώπων ακόλουθων καρτέ συσχετίζονται με το χαρακτηριστικό καρτέ της κάθε κλάσης και συγκρατείται το πλέον συσχετισμένο για κάθε καρτέ. Εφόσον μία γωνία περιστροφής που έχει υπολογιστεί για το τρέχον καρτέ είναι εκτός των ορίων της πιο συσχετισμένης κλάσης, η οριζόντια και κάθετη γωνία περιστροφής ξαναυπολογίζεται με βάση προηγούμενα καρτέ, έντονα συσχετισμένα με το τρέχον. Με βάση αυτή τη λογική, η οριζόντια και κάθετη περιστροφή υπολογίζονται με βάση τις εξισώσεις 4.10 και 4.11.

$$Y(i) = \frac{\sum_{j=1}^N f(\text{corr}(i, i-j), \text{threshold}) \cdot Y(i-j)}{\sum_{j=1}^N f(\text{corr}(i, i-j), \text{threshold})} \quad (4.10)$$

$$P(i) = \frac{\sum_{j=1}^N f(\text{corr}(i, i-j), \text{threshold}) \cdot P(i-j)}{\sum_{j=1}^N f(\text{corr}(i, i-j), \text{threshold})} \quad (4.11)$$

όπου $f(x, y) = x$ για $x > y$ και 0 αλλιώς, με i το τρέχον καρτέ, $\text{corr}(i, j)$ η συσχέτιση ανάμεσα στο οριοθετικό πλαίσιο του προσώπου του τρέχοντος καρτέ και σε προηγούμενα καρτέ j ³. Για την κατακόρυφη περιστροφή, βρέθηκε ότι το παραπάνω βήμα, στη συγκεκριμένη βάση, δίνει καλά αποτελέσματα για χαμηλό *threshold* (0.5), δηλ. λαμβάνοντας τιμές από πολλά προηγούμενα καρτέ ενώ, αντίθετα, για οριζόντια περιστροφή, τα αποτελέσματα σχεδόν παραμένουν αμετάβλητα, ενώ μόνο για μεγάλη τιμή του *threshold* (0.95) παρουσίασαν πολύ μικρή βελτίωση. Τα αποτελέσματα που προέκυψαν από την χρήση των χαρακτηριστικών καρτέ, φαίνονται στον πίνακα 4.4. Σύγκριση με τον πίνακα 4.3 καταδεικνύει μικρή βελτίωση, ειδικά στην περίπτωση της κατακόρυφης γωνίας περιστροφής.

4.3 Χρήση ολιστικής πληροφορίας με Συνελικτικά Νευρωνικά Δίκτυα

Για την εκμετάλλευση ολιστικής πληροφορίας για εκτίμηση περιστροφής κεφαλιού, γίνεται χρήση ενός συγκεκριμένου τύπου Νευρωνικών δικτύων, τα Συνελικτικά Νευρωνικά Δίκτυα (ΣΝΔ - Convolutional Neural Networks). Ο κύριος λόγος αυτής της επιλογής είναι ότι το κλασικό πολυεπίπεδο perceptron παρουσιάζει τα ακόλουθα μειονεκτήματα στην εκτίμηση περιστροφής προσώπου: Πρώτον, η πολυπλοκότητα της εκπαίδευσης όλων των βαρών σε ένα πλήρως συνδεδεμένο δίκτυο θα ήταν πολύ υψηλή, καθώς θα απαιτούνταν εκπαίδευση ενός πολύ μεγάλου αριθμού ελεύθερων παραμέτρων. Όπως θα φανεί και παρακάτω, η προτεινόμενη αρχιτεκτονική χρησιμοποιεί

³τα οριοθετικά πλαίσια έρχονται σε διάσταση 50×50 και ο χρόνος για τον υπολογισμό μιας ετεροσυσχέτισης είναι λιγότερος από 1ms σε έναν επεξεργαστή 3.2GHz. Θεωρήθηκε $N=20$

ένα μεγάλο αριθμό ταξινομητών, οι οποίοι δίνουν στην έξοδο βαθμό συμμετοχής συγκεκριμένων κλάσεων περιστροφής. Ο περιορισμός αυτός μείωσε τον αριθμό των διαθέσιμων εικόνων εκπαίδευσης για κάθε ταξινομητή ενώ, εκπαίδευση πολύπλοκων δικτύων με χρήση λίγων δεδομένων εκπαίδευσης ενέχει τον κίνδυνο υπερεκπαίδευσης. Τα ΣΝΔ, χάρη στο μοίρασμα βαρών (*weight sharing*) μειώνουν δραστικά τον αριθμό των παραμέτρων εκπαίδευσης, σε σχέση με ένα τυπικό νευρωνικό δίκτυο, κι έτσι, είναι κατάλληλα για περιπτώσεις περιορισμένου αριθμού δεδομένων εκπαίδευσης. Δεύτερον, ένα τυπικό νευρωνικό δίκτυο θα απαιτούσε πολύ καλή αντιστοίχιση χαρακτηριστικών των εικόνων εκπαίδευσης με τις εικόνες ελέγχου. Καθώς ο στόχος της παρούσας διατριβής είναι ακολουθίες βίντεο σε μη ελεγχόμενα και αυθόρμητα περιβάλλοντα, ο αλγόριθμος χειρισμού ολιστικής πληροφορίας απαιτεί ικανοποιητική ανοχή σε μικρές αποκλίσεις των δεδομένων από συγκεκριμένη ευθυγράμμιση. Τρίτον, τα τυπικά νευρωνικά δίκτυα χρησιμοποιούν τοπική πληροφορία γύρω από κάθε χαρακτηριστικό, ενώ δε λαμβάνουν υπόψη χωροταξικές σχέσεις ανάμεσα στα χαρακτηριστικά του προσώπου. Όπως θα αναλυθεί και παρακάτω, τα Συνελικτικά Νευρωνικά Δίκτυα, σε μεγάλο βαθμό, ικανοποιούν τις παραπάνω απαιτήσεις, κι έτσι είναι ιδανικά στη χρήση ολιστικής πληροφορίας.

Τα Συνελικτικά Νευρωνικά Δίκτυα (ΣΝΔ) [92] αποτελούν μία κλάση ταξινομητών εμπνευσμένοι από τη βιολογία, και πιο συγκεκριμένα, τα κύτταρα των οπτικών νεύρων της γάτας, τα οποία έχουν μία ειδική ικανότητα στο να ξεχωρίζουν προσανατολισμένες ακμές. Είναι γνωστή και ευρεία η χρήση τους στην αναγνώριση γραπτού χαρακτήρα [92], ενώ έχει μελετηθεί και η δυνατότητα αναγνώρισης εκφράσεων προσώπου με χρήση Συνελικτικών Νευρωνικών Δικτύων [43]. Στο [123] παρουσιάζεται, επίσης, προσπάθεια χρήσης των Συνελικτικών Δικτύων στον ταυτόχρονο εντοπισμό προσώπου και εκτίμηση περιστροφής κεφαλιού, σε ένα κοινό πλαίσιο. Στο [123], οι συγγραφείς κάνουν χρήση ΣΝΔ για τον εντοπισμό προσώπων, ανεξαρτήτως προσανατολισμού τους. Έτσι, εκπαιδεύονται ΣΝΔ, τα οποία αντιστοιχίζουν εικόνες εισόδου X , σε υποχώρους. Η "καμπύλη προσώπων" που αντιστοιχεί σε αυτούς τους υποχώρους αντιπροσωπεύεται από ένα πολύπτυχο (*manifold*), με την κατάλληλη παραμετροποίηση ώστε κατά μήκος του να αντιστοιχίζονται διαφορετικά είδη περιστροφής κεφαλιού. Το ΣΝΔ προβάλλει την εικόνα εισόδου στον υποχώρο και, αν η απόσταση της προβολής από το πολύπτυχο είναι μικρότερη από ένα προκαθορισμένο κατώφλι, τότε η εικόνα ταξινομείται ως πρόσωπο. Το κοντινότερο σημείο του πολύπτυχο στην προβολή της εικόνας χρησιμοποιείται για την ανάκτηση της πόζας του προσώπου. Στην παρούσα διατριβή, ακολουθώντας διαφορετική αρχιτεκτονική και διαδικασία συλλογιστικής, δημιουργήθηκαν μοντέλα υπολογισμού ακριβούς περιστροφής κεφαλιού με χρήση μιας σειράς υπο-ταξινομητών Συνελικτικών Δικτύων. Ποιοτική σύγκριση [123] (σε διαφορετικά δεδομένα) των δύο τύπων αρχιτεκτονικών, έδειξε ότι η προτεινόμενη αρχιτεκτονική μπορεί να υπολογίσει με μεγάλο βαθμό αξιοπιστίας την περιστροφή του κεφαλιού στον οριζόντιο και τον κάθετο άξονα, ενώ για την εκπαίδευση απαιτούνται συγκεκριμένες αντιπροσωπευτικές κλάσεις γωνιών περιστροφής. Επιπλέον, στο [123] δεν αναφέρονται αποτελέσματα ελέγχου της προτεινόμενης μεθοδολογίας για κατακόρυφη περιστροφή του κεφαλιού.

Η αρχιτεκτονική ενός τυπικού Συνελικτικού Νευρωνικού Δικτύου είναι παρόμοια με αυτήν ενός τυπικού Νευρωνικού Δικτύου, με την έννοια ότι αποτελείται από επίπεδα μετασχηματισμένων εκδόσεων της εισόδου. Πιο συγκεκριμένα, για ένα γενικό πρόβλημα αναγνώρισης αντικειμένου από εικόνα, ας θεωρηθεί η χρήση μιας $N \times N$ εικόνας στην είσοδο στο πρώτο στρώμα. Η εικόνα συνελίσσεται με μια σειρά φίλτρων

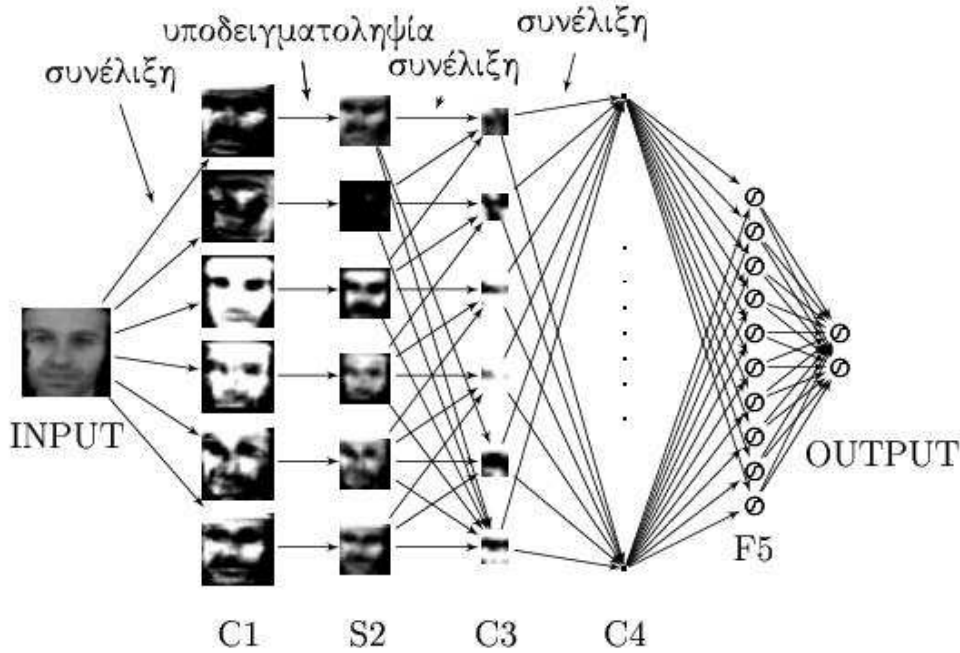
Πίνακας 4.5: Πίνακας συνδεσμολογίας επιπέδου $S2$ (γραμμές) με στρώμα συνέλιξης $C3$ (στήλες). Κάθε X αντιστοιχεί σε υπάρχουσα σύνδεση μεταξύ εξόδου του $S2$ και εισόδου του $C3$

	0	1	2	3	4	5
0	X				X	X
1	X	X				X
2	X	X	X			X
3		X	X	X		X
4			X	X	X	X
5				X	X	X

μεγέθους $p \times p$ στο δεύτερο στρώμα ($C1$), καταλήγοντας σε μετασχηματισμούς διάστασης $N - p + 1 \times N - p + 1$. Οι παράμετροι των φίλτρων (ίδιες για όλα) βρίσκονται αυτόματα, με την εκπαίδευση του δικτύου. Στο τρίτο επίπεδο $S2$, οι μετασχηματισμοί αυτοί υπόκεινται σε υπο-δειγματοληψία κατά έναν εκπαιδευσιμο από το δίκτυο παράγοντα. Τα παραπάνω χαρακτηριστικά των επιπέδων $C1$ και $S2$ εγγυώνται τα ακόλουθα: Πρώτον, η ικανότητα του δικτύου να εκπαιδευτεί με μικρό αριθμό δεδομένων, σε μικρό χρονικό διάστημα, είναι εφικτή, καθώς ο αριθμός των ελεύθερων παραμέτρων είναι σημαντικά μειωμένος (σε σχέση με το κλασικό πολυεπίpedo πέρσεπτρον), αφού οι μετασχηματισμοί στο επίπεδο $C1$ "μοιράζονται" τα ίδια βάρη. Δεύτερον, η υποδειγματοληψία κάνει το δίκτυο πιο ανθεκτικό σε μικρές παραμορφώσεις ή μετατοπίσεις του προτύπου προς αναγνώριση στην εικόνα εισόδου. Ακόμη, εναλασσόμενες αλληλουχίες επιπέδων συνέλιξης και υποδειγματοληψίας επιτρέπουν τον εντοπισμό απλών χαρακτηριστικών στα πρώτα στρώματα (πχ ακμές, γωνίες) τα οποία συνδυάζονται σε ανώτερα στρώματα, με σκοπό την εκμετάλλευση συνδυαστικής πληροφορίας, η οποία προέρχεται και από χωροταξικές σχέσεις μεταξύ των χαρακτηριστικών.

4.3.1 Προτεινόμενη αρχιτεκτονική

Η αρχιτεκτονική που υιοθετείται στην παρούσα διατριβή, είναι ένα Συνελικτικό Νευρωνικό Δίκτυο 6 στρωμάτων, με τα επίπεδα να ακολουθούν τη σειρά $C1$, $S2$, $C3$, $C4$, $F5$, $F6$ (έξοδος). Η έξοδος είναι ένα διδιάστατο διάνυσμα. Τα επίπεδα συνέλιξης χρησιμοποιούν φίλτρα 7×7 , ενώ το επίπεδο υποδειγματοληψίας χρησιμοποιεί παράγοντα υποδειγματοληψίας ίσο με 2. Τα στρώματα $F5$ και $F6$ αποτελούνται από 10 και από 2 (έξοδος) πλήρως συνδεδεμένους τυπικούς νευρώνες, αντίστοιχα. Για τα επίπεδα $C1$ και $C3$ χρησιμοποιήθηκαν 6 μετασχηματισμοί, και 80 για τον $C4$. Στην παρούσα διατριβή, η κύρια διαφορά από ένα ΣΝΔ είναι ότι, μεταξύ του τρίτου και του τέταρτου επιπέδου δεν χρησιμοποιήθηκε επίπεδο υποδειγματοληψίας, καθώς τα αποτελέσματα αποδείχτηκαν πιο ακριβή. Κάτι τέτοιο δικαιολογείται άλλωστε, από την ήδη μικρή ανάλυση των εικόνων εισόδου (32×32) η οποία, μετά από συνελίξεις και υποδειγματοληψίες, θα εκφυλιζόταν σε ελάχιστα εικονοστοιχεία. Επιπλέον, όπως προτείνεται και στο [93], χρησιμοποιήθηκε μη πλήρης συνδεσμολογία ανάμεσα στα επίπεδα $S2$ και $C3$ (πίνακας 4.5). Μία σχηματική αναπαράσταση του συνολικού δικτύου φαίνεται στο σχήμα 4.11.

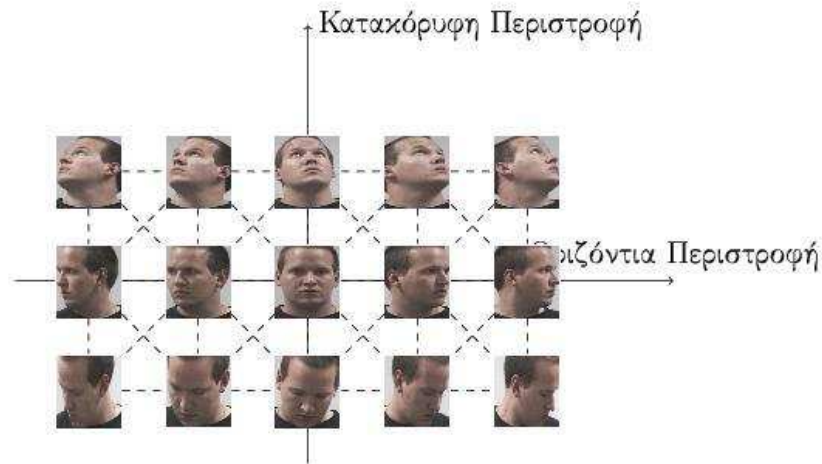


Σχήμα 4.11: Αρχιτεκτονική Συνελικτικού Νευρωνικού Δικτύου για εκτίμηση περιστροφής κεφαλιού

4.3.2 Διαδικασία εκπαίδευσης

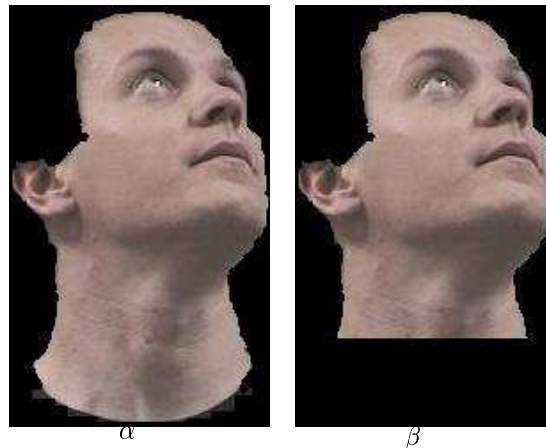
Το δίκτυο εκπαιδεύτηκε με χρήση του στοχαστικού αλγορίθμου Levenberg-Marquardt [94] (με παράμετρο $\mu=10^{-4}$, global ρυθμό μάθησης $5 \cdot 10^{-4}$, παράμετρο "μνήμης" στον υπολογισμό δεύτερων παραγώγων $\gamma=0.1$, καθώς και μέγιστο αριθμό εποχών ίσο με 10), ενώ, σαν συνάρτηση ενεργοποίησης σε όλη την έκταση του δικτύου χρησιμοποιήθηκε η υπερβολική εφαπτομένη. Ως δεδομένα εκπαίδευσης χρησιμοποιήθηκαν οι εικόνες της βάσης στο [51], η οποία αποτελείται από στατικές εικόνες ατόμων με περιστροφή κεφαλιού σε διάφορες, αλλά συγκεκριμένες γωνίες του οριζόντιου και κάθετου άξονα. Τα δεδομένα εκπαίδευσης μετατοπίστηκαν κατά ένα έως τρία εικονοστοιχεία προς όλες τις κατευθύνσεις, έτσι ώστε, στη συμπεριφορά του δικτύου να εισαχθεί ανεκτικότητα σε μικρές μεταβολές από την ευθυγράμμιση των απόλυτων ορίων του προσώπου. Επίσης, οι εικόνες υπέστησαν αντιστροφή ως προς τον κάθετο άξονα (mirroring), με σκοπό να αυξηθούν τα δεδομένα εκπαίδευσης. Τέλος, στην εκπαίδευση τα κέντρα των κλάσεων ήταν $\{-60^\circ, 0, 60^\circ\}$ στις κάθετες και $\{-90^\circ, -45^\circ, 0, 45^\circ, 90^\circ\}$ στις οριζόντιες γωνίες. Ωστόσο, στα δεδομένα εκπαίδευσης, αντιπροσωπευτικές εικόνες της κάθε κλάσης, θεωρήθηκαν κι αυτές που απείχαν $\pm 15^\circ$ (ή $\pm 30^\circ$ για κατακόρυφες γωνίες με κέντρα κλάσεων $\pm 60^\circ$) από τα αντίστοιχα κέντρα των κλάσεων.

Με τον παραπάνω καταμερισμό, δημιουργήθηκε ένας χώρος περιστροφής κεφαλιού 15 συνδυασμών (βλ. σχήμα 4.12). Για ευκολία στο συμβολισμό, οι κλάσεις ήταν: $C_1 : \{-90, -60\}$, $C_2 : \{-45, -60\}$, $C_3 : \{0, -60\}$, $C_4 : \{45, -60\}$, $C_5 : \{90, -60\}$, $C_6 : \{-90, 0\}$, $C_7 : \{-45, 0\}$, $C_8 : \{0, 0\}$, $C_9 : \{45, 0\}$, $C_{10} : \{90, 0\}$, $C_{11} : \{-90, 60\}$, $C_{12} : \{-45, 60\}$, $C_{13} : \{0, 60\}$, $C_{14} : \{45, 60\}$, $C_{15} : \{90, 60\}$, με το πρώτο στοιχείο να αντιστοιχεί στην οριζόντια περιστροφή, και το δεύτερο στην κατακόρυφη. Για κάθε συνδυασμό γειτονικών κλάσεων εκπαιδεύτηκε ένα ΣΝΔ, καταλήγοντας, με αυτόν τον τρόπο, σε ένα σύνολο 38 ταξινομητών. Σε αυτό το στάδιο, λοιπόν, η εκπαίδευση έγινε χρησιμοποιώντας ζεύγη εικόνων που αντιστοιχούν σε διαφορετικά, αλλά γειτο-



Σχήμα 4.12:

Κλάσεις περιστροφής κεφαλιού για εκπαίδευση των ΣΝΔ. Κάθε ταξινομητής συμβολίζεται με μια διακεκομμένη γραμμή.



Σχήμα 4.13: α) Περιοχή προσώπου με βάση το χρωματικό πρότυπο β) Μέρος περιοχής προσώπου, όπως χρησιμοποιείται από τα ΣΝΔ

νικά σημεία στο χώρο περιστροφής κεφαλιού, ενώ οι τιμές στόχου για την έξοδο ήταν $\{-1, 1\}$ ή $\{1, -1\}$, ανάλογα με το ζεύγος εκπαίδευσης. Πριν το στάδιο της εκπαίδευσης, όλες οι εικόνες κανονικοποιήθηκαν ώστε να έχουν μέσο όρο τιμών φωτεινότητας 0 και τυπική απόκλιση 1, η κάθε μία. Επιπλέον, για να επιτευχθεί ομοιομορφία στα δεδομένα, όλες οι εικόνες προσώπων ήρθαν σε τέτοια μορφή ώστε το μήκος τους να είναι 1.3 φορές μεγαλύτερο από το πλάτος, και κατόπιν ήρθαν σε διάσταση 32×32 εικονοστοιχεία (βλ σχήμα 4.13). Ο συνολικός αριθμός εικόνων εκπαίδευσης για κάθε υπο-ταξινομητή, ποικίλε από 29400 έως 52920 εικόνες, ανάλογα με την τοπολογική θέση της κάθε κλάσης στο χώρο περιστροφής κεφαλιού.

4.3.3 Εκτίμηση οριζόντιας και κατακόρυφης περιστροφής κεφαλιού με χρήση Συνελικτικών Νευρωνικών Δικτύων

Για την εκτίμηση της περιστροφής του κεφαλιού, με βάση την ολιστική πληροφορία από την περιοχή του προσώπου στα Συνελικτικά Νευρωνικά Δίκτυα, η περιοχή του



α



β

Σχήμα 4.14: Περιστροφή προσώπου στο επίπεδο της κάμερας, για εξαγωγή γωνιών γύρω από τον οριζόντιο και κατακόρυφο άξονα

προσώπου διαμορφώνεται όπως φαίνεται στο σχήμα 4.13 και, αφού έρχεται σε διάσταση 32×32 εικονοστοιχεία, οι τιμές φωτεινότητας κανονικοποιούνται ώστε να έχουν μέσο όρο ίσο με 0 και τυπική απόκλιση ίση με 1. Κατόπιν, ενεργοποιούνται εκείνα τα n δίκτυα που περιέχουν την κλάση C_c , της οποίας το κέντρο είναι το κοντινότερο στην οριζόντια και κατακόρυφη γωνία περιστροφής που αντιστοιχεί στο προηγούμενο καρέ. Με αυτόν τον τρόπο, για κάθε καρέ χρησιμοποιείται ένα υποσύνολο των ΣΝΔ, κι έτσι το σύστημα καθίσταται πιο γρήγορο και αξιόπιστο, αφού η πιθανότητα σφάλματος ελαχιστοποιείται, με τη θεώρηση γειτονικών κλάσεων με την πραγματική γωνία περιστροφής. Σαν παράδειγμα, εάν η οριζόντια και κατακόρυφη γωνία είναι πιο κοντά στο κέντρο της κλάσης C_8 , τα συνελικτικά δίκτυα που χρησιμοποιούνται είναι τα $C_8 - C_3, C_8 - C_2, C_8 - C_4, C_8 - C_6, C_8 - C_9, C_8 - C_{12}, C_8 - C_{13}, C_8 - C_{14}$. Στη γενική περίπτωση, ας θεωρηθεί ότι το πρώτο στοιχείο του διανύσματος εξόδου του κάθε ταξινομητή συμβολίζεται με out_1 , και το δεύτερο στοιχείο, με out_2 . Εδώ, θεωρήθηκε η διαφορά των δύο εξόδων ώστε να συγκριθεί η κλάση C_c με τις γειτονικές της και, έτσι, αφού αναλύεται κάθε καρέ, η συνολική έξοδος, είναι ένα διάνυσμα αποτελούμενο από οκτώ στοιχεία ($out_{up}, out_{down}, out_{left}, out_{right}, out_{up,left}, out_{up,right}, out_{down,left}, out_{down,right}$, ανάλογα με την τοπολογική σχέση της κλάσης C_c με κάθε κλάση με την οποία συγκρίνεται⁴). Στην εκτίμηση της οριζόντιας περιστροφής, οι τιμές out_{up} και out_{down} αποκλείστηκαν από το μοντέλο, όπως αντίστοιχα στην περίπτωση εκτίμησης κατακόρυφης περιστροφής, αποκλείστηκαν οι τιμές out_{left} και out_{right} , δίνοντας, με αυτόν τον τρόπο, μία συνολική έξοδο 7 διαστάσεων, συμπεριλαμβανομένου του κέντρου της κλάσης C_c . Η τελική εκτίμηση της οριζόντιας (κάθετης) περιστροφής γίνεται με χρήση γραμμικών μοντέλων παλινδρόμησης στα στοιχεία του παραπάνω διανύσματος, που αντιστοιχούν στις εξόδους των ταξινομητών, τα οποία εκπαιδεύτηκαν σε επισημειωμένα δεδομένα (βλ υποκεφάλαιο 4.4.1). Στο σημείο αυτό, θα πρέπει να σημειωθεί ότι, καθώς δεν θεωρήθηκαν περιπτώσεις προσώπων με περιστροφή παράλληλη στο επίπεδο της εικόνας, κατά την εκπαίδευση, όλα τα πρόσωπα στις εικόνες εισόδου, περιστρέφονται ώστε το ευθύγραμμο τμήμα που ενώνει τα μάτια, να είναι παράλληλο με τον οριζόντιο άξονα (βλ σχήμα 4.14).

⁴Στην περίπτωση που η κλάση C_c βρίσκεται στα όρια του χώρου περιστροφή κεφαλιού (βλ. σχήμα 4.12), χρησιμοποιούνται υποθετικοί ταξινομητές, οι οποίοι δίνουν έξοδο ίση με 2 (θέτοντας $out_1=1$ και $out_2=-1$ για την υπάρχουσα κλάση C_c και τη μη υπάρχουσα, αντίστοιχα) για τις μη υπάρχουσες, πέραν των ορίων κλάσεις

4.4 Σύντηξη ολιστικής με τοπική πληροφορία με χρήση σταθμισμένου μέσου

Για να γίνει εκμετάλλευση πληροφορίας προερχόμενη κι από τις δύο τεχνικές (Διανυσματικά Πεδία Απόστασης και Συνελικτικά Νευρωνικά Δίκτυα), συγχωνεύτηκαν αποτελέσματα και από τις δύο μεθόδους. Για την συγχώνευση χρησιμοποιήθηκε σταθμισμένος μέσος των τιμών της οριζόντιας και της κατακόρυφης περιστροφής (εξισώσεις 4.12 και 4.13, αντίστοιχα), $y_{DFV,i}$, $p_{DFV,i}$, όπως υπολογίζονται από τις συντεταγμένες των χαρακτηριστικών, και τις αντίστοιχες γωνίες $y_{CNN,i}$, $p_{CNN,i}$ (υποκεφάλαιο 4.3.3), όπως υπολογίζονται από τα ΣΝΔ, στο καρέ i .

$$yaw_i = (1 - a) \times y_{DFV,i} + a \times y_{CNN,i} \quad (4.12)$$

$$pitch_i = (1 - a) \times p_{DFV,i} + a \times p_{CNN,i} \quad (4.13)$$

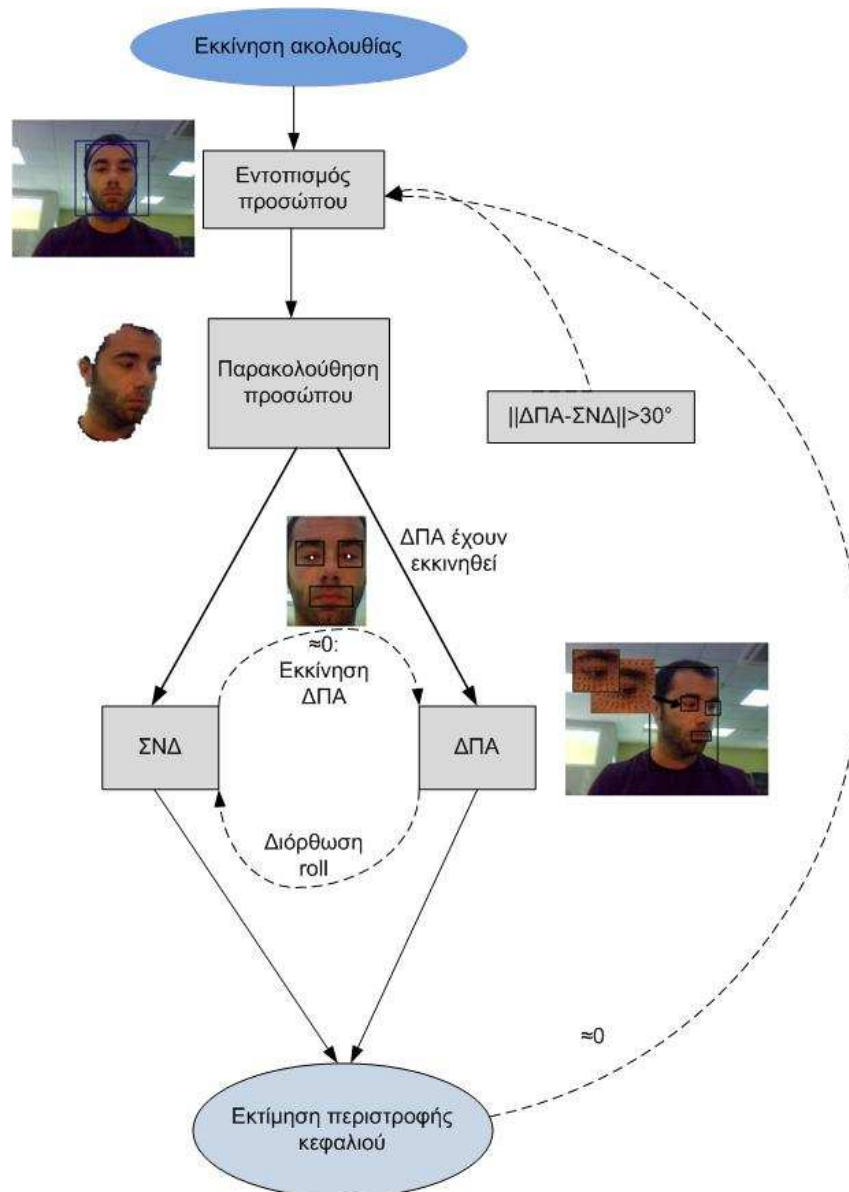
χρησιμοποιήθηκε $a=0.4$ στα πειράματα.

Να σημειωθεί πως, με σκοπό το σύστημα να μπορεί να εκκινείται από αυθαίρετες γωνίες περιστροφής του προσώπου (όχι, όμως, με περιστροφή παράλληλη στο επίπεδο της κάμερας), αρκεί στα πρώτα καρέ να χρησιμοποιείται μόνο ολιστική πληροφορία. Στο καρέ εκείνο, που υπολογίζεται περιστροφή προσώπου χαμηλότερη από ένα κατώφλι, μπορεί να εκκινηθεί ο αλγόριθμος τοπικής πληροφορίας. Ωστόσο, στις βάσεις που χρησιμοποιήθηκαν, όλες οι ακολουθίες ξεκινούσαν από μετωπική στάση, με αποτέλεσμα να εκκινούνται και οι δύο τεχνικές ταυτοχρόνως. Ένας επιπλέον κανόνας επανεκκίνησης που τέθηκε, είχε να κάνει με το παρακάτω σενάριο: Εάν η ολιστική και η τοπική πληροφορία διαφέρουν σημαντικά ($>30^\circ$), η μέθοδος επανεκκινείται με ανίχνευση προσώπου. Εφόσον πρόσωπο δεν ανιχνευθεί [171], ο αλγόριθμος επαναλαμβάνεται, μέχρι το πρόσωπο να έρθει σε μετωπική στάση και ανιχνευθεί. Για μια συγκεντρωτική εσπερία του συνολικού συστήματος εύρεσης περιστροφής προσώπου με χρήση ολιστικής και τοπικής πληροφορίας, προερχόμενες από Συνελικτικά Νευρωνικά Δίκτυα και Διανυσματικά Πεδία Απόστασης, το σχήμα 4.15 είναι διαφωτιστικό.

4.4.1 Αποτελέσματα στη βάση του Πανεπιστημίου της Βοστώνης

Έγιναν τρία διαφορετικά πειράματα στη βάση του Πανεπιστημίου της Βοστώνης [22], σε ό,τι αφορά στην εκτίμηση περιστροφής κεφαλιού: Με χρήση μόνο Διανυσματικών Πεδίων Απόστασης, χρήση μόνο Συνελικτικών Νευρωνικών Δικτύων και σύντηξη των δύο. Χρησιμοποιήθηκαν τα μοντέλα παλινδρόμησης που περιγράφηκαν στα υποκεφάλαια 4.2.3 και 4.3.3 για τα ΔΠΑ και τα ΣΝΔ αντίστοιχα, ενώ για τη χρήση τους ακολουθήθηκε πρωτόκολλο παράλειψης στιγμιτύπου, εκπαιδευόντάς τα με τις ακολουθίες όλων των ατόμων της βάσης εκτός εκείνων των οποίων οι ακολουθίες ελέγχονται.

Στους πίνακες 4.6 και 4.7 φαίνονται τα μέσα τετραγωνικά και μέσα απόλυτα, αντίστοιχα, σφάλματα στην εκτίμηση περιστροφής κεφαλιού με χρήση Διανυσματικών Πεδίων Απόστασης, Συνελικτικών Νευρωνικών Δικτύων, καθώς και με σύντηξη των δύο. Τα αποτελέσματα συγκρίνονται και με αυτά άλλων πρόσφατων μεθόδων στη βιβλιογραφία, στις οποίες, όμως, δηλώνεται ότι χρησιμοποιείται το αντίστοιχο σφάλμα στην ίδια βάση. Από τα αποτελέσματα γίνεται εμφανές ότι η προτεινόμενη τεχνική είναι συγκρίσιμη και, κατά περιπτώσεις, πιο ακριβής σε σχέση με άλλες μεθόδους.



Σχήμα 4.15: Μέθοδος εκτίμησης περιστροφής προσώπου με χρήση Διανυσματικών Πεδίων Απόστασης και Συνελικτικών Νευρωνικών Δικτύων: Αρχικά εξάγεται η περιοχή του προσώπου και, βάσει δείγματος από το πρόσωπο, παρακολουθείται η περιοχή του προσώπου, η οποία χρησιμοποιείται σαν είσοδος στα Συνελικτικά Δίκτυα. Όταν η περιστροφή είναι σχεδόν μηδενική, εκκινείται εντοπισμός/παρακολούθηση προσώπου με Διανυσματικά Πεδία Απόστασης. Ο αλγόριθμος μπορεί να επανεκκινείται όταν η περιστροφή υπολογίζεται μηδενική. Επιπλέον, μεγάλες διαφορές στα αποτελέσματα ολιστικής από τοπική μέθοδο ομαλοποιούνται επανεκκινώντας το σύστημα σε κατάλληλη χρονική στιγμή.

Πίνακας 4.6: Μέσα Τετραγωνικά Σφάλματα στη BU για Διανυσματικά Πεδία Απόστασης, Συνελικτικά Νευρωνικά Δίκτυα, σύνηξη των δύο, και άλλες μεθόδους της βιβλιογραφίας.

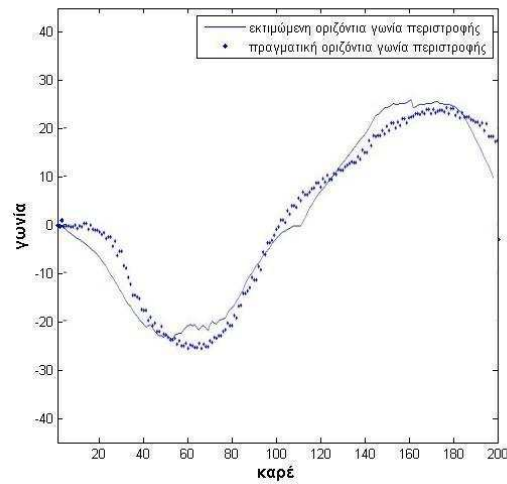
	ΔΠΑ	ΣΝΔ	ΔΠΑ+ΣΝΔ	[168]	[158]
	Σφάλμα STD	Σφάλμα STD	Σφάλμα STD	Σφάλμα STD	Σφάλμα
οριζ. περιστροφή	5.72° 4.90°	7.07° 6.45°	5.65° 5.02°	6.10° 5.79°	5.40°
κατακ. περιστροφή	4.89° 4.04°	5.81° 5.13°	4.78° 4.11°	5.26° 4.67°	5.60°
περιστροφή στο επίπεδο της εικόνας	3.56° 3.05°	—	3.56° 3.05°	3.00° 2.82°	3.10°

Πίνακας 4.7: Μέσα απόλυτα Σφάλματα στη BU για Διανυσματικά Πεδία Απόστασης, Συνελικτικά Νευρωνικά Δίκτυα, σύνηξη των δύο, και άλλες μεθόδους της βιβλιογραφίας.

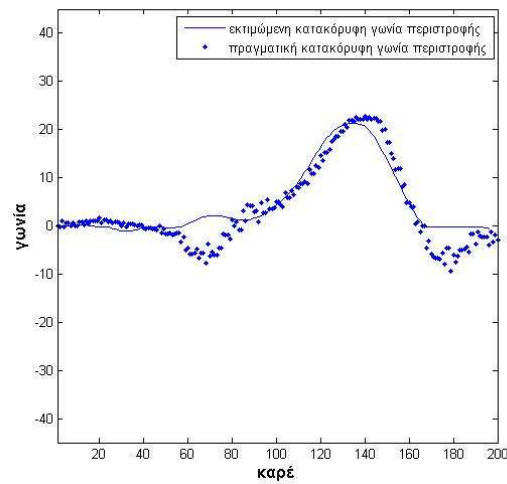
	Σφάλμα: ΔΠΑ	ΣΝΔ	ΔΠΑ+ΣΝΔ	[110]
οριζ. περιστροφή	4.39°	5.63°	4.38°	4.97° (mae)
κατακ. περιστροφή	3.87°	4.74°	3.82°	3.67° (mae)
περιστροφή στο επίπεδο της εικόνας	2.61°	—	2.61°	2.91° (mae)

Ωστόσο, το πλεονέκτημα της προτεινόμενης τεχνικής είναι ότι δεν απαιτείται εκ των προτέρων γνώση σχετικά με τις παραμέτρους της κάμερας (αρκεί να μην είναι στενής ή ευρείας γωνίας θέασης) ή την απόσταση του χρήστη σε σχέση με την κάμερα. Ακόμη, σε αντίθεση με άλλες μεθόδους, δεν υπάρχει η απαίτηση το μέγεθος του προσώπου, όπως προβάλλεται στην κάμερα, να φαίνεται σταθερό, αφού είναι ενσωματωμένος μηχανισμός επανεκκίνησης του συστήματος, καθώς και κανονικοποίηση με σχετικά μεγέθη, ώστε να μειώνεται η επίδραση του παράγοντα κλιμάκωσης.

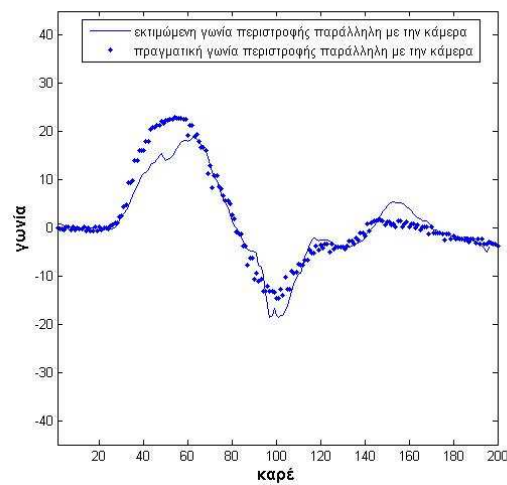
Οι μέθοδοι στα άρθρα [168] και [158] χρησιμοποιούν κυλινδρικά μοντέλα του κεφαλιού, κάνοντας υποθέσεις, είτε σχετικά με το εστιακό μήκος της κάμερας, είτε την σταθερότητα του χρήστη σε ό,τι αφορά την απόστασή του από την κάμερα [3]. Επιπλέον, στην παρούσα διατριβή, κάνοντας χρήση των Συνελικτικών Νευρωνικών Δικτύων στα πρώτα καρέ, εξαλείφεται και η απαίτηση μετωπικής στάσης σε περιστροφή αναφοράς στο ξεκίνημα. Ακόμη, εφόσον η μέθοδος χρησιμοποιεί σαν είσοδο την περιοχή του προσώπου, και επανεκκινείται όταν ικανοποιούνται συγκεκριμένοι περιορισμοί, μπορεί να παρακολουθήσει, με μεγάλη ακρίβεια, περίπλοκες κινήσεις του κεφαλιού, ακόμα και αν αυτές εμπεριέχουν κάθετες στο επίπεδο της κάμερας κινήσεις. Στις γραφικές παραστάσεις του σχήματος 4.16 φαίνονται χαρακτηριστικά παραδείγματα γωνιών περιστροφής κατά τη διάρκεια τριών διαφορετικών ακολουθιών, συνοδευόμενα από τα δεδομένα αναφοράς, για σύγκριση.



α



β



γ

Σχήμα 4.16: Εκτιμώμενες γωνίες περιστροφής στη βάση BU και αντίστοιχες τιμές δεδομένων αναφοράς, με χρήση ολιστικής και τοπικής πληροφορίας.

4.5 Σύντηξη ολιστικής με τοπική πληροφορία με χρήση μπεϋζιανών δικτύων

Με βάση την παρατήρηση ότι κάθε ένα από τα δύο επίπεδα πληροφορίας (τοπική και ολιστική) εμφανίζει διαφορετική αξιοπιστία, ανάλογα με τη φάση της διάδρασης, στο πλαίσιο της παρούσας διατριβής μελετήθηκε η δυνατότητα να ληφθεί υπόψη αυτή η παράμετρος στη διαδικασία σύντηξής τους. Για το σκοπό αυτό έγινε χρήση Μπεϋζιανής Σύντηξης επιπέδων πληροφορίας (Bayesian Modality Fusion), ώστε να μοντελοποιηθεί η αξιοπιστία του κάθε επιπέδου πληροφορίας ανάλογα με τη φάση διάδρασης. Πιο συγκεκριμένα, έγινε εκπαίδευση Μπεϋζιανών Δικτύων, ώστε να υπολογιστούν οι πιθανοτικές σχέσεις ανάμεσα στην πραγματική γωνία περιστροφής του κεφαλιού, και τις τιμές που προκύπτουν από τους αλγορίθμους υπολογισμού της.

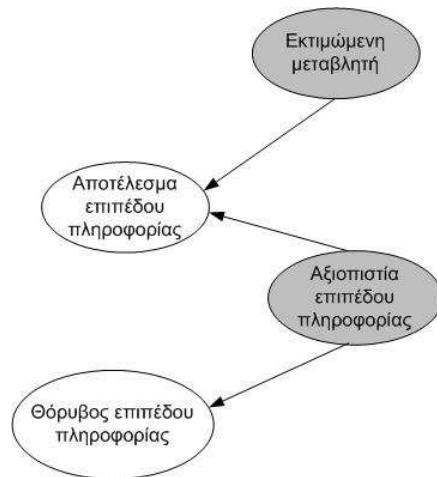
Με την έννοια *μπεϋζιανό δίκτυο* στη βιβλιογραφία ορίζεται ένας κατευθυντικός ακυκλικός γράφος, ο οποίος αναπαριστά την από κοινού κατανομή πιθανότητας για ένα σύνολο μεταβλητών [60],[74],[126]. Σε ένα τέτοιο δίκτυο, κόμβοι είναι τυχαίες μεταβλητές, και ακμές οι πιθανοτικές εξαρτήσεις ανάμεσα σε ζεύγη μεταβλητών. Τέτοιου είδους εξαρτήσεις σε ένα μπεϋζιανό δίκτυο μοντελοποιούν αιτιατές αλληλεπιδράσεις ανάμεσα σε μεταβλητές.

Στην παρούσα διατριβή, κατά τη δημιουργία του δικτύου για τον υπολογισμό γωνίας περιστροφής του κεφαλιού, η πραγματική γωνία θεωρείται τυχαία, παρατηρήσιμη μεταβλητή, η οποία επηρεάζει τις (παρατηρούμενες) μεταβλητές που αντιστοιχούν στα αποτελέσματα εκτίμησης περιστροφής από τοπική και ολιστική πληροφορία. Για την προσαρμογή των επιπέδων πληροφορίας στον εκάστοτε τύπο περιεχομένου, πέρα από τις παρατηρήσιμες μεταβλητές αποτελεσμάτων, χρησιμοποιούνται και άλλες, είτε παρατηρήσιμες, είτε κρυφές.

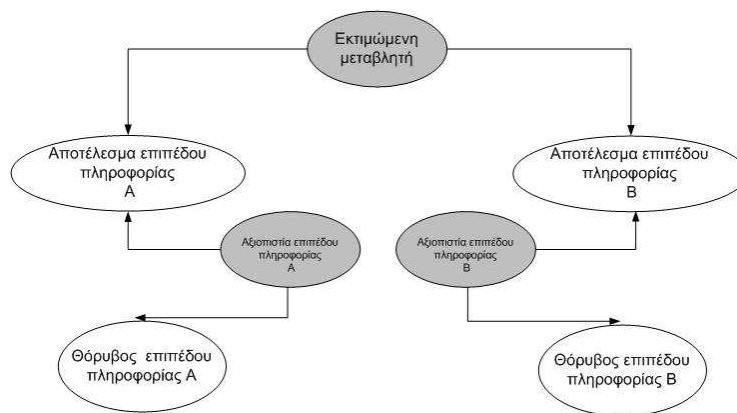
4.5.1 Μπεϋζιανά μοντέλα συλλογιστικής από ένα επίπεδο πληροφορίας

Στην εικόνα 4.17 φαίνεται η γενική μορφή συλλογιστικής με μπεϋζιανό δίκτυο, για ένα μόνο επίπεδο πληροφορίας. Οι κόμβοι του γράφου αναπαριστούν μεταβλητές ενδιαφέροντος, με τους λευκούς να αντιστοιχούν σε παρατηρήσιμες μεταβλητές, και τους γκριζούς σε κρυφές. Ο κόμβος *εκτιμώμενη μεταβλητή* είναι εκείνη η μεταβλητή που, με βάση την κατανομή πιθανότητάς της στην έξοδο, γίνεται η συλλογιστική της τελικής τιμής εξόδου του συστήματος.

Από μπεϋζιανής πλευράς, η πραγματική τιμή του συστήματος (*εκτιμώμενη μεταβλητή*) μια δεδομένη στιγμή είναι αυτό το οποίο επηρεάζει το αποτέλεσμα της εξόδου της ροής πληροφορίας. Αυτή η σχέση αιτίου-αιτιατού μοντελοποιείται από μια ακμή που ξεκινάει από την *εκτιμώμενη μεταβλητή* και καταλήγει στο *αποτέλεσμα επιπέδου (ροής) πληροφορίας*. Το *αποτέλεσμα επιπέδου πληροφορίας*, επιπλέον, επηρεάζεται από την *αξιοπιστία* του (την ακρίβεια με την οποία μπορεί να εκτιμήσει την τιμή της εξόδου). Ωστόσο, η αξιοπιστία είναι ένα μέγεθος το οποίο δε μπορεί να μετρηθεί απευθείας. Η ιδέα πίσω από τη Μπεϋζιανή λογική είναι ότι, τόσο η αξιοπιστία του επιπέδου πληροφορίας, όσο και τα αποτελέσματά του, είναι μεγέθη που διαφοροποιούνται ανάλογα με το περιεχόμενο της σκηνής. Για να δημιουργηθούν εκείνα τα πλαίσια στα οποία θα γίνεται δυναμική αξιοποίηση πληροφοριών, ανάλογα με το περιεχόμενο της σκηνής, θεωρήσαμε την αξιοπιστία του επιπέδου πληροφορίας ως μεταβλητή, και τη μοντελοποιήσαμε με τέτοιο τρόπο, ώστε να προσεγγίζεται από μετρήσιμες ποσότητες.



Σχήμα 4.17: Γενική μορφή μπεϋζιανού δικτύου για ένα επίπεδο πληροφορίας



Σχήμα 4.18: Μπεϋζιανό Δίκτυο για δύο επίπεδα πληροφορίας

Στην εικόνα 4.17, το μέγεθος αυτό αντιστοιχεί στο θόρυβο επιπέδου πληροφορίας.

Συνολικά, η πραγματική τιμή εξόδου του συστήματος υπολογίζεται με γνώση των παρατηρήσιμων μεταβλητών των αποτελεσμάτων και θορύβου μέτρησης. Ο υπολογισμός της εξόδου γίνεται από τη συνάρτηση κατανομής πιθανότητας στον κόμβο της πραγματικής γωνίας (εκτιμώμενη μεταβλητή).

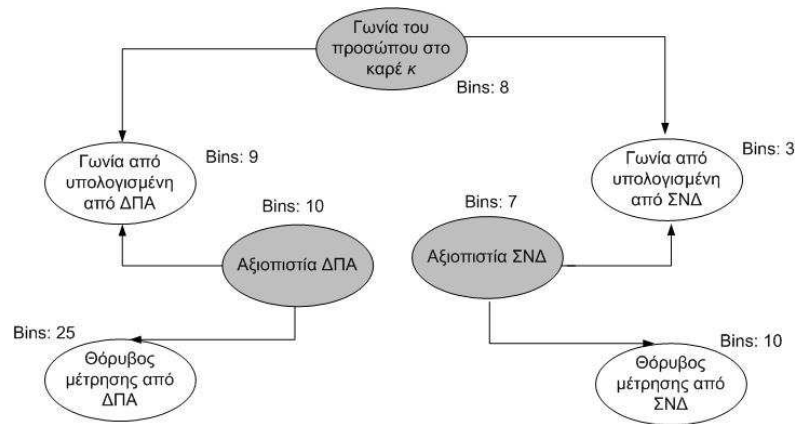
4.5.2 Μπεϋζιανά μοντέλα συλλογιστικής από περισσότερα επίπεδα πληροφορίας

Στην εικόνα 4.18 φαίνεται γενίκευση των όσων περιγράφηκαν προηγουμένως, για δυο επίπεδα πληροφορίας, ενώ στο σχήμα 4.19 δίνεται μια σχηματική αναπαράσταση του μοντέλου που χρησιμοποιήθηκε.

4.5.3 Θόρυβος μέτρησης από τοπική πληροφορία

Ως θόρυβος μέτρησης $noiseDVF_{y,i}$ της τοπικής πληροφορίας στο καρέ i θεωρήθηκε ο λόγος ανάμεσα στην κάθετη απόσταση μεταξύ ματιών και στόματος και ματιών μεταξύ τους.

$$noiseDVF_{y,i} = \frac{dist(Eyes_{middle}, Mouth_{middle})}{dist(Eyes_{right}, Eyes_{left})} \quad (4.14)$$



Σχήμα 4.19: Αρχιτεκτονική μπεϋζιανού δικτύου που χρησιμοποιείται για σύντηξη ολιστικής (ΣΝΔ) με τοπική πληροφορία (ΔΠΑ)

4.5.4 Θόρυβος μέτρησης από ολιστική πληροφορία

Για κάθε τιμή $y_{CNN,i}$ χρησιμοποιήθηκε ο παράγοντας εμπιστοσύνης σύμφωνα με την εξίσωση 4.15, ως θόρυβος της μέτρησης από ΣΝΔ.

$$noiseCNN_{y,i} = 1 - \frac{|y_{CNN,i} - m_{y,i:i-n+1}|}{std_{y,i:i-n+1}} \quad (4.15)$$

με $m_{y,i:i-n+1}$ και $std_{y,i:i-n+1}$ οι μέσες τιμές και οι τυπικές αποκλίσεις της οριζόντιας περιστροφής, σε χρονικά παράθυρα μήκους n προηγούμενων καρέ. Τιμές του n από 5 έως 25 δεν είχαν μεγάλη επίδραση στα αποτελέσματα (εδώ, χρησιμοποιήθηκε $n=5$).

Οι παράμετροι του θορύβου θεωρείται ότι, υπό κανονικές συνθήκες, παίρνουν τιμές σε συγκεκριμένο εύρος ενώ, όταν η αξιοπιστία του εκάστοτε επιπέδου πληροφορίας είναι χαμηλή, οι τιμές που αντιστοιχούν στις ποσότητες 4.14 και 4.15 μπορεί να είναι αυθαίρετες [95].

4.5.5 Παράμετροι και εκπαίδευση δικτύου

Η εκπαίδευση παραμέτρων του μπεϋζιανού δικτύου έγινε με εκπαίδευση των υπό συνθήκη πιθανοτήτων πινάκων των κόμβων. Στην εκπαίδευση χρησιμοποιήθηκε η βάση του Πανεπιστημίου της Βοστώνης [22] και σαν τιμές στους κόμβους που αντιστοιχούν στην πραγματική γωνία του προσώπου χρησιμοποιήθηκαν τα δεδομένα αναφοράς της βάσης. Το μπεϋζιανό δίκτυο μπορεί να θεωρηθεί σαν ένα σύνολο υπό συνθήκη πιθανοτήτων πινάκων, οι οποίοι προκύπτουν με εκτίμηση μέγιστης πιθανοφάνειας.

Μετά την εκπαίδευση, το δίκτυο μπορεί να χρησιμοποιηθεί για τη σύντηξη των επιπέδων πληροφορίας από την ολιστική και την τοπική μέθοδο. Για το σκοπό αυτό, χρησιμοποιούνται τα αποτελέσματα των εκτιμήσεων των γωνιών από κάθε επίπεδο, καθώς και οι μεταβλητές που αντιστοιχούν στο θόρυβο μέτρησης της κάθε πληροφορίας (δείκτες αξιοπιστίας). Τελικά, το δίκτυο υπολογίζει την κατανομή πιθανότητας γύρω από τη γωνία που θα δινόταν από τα δεδομένα αναφοράς, δοθεισών των μετρήσεων, και υπολογίζει την τελική τιμή της από το ολοκλήρωμά της. Η κατανομή αυτή βρίσκεται με χρήση του αλγορίθμου διάδοσης πιθανότητας σε δένδρα διακλάδωσης [61]. Ο αλγόριθμος αυτός δρα σε μια μορφή μη κατευθυντικών γράφων, τα οποία

Πίνακας 4.8: Σύγκριση διαφόρων τεχνικών σύντηξης ολιστικής και τοπικής πληροφορίας για υπολογισμό οριζόντιας περιστροφής κεφαλιού

Μέθοδος Σύντηξης ΔΠΑ+ΣΝΔ	Σφάλμα	STD
ΔΠΑ μόνο	5.72° (rmse) / 4.39° (mae)	4.90°
ΣΝΔ μόνο	7.07° (rmse) / 5.63° (mae)	6.45°
Σταθμ. Μέσος ΔΠΑ και ΣΝΔ	5.65° (rmse) / 4.38° (mae)	5.02°
Μπεϋζιανή σύντηξη	5.66° (rmse) / 4.29° (mae)	5.02°

είναι δένδρα ομάδων μεταβλητών. Αποτελούν δίκτυα τα οποία έχουν ως κόμβους τις κλίκες ενός αρχικού δικτύου. Κλίκες (cliques-clusters) ονομάζονται τα σύνολα των μεταβλητών που αποτελούνται από γονείς και τα παιδιά τους και είναι σημαντικές για το λόγο ότι βάσει αυτών κατασκευάζονται τα δένδρα διακλάδωσης. Οι κλίκες αποτελούν ένα τριγωνοποιημένο γράφημα, το οποίο είναι ένα ειδικό γράφημα που κατασκευάζεται σχετικά με το εκάστοτε δίκτυο. Κάθε κλίκα διατηρεί έναν πίνακα που σχετίζεται με τη μορφή των μεταβλητών της, ενώ η διαδικασία αναπαραγωγής μέσω πολλαπλασιασμού (propagation) αυτών των πινάκων εμπεριέχει μία σειρά από λειτουργίες, κατά τη διάρκεια εκτέλεσης των διαδικασιών.

Οι πίνακες υπό συνθήκη πιθανοτήτων του δικτύου εκπαιδεύτηκαν με διακριτοποίηση των μεταβλητών του. Η διαμέριση που έδωσε τα καλύτερα αποτελέσματα φαίνεται στο σχήμα 4.19. Για την εκπαίδευση εφαρμόστηκε πρωτόκολλο αφαίρεσης στιγμιοτύπου, με σκοπό τα δεδομένα ελέγχου να μην ενέχονται στην εκπαίδευση του μπεϋζιανού δικτύου στο οποίο θα ελεγχθούν. Τέλος, όπως και στην έξοδο των επιμέρους επιπέδων πληροφορίας, έτσι και η τελική έξοδος, φιλτράρεται (βλ. 4.2.3).

4.5.6 Πειραματικά αποτελέσματα: Υπολογισμός οριζόντιας περιστροφής κεφαλιού

Στον πίνακα 4.8 φαίνονται συγκεντρωτικά τα αποτελέσματα για οριζόντια περιστροφή, με χρήση τοπικής, ολιστικής πληροφορίας, σταθμισμένο μέσο, καθώς και χρήση μπεϋζιανού δικτύου σύντηξης.

Θεώρηση των αποτελεσμάτων των επιπέδων πληροφορίας και της μεταβλητής της πραγματικής γωνίας περιστροφής ως συνεχόμενες μεταβλητές κανονικής κατανομής κατέληξε σε σφάλματα στον υπολογισμό οριζόντιας περιστροφής 5.75° (rmse) / 4.51° (mae), ενώ αποδείχτηκε ότι η ταχύτητα υπολογισμού των πιθανοτήτων ήταν αρκετά χαμηλότερη σε σχέση με τη διακριτοποίηση του συνόλου των μεταβλητών στο μπεϋζιανό δίκτυο.

4.6 Συμπεράσματα

Στο κεφάλαιο αυτό μελετήθηκε το πρόβλημα υπολογισμού της οριζόντιας, κατακόρυφης, και στο επίπεδο της κάμερας περιστροφής του κεφαλιού, με χρήση παρακολούθησης χαρακτηριστικών από Διανυσματικά Πεδία Απόστασης και Συνελικτικά Νευρωνικά Δίκτυα (τοπική και ολιστική πληροφορία, αντίστοιχα). Στην πρώτη περίπτωση, προτάθηκε και μελετήθηκε ένας νέος, εύρωστος μετασχηματισμός που κωδικοποιεί γεωμετρική πληροφορία από κάθε εικονοστοιχείο των περιοχών αναζήτησης. Με αυτόν τον τρόπο, επιτυγχάνεται ευρωστία σε ποικίλες συνθήκες φωτισμού. Η παρακολούθηση χαρακτηριστικών περιοχών προσώπου συγκρίθηκε με την κλασική

μέθοδο οπτικής ροής και, για τις δοθείσες παραμέτρους, βρέθηκε ότι δίνει καλύτερα αποτελέσματα στην εκτίμηση περιστροφής προσώπου. Ακόμη, προτάθηκε μία μέθοδος παρακολούθησης προσώπου, για την επακριβή ανάκτηση των ορίων του, με σκοπό να περιορίζεται η παρακολούθηση των χαρακτηριστικών μέσα σε αυτά, αλλά και να υπολογίζονται οι παράμετροι περιστροφής προσώπου με αναφορά σε αυτά τα όρια. Η μέθοδος στηρίζεται σε προσαρμοστική κατωφλίωση, η οποία λαμβάνει υπόψη το χρώμα του δέρματος, αλλά και το μέγεθος του προσώπου του εκάστοτε χρήστη. Ικανοποιητική βελτίωση των αποτελεσμάτων (ιδιαίτερα της κατακόρυφης περιστροφής) φαίνεται να επιτυγχάνεται με χρήση χαρακτηριστικών καρέ, τα οποία εξάγονται και αποθηκεύονται κατά τη διάρκεια των πειραμάτων.

Τα όρια του προσώπου ορίζουν την περιοχή που χρησιμοποιείται για εκμετάλλευση πληροφορίας υψής (ολιστική) του προσώπου, για την ανάκτηση της περιστροφής του. Για το σκοπό αυτό, προτείνεται μια αρχιτεκτονική ταξινομητών, υλοποιημένων με Συνελικτικά Νευρωνικά Δίκτυα. Η επιλογή της λογικής των συνελικτικών βασίζεται στην εύκολη εκπαίδευσή τους, αλλά και στο γεγονός ότι δεν απαιτούν απόλυτη αντιστοίχιση χαρακτηριστικών ανάμεσα στα δεδομένα εκπαίδευσης και τα δεδομένα ελέγχου. Η τελευταία ιδιότητα είναι ιδιαίτερος σημαντική, δεδομένου ότι τα πειράματα έγιναν σε συνθήκες φυσιολογικού φωτισμού και ανεπιτήδευτων κινήσεων από μέρους των χρηστών. Ταυτόχρονα, προτείνεται ένα δίκτυο υποταξινομητών τέτοιων, που να διακρίνουν μονάχα ανάμεσα σε γειτονικές κλάσεις. Το νόημα αυτής της λογικής είναι πως, με αυτόν τον τρόπο, ελαχιστοποιείται το σφάλμα, αφενός, δεδομένου ότι ο κάθε υπο-ταξινομητής διακρίνει μόνο ανάμεσα σε δύο (γειτονικές) κλάσεις, αλλά υπάρχει και μεγαλύτερη ταχύτητα, αφού, σε κάθε καρέ ενεργοποιούνται μόνο εκείνοι οι ταξινομητές που ενέχουν στις κλάσεις τους τη γωνία που αντιστοιχεί στο προηγούμενο καρέ.

Τέλος, στο κεφάλαιο αυτό μελετήθηκαν διάφορες μέθοδοι σύντηξης του ολιστικού και τοπικού επιπέδου πληροφορίας, με σκοπό το βέλτιστο υπολογισμό περιστροφής κεφαλιού. Απλώς σταθμισμένος μέσος, μέσα από τα πειράματα έγινε σαφές ότι δίνει αρκετά καλά αποτελέσματα. Επιπλέον, έγιναν πειράματα στην ανάκτηση της οριζόντιας γωνίας και, μέσα από αυτά, φάνηκε ότι ένας "έξυπνος" τρόπος σύντηξης των δύο επιπέδων πληροφορίας με χρήση μοντελοποιημένης γνώσης περιεχομένου, μπορεί να επιφέρει αρκετά καλά αποτελέσματα σε σχέση με την τρέχουσα βιβλιογραφία. Αυτό οφείλεται στους εξής λόγους: Πρώτον, τοπικές τεχνικές εξασφαλίζουν αρκετά καλή ακρίβεια όταν η παρακολούθηση χαρακτηριστικών γίνεται με καλή ακρίβεια. Δεύτερον, οι ολιστικές δεν απαιτούν ακριβή προσδιορισμό συγκεκριμένων σημείων, με την προϋπόθεση ότι επιτυγχάνεται αποτελεσματική παρακολούθηση του περιγράμματος του προσώπου. Σύντηξη των παραπάνω πηγών πληροφορίας, ιδιαίτερα με τρόπο τέτοιο που να μοντελοποιούνται οι συνθήκες κάτω από τις οποίες υπερτερεί η μία έναντι της άλλης μεθόδου, λαμβάνοντας υπόψιν το περιεχόμενο της κίνησης, επέφερε βελτίωση του σφάλματος. Επιπλέον, συνύπαρξη των δύο πηγών πληροφορίας επιτρέπει την επιβολή κανόνων για επανένταξη του συστήματος. Ακόμη, η ολιστική πληροφορία μπορεί να χρησιμοποιηθεί καθ' αποκλειστικότητα στα πρώτα καρέ, έως ότου υπολογιστεί πολύ μικρή περιστροφή, όποτε εκκινείται και η τοπική μέθοδος. Με αυτόν τον τρόπο, απαλείφεται η απαίτηση ο χρήστης να είναι μετωπικά στραμμένος στα πρώτα καρέ, ώστε να αποκτηθεί γεωμετρία αναφοράς.

Κεφάλαιο 5

Εκτίμηση Κατάστασης-Ενασχόλησης Χρήστη με βάση οπτικά χαρακτηριστικά σε περιβάλλον επικοινωνίας Ανθρώπου-Εικονικού πράκτορα

Στο κεφάλαιο αυτό χρησιμοποιείται το σύστημα ανίχνευσης κίνησης κεφαλιού που περιγράφηκε στο κεφάλαιο 3 για να διαπιστωθεί η δυνατότητα μοντελοποίησης του ενδιαφέροντος και της προσοχής του χρήστη σε συνθήκες αλληλεπίδρασης. Καθώς στα περισσότερα προτεινόμενα συστήματα της βιβλιογραφίας η μοντελοποίηση με βάση το βλέμμα γίνεται, είτε με χρήση μαγνητικών συσκευών, είτε εξειδικευμένου εξοπλισμού (μικροκάμερες, κάμερες ευαίσθητες στο υπέρυθρο φως) σκοπός του παρόντος κεφαλαίου είναι η θεώρηση διαφορετικών σεναρίων στα οποία η μοντελοποίηση της προσοχής και του ενδιαφέροντος με βάση οπτικά χαρακτηριστικά να προέρχεται αποκλειστικά από μια απλή δικτυακή κάμερα, χωρίς πρότερη γνώση αναφορικά με τη θέση ή τα χαρακτηριστικά του χρήστη. Αρχικά, έγινε μελέτη της αντίληψης από την πλευρά του χρήστη, της συμπεριφοράς ψηφιακού πράκτορα, κατά τη διάρκεια σεναρίου προσοχής από κοινού (*shared attention scenarios*). Σκοπός της έρευνας αυτής, είναι να μελετηθούν οι παράγοντες εκείνοι που πρέπει να ληφθούν υπόψη στην εκπαίδευση μοντέλων ενασχόλησης (*engagement models*), χρησιμοποιώντας, όχι μόνο τη συμπεριφορά του χρήστη, αλλά και το ίδιο το περιεχόμενο της αλληλεπίδρασης, κάτι το οποίο εφαρμόζει απόλυτα σε σενάρια προσοχής από κοινού.

Επιπλέον, αναζητείται ο βαθμός καταλληλότητας των αλγορίθμων που περιγράφονται στα κεφάλαια 3 για την εκτίμηση ενός συνόλου καταστάσεων ατόμων σε περιβάλλοντα τέτοια, όπου το ζητούμενο είναι ο βαθμός προσοχής (*attention levels*) τους, σε κάποιο πολυμεσικό υλικό. Επίσης, ενσωματώνεται και η έννοια του βλέμματος από την κίνηση των ματιών σαν πηγή πληροφορίας, αλλά και άλλων χαρακτηριστικών κινήσεων σημείων του προσώπου. Στη βιβλιογραφία λίγες είναι οι εργασίες που ασχολούνται με την ανάκτηση του βλέμματος δια μέσω συνδυασμού περιστροφής προσώπου και κατευθυντικότητας βλέμματος, ειδικά σε περιβάλλοντα μίας, μη βαθμονομημένης μηχανής λήψης, με ανεπιτήδευτο φωτισμό και φόντο. Επιπλέον, ακόμα και σε επίπεδο υλικού, κατά τη διάρκεια συγγραφής της παρούσας διατριβής, ο συνδυασμός των δύο εισόδων εξακολουθεί να μην έχει επιτευχθεί. Γίνεται μελέτη των μη προφορικών

εκείνων παραγόντων που συμμετέχουν στην αντίληψη της έννοιας της προσοχής και, μέσω ασαφούς λογικής, προτείνονται συστήματα τέτοια που στόχο έχουν να πλησιάσουν την ανθρώπινη αντίληψη της έννοιας της προσοχής. Η αντιστοίχιση μεγεθών όπως περιστροφή προσώπου, κατευθυντικότητα βλέμματος και μεταβολές γεωμετρικών σχέσεων ανάμεσα σε χαρακτηριστικά προσώπου, σε ανώτερες έννοιες όπως αυτή της προσοχής, κατά τη διάρκεια της αλληλεπίδρασης, μοντελοποιούνται με στόχο να προταθούν συγκεκριμένα συστήματα, των οποίων οι παράμετροι αναγράφονται στο παράρτημα της διατριβής.

5.1 Ερευνητικό πλαίσιο

Οι άνθρωποι παρουσιάζουν αρκετά μεγάλη ευαισθησία στην κατευθυντικότητα και φύση του βλέμματος των άλλων, συμπεριφορά η οποία μπορεί να προσφέρει πληροφορία σχετικά με τις προθέσεις τους και τα ενδιαφέροντά τους, καθώς και να λειτουργήσει καταλυτικά στη διαχείριση αλληλεπίδρασης. Στη βιβλιογραφία παρουσιάζονται πολλές δουλειές στις οποίες έχει γίνει προσπάθεια να μοντελοποιηθούν θεωρητικά, πνευματικές και συναισθηματικές καταστάσεις βασισμένες σε συμπεριφορικούς παράγοντες, όπως το βλέμμα και έκφραση του προσώπου [78], αλλά και η στάση σώματος [112].

Ενώ τέτοιες μελέτες εστίασαν στην εκτίμηση τέτοιων παραγόντων σε καταστάσεις όπου ο χρήστης δεν αλληλεπιδρά, άλλες εστίασαν σε περιπτώσεις έντονης αλληλεπίδρασης. Για παράδειγμα, στο [24], ερευνάται αλληλεπίδραση μεταξύ παιδιών και ρομποτικών πρακτόρων, ενώ στο [79] χρησιμοποιείται πολυτροπική προσέγγιση, μαζί με πληροφορία περιεχομένου που αφορά σε στόχο παιχνιδιού, για να ταξινομηθεί το ενδιαφέρον παιδιών κατά τη διάρκεια κάποιων παιχνιδιών. Πληροφορία περιεχομένου, επίσης, είναι σημαντική, όταν πρέπει να λαμβάνονται αποφάσεις σε συνθήκες εκτάκτου ανάγκης, πλήθη, κλπ. Σε τέτοιες περιπτώσεις, η παρακολούθηση του βλέμματος αυτού με τον οποίο αλληλεπιδρά ο χρήστης, ενδέχεται να είναι κρίσιμης σημασίας [90]: σε συνθήκες κινδύνου, η παρακολούθηση του βλέμματος τρίτων μπορεί να οδηγήσει στον άμεσο εντοπισμό κινδύνων και την αποφυγή τους. Επιπλέον, σε κοινωνικές συναθροίσεις, αντικείμενα η γεγονότα αποτελούν το κέντρο της προσοχής του πλήθους, ακόμα κι αν δεν απαιτείται για την εξαγωγή κάποιας πληροφορίας. Με αυτόν τον τρόπο, «μοιράζεται» η εμπειρία. Αυτός είναι και ένας κλάδος ο οποίος μελετάται ενδελεχώς τα τελευταία χρόνια στον τομέα της ρομποτικής [36], και έχει μεγάλη σχέση με τις συμπεριφορές μιμητισμού [21]. Ο ρόλος του βλέμματος έχει επίσης μελετηθεί σε συνθήκες αλληλεπίδρασης με εικονικές οντότητες, όπως συνομιλητικούς πράκτορες (ECA's). Για παράδειγμα, πράκτορες που κάνουν παρουσιάσεις προϊόντων, λαμβάνουν υπόψιν το ενδιαφέρον του χρήστη [139], και εξαρτώνται από το βλέμμα τους. Με αυτόν τον τρόπο μπορεί να προσαρμόζεται και η συμπεριφορά των πρακτόρων. Με παρόμοιο τρόπο, στα [66] και [132] έχουν μελετηθεί συστήματα για την εκτίμηση της ενασχόλησης του χρήστη με βάση το βλέμμα, όταν «συνδιαλέγεται» με έναν εικονικό πράκτορα, ενώ στο [15], ο ρόλος του βλέμματος χρησιμοποιήθηκε για να βοηθηθεί ένας πράκτορας με ικανότητες «ακοής» να αντιδρά. Τέτοιου είδους μελέτες εξαρτώνται σε μεγάλο βαθμό, τόσο από την ικανότητα ενός συστήματος να εξάγει πληροφορία βλέμματος με αξιοπιστία, όσο και από τη φύση του ίδιου του συστήματος ώστε να μην είναι έντονα παρεμβατικό και επηρεάζει, έστω και ακούσια, τη συμπεριφορά του χρήστη.

Ενώ στη βιβλιογραφία έχει προταθεί πλειάδα εργασιών γύρω από το θέμα της

εκτίμησης των επιπέδων ενασχόλησης του χρήστη με κάποιο πολυμεσικό υλικό, αλληλεπίδραση με εικονικό πράκτορα ή ρομπότ και οδήγηση, δεν υπάρχει κάποιου είδους ενιαία αντιμετώπιση-ερευνητικό πλαίσιο. Αντίθετα, κάθε μελέτη επικεντρώνεται στη συγκεκριμένη περίπτωση που έχει να αντιμετωπίσει, ορίζοντας διαφορετικές κλάσεις καταστάσεων, και εκτιμώντας τις λαμβάνοντας υπόψη διαφορετικές ροές πληροφορίας. Επιπλέον, δεν υπάρχουν επισημειωμένα δεδομένα που να επιτρέπουν την επαλήθευση και σύγκριση μεθοδολογιών γύρω από το θέμα της ενασχόλησης-προσοχής χρήστη. Παρ' όλα αυτά, υπάρχει, έστω και σε περιορισμένο βαθμό, εννοιολογικό πλαίσιο, επάνω στο οποίο έχουν στηριχθεί πολλές έρευνες, και το οποίο διαφέρει από εφαρμογή σε εφαρμογή. Έτσι, στην περίπτωση επικοινωνίας Ανθρώπου-Εικονικού πράκτορα, έχουν οριστεί τα IScores [140], ενώ στην περίπτωση οδήγησης αυτοκινήτου, το μέτρο PERCLOS [38] έχει χρησιμοποιηθεί από αρκετές ερευνητικές ομάδες, ενώ έχουν προταθεί και αρκετές παραλλαγές του. Επιπλέον, ενώ η έρευνα γύρω από συνθήκες οδήγησης φαίνεται να ακολουθεί ανεξάρτητο δρόμο και επικεντρώνεται περισσότερο σε τεχνικά ζητήματα (μη παρεμβατικές συσκευές, συστήματα εύρυθμης λειτουργίας σε οποιοδήποτε συνθήκες φωτισμού, πλήθος εισόδων σε τέτοια συστήματα, κλπ), φαίνεται ότι, στην περίπτωση ανθρώπου-υπολογιστή, μεγάλο μέρος του εννοιολογικού πλαισίου είναι δανεισμένο από την ψυχολογία και θεωρίες σχετικά με συμπεριφορές που αφορούν σε ομαδικές συνεργατικές διαδικασίες με κοινό στόχο στο φυσικό κόσμο.

Τέτοιου είδους συμπεριφορές [115] συναντώνται συχνά όταν ο στόχος απαιτεί αναφορά ή κοινή δράση γύρω από συγκεκριμένο αντικείμενο στο φυσικό κόσμο [178]. Σε πολλές εργασίες συσχετίζεται, είτε άμεσα, είτε με έμμεσο τρόπο (λαμβάνοντας υπόψη το περιεχόμενο της αλληλεπίδρασης) η έννοια της ενασχόλησης με αυτήν της από κοινού προσοχής, η οποία, στο φυσικό κόσμο, νοείται ως εκείνη η διαδικασία κατά την οποία όλοι η συμμετέχοντες σε μια συζήτηση έχουν στραμμένη την προσοχή τους στο υπό συζήτηση αντικείμενο. Στο [6] αναφέρεται η σύγκριση της κατάστασης όπου, σε ομιλία σε ακροατήριο, δεν υπήρχε φυσικό αντικείμενο αναφοράς, με αυτήν όπου υπήρχε ένα σχετικά περίπλοκο αντικείμενο (π.χ. ένας χάρτης). Αναφέρεται ότι η χρονική διάρκεια που οι συμμετέχοντες κοίταζαν ο ένας τον άλλον έπεσε από 76,6% σε 6,4% στη δεύτερη περίπτωση. Πειράματα που περιγράφονται στο [169], έδειξαν ότι η συμπεριφορά του βλέμματος, ακόμα και με τα ίδια επίπεδα ενδιαφέροντος, έχει άμεση σχέση με τον στόχο της εκάστοτε διαδικασίας.

Αρκετή έρευνα, έχει γίνει και γύρω από θέματα σχετικά με περιβάλλοντα συνεδριάσεων, όπου, εκεί, λιγότερη έμφαση δίνεται στην εκτίμηση των επιπέδων προσοχής των συμμετεχόντων, και η μελέτη επικεντρώνεται κυρίως γύρω από τη διερεύνηση «ακρυφών» χαρακτηριστικών, όπως είναι ο ρόλος που έχει ο καθένας [125, 157]. Σε τέτοιου είδους πειράματα, συνήθως, εκτός από οπτική πληροφορία, λαμβάνονται υπόψιν και χαρακτηριστικά φωνής. Ωστόσο, σε συνθήκες συνεδριάσεων, το να έχει κάποιος ακροατής το βλέμμα του στραμμένο προς τον εκάστοτε ομιλητή είναι ένδειξη ενασχόλησης με τη συζήτηση [27], ενώ σημαντικό ρόλο παίζει το βλέμμα μεταξύ δύο συνομιλητών σαν επικοινωνιακό σήμα, στο να ανταλλάσσονται οι ρόλοι σε μια συζήτηση [41]. Πολύ σημαντικός, επίσης, σε συνεδριάσεις πολλών ατόμων, είναι και ο ρόλος του βλέμματος στο να αποδοθεί ο λόγος σε συγκεκριμένο άτομο από τον ομιλητή [5]. Πιο πρόσφατα πειράματα [170], με χρήση συσκευών παρακολούθησης βλέμματος, έδειξαν ότι το βλέμμα των συμμετεχόντων σε μία συνεδρίαση, είναι άριστο μέσο στην πρόβλεψη σχετικά με το ποιος είναι ο ομιλητής, αφού αποδείχτηκε ότι ένας ομιλητής κοιτάζεται από τον κόσμο 7 φορές περισσότερο από ό,τι οι υπόλοιποι ακροατές, ενώ,

εάν ο ομιλητής φαίνεται να απευθύνεται σε συγκεκριμένο συμμετέχοντα-αχροατή, τότε ο κόσμος κοιτάζει τον συγκεκριμένο συμμετέχοντα-αχροατή, 3 φορές περισσότερο από τους υπόλοιπους.

Παρακάτω παρουσιάζεται μία σειρά από χαρακτηριστικές προτάσεις στη βιβλιογραφία, οι οποίες μοντελοποιούν την ανθρώπινη συμπεριφορά με βάση χαρακτηριστικά βλέμματος.

5.1.1 Επικοινωνία ανθρώπου με εικονικό πράκτορα-ρομπότ

Στο πλαίσιο της επικοινωνίας Ανθρώπου-Εικονικού πράκτορα, οι συγγραφείς στο [141] ερευνούν τη σχέση μεταξύ της μη-προφορικής συμπεριφοράς του χρήστη και των επιπέδων ενασχόλησής του κατά την αλληλεπίδρασή του με εικονικό πράκτορα. Για το σκοπό αυτό, στήθηκαν πειράματα, στα οποία συμμετείχαν 10 άτομα, και από τα οποία ζητήθηκε να παρακολουθήσουν την επεξήγηση κινητών τηλεφώνων από ψηφιακό πράκτορα. Ταυτόχρονα με τους χρήστες, την αλληλεπίδραση παρακολουθούσαν 7 παρατηρητές. Και στους μεν και στους δε, δόθηκε μία συσκευή, την οποία καλούνταν να ενεργοποιήσουν κάθε φορά που οι χρήστες ένιωθαν (ή, στην περίπτωση των παρατηρητών, έκριναν ότι οι χρήστες ήταν) μη αρκετά ενασχολημένοι στη διαδικασία, λόγω του ότι την έβρισκαν ανιαρή. Στη συνέχεια, συλλέχθηκαν αυτά τα δεδομένα, μαζί με τις συντεταγμένες του βλέμματος των χρηστών στην οθόνη (τα δεδομένα βλέμματος ματιού συλλέχθηκαν με χρήση της συσκευής παρατήρησης Tobii X-120). Κατόπιν, τα δεδομένα του βλέμματος ομαδοπήθηκαν σε αντικείμενα (προϊόν προς περιγραφή, κεφάλι πράκτορα, σώμα πράκτορα, άλλα αντικείμενα) και αντιστοιχήθηκαν ακολουθίες των τριών αντικειμένων σε επίπεδα ενασχόλησης, όπως προέκυψαν από την κρίση των παρατηρητών και τις δηλώσεις των ίδιων των χρηστών. Βρέθηκε ότι υπάρχει σχέση μεταξύ ενασχόλησης-ακολουθίας βλέμματος, η οποία, μάλιστα, μπορεί να οδηγήσει σε αρκετά ασφαλή συμπεράσματα, αν λάβει κανείς υπόψη του το προφίλ του κάθε χρήστη και κάνει ξεχωριστή αποτίμηση για τον καθένα.

Στο [140], στα πλαίσια επικοινωνίας ανθρώπου-εικονικών χαρακτήρων, ορίζονται δύο κριτήρια: 1) Το σκορ ενδιαφέροντος IScore και 2) το σκορ ενδιαφέροντος FIScore. Το πρώτο αναφέρεται στην πιθανότητα ο χρήστης να ενδιαφέρεται για ένα συγκεκριμένο αντικείμενο, το οποίο και χαρακτηρίζεται ενεργό, εάν κάποια στιγμή το ISCORE ξεπεράσει κάποιο κατώφλι. Το FIScore υπολογίζει το «ποσό» ενδιαφέροντος για ένα ενεργό αντικείμενο κατά τη διάρκεια του χρόνου. Αυτό μπορεί να αναφέρεται στο κατά πόσο κάποιος, παράλληλα με το αντικείμενο υπό εξέταση, παρακολουθεί και σχετικά αντικείμενα. Το βασικό συστατικό του IScore είναι $p = T_{ISon}/T_{IS}$, με τη μεταβλητή T_{ISon} να αναφέρεται στη συνολική διάρκεια που ένα αντικείμενο έχει στραμμένο το βλέμμα του χρήστη επάνω του για χρονική διάρκεια T_{IS} . Στο [140] προτάθηκαν παραλλαγές του p , ώστε να ληφθούν υπόψη κι άλλοι παράγοντες, όπως το φυσικό μέγεθος ενός αντικειμένου, και ο αριθμός των φορών που ο χρήστης έδειξε ενδιαφέρον στο αντικείμενο αυτό (συχνότητα με την οποία κοιτάζε το αντικείμενο). Οι διαδικασίες αυτές αφορούν σε στατική μοντελοποίηση κάποιου περιβάλλοντος και δε μοντελοποιούν δυναμικής φύσεως αλληλεπιδράσεις (όπως, για παράδειγμα, μία εικονική παρουσίαση δυναμικού υλικού), όπου ο ρόλος των αντικειμένων εξαρτάται σε μεγάλο βαθμό από το ίδιο το περιεχόμενο της παρουσίασης, το ρόλο του εκάστοτε αντικειμένου, προηγούμενες εκφάνσεις του βλέμματος, κλπ.

Στη δουλειά που αναφέρεται στο [20], οι συγγραφείς κάνουν διάκριση μεταξύ στατικών και δυναμικών περιβάλλοντων. Για τη μοντελοποίηση του συνολικού ενδιαφέ-

ροντος του χρήστη σε μια εικονική παρουσίαση, αντιστοιχίζεται σε κάθε αντικείμενο, ένα Δυναμικό Bayes Δίκτυο (DBN). Για το δίκτυο αυτό, λαμβάνονται υπόψη, εκτός από το αν τη συγκεκριμένη χρονική στιγμή ο χρήστης κοιτάζει το αντίστοιχο αντικείμενο, και άλλες παράμετροι, όπως το αν την προηγούμενη χρονική στιγμή το βλέμμα ήταν στραμμένο επάνω στο αντικείμενο, και το περιεχόμενο της παρουσίασης τη δεδομένη χρονική στιγμή. Πιο συγκεκριμένα, με τον όρο περιεχόμενο, εννοείται το αντικείμενο στο οποίο γίνεται αναφορά ανά πάσα χρονική στιγμή. Κάτι τέτοιο λαμβάνεται υπόψη στο να κριθεί εάν η κατευθυντικότητα του βλέμματος είναι τέτοια που να δηλώνει ενδιαφέρον στην παρουσίαση. Έτσι, με κόμβους παρατήρησης το περιεχόμενο της παρουσίασης και τα δεδομένα του βλέμματος (όπως εξήχθη με χρήση του λογισμικού FaceLAB¹) τη δεδομένη χρονική στιγμή, και κρυφούς κόμβους το κατά πόσο ο χρήστης κοιτάζει προς το αντικείμενο, το κατά πόσο ο χρήστης ενδιαφέρεται για το αντικείμενο και το κατά πόσο ο χρήστης ενδιαφέρεται συνολικά για την παρουσίαση, έγινε μια συνολική μοντελοποίηση της ενασχόλησης του χρήστη με μοναδική είσοδο δεδομένα από την κατευθυντικότητα του βλέμματός του.

Στο [101], οι συγγραφείς χρησιμοποιούν ένα σύστημα δύο καμερών για την ανάκτηση της συνολικής κατευθυντικότητας του βλέμματος, με βάση, τόσο την περιστροφή του κεφαλιού, ταιριάζοντας ένα τρισδιάστατο μοντέλο χαρακτηριστικών προσώπου, όσο και μελετώντας την τρισδιάστατη γεωμετρία του ματιού. Από αυτές τις τροπικότητες, οι συγγραφείς εξάγουν το συνολικό διάνυμα κατευθυντικότητας βλέμματος. Η μοντελοποίηση της προσοχής του χρήστη γίνεται καθαρά σε τοπολογικό επίπεδο, χωρίζοντας τη σκηνή σε περιοχές (οθόνη, κάμερα, παράθυρο, τοίχος, κλπ), αλλά και με χρήση του αλγορίθμου που περιγράφεται στο [159] για κατηγοριοποίηση των κινήσεων του κεφαλιού σε λέξεις.

Στο [152], οι συγγραφείς, στα πλαίσια της μελέτης τους σχετικά με θέματα ενασχόλησης στην επικοινωνία ανθρώπου-ρομπότ, ερευνούν τις κινήσεις του κεφαλιού εκείνες, οι οποίες παίζουν πρωταρχικό ρόλο στην αντίληψη ότι ο συνομιλητής είναι ενασχολημένος με την αλληλεπίδραση. Δανειζόμενοι στοιχεία από την μεταξύ ανθρώπων επικοινωνία, μελέτησαν, εκτός από ηχητικά σήματα, εκείνες τις κινήσεις του προσώπου που μπορεί να δηλώνουν ενασχόληση στο αντικείμενο προς συζήτηση. Τέτοιες κινήσεις είναι καταφατικά νεύματα, κατεύθυνση βλέμματος προς το αντικείμενο συζήτησης, βλέμμα προς τον ίδιο τον εικονικό πράκτορα. Ενδιαφέρον παρουσιάζει το ότι, τουλάχιστον στην αρχή της αλληλεπίδρασης, οι συγγραφείς αντιμετωπίζουν τα γενικά βλέμματα των χρηστών στο χώρο, ως ενδείξεις ενασχόλησης, αφού είναι φυσιολογικό να υπάρχει ανάγκη διερεύνησης του «χώρου» στο ξεκίνημα κάθε αλληλεπίδρασης.

5.1.2 Μοντελοποίηση ανθρώπινης συμπεριφοράς σε συνθήκες οδήγησης

Στο [35], οι συγγραφείς εξάγουν χαρακτηριστικά για τη μοντελοποίηση των επιπέδων εγρήγορσης ανθρώπων σε συνθήκες οδήγησης. Για το σκοπό αυτό, και με τη χρήση μιας κάμερας τοποθετημένης σε κεντρικό σημείο στο ταμπλό του αυτοκινήτου, οι συγγραφείς προτείνουν αρχικά μία μεθοδολογία για την εύρεση των ματιών, ενώ, ακολούθως, χρησιμοποιούν τη διάρκεια κατά την οποία τα μάτια είναι κλειστά, καθώς και τη συχνότητα που ανοιγοκλείνουν τα μάτια, για να εκτιμήσουν το κατά πόσο

¹<http://www.seeingmachines.com/product/facelab/>

κάποιος οδηγός είναι προσεκτικός ή όχι. Θεωρώντας ότι είναι γνωστή μία αρχική εκτίμηση της απόστασης του οδηγού από την κάμερα, οι συγγραφείς κάνουν αναζήτηση περιοχών ματιών σε όλη την εικόνα. Η εύρεση γίνεται με χρήση μετασχηματισμού Hough [12] για μεμονωμένες ίριδες και, ακολούθως, χρήση πληροφορίας συμμετρίας για το άλλο μάτι. Οι υποψήφιας περιοχές επαληθεύονται με νευρωνικά δίκτυα. Έξοδος του Νευρωνικού Δικτύου ως μη-μάτια για όλο το καρέ σημαίνει, είτε ότι τα μάτια είναι κλειστά, είτε ότι έχουν αποκρυφτεί. Με βάση την έξοδο αυτή, οι συγγραφείς μοντελοποιούν τη συμπεριφορά του οδηγού. Έχοντας χρησιμοποιήσει ακολουθίες 20 λεπτών από δύο άτομα, μοντελοποιήθηκε η διάρκεια και η συχνότητα κλεισίματος των ματιών (σε αντιδιαστολή με το μέτρο PERCLOS που υπολογίζει σχέση διάρκειας ανοιχτών σε σχέση με κλειστά μάτια). Η μοντελοποίηση έγινε με σκοπό να εξαχθεί ένα είδος προφίλ για κάθε χρήστη, το οποίο θα χαρακτηρίζει τη συμπεριφορά του σε φυσιολογικές συνθήκες. Το προφίλ μοντελοποιείται με γκαουσιανές κατανομές, και οι συνθήκες οι οποίες συγκρίθηκαν είναι περίοδοι κατά τις οποίες ο χρήστης μιλάει στο κινητό, συνομιλεί με άλλους επιβάτες, ή κάνει άλλες δραστηριότητες.

Στο [154], οι συγγραφείς κάνουν διαχωρισμό μεταξύ της έννοιας της προσοχής (attention) και εγρήγορης (alertness), χαρακτηρίζοντας την πρώτη, ως εκείνη την κατάσταση όπου ο οδηγός κοιτάζει μπροστά του (χωρίς, όμως, να μπορεί να γίνει διάκριση σχετικά με τον αν όντως είναι συγκεντρωμένος, αφού αυτό απαιτεί άλλου είδους χαρακτηριστικά) και την δεύτερη, την κατάσταση εκείνη όπου ο οδηγός παρουσιάζει σημάδια κόπωσης και έχει τάση να κοιμηθεί ή άλλα παρόμοια συμπτώματα. Για το λόγο αυτό, στο συγκεκριμένο άρθρο, γίνεται προσπάθεια συγκεντρωτικής χρήσης μιας σειράς χαρακτηριστικών, η οποία κυρίως χρησιμοποιεί την περιστροφή του κεφαλιού και το άνοιγμα-κλείσιμο των ματιών. Οι συγγραφείς, βασισμένοι σε μελέτες [138] ότι μεγάλος αριθμός ατυχημάτων δεν οφείλεται αναγκαστικά σε κούραση, αλλά σε απλή αλλαγή βλέμματος, ενδεχομένως, επεκτείνουν τα χαρακτηριστικά πέρα από το PERCLOS. Για τα πειράματα χρησιμοποίησαν 8 ακολουθίες ατόμων, σε συνθήκες πραγματικής και υποτιθέμενης οδήγησης, ενώ τα πειράματα έγιναν σε συνθήκες ημέρας. Το σύστημα βασίζεται σε παρακολούθηση ματιών και στόματος. Για τον εντοπισμό και παρακολούθηση των χαρακτηριστικών, χρησιμοποιήθηκαν μοντέλα χρωματικότητας [82] και μορφολογικοί μετασχηματισμοί, καθώς και αφφινική παρακολούθηση των ματιών, στην περίπτωση που χρωματικά μοντέλα δεν έδιναν αποτελέσματα τέτοια που να ικανοποιούν συγκεκριμένους γεωμετρικούς περιορισμούς. Χρωματικά μοντέλα χρησιμοποιήθηκαν και για την παρακολούθηση του υπόλοιπου προσώπου. Για το χαρακτηρισμό περιστροφής κεφαλιού παρατηρήθηκαν οι αλλαγές στις αποστάσεις μεταξύ χαρακτηριστικών και ορίων προσώπου, ενώ το άνοιγμα-κλείσιμο υπολογίστηκε με βάση τα φωτεινότερα εικονοστοιχεία στην περιοχή του ματιού, ενώ έγινε προσέγγιση της προβολής της θέσης των ματιών στο πίσω μέρος του κεφαλιού για ανάκτηση γωνίας περιστροφής. Τα παραπάνω χρησιμοποιήθηκαν για τη δημιουργία τριών μηχανών πεπερασμένων καταστάσεων. Τα χαρακτηριστικά που χρησιμοποιήθηκαν για τη μετάβαση μεταξύ των καταστάσεων ήταν η περιστροφή του κεφαλιού και το άνοιγμα-κλείσιμο των ματιών, ενώ η τελική απόφαση σε σχέση με την κατάσταση του οδηγού δίνεται από τη σύνθεση των τριών μηχανών. Ωστόσο, όπως τονίζουν και οι ίδιοι οι συγγραφείς, δεν υπάρχουν καθολικά κατώφλια τιμών για τη μετάβαση μεταξύ των καταστάσεων, ώστε να προταθεί συγκεκριμένο μοντέλο.

Οι συγγραφείς στη δουλειά που παρουσιάζεται στο [40] αναλύουν την ανθρώπινη, μη προφορική συμπεριφορά όπως μπορεί να μοντελοποιηθεί με χρήση κατευθυντικότητας βλέμματος ματιών και περιστροφής κεφαλιού. Πιο συγκεκριμένα, η δουλειά

αυτή βασίζεται σε προηγούμενες μελέτες αναφορικά με τον τρόπο που ο άνθρωπος χειρίζεται το βλέμμα του σε διαφορετικά είδη καταστάσεων [69, 116, 134, 77, 145] και, ιδιαίτερα, για καταστάσεις στις οποίες ο χρήστης συνειδητά και με σκοπό αλλάζει το βλέμμα του προς μια συγκεκριμένη κατεύθυνση, και καταστάσεις όπου απρόσμενα ερεθίσματα τον παρότρυναν να κοιτάξει προς τα κάπου (είτε πρόκειται για ενδιαφέρον ερέθισμα, είτε λόγω, απλά, διάσπασης της προσοχής). Για το σκοπό αυτό, οι συγγραφείς πραγματοποίησαν δύο πειράματα: Ένα σε «έξυπνο δωμάτιο» και ένα σε συνθήκες οδήγησης. Στην πρώτη περίπτωση παρακολούθησαν την κίνηση των ματιών και του κεφαλιού με χρήση στερεοσκοπικών μεθόδων και στη δεύτερη με χρήση οπτικής ροής για το κεφάλι και χειροκίνητης επισημείωσης για τα μάτια. Και τα δύο πειράματα έδειξαν πως, όταν υπήρχε ηθελημένος σκοπός (πχ, αλλαγή λωρίδας σε συνθήκες οδήγησης), η κίνηση του κεφαλιού προηγούνταν αυτής των ματιών, ενώ για απρόβλεπτα ερεθίσματα, τα μάτια στρέφονταν πρωτίστως στην κατεύθυνση του ερεθίσματος και ακολουθούσε το κεφάλι. Αντίστοιχα πειράματα και μελέτες έγιναν και στα [106, 128]. Η πιθανή εξήγηση για αυτούς τους τύπους κινήσεων είναι ότι, υποσυνειδητά, σε περιπτώσεις γνωστών κινήσεων (στόχων) ο χρήστης γνωρίζει ότι θα χρειαστεί να κινήσει το κεφάλι του (λόγω της περιορισμένης γωνίας όρασης που έχουν τα μάτια από μόνα τους), οπότε είναι η πρώτη κίνηση που κάνει. Σε περίπτωση απρόβλεπτων ερεθισμάτων, ο άνθρωπος, από τη μία δεν γνωρίζει κατά πόσο θα πρέπει να στρέψει το βλέμμα του και, από την άλλη, η κίνηση των ματιών είναι πολύ εύκολο να πραγματοποιηθεί, καθώς και πολύ πιο σύντομη και, συνεπώς «προτιμάται» σε ξαφνικές και μη γνωστές καταστάσεις.

5.1.3 Βιβλιογραφική επισκόπηση μεθόδων κατευθυντικότητας βλέμματος ματιών

Στο [174], οι συγγραφείς παρουσιάζουν μία μέθοδο με σκοπό τον υπολογισμό της περιοχής εστίασης προσοχής, παίρνοντας ως είσοδο, τόσο την πόζα του κεφαλιού, όσο και την περιστροφή των ματιών. Για τον υπολογισμό της περιστροφής του κεφαλιού, οι συγγραφείς χρησιμοποιούν, είτε ελαστικούς γράφους [87], είτε μία τεχνική ταιριάσματος μοντέλων [49], η οποία χρησιμοποιεί χαρακτηριστικά προσώπου και τον άξονα συμμετρίας του προσώπου. Και με τις δύο τεχνικές, επιδιώκεται υπολογισμός μονάχα της περιστροφής γύρω από τον κάθετο άξονα. Για την εκτίμηση της κατευθυντικότητας των ματιών, χρησιμοποιείται συνέλιξη με φίλτρα Gabor, και χρησιμοποιείται η φάση:

$$phase = atan2(I * G_{sin}, I * G_{cos}) \quad (5.1)$$

όπου I είναι η εικόνα εισόδου, $*$ ο τελεστής συνέλιξης, και G_{sin} , G_{cos} τα φανταστικά και πραγματικά μέρη του φίλτρου Gabor. Το μήκος κύματος των φίλτρων επιλέχθηκε να έχει το μέγεθος των ματιών. Για τη σύνδεση της περιστροφής του κεφαλιού με την κατευθυντικότητα των ματιών, δημιουργήθηκε ένας πίνακας αναζήτησης, ο οποίος αποτελείται από την φάση (εξίσωση 5.1) και την περιστροφή του κεφαλιού. Με αυτόν τον τρόπο, συμπεραίνεται μία καθολική εκτίμηση της εστίασης προσοχής του χρήστη. Η μέθοδος ελέγχθηκε σε μία βάση που αποτελείται από 5 άτομα. Η βάση συνοδεύεται από την καθολική αλήθεια σε ό,τι αφορά την πραγματική περιστροφή του κεφαλιού, η οποία εξήχθη με τη χρήση ενός δείκτη λέιζερ, στερεωμένος στο κεφάλι του καθενός, αλλά και την κατευθυντικότητα του βλέμματος. Οι συγγραφείς αναφέρουν σφάλμα της τάξεως των 10° μόνο στην εκτίμηση περιστροφής

κεφαλιού, και παρόμοια αποτελέσματα στον έλεγχο όλου του συστήματος συνολικά (περιστροφή κεφαλιού και κατευθυντικότητα βλέμματος).

Στο [83], οι συγγραφείς, μέσα από μία σειρά ευριστικών κανόνων βασισμένων στη χρωματικότητα και φωτεινότητα της ευρύτερης περιοχής των ματιών, και της ίριδας, ανιχνεύουν το κέντρο του ματιού μέσα στην περιοχή του προσώπου. Κατόπιν, βασισμένοι στη χρωματικότητα της περιοχής του ματιού, ξεχωρίζουν το άνοιγμα από το κλείσιμο του ματιού, ενώ ξεχωρίζουν εάν πρόκειται με βεβαιότητα για κλείσιμο, και όχι για βλέμμα προς τα κάτω, βασισμένοι σε χρονικά χαρακτηριστικά. Αντιστοίχως, χρησιμοποιούν φωτεινότητα και χαρακτηριστικά χρώματος για να ελέγξουν βλέμμα προς τα επάνω ή δεξιά-αριστερά. Στην τελευταία περίπτωση, οι συγγραφείς συγκρίνουν τη θέση της ίριδας με αυτή του μέσου της περιοχής του προσώπου, όπως υπολογίστηκε από τον αλγόριθμο εντοπισμού προσώπου, υπονοώντας, όμως, έτσι, ότι η προτεινόμενη μέθοδος προϋποθέτει το πρόσωπο να είναι προσανατολισμένο προς την κάμερα. Επιπλέον, οι συγγραφείς αναφέρουν ότι τα διάφορα κατώφλια μαθαίνονται κατά τη διάρκεια εκπαίδευσης για κάθε χρήστη, αφήνοντας να εννοηθεί ότι απαιτείται κάποιο είδος βαθμονόμησης. Ποιοτική ανάλυση των αποτελεσμάτων, τέλος, γίνεται, με τη χρήση δεδομένων που συλλέχθηκαν για τους σκοπούς της συγκεκριμένης διατριβής.

Οι συγγραφείς, στη δουλειά που αναλύεται στο [160], χρησιμοποιούν ολιστική πληροφορία από τη γειτονιά του ματιού για να εκτιμήσουν την κατευθυντικότητα του βλέμματος σε επίπεδο μοίρας. Χωρίς να γίνεται άμεσα προφανές, θεωρείται δεδομένο ότι το κεφάλι είναι στραμμένο εμπρόσθια ως προς την κάμερα. Οι συγγραφείς ανέπτυξαν ένα σύνολο δεδομένων για τους σκοπούς των πειραμάτων τους, ζητώντας από τρεις χρήστες να παρακολουθούν έναν κέρσορα σε μια οθόνη και, τη στιγμή που θεωρούσαν ότι το βλέμμα είναι σταθερό, το σύστημα κατέγραφε τον χρήστη, κρατώντας την θέση του κέρσορα τη στιγμή εκείνη. Αξίζει να ειπωθεί ότι χρησιμοποιήθηκε κάμερα ευαίσθητη, τόσο σε ορατό, όσο και σε υπέρυθρο φως, με κύριο στόχο την αύξηση της αντίθεσης στην περιοχή γύρω από τα μάτια. Χρησιμοποιώντας συγκεκριμένο κατώφλι στη φωτεινότητα, οι συγγραφείς εντόπισαν το μεγαλύτερο συνεκτικό κομμάτι και θεώρησαν ότι αντιστοιχεί στην περιοχή του ματιού. Με καταγεγραμμένες εικόνες ματιών σε διάσταση 200×200 , όσο δηλαδή το μέγεθός τους, οι συγγραφείς δημιούργησαν μία καμπύλη σε χώρο 40000 διαστάσεων και για κάθε καινούρια είσοδο, υπολόγιζαν τα σημεία της καμπύλης που είναι κοντύτερα σε αυτήν. Με παλινδρόμηση υπολόγιζαν εκείνα τα βάρη που, από τα σημεία της καμπύλης, προκύπτει η καινούρια εικόνα. Τέλος, χρησιμοποιώντας τα ίδια βάρη στις κατευθυντικότητες βλέμματος που αντιστοιχούν στα σημεία της καμπύλης, υπολόγιζαν την κατευθυντικότητα του βλέμματος της καινούριας εικόνας. Οι συγγραφείς, αναφέρουν ότι, παρόλο που οι περιοχές των ματιών δεν εντοπίζονταν με αυστηρά όρια (σε κάποιους χρήστες ενσωματώνονται και τα φρύδια), ωστόσο, τα αποτελέσματα είναι αρκετά επιτυχημένα.

Στη δουλειά που περιγράφεται στο [76], οι συγγραφείς περιγράφουν ένα σύστημα για τον εντοπισμό χρονικών περιόδων κατά τις οποίες, ο οδηγός ενός αυτοκινήτου είναι κουρασμένος ή αφηρημένος από την οδήγησή του. Για το σκοπό αυτό, χρησιμοποιούν τρία χαρακτηριστικά: το άνοιγμα/κλείσιμο των βλεφάρων, την περιστροφή του κεφαλιού και την κατευθυντικότητα των ματιών. Το προτεινόμενο σύστημα χρησιμοποιεί δύο κάμερες, η μία εκ των οποίων είναι ευρυγώνια και χρησιμοποιείται για την παρακολούθηση του προσώπου, ενώ η άλλη είναι στενής γωνίας θέασης, και εστιασμένη στα μάτια, για την παρακολούθηση της κίνησης των βλεφάρων. Το όλο σύστημα βασίζεται στον εντοπισμό και παρακολούθηση της κόρης των οφθαλμών. Ο

εντοπισμός γίνεται με χρήση του φαινομένου της φωτεινής κόρης οφθαλμών, όταν προσπίπτει φως κύματος κοντά στην υπέρυθρη ακτινοβολία. Για τον αποτελεσματικό εντοπισμό των κορών, χρησιμοποιήθηκε σύστημα LEDs τέτοιο ώστε, ανά δύο καρέ, να μην προσπίπτει υπέρυθρο φως. Με αφαίρεση καρέ, επιτεύχθηκε αποτελεσματική εξαγωγή των θέσεων της ίριδας. Για την παρακολούθηση των θέσεων της κόρης των ματιών, οι συγγραφείς έκαναν χρήση φίλτρων Kalman. Καθώς η κίνηση των βλεφάρων είναι ένα από τα χαρακτηριστικά που χρησιμοποιούν οι συγγραφείς, υπολόγισαν, σαν κριτήρια της κούρασης του οδηγού, δύο δείκτες: τον δείκτη PERCLOS [38] και τον προτεινόμενο δείκτη AECS σε χρονικά παράθυρα των 30 δευτερολέπτων. Ο δείκτης PERCLOS είναι μία ποσοτική ένδειξη του ποσοστού κατά το οποίο τα μάτια είναι κλειστά, ενώ ο δείκτης AECS μοντελοποιεί την ταχύτητα κίνησης των βλεφάρων. Και οι δύο δείκτες εξήχθησαν βάσει του ποσοστού της κόρης που γίνεται ορατή, σε σχέση με κατάσταση μη-κούρασης. Για την περιστροφή του κεφαλιού, οι συγγραφείς χρησιμοποιούν τις σχετικές θέσεις, το σχήμα και τη φωτεινότητα των κορών των ματιών, ενώ για την κατευθυντικότητα του βλέμματος, βασίζονται στην απόσταση της κόρης από την αντανάκλαση κανονικού φωτός στο μάτι [62], [63].

Στο [98], οι συγγραφείς προτείνουν μια μέθοδο που βασίζεται αποκλειστικά σε μεθοδολογίες επεξεργασίας εικόνας, για τον υπολογισμό της κατευθυντικότητας του βλέμματος. Τα βήματα του αλγορίθμου που χρησιμοποιούν αποτελούνται από εντοπισμό και παρακολούθηση προσώπου και, ακολούθως, ανίχνευση των περιοχών των οφθαλμών. Για τον εντοπισμό των οφθαλμών χρησιμοποιούνται απλές ανθρωπομετρικές μετρήσεις, με σκοπό να γίνει μία αρχική εκτίμηση της περιοχής τους. Μέσα σε αυτές τις περιοχές γίνεται ανίχνευση κίνησης, ενώ εντοπίζονται και οι θέσεις των γωνιών των ματιών με σύγκριση με πρότυπα εικονίδια. Με βάση τις θέσεις των γωνιών των ματιών βρίσκεται το κέντρο. Ακολούθως, και με βάση την πιο αναλυτική περιοχή του κάθε ματιού, συγκρίνονται οι περιοχές του αριστερού και του δεξιού ματιού, μετά την αντιστροφή του ενός. Το μέτρο της κατευθυντικότητας του βλέμματος παίρνει μεγαλύτερες τιμές, όσο μεγαλύτερη είναι αυτή η διαφορά που προκύπτει από τη σύγκριση, ενώ το πρόσημο της διαφοράς παρέχει πληροφορία και σχετικά με τη φορά του βλέμματος. Ο αλγόριθμος δεν παρέχει δυνατότητα υπολογισμού κατακόρυφης συνιστώσας βλέμματος, ενώ θεωρείται ότι το πρόσωπο είναι συνεχώς στραμμένο προς την κάμερα. Όπως αναφέρουν οι συγγραφείς, βρέθηκε ότι τα μάτια πρέπει να είναι περιστραμμένα τουλάχιστον για 30° για να εξαχθούν ασφαλή συμπεράσματα. Για τους σκοπούς της συγκεκριμένης δουλειάς δημιουργήθηκε ένα σύνολο δεδομένων, στο οποίο αναφέρονται αποτελέσματα σχετικά με δεξιά ή αριστερή κατευθυντικότητα του βλέμματος. Η συνολική απόδοση του συστήματος (λαμβάνοντας υπόψη, τόσο την εύρεση των ματιών, όσο και το σωστό υπολογισμό της κατευθυντικότητας του βλέμματος) είναι 79.7%. Για το χειρισμό περιπτώσεων μη ισορροπημένου φωτισμού, οι συγγραφείς προτείνουν μικρή προσαρμογή στις τιμές της φωτεινότητας. Στην περίπτωση αυτή, η συνολική απόδοση του συστήματος είναι 72.6%.

Στην εργασία [167], οι συγγραφείς χρησιμοποιούν την έννοια των ισόφωτων [166] για τον εντοπισμό και παρακολούθηση των κέντρων και γωνιών των ματιών. Τα ισόφωτα είναι καμπύλες που ενώνουν γειτονικά εικονοστοιχεία ίσης φωτεινότητας. Σύμφωνα με τους συγγραφείς, θα μπορούσαν να παρομοιαστούν με καμπύλες οι οποίες υπολογίστηκαν με χωρισμό της εικόνας σε διακριτά επίπεδα φωτεινότητας. Με χρήση αυτής της ιδέας, οι συγγραφείς εντοπίζουν κέντρα και άκρα ματιών, χρησιμοποιώντας σαν ισόφωτες καμπύλες την ίριδα και τον κερατοειδή. Οι αποστάσεις του κέντρου του κάθε ματιού από τα άκρα του, είναι η πληροφορία που χρησιμοποιούν

οι συγγραφείς για τον υπολογισμό της κατευθυντικότητας του βλέμματος. Για την αποτελεσματική εκμετάλλευση των αποστάσεων αυτών, γίνεται βαθμονόμηση, ώστε να αντιστοιχιστούν με συγκεκριμένα σημεία σε μια οθόνη. Ενώ η μέθοδος δε συνεπάγεται εξοπλισμό που ο χρήστης πρέπει να φοράει, και τα αποτελέσματα φαίνονται ενθαρρυντικά, αναφέρεται ότι υπάρχει μεγάλη ευαισθησία, ακόμα και σε μικρές κινήσεις του κεφαλιού, και ότι σε τέτοιες περιπτώσεις απαιτείται εκ νέου βαθμονόμηση.

5.2 Προσοχή και ενασχόληση με χρήση κίνησης κεφαλιού

Ο όρος *σενάριο προσοχής από κοινού*, στη βιβλιογραφία, αναφέρεται στην κοινή εμπειρία δύο ή περισσότερων συμμετεχόντων, καθώς παρατηρούν αντικείμενα ή γεγονότα μαζί, και αναγνωρίζεται σαν καίρια ικανότητα της διαδικασίας κοινωνικής κατανόησης του ανθρώπου (social understanding) [13, 146]. Ο μηχανισμός αυτός είναι ουσιαστικής σημασίας για τον εφοδιασμό της μηχανής με θεμελιώδεις κανόνες κατανόησης της κοινωνικής συμπεριφοράς του χρήστη. Επιπλέον, αποτελεί συνδυαστικό κρίκο μεταξύ χαμηλού επιπέδου τροπικότητων, όπως το βλέμμα ή/και η έκφραση, με λεπτομέρειες περιεχομένου, με σκοπό την καλύτερη κατανόηση της αλληλεπίδρασης. Αυτό είναι αναγκαίο καθώς, σενάρια από κοινού προσοχής, δεν απαιτούν μονάχα μια απομονωμένη εκτίμηση της κατάστασης του χρήστη, αλλά τη συσχέτιση της κατάστασής του με κάποιον άλλο χρήστη, ή/και το περιβάλλον. Για παράδειγμα, σε πολλά συστήματα, για να συμπεραίνεται ενδιαφέρον από πλευράς κάποιου χρήστη προς έναν πράκτορα, αρκεί ο χρήστης να κοιτάζει προς τον πράκτορα. Σε σενάρια από κοινού προσοχής, ωστόσο, ενδέχεται να ισχύει το αντίθετο: Το συνεχόμενο βλέμμα προς κάποιον πράκτορα, ο οποίος περιγράφει κάποιο αντικείμενο, κοιτάζοντάς το ή δείχνοντάς το, μπορεί να σημαίνει έλλειψη ενδιαφέροντος, ή προσοχής προς τα λεγόμενα του πράκτορα. Η θεώρηση τέτοιων καταστάσεων, έδωσε ώθηση στην αντίστοιχη έρευνα, ώστε να θεωρηθούν εκείνοι οι παράγοντες που συνεισφέρουν στις διάφορες παραμέτρους τις έννοιες *ενασχόληση*, όπως αναφέρεται στη βιβλιογραφία [133], με σκοπό τη δημιουργία πιο πολύπλοκων μοντέλων.

Στη συγκεκριμένη διατριβή μελετάται το σενάριο προσοχής από κοινού, μεταξύ ενός χρήστη, ενός ψηφιακού πράκτορα και μίας σειράς εικονικών αντικειμένων στην οθόνη του υπολογιστή. Κατά τη διάρκεια που ο πράκτορας περιέγραφε στον χρήστη τα εικονικά αντικείμενα, υπολογιζόταν το διάνυσμα περιστροφής κεφαλιού, με τη βοήθεια μίας απλής δικτυακής κάμερας. Συνεπώς, ο στόχος της συγκεκριμένης δουλειάς, ήταν διττής σημασίας:

1. Κατά πρώτον, η εκτίμηση της καταλληλότητας του συστήματος εκτίμησης περιστροφής κεφαλιού, ως ένδειξη κατευθυντικότητας βλέμματος σε σενάριο από κοινού προσοχής, όπου μελετάται η κατευθυντικότητα βλέμματος.
2. Δεύτερον, να χρησιμοποιηθεί το διάνυσμα περιστροφής κεφαλιού ως μέσο έρευνας της αντίληψης των χρηστών όσον αφορά τον πράκτορα σε τέτοιου είδους σενάρια.

Το διάνυσμα περιστροφής κεφαλιού, ανάλογα με τις τιμές που παίρνει, αντιστοιχίζεται σε συντεταγμένες επάνω στην οθόνη του υπολογιστή. Αυτό είναι απλά το πρώτο επίπεδο συμπερασματολογίας σχετικά με την ενασχόληση του χρήστη. Πέρα

από αυτήν την αντιστοίχιση, απαιτούνται επιπλέον επίπεδα, κάθε ένα από τα οποία είναι και γενικότερο από το προηγούμενό του, παρέχοντας πληροφορίες προερχόμενες από δεδομένα μεγαλύτερης χρονικής διάρκειας από το προηγούμενό του [131], ξεκινώντας, για παράδειγμα, με την ικανότητα να εντοπιστούν αντικείμενα τα οποία ο χρήστης κοιτάζει μια δεδομένη στιγμή, μέχρι το επίπεδο αποτίμησης της συνολικής ενασχόλησης του χρήστη με τη θεώρηση αντικειμένων σε σχέση με το περιεχόμενο του διαλόγου.

5.2.1 Αντιστοίχιση με συντεταγμένες οθόνης

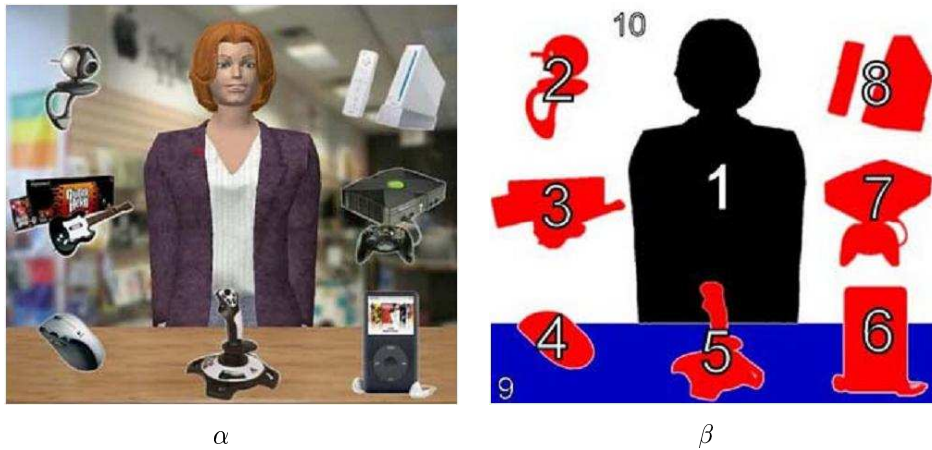
Για τη διαδικασία μετατροπής προσανατολισμού κεφαλιού σε συντεταγμένες οθόνης, υλοποιήθηκε μία διαδικασία βαθμονόμησης, η οποία καλείται στην αρχή της συνεδρίας με τον κάθε χρήστη. Συνοπτικά, ζητείται από τον χρήστη να στρέψει το κεφάλι του προς οκτώ χαρακτηριστικά σημεία της οθόνης: Δεξί, αριστερό, πάνω, κάτω όριο, καθώς και τις τέσσερις γωνίες της οθόνης. Καταγράφηκαν οι τιμές του διανύσματος περιστροφής και, κατά τη διάρκεια της αλληλεπίδρασης με τον ενσωματωμένο συνομιλητικό πράκτορα (Embodied Conversational Agent - ECA), οι υπολογιζόμενες τιμές του συγκρίνονται με τις καταγεγραμμένες που αντιστοιχούν στα χαρακτηριστικά σημεία και, έτσι, εξάγονται οι συντεταγμένες. Είναι χαρακτηριστικό, όμως, ότι η βαθμονόμηση αυτή, ακόμη και αν γίνει για διαφορετικά άτομα από τον εκάστοτε χρήστη, αρκεί για να γίνεται ικανοποιητική αντιστοίχιση μεταξύ περιστροφής κεφαλιού και συντεταγμένων οθόνης. Συνεπώς, για να αποφευχθεί η διαδικασία βαθμονόμησης για κάθε συμμετέχοντα, η αντιστοίχιση μπορεί να γίνει εκ των προτέρων.

Σε κάθε καρέ, με βάση το διάνυσμα περιστροφής κεφαλιού, υπάρχουν δύο γενικές περιπτώσεις: ο χρήστης να κοιτάζει εντός ή εκτός της οθόνης. Η περιοχή της οθόνης θεωρείται το πεδίο δράσης, μέσα στο οποίο συμβαίνουν τα γεγονότα και τα αντικείμενα της διάδρασης. Εάν ο χρήστης δεν κοιτάζει μέσα στο πεδίο αυτό, μπορεί να θεωρηθεί ότι, είτε δεν ασχολείται με το σενάριο, είτε δεν τον ενδιαφέρει, έστω και σε κάποιο βαθμό. Η δομή δεδομένων που περιέχει τις τελικές συντεταγμένες του διανύσματος περιστροφής, περιέχει και μία δυαδική μεταβλητή που δηλώνει εάν ο χρήστης κοιτάζει εντός ή εκτός του πεδίου δράσης. Επιπλέον, το αντίστοιχο σημείο της οθόνης σε x,y συντεταγμένες (εφόσον ο χρήστης κοιτάζει εντός της οθόνης) αποτελεί μέρος της δομής.

Μπορούν να ληφθούν υπόψη και τα διαφορετικά όρια της οθόνης και τους αποδόθηκε διαφορετικό νόημα. Αυτό έγινε με σκοπό να συλλεχθούν περισσότερα δεδομένα σχετικά με το λόγο της μη-ενασχόλησης του χρήστη με το σενάριο σε κάποια δεδομένη στιγμή. Για παράδειγμα, κάποιος χρήστης ο οποίος, για κάποια στιγμή κοιτάζει επάνω από το ανώτατο όριο της οθόνης, μπορεί να θεωρηθεί ότι, ενώ είναι μη-προσηλωμένος, ο λόγος του ενδέχεται να μην οφείλεται σε αδιαφορία, αλλά στο ότι σκέφτεται. Κάτι τέτοιο αποτελεί έναν ενδογενή παράγοντα μη-προσήλωσης σε κάτι, ο οποίος διαφοροποιεί το είδος μη-προσήλωσης σε σχέση με αυτά που οφείλονται σε εξωγενείς παράγοντες. Δυστυχώς, δε στάθηκε δυνατό, ωστόσο, να μοντελοποιηθούν τέτοιες περιπτώσεις, καθώς δεν υπήρχαν επαρκή δεδομένα.

5.2.2 Κατευθυντικότητα και επίπεδα προσήλωσης

Σε ένα κατώτατο επίπεδο, ορίζεται η μετρική της *κατευθυντικότητας*, σαν έννοια που αναφέρεται στον στιγμιαίο προσανατολισμό του χρήστη προς κάποια οντότητα ή



Σχήμα 5.1: Απεικόνιση του σεναρίου εικονικού πωλητή και αντικειμένων προς πώληση και των περιεχομένων σκηνής με συμβολισμό Αντικειμένων Εικονικής Προσήλωσης. Κάθε Αντικείμενο Εικονικής Προσήλωσης κρατάει ιστορικό των χρονικών στιγμών και το βαθμό στον οποίο κάποιος χρήστης ενδιαφέρθηκε για αυτό.

αντικείμενο. Αυτό βασίζεται στη μελέτη που έγινε με τα συστήματα ελέγχου προσανατολισμού βλέμματος, κεφαλιού και σώματος [13], και τη σχετική δουλειά [90, 42], η οποία βασίζεται σε νευροφυσιολογικές ενδείξεις [130]. Για παράδειγμα, εάν ο χρήστης κατευθύνει το βλέμμα του προς ένα αντικείμενο, τότε θεωρείται ότι έχει υψηλά επίπεδα κατευθυντικότητας ως προς αυτό. Στις μελέτες αυτές, ωστόσο, ο χρήστης περιορίζεται σε στατικές πόζες, καθισμένος μπροστά στην οθόνη και τον πράκτορα. Στην περίπτωση της παρούσας έρευνας, η κατευθυντικότητα μετράται μονάχα από τον εντοπισμό του προσανατολισμού του κεφαλιού.

Ωστόσο, η έννοια της κατευθυντικότητας είναι στιγμιαία και, από μόνη της, είναι αναξιόπιστη πηγή πληροφόρησης σχετικά με τα επίπεδα προσήλωσης του χρήστη. Για παράδειγμα, εάν δειγματοληπτούνται δεδομένα τη στιγμή που ο χρήστης μεταφέρει το βλέμμα του από ένα αντικείμενο σε ένα άλλο, το αποτέλεσμα θα είναι παραπλανητικό. Τα επίπεδα προσήλωσης, λοιπόν, χρησιμοποιούνται για να αναφερθούν σε κατευθυντικότητα προς περιοχές, για χρονικά διαστήματα αρκετών δειγμάτων. Ένα σημαντικό θέμα που εγείρεται στο σημείο αυτό, είναι η οριοθέτηση των περιοχών, κάτι το οποίο επιτυγχάνεται με τη χρήση Αντικειμένων Εικονικής Προσήλωσης.

5.2.3 Αντικείμενα Εικονικής Προσήλωσης

Όσον αφορά σε συντεταγμένες οθόνες, ο χώρος εστίασης προσοχής του χρήστη ενδέχεται να φαίνεται δυναμικός λαμβάνοντας υπόψιν μόνο την κατευθυντικότητα του βλέμματος. Για το λόγο αυτό, απαιτείται ένα ανώτερο επίπεδο αναπαράστασης, καθώς ο χρήστης ενδέχεται να παρατηρεί κάποιο συγκεκριμένο αντικείμενο, ενώ το βλέμμα του να φαίνεται δυναμικό. Στο σημείο αυτό, και για να απλουστευθεί η ανάλυση σχετικά με το τί κοιτάζει ο χρήστης στην οθόνη, ορίζονται τα Αντικείμενα Εικονικής Προσήλωσης (Virtual Attention Objects - VAO) (σχήμα 5.1)

Για κάθε αντικείμενο της σκηνής ορίζεται και ένα VAO, για το οποίο θα υπολογιστούν τα επίπεδα προσήλωσης των χρηστών. Για παράδειγμα, στο σενάριο προσοχής από κοινού, αντιστοιχίζεται ένα VAO για τον εικονικό πράκτορα, ένα για κάθε ένα από τα αντικείμενα που απεικονίζονται, ένα για το φόντο της σκηνής και ένα για το χώρο εκτός της οθόνης. Με απλή διαδικασία μπορεί να γίνει αντιστοίχιση των τιμών

κατευθυντικότητας βλέμματος με τα VAO: Εάν οι συντεταγμένες οθόνης στις οποίες αντιστοιχεί το διάνυσμα βλέμματος βρίσκονται εντός κάποιου αντικειμένου, τότε το πεδίο προσοχής βρίσκεται μέσα σε αυτό, και τα αντίστοιχα επίπεδα προσοχής αυξάνονται. Συνεπώς, καθώς το διάνυσμα κατευθυντικότητας βλέμματος «μετακινείται» επάνω στην οθόνη, κάθε αντικείμενο (συμπεριλαμβανομένου του εικονικού πράκτορα) κρατά ένα ιστορικό της προσήλωσης από τον χρήστη.

5.2.4 Επίπεδα Ενδιαφέροντος

Σε μεγαλύτερα χρονικά πλαίσια υπολογίζεται ένας συνολικός βαθμός ενδιαφέροντος του χρήστη στο σενάριο. Σε αντίθεση με τα επίπεδα προσοχής στα VAO, η μετρική αυτή υπολογίζεται με βάση τα επίπεδα προσοχής που αντιστοιχούν μέχρι τη συγκεκριμένη χρονική στιγμή σε όλα τα αντικείμενα. Κάθε μέλος του συνόλου των αντικειμένων κατηγοριοποιείται ανάλογα με το ρόλο του στο σενάριο (αντικείμενο της σκηνής, πράκτορας, φόντο, περιοχή εκτός της οθόνης). Σε αυτό το σημείο λαμβάνεται υπόψη το περιεχόμενο του σεναρίου: Ορίζοντας δυναμικά το σύνολο εκείνων των αντικειμένων που έχουν κάποιο ρόλο στο σενάριο τη συγκεκριμένη χρονική στιγμή, εξετάζεται το διάνυσμα κατευθυντικότητας του χρήστη και, εάν συμπίπτει με τα αντικείμενα αυτά, τα επίπεδα ενδιαφέροντος (ενασχόλησης) στο σενάριο αυξάνονται. Η ενασχόληση έχει οριστεί σαν τη διαδικασία κατά την οποία δύο (ή περισσότεροι) συμμετέχοντες εγκαθιδρύουν, διατηρούν και ολοκληρώνουν μία αντιληπτή επικοινωνία, κατά τη διάρκεια δραστηριοτήτων που όλες οι πλευρές αναλαμβάνουν [151]. Ένας άλλος ορισμός που έχει δοθεί είναι ο εξής: Πρόκειται για το βαθμό στον οποίο ένας συμμετέχοντας σε μια δραστηριότητα αφιερώνεται στο να διατηρηθεί η διάδραση με τις άλλες πλευρές [29, 137]. Στο σενάριο που εξετάζεται εδώ, τα επίπεδα ενασχόλησης είναι άμεση συνάρτηση του κατά πόσο ο χρήστης κοιτάζει προς κατάλληλα αντικείμενα την κατάλληλη στιγμή. Τέτοια αντικείμενα είναι αυτά στα οποία έχει γίνει πρόσφατη αναφορά από τον πράκτορα, είτε λεκτικά, είτε δείχνοντάς τα. Τέτοιες μετρικές είναι εφικτές με τη δημιουργία συνόλων αντικειμένων στα οποία έχει γίνει πρόσφατη αναφορά. Όταν ο πράκτορας δεν αναφέρεται σε κανένα αντικείμενο, τότε αυτός είναι το μόνο μέλος του συνόλου και, όταν σταματήσει να μιλάει, το σύνολο αυτό είναι κενό.

Θα πρέπει να τονιστεί η διαφορά μεταξύ προσήλωσης και ενασχόλησης κατά τη διάρκεια της επικοινωνίας με τον εικονικό πράκτορα. Για το σκοπό αυτό, πέρα από τα επίπεδα ενασχόλησης, εισάγεται εδώ και η έννοια της ποιότητας ενασχόλησης, κάτι το οποίο προσφέρει μία πιο λεπτομερειακή αποτίμηση του τύπου της ενασχόλησης του χρήστη με τον πράκτορα. Για παράδειγμα, ένας χρήστης ο οποίος δεν ενδιαφέρεται ιδιαίτερα για το σενάριο, δεν κοιτάζει αναγκαστικά εκτός του πεδίου της οθόνης. Αντίθετα, ενδέχεται να την κοιτάζει με ένα πιο επιφανειακό τρόπο, πχ κοιτάζοντας αντικείμενα μη σχετικά με την διάδραση. Για το σκοπό αυτό, ορίζονται τρία βασικά επίπεδα ποιότητας: i) ενασχόληση με το σενάριο, ii) επιφανειακή ενασχόληση iii) καθόλου ενδιαφέρον. Με τον τρόπο αυτό, η συμπεριφορά του χρήστη δεν αντιμετωπίζεται ανεξάρτητα, αλλά πάντα σε σχέση με το περιεχόμενο της διάδρασης. Έτσι, για παράδειγμα, εάν ο πράκτορας περιγράφει κάτι σημαντικό και ο χρήστης δεν δείχνει κανένα ενδιαφέρον, τότε η μη ενασχόληση από πλευράς χρήστη είναι πιο σοβαρή από ότι εάν ο πράκτορας δεν έκανε τίποτα.

5.2.5 Πειραματικά αποτελέσματα

Στο πλαίσιο του σεναρίου επικοινωνίας ανθρώπου-ψηφιακού πράκτορα, έγιναν δύο πειράματα. Στο πρώτο έγινε προσομοίωση του περιβάλλοντος της οθόνης και, με χρήση του εντοπισμού περιστροφής κεφαλιού, ελέγχθηκε η καταλληλότητα του συστήματος για τη συγκεκριμένη εφαρμογή. Στο δεύτερο πείραμα θεωρήθηκε το σενάριο από κοινού προσοχής, με σκοπό να μελετηθεί η συμπεριφορά του χρήστη για τέσσερα διαφορετικά σενάρια αλληλεπίδρασης.

Λεπτομέρειες συστήματος: Το σύστημα που χρησιμοποιήθηκε στα πειράματα αποτελείται από δύο βασικά δομικά στοιχεία: το σύστημα εκτίμησης περιστροφής κεφαλιού από τη μία και το εικονικό περιβάλλον του σεναρίου από την άλλη. Τα δύο αυτά συστατικά επικοινωνούν με χρήση της πλατφόρμας Psyclone², λογισμικό που χρησιμοποιείται για τη δημιουργία μεγάλης κλίμακας, πολυτροπικών συστημάτων. Ο αλγόριθμος που χρησιμοποιήθηκε για την εκτίμηση της κατευθυντικότητας του κεφαλιού σαν ένδειξη του βλέμματος περιγράφεται στο κεφάλαιο 3, ενώ το εικονικό περιβάλλον αποτελείται από τον εικονικό πράκτορα Greta [15], και τη σκηνή των αντικειμένων προς πώληση.

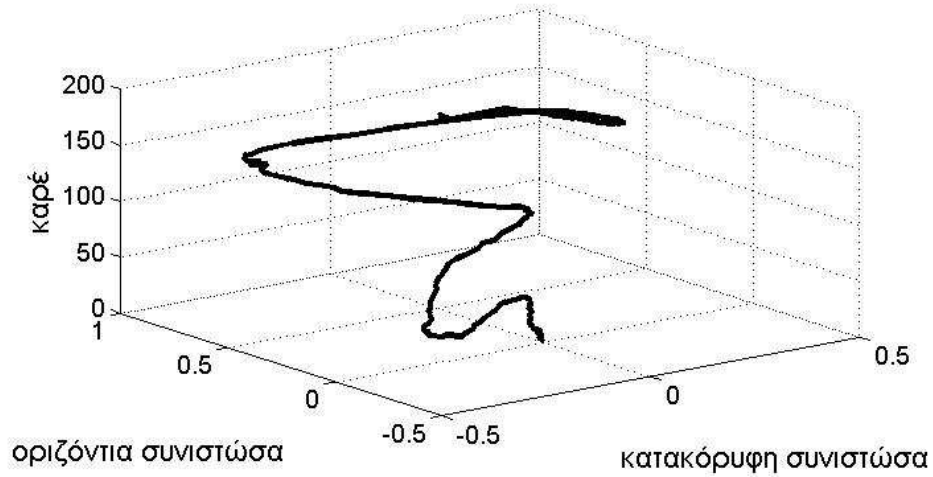
5.2.5.1 Πείραμα 1

Για να επαληθευθεί η καταλληλότητα του συστήματος εκτίμησης περιστροφής κεφαλιού σε σενάριο όπου ο χρήστης αλληλεπιδρά με εικονικό πράκτορα, έγινε το παρακάτω πείραμα: Ζητήθηκε από 7 άτομα να καθίσουν μπροστά στον υπολογιστή και, με μια δικτυακή κάμερα, παρακολούθηθηκε η κίνηση περιστροφής του κεφαλιού τους. Για κάθε έναν από τους συμμετέχοντες, καταγράφηκε η κίνηση του κεφαλιού του για ακολουθίες 200 καρτέ, δείχνοντάς του ένα αντικείμενο στην οθόνη και ζητώντας του να ακολουθήσει τη θέση του με την κίνηση του κεφαλιού του. Αν και η κίνηση του κεφαλιού δεν εμπίπτει στις αναμενόμενες κινήσεις που θα έκανε κάποιος κοιτάζοντας κάποιο πολυμεσικό υλικό στην οθόνη του υπολογιστή του, αποδείχτηκε ότι οι χρήστες εξοικειώθηκαν πολύ γρήγορα με την κίνηση του κεφαλιού ως ένδειξη κατευθυντικότητας βλέμματος. Πιο συγκεκριμένα, στο ξεκίνημα, εμφανιζόταν ένα εικονίδιο στο κέντρο της οθόνης, το οποίο, ακολουθώντας, κινούταν στην κάτω-αριστερή γωνία της οθόνης, και μετά πάλι πίσω στο κέντρο. Ακολουθώντας, το εικονίδιο εμφανιζόταν στην κάτω δεξιά γωνία της οθόνης και, τελικά, ζητούνταν από τον χρήστη να κοιτάξει ελεύθερα έξω από την οθόνη.

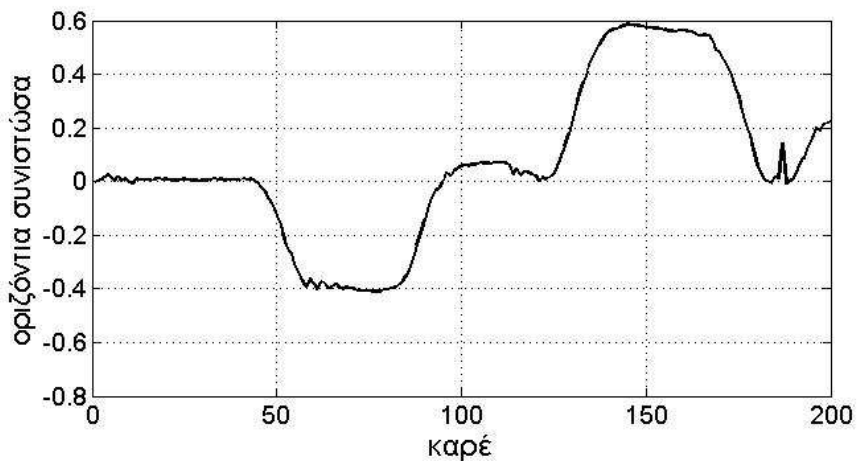
Με αυτόν τον τρόπο, μοντελοποιήθηκε ένα απλό σενάριο ενός υποθετικού πράκτορα (που μπορεί να βρίσκεται στο κέντρο της οθόνης), δύο εικονικών αντικειμένων (στα άκρα της οθόνης) και του περιβάλλοντα χώρου (περίπτωση που θα μπορούσε να αντιστοιχηθεί στην κατάσταση όπου ο χρήστης δεν προσέχει στην οθόνη). Οι συμμετέχοντες κάθισαν μπροστά από την οθόνη του υπολογιστή, ενώ η δικτυακή κάμερα τοποθετήθηκε απέναντί τους (σε απόσταση περίπου 60-70cm από το επίπεδο των ματιών τους). Όσον αφορά στις συνθήκες φωτισμού, ήταν αντίστοιχες με τις αναμενόμενες ενός εργασιακού χώρου. Να σημειωθεί ότι τοποθέτηση των συμμετεχόντων σε διαφορετικές θέσεις θα απαιτούσε, όπως θα φανεί παρακάτω, διαφορετικά κατώφλια για την ταξινόμηση των περιοχών εστίασης.

Συμβολίζοντας το διάνυσμα περιστροφής κεφαλιού με $\mathbf{p} = [p_x \ p_y]$, με p_x και p_y οι οριζόντια και η κάθετη συνιστώσα, αντίστοιχα, οι μέσες τιμές τους ανά συμμετέχοντα, φαίνονται στο σχήμα 5.2.

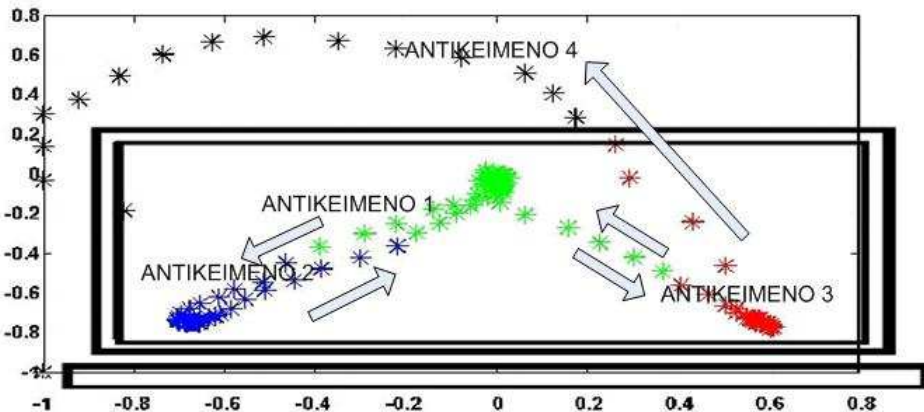
²<http://www.cmlabs.com/psyclone/>



Σχήμα 5.2: 3Δ καμπύλη που απεικονίζει τη μέση κίνηση ανά χρήση



Σχήμα 5.3: Μέση τιμή οριζόντιας συνιστώσας διανύσματος περιστροφής κεφαλιού



Σχήμα 5.4: Αντιστοίχιση διανύσματος περιστροφής κεφαλιού σε συντεταγμένες οθόνης

Η μέση τιμή της οριζόντιας συνιστώσας, υπολογισμένη από όλους τους χρήστες, φαίνεται στο σχήμα 5.3. Φαίνεται ότι, παρόλο που το εικονίδιο άλλαζε θέση κάθε 40 καρέ, υπήρχε μια μικρή καθυστέρηση στην απόκριση του βλέμματος του κόσμου (≈ 10 καρέ), κάτι το οποίο ελήφθη υπόψη στην εξαγωγή των αποτελεσμάτων. Για την αποτίμηση της μεθόδου, θεωρήθηκαν οι παρακάτω κανόνες: ANTIKEIMENO1=κέντρο οθόνης, ANTIKEIMENO2=κάτω αριστερή γωνία οθόνης, ANTIKEIMENO3=κάτω δεξιά γωνία οθόνης, ANTIKEIMENO4=αλλού. Τυπικό παράδειγμα διανύσματος περιστροφής κεφαλιού συμμετέχοντα, ο οποίος ακολούθησε την πορεία ANTIKEIMENO1 \rightarrow ANTIKEIMENO2 \rightarrow ANTIKEIMENO1 \rightarrow ANTIKEIMENO3 \rightarrow ANTIKEIMENO4 φαίνεται στο σχήμα 5.4.

Από καμπύλες σαν αυτή του σχήματος 5.4, εξήχθησαν οι ακόλουθοι κανόνες, με σκοπό να αντιστοιχηθούν οι τιμές του διανύσματος περιστροφής κεφαλιού με τα εικονικά αντικείμενα.

- If $-thr_1 < p_y < thr_1$ AND $-thr_3 < p_x < thr_3$ ANTIKEIMENO1
- If $p_y < -thr_1$ AND $p_y > -thr_2$ AND $p_x < -thr_3$ AND $p_x > -thr_4$ ANTIKEIMENO2
- If $p_y > thr_1$ AND $p_y < thr_2$ AND $p_x < -thr_3$ AND $p_x > -thr_4$ ANTIKEIMENO3
- ELSE ANTIKEIMENO4

Στα πειράματα χρησιμοποιήθηκαν κατώφλια: $thr_1=0.2$, $thr_2=0.7$, $thr_3=0.2$ και $thr_4=0.7$. Όπως αναφέρθηκε και προηγουμένως, ενώ το εικονίδιο, από τη μια θέση στην άλλη μετέβαινε ακαριαία, δε συνέβαινε το ίδιο και με το βλέμμα του εκάστοτε χρήστη, κάτι το οποίο ποίκιλλε, μάλιστα, ανάλογα με το αντικείμενο. Για το λόγο αυτό, στην αποτίμηση των αποτελεσμάτων, δεν ελήφθη υπόψη το διάνυσμα περιστροφής κεφαλιού στα πρώτα 6 με 14 καρέ εμφάνισης κάθε αντικειμένου (έτσι, για κάθε χρήστη, τα δεδομένα που χρησιμοποιήθηκαν, αντιστοιχούσαν σε 156 καρέ). Ο πίνακας 5.1 είναι ο πίνακας σύγκρισης της ταξινόμησης κάθε διανύσματος περιστροφής κεφαλιού στα διάφορα αντικείμενα, με χρήση των κατωφλίων thr_1 , thr_2 , thr_3 και thr_4 .

Οι τελευταίες δύο στήλες του πίνακα 5.1 αντιστοιχούν στην Ακρίβεια Παραγωγού (ΑΠ) και στην Ακρίβεια Χρήστη (ΑΧ). Το πρώτο κριτήριο αντιστοιχεί στην ικανότητα

Πίνακας 5.1: Πίνακας σύγχυσης σφαλμάτων για όλους τους χρήστες

	A1	A2	A3	A4	ΑΠ(%)	ΑΧ (%)
A1	493	0	0	13	95.17	97.43
A2	0	143	0	0	70.44	100
A3	12	54	173	1	85.22	72.08
A4	13	6	30	154	91.67	75.86
Σύνολο	518	203	203	168		
Συνολική ακρίβεια: 88.2% Khat=0.83,						
A==ANTIKEIMENO						

ενός ταξινομητή να ταξινομήσει ένα πρότυπο σωστά, σε σχέση με τις πραγματικές εμφανίσεις του, και το δεύτερο είναι το ποσοστό των σωστών ταξινομήσεων σε μια κλάση, σε σχέση με το σύνολο των προτύπων που βρέθηκε να σχετίζονται με αυτήν. Όπως φαίνεται, η συνολική ακρίβεια του συστήματος είναι 88.2%. Ωστόσο, εδώ υπολογίστηκε και ο δείκτης Khat (εξίσωση 5.2), ως μέτρο της διαφοροποίησης του ταξινομητή από την τυχαία ταξινόμηση. Στο πείραμα αυτό, επιτεύχθηκε τιμή Khat μεγαλύτερη από 0,8, κάτι το οποίο δείχνει ότι υπάρχει ισχυρή συμφωνία [28] μεταξύ των δεδομένων αναφοράς (ANTIKEIMENO) και της ταξινόμησης, όπως προέκυψε από τον πίνακα σύγχυσης. Για κάθε χρήστη ξεχωριστά, οι παραπάνω μετρήσεις, φαίνονται στον πίνακα 5.2.

$$Khat = \frac{N \sum_{i=1}^r x_{ii} - \sum_{i=1}^r (x_{i+} x_{+i})}{N^2 - \sum_{i=1}^r (x_{i+} x_{+i})} \quad (5.2)$$

όπου r είναι ο αριθμός των κλάσεων, x_{ii} είναι ο αριθμός των στοιχείων στην γραμμή i και στήλη j , x_{i+} και x_{+i} είναι το άθροισμα των τιμών της γραμμής i και της στήλης i αντιστοίχως, και N είναι ο συνολικός αριθμός των στοιχείων του συνόλου που εξετάζεται με τον πίνακα λαθών. Οι τιμές του Khat βρίσκονται στο πεδίο $[0,1]$ όπου με 1 χαρακτηρίζεται η βέλτιστη ταξινόμηση με απόλυτη συμφωνία μεταξύ της ταξινόμησης και των στοιχείων αναφοράς. Συνήθως, τιμές μεγαλύτερες από 0.8 αντιστοιχούν σε ισχυρή συμφωνία, τιμές στο $[0.4,0.8]$ αντιστοιχούν σε μεσαία συμφωνία, ενώ τιμές κάτω από 0.4 αντιστοιχούν σε μικρή συμφωνία.

Από τον πίνακα 5.2 προκύπτει ότι το κριτήριο Khat είναι πολύ υψηλό για τους χρήστες 1,2,4 και 7, ενώ για τους χρήστες 3,5 και 6 δείχνουν μέτρια συμφωνία μεταξύ δεδομένων αναφοράς και μετρήσεων [28]. Για τους χρήστες 3 και 6, αυτό οφείλεται κυρίως στο ότι το ANTIKEIMENO2 ταξινομήθηκε λανθασμένα ως ANTIKEIMENO4, ενώ, για τον χρήστη 5, το ANTIKEIMENO3 ταξινομήθηκε λανθασμένα ως ANTIKEIMENO4. Γενικά, η συνολική ακρίβεια του συστήματος ήταν 88.2%, τιμή αρκετά ενθαρρυντική για σύστημα που στηρίζεται σε απλό και μη παρεμβατικό εξοπλισμό, χωρίς πρότερη γνώση χρηστών και εσωτερικών παραμέτρων της κάμερας. Τα κατώφλια που χρησιμοποιήθηκαν είναι μετρίως αυστηρά, με την έννοια ότι, για το πρόβλημα της ταξινόμησης των τριών αντικειμένων στην οθόνη, θα μπορούσαν να θεωρηθούν μεγαλύτεροι χώροι για κάθε αντικείμενο (σχήμα 5.4). Έτσι, θέτοντας για τους κανόνες

Πίνακας 5.2: Πίνακας Σύγκρισης για κάθε χρήστη ξεχωριστά

	Κριτήριο Khat	Συνολική ακρίβεια (%)
Χρήστης 1	0.96	97.44
Χρήστης 2	0.96	97.44
Χρήστης 3	0.63	73.72
Χρήστης 4	1.00	100.0
Χρήστης 5	0.62	74.36
Χρήστης 6	0.65	75.00
Χρήστης 7	0.99	99.36
Συνολική Ακρίβεια: 88.2% Khat=0.83		

των αντικειμένων 2 και 3, $thr_1=0$, τότε αυτά καταλαμβάνουν περισσότερο χώρο και, άρα, η ταξινόμηση στις αντίστοιχες κλάσεις θα είναι συχνότερη, χωρίς να επηρεάζονται οι υπόλοιπες. Σε αυτήν την περίπτωση, ο αλγόριθμος θα επιτύγχανε, συνολικά, $khat=0.88$ και συνολική ακρίβεια 92.2%. Αυτά τα ποσοστά είναι συγκρίσιμα με τα κατώφλια που εξετάστηκαν προηγουμένως, δείχνοντας, με αυτόν τον τρόπο, ότι η ταξινόμηση βλέμματος με βάση την κίνηση του κεφαλιού μπορεί να πετύχει αρκετά ικανοποιητικά αποτελέσματα, για σχετικά λεπτομερή καταμερισμό της περιοχής της οθόνης.

5.2.5.2 Πείραμα 2

Στο δεύτερο πείραμα έγινε πιο εκτενής μελέτη των συνθηκών του σεναρίου από κοινού προσοχής. Σκοπός ήταν να μελετηθεί η συμπεριφορά του χρήστη για μικρές διαφορές στον τρόπο που εκφράζεται ο εικονικός πράκτορας, η Greta. Το πείραμα έγινε σε περιβάλλον εργαστηρίου, και συμμετείχαν επτά άτομα (4 άντρες, 3 γυναίκες), από τα οποία ζητήθηκε να καθίσουν μπροστά σε μία οθόνη στο πάνω μέρος της οποίας είχε εφαρμοστεί δικτυακή κάμερα, ενώ η οθόνη του υπολογιστή είχε ανάλυση 1280×1024 .

Σενάριο: Κατά τη διάρκεια του σεναρίου, ο εικονικός πράκτορας παίζει το ρόλο πωλητή σε κατάστημα ειδών υπολογιστή, και παρουσιάζει ένα σύνολο διαφορετικών προϊόντων προς πώληση (βλ. σχήμα 5.5)

Κάθε ένα από τα προϊόντα απεικονίζεται σαν γραφικό στοιχείο στην οθόνη, δίπλα στον εικονικό πράκτορα, ενώ αλγοριθμικά αναπαρίσταται σαν VAO. Ο πράκτορας ξεκινά την περιγραφή ενός τυχαίου αντικειμένου, ενώ ο χρήστης αναμένεται να ακούει την περιγραφή, με το διάνυσμα περιστροφής κεφαλιού να δίνει πληροφορία σχετικά με το ενδιαφέρον του στην αλληλεπίδραση. Κάθε χρήστης επανέλαβε την αλληλεπίδραση σε τέσσερις διαφορετικές εκδοχές, από τρία περίπου λεπτά η κάθε μία (συνολικά 80 περίπου λεπτά δεδομένων). Κάθε εκδοχή διέφερε από τις υπόλοιπες σε σχέση με τον τρόπο που ο εικονικός πράκτορας κοίταζε προς τα αντικείμενα για τα οποία μιλούσε ανά πάσα στιγμή. Έτσι, οι συνθήκες είναι οι εξής:

1. *Συνθήκη 1:* Μετωπικό βλέμμα. Ο εικονικός πράκτορας κοιτάζει μετωπικά προς τον χρήστη, καθόλη τη διάρκεια της αλληλεπίδρασης.



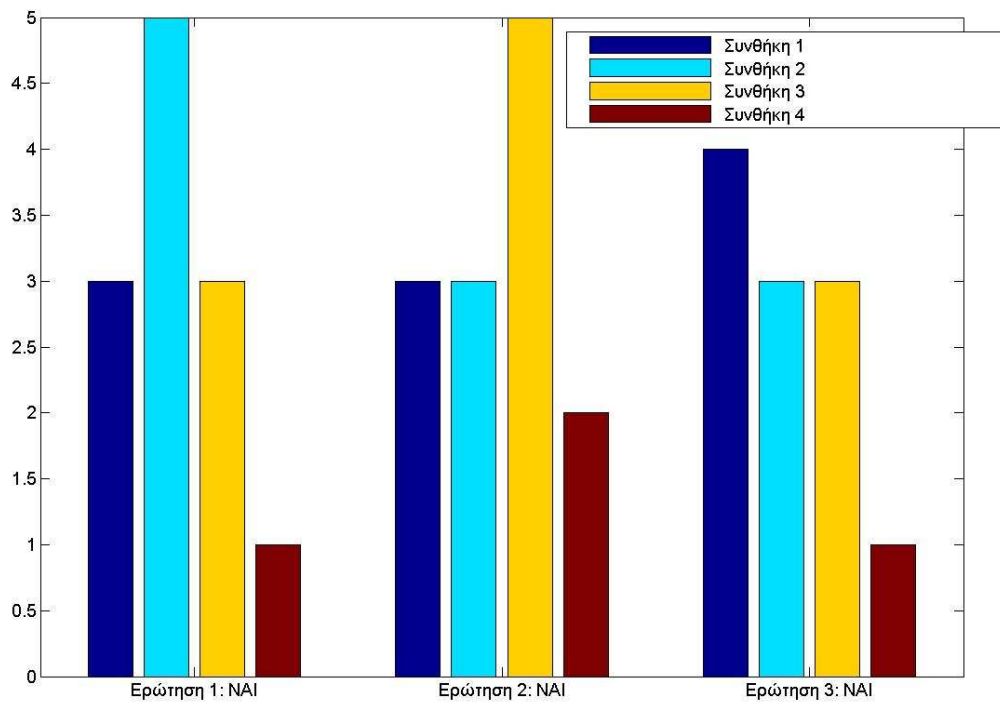
Σχήμα 5.5: Απεικόνιση του σεναρίου, όπου ο πράκτορας ακολουθάει διαφορετικά πρότυπα συμπεριφοράς όσον αφορά το βλέμμα του σε σχέση με τα αντικείμενα για τα οποία μιλάει. Στο κέντρο της οθόνης, ο χρήστης έχει τη δυνατότητα να έχει οπτική επαφή με την εικόνα του, καθώς το αντικείμενο προς το οποίο κοιτάζει είναι μαρκαρισμένο με ένα κόκκινο κύκλο. Δεξιά φαίνεται το ιστορικό της κατευθυντικότητας του βλέμματος ενός χρήστη κατά τη διάρκεια ενός σεναρίου.

2. *Συνθήκη 2:* Συνεπής κατευθυντικότητα βλέμματος. Σε αυτή την εκδοχή, ο εικονικός πράκτορας κοιτάζει μετωπικά όσο δε μιλάει για κάποιο αντικείμενο, ενώ κοιτάζει πάντα το αντικείμενο για το οποίο μιλάει.
3. *Συνθήκη 3:* Παροδικά συνεπής κατευθυντικότητα βλέμματος. Η εκδοχή αυτή μοιάζει με την 2, με τη διαφορά ότι ο πράκτορας επιστρέφει το βλέμμα του στον χρήστη, 6 δευτερόλεπτα αφού ξεκινήσει να περιγράφει ένα αντικείμενο.
4. *Συνθήκη 4:* Ασυνεπής κατευθυντικότητα βλέμματος. Σε αυτήν την περίπτωση, ο πράκτορας κοιτάζει τυχαία προς αντικείμενα διαφορετικά από αυτά που περιγράφει.

Οι διαφορές ανάμεσα στις τρεις πρώτες εκδοχές ενδεχομένως να μην είναι τόσο μεγάλες, όμως, κύριος στόχος του πειράματος ήταν να μελετηθεί η συνθήκη 4. Έτσι, σκοπός ήταν να βρεθεί εάν η κατευθυντικότητα του βλέμματος των χρηστών επηρεάζονταν από την ακουστική ή την οπτική ανάδραση που είχαν από τον πράκτορα.

Καθώς ο σκοπός της μελέτης ήταν να θεωρηθούν, όχι αναγκαστικά συνθήκες έντονου ενδιαφέροντος από πλευράς χρήστη, αλλά να ερευνηθεί πληθώρα συνθηκών (οι χρήστες θα έχουν διαφορετικά επίπεδα ενδιαφέροντος, ανάλογα και με τη συμπεριφορά του πράκτορα), το σενάριο δεν επιδίωκε να κρατήσει το ενδιαφέρον του χρήστη: αν και τα αντικείμενα επιλέγονταν τυχαία, η θέση των αντικειμένων παρέμενε η ίδια για όλες τις εκδοχές, ενώ δεν υπήρχε καμία αλλαγή στο περιβάλλον του πράκτορα. Επίσης, σκοπίμως επιλέχθηκε ένα περιβάλλον καταστήματος προϊόντων γνώριμων (και άρα χωρίς ενδιαφέρον) στους χρήστες, ώστε να μην υπάρξει επιτηδευμένο ερέθισμα να τραβηχθεί η προσοχή των συμμετεχόντων. Ακόμη, καθώς το πείραμα έλαβε χώρα σε ένα μη ελεγχόμενο περιβάλλον εργαστηρίου, με έντονη δραστηριότητα, αναμενόταν (και στην πραγματικότητα ήταν επιθυμητό) οι συμμετέχοντες να αφαιρούνται κάποιες στιγμές από το πείραμα, ή να μην έχουν μεγάλο βαθμό ενασχόλησης με το σενάριο.

Σε κάθε εκδοχή, ο πράκτορας επέλεγε πέντε αντικείμενα τυχαία, και τα περιέγραφε, σύμφωνα με την εκάστοτε συνθήκη (1 έως 4). Επιπλέον, οι εκδοχές του σεναρίου παρουσιάζονταν στο χρήστη με τυχαία σειρά. Στο τέλος κάθε εκδοχής του σεναρίου, ζητούνταν από κάθε συμμετέχοντα να συμπληρώσει ένα ερωτηματολόγιο με τρεις ερωτήσεις τύπου ΝΑΙ/ΟΧΙ:



Σχήμα 5.6: Θετικές απαντήσεις στα ερωτηματολόγια στην ερώτηση (αριστερά) αν ενδιαφέρεται ο πράκτορας για τον πράκτορα, αν (κέντρο) ο πράκτορας ενδιαφέρεται για τον χρήστη και (δεξιά) αν ο χρήστης ενδιαφέρεται για τον πράκτορα. Οι απαντήσεις αφορούν και στις τέσσερις συνθήκες του πειράματος

1. Ο πράκτορας ενδιαφέρεται για τον εαυτό του; Στην ερώτηση αυτή ο συμμετέχων καλούνταν να απαντήσει στο εάν πίστευε ότι ο πράκτορας έδειχνε ο ίδιος ενδιαφέρον στα λεγόμενά του
2. Ο πράκτορας ενδιαφέρεται για τον χρήστη; Στην ερώτηση αυτή ο συμμετέχων καλούνταν να απαντήσει στο εάν πίστευε ότι ο πράκτορας έδειχνε ενδιαφέρον στον χρήστη κατά τη διάρκεια της περιγραφής
3. Ο χρήστης ενδιαφέρεται για τον πράκτορα; Η ερώτηση αυτή είχε να κάνει με το κατά πόσο ο κάθε χρήστης, ανά εκδοχή του σεναρίου, έδειχνε ενδιαφέρον σε αυτά που περιέγραφε ο πράκτορας

Τα αποτελέσματα των ερωτηματολογίων φαίνονται στο σχήμα 5.6 και συζητιούνται παρακάτω.

5.2.6 Πορίσματα

Η ανάλυση των αποτελεσμάτων όσον αφορά στην περιστροφή του κεφαλιού έδειξε ότι οι συμμετέχοντες είχαν αρκετά ανεπτυγμένη την ευχέρεια να κινήσουν το κεφάλι τους σαν ένδειξη του ενδιαφέροντός τους. Πιο συγκεκριμένα, αν και απλή περιστροφή του βλέμματος των ματιών είναι αρκετή για την «εξερεύνηση» του τί βρίσκεται στην οθόνη του υπολογιστή, εν τούτοις, οι χρήστες μπόρεσαν με αρκετά μεγάλη φυσικότητα να χρησιμοποιήσουν την κίνηση του κεφαλιού για να κατευθύνουν το βλέμμα τους. Από

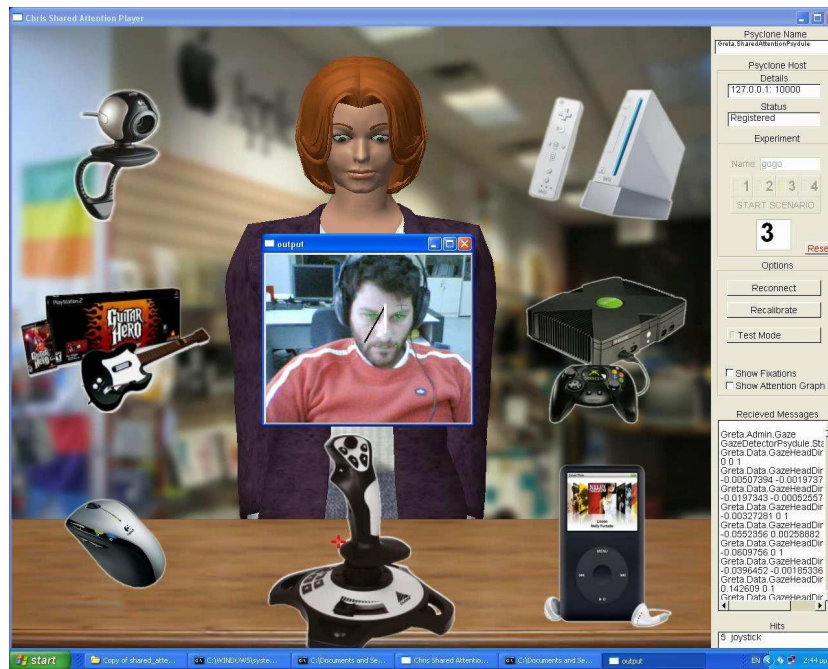
τα αποτελέσματα, επίσης, βγήκε το συμπέρασμα ότι οι χρήστες κοίταζαν τον πράκτορα όσο δεν αναφερόταν σε κάποιο αντικείμενο, ενώ είχαν το βλέμμα τους στραμμένο προς το προϊόν, όσο γινόταν λόγος γι αυτό.

Μέσα στα μειονεκτήματα των συνθηκών του πειράματος, και μετά από παρατήρηση των αποτελεσμάτων, συγκαταλέγεται το γεγονός ότι, καθώς υπήρχε οπτική ανάδραση του VAO στο οποίο είναι επικεντρωμένος ο χρήστης (με χρήση ενός κόκκινου σταυρού), φάνηκε πως υπήρχε μεγάλη προσπάθεια από μέρους του να επικεντρωθεί εκεί, κάτι το οποίο θα μπορούσε να ερμηνευθεί ως ανασταλτικός παράγοντας του αυθορμητισμού. Επιπλέον, γνωρίζοντας οι χρήστες πως το βλέμμα τους παρακολουθείται, φάνηκε να καταβάλουν αρκετά μεγάλη προσπάθεια ούτως ώστε να μην αφαιρεθούν από το σενάριο. Κάτι τέτοιο ήρθε σε αντίθεση με το γεγονός ότι οι περισσότεροι χρήστες δήλωσαν, μετά την εκτέλεση του πειράματος, ότι η διαδικασία τους φάνηκε κουραστική.

Σε ό,τι αφορά στις απαντήσεις των χρηστών στα ερωτήματα, δεν διαφάνηκαν σημαντικές διαφορές ανάμεσα στις τέσσερις συνθήκες. Κάτι τέτοιο, σε γενικές γραμμές, δείχνει ότι οι χρήστες δεν παρατήρησαν έντονες διαφορές ανάμεσα σε περιπτώσεις όπου ο πράκτορας κοίταζε ή όχι το αντικείμενο στο οποίο αναφερόταν. Ωστόσο, ανάμεσα στις συνθήκες 2 και 4, υπήρχε αρκετά μεγάλη στατιστική διαφορά ($p < 0.05$) στην ερώτηση εάν ο χρήστης πίστευε ότι ο πράκτορας ενδιαφερόταν για αυτό που περιέγραφε. Οι συνθήκες αυτές, έτσι κι αλλιώς, έρχονται σε μεγάλη αντίθεση, αφού στην 2, ο πράκτορας κοιτάζει το προϊόν για το οποίο μιλάει, καθ' όλη τη διάρκεια που αναφέρεται σε αυτό, ενώ στην 4 κοιτάζει τυχαία προς άλλα προϊόντα. Το συμπέρασμα αυτό είναι σημαντικό, καθώς σηματοδοτεί πως οι χρήστες όντως παρατήρησαν διαφορές στη συμπεριφορά του πράκτορα για αυτές τις δύο ακραίες περιπτώσεις. Πιο συγκεκριμένα, ανάλυση του συστήματος εκτίμησης περιστροφής κεφαλιού έδειξε ότι οι χρήστες, στο σενάριο της τέταρτης συνθήκης, είχαν την τάση να ακολουθούν το βλέμμα του πράκτορα πριν συνειδητοποιήσουν ότι αναφέρεται σε προϊόν που δεν κοιτάζει, μέχρι να στρέψουν το βλέμμα τους στο προϊόν που περιγράφει ο πράκτορας (βλ. σχήμα 5.7).

5.3 Από το βλέμμα στην αναγνώριση κατάστασης χρήστη

Η κατανόηση των προθέσεων των ανθρώπων αποτελεί δομική ικανότητα και συνεπάγεται την αναπαράσταση συναισθηματικών καταστάσεων τρίτων, στην περίπτωση μας, από ένα σύστημα μηχανικής εκπαίδευσης. Οι αρχές και οι τεχνικές που εκμεταλλεύονται οι άνθρωποι προκειμένου να κατανοήσουν, να προβλέψουν και να χειριστούν τις συμπεριφορές άλλων ανθρώπων, αναφέρονται στη βιβλιογραφία σαν «Θεωρία του Νου» (Theory of Mind - ToM) [14]. Σε σενάρια επικοινωνίας ανθρώπου-μηχανής, κάτι τέτοιο μεταφράζεται σε παρακολούθηση της συμπεριφοράς χρηστών όσον αφορά το υλικό που μελετούν, ανιχνεύοντας συγκεκριμένα χαρακτηριστικά (πχ. προσοχή σε συγκεκριμένα σημεία, ή αποστροφή του βλέμματος από την οθόνη) και προσαρμογή του συστήματος ώστε να ανταποκρίνεται στις απαιτήσεις της εκάστοτε κατάστασης του χρήστη. Στα [18] και [100] αναλύονται συγκεκριμένα μοντέλα ToM. Συνήθως, τέτοια μοντέλα λαμβάνουν υπόψη επιστημικές (πχ απόψεις) και/ή καταστάσεις που σχετίζονται με τα κίνητρα (πχ επιθυμίες, προθέσεις), ενώ επιχειρούν, επίσης, ενσωμάτωση του συναισθήματος [48]. Στη συγκεκριμένη διατριβή δεν γίνεται προσπάθεια



Σχήμα 5.7: Παράδειγμα περίπτωσης συνθήκης 4, στο οποίο φαίνεται ότι ο χρήστης ακολούθησε τη μη-προφορική πληροφορία του πράκτορα για να στρέψει το βλέμμα του. Στη συγκεκριμένη περίπτωση, ο χρήστης στρέφει το βλέμμα του στο παιχνιδιόμχλιο (κάτω-κέντρο) ενώ ο πράκτορας αναφέρεται στο ποντίκι (κάτω-αριστερά). Αργότερα, ο χρήστης έστρεψε το βλέμμα του στο σωστό αντικείμενο.

εντοπισμού τέτοιων παραγόντων, δηλαδή να ανιχνευθεί τί κρύβεται πίσω από συγκεκριμένες καταστάσεις (γιατί ο χρήστης, για παράδειγμα, δεν είναι προσεχτικός), το οποίο θα απαιτούσε μοντελοποίηση των απόψεων, επιθυμιών και προθέσεων του, αλλά να χρησιμοποιηθούν οι καταστάσεις με σκοπό την προσαρμογή της μαθησιακής διαδικασίας. Πιο συγκεκριμένα, χρησιμοποιούνται τα συμπεράσματα από τη Θεωρία του Νου για να συσχετισθούν τα διάφορα χαρακτηριστικά που εντοπίζει το σύστημα σχετικά με την τρέχουσα κατάσταση του χρήστη. Για παράδειγμα, η κατεύθυνση του βλέμματος του χρήστη προς την οθόνη, καθώς αυτός διαβάζει κάποιο ηλεκτρονικό κείμενο που προβάλλεται σε αυτήν, μπορεί να συνδεθεί με το επίπεδο του ενδιαφέροντος του χρήστη για αυτό που διαβάζει εκείνη τη στιγμή. Επιπλέον, ξαφνικές και απότομες κινήσεις της κεφαλής, σύμφωνα με τη Θεωρία του Νου, μπορούν να συνδεθούν με τη νευρικήτητα και την απογοήτευση του χρήστη.

Στα πλαίσια της παρούσας διατριβής, και της ανάγκης να διαπιστωθεί ότι το σύστημα που περιγράφεται στο κεφάλαιο 3 μπορεί να χρησιμοποιηθεί σε πραγματικές εφαρμογές, προτείνεται σύστημα αναγνώρισης κατάστασης μαθητών σε συνθήκες μάθησης. Η πληροφορία που προέρχεται από την ανάλυση του προσώπου μπορεί να αναχθεί σε χαρακτηριστικά σχετιζόμενα με συμπεριφορά (προσοχή, ενδιαφέρον, νευρικήτητα) χρηστών οι οποίοι διαβάζουν κάποιο πολυμεσικό υλικό σε οθόνη υπολογιστή, και των οποίων οι κινήσεις καταγράφονται από δικτυακή κάμερα. Η ανάλυση αυτών των δεδομένων γίνεται με χρήση των Θεωριών του Νου και την αντίστοιχη επισήμειωση πληροφορίας κατάστασης χρήστη σε μία βάση από ανήλικα παιδιά, ενώ χρησιμοποιήθηκε νευροασαφές σύστημα για την αναγωγή των χαρακτηριστικών (περιστροφή κεφαλιού, κατευθυντικότητα βλέμματος κλπ) στις καταστάσεις συμπεριφοράς. Αποδείχθηκε από το σύστημα και την αξιολόγησή του (τόσο ποσοτική, όσο και ποιο-

τική) ότι το μη-παρεμβατικό σύστημα που περιγράφεται μπορεί να λειτουργήσει σα μέσο καταγραφής της κατάστασης του παιδιού απέναντι στο κείμενο που διαβάσει και, μάλιστα, χωρίς να αντιλαμβάνεται ότι οι κινήσεις του καταγράφονται, ώστε να μειωθεί, ενδεχομένως, ο αυθορμητισμός του.

Η πλατφόρμα δοκιμών [165] ανάγνωσης δείχνει στον χρήστη ένα έγγραφο τύπου pdf (Portable Document Format) και ο στόχος είναι να εντοπιστούν προβλήματα ανάγνωσης σχετικά με δυσλεξία. Τέτοια προβλήματα [163], στην πλειοψηφία τους αφορούν στην ανάγνωση λέξεων με λανθασμένη σειρά, παράλειψη λέξεων ή δυσκολίες στην ανάλυση των λέξεων σε ακουστικές μονάδες συλλαβικής βάσης και στη σύνθεση συλλαβικών ακουστικών μονάδων σε λεξικά σύνολα με εννοιολογικό περιεχόμενο. Για να ελαχιστοποιηθούν τέτοια φαινόμενα, το πολυμεσικό υλικό επι φωτίζει τις λέξεις μέσα στο κείμενο με τη σωστή τους σειρά, ώστε το παιδί να μη χάσει τη σωστή ακολουθία. Σε περίπτωση που εντοπιστεί αλλαγή στην κατάσταση του χρήστη, το σύστημα προσαρμόζει τη λειτουργία αυτή αναλόγως, ακόμα και παύοντάς την εάν εντοπιστεί ότι το παιδί δεν έχει στραμμένη την προσοχή του προς την οθόνη. Το σύστημα, επιπλέον, λαμβάνει μέριμνα ούτως ώστε, για περιπτώσεις επαναλαμβανόμενης απουσίας προσοχής, να αλλάζει ακόμα και το στυλ γραμματοσειράς για να ενθαρρύνει την ανάγνωση. Η δουλειά στο [165] παρουσιάζει μια πιο πλήρη περιγραφή της πλατφόρμας δοκιμών. Λειτουργικές λεπτομέρειες ξεφεύγουν από το σκοπό της παρούσας δουλειάς, γι' αυτό και στην παρούσα διατριβή έχουν σκοπίμως παραλειφθεί.

Η εκπαίδευση του συστήματος αναγνώρισης καταστάσεων έγινε με χρήση δεδομένων που προήλθαν κατά τη χρησιμοποίηση της παραπάνω διεπαφής, και τα οποία επισημειώθηκαν από ειδικούς στην ηλεκτρονική εκπαίδευση, στο πλαίσιο του έργου Agent-Dysl FP6 Project [2], κάνοντας χρήση του λογισμικού ANVIL [81]. Καταγράφηκαν ακολουθίες με συγκεκριμένα χαρακτηριστικά καταστάσεων (προσεκτικός/μη προσεκτικός νευρικός/ήρεμος) τα οποία συνδέθηκαν με ανιχνεύσιμα και μετρήσιμα χαρακτηριστικά. Για παράδειγμα, στην επισημείωση ότι κάποιο παιδί είναι αφηρημένο, στα δεδομένα αναφοράς δίνεται η εξήγηση του λόγου, ο οποίος, για παράδειγμα, μπορεί να είναι ότι το κεφάλι είναι περιστραμμένο σε σχέση με την οθόνη, συνδέοντας, με αυτόν τον τρόπο, μία έννοια που περιγράφεται στη Θεωρία του Νου, με ένα απόλυτα ανιχνεύσιμο και μετρήσιμο μέγεθος της επιστήμης όρασης υπολογιστών. Ως εκ τούτου, χρησιμοποιήθηκε η επισημείωση μιας βάσης ακολουθιών από παιδιά που διαβάζουν σχολικά κείμενα για να εκπαιδευτεί ένα νευροασαφές δίκτυο, το οποίο εν συνεχεία χρησιμοποιείται για να εντοπίσει συμπεριφορικές καταστάσεις με βάση μετρήσιμα χαρακτηριστικά.

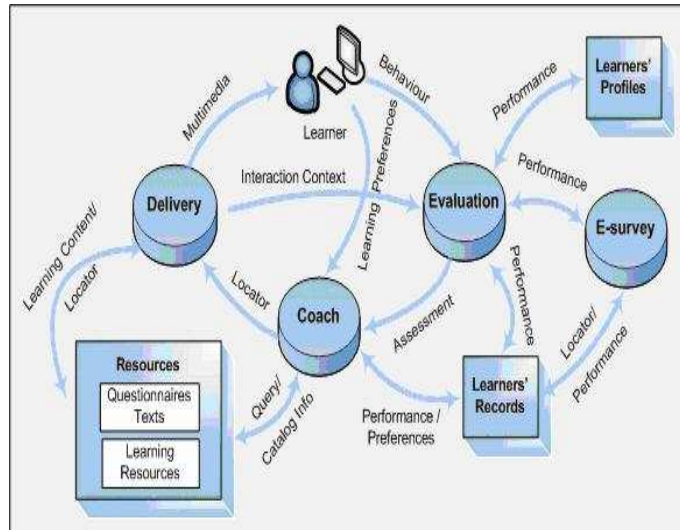
Οι καταστάσεις που ανιχνεύονται χρησιμοποιούνται για να ενημερώσουν το προφίλ του κάθε χρήστη, με χρήση μιας επέκτασης [164] του μοντέλου αναφοράς της IEEE (WG) της Επιτροπής Προτύπων Τεχνολογίας Εκπαίδευσης (Learning Technology Standards Committee), στην οποία είναι ενσωματωμένη μία διαδικασία προφίλ χρηστών. Έτσι, οι συμπεριφορικές καταστάσεις του χρήστη χρησιμοποιούνται για να χτιστεί το προφίλ του και στο να προσαρμοστεί η παρουσίαση των πολυμεσικών δεδομένων.

5.3.1 Περιγραφή της εφαρμογής

Μία από τις τεχνικές πρωτοτυπίες του συγκεκριμένου συστήματος ηλεκτρονικής μάθησης είναι ο χειρισμός των εκπαιδευομένων με προσωπικό τρόπο, χτίζοντας ξεχωριστά για τον καθένα προφίλ, σύμφωνα με τη συμπεριφορά του. Στο σχήμα 5.8 φαίνεται η

Πίνακας 5.3: Κατηγοριοποίηση Πιθανών Καταστάσεων Χρήστη

Ενδείξεις	Κατάσταση
- Συνοφρύωση	Απογοητευμένος / Δυσκολία στην ανάγνωση
- Κάποιος μιλάει στο χρήστη - Ακούγεται κάποιος θόρυβος - Κάνει σπασμωδικές κινήσεις	Νευρικός
- Χασμουριέται - Προσπαθεί να σταματήσει το διάβασμα μιλώντας σε κάποιον	Κουρασμένος/Νυσταγμένος
- Δεν κοιτούν στην οθόνη - Μιλάει σε κάποιον - Σταματά το διάβασμα	Με αποσπασμένη προσοχή
- Κοιτάζει επίμονα στην οθόνη - Κάνει μια αξιόλογη προσπάθεια να αλληλεπιδράσει (συνοφρύωση) - Αντιμετωπίζει μεγάλες δυσκολίες (έρχεται πολύ κοντά στην οθόνη)	Προσηλωμένος / Προσεκτικός
- Κοιτάζει επίμονα στην οθόνη - Βρίσκει κάτι αξιοσημείωτο	Με μεγάλο ενδιαφέρον



Σχήμα 5.8: Επέκταση του συστήματος εκμάθησης LTSA IEEE

επέκταση του LTSE IEEE Μοντέλου Αναφοράς [164]. Πρόκειται για ένα ολοκληρωμένο σύστημα ενσωμάτωσης προφίλ χρηστών, το οποίο προσαρμόζει το περιεχόμενο του υλικού, στις συνθήκες της αλληλεπίδρασης με το χρήστη. Το σύστημα είναι ικανό να αυτοπροσαρμόζεται συνεχώς στις καταστάσεις των χρηστών (παιδιών), ενώ έχει «μνήμη» του ιστορικού των καταστάσεών τους.

Παρακάτω περιγράφεται η αρχιτεκτονική του προτεινόμενου συστήματος ηλεκτρονικής μάθησης. Σε γενικές γραμμές, το προτεινόμενο σύστημα ηλεκτρονικής μάθησης αποτελείται από τρεις τύπους οντοτήτων: τις διαδικασίες (οβάλ σχήμα), τις αποθήκες (τετράγωνα) και τις ροές (βέλη). Κάθε μία από αυτές τις οντότητες αποτελεί ένα ενεργό σύστημα, το οποίο επιτελεί κάποιου είδους μετασχηματισμό στην είσοδό του και την επιστρέφει στην έξοδο. Πιο συγκεκριμένα, η πρώτη οντότητα επεξεργάζεται την πληροφορία που προέρχεται μέσω των ροών. Οι αποθήκες είναι χώροι στους οποίους κρατιέται πληροφορία. Στο [164], οι συγγραφείς αναλύουν τη συμπεριφορά του χρήστη με το μοντέλο αυτό, επεξεργαζόμενοι τις απαντήσεις του σε ερωτηματολόγια. Στην παρούσα δουλειά, η πληροφορία της συμπεριφοράς (καταστάσεων) του χρήστη, προέρχεται από την ανάλυση που γίνεται στα χαρακτηριστικά του προσώπου του, όπως προκύπτουν από την είσοδο στη δικτυακή κάμερα. Η συμπεριφορά του χρήστη δίνεται ως είσοδος στη διαδικασία της αξιολόγησης, η οποία δημιουργεί ροή πληροφορίας σχετική με την απόδοση του χρήστη, και καταλήγει στα αρχεία χρήστη. Η διαδικασία αξιολόγησης εμπλουτίστηκε με την πληροφορία καταστάσεων του χρήστη. Η πληροφορία απόδοσης μπορεί να προέλθει, τόσο από τη διαδικασία αξιολόγησης, όσο και από τη διαδικασία διδασκαλίας (πχ βαθμολογίες). Ο χώρος αποθήκευσης αρχείου χρήστη εμπεριέχει πληροφορία σχετιζόμενη, τόσο με παρελθοντικές καταστάσεις του χρήστη, όσο παροντικές, αλλά και μελλοντικές συνθήκες (πχ ρυθμός παρουσίασης).

Στη γενική περίπτωση, τα προφίλ των εκπαιδευομένων ορίζονται αρχικά από τους ειδικούς, και κρατιούνται στις «αποθήκες» των προφίλ των εκπαιδευομένων. Σε αυτούς τους χώρους βρίσκεται η πληροφορία που σχετίζεται με τα χαρακτηριστικά των εκπαιδευομένων, τις μαθησιακές τους ανάγκες και τις προτιμήσεις τους. Η ροή πληροφορίας που παρέχει η διαδικασία αξιολόγησης επιλέγει το προφίλ εκείνο που ταιριάζει καλύτερα με τις τρέχουσες καταστάσεις του εκπαιδευόμενου. Με αυτόν τον τρόπο,

όταν ένας εκπαιδευόμενος αντιστοιχιστεί σε συγκεκριμένο προφίλ, η διαδικασία της διδασκαλίας (coach) χρησιμοποιεί αυτήν την πληροφορία για να αναθέσει το ανάλογο πολυμεσικό υλικό.

Για να εξαχθεί το προφίλ του χρήστη, η διαδικασία αξιολόγησης λαμβάνει υπόψιν την τρέχουσα συμπεριφορά χρήστη, και τα περιεχόμενα του χώρου αποθήκευσης προφίλ του. Κατόπιν, εκτιμάται το επίπεδο προσοχής του, κατηγοριοποιώντας το προφίλ του, το οποίο πηγαίνει στο χώρο αρχειοθέτησης. Επιπλέον, ο χώρος αποθήκευσης των προφίλ των χρηστών περιέχει τα τρέχοντα προφίλ του συστήματος. Έτσι, όταν ένας χρήστης αντιστοιχηθεί σε κάποιο προφίλ, η διαδικασία διδασκαλίας εντοπίζει εκείνο το υλικό από τους αντίστοιχους χώρους αποθήκευσης που είναι το πλέον κατάλληλο στο προφίλ του χρήστη, και του το εκθέτει με τον κατάλληλο τρόπο ώστε να ανταποκρίνεται στις μαθησιακές του δυνατότητες. Για παράδειγμα, εάν ένας χρήστης είναι νευρικός, ο ρυθμός επιφώτισης λέξεων μειώνεται, ενώ μεγαλώνουν και τα γράμματα. Εάν ο χρήστης φαίνεται κουρασμένος ή δεν είναι προσεκτικός, ακούγεται κάποιος ήχος, με σκοπό να τον επαναφέρει στο υλικό διδασκαλίας. Εάν η κατάσταση του χρήστη είναι «προσεκτικός», η παρουσίαση θα συνεχίζεται με τον ίδιο ρυθμό. Αλλαγή των προφίλ των χρηστών μπορεί να γίνεται σε πραγματικό χρόνο, κατά τη διάρκεια της εκπαίδευσής τους, ενημερώνοντας το χώρο αποθήκευσης των αρχείων χρήστη. Επίσης, μέσω στατιστικών διεργασιών μπορούν να δημιουργούνται καινούρια προφίλ, ή τα ήδη υπάρχοντα να τροποποιούνται. Καινούρια και τροποποιημένα προφίλ, αποθηκεύονται στο χώρο προφίλ χρηστών.

Η οντότητα της διδασκαλίας λαμβάνει πληροφορία σχετική με την απόδοση του χρήστη, από το χώρο στον οποίο βρίσκονται τα αρχεία. Με βάση αυτήν την πληροφορία, η διαδικασία της διδασκαλίας προωθεί ερωτήματα στο χώρο αποθήκευσης εκπαιδευτικού υλικού, με σκοπό να αντλήσει το κατάλληλο. Τέλος, η διαδικασία παράδοσης μετατρέπει αυτό το υλικό σε κατάλληλη πολυμεσική παρουσίαση για τον χρήστη.

Ο χώρος αποθήκευσης των πόρων εκμάθησης είναι μια βάση που αναπαριστά γνώση, πληροφορία και άλλους πόρους απαραίτητους στη διαδικασία μάθησης. Ανταποκρίνεται στα ερωτήματα της διαδικασίας διδασκαλίας με λίστες, τις οποίες χρησιμοποιεί η διαδικασία παράδοσης για να εξάγει το μαθησιακό υλικό από την αντίστοιχη αποθήκη.

5.3.2 Καταγραφή και επισημείωση της πλατφόρμας δοκιμών

Στο πλαίσιο του έργου Agent-Dysl στήθηκε μία βάση από παιδιά με μαθησιακές δυσκολίες. Στο πείραμα συμμετείχαν παιδιά ηλικίας από 8 έως 10 ετών, τα οποία επιλέχθηκαν από ειδικούς. Τα κριτήρια της επιλογής ήταν συγκεκριμένες μαθησιακές δυσκολίες που εμφάνιζαν στην αποκωδικοποίηση γραπτού λόγου και στην παρουσίαση κειμένου στο οποίο υπάρχουν πολλά σημεία αδυναμίας των μαθητών (συγκεκριμένες λέξεις που προφέρονται λάθος, ή συνδυασμός αυτών). Οι συγκεκριμένοι μαθητές είχαν συμμετάσχει, σε πανελλήνιο επίπεδο, σε διαδικασία αξιολόγησης αποκωδικοποίησης γραπτού λόγου και παρουσίασαν τις πιο χαμηλές βαθμολογίες.

Τα βίντεο επισημειώθηκαν από ειδικούς στις εξής καταστάσεις: *Απογοητευμένος, Νευρικός, Κουρασμένος/Νυσταγμένος, Αφηρημένος, Προσηλωμένος/Προσεκτικός, Με μεγάλο ενδιαφέρον*. Αυτές οι καταστάσεις χρησιμοποιήθηκαν από τον αλγόριθμο αντιστοίχισης χαρακτηριστικών-καταστάσεων σε δεδομένα αναφοράς. Στον Πίνακα 5.3 παρουσιάζεται η κατηγοριοποίηση των πιθανών καταστάσεων του χρήστη, σύμ-

φωνα με την επισημείωση. Ειδικότερα, η κάθε μία από τις έξι καταστάσεις που μπορεί να βρεθεί ο χρήστης χαρακτηρίζεται από ένα σύνολο ενδείξεων συμπεριφοράς του χρήστη. Για παράδειγμα, στην περίπτωση που ο χρήστης φανεί ότι κινείται σπασμωδικά, η κατάσταση του χαρακτηρίζεται ως νευρικός.

Ωστόσο, υπήρξαν αρκετές δυσκολίες στη χρησιμοποίηση της βάσης: Στο συντριπτικά μεγαλύτερο μέρος της βάσης, τα παιδιά ήταν αρκετά προσηλωμένα στο καθήκον της μελέτης τους, κάνοντας, έτσι, οποιοδήποτε αλγόριθμο εκπαίδευσης να πολωθεί εάν δεν γινόταν προσεκτική επιλογή των ακολουθιών εκπαίδευσης. Επιπλέον, πολλές επισημειωμένες καταστάσεις είχαν αλληλοεπικάλυψη (πχ προσεκτικός και νυσταγμένος) σε κάποια καρέ, οπότε και εκπαιδευτήκαν διαφορετικοί ταξινομητές για κάθε κατάσταση. Ακόμη, στην επισημείωση, τα χρονικά όρια στα οποία ξεκινούσε μία κατάσταση δεν ήταν αυστηρώς ορισμένα σε σχέση με το οπτικό αποτέλεσμα που τη χαρακτήριζε. Για παράδειγμα, εάν κάποιος χαρακτηριζόταν ως *αφηρημένος* για κάποια καρέ, επειδή κοίταζε έξω από την οθόνη, τα αντίστοιχα καρέ, ενδέχεται να εμπεριείχαν στιγμιότυπα από την προηγούμενη κατάσταση. Τέλος, σε κάποιες περιπτώσεις, η επισημείωση γινόταν από παράγοντες μη σχετικούς με οπτικά χαρακτηριστικά (για παράδειγμα, το χτύπημα ενός τηλεφώνου έκανε τον ειδικό να χαρακτηρίσει το παιδί απρόσεχτο εκείνα τα δευτερόλεπτα, χωρίς οπτικά να φαίνεται κάτι τέτοιο). Οι περιορισμοί αυτοί οδήγησαν στην επιλογή λίγων, μόνο, στιγμιότυπων από τη βάση, για να γίνει εκπαίδευση των ταξινομητών, όπως επίσης περιόρισαν τον αριθμό των καταστάσεων σε τέσσερις που, για ανά δύο αλληλοαποκλειόμενες, εκπαιδευτήκε από ένα νευροσαφές σύστημα: *Προσεκτικός-Απρόσεκτος και Νευρικός-Ήρεμος*. Έτσι, συνολικά, προέκυψαν 10000 και 12250 καρέ για τον ταξινομητή *Προσεκτικός-Απρόσεκτος και Νευρικός-Ήρεμος* αντίστοιχα. Επιπλέον, κάποιες καταστάσεις ή ο αριθμός τους δεν ήταν επαρκής για την εκπαίδευση συστήματος εκτίμησης κατάστασης του μαθητή. Για παράδειγμα, παρατηρήθηκε ότι περιστροφή του κεφαλιού, με ταυτόχρονη περιστροφή του βλέμματος προς αντίθετη κατεύθυνση ήταν πολύ σπάνια μέσα στη βάση, κάτι το οποίο μας ώθησε στη θεώρηση μονάχα απόλυτων τιμών των ίδιων μεγεθών.

5.3.3 Υπολογισμός χαρακτηριστικών ανάλυσης κίνησης κεφαλιού και προσώπου, για εκτίμηση επιπέδων προσοχής και νευρικότητας

Η διαδικασία εξαγωγής χαρακτηριστικών προσώπου, παρακολούθησης των θέσεων και υπολογισμού του μέτρου της περιστροφής του κεφαλιού (στο εξής περιστροφή κεφαλιού), έγινε σύμφωνα με τις αντίστοιχες μεθοδολογίες που περιγράφονται στα κεφάλαια 2 και 3.

Στη συγκεκριμένη δουλειά, όπως εξηγήθηκε και παραπάνω, χρησιμοποιούνται περισσότερο χαρακτηριστικά, πέρα από την περιστροφή του προσώπου. Καθώς τα πειράματα έγιναν σε πραγματικές συνθήκες και, μάλιστα, με ένα σύνολο ιδιαίτερα κινητικών και εκφραστικών χρηστών, η πρόκληση της χρήσης μεταβλητών που εξαρτώνται από λεπτομερείς κινήσεις του προσώπου, ήταν αρκετά έντονη. Αν και δεν υπήρξε τρόπος να ελεγχθεί η ακρίβεια της παρακολούθησης του κάθε χαρακτηριστικού, μία συνολική αποτίμηση θα γίνει, έστω και έμμεσα, παρακάτω σε αυτό το κεφάλαιο, όπου θα φανεί, μέσα από στατιστική αποτίμηση, η σπουδαιότητα του κάθε ενός, ή μη. Πάντως, πρέπει να τονιστεί ότι, μία ποιοτική κρίση των αποτελεσμάτων στην εκτίμηση βλέμματος, στις κινήσεις των φρυδιών και του στόματος, έδειξε ότι, ειδικά για μετωπικές στάσεις του προσώπου, τα αποτελέσματα είναι αρκετά αξιόπιστα, ενώ η ευρωστία εξαρτάται

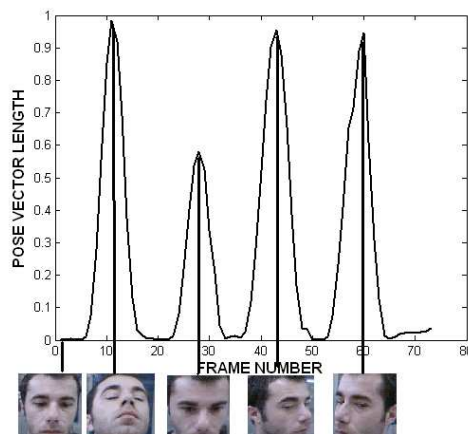
σε μεγάλο βαθμό από την επιτυχία του αλγορίθμου παρακολούθησης [19].

5.3.3.1 Εκτίμηση βλέμματος ματιών

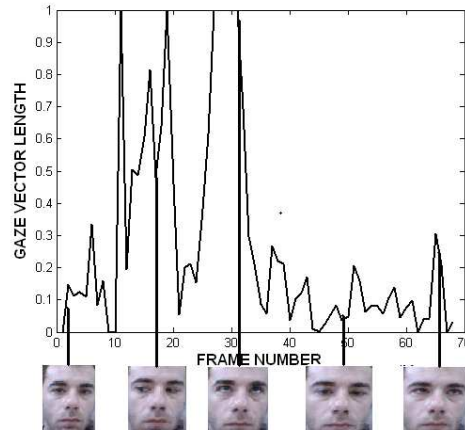
Η εύρεση της κατευθυντικότητας του βλέμματος μπορεί να γίνει με την παρακολούθηση του κέντρου του ματιού, σε σχέση με ένα σύστημα συντεταγμένων ορισμένο από τα τέσσερα σημεία που ορίζουν την περιοχή του ματιού σε κάθε καρέ (σχήμα 5.10). Όπως και στην περίπτωση της περιστροφής κεφαλιού, προκύπτει ένα διάνυσμα, το οποίο κανονικοποιείται με την απόσταση μεταξύ των οφθαλμών, όπως υπολογίζεται σε καρέ εκκίνησης, ή επανεκκίνησης του συστήματος. Με αυτόν τον τρόπο, επιτυγχάνεται η αποσύνδεση του συστήματος από τον παράγοντα κλιμάκωσης του προσώπου στην εικόνα, και από την κανονικοποιημένη μετατόπιση του κέντρου, προκύπτει το διάνυσμα $\underline{g} = (\underline{g}_x, \underline{g}_y)$, όπου g_x και g_y είναι οι οριζόντιες και κάθετες συνιστώσες (σε απόλυτη τιμή), αντίστοιχα.

5.3.3.2 Υπολογισμός περισσότερων χαρακτηριστικών

Στο σύστημα μπορούν να εισηχθούν και οι κινήσεις των φρυδιών, οι οποίες μετρώνται με σημείο αναφοράς το ανώτερο σημείο του κάθε ματιού και την απόσταση από αυτό σε σχέση με καρέ εκκίνησης (ή επανεκκίνησης). Αντίστοιχες μετρήσεις έγιναν και για το οριζόντιο και κατακόρυφο άνοιγμα του στόματος. Επιπλέον, συγκρίνοντας την απόσταση μεταξύ των κέντρων των ματιών (με δεδομένη μετωπική στάση του κεφαλιού σε σχέση με την κάμερα) σε σχέση με αυτήν σε καρέ εκκίνησης (ή επανεκκίνησης), μπορεί να γίνει ποιοτική εκτίμηση των μετατοπίσεων του χρήστη σε σχέση με το επίπεδο της κάμερας. Για όλα τα παραπάνω, τα μετρούμενα μεγέθη συνίστανται στο λόγο της τρέχουσας μέτρησης του εκάστοτε μεγέθους, σε σχέση με την αρχική, όπως μετρήθηκε στο αντίστοιχο καρέ εκκίνησης. Τέλος, το σύστημα χρησιμοποιεί και το μέτρο της ταχύτητας κίνησης του κεφαλιού (στο εξής ταχύτητα κίνησης του κεφαλιού) (σε εικονοστοιχεία/καρέ), ως μέσο όρο της κίνησης όλων των χαρακτηριστικών σημείων.



Σχήμα 5.9: Μέτρο διανύσματος περιστροφής κεφαλιού κατά τη διάρκεια ακολουθίας.



Σχήμα 5.10: Μέτρο διανύσματος κατευθυντικότητας βλέμματος ματιών κατά τη διάρκεια ακολουθίας.

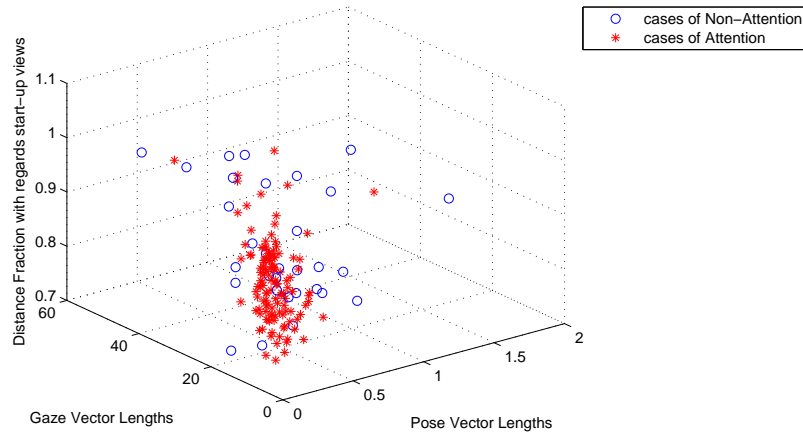
5.3.4 Στατιστικός υπολογισμός καταλληλότερων χαρακτηριστικών

Για να αξιολογηθεί η καταλληλότητα καθενός από τα χαρακτηριστικά, χρησιμοποιήθηκε ο ακριβής έλεγχος Fisher [45]. Για το σκοπό αυτό, υπολογίστηκε το ιστογράμμα τριών κλάσεων του κάθε χαρακτηριστικού, και συγκρίθηκε με την αντίστοιχη κατανομή του ίδιου χαρακτηριστικού για τις περιπτώσεις εκείνες που η κατάσταση ήταν θετική (προσεκτικός, ήρεμος). Τα ιστογράμματα κανονικοποιήθηκαν στην κλίμακα επί τοις εκατό ώστε να είναι συγκρίσιμα. Ο ακριβής έλεγχος Fisher επιλέχθηκε ως ο καταλληλότερος (πχ έναντι της chi-square μεθόδου), καθώς είναι ιδανικός για δεδομένα μικρής κλίμακας. Πράγματι, σε αρκετές περιπτώσεις (για παράδειγμα, οριζόντια ταχύτητα κεφαλιού όταν ο χρήστης είναι νευρικός) κάποιες κλάσεις του αντίστοιχου ιστογράμματος αντιστοιχίζονται σε πολύ χαμηλές τιμές. Ο ακριβής έλεγχος Fisher είναι ιδανικός για μικρής κλίμακας δεδομένα.

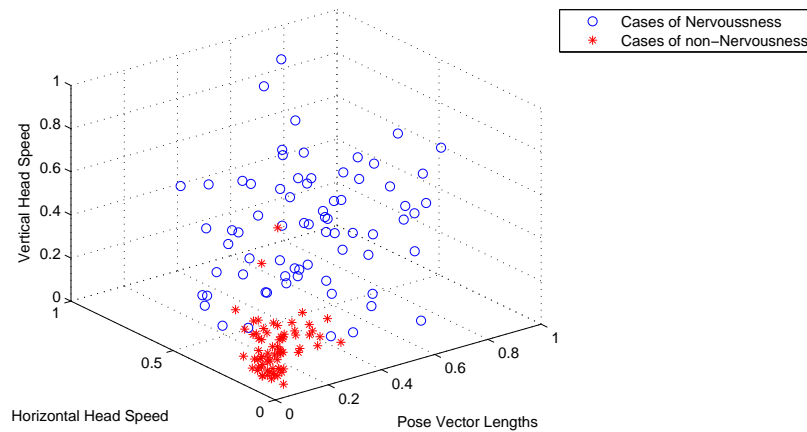
Ο ακριβής έλεγχος Fisher έδειξε ότι, για τα πειράματα καταστάσεων Προσεκτικός-Απόσεκτος, και τα χαρακτηριστικά Περιστροφή κεφαλιού, βλέμμα ματιών, αλλαγές στην απόσταση ανάμεσα στους οφθαλμούς, ταχύτητα κίνησης κεφαλιού, η μηδενική υπόθεση (ότι οι παρατηρούμενες και οι αναμενόμενες κατανομές δε διαφέρουν) θα πρέπει να απορριφθεί με μεγάλο βαθμό εμπιστοσύνης ($p < 0.05$) σε σχέση με τα υπόλοιπα χαρακτηριστικά. Για τα πειράματα Νευρικός-Ήρεμος, τα χαρακτηριστικά Περιστροφή Κεφαλιού, Οριζόντια και Κάθετη ταχύτητα κεφαλιού και κατευθυντικότητα βλέμματος ματιών είναι χαρακτηριστικά που θα πρέπει να ληφθούν υπόψη ($p < 0.05$). Στα σχήματα 5.11α και 5.11β φαίνονται παραδείγματα δεδομένων για την κάθε κλάση.

5.3.5 Πειραματικά αποτελέσματα

Για την εκπαίδευση ενός συστήματος ταξινόμησης των χαρακτηριστικών σε καταστάσεις (κλάσεις), χρησιμοποιήθηκαν ασαφή συστήματα συγκερασμού. Η ιδέα πίσω από τη χρήση ασαφούς λογικής έγκειται στο γεγονός ότι ο βαθμός συμμετοχής σε κάποια κατάσταση δε μπορεί να παίρνει αυστηρά τιμή 0 ή 1, αλλά να έχει διαβαθμίσεις. Επιπλέον, καθώς είναι η πρώτη φορά που επιχειρείται να γίνει αντιστοίχιση μεταξύ χαρακτηριστικών όπως περιστροφή και ταχύτητα κίνησης κεφαλιού σε έννοιες όπως «αφηρημένος», «νευρικός», με στατιστικό τρόπο, είναι αβέβαιη η υπακοή των



α



β

Σχήμα 5.11: (α) Χαρακτηριστικά που χρησιμοποιούνται στο σενάριο Προσεκτικός-Απόσεκτος (β) Χαρακτηριστικά που χρησιμοποιούνται στο σενάριο Νευρικός-Ήρεμος

χαρακτηριστικών σε συγκεκριμένες κατανομές ή γραμμικές σχέσεις σε σχέση με τις καταστάσεις.

5.3.5.1 Μοντέλο Takagi-Sugeno-Kang

Η ασαφής μέθοδος συγκερασμού Takagi-Sugeno-Kang προτάθηκε το 1985 [73] και αποτελείται από κανόνες της μορφής (για μία έξοδο):

Κανόνας R^k : $IF x_1 is M_1^k AND \dots AND x_m is M_m^k THEN$
 $y^k(x) = w_0^k + \sum_{i=1}^m w_i^k x_i$

Η έξοδος y^k του κάθε κανόνα προέρχεται από τη γραμμική συνάρτηση που την χαρακτηρίζει, ενώ στην τελική έξοδο πολλαπλασιάζεται με το βάρος w^k του κάθε κανόνα. Το βάρος w^k προκύπτει από τον τελεστή AND σε όλες τις εισόδους (εξίσωση 5.3).

$$w^k = \text{AndMethod}(M_1^k(x_1) \dots M_n^k(x_n)) \quad (5.3)$$

όπου n ο αριθμός των εισόδων, και $M_j^k(x_j)$ η τιμή που παίρνει η συνάρτηση συμμετοχής που αντιστοιχεί στον κανόνα k και τη μεταβλητή x_j .

Η τελική έξοδος του συστήματος είναι ο σταθμισμένος μέσος των εξόδων όλων των κανόνων (εξίσωση 5.4)

$$\text{Output} = \frac{\sum_{k=1}^K w^k y^k}{\sum_{k=1}^K w^k} \quad (5.4)$$

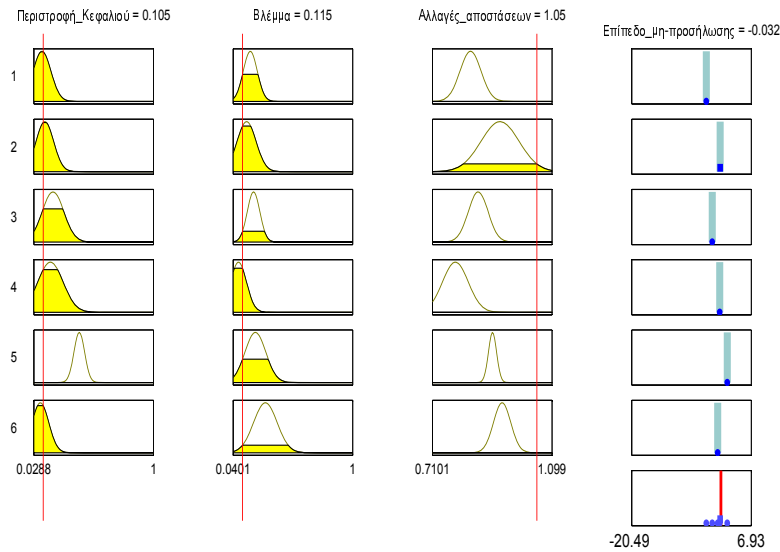
όπου K το σύνολο των κανόνων.

Πριν την εκπαίδευση των παραμέτρων του συστήματος συμπερασμού (συναρτήσεις συμμετοχής εισόδων, παράμετροι εξόδου του κάθε κανόνα), αποφασίζεται ο βέλτιστος αριθμός κανόνων, με βάση τον αλγόριθμο ταξινόμησης υποσυστάδας [26]. Σύμφωνα με αυτόν, κάθε σημείο εισόδου θεωρείται πιθανό κέντρο συστάδας, και υπολογίζεται η αντίστοιχη πιθανότητα με βάση την πυκνότητα των σημείων που το περιβάλλουν. Κατόπιν, ο αλγόριθμος επιλέγει εκείνο το σημείο με την υψηλότερη πιθανότητα να αποτελεί κέντρο συστάδας, και αφαιρεί από την διαδικασία τα γειτονικά του (σε ακτίνα r). Η διαδικασία επαναλαμβάνεται με τα υπόλοιπα σημεία, μέχρις ότου όλα ανήκουν σε κάποια συστάδα. Ο αριθμός των συστάδων είναι και ο αριθμός των κανόνων του TSK συστήματος.

Η εκπαίδευση των παραμέτρων (συναρτήσεις συμμετοχής και τμήμα συμπερασμού των κανόνων) γίνεται με τη μέθοδο «Ασαφής Συστήματα Συμπερασμού βασισμένα σε Προσαρμοστικά Δίκτυα» (Adaptive-Network-based Fuzzy Inference Systems - ANFIS) [73]. Με αυτόν τον τρόπο, η εκπαίδευση γίνεται με έναν υβριδικό τρόπο: Χρήση ελαχίστων τετραγώνων και βελτιστοποίηση επικλινούς καθόδου με οπισθοδιάδοση σφάλματος (backpropagation gradient descent). Η χρήση ελαχίστων τετραγώνων γίνεται για τον προσδιορισμό των παραμέτρων συμπερασμού του κάθε κανόνα (αρκετά δεδομένα εκπαίδευσης να είναι αρκετά περισσότερα από τις προς εκπαίδευση παραμέτρους), αφού υπάρχει γραμμική σχέση μεταξύ της εξόδου του συστήματος και των παραμέτρων w_i^k . Αυτή είναι η σάρωση προς τα εμπρός, και γίνεται με σταθερές τις υπόλοιπες παραμέτρους. Στην οπισθοδιάδοση σφαλμάτων υπολογίζονται οι υπόλοιπες παράμετροι (κέντρα και εύρη γκαουσιανών συναρτήσεων συμμετοχής) με χρήση επικλινούς μεθόδου βελτιστοποίησης. Η χρήση του υβριδικού αυτού μοντέλου (και όχι αποκλειστική χρήση της επικλινούς καθόδου) έγινε με σκοπό την αύξηση της ανθεκτικότητας του συστήματος [73] σε τοπικά ελάχιστα, της μείωσης του εύρους αναζήτησης παραμέτρων βελτιστοποίησης των παραμέτρων των συναρτήσεων συμμετοχής, και την συνακόλουθη μείωση σε υπολογιστική πολυπλοκότητα.

5.3.5.2 Ασαφές σύστημα συμπερασμού για εκτίμηση επιπέδων προσήλωσης

Για τα χαρακτηριστικά που μετρήθηκαν (Περιστροφή κεφαλιού, κατευθυντικότητα βλέμματος, οριζόντια/κάθετη ταχύτητα κίνησης κεφαλιού, κινήσεις φρυδιών, οριζόντιο και κάθετο άνοιγμα στόματος, κινήσεις κάθετα στο επίπεδο της κάμερας), υπολογίστηκε ο μέσος όρος τους σε διαστήματα των 25 καρτέ, και έγινε η αντιστοίχιση με τα επισημειωμένα δεδομένα αναφοράς. Κατόπιν, οι μετρήσεις που χρησιμοποιούνται



Σχήμα 5.12: Ασαφές TSK και αντίστοιχη έξοδος για είσοδο Περιστροφή Κεφαλιού=0.105, Βλέμμα=0.115 και Απόσταση=1.05

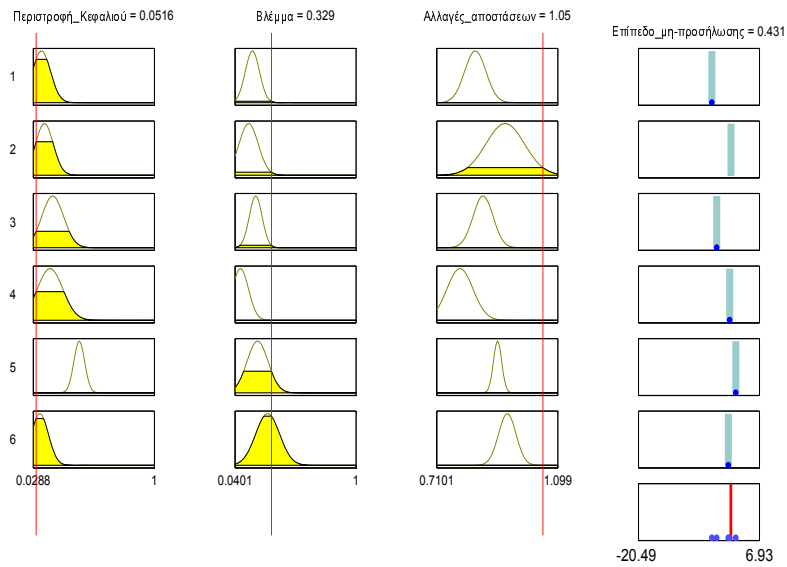
για εκπαίδευση κανονικοποιήθηκαν από 0 μέχρι 1^3 και η κανονικοποίηση αυτή γίνεται και για τα δεδομένα επαλήθευσης. Για το πρόβλημα Προσεκτικός-Απρόσεκτος η τιμή 0 της εξόδου αντιστοιχίζεται στην κλάση Προσεκτικός, ενώ η τιμή 1 στην κλάση Απρόσεκτος. Για την εκπαίδευση χρησιμοποιήθηκαν γκαουσιανές συναρτήσεις συμμετοχής και ο αλγόριθμος εκπαίδευσης έτρεξε για 100 επαναλήψεις. Η ακτίνα που χρησιμοποιήθηκε για την ομαδοποίηση των χαρακτηριστικών σε υποσυστάδες [26] ήταν 0.2 (δυσνητικά, σε κανονικοποιημένα δεδομένα, πχ. τριών εισόδων, ο μέγιστος δυνατός αριθμός κανόνων είναι $5^3=125$), ενώ, λόγω του γεγονότος ότι τα δεδομένα είναι λίγα, γινόταν διαφορετική εκπαίδευση σε κάθε ακολουθία παιδιού, χρησιμοποιώντας τις ακολουθίες από τα υπόλοιπα (*leave-one-out*).

Ο αλγόριθμος συσταδοποίησης των δεδομένων από όλα τα παιδιά, και με χρήση των χαρακτηριστικών Περιστροφή Κεφαλιού, Βλέμμα, Αλλαγές Αποστάσεων ως προς την κάμερα, οδήγησε σε ασαφές TSK σύστημα 6 κανόνων, της μορφής που φαίνεται στα σχήματα 5.12-5.14.

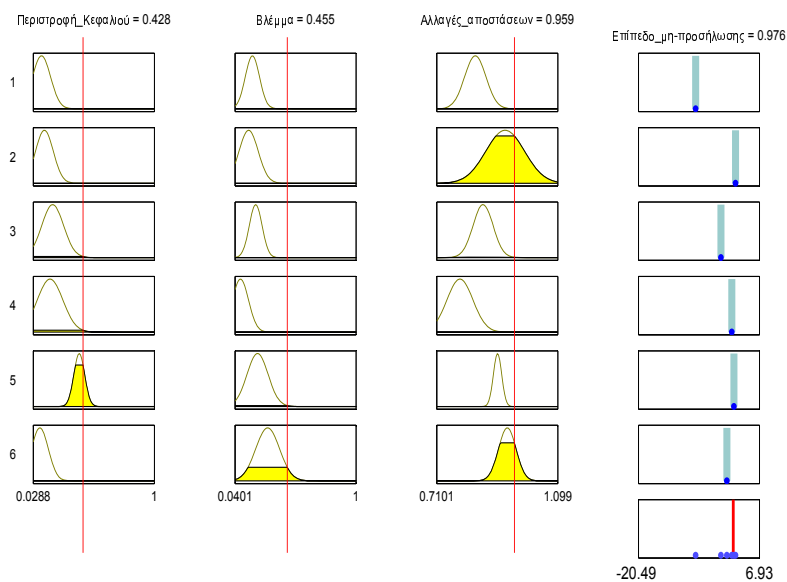
Στα σχήματα 5.15-5.17 φαίνονται οι τιμές που παίρνει η έξοδος (επίπεδα μη προσήλωσης) για διάφορους συνδυασμούς εισόδων. Στην εικόνα 5.15, για παράδειγμα, μεγάλες τιμές στο διάνυσμα που αντιστοιχεί στην κατευθυντικότητα βλέματος αντιστοιχούν σε υψηλές τιμές μη-προσήλωσης, ενώ μικρές τιμές αντιστοιχούν σε χαμηλές τιμές της παραμέτρου μη-προσήλωσης. Αξίζει, εδώ, να σημειωθεί ότι, λόγω των δεδομένων, δεν υπήρχαν στιγμιότυπα με μεγάλη περιστροφή κεφαλιού και ταυτόχρονη μικρή περιστροφή του βλέματος των ματιών. Σε αυτήν την περίπτωση, εισαγωγή γνώσης από ανθρώπους στην διαμόρφωση των κατάλληλων συναρτήσεων συμμετοχής [73] θα ήταν απαραίτητη για την ενσωμάτωση ενός τέτοιου κανόνα. Αντίστοιχο πρόβλημα προέκυψε και στη συλλογή δεδομένων τέτοιων, όπου να υπάρχει κίνηση κεφαλιού και ματιών προς αντίθετες φορές.

Στο σχήμα 5.16 φαίνεται ότι μικρές αλλαγές στην απόσταση του χρήστη από την

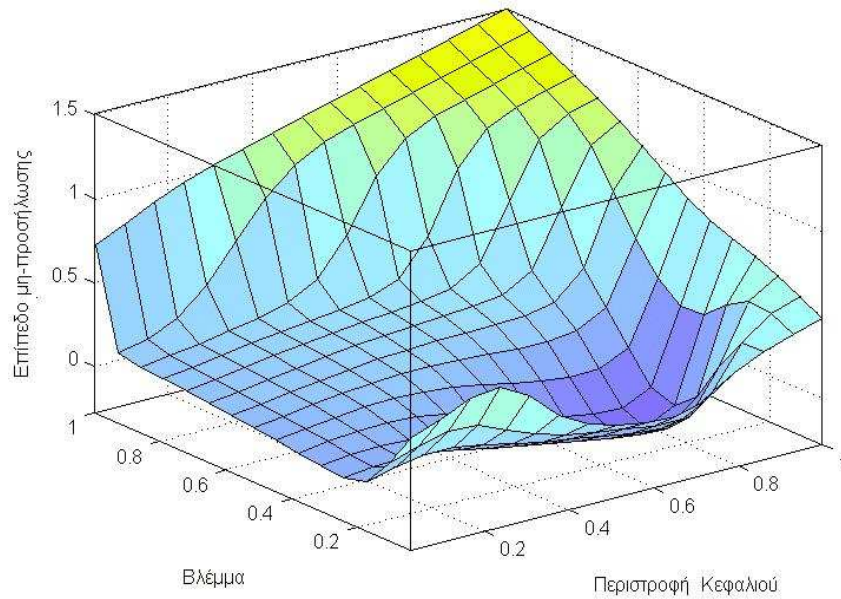
³κανονικοποιήθηκαν όλες οι παράμετροι, εκτός από αυτήν που αντιστοιχεί στις αλλαγές αποστάσεων, και η οποία, εξ ορισμού, παίρνει τιμές κοντά στη μονάδα.



Σχήμα 5.13: Ασαφές TSK και αντίστοιχη έξοδος για είσοδο Περιτροφή Κεφαλιού=0.052, Βλέμμα=0.33 και Απόσταση=1.05



Σχήμα 5.14: Ασαφές TSK και αντίστοιχη έξοδος για είσοδο Περιτροφή Κεφαλιού=0.428, Βλέμμα=0.455 και Απόσταση=0.96

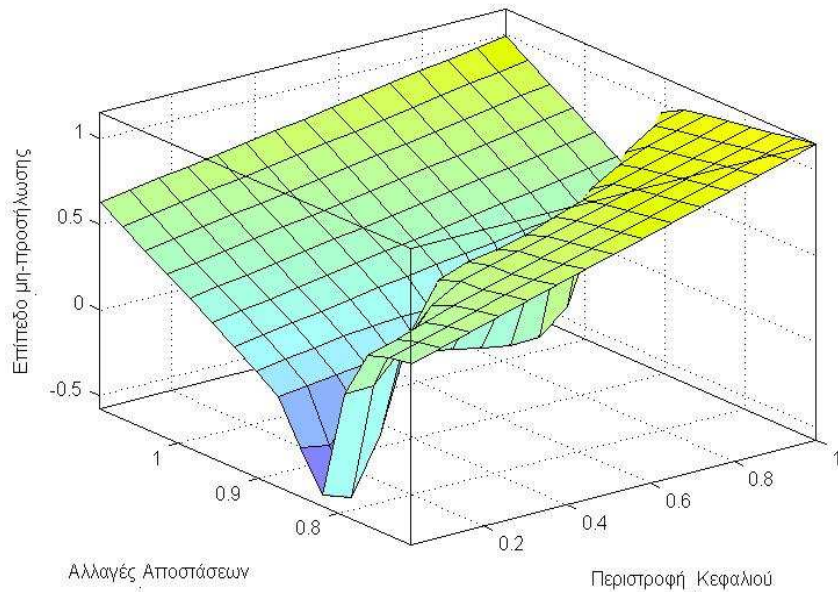


Σχήμα 5.15: Αντιστοίχιση Περιστροφής Κεφαλιού και Βλέμματος με έξοδο Μη-Προσήλωση

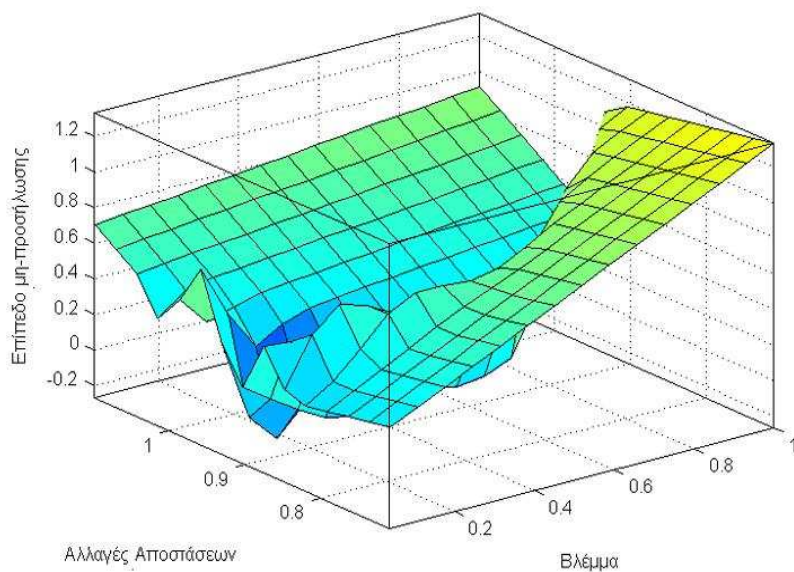
κάμερα (τιμές κοντά στη μονάδα), σε συνδυασμό με μικρές τιμές του διανύσματος περιστροφής κεφαλιού, οδηγούν σε χαμηλά επίπεδα του συμπεράσματος *Μη-Προσήλωση*. Αντίθετα, για μεγάλες αποκλίσεις από την αρχική απόσταση του κεφαλιού από την κάμερα, συμπεραίνεται μεγάλος βαθμός μη-προσήλωσης, ο οποίος δυσχεραίνεται ακόμη περισσότερο όσο αυξάνει η περιστροφή του κεφαλιού.

Αντίστοιχη μορφή παίρνει και η επιφάνεια που απεικονίζεται στο σχήμα 5.17 όπου, αντί για την περιστροφή κεφαλιού, χρησιμοποιείται σαν είσοδος η κατευθυντικότητα του βλέμματος. Και εδώ, μικρές αποκλίσεις από την αρχική απόσταση χρήστη-κάμερας, σε συνδυασμό με βλέμμα στραμμένο προς την οθόνη, αντιστοιχίζονται σε μικρά επίπεδα μη-προσήλωσης, ενώ, μεγάλες αποκλίσεις έχουν σαν αποτέλεσμα υψηλά επίπεδα μη-προσήλωσης, τα οποία γίνονται ακόμη υψηλότερα, όσο αυξάνει σε μέγεθος το διάνυσμα κατευθυντικότητας βλέμματος.

Στον πίνακα 5.4 φαίνονται τα συνολικά ποσοστά επιτυχίας στην ταξινόμηση καταστάσεων σύμφωνα με τα δεδομένα αναφοράς. Έγιναν δοκιμές με χρήση των χαρακτηριστικών *Περιστροφή Κεφαλιού*, *Βλέμμα ματιών*, *Αλλαγές αποστάσεων* και με χρήση των παραπάνω συν την *ταχύτητα κίνησης του κεφαλιού*. Έξοδοι του ασαφούς μεγαλύτερες από 0.5 αντιστοιχήθηκαν σε τιμή 1 (*μη προσεκτικός*). Τα αποτελέσματα, χωρίς χρήση της ταχύτητας, εμφανίζονται ελαφρώς βελτιωμένα. Αυτό μπορεί να οφείλεται στο ότι, μεγάλες ταχύτητες αποτελούν πάντα μέρος των χρονικών περιόδων μη προσεκτικής κατάστασης (οπότε και το p-value του στατιστικού ελέγχου είναι χαμηλό). Ωστόσο, κάτι τέτοιο δε σημαίνει ότι, όση ώρα το παιδί ήταν μη προσεκτικό, η ταχύτητα κίνησης έπαιρνε μεγάλες τιμές. Θα ήταν αναμενόμενο, λοιπόν, αυτή η γνώση να δημιουργούσε κανόνες της μορφής: *ΜΕΓΑΛΗ ΤΑΧΥΤΗΤΑ AND ΟΠΟΙΕΣΔΗΠΟΤΕ ΤΙΜΕΣ ΓΙΑ ΤΑ ΥΠΟΛΟΙΠΑ == ΜΗ ΠΡΟΣΗΛΩΣΗ*. Κάτι τέτοιο δεν συνέβη, δυστυχώς, με τις παραμέτρους με τις οποίες έγινε η εκπαίδευση, ίσως λόγω και του χαμηλού αριθμού των δεδομένων εκπαίδευσης.



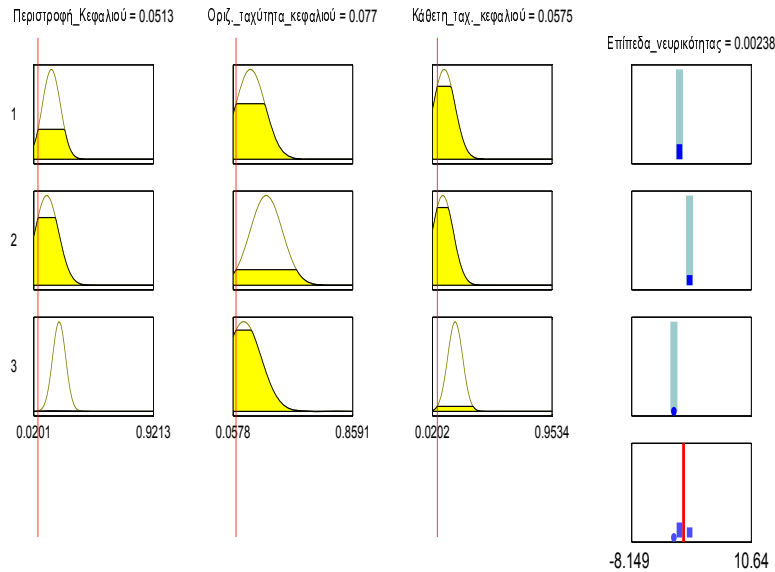
Σχήμα 5.16: Αντιστοίχιση Περιστροφής Κεφαλιού και Αλλαγές αποστάσεων χρήστη-κάμερας με έξοδο Μη-Προσέγγιση



Σχήμα 5.17: Αντιστοίχιση Βλέμματος και Αλλαγές αποστάσεων χρήστη-κάμερας με έξοδο Μη-Προσέγγιση

Πίνακας 5.4: Ακρίβεια νευρο-ασαφούς στην εκτίμηση επιπέδων προσέγγισης στη βάση του Agent-Dysl με χρήση δύο διαφορετικών συνόλων χαρακτηριστικών

	Συνολικά Ποσοστά επιτυχίας
Περιστροφή Κεφαλιού, Κατευθυντικότητα βλέμματος, Αλλαγές Αποστάσεων, Ταχύτητα κίνησης κεφαλής	84.00%
Περιστροφή Κεφαλιού, Κατευθυντικότητα βλέμματος, Αλλαγές Αποστάσεων	88.00%



Σχήμα 5.18: Ασαφές TSK και αντίστοιχη έξοδος για είσοδο Περιστροφή Κεφαλιού=0.051,οριζόντια ταχύτητα=0.077 και κάθετη ταχύτητα=0.058

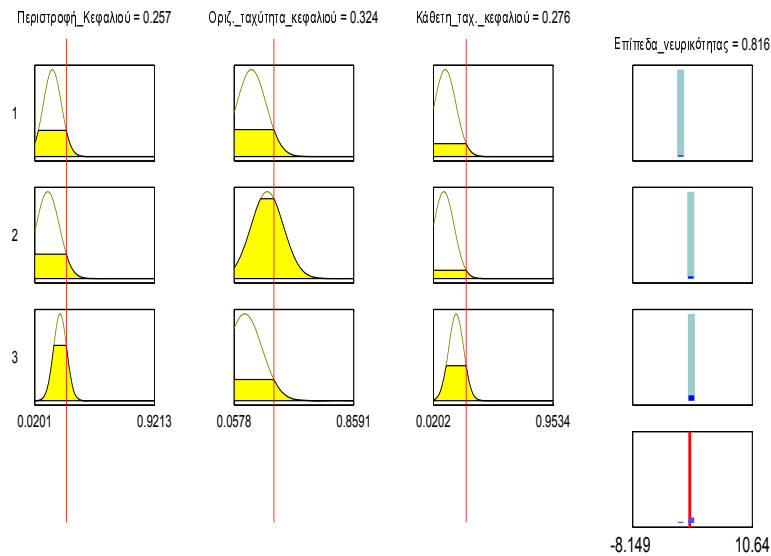
5.3.5.3 Ασαφές σύστημα συμπερασμού για εκτίμηση επιπέδων νευρικότητας

Παρόμοια με το πρόβλημα ταξινόμησης *Προσεκτικός-Απρόσεκτος*, για το πρόβλημα *Νευρικότητα-Ηρεμία* η τιμή 0 της εξόδου αντιστοιχίζεται στην κλάση *Ηρεμία*, ενώ η τιμή 1 στην κλάση *Νευρικότητα*. Για την εκπαίδευση χρησιμοποιήθηκαν γκαουσιανές συναρτήσεις συμμετοχής και ο αλγόριθμος εκπαίδευσης έτρεξε για 100 επαναλήψεις. Η ακτίνα που χρησιμοποιήθηκε για την ομαδοποίηση των χαρακτηριστικών σε υποσυστάδες [26] ήταν 0.2 (δυνητικά, σε κανονικοποιημένα δεδομένα, πχ. τριών εισόδων, ο μέγιστος αριθμός κανόνων είναι $5^3=125$), ενώ, λόγω του γεγονότος ότι τα δεδομένα μας είναι λίγα, γινόταν διαφορετική εκπαίδευση σε κάθε ακολουθία παιδιού, χρησιμοποιώντας τις ακολουθίες από τα υπόλοιπα (*leave-one-out*).

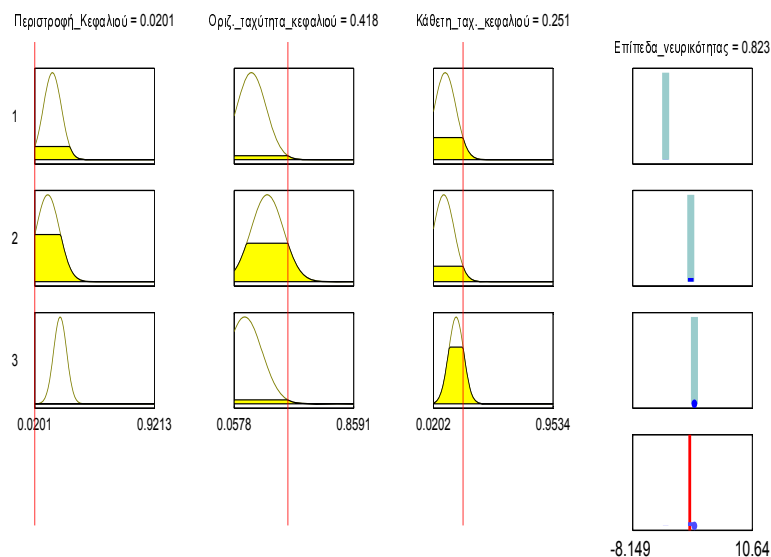
Ο αλγόριθμος συσταδοποίησης δεδομένων από όλα τα παιδιά, και με χρήση των χαρακτηριστικών Περιστροφή Κεφαλιού, οριζόντια και κάθετη ταχύτητα κίνησης κεφαλιού, οδήγησε σε ένα ασαφές TSK σύστημα 3 κανόνων, της μορφής που φαίνεται στα σχήματα 5.18-5.20.

Στα σχήματα 5.21-5.22 φαίνονται οι τιμές που παίρνει η έξοδος (επίπεδα νευρικότητας) για διάφορους συνδυασμούς εισόδων. Πιο συγκεκριμένα, στο σχήμα 5.21 φαίνεται ότι, κάθετη κίνηση κεφαλιού και μικρή περιστροφή κεφαλιού ευνοεί την κατάσταση μη-νευρικότητα (ίσως στην επισημείωση μια τέτοια κίνηση να ήταν δείγμα κατάστασης *κατανόησης*), ενώ στο σχήμα 5.22 φαίνεται ότι μέτρια περιστροφή κεφαλιού και μικρή οριζόντια ταχύτητα ευνοεί ταξινόμηση του χρήστη ως νευρικός, ενώ μεγάλη περιστροφή κεφαλιού, ή μεγάλη οριζόντια ταχύτητα ευνοούν την κατάσταση νευρικότητας.

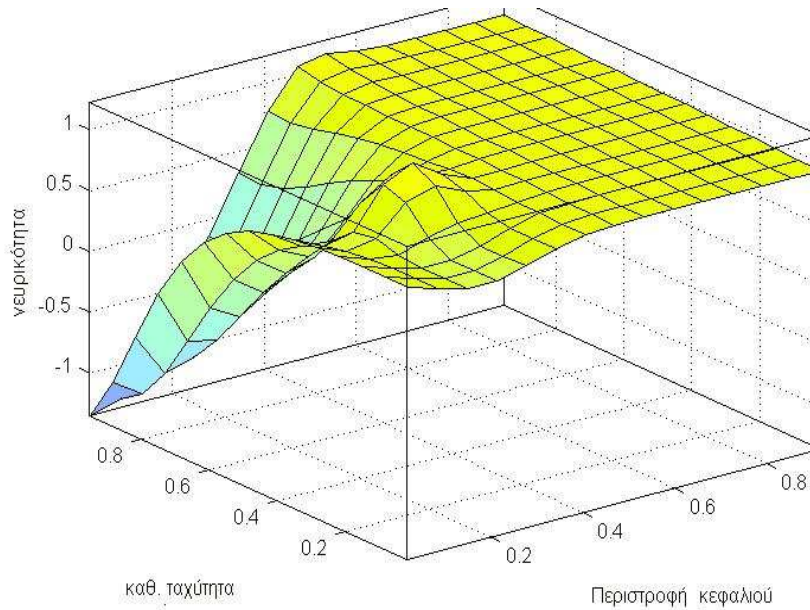
Στον πίνακα 5.5 φαίνονται τα συνολικά ποσοστά επιτυχίας στην ταξινόμηση καταστάσεων σύμφωνα με τα δεδομένα αναφοράς. Έγιναν δοκιμές με χρήση των χαρακτηριστικών *Περιστροφή Κεφαλιού*, *οριζόντια και κάθετη ταχύτητα* και με χρήση των παραπάνω συν την *κατευθυντικότητα του βλέμματος*. Όπως και προηγουμένως, έτσι κι εδώ, έξοδοι του ασαφούς συστήματος μεγαλύτερες από 0.5 αντιστοιχήθηκαν σε κατάσταση νευρικότητας, και μικρότερες από 0.5, σε κατάσταση ηρεμίας. Από τον



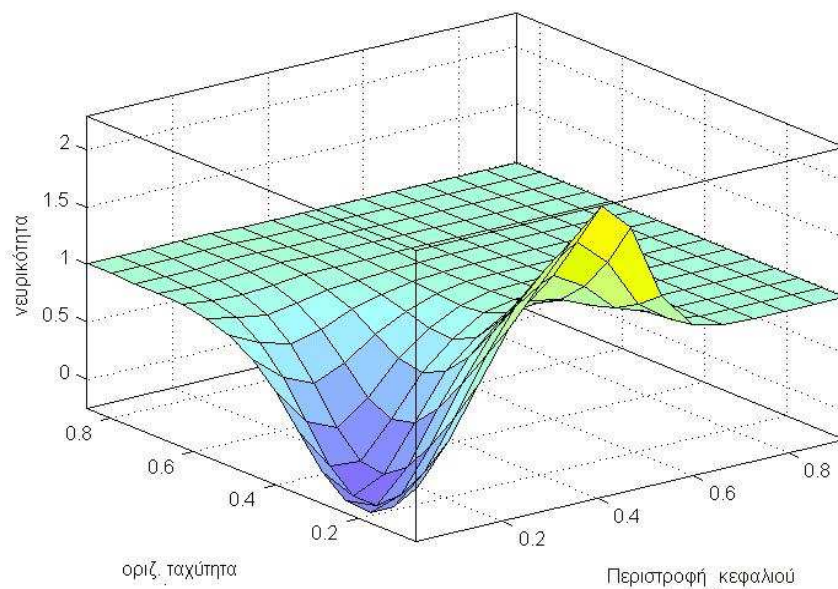
Σχήμα 5.19: Ασαφές TSK και αντίστοιχη έξοδος για είσοδο Περιστροφή Κεφαλιού=0.257, οριζόντια ταχύτητα=0.324 και κάθετη ταχύτητα=0.276



Σχήμα 5.20: Ασαφές TSK και αντίστοιχη έξοδος για είσοδο Περιστροφή Κεφαλιού=0.020, οριζόντια ταχύτητα=0.418 και κάθετη ταχύτητα=0.251



Σχήμα 5.21: Αντιστοίχιση Περιστροφής Κεφαλιού και κάθετης ταχύτητας κίνησης κεφαλιού με έξοδο Νευρικότητα



Σχήμα 5.22: Αντιστοίχιση Περιστροφής Κεφαλιού και οριζόντιας ταχύτητας κίνησης κεφαλιού με έξοδο Νευρικότητα

Πίνακας 5.5: Ακρίβεια νευρο-ασαφούς στην εκτίμηση επιπέδων νευρικότητας στη βάση του *Agent-Dysl* με χρήση δύο διαφορετικών συνόλων χαρακτηριστικών

	Συνολικά Ποσοστά επιτυχίας
Περιστροφή Κεφαλιού, οριζόντια συνιστώσα ταχύτητας κίνησης κεφαλιού, κάθετη συνιστώσα ταχύτητας κίνησης κεφαλιού	82.00%
Περιστροφή Κεφαλιού, οριζόντια συνιστώσα ταχύτητας κίνησης κεφαλιού, κάθετη συνιστώσα ταχύτητας κίνησης κεφαλιού, βλέμμα ματιών	80.63%

πίνακα, φαίνεται ότι η χρήση της κατευθυντικότητας των ματιών, δεν άλλαξε σημαντικά τα αποτελέσματα. Κάτι τέτοιο, ύστερα από παρατήρηση των δεδομένων, φάνηκε ότι οφείλεται στο ότι η κατευθυντικότητα του βλέμματος ήταν αρκετά συσχετισμένη με το διάνυσμα περιστροφής κεφαλιού ($r=0.53$, $p < 0.05$).

5.4 Συμπεράσματα

Στο κεφάλαιο αυτό παρουσιάστηκε μία διεπαφή διερεύνησης συμπεριφορών από κοινού προσοχής με εικονικό πράκτορα, σε μη παρεμβατικό περιβάλλον παρακολούθησης βλέμματος. Κατά τη διάρκεια των πειραμάτων ο εικονικός πράκτορας μιλούσε στον χρήστη, παρουσιάζοντάς του διάφορα αντικείμενα. Για τον υπολογισμό της κατευθυντικότητας του βλέμματος του χρήστη χρησιμοποιήθηκε απλή δικτυακή κάμερα, σε πραγματικό χρόνο. Κάτι τέτοιο ήταν προτεραιότητα στην εφαρμογή, αφού υπήρχε απαίτηση να μεγιστοποιείται η φυσικότητα της αλληλεπίδρασης, με χρήση μη παρεμβατικών, φθηνών μηχανισμών, ευρέως διαδεδομένων στο εμπόριο. Αποδείχτηκε ότι η μέθοδος ανάκτησης βλέμματος είναι αρκετά αποτελεσματική για τις ανάγκες του σεναρίου που εξετάστηκε, αποβάλλοντας την ανάγκη για παρεμβατικές συσκευές. Επιπλέον, στα πειράματα μελετήθηκαν θέματα αναφορικά με αλληλεπίδραση ανθρώπου με εικονικό πράκτορα σε συνθήκες από κοινού προσοχής: Ενώ οι συμμετέχοντες εμφανίστηκαν να ακολουθούν, πρωτίστως, την προφορική πληροφορία από τον πράκτορα, για να κατευθύνουν το βλέμμα τους στα αντικείμενα, τα πειράματα παρέχουν απόδειξη ότι είναι κάθε άλλο παρά αδιάφοροι στην οπτική πληροφορία που τους παρέχει ο ίδιος ο πράκτορας. Ανάλογα με το που κατηύθυνε το βλέμμα του, οι χρήστες κοιτούσαν προς τα εκεί, μέχρι να διαπιστώσουν εάν ήταν αναγκαίο να ψάξουν το σωστό αντικείμενο στη σκηνή.

Μελετήθηκε ο βαθμός ενασχόλησης ή ενδιαφέροντος ενός χρήστη μπροστά σε μαθησιακό υλικό, μια ιστοσελίδα, ή οποιοδήποτε άλλο είδος ψηφιακού υλικού, με βάση μη προφορική πληροφορία από ακολουθίες βίντεο στις οποίες συμμετείχαν παιδιά. Η κατευθυντικότητα του βλέμματος, προερχόμενη, τόσο από την περιστροφή του κεφαλιού, όσο και από την κατευθυντικότητα των ματιών, αλλά και άλλα χαρακτηριστικά, είναι ζωτικής σημασίας για τέτοιου είδους εφαρμογές, όπως αναλύεται και από τη *Θεωρία*

του Νου των Baron-Cohen [14], όπου η κατευθυντικότητα του βλέμματος είναι άμεσα συσχετισμένη με τα επίπεδα προσήλωσης του χρήστη. Όπως αναλύθηκε και σε προηγούμενο κεφάλαιο, στη βιβλιογραφία δεν υπάρχει πρόταση συγκεκριμένου μοντέλου αναπαράστασης κατάστασης χρήστη σε ό,τι αφορά τα επίπεδα προσοχής σε κάποιο υλικό. Στην παρούσα διατριβή, υπολογίζεται μια σειρά βιομετρικών δεδομένων, με μη παρεμβατικό τρόπο και, ύστερα από κατάλληλη στατιστική επεξεργασία, γίνεται η επιλογή των καταλληλότερων από αυτά. Επιπλέον, προτείνεται η μοντελοποίηση της συμπεριφοράς με χρήση ασαφούς λογικής για τον υπολογισμό των επιπέδων προσήλωσης/νευρικής/νευρικής. Αν και τα δεδομένα παρουσίασαν αρκετές δυσκολίες, κυρίως σε σχέση με τα δεδομένα αναφοράς και την μη προφανή συσχέτισή τους, πάντα, με τα χαρακτηριστικά που εξάγονται, κατέστη δυνατό να διαφανεί το σύνολο εκείνων των μη παρεμβατικών μετρήσεων στην κίνηση του προσώπου που μπορούν να μεταφέρουν πληροφορία σχετικά με την κατάσταση του χρήστη. Το σύστημα που αναπτύχθηκε με βάση τις μετρήσεις που έγιναν στις ακολουθίες παιδιών, ήδη χρησιμοποιείται πιλοτικά σε σχολεία, στα πλαίσια μιας γενικότερης πλατφόρμας που έχει σα σκοπό τη συσχέτιση της απόδοσης του παιδιού με τις μαθησιακές ιδιαιτερότητές του καθώς και να προσφέρει μετρήσιμη ανάδραση της προόδου του.

Κεφάλαιο 6

Χρήση γνώσης προσοχής χρήστη για
εύρεση περιστροφής κεφαλιού.

Επέκταση με χρήση κυλινδρικών
μοντέλων για ανάκτηση βλέμματος
ματιών

Στο κεφάλαιο αυτό ελέγχεται εάν η έννοια της προσοχής, ως γνώση κατάστασης, μπορεί να υποστηρίξει τη βελτίωση της ακρίβειας του μπεϋζιανού δικτύου σύντηξης των δύο επιπέδων πληροφορίας (τοπική και ολιστική) που προτάθηκε στο κεφάλαιο 4. Στο πλαίσιο αυτό, η προσοχή θεωρείται ότι εξαρτάται και από γεωμετρικά χαρακτηριστικά του προσώπου, και μοντελοποιείται η κατανομή επιπέδων προσοχής σε 3 επίπεδα. Το προτεινόμενο σύστημα έχει σα στόχο να λαμβάνει υπόψη του δεδομένα αλληλεπίδρασης, τα οποία έχουν άμεση σχέση με τα αναμενόμενα επίπεδα προσοχής και, με αυτό τον τρόπο, να γίνεται εκτίμηση των πιθανών επιπέδων ενδιαφέροντος από πλευράς χρήστη, με σκοπό τη βελτιστοποίηση στην εκτίμηση περιστροφής κεφαλιού. Η γνώση της περιστροφής κεφαλιού, στα πλαίσια της αλληλεπίδρασης, μπορεί να χρησιμοποιηθεί στην εκτίμηση συγκεκριμένων κινήσεων του κεφαλιού, οι οποίες μπορούν να εμπλουτίσουν την αλληλεπίδραση. Κάνοντας χρήση της σύντηξης ολιστικής με τοπική πληροφορία, ακόμη, σε συνδυασμό με εκτίμηση κατευθυντικότητας βλέμματος ματιών, στο τέλος του κεφαλαίου γίνεται χρήση της νευροασαφούς λογικής που προτάθηκε στα προηγούμενα, με σκοπό την συνολική εκτίμηση του ενδιαφέροντος του χρήστη. Για τα παραπάνω πειράματα, στήθηκε διαδικτυακή μελέτη, στην οποία ζητήθηκε από 189 άτομα να δηλώσουν τα επίπεδα προσοχής τα οποία αποδίδουν στους συμμετέχοντες στη βάση του Πανεπιστημίου της Βοστόνης [22].

6.1 Ερευνητικό πλαίσιο

Στη βιβλιογραφία υπάρχει πληθώρα εργασιών που έχουν ασχοληθεί με το ζήτημα της χρήσης πρότερης γνώσης για βελτιστοποίηση μεθόδων αναγνώρισης. Χαρακτηριστική είναι η εργασία στο [122], όπου η γραμματική του αναγνωριστή ομιλίας αλλάζει δυναμικά, ανάλογα με την προηγούμενη πράξη ή πρόταση του εικονικού πράκτορα. Επίσης, στο [153], οι συγγραφείς χρησιμοποιούν ασαφή λογική για την αξιολόγηση εξόδου νευρωνικού δικτύου, με κανόνες οι οποίοι δημιουργήθηκαν βασισμένοι σε συ-

γκεκριμένη γνώση. Με βάση την έξοδο της ασαφούς μηχανής, προσαρμόζεται το νευρωνικό δίκτυο, οδηγώντας σε βελτιστοποίηση σημασιολογικής κατάταξης εικόνας. Γνώση που προέρχεται από Μπεϋζιανό δίκτυο, το οποίο μοντελοποιεί σχέσεις αντικειμένων-χώρων, αξιοποιείται και στο πρόβλημα αναγνώρισης σκηνής [64]. Τα αντικείμενα αναγνωρίζονται με χρήση περιγραφητών SIFT [96], και η παραγόμενη πληροφορία χρησιμοποιείται στο Μπεϋζιανό δίκτυο. Χαρακτηριστική είναι η δουλειά που περιγράφεται στο [181], όπου, με χρήση καμερών σε "έξυπνα δωμάτια", και, θεωρώντας ως γνώση πλαισίου την ανθρώπινη δραστηριότητα, οι συγγραφείς προχωράνε στην αναγνώριση αντικειμένων. Η μοντελοποίηση της γνώσης για αναγνώριση αντικειμένου γίνεται με χρήση Δικτύων Μαρκοβιανής Λογικής [144].

Στα πλαίσια θεμάτων επικοινωνίας ανθρώπου-υπολογιστή, ωστόσο, οι περισσότερες εργασίες αφορούν σε θέματα, στα οποία γίνεται χρήση γνώσης αλληλεπίδρασης με σκοπό περισσότερο τη *μοντελοποίηση* και λιγότερο την *αναγνώριση* [141],[20], [79]. Χρήση γνώσης προερχόμενης από το περιεχόμενο της αλληλεπίδρασης για βελτιστοποίηση θεμάτων αναγνώρισης έγινε στο [109], όπου οι συγγραφείς μελετούν την αναγνώριση κινήσεων κεφαλιού και, πιο συγκεκριμένα, διακρίνουν θετικά από αρνητικά νεύματα κατά τη διάρκεια αλληλεπίδρασης με εικονικό πράκτορα. Για το σκοπό αυτό, χτίζουν πιθανοτικά μοντέλα πρόβλεψης κίνησης κεφαλιού, βασισμένοι αποκλειστικά σε γνώση σχετικά με χαρακτηριστικά του εικονικού πράκτορα (παύσεις, ακολουθίες λέξεων, κλπ), με τη βοήθεια Μηχανών Υποστήριξης Διανυσμάτων. Με χρήση του συστήματος που περιγράφεται στο [1], γίνεται αναγνώριση κίνησης κεφαλιού αποκλειστικά από οπτικά χαρακτηριστικά. Ακολουθώντας διαδικασία εκ των υστέρων σύντηξης (*late fusion*) των δύο πηγών πληροφορίας (πρόβλεψη κίνησης βάσει διαλογικού περιεχομένου και αναγνωριστή κίνησης κεφαλιού), οι συγγραφείς εκτιμούν την τελική κίνηση του κεφαλιού, με σκοπό τον εμπλουτισμό της αλληλεπίδρασης. Επιπλέον, χρήση περιεχομένου, προερχόμενο από διαφορετικές πληροφορίες (πόζα κεφαλιού, χρώμα ρούχων, τοπολογική σχέση ανάμεσα σε ανθρώπους) χρησιμοποιείται με χρήση Μπεϋζιανών μοντέλων, με στόχο αναγνώριση ταυτότητας, ηλικία, φύλο, κλπ στο [46]. Στο [107] οι συγγραφείς χρησιμοποιούν γνώση από συμμετέχοντες σε αλληλεπιδράσεις μεταξύ περισσότερων των δύο ατόμων (πχ συνεδριάσεις) για να υποστηρίξουν την πρόβλεψη κίνησης κεφαλιού. Κάνοντας χρήση γράφων συμπτώσεων (*co-occurrence graphs*) μεταξύ πληροφοριών πλαισίου (παύσεις, συγκεκριμένες λέξεις, κλπ) και αντίστοιχων κινήσεων στο κεφάλι, οι συγγραφείς δημιουργούν πιθανοτικά μοντέλα για πρόβλεψη κίνησης κεφαλιού. Τα αποτελέσματα αλγορίθμων επεξεργασίας εικόνας συνδέονται σε πραγματικό χρόνο με τις πιθανότητες αυτές, με χρήση του αλγορίθμου LDCRF [111], ως μεταβλητές του ίδιου διανύσματος εισόδου.

6.2 Χρήση γνώσης προσοχής χρήστη για εύρεση περιστροφής κεφαλιού

Στο κεφάλαιο 5 συζητήθηκε η σημαντικότητα της κατευθυντικότητας του κεφαλιού, των ματιών και των εναλλαγών στην απόσταση των δύο ματιών, όπως αυτή προβάλλεται στην εικόνα, για συμπερασματολογία σχετικά με τα επίπεδα προσοχής του χρήστη σε κάποιο στόχο που υποτίθεται ότι βρίσκεται μπροστά του. Στο κεφάλαιο αυτό (εξαιρώντας την επίδραση της κατευθυντικότητας των ματιών) μελετάται η δυνατότητα να διαπιστωθεί εάν η γνώση σχετικά με τα επίπεδα προσοχής του χρήστη μπορεί να συνεισφέρει στον υπολογισμό της γωνίας περιστροφής του κεφαλιού του.

Θεωρώντας ότι ο υπολογισμός των επιπέδων προσοχής έχει να κάνει, τόσο με την περιστροφή του κεφαλιού, αλλά και άλλες μετρικές (εδώ, εναλλαγές στην απόσταση ανάμεσα στα μάτια) έγιναν πειράματα στα οποία, στο μπεϋζιανό δίκτυο σύντηξης ολιστικής και τοπικής πληροφορίας, που περιγράφηκε στο κεφάλαιο 4, θεωρήθηκε ένας ακόμη κόμβος-γονέας της πραγματικής γωνίας περιστροφής, και ο οποίος αντιστοιχεί στα επίπεδα προσοχής του χρήστη στο προηγούμενο καρέ.

6.2.1 Υπολογισμός επιπέδων προσοχής χρήστη

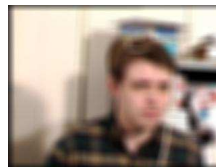
Η επισημείωση που χρησιμοποιήθηκε στο κεφάλαιο 5 αντιστοιχούσε σε σαφείς αποφάσεις σχετικά με το αν τα άτομα ήταν προσεκτικά ή όχι. Εδώ, χρησιμοποιήθηκε επισημείωση που έγινε στη βάση του Πανεπιστημίου της Βοστώνης, και σε κάθε εικόνα αποδόθηκε μία ασαφής τιμή από 0 έως 1, ανάλογα με το βαθμό στον οποίο πολλοί επισημειωτές έκριναν ότι τα άτομα είναι προσεκτικά ή όχι. Με βάση αυτές τις ασαφείς τιμές, πλέον, η εκπαίδευση ενός ασαφούς συστήματος είναι πιο ασφαλής.

6.2.2 Επισημείωση επιπέδων προσοχής

Χρησιμοποιήθηκε το σύνολο ακολουθιών της βάσης του Πανεπιστημίου της Βοστώνης. Καθώς διατίθεται μόνο με επισημείωση σχετικά με την πόζα του κεφαλιού, στήθηκε μία μελέτη στο διαδίκτυο, στην οποία κλήθηκε κόσμος διαφόρων ηλικιών, εθνικοτήτων και απασχόλησης, να επισημειώσει το κατά πόσο θεωρούσε ότι οι απεικονιζόμενοι ήταν προσεκτικοί ή όχι στην κάμερα, με κλίμακα από 0 έως 1. Πιο συγκεκριμένα, για κάθε ακολουθία των 200 καρέ, χρησιμοποιήθηκαν 14 εικόνες, οι οποίες ελήφθησαν ανά τακτά διαστήματα των 15 καρέ (μισό δευτερόλεπτο, καθώς οι ακολουθίες εγγράφηκαν με ρυθμό 30 fps). Έτσι, συνολικά, προέκυψαν 630 εικόνες και στον κάθε επισημειωτή ζητήθηκε να αξιολογήσει έως 60 τυχαίες εικόνες. Οι εικόνες αυτές φορτώθηκαν σε κάποιο διακομιστή, ώστε να υπάρχει ανοιχτή πρόσβαση σε αυτές, και συλλέχθηκαν οι απαντήσεις 87 ατόμων. Τέλος, σε κάθε εικόνα αποδόθηκε ο μέσος όρος των απαντήσεων που δόθηκαν για αυτή. Καθώς οι εικόνες που ελήφθησαν προήλθαν από δειγματοληψία, εκτιμήθηκαν τα επίπεδα προσοχής στα υπόλοιπα καρέ με παρεμβολή. Οι εικόνες που φορτώθηκαν στο διακομιστή πέρασαν από χαμηλοπερατά φίλτρα ώστε να θολώσουν και να μη διακρίνεται η κατευθυντικότητα του βλέμματος. Με αυτόν τον τρόπο, οι επισημειωτές είχαν σα μοναδικό κριτήριο τον προσανατολισμό του κεφαλιού. Χαρακτηριστικά στιγμιότυπα φαίνονται στο σχήμα 6.1, ενώ στο σχήμα 6.2 φαίνεται χαρακτηριστικό στιγμιότυπο της διαδικασίας επισημείωσης.

6.2.3 Εκπαίδευση νευρο-ασαφούς συστήματος για εκτίμηση επιπέδων προσοχής

Χρησιμοποιήθηκε ο ίδιος τύπος συστήματος συμπερασμού, όπως στα προηγούμενα. Για το σχηματισμό των κανόνων, και εδώ χρησιμοποιήθηκε ο αλγόριθμος υποσυστάδας [26] με ακτίνες αναμενόμενης συστάδας 30, 0,15 και 0,3 για τη γωνία του προσώπου, τις μεταβολές στο κλάσμα της απόστασης ανάμεσα στα μάτια (βλ. ενότητα 5.3.3.2) και τα επίπεδα προσοχής του χρήστη. Η εκπαίδευση των παραμέτρων (συναρτήσεις συμμετοχής και τμήμα συμπερασμού των κανόνων) έγινε και εδώ, με



α



β

Σχήμα 6.1: Χαρακτηριστικές εικόνες της βάσης του Πανεπιστημίου της Βοστώνης, στις οποίες οι επισημειωτές θεώρησαν ότι ο χρήστης στην (α) έχει βαθμό προσοχής 7% και στη (β) 10%.

Εικόνα 1/60

ΜΗ ΠΡΟΣΗΛΩΣΗ ΠΡΟΣΗΛΩΣΗ

Επόμενο

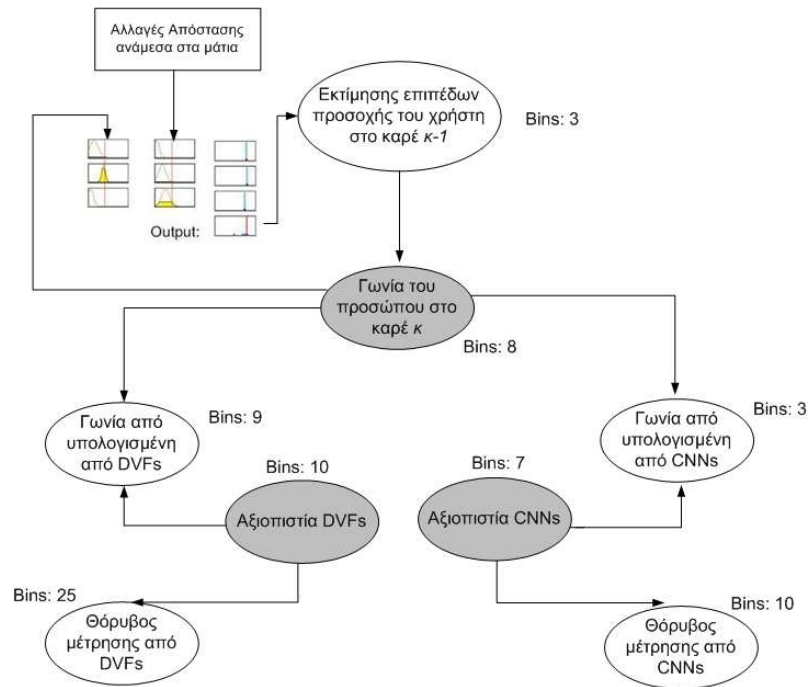
Κοιτάξτε τον άνθρωπο στην εικόνα και κρίνετε το κατά πόσο είναι προσηλωμένος-προσεκτικός στην κάμερα που βρίσκεται τοποθετημένη μπροστά του.

Αξιολογήστε τον βαθμό προσοχής μετακινώντας τη μπάρα ολίσθησης. Η μπάρα αρχικά βρίσκεται τοποθετημένη στη μέση. Όσο πιο δεξιά την τοποθετήσετε, τόσο περισσότερο κρίνετε πως ο άνθρωπος είναι προσεκτικός και, όσο πιο αριστερά, το αντίθετο.

Τέλος, προχωρήστε στην επόμενη εικόνα πατώντας το κουμπί "Επόμενο". **ΠΡΟΣΟΧΗ: Θέλουμε να μελετήσουμε την προσοχή του ατόμου ως προς αυτό που είναι τοποθετημένο μπροστά του και, άρα, δεν ενδιαφερόμαστε για το εάν φαίνεται προσεχτικός σε κάτι το οποίο βρίσκεται σε άλλη κατεύθυνση σε σχέση με την οθόνη.**

Εάν επιθυμείτε μπορείτε να ξαναδιαβάσετε τις αναλυτικές οδηγίες σε νέο παράθυρο κάνοντας κλικ [ΕΔΩ](#).

Σχήμα 6.2: Χαρακτηριστικό στιγμιότυπο διαδικασίας επισημείωσης της βάσης του Πανεπιστημίου Βοστώνης, ως προς τα επίπεδα προσοχής (η επίδραση της κατευθυντικότητας βλέμματος έχει αποκλεισθεί).



Σχήμα 6.3: Μπεϋζιανό Δίκτυο με χρήση πληροφορίας προσοχής χρήστη

τη μέθοδο «Ασαφή Συστήματα Συμπερασμού βασισμένα σε Προσαρμοστικά Δίκτυα» (Adaptive-Network-based Fuzzy Inference Systems - *ANFIS*) [73].

6.2.4 Χρήση νευρο-ασαφούς συστήματος για δυναμική σύντηξη ολιστικής και τοπικής πληροφορίας

Θεωρώντας ως κόμβο-γονέα της πραγματικής γωνίας περιστροφής (βλ. υποενότητα 4.5.5) τα εκτιμώμενα επίπεδα προσοχής στο αμέσως προηγούμενο καρέ, χρησιμοποιείται το μπεϋζιανό δίκτυο του κεφαλαίου 4. Στο σχήμα 6.3 φαίνεται η αρχιτεκτονική που χρησιμοποιήθηκε στην εκπαίδευση του μπεϋζιανού. Η έξοδος του μπεϋζιανού σε ό,τι αφορά την εκτίμηση της πραγματικής περιστροφής του κεφαλιού, καθώς και πληροφορία σχετικά με την απόσταση των ματιών σε σχέση με το πιο πρόσφατο καρέ αρχικοποίησης του συστήματος τροφοδοτούν το εκπαιδευμένο νευροασαφές σύστημα, η έξοδος του οποίου χρησιμοποιείται σαν παρατηρήσιμη μεταβλητή στο μπεϋζιανό του επόμενου καρέ, και επηρεάζει την πραγματική γωνία περιστροφής.

6.2.5 Πειραματικά αποτελέσματα: Υπολογισμός οριζόντιας περιστροφής κεφαλιού

Η χρήση της προτεινόμενης αρχιτεκτονικής για εκτίμηση οριζόντιας περιστροφής κεφαλιού, έχοντας κάνει εκπαίδευση του μπεϋζιανού δικτύου, για κάθε ακολουθία, ακολουθώντας πρότυπο αφαίρεσης στιγμιοτύπου, οδήγησε σε ακρίβεια 4.21° MAE, (5.57° RMS / 4.94° STD).

Η περαιτέρω βελτίωση σε σχέση με προηγούμενα πειράματα στην ίδια βάση (βλ. 4.5.6) επήλθε λαμβάνοντας υπόψη χρονική πληροφορία και, μάλιστα, όταν αυτή έχει να κάνει με τα επίπεδα προσοχής του χρήστη, εξαρτώμενα από την περιστροφή του κεφαλιού και την θέση των ματιών.

Χρησιμοποιώντας ασαφή λογική, η έννοια της προσοχής κατέστη ικανοποιητικά εύρωστη σε επιμέρους σφάλματα της μίας ή της άλλης εισόδου, καθώς η εκπαίδευση του αντίστοιχου ασαφούς συστήματος κανόνων έγινε με μετρούμενα από το σύστημα μεγέθη (οπότε, σφάλματα επίσης μοντελοποιήθηκαν).

Αντικαθιστώντας τον κόμβο που αντιστοιχεί στα επίπεδα προσοχής με την εκτιμώμενη οριζόντια περιστροφή στο προηγούμενο καρέ, το σφάλμα παρουσιάζεται αρκετά αυξημένο (8.08° (rmse) / 6.21° (mae) / 6.28° (std)). Παρατηρήθηκε, λοιπόν, πως μοντελοποίηση με ανάδραση επιπέδων προσοχής από προηγούμενο καρέ, έδωσε καλύτερα αποτελέσματα από ό,τι με χρήση της εκτίμησης περιστροφής από προηγούμενο καρέ. Η πιο πιθανή εξήγηση σε αυτό είναι ότι, στη δεύτερη περίπτωση εισάγεται μεταβλητή που, όταν προέρχεται από λάθος εκτιμήσεις, επηρεάζει το σύστημα με τρόπο τέτοιο που το σφάλμα συσσωρεύεται, αφού δεν πρόκειται για σωστά μοντελοποιημένη πληροφορία.

6.3 Κατάσταση χρήστη με χρήση τοπικής και ολιστικής πληροφορίας για εκτίμηση περιστροφής κεφαλιού, και κυλινδρικών μοντέλων για υπολογισμό κατευθυντικότητας βλέμματος


Με στόχο να αποδοθούν ασαφείς τιμές στα επίπεδα προσοχής που χαρακτηρίζουν τη βάση ακολουθιών του Πανεπιστημίου της Βοστώνης [22], έγινε επισημείωση λαμβάνοντας, όμως, εδώ, υπόψιν, και την επίδραση του βλέμματος των ματιών. Επιπλέον, ως πληροφορία περιστροφής κεφαλιού, χρησιμοποιήθηκε η έξοδος που δίνει η σύντηξη της ολιστικής με τοπική πληροφορία (βλ. 4.4). Η κατευθυντικότητα των ματιών υπολογίζεται με χρήση κυλινδρικών μοντέλων. Η περιοχή της υφής γύρω από τα μάτια μοντελοποιείται από έναν κύλινδρο, ο οποίος περιστρέφεται ώστε να μηδενιστεί η περιστροφή του κεφαλιού (εδώ θεωρήθηκαν μόνο οριζόντιες περιστροφές). Οι νέες θέσεις των ματιών στην καινούρια (μηδενικής περιστροφής κεφαλιού) εικόνα, χρησιμοποιούνται για να εκτιμηθεί η κατευθυντικότητα του βλέμματος. Όπως και στην ενότητα 5.3.5.2, έτσι κι εδώ, χρησιμοποιήθηκαν οι αλλαγές της απόστασης ανάμεσα στα μάτια σε σχέση με το πιο πρόσφατο καρέ αρχικοποίησης και, με χρήση των παραπάνω μεγεθών, εκπαιδεύτηκαν νευροασαφή συστήματα, με σκοπό να ελεγχθεί η δυνατότητα της μεθόδου να προσεγγίσει την ανθρώπινη αντίληψη σχετικά με τα επίπεδα προσοχής κάποιου. Η αρχιτεκτονική που υλοποιήθηκε φαίνεται στο σχήμα 6.5.

6.3.1 Εύρεση κατευθυντικότητας ματιών με χρήση κυλινδρικών μοντέλων

Στο κεφάλαιο 5, η εκτίμηση της κατευθυντικότητας του βλέμματος εξαρτάται σε μεγάλο βαθμό από τη σωστή παρακολούθηση των ορίων του κάθε ματιού, έτσι ώστε να οριοθετείται η περιοχή του και να γίνονται σωστές μετρήσεις αναφορικά με τη θέση της ίριδας μέσα σε αυτό. Επιπλέον, οι μετρήσεις αγνοούσαν την επίδραση της προοπτικής απεικόνισης, θεωρώντας ότι το μήκος του ματιού δε μεταβάλλεται όσο περιστρέφεται το κεφάλι. Κάτι τέτοιο είχε ως συνέπεια, σε περιπτώσεις που το κεφάλι ήταν περιστραμμένο, το προκύπτον ασαφές σύστημα κανόνων, να θεωρεί, σε αρκετές περιπτώσεις, ότι ο χρήστης είναι μη προσεκτικός. Ωστόσο, ενδέχεται μεγάλη

Κεφάλαιο 6. Χρήση γνώσης προσοχής χρήστη για εύρεση περιστροφής κεφαλιού.
Επέκταση με χρήση κυλινδρικών μοντέλων για ανάκτηση βλέμματος ματιών

Εικόνα 21/60



ΜΗ ΠΡΟΣΗΛΩΣΗ ΠΡΟΣΗΛΩΣΗ

Επόμενο

Κοιτάζτε τον άνθρωπο στην εικόνα και κρίνετε το κατά πόσο είναι προσηλωμένος-προσεκτικός στην κάμερα που βρίσκεται τοποθετημένη μπροστά του.

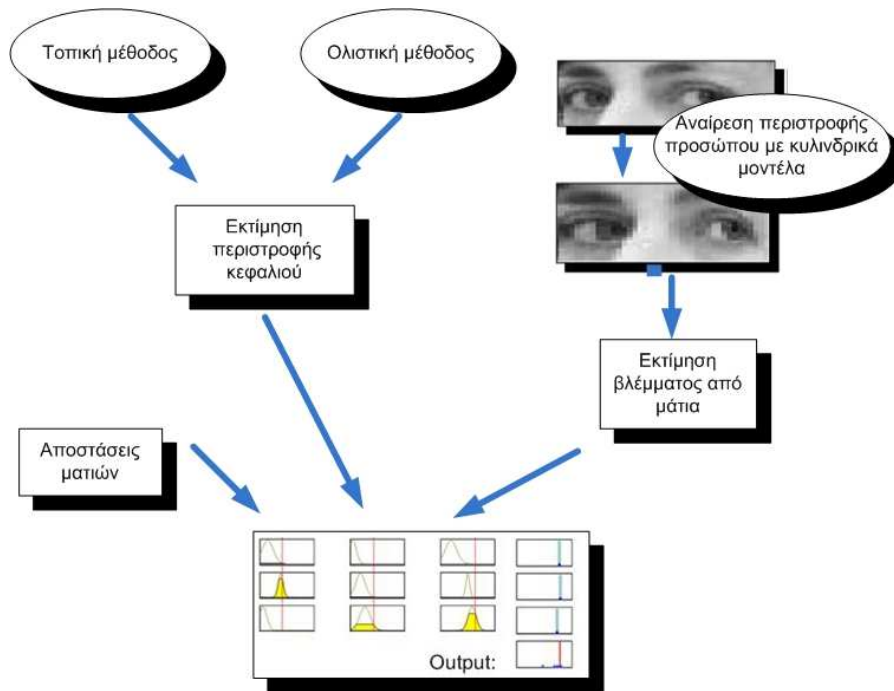
Αξιολογήστε τον βαθμό προσοχής μετακινώντας τη μπάρα ολίσθησης. Η μπάρα αρχικά βρίσκεται τοποθετημένη στη μέση. Όσο πιο δεξιά την τοποθετήσετε, τόσο περισσότερο κρίνετε πως ο άνθρωπος είναι προσεκτικός και, όσο πιο αριστερά, το αντίθετο.

Τέλος, προχωρήστε στην επόμενη εικόνα πατώντας το κουμπί "Επόμενο". **ΠΡΟΣΟΧΗ:** Θέλουμε να μελετήσουμε την προσοχή του ατόμου ως προς αυτό που είναι τοποθετημένο μπροστά του και, άρα, δεν ενδιαφερόμαστε για το εάν φαίνεται προσεκτικός σε κάτι το οποίο βρίσκεται σε άλλη κατεύθυνση σε σχέση με την οθόνη. Με άλλα λόγια, υποθέτουμε ότι στο σημείο της κάμερας υπάρχει μία οθόνη και ενδιαφερόμαστε για το εάν το άτομο την κοιτάζει-την προσέχει.

Εαν επιθυμείτε μπορείτε να ξαναδιαβάσετε τις αναλυτικές οδηγίες σε νέο παράθυρο κάνοντας κλικ [ΕΔΩ](#).

Done

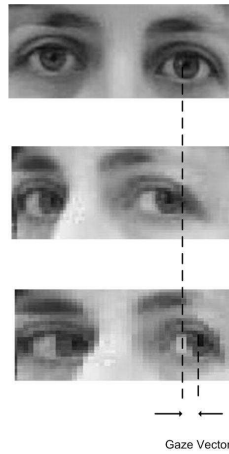
Σχήμα 6.4: Χαρακτηριστικό στιγμιότυπο διαδικασίας επισημείωσης της βάσης του Πανεπιστημίου Βοστώνης, ως προς τα επίπεδα προσοχής (η επίδραση της κατευθυντικότητας βλέμματος δεν έχει αποκλεισθεί).



Σχήμα 6.5: Σύνοψη προτεινόμενης αρχιτεκτονικής με χρήση ολιστικής και τοπικής πληροφορίας για εκτίμηση περιστροφής κεφαλιού και κατευθυντικότητα βλέμματος με χρήση κυλινδρικών μοντέλων.

περιστροφή του κεφαλιού να συνοδεύεται, αντίστοιχα, με μεγάλο σε μέτρο, αλλά αντίθετης κατεύθυνσης, διάνυσμα κατευθυντικότητας βλέμματος, οπότε και η συνολική έξοδος είναι προσεκτικός. Για το λόγο αυτό, στην παρούσα ενότητα, προτείνεται ένα μοντέλο εύρεσης κατευθυντικότητας βλέμματος, με αναίρεση της περιστροφής κεφαλιού. Για την εκτίμηση της κατευθυντικότητας των ματιών, η περιοχή του προσώπου γύρω από τα μάτια (Σχήμα 6.6) μοντελοποιείται από κυλινδρικό σχήμα, με παραμέτρους $p = [\omega_x, \omega_y, \omega_z, t_x, t_y, t_z]$, όπου $\omega_x, \omega_y, \omega_z$ οι γωνίες περιστροφής του κυλίνδρου στους τρεις άξονες και t_x, t_y, t_z οι παράμετροι μετατόπισης. Καθώς σαν εικόνα εισόδου θεωρείται μόνο η περιοχή γύρω από τα μάτια, εδώ θεωρήσαμε μηδενικά t_x και t_y , ενώ το t_z προσεγγίστηκε στα 80cm. Παρομοίως, η ω_x (κατακόρυφη γωνία) θεωρείται μηδενική, και ω_z (γωνία περιστροφής παράλληλη με το επίπεδο της εικόνας) θεωρείται μηδενική, καθώς αυτή μπορεί να εξαλειφθεί, περιστρέφοντας την περιοχή με τέτοιο τρόπο, ώστε τα μάτια να βρίσκονται στο ίδιο οριζόντιο επίπεδο. Η ω_y αντιστοιχεί στην οριζόντια περιστροφή, όπως υπολογίστηκε στην ενότητα 4.4. Επιπλέον, θεωρήθηκε εστιακή απόσταση της κάμερας $f=500$ (σε εικονοστοιχεία), και η διάμετρος κεφαλιού τέθηκε 14 εκατοστά [190].

Ακολούθως, η εικόνα εισόδου "παραμορφώνεται", έτσι ώστε η ω_y του κυλίνδρου να μηδενιστεί (Σχήμα 6.6). Από τις δύο νέες θέσεις των κέντρων των ματιών, επιλέγεται εκείνη που αντιστοιχεί στο μάτι που βρίσκεται πιο κοντά στην κάμερα, καθώς το σφάλμα που αντιστοιχεί σε αυτό, λόγω της προοπτικής απεικόνισης, είναι μικρότερο. Η θέση του στον οριζόντιο άξονα αφαιρείται από εκείνη που είχε στην αρχική μετωπική θέση του ματιού, με σκοπό τον υπολογισμό του διανύσματος κατευθυντικότητας βλέμματος. Το αποτέλεσμα, τέλος, κανονικοποιείται με την απόσταση ανάμεσα στους δύο οφθαλμούς, στο καρέ αρχικοποίησης, με σκοπό να αποφευχθούν προβλήματα που εισάγονται από τις μεταβολές στη μετατόπιση κάθετα στο επίπεδο της κάμερας.



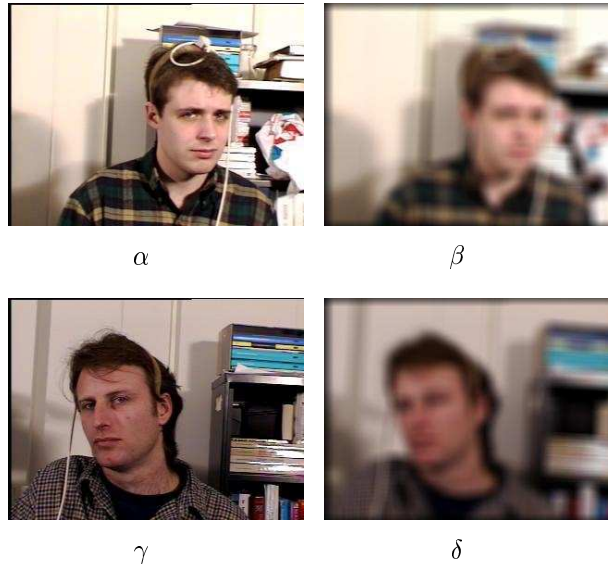
Σχήμα 6.6: Υπολογισμός διανύσματος κατευθυντικότητας βλέμματος ματιών. Οι θέσεις των ματιών στην περιστραμμένη εικόνα συγκρίνονται με αυτές στη μετωπική θέση, αφού εξαλειφθεί η επίδραση της περιστροφής του κεφαλιού.

6.3.2 Επισημείωση επιπέδων προσοχής της Βάσης του Πανεπιστημίου της Βοστώνης, με χρήση κατευθυντικότητας βλέμματος ματιών

Στα προηγούμενα, η βάση του Πανεπιστημίου της Βοστώνης (BU) επισημειώθηκε με τέτοιο τρόπο, ώστε η κρίση των εθελοντών να μην επηρεάζεται από την επίδραση της κατευθυντικότητας των ματιών. Στην παρούσα μελέτη, έχοντας εξαγάγει μετρήσεις σχετικά με το πού κοιτάζουν τα μάτια, η πληροφορία αυτή ήταν απαραίτητη στην επισημείωση των εθελοντών όσον αφορά στην αντίληψή τους σχετικά με τα επίπεδα προσοχής των συμμετεχόντων στη βάση. Συνεπώς, σε μια ξεχωριστή μελέτη, ζητήθηκε από ένα σύνολο διαφορετικών εθελοντών από την προηγούμενη αντίστοιχη μελέτη να επισημειώσουν τη βάση όπως είναι (με την κατευθυντικότητα των ματιών να συμμετέχει στη βάση). Πιο συγκεκριμένα, όπως και προηγουμένως, για κάθε ακολουθία, χρησιμοποιήθηκαν 14 εικόνες, οι οποίες ελήφθησαν με περιοδικότητα 15 καρέ σε κάθε ακολουθία (μισό δευτερόλεπτο, καθώς οι ακολουθίες εγγράφηκαν με ρυθμό 30 fps). Έτσι, συνολικά, προέκυψαν 630 εικόνες. Οι εικόνες αυτές φορτώθηκαν σε κάποιο διακομιστή, ώστε να υπάρχει ανοιχτή πρόσβαση σε αυτές, και συλλέχθηκαν οι απαντήσεις 102 ατόμων, ως προς την αντίληψή τους σχετικά με τα επίπεδα προσοχής των ατόμων που απεικονίζονται στη βάση, με κλίμακα από 0 έως 1 (κάθε εθελοντής μπορούσε να επισημειώσει έως και 60 εικόνες). Τέλος, σε κάθε εικόνα αποδόθηκε ο μέσος όρος των απαντήσεων που δόθηκαν για αυτή. Καθώς οι εικόνες που ελήφθησαν προήλθαν από δειγματοληψία, εκτιμήθηκαν τα επίπεδα προσοχής στα υπόλοιπα καρέ με παρεμβολή. Χαρακτηριστικές εικόνες, τόσο στη φιλτραρισμένη, όσο και στην πρωτότυπη μορφή τους, με τις αντίστοιχες επισημειώσεις που προέκυψαν, φαίνονται στο Σχήμα 6.7.

6.3.3 Στατιστική σημαντικότητα κατευθυντικότητας ματιών και περιστροφής κεφαλιού

Μετά από σύγκριση της επισημείωσης των δύο εκδόσεων της βάσης, ο έλεγχος ANOVA στην απόλυτη διαφορά τους, έδειξε ότι οι πιο σημαντικές διαφοροποιήσεις συμβαίνουν σε εκείνες τις περιπτώσεις όπου ο αλγόριθμος εύρεσης κατευθυντικότητας



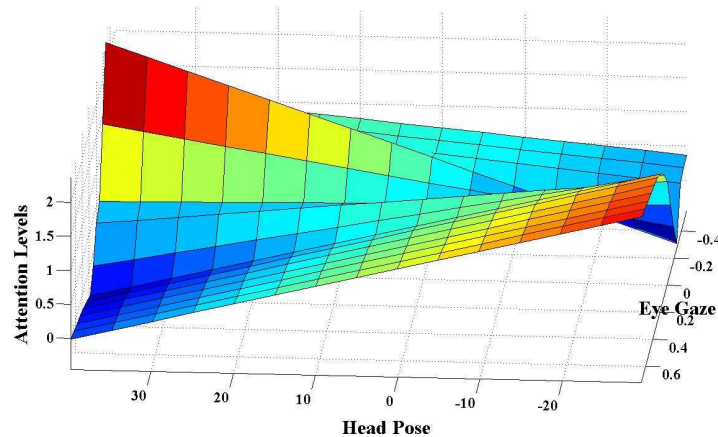
Σχήμα 6.7: Παραδείγματα επισημειωμένων εικόνων, με χρήση, τόσο περιστροφής κεφαλιού, όσο και κατευθυντικότητας ματιών (α,γ), και αποκλειστικά περιστροφής κεφαλιού (β,δ). Τα άτομα στις εικόνες (α) και (γ) χαρακτηρίστηκαν με μέσο βαθμό επιπέδων προσοχής 53% και 60%, αντίστοιχα, ενώ οι αντίστοιχες φιλτραρισμένες εκδόσεις (β) και (δ) με 7% και 10%.

βλέμματος έδινε υψηλές απόλυτες τιμές, ενώ για χαμηλές διαφορές στην επισημείωση, η κατευθυντικότητα βλέμματος έπαιρνε μικρές τιμές ($p < 0.05$).

6.3.4 Πειραματικά αποτελέσματα

Χρησιμοποιώντας την επισημείωση που περιγράφηκε πιο πάνω, εκπαιδεύτηκαν συστήματα ασαφούς συμπερασμού, τύπου Takagi-Sugeno [73], με προεπεξεργασία των δεδομένων σε συστάδες [26] για την εύρεση του βέλτιστου αριθμού κανόνων. Η κατηγοριοποίηση έγινε με ακτίνες 25° για την οριζόντια περιστροφή του κεφαλιού, 0,15 για τα διανύσματα κατευθυντικότητας του βλέμματος, 0,45 για τα κλάσματα μεταβλητότητας των αποστάσεων των ματιών και 0,4 για τα επίπεδα προσοχής (έξοδος). Όπως και στα προηγούμενα, μετά τον καθορισμό της αρχιτεκτονικής του συστήματος, οι παράμετροί του (κέντρα και πλάτη συναρτήσεων συμμετοχής) καθορίστηκαν με χρήση ελαχίστων τετραγώνων και βελτιστοποίηση επικλινούς καθόδου με οπισθοδιάδοση σφάλματος (backpropagation gradient descent, βλ. 5.3.5.1).

Κατά την εκπαίδευση, οι ακολουθίες που αντιστοιχούν στον κάθε συμμετέχοντα στη βάση, ελέγχθηκαν σε συστήματα τα οποία εκπαιδεύτηκαν με χρήση των υπόλοιπων ακολουθιών. Με αυτόν τον τρόπο, προέκυψαν συστήματα συμπερασμού, των οποίων οι κανόνες ήταν από 3 έως 6. Τα πειράματα έδειξαν ότι οι κανόνες που προέκυψαν προσέγγισαν την επισημείωση με ακρίβεια σφάλματος 0.16. Το αντίστοιχο σφάλμα στην επισημείωση της φιλτραρισμένης έκδοσης της βάσης έδωσε σφάλμα 0.18. Κάτι τέτοιο ήταν αναμενόμενο, καθώς, σε αυτήν την περίπτωση, κατά τη διάρκεια της επισημείωσης, δεν λαμβανόταν υπόψη το βλέμμα των ματιών (Σχήμα 6.7), το οποίο, ωστόσο, επιδρούσε στην έξοδο του νευροασαφούς. Επιπλέον, οι έξοδοι των ασαφών συστημάτων θεωρήθηκαν σαν ακόμη ένας επισημειωτής και υπολογίστηκε ο δείκτης Krippendorff's alpha [86]. Πρόκειται για ένα στατιστικό μέτρο το οποίο εκτιμά την αξιοπιστία στη συνέπεια των απαντήσεων ενός συνόλου επισημειωτών, και είναι γε-



Σχήμα 6.8: Έξοδος συστήματος συμπερασμού, εξαρτώμενη από τις τιμές της οριζόντιας περιστροφής και της κατευθυντικότητας βλέμματος.

νίκευση άλλων μετρικών, αφού μπορεί να αξιολογήσει, τόσο κατηγορικά δεδομένα, όσο και αριθμητικά. Επίσης, καθώς μπορεί να διαχειριστεί ελλιπή δεδομένα, κρίθηκε ιδανικό στην περίπτωση μας. Ο κάθε επισημειωτής αξιολόγησε ένα τυχαίο σύνολο εικόνων και, μάλιστα, μερικοί δεν εξάντλησαν τον αριθμό αυτόν, λήγοντας την έρευνα νωρίτερα. Έτσι, ο αριθμός των αξιολογήσεων ποίκιλε από εικόνα σε εικόνα. Οι τιμές που μπορεί να πάρει ο δείκτης Krippendorff's alpha α έχει τις ακόλουθες ερμηνείες:

- $\alpha=1$: τέλεια συμφωνία
- $\alpha=0$: απώλεια αξιοπιστίας. Οι απαντήσεις είναι ασυσχέτιστες
- $\alpha < 0$: Οι ασυμφωνίες είναι συστηματικές και ξεπερνούν το στατιστικά αναμενόμενο.

Στην περίπτωση της επισημείωσης της βάσης, ο συνολικός δείκτης ήταν 0.48, και για τις δύο εκδόσεις της βάσης. Εισάγοντας το αποτέλεσμα του παραπάνω συστήματος, σαν επιπλέον επισημειωτή, η συνολική συμφωνία διαμορφώνεται στο 0.45, θεωρώντας τις επισημειώσεις που λαμβάνουν υπόψιν την κίνηση των ματιών και 0.43 θεωρώντας εκείνες που δεν τις λαμβάνουν.

Στο Σχήμα 6.8 απεικονίζεται μία τυπική τρισδιάστατη επιφάνεια που αντιστοιχεί σε ένα εκπαιδευμένο σύστημα. Φαίνεται ότι, για μικρές τιμές της οριζόντιας περιστροφής, τα επίπεδα προσοχής είναι υψηλότερα όσο τα διανύσματα κατευθυντικότητας βλέμματος παίρνουν χαμηλότερες τιμές (το πρόσωπο είναι στραμμένο μετωπικά ως προς την κάμερα, και ομοίως και τα μάτια). Για υψηλότερες απόλυτες τιμές της οριζόντιας περιστροφής, τα επίπεδα προσοχής τείνουν να είναι υψηλότερα, όσο υψηλότερη είναι η τιμή της κατευθυντικότητας βλέμματος (μη απόλυτες τιμές του αντίθετου προσήμου, σημαίνοντας ότι το άτομο έχει στραμμένο το κεφάλι του προς μια κατεύθυνση, με τα μάτια στραμμένα στην αντίθετη) ¹.

¹Μεγαλύτερος αριθμός δεδομένων εκπαίδευσης θα ήταν αναμενόμενο να περιέχει περισσότερες περιπτώσεις, οδηγώντας σε μια πιο φυσική σχέση ανάμεσα στις έννοιες "περιστροφής κεφαλιού", "κατευθυντικότητα βλέμματος" και "επίπεδα προσοχής". Το γράφημα του Σχήματος 6.8, ωστόσο, είναι ενδεικτικό της αναμενόμενης συμπεριφοράς, ειδικά εάν ληφθεί υπόψη ο μικρός αριθμός των δεδομένων εκπαίδευσης και ο θόρυβος των μετρήσεων

6.4 Συμπεράσματα

Στο κεφάλαιο αυτό εξετάστηκε η δυνατότητα χρήσης γνώσης σχετικά με τα επίπεδα προσοχής χρήστη, με σκοπό τη βελτίωση της εκτίμησης της περιστροφής του προσώπου. Για το σκοπό αυτό στήθηκε διαδικτυακή μελέτη στην οποία επισημειώθηκε η βάση του Πανεπιστημίου της Βοστώνης σχετικά με τα επίπεδα προσοχής που έχουν οι συμμετέχοντές της, και εξετάστηκε εάν η τεχνική προσδιορισμού περιστροφής κεφαλιού που παρουσιάστηκε στο κεφάλαιο 4, μπορεί να βελτιωθεί περαιτέρω, σε μια δυναμική σύντηξη τοπικής και ολιστικής ροής πληροφορίας, με χρήση μπεϋζιανών δικτύων που μοντελοποιούν πληροφορία γνώσης επιπέδων ενδιαφέροντος. Μέσα από τα πειράματα φάνηκε ότι η έννοια της προσοχής του χρήστη (εξαιρώντας την επίδραση του βλέμματος των ματιών) είναι παράγοντας ο οποίος μπορεί να δράσει καταλυτικά στη σωστή εκτίμηση περιστροφής κεφαλιού. Έτσι, προεκτείνοντας αυτή τη δουλειά, γνώση του περιεχομένου αλληλεπίδρασης, ανάλογα με το κατά πόσο αυτό θεωρείται ενδιαφέρον ή όχι, μπορεί να αντικαταστήσει τις ασαφείς εκτιμήσεις ως προς τα επίπεδα προσοχής, και να αυξομειώνει την πιθανότητα η περιστροφή του κεφαλιού να βρίσκεται σε συγκεκριμένα όρια.

Προτάθηκε μία τεχνική προσδιορισμού κατευθυντικότητας βλέμματος με χρήση κυλινδρικών μοντέλων, έχοντας αναιρέσει την επίδραση της περιστροφής του κεφαλιού. Επισημειώνοντας ως προς τα επίπεδα προσοχής των χρηστών, την ίδια βάση με αξιοποίηση της κατευθυντικότητας των ματιών, και χρησιμοποιώντας ασαφή λογική ελέγχθηκε η απόδοση του συστήματος και επιβεβαιώθηκε ότι οι προτεινόμενοι αλγόριθμοι μπορούν να παίξουν σημαντικό ρόλο σε συστήματα αλληλεπιδράσεων. Η σύνθεση πληροφορίας προερχόμενη από κίνηση-περιστροφή κεφαλιού και κατευθυντικότητα βλέμματος, αξιολογήθηκε μέσα από την ικανότητα του ασαφούς συστήματος να πλησιάζει την σχετική επισημειωμένη ανθρώπινη αντίληψη.

Το πλεονέκτημα των συστημάτων που αναλύθηκαν είναι ότι είναι μη παρεμβατικά, λειτουργούν σε πραγματικό χρόνο και, έτσι, είναι σε θέση να υποστηρίζουν εφαρμογές στις οποίες ο αυθορμητισμός είναι μεγάλης σημασίας. Πολύ μεγάλης ιδιαιτερότητας είναι, ακόμη, το γεγονός ότι τα συστήματα είναι σχεδιασμένα με τέτοιο τρόπο, ώστε να λειτουργούν κάτω από σχετικά περίπλοκες συνθήκες σε ό,τι αφορά τον φωτισμό και τη φυσική κίνηση των χρηστών. Κατ' επέκταση, ένα τέτοιο σύστημα μπορεί να προσαρμοστεί/χρησιμοποιηθεί σε μεγάλο αριθμό λοιπών εφαρμογών, καθώς δύναται να προσφέρει μη προφορική ανάδραση, σε πραγματικό χρόνο, με μη παρεμβατικό τρόπο.

Κεφάλαιο 7

Συμπεράσματα και μελλοντικές επεκτάσεις

7.1 Συμπεράσματα

Στη διατριβή αυτή ερευνήθηκαν θέματα αναγνώρισης και μοντελοποίησης προσοχής χρήστη σε περιβάλλοντα αλληλεπίδρασης με πολυμεσικό υλικό, με χρήση απλής, μη βαθμονομημένης κάμερας, πολύπλοκο φόντο, αυθόρμητη κίνηση, και ανεπιτήδευτο φωτισμό. Ερευνήθηκαν και προτάθηκαν αλγόριθμοι που, λαμβάνοντας υπόψιν τους παραπάνω περιορισμούς, είναι μέρος μιας γενικότερης αρχιτεκτονικής τέτοιας που να προσεγγίζει την ανθρώπινη αντίληψη στα επίπεδα προσοχής κάποιου, λαμβάνοντας υπόψη χαρακτηριστικά της υφής και της κίνησης του προσώπου. Οι επιμέρους αλγόριθμοι εξετάστηκαν σε διάφορες βάσεις, στις περισσότερες των οποίων συγκεντρώνονται όλοι οι παραπάνω περιορισμοί. Οι προτεινόμενες αρχιτεκτονικές ενέχουν όλα τα βήματα (από τον εντοπισμό προσώπου έως την τελική μοντελοποίηση των επιπέδων προσήλωσης του χρήστη) και τα αναφερόμενα σφάλματα είναι του συνολικού συστήματος. Προτάθηκαν μοντέλα αντιστοίχισης κίνησης κεφαλιού και βλέμματος, σε καταστάσεις όπως προσήλωση και νευρικότητα.

7.1.1 Αναγνώριση χαρακτηριστικών

Αναπτύχθηκε, αρχικά, μία μέθοδος βασισμένη σε αποτελεσματικές τεχνικές επεξεργασίας εικόνας και βίντεο ώστε να εξαχθεί η περιστροφή του κεφαλιού. Εξ αρχής δόθηκε βαρύτητα σε μία αποτελεσματική, μη πολύπλοκη υπολογιστικά μέθοδο εντοπισμού χαρακτηριστικών σημείων, η οποία θα ήταν κατά το δυνατόν ανεκτική σε διάφορες συνθήκες φωτισμού, και παράγοντες κλιμάκωσης. Αυτά επιτεύχθηκαν με χρήση των Διανυσματικών Πεδίων Απόστασης, με τα οποία γίνεται κωδικοποίηση τοπολογικής πληροφορίας σε όλη την έκταση των περιοχών αναζήτησης, και με τη χρήση ανιχνευτών περιοχής κεφαλιού, έτσι ώστε οι υπολογισμένες περιοχές να έρχονται σε συγκεκριμένες διαστάσεις και να διευκολύνεται ο υπολογισμός των χαρακτηριστικών.

Στη συνέχεια πραγματοποιήθηκε μελέτη αλγορίθμων υπολογισμού περιστροφής κεφαλιού και δόθηκε βαρύτητα σε εκείνες τις μεθόδους που είναι μη παρεμβατικές στον χρήστη και, κατά δεύτερον, δεν απαιτούν βαθμονόμηση και εκπαίδευση για κάθε χρήστη ξεχωριστά. Έχοντας κατά νου το συγκεκριμένο πλαίσιο αλληλεπίδρασης, αναπτύχθηκε η μεθοδολογία υπολογισμού της περιστροφής του κεφαλιού που περιγράφεται στο κεφάλαιο 3. Η μέθοδος έδωσε ικανοποιητικά αποτελέσματα και παρείχε

ικανοποιητικό πλαίσιο ώστε να καθίσταται δυνατή η συμμόρφωση με φυσιολογικές και αυθόρμητες κινήσεις.

Με στόχο την ανάπτυξη ενός συστήματος που θα δίνει μεγαλύτερη ακρίβεια στις εκτιμήσεις, στο κεφάλαιο 4 ερευνήθηκε η ανάπτυξη υβριδικού μοντέλου συνδυασμού τοπικής με ολιστική πληροφορία. Οι υβριδικές μέθοδοι είναι αυτές στις οποίες γίνεται συστηματική προσπάθεια εξάλειψης των μειονεκτημάτων μεθοδολογιών μιας κατηγορίας, με χρήση των πλεονεκτημάτων μεθοδολογιών άλλης κατηγορίας. Στα πλαίσια ανάπτυξης τοπικής τεχνικής αναζητήθηκε μεθοδολογία παρακολούθησης χαρακτηριστικών πιο αποτελεσματική από τη μέθοδο της οπτικής ροής, ενώ, υποστηρικτικά, χρησιμοποιήθηκε και η θέση του στόματος. Επίσης, με στόχο να υπάρχει μεγαλύτερη επιτυχία και αντικειμενικότητα στο διαχωρισμό μεταξύ μεταφορικής και περιστροφικής κίνησης, αναζητήθηκε τρόπος εκτίμησης των γωνιών σχετικά με τα όρια του προσώπου. Επιπλέον, υιοθετήθηκε γκαουσιανό μοντέλο προσομοίωσης του λόγου των αποστάσεων μεταξύ των ματιών και μεταξύ των ματιών και του στόματος, με σκοπό να δοθεί περισσότερη έμφαση σε αναμενόμενες και φυσιολογικές θέσεις των ματιών. Ωστόσο, καθώς η εκτίμηση περιστροφής κεφαλιού με αποκλειστική χρήση τοπικής πληροφορίας εξαρτάται, σε μεγάλο βαθμό, από τη σωστή παρακολούθηση των χαρακτηριστικών, ερευνήθηκε η ανάπτυξη μεθοδολογίας που βασίζεται, τόσο στα Διανυσματικά Πεδία Απόστασης, όσο και σε μέθοδο ολιστικής φύσης. Λαμβάνοντας, όμως, υπόψη τις μη ελεγχόμενες σε φωτισμό και φόντο συνθήκες, η χρήση μιας ολιστικής μεθόδου εισάγει μεγάλο βαθμό περιορισμών: Όπως εξηγήθηκε στην ενότητα 3.1, μέθοδοι εμφάνισης (ολιστικές) απαιτούν πολύ καλή ευθυγράμμιση της περιοχής του προσώπου με τα δεδομένα εκπαίδευσης. Σε περιπτώσεις χαμηλής ανάλυσης, απρόβλεπτου φωτισμού και πολύπλοκου φόντου, η κατάτμηση της περιοχής του προσώπου ενδέχεται να ποικίλλει από καρέ σε καρέ. Για το λόγο αυτό, έγινε χρήση Συνελικτικών Νευρωνικών Δικτύων, τα οποία είναι ανεκτικά σε τέτοιου είδους διαφοροποίηση ανάμεσα στα δεδομένα εκπαίδευσης και ελέγχου. Ερευνήθηκε μία αρχιτεκτονική βασισμένη σε υπο-ταξινομητές γειτονικών κλάσεων. Σε κάθε καρέ ενεργοποιείται μόνο ένα υποσύνολο από αυτούς. Με αυτόν τον τρόπο, μεγιστοποιήθηκε η ακρίβεια και η ταχύτητα. Ο υβριδικός χαρακτήρας του τελικού συστήματος (σύντηξη τοπικού με ολιστικό επίπεδο πληροφορίας) μελετήθηκε με χρήση γραμμικής παρεμβολής και πιθανοτικών δικτύων, με στόχο την αξιοποίηση γνώσης πλαισίου (context) για δυναμική σύντηξη των δύο.

Στα πλαίσια της αναγνώρισης, ερευνήθηκε και το πρόβλημα του υπολογισμού κατευθυντικότητας βλέμματος, με ταυτόχρονη περιστροφή του κεφαλιού, κάτι το οποίο δεν έχει αντιμετωπιστεί, σε ικανοποιητικό βαθμό, από την τρέχουσα βιβλιογραφία, την ώρα συγγραφής της διατριβής, ιδιαίτερα λαμβάνοντας υπόψη τις συνθήκες κάτω από τις οποίες επιδιώκεται να λειτουργήσει η προτεινόμενη αρχιτεκτονική. Αν και η χρήση της σχετικής θέσης του κέντρου του κάθε ματιού σε σχέση με τα άκρα του, έδωσε ικανοποιητικά αποτελέσματα σε προβλήματα μοντελοποίησης, εξακολουθούσε να υφίσταται το πρόβλημα της προσέγγισης λόγω προοπτικής απεικόνισης, καθώς και η ανάγκη ακριβούς παρακολούθησης των γωνιών των ματιών. Για το σκοπό αυτό, έγινε εφαρμογή κυλινδρικών μοντέλων σε περιστραμμένα πρόσωπα, και αναίρεση της περιστροφής με τη βοήθειά τους. Στις εικόνες που προκύπτουν, οι οποίες αντιστοιχούν σε μηδενική περιστροφή κεφαλιού, η καινούρια (εικονική) θέση των ματιών, σε σχέση με καρέ αναφοράς έδωσε πληροφορία κατευθυντικότητας βλέμματος, η οποία αξιολογήθηκε σε προβλήματα μοντελοποίησης.

7.1.2 Μοντελοποίηση επιπέδων προσήλωσης χρήστη

Η εφαρμογή της αλληλεπίδρασης με τον εικονικό πράκτορα, στο κεφάλαιο 5 έδειξε ότι το σύστημα παρακολούθησης κατευθυντικότητας βλέμματος με χρήση προσανατολισμού κεφαλιού έδωσε ικανοποιητικά πορίσματα, τόσο από τεχνικής, όσο και από θεωρητικής πλευράς. Οι χρήστες αρκετά εύκολα μπόρεσαν να χρησιμοποιήσουν ολόκληρο το κεφάλι τους σαν μέθοδο αλληλεπίδρασης με τον εικονικό πράκτορα, ακόμα και αν θα ήταν προφανές ότι η κατευθυντικότητα των ματιών είναι πιο κατάλληλη για τέτοιου είδους εφαρμογές. Οι άνθρωποι είναι αρκετά εξοικειωμένοι με την περιστροφή του κεφαλιού σαν ένδειξη κατευθυντικότητας βλέμματος. Ακόμη, μέσα από ερωτηματολόγια έγινε φανερό η προτίμηση των συμμετεχόντων στα πειράματα εκείνα όπου η μη προφορική ανάδραση από τον πράκτορα συμφωνούσε με το περιεχόμενο της αλληλεπίδρασης. Επίσης, φάνηκε πως η κατευθυντικότητα του βλέμματος των χρηστών επηρεαζόταν σε μεγάλο βαθμό από την αντίστοιχη του εικονικού πράκτορα. Ταυτόχρονα με την παραπάνω εργασία, μελετήθηκε και το πρόβλημα της μοντελοποίησης της συμπεριφοράς παιδιών με μαθησιακές δυσκολίες. Τα δεδομένα από μόνα τους ενείχαν αρκετά μεγάλες δυσκολίες, αφού η επισημείωση δεν συμφωνούσε απαραίτητα με οπτικά χαρακτηριστικά. Για παράδειγμα, σε κάποια σημεία, ήταν επισημειωμένο ότι ένα παιδί ήταν *μη-προσεκτικό* αλλά και μόνο εξαιτίας ενός χτυπήματος του τηλεφώνου. Επιπλέον, πολλές καταστάσεις δεν απέκλειαν τις άλλες (στην επισημείωση αναφερόταν ότι κάποιο παιδί μπορούσε να είναι προσεκτικό αλλά και νευρικό την ίδια στιγμή). Σε τεχνικό επίπεδο, μία δυσκολία που προέκυψε ήταν στην εκτίμηση της κατευθυντικότητας βλέμματος ματιών για μεγάλες περιστροφές του κεφαλιού. Καθώς, επίσης, η εκτίμηση της περιστροφής του κεφαλιού εξαρτώταν από τις θέσεις των κέντρων των ματιών, υπήρχε κάποια συσχέτιση μεταξύ των δύο, η οποία παρατηρήθηκε εξαιτίας του τρόπου υπολογισμού τους. Ωστόσο, η στατιστική επεξεργασία των χαρακτηριστικών, όπως επίσης και τα αποτελέσματα που προέκυψαν από τα νευρο-ασαφή συστήματα, έδειξαν ότι ένα τέτοιο μη παρεμβατικό σύστημα εκτίμησης καταστάσεων χρήστη μπορεί να δώσει, με αρκετά μεγάλη ακρίβεια, ενδείξεις σχετικά με την κατάσταση των παιδιών. Οι αρχιτεκτονικές αυτές χρησιμοποιήθηκαν, ακόμα, για την εκτίμηση των επιπέδων προσοχής με χρήση των τεχνικών που περιγράφηκαν στις ενότητες 4.4 και 6.3.1. Για το σκοπό αυτό επισημειώθηκε η βάση του Πανεπιστημίου της Βοστώνης από ομάδα εθελοντών. Μία ακόμη συμβολή του κεφαλαίου αυτού έγκειται στο ότι για πρώτη φορά έγινε στατιστική μελέτη και επιλογή μετρήσεων στο πρόσωπο, με μη παρεμβατικό τρόπο, για τη μοντελοποίηση κατάστασης χρήστη σε ό,τι αφορά την ενασχόλησή του (*engagement*) σε κάποιο πολυμεσικό υλικό. Επίσης, αναφορικά με το μοντέλο μη παρεμβατικής φύσεως, για εκτίμηση επιπέδων προσήλωσης, στη βιβλιογραφία δεν υπάρχει ένας κοινός, μετρήσιμος τύπος αναφοράς σχετικά με το πώς ορίζεται η προσοχή με μη παρεμβατικά μέσα. Μέσα από στατιστική αξιολόγηση και πρόταση ασαφούς μοντελοποίησης, στη διατριβή αυτή προτείνεται η εγκαθίδρυση ενός σημείου αναφοράς, το οποίο ελπίζεται ότι θα σταθεί αφετηρία για τέτοιου είδους προβλήματα.

Επιπλέον, για την αποτελεσματικότερη σύντηξη της ολιστικής και τοπικής τεχνικής, χρησιμοποιήθηκε πληροφορία γνώσης επιπέδων προσήλωσης, όπως υπολογίστηκε από προηγούμενο καρέ. Για το σκοπό αυτό, χρησιμοποιήθηκε η βάση του Πανεπιστημίου της Βοστώνης, στην οποία, όμως, δεν γινόταν αντιληπτή πληροφορία κατευθυντικότητας ματιών, ώστε να μην επηρεάζεται η κρίση των επισημειωτών. Τα αποτελέσματα που προέκυψαν από αυτήν την μελέτη ήταν τα εξής: Κατ' αρχάς βελτιώθηκε η ακρίβεια υπολογισμού περιστροφής κεφαλιού με χρήση γνώσης επιπέδων προσοχής

χρήστη από προηγούμενο καρέ. Επίσης, φάνηκε ότι υπάρχουν εκείνες οι συνθήκες κάτω από τις οποίες, οι μεν μέθοδοι υπερτερούν έναντι των δε. Κάτι τέτοιο υποστηρίζει το επιχείρημα ότι μία 'έξυπνη' πολυτροπική σύντηξη μπορεί να καταλήξει στο να μειωθεί η επίδραση των μειονεκτημάτων του ενός επιπέδου πληροφορίας, ενισχύοντας δυναμικά το άλλο, χρησιμοποιώντας πληροφορία γνώσης. Από τα παραπάνω προκύπτει το συμπέρασμα πως γνώση πληροφορίας περιεχομένου αλληλεπίδρασης μπορεί να δράσει καταλυτικά στη βελτίωση της αναγνώρισης κίνησης κεφαλιού από την πλευρά του χρήστη, εμπλουτίζοντας και βελτιστοποιώντας, με αυτόν το τρόπο, την εμπειρία της επικοινωνίας.

7.2 Μελλοντικές επεκτάσεις

Η έρευνα που διεξήχθη στα πλαίσια της διατριβής, στο επίπεδο της αναγνώρισης, κινήθηκε με βασικό γνώμονα ανεπιτήδευτα περιβάλλοντα, μη παρεμβατικούς μηχανισμούς και φυσική κίνηση εκ μέρους του χρήστη. Θεωρώντας ότι τέτοια συστήματα είναι ιδανικά για αύξηση του αυθορμητισμού των χρηστών, συστήματα μελέτης συμπεριφοράς σε παρόμοια περιβάλλοντα θα μπορούσαν να επωφεληθούν από την έρευνα αυτή. Για παράδειγμα, ήδη σε ερευνητικό επίπεδο εκδηλώνεται ενδιαφέρον για παρακολούθηση συμπεριφοράς του επισκέπτη σε μουσείο, εγκαθιστώντας μία κάμερα κοντά σε κάποιο έργο, και παρακολουθώντας τα σημεία στα οποία εστιάζει ο θεατής [135]. Χρήση χαρακτηριστικών όπως προσανατολισμός κεφαλιού, ταχύτητα κίνησής του, κλπ, θα ήταν ενδιαφέρον να μελετηθούν σε πραγματικές συνθήκες έκθεσης εμπορευμάτων, ούτως ώστε να μελετηθεί η συμπεριφορά υποψήφιων πελατών: τοπολογικά, σε ποιο σημείο εστιάζεται το ενδιαφέρον και, εάν αυτό σχετίζεται με το προϊόν που εκτίθεται ή έχει να κάνει με αυθόρμητες κινήσεις του βλέμματος του κόσμου μπροστά σε ένα χώρο όπου εκτίθενται εμπορεύματα. Διασύνδεση του συστήματος εκτίμησης βλέμματος ματιών με πολυμεσικά περιβάλλοντα στα οποία εκτίθεται συγκεκριμένο περιεχόμενο (πχ φυλλομετρητές) [23] μπορεί να αποτελέσει ιδανικό εργαλείο επανασχεδιασμού ιστοσελιδών στοχευμένες σε εμπόριο, με σκοπό την προσαρμογή τους στις πραγματικές προτιμήσεις και τη διευκόλυνση των χρηστών.

Όπως αναφέρθηκε στο κεφάλαιο 5, σημαντικό ρόλο στην έννοια της ενασχόλησης παίζει το πλαίσιο (*context*) στο οποίο λαμβάνει χώρα η αλληλεπίδραση με εικονικούς πράκτορες. Επέκταση της μελέτης που έγινε στο αντίστοιχο πείραμα μπορεί να γίνει στα πλαίσια επικοινωνίας ανθρώπου-ρομπότ. Ήδη είναι γνωστό πως οι αντιδράσεις του ανθρώπου, όταν έρχεται σε επικοινωνία με ρομπότ είναι τόσο πιο έντονες, όσο πιο φυσική είναι η συμπεριφορά του ρομπότ. Θεωρώντας σύνολο σεναρίων αλληλεπίδρασης, μετρήσιμα στοιχεία από την ανάλυση προσώπου μπορούν να εισέλθουν στο εννοιολογικό πλαίσιο της αλληλεπίδρασης και να υποστηρίξουν την έννοια της ενασχόλησης λαμβάνοντας υπόψιν τη συμπεριφορά του ρομπότ, το περιεχόμενο της αλληλεπίδρασης, χρονικές παραμέτρους, κλπ [143],[11]. Επέκταση των νευρο-ασαφών μοντέλων που προτάθηκαν στο αντίστοιχο κεφάλαιο, μπορεί να λάβει υπόψιν τους παραπάνω παράγοντες, και να αποδίδεται, πλέον, στην έξοδο, η ενασχόληση αντί της προσοχής.

Τα παραπάνω, πέρα από το εννοιολογικό περιεχόμενο του πλαισίου, είναι θέματα τα οποία έχουν, σε μεγάλο βαθμό, να κάνουν με την έννοια της προσωποποίησης. Τα συστήματα που προτάθηκαν στο κεφάλαιο 6 έχουν εκπαιδευτεί ακολουθώντας πρωτόκολλο αφαίρεσης στιγμιοτύπου, ώστε να ελέγχεται η δυνατότητά τους να εκτιμήσουν τα επίπεδα προσοχής άγνωστων χρηστών. Άλλωστε και η επισημείωση των ίδιων

των δεδομένων έγινε από εθελοντές οι οποίοι δεν είχαν πρότερη γνώση σχετικά με το πώς αντιλαμβάνεται την έννοια της προσοχής ο κάθε συμμετέχων στις βάσεις που εξετάστηκαν. Μελετώντας το πώς εκφράζονται τα διάφορα επίπεδα προσοχής για τον κάθε χρήστη, και συσχετίζοντάς τα με μετρήσιμα μεγέθη του προσώπου, ασαφής λογική μπορεί να χρησιμοποιηθεί προσαρμοστικά για τον κάθε χρήστη, μέσα από ένα θησαυρό γνώσης παραμέτρων των δικτύων.

Βιβλιογραφία

- [1] L.P. Morency N. Checka A. Rahimi and T. Darrell. Adaptive-view-based appearance model. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 375–380. IEEE Computer Society, 2003.
- [2] AgentDysl. *The AgentDysl Project*. <http://www.agent-dysl.eu/>.
- [3] G. Aggarwal, A. Veeraraghavan, and R. Chellappa. 3d facial pose tracking in uncalibrated videos. In *PReMI*, pages 515–520, 2005.
- [4] J. G. Allen, R. Y. D. Xu, and J. S. Jin. Object tracking using camshift algorithm and multiple quantized feature spaces. In *Proceedings of the Pan-Sydney area workshop on Visual information processing*, pages 3–7, Darlinghurst, Australia, 2004. Australian Computer Society, Inc.
- [5] M. Argyle, M. Cook, and M. Argyle. *Gaze and mutual gaze*. Cambridge University Press Cambridge, 1976.
- [6] M. Argyle and J.A. Graham. The central Europe experiment: Looking at persons and looking at objects. *Journal of Non-verbal Behavior*, 1(1):6–16, 1976.
- [7] S. Asteriadis, N. Nikolaidis, and I. Pitas. Facial feature detection using distance vector fields. *Pattern Recognition*, 42(7):1388–1398, 2009.
- [8] S. Asteriadis, N. Nikolaidis, I. Pitas, and M. Pardas. Detection of facial characteristics based on edge information. In *Proceedings of the Second International Conference on Computer Vision Theory and Applications (VISAPP), Barcelona, Spain*, volume 2, pages 247–252, 2007.
- [9] S. Asteriadis, D. Soufleros, K. Karpouzis, and S. Kollias. A natural head pose and eye gaze dataset. Proceedings of the International Workshop on Affective-Aware Virtual Agents and Social Robots, November 2-6, Boston, MA., 2009.
- [10] S. Asteriadis, P. Tzouveli, K. Karpouzis, and S. Kollias. Estimation of behavioral user state based on eye gaze and head pose - application in an e-learning environment. *Multimedia Tools and Applications, Springer*, 41(3):469 – 493, 2009.
- [11] T. Bader and J. Beyerer. Putting gaze into context: A framework for analyzing gaze behavior in interactive and dynamic environments. In *Proceedings of the IUI2010, Workshop on Eye Gaze in Intelligent Human Machine Interaction*, 2010.

-
- [12] D.H. Ballard. Generalizing the Hough transform to detect arbitrary shapes. *Pattern recognition*, 13(2):111–122, 1981.
- [13] Baron-Cohen. How to build a baby that can read minds: Cognitive mechanisms in mindreading. *Cahiers de Psychologic*, 13:513–552, 1994.
- [14] Baron-Cohen. *Mindblindness*. MIT press, 1995.
- [15] E. Bevacqua, M. Mancini, and C. Pelachaud. A listening agent exhibiting variable behaviour. *Lecture Notes in Computer Science*, 5208:262–269, 2008.
- [16] M. Bhuiyan, V. Ampornaramveth, S. Muto, and H. Ueno. Face Detection and Facial Feature Localization for Human-machine Interface. *National Institute of Informatics Journal*, 5:25–39, 2003.
- [17] BioID. *The BioID Face Database*. <http://www.bioid.com/downloads/facedb/facedatabase.html>.
- [18] T. Bosse, Z. A. Memon, and J. Treur. A two-level bdi-agent model for theory of mind and its use in social manipulation. In *AISB 2007 Workshop on Mindful Environments*, pages 335–342, 2007.
- [19] J.Y. Bouguet et al. Pyramidal implementation of the lucas kanade feature tracker description of the algorithm. *Intel Corporation, Microprocessor Research Labs, OpenCV Documents*, 1999.
- [20] B. Brandherm, H. Prendinger, and M. Ishizuka. Interest estimation based on dynamic bayesian networks for visual attentive presentation agents. In *Proceedings of the 9th international conference on Multimodal interfaces*, pages 346–349, New York, NY, USA, 2007. ACM.
- [21] C. Breazeal and B. Scassellati. Challenges in building robots that imitate people. In *Imitation in Animals and Artifacts*, pages 363–390. MIT Press, 2002.
- [22] M. L. Cascia, S. Sclaroff, and V. Athitsos. Fast, reliable head tracking under varying illumination: An approach based on robust registration of texture-mapped 3d models. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22:322–336, 2000.
- [23] S. Castagnos and P. Pu. Consumer decision patterns through eye gaze analysis. In *Proceedings of the IUI2010, Workshop on Eye Gaze in Intelligent Human Machine Interaction*, 2010.
- [24] G. Castellano, A. Pereira, I. Leite, A. Paiva, and P. W. McOwan. Detecting user engagement with a robot companion using task and social interaction-based features. In *Proceedings of the 2009 international conference on Multimodal interfaces*, pages 119–126, New York, NY, USA, 2009. ACM.
- [25] S.C.Y. Chan and P.H. Lewis. A pre-filter enabling fast frontal face detection. *Lecture notes in computer science*, pages 777–784, 1999.

-
- [26] S. L. Chiu. Fuzzy Model Identification Based on Cluster Estimation. *Journal of Intelligent and Fuzzy Systems*, 2(3), 1994.
- [27] H.H. Clark. Using language. *Computational Linguistics*, 23(4), 1996.
- [28] R. G. Congalton and K. Green. *Assessing the accuracy of remotely sensed data: Principles and practices*. Lewis Publishers, Boca Raton, Florida, 1999.
- [29] R. Conte and C. Castelfranchi. *Cognitive and social action*. Garland Science, 1995.
- [30] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active Appearance Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):681–685, 2001.
- [31] T.F. Cootes, K. Walker, and C.J. Taylor. View-based Active Appearance Models. In *Fourth IEEE International Conference on Automatic Face and Gesture Recognition*, pages 227–232, 2000.
- [32] M. Cordea, E. Petriu, and D. Petriu. Three-dimensional head tracking and facial expression recovery using an anthropometric muscle-based active appearance model. *IEEE transactions on instrumentation and measurement*, 57(8):1578–1588, 2008.
- [33] M. Cordea and E.M. Petriu. A 3-D anthropometric-muscle-based active appearance model. *IEEE Transactions on Instrumentation and Measurement*, 55(1):91–98, 2006.
- [34] D. Cristinacce, T. Cootes, and I. Scott. A multi-stage approach to facial feature detection. In *15th British Machine Vision Conference, 231-240*, 2004.
- [35] T. D’Orazio, M. Leo, C. Guaragnella, and A. Distanto. A visual approach for driver inattention detection. *Pattern Recognition*, 40(8):2341–2355, 2007.
- [36] M. Hoffman David, D. B. Grimes, A. P. Shon, and R. P. N. Rao. A probabilistic model of gaze imitation and shared attention. *Neural Networks*, 19(3):299–310, 2006.
- [37] L. Ding and A.M. Martinez. Precise detailed detection of faces and facial features. *Proceedings of the IEEE Computer Vision and Pattern Recognition (CVPR)*. Anchorage, AK, 2008.
- [38] DF Dinges, M. Mallis, G. Maislin, and JW Powell. Evaluation of techniques for ocular measurement as an index of fatigue and the basis for alertness management. US Department of Transportation, National Highway Traffic Safety Administration, Contract No. 1998.
- [39] T. D’Orazio, M. Leo, G. Cicirelli, and A. Distanto. An algorithm for real time eye detection in face images. In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, volume 3, 2004.
- [40] A. Doshi and M.M. Trivedi. Head and Gaze Dynamics in Visual Attention and Context Learning. In *Proceedings of the CVPR Workshop on Visual and Contextual Learning*, 2009.

-
- [41] S. Duncan. Some signals and rules for taking speaking turns in conversations. *Journal of personality and social psychology*, 23(2):283–292, 1972.
- [42] N.J. Emery. The eyes have it: the neuroethology, function and evolution of social gaze. *Neuroscience and biobehavioral reviews*, 24(6):581–604, 2000.
- [43] B. Fasel. Multiscale facial expression recognition using convolutional neural networks, 2002.
- [44] R.S. Feris, J. Gemmell, K. Toyama, and V. Kruger. Hierarchical wavelet networks for facial feature localization. In *ICCV Workshop on Recognition, Analysis and Tracking of Faces and Gestures in Real-Time Systems*, 2001.
- [45] R.A. Fisher. *Statistical Methods for Research Workers*. Hafner Publishing, 14th edition, 1970.
- [46] A. C. Gallagher. *A Framework for Using Context to Understand Images of People*. PhD thesis, Ph.D. Thesis, Department of Electrical and Computer Engineering, Carnegie Mellon University, 2009.
- [47] W. Gao, B. Cao, S. Shan, X. Chen, D. Zhou, X. Zhang, and D. Zhao. The CAS-PEAL large-scale chinese face database and baseline evaluations. *IEEE Transactions on Systems, Man, and Cybernetics, Part A*, 38(1):149–161, 2008.
- [48] P. Grdenfors. *Slicing the Theory of Mind*. Museum Tusulanum Press, 2001.
- [49] A. Gee and R. Cipolla. Determining the gaze of faces in images. *Image and Vision Computing*, 12:639–647, 1994.
- [50] A. Gee and R. Cipolla. Non-intrusive gaze tracking for human-computer interaction. In *International Conference on Mechatronics and Machine Vision in Practice*, pages 112–117, Toowoomba, Australia, 1994.
- [51] N. Gourier, D. Hall, and J. Crowley. Estimating face orientation from robust detection of salient facial features. In *International Workshop on Visual Observation of Deictic Gestures (ICPR)*, Cambridge, UK, 2004.
- [52] N. Gourier, D. Hall, and J.L. Crowley. Facial features detection robust to pose, illumination and identity. In *International Conference on Systems Man and Cybernetics*. Citeseer, 2004.
- [53] N. Gourier and J. Letessier. The pointing 04 data sets. In *Pointing 2004, ICPR, Visual Observation of Deictic Gestures*, Cambridge, UK.
- [54] N. Gourier, J. Maisonnasse, D. Hall, and J. L. Crowley. Head pose estimation on low resolution images. In *CLEAR Workshop, In Conjunction with Face and Gesture*. Springer Verlag, April 2006.
- [55] Z. Gui and C. Zhang. 3d head pose estimation using non-rigid structure-from-motion and point correspondence. In *Proceedings of the IEEE Region 10 Conference TENCN*, pages 1–3, Cambridge, UK, 2006.

-
- [56] G. D. Hager and P. N. Belhumeur. Efficient region tracking with parametric models of geometry and illumination. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(10):1025–1039, 1998.
- [57] C.C. Han, H.Y.M. Liao, G.J. Yu, and L.H. Chen. Fast face detection via morphology-based pre-processing. *Pattern Recognition*, 33(10):1701–1712, 2000.
- [58] R. M. Haralick and L. G. Shapiro. *Computer and Robot Vision*, volume 1. Addison-Wesley, 1992.
- [59] T. Horprasert, Y. Yacoob, and L. S. Davis. Computing 3-d head orientation from a monocular image sequence. pages 242–247, 1996.
- [60] E.J. Horvitz, J. S. Breese, and M. Henrion. Decision theory in expert systems and artificial intelligence. *International Journal of Approximate Reasoning*, 2(3):247–302, 1988.
- [61] C. Huang and A. Darwiche. Inference in belief networks: A procedural guide. *Int. J. Approx. Reasoning*, 15(3):225–263, 1996.
- [62] T. E. Hutchinson. Eye movement detection with improved calibration and speed. *United States Patent*, [19](4,950,069):357–377, 1988.
- [63] T. E. Hutchinson, K. W. Preston, J. R. Worth, N. Martin, C. Kelly, R. Lisa, and A. Frey. Human-computer interaction using eye-gaze input. *IEEE transactions on Systems, Man and Cybernetics*, 19:1527–1533, 1989.
- [64] S.-B. Im and S.-B. Cho. Context-based scene recognition using bayesian networks with scale-invariant feature transform. In *Advanced Concepts for Intelligent Vision Systems*, volume 4179/2006, pages 1080–1087. Springer Berlin / Heidelberg, 2006.
- [65] S. Ioannou, G. Caridakis, K. Karpouzis, and S. Kollias. Robust Feature Detection for Facial Expression Recognition. *EURASIP Journal on Image and Video Processing*, 2007(2), 2007.
- [66] R. Ishii and Y. I. Nakano. Estimating user’s conversational engagement based on gaze behaviors. In *IVA*, pages 200–207, 2008.
- [67] T. Ishikawa, S. Baker, I. Matthews, and T. Kanade. Passive driver gaze tracking with active appearance models. Technical Report CMU-RI-TR-04-08, Robotics Institute, Pittsburgh, PA, February 2004.
- [68] P. Isokoski, M. Joos, O. Spakov, and B. Martin. Gaze controlled games. *Universal Access in the Information Society*, pages 1–15, 2009.
- [69] L. Itti and C. Koch. A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision research*, 40(10-12):1489–1506, 2000.
- [70] A. Jacquin and A. Eleftheriadis. Automatic location tracking of faces and facial features in video sequences. In *Proceedings of the International Workshop on Automatic Face and Gesture Recognition*, pages 142–147. Citeseer, 1995.

- [71] V. Jain and A. Mukherjee. Indian face database. 2002.
- [72] J.-S. Jang and T. Kanade. Robust 3d head tracking by online feature registration. In *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition*, September 2008.
- [73] J.-S. Roger Jang. ANFIS: Adaptive-Network-Based Fuzzy Inference System. *IEEE Transactions on Systems, Man, and Cybernetics*, 23:665–684, 1993.
- [74] F V Jensen. *An Introduction to Bayesian Networks*. Springer-Verlag, 1996.
- [75] O. Jesorsky, K.J. Kirchberg, and R.W. Frischholz. Robust face detection using the hausdorff distance. *Lecture Notes in Computer Science*, pages 90–95, 2001.
- [76] Q. Ji and X. Yang. Real-time eye, gaze, and face pose tracking for monitoring driver vigilance. *Real-Time Imaging*, 8(5):357–377, 2002.
- [77] J. Jovancevic, B. Sullivan, and M. Hayhoe. Control of attention and gaze in complex environments. *Journal of Vision*, 6(12):9, 2006.
- [78] R. El Kaliouby and P. Robinson. Generalization of a vision-based computational model of mind-reading. In *Proceedings of the first International Conference on Affective Computing and Intelligent Interaction, LNCS 3784*, pages 582–589. Springer, 2005.
- [79] A. Kapoor and R. W. Picard. Multimodal affect recognition in learning environments. In *Proceedings of the 13th annual ACM international conference on Multimedia*, pages 677–682, New York, NY, USA, 2005. ACM.
- [80] S. Kawato and N. Tetsutani. Detection and tracking of eyes for gaze-camera control. 22(12):1031–1038, October 2004.
- [81] M. Kipp. Anvil-a generic annotation tool for multimodal dialogue. In *Seventh European Conference on Speech Communication and Technology*. ISCA, 2001.
- [82] R. Kjeldsen and J. Kender. Finding skin in color images. In *Proceedings of the International Conference on Automatic Face and Gesture Recognition*, pages 312–317, 1996.
- [83] L.G. Kourkoutis, K.I. Panoulas, and L.J. Hadjileontiadis. Automated Iris and Gaze Detection Using Chrominance: Application to Human-Computer Interaction Using a Low Resolution Webcam. In *Proceedings of the 19th IEEE International Conference on Tools with Artificial Intelligence*, volume 01, pages 536–539. IEEE Computer Society, 2007.
- [84] J. Kovac, P. Peer, and F. Solina. Human skin colour clustering for face detection. In *IEEE International Conference on Computer as a Tool*, volume 2, 2003.
- [85] T. Kozakaya, T. Shibata, M. Yuasa, and O. Yamaguchi. Facial feature localization using weighted vector concentration approach. *Image and Vision Computing*, 2009.

-
- [86] K Krippendorff. *Content analysis: an introduction to its methodology*. Thousand Oaks, CA: Sage, 2004.
- [87] N. Kruger, M. Potzsch, and C. von der Malsburg. Determination of face position and pose with a learned representation based on labelled graphs. *Image Vision Computing*, 15(8):665–673, 1997.
- [88] J. Ng Sing Kwong and S. Gong. Composite support vector machines for detection of faces across views and pose estimation. *Image and Vision Computing*, 20(5-6):359–368, 2002.
- [89] J.H. Lai, P.C. Yuen, W.S. Chen, S. Lao, and M. Kawade. Robust facial feature point detection under nonlinear illuminations. In *Proceedings of the IEEE ICCV Workshop on Recognition, Analysis and Tracking of Faces and Gestures in Real-time Systems, Canada*, pages 168–174, 2001.
- [90] S. R. Langton, R. J. Watt, and V. Bruce. Do the eyes have it? cues to the direction of social attention. *Trends in Cognitive Neuroscience*, 4(2), 2000.
- [91] S.R.H. Langton and V. Bruce. You must see the point: Automatic processing of cues to the direction of social attention. *Journal of Experimental Psychology Human Perception and Performance*, 26(2):747–757, 2000.
- [92] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel. Handwritten digit recognition with a back-propagation network. In *Advances in Neural Information Processing Systems*, pages 396–404. Morgan Kaufmann, 1990.
- [93] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. In *Proceedings of the IEEE*, volume 86, pages 2278–2324, 1998.
- [94] Y. LeCun, L. Bottou, G. Orr, and K. Muller. Efficient backprop. In G. Orr and K. Muller, editors, *Neural Networks: Tricks of the trade*. Springer, 1998.
- [95] F. Liu, X. Lin, S. Z. Li, and Y. Shi. Multi-modal face tracking using bayesian network. In *Proceedings of IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, 2003.
- [96] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60:91–110, 2004.
- [97] B. Ma, S. Shan, X. Chen, and W. Gao. Head yaw estimation from asymmetry of facial appearance. *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, 38(6):1501–1512, 2008.
- [98] J. J. Magee, M. Betke, J. Gips, M. R. Scott, and B. N. Waber. A human-computer interface using symmetry between eyes to detect gaze direction. *IEEE Transactions on Systems, Man, and Cybernetics, Part A*, 38(6):1–1261, 2008.

-
- [99] Y. Mao, C. Y. Suen, C. Sun, and C. Feng. Pose estimation based on two images from different views. In *Proceedings of the eighth IEEE Workshop on Applications of Computer Vision*, page 9, Washington, DC, USA, 2007. IEEE Computer Society.
- [100] S. C. Marsella and D. V. Pynadath. Psychsim: Agent-based modeling of social interactions and influence. In *Proceedings of the International Conference on Cognitive Modeling*, pages 243–248, 2004.
- [101] Y. Matsumoto, T. Ogasawara, and A. Zelinsky. Behavior recognition based on head pose and gaze direction measurement. In *Proceedings of the IEEE International Conference on Intelligent Robots and Systems*, 2000.
- [102] P. S. Maybeck. *Stochastic models, estimation, and control*. Academic Press, 1979.
- [103] A. McKenzie, S. Lombeyda, and M. Desbrun. Vector field analysis and visualization through variational clustering. In *Eurographics-IEEE VGTC Symposium on Visualization*. Citeseer, 2005.
- [104] K. Messer, J. Kittler, M. Sadeghi, S. Marcel, C. Marcel, S. Bengio, F. Cardinaux, J. Czyz, S. Srisuk, M. Petrou, W. Kurutach, E. Kadyrov, B. Kepenekci, F. B. Tek, G. B. Akar, and F. Deravi. Face verification competition on the xm2vts database. In *Proceedings of the International Conference on Audio and Video Based Biometric Person Authentication*, pages 964–974, 2003.
- [105] M. Morales, P. Mundy, C. delgado, M. Yale, R. Neal, and H. Schwartx. Gaze following, temperament, and language development in 6-month-olds: A replica and extension. *Infant Behavior and Development*, 23(2):231–236, 2000.
- [106] P. Morasso, G. Sandini, V. Tagliasco, and R. Zaccaria. Control strategies in the eye-head coordination system. *IEEE Transactions on Systems, Man, and Cybernetics*, 1977.
- [107] L.-P. Morency. Co-occurrence graphs: Contextual representation for head gesture recognition during multi-party interactions. In *Proceedings of the ICMI workshop on Use of Context in Vision Processing*, 2009.
- [108] L.-P. Morency, A. Rahimi, and T. Darrell. Adaptive view-based appearance model. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 803–810, 2003.
- [109] L.-P. Morency, C. L. Sidner, C. Lee, and T. Darrell. Contextual recognition of head gestures. In *ICMI*, pages 18–24, 2005.
- [110] L.-P. Morency, J. Whitehill, and J. Movellan. Generalized adaptive view-based appearance model: Integrated framework for monocular head pose estimation. In *Proceedings of the IEEE International Conference on Face and Gesture Recognition*, 2008.
- [111] L.P. Morency and T. Darrell. Conditional sequence model for context-based recognition of gaze aversion. In *MLMI'07*, pages 11–23, 2007.

-
- [112] S. Mota and R. W. Picard. Automated posture analysis for detecting learner's interest level. In *Computer Vision and Pattern Recognition workshop on HCI*. IEEE Comput Society, 2003.
- [113] E. Murphy-Chutorian, A. Doshi, and M. M. Trivedi. Head pose estimation for driver assistance systems: A robust algorithm and experimental evaluation. In *Proceedings of the IEEE Conference on Intelligent Transportation Systems*, pages 709–714, 2007.
- [114] E. Murphy-Chutorian and M. M. Trivedi. Head pose estimation in computer vision: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.*, 31(4):607–626, 2009.
- [115] Y.I. Nakano and T. Nishida. Interpreting user's attention as engagement cues in multiparty immersive conversations. 2005.
- [116] V. Navalpakkam and L. Itti. Modeling the influence of task on attention. *Vision Research*, 45(2):205–231, 2005.
- [117] M. H. Nguyen, J. P., and F. De la Torre. Facial feature detection with optimal pixel reduction svm. In *Proceedings of the eighth IEEE International Conference on Automatic Face and Gesture Recognition*, 2008.
- [118] M.H. Nguyen, J. Perez, and F. De la Torre. Facial feature detection with optimal pixel reduction SVMs. In *Proceedings of the International Conference on Automatic Face and Gesture Recognition*, 2008.
- [119] A. Nikolaidis and I. Pitas. Facial feature extraction and determination of pose. *Pattern Recognition*, 33:1783–1791, 2000.
- [120] S. Ohayon and E. Rivlin. Robust 3d head tracking using camera pose estimation. In *Proceedings of the 18th International Conference on Pattern Recognition*, pages 1063–1066. Citeseer.
- [121] Y. Ohmoto, K. Ueda, and T. Ohno. Real-time system for measuring gaze direction and facial features: towards automatic discrimination of lies using diverse nonverbal information. *AI Soc.*, 23(2):187–200, 2008.
- [122] E. Gruenstein Ol. Lemon, L. Cavedon and S. Peters. Collaborative dialogue for controlling autonomous systems. In *In proceedings of AAAI Fall Symposium*, 2002.
- [123] M. Osadchy, Y. LeCun, and M. Miller. Synergistic face detection and pose estimation with energy-based models. *Journal of Machine Learning Research*, 8:1197–1215, May 2007.
- [124] N. Otsu. A threshold selection method from gray-level histograms. *Automatica*, 11:285–296, 1975.
- [125] K. Otsuka, Y. Takemae, and J. Yamato. A probabilistic inference of multiparty-conversation structure based on Markov-switching models of gaze patterns, head directions, and utterances. In *Proceedings of the 7th international conference on Multimodal interfaces*, page 198. ACM, 2005.

-
- [126] J. Pearl. *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann, 1988.
- [127] P. Peer. *CVL Face Database*. Online, 1999.
- [128] J. Pelz, M. Hayhoe, and R. Loeber. The coordination of eye, head, and hand movements in a natural task. *Experimental Brain Research*, 139(3):266–277, 2001.
- [129] V. Perlibakas. Automatical detection of face features and exact face contour. *Pattern Recognition Letters*, 24(16):2977–2985, 2003.
- [130] D.I. Perrett and N.J. Emery. Understanding the intentions of others from visual signals: neurophysiological evidence. *Cahiers de Psychologie Cognitive*, 13(5):683–694, 1994.
- [131] C. Peters. A perceptually-based theory of mind for agent interaction initiation. *International Journal of Humanoid Robotics*, 3(3):321–340, 2006.
- [132] C. Peters, S. Asteriadis, K. Karpouzis, and E. de Sevin. Towards a real-time gaze-based shared attention for a virtual agent. In *Workshop on Affective Interaction in Natural Environments (AFFINE), Crete, Greece*, 2008.
- [133] C. Peters, G. Castellano, and S. de Freitas. An exploration of user engagement in hci. In *AFFINE '09: Proceedings of the International Workshop on Affective-Aware Virtual Agents and Social Robots*, pages 1–3, New York, NY, USA, 2009. ACM.
- [134] R.J. Peters and L. Itti. Beyond bottom-up: Incorporating task-dependent influences into a computational model of spatial attention. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. Citeseer, 2007.
- [135] PIL. *PIL project: Personal Experience with Active Cultural Heritage - Israel*. <http://www.cri.haifa.ac.il/connections/pil/>.
- [136] C. Poelman and T. Kanade. A paraperspective factorization method for shape and motion recovery. Technical Report CMU-CS-93-219, Computer Science Department, Pittsburgh, PA, December 1993.
- [137] I. Poggi. *Mind, hands, face and body: a goal and belief view of multimodal communication*. Weidler, 2007.
- [138] I. Potamitis, N. Fakotakis, and G. Kokkinakis. Traffic safety facts 2000: A compilation of motor vehicle crash data from the fatality analysis reporting system and the general estimates system. In *DOT HS 809 337, U.S. Department of Transportation, National Highway Traffic Safety Administration, National Center for Statistics and Analysis*, 2001.
- [139] H. Prendinger, T. Eichner, E. André, and M. Ishizuka. Gaze-based infotainment agents. In *Proceedings of the international conference on Advances in computer entertainment technology*, page 90. ACM, 2007.

- [140] P. Qvarfordt and S. Zhai. Conversing with the user based on eye-gaze patterns. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, page 230. ACM, 2005.
- [141] R. Ishii and I. Y. Nakano . An empirical study of eye-gaze behaviors: Towards the estimation of conversational engagement in human-agent communication. In *Proceedings of the IUI2010, Workshop on Eye Gaze in Intelligent Human Machine Interaction*, 2010.
- [142] M.J.T. Reinders, RWC Koch, and J.J. Gerbrands. Locating facial features in image sequences using neural networks. In *Proceedings of the second International Conference on Automatic Face and Gesture Recognition*, pages 230–236. Citeseer, 1996.
- [143] C. Rich, B. Ponsleur, A. Holroyd, and C. L. Sidner. Recognizing engagement in human-robot interaction. In *Human-Robot Interaction 2010*, pages 375–382. ACM, 2010.
- [144] M. Richardson and P. Domingos. Markov logic networks. *Machine Learning*, 53(2):169–191, 2003.
- [145] C.A. Rothkopf, D.H. Ballard, and M.M. Hayhoe. Task and context determine where you look. *Journal of Vision*, 7(14):12, 2007.
- [146] B. Scassellati. Mechanisms of shared attention for a humanoid robot. In *in Embodied Cognition and Action: Papers from the 1996 AAAI Fall Symposium*, AAAI. Press, 1996.
- [147] G. A. Seber and C. J. Wild. *Nonlinear Regression*. Wiley, 1989.
- [148] K. Seo, I. Cohen, S. You, and U. Neumann. Face pose estimation system by combining hybrid ica-svm learning and re-registration. In *5th Asian Conference on Computer Vision*, Jeju, Korea, 2004.
- [149] F.Y. Shih and C.F. Chuang. Automatic extraction of head and face boundaries and facial features. *Information Sciences*, 158:117–130, 2004.
- [150] S. Shrestha and J.W. Owens. Eye movement patterns on single and dual-column web pages. *Usability News*, 10, 2008.
- [151] C. Sidner, C. Kidd, C. Lee, and N. Lesh. Where to look: A study of human-robot interaction. In *Intelligent User Interfaces Conference, ACM, New York*, pages 78–84, 2004.
- [152] C.L. Sidner, C. Lee, C.D. Kidd, N. Lesh, and C. Rich. Explorations in engagement for humans and robots. *Artificial Intelligence*, 166(1-2):140–164, 2005.
- [153] N. Simou, T. Athanasiadis, S. Kollias, and G. Stamou. Semantic adaptation of neural network classifiers in image segmentation. 2009.
- [154] P. Smith, M. Shah, and N. da Vitoria Lobo. Determining driver visual attention with one camera. *IEEE transactions on intelligent transportation systems*, 4(4):205–218, 2003.

- [155] O. Špakov and D. Miniotas. Visualization of Eye Gaze Data using Heat Maps. *Electronics and Electrical Engineering*, 74(2):55–58, 2007.
- [156] R. Stiefelhagen. Estimating Head Pose with Neural Networks - Results on the Pointing04 ICPR Workshop Evaluation Data. In *Pointing 04 Workshop (ICPR)*, Cambridge, UK, 2004.
- [157] R. Stiefelhagen, J. Yang, and A. Waibel. Estimating focus of attention based on gaze and sound. In *Proceedings of the 2001 workshop on Perceptive user interfaces*, pages 1–9. ACM New York, NY, USA, 2001.
- [158] J. Sung, T. Kanade, and D. Kim. Pose robust face tracking by combining active appearance models and cylinder head models. *International Journal of Computer Vision*, 80(2):260–274, 2008.
- [159] K. Takahashi, S. Seki, H. Kojima, and R. Oka. Spotting recognition of human gestures from time-varying images. In *Proceedings of the second International Conference On Automatic Face-and Gesture-Recognition*, pages 318–322, 1996.
- [160] K. Tan, D. Kriegman, and N. Ahuja. Appearance-based eye gaze estimation. In *IEEE Workshop on Applications of Computer Vision.*, 2002.
- [161] S. Terence, S. Baker, and M. Bsat. The CMU pose, illumination, and expression (pie) database. In *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition*, May 2002.
- [162] C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: A factorization method. 9(2):137–154, November 1992.
- [163] P. Tzouveli, E. Mitropoulou, K. Ntalianis, S. Kollias, and A. Symvonis. Design of an accommodative intelligent educational environments for dyslexic learners. 11th Conference on Learning Difficulties in the Framework of School Education, November 23-25, 2007 , Athens, Greece, 2007.
- [164] P. Tzouveli, Ph. Mylonas, and S. Kollias. An intelligent e-learning system based on learner profiling and learning resources adaptation. *Computers and Education*, 51(1):224–238, August 2008.
- [165] P. Tzouveli, A. Schmidt, M. Schneider, A. Symvonis, and S. Kollias. Adaptive reading assistance for the inclusion of students with dyslexia: The agent-dysl approach. In *Proceedings of the 2008 Eighth IEEE International Conference on Advanced Learning Technologies*, pages 167–171, Washington, DC, USA, 2008. IEEE Computer Society.
- [166] R. Valenti and T. Gevers. Accurate eye center location and tracking using isophote curvature. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2008.
- [167] R. Valenti, N. Sebe, and T. Gevers. Simple and efficient visual gaze estimation. In *Workshop on Multimodal Interactions Analysis of Users in a Controlled Environment*, 2008.

- [168] R. Valenti, Z. Yucel, and T. Gevers. Robustifying eye center localization by head pose cues. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2009.
- [169] R. Vertegaal. The GAZE groupware system: mediating joint attention in multiparty communication and collaboration. In *Proceedings of the SIGCHI conference on Human factors in computing systems: the CHI is the limit*, pages 294–301. ACM New York, NY, USA, 1999.
- [170] R. Vertegaal, R. Slagter, G. van der Veer, and A. Nijholt. Eye gaze patterns in conversations: there is more to conversational agents than meets the eyes. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, page 308. ACM, 2001.
- [171] P. A. Viola and M. J. Jones. Rapid object detection using a boosted cascade of simple features. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*, volume 1, pages 511–518, 2001.
- [172] M. Voit, K. Nickel, and R. Stiefelhagen. Multi-view head pose estimation using neural networks. In *Proceedings of the second Canadian Conference on Computer and Robot Vision (CRV)*, pages 347–352, Victoria, BC, Canada, 2005. IEEE Computer Society.
- [173] J. G. Wang. Eye gaze estimation from a single image of one eye. In *IEEE International conference on Computer Vision*, pages 136–143, 2003.
- [174] U. Weidenbacher, G. Layher, P. Bayerl, and H. Neumann. Detection of head pose and gaze direction for human-computer interaction. In *Perception and Interactive Technologies*, pages 9–19, 2006.
- [175] U. Weidenbacher, G. Layher, PM Strauss, and H. Neumann. A comprehensive head pose and gaze database. In *3rd IET International Conference on Intelligent Environments*, 2007.
- [176] M. T. Wenzel and W. H. Schiffmann. Head pose estimation of partially occluded faces. In *Proceedings of the second Canadian conference on Computer and Robot Vision*, pages 353–360, Washington, DC, USA, 2005. IEEE Computer Society.
- [177] J. Whitehill and J.R. Movellan. A discriminative approach to frame-by-frame head pose tracking. In *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition*, 2008.
- [178] S. Whittaker. Theories and methods in mediated communication. *The handbook of discourse processes*, pages 243–286, 2003.
- [179] H. Wilson, F. Wilkinson, L. M. Lin, and M. Castillo. Perception of head orientation. *Vision research*, 40(5):459–472, 2000.
- [180] L. Wiskott, J.-M. Fellous, N. Kruger, and C. von der Malsburg. Face recognition by elastic bunch graph matching. *IEEE Transactions Pattern Analysis and Machine Intelligence*, 19:775–779, 1997.

-
- [181] C. Wu and H. Aghajan. Using vcontext with statistical relational models-object recognition from observing user activity in home environment. In *Proceedings of the ICMI workshop on Use of Context in Vision Processing*, 2009.
- [182] J. Wu and M. M. Trivedi. A two-stage head pose estimation framework and evaluation. *Pattern Recognition*, 41(3):1138–1158, 2008.
- [183] J. Wu and Z.H. Zhou. Efficient face candidates selector for face detection. *Pattern Recognition*, 36(5):1175–1186, 2003.
- [184] J. Xiao, S. Baker, I. Matthews, and T. Kanade. Real-time combined 2d+3d active appearance models. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 535 – 542, June 2004.
- [185] J. Xiao and J. F. Cohn. Robust full-motion recovery of head by dynamic templates and re-registration techniques. *International Journal of Imaging Systems and Technology*, 13:85–94, 2003.
- [186] Yale. *Yale Face Database*. <http://cvc.yale.edu/projects/yalefaces/yalefaces.html>.
- [187] H. Yamazoe, A. Utsumi, K. Hosaka, and M. Yachida. A body-mounted camera system for head-pose estimation and user-view image synthesis. *Image and Vision Computing*, 25(12):1848–1855, 2007.
- [188] P. Yao, G. Evans, and A. Calway. Face tracking and pose estimation using affine motion parameters. In *Proceedings of the 12th Scandinavian Conference on Image Analysis*, pages 531–536. Norwegian Society for Image Processing and Pattern Recognition, 2001.
- [189] A. Yilmaz and M. Shah. Automatic feature detection and pose recovery for faces. In *Proceedings of the fifth Asian Conference on Computer Vision*, pages 284–289, 2002.
- [190] J. W. Young. Head and face anthropometry of adult u.s. civilians, 1963-1993, 1993.
- [191] Y. Zhou. Bayesian tangent shape model: Estimating shape and pose parameters via bayesian inference. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 16–22, 2003.
- [192] Z.-H. Zhou and X. Geng. Projection functions for eye detection. *Pattern Recognition*, 37(5):1049–1056, 2004.
- [193] M. Zhu and A.M. Martinez. Subclass discriminant analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(8):1274, 2006.

Κατάλογος δημοσιεύσεων συγγραφέα

Περιοδικά

Δημοσιευμένα

- [J1] S. Asteriadis, P. Tzouveli, K. Karpouzis, and S. Kollias. Estimation of behavioral user state based on eye gaze and head pose - application in an e-learning environment. *Multimedia Tools and Applications, Springer*, 41(3):469 – 493, 2009.
- [J2] C. Peters, S. Asteriadis, and K. Karpouzis. Investigating shared attention with a virtual agent using a gaze-based interface. *Journal on Multimodal User Interfaces, Springer*, 3:119 – 130, 2009.

Υποβεβλημένα

- [S1] S. Asteriadis, K. Karpouzis, and S. Kollias. Head Pose Estimation: A Hybrid Approach in Non-calibrated Environments. Submitted to: *Image and Vision Computing*. Elsevier

Συνέδρια

Δημοσιευμένα

- [C1] S. Asteriadis, K. Karpouzis, and S. Kollias. Face Tracking and Head Pose Estimation using Convolutional Neural Networks. *ACM / SSPNET 2nd International Symposium on Facial Analysis and Animation (FAA 2010)*, October 21, Edinburgh, UK, 2010.
- [C2] G. Caridakis, S. Asteriadis, and K. Karpouzis. User modeling via gesture and head pose expressivity features. *5th International Workshop on Semantic Media Adaptation and Personalization (SMAP 2010)*, December 9-10, Limassol, Cyprus, 2010.
- [C3] S. Asteriadis, K. Karpouzis, and S. Kollias. Head pose estimation with one camera, in uncalibrated environments. *International Workshop on Eye Gaze in Intelligent Human Machine Interaction, IUI*, February 7-10, Hong Kong, 2010.
- [C4] Stylianos Asteriadis, Dimitris Soufleros, Kostas Karpouzis, and Stefanos Kollias. A natural head pose and eye gaze dataset. *Proceedings of the International Workshop on Affective-Aware Virtual Agents and Social Robots*, November 2-6, Boston, MA., 2009.

- [C5] C. Peters, S. Asteriadis, and G. Rebolledo-Mendez. Modelling user attention for human-agent interaction. WIAMIS, May 6-8, London, 2009.
- [C6] S. Asteriadis, K. Karpouzis, and S. Kollias. Feature extraction and selection for inferring user engagement in an hci environment. HCI International, July 19-24, San Diego, CA, 2009.
- [C7] S. Asteriadis, K. Karpouzis, and S. Kollias. A neuro-fuzzy approach to user attention. 18th International Conference on Artificial Neural Networks (ICANN), September 3-6, Prague, 2008.
- [C8] S. Asteriadis, P. Tzouveli, K. Karpouzis, and S. Kollias. A non-intrusive method for user focus of attention estimation in front of a computer monitor. IEEE Conference on Automatic Face and Gesture Recognition (FG 2008), September 17-19, Amsterdam, The Netherlands, 2008.
- [C9] C. Peters, S. Asteriadis, K. Karpouzis, and E. de Sevin. Towards a real-time gaze-based shared attention for a virtual agent. In Workshop on Affective Interaction in Natural Environments (AFFINE), October 24, Crete, Greece, 2008.
- [C10] S. Asteriadis, P. Tzouveli, K. Karpouzis, and S. Kollias. Non-verbal feedback on user interest based on gaze direction and head pose. 2nd International Workshop on Semantic Media Adaptation and Personalization (SMAP 2007), December 17-18, London, United Kingdom, 2007.

□

Παράρτημα Α

Παράμετροι Ασαφών Συστημάτων

Παράμετροι γενικού νευροασαφούς, που χρησιμοποιήθηκε στην ενότητα 5.3.5.2 (η περιστροφή του κεφαλιού και το βλέμμα των ματιών έχουν κανονικοποιηθεί):

Πίνακας Α.1: Ασαφές σύστημα ενότητας 5.3.5.2: Είσοδοι, Κανόνας 1

Είσοδος	Συνάρτηση Συμμετοχής	Κέντρο	Τυπική απόκλιση
Περιστροφή κεφαλιού	γκαουσιανή	0.095	0.072
Βλέμμα ματιών	γκαουσιανή	0.178	0.056
Αλλαγές αποστάσεων ματιών	γκαουσιανή γκαουσιανή	0.833	0.034

Πίνακας Α.2: Ασαφές σύστημα ενότητας 5.3.5.2: Έξοδος, Κανόνας 1

Έξοδος	Συνάρτηση Συμμετοχής	Παράμετροι
1=απρόσεχτος 0=προσεκτικός	γραμμική	-10.99 -4.978 -11.86 10.8

Πίνακας Α.3: Ασαφές σύστημα ενότητας 5.3.5.2: Είσοδοι, Κανόνας 2

Είσοδος	Συνάρτηση Συμμετοχής	Κέντρο	Τυπική απόκλιση
Περιστροφή κεφαλιού	γκαουσιανή	0.118	0.071
Βλέμμα ματιών	γκαουσιανή	0.148	0.077
Αλλαγές αποστάσεων ματιών	γκαουσιανή γκαουσιανή	0.930	0.062

Πίνακας Α.4: Ασαφές σύστημα ενότητας 5.3.5.2: Έξοδος, Κανόνας 2

Έξοδος	Συνάρτηση Συμμετοχής	Παράμετροι
1=απρόσεχτος 0=προσεκτικός	γραμμική	0.8981 3.321 -3.262 2.742

Πίνακας Α.5: Ασαφές σύστημα ενότητας 5.3.5.2: Είσοδοι, Κανόνας 3

Είσοδος	Συνάρτηση Συμμετοχής	Κέντρο	Τυπική απόκλιση
Περιστροφή κεφαλιού	γκαουσιανή	0.184	0.088
Βλέμμα ματιών	γκαουσιανή	0.204	0.051
Αλλαγές αποστάσεων ματιών	γκαουσιανή γκαουσιανή	0.859	0.032

Πίνακας Α.6: Ασαφές σύστημα ενότητας 5.3.5.2: Έξοδος, Κανόνας 3

Έξοδος	Συνάρτηση Συμμετοχής	Παράμετροι
1=απρόσεχτος 0=προσεκτικός	γραμμική	-0.6938 -3.664 -18.32 17.73

Πίνακας Α.7: Ασαφές σύστημα ενότητας 5.3.5.2: Είσοδοι, Κανόνας 4

Είσοδος	Συνάρτηση Συμμετοχής	Κέντρο	Τυπική απόκλιση
Περιστροφή κεφαλιού	γκαουσιανή	0.163	0.103
Βλέμμα ματιών	γκαουσιανή	0.083	0.064
Αλλαγές αποστάσεων ματιών	γκαουσιανή γκαουσιανή	0.785	0.042

Πίνακας Α.8: Ασαφές σύστημα ενότητας 5.3.5.2: Έξοδος, Κανόνας 4

Έξοδος	Συνάρτηση Συμμετοχής	Παράμετροι
1=απρόσεχτος 0=προσεκτικός	γραμμική	0.246 2.207 -1.159 0.6467

Πίνακας Α.9: Ασαφές σύστημα ενότητας 5.3.5.2: Είσοδοι, Κανόνας 5

Είσοδος	Συνάρτηση Συμμετοχής	Κέντρο	Τυπική απόκλιση
Περιστροφή κεφαλιού	γκαουσιανή	0.398	0.043
Βλέμμα ματιών	γκαουσιανή	0.219	0.084
Αλλαγές αποστάσεων ματιών	γκαουσιανή γκαουσιανή	0.906	0.013

Πίνακας Α.10: Ασαφές σύστημα ενότητας 5.3.5.2: Έξοδος, Κανόνας 5

Έξοδος	Συνάρτηση Συμμετοχής	Παράμετροι
1=απρόσεχτος 0=προσεκτικός	γραμμική	1.789 0.9116 13.22 -12.74

Πίνακας Α.11: Ασαφές σύστημα ενότητας 5.3.5.2: Είσοδοι, Κανόνας 6

Είσοδος	Συνάρτηση Συμμετοχής	Κέντρο	Τυπική απόκλιση
Περιστροφή κεφαλιού	γκαουσιανή	0.082	0.068
Βλέμμα ματιών	γκαουσιανή	0.299	0.095
Αλλαγές αποστάσεων ματιών	γκαουσιανή	0.937	0.028

Πίνακας Α.12: Ασαφές σύστημα ενότητας 5.3.5.2: Έξοδος, Κανόνας 6

Έξοδος	Συνάρτηση Συμμετοχής	Παράμετροι
1=απρόσεχτος 0=προσεκτικός	γραμμική	-1.974 2.387 -1.362 0.6025

Παράρτημα Α. Παράμετροι Ασαφών Συστημάτων

Παράμετροι γενικού νευροασαφούς, που χρησιμοποιήθηκε στην ενότητα 5.3.5.3 (η περιστροφή του κεφαλιού και οι ταχύτητες έχουν κανονικοποιηθεί):

Πίνακας Α.13: Ασαφές σύστημα ενότητας 5.3.5.3: Είσοδοι, Κανόνας 1

Είσοδος	Συνάρτηση Συμμετοχής	Κέντρο	Τυπική απόκλιση
Περιστροφή κεφαλιού	γκαουσιανή	0.152	0.068
Οριζ. Ταχύτητα	γκαουσιανή	0.174	0.098
Κατακόρυφη Ταχύτητα	γκαουσιανή	0.112	0.084

Πίνακας Α.14: Ασαφές σύστημα ενότητας 5.3.5.3: Έξοδος, Κανόνας 1

Έξοδος	Συνάρτηση Συμμετοχής	Παράμετροι
1=νευρική 0=ηρεμία	γραμμική	6.884 -7.234 1.81 -0.5438

Πίνακας Α.15: Ασαφές σύστημα ενότητας 5.3.5.3: Είσοδοι, Κανόνας 2

Είσοδος	Συνάρτηση Συμμετοχής	Κέντρο	Τυπική απόκλιση
Περιστροφή κεφαλιού	γκαουσιανή	0.118	0.088
Οριζ. Ταχύτητα	γκαουσιανή	0.280	0.108
Κατακόρυφη Ταχύτητα	γκαουσιανή	0.101	0.081

Πίνακας Α.16: Ασαφές σύστημα ενότητας 5.3.5.3: Έξοδος, Κανόνας 2

Έξοδος	Συνάρτηση Συμμετοχής	Παράμετροι
1=νευρική 0=ηρεμία	γραμμική	-0.0186 0.0009246 0.112 0.951

Πίνακας Α.17: Ασαφές σύστημα ενότητας 5.3.5.3: Είσοδοι, Κανόνας 3

Είσοδος	Συνάρτηση Συμμετοχής	Κέντρο	Τυπική απόκλιση
Περιστροφή κεφαλιού	γκαουσιανή	0.211	0.049
Οριζ. Ταχύτητα	γκαουσιανή	0.129	0.116
Κατακόρυφη Ταχύτητα	γκαουσιανή	0.197	0.058

Πίνακας Α.18: Ασαφές σύστημα ενότητας 5.3.5.3: Έξοδος, Κανόνας 3

Έξοδος	Συνάρτηση Συμμετοχής	Παράμετροι
1=νευρικότητα 0=ηρεμία	γραμμική	-1.131 4.363 7.967 -2.249

Παράμετροι γενικού νευροασαφούς, που χρησιμοποιήθηκε στην ενότητα 6.3.4:

Πίνακας Α.19: Ασαφές σύστημα ενότητας 6.3.4: Είσοδοι, Κανόνας 1

Είσοδος	Συνάρτηση Συμμετοχής	Κέντρο	Τυπική απόκλιση
Περιστροφή κεφαλιού	γκαουσιανή	1.837	36.79
Βλέμμα ματιών	γκαουσιανή	0.037	0.119
Αλλαγές αποστάσεων ματιών	γκαουσιανή γκαουσιανή	0.918	0.256

Πίνακας Α.20: Ασαφές σύστημα ενότητας 6.3.4: Έξοδος, Κανόνας 1

Έξοδος	Συνάρτηση Συμμετοχής	Παράμετροι
1=απρόσεχτος 0=προσεκτικός	γραμμική	0.007365 -8.642 0.5466 5.448

Πίνακας Α.21: Ασαφές σύστημα ενότητας 6.3.4: Είσοδοι, Κανόνας 2

Είσοδος	Συνάρτηση Συμμετοχής	Κέντρο	Τυπική απόκλιση
Περιστροφή κεφαλιού	γκαουσιανή	0.016	36.79
Βλέμμα ματιών	γκαουσιανή	0.003	0.102
Αλλαγές αποστάσεων ματιών	γκαουσιανή γκαουσιανή	0.995	0.281

Πίνακας Α.22: Ασαφές σύστημα ενότητας 6.3.4: Έξοδος, Κανόνας 2

Έξοδος	Συνάρτηση Συμμετοχής	Παράμετροι
1=απρόσεχτος 0=προσεκτικός	γραμμική	0.01263 -50.9 3.863 -4.993

Πίνακας Α.23: Ασαφές σύστημα ενότητας 6.3.4: Είσοδοι, Κανόνας 3

Είσοδος	Συνάρτηση Συμμετοχής	Κέντρο	Τυπική απόκλιση
Περιστροφή κεφαλιού	γκαουσιανή	6.239	36.79
Βλέμμα ματιών	γκαουσιανή	0.135	0.074
Αλλαγές αποστάσεων ματιών	γκαουσιανή γκαουσιανή	0.894	0.327

Πίνακας Α.24: Ασαφές σύστημα ενότητας 6.3.4: Έξοδος, Κανόνας 3

Έξοδος	Συνάρτηση Συμμετοχής	Παράμετροι
1=απρόσεχτος 0=προσεκτικός	γραμμική	-0.03247 -9.237 0.8349 2.181

Πίνακας Α.25: Ασαφές σύστημα ενότητας 6.3.4: Είσοδοι, Κανόνας 4

Είσοδος	Συνάρτηση Συμμετοχής	Κέντρο	Τυπική απόκλιση
Περιστροφή κεφαλιού	γκαουσιανή	-1.37	36.79
Βλέμμα ματιών	γκαουσιανή	-0.094	0.091
Αλλαγές αποστάσεων ματιών	γκαουσιανή γκαουσιανή	0.962	0.282

Πίνακας Α.26: Ασαφές σύστημα ενότητας 6.3.4: Έξοδος, Κανόνας 4

Έξοδος	Συνάρτηση Συμμετοχής	Παράμετροι
1=απρόσεχτος 0=προσεκτικός	γραμμική	-0.02031 -8.906 -1.614 -4.176