



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΣΗΜΑΤΩΝ, ΕΛΕΓΧΟΥ ΚΑΙ ΡΟΜΠΟΤΙΚΗΣ

Ανίχνευση Προσώπου και Εκτίμηση της Πόζας
με Εφαρμογή στην Αυτόματη Ανάλυση
Νοηματικής Γλώσσας

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΤΟΥ

ΣΤΡΑΤΗ Β. ΤΖΟΥΜΑ

Επιβλέπων: Πέτρος Μαραγκός
Καθηγητής Ε.Μ.Π.

ΕΡΓΑΣΤΗΡΙΟ ΟΡΑΣΗΣ ΥΠΟΛΟΓΙΣΤΩΝ, ΕΠΙΚΟΙΝΩΝΙΑΣ ΛΟΓΟΥ ΚΑΙ ΕΠΕΞΕΡΓΑΣΙΑΣ ΣΗΜΑΤΩΝ
Αθήνα, Ιούλιος 2011



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

Ανίχνευση Προσώπου και Εκτίμηση της Πόζας με Εφαρμογή στην Αυτόματη Ανάλυση Νοηματικής Γλώσσας

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

ΣΤΡΑΤΗ Β. ΤΖΟΥΜΑ

Επιβλέπων: Πέτρος Μαραγκός
Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 22η Ιουλίου 2011.

(Υπογραφή)

(Υπογραφή)

(Υπογραφή)

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

.....
Κων/νος Τζαφέστας
Επικ. Καθ. Ε.Μ.Π.

.....
Δρ. Γεράσιμος Ποταμιάνος
Δ/ντής Ερευνών Δημόκριτος.

Αθήνα, Ιούλιος 2011

(Υπογραφή)

.....
ΤΖΟΥΜΑΣ ΣΤΡΑΤΗΣ

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

© 2011 – All rights reserved



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

Copyright ©–All rights reserved Τζούμας Στρατής, 2011.

Με επιφύλαξη παντός δικαιώματος.

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.

Ευχαριστίες

Αυτή η διπλωματική εργασία πέρασε από πολλά στάδια και από πολλές χώρες πριν φτάσει στην τελική της μορφή, και αντιπροσωπεύει ένα ταξίδι ενός χρόνου, στον κόσμο και στην έρευνα. Τμήματα της εργασίας αναπτύχθηκαν, παράλληλα με άλλες ερευνητικές δραστηριότητες, σε τρεις διαφορετικές Ηπείρους, ξεκινώντας από το Ισραήλ, με ένα ενδιάμεσο διάλειμμα στην Ελλάδα, και καταλήγοντας στις ΗΠΑ. Σε αυτό το σημείο θα ήθελα να ευχαριστήσω κάποια άτομα που συνέβαλαν καταλυτικά τόσο στην ανάπτυξη αυτής της εργασίας όσο και στην απόφασή μου να συνεχίσω την πορεία μου στον χώρο της έρευνας.

Αρχικά ευχαριστώ τον καθηγητή μου Πέτρο Μαραγκό που με τις διαλέξεις του με ενέπνευσε να ακολουθήσω τον γοητευτικό τομέα της Όρασης Υπολογιστών, και μου έδωσε τις πρώτες σημαντικές ερευνητικές κατευθύνσεις. Επίσης ευχαριστώ τους ερευνητές Βασίλη, Σταύρο και Ισίδωρο για την καθοδήγηση που μου έδωσαν σε αρκετά σημεία της διπλωματικής μου.

Ευχαριστώ τους Yair Moshe και Nimrod Peleg για την συνεργασία που είχαμε κατά την πρακτική μου στο Technion Israel Institute of Technology, καθώς και τον Peng Wang για την συνεργασία μας κατά την πρακτική μου στην Siemens Corporate Research, και για την στήριξη και τις συμβουλές του όσον αφορά το επόμενο βήμα της επαγγελματικής μου πορείας.

Τέλος θα ήθελα να ευχαριστήσω τους φίλους μου Ιωάννα, Παναγιώτη και Κώστα για την βοήθεια και στήριξή τους κατά την τελευταία πιεστική περίοδο.

Περίληψη

Ο στόχος της διπλωματικής αυτής εργασίας είναι η ανάπτυξη μιας εύρωστης μεθόδου για την ανίχνευση του προσώπου και την εκτίμηση της πόζας του ομιλητή κατά την Νοηματική Γλώσσα. Η θέση του προσώπου και η πόζα του ομιλητή είναι απαραίτητα για την περαιτέρω ανάλυση της έκφρασης του, ενώ κρύβουν και αυτά γλωσσική πληροφορία. Οι ιδιαίτερες προκλήσεις που αντιμετωπίζονται σε αυτό το πρόβλημα είναι η συχνή και εκτεταμένη επικάλυψη του προσώπου από τα χέρια, και η σχετικά χαμηλή ανάλυση της εικόνας στην περιορισμένη περιοχή του προσώπου. Για τις ανάγκες της εφαρμογής, σύγχρονες μέθοδοι μελετώνται και προσαρμόζονται στις ιδιαιτερότητες του συγκεκριμένου προβλήματος. Η εργασία διαιρείται σε τρία κεφάλαια:

Στο πρώτο κεφάλαιο μελετάται η πλέον σύγχρονη μέθοδος ανίχνευσης προσώπων, υλοποιείται και προσαρμόζεται στο πρόβλημα. Το δεύτερο κεφάλαιο εισάγει ένα στοχαστικό πλαίσιο συνεχούς ανίχνευσης του προσώπου που βασίζεται στο φίλτρο Κάλμαν. Στο τρίτο κεφάλαιο παρουσιάζονται οι κυριότερες μέθοδοι για την εκτίμηση της πόζας, με ιδιαίτερη έμφαση στους ευέλικτους γράφους προσώπου. Τελικά αναπτύσσεται μια πρωτότυπη μέθοδος για την εκτίμηση της πόζας, η οποία εκμεταλλεύεται την ανίχνευση του στενού πλαισίου του προσώπου, και αποδεικνύεται ιδιαίτερα εύρωστη σε επικαλύψεις.

Λέξεις Κλειδιά

Ανίχνευση προσώπου, εκτίμηση της πόζας, νοηματική γλώσσα, ανίχνευση υπο επικάλυψη, όραση υπολογιστών, εκμάθηση μηχανής

Abstract

The goal of this thesis is to develop a robust method for face tracking and pose estimation for the problem of Automatic Sign Language Analysis. The position and pose of signer's head are necessary for the further investigation of his facial expression but they also hide linguistic information themselves. The particular challenges faced, are the extended and frequent occlusion by the hands and the relatively low resolution of the face region. Current state-of-the-art methods are used and adapted to fit the specificities of this problem. The thesis is divided into three chapters:

In the first chapter, the method of face detection based on boosted Haar features is presented and theoretically analyzed. The second chapter introduces a tracking framework for the face, based on the Kalman filter. In the third chapter, the most important methods for pose estimation are presented and evaluated, with special focus on the ones based on Active Appearance Models. A novel method is finally developed that estimates the head pose from the relative position of the head centroid to the face center. This method exploits the good face tracking results, and proves to be much more resilient to occlusion.

Keywords

Face tracking, pose estimation, sign language, tracking under occlusion, computer vision, machine learning

Περιεχόμενα

Ευχαριστίες	1
Περίληψη	3
Abstract	5
Περιεχόμενα	9
Κατάλογος Σχημάτων	12
1 Εισαγωγή	13
1.1 Αντικείμενο της διπλωματικής	13
1.2 Συνεισφορά	14
1.3 Οργάνωση του τόμου	15
2 Ανίχνευση Προσώπων	17
2.1 Εισαγωγή	17
2.2 Ανίχνευση Προσώπων με την Βοήθεια του AdaBoost και των Haar Χαρακτηριστικών	18
2.2.1 Haar-like χαρακτηριστικά	18
2.2.2 Ο αλγόριθμος AdaBoost	20
2.2.3 Η συνεισφορά των P. Viola και M. Jones	24
2.3 Υλοποίηση και Προσαρμογή στο Συγκεκριμένο Πρόβλημα	26
2.3.1 Υλοποίηση	26
2.3.2 Προσαρμογή στο συγκεκριμένο πρόβλημα	29
2.4 Κατασκευή και Πειραματισμός με την Βάση Δεδομένων Εκπαίδευσης	30
2.4.1 Εισαγωγή	30
2.4.2 Πειράματα με την βάση δεδομένων εκπαίδευσης	31
2.4.3 Συμπεράσματα	34
2.5 Αξιολόγηση της Μεθόδου και Εξελίξεις στην Ανίχνευση Προσώπων	36
2.5.1 AdaTree	37
2.5.2 Probabilistic boosting tree	38

3	Συνεχής Ανίχνευση σε Ακολουθίες Βίντεο της Νοηματικής	41
3.1	Εισαγωγή	41
3.2	Το πρόβλημα Observer Design	42
3.3	Γενικευμένη Λύση	42
3.4	Επιλογή του Φίλτρου Kalman	43
3.4.1	Υπόθεση	43
3.4.2	Προσομοίωση	43
3.4.3	Συμπεράσματα και επιπλέον παραδοχές	44
3.5	Σχεδιασμός και Υλοποίησης ενός Φίλτρου Kalman	46
3.5.1	Φίλτρο Kalman	46
3.5.2	Σχεδιαστικές αποφάσεις για την υλοποίηση του φίλτρου Kalman	47
3.5.3	Κινητικά μοντέλα	48
3.6	Απόδοση Εμπιστοσύνης στην Ανίχνευση	51
3.7	Συμπεράσματα	52
3.7.1	Ποσοστά ανίχνευσης	53
3.7.2	Ταχύτητα ανίχνευσης	53
3.7.3	Εμπιστοσύνη και απόφαση σχετικά με την ποιότητα της ανίχνευσης	54
4	Εκτίμηση της Πόζας	55
4.1	Εισαγωγή	55
4.2	Διαφορετικές Προσεγγίσεις για την Εκτίμηση της Πόζας	56
4.2.1	Κατηγοριοποίηση με βάση την εμφάνιση	56
4.2.2	Εφαρμογή ευέλικτων μοντέλων γράφων προσώπου	58
4.2.3	Γεωμετρικές μέθοδοι	62
4.3	Ανίχνευση της Πόζας μέσω της Συνεχούς Ανίχνευσης του Στενού Μοτίβου του Προσώπου	66
4.3.1	Εισαγωγή	66
4.3.2	Ψυχοφυσική και γεωμετρική ερμηνεία της ιδέας	66
4.3.3	Περιγραφή της μεθόδου	68
4.3.4	Αποτελέσματα και προτεινόμενες επεκτάσεις	70
5	Παράρτημα	73
5.1	Στατιστική Ερμηνεία του Αλγορίθμου AdaBoost	73
5.1.1	Προσθετικό Μοντέλο	73
5.1.2	Λογιστική Παλινδρόμηση	73
5.1.3	Συμπεράσματα	74
5.2	Εκτίμηση των γωνιών Euler από τρισδιάστατα σημεία του χώρου (landmarks)	74
5.3	Ανάλυση σε Πρωτεύουσες Συνιστώσες	75
5.4	Active Shape και Active Appearance Models	76
5.4.1	Active Shape Models	76
5.4.2	Independent Active Appearance Models	77

5.4.3 Combined Active Appearance Models	78
6 Επίλογος	79
6.1 Σύνοψη και Συνεισφορά	79
6.2 Κατευθύνσεις για Μελλοντική Έρευνα	80
Βιβλιογραφία	82
7 Μεταφράσεις Ξένων όρων	87

Κατάλογος Σχημάτων

1.1	Παράδειγμα ανίχνευσης προσώπου και εκτίμησης της πόζας σε εικόνες της νοηματικής γλώσσας	13
2.1	Τα 4 βασικά είδη των Haar-like χαρακτηριστικών [26]	20
2.2	Ένα επεκτεταμένο σύνολο από Haar Features [16] [25]	20
2.3	Οπτικοποίηση του υπολογισμού αθροίσματος τετραγώνου μέσω της ολοκληρωμένης εικόνας	24
2.4	Αποτελέσματα ανίχνευσης προσώπου στην ομάδα SIPL (1)	29
2.5	Αποτελέσματα ανίχνευσης προσώπου στην ομάδα SIPL (2)	29
2.6	Μεταβάλλοντας την περιοχή εστίασης του προσώπου (4 διαφορετικά μοντέλα)	32
2.7	Ποσοστά Ανίχνευσης για τα 4 διαφορετικά μοντέλα εστίασης του προσώπου	33
2.8	Ποσοστά Ανίχνευσης με διαφορετικά ποσοστά προσώπων προφίλ στην βάση εκπαίδευσης	34
2.9	Ποσοστά ανίχνευσης μετά την προσθήκη χειρομορφών στην κλάση μη-προσώπων	35
2.10	Περιπτώσεις σωστών και λάθος ανιχνεύσεων σε μεμονωμένες εικόνες της νοηματικής	37
2.11	Οπτικοποίηση της διαδικασίας κατηγοριοποίησης με την βοήθεια του αλγορίθμου Probabilistic Boosting Tree [24]	38
3.1	Περιοχή προσομοίωσης	44
3.2	Οπτική αναπαράσταση της διδιάστατης ανίχνευσης, και μοντελοποίηση με γκαουσιανή	45
3.3	Πειραματισμός με το Μοντέλο 1 της βάσης εκπαίδευσης	45
3.4	Ποσοστά ανίχνευσης υπο αυξανόμενες προσημειωμένες ταχύτητες στο ακίνητο μοντέλο, στα δύο benchmarks	49
3.5	Ποσοστά ανίχνευσης υπο αυξανόμενες προσομοιωμένες ταχύτητες στο απλό κινητικό μοντέλο, στα δύο benchmarks	50
3.6	Ποσοστά ανίχνευσης υπο αυξανόμενες προσομοιωμένες ταχύτητες στο σύνθετο κινητικό μοντέλο, στα δύο benchmarks	51
3.7	Παρουσίαση των τυπικών αποκλίσεων σε περιπτώσεις σωστών και λάθος ανιχνεύσεων	52

3.8	Παρουσίαση των τιμών ανίχνευσης σε περιπτώσεις σωστών και λάθος ανιχνεύσεων	53
3.9	Αποτελέσματα συνεχούς ανίχνευσης υπό μεταβαλλόμενη πόζα και εκτενή επί- κάλυψη	54
4.1	Βαθμοί Ελευθερίας του Ανθρώπινου Προσώπου [8].	56
4.2	Εκτίμηση της πόζας με διακριτή κατηγοριοποίηση με βάση την εμφάνιση [8].	57
4.3	Κύριες συνιστώσες του σχήματος του γράφου προσώπου ASM [14].	59
4.4	Αποτελέσματα εκτίμησης της πόζας με την μέθοδο View-Based AAMs [4].	60
4.5	Μεταβάλλοντας την πρώτη κύρια συνιστώσα του σχήματος από $-3\sqrt{\lambda}$ το $3\sqrt{\lambda}$	62
4.6	Μεταβάλλοντας την δεύτερη κύρια συνιστώσα του σχήματος από $-3\sqrt{\lambda}$ το $3\sqrt{\lambda}$	63
4.7	Μεταβάλλοντας την τρίτη και τέταρτη κύρια συνιστώσα του σχήματος	63
4.8	Κύριες συνιστώσες μεταβολής του σχήματος [17].	64
4.9	Άξονες στρέψης Yaw, Pitch και Roll	65
4.10	Οπτική αναπαράσταση των διανυσμάτων v , v_1 και v_2 πάνω στο ανθρώπινο πρόσωπο.	67
4.11	Οπτικοποίηση προοπτικής και ορθογραφικής προβολής διανύσματος.	68
4.12	Σημεία ενδιαφέροντος πάνω στο πρόσωπο.	69
4.13	Κέντρο προσώπου και κεντροειδές κεφαλιού υπο μεταβολή της πόζας $[-45^\circ, 45^\circ]$	70
4.14	Συναρτήσεις αντιστοίχισης των κανονικοποιημένων διανυσμάτων V_{Yaw} και V_{Pitch} στις γωνίες Yaw και Pitch	71
4.15	Yaw = 26 , Pitch = -23	72
4.16	Yaw = 20 , Pitch = -10	72
4.17	Yaw = 26 , Pitch = 15	72
4.18	Yaw = 22 , Pitch = 9	72
4.19	Yaw = -17 , Pitch = -9	72
4.20	Yaw = 14 , Pitch = -45	72
4.21	Αποτελέσματα ανίχνευσης της πόζας (σε μοίρες) σε ακολουθίες εικόνων της νοηματικής	72

Κεφάλαιο 1

Εισαγωγή

1.1 Αντικείμενο της διπλωματικής



Σχήμα 1.1: Παράδειγμα ανίχνευσης προσώπου και εκτίμησης της πόζας σε εικόνες της νοηματικής γλώσσας

Αντικείμενο της διπλωματικής αυτής εργασίας είναι η μελέτη, αξιολόγηση, προσαρμογή και περαιτέρω ανάπτυξη μεθόδων και τεχνικών Όρασης Υπολογιστών (Computer Vision) και Αναγνώρισης Προτύπων (Pattern Classification) για την επιτυχή ανίχνευση του προσώπου και εκτίμηση της πόζας, με εφαρμογή στο πρόβλημα της αυτόματης ανάλυσης της νοηματικής γλώσσας.

Στην νοηματική γλώσσα, το πρόσωπο αποτελεί έναν ιδιαίτερα σημαντικό παράγοντα επικοινωνίας και μετάδοσης πληροφορίας. Κατά την ανάλυση της νοηματικής γλώσσας, το πρόσωπο αποτελεί το κέντρο βάσει του οποίου υπολογίζεται η σχετική θέση των χεριών, και καταγράφεται και μοντελοποιείται η κίνησή τους [6]. Γι αυτό τον λόγο η σωστή και ακριβής ανίχνευση του προσώπου είναι απαραίτητη για κάθε περαιτέρω μελέτη των χειρονομιών του νοηματιστή. Επίσης η ακριβής ανίχνευση του προσώπου αποτελεί το πρώτο βήμα για την περαιτέρω ανάλυση της έκφρασης του προσώπου, για την εξαγωγή του γλωσσικού νόηματος που αυτή κρύβει.

Όσον αφορά την πόζα του κεφαλιού του νοηματιστή, αυτή κρύβει από μόνη της σημαντικό γλωσσικό νόημα [27]. Για παράδειγμα ερωτήσεις, επιβεβαιώσεις, αρνήσεις και υποθετικές

προτάσεις επικοινωνούνται κυρίως βάσει της πόζας του κεφαλιού. Επίσης η πόζα του κεφαλιού συνήθως χρησιμοποιείται για την αναφορά σε χρονικές περιόδους. Πιο συγκεκριμένα η αναφορά σε σύντομα χρονικά διαστήματα συνοδεύεται συνήθως από μικρές μεταβολές της πόζας, ενώ αναφορές σε μεγάλα χρονικά διαστήματα γίνονται από την εκτεταμένη μεταβολή της πόζας σε κατεύθυνση αντίθετη από αυτή της χειρονομίας.

Οι ιδιαίτερες προκλήσεις που αντιμετωπίζονται κατά την ανίχνευση του προσώπου και την εκτίμηση της πόζας στην περίπτωση της νοηματικής γλώσσας είναι η σχετικά χαμηλή ανάλυση στην περιοχή του προσώπου, και η συχνή και εκτεταμένη επικάλυψη του προσώπου από τα χέρια. Η εργασία αυτή μελετά τα επιμέρους προβλήματα της Όρασης Υπολογιστών στην γενική τους μορφή, ενώ τελικά προσφέρει λύσεις προσαρμοσμένες στις ιδιαίτερες προκλήσεις της συγκεκριμένης εφαρμογής.

1.2 Συνεισφορά

Η συνεισφορά της εργασίας είναι η κατασκευή ενός πλήρους ενοποιημένου πλαισίου για την ανίχνευση προσώπου και εκτίμηση της πόζας, που προσανατολίζεται και εφαρμόζεται στην επι μέρους περίπτωση της ανάλυσης του προσώπου κατά την νοηματική γλώσσα. Κατά την κατασκευή αυτού του πλαισίου αναλύονται, αξιολογούνται και προσαρμόζονται σύγχρονοι αλγόριθμοι και μαθηματικά μοντέλα, ενώ επίσης αναπτύσσεται μια πρωτότυπη μέθοδος για την εκτίμηση της πόζας, κατάλληλη για το συγκεκριμένο πρόβλημα. Το πλαίσιο που αναπτύσσεται αξιολογείται πάνω σε ακολουθίες εικόνων της νοηματικής γλώσσας και υπόσχεται ιδιαίτερα εύρωστα αποτελέσματα. Πιο συγκεκριμένα οι συνεισφορές της διπλωματικής αυτής εργασίας συνοψίζονται σε τρία βασικά σημεία:

- Αρχικά γίνεται αναλυτική μελέτη, περιγραφή και υλοποίηση μιας εκ των σημαντικότερων, βασισμένων-σε-εκπαίδευση, μεθόδων της Όρασης Υπολογιστών για την ανίχνευση προσώπων και γενικότερα αντικειμένων. Για την εφαρμογή της μεθόδου κατασκευάζεται εκ του μηδενός μια βάση δεδομένων επαιδευσεις, πάνω στην οποία γίνονται πειραματισμοί που προσφέρουν ιδιαίτερα ενδιαφέροντα συμπεράσματα.
- Σε επόμενο στάδιο η μέθοδος αυτή συνδυάζεται με ένα μαθηματικό μοντέλο στοχαστικής εκτίμησης, με σκοπό την εκμετάλλευση πρότερης χρονικής πληροφορίας για συνεχή ανίχνευση σε ακολουθίες βίντεο. Αρχικά δικαιολογείται η επιλογή του φίλτρου Kalman ως του καταλληλότερου για το συγκεκριμένο πρόβλημα, και στην συνέχεια λαμβάνονται αρκετές σχεδιαστικές αποφάσεις μέσω πειραματισμών για την επιτυχία και την ευστάθεια του συστήματος. Επίσης οι στοχαστικές παράμετροι που προκύπτουν από αυτή την ανάλυση χρησιμοποιούνται για την απόδοση εμπιστοσύνης την ανίχνευση.
- Τέλος γίνεται μια βιβλιογραφική ανασκόπηση κάποιων εκ των κυριότερων μεθόδων που χρησιμοποιούνται σήμερα για το γενικότερο πρόβλημα εκτίμησης της πόζας, ενώ προτείνεται και αναπτύσσεται μια καινοτόμος μέθοδος η οποία εκμεταλλεύεται τα χαρακτηριστικά του συγκεκριμένου προβλήματος. Η μέθοδος αυτή βασίζεται στην εύρωστη

ανίχνευση δύο μόνο καθολικών σημείων του ανθρώπινου κεφαλιού, και την μετάφρασή τους σε γωνίες στρέψης, και για αυτό τον λόγο ξεπερνά τις προκλήσεις που αντιμετωπίζονται στην νοηματική γλώσσα κυρίως λόγω της επικάλυψης του προσώπου από τα χέρια.

1.3 Οργάνωση του τόμου

Ο τόμος οργανώνεται σε τρία κύρια κεφάλαια, και ένα παράρτημα ως εξής:

Στο Κεφάλαιο 2 παρουσιάζεται το γενικευμένο θεωρητικό πλαίσιο για την ανίχνευση προσώπων, που επιλέχτηκε ως το πιο κατάλληλο, και υλοποιήθηκε για την επίλυση του παραπάνω προβλήματος. Το κεφάλαιο ξεκινά περιγράφοντας την πορεία εξέλιξης του πλαισίου στην τελική του μορφή, ενώ παράλληλα γίνεται θεωρητική ανάλυση των βασικών χαρακτηριστικών του, και του αλγόριθμου που χρησιμοποιεί. Έπειτα γίνεται μια σύντομη περιγραφή της υλοποίησης και της προσαρμογής του πλαισίου στο συγκεκριμένο πρόβλημα, και παρουσιάζονται κάποια πειράματα που εκτελέστηκαν, καθώς και τα συμπεράσματα που προέκυψαν. Τέλος γίνεται συζήτηση σχετικά με την επιτυχία του παραπάνω πλαισίου καθώς και τους περιορισμούς του, και περιγράφονται συνοπτικά κάποιες ενδιαφέρουσες εξελίξεις στην ανίχνευση προσώπου που χρησιμοποιούνται σήμερα.

Το Κεφάλαιο 3 περιγράφει την διαδικασία που ακολουθήθηκε για την ανάπτυξη ενός στοχαστικού πλαισίου συνεχούς ανίχνευσης (tracking) του προσώπου πάνω σε ακολουθίες εικόνων της νοηματικής γλώσσας. Αρχικά παρέχονται στοιχεία που δικαιολογούν την επιλογή του μαθηματικού μοντέλου (Kalman Filter) ως του καταλληλότερου για αυτό τον σκοπό, ενώ στην συνέχεια περιγράφεται αναλυτικά ο σχεδιασμός του, ο συνδυασμός του με το υπάρχον σύστημα ανίχνευσης του Κεφαλαίου 2, και η εφαρμογή του στο συγκεκριμένο πρόβλημα. Τέλος η μέθοδος αξιολογείται πάνω σε ακολουθίες εικόνων της νοηματικής γλώσσας.

Το Κεφάλαιο 4 απευθύνεται στο πρόβλημα της εκτίμησης της πόζας του προσώπου του νοηματιστή. Αρχικά γίνεται μια βιβλιογραφική ανάλυση των διαφορετικών προσεγγίσεων και μεθόδων που έχουν χρησιμοποιηθεί επιτυχώς για το γενικότερο πρόβλημα εκτίμησης της πόζας, και γίνεται συζήτηση σχετικά με την καταλληλότητα της κάθε μεθόδου για την συγκεκριμένη εφαρμογή. Στην συνέχεια προτείνεται και υλοποιείται μια απλή, πρωτότυπη μέθοδος που εκμεταλλεύεται την επιτυχημένη συνεχή ανίχνευση του στενού μοτίβου του προσώπου και αποδεικνύεται ιδιαίτερα εύρωστη στις προκλήσεις που παρουσιάζει το συγκεκριμένο πρόβλημα.

Τέλος ακολουθεί ένα παράρτημα όπου εξηγούνται κάποιοι αλγόριθμοι και μαθηματικές μέθοδοι που χρησιμοποιούνται ή συζητώνται στην εργασία.

Κεφάλαιο 2

Ανίχνευση Προσώπων

2.1 Εισαγωγή

Το πρόβλημα της ανίχνευσης προσώπων αποτελεί μια από τις σημαντικότερες προκλήσεις στην όραση υπολογιστών, έχει μελετηθεί από πολλούς ερευνητές και έχει προσεγγιστεί από πολλές διαφορετικές οπτικές γωνίες. Η ιδιαίτερη σημασία του δεν έγκειται μόνο στις αναρίθμητες εμπορικές του εφαρμογές, αλλά το γεγονός ότι αποτελεί μια χαρακτηριστική εξειδίκευση του γενικότερου προβλήματος ανίχνευσης αντικειμένων σε εικόνες. Τα πρόσωπα είναι μια κλάση σύνθετων αντικειμένων που παρουσιάζουν μικρές διαφοροποιήσεις όσον αφορά το σχήμα τους, αλλά μπορούν να παρουσιάσουν μεγάλες διαφοροποιήσεις όσον αφορά την εμφάνιση τους (χρώμα, υφή), από άνθρωπο σε άνθρωπο, ή λόγω μεταβολής της πόζας.

Η ανίχνευση προσώπων για αρκετό καιρό προσεγγίζονταν από την σκοπιά της ανίχνευσης του χρώματος του δέρματος (δεν γενικεύεται σε άλλα αντικείμενα), ή από την σκοπιά του ταιριάσματος προτύπων. Η δυσκολία που παρουσιάζει αυτό το πρόβλημα έγκειται τόσο στην κατηγοριοποίηση (classification) όσο και στην μοντελοποίηση. Τα πρόσωπα είναι σύνθετα αντικείμενα, που είναι αρκετά δύσκολο να μοντελοποιηθούν ώστε να μειωθεί η διάστασή τους και να μπορεί να γίνει χρήση κάποιας από τις κλασικές μεθόδους κατηγοριοποίησης. Επίσης το πρόβλημα της κατηγοριοποίησης μεταξύ των διακριτών κλάσεων πρόσωπο και μη-πρόσωπο, παρουσιάζει δυσκολίες καθώς η κλάση μη-πρόσωπο είναι άπειρη και μη αυστηρώς ορισμένη, ενώ η κλάση πρόσωπο μπορεί να παρουσιάζει μεγάλη μεταβλητότητα.

Μετά την καταλυτική συμβολή των Viola & Jones [26] το 2001, το πρόβλημα της ανίχνευσης προσώπων θεωρείται πλέον λυμένο, με πολύ μεγάλα ποσοστά επιτυχίας (ανεξαρτήτως της πόζας) και μικρό υπολογιστικό κόστος που επιτρέπει real-time εφαρμογές. Το θεωρητικό πλαίσιο που αναπτύχθηκε για αυτό τον σκοπό εφαρμόστηκε επιτυχώς και σε πολλά άλλα προβλήματα ανίχνευσης αντικειμένων, σε κάθε πεδίο εφαρμογής της όρασης υπολογιστών, όπως πραγματικές, δισδιάστατες και τρισδιάστατες ιατρικές εικόνες, εικόνες δορυφόρου κ.α..

2.2 Ανίχνευση Προσώπων με την Βοήθεια του AdaBoost και των Haar Χαρακτηριστικών

Το 1998 οι Papageorgiou & Poggio πρότειναν μια νέα μέθοδο μοντελοποίησης του προσώπου [20] που βασίζεται στην χρήση Haar χαρακτηριστικών, η οποία αποδείχτηκε ιδιαίτερα επιτυχημένη τόσο για το συγκεκριμένο πρόβλημα όσο και για την γενικότερη ανίχνευση κλάσεων αντικειμένων που παρουσιάζουν ομοιότητες με τα πρόσωπα. Οι Papageorgiou et al, επιλέγουν για την μοντελοποίηση του προσώπου έναν πεπερασμένο αριθμό Haar χαρακτηριστικών (βλ. ενότητα 2.2.1), από μια υπερπλήρη βάση μέσω μιας στατιστικής διαδικασίας, και μετά εκπαιδεύουν μια συνάρτηση κατηγοριοποίησης μέσω του αλγορίθμου Support Vector Machine (SVM) από τις τιμές αυτών των χαρακτηριστικών πάνω σε εικόνες προσώπων και σε εικόνες μη-προσώπων.

Αυτή η ιδέα σχετικά με την μοντελοποίηση υιοθετήθηκε αργότερα (2001) από τους Viola & Jones [26], και συνδυάστηκε με τον αλγόριθμο AdaBoost (βλ. ενότητα 2.2) για την ανάπτυξη ενός νέου ιδιαίτερα επιτυχημένου γενικευμένου πλαισίου ανίχνευσης προσώπων και αντικειμένων. Οι Viola & Jones δεν επιλέγουν τα Haar χαρακτηριστικά με κάποια στατιστική διαδικασία, αλλά αυτά επιλέγονται αυτόματα από τον αλγόριθμο AdaBoost κατά την εκπαίδευση της συνάρτησης κατηγοριοποίησης, ενοποιώντας έτσι την διαδικασία της μοντελοποίησης με την διαδικασία της εκπαίδευσης. Τα Haar χαρακτηριστικά και ο αλγόριθμος AdaBoost με μικρές διαφοροποιήσεις, ανάλογα με την περίπτωση, χρησιμοποιούνται ακόμη και σήμερα στις μεγαλύτερες εμπορικές εφαρμογές αναγνώρισης αντικειμένων.

2.2.1 Haar-like χαρακτηριστικά

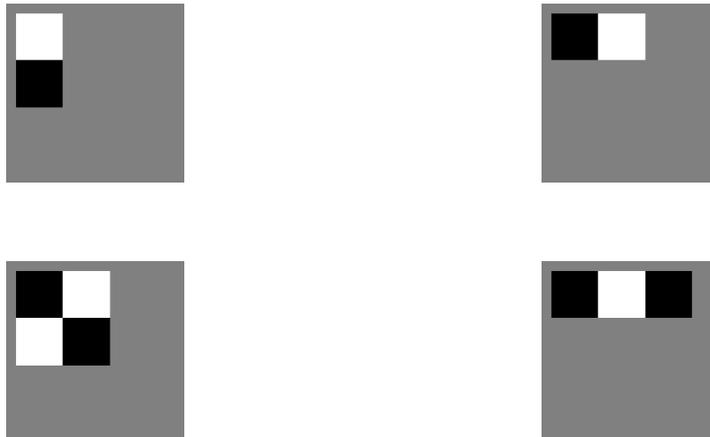
Haar-Wavelet Μετασχηματισμός Τα Haar είναι τα πρώτα γνωστά κυματίδια (wavelets), προτάθηκαν από τον Ούγγρο μαθηματικό Haar το 1910, και αποτελούν μια ορθοκανονική βάση σε έναν Hilbert χώρο. Στον διδιάστατο χώρο τα Haar χαρακτηριστικά παίρνουν την μορφή των τριών πρώτων στοιχείων του Σχήματος 2.1. Όταν ένα Haar χαρακτηριστικό εφαρμόζεται σε κάποια εικόνα, οι εντάσεις των περιοχών της εικόνας που ανήκουν στην λευκή και την μαύρη περιοχή αθροίζονται, και η διαφορά αυτών των δύο αθροισμάτων δίνει την τιμή του Haar χαρακτηριστικού στο συγκεκριμένο σημείο. Ως κυματίδια (wavelets) τα Haar χαρακτηριστικά συναντώνται σε όλες τις διεσταλμένες μορφές των αρχικών προτύπων (τρία πρώτα στοιχεία, Σχήματος 2.1), με πλάτος 2^n και ύψος 2^k , και σε όλες τις καβντισμένες μετατοπίσεις ως προς n (κατά πλάτος) και k (κατά ύψος). Το σύνολο αυτών των διεσταλμένων και μετατοπισμένων στοιχείων αποτελούν μια ορθοκανονική βάση, και ορίζουν έναν αντιστρέψιμο μετασχηματισμό της εικόνας, από τον χώρο των pixels, στον χώρο των Haar χαρακτηριστικών.

Quadruple Haar Transform: Οι Papageorgiou & Poggio [20] παρατήρησαν την ιδιότητα των Haar χαρακτηριστικών να εντοπίζουν ακμές και γωνίες (όταν έχουν μεγάλες τιμές) καθώς και ομοιόμορφες περιοχές (όταν έχουν μικρές τιμές), και πρότειναν τον χώρο των Haar χαρακτηριστικών ως πολύ πιο κατάλληλο από των pixels

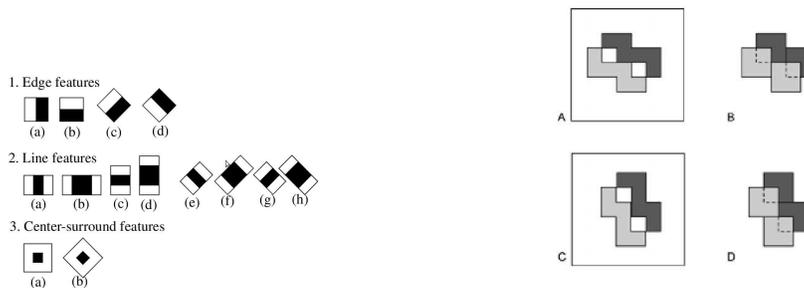
για την μοντελοποίηση αντικειμένων. Οι ίδιοι πρότειναν μια επέκταση του κλασσικού μετασχηματισμού Haar (Haar Transform), όπου τα Haar χαρακτηριστικά δεν μετατοπίζονται κατά n , αλλά κατά $\frac{1}{4}2^n$, γεγονός που τα κάνει πιο πυκνά στις εικόνες 19×19 όπου γίνεται ο μετασχηματισμός. Αυτός ο μετασχηματισμός ονομάστηκε Quadruple Haar Transform, και δεν αποτελεί μια ορθοκανονική βάση, αλλά μια υπερπλήρη βάση που προσδίδει μεγαλύτερη ευκινησία στα Haar χαρακτηριστικά και τους επιτρέπει να προσαρμόζονται και να περιγράφουν τα ιδιαίτερα χαρακτηριστικά των αντικειμένων της εικόνας (ακμές, γωνίες, ομοιόμορφες περιοχές). Μετά από αυτόν τον μετασχηματισμό οι Parageorgiou & Poggio επιλέγουν ένα μικρό σύνολο με τα πιο 'περιγραφικά' από αυτά τα χαρακτηριστικά, δηλαδή αυτά με τις στατιστικά μεγαλύτερες και τις μικρότερες τιμές, για την μοντελοποίηση των προσώπων.

Υπερπλήρης μετασχηματισμός στον χώρο Haar: Οι Viola & Jones [26] πηγαίνουν ένα βήμα παραπέρα στην υπερπληρότητα και ορίζουν έναν νέο μετασχηματισμό της εικόνας όπου τα Haar χαρακτηριστικά μπορούν πια να πάρουν όλα τα πιθανά μεγέθη (όχι δυνάμεις του 2), και μετατοπίζονται σε όλες τις πιθανές θέσεις πάνω στην εικόνα. Επίσης εισάγουν και ένα επιπλέον Haar-like χαρακτηριστικό (το τέταρτο του σχήματος 2.1). Με αυτόν τον τρόπο προσδίδεται απόλυτη ελευθερία στα Haar-like χαρακτηριστικά να προσαρμοστούν στα χαρακτηριστικά του αντικειμένου της εικόνας, ενώ τώρα πια η βάση γίνεται μακράν μεγαλύτερη, αφού από τα 575 pixels (ο μετασχηματισμός γίνεται σε 24×24 εικόνες) πηγαίνουμε σε περισσότερα από 160,000 Haar-like χαρακτηριστικά. Ένας τέτοιος υπερπλήρης μετασχηματισμός θα ήταν αδύνατος υπολογιστικά αν οι Viola & Jones δεν είχαν εισάγει την έννοια της ολοκληρωμένης εικόνας (integral image) που κάνει τον υπολογισμό των Haar και Haar-like χαρακτηριστικών πολύ γρήγορο (σταθερή υπολογιστική πολυπλοκότητα $O(1)$).

Περισσότερα Haar Χαρακτηριστικά: Οι Lienhart et al [16] επέκτειναν το σύνολο των Haar χαρακτηριστικών με νέα Haar-like σχήματα (Σχήμα 2.2 αριστερά) αλλά και με τα αυθεντικά χαρακτηριστικά υπό περιστροφή 45 μοιρών. Τα νέα Haar-like χαρακτηριστικά μπορούν επίσης να υπολογιστούν πάνω σε μια εικόνα με σταθερή υπολογιστική πολυπλοκότητα, ενώ αυξάνουν αισθητά τις επιδόσεις στην κατηγοριοποίηση. Επίσης οι Viola & Jones στην επέκταση του αλγορίθμου για την ανίχνευση προσώπων υπό προφίλ χρησιμοποίησαν δύο επιπλέον Haar-like χαρακτηριστικά που φαίνονται στο Σχήμα 2.2 (δεξιά). Γενικά σε προβλήματα ανίχνευσης αντικειμένων, έχουν προταθεί και χρησιμοποιηθεί πολλά διαφορετικά Haar χαρακτηριστικά ή διαφορετικές εκδοχές των προηγούμενων, ανάλογα με το πρόβλημα και την γεωμετρία του αντικειμένου προς ανίχνευση.



Σχήμα 2.1: Τα 4 βασικά είδη των Haar-like χαρακτηριστικών [26]



Σχήμα 2.2: Ένα επεκτεταμένο σύνολο από Haar Features [16] [25]

2.2.2 Ο αλγόριθμος AdaBoost

Εισαγωγή

Το 1995 οι Yoav Freund και Robert Schapire πρότειναν έναν νέο αλγόριθμο που έχει την ικανότητα να συνδυάζει πολλούς παρόμοιους αδύναμους classifiers σε έναν ισχυρό [9]. Ο αλγόριθμος αυτός σε κάθε βήμα βρίσκει τον καλύτερο αδύναμο classifier βάσει κάποιας στρατηγικής, και τον συνδυάζει με τον βέλτιστο τρόπο με αυτούς που επιλέχτηκαν σε προηγούμενα βήματα ώστε να παράγει τον τελικό δυνατό classifier ως συνδυασμό αυτών. Οι Freund & Schapire ονόμασαν αυτόν τον νέο αλγόριθμο AdaBoost γιατί ενισχύει (boosts) την επίδοση των αδύναμων classifiers συνδυάζοντάς τους, ενώ παράλληλα προσαρμόζει (adapt) τους νέους classifiers που προσθέτονται, έτσι ώστε αυτοί να δίνουν έμφαση στα στοιχεία που παρουσίασαν δυσκολίες στην κατηγοριοποίηση τους στα προηγούμενα βήματα.

Οι Freund και Schapire δεν πρότειναν απλώς έναν ακόμη αλγόριθμο που στόχο έχει την κατηγοριοποίηση, αλλά απέδειξαν και θεωρητικά ότι ο αλγόριθμος αυτός συγκλίνει ασυμπτω-

τικά σε μηδενικό λάθος (στο σύνολο εκπαίδευσης) με το πέρας των επαναλήψεων, έδωσαν ενδείξεις για την επιτυχία του αλγορίθμου στην γενική περίπτωση, και θεμελίωσαν και την πιθανοτική ερμηνεία του αλγορίθμου βάσει της Bayesian θεωρίας. Στις επόμενες παραγράφους εξηγείται ο αλγόριθμος σε βάθος και παρατίθενται τα σημαντικότερα χαρακτηριστικά του.

Περιγραφή του αλγορίθμου

Έστω X ένα σύνολο από αντικείμενα προς κατηγοριοποίηση, που υπακούει σε μια πιθανοτική κατανομή D . Κάθε αντικείμενο x ανήκει σε μια κατηγορία $\{0, 1\}$ σύμφωνα με την άγνωστη συνάρτηση (oracle) $c : X \rightarrow \{0, 1\}$. Αυτό που αναζητούμε είναι μια συνάρτηση $h : X \rightarrow [0, 1]$, όπου $h(x)$ εκφράζει την πιθανότητα το στοιχείο x να ανήκει στην κατηγορία 1, ενώ η ποσότητα $1 - h(x)$ είναι η πιθανότητα να ανήκει στην κατηγορία 0. Το λάθος της υπόθεσης h είναι η αναμενόμενη τιμή $E(|h(x) - c(x)|)$, όπου το x επιλέγεται σύμφωνα με την κατανομή D . Το λάθος αυτό θέλουμε να ελαχιστοποιήσουμε.

Ο αλγόριθμος αρχικοποιεί την κατανομή D ως ομοιόμορφη και διαμορφώνει μια αρχική υπόθεση h_1 καλώντας μια συνάρτηση $\text{WeakLearn}(X, D)$ με αυτή την κατανομή. Σύμφωνα με το λάθος που παράγει η υπόθεση αυτή, η κατανομή D ανανεώνεται ώστε να δίνει μεγαλύτερη πιθανότητα στα αντικείμενα που κατηγοριοποιήθηκαν λάθος και μικρότερη σε αυτά που κατηγοριοποιήθηκαν σωστά. Στο επόμενο βήμα η ίδια συνάρτηση WeakLearn ξανακαλείται με την ανανεωμένη πλέον κατανομή ως είσοδο, και δημιουργεί μια νέα υπόθεση h_2 , κ.ο.κ. Στο τέλος όλες αυτές οι υποθέσεις συνδυάζονται (αθροίζονται με κατάλληλους συντελεστές) με την κάθε μια να επηρεάζει το τελικό αποτέλεσμα ανάλογα με την επίδοσή της στην κατηγοριοποίηση (weighted majority voting). Ο αλγόριθμος παρουσιάζεται παρακάτω:

ADABOOST:

Είσοδος:

1. Μια ακολουθία n δειγμάτων (x) με τις ετικέτες της κατηγορίας στην οποία ανήκουν $(y) < (x_1, y_1), \dots, (x_n, y_n) >$
2. Η αρχική κατανομή D πάνω στα δείγματα x . (Συνήθως είναι ομοιόμορφη)
3. Ένας αλγόριθμος WeakLearn που παράγει αδύναμους classifiers για δεδομένη κατανομή D
4. Ένας αχέραιος T που προσδιορίζει τον αριθμό των επαναλήψεων

Αλγόριθμος:

Αρχικοποίησε τα βάρη $w_i^1 = D(i)$ για $i = 1, \dots, N$

Για $t = 1, 2, \dots, T$:

1. Ανανέωσε την πιθανοτική κατανομή $p^t = \frac{w_i^t}{\sum_{i=1}^N w_i^t}$.

2. Κάλεσε τον αλγόριθμο WeakLearn με είσοδο αυτή την πιθανοτική κατανομή, ο οποίος επιστρέφει την βέλτιστη υπόθεση $h_t : X \rightarrow [0, 1]$
3. Υπολόγισε το λάθος της κατηγοριοποίησης της υπόθεσης h_t $\epsilon_t = \sum_{i=1}^N p_i^t |h_t(x_i) - y_i|$.
4. $\beta_t = \epsilon_t / (1 - \epsilon_t)$ (μεγαλώνει όταν μεγαλώνει το λάθος)
5. Ανανέωσε τα βάρη: $w_i^{t+1} = w_i^t \beta_i^{1 - |h_t(x_i) - y_i|}$ (όταν έχουμε λάθος κατηγοριοποίηση το βάρος μεγαλώνει, αλλιώς παραμένει ίδιο)

Έξοδος

Η τελική ισχυρή υπόθεση δίνεται από το σταθμισμένο άθροισμα των αδύναμων υποθέσεων:

$$h_f(x) = \sum_{t=1}^T \log(1/\beta_t) h_t(x) \geq \frac{1}{2} \sum_{t=1}^T \log(1/\beta_t) \quad (2.1)$$

Από την παραπάνω εξίσωση φαίνεται ότι οι συντελεστές των επιμέρους αδύναμων υποθέσεων h_t μεγαλώνουν όσο το λάθος των υποθέσεων αυτών μικραίνει.

Θεωρητική ανάλυση

Για τον αλγόριθμο Adaboost όπως περιγράφηκε παραπάνω, οι Yoan & Schapire [9] απέδειξαν ότι ανεξαρτήτως του αλγορίθμου που δημιουργεί τις αδύναμες υποθέσεις h_t (WeakLearn), όσο αυτός δίνει λάθος κατηγοριοποίησης μικρότερο του 50%, ισχύει το παρακάτω θεώρημα:

Θεώρημα AdaBoost. Έστω ο αλγόριθμος WeakLearn ο οποίος κάθε φορά που καλείται επιστρέφει μια αδύναμη υπόθεση h_t με λάθος $\epsilon_1, \epsilon_2, \dots, \epsilon_T$, όπως περιγράφηκε προηγουμένως. Τότε το συνολικό λάθος της τελικής υπόθεσης h_f που προέρχεται από τον AdaBoost, $\epsilon = Pr_D[h_f(x) \neq y_i]$ είναι άνω φραγμένο σύμφωνα με την σχέση:

$$\epsilon \leq 2^T \prod_{t=1}^T \sqrt{\epsilon_t(1 - \epsilon_t)} \quad (2.2)$$

Σύμφωνα με το παραπάνω θεώρημα αποδεικνύεται και μαθηματικά ότι ο αλγόριθμος Adaboost έχει την ικανότητα να ενισχύει επιτυχώς αδύναμους classifiers και να τους συνδυάζει σε έναν δυνατό, του οποίου το λάθος συγκλίνει ασυμπτωτικά στο μηδέν με το πέρασμα των επαναλήψεων. Για την απόδειξη του παραπάνω θεωρήματος, παραπέμπουμε στην αντίστοιχη δημοσίευση [9].

Το λάθος που τείνει στο μηδέν είναι το λάθος πάνω στα δεδομένα εκπαίδευσης και όχι το γενικευμένο λάθος κατηγοριοποίησης, το οποίο μπορεί να μεγαλώσει με την αύξηση των επαναλήψεων T (φαινόμενο overfitting). Παρόλα αυτά ο αλγόριθμος AdaBoost έγινε ιδιαίτερα διάσημος και εφαρμόστηκε επιτυχώς σε πάρα πολλά προβλήματα κατηγοριοποίησης λόγω

ακριβώς της ιδιότητάς του να αποφεύγει αυτό το φαινόμενο της υπερεκπαίδευσης (overfitting). Συγκεκριμένα στην συντριπτική πλειοψηφία των πειραμάτων που έγιναν με τον AdaBoost, το γενικευμένο λάθος μειώνεται με την αύξηση των επαναλήψεων. Αυτή η ιδιότητα του AdaBoost που τον καθιστά τόσο επιτυχημένο εξηγήθηκε αργότερα από τους Friedman και Hastie [10].

Οι Friedman και Hastie [10] εξέτασαν τον AdaBoost από την σκοπιά της στατιστικής και έδωσαν μια πιο συμπαγή ερμηνεία στο γιατί ο αλγόριθμος αυτός είναι τόσο επιτυχημένος στο να αποφεύγει το overfitting. Οι Friedman και Hastie απέδειξαν ότι ο AdaBoost, με το να ελαχιστοποιεί σε κάθε επανάληψη ένα εκθετικό κριτήριο ($e^{yF(x)}$, $y \in \{-1, 1\}$), ουσιαστικά αποτελεί μια επαναληπτική διαδικασία για την εκπαίδευση ενός αθροιστικού μοντέλου λογιστικής παλινδρόμησης (additive logistic regression model) (βλ. Παράρτημα). Αυτό το γεγονός εξηγεί από μόνο του την επιτυχία του AdaBoost απέναντι στο overfitting, αφού η εφαρμογή ενός αθροιστικού μοντέλου απλών συναρτήσεων επεκτείνει την κλάση των συναρτήσεων που μπορούν να εκτιμηθούν, σε κάθε βήμα. Επίσης η ελαχιστοποίηση του εκθετικού κριτηρίου ($e^{yF(x)}$, $y \in \{-1, 1\}$) συμπίπτει (προσεγγιστικά) με την μεγιστοποίηση της πιθανοφάνειας, και παράγει πιο αγνές λύσεις αποφεύγοντας το φαινόμενο της υπερ-εκπαίδευσης (overfitting).

Από την στατιστική ερμηνεία του AdaBoost αποδεικνύεται επίσης ότι ο αλγόριθμος αυτός προσεγγίζει την ύστερη πιθανοτική κατανομή $p(y|x)$ μέσω της σχέσης:

$$p(y|x) = \frac{e^{y \sum_{j=1}^T \alpha_j f_j(x,y)}}{e^{-y \sum_{j=1}^T \alpha_j f_j(x,y)} + e^{y \sum_{j=1}^T \alpha_j f_j(x,y)}}, y \in \{-1, 1\}, \quad (2.3)$$

όπου $\sum_{j=1}^T \alpha_j f_j(x, y)$ είναι ο τελικός ισχυρός classifier εκπαιδευμένος από τον αλγόριθμο AdaBoost.

Ιδιότητες και επεκτάσεις

Ο AdaBoost όπως περιγράφηκε παραπάνω μπορεί να ερμηνευτεί από αλγοριθμική σκοπιά σαν ένας αλγόριθμος που ενισχύει την επίδοση αδύναμων όμοιων classifiers, κάθε ένας από τους οποίους απευθύνεται σε διαφορετική πιθανοτική κατανομή των δειγμάτων εισόδου, αθροίζοντας τους με κατάλληλους συντελεστές (weighted majority voting). Επίσης από μαθηματικής πλευράς μπορεί να ερμηνευτεί σαν την εφαρμογή ενός αθροιστικού μοντέλου λογιστικής παλινδρόμησης στα δείγματα εισόδου, που σε κάθε βήμα ελαχιστοποιεί το εκθετικό κριτήριο ($e^{yF(x)}$, $y \in \{-1, 1\}$). Πολλές διαφορετικές εκδοχές του αλγορίθμου AdaBoost που διατηρούν τα βασικά του αυτά χαρακτηριστικά έχουν προταθεί, οι σημαντικότερες από τις οποίες παρουσιάζονται με συντομία παρακάτω:

Discrete AdaBoost Είναι ο αυθεντικός αλγόριθμος AdaBoost όπως ακριβώς περιγράφηκε παραπάνω, μόνο που οι αδύναμες υποθέσεις h_t επιστρέφουν διακριτές τιμές $\{-1, 1\}$ [10].

Real AdaBoost Το πρόσημο του τελικού classifier αποφασίζει την κλάση στην οποία κατηγοριοποιείται η είσοδος, ενώ το μέτρο του δείχνει την εμπιστοσύνη αυτού του αποτελέσματος [10].

Gentle AdaBoost Οι αδύναμες υποθέσεις h_t παράγονται από την σταθμισμένη μέθοδο ελαχίστων τετραγώνων των y_i πάνω στα x_i . Στο τελικό άθροισμα δεν συμπεριλαμβάνονται συντελεστές [10].

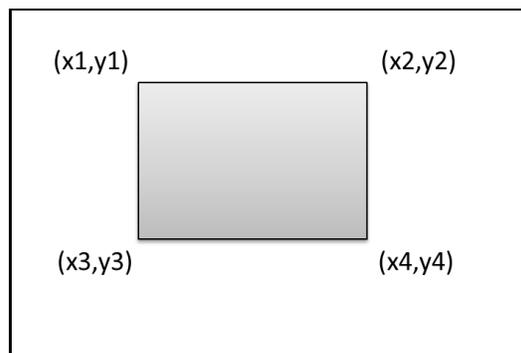
Συμπεράσματα

Συνοψίζοντας ο αλγόριθμος AdaBoost είναι ένας πολύ ισχυρός αλγόριθμος κατηγοριοποίησης που χρησιμοποιείται με μεγάλη επιτυχία σε αναρίθμητες εφαρμογές, ενώ αποτελεί ίσως την πιο επιτυχημένη μέθοδο στο πεδίο της ανίχνευσης αντικειμένων σήμερα. Ο ίδιος ή παραλλαγές του αλγορίθμου που τον συνδυάζουν με δένδρα αποφάσεων, όπως ο αλγόριθμος AdaTree [11] ή ο αλγόριθμος Probabilistic Boosting Tree (PBT) [24] έχουν προταθεί και χρησιμοποιούνται σήμερα σε πραγματικές εμπορικές εφαρμογές ανίχνευσης αντικειμένων.

2.2.3 Η συνεισφορά των P. Viola και M. Jones

Παρόλο το γεγονός ότι και η ιδέα της χρήσης Haar χαρακτηριστικών για την μοντελοποίηση των αντικειμένων, αλλά και ο αλγόριθμος AdaBoost προϋπήρχαν και θεωρούνταν ήδη ιδιαίτερα επιτυχημένα, η συνεισφορά των Viola & Jones είναι ιδιαίτερα σημαντική και αποφασιστική στην ανάπτυξη ενός θεωρητικού πλαισίου για την ανίχνευση προσώπων και γενικότερα αντικειμένων. Η προσφορά των Viola & Jones συνοψίζεται σε τρεις βασικές ιδέες. Οι τρεις αυτές ιδέες, καθώς και η σημασία τους για την ανίχνευση προσώπων σε πραγματικό χρόνο αναλύονται παρακάτω:

Ολοκληρωμένη εικόνα (integral image)



Σχήμα 2.3: Οπτικοποίηση του υπολογισμού αθροίσματος τετραγώνου μέσω της ολοκληρωμένης εικόνας

Η πρώτη βασική καινοτομία που πρότειναν οι Viola & Jones είναι η ολοκληρωμένη εικόνα (integral image). Η ολοκληρωμένη εικόνα ii έχει το ίδιο μέγεθος με την αυθεντική εικόνα i και σε κάθε σημείο (x, y) παίρνει την τιμή του αθροισμάτων των εντάσεων της εικόνας i στην περιοχή $(x_i, y_i) : 0 \leq x_i < x, 0 \leq y_i < y$. Η ολοκληρωμένη εικόνα μπορεί να υπολογιστεί με

ένα πέρασμα πάνω στην αυθεντική εικόνα, και με την βοήθεια της μπορούμε να υπολογίσουμε το άθροισμα των εντάσεων των στοιχείων της εικόνας μέσα σε ένα αυθαίρετο ορθογώνιο σε σταθερό χρόνο. Πιο συγκεκριμένα το άθροισμα των εντάσεων μέσα στο ορθογώνιο με ακμές $(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4)$ (Σχήμα 2.3) προκύπτει απλώς από την αναφορά σε τέσσερα σημεία της ολοκληρωμένης εικόνας: $Sum = ii(x_1, y_1) + ii(x_4, y_4) - ii(x_2, y_2) - ii(x_3, y_3)$.

Με αυτόν τον τρόπο μπορούμε να υπολογίσουμε την τιμή των Haar-like χαρακτηριστικών πάνω σε μια εικόνα σε σταθερό χρόνο, πράγμα που μας επιτρέπει να χρησιμοποιούμε μια τεράστια υπερπλήρη βάση από Haar-like χαρακτηριστικά για την καλύτερη μοντελοποίηση των αντικειμένων, χωρίς να επιβαρύνουμε την εφαρμογή με μεγάλο υπολογιστικό κόστος.

Συνδυασμός των Haar-like χαρακτηριστικών με τον αλγόριθμο AdaBoost

Η δεύτερη βασική καινοτομία είναι η χρήση του αλγορίθμου AdaBoost με Haar-like χαρακτηριστικά (συνδυασμένα με κάποιο κατώφλι κατηγοριοποίησης) ως αδύναμους classifiers για την διαδικασία της κατηγοριοποίησης. Ο αλγόριθμος AdaBoost επιλέγει σε κάθε βήμα το πιο περιγραφικό Haar-like χαρακτηριστικό για την μοντελοποίηση του προσώπου, και τελικά συνδυάζει όλα τα επιλεγμένα χαρακτηριστικά σε έναν δυνατό classifier. Με αυτόν τον τρόπο έκτος του ότι αξιοποιείται η ικανότητα του AdaBoost να παράγει ισχυρούς classifiers με μεγάλη δυνατότητα γενίκευσης, η διαδικασία της μοντελοποίησης (επιλογή των Haar χαρακτηριστικών) αυτοματοποιείται και ενοποιείται με την διαδικασία της εκπαίδευσης του τελικού classifier.

Οι Viola & Jones χρησιμοποίησαν τον κλασικό Discrete AdaBoost για την εκπαίδευση του classifier. Αργότερα οι Lienhart et al πειραματίστηκαν με διαφορετικές εκδοχές του AdaBoost (Discrete, Gentle, Real AdaBoost) [15] και πέτυχαν αρκετά καλύτερα αποτελέσματα χρησιμοποιώντας τον Gentle AdaBoost.

Αλυσίδα από classifiers (Cascade Classifiers)

Η τρίτη και ιδιαίτερα σημαντική συνεισφορά που έχει να κάνει με την ταχύτητα της ανίχνευσης προσώπων σε εικόνες, και επιτρέπει την εφαρμογή του αλγορίθμου σε πραγματικό χρόνο (real time), είναι η εισαγωγή της έννοιας της αλυσίδας classifiers. Η βασική ιδέα είναι ότι κάθε τμήμα της εικόνας που πρόκειται να κατηγοριοποιηθεί ως πρόσωπο ή μη-πρόσωπο δεν περνάει από έναν αλλά από μια σειρά από classifiers (στην τελική εφαρμογή του αλγορίθμου των Viola & Jones χρησιμοποιούνται 38 βήματα classification). Οι αρχικοί classifiers στην αλυσίδα είναι πολύ απλοί (συνδυασμός λίγων αδύναμων classifiers) και ελαστικοί, και έχουν στόχο την γρήγορη απόρριψη του μεγαλύτερου μέρους των τμημάτων της εικόνας που σίγουρα δεν αποτελούν πρόσωπα. Τα υπόλοιπα τμήματα της εικόνας περνάνε στους πιο σύνθετους και αυστηρούς classifiers των μεγαλύτερων επιπέδων ώστε να επιτευχθεί σωστή κατηγοριοποίηση. Η αλυσίδα classifiers αποτελεί ουσιαστικά ένα εκφυλισμένο δένδρο αποφάσεων, όπου σε κάθε βήμα απορρίπτεται ένα δεδομένο μέρος από μη-πρόσωπα, και επιτυγχάνει την επιτάχυνση του αλγορίθμου σε σημαντικό βαθμό.

2.3 Υλοποίηση και Προσαρμογή στο Συγκεκριμένο Πρόβλημα

Μέρος της διπλωματικής αυτής εργασίας είναι η υλοποίηση του αλγορίθμου των Viola & Jones για την ανίχνευση προσώπων σε εικόνες, και η προσαρμογή του στις ιδιαιτερότητες του δεδομένου προβλήματος. Στην παρούσα ενότητα περιγράφονται με συνομία κάποια βασικά σημεία της υλοποίησης όσον αφορά την εκπαίδευση του classifier και την εφαρμογή του σε εικόνες της νοηματικής γλώσσας. Στο τέλος της ενότητας περιγράφεται η διαδικασία κατασκευής μιας μικρής βάσης δεδομένων εκπαίδευσης, καθώς και κάποια πειράματα που έγιναν με αυτή.

2.3.1 Υλοποίηση

Εκπαίδευση

Πριν την διαδικασία της εκμάθησης έχουν κατασκευαστεί δύο βάσεις δεδομένων εκπαίδευσης, μια με πρόσωπα και μία με μη-πρόσωπα. Τόσο τα πρόσωπα όσο και τα μη-πρόσωπα αποτελούνται από εικόνες που έχουν σμικρυνθεί σε κλίμακα 19×19 pixels.

Για να ελαχιστοποιήσουμε την μεταβλητότητα που οφείλεται στην διαφορά των συνθηκών φωτισμού στις διαφορετικές εικόνες προσώπων χρειάζεται να κανονικοποιήσουμε τις εικόνες αφαιρώντας την μέση τιμή της έντασης και διαιρώντας με την τυπική απόκλιση: $I'(x, y) = \frac{I(x, y) - \mu}{\sigma}$. Αυτή η διαδικασία γίνεται τόσο κατά την εκπαίδευση όσο και κατά την ανίχνευση και για αυτό πρέπει να γίνεται με μικρή υπολογιστική πολυπλοκότητα. Αυτό γίνεται με την βοήθεια της τετραγωνικής ολοκληρωμένης εικόνας (squared itegral image) $ii^2(x_i, y_i) = \sum_{x < x_i, y < y_i} i^2(x, y)$, η οποία επιτρέπει τον υπολογισμό της τυπικής απόκλισης της εικόνας σε σταθερό χρόνο.

Μετά την προεπεξεργασία των εικόνων εκπαίδευσης, εφαρμόζεται πάνω τους το υπερπλήρες σύνολο από Haar-like χαρακτηριστικά και αποθηκεύονται τα ενδιάμεσα αποτελέσματα. Η βάση των χαρακτηριστικών περιορίζεται σε αυτά που είναι μεγαλύτερα ή ίσα από 4×4 pixels. Αυτός ο περιορισμός γίνεται για να πετύχουμε όσο το δυνατόν καλύτερη γενίκευση δεδομένου ότι η βάση εκπαίδευσης είναι πολύ μικρή. Μετά την εφαρμογή των χαρακτηριστικών πάνω στις εικόνες τα αποτελέσματα αποθηκεύονται για να χρησιμοποιηθούν για την εκπαίδευση του classifier.

Στο βασικό στάδιο της εκπαίδευσης τα δεδομένα αυτά ανακαλούνται και δίνονται ως είσοδο στον αλγόριθμο AdaBoost, ο οποίος εκπαιδεύει την τελική συνάρτηση κατηγοριοποίησης (classifier) ως ένα σταθμισμένο άθροισμα από αδύναμους classifiers με διαφορετικούς συντελεστές. Για την εκπαίδευση χρησιμοποιείται ο αλγόριθμος Discrete AdaBoost με $T = 50$ weak classifiers. Παρακάτω περιγράφεται η διαδικασία εκπαίδευσης σε μορφή ψευδοκώδικα.

Training Pseudocode:

1. Produce all possible haar-like features for 19x19 images

2. Load all images

Convert to grayscale

Normalize images by subtracting the mean and dividing standard deviation

Calculate the corresponding integral images

3. Apply all Haar-like features on integral images

Save the results

4. Load the results

Use AdaBoost to create the strong classifier

Save the classifier

Ανίχνευση

Μετά την παραγωγή του classifier από το στάδιο της εκπαίδευσης αυτός πρέπει να εφαρμοστεί σε όλα τα υποτμήματα 19×19 και σε όλες τις κλίμακες της νέας εικόνας στην οποία θέλουμε να αναγνωρίσουμε τα πρόσωπα. Αν το υποτμήμα της εικόνας που περνάει από τον classifier έχει πιθανότητα μεγαλύτερη από ένα κατώφλι τότε αυτό αναγνωρίζεται θετικά ως πρόσωπο, αλλιώς απορρίπτεται.

Όλα τα υποτμήματα της εικόνας πριν μπουν στην διαδικασία της κατηγοριοποίησης πρέπει να κανονικοποιηθούν ως προς την φωτεινότητά τους, όπως περιγράφεται στην προηγούμενη παράγραφο, και να υπολογιστεί η ολοκληρωμένη τους μορφή.

Αντί για την διαδικασία της αλυσίδας classifiers υλοποιείται μια απλούστερη μέθοδος για την εύκολη απόρριψη τμημάτων της εικόνας που σίγουρα δεν αποτελούν πρόσωπα. Η μέθοδος βασίζεται αποκλειστικά στην μεταβλητότητα της εικόνας και σε στατιστικά στοιχεία που αποκτήθηκαν από το σύνολο εκπαίδευσης. Πιο συγκεκριμένα, από το σύνολο εκπαίδευσης εκτιμούμε ένα εύρος τυπικών αποκλίσεων που αντιστοιχούν σε εικόνες προσώπων. Αν η τυπική απόκλιση του τμήματος προς κατηγοριοποίηση δεν ανήκει σε αυτό το εύρος τότε απορρίπτεται

χωρίς να περάσει την διαδικασία classification. Αυτή η μέθοδος επιτρέπει την εύκολη απόρριψη ομοιόμορφων περιοχών του φόντου και επιταχύνει την διαδικασία ανίχνευσης.

Τέλος μετά την διαδικασία ανίχνευσης, ακολουθεί η διαδικασία κλαδέματος (pruning) των περιοχών αναγνώρισης. Πιο συγκεκριμένα αν δύο ή περισσότερες επιτυχείς αναγνώρισεις επικαλύπτονται σε ποσοστό μεγαλύτερο του 50% τότε διατηρείται μόνο η αναγνώριση με την μεγαλύτερη πιθανότητα.

Detection Pseudocode:

for all candidate scales

 Rescale image

 Calculate the integral and the square integral image

 for all the 19x19 subimages of the image

 Calculate the mean and standard deviation of the subimage
 (using the integral and square intergal image)

 if std is within face-image region

 Normalize the subimage

 Calculate the integral subimage
 (using the original integral image)

 Classify subimage

 if probability>threshold

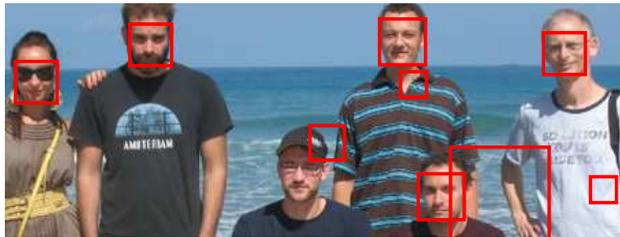
 Save detection

Prune detections

Η εφαρμογή υλοποιήθηκε με την βοήθεια της γλώσσας προγραμματισμού Matlab. Η υλοποίηση έγινε κατά την διάρκεια εκπόνησης πρακτικής στην ομάδα Signal and Image Processing

Laboratory (SIPL), Technion Israel Institute of Technology. Συνολικά για την εφαρμογή χρησιμοποιήθηκαν τεχνικές vector computing για καλύτερη αποδοτικότητα στην δεδομένη γλώσσα προγραμματισμού.

Αποτελέσματα



Σχήμα 2.4: Αποτελέσματα ανίχνευσης προσώπου στην ομάδα SIPL (1)



Σχήμα 2.5: Αποτελέσματα ανίχνευσης προσώπου στην ομάδα SIPL (2)

2.3.2 Προσαρμογή στο συγκεκριμένο πρόβλημα

Ο γενικευμένος αλγόριθμος για την ανίχνευση προσώπων απλοποιήθηκε και προσαρμόστηκε στο δεδομένο πρόβλημα της νοηματικής γλώσσας σύμφωνα με τις ιδιαιτερότητες που αυτό παρουσιάζει:

Μοναδικό πρόσωπο στην εικόνα: Εχμεταλλευόμενοι την πληροφορία ότι στα video της νοηματικής γλώσσας υπάρχει ένας μοναδικός νοηματιστής, και επομένως ένα μοναδικό πρόσωπο επιλέγουμε ως επιτυχημένη αναγνώριση το κομμάτι της εικόνας με την μεγαλύτερη πιθανότητα σύμφωνα με τον classifier. Επομένως δεν χρειάζεται να αποφασίσουμε σχετικά με κάποιο κατώφλι αναγνώρισης.

Μικρή κίνηση του προσώπου: Το πρόσωπο του νοηματιστή κινείται σε ένα περιορισμένο πλαίσιο σε σχέση με την συνολική εικόνα. Αφού βρεθεί επιτυχώς το πρόσωπο

στην πρώτη εικόνα του video, στις επόμενες εικόνες η αναζήτηση περιορίζεται σε ένα περιορισμένο πλαίσιο με κέντρο το πρώτο πρόσωπο και έτσι περιορίζεται το υπολογιστικό κόστος σε κάθε εικόνα της ακολουθίας.

Σταθερή απόσταση του προσώπου ως προς την κάμερα: Η σχετική απόσταση του προσώπου ως προς την κάμερα δεν μεταβάλλεται ιδιαίτερα κατά την διάρκεια της ομιλίας του νοηματιστή, πράγμα που μας επιτρέπει να περιορίσουμε και τις κλιμακώσεις της εικόνας σε τιμές κοντά σε αυτήν της πρώτης ανίχνευσης.

Οι παραπάνω απλοποιήσεις βελτιστοποιούν την απόδοση του αλγορίθμου όσον αφορά το αποτέλεσμα, αλλά κυρίως όσον αφορά την υπολογιστική πολυπλοκότητα. Επίσης η αναζήτηση σε περιορισμένη περιοχή καθιστά μη-αναγκαία την υλοποίηση της αλυσίδας classifiers για γρήγορα αποτελέσματα, ενώ η αναζήτηση σε μικρό εύρος κλιμακώσεων καθιστά μη αναγκαία την κλιμάκωση των χαρακτηριστικών αντί της εικόνας.

2.4 Κατασκευή και Πειραματισμός με την Βάση Δεδομένων Εκπαίδευσης

2.4.1 Εισαγωγή

Η δημιουργία μιας καλής βάσης δεδομένων εκπαίδευσης είναι ίσως πιο ο σημαντικός παράγοντας που θα καθορίσει την επίδοση του ανιχνευτή προσώπων, δεδομένου ότι ο αλγόριθμος δεν διαφέρει συνήθως σημαντικά από τον αυθεντικό που πρότειναν οι Viola και Jones. Ο πειραματισμός με διάφορα στοιχεία, αυτής της βάσης έχει μεγάλο ενδιαφέρον και επηρεάζει σημαντικά το τελικό αποτέλεσμα.

Στην κατασκευή μιας σωστής τέτοιας βάσης παρουσιάζονται αρκετές προκλήσεις που κυρίως έχουν να κάνουν με το τεράστιο μέγεθός της. Οι Viola & Jones [26] για την εκπαίδευση του αλγορίθμου τους χρησιμοποίησαν μια βάση δεδομένων που αποτελείται από 5,000 πρόσωπα και 10,000 μη-πρόσωπα. Οι Lienhart & Kuranov [15] χρησιμοποίησαν βάσεις αντίστοιχου μεγέθους για να πειραματιστούν με διαφοροποιήσεις του αρχικού αλγορίθμου. Τέλος η υλοποίηση της OpenCV είναι εκπαιδευμένη με μια αντίστοιχα μεγάλη βάση δεδομένων, οι κατασκευαστές δεν παρέχουν όμως πληροφορίες σχετικά με το πόσες και ποιες εικόνες χρησιμοποίησαν.

Όλα αυτά τα πρόσωπα πρέπει να ευθυγραμμιστούν, να κοπούν και να αποθηκευτούν χειρωνακτικά καθώς τέτοιες μεγάλες βάσεις δεν παρέχονται ελεύθερες. Επίσης για να εκπαιδευτεί ένας classifier με μια βάση δεδομένων τέτοιας τάξης μεγέθους χρειάζεται περίπου μια εβδομάδα χρόνου εκτέλεσης, εάν δεν έχει υπάρξει κατάλληλη παραλληλοποίηση.

Για αυτούς τους λόγους, σε αυτή την διπλωματική εργασία, περιοριζόμαστε σε μια πολύ μικρή βάση εκπαίδευσης που αποτελείται από 140 μόνο εικόνες προσώπων και από 300 περίπου εικόνες μη-προσώπων για τον πειραματισμό σχετικά με παραμέτρους που αφορούν τις εικόνες εκπαίδευσης. Τέλος με μια βάση τέτοιας τάξης μεγέθους εκπαιδεύεται ένας ανιχνευτής προσώπου που δίνει πολύ καλά αποτελέσματα για το συγκεκριμένο πρόβλημα. Σημειώνεται ότι στις βάσεις δεδομένων εκπαίδευσης δεν χρησιμοποιήθηκαν πρόσωπα ή τμήματα του φόντου

από τις εικόνες της νοηματικής γλώσσας στις οποίες εφαρμόστηκε η ανίχνευση προσώπου. Τα πρόσωπα ανακτήθηκαν από εικόνες στο internet και από την βάση προσώπων CUAVE [22].

2.4.2 Πειράματα με την βάση δεδομένων εκπαίδευσης

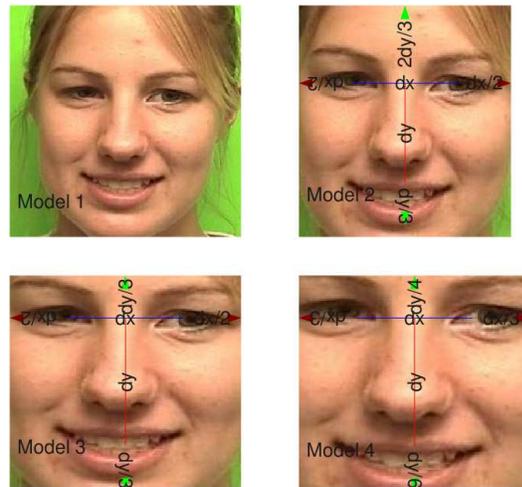
Σε αυτή την υποενότητα περιγράφονται κάποια πειράματα που εκτελέστηκαν πάνω στην βάση δεδομένων που χρησιμοποιείται για την εκπαίδευση του classifier. Τα πειράματα αυτά έχουν σκοπό τόσο την βελτίωση των ποσοστών επιτυχίας της ανίχνευσης, όσο και την καλύτερη κατανόηση των δυνατοτήτων και περιορισμών του παραπάνω πλαισίου. Από τα πειράματα αυτά προκύπτουν αρκετά ενδιαφέροντα συμπεράσματα, τα οποία αναλύονται στην τελευταία υποενότητα.

Για την εκτίμηση την επίδοση του Ανιχνευτή Προσώπων κατά τους πειραματισμούς, δημιουργήθηκαν 2 benchmarks που αποτελούνται από συνεχόμενα frames της νοηματικής γλώσσας, που ανακτήθηκαν από την βάση δεδομένων BU400 [19]. Σε κάθε frame σημειώθηκαν οι θέσεις των ματιών και του στόματος του προσώπου. Αν τα τρία αυτά σημεία βρίσκονται μέσα στο τετράγωνο που επιστρέφει ως αποτέλεσμα η ανίχνευση, τότε αυτό θεωρείται επιτυχημένο. Αν έστω και ένα από αυτά τα τρία σημεία βρίσκεται εκτός τετραγώνου τότε η ανίχνευση θεωρείται αποτυχημένη. Το benchmark 1 αποτελείται από τα 370 διαδοχικά frames της νοηματικής και χαρακτηρίζεται από σημαντικές αλλαγές στην πόζα και επικάλυψη του προσώπου από τα χέρια, ενώ το benchmark 2 αποτελείται από 500 frames και έχει μικρότερες αλλαγές στην πόζα αλλά περιλαμβάνει και αυτό σημαντική επικάλυψη.

Πείραμα 1: Μεταβολή της περιοχής εστίασης του προσώπου

Το πρώτο και σημαντικότερο πείραμα πάνω στις εικόνες προσώπων της βάσης εκπαίδευσης έχει να κάνει με την εστίαση της περιοχής του προσώπου. Δοκιμάστηκαν τέσσερα διαφορετικά μοντέλα στα οποία η περιοχή του προσώπου μεταβάλλεται από ολόκληρο το κεφάλι, μέχρι το στενό μοτίβο μάτια-στόμα (Σχήμα 2.6). Το πρώτο μοντέλο προέρχεται από την ανίχνευση μέσω του ανιχνευτή της OpenCV, και ανταποκρίνεται στην αντίστοιχη βάση δεδομένων εκπαίδευσης (αντίστοιχα στην βάση εκπαίδευσης των Viola & Jones, όπως και των Rainer Lienhart & Alexander Kuranov). Τα επόμενα μοντέλα προέκυψαν μετά από προσεκτική σημείωση των ματιών και του στόματος σε όλες τις εικόνες και κατάλληλη τομή. Με αυτό τον τρόπο και τα τρία αυτά μοντέλα είναι καλά ευθυγραμμισμένα, με την έννοια ότι τα μάτια και το στόμα σε όλες τις εικόνες είναι στο ίδιο επίπεδο. Στο Σχήμα 2.6 εκτός από την οπτική περιγραφή και απεικόνιση των τεσσάρων αυτών μοντέλων φαίνεται και η ελάχιστη, η μέγιστη, και η μέση τυπική απόκλιση που παρουσιάζουν. Η τυπική απόκλιση της έντασης στα pixel της εικόνας είναι ένας καλός δείκτης της πολυπλοκότητας μιας εικόνας. Το πρώτο μοντέλο παρουσιάζει σημαντικά μεγαλύτερες τυπικές αποκλίσεις από τα τρία επόμενα, και επομένως μεγαλύτερη πολυπλοκότητα, λόγω των μαλλιών και των περιοχών του φόντου που περιλαμβάνονται.

Στο Σχήμα 2.7 φαίνεται το ποσοστό επιτυχίας που πέτυχε ο αλγόριθμος εκπαιδευμένος με κάθε ένα από τα τέσσερα αυτά μοντέλα και στα δύο benchmarks. Είναι αξιοσημείωτη η

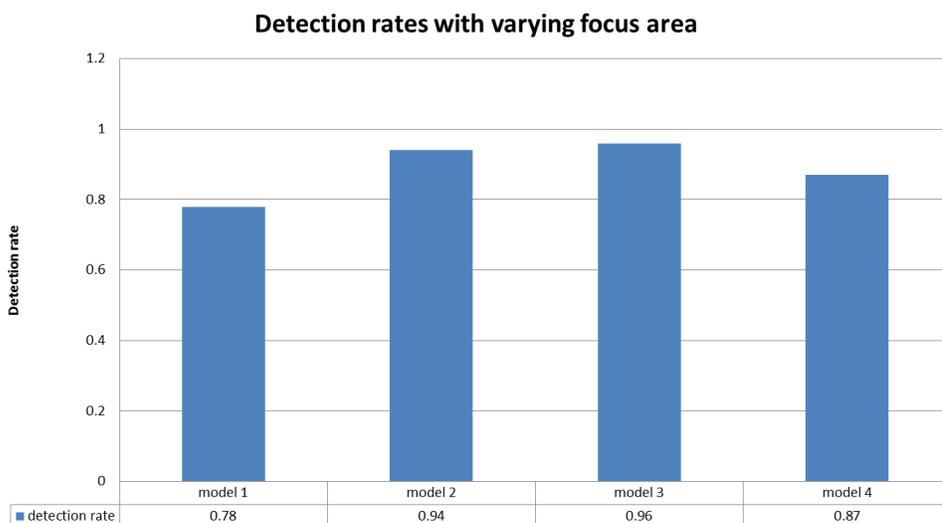


Model 1: face cropped by OpenCV			
Standard Deviation: min: 18	mean: 43.5	max: 75	
Model 2: aligned by annotation of eyes and mouth			
Standard Deviation: min: 14.5	mean: 34	max: 56.5	
Model 3: aligned and squared			
Standard Deviation: min: 13.4	mean: 31.2	max: 49	
Model 4: aligned small square pattern			
Standard Deviation: min: 12.3	mean: 30.6	max: 48	

Σχήμα 2.6: Μεταβάλλοντας την περιοχή εστίασης του προσώπου (4 διαφορετικά μοντέλα)

διακύμανση της επίδοσης του ανιχνευτή σε σχέση με την περιοχή εστίασης του προσώπου (και συνεπώς με την μεταβλητότητα των υποεικόνων). Τα χειρότερα αποτελέσματα τα δίνει το πρώτο μοντέλο το οποίο προφανώς εμπεριέχει περιττή πληροφορία (μαλλιά, φόντο) και περιττή πολυπλοκότητα πράγμα που αποσυγκεντρώνει τα επιλεγμένα Haar χαρακτηριστικά, πολλά από τα οποία επικεντρώνονται στον φόντο. Το τελευταίο μοντέλο επίσης δεν δίνει καλά αποτελέσματα πράγμα που οφείλεται ίσως στην ελλιπή πληροφορία αυτή τη φορά, αφού το μοτίβο του προσώπου είναι υπερβολικά στενό. Από το μοντέλο 2, το 3 δίνει ελαφρώς καλύτερα αποτελέσματα πράγμα που πιθανότατα οφείλεται στο γεγονός ότι είναι ήδη τετραγωνισμένο, και επομένως οι εικόνες δεν υπόκεινται σημαντική παραμόρφωση όταν μετασχηματίζονται σε τετράγωνα 19×19 .

Στα επόμενα πειράματα και εφαρμογές θα χρησιμοποιείται μόνο το μοντέλο 3, το οποίο είναι και το πιο επιτυχημένο.



Σχήμα 2.7: Ποσοστά Ανίχνευσης για τα 4 διαφορετικά μοντέλα εστίασης του προσώπου

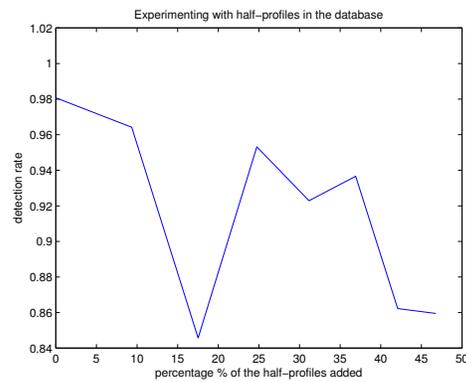
Πείραμα 2: Πρόσθεση εικόνων προφίλ στην βάση δεδομένων

Το δεύτερο πείραμα που διεξήχθη έχει να κάνει με την μεταβολή της πόζας του προσώπου κατά την ομιλία του νοηματιστή. Η αρχική υπόθεση είναι ότι με την πρόσθεση προσώπων υπό διαφορετικές πόζες στην βάση δεδομένων εκπαίδευσης, θα καταφέρουμε να αυξήσουμε τα ποσοστά επιτυχίας του ανιχνευτή. Μην γνωρίζοντας σε τι ποσοστό αυτά τα πρόσωπα θα είναι ευεργετικά για την βάση εκπαίδευσης, δοκιμάζεται η πρόσθεση διαφορετικών ποσοστών προσώπων υπό διαφορετικές γωνίες προφίλ. Τα νέα πρόσωπα που προστέθηκαν κυμαίνονται μέχρι γωνίες 45°, και ανακτήθηκαν από διάφορες βάσεις δεδομένων, ενώ ευθυγραμμίστηκαν σύμφωνα με το μοντέλο 3. Τα αποτελέσματα του πειράματος που περιγράφηκε παραπάνω φαίνονται στο Σχήμα 2.8.

Τα αποτελέσματα αυτά δείχνουν ότι η προσθήκη των εικόνων προσώπων υπό διαφορετικές γωνίες δεν βελτιώνει συνήθως την επίδοση της ανίχνευσης όπως θα ήταν αρχικά αναμενόμενο, αλλά αντιθέτως τις περισσότερες φορές την χειροτερεύουν. Αυτό αποδίδεται στο γεγονός ότι το στενό μοτίβο του προσώπου που χρησιμοποιείται εφαρμόζεται καλά και στην ελαφριά μεταβολή της πόζας, ενώ η προσθήκη των προφίλ χαλάει την ομοιομορφία της βάσης εκπαίδευσης και καθιστά δυσκολότερη την μοντελοποίηση του προσώπου από τον αλγόριθμο AdaBoost.

Πείραμα 3: Πρόσθεση εικόνων χειρομορφών στην βάση των μη-προσώπων

Για την κατασκευή της βάσης δεδομένων των μη-προσώπων χρησιμοποιήθηκαν τυχαία κομμάτια από τον φόντο των εικόνων από τις οποίες ανακτήθηκαν τα πρόσωπα. Για έναν ισχυρό ανιχνευτή προσώπων, η βάση μη-προσώπων θα πρέπει να περιλαμβάνει μια μεγάλη ποικιλία εικόνων. Όπως έχει αναφερθεί παραπάνω, η κλάση μη-πρόσωπο δεν είναι αυστηρώς ορισμένη, και επομένως σε αυτήν θα πρέπει να περιλαμβάνεται μια μεγάλη ποικιλία εικόνων,



Σχήμα 2.8: Ποσοστά Ανίχνευσης με διαφορετικά ποσοστά προσώπων προφίλ στην βάση εκπαίδευσης

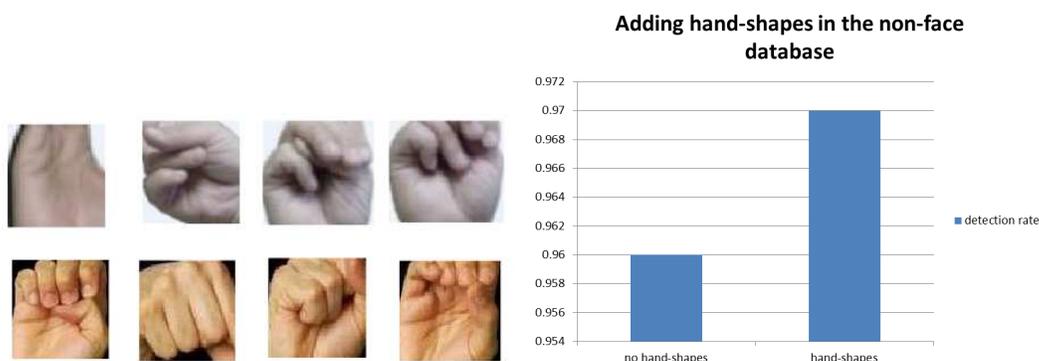
καθώς και εικόνες που λαμβάνονται συχνά λανθασμένα ως πρόσωπα. Συνήθως για αυτόν τον λόγο εφαρμόζονται bootstrapping τεχνικές, οι οποίες συμπεριλαμβάνουν στην κλάση μη-πρόσωπο τμήματα εικόνας που ο ανιχνευτής αναγνώρισε λάθος ως πρόσωπα σε προηγούμενο στάδιο. Για τον συγκεκριμένο ανιχνευτή που απευθύνεται σε ένα τόσο συγκεκριμένο πρόβλημα τα μη-πρόσωπα θα ήταν ιδανικά κομμάτια από το σώμα του ανθρώπου που θα μπορούσαν να μοιάζουν με πρόσωπο, καθώς και κάποιες γενικές εικόνες που θα απευθύνονται σε πιθανούς φόντους.

Τα κομμάτια του ανθρώπινου σώματος που συχνότερα λαμβάνονται λανθασμένα ως πρόσωπα είναι τα χέρια του, ιδιαίτερα στην περίπτωση της νοηματικής γλώσσας, όπου τα χέρια επιδίδονται σε διάφορες χειρομορφές πολλές από τις οποίες λόγω σχήματος και σκιών μοιάζουν αρκετά με πρόσωπο.

Για αυτόν τον λόγο έγινε προσπάθεια να περιληφθούν μέρη των διάφορων χειρομορφών από το αλφάβητο της νοηματικής που μοιάζουν σχετικά με πρόσωπα στην βάση των μη-προσώπων. Μετά την εκπαίδευση με την ανανεωμένη βάση, τα αποτελέσματα των ποσοστών ανίχνευσης πάνω στα δύο benchmarks φαίνονται στο Σχήμα 2.9. Τα αποτελέσματα δείχνουν μια μικρή βελτίωση καθώς κάποιες περιπτώσεις λανθασμένης αναγνώρισης χειρομορφών ως πρόσωπα πλέον απορρίπτονται. Παρ' όλα αυτά η κλάση όλων των πιθανών χειρομορφών είναι τεράστια και προφανώς δεν μπορεί να ληφθεί εξ' ολοκλήρου υπόψιν κατά την εκπαίδευση.

2.4.3 Συμπεράσματα

Από τα παραπάνω πειράματα και την γενικότερη διαδικασία κατασκευής μιας βάσης εκπαίδευσης για την ανίχνευση προσώπων προκύπτουν αρκετά ενδιαφέροντα συμπεράσματα σχετικά με την μεταβλητότητα που παρουσιάζουν τα ανθρώπινα πρόσωπα, καθώς και την εξάρτησή της με το απαιτούμενο μέγεθος της βάσης εκπαίδευσης.



Σχήμα 2.9: Ποσοστά ανίχνευσης μετά την προσθήκη χειρομορφών στην κλάση μη-προσώπων

Παράγοντες μεταβλητότητας των προσώπων

Μια καλή βάση εκπαίδευσης έχει σκοπό την σωστή μοντελοποίηση του ανθρώπινου προσώπου. Κατά την κατασκευή της πρέπει να ληφθούν υπόψιν οι σημαντικότεροι παράγοντες μεταβλητότητας της εμφάνισης του ανθρώπινου προσώπου σε εικόνες, που είναι οι παρακάτω:

Διαφοροποιήσεις από άνθρωπο σε άνθρωπο: Ο τρόπος να λάβουμε υπόψιν αυτές τις διαφοροποιήσεις είναι απλώς να συμπεριλάβουμε όσο το δυνατόν περισσότερα ενδεχόμενα στην βάση εκπαίδευσης. Μια καλή βάση εκπαίδευσης πρέπει να παίρνει υπ όψιν διαφοροποιήσεις λόγω φύλου, φυλής, έκφρασης του προσώπου, καθώς και ιδιαίτερα στοιχεία που μπορεί να παρουσιάζονται σε ένα πρόσωπο (γυαλιά, μούσι).

Διαφοροποιήσεις λόγω μεταβολής της πόζας: Παρόλο που έχουν γίνει προσπάθειες για την κατασκευή καθολικών ανιχνευτών για όλες τις πόζες του ανθρώπινου προσώπου, αυτό γενικά είναι δύσκολο με έναν μόνο classifier. Η πρόσθεση στην βάση δεδομένων εικόνων από όλες τις πιθανές πόζες καθιστά δύσκολη την μοντελοποίηση του προσώπου (καθώς η κλάση πρόσωπο αντιπροσωπεύει πλέον ένα αντικείμενο με πολύ μεγάλη μεταβλητότητα τόσο στην εμφάνιση όσο και στο σχήμα), και δίνει γενικά χειρότερα αποτελέσματα στην ανίχνευση, γεγονός που αντικατοπτρίζεται και από το πείραμα 2. Το πρόβλημα της μεταβολής της πόζας αντιμετωπίζεται εδώ με την στενή εστίαση του προσώπου (μοντέλο 3). Αυτό το στενό μοντέλο δεν μεταβάλλεται ιδιαίτερα με την μεταβολή της πόζας πράγμα που επιτρέπει την σωστή μοντελοποίηση με έναν μόνο classifier σε ένα σχετικά μεγάλο εύρος πόζας, γεγονός που επιβεβαιώνεται και από τα ποσοστά ανίχνευσης.

Διαφοροποιήσεις εξαιτίας του φωτισμού: Αυτή η κατηγορία περιλαμβάνει τόσο την ένταση (intensity) του φωτισμού όσο και την κατεύθυνση πρόσπτωσης του φωτός (illumination). Οι διαφοροποιήσεις λόγω γωνίας πρόσπτωσης του φωτός δεν λαμβάνονται συνήθως υπ όψιν ως αμελητέες, ενώ οι διαφοροποιήσεις στην ένταση απαλείφονται κατά την κανονικοποίηση της εικόνας, διαδικασία που περιγράφεται στην ενότητα 2.3.1.

Σχέση μεταβλητότητας του αντικειμένου - μεγέθους της βάσης

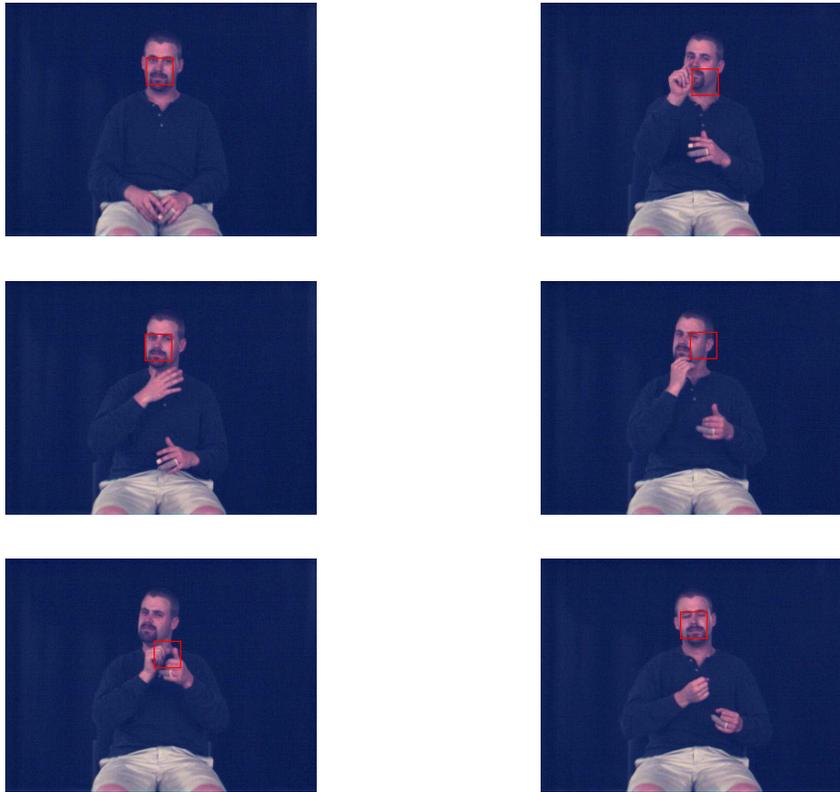
Όσον αφορά τις δυνατότητες του εν λόγω πλαισίου στην μοντελοποίηση των αντικειμένων, από τα παραπάνω πειράματα γίνεται ξεκάθαρο ότι υπάρχει μια ισχυρή σύνδεση της πολυπλοκότητας του αντικειμένου και του μεγέθους της απαιτούμενης βάσης εκπαίδευσης. Με το τέχνασμα να περιορίσουμε την περιοχή εστίασης του προσώπου και να ευθυγραμμίσουμε σωστά τα πρόσωπα, η μεταβλητότητα του προσώπου μειώθηκε, επιτυγχάνοντας πολύ καλά αποτελέσματα με μια μικρή βάση εκπαίδευσης. Για να πετύχουμε αντίστοιχα ποσοστά επιτυχίας εφαρμόζοντας το μοντέλο 1, το οποίο παρουσιάζει σημαντικά μεγαλύτερη πολυπλοκότητα, θα έπρεπε να χρησιμοποιηθεί μια πολύ μεγαλύτερη βάση εκπαίδευσης.

Παρόλα αυτά, ανεξάρτητα από το μέγεθος και την ποιότητα της βάσης δεδομένων εκπαίδευσης, το πλαίσιο ανίχνευσης αντικειμένων που παρουσιάστηκε σε αυτή την ενότητα αντιμετωπίζει πολύ συγκεκριμένους περιορισμούς. Αν το αντικείμενο που επιδιώκουμε να ανιχνεύσουμε παρουσιάζει μεγάλη μεταβλητότητα τόσο στο σχήμα και στην εμφάνιση (πχ πρόσωπα σε όλες τις πιθανές πόζες, ή χειρομορφές), τότε η παραπάνω μέθοδος είναι μάλλον ακατάλληλη. Διαφοροποιήσεις και εξελίξεις της παραπάνω μεθόδου, που επιτρέπουν την μοντελοποίηση και ανίχνευση πιο πολύπλοκων αντικειμένων, και αντικειμένων που παρουσιάζουν μεταβολές στο σχήμα αναφέρονται στην επόμενη ενότητα.

2.5 Αξιολόγηση της Μεθόδου και Εξελίξεις στην Ανίχνευση Προσώπων

Το framework για την ανίχνευση προσώπων (και γενικότερα αντικειμένων) που περιγράφεται σε αυτό το κεφάλαιο, και βασίζεται στον συνδυασμό του αλγορίθμου AdaBoost με τα Haar-like χαρακτηριστικά, θεωρείται ακόμη και σήμερα ίσως το πιο επιτυχημένο για την ανίχνευση, σε πραγματικό χρόνο, αντικειμένων με μικρή μεταβλητότητα στο σχήμα. Δεδομένων και των ποσοστών επιτυχίας που παρουσιάστηκαν στην προηγούμενη ενότητα, μπορούμε να ισχυριστούμε ότι το παραπάνω framework είναι το πλέον κατάλληλο για την ανίχνευση προσώπων σε βίντεο της νοηματικής γλώσσας. Εφόσον συνήθως η πόζα του ομιλητή δεν μεταβάλλεται έντονα (το πολύ μέχρι 45° προφίλ) ο αλγόριθμος AdaBoost παρέχει την δυνατότητα καλής μοντελοποίησης και κατηγοριοποίησης. Οι μόνες περιπτώσεις όπου η μέθοδος αποτυγχάνει να ανιχνεύσει σωστά το πρόσωπο, είναι εικόνες όπου παρουσιάζεται σημαντική επικάλυψη του προσώπου από τα χέρια. Αυτό το κώλυμα, δεν μπορεί να αντιμετωπιστεί με καλύτερο classifier παρά μόνο με μια στοχαστική μέθοδο συνεχούς ανίχνευσης (tracking), που αναπτύσσεται στο επόμενο κεφάλαιο.

Σε αυτή την ενότητα παρουσιάζονται σύντομα, για βιβλιογραφικούς λόγους, κάποιες ενδιαφέρουσες εξελίξεις του παραπάνω framework που έχουν προταθεί τα τελευταία χρόνια, και σκοπό έχουν την επιτάχυνση της διαδικασίας ή την καλύτερη μοντελοποίηση πιο πολύπλοκων αντικειμένων. Και οι δύο νέοι αλγόριθμοι που παρουσιάζονται βασίζονται στο παραπάνω framework, και συνδυάζουν τον αλγόριθμο AdaBoost με μια δενδρική δομή.



Σχήμα 2.10: Περιπτώσεις σωστών και λάθος ανιχνεύσεων σε μεμονωμένες εικόνες της νοηματικής

2.5.1 AdaTree

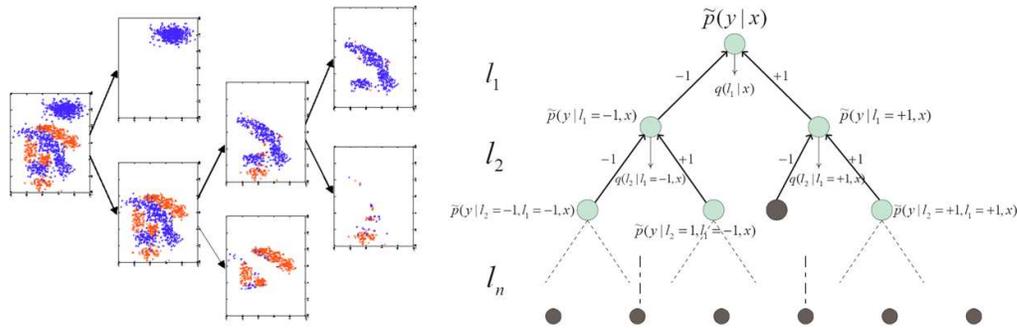
Το 2004 ο Etienne Grossman προτείνει τον αλγόριθμο AdaTree [11] που συνδυάζει τον AdaBoost με τα δένδρα αποφάσεων (Decision Trees). Αυτός ο αλγόριθμος βασίζεται στην ίδια λογική του συνδυασμού αδύναμων classifiers σε έναν ισχυρό, αλλά αυτή την φορά ο τελικός classifier έχει δενδρική δομή, με κάθε κόμβο του δένδρου να αποτελείται από έναν αδύναμο classifier. Κατά τη διαδικασία του classification το νέο πρότυπο περνάει από τον classifier που βρίσκεται στην ρίζα του δένδρου. Αν αυτός επιστρέψει θετική τιμή, τότε η διαδικασία συνεχίζεται με τον κόμβο στα δεξιά του, αλλιώς, με αυτόν στα αριστερά του. Στο τέλος οι τιμές των αδύναμων classifiers από τους οποίους πέρασε το άγνωστο πρότυπο αθροίζονται με κατάλληλους συντελεστές, και η τελική αυτή τιμή καθορίζει την απόφαση σχετικά με την κατηγορία στην οποία ανήκει το πρότυπο. Έτσι, σε αντίθεση με τα decision trees η κατηγορία στην οποία αντιστοιχίζεται το αντικείμενο δεν εξαρτάται μόνο από τον τελευταίο κόμβο του δένδρου, αλλά από το σταθμισμένο άθροισμα όλων των αδύναμων classifiers από τους οποίους

πέρασε το αντικείμενο.

Κατά την εκπαίδευση του αλγορίθμου ακολουθείται η ίδια λογική που επικρατεί και στον AdaBoost σχετικά με την ανανεώσιμη κατανομή πυκνότητας πιθανότητας του συνόλου δεδομένων εκπαίδευσης, αλλά σε κάθε κόμβο το σύνολο εκπαίδευσης περιορίζεται μόνο στα δεδομένα που θα κατέληγαν εκεί από τους classifiers των προηγούμενων επιπέδων. Με αυτόν τον τρόπο ο αλγόριθμος AdaTree διατηρεί την πολύ σημαντική ιδιότητα του AdaBoost σχετικά με το φραγμένο λάθος της τελικής υπόθεσης, ενώ ταυτόχρονα εισάγει το πλεονέκτημα της ταχύτητας που έχουν τα decision trees

Ο αλγόριθμος AdaTree έχει δοκιμαστεί για την περίπτωση της αναγνώρισης προσώπων και προσφέρει αντίστοιχα αποτελέσματα με τον AdaBoost σε μεγαλύτερες ταχύτητες, πράγμα που τον κάνει ακόμη πιο κατάλληλο για εφαρμογές πραγματικού χρόνου.

2.5.2 Probabilistic boosting tree



Σχήμα 2.11: Οπτικοποίηση της διαδικασίας κατηγοριοποίησης με την βοήθεια του αλγορίθμου Probabilistic Boosting Tree [24]

Ένας ακόμη συνδυασμός του αλγορίθμου AdaBoost με δενδρική δομή είναι ο αλγόριθμος Probabilistic Boosting Tree [24] που αναπτύχθηκε το 2005 στην Siemens Corporate Research και χρησιμοποιείται σήμερα με μεγάλη επιτυχία για την ανίχνευση αντικειμένων των οποίων το σχήμα και η εμφάνιση μπορούν να παρουσιάσουν σημαντικές μεταβολές. Ο αλγόριθμος αυτός και πάλι ακολουθεί τις βασικές αρχές του AdaBoost, αλλά αντί να συνδυάζει όλους του αδύναμους classifiers σε έναν ισχυρό μέσω ενός απλού σταθμισμένου αθροίσματος, ακολουθεί μια πιο πολύπλοκη 'διαίρει και βασίλευε' (divide and conquer) στρατηγική.

Κάθε κόμβους του δένδρου σε αυτή την περίπτωση αποτελείται από έναν ισχυρό classifier εκπαιδευμένο με την βοήθεια του αλγορίθμου AdaBoost. Η ρίζα του δένδρου εκπαιδεύεται με ολόκληρο το σύνολο εκπαίδευσης, ενώ σε κάθε βήμα, ο κόμβος στα δεξιά εκπαιδεύεται με τα δείγματα που κατηγοριοποιούνται ως θετικά, ενώ ο κόμβος στα αριστερά με αυτά που κατηγοριοποιούνται ως αρνητικά. Αυτή η διαδικασία μοιάζει σε μεγάλο βαθμό με τον αλγόριθμο AdaTree, αλλά η ουσιαστική διαφορά δεν έγκειται στην εκπαίδευση, αλλά στην διαδικασία της κατηγοριοποίησης.

Ο αλγόριθμος Probabilistic Boosting Tree υπολογίζει την ύστερη πιθανότητα $p(y|x)$ όπου y είναι η κατηγορία και x το άγνωστο δείγμα, με μια αναδρομική διαδικασία η οποία ανατρέχει όλους τους κόμβους του δένδρου. Σε κάθε κόμβο η πιθανότητα υπολογίζεται μέσω της αναδρομικής σχέσης:

$$p(y|x) = q(+1|x)p_{right}(y) + q(-1|x)p_{left}(y),$$

όπου $q(+1|x)$ είναι η τιμή που επιστρέφει ο classifier του συγκεκριμένου κόμβου για την αντίστοιχη κατηγορία και p_{right} (p_{left}) είναι η ύστερη πιθανότητα του στα δεξιά (αριστερά αντίστοιχα), η οποία υπολογίζεται με την ίδια διαδικασία. Έτσι ξεκινώντας από την ρίζα, ανατρέχεται κάθε κόμβος του δένδρου, και υπολογίζεται η συνολική ύστερη πιθανότητα της κλάσης στην οποία ανήκει το άγνωστο δείγμα. Μια οπτική αναπαράσταση της στρατηγικής κατηγοριοποίησης που εφαρμόζει ο εν λόγω αλγόριθμος παρουσιάζεται στο Σχήμα 2.11.

Όπως αναφέρεται και παραπάνω, ο αλγόριθμος αυτός έχει την ικανότητα να μοντελοποιεί καλά αντικείμενα τα οποία παρουσιάζουν σημαντικές διαφοροποιήσεις στο σχήμα ή την εμφάνιση. Μία από τις χαρακτηριστικές εφαρμογές του αλγορίθμου Probabilistic Boosting Tree είναι η ανίχνευση προσώπων για όλες τις πόζες, πρόβλημα στο οποίο ο κλασικός αλγόριθμος AdaBoost αποτυγχάνει. Το μειονέκτημά του είναι ότι είναι συνήθως πιο αργός από τον AdaBoost δεδομένων όλων των κόμβων - classifiers από τους οποίους πρέπει να περάσει το άγνωστο πρότυπο πριν υπολογιστεί η ύστερη πιθανότητα σχετικά με την κλάση στην οποία ανήκει.

Κεφάλαιο 3

Συνεχής Ανίχνευση σε Ακολουθίες Βίντεο της Νοηματικής

3.1 Εισαγωγή

Στο Κεφάλαιο 2 έγινε μια εκτενής μελέτη και υλοποίηση ενός ανιχνευτή προσώπων, ο οποίος εφαρμόστηκε σε μεμονωμένες εικόνες της νοηματικής με αρκετά καλά αποτελέσματα. Παρ' όλα αυτά το πλαίσιο που αναπτύχθηκε αντιμετωπίζει πολύ συγκεκριμένους περιορισμούς, και αποτυγχάνει σε περιπτώσεις όπου το πρόσωπο επικαλύπτεται σημαντικά από τα χέρια του νοηματιστή. Σε αυτό το κεφάλαιο επιδιώκεται να αντιμετωπιστεί ο παραπάνω περιορισμός, εφαρμόζοντας συνεχή ανίχνευση (tracking) στις ακολουθίες βίντεο.

Σε μια ακολουθία διαδοχικών εικόνων (frames) μπορούμε να χρησιμοποιήσουμε την πρότερη πληροφορία σχετικά με την προηγούμενη σωστή ανίχνευση ώστε να επιτύχουμε κάποια πρόβλεψη για την επόμενη. Συνδυάζοντας αυτή την πρόβλεψη με την νέα ανίχνευση, μπορούμε να πάρουμε πιο ακριβή αποτελέσματα, ακόμη και σε περιπτώσεις όπου το πρόσωπο επικαλύπτεται σημαντικά ή είναι δύσκολο να ανιχνευτεί. Υιοθετώντας αυτή την πορεία, αναφερόμαστε πλέον σε ένα στοχαστικό μοντέλο όπου η πρόβλεψη όπως και η μέτρηση δεν είναι απόλυτες τιμές αλλά εκφράζονται από κάποια συνάρτηση πυκνότητας πιθανότητας πάνω στον δισδιάστατο χώρο.

Στο μαθηματικό μοντέλο που θα περιγραφεί παρακάτω κυριαρχούν τρεις στοχαστικές μεταβλητές: Η $x(k|k)$ που αντιπροσωπεύει την εσωτερική κατάσταση του συστήματος (πραγματική θέση του προσώπου) την χρονική στιγμή k , η $x(k|k-1)$ που εκφράζει την πρόβλεψη σχετικά με την χρονική στιγμή k , και η $z(k)$ που αντιπροσωπεύει την μέτρηση σχετικά με το που βρίσκεται το πρόσωπο (τιμή που επιστρέφει ο ανιχνευτής του Κεφαλαίου 2). Οι μεταβλητές αυτές μοντελοποιούνται ως χωρικές συναρτήσεις πυκνότητας πιθανότητας, οι οποίες μεταβάλλονται χρονικά. Στόχος του κεφαλαίου είναι η μοντελοποίηση και ο συνδυασμός αυτών των στοχαστικών μεταβλητών, με σκοπό την στοχαστική εκτίμηση της εσωτερικής κατάστασης, ώστε να βελτιωθεί ποιοτικά και χρονικά η ανίχνευση.

3.2 Το πρόβλημα Observer Design

Το μαθηματικό μοντέλο που υιοθετείται για την εκτίμηση της εσωτερικής κατάστασης του συστήματος (θέση του προσώπου) βάση κάποιας μέτρησης και κάποιας πρόβλεψης, είναι το μοντέλο State Observer της Θεωρίας Ελέγχου [28]. Η βασική παραδοχή που γίνεται είναι ότι το σύστημα είναι γραμμικό, δηλαδή η εσωτερική κατάσταση x_k εξαρτάται γραμμικά τόσο από την προηγούμενη εσωτερική κατάσταση όσο και από την μέτρηση y_k . Υπό αυτή την παραδοχή, οι εξισώσεις του συστήματος είναι οι 3.1 και 3.2 που παρουσιάζονται και αναλύονται παρακάτω.

Το νέο στοχαστικό μοντέλο χαρακτηρίζεται από μία εσωτερική κατάσταση (στοχαστική μεταβλητή x_k) αόρατη στον εξωτερικό κόσμο, την οποία προσπαθούμε να εκτιμήσουμε με την βοήθεια δύο εξισώσεων, την εξίσωση 3.1 (Process Model) και την εξίσωση 3.2 (Measurement Model). Ουσιαστικά αυτό που επιδιώκουμε να εκτιμήσουμε, είναι η συνάρτηση πυκνότητας πιθανότητας της εσωτερικής κατάστασης, δεδομένων όλων των μέχρι τώρα μετρήσεων: $P(x_k|z_{1,2,\dots,k})$.

$$x_k = Ax_{k-1} + Bu_k + w_{k-1} \quad (3.1)$$

$$y_k = Hx_k + v_k \quad (3.2)$$

Στο μοντέλο πρόβλεψης (Process Model), η εσωτερική κατάσταση x_k εκτιμάται με βάση την προηγούμενη εσωτερική κατάσταση x_{k-1} και κάποιες εισόδους στο σύστημα u_k (όπως πχ μια εξωτερική ταχύτητα ή δύναμη που ασκείται στο σύστημα), ενώ σε αυτήν προσθέτεται πάντα και κάποια αβεβαιότητα (τυχαία μεταβλητή θορύβου πρόβλεψης w_{k-1})

Στο μοντέλο μέτρησης (Measurement Model), η εξωτερική μέτρηση που παίρνουμε από τον ανιχνευτή προσώπων y_k εκφράζεται συναρτήσει της πραγματικής εσωτερικής κατάστασης x_k και κάποιας τυχαίας μεταβλητής θορύβου (μηδενικής μέσης τιμής) v_k που αντιπροσωπεύει το λάθος μέτρησης.

3.3 Γενικευμένη Λύση

Στην γενική περίπτωση, θεωρητικά βέλτιστη λύση σε αυτό το πρόβλημα δίνει το φίλτρο Bayes (Recursive Bayesian Estimator). Το φίλτρο Bayes είναι μια επαναληπτική διαδικασία, η οποία εκτιμά την εσωτερική κατάσταση μέσω δύο βημάτων (predict & update) σε κάθε επανάληψη. Το βήμα πρόβλεψης (predict) (Εξίσωση 3.3) υπολογίζει μια πρότερη συνάρτηση πυκνότητας πιθανότητας για την μεταβλητή x_k δεδομένων των μέχρι εδώ μετρήσεων: $P(x_k|z_{1,2,\dots,k-1})$, ενώ το βήμα διόρθωσης (update) (Εξίσωση 3.4) βρίσκει την ύστερη συνάρτηση πυκνότητας πιθανότητας (pdf) $P(x_k|z_{1,2,\dots,k})$ συναρτήσει της πρότερης συνάρτησης και της μέτρησης $P(z_k|x_k)$:

$$P(x_k|Z_{k-1}) = \int p(x_k|x_{k-1})p(x_{k-1}|Z_{k-1}) \quad (3.3)$$

$$P(x_k|Z_k) = \frac{p(z_k|x_k)p(x_k|Z_{k-1})}{p(z_k|Z_{k-1})} \quad (3.4)$$

Εαν οι συναρτήσεις πυκνότητας πιθανότητας τόσο της αρχικής εσωτερικής κατάστασης x_0 όσο και των σφαλμάτων μέτρησης και πρόβλεψης μπορούν να μοντελοποιηθούν ως πολυμεταβλητές γκαουσιανές, τότε το φίλτρο Bayes εκφυλίζεται ουσιαστικά στο φίλτρο Kalman, το οποίο δίνει τότε την βέλτιστη λύση. Αν μια τέτοια μοντελοποίηση δεν μπορεί να στηριχθεί, τότε το Kalman φίλτρο δεν θα έδινε καλά αποτελέσματά. Σε τέτοιες περιπτώσεις χρησιμοποιείται συνήθως η μέθοδος Particle filters (Condensation Algorithm) [13] η οποία χρησιμοποιεί διακριτή συνάρτηση πυκνότητας πιθανότητας για την μοντελοποίηση της εσωτερικής κατάστασης.

3.4 Επιλογή του Φίλτρου Kalman

3.4.1 Υπόθεση

Σε αυτή την ενότητα ερευνάται και δικαιολογείται η υπόθεση σχετικά με την βελτιστότητα του φίλτρου Kalman για την συνεχή ανίχνευση του προσώπου στις ακολουθίες βίντεο. Όπως εξηγήθηκε και προηγουμένως, για να στηριχθεί μια τέτοια υπόθεση πρέπει να δοθούν ενδείξεις, ότι όλοι οι παράμετροι του μοντέλου State Observer μπορούν να μοντελοποιηθούν επιτυχώς με δισδιάστατες χωρικές γκαουσιανές συναρτήσεις πυκνότητας πιθανότητας. Σε αυτήν την ενότητα προσομοιώνεται, με την βοήθεια του ανιχνευτή προσώπου, η χωρική συνάρτηση πυκνότητας πιθανότητας της μέτρησης και εκτιμώνται οι βασικές παράμετροι της.

3.4.2 Προσομοίωση

Όπως προκύπτει από την θεωρία σχετικά με τον αλγόριθμο AdaBoost, ο ανιχνευτής προσώπου που περιγράφηκε στην προηγούμενη ενότητα δίνει μια καλή εκτίμηση σχετικά με την πιθανότητα ύπαρξης ή μη προσώπου στο κομμάτι της εικόνας που εφαρμόζεται. Εφαρμόζοντας λοιπόν τον ανιχνευτή σε κάθε σημείο μιας περιοχής της εικόνας κοντά στο πρόσωπο (Σχήμα 3.1), και αποτυπώνοντας την τιμή (πιθανότητα) που αυτός δίνει, μπορούμε να πάρουμε μια εκτίμηση σχετικά με την χωρική συνάρτηση πυκνότητας πιθανότητας για την ύπαρξη ή μη προσώπου. Το αποτέλεσμα φαίνεται στο Σχήμα 3.2 αριστερά. Βλέπουμε ότι σε μια περιοχή κοντά στο πρόσωπο ο ανιχνευτής προσώπου δίνει αποτελέσματα τα οποία μοιάζουν σε μεγάλο βαθμό με μια δισδιάστατη γκαουσιανή. Μάλιστα κινούμενοι κατά μήκος των αξόνων x και y μπορούμε να κάνουμε και μια αρχική εκτίμηση της τυπικής απόκλισης αυτής της γκαουσιανής σε κάθε άξονα (σ_x , σ_y pixels) η οποία εκφράζεται σε pixels και εξαρτάται φυσικά τόσο από την κλίμακα της εικόνας όσο και από το μέγεθος του προσώπου (για παράδειγμα στην παρούσα εφαρμογή είναι $\sigma_x = 6.3$, $\sigma_y = 6.9$ pixels).

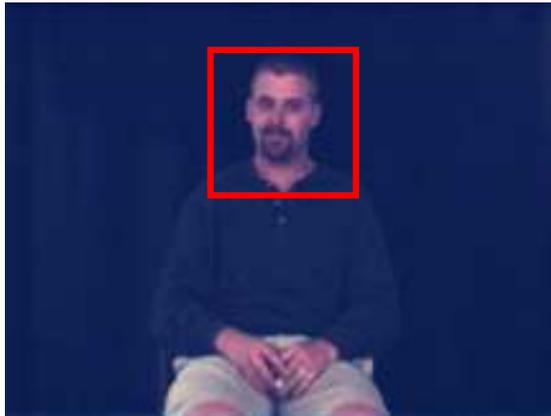
Με αυτό τον τρόπο κατασκευάζεται μια δισδιάστατη γκαουσιανή συνάρτηση (Σχήμα 3.2 δεξιά) η οποία μοντελοποιεί την χωρική συνάρτηση πυκνότητας πιθανότητας της μέτρησης που προκύπτει από τον ανιχνευτή προσώπων. Όπως γίνεται οπτικά φανερό από το Σχήμα 3.2,

οι δύο δισδιάστατες κατανομές δεν διαφέρουν και πολύ σε περιοχές κοντά στο πρόσωπο, και επομένως η παραπάνω μοντελοποίηση δεν απέχει πολύ από την πραγματικότητα.

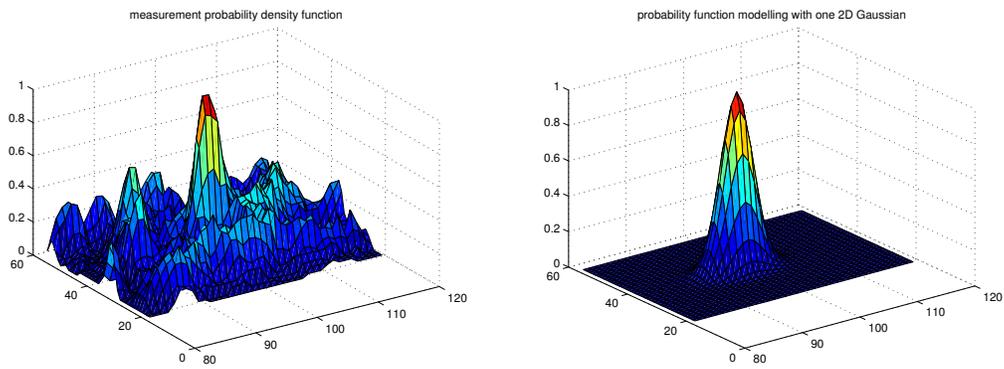
Στο Σχήμα 3.3 παρουσιάζονται τα αποτελέσματα του ίδιου πειράματος, όπου ο ανιχνευτής προσώπου έχει εκπαιδευτεί με την βάση εκπαίδευσης του μοντέλου 1 (βλ. ενότητα 2.4.2). Και σε αυτή την περίπτωση η μέτρηση προσεγγίζει μια δισδιάστατη γκαουσιανή με μεγαλύτερο όμως εύρος (μεγαλύτερη αβεβαιότητα) ιδιαίτερα στον άξονα των x . Αυτή η παρατήρηση επιβεβαιώνει και πάλι την υπεροχή του μοντέλου 3 της βάσης εκπαίδευσης, όσον αφορά την ακρίβεια των αποτελεσμάτων.

3.4.3 Συμπεράσματα και επιπλέον παραδοχές

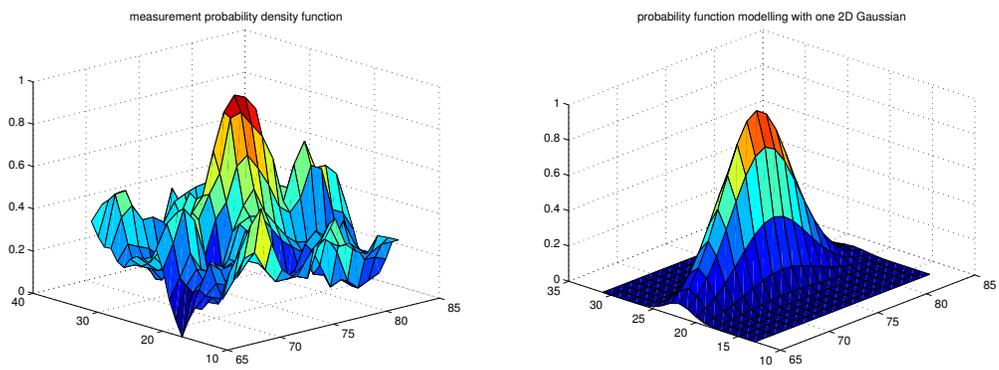
Το παραπάνω πείραμα δικαιολογεί την μοντελοποίηση τόσο του σφάλματος μέτρησης όσο και της αρχικής εσωτερικής κατάστασης με δισδιάστατες χωρικές γκαουσιανές συναρτήσεις πυκνότητας πιθανότητας, για τις οποίες έγινε επίσης εκτίμηση των παραμέτρων μέσης τιμής και τυπικής απόκλισης. Επομένως κάνοντας την επιπλέον παραδοχή ότι και το σφάλμα πρόβλεψης θα έχει την μορφή δισδιάστατης γκαουσιανής (οι παράμετροι της οποίας θα εκτιμηθούν πειραματικά) μπορεί να υιοθετηθεί η εκδοχή του φίλτρου Kalman ως η βέλτιστη για την συνεχή ανίχνευση του προσώπου στις ακολουθίες βίντεο.



Σχήμα 3.1: Περιοχή προσομοίωσης



Σχήμα 3.2: Οπτική αναπαράσταση της δισδιάστατης ανίχνευσης, και μοντελοποίηση με γκαουσιανή



Σχήμα 3.3: Πειραματισμός με το Μοντέλο 1 της βάσης εκπαίδευσης

3.5 Σχεδιασμός και Υλοποίησης ενός Φίλτρου Kalman

3.5.1 Φίλτρο Kalman

Στην προηγούμενη ενότητα δικαιολογήθηκε η γκαουσιανή μοντελοποίηση της μέτρησης που υπαγορεύει την επιλογή του φίλτρου Kalman, ως τη βέλτιστη μέθοδο για τον συνδυασμό (fusion) της μέτρησης με την πρόβλεψη. Το φίλτρο Kalman [3] ως μια υποπερίπτωση του φίλτρου Bayes, είναι και αυτό ένας αναδρομικός αλγόριθμος δύο βημάτων (πρόβλεψης και ανανέωσης) σε κάθε χρονική στιγμή. Στο βήμα πρόβλεψης εκτιμάται η εσωτερική κατάσταση της χρονικής στιγμής t σαν συνάρτηση της εσωτερικής κατάστασης στην $t - 1$, ενώ το βήμα ανανέωσης, η εσωτερική κατάσταση ανανεώνεται με βάση την μέτρηση της χρονικής στιγμής t . Οι εξισώσεις του φίλτρου Kalman σε αυτά τα δύο βήματα φαίνονται παρακάτω:

Time Update

$$x_{t|t-1} = Ax_{t-1|t-1} + Bu_t \quad (3.5)$$

$$P_{t|t-1} = AP_{t-1}A^T + Q \quad (3.6)$$

Measurement Update

$$K_t = P_{t|t-1}H^T(HP_{t|t-1}H^T + R)^{-1} \quad (3.7)$$

$$x_{t|t} = x_{t|t-1} + K_t(y_t - Hx_{t|t-1}) \quad (3.8)$$

$$P_{t|t} = P_{t|t-1} - K_tHP_{t|t-1} \quad (3.9)$$

$x_{t|t}$: είναι η μέση τιμή της συνάρτησης πυκνότητας πιθανότητας (pdf) της εσωτερικής κατάστασης την στιγμή t .

$x_{t|t-1}$ είναι η εκτίμηση της μέση τιμή της συνάρτησης πυκνότητας πιθανότητας της εσωτερικής κατάστασης την στιγμή t μέσω της πρόβλεψης.

$P_{t|t}$: είναι η μεταβλητότητα (variance) της συνάρτησης πυκνότητας πιθανότητας της εσωτερικής κατάστασης την στιγμή t .

$P_{t|t-1}$: είναι η εκτίμηση της μεταβλητότητας της συνάρτησης πυκνότητας πιθανότητας της εσωτερικής κατάστασης την στιγμή t μέσω της πρόβλεψης.

y_t : η μέση τιμή της χωρικής pdf της μέτρησης (εδώ η θέση που επιστρέφει ο ανιχνευτής προσώπου).

u_t : η είσοδος του συστήματος στο βήμα της πρόβλεψης (μπορεί να αναφέρεται σε εξωτερική ταχύτητα ή δυναμική).

A : εδώ ο μοναδιαίος πίνακας I.

H : ο πίνακας που συνδέει την μέτρηση με την εσωτερική κατάσταση.

B : ο πίνακας με τον οποίον θα πολλαπλασιαστεί η είσοδος (εδώ ο μοναδιαίος πίνακας I).

Q : Η μεταβλητότητα της χωρικής pdf της πρόβλεψης.

R : Η μεταβλητότητα της χωρικής pdf της μέτρησης.

3.5.2 Σχεδιαστικές αποφάσεις για την υλοποίηση του φίλτρου Kalman

Για να υλοποιηθεί το φίλτρο Kalman και να ενταχθεί επιτυχώς στο υπάρχον πλαίσιο ανίχνευσης προσώπου, πρέπει να παρθούν κάποιες σχεδιαστικές αποφάσεις και να εκτιμηθούν οι συντελεστές των εξισώσεων που περιγράφηκαν στην προηγούμενη ενότητα. Σε αυτή την ενότητα αναλύονται κάποιες παραδοχές που έγιναν σχετικά με τις μεταβλητές της αναδρομικής διαδικασίας, καθώς και οι τιμές των συντελεστών που επιλέχθηκαν ως οι καταλληλότερες για την υλοποίηση του φίλτρου:

Εσωτερική κατάσταση (x, P)

Εφόσον δεν διαθέτουμε καμία πληροφορία για την μεταβλητότητα της ταχύτητας, ούτε έχουμε κάποια μέτρηση της ταχύτητας, είναι πιο σωστό να μην την συμπεριλάβουμε στην εσωτερική κατάσταση (x, P) . Εξάλλου η κίνηση του κεφαλιού είναι σχετικά μικρή και σπασμωδική, γεγονός που υποδεικνύει ένα απλό κινητικό (σταθερή ταχύτητα ως είσοδος στην πρόβλεψη) παρά κάποιο δυναμικό μοντέλο (επιτάχυνση ως είσοδος). Επομένως στην εσωτερική κατάσταση συμπεριλαμβάνεται μόνο η δισδιάστατη θέση της ανίχνευσης. Η εσωτερική κατάσταση αρχικοποιείται με την θέση (x) της ανίχνευσης στην πρώτη εικόνα της ακολουθίας βίντεο, και τον πίνακα P να είναι διαγώνιος, με τα στοιχεία της διαγωνίου να είναι τα σ_x^2 και σ_y^2 όπως αυτά υπολογίστηκαν από την προσομοίωση της προηγούμενης ενότητας.

Μέτρηση y

Η μέτρηση y είναι φυσικά η θέση που επιστρέφει ο ανιχνευτής προσώπων, και επομένως ο πίνακας H είναι ο μοναδιαίος 2×2

Είσοδος στο σύστημα u

Εφόσον η εσωτερική κατάσταση περιλαμβάνει μόνο την θέση του αντικειμένου, η είσοδος στο βήμα της πρόβλεψης αντιπροσωπεύει κάποια εκτίμηση σχετικά με το πως θα κινηθεί το αντικείμενο την επόμενη χρονική στιγμή (ταχύτητα). Σχετικά με αυτή την μεταβλητή, δοκιμάστηκαν τρία κινητικά μοντέλα τα οποία περιγράφονται αναλυτικά στην επόμενη ενότητα.

Μεταβλητότητα του θορύβου μέτρησης R

Όσον αφορά την μεταβλητότητα της μέτρησης, η προσομοίωση που επιχειρήθηκε στην προηγούμενη ενότητα δίνει κάποια ενόραση σχετικά με την τιμή (σε pixels) που μπορεί να πάρει. Παρόλα αυτά, ο θόρυβος αυτός διαφέρει από εικόνα σε εικόνα, και είναι λογικό να εξαρτάται από την τιμή που επιστρέφει ο ανιχνευτής προσώπων σε κάθε περίπτωση. Για αυτόν τον λόγο υιοθετείται ένα μοντέλο όπου η μεταβλητότητα της μέτρησης είναι πάντα ένας

διαγώνιος πίνακας με στοιχεία ($\sigma_x^2/detection\ value$) και ($\sigma_y^2/detection\ value$) όπου *detection value* είναι η τιμή (από 0 μέχρι 1) που επιστρέφει ο ανιχνευτής προσώπων, και σ_x , σ_y οι τιμές της προσομοίωσης. Με βάση το παραπάνω μοντέλο, η μεταβλητότητα είναι μικρή όταν η ανίχνευση επιστρέφει μια μεγάλη τιμή, αλλά μεγαλώνει πολύ αν η ανίχνευση είναι εντελώς λάθος. Με αυτό τον τρόπο, σε τέτοιες περιπτώσεις, δίνεται μεγαλύτερη εμπιστοσύνη στην πρόβλεψη, παρά στην μέτρηση.

Μεταβλητότητα του θορύβου στο βήμα πρόβλεψης Q

Ο θόρυβος του βήματος πρόβλεψης θεωρείται γκαουσιανός με μεταβλητότητα επίσης διαγώνιο πίνακα. Οι τιμές των στοιχείων της διαγώνιου εξαρτώνται από το κινητικό μοντέλο που υιοθετείται (βλ. επόμενη ενότητα), και υπολογίζονται πειραματικά.

Ενσωμάτωση του Φίλτρου στο υπάρχων πλαίσιο ανίχνευσης

Με βάση τα παραπάνω και με τα κινητικά μοντέλα που θα αναλυθούν στην επόμενη παράγραφο είναι δυνατή η ενσωμάτωση της πρόβλεψης και του φίλτρου Kalman στο υπάρχον πλαίσιο ανίχνευσης προσώπου. Πριν την ανίχνευση σε κάθε εικόνα, προηγείται το στάδιο της πρόβλεψης, όπου με βάση την εσωτερική κατάσταση της προηγούμενης χρονικής στιγμής, υπολογίζεται η πιθανότερη θέση της επόμενης σαν μια συνάρτηση πυκνότητας πιθανότητας. Εφόσον υπάρχει εμπιστοσύνη στην πρόβλεψη, η ανίχνευση επιχειρείται πλέον μόνο στην περιοχή όπου η πιθανότητα, βάσει της πρόβλεψης, είναι μη αμελητέα. Πιο συγκεκριμένα, επειδή σχεδόν το σύνολο της πιθανότητας περιορίζεται στην περιοχή μεταξύ $x + 3\sigma$ και $x - 3\sigma$, αυτή θα είναι και η περιοχή όπου θα επιχειρηθεί ανίχνευση στην επόμενη εικόνα. Με αυτό τον τρόπο επιτυγχάνεται σημαντική επιτάχυνση της συνολικής διαδικασίας,

3.5.3 Κινητικά μοντέλα

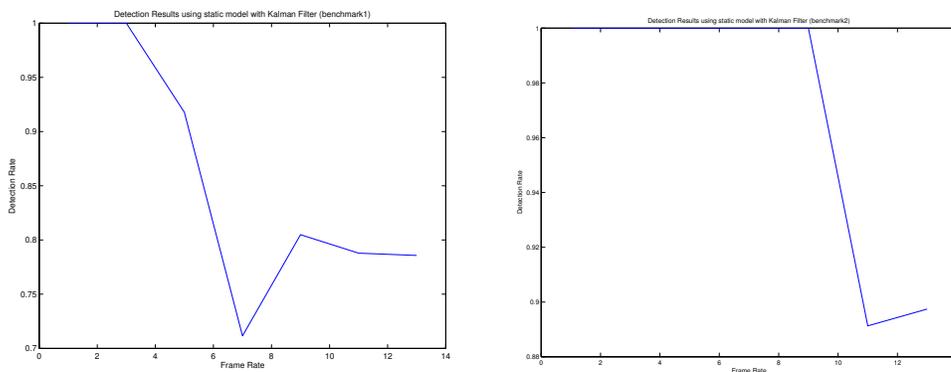
Όσον αφορά το στάδιο της πρόβλεψης, υλοποιήθηκαν και δοκιμάστηκαν 3 διαφορετικά κινητικά μοντέλα. Δεδομένου ότι η κίνηση του προσώπου είναι σχετικά αργή, το πρώτο μοντέλο υποθέτει ακινησία, προβλέποντας την επόμενη θέση του προσώπου ακριβώς στο σημείο της προηγούμενης. Στην συνέχεια προτείνονται δύο πιο σύνθετα μοντέλα, τα οποία υποθέτουν διατήρηση της υπάρχουσας κίνησης, ενώ γίνεται πειραματισμός με τις παραμέτρους τους ώστε να αποφευχθεί η αστάθεια του συστήματος.

Για την αξιολόγηση των τριών αυτών κινητικών μοντέλων, και συνολικά του πλαισίου συνεχούς ανίχνευσης που αναπτύσσεται σε αυτή την ενότητα, χρησιμοποιούνται τα δύο μειωμένα benchmarks ακολουθιών βίντεο που χρησιμοποιήθηκαν και στο προηγούμενο κεφάλαιο για την αξιολόγηση της ανίχνευσης. Όπως αναφέρθηκε και προηγουμένως, το πρόσωπο σε αυτά τα βίντεο κινείται σχετικά αργά. Για να προσομοιωθούν πιο γρήγορες κινήσεις και να κατασκευαστεί ένα καλό φίλτρο για την γενική περίπτωση, κατασκευάζονται πιο αραιές ακολουθίες, χρησιμοποιώντας ένα ανά n (όπου η παράμετρος n κυμαίνεται από 1 μέχρι 13) διαδοχικά frames των ακολουθιών βίντεο. Το φίλτρο που κατασκευάζεται με κάθε ένα από τα παρακάτω κινητικά μοντέλα εφαρμόζεται πάνω σε όλες αυτές τις ακολουθίες και αξιολογείται

με αυτόν τον τρόπο σε ένα μεγάλο φάσμα ταχυτήτων. Σε κάθε μία από τις παρακάτω υποενότητες αναλύονται οι σχεδιαστικές αποφάσεις που αφορούν το εκάστοτε κινητικό μοντέλο, και παρουσιάζονται τα ποσοστά ανίχνευσης που επιτυγχάνονται υπό τις διαφορετικές ταχύτητες κίνησης του προσώπου.

Ακίνητο μοντέλο

Σε αυτό το κινητικό μοντέλο η πρόβλεψη σχετικά με την επόμενη θέση του προσώπου είναι ακριβώς η θέση του στην προηγούμενη εικόνα της ακολουθίας βίντεο. Μετά από πειραματισμούς πάνω στην αβεβαιότητα της πρόβλεψης (θόρυβος w), επιλέχθηκε η τιμή των 6 pixels και ως προς τους δύο άξονες ως κατάλληλη για την τυπική απόκλιση. Όταν αυτό το μοντέλο εφαρμοστεί και στα 2 benchmarks δίνει 100% σωστά αποτελέσματα, επιτυγχάνοντας σωστή ανίχνευση ακόμη και στις περιπτώσεις όπου το πρόσωπο επικαλύπτεται πλήρως από τα χέρια. Παρ' όλα αυτά, δοκιμάζοντας το παραπάνω μοντέλο σε μεγαλύτερες ταχύτητες τα αποτελέσματα είναι αρκετά χειρότερα, καθώς η ανίχνευση αδυνατεί να ακολουθήσει την ταχύτερη κίνηση του προσώπου. Τα ποσοστά επιτυχίας για όλες τις ταχύτητες παρουσιάζονται στο Σχήμα 3.4.

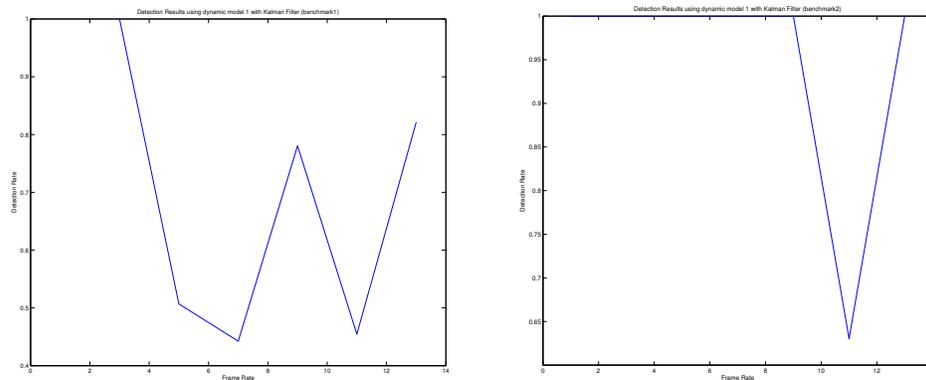


Σχήμα 3.4: Ποσοστά ανίχνευσης υπο αυξανόμενες προσημειωμένες ταχύτητες στο ακίνητο μοντέλο, στα δύο benchmarks

Απλό κινητικό μοντέλο

Για την βελτίωση των αποτελεσμάτων σε περιπτώσεις όπου η κίνηση του προσώπου είναι γρήγορη, δοκιμάστηκε ένα κινητικό μοντέλο που υποθέτει ότι το αντικείμενο προς ανίχνευση θα συνεχίσει να κινείται προς την ίδια κατεύθυνση, και με την ίδια ταχύτητά. Η τυπική απόκλιση της πρόβλεψης (αβεβαιότητα) αρχικά διατηρείται στα 6 pixels όπου το μοντέλο έδινε καλά αποτελέσματα για μικρές ταχύτητες. Από το διάγραμμα ποσοστών ανίχνευσης (Σχήμα 3.5) φαίνεται ότι και αυτό το μοντέλο αποτυγχάνει σε μεγάλες ταχύτητες, για διαφορετικούς λόγους αυτή την φορά. Τώρα το σύστημα γίνεται ασταθές, με την έννοια ότι όταν η ταχύτητα

γίνει μεγάλη, η ανίχνευση χάνει το πρόσωπο και μετά ταλαντώνεται γύρω του χωρίς να μπορεί να σταθεροποιηθεί ξανά. Κάνοντας πειράματα με μεγαλύτερες τιμές της τυπικής απόκλισης της πρόβλεψης, το φαινόμενο αυτό γίνεται λιγότερο σύννηθες αλλά συνεχίζει να υπάρχει, ενώ παράλληλα τα αποτελέσματα γίνονται χειρότερα στην περίπτωση των μικρών ταχυτήτων. Το γεγονός αυτό υποδεικνύει ότι η αβεβαιότητα της πρόβλεψης δεν πρέπει να είναι σταθερή, αλλά πρέπει να εξαρτάται από το μέτρο της ταχύτητας. Έτσι καταλήγουμε στο τελευταίο, πιο πολύπλοκο κινητικό μοντέλο που παρουσιάζεται στην επόμενη ενότητα.



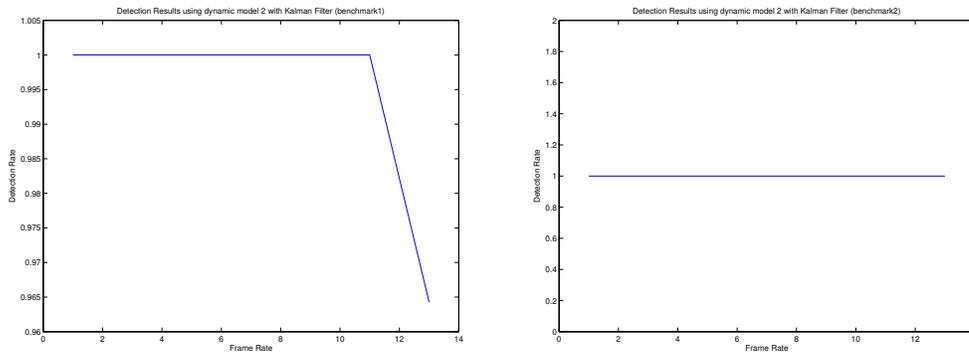
Σχήμα 3.5: Ποσοστά ανίχνευσης υπο αυξανόμενες προσομοιωμένες ταχύτητες στο απλό κινητικό μοντέλο, στα δύο benchmarks

Σύνθετο κινητικό μοντέλο

Το τελευταίο κινητικό μοντέλο που δοκιμάζεται αποτελεί έναν συνδυασμό των δύο προηγούμενων. Επιδιώκει να εκμεταλλευτεί τα πολύ καλά αποτελέσματα του στατικού μοντέλου για μικρές ταχύτητες, και να τα γενικεύσει στις μεγάλες. Και αυτό το μοντέλο υποθέτει την διατήρηση της κίνησης, αλλά υιοθετεί μια πιο μετριοπαθή πρόβλεψη για την ταχύτητα, η οποία είναι τώρα ένα μέρος μόνο (εδώ το ένα τρίτο) της προηγούμενης. Με αυτόν τον τρόπο η πρόβλεψη δεν θα απομακρύνεται επικίνδυνα σε περιπτώσεις απότομης εναλλαγής της κίνησης.

Μία επιπλέον εξέλιξη σε σχέση με το προηγούμενο κινητικό μοντέλο είναι ότι πλέον η αβεβαιότητα της πρόβλεψης δεν είναι μια σταθερά αλλά εξαρτάται από το μέτρο της προβλεπόμενης ταχύτητας. Πιο συγκεκριμένα οι τυπικές αποκλίσεις και ως προς τους δύο άξονες θα είναι ίσες με μια σταθερά (6 pixels) συν το τριπλάσιο της συνιστώσας της ταχύτητας προς αυτή την κατεύθυνση. Έτσι ακόμη και αν η υπόθεση είναι λάθος, και το πρόσωπο κινηθεί προς την αντίθετη κατεύθυνση με την ίδια ταχύτητα, πάλι θα βρίσκεται μέσα στα όρια της αβεβαιότητας.

Αυτό το κινητικό μοντέλο σημειώνει 100% επιτυχία σχεδόν σε ολόκληρο το φάσμα των ταχυτήτων (Σχήμα 3.6). Το γεγονός ότι έχουμε μεγάλη επιτυχία ανεξαρτήτως ταχύτητας, καθιστά το μοντέλο αξιόπιστο στην γενική περίπτωση.



Σχήμα 3.6: Ποσοστά ανίχνευσης υπο αυξανόμενες προσομοιωμένες ταχύτητες στο σύνθετο κινητικό μοντέλο, στα δύο benchmarks

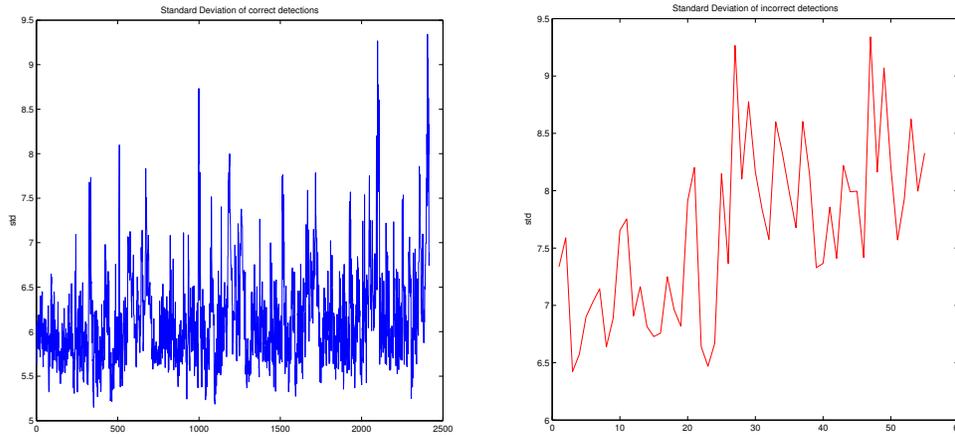
3.6 Απόδοση Εμπιστοσύνης στην Ανίχνευση

Το πλαίσιο για απλή και συνεχή ανίχνευση προσώπου που αναπτύχθηκε σε αυτό και στο προηγούμενο κεφάλαιο, επιστρέφει πάντα ως αποτέλεσμα μία περιοχή της εικόνας, η οποία αποτελεί το πιο πιθανό πρόσωπο, ακόμη και αν το αποτέλεσμα αυτό είναι αρκετά αβέβαιο. Το γεγονός αυτό μπορεί να αποτελέσει πρόβλημα σε πιο πολύπλοκες εφαρμογές που στηρίζονται στην σωστή ανίχνευση του προσώπου για περαιτέρω ανάλυση, όπως για παράδειγμα η ανίχνευση της πόζας (βλ. επόμενο κεφάλαιο). Σε τέτοιες εφαρμογές είναι ίσως προτιμότερο να μην επιστραφεί κανένα αποτέλεσμα (και να χρησιμοποιηθεί απλά πρότερη γνώση), από το να δοθεί μία εντελώς λάθος ανίχνευση. Για αυτό τον λόγο, σε αυτή την ενότητα επιχειρείται η απόδοση εμπιστοσύνης στο αποτέλεσμα της ανίχνευσης, ώστε να μπορούν να απορριφθούν αποτελέσματα που πιθανώς να είναι λανθασμένα.

Οι μεταβλητές που συνδέονται με την ποιότητα της ανίχνευσης είναι η τιμή που επιστρέφει ο ανιχνευτής προσώπων (*detection value*) που περιγράφηκε στο πρώτο κεφάλαιο, και η μεταβλητότητα της εσωτερικής κατάστασης (P) που ορίστηκε κατά την στοχαστική ανάλυση αυτού του κεφαλαίου. Από την τιμή μίας εκ των δύο αυτών μεταβλητών μπορούμε να πάρουμε κάποια ενόραση σχετικά με την εμπιστοσύνη που αποδίδεται στην ανίχνευση.

Για την παραγωγή στατιστικών στοιχείων σχετικά με τις τιμές των δύο αυτών μεταβλητών σε περιπτώσεις σωστών και λάθος ανιχνεύσεων κατασκευάζεται ένα ακόμη πιο δύσκολο benchmark ακολουθίας εικόνων της νοηματικής, όπου η συνεχής ανίχνευση αποτυγχάνει κάποιες φορές σε μεγάλες ταχύτητες. Οι τιμές τόσο της μεταβλητής ανίχνευσης (*detection value*) όσο και της τυπικής απόκλισης της εσωτερικής κατάστασης (P) αποθηκεύονται τόσο σε περιπτώσεις σωστής όσο και σε περιπτώσεις λάθος ανίχνευσης και χρησιμοποιούνται για την εκπαίδευση απλών γραμμικών classifiers που αποφασίζουν σχετικά με την ποιότητα της επιστρεφόμενης ανίχνευσης. Ανάλογα με την ανοχή σε λάθος ανιχνεύσεις που μπορούμε να επιτρέψουμε, μπορούμε να αποδώσουμε βάρη στις σωστές και λάθος αναγνωρίσεις κατά την διαδικασία της εκπαίδευσης.

Στην συγκεκριμένη εφαρμογή εκπαιδεύονται 2 classifiers, ένας από την τυπική απόκλιση της εσωτερική κατάσταση P , και ένας από την τιμή που επιστρέφει ο ανιχνευτής προσώπων του Κεφαλαίου 1. Από τα ποσοστά ανίχνευσης που παρουσιάζονται στον παρακάτω πίνακα, γίνεται φανερό ότι η τυπική απόκλιση της στοχαστικής μεταβλητής της εσωτερικής κατάστασης είναι ένας πιο ισχυρός δείκτης της ποιότητας της ανίχνευσης, από την τιμή που επιστρέφει ο ανιχνευτής που εκπαιδεύτηκε βάσει του αλγορίθμου AdaBoost.



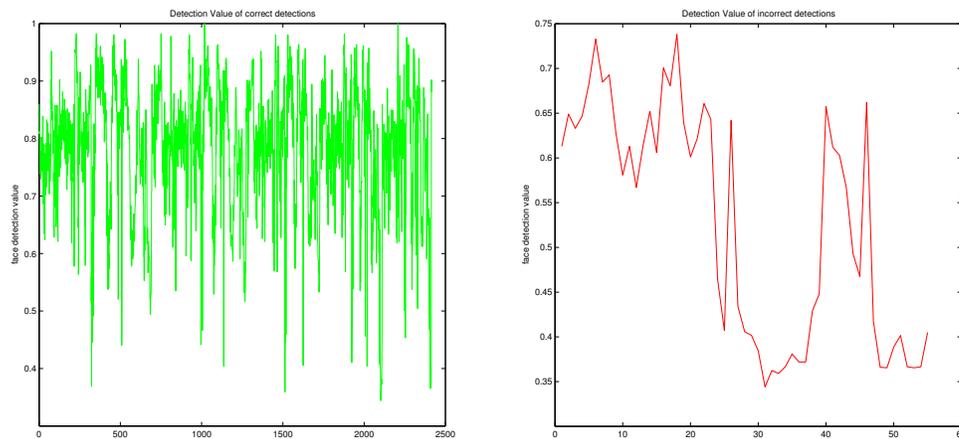
Σχήμα 3.7: Παρουσίαση των τυπικών αποκλίσεων σε περιπτώσεις σωστών και λάθος ανιχνεύσεων

Classification feature	Classification error (weighted)	Classification error (unweighted)
Inner State Std	10.3%	17.5%
Detection Value	11.91%	23.26%

3.7 Συμπεράσματα

Συνοψίζοντας, σε αυτό το κεφάλαιο έγινε μελέτη, σχεδιασμός και υλοποίηση μιας στοχαστικής μεθόδου συνεχούς ανίχνευσης (tracking) του προσώπου, βασισμένης στο φίλτρο Kalman. Αρχικά αναλύθηκε το πρόβλημα και δικαιολογήθηκε η επιλογή του φίλτρου Kalman ως του βέλτιστου. Στην συνέχεια έγινε εκτεταμένη περιγραφή των παραδοχών και των σχεδιαστικών αποφάσεων που πάρθηκαν κατά την υλοποίηση του φίλτρου, ενώ παράλληλα η μέθοδος αξιολογήθηκε πάνω στα benchmarks της νοηματικής. Τέλος, μεγέθη αυτού του στοχαστικού μοντέλου χρησιμοποιήθηκαν για την απόδοση εμπιστοσύνης στην ανίχνευση. Τόσο η επιλογή του φίλτρου Kalman, όσο και οι παραδοχές που πάρθηκαν κατά τον σχεδιασμό και την υλοποίησή του δικαιώνονται από τα τελικά ποσοστά ανίχνευσης.

Με την εφαρμογή του παραπάνω στοχαστικού φίλτρου, οι επιδόσεις του ανιχνευτή βελτιώθηκαν στους παρακάτω τομείς:



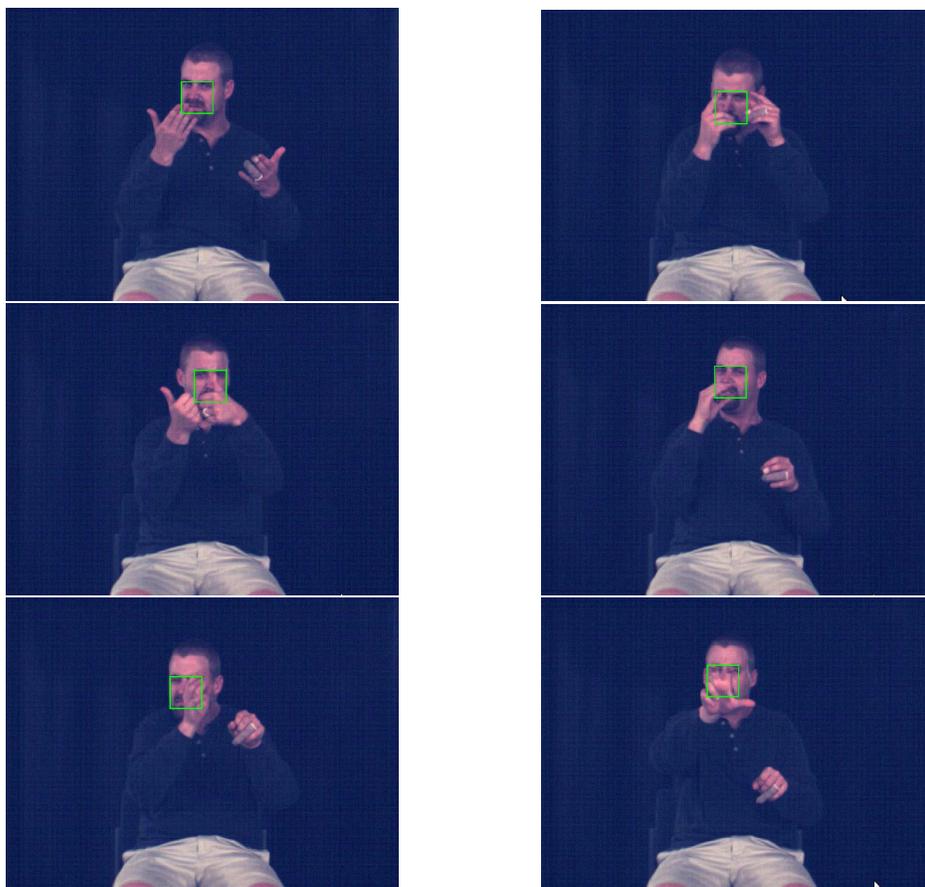
Σχήμα 3.8: Παρουσίαση των τιμών ανίχνευσης σε περιπτώσεις σωστών και λάθος ανιχνεύσεων

3.7.1 Ποσοστά ανίχνευσης

Όπως φάνηκε και από τα παραπάνω διαγράμματα τα ποσοστά ανίχνευσης και στα 2 σημειωμένα benchmarks, που κατασκευάστηκαν και χρησιμοποιήθηκαν στο Κεφάλαιο 1, ανέβηκαν φτάνοντας το 100% σχεδόν για όλες τις προσομοιωμένες ταχύτητες. Αυτό δεν σημαίνει οτι η συνεχής ανίχνευση έχει πάντα πλήρη επιτυχία, αλλά στις περισσότερες περιπτώσεις τα αποτελέσματα είναι πολύ ικανοποιητικά. Αν το πρόσωπο κινείται συνήθως με μικρές ταχύτητες (όπως συμβαίνει συνήθως στα βίντεο της νοηματικής), τότε το απλό ακίνητο μοντέλο δίνει πολύ ικανοποιητικά και αρκετά σταθερά αποτελέσματα. Για μεγαλύτερες ταχύτητες καλύτερα αποτελέσματα δίνει το σύνθετο κινητικό μοντέλο, αλλά υπάρχει πάντα ο κίνδυνος της αστάθειας του συστήματος.

3.7.2 Ταχύτητα ανίχνευσης

Ο ανιχνευτής προσώπων μετά την στοχαστική ανάλυση που προηγήθηκε και βάση του γεγονότος οτι πλέον η ανίχνευση γίνεται μόνο σε περιοχές $\pm 3\sigma$ από την προβλεπόμενη επόμενη θέση, γίνεται πολύ πιο γρήγορος από πριν. Ενώ στο προηγούμενο βήμα, και μετά από όλες τις προσπάθειες βελτιστοποίησης η ταχύτητα είχε φτάσει από τα 0.15 frames/sec (απλή ανίχνευση στο σύνολο της εικόνας και σε όλες τις κλίμακες) στα 2.85 frames/sec (ανίχνευση σε περιορισμένη περιοχή κίνησης του προσώπου και μόνο σε κοντινές κλίμακες), τώρα η ταχύτητα ανίχνευσης φτάνει τα 4.15 frames/sec. Η τιμή των 4 frames/sec είναι κοντά σε πραγματικό χρόνο, και δεδομένου οτι η εφαρμογή είναι υλοποιημένη σε γλώσσα προγραμματισμού Matlab, υπονοεί ελάχιστο υπολογιστικό κόστος σε μια πιο χαμηλού επιπέδου γλώσσα.



Σχήμα 3.9: Αποτελέσματα συνεχούς ανίχνευσης υπό μεταβαλλόμενη πόζα και εκτενή επικάλυψη

3.7.3 Εμπιστοσύνη και απόφαση σχετικά με την ποιότητα της ανίχνευσης

Στο τελευταίο μέρος αυτής την μελέτης έγινε φανερό ότι η τυπική απόκλιση της εσωτερικής κατάστασης του στοχαστικού μοντέλου, είναι ένας καλύτερος δείκτης της ποιότητας της ανίχνευσης από την τιμή που επιστρέφει ο ανιχνευτής. Χρησιμοποιώντας αυτό το μέγεθος πλέον, και έναν απλό classifier, είναι δυνατόν να επιλέξουμε αν θα εμπιστευτούμε το αποτέλεσμα της ανίχνευσης για το επόμενο στάδιο της εφαρμογής, ή αν θα βασιστούμε σε πρότερη πληροφορία στην περίπτωση που η ανίχνευση εμπεριέχει μεγάλη πιθανότητα λάθους.

Κεφάλαιο 4

Εκτίμηση της Πόζας

4.1 Εισαγωγή

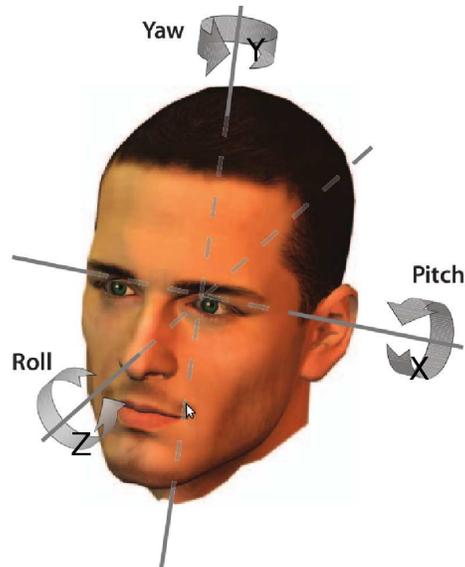
Η αυτόματη εκτίμηση της τρισδιάστατης πόζας του ανθρώπινου προσώπου (Head Pose Estimation) από υπολογιστή, μια λειτουργία τόσο εύκολη για τον άνθρωπο, έχει απασχολήσει πολλούς μελετητές και έχει προσεγγιστεί με πολλούς διαφορετικούς τρόπους τα τελευταία χρόνια. Μαζί με τον εντοπισμό και την αναγνώριση του προσώπου, αποτελεί ένα από τα τρία κύρια προβλήματα της μοντελοποίησης του ανθρώπινου προσώπου με αναρίθμητες εφαρμογές, από έξυπνα συστήματα για γρήγορο ή εναλλακτικό χειρισμό Η/Υ μέχρι την παρακολούθηση της προσοχής του οδηγού για την αποφυγή ατυχημάτων.

Το ανθρώπινο πρόσωπο έχει τρεις βαθμούς ελευθερίας όσον αφορά την στρέψη. Στην βιβλιογραφία οι τρεις αυτές στρέψεις αναφέρονται ως Pitch, Yaw και Roll. Η στρέψη Roll, είναι η περιστροφή του προσώπου πάνω στο δισδιάστατο επίπεδο της εικόνας, το Yaw είναι η τρισδιάστατη στρέψη από προφίλ σε προφίλ, ενώ το Pitch είναι η κάθετη στις άλλες δύο τρισδιάστατη στρέψη. Οι τρεις αυτές στρέψεις φαίνονται παραστατικά στο Σχήμα 4.1.

Όσον αφορά την νοηματική γλώσσα, η πόζα του προσώπου αποτελεί σημαντικό στοιχείο πληροφορίας για τα λεγόμενα του νοηματιστή. Για παράδειγμα ερωτήσεις, αρνήσεις και υποθετικές προτάσεις εκφράζονται πάντα με την βοήθεια της πόζας του κεφαλιού. Επίσης η εκτίμηση της πόζας είναι ένα σημαντικό πρώτο βήμα για την εκτίμηση του βλέμματος του ομιλητή (gaze), στοιχείο που επίσης μπορεί να επηρεάσει σημαντικά τα λεγόμενα του.

Οι ιδιαιτερότητες που παρουσιάζονται στις ακολουθίες εικόνων της νοηματικής γλώσσας είναι ότι συνήθως δεν υπάρχει εστίαση στο πρόσωπο και επομένως οι εικόνες του προσώπου είναι χαμηλής ανάλυσης, ενώ πολύ συχνά παρουσιάζεται το φαινόμενο μερικής ή ολικής επικάλυψης του προσώπου από τα χέρια. Τα δύο αυτά χαρακτηριστικά καθιστούν την αυτόματη αναγνώριση της πόζας σε αυτό το πρόβλημα αρκετά δύσκολη, και οι περισσότερες κλασσικές προσεγγίσεις αποτυγχάνουν. Αντίθετα, ο νοηματιστής περιορίζεται σε σχετικά μικρές στρέψεις Yaw και Pitch (μέχρι 45°), ενώ η στρέψη Roll είναι συνήθως μικρή έως αμελητέα.

Στο κεφάλαιο αυτό γίνεται αρχικά μια συνοπτική παρουσίαση και κατηγοριοποίηση κάποιων από τις σημαντικότερες μεθόδους για την εκτίμηση της πόζας του ανθρώπινου προσώπου, με ιδιαίτερη έμφαση σε προσεγγίσεις ευέλικτων γράφων προσώπου, όπως τα Active Appearance



Σχήμα 4.1: Βαθμοί Ελευθερίας του Ανθρώπινου Προσώπου [8].

Models. Στο τέλος της ενότητας προτείνεται μια εναλλακτική ιδέα για την ανίχνευση της πόζας, που εκμεταλλεύεται τις απλοποιήσεις όσον αφορά την νοηματική γλώσσα (μικρό εύρος γωνιών στρέψης), και αποδεικνύεται ιδιαίτερα εύρωστη στην περίπτωση της επικάλυψης από τα χέρια.

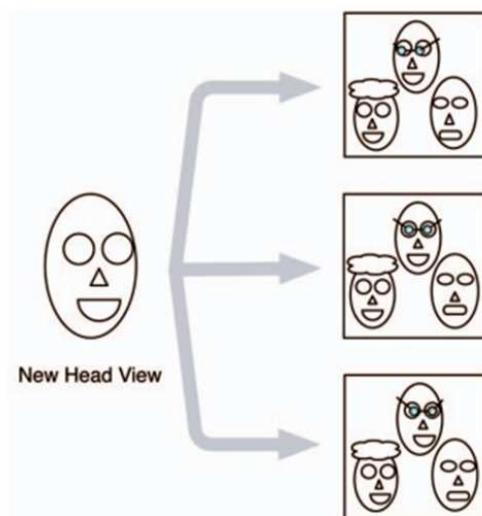
4.2 Διαφορετικές Προσεγγίσεις για την Εκτίμηση της Πόζας

Η παρούσα ενότητα κάνει μια βιβλιογραφική ανασκόπηση κάποιων εκ των σημαντικότερων μεθόδων που έχουν προταθεί και χρησιμοποιηθεί επιτυχώς για την εκτίμηση της πόζας του ανθρώπινου προσώπου. Οι διαφορετικές μέθοδοι μπορούν να κατηγοριοποιηθούν σε ομάδες ανάλογα με την σκοπιά από την οποία προσεγγίζουν το πρόβλημα. Η κάθε ομάδα μεθόδων αντιμετωπίζει διαφορετικές δυσκολίες και περιορισμούς, ενώ παρουσιάζει και ανάλογα πλεονεκτήματα [8].

4.2.1 Κατηγοριοποίηση με βάση την εμφάνιση

Αυτή η κατηγορία αντιμετωπίζει το πρόβλημα από την σκοπιά της κατηγοριοποίησης των εικόνων προσώπων σε διακριτές κλάσεις. Οι κλάσεις αυτές αντιστοιχούν σε κάποιες προεπιλεγμένες διακριτές πόζες, ενώ για την κατηγοριοποίηση έχουν προταθεί αρκετές διαφορετικές μέθοδοι.

Η πιο απλή μέθοδος κατηγοριοποίησης είναι η διαδοχική σύγκριση της εικόνας του προσώπου με εικόνες προσώπων σε διακριτές προσημειωμένες πόζες. Για κάθε πιθανή διακριτή πόζα



- Iterative Comparison
- Detector Arrays
- Pose Eigenspaces

Σχήμα 4.2: Εκτίμηση της πόζας με διακριτή κατηγοριοποίηση με βάση την εμφάνιση [8].

κατασκευάζεται μια κλάση που αποτελείται από ένα σύνολο διαφορετικών προσώπων σε αυτή την πόζα. Το νέο, προς αναγνώριση, πρόσωπο συγκρίνεται στην συνέχεια με κάθε εικόνα σε κάθε μία από τις διακριτές αυτές κλάσεις, και επιλέγεται ως σωστή η κλάση όπου επιτυγχάνεται το καλύτερο ταιρίασμα [12]. Για να δοθεί έμφαση στην πόζα, παρά στην εμφάνιση του προσώπου κατά την κατηγοριοποίηση, οι εικόνες φιλτράρονται με κατευθυντικά φίλτρα Gabor. Με αυτόν τον τρόπο αμβλύνονται οι λεπτομέρειες της εμφάνισης, ενώ εντείνονται κάποια κατευθυνόμενα χαρακτηριστικά του προσώπου όπως η μύτη ή το στόμα.

Σε αυτή την κατηγορία μεθόδων, συμπεριλαμβάνονται και οι μέθοδοι πολλαπλών ανιχνευτών (detector arrays), όπου μια σειρά από ανιχνευτές κάθε ένας από τους οποίους έχει εκπαιδευτεί για διαφορετική διακριτή πόζα εφαρμόζονται στο πρόσωπο, και ως σωστή πόζα επιλέγεται αυτή της οποίας ο αντίστοιχος ανιχνευτής δίνει θετική απάντηση. Οι ανιχνευτές εκπαιδεύονται με βάση κάποιον αλγόριθμο εκμάθησης, όπου τα θετικά δείγματα είναι εικόνες προσώπων σε μια συγκεκριμένη πόζα, και τα αρνητικά δείγματα αποτελούνται από πρόσωπα σε γειτονικές πόζες. Οι αλγόριθμοι εκμάθησης που έχουν χρησιμοποιηθεί για αυτή την μέθοδο είναι οι AdaBoost, SVM όπως επίσης και Neural Networks.

Μια εναλλακτική προσέγγιση στην διακριτή κατηγοριοποίηση της πόζας, προτείνει η μέθοδος ιδιοχώρων της πόζας (pose eigenspaces) [23]. Με βάση αυτή την ιδέα, κατασκευάζεται ένας υποχώρος για κάθε διακριτή πόζα εφαρμόζοντας τον μετασχηματισμό Ανάλυσης Πρωτεύουσων Συνιστωσών (Principal Component Analysis) (βλ. Παράρτημα) πάνω σε εικόνες

προσώπων αυτής της πόζας. Η νέα εικόνα προσώπου προβάλλεται πάνω σε κάθε έναν από τους διακριτούς αυτούς υποχώρους, και ως σωστή πόζα επιλέγεται αυτή όπου η ενέργεια της προβολής μεγιστοποιείται.

Οι παραπάνω μέθοδοι έχουν δοκιμαστεί σε μεγάλο βαθμό και αποδίδουν σχετικά καλά σε εικόνες τόσο υψηλής όσο και χαμηλής ανάλυσης. Παρ' όλα αυτά έχουν το βασικό μειονέκτημα ότι μπορούν να δώσουν μόνο διακριτά αποτελέσματα όσον αφορά την πόζα, είναι αρκετά αργές κατά την εφαρμογή τους, και προφανώς, καθώς βασίζονται στην εμφάνιση του προσώπου, επηρεάζονται σε πολύ μεγάλο βαθμό από την πιθανή επικάλυψη.

4.2.2 Εφαρμογή ευέλικτων μοντέλων γράφων προσώπου

Εισαγωγή

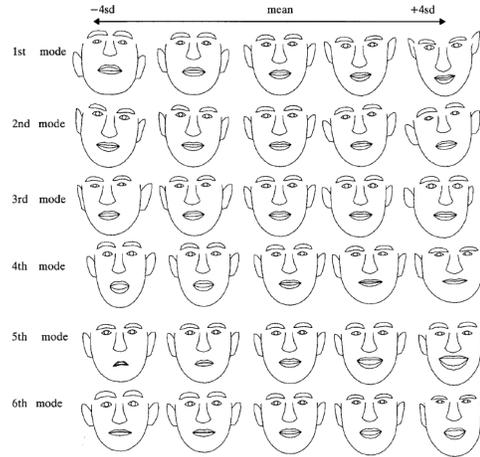
Για την μοντελοποίηση του ανθρώπινου προσώπου, έχουν προταθεί αρκετά διαφορετικά μοντέλα ευέλικτων γράφων, οι οποίοι έχουν την ιδιότητα να προσαρμόζονται στο ανθρώπινο πρόσωπο και να μοντελοποιούν το σχήμα ή και την εμφάνισή του. Δύο από τα σπουδαιότερα αυτά μοντέλα είναι τα Active Shape Models (ASMs) [14] και τα Active Appearance Models (AAMs) [5] [17], τα οποία χρησιμοποιούν τον μετασχηματισμό Principal Component Analysis για αυτή την μοντελοποίηση. Οι ιδέες των ASM και των AAM στηρίζονται στο γεγονός ότι το σχήμα και η εμφάνιση του ανθρώπινου προσώπου έχουν πεπερασμένη μεταβλητότητα, και μπορούν να περιγράφουν από έναν χώρο σχετικά μικρής διάστασης. Ο μετασχηματισμός Principal Component Analysis (PCA) πάνω σε γράφους σχήματος ή σε εικόνες προσώπων έχει την ιδιότητα να κατασκευάζει τέτοιους χώρους περιγραφής. Με αυτό τον τρόπο, τόσο το σχήμα όσο και η εμφάνιση του προσώπου, μπορούν να παρασταθούν με ένα γραμμικό μοντέλο ως ένα σταθμισμένο άθροισμα από τις κύριες συνιστώσες του χώρου πάνω στον οποίον προβάλλει ο μετασχηματισμός Principal Component Analysis. Ο μετασχηματισμός αυτός καθώς και οι βασικές ιδέες για την μοντελοποίηση και την εφαρμογή των Active Shape Models και Active Appearance Models περιγράφονται στο Παράρτημα. Σε αυτή την υποενότητα συζητιέται η δυνατότητά τους να παρέχουν εύρωστη και ακριβή εκτίμηση της πόζας, μέσω του ελέγχου των συνιστωσών τους στον χώρο προβολής που δημιουργείται από τον μετασχηματισμό PCA.

Active Shape Models

Η εφαρμογή των ευέλικτων γράφων Active Shape Models (βλ. Παράρτημα) πάνω στο ανθρώπινο πρόσωπο παρέχει την δυνατότητα επιτυχούς αναπαράστασης των πρωτογενών παραγόντων παραμόρφωσης του σχήματός του. Η πόζα του προσώπου αποτελεί έναν από τους κύριους παράγοντες παραμόρφωσης του σχήματος, και όπως φαίνεται στο Σχήμα 4.3, η παραμόρφωση λόγω μεταβολής της πόζας συγκεντρώνεται στις πρώτες κύριες συνιστώσες που προκύπτουν από τον μετασχηματισμό Principal Component Analysis. Πιο συγκεκριμένα, η εφαρμογή του PCA πάνω στους γράφους προσώπου (ASM) τείνει να διαχωρίζει τις τρεις στρέψεις του προσώπου Yaw, Pitch και Roll και να τις αναπαριστά από τις τρεις πρώτες κύριες συνιστώσες. Αυτή η παρατήρηση προτείνει ότι απλώς ελέγχοντας τις τρεις πρώτες κύριες

συνιστώσες του γράφου σχήματος μπορεί να δοθεί μια συμπαγής εκτίμηση της πόζας.

Παρ' όλα αυτά αυτός ο διαχωρισμός είναι δυνατός μόνο όταν η μεταβολή της πόζας είναι ιδιαίτερα μικρή (από -20 μέχρι 20 μοίρες). Όπως θα παρουσιαστεί παρακάτω, στις περιπτώσεις που η πόζα του προσώπου μεταβάλλεται σημαντικά, ο μετασχηματισμός PCA διαχωρίζει μεν την παραμόρφωση λόγω στρέψης από άλλου είδους παραμορφώσεις, αλλά δεν παρέχει την δυνατότητα διαχωρισμού των επιμέρους συνιστωσών της στρέψης, καθιστώντας έτσι το πρόβλημα εκτίμησης της πόζας πιο δύσκολο. Επίσης, τα Active Shape Models δεν είναι ιδιαίτερα εύρωστα στην εφαρμογή τους, ιδιαίτερα σε περιπτώσεις επικάλυψης.



Σχήμα 4.3: Κύριες συνιστώσες του σχήματος του γράφου προσώπου ASM [14].

View-Based combined Active Appearance Models

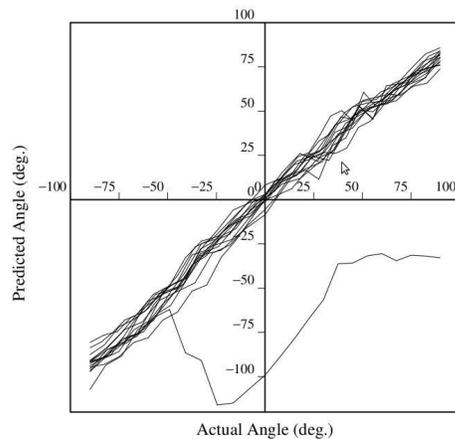
Η μέθοδος των συνδυασμένων (combined) AAMs (βλ. Παράρτημα), παραμετροποιεί με την βοήθεια του μετασχηματισμού PCA τις μεταβολές στο σχήμα και την εμφάνιση του προσώπου σε ένα μοναδικό διάνυσμα κυρίων συνιστωσών c το οποίο τις περιγράφει από κοινού. Εφόσον τα Active Appearance Models χρησιμοποιούν πια την εμφάνιση σαν παράγοντα για την εφαρμογή τους, γίνονται πολύ πιο εύρωστα στην εφαρμογή τους σε νέα πρόσωπα (fitting) και μπορούν να παρέχουν πιο ακριβή περιγραφή του σχήματος. Τα συνδυασμένα AAMs, όπως και τα ASMs, μπορούν να χρησιμοποιηθούν για την εκτίμηση της πόζας από τον έλεγχο του διανύσματος συνιστωσών c το οποίο περιγράφει σε αυτή την περίπτωση τόσο το σχήμα όσο και την εμφάνιση του προσώπου [4].

Υποθέτοντας ότι το το διάνυσμα των παραμέτρων c θα εξαρτάται από την γωνία στρέψης του ανθρώπινου προσώπου σύμφωνα με την Εξίσωση 4.1, είναι δυνατόν να εκτιμηθούν οι παράμετροι c_0 , c_x και c_y από ένα σύνολο προσώπων εκπαίδευσης υπό γνωστή στρέψη.

$$c = c_0 + c_x \cos(\theta) + c_y \sin(\theta) \quad (4.1)$$

Με αυτή την μοντελοποίηση γίνεται δυνατή η εκτίμηση της πόζας του ανθρώπινου προσώπου αλλά με ορισμένους περιορισμούς. Η εφαρμογή των AAMs καθίσταται συνήθως ανεπιτυχής όταν το πρόσωπο βρίσκεται υπό στρέψη μεγαλύτερη από 45° . Αυτό συμβαίνει λόγω της επικάλυψης απαραίτητων σημείων του γράφου υπό μεγαλύτερες γωνίες. Για ξεπεραστεί αυτός ο περιορισμός εκπαίδευεται μια σειρά από AAMs, κάθε ένα από τα οποία απευθύνεται σε διαφορετικό εύρος γωνιών ώστε να καλύπτεται ολόκληρο το φάσμα ($-90^\circ, 90^\circ$).

Η παραπάνω ιδέα, ονομάστηκε View-Based AAMs [4] και έχει εφαρμοστεί για την εκτίμηση της πόζας του ανθρώπινου προσώπου. Η επίδοσή τους στην περίπτωση εκτίμησης της στρέψης Yaw, φαίνεται στο Σχήμα 4.4, ενώ με ανάλογη διαδικασία είναι δυνατή και η εκτίμηση της κάθετης στρέψης Pitch.



Σχήμα 4.4: Αποτελέσματα εκτίμησης της πόζας με την μέθοδο View-Based AAMs [4].

Στην γενική περίπτωση η εκτίμηση γίνεται με απόκλιση 10° , αλλά όπως φαίνεται στο διάγραμμα το λάθος μπορεί πολλές φορές να φτάσει και τις 20° . Εκτός αυτού η εφαρμογή των συνδιασμένων AAMs στο ανθρώπινο πρόσωπο είναι αρκετά πολύπλοκη και σχετικά αργή διαδικασία, ενώ η επιτυχία αναγνώρισης της πόζας βασίζεται πάντα στην επιτυχημένη εφαρμογή (fitting) του μοντέλου, γεγονός που δεν είναι εγγυημένο ιδιαίτερα σε περιπτώσεις χαμηλής ανάλυσης και εκτεταμένης επικάλυψης. Τέλος δεν έχει δοκιμαστεί πως αυτή η τεχνική συμπεριφέρεται σε κάποιον συνδυασμό Yaw και Pitch. Για όλους αυτούς τους λόγους η μέθοδος εκτίμησης της πόζας από τις παραμέτρους των combined AAMs δεν εφαρμόζεται συνήθως στην πράξη.

Independent AAMs

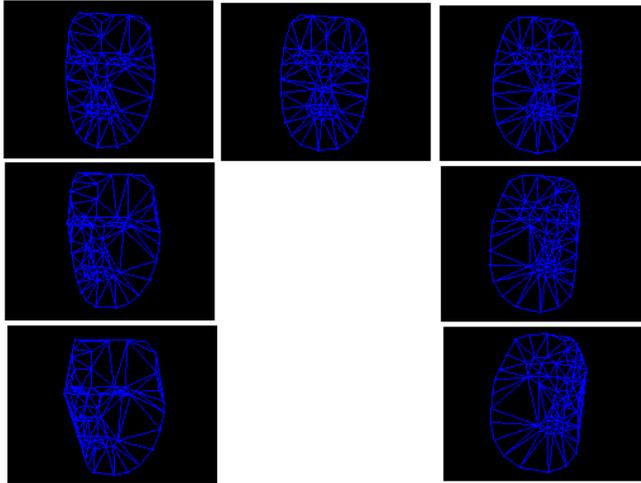
Τα independent AAMs (βλ. Παράρτημα) παραμετροποιούν ανεξάρτητα το σχήμα και την εμφάνισή του ανθρώπινου προσώπου, με την βοήθεια του μετασχηματισμού PCA και στις δύο περιπτώσεις. Το σχήμα περιγράφεται από έναν ευέλικτο γράφο, ο οποίος εκφράζεται σαν ένα σταθμισμένο άθροισμα των κύριων συνιστωσών παραμόρφωσης (όπως στην περίπτωση των

ASMs), ενώ η εμφάνιση ως σταθμισμένο άθροισμα ιδιοπροσώπων. Με την βοήθεια του αλγορίθμου Inverse Compositional Image Alignment, τα independent AAMs έχουν την ικανότητα να εφαρμόζουν πολύ καλύτερα στο ανθρώπινο πρόσωπο ακόμα και σε περιπτώσεις εκτεταμένης μεταβολής της πόζας (self-occlusion) ή επικάλυψης από άλλα αντικείμενα (occlusion). Για το πρόβλημα της πόζας μπορούμε να επικεντρωθούμε στο σχήμα του γράφου και να αδιαφορήσουμε σχετικά με την εμφάνιση, εφόσον πια οι δύο αυτές παράμετροι αντιμετωπίζονται ανεξάρτητα.

Εφόσον η μεταβολή της πόζας αποτελεί τον σημαντικότερο λόγο παραμόρφωσης του σχήματος του ανθρώπινου προσώπου, αναμένεται ότι και σε αυτή την περίπτωση οι πρώτες συνιστώσες του μετασχηματισμού PCA θα σχετίζονται με την πόζα. Μεταβάλλοντας τις πρώτες κύριες συνιστώσες του σχήματος σε AAMs που εκπαιδεύτηκαν από εικόνες προσώπου της νοηματικής γλώσσας προκύπτουν γράφοι προσώπου που φαίνονται στα Σχήματα 4.5, 4.6 και 4.7. Είναι εμφανές ότι οι πρώτες δύο συνιστώσες εμπεριέχουν την πληροφορία της πόζας του προσώπου, ενώ οι επόμενες (που είναι και λιγότερο σημαντικές) εμπεριέχουν πληροφορία σχετικά με τις διαφορετικές εκφράσεις του προσώπου. Αξίζει να σημειωθεί, ότι ακριβώς η ίδια συμπεριφορά σχετικά με τις πρώτες συνιστώσες του σχήματος του προσώπου επιδεικνύεται και στην αυθεντική δημοσίευσή των independent AAMs [17] (Σχήμα 4.8), γεγονός που φανερώνει ότι ανεξαρτήτως του συνόλου εκπαίδευσης, με την προϋπόθεσή ότι αυτό περιλαμβάνει αρκετά μεγάλη μεταβολή της πόζας του προσώπου, ο μετασχηματισμός PCA έχει την ικανότητα να απομονώνει την παραμόρφωση στο σχήμα που οφείλεται στην μεταβολή της πόζας, και να την συγκεντρώνει στις δύο πρώτες συνιστώσες. Για την εκπαίδευση και εφαρμογή των AAMs στα βίντεο της νοηματικής χρησιμοποιήθηκε ο ανοιχτός κώδικας των Papandreou et. al. [21] και η βάση δεδομένων BU400 [19]

Παρ' όλα αυτά, όπως είναι εμφανές από τα σχήματα, ο μετασχηματισμός PCA δεν διαχωρίζει τις γωνίες στρέψης Yaw και Pitch, και επομένως η εκτίμηση της πόζας απευθείας από την προβολή πάνω στον ιδιοχώρο του σχήματος δεν είναι αυτόματη. Η στρέψη Roll δεν εμφανίζεται καν σαν παράγοντας παραμόρφωσης του σχήματος στις κύριες συνιστώσες. Αυτό οφείλεται στο γεγονός ότι κατά την εκπαίδευση, το Roll (που εκφράζει την περιστροφή στο επίπεδο της εικόνας) έχει αφαιρεθεί με την βοήθεια του αλγορίθμου Procrustis Analysis, ο οποίος εφαρμόζεται πριν τον PCA και αφαιρεί όλες τις στεραίες (rigid) παραμορφώσεις του σχήματος (μετατόπιση και δισδιάστατη περιστροφή στο επίπεδο της εικόνας). Αυτές οι παραμορφώσεις του σχήματος αντιμετωπίζονται ανεξάρτητα και κατά την διαδικασία της εφαρμογής του μοντέλου (fitting) και μπορούν εύκολα να δώσουν με καλή ακρίβεια την γωνία Roll.

Επειδή οι στρέψεις Yaw και Pitch δεν είναι διαχωρίσιμες από τον μετασχηματισμό PCA δεν είναι δυνατή η αυτόματη εκτίμησή της πόζας από τους συντελεστές των κύριων συνιστωσών του σχήματος. Παρόλα αυτά, μετά την επιτυχή εφαρμογή των AAMs στο πρόσωπο, μπορεί να εκπαιδευτεί μια δισδιάστατη συνάρτηση αντιστοίχισης των πρωτεύουσων συνιστωσών στις πραγματικές γωνίες από ένα σύνολο προσώπων εκπαίδευσης υπό γνωστές στρέψεις, ή να χρησιμοποιηθεί κάποια γεωμετρική μέθοδος για την εκτίμηση της πόζας βάσει της θέσης των σημείων του εφαρμοσμένου γράφου. Τέτοιες γεωμετρικές μέθοδοι αναφέρονται παρακάτω.



Σχήμα 4.5: Μεταβάλλοντας την πρώτη κύρια συνιστώσα του σχήματος από $-3\sqrt{\lambda}$ το $3\sqrt{\lambda}$.

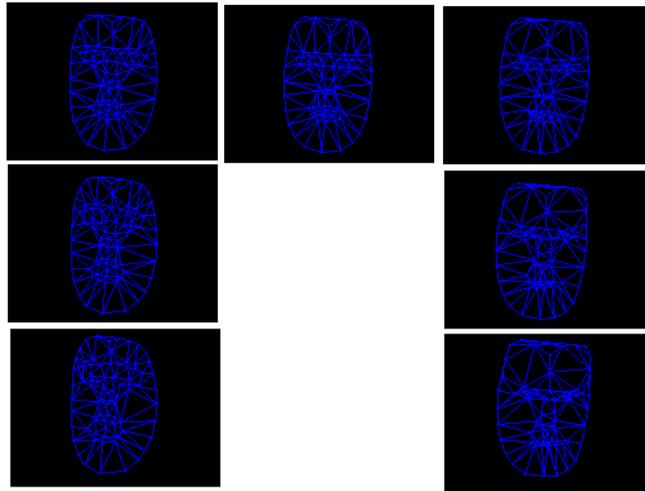
4.2.3 Γεωμετρικές μέθοδοι

Γεωμετρικά το θέμα της πόζας του ανθρώπινου προσώπου μπορεί να προσεγγιστεί σαν ένα πρόβλημα εύρεσης της τρισδιάστατης περιστροφής των τοπικών αξόνων του προσώπου ως προς τους άξονες της κάμερας. Το πρόσωπο αντιμετωπίζεται σαν ένα μη παραμορφώσιμο στερεό σώμα οι τοπικοί άξονες του οποίου φαίνονται στο Σχήμα 4.1 του κεφαλαίου, ώστε να συμπίπτουν με τους άξονες της κάμερας όταν το πρόσωπο κοιτάει ευθέως την κάμερα. Έτσι στην ευθεία θέση οι τρεις γωνίες περιστροφής γύρω από τους άξονες είναι μηδενικές, ενώ αν το πρόσωπο περιστραφεί οι τοπικοί του άξονες αλλάζουν προσανατολισμό σε σχέση με αυτούς της κάμερας και υπολογίζοντας τις γωνίες αυτές της περιστροφής βρίσκουμε αμέσως την πόζα του προσώπου.

Τρισδιάστατες γεωμετρικές μέθοδοι

Στην γενική περίπτωση, αρκεί η αντιστοίχιση τριών σημείων (landmarks) του χώρου στην αρχική θέση με τρία σημεία στην τελική θέση για να προσδιοριστεί τόσο η μετατόπιση όσο και περιστροφή του στερεού. Στην ουσία επειδή πάντα ο προσδιορισμός τρισδιάστατων σημείων από δισδιάστατη κάμερα εμπεριέχει σε μεγάλο βαθμό κάποιο σφάλμα, αντιστοιχίζεται ένα σύνολο πολλών σημείων, και η τελική περιστροφή βρίσκεται χρησιμοποιώντας κάποια μέθοδο ελαχίστων τετραγώνων.

Για να προσδιοριστεί η τρισδιάστατη θέση σημείων πάνω στο ανθρώπινο πρόσωπο, έχουν προταθεί και χρησιμοποιηθεί αρκετές μέθοδοι. Κάποιες από τις βασικότερες που έχουν χρη-



Σχήμα 4.6: Μεταβάλλοντας την δεύτερη κύρια συνιστώσα του σχήματος από $-3\sqrt{\lambda}$ το $3\sqrt{\lambda}$.

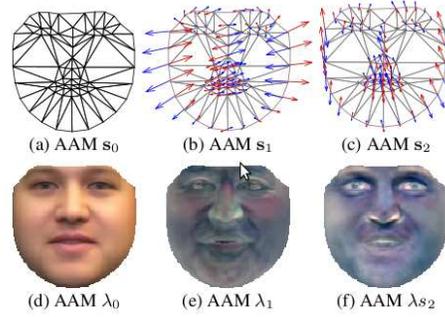


Σχήμα 4.7: Μεταβάλλοντας την τρίτη και τέταρτη κύρια συνιστώσα του σχήματος

σιμοποιηθεί για την εκτίμηση της πόζας περιγράφονται με συνομία παρακάτω:

3D Morphable Models Τα 3D Morphable Models [2] αποτελούν τρισδιάστατα μοντέλα προσώπου που εκπαιδεύονται από σαρώσεις laser πάνω σε πρόσωπα, και εφαρμόζουν σε καινούργια πρόσωπα παίρνοντας την τρισδιάστατη επιφάνεια τους. Με αυτόν τον τρόπο δίνουν πληθώρα τρισδιάστατων σημείων πάνω στην επιφάνεια του προσώπου με αρκετά καλή ακρίβεια του βάθους από μία και μοναδική εικόνα. Το μειονέκτημά τους για μια τέτοιου είδους εφαρμογές είναι ότι είναι πολύ πολύπλοκα στην υλοποίηση, ευαίσθητα σε επικαλύψεις και πάρα πολύ αργά κατά την εφαρμογή τους, μην επιτρέποντας έτσι αποτελέσματα σε πραγματικό χρόνο.

3D AAMs Τα Active Appearance Models είναι δυνατόν να επεκταθούν σε τρεις διαστάσεις με την βοήθεια κάποιου αλγόριθμου structure from motion [1], κρατώντας όλα τα υπόλοιπα πλεονεκτήματα που αυτά προσφέρουν, δηλαδή την ευρωστία σε μεταβολές της πόζας και επικαλύψεις. Η τρισδιάστατη εκδοχή των AAMs εφαρμόζει έναν τρισδιάστα-



Σχήμα 4.8: Κύριες συνιστώσες μεταβολής του σχήματος [17].

το γράφο πάνω στο πρόσωπο, παρέχοντας με αυτόν τον τρόπο την τρισδιάστατη θέση σημείων (landmarks) πάνω στο πρόσωπο.

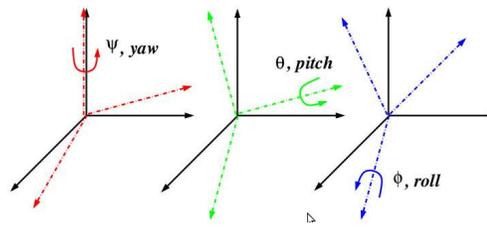
Εφόσον κάποια χαρακτηριστικά σημεία του τρισδιάστατου χώρου 3D landmarks έχουν προσδιοριστεί πάνω στο πρόσωπο, η πόζα του προσώπου μπορεί να βρεθεί με άμεσο τρόπο χρησιμοποιώντας κάποια μέθοδο ελαχίστων τετραγώνων. Το πρόβλημα εκφυλίζεται πλέον στην περιστροφή του συστήματος αξόνων της κάμερας (που συμπίπτει με το σύστημα αξόνων του προσώπου στην μηδενική πόζα) (E_1, E_2, E_3) ώστε να συμπίσει με το σύστημα αξόνων του προσώπου στην αυθαίρετη πόζα (e_1, e_2, e_3) . Το δεύτερο σύστημα προκύπτει από το πρώτο μετά από τον πολλαπλασιασμό με τον κατάλληλο Euler πίνακα περιστροφής (Euler Rotation Matrix). Ο πίνακας περιστροφής Euler μπορεί να πάρει διαφορετικές μορφές ανάλογα με τους άξονες περιστροφής και τις αντίστοιχες γωνίες Euler που έχουμε επιλέξει. Για την άμεση εκτίμηση των γωνιών Yaw, Pitch, Roll μπορούμε να επιλέξουμε τις γωνίες Euler 3-2-1, και τον αντίστοιχο πίνακα περιστροφής. Σε αυτή την περίπτωση οι άξονες περιστροφής φαίνονται στο Σχήμα 4.9, ενώ ο Euler πίνακας περιστροφής Q δίνεται από την σχέση:

$$\begin{pmatrix} e_1 \\ e_2 \\ e_3 \end{pmatrix} = Q^T \times \begin{pmatrix} E_1 \\ E_2 \\ E_3 \end{pmatrix}$$

με

$$Q = \begin{bmatrix} \cos\theta\cos\psi & \sin\theta\sin\phi\cos\psi - \cos\phi\sin\psi & \cos\phi\sin\theta\cos\psi + \sin\phi\sin\psi \\ \cos\theta\sin\psi & \sin\theta\sin\phi\sin\psi - \cos\phi\cos\psi & \cos\phi\sin\theta\sin\psi + \sin\phi\cos\psi \\ -\sin\theta & \sin\phi\cos\theta & \cos\phi\cos\theta \end{bmatrix}$$

Αν γνωρίζουμε τις συντεταγμένες στην μηδενική $X = [x_1, \dots, x_n]$ και στην τυχαία περιστροφή $X' = [x'_1, \dots, x'_n]$, ο πίνακας περιστροφής Euler ορίζει την σχέση $X' = QX$ και μπορεί να βρεθεί με την βοήθεια της μεθόδου ελαχίστων τετραγώνων μέσω της σχέσης



Σχήμα 4.9: Άξονες στρέψης Yaw, Pitch και Roll

$Q = X'X^T(XX^T)^{-1}$, και από τον πίνακα περιστροφής μπορούν να προσδιοριστούν οι γωνίες ψ, θ, ϕ που δίνουν τα Yaw, Pitch και Roll αντίστοιχα.

Το σύστημα είναι υπερορισμένο καθώς ο πίνακας περιστροφής έχει 9 αγνώστους, ενώ οι γωνίες περιστροφής είναι μόνο 3. Επειδή δεν ενδιαφερόμαστε για τον πίνακα περιστροφής αλλά συγκεκριμένα για τις τρεις γωνίες, μπορούμε αντί να επιλύσουμε το υπερορισμένο σύστημα να προσδιορίσουμε άμεσα τις γωνίες από τα τρισδιάστατα σημεία χρησιμοποιώντας μη γραμμική μέθοδο ελαχίστων τετράγωνο (non-linear least squares method) (βλ. Παράρτημα).

Δισδιάστατες γεωμετρικές μέθοδοι

Εκτός από τρισδιάστατες γεωμετρικές μεθόδους, οι οποίες απαιτούν τον προσδιορισμό τρισδιάστατων σημείων πάνω στο ανθρώπινο πρόσωπο, υπάρχουν και δισδιάστατες γεωμετρικές μέθοδοι για την εκτίμηση της πόζας. Αυτές οι μέθοδοι ανιχνεύουν και πάλι σημεία (landmarks) πάνω στο ανθρώπινο πρόσωπο υπό γνωστή γεωμετρία, και χρησιμοποιούν κάποια γεωμετρική μέθοδο για την εξαγωγή της πόζας υποθέτωντας συνήθως προοπτική προβολή. Τέτοια σημεία μπορεί να είναι για παράδειγμα οι εσωτερικές και εξωτερικές γωνίες των ματιών και του στόματος (συνεπίπεδα) και η κορυφή της μύτης ή το σύνολο των σημείων του γράφου προσώπου, όπως αυτά βρίσκονται από την εφαρμογή ενός μοντέλου ευέλικτου γράφου.

Προϋποθέτοντας γνώση σχετικά με την γεωμετρία των σημείων αυτών πάνω στο πρόσωπο, η πόζα μπορεί να εκτιμηθεί από κάποια λύση κλειστού τύπου ή από κάποιον αλγόριθμο που χρησιμοποιεί ένα υπερ-πλήρες σύνολο σημείων για πιο εύρωστες, στην λάθος ανίχνευση, λύσεις. Τέτοιες λύσεις μπορεί να είναι από την απλή εύρεση του μετασχηματισμού ομογραφίας (planar homography) από τέσσερα συνεπίπεδα σημεία (γωνίες των ματιών και το στόματος), μέχρι την εφαρμογή ενός καθολικού αλγορίθμου ανίχνευσης της πόζας όπως ο αλγόριθμος POSIT, ο οποίος απαιτεί όμως γνώση σχετικά με την τρισδιάστατη γεωμετρία του προσώπου, όπως και των εσωτερικών χαρακτηριστικών της κάμερας.

Αυτές οι μέθοδοι είναι συνήθως απλές και γρήγορες στην υλοποίησή τους, και μπορούν να εκτιμήσουν με αρκετά μεγάλη ακρίβεια την πόζα. Παρ' όλα αυτά, για την επιτυχή υλοποίησή τους απαιτείται ο εντοπισμός ενός συνόλου σημείων, με πολύ μεγάλη ακρίβεια, γεγονός που τις καθιστά ακατάλληλες σε περιπτώσεις χαμηλής ανάλυσης ή επικάλυψης.

4.3 Ανίχνευση της Πόζας μέσω της Συνεχούς Ανίχνευσης του Στενού Μοτίβου του Προσώπου

4.3.1 Εισαγωγή

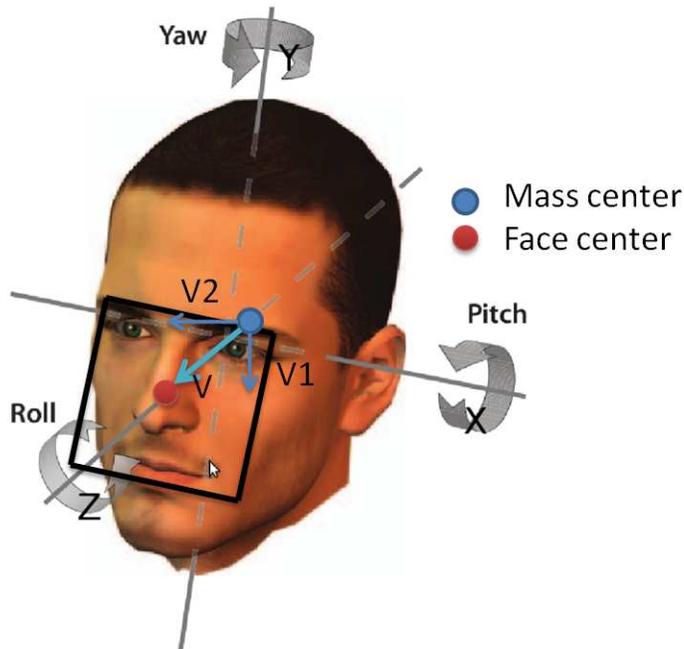
Στην προηγούμενη ενότητα περιγράφηκαν κάποιες από τις σημαντικότερες μεθόδους για την εκτίμηση της πόζας του ανθρώπινου προσώπου που συναντώνται στην βιβλιογραφία, κατηγοριοποιημένες σε τρεις υποενότητες ανάλογα με την σκοπιά από την οποία αντιμετωπίζουν το πρόβλημα. Οι μέθοδοι που βασίζονται στην κατηγοριοποίηση της εμφάνισης είναι συνήθως εύκολες στην υλοποίησή τους, άλλα μπορούν να δώσουν μόνο διακριτά και ανακριβή αποτελέσματα, ενώ είναι σχετικά αργές. Ειδικά στην περίπτωση της νοηματικής γλώσσας όπου οι γωνίες στρέψης είναι σχετικά μικρές (μικρότερες από 40°) μια τέτοια μέθοδος θα αδυνατούσε αν δώσει επιτυχή αποτελέσματα.

Τα ευέλικτα μοντέλα γράφων, παρόλο που παρέχουν την δυνατότητα απομόνωσης της πόζας από τους υπόλοιπους παράγοντες παραμόρφωσης του σχήματος, δεν επιτυγχάνουν τον διαχωρισμό των τριών συνιστωσών στρέψης. Έτσι, απαιτείται περαιτέρω επεξεργασία μέσω της εκπαίδευσης κάποιας δισδιάστατης συνάρτησης που αντιστοιχεί τις ιδιοτιμές σε γωνίες, ή κάποιας γεωμετρικής μεθόδου που θα ανακτήσει την γωνία από τις σχετικές θέσεις των σημείων (landmarks) του γράφου προσώπου. Σε περιπτώσεις όμως χαμηλής ανάλυσης και εκτεταμένης επικάλυψης, όπως είναι η περίπτωση που μελετάται, η ακριβής εφαρμογή του γράφου προσώπου καθίσταται συχνά δύσκολη ή αδύνατη, γεγονός που θα παρήγαγε ανακριβή και μη αξιόπιστη εκτίμηση της πόζας.

Σε αυτή την ενότητα προτείνεται μια καινοτόμος μέθοδος η οποία προσεγγίζει το πρόβλημα από μια σκοπιά που συνδυάζει την γεωμετρική γνώση με την εκπαίδευση μιας συνάρτησης (regression) για την εκτίμηση της πόζας του ανθρώπινου προσώπου. Η μέθοδος βασίζεται στον εντοπισμό δύο καθολικών σημείων του κεφαλιού (landmarks), το κέντρο του προσώπου, και το κεντροειδές (κέντρο μάζας) του συνολικού κεφαλιού. Εφόσον αυτά τα σημεία είναι καθολικά, μπορούν να ανιχνευτούν εύρωστα ακόμη και σε περιπτώσεις χαμηλής ανάλυσης και εκτεταμένης επικάλυψης, γεγονός που καθιστά αυτή την μέθοδο ιδιαίτερα επιτυχημένη για το συγκεκριμένο πρόβλημα. Η μέθοδος που θα περιγραφεί υποθέτει μηδενικές γωνίες Roll, οι οποίες είναι άλλωστε συνήθως αμελητέες στην περίπτωση της νοηματικής γλώσσας.

4.3.2 Ψυχοφυσική και γεωμετρική ερμηνεία της ιδέας

Έχει παρατηρηθεί ότι ο άνθρωπος εγκέφαλος αντιλαμβάνεται την μεταβολή της πόζας του ανθρώπινου προσώπου από την σχετική θέση της στενής περιοχής του προσώπου ως προς το συνολικό κεφάλι [8]. Η μέθοδος που θα αναπτυχθεί σε αυτό το κεφάλαιο εκμεταλλεύεται αυτή την διαισθητική αντίληψη για την ανάπτυξη μιας καινοτόμας προσέγγισης για την συνεχή εκτίμηση των γωνιών Yaw και Pitch του ανθρώπινου προσώπου σε ακολουθίες εικόνων της νοηματικής. Η μέθοδος αυτή, αν και σχετικά απλή στην υλοποίησή της, αποδεικνύεται αρκετά ακριβής και αξιόπιστη, ιδιαίτερα σε περιπτώσεις επικάλυψης με τα χέρια, όπου πιο πολύπλοκες μέθοδοι αποτυγχάνουν.



Σχήμα 4.10: Οπτική αναπαράσταση των διανυσμάτων v , $v1$ και $v2$ πάνω στο ανθρώπινο πρόσωπο.

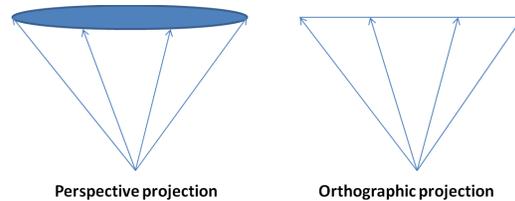
Στο Σχήμα 4.10 παριστάνεται μια άποψη του ανθρώπινου κεφαλιού υπό περιστροφή, όπου έχει σημειωθεί το κέντρο του προσώπου, καθώς και το κέντρο μάζας του κεφαλιού. Το διάνυσμα v είναι το διάνυσμα του τρισδιάστατου χώρου από το κέντρο μάζας στο κέντρο του προσώπου, ενώ $v1$ και $v2$ είναι οι συνιστώσες της προβολής του πάνω στην δισδιάστατη εικόνα. Αν θεωρήσουμε μηδενική γωνία στρέψης Roll, και περιορίσουμε την στρέψη σε δύο βαθμούς ελευθερίας (Yaw και Pitch), τότε υποθέτοντας ορθογραφική προβολή μπορούμε να εκτιμήσουμε άμεσα τις δύο αυτές γωνίες από τα διανύσματα v , $v1$ και $v2$ από τους παρακάτω τύπους:

$$\theta_{Yaw} = \arcsin\left(\frac{v2}{\|v\|}\right) \quad (4.2)$$

$$\theta_{Pitch} = \arcsin\left(\frac{v1}{\|v\|}\right) \quad (4.3)$$

Η υπόθεση ορθογραφικής προβολής εγγυάται την ανεξαρτησία των διανυσμάτων $v1$ και $v2$ (βλ. Σχήμα 4.11), και απλοποιεί την εκτίμηση των γωνιών στις παραπάνω απλές εξισώσεις. Αν και η παραπάνω ιδέα για την εκτίμηση της πόζας φαίνεται απλούστατη στην εφαρμογή της, και δίνει έναν κλειστό τύπο για τις γωνίες Yaw και Pitch, στην ουσία είναι αδύνατον να εφαρμοστεί ακριβώς, καθώς είναι αδύνατον να ανιχνευθεί το κέντρο μάζας του κεφαλιού,

όπως και το διάνυσμα του τρισδιάστατου χώρου v . Μια μέθοδος που επιδιώκει να προσεγγίσει αυτή την ιδέα προτείνεται στην επόμενη ενότητα.



Σχήμα 4.11: Οπτικοποίηση προοπτικής και ορθογραφικής προβολής διανύσματος.

4.3.3 Περιγραφή της μεθόδου

Όπως περιγράφηκε παραπάνω η απλή γεωμετρική ιδέα που προτάθηκε, και εκτιμά τις γωνίες Yaw και Pitch μέσω δύο κλειστών τύπων είναι αδύνατον να υλοποιηθεί ακριβώς, εφόσον δεν μπορεί να βρεθεί το κέντρο μάζας του κεφαλιού, ή το τρισδιάστατο διάνυσμα v . Σε αυτή την ενότητα προτείνεται μια υβριδική μέθοδος, που επιδιώκει να προσεγγίσει την προηγούμενη ιδέα. Το κέντρο μάζας του κεφαλιού, προσεγγίζεται από το κεντροειδές του κεφαλιού, ενώ η απουσία γνώσης σχετικά με το τρισδιάστατο διάνυσμα v αντισταθμίζεται με την εκπαίδευση μιας συνάρτησης αντιστοίχισης των διανυσμάτων του επιπέδου στις γωνίες στρέψης, από διαφορετικά πρόσωπα σε προσημειωμένες πόζες.

Η μέθοδος μπορεί να χωριστεί σε δύο μέρη. Στο πρώτο μέρος εξάγονται τα δύο σημεία ενδιαφέροντος από την εικόνα του κεφαλιού, και προσδιορίζεται το δισδιάστατο διάνυσμα που τα ενώνει. Σε ένα δεύτερο βήμα, το διάνυσμα αυτό μεταφράζεται στις γωνίες Yaw και Pitch μέσω της συνάρτησης που εκπαιδεύτηκε offline από πρόσωπα σε προσημειωμένες πόζες. Τα δύο αυτά βήματα περιγράφονται στις επόμενες υποενότητες.

Εντοπισμός των σημείων ενδιαφέροντος και εξαγωγή του χαρακτηριστικού διανύσματος (V_{Yaw}, V_{Pitch})

Στο Σχήμα 4.12 φαίνεται το κεφάλι με σημειωμένα τα σημεία ενδιαφέροντος. Το κέντρο του προσώπου (F_x, F_y) αναχτάται εύκολα, ως το κέντρο της περιοχής που επιστρέφεται από την συνεχή ανίχνευση προσώπου που περιγράφηκε στο Κεφάλαιο 2. Για να προσεγγιστεί το κέντρο μάζας (M_x, M_y), ανιχνεύεται πρώτα το κεντροειδές του κεφαλιού (C_x, C_y).

Το κεντροειδές του κεφαλιού εκτιμάται ως το κέντρο τεσσάρων σημείων, δηλαδή του κατώτατου μέρους του προσώπου, και των άνω, δεξιά και αριστερά ακρότατων του προσώπου, όπως φαίνεται στο σχήμα. Το κατώτατο μέρος του προσώπου αναχτάται και εδώ από την ανίχνευση του προσώπου, ενώ τα υπόλοιπα ακρότατά σημεία του κεφαλιού αναχτώνται, σε αυτή την υλοποίηση, μετά από κατωφλίωση και ανίχνευση των ακμών μέσω του αλγορίθμου Canny. Για την ανάκτηση των σημείων αυτών σε κάποια πιο δύσκολη εφαρμογή θα μπορούσε



Σχήμα 4.12: Σημεία ενδιαφέροντος πάνω στο πρόσωπο.

να χρησιμοποιηθεί κάποιος πιο πολύπλοκος αλγόριθμος Background Substraction ή Active Countors για τον συνεχή εντοπισμό του περιγράμματος του κεφαλιού.

Αφού προσεγγιστεί το κεντροειδές του κεφαλιού (C_x, C_y) , το κέντρο μάζας προσεγγίζεται μέσω των σχέσεων $M_x = C_x$ και $M_y = C_y - N$, όπου η παράμετρος κανονικοποίησης N ορίζεται ως η διαφορά $C_y - F_y$ όταν το πρόσωπο βρίσκεται στην μηδενική πόζα. Αφού προσδιοριστούν τα σημεία (F_x, F_y) και (M_x, M_y) , ορίζεται το δισδιάστατο διάνυσμα απόστασης τους ως:

$$v = (v_x, v_y) = (C_x - M_x, C_y - M_y) \quad (4.4)$$

Τα διανύσματα v_x και v_y κανονικοποιούνται με σταθερές όπως το πλάτος του προσώπου και η παράμετρος κανονικοποίησης N ώστε να απαλειφθεί ο παράγοντας απόστασης του προσώπου από την κάμερα, και διαφοροποιήσεις λόγω της γεωμετρίας του προσώπου:

$$V_{Yaw} = \frac{v_x}{width(head)} \quad (4.5)$$

$$V_{Pitch} = \frac{v_y}{N} \quad (4.6)$$

Τα νέα κανονικοποιημένα διανύσματα V_{Yaw} και V_{Pitch} είναι μηδέν στην ευθεία πόζα, και κυμαίνονται μέσα στο διάστημα $(-1,1)$ για μικρές γωνίες στρέψης.

Εκπαίδευση συναρτήσεων για την μετάφραση στις γωνίες Yaw και Pitch

Για την συνεχή αντιστοίχιση των κανονικοποιημένων διανυσμάτων V_{Yaw} και V_{Pitch} στις πραγματικές γωνίες θ_{Yaw} και θ_{Pitch} εκπαιδεύονται δύο συναρτήσεις, μία για κάθε στρέψη, από εικόνες προσώπων σε γνωστή προσημειωμένη πόζα.

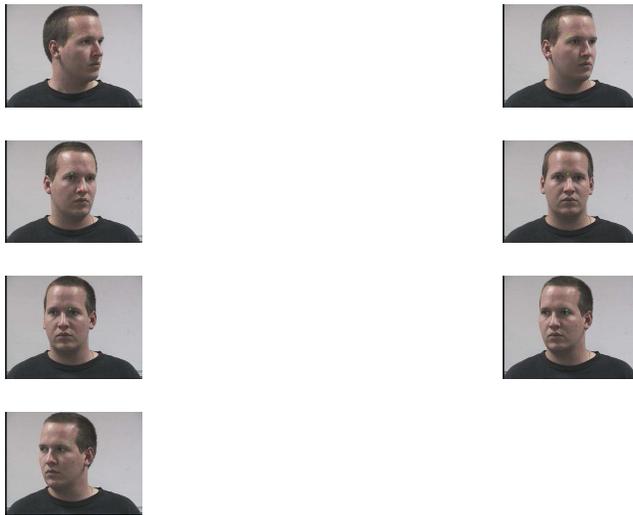
Για την εκπαίδευση των συναρτήσεων αντιστοίχισης, χρησιμοποιήθηκε μια βάση εικόνων από τέσσερα διαφορετικά πρόσωπα, υπό γνωστές προσημειωμένες πόζες, οι οποίες ανακτήθηκαν από την βάση δεδομένων INRIA [18]. Στις εικόνες αυτές, σημειώθηκε το κέντρο του

προσώπου και το κέντρο μάζας, όπως αυτά ορίστηκαν στην προηγούμενη ενότητα (βλ. Σχήμα 4.13) και ανακτήθηκαν τα κανονικοποιημένα διανύσματα V_{Yaw} και V_{Pitch} . Από τα δεδομένα αυτά, και με την βοήθεια γραμμικής παλινδρόμησης ελαχίστων τετραγώνων, εκπαιδεύτηκε μια γραμμική συνάρτηση για κάθε έναν από τους δύο βαθμούς ελευθερίας Yaw και Pitch, η οποία αντιστοιχεί τα κανονικοποιημένα διανύσματα V_{Yaw} και V_{Pitch} στις πραγματικές γωνίες στρέψης θ_{Yaw} και θ_{Pitch} . Χρησιμοποιείται απλή γραμμική παλινδρόμηση, υπο την παραδοχή ότι για μικρές γωνίες στρέψης το ημίτονο που θεωρητικά αντιστοιχεί το διάνυσμα στην γωνία (Εξισώσεις 4.2 και 4.3) μπορεί να προσεγγιστεί από μια γραμμική σχέση:

$$\theta_{Yaw} = a_1 V_{Yaw} \quad (4.7)$$

$$\theta_{Pitch} = a_2 V_{Pitch} \quad (4.8)$$

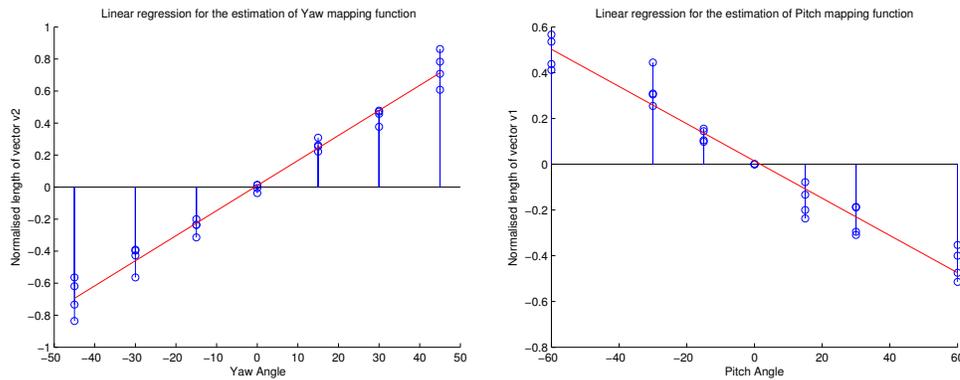
Όπου οι συντελεστές a_1 και a_2 ανακτώνται από την γραμμική παλινδρόμηση ελαχίστων τετραγώνων (βλ. Σχήμα 4.14).



Σχήμα 4.13: Κέντρο προσώπου και κεντροειδές κεφαλιού υπο μεταβολή της πόζας $[-45^\circ, 45^\circ]$

4.3.4 Αποτελέσματα και προτεινόμενες επεκτάσεις

Συνοψίζοντας σε αυτή την ενότητα προτείνεται μια καινοτόμος μέθοδος για την συνεχή εκτίμηση της πόζας του ανθρώπινου προσώπου, η οποία βασίζεται στην συνεχή ανίχνευση του στενού πλαισίου του προσώπου που περιγράφηκε στο Κεφάλαιο 3. Η μέθοδος αυτή υπερέρχει άλλων γνωστών μεθόδων στο γεγονός ότι είναι πολύ εύρωστη στην επικάλυψη, και σε μικρές αναλύσεις, καθώς βασίζεται στον εντοπισμό καθολικών σημείων του ανθρώπινου κεφαλιού, παρά μεμονωμένων χαρακτηριστικών σημείων που απαιτούν λεπτομέρεια. Τα μειονεκτήματα



Σχήμα 4.14: Συναρτήσεις αντιστοίχισης των κανονικοποιημένων διανυσμάτων V_{Yaw} και V_{Pitch} στις γωνίες Yaw και Pitch

της μεθόδου είναι ότι αγνοεί την στρέψη Roll, ενώ απαιτεί αρχικοποίηση στην μηδενική πόζα για την εκτίμηση του διανύσματος κανονικοποίησης N . Επίσης, καθώς βασίζεται στην συνεχή ανίχνευση του προσώπου, δεν μπορεί να εφαρμοστεί σε μεμονωμένες εικόνες, παρά μόνο σε ακολουθίες βίντεο.

Στο επόμενο σχήμα φαίνονται κάποια αποτελέσματα εφαρμογής αυτής της μεθόδου σε ακολουθίες εικόνων της νοηματικής. Η οπτική αναπαράσταση των αποτελεσμάτων δείχνει ότι η μέθοδος είναι αρκετά επιτυχημένη, και οι γωνίες που ανιχνεύονται συμβαδίζουν με την ανθρώπινη διάθεση. Για την πλήρη αξιολόγηση της μεθόδου, εφόσον αυτή βασίζεται στην συνεχή ανίχνευση και δεν μπορεί να εφαρμοστεί σε μεμονωμένες εικόνες, απαιτείται μια ακολουθία βίντεο με προσημειωμένες γωνίες στρέψης του προσώπου, κάτι που είναι αρκετά δύσκολο να κατασκευαστεί. Παρ' όλα αυτά οι συναρτήσεις παλινδρόμησης που παρουσιάζονται στο Σχήμα 4.14, προτείνουν ότι το λάθος εκτίμησης, δεδομένης της σωστής ανίχνευσης των δύο χαρακτηριστικών σημείων, κυμαίνεται μέχρι 5° για το Yaw και μέχρι 10° για το Pitch. Αυτό δείχνει ότι η μέθοδος είναι αρκετά ακριβής, και συγκρίσιμη με άλλες μεθόδους ανίχνευσης της πόζας.

Η καινοτόμος ιδέα που περιγράφηκε σε αυτό το κεφάλαιο δεν αποτελεί μια πλήρη μέθοδο για την ανίχνευση της πόζας, παρά μια πρώτη προσέγγιση η οποία όμως φαίνεται αρκετά επιτυχής, ιδιαίτερα σε περιπτώσεις χαμηλής ανάλυσης και επικάλυψης όπως η παρούσα. Για να αποτελέσει ολοκληρωμένη μέθοδο θα πρέπει να γίνει πιο εμπειριστατωμένη μελέτη, τόσο στην εξαγωγή των χαρακτηριστικών σημείων, όσο και στην εκπαίδευση της συνάρτησης αντιστοίχισης του διανύσματος σε γωνίες. Όσον αφορά την ανίχνευση του κεντροειδούς του κεφαλιού, μια μελλοντική εργασία θα μπορούσε να επιχειρήσει κάποια πιο ολοκληρωμένη μέθοδο αφαίρεσης φόντου ή ενεργών καμπυλών, καθώς και ένα στοχαστικό πλαίσιο συνεχούς ανίχνευσης του κεντροειδούς, το οποίο θα περιλαμβάνει πρότερη γνώση σχετικά με την γεωμετρία του κεφαλιού. Όσον αφορά τις συναρτήσεις αντιστοίχισης των διανυσμάτων στις γωνίες στρέψης, θα πρέπει να γίνει μια πιο εμπειριστατωμένη μελέτη με πολύ μεγαλύτερο δείγμα εκπαίδευσης, ώστε να μελετηθεί κατά πόσον η μέθοδος επηρεάζεται, παρά τις κανονικοποιήσεις,

από τις διαφορετικές γεωμετρίες προσώπων. Επίσης, οι συναρτήσεις αντιστοίχισης των χαρακτηριστικών διανυσμάτων σε γωνίες, θα μπορούσαν να εκπαιδευτούν με βάση το Γενικευμένο Γραμμικό Μοντέλο (Generalized Linear Model) με συνάρτηση συνδέσμου (link function) το ημίτονο, ώστε να συμβαδίζουν με την γεωμετρική υπόθεση. Τέλος η μέθοδος θα πρέπει να αξιολογηθεί πάνω σε ακολουθίες βίντεο με γνωστή πόζα σε κάθε εικόνα.



Σχήμα 4.15: Yaw = 26 , Pitch = -23



Σχήμα 4.16: Yaw = 20 , Pitch = -10



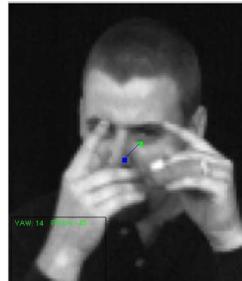
Σχήμα 4.17: Yaw = 26 , Pitch = 15



Σχήμα 4.18: Yaw = 22 , Pitch = 9



Σχήμα 4.19: Yaw = -17 , Pitch = -9



Σχήμα 4.20: Yaw = 14 , Pitch = -45

Σχήμα 4.21: Αποτελέσματα ανίχνευσης της πόζας (σε μοίρες) σε ακολουθίες εικόνων της νοηματικής

Κεφάλαιο 5

Παράρτημα

5.1 Στατιστική Ερμηνεία του Αλγορίθμου AdaBoost

Οι Friedman et al. απέδειξαν [10] ότι ο αλγόριθμος AdaBoost αποτελεί ένα προσθετικό μοντέλο λογιστικής παλινδρόμησης (Additive Logistic Regression Model), γεγονός στο οποίο αποδίδεται η υψηλή ικανότητα του για γενίκευση. Σε αυτή την παράγραφο επιδιώκουμε να δώσουμε μια ενόραση σε αυτή την στατιστική ερμηνεία του AdaBoost, καθώς και στην απόδειξη των Friedman et al., χωρίς όμως να προχωρήσουμε σε μαθηματικές λεπτομέρειες. Για την πλήρη απόδειξη και περαιτέρω λεπτομέρειες παραπέμπουμε στην αντίστοιχη δημοσίευση [10].

5.1.1 Προσθετικό Μοντέλο

Η απόκριση του αλγορίθμου AdaBoost δεν είναι παρά το σταθμισμένο άθροισμα των αδύναμων classifiers:

$$F(x) = c_1 f_1 + \dots + c_n f_n$$

Αυτό αποτελεί ουσιαστικά ένα αθροιστικό μοντέλο όπου μια πολύπλοκη συνάρτηση $F(x)$ προσδιορίζεται ως το άθροισμα απλών συναρτήσεων $f_i(x)$ με κατάλληλους συντελεστές c_i .

5.1.2 Λογιστική Παλινδρόμηση

Οι συντελεστές $c_i = \log\left(\frac{1-e_i}{e_i}\right)$ που επιλέγει ο AdaBoost σε κάθε βήμα έχουν την ιδιότητα να ελαχιστοποιούν το εκθετικό κριτήριο $e^{y(F(x)+c_i f_i)}$ σε κάθε επανάληψη i . Αυτό αποδεικνύεται από τους Friedman et al. στην αντίστοιχη δημοσίευση, με την αντίστροφη διαδικασία, δηλαδή αναζητώντας έναν συντελεστή c τέτοιο ώστε να ελαχιστοποιείται το κριτήριο $e^{y(F(x)+c f_i)}$, $y \in \{-1, 1\}$, κατά την διαδικασία εφαρμογής ενός προσθετικού μοντέλου σε μια άγνωστη συνάρτηση, προκύπτει ακριβώς ο συντελεστής c_i που επιλέγει ο AdaBoost.

Τέλος αποδεικνύεται ότι η συνάρτηση που ελαχιστοποιεί το εκθετικό κριτήριο $J(F) = e^{yF(x)}$, δεν είναι άλλη από την $F(x) = \frac{1}{2} \log\left(\frac{P(y=1|x)}{P(y=-1|x)}\right)$ και επομένως:

$$P(y=1|x) = p(x) = \frac{e^{F(x)}}{e^{-F(x)} + e^{F(x)}} = \frac{e^{2F(x)}}{1 + e^{2F(x)}}$$

Η συνάρτηση της πιθανότητας είναι η λογιστική (Logit) συνάρτηση σε δεύτερη δύναμη, και επομένως ο AdaBoost σε κάθε ανανέωση ελαχιστοποιεί το κριτήριο $E(e^{-yF(x)})$, που είναι ουσιαστικά μια βηματική διαδικασία για την εκπαίδευση ενός αθροιστικού μοντέλου λογιστικής παλινδρόμησης.

5.1.3 Συμπεράσματα

Συνοψίζοντας, στον αλγόριθμο AdaBoost, τόσο η ανανέωση των πιθανοτικών κατανομών όσο και οι συντελεστές που αποδίδονται στους αδύναμους classifiers για την παραγωγή του τελικού ισχυρού, δεν είναι τυχαίοι. Η ανανέωση των πιθανοτικών κατανομών είναι τέτοια ώστε αν εφαρμοστεί πάνω τους ο προηγούμενος αδύναμος classifier θα δώσει μόνο 50% σωστή κατηγοριοποίηση. Επομένως τα βάρη ανανεώνονται έτσι ώστε να κάνουν το πρόβλημα της κατηγοριοποίησης μέγιστα δύσκολο για τον επόμενο αδύναμο classifier. Επίσης, οι συντελεστές που αποδίδονται στους αδύναμους classifiers για την κατασκευή του τελικού ισχυρού, είναι τέτοιοι ώστε να ελαχιστοποιείται το εκθετικό κριτήριο $J(F) = e^{yF(x)}$, και γι' αυτό τον λόγο ο αλγόριθμος AdaBoost προσομοιώνει ένα αθροιστικό μοντέλο λογιστικής παλινδρόμησης (additive logistic regression model) της συνάρτησης πιθανότητας του αποτελέσματος y , από τις επιμέρους συναρτήσεις των αδύναμων υποθέσεων.

5.2 Εκτίμηση των γωνιών Euler από τρισδιάστατα σημεία του χώρου (landmarks)

Η τρισδιάστατη περιστροφή ενός συστήματος αξόνων (E_1, E_2, E_3) ώστε να συμπίσει με το σύστημα αξόνων (e_1, e_2, e_3) μπορεί να προσδιοριστεί μέσω των γωνιών Euler, οι οποίες εξαρτώνται από τους άξονες περιστροφής που έχουμε επιλέξει και την σειρά τους. Ανάλογα με τις γωνίες Euler που επιλέγουμε μπορούμε να προσδιορίσουμε τον πίνακα περιστροφής Euler:

$$Q = \begin{bmatrix} f_1(\phi, \psi, \theta) & f_2(\phi, \psi, \theta) & f_3(\phi, \psi, \theta) \\ f_4(\phi, \psi, \theta) & f_5(\phi, \psi, \theta) & f_6(\phi, \psi, \theta) \\ f_7(\phi, \psi, \theta) & f_8(\phi, \psi, \theta) & f_9(\phi, \psi, \theta) \end{bmatrix}$$

Από τα τρισδιάστατα σημεία X (μηδενικές γωνίες) και X' (αυθαίρετες γωνίες που θέλουμε να προσδιορίσουμε) μπορούμε να προσδιορίσουμε άμεσα τον πίνακα Q μέσω της γραμμικής μεθόδου ελαχίστων τετραγώνων και έπειτα να λύσουμε το υπερπροσδιορισμένο σύστημα ώστε να αποκτήσουμε τις Euler γωνίες περιστροφής. Ο πίνακας περιστροφής Q μπορεί να βρεθεί μέσω των 3D συνιστωσών των σημείων μέσω της σχέσης:

$$Q = X'X^T(XX^T)^{-1} \Rightarrow \begin{bmatrix} f_1(\phi, \psi, \theta) & f_2(\phi, \psi, \theta) & f_3(\phi, \psi, \theta) \\ f_4(\phi, \psi, \theta) & f_5(\phi, \psi, \theta) & f_6(\phi, \psi, \theta) \\ f_7(\phi, \psi, \theta) & f_8(\phi, \psi, \theta) & f_9(\phi, \psi, \theta) \end{bmatrix} = \begin{bmatrix} y_1 & y_2 & y_3 \\ y_4 & y_5 & y_6 \\ y_7 & y_8 & y_9 \end{bmatrix}$$

Εφόσον όμως δεν ενδιαφερόμαστε για τον πίνακα περιστροφής πάρα μόνο για τις γωνίες περιστροφής μπορούμε να τις προσδιορίσουμε άμεσα μέσω της μη γραμμικής μεθόδου ελαχίστων

τετραγώνων η οποία μετά από διαδοχικές επαναλήψεις συγχλίνει στις γωνίες περιστροφής.

Αρχικά επιλέγουμε κάποιες αυθαίρετες τιμές για τις γωνίες $(\phi_0, \psi_0, \theta_0)$ και υπολογίζουμε τις τιμές των συναρτήσεων f_1, \dots, f_9 . Ορίζουμε το σφάλμα $\beta_i = y_i - f_i$ και από αυτό το σφάλμα παίρνουμε την γραμμική εκτίμηση για την μεταβολή των γωνιών $d\lambda = (d\phi, d\psi, d\theta)$ μέσω των σχέσεων:

$$d\beta_i = \frac{\partial f}{\partial \phi} d\phi + \frac{\partial f}{\partial \psi} d\psi + \frac{\partial f}{\partial \theta} d\theta$$

Οι παραπάνω εξισώσεις γράφονται στην μορφή:

$$d\beta_i = A_{ij} d\lambda_j$$

με τον 9×3 πίνακα

$$A = \begin{bmatrix} \frac{\partial f_1}{\partial \phi} & \frac{\partial f_1}{\partial \psi} & \frac{\partial f_1}{\partial \theta} \\ \frac{\partial f_2}{\partial \phi} & \frac{\partial f_2}{\partial \psi} & \frac{\partial f_2}{\partial \theta} \\ \dots & \dots & \dots \\ \frac{\partial f_9}{\partial \phi} & \frac{\partial f_9}{\partial \psi} & \frac{\partial f_9}{\partial \theta} \end{bmatrix}$$

και επομένως μπορούμε να βρούμε τις διορθώσεις των γωνιών $d\lambda = (d\phi, d\psi, d\theta)$ λύνοντας το σύστημα:

$$A^T A d\lambda = A^T d\beta$$

Ακολουθώντας αυτή την διαδικασία επαναληπτικά, τελικά οι γωνίες συγχλίνουν σε αυτές που επιλύουν το σύστημα, και αποτελούν τις γωνίες περιστροφής του αρχικού συστήματος ώστε να συμπέσει το νέο.

5.3 Ανάλυση σε Πρωτεύουσες Συνιστώσες

Ο μετασχηματισμός ανάλυσης πρωτεύουσων συνιστωσών (Principal Component Analysis) αποτελεί έναν ορθογώνιο μετασχηματισμό που μεταφέρει ένα σύνολο από συσχετισμένες παρατηρήσεις σε έναν χώρο όπου αυτές παρουσιάζονται ασυσχέτιστες. Η διάσταση του νέου χώρου είναι μικρότερη ή ίση από το πλήθος των αρχικών μεταβλητών. Ο μετασχηματισμός ορίζεται έτσι ώστε η μεταβλητότητα των δεδομένων να συγκεντρώνεται στις πρώτες συνιστώσες του νέου χώρου.

Έστω ότι έχουμε k δείγματα δεδομένων $\{x_1, x_2, \dots, x_k\}$ διάστασης n , με μέσο:

$$x_{mean} = \frac{1}{k} \sum_{i=1}^k x_i$$

Ο πίνακας συμμεταβλητότητας του συνόλου αυτού είναι:

$$S = \sum_{i=1}^k (x_i - x_{mean})(x_i - x_{mean})^T$$

Ο χώρος προβολής που ορίζει ο μετασχηματισμός PCA κατασκευάζεται από τα ιδιοδιανύσματα (Eigenvectors) e του πίνακα συμμεταβλητότητας S :

$$Se = \lambda e$$

Αποδεικνύεται ότι σε αυτόν τον χώρο η ενέργεια των δειγμάτων συγκεντρώνεται στις πρώτες συνιστώσες, ενώ τα δείγματα είναι πλέον ασυσχέτιστα. Για περισσότερες πληροφορίες και αποδείξεις σχετικά με τον μετασχηματισμόν Πρωτεύουσων Συνιστωσών παραπέμπουμε στο σύγγραμμα [7].

5.4 Active Shape και Active Appearance Models

Τα **Active Shape Models** προτάθηκαν από τον Tim Cootes το 1994 [14] και αποτελούν ευέλικτους γράφους που μοντελοποιούν το σχήμα και προσαρμόζονται στα ανθρώπινα πρόσωπα. Για την εκπαίδευση τους σημειώνονται οι κόμβοι του γράφου σε ένα σύνολο από πρόσωπα. Το σχήμα του γράφου παραμετροποιείται ως το άθροισμα των πρώτων κύριων συνιστωσών που προκύπτουν από τον PCA με διαφορετικούς συντελεστές η κάθε μία. Επίσης αποθηκεύεται το πρότυπο της εμφάνισης της γειτονιάς κάθε χαρακτηριστικού ώστε να μπορεί να εντοπιστεί σε μια καινούργια εικόνα. Έτσι κατά την εφαρμογή του μοντέλου σε ένα νέο πρόσωπο, το σχήμα αναανώνεται διαδοχικά (ως γραμμικός συνδυασμός των κύριων συνιστωσών) ώστε το προφίλ κάθε κόμβου να ταιριάζει με το αποθηκευμένο πρότυπο.

Τα **Active Appearance Models** επεκτείνουν την ιδέα των ASM και μοντελοποιούν τόσο το σχήμα όσο και την εμφάνιση του ανθρώπινου προσώπου με την βοήθεια του PCA. Τα AAMs προτάθηκαν αρχικά από τον Tim Cootes το 1998 [5], αλλά επεκτάθηκαν και βελτιώθηκαν από τους Matthews και Baker ώστε να πάρουν την σημερινή τους μορφή το 2004 [17]. Σε αυτή την ενότητα περιγράφεται ο τρόπος με το οποίο τα ASMs και τα AAMs μοντελοποιούν το σχήμα και την εμφάνιση του γράφου προσώπου, ενώ δεν παρέχονται αναλυτικές πληροφορίες σχετικά με τους αλγόριθμους που χρησιμοποιούνται για την εφαρμογή τους στο πρόσωπο. Για περισσότερες πληροφορίες σχετικά με τα Active Shape και Active Appearance Models παραπέμπουμε στις αντίστοιχες δημοσιεύσεις [14] [5] [17].

5.4.1 Active Shape Models

Το σχήμα του γράφου προσώπου στην περίπτωση των ASMs ορίζεται από ένα υπερδιάνυσμα που περιλαμβάνει όλες τις συνιστώσες όλων των σημείων:

$$x = (x_1, x_2, \dots, x_n, y_1, y_2, \dots, y_n)$$

Ορίζοντας το μέσο σχήμα προσώπου ως:

$$x_{mean} = \frac{1}{k} \sum_{i=1}^k x_i,$$

και εφαρμόζοντας PCA στο σχήμα του γράφου βρίσκοντας τα ιδιοδιανύσματα ϕ_i και τις ιδιοτιμές λ_i του πίνακα συμεταβλητότητας:

$$S = \sum_{i=1}^k (x_i - x_{mean})(x_i - x_{mean})^T,$$

μπορούμε πλέον να μοντελοποιήσουμε το σχήμα του γράφου ως έναν γραμμικό συνδυασμό του μέσου σχήματος, και των πρωτεύουσων συνιστωσών:

$$x = x_{mean} + \Phi b$$

Το διάνυσμα b παραμετροποιεί το σχήμα του γράφου, ενώ η ιδιοτιμή λ_i κάθε ιδιοδιάστασης ϕ_i δείχνει την μεταβλητότητα της διάστασης αυτής στο σύνολο εκπαίδευσης. Μεταβάλλοντας κάθε διάσταση του διανύσματος b σε τιμές μέσα στο πεδίο $(-3\sqrt{\lambda_i}, 3\sqrt{\lambda_i})$ δημιουργούμε σχήματα αντίστοιχα με αυτά του συνόλου εκπαίδευσης.

5.4.2 Independent Active Appearance Models

Τα AAMs, όπως ειπώθηκε και παραπάνω, παραμετροποιούν τόσο το σχήμα όσο και την εμφάνιση μέσω του μετασχηματισμού PCA. Το σχήμα παραμετροποιείται όπως ακριβώς περιγράφηκε και στην περίπτωση των ASMs, ενώ η εμφάνιση του προσώπου παραμετροποιείται μόνο πάνω στο μέσο σχήμα x_{mean} . Οι γράφοι προσώπου σε αυτή την περίπτωση είναι τριγωνοποιημένοι, και επομένως κάθε πρόσωπο του συνόλου μπορεί να μετασχηματιστεί μέσω ενός τμηματικού αφινικού μετασχηματισμού (piecewise affine transform) στο μέσο σχήμα του. Επίσης κάθε άκαμπτη παραμόρφωση (rigid deformation) από το μέσο σχήμα απαλείφεται μέσω του μετασχηματισμού Procrustes Analysis.

Επομένως μετά από την προεπεξεργασία μέσω του μετασχηματισμού Procrustes Analysis και του τμηματικού αφινικού μετασχηματισμού, όλα τα πρόσωπα του συνόλου εκπαίδευσης μετασχηματίζονται στο μέσο σχήμα, και η εμφάνιση, που αποτελείται από τα pixels εντός του μέσου σχήματος παραμετροποιείται μέσω του μετασχηματισμού PCA στην μορφή:

$$A(x) = A_0(x) + \sum_{i=1}^m \lambda_i A_i(x)$$

όπου A_0 είναι η μέση εμφάνιση, ενώ $A_i(x)$ είναι οι ιδιοεμφανίσεις, ή ιδιοπρόσωπα (eigen-faces).

Εφόσον στα ανεξάρτητα AAMs το σχήμα και η εμφάνιση παραμετροποιούνται ξεχωριστά, μπορούμε να επεξεργαστούμε το σχήμα ανεξάρτητα για την εξαγωγή της πόζας. Επίσης, εφόσον πριν την παραμετροποίηση του σχήματος όλες οι άκαμπτες παραμορφώσεις έχουν εξαλειφθεί, αυτές δεν μοντελοποιούνται στις κύριες συνιστώσες του σχήματος. Μια τέτοια παραμόρφωση είναι η στρέψη Roll, η οποία μπορεί να υπολογιστεί ανεξάρτητα σε ένα πρώιμο βήμα κατά την εφαρμογή των AAMs στο ανθρώπινο πρόσωπο.

5.4.3 Combined Active Appearance Models

Τα συνδυασμένα AAMs παραμετροποιούν το σχήμα του γράφου και την εμφάνιση του προσώπου μαζί σε ένα μοναδικό διάνυσμα c , στο οποίο οφείλεται η συνδυασμένη μεταβολή τους. Μετά την εφαρμογή του PCA πάνω στο σχήμα του γράφου και στην εμφάνιση του προσώπου, για την ανάκτηση του διανύσματος b για την περιγραφή του σχήματος, και του διανύσματος λ για την περιοχή της εμφάνισης, τα δύο αυτά διανύσματα που περιγράφουν τα δεδομένα εκπαίδευσης συνδυάζονται σε ένα νέο υπερδιάνυσμα

$$s = (b_1, b_2, \dots, b_n, \lambda_1, \lambda_2, \dots, \lambda_n)$$

Στο διάνυσμα s εφαρμόζεται εκ νέου ο μετασχηματισμός PCA:

$$s = P_c c$$

Τέλος, το διάνυσμα c είναι το συνδυασμένο διάνυσμα που ελέγχει πλέον τόσο το σχήμα όσο και την εμφάνιση

Κεφάλαιο 6

Επίλογος

6.1 Σύνοψη και Συνεισφορά

Συνοψίζοντας, αυτή η εργασία μελετά τα προβλήματα πρώιμης Όρασης που σχετίζονται με την ανίχνευση του προσώπου και την εκτίμηση της πόζας του νοηματιστή, κατά την εκτέλεση νοηματικής γλώσσας. Η εργασία δεν ασχολείται με την περαιτέρω αξιοποίηση της σχετικής πληροφορίας για την εξαγωγή νοήματος από τα λεγόμενα του νοηματιστή, παρά μόνο με την εύρωστη και ακριβή εξαγωγή αυτής καθ' αυτής της πληροφορίας.

Τα προβλήματα που σχετίζονται με την ανίχνευση προσώπων, την συνεχή ανίχνευση βάσει στοχαστικής εκτίμησης, και την εκτίμηση της τρισδιάστατης πόζας του προσώπου, αποτελούν τρία από τα πιο διάσημα και κλασσικά προβλήματα της Όρασης Υπολογιστών. Το ανθρώπινο πρόσωπο αποτελεί ίσως το πιο αγαπημένο αντικείμενο επεξεργασίας της Όρασης, ενώ τεχνικές που αναπτύσσονται και εφαρμόζονται αρχικά στο ανθρώπινο πρόσωπο (όπως το πλαίσιο ανίχνευσης που μελετάται σε αυτή την εργασία), γενικεύονται συνήθως αργότερα για την επίλυση μιας πληθώρας προβλημάτων που αναφέρονται τόσο σε πραγματικές όσο και σε ιατρικές ή δορυφορικές εικόνες.

Η εργασία μελετά τα γενικευμένα προβλήματα της Όρασης που αφορούν στην ανίχνευση αντικειμένων, στοχαστική εκτίμηση για συνεχή ανίχνευση, και ανίχνευση της πόζας στην γενική τους μορφή, ενώ παράλληλα προσαρμόζει και υλοποιεί γνωστές μεθόδους, και προτείνει καινοτόμες λύσεις που απευθύνονται στις ιδιαιτερότητες της συγκεκριμένης εφαρμογής της νοηματικής γλώσσας.

Τελικά, η εργασία προσφέρει μια εμπειριστατωμένη μελέτη και υλοποίηση ενός καθολικού ενοποιημένου πλαισίου ανίχνευσης του προσώπου και εκτίμησης της πόζας το οποίο εφαρμόζεται και αξιολογείται πάνω σε βίντεο της νοηματικής γλώσσας. Τα αποτελέσματα δείχνουν ότι το πλαίσιο αυτό είναι ιδιαίτερα ακριβές και αξιόπιστο, ενώ επιτυγχάνει να ξεπεράσει τις προκλήσεις που παρουσιάζονται στο συγκεκριμένο πρόβλημα, και σχετίζονται κυρίως με την επικάλυψη του προσώπου από τα χέρια. Το πλαίσιο που προτείνεται, δεν περιορίζεται μόνο στο επιμέρους πρόβλημα της νοηματικής, αλλά μπορεί να γενικευτεί σε όλες τις περιπτώσεις όπου αντιμετωπίζονται αντίστοιχα χαρακτηριστικά όπως συχνή και εκτεταμένη επικάλυψη.

6.2 Κατευθύνσεις για Μελλοντική Έρευνα

Οι κατευθύνσεις για μελλοντική έρευνα προσανατολίζονται τόσο στα γενικότερα προβλήματα Όρασης Υπολογιστών που αντιμετωπίζονται σε αυτή την εργασία, όσο και στην αντιμετώπιση επόμενων προκλήσεων που αφορούν στην νοηματική γλώσσα. Οι κατευθύνσεις που προτείνονται, συνοψίζονται στα επόμενα σημεία:

- Όσον αφορά την ανίχνευση της πόζας, η ιδέα που προτείνεται και υλοποιείται δίνει ιδιαίτερα αξιόλογα αποτελέσματα και αντιμετωπίζει την πρόκληση της επιάλυψης πολύ καλύτερα από άλλες μεθόδους, δεν αποτελεί όμως ακόμη μια πλήρη και ολοκληρωμένη μέθοδο. Η ιδέα θα μπορούσε να εξελιχθεί τόσο στην κατεύθυνση της συνεχούς ανίχνευσης (tracking) των χαρακτηριστικών σημείων, όσο και στην μετάφραση των διανυσμάτων σε γωνίες, με την εκπαίδευση συναρτήσεων αντιστοίχισης από πολλά περισσότερα πρόσωπα με διαφορετικές γεωμετρίες. Τέλος η μέθοδος πρέπει να εφαρμοστεί πάνω σε ένα σύνολο συνεχόμενων εικόνων προσώπων με προσημειωμένη πόζα για την αξιολόγησή της και την σύγκρισή της με άλλες μεθόδους.
- Μια ακόμη ενδιαφέρουσα προοπτική είναι η μελέτη και ανάπτυξη ενός αντίστοιχου, βασισμένου-σε-εκπαίδευση (learning-based) πλαισίου για την ανίχνευση των χεριών. Εφόσον τα χέρια αποτελούν παραμορφώσιμα αντικείμενα, το συγκεκριμένο πλαίσιο με τα Haar χαρακτηριστικά θα αποτύγχανε, ο αλγόριθμος AdaBoost θα μπορούσε όμως να συνδυαστεί με άλλου είδους χαρακτηριστικά που συναντώνται στην βιβλιογραφία, και έχουν αναπτυχθεί ειδικά για παραμορφώσιμα αντικείμενα.
- Τέλος όσον αφορά αυτή καθ' αυτή την ανάλυση της νοηματικής γλώσσας, η ακριβής ανίχνευση του προσώπου θα μπορούσε να συνδυαστεί με την ανίχνευση των χεριών σε περιπτώσεις επιάλυψης, για την εξαγωγή της πληροφορίας σχετικά με το σημείο εναπόθεσης του χεριού πάνω στο πρόσωπο (place of articulation), η οποία κρύβει επίσης σημαντικό γλωσσικό νόημα. Επίσης, η θέση του προσώπου και η εκτίμηση της πόζας, θα μπορούσαν να χρησιμοποιηθούν σαν πρότερη γνώση για την πιο εύρωστη εφαρμογή ευέλικτων γράφων στο ανθρώπινο πρόσωπο, για την περαιτέρω ανάλυση της έκφρασής του.

Βιβλιογραφία

- [1] Simon Baker, Iain Matthews, Jing Xiao, Ralph Gross, Takahiro Ishikawa και Takeo Kanade. Real-time non-rigid driver head tracking for driver mental state estimation, 2004.
- [2] Volker Blanz και Thomas Vetter. A morphable model for the synthesis of 3d faces. Στο *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, SIGGRAPH '99, σελίδες 187–194, New York, NY, USA, 1999. ACM Press/Addison-Wesley Publishing Co.
- [3] Howie Choset, Kevin M. Lynch, Seth Hutchinson, George Kantor, Wolfram Burgard, Lydia E. Kavraki και Sebastian Thrun. *Principles of Robot Motion: Theory, Algorithms, and Implementations (Intelligent Robotics and Autonomous Agents)*. The MIT Press, 2005.
- [4] T. F. Cootes, G. V. Wheeler, K. N. Walker και C. J. Taylor. View-based active appearance models. *Image and Vision Computing*, 20(9-10):657 – 664, 2002.
- [5] Timothy F. Cootes, Gareth J. Edwards και Christopher J. Taylor. Active appearance models. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23(6):681–685, 2001.
- [6] Liya Ding και Aleix M. Martinez. Modelling and recognition of the linguistic components in american sign language. *Image and Vision Computing*, 27(12):1826–1844, 2009.
- [7] Richard O. Duda, Peter E. Hart και David G. Stork. *Pattern Classification (2nd Edition)*. Wiley-Interscience, 2η έκδοση, 2001.
- [8] IEEE Erik Murphy-Chutorian, Student Member και IEEE Mohan Manubhai Trivedi, Senior Member. Head pose estimation in computer vision: A survey. *IEEE TPAMI*, 31(4):607–627, 2009.
- [9] Yoav Freund και Robert E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. Στο *Proceedings of the Second European Conference on Computational Learning Theory*, σελίδες 23–37, London, UK, 1995. Springer-Verlag.

-
- [10] Jerome Friedman, Trevor Hastie και Robert Tibshirani. Additive logistic regression: a statistical view of boosting. *Annals of Statistics*, 28, 2000.
- [11] Etienne Grossmann. Adatree: Boosting a weak classifier into a decision tree. *Computer Vision and Pattern Recognition Workshop*, 6:105, 2004.
- [12] Jeffrey Huang, Xuhui Shao και Harry Wechsler. Face pose discrimination using support vector machines (svm). Στο *Proceedings of the 14th International Conference on Pattern Recognition-Volume 1 - Volume 1*, ICPR '98, σελίδες 154–, Washington, DC, USA, 1998. IEEE Computer Society.
- [13] Michael Isard και Andrew Blake. Condensation conditional density propagation for visual tracking. *International Journal of Computer Vision*, 29:5–28, 1998.
- [14] Andreas Lanitis, Christopher J. Taylor και Timothy F. Cootes. Automatic interpretation and coding of face images using flexible models. *IEEE Trans. Pattern Anal. Mach. Intell.*, 19(7):743–756, 1997.
- [15] Rainer Lienhart, Alexander Kuranov και Vadim Pisarevsky. Empirical analysis of detection cascades of boosted classifiers for rapid object detection. Στο *Pattern Recognition*, τόμος 2781 στο *Lecture Notes in Computer Science*, σελίδες 297–304. Springer Berlin / Heidelberg, 2003.
- [16] Rainer Lienhart και Jochen Maydt. An extended set of haar-like features for rapid object detection. 2002.
- [17] Iain Matthews και Simon Baker. Active appearance models revisited. *International Journal of Computer Vision*, 60(2):135–164, 2004.
- [18] J. L. Crowley N. Gourier, D. Hall. Estimating face orientation from robust detection of salient facial features. *Proceedings of Pointing , ICPR, International Workshop on Visual Observation of Deictic Gestures, Cambridge, UK*, 2004.
- [19] Carol Neidle, Vassilis Athitsos, Stan Sclaroff και Hermann Ney. Benchmark databases for video-based automatic sign language recognition, *χ.χ.*
- [20] Constantine P. Papageorgiou, Michael Oren και Tomaso Poggio. A general framework for object detection. *Computer Vision, IEEE International Conference on*, 0:555, 1998.
- [21] George Papandreou και Petros Maragos. Adaptive and constrained algorithms for inverse compositional active appearance model fitting. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 0:1–8, 2008.
- [22] E. K. Patterson, S. Gurbuz, Z. Tufekci και J. N. Gowdy. Cuave: A new audio-visual database for multimodal human-computer interface research. Στο *In Proc. ICASSP*, σελίδες 2017–2020, 2002.

-
- [23] Sujith Srinivasan και Kim L. Boyer. Head pose estimation using view based eigenspaces. *Pattern Recognition, International Conference on*, 4:40302, 2002.
- [24] Zhuowen Tu. Probabilistic boosting-tree: Learning discriminative models for classification, recognition, and clustering. *Computer Vision, IEEE International Conference on*, 2:1589–1596, 2005.
- [25] M. Viola, Michael J. Jones και Paul Viola. Fast multi-view face detection. Στο *Proc. of Computer Vision and Pattern Recognition*, 2003.
- [26] Paul Viola και Michael J. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57:137–154, 2004.
- [27] Ulrich von Agris, Jörg Zieren, Ulrich Canzler, Britta Bauer και Karl Friedrich Kraiss. Recent developments in visual sign language recognition. *Universal Access in the Information Society*, 6:323–362, 2008.
- [28] Greg Welch και Gary Bishop. An Introduction to the Kalman Filter. Τεχνική Αναφορά υπ. αριθμ. TP 95-041, 2004.

Κεφάλαιο 7

Μεταφράσεις Ξένων όρων

Μετάφραση

ταξινομητής
υπερεκπαίδευση
λογαριθμική πιθανοφάνεια
παλινδρόμηση
λογιστική παλινδρόμηση
παλινδρόμηση ελαχίστων τετραγώνων
ανίχνευση
συνεχής ανίχνευση
δένδρα αποφάσεων
σύνολο αξιολόγησης
χαρακτηριστικό
φωτισμός
ταίριασμα προτύπων
ιδιοδιανύσματα
ιδιοτιμές
ιδιοχώρος

Αγγλικός όρος

classifier
overfitting
log likelihood
regression
logistic regression
least squares regression
detection
tracking
decision tree
benchmark
feature
illumination
template matching
eigenvectors
eigenvalues
eigenspace

